

Washington University School of Medicine

Digital Commons@Becker

---

Open Access Publications

---

2020

**FlyNet 2.0: Drosophila heart 3D (2D + time) segmentation in optical coherence microscopy images using a convolutional long short-term memory neural network**

Zhao Dong

Jing Men

Zhiwen Yang

Jason Jerwick

Airong Li

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

---

---


**Authors**

Zhao Dong, Jing Men, Zhiwen Yang, Jason Jerwick, Airong Li, Rudolph E Tanzi, and Chao Zhou

---



# FlyNet 2.0: drosophila heart 3D (2D + time) segmentation in optical coherence microscopy images using a convolutional long short-term memory neural network

ZHAO DONG,<sup>1,2</sup> JING MEN,<sup>2,3</sup> ZHIWEN YANG,<sup>2,4</sup>  JASON JERWICK,<sup>1,2</sup> AIRONG LI,<sup>5</sup> RUDOLPH E. TANZI,<sup>5</sup> AND CHAO ZHOU<sup>1,2,3,\*</sup> 

<sup>1</sup>Department of Electrical and Computer Engineering, Lehigh University, 27 Memorial Drive W, Bethlehem, PA 18015, USA

<sup>2</sup>Department of Biomedical Engineering, Washington University in St. Louis, 1 Brookings Dr., St. Louis, MO 63130, USA

<sup>3</sup>Department of Bioengineering, Lehigh University, 111 Research Drive, Bethlehem, PA 18015, USA

<sup>4</sup>ShenYuan Honors College, Beihang University, Beijing 100191, China

<sup>5</sup>Genetics and Aging Research Unit, McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, 55 Fruit St, Boston, MA 02114, USA

\*chaozhou@wustl.edu

**Abstract:** A custom convolutional neural network (CNN) integrated with convolutional long short-term memory (LSTM) achieves accurate 3D (2D + time) segmentation in cross-sectional videos of the *Drosophila* heart acquired by an optical coherence microscopy (OCM) system. While our previous FlyNet 1.0 model utilized regular CNNs to extract 2D spatial information from individual video frames, convolutional LSTM, FlyNet 2.0, utilizes both spatial and temporal information to improve segmentation performance further. To train and test FlyNet 2.0, we used 100 datasets including 500,000 fly heart OCM images. OCM videos in three developmental stages and two heartbeat situations were segmented achieving an intersection over union (IOU) accuracy of 92%. This increased segmentation accuracy allows morphological and dynamic cardiac parameters to be better quantified.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

*Drosophila melanogaster*, widely known as the fruit fly, shows many similarities with vertebrates in the early stages of heart development [1]. About 75% of disease-causing genes in humans have been estimated to have functional orthologs in *Drosophila* [2–4], and *Drosophila* has been widely used to investigate mechanisms associated with human cardiac disorders. As a powerful genetic model, *Drosophila* has been used to investigate heart development and cardiac diseases [5–9]. Recently, *Drosophila* was also used to develop a novel optogenetic pacing technique that can noninvasively control the heart rhythm [10].

Optical coherence tomography (OCT) [11–14] is an emerging biomedical imaging technology that enables noninvasive micron-scale, cross-sectional, 3D imaging of biological tissues. OCT is widely used for clinical applications, including ophthalmology [15–17], cardiology [18], endoscopy [19], dermatology [20,21], and dentistry [22]. Optical coherence microscopy (OCM) [10], which integrates OCT and confocal detection, can achieve high resolution. The penetration image depth of an OCM system is around 500  $\mu\text{m}$ , deep enough to image a fruit fly heart, which is around 200  $\mu\text{m}$  below the surface of the fly's cuticle. In this project, the OCM system acquires cross-sectional videos of the fruit fly heart *in vivo*, each containing either 4,000 or 6,000 OCM cross-sectional images that cover over a hundred heartbeat cycles. 2D OCM image acquisition

is less sensitive to sample motion and A-scan position of M-mode image acquisition. With accurate fruit fly heart segmentation from 2D OCM images, dynamic cardiac parameters, such as fly heart area, heart rate, end-diastolic diameter (EDD), end systolic diameter (ESD), and fraction shortening (FS), can be accurately measured. However, due to the large data size of the *Drosophila* cardiac OCM recording, a robust and fast method to segment the fruit fly heart is needed.

Different methods have been utilized for object detection and segmentation over the last several decades. These include traditional methods, like the histogram of oriented gradients (HOG) [23,24], the scale-invariant feature transform (SIFT) [25], and the features from accelerated segment test (FAST) [26]. Thresholding methods have also been widely used for grayscale image segmentation [27–29]. In addition, k-means [7] and support vector machines (SVMs) [30] have been employed to segment images. However, high segmentation accuracy and universality are difficult to achieve with these traditional methods.

The success of deep neural networks (DNN) in various computer vision tasks has inspired researchers to apply DNN in image segmentation, and many network structures have been designed for this application. The fully convolutional network (FCN) [31] is one of the most successful DNNs for semantic segmentation. It employs a convolutional neural network (CNN) [32] to extract image feature maps and achieves a significant improvement in training accuracy compared to traditional methods. Later, more advanced convolutional networks were designed to improve performance further. The SegNet [33] and U-Net [34] networks added a decoder stage, which is a symmetrically expanding path, to accurately localize features. With a much better model structure, DNN has much-improved segmentation accuracy over traditional methods.

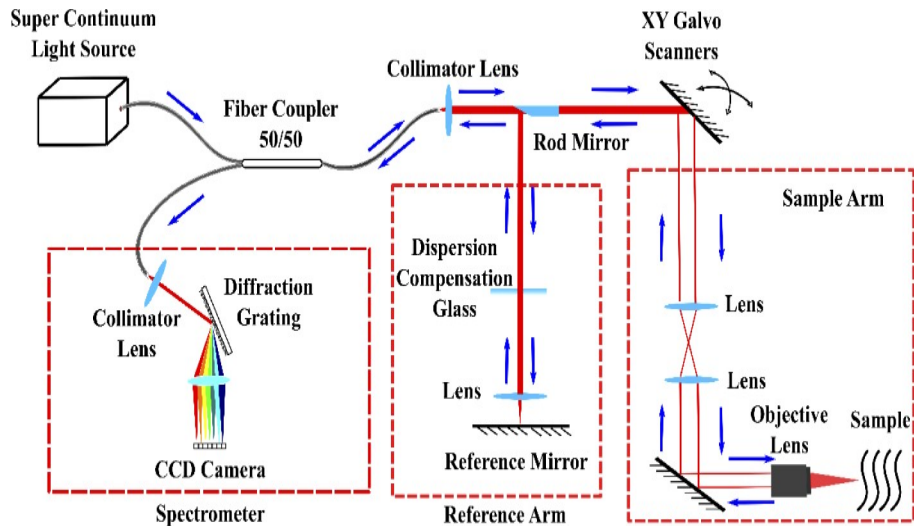
As a starting point, we employed U-Net based convolutional neural networks to segment fly heart OCM videos and called the software FlyNet 1.0 [35]. Traditional CNNs extract only 2D spatial information from each OCM image, but fruit fly heartbeat videos contain time sequence information between adjacent images which can be used to improve the segmentation performance further. Long short-term memory (LSTM) [36] is an artificial recurrent neural network architecture that has feedback connections between time sequences. LSTM networks have internal contextual state cells that act as long-term or short-term memory cells. The time-related OCM input images of the LSTM network are modulated by the states of these cells. This property allows the historical state of the input image to affect the segmentation prediction. In this project, instead of segmenting the fly heart only from 2D OCM images, we further utilized the time-dependent information contained in the fly heart OCM videos. Employing the convolutional LSTM network found correlations between single fly heart OCM images separated by many frames, and these correlations can help us extract time-dependent information features of the fly heartbeat. With this well-trained FlyNet 2.0, we examined fly heart OCM videos taken in three developmental stages and two heartbeat situations. Further, dynamic cardiac parameters, such as the EDD, ESD, heart area, and heart rate, were calculated to further demonstrate that FlyNet 2.0 can segment fly heart OCM videos accurately.

## 2. Methods

### 2.1. Training data and training strategy

We acquired fruit fly heartbeat videos with the custom OCM system [10] shown in Fig. 1. The OCM system uses a broadband supercontinuum laser source with a central wavelength at 840 nm and has a sensitivity of ~94 dB. The axial resolution was measured to be ~1.5  $\mu\text{m}$ , with a transverse resolution of ~5  $\mu\text{m}$ .

We collected OCM images from 100 different flies in three different developmental stages (larva, early pupa, and adult). The datasets include 35 larva, 35 early pupa, and 30 adult flies. The fly heartbeat videos for different developmental stages were equally distributed for training, validating, and testing dataset. Table 1 shows how datasets were distributed for training, validating,



**Fig. 1.** The OCM setup uses a super continuum light source with a central wavelength of 840 nm. A rod mirror splits the laser beam into sample and reference arms.

and testing with different developmental stages. To record fly heartbeats over time, each OCM video includes either 4000 or 6000 frames of 2D fly heart images which were continuously imaged over 35 or 60 seconds, respectively. We generated a large training database for fruit fly heart segmentation, including 500,000 well-masked fruit fly heart images. Our strategy to generate the training dataset was first to mask 20,000 fruit fly heart images with Amira software manually, then to use these 20,000 images as a starting point to train with the CNN model, an updated version of FlyNet 1.0 [35]. Then we segmented the fly heartbeat OCM videos with this CNN based model. However, the IOU segmentation accuracies of these videos were only around 80%, not good enough for FlyNet 2.0 training; hence we wrote segmentation software in C++ to help manually revise the segmentation results from the CNN model. The corrected masks and OCM videos were used as training and testing data for FlyNet 2.0. With this method, we quickly generated a large training database for fruit fly heart segmentation, including 500,000 well-masked fruit fly heart images as ground truth.

**Table 1. Datasets for different developmental stages of fly**

	Training	Validating	Testing
Larva	140,000	20,000	15,000
Early Pupa	140,000	20,000	15,000
Adult	120,000	15,000	15,000

Images from the same long OCM videos were used as either training or testing data. No testing data were used during the training process. We resized every fly video into 10 or 15 short videos, each including 400 frames. For each short video, we selected 32 temporally adjacent frames as one training data batch. In order to increase the number of datasets, the data was randomly augmented spatially and temporally by one of the following four methods (1) random image flipping, (2) random image rotation (3) contrast adjustment (4) switching the temporal down-sampling factor (indicating the number between the adjacent frames of the training input) between 1 ~ 4. We adjusted the contrast by multiplying and adding gain and bias parameters for

each pixel in the OCM images. The equation was shown below:

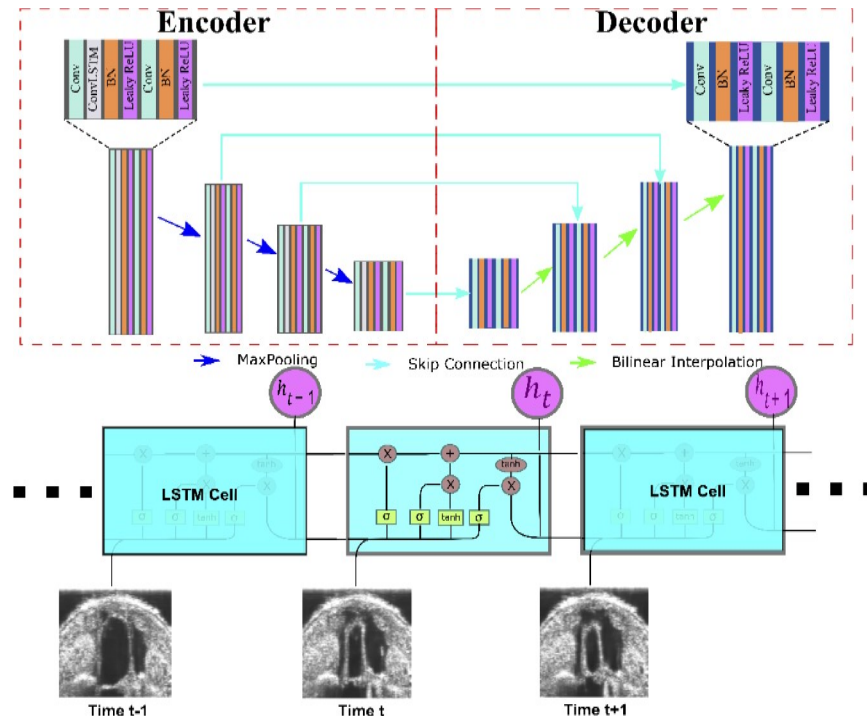
$$g(i,j) = \alpha f(i,j) + \beta, \quad (1)$$

where  $i$  and  $j$  indicate that the pixel is in the  $i$ -th row and  $j$ -th column,  $\alpha$  is the gain parameter, and the  $\beta$  is the bias parameter. So, after image augmentation, in total, we had 2000 training videos including 800,000 images, 275 validating videos including 110,000 images, and 225 testing videos including 90,000 images.

## 2.2. Network architecture

Our FlyNet 2.0 integrates convolutional LSTM blocks into the U-Net architecture [34,37]. The U-Net architecture, built as an encoder-decoder structure with skip connections, enables the extraction of spatial information at different image scales. By retaining extracted representations in the LSTM memory units, LSTM allows the consideration of the former cell appearances at multiple scales. We use a combination of convolutional LSTM layers in every scale of the encoder of the U-Net.

As shown in Fig. 2, FlyNet 2.0 includes four encoder and decoder blocks [37]. Each block in the encoder includes one convolutional LSTM layer, two leaky ReLU layers with a 0.3 negative slope, two convolutional layers, two batch normalization (BN) layers, and each blocks finally downsampled using the MaxPooling operation. The decoder blocks consist of a bi-linear interpolation, a concatenation with the parallel encoder block, and two convolutional layers, batch normalization, and leaky ReLU. Each convolutional layer has a kernel size of  $3 \times 3$ , with layer



**Fig. 2.** FlyNet 2.0 structure for fruit fly heart segmentation. The left four plate blocks indicate the encoder, and the right four plate blocks indicate the decoder. Each plate block in the encoder includes the LSTM layer to extract time sequence information from fly heart videos. Time sequenced input OCM images are connected by the LSTM cells.

depths of 128, 256, 512, and 1024. Each MaxPooling layer uses a kernel size of  $2 \times 2$  without overlap. Each convolutional LSTM kernel is  $5 \times 5$ , with layer depths of 128, 256, 512, and 1024. The number of filters is symmetric between encoder and decoder. The last convolutional layer is input to a softmax layer to produce the final probabilities.

### 2.3. Loss functions and training parameters

For the training process, intersection over union (IOU) was used as the accuracy metric to evaluate the performance of the FlyNet 2.0 network. IOU, as shown below, describes the similarity between the ground truth and the predicted result:

$$IOU = \frac{I_{predicted} \cap I_{GT}}{I_{predicted} \cup I_{GT}}, \quad (2)$$

where  $I_{predict}$  and  $I_{GT}$  represent the predicted result and ground truth result, respectively. A differentiable soft IOU score was used as our loss function during the training process and the Adam optimizer [38] was used to optimize the training process. The learning rate for FlyNet 2.0 was set to  $10^{-5}$ . Early stopping strategy was employed to improve training efficiency. The program stopped training if the validating loss value improved less than  $10^{-4}$  for 3 epochs. We saved the final model with the lowest validation loss.

## 3. Results

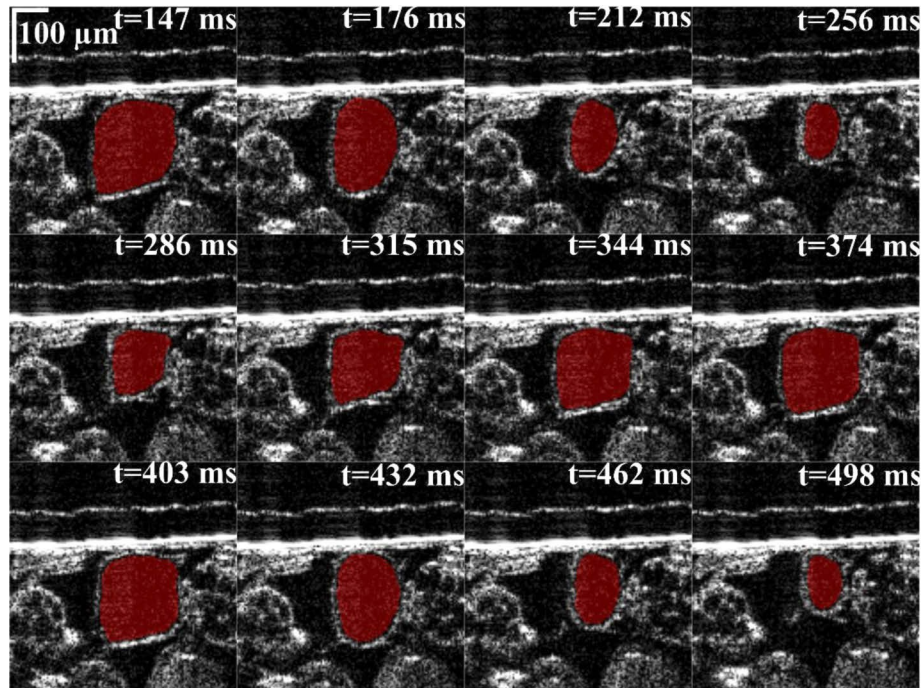
We trained fly heart OCM images on a single NVIDIA GeForce GTX 1080 GPU with 8 GB of memory. With reference to the IOU test, the segmentation accuracy of FlyNet 2.0 was 92% which improved over the 86% value of the FlyNet 1.0 (re-trained with the same data as the FlyNet 2.0). FlyNet 2.0 employs convolutional LSTM layers to extract the temporal information contained between adjacent fly heart images. The convolutional LSTM layers were inserted in the encoder, as shown in the network structure. However, different combinations of convolutional LSTM layers were tested to compare their performances during the training process. Three different network structures were tested: (1) the convolutional LSTM layers only in the encoder, (2) the convolutional LSTM layers only in the decoder, (3) the convolutional LSTM layers in both the encoder and decoder. We tested all three structures with the same training dataset, and the loss functions were consistent. The testing IOU accuracy results for all three structures are shown below in Table 2. The convolutional LSTM layer in the encoder achieved the best segmentation accuracy. Hence in our FlyNet 2.0, we inserted the convolutional LSTM layers in the encoder to achieve higher training accuracy.

**Table 2. IOU accuracy of three different network structures**

	LSTM in encoder	LSTM in decoder	LSTM in both
IOU Accuracy	92%	81%	85%

We tested FlyNet 2.0 with fly heartbeat OCM videos from different developmental stages and with different heartbeat situations. Figure 3, 4, and 5 show segmentation results of several heartbeat cycles for larva, early pupa, and adult fruit flies, respectively. For each developmental stage, twelve OCM images were selected to demonstrate heart contraction and dilation. Figure 3 shows the segmentation result of a larva's heartbeat cycles. In the larva stage, the shape of the fly heart changes during the heartbeat. With FlyNet 2.0, the fly heart can be accurately labeled at different time stamps, even though positions and shapes have changed. To characterize *Drosophila*'s heart fitness, it is important to analyze certain functional parameters [39], such as the end-diastolic diameter (EDD), indicating the diameter during the heart dilation, and the end-systolic diameter (ESD), indicating the diameter during heart contraction. These parameters

can be calculated based on the segmentation output of the neural network. As shown in Fig. 3, at 147 ms the fly heart dilation area achieves its maximum. The EDD value for this beating cycle is equal to the maximum distance of the red labeled pixels in the vertical direction. At 256 ms within the same beating cycle, the fly heart contraction area achieves its minimum. The ESD value is equal to the maximum distance of the red labeled pixels in the vertical direction. The EDD and ESD values of the larva stage at 147 ms and 256 ms are 159  $\mu\text{m}$  and 92  $\mu\text{m}$ , respectively. In comparison, the EDD and ESD values of the ground truth are 164  $\mu\text{m}$  and 95  $\mu\text{m}$ .



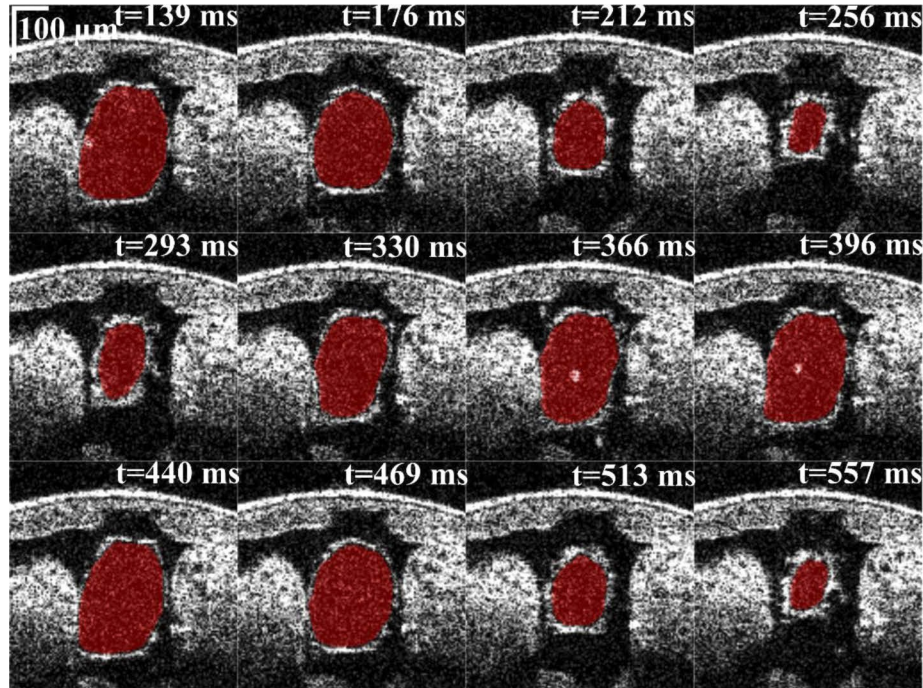
**Fig. 3.** FlyNet 2.0 segmentation results for larva heartbeat cycles (see [Visualization 1](#)). Twelve frames from [Visualization 1](#) are shown.

Figure 4 shows the segmentation results for a fly heart contraction in the early pupa stage. With FlyNet 2.0, the hearts were accurately masked for all contraction stages, as shown in Fig. 4. The EDD and ESD values of the early pupa stage at 139 ms and 256 ms are 247  $\mu\text{m}$  and 114  $\mu\text{m}$ , the same as the ground truth values.

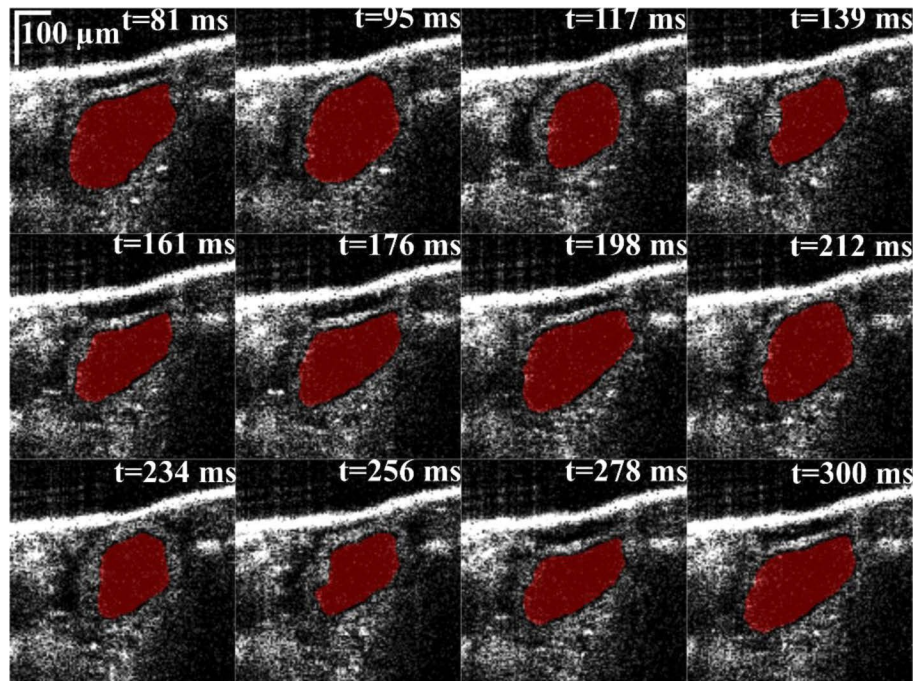
Figure 5 shows the segmentation results for an adult fly. The adult fly has a dark cuticle which increases light absorption and scattering. Therefore, the backscattered intensity from the fly inner organ is much lower compared with the larva and early pupa flies. Lower backscattered intensity influences the signal-to-noise ratio and contrast of the OCM images. Thus, the boundaries of adult fly hearts are not easy to observe, making it difficult to detect and label the fly heart from only spatial information. By adding temporal information, FlyNet 2.0 can accurately segment the adult fly heart at different contraction stages. The EDD and ESD values of an adult fly at 198 ms and 140 ms are 198  $\mu\text{m}$  and 140  $\mu\text{m}$ , which are slightly less than the ground truth results of 203  $\mu\text{m}$  and 144  $\mu\text{m}$ .

Besides testing flies at different development stages, two different heartbeat situations were selected to test with FlyNet 2.0. The first situation was a steady heartbeat throughout the whole OCM video, and the second was a heartbeat frequency that changed during the whole OCM video due to optogenetic pacing [40]. For each situation, a 4000 or 6000 frames fly heartbeat video was segmented with FlyNet 2.0. Based on the segmentation results, the fly heart area was





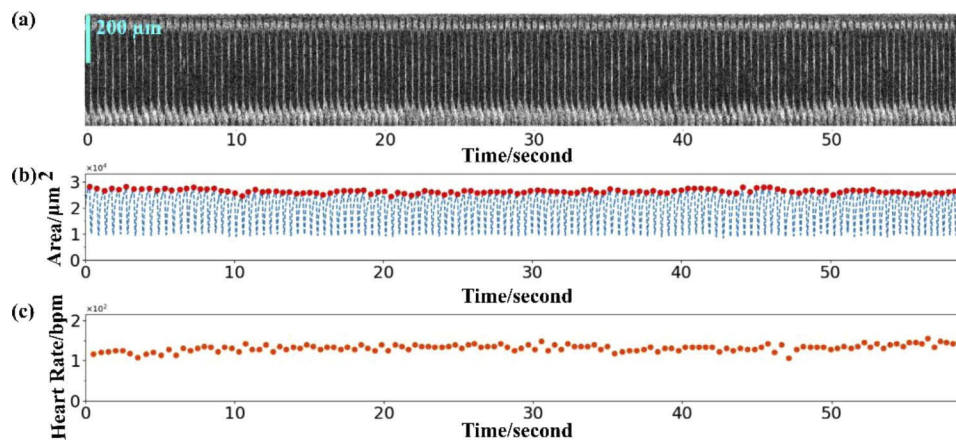
**Fig. 4.** FlyNet 2.0 segmentation results for early pupa heartbeat cycles (see [Visualization 2](#)). Twelve frames from [Visualization 2](#) are shown.



**Fig. 5.** FlyNet 2.0 segmentation results for adult heartbeat cycles (see [Visualization 3](#)). Twelve frames from [Visualization 3](#) are shown.

calculated for each frame. The area of the fly heart lumen in each OCM image can be calculated as the summation of the red labeled pixels. From the area plot, we can characterize the heart rate. The peak values in the area plot represent the maximally dilated heart area for every contraction cycle. The contraction period is the distance between two adjacent peak points in the area plot, and the heart rate can be calculated as the reciprocal of the contraction period. For each input video, we concatenated all the OCM frames to generate an M-mode image showing how the fly heart changed during the long recording time. Based on the segmentation results, the fly heart area value and its corresponding heart rate for each frame were plotted to see if the segmentation results were consistent with the M-mode image.

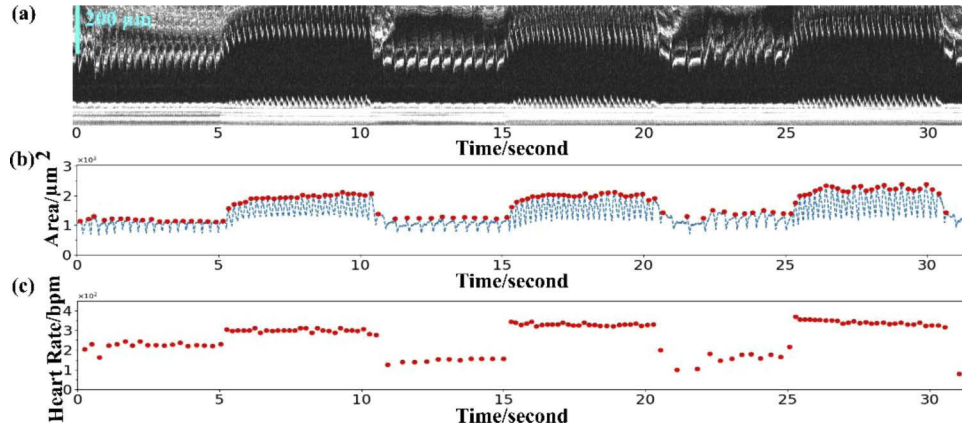
Figure 6 shows the first heartbeat situation, in which the fly heartbeat frequency, at the resting heart rate, is uniformly distributed with time. The fly heart video includes 6000 frames recording the heartbeat for 60 seconds. Figure 6(a) is an M-mode image of the early pupa fly heartbeat video, showing that the fly heart contracts regularly and is not influenced by the laser beam used for OCM imaging. The shapes and areas for all the contraction cycles are similar. FlyNet 2.0 was employed to segment the whole OCM video. Based on the segmentation results, we calculated the heart area for each frame. Figure 6(b) shows the area plot for the 6000 frames of the fly heartbeat video. Fly heart areas are regularly distributed for 60 seconds, a result consistent with the M-mode image. Figure 6(c) shows the heart rate plot based on the segmentation results. The heart rate is around 100 beats per minute for the fly heartbeat video.



**Fig. 6.** Fly heart OCM image, area plot, and heart rate plot with a uniform fly heartbeat frequency (see Visualization 4). (a) M-mode image generated from a 2D OCM video of an early pupa. (b) Fly heart area plot. (c) Fly heart rate plot.

The second example involves the alteration of the heartbeat frequency over the course of the OCM video of a larva, driven by optogenetic pacing [40]. The OCM video includes 4000 frames recorded over 35 seconds. Figure 7(a) clearly shows that the frequency changed three times during the recording. From the 5<sup>th</sup> second to the 10<sup>th</sup> second, from the 15<sup>th</sup> second to the 20<sup>th</sup> second, and from the 25<sup>th</sup> second to the 30<sup>th</sup> second, we employed optogenetic pacing to speed up the heartbeat frequency. During the pacing time, not only was the OCM laser beam used to image the fly heart, but we also illuminated the fly heart with a flashing 617 nm LED light at 5 Hz, 5.5 Hz, and 6 Hz, respectively. When this red LED light was on, the fly heartbeats at the same pace as the flashing frequency of the LED. Figure 7(b) and (c) respectively show the fly heart area and heart rate plots based on the segmentation results. The plots clearly show that the fly heartbeat frequency increases during the pacing time. From the 10<sup>th</sup> second to the 15<sup>th</sup> second and from the 20<sup>th</sup> second to the 25<sup>th</sup> second, the fly heart needs time to recover to its resting rate. After each tachypacing, the heart rate slows down below the resting heart rate then

gradually recovers. Immediately after the repeated pacing challenge, the heartbeat further slows down and takes a longer time to recover [40]. Tachypacing challenges the heartbeat by forcing the fly heart to follow the pacing frequencies, which induces stress on the heart. After the second tachypacing, the cumulative stress from multiple challenges causes the heart to take a longer time to recover and leads to a larger variation. Figure 7 clearly demonstrates that FlyNet 2.0 can accurately segment fly heart OCM videos, even when the fly heartbeat frequency changes.



**Fig. 7.** Fly heart OCM image, area plot, and heart rate plot with fly heartbeat frequency changes (see Visualization 5). (a) M-mode image generated from a 2D OCM video of a larva. (b) Fly heart area plot. (c) Fly heart rate plot.

#### 4. Discussion

Our custom FlyNet 2.0 software employs both spatial and temporal information to segment images of the fruit fly heart in OCM videos. Compared with FlyNet 1.0, FlyNet 2.0 has two improvements in its neural network structure design. First, for each convolutional block in both the encoder or decoder of the U-Net, we added two batch normalization layers and two leaky Relu layers. Leaky Relu layers [41,42] can accelerate the training process and keep the back-propagation optimization from getting stuck [42]. The batch normalization layers [43] provide regularization and also speed up the training time. The second improvement adds convolutional LSTM layers in the encoder of the network. We decided to insert convolutional LSTM layers in the encoder by trial-and-error. There are two possible reasons that the network with the convolutional LSTM layers in the encoder achieves the best IOU accuracy. First, adding the convolutional LSTM layers to the encoder adds more degrees of freedom since skip connections were used to sum the outputs of the encoder at different scales with the corresponding outputs of the decoder [44,45]. Second, the decoder employs the bilinear interpolation to increase the image scales [46]. Adding the convolutional LSTM layers in the decoder increase the loss value for our application. With the convolutional LSTM layers, both spatial and temporal information can be extracted to increase the segmentation accuracy of the fly heart OCM videos.

Time sequence information is widely used in image segmentation to achieve high segmentation accuracy. A recurrent neural network (RNN) [47] copies the same network multiple times for adjacent time-sequence images, and each copy of the network passes the output information to its successor. However, vanishing or exploding gradient problems can lead to training fails for RNN, due to the large time sequence datasets [48]. The Gated Recurrent Unit (GRU) [49] is a simple version of LSTM that employs time sequences information. Due to the smaller number of training parameters, the GRU model is faster than LSTM. On the other hand, with enough computational power and training data, LSTM achieves better performance than GRU, as

shown in empirical studies [50]. Convolutional LSTM [51] can extract both spatial and temporal features and is widely used for image segmentation with time sequence inputs. Convolutional LSTM can be integrated with different classical segmentation CNN networks, such as a Fully Convolutional network [52], and SegNet [53]. On the other hand, LSTMs can be separated from the CNN network: for example, Gopinath et al. [54] built three cascaded networks to segment retinal layers. Two CNN-based networks were used for feature extraction and edge detection, and the LSTM-based network was used for boundary tracing. For FlyNet 2.0, convolutional LSTM layers were integrated with the U-Net structure to generate a compact network. Hence FlyNet2.0 has the advantages of both LSTM and U-Net, achieving high segmentation accuracy with both spatial and temporal information.

Another important parameter for FlyNet 2.0 is processing time. We tested a fly heartbeat video with 4000 frames ( $128 \times 128$  pixels for each frame) on a single NVIDIA GeForce GTX 1080 GPU with 8 GB of memory. The total processing time using FlyNet 2.0 was 12 seconds. Including the functional parameter calculation based on the segmentation results, the prediction took about 14 seconds in all. These results demonstrate that FlyNet 2.0 is fast and sufficiently robust for fly heartbeat segmentation applications.

Although the FlyNet 2.0 network has demonstrated better fly heart segmentation results than FlyNet 1.0, there still is room to improve its performance. In future work, more training dataset of fly OCM videos with more accurate masks, in different development stages and with different heartbeat situations, can be added to further improve the segmentation accuracy. In addition, Field Programmable Gate Arrays (FPGA) [55,56] has been used successfully for various deep-learning applications. We can use FPGA to achieve even faster image processing and reduce latency in image processing.

## 5. Conclusion

In conclusion, we have proposed a segmentation model, FlyNet 2.0, to segment fruit fly heart OCM videos. Based on the previous FlyNet model, this new FlyNet 2.0 utilizes LSTM to extract both spatial and temporal information to improve segmentation accuracy. Different fruit fly heart OCM videos with different heartbeat situations were segmented using FlyNet 2.0. The segmentation IOU accuracy of FlyNet 2.0 was improved over the 86% value of the initial CNN model, achieving 92%. With these accurate segmentation results, fruit fly heart functional parameters, such as the EDD, ESD, heart area, and heart rate, can be accurately and efficiently calculated for cardiac disease research.

## Funding

National Institutes of Health (R01EB025209, R15EB019704); National Science Foundation (DBI-1455613, IIP-1640707).

## Acknowledgment

The authors would like to thank James Ballard, Guangming Ni, Jinyun Zou, Hongwu Liang, Boyang Zhang, and Suya Li for helpful discussions and assistance.

## Disclosures

The authors declare that there are no conflicts of interest related to this article.

## References

1. R. Bodmer and T. V. Venkatesh, "Heart development in *Drosophila* and vertebrates: Conservation of molecular mechanisms," *Dev. Genet.* **22**(3), 181–186 (1998).

2. N. Piazza and R. J. Wessells, "Drosophila models of cardiac disease," *Prog. Mol. Biol. Transl. Sci.* **100**, 155–210 (2011).
3. U. B. Pandey and C. D. Nichols, "Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery," *Pharmacol. Rev.* **63**(2), 411–436 (2011).
4. L. T. Reiter, L. Potocki, S. Chien, M. Gribskov, and E. Bier, "A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*," *Genome Res.* **11**(6), 1114–1125 (2001).
5. M. Wolf, H. Amrein, J. Izatt, M. Choma, M. Reedy, and H. Rockman, "Drosophila as a model for the identification of genes causing adult human heart disease," *Proc. Natl. Acad. Sci. U. S. A.* **103**(5), 1394–1399 (2006).
6. 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).
7. A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recogn.* **36**(2), 451–461 (2003).
8. 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. Visualized Exp.* **118**, 55002 (2016).
9. J. Men, Y. Huang, J. Solanki, X. Zeng, A. Alex, J. Jerwick, Z. Zhang, R. E. Tanzi, A. Li, and C. Zhou, "Optical Coherence Tomography for Brain Imaging and Developmental Biology," *IEEE J. Sel. Top. Quantum Electron.* **22**(4), 1–13 (2016).
10. A. Alex, A. Li, R. E. Tanzi, and C. Zhou, "Optogenetic pacing in *Drosophila melanogaster*," *Sci. Adv.* **1**(9), e1500639 (2015).
11. 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).
12. M. A. Choma, M. V. Sarunic, C. Yang, and J. A. Izatt, "Sensitivity advantage of swept source and Fourier domain optical coherence tomography," *Opt. Express* **11**(18), 2183–2189 (2003).
13. J. F. de Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," *Opt. Lett.* **28**(21), 2067–2069 (2003).
14. M. Wojtkowski, "High-speed optical coherence tomography: basics and applications," *Appl. Opt.* **49**(16), D30–D61 (2010).
15. T. Klein, W. Wieser, L. Reznicek, A. Neubauer, A. Kampik, and R. Huber, "Multi-MHz retinal OCT," *Biomed. Opt. Express* **4**(10), 1890–1908 (2013).
16. I. Grulkowski, J. J. Liu, B. Potsaid, V. Jayaraman, C. D. Lu, J. Jiang, A. E. Cable, J. S. Duker, and J. G. Fujimoto, "Retinal, anterior segment and full eye imaging using ultrahigh speed swept source OCT with vertical-cavity surface emitting lasers," *Biomed. Opt. Express* **3**(11), 2733–2751 (2012).
17. T. Klein, W. Wieser, C. M. Eigenwillig, B. R. Biedermann, and R. Huber, "Megahertz OCT for ultrawide-field retinal imaging with a 1050 nm Fourier domain mode-locked laser," *Opt. Express* **19**(4), 3044–3062 (2011).
18. J. Reiber, S. Tu, J. Tuinenburg, G. Koning, J. Janssen, and J. Dijkstra, "QCA, IVUS and OCT in interventional cardiology in 2011," *Cardiovasc. Diagn. Ther.* **1**(1), 57–70 (2011).
19. T.-H. Tsai, B. Potsaid, Y. K. Tao, V. Jayaraman, J. Jiang, P. J. S. Heim, M. F. Kraus, C. Zhou, J. Hornegger, H. Mashimo, A. E. Cable, and J. G. Fujimoto, "Ultrahigh speed endoscopic optical coherence tomography using micromotor imaging catheter and VCSEL technology," *Biomed. Opt. Express* **4**(7), 1119–1132 (2013).
20. T. Gambichler, G. Moussa, M. Sand, D. Sand, P. Altmeyer, and K. Hoffmann, "Applications of optical coherence tomography in dermatology," *J. Dermatol. Sci.* **40**(2), 85–94 (2005).
21. J. Welzel, "Optical coherence tomography in dermatology: a review," *Skin Res. Technol.* **7**(1), 1–9 (2001).
22. L. L. Otis, M. J. Everett, U. S. Sathyam, and B. W. Colston, "Optical coherence tomography: a new imaging: technology for dentistry," *J. Am. Dent. Assoc.* **131**(4), 511–514 (2000).
23. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, 886–893 vol. 881.
24. L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting people using mutually consistent poselet activations," in *Proceedings of the 11th European conference on Computer vision: Part VI*, (Springer-Verlag, Heraklion, Crete, Greece, 2010), pp. 168–181.
25. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
26. E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, 1508–1515 Vol. 1502.
27. L. Zheng, G. Li, and Y. Bao, "Improvement of grayscale image 2D maximum entropy threshold segmentation method," in *2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)*, 2010, 324–328.
28. S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imaging* **20**(6), 490–498 (2001).
29. 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, 703–706.

30. J. A. K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Process. Lett.* **9**(3), 293–300 (1999).
31. E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 640–651 (2017).
32. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* **60**(6), 84–90 (2017).
33. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
34. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Springer International Publishing, 2015), 234–241.
35. L. Duan, X. Qin, Y. He, X. Sang, J. Pan, T. Xu, J. Men, R. Tanzi, A. Li, Y. Ma, and C. Zhou, "Segmentation of Drosophila Heart in Optical Coherence Microscopy Images Using Convolutional Neural Networks," *J. Biophoton.* **11**, e201800146 (2018).
36. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," in *IET Conference Proceedings*, (Institution of Engineering and Technology, 1999), pp. 850–855.
37. A. Arbellet and T. R. Raviv, "Microscopy Cell Segmentation via Convolutional LSTM Networks," arXiv preprint arXiv:1805.11247 (2018).
38. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations (2014).
39. C. D. Nichols, J. Becnel, and U. B. Pandey, "Methods to assay Drosophila behavior," *J. Visualized Exp.* **61**, 3795 (2012).
40. J. Men, A. Li, J. Jerwick, Z. Li, R. Tanzi, and C. Zhou, "Non-invasive red-light optogenetic control of Drosophila cardiac function," bioRxiv, <https://doi.org/10.1101/2020.01.26.920132> (2020).
41. G. Xavier, B. Antoine, and B. Yoshua, "Deep Sparse Rectifier Neural Networks," (PMLR, 2011), pp. 315–323.
42. B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," (2015).
43. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," arXiv:1502.03167 (2015).
44. M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications* (Springer, 2016), pp. 179–187.
45. A. E. Orhan and X. Pitkow, "Skip connections eliminate singularities," arXiv preprint arXiv:1701.09175 (2017).
46. H. Dianyan, "Comparison of Commonly Used Image Interpolation Methods," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, (Atlantis Press, 2013).
47. J. L. Elman, "Finding structure in time," *Cogn. Sci.* **14**(2), 179–211 (1990).
48. S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *Int. J. Unc. Fuzz. Knowl. Based Syst.* **6**(2), 107–116 (1998).
49. J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," CoRR **abs/1412.3555** (2014).
50. R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, (JMLR.org, Lille, France, 2015), pp. 2342–2350.
51. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM Network: a machine learning approach for precipitation nowcasting," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, (MIT Press, Montreal, Canada, 2015), pp. 802–810.
52. Y. Gao, J. M. Phillips, Y. Zheng, R. Min, P. T. Fletcher, and G. Gerig, "Fully convolutional structured LSTM networks for joint 4D medical image segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (2018), 1104–1108.
53. A. Pfeuffer, K. Schulz, and K. Dietmayer, "Semantic Segmentation of Video Sequences with Convolutional LSTMs," arXiv:1905.01058 (2019).
54. K. Gopinath, S. B. Rangrej, and J. Sivaswamy, "A Deep Learning Framework for Segmentation of Retinal Layers from OCT Images," in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, (2017), 888–893.
55. C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, "Optimizing FPGA-based Accelerator Design for Deep Convolutional Neural Networks," in *Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, (ACM, Monterey, California, USA, 2015), pp. 161–170.
56. S. Asano, T. Maruyama, and Y. Yamaguchi, "Performance comparison of FPGA, GPU and CPU in image processing," in *2009 International Conference on Field Programmable Logic and Applications*, (2009), 126–131.