Title

Assessing Cardiac Functions of Zebrafish from Echocardiography Using Deep Learning

Author

Huang, Mao-Hsiang

Publication Date

2023

UNIVERSITY OF CALIFORNIA,
IRVINE

Assessing Cardiac Functions of Zebrafish from Echocardiography Using Deep Learning

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Electrical and Computer Engineering

by

Mao-Hsiang Huang

Thesis Committee:
Associate Professor Hung Cao, Chair
Chancellor's Professor Nikil Dutt
Assistant Professor of Teaching Quoc-Viet Dang

2023

DEDICATION

To my parents and friends

in recognition of their worth

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Foremost, I extend my profound gratitude to my committee chair, Dr. Hung Cao, whose intellectual prowess and insightful direction have been instrumental in completing this study. This thesis came to fruition with his unwavering guidance, continuous support, and patient mentoring. I am deeply indebted to him for his invaluable contributions.

Next, I would like to acknowledge the significant contributions of my committee members, Dr. Nikil Dutt and Dr. Quoc-Viet Dang, who provided invaluable insights that have been pivotal to my work.

I am also immensely grateful to Dr. Xiaolei Xu and Dr. Ping Zhu, both esteemed members of the Mayo Clinic. Their generosity in providing access to crucial data, suggestions for necessary revisions, and overall support throughout the process has been of immense help.

Finally, I would like to extend my appreciation to the dedicated members of our lab - Amir, Xing, and Khuong. Their detailed reviews and constructive comments significantly improved the quality of the final manuscript. Their experimental support, teamwork, and suggestions for research resources were indispensable for the successful execution of this research project.

ABSTRACT OF THE THESIS

Assessing Cardiac Functions of Zebrafish from Echocardiography Using Deep Learning

by

Mao-Hsiang Huang

Master of Science in Electrical and Computer Engineering

University of California, Irvine, 2023

Professor Hung Cao, Chair

Zebrafish, revered as an invaluable model organism, is widely employed in cardiovascular disease research. It facilitates a comprehensive understanding of cardiac behaviors and conditions by appraising cardiac functions extracted from heartbeat echo-videos. Researchers often scrutinize the ejection fraction, a metric indicative of heart performance. Nevertheless, current techniques for such evaluations grapple with numerous challenges. These methods are laborious, time-consuming, and prone to errors, making them unfavorable for large-scale investigations. These limitations are particularly troublesome when dealing with massive datasets or when detailed assessments are required, hindering their effectiveness in applications like high-throughput screening for drug discovery. Addressing these constraints, an approach was conceived to enhance cardiac function analysis in zebrafish. This method utilizes a deep learning model to enable the automatic evaluation of ejection fractions from heart echo-videos.

This thesis outlines an approach hinging on a specific deep learning model architecture. The model's accuracy, confirmed using the Dice coefficient and the Intersection over Union (IoU) score, stood at a robust 0.967 and a significant 0.937, respectively. The testing phase yielded a promising error rate range from 0.11% to 16.96%, averaging 5.13%, attesting to the method's accuracy and reliability. Furthermore, this method can be assimilated into existing lab settings, synergizing with binary recordings to optimize large-scale video analysis and improve high-throughput screenings' efficacy. Compared to traditional techniques, this deep learning-based method simplifies zebrafish cardiac function monitoring and quantification, thus boosting laboratory efficiency. In conclusion, this approach heralds a notable advancement in zebrafish cardiovascular research. Enhancing the speed, accuracy, and ease of cardiac function analysis holds significant potential to transform the study of cardiovascular diseases, serving as a crucial tool for researchers in this pivotal field.

CHAPTER 1 – INTRODUCTION

Cardiovascular disease (CVD) remains an escalating global health concern, representing a significant public health challenge. It is a leading cause of morbidity and mortality, with an alarming 17.9 million deaths annually, equating to roughly 32% of all global fatalities [1]. CVD persists despite the significant financial resources dedicated to medical research to curb this health crisis. The persistently high rates of CVD incidence necessitate ceaseless efforts from the scientific community. Researchers consistently seek innovative strategies to understand heart diseases and potential therapies further. Zebrafish (Danio rerio) have surfaced in this endeavor as a beneficial and unique model organism to bolster cardiovascular research [2]. Zebrafish offer multiple benefits over traditional animal models in cardiovascular studies. First, these tropical fish mature rapidly within 2-3 months, making them ideal for high-throughput genetic and drug screening studies. Second, they require standard aquarium conditions and regular feeding, making maintenance straightforward and cost-effective.

Notably, zebrafish exhibit a high degree of genetic similarity to humans, with about 70% of human genes having a zebrafish analog. This facilitates a comprehensive exploration of genetic factors contributing to CVD. Zebrafish stand out due to the optically transparent nature of their embryos, enabling non-invasive observation of cardiac development. This transparency allows real-time heart formation observation using standard microscopic techniques [3]. Such visualization and tracking of cardiac development and function allow dynamic and longitudinal monitoring of disease progression and therapeutic interventions. Furthermore, zebrafish lay hundreds of eggs weekly, accelerating genetic studies. The swift generation of large offspring numbers allows efficient screening of genetic variations and deepens understanding of their impact on cardiac function. In summary, zebrafish offer multiple benefits, such as rapid development, easy maintenance, genetic similarity to humans, optical transparency, and high reproduction, making them indispensable for advancing cardiovascular disease knowledge. The use of zebrafish in cardiac research continues to grow as scientists uncover therapeutic possibilities against cardiovascular diseases.

Zebrafish and humans share striking physiological similarities, making these tropical fish invaluable for medical research, particularly cardiovascular studies. Notably, due to a similar prolonged plateau phase in both species, zebrafish's cardiac action potential phenotype closely mirrors humans' [4]. This similarity enables a deeper understanding of cardiovascular processes that are usually challenging to study in humans. Remarkably, zebrafish can regenerate damaged heart tissue, a characteristic they share with certain amphibian species. Unlike humans, where heart damage usually results in non-functional scar tissue, zebrafish can fully restore their cardiac functionality after injury. This regenerative capacity offers a powerful model for studying heart repair and regeneration, potentially leading to innovative therapeutic approaches for heart disease [5]. Zebrafish are invaluable tools for preclinical drug testing due to their genetic tractability and experimental versatility. Researchers can induce heart disease in zebrafish using genetic manipulation or chemical exposure, then thoroughly evaluate new drugs' therapeutic potential. This includes assessing the drugs' ability to alleviate disease symptoms, restore normal heart function, and promote cardiac repair [6]. An example of zebrafish's utility in therapeutic testing is antiarrhythmic drug development. Arrhythmias, irregular heart rhythms, can lead to severe complications, including heart failure and stroke. Zebrafish have successfully been used to evaluate potential antiarrhythmic drugs' efficacy [7]. In conclusion, zebrafish's physiological homology with humans, regenerative cardiac ability, and suitability for disease induction and drug testing make them invaluable model organisms in cardiovascular research. Their use is expected to grow, driving advancements in the understanding and therapeutic development of heart disease.

Over the past decade, scientists have developed various cardiomyopathy models in adult zebrafish, including models for inherited cardiomyopathies. These are linked to causative genes such as Myosin heavy chain 7 (MYH7) and Lysosome Associated Membrane Protein 2 (LAMP2) [8-10]. These models offer precious insights into heart disease's genetic and molecular mechanisms and are helpful for testing potential therapeutic interventions. The creation and subsequent analysis of cardiomyopathy models in zebrafish often critically depend on the availability of medical imaging data, which can sometimes prove limiting. This constraint is especially pronounced for echo-videos, invaluable tools for assessing cardiac structure and function. Echo-videos facilitate real-time visualization and measurement of diverse cardiac parameters, providing essential information on the heart's overall health and performance. A significant challenge posed by echo-video analysis is the requirement for manual annotation. This particular task, usually carried out by expert biologists, involves precisely labeling different video aspects. Due to the subjective nature of manual annotation, this process is labor-intensive, time-consuming, and susceptible to inconsistencies. Furthermore, when performed by untrained individuals, it can lead to errors, introducing potential challenges for subsequent data validation. However, the recent strides in deep learning methodologies offer a promising solution to these hurdles.

Deep learning, an artificial intelligence subfield, leverages the power of intricate algorithms and neural networks to facilitate continuous learning and improvement based on experiential data. It has shown tremendous potential in developing automated segmentation tools explicitly designed for medical imaging data, including echo-videos. Deep learning algorithms can be rigorously trained in medical imaging to recognize and segment-specific structures within images, such as accurately identifying the cardiac region in echo-videos [11]. This process, known as image segmentation, is crucial in computer vision and medical imaging. It involves partitioning an image into distinct regions based on semantic meaning, greatly assisting in image analysis and interpretation. In recent years, deep learning-driven image segmentation techniques have exhibited remarkable effectiveness across various applications.

The encoder-decoder network stands out among the diverse deep learning architectures employed for this task. This architecture comprises two distinct components: an encoder network, which extracts high-level features from the image, and a decoder network, which uses these features to produce a pixel-wise segmentation map [12]. In conclusion, while generating and analyzing zebrafish cardiomyopathy models present specific challenges, particularly regarding medical imaging data, recent advancements in deep learning offer a promising avenue. Utilizing sophisticated algorithms and neural networks, deep learning holds the potential to revolutionize medical imaging data analysis and interpretation. This could lead to a more accurate understanding of heart diseases, ultimately paving the way for developing innovative therapeutic strategies.

The Unet architecture, a seminal structure frequently deployed for image segmentation tasks across a broad range of domains, has demonstrated efficacy and robustness that have propelled it to become a standard reference for evaluating and comparing the performance of various image segmentation algorithms [13]. Comprised of two primary components—namely a contracting or "downsampling" path and an expanding or "upsampling" path, symmetric to one another—the Unet architecture presents a carefully balanced design. The contracting path endeavors to capture the context from the input image. This is accomplished through convolutional layers, followed by max-pooling layers, which progressively diminish the spatial dimensions of the feature maps while concurrently increasing their depth. By concentrating on the broader context, this network segment absorbs and abstracts the overarching characteristics and patterns present within the image. Conversely, the symmetric expanding path places its focus on precise localization. This pathway employs a series of up-convolution or transposed convolution operations, which progressively enlarge the spatial dimensions of the feature maps, thus empowering the network to generate a detailed segmentation map. Unet's unique design configuration enables it to capture global context information while preserving detailed spatial information. This capability renders the architecture highly suitable for segmenting objects of diverse sizes and shapes, offering a combination of generalization and specificity in its output.

Notwithstanding the proven efficacy of Unet, the potential for further enhancement remained. In pursuit of augmenting the segmentation performance of Unet, an advanced iteration—aptly named Unet++—was proposed [14]. This novel architecture introduces an innovative series of nested and dense skip connections between the encoder and decoder networks. These skip connections permit Unet++ to capture and utilize multi-scale features, resulting in heightened localization accuracy compared to the original Unet. Moreover, Unet++ integrates the concept of deep supervision into its design. This technique incorporates auxiliary segmentation branches into the network's intermediate layers, enabling Unet++ to provide earlier and more frequent feedback during the training process and assist the network in learning more efficiently. This methodology has enhanced training stability, accelerated the learning process's convergence, and ultimately contributed to superior segmentation performance. In essence, Unet++ represents a significant leap forward from the original Unet architecture, offering enhancements in feature extraction, localization accuracy, and training efficiency. These improvements render Unet++ an even more potent tool for image segmentation, capable of tackling complex segmentation tasks with remarkable accuracy and robustness.

In conjunction with these supervised learning architectures, unsupervised image and video segmentation models have increasingly drawn interest. Unsupervised learning models, which operate independently of labeled training data, have been effectively implemented with medical datasets, thus offering an alternative technique for segmenting medical images. Utilizing unsupervised segmentation, medical images can be processed in a manner that is both efficient and precise with skipping the labeling process.

In this thesis, experiments involving unsupervised segmentation and supervised image segmentation were conducted. Among these diverse methodologies, supervised image segmentation emerged as the most promising technique for the task of heart segmentation in echo-videos and the evaluation of the ejection fraction. Within this context, the supervised image segmentation approach exhibited commendable accuracy, furnishing a robust and reliable method for the evaluation of the ejection fraction. In summation, deep learning-based image segmentation architectures, such as Unet and Unet++, have displayed extraordinary promise in medical imaging applications, notably in tasks related to object segmentation and ejection fraction evaluation. While unsupervised models also present unique advantages, supervised image segmentation has demonstrated itself to be particularly effective in tasks that demand high levels of precision and accuracy.

CHAPTER 2 – Materials and Methods

2.1 Experimental animals

Zebrafish (Danio rerio) were maintained under a 14 h light/10 h dark cycle at 28.5 ◦ C. All animal study procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals published by the U.S. National Institutes of Health (NIH Publication No. 85-23, revised 1996). Animal study protocols were approved by the Mayo Clinic Institutional Animal Care and Use Committee (IACUC #A00002783-17-R20).

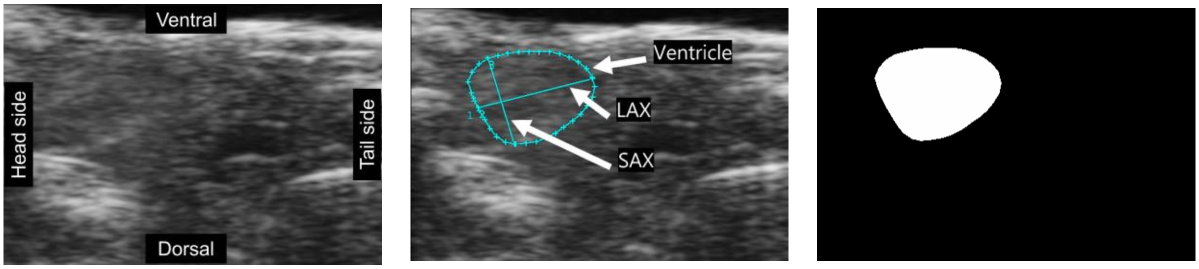
2.2 Dataset

The present study utilized a dataset consisting of echo-videos of zebrafish. The production of video files was achieved by utilizing the Vevo 3100 high-frequency imaging system, which is equipped with a 50 MHz (MX700) linear array transducer (manufactured by FUJIFILM VisualSonics Inc). This advanced imaging system was employed to precisely measure cardiac function indices in adult zebrafish of varying ages and mutant types. To ensure the highest level of image clarity and detail, acoustic gel (specifically, Aquasonic® 100, produced by Parker Laboratories) was applied to the surface of the transducer, promoting optimal coupling between the transducer and the tissue interface. To obtain these images, the adult zebrafish were anesthetized by using a 0.02% tricaine concentration, which lasted approximately five minutes. Subsequently, each zebrafish was placed ventral side up and held firmly, yet gently, in place using a soft sponge stage.

Regarding image acquisition, the 50 MHz (MX700) transducer was positioned directly above the zebrafish, allowing for the clear capture of images from the sagittal imaging plane of the heart. B-mode images were within an imaging field of view of 9.00 mm in the axial direction and 5.73 mm in the lateral direction. Additionally, the frame rate was 123 Hz, used medium persistence, and set the transmit focus at the heart's center to ensure the utmost image clarity and accuracy. The data acquisition and subsequent processing followed protocols as outlined in a report available in the existing literature [15]. The application of doxorubicin was also involved, a compound recognized for its capacity to induce cardiomyopathy in adult zebrafish [16,17]. Doxorubicin was administered intraperitoneally at a dosage of 20 mg/kg. The ejection fraction decline, a key measure of cardiac function, became detectable via echocardiography 56 days post-injection (dpi). In the orientation of the zebrafish within the echocardiography apparatus, a consistent protocol whereby the zebrafish was oriented with its head to the left and ventral side facing upwards was adhered to. This consistent positioning facilitated reproducibility and standardization across all imaging procedures.

The present study utilized a dataset composed of 1005 meticulously curated frames sourced from 164 echo-videos, with individual videos contributing between 2 to 10 frames. This ensured a diverse and representative data aggregation for the research study, as depicted in **Figure 1**. The selection process was accounting for various influential factors such as the visibility and clarity of the heart and ventricular boundaries, alongside the diversity of heart states and rhythms. Once the frame selection process was completed, each chosen frame underwent a comprehensive manual annotation process. An expert collaborative researcher carried out this task—a highly trained biologist with considerable field experience. Vevo Lab software was the tool of choice for this crucial step. This software, noted for its precision and extensive features, empowered the biologist to accurately delineate the requisite anatomical boundaries within each frame. The resultant annotated images featured cyan lines to depict the ventricular boundaries and denote the ventricle's long axis (LAX) and short axis (SAX). Such annotations are essential for training the deep learning model, as they function as the 'ground truth' from which the algorithm derives its learning. Upon completing the annotation phase, corresponding masks representing the ventricular area were generated. This was facilitated by ImageJ software, an open-source platform with a broad spectrum of features apt for the analysis and processing of scientific images. These masks offer a binary representation of the ventricular area, a fundamental input for the segmentation model training process.

To augment the richness and diversity of the dataset—thus enhancing the robustness of the deep learning model—this study incorporated data augmentation techniques. These techniques encompass operations such as rotation, random brightness, sharpening, elastic transform, etc., and serve to inflate the dataset by creating modified versions of the existing images. This strategy led to a four-fold increase in the size of the dataset, supplying the model with a broader, more comprehensive range of data from which to learn. Additionally, for data splitting and model evaluation, group k-fold cross-validation was employed, with k assigned the value of 5. This approach ensures that the same group, in this instance, frames from an identical video, does not feature in training and validation sets. Each validation set was composed of frames sourced from videos that were unequivocally different from those included in the corresponding training set. This precaution was of utmost importance to mitigate the risk of overfitting—a phenomenon that arises when a model acquires a deep familiarity with the training data to the extent that it captures noise or random fluctuations inherent in the data. This, in turn, results in subpar performance on unseen data. Ensuring the validation set incorporated frames from unique videos prevented the model from achieving high accuracy simply by memorizing familiar patterns and signal noise. This promoted the model's capacity to generalize its learning to new, previously unseen data.



(a) (b) (c)

Figure 1. (a) A cropped frame extracted from echo-videos; (b) A cropped image labeled by VEVOLAB; (c) A mask created by ImageJ and image processing method.

2.3 Cardiac function assessment

Ejection fraction (EF), an essential metric to evaluate heart function, is quantified as the ratio of blood ejected from the ventricle with each heartbeat and can be mathematically expressed as follows:

$$EF\% = \frac{(EDV - ESV)}{EDV} \times 100\%, \tag{1}$$

the end-diastolic volume (EDV) and end-systolic volume (ESV) represent the ventricular volumes at the end of diastole and end-systole, respectively. The area-length method is frequently utilized to calculate EDV and ESV [18] using the following formula:

$$EDV = \frac{8 \times Area(diastolic)^2}{3\pi \times LAX(diastolic)}, \tag{2}$$

$$ESV = \frac{8 \times Area(systolic)^2}{3\pi \times LAX(systolic)}, \tag{3}$$

this method involves the measurement of the area of the ventricle and the length of the ventricular long axis (LAX), which is the line connecting the middle of the base of the heart to its tip.

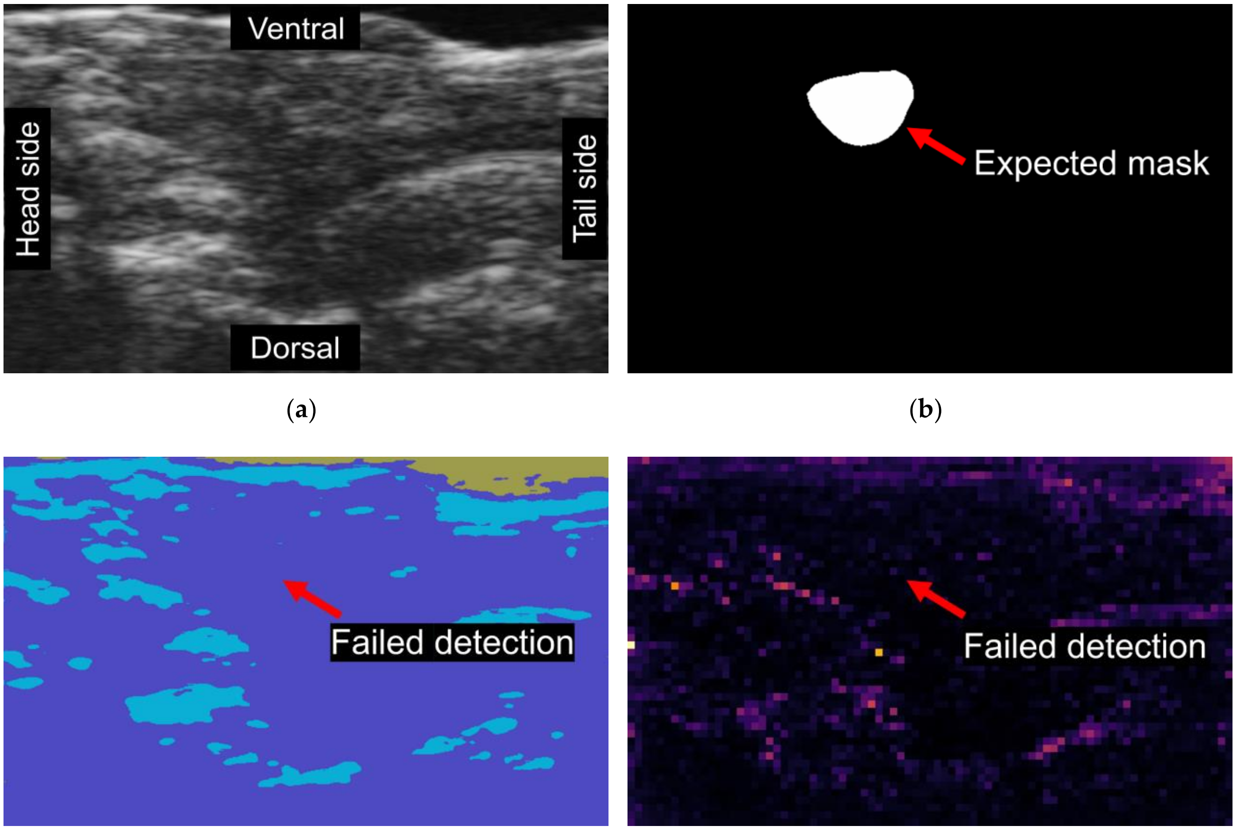
2.4 Unsupervised segmentation approach

Unsupervised image and video segmentation methodologies, which circumvent the need for pre-labeled training data, have progressively gained traction in computer vision. These techniques harbor considerable potential across a gamut of applications. However, the formidable challenge lies in discerning and tailoring an appropriate method for extracting features from zebrafish echo-videos.

A method that has drawn attention entails the utilization of modified convolutional neural networks (CNNs) [19], a particular category of deep learning models that have made notable strides in image analysis tasks. These specially designed CNNs undertake the task of ascribing labels to pixels in an unsupervised manner, thereby facilitating image segmentation without needing pre-labeled training data. This unsupervised approach's feasibility is rooted in optimizing pixel labels and feature representations via gradient descent—a widely used iterative optimization algorithm within the machine learning sphere aimed at function minimization. This procedure involves iterative updates to the network parameters, enabling the model to learn how to assign labels to pixels most effectively.

A distinct and promising instantiation of unsupervised learning within the video segmentation context is exemplified by Dino [20], a self-supervised learning method predicated on Vision Transformers (ViTs). Vision Transformers represent a novel advancement in computer vision, utilizing transformer models—originally crafted for natural language processing tasks—for image classification. Dino employs an innovative approach to self-supervised learning, aiming to predict the output of a teacher network—a well-trained model that serves as the guiding beacon throughout the learning process. It comprises a momentum encoder, which maintains a moving average of the model parameters to ensure learning stability. It employs a standard cross-entropy loss function for optimization.

What distinguishes Dino from CNNs and supervised ViTs is its innate capability to explicitly encode semantic information pertinent to image segmentation, such as scene layout and object boundaries. This unique functionality enables it to discern the structural layout of the scene and distinguish between different objects, offering potential benefits for tasks necessitating an understanding of spatial relationships within an image. Nevertheless, despite the promise held by these unsupervised methods, they come with limitations, particularly when confronted with the inherent challenges of echo-videos, such as blurring and signal noise (**Figure 2**). These factors can severely impede the effectiveness of unsupervised segmentation methods, compromising their ability to accurately segment the heart region and assess the ejection fraction within zebrafish echo-videos. As a result, supervised image segmentation, which harnesses the power of pre-labeled training data, has been the solitary method demonstrating adequate accuracy within the context of this study. Supervised learning methods hold the advantage of learning directly from labeled exemplars, enabling them to develop a more precise comprehension of the segmentation task at hand. Consequently, despite the potential shown by unsupervised methods, supervised image segmentation has proven to be the most reliable and accurate approach for this specific task.



(a) (b) (c) (d)

Figure 2. Unsupervised segmentation trials utilizing different methods failed to detect expected masks. Each image is represented by a distinct color, signifying the model-generated segmentation regions. (a) An original frame, where individuals without training could not accurately classify the ventricle's location due to the lack of ground truth labels; (b) An expected mask created manually; (c) A result frame obtained using modified CNNs revealing that the model is susceptible to color interference and, as such, is unsuitable for echo-videos; (d) A result of Dino demonstrating the model was incapable of detecting the ventricle due to background noise.

2.5 Supervised image segmentation approach

In previous work, members of our lab proposed the ZACAF, a framework based on a deep learning model for the automated assessment from bright field microscopy videos [21]. This thesis incorporated the usage of segmentation model modules that are readily accessible on PyTorch, an established machine-learning framework renowned for its widespread recognition. The models selected for the encoding phase included Resnet [22], Efficientnet [23], ResNeXt [24], and Mobilenet [25]. These models were carefully handpicked, not merely due to their robust capabilities but also their distinct attributes, which catered to a wide array of approaches to address the problem. Resnet34, Efficientnet-b4, and ResNeXt-50-32x4d were meticulously singled out for a more granular examination. This approach paved the way for a comparative exploration to identify which varied architectures demonstrated superior proficiency at extracting pertinent features from the dataset under scrutiny while maintaining parameter sizes within relatively consistent boundaries.

Resnet, the first among these models, has garnered widespread popularity in deep learning. Characterized by its relatively shallow depth, Resnet has the upper hand in alleviating issues such as overfitting and vanishing gradients, common stumbling blocks in deep learning. Furthermore, Resnet distinguishes itself through its 'residual learning' approach, which learns from residuals or errors instead of the unaltered target.

Efficientnet, the second architecture, has demonstrated impressive efficacy across various image classification tasks. This performance can be attributed to its hierarchical structure, which facilitates balanced handling of the network's depth, width, and resolution. This unique architectural design empowers Efficientnet to manage more complex models and utilize resources more efficiently, conferring an advantage in image processing tasks.

ResNeXt, on the other hand, deploys a unique strategy known as the split-transform-merge methodology. In this approach, information from multiple pathways is initially split and then independently transformed, post which it is recombined. This procedure bestows ResNeXt a high degree of flexibility and adaptability, enabling it to grapple with intricate data structures.

Finally, Mobilenet-v2 was incorporated into this study, given its compact yet potent design. With its modest size and reduced computational demands, it has proven to be an optimal choice for mobile and embedded devices, ensuring practicality and efficiency. It utilizes depthwise separable convolutions, which render it significantly smaller and faster than other models while retaining a high degree of accuracy, thereby making it a crucial addition to the comparative analysis.

For the decoder portion of the image segmentation pipeline, Unet (**Figure 3**) and Unet++ were chosen, both highly regarded and extensively employed architectures in deep learning for image segmentation. The Unet model is valued for its effectiveness in biomedical image segmentation, owing to its expansive symmetric path that accurately captures context and location information. Unet++, a more advanced variant of Unet, encompasses nested, dense skip pathways to promote feature reuse, thereby providing more precise localization and enhancing the model's performance in segmenting complex images.

The model training procedure in this research was accelerated by utilizing NVidia's A10 and H100 GPUs' computational capabilities, which were available via an online cloud service platform. Harnessing the potency of these GPUs expedited the procedure, thus rendering the training phase both speedier and more efficient. This acceleration is critically important considering the inherently complex nature of deep learning models. In this investigation, the dice loss function was employed, a frequently used loss function specifically tailored for semantic image segmentation tasks [26]. This loss function is distinguished by its proficiency in tackling typical challenges that surface in medical image segmentation tasks. A prominent issue in this field is the imbalance between foreground and background classes, potentially resulting in suboptimal segmentation outcomes. The dice loss function proposes an elegant resolution to this issue. Its design encourages the equal penalization of false negative and false positive errors, thereby promoting a balanced methodology and assuring superior precision in the segmentation outcomes.

The experimental findings revealed that using pre-trained weights, a technique routinely employed to enhance the performance of deep learning models, had a detrimental impact on the model's accuracy. Pre-trained weights represent a form of transfer learning, a process whereby knowledge acquired from training on large-scale datasets is transferred to boost the model's performance. Nevertheless, in this specific scenario, this technique proved to be counterproductive. This unexpected outcome might be attributable to the domain shift. In this phenomenon, the distribution of data changes between the extensive datasets on which the weights were pre-trained and the specific medical datasets employed in this research. Within deep learning, domain shift poses a considerable challenge as models trained on a particular dataset may not perform optimally on another due to variations in data distribution. Consequently, while pre-trained weights can prove beneficial in many scenarios, this finding suggest that their usage only sometimes enhances performance. It may even have a harmful effect when a significant domain shift occurs, as was evident in this research. This underscores the necessity of exercising caution and mindfulness when implementing transfer learning, particularly within highly specialized domains such as medical image analysis.

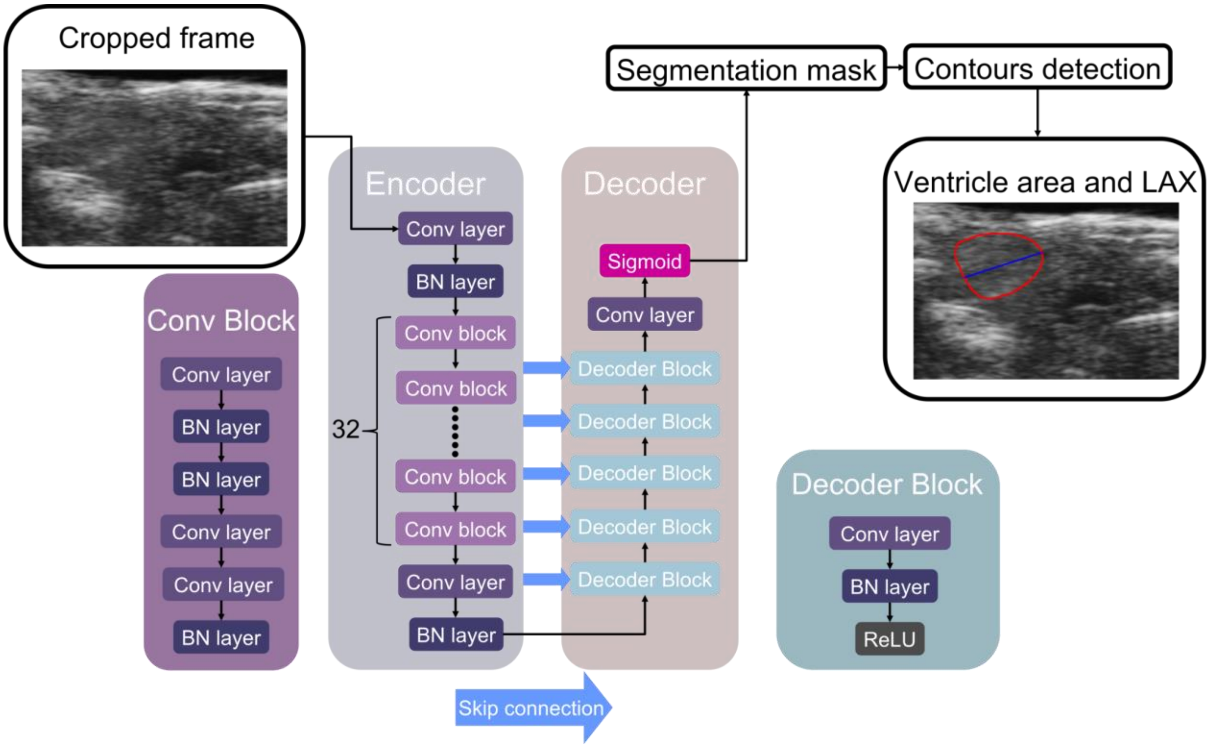


Figure 4. The selected example supervised image segmentation model architecture and its associated pro-cess flow. Specifically, Efficientnet-b4 as the Encoder and Unet as the Decoder were utilized in this figure. Initially, the frame was cropped from echocardiography videos and fed it into the Encoder. The Encoder comprised a convolution layer (Conv layer), batch normalization layer (BN layer), and Convolution block (Conv Block), each consisting of several convolution layers and batch normalization layers with varying scales. Next, the Decoder was utilized, which included the De-coder block, convolution layer, and sigmoid function. Each decoder block contains a convolution layer, batch normalization layer, and rectified linear unit (ReLU) activation function. Upon completing the deep learning model, a segmentation mask was obtained. Subsequently, contour detection was utilized to locate the ventricle, with the LAX determined via the midpoint of the left base connected to the right tip of the heart.

2.6 Quantitative comparison of approaches

2.6.1 Dice coefficient

The dice coefficient (DC) is a commonly used evaluation metric in image segmentation tasks. It measures the degree of similarity between two objects, where a score of 1 denotes perfect agreement or complete overlap, and a score of 0 indicates no overlap. The calculation of the dice coefficient is obtained by taking twice the intersection of the two objects and dividing it by the sum of the pixels in both objects. In the case of binary segmentation, the formula can be expressed as:

$$Dice = \frac{2|A \cap B|}{|A| + |B|}, \tag{4}$$

where A and B represent the two objects being compared, and the absolute values of A and B denote the total number of pixels in each object. The intersection of the two objects is the number of pixels that both objects have in common.

2.6.2 Intersection over union

Intersection over union (IoU), also known as the Jaccard Index, is a widely used metric for evaluating the performance of image segmentation algorithms. This metric is calculated by taking the ratio of the area of overlap between the predicted segmentation and the ground truth segmentation to the area of union between them. IoU ranges from 0 to 1, with 0 indicating no overlap and 1 indicating perfect overlap. For binary segmentation, the IoU can be computed using the following formula:

$$J = \frac{|A \cap B|}{|A \cup B|}, \tag{5}$$

where A and B represent the predicted and ground truth segmentation masks, respectively. The intersection between A and B refers to the set of pixels where both A and B have a non-zero value, while the union between A and B refers to the set of pixels where either A or B has a non-zero value. This metric is commonly used in deep learning-based segmentation models, as it provides a reliable measure of the accuracy of the model's predictions compared to the ground truth segmentation.

2.6.3 Receiver operating characteristic

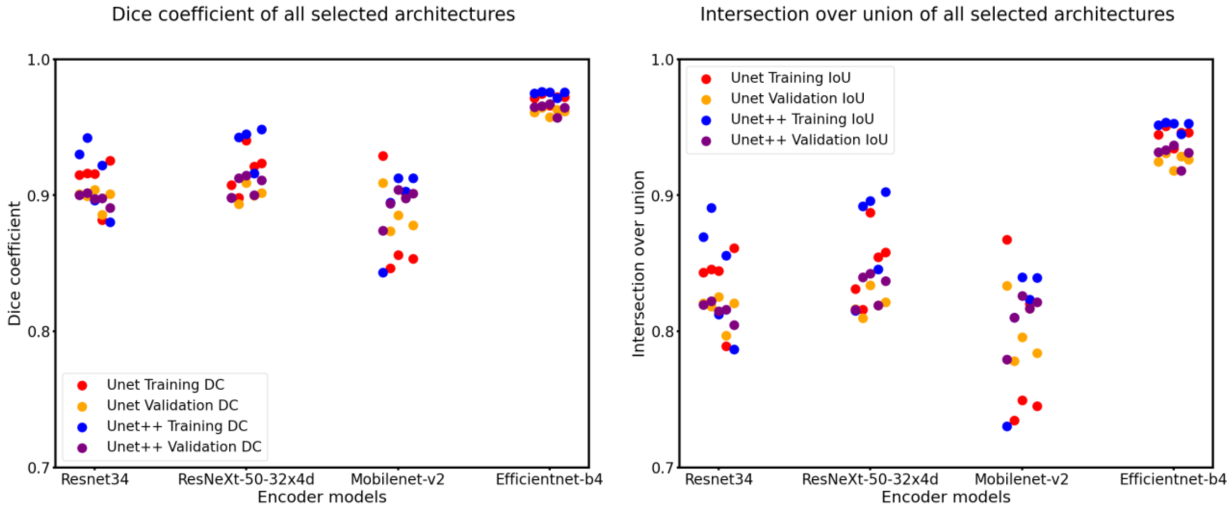
The receiver operating characteristic (ROC) curve serves as a graphic representation of a binary classifier's diagnostic competence as the discrimination threshold is adjusted. For the incorporation of ROC curves in this image segmentation model, the preliminary step was the calculation of a probability map corresponding to the target segmentation. Herein, each pixel is attributed with a probability value indicating its likelihood of being a part of the foreground. As the next step, the threshold was adjusted to classify a pixel as foreground or background, facilitating the computation of the true positive rate and the false positive rate at each respective threshold.

The area under the ROC curve (AUC), a critical metric, provides a quantitative measure of the model's precision in pixel classification, irrespective of the selected threshold. In an ideal scenario, a flawless classification model would be characterized by an AUC of 1, whereas a model whose performance equates to that of random classification would exhibit an AUC of 0.5. The ROC curve and the AUC together serve as indicators the image segmentation models'.

CHAPTER 3 – Results

The results provided in **Figure 4** elucidated the model's performance characteristics when trained using a combination of the dice loss function and the Adam optimizer, a highly respected optimization algorithm within the realm of deep learning. The model was configured with a learning rate of 0.001, a pivotal hyperparameter that prescribes the magnitude of the step during the gradient descent process. When combined with a decay rate of 0.8 that was applied after every 20 epochs, the chosen learning rate facilitated the sculpting of a finely calibrated learning curve that promoted superior model optimization throughout the training. During the model training phase, this study incorporated an early stopping strategy. This technique, frequently employed in machine learning paradigms, safeguards against overfitting by prematurely terminating the training process when no significant improvement in the model's performance on the validation set is observed over successive iterations. Beyond its primary role in preventing overfitting, this strategy is also instrumental in conserving computational resources, a key consideration given the often-extensive computational requirements of deep learning models.

Among the encoder-decoder architectures utilized in this research, the Unet++ and Efficientnet structures emerged as distinct frontrunners. These architectures achieved exceptional validation Dice coefficients of 0.967 and validation IoU scores of 0.937, respectively. These scores underline Efficientnet's proficiency in skillfully extracting pertinent features from the specific frames in this dataset, even when maintaining parameter sizes comparable to other encoder architectures. In a noteworthy observation, the Mobilenet_v2 architecture, despite its comparatively lower parameter count, achieved an IoU score comparable to those of other encoders. This result underscores its potential as a practical choice for lightweight segmentation tasks, suggesting its possible utility in scenarios where computational resources might be limited.



(**a**) (**b**)

Figure 5. The dice coefficient (a) and intersection over union (b) results with k-fold cross-validation for all selected architectures, with the x-axis representing different encoder architectures and the y-axis denoting the metrics. The model selection process was based on the higher validation result achieved by each architecture. The Unet++ and Efficientnet architectures outperformed the rest, achieving a validation Dice coefficient of 0.967 and validation IoU score of 0.937. Notably, over-fitting was observed, as indicated by the difference in training and validation metrics across all architectures.

The model consistently demonstrated high-performance levels throughout the application of the k-fold cross-validation training process, a method renowned for providing a robust estimation of a model's performance. The Area exemplified this consistency Under the Curve (AUC) values surpassing 0.93 in all instances. Along with these remarkable AUC scores, the detailed outcomes of the Receiver Operating Characteristic (ROC) analysis were presented in **Figure 5**.

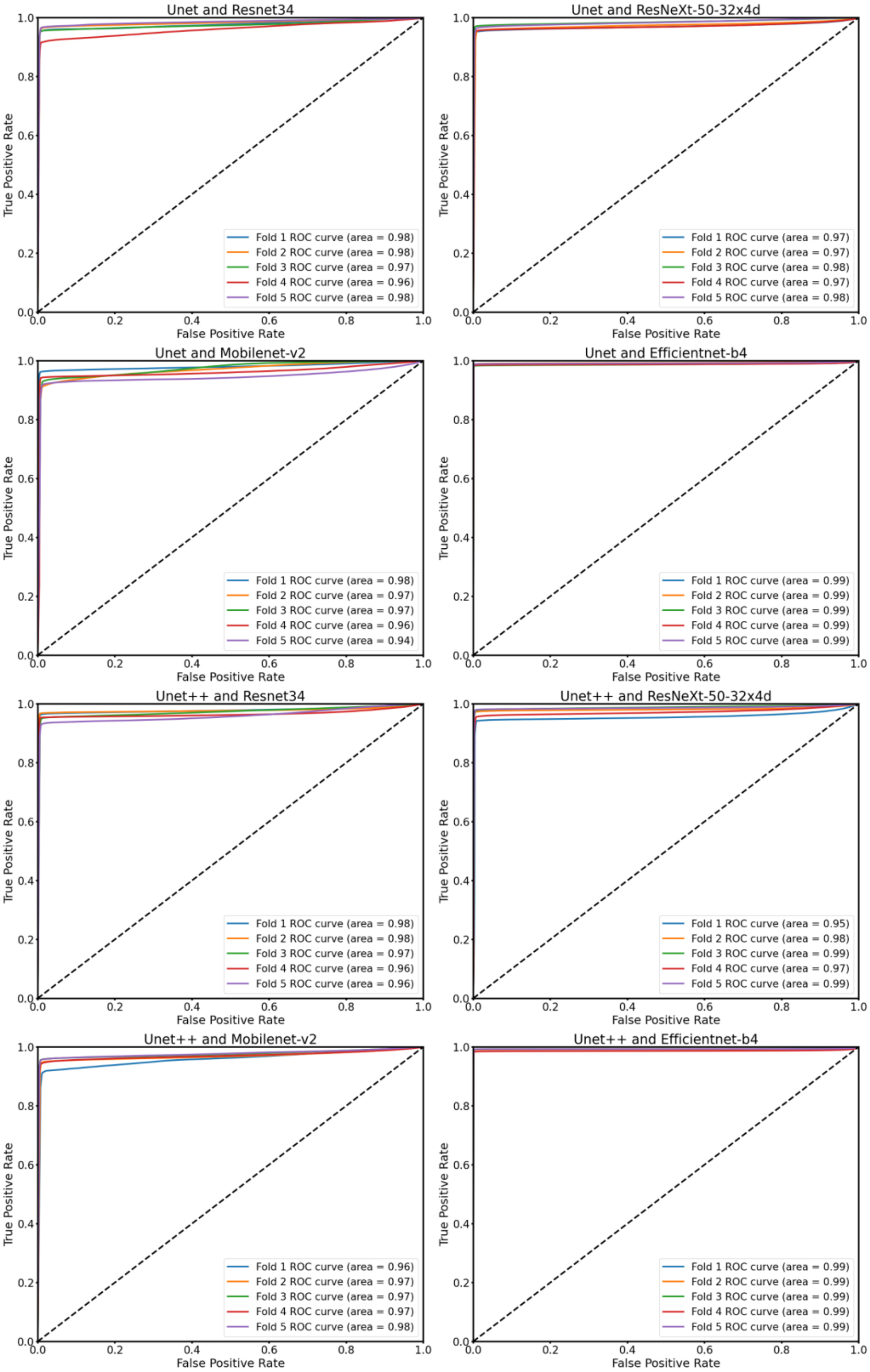


Figure 6. The receiver operating characteristic (ROC) curve and the area under the ROC curve during the k-fold cross-validation training process. AUC values exceeding 0.93 in all instances.

The principal aim of this research endeavor was to design a pioneering methodology that facilitates the automatic calculation of the ejection fraction in fish populations. In this scenario, this research sought to deploy image processing techniques as an innovative and potentially influential solution to this challenge. The methodology relies primarily on identifying the ventricular area within the fish's heart. After this, measurements of the identified ventricle's size were conducted in parallel with determinations of the left axis length (LAX) extracted from meticulously selected frames. The left axis length is a cardinal parameter as it offers insights into the heart's dimensions and the potential capacity of the ventricle (**Figure 6**).

To evaluate the proposed method's performance, a dataset with 27 videos was utilized. These videos covered three distinct categories of fish, thus offering a varied range of data to scrutinize. For analysis, frames were manually selected and labeled to ensure the most informative snapshots were employed to assess this method's capabilities. The findings obtained from this investigation were illuminating. The method's error rate was discovered to span between a minimum of 0.11% and a maximum of 16.96%, as depicted in **Figure 7**. This range provides insights into the variability of the results, suggesting that while the model is generally accurate, there can be instances where the error is marginally higher. Perhaps of greater significance, the mean error across all the conducted tests was determined to be a mere 5.13%. This statistic affirms the exceptional accuracy of the automated EF assessment method. Such a low average error implies that this method is robust and reliable, consistently delivering precise results. This research endeavor illuminated the potential of the proposed method in revolutionizing EF calculation in zebrafish and possibly in other fish species. The indications suggested that this method can significantly augment both efficiency and accuracy of EF computations, thus reducing the time consumed in manual evaluations and mitigating the propensity for human error.

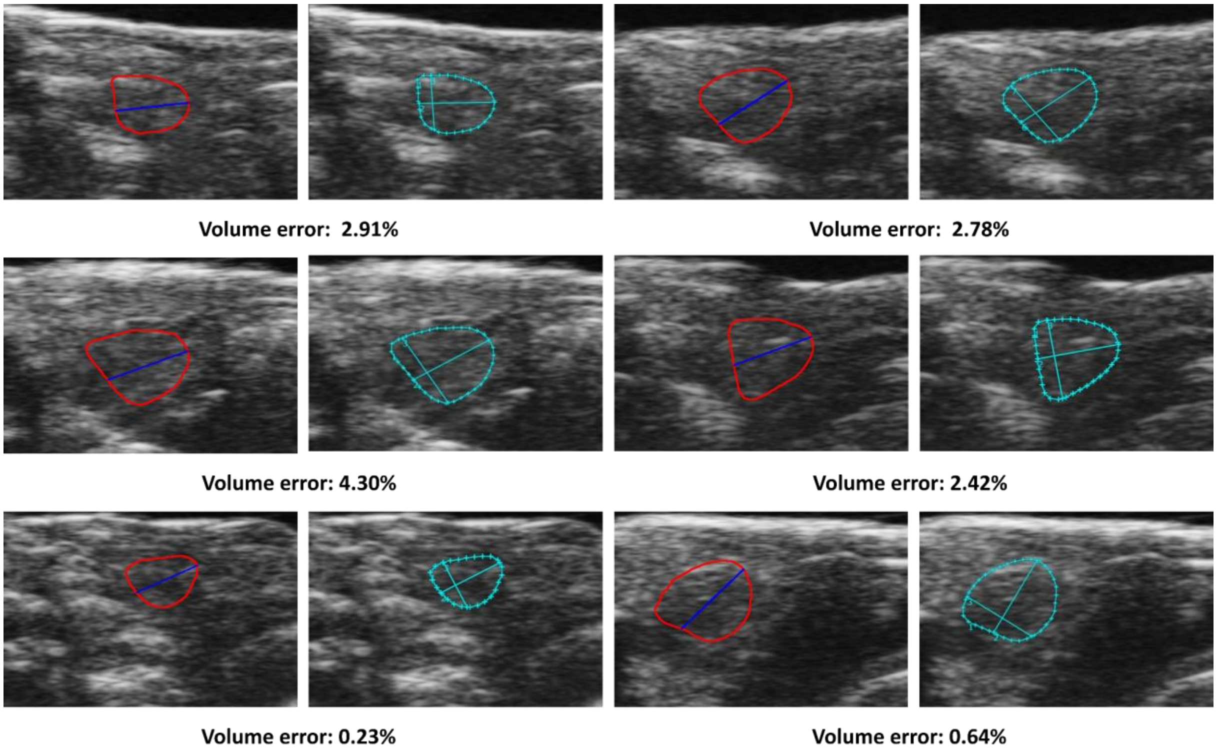


Figure 7. Comparison of automatic and manual volumes measurements. The solid lines in the figure corresponded to the ventricle contour and the LAX, detected by the proposed method. The cropped images, labeled manually with dotted lines, correspond to the manual label for each frame. The figures in the same row were selected from the same fish group, and the measurements are in pixels.

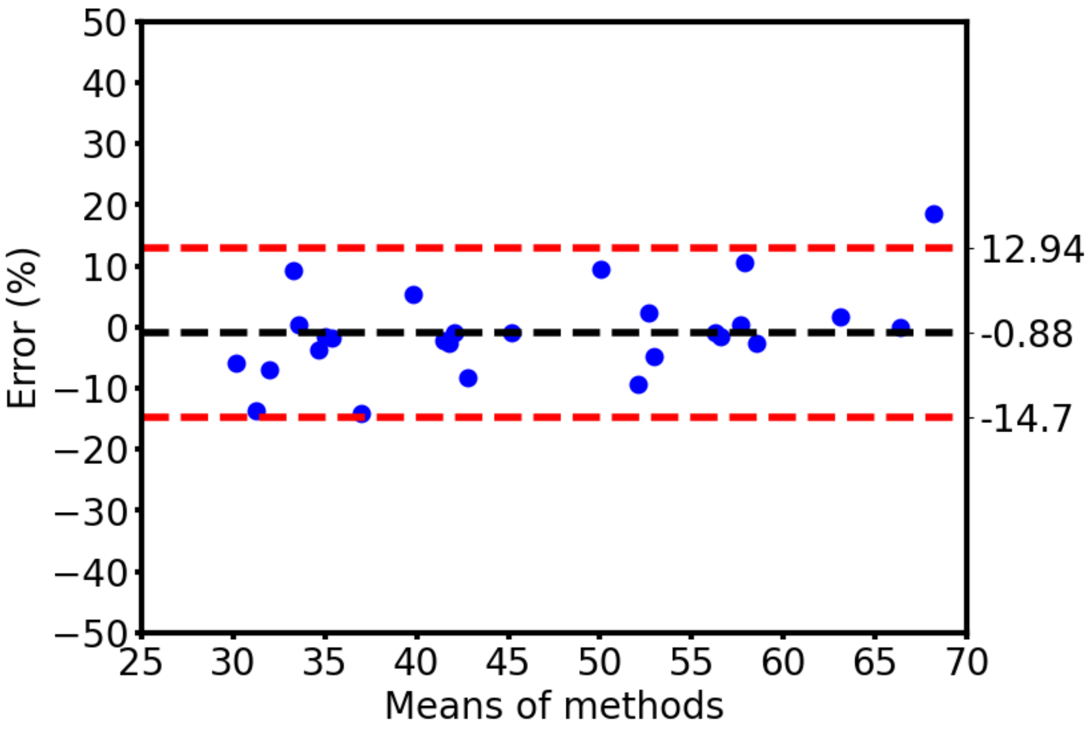


Figure 8. Bland-Altman plot for 27 sets of measurements of the EF evaluation using manual and automatic methods. Each point in the figure represents the assessment result from a different video, and the measurements are in pixels.

CHAPTER 4 – Discussion

The critical role of medical image segmentation in evaluating and analyzing cardiac function cannot be overstated. This process carries significant weight in the realm of avant-garde scientific research and within the confines of clinical settings. Traditional manual segmentation demands a prodigious amount of time and meticulous precision, two factors that often hinder its practicability and utility. Furthermore, the outcome of this labor-intensive manual process is frequently characterized by substantial variability, a discrepancy that arises from differences in interpretation among human evaluators. This inherent variability has a pronounced effect on the reliability and reproducibility of the results, thereby underscoring the urgent need for and the desirability of automated segmentation methodologies. In the study, an exploratory expedition to scrutinize the segmentation of cardiac imaging was embarked on, specifically of zebrafish.

In the recent past, unsupervised learning methodologies have captivated the research community's interest and experienced a surge in popularity in the domain of image and video segmentation. It has shown substantial promise when applied to more voluminous and diverse datasets, such as the Densely Annotated Video Segmentation (DAVIS) dataset [27]. The main allure of these methodologies stems from their innate ability to extract pertinent features from datasets without the necessity for labeled data. This capability is particularly beneficial when data labeling becomes an onerous task due to the sheer magnitude of the data involved or where labeling is infeasible due to other constraints. However, the efficacy of unsupervised learning algorithms has its limitations. It is intrinsically tied to the quality and complexity of the dataset involved, as highlighted by previous research [28]. These factors become exponentially critical in medical imaging, wherein a high level of detail and complexity often characterizes the images. Substandard dataset quality or limitations in its complexity can significantly impair the effectiveness of unsupervised learning algorithms.

In this study, the quality of the zebrafish cardiac imaging dataset posed several impediments that negatively influenced the performance of the unsupervised learning methods under scrutiny. To begin with, the relatively limited size of the dataset, coupled with the blurred images, posed formidable challenges for unsupervised learning methodologies. These dataset characteristics crippled the algorithms' ability to identify pertinent features and patterns effectively, leading to less-than-optimal segmentation results.

Secondly, the high degree of variability inherent in the dataset presented another obstacle, especially regarding classification tasks. This proved particularly challenging for individuals who need a background or expertise in the field. The variability was not merely confined to discrepancies between videos sourced from different fish; it was also observed within the same fish group. Videos obtained from the same group often exhibited divergent visual characteristics, adding complexity to the classification task. Such visual appearance variations could perplex algorithms, making it arduous to establish consistent classification rules and, consequently, diminish the effectiveness of the unsupervised learning methods.

In conclusion, despite the promising potential that unsupervised learning methods present for automating image and video segmentation tasks, their effectiveness is inextricably linked to the quality and complexity of the data they are applied to. In the case of the zebrafish cardiac imaging dataset, several challenges that emphasize the necessity for meticulous consideration of data quality as dataset characteristics when deploying these methods were identified.

During the investigation, integrating the Efficientnet and Unet++ architectures yielded the most favorable segmentation outcomes when applied to the zebrafish heart imaging dataset was observed. This compelling discovery underscores the potential advantages of combining different architectures in a hybrid model to leverage their strengths. In this specific scenario, the objective was to enhance the performance of automated image segmentation tasks, which was evidently achieved by amalgamating these two architectures.

In this study, to counteract the common issue of overfitting, this approach involved implementing data augmentation techniques, a strategy widely utilized in the domain of deep learning. This technique involves the generation of new training samples by applying various transformations to existing data. This strategy intends to enhance the dataset's diversity and size, thereby bolstering the model's capability to generalize. Regrettably, despite the proactive measures to combat overfitting through data augmentation, this issue persisted in the remaining architectural models. This inability of data augmentation to improve performance may be attributable to the highly specialized nature of the dataset, the temporal information embedded in the videos, and the limited variability within the dataset. Cumulatively, these factors have rendered the data augmentation strategies impotent in improving the model's consistency across the dataset.

An intriguing finding that emerged from this study was the substantial improvement in model accuracy observed when the encoder component of the model was conjoined with the Efficientnet architecture. This result implies that Efficientnet displayed a particular prowess in extracting valuable features from the zebrafish heart imaging dataset, a strength that consequently improved the model's overall performance. Turning the attention to the decoder architecture, the Unet model generated results that were on par with those of Unet++, despite a substantial disparity in their training times. It is noteworthy that Unet++ demanded more than a third of the total training time required by Unet, thereby establishing Unet as a more time-efficient choice for this specific task. This finding holds particular significance when working with nascent or smaller datasets, where the minimization of training time can be a critical factor to consider.

The Intersection over Union (IoU) score represents a critical performance metric in image segmentation tasks. Its pivotal role is particularly emphasized in biomedical image analysis, where high precision and accuracy are not only expected but fundamentally required. An IoU score in the lower range signifies a diminished overlap between the predicted segmentation and the ground truth. This suggests a divergence in identifying and delineating an object or area within the image. This discrepancy can lead to inaccuracies in subsequent calculations, including, but not limited to, measurements of area and the length of the left axis (LAX). This study identified a potential complication that manifests as an inaccurate representation of the ventricular shape by the segmented area. Should the segmentation model need to capture the precise shape and dimensions of the ventricle, discrepancies could emerge in the subsequent EF evaluations. This concern warrants substantial attention, especially in the context of biological structures, given their inherent complexity and wide range of variability. The dice loss function, which enjoys widespread application in image segmentation tasks, quantifies the similarity between the predicted and ground truth segments, focusing exclusively on their intersection. Despite its extensive usage, the Dice loss function might exhibit potential limitations, particularly its ability to accurately account for the shape patterns inherent within the images. This could potentially culminate in instances where the model fails to acknowledge some of the subtler aspects of the ventricular structure, consequently jeopardizing the accuracy of the segmentation results. This study achieved an impressive IoU score of 0.937, which hints at a high level of accuracy in the segmentation results. However, it is vital to acknowledge these potential limitations, as they offer insights into areas ripe for refinement and enhancement. In conclusion, this research underscores the importance of meticulous consideration of both the unique attributes of the dataset and the complexity of the machine learning model architecture. This is especially significant when working with smaller, specialized datasets like the zebrafish heart imaging dataset. Understanding these nuances and potential pitfalls is vital to engineer accurate and reliable segmentation models. This understanding emphasizes the need for a delicate balancing act, wherein both the dataset's characteristics and the complexity of the machine learning model must be optimally attuned to achieve the best possible results.

Considering the limitations above, it becomes evident that future research endeavors should strategically focus on examining larger and more comprehensive datasets of superior quality. Expanding the dataset's size and enhancing its quality could augment the efficacy of unsupervised learning methods. Such improvements could empower unsupervised models to derive more sophisticated patterns and insights, enhancing their competency in performing image segmentation tasks. In addition to amplifying the dataset's scope and quality, future studies could combine supervised and unsupervised learning methodologies to bolster the accuracy of medical image segmentation. This integrated approach could potentially harness the advantages of both paradigms, enabling models to learn from labeled and unlabeled data, which, in turn, could enhance their overall performance [29].

One feasible trajectory for further research could be implementing a supervised segmentation model predicated on the video model. This approach holds considerable promise to boost segmentation accuracy substantially. Nevertheless, this pathway necessitates substantial volumes of labeled data, implying that extensive labeling efforts would be required. Given the limitations posed by the original scope of this study, such a pursuit was deemed infeasible. Therefore, subsequent work could also target the selection of the most suitable unsupervised algorithm, capable of extracting features that can provide robust support to model training effectively. By synthesizing all these strategies, it might be possible to navigate the multifaceted challenges posed by smaller, specialized datasets, propelling the field of automated medical image segmentation.

CHAPTER 5 – Conclusion

The prospect of harnessing the power of deep learning algorithms for the segmentation and assessment of cardiovascular metrics presents a promising direction for future research endeavors. The utilization of such technology carries distinct advantages, particularly in its capacity to substantially enhance the accuracy, efficiency, and objectivity of interpreting and analyzing complex biomedical data. The complexity and remarkable specificity inherent to such data often pose formidable challenges to traditional manual methods, accentuating the potential benefits and transformative impact of incorporating cutting-edge machine learning techniques. Furthermore, integrating automation into image processing and analysis can significantly alleviate the burden placed on researchers. It effectively mitigates the time and effort typically expended on labor-intensive manual tasks. Such a stride forward in technological advancement fuels the acceleration of research and amplifies the precision and replicability of data interpretation.

However, despite these substantial benefits, this investigation illuminates the inherent challenges and constraints of applying deep learning methodologies on compact, domain-specific datasets, such as those deployed in zebrafish studies. These datasets' confined size and specialized nature pose significant hurdles for deep learning models. The need for more training data may render these models susceptible to overfitting, a prevalent issue where the model excessively learns the training data, thereby performing poorly on unseen data. This predicament can lead to diminished generalization capabilities and decreased overall predictive accuracy, undermining the models' effectiveness and reliability. Future research endeavors will benefit significantly from investigating larger, superior-quality datasets. By expanding the size and enhancing the quality of datasets, the effectiveness of unsupervised learning methods can be increased. Such improvements would enable unsupervised models to uncover more sophisticated patterns and insights, enhancing their proficiency in image segmentation tasks. Moreover, an integrated approach that combines the merits of supervised and unsupervised learning methodologies can provide a valuable avenue for augmenting the accuracy of medical image segmentation. This integration could harness the potential of both paradigms, facilitating models to learn from labeled and unlabeled data, which could amplify their overall performance.

On a broader scale, developing an automated system for segmenting zebrafish embryos from echo-videos utilizing supervised deep learning methods signifies a substantial progression in biomedical research. Further optimization and refinement of these deep learning models are anticipated to pave the way for accurate and efficient evaluation of cardiovascular metrics derived from echo-videos, thus enhancing research efforts in cardiovascular development.

References

1. Cardiovascular diseases (CVDs). Available online: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on
2. Khan, F.R.; Alhewairini, S.S. Zebrafish (Danio rerio) as a model organism. *Current trends in cancer management* **2018**, *27*, 3-18.
3. Martin, W.K.; Tennant, A.H.; Conolly, R.B.; Prince, K.; Stevens, J.S.; DeMarini, D.M.; Martin, B.L.; Thompson, L.C.; Gilmour, M.I.; Cascio, W.E. High-throughput video processing of heart rate responses in multiple wild-type embryonic zebrafish per imaging field. *Scientific reports* **2019**, *9*, 145.
4. Vornanen, M.; Hassinen, M. Zebrafish heart as a model for human cardiac electrophysiology. *Channels* **2016**, *10*, 101-110.
5. Asnani, A.; Peterson, R.T. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *Disease models & mechanisms* **2014**, *7*, 763-767.
6. Narumanchi, S.; Wang, H.; Perttunen, S.; Tikkanen, I.; Lakkisto, P.; Paavola, J. Zebrafish heart failure models. *Frontiers in Cell and Developmental Biology* **2021**, *9*, 662583.
7. Pott, A.; Rottbauer, W.; Just, S. Functional genomics in zebrafish as a tool to identify novel antiarrhythmic targets. *Current medicinal chemistry* **2014**, *21*, 1320-1329.
8. Bu, H.; Ding, Y.; Li, J.; Zhu, P.; Shih, Y.-H.; Wang, M.; Zhang, Y.; Lin, X.; Xu, X. Inhibition of mTOR or MAPK ameliorates vmhcl/myh7 cardiomyopathy in zebrafish. *JCI insight* **2021**, *6*.
9. Ding, Y.; Bu, H.; Xu, X. Modeling inherited cardiomyopathies in adult zebrafish for precision medicine. *Frontiers in Physiology* **2020**, *11*, 599244.
10. Dvornikov, A.V.; Wang, M.; Yang, J.; Zhu, P.; Le, T.; Lin, X.; Cao, H.; Xu, X. Phenotyping an adult zebrafish lamp2 cardiomyopathy model identifies mTOR inhibition as a candidate therapy. *Journal of molecular and cellular cardiology* **2019**, *133*, 199-208.
11. Sirjani, N.; Moradi, S.; Oghli, M.G.; Hosseinsabet, A.; Alizadehasl, A.; Yadollahi, M.; Shiri, I.; Shabanzadeh, A. Automatic cardiac evaluations using a deep video object segmentation network. *Insights into Imaging* **2022**, *13*, 1-14.
12. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **2017**, *39*, 2481-2495.
13. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015; pp. 234-241.

14. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, 2018; pp. 3-11.

15. Wang, L.W.; Huttner, I.G.; Santiago, C.F.; Kesteven, S.H.; Yu, Z.-Y.; Feneley, M.P.; Fatkin, D. Standardized echocardiographic assessment of cardiac function in normal adult zebrafish and heart disease models. *Disease models & mechanisms* **2017**, *10*, 63-76.

16. Ma, X.; Ding, Y.; Wang, Y.; Xu, X. A doxorubicin-induced cardiomyopathy model in adult zebrafish. *Journal of visualized experiments: JoVE* **2018**.

17. Wang, Y.; Lu, X.; Wang, X.; Qiu, Q.; Zhu, P.; Ma, L.; Ma, X.; Herrmann, J.; Lin, X.; Wang, W. atg7-based autophagy activation reverses doxorubicin-induced cardiotoxicity. *Circulation research* **2021**, *129*, e166-e182.

18. González-Rosa, J.M.; Guzman-Martinez, G.; Marques, I.J.; Sanchez-Iranzo, H.; Jiménez-Borreguero, L.J.; Mercader, N. Use of echocardiography reveals reestablishment of ventricular pumping efficiency and partial ventricular wall motion recovery upon ventricular cryoinjury in the zebrafish. *PloS one* **2014**, *9*, e115604.

19. Kanezaki, A. Unsupervised image segmentation by backpropagation. In Proceedings of the 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2018; pp. 1543-1547.

20. Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; Joulin, A. Emerging properties in self-supervised vision transformers. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021; pp. 9650-9660.

21. Naderi, A.M.; Bu, H.; Su, J.; Huang, M.-H.; Vo, K.; Torres, R.S.T.; Chiao, J.-C.; Lee, J.; Lau, M.P.; Xu, X. Deep learning-based framework for cardiac function assessment in embryonic zebrafish from heart beating videos. *Computers in biology and medicine* **2021**, *135*, 104565.

22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp. 770-778.

23. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020; pp. 10781-10790.

24. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 1492-1500.

25. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 4510-4520.

26. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), 2020; pp. 1-7.

27. Caelles, S.; Pont-Tuset, J.; Perazzi, F.; Montes, A.; Maninis, K.-K.; Van Gool, L. The 2019 davis challenge on vos: Unsupervised multi-object segmentation. *arXiv preprint arXiv:1905.00737* **2019**.

28. Locatello, F.; Bauer, S.; Lucic, M.; Raetsch, G.; Gelly, S.; Schölkopf, B.; Bachem, O. Challenging common assumptions in the unsupervised learning of disentangled representations. In Proceedings of the international conference on machine learning, 2019; pp. 4114-4124.

29. Bortsova, G.; Dubost, F.; Hogeweg, L.; Katramados, I.; De Bruijne, M. Semi-supervised medical image segmentation via learning consistency under transformations. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22, 2019; pp. 810-818.