

Washington University in St. Louis

Washington University Open Scholarship

McKelvey School of Engineering Theses & Dissertations

McKelvey School of Engineering

12-22-2023

Improving and Integrating Quantitative Single Cell Spatial Biology for Theranostic Medicine

Peng Lu

Washington University – McKelvey School of Engineering

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Recommended Citation

Lu, Peng, "Improving and Integrating Quantitative Single Cell Spatial Biology for Theranostic Medicine" (2023). *McKelvey School of Engineering Theses & Dissertations*. 996.
https://openscholarship.wustl.edu/eng_etds/996

This Dissertation is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in McKelvey School of Engineering Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Dissertation Examination Committee:

Daniel L.J. Thorek, Chair

Aimilia Gastounioti

Abhinav K. Jha

Stephen T. Oh

Joseph A. O'Sullivan

Improving and Integrating Quantitative Single Cell Spatial Biology
for Theranostic Medicine

by

Peng Lu

A dissertation presented to
the McKelvey School of Engineering
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2023
St. Louis, Missouri

© 2023, Peng Lu

Table of Contents

List of Figures.....	vii
List of Tables	xx
Acknowledgments.....	xxi
Abstract	xxiii
Chapter 1: Introduction.....	1
1.1 DAR and IMC in Theranostic Medicine	2
1.2 Objectives and Structure of This Dissertation	4
1.3 Dissemination of Research Findings in the Dissertation	6
Chapter 2: Dose Measurement and Microenvironment Analysis Techniques in Theranostic Medicine	8
2.1 DAR Imaging Technique.....	8
2.1.1 DAR imaging process	8
2.1.2 Imaging artifacts of DAR and the corresponding restoration techniques	9
2.2 IMC Imaging Technique	11
2.2.1 IMC imaging process and noise sources	11
2.2.2 IMC denoising techniques	15
2.3 Multiplexed Imaging Data Analysis Techniques	15
2.4 On-Tissue Dose Analysis Techniques	17
Chapter 3: PG-PEM: A Blind Image Restoration Algorithm to Enhance Digital Autoradiography	18
3.1 Introduction.....	18
3.2 PG-PEM Algorithm	19
3.2.1 DAR imaging model	19
3.2.2 Patch-based estimation of noise parameters	21

3.2.3	MLEM algorithm for the mixed Poisson-Gaussian model	23
3.2.4	Regularization for \mathbf{h} and \mathbf{X}	27
3.2.5	PSF model.....	30
3.2.6	The impact of the scaling factor α	31
3.2.7	PG-PEM algorithm summary	33
3.2.8	Runtime	33
3.3	Reference DAR Restoration Methods	35
3.3.1	RL algorithm	36
3.3.2	RD algorithm	36
3.3.3	SP algorithm	37
3.3.4	NP and TV algorithms	38
3.4	Accuracy Metrics and Statistical Analysis	38
3.5	Simulations.....	41
3.5.1	Simulated data generation.....	41
3.5.2	Impact of the regularization parameters	42
3.5.3	Comparison with reference algorithms	45
3.6	Experiments	48
3.6.1	Material preparation	48
3.6.2	Assessment of image enhancement on experimental data.....	49
3.6.3	PG-PEM improves DAR of diagnostic radiopharmaceuticals	51
3.6.4	Enhanced Targeted Alpha Particle Radiotherapy Evaluation by PG-PEM	54
3.7	Discussion	57
3.8	Conclusion.....	59
Chapter 4: IMC-Denoise: A Content-Aware Pipeline to Enhance Imaging Mass Cytometry		60
4.1	Introduction.....	60
4.2	IMC-Denoise Framework	61
4.2.1	DIMR algorithm for hot pixel removal	61
4.2.2	DeepSNiF for image quality enhancement.....	66
4.2.3	DeepSNiF network implementation	74

4.3	Reference IMC Denoising Methods	76
4.3.1	Hot pixel removal methods.....	76
4.3.2	Deep learning-based shot noise filtering methods.....	77
4.3.3	Traditional statistics-based shot noise filtering methods.....	80
4.4	Analyzing IMC Image Quality Enhancement through Simulation Verification	81
4.4.1	Simulated data generation.....	82
4.4.2	Accuracy metrics and statistical analysis	83
4.4.3	Hot pixel removal methods evaluation	84
4.4.4	Shot noise filtering methods evaluation	87
4.5	Experimental Verification of Enhanced Image Quality in IMC	93
4.5.1	Material preparation	93
4.5.2	Accuracy metrics and statistical analysis	95
4.5.3	DIMR removes hot pixels effectively in IMC images	95
4.5.4	DeepSNiF enhances image quality significantly in IMC	97
4.6	Experimental Verification of Enhanced Downstream Analysis in IMC	102
4.6.1	Reference methods in downstream analysis	102
4.6.2	Accuracy metrics and statistical analysis	103
4.6.3	IMC-Denoise enables background noise removal of IMC images and enhancement of IMC downstream analysis.....	105
4.6.4	DeepSNiF in IMC-Denoise enhances automated cell phenotyping	112
4.6.5	DeepSNiF in IMC-Denoise enhances lymphocyte analysis	117
4.7	Discussion	123
4.8	Conclusion.....	127

Chapter 5: InterSTELLAR: An Interpretable Spatial Cell Learning Framework to Enhance the Characterization of Highly Multiplexed Imaging Data.....	129
5.1 Introduction.....	129
5.2 InterSTELLAR Framework	130
5.2.1 Dataset description and pre-processing	130
5.2.2 Graph construction.....	131
5.2.3 Network structure	133

5.2.4	Model training	136
5.3	Accuracy Metrics and Statistical Analysis	137
5.4	Experiments	138
5.4.1	InterSTELLAR achieves accurate clinical type classification and cell-scale characterization	138
5.4.2	Attention Mapping by InterSTELLAR Across Cancer Tissue Types ...	143
5.4.3	InterSTELLAR captures tissue microenvironmental features from different clinical subtypes	145
5.4.4	InterSTELLAR uncovers single-cell pathology groups associated with patient survival	146
5.5	Discussion	152
5.6	Conclusion	154
Chapter 6: Development of New Theranostic Approaches by Integrating Multi-modality Imaging Data		155
6.1	Introduction	155
6.2	Multi-modality Image Integration Framework	156
6.2.1	DAR and H&E image registration framework	156
6.2.2	IMC and H&E image registration framework	161
6.3	Experiments	163
6.3.1	DAR and H&E image registration results	163
6.3.2	IMC and H&E image registration results	166
6.4	Discussion and Conclusion	167
Chapter 7: Conclusions and Future Work		168
7.1	Summary	168
7.2	Future Work	171
7.2.1	Validating PG-PEM with DAR from a silicon strip detector autoradiography system	171
7.2.2	Advancing IMC-Denoise for larger hot pixel clusters and better image quality	172
7.2.3	Validating InterSTELLAR with larger and other multiplexed imaging datasets	172

7.2.4 Validating and advancing data integration framework for DAR and IMC	173
References	174
Appendix A: Supplementary Tables for Chapter 4	[186]

List of Figures

Figure 2.1: DAR imaging process. (a) Expose the plate to the radioactive section to produce a latent image through a trapping of electron-hole. The latent image is formed by x-rays, gamma rays, beta particles, and alpha particles generated in the exposing process. Here, S is a point source and S' is a possible destination of S . d is the distance between the tissue section and the phosphor plate. In uv -plane, the coordinate of S and S' are (u_o, v_o) and (u', v') separately. (b) Scan a laser over the latent image on the phosphor plate to stimulate release of photons so as to generate the corresponding digital autoradiographic image. The pure signal is blurred mainly during the exposing process while contaminated by noise during the scanning process. 9

Figure 2.2: IMC workflow: Stable metal isotope-labeled antibodies are bound to molecular epitopes in tissue sample (fresh frozen, fixed or formalin-fixed paraffin-embedded) before rastering of an ablative UV laser with a spot size $< 1 \mu\text{m}$. The ionized material is coupled to a Hyperion time-of-flight (TOF) mass cytometer to detect mid-range metal masses ($> 80 \text{ Da}$). Unlike traditional microscopy, there is no background as the masses are detected in non-overlapping distinct channels (m/z). Finally, more than 40 channels corresponding to molecular specific biological features are acquired for the defined region of interest tissue slide. 12

Figure 3.1: Poisson distributions approximated as Gaussian distributions. (a) Poisson-distributed data (dots) for different values of photon flux λ_P and the corresponding Gaussian fitting functions $\mathcal{N}(\lambda_P, \lambda_P)$ (solid lines). (b) The relative error between the two distributions, in which σ_P and σ_G are the standard deviation of the Poisson- and Gaussian-distributed data, respectively. Based on these results, Poisson distribution can be feasibly approximated as a Gaussian distribution when $\lambda_P > 3$ 20

Figure 3.2:	Noise parameters estimation. (a) A raw DAR image. (b) The histogram of the raw image. (c) Split the raw image into multiple patches and calculate the mean, standard deviation, skewness and kurtosis of each patch. (d) Plot of the sorted (minPts-1)-th nearest distance of every patch. (e) DBSCAN results for the patches. (f) The histograms of the extracted background field which can be fitted by a Gaussian distribution; while (b) cannot be. The green dotted lines in (d) correspond to the ϵ used in (e). Scale bar: 2.3 mm.	22
Figure 3.3:	PG-PEM algorithmic framework summary. (1) Noise parameters estimation. (2) PSF and specimen image estimation. Scale bar: big figure: 2.3 mm, small figures: 0.54 mm.	34
Figure 3.4:	Effective resolution estimation by decorrelation analysis. (a) A raw DAR images with low noise level. (d) The corresponding PG-PEM restored image. (b) and (e) The corresponding decorrelation analysis of (a) and (d). (c) and (f) The log-scale frequency map of (a) and (d) labeled with cut-off frequency estimated from (b) and (e), respectively. Scale bar: 1.8 mm.	40
Figure 3.5:	Data generation for simulation and accuracy metrics for different parameters. (a) The simulated ground truth image and the corresponding PSF. (b) Blurred image with background. (c) Noisy image. (d) RMSE, SSIM and SNR for all the parameters. (e) RMSE, SSIM and SNR with different $\lambda_{\mathbf{h}}$ when $\lambda_{\mathbf{x}} = 0.001$. (f) RMSE, SSIM and SNR with different $\lambda_{\mathbf{x}}$ when $\lambda_{\mathbf{h}} = 1.2$. Note that in the figures the dots indicate the optimal values.	42
Figure 3.6:	Restoration images and corresponding PSFs using PG-PEM algorithm with $\lambda_{\mathbf{h}}$ from 0 to 10 and $\lambda_{\mathbf{x}}$ from 0 to 0.1. With $\lambda_{\mathbf{h}}$ increasing, the kernel size of the estimated PSF become larger and larger, and the restored images tend to become less noisy. But when the PSF kernel is large enough, the restored images become blurry. With $\lambda_{\mathbf{x}}$ increasing, the image noise is also suppressed. But with too large $\lambda_{\mathbf{x}}$, the images lose fine details and become blurry.	43
Figure 3.7:	Evaluation results for different parameters and methods when $\alpha = 20$ and 100. (a) RMSE evaluation when $\alpha = 20$. (b) SSIM evaluation when $\alpha = 20$. (c) SNR evaluation when $\alpha = 20$. (d) RMSE evaluation when $\alpha = 100$. (e) SSIM evaluation when $\alpha = 100$. (f) SNR evaluation when $\alpha = 100$. From the figures, PG-PEM is the best performer among all the conditions regarding RMSE, SSIM and SNR.	45
Figure 3.8:	Restoration results from different noise levels and methods when $\alpha = 20$	46

Figure 3.9:	Restoration results from different noise levels and methods when $\alpha = 20$ (log scale). It is easier to compare the denoising ability between different methods using the log-scale images.	47
Figure 3.10:	Blind restoration and background segmentation improve DAR image quality. (a) A raw DAR image from the mouse hindlimb following ^{18}F NaF PET imaging and its restoration results using modified restoration algorithms, PSF estimation and patch-based segmentation. The corresponding estimated PSFs are inset in grey scale. (b) Magnified images of the corresponding dashed boxed regions in (a). (c) Log-scale transformed images from (a) for background appraisal. (d) Log-scale amplitude of the Fourier transform of the raw and restored images in (a). Scale bar: (a) and (c): 4.95 mm, (b): 0.86 mm.....	50
Figure 3.11:	Quantitative evaluation of the DAR image quality enhancement. (a) Line profiles of the images in Figure 3.10a along the corresponding dashed lines. (b) STD, CNR and effective resolution assessment of the DAR images. Different colors correspond to raw and restored data. Box plot: Center line, median; limits, 75% and 25%; whiskers, maximum and minimum. DLU: digital light unit. We used the two-sided Student's t-test to compare the effective resolution of Raw and PG-PEM restored DAR images, and the one-way analysis of variance (ANOVA) to compare the STD and CNR of PG-PEM and all the other restoration algorithms. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$	52
Figure 3.12:	Content-adaptive blind restoration improves mice tumor and heart DAR. (a) and (b): Raw and PG-PEM restored DAR images of mice tumor and heart sections, which were treated with ^{18}F FDG. DLU: digital light unit. Scale bar: (a): 4.2 mm, (a1): 1.4 mm; (b): 3.7 mm, (b1): 0.75mm.	52
Figure 3.13:	Content-adaptive blind restoration improves mice femur DAR. From left to right, H&E stained histopathological, raw and PG-PEM restored DAR images of mice femur section, which was treated with ^{18}F -NaF. The images from the middle and bottom rows correspond to their labeled zoomed-in regions of the images from the top row. DLU: digital light unit. Scale bar: H&E: top: 5 mm, middle and bottom: 1.2 mm; Raw: top: 5 mm, middle and bottom: 1.2 mm.	53

Figure 3.14: STDB, CNR and effective resolution of the DAR images before and after restoration by PG-PEM on pre-clinical data. Every kind of tissue consists of 10 sections. Box plot: Center line, median; limits, 75% and 25%; whiskers, maximum and minimum. We used the two-sided Student's t-test to compare the data of Raw and PG-PEM restored DAR images in (h)–(j). **** $P < 0.0001$	54
Figure 3.15: PG-PEM restoration in clinical targeted alpha particle radiotherapy specimens. (a) H&E stained histopathological image of a bone biopsy section obtained from Radium-223 dichloride treated castrate resistant prostate cancer patient metastasis. (b) and (c) Raw and PG-PEM restored DAR image of the same representative section in (a). (d) The bone area was manually segmented and registered with raw and restored DAR based on their mutual information to produce (e) and (f) fused DAR and pathology images. In (b)–(f) magnified regions delineated by dashed lines are shown below the full-section image. Scale bar: (a) 1 mm. (b) upper image: 2.3 mm, lower images 0.5 mm.	55
Figure 3.16: Quantitative evaluation of the DAR image quality enhancement in clinical targeted alpha particle radiotherapy specimens. (a) Line profiles of raw and PG-PEM DAR data (as defined in Figure 3.15(b) and (c)). (b)–(d) STD, CNR and effective resolution assessment of DAR images. (e) and (f) Comparison of the accuracy of alignment of DAR and histopathology data for SSIM and fusion indices for raw and restored images ($N = 10$). DLU: digital light units. Box plot: Center line, median, limits, 75% and 25%; whiskers, maximum and minimum. We used the two-sided Student's t-test to compare the data of original and enhanced images in (b)–(f). **** $P < 0.0001$	56
Figure 3.17: PSF comparison. (a) The estimated PSF from a $^{223}\text{RaCl}_2$ treated human bone biopsy image. (b) The estimated PSF from a $^{18}\text{F-NaF}$ treated mouse hindlimb image. To increase the image contrast, we have applied a false coloured “jet” colormap. The estimated full-width half-maximum and full-width tenth-maximum of (a) are approximately 0.159 mm and 0.344 mm, while those of (b) are 0.216 mm and 0.66 mm, respectively. These results are consistent with the physics of alpha/positron transport: alpha particles have a significantly shorter path length than positrons. This is also why the raw autoradiographic images from ^{223}Ra treated human bone biopsy have higher resolution than those from ^{18}F treated sections. It should be noted that while several beta particles are produced by its daughters, the vast majority of the ^{223}Ra decay energy is in the form of alpha particles.	58
Figure 4.1: The schematic of the DIMR algorithm.	65

Figure 4.2:	The proposed DeepSNiF network structure.	75
Figure 4.3:	DIMR algorithm evaluations on the simulated DNA dataset ($N = 50$ per noise setting). (a) Evaluation of DIMR with N_{iter} from 1 to 5. “Same results” indicate the iteration stops beforehand. For instance, in noise setting 3, DIMR stops after the second iteration even though N_{iter} is larger than 2. (b) DIMR running time evaluations under different noise settings with $N_{\text{iter}} = 3$	85
Figure 4.4:	Evaluation of the hot pixel removal methods on the simulated CD14 dataset ($N = 50$ per noise setting). (a) RMSE comparison under different noise settings. (b) The optimal hyper-parameters of NTHM and MTHM under different noise settings. (c) Visual inspection of the three hot pixel removal methods on a CD14 image with the highest hot pixel density under noise setting 1. (d) Visual inspection of the DIMR algorithm under different noise settings. Scale bar: (c) $20 \mu\text{m}$, (d) $20 \mu\text{m}$	86
Figure 4.5:	Evaluation of the hot pixel removal methods on the simulated DNA dataset ($N = 50$ per noise setting). (a) RMSE comparison under different noise settings. (b) The optimal hyper-parameters of NTHM and MTHM under different noise settings. (c) Visual inspection of the three hot pixel removal methods on a DNA image with the highest hot pixel density under noise setting 1. (d) Visual inspection of the DIMR algorithm under different noise settings. Scale bar: (c) $20 \mu\text{m}$, (d) $20 \mu\text{m}$	88
Figure 4.6:	Comparison of DeepSNiF-NR with N2V and MN2V on the simulated CD14 dataset with 4 noise levels ($N = 25$ per level). (a) PSNR and SSIM evaluation on the three algorithms. (b) Visual inspection of the three algorithms and N2T on denoising a low SNR image. (c) Visual inspection of the three algorithms and N2T on denoising a very low SNR image. Scale bar: (b) $20 \mu\text{m}$, (c) $20 \mu\text{m}$	89
Figure 4.7:	The effect of Hessian norm regularization on the simulated CD14 image denoising ($N = 25$ per level). (a) PSNR and SSIM evaluation for DeepSNiF with different Hessian norm regularization parameters λ_{Hessian} . (b) Visual inspection of DeepSNiF with different Hessian norm regularization parameters λ_{Hessian} on a simulated CD14 image with medium SNR. (c) Visual inspection of DeepSNiF when $\lambda_{\text{Hessian}} = 0$ (NR) and $\lambda_{\text{Hessian}} = 3\text{e-}6$ on simulated CD14 images with different noise levels. Scale bar: (b) $20 \mu\text{m}$, (c) $20 \mu\text{m}$	90

Figure 4.8:	Comparison of DeepSNiF with GAUSS, NLM and BM3D algorithms on the simulated CD14 dataset with 4 noise levels ($N = 25$ per level). (a) PSNR and SSIM evaluation on the algorithms. (b) Visual inspection of the algorithms on denoising the images with different noise levels. Scale bar: $20 \mu\text{m}$	92
Figure 4.9:	Validation of DIMR on the human bone marrow IMC dataset. (a) The fractions of detected hot pixels by DIMR in selected channels. (b) DIMR removes hot pixels in DNA intercalator channel effectively. Left: Comparison of the raw and DIMR-processed images; and the difference between the images, in which Residual corresponds to the detected hot pixels. Upper right: The corresponding histograms of the raw and DIMR-processed images. Lower right: Comparisons between the raw, NTHM, MTHM and DIMR processed images. Scale bar: left: $75 \mu\text{m}$, right: $8 \mu\text{m}$	96
Figure 4.10:	DIMR slightly enhances single cell segmentation. (a) Comparison between the raw and DIMR-processed DNA images. (b) The probability maps of the raw and DIMR-processed images generated by the Ilastik software, and their difference map (blue: nuclei, green: cytoplasm, and red:background). (c) The segmentation masks of the raw and DIMR-processed images generated by Cell Profiler software, and their overlaid comparison map. By comparing the figures in the third column, the different segmented masks between raw and DIMR images are frequently caused by the presence of hot pixels. Even with a cell size threshold, the hot pixels can still split cells, and falsely expand normal cell borders. Correspondingly, the raw image segmented a little more cells than that of DIMR (906 to 994). At the same time, the average area of cells from the raw image is a little smaller (29.57 to 29.89), but the summed area is slightly larger (26790 to 26442). Scale bar: $47 \mu\text{m}$	97
Figure 4.11:	Validation of DeepSNiF on IMC images labeled with Collagen III. (a) Visual inspection of DeepSNiF and other statistics-based denoising algorithms. (b) DeepSNiF performs significantly better than other algorithms ($N = 12$) on denoising Collagen III-labeled IMC images in terms of STDB and CNR. Scale bar: $50 \mu\text{m}$	98
Figure 4.12:	Further inspection of DeepSNiF on experimental data. (a) Visual inspection of DeepSNiF denoised IMC images labeled with other markers. (b) DeepSNiF improves the Pearson correlations between Collagen III-labeled IMC images with low and high SNR significantly ($N = 10$). Scale bar: (a) $45 \mu\text{m}$. (b) $100 \mu\text{m}$	99

Figure 4.13: IMC-Denoise enhances both IMC image quality and the Pearson correlations of IMC images labeled with the same markers. (a) A tissue was stained by ^{143}Nd , ^{146}Nd conjugated TP53, and ^{158}Gd , ^{173}Yb conjugated CD56, respectively, with different SNRs. The IMC images were firstly processed by DIMR to remove hot pixels. Then DeepSNiF was employed to improve the image quality of all the images, because the qualities of the higher SNR images are still sub-optimal. (b) After DeepSNiF processing, the Pearson correlation coefficients (PCC) improved, in which those of the double DeepSNiF-processed images are the highest. Notably, the DeepSNiF trained by the CD3 images from the human bone marrow dataset (Tables A.4 and A.7) was used to denoise the DIMR-processed TP53 and CD56 images, due to their highly similar features and the lacking of sufficient TP53 and CD56 training sets. Scale bar: $75\ \mu\text{m}$ 100

Figure 4.14: IMC-Denoise enhances both IMC image quality and the Pearson correlations between IMC and IF images. (a) The same tissues were stained with CD3, CD4, CD61 and CD169 by IMC and IF, respectively. The low SNR IMC images were processed by DIMR to remove hot pixels and then by DeepSNiF to improve image quality. (b) After DeepSNiF processing, the PCC between IMC and IF improved, indicating DeepSNiF is able to improve the IMC image quality. Scale bar: CD3: $98\ \mu\text{m}$. CD4: $110\ \mu\text{m}$. CD61: $69\ \mu\text{m}$. CD169: $87\ \mu\text{m}$ 101

Figure 4.15: DeepSNiF performs better than MAUI on DIMR data to filter background noise. (a) DIMR and DeepSNiF-processed IMC images labeled with CD34, respectively. The DeepSNiF-processed image was binarized by the threshold value 1 and then overlaid with manual annotated ground truth. (b) The DIMR-processed image was processed by the MAUI software package with a wide range of parameters to select the best background noise removal result and also overlaid with the manual annotated ground truth. The F1 score of the DeepSNiF_thresh (lower right corner) result is always higher than that of the MAUI results (upper left corner in every image), indicating DeepSNiF is better than MAUI on DIMR data in terms of background noise removal. Scale bar: $96\ \mu\text{m}$ 104

- Figure 4.16: IMC-Denoise enables background noise removal of the human bone marrow IMC dataset. (a) Examples of DIMR and DeepSNiF-processed IMC images labeled with CD34 and Collagen III. (b) Visual inspection of background removal results of DIMR and DeepSNiF-processed images, in which DIMR_thresh and DeepSNiF_thresh are binarized with the optimal thresholds, DIMR_Ilastik and DeepSNiF_Ilastik are segmented by the Ilastik software package, and MAUI results are the DIMR images processed by the MAUI software package, respectively. Manual annotated images are served as ground truths. (c) After DeepSNiF denoising, the background removal accuracy improves significantly in terms of F1 score, for both CD34 ($N = 15$) and Collagen III-labeled images ($N = 12$). Notably, DeepSNiF_Ilastik achieves the highest accuracy, while DeepSNiF_thresh performs better than all the background removal results from DIMR images. Scale bar: Top: 50 μm , bottom: 35 μm 106
- Figure 4.17: The impact of DIMR on single cell data extracted from DeepSNiF-based cell segmentaton masks. Each sub-figure represents the one-on-one relationship between the raw and DIMR data of a particular marker in single cell scale. The bottom right value in each sub-figure represents the percentage of the difference between the raw and DIMR data. 107
- Figure 4.18: The impact of DeepSNiF on single cell data. (a) Each sub-figure represents the one-on-one relationship between the DIMR and DeepSNiF data of a particular marker in single cell scale. The bottom right value in each sub-figure represents the slope of the line fitting results and the PCC between the DIMR and DeepSNiF data. (b) The STD of the normalized positive marker differences between DIMR and DeepSNiF for DeepSNiF impact evaluation. The data was 99th-percentile normalized for extreme value mitigation..... 109
- Figure 4.19: IMC-Denoise enhances downstream analysis of the human bone marrow IMC dataset. (a) Visual inspection of DeepSNiF and DIMR_Ilastik-based denoising results on different markers-labeled IMC images. (b)–(e) Evaluations of denoising algorithms with manual gating strategies on single cell data. The numbers in these panels are the cell percentages of the corresponding ranges. DIMR slightly enhances the single cell analysis over raw data, while DeepSNiF further enhances the DIMR results and overall performs better than semi-automated DIMR_Ilastik-processing. Scale bar: 107 μm 110

Figure 4.20: <i>t</i> -SNE plots and relative change of DIMR and DeepSNiF data. (a) <i>t</i> -SNE plots of DIMR and DeepSNiF with cell phenotyping results. (b) The relative change in cell phenotypes before and after DeepSNiF.	114
Figure 4.21: Comparisons of DIMR and DeepSNiF-processed IMC images labeled with different cell markers, and the corresponding cell annotation results. The sub-panels (i)–(iv) in (a) and the bottom row in (b) correspond to the white dashed box region selection in their first panels, respectively. The white contours represent the differential phenotyping results between DIMR and DeepSNiF. Scale bar: (a) 110 μm . (b) Top: 145 μm , bottom: 50 μm	115
Figure 4.22: Deep-SNiF enhances the sensitivity of cell phenotyping. After Deep-SNiF processing, the non-specific marker signals reduce while the specific ones enrich in the cell types, respectively. The circle size indicates the positive marker percentage in a particular phenotype of DIMR, and the circle colour indicates the relative changes of the positive rate for the particular markers after DeepSNiF enhancement.	117
Figure 4.23: DeepSNiF enhances the specificity of cell phenotyping. With Deep-SNiF denoising, the ratios of specific phenotypes increase while those of non-specific phenotypes decrease in the positive markers. The relative change is the difference in percentage composition of each cell type before and after DeepSNiF enhancement.	118
Figure 4.24: DeepSNiF enhances lymphocyte annotation accuracy. (a) Manual annotations for lymphocytes and comparisons with DIMR and Deep-SNiF phenotyping results with DeepSNiF-based cell masks. The white contours represent the differential phenotyping results between the annotated and DIMR/DeepSNiF results. (b) Annotation evaluations of DIMR and DeepSNiF by both Jaccard and F1 scores. Scale bar: 85 μm	119
Figure 4.25: Representative images of lymphocyte markers after DeepSNiF denoising from specimens of normal (upper left), MDS (upper right), AML (lower left) and AML with lymphoid aggregate (lower right) tissue samples. Scale bar: 112 μm	120

Figure 4.26: DeepSNiF enhances lymphocyte analysis. (a) Nearest distance comparisons between different cell types of normal, MDS and AML tissues from manual, DIMR and DeepSNiF phentyping results. The box plots are shown as the tukey format. (b) Cell densities comparisons of normal, MDS and AML tissues from manual, DIMR and DeepSNiF phentyping results. Single dots and the means with 95% confidence interval are shown in these figures. (c) and (d): Correlation analysis between CD4 T cell and B cell, monocyte/macrophage densities per tissue from manual, DIMR and DeepSNiF phentyping results. The data from the reference group in (c) comes from annotated data; while that from the reference in (d) comes from annotated (CD4 T cells) and DeepSNiF (monocytes/macrophages) results, separately. The dashed lines represent 95% confidence interval.	122
Figure 4.27: DeepSNiF works on multiple markers training. The DIMR-processed IMC images were trained by DeepSNiF with single marker in each network (DeepSNiF_single) and all the markers in a single network (DeepSNiF_combo), respectively. The denoising results indicate both approaches enables IMC image quality improvement. Scale bar: 48 μm	124
Figure 4.28: The limitation of the DIMR algorithm. (a) Success cases of DIMR on challenging hot pixels. (b) Failure cases of DIMR on hot clusters. DIMR is able to remove line-style consecutive hot pixels while fails on hot clusters. Scale bar: (a) Top: 8 μm , bottom: 24 μm . (b) Top: 40 μm , bottom: 20 μm	127
Figure 5.1: Overview of InterSTELLAR. (a) Undirected graph construction from highly multiplexed images. (b) Single cell feature matrix \mathbf{Z} , adjacent matrix \mathbf{A} and tissue label Y per tissue are fed into InterSTELLAR. (c) Self-attention pooling module with interpretable learning. (d) A cell-based attention heatmap generated by InterSTELLAR, in which the attention scores per cell are positively correlated with their contribution to the tissue classification results. Scale bar: 172 μm	133
Figure 5.2: The relationship between tissue-scale weight η and clinical type classification performance.	136
Figure 5.3: Box plots of cell density per phenotype in high- and low-attention regions from healthy, TNBC and non-TNBC cancerous tissues.	139

Figure 5.4:	Evaluations of InterSTELLAR on tissue classification and cell community detection. (a) InterSTELLAR is more accurate than FNN, Random Forest and SVM algorithms on tissue clinical type classification, validated by 10-fold cross validation ($N = 30$ per fold) and an independent test set ($N = 73$). (b) Highly multiplexed IMC images and the corresponding attention heatmaps generated by InterSTELLAR and FNN. (c) Moran's I and Geary's C statistics indicate InterSTELLAR achieves higher spatial correlation than FNN in terms of attention scores. Scale bar: $145 \mu m$	140
Figure 5.5:	Evaluations of InterSTELLAR on tumor region detection. (a) Schematic of generating tumor and stroma masks. Embedded features are extracted from a trained network and then classified by unsupervised clustering algorithms. Here, K-means clustering with $K = 2$ is utilized. (b) Highly multiplexed IMC images and the corresponding tumor and stroma masks generated by manual annotations, InterSTELLAR and FNN. (InterSTELLAR performs better than FNN on tumor region identifications of both TNBC and non-TNBC cancerous tissues in terms of F1 score. Scale bar: $158 \mu m$	142
Figure 5.6:	InterSTELLAR characterizes the breast cancer tissue microenvironments from different clinical subtypes. (a) Highly multiplexed IMC images of healthy, TNBC and non-TNBC cancerous tissues as well as their corresponding attention heatmaps and segmented high attention region masks. (b) The percentages of immune, stroma and epithelial cells in high and low attention regions from healthy, TNBC and non-TNBC cancerous tissues. (c) The percentages of tumor cells and median distance of immune cells to their nearest tumor cells in high and low attention regions from cancer tissues. Scale bars: Healthy: $195 \mu m$; TNBC: $175 \mu m$; Non-TNBC cancers: $183 \mu m$	144
Figure 5.7:	InterSTELLAR characterizes tissue microenvironments from different clinical subtypes. (a) Mean cell density per phenotype from high and low attention regions of all the tissues. (b) KL divergence between high and low attention regions regarding distributions of cell density per phenotype.	146
Figure 5.8:	InterSTELLAR characterizes distinct clinical outcomes for SCP subgroups. (a) Mean cell density per phenotype of high attention regions from various SCP subgroups. (b) UMAP plot of the tissues labeled with their corresponding SCP indexes.....	147

Figure 5.9: Survival analysis for SCP subgroups. (a) Kaplan-Meier curves of overall survival for each subgroup ($N = 283$) on the basis of cell density per phenotype of high attention regions, with $*P < 0.05$. $\#P < 0.05$, $\#\#P < 0.01$ represent the statistical significance of a single subgroup compared to all other samples. (b) P values for overall survival analysis from different clustering strategies. In (a) and (b), P values were calculated through two-sided log-rank test. (c) Relative hazard ratios and 95% confidence intervals of disease-specific overall survival for cell densities per phenotype and SCP subgroups estimated using a Cox proportional hazards model. Reference group 1: Mixed for SCP subgroups. 148

Figure 5.10: Comparisons of the patient group classified by SCP subgroup, clinical subtype and tumor grade. (a) UMAP plots of the tissues labeled with their corresponding SCP indexes, clinical subtypes and tumor grades. (b) Kaplan-Meier curves of overall survival for each patient group on the basis of SCP subgroup, clinical subtype and tumor grade. 149

Figure 5.11: Kaplan-Meier curves of overall survival from different clustering strategies and their corresponding P values of log-rank tests. 150

Figure 5.12: Kaplan-Meier curves of overall survival of each subgroup on the basis of cell density per phenotype of high attention regions. The black curve represents the survival curve of all samples. Each P value represents the statistical significance of a single subgroup compared to all the other samples. All the P values were calculated through two-sided log-rank tests. 151

Figure 6.1: Schematic of theranostic analysis by integrating histological images, DAR and IMC data. The DAR images and IMC images are restored by PG-PEM and IMC-Denoise, respectively. Then, the histological images are co-registered with the restored DAR and IMC images, respectively. Subsequently, downstream analysis including dose distribution measurement and microenvironment analysis can be conducted, in which the correlations between tissue disease states, dose distribution and microenvironment are analyzed. Finally, theranostic medicine can be enhanced based on the results. 157

Figure 6.2: The segmentation pipeline: The raw image was converted to CIELAB space. After Z-score normalization, blocks were sampled, and hand-crafted features were extracted from them. The non-osseous tissues were predicted based on K-means clustering method. The bone surfaces were predicted using a network combining the extracted features and CNN trained features. Scale bar: $640 \mu\text{m}$ 158

Figure 6.3:	Network structure, in which the upper part can be seen as a FNN, the lower part is a CNN and the whole network is a MNN.	160
Figure 6.4:	Registration process of DAR and H&E images. (a) Segmentation result. (b) Restoration result. (c) Fusion of (a) and (b). (d) Initial registration result. (e) Final registration result. Scale bar: (a) 1 mm, (b) 2.3 mm.....	161
Figure 6.5:	Registration process of IMC and H&E images. (a) A H&E image from the human bone marrow data of Chapter 4 and its zoomed-in version. (b) Nuclei segmentation mask. (c) A DNA channel image of IMC data. (d) The final registered H&E image. Scale bar: (a) left: 1.4 mm, right: 356 μm , (c) 138 μm	162
Figure 6.6:	Tissue segmentation results for human bone biopsy. (a)–(d) corresponds to 4 different slides. For each group, the left image is the raw histopathology image and the right one is the segmentation result. Scale bar: 1 mm.	163
Figure 6.7:	Training and prediction results of different networks. (a)–(c) are the loss curves for FNN, CNN and MNN, respectively. The validation for (a)–(c) are 0.0741, 0.0723 and 0.0709, respectively. (d) Fuzzy Dice score for the prediction results. (e) Pixel wise MSE for the prediction results. (f)–(i): Examples of bone segmentation results for human bone biopsy using different networks. For each group, the left image is the raw histopathology image, the right part: upper left: FNN, upper right: CNN, lower left: MNN and lower right: manually labeled data. Scale bar: 1 mm.	164
Figure 6.8:	Examples of registration results between DAR and H&E images. (a)–(c) are the results from three slides; From left to right in every row: H&E image, segmented image before registration, segmented image after registration, restored DAR image and fusion of the two images. .	165
Figure 6.9:	An example of registration result between IMC and H&E images. Note that the IMC images have been restored by IMC-Denoise in Chapter 4. Scale bar: 138 μm	166

List of Tables

Table 3.1:	Computational time of PG-PEM in different conditions	35
Table 4.1:	Simulation parameters	83
Table 5.1:	Cell markers used in training and downstream analysis.	131
Table 5.2:	P values for overall survival analysis from different clustering strategies.	150
Table A.1:	List of cell markers used for Collagen III-labeled tissues in Figures 4.12a and 4.16a.....	[186]
Table A.2:	List of cell markers for the tissue staining in Figure 4.12b.....	[188]
Table A.3:	List of cell markers for the tissue staining in Figure 4.13	[189]
Table A.4:	List of cell markers used for other IMC images from the human bone marrow IMC dataset.....	[190]
Table A.5:	Training details for the simulation datasets.....	[193]
Table A.6:	Training details for the Collagen III-labeled images in Figures 4.11, 4.12 and 4.16.....	[193]
Table A.7:	Training details for the other markers-labeled images from the human bone marrow IMC dataset	[194]
Table A.8:	Training details for the images from human breast cancer IMC dataset.....	[195]
Table A.9:	Training details for the images from the human pancreatic cancer IMC dataset	[196]
Table A.10:	Training details for the images from the IMC dataset	[197]
Table A.11:	The estimated thresholds for positive markers	[197]

Acknowledgments

I would like to express my deepest and heartfelt gratitude to my advisor, Dr. Daniel Thorek, for his invaluable guidance and unwavering support throughout my doctoral research journey. Dr. Thorek has been an exceptional advisor, showing me how to develop critical thinking skills and grow as a dedicated researcher. His expertise, passion, and dedication to the field of medical imaging have been truly inspiring me and I feel extremely fortunate to have had the opportunity to work alongside him.

I am grateful to my dissertation committee members, Dr. Joseph O’Sullivan, Dr. Abhinav Jha, Dr. Stephen Oh, and Dr. Aimilia Gastounioti, for their insightful feedback and valuable contributions to my research. Their expertise and diverse perspectives have substantially improved the quality of my research.

I would like to acknowledge the funding agencies that supported this research: National Institute of Health R01CA229893, R01CA201035, and R01CA24071. I would also like to thank the support from the student research award from Society of Nuclear Medicine and Molecular Imaging (SNMMI).

Last but not least, I am indebted to my family, Weiting Hu, Yanbo Wen, and Shengdong Lu, for their unconditional love, encouragement, support, and inspiring me to pursue my passions and dreams.

Peng Lu

Washington University in St. Louis

December 2023

Dedicated to my wife, Weiting Hu, and my parents, Yanbo Wen and Shengdong Lu

ABSTRACT OF THE DISSERTATION

Improving and Integrating Quantitative Single Cell Spatial Biology
for Theranostic Medicine

by

Peng Lu

Doctor of Philosophy in Imaging Science

Washington University in St. Louis, 2023

Professor Daniel L.J. Thorek, Chair

Disease states are the result of a complex interplay of many different cell types interacting in close proximity in the context of often heterogeneous tissues. Alpha particles are drawing intense research and clinical interest because of their potent cytotoxic effects and their short path lengths. Analyzing the dose distribution and tissue micro-environment for alpha therapy plays a key role in predicting the efficacy of this targeted radiotherapy. However, to date there have been no direct on-tissue analytical methods for alpha dose distributions and the corresponding tissue microenvironments.

Therefore, we have developed a pipeline to overcome this limitation by utilizing quantitative imaging techniques including digital autoradiography (DAR) and imaging mass cytometry (IMC). DAR is an imaging technique that provides information on the distribution of radionuclides in tissue sections and is widely used in drug development. IMC is an emerging highly multiplexed molecularly specific histological method that can report the protein expression profiles of up to 40 markers on processed tissue sections.

DAR and IMC have their own deficiencies that have challenged their utility in previous applications. Firstly, DAR suffers from low image resolution and significant background noise,

which can lead to poor correlation and, in some cases, errors in determining the relationship between radiotracer distribution, anatomical structure, and molecular expression profiles. Secondly, despite extensive optimization of staining conditions, IMC images may exhibit low signal-to-noise ratios (SNR) for specific markers, and the presence of pixel intensity artifacts can detrimentally affect image quality and subsequent downstream analysis. Lastly, the spatial information of cells obtained from IMC images has not been adequately harnessed, resulting in a limited understanding of higher-order patterns such as tissue organization.

To enable direct on-tissue analysis of dose distribution and the tissue microenvironment, we propose the development of algorithms and pipelines to address these challenges. First, we introduce a Poisson-Gaussian penalized expectation maximization (PG-PEM) algorithm to blindly enhance DAR images, thereby improving dose measurement accuracy. Subsequently, we present IMC-Denoise, a content-aware pipeline designed to enhance IMC cell phenotyping outcomes. It includes a differential intensity-based restoration algorithm (DIMR) for outlier pixel removal and a self-supervised deep learning algorithm for shot noise image filtering (DeepSNiF). Furthermore, we introduce an interpretable spatial cell learning framework called InterSTELLAR, which classifies tissues into distinct clinical types. InterSTELLAR incorporates an attention-based pooling module for cell-level interpretable learning. Finally, with the improved data and newly developed algorithms, we establish an integrated pipeline for the automated analysis of data from different modalities. These innovative approaches are poised to enhance our understanding of dose distribution and tissue microenvironment responses, benefiting target engagement studies in drug development and enabling more precise theranostic medicine.

Chapter 1

Introduction

The development of improved biomedical imaging methods has been a critical component in our improved understanding of healthy and diseased tissues. This is true at the microscopic scale – the ability to localize structures within a single cell, to the tissue scale – to visualize and compare the organization of an organ or tumor, and through to the organismal scale – for noninvasive imaging for disease detection and monitoring. In this thesis, we are concerned with further improvements to imaging methods for the evaluation of drug and cell type distributions using digital autoradiography (DAR) [1, 2, 3, 4] and imaging mass cytometry (IMC) [5, 6, 7], respectively. As will be detailed below, these are powerful techniques to quantitatively characterize tissues and biological systems, they are limited in several fundamental ways. We have leveraged advances in imaging science and computational methods in order to improve our extraction of quantitative and qualitative data from these two techniques that enables further understanding of biological systems of interest.

1.1 DAR and IMC in Theranostic Medicine

Radionuclides have been used extensively in imaging and therapeutics since the initial discovery by the Curies. Widely used technologies including single photon emission computed tomography (SPECT) [8] and positron emission tomography (PET) [9] provide highly sensitive deep tissue penetration and accurate assessment of tracer distribution to detect aberrant cellular activity often with molecular specificity. In recent years, the interest in use of isotopes for systemic administration for anticancer therapy has increased [10, 11, 12, 13, 14]. This is reflected in approval of $^{223}\text{RaCl}_2$ [10, 11], ^{177}Lu -octreotate [12] and ^{177}Lu -PSMA-617 [13].

Alpha particles have become the focus of extensive research and clinical attention due to their powerful cytotoxic effects and limited penetration distances [15, 16]. Assessing the dose distribution for alpha therapy at the cellular level is crucial for predicting the effectiveness of this targeted radiotherapy, particularly given the microscopic scale of alpha particle paths. However, current small-scale dosimetry methods are predominately based upon idealized computational anatomical models [17, 18]. While useful, these provide limited real-world information in heterogeneous patient populations. Autoradiography is a powerful, high resolution and quantitative molecular imaging technique used to study the tissue distribution of radioisotopes in biologic systems and for analytic assays, which is able to overcome the deficiencies of ideal model-based methods. Originally, radioactivity distributions were acquired using photographic emulsions, which are high resolution, but require time consuming, fickle, and variable processes. Currently, phosphor imaging plate-based DAR has supplanted film due to its linear activity response, non-destructive approach, no chemical-processing requirement, large dynamic range and considerable sensitivity [2, 4, 19].

In another aspect, disease states arise from intricate interactions among numerous diverse cell types, occurring in close proximity within the frequently heterogeneous milieu of tissues. Traditional approaches to assess the microenvironment at the tissue scale have been limited in the number of specific markers that can be acquired to robustly resolve distinct cell types. Flow cytometry, perhaps the most widely used technique to study cell populations and states in this milieu, requires single cell disaggregation of the tissue resulting in complete loss of spatial context [20, 21]. Highly multiplexed imaging provides a means to assess these events at cellular resolution in *situ*, with extensive protocol development in progress[22]. Such techniques consist of tissue-based cyclic immunofluorescence (*t*-CyCIF) [23], co-detection by indexing (CODEX) [24], multiplexed ion beam imaging (MIBI) [25, 26] and IMC. Among these novel imaging techniques, IMC stands out with its own advantages including very low background signal and channel crosstalk [5, 6, 7]. It allows for the detection of more than 40 antigens simultaneously to facilitate single-cell, spatially resolved, highly multiplexed analysis of solid tissues. This provides essential information on the distribution of transcripts, proteins, and protein modifications within single cells, microenvironments, and entire tissues [5, 6, 7, 27, 28, 29, 30, 31, 32].

While DAR and IMC serve as powerful tools for the on-tissue alpha dose and tissue microenvironment analyses separately, there have been a lack of direct on-tissue analytical methods for aligning the dose distributions with their respective tissue microenvironments. Besides, DAR and IMC images suffer from modality-intrinsic artifacts, to name a few, low image resolution and noise sources. These effects may limit their powers in downstream on-tissue analysis. Additionally, the cell spatial information from IMC images, as well as other multiplexed imaging data, are not adequately utilized. Therefore, higher-order patterns such as tissue organization cannot be well investigated. This may further hinder the understanding of specific relationship between particular dose distributions and tissue organizations.

In this dissertation, we develop several algorithms and frameworks to improve the image quality of DAR and IMC, the spatial analysis of IMC data and other multiplexed imaging data, and integrate all these information to achieve on-tissue dose and microenvironment analysis. With these approaches, improved understandings of dose distribution and tissue microenvironment responses are expected to benefit target engagement studies for drug development and to enable more accurate theranostic medicine.

1.2 Objectives and Structure of This Dissertation

In Chapter 2, we begin with giving general introductions on DAR and IMC, the two imaging modalities used in our theranostic medicine pipeline. We first provide an overview of their imaging principles, including hardware and data processing pipelines. We then elucidate the common imaging artifacts associated with DAR, such as low resolution, shot (Poisson) noise, and Gaussian noise, as well as those pertaining to IMC, including the occurrence of hot pixels and shot noise. In particular, we provide a detailed explanation of where the artifact sources come from and how existing methods used to solve them. Next, we introduce current methods for multiplexed imaging data processing, and discuss their shortcomings. Finally, we discuss the limitations of current on-tissue dose analysis techniques.

In Chapter 3, we focus on DAR image restoration. Differing from conventional optical systems, the point-spread function (PSF) in DAR is determined by properties of radioisotope decay, phosphor, and digitizer. Calibration of an experimental point-spread function a priori is difficult, prone to error, and impractical. To address this challenge along with the blurry and noise effects in DAR, we develop a penalized maximum-likelihood expectation-maximization algorithm (PG-PEM) to blindly restore DAR images. PG-PEM implements a patch-based estimation algorithm with density-based spatial clustering of applications with

noise (DBSCAN) [33] to estimate noise parameters and uses $L2$ and Hessian Frobenius norms as regularization functions to improve performance. We will implement PG-PEM on pre- and clinical DAR images including mice sections and human bone biopsy labeled with various radioisotopes. We will validate that PG-PEM is able to increase resolution and contrast comparing to existing methods, while robustly accounting for DAR noise and demonstrates the capacity to be widely implemented to improve pre- and clinical DAR imaging of radio-pharmaceutical distribution.

In Chapter 4, we demonstrate an automated content-aware pipeline, IMC-Denoise, for IMC image restoration. This framework deploys a differential intensity map-based restoration (DIMR) algorithm for removing hot pixels, and a self-supervised deep learning algorithm for shot noise image filtering (DeepSNiF). We will demonstrate that IMC-Denoise outperforms existing methods for adaptive hot pixel and background noise removal, with significant image quality improvement in modeled data and datasets from multiple pathologies, particularly for a technically challenging human bone marrow IMC dataset. Besides, we will illustrate IMC-Denoise can enhance manual gating and automated phenotyping with cell-scale downstream analyses. Furthermore, we will conduct spatial and density analysis on targeted cell groups to reveal more subtle but significant enhancement by IMC-Denoise.

Multiplexed imaging technologies enable highly resolved spatial characterization of cellular environments. Nevertheless, even with enhanced image quality, exploiting these rich spatial cell datasets for biological insight is a considerable analytical challenge. In particular, effective approaches to define disease-specific microenvironments on the basis of clinical outcomes is a complex problem with immediate pathological value. In Chapter 5, we present an interpretable spatial cell learning framework (InterSTELLAR), a geometric deep learning approach for multiplexed imaging data analysis, to directly link tissue subtypes with corresponding cell communities that have clinical relevance. Using a publicly available breast

cancer IMC dataset, we will demonstrate that InterSTELLAR allows simultaneous tissue type prediction and interested community detection, with improved performance over conventional methods. Besides, we will validate that InterSTELLAR enables capture of specific pathological features from different clinical cancer subtypes, and reveal potential relationships between these regions and patient prognosis.

With improved image quality and spatial analysis method of DAR and IMC, we propose to investigate the relationship between dose distributions and tissue microenvironments. To achieve this goal, in Chapter 6, we develop a pipeline to automatically integrate and analyze classical histological image information along with DAR and IMC data. We firstly develop the framework based on Hematoxylin and eosin (H&E) stained histopathological images and DAR from the human bone biopsy dataset in Chapter 3. In particular, we develop a deep learning-based algorithm to segment bone area in H&E images and then register the two modalities automatically based on the segmentation results and dose distribution in DAR. Thereafter, the registration results will be used for on-tissue dose analysis. Next, we propose to further embed IMC images into the developed pipeline. Here, the H&E and IMC images are co-registered by the cell nuclei segmentation results of H&E and DNA channel images of IMC, so that IMC and DAR images can be integrated for theranostic medicine downstream analysis.

1.3 Dissemination of Research Findings in the Dissertation

We will make all the advancements achieved in this dissertation openly available to other researchers in the imaging science community through publications and reports. The work on development of PG-PEM for DAR image restoration in Chapter 3 has been published

in The Journal of Nuclear Medicine [34]. The source code of PG-PEM has also been made available at https://github.com/PENGLU-WashU/PG_PEM. The work on development IMC-Denoise for IMC image restoration in Chapter 4 has been published in Nature Communications [35]. The source code of IMC-Denoise has been made available at https://github.com/PENGLU-WashU/IMC_Denoise. The work on development of InterSTELLAR for multiplexed imaging data processing has been uploaded on bioRxiv [36] and has also been partly summarized and published as a proceedings paper at SPIE Medical Imaging 2023 Conference [37]. The corresponding source code has been made available at <https://github.com/PENGLU-WashU/InterSTELLAR>. Partial work in Chapter 6 has been published in The Journal of Nuclear Medicine [38]. Additionally, our research has been presented at premier conferences that include Biomedical Engineering Society (BMES) Annual Meeting 2022, and The Society of Nuclear Medicine and Molecular Imaging (SNMMI) Annual Meeting 2020 [39, 40].

Chapter 2

Dose Measurement and Microenvironment Analysis Techniques in Theranostic Medicine

2.1 DAR Imaging Technique

2.1.1 DAR imaging process

Generally, DAR imaging process starts from the placement of tissue samples containing radioactivity in close proximity to a phosphor screen. The phosphor screen absorbs ionizing radiation and effectively captures the energy emitted by the radioactive substance, resulting in the formation of a “latent image” that represents the distribution of radioactive activity (Figure 2.1a). Except for very low energy beta emitters (tritium), the phosphor layer and the specimens are typically separated by low-attenuation film to prevent contamination of the

screen itself, and exposure lasts hours to days. The phosphor plate is raster scanned with a small focal-spot red laser, and the photostimulated light is collected by a photomultiplier tube to form a digital image (Figure 2.1b). The intensity of emitted light is proportional to the amount of radioactivity in the tissue sample.

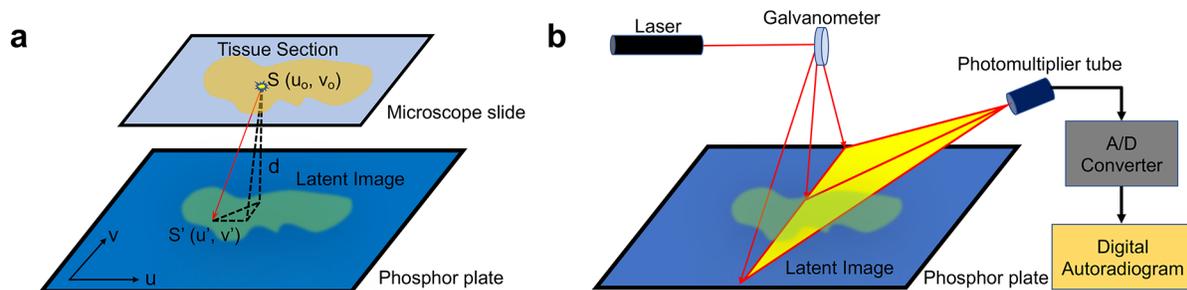


Figure 2.1: DAR imaging process. (a) Expose the plate to the radioactive section to produce a latent image through a trapping of electron-hole. The latent image is formed by x-rays, gamma rays, beta particles, and alpha particles generated in the exposing process. Here, S is a point source and S' is a possible destination of S . d is the distance between the tissue section and the phosphor plate. In uv -plane, the coordinate of S and S' are (u_o, v_o) and (u', v') separately. (b) Scan a laser over the latent image on the phosphor plate to stimulate release of photons so as to generate the corresponding digital autoradiographic image. The pure signal is blurred mainly during the exposing process while contaminated by noise during the scanning process.

2.1.2 Imaging artifacts of DAR and the corresponding restoration techniques

Image quality and resolution from DAR are suboptimal for the purposes of radioligand target-cell engagement, diagnostic tracer distribution and the microscopic distribution of radiotherapeutics in pre- and clinical tissue samples. Unlike conventional optical microscopy systems, DAR does not utilize an aperture or a collimator, and the solid angle subtended at the samples by the imaging plate is almost 2π . Therefore, the PSF results from isotropic

radiation emission and is dependent on a combination of the dispersion of the energy deposited in the phosphor plate, characteristics of the phosphor plate (lattice and grain size) and readout laser. Because the radiative emissions vary from isotope to isotope, the PSF is in general isotope specific. Additionally, replicating relevant features of the radiative signal for DAR acquisitions in a phantom is difficult, and decay of the isotopes further complicates calibration. In aggregate, it is thus not practical to calibrate the PSF beforehand.

Apart from blurring effects caused by the PSF, background signal is always present in the imaging process caused by environmental radiation. DAR image noise can be attributed to multiple sources: Poisson noise exists in the photon counting imaging system; Gaussian noise comes from the imaging reader readout process, phosphor sheet inhomogeneities and grain [41]. A limited number of approaches have been tested to restore autoradiographs to overcome noise and blur-related artifacts. These include a regularized iteration method after noise filtration [42] and the modeling of noise features [43]. The results from these investigations are not ideal and have not been widely adopted, in part, because several have used an emulsion film-based system (the predecessor to phosphor storage plate technology) and several have caused noise amplification effects. Common to these approaches, the authors have calibrated the PSF by building a non-ideal resolution phantom.

In order to model the noise in DAR systems of many different isotopes, a blind estimation approach for restoration is preferred. Recently, a mixed noise model has been employed to denoise digital images, which can improve the quality of images contaminated by Poisson and Gaussian noise sources [44, 45, 46, 47]. A key step in such a model is the estimation of the noise parameters. For single image restoration, patch-based [44], segmentation-based [46] or Fourier-based [47] methods have been developed. Additionally, several blind and non-blind image restoration techniques for biomedical images have been advanced [48, 49, 50, 51, 52, 53, 54]. For the specific task of blind restoration, the regularizations for PSF and specimen

are considered in some of these methods, providing a path forwards for blind DAR image estimation.

2.2 IMC Imaging Technique

2.2.1 IMC imaging process and noise sources

In IMC imaging process, tissue sections are stained with a panel of metal-conjugated antibodies, and data is acquired by UV-laser raster ablation of the section in 1-micron pixels for cytometry by time-of-flight (CyTOF) mass analyzer (Figure 2.2). This novel imaging technology allows for the detection of more than 40 antigens simultaneously to facilitate single-cell, spatially resolved, highly multiplexed analysis of solid tissues. This provides essential information on the distribution of transcripts, proteins, and protein modifications within single cells, microenvironments, and entire tissues [5, 6, 7, 28, 29, 30, 31, 32, 55]. The pixel data is processed into an image, thereby allowing the visualization of phenotypes and incorporation of spatial information in subsequent analyses. These properties make it a unique tool for the evaluation of complex biological systems.

Despite the wide applications in pre- and clinical research using this state-of-the-art multiplexed imaging technique, there exist specific technical noise sources in IMC, which include hot pixels, channel spillover and shot noise [5, 6, 27, 31, 56, 57]. Hot pixels are concentrated areas of high counts which are uncorrelated with any biological structures. Putatively, these can result from deposition of metal-stained antibody aggregates. In IMC images, single hot pixels are the most common outliers, and small hot clusters with several consecutive pixels may also exist. Channel spillover refers to scenarios where the signal of a source channel contaminates a target channel or is correlated with such contamination. The spillover in IMC

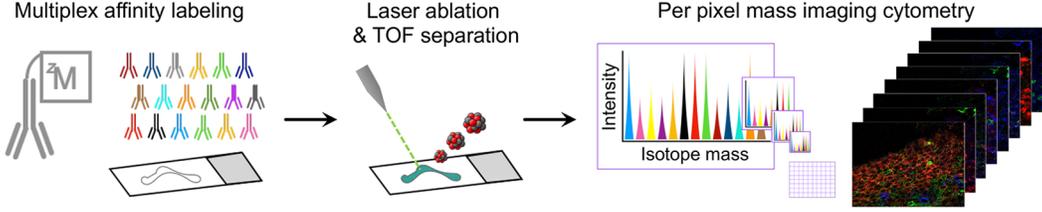


Figure 2.2: IMC workflow: Stable metal isotope-labeled antibodies are bound to molecular epitopes in tissue sample (fresh frozen, fixed or formalin-fixed paraffin-embedded) before rastering of an ablative UV laser with a spot size $< 1 \mu\text{m}$. The ionized material is coupled to a Hyperion time-of-flight (TOF) mass cytometer to detect mid-range metal masses (> 80 Da). Unlike traditional microscopy, there is no background as the masses are detected in non-overlapping distinct channels (m/z). Finally, more than 40 channels corresponding to molecular specific biological features are acquired for the defined region of interest tissue slide.

can occur from a variety of reasons, such as instrument properties (abundance sensitivity), isotopic impurities and oxidation. Finally, shot noise exists because of ion counting imaging processes, which are pixel-independent, signal-dependent and usually modeled as a Poisson process. Additionally, noise levels are related to multiple other factors, including variations in conjugated metal isotopes, antibody concentration and arrangement.

Here, we model the ion counting imaging as a Poisson process and hot pixels as outliers with much larger intensity than their adjacent pixels. As a result, the IMC imaging model is built as Equation (2.1).

$$\mathbf{R} = \mathcal{P}[\mathbf{X} + \mathbf{X}^{\text{spillover}}] + \mathbf{Q}, \quad (2.1)$$

where $\mathbf{R} = \{r_p\}$ is the raw image set, $\mathbf{X} = \{x_p\}$ the true signals without noise, $\mathbf{X}^{\text{spillover}} = \{x_p^{\text{spillover}}\}$ the spillover signals without noise, $\mathbf{Q} = \{q_p\}$ the hot pixels, p the pixel index and $\mathcal{P}[x]$ the Poisson noise with mean x .

In our research, we only consider the hot pixels and ion shot noise in raw IMC images, as the impact of spillover is usually minor [56]. Meanwhile, if the marker panel employed is well-designed and titrated, spillover can be even avoided. If spillover is observed, it should be corrected after the restoration of these two noise sources, as its signal is contaminated by them as well. Therefore, the IMC imaging model is simplified as Equation (2.2).

$$\mathbf{R} = \mathcal{P}[\mathbf{X}] + \mathbf{Q}, \quad (2.2)$$

Noise impact on pixel scale

First we consider a signal pixel s and a background pixel b . Assume that the intensity of the background pixel $x_b > 0$ because of unspecific staining or staining artifacts. Without noise, we have $x_s > x_b$. However, with the shot noise and hot pixel artifact, the two pixel values become $r_s = \mathcal{P}[x_s] + q_s$ and $r_b = \mathcal{P}[x_b] + q_b$. Because the Poisson model can be feasibly estimated as a Gaussian process [34], Equation (2.2) can be converted as

$$\mathbf{R} = \mathcal{N}(\mathbf{X}, \mathbf{X}) + \mathbf{Q}. \quad (2.3)$$

As a result, the two pixel values are approximated as $x_s + \mathcal{N}(0, x_s) + q_s$ and $x_b + \mathcal{N}(0, x_b) + q_b$. Thus, the noise terms of pixels s and b are $\mathcal{N}(0, x_s) + q_s$ and $\mathcal{N}(0, x_b) + q_b$, respectively. With the impact of the noise terms, it is possible that $r_s < r_b$, resulting in error detection of IMC signals. Specifically, the signal-to-noise ratio (SNR) for the shot noise can be defined as

$$\text{SNR} = \frac{x}{\sqrt{x}} = \sqrt{x}. \quad (2.4)$$

Therefore, the lower the signal x , the lower SNR and the higher chance $r_s < r_b$ will be.

Noise impact on cell scale

Subsequently, we consider a positive cell with marker intensity $\frac{1}{M} \sum_{i=1}^M x_i$ and a negative cell with intensity $\frac{1}{N} \sum_{j=1}^N x_j$, respectively. Without the noise impact, we can assume $\frac{1}{M} \sum_{i=1}^M x_i > \frac{1}{N} \sum_{j=1}^N x_j$ even with the existence of unspecific staining. However, after contaminated by the shot and hot pixel noises, the two terms become

$$\frac{1}{M} \sum_{i=1}^M r_i = \frac{1}{M} \sum_{i=1}^M (\mathcal{P}[x_i] + q_i). \quad (2.5)$$

$$\frac{1}{N} \sum_{j=1}^N r_j = \frac{1}{N} \sum_{j=1}^N (\mathcal{P}[x_j] + q_j). \quad (2.6)$$

Similar to pixel scale analysis, these two terms can be further approximated as

$$\frac{1}{M} \sum_{i=1}^M r_i = \frac{1}{M} \sum_{i=1}^M x_i + \mathcal{N}(0, \frac{1}{M} \sum_{i=1}^M x_i) + \frac{1}{M} \sum_{i=1}^M q_i. \quad (2.7)$$

$$\frac{1}{N} \sum_{j=1}^N r_j = \frac{1}{N} \sum_{j=1}^N x_j + \mathcal{N}(0, \frac{1}{N} \sum_{j=1}^N x_j) + \frac{1}{N} \sum_{j=1}^N q_j. \quad (2.8)$$

Under this condition, it is possible that $\frac{1}{M} \sum_{i=1}^M r_i < \frac{1}{N} \sum_{j=1}^N r_j$ such that a positive cell is falsely regarded as a negative one or a negative cell is falsely detected as a positive one. Also similar to the pixel scale analysis, for shot noise the lower the signal values, the higher chance the detection errors will occur.

To summarize, the noise sources can result false detection of signal and positive cell markers, so as to impact downstream analysis. Therefore, it is essential to develop algorithms to filter the hot pixel artifact and to account for the shot noise.

2.2.2 IMC denoising techniques

A number of studies have attempted to address the unique imaging data features of IMC. Hot pixels can be corrected by thresholding methods [27, 30, 31, 58]; however, due to the differences between marker channels and tissues, a threshold needs to be pre-set carefully. An inappropriate threshold may lead to unsatisfactory results. Regarding the channel spillover correction, post-acquisition methods [27, 57] and a bead-based compensation workflow [56] have been proposed. As mentioned, spillover can be neglected when using low concentrations of staining antibodies, which however further lowers SNR. To account for the impact of shot noise, MAUI [26, 57] and a semi-automated Ilastik-based method [59] have been used for background noise removal. These approaches require finely tuned parameters or manually annotated background regions, requiring preprocessing expertise. When working with tissues exhibiting low marker signals, complex mixtures of cell populations, or challenging immunostaining procedures, establishing clear thresholds can prove to be a time-consuming task, often characterized by significant inter-user subjectivity. This, in turn, can lead to suboptimal image quality, thereby adding complexity to subsequent analytical processes.

2.3 Multiplexed Imaging Data Analysis Techniques

Highly multiplexed imaging techniques provides a means to assess tissue microenvironments with both molecularly specific features and cell location information. Currently, sophisticated multiplexed imaging protocols are being developed, including IMC, CODEX, *t*-CyCIF and MIBI. With rich cellular and neighbourhood information captured, proper analysis of these spatial data has become a new challenge. Traditional analysis methods cluster cells into distinct communities using unsupervised machine learning algorithms, on the basis of the cell-type composition mixtures of their neighbours [7, 29, 60]. However, these strategies

overlook spatial inter-cellular relationships from tissues with different topological structures; hence, they can only provide a highly resolved view of cellular heterogeneity in tissues. This limits higher-order cellular community identification, such as detection of disease relevant areas.

To overcome this challenge, there has been increased interest in applying graph neural networks (GNN) [61, 62] to spatial cell analysis, in which both cell marker expressions and spatial information are taken into consideration [63, 64, 65, 66, 67, 68, 69]. Some methods focus only on tissue-scale classification [63] or cell phenotype annotation [68]; other methods have been developed for microenvironment analysis [64, 65]; and finally, some are directed at general tissue structure classification, through integrating GNN and unsupervised learning algorithms [66, 67]. However, these methods solely focus on either patient-level outcome [63] or cell-scale analysis [64, 65, 66, 67, 68].

Connecting the emerging single-cell rich information with spatially relevant contextual information is a challenge – specifically as it relates to correlating outcomes and cell communities of interest. SPACE-GM [69] solved this issue by training a GNN with tissue-scale labels, then combining trained latent features and K-means clustering to identify disease relevant microenvironments. Though powerful, this framework requires downstream unsupervised clustering to locate potential interested communities related to the corresponding tissue-level outcomes.

2.4 On-Tissue Dose Analysis Techniques

Current small-scale dosimetry methods are predominately based upon idealized computational anatomical models and Monte Carlo simulations [17, 18], often informed by noninvasive imaging. While useful, these provide limited real-world information in heterogeneous patient populations.

To the best of our knowledge, currently there are no direct on-tissue analytical methods for alpha emitter and dose distributions nor the corresponding microenvironment analysis of the treated tissues. In part, this is due to a lack of imaging and analytical tools that enable integration of these orthogonal information sets. As we have shown for resolution-recovered DAR images fused with conventional H&E histologically stained tissues, we achieve greater accuracy in co-registration of these two modalities. Bringing this increased accuracy to bear in molecularly-specific immuno-labeled IMC has the potential to be a transformative advance in understanding how drug distribution is mediated by cellular and extracellular composition of tissues, and how these components respond to localized ionizing radiation.

Chapter 3

PG-PEM: A Blind Image Restoration Algorithm to Enhance Digital Autoradiography

3.1 Introduction

To address the challenges of DAR image restoration, we present PG-PEM in this chapter. We first describe this model in the context of the DAR imaging process along with a patch-based noise parameters estimation method. This approach is incorporated with a penalized maximum-likelihood expectation maximization (MLEM) algorithm to jointly estimate the restored specimen image and corresponding PSF. L_2 and Hessian Frobenius norms are implemented for the PSF and the specimen signal separately to improve the quality of

Content of this chapter is extracted/adapted from the author's published journal article [34].

the restored image. PG-PEM can improve the resolution and contrast of DAR while suppressing the noise more effectively than contemporary blind image restoration approaches. Using this method, we demonstrate significant improvements in both pre- and clinical DAR imaging of diagnostic and therapeutic radiopharmaceuticals.

3.2 PG-PEM Algorithm

3.2.1 DAR imaging model

According to the DAR imaging process, its physical model can be expressed as:

$$\mathbf{R} = \alpha\mathbf{Q} + \mathbf{N}, \mathbf{Q}_p \sim \mathcal{P}[(\mathbf{X} * \mathbf{h})_p + b_p], \mathbf{N}_p \sim \mathcal{N}(0, \sigma_G^2), \quad (3.1)$$

where p is the pixel index ($p \in \{1, 2, \dots, P\}$), \mathbf{R} is the raw DAR image of a tissue, α is a scaling factor corresponding to the gain of the imaging system, \mathbf{X} is the clean image, \mathbf{h} is the PSF, b is the mean of background, $\mathcal{P}[x]$ refers to the Poisson noise with mean x , $\mathcal{N}(0, \sigma_G^2)$ represents Gaussian noise and σ_G is its standard deviation. Because \mathbf{Q} and \mathbf{N} are both random fields in Equation (3.1), \mathbf{R} is also a random field. Thus, we define \mathbf{r} as the available observations of \mathbf{R} .

According to the definition of Poisson process,

$$\mathcal{P}[(\mathbf{X} * \mathbf{h})_p + b_p] = \mathcal{P}[(\mathbf{X} * \mathbf{h})_p] + \mathcal{P}[b_p]. \quad (3.2)$$

When $b > 3$, $\mathcal{P}[b] \approx \mathcal{N}(b, b)$ [70] (Figure 3.1). In the DAR images, b is normally assumed to be spatially invariant around the tissue and larger than 10. Therefore, Equation (3.3) is derived from Equations (3.1) and (3.2),

$$\begin{aligned} \mathbf{R} &= \alpha \mathcal{P}[\mathbf{X} * \mathbf{h}] + \alpha \mathcal{N}(b, b) + \mathcal{N}(0, \sigma_G^2) \\ &= \alpha \mathcal{P}[\mathbf{X} * \mathbf{h}] + \mathcal{N}(\mu_N, \sigma_N^2). \end{aligned} \quad (3.3)$$

where $\mu_N = \alpha b$ and $\sigma_N = \sqrt{\alpha^2 b + \sigma_G^2}$. Notably, the raw image can be splitted into Poisson-distributed signal and Gaussian-distributed noise. To estimate the noise parameters μ_N and σ_N , the background part needs to be extracted.

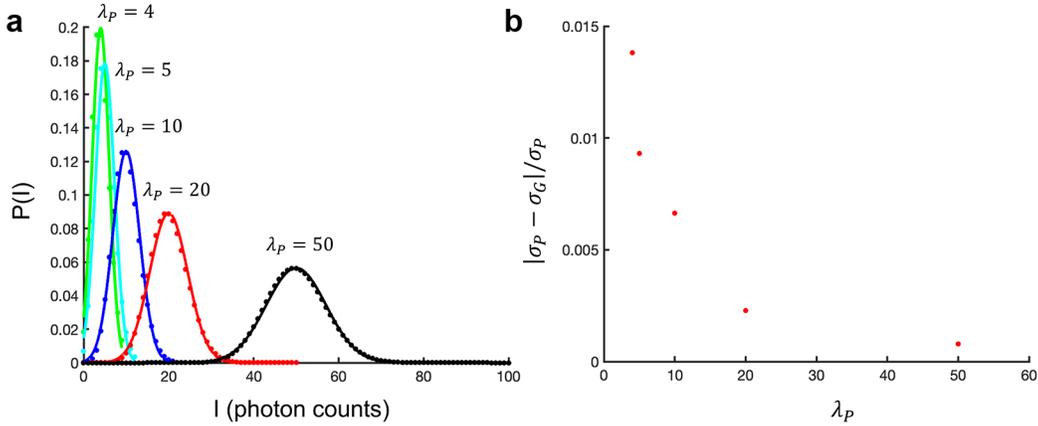


Figure 3.1: Poisson distributions approximated as Gaussian distributions. (a) Poisson-distributed data (dots) for different values of photon flux λ_P and the corresponding Gaussian fitting functions $\mathcal{N}(\lambda_P, \lambda_P)$ (solid lines). (b) The relative error between the two distributions, in which σ_P and σ_G are the standard deviation of the Poisson- and Gaussian-distributed data, respectively. Based on these results, Poisson distribution can be feasibly approximated as a Gaussian distribution when $\lambda_P > 3$.

3.2.2 Patch-based estimation of noise parameters

As Equation (3.3), the raw image can be splitted into Poisson-distributed signal $\alpha\mathcal{P}[\mathbf{X}_p * \mathbf{h}]$ and Gaussian-distributed noise $\mathcal{N}(\mu_N, \sigma_N^2)$. Because of the unknown true PSF and the non-white noise, the frequency domain-based noise parameters estimation method in [47] cannot be used. To estimate the noise parameters, the background part needs to be extracted. Due to the continuity of the histogram of DAR images (Figure 3.2b, h), simple thresholding method based on intensity values will result in signal pixels aberrantly being classified as background, known as false negative in detection theory. In fact, most of the areas without tissues in DAR images normally should not have radioactive signal from the tissues. Thus, these areas should only have background and noise and be highly similar to each other. Based on this assumption, we propose a patch-based estimation algorithm with DBSCAN by searching patches with similar features to robustly estimate the noise parameters. The process is as Algorithm 1 shows.

Algorithm 1 Patch-based estimation of noise parameters

Input: Observation of the raw image, \mathbf{r} ;

Patch size, (M, N) ;

Output: Mean, ab ;

Standard deviation, $\sqrt{\alpha^2 b + \sigma_G^2}$;

- 1: Split \mathbf{r} (Figure 3.2a) into multiple patches with size of M rows and N columns (Figure 3.2c);
 - 2: Calculate the mean, standard deviation, skewness and kurtosis of every patch;
 - 3: Use the Z-score means method [71] to normalize the data and cluster them by DBSCAN (Figure 3.2d);
 - 4: Select the background patches based on the cluster result, and form a new dataset: $\{\mathbf{X}_p | p = 1, 2, \dots, T\}$, where T is the total number of the extracted background pixels;
 - 5: The final values are estimated by MLEM algorithm based on the histograms of the extracted background (Figure 3.2e).
 - 6: **return** $ab, \sqrt{\alpha^2 b + \sigma_G^2}$.
-

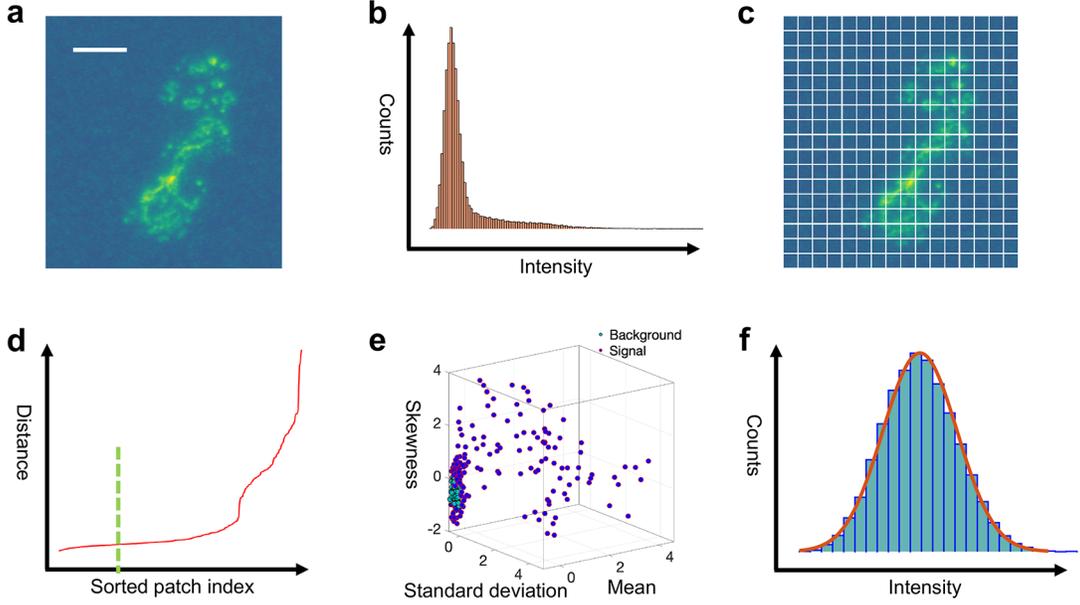


Figure 3.2: Noise parameters estimation. (a) A raw DAR image. (b) The histogram of the raw image. (c) Split the raw image into multiple patches and calculate the mean, standard deviation, skewness and kurtosis of each patch. (d) Plot of the sorted (minPts-1)-th nearest distance of every patch. (e) DBSCAN results for the patches. (f) The histograms of the extracted background field which can be fitted by a Gaussian distribution; while (b) cannot be. The green dotted lines in (d) correspond to the ϵ used in (e). Scale bar: 2.3 mm.

In this algorithm, M and N are usually set as 10 for our dataset. When implementing DBSCAN, its two parameters ϵ and minPts needs to be manually set, in which ϵ is a parameter specifying the radius of a neighborhood with respect to the same point and minPts is the minimum number of points required to form a dense region. As a rule of thumb, minPts is usually twice as the dimension number of the features. Therefore, we set minPts as 8 in PG-PEM. Meanwhile, we use K-nearest neighbour algorithm to determine ϵ . In detail, a K-distance graph is first built based on the data, and then we find every point's (minPts-1)-th nearest distance, sort them in the order from low to high and plot them (Figure 3.2d). Next, unlike the approach [33] which select the “elbow” point as ϵ , we experimentally select the distance before the first 25% position in the range as ϵ , which is far from the “elbow” point for our DAR images. This approach could ensure enough points to form the Gaussian-shape

histogram in Figure 3.2f. Simultaneously, it could also avoid classify signal as background by mistake.

3.2.3 MLEM algorithm for the mixed Poisson-Gaussian model

By simple variable substitutions using Equation (3.4), Equation (3.3) can be simplified as Equation (3.5), where \mathbf{R}' and \mathbf{X} are both 2D matrices with P pixels and \mathbf{h} is a 2D kernel with S pixels. For our DAR images, normally $S \ll P$.

$$\mathbf{R}' = \frac{\mathbf{R}}{\alpha}, \mu'_N = \frac{\mu_N}{\alpha} = b, \sigma'_N = \frac{\sigma_N}{\alpha} = \sqrt{b + \frac{\sigma_G^2}{\alpha^2}}, \quad (3.4)$$

$$\mathbf{R}' = \mathcal{P}[\mathbf{X} * \mathbf{h}] + \mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2}). \quad (3.5)$$

Here we donate \mathbf{Q} as $\mathcal{P}[\mathbf{X} * \mathbf{h}]$ and \mathbf{U} as $\mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2})$, whose elements follow Poisson and Gaussian distributions, respectively. Our goal is to estimate the unknown parameters $\theta = (\mathbf{X}, \mathbf{h})$ with the given parameters \mathbf{r}' (the available observations of \mathbf{R}'), b and $b + \frac{\sigma_G^2}{\alpha^2}$ by following the assumptions for the Poisson-distributed signal and Gaussian-distributed noise in Equation (3.5): 1) they are mutually independent; 2) their components are independent. Under these assumptions, Equation (3.6) is obtained by applying Bayes rule, which is the mixed continuous-discrete probability distribution of $(\mathbf{R}'_p, \mathbf{Q}_p)$ for every $p \in \{1, 2, \dots, P\}$.

$$\begin{aligned} \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{r}'_p, \mathbf{q}_p | \theta) &= \Pr(\mathbf{Q}_p = \mathbf{q}_p | \theta) f_{\mathbf{R}'_p | \mathbf{Q}_p = \mathbf{q}_p}(\mathbf{r}'_p | b, b + \frac{\sigma^2}{\alpha^2}) \\ &= \Pr(\mathbf{Q}_p = \mathbf{q}_p | \theta) f_{\mathbf{U}_p}(\mathbf{r}'_p - \mathbf{q}_p | b, b + \frac{\sigma^2}{\alpha^2}). \end{aligned} \quad (3.6)$$

where $f_{\mathbf{R}'_p|\mathbf{Q}_p=\mathbf{q}_p}(\cdot|b, b + \frac{\sigma^2}{\alpha^2})$ is the conditional probability density function (PDF) of \mathbf{R}'_p knowing that $\mathbf{Q}_p = \mathbf{q}_p$ and $f_{\mathbf{U}_p}(\cdot|b, b + \frac{\sigma^2}{\alpha^2})$ is the PDF of \mathbf{U}_p . In detail,

$$\Pr(\mathbf{Q}_p = \mathbf{q}_p|\theta) = \exp[(-\mathbf{X} * \mathbf{h})_p] \frac{(\mathbf{X} * \mathbf{h})_p^{\mathbf{q}_p}}{\mathbf{q}_p!}, \quad (3.7)$$

$$f_{\mathbf{U}_p}(\mathbf{r}'_p - \mathbf{q}_p|b, b + \frac{\sigma^2}{\alpha^2}) = \frac{1}{(2\pi)^{1/2}(b + \frac{\sigma_G^2}{\alpha^2})^{1/2}} \exp\left(-\frac{(\mathbf{r}'_p - \mathbf{q}_p - b)^2}{2(b + \frac{\sigma_G^2}{\alpha^2})}\right). \quad (3.8)$$

Therefore, the likelihood of Equation (3.5) takes the form as Equation (3.9):

$$\begin{aligned} f_{\mathbf{R}'}(\mathbf{r}'|\theta) &= \prod_{p=1}^P \sum_{\mathbf{q}_p=0}^{+\infty} \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{r}'_p, \mathbf{q}_p|\theta) \\ &= \frac{1}{(2\pi)^{P/2}(b + \frac{\sigma_G^2}{\alpha^2})^{P/2}} \prod_{p=1}^P \exp[(-\mathbf{X} * \mathbf{h})_p] \\ &\quad \sum_{\mathbf{q}_p=0}^{+\infty} \exp\left(-\frac{(\mathbf{r}'_p - \mathbf{q}_p - b)^2}{2(b + \frac{\sigma_G^2}{\alpha^2})}\right) \frac{(\mathbf{X} * \mathbf{h})_p^{\mathbf{q}_p}}{\mathbf{q}_p!}. \end{aligned} \quad (3.9)$$

To solve the parameters θ in Equation (3.9), an iterative MLEM approach is utilized:

$$\theta^{(n+1)} = \operatorname{argmax}_{\theta} J(\theta|\theta^{(n)}), \quad (3.10)$$

where $J(\theta|\theta^{(n)}) = E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}', \theta^{(n)}}[\log \Pr_{\mathbf{R}', \mathbf{Q}}(\mathbf{R}', \mathbf{Q}|\theta)]$ and $\Pr_{\mathbf{R}', \mathbf{Q}}(\mathbf{R}', \mathbf{Q}|\theta) = \prod_{p=1}^P \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{R}'_p, \mathbf{Q}_p|\theta)$.

According to Equation (3.6),

$$\begin{aligned} \log \Pr_{\mathbf{R}', \mathbf{Q}}(\mathbf{R}', \mathbf{Q} | \theta) = & -\frac{1}{2(b + \frac{\sigma_G^2}{\alpha^2})^{P/2}} \sum_{p=1}^P (\mathbf{R}'_p - \mathbf{Q}_p - b)^2 - \frac{P}{2} \log(2\pi(b + \frac{\sigma_G^2}{\alpha^2})) \\ & - \sum_{p=1}^P (\mathbf{X} * \mathbf{h})_p + \sum_{p=1}^P \log(\mathbf{X} * \mathbf{h})_p \mathbf{Q}_p - \sum_{p=1}^P \log(\mathbf{Q}_p!) \end{aligned} \quad (3.11)$$

By dropping the terms that are independent of θ , Equation (3.10) is simplified as:

$$\theta^{(n+1)} = \operatorname{argmin}_{\theta} \tilde{J}(\theta | \theta^{(n)}), \quad (3.12)$$

where

$$\tilde{J}(\theta | \theta^{(n)}) = \sum_{p=1}^N (\mathbf{X} * \mathbf{h})_p - \sum_{p=1}^N \log(\mathbf{X} * \mathbf{h})_p E_{\mathbf{Q}_p | \mathbf{R}'_p = \mathbf{r}'_p, \theta^{(n)}}(\mathbf{Q}_p). \quad (3.13)$$

For every $p \in \{1, 2, \dots, P\}$, we have

$$\begin{aligned} E_{\mathbf{Q}_p | \mathbf{R}'_p = \mathbf{r}'_p, \theta^{(n)}}(\mathbf{Q}_p) &= \sum_{\mathbf{q}_p=0}^{+\infty} \mathbf{q}_p \Pr(\mathbf{Q}_p = \mathbf{q}_p | \mathbf{R}'_p = \mathbf{r}'_p, \theta^{(n)}) \\ &= \frac{\sum_{\mathbf{q}_p=0}^{+\infty} \mathbf{q}_p \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{r}'_p, \mathbf{q}_p | \theta^{(n)})}{f_{\mathbf{R}'_p}(\mathbf{r}'_p | \theta^{(n)})} \\ &= \frac{\sum_{\mathbf{q}_p=0}^{+\infty} \mathbf{q}_p \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{r}'_p, \mathbf{q}_p | \theta^{(n)})}{\sum_{\mathbf{q}_p=0}^{+\infty} \Pr_{\mathbf{R}'_p, \mathbf{Q}_p}(\mathbf{r}'_p, \mathbf{q}_p | \theta^{(n)})}. \end{aligned} \quad (3.14)$$

By combining Equations (3.6)–(3.8), Equation (3.14) simplifies to:

$$E_{\mathbf{Q}_p | \mathbf{R}'_p = \mathbf{r}'_p, \theta^{(n)}}(\mathbf{Q}_p) = \frac{\sum_{\mathbf{q}_p=1}^{+\infty} \exp\left(-\frac{(\mathbf{r}'_p - \mathbf{q}_p - b)^2}{2(b + \sigma_G^2/\alpha^2)}\right) \frac{(\mathbf{X} * \mathbf{h})_p^{\mathbf{q}_p}}{(\mathbf{q}_p - 1)!}}{\sum_{\mathbf{q}_p=0}^{+\infty} \exp\left(-\frac{(\mathbf{r}'_p - \mathbf{q}_p - b)^2}{2(b + \sigma_G^2/\alpha^2)}\right) \frac{(\mathbf{X} * \mathbf{h})_p^{\mathbf{q}_p}}{\mathbf{q}_p!}}. \quad (3.15)$$

Therefore, a 2D matrix $E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})$, sharing the same shape with \mathbf{R}' and \mathbf{X} , is acquired. When implementing Equation (3.15), there is an issue calculating the infinite terms of \mathbf{q}_p on numerator and denominator. Nonetheless, the infinite terms are bounded and the Lambert W Function is used to estimate the terms of summarizations, as described previously [45]. The iteration process is done by differentiating Equation (3.13) with respect to \mathbf{X} and \mathbf{h} and setting the derivative to zero [48, 49, 51]. The resulting iterative scheme is given by alternating:

$$\hat{\mathbf{h}}^{(n+1)} = \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{X}}^{(n),m} \right], \quad (3.16)$$

$$\hat{\mathbf{X}}^{(n+1)} = \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{h}}^{(n),m} \right], \quad (3.17)$$

where $\hat{\mathbf{h}}^{(n),m}$ and $\hat{\mathbf{X}}^{(n),m}$ are the mirrored results of $\hat{\mathbf{h}}^{(n)}$ and $\hat{\mathbf{X}}^{(n)}$, respectively. In addition, $\hat{\mathbf{h}}$ has three constraints: circularly symmetric constraint, non-negativity ($\forall s, \hat{\mathbf{h}}_s > 0$) and norm to 1 ($\sum_{s=1}^S \hat{\mathbf{h}}_s = 1$). To ensure the circularly symmetric constraint, $\hat{\mathbf{h}}$ is averaged in the angular direction as Equation (3.18), where r is the magnitude and ϕ is the phase in the polar coordinate system; To ensure $\sum_{s=1}^S \hat{\mathbf{h}}_s = 1$, Equation (3.19) is conducted; To ensure $\forall s, \hat{\mathbf{h}}_s > 0$, Equation (3.20) is conducted.

$$\hat{\mathbf{h}}'(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\mathbf{h}}(r \cos \phi, r \sin \phi) d\phi. \quad (3.18)$$

$$\hat{\mathbf{h}} = \frac{\hat{\mathbf{h}}'}{\sum_{s=1}^S \hat{\mathbf{h}}'_s}. \quad (3.19)$$

$$\exists \hat{\mathbf{h}}_s < 0, \text{ set } \hat{\mathbf{h}}_s = 0. \quad (3.20)$$

3.2.4 Regularization for \mathbf{h} and \mathbf{X}

In practice, the blind deconvolution problem is highly ill-posed, so regularizations for both \mathbf{h} and \mathbf{X} are needed. Through the iteration process \mathbf{h} tends to converge as a delta function. To avoid this trivial solution and considering the smooth characteristics of \mathbf{h} , \mathbf{h} is regularized by $L2$ norm. The noise of \mathbf{X} may amplify in the iteration process, so total variation (TV) is normally added for \mathbf{X} [51, 54, 72]. Nevertheless, TV penalty always oversharpens the boundaries between different regions, generating “staircase” effect. To avoid this, we adopted a Hessian Frobenius penalty for the estimated \mathbf{X} to enable smoother transitions between different regions and to suppress noise simultaneously [50, 52, 53]. As a result, this leads to a penalized MLEM algorithm as Equation (3.21), where $\lambda_{\mathbf{h}}$ and $\lambda_{\mathbf{X}}$ are the regularization parameters for \mathbf{h} and \mathbf{X} , $P_{\mathbf{h}} = \exp(-\frac{1}{2} \sum_{s=1}^S \mathbf{h}_s^2)$ and $P_{\mathbf{X}} = \exp(-\sum_{p=1}^P |\mathcal{H}\mathbf{X}|_p)$ are the prior probability functions for \mathbf{h} and \mathbf{X} , respectively. Here, \mathcal{H} is a Hessian operator and defined as $[\partial_{xx}, \partial_{xy}; \partial_{xy}, \partial_{yy}]$, where $\partial_{xx} = \partial^2/\partial x^2$, $\partial_{xy} = \partial^2/\partial x\partial y$ and $\partial_{yy} = \partial^2/\partial y^2$. $|\mathcal{H}\mathbf{X}|$ is the Hessian Frobenius (HF) norm for \mathbf{X} and defined as $|\mathcal{H}\mathbf{X}| = \sqrt{(\partial_{xx}\mathbf{X})^2 + (\partial_{yy}\mathbf{X})^2 + 2(\partial_{xy}\mathbf{X})^2}$.

$$\theta^{(n+1)} = \operatorname{argmin}_{\theta} [\tilde{J}(\theta|\theta^{(n)}) + \lambda_{\mathbf{h}} \sum_{s=1}^S \mathbf{h}_s^2 + 2\lambda_{\mathbf{X}} \sum_{p=1}^P |\mathcal{H}\mathbf{X}|_p]. \quad (3.21)$$

The penalized optimization process for \mathbf{h} and \mathbf{X} can both be implemented by a forward-backward splitting algorithm similar to [73]. In [73], this algorithm was originally designed for TV regularization. Nevertheless, since the regularization terms for \mathbf{h} and \mathbf{X} are all convex and the data fidelity term can be extended to a Kullback-Leibler (KL) functional without

affecting the stationary points [74], which is the same as the condition of TV regularization, the algorithm framework still works for our problem. Therefore, the original Equation (3.16) is modified as expectation maximization (EM) step and $L2$ norm regularization step:

$$\begin{cases} \hat{\mathbf{h}}^{(n+\frac{1}{2})} = \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{X}}^{(n),m} \right] & \text{(EM step)} \\ \hat{\mathbf{h}}^{(n+1)} = \operatorname{argmin}_{\mathbf{h}} \left\{ \sum_{s=1}^S \frac{(\mathbf{h}_s - \hat{\mathbf{h}}_s^{(n+\frac{1}{2})})^2}{\hat{\mathbf{h}}_s^{(n)}} + \lambda_{\mathbf{h}} \sum_{s=1}^S \hat{\mathbf{h}}_s^2 \right\} & \text{($L2$ norm regularization step)}. \end{cases} \quad (3.22)$$

Expand the $L2$ norm regularization step:

$$\hat{\mathbf{h}}^{(n+1)} = \operatorname{argmin}_{\mathbf{h}} \left\{ \sum_{s=1}^S \frac{1}{\hat{\mathbf{h}}_s^{(n)}} \left[\left(1 + \lambda_{\mathbf{h}} \hat{\mathbf{h}}_s^{(n)} \right) \hat{\mathbf{h}}_s^2 - 2\hat{\mathbf{h}}_s^{(n+\frac{1}{2})} \hat{\mathbf{h}}_s + \text{const} \right] \right\}. \quad (3.23)$$

By setting the derivative of every $\hat{\mathbf{h}}_s$ to be zero, we get

$$\hat{\mathbf{h}}_s^{(n+1)} = \frac{\hat{\mathbf{h}}_s^{(n+\frac{1}{2})}}{1 + \lambda_{\mathbf{h}} \hat{\mathbf{h}}_s^{(n)}}. \quad (3.24)$$

Therefore, the solution of Equation (3.22) is

$$\hat{\mathbf{h}}^{(n+1)} = \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{X}}^{(n),m} \right] / \left(1 + \lambda_{\mathbf{h}} \hat{\mathbf{h}}^{(n)} \right). \quad (3.25)$$

Likewise, the original Equation (3.17) is modified as EM step and HF norm regularization step:

$$\begin{cases} \hat{\mathbf{X}}^{(n+1)} = \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=r',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{h}}^{(n),m} \right] & \text{(EM step)} \\ \hat{\mathbf{X}}^{(n+1)} = \operatorname{argmin}_{\mathbf{X}} \left\{ \sum_{p=1}^P \frac{(\hat{\mathbf{X}}_p - \hat{\mathbf{X}}_p^{(n+\frac{1}{2})})^2}{\hat{\mathbf{X}}_p^{(n)}} + 2\lambda_{\mathbf{X}} \sum_{p=1}^P |\mathcal{H}\hat{\mathbf{X}}|_p \right\} & \text{(HF norm regularization step)}. \end{cases} \quad (3.26)$$

The HF norm regularization step can be solved by the *majorization – minimization* (MM) framework [50]. Based on this method, the point-wise regularization term can be transformed as Equation (3.27). Its equality holds if and only if $|\mathcal{H}\hat{\mathbf{X}}|_p = |\mathcal{H}\hat{\mathbf{X}}^{(n)}|_p$, which can be achieved when the iteration converges.

$$|\mathcal{H}\hat{\mathbf{X}}|_p \leq \frac{|\mathcal{H}\hat{\mathbf{X}}^{(n)}|_p}{2} + \frac{|\mathcal{H}\hat{\mathbf{X}}|_p^2}{2|\mathcal{H}\hat{\mathbf{X}}^{(n)}|_p}. \quad (3.27)$$

Thus, minimizing the HF norm regularization step in Equation (3.26) can be conducted by minimizing a surrogate function as Equation (3.28), where $\mathcal{W} = [\partial_{xx}, \partial_{yy}, \sqrt{2}\partial_{xy}]^T$.

$$\hat{\mathbf{X}}^{(n+1)} = \operatorname{argmin}_{\mathbf{X}} \left\{ \sum_{p=1}^P \frac{(\hat{\mathbf{X}}_p - \hat{\mathbf{X}}_p^{(n+\frac{1}{2})})^2}{\hat{\mathbf{X}}_p^{(n)}} + \lambda_{\mathbf{X}} \sum_{p=1}^P \frac{|\mathcal{H}\hat{\mathbf{X}}|_p^2}{|\mathcal{H}\hat{\mathbf{X}}^{(n)}|_p} + \text{const} \right\}. \quad (3.28)$$

Expanding Equation (3.28), we have:

$$\hat{\mathbf{X}}^{(n+1)} = \operatorname{argmin}_{\mathbf{X}} \left\{ \sum_{p=1}^P \frac{1}{\hat{\mathbf{X}}_p^{(n)}} \left[\left(1 + 2\lambda_{\mathbf{X}} \left(\mathcal{W}^T \frac{\mathcal{W}\hat{\mathbf{X}}^{(n)}}{|\mathcal{H}\hat{\mathbf{X}}^{(n)}|_p} \right) \right) \hat{\mathbf{X}}_p^2 - 2\hat{\mathbf{X}}_p \hat{\mathbf{X}}_p^{n+\frac{1}{2}} + \text{const} \right] \right\}. \quad (3.29)$$

By setting the derivative of $\hat{\mathbf{X}}_p$ to be zero, Equation (3.28) is solved as Equation (3.30), where $\Xi = \frac{\partial}{\partial x x} + \frac{\partial}{\partial y y} + \sqrt{2} \frac{\partial}{\partial x y}$.

$$\hat{\mathbf{X}}_p^{(n+1)} = \frac{\hat{\mathbf{X}}_p^{(n+1/2)}}{1 + \lambda_{\mathbf{X}} \left(\Xi \frac{\mathcal{W}\hat{\mathbf{X}}^{(n)}}{|\mathcal{H}\hat{\mathbf{X}}^{(n)}|} \right)_p}. \quad (3.30)$$

Therefore, the solution of Equation (3.26) is derived as:

$$\hat{\mathbf{X}}^{(n+1)} = \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)}} \right) * \hat{\mathbf{h}}^{(n),m} \right] / \left[1 + \lambda_{\mathbf{X}} \left(\Xi \frac{\mathcal{W}\hat{\mathbf{X}}^{(n)}}{|\mathcal{H}\hat{\mathbf{X}}^{(n)}|} \right) \right]. \quad (3.31)$$

3.2.5 PSF model

The DAR imaging system involves exposing the plate to the radioactive section which produces a latent image through a trapping of electron-hole (Figure 2.1a), and then scanning a laser over the latent image on the phosphor plate to stimulate release of photons (Figure 2.1b). As discussed in this section, \mathbf{h} is circularly symmetric, also affected by the finite focal point of the laser scanner and the modulation transfer function (MTF) of the phosphor plate. However, in blind restoration it is not needed to consider all these points to initialize a PSF. Instead, we model it based on the scattering effect using the inverse square law [75]. In Figure 2.1a, the energy of one point in the latent image can be expressed as Equation (3.32), where E_S is the total energy of its source S , E_C is the cut-off energy of the phosphor plate. As a result, \mathbf{h} is approximated as Equation (3.33) when initializing, in which a controls its size. In fact, because of the regularization for \mathbf{h} , the initial value of a does not have a good impact on the result. In our application, we set a as 1.

$$E_{S'} = \begin{cases} \frac{E_S}{4\pi[(u'-u_o)^2+(v'-v_o)^2+d^2]} & \text{when } E_{S'} > E_C \\ 0 & \text{else} \end{cases}. \quad (3.32)$$

$$\mathbf{h} = \begin{cases} \frac{1}{a(u^2+v^2)+1} & \text{when } E_{S'} > E_C \\ 0 & \text{else} \end{cases}. \quad (3.33)$$

3.2.6 The impact of the scaling factor α

According to Equations (3.1), (3.4) and (3.5), the scaling factor α needs to be pre-calibrated. However, this process is very time-consuming and may not be robust. Here, we investigate the impact of α on the restoration result. Assuming α is not calibrated correctly and the calibrated scaling factor is $\alpha^* = \beta\alpha$. Then Equation (3.5) is transformed as Equation (3.34), where \mathbf{X}' is the estimated image under this condition.

$$\frac{\mathbf{R}'}{\beta} = \mathcal{P}[\mathbf{X}' * \mathbf{h}] + \frac{\mathcal{N}(b, b + \frac{\sigma_G^2}{\alpha^2})}{\beta}. \quad (3.34)$$

By using Poisson distribution to estimate Gaussian distribution, Equation (3.34) is estimated as a shift-Poisson format [53] as Equation (3.35), in which b/β and $(b + \frac{\sigma_G^2}{\alpha^2})/\beta^2$ can both be estimated from \mathbf{R}'/β with Algorithm 1.

$$\frac{\mathbf{R}' - b}{\beta} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} = \mathcal{P}[\mathbf{X}' * \mathbf{h} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}]. \quad (3.35)$$

Based on Equation (3.35), the likelihood of the Poisson statistics is as Equation (3.36), where $\mathbf{k}_p = \frac{\mathbf{r}'_p - b}{\beta} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2}$ ($p \in \{1, 2, \dots, P\}$).

$$f_{\mathbf{K}}(\mathbf{k}|\theta) = \prod_{p=1}^P \exp \left[-(\mathbf{X}' * \mathbf{h})_p - \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} \right] \frac{\left[(\mathbf{X}' * \mathbf{h})_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} \right]^{\mathbf{k}_p}}{\mathbf{k}_p!}. \quad (3.36)$$

Therefore, the negative log-likelihood of Equation (3.36) is:

$$-\log f_{\mathbf{K}}(\mathbf{k}|\theta) = \sum_{p=1}^P \left\{ (\mathbf{X}' * \mathbf{h})_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} - \mathbf{k}_p \log \left[(\mathbf{X}' * \mathbf{h})_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} \right] + \log(\mathbf{k}_p!) \right\}. \quad (3.37)$$

By setting the derivative of every $(\mathbf{X}' * \mathbf{h})_p$ to be 0, we have:

$$1 - \left[(\mathbf{X}' * \mathbf{h})_p + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} \right] / \left(\frac{\mathbf{r}'_p - b}{\beta} + \frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} \right) = 0. \quad (3.38)$$

Under the MLEM iterative framework, the result of Equation (3.38) can be expressed as Equation (3.39), where $E(\mathbf{r}'_p)$ is the expectation of \mathbf{r}'_p .

$$(\mathbf{X}' * \mathbf{h})_p = \frac{E(\mathbf{r}'_p) - b}{\beta}. \quad (3.39)$$

Thus, when the estimated PSF \mathbf{h} is the same, the relationship between the results using α^* and α can be approximated as Equation (3.40),

$$\mathbf{X} = \beta \mathbf{X}'. \quad (3.40)$$

Based on these results, the scaling factor α does not have an appreciable impact on the restoration result except as a multiplier parameter β . The reason is that the Gaussian-distributed noise parameters μ_N and σ_N are estimated directly from the background, and this condition is quite similar to that of pure Poisson noise, whose scaling parameter is not required to be calibrated. In fact, digital light unit (DLU) itself is meaningless. When implementing dosimetry calculation, several phantoms should be built for calibrating the mapping relationship between DLU and dosimetry values. In this sense, α will not affect the result of Equation (3.15) and is not needed to be pre-calibrated for DAR image restoration if Equations (3.34) and (3.35) are approximately equal to each other. However, when α^* is too large, the computation process in Equation (3.15) will generate sampling errors, especially for the low intensity regions of the images. To maintain this accurate Gaussian-Poisson (continuous to discrete) transformation, we should make sure $\frac{b + \frac{\sigma_G^2}{\alpha^2}}{\beta^2} = \frac{b}{\beta^2} + \left(\frac{\sigma_G}{\alpha^*}\right)^2 \gg 1$. Therefore, we empirically set $\frac{\sigma_G}{\alpha^*} \geq 10$. In real applications, we can directly use α^* to replace α . Before the restoration process, the raw image could be divided by a large α to decrease the number of summations in Equation (3.15) and then multiply the same α after estimation to maintain the final result with the same scale with the raw image. Then, we set $\alpha \leq \frac{\sigma_G}{10}$. In practice, we set α as $\frac{\sigma_G}{10}$ considering both the computation efficiency and the accuracy of PG-PEM.

3.2.7 PG-PEM algorithm summary

The algorithm is summarized as Figure 3.3 and Algorithm 2.

3.2.8 Runtime

All the codes for PG-PEM were written in Matlab 2019a (MathWorks). The runtime of PG-PEM mainly depends on the image size, the number of summation terms in Equation

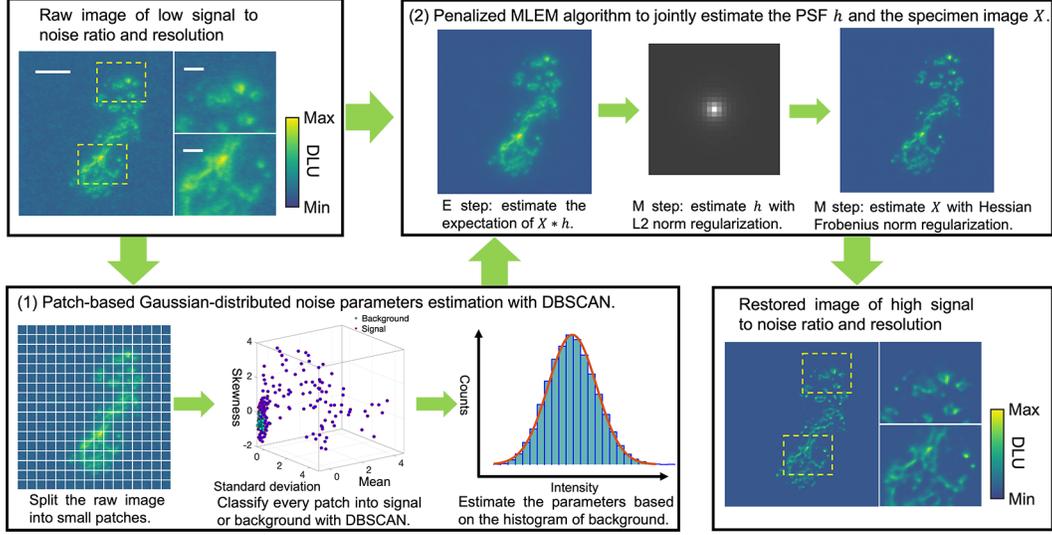


Figure 3.3: PG-PEM algorithmic framework summary. (1) Noise parameters estimation. (2) PSF and specimen image estimation. Scale bar: big figure: 2.3 mm, small figures: 0.54 mm.

Algorithm 2 PG-PEM algorithm

Initialization:

Estimate ab and $\alpha^2 b + \sigma_G^2$ using Algorithm 1.

Rescale \mathbf{r} , ab and $\alpha^2 b + \sigma_G^2$ to \mathbf{r}' , b and $b + \sigma_G^2/\alpha^2$ using the pre-set α and (3.4).

Set the raw image observation \mathbf{r}' as the initial \mathbf{X} .

Initialize the PSF \mathbf{h} using Equation (3.33).

Set the regularization parameters $\lambda_{\mathbf{X}}$ and $\lambda_{\mathbf{h}}$.

Iteration:

- 1: **for** each $i = 1, 2, \dots, N$ **do**
 - 2: E step: Use Equation (3.15) to estimate $E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})$.
 - 3: M step: Use Equation (3.31) to estimate \mathbf{X} .
 - 4: Use Equation (3.25) to estimate \mathbf{h} .
 - 5: Use Equations (3.18), (3.19) and (3.20) to normalize \mathbf{h} .
 - 6: **if** $\sqrt{\sum_{p=1}^P (\hat{\mathbf{X}}_p^{(n+1)} - \hat{\mathbf{X}}_p^{(n)})^2} / \sqrt{\sum_{p=1}^P (\hat{\mathbf{X}}_p^{(n)})^2} < \text{threshold}$ **then**
 - 7: break
 - 8: **end if**
 - 9: **end for**
-

(3.15) and the value of the threshold in Algorithm 2. In our experiments, we typically set the value of the threshold from 0.0005 to 0.001. In such a range, the iteration numbers are usually around 100. Here we report the performance of our software under various image

sizes both with and without graphics processing unit (GPU, NVIDIA Quadro RTX 6000) on a workstation using 12 cores Intel(R) Xeon(R) W-2133, 3.60GHz central processing unit (CPU). Specifically, we set the iteration numbers as 100 for all the groups. Besides, the number of summation terms in Equation (3.15) is set as around 700, which can satisfy the criteria in Subsection 3.2.6. The results are presented in Table 3.1.

Table 3.1: Computational time of PG-PEM in different conditions

Image Size (pixel)	CPU (s)	GPU (s)
250×250	12.3	42.6
500×500	85.5	62.6
750×750	193.6	90.7

From the table, when the image size is small, CPU-based restoration runs fast. However, as the image sizes grow larger, GPU-based restoration performs better than CPU. Fortunately, the size of all the DAR images utilized in research practice are smaller than 500×500 pixels (pixel size: 0.042×0.042 mm²). Therefore, CPU can fulfill most of the requirements. Notably, PG-PEM could be slower than the algorithm based on the shifted-Poisson model [53] due to the computation of Equation (3.15). Nevertheless, PG-PEM achieves much better denoising performance especially for the low SNR images. Meanwhile, compared to the very long exposure time of the DAR imaging process (several hours to several days), our PG-PEM algorithm still runs very fast (from several seconds to less than 2 minutes).

3.3 Reference DAR Restoration Methods

We modified five restoration algorithms as referenced to blind restore DAR images, namely Richardson-Lucy (RL) [48], RL with wavelet-based residue denoising (RD) [76], Shift-Poisson (SP) [53], PG-PEM with no regularization for \mathbf{X} (NP) and PG-PEM with TV regularization

(TV). Note that RL, RD and SP have the same regularization strategies for both \mathbf{h} and \mathbf{X} with PG-PEM, while NP and TV have the same regularization strategy for \mathbf{h} . Further, similar to PG-PEM, the scaling parameter α here does not impact the estimation result except a multiplier parameter.

3.3.1 RL algorithm

RL algorithm is fit for Poisson-distributed data. For DAR image restoration problem, it assumes the imaging model is:

$$\mathbf{R} = \alpha \mathcal{P}[\mathbf{X} * \mathbf{h} + b]. \quad (3.41)$$

Based on this model, Equation (3.42) demonstrates the iterative deconvolution scheme without regularization.

$$\begin{aligned} \hat{\mathbf{h}}^{(n+1)} &= \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{\mathbf{r}'}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b} \right) * \hat{\mathbf{X}}^{(n),m} \right], \\ \hat{\mathbf{X}}^{(n+1)} &= \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{\mathbf{r}'}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b} \right) * \hat{\mathbf{h}}^{(n),m} \right]. \end{aligned} \quad (3.42)$$

3.3.2 RD algorithm

Different from RL, RD algorithm performs residual denoising during each iteration. In [49, 51], the authors utilized median filter as the denoising algorithm. Not the same as them, wavelet denoising algorithm [77] is applied here. The basic scheme of RD is as Equation (3.43), where Denoise() represents wavelet denoising algorithm.

$$\begin{aligned}
\mathbf{r}'^{(n)} &= \mathbf{r}' - \hat{\mathbf{h}} * \hat{\mathbf{X}}^{(n)}, \\
\overline{\mathbf{r}'^{(n)}} &= \text{Denoise}(\mathbf{r}'^{(n)}), \\
\hat{\mathbf{h}}^{(n+1)} &= \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + \overline{\mathbf{r}'^{(n)}}}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b} \right) * \hat{\mathbf{X}}^{(n),m} \right], \\
\hat{\mathbf{X}}^{(n+1)} &= \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + \overline{\mathbf{r}'^{(n)}}}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b} \right) * \hat{\mathbf{h}}^{(n),m} \right].
\end{aligned} \tag{3.43}$$

In the wavelet denoising algorithm, the input image is decomposed for 7 levels, and Stein's Unbiased Risk Estimate (SURE) and a soft-thresholding approach are conducted.

3.3.3 SP algorithm

As discussed in Section 3.2.6, Equation (3.5) can be transformed to a shifted-Poisson distribution:

$$\mathbf{R}' + \frac{\sigma_G^2}{\alpha^2} = \mathcal{P}[\mathbf{X} * \mathbf{h} + b + \frac{\sigma_G^2}{\alpha^2}]. \tag{3.44}$$

Based on this equation, the iteration scheme can be conducted as Equation (3.45).

$$\begin{aligned}
\hat{\mathbf{h}}^{(n+1)} &= \hat{\mathbf{h}}^{(n)} \cdot \left[\left(\frac{\mathbf{r}' + \frac{\sigma_G^2}{\alpha^2}}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b + \frac{\sigma_G^2}{\alpha^2}} \right) * \hat{\mathbf{X}}^{(n),m} \right], \\
\hat{\mathbf{X}}^{(n+1)} &= \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{\mathbf{r}' + \frac{\sigma_G^2}{\alpha^2}}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}^{(n)} + b + \frac{\sigma_G^2}{\alpha^2}} \right) * \hat{\mathbf{h}}^{(n),m} \right].
\end{aligned} \tag{3.45}$$

3.3.4 NP and TV algorithms

Note that NP and TV have almost the same framework with PG-PEM except the regularization for \mathbf{X} . We aim to show the competitive performance of Hessian Frobenius norm regularization in DAR images by comparing it with NP and TV. NP does not have regularization for \mathbf{X} while TV algorithm utilizes TV norm as its regularization for \mathbf{X} as Equation (3.46), in which ∇ represents $[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ and $\text{div} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$.

$$\hat{\mathbf{X}}^{(n+1)} = \hat{\mathbf{X}}^{(n)} \cdot \left[\left(\frac{E_{\mathbf{Q}|\mathbf{R}'=\mathbf{r}',\theta^{(n)}}(\mathbf{Q})}{\hat{\mathbf{h}}^{(n)} * \hat{\mathbf{X}}_p^{(n)}} \right) * \hat{\mathbf{h}}^{(n),m} \right] / \left[1 - \lambda_{\mathbf{X}} \text{div} \left(\frac{\nabla \hat{\mathbf{X}}^{(n)}}{|\nabla \hat{\mathbf{X}}^{(n)}|} \right) \right]. \quad (3.46)$$

3.4 Accuracy Metrics and Statistical Analysis

The root mean squared error (RMSE) [78, 79] is an pixel-wise difference between two inputs, where the ideal value is zero. It is computed as Equation (3.47), where \mathbf{Y}^{est} is the estimated image, \mathbf{Y}^{true} is the ground truth, p is the pixel index and P is the total pixel number for every image.

$$\text{RMSE}(\mathbf{Y}^{\text{est}}, \mathbf{Y}^{\text{true}}) = \sqrt{\frac{1}{P} \sum_{p=1}^P (\mathbf{Y}_p^{\text{est}} - \mathbf{Y}_p^{\text{true}})^2}. \quad (3.47)$$

The SNR indicates the ratio of the power of a signal to the power of background noise. It is defined as Equation (3.48), where \mathbf{Y}^{est} is the estimated image from restoration algorithms and \mathbf{Y}^{true} is the ground truth.

$$\text{SNR}(\mathbf{Y}^{\text{est}}, \mathbf{Y}^{\text{true}}) = 10 \log \frac{\sum_{p=1}^P (\mathbf{Y}_p^{\text{true}})^2}{\sum_{p=1}^P (\mathbf{Y}_p^{\text{est}} - \mathbf{Y}_p^{\text{true}})^2} \quad (3.48)$$

The structure similarity [80] (SSIM) is a perception-based model that considers image degradation as perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms. Compared to RMSE and SNR, it is supposed to give more information about image distortion by the computation of local image structure, luminance and contrast into a single local quality score. In this section, the luminance and contrast are normalized and SSIM is defined as Equation (3.49),

$$\text{SSIM}(\mathbf{Y}^{\text{est}}, \mathbf{Y}^{\text{true}}) = \frac{2\mu_{\mathbf{Y}^{\text{est}}}\mu_{\mathbf{Y}^{\text{true}}} + C_1}{\mu_{\mathbf{Y}^{\text{est}}}^2 + \mu_{\mathbf{Y}^{\text{true}}}^2 + C_1} \cdot \frac{2\sigma_{\mathbf{Y}^{\text{est}}\mathbf{Y}^{\text{true}}} + C_2}{\sigma_{\mathbf{Y}^{\text{est}}}^2 + \sigma_{\mathbf{Y}^{\text{true}}}^2 + C_2} \quad (3.49)$$

where \mathbf{Y}^{est} is the estimated image from restoration algorithms, \mathbf{Y}^{true} is the ground truth, $\mu_{\mathbf{Y}^{\text{est}}}$, $\mu_{\mathbf{Y}^{\text{true}}}$, $\sigma_{\mathbf{Y}^{\text{est}}}$, $\sigma_{\mathbf{Y}^{\text{true}}}$ and $\sigma_{\mathbf{Y}^{\text{est}}\mathbf{Y}^{\text{true}}}$ are the local means, standard deviations and cross-covariance for images \mathbf{Y}^{est} and \mathbf{Y}^{true} , C_1 and C_2 are the regularization constants to avoid instability for image regions where the local mean or standard deviation is close to zero.

Contrast to noise ratio (CNR) is defined as Equation (3.50),

$$\text{CNR} = (C_{\text{sig}} - C_{\text{bg}}) / \sigma_{\text{bg}} \quad (3.50)$$

where C_{sig} and C_{bg} are the mean of the signal and background and σ_{bg} is the standard deviation of the background. In this metric, the signal area is defined as the high activity region in the DAR images, and the background is extracted by our patch-based estimation method.

Effective resolution is estimated based on the recently published decorrelation-based method [81]. This method estimates the highest frequency with sufficiently high signal in relation to noise (Figure 3.4). Quantitative data are presented as box-and-whisker plots (center line,

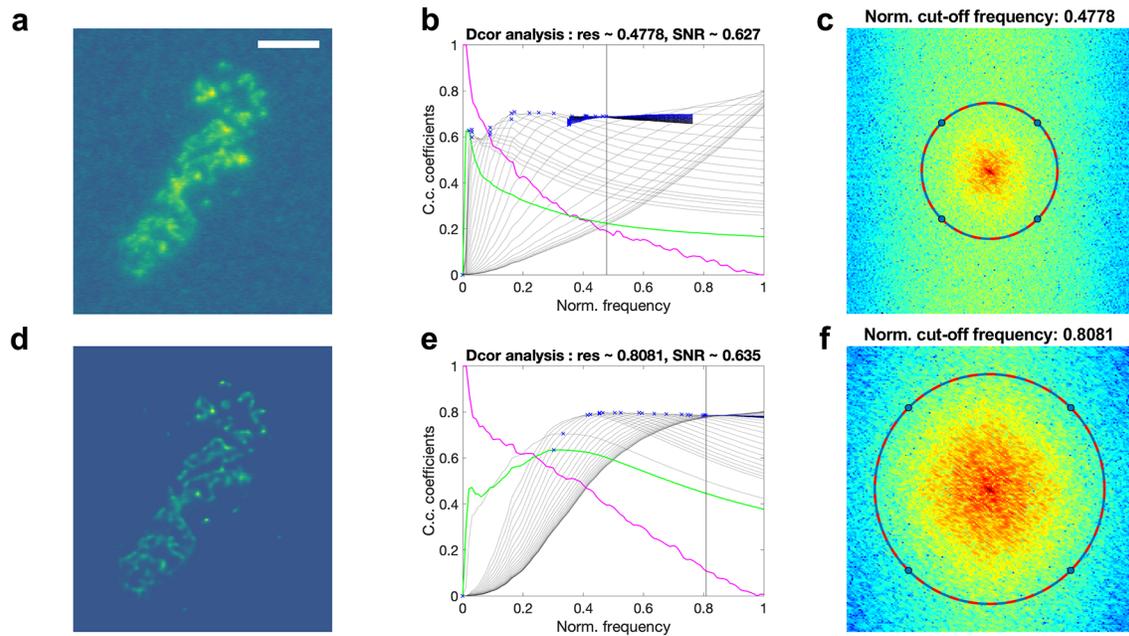


Figure 3.4: Effective resolution estimation by decorrelation analysis. (a) A raw DAR images with low noise level. (d) The corresponding PG-PEM restored image. (b) and (e) The corresponding decorrelation analysis of (a) and (d). (c) and (f) The log-scale frequency map of (a) and (d) labeled with cut-off frequency estimated from (b) and (e), respectively. Scale bar: 1.8 mm.

median; limits, 75% and 25%; whiskers, maximum and minimum). We used the paired two-side Student's t-test to compare the data of Raw and PG-PEM restored DAR images, and the paired one-way analysis of variation to compare all the other data with Prism 8 (GraphPad Software Inc.). Statistical significance at $P < 0.05$, 0.01, 0.001 and 0.0001 are denoted by *, **, *** and ****, respectively.

3.5 Simulations

3.5.1 Simulated data generation

We use Equation (3.51) to generate simulated data, in which p is the pixel index, \mathbf{X} the ground truth, \mathbf{h} the pre-set PSF, b the background, α the scaling factor and $\mathcal{N}(0, \sigma_G^2)$ the Gaussian noise with mean of 0 and standard deviation of σ_G . Here, α and σ_G control the level of Poisson and Gaussian noises, respectively. Note that we divide $\mathbf{X} * \mathbf{h} + b$ with α to ensure the generated images with the same range with the ground truth.

$$\mathbf{R} = \alpha \mathcal{P} \left[\frac{\mathbf{X} * \mathbf{h} + b}{\alpha} \right] + \mathcal{N}(0, \sigma_G^2) \quad (3.51)$$

Because it is impossible to acquire a “clean” raw DAR image without noise and blurring effect, the ground truth image needs to be generated. To do so, we selected a raw DAR image with high SNR and low blurring effect. Then, it was blindly restored by PG-PEM. The regularization parameters were carefully tuned so that the restored image could achieve its best quality. After restoration, the background of the image was cleared to further remove background noise. The pre-set PSF was generated using a more blurred raw DAR image so that the kernel size of the PSF is larger. To do this, we aim to better test the deblurring ability of the restoration algorithms. The generated ground truth image and PSF are both shown in Figure 3.5a.

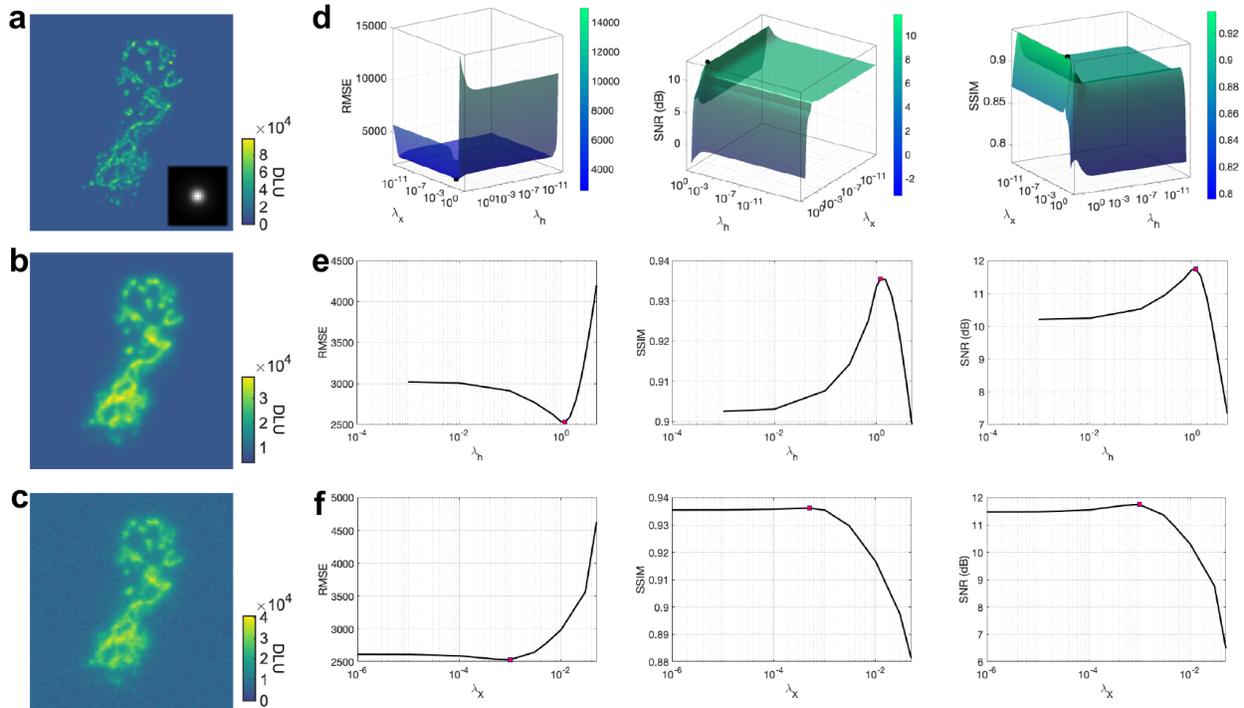


Figure 3.5: Data generation for simulation and accuracy metrics for different parameters. (a) The simulated ground truth image and the corresponding PSF. (b) Blurred image with background. (c) Noisy image. (d) RMSE, SSIM and SNR for all the parameters. (e) RMSE, SSIM and SNR with different λ_h when $\lambda_x = 0.001$. (f) RMSE, SSIM and SNR with different λ_x when $\lambda_h = 1.2$. Note that in the figures the dots indicate the optimal values.

3.5.2 Impact of the regularization parameters

To test the impact of regularization parameters, different λ_x (0 to 0.1) and λ_h (0 to 10) were selected to test the performance of the algorithm using a simulated image. To generate the image, the ground truth image was convoluted with the pre-set PSF, to which a constant background was added ($b = 4000$) and then corrupted with Poisson noise ($\alpha = 20$) and Gaussian noise with standard deviation (σ_G) of 1500. The generated image (Figure 3.5c) was restored using the PG-PEM algorithm with the different regularization parameters. The results are shown in Figure 3.5d–f. Meanwhile, several restoration results and their corresponding PSFs with different regularization parameters are shown in Figure 3.6.

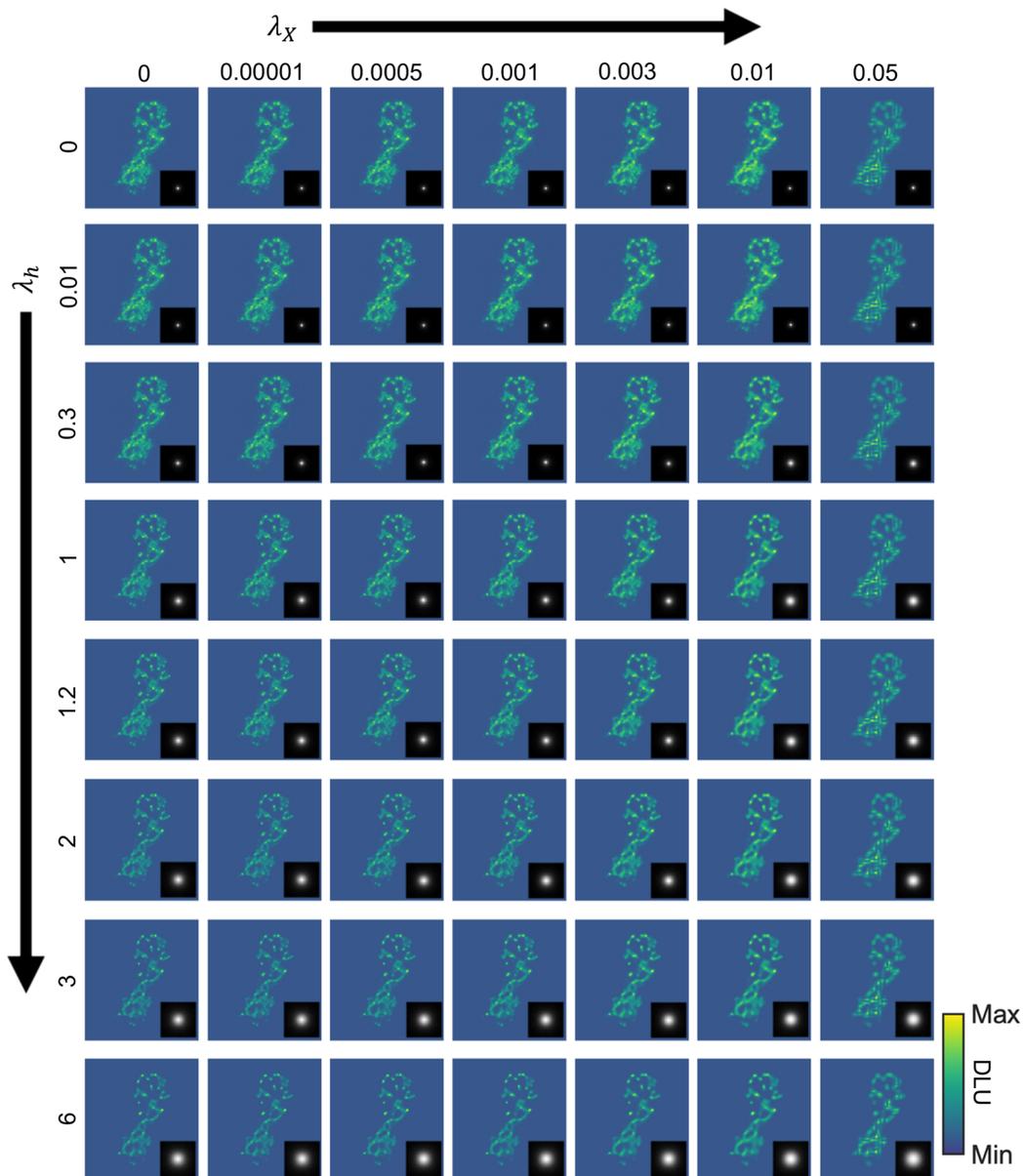


Figure 3.6: Restoration images and corresponding PSFs using PG-PEM algorithm with λ_h from 0 to 10 and λ_X from 0 to 0.1. With λ_h increasing, the kernel size of the estimated PSF become larger and larger, and the restored images tend to become less noisy. But when the PSF kernel is large enough, the restored images become blurry. With λ_X increasing, the image noise is also suppressed. But with too large λ_X , the images lose fine details and become blurry.

From the results, the parameters $\lambda_{\mathbf{h}}$ and $\lambda_{\mathbf{X}}$ control the strength of regularization on the PSF \mathbf{h} and specimen image \mathbf{X} , respectively. With larger $\lambda_{\mathbf{h}}$ and $\lambda_{\mathbf{X}}$, the restored images tend to become less noisy. But when they are large enough, the restored images will lose fine details and become blurry. Therefore, it is essential to set the two regularization parameters properly in order to restore the raw images with a good quality. Empirically, we set $\lambda_{\mathbf{X}}$ around 0.001 (0.0005 to 0.0015). Different from $\lambda_{\mathbf{X}}$, $\lambda_{\mathbf{h}}$ controls the size of the PSF. Because of the existence of the noise, the PSF tends to amplify its high frequency portion during the iteration. The higher noise level, the more high frequencies the PSF amplifies. With the support from the high frequency portion of the PSF, the high frequency noises cannot be suppressed effectively. Therefore, without the regularization for \mathbf{h} , the restored results are still very noisy. In order to reduce the impact of noise on the estimation process, $\lambda_{\mathbf{h}}$ should be positively correlated to the noise level of the raw images. For our DAR images, the parameter $\lambda_{\mathbf{h}}$ is normally set between 0 and 4, determined by the noise level of the input images. In our work, we define the noise level as the reciprocal of the CNR of the images.

$$\text{Noise level} = \sigma_{\text{bg}} / (C_{\text{sig}} - C_{\text{bg}}), \quad (3.52)$$

where C_{sig} and C_{bg} are the mean of the signal and background and σ_{bg} is the standard deviation of the background. The same is done for CNR; the signal area is defined as the high activity region in the DAR images, and the background is extracted by our patch-based estimation method. Reasonably, this can reflect the normalized noise after considering both the noise power and the averaged signal energy. We also found that $\lambda_{\mathbf{h}}$ grows slower with the noise level increasing because the E step in Equation (3.15) can also suppress parts of the noise. We thus empirically set $\lambda_{\mathbf{h}}$ as:

$$\lambda_{\mathbf{h}} = 4\sqrt{\text{Noise level}}. \quad (3.53)$$

In fact, according to the results in Figures 3.5 and 3.6, the restored data are quite similar along a range of $\lambda_{\mathbf{h}}$, which brings more flexibility to our empirical setting of $\lambda_{\mathbf{h}}$ in Equation (3.53). We can also manually tune the $\lambda_{\mathbf{h}}$, for example, by visual assessment or with the help of the decorrelation analysis method [81]. The most optimal $\lambda_{\mathbf{h}}$ should correspond to the point with the best signal-noise trade-off.

3.5.3 Comparison with reference algorithms

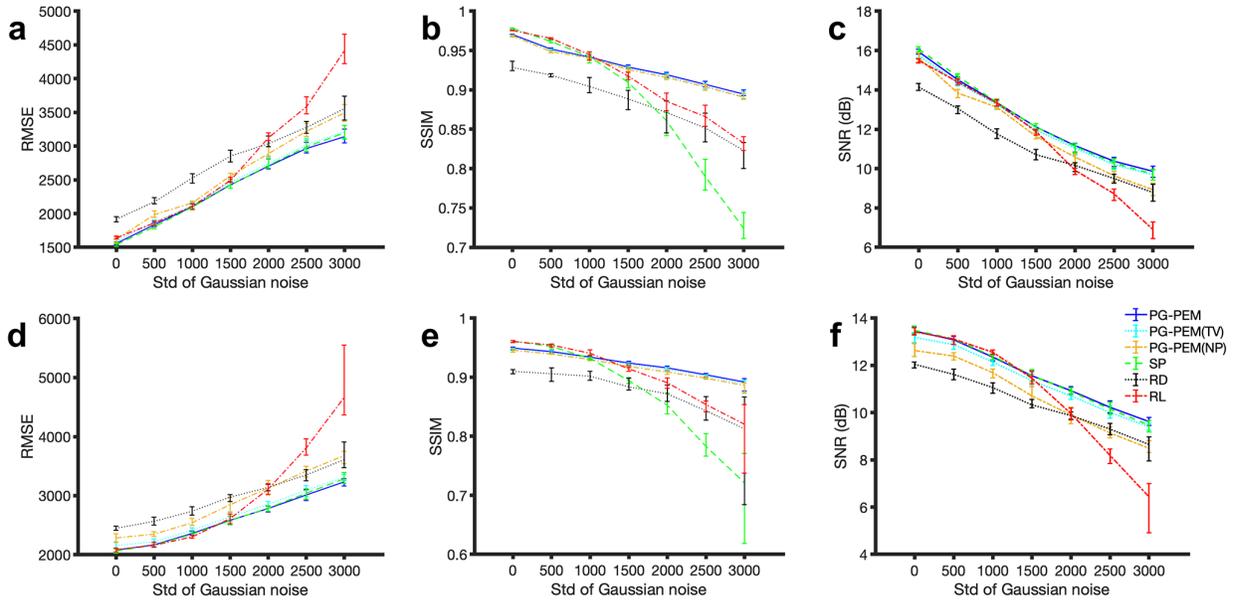


Figure 3.7: Evaluation results for different parameters and methods when $\alpha = 20$ and 100. (a) RMSE evaluation when $\alpha = 20$. (b) SSIM evaluation when $\alpha = 20$. (c) SNR evaluation when $\alpha = 20$. (d) RMSE evaluation when $\alpha = 100$. (e) SSIM evaluation when $\alpha = 100$. (f) SNR evaluation when $\alpha = 100$. From the figures, PG-PEM is the best performer among all the conditions regarding RMSE, SSIM and SNR.

Next, to evaluate the denoising performance of PG-PEM algorithm, the simulated image had added to it a constant background ($b = 4000$), blurred by the same PSF again and corrupted with two different levels of Poisson noise ($\alpha = 20$ and 100) and Gaussian noise with different standard deviations (σ_G) from 0 to 3000 with interval of 500. We generated 10 groups of data for each noise level. In this way, by generating simulated images with different Poisson and Gaussian noise levels, we aim to have a thorough comparison of the algorithms under different conditions. The PG-PEM algorithm was compared with RL, RD, SP, NP and TV.

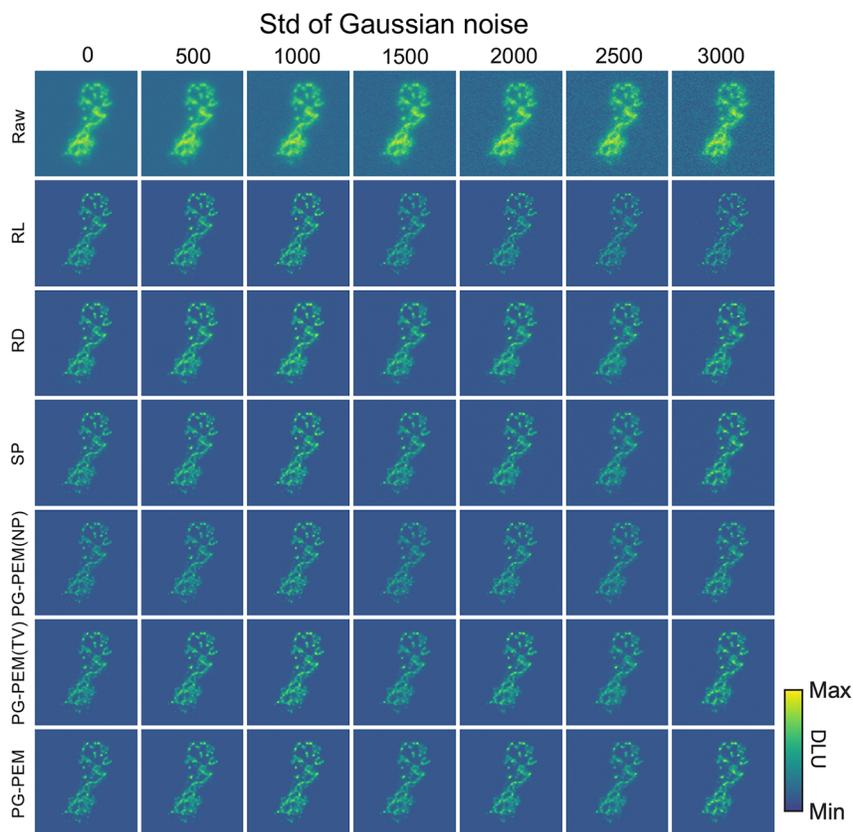


Figure 3.8: Restoration results from different noise levels and methods when $\alpha = 20$.

In this simulation, $\lambda_{\mathbf{X}}$ is set as 0.001. RL, RD and SP shared the same $\lambda_{\mathbf{X}}$ but different $\lambda_{\mathbf{h}}$ to make sure the PSFs of these algorithms have the same shape. We did this to eliminate the

impact of PSF and only focused on the different frameworks of these models. RMSE, SSIM and SNR were used to compare the restoration performance between different algorithms. The results are shown in Figure 3.7.

$$\mathbf{R}' = \mathcal{P}[\mathbf{X} * \mathbf{h}] + \gamma \mathcal{P}(b/\gamma) \quad (3.54)$$

From Figure 3.7, in the parametric space where the standard deviation of the Gaussian noise

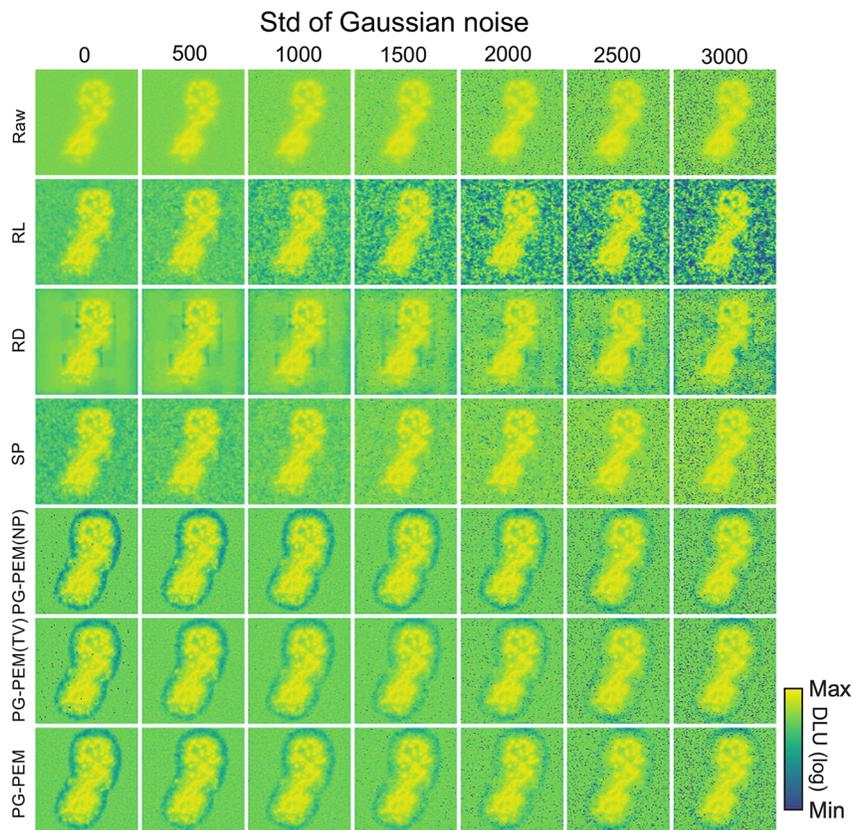


Figure 3.9: Restoration results from different noise levels and methods when $\alpha = 20$ (log scale). It is easier to compare the denoising ability between different methods using the log-scale images.

is small, RL outperforms RD and is similar to SP and PG-PEM. However, with additional

noise, RD algorithm performs better than RL. This can be interpreted using Equation (3.54) estimated from Equation (3.5), where $(\gamma - 1)b = \sigma_G^2/\alpha^2$. When σ^2 is small, the Gaussian noise part can be approximated as a Poisson distribution; while σ^2 is large enough, $\frac{b}{\gamma} < 10$, the approximation will cause large errors. This is why as Gaussian noise increases above a level, the performance of RL decreases dramatically. Parts of the restored images with their log-scale versions are shown in Figures 3.8–3.9. The same as the results from the accuracy metrics, PG-PEM outperforms alternative methods by means visual inspection.

3.6 Experiments

3.6.1 Material preparation

Mice tumor, heart and femur preparation

All radioisotope handling, tissue processing and animal experiments were performed in accordance with approved protocols from the Environmental Health and Safety and Division of Comparative Medicine of Washington University in St. Louis. Male C57Bl/6 mice (6–10 weeks) were acquired from Jackson Laboratory (Bar Harbor, Maine) for animal experiments. For positron emission tomography (PET) tracer analyses, animals were administered approximately 200 μCi (7.4 MBq) of either ^{18}F -FDG or ^{18}F -NaF. For radioglucose, animals were kept sedated following administration. After 1 hour, animals were imaged by PET (R4, Concorde Microsystems) and then sacrificed by CO_2 asphyxiation. Tissues were removed immediately, embedded in optimal cutting temperature media, frozen on dry ice, and sectioned at 8 μm by cryostat (CM188, Leica). For all radiographic exposures, MS phosphor plates were exposed at $-20\text{ }^\circ\text{C}$ and read using a Cyclone Plus (Perkin Elmer). We then used

ImageJ software [82] (National Institute of Health) to crop the region of interest (ROI) of the scanned images.

Human bone biopsy preparation

Bone biopsies were obtained from seven mCRPC patients under fluoroscopic guidance following a bone scan, 24 hours after injection of ^{223}Ra . Determination of the location of biopsy was done in collaboration with the interventional radiologist. The biopsy were fixed in 4% paraformaldehyde for 24 hours and transferred in 30% sucrose for 24 hours. The biopsy were frozen in OCT and cryosectioned. Each section is 8 μm thick. Autoradiography was performed on the sections using phosphor sheet. The sections were exposed between 60 and 90 hours.

H&E staining and slides scanning

All the sections were finally stained with H&E. Stained sections were mounted under coverslip with 30% glycerol. All the slides were finally scanned using bright field microscopes with 10 \times objective (Nikon Eclipse Ti2 for the mice tumor, heart and femur slides; Zeiss Axio Scan Z1 Brightfield/Fluorescence Slide Scanner for human bone biopsy slides).

3.6.2 Assessment of image enhancement on experimental data

Apart from the simulations, we have also benchmarked performance of the proposed restoration frameworks on the experimental data. DAR images ($N = 10$) acquired from the mouse hindlimb following ^{18}F -NaF PET imaging were used as experimental data and to evaluate the performance of image restoration approaches. As is standard for short lived diagnostic radioisotopes and the required tissue-processing, sectioning and exposure times, the SNR

of the raw images are low; providing a model setting for comparison. Visual assessment and analyses (Figure 3.10) show implementation of restoration algorithms improved reso-

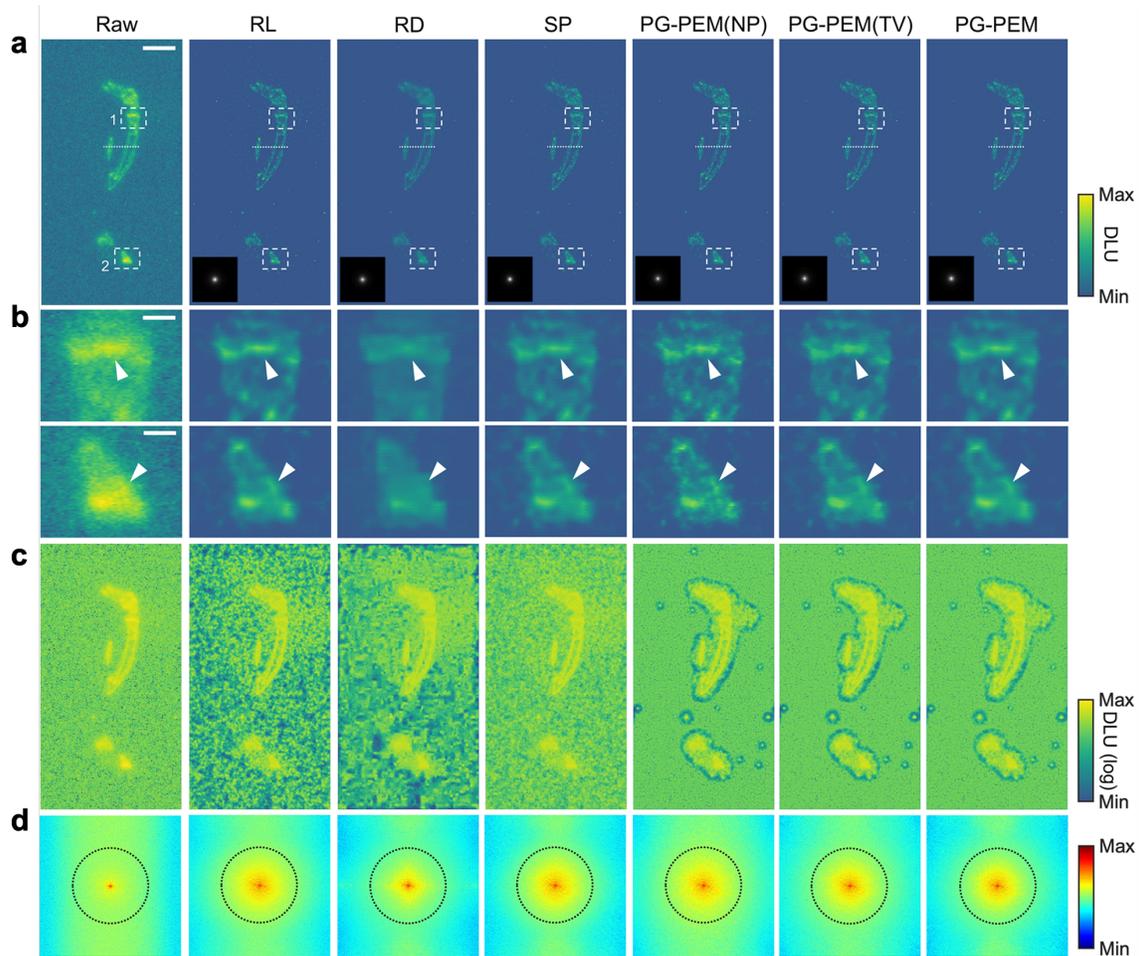


Figure 3.10: Blind restoration and background segmentation improve DAR image quality. (a) A raw DAR image from the mouse hindlimb following $[^{18}\text{F}]\text{NaF}$ PET imaging and its restoration results using modified restoration algorithms, PSF estimation and patch-based segmentation. The corresponding estimated PSFs are inset in grey scale. (b) Magnified images of the corresponding dashed boxed regions in (a). (c) Log-scale transformed images from (a) for background appraisal. (d) Log-scale amplitude of the Fourier transform of the raw and restored images in (a). Scale bar: (a) and (c): 4.95 mm, (b): 0.86 mm.

lution and suppressed noise to varying magnitudes. Log-scale images reveal NP, TV and PG-PEM have more homogeneous background than other methods, a result of splitting the

image components into Poisson-distributed signal and Gaussian-distributed noise. The non-homogeneous background in RL, RD and SP correspond to noise and false positive signal generated in their restoration process.

Next we assessed the log-scale amplitude of the Fourier space. Because h is isotropic, the resolution of DAR images should be at least quasi-isotropic. Curiously, we observed that high frequencies tended along the horizontal direction and are highly non-isotropic, which correspond to the noise. By comparing the frequency maps of NP, TV and PG-PEM, the non-isotropic components of NP have the highest energy. TV produces a broader non-isotropic frequency portion than PG-PEM and a “staircase” effect. These along with STDB and CNR indicate that PG-PEM is the best denoiser. Meanwhile, RL, SP, NP, TV and PG-PEM share similar quasi-isotropic areas in the dotted black circles, while that of RD has the lowest energy. In fact, the resolution of RD is the lowest because the wavelet-denoising process removes fine details. With a MLEM restoration framework (and the same regularization strategy for the PSF \mathbf{h}) RL, SP, NP, TV and PG-PEM share similar resolutions. Notably, due to the lack of a regularization strategy for \mathbf{X} , the resolution of NP may be slightly higher than those of the other methods, which can be neglected due to the impact of noise. The effective resolution improves at least 5-fold after restoration by PG-PEM ($P < 0.0001$). These data along with the simulation results demonstrate that PG-PEM is the best performer for blind restoration of DAR images.

3.6.3 PG-PEM improves DAR of diagnostic radiopharmaceuticals

We were interested to determine whether PG-PEM could improve the quality of DAR images in diagnostic radiopharmaceuticals. We have investigated the distribution of the widely used metabolic tracer ^{18}F -FDG, and the bone seeking ^{18}F -NaF, in tissue samples from mice tumor, heart and femur ($N = 10$ per group).

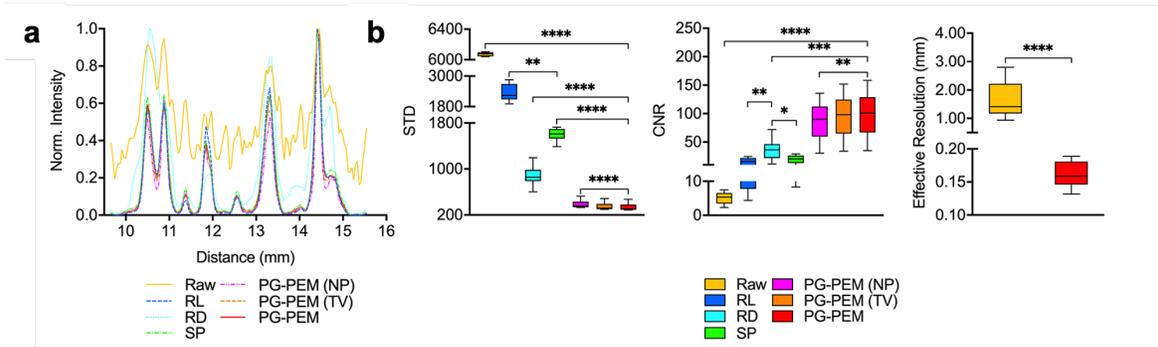


Figure 3.11: Quantitative evaluation of the DAR image quality enhancement. (a) Line profiles of the images in Figure 3.10a along the corresponding dashed lines. (b) STD, CNR and effective resolution assessment of the DAR images. Different colors correspond to raw and restored data. Box plot: Center line, median; limits, 75% and 25%; whiskers, maximum and minimum. DLU: digital light unit. We used the two-sided Student's t-test to compare the effective resolution of Raw and PG-PEM restored DAR images, and the one-way analysis of variance (ANOVA) to compare the STD and CNR of PG-PEM and all the other restoration algorithms. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

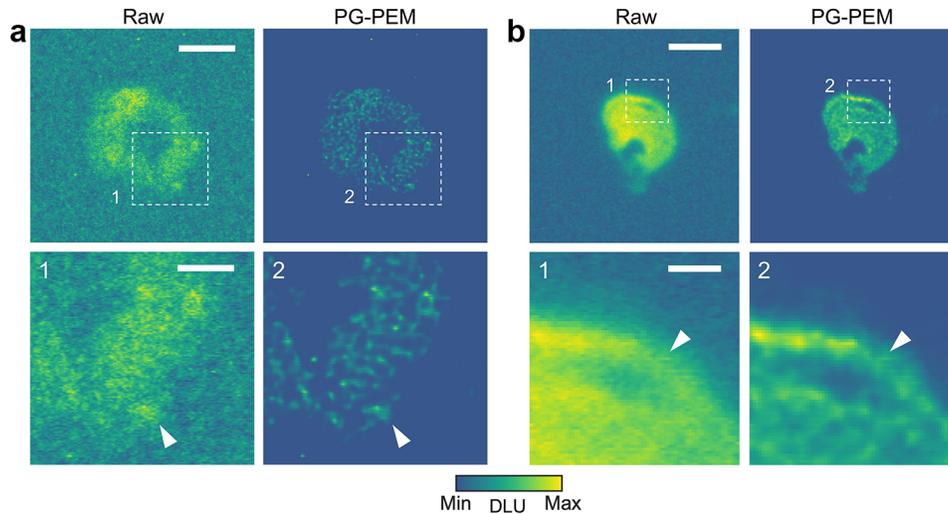


Figure 3.12: Content-adaptive blind restoration improves mice tumor and heart DAR. (a) and (b): Raw and PG-PEM restored DAR images of mice tumor and heart sections, which were treated with $[^{18}\text{F}]\text{FDG}$. DLU: digital light unit. Scale bar: (a): 4.2 mm, (a1): 1.4 mm; (b): 3.7 mm, (b1): 0.75mm.

We used PG-PEM to restore these data, calculated STDB, CNR, and effective resolution for comparison to the raw images. These results demonstrate the image quality improvement after restoration. Notably, the presented DAR image of a non-glycolytic (prostate) tumor section which takes up little ^{18}F -FDG has extremely low SNR. Nevertheless, PG-PEM suppresses background noise and improves resolution of regions of uptake (Figure 3.12a). The corresponding STDB and CNR reveal that PG-PEM is superior to restore DAR images under extremely low SNR conditions, with a $P < 0.0001$ (Figure 3.14).

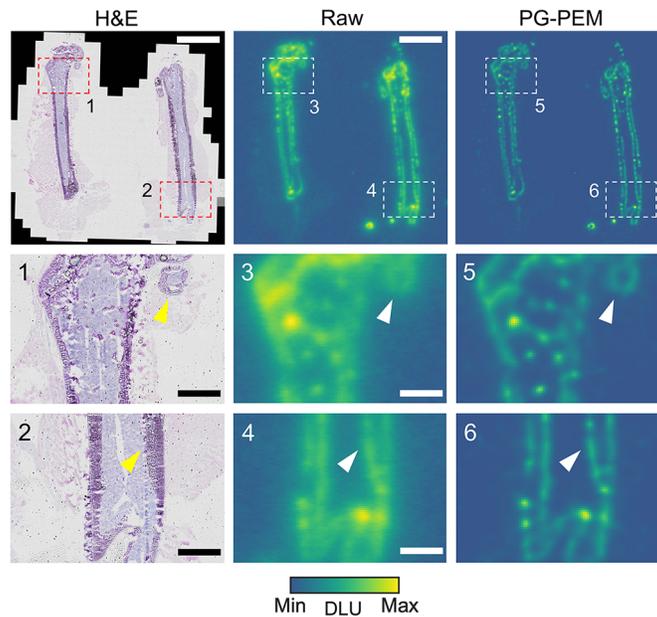


Figure 3.13: Content-adaptive blind restoration improves mice femur DAR. From left to right, H&E stained histopathological, raw and PG-PEM restored DAR images of mice femur section, which was treated with ^{18}F -NaF. The images from the middle and bottom rows correspond to their labeled zoomed-in regions of the images from the top row. DLU: digital light unit. Scale bar: H&E: top: 5 mm, middle and bottom: 1.2 mm; Raw: top: 5 mm, middle and bottom: 1.2 mm.

We next asked if higher SNR images, from ^{18}F -FDG in the heart and ^{18}F -NaF in the bone, could likewise be improved by PG-PEM. From the raw cardiac images, radioisotope signal is

almost homogeneous. By contrast, the PG-PEM restored data has higher resolution and improved contrast which may better reflect the spatial distribution of the tracer (Figure 3.12b). We further compared the H&E, raw and restored DAR images of the murine femur (Figure 3.13). After restoration the endosteal and periosteal surfaces are clearly visualized, and the proximal head of the femur is resolved. Because the positron range of ^{18}F is considerable, the corresponding DAR images more blurred than those treated by lower energy beta emitters or high-linear energy transfer alphas, which hinders the assessment of the radiopharmaceuticals distribution. Our results indicate that PG-PEM can ameliorate this issue, underscoring pre-clinical utility (Figure 3.14).

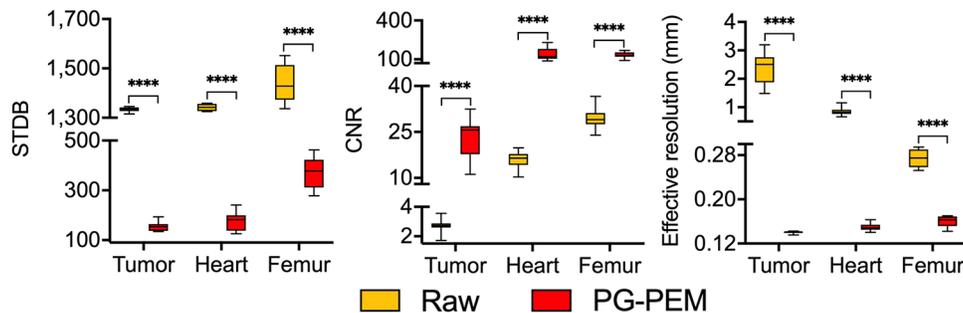


Figure 3.14: STDB, CNR and effective resolution of the DAR images before and after restoration by PG-PEM on pre-clinical data. Every kind of tissue consists of 10 sections. Box plot: Center line, median; limits, 75% and 25%; whiskers, maximum and minimum. We used the two-sided Student's t-test to compare the data of Raw and PG-PEM restored DAR images in (h)–(j). **** $P < 0.0001$.

3.6.4 Enhanced Targeted Alpha Particle Radiotherapy Evaluation by PG-PEM

We next investigated alpha particle emitter activity distributions from a dataset of 10 bone biopsy slides of mCRPC patients treated with $^{223}\text{RaCl}_2$ (Figures 3.15 and 3.16). The raw

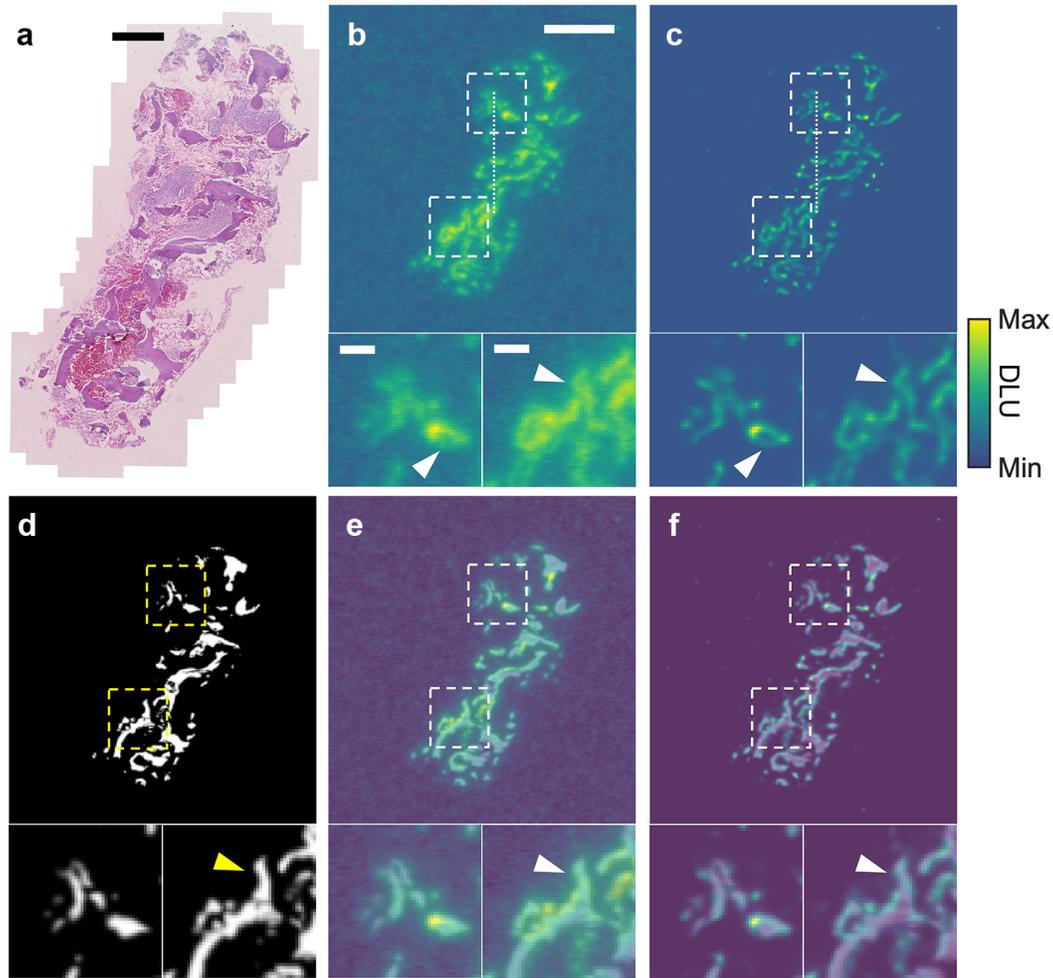


Figure 3.15: PG-PEM restoration in clinical targeted alpha particle radiotherapy specimens. (a) H&E stained histopathological image of a bone biopsy section obtained from Radium-223 dichloride treated castrate resistant prostate cancer patient metastasis. (b) and (c) Raw and PG-PEM restored DAR image of the same representative section in (a). (d) The bone area was manually segmented and registered with raw and restored DAR based on their mutual information to produce (e) and (f) fused DAR and pathology images. In (b)–(f) magnified regions delineated by dashed lines are shown below the full-section image. Scale bar: (a) 1 mm. (b) upper image: 2.3 mm, lower images 0.5 mm.

DAR images suffer from blur and noise due to the imaging process, distorting the true radio-tracer distribution. This can cause large errors in the registration, and degrades treatment response assessment and toxicity analysis. ^{223}Ra will adsorb on the bone surface [83] and

the high activity regions should be located at the bone. Based on this knowledge, DAR and histopathology images can be co-registered, and restoration algorithms evaluated.

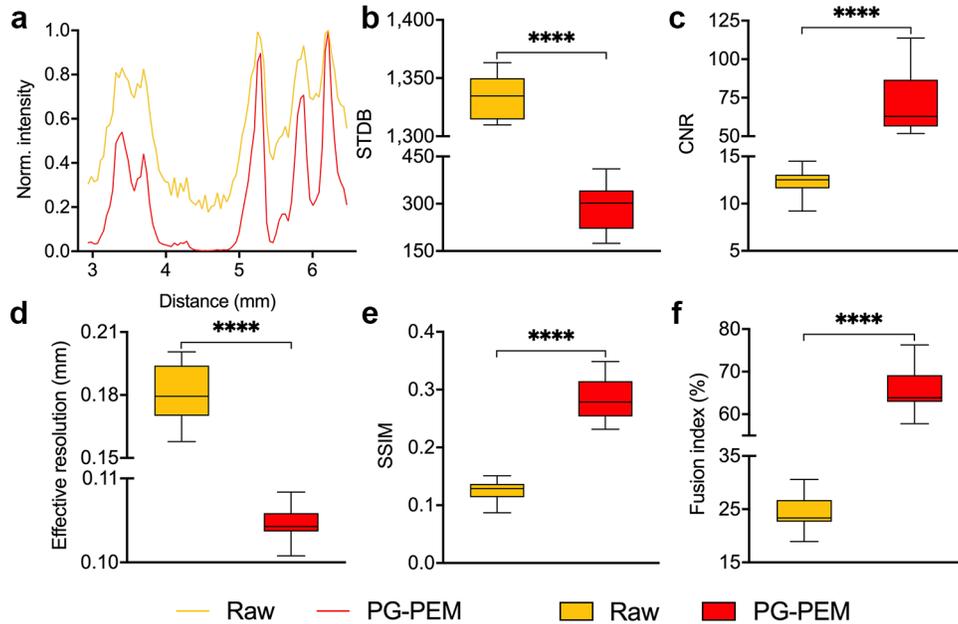


Figure 3.16: Quantitative evaluation of the DAR image quality enhancement in clinical targeted alpha particle radiotherapy specimens. (a) Line profiles of raw and PG-PEM DAR data (as defined in Figure 3.15(b) and (c)). (b)–(d) STD, CNR and effective resolution assessment of DAR images. (e) and (f) Comparison of the accuracy of alignment of DAR and histopathology data for SSIM and fusion indices for raw and restored images ($N = 10$). DLU: digital light units. Box plot: Center line, median, limits, 75% and 25%; whiskers, maximum and minimum. We used the two-sided Student’s t-test to compare the data of original and enhanced images in (b)–(f). **** $P < 0.0001$.

Based on the registered result, we observe that PG-PEM can not only improve the resolution and remove the noise of DAR images, but also results in more accurate correlation with histopathology and underlying anatomy. To quantify these improvements, we first computed STDB, CNR and effective resolution. All of them improve and the effective resolution increases by approximately 1.7 folds over raw data. We then calculated the SSIM between the high activity regions of DAR images with their segmented bone masks, and evaluated a

Fusion Index, defined as the ratio of total activity at bone surfaces. Note that the higher SSIM and Fusion Index are, the better the correlation between the modalities. The evaluation results show PG-PEM is able to improve these two accuracy metrics significantly ($P < 0.0001$). Consequently, PG-PEM can be of great use in personalized targeted alpha particle radiotherapy assessment.

3.7 Discussion

Autoradiography is an important technique in drug development and in evaluation of radio-labeled compounds for imaging and targeted therapy [84, 85, 86, 87, 88, 89]. In particular, there is considerable academic, pharmaceutical industry and clinical interest to use it to evaluate targeted alpha and beta particle emitters for endotherapy. Unlike external beam radiation delivery, systemically administered radionuclides have the capacity to irradiate all tissues in the body. The localized distribution is central to calculate absorbed doses and to predict both treatment response and off-target toxicities. Conventional image formation methods using DAR suffer from noise and other image artifacts. In this work we have defined and implemented a novel PG-PEM algorithm to restore blurred and noisy DAR images.

PG-PEM is based on the DAR imaging process and a mixed Poisson-Gaussian noise model. The noise parameters are estimated with a patch-based algorithm after a Poisson-Gaussian distribution conversion. A penalized MLEM approach is then used to jointly estimate the specimen image and its corresponding PSF, simultaneously. Specifically, we used $L2$ norm to regularize the PSF to ensure its smoothness and avoid the trivial solution; and HF norm to regularize the estimated specimen image to ensure its continuity and to suppress noise. Notably, this approach can effectively eliminate the “staircase” effect caused by TV regularization. As a consequence, even low SNR images are robustly restored. To the best of our

knowledge, this is the first attempt to combine MLEM with Hessian norm-based regularization.

After analyzing the scaling factor α , we prove that it is free of pre-calibration in PG-PEM. Subsequently, the algorithm was quantitatively compared against alternative approaches across multiple datasets. Consistent with other EM-based algorithms, PG-PEM is not a convex problem and we cannot guarantee it can converge to a global solution. Nevertheless,

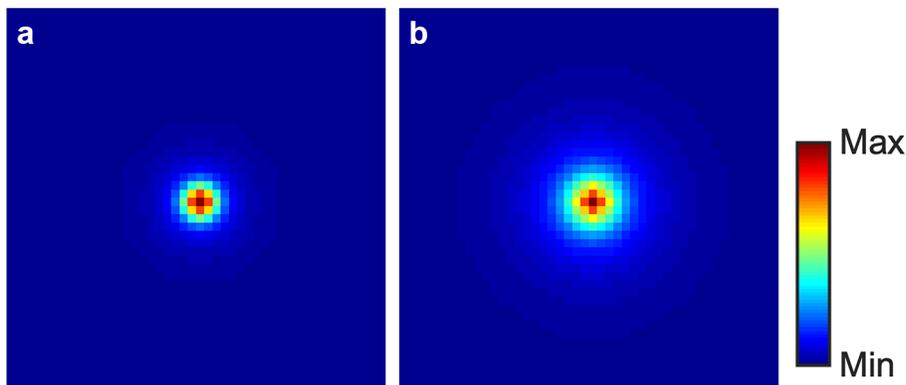


Figure 3.17: PSF comparison. (a) The estimated PSF from a $^{223}\text{RaCl}_2$ treated human bone biopsy image. (b) The estimated PSF from a $^{18}\text{F-NaF}$ treated mouse hindlimb image. To increase the image contrast, we have applied a false coloured “jet” colormap. The estimated full-width half-maximum and full-width tenth-maximum of (a) are approximately 0.159 mm and 0.344 mm, while those of (b) are 0.216 mm and 0.66 mm, respectively. These results are consistent with the physics of alpha/positron transport: alpha particles have a significantly shorter path length than positrons. This is also why the raw autoradiographic images from ^{223}Ra treated human bone biopsy have higher resolution than those from ^{18}F treated sections. It should be noted that while several beta particles are produced by its daughters, the vast majority of the ^{223}Ra decay energy is in the form of alpha particles.

simulation and experimental results show that PG-PEM is the lead performer, providing improved correlation between signal and tissue features. Interestingly, even though both SP and PG-PEM are based upon the mixed Poisson-Gaussian noise model, PG-PEM has lower noise and reduced background false positive signal. This difference comes from the iteration process: PG-PEM first filters Gaussian-distributed noise in the E step and then

filters Poisson-distributed noise in the M step. In addition, we have also compared the PSFs estimated from different isotopes ($^{223}\text{RaCl}_2$ treated human bone biopsy and $^{18}\text{F-NaF}$ treated mouse hindlimb). Clearly, the kernel size of the PSF from the hindlimb is larger than that from the biopsy (Figure 3.17), consistent with the physics of alpha/positron travel, further validating the blind restoration approach.

Recently, convolutional neural networks have proved effective in blind image restoration and reconstruction [90, 91, 92, 93]. However, it may not be well suited for DAR restoration because of multiparametric factors influencing PSF, noise characteristics for each isotope and tissue, and the lack of clean data as label.

3.8 Conclusion

We have developed the PG-PEM algorithm for improved DAR image quality. Predicated on a complete image formation model for DAR and implementation of a signal and background segmentation approach, this blind image restoration approach reduced background noise and image blur in simulated and primary image samples. For both high and low SNR datasets, of diagnostic and therapeutic radionuclides, there were significant improvements in DAR resolution, contrast and accuracy of localization. This method will be widely applicable to both pre-clinical and clinical sample autoradiograms to improve radiotracer and radiotherapy agent evaluation.

Chapter 4

IMC-Denoise: A Content-Aware Pipeline to Enhance Imaging Mass Cytometry

4.1 Introduction

In this section, we develop and apply IMC-Denoise, a content aware denoising pipeline to enhance IMC images through an automated process. Specifically, IMC-Denoise invokes a novel DIMR algorithm to account for hot pixels, and another novel DeepSNiF algorithm for shot noise. We deploy the pipeline on a technically challenging unique human bone marrow dataset. We benchmark our approach against existing hot pixel removal methods [27, 30, 31, 58] and other advanced biomedical imaging denoising algorithms, such as non-local means filtering (NLM) [94], batch matching and 3D filtering (BM3D) [95] and Noise2Void (N2V)

Content of this chapter is extracted/adapted from the author’s published journal article [35].

[96], which is used in IMC here for the first time. We demonstrate that the image formation model derived IMC-Denoise pipeline produces image quality enhancements that are best-in-class and leads to improved downstream analysis, with limited manual user manipulation. Qualitative improvements in images enhances their interpretation, and quantitatively improve molecularly-defined phenotyping. Results from the IMC-Denoise pipeline are suitable for further downstream analysis, such as Mesmer/DeepCell and ark-analysis [97] or MCMicro [98]. We provide this tool to augment studies that seek to more deeply characterize the complex and diverse tissue microenvironment.

4.2 IMC-Denoise Framework

4.2.1 DIMR algorithm for hot pixel removal

Hot pixel unmixing

From Equation (2.3), larger true signal \mathbf{X} will result in larger variance. Thus, the contaminated signal $\mathcal{P}[\mathbf{X}]$ with larger \mathbf{X} is more likely to be considered as hot pixels \mathbf{Q} and vice versa. To avoid such false detection, we stabilize the variance of the signal \mathbf{X} with the Anscombe transformation [99] as Equation (4.1).

$$\mathbf{R}' = \mathbf{X}' + \mathcal{N}(0, 1) + \mathbf{Q}', \quad (4.1)$$

where $\mathcal{N}(0, 1)$ is the additive noise with standard Gaussian distribution, and \mathbf{R}' , \mathbf{X}' and \mathbf{Q}' are the transformed raw image, “clean” signal and hot pixels, respectively. As hot pixels are local maxima in IMC images, we detect them by comparing adjacent pixels in a 3×3 sliding window. Considering the nonlinearity of the Anscombe transformation [100], pixels

with intensities lower than 4 in \mathbf{R} are omitted directly in order to exclude the impact of background regions, which cannot be outliers. Additionally, the difference between adjacent pixels can be fitted as a generalized Gaussian distribution [101]. Thus, in the 3×3 sliding window, we derive Equation (4.2) by calculating the differences between the center pixel and its 8 neighbours. In Equation (4.2), i is the neighbour index in the sliding window ($i \in \{1, 2, \dots, 8\}$), $\mathcal{G}(\mu, \alpha, \beta)$ is a generalized Gaussian distribution with location μ , scale α and shape β , and $\mathbf{R}' - \mathbf{R}''_i$, $\mathbf{X}' - \mathbf{X}''_i$ and $\mathbf{Q}' - \mathbf{Q}''_i$ are the differences of the raw image, “clean” signal and hot pixels in the i -th direction, respectively. Without the hot pixel component $\mathbf{Q}' - \mathbf{Q}''_i$, \mathbf{D}_i can also be approximated as a generalized Gaussian distribution.

$$\begin{aligned} \mathbf{D}_i &= \mathbf{R}' - \mathbf{R}''_i = \mathbf{X}' - \mathbf{X}''_i + \mathcal{N}(0, 2) + \mathbf{Q}' - \mathbf{Q}''_i \\ &= \mathcal{G}(\mu_i, \alpha_i, \beta_i) + \mathcal{N}(0, 2) + \mathbf{Q}' - \mathbf{Q}''_i. \end{aligned} \tag{4.2}$$

Similar to fluorescence microscopy, in IMC images the tissue or background pixels should be continuous. Thus, for any normal pixel p the distance between d_i^p and the mean μ_i is always less than that of a single hot pixel, where d_i^p is the pixel p 's value in the distribution \mathbf{D}_i . Therefore, we can define $\Delta_i^p = |d_i^p - \mu_i|$ as the measure to determine whether a pixel p is a hot pixel. However, this might not hold for consecutive hot pixels. For instance, if two consecutive hot pixels sharing similar intensities, their difference may be very close to μ_i . To detect consecutive hot pixels, it is reasonable to assume there are at least l pixels close to a normal pixel p . l is normally set as 4, which corresponds to half neighbours. Consequently, we sort Δ_i^p for every pixel p in an ascending direction as Equation (4.3).

$$\Delta_{(i)}^p = \text{sort}(\Delta_i^p), \tag{4.3}$$

where (i) is the sorted index and $i \in \{1, 2, \dots, 8\}$. Then we define the sum of the first l $\Delta_{(i)}^p$ as Equation (4.4).

$$s_l^p = \sum_{i=1}^l \Delta_{(i)}^p, \quad (4.4)$$

For a normal pixel p , its s_l^p should be less than that of a hot pixel. Because s_l^p measures the relationship between the center pixel and its multiple neighbours, it is more robust than a single Δ_i^p , especially for consecutive hot pixels.

Due to the spatial continuity and isotropic resolution of IMC images, μ_i from different directions can be regarded as equal to each other. For the sake of simplicity, we define $\mu_i = \mu$ for $i \in \{1, 2, \dots, 8\}$. To further separate normal and hot pixels, Equation (4.5) is consequently derived based on the triangle inequality [102].

$$s_l^p = \sum_{i=1}^l |d_{(i)}^p - \mu| \geq |t_l^p - l\mu|, \quad (4.5)$$

where $t_l^p = \sum_{i=1}^l d_{(i)}^p$. The equality holds only when $d_{(i)}^p \geq \mu$ or $d_{(i)}^p \leq \mu$ for all $i \in \{1, 2, \dots, l\}$. The first case is always true for single hot pixels and pixels with the largest intensities in hot clusters. Otherwise, t_l^p will shrink towards $l\mu$. Therefore, t_l^p will further unmix hot pixels from normal ones. Combining all the t_l^p , a new distribution \mathbf{T}_l is generated, and outliers are located beyond its right tail. In fact, some consecutive hot pixels can shrink towards $n\mu$ as well. For example, in the case of a hot pixel that is larger than all of its normal neighbours but smaller than the largest hot pixel in the 3×3 sliding window, it is possible that $t_{l-1}^p - (l-1)\mu = \mu - d_{(l)}^p > 0$. To solve this issue, we implement multiple iterations of the sorting to adequately remove the hot pixel noise. Normally the iteration number N_{iter} is set as 3 such that at least 3 consecutive hot pixels will be removed after 3 iterations.

Hot pixel detection

In order to search for outliers, the shape of \mathbf{T}_l should be investigated first. Therefore, let us also define $u_l^p = \sum_{i=1}^l d_i^p$, and combine u_l^p from all pixels to form a distribution \mathbf{U}_l . Without hot pixels, \mathbf{U}_l is approximately a generalized Gaussian distribution with mean $l\mu$. Because

$$\sum_p \sum_{i=1}^l (\Delta_{(i)}^p)^2 \leq \sum_p \sum_{i=1}^l (\Delta_i^p)^2, \quad (4.6)$$

where the two items are proportional to the variances of \mathbf{T}_l and \mathbf{U}_l , respectively, \mathbf{T}_l can be regarded as a generalized Gaussian distribution with mean $n\mu$ and smaller variance than \mathbf{U}_l .

Because the histogram of \mathbf{T}_l is discretized, we apply the kernel density estimation algorithm [103], as Equation (4.7) shows, to fit a continuous curve.

$$\hat{g}_h(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \quad (4.7)$$

where K is a Gaussian kernel, h the bandwidth, m the point number and x_i the sampled points. The bandwidth h is set as $1.06\hat{\sigma}l^{-\frac{1}{5}}$ [103], where $\hat{\sigma}$ is the standard deviation of the sampled points. The point interval is set as 1 for adequate sampling as well as avoiding comb-like structures. Subsequently, a moving mean filter with window size of 3 is optionally utilized to further eliminate minor fluctuations of the fitted curve. With the fitted curve $(x, \hat{g}_h(x))$, a threshold point x_T is defined, and any points $x > x_T$ are considered as outliers.

Because outliers are located beyond the right tail of \mathbf{T}_l , it is reasonable to set x_T when $\frac{d\hat{g}_h(x)}{dx} \rightarrow 0$, which means the current distribution ends. Likewise, the shape of the curve will not change from convex to concave on its right tail. Thus, it is also reasonable to set x_T when $\frac{d^2\hat{g}_h(x-\Delta x)}{dx^2} \geq 0$ and $\frac{d^2\hat{g}_h(x)}{dx^2} \leq 0$, where Δx represents a small value. Because the pixel

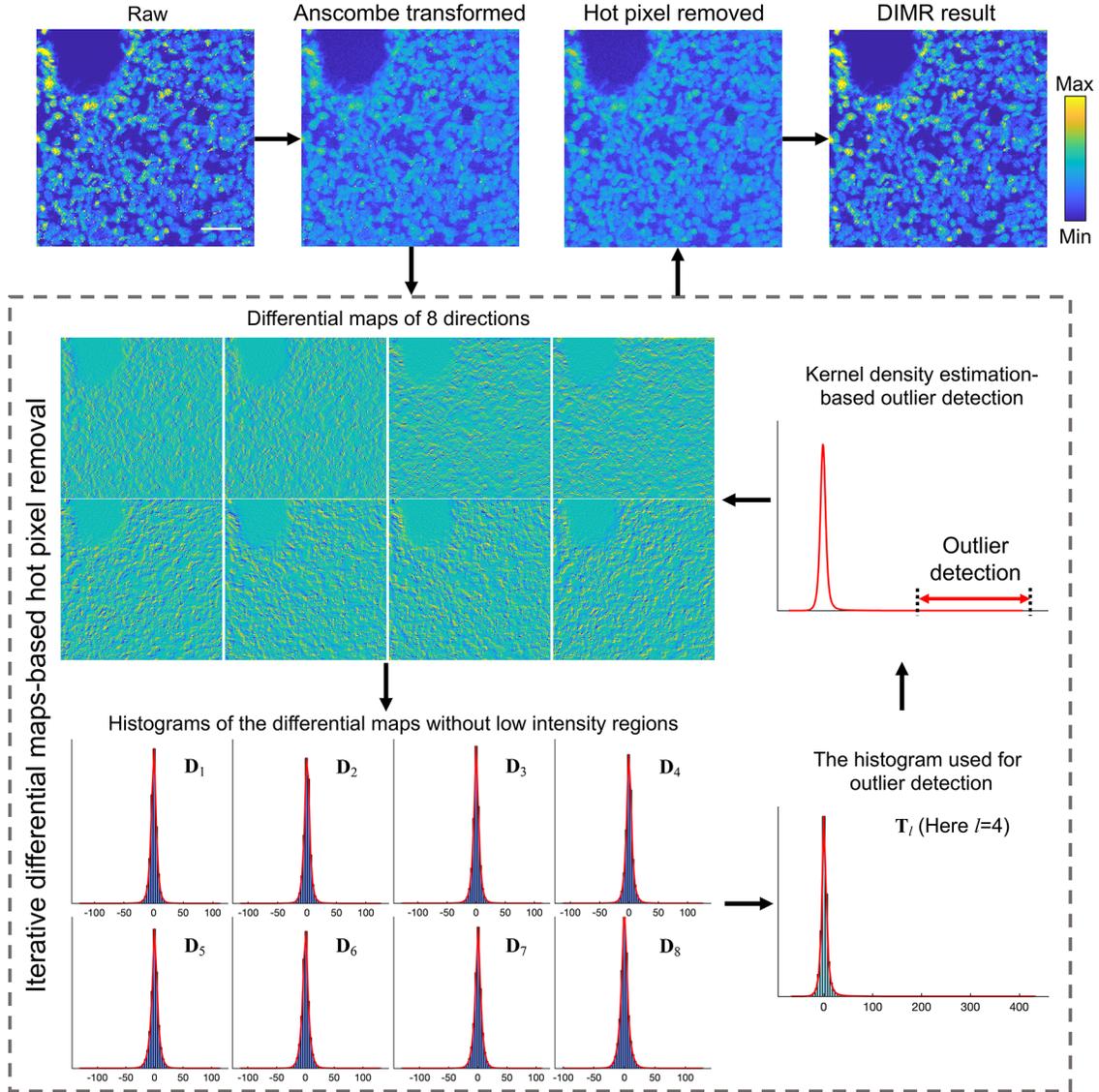


Figure 4.1: The schematic of the DIMR algorithm.

values of the raw images are discrete, Δx is normally set as 1. The detected outliers are filtered by a 3×3 median filter in each iteration. When the iterations terminate, the image is transformed to its original scale by the direct algebraic inverse Anscombe transformation [104]. Additionally, we substitute the mean μ_i^p of each distribution D_i with the median $\tilde{\mu}_i^p$ when implementing this algorithm, as median is more robust than mean when encountering outliers.

The DIMR algorithm is summarized as Algorithm 3 and Figure 4.1.

Algorithm 3 DIMR algorithm

Initialization: Raw image \mathbf{R} ;
Hyper-parameters l and N_{iter} ;
Output: Hot pixel removed image;

- 1: Apply the Anscombe transformation to \mathbf{R} ;
- 2: **for** each $k = 1, 2, \dots, N_{\text{iter}}$ **do**
- 3: Calculate the 8 differential maps of \mathbf{R}' in a sliding 3×3 window;
- 4: Remove the pixels with value less than 4 for each differential map and form the distributions \mathbf{D}_i where $i \in \{1, 2, \dots, 8\}$;
- 5: For every remaining pixel p , calculate $\Delta_i^p = |d_i^p - \tilde{\mu}_i|$, where $\tilde{\mu}_i$ is the median of \mathbf{D}_i ;
- 6: Sort all the Δ_i^p and get the corresponding sorted index (i) for every pixel p . With the index (i), calculate t_i^p for each pixel p to form a new distribution \mathbf{T}_l ;
- 7: Apply the kernel density estimation algorithm (Equation (4.7)) to generate a continuous curve $(x, \hat{g}_h(x))$ from the histogram of \mathbf{T}_l , followed by moving mean filtering with window size of 3 (optional).
- 8: Starting from the right tail of the curve, if (1) $\frac{d\hat{g}_h(x)}{dx} \rightarrow 0$ or (2) $\frac{d^2\hat{g}_h(x-\Delta x)}{dx^2} \geq 0$ and $\frac{d^2\hat{g}_h(x)}{dx^2} \leq 0$, set the point x as x_T ;
- 9: **if** There is no $x > x_T$ **then**
- 10: Break
- 11: **else**
- 12: Set any points $x > x_T$ as outliers and filter the corresponding pixels with a 3×3 median filter.
- 13: **end if**
- 14: **end for**
- 15: Apply the direct algebraic inverse Anscombe transformation to the processed image.

4.2.2 DeepSNiF for image quality enhancement

Optimal loss function derivation

After hot pixel removal, the IMC imaging model is simplified as Equation (4.8).

$$\mathbf{R} = \mathcal{P}[\mathbf{X}]. \quad (4.8)$$

In a supervised learning framework, the noisy and clean images \mathbf{R} and \mathbf{X} are both known. Therefore, for all the pixels p from 1 to L , pairs of (r_p, x_p) are formed as a training set. A deep convolutional neural network with forward model $\mathcal{F}_\theta (\mathbb{E}^L \mapsto \mathbb{E}^L)$ is then built to filter the Poisson noise such that $\mathbf{X} = \mathcal{F}_\theta[\mathbf{R}]$, where θ reflects the parameters of the neural network. Ideally, $\mathcal{F}_\theta[\mathbf{R}]_p$ should be identical to that of x_p for all pixels. In this sense, the optimal estimator for denoising can be derived through a binary signal detection task. This task requires an observer to classify the denoised signals $\mathcal{F}_\theta[\mathbf{R}]_p$ under a hypothesis H_0 from the true clean signals x_p under another hypothesis H_1 . Given noisy data r_p , these two hypotheses can be described as:

$$\begin{aligned} H_0 : \text{denoised signals} : \mathcal{F}_\theta[\mathbf{R}]_p, 1 \leq p \leq L \\ H_1 : \text{true signals} : x_p, 1 \leq p \leq L \end{aligned} \quad (4.9)$$

Because the IMC data follows a Poisson distribution, the corresponding likelihood functions for these two hypotheses are expressed as Supplementary Equations (4.10) and (4.11). From Equation (4.11), x_p can be regarded as the maximum likelihood estimation (MLE) solution of r_p .

$$\Pr(\mathbf{R}|\mathcal{F}_\theta[\mathbf{R}]) = \prod_{p=1}^L \exp(-\mathcal{F}_\theta[\mathbf{R}]_p) \mathcal{F}_\theta[\mathbf{R}]_p^{r_p} / r_p!. \quad (4.10)$$

$$\Pr(\mathbf{R}|\mathbf{X}) = \prod_{p=1}^L \exp(-x_p) x_p^{r_p} / r_p!. \quad (4.11)$$

The log-likelihood functions of the two hypotheses are derived as Supplementary Equations (4.12) and (4.13).

$$\log[\Pr(\mathbf{R}|\mathcal{F}_\theta[\mathbf{R}])] = \sum_{p=1}^L (r_p \log(\mathcal{F}_\theta[\mathbf{R}]_p) - \mathcal{F}_\theta[\mathbf{R}]_p - \log r_p!). \quad (4.12)$$

$$\log[\Pr(\mathbf{R}|\mathbf{X})] = \sum_{p=1}^L (r_p \log(x_p) - x_p - \log r_p!). \quad (4.13)$$

Then, we define the log-likelihood ratio $\mathcal{L}(H_0, H_1)$ to measure the difference between the two hypotheses:

$$\begin{aligned} \mathcal{L}(H_0, H_1) &= \log[\Pr(\mathbf{R}|\mathbf{X})] - \log[\Pr(\mathbf{R}|\mathcal{F}_\theta[\mathbf{R}])] \\ &= \sum_{p=1}^L (r_p \log \frac{x_p}{\mathcal{F}_\theta[\mathbf{R}]_p} - x_p + \mathcal{F}_\theta[\mathbf{R}]_p). \end{aligned} \quad (4.14)$$

The learning process aims to train \mathcal{F}_θ such that $x_p = \mathcal{F}_\theta[\mathbf{R}]_p$ for all the p from 1 to L . Thus, the expectation of this loss function \mathcal{L} under hypothesis H_1 , $E[\mathcal{L}(H_0, H_1)|H_1]$, as Equation (4.15) demonstrates, should be as small as possible. In that sense, the observer in the binary detection task will be more difficult to differ these two hypotheses as $E[\mathcal{L}(H_0, H_1)|H_1]$ decreases. When $E[\mathcal{L}(H_0, H_1)|H_1] = 0$, the two hypotheses are identical and the denoised signal $\mathcal{F}_\theta[\mathbf{R}]_p$ will be identical to the true signal x_p for all the pixels, which means $\mathcal{F}_\theta[\mathbf{R}]_p$ will be the MLE solution of r_p as well.

$$\begin{aligned} E[\mathcal{L}(H_0, H_1)|H_1] &= \sum_{p=1}^L (E[r_p|H_1] \log \frac{x_p}{\mathcal{F}_\theta[\mathbf{R}]_p} - x_p + \mathcal{F}_\theta[\mathbf{R}]_p) \\ &= \sum_{p=1}^L (x_p \log \frac{x_p}{\mathcal{F}_\theta[\mathbf{R}]_p} - x_p + \mathcal{F}_\theta[\mathbf{R}]_p). \end{aligned} \quad (4.15)$$

Equation (4.15) is generally known as I-divergence [105]. Here, we set it as the loss function, and thus the optimal parameter of the network $\hat{\theta}^*$ is defined as Equation (4.16).

$$\hat{\theta}^* = \operatorname{argmin}_\theta \sum_{p=1}^L (x_p \log \frac{x_p}{\mathcal{F}_\theta[\mathbf{R}]_p} - x_p + \mathcal{F}_\theta[\mathbf{R}]_p). \quad (4.16)$$

Due to the difficulties associated with acquiring high SNR images as ground truths and the impossibility to repetitively scan the same tissue in IMC, a supervised learning approach is unavailable here. Fortunately, we find self-supervised approaches are also qualified for this denoising task. Let us define a function f demonstrating a random pixel masking approach for images such as those in Noise2Void [96] and Noise2Self [106]. In such strategies, multiple pixels are randomly masked and replaced by their adjacent pixels or random values. The new pixel value $f(\mathbf{R})_p$ at pixel p can be regarded as the true value x_p contaminated by another noise process. In this case, the true value x_p can be predicted by the adjacent pixels because of the spatial continuity of IMC images. Therefore, in a self-supervised learning problem, the loss function will be

$$\begin{aligned}
\mathcal{L}(\mathbf{R}, \mathcal{F}_\theta[f(\mathbf{R})]) &= \sum_{p=1}^L (r_p \log \frac{r_p}{\mathcal{F}_\theta[f(\mathbf{R})]_p} - r_p + \mathcal{F}_\theta[f(\mathbf{R})]_p) \\
&= \sum_{p=1}^L (r_p \log r_p - (x_p + r_p - x_p) \log \mathcal{F}_\theta[f(\mathbf{R})]_p - (x_p + r_p - x_p) + \mathcal{F}_\theta[f(\mathbf{R})]_p) \\
&= \sum_{p=1}^L (x_p \log \frac{x_p}{\mathcal{F}_\theta[f(\mathbf{R})]_p} - x_p + \mathcal{F}_\theta[f(\mathbf{R})]_p) + \sum_{p=1}^L (r_p \log r_p - x_p \log x_p) \\
&\quad - \sum_{p=1}^L (r_p - x_p) - \sum_{p=1}^L (r_p - x_p) \log \mathcal{F}_\theta[f(\mathbf{R})]_p. \tag{4.17}
\end{aligned}$$

In order to ensure that the training process with the self-supervised learning loss is identical to that with known true signals, the last three terms in Equation (4.17) should be constants. Obviously, the second term $\sum_{p=1}^L (r_p \log r_p - x_p \log x_p)$ fulfills this requirement. On the other hand, in the last two terms, $r_p - x_p$ and $\mathcal{F}_\theta[f(\mathbf{R})]_p$ are the noise component and predicted value of pixel p , respectively. Because the Poisson noise is a pixel-independent stochastic process and $\mathcal{F}_\theta[f(\mathbf{R})]_p$ is determined by its neighbours in the proposed self-supervised learning framework, they are uncorrelated with each other, except the masked pixel p is replaced by itself. As a result, the last two term in Equation (4.17) can be approximated as Equation

(4.18).

$$\begin{aligned} \sum_{p=1}^L (r_p - x_p) + \sum_{p=1}^L (r_p - x_p) \log \mathcal{F}_\theta[f(\mathbf{R})]_p &= \sum_{p=1}^L (r_p - x_p) + \sum_{p=1}^L (r_p - x_p) \sum_{p=1}^L \log \mathcal{F}_\theta[f(\mathbf{R})]_p \\ &= \sum_{p=1}^L (r_p - x_p) \left[1 + \sum_{p=1}^L \log \mathcal{F}_\theta[f(\mathbf{R})]_p \right]. \end{aligned} \quad (4.18)$$

Because $E[r_p] = E[x_p]$ for a Poisson noise model, Equation (4.18) is approximated as 0 when L is large enough, also suggested by [106]. From this viewpoint, training with noisy images and their masked pairs is identical to training with noisy and clean image pairs, as long as the masked pixels are not replaced by themselves. To avoid learning identity in the training, only the loss of the masked pixels are considered, as Equation (4.19) shows.

$$\mathcal{L}(\mathbf{R}, \mathcal{F}_\theta[f(\mathbf{R})]) = \sum_p \mathbf{M}_p \cdot \left[r_p \log \frac{r_p}{\mathcal{F}_\theta[f(\mathbf{R})]_p} - r_p + \mathcal{F}_\theta[f(\mathbf{R})]_p \right] / \sum_p \mathbf{M}_p, \quad (4.19)$$

where \mathbf{M}_p is the pixel mask ($\mathbf{M}_p \in \{0, 1\}$). If $\mathbf{M}_p = 1$, then the pixel p is masked; otherwise, it is not. To guarantee the output is non-negative, a softplus function $\log(1 + \exp(x))$ is set as the activation of the network's output layer.

Hessian norm regularization as a booster for the denoising task

Even with the optimal estimator, the denoising performance is still sub-optimal: in the training process, the information of the masked pixels is always neglected; and only partial pixels are utilized. These limitations hinder the trained denoiser to further enhance its performance. To solve this issue, the characteristics of the IMC images should be utilized in the training process. As an imaging modality similar to traditional fluorescence microscopy, IMC is always used for detecting the phenotype of biological structures. Thus, the spatial

continuity between biological structures can be used as a *priori*. In fact, the Hessian norm regularization is widely used to describe such spatial continuity in biological imaging data [52, 34, 107]. However, to the best of our knowledge, this advanced statistical prior has never been used in deep learning-based denoising task. To implement this prior in IMC images, let us first define the Hessian operator $\mathcal{R}_{\text{Hessian}}$ as Equation(4.20).

$$\mathcal{R}_{\text{Hessian}} = \begin{bmatrix} \partial_{xx} & \partial_{xy} \\ \partial_{yx} & \partial_{yy} \end{bmatrix}, \quad (4.20)$$

where $\partial_{xx} = \partial^2/\partial x^2$, $\partial_{xy} = \partial_{yx} = \partial^2/\partial x\partial y$, $\partial_{yy} = \partial^2/\partial y^2$, and x and y are the horizontal and vertical directions of an image, respectively. Then, for any estimated image $\mathcal{F}_\theta[f(\mathbf{R})]$, the corresponding Hessian norm regularization term is defined as Equation (4.21).

$$\|\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})])\| = |\partial_{xx}\mathcal{F}_\theta[f(\mathbf{R})]| + |\partial_{yy}\mathcal{F}_\theta[f(\mathbf{R})]| + \sqrt{2}|\partial_{xy}\mathcal{F}_\theta[f(\mathbf{R})]| \quad (4.21)$$

When involving the Hessian norm regularization as a prior, the optimization problem will convert from Equation (4.15) to Equation (4.22).

$$\hat{\theta}^* = \operatorname{argmin}_\theta \sum_p [E[\mathcal{L}(H_0, H_1)|H_1] - \lambda_{\text{Hessian}} \log[\operatorname{Pr}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})]_p; \theta)]], \quad (4.22)$$

where $\operatorname{Pr}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})]_p; \theta) = \exp(-\|\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})])\|_p)$ and λ_{Hessian} is the Hessian norm regularization parameter. Therefore, the loss function Equation (4.19) will be

$$\begin{aligned} \mathcal{L}(\mathbf{R}, \mathcal{F}_\theta[f(\mathbf{R})]) &= \sum_p \mathbf{M}_p \cdot \left[r_p \log \frac{r_p}{\mathcal{F}_\theta[f(\mathbf{R})]_p} - r_p + \mathcal{F}_\theta[f(\mathbf{R})]_p \right] / \sum_p \mathbf{M}_p \\ &+ \lambda_{\text{Hessian}} \sum_p \|\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})])\|_p / \sum_p. \end{aligned} \quad (4.23)$$

Note that the first term works only for the selected masked pixels, while the second regularization term utilizes all the information of images. Thus, Equation (4.23) overcomes the limitations of Equation (4.19) and further enhances the performance of the denoiser.

Image normalization

For both training and prediction, it is important to normalize the input images to a common range. However, some cells or structures may still exhibit extremely bright signal even after hot pixel removal. Consequently, all input and output data are percentile-normalized between 0 and 1. IMC images are always non-negative and consist of pixels with 0 value, this normalization is defined as Equation (4.24).

$$\text{Norm}(\mathbf{R}; q) = \frac{\mathbf{R}}{1.1 \times \text{perc}(\mathbf{R}, q)}, \quad (4.24)$$

where $\text{perc}(\mathbf{R}; q)$ is the q -th percentile of all pixel values in the training set. We typically use values of $q \in (99.9, 99.999)$. Note that only in training phase any values which are larger than 1 are set as 1. After prediction, the denoised images are re-transformed to their original scales.

This normalization approach does not affect the selection of regularization parameter λ_{Hessian} . The Hessian operator (4.20) is a linear operator, so we have

$$|\mathcal{R}_{\text{Hessian}}(\alpha_{\text{scale}} \mathcal{F}_\theta[f(\mathbf{R})])| = \alpha_{\text{scale}} |\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})])|, \quad (4.25)$$

where $\alpha_{\text{scale}} = 1/(1.1 \times \text{perc}(\mathbf{R}; q))$. From here, we are able to derive

$$\begin{aligned}
& \mathcal{L}(\alpha_{\text{scale}}\mathbf{R}, [\mathcal{F}_\theta[f(\mathbf{R})]]^{\text{scaled}}) \\
&= \sum_p \mathbf{M}_p \cdot \left[\alpha_{\text{scale}} r_p \log \frac{\alpha_{\text{scale}} r_p}{[\mathcal{F}_\theta[f(\mathbf{R})]]_p^{\text{scaled}}} - \alpha_{\text{scale}} r_p + [\mathcal{F}_\theta[f(\mathbf{R})]]_p^{\text{scaled}} \right] \bigg/ \sum_p \mathbf{M}_p \\
&\quad + \lambda_{\text{Hessian}} \sum_p \|\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})])\|_p \bigg/ \sum_p \\
&= \alpha_{\text{scale}} \left\{ \sum_p \mathbf{M}_p \cdot \left[r_p \log \frac{r_p}{[\mathcal{F}_\theta[f(\mathbf{R})]]_p^{\text{scaled}} / \alpha_{\text{scale}}} - r_p + [\mathcal{F}_\theta[f(\mathbf{R})]]_p^{\text{scaled}} / \alpha_{\text{scale}} \right] \bigg/ \sum_p \mathbf{M}_p \right. \\
&\quad \left. + \lambda_{\text{Hessian}} \sum_p \|\mathcal{R}_{\text{Hessian}}(\mathcal{F}_\theta[f(\mathbf{R})] / \alpha_{\text{scale}})\|_p \bigg/ \sum_p \right\}. \tag{4.26}
\end{aligned}$$

Therefore, $[\mathcal{F}_\theta[f(\mathbf{R})]]_p^{\text{scaled}} = \alpha_{\text{scale}} \mathcal{F}_\theta[f(\mathbf{R})]_p$ and the normalization does not affect the strength of regularization.

The DeepSNiF algorithm is summarized as Algorithm 4.

Algorithm 4 DeepSNiF algorithm

Initialization: Hot pixel removed images \mathbf{R} ;

Output: Noise filtered images;

- 1: Generate a training set for a specific hot pixel removed marker channel, in which all the images are percentile normalized between 0 and 1 with Equation (4.24);
 - 2: Train a denoising network for the marker channel with Equation (4.23) as the loss function;
 - 3: Normalize the hot pixel removed images with the pre-calculated maximum of the training set;
 - 4: Filter shot noise for the normalized images with the trained network;
 - 5: De-normalize the predicted images to their original scales.
-

4.2.3 DeepSNiF network implementation

The DeepSNiF structure follows the U-Net architecture [108] with Res-block modules [109], in which the input and output images share the same size (Figure 4.2). U-Net architecture is widely used for image deblurring and denoising [91, 96]. In general, the network is composed of an encoder and a decoder. Starting with the input, the encoder path gradually condenses the spatial information into high-level feature maps with growing depths; the decoder path reverses this process by recombining the information into feature maps with gradually increased lateral details. The information in adjacent feature maps transfers by convolving with 3×3 convolutional filters. The down-sampling and up-sampling are used in encoder and decoder for compressing and reconstructing features, performed here by 2×2 max-pooling and 2×2 up-sampling operations, respectively. Res-blocks are applied to facilitate efficient training. Each Res-block contains a convolution layer, batch normalization and the rectified linear unit (ReLU) nonlinear activation, in which the batch normalization layer aims to speed up training process, ReLU could provide non-linearity in the network. Drop out layers are also added with 0.5 dropout rate after the central two Res-blocks to mitigate overfitting. The skip connections link low-level features and high-level features by concatenating their feature maps. We use the softplus function ($\log(1 + \exp(x))$) as the activation function of the final layer and Equation (4.26) as the loss function so that the output of the network is guaranteed to be non-negative.

The hot pixel-removed images are split into multiple 64×64 patches. Then, the patches are rotated by 90° , 180° and 270° , and flipped as a data augmentation approach. In IMC images, foreground objects of interest might be distributed sparsely. In this case, the model might overfit the background areas and fail to learn the structure of foreground objects if the entire image is used indiscriminately for training. Therefore, patches from the background regions

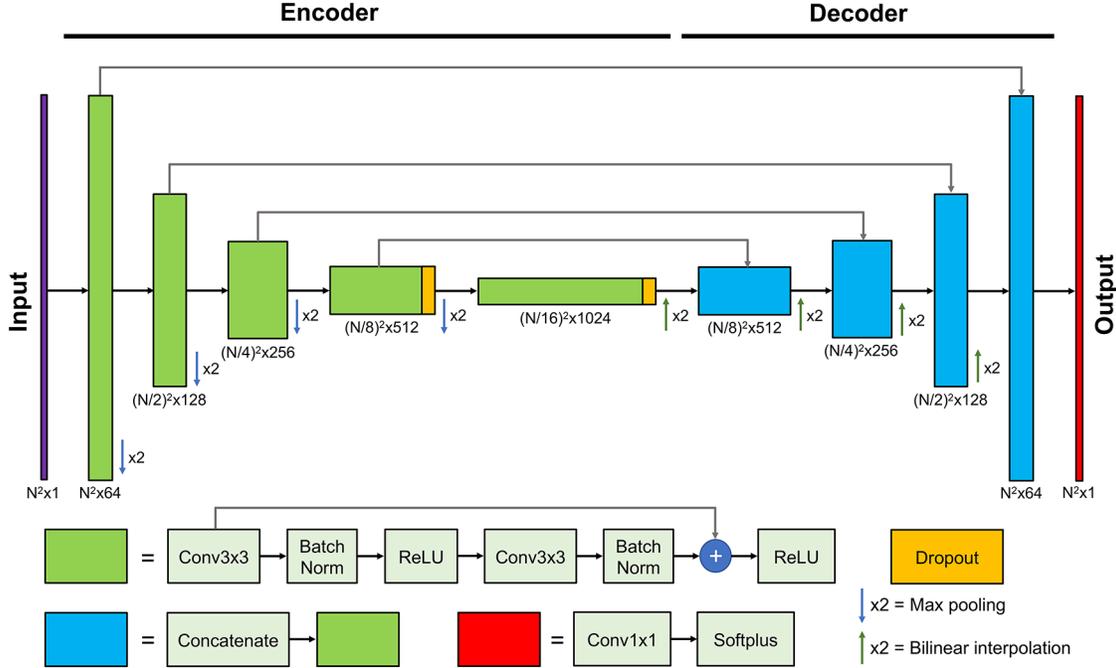


Figure 4.2: The proposed DeepSNiF network structure.

are excluded from training. In IMC images, pixels with intensity value 0 are considered as background. Afterwards, we define the background pixel ratio r as ratio of the number of background pixels and that of total pixels in the patch. Patches are considered as the background regions if $r \leq \rho$, where ρ is the threshold and set from 0.2 to 0.99 for different channels and datasets. We applied a smaller ρ for the datasets less sparse images and vice versa. For good generalization ability of the network, we recommend at least 5000 patches for training. Before training, all the generated patches were percentile normalized (99.9 to 99.999). The percentile of 99.9 was applied for those training sets with extremely bright markers and larger percentile with relatively homogeneous intensity distributions. To balance the training efficiency and accuracy, 0.2% pixels of each patch are masked and replaced by their neighbours using a stratified sampling strategy [96]. Finally, 85% of the patches are set as training set and the rest as validation set.

All models were trained using Keras [110] (version 2.3.1) on a single NVIDIA Quadro RTX 6000 GPU with 24 GB of VRAM. Adam optimizer [111] was applied as the optimization algorithm with a initial learning rate of 0.001 for 200 epochs and batch size of 128. Learning rate is multiplied by 0.6 if validation loss does not improve for 20 epoches. The training details for all the datasets are summarized as Tables A.5–A.10.

Given a trained denoising model, we denoise full-size IMC images to avoid edge stitching effects. In order to achieve end-to-end prediction, we pad pixels around each image so their width and height are the multiples of 16 with reference to the network architecture (Figure 4.2). The padding pixels are the replications of the border pixels. Before prediction, the IMC images are normalized by the pre-calculated maximum of the corresponding channels in the training set. The outputs of the network are re-scaled and set as the denoised images. Given the trained denoising model, inference is fast. We are able to denoise IMC images with pixels of 1000 by 1000 less than 1 second per image on a single NVIDIA Quadro RTX 6000 GPU.

4.3 Reference IMC Denoising Methods

4.3.1 Hot pixel removal methods

Currently, two thresholding methods are mostly applied to remove hot pixels, which are neighbour-based threshold hot pixel removal method [58, 30, 31] (NTHM) and median-based threshold hot pixel removal method [27] (MTHM). NTHM is simple and straightforward. It pre-sets a threshold σ_{thresh} . Then single pixels with intensity greater than this threshold of the maximum value in its 3×3 neighbourhood are detected as outliers. This method is very similar to our DIMR algorithm with hyper-parameter $l = 1$, which will always overlook

consecutive hot pixels. In addition, σ_{thresh} needs to be manually set. However, the hot pixels in different tissues and channels have different scales. That is, one fixed threshold does not work for different images. These limitations downgrades its performance in real applications. The NTHM algorithm is summarized as Algorithm 5, and the default value of σ_{thresh} is set 50. MTHM is an automated approach and can remove consecutive hot pixels, of which two

Algorithm 5 NTHM for hot pixel removal

Initialization: Raw image \mathbf{R} ;

Hyper-parameter σ_{thresh} ;

Output: Hot pixel removed image;

- 1: In a sliding 3×3 window, select the maximum value after excluding the center pixel;
 - 2: **if** The difference between the center pixel and the maximum value is larger than σ_{thresh}
then
 - 3: Substitute the center pixel's value with the maximum value.
 - 4: **end if**
-

thresholds σ_{thresh1} and σ_{thresh2} are needed to be manually set. This method first searches the top σ_{thresh1} pixels and regards any pixels as outliers if they are σ_{thresh2} times higher than the median in a sliding 5×5 window. However, this method may falsely remove normal pixels located at the border between tissues and background. Additionally, with a larger σ_{thresh1} or smaller σ_{thresh2} , false negatives may also be generated. Normally, σ_{thresh1} and σ_{thresh2} are set as 2% and 4 separately [27]. This algorithm is summarized as Algorithm 6. The default values of σ_{thresh1} and σ_{thresh2} are set as 2% and 4, respectively.

4.3.2 Deep learning-based shot noise filtering methods

N2V algorithm

N2V uses noisy images as the input and output to train a denoising network. It randomly masks several pixels and replaces them with their neighbours. As Equation (4.27) shows,

Algorithm 6 MTHM for hot pixel removal

Initialization: Raw image \mathbf{R} ;

Hyper-parameters σ_{thresh1} and σ_{thresh2} ;

Output: Hot pixel removed image;

- 1: Search all the pixels with top σ_{thresh1} values in an image;
 - 2: **if** The pixels are larger than σ_{thresh2} times of the medians in their 5×5 window **then**
 - 3: Substitute the pixels' values with the medians.
 - 4: **end if**
-

minimizing the mean squared error (MSE) between the noisy image pairs is equal to minimizing that of the noisy and clean image pairs plus the variance of the noise. This equality holds as long as the function \mathcal{F} is “ \mathcal{J} -invariant” [106], which means $\mathcal{F}[f(\mathbf{R})]_p - x_p$ and $r_p - x_p$ are independent for any pixel p . Notably, the current masking strategies always fulfill this requirement if the noise is pixel-independent and the pixels are not replaced by themselves.

$$\mathcal{L}_{\text{MSE}}(\mathbf{R}, \mathcal{F}_\theta[\mathbf{R}_{\text{masked}}]) = E(\mathcal{F}_\theta[f(\mathbf{R})]_p - r_p)^2 \quad (4.27)$$

$$= E(\mathcal{F}_\theta[f(\mathbf{R})]_p - x_p)^2 + E(r_p - x_p)^2 \quad (4.28)$$

In Equation (4.28), the second term corresponds to the noise variance at pixel p , which is a constant for a Gaussian noise model. In this condition, it can be neglected in the training process, and hence the self-supervised learning loss is equal to a supervised one.

However, this statement is not true for a Poisson noise model. The variance of Poisson noise is equal to its true signal intensity, i.e., $E(r_p - x_p)^2 = E(x_p)$. Therefore, the pixels with higher signal will gain more attention in Equation (4.28), resulting in sub-optimal denoising for signal-limited area. In fact, even in a supervised learning task, MSE is still not an unbiased estimator for Poisson noise. When true signals are available, the optimizer can be derived by the MLE framework:

$$\hat{\theta}_{\text{MLE}} = \operatorname{argmin}_\theta E[-\log(\Pr(\mathcal{F}_\theta(\mathbf{R})_p | x_p; \theta))]. \quad (4.29)$$

Approximating a Poisson process as Gaussian distribution, Equation (4.29) is converted as

$$\hat{\theta}_{\text{MLE}} = \operatorname{argmin}_{\theta} E \left[\frac{1}{2} (\mathcal{F}_{\theta}(\mathbf{R})_p - x_p)^2 / x_p + \frac{1}{2} \log(2\pi x_p) \right]. \quad (4.30)$$

Obviously, minimizing this loss in Equation (4.30) is not equivalent to minimizing the MSE between $\mathcal{F}_{\theta}(\mathbf{R})_p$ and x_p , because x_p varies with different pixels p . As a result, MSE loss will generate bias in Poisson denoising. Beside this issue, in the original version of N2V, a linear activation is used as the output layer’s activation function of the network, which violates the non-negativity of IMC images.

Modified N2V algorithm with the Anscombe transformation and ReLU activation (MN2V)

We propose a MN2V algorithm to correct the bias of N2V. The Anscombe transformation is first applied to the IMC data, so that the Poisson noise is approximated as a Gaussian-distributed noise model. In this sense, the noise variance term in Equation (4.27) can be regarded as a constant and the bias from MSE loss will be mitigated. By substituting linear activation with ReLU [112], the non-negativity of IMC images is satisfied. Note that the denoised images should be re-transformed to their original scale by the exact unbiased inverse Anscombe transformation [104]. The exact unbiased inverse Anscombe transformation software package was downloaded from <https://webpages.tuni.fi/foi/invansc/>.

However, the Anscombe transformation is not accurate for very low values [100]. Unfortunately, there are usually a portion of pixels in IMC images suffering from extremely low counts. The variances of these pixels are still positively correlated with the counts. Thus, even this approach reduces the bias of N2V, some errors are still inevitable.

Compared to the MSE loss used by N2V and Noise2Self, our derived loss function is the optimal estimator for Poisson denoising and the corresponding outputs are inherently non-negative with a softplus function. Consequently, it does not generate any biases for IMC denoising.

Noise2True (N2T) algorithm

In N2T [91], clean images are available as ground truths, so that supervised learning is possible. To achieve a learning with the optimal estimator, we use our derived I-divergence Equation (4.15) as the loss function in N2T. In simulation, we compare all the above denoising methods with N2T. We are especially curious as to how our Hessian norm regularization could help boost self-supervised learning to approach the performance of N2T. The ground truths of IMC images can only be acquired in simulation. Thus, N2T is in fact not possible in real IMC denoising.

4.3.3 Traditional statistics-based shot noise filtering methods

Gaussian filter-based denoising algorithm

Gaussian filter might be the most widely used noise filter. For all of our IMC images, we apply a Gaussian filter with kernel size of 5×5 and standard deviation of 0.8. Gaussian filter can only remove the high-frequency noise with the risk of filtering lots of true signal. As a result, its performance is sub-optimal than other smarter denoising algorithms.

NLM algorithm

NLM algorithm [94] takes a mean of all pixels in an image, weighted by how similar these pixels are to the target pixel. In this section, this algorithm is implemented using the Matlab built-in function “imnlmfilt”. All the parameters are set as default.

BM3D algorithm with Anscombe transformation

BM3D [95] is modified from NLM algorithm, which is usually regarded as a state-of-the-art denoising algorithm due to its superior performance in multiple applications. Instead of simply averaging the pixel values, it collects similar patches of an image, and then applies hard thresholding and Wiener filtering in two stages. BM3D is built on a white Gaussian noise model. Therefore, the Anscombe transformation is applied to approximate the Poisson noise in IMC images to Gaussian noise. The same applies for the MN2V algorithm, the exact unbiased inverse Anscombe transformation is used to rescale the denoised images.

The BM3D algorithm software package was downloaded from https://webpages.tuni.fi/foi/GCF-BM3D/index.html#ref_software. When implementing BM3D, we set $\sigma_{\text{noise}} = 1$, $N_2 = 8$, $N_s = 17$, $\tau_{\text{match}} = 2500$, $\lambda_{\text{thr3D}} = 1$, $N_{S_{\text{wiener}}} = 25$ and $\tau_{\text{match-wiener}} = 600$. All the other parameters were set as default.

4.4 Analyzing IMC Image Quality Enhancement through Simulation Verification

Due to the difficulty to acquire ground truths, it is infeasible to quantitatively evaluate the accuracy of hot pixel removal methods and the shot noise filtering algorithms using real IMC

images. As a consequence, we propose to conduct a comprehensive and accurate quantitative evaluation with simulated data.

4.4.1 Simulated data generation

We modify Equation (2.2) as Equation (4.31) to generate simulated data.

$$\mathbf{R} = \mathcal{P}[\mathbf{X}_{\text{origin}}/\gamma] + \mathbf{Q}, \quad (4.31)$$

where $\mathbf{X}_{\text{origin}}$ represent the original clean images used for simulation and γ is the scale factor to control the overall ion counts level. A larger α indicates the overall ion counts of the ground truth image $\mathbf{X}_{\text{GT}} = \mathbf{X}_{\text{origin}}/\gamma$ is lower, and hence the shot noise level is higher. On the other hand, we assume hot pixels \mathbf{Q} obey a negative binomial distribution $\text{NB}(\tau, \eta)$. Additionally, another parameter ω is used to indicate the density of hot pixels.

Here we select the original clean images $\mathbf{X}_{\text{origin}}$ from the *t*-CyCIF dataset [113] because of their very high SNR and similar resolution (1.06 μm) with IMC images. In particular, we choose a cell marker CD14 from a lung tissue to evaluate the shot noise filtering algorithms. These two channels along with a DNA channel are used to evaluate the hot pixel removal methods.

We have four parameters γ , τ , η and ω to determine the noise conditions of the simulated images. For each channel, we set 4 different γ which correspond to 4 different SNR levels including high, medium, low and very low. With different noise levels we set different τ , η and ω to represent different conditions of hot pixels. All the parameters are listed in Table 4.1.

Table 4.1: Simulation parameters

Noise settings	1				2			3			4		
SNR levels	high				medium			low			very low		
Parameters	τ	γ	η	ω									
CD14	3	2000	0.05	1%	5000	0.1	0.1%	10000	0.2	0.01%	16000	0.3	0.01%
DNA	5	500	0.05	3%	1000	0.1	1%	2000	0.2	0.1%	5000	0.25	0.01%

4.4.2 Accuracy metrics and statistical analysis

In simulation, RMSE is used to evaluate the performance of the hot pixel removal methods, as Equation (4.32) shows, in which p is the pixel index, $\mathbf{Y}^{\text{clean}}$ and \mathbf{Y}^{HM} are the simulated images without hot pixels and hot pixel removed images, separately.

$$\text{RMSE}(\mathbf{Y}^{\text{HM}}, \mathbf{Y}^{\text{clean}}) = \sqrt{\frac{1}{P} \sum_{p=1}^P (\mathbf{Y}_p^{\text{HM}} - \mathbf{Y}_p^{\text{clean}})^2} \quad (4.32)$$

The peak SNR (PSNR) and SSIM are used to evaluate the performance of the shot noise filtering algorithms. PSNR indicates the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is defined as Equation (4.33), where \mathbf{Y}^{est} is the estimated image from denoising algorithms and \mathbf{Y}^{true} is the ground truth.

$$\text{PSNR}(\mathbf{Y}^{\text{est}}, \mathbf{Y}^{\text{true}}) = 20 \log \frac{\max(\mathbf{Y}^{\text{true}})}{\sqrt{\frac{1}{P} \sum_{p=1}^P (\mathbf{Y}_p^{\text{est}} - \mathbf{Y}_p^{\text{true}})^2}} \quad (4.33)$$

In this chapter, the SSIM is defined as Equation (4.34),

$$\text{SSIM}(\mathbf{Y}^{\text{est}}, \mathbf{Y}^{\text{true}}) = \frac{2\mu_{\mathbf{Y}^{\text{est}}}\mu_{\mathbf{Y}^{\text{true}}} + C_1}{\mu_{\mathbf{Y}^{\text{est}}}^2 + \mu_{\mathbf{Y}^{\text{true}}}^2 + C_1} \cdot \frac{2\sigma_{\mathbf{Y}^{\text{est}}\mathbf{Y}^{\text{true}}} + C_2}{\sigma_{\mathbf{Y}^{\text{est}}}^2 + \sigma_{\mathbf{Y}^{\text{true}}}^2 + C_2} \quad (4.34)$$

where \mathbf{Y}^{est} is the estimated image from denoising algorithms, \mathbf{Y}^{true} is the ground truth, $\mu_{\mathbf{Y}^{\text{est}}}$, $\mu_{\mathbf{Y}^{\text{true}}}$, $\sigma_{\mathbf{Y}^{\text{est}}}$, $\sigma_{\mathbf{Y}^{\text{true}}}$ and $\sigma_{\mathbf{Y}^{\text{est}}\mathbf{Y}^{\text{true}}}$ are the local means, standard deviations and cross-covariance for images \mathbf{Y}^{est} and \mathbf{Y}^{true} , C_1 and C_2 are the regularization constants to avoid instability for image regions where the local mean or standard deviation is close to zero.

In simulation, all the RMSE, PSNR and SSIM data are presented as box-and-whisker plots (center line, median limits, 75% and 25% whiskers, maximum and minimum) along with all the data points. We use the paired one-way analysis of variation to do the multiple comparisons of these accuracy metrics. All the statistical tests are implemented with Prism 9 (GraphPad Software Inc.). Statistical significance at $P < 0.05$, 0.01, 0.001 and 0.0001 are denoted by *, **, *** and ****, respectively. “ns” means “no significance”.

4.4.3 Hot pixel removal methods evaluation

Optimal iteration number selection and running time evaluation

First, we evaluated the impact of N_{iter} to the hot pixel removal results in DIMR. To achieve this, we utilized the simulated DNA dataset with 4 hot pixel conditions, in which each condition contains 50 images (4.1). N_{iter} was set from 1 to 5. Besides, l was empirically set as 4 as we consider half adjacent pixels should be close to a center pixel in a 3×3 window [114, 115]. RMSE was utilized to evaluate the results, which is listed in Figure 4.3a. From the results, the accuracy almost does not improve after 3 iterations in all the noise settings. In particular, the DIMR algorithm stops the iteration before it reaches to the pre-set iteration times in low hot pixel density conditions (noise settings 3 and 4). Therefore, we recommend

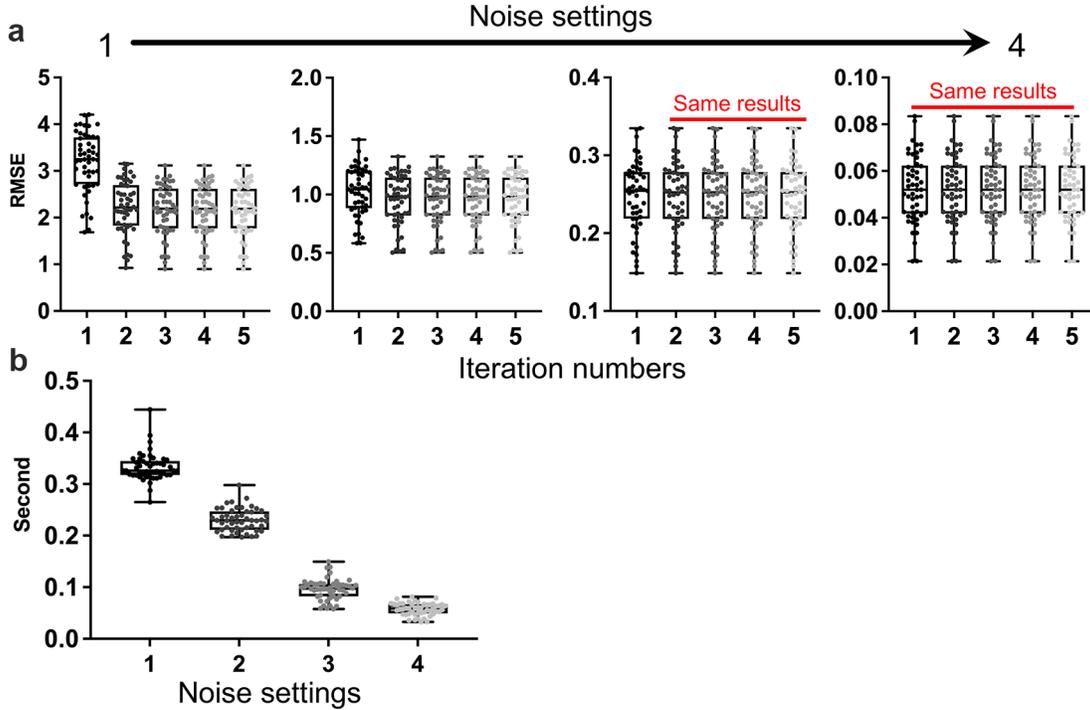


Figure 4.3: DIMR algorithm evaluations on the simulated DNA dataset ($N = 50$ per noise setting). (a) Evaluation of DIMR with N_{iter} from 1 to 5. “Same results” indicate the iteration stops beforehand. For instance, in noise setting 3, DIMR stops after the second iteration even though N_{iter} is larger than 2. (b) DIMR running time evaluations under different noise settings with $N_{\text{iter}} = 3$.

N_{iter} as 3 to adequately remove hot pixels. We also evaluate the running time of DIMR under different hot pixels conditions with $N_{\text{iter}} = 3$ (Figure 4.3b). The results indicate that it takes from approximately 0.05 to 0.4 second for DIMR to remove hot pixels, depending on the hot pixel densities.

Benchmarking DIMR to other hot pixel removal methods

Subsequently, we benchmarked our DIMR algorithm with NTHM and MTHM on the simulated CD14 and DNA datasets. Each marker contains 4 hot pixel conditions and each condition contains 50 images. We set the hyperparameters of DIMR, l and N_{iter} , as 4 and

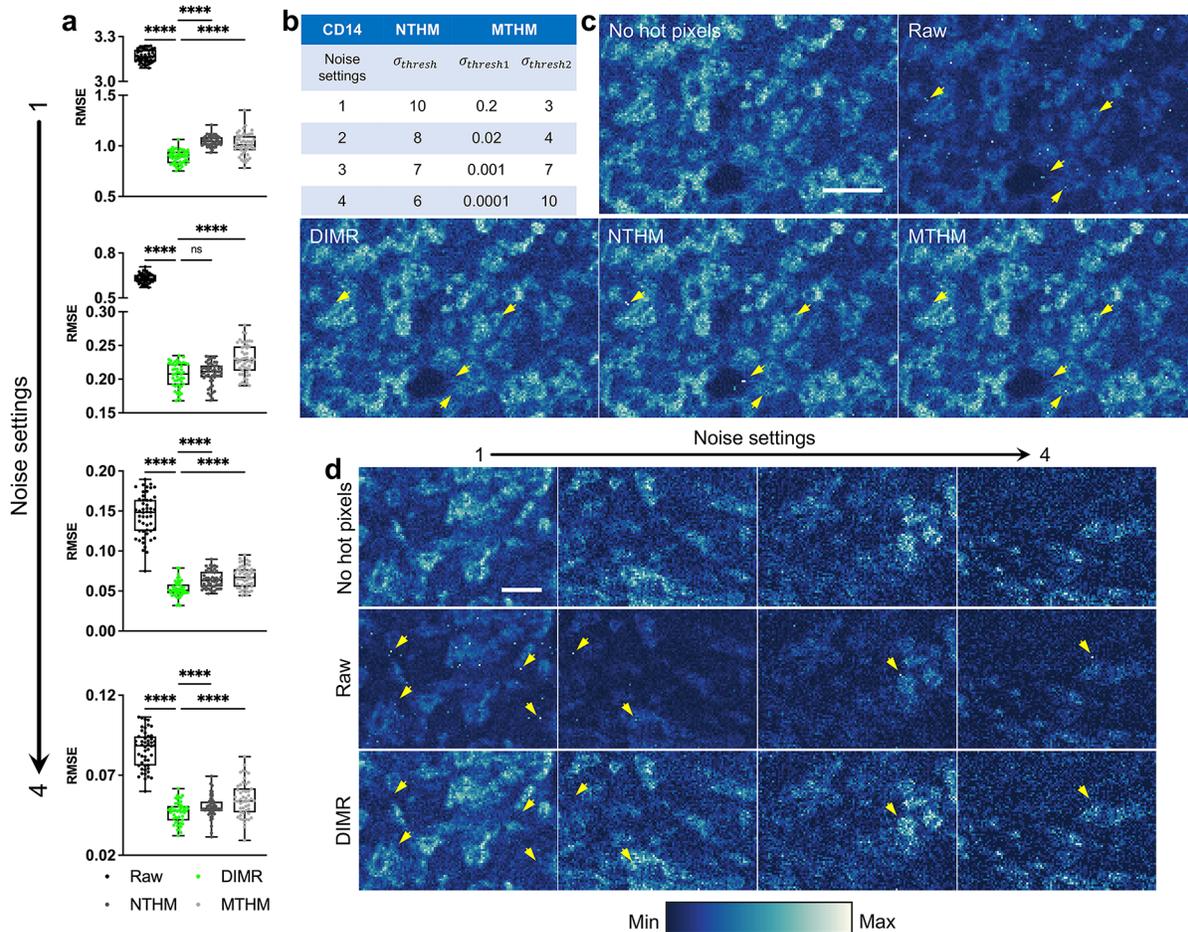


Figure 4.4: Evaluation of the hot pixel removal methods on the simulated CD14 dataset ($N = 50$ per noise setting). (a) RMSE comparison under different noise settings. (b) The optimal hyper-parameters of NTHM and MTHM under different noise settings. (c) Visual inspection of the three hot pixel removal methods on a CD14 image with the highest hot pixel density under noise setting 1. (d) Visual inspection of the DIMR algorithm under different noise settings. Scale bar: (c) $20 \mu\text{m}$, (d) $20 \mu\text{m}$.

3 for all the datasets. The hyperparameters of NTHM and MTHM were optimized for each noise setting to guarantee their best performance.

The evaluation results for these two markers are listed in Figures 4.4 and 4.5, respectively. As these figures suggest, DIMR always removes hot pixels effectively and outperforms NTHM

and MTHM on RMSE under different hot pixel conditions. Besides, the optimal hyper-parameters of NTHM and MTHM vary in a wide range with different settings of hot pixels (Figures 4.4b and 4.5b), while those of DIMR remain the same ($l = 4$ and $N_{\text{iter}} = 3$). In fact, the hot pixel conditions vary under different images and markers. Therefore, it is labor-intensive to tune the hyper-parameters for every image. In comparison, the outlier detection of DIMR is based on the overall statistical features of the images, so that it is not essential to tune its hyper-parameters. To summarize, the DIMR algorithm is more accurate and flexible for removing hot pixels in IMC images.

4.4.4 Shot noise filtering methods evaluation

Compare DeepSNiF with N2V and MN2V

We have also compared our DeepSNiF algorithm to N2V, MN2V and N2T on the simulated CD14 images under 4 different noise levels. For each condition, we generated training sets and trained separate networks with 25 images for DeepSNiF, N2V, MN2V and N2T, and restored the other 25 images with the trained denoisers. Note that here DeepSNiF with no regularization (DeepSNiF-NR) was applied because we aimed to compare the performances of MSE and Equation (4.15) as loss functions in the IMC denoising task. The denoised images were evaluated both visually and quantitatively, shown in Figure 4.6.

In Figure 4.6a, N2T is always the best performer because of the availability of ground truths in training. Nevertheless, in self-supervised learning algorithms, DeepSNiF-NR wins over N2V and MN2V on both PSNR and SSIM. The restored image qualities are also reflected in Figures 4.6b, c, in which saturated maps are applied to better visualize the differences between the algorithms. The restored images of N2V and MN2V are noisier than the DeepSNiF ones, which is more noticeable on low intensity regions. As we analyzed, this is because

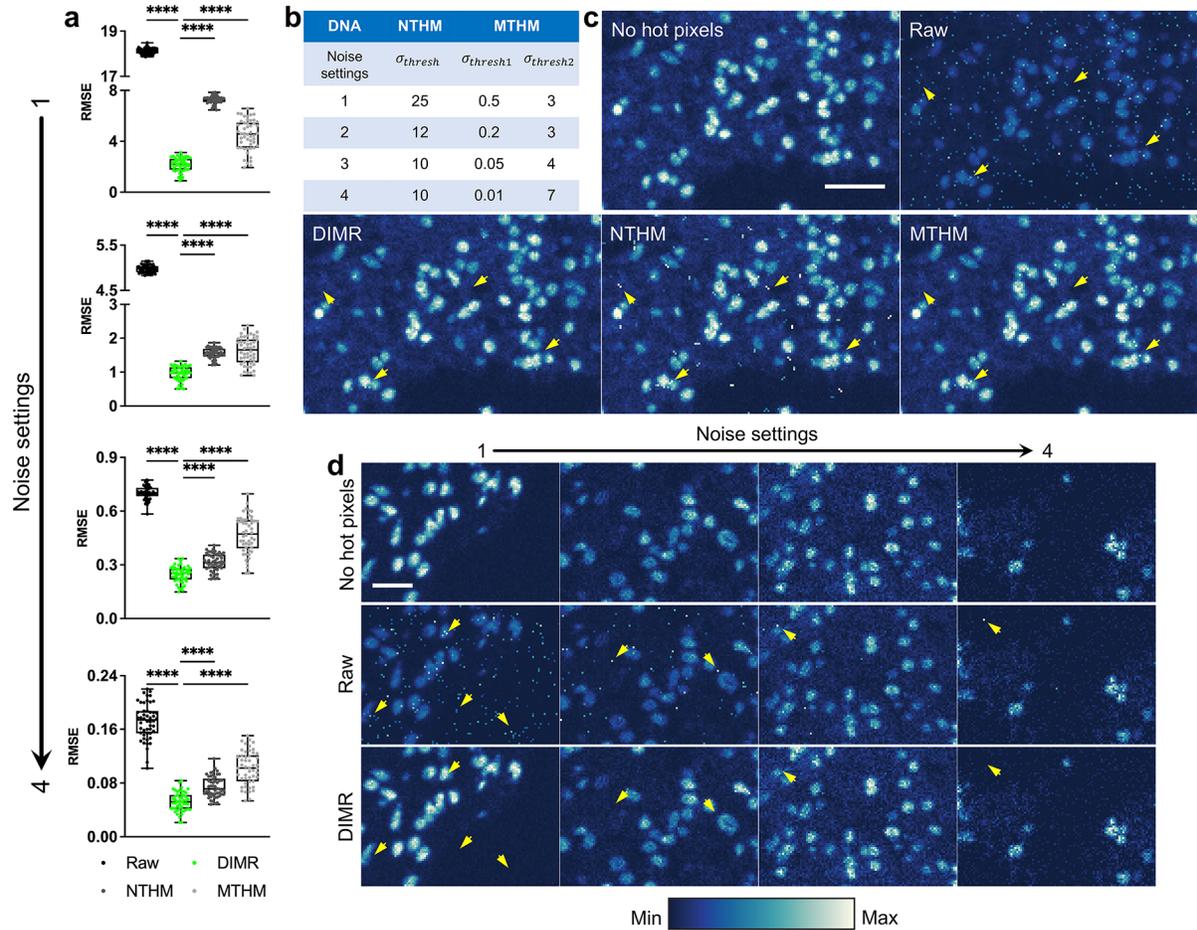


Figure 4.5: Evaluation of the hot pixel removal methods on the simulated DNA dataset ($N = 50$ per noise setting). (a) RMSE comparison under different noise settings. (b) The optimal hyper-parameters of NTHM and MTHM under different noise settings. (c) Visual inspection of the three hot pixel removal methods on a DNA image with the highest hot pixel density under noise setting 1. (d) Visual inspection of the DIMR algorithm under different noise settings. Scale bar: (c) $20 \mu\text{m}$, (d) $20 \mu\text{m}$.

MSE is a biased estimator for Poisson noise, even with the Anscombe transformation. To conclude, we have verified that our derived I-divergence is a better estimator than MSE on Poisson denoising task, and our derived DeepSNiF framework is more capable to restore IMC images than N2V and MN2V.

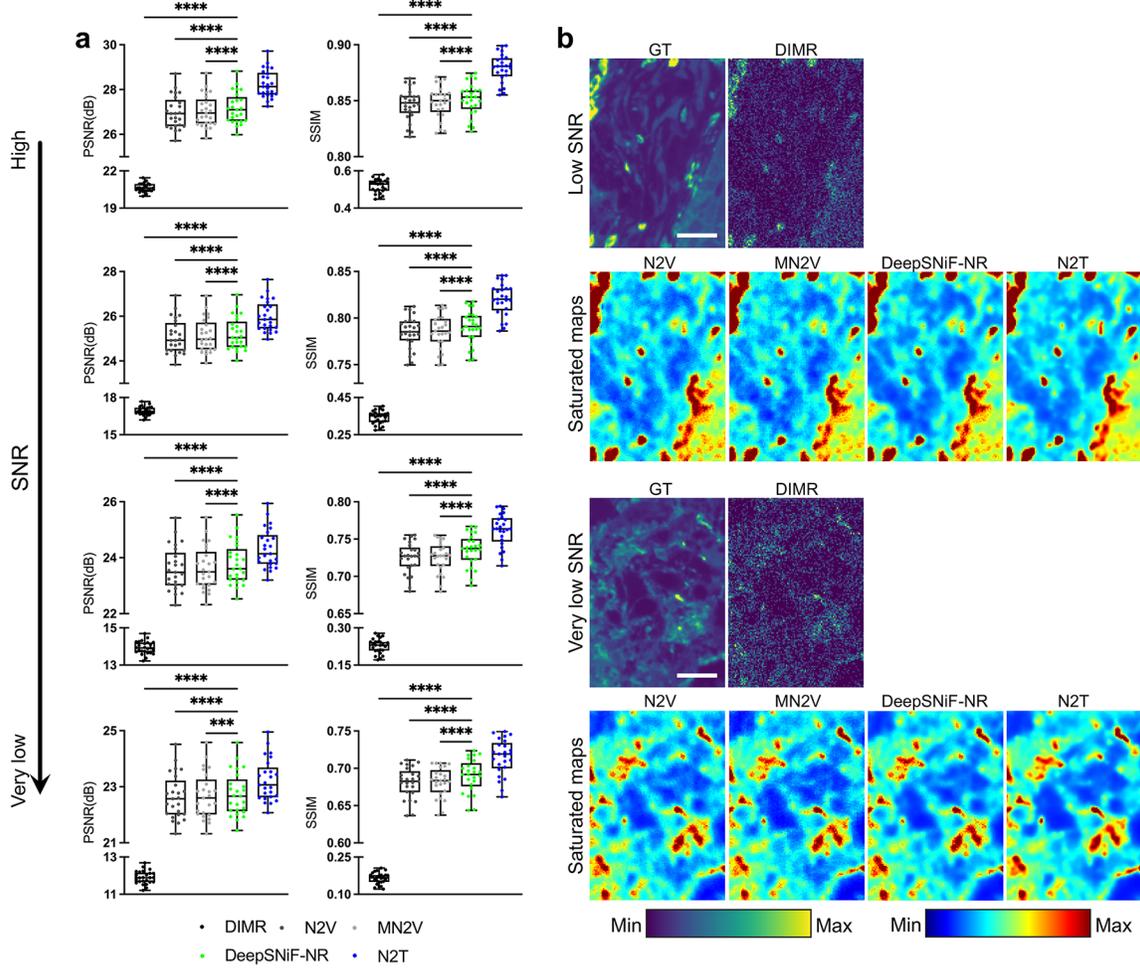


Figure 4.6: Comparison of DeepSNiF-NR with N2V and MN2V on the simulated CD14 dataset with 4 noise levels ($N = 25$ per level). (a) PSNR and SSIM evaluation on the three algorithms. (b) Visual inspection of the three algorithms and N2T on denoising a low SNR image. (c) Visual inspection of the three algorithms and N2T on denoising a very low SNR image. Scale bar: (b) $20 \mu\text{m}$, (c) $20 \mu\text{m}$.

The effect of Hessian norm regularization on DeepSNiF

Even though DeepSNiF-NR is superior to N2V and MN2V for the task of IMC denoising, it still suffers from discontinuous artifacts and there remains a big gap between the performance of DeepSNiF and N2T in terms of PSNR and SSIM (Figure 4.7). To further enhance

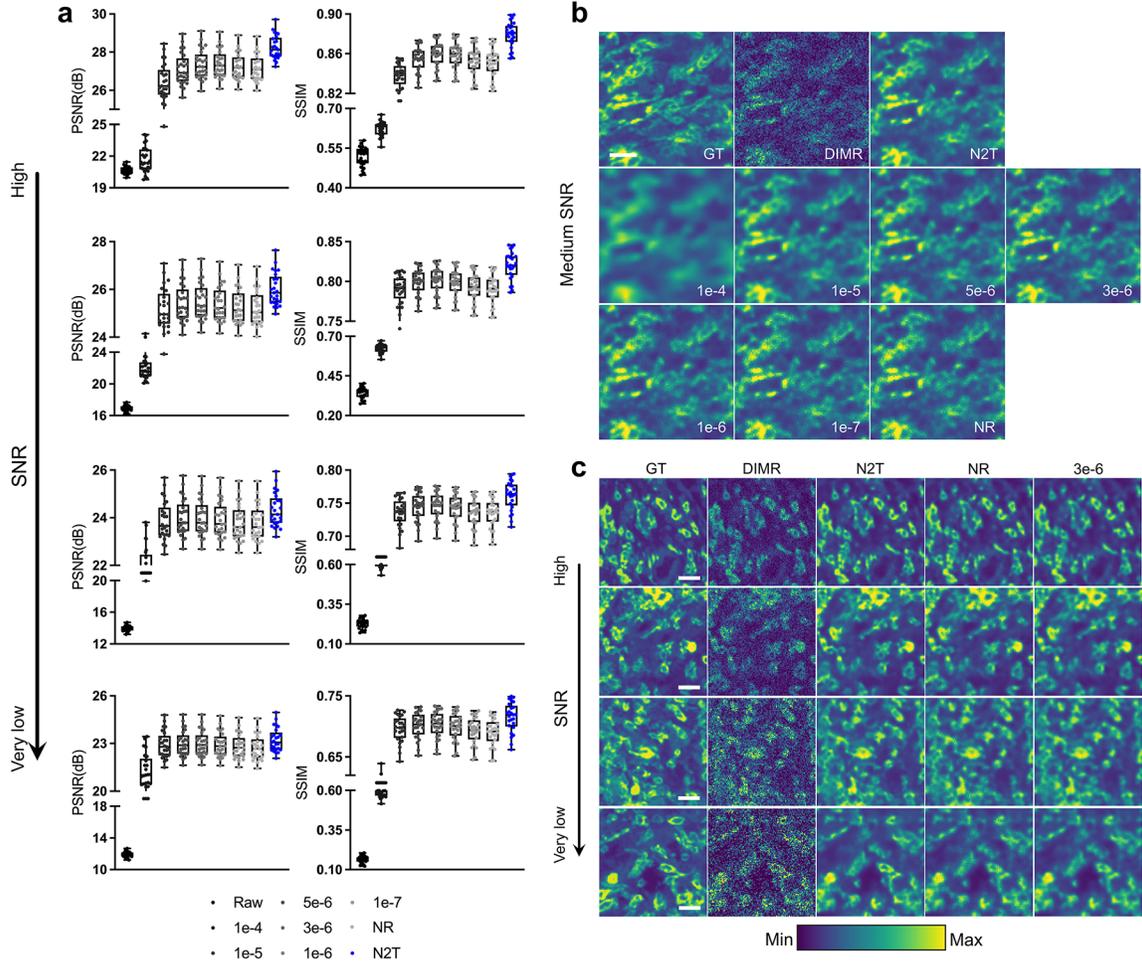


Figure 4.7: The effect of Hessian norm regularization on the simulated CD14 image denoising ($N = 25$ per level). (a) PSNR and SSIM evaluation for DeepSNiF with different Hessian norm regularization parameters λ_{Hessian} . (b) Visual inspection of DeepSNiF with different Hessian norm regularization parameters λ_{Hessian} on a simulated CD14 image with medium SNR. (c) Visual inspection of DeepSNiF when $\lambda_{\text{Hessian}} = 0$ (NR) and $\lambda_{\text{Hessian}} = 3e-6$ on simulated CD14 images with different noise levels. Scale bar: (b) $20 \mu\text{m}$, (c) $20 \mu\text{m}$.

DeepSNiF, the Hessian norm regularization is applied in the loss function (4.23). In particular, λ_{Hessian} determines the strength of regularization and thus further affects the denoising performance. To choose a proper λ_{Hessian} , we trained a series of networks with a range of λ_{Hessian} from $1e-4$ to $1e-7$ and no regularization (NR) on the simulated CD14 images. Here, 25 images were used for training and the other 25 for testing (in each dataset).

The quantitative and qualitative results are presented in Figure 4.7. The PSNR and SSIM (Figure 4.7a) indicate Hessian norm regularization could enhance the performance of DeepSNiF with a proper λ_{Hessian} . Specifically, we found DeepSNiF achieves its optimal performance when λ_{Hessian} is approximately $3\text{e-}6$. We then investigated the qualities of the restored images from the medium SNR datasets (Figure 4.7b). The larger λ_{Hessian} , the stronger regularization will be. Therefore, the images look more blurry when $\lambda_{\text{Hessian}} = 1\text{e-}4$ and noisier with no regularization. With $\lambda_{\text{Hessian}} = 3\text{e-}6$, the restored images achieve their best qualities. Subsequently, by observing the denoised images from different noise levels (Figure 4.7c), the Hessian norm regularization effectively modifies the discontinuous artifacts in the images without regularization. We conclude that the Hessian norm regularization effectively improves the performance of DeepSNiF, both quantitatively and qualitatively, even boost the performance of DeepSNiF close to that of N2T. As a consequence, we set DeepSNiF with $\lambda_{\text{Hessian}} = 3\text{e-}6$ for the subsequent experiments due to its optimal accuracy.

Compare DeepSNiF with traditional denoising methods

We are also interested in comparing DeepSNiF with traditional statistical denoising methods, including Gaussian filter (GAUSS), NLM and BM3D algorithms. These three algorithms along with DeepSNiF were benchmarked on the simulated CD14 datasets. Notably, they were only applied on the 25 test images of DeepSNiF. The results are presented in Figure 4.8.

From these quantitative evaluations (Figure 4.8a), we can observe that DeepSNiF is frequently better than the other algorithms on PSNR and SSIM, except for high SNR images. This is also in accordance with the visual inspection (Figure 4.8b). For high SNR images, BM3D performs slightly better than DeepSNiF. Nevertheless, as SNR becomes lower, BM3D

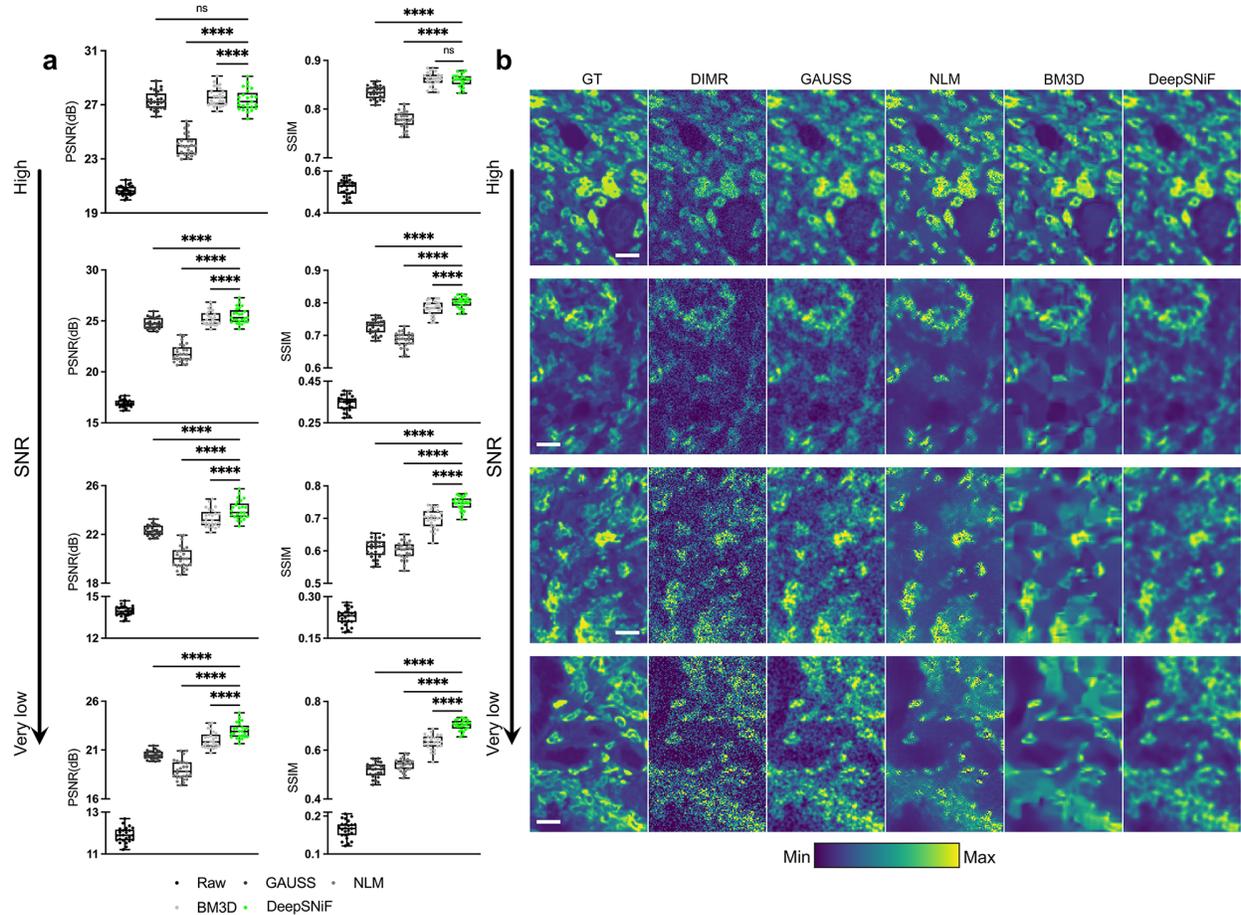


Figure 4.8: Comparison of DeepSNiF with GAUSS, NLM and BM3D algorithms on the simulated CD14 dataset with 4 noise levels ($N = 25$ per level). (a) PSNR and SSIM evaluation on the algorithms. (b) Visual inspection of the algorithms on denoising the images with different noise levels. Scale bar: $20 \mu\text{m}$.

tends to over-smooth the restored images and details are lost. In fact, the SNR of IMC images are always low because of the limited ion counts. As a result, BM3D is not fit for IMC denoising, even though its PSNR and SSIM are the closest to those of DeepSNiF. GAUSS and NLM are outperformed by BM3D, and therefore not of interest for IMC data processing. In particular, the GAUSS algorithm can be regarded as a low-pass filter and only the high frequency noises are filtered. Similar to BM3D, NLM algorithm also blurs the restored images. However, it still leaves noisy regions and its denoising ability is even weaker than GAUSS. To

conclude, traditional statistical denoising algorithms are not suited to IMC denoising, and deep learning-based algorithms such as DeepSNiF may be preferred.

4.5 Experimental Verification of Enhanced Image Quality in IMC

4.5.1 Material preparation

Human bone marrow dataset

Sections were cut in 4–6 micron thickness from formalin-fixed paraffin-embedded (FFPE) blocks of ethylenediaminetetraacetic acid (EDTA)-decalcified bone marrow trephine biopsy specimens. Three patients demonstrated normal morphology, and 4 patients were diagnosed with myelodysplastic syndromes (MDS), with additional later timepoints obtained at disease progression including acute myeloid leukemia (AML). Use of specimens for secondary analysis in this study was approved by the Washington University in St. Louis Institutional Review Board (#201912110).

Tissue staining and IMC data acquisition

Descriptions of cell markers and isotope tags are provided in Tables A.1–A.4. Staining was performed according to Fluidigm IMC recommendations for FFPE as follows. Briefly, tissue sections were dewaxed in xylene and rehydrated in a graded series of alcohol. Epitope retrieval was conducted in a water bath at 96 °C in Tris-EDTA buffer at pH 9 for 30 minutes, then cooled and washed in metal-free PBS. Blocking with Superblock (ThermoFisher) plus 5% FcX TruBlock (Biolegend) was followed by staining with antibody cocktail prepared in

0.5% BSA and metal-free PBS overnight at 4 °C. Sections were washed in 0.02% TritonX100 followed by metal-free PBS, then nuclear staining was performed using 1:200 or 1:300 dilution of Intercalator-Ir (125 μ M, Fluidigm) solution for 30 minutes, followed by ddH₂O for 5 minutes. Slides were air-dried before IMC measurement.

The abundance of bound antibody was quantified using the Hyperion imaging system (Fluidigm) controlled by CyTOF Software (version 7.0.8493), with UV-laser set at 200 Hz. Count data were then converted to tiff image stacks for further analysis using MCD Viewer (version 1.0.560.6, Fluidigm) or imctools (Bodenmiller lab, <https://github.com/BodenmillerGroup/imctools>).

Tissue staining and IF data acquisition

For IF staining, tissue was prepared as described above, then stained overnight at 4 °C with a single metal-conjugated primary antibody (CD3, CD4, CD169 or CD61 in Table A.4). The single-stained tissue was washed, then stained with secondary antibody (donkey anti-rabbit AF647 or goat anti-mouse AF750, Invitrogen, 2mg/mL diluted 1:400 in 0.5% BSA in PBS) at room temperature 1 hour, followed by additional washing in PBS and DAPI (1 μ g/mL) staining. Slides were mounted with SlowFade Glass antifade reagent (ThermoFisher) and # 1 1/2 coverslips. Images were acquired using Leica DMi8 inverted widefield microscope with Lumencor SOLA SE U-nIR light engine, DAPI/FITC/TRITC/Cy5/Cy7 filters, DFC9000 GT sCMOS camera, PL APO 20x/0.80 objective and LAS X software (version 3.7.3.23245). After image acquisition, coverslips were removed with gentle agitation in PBS, then Ir-intercalator staining, washing and drying performed as above for subsequent Hyperion data acquisition.

4.5.2 Accuracy metrics and statistical analysis

For the real experimental data, five types of metrics were used for quantitative evaluations. STDB and CNR (Equation 3.50) were used to evaluate the noise level and contrast of IMC images. In CNR, the signal and background regions of IMC images are manually annotated. Pearson’s correlation coefficient (PCC) was used as the metric to reflect the similarity between two groups of data. The PCC between measured data \mathbf{Y} and the reference \mathbf{Y}_{ref} is defined as Equation (4.35),

$$\text{PCC}(\mathbf{Y}, \mathbf{Y}_{\text{ref}}) = \frac{E[(\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y}_{\text{ref}} - \mu_{\mathbf{Y}_{\text{ref}}})]}{\sigma_{\mathbf{Y}}\sigma_{\mathbf{Y}_{\text{ref}}}}, \quad (4.35)$$

where $\mu_{\mathbf{Y}}$ and $\mu_{\mathbf{Y}_{\text{ref}}}$ are the mean values of images \mathbf{Y} and \mathbf{Y}_{ref} , respectively; $\sigma_{\mathbf{Y}}$ and $\sigma_{\mathbf{Y}_{\text{ref}}}$ are the standard deviations of \mathbf{Y} and \mathbf{Y}_{ref} , respectively; and E represents arithmetic mean.

Here, the STDB, CNR and PCC data are presented as box-and-whisker plots (center line, median limits, 75% and 25% whiskers, maximum and minimum) along with all the data points. We use the paired one-way analysis of variation to do the multiple comparisons of these accuracy metrics. All the statistical tests are implemented with Prism 9 (GraphPad Software Inc.). Statistical significance at $P < 0.05$, 0.01, 0.001 and 0.0001 are denoted by *, **, *** and ****, respectively. “ns” means “no significance”.

4.5.3 DIMR removes hot pixels effectively in IMC images

We initially tested our DIMR algorithm on selected markers of a human bone marrow dataset. Here, inherently high autofluorescence and tissue features (fragile haematopoietic stroma intermixed with dense cortical bone) excluded other spatial biology methods, even after

substantial pre-processing. Figure 4.9a enumerates the proportion of hot pixels detected by DIMR for each marker. We then selected DNA intercalator (Figure 4.9b) to evaluate DIMR due to the high hot pixel density. By comparing the images and the corresponding histograms, hot pixels are effectively eliminated by DIMR. We further compared DIMR with NTHM and MTHM with default parameters on the data, to benchmark its performance. From the results, DIMR performs better than the other methods by visual inspection. The results along with the simulation data results demonstrate the versatility and accuracy of DIMR. DIMR also results in the additional benefit of moderately improved cell segmentation, the result of robust removal of artifacts caused by hot pixels (Figure 4.10).

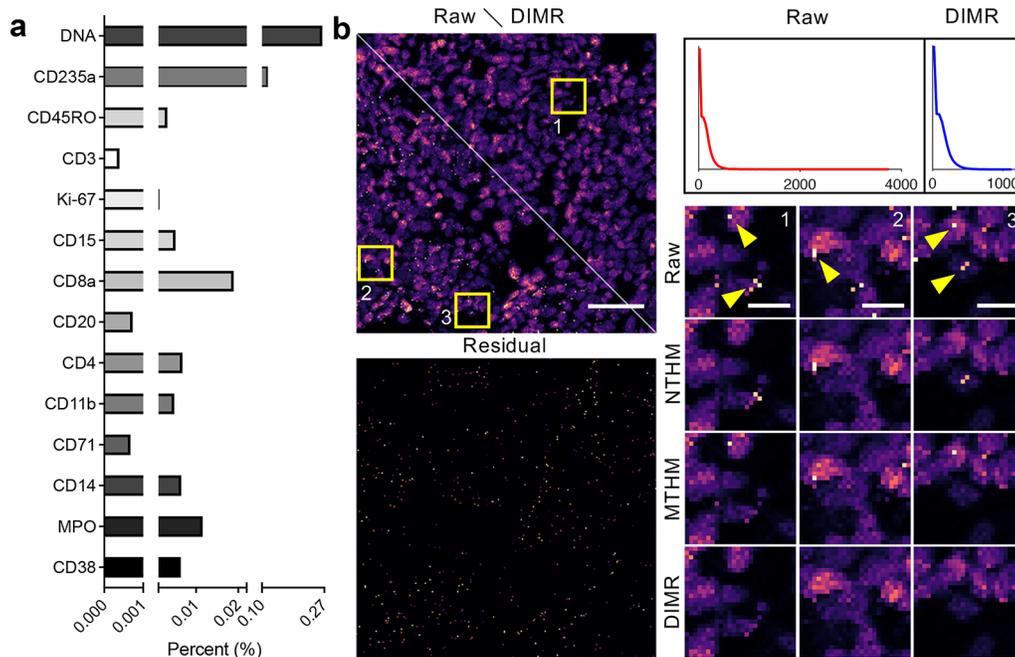


Figure 4.9: Validation of DIMR on the human bone marrow IMC dataset. (a) The fractions of detected hot pixels by DIMR in selected channels. (b) DIMR removes hot pixels in DNA intercalator channel effectively. Left: Comparison of the raw and DIMR-processed images; and the difference between the images, in which Residual corresponds to the detected hot pixels. Upper right: The corresponding histograms of the raw and DIMR-processed images. Lower right: Comparisons between the raw, NTHM, MTHM and DIMR processed images. Scale bar: left: 75 μm , right: 8 μm .

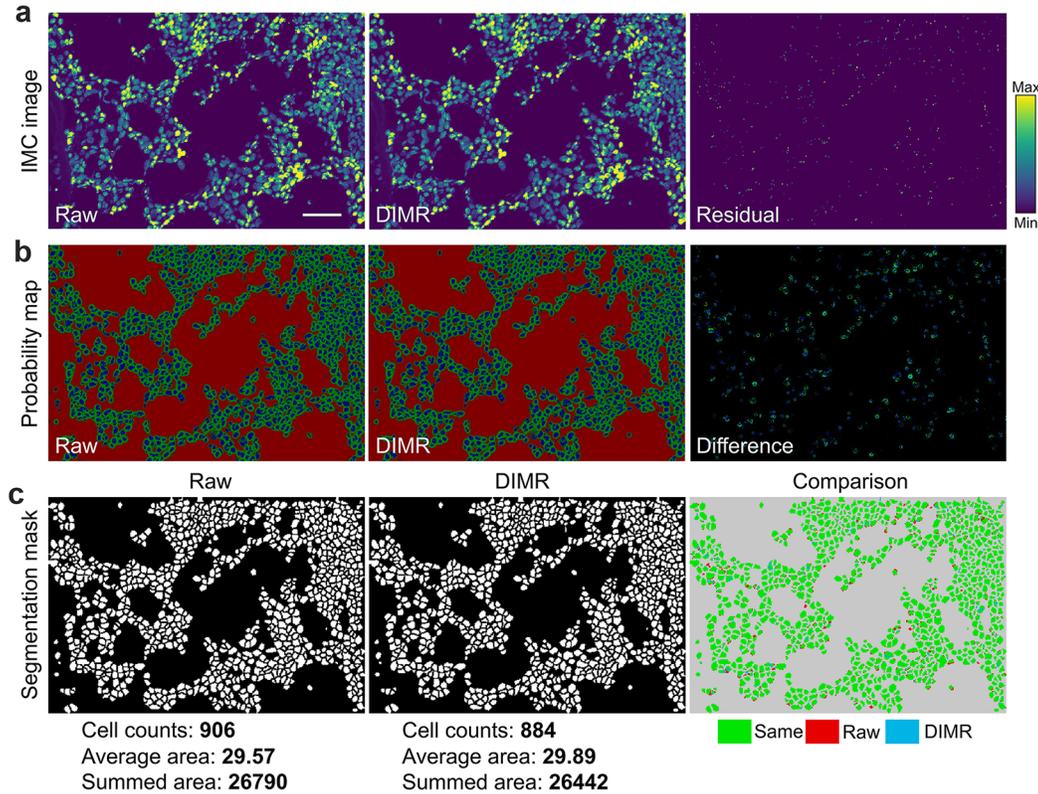


Figure 4.10: DIMR slightly enhances single cell segmentation. (a) Comparison between the raw and DIMR-processed DNA images. (b) The probability maps of the raw and DIMR-processed images generated by the Ilastik software, and their difference map (blue: nuclei, green: cytoplasm, and red: background). (c) The segmentation masks of the raw and DIMR-processed images generated by Cell Profiler software, and their overlaid comparison map. By comparing the figures in the third column, the different segmented masks between raw and DIMR images are frequently caused by the presence of hot pixels. Even with a cell size threshold, the hot pixels can still split cells, and falsely expand normal cell borders. Correspondingly, the raw image segmented a little more cells than that of DIMR (906 to 994). At the same time, the average area of cells from the raw image is a little smaller (29.57 to 29.89), but the summed area is slightly larger (26790 to 26442). Scale bar: 47 μm .

4.5.4 DeepSNiF enhances image quality significantly in IMC

With hot pixels removed from image data, we next benchmarked the denoising performance of DeepSNiF along with DIMR and other statistics-based methods on the experimental dataset. First we tested IMC images labeled with Collagen III from the human bone marrow

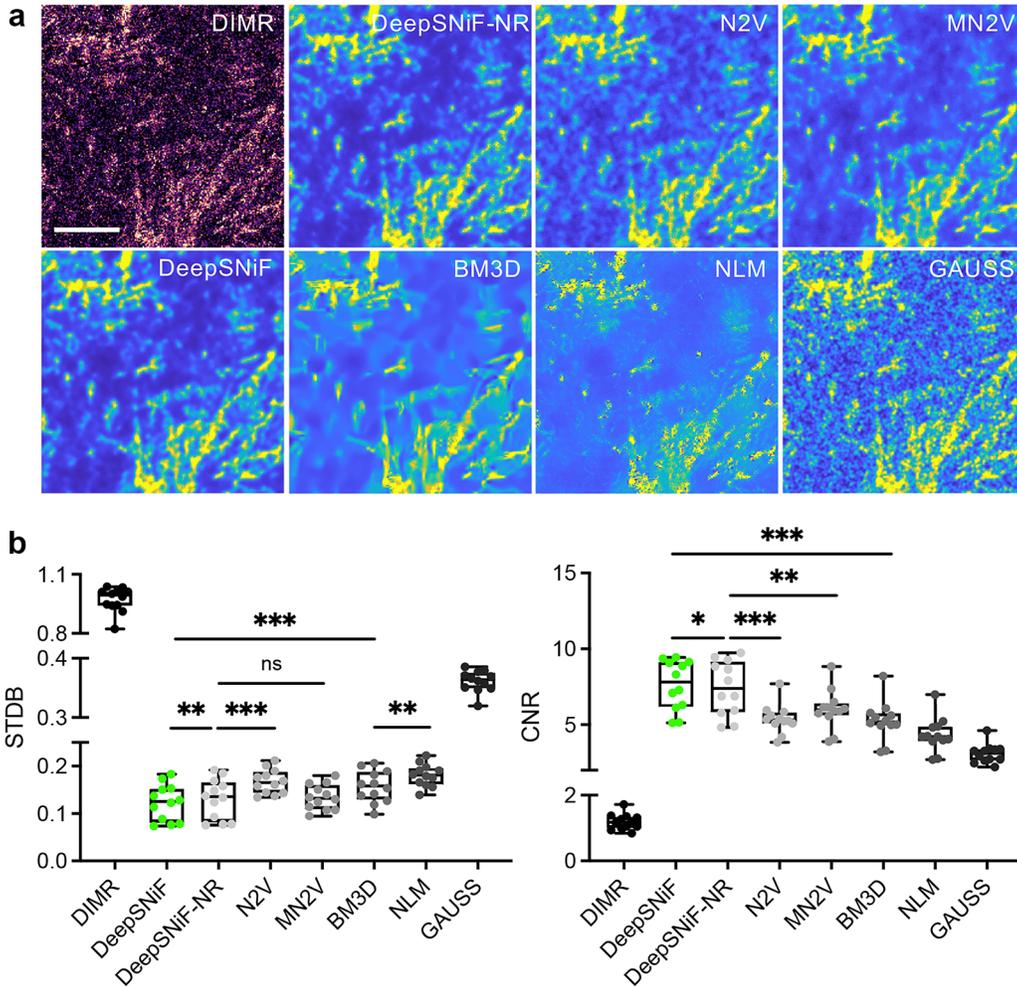


Figure 4.11: Validation of DeepSNiF on IMC images labeled with Collagen III. (a) Visual inspection of DeepSNiF and other statistics-based denoising algorithms. (b) DeepSNiF performs significantly better than other algorithms ($N = 12$) on denoising Collagen III-labeled IMC images in terms of STDB and CNR. Scale bar: $50 \mu\text{m}$.

dataset (Figure 4.11). We found all the algorithms enhanced the DIMR data even though variant performances were achieved. GAUSS lowers the noise level by sacrificing resolution. NLM is effective at background denoising but does not account adequately for the noise components of signal. BM3D improves NLM further by its cooperative denoising procedure. However, we found it tended to over-smooth foreground and distorted cell shapes. N2V always generates artifacts because of an inappropriate noise model. DeepSNiF-NR performs

better than MN2V because the Anscombe transformation in MN2V may generate some bias for extremely low counts; both of which are better than GAUSS, NLM and BM3D. DeepSNiF further enhances these results by mitigating the discontinuities in the DeepSNiF-NR output, and furthermore retains cell morphology features. We then quantitatively compared the differently processed images across STDB and CNR. All results indicated DeepSNiF enables the optimal denoising performance among these algorithms. In particular, the noise level (STDB) decreased by 87% and CNR increased by 5.6-fold after DeepSNiF (0.9938 to 0.1254 and 1.1749 to 7.8065, median value).

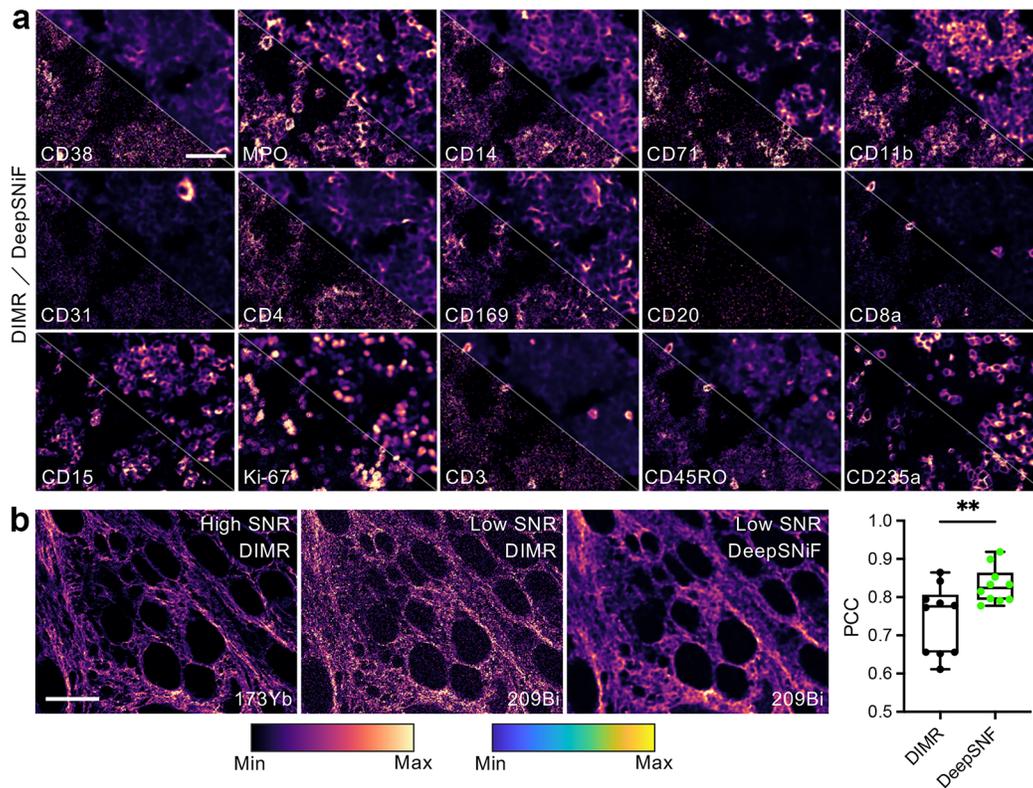


Figure 4.12: Further inspection of DeepSNiF on experimental data. (a) Visual inspection of DeepSNiF denoised IMC images labeled with other markers. (b) DeepSNiF improves the Pearson correlations between Collagen III-labeled IMC images with low and high SNR significantly ($N = 10$). Scale bar: (a) $45 \mu\text{m}$. (b) $100 \mu\text{m}$.

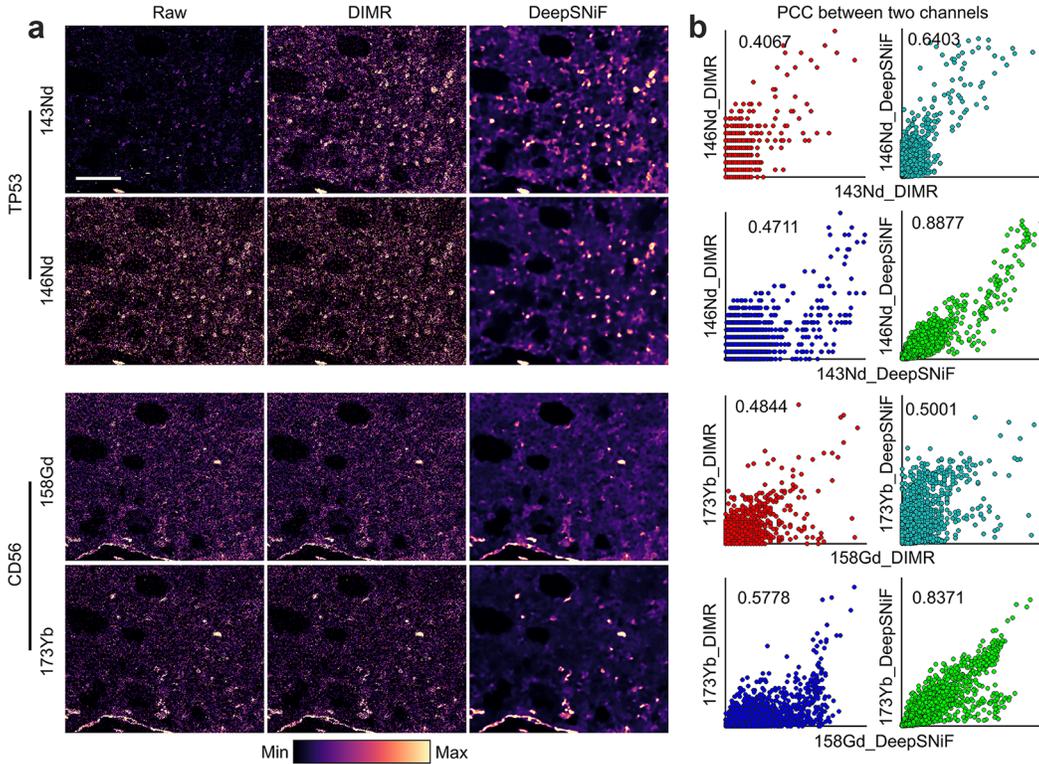


Figure 4.13: IMC-Denoise enhances both IMC image quality and the Pearson correlations of IMC images labeled with the same markers. (a) A tissue was stained by 143Nd, 146Nd conjugated TP53, and 158Gd, 173Yb conjugated CD56, respectively, with different SNRs. The IMC images were firstly processed by DIMR to remove hot pixels. Then DeepSNiF was employed to improve the image quality of all the images, because the qualities of the higher SNR images are still sub-optimal. (b) After DeepSNiF processing, the Pearson correlation coefficients (PCC) improved, in which those of the double DeepSNiF-processed images are the highest. Notably, the DeepSNiF trained by the CD3 images from the human bone marrow dataset (Tables A.4 and A.7) was used to denoise the DIMR-processed TP53 and CD56 images, due to their highly similar features and the lacking of sufficient TP53 and CD56 training sets. Scale bar: 75 μm .

We further visually inspect the denoising results of IMC-Denoise on multiple datasets on human bone marrow images (Figure 4.12a). Image quality improvements that enhance image interpretation are apparent, in particular for low SNR channels. Two orthogonal staining approaches were pursued in order to provide further validation of these image quality improvements. Firstly, the same antibody was conjugated to two different metals and co-stained

on the same tissue for detection in high and low sensitivity channels, without spillover. IMC-Denoise was employed on the low signal channel (209Bi) and was able to restore the image quality to match the high sensitivity channel (173Yb) with the Pearson correlation coefficient (PCC) improved as high as 0.16, as shown in representative images (Figure 4.12b).

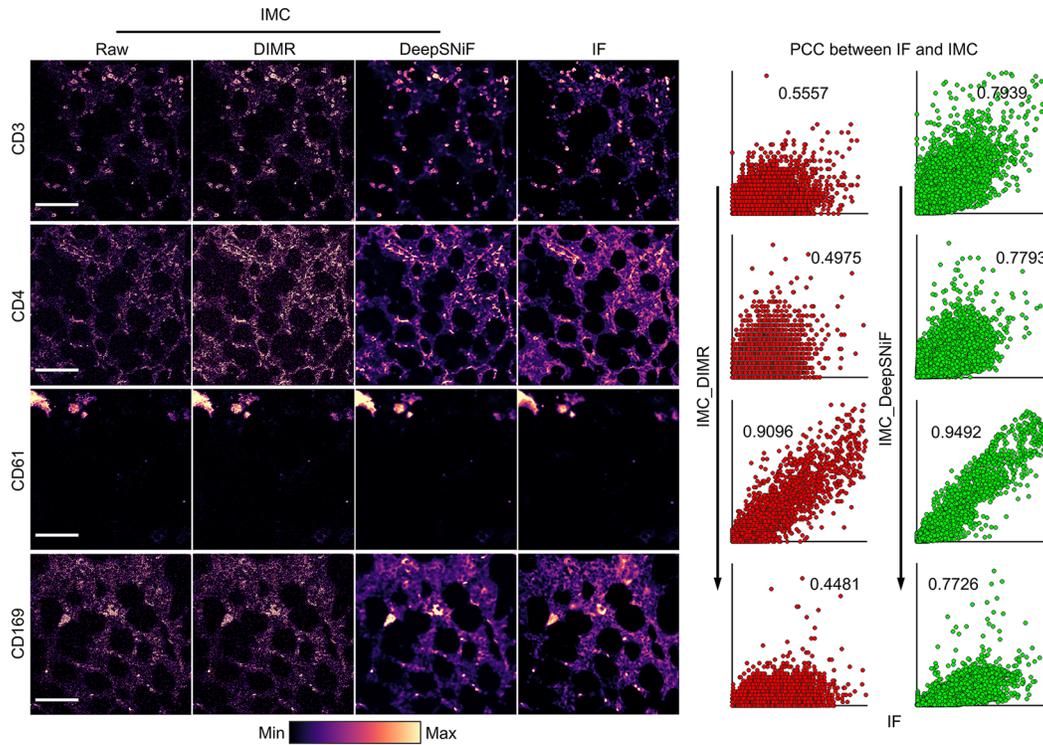


Figure 4.14: IMC-Denoise enhances both IMC image quality and the Pearson correlations between IMC and IF images. (a) The same tissues were stained with CD3, CD4, CD61 and CD169 by IMC and IF, respectively. The low SNR IMC images were processed by DIMR to remove hot pixels and then by DeepSNiF to improve image quality. (b) After DeepSNiF processing, the PCC between IMC and IF improved, indicating DeepSNiF is able to improve the IMC image quality. Scale bar: CD3: 98 μm . CD4: 110 μm . CD61: 69 μm . CD169: 87 μm .

Similar conclusions can also be drawn from other channels with increased PCC by more than 0.48 and 0.35, respectively (Figure 4.13). Secondly, tissue sections stained with metal-conjugated antibodies (for CD3, CD4, CD61 and CD169) were probed with a fluorophore-conjugated secondary antibody for IF, individually. We then followed IF imaging by ablative-IMC (Figure 4.14). The additional handling and washing after IF imaging often leads to extremely low remaining metal isotope signal; however, enhancement in image quality can still be observed to restore the image to correlate to the IF. Specifically, the PCC quantitatively verified the image quality improvement of DeepSNiF (CD3: 0.5557 to 0.7939, CD4: 0.4975 to 0.7793, CD61: 0.9096 to 0.9492, and CD169: 0.4481 to 0.7726).

4.6 Experimental Verification of Enhanced Downstream Analysis in IMC

4.6.1 Reference methods in downstream analysis

Semi-automated Ilastik-based background noise removal

The semi-automated strategy in [59] utilizes Ilastik segmentation [116] to remove background noise in IMC images. An expert annotates signal or background regions of IMC images, and then Ilastik trains a random forest classifier for background noise removal. To achieve good denoising quality, large areas of background require manual labeling, which is labor-intensive. Furthermore, low image quality may affect the accuracy of this method as well. After background removal, the images are binarized to solve batch effect issues. Then the single cell information is calculated by counting the positive signal frequency rather than the mean intensity of every single cell. Here, we only utilized Ilastik (version 1.3.2post1) for

background noise removal of IMC images, and still applied the mean intensities as the single cell profiles. To better reveal the enhancement by DeepSNiF, we applied the same labels for the trainings of DIMR and DeepSNiF-processed images.

MAUI

MAUI software package [26, 57] includes spillover correction, noise removal and aggregate removal. All three steps require expert observation, which is also labor-intensive. Here, we only benchmarked the noise removal method in MAUI with our DeepSNiF algorithm. Briefly, it calculates the distances between a non-zero pixel and its K nearest non-zero neighbours, then builds a histogram based on the summations of the distances for all the non-zero pixels. Next, a threshold is manually selected to remove the pixels with larger summations by observing the distribution of the histogram. This method is based on the assumption that noisy regions look more sparse than normal regions. MAUI was implemented by the software package from <https://github.com/angelolab/MAUI>. The parameter K and the threshold were manually tuned to guarantee the best performance of MAUI (Figure 4.15).

4.6.2 Accuracy metrics and statistical analysis

In this section, F1 score was used to evaluate the accuracy of background noise removal. F1 score and Jaccard score were used to evaluate the accuracy of cell annotation of B, CD8 T and CD4 T cells, which can be formulated as Equation (4.36) and Equation (4.37), respectively.

$$\text{F1 score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (4.36)$$

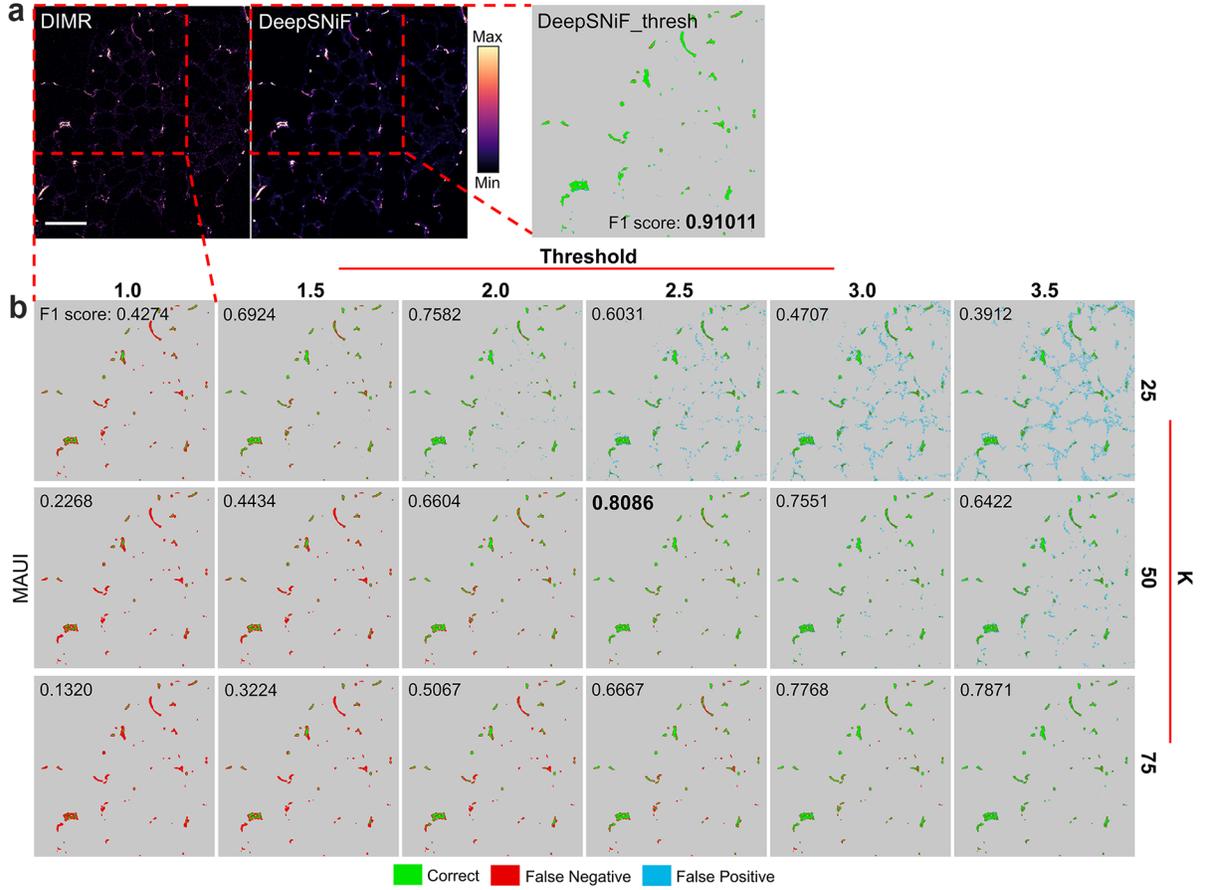


Figure 4.15: DeepSNiF performs better than MAUI on DIMR data to filter background noise. (a) DIMR and DeepSNiF-processed IMC images labeled with CD34, respectively. The DeepSNiF-processed image was binarized by the threshold value 1 and then overlaid with manual annotated ground truth. (b) The DIMR-processed image was processed by the MAUI software package with a wide range of parameters to select the best background noise removal result and also overlaid with the manual annotated ground truth. The F1 score of the DeepSNiF_thresh (lower right corner) result is always higher than that of the MAUI results (upper left corner in every image), indicating DeepSNiF is better than MAUI on DIMR data in terms of background noise removal. Scale bar: 96 μm .

$$\text{Jaccard score} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (4.37)$$

where TP, FP and FN are the pixel number of true positives, false positives and false negatives, respectively. All of the evaluation process was implemented with customized MATLAB

(R2021a, MathWork) scripts. RMSE, PSNR, SSIM, PCC, F1 score and Jaccard score were computed using MATLAB built-in functions.

Here, the F1 score and Jaccard score data are presented as box-and-whisker plots (center line, median limits, 75% and 25% whiskers, maximum and minimum) along with all the data points. We use the paired one-way analysis of variation to do the multiple comparisons of these accuracy metrics. All the statistical tests are implemented with Prism 9 (GraphPad Software Inc.). Statistical significance at $P < 0.05$, 0.01, 0.001 and 0.0001 are denoted by *, **, *** and ****, respectively. “ns” means “no significance”.

4.6.3 IMC-Denoise enables background noise removal of IMC images and enhancement of IMC downstream analysis

We next evaluated the ability of DeepSNiF in IMC-Denoise to remove background noise of IMC images. We manually annotated 15 images labeled with CD34 and 12 IMC images labeled with Collagen III (Figure 4.16a). The single threshold-based method and semi-automated Ilastik-based method [59] were applied on both DIMR and DeepSNiF-processed CD34 and Collagen III images (DIMR_thresh, DeepSNiF_thresh, DIMR_illastik and DeepSNiF_illastik, respectively), and MAUI was only applied on DIMR images. The results were compared with the manually annotated ground truths (Figure 4.16b), and F1 score was set as the accuracy metric to quantitatively assess the results (Figure 4.16c). Note that in threshold-based methods, optimal thresholds from 1 to 4 were selected for individual DIMR-processed images per marker for fair comparison. Nevertheless, the single threshold 1 was selected for all the images per marker for DeepSNiF-processing, without the need of further tuning.

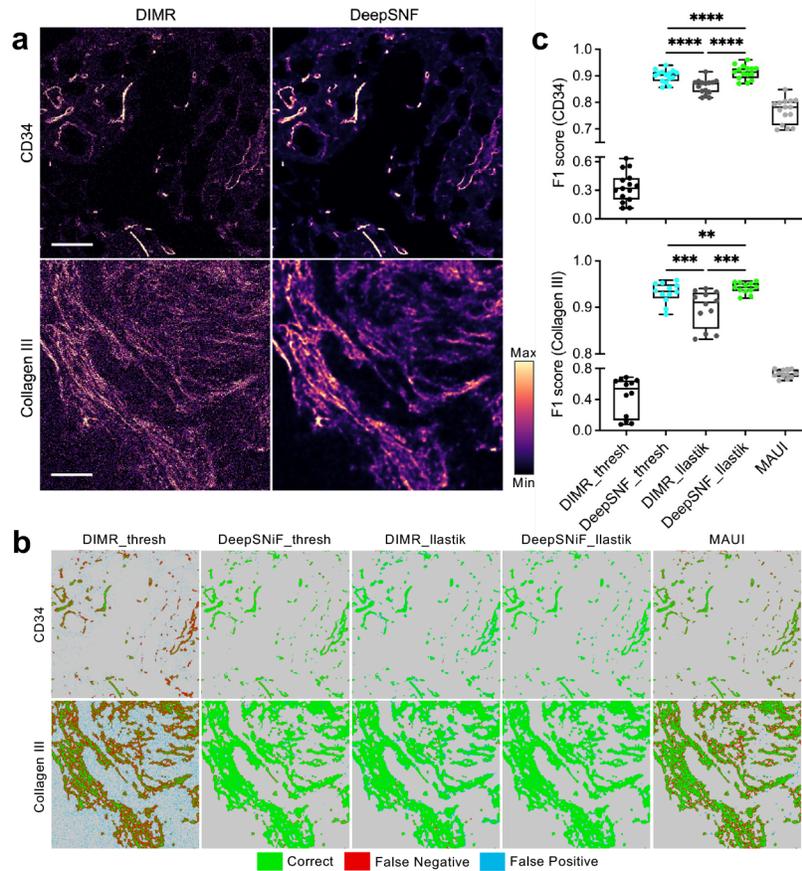


Figure 4.16: IMC-Denoise enables background noise removal of the human bone marrow IMC dataset. (a) Examples of DIMR and DeepSNiF-processed IMC images labeled with CD34 and Collagen III. (b) Visual inspection of background removal results of DIMR and DeepSNiF-processed images, in which DIMR_thresh and DeepSNiF_thresh are binarized with the optimal thresholds, DIMR_Ilastik and DeepSNiF_Ilastik are segmented by the Ilastik software package, and MAUI results are the DIMR images processed by the MAUI software package, respectively. Manual annotated images are served as ground truths. (c) After DeepSNiF denoising, the background removal accuracy improves significantly in terms of F1 score, for both CD34 ($N = 15$) and Collagen III-labeled images ($N = 12$). Notably, DeepSNiF_Ilastik achieves the highest accuracy, while DeepSNiF_thresh performs better than all the background removal results from DIMR images. Scale bar: Top: $50 \mu\text{m}$, bottom: $35 \mu\text{m}$.

Overlaid masks and F1 scores for both markers indicated DeepSNiF_Ilastik achieves the highest accuracy while DIMR_thresh is the weakest performer (CD34: 0.9143 to 0.4155, and Collagen III: 0.9434 to 0.5378, median value). Surprisingly, DeepSNiF_thresh is a better method for background noise removal than the semi-automated DIMR_Ilastik (CD34:

0.9040 to 0.8716, and Collagen III: 0.9345 to 0.9108, median value), and its F1 score was improved by approximately twofold compared to DIMR_thresh. We infer that DeepSNiF is capable of unmixing the signal and background, while the shot noise in DIMR images hinders the performances of the Ilastik-based method. MAUI was able to account for the background noise at the cost of false negative generation (CD34: 0.7824 and Collagen III: 0.7305, median value). Indeed, the signal has been unmixed from background through DeepSNiF because we have proved a simple thresholding can remove background accurately (Figure 4.16b and c). These findings support replacement of tedious semi-automated approaches by automated DeepSNiF.

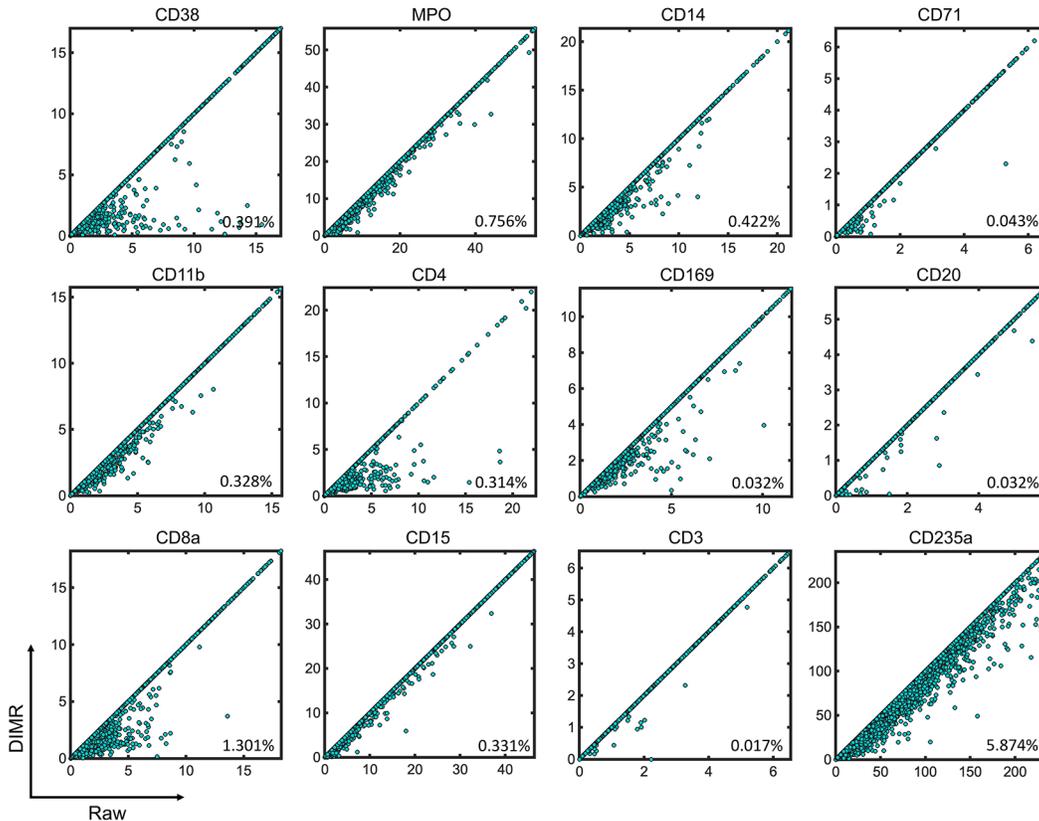


Figure 4.17: The impact of DIMR on single cell data extracted from DeepSNiF-based cell segmentation masks. Each sub-figure represents the one-on-one relationship between the raw and DIMR data of a particular marker in single cell scale. The bottom right value in each sub-figure represents the percentage of the difference between the raw and DIMR data.

Next, we were curious to evaluate the impact of IMC-Denoise on single cell profiles. In single cell segmentation, the pixels in each image were defined as belonging to the nucleus, cytoplasm, or background compartment using the pixel classification module of Ilastik as described in <https://github.com/BodenmillerGroup/ImcSegmentationPipeline>. In our experiments, the DIMR and DeepSNiF-processed images were both set as inputs for cell mask segmentation. The DeepSNiF-based cell masks were primarily used for further analysis, while those based on DIMR were validated for robustness in some cases. The Random Forest classifier was trained on the channels including CD38, MPO, CD14, CD71, CD11b, CD4, CD20, CD8a, CD15, Ki-67, CD3, CD45RO CD235a, Histone-H3 and Iridium. This allowed for the generation of maps that integrate for each pixel the probability of belonging to each of three compartments. Based on the trained classifier, probability maps were generated for the whole dataset and exported as tiff files in batch mode. Subsequently, CellProfiler [117] (version 3.1.8) was used to define cell masks for marker expression quantification. To define cell borders, nuclei were first identified as primary objects based on Ilastik probability maps and expanded through the cytoplasm compartment until either a neighboring cell or the background compartment was reached. Cell masks were generated for identification of single cells and used to extract single-cell information from IMC images.

Based on the IMC images and their segmentation masks, we used HistoCAT [118] (version 1.7.6) to extract single-cell marker profiles. Note that all the data were not transformed and used directly. In total, we extracted the cell intensities of CD38, MPO, CD14, CD71, CD11b, CD4, CD169, CD20, CD8a, CD15, CD3 and CD235a markers for 96232 cells. Please note that segmentation masks were identical for each comparison, using masks generated from DeepSNiF, so that the impact of variability in segmentation algorithms can be neglected. In Figures 4.17 and 4.18a, the comparison of the single cell profiles of raw, DIMR and DeepSNiF data show that DIMR has the potential to correct false positive data, and DeepSNiF corrects

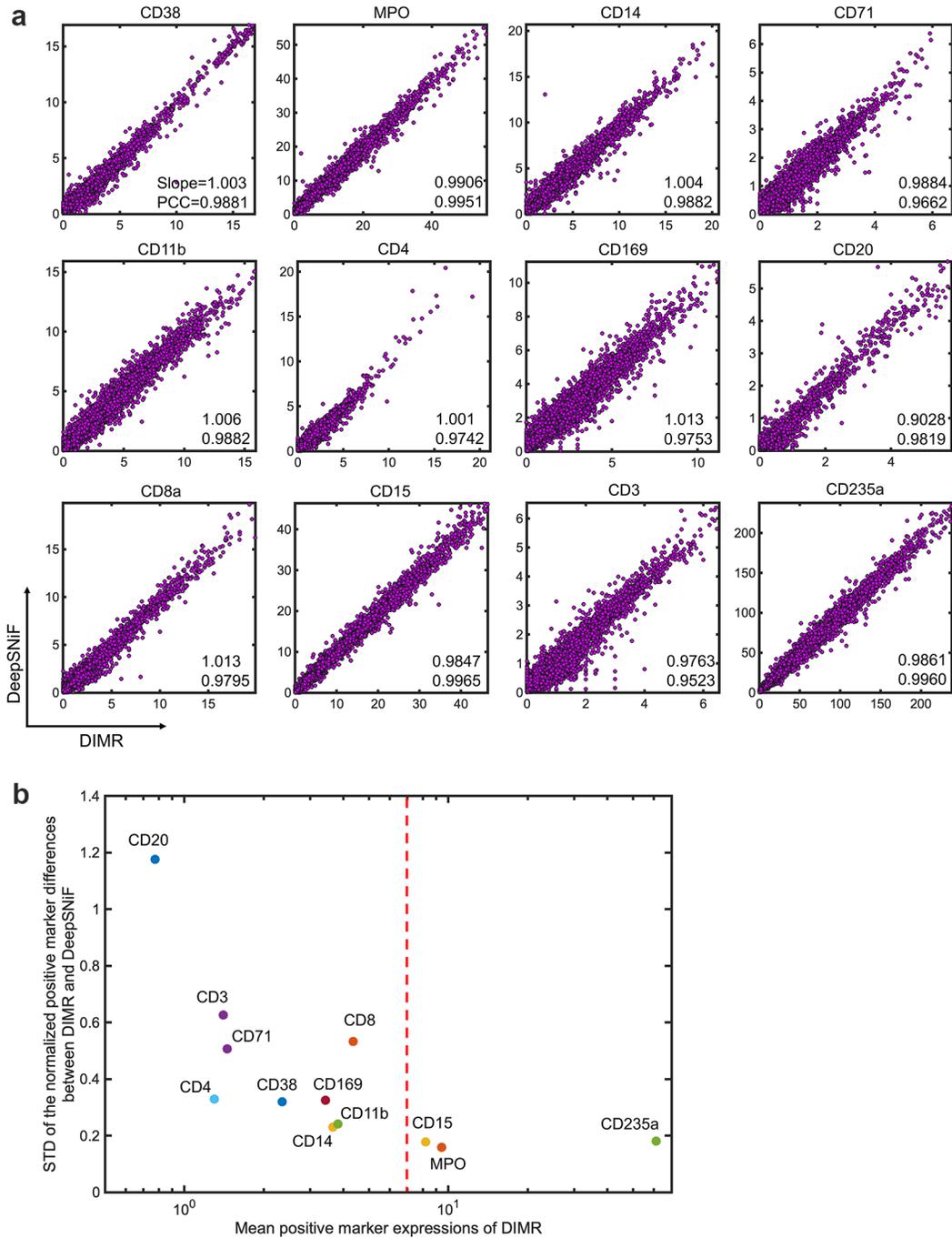


Figure 4.18: The impact of DeepSNiF on single cell data. (a) Each sub-figure represents the one-on-one relationship between the DIMR and DeepSNiF data of a particular marker in single cell scale. The bottom right value in each sub-figure represents the slope of the line fitting results and the PCC between the DIMR and DeepSNiF data. (b) The STD of the normalized positive marker differences between DIMR and DeepSNiF for DeepSNiF impact evaluation. The data was 99th-percentile normalized for extreme value mitigation.

all cell profiles. Meanwhile, we conducted bisquare line fitting for the extracted DIMR and DeepSNiF-processed single cell data with customized MATLAB (R2021a, MathWork) scripts, and calculated their PCC (Figure 4.18a). The results indicate DeepSNiF has not changed the single cell intensity scale nor biased the overall linearity of the data. Furthermore, larger mean positive marker expressions lead to lighter corrections by DeepSNiF (Figure 4.18b). This follows from the logic that larger ion counts have lower shot noise levels.

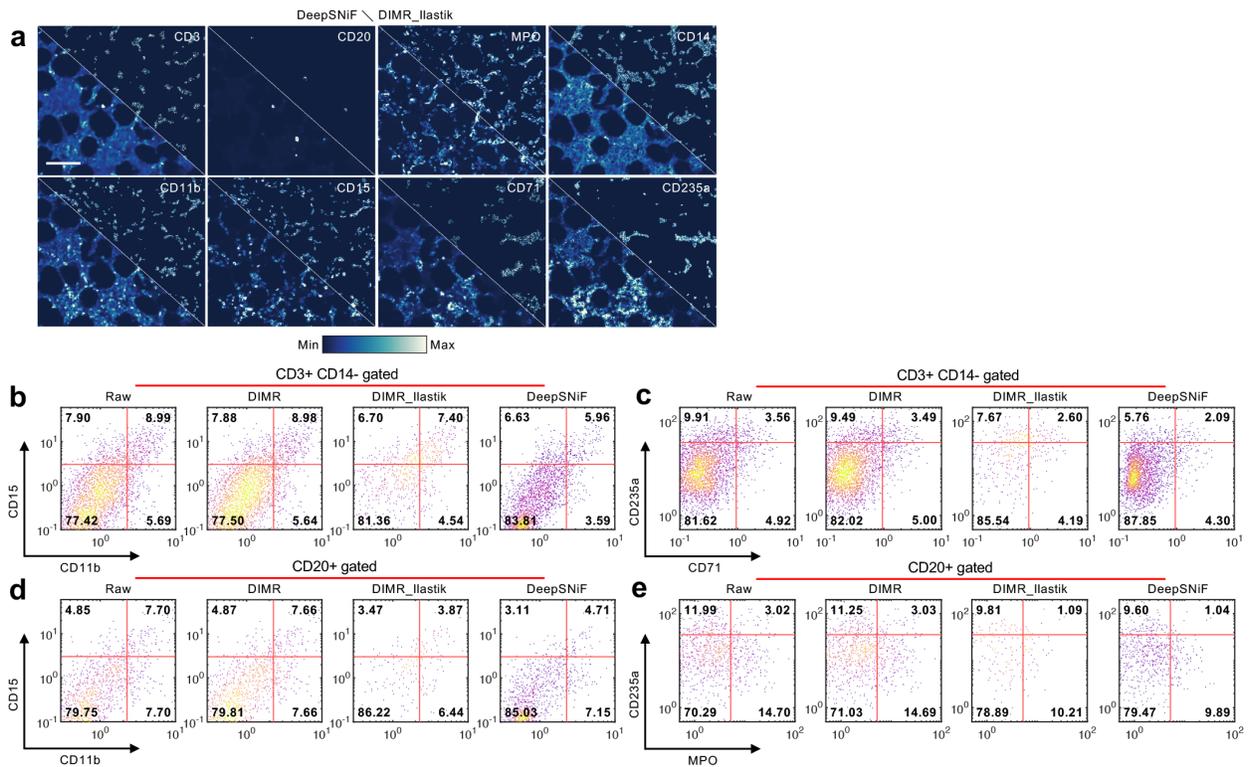


Figure 4.19: IMC-Denoise enhances downstream analysis of the human bone marrow IMC dataset. (a) Visual inspection of DeepSNiF and DIMR_IIastik-based denoising results on different markers-labeled IMC images. (b)–(e) Evaluations of denoising algorithms with manual gating strategies on single cell data. The numbers in these panels are the cell percentages of the corresponding ranges. DIMR slightly enhances the single cell analysis over raw data, while DeepSNiF further enhances the DIMR results and overall performs better than semi-automated DIMR_IIastik-processing. Scale bar: 107 μm .

Subsequently, we benchmarked the single cell data from the raw, DIMR, DIMR_Ilustik and DeepSNiF-processed images (Figure 4.19a). To achieve this, we first identified thresholds for positive cell markers. We modified the method described in [30]. Briefly, univariate Gaussian mixture models with scikit-learn [119] (version 1.0.2) were used to estimate the positive thresholds of each marker. Before threshold estimation, all data were 99th-percentile normalized so that the impact of extremely bright cells can be eliminated. For each channel, we performed model selection with models with 6 to 15 mixtures for DIMR data, in order to estimate the positive threshold accurately. We selected the model on the basis of the Davies-Bouldin index [120] and identified a positive threshold for a given channel by considering both the distributions of cell profiles and the overall IMC image intensities. The estimated positive thresholds of single cell data are summarized in Table A.11.

Manual gating approaches with prior knowledge of cell markers were applied to DIMR, DIMR_Ilustik and DeepSNiF on IMC data (Figures 4.19b–e). For example, among T cells (CD3-positive, CD14-negative), myeloid (CD11b, CD15) and erythroid (CD71, CD235a) markers should be absent. However, this condition may not hold because: segmentation and staining artifacts are unavoidable, and because hot pixels are present in the raw data. With the presence of shot noise, the single cell data could be further biased. In Figures 4.19b and c, the false positive of myeloid and erythroid markers decrease slightly after DIMR correction (0.8% and 0.4%). DIMR_Ilustik and DeepSNiF further removed false positive myeloid (3.86% and 6.31%) and erythroid markers (3.52% and 5.83%) after the slight improvement of DIMR. Similarly, among B cells (CD20-positive), myeloid and erythroid markers (CD11b, CD15, MPO and CD235a) should be absent as well. In Figures 4.19d and e, the false positive markers decrease slightly after DIMR correction (0.06% and 0.74%). Compared to DIMR, DIMR_Ilustik and DeepSNiF removed more false positive markers (DIMR_Ilustik: 6.41% and 7.86%, DeepSNiF: 5.22% and 8.44%). Overall, as expected DIMR could enhance

the single cell analysis to a limited extent. DeepSNiF and DIMR_IIlastik enable further enhancement, and overall the former achieves better performance than the latter on this task.

4.6.4 DeepSNiF in IMC-Denoise enhances automated cell phenotyping

Cell phenotype annotation plays a key role in tissue microenvironment analysis. Indeed, false annotation of cell phenotypes has the potential to lead to false biological or clinical conclusions. Hot pixel removal is normally conducted before automated cell phenotyping [30, 31, 55]. Therefore, we focused on whether DeepSNiF in IMC-Denoise could impact phenotypic annotation of cell types.

Here, the extracted single cell data with DeepSNiF-based segmentation masks from the human bone marrow dataset were used for phenotypic annotation, including CD38, MPO, CD14, CD71, CD11b, CD4, CD169, CD20, CD8a, CD15, CD3 and CD235a channels. Before analysis, data were 99th-percentile normalized followed by Z-score normalization. Then the DIMR data was clustered by the Phenograph algorithm with 20 nearest neighbours of each cell [121] with the Leiden community detection algorithm [122] with resolution of 6.0, which resulted in over clustering with 117 clusters. The generated clusters were manually labelled with a broad ontology and the channels that were most abundant in each cluster, resulting in 9 cell types, including immune cell subsets (B cell, CD4+ T cell, CD8+ T cell and plasma cell, monocyte/macrophages), erythroid, myeloid, and other CD4+ cells and others. To better demonstrate the modifications of DeepSNiF denoising, the DeepSNiF data clustering and annotation utilized a weighted K-nearest neighbour (KNN) approach ($K=20$) to map the DeepSNiF data into the DIMR clusters. It first constructs a Jaccard graph

between each cell from DeepSNiF and all the DIMR cells, and then maps the DeepSNiF data into the DIMR clusters with the shortest weighted distance. The Phenograph with Leiden algorithms were implemented by the software packages from <https://github.com/jacoblevine/PhenoGraph> and <https://github.com/vtraag/leidenalg>. DeepSNiF data annotation was implemented with customized python scripts.

Notably, multiple strategies were applied to reduce the noise impact during DIMR clustering: (1) Z-score normalization is consistent for handling different sources of noise in multiplexed cell data, including low intensity signal, high background signal, segmentation noise, and imaging artifacts, as verified by [123]; (2) the Jaccard graph construction in Phenograph is robust to noise, which is verified in [121]; and (3) over-clustering could improve the clustering accuracy [123]. Besides, we didn't annotate the DeepSNiF data with the same approach of DIMR because (1) The community detection results by Leiden algorithm is random so that it is very difficult to compare the annotations from different data; and (2) the weighted KNN method for DeepSNiF clustering could clearly reveal the differences before and after the processing.

For visualization, high-dimensional single-cell data of DIMR and DeepSNiF were reduced to two dimensions using the nonlinear dimensionality reduction algorithm fast interpolation-based *t*-SNE [124]. This algorithm was implemented by the software package in <https://github.com/KlugerLab/Flt-SNE>. Before the analysis, data were 99th-percentile normalized followed by Z-score normalization. The *t*-SNE parameters with perplexity of 50 and theta of 0.5 were used. The random seeds for the individual runs were recorded. The assigned phenotypes of DIMR and DeepSNiF datasets are demonstrated in Figure 4.20a and the relative changes of each cell sub-population after DeepSNiF processing is shown in Figure 4.20b. After DeepSNiF processing, B cells, CD8 T cells, plasma cells, CD4 T cells and other CD4+ cells decrease (20.86%, 3.70%, 13.44%, 18.49% and 8.23%, respectively), the

monocytes/macrophages increase (4.82%), while erythroid, myeloid and other cells remain largely unchanged.

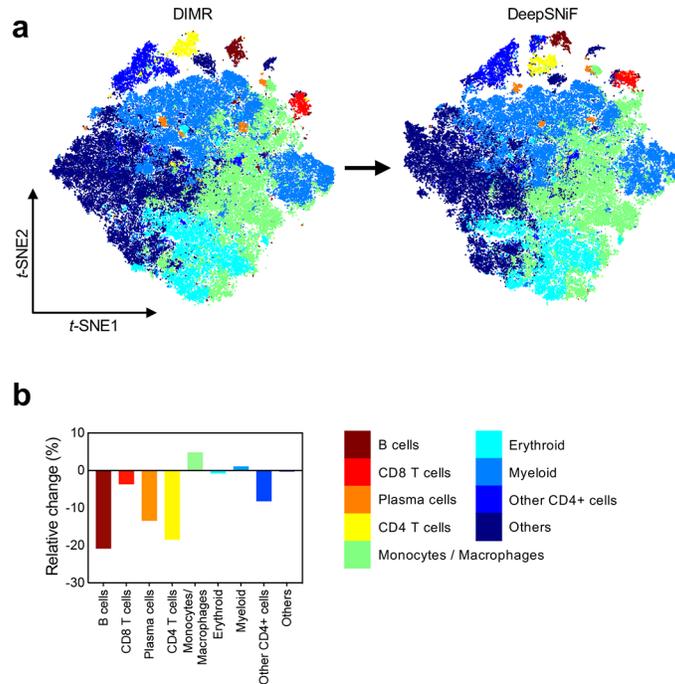


Figure 4.20: *t*-SNE plots and relative change of DIMR and DeepSNiF data. (a) *t*-SNE plots of DIMR and DeepSNiF with cell phenotyping results. (b) The relative change in cell phenotypes before and after DeepSNiF.

The phenotyping results of DIMR and DeepSNiF were mapped back into their segmentation masks and images (Figure 4.21). To highlight cells where DeepSNiF changes the cell phenotyping results, conflicting annotations between DIMR and DeepSNiF were labeled with white contours, and the changes were quantified for cell phenotype and marker enrichments. To evaluate the effect of positive cell marker enrichment after DeepSNiF, the cell-type annotations before and after the processing were selected, and the percentage of positive markers on each cell types was calculated. The relative change was then defined as the difference between the percentage of positive markers after and before the processing. To evaluate the

effect of cell type enrichment after DeepSNiF, the positive cells for a given marker before and after the processing were selected, and the percentage of each cell type based on cell-type annotation was calculated. The relative change was then defined as the difference between the percentage of cell-type composition after and before the processing.

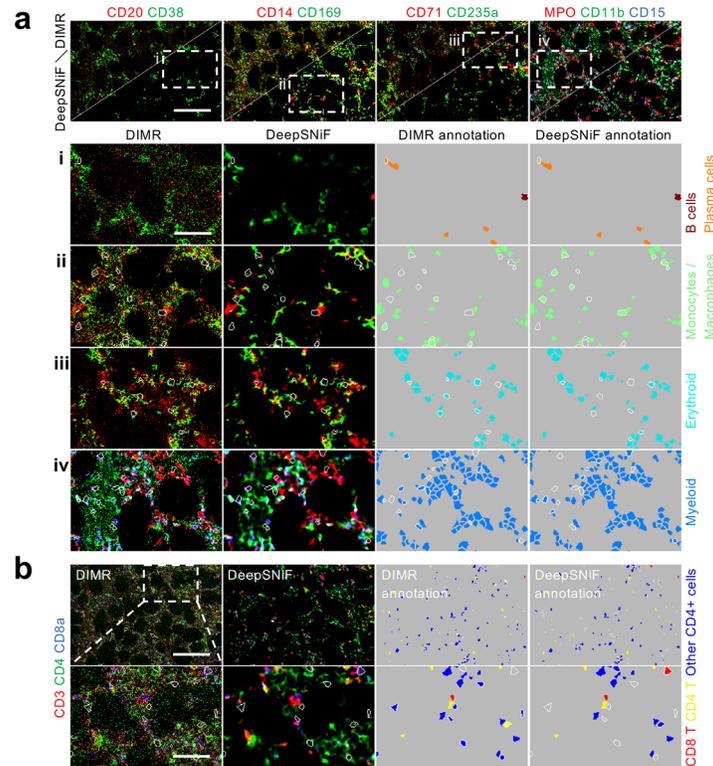


Figure 4.21: Comparisons of DIMR and DeepSNiF-processed IMC images labeled with different cell markers, and the corresponding cell annotation results. The sub-panels (i)–(iv) in (a) and the bottom row in (b) correspond to the white dashed box region selection in their first panels, respectively. The white contours represent the differential phenotyping results between DIMR and DeepSNiF. Scale bar: (a) 110 μm . (b) Top: 145 μm , bottom: 50 μm .

After DeepSNiF denoising, non-specific markers are reduced, while specific markers are enriched within the cell phenotypes (Figure 4.22). For example, we observed the positive rate increased for CD20 in B cells (10.53%), CD8a in CD8 T cells (2.32%), CD3 and CD4 in CD4 T cells (6.84% and 4.64%), CD38 in plasma cells (6.21%) and CD4 in other CD4+ cells

(4.26%). Conversely, we observed a decrease of non-specific markers, such as CD38, MPO and CD14 in B cells (5.24%, 8.11% and 5.64%), CD3 in erythroid (1.79%) and myeloid (1.62%) cells, and all marker signals in “other” cells. Furthermore, the identified cell types were enriched in a marker-specific manner after DeepSNiF (Figure 4.22). For instance, we observed a post-DeepSNiF enrichment of monocytes/macrophages in CD14+ cells (2.36%), CD11b+ cells (1.70%) and CD169+ cells (2.21%), and enrichment of B cells in CD20+ cells (5.42%) and CD8 T cells in CD8a+ cells (3.15%). Similarly, myeloid cells were enriched in MPO+ (1.26%), and erythroid cells in CD71+, CD235a+ cells (2.45% and 1.48%). DeepSNiF also yielded an enriched composition of CD8 and CD4 T cells (3.50% and 2.54%), and reduced composition of myeloid and erythroid cells (2.47% and 4.27%) in CD3+ cells. However, we noticed the enrichment of erythroid cells in CD169+ cells (1.03%), which may result from an artifact of the current segmentation approach due to the close relationship and irregular morphology at the boundaries between erythroids and macrophages within the bone marrow [125]. Cell phenotyping by immunostaining of FFPE tissues is also inherently limited by antibody specificity and antigen retrieval protocols. In this tissue, CD38+ and CD14+ antibody staining is not strictly restricted to single lineages, and these markers can be aberrantly expressed in myeloid neoplasms included in this data set. On manual inspection, DeepSNiF improves the ability to identify co-localization of cell surface markers (Figure 4.21). Overall, DeepSNiF enhances the sensitivity and specificity of cell phenotyping.

We observed that the enhancements in cell phenotyping and marker enrichments in Figures 4.22 and 4.23 are related to the noise level of the IMC images. Specifically, DeepSNiF has the highest impact on CD20 and CD3 related phenotypes, improvement for CD15, MPO and CD235a related phenotypes is limited, with moderate changes for other cell classes. These findings agree with Figure 4.18b, where we plot the STD of the normalized positive marker differences between DIMR and DeepSNiF against intensity.

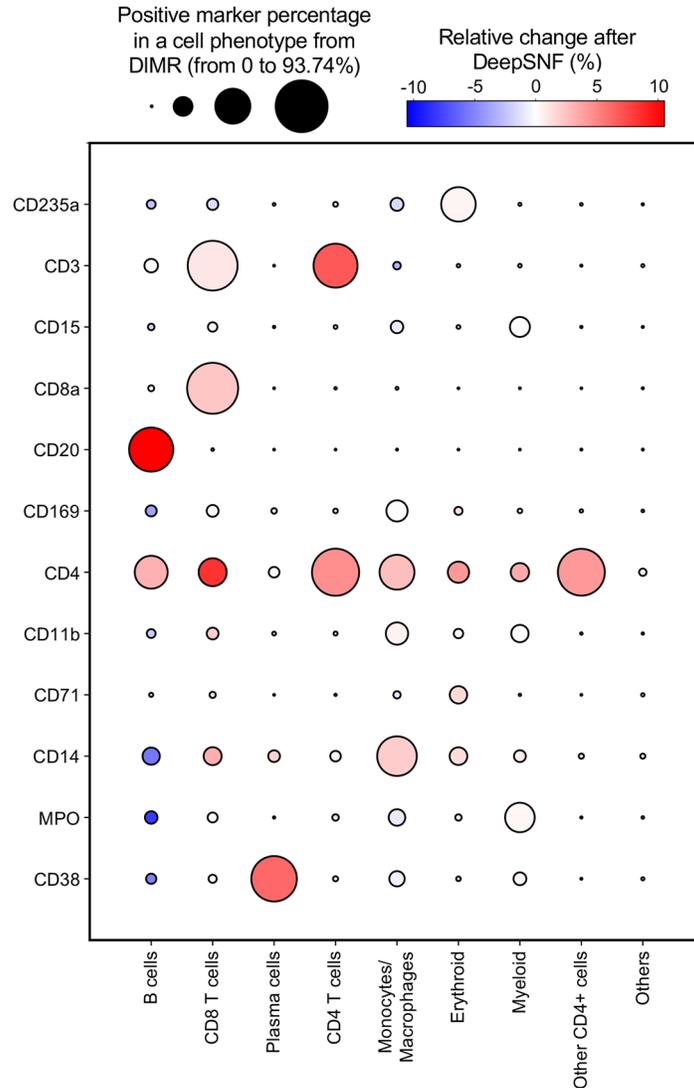


Figure 4.22: Deep-SNiF enhances the sensitivity of cell phenotyping. After DeepSNiF processing, the non-specific marker signals reduce while the specific ones enrich in the cell types, respectively. The circle size indicates the positive marker percentage in a particular phenotype of DIMR, and the circle colour indicates the relative changes of the positive rate for the particular markers after DeepSNiF enhancement.

4.6.5 DeepSNiF in IMC-Denoise enhances lymphocyte analysis

Cell-cell interactions of immune cells within the tumor microenvironment is of broad interest for many clinical pathology specimens. In myeloid malignancies, immune infiltrates

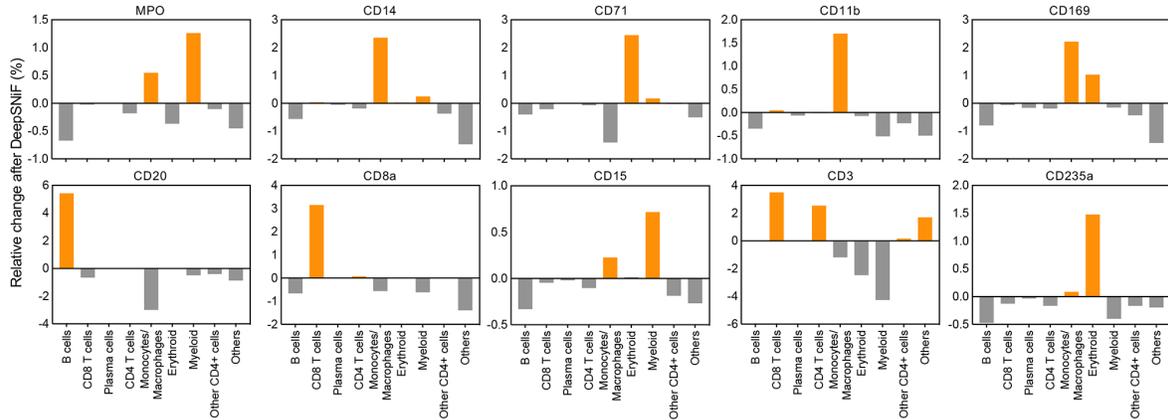


Figure 4.23: DeepSNiF enhances the specificity of cell phenotyping. With DeepSNiF denoising, the ratios of specific phenotypes increase while those of non-specific phenotypes decrease in the positive markers. The relative change is the difference in percentage composition of each cell type before and after DeepSNiF enhancement.

are most commonly assessed by flow cytometry and are an active area of interest in therapeutic clinical trials [126]. However, in situ spatial context of cell-cell interaction mediated immune responses cannot be directly measured through this approach. We quantified the enhancement of lymphocyte spatial analysis for B cells, CD8+ T cells and CD4+ T cells by DeepSNiF, and compared these to a manually curated set of image annotations based on DeepSNiF-based cell masks (Figure 4.24a). The manual cell-type annotation was based on the DeepSNiF-based cell segmentation masks. Briefly, DIMR and DeepSNiF images were overlaid with the cell masks in FIJI [82]. In some extremely noisy cases, the DIMR images were denoised by Gaussian filters to improve the annotation accuracy. Based on the signal in each cell mask, the cells were classified as B, CD8 T, CD4 T cells and other cells. Some positive signals were identified as hot clusters and discarded. The annotation results were manually recorded. CD3, CD4 and CD20-stained images are more easily contaminated by shot noise than others. Therefore, this approach can further validate the shot noise accounting ability of DeepSNiF as well. The phenotyping accuracy of DIMR and DeepSNiF as evaluated by the Jaccard score and F1 score indicate a significant improvement by

DeepSNiF denoising (Figure 4.24b); and DeepSNiF denoised data closely recapitulates gold-standard but laborious manual annotation. Specifically, the overall Jaccard scores improve from 0.6785, 0.8229 and 0.6781 to 0.9201, 0.8922 and 0.8860 for B cells, CD8+ T cells and CD4+ T cells, respectively. Similarly, the F1 scores improve from 0.8085, 0.9029 and 0.8082 to 0.9584, 0.9430 and 0.9396 for these cell types, respectively.

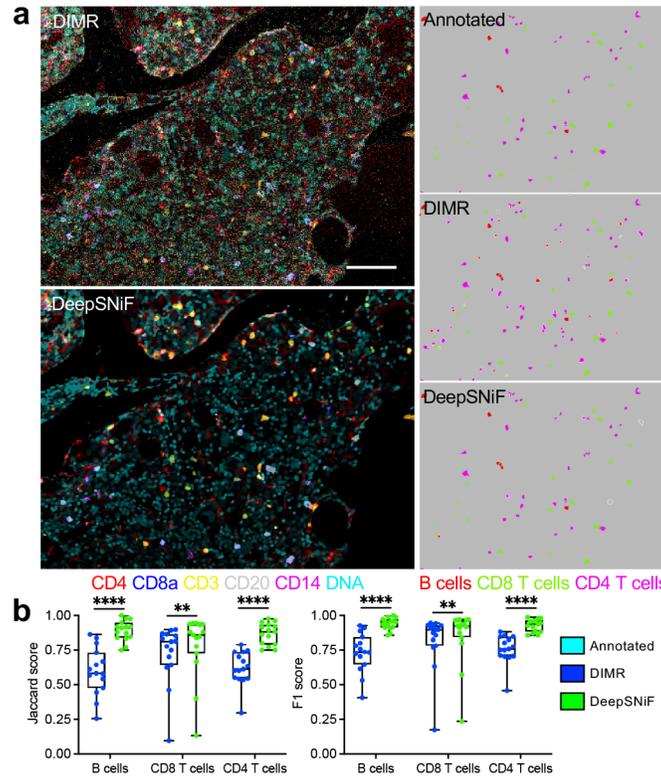


Figure 4.24: DeepSNiF enhances lymphocyte annotation accuracy. (a) Manual annotations for lymphocytes and comparisons with DIMR and DeepSNiF phenotyping results with DeepSNiF-based cell masks. The white contours represent the differential phenotyping results between the annotated and DIMR/DeepSNiF results. (b) Annotation evaluations of DIMR and DeepSNiF by both Jaccard and F1 scores. Scale bar: 85 μm .

Subsequently, the tissues were classified as normal morphology (Normal), myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML). The improvements in visual quality afforded by DeepSNiF denoising facilitated manual review of lymphocyte staining patterns

for annotation of lymphocyte subpopulations (Figure 4.25). B and T cell populations are scattered throughout the bone marrow cellularity in normal and malignant specimens, and lymphoid aggregates are occasionally present. To characterize the density and distance relationships between lymphocyte subpopulations, samples were analyzed in cohorts of extent of malignant blast involvement, after exclusion of the lymphoid aggregate outlier. Nearest neighbor distances between B cells, CD4 T and B cells, and CD4 and CD8 T cells were calculated for different disease tissues (Figure 4.26a). Overall, the distributions from DeepSNiF are more concordant with annotated data. By contrast, those from DIMR are biased, with significant differences to the annotations due to cell misclassifications.

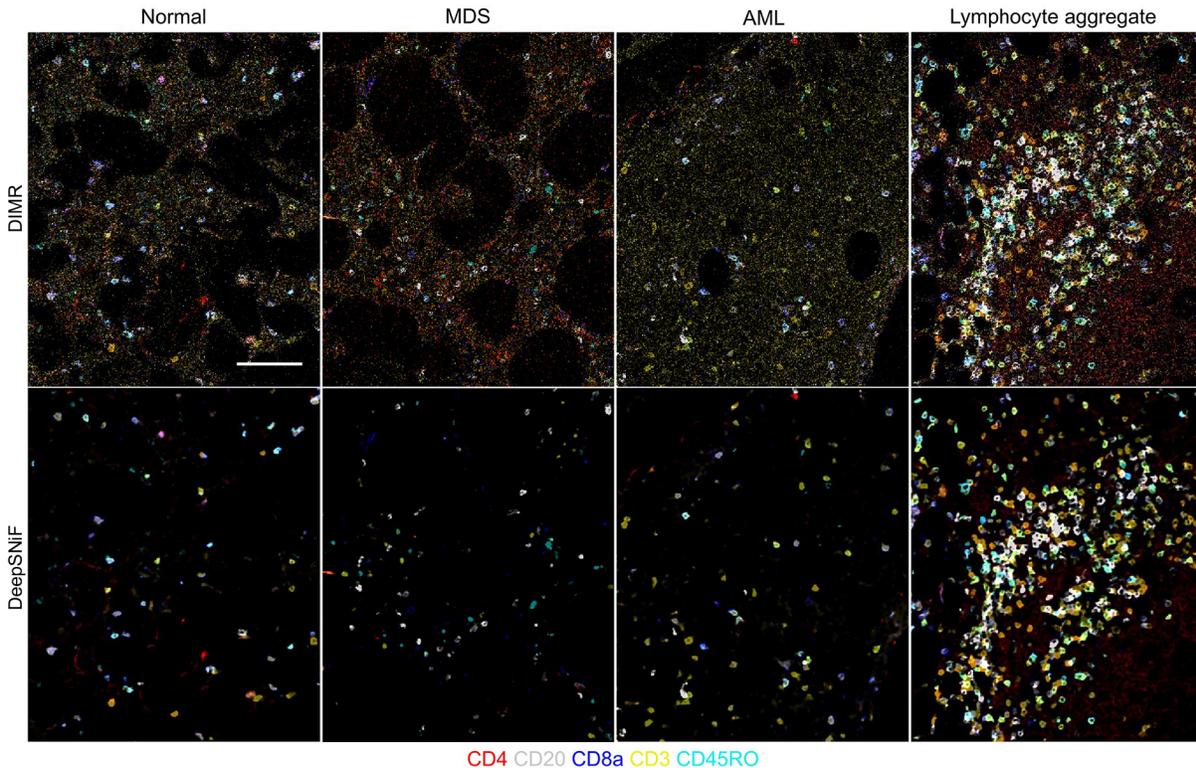


Figure 4.25: Representative images of lymphocyte markers after DeepSNiF denoising from specimens of normal (upper left), MDS (upper right), AML (lower left) and AML with lymphoid aggregate (lower right) tissue samples. Scale bar: 112 μm .

Automated DeepSNiF denoising reveals that as disease develops, the B cell distances become larger ($P < 0.01$); The distances between CD4 T and B cells in normal and MDS tissues are greater than those of AML ($P < 0.0001$); And CD4 and CD8 T cells in normal tissues trend towards longer distances than those in MDS ($P = 0.0916$). Interestingly, the overall distances between CD4 T and B cells in MDS tissues are greater than those of Normal samples ($P < 0.01$). These findings hold for DeepSNiF denoised data in distances between B cells from MDS to AML ($P < 0.05$), and from Normal to MDS samples for CD4 and CD8 T cells ($P = 0.0929$) and CD4 T to B cells ($P < 0.05$). However in non-DeepSNiF denoised DIMR data, the trends between B cells has been violated from MDS to AML by DIMR data ($P = 0.4923$), as well as those from Normal to MDS between CD4 and CD8 T cells ($P = 0.4762$), and CD4 T to B cells ($P = 0.6685$). From this point, DeepSNiF is able to correct the distorted cell spatial distributions from less accurate annotations caused by noise, which may further enhance downstream cell-specific spatial analyses.

We also calculated the cell densities per tissue of these lymphocytes (Figure 4.26b). The DeepSNiF results are closer to those annotated data for B cell and CD4 T cell. By contrast, the CD8 T cell densities from DIMR, DeepSNiF and annotated data are close to each other. Additionally, we observe higher B cell and CD4 T cell densities in Normal tissues than others, and higher B cell density of MDS than that of AML. No obvious developing trend for CD8 T cells are observed as the disease status changes. Furthermore, we have analyzed the correlations between the densities from different cell types (Figures 4.26c and d). Note that in the reference groups, the B and CD4 T cell densities are generated from the annotated data, while the monocyte/macrophage density comes from DeepSNiF data. This is because the relative change of the monocytes/macrophages by DeepSNiF is smaller compared to those of B and CD4 T cells (Figure 4.20b) and because DeepSNiF achieves higher accuracy than DIMR for the cell phenotyping (Figures 4.22 and 4.23). From the

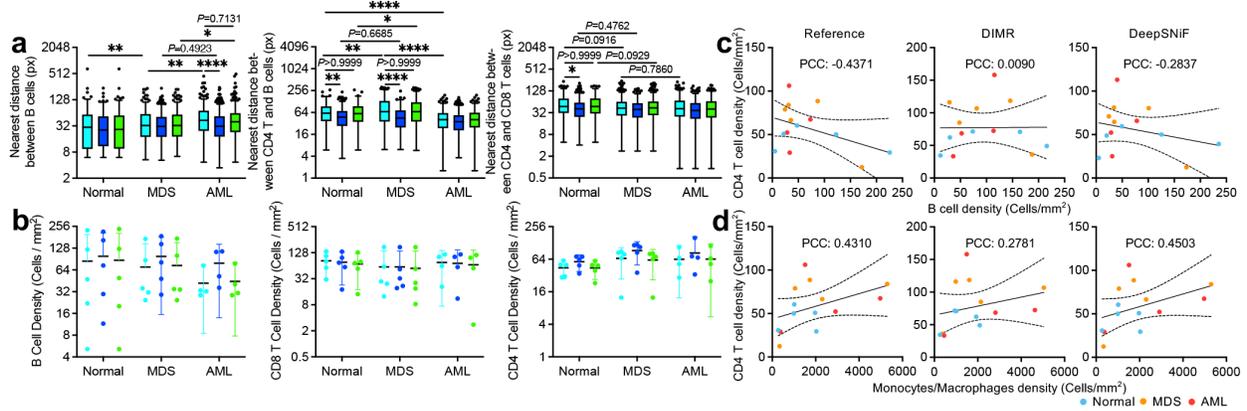


Figure 4.26: DeepSNiF enhances lymphocyte analysis. (a) Nearest distance comparisons between different cell types of normal, MDS and AML tissues from manual, DIMR and DeepSNiF phentyping results. The box plots are shown as the tukey format. (b) Cell densities comparisons of normal, MDS and AML tissues from manual, DIMR and DeepSNiF phentyping results. Single dots and the means with 95% confidence interval are shown in these figures. (c) and (d): Correlation analysis between CD4 T cell and B cell, monocyte/macrophage densities per tissue from manual, DIMR and DeepSNiF phentyping results. The data from the reference group in (c) comes from annotated data; while that from the reference in (d) comes from annotated (CD4 T cells) and DeepSNiF (monocytes/macrophages) results, separately. The dashed lines represent 95% confidence interval.

reference group in Figure 4.26c, the densities of CD4 T and B cells are negatively correlated with each other (PCC: -0.4371). Nevertheless, DIMR result indicates no correlation between the cell densities (PCC: 0.0090), which demonstrates false annotations hinder true relational definition between different cell types. Again, the negatively correlated relationship can be uncovered using automated DeepSNiF (PCC: -0.2837). Likewise, the DIMR data fails to detect the correlation between the densities of CD4 T cell and monocyte/macrophages (PCC: 0.2781) (Figure 4.26d). Corrected by DeepSNiF, the measured correlation (PCC: 0.4503) approximates the reference finding (PCC: 0.4310).

4.7 Discussion

With the rise of novel multiplexed technologies for the characterization of cellular context in health and disease, IMC has emerged as a valuable tool to investigate immunophenotypes while preserving spatial information [5, 6, 7, 27, 28, 29, 30, 31, 32]. Differing from traditional multiplexed imaging approaches based upon fluorescence microscopy, IMC allows for simultaneous acquisition of more than 40 cell-specific markers with greatly suppressed channel crosstalk, and avoids tissue and marker degradation in multi-round staining protocols. Furthermore, it eliminates autofluorescence and background signal issues that are inherent in fluorescent microscopy. The high-dimensional datasets then enable complex microenvironment analysis. However, IMC suffers from unique hot pixel and shot noise features. Analyzing raw IMC data without further restoration may lead to distortions, even errors, in downstream analysis. Contemporary denoising strategies [27, 30, 31, 57, 59] are usually not adaptive or effective for these particular noise conditions. For example, the parameters of some methods must be tuned manually, which is not suitable for large datasets and may cause subjective, batch, and channel-specific errors.

In this work, we propose IMC-Denoise to account for the specific technical noise present in IMC images. In this pipeline, the DIMR algorithm is first applied to adaptively remove hot pixels. It does not use a user-defined intensity threshold or range to define hot pixels, eliminating the impact of the density and intensity variations of hot pixels in different datasets or markers. Instead, it builds a histogram from the differential maps of raw images followed by an iterative outlier detection algorithm. In comparison with other methods, DIMR achieves more robust hot pixel detection capability and normal pixel preserving performance. After hot pixel removal, the DeepSNiF algorithm is proposed to restore image quality. I-divergence is derived as the optimal loss function for this denoising task. Due to the absence of noise-free

IMC images and incompatibility with repeated scanning to generate training labels [5], we applied a masking strategy with stratified sampling from N2V. This enabled self-supervised training for this denoising task, in which multiple pixels are randomly masked and replaced by its adjacent pixels. With the continuity of antibody signals in IMC, Hessian norm regularization [34, 52, 107] is added in the loss function to boost the denoising performance. In DeepSNiF, we train a single network for a single marker, which reduces the memory allocated for training. Nevertheless, we note that DeepSNiF also works on multi-marker training (Figure 4.27).

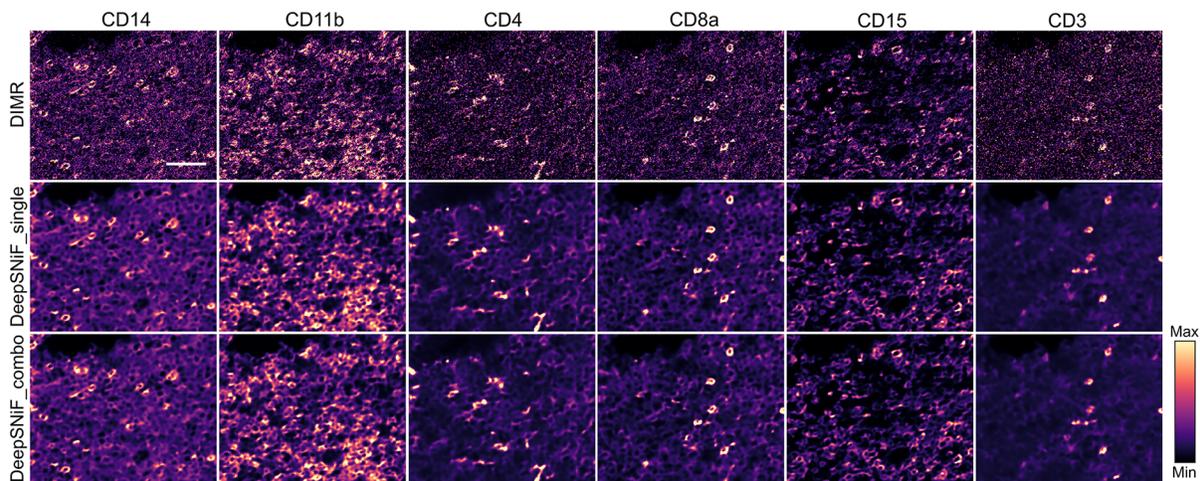


Figure 4.27: DeepSNiF works on multiple markers training. The DIMR-processed IMC images were trained by DeepSNiF with single marker in each network (DeepSNiF_single) and all the markers in a single network (DeepSNiF_combo), respectively. The denoising results indicate both approaches enables IMC image quality improvement. Scale bar: 48 μm .

In another aspect, this demonstrates that DeepSNiF works on the markers stained for morphologically heterogeneous markers, since the variant features have been learned in the training process. In addition, monocytes/macrophages are morphologically heterogeneous so that the successful denoising of CD14/CD169 (Figure 4.26a) validates the adaptability of DeepSNiF as well. In fact, the networks are able to learn all the features existing in the training

images but not focus on any specific structures. As a result, markers with interstitial staining patterns (e.g. vessels, fibrosis, reticular cytoplasmic projections) can be well restored (CD31, CD34 and Collagen III in Figures 4.12 and 4.16). However, small areas of staining at the size of a sub-cellular synapse (e.g. 1–2 μm diameter) will not be successfully distinguished by IMC due to its relatively low resolution of 1 μm . Therefore, the network cannot learn the features of such small structures. The trained network can be employed to other datasets which share similar features (Figure 4.13).

To determine the applicability of our approach, reference denoising algorithms were utilized to rigorously evaluate IMC-Denoise on both simulated data and multiple pathological patient datasets. Compared to other methodologies, both DIMR and DeepSNiF achieve the best denoising performance, qualitatively and quantitatively. Orthogonal approaches that have not been previously tested in evaluation of IMC restoration are also used to verify the image quality improvement by IMC-Denoise. This pipeline can be further extended by existing analytical processing pipelines including Mesmer/DeepCell and ark-analysis [97] or MCMicro [98]. If warranted, one may [56] address spillover issues after hot pixel removal and shot noise filtering, as indicated in Equation (4.26). A related modality, MIBI [25, 26], shares several image formation and noise features with IMC, and the denoising pipeline deployed here may also enhance MIBI datasets.

IMC-Denoise is effective at removing background noise and enhancing downstream analysis of IMC data with limited, subjective, user-input. Multiple datasets processed by DIMR and DeepSNiF were compared with the state-of-the-art IMC background removal methods, including single threshold binarization, semi-automated Ilastik-based [59], and MAUI [57], using the F1 score as the accuracy metric to evaluate the results. The qualitative and quantitative results indicate DeepSNiF can affect significant background noise removal, and is superior to tedious semi-automated approaches. In particular, DeepSNiF is capable of

unmixing specific IMC staining signal from background noise. This means that even the thresholding approach for background removal is not essential after DeepSNiF denoising.

Conventional workflows typically use manual gating strategies combined with prior cell marker knowledge to identify and compare cell types in pathological samples. We used real world data and these methods to evaluate the IMC denoising algorithm for single cell analyses, and compared to DIMR, DIMR_Ilstik and DeepSNiF. Automated IMC-Denoise performs equally or superior to the semi-manual Ilstik-based method in downstream single cell analysis, and DeepSNiF notably enhances cell clustering and annotation. Quantitative evaluations of cell phenotyping results indicate the improvement of sensitivity and specificity by DeepSNiF denoising. Further validations with DIMR-based cell masks demonstrate the robustness of IMC-Denoise to variant cell segmentation results. For lymphocyte annotation, Jaccard and F1 scores demonstrate that DeepSNiF performs significantly better than DIMR on phenotyping of B, CD8-positive T and CD4-positive T cells. Further, spatial distribution and cell density correlation analysis indicate less accurate annotations by the data denoised solely by DIMR, leading to biased conclusions. With the data denoised by DeepSNiF, such distortions can be corrected and more accurate downstream analysis is achieved.

As noted, DeepSNiF enhances all the markers and their downstream analysis. However, the marker channels with high noise levels benefit to a larger degree. In theory, there is no maximum noise level present for denoising algorithms. Even under some extremely noisy conditions (CD20 and CD3 in Figure 4.12a), DeepSNiF improves the image quality. Nevertheless, lower SNR in raw images means lower specific information content and thus the quality of the restored images are lower. Because of the signal-dependent characteristics of shot noise, the noise components of high SNR channels contribute less to overall image quality, and thus have lower impact on downstream analysis. Empirically, we find that denoising by DeepSNiF can be omitted when the mean expressions of positive markers are

larger than 7 (MPO, CD15 and CD235a), however denoising all marker channels improves performance and is not computationally intensive.

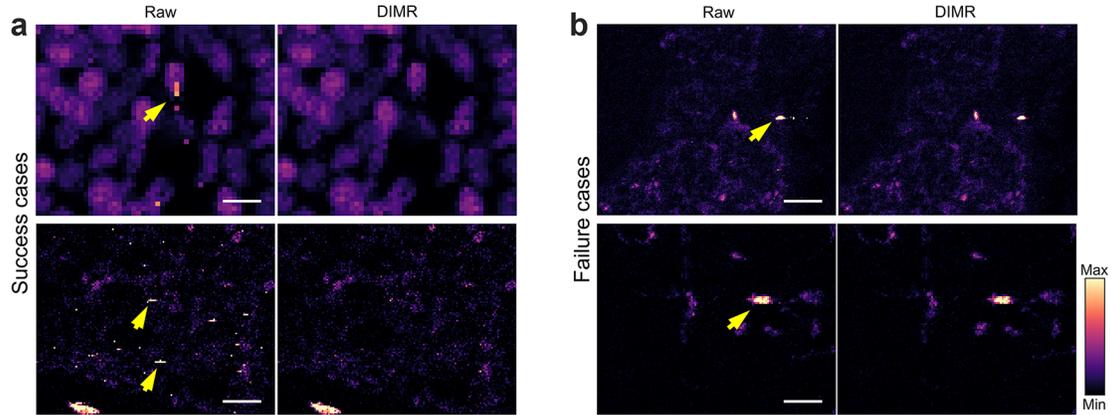


Figure 4.28: The limitation of the DIMR algorithm. (a) Success cases of DIMR on challenging hot pixels. (b) Failure cases of DIMR on hot clusters. DIMR is able to remove line-style consecutive hot pixels while fails on hot clusters. Scale bar: (a) Top: $8 \mu\text{m}$, bottom: $24 \mu\text{m}$. (b) Top: $40 \mu\text{m}$, bottom: $20 \mu\text{m}$.

Limitations of IMC-Denoise include the inability to remove large hot pixel clusters, as DIMR cannot discriminate these larger areas of outliers from signal (Figure 4.28). Further the self-supervised DeepSNiF algorithm cannot reach the accuracy of supervised denoising methods due to unavailability of ground truths (Figure 4.7). Nevertheless, DIMR can remove single hot pixels and small hot clusters of several consecutive pixels, and DeepSNiF performs better than other unsupervised and self-supervised denoising methods on IMC datasets.

4.8 Conclusion

To conclude, we have developed the content aware IMC-Denoise for improved IMC image quality and quantitative accuracy. Predicated on a novel combination of differential map-based and self-supervised deep learning-based algorithms, this open source pipeline removes

hot pixels and effectively suppresses shot noise in multiplexed IMC data. Multiple image and cell-based analyses from different IMC datasets verified the enhancements brought by this approach, with the ability to resolve significant cellular phenotypic and spatial information approximating manual annotation. We have provided tutorials to help users implement IMC-Denoise (https://github.com/PENGLU-WashU/IMC_Denoise). We expect IMC-Denoise to become a widely used pipeline in IMC analysis due to its adaptability, effectiveness and flexibility.

Chapter 5

InterSTELLAR: An Interpretable Spatial Cell Learning Framework to Enhance the Characterization of Highly Multiplexed Imaging Data

5.1 Introduction

In this section, we present InterSTELLAR, a geometric deep learning framework for multiplexed imaging data, to link the outcomes of tissue and microenvironments directly without downstream processing algorithms. By employing weakly supervised learning methods based on tissue-scale labels, InterSTELLAR is designed to simultaneously predict tissue outcomes and detect disease relevant microenvironments. We apply InterSTELLAR to an open-source

Content of this chapter is extracted/adapted from the author's preprint manuscript [36] and published conference article [37].

breast cancer IMC dataset [7] and show that it can accurately characterize patient tissue clinical subtypes. Moreover, by utilizing identified cell communities with high diagnostic value, InterSTELLAR can benefit microenvironment exploration and correlative patient outcomes. We demonstrate the InterSTELLAR workflow using the breast cancer IMC dataset [7], but the method can be easily modified to analyze any other types of highly multiplexed imaging data, such as CODEX, *t*-CyCIF and MIBI.

5.2 InterSTELLAR Framework

5.2.1 Dataset description and pre-processing

Here, we focus on a key feature set in breast cancer pathology assessment, including hormone status and growth factor receptor expression, to investigate these cell and cell organizations in aggressive triple negative breast cancer (TNBC). We applied the InterSTELLAR framework on an open-source breast cancer IMC dataset [7]. The dataset consists of 381 tissues with 35 cell markers, as well as segmentation masks, single cell data, cell phenotypes, tumor-stroma masks, tissue clinical subtype and patient survival information. To focus on the effects of antibody markers and to increase robustness, DNA markers and tissues with too few cells (less than 50) were removed, leaving 366 tissues with 30 markers (Table 5.1). Specifically, there are 83 healthy, 49 TNBC and 234 non-TNBC cancerous tissues.

For a specific cell marker c , denote the single cell data $x_i^{(k)}$ as the mean expression value of each cell i , and the array of all expression values $\{s_1^{(c)}, s_2^{(c)}, \dots\}$ as $\mathbf{S}^{(c)}$. We then normalized the single cell data using log transformation:

$$f(s_i^{(c)}) = \log(s_i^{(c)} + \epsilon), \quad (5.1)$$

Table 5.1: Cell markers used in training and downstream analysis.

Metal Tag	Target	Metal Tag	Target
141Pr	Cytokeratin5	159Tb	p53
142Nd	Fibronectin	160Gd	CD44
143Nd	Cytokeratin19	162Dy	CD45
144Nd	Cytokeratin8/18	163Dy	GATA3
145Nd	Twist	164Dy	CD20
146Nd	CD68	166Er	Carbonic Anhydrase IX
147Sm	Cytokeratin14	167Er	E/P-Cadherin
148Nd	SMA	168Er	Ki-67
149Sm	Vimentin	169Tm	EGFR
150Nd	c-Myc	170Er	S6
151Eu	c-erbB-2-Her2	172Yb	vWF
152Sm	CD3	173Yb	mTOR
155Gd	Slug	174Yb	Cytokeratin7
156Gd	Rabbit IgG H L	175Lu	panCytokeratin
158Gd	Progesterone Receptor A/B	176Yb	cleaved PARP

where ϵ represents a small value. Here, we set it as 10^{-4} . Next, we calculated the Z scores of the normalized expression values:

$$z(s_i^{(c)}) = \frac{f(s_i^{(c)}) - \text{MEAN}(f(\mathbf{S}^c))}{\text{STD}(f(\mathbf{S}^c))}. \quad (5.2)$$

5.2.2 Graph construction

Through the modeling of cell spatial interactions of patient tissue specimens, InterSTELLAR aims to predict clinically relevant tissue subtypes and corresponding cell communities. As shown in Figure 5.1a, InterSTELLAR is built upon undirected graphs. To construct graphs from tissues, the set of cells are represented by a set of discrete points located at cellular

centroids. The 2D coordinates of these cellular centroids (x, y) are determined by the segmentation masks of the corresponding cells. Then, we regard each tissue as a single graph, in which each cell is a node of the graph. The node features are the matrix of marker expressions $\mathbf{Z}^{N \times F}$. The edge between any two nodes determines whether the nodes are connected. The Euclidean distance between any two nodes u and v is calculated as

$$d(u, v) = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2}. \quad (5.3)$$

Considering any two cells in a tissue, the longer their distance, the less their inter-communications will be. We define the weight of each edge (u, v) , which is negatively associated with their distance $d(u, v)$:

$$w(u, v) = \begin{cases} \exp(-\frac{d(u,v)^2}{l}) & \text{when } d(u, v) < T \\ 0 & \text{else} \end{cases}. \quad (5.4)$$

In Equation (5.4), the hyperparameter l and T determines how rapidly the weight decays as a function of distance. Here, we set T as $40 \mu m$. This is approximately twice the size of a regular cell as it assumes that cells have to be within reach to each other to interact [63]. l is empirically set as $\frac{1600}{\log 40}$ so that $w(u, v)$ approaches $\frac{1}{40}$ when $d(u, v)$ approaches the threshold. Therefore, the graph adjacency matrix $\mathbf{A}^{N \times N}$ with a shape of $N \times N$ is built, in which $w(u, v)$ is the matrix element. Finally, there are a node feature matrix $\mathbf{Z}^{N \times F}$, a graph adjacency matrix $\mathbf{A}^{N \times N}$ and a clinical type label $Y \in \{\text{Healthy, TNBC, Non-TNBC cancers}\}$ per graph as inputs to InterSTELLAR.

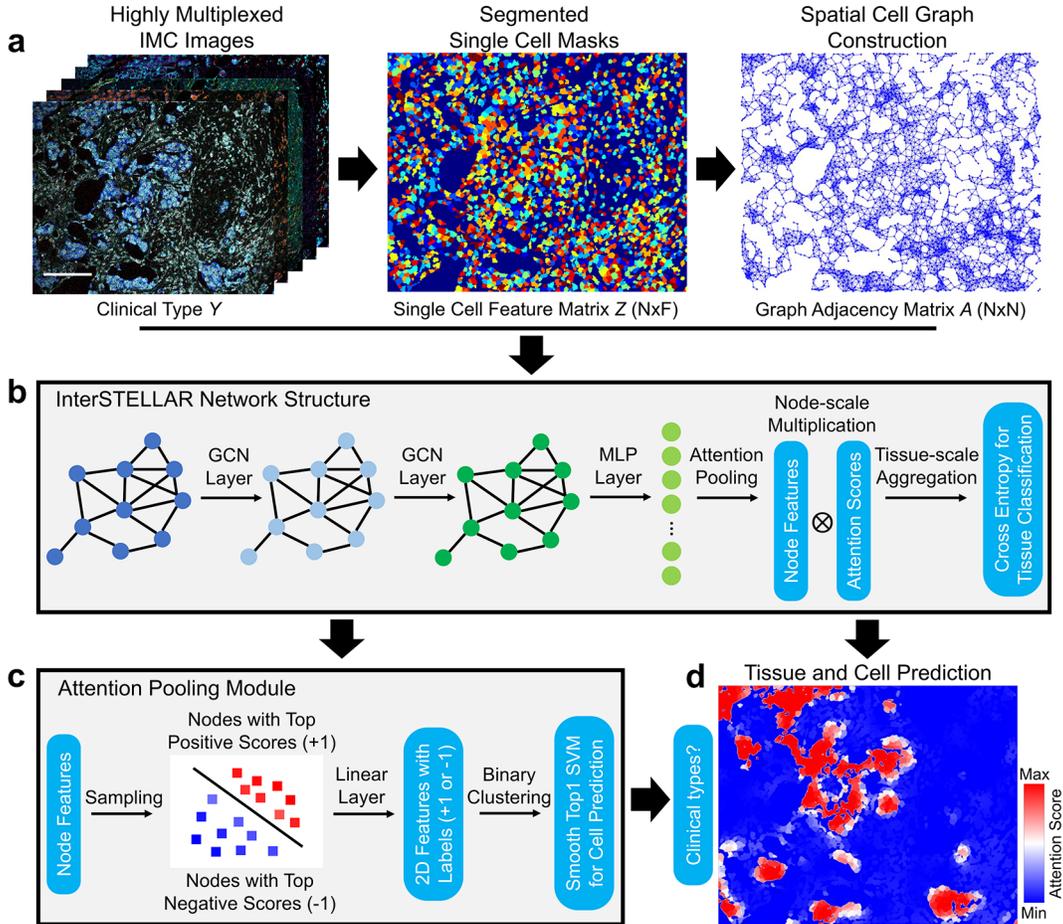


Figure 5.1: Overview of InterSTELLAR. (a) Undirected graph construction from highly multiplexed images. (b) Single cell feature matrix \mathbf{Z} , adjacent matrix \mathbf{A} and tissue label Y per tissue are fed into InterSTELLAR. (c) Self-attention pooling module with interpretable learning. (d) A cell-based attention heatmap generated by InterSTELLAR, in which the attention scores per cell are positively correlated with their contribution to the tissue classification results. Scale bar: 172 μm .

5.2.3 Network structure

For each constructed graph, a collection of $(\mathbf{Z}^{N \times F}, \mathbf{A}^{N \times N}, Y)$ are fed into InterSTELLAR (Figure 5.1b). The inputs $\mathbf{Z}^{N \times F}$ and $\mathbf{A}^{N \times N}$ first pass through two graph convolutional modules, and the node feature matrix is transformed as $\mathbf{h}_1^{N \times F_1}$ and $\mathbf{h}_2^{N \times F_2}$, respectively. Each module consists of a graph convolutional layer [127], a layer normalization module

[128] and a scaled exponential linear unit (SELU) [129]. Subsequently, $\mathbf{h}_2^{N \times F_2}$ is fed into a fully connected layer followed by another SELU module, and the output is $\mathbf{h}_3^{N \times F_3}$. In this sense, the two graph convolutional layers not only learn from the highly multiplexed cellular data, but also exploit the features from the cell interactions. This strategy enables the identification of spatial domains related to specific clinical subtypes.

To achieve tissue-scale classification and cell-scale interpretable learning, a self-attention pooling module, modified from [130], is embedded between $\mathbf{h}_3^{F_3 \times N}$ (the transpose of $\mathbf{h}_3^{N \times F_3}$) and the final output (Figure 5.1c). The attention score of the i -th cell is defined as Equation (5.5). As a result, the tissue-scale representation aggregated per the attention score distribution is defined as Equation (5.6), in which \otimes denotes element-wise multiplication.

$$a_i = \frac{\exp\{\mathbf{W}_1^{1 \times F_4}(\tanh(\mathbf{W}_2^{F_4 \times F_3} \mathbf{h}_{3,i}^{F_3 \times 1}) \otimes \text{sigmoid}(\mathbf{W}_3^{F_4 \times F_3} \mathbf{h}_{3,i}^{F_3 \times 1}))\}}{\sum_{i=1}^N \exp\{\mathbf{W}_1^{1 \times F_4}(\tanh(\mathbf{W}_2^{F_4 \times F_3} \mathbf{h}_{3,i}^{F_3 \times 1}) \otimes \text{sigmoid}(\mathbf{W}_3^{F_4 \times F_3} \mathbf{h}_{3,i}^{F_3 \times 1}))\}}. \quad (5.5)$$

$$\mathbf{h}_4^{F_3 \times 1} = \sum_{i=1}^N a_i \mathbf{h}_{3,i}^{F_3 \times 1}. \quad (5.6)$$

The attention score a_i is the cell-scale contribution to the final tissue-scale output (Figure 5.1d). It can be interpreted with the tissue-scale aggregation rule of attention-based pooling, which computes the whole tissue representation as the weighted average of all cells in the tissue by their respective attention score. In this sense, the higher the attention score, the greater the contributions of the corresponding cells to the tissue representation will be, and vice versa. Therefore, the attention scores can quantify the diagnostic value of cell communities to delineate tissue clinical types. Finally, $\mathbf{h}_4^{F_3 \times 1}$ is further transformed as $\mathbf{h}_5^{K \times 1} = \text{softmax}(\mathbf{W}_4^{K \times F_3} \mathbf{h}_4^{F_3 \times 1})$, and cross entropy is set as the loss function $\mathcal{L}_{\text{tissue}}$ between the tissue-scale prediction and the label Y , in which K is the tissue class number.

In the attention pooling module, an additional binary clustering objective is introduced so that class-specific features can be learnt [130]. During training, a collection of $(a_i, \mathbf{h}_{3,i}^{F_3 \times 1})$ is sorted according to the value of a_i , in which $i = 1, 2, \dots, N$. Then, the pairs of $(a_i, \mathbf{h}_{3,i}^{F_3 \times 1})$ with Q top highest and lowest a_i are selected. Next, K separate fully connected layers are utilized to process the selected $\mathbf{h}_{3,i}^{F_3 \times 1}$ for each class, respectively ($k = 1, 2, \dots, K$):

$$\mathbf{h}_{6,i,k}^{2 \times 1} = \mathbf{W}_{5,k}^{2 \times F_3} \mathbf{h}_{3,i}^{F_3 \times 1}. \quad (5.7)$$

Regarding the training of each classifier, the $\mathbf{h}_{6,i,k}^{2 \times 1}$ with the Q top highest a_i are attached with positive labels (+1) while those with the Q top lowest a_i are attached with negative labels (−1). The smooth top-1 Support Vector Machine (SVM) loss is selected as the loss function $\mathcal{L}_{\text{cell}}$ for this cell-scale learning task, because it has been empirically shown to reduce over-fitting under the conditions of noisy data labels or limited data [131]. Note that the labels are independently generated in each iteration. Intuitively, the sub-training task in each of the K classes is supervised by the corresponding tissue-scale label. Consequently, the cell communities with high attention scores are expected to be positive evidence for its current tissue label; By contrast, the communities with low attention scores are the negative evidence. Therefore, this sub-training task can be regarded as a constraint for the cell-scale feature $\mathbf{h}_{3,i}^{F_3 \times 1}$, such that the features favoring the correct outcome are linearly separable from those uncorrelated ones. The overall loss function \mathcal{L} , as Equation (5.8) shows, is the weighted sum of $\mathcal{L}_{\text{tissue}}$ and $\mathcal{L}_{\text{cell}}$, in which $\eta \in [0, 1]$ is the tissue-scale weight.

$$\mathcal{L} = \eta \mathcal{L}_{\text{tissue}} + (1 - \eta) \mathcal{L}_{\text{cell}}. \quad (5.8)$$

5.2.4 Model training

We preset that $F = 30$ and $K = 3$ for the breast cancer IMC dataset training task. Other hyper-parameters were set as $F_1 = 40$, $F_2 = 40$, $F_3 = 20$, $F_4 = 10$ and $Q = 8$. The optimal η was found as approximately 0.85 so that the highest accuracy is achieved (Figure 5.2).

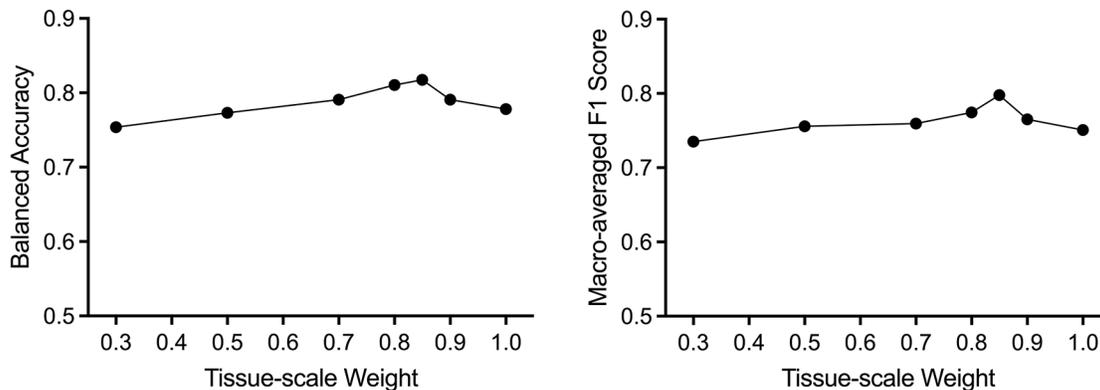


Figure 5.2: The relationship between tissue-scale weight η and clinical type classification performance.

The network was trained using Pytorch [132] (version 1.10.2) and Pytorch-Geometric [133] (version 2.0.4) on a single NVIDIA Quadro RTX 6000 GPU with 24 GB of VRAM. To mitigate the class imbalance in the training set, the sampling probability of each tissue was inversely proportional to the frequency of its label. Of the 366 tissues, 73 tissues were selected as test set, the remaining tissues were trained with a 10-fold cross validation strategy. During the training, the model parameters were updated via Adam optimizer with a learning rate of 3×10^{-4} and $L1$ weight decay of 3×10^{-5} . All the other parameters of the Adam optimizer were utilized with default values. A step learning rate strategy was also applied so that the learning rate was multiplied by 0.9 after each 5 epoches. The network was trained with 30 epoches with a batch size of 8. Each fold took approximately 166 seconds for training. The trained model with the highest validation accuracy was saved for each fold.

5.3 Accuracy Metrics and Statistical Analysis

To access the tissue-scale classification performance considering the label imbalance effect, balanced accuracy and macro-averaged F1 score were implemented with scikit-learn package [119] (version 1.0.2). Segmentation performance for tumor regions was accessed by F1 score with scikit-learn package.

Because InterSTELLAR aims to detect the cell communities with high diagnostic values, the predicted attention scores should be continuous through the cell neighbourhood, such that adjacent cells from same communities make similar contributions to the final output. In another word, the attention scores of the adjacent cells should be spatial correlated. We evaluated the correlation by Moran’s I [134] and Geary’s C statistics. Moran’s I metric is a correlation coefficient that measures how one spot is similar to other spots surrounding it. Its value ranges from -1 to 1 . The higher the value, the higher the spatial correlation relationship will be. For the given attention scores, we define the Moran’s I using the following formula,

$$I = \frac{N \sum_i \sum_j [m_{ij}(x_i - \bar{x})(x_j - \bar{x})]}{M \sum_i (x_i - \bar{x})^2}, \quad (5.9)$$

where x_i and x_j are the attention scores of cells i and j , \bar{x} is the mean attention score of a tissue, \mathbf{m}_{ij} is spatial weight between cells i and j calculated using the 2D spatial coordinates of the spots, and M is the sum of m_{ij} . For each cell, 4 nearest neighbours are selected based on the Euclidean distance between cells. If cell j is the nearest neighbour of cell i , m_{ij} is assigned as 1; otherwise, $m_{ij} = 0$.

Geary’s C can be also used for the spatial autocorrelation evaluation, which is calculated as

$$C = \frac{N \sum_i \sum_j [m_{ij}(x_i - x_j)^2]}{M \sum_i (x_i - \bar{x})^2}. \quad (5.10)$$

The value of Geary’s C ranges from 0 to 2. We transform it as $C^* = 1 - C$ so that its range will be $[-1, 1]$ [66]. Similar to Moran’s I , the higher the value of C^* , the higher the spatial correlation relationship will be between cells in the same neighbourhood.

Other than specially stated, quantitative data are presented as box-and-whisker plots (center line, median limits, 75% and 25% whiskers, maximum and minimum). The two-sided log-rank tests were implemented with Prism 9 (GraphPad Software Inc.). Statistical significance at $P < 0.05$, 0.01 are denoted by * and ** or # and ##, respectively.

5.4 Experiments

5.4.1 InterSTELLAR achieves accurate clinical type classification and cell-scale characterization

We first evaluated InterSTELLAR on clinical tissue subtype classification. From the data, 73 samples were selected as a test set, and the remaining 293 tissue were trained with a 10-fold cross validation strategy [135]. For reference, InterSTELLAR was benchmarked in comparison with a fully connected neural network (FNN), Random Forest algorithm, and SVM algorithm.

The FNN method is similar to InterSTELLAR – both tissue and cell-scale predictions can be conducted. However, the two graph convolutional layers are replaced by two fully connected layers. In this case, the spatial locations of cells are not taken into consideration so that the adjacent matrix $\mathbf{A}^{N \times N}$ is neglected. Thus, FNN can only utilize the single cell data without the information of spatial cell interactions. Random Forest and SVM algorithms are conducted on the basis of composition vector inputs, which are the cell densities of each phenotype per tissue. There are 25 cell phenotypes of the breast cancer IMC dataset (Figure

5.3). As a consequence, the input per tissue is a 25×1 vector. Before training and inference, the inputs are Z-score normalized. Compared to InterSTELLAR and FNN methods, only tissue-scale classification is available for Random Forest and SVM, due to the loss of cell-scale information. Note that both Random Forest and SVM were implemented using scikit-learn [119] (version 1.0.2) with default settings.

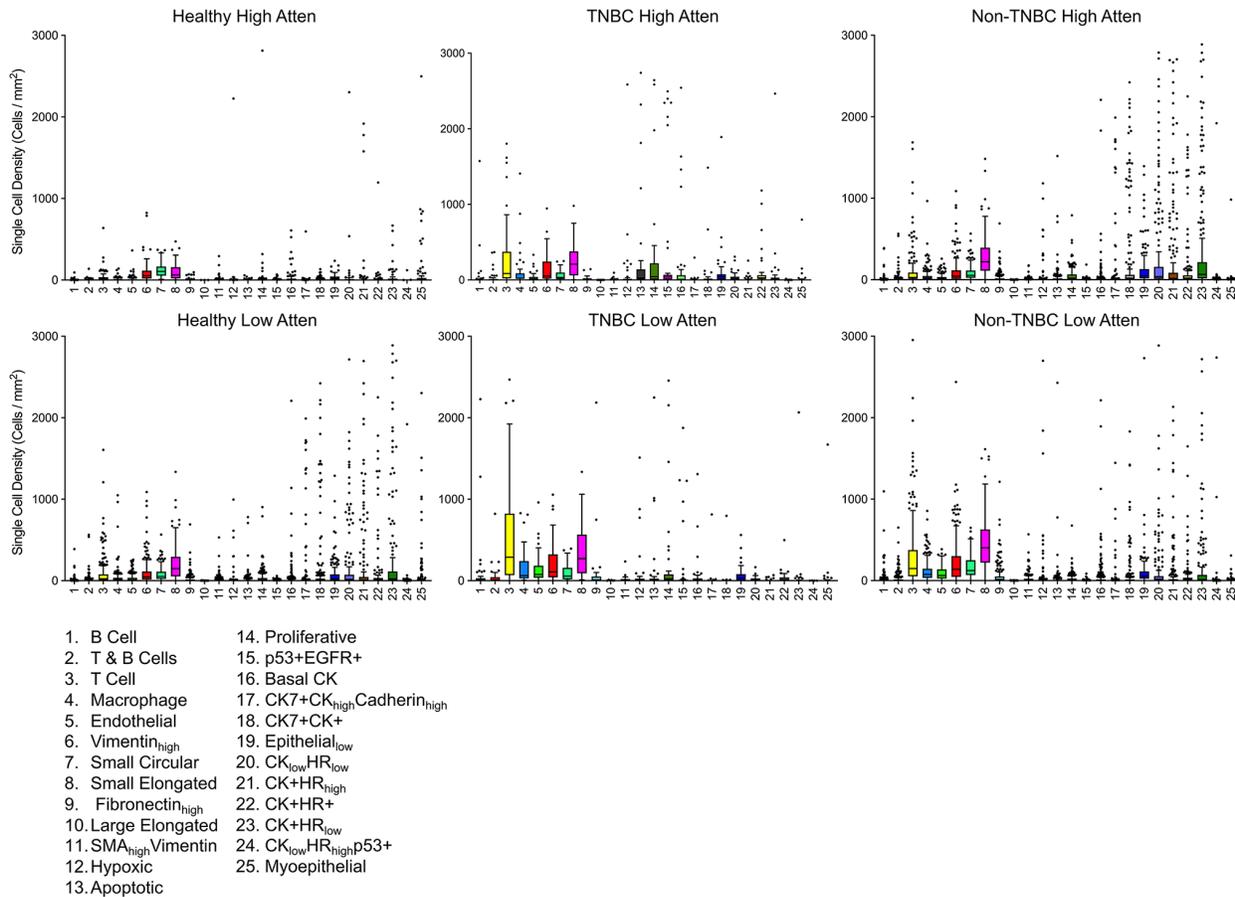


Figure 5.3: Box plots of cell density per phenotype in high- and low-attention regions from healthy, TNBC and non-TNBC cancerous tissues.

To account for the label imbalance issue (83 healthy, 49 TNBC and 234 non-TNBC cancerous tissues), both balanced accuracy and macro-averaged F1 score were used to evaluate the classification performance. As demonstrated in Figure 5.4a, InterSTELLAR outperforms

the three comparator algorithms in both cross validations and testing results. FNN most closely compares to InterSTELLAR of the three; however, FNN lacks information about the relative spatial arrangement of cells within a tissue, so that only the marker features are available in the learning task. Under this condition, some tissues with clinical outcome-relevant spatial structures may be misclassified. As a result, the FNN approach cannot outperform InterSTELLAR. For Random Forest and SVM, performances in cross validation are close to FNN; however, their accuracies decline severely in the independent test set. We infer that the loss of cell-scale information results in the poor performance of generalization of these two algorithms.

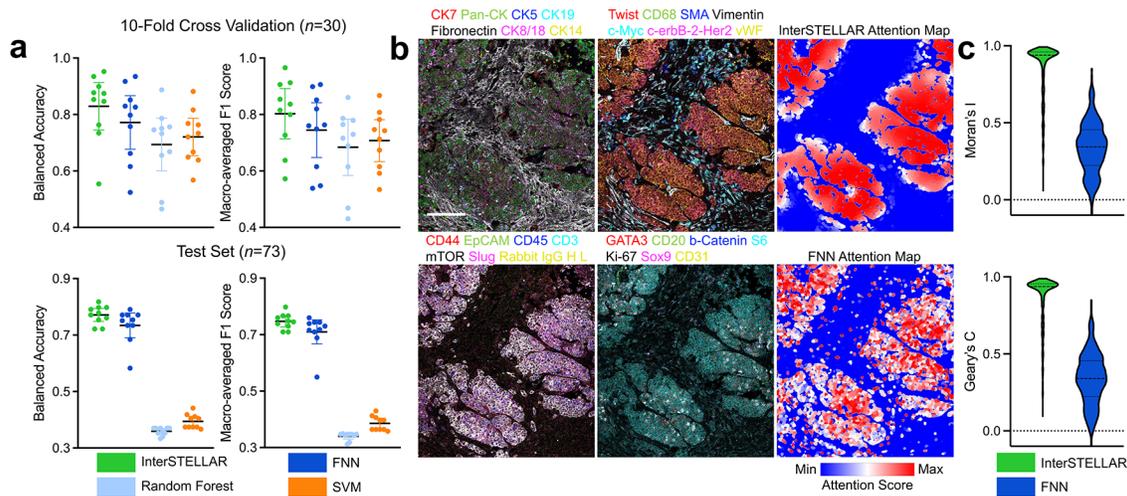


Figure 5.4: Evaluations of InterSTELLAR on tissue classification and cell community detection. (a) InterSTELLAR is more accurate than FNN, Random Forest and SVM algorithms on tissue clinical type classification, validated by 10-fold cross validation ($N = 30$ per fold) and an independent test set ($N = 73$). (b) Highly multiplexed IHC images and the corresponding attention heatmaps generated by InterSTELLAR and FNN. (c) Moran's I and Geary's C statistics indicate InterSTELLAR achieves higher spatial correlation than FNN in terms of attention scores. Scale bar: $145 \mu m$.

Subsequently, we analyzed the predicted cell-scale attention scores of InterSELLAR, and again benchmarked performance to FNN. Random Forest and SVM are omitted from benchmarking, because they are not feasible methods to predict tissue and cell-scale outcomes simultaneously, due to heterogeneous cell number per field of view in the tissue data set. Both InterSTELLAR and FNN can construct cell-based heatmaps through attention score values (Figure 5.4b). However, with more homogeneous distribution of attention scores, InterSTELLAR is superior than FNN on interested community identification. In fact, the attention scores in the same community should be continuous due to the cell-cell communications, such that the cells from the neighbourhood make similar contributions. That is, the attention score of a single cell should be spatial correlated with those of its neighbours. To evaluate the spatial cell correlation per tissue, we benchmarked InterSTELLAR and FNN in terms of Moran’s I and Geary’s C statistics. Moran’s I measures how one region is similar to other spots surrounding it. If the regions are attracted by each other, it implies the regions are not independent. Therefore, it is positively correlated with the spatial correlation relationship. Similar to Moran’s I , Geary’s C is also used for the evaluation of spatial autocorrelation. Both the quantitative results of these two metrics in Figure 5.4c validate that InterSTELLAR achieves higher spatial correlation, which endorses its improved ability to identify interested communities.

With cell neighbourhood information integrated in training, we hypothesized that the latent embedding features of InterSTELLAR could reflect biologically meaningful information about the tissue structure. Therefore, we proposed to classify all the cells per cancer tissue to either tumor or stromal regions, and then compared the segmentation results with manually labeled tumor region masks by Ilastik [7, 136] (Figure 5.5a). Specifically, a K-means clustering algorithm with $K=2$ was utilized to cluster the latent cell embedding features from the last hidden layers of the trained InterSTELLAR network (the input features to the

attention pooling module). As a reference, the latent cellular features from the FNN were also clustered with the same approach. Visual inspection and quantitative evaluation confirm InterSTELLAR is superior at capturing tumor regions in both TNBC and non-TNBC cancerous tissues (Figures 5.5b and c). Indeed, the assembled tumor organizations from InterSTELLAR are much closer to the manual annotations. Interestingly, the FNN has a particular weakness for errantly identifying isolated epithelial cells as tumor cells within the stromal region. Together, these analyses reveal the capabilities of InterSTELLAR over reference methods to identify and interpret class-level features.

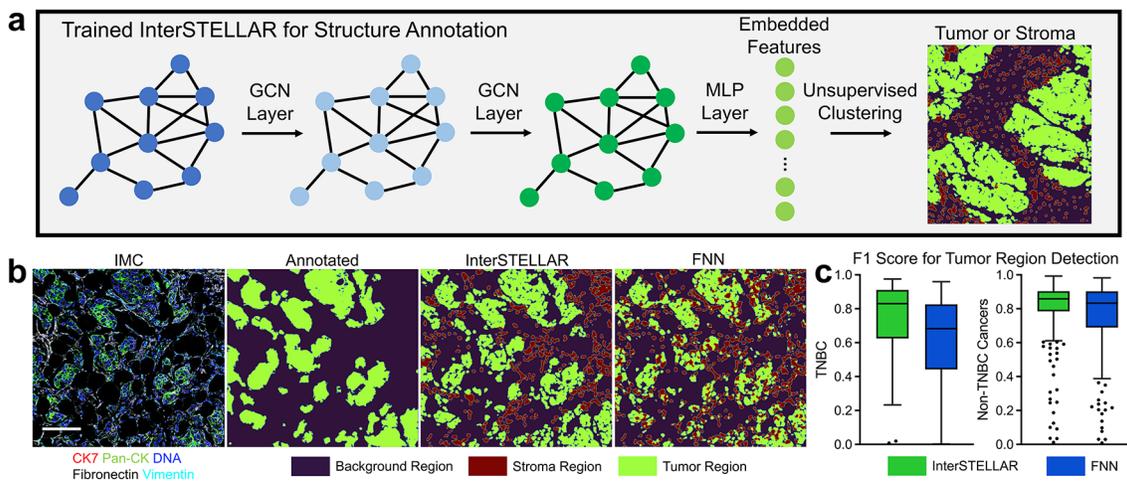


Figure 5.5: Evaluations of InterSTELLAR on tumor region detection. (a) Schematic of generating tumor and stroma masks. Embedded features are extracted from a trained network and then classified by unsupervised clustering algorithms. Here, K-means clustering with $K = 2$ is utilized. (b) Highly multiplexed IHC images and the corresponding tumor and stroma masks generated by manual annotations, InterSTELLAR and FNN. (InterSTELLAR performs better than FNN on tumor region identifications of both TNBC and non-TNBC cancerous tissues in terms of F1 score. Scale bar: $158 \mu\text{m}$.)

5.4.2 Attention Mapping by InterSTELLAR Across Cancer Tissue Types

Pathological categorization of breast cancer is critical for patient care and is typically accomplished with well-established markers for hormone status, HER2 expression and tissue morphology. We next investigated whether InterSTELLAR enables microenvironment characterization for pathologically distinct clinical types of breast cancer from the same IMC dataset. We defined high- and low-attention regions by segmenting the attention heatmaps with the median attention score of each sample (Figure 5.6a). With these binary masks, we calculated the percentages of immune, stromal and epithelial cells in high- and-low attention regions for healthy, TNBC, and non-TNBC cancerous tissues. The attention regions for each tissue type show distinguishing compositions for these cell types (Figure 5.6b). First, we observed that healthy tissues have higher stromal cell proportion in both attention regions than any cancer tissues. Importantly, high-attention regions within cancer tissues contain more epithelial cells than of all the remaining regions. Comparatively, the low-attention regions correspond with an increased stromal cell presence than the high-attention regions from the same tissues. Low-attention regions from all three clinical types contain similar portions of epithelial cells. Interestingly, TNBC tissues had the highest proportion of immune cells compared to healthy or non-TNBC cancerous tissues, revealing higher immune activity in the TNBC tissue microenvironment. In non-TNBC cancerous specimens, low-attention regions have a higher proportion of immune cells than healthy tissues, but there are no appreciable differences in immune cell percentages of the high-attention regions. In sum, the variations of cell composition by attention scores can distinguish the microenvironments from different clinical subtypes of breast cancer.

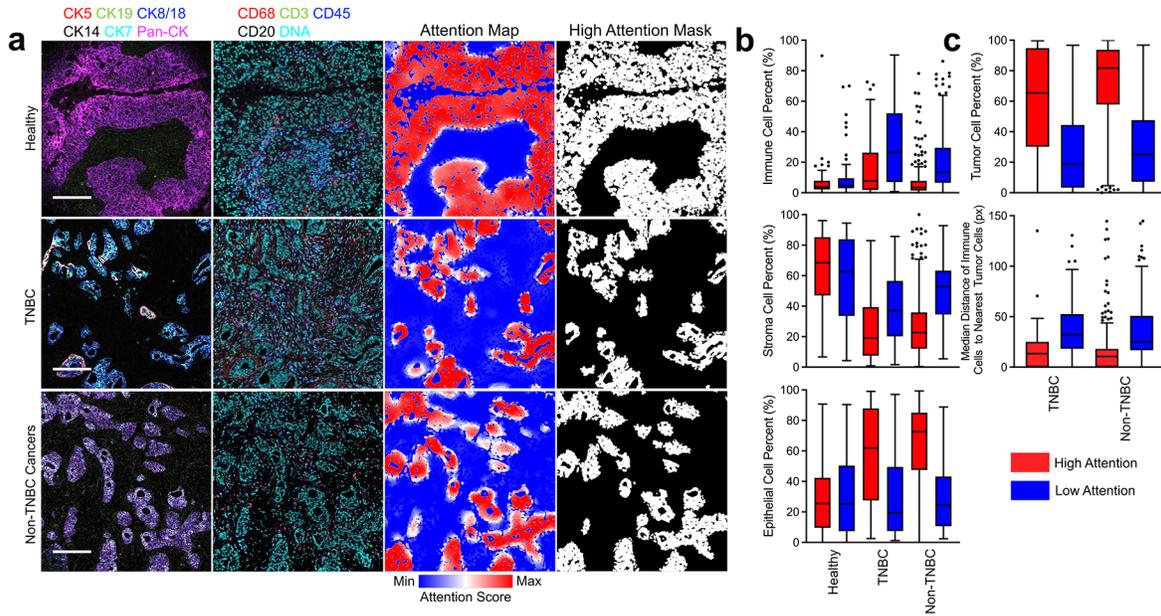


Figure 5.6: InterSTELLAR characterizes the breast cancer tissue microenvironments from different clinical subtypes. (a) Highly multiplexed IMC images of healthy, TNBC and non-TNBC cancerous tissues as well as their corresponding attention heatmaps and segmented high attention region masks. (b) The percentages of immune, stroma and epithelial cells in high and low attention regions from healthy, TNBC and non-TNBC cancerous tissues. (c) The percentages of tumor cells and median distance of immune cells to their nearest tumor cells in high and low attention regions from cancer tissues. Scale bars: Healthy: $195 \mu m$; TNBC: $175 \mu m$; Non-TNBC cancers: $183 \mu m$.

Next, we calculated the percent of tumor cells in high- and low-attention regions from cancer tissue specimens (Figure 5.6c). As expected, high-attention regions are occupied by a greater number of tumor cells than low-attention ones. As an example, previous reports have suggested that tumor-infiltrating lymphocytes may be an indication of TNBC [137]. Additionally, high-attention regions have higher ratios of tumor cells in non-TNBC cancerous tissues than TNBC, which is inversely related to the immune cell ratios for these tissues. This suggests more frequent immune-tumor cell interactions in TNBC tissues. Moreover, we calculated the median distance of immune cells from both high- and low-attention regions

to their nearest tumor cells (Figure 5.6c). The results reveal that immune cells in high-attention regions are in closer proximity to tumor cells. Therefore, more frequent immune-tumor interactions, such as tumor-infiltrating lymphocytes, are expected in these regions.

5.4.3 InterSTELLAR captures tissue microenvironmental features from different clinical subtypes

Different clinical types of breast tissue are characterized by distinctive cell phenotypes compositions, for example, myoepithelial cells in healthy tissues and extensive proliferative cells in TNBC. Now, we ask whether distinct attention regions can also be characterized by distinct cell-type compositions, even in the same tissue. We adopted a cell phenotype approach for breast cancer tissues as previously described [7]. To conduct more detailed phenotype analysis, we calculated the cell density per phenotype for all the attention regions (Figure 5.3). The normalized distributions of the mean cell density of each phenotype are summarized in Figure 5.7a. This demonstrates that in TNBC high-attention regions there are more Basal CK and Epithelial_{low} cells, but fewer myoepithelial cells, compared to TNBC low-attention counterparts. To quantify such differences, we calculated the KL divergence of the phenotype distributions between high- and low-attention regions per sample (Figure 5.7b). The results reveal that almost all the KL divergence values are away from 0, indicating that cell phenotype compositions enable expected attention score differentiation. Interestingly, more diverse distributions are noticed in healthy tissues rather than cancer types. We inferred that this reveals more heterogeneity in the microenvironments in healthy tissues.

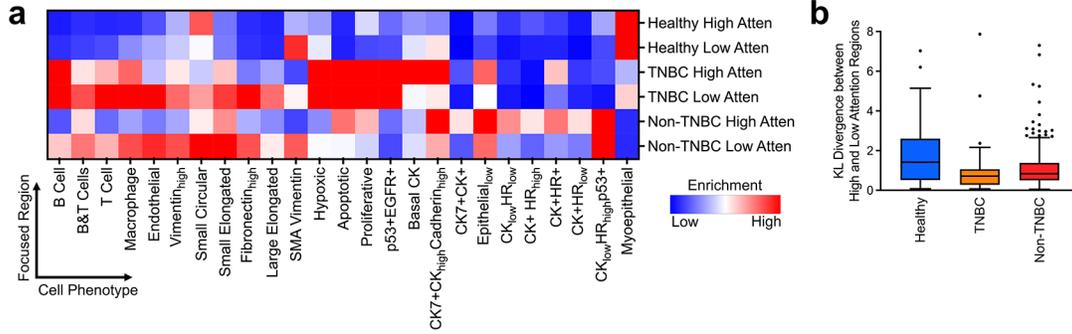


Figure 5.7: InterSTELLAR characterizes tissue microenvironments from different clinical subtypes. (a) Mean cell density per phenotype from high and low attention regions of all the tissues. (b) KL divergence between high and low attention regions regarding distributions of cell density per phenotype.

5.4.4 InterSTELLAR uncovers single-cell pathology groups associated with patient survival

Beyond the microenvironment and spatial characterization for various clinical types, we were curious whether InterSTELLAR could benefit single-cell pathology (SCP) analysis. Using the unsupervised Phenograph algorithm with Leiden algorithm [121, 122], we grouped patient cancer tissues on the basis of cell phenotype densities within high-attention regions. These algorithms were implemented by Scanpy package [138] (version 1.8.2) with 20 nearest neighbours and resolution of 1. All the other parameters were utilized with default settings. These parameters were chosen such that groups of patients from distinct cell type compositions can be successfully separated without limiting statistical power for group comparisons. Similar unsupervised clusterings were also conducted on the basis of the cell densities of all the phenotypes in low attention and all regions, and epithelial cell densities from high attention and all regions. The parameters of Phenograph kept unchanged. The random seeds for the individual runs were recorded.

With the algorithms, 8 SCP subgroups were identified based on cell phenotype densities within high-attention regions, which are named according to their dominant phenotypes (Figure 5.8a). The dimensions of the density data were reduced by uniform manifold approximation and projection (UMAP) algorithm for visualization. This algorithm was implemented by the umap package (version 0.5.2) after all the inputs were Z-score normalized. All the parameters were utilized with default settings. The random seeds for the individual run was recorded. Through inspection, the tissues are clustered by their distinct phenotype compositions (Figure 5.8b).

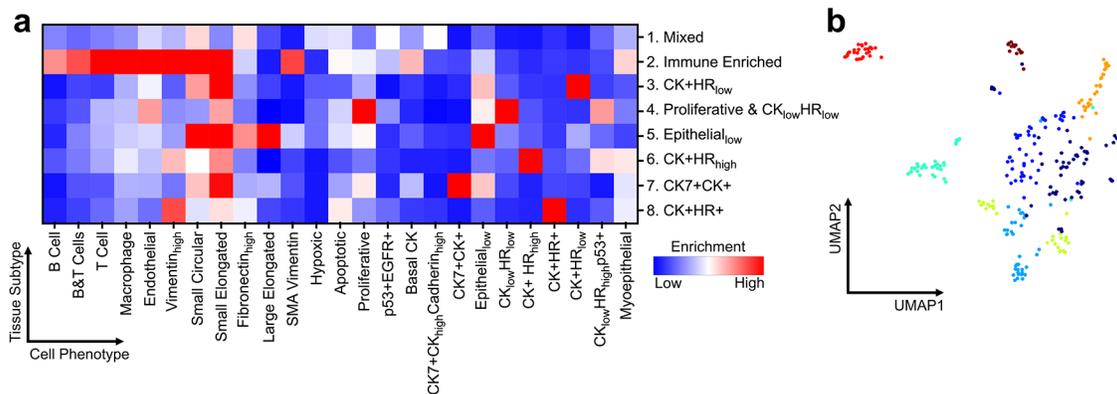


Figure 5.8: InterSTELLAR characterizes distinct clinical outcomes for SCP subgroups. (a) Mean cell density per phenotype of high attention regions from various SCP subgroups. (b) UMAP plot of the tissues labeled with their corresponding SCP indexes.

Importantly, the presence of these subgroups have significantly different clinical outcomes in overall survival (Figure 5.9a), validated by a log-rank test with $P = 0.0425$ on Kaplan-Meier curves using Prism 9 (GraphPad Software Inc.). In particular, the CK7+CK+ subset in high attention regions defined patients with favorable clinical outcome, while the presence of proliferative CK_{low}HR_{low} was associated with adverse overall survival. Hormone receptor and HER2 subtypes and tumor grade were associated with prognosis, as expected (Figure 5.10). Most remarkably, SCP features identified cohorts that were independent from clinical

subtype or tumor grade (Figure 5.10a) with distinct survival results (Figure 5.10b). Also compared to clinical subtype and tumor grade results, SCP subgroup analysis with InterSTELLAR allows a higher-resolution tissue characterization paradigm (4 and 3 vs 8).

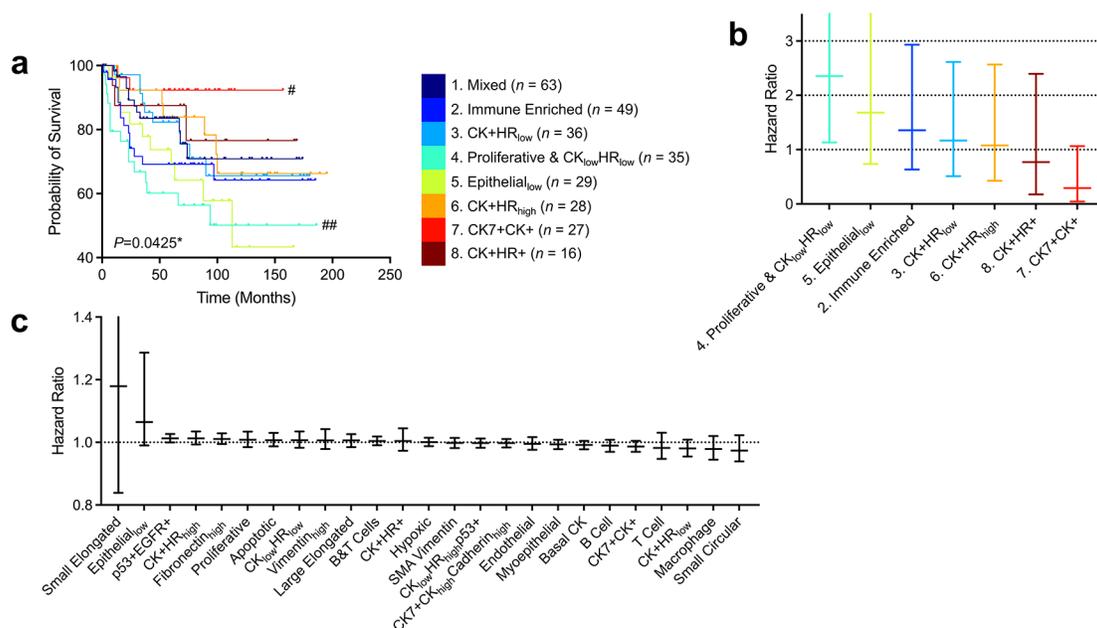


Figure 5.9: Survival analysis for SCP subgroups. (a) Kaplan-Meier curves of overall survival for each subgroup ($N = 283$) on the basis of cell density per phenotype of high attention regions, with $*P < 0.05$. $\#P < 0.05$, $\#\#P < 0.01$ represent the statistical significance of a single subgroup compared to all other samples. (b) P values for overall survival analysis from different clustering strategies. In (a) and (b), P values were calculated through two-sided log-rank test. (c) Relative hazard ratios and 95% confidence intervals of disease-specific overall survival for cell densities per phenotype and SCP subgroups estimated using a Cox proportional hazards model. Reference group 1: Mixed for SCP subgroups.

Based on the observation that SCP analysis within high-attention regions appeared associated with clinically significant outcomes, we examined whether this association was also present for low-attention regions or whole tissue analysis. Applying the approach above for low-attention regions ($P = 0.8424$) or whole regions ($P = 0.1381$), there was no statistically significant difference in assessing outcome (Figure 5.11 and Table 5.2). We conclude that

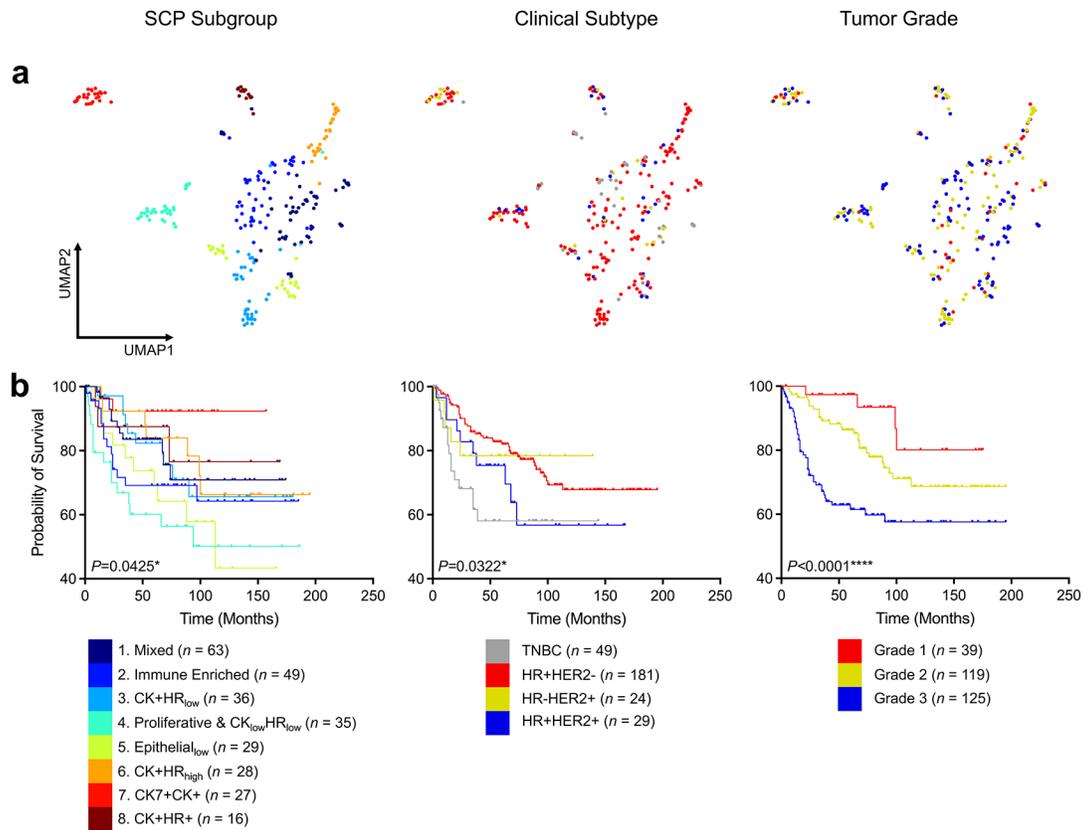


Figure 5.10: Comparisons of the patient group classified by SCP subgroup, clinical subtype and tumor grade. (a) UMAP plots of the tissues labeled with their corresponding SCP indexes, clinical subtypes and tumor grades. (b) Kaplan-Meier curves of overall survival for each patient group on the basis of SCP subgroup, clinical subtype and tumor grade.

high-attention regions have greater diagnostic value such that they are more relevant to contribute to patient outcomes. Since high-attention regions contain a higher proportions of epithelial cells, we used the same unsupervised clustering approach based on only epithelial cells from high-attention regions, or clustering using epithelial cells from the entire tissue sample (Figure 5.11 and Table 5.2). Both of these approaches based on using only epithelial cells did not reach the significance of all cell types in the high-attention region, indicating that indeed the spatial organization of multiple cell types are essential in microenvironment analysis and interpretation.

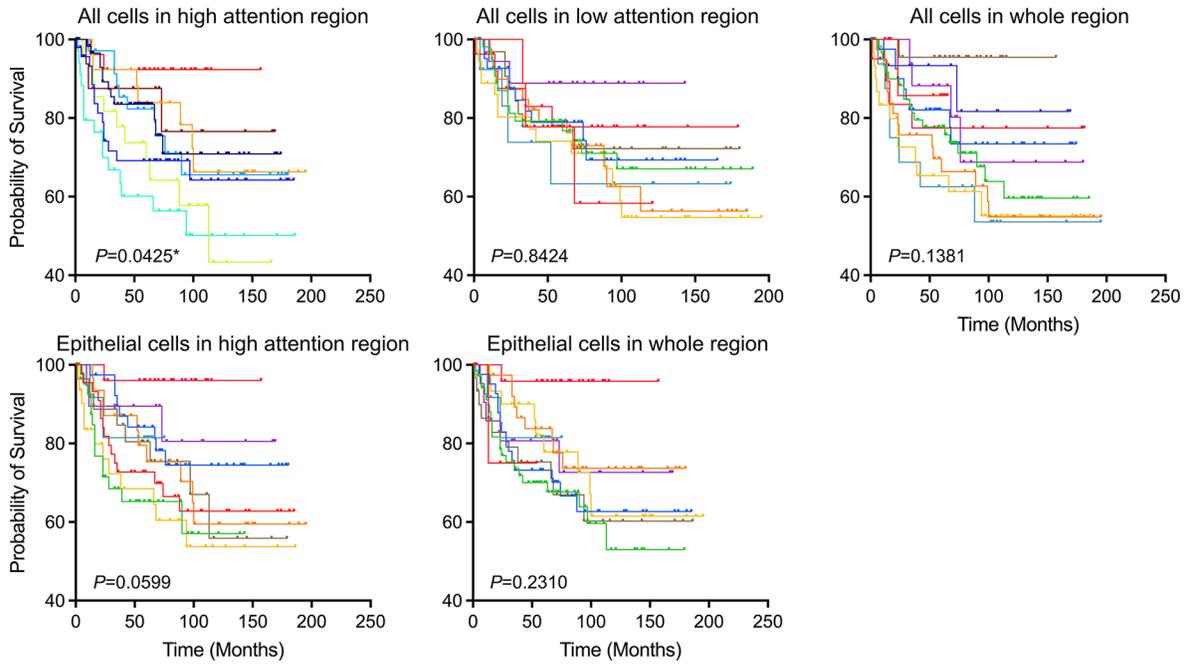


Figure 5.11: Kaplan-Meier curves of overall survival from different clustering strategies and their corresponding P values of log-rank tests.

Table 5.2: P values for overall survival analysis from different clustering strategies.

Cell Types	Region	P Value
All	High attention	0.0425*
All	Low attention	0.8424
All	Whole	0.1381
Epithelial cells	High attention	0.0599
Epithelial cells	Whole	0.2310

Inspecting the subgroups identified through SCP analysis of high-attention regions, prognostic groups become apparent (Figure 5.12). SCP Group 4 Proliferative & CK_{low}HR_{low} has very poor prognosis with less than 70 percent overall survival at 5 years (Cox proportional hazard HR=2.36, CI: 1.13–4.99, Figure 5.9b), while SCP Group 5 Epithelial_{low} (HR=1.68, CI: 0.74–3.76) and Group 2 Immune Enriched (HR=1.36, CI: 0.63–2.93) are moderately unfavorable. In contrast, SCP Group 7 CK7+CK+ exceeds 90 percent overall survival beyond 10 years (HR=0.29, CI: 0.05–1.07). SCP Groups 1, 3, 6 and 8 have intermediate risk. Alternatively, when analyzing each single cell phenotype within the high-attention region but without clustering by tissue composition, no individual cell type was directly correlated with outcomes (Figure 5.9c). Thus, SCP subgroups clustered by the tissue composition from high-attention regions provide an innovative approach to inform prognosis.

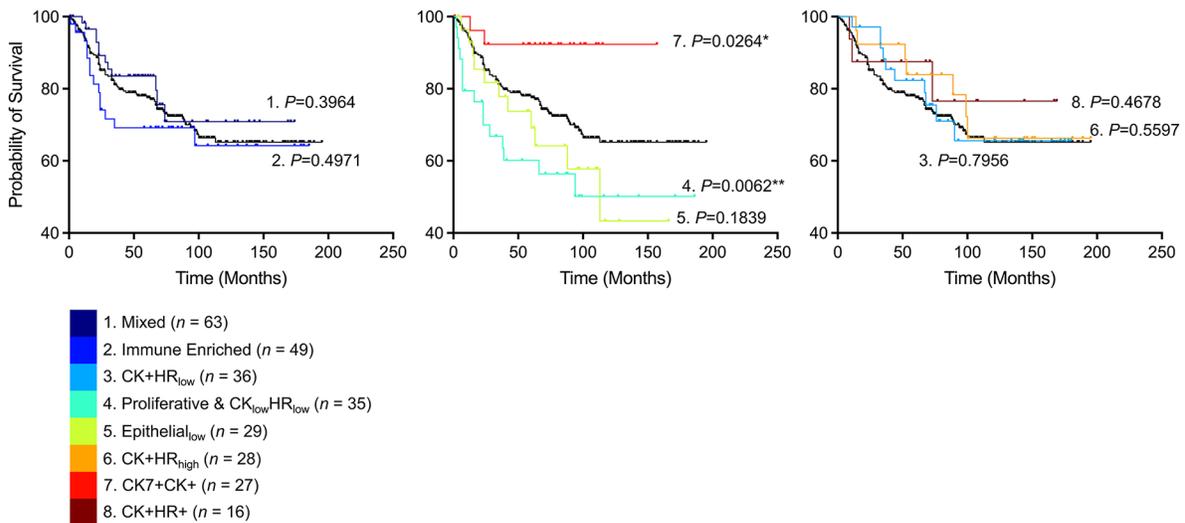


Figure 5.12: Kaplan-Meier curves of overall survival of each subgroup on the basis of cell density per phenotype of high attention regions. The black curve represents the survival curve of all samples. Each P value represents the statistical significance of a single subgroup compared to all the other samples. All the P values were calculated through two-sided log-rank tests.

5.5 Discussion

With the expansion of novel multiplexed technologies for the characterization of cellular context in health and disease, graph-based deep learning algorithms have begun to be investigated on high dimensional single cell data. In this section we present InterSTELLAR, a GNN framework that can predict patient tissue outcomes and disease relevant communities simultaneously. We introduce and evaluate InterSTELLAR using an open-source breast cancer IMC dataset. Cell microenvironments per tissue are first modeled as graphs, and the nodes represent cells and the edges represent inter-cell communications. Then, graph convolutional layers are applied to extract cell interaction features, followed by a self-attention pooling module to learn the cell-based contribution to clinical outcomes. InterSTELLAR achieves higher accuracy than traditional machine learning algorithms, including Random Forest, SVM, and a FNN framework which neglects spatial cell information. Moreover, InterSTELLAR performs better than FNN in terms of interested community identification. We infer that modeling cell communications is crucial to accurate tissue characterization.

We performed validation studies indicating that InterSTELLAR can capture distinct tissue microenvironment features from healthy and breast cancer tissues. With high- and low-attention regions segmented by median attention score values, we observed disease-specific composition of immune, stroma and epithelial cells per tissue. This approach removes manual segmentation of regions of interest by an expert pathological reader to identify regions within a field of view of high (and low) value for correlating with microenvironmental subclasses. Our analyses revealed that greater numbers of tumor cells are localized in high-attention regions of cancer tissues, which potentially distinguish communities by diagnostic value. Proximity of immune and tumor cells from high-attention regions is indicative of active immune-tumor interactions in these domains. Furthermore, the heterogeneity of low- and

high-attention regions based on cell phenotypes reveals the ability of InterSTELLAR to discriminate clinically important aspects of the tumor microenvironments.

Through SCP subgroup analysis, we find that InterSTELLAR can establish a mapping relationship between tissue microenvironment and breast cancer patient prognosis. Specifically, each subgroup has unique cell community features in high attention regions, which are significantly associated with the survival outcomes. By contrast, other attention regions or areas of epithelial cells do not have this association. Moreover, both Kaplan-Meier curves and Cox proportional hazards modeling suggest distinct survival outcomes of the subgroups.

The results discussed in this section demonstrate the value of InterSTELLAR for high-dimensional spatial cell data to investigate cellular communities of interest. A particular strength of the InterSTELLAR approach is the flexibility to apply this method across multiple platforms for highly multiplexed imaging, because it is capable of extracting cell marker features and tissue graphs independent of the imaging modality. As a supervised learning framework, InterSTELLAR requires substantial training data to guarantee the generalization of the trained model and careful minimization of batch effects on marker staining to avoid overfitting and to prevent degradation of prediction performance. The current study analyzes a large, recent cohort of patient tissue data, but still remains limited in providing sufficient size for validation of the intriguing findings suggested by SCP analysis. Future advances in imaging platforms and staining methods that improve the ability to generate affordable, very large data sets will be an exciting opportunity for wider implementation of the InterSTELLAR approach. The proposed framework can be applied to other highly multiplexed imaging techniques and diseases for enhanced downstream analysis.

5.6 Conclusion

To conclude, InterSTELLAR is a versatile GNN framework for highly multiplexed imaging data that simultaneously classifies tissue types by clinical classes and predicts disease-relevant cell communities. Most importantly, by exploiting cell communities with high diagnostic values, it enhances the characterization of patient tissue microenvironments.

Chapter 6

Development of New Theranostic Approaches by Integrating Multi-modality Imaging Data

6.1 Introduction

In the previous three chapters, we have developed algorithms to restore DAR and IMC images, as well as graph-based downstream analysis framework for multiplexed imaging data. With the enhanced data and novel framework, we propose to develop new theranostic analysis approaches by integrating multi-modality imaging data in this chapter, including DAR, IMC and H&E histopathological images. We first develop and verify the integration

Partial content of this chapter has been published in The Journal of Nuclear Medicine [38] and as a conference article [40].

procedure of DAR and H&E data, by utilizing restored DAR and the corresponding H&E stained histopathological images from the human bone biopsies treated with $^{223}\text{RaCl}_2$ in Chapter 3. Thereafter, we develop and verify the procedure of IMC and H&E data, by utilizing restored IMC and the corresponding H&E images from the human bone marrow data in Chapter 4. With the H&E data as reference, DAR and IMC images can then be co-registered for dose and microenvironment correlation analysis.

6.2 Multi-modality Image Integration Framework

Before data integration, the raw DAR images are restored by the PG-PEM algorithm and the raw IMC images are denoised by the IMC-Denoise pipeline, respectively. Next, the H&E stained histopathological images are set as references to co-register DAR and IMC data after tissue structure segmentation. Specifically, the H&E and DAR images are co-registered with the structural features of tissues, while the H&E and IMC images are co-registered with the segmented nuclei. Consequently, the DAR and IMC data are co-registered with the H&E images as references. After the registrations, dose distribution measurement can be then conducted based on the restored DAR images and tissue microenvironment can be analyzed based on the denoised IMC images. The whole procedure is illustrated as Figure 6.1.

6.2.1 DAR and H&E image registration framework

Based on prior knowledge, ^{223}Ra tends to adhere to the bone surface [83], suggesting that the regions of high activity are likely concentrated within the bone tissue. In light of this, we intend to co-register DAR and H&E images by prioritizing the high activity areas identified in DAR alongside a bone mask extracted from H&E images. To facilitate successful registration

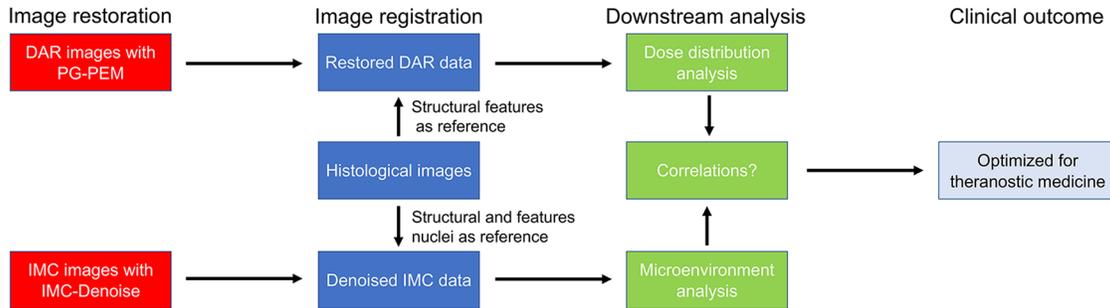


Figure 6.1: Schematic of theranostic analysis by integrating histological images, DAR and IMC data. The DAR images and IMC images are restored by PG-PEM and IMC-Denoise, respectively. Then, the histological images are co-registered with the restored DAR and IMC images, respectively. Subsequently, downstream analysis including dose distribution measurement and microenvironment analysis can be conducted, in which the correlations between tissue disease states, dose distribution and microenvironment are analyzed. Finally, theranostic medicine can be enhanced based on the results.

and enable on-tissue dose analysis, our initial focus is on the development of segmentation methods for both bone and soft tissues.

Hand-crafted features extraction and non-osseous tissue segmentation

The segmentation pipeline is summarized as Figure 6.2. Color variations from staining difference can impact segmentation results. To compensate this error, the raw H&E stained images in RGB space are converted into CIELAB space [139] and then normalized using Z-score means method. Additionally, the size of the histopathology images is large (about 20000×15000 pixels) relative to that of the autoradiograms (at approximately 300×200 pixels). As a result, pixel-wise segmentation is not needed in terms of registration and dose calculation. Here, we propose a patch-based segmentation pipeline: each patch is considered as a pixel in the segmented image. Thus, this problem is simplified as an image classification problem. For non-osseous tissue segmentation, K-means clustering method [140] is used; for bone segmentation, we propose to test the performance of network.

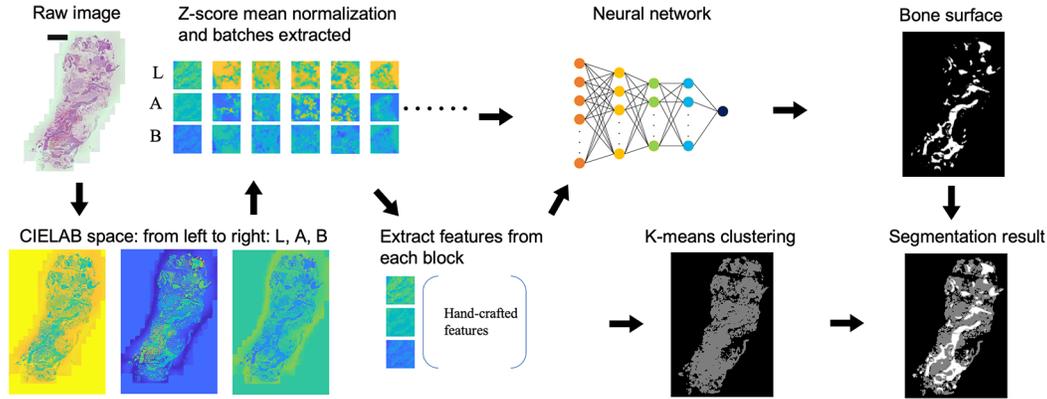


Figure 6.2: The segmentation pipeline: The raw image was converted to CIELAB space. After Z-score normalization, blocks were sampled, and hand-crafted features were extracted from them. The non-osseous tissues were predicted based on K-means clustering method. The bone surfaces were predicted using a network combining the extracted features and CNN trained features. Scale bar: 640 μm .

To segment the image using K-means clustering algorithm, features need to be extracted from every patch. In this part, three kinds of features are extracted from LAB channels: 1) Low order features (LOF) including mean, standard deviation, kurtosis and skewness (12 features); 2) Local binary pattern [141] (LBP) with radius of 1 and neighbors of 8 (177 features); 3) Rotation-invariant gray-level co-occurrence matrix-based [142] features (GLCM, 225 features): Each patch is binned into 16 gray-scale values. Subsequently, 20 gray-level co-occurrence matrices are extracted from 4 directions with 5 distances. Next the matrices are combined into 5 rotation-invariant matrices and the features are extracted. The features include: contrast, dissimilarity, correlation, energy, homogeneity, auto correlation, cluster shade, cluster prominence, maximum probability, sum of squares variance, sum average, sum variance, sum entropy, difference entropy, difference variance.

The soft tissue regions of the tissue sections are difficult to label manually, however the tissue and background are easily separated based on the low order features. Here a K-means clustering method is used to segment tissue and background. Once the tissue and bone are both segmented, the non-osseous regions can also be segmented. Note that here we do

not consider supervised ML methods because: 1) Compared to bone, the non-osseous tissue is more difficult to label manually; 2) Using unsupervised method like K-means clustering method, the segmentation results are adequate based on visual evaluation; 3) In the aspect of dose calculation, bone regions have greater importance than non-osseous tissue.

Bone segmentation

We manually labeled 12 slides as the training set. If every patch is simply set as bone or background, some information will be lost. For example, some bone boundary pixels are the mixture of bone and background, so they cannot simply be labeled as bone or background. This is known as tissue-fraction effects. To solve this problem, we set the ratio of the bone as the label of every patch. Then, this problem becomes a fuzzy segmentation problem.

Here, we utilize a network for the fuzzy segmentation and estimation task [143]. We have tested three different network structures as Figure 6.3 shows: 1) fully-connected neural network (FNN), which utilizes the hand-crafted features as the input; 2) CNN, which utilizes the patches as the input; and 3) Multi-features neural network (MNN), which utilizes the patches and the corresponding hand-crafted features as the input. The loss functions of the networks are all binary cross entropy (BCE) [143]. Based on the result in [143], this loss function is able to do unbiased estimation for fuzzy segmentation tasks.

Registration between the histopathology and DAR images

After H&E image segmentation, the processed images can be registered based on mutual information [144] with the restored DAR images by PG-PEM in Chapter 3. Fiducial markers are not needed here based on the assumption and prior information: most radium should fix on bone or its neighbors. In the registration process, scaling, rotation and translation need to be performed for the segmented image to correlate with the restored DAR image.

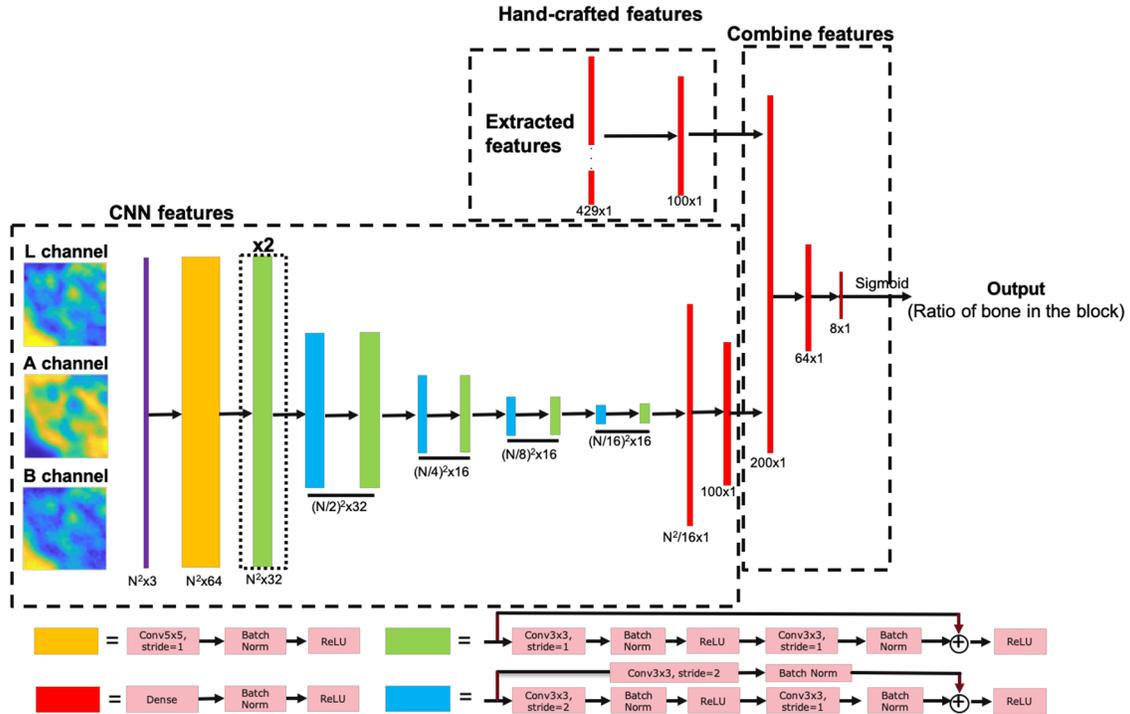


Figure 6.3: Network structure, in which the upper part can be seen as a FNN, the lower part is a CNN and the whole network is a MNN.

The registration algorithm is very easy to converge to local minimum so initial registration is needed.

The registration procedure of DAR and H&E images is summarized as follows (Figure 6.4):

- 1) Find the bounding boxes, centroids and orientation directions of the segmented bone mask from H&E image and high activity region from restored DAR image, respectively. Note that the high activity region can be segmented by a simple K-means clustering algorithm with $K=2$ or 3.
- 2) Use the bounding boxes to estimate their areas and then estimate the scale value. Use the orientation directions to estimate the rotation angle. Use the centroids to estimate the translation distance.

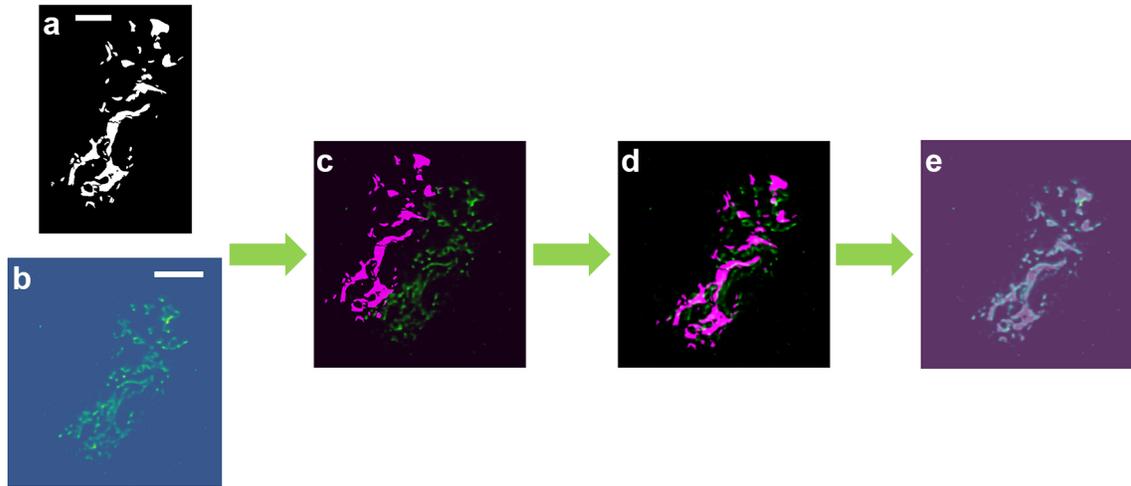


Figure 6.4: Registration process of DAR and H&E images. (a) Segmentation result. (b) Restoration result. (c) Fusion of (a) and (b). (d) Initial registration result. (e) Final registration result. Scale bar: (a) 1 mm, (b) 2.3 mm.

3) Do the initial registration.

4) Do the registration based on mutual information using gradient descent or one plus one evolutionary method as the optimization method.

6.2.2 IMC and H&E image registration framework

In H&E images, the nuclei are stained with haematoxylin. Consequently, the registration of IMC and H&E images becomes feasible through the segmentation of nuclei in H&E images and the corresponding DNA channel images in IMC (Figure 6.5). To accomplish this alignment, the nuclei in the H&E images are initially segmented using Ilastik. As the slides are mounted during H&E and IMC staining, the alignment of the segmented H&E image with the DNA image of IMC in the registration process requires only scaling and translation operations. It is important to note that the scaling operation is initiated by computing the pixel size difference between the H&E and IMC images.

The registration process of IMC and H&E images is summarized as follows (Figure 6.5):

- 1) Use Ilastik to segment the nuclei in H&E images.
- 2) Calculate the scale difference between IMC and H&E images, and then scale the H&E images.
- 3) Do the registration based on mutual information using gradient descent or one plus one evolutionary method as the optimization method.

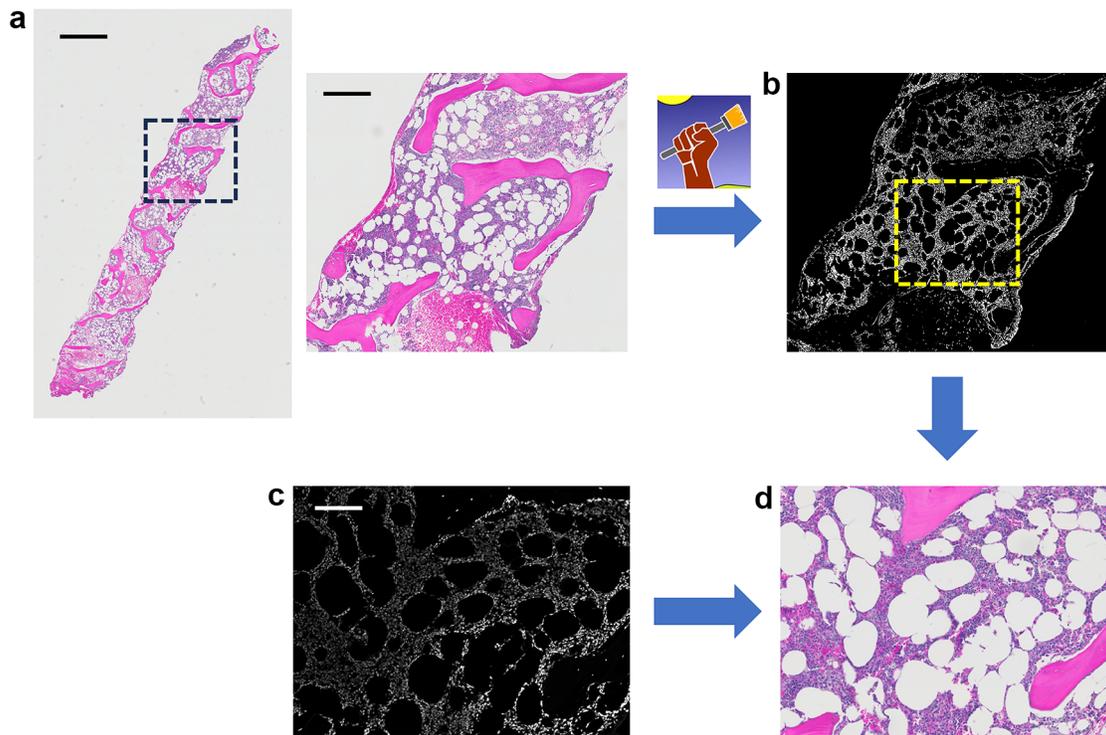


Figure 6.5: Registration process of IMC and H&E images. (a) A H&E image from the human bone marrow data of Chapter 4 and its zoomed-in version. (b) Nuclei segmentation mask. (c) A DNA channel image of IMC data. (d) The final registered H&E image. Scale bar: (a) left: 1.4 mm, right: 356 μm , (c) 138 μm .

6.3 Experiments

6.3.1 DAR and H&E image registration results

Tissue segmentation

As stated in last section, we used K-means clustering method to segment the tissue and background. The results are shown in Figure 6.6. Visual inspection suggest that the method is able to segment the tissue regions correctly.

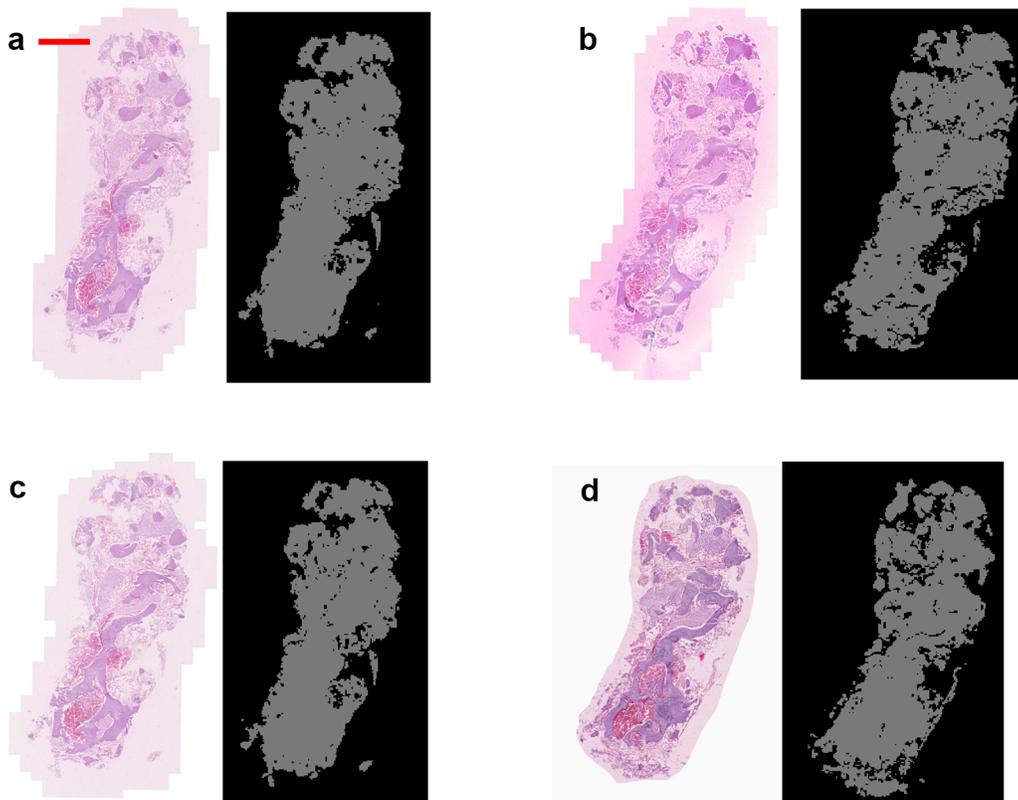


Figure 6.6: Tissue segmentation results for human bone biopsy. (a)–(d) corresponds to 4 different slides. For each group, the left image is the raw histopathology image and the right one is the segmentation result. Scale bar: 1 mm.

Bone segmentation

To compensate the tissue fraction effect, we implemented fuzzy segmentation using neural networks. The training and prediction results are as Figure 6.7 shows. Fuzzy Dice score (fDsc [145], Equation (6.1)) and pixel wise mean squared error (MSE, Equation (6.2)) are utilized as the accuracy metrics, in which N is the total pixel number and α_i and β_i are the predicted and labeled values separately. Comparing the loss curve and validation errors, MNN performs the best. However, by comparing the prediction results, the performances of CNN and MNN are similar. Both of them are better than FNN. This may be due to manual labeling bias of the dataset or lacking generalization ability of MNN.

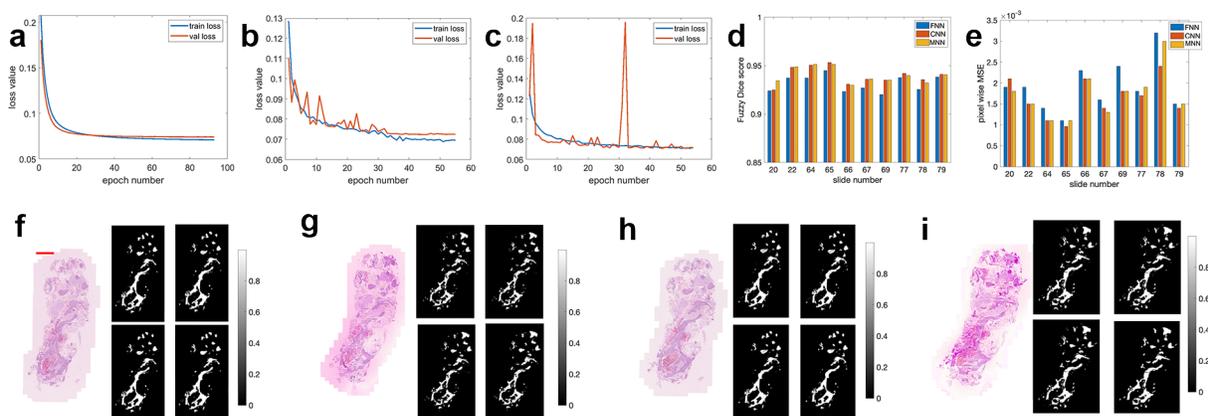


Figure 6.7: Training and prediction results of different networks. (a)–(c) are the loss curves for FNN, CNN and MNN, respectively. The validation for (a)–(c) are 0.0741, 0.0723 and 0.0709, respectively. (d) Fuzzy Dice score for the prediction results. (e) Pixel wise MSE for the prediction results. (f)–(i): Examples of bone segmentation results for human bone biopsy using different networks. For each group, the left image is the raw histopathology image, the right part: upper left: FNN, upper right: CNN, lower left: MNN and lower right: manually labeled data. Scale bar: 1 mm.

$$\text{fDsc} = \frac{2 \sum_{i=1}^N \min(\alpha_i, \beta_i)}{\sum_{i=1}^N \min(\alpha_i, \beta_i) + \sum_{i=1}^N \max(\alpha_i, \beta_i)}. \quad (6.1)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\alpha_i - \beta_i)^2. \quad (6.2)$$

Registration between the H&E and DAR images

Using the registration algorithm, several results are shown in Figure 6.8. Because of the high correlation between the bone area and the radio activity, the registration can be achieved without fiducial markers.

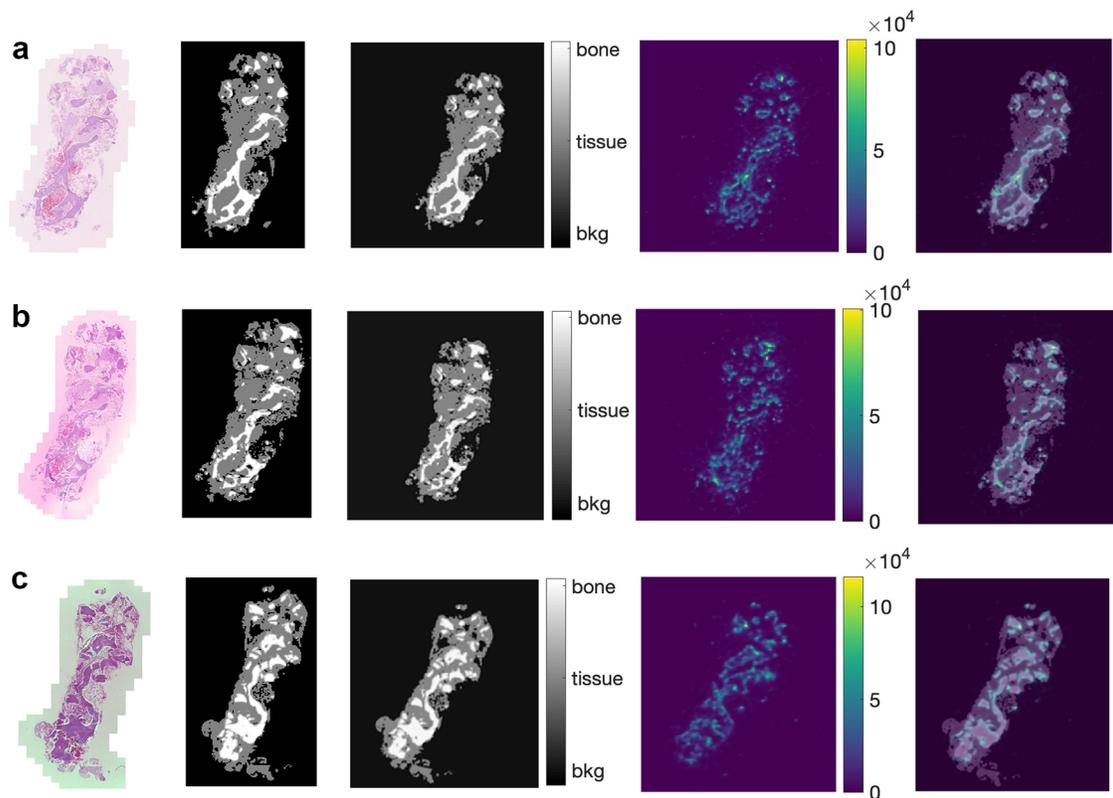


Figure 6.8: Examples of registration results between DAR and H&E images. (a)–(c) are the results from three slides; From left to right in every row: H&E image, segmented image before registration, segmented image after registration, restored DAR image and fusion of the two images.

6.3.2 IMC and H&E image registration results

Using the proposed registration algorithm, a group of co-registered H&E and IMC images are shown in Figure 6.9. Because of the high correlation between the nuclei between the two modalities, the registration performs well.

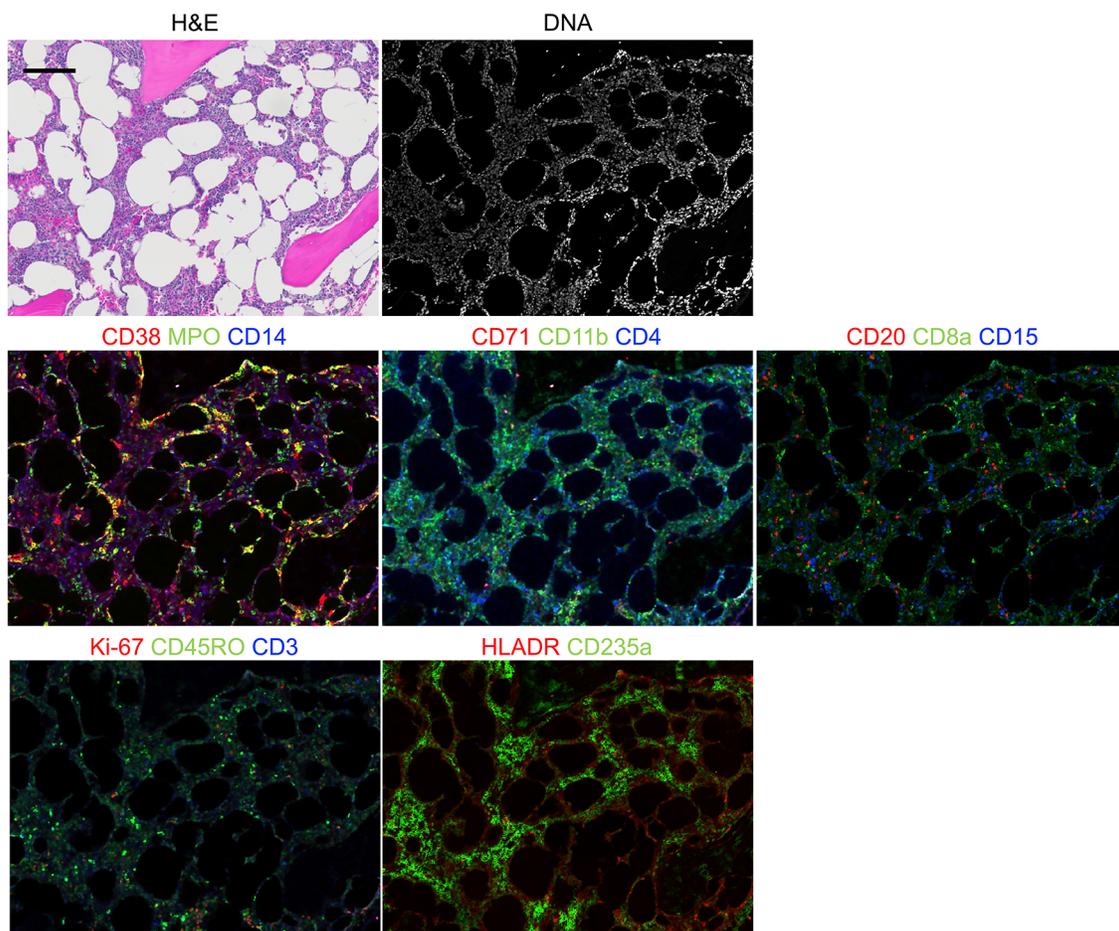


Figure 6.9: An example of registration result between IMC and H&E images. Note that the IMC images have been restored by IMC-Denoise in Chapter 4. Scale bar: 138 μm .

6.4 Discussion and Conclusion

In this chapter, we introduce computational pipelines designed for integrating multi-modality imaging data, paving the way for the development of novel theranostic approaches.

We first developed a pipeline for DAR and H&E image registration on human bone biopsy treated with $^{223}\text{RaCl}_2$. For these two modalities, we splitted the work into three parts: H&E image segmentation, DAR image restoration and image registration of the two modalities. For accurate image segmentation, K-means clustering and neural networks have been developed separately for whole tissue and bone area. Notably, a fuzzy segmentation approach has been applied for bone segmentation to account for tissue fraction effect. Due to the blur and noise effects generated in DAR imaging process, PG-PEM, developed in Chapter 3 , has been used to restore DAR images. For image registration, the histopathological and DAR images were co-registered based on their mutual information without the use of fiducial markers, including initial and fine-tune steps. Based on the results, small scale radiopharmaceutical distribution can be estimated [39].

We further developed a pipeline for IMC and H&E image registration on human bone marrow data. Here, the two modalities are co-registered by the segmented cell nuclei of the H&E image and DNA channel from IMC. Due to the presence of hot pixels and shot noise, all IMC images are restored by IMC-Denoise developed in Chapter 4. The co-registered results indicate that the pipeline performs well based on the mutual information between the two modalities. With the H&E image as reference, the relationship between dose, cell phenotypes and communities can be analyzed after registration.

Chapter 7

Conclusions and Future Work

7.1 Summary

In this dissertation, we have developed a pipeline that address a prominent challenge of theranostic medicine – on-tissue dose and microenvironment analysis. Specifially, DAR has been applied for activity distribution and dose measurement, and IMC for microenvironmental characterization. Several algorithms and frameworks have been developed to improve the data quality and downstream analysis for these two modalities. Finally, we have proposed data integration approaches with H&E, DAR and IMC images for tissue assessment.

In Chapter 3, we have developed the PG-PEM algorithm for improved DAR image quality. By employing a physics-based image formation model for DAR and implementing a DBSCAN-based signal and background segmentation approach, our blind image restoration technique has successfully mitigated background noise and image blur in both simulated and primary image samples. Our approach has shown remarkable results across a range of

datasets, encompassing high- and low-SNR scenarios for diagnostic and therapeutic radionuclides. The versatility of this method extends its utility to both pre- and clinical sample autoradiograms. As such, it promises to significantly enhance the evaluation of radiotracers and radiotherapy agents, thereby advancing the capabilities of radiological imaging and treatment.

In Chapter 4, we have developed IMC-Denoise to account for the specific technical noise present in IMC images. In this pipeline, the DIMR algorithm is first applied to adaptively remove hot pixels. Then, the self-supervised deep learning-based DeepSNiF algorithm is proposed to improve image quality. Reference denoising algorithms were utilized to rigorously evaluate IMC-Denoise on both simulated data and multiple pathological patient datasets. Compared to other methodologies, our DIMR and DeepSNiF algorithms consistently outperformed them, showcasing superior denoising capabilities. Additionally, IMC-Denoise stands out when compared to the state-of-the-art IMC background removal methods, effectively reducing background noise while minimizing the need for subjective user input. In real-world data evaluations, IMC-Denoise demonstrated equal or superior performance to the current semi-manual methods in downstream single-cell analysis. DeepSNiF, in particular, excelled in enhancing cell clustering and annotation, as indicated by quantitative evaluations of cell phenotyping results that displayed improved sensitivity and specificity following DeepSNiF denoising.

For lymphocyte annotation, the Jaccard and F1 scores clearly highlight DeepSNiF's significant advantage in phenotyping B, CD8-positive T, and CD4-positive T cells. Furthermore, our spatial distribution and cell density correlation analysis revealed less accurate annotations when using data denoised solely by DIMR, which could lead to biased conclusions. However, data denoised by DeepSNiF corrected these distortions, facilitating more accurate downstream analysis. We anticipate that IMC-Denoise will be widely adopted in the IMC

analysis community due to its adaptability, effectiveness, and flexibility, ultimately enhancing the quality of IMC data analysis and contributing to more precise research outcomes.

In Chapter 5, we have introduced and evaluated InterSTELLAR using an open-source breast cancer IMC dataset. In InterSTELLAR, tissue cell microenvironments are represented as graphs, where nodes represent cells and edges signify cell-to-cell interactions. Graph convolutional layers extract these interaction features, and a self-attention pooling module learns cell-based contributions to clinical outcomes. InterSTELLAR surpasses traditional machine learning algorithms in accuracy and outperforms FNN for community identification. Validation studies show that InterSTELLAR effectively captures tissue microenvironment features in healthy and breast cancer tissues, particularly high-attention regions containing more tumor cells. These regions reveal distinct immune-tumor interactions. InterSTELLAR’s ability to discriminate clinically relevant aspects of tumor microenvironments is evident in heterogeneity analysis of low- and high-attention regions based on cell phenotypes. Subgroup analysis using SCP establishes a mapping between tissue microenvironment and breast cancer patient prognosis. To conclude, InterSTELLAR is a versatile GNN framework that classifies tissue types and predicts disease-relevant cell communities, enhancing the characterization of patient tissue microenvironments.

In Chapter 6, we have introduced computational pipelines for integrating multi-modality imaging data to develop new theranostic approaches. Our initial pipeline focuses on human bone biopsy treated with $^{223}\text{RaCl}_2$, involving DAR and H&E stained images. We divided the workflow into three key components: H&E image segmentation, DAR image restoration, and image registration between these two modalities. For precise image segmentation, we implemented K-means clustering and deep learning algorithm for whole tissue and bone area, respectively. To address the blurriness and noise in DAR images, we utilized the PG-PEM method described in Chapter 3 for image restoration. In terms of image registration, we

successfully aligned the histopathological and DAR images, achieving good co-registration based on mutual information, without the need for fiducial markers. This co-registration enabled us to estimate small-scale radiopharmaceutical distribution within the tissue. Subsequently, we focus on registration between IMC and H&E stained histopathological images, which was achieved using segmented cell nuclei in H&E image and DNA channel of IMC. As H&E images can serve as references for DAR and IMC registration, we have the opportunity to explore the intricate interplay between doses, cell phenotypes, and cellular communities. This analysis can yield valuable insights, contributing to the advancement of theranostic developments.

7.2 Future Work

In this section, we provide a summary of the key steps moving forward with elaboration of the work in Chapters 3–6.

7.2.1 Validating PG-PEM with DAR from a silicon strip detector autoradiography system

Other than the most widely used phosphor-based DAR imaging system, some other DAR imaging systems have been developed in recently years. In specific, we are curious whether our PG-PEM algorithm can be applied to restore DAR images from a silicon strip detector autoradiography system [146, 147]. We will modify PG-PEM based on the imaging model of the silicon strip detector autoradiography system, then apply the modified PG-PEM algorithm on the new DAR images. Qualitative and quantitative assessments will be used for the restored images.

7.2.2 Advancing IMC-Denoise for larger hot pixel clusters and better image quality

As discussed in the Discussion Section of Chapter 4, IMC-Denoise does come with certain limitations. One such limitation is its inability to effectively remove large hot pixel clusters, as the DIMR component struggles to distinguish these sizable outlier areas from genuine signal. Additionally, the self-supervised DeepSNiF algorithm may not attain the same level of accuracy as supervised denoising methods, primarily because it lacks access to ground truths for training and reference. In the future work, we will modify DIMR for larger hot pixel clusters removal. Besides, we will also modify the network structure, loss function and training strategy of DeepSNiF, so that its performance can be enhanced and even reach the accuracy of supervised learning-based denoising.

7.2.3 Validating InterSTELLAR with larger and other multiplexed imaging datasets

As a supervised learning framework, InterSTELLAR necessitates a substantial volume of training data to ensure the robust generalization of the model. It is imperative to exercise meticulous care in minimizing batch effects associated with marker staining to avert overfitting and safeguard against the degradation of prediction performance. Although our current study draws upon a substantial, up-to-date cohort of patient tissue data, it is important to acknowledge that the dataset's size remains somewhat limited for the comprehensive validation of the intriguing findings suggested by SCP analysis. Furthermore, it's worth noting that our proposed framework is not limited to a specific imaging modality or disease. It can

be readily applied to other highly multiplexed imaging techniques and diverse medical conditions. As a consequence, InterSTELLAR needs to be further validated by larger datasets as well as other multiplexed imaging data, such as MIBI, *t*-CyCIF and CODEX.

7.2.4 Validating and advancing data integration framework for DAR and IMC

In Chapter 6, we have proposed a pipeline to integrate IMC and DAR data for downstream analysis. However, it is challenging and time-consuming to collect enough pre- and clinical data. As a consequence, we will work on validating the proposed DAR and IMC integration framework once enough data is collected.

After data integration framework is validated, we propose to use small scale monte carlo simulations of alpha and beta particle disintegrations for dose assessment, using MIRDCell simulations. Next, we will analyze whether any dose distributions and specific microenvironment are correlated, and what are the impacts of alpha therapy to tissue microenvironment. Based on the pre-clinical results, we will ask and optimize the overall alpha therapy process.

References

- [1] Eric G Solon. “Autoradiography: high-resolution molecular imaging in pharmaceutical discovery and development.” In: *Expert opinion on drug discovery* 2.4 (2007), pp. 503–514.
- [2] Peter Johnström, Joseph L Bird, and Anthony P Davenport. “Quantitative phosphor imaging autoradiography of radioligands for positron emission tomography.” In: *Receptor Binding Techniques*. Springer, 2012, pp. 205–220.
- [3] Tom Bäck and Lars Jacobsson. “The α -camera: a quantitative digital autoradiography technique using a charge-coupled device for ex vivo high-resolution bioimaging of α -particles.” In: *Journal of Nuclear Medicine* 51.10 (2010), pp. 1616–1623.
- [4] Nane Griem-Krey, Anders Bue Klein, Matthias Herth, and Petrine Wellendorph. “Autoradiography as a simple and powerful method for visualization and characterization of pharmacological targets.” In: *JoVE (Journal of Visualized Experiments)* 145 (2019), e58879.
- [5] Charlotte Giesen et al. “Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry.” In: *Nature methods* 11.4 (2014), pp. 417–422.
- [6] Heeva Baharlou, Nicolas P Canete, Anthony L Cunningham, Andrew N Harman, and Ellis Patrick. “Mass cytometry imaging for the study of human diseases applications and data analysis strategies.” In: *Frontiers in immunology* 10 (2019), p. 2657.
- [7] Hartland W Jackson, Jana R Fischer, Vito RT Zanotelli, et al. “The single-cell pathology landscape of breast cancer.” In: *Nature* 578 (2020), pp. 615–620.
- [8] Thomas A Holly et al. *Single photon-emission computed tomography*. 2010.
- [9] Dale L Bailey, Michael N Maisey, David W Townsend, and Peter E Valk. *Positron emission tomography*. Vol. 2. Springer, 2005.
- [10] Øyvind S Bruland, Sten Nilsson, Darrell R Fisher, and Roy H Larsen. “High-linear energy transfer irradiation targeted to skeletal metastases by the α -emitter ^{223}Ra : adjuvant or alternative to conventional modalities?” In: *Clinical cancer research* 12.20 (2006), 6250s–6257s.

- [11] Gjermund Henriksen, Knut Breistøl, Øyvind S Bruland, Øystein Fodstad, and Roy H Larsen. “Significant antitumor effect from bone-seeking, α -particle-emitting ^{223}Ra demonstrated in an experimental skeletal metastases model.” In: *Cancer research* 62.11 (2002), pp. 3120–3125.
- [12] Amir Sabet et al. “Long-term hematotoxicity after peptide receptor radionuclide therapy with ^{177}Lu -octreotate.” In: *Journal of Nuclear Medicine* 54.11 (2013), pp. 1857–1861.
- [13] Kambiz Rahbar et al. “German multicenter study investigating ^{177}Lu -PSMA-617 radioligand therapy in advanced prostate cancer patients.” In: *Journal of Nuclear Medicine* 58.1 (2017), pp. 85–90.
- [14] Abbie Hasson et al. “Radiochemical quality control methods for radium-223 and thorium-227 radiotherapies.” In: *Cancer Biotherapy & Radiopharmaceuticals* 38.1 (2023), pp. 15–25.
- [15] Kwamena E Baidoo, Kwon Yong, and Martin W Brechbiel. “Molecular pathways: targeted α -particle radiation therapy.” In: *Clinical cancer research* 19.3 (2013), pp. 530–537.
- [16] Diane Abou, Nadia Benabdallah, Wen Jiang, Lu Peng, Hanwen Zhang, Alexandria Villmer, Mark S Longtine, and Daniel LJ Thorek. “Prostate Cancer Theranostics-An Overview.” In: *Frontiers in Oncology* 10 (2020), p. 884.
- [17] George Sgouros, Robert F Hobbs, and Hong Song. “Modelling and dosimetry for alpha-particle therapy.” In: *Current radiopharmaceuticals* 4.3 (2011), pp. 261–265.
- [18] Robert F Hobbs, Hong Song, Christopher J Watchman, Wesley E Bolch, Anne-Kirsti Aksnes, Thomas Ramdahl, Glenn D Flux, and George Sgouros. “A bone marrow toxicity model for ^{223}Ra alpha-emitter radiopharmaceutical therapy.” In: *Physics in Medicine & Biology* 57.10 (2012), p. 3207.
- [19] Minoru Sonoda, Masao Takano, Junji Miyahara, and Hisatoyo Kato. “Computed radiography utilizing scanning laser stimulated luminescence.” In: *Radiology* 148.3 (1983), pp. 833–838.
- [20] Yvan Saeys, Sofie Van Gassen, and Bart N Lambrecht. “Computational flow cytometry: helping to make sense of high-dimensional immunology data.” In: *Nature Reviews Immunology* 16.7 (2016), pp. 449–462.
- [21] Aysun Adan, Günel Alizada, Yamur Kiraz, Yusuf Baran, and Ayten Nalbant. “Flow cytometry: basic principles and applications.” In: *Critical reviews in biotechnology* 37.2 (2017), pp. 163–176.
- [22] John W Hickey et al. “Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging.” In: *Nature methods* 19.3 (2022), pp. 284–295.

- [23] Jia-Ren Lin, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin M Shah, Sandro Santagata, and Peter K Sorger. “Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes.” In: *Elife* 7 (2018).
- [24] Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P Nolan. “Deep profiling of mouse splenic architecture with CODEX multiplexed imaging.” In: *Cell* 174.4 (2018), pp. 968–981.
- [25] Michael Angelo et al. “Multiplexed ion beam imaging of human breast tumors.” In: *Nature medicine* 20.4 (2014), pp. 436–442.
- [26] Leeat Keren et al. “A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging.” In: *Cell* 174 (2018), pp. 1373–1387.
- [27] Yue J Wang et al. “Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes.” In: *Cell metabolism* 29.3 (2019), pp. 769–783.
- [28] Nicolas Damond et al. “A map of human type 1 diabetes progression by imaging mass cytometry.” In: *Cell metabolism* 29.3 (2019), pp. 755–768.
- [29] H Raza Ali et al. “Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer.” In: *Nature Cancer* 1.2 (2020), pp. 163–175.
- [30] André F Rendeiro, Hiranmayi Ravichandran, Yaron Bram, et al. “The spatial landscape of lung pathology during COVID-19 progression.” In: *Nature* 593 (2021), pp. 564–569.
- [31] Minghui Wu et al. “Single-cell analysis of the human pancreas in type 2 diabetes using multi-spectral imaging mass cytometry.” In: *Cell reports* 37.5 (2021), p. 109919.
- [32] Dan Moldoveanu et al. “Spatially mapping the immune landscape of melanoma using imaging mass cytometry.” In: *Science Immunology* 7.70 (2022), eabi5072.
- [33] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *Kdd*. Vol. 96. 1996, pp. 226–231.
- [34] Peng Lu et al. “Blind Image Restoration Enhances Digital Autoradiographic Imaging of Radiopharmaceutical Tissue Distribution.” In: *Journal of Nuclear Medicine* (2021).
- [35] Peng Lu et al. “IMC-Denoise: a content aware denoising pipeline to enhance Imaging Mass Cytometry.” In: *Nature communications* 14.1 (2023), p. 1601.
- [36] Peng Lu, Karolyn A Oetjen, Stephen T Oh, and Daniel LJ Thorek. “Interpretable spatial cell learning enhances the characterization of patient tissue microenvironments with highly multiplexed imaging data.” In: *bioRxiv* (2023).

- [37] Peng Lu, Karolyn A Oetjen, and Daniel LJ Thorek. “Interpretable graph convolutional network enables triple negative breast cancer detection in imaging mass cytometry.” In: *Medical Imaging 2023: Digital and Computational Pathology*. Vol. 12471. SPIE. 2023, pp. 72–79.
- [38] Nadia Benabdallah et al. “Beyond average: α -particle distribution and dose heterogeneity in bone metastatic prostate cancer.” In: *Journal of Nuclear Medicine* (2023). DOI: 10.2967/jnumed.123.266571. eprint: <https://jnm.snmjournals.org/content/early/2023/12/14/jnumed.123.266571.full.pdf>.
- [39] Nadia Benabdallah, Peng Lu, Jeff Michalski, Brian Baumann, Russell Pachynski, Jack Jennings, Robert Hobbs, and Daniel Thorek. “Radium-223 Treated Primary Patient Bone Biopsy Analysis: Macro to Microscale Analyses and Dosimetry.” In: *Journal of Nuclear Medicine* 61.supplement 1 (2020), pp. 531–531. eprint: <https://jnm.snmjournals.org/content>.
- [40] Peng Lu et al. “Automated small-scale radiopharmaceutical distribution analysis for patient bone biopsy based on histopathological and ²²³Ra autoradiographic images.” In: *Journal of Nuclear Medicine* 61.supplement 1 (2020), pp. 529–529. eprint: <https://jnm.snmjournals.org/content>.
- [41] Charles Boncelet. “Image noise models.” In: *The Essential Guide to Image Processing*. Elsevier, 2009, pp. 143–167.
- [42] JA Goyette, GD Lapin, MG Kang, and Aggelos K Katsaggelos. “Regularized iterative image restoration algorithms applied to autoradiography.” In: *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Publ by IEEE. 1993, pp. 490–491.
- [43] Mutian Zhang, Qing Chen, Xiao-Feng Li, Joseph O’Donoghue, Shutian Ruan, Pat Zanzonico, C Clifton Ling, and John L Humm. “Image deconvolution in digital autoradiography: A preliminary study.” In: *Medical physics* 35.2 (2008), pp. 522–530.
- [44] Anna Jezierska, J-C Pesquet, Hugues Talbot, and Caroline Chaux. “Iterative Poisson-Gaussian noise parametric estimation for blind image denoising.” In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, pp. 2819–2823.
- [45] Emilie Chouzenoux, Anna Jezierska, Jean-Christophe Pesquet, and Hugues Talbot. “A Convex Approach for Image Restoration with Exact Poisson–Gaussian Likelihood.” In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2662–2682.
- [46] Alessandro Foi, Mejdî Trimeche, Vladimir Katkovnik, and Karen Egiazarian. “Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data.” In: *IEEE Transactions on Image Processing* 17.10 (2008), pp. 1737–1754.
- [47] Biagio Mandracchia, Xuanwen Hua, Changliang Guo, Jeonghwan Son, Tara Urner, and Shu Jia. “Fast and accurate sCMOS noise correction for fluorescence microscopy.” In: *Nature communications* 11.1 (2020), pp. 1–12.

- [48] DA Fish, AM Brinicombe, ER Pike, and JG Walker. “Blind deconvolution by means of the Richardson–Lucy algorithm.” In: *JOSA A* 12.1 (1995), pp. 58–65.
- [49] Tal Kenig, Zvi Kam, and Arie Feuer. “Blind image deconvolution using machine learning for three-dimensional microscopy.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (2010), pp. 2191–2204.
- [50] Stamatios Lefkimmiatis, Aurélien Bourquard, and Michael Unser. “Hessian-based norm regularization for image restoration with biomedical applications.” In: *IEEE Transactions on Image Processing* 21.3 (2011), pp. 983–995.
- [51] Margret Keuper, Thorsten Schmidt, Maja Temerinac-Ott, Jan Padeken, Patrick Heun, Olaf Ronneberger, and Thomas Brox. “Blind deconvolution of widefield fluorescence microscopic data by regularization of the optical transfer function (OTF).” In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 2179–2186.
- [52] Xiaoshuai Huang et al. “Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy.” In: *Nature biotechnology* 36.5 (2018), p. 451.
- [53] Hayato Ikoma, Michael Broxton, Takamasa Kudo, and Gordon Wetzstein. “A convex 3D deconvolution algorithm for low photon count fluorescence imaging.” In: *Scientific reports* 8.1 (2018), pp. 1–12.
- [54] Sami Koho, Giorgio Tortarolo, Marco Castello, Takahiro Deguchi, Alberto Diaspro, and Giuseppe Vicidomini. “Fourier ring correlation simplifies image restoration in fluorescence microscopy.” In: *Nature communications* 10.1 (2019), pp. 1–9.
- [55] Laura Kuett, Raúl Catena, Alaz Özcan, Alex Plüss, Peter Schraml, Holger Moch, Natalie de Souza, and Bernd Bodenmiller. “Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment.” In: *Nature Cancer* 3 (2022), pp. 122–133.
- [56] Stéphane Chevrier, Helena L Crowell, Vito RT Zanutelli, Stefanie Engler, Mark D Robinson, and Bernd Bodenmiller. “Compensation of signal spillover in suspension and imaging mass cytometry.” In: *Cell Systems* 6.5 (2018), pp. 612–620.
- [57] Alex Baranski, Idan Milo, Shirley Greenbaum, John-Paul Oliveria, Dunja Mrdjen, Michael Angelo, and Leeat Keren. “MAUI (MBI Analysis User Interface) An image processing pipeline for Multiplexed Mass Based Imaging.” In: *PLOS Computational Biology* 17.4 (2021), e1008887.
- [58] VRT Zanutelli and B Bodenmiller. “IMC Segmentation Pipeline: a pixel classification based multiplexed image segmentation pipeline.” In: *Zenodo* <https://doi.org/10.5281/zenodo.3841960> (2017).
- [59] Marieke E Ijsselsteijn, Antonios Somarakis, Boudewijn PF Lelieveldt, Thomas Höllt, and Noel FCC de Miranda. “Semi-automated background removal limits data loss and normalizes imaging mass cytometry data.” In: *Cytometry Part A* (2021).

- [60] Christian M Schürch et al. “Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front.” In: *Cell* 182 (2020), pp. 1341–1359.
- [61] Jie Zhou et al. “Graph neural networks: A review of methods and applications.” In: *AI Open* 1 (2020), pp. 57–81.
- [62] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. “A comprehensive survey on graph neural networks.” In: *IEEE Trans. Neural Netw. Learn. Syst.* 32 (2020), pp. 4–24.
- [63] Paula Martin-Gonzalez, Mireia Crispin-Ortuzar, and Florian Markowetz. “Predictive modelling of highly multiplexed tumour tissue images by graph neural networks.” In: *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*. Springer, 2021, pp. 98–107.
- [64] Christopher Innocenti et al. “An Unsupervised Graph Embeddings Approach to Multiplex Immunofluorescence Image Exploration.” In: *bioRxiv* (2021).
- [65] David S Fischer, Anna C Schaar, and Fabian J Theis. “Modeling intercellular communication in tissues using spatial graphs of cells.” In: *Nat. Biotechnol.* (2022), pp. 1–5.
- [66] Jian Hu, Xiangjie Li, Kyle Coleman, et al. “SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network.” In: *Nat. methods* 18 (2021), pp. 1342–1351.
- [67] Junbum Kim, Samir Rustam, Juan Miguel Mosquera, Scott H Randell, Renat Shaykhiev, André F Rendeiro, and Olivier Elemento. “Unsupervised discovery of tissue architecture in multiplexed imaging.” In: *Nat. Methods* 19 (2022), pp. 1653–1661.
- [68] Maria Brbi, Kaidi Cao, John W Hickey, Yuqi Tan, Michael P Snyder, Garry P Nolan, and Jure Leskovec. “Annotation of spatially resolved single-cell data with STELLAR.” In: *Nat. Methods* 19 (2022), pp. 1411–1418.
- [69] Zhenqin Wu et al. “Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens.” In: *Nat. Biomed. Eng* 6 (2022), pp. 1435–1448.
- [70] Markku Makitalo and Alessandro Foi. “Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise.” In: *IEEE transactions on image processing* 22.1 (2012), pp. 91–103.
- [71] Ismail Bin Mohamad and Dauda Usman. “Standardization and its effects on K-means clustering algorithm.” In: *Research Journal of Applied Sciences, Engineering and Technology* 6.17 (2013), pp. 3299–3303.

- [72] Elad Shaked, Sudipto Dolui, and Oleg V Michailovich. “Regularized Richardson-Lucy algorithm for reconstruction of Poissonian medical images.” In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. 2011, pp. 1754–1757.
- [73] Christoph Brune, Alex Sawatzky, and Martin Burger. “Bregman-EM-TV methods with application to optical nanoscopy.” In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 2009, pp. 235–246.
- [74] Elena Resmerita and Robert S Anderssen. “Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems.” In: *Mathematical Methods in the Applied Sciences* 30.13 (2007), pp. 1527–1544.
- [75] Patrizio Campisi and Karen Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2017.
- [76] Nicolas Bousson, C Cheze Le Rest, Mathieu Hatt, and Dimitris Visvikis. “Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging.” In: *European journal of nuclear medicine and molecular imaging* 36.7 (2009), pp. 1064–1075.
- [77] Jean-Luc Starck, E Pantin, and F Murtagh. “Deconvolution in astronomy: A review.” In: *Publications of the Astronomical Society of the Pacific* 114.800 (2002), p. 1051.
- [78] Peng Lu, Changku Sun, Bin Liu, and Peng Wang. “Accurate and robust calibration method based on pattern geometric constraints for fringe projection profilometry.” In: *Applied optics* 56.4 (2017), pp. 784–794.
- [79] Peng Lu, Changku Sun, and Peng Wang. “Fringe projection phase-to-height mapping model and its calibration method.” In: *Acta Optica Sinica* 38.2 (2018), p. 0212004.
- [80] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. “Image quality assessment: from error visibility to structural similarity.” In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.
- [81] Adrien Charles Descloux, Kristin Stefanie Grussmayer, and Aleksandra Radenovic. “Parameter-free image resolution estimation based on decorrelation analysis.” In: *Nature Methods* 16.ARTICLE (2019), pp. 918–924.
- [82] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis.” In: *Nature methods* 9.7 (2012), pp. 676–682.
- [83] Diane S Abou, David Ulmert, Michele Doucet, Robert F Hobbs, Ryan C Riddle, and Daniel LJ Thorek. “Whole-body and microenvironmental localization of radium-223 in naive and mouse models of prostate cancer metastasis.” In: *Journal of the National Cancer Institute* 108.5 (2016), djv380.
- [84] Eric G Solon and Lori Kraus. “Quantitative whole-body autoradiography in the pharmaceutical industry: Survey results on study design, methods, and regulatory compliance.” In: *Journal of pharmacological and toxicological methods* 46.2 (2001), pp. 73–81.

- [85] Eric G Solon. “Use of radioactive compounds and autoradiography to determine drug tissue distribution.” In: *Chemical research in toxicology* 25.3 (2012), pp. 543–555.
- [86] Andrew B McEwen, Claire M Henson, and Stuart G Wood. “Quantitative whole-body autoradiography, LC–MS/MS and MALDI for drug-distribution studies in biological samples: the ultimate matrix trilogy.” In: *Bioanalysis* 6.3 (2014), pp. 377–391.
- [87] Michael R McDevitt et al. “Feed-forward alpha particle radiotherapy ablates androgen receptor-addicted prostate cancer.” In: *Nature communications* 9.1 (2018), pp. 1–11.
- [88] Andrew C Yang et al. “Physiological blood–brain transport is impaired with age by a shift in transcytosis.” In: *Nature* 583.7816 (2020), pp. 425–430.
- [89] Hanwen Zhang et al. “[18F]-Labeled PARP-1 PET imaging of PSMA targeted alpha particle radiotherapy response.” In: *Scientific reports* 12.1 (2022), p. 13034.
- [90] Ruomei Yan and Ling Shao. “Blind image blur estimation via deep learning.” In: *IEEE Transactions on Image Processing* 25.4 (2016), pp. 1910–1921.
- [91] Martin Weigert et al. “Content-aware image restoration: pushing the limits of fluorescence microscopy.” In: *Nature methods* 15.12 (2018), pp. 1090–1097.
- [92] Min Guo et al. “Rapid image deconvolution and multiview fusion for optical microscopy.” In: *Nature Biotechnology* (2020), pp. 1–10.
- [93] Tingting Wu, Peng Lu, Md Ashequr Rahman, Xiao Li, and Matthew D Lew. “Deep-SMOLM: deep learning resolves the 3D orientations and 2D positions of overlapping single molecules with optimal nanoscale resolution.” In: *Optics Express* 30.20 (2022), pp. 36761–36773.
- [94] Antoni Buades, Bartomeu Coll, and J-M Morel. “A non-local algorithm for image denoising.” In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Vol. 2. IEEE. 2005, pp. 60–65.
- [95] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. “Image denoising by sparse 3-D transform-domain collaborative filtering.” In: *IEEE Transactions on image processing* 16.8 (2007), pp. 2080–2095.
- [96] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. “Noise2void-learning denoising from single noisy images.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2129–2137.
- [97] Noah F Greenwald et al. “Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning.” In: *Nature biotechnology* 40.4 (2022), pp. 555–565.
- [98] Denis Schapiro et al. “MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging.” In: *Nature methods* 19.3 (2022), pp. 311–315.
- [99] Francis J Anscombe. “The transformation of Poisson, binomial and negative-binomial data.” In: *Biometrika* 35.3/4 (1948), pp. 246–254.

- [100] Shaul K Bar-Lev and Peter Enis. “On the classical choice of variance stabilizing transformations and an application for a Poisson variate.” In: *Biometrika* 75.4 (1988), pp. 803–804.
- [101] Marshall F Tappen Bryan C Russell and William T Freeman. “Exploiting the sparse derivative prior for super-resolution and image demosaicing.” In: *Proceedings of the Third International Workshop Statistical and Computational Theories of Vision*. 2003, pp. 1–28.
- [102] Walter Rudin et al. *Principles of mathematical analysis*. Vol. 3. McGraw-hill New York, 1964.
- [103] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
- [104] Markku Makitalo and Alessandro Foi. “Optimal inversion of the Anscombe transformation in low-count Poisson image denoising.” In: *IEEE transactions on Image Processing* 20.1 (2010), pp. 99–109.
- [105] Lorenzo Finesso and Peter Spreij. “Nonnegative matrix factorization and I-divergence alternating minimization.” In: *Linear Algebra and its Applications* 416.2-3 (2006), pp. 270–287.
- [106] Joshua Batson and Loic Royer. “Noise2self: Blind denoising by self-supervision.” In: *International Conference on Machine Learning*. PMLR. 2019, pp. 524–533.
- [107] Weisong Zhao et al. “Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy.” In: *Nature biotechnology* (2021), pp. 1–12.
- [108] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation.” In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [109] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Identity mappings in deep residual networks.” In: *European conference on computer vision*. Springer. 2016, pp. 630–645.
- [110] François Chollet. *keras*. <https://github.com/fchollet/keras>. 2015.
- [111] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization.” In: *arXiv preprint arXiv:1412.6980* (2014).
- [112] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. “Empirical evaluation of rectified activations in convolutional network.” In: *arXiv preprint arXiv:1505.00853* (2015).
- [113] Rumana Rashid et al. “Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer.” In: *Scientific data* 6.1 (2019), pp. 1–10.
- [114] Roman Garnett, Timothy Huegerich, Charles Chui, and Wenjie He. “A universal noise removal algorithm with an impulse detector.” In: *IEEE Transactions on image processing* 14.11 (2005), pp. 1747–1754.

- [115] Yiqiu Dong, Raymond H Chan, and Shufang Xu. “A detection statistic for random-valued impulse noise.” In: *IEEE transactions on image processing* 16.4 (2007), pp. 1112–1120.
- [116] Christoph Sommer, Christoph Straehle, Ullrich Koethe, and Fred A Hamprecht. “Ilastik: Interactive learning and segmentation toolkit.” In: *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE. 2011, pp. 230–233.
- [117] Lee Kametsky et al. “Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software.” In: *Bioinformatics* 27.8 (2011), pp. 1179–1180.
- [118] Denis Schapiro et al. “histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data.” In: *Nature methods* 14.9 (2017), p. 873.
- [119] Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python.” In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.
- [120] David L Davies and Donald W Bouldin. “A cluster separation measure.” In: *IEEE transactions on pattern analysis and machine intelligence* (1979), pp. 224–227.
- [121] Jacob H Levine et al. “Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis.” In: *Cell* 162.1 (2015), pp. 184–197.
- [122] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. “From Louvain to Leiden: guaranteeing well-connected communities.” In: *Scientific reports* 9.1 (2019), pp. 1–12.
- [123] John W Hickey, Yuqi Tan, Garry P Nolan, and Yury Goltsev. “Strategies for accurate cell type identification in CODEX multiplexed imaging data.” In: *Frontiers in Immunology* (2021), p. 3317.
- [124] George C Linderman, Manas Rachh, Jeremy G Hoskins, Stefan Steinerberger, and Yuval Kluger. “Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data.” In: *Nature methods* 16.3 (2019), pp. 243–245.
- [125] Andrew Chow et al. “CD169+ macrophages provide a niche promoting erythropoiesis under homeostasis and stress.” In: *Nature medicine* 19.4 (2013), pp. 429–436.
- [126] Patrick Williams et al. “The distribution of T-cell subsets and the expression of immune checkpoint receptors and ligands in patients with newly diagnosed and relapsed acute myeloid leukemia.” In: *Cancer* 125.9 (2019), pp. 1470–1481.
- [127] Thomas N Kipf and Max Welling. “Semi-supervised classification with graph convolutional networks.” In: *arXiv preprint arXiv:1609.02907* (2016).
- [128] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. “Layer normalization.” In: *arXiv preprint arXiv:1607.06450* (2016).
- [129] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. “Self-normalizing neural networks.” In: *Adv. Neural Inf. Process. Syst.* 30 (2017).

- [130] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, et al. “Data-efficient and weakly supervised computational pathology on whole-slide images.” In: *Nat. Biomed. Eng.* 5 (2021), pp. 555–570.
- [131] Leonard Berrada, Andrew Zisserman, and M Pawan Kumar. “Smooth loss functions for deep top-k classification.” In: *arXiv preprint arXiv:1802.07595* (2018).
- [132] Adam Paszke et al. “Pytorch: An imperative style, high-performance deep learning library.” In: *Adv. Neural Inf. Process. Syst.* 32 (2019).
- [133] Matthias Fey and Jan Eric Lenssen. “Fast graph representation learning with PyTorch Geometric.” In: *arXiv preprint arXiv:1903.02428* (2019).
- [134] Hongfei Li, Catherine A Calder, and Noel Cressie. “Beyond Moran’s I: testing for spatial dependence based on the spatial autoregressive model.” In: *Geogr. Anal.* 39 (2007), pp. 357–375.
- [135] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.
- [136] Stuart Berg et al. “Ilastik: interactive machine learning for (bio) image analysis.” In: *Nat. Methods* 16 (2019), pp. 1226–1232.
- [137] S Loi et al. “Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial.” In: *Ann. Oncol.* 25 (2014), pp. 1544–1550.
- [138] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. “SCANPY: large-scale single-cell gene expression data analysis.” In: *Genome Biol.* 19 (2018), pp. 1–5.
- [139] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. “Color transfer between images.” In: *IEEE Computer graphics and applications* 21.5 (2001), pp. 34–41.
- [140] HP Ng, SH Ong, KWC Foong, PS Goh, and WL Nowinski. “Medical image segmentation using k-means clustering and improved watershed algorithm.” In: *2006 IEEE southwest symposium on image analysis and interpretation*. IEEE. 2006, pp. 61–65.
- [141] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. “Face description with local binary patterns: Application to face recognition.” In: *IEEE transactions on pattern analysis and machine intelligence* 28.12 (2006), pp. 2037–2041.
- [142] Robert M Haralick, Karthikeyan Shanmugam, and Its’ Hak Dinstein. “Textural features for image classification.” In: *IEEE Transactions on systems, man, and cybernetics* 6 (1973), pp. 610–621.
- [143] Ziping Liu, Joyce C Mhlanga, Richard Laforest, Paul-Robert Derenoncourt, Barry A Siegel, and Abhinav K Jha. “A Bayesian approach to tissue-fraction estimation for oncological PET segmentation.” In: *Physics in Medicine & Biology* 66.12 (2021), p. 124002.

- [144] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. “Mutual-information-based registration of medical images: a survey.” In: *IEEE transactions on medical imaging* 22.8 (2003), pp. 986–1004.
- [145] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B Blaschko. “Optimizing the Dice score and Jaccard index for medical image segmentation: Theory and practice.” In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 92–100.
- [146] Anders Örbom, Sophie E Eriksson, Erika Elgström, Tomas Ohlsson, Rune Nilsson, Jan Tennvall, and Sven-Erik Strand. “The intratumoral distribution of radiolabeled ¹⁷⁷Lu-BR96 monoclonal antibodies changes in relation to tumor histology over time in a syngeneic rat colon carcinoma model.” In: *Journal of Nuclear Medicine* 54.8 (2013), pp. 1404–1410.
- [147] Anders Örbom et al. “Characterization of a double-sided silicon strip detector autoradiography system.” In: *Medical Physics* 42.2 (2015), pp. 575–584.

Appendix A

Supplementary Tables for Chapter 4

Table A.1: List of cell markers used for Collagen III-labeled tissues in Figures 4.12a and 4.16a

Isotope	Metal	Epitope	Clone	Source	Catalog #	Dilution
141	Pr	CD235 *	HIR2	Fluidigm	3141001B	1:200
142	Nd	MPO *	polyclonal	Dako	A0398	1:800
144	Nd	CD14 *	EPR3653	Fluidigm	3144025D	1:800
145	Nd	CD117 *	YR145	Abcam	ab216450	1:100
147	Sm	CD163 *	EDHu-1	Fluidigm	3147021D	1:100
148	Nd	CD71 *	MRQ-48	eBiosciences	14-0718-93	1:200
149	Sm	CD11b *	EPR1344	Fluidigm	3149028D	1:150
151	Eu	CD31 *	EPR3094	Fluidigm	3151025D	1:100
152	Sm	CD34 *	QBend/10	ThermoFisher	MA1-10202	1:400
153	Eu	pSTAT5	47	BD	custom quote	1:100
154	Sm	TNF α	TNF706 + P/T2	Abcam	ab212899	1:150

Table A.1 continued from previous page

155	Gd	IL8	807	Abcam	custom quote	1:200
156	Gd	CD4 *	EPR6855	Fluidigm	3156033D	1:500
157	Gd	IL6	1936	R&D	MAB2061	1:100
158	Gd	pSTAT3	4/P-STAT3	Fluidigm	3158030D	1:100
159	Tb	CD90	5.00E+10	Fluidigm	3159007B	1:100
160	Gd	CD61 *	2f2	Sigma	custom quote	1:400
161	Dy	CD20 *	H1	Fluidigm	3161029D	1:400
162	Dy	CD8a *	C8/144B	Fluidigm	3162034D	1:300
163	Dy	TGFb	TB21	Invitrogen	MA1-21595	1:800
164	Dy	CD15 *	W6D3	Fluidigm	3164001B	1:150
165	Ho	pCREB	87G3	Fluidigm	3165009A	1:400
166	Er	p65 pS529	K10x	Fluidigm	3166006A	1:200
167	Er	RELA	2A12A7	ThermoFisher	33-9900	1:100
168	Er	Ki-67 *	B56	Fluidigm	3168022D	1:200
169	Tm	pIKK α/β	16A6	CST	2697BF	1:200
170	Er	CD3 *	polyclonal	Fluidigm	3170019D	1:200
172	Yb	Cleaved casp 3	5A1E	Fluidigm	3172027D	1:300
174	Yb	pERK1/2	D13.14.4E	Fluidigm	3171021D	1:100
175	Lu	pS6	N7-548	Fluidigm	3175009A	1:400
176	Yb	Histone H3 *	D1H2	Fluidigm	3176023D	1:2000
209	Bi	Collagen III *	polyclonal	Southern Biotech	1330-01	1:100
191/193	Ir	intercalator *				1:300

* Denotes markers with validated staining patterns on this tissue

Table A.2: List of cell markers for the tissue staining in Figure 4.12b

Isotope	Metal	Epitope	Clone	Source	Catalog #	Dilution
142	Nd	MPO *	polyclonal	Dako	A0398	1:400
145	Nd	CD117 *	YR145	Abcam	ab216450	1:50
150	Nd	CXCL12	79018	Novus	MAB350	1:75
153	Eu	IFNg	IFNG/466	Novus	NBP2-54394	1:25
154	Sm	TNFa	TNF706 + P/T2	Abcam	ab212899	1:75
155	Gd	IL8	807	Abcam	custom quote	1:25
157	Gd	IL6	1936	R&D	MAB2061	1:25
158	Gd	pSmad4	polyclonal	ThermoFisher	PA5-12695	1:150
159	Tb	CD169 *	SP213	Abcam	ab245735	1:100
163	Dy	CD271	EP1039Y	Abcam	ab256584	1:50
173	Yb	Collagen III *	polyclonal	Southern Biotech	1330-01	1:400
209	Bi	Collagen III *	polyclonal	Southern Biotech	1330-01	1:200
191/193	Ir	intercalator				1:200

* Denotes markers with validated staining patterns on this tissue

Table A.3: List of cell markers for the tissue staining in Figure 4.13

Isotope	Metal	Epitope	Clone	Source	Catalog #	Dilution
142	Nd	MPO *	polyclonal	Dako	A0398	1:400
143	Nd	TP53 *	DO-7	Fluidigm	3143026D	1:50
146	Nd	TP53 *	DO-7	Biolegend	345102	1:50
150	Nd	CXCL12	79018	Novus	MAB350	1:50
158	Gd	CD56 *	MRQ-42	CellMarque	custom quote	1:100
163	Dy	CD271	EP1039Y	Abcam	ab256584	1:50
167	Er	GranzymeB	EPR20129-217	Fluidigm	3167021D	1:600
173	Yb	CD56 *	MRQ-42	CellMarque	custom quote	1:100
191/193	Ir	intercalator				1:300

* Denotes markers with validated staining patterns on this tissue

Table A.4: List of cell markers used for other IMC images from the human bone marrow IMC dataset

Isotope	Metal	Epitope	Clone	Source	Catalog #	Dilution
89	Yb	Alpha-SMA *	1A4	Bio-Rad	MCA5781GA	1:100
115	In	perilipin *	D1D8	CST	9349 custom	1:50
139	La	VCAM1	EPR5047	Abcam	ab215380	1:50
141	Pr	CD38 *	EPR4106	Fluidigm	3141018D	1:50
142	Nd	MPO *	polyclonal	Dako	A0398	1:400
143	Nd	vimentin *	RV202	Fluidigm	3143029D	1:200
144	Nd	CD14 *	EPR3653	Fluidigm	3144025D	1:400
145	Nd	CD117 *	YR145	Abcam	ab216450	1:50
146	Nd	CD16	EPR16784	Fluidigm	3146020D	1:150
147	Sm	CD163 *	EDHu-1	Fluidigm	3147021D	1:100
148	Nd	CD71 *	MRQ-48	eBiosciences	14-0718-93	1:50
149	Sm	CD11b *	EPR1344	Fluidigm	3149028D	1:150
150	Nd	CXCL12	79018	Novus	MAB350	1:25
151	Eu	CD31 *	EPR3094	Fluidigm	3151025D	1:50

Table A.4 continued from previous page

152	Sm	CD34 *	QBend/10	ThermoFisher	MA1-10202	1:50
153	Eu	IFNg	IFNG/466	Novus	NBP2-54394	1:25
154	Sm	TNFa	TNF706 + P/T2	Abcam	ab212899	1:100
156	Gd	CD4 *	EPR6855	Fluidigm	3156033D	1:200
157	Gd	IL6	1936	R&D	MAB2061	1:25
158	Gd	pSmad4	polyclonal	ThermoFisher	PA5-12695	1:150
159	Tb	CD169 *	SP213	Abcam	ab245735	1:100
160	Gd	CD61 *	2f2	Sigma	custom quote	1:100
161	Dy	CD20 *	H1	Fluidigm	3161029D	1:400
162	Dy	CD8a *	C8/144B	Fluidigm	3162034D	1:300
163	Dy	CD271	EP1039Y	Abcam	ab256584	1:50
164	Dy	CD15 *	W6D3	Fluidigm	3164001B	1:150
165	Ho	pH2AX *	N1-431	Fluidigm	3165036D	1:150
166	Er	p65 pS529	K10x	Fluidigm	3166006A	1:25
167	Er	SCF	polyclonal	ThermoFisher	PA5-20746	1:25
168	Er	Ki-67 *	B56	Fluidigm	3168022D	1:100

Table A.4 continued from previous page

169	Tm	Collagen I *	polyclonal	Fluidigm	3169023D	1:2000
170	Er	CD3 *	polyclonal	Fluidigm	3170019D	1:100
171	Yb	pERK1/2	D13.14.4E	Fluidigm	3171021D	1:50
172	Yb	Cleaved casp 3 *	5A1E	Fluidigm	3172027D	1:25
173	Yb	CD45RO *	UCHL1	Fluidigm	3173016D	1:500
174	Yb	HLA-DR *	YE2/36HLK	Fluidigm	3174023D	1:100
175	Lu	CD235a *	HIR2	Fluidigm	3175029D	1:200
176	Yb	Histone H3 *	D1H2	Fluidigm	3176023D	1:2000
209	Bi	Collagen III *	polyclonal	Southern Biotech	1330-01	1:75
191/193	Ir	intercalator				1:200

Note: The tissues with headers of K, L do not have CXCL12.

* Denotes markers with validated staining patterns on this tissue.

Table A.5: Training details for the simulation datasets

#Patches	Normalized percentile	training time
12000	99.999	89 min

Table A.6: Training details for the Collagen III-labeled images in Figures 4.11, 4.12 and 4.16

Marker	#Patches	Normalized percentile	Background thresh ρ	training time
Collagen III	1992	99.9	0.55	17 min

Table A.7: Training details for the other markers-labeled images from the human bone marrow IMC dataset

Marker	#Patches	Normalized percentile	Background thresh ρ	training time
CD38	21768	99.999	0.9	160 min
MPO	20800	99.999	0.8	153 min
CD14	21784	99.999	0.9	160 min
CD71	14960	99.999	0.9	110 min
CD11b	20096	99.999	0.9	147 min
CD31	9040	99.99	0.75	67 min
CD34	15208	99.9	0.85	114 min
CD4	20832	99.9	0.9	154 min
CD169	20800	99.999	0.9	153 min
CD61	3360	99.9	0.75	27 min
CD20	12304	99.9	0.95	90 min
CD8a	19360	99.999	0.9	144 min
CD15	17456	99.99	0.9	127 min
Ki-67	18032	99.999	0.9	134 min
CD3	16728	99.99	0.9	124 min
CD45RO	11672	99.99	0.75	87 min
CD235a	21144	99.999	0.7	154 min
Histone H3	14952	99.999	0.5	108 min
DNA2	22136	99.999	0.4	161 min
Combinations of CD4, CD8a, CD3, CD14, CD11b, CD71 and CD15	131216	99.9		15.8 h

Table A.8: Training details for the images from human breast cancer IMC dataset

Marker	#Patches	Normalized percentile	Background thresh ρ	training time
CD3	12592	99.9	0.95	94 min
CD20	17592	99.9	0.95	133 min
CD45	9368	99.999	0.9	72 min
CD68	13024	99.999	0.85	97 min
c-Myc	14056	99.999	0.85	104 min
EGFR	17624	99.999	0.85	135 min
EpCAM	15064	99.999	0.8	112 min
Ki-67	10824	99.999	0.9	80 min
Rabbit IgG H L	10928	99.999	0.9	84 min
Slug	11856	99.999	0.9	90 min
Twist	14496	99.999	0.9	107 min
vWF	17928	99.999	0.95	136 min

Table A.9: Training details for the images from the human pancreatic cancer IMC dataset

Marker	#Patches	Normalized percentile	Background thresh ρ	training time
CD3	16304	99.99	0.75	120 min
CD4	14344	99.999	0.7	107 min
CD8	9792	99.9	0.77	74 min
CD11b	21120	99.99	0.3	154 min
CD14	10896	99.99	0.5	82 min
CD31	17088	99.99	0.75	126 min
CD44	17392	99.99	0.5	127 min
CD45	21984	99.99	0.7	161 min
CD45RO	18056	99.99	0.6	134 min
CD56	11168	99.99	0.5	84 min
Foxp3	9240	99.99	0.4	70 min
pS6	19440	99.99	0.2	144 min

Table A.10: Training details for the images from the IMC dataset

Marker	#Patches	Normalized percentile	Background thresh ρ	training time
CD3	5096	99.999	0.95	43 min
CD4	4120	99.999	0.95	34 min
CD8	7472	99.999	0.98	57 min
CD11b	11400	99.999	0.95	87 min
CD11c	6720	99.999	0.96	50 min
CD14	9768	99.999	0.95	74 min
CD20	13184	99.999	0.95	97 min
CD31	7272	99.999	0.99	57 min
CD45	13792	99.999	0.95	104 min
CD68	8368	99.999	0.98	64 min
CD206	10672	99.999	0.98	80 min
HLA-DR	13296	99.999	0.95	100 min

Table A.11: The estimated thresholds for positive markers

Marker	Value	Marker	Value
CD38	1.3981	CD169	2.5625
MPO	5.2272	CD20	0.6300
CD14	2.3947	CD8a	3.3559
CD71	0.9478	CD15	3.0734
CD11b	2.2740	CD3	0.9180
CD4	0.7830	CD235a	35.7895