Spring 5-11-2024

An Attention LSTM U-Net Model for Drosophila Melanogaster Heart Tube Segmentation

Xiangping Ouyang
Washington University – McKelvey School of Engineering

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering
Department of Electrical & Systems Engineering

Thesis Examination Committee:
Chao Zhou, Chair
Song Hu
Quing Zhu

An Attention LSTM U-Net Model for Drosophila Melanogaster Heart Tube Segmentation

by
Xiangping Ouyang

A thesis presented to
the McKelvey School of Engineering
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Science

May 2024
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Throughout my three-year academic journey at WashU, I have been fortunate to receive support, guidance, and encouragement from many individuals and organizations.

First, I would like to express my sincere gratitude to my thesis advisor, Professor Chao Zhou, for providing me with the opportunity to work and learn in his lab. Before I entered the Z-lab, I knew nothing about optical coherence tomography or *Drosophila melanogaster*. Researching in a field that I am not familiar with is full of thistles and thorns, but Professor Chao Zhou enlightened me and guided me into the field of biomedical optical image analysis. I am more than grateful to him for all the patience, trust, and encouragement he has given to me. I remember the morning he encouraged me to believe in myself, work harder, and learn more about the project. In a research form, he included a sentence claiming that I was expected to present on the SPIE Photonics West and draft a manuscript for a scientific journal. This opportunity provided me with an unprecedented ability to move forward in the research field. I remember how fast my heart beat when he stated I could publish a paper based on my current work. I could never have imagined having this opportunity before I joined the Z-lab. Thanks to Professor Chao Zhou for providing me with the chance to attend an international conference. I will remember the invaluable guidance and support from him forever and how he raised me into the position I now occupy.

I also want to express my heartfelt appreciation to Senyue Hao for being my student advisor for my master's research and to Doctor Elena Gracheva for her guidance in conducting academic reports. Senyue led me to the door of conducting research in biomedical image analysis. I remember the several occasions on which he left plenty of comments and helped me reorganize my poster again and again. Elena taught me a great deal about how to conduct presentations and

how to structure manuscripts in biomedical research academia. I remember the morning she drew a diagram and showed me the procedures. Thanks to Senyue Hao and Doctor Elena Gracheva for all the effort they have dedicated to training me to become a researcher.

I am grateful to Abby Matt and Fei Wang for their guidance in collecting data, imaging the fly, and examining the FlyNet system. Special thanks to Matthew Fishman for providing me with the starting points for optimizing the FlyNet model and answering all questions I have about the dataset and codes. Doing research across disciplines is challenging, and I am sincerely thankful for the help given by Abby. She showed me detailed information about the OCM dataset including useful biological background and led me into the world of cardiac research by showing me how to manually identify the heart region and many useful heart function analysis methods. She also taught me how to prepare and image the fruit fly. I also remember the busy Saturday during which we collected more images with artifacts. She and Fei helped me immensely in formulating and editing the manuscript. Fei also helped me to think of additional FlyNet applications to strengthen the contribution of the novel *Drosophila* heart segmentation network in the academic field. Whenever I face any questions, they are always there and willing to help me, from the basic biological foundations of cells to the naming and usage of specific opsins or genotypes. Thanks to Abby and Fei for their support in helping me with writing and revising my manuscript and filling the gap in my biology background.

I also want to thank my other committee members, Professor Song Hu and Professor Qing Zhu, for spending valuable time reviewing my thesis.

I would like to thank all the lab members in Z-lab, including Chao Ren, Jiawei Meng, AJ Adkins, Yitian Zhang, Weiyan Zhou, Jonathan Lin, Hongwu Liang, and Michael Gao, and former lab

members including Marcello Amaral, Kibeom Park, Jiantao Zhu, Yuxuan Wang, Siming Yin, and Yilin Li, for their unwavering support and inspiration, which I have integrated into my academic journey.

I would like to express special thanks to my lovely parents, my mother Tianyun Wang and my father Changpeng Ouyang, for their unrequited support whenever I hesitated. Their encouragement brought me to where I am and accompanied me throughout my master's journey.

I would also like to express my appreciation for the financial support that has enabled this research. This work was supported by a start-up fund from Washington University in St. Louis and the National Institutes of Health (NIH) grants R01-HL156265 and R21EB03268401 and the Clayco Foundation Innovative Research Award. These funding sources have been instrumental in the progress and success of this project.

Xiangping Ouyang

Washington University in St. Louis

May 2024

Dedicated to my parents.

ABSTRACT OF THE THESIS

An Attention LSTM U-Net Model for *Drosophila Melanogaster* Heart Tube Segmentation

by

Xiangping Ouyang

Master of Science in Computer Engineering

Washington University in St. Louis, 2024

Professor Chao Zhou, Chair

Machine learning is commonly used in biomedical image analysis, as it allows automated image segmentation and identification that minimizes the need for tedious human involvement. *Drosophila melanogaster* is often used as a cardiac disease model, where optical coherence microscopy (OCM) is used to image and analyze its beating dynamics. As OCM often generates a large volume of images, automated image segmentation is necessary to quantify the heart beating efficiently. Our most recent heart segmentation model, FlyNet 2.0+, is a fully convolutional LSTM U-Net model [1]. However, the performance of the model diminishes in the presence of artifacts, such as image reflection and heart movement, resulting in time-consuming manual intervention for mask correction. Therefore, we developed the FlyNet 3.0 model with integrated attention gates in skip connections between each level of the LSTM U-Net model [2]. The attention model adaptively adjusts and automatically learns to focus on the target structure, the heart area. Compared to the previous model, Flynet 3.0 increases the prediction intersection over union (IOU) accuracy from 0.86 to 0.89 for images with reflection artifacts and from 0.81 to 0.89 for those depicting heart movement. Furthermore, we have expanded the functionalities of OCM analyses through automated and dynamic heart wall thickness measurements, which we have validated using a *Drosophila* model of cardiac hypertrophy.

Chapter 1: Introduction

Drosophila melanogaster, or the fruit fly, is a powerful model organism for studying the developmental processes and biological mechanisms of human diseases. Approximately 75% of disease-causing genes in humans have been discovered to possess a homologous counterpart in this fruit fly [3,4]. The fruit fly's short life cycle makes it convenient for studying developmental processes, and during development, the heart tube resembles the human embryonic heart. Moreover, the *Drosophila* genome is fully sequenced, and the methods for manipulating the genes are well-established [5]. These characteristics make *Drosophila melanogaster* a perfect candidate for cardiac research. A significant number of cardiology studies have been done on *Drosophila* over the past 30 years, including exploring heartbeat characteristics [6,7], performing cardiac simulation [8,9], and building and analyzing the disease model [10–15].

Optical coherence microscopy (OCM) is chosen to image *Drosophila*'s cardiac structure and function *in vivo* because it is a noninvasive imaging modality, so that no exogenous contrast agents are involved, and samples can be returned after being used. The principles of OCM are the same as optical coherence tomography (OCT), which relies on light coherence and interference to obtain a resulting image. OCM inherits the imaging ability of OCT and further extends its resolution. The resulting image from OCM has a high spatial resolution, which can image up to 1 mm in depth, and has a high temporal resolution to capture the beating dynamics.

In the past decades, OCM has been used to study various characteristics of the *Drosophila*'s heart during different health conditions [16–18]. In most studies, the identifying and marking processes of the heart region were manually completed by human experts. As OCM quickly generates a large quantity of data for analysis of the *Drosophila*'s heart model in cardiac research, manually

performing heart segmentation on a large volume of OCM videos is time-consuming and laborious [8]. Therefore, the automatic image segmentation process is beneficial.

Previously, our group developed multiple versions of the FlyNet models, aiming to develop an automated *Drosophila*'s heart segmentation algorithm that minimizes the need for human involvement. Through these models, we have created a trustworthy heart function analysis pipeline. Although these models performed well in most cases and achieved an impressive prediction speed, they performed poorly when imaging artifacts were presented, including normal movements in the larval stage and noisy light reflection during the imaging process. Therefore, we implemented an attention LSTM model, named FlyNet 3.0, which integrated attention gates into the skip connections between each level of the LSTM U-Net model to improve segmentation accuracy in the presence of common artifacts. Additionally, we developed an automated *in vivo* heart wall thickness measurement based on the prediction mask to expand the current characterization of the heart using OCM. Altogether, we aim to provide an image processing platform that reduces the need for manual segmentation, allowing for accurate rapid segmentation regardless of artifacts to facilitate heart morphology and functional characteristic assessments.

Chapter 2 discusses more background about the OCT and OCM imaging systems, characteristics of *Drosophila* heart function and how it has been investigated using OCM, and previous work in automatic *Drosophila* heart segmentation. In Chapter 3, we introduce our novel FlyNet 3.0 model, including the dataset acquisition, dataset preparation, network structure, and other network settings, with a special focus on the newly added attention gate structure. The model's prediction results are discussed in Chapter 4, which includes attention weight visualization, prediction results analysis, and comparison between diverse imaging qualities. The next chapter outlines the application of this model on *in vivo* heart wall thickness measurements. The thesis concludes with

a discussion of the model's novelty and limitations in Chapter 6 followed by a conclusion and its future development in Chapter 7.

Chapter 2: Background

In this section, we will review the background of optical coherence tomography (OCT) and optical coherence microscopy (OCM) imaging systems, and how they were employed in the research project. This section will also introduce the importance of automating *Drosophila*'s heart segmentation in our cardiac research and the history of this automation process.

2.1 Optical Coherence Tomography (OCT)

In the development of medical imaging, it is beneficial to attempt to eliminate and substitute invasive methods that could create a negative impact on the examined object, while still acquiring useful images. To fill the need for noninvasive imaging that relies on endogenous contrast with high resolution, OCT was invented in 1991 by Fujimoto's group, which enabled researchers to obtain cross-sectional images of tissue by measuring the echo time delay and the magnitude of back-reflected light [19]. **Figure 1** shows an example of the OCT system, which is subdivided into 5 parts, the light source, the fiber coupler as the beam splitter, the reference arm, the sample arm, and the detector. The light emitted from the light source is guided to the fiber coupler and splits into the sample beam and the reference beam [20]. The reference beam travels to the mirror along the reference arm and directly reflects back to the fiber coupler, and the beam in the sample arm interacts with the backscatter of the sample. Both back-reflected lights combine and interfere with each other and are recorded by the detector. The resulting interference pattern is processed through a computer and generates a complete depth profile of the sample reflectivity called an axial scan or A-scan. A cross-sectional image or B-scan can be received by acquiring several A-scans at different transverse positions. Similarly, a three-dimensional data set or a 3D-OCT image can be acquired by performing repeated B-scans with a change in the Y direction. **Figure 2** shows examples of A-scan, B-scan, and 3D-OCT scan. OCT's high resolution, noninvasive, and

noncontact imaging characteristics make OCT popular in many research fields including ophthalmology and cardiovascular research.

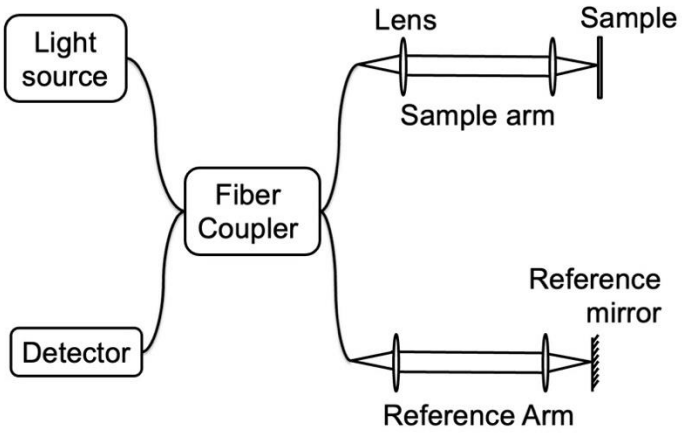

Figure 1: Schematic of OCT Imaging System. Each OCT system contains five parts, the light source, the fiber coupler as the beam splitter, the reference arm, the sample arm, and the detector.

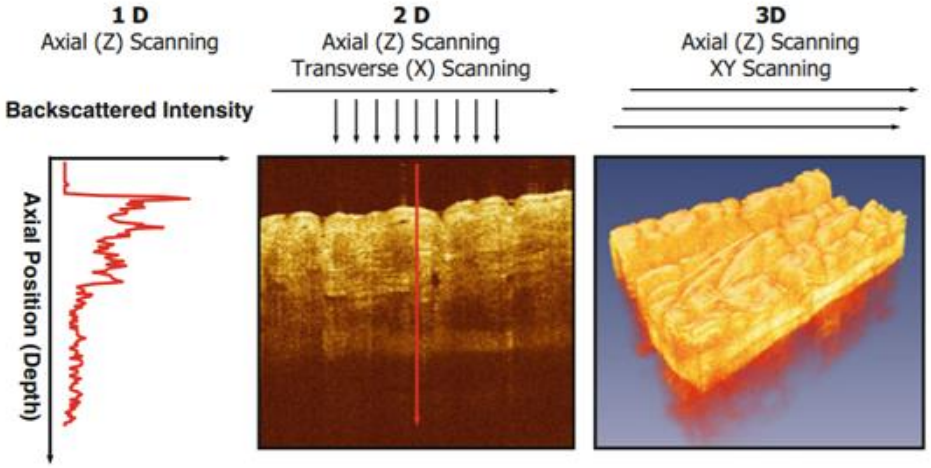

Figure 2: Examples of axial scan (A-scan), cross-sectional image (B–scan), and three-dimensional data set (3D–OCT) [21]. A-scan measures the backscattered intensity with respect to depth. Repeated A-scans at different transverse positions generate a B-scan. A 3D-OCT is captured by performing a series of B-scans along the Y direction.

OCT is mainly configured into two categories: Time-domain OCT (TD-OCT) and Fourier-domain OCT (FD-OCT). FD-OCT can be further separated into spectral-domain OCT (SD-OCT) [22] and swept-source OCT (SS-OCT) [23].

2.2.1 Time-domain OCT

The first OCT developed by Fujimoto's group is a TD-OCT system with a block diagram shown in **Figure 3**. The echo time delay and the intensity of backscattered light from biological tissues are measured by constantly moving the reference mirror to match the depth of the sample. The light signals are captured through the detector, a photodiode, to extract the A-scan information along the z direction.

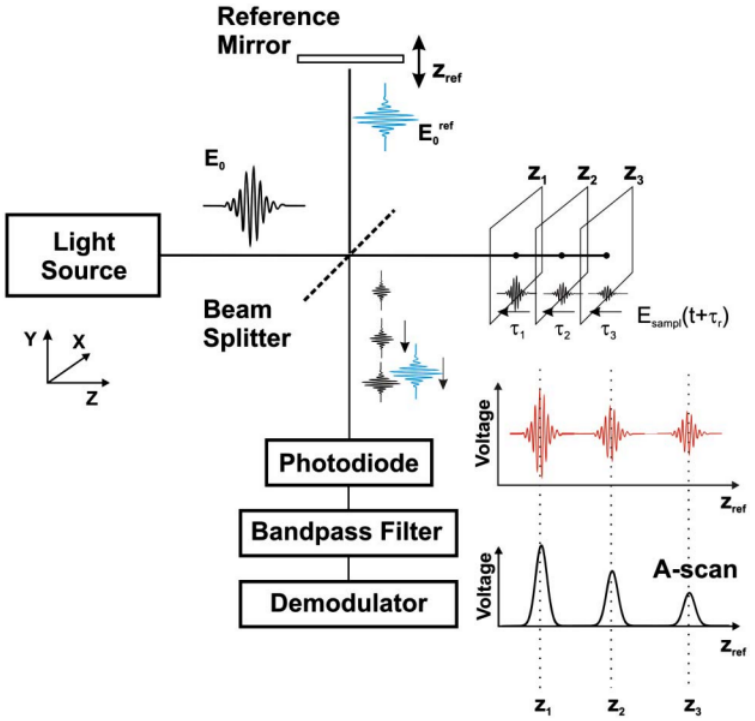

Figure 3: Block diagram of the time-domain OCT system [24]. The reference mirror is persistently moving to capture A-scans along the z direction.

2.2.2 Fourier-domain OCT

However, manually moving the reference mirror is laborious and greatly reduces the imaging speed. Therefore, researchers invented FD-OCT to improve image speeds, sensitivity, and axial resolution [25]. The key idea for FD-OCT is to fix the reference mirror and capture the spectral information of the backscattered light from the tissue, which is extracted through the Fourier transform. There are two approaches to achieve this idea, which developed into the SD-OCT and

SS-OCT. **Figure 4a** shows the block diagram for the SD-OCT. The main innovation for SD-OCT is that it uses a spectrometer instead of the photodiode as the detector to receive signals with different wavelengths. **Figure 4b** shows the block diagram for the SS-OCT. The SS-OCT keeps the photodiode as the detector but changes the light source to a tunable swept laser to allow signals with different wavelengths ejected with respect to time. Therefore, in both cases, the signal intensity and the echo time delay can be directly measured without moving the reference mirror. Moreover, FD-OCT has higher sensitivity because its parallel detection enables all depth information acquired simultaneously and improves the signal-to-noise ratio.

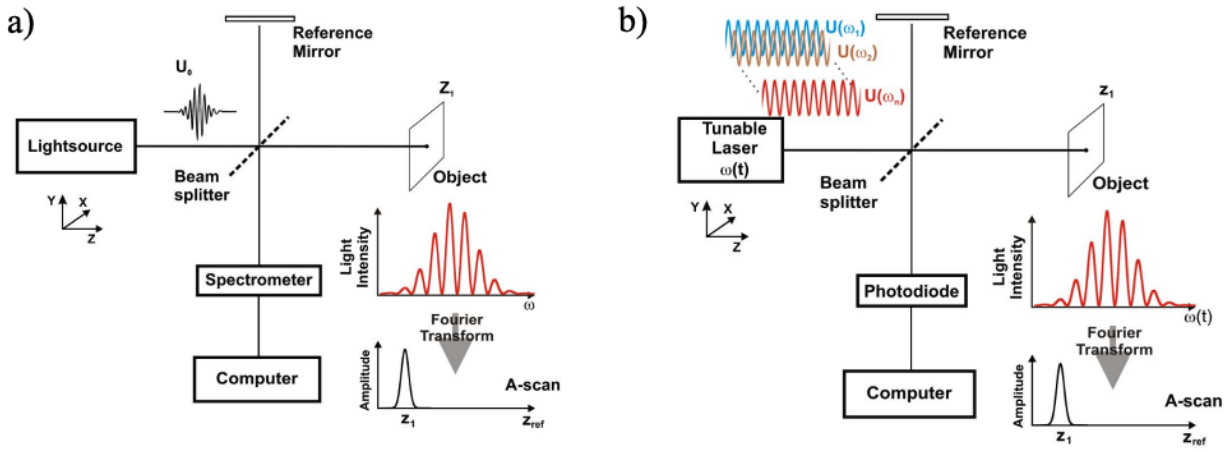

Figure 4: Block Diagrams of spectral-domain (SD-OCT) Imaging System and swept-source (SS-OCT) Imaging System [24]. A) The SD-OCT system contains the light source, the fiber coupler, the reference arm, the sample arm, and the spectrometer as the detector. B) The SS-OCT system keeps the photodiode from the TD-OCT system but changes the light source to a tunable laser that changes the source signal wavelength by time so that the signal can be directly measured without moving the reference mirror.

In general, FD-OCT has a more stable mechanism because of its fixed reference mirror, faster imaging speed, and higher sensitivity compared to TD-OCT [26]. Various adaptations and extensions have been made to the OCT system to further utilize it in a variety of research scopes, such as optical coherence microscopy (OCM) [19].

2.2 Optical Coherence Microscopy (OCM)

OCM is a well-established imaging modality that is modified based on OCT to provide a better transverse spatial resolution. OCM employs an objective lens with a greater numerical aperture compared to the conventional OCT, which enables high-resolution cellular imaging, sufficient to capture the beating dynamic of the fruit fly in this project. In our experiment, instead of monitoring the 3D-OCT images, we used OCM to scan repeated frames of the same cross-section over time. **Figure 5a** displays the OCM imaging system setup in our lab and **Figure 5b** shows the corresponding schematic.

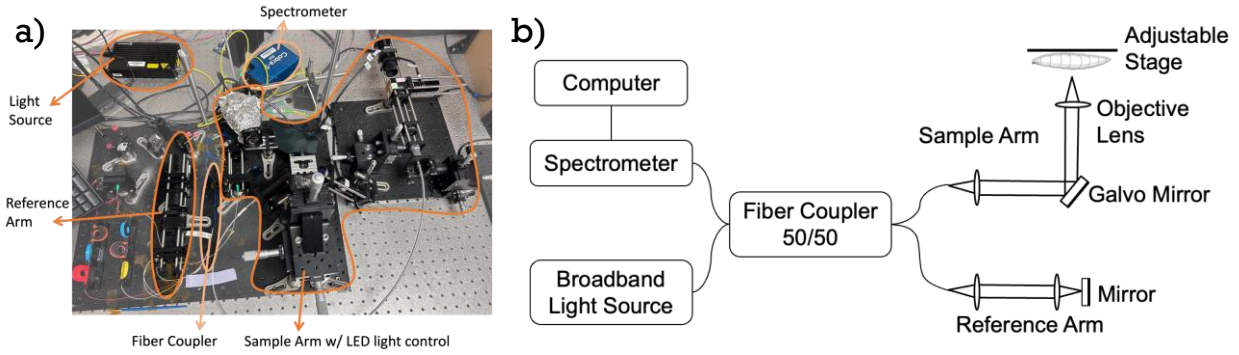

Figure 5: OCM Imaging System Setup and its Schematic. A) Experiment setup for OCM with LED light control of the sample. B) Experiment setup schematic for OCM, which includes a light source, fiber coupler, sample arm, reference arm, and spectrometer.

Previous OCM studies have examined cardiac disease's morphological and functional effects on *Drosophila'*s cardiac function, as OCM enables *in vivo*, label-free, high spatiotemporal resolution imaging [16–18]. Since the *Drosophila's* heart tube is located ~200μm beneath its surface, the light beam of OCM does not greatly attenuate in the tissue. Also, OCM is a noninvasive and noncontact imaging modality that utilizes the endogenous contrast naturally provided by light backscattering from the tissue. Moreover, OCM can reach up to a single-micron resolution and support a frame rate over 100 Hz to capture the heart-beating dynamics. These characteristics make OCM a perfect imaging modality to capture *Drosophila's* heart function. For example, in 2015,

Alex et al. utilized OCM to analyze the functional and morphological changes in the *Drosophila* heart throughout the entire post-embryonic lifecycle [27].

2.3 *Drosophila* Cardiac Study

The *Drosophila* heart is a valuable model for heart development due to its short life cycle, simple genome, well-established physiological processes and disease models, and similarities to the human embryonic heart [8]. The first cardiac study on *Drosophila melanogaster* was conducted 30 years ago, claiming that "the gene *tinman* is required for specification of the heart and visceral muscles in Drosophila [28]." Since then, *Drosophila* cardiac research has continued to grow. For example, in 2009, Holmes used OCT with M-mode to study *Drosophila*'s cardiovascular dynamics, including end systolic diameter (ESD), end diastolic diameter (EDD), and percentage fractional shortening [29]. The paper also suggested *Drosophila* as an important cardiac model for human disease, including disturbances of heart rate, rhythm, remodeling, and contractile function. In the same year, Fink et al. developed a *Drosophila* heart analysis algorithm for functional parameter quantifications [30]. They created a semi-automated heartbeat analysis pipeline to detect and quantify the heart rate, diastolic and systolic indices, ESD, EDD, percentage fractional shortening, contraction wave velocity, and arrhythmia index in adult and larval flies.

To monitor the heart function in real-time using OCM, we used M-mode images with repeated B-scans showing a cross-section of the heart at the same location over time to capture heart motion. **Figure 6** shows an example of the cropped OCM image of the *Drosophila* heart in both cross-sectional view and M-mode image. Several cardiac research projects are going on in parallel in our lab, including optogenetics projects that use light to control heart rhythm and the TREX project that uses *Drosophila* as a high throughput and a simpler model of disease, and each of them produces plenty of frames that need to be segmented and analyzed. A frame rate of 125 Hz was

applied to capture the beating dynamics of the heart. Different number frames were recorded for different developmental stages. In general, we obtained 2000 frames for larvae and early pupae and 4000 frames for adult flies.

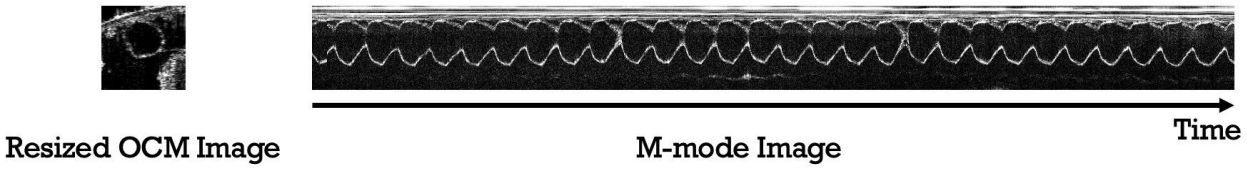

Figure 6: Example Resized OCM Image and M-mode Image. The M-mode image was acquired by repeatedly requiring the cross-section of the heart at the same location over time to capture motion.

2.4 Previous *Drosophila Melanogaster* Heart Tube Segmentation Model

In the early studies of examining the morphological and functional effects of cardiac disease on *Drosophila'*s cardiac function, human experts needed to identify the heart region manually or by using various image segmentation techniques, such as pacing pulses, measuring the dimensions of the heart chamber in each frame using custom-written MATLAB codes, and applying the magic wand algorithm to perform heart characteristic analysis [18,27,31,32]. However, these methods fail to offer a swift and comprehensive characterization of both the structure and function of the heart, and a great amount of human involvement is demanded. The large amounts of cross-sectional beating heat images time rapidly generated by the OCM further motivate and underscore the need for advancing automated segmentation techniques for the efficient assessment of heart function.

Over the past few decades, various object detection and image segmentation techniques have been employed to semi-automatically detect the *Drosophila*'s heart region in the OCM images, such as thresholding [33], Markov random field [34], region growing [35], active contour [36], edge

detection [37], and clustering [38]. However, the issue of automatically attaining both high segmentation accuracy and universality remains challenging.

As the deep neural network (DNN) continues to demonstrate success across diverse computer vision tasks, researchers are excitingly exploring and deploying various DNN architectures for segmenting biomedical images. Among these, the fully convolutional network stands out as one of the most successful DNNs for semantic segmentation [39]. In 2018, Duan et al. introduced a highly accurate model for identifying and marking the heart region of *Drosophila* in cross-sectional images [40]. This model utilized a fully convolutional U-Net architecture, featuring a U-shaped network structure and convolutional layers for both down-sampling and up-sampling. This model, named FlyNet 1.0, achieved an intersection over union (IOU) accuracy of 86%. In 2020, Dong et al. enhanced the FlyNet architecture to version 2.0 by incorporating convolutional long short-term memory (LSTM) to enable accurate 3D segmentation [41]. LSTM is an artificial recurrent neural network with feedback connections between time sequences [42]. FlyNet 2.0 utilizes both spatial and temporal information to improve segmentation performance, resulting in an improved IOU accuracy of 92%. In 2023, Fishman et al. optimized the FlyNet model with better hardware and GPU utilization [1]. The encoder blocks start with spatiotemporal encoders and are followed by spatial encoders. The spatiotemporal encode block keeps the LSTM 2D convolution layer to capture temporal information. The spatial encoders are encapsulated in a time-distributed layer to enable parallel computing to frame without temporal dependencies. The decoder block has no LSTM layer and therefore operates in parallel across all time slices. With using an Nvidia RTX 3090 GPU, FlyNet 2.0+ reaches an average IOU accuracy of 93%.

Chapter 3: Methods

This section will introduce the methods we used to gather, process, and analyze the *Drosophila* OCM dataset. This project was a teamwork collaboration with lab mates from Z-lab and the work has been published and available at [43].

3.1 Data Acquisition

The dataset used in this research project was acquired and developed based on the publication by Fishman et al. [1]. Therefore, most of the raw images were collected by the lab for the use of other projects. Extra images were gathered to ensure enough diversity in the dataset. **Figure 7** shows the developmental stages of *Drosophila melanogaster*. Our experiments are most interested in flies at larva, early pupa, and adult developmental stages. To collect the data, samples were mounted onto a glass slide, in which double-sided tape was used for each larval and pupal sample, and rubber cement was used for adult flies to stabilize their wings to the slide. The glass slide was placed in the sample holder of the OCM system. In our experiment, we used the customized OCM system described in **Figure 5** with a central wavelength of 850nm and a bandwidth of ~175 nm from a broadband light source, achieving a ~3.3 μm axial resolution in tissue. With a 10x objective lens, reaching a lateral resolution of ~2.8 μm, and a frame rate of 125 Hz, the system provides us with a high spatiotemporal resolution imaging platform. **Figure 8a** shows the example of an adult fly mounted on the glass slide, and a zoomed-in view of the fly under a stereo microscope is displayed in **Figure 8b**. During the imaging process, we chose to image the largest heart segmentation in the heart tube, which is the A7 segment of the heart in larval and early pupal flies and the A1 segment for adult flies, as **Figure 9** shows.

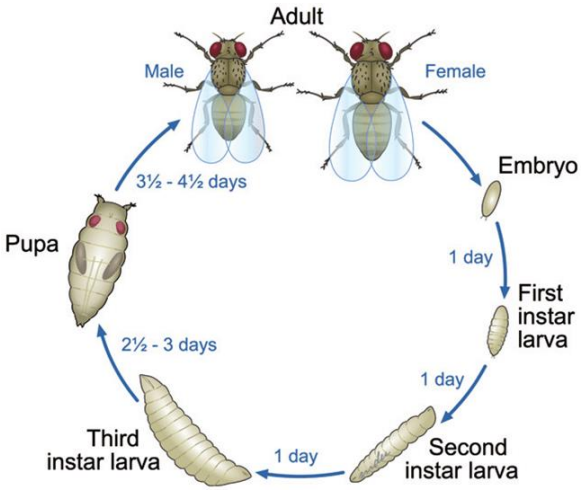

Figure 7: Developmental Stages for *Drosophila Melanogaster* [44].

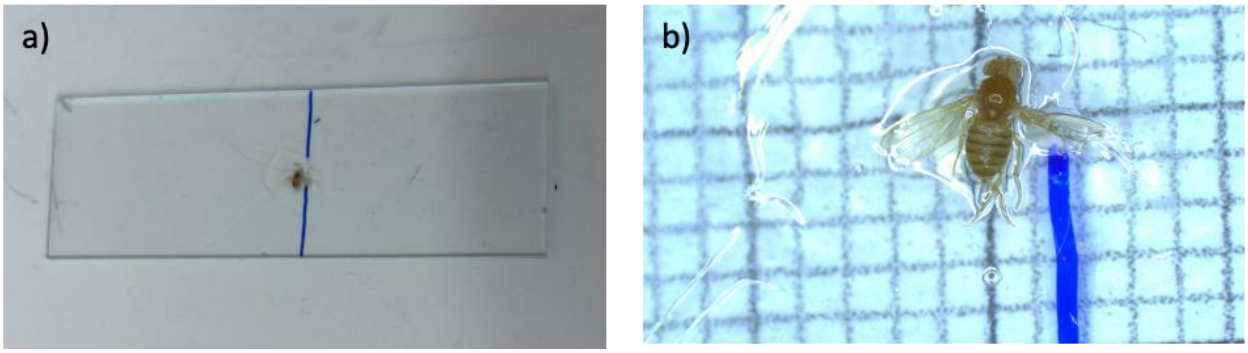

Figure 8: Example of Adult Fly being Stabled on Glass Slide. A) Adult fly mounted on the glass slide by rubber cement. B) Zoomed in on the adult fly under the microscope.

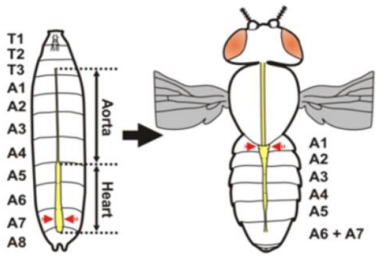

Figure 9: Schematic Representation of *Drosophila* Heart Metamorphosis on larva and adult flies [31]. The yellow section represents the heart tube of the fruit fly, and the red arrow points to the largest cross-sectional heart area we chose to image under the OCM system.

3.2 Data Preparation

To generate a model that is generalizable for automatic image segmentation, we prepared a dataset of a wide range of genotypes, involving disease models and optogenetic pacing data at different developmental stages, consisting of larva, early pupa, and adult Drosophila [3,32,45]. We

additionally collected a set of images with different degrees of image artifacts, including reflection and movement artifacts, for different developmental stages and added them on top of the existing dataset for training. All the datasets were manually assigned to one of the three categories: “reflection”, “movement”, and “normal.” As **Figure 10** shows, “Normal” images present a clear and well-defined heart region present throughout the whole video. Images with reflection artifacts were categorized according to the extent to which the artifact passes through or blocks part of the heart region. As **Figure 11** shows, the orange arrow points to the vertical white reflection passing through the right side of the heart. **Figure 12** shows an example from the “Movement” category. “Movement” sets contain images with the heart region moving positions occasionally within the frame, which are dominated by larval samples. The example demonstrates a change in heart position ~1s apart.

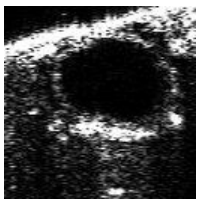

Figure 10: OCM Image of Adult Example from "Normal" Category in Cross-section view.

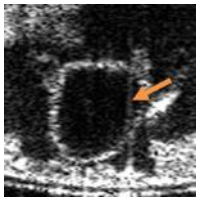

Figure 11: OCM Image of Early Pupa Example from "Reflection" Category in Cross-section view. The orange arrow points to a vertical white reflection artifact that cuts off the right side of the heart.

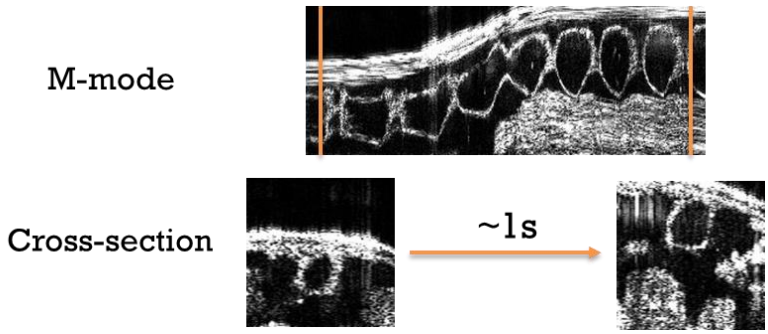

Figure 12: OCM Images of Larva Example from "Movement" Category. The top is an M-mode image acquired from the OCM imaging system. The bottom are cross-sectional images of the orange frames with a time change of ~1s. In the example, the heart moves to a different position within the same frame position.

Overall, we gathered 710,000 frames of OCM images, which included 198,000 reflection frames, 48,000 movement frames, and 464,000 normal frames, as shown in **Table 1**. To train the neural network model, we randomly split the dataset into three sets: 60% for training, 5% for validation, and 35% for testing.

Table 1: Fly Heart OCM Videos Dataset

	Larva	Early Pupa	Adult	Total	# Frames
Reflection	15	17	35	67	198,000
Movement	12	0	0	12	48,000
Normal	30	35	102	167	464,000
Total	57	52	137	246	710,000

To predict a mask, the raw 128 x 701 pixel full-size images were cropped and resized into a 128 x 128 pixel image using the first neural network to identify the general heart region of the video. The resulting image was fed into our FlyNet 3.0 model for precise heart segmentation.

3.3 General Network Structure

Figure 13 presents a comprehensive overview of the proposed heart region segmentation network. We utilized the general structure of the FlyNet 2.0+ model, which consists of four convolution levels and three skip connections bridging the corresponding encoder and decoder blocks, and built the FlyNet 3.0 model by including the attention gate at each skip connection, as shown in red in

Figure 13 [1]. Inside each convolution block, there are two pairs of long short-term memory (LSTM) convolution layers, batch normalization (BN) layers, and Leaky rectified linear unit (ReLU) activation layers. Each LSTM 2D convolution layer from the encoder captures the spatial and temporal information of the input image. The skip connection passed the information to the decoder blocks and contributed to the final prediction. Under this model, we treated all the information as equally important and contributed to the final prediction at the same level. However, the spatial information containing the heart region needed to be further emphasized. Therefore, we added the attention gates on the skip connection to assign attention weight to each pixel, emphasizing the importance of these features. **Figure 13B** shows the inside structure of the newly added attention gates, which takes the skip connection signal as the horizontal dashed purple input "x", the gating signal from the previous layer as the vertical dashed pink input "g," and the output the weighted signal "s" into the skip connection. The attention model adaptively adjusted and automatically learned to focus on the target structure, the heart region. To locate the heart area precisely, the model placed more weight on the pixels containing the salient features of the heart, like the heart area and the heart boundary, and less weight on irrelevant regions, like the background tissues, in the input images.

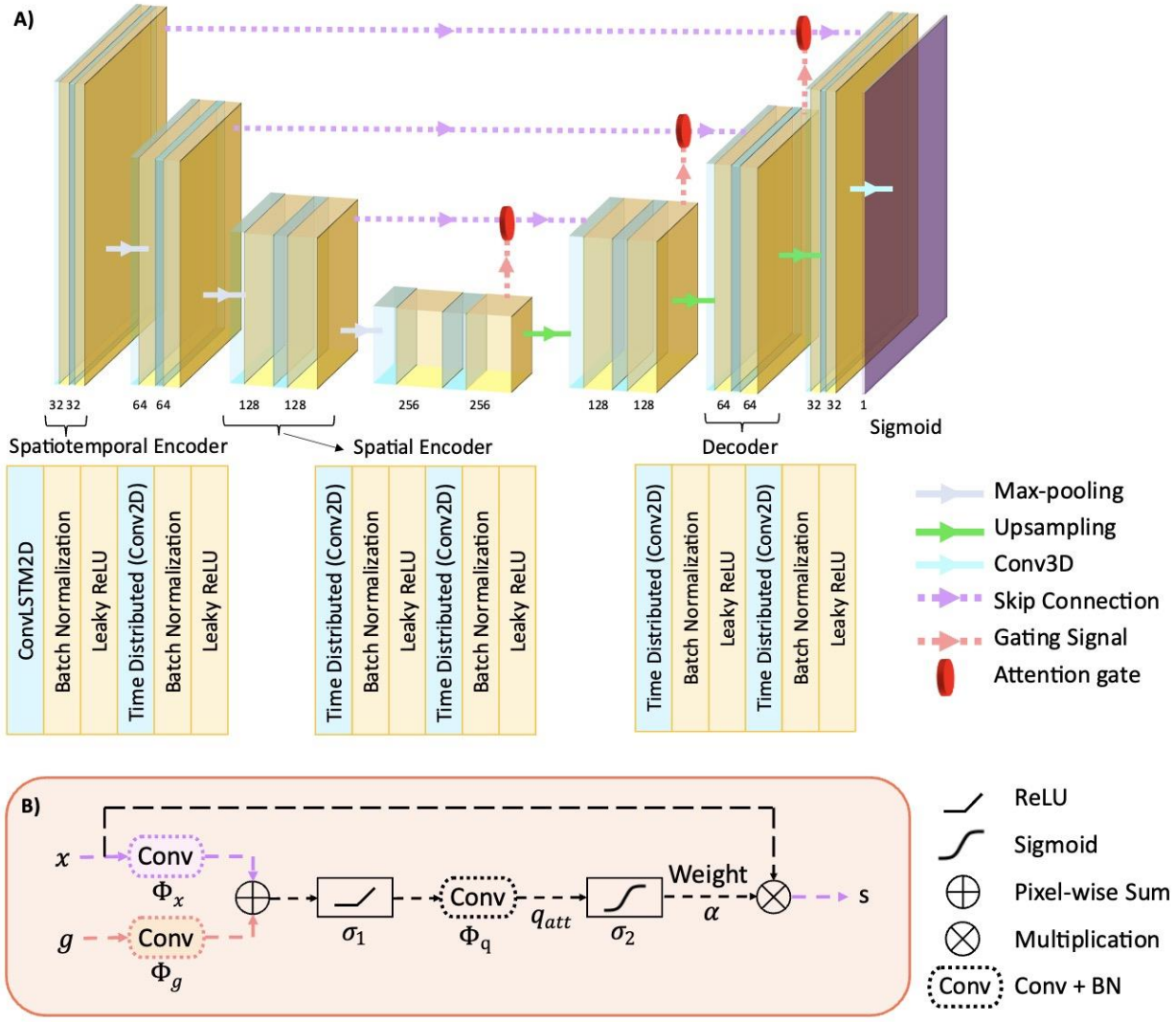

Figure 13: The Flynet 3.0 Model Architecture [43]. A) There are four encoders and three decoder blocks, followed by a sigmoid layer. Each block contains two groups of 2D convolutional layers with a kernel size of 5x5, shown in blue, a batch normalization layer in yellow, and a ReLU activation function in yellow. The numbers shown below each block denote the number of filters for each group. The encoder blocks include a spatiotemporal encoder, in which the 2D convolutional layer is wrapped with an LSTM layer, followed by spatial encoding blocks, in which the 2D convolutional layer is encapsulated in a time-distributed wrapper. Similarly, all the decoders are spatial decoders, except for the last one, which is a spatiotemporal decoder. The gray solid arrow represents a 2D max-pooling layer with a pool size of 2x2. The green solid arrow represents a 2D transposed convolutional layer with a kernel size of 2x2 and a stride of 2. The purple dashed arrow indicates where a skip connection occurs. The pink dashed arrow stands for the gating signal. The gating signal and the skip connections are copied to the attention gate, shown in the red cylinder, and the output concatenates to the next layer. The last layer is a 3D convolutional layer with a kernel size of 1x1x1, followed by a sigmoid activation function. B) Detailed structure of the attention gate. The input g is from the gating signal, x is from the skip connection, and the output s is copied to the input of the corresponding decoder.

In summary, FlyNet 3.0 takes in a resized OCM video as the input to the model and outputs a binary image of the heart region mask, as shown in **Figure 14**. The FlyNet 3.0 model was implemented in Python and the code is publicly available on GitHub [46].

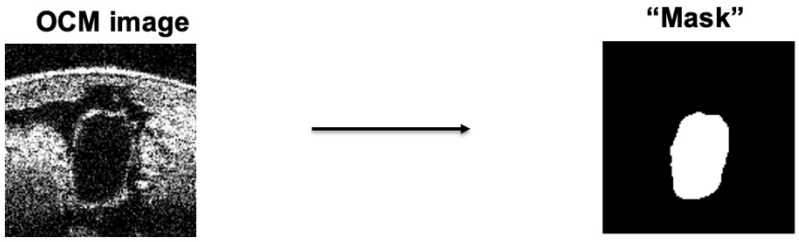

Figure 14: FlyNet 3.0 Input and Output. The input to the FlyNet model is a grayscale resized OCM video, and the model predicts and outputs a binary mask of the heart region.

3.4 Model Training and Evaluation

To create the model with the best segmentation performance of the *Drosophila* heart, we trained the FlyNet 3.0 model using the Adam optimizer with a learning rate of 0.0001. The loss function was chosen as the Log-Cosh Dice Loss function, which was developed from the Dice Coefficient. The Dice Coefficient follows:

$$Dice\ Coefficient = \frac{2(predict\ area \cap ground\ area)}{predict\ area + ground\ area}, \quad (1)$$

where *predict area* is the sum of the area from the predicted mask from the FlyNet 3.0 model and *ground area* is the sum of the ground-truth mask area. The purpose of the Dice Coefficient is to optimize the segmented area to the ground truth as closely as possible [47]. However, the non-convex nature of the Dice Coefficient can potentially lead to failure when attaining optimal outcomes, because the result may not reach convergence. Inspired by the Lovsz-Softmax loss function, researchers developed the Log-Cosh approach [48]:

$$Dice\ Loss = 1 - Dice\ Coefficient, \quad (2)$$

$$L_{lc-dce} = \log\left[\cosh(Dice\ Loss)\right]. \quad (3)$$

To evaluate the accuracy of the attention model in generating the 2D mask over times of the heart from the resized input image, we chose to use the IOU as the evaluation criteria to compare the

predicted mask with the ground truth mask. IOU is calculated as the intersection region of the predicted mask and the ground-truth mask over the union region of the two masks, as shown below:

$$IOU = \frac{predict\ area \cap ground\ area}{predict\ area \cup ground\ area}. \quad (4)$$

The model was trained for 80 epochs with 193 steps per iteration and a batch size of 32. We monitored the change in learning performance throughout the epochs, after the loss value reached stability, and determined the best model based on the lowest loss value and the highest IOU over the full validation dataset. The model was trained and predicted using TensorFlow 2.10.1 and Python 3.7.4 on a workstation equipped with an NVIDIA GeForce RTX 3090 card.

All OCM videos were assumed to contain the whole heart region within the frame, and it was contained and continuous in each frame of the OCM cross-sectional view images. Therefore, we applied postprocessing steps on the predicted mask to improve the segmentation accuracy by keeping the largest connected component and smoothing the boundary. All results for both FlyNet 2.0+ and FlyNet 3.0 are shown with post-processed images in this report.

3.5 Attention Gate Structure

Figure 13B shows the structure of the attention gate, which computes the attention weight for the input feature, x. The attention weight aims to identify the significant image region and determine the focal area [49]. The attention gate inputs two parameters: input x from the skip connection and the gating signal g from upsampling of the previous block. The attention weight is a single scalar weight vector calculated for each pixel vector at each frame. The output of the attention gate is the element-wise multiplication between the input feature mapping and the calculated attention weight, which ranges from 0 to 1.

To calculate the attention weight, we first need to capture the characteristics of the skip connection input and the gating signal through the convolutional and BN layers, with a kernel size of 2x2 for the convolutional layers and a kernel size of 1x1 for the BN layers. The next step is to add the result from the gating signal to the result from the skip connection. Adding the gating signal's characteristics helps determine the focal region of the attention weight. Then, the ReLU activation function is applied to the pixel-wise summation to bring non-linearity into the model to introduce complexity [48]. Another linear transformation was applied to the result by passing the signals through an additional convolutional and BN layer with a kernel size of 1x1. This would output the initial attention weight coefficient of the input as follows:

$$q_{att}(x, g) = \Phi_q\{\sigma_1[\Phi_x(x) + \Phi_g(g)]\}, \quad (5)$$

where Φ_x and Φ_g are the convolutional and BN layers for inputs x and g respectively, σ_1 is the ReLU activation function, and Φ_q represents the last convolutional and BN layers for the initial model after passing through the first activation function.

The coefficient needs to be rescaled between 0 and 1 by applying an activation function before applying the initial attention weight coefficient to the input x:

$$\alpha = \sigma_2[q_{att}(x, g)], \quad (6)$$

where σ_2 is the sigmoid activation function.

Lastly, the attention weight, α, is applied to and rescales the input features, x, as follows:

$$s = \alpha x. \quad (7)$$

The result, s, will be used as the new skip connection signal feeding to the next block.

Adding the gating characteristics also eliminates noisy and unrelated responses from the skip connections, because the gating signal is determined based on pertinent information extracted from the previous prediction layer. Furthermore, only filtering neuronal activations and element-wise multiplication were performed prior to the merge operations of the attention gate, ensuring all features extracted are retained. All the operations in the attention gates are linearly transformed without introducing any spatial modifications, making them a suitable complement to the LSTM model we previously employed.

3.6 Statistical Analysis

We performed the statistical analysis using a two-sided student's t-test. The results were considered significant when $p<0.05$.

Chapter 4: Results

This section presents the experimental results of testing the model and attention weight visualization to help understand the model mechanism. Several segmentation result comparisons were made to demonstrate the improvement of our model.

4.1 Attention Weight Visualization

To facilitate a better understanding of the training process and the attention gate algorithm, the attention weight vector obtained from different training epochs was visualized in **Figure 15.** The demo ground-truth image shown on the left covers a red ground-truth mask on its corresponding grayscale resized image. The rest of the images are the visualization of the attention weight vector with respect to different training epochs. All the weight vectors represented in the heatmap were normalized to the attention weight of the whole video. In the heatmap, a red pixel in the visualization indicates high intensity in weight, in contrast to a blue pixel. In the beginning epochs, the broad distribution of the heatmap generated at the 3rd epoch indicates a uniform distribution of the attention weight over the entire image frame, and the attention weight passes through features at all locations. Then, the attention weight gradually learned and focused on the pattern of the heart region, like the heart area or the heart boundary, depending on the area of the heart. **Figure 15A** and **Figure 15B-C** have shown example images for both scenarios. If the heart's area is relatively large along all the frames in a resized OCM video, as shown in **Figure 15A**, the attention weight gradually focuses on the heart region and penalizes the rest of the background pixels by giving them low intensities. On the other hand, if an obvious variation in the size of the heart region occurs between the systolic and diastolic phases, as shown in **Figure 15B-C**, the attention weight uses the empty surrounding region of the heart bordered by tissues around the heart region in the larval body to help to locate the heart position.

When the evaluation criteria reached stability, the best training model was determined based on the lowest loss value. In this experiment, the model reaches an optimal focus on the heart region around the 47th epoch. Further training of the model does not help with refining the model weights to capture additional useful information about the heart region. When comparing attention weight between the 47th epoch to the later epochs, such as epochs 47 and 63 found in **Figure 15**, the attention weight demonstrates a higher intensity in the surrounding heart region in the 63rd epoch compared to epoch 47. The model also overtrained and produced an attention weight with a larger spread throughout the frame, indicating less focus around the heart region. In contrast, the attention weight at the 47th epoch provided a rough outline of the heart, which is gradually refined at finer resolution at coarser scales. This is proven in **Figure 15D** when the filtered-out low-intensity pixels of the optimal attention weight overlap the resized image.

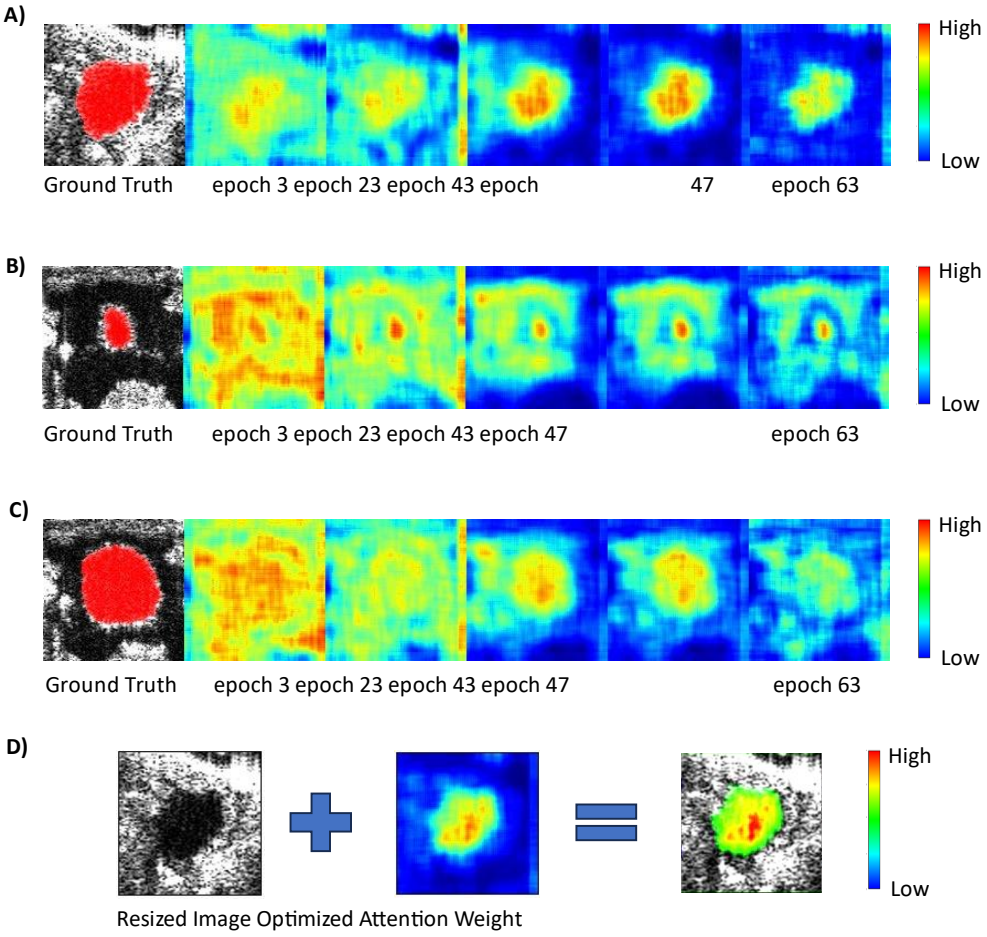

Figure 15: Attention weight visualization [43]. A) Example of the attention weight learning process on adult fly images over a range of epochs. B) and C) Attention weight learning process on the same fly larva images over the same range of epochs of systolic and diastolic phases, respectively. D) Resized input image to the model and the corresponding optimized attention weight. After overlapping them and filtering out the low-intensity region, the ideal attention weight is focused on the heart area.

4.2 Test Results

To reach an objective evaluation, a variety of datasets were tested on our model, including images with no artifacts presented (normal), images with reflection artifacts intersecting the heart region (reflection), and images in which the heart moves horizontally or vertically during the frame sequence (movement). The average IOU results for both FlyNet 2.0+ and FlyNet 3.0 in each category and the overall performance, with the use of the same testing dataset, were demonstrated in **Table 2**. The overall performance increased from 88% to 92%, and improvements in accuracy

happened in all the categories. The ‘movement’ group showed the largest enhancement, with the accuracy raised from 81% to 89%.

Table 2: Comparative Test Results for FlyNet 3.0 and 2.0+

Average IOU	Normal	Reflection	Movement	Overall*
FlyNet 3.0	0.93	0.89	0.89	0.92
FlyNet 2.0+	0.91	0.86	0.81	0.88

* “Overall” = 67% Normal Frames + 24% Reflection Frames + 9% Movement Frames

4.2.1 Normal

Images with no obvious artifacts were categorized into the normal dataset. The prediction result from the normal group remains excellent from the FlyNet 2.0+ model to the FlyNet 3.0 model. **Figure 16A** shows an example of resized images and heart mask combined images of the heart on the left and the overlapping comparison image between the ground truth and both prediction models in a cross-sectional view on the right column. All the ground truth masks are shown in red, prediction masks from both models are in green, and the yellow region represents the overlapping area between the ground truth and prediction results. An almost yellow-covered overlapping image indicates a high accuracy in segmenting the mask region from the OCM image. **Figure 16B** shows the M-mode image of the ground truth mask and both overlayed comparison masks overlayed on the M-mode resize image. The heart movement was clearly captured. The heart regions of the overlapping images in both **Figure 16A-B** were almost completely overridden by yellow, which represents an accurate automatic heart segmentation. This example image has an accuracy of 96% for the FlyNet 3.0 model and 95% for the FlyNet 2.0+ model.

The precise result for both models is further confirmed by the IOU traces shown in **Figure 16C**, where IOU remains close to 100%, with some periodic oscillation due to the smaller area in systole. **Figure 16D** further compares the overall IOU for samples in the ‘normal’ group. The statistical

analysis result further proves that FlyNet 3.0 improves segmentation accuracy by 2% over FlyNet 2.0+.

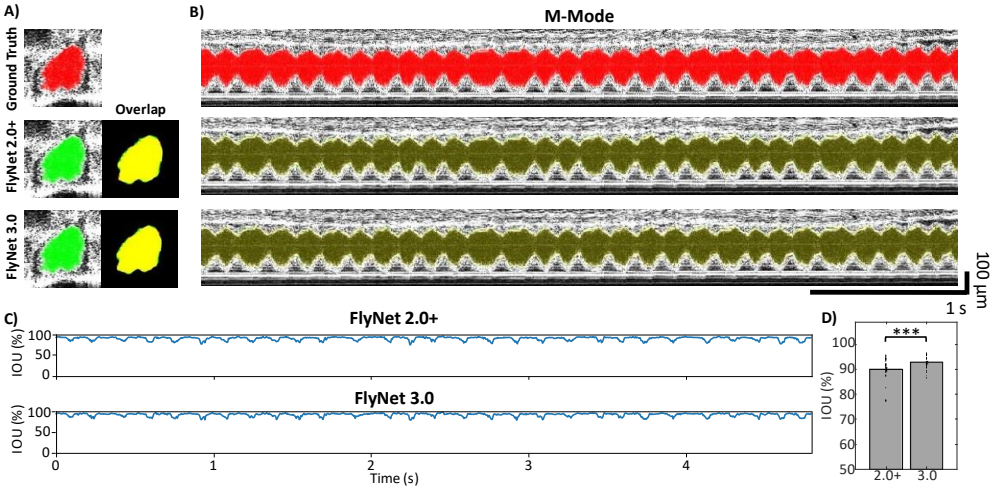

Figure 16: Flynet 3.0 performance on normal group example [47]. A) Cross-sectional comparison of the ground truth mask (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0. The overlap between the two masks is shown in yellow. B) M-mode images comparison of ground truth masks (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0. The overlap between the two masks is shown in yellow. C) IOU comparison of FlyNet 2.0+ and FlyNet 3.0 for the example shown in A-B. D) Group comparison between FlyNet 2.0+ and FlyNet 3.0, *** $p<0.001$.

4.2.2 Reflection

One common artifact in optical images is light reflection from noisy parts of the sample. **Figure 17A** shows an example from the "reflection" group in the cross-sectional view, including the combined resized OCM and heart mask images on the left column and overlapping comparison images of the ground truth mask and prediction models. The corresponding M-model images are shown in **Figure 17B**. The blue arrow indicates where the reflection occurs, which appears as vertical and horizontal white lines in the cross-section. By looking at the overlapping image produced by FlyNet 2.0+ in cross-section, we can easily point out the red area next to the small yellow overlapping region. The M-mode image further demonstrates the misclassification region for the entire image, where the presence of the artifact reduces the size of the FlyNet 2.0+ predicted mask, especially in the systolic phases. These red pixels indicate an inaccurate prediction of the mask area due to the presence of a reflection artifact. The mask overlayed images show FlyNet

2.0+ falsely treats the reflection artifacts as the heart boundary. With the attention model added to FlyNet, the prediction mask from the FlyNet 3.0 model covers the entire heart region, and the yellow area from the overlapping image accurately represents the precise prediction result of the FlyNet 3.0 model. Both almost yellow-covered overlayed images from the bottom row of **Figure 17A-B** indicate high fidelity between the ground truth and predicted masks.

Figure 17C further confirms this result by showing the IOU trace for this example dataset. The IOU drops down to 0 in some systolic phases in FlyNet 2.0+ due to the reflection artifact causing the model to be unable to recognize any of the heart areas. In contrast, the IOU performance for FlyNet 3.0 remains close to 100% for the entire image. **Figure 17D** compares the overall performance of the two models for the entire 'reflection' group, where FlyNet 3.0 improves IOU over FlyNet 2.0+ by 3%.

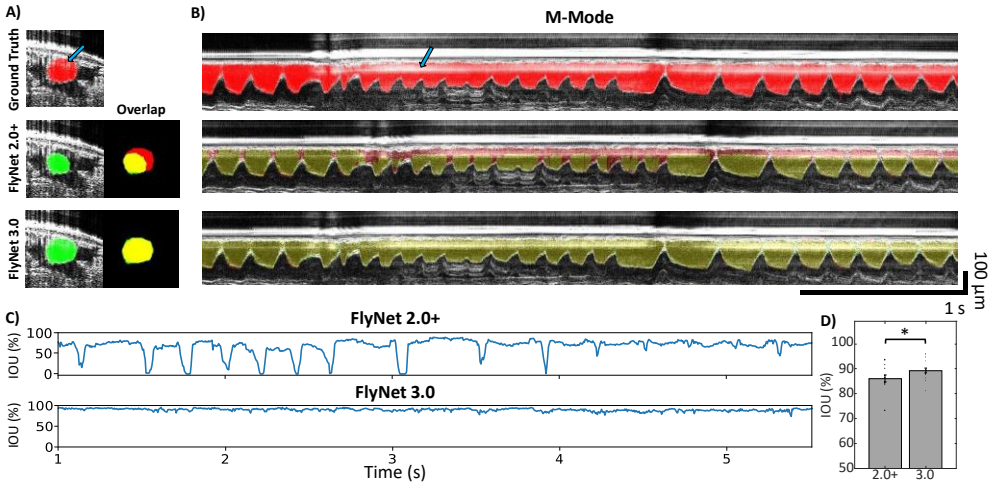

Figure 17: Flynet 3.0 performance on reflection group example [43]. A) Cross-sectional comparison of the ground truth mask (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0. The overlap between the two masks is shown in yellow. Reflection artifacts are indicated in the region by the white arrow. B) M-mode images comparison of ground truth masks (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0. The overlap between the two masks is shown in yellow. An example reflection artifact is indicated by the blue arrow. C) IOU comparison of FlyNet 2.0+ and FlyNet 3.0 for the example shown in A-B. D) Group comparison between FlyNet 2.0+ and FlyNet 3.0, * $p<0.05$.

4.2.3 Movement

When *Drosophila melanogaster* is in the larval stage, it can move its position from time to time. This creates the movement artifacts shown in this category. **Figure 18** shows an example from the 'movement' group. The cross-sectional view OCM image shown in **Figure 18A** is selected from the frame soon after the movement occurs. The heart region changes position within the frame in seconds due to movement. When resizing an image to 128x128 for prediction, the resized image has a fixed cropped region and must contain the whole heart for the duration of the recording. Thus, the resulting resized images in the movement dataset do not focus as tightly on the heart region, as the region the heart occupies during the recording is large. FlyNet 2.0+ performs poorly with smaller and less centered heart regions, which leads to a high potential of experiencing lower accuracy for identifying the mask using the FlyNet 2.0+ model. The additional green area in the overlap image from the FlyNet 2.0+ prediction, which cannot be removed by post-processing due to it being connected to the true heart region, points out the misidentification area for causing a low IOU value for this frame. By adding the attention gate to the model, this area is less likely to be recognized by our FlyNet 3.0 model. By looking at the M-Mode overlap images shown in **Figure 18B**, the dense green areas in the FlyNet 2.0+ row indicate it struggles in this scenario and continues the pattern of recognizing additional areas that are not part of the heart region. The white arrows point to approximately when the heart movement starts and ends. Heart movement diminishes the contribution of the LSTM model, resulting in poor prediction accuracy. The Flynet 3.0 model aims to overcome this issue by placing more attention on the heart region, making it more likely to identify the heart region correctly. Therefore, the overlapping mask in the M-mode for FlyNet 3.0 shows a mainly yellow area, with a few areas of green, indicating a more accurate prediction.

These observations and explanations are further proven graphically in **Figure 18C**, where the IOU plot for FlyNet 2.0+ has more discontinuities than the plot for FlyNet 3.0. The IOU per frame remains around 100% for FlyNet 3.0 except for a few, but the IOU changes dramatically for FlyNet 2.0+ after the change in heart position. Overall statistical results for the movement group, shown in **Figure 18D**, indicate that FlyNet 3.0 significantly improves performance over FlyNet 2.0+ by 4%.

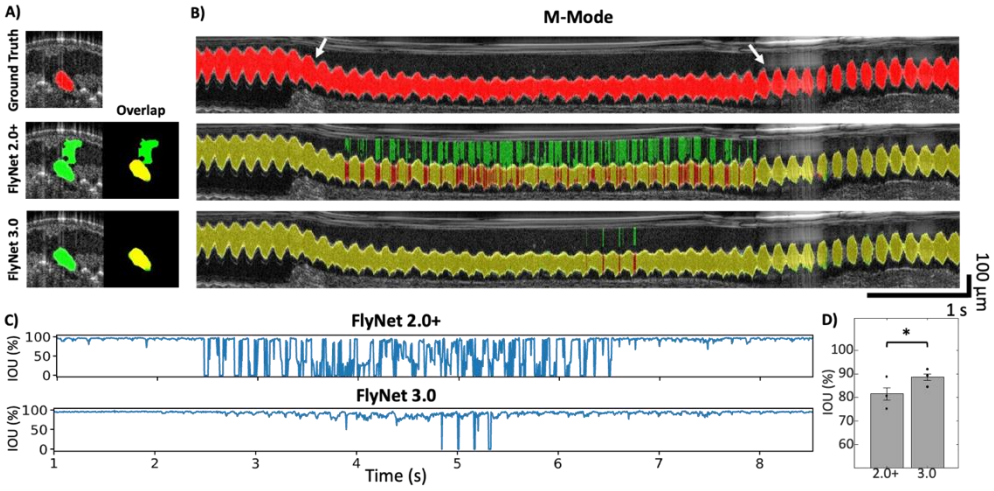

Figure 18: Flynet 3.0 performance on reflection group example [43]. A) Cross-sectional comparison of the ground truth mask (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0 during a period of movement. The overlap between the two masks is shown in yellow. B) M-mode images comparison of ground truth masks (red) and prediction results (green) from FlyNet 2.0+ and FlyNet 3.0. The overlap between the two masks is shown in yellow. Areas with movement are indicated by the white arrow. C) IOU comparison of FlyNet 2.0+ and FlyNet 3.0 for the example shown in A-B. D) Group comparison between FlyNet 2.0+ and FlyNet 3.0, * $p<0.05$.

Chapter 5: Application – *In vivo* Heart Wall Thickness Measurements

In addition to training a robust model that accurately identifies the heart region while mitigating the influence of artifacts, we aim to create a more comprehensive package for analyzing both morphological and dynamic features of the *Drosophila* heartbeat and ultimately contribute to cardiac research. Our previous FlyNet models have enabled many useful phenotyping tools, including automatic analysis of the characteristics of heart rate, fractional shortening, and heart areas. In our updated model, given a more accurate prediction, we propose an additional tool for measuring the heart wall thickness *in vivo*.

5.1 Methods

Figures 19A-F show the calculation examples of a larval sample. To measure the thickness of the heart wall, we begin with an accurate prediction mask on the raw OCM image. After receiving the 128x128 pixel resized image of the heart region, the first step is to crop the raw OCM image according to the resize parameters generated by the resize network. Then, we resize the predicted mask according to the same parameters to match the size of the cropped OCM image. This step is critical in keeping the precision of heart wall thickness measurement. Then, we rescale both the cropped OCM image and the resized mask according to the pixel-to-micron ratio, which was found by imaging a USAF target with the same scan settings. Therefore, we can directly measure the thickness along the heart wall in the image to get the result, as each pixel in the rescaled image represents one micron.

All the images demonstrated in **Figure 19** are rescaled to their physical distance representation. We are most interested in the heart wall thickness during the systolic and diastolic stages. **Figures**

19A, 19C, and **19E** show example measurement procedures during the systolic stage, and **Figures 19B, 19D,** and **19F** from the diastolic stage. The end-systolic and end-diastolic measuring points are identified by the peaks and valleys calculated through the heart area measurement over time. **Figures 19A-B** show the scaled combined images in the systolic and diastolic phases respectively, with the OCM image in greyscale and the prediction mask of the heart in red. It is critical to know the position of the heart wall precisely. Luckily, we can retrieve this information based on the heart boundary of the predicted mask. The pink curves outlined in **Figure 19C-D** depict the relative position of the heart wall. Then, we smooth out the curve to decrease the effect brought by the sharpness of the predicted heart. For each pixel position along the smoothed curve of the heart boundary, we drew a line perpendicular to the heart wall to obtain the OCM intensity along the line. The exampled two neighboring red dots in **Figure 19C-D** are used to determine the slope of the green point of interest, which results in the blue direction. We allow the step for all the pixels along the pink curve except for those that border other tissues in the body. We choose to exclude those points because there are only a minority of them per frame in the video and it is hard to accurately differentiate the heart wall computationally for these points. Points are also excluded if artifacts are presented in the image that may reduce the imaging quality of the heart wall. Since other tissues near the heart and artifacts have different abilities to reflect light, we apply thresholding, prominence, peak detection, and other methods to determine the bad points automatically. For each frame, after obtaining the intensity measurement along the perpendicular direction for all the selected points along the heart boundary, we align all the measurement profiles and average them. This results in the blue curve in **Figure 19E-F**. We employed the full-width at half-maximum (FWHM) method to measure and determine the heart wall thickness, as shown in the green line segment of **Figure 19E-F**. Therefore, to optimize the alignment and averaging

algorithm, we use the center position of the FWHM measurement as the reference to align them together in a single frame. The averaging of the intensity distribution along all the points in a single frame produces a trace that follows the Gaussian distribution. These steps are repeated for all the frames, and the FWHM values are recorded accordingly. In this example, the resulting average heart wall thickness during systole is ~13.0 μm, and ~10.5 μm during diastole, which is ~2.4 μm in difference on average, which aligns with the hypothesis that the heart wall is thicker during systole than diastole in general.

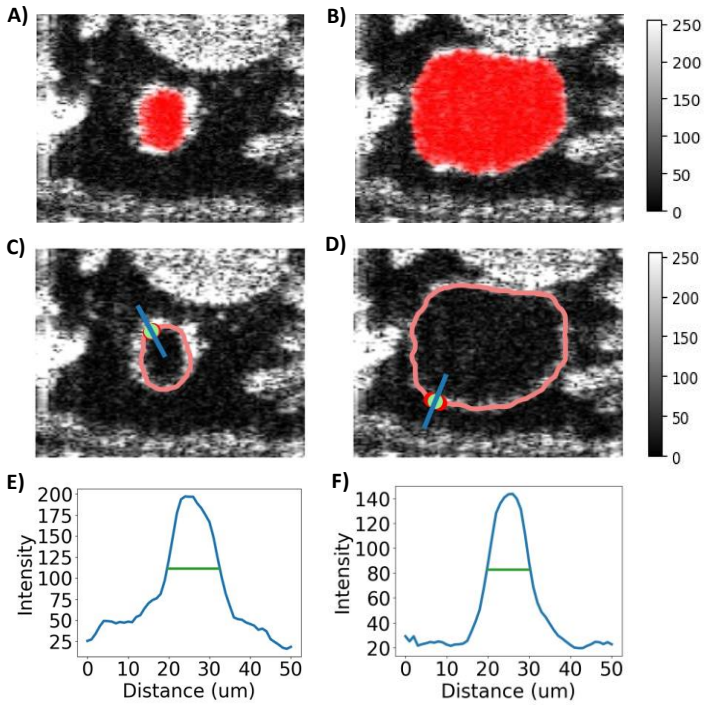

Figure 19: Larval heart wall thickness calculation example [43]. The left column is the systolic measurement data, and the right column is the diastolic. A) and B) Combined OCT and heart segmentation image; C) and D) Heart wall boundary and perpendicular trace example along the boundary to measure the heart wall thickness; E) and F) Averaged intensities for all the points along the boundary in C) and D) respectively, with the corresponding FWHM values.

5.2 Algorithm Validation

5.2.1 Dataset

We utilized the OCM dataset of a *Drosophila* model of cardiac hypertrophy published by Migunova et al. in 2023 to verify the heart wall thickness measurements algorithm [50]. Here, CRISPR technology was used to create *Drosophila* models that developed heart hypertrophy and

systolic dysfunction. For this study, we acquired OCM videos of n=34 heart-specific knockouts of RNase Z, tinC-Cas9>RNZ^{KO}, and n=16 control, tinC-Cas9>gZ^{+} for in vivo heart wall thickness analysis. The tinC-Cas9>RNZ^{KO} larvae were expected to have a thicker heart wall hypertrophy compared to the tinC-Cas9>gZ^{+}.

5.2.2 Results

To validate our algorithms, we characterized the heart wall thickness of tinC-Cas9>gZ^{+} and tinC-Cas9>RNZ^{KO} larvae from our previous study using this method, where the heart wall thickness measurements were performed by manually measuring the transverse histological sections [50]. The average heart wall thickness was measured at the end-diastolic phase and the end-systolic phase separately for the two larvae genotypes, as **Figure 20** shows. For both end-diastolic and end-systolic phases, tinC-Cas9>RNZ^{KO} larvae exhibited a significant increase in cardiac wall thickness compared to those of tinC-Cas9>gZ^{+} control ($p<0.05$). The histology analysis of the larvae dataset indicates an average heart wall thickness of ~4.3 μm for tinC-Cas9>gZ^{+} and ~7.2 μm for tinC-Cas9>RNZ^{KO}. Though our measurements are larger than the histology measurement, the results are still consistent with the findings of the original publication, where histologic treatment of tissue may cause shrinkage, explaining the discrepancy in measurements [50].

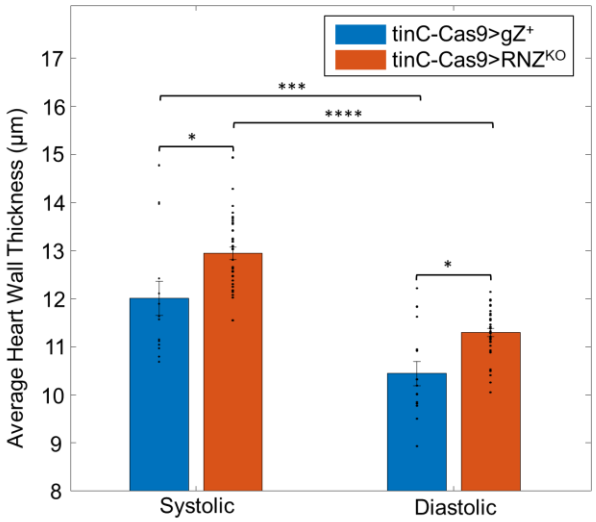

Figure 20: Heart wall thickness quantification measurement [43]. Quantification of heart wall thickness measurements in between a cardiac hypertrophy model (tinC-Cas9>RNZ^{KO}) and control (tinC-Cas9>gZ^{+}). Error bars indicate the mean ±s.e.m. * $p<0.05$, *** $p<0.001$, **** $p<0.0001$.

Chapter 6: Discussion

Our customized FlyNet 3.0 segmentation network leverages the attention mechanism to capture and concentrate on characteristics of the *Drosophila* heart region in the cross-sectional OCM videos, aiming to minimize the need for human supervision and interference. In contrast to FlyNet 2.0+, FlyNet 3.0 improved the design of the neural network structure by seamlessly integrating attention gates into the skip connections at every level of the LSTM U-Net model. The attention mechanism, introduced by Bahdanau et al. in 2015 [49], enables the autonomous learning of the objective features through attention gates, by focusing on the characteristics of target structures without the need for additional supervision. The incorporation of attention gates effectively elevates the model sensitivity and accuracy of dense label predictions without imposing significant computational overhead [2]. This improvement is achieved by introducing the attention weight vector and suppressing feature activations in irrelevant regions, thereby accentuating the important features of the heart region. FlyNet 3.0 takes in these advantages and further improves the prediction performance, especially for images with artifacts, because the heart region demands special attention for accurate localization. Models trained with attention blocks self-learn to progressively focus on the *Drosophila* heart region, resulting in enhanced segmentation accuracy for fly heart OCM videos.

The high segmentation accuracy allows Flynet 3.0 to increase its ability in automatic cardiac analysis by introducing the ability to measure the heart wall thickness dynamically *in vivo*. Although our model (**Figure 20**) showed the same measurement trend as in the source publication, the measurement numbers for histology and in vivo OCM imaging are quite different. The histological measurements of the average heart wall thickness present a difference of ~3 μm between the tinC-Cas9>RNZ^{KO} and tinC-Cas9>gZ^{+} larvae, whereas our OCM measurements

differ by ~1 μm during the same systolic and diastolic phases. Additionally, our calculated result from *in vivo* imaging was greater than the histological measurement. This difference originates from the degree of sample or tissue processing required for each method. Tissue distortion or artifacts could occur for a histological measurement because tissues must be fixed, paraffin-embedded, thinly sliced, and stained. Using OCM as a method to noninvasively image and measure the heart wall thickness is dynamic and vividly shows the change in heart wall thickness at each point in the cardiac cycle, providing a comparison between systolic and diastolic heart wall thicknesses. The heart wall is thicker in the systolic measurement, as the tissue is contracted, and thinner in the diastolic measurement, as the tissue is expanded.

Although FlyNet 3.0 has demonstrated significant improvements compared to FlyNet 2.0+, especially for images with artifacts, there are still some limitations. First, our model assumes all the images contain the whole heart region throughout the entire video. Therefore, datasets containing the heart moving out of frame, most common in the larval stage, or the heart region completely covered by reflections cannot be used. Repeated collections are recommended to decrease the occurrence of these situations. Next, for a dataset to be used, human experts must be able to identify the heart region to verify the quality of the segmentation. Additionally, adding the attention gate module also slightly increases the computational cost and requires more GPU resources, as it introduces the attention weight parameters to the model.

For using the proposed method to calculate the heart wall thickness, the heart wall is required to be differentiated from surrounding tissues in most places. Frames are excluded if the calculated thickness is outside of a stable range, defined as within a standard deviation of the mean, or below a certain degree of prominence, which helps exclude misleading measurements where the heart is in contact with other tissue or experiencing strong artifacts and a valid measurement cannot be

performed. Additionally, image rescaling from pixels to physical distance in microns introduces a pixelation effect, where all pixelwise measurements must be rounded to an integer pixel.

In future work, more training datasets of fly OCM videos with more accurate masks and different image qualities, including artifacts in different development stages, can be added to further boost the segmentation accuracy. In addition, other model structures, such as the transformer U-Net, may provide even faster image processing and utilize long-range dependencies among pixels in the input image, potentially advancing the overall performance of the segmentation model [51–53].

Chapter 7: Conclusion

We have proposed an LSTM attention model, FlyNet 3.0, designed to segment the cross-section heart region of the *Drosophila melanogaster* more accurately in OCM videos. Evolved from the earlier versions of FlyNet, FlyNet 3.0 incorporates a learning attention mechanism to focus on image features extracted from the heart region. The new model was tested on a diverse *Drosophila* heart OCM video dataset, varying in image quality, developmental stage, and heart rhythmicity. FlyNet 3.0 demonstrated notable improvements in segmentation as measured by IOU accuracy, particularly in scenarios involving artifacts. It achieved an average accuracy of 89% for images with reflection artifacts and 89% for those depicting occasional heart movement. These refined segmentation results enable the accurate and efficient calculation of heart wall thickness, ensuring reliable performance in cardiac analysis. In conclusion, FlyNet 3.0 produces more accurate masks that reduce the need for manual correction for the heart region. Combined with the new *in vivo* heart wall thickness measurement algorithm, it will make cardiac disease analysis in *Drosophila* more comprehensive, as demonstrated with the RNase Z model of cardiac hypertrophy.

References

1. M. Fishman, A. Matt, F. Wang, E. Gracheva, J. Zhu, X. Ouyang, A. Komarov, Y. Wang, H. Liang, and C. Zhou, "A Drosophila heart optical coherence microscopy dataset for automatic video segmentation," Sci. Data **10**(1), 886 (2023).

2. P. Harsh, R. Chakraborty, S. Tripathi, and K. Sharma, "Attention U-Net Architecture for Dental Image Segmentation," in *2021 International Conference on Intelligent Technologies (CONIT)* (2021), pp. 1–5.

3. E. Gracheva, F. Wang, A. Matt, H. Liang, M. Fishman, and C. Zhou, "Developing Drosophila melanogaster Models for Imaging and Optogenetic Control of Cardiac Function," J. Vis. Exp. JoVE (186), (2022).

4. B. Ugur, K. Chen, and H. J. Bellen, "Drosophila tools and assays for the study of human diseases," Dis. Model. Mech. **9**(3), 235–244 (2016).

5. "Fruitfly Genome Sequencing," https://www.genome.gov/11008080/fruitfly-genome-sequencing.

6. K. Ocorr, G. Vogler, and R. Bodmer, "Methods to assess Drosophila heart development, function and aging," Methods San Diego Calif **68**(1), 265–272 (2014).

7. C. E. Petersen, B. A. Tripoli, T. A. Schoborg, and J. T. Smyth, "Analysis of Drosophila cardiac hypertrophy by microcomputerized tomography for genetic dissection of heart growth mechanisms," Am. J. Physiol.-Heart Circ. Physiol. **322**(2), H296–H309 (2022).

8. A. Souidi and K. Jagla, "Drosophila Heart as a Model for Cardiac Development and Diseases," Cells **10**(11), 3078 (2021).

9. M. Nishimura, K. Ocorr, R. Bodmer, and J. Cartry, "Drosophila as a model to study cardiac aging," Exp. Gerontol. **46**(5), 326–330 (2011).

10. N. Piazza and R. J. Wessells, "Drosophila Models of Cardiac Disease," Prog. Mol. Biol. Transl. Sci. **100**, 155–210 (2011).

11. M. Cozhimuttam Viswanathan, G. Kaushik, A. J. Engler, W. Lehman, and A. Cammarato, "A Drosophila melanogaster Model of Diastolic Dysfunction and Cardiomyopathy Based on Impaired Troponin-T Function," Circ. Res. **114**(2), e6–e17 (2014).

12. I.-M. Kim and M. J. Wolf, "Serial Examination of an Inducible and Reversible Dilated Cardiomyopathy in Individual Adult Drosophila," PLOS ONE **4**(9), e7132 (2009).

13. O. Taghli-Lamallem, T. Akasaka, G. Hogg, U. Nudel, D. Yaffe, J. S. Chamberlain, K. Ocorr, and R. Bodmer, "Dystrophin deficiency in Drosophila reduces lifespan and causes a dilated cardiomyopathy phenotype," Aging Cell **7**(2), 237–249 (2008).

14. M. J. Wolf, "Modeling Dilated Cardiomyopathies in *Drosophila*," Trends Cardiovasc. Med. **22**(3), 55–61 (2012).

15. W. A. Kronert, K. M. Bell, M. C. Viswanathan, G. C. Melkani, A. S. Trujillo, A. Huang, A. Melkani, A. Cammarato, D. M. Swank, and S. I. Bernstein, "Prolonged cross-bridge binding triggers muscle dysfunction in a Drosophila model of myosin-based hypertrophic cardiomyopathy," eLife **7**, e38064 (2018).

16. M. A. Choma, S. D. Izatt, R. J. Wessells, R. Bodmer, and J. A. Izatt, "In Vivo Imaging of the Adult Drosophila melanogaster Heart With Real-Time Optical Coherence Tomography," Circulation **114**(2), e35–e36 (2006).

17. L. Ma, A. Bradu, A. G. Podoleanu, and J. W. Bloor, "Arrhythmia Caused by a Drosophila Tropomyosin Mutation Is Revealed Using a Novel Optical Coherence Tomography Instrument," PLOS ONE **5**(12), e14348 (2010).

18. A. Li, C. Zhou, J. Moore, P. Zhang, T.-H. Tsai, H.-C. Lee, D. M. Romano, M. L. McKee, D. A. Schoenfeld, M. J. Serra, K. Raygor, H. F. Cantiello, J. G. Fujimoto, and R. E. Tanzi, "Changes in the Expression of the Alzheimer's Disease-Associated Presenilin Gene in Drosophila Heart Leads to Cardiac Dysfunction," Curr. Alzheimer Res. **8**(3), 313–322 (2011).

19. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical Coherence Tomography," Science **254**(5035), 1178–1181 (1991).

20. S. Aumann, S. Donner, J. Fischer, and F. Müller, "Optical Coherence Tomography (OCT): Principle and Technical Realization," in *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*, J. F. Bille, ed. (Springer, 2019).

21. J. G. Fujimoto and W. Drexler, "Introduction to OCT," in *Optical Coherence Tomography: Technology and Applications*, W. Drexler and J. G. Fujimoto, eds. (Springer International Publishing, 2015), pp. 3–64.

22. M. Wojtkowski, A. Kowalczyk, R. Leitgeb, and A. F. Fercher, "Full range complex spectral optical coherence tomography technique in eye imaging," Opt. Lett. **27**(16), 1415–1417 (2002).

23. S. R. Chinn, E. A. Swanson, and J. G. Fujimoto, "Optical coherence tomography using a frequency-tunable optical source," Opt. Lett. **22**(5), 340–342 (1997).

24. M. Wojtkowski, "High-speed optical coherence tomography: basics and applications," Appl. Opt. **49**(16), D30–D61 (2010).

25. K. A. Townsend, Gadi Wollstein, and Joel S. Schuman, "Fourier-Domain Optical Coherence Tomography," https://glaucomatoday.com/articles/2007-nov-dec/GT1107_08-php.

26. R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher, "Performance of fourier domain vs. time domain optical coherence tomography," Opt. Express **11**(8), 889–894 (2003).

27. A. Alex, A. Li, X. Zeng, R. E. Tate, M. L. McKee, D. E. Capen, Z. Zhang, R. E. Tanzi, and C. Zhou, "A Circadian Clock Gene, Cry, Affects Heart Morphogenesis and Function in Drosophila as Revealed by Optical Coherence Microscopy," PLoS ONE **10**(9), e0137236 (2015).

28. R. Bodmer, "The gene tinman is required for specification of the heart and visceral muscles in Drosophila," Dev. Camb. Engl. **118**(3), 719–729 (1993).

29. J. Holmes, "In vivo real-time optical coherence tomography imaging of Drosophila for cardiovascular research," Nat. Methods **6**(10), iii–iv (2009).

30. M. Fink, C. Callol-Massot, A. Chu, P. Ruiz-Lozano, J. C. Izpisua Belmonte, W. Giles, R. Bodmer, and K. Ocorr, "A new method for detection and quantification of heartbeat parameters in Drosophila, zebrafish, and embryonic mouse hearts," BioTechniques **46**(2), 101–113 (2009).

31. J. Men, J. Jerwick, P. Wu, M. Chen, A. Alex, Y. Ma, R. E. Tanzi, A. Li, and C. Zhou, "Drosophila Preparation and Longitudinal Imaging of Heart Function In Vivo Using Optical Coherence Microscopy (OCM)," J. Vis. Exp. JoVE (118), 55002 (2016).

32. A. Alex, A. Li, R. E. Tanzi, and C. Zhou, "Optogenetic pacing in Drosophila melanogaster," Sci. Adv. **1**(9), e1500639 (2015).

33. A. Xu, L. Wang, S. Feng, and Y. Qu, "Threshold-Based Level Set Method of Image Segmentation," in *2010 Third International Conference on Intelligent Networks and Intelligent Systems* (2010), pp. 703–706.

34. G. Xiao, M. Brady, J. A. Noble, and Y. Zhang, "Segmentation of ultrasound B-mode images with intensity inhomogeneity correction," IEEE Trans. Med. Imaging **21**(1), 48–57 (2002).

35. A. Madabhushi and D. N. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," IEEE Trans. Med. Imaging **22**(2), 155–169 (2003).

36. X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, "Learning Active Contour Models for Medical Image Segmentation," in (2019), pp. 11632–11640.

37. H. G. Kaganami and Z. Beiji, "Region-Based Segmentation versus Edge Detection," in *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (2009), pp. 1217–1221.

38. D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," Fuzzy Sets Syst. **161**(4), 522–543 (2010).

39. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 3431–3440 (2015).

40. L. Duan, X. Qin, Y. He, X. Sang, J. Pan, T. Xu, J. Men, R. E. Tanzi, A. Li, Y. Ma, and C. Zhou, "Segmentation of Drosophila heart in optical coherence microscopy images using convolutional neural networks," J. Biophotonics **11**(12), e201800146 (2018).

41. Z. Dong, J. Men, Z. Yang, J. Jerwick, A. Li, R. E. Tanzi, and C. Zhou, "FlyNet 2.0: drosophila heart 3D (2D + time) segmentation in optical coherence microscopy images using a convolutional long short-term memory neural network," Biomed. Opt. Express **11**(3), 1568–1579 (2020).

42. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," in *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)* (1999), **2**, pp. 850–855 vol.2.

43. X. Ouyang, A. Matt, F. Wang, E. Gracheva, E. Migunova, S. Rajamani, E. B. Dubrovsky, and C. Zhou, "Attention LSTM U-Net model for Drosophila melanogaster heart tube segmentation in optical coherence microscopy images," Biomed. Opt. Express **15**(6), 3639–3653 (2024).

44. C. Ong, L.-Y. L. Yung, Y. Cai, B.-H. Bay, and G.-H. Baeg, "Drosophila melanogaster as a model organism to study nanotoxicity," Nanotoxicology **9**(3), 396–403 (2015).

45. J. Men, A. Li, J. Jerwick, Z. Li, R. E. Tanzi, and C. Zhou, "Non-invasive red-light optogenetic control of Drosophila cardiac function," Commun. Biol. **3**(1), 1–10 (2020).

46. Xiangping Ouyang, "Alli-lolly/FlyNet3.0," (2024).

47. S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (2020), pp. 1–7.

48. M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks," in (IEEE Computer Society, 2018), pp. 4413–4421.

49. D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," (2016).

50. E. Migunova, S. Rajamani, S. Bonanni, F. Wang, C. Zhou, and E. B. Dubrovsky, "Cardiac RNase Z edited via CRISPR-Cas9 drives heart hypertrophy in Drosophila," PLOS ONE **18**(5), e0286214 (2023).

51. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation," in *Computer Vision – ECCV 2022 Workshops*, L. Karlinsky, T. Michaeli, and K. Nishino, eds., Lecture Notes in Computer Science (Springer Nature Switzerland, 2023), pp. 205–218.

52. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in (IEEE Computer Society, 2022), pp. 1748–1758.

53. S. He, R. Bao, P. E. Grant, and Y. Ou, "U-Netmer: U-Net meets Transformer for medical image segmentation," (2023).