



Universiteit
Leiden
The Netherlands

Opleiding Bioinformatica

Evaluation and Improvement of Methods and Models

for Segmentation of Zebrafish Larvae

Wilco de Boer

Supervisors:

Fons Verbeek & Kristian Rietveld, Lu Cao

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

liacs.leidenuniv.nl

19/09/2019

This page intentionally left blank.

Abstract

The zebrafish (*Danio rerio*) is a popular model organism in biomedical and pharmaceutical research. In order to support this research, high-throughput methods for the analysis of large amounts of zebrafish are continuously being developed. Among these are methods for segmentation, which is an especially troublesome process due to the high transparency of the zebrafish. Over the course of 2 years, the Vertebrate Automated Screening Technology (VAST) BioImager has been pivotal for imaging a sizable amount of zebrafish using the system's integrated positioning camera. The use of a Leica microscope camera was important as well, and a smaller amount of zebrafish have been imaged after adding one to the setup. For the majority of these images, segmentations have been created using multiple low throughput methods (a MATLAB script, a Python script, and manual tracing). The resulting segmentations were used to train a convolutional deep neural network in several ways. By analyzing the predictions of these different networks, we show that only a small amount of training data is required to achieve sufficient accuracy, and that adding too much or the wrong training images can have adverse effects. Additionally, we show that applying a grey erosion filter can improve predictions under certain circumstances, but has a detrimental effect under others. Finally, it has been determined that images from the VAST camera and the Leica microscope require a different model for segmentation, and for both cases an optimal model is created.

This page intentionally left blank.

Contents

1	Introduction	1
1.1	Research Questions	3
1.2	Thesis Structure	5
2	Materials & Methods	7
2.1	Zebrafish	7
2.2	VAST BioImager & Leica Microscope	8
2.3	Image Processing	9
2.3.1	Morphological Erosion	9
2.3.2	Methods used for Image Resampling	9
2.3.3	Simple Thresholding	10
2.3.4	Adaptive Thresholding	11
2.4	Image Properties	11
2.4.1	Moments	12
2.4.2	Centroid	12
2.4.3	Convex Hull	13
2.4.4	Convexity Defects	13
2.4.5	Solidity	13
2.5	Violin Plots	14
2.6	Deep Neural Network	14
3	Implementation	17
3.1	Data Collection	18
3.1.1	Old VAST Images	18
3.1.2	New VAST Images	18
3.1.3	Newer VAST Images	21
3.1.4	Newer BFM Images	22
3.1.5	Dataset Summary	27
3.2	Segmentation	29
3.2.1	Model Generation and Training	29

3.2.2	Preprocessing and Postprocessing	29
3.2.3	Metrics	30
3.2.4	Evaluations	31
4	Results	33
4.1	Evaluating with increasing amounts of training data	33
4.2	Evaluation of models based on different training data	34
4.3	Determining the impact of the aspect ratio	36
4.4	Using bad data as testing samples	38
4.5	Using bad data as training samples	39
4.6	Explaining the functionality of the DNN-ZF method	40
5	Conclusions & Discussions	51
5.1	How can the functionality of the DNN-ZF be explained?	51
5.2	How well does the existing hybrid segmentation method perform on new datasets?	52
5.3	How does the quality of a model change as it is being trained on more data?	52
5.4	Does the aspect ratio of an image affect the quality of its segmentation?	52
5.5	How does the DNN-ZF method perform on 'bad' input data?	53
5.6	Can 'bad' samples still be used as training data?	53
5.7	Can a single model make accurate predictions for both VAST and BFM images?	54
5.8	What effect does adding a filter to the images have?	54
5.9	How can the performance of the DNN-ZF for the segmentation of BFM images be improved?	54
5.10	How can an assessment of methods and models for segmentation of zebrafish larvae contribute to an optimal processing pipeline?	55
	Bibliography	56

Chapter 1

Introduction

The zebrafish (*Danio rerio*) is a small freshwater fish, commonly found in aquariums around the world. Zebrafish however are native to South Asia, where they inhabit moderately flowing or stagnant water. When the zebrafish was first described in literature in 1822 it was noted that they were “beautiful fish”, but that they were “insipid” and thus of little economical value beyond their beauty [HS22]. However, since their first use as a model organism in the 1960s, the zebrafish has become increasingly popular among researchers. This popularity is in part due to the simple and cheap logistics of the zebrafish. They are small (with adults only reaching a length of up to 4 cm), breed year-round, and are capable of producing hundreds of larvae each time they mate. This makes it easy to both acquire and store them in bulk. Zebrafish are also among the aquarium fish which are not very particular about their surroundings which make them much cheaper and easier to maintain than other fish and mammals. Finally, the larvae and eggs are transparent, which makes it possible to observe the internals of live specimen using any form of microscopy. Another big reason for the popularity of the zebrafish is that 84% of human genes associated with disease have an orthologue in the genome of the zebrafish. A total of 70% of human genes have an orthologue available. This makes the zebrafish very useful for research into diseases and cures, toxicology, and development. Additionally, zebrafish have a high regenerative capacity and are capable of repairing damage to almost all their tissues. This includes the heart and nervous system. Given their similarity to humans, further understanding of this regeneration could lead to new treatments for tissue damage in humans.

Since the zebrafish is such a useful model organism, several different methods for high-throughput processing have been devised. One of these methods is the VAST (Vertebrate Automated Screening Technology) BioImager developed by Union Biometrica. This system can be loaded up with a large amount (100) of prepared zebrafish larvae, after which the VAST BioImager will image each specimen fully automatically, without killing the specimen. This imaging involves rotating the specimen as a lot (100) images are being taken from different angles. These images can be taken by both the camera integrated into the VAST BioImager which is ordinarily used to position the zebrafish larvae, or any external optical solution such as a bright field microscope with a camera. After these images are taken, however, they still need to be processed to obtain useful information about the zebrafish fed to the VAST BioImager, such as the length, surface area, volume, perceived age, etc.

This consist of three different steps:

1. First, each image is individually segmented in order to separate the fish from the capillaries and background.
2. Then, all these segmentations for a particular zebrafish larva are combined into a 3D reconstruction of the concerned specimen.
3. Finally, several properties can be extracted from this 3D reconstruction (length, surface area, volume, etc.) and others from the images or segmentations themselves (perceived age).

The requirement of these processing steps has led to the envisioning of a fully automated pipeline outlined in Figure 1.1, which is currently being developed by the section Imaging and Bioinformatics at the Leiden University. This pipeline begins with a researcher working with zebrafish larvae, who requires the provided metrics for their zebrafish larvae. After processing their zebrafish larvae with the VAST BioImager, the images are uploaded through a web portal, indicated by the first two steps in Figure 1.1. A computing cluster will then automatically perform the three steps mentioned above, that of segmenting, 3D reconstruction, and metric extraction. This is indicated by the next three steps in Figure 1.1. Finally, these metrics will be presented to the researcher through the same web portal, after which they can be used to give quantifiable results for a large number of zebrafish larvae. This is once again indicated by the last three steps in Figure 1.1.

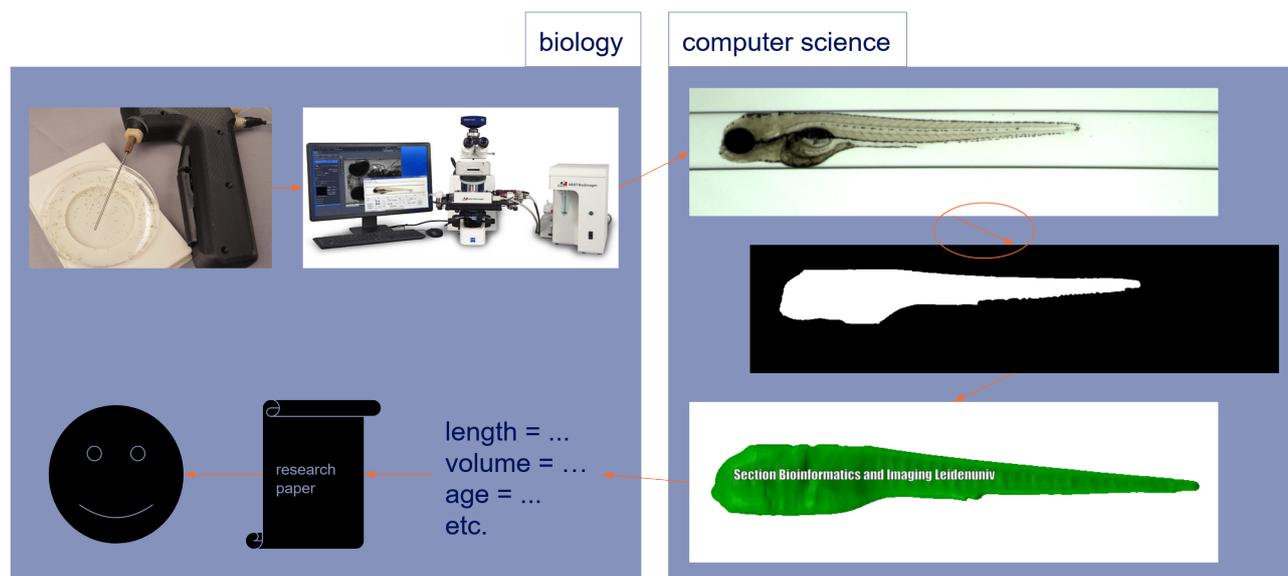


Figure 1.1: Rough overview of a pipeline for the processing of images taken by the VAST BioImager, as envisioned by the section Imaging and Bioinformatics at the Leiden University.

Naturally, each of the three steps mentioned above is extremely complicated, and a single sentence does not do them justice. This paper in particular will focus solely on the step of segmenting the images of zebrafish larvae, which is encircled in Figure 1.1. Whilst many good methods for image segmentation already exist, these can not be applied to images of zebrafish larvae. First of all, images of zebrafish larvae feature a partially transparent foreground. For the most part, enough contrast between the specimen and the background remains for proper segmentation. The tail of the zebrafish larva, however, contains sections with the same colour as

the background due to its transparency. Additionally, the edges of the capillary which holds the zebrafish larva during imaging are visible, and have the same colour as parts of the specimen. Proper segmentation of zebrafish larvae would require a solution for these two features which would pose a problem to generic segmentation methods.

Prior research on this particular topic has already been done [Guo17] and shows promising solutions to these problems. However, this research and the resulting method have only taken the single dataset available at the time into consideration. And although the resulting method works amazingly on this particular dataset, it might not do so on datasets created under slightly different conditions. Additionally, the created method relies on MATLAB, and is thus not suited to be transferred to a computing cluster. Part of this problem has been solved by creating a new method based on deep neural networks, which would make it more flexible towards future datasets [Ver18]. This new method is also written in Python with the computing cluster taken into account. However, this method is far from perfect and is still built and tested against the single available dataset. As such, more research into this topic is still required. And with the recent availability of three additional datasets, this can also be conducted effectively.

1.1 Research Questions

As mentioned before, the research presented in this paper will focus on the segmentation of the images of zebrafish larvae. As several methods for this process already exist however, the focus will not be on the creation of another method. Instead, the three newly available datasets will be used to improve and evaluate the existing methods. The final goal in mind is to improve and support the processing pipeline explained above. This leads us to the following primary research question:

RQ: How can an assessment of methods and models for segmentation of zebrafish larvae contribute to an optimal processing pipeline?

As one of the methods that we are dealing with is based on a deep neural network, it will be very informative to figure out the principles behind its operation. This can potentially lead to further improvements of the segmentation method. As such, we include the following research question:

SRQ1: How can the functionality of the DNN-ZF be explained?

Another of the prior methods is the hybrid segmentation method. As this is the oldest and most rigid of the two, it is relevant to know how well it performs on these three new datasets, leading to the following research question:

SRQ2: How well does the existing hybrid segmentation method perform on new datasets?

When a deep neural network is trained, the assumption is that the quality increases over time. At some point however, this quality will reach a point where additional training will not lead to much of an increase. It is relevant to prove this is indeed the case for our methods. Additionally, determining the point at which enough training data is supplied is relevant to determine if specific models have been trained sufficiently. This leads us to the following research question:

SRQ3: How does the quality of a model change as it is being trained on more data?

Within the total of four available datasets, there is one dataset of which the images have a different aspect ratio than the other three. It is relevant to determine if this will present any issues when these images are segmented, leading to the following research question:

SRQ4: Does the aspect ratio of an image affect the quality of its segmentation?

Additionally, among these four datasets, there are many images which have been considered to be 'bad'. This category includes images where the zebrafish larva is physically damaged, or the zebrafish larva has been imaged improperly by the VAST BioImager. Additionally, when training segmentations were created for use with the deep neural network, a large number of them were segmented incorrectly. Despite the irrelevance of these images to the general process, it is still interesting to find out what happens when these images are predicted by the DNN-ZF method. This leads us to the following research question:

SRQ5: How does the DNN-ZF method perform on 'bad' input data?

Similarly, although such 'bad' images should never be used as training data, it is still interesting to find out what happens when they are. By doing so interesting insights could be gained, but even just confirming something seemingly obvious proves that everything works as intended. We also have to consider the situation where there is not enough good training data available and we have to rely on such 'bad' samples. As such, the following research question is included:

SRQ6: Can 'bad' samples still be used as training data?

Among the four available datasets, one of them not only has a different aspect ratio, but also comes from a different source. Where the other three datasets were taken using the camera integrated into the VAST BioImager (denoted as **VAST images**), this fourth dataset was made using the aforementioned external optical solution: a bright field microscope with an attached camera (denoted as **BFM images**). These images are completely different than the others. Not only do they have a different resolution and aspect ratio, but they also have a different brightness, contrast, and general quality. It is relevant to know if the deep neural network can handle both sources at once, or if they need to be treated separately. This leads to the following research question:

SRQ7: Can a single model make accurate predictions for both VAST and BFM images?

One of the possible ways to improve the deep neural network is to apply a grey erosion filter to both the images used for training and those that will be predicted. Whether such a filter actually improves results or not needs to be tested, which naturally leads to the following research question:

SRQ8: What effect does adding a filter to the images have?

Only a single dataset out of the four available consists of the higher quality microscope images. This brings us to the same issue outlined before this section, which is that any method made for and tested on this dataset will be relatively biased. Additionally, adapting the existing methods to very different images might present challenges of its own. These issues are covered under the following research question:

SRQ9: How can the performance of the DNN-ZF for the segmentation of BFM images be improved?

1.2 Thesis Structure

In this section, a general overview of the structure of this paper will be provided. Chapter 1 introduces the scope of the performed research and puts it in context to previous works. Additionally, the research questions are introduced.

Chapter 2 introduces several theoretical principles which are important to one's understanding of the content of the paper. These principles were used at vital points during the research and will be referenced several times. Additionally, the specific deep neural network architecture evaluated in this paper is introduced. Finally, the systems used to image the zebrafish larvae, as well as the species itself, is explained in greater detail.

Chapter 3 details the proceedings of the performed research. First, the research is divided into several steps which are presented in a flowchart. Then, each section of the flowchart is expanded upon separately. Choices regarding the use of certain methods are included in these explanations, as are problems which were faced and resolved during the research. Additionally, both a concise and detailed overview of all the data used in this research is presented in this chapter. It should be noted that no specific experiments are described in this chapter, but that the general methodology for all final experiments is introduced instead, as the only difference between these experiments is the data which is used for training and testing.

Chapter 4 describes these various final experiments and presents their results. Additionally, these results are discussed and a preliminary conclusion limited to each experiment is made. Finally, several images are presented and discussed that will grant a greater understanding of the functionality of the specific deep neural network architecture evaluated in this paper.

Chapter 5 answers all the research question presented in Section 1.1 using the results and conclusions from Chapter 4. These answers are further discussed and, where appropriate, accompanied by recommendations for further research. The chapter ends with a few concise recommendations in regards to the future treatment of the most insightful conclusions.

Chapter 2

Materials & Methods

This chapter will introduce several theoretical principles which are used throughout the presented research. Most of these are related to image processing, but a more obscure type of plots is also explained. Additionally, more information is given about the zebrafish as a model organism, and the system used to image them is thoroughly explained.

2.1 Zebrafish

The zebrafish (*Danio rerio*) has become an extremely popular model organism over the last few decades, mostly in biomedical and pharmaceutical research.

Zebrafish larvae are mostly transparent and develop in eggs outside of the parents. This makes it extremely easy to observe the internals of the fish with both bright field microscopy and fluorescent microscopy. As with many other model organisms, the genome of the zebrafish has been fully sequenced and is widely available. What is interesting about the genome of the zebrafish in particular is that about 70% of human genes have an orthologue in this genome [HCT⁺13]. Furthermore, 84% of the human genes associated with disease are found to have an orthologous gene in the zebrafish. Combined with the fact that zebrafish are vertebrates just like humans, and thus feature very similar anatomy, this makes them ideal for research into diseases and their cures. Another very interesting property of zebrafish is that they are capable of regenerating the vast majority of their tissues, including the heart, brain, and spinal cord [GBHP13]. Given the similarity of zebrafish and humans, a more complete understanding of the processes behind this regeneration can lead to novel treatment methods applicable to humans.

In addition to these features which makes the zebrafish useful as a model organism, they are also easy to handle logistically [Mey18]. As zebrafish are a common freshwater aquarium fish, they can be obtained quite easily, and methods for their care are well known and distributed. Additionally, zebrafish are a fecund species which breeds all year round, and develops rapidly, which allows for rapid growth and proliferation of modified

lines of fish. Due to the small size of the zebrafish, and the even smaller size of their larvae, they are easy to handle and space-efficient. Combined with their rapid proliferation, any experiment can easily be performed in bulk to average out inherent biological randomness. Figure 2.1 should give a good idea of the size of the zebrafish larva, as well as the ease with which experiments can be scaled up.

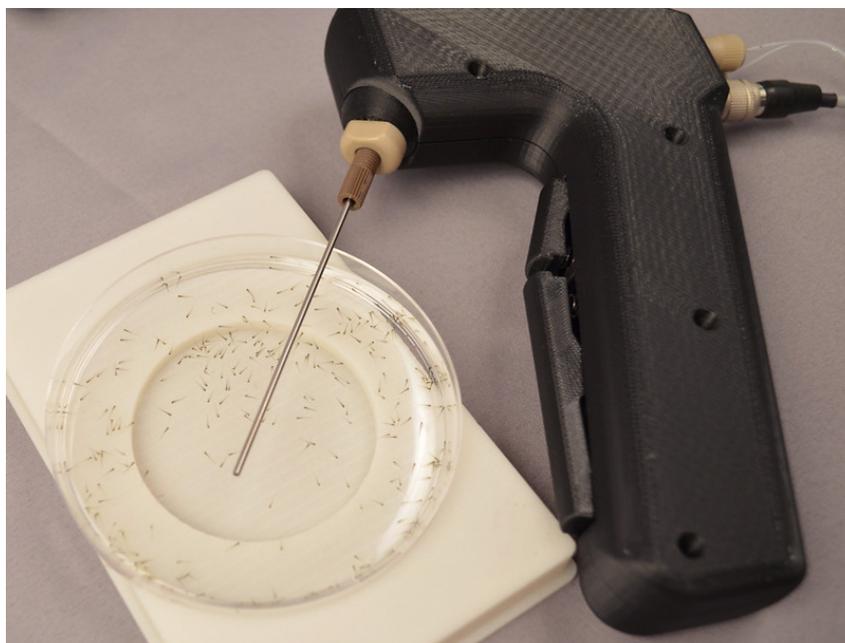


Figure 2.1: A large amount of zebrafish larvae in a petri dish. The device visible to the right is designed to be hand held. Source: <https://www.unionbio.com/vast/>.

2.2 VAST BioImager & Leica Microscope

The VAST (Vertebrate Automated Screening Technology) BioImager¹ is a system developed by Union Biometrica. The system is designed to aid researchers in the process of imaging large quantities of zebrafish larvae. The core of the VAST BioImager consists of a capillary, as can be seen in Figure 2.2. An external feeding system is used to load zebrafish larvae into this capillary, one at a time. A system of stepper motors can rotate this capillary at increments of 3° [Uni14], along with the contained specimen. The VAST BioImager also contains a simple camera (Allied Vision Systems, Pro Silica GE 1050 [Guo17]), which observes the capillary and its content through a prism. The primary purpose of this camera is to direct the feeding system such that each zebrafish larva is positioned at the proper location of the capillary. In practice, this camera is also capable of capturing usable images of the entire specimen loaded into the capillary, at any desired rotation. Due to the camera's low resolution ($10\mu\text{m}$ [Uni14]) and the characteristics of its lenses however, it will do so at a sub-optimal quality.

In order to use the VAST BioImager, an external feeding system is attached to provide it with zebrafish larvae. Additionally, the system is mounted onto a Leica microscope. Electrical connections to this microscope allow it to be controlled by the VAST system. This gives rise to the MM-HTAI architecture as envisioned by Y. Guo.

¹<https://www.unionbio.com/vast/>



Figure 2.2: The core of the VAST BioImager, mounted on a microscope. Source: <https://www.unionbio.com/vast/>.

With this architecture, samples are provided from a single conical sample cup, mounted as seen on the right of Figure 2.3. Imaging of the zebrafish larvae is done both by the positioning camera and the Leica microscope. The positioning camera is used to capture images of the entire specimen at a low quality. The Leica microscope is used to capture high-quality images of the loaded specimen, although at high magnifications it is not capable of capturing the entire specimen at once. Additionally, the Leica microscope is capable of operating with both bright field microscopy and fluorescent microscopy.

2.3 Image Processing

Several methods of processing images, that is the process of transforming one image into a similar but different image, have been used.

2.3.1 Morphological Erosion

Morphological erosion [IC18] is the process of computing a local minimum over some kernel for each pixel in the image. These minima will form a new image of the same dimensions. In effect, this leads to dark areas becoming dilated, whilst light areas become eroded, as can be observed in Figure 2.4.

2.3.2 Methods used for Image Resampling

Image resampling filters can be used to convert images from one coordinate system into another. These would take the place of the mapping function. The following filters will be used in this paper:



Figure 2.3: A usable setup consisting of the VAST BioImager, a microscope, and a sample feeder. Source: <https://www.unionbio.com/vast/>.

Nearest-neighbor interpolation simply takes the nearest pixel from the input for each pixel in the output.

Bilinear interpolation takes a weighted average over the closest 2x2 pixels from the input for each pixel in the output.

Bicubic interpolation behaves just like bilinear interpolation, but over a 4x4 area of pixels.

Box sampling considers each pixel in the output as a box over the input, and gives them the value of the weighted average of all pixel in this box.

The Hamming filter is a windowing filter using the cardinal sine function, with a tapered bell-shaped curve.

The Lanczos filter is a windowing filter using the cardinal sine function, with an untapered lobe-shaped curve.

A comparison between these filters is presented in Figure 3.9. The filter with the best results in the context of this research will be selected later.

2.3.3 Simple Thresholding

Segmenting is the process of dividing an image into the foreground and background. The simplest way to segment an image is to convert it to grayscale first. Then, an arbitrary value can be chosen as a threshold. Any pixel with a higher value than this threshold is assigned to be one extreme value (black or white), and any other pixel is assigned to be the other extreme.



Figure 2.4: Morphological erosion applied to an image with a dark foreground and an image with a dark background. Source: docs.opencv.org.

2.3.4 Adaptive Thresholding

In the case that lighting conditions differ throughout an image, using a single global value as a threshold would be unsuitable. This threshold might work for part of the image, but would not work for darker or lighter areas. In this case, adaptive thresholding can be used, which calculates a threshold for each pixel individually. A distinction is made between two variations:

Adaptive mean thresholding [IC18] calculates the threshold value as a mean of some neighbouring area, with an arbitrary constant subtracted from this. This constant can also be zero or negative.

Adaptive Gaussian thresholding [IC18] calculates the threshold value as the weighted sum over some neighbouring area, where the weights are a Gaussian window. An arbitrary constant is then subtracted from this.

After calculating the local threshold for a pixel, it is compared to the grayscale value of that pixel to determine which of the two extremes should be assigned. In Figure 2.5, the two methods listed above are compared to simple thresholding.

2.4 Image Properties

Images and contours have a large number of properties, some of which require further explanation.

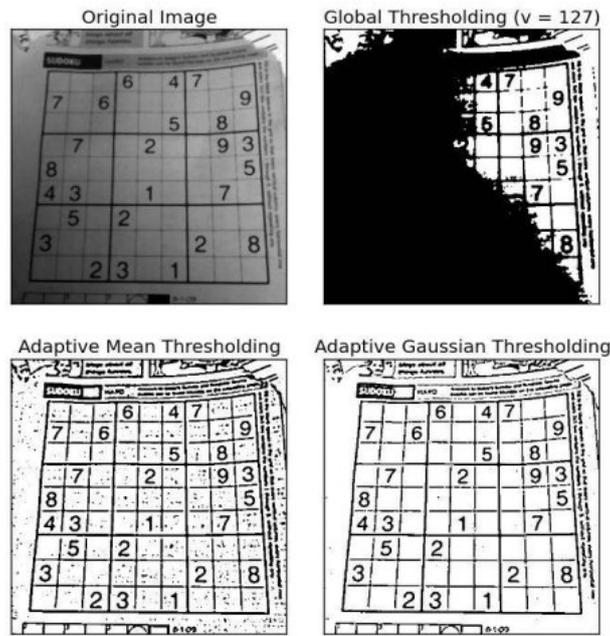


Figure 2.5: Comparison between global and adaptive thresholding. Source: docs.opencv.org.

2.4.1 Moments

Image moments [IC18] are very specific weighted sums over all pixels of an image, where the exact meaning depends on how the weights are chosen. For digital images, the moment is defined as such:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y) \quad (2.1)$$

where $I(x, y)$ is the intensity of the pixel at location (x, y) .

If the weights i and j are both chosen to be 0, the moment M_{00} is calculated as follows:

$$M_{00} = \sum_x \sum_y x^0 y^0 I(x, y) \quad (2.2)$$

$$M_{00} = \sum_x \sum_y I(x, y) \quad (2.3)$$

which is simply the sum of all pixel intensities of the entire image. In the case of a binary image, this is equivalent to the area.

2.4.2 Centroid

The centroid [IC18] of an image is its centre of mass. It is calculated from the image moments according to the following formula:

$$\{\bar{x}, \bar{y}\} = \left\{ \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right\} \quad (2.4)$$

2.4.3 Convex Hull

An arbitrary curve can have sections which are flat, concave or convex. A convex hull [IC18] for such a curve is another curve consisting only of convex or flat sections, which matches the original curve as closely as possible. A convex hull can be created by replacing all concave sections of a curve with straight lines. In other words, a convex hull for a curve is what would be obtained after stretching a rubber band around it.

2.4.4 Convexity Defects

Any deviations between a curve and its convex hull are called convexity defects [IC18]. For each section of the curve, this can be represented as a point on the curve located the furthest from the convex hull. An example of a curve, convex hull, and associated convexity defects is given in Figure 2.6.

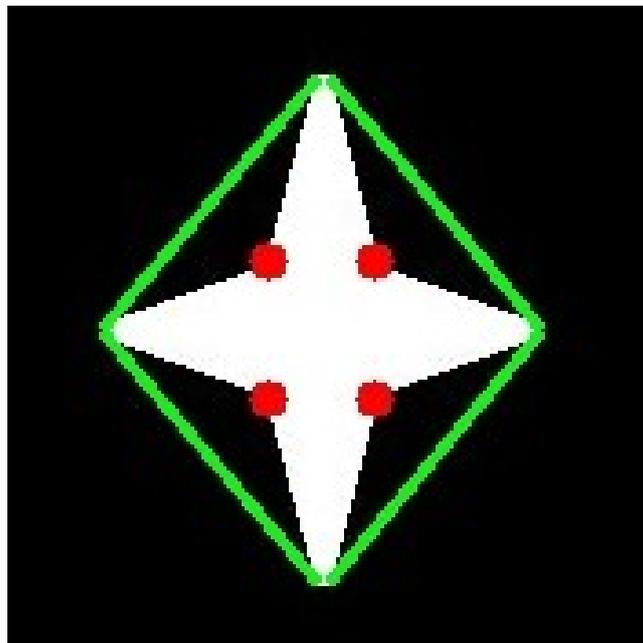


Figure 2.6: A figure displaying convexity defects. The white object represents the original curve. The green lines represent the convex hull. The red dots represent the convexity defects. Source: docs.opencv.org.

2.4.5 Solidity

The solidity [IC18] of an object is the ratio of the object area to the object's convex hull area. This is taken as the fraction of the object's convex hull which is actually occupied by the object. Thus, the more concave an object is, or the more holes it contains, the lower the solidity becomes. This gives an idea of how 'solid' an object is.

2.5 Violin Plots

When using box plots to plot quantitative values across a small number of categories, only the four quartiles are presented. This gives a general overview of the distribution of the data, but the distribution within each quartile remains unknown. When the distribution of data contains more than one peak, conclusions based on box plots might be very misguided. Violin plots [HN98] are a variation of box plots. In addition to displaying the aforementioned four quartiles, violin plots show the complete distribution of data per category. This distribution is usually smoothed by a kernel density estimate. Figure 2.7 shows how a box plot and kernel density estimate can be combined into a single graph.

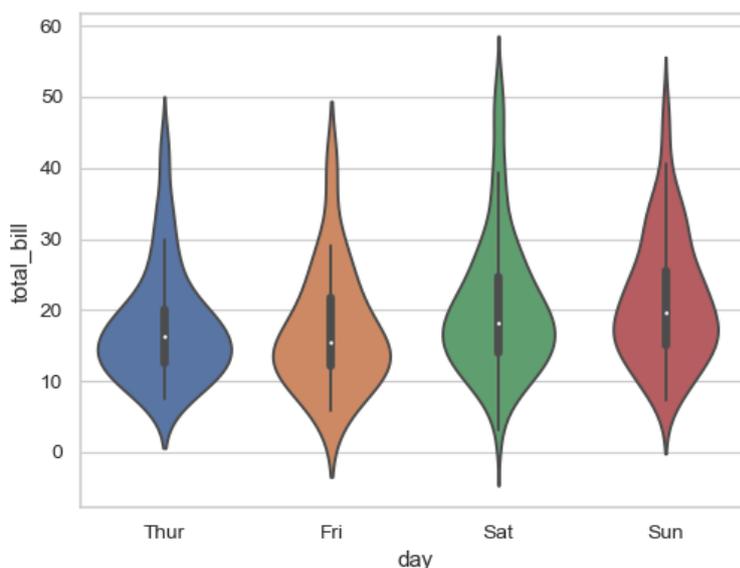


Figure 2.7: Example of a violin plot which plots the values for `total_bill` against the category `day`. Source: `seaborn.pydata.org`.

2.6 Deep Neural Network

Previous research performed by W. Verhoef explored the use of deep neural networks (DNNs) for the purpose of the segmentation of zebrafish larvae [Ver18]. Despite using the term deep neural networks, the neural networks presented in that research are both deep and convolutional. Experiments were performed on two different architectures, and one was determined to perform better than the other. This architecture, called the residual architecture [HZRS15], is a total of 31 layers deep. To improve the propagation of information, various lateral connections and skip-connections have been added. Altogether, the total amount of parameters is 4,062,689, with the number of trainable parameters being 4,060,705. A more elaborate description of this architecture can be found in Section 3.2 of [Ver18]. This residual architecture will be further evaluated and analyzed in this paper and will be referred to with the term **DNN-ZF**. It should be noted that from the perspective of this paper, this architecture will be treated as a black-box algorithm. As such, a detailed analysis

of the structure of the residual architecture will be omitted here. Any other work which can only be done by interacting with the content of the architecture will also be ignored.

Chapter 3

Implementation

The research presented in this paper is divided into two very distinct blocks. The first block concerns the gathering, processing, and organizing of various datasets. The second block concerns the use of these datasets to evaluate and improve the DNN-ZF method created by W. Verhoef. Both blocks are themselves divided into multiple steps, leading to the flowchart presented in Figure 3.1.

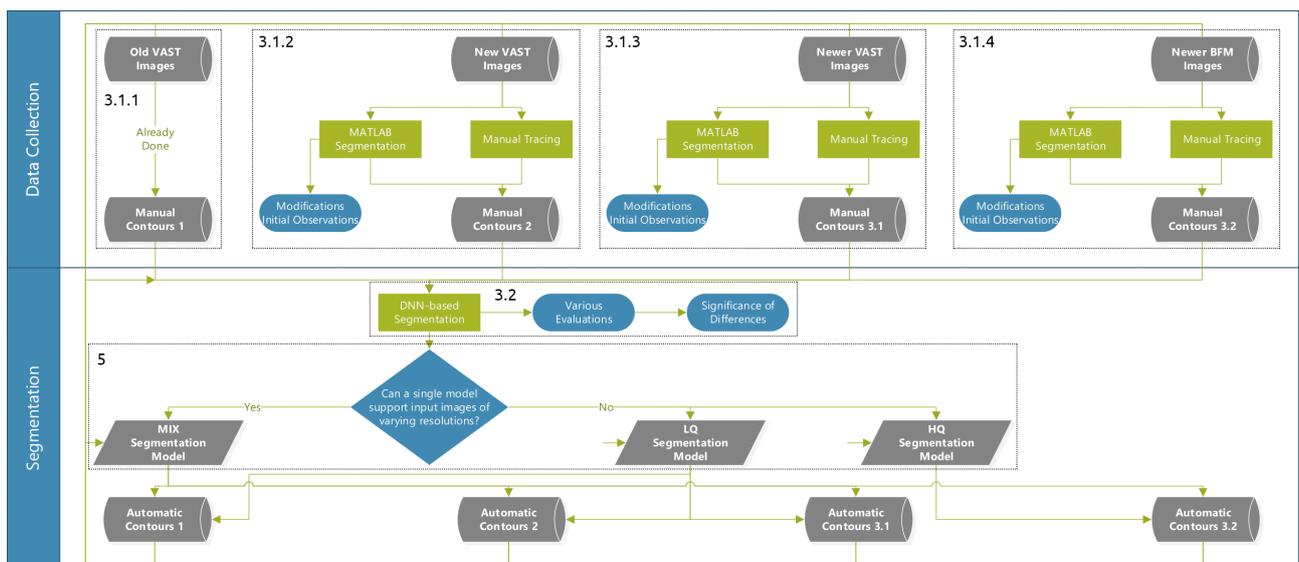


Figure 3.1: Flowchart outlining the performed research.

The first block concerning the gathering of data is fully explained in Section 3.1. Within this block, images from four different sources are considered separately and are outlined as such in the flowchart. All four sources are explained in a dedicated subsection, of which the numbers are also included in the flowchart.

The second block regarding the work with the DNN-ZF method is a lot more involved. In Section 3.2, a full explanation regarding the functional implementation of the DNN-ZF method is given. Additionally, the methods which are used to evaluate the DNN-ZF method are described here. This section covers the outlined area labelled as such in the flowchart. The next phase of this block concerns the answering of one of the research questions. Based on this answer, one or two ideal models are selected for future use. This

is also outlined in the flowchart and is expanded upon throughout Chapter 5. The remaining items on the flowchart indicate the new high-throughput segmentations generated by the ideal model or models. These are of relevance to future research but do not warrant any special attention in this paper.

3.1 Data Collection

Available for use were four datasets consisting out of images of zebrafish larvae, which were created using the VAST BioImager and Leica microscope. The zebrafish larvae used for imaging varied between 3dpf and 5dpf in age. Even though the ages of the zebrafish larvae were known at the time of imaging, there are still variances in their phenotype due to genetic and environmental factors, as would be expected from any organism. Research by H. Spaink [Spa18] has shown that within each age category outliers exists which present a phenotype more typical of another age category.

3.1.1 Old VAST Images

This dataset was created in 2017 using the previously outlined MM-HTAI architecture. For this dataset, the internal VAST camera was used to obtain images of the zebrafish larvae. Each zebrafish larva has been imaged a total of 84 times, with those images being equally distributed around a full rotation of the specimen. All images were saved as TIFF [GP02] images with a resolution of 1024px by 250px. A total of 60 different zebrafish larva were imaged in this way, of which 12 specimens were 3dpf, 24 specimens were 4dpf, and 24 specimens were 5dpf. For one such 4dpf specimen, four of these images are presented in Figure 3.2.

As part of the research performed by Y. Guo in 2017 [Guo17], these images have been segmented using a hybrid method implemented in MATLAB. This produces segmentations such as those presented in Figure 3.2. The hybrid method first creates two different segmentation candidates, one using the mean shift algorithm, and one using an improved level set method [ZMSM08]. It then combines and refines these two candidates to create the final hybrid segmentation.

3.1.2 New VAST Images

This dataset was created in 2018 using the same MM-HTAI architecture and internal VAST camera. For this dataset, a total of 123 zebrafish larvae were imaged, with 39 specimens being 3dpf, 36 specimens being 4dpf, and 48 specimens being 5dpf. This time, the majority of zebrafish larvae were imaged a total of 100 times distributed around one full rotation. A total of three specimens, however, were improperly positioned by the VAST BioImager which caused the imaging to fail for a few of their angles. Once again, all images were saved as TIFF images with a resolution of 1024px by 250px. For one of the properly imaged 4dpf specimens, four of these images are presented in Figure 3.3.

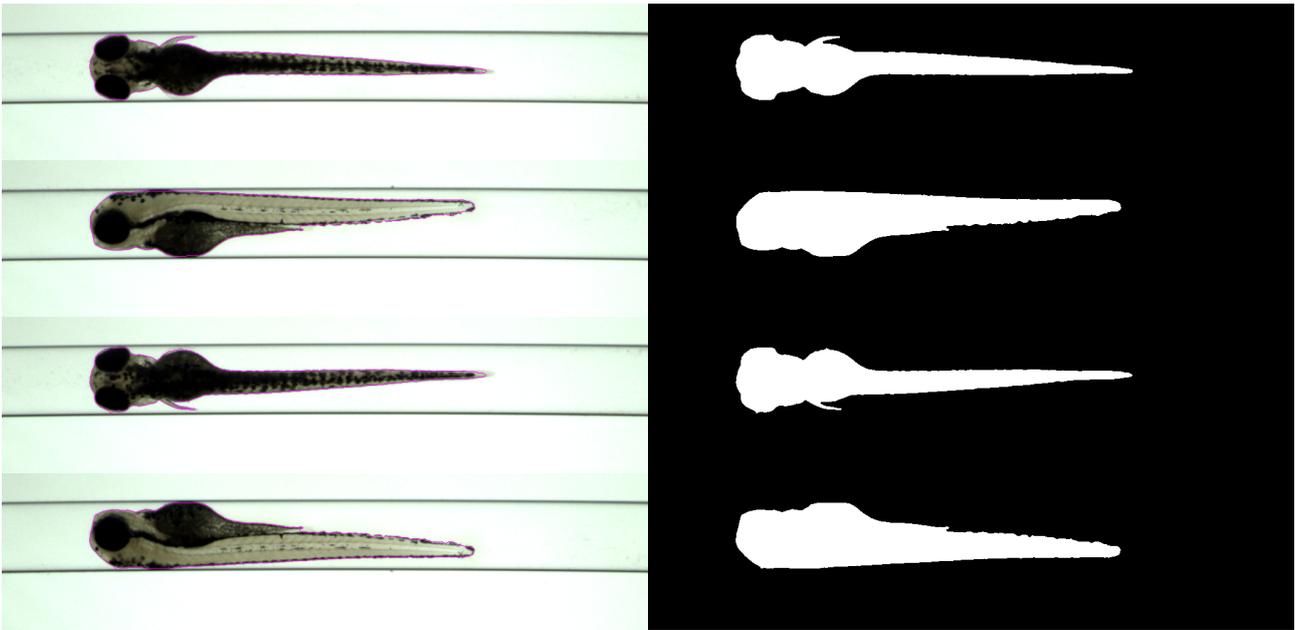


Figure 3.2: Bright field images of four views (ventral, left lateral, dorsal, right lateral) from a typical 4dpf zebrafish larva in the first dataset, as well as their corresponding low-throughput segmentations. The purple line in the bright field images indicates the boundary of the segmentation.

Of particular note is that this dataset contained several specimens for which the images were classified as useless due to several different reasons:

- Certain zebrafish larvae were accompanied by a large amount of debris after being positioned in the capillary, such as the specimen in Figure 3.4a.
- Other zebrafish larvae, after being positioned into the capillary, were closely accompanied by one or more other zebrafish. This would generally cause parts of multiple zebrafish larvae to show up in the same image, such as in Figure 3.4b. In other cases, the VAST BioImager would switch back and forth between the two specimens as it was taking the images at different angles, which caused the set of images to become inconsistent.
- Due to the forces exerted on the zebrafish larvae by the MM-HTAI architecture several of them were damaged severely, such as the specimen in Figure 3.4c.

Since image sets for specimens with one of the properties listed above would either significantly deviate from the norm or be unsuitable for the construction of a sensible three-dimensional model they were discarded and labelled as a separate dataset. A total of 29 image sets have been discarded in this way.

To use this dataset to train and validate the DNN-ZF, ground truth segmentations need to be made. As the accuracy of the hybrid segmentation method developed by Y. Guo has already been proven, the decision was made to use it to segment this new dataset. This method, however, was developed and optimized for only the single dataset created in 2017. As such, its performance on this new dataset was less than desirable. The five initial problems are outlined below:

1. Some segmentations included the capillary as part of the fish. This can be observed in the first image of

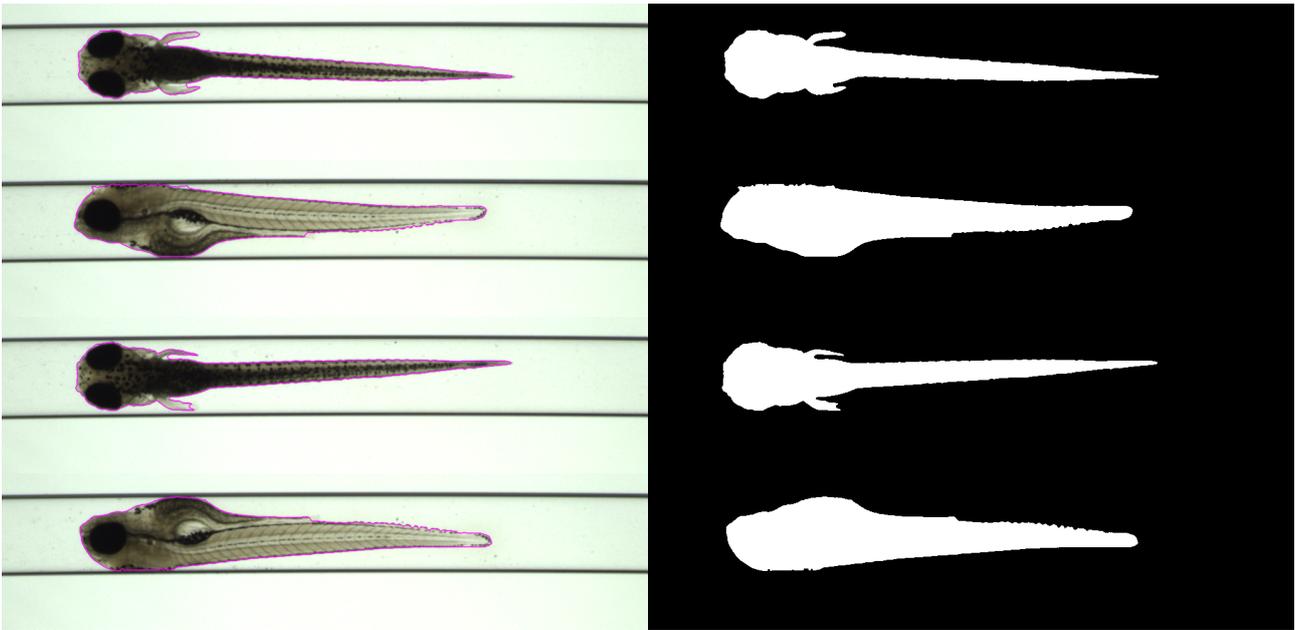


Figure 3.3: Bright field images of four views (ventral, left lateral, dorsal, right lateral) from a typical 4dpf zebrafish larva in the second dataset, as well as their corresponding low-throughput segmentations. The purple line in the bright field images indicates the boundary of the segmentation.

Figure 3.5a.

2. Some segmentations were off by a sliver of fairly constant height. This can be observed in the first three images of Figure 3.5a.
3. Sometimes only a very small section was segmented as part of the fish. This can be observed in the last image of Figure 3.5a. It should be noted that this effect is not limited to only the tip of the tail.
4. Sometimes the very tip of the tail was incorrectly segmented. This can be observed in the second image of Figure 3.5a.
5. Some images failed to be segmented at all, these were mostly images where the zebrafish larva was positioned too close to the edge of the view. Out of the 12288 images which were available in this dataset, only 10841 were successfully segmented.

It was also discovered that the preprocessing code of the hybrid segmentation method did not properly remove the capillaries from images in this new dataset. In some cases, this would lead to the capillaries not being entirely removed, which in turn would cause the first issue outlined above. In other cases, this would lead to part of the fish being removed along with the capillary, causing the second and third issues outlined above. After fixing the code responsible for removing the capillaries, these three issues were also fixed, leading to the segmentations displayed in Figure 3.5b.

Even though the hybrid segmentation method has solved the issue of segmenting the extremely transparent tail of the zebrafish larvae, it has only done so for the single dataset from 2017. On the new dataset, this method still has trouble with segmenting the transparent section of the tail. This indicates that the hybrid segmentation method is not flexible enough to be used on any future datasets as is, and this is what prompted

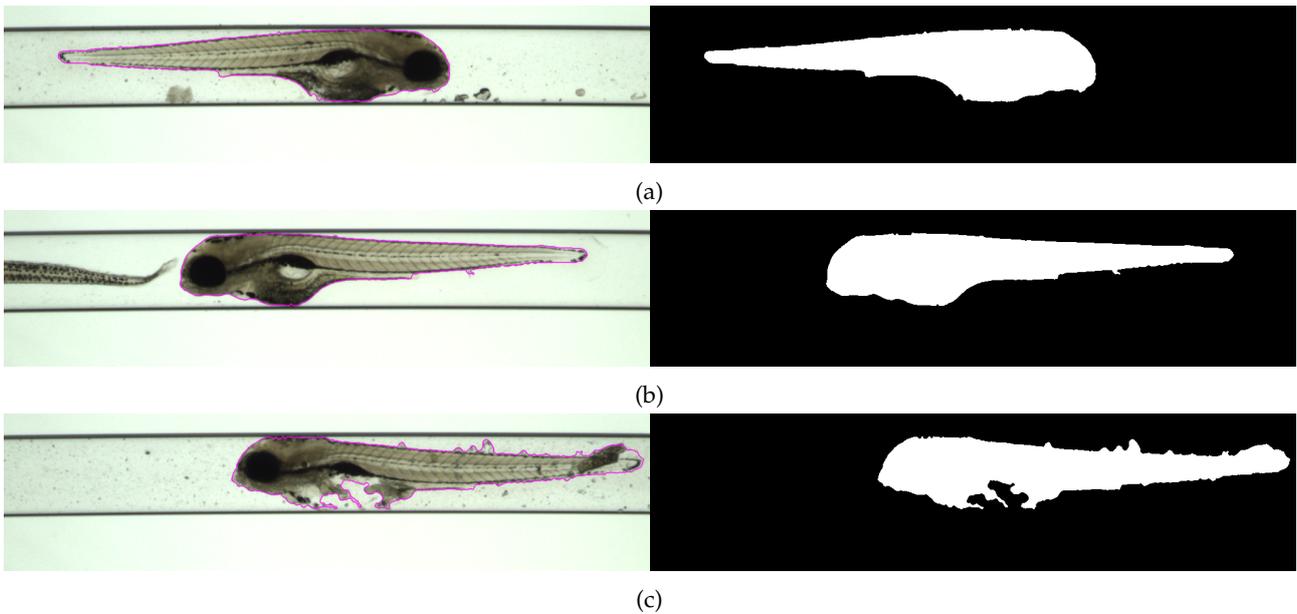


Figure 3.4: Lateral bright field images of three 4dpf specimens from the second dataset which have been discarded for different reasons, as well as their corresponding low-throughput segmentations. The purple line in the bright field images indicates the boundary of the segmentation. Specimen (a) has been discarded due to an overabundance of debris in the capillary. Specimen (b) has been discarded due to the presence of another zebrafish larva in the images. Specimen (c) has been discarded due to extensive damage.

the development of an alternative, more general method. As can be seen in the right lateral views of Figure 3.5, the tail is segmented improperly even after modifying the hybrid segmentation method for this new dataset.

As such, a large number of images in this new dataset are in fact improperly segmented and thus unsuitable as reliable training data. This mostly affects lateral views, in which the tail is the most transparent. To resolve this issue, all images were manually verified and any improperly segmented views were separated into a different dataset. A total of 81 zebrafish larvae were affected (out of 123), and 2790 segmentations were separated (out of 10841). Additionally, to compensate for the comparatively small amount of lateral views, 50 of the lateral images which were improperly segmented, spread out over 23 specimens, were manually traced and segmented by R. Megens. An additional four zebrafish larvae were selected and had all improperly segmented views manually segmented using Sefexa, totalling an additional 194 segmentations. One such specimen is presented in Figure 3.6.

3.1.3 Newer VAST Images

Similar to the previously described dataset, this dataset was created later in 2018 using the same MM-HTAI architecture and internal VAST camera. A total of 35 zebrafish larvae of unknown age have been imaged, with each specimen being imaged 100 times distributed around one full rotation. This time around, a total of six zebrafish larvae had the images for several angles missing as they were being positioned improperly by the VAST BioImager. Once again, all images were saved as TIFF images with a resolution of 1024px by 250px. An overview of one of the properly imaged zebrafish larvae from this dataset is presented in Figure 3.7.

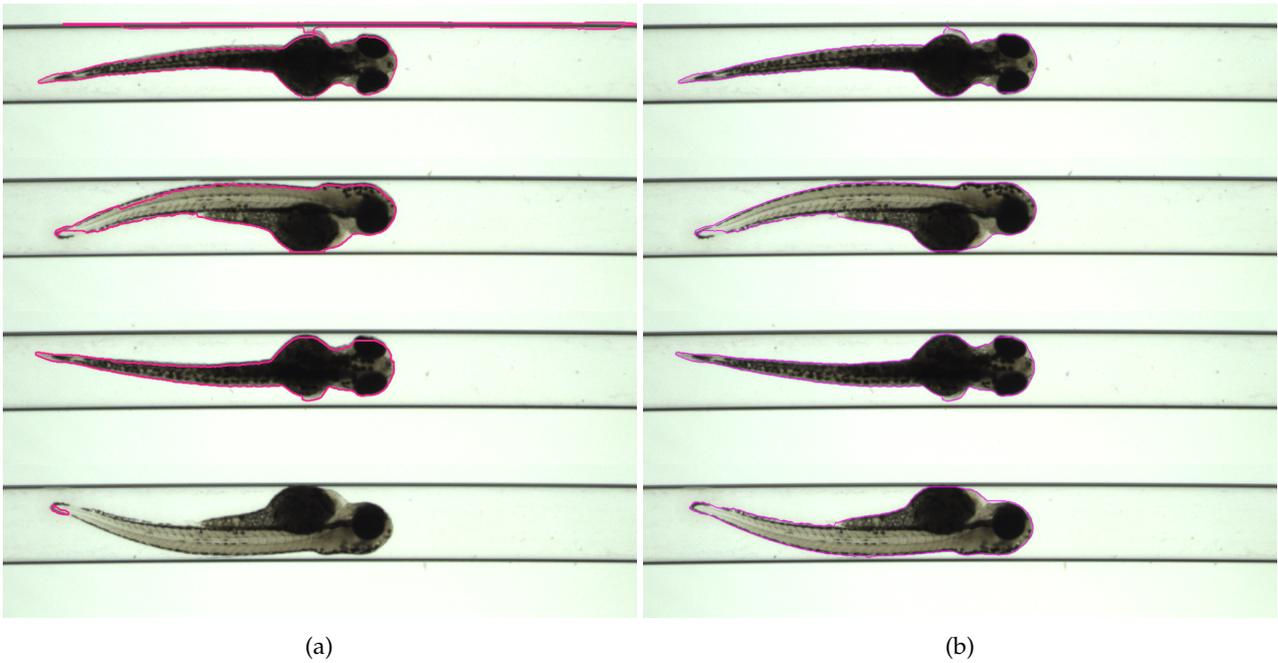


Figure 3.5: Bright field images of four views (ventral, left lateral, dorsal, right lateral) from a typical 3dpf zebrafish larva in the second dataset. The purple line in the bright field images indicates the boundary of the segmentation. The images of segmentation (a) were made with the original version of the hybrid segmentation method. The images of segmentation (b) were made with the modified version.

As described in the previous section, the hybrid segmentation method that is used to generate the training segmentations for the DNN-ZF generalizes poorly to new datasets. Additionally, at the moment this dataset became available it had already been determined that adding more training data beyond a certain point would not improve the quality of the model, as concluded from Figure 4.2. The previous datasets already provide more than enough training data sourced from the VAST BioImager’s internal camera to train any model to the maximum accuracy. As the creation of a new set of training data is extremely time-consuming and not necessary in this case, the decision was made to not segment this dataset using the hybrid segmentation method and instead use it to solely evaluate the metrics which do not require the presence of a ‘truth’ segmentation.

3.1.4 Newer BFM Images

This dataset was created in 2018, using the same specimens as the previous dataset (Newer VAST Images). Unlike the other datasets discussed until this point, this dataset was not created using the internal positioning camera of the VAST BioImager. Instead, the MM-HTAI architecture was fully realized by using the imaging capabilities of the Leica microscope that the system is mounted on. A total of 32 zebrafish larvae of unknown age were imaged by the Leica microscope. This time, each zebrafish larva was imaged 101 times distributed around a full rotation, with eight of the specimens partially failing and not having images provided for all their angles. All images were saved as TIFF images with a resolution of 2560px by 1920px.

The deep neural network created by W. Verhoef was made to accept images from the VAST BioImager’s

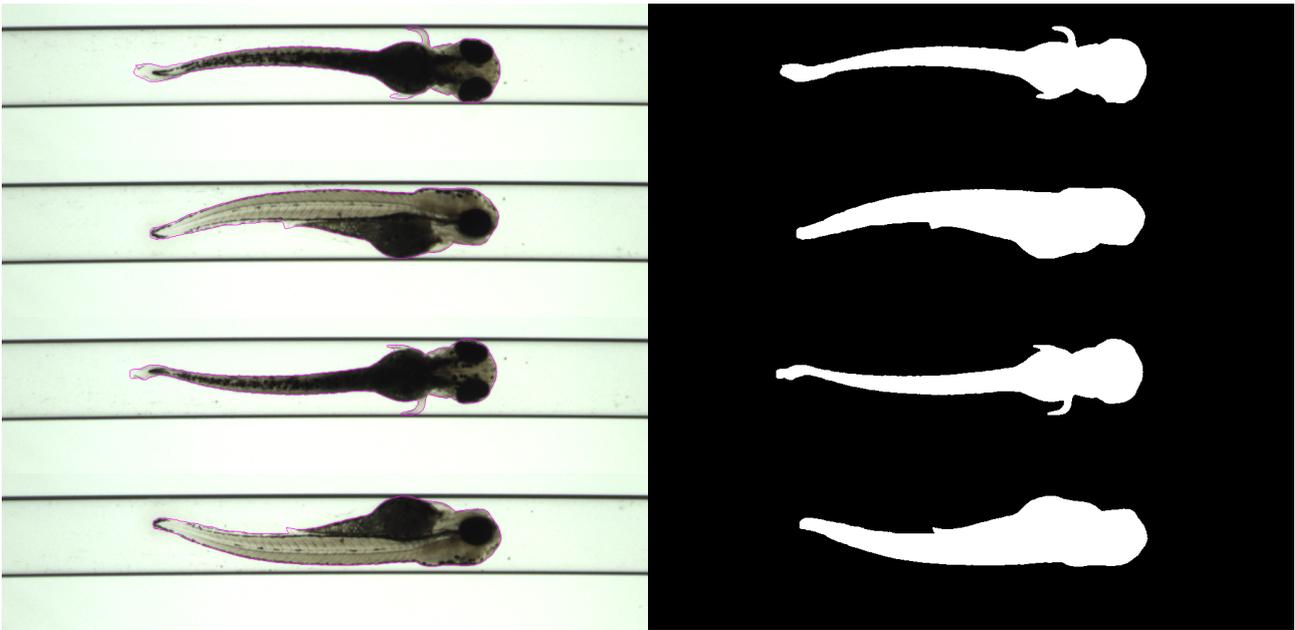


Figure 3.6: Bright field images of four views (ventral, left lateral, dorsal, right lateral) from a typical 3dpf zebrafish larva in the second dataset, as well as their corresponding manual segmentations. The purple line in the bright field images indicates the boundary of the segmentation.

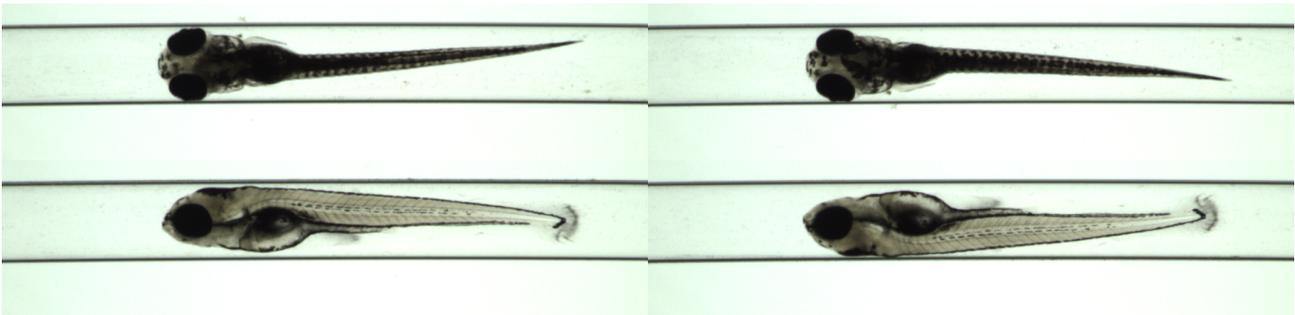


Figure 3.7: Bright field images of four views (ventral, left lateral, dorsal, right lateral) from an typical zebrafish larva in the third dataset. Note that these images are taken from a specimen also displayed in Figure 3.8.

internal camera. As the amount of input nodes for any given neural network is static, this DNN-ZF used will only accept images with a resolution of 1024px by 250px. Since the images of this dataset have a resolution of 2560px by 1920px, they need to be downsampled before they can be used in conjunction with this DNN-ZF. Additionally, the segmentations created by the DNN-ZF will also have a resolution of 1024px by 250px, and will thus need to be upsampled afterwards. To select the most suitable resampling algorithm for these tasks, an arbitrary BFM image was downsampled to the input size of the DNN-ZF, and upsampled back to its original size, for each algorithm supported by OpenCV version 3.4.5.20. The resulting images are organized in Figure 3.9. Since the image that was processed with the Lanczos resampling algorithm looks the most like the original, this was selected for use in any resampling tasks performed on input or output for the DNN-ZF.

Not only do the BFM images from this dataset have a different resolution than the input of the DNN-ZF, but they also have a different aspect ratio. While the aspect ratio of the images from the VAST BioImager's internal camera and thus the input of the DNN-ZF is 512:125 or 4.096:1, the aspect ratio of the images produced by the Leica microscope is 4:3 or 1.33:1. This means that the images are squashed vertically by a factor of 3.072

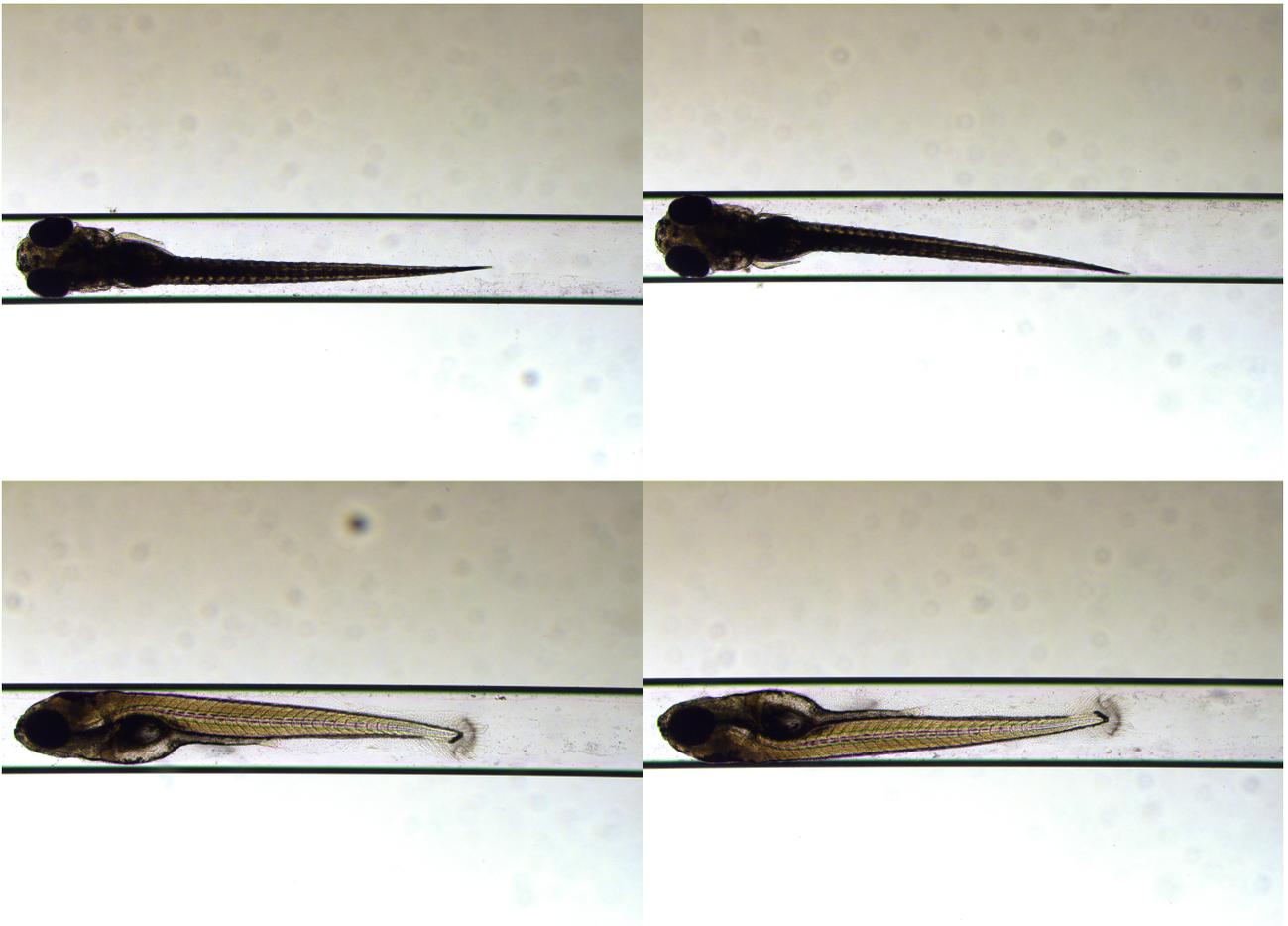


Figure 3.8: Bright field microscopy images of four views (ventral, left lateral, dorsal, right lateral) from an typical zebrafish larva in the fourth dataset. Note that these images are taken from a specimen also displayed in Figure 3.7.

when they are resized to match the input of the network. This changes the shape of the zebrafish larvae in the images relative to those in the other datasets, potentially affecting the quality of the model or the predictions. Additionally, as can be observed in Figure 3.10a, some of the BFM images contain a digitally introduced scale in the top left corner. As such, it is relevant to determine if these properties play a role in the quality of the predicted segmentations for these images, or in the effectiveness of these images as training data. For this purpose, a copy of this dataset was created with all of the BFM images cropped to the aspect ratio of the VAST images, such as the image presented in Figure 3.10b.

At this point, it was obvious that the hybrid segmentation method would not be able to provide acceptable training segmentations for this dataset. However, due to the novelty of this dataset over the others, it could not simply be ignored like the previously discussed dataset. As such, it was necessary to develop a new segmentation method which would work on the majority of this new dataset. The new method is outlined in Figure 3.11, and explained in detail below. This method was used to generate training segmentations for this dataset of BFM images, as well as the dataset of cropped BFM images.

original image The process starts by ensuring that all BFM images are cropped to an aspect ratio of 512:125, the same aspect ratio as the VAST images. This does not affect images from the pre-cropped BFM dataset,

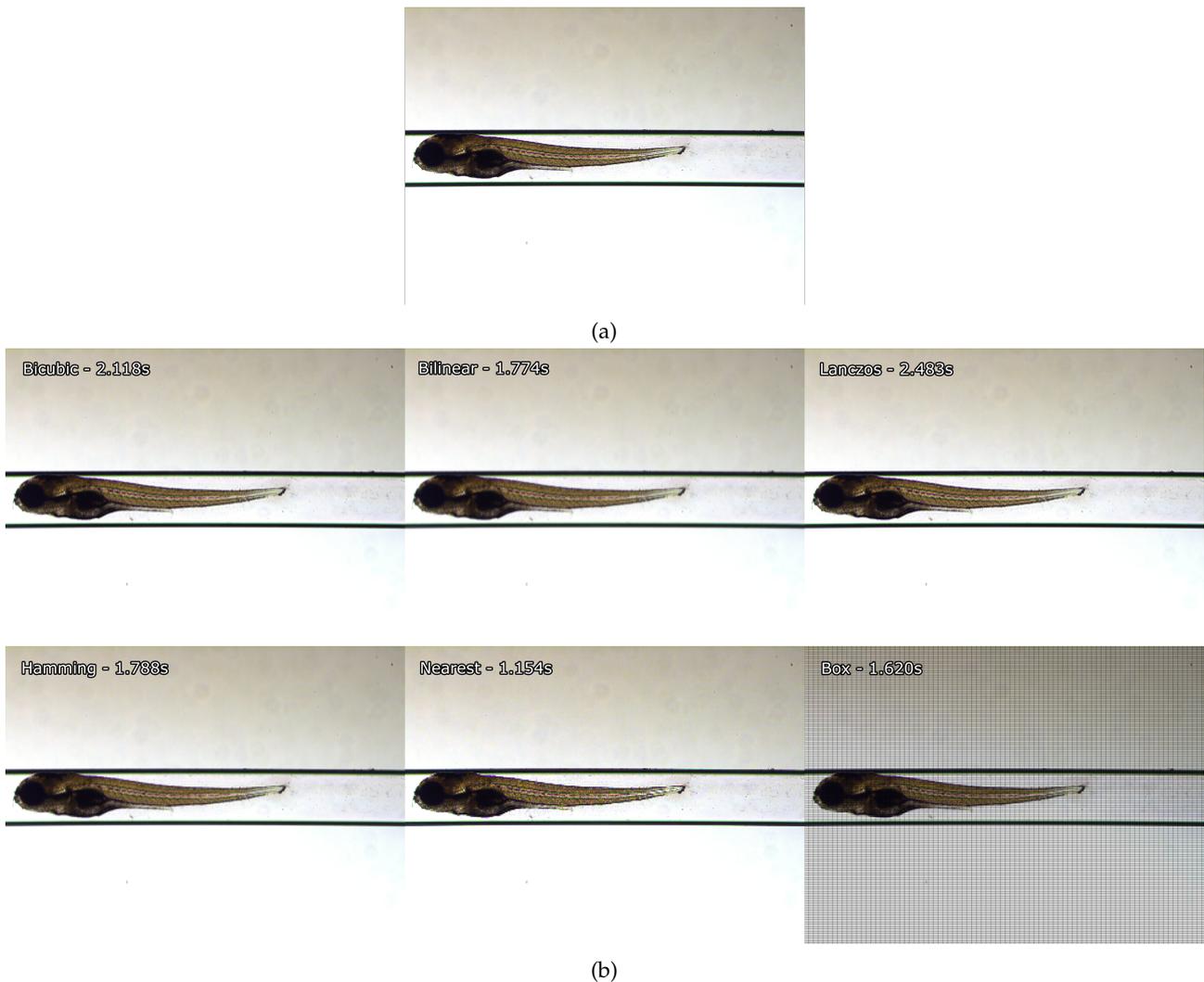


Figure 3.9: Comparison of various algorithms for the interpolation of BFM images of zebrafish larvae. (a) presents the original image which has been interpolated, with a resolution of 2560px by 1920px. (b) presents several images which have been down sampled to 1024px by 250px and up sampled back to the size of the original, both steps with the marked algorithm. The presented times are the duration of ten individual rounds of down sampling followed by up sampling for each marked algorithm.

as they already are at this aspect ration. In addition to ensuring consistency between the two variants of the dataset, this cropping step removes a large amount of the background as well as the digitally introduced scale, which would otherwise be detrimental to the later steps.

grayscale Afterwards, the images are converted to grayscale. For this process, there is no particular value in having three different colour channels over just one, and this conversion simplifies the following steps.

remove capillaries Now, the capillaries are removed from the image in such a way that any overlapping parts of the zebrafish larva are not affected. This is done separately for both sides (top and bottom) of the image. For each vertical line of pixels, a search starts for the outer edge of the capillary. This edge is determined by searching for a sharp decrease in pixel intensity relative to the previous pixel. Additionally, to prevent issues with dark pieces of debris or air bubbles outside the capillaries, the location of this new edge is clamped within a certain range of the edge found in the previous column.

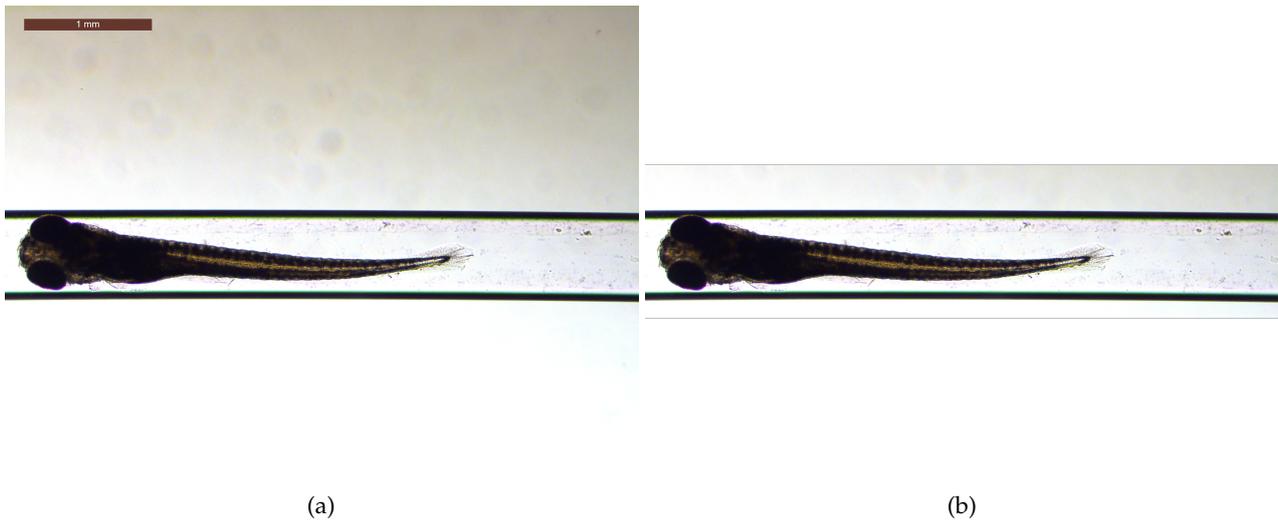


Figure 3.10: An image from the fourth dataset before and after cropping. (a) presents a BFM image as it was originally captured, with a resolution of 2560px by 1920px. (b) presents a BFM image after cropping an equal height from the top and bottom, with a resolution of 2560px by 625px.

Afterwards, an interval starting at this edge and going inwards for a certain distance is considered, and the location of the pixel with the highest intensity in this interval is extracted. The length of the interval depends on the average pixel intensity of another static interval inwards. This process is designed such that when a zebrafish larva partially overlaps the capillary, such as in the example in Figure 3.11, the dynamic interval is much shorter. In turn, this causes the selected lightest pixel to fall in between the zebrafish and the capillary, instead of falling somewhere within the body of the zebrafish. Finally, an appropriate background colour is calculated, and the interval from the edge of the image to the selected lightest pixel is set to this colour.

grey erosion After removing the capillaries, grey erosion is applied. This erosion will cause the black areas of the image to expand according to a kernel, thereby filling any holes which might be present in the specimen. To conform to the morphology of the zebrafish larvae, a circular kernel is used.

adaptive Gaussian thresholding After the capillaries are removed and any holes in the specimen are filled, thresholding can occur. After some experimentation, it was determined that using adaptive Gaussian thresholding with a block size of 1111 provides the best results.

revert grey erosion Now, the earlier applied grey erosion needs to be reverted. This is done by eroding the foreground of the threshold with the same kernel used earlier. In doing so, the threshold belonging to the zebrafish larva is reduced to the actual size of the specimen.

select fish segmentation Finally, the full segmentation can be generated. First, a new black image with the same dimensions as the input is created, thereby undoing any cropping. Afterwards, the threshold belonging to the fish is selected from among the several disconnected thresholds and is placed on the new background at the appropriate location.

This method ended up working extremely well for the majority of BFM images. However, with certain images the novel method failed spectacularly, such as for the image in Figure 3.12, indicating reliability issues. As such,

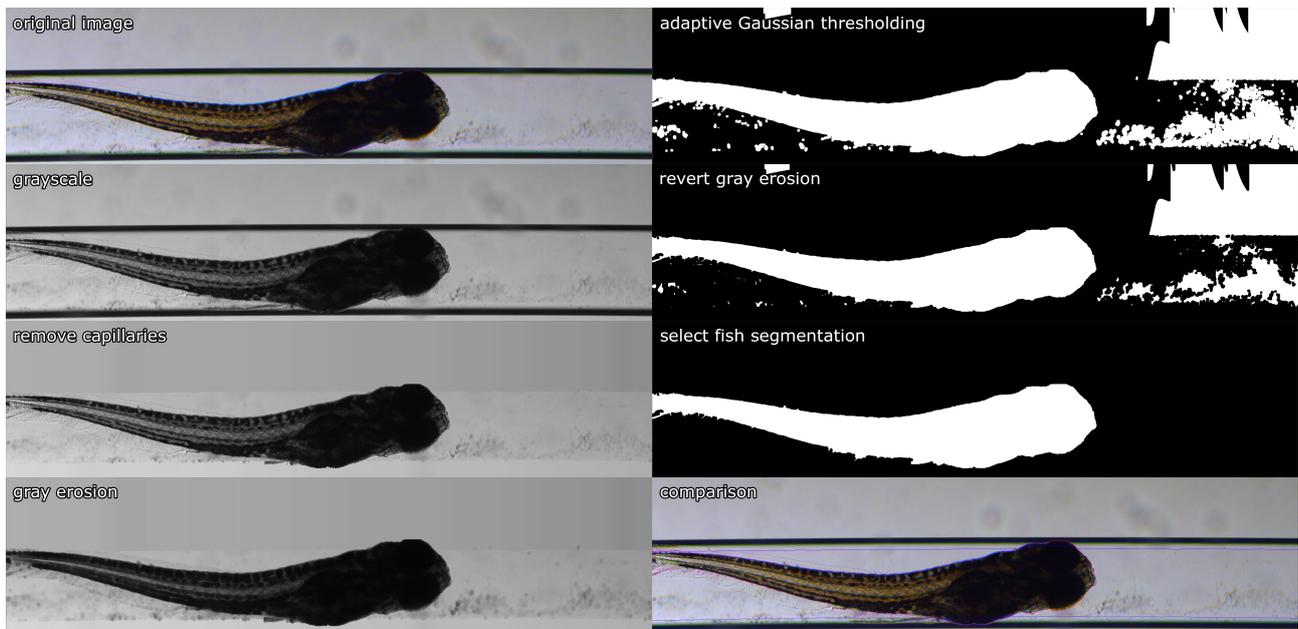


Figure 3.11: An overview of a novel method for the segmentation of zebrafish larvae in BFM images. (1) if necessary, the image is cropped to an aspect ratio of 512:125. (2) the image is converted to grayscale. (3) the capillaries are removed from the image, without affecting any overlapping parts of the larva. (4) grey erosion is applied to fill any holes. (5) adaptive Gaussian thresholding is used for an initial segmentation of the image. After this, the initial cropping is also reverted. (6) the previously applied grey erosion is reverted. (7) the segmentation belonging to the zebrafish larva is isolated from the others. (8) an overview is created from the original BFM image. The purple line represents the boundary of the segmentation. The blue lines represent a start and end point related to removing the capillary.

it is more desirable to train the DNN-ZF method on all the successful segmentations, and use that method instead.



Figure 3.12: Failed segmentation of a BFM image using the novel method, with the corresponding bright field image. The purple line in the bright field image indicates the boundary of the segmentation.

3.1.5 Dataset Summary

To make the remainder of this paper easier to understand, a concise summary of the processed datasets is provided.

Source	Name	# of Specimen	# of Images	Has Segmentations?
Old VAST Images	<code>raw_data</code>	60	5040	✓
New VAST Images	<code>raw_new_data</code>	123	12288	✗
	<code>raw_new_data_bad</code>	81	2790	✓
	<code>raw_new_data_broken</code>	29	1926	✓
	<code>raw_new_data_good</code>	79	6125	✓
	<code>raw_new_data_manual</code>	27	271	✓
Newer VAST Images	<code>raw_newer_data_vast</code>	35	3403	✗
Newer BFM Images	<code>raw_newer_data_bfm</code>	32	2730	✓
	<code>raw_newer_data_bfm_cropped</code>	32	2730	✓

Table 3.1: A listing of all the available datasets along with their most important properties.

`raw_data` A dataset of VAST images which was finalized prior to this research.

`raw_new_data` A dataset of VAST images with no segmentations.

`raw_new_data_bad` A dataset of VAST images with incorrect segmentations made by the hybrid segmentation method.

`raw_new_data_broken` A dataset of VAST images containing physically damaged zebrafish larvae with segmentations made by the hybrid segmentation method.

`raw_new_data_good` A dataset of VAST images with segmentations made by the hybrid segmentation method.

`raw_new_data_manual` A dataset of VAST images with segmentations which were manually traced.

`raw_newer_data_vast` A dataset of VAST images with no segmentations.

`raw_newer_data_bfm` A dataset of unmodified BFM images with segmentations made by the novel method.

`raw_newer_data_bfm_cropped` A dataset of cropped (to the aspect ratio of the VAST images) BFM images with segmentations made by the novel method.

As these datasets were being created, four collections of datasets were designed which would likely work well as training data based on prior knowledge and initial observations. Starting with only the old VAST images, each further collection would add one additional dataset. These promising collections are as follows:

1. `raw_data`
2. `raw_data+raw_new_data_good`
3. `raw_data+raw_new_data_good+raw_new_data_manual`
4. `raw_data+raw_new_data_good+raw_new_data_manual+raw_newer_data_bfm_cropped`

Additionally, a fifth collection was considered containing only BFM images, as using VAST images for training could potentially be detrimental for the prediction of BFM images:

5. `raw_newer_data_bfm_cropped`

3.2 Segmentation

After the creation of all the datasets was completed, the DNN-ZF method was evaluated. This was done by generating many different sets of training and testing data from those datasets in several different ways.

3.2.1 Model Generation and Training

To evaluate any deep neural network with a certain architecture, models need to be created, trained, and tested on certain metrics. The properties of these models will be outlined below:

Architecture The architecture which will be evaluated is a residual architecture developed by W. Verhoef, which is described in Section 4.1 of [Ver18].

Input Dimensions The input to the network is three dimensional, with a width of 1024, a height of 256, and a depth of 3. This corresponds to an image with a resolution of 1024px by 256px, and three colour channels.

Output Dimensions The output from the network is two dimensional, with a width of 1024 and a height of 256. This corresponds to a grayscale image with a resolution of 1024px by 256px.

Optimizer The optimizer used when training the models is one called Adam [KB15]. The only parameter changed from the defaults as discussed in the paper is the learning rate, which has been decreased to 0.0001.

Loss Function The loss function used when training the models is the binary cross-entropy¹. This function works well in networks where each output node predicts a value between 0 and 1.

When training the model, it is always trained one image at a time. Additionally, each of those images is given a separate epoch, thus there is only one step per epoch.

3.2.2 Preprocessing and Postprocessing

Before any images were used as input for the DNN-ZF, they had to be processed, which was done in two different ways. Images and segmentations which were used to train the DNN-ZF would undergo data augmentation first. This included shifting, rotating, zooming, and flipping. Naturally, these modifications

¹https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html#cross-entropy

would be the same for an image and its related segmentation. Afterwards, both the image and the segmentation would be resized to a resolution of 1024px by 256px using the Lanczos interpolation algorithm, which was required to match the input dimensions of the DNN-ZF. Images and segmentations which were used to test the DNN-ZF only underwent resizing to the input dimensions of the DNN-ZF and did not undergo data augmentation.

Of particular note is that every evaluation on this DNN-ZF was performed twice. Once with the preprocessing as described above, and once with a filter applied. This filter was applied before the data augmentation and resizing and would be applied to both the training and testing image, but not the segmentations. The filter used was a gray erosion filter with a circular kernel. For the images used, this filter would increase the area of any dark parts, causing the effect observed in Figure 3.13. Since this filter was not applied to the segmentations used as training or evaluation data, any model trained with this filter enabled would still be motivated to predict segmentations of the correct size and shape.

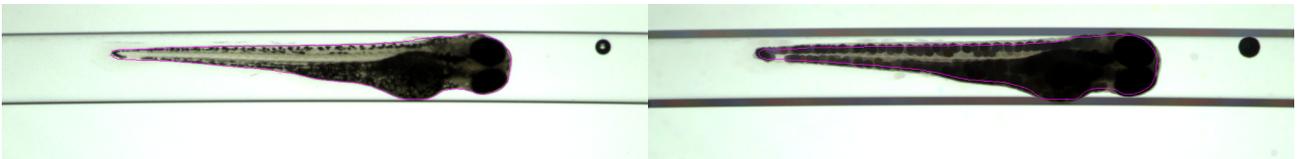


Figure 3.13: Image of a zebrafish larva before and after the application of a gray erosion filter with a circular kernel. The purple lines indicate the boundaries of the segmentations predicted by a model trained on only the `raw_data` dataset.

The output which was predicted by any trained model also needed to be processed before it would be a proper segmentation. First of all, the predicted output consisted of floating-point values between 0 and 1. To create a binary segmentation, a simple rounding operation was performed. Any pixel with a value of ≤ 0.5 was rounded off to 0, indicating the background. Any pixel with a value > 0.5 was rounded off to 1, indicating the foreground. Afterwards, all separated parts of the segmentation would be identified, and only the part with the biggest area would be selected as the final segmentation. Additionally, any fully enclosed holes in this segmentation would be filled in.

3.2.3 Metrics

To evaluate the predictions made by any particular model, several different metrics were implemented. These metrics were calculated from the final segmentation after the postprocessing step described above. Special care was taken to provide metrics which are relevant in regards to the contents of the predictions, rather than just the values of the different pixels.

Accuracy

An accuracy metric was implemented to get a grasp on the similarity between the segmentations made by the model and the curated segmentations made by the various other methods. This metric was implemented as

the fraction of pixels with the same value between these two segmentations. Note that after the postprocessing, both segmentations are binary images.

Tail Solidity

As the most prevalent problem of the various segmentation methods is the segmentation of the tail, which is extremely transparent, a metric was implemented to measure the quality of the segmentation in this area. To calculate this metric, the orientation of the zebrafish larva needs to be determined first. This is done by taking the centroid of the segmentation, which is generally located near the yolk sac. If this point is located to the left of the centre of the bounding box of the segmentation, it indicates that the zebrafish larva is orientated with its head on the left side. If the centroid is located on the right, the opposite is true. Once the orientation of the fish is known, the segmentation of the tail of the larva is isolated. More specifically, the image is cut vertically at 15% inwards from the tail end of the bounding box. Once only the segmentation of the tail is remaining, the area of that segmentation as well as the area of the convex hull of that segmentation is determined, from which the solidity is calculated. A solidity close to 1 will indicate that the entirety of the tail has been detected in the segmentation, as in Figure 3.14b. If the tail has been segmented incorrectly, it will lead to a notch missing from the segmentation of the tail, leading to a lower value for this metric, as the tail is now less 'solid'. This can be observed in Figure 3.14a.

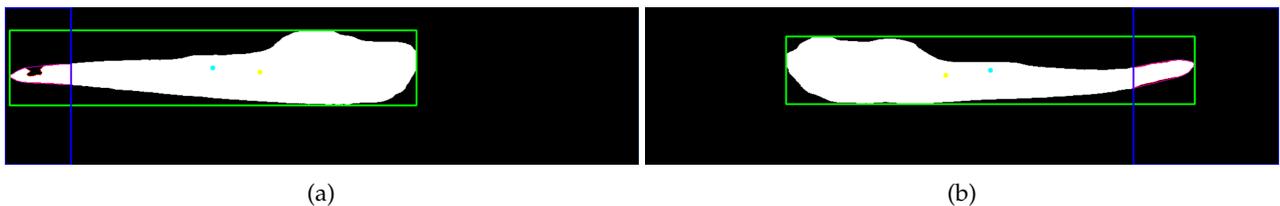


Figure 3.14: Overview of the various image properties used during the calculation of the tail solidity metric. The green rectangle represents the bounding box of the segmentation. The cyan dot is the centre of this bounding box. The yellow dot represents the centroid of the segmentation. The blue rectangle represents the part of the image which has been determined to contain the tail of the zebrafish larva. Finally, the red line outlines the segmentation of this tail, and the purple line represents the convex hull of this segmentation.

3.2.4 Evaluations

Three different methods were used to train and evaluate models for this DNN-ZF, which will be described below.

K-fold cross-validation

Due to its versatility, 5-fold cross-validation was used to evaluate the DNN-ZF. For this method, any single dataset or combination of datasets could be used. All the samples in the selected datasets would be separated into five folds. Here, a sample refers to all the images taken of a single zebrafish larva, and not an individual

image/segmentation pair. For each fold, a new model would be created and trained on all the samples for that fold. Then, this model would be evaluated on all the samples from the four other folds.

The results of the evaluations were saved in a .csv file, in which each evaluated image/segmentation pair received its own entry. Since the separation into folds was done at the level of specimens rather than images, and the number of images per specimen contained some variation, the models for certain folds had more training data available than others. As such, this information was also saved with each entry so that any further processing could be weighted depending on the amount of training data used.

Incremental k-fold cross-validation

To get a good image of how the quality of a model changes as more training data is added, a variation on 5-fold cross-validation was used. For this incremental variation, the selected datasets were divided into folds as described above. Once again, the processing of each fold starts with the creation of a new model. Then, until all the samples in the fold have been processed, this model undergoes many rounds of training and testing. For each round, the model is first trained on all images belonging to only a single sample. Then, this partially trained model is evaluated on all samples from the four remaining folds. Each iteration of this process builds upon the model from the previous rounds.

The results of this process are saved in the same manner as the non-incremental k-fold cross-validation, which is described above. This means that every entry also contains the number of training images used for the model which was evaluated.

Train/Test split

As many of the available datasets are completely independent of each other, it is possible to split training from testing data using this existing separation of datasets. For this method, any single dataset or combination of datasets was selected as training data. Then, any single dataset or combination of datasets that does not intersect the training data was selected as testing data. After the selection of training and testing data, a single model was created, trained, and evaluated.

The evaluation results for this method are saved in the same manner as the other two methods, as described above. The amount of images used to train the model is also included.

Chapter 4

Results

In this chapter all our experimental results will be presented. Most of these results come from the experimental setup explained in Chapter 3. Depending on the goal of the evaluation, different datasets are used for training and testing, which will allow us to draw varying conclusions. Additionally, a set of experiments is presented where metrics were gathered on models trained with increasing amounts of data. Finally, an experiment is presented which will give some insight into the functionality of the DNN-ZF method.

4.1 Evaluating with increasing amounts of training data

In order to visualize what happens to the model as more training data is added, as well as to find a potential maximum for the amount of training data needed, incremental k-fold cross-validation was used with the five promising collections of datasets. The results for the `raw_data` dataset are visualized in Figure 4.1, those for `raw_data+raw_new_data_good` in Figure 4.2, for `raw_data+raw_new_data_good+raw_new_data_manual` in Figure 4.3, for `raw_data+raw_new_data_good+raw_new_data_manual+raw_newer_data_bfm_cropped` in Figure 4.4, and for `raw_newer_data_bfm_cropped` in Figure 4.5. In each of these figures, the results from the five different folds are combined into a single line, with an error band around each line indicating the confidence interval.

As can be seen in Figure 4.2, Figure 4.3, and Figure 4.4, all metrics stabilize after the model has been trained on 3000 images. This is equivalent to roughly 30 fish. The results in Figure 4.1 also match this observation, although with lower certainty. After this point, training on additional images will provide no significant changes to the quality of the model. Each new image used as training data will still cause changes in the model, but over time this will lead to oscillations in the metrics, rather than a change towards either direction, as can be observed in the figures.

The model evaluated in Figure 4.5 however, which is being trained on BFM images only, never reaches 3000 training samples. Generally speaking, this graph follows the same trend as the others, and will most likely exhibit the same behaviour after reaching 3000 training images. A big difference, however, is that the

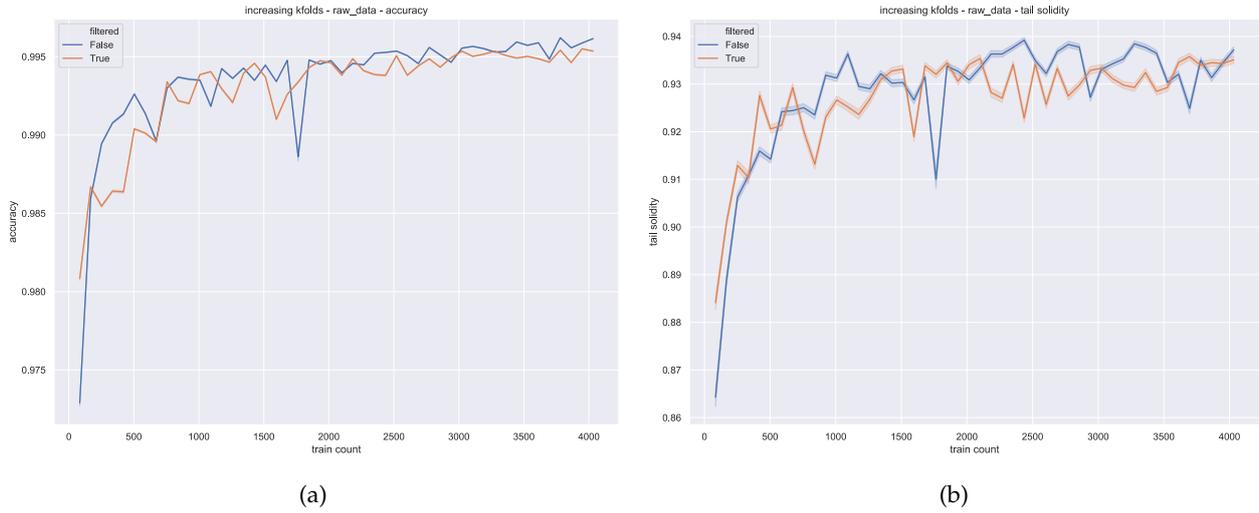


Figure 4.1: Results for evaluating a model using different metrics during the training process. This model is being trained with the `raw_data` dataset.

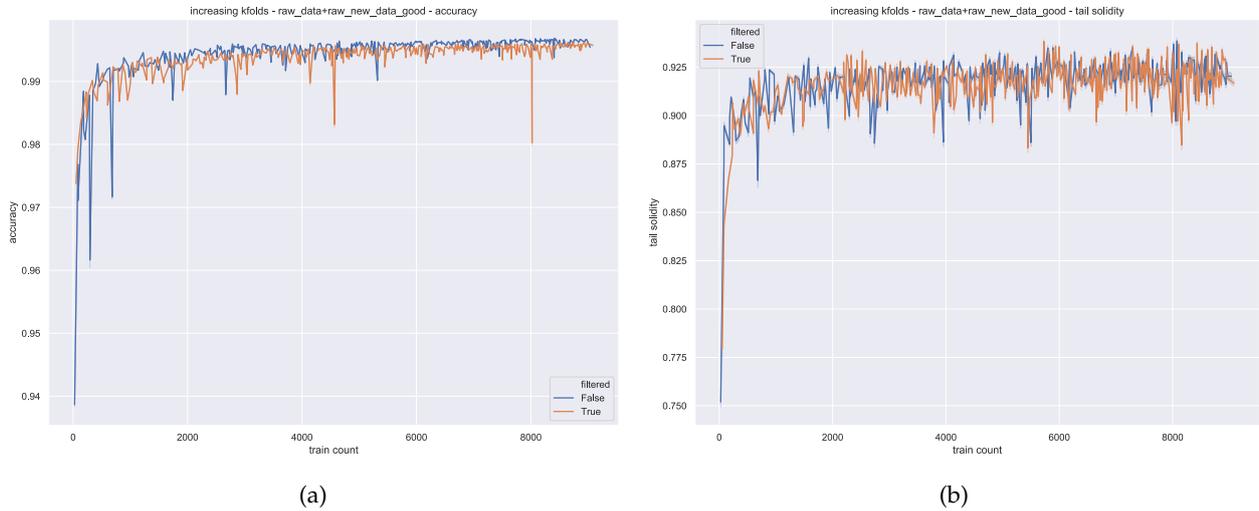


Figure 4.2: Results for evaluating a model using different metrics during the training process. This model is being trained with the `raw_data` and `raw_new_data_good` datasets.

oscillations of the metrics in this model are much more significant than those in the other models.

4.2 Evaluation of models based on different training data

One of the most important questions that this research aims to answer is if one model is sufficient for the prediction of both images from the VAST and the BFM source, or if two separate models are required. Additionally, the most ideal or two most ideal models need to be identified. For this purpose, the five models trained with the most promising datasets were evaluated on only VAST images in Figure 4.6, and evaluated on only BFM images in Figure 4.7. After combining all the relevant data, each of the five models is associated with a large number of numbers for both metrics, which were plotted as violin plots. This way, in addition to the four quartiles, information about the distribution of data both within and outside these quartiles is

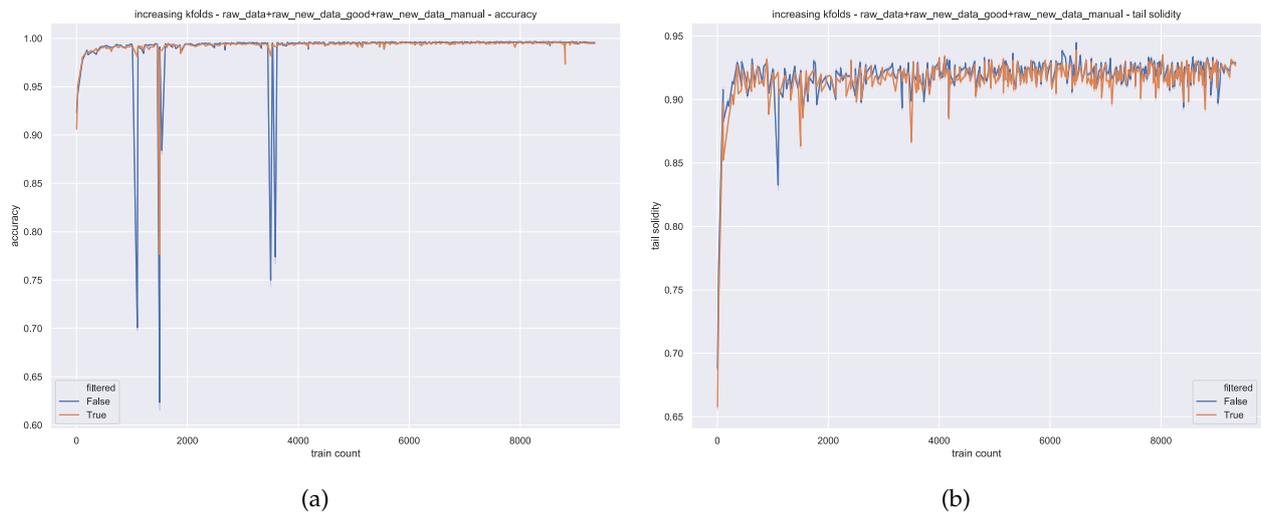


Figure 4.3: Results for evaluating a model using different metrics during the training process. This model is being trained with the `raw_data`, `raw_new_data_good`, and `raw_new_data_manual` datasets.

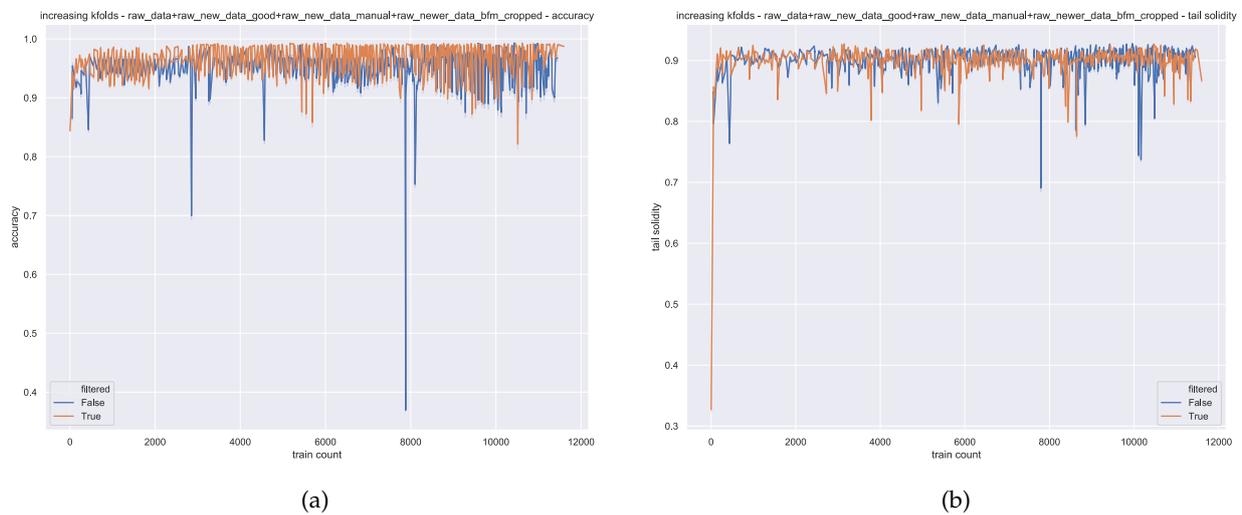


Figure 4.4: Results for evaluating a model using different metrics during the training process. This model is being trained with the `raw_data`, `raw_new_data_good`, `raw_new_data_manual`, and `raw_newer_data_bfm_cropped` datasets.

available, which allows for more precise and detailed conclusions.

Figure 4.6 applies to the case where we only consider images taken by the VAST's positioning camera as a target for segmentation. In Figure 4.6a it can be seen that all models trained using VAST images provide a similar accuracy, which is in line with our earlier observation that only a small amount of fish are necessary to provide a stable model. Of particular note is the fact that the filtered model trained on only BFM images also provides high accuracy and high tail solidity, and would thus be suitable to segment VAST images.

Figure 4.7 applies to the case where we only consider images taken by the Leica microscope camera as a target for segmentation. As can be seen in Figure 4.7a and Figure 4.7b models trained using VAST images provide a terrible accuracy and tail solidity for these BFM images, and would be unsuitable to segment them. Both variants of the model trained with BFM images have much better accuracy and tail solidity. The distribution for the accuracy for the filtered variant is much narrower and shorter, suggesting this model is more suitable

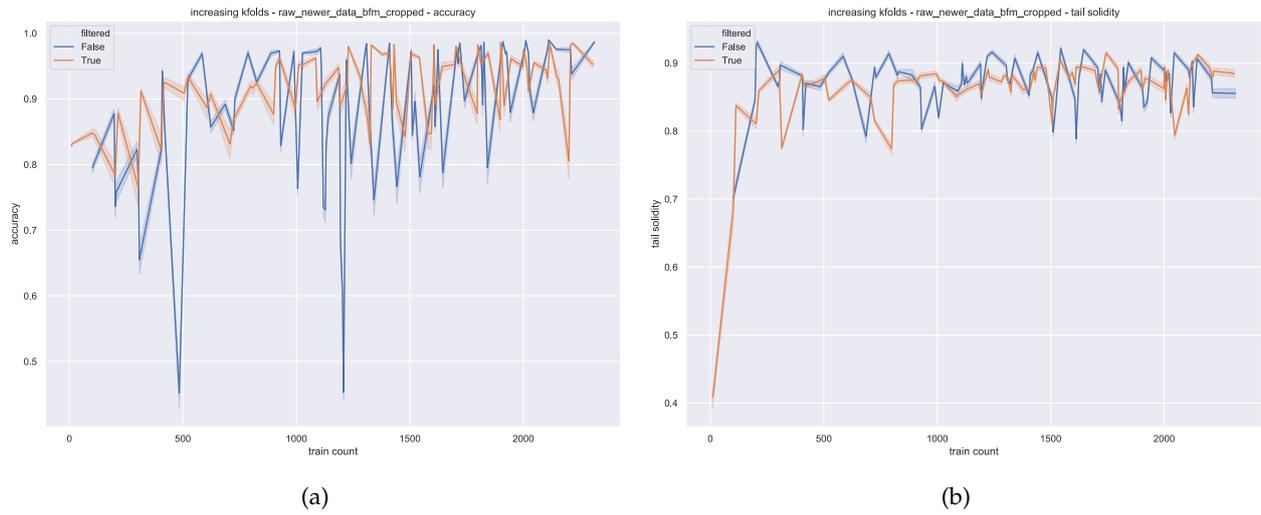


Figure 4.5: Results for evaluating a model using different metrics during the training process. This model is being trained with the `raw_newer_data_bfm_cropped` dataset.

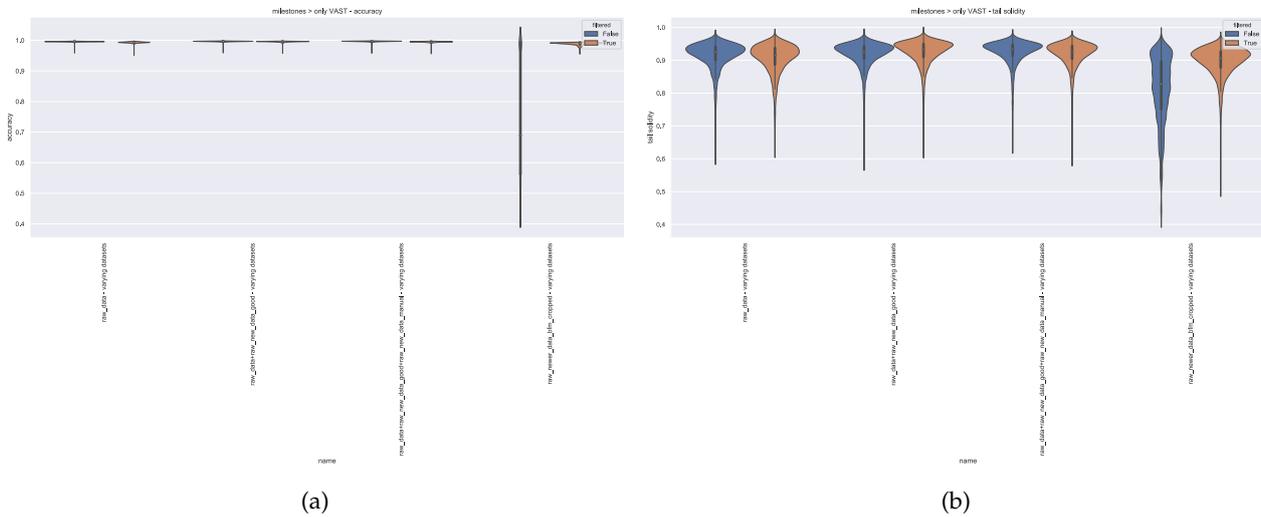


Figure 4.6: Overview of all evaluations on only VAST images for several different models.

to segment BFM images.

4.3 Determining the impact of the aspect ratio

Since the images taken by the Leica microscope camera not only have a different resolution than the VAST images, but also have a different aspect ratio, the impact of the asymmetrical scaling was evaluated. This would allow us to conclude whether a cropping step needs to be added to any preprocessing. To evaluate this impact, several different models were separately evaluated on BFM images as they were originally taken and BFM images which were cropped to the same aspect ratio as the VAST images. During testing, images of both datasets are scaled down to the input of the network, where only the BFM images not cropped in advance would be scaled asymmetrically. Both of these evaluations are associated with a large number of numbers for each metric, which are once again plotted as violin plots. The results are displayed in Figure 4.8.

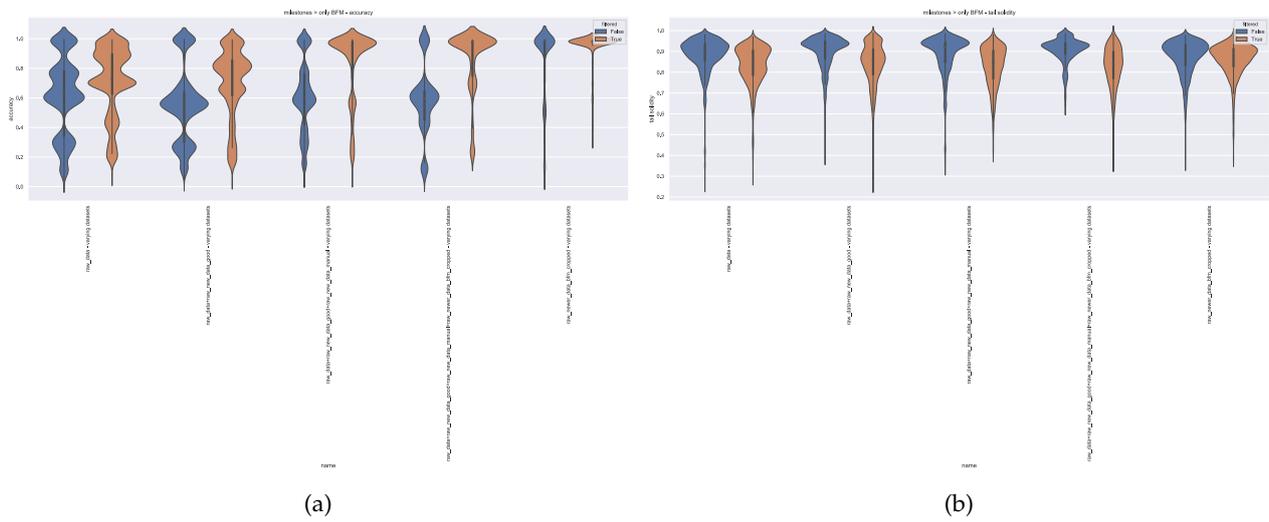


Figure 4.7: Overview of all evaluations on only BFM images for several different models.

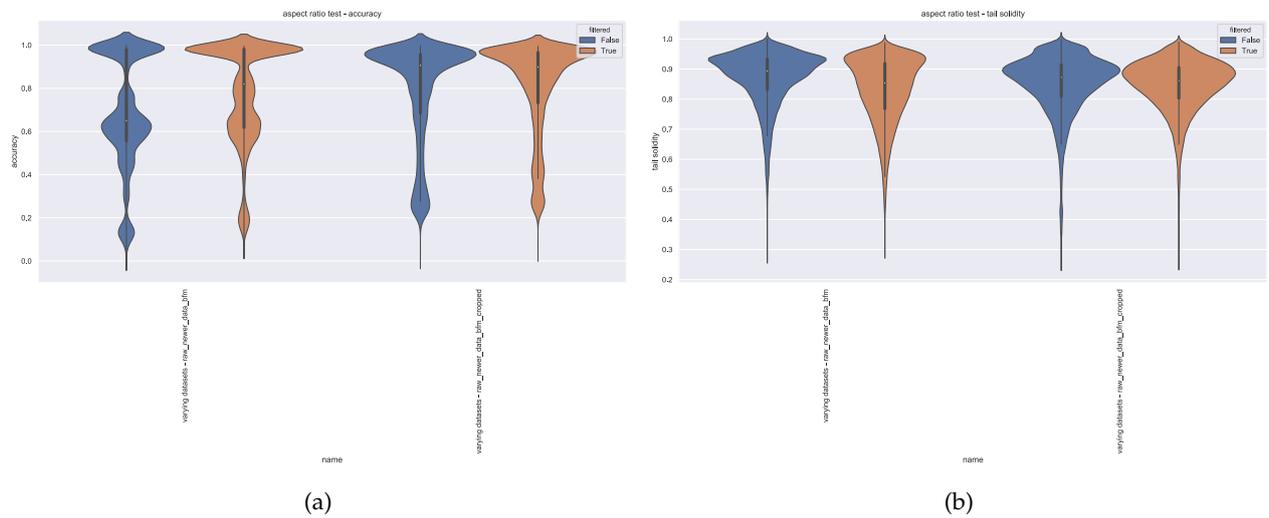


Figure 4.8: Results when using two datasets with different aspect ratios as testing data.

Using images with a different aspect ratio than the input of the network causes these images to be scaled to this new aspect ratio. This causes the shape of the fish to change. As can be seen in Figure 4.8a, using such images as testing data leads to a distribution where a lot of images are segmented with low accuracy. If these images are cropped to the correct aspect ratio before using them as testing data, the shape of the fish is maintained. This leads to a much nicer distribution with a higher median for the accuracy. In Figure 4.8b it can be seen that the distribution for the tail solidity also has a much flatter tail for cropped images and that more images are segmented with a high tail solidity. This leads us to believe that cropping images to the aspect ration of the network to maintain the shape of the fish is a necessary step to improve the quality of the segmentation.

4.4 Using bad data as testing samples

The second VAST data set had been divided into four parts based on the quality of the low-throughput segmentations.

1. Images which were segmented incorrectly by the low-throughput method. Such images were generally segmented like the image in Figure 4.10.
2. Images which contained a physically damaged or otherwise abnormal fish.
3. Images which were normal and segmented correctly by the low-throughput method.
4. Images which had a manually traced segmentation available.

Images which were segmented incorrectly were ignored in all other experiments, as their incorrect segmentations would be detrimental to any models trained with them, as well as meaninglessly lower the results of any evaluations in which they were included. However, it is still relevant to quantify the loss in metrics which would occur when using such images as testing data. Images which contained a physically damaged fish were also excluded from all other experiments, as their irregular nature and often incorrect segmentations could once again be detrimental to both evaluations were they are used as training data, as those where they are used as testing data. Furthermore, all images for a damaged fish hold little meaning altogether, as the damage to the fish would cause it to be discarded from the original experiment in its entirety, and its images would never make it to the segmentation step. It is however still interesting to measure the accuracy with which such images would have their segmentation predicted, as this gives us more information about the DNN-ZF method itself. For these reasons, several models were evaluated separately on the `raw_new_data_bad` and `raw_new_data_broken` datasets. The same models were also evaluated on the two remaining datasets created from the new VAST images combined, namely `raw_new_data_good+raw_new_data_manual`, for the sake of providing a control. Once again, each of these evaluations is associated with a large number of numbers for both metrics, which are plotted as violin plots. The results are displayed in Figure 4.9.

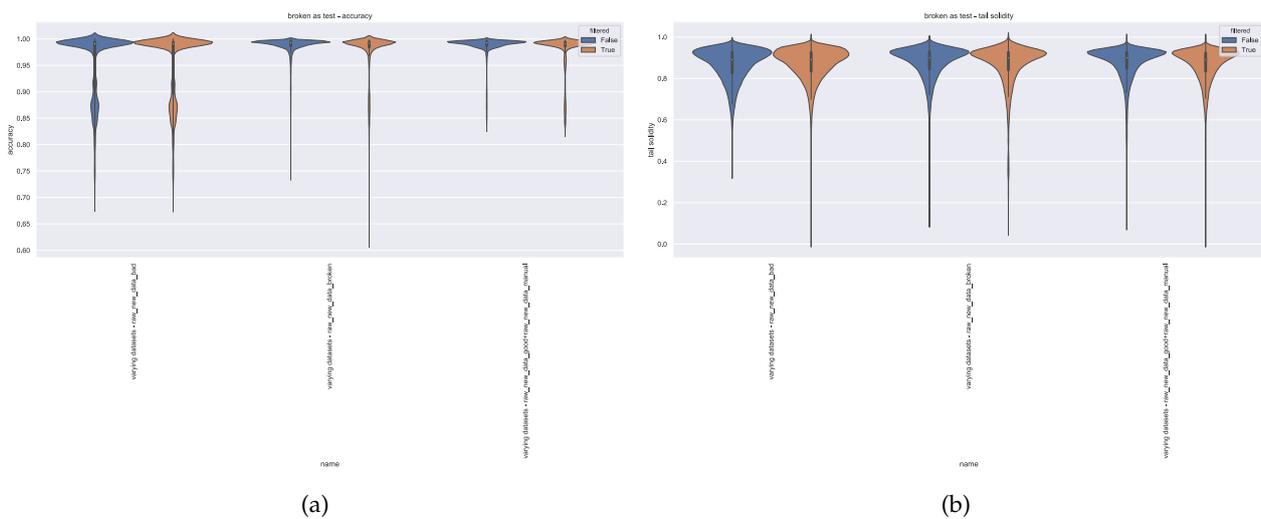


Figure 4.9: Comparison between good and bad samples as testing data of the same data set.

As can be seen in Figure 4.9, segmenting images with physically damaged fish leads to only a small decrease

in accuracy and a similar tail solidity. This indicates that our method would have no additional issues with segmenting such images.

Images for which the predicted segmentations are validated against incorrect 'truth' segmentations would be expected to give a much lower accuracy than normal. This would happen regardless of how correct the predicted segmentations are. This can indeed be observed in Figure 4.9a. As the tail solidity metric is calculated solely from the predicted output and does not use this incorrect 'truth' data, it is still indicative of the quality of the predictions. In Figure 4.9b it is observed this metric is lower and has a longer tail than the control, which indicates that this new method performs worse on images which were segmented incorrectly by the low-throughput method.



Figure 4.10: Example of an image with an incorrect segmentation. The purple line indicates where the image has been segmented.

4.5 Using bad data as training samples

Not only is it relevant to quantify the loss in metrics when incorrectly segmented images are used as testing data, but it is also relevant to quantify the loss of quality when a model is trained using such images. Similarly, it is also interesting to observe the effect the use of images containing damaged fish as training data has on the quality of a model. Not only will this give us more insight into the DNN-ZF method itself, but it will also allow us to determine if such images would be acceptable to use as training data, if the alternatives are not of sufficient quality or quantity. In order to accomplish this, a model trained on the `raw_new_data_bad` dataset and a model trained on the `raw_new_data_broken` dataset were evaluated on all other segmented images. Another model trained on the remaining datasets created from the new VAST images combined, namely `raw_new_data_good+raw_new_data_manual` was also evaluated to provide a control. Once again, each of these evaluations is associated with a large number of numbers for both metrics, which are plotted as violin plots. The results are displayed in Figure 4.11.

As can be seen in Figure 4.11, the model trained using images of physically damaged fish provides a similar accuracy and tail solidity for segmentation as the model trained using normal images. This indicates that images with physically damaged fish are still suitable for training a good model.

When a model that has been trained on images with incorrect training segmentations is evaluated on images with good 'truth' segmentations, the expectation is that the predicted segmentations are also incorrect and will

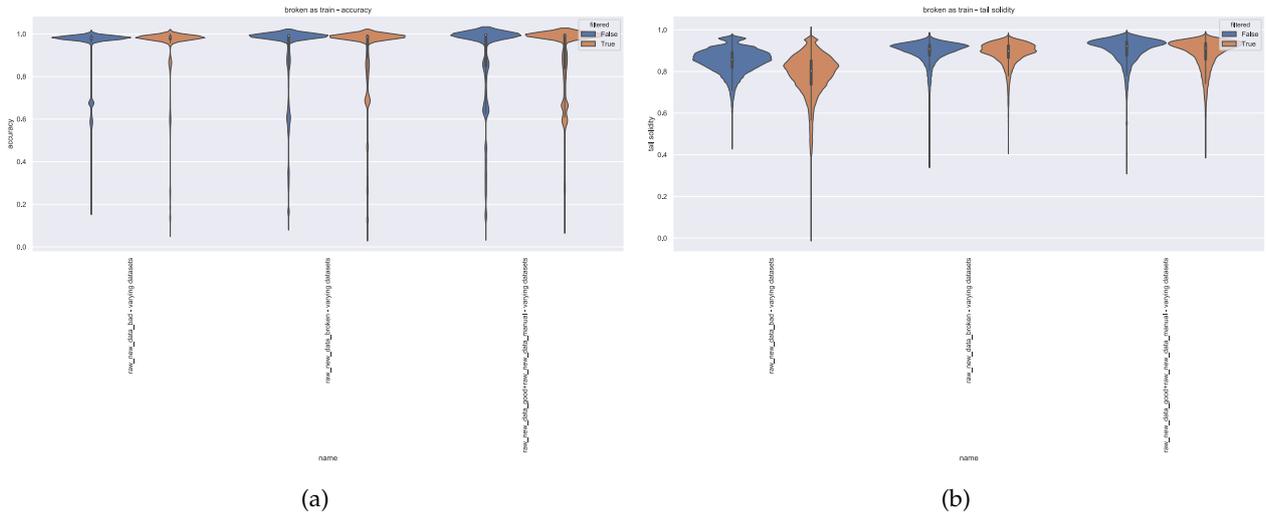


Figure 4.11: Comparison between good and bad samples as training data of the same data set.

cause a significant drop in accuracy. But as can be seen in Figure 4.11a, the accuracy remains similar to that of a normal model evaluated on the same data. However, in Figure 4.11b it is made clear that the tail solidity for the predictions of this model are much lower than those for the normal model, indicating that many of these segmentations are incorrect none the less.

4.6 Explaining the functionality of the DNN·ZF method

The inner workings of any deep neural network are always vast and complex. In the case of the architecture evaluated in this paper, a total of 4,060,705 variables are involved. Additionally, the connections between these variables need to be considered as well. It should be easy to understand that by simply looking at the structure and weights of such a large neural network, its processes can not be easily understood. Several methods already exist for the visualization and understanding of large neural networks [ZF14] [SVZ14]. However, as these methods interact with the content of the architecture, they are deemed unsuitable for the presented research.

Even so, an attempt was still made at understanding the method by which this architecture performs segmentations. This was done by providing some models with simplified or modified inputs and observing the properties of the produced segmentations. To improve the visualization of the predicted segmentations, the postprocessing step which keeps and fills only the largest segmented part has been disabled for these experiments.

The first experiment was performed by considering the model trained on only the dataset containing old VAST images. A series of images which start with a rough approximation of a zebrafish larva and gradually become more refined were provided as input to this model, and the predicted segmentations are visualized in Figure 4.12. Additionally, this series was created on a background both with and without any visible capillaries in an attempt to discover the reaction of the DNN·ZF. As can be seen in Figure 4.12, the images containing

only the background have not been segmented at all. In the images containing a box and a thin outline, it can be observed that the DNN-ZF method aggressively tries to find the larva at both sides of the image, but does not care that much about the thin lines at the top and bottom which have the same thickness as the line segment on the right. Only once the thickness of these lines have been doubled, they are included in the segmentation. Additionally, the segmentation extends well beyond the line segment on the right of the images, but not beyond the lines at the top and bottom, or the box on the left. Once we start to more closely emulate the anatomy of the zebrafish larva with only the two darkest areas (eye and yolk sac) as well as an outline, these results stay mostly the same. Both of the dark areas are selected as part of the segmentation. Additionally, the outline at the right of the image, as well as an area extended inwards is included. The top and bottom are only fully included when the thickness of the line has been increased. Once a gradient is added to the inside of the outline in a fashion similar to that of the actual zebrafish larva, the entire outlined area is included in the segmentation, regardless of the thickness of the line. The final observation we can make is that when the series of images without the capillaries is considered relative to the series with capillaries, all of the segmentations have become more 'narrow'. As in, they extend less far upwards and downwards.

The same model was also provided with a collection of images modified from a single base image. In addition to gaining insight into the functionality of the DNN-ZF, this will also help determine features which could prevent the model in question from working with future datasets. The predicted segmentations for these images are visualized in Figure 4.13. As can be seen in the figure, the original image, which is displayed in the top left, is segmented as desired. In this particular case, the especially troublesome tail is also properly segmented. Once several black bars are added, they are simply included in the segmentation. This does however cause a hole to form just before the tip of the tail. Blurring the image does not affect the segmentation much, with the exception of the appearance of a notch in the tail. Decreasing the brightness will cause the segmentation to include much more than just the zebrafish larva. In addition to including some of the now much darker background noise, parts of the capillaries are also included. Interesting is that some of the included parts of the background noise appear green, and have low values on the blue and red colour channels. If the capillaries are manually cleaned from the image, the same effect described in the previous paragraph is observed. That is, the segmentation extends less far up and down. These missing sections also cause a hole to form in the tail. Raising the contrast of the image causes parts of it to become much brighter than before, which are then excluded from the segmentation. Inverting the image does not cause the segmentation to invert. Instead, only the previously darkest areas are excluded from this new segmentation. If the zebrafish larva is deformed in regards to its shape then the segmentation properly fits this new shape, save for the areas which are blurred in addition to deformed. Filling in the tip of the tail of the zebrafish larva causes the slight dimple in its segmentation to disappear, but does not affect the segmentation otherwise. Finally, if the darkest areas of the zebrafish larva are removed, the rest is still properly segmented. The only exception is the few lines which have become extremely thin as a result, which matches the observation regarding the thickness of lines made in the previous paragraph.

The same collections of images have also been segmented using a model trained with much more training images. In addition to the old VAST images, this also includes some of the new VAST images and the newer

BFM images. The BFM images especially have a much different contrast and brightness when compared to the VAST images. The predicted segmentations for the series of approximations for this new model are visualized in Figure 4.14, and are mostly the same as those for the previous model. Some differences are still present however. Regarding the rectangular approximations, the segmentations have changed to no longer include an area inwards from the right line segment. The top lines are also no longer included. Instead, a part of the bottom line is included in the segmentation, where the thinner lines include a greater length. Regarding the zebrafish larva shaped approximation, the only change is that in the variant without gradient, a much larger section of the tail is included in the segmentation.

Naturally, this model has also been used to segment the collection of modified images, visualized in Figure 4.15. Once again, the majority of the segmentations have not undergone any notable changes between the models. The segmentations for the images where the brightness and contrast have been modified however have improved quite a bit. The segmentation of the tail area has also been performed better overall.

The final model on which this experiment has been performed is a model only trained with the newer BFM images. Additionally, the grey erosion preprocessing filter had been enabled. When the segmentations for the series of approximations are predicted using this setup, this leads to the visualization in Figure 4.16. The segmentations predicted by this model are very different from the previous two models. In all of the rectangular approximations, the entirety of the solid box is still included in the segmentation. This time however, in the variant with thin lines only the bottom right corner is included, whereas in the variant with thicker lines both the bottom and right lines are included entirely. Once the shape of the zebrafish larva is approximated with an outline and two dark areas, the opposite happens. In addition to including both dark areas in the segmentation, like with the previous models, the top of the outline is included as well. This is unlike the results for the other two models, and also unlike the other results for this model, where the segmentations tended to include the bottom of the outline. Furthermore, the segmentation of the tail no longer extends inwards. When a gradient is added to the inside of the outline, the segmentations made by the other two models included the zebrafish larva in its entirety. The segmentation made by this model, however, misses a significant portion of the bottom of the tail.

Once again, this same model has been used to predict segmentations for the collection of modified images, leading to the visualization in Figure 4.17. Unlike with the previous two models, the segmentation predicted by this model for the original image is slightly off from what it should be, but correctly segments the tip of the tail. When black bars are added, the same effect as with the other two models is observed, where the bars are included but the tip of the tail is separated from the rest of the segmentation. Where blurring the image provided a detrimental effect to the segmentations predicted by the other models, the prediction made by this model is unaffected. Decreasing the brightness provides the same effect on the segmentation as with the first model, where parts of the background and the capillaries are now included. If the capillaries are removed, the segmentation extends less far up and down, and the quality of the segmentation of the tail decreases slightly. This is the same for the other two models. Increasing the contrast, just as with the other two models, causes parts of the image to become much brighter than before, which are subsequently excluded from the segmentation. Whereas segmenting the inverted image with the previous two models only caused the

now brightest two areas of the zebrafish larva to be excluded from the new segmentation, the segmentation predicted by this model excludes almost the entire fish. Only the now darkest part of the tail is still included. When the shape of the zebrafish larva is deformed the segmentation follows these deformations pretty well with only some issues caused by the additional blurring, just as with the previous models. As the very tip of the tail is already properly segmented in the original image, filling it in provides no significant change. Of particular note however is that the tiny hole present in the original segmentation is still present after the tail is filled in, despite its proximity to the now pitch-black area. Finally, the removal of the two darkest areas of the zebrafish larva causes those areas to be excluded from the segmentation, along with any parts which have turned into thin lines as a result, just like with the previous two models. Unlike the segmentation predicted by those models however, the segmentation predicted by this model is also affected in other ways. More specifically, the dip in the segmentation above the yolk sac has disappeared, and the segmentation also includes a part of the capillary extending beyond the head. Overall, the segmentation directly above the missing parts has been increased in height, whilst the segmentation of the tail is entirely unaffected.

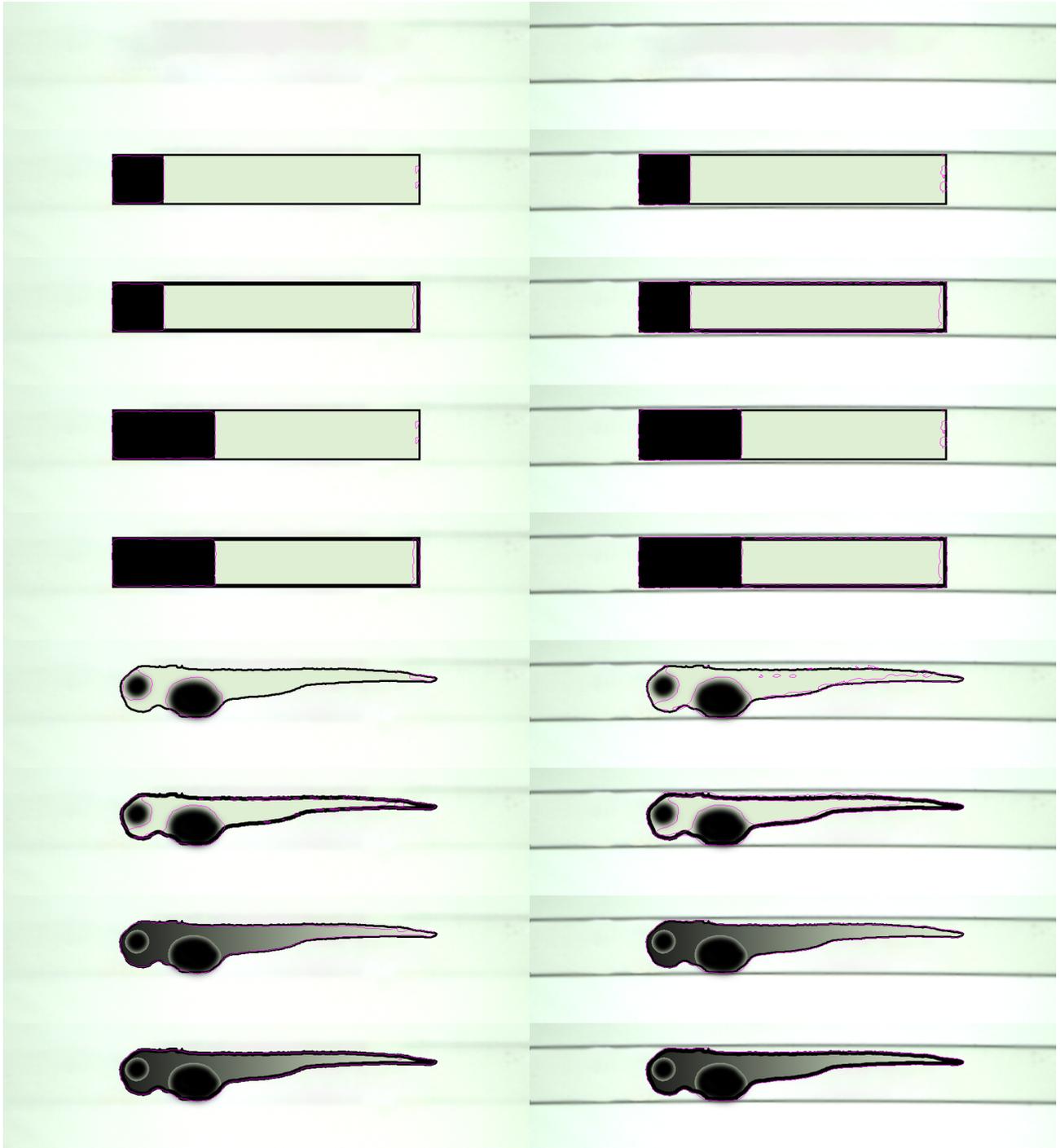


Figure 4.12: Results of segmentations performed by a model trained on the raw_data dataset. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several shapes approximating that of a zebrafish larva.

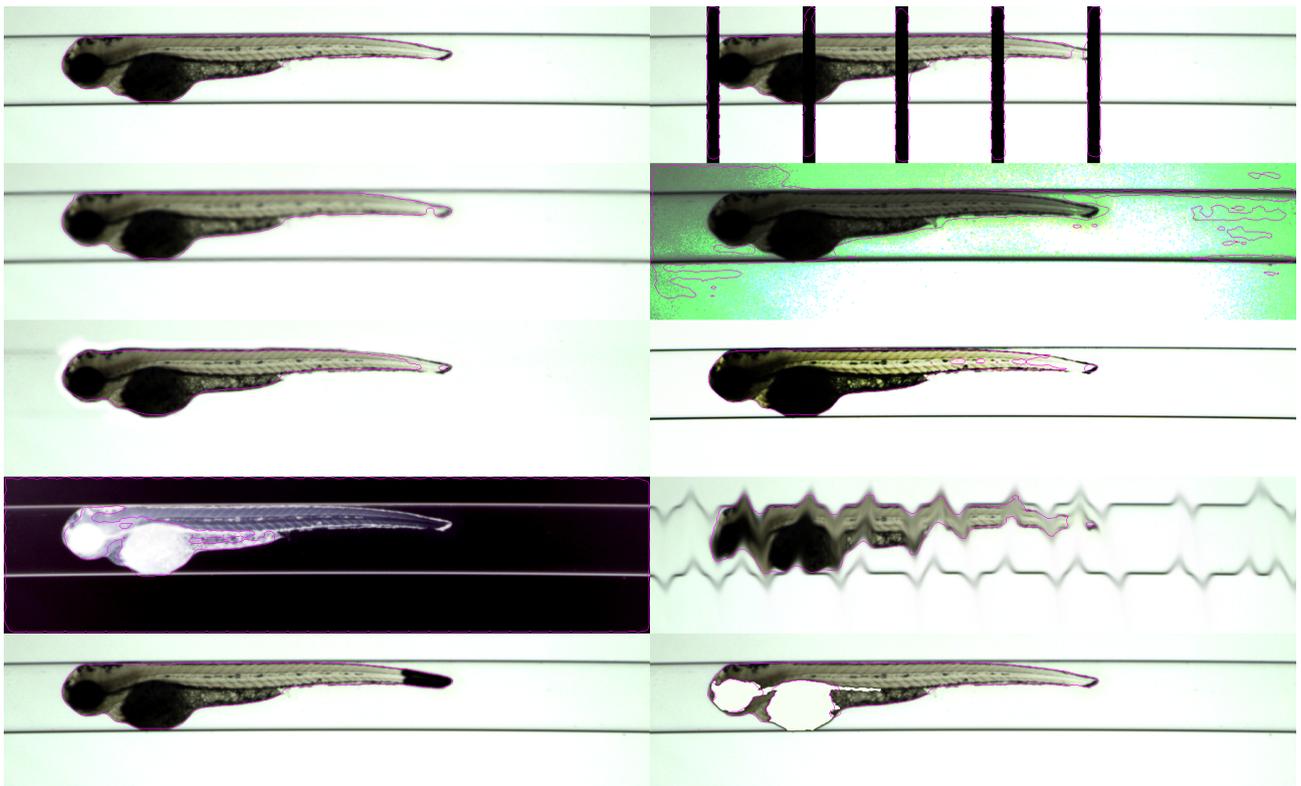


Figure 4.13: Results of segmentations performed by a model trained on the raw.data dataset. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several modified images of a zebrafish larva. The image in the top left is the original image.

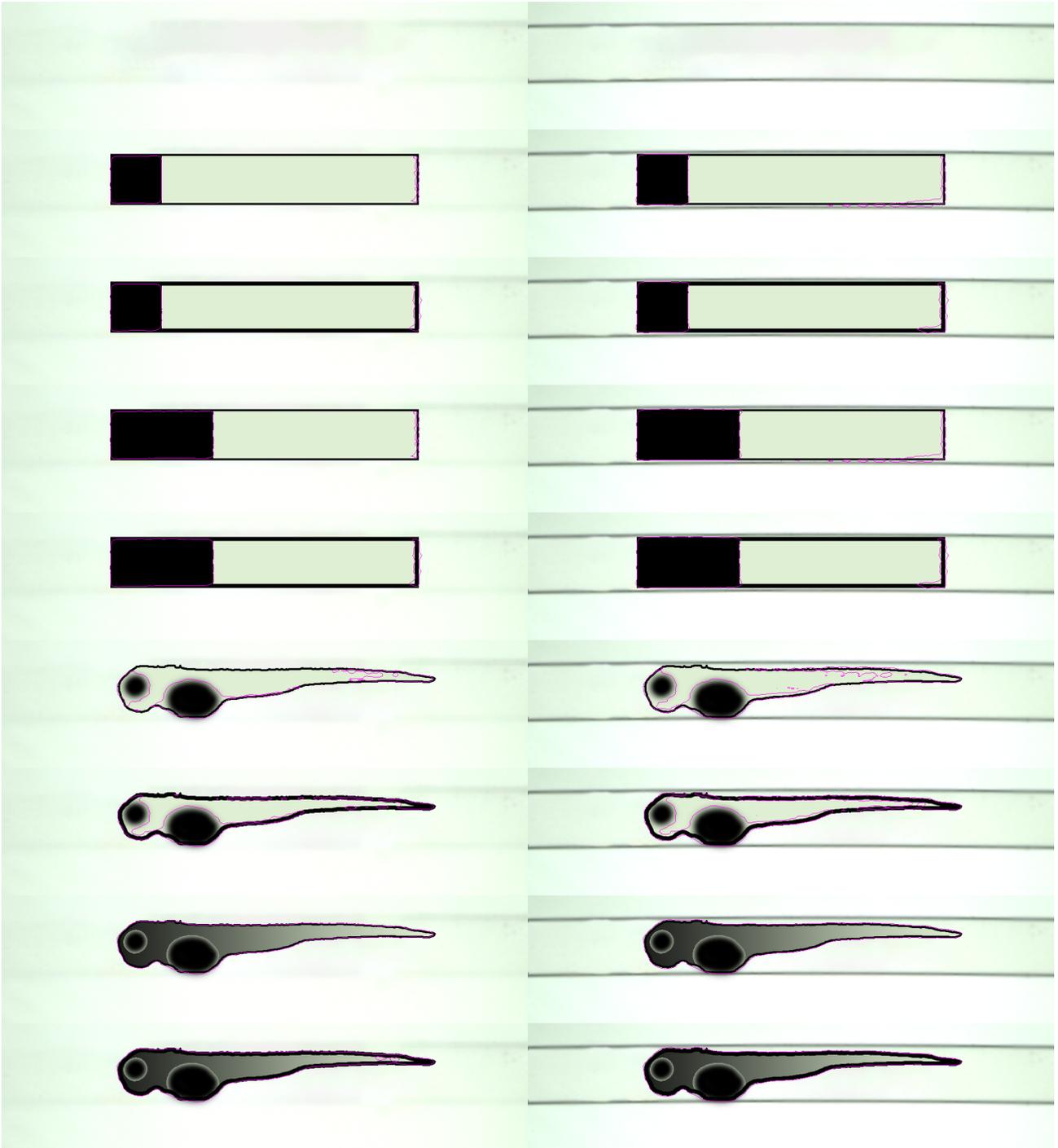


Figure 4.14: Results of segmentations performed by a model trained on the `raw_data`, `raw_new_data_good`, `raw_new_data_manual`, and `raw_newer_data_bfm_cropped` datasets. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several shapes approximating that of a zebrafish larva.

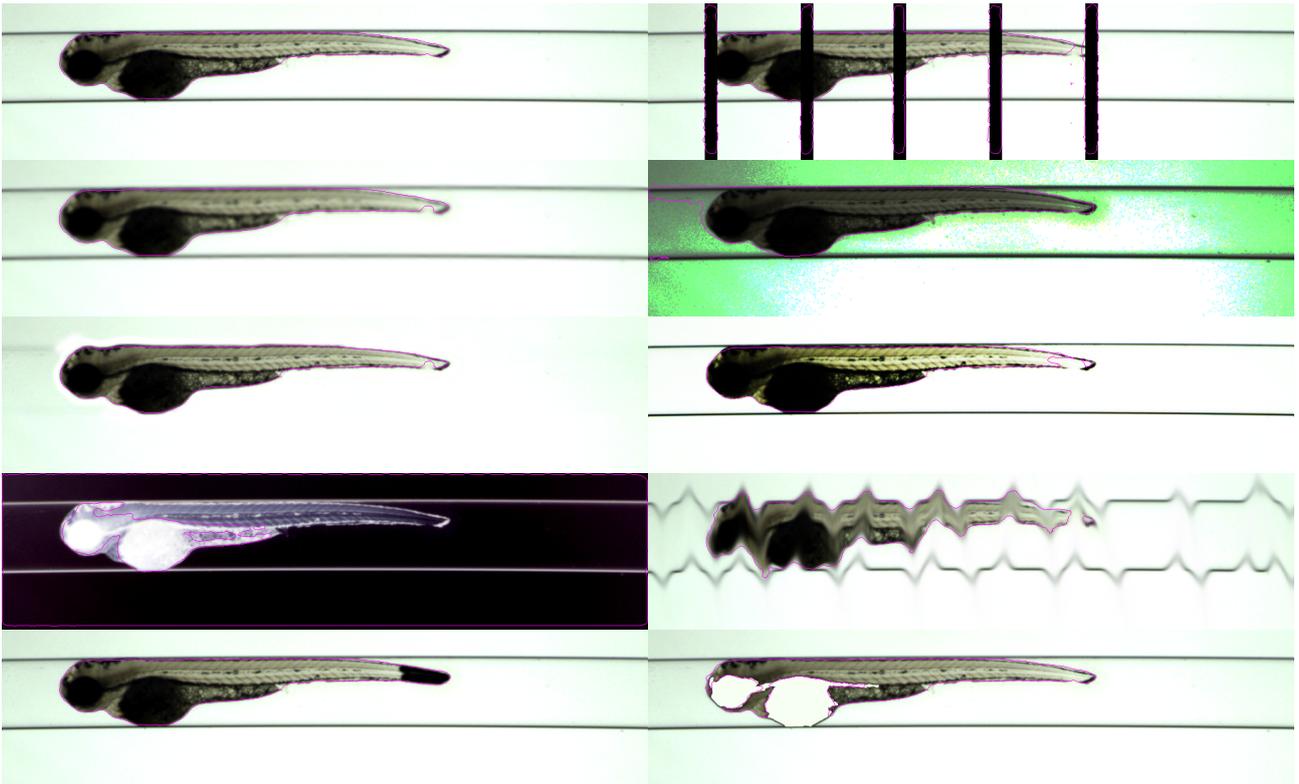


Figure 4.15: Results of segmentations performed by a model trained on the `raw_data`, `raw_new_data_good`, `raw_new_data_manual`, and `raw_newer_data_bfm_cropped` datasets. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several modified images of a zebrafish larva. The image in the top left is the original image.

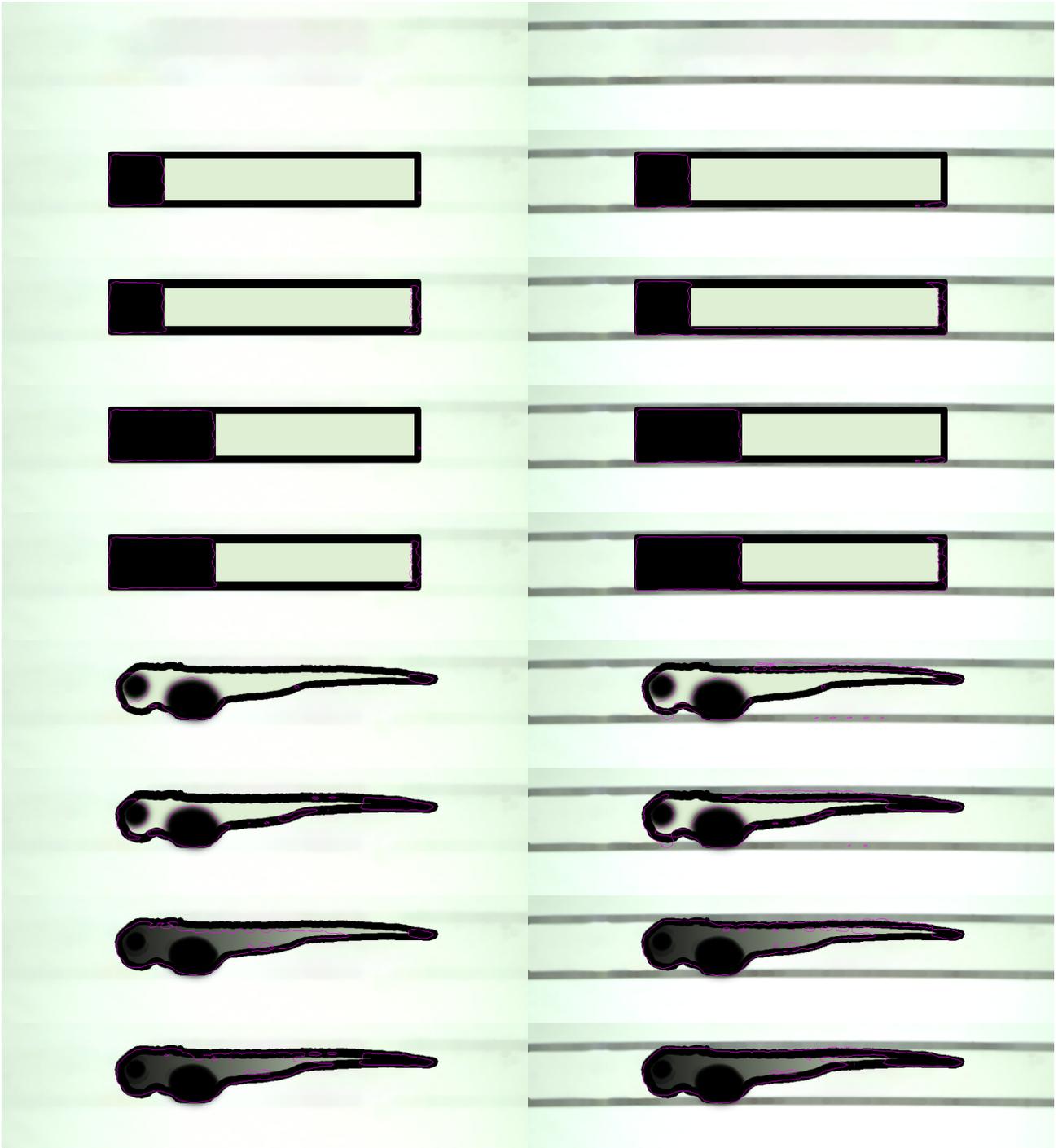


Figure 4.16: Results of segmentations performed by a model trained on the `raw_newer_data_bfm_cropped` dataset, with the gray erosion filter applied. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several shapes approximating that of a zebrafish larva.

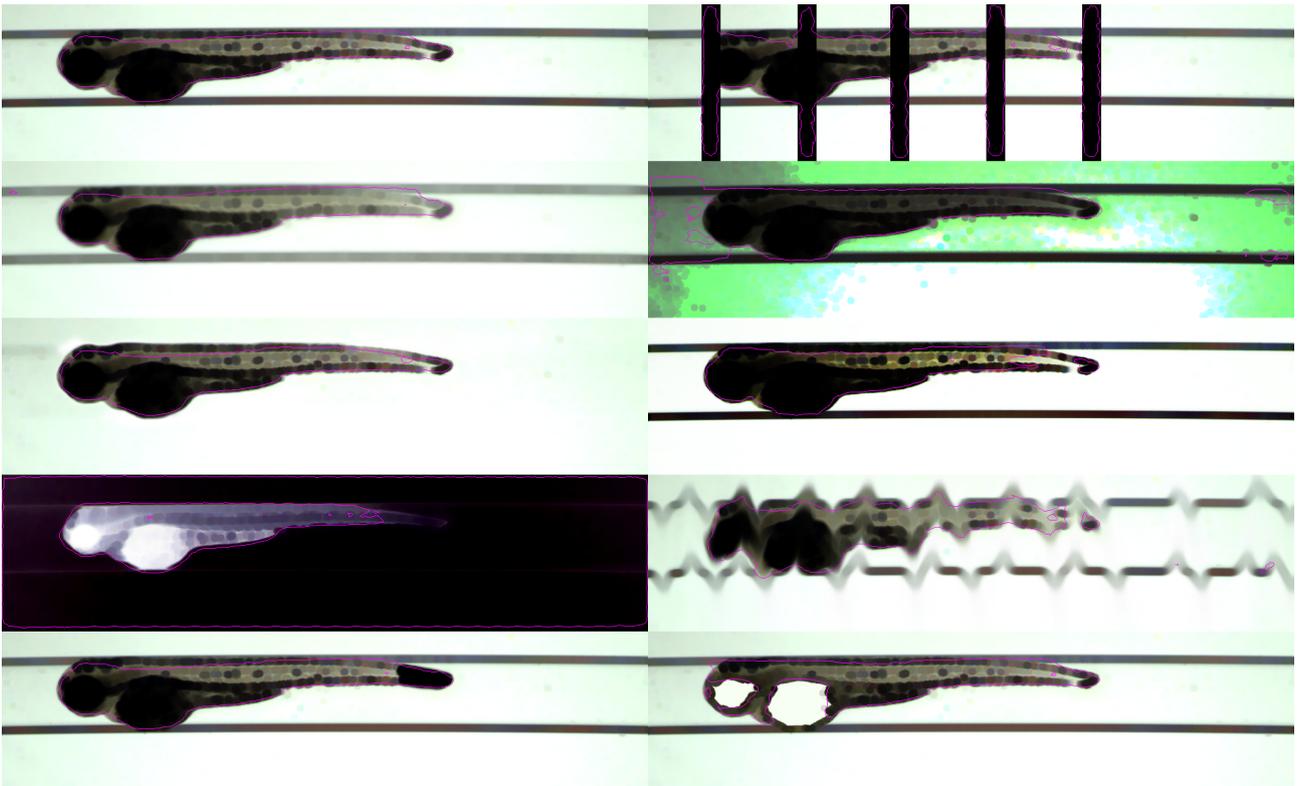


Figure 4.17: Results of segmentations performed by a model trained on the `raw_newer_data_bfm_cropped` dataset, with the gray erosion filter applied. The purple lines indicate the boundaries of the segmentations. The segmentations have been performed on several modified images of a zebrafish larva. The image in the top left is the original image.

Chapter 5

Conclusions & Discussions

Using the results presented in Chapter 4, we can answer the research questions asked in Chapter 1. Each research question will be answered and discussed separately. Recommendations for future work are also made. This chapter concludes with a concise summary of the most important results and recommendations.

5.1 How can the functionality of the DNN-ZF be explained?

Combining the information obtained from the various collections of testing images, we can reason about the principles that the DNN-ZF bases its segmentations on. A consistent action across all images was that large 'dark' areas of the image are always included in the segmentation. Areas which are mostly 'dark' on the green channel count as well. After this, the somewhat darker outline of the zebrafish larva is included as well, with a specific focus on the bottom corners. From there, a part of the area enclosed by the outline is included, which will help with the segmentation of the transparent tail.

It has also been discovered that the shape of the zebrafish larva is *not* crucial to a successful segmentation.

By comparing the segmentations of images with capillaries to those with the capillaries removed, we can also reason how the DNN-ZF deals with these capillaries. A notable difference when considering the images without capillaries is that their segmentations not only extend less far up and down, but some of the lighter areas which would otherwise be enclosed are excluded as well. This indicates that the segmentation is not simply trimmed down after the fact. Instead, the segmenting itself outright ignores certain parts of the image. This would cause the outer edge of the content to be ignored when the capillaries are missing, which in turn causes certain areas to no longer be enclosed, leading to the observed changes.

An understanding of the functionality of the DNN-ZF method can lead to solutions to a problem which is still partially present, which is that of a notch missing from the segmentation of the tail. Given our current understanding, priority should be given to enclosing this section of the tail with some sort of outline. Alternatively, the very tip of the tail which is generally segmented and thus located correctly can be

coloured black. Applying a filter that accomplishes either of these properties could improve the quality of the segmentations. Further research could determine if this is indeed the case.

5.2 How well does the existing hybrid segmentation method perform on new datasets?

As his research has shown, the hybrid segmentation method developed by Y. Guo provides exemplary results when used to segment the dataset it was developed for (old VAST images). However, it is completely unsuitable for the segmentation of any other datasets, as discussed in Chapter 3.

As such, the hybrid segmentation method should be abandoned in favour of the DNN·ZF method. However, there is still a need for the creation of low throughput training segmentations. Given that a relatively small number of training images is already sufficient, our advise is that the hybrid segmentation method is applied to any images it happens to work on and that the remainder is segmented by manual tracing.

5.3 How does the quality of a model change as it is being trained on more data?

Initially, after more images are used to train a network, its accuracy increases. After about 3000 images, however, the maximum accuracy of the model will be reached, and further training will not have much of an effect. This amount of images is equivalent to about 30 fish to be used for training. Because such a small amount of fish is sufficient to reach the maximum accuracy of the network, it is possible to manually select the 30 most desirable fish from all available data sets to train a more desirable network. It is also very doable to create manual segmentations for all 3000 images to ensure the resulting model produces the most ideal predictions. Future research could use the data produced by this research to determine which samples caused the largest increase in metrics and are thus the most desirable.

5.4 Does the aspect ratio of an image affect the quality of its segmentation?

When images with a different resolution than the input of the network are used they are resized to match this input. In the case that the aspect ratio differs between the two, changes will occur in the zebrafish larva during the resizing. As has been concluded from Figure 4.8, this negatively impacts both the accuracy and quality of the predicted segmentations and should be avoided. This can be achieved by simply cropping each image to the aspect ratio of the network before using them.

When an image is rescaled, the shape of the zebrafish larva in it changes. As it has already been concluded that the shape of the zebrafish larva does not affect the quality of the predicted segmentation, the observed drop in quality would be the result of other changes in the image. For example, the change in the shape in the capillaries, or a loss of quality in an unexpected way. The exact cause of the loss in quality could be determined in future research, although we reason that whatever the conclusion might be, it will not provide useful insight into improving the segmentations.

5.5 How does the DNN-ZF method perform on 'bad' input data?

As has been outlined previously, there are two kinds of 'bad' samples:

1. Images which were segmented incorrectly by the low-throughput method. Such images will have a 'truth' segmentation which does not match the desired result.
2. Images which contained a physically damaged or otherwise abnormal fish. The majority of these images have been properly segmented, but the fish themselves do not hold much meaning.

As has been concluded from Figure 4.9, this method has no additional issues with segmenting images of damaged fish. Since the analysis of the functionality of the network has shown that the shape of the fish does not matter, this comes at no surprise. However, as damaged fish do not contribute accurate data to the original research, they should be discarded before they ever reach the segmentation step. The method with which images will be discarded as well as its implementation into the pipeline is a matter for future research.

Images which happened to be incorrectly segmented by the low-throughput method also seem to perform worse than average with this new method, as has been concluded from Figure 4.9b. The accuracy metric for these images is uninformative as any prediction which would give a high accuracy is an incorrect prediction, and vice versa. As the tail solidity metric does not depend on these incorrect segmentations, it is still valid and shows a decrease in performance. This indicates that the properties of these images that made the low-throughput method fail to correctly segment them (high tail transparency) still affect this new method, albeit to a much smaller degree. This indicates that the DNN-ZF method in its current state still has one significant drawback, which needs to be solved in the future. Section 5.1 has outlined several possible solutions.

5.6 Can 'bad' samples still be used as training data?

From Figure 4.11 it has been concluded that images of damaged fish are still suitable as training data. However, as long as the trained model will not be used to segment images of damaged fish, there is no need to use such images for training. These images should only be used if no other training data is available, but given that the images of 30 fish are already sufficient, this will never be the case.

It should be clear that using images with incorrect 'truth' segmentations as training data will lead to a decrease in the quality of the model, as this will cause the model to be explicitly trained to give incorrect results. From

Figure 4.11 it has been concluded that this is indeed the case. As such, incorrectly segmented images need to be manually removed from any training data, or be manually re-traced. Future research could lead to better methods for the creation of training segmentations. For example, the tail solidity metric could be used to identify and discard incorrect segmentations.

5.7 Can a single model make accurate predictions for both VAST and BFM images?

Out of the currently available models, the model trained only on the cropped BFM images and with the grey erosion filtered is the *only* model which provides acceptable results when segmenting the BFM images. Other models are unsuitable for this task. This model provides decent results when used to segment VAST images, but is outdone by the other available models. More specifically, for the segmentation of VAST images the model trained on the `raw_data` and `raw_new_data_good` datasets with the gray erosion filter enabled seems to perform best.

At the moment, it is advisable to use a separate model for the segmentation of VAST and BFM images. The aforementioned model most suitable for the segmentation of BFM images, however, comes close to being able to handle both cases. Further research into this model, along with the availability of more BFM training data, could lead to the creation of a single general-purpose model in the future.

5.8 What effect does adding a filter to the images have?

The effect varies wildly per model. Some combinations of datasets for training benefit from the use of the grey erosion filter, whereas others do not. There is however no clear correlation between the data used for training and the usefulness of the grey erosion filter. Whether the filter should be used or not must be separately evaluated for each set of training data.

Future research can identify filters which do provide a positive effect at all times. In addition to the candidates mentioned in Section 5.1, mean shift and texture-based filters might prove useful.

5.9 How can the performance of the DNN-ZF for the segmentation of BFM images be improved?

When comparing the evaluation of the model most suitable for the segmenting of VAST images with the evaluation of the model most suitable for segmenting BFM images, we can see that the segmentation of BFM images is significantly less accurate. In its current state, the DNN-ZF method works extremely well for the

segmentation of VAST images, but performs only adequately when segmenting BFM images. Part of the issue is that the number of training images does not reach the recommended 3000 minimum. This could be solved by using data augmentation or manual tracing which would increase both the number and variety of training images.

Another major problem is that the structure of the architecture used in the DNN-ZF had been designed around the use of images with a resolution of 1024px by 250px, the resolution of the VAST images. This means that BFM images have to be scaled down before they can be processed by the DNN-ZF method, losing much detail in the process. By redesigning the network such that the input coincides with the resolution of the BFM images this detail can be kept, which could lead to significant improvements in the quality of the segmentations. More research is required into this matter however, as an entirely new architecture has to be designed and evaluated.

It should also be noted that in the future the aim is to take BFM images at a higher magnification. As a higher magnification means that less of the specimen is visible, multiple images would need to be taken and stitched together. This process might introduce several artefacts into the final image. Additionally, each of the stitched segments would have the same resolution as the current BFM images, which means that the final image is of a much higher resolution. Research needs to be done to determine whether these artefacts introduce additional problems. More importantly, as the input of a neural network has a fixed size, an ideal input resolution needs to be determined which can handle VAST images, BFM images, and stitched BFM images.

5.10 How can an assessment of methods and models for segmentation of zebrafish larvae contribute to an optimal processing pipeline?

Finally, we will provide a concise answer to the primary research question introduced at the beginning of the paper. In addition to presenting the two most suitable models for segmentation, we will provide a list of recommendations regarding the implementation and further improvements of these models.

First of all, we will once again focus on the only question in the flowchart from Figure 3.1. At the moment, the answer is that a separate model is needed for both image sources. The most ideal models are as follows:

VAST images The model trained on the `raw_newer_data_bfm_cropped` dataset *with* the gray erosion filter.

BFM images The model trained on the `raw_data+raw_new_data_good` datasets *with* the gray erosion filter.

Secondly, we will present recommendations regarding the implementation of this model into the processing pipeline:

1. Any images presented to the segmentation step must be cropped to the aspect ratio of the network input (4.096:1).

Thirdly, we will provide a list of crucial optimizations which should be made to the processing pipeline and the DNN-ZF method.

1. Physically damaged zebrafish larvae need to be detected and discarded early in the pipeline.
2. The architecture comprising the DNN-ZF method needs to be updated such that it supports images of a much higher resolution. When doing so, consideration needs to be made for any future increases in resolution as increasingly magnified images are stitched together.
3. New models or even a single new model need to be trained, using only manually traced images from 30 randomly selected undamaged specimen.
4. Other image filters need to be evaluated, possibly providing features similar to those discussed in Section 5.1.

During the creation of these modifications it is important to be aware of how the segmentations will be used. As the next step in the processing pipeline is the generation of 3D models from these segmentations, they need to exhibit properties suitable for this process.

Bibliography

- [GBHP₁₃] M. Gemberling, T. J. Bailey, D. R. Hyde, and K. D. Poss. The zebrafish as a model for complex tissue regeneration. *Trends Genet.*, 29(11):611–620, Nov 2013.
- [GP02] J. Rafferty G. Parsons. Tag image file format (tiff) - image/tiff mime sub-type registration. RFC 3302, RFC Editor, September 2002.
- [Guo17] Yuanhao Guo. *Shape Analysis for Phenotype Characterisation from High-throughput Imaging*. PhD thesis, LIACS, 2017.
- [HCT⁺₁₃] Kerstin Howe, Matthew D. Clark, Carlos F. Torroja, James Torrance, Camille Berthelot, Matthieu Muffato, John E. Collins, Sean Humphray, Karen McLaren, Lucy Matthews, Stuart McLaren, Ian Sealy, Mario Caccamo, Carol Churcher, Carol Scott, Jeffrey C. Barrett, Romke Koch, Gerd-Jörg Rauch, Simon White, William Chow, Britt Kilian, Leonor T. Quintais, José A. Guerra-Assunção, Yi Zhou, Yong Gu, Jennifer Yen, Jan-Hinnerk Vogel, Tina Eyre, Seth Redmond, Ruby Banerjee, Jianxiang Chi, Beiyuan Fu, Elizabeth Langley, Sean F. Maguire, Gavin K. Laird, David Lloyd, Emma Kenyon, Sarah Donaldson, Harminder Sehra, Jeff Almeida-King, Jane Loveland, Stephen Trevanion, Matt Jones, Mike Quail, Dave Willey, Adrienne Hunt, John Burton, Sarah Sims, Kirsten McLay, Bob Plumb, Joy Davis, Chris Clee, Karen Oliver, Richard Clark, Clare Riddle, David Elliott, Glen Threadgold, Glenn Harden, Darren Ware, Sharmin Begum, Beverley Mortimore, Giselle Kerry, Paul Heath, Benjamin Phillimore, Alan Tracey, Nicole Corby, Matthew Dunn, Christopher Johnson, Jonathan Wood, Susan Clark, Sarah Pelan, Guy Griffiths, Michelle Smith, Rebecca Glithero, Philip Howden, Nicholas Barker, Christine Lloyd, Christopher Stevens, Joanna Harley, Karen Holt, Georgios Panagiotidis, Jamieson Lovell, Helen Beasley, Carl Henderson, Daria Gordon, Katherine Auger, Deborah Wright, Joanna Collins, Claire Raisen, Lauren Dyer, Kenric Leung, Lauren Robertson, Kirsty Ambridge, Daniel Leongamornlert, Sarah McGuire, Ruth Gilderthorp, Coline Griffiths, Deepa Manthravadi, Sarah Nichol, Gary Barker, Siobhan Whitehead, Michael Kay, Jacqueline Brown, Clare Murnane, Emma Gray, Matthew Humphries, Neil Sycamore, Darren Barker, David Saunders, Justene Wallis, Anne Babbage, Sian Hammond, Maryam Mashreghi-Mohammadi, Lucy Barr, Sancha Martin, Paul Wray, Andrew Ellington, Nicholas Matthews, Matthew Ellwood, Rebecca Woodmansey, Graham Clark, James D. Cooper, Anthony Tromans, Darren Grafham, Carl Suke, Richard Pandian, Robert Andrews, Elliot Harrison, Andrew Kimberley, Jane Garnett, Nigel

Fosker, Rebekah Hall, Patrick Garner, Daniel Kelly, Christine Bird, Sophie Palmer, Ines Gehring, Andrea Berger, Christopher M. Dooley, Zübeyde Ersan-Ürün, Cigdem Eser, Horst Geiger, Maria Geisler, Lena Karotki, Anette Kirn, Judith Konantz, Martina Konantz, Martina Oberländer, Silke Rudolph-Geiger, Mathias Teucke, Christa Lanz, Günter Raddatz, Kazutoyo Osoegawa, Baoli Zhu, Amanda Rapp, Sara Widaa, Cordelia Langford, Fengtang Yang, Stephan C. Schuster, Nigel P. Carter, Jennifer Harrow, Zemin Ning, Javier Herrero, Steve M. J. Searle, Anton Enright, Robert Geisler, Ronald H. A. Plasterk, Charles Lee, Monte Westerfield, Pieter J. de Jong, Leonard I. Zon, John H. Postlethwait, Christiane Nüsslein-Volhard, Tim J. P. Hubbard, Hugues Roest Crollius, Jane Rogers, and Derek L. Stemple. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496:498 EP –, Apr 2013.

- [HN98] Jerry L. Hintze and Ray D. Nelson. Violin plots: A box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.
- [HS22] Francis Hamilton and John Swaine. *An account of the fishes found in the river Ganges and its branches*, volume [Text] (1822). Edinburgh :Hurst, Robinson, and Co., 1822. <https://www.biodiversitylibrary.org/bibliography/59540> — Atlas has no t.p. — The atlas plates are signed by the engraver J. Swaine. — Errata slip tipped-in after p. vii. — Colophon: Printed by George Ramsay & Co. Edinburgh, 1822. — Includes bibliographical references and index. — Wood, C.A. Vertebrate zoology, — p. 266 — Nissen, C. Zoologische Buchillustration, — 1809.
- [HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [IC18] Itseez Intel Corporation, Willow Garage. Image processing in opencv. https://docs.opencv.org/3.4.5/d2/d96/tutorial_py_table_of_contents_imgproc.html, 2018. Last Accessed: 10/2/2019.
- [KB15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [Mey18] Jason R. Meyers. Zebrafish: Development of a vertebrate model organism. *Current Protocols Essential Laboratory Techniques*, 16(1):e19, 2018.
- [Spa18] Hermes Spaink. Age classification of zebrafish larvae using machine learning from hog features, 2018.
- [SVZ14] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings*, 2014.
- [Uni14] Union Biometrica, Inc. *VAST BioImager System Specifications*, October 2014.

- [Ver18] Wilco Verhoef. On automatic segmentation and classification on age groups of zebrafish larvae microscopy images by deep neural networks, 2018.
- [ZF14] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham, 2014. Springer International Publishing.
- [ZMSMo8] Y. Zhang, B. J. Matuszewski, L. Shark, and C. J. Moore. Medical image segmentation using new hybrid level-set method. In *2008 Fifth International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics*, pages 71–76, July 2008.