Computational optimisation of optical projection tomography for 3D image analysis

Tang, X.

Chapter 5

Automated Detection of Reference Structures for Fluorescent Signals in Zebrafish with a Case Study in Tumour Quantification

This chapter is based on the following publication:

Tang X., Wijk, R.C. van., He S., Spaink H., Spaink H.P & Verbeek F.J. Automated Detection of Reference Structures for Fluorescent Signals in Zebrafish with a Case Study in Tumour Quantification. *IEEE Journal of Biomedical and Health Informatics*. (under review)

Chapter summary

Zebrafish as a vertebrate model plays an important role in biomedical researches such as development, disease, toxicological and drug discovery studies. In this chapter we assume that fluorescent markers represent a specific signal of interest. We aim to quantify these signals in zebrafish, to provide accurate experimental information, for e.g. drug discovery, in an automated and efficient way. We first define the quantification approaches with a case study in tumour growth. Based on the definition, the reference structures for the quantification, obtained from bright-field images, are studied.

In order to automatically detect the reference structures from 3D bright-field images with a high performance, we use the deep learning approach to obtain a segmentation of the reference structures for each sample. The 3D images are obtained and reconstructed from the optical projection tomography imaging. According to our experiments, the automated approach for detecting reference structures is a promising method for the relative quantification of fluorescent signals in zebrafish.

5.1 Introduction

In this chapter we will focus on the application of imaging of zebrafish in OPT for disease modelling. In biomedical research, zebrafish has become a widely used model organism in the last decade because of its fecundity, its physiological and genomic similarity to mammals, the existence of many genomic tools and the ease with which large, phenotype-based screens can be performed [92]. Mammalian models of absorption, distribution, metabolism and excretion (ADME) or pharmacokinetics and efficacy, are considered expensive and laborious and consume quantities of precious compounds. Compared to this, zebrafish is more cost-effective and can therefore be a useful alternative to mammalian models.

5.1.1 Research questions

Drug discovery involves a complex iterative process of biochemical and cellular assays, working up to *in vivo* validation in animal models and ultimately in humans. Zebrafish is considered promising in accelerating the process of drug discovery with a comprehensive advantage of scale, high-throughput screening and physiological complexity. In disease modelling and treatment, e.g. drug discovery for tumour treatment, zebrafish has revealed its effectiveness and advantages [93]. Using zebrafish as disease model for tumour means exposures to different levels of drug treatment. The performance of this drug treatment can be expressed in qualitative terms, i.e. using the visual signal, as well as in quantitative terms, i.e. measurements of the intensities and extend of the signal.

To this end a specific fluorescent signal, i.e. the expression of a representative gene is used. However, compared to visual qualitative assays, quantitative assessment is more comparable and transferable and hence more convincing. The quantification of the expression of specific disease within zebrafish such as tumour [93] can give a direct and accurate insight of tumour size and shape, as well as make the precise comparison of different treatment groups.

Depending on different research demands and available facilities, the disease phenotype can be represented as 2D or 3D microscope images. For a whole-mount zebrafish, 2D images from a stereo-fluorescence microscope can provide fast information on the structure of a tumour in a single specimen; de facto this is a projection of 3D information. For measurement and phenotypical description, this should be considered with caution. Projecting and imaging a zebrafish from a different angle will result in a different 2D image. This difference can result in inaccurate and biased quantification. In contrast, 3D imaging of disease phenotype reconstructs the 3D structure, therefore potentially more accurate for disease phenotype quantification. Therefore, we take it as a research question: given a 3D image, what are the possible solutions for the quantification of disease phenotype and to what extent the quantification process can be automated.

5.1.2 OPT as a solution for whole-mount imaging

As mentioned, zebrafish is a good model system for the disease studies. For overview of disease and disease progression in zebrafish, whole-mount imaging is indispensable. With confocal laser scanning microscopy (CLSM), light microscopy (LM) and, to a lesser extent, scanning electron microscopy (SEM) the size of the specimen limits the application of whole-mount imaging [15], [94]. With MRI [16] the strength of the magnetic field determines the resolution that can be obtained for whole-mount imaging and a *mm*-scale object requires quite a strong magnet. With a right choice of optics, optical projection tomography (OPT) [95] can conveniently operate with *mm*-scale objects. It can display gene expression or a specific staining in the bright-field or fluorescence channel, while the specimen as a whole can be visualized. In this manner OPT adds an important range of scale that can be investigated. It allows for the acquisition of whole-mount images of animal/plant tissues as well as organs/organisms [18], [19]. OPT has also been studied for its capability of imaging with excellent spatial resolution and contrast and minimal shadowing artefacts produced from back-projection reconstruction after tomogram acquisition. Therefore, we take OPT into account to assess the whole-mount 3D imaging of zebrafish.

5.1.3 Multi-channel analysis of whole-mount zebrafish

In order to exclude biological variation and individual differences, large scale analysis of zebrafish for disease treatment is necessary. This means that multiple zebrafishes will need to be quantified and averaged to describe the disease progression or the performance of drug exposure at different time-points in disease.

To our best knowledge, two methodologies are used for the quantifying of a disease model. One is measuring a read-out in absolute sizes of the disease marker in either 2D or 3D, named as absolute quantification. This involves a segmentation of the signal that is representing the disease, i.e. the marker, and pixel/voxel size calibration. For absolute quantification, only the fluorescence channel, with fluorescent disease marker, is required. In this particular case, the pixel/voxel size calibration for the imaging system is required so as to make measurements comparable and transferrable between different systems. Another approach is calculating the relative ratio of disease phenotype, i.e. fluorescent signal, referring to a specific structure such as *Body* or *Eye*. This is mostly depending on detection of the reference structure (RS) and has generality across imaging systems and between specimens. In this case often both modalities, i.e. fluorescence and bright-field, are required; with one for the fluorescent marker and the other for the RS. Taking advantage from computational techniques and resources, in this chapter we construct and detect RS for relative quantification of disease phenotype. To that end we introduce two definitions so as to support our approach for relative quantification.

Definition 1 Phenotype: The total appearance of an organism determined by interaction during development between its genetic constitution (genotype) and the environment [96].

Definition 2 Reference Structure (RS): a structure that is a part of the organism under study that is used for a relative comparison over scale and/or time. Usually, the RS is used in a normalization to establish an effect in a phenotype.

Relative quantification of 2D disease phenotype for each zebrafish can be achieved by standard image analysis tools [97]. When considering throughput of the data, a more automated approach is preferred. However, quantification of 3D disease phenotypes for a zebrafish specimen is much more difficult to obtain than it is in 2D. Therefore, in order to prevent manual labor and enable application of analysis on a larger scale, automated image analysis is necessary. We focus on the automation of obtaining RS to accelerate the throughput of the 3D analysis of zebrafish.

The research presented in this chapter concentrates on relative quantification of disease phenotyping, specifically constructing 3D reference structures in zebrafish. We have chosen to work with two RSs that are always visible: the *Body* and the *Eye*. The *Body* represents the overall phenotype of a sample, providing a normalization standard for all the samples. The *Eye* is a local RS with less deviation among specimens of the same stage, it is also easier to detect because of its clear texture. The goal of this research is to automatically detect both RSs and using them for the calculation of the relative quantification of fluorescent signals. Here we focus on the 3D quantification of signals from a disease. The automated RS detection can be generalized to other 3D fluorescent signals in zebrafish.

The RSs we need for the 3D relative quantification are automatically detected from the bright-field 3D image by using segmentation techniques. Concerning the large-scale requirement, we are into exploring an approach for automated detection of the RSs. The transparency and inhomogeneity of the zebrafish make segmentation performance cumbersome, particularly in 3D, when using traditional segmentation algorithms; i.e. threshold-based, region growth, graph cut and traditional machine learning techniques. The challenge must be seen in the high similarity of intensity between voxels inside and outside zebrafish, as well as the edge/surface discontinuities [98]. Fortunately compared to *Body*, *Eye* has a more dense tissue, therefore more discriminative. To meet our requirements for automated RS detection, advanced segmentation approaches will have to be explored.

5.1.4 Related work

During the last twenty years, to our best knowledge there were just a few research topics on automated analysis in zebrafish. Mikut *et al.* [99] contributed a survey on automated processing of zebrafish-related data and generalized the workflow for analysis of biomedical research on zebrafish model. They showed some examples of automated

image analysis, including cell tracking during embryogenesis, heartbeat detection, anatomical landmarks, dead embryo detection, recognition of tissues, and quantification of behavioral patterns. In general, the microscopy images could be classified into two categories as mentioned: bright-field images and fluorescence images. Analysis of the fluorescence images is relatively easy compared to bright-field because fluorescent intensity typically reflects specific information of interest.

Previous work on fluorescent imaging varies with the specific research topic. Understanding bacterial infection was accomplished with a template-based segmentation method through which the shape of a zebrafish larva was detected. The bacterial load was obtained from the fluorescent channel and normalized to the size of the larva, or specific parts thereof [100]–[103]. An automated segmentation was utilized to zebrafish heart based on 2D light-sheet fluorescent images, accompanied by 3D reconstruction in the second stage [104]. They followed the pipeline that 3D volume is reconstructed based on the segmentation results of tomograms. Segmentation of the axial skeleton and spine of the zebrafish are also common in developmental research [105],[106]. Their segmentation was implemented on images of fluorescent maker in notochord sheath cells and with conventional segmentation techniques acceptable results have been achieved [105], [106]. More lately segmentation of developing zebrafish vasculature was proposed from light sheet fluorescence microscopy imaging [107]. They used the open source software Fiji [97] for segmentation of the fluorescent marker and achieved satisfied results [107]. Different from the aforementioned non-learning algorithms, Zhang *et al.* [108] first brought the deep learning technique to vessel segmentation on images from 3D confocal imaging in 2019. They obtained promising results on accurate segmentation of challenging vessel data that were labelled with green fluorescent protein (GFP). The 3D image data were acquired with confocal microscopy and the segmentation was implemented on the reconstructed slices [108].

In addition, whole-mount specimen segmentation can be achieved using the bright-field image of the sample. It is typically related to phenotype and behavior analysis in the research of development and drug discovery [98], [109]–[111],. Wu *et al.* [109] proposed a hybrid method which integrates region and boundary information into an active contour model considering the ambiguity of edges for 2D image segmentation. Later, Xiong *et al.* [110] presented the level-set model to segment zebrafish on image slices from confocal microscopy and achieved promising results on 3D images. Inspired by the good performance of level-set model, Guo *et al.* [98] integrated mean shift to level-set model for accurate 2D zebrafish image segmentation in bright-field channel, and then used the segmented masks for 3D reconstruction of zebrafish surface. Recently, instead of phenotype analysis, Ishaq *et al.* [111] classified zebrafish deformation, i.e. normality or deformation, based on 2D bright-field images for drug discovery using a deep neural network.

5.1.5 Structure of this chapter

The research presented in this chapter is to explore a segmentation approach, aiming at automated detection of the RSs for fluorescent signals in zebrafish. The detected RS is able to help with relative quantification of fluorescent signals in zebrafish. In section 5.2 we present the materials and methods used for our research. In section 5.3 we further elaborate the design and implementation of the segmentation approach for automated detection of the two 3D RSs that we use here. In section 5.4 the experiments and results will be presented, with a case study in tumour quantification, followed by conclusions and discussion in section 5.5.

5.2 Materials and methods

Here, we will first explain the specimen and sample preparation of zebrafish used for automated RS detection. The 3D imaging and reconstruction framework will follow afterwards. Based on the reconstructed 3D image, the approach to the relative quantification of fluorescent signals will be formed.

5.2.1 Zebrafish

Both for the *Body* and *Eye* reference structures, 38 zebrafish samples are used to learn parameters of the segmentation approach for automated detection. These zebrafishes are from three different stages including 5 *dpf*, 6 *dpf* and 7 *dpf* and all of them are cleared with the BABB protocol (cf. [20]). As we are only interested in the bright-field image for the detection, the samples are not necessarily stained or marked with fluorescent markers. For our experiments we have eight 5 *dpf* zebrafishes, fifteen 6 *dpf* and fifteen 7 *dpf* zebrafish without staining.

5.2.2 OPT imaging and reconstruction

3D imaging with OPT is suitable for zebrafish imaging as it can deal with the size range to produce whole-mount images. With OPT both bright-field and fluorescence modalities can be accomplished and the acquisition of these channels is done in a sequential manner. In confocal microscopy the details are inspected at cellular level but it compromises information at the global level [112],[113]. The strength of OPT is that it enables observation of the whole specimen, i.e. the zebrafish, at a tissue level with depth ranging from millimeter to centimeter. This character of a large depth and field of view (FoV) enables the analysis the zebrafish as a whole on volumetric level. This also explains the conditions of our research work for detecting the RSs for fluorescent signals in zebrafish.

The bright-field images from the 38 samples are acquired using the OPT imaging system as depicted in chapter 1, cf. § 1.2, producing a 3D tomogram of $1360 \times 1036 \times 400$ for each sample; 1360×1036 per image over 400 angles in full revolution. The 3D bright-field image of each zebrafish is obtained from the corrected tomogram set by

using the iterative reconstruction algorithm as described in chapter 4, cf. § 4.2. The transform from original OPT tomogram set to corrected ones are accomplished by applying the centre of rotation (CoR) correction approach in chapter 2, cf. § 2.3. The application of the iterative reconstruction [28] results in a 3D image sized $R \times R \times 1360$, where R is the determined by the CoR value and 1360 refers to the number of slices in 3D. The 3D images of all the 38 zebrafishes are subsequently used to learn the parameters of segmentation approach for RS detection. This design of the automated RS detection will be elaborated in the section on design and implementation.

5.2.3 Relative quantification

The relative quantification of fluorescent signals is based on the reconstructed 3D images from both the bright-field and fluorescence channel. Specifically, the fluorescent signals from the 3D image of the fluorescence channel are first segmented through a threshold-based algorithm. In this manner, a sub-volume is produced from which the fluorescent signal is quantified. For RS detection, we need first to segment or identify the RS from the 3D bright-field image, producing the sub-volume of the reference structure. The ratio of the two volumes obtained from the segmentation results is defined as the relative quantification of the fluorescent signal.

5.3 Design and implementation

In this section, we focus on the design and implementation of automated RS detection for 3D quantification. With the 3D bright-field image of zebrafish, cf. § 5.2.2, the RS, i.e. *Body* or *Eye* will be identified by the supervised segmentation approach. We investigate how this can be accomplished by a convolutional neural network. For each of the RSs binary ground truth is realized with annotation software, such as TDR or Amira. In order to train a high-performance segmentation network for each RS, we employ U-net segmentation network [29], [114] implemented in both 2D and 3D image space.

5.3.1 Segmentation of reference structures

In order to reduce the computational load, we rescale the 3D image to $512 \times 512 \times 680$ for the 2D U-net segmentation and to $128 \times 128 \times 340$ for the 3D U-net segmentation, thereby compromising resolution. To our best knowledge, 3D U-net has the highest performance when each 3D image is directly fed into the network. However, constrained by the memory, this is not feasible for our 3D image due to its large size. Alternatively, we resized and cropped the 3D image as $128 \times 128 \times 340$ and feed the 3D U-net in a patch way. From the 38 samples as introduced in§ 5.2.1, 35 samples are used for training and validating the segmentation network, whilst 3 samples for testing or evaluating the performance of the segmentation approach. So, there are 23800 slice samples in total for training and validating the 2D U-net, but much less volume patches for 3D U-net training. In order to increase the sample size for 3D U-net, we decrease the patch size to $64 \times$

64×64 with an overlap of 16 to cover 3D context well. This results in 1113 cubic patch samples for training and validation. Concerning the high imbalance of voxels between RS and background, in particular for *Eye*, we exclude the patches that have no object in the ground truth data. For the remaining patch samples, data augmentation including distortion and flip is employed before being fed into the network.

1) 2D Unet

The 2D U-net network [29] feeds 2D images in the *Input* layer. As our starting point is 3D images, we need a slice extractor to provide the 2D images to the net. In this manner, the slice extractor also contributes in stacking all slices back to 3D volumetric images after *Output* layer. The network structure has been elaborated in chapter 4, cf. § 4.3.2.

2) 3D Unet

In the 3D U-net [114], different from the 2D U-net, 3D image are used as *Input* layer. This is schematically depicted in Figure 5.1. As we now work in 3D, a slice extractor is no longer required, and instead a cubic patch is used. The equation operation fits the cubic patch into the *Input* or *Output* layer. The *Output* layer of the network is named as the segmentation map. This will, to some extent simplify the description of network layers including 3D Input and number of kernels. Besides the features on each reconstruction slice, the encoder and decoder in 3D U-net also considers the correlation between adjacent slices by using 3D *Maxpooling*, *Convolution* and *Upsampling* layers. Compared to a 2D U-net approach, this produces smoother segmentation results. Similar to 2D U-net there is one more *Convolution* layer before each *Maxpooling* or *Upsampling* layer (not shown in Figure 5.1). The *Merge* layer after each *Upsampling* layer integrates shadow layer into a deeper one, this yielding even more informative layers. The segmentation result of the volumetric patch is achieved based on the thresholding of 3D segmentation map, the *Output* layer of network.

5.3.2 Learning scheme

A CNN can only be successful when it is properly designed and parameterized. For this, the proper loss functions, optimizers and learning rate will be employed. In this section, we elaborate on these schemes and how they should be applied.

1) Loss & Metrics

Loss and metrics functions play an important role in training networks, because they provide a criterion for measuring the similarities between prediction and truth, and determine the level of convergence for the training process. With the Sigmoid activation function at the *Output* layer and the binary segmentation problem for both RSs, we first

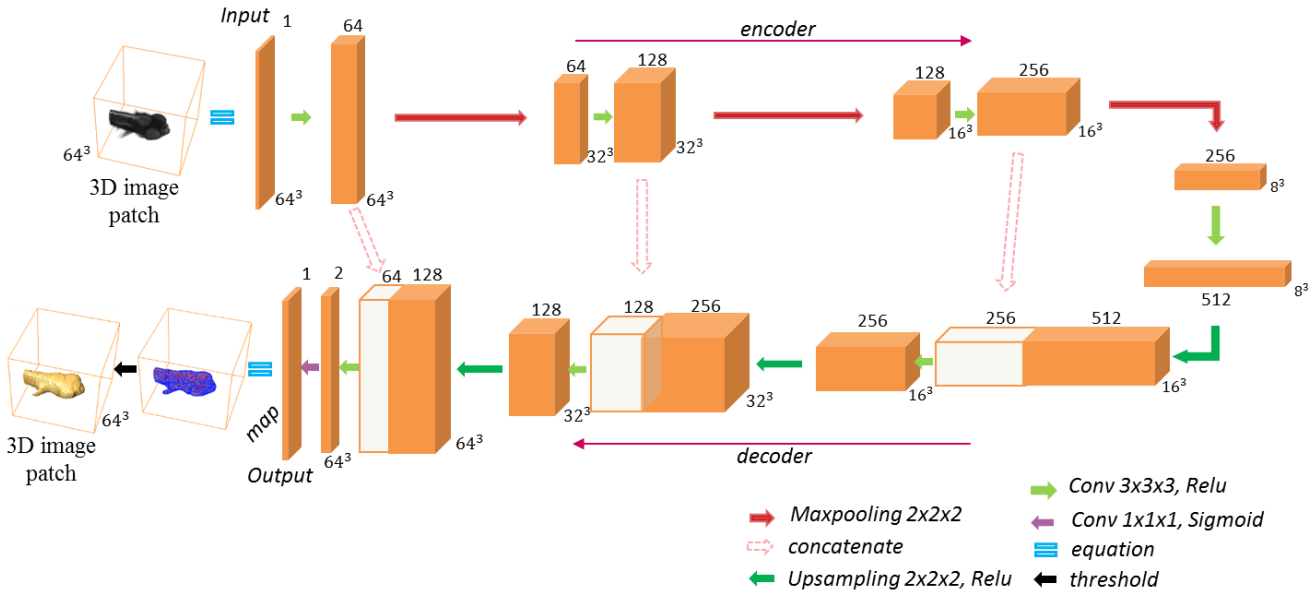

Figure 5.1. The 3D U-net framework for volumetric segmentation of the *Body* RS in zebrafish. A volumetric patch of reconstructed 3D image, i.e. 64x64x64, is fed into the network as *Input* layer. It inherits the characteristics of *Max pooling*, *Upsampling* and *Convolution* layers; moreover, it includes the concatenation operation to merge similar layers of different depths. In this structure, every operation is implemented in 3D. The equations operation at *Input* and *Output* layer help to overcome the visualization gap between 3D image and 4D network layers. The values of the segmentation map at the *Output* layer are in the range of [0, 1], representing the network response to the *Body* RS. By applying threshold=0.5 to the segmentation map, the binary 3D *Body* RS is obtained as shown in the 3D image patch.

use the binary cross entropy loss [115],[116] and accuracy metrics. In this case, the segmentation performance of both background and RS, i.e. foreground, are taken into account for updating the network weights according to:

$$EL = \sum_{i=0}^{N} g_i^r \log(p_i^r) + (1 - g_i^r) \log(1 - p_i^r) \quad (1)$$

where EL is the loss, whilst p_i^r represents the probability of voxel i being predicted as the RS r, and g_i^r symbolizes the corresponding ground truth. This normally results in very small losses and high accuracies during training process when the classes are highly imbalanced. In classification problems, class imbalance exists if the number of samples for each class varies. Such imbalance may have a different impact on the classification results depending on the level of imbalance. For the 3D images in our work, the number of voxels for RS class is generally smaller compared to the background class.

With respect to segmentation map in both networks, it only records the response of the positive class, i.e the RS. This means that the network can be updated according to the assessment of the segmentation performance for the positive class. To this end, the Dice coefficient and Dice loss [117] are applied for the training of the network. With the Sigmoid activation, the *Output* layer consists of one plane for the RS. Now P and G are

the set of predicted and ground truth binary labels. The Dice similarity coefficient between two volumes is defined as [117], [118]:

$$D(P,G) = \frac{2|P \cap G|}{|P| + |G|} \tag{2}$$

It weighs FP and FN (precision and recall) equally. The Dice coefficient loss is then defined as follows:

$$DL = 1 - \frac{2\sum_{i=1}^{N} p_i^r g_i^r + \epsilon}{\sum_{i=1}^{N} p_i^r + \sum_{i=1}^{N} g_i^r + \epsilon} \tag{3}$$

where p_i^r represents the probability of voxel i being predicted as the RS and g_i^r symbolizes that of the ground truth. ϵ is a secondary functional term which helps the loss function converge more effectively.

Considering the high imbalance of 3D image for RS detection, in addition the Tversky loss function [119] is also applied. This metric was formulated based on the Tversky index [120], which gives FP higher weights than FN in the training of the network. The Tversky index between prediction and ground truth volume is formulated as:

$$T(P,G;\ \alpha,\beta) = \frac{|P \cap G|}{|P \cap G| + \alpha|P - G| + \beta|G - P|} \tag{4}$$

where α and β control the magnitude of penalties for FP and FN, respectively. Accordingly, the Tversky loss function is formulated as:

$$TL = 1 - \frac{\sum_{i=1}^{N} p_i^r g_i^r + \epsilon}{\sum_{i=1}^{N} p_i^r g_i^r + \alpha\sum_{i=1}^{N} p_i^r g_i^0 + \beta\sum_{i=1}^{N} p_i^0 g_i^r + \epsilon} \tag{5}$$

Here p_i^r and g_i^r have the same meaning with Dice coefficient loss, and p_i^0 and g_i^0 are separately the probability of voxel i belonging the background (label = 0) in prediction and ground truth. According to the literature [119], for $\alpha = 0.3$ and β=0.7 the Tversky loss function has the best performance in managing highly imbalanced data. Therefore, we adopt these values to the CNN for our segmentation.

The aforementioned three state-of-the-art metrics and loss functions are widely used in biomedical image segmentation [121] because of their stability and robustness. In the results section we will apply the three metrics and loss functions to our data and segmentation network, and explore how they perform on the results.

2) Optimizer & Learning rate

An Optimizer is an optimisation algorithm that regulates and determines the route of converging for the loss function. Our volumetric data exhibit sparseness, i.e. compared to the object voxels the ratio of the background voxels is high. The Adam optimizer [122] is typically useful for such sparse data. Adam was designed to combine the advantages of Adagrad [123],[124] and RMSprop [125],[126] with momentum [122] as an improved version of stochastic gradient decent (SGD) [87] for training deep learning models. This makes it suitable to work with sparse gradients on noisy data. Another advantage of Adam is that

the rule for step size updating, is invariant to the magnitude of gradient. This helps to go through areas with low gradients such as saddle points and ravines [122].

In order to accelerate training process and to some extent improve the performance of the deep network [127], we employ the stochastic gradient descent with warm restarts (SGDR) [128] in a cyclical learning rate scheme. The idea of this strategy is to decay the learning rate from maximum l^i_{max} to minimum l^i_{min} in a cyclic fashion, using:

$$l_t = l^i_{min} + \frac{1}{2}\left(l^i_{max} - l^i_{min}\right)\left(1 + \cos\left(\frac{T_t}{T_i}\pi\right)\right) \quad (6)$$

where *i* is the current cycle for learning decay. The beginning of each cycle is referred to as a restart. T_i is the length of the i^{th} cycle, deciding the number of epochs in this cycle. Experimentally, it is preferred to increase T_i as *i* increases during training. The increasing step for each restart cycle is controlled by T_{mult}. T_t accounts for the number of epochs that have been performed since the last restart or in the current cycle *i*. This means that the learning rate will decay for each epoch within each cycle, and the decay speed will decrease as the cycle progresses. Experimentally we set $l^i_{max} = 10^{-4}$, $l^i_{min} = 10^{-6}$, $T_{mult} = 1.5$ and number of epochs as 500. In this manner we are achieving reasonable results. For more information about the learning rate, we refer to [128].

5.4 Experiments and results

The experiments are first applied to the 3D bright-field images of the 38 zebrafish samples to train and evaluate the RS segmentation network (cf. § 5.3.1). Both networks and different metrics are used for comparison to achieve the highest performance of the segmentation network. This network is then employed for the case study in tumour quantification (cf. § 5.2.4) as a test for automated RS detection and phenotype quantification.

5.4.1 Detection of 3D reference structures

In order to use the segmentation network for automated detection of the RSs from the 3D bright-field image, we first need to train and optimize the segmentation network with the 35 training samples. This means that the performance of the segmentation network needs to be evaluated first on the 3 testing samples with evaluation metrics, before it can be used further for 3D quantification. To this end, we introduce the evaluation metrics, followed by optimisation and evaluation experiments implemented on both RSs for this study.

1) Evaluation metrics

To evaluate the performance of different loss functions on both networks for each RS, we spilt the 38 samples into three sets: 28 samples for training the networks, 7 samples for validation and 3 samples for testing. The performance is compared by

applying five different evaluation metrics to the 3 test samples from individual prediction. The evaluation metrics we employ include the Dice similarity coefficient (DSC) [117] or *F*1 score, sensitivity, specificity, *F*2 score and area under the Precision-Recall curve, i.e. APR score [129]–[131]:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$F2 = \frac{5TP}{5TP + FP + 4FN} \tag{10}$$

where *TP*, *TN*, *FP* and *FN* are the true positive, true negative, false positive and false negative rates, respectively. Sensitivity or recall, measures the proportion of actual positives that are correctly identified as such and it also quantifies the ability to avoid false negatives. Specificity or precision, quantifies the ability to avoid false positive. The *F*2 score is an effective measure for cases where recall is more important than precision compared to *F*1 that equally measures the recall and precision. To critically evaluate the segmentation performance of different networks for highly unbalanced data, in our case in particular for *Eye*, we use the APR score.

2) Detection of 3D Body reference structure

Compared to 2D, the 3D segmentation of the *Body* RS is more complicated. Specifically, 3D data contains both image information of each slice and correlation between adjacent slices. By using 3D based segmentation techniques such as 3D U-net, we can collect and extract, to some extent, both kinds of information. To this respect it offers more features compared to the 2D *Body* RS. However, because of the transparency of the specimen in the bright-field channel, the difficulty of segmentation in zebrafish, especially the surface, also increases; even more so from 2D to 3D. Figure 5.2 (a) gives us an intuitive idea of what transparency means in a single reconstructed slice of a 3D image. Due to the intensity similarity between background and transparent tissue, the difficulty of segmentation with transparencies in the specimen on a single slice can be assessed by comparing Figure 5.2 (a) and (b).

To investigate the ability of the U-net based segmentation network on transparent specimens, we evaluate both 2D U-net network excluding correlations between adjacent slices and 3D U-net network including correlations between adjacent slices. Both are trained based on the three different loss functions (cf. § 5.3.2). Figure 5.3 shows the results of a test sample achieved from different methods, i.e. two networks with three loss

functions, and the corresponding errors (yellow for *FP* and red for *FN*), compared to the ground truth. From these qualitative results, it is difficult to conclude which methodology performs best on the data. However, we observe that the loss function has an impact on the segmentation errors. From binary cross-entropy loss [116] to Tversky loss [119], regardless of different networks used, the *FN* error decreases whilst the *FP* error increases. This means that both segmentation networks have the highest ability to avoid *FN* errors when using the aforementioned Tversky loss, and the highest capability to avoid *FP* errors when using binary cross-entropy loss.

(a) Reconstructed slice (b) Groundtruth of segmentation

Figure 5.2. An Example for the reconstructed slice of 3D zebrafish image and the corresponding segmentation result for the slice of the *Body*.

To further quantify the performances of different methods and loss functions, five evaluation metrics are reported in Table 5.1 based on the average performance of the three tested samples. The best result of each individual metric across different segmentation networks and loss functions is presented in bold. Since in our work avoiding FP errors is equally important as avoiding FP errors, the combination of network and loss function that has the most bold results in the DSC and APR metrics, is regarded as the overall best performing method. By evaluating and assessing the results in Table 5.1, we concluded that 3D U-net with Dice loss has the best performance on the current dataset. It achieves the highest DSC/F1 and APR score as 93.9% and 88.4% separately.

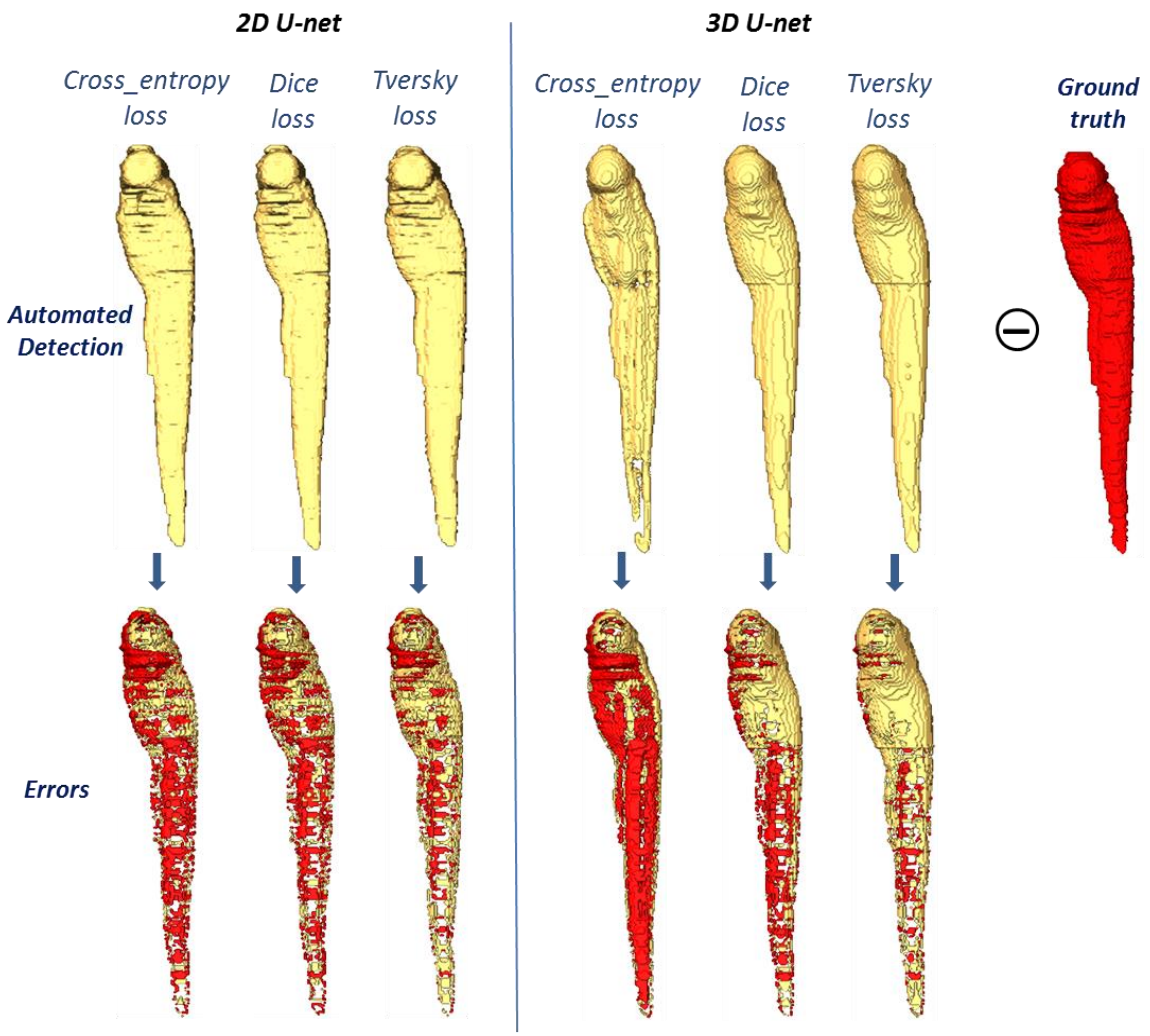

Figure 5.3. Comparison of the detected *Body* (volume rendering), as a RS on a zebrafish sample. The results from different segmentation networks and loss functions are displayed on the top row, as well as the ground truth model manually labelled on the top right. On the bottom there are six errors (volume rendering), corresponding to different segmentation methods or metrics, with yellow showing *FP* and red for *FN*.

3) Detection of 3D Eye reference structure

The tissues in *Eye* are relatively dense, which guarantees a sufficiently distinct range of image intensities for *Eye*. One should, however, be aware that similar intensity patterns also exist in other parts of the zebrafish. A good segmentation method is supposed to discriminate *Eye* from other tissues considering the 3D specific configuration of that shape. The experiments for *Eye* segmentation are implemented according to the description in § 5.3. Ground truth labels for *Eye* are obtained through manual labelling using annotation software; i.e. TDR [132] and Amira [63].

Figure 5.4 presents the comparisons of segmentation results on the same sample, when applying the 2D/3D networks and various loss functions. When assessing the segmentation results, we conclude that 2D U-net with Dice loss and 3D U-net with binary cross entropy loss fail to identify or segment the volume of *Eye* correctly. 2D U-

net with binary cross entropy loss is able to identify parts of *Eye*, whilst with Tversky loss there is an over-segmentation of the volume. Similar to the *Body*, the best performing methods for the *Eye* are also 3D U-net with Dice loss, with much less errors both in the volume and on the surface. In order to identify the best segmentation performance for *Eye*, a quantified evaluation on the dataset is required.

Table 5.1. The average evaluation results of the *Body* detection on three test samples. We compare 6 methods including 2D U-net and 3D U-net across three different loss functions. In order to evaluate the results in a comprehensive and critical way, five evaluation metrics are employed. Referring to the visible errors in Figure 5.3, it is easier to understand the differences. The bold number represents the highest accuracy among 6 approaches for each individual evaluation metric.

Network	Metrics	Loss_function		
		Cross_entropy[116]	Dice [117]	Tversky[119]
2D U-net	DSC	92.80%	93.03%	92.47%
	Sensitivity	92.37%	94.20%	95.70%
	Specificity	**99.88%**	99.84%	99.79%
	F2	92.53%	93.70%	94.40%
	APR	86.37%	86.73%	85.83%
3D U-net	DSC	90.60%	**93.90%**	92.37%
	Sensitivity	90.53%	94.47%	**97.47%**
	Specificity	99.83%	99.87%	99.74%
	F2	90.50%	94.27%	**95.40%**
	APR	82.60%	**88.40%**	85.63%

Table 5.2. The evaluation of *Eye* detection on three test samples. Five evaluations with 6 approaches for segmentation are presented. The bold number in represents the highest accuracy among 6 approaches for each individual evaluation metric.

Network	Metrics	Loss_function		
		Cross_entropy [116]	Dice[117]	Tversky[119]
2D U-net	DSC	54.70%	0%	35.00%
	Sensitivity	43.00%	0%	**99.70%**
	Specificity	99.98%	100%	99.21%
	F2	46.87%	0%	55.47%
	APR	34.43%	0.20%	22.03%
3D U-net	DSC	0%	**91.80%**	90.64%
	Sensitivity	0%	90.78%	93.62%
	Specificity	100%	**99.99%**	99.98%
	F2	0%	91.17%	**92.40%**
	APR	0.20%	**84.53%**	82.40%

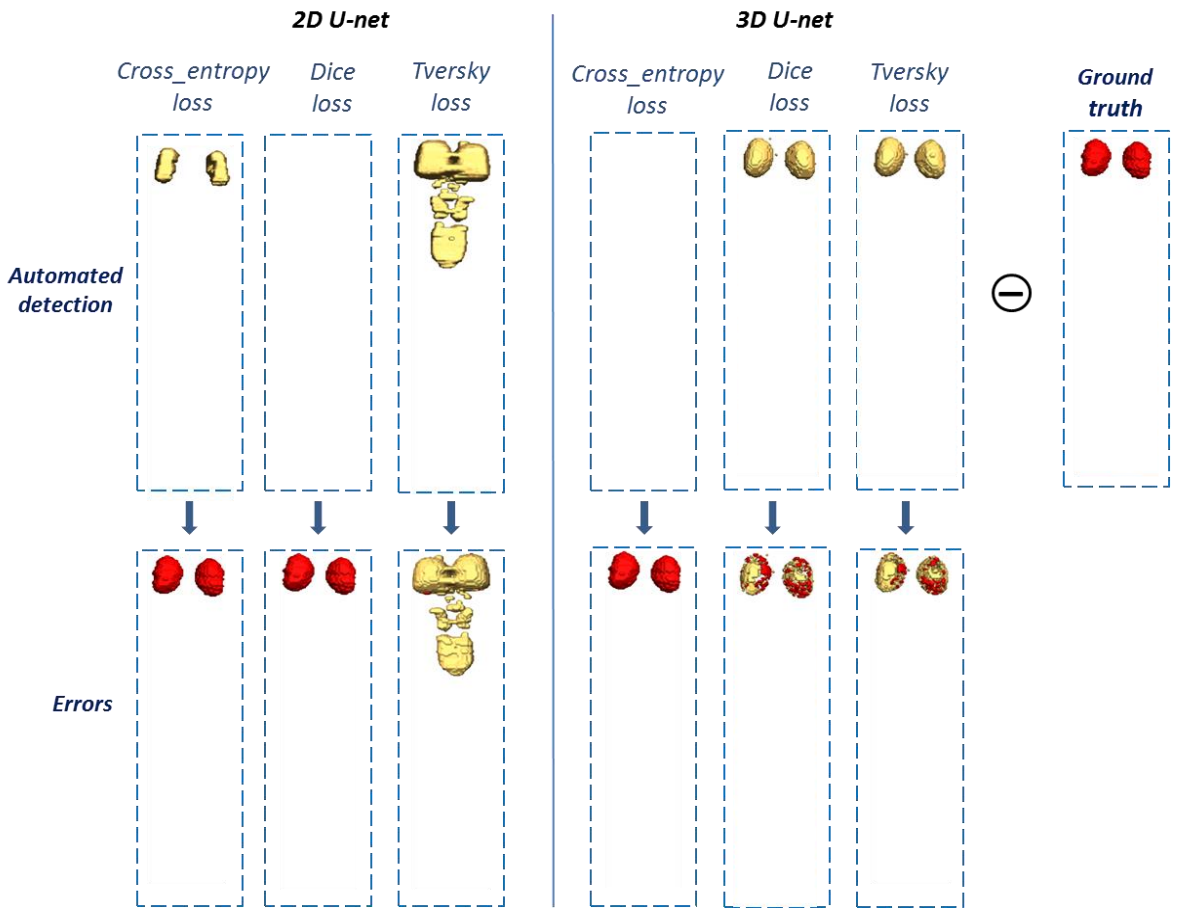

Figure 5.4. Comparison of the detected *Eye* (volume rendering), as a RS on the same zebrafish sample. Similar to Figure 5.3, the detected results and ground truth are presented on the top, while the segmentation errors are showing on the bottom with yellow showing *FP* and red for *FN*.

A qualitative report on the evaluation results on the three test samples is given in Table 5.2. Because the 2D U-net with Dice loss and 3D U-net with binary cross entropy loss fail to detect *Eye*, the accuracies of DSC, sensitivity, F2 and APR are 0%. But specificities are reported 100% because they are able to avoid FP errors. 2D U-net with Tversky loss performs best for sensitivity at 99.7%, but not best for specificity at 99.21%, meaning it has the highest ability to prevent FN errors but lowest ability to avoid FP errors. 3D U-net with Tversky loss has best performance for F2 score at 92.40%. However, overall we conclude that the 3D U-net with Dice loss performs best, achieving a DSC score of 91.80%, specificity of 99.99%, and APR of 84.53%.

5.4.2 Case study in tumour

Tumour growth and remission of neuroblastoma in zebrafish can be observed in a controlled experiment with and without a tumour inhibitor. The experimental condition includes a longitudinal exposure to the tumour inhibitor isotretinoin so that the relative quantification can be used for determining the tumour size at a specific stage of treatment. In this manner the measurements are independent of the variation in the individual

samples and imaging environment (e.g. exposure time and magnification). The performance of the treatment at that stage will be statistically averaged based on the quantification of multiple samples. In this section we describe the 3D relative quantification of tumour as a case study. Prior that, 2D relative quantification is explained for comparison.

1) 2D relative quantification with manual labelling

2D quantification provides a fast and intuitive view of fluorescent signals (tumour) and the RS, i.e. *Body* or *Eye*, in zebrafish in terms of a projection from the sample in both channels. We wish to obtain a measurement with which we can compare samples. Therefore, in 2D we normalize a tumour for a sample *n* by dividing the tumour area t_n by area of a RS f_n, achieving the area ratio $r_n = t_n/f_n$, which is depicted in Figure 5.5 (A) and (B) for a different RS. In the example of Fish 1, the area ratios from the two different RSs are separately 0.3694% and 3.5006%, with threshold-based segmentation for the tumour and manual labelling for the RSs. The performance of a treatment at stage *i* is determined by the average ratio of the *N* samples, $R_i = \frac{1}{N}\sum_{n=1}^{N} r_n$. A satisfactory segmentation of the tumour in the fluorescence channel is basically easy to achieve by using traditional threshold-based algorithms. However, for the segmentation of the RSs, more advanced methods [109],[98],[111] are needed.

When averaging the relative ratios of *N* samples, the treatment performance of all samples R_i should be calculated at the same projection angle, which is difficult to achieve. Another drawback of 2D quantification is that it fails to provide information about the shape of the tumour and the RSs. This limits its capability of representation in true 3D space. Therefore, we change to 3D quantification.

2) 3D relative quantification with manual labelling

Before 3D quantification, the OPT tomograms from individual channel are reconstructed to a 3D image, cf. § 5.2.2. This 3D reconstruction is required for the volumetric analysis. The pipeline of 3D relative quantification for each sample is similar to that of the 2D quantification, so is tumour segmentation in the fluorescence channel. The challenge of the 3D quantification lies in obtaining the volumetric RS. Specifically, the large number of reconstructed slices in 3D image makes manual labelling of a 3D RS impractical. Thus there is a demand for automated RS detection. Figure 5.6 presents an example of workflow for 3D tumour quantification at treatment stage *i* using the individual RS. In the next section, the state-of-the-art automated segmentation strategy based on convolutional neural network (CNN), cf. § 5.4.1, will be presented and implemented. In the same example of Fish 1, the 3D relative quantification results of tumour from the two different RSs are separately 0.1865% and 2.6962%.

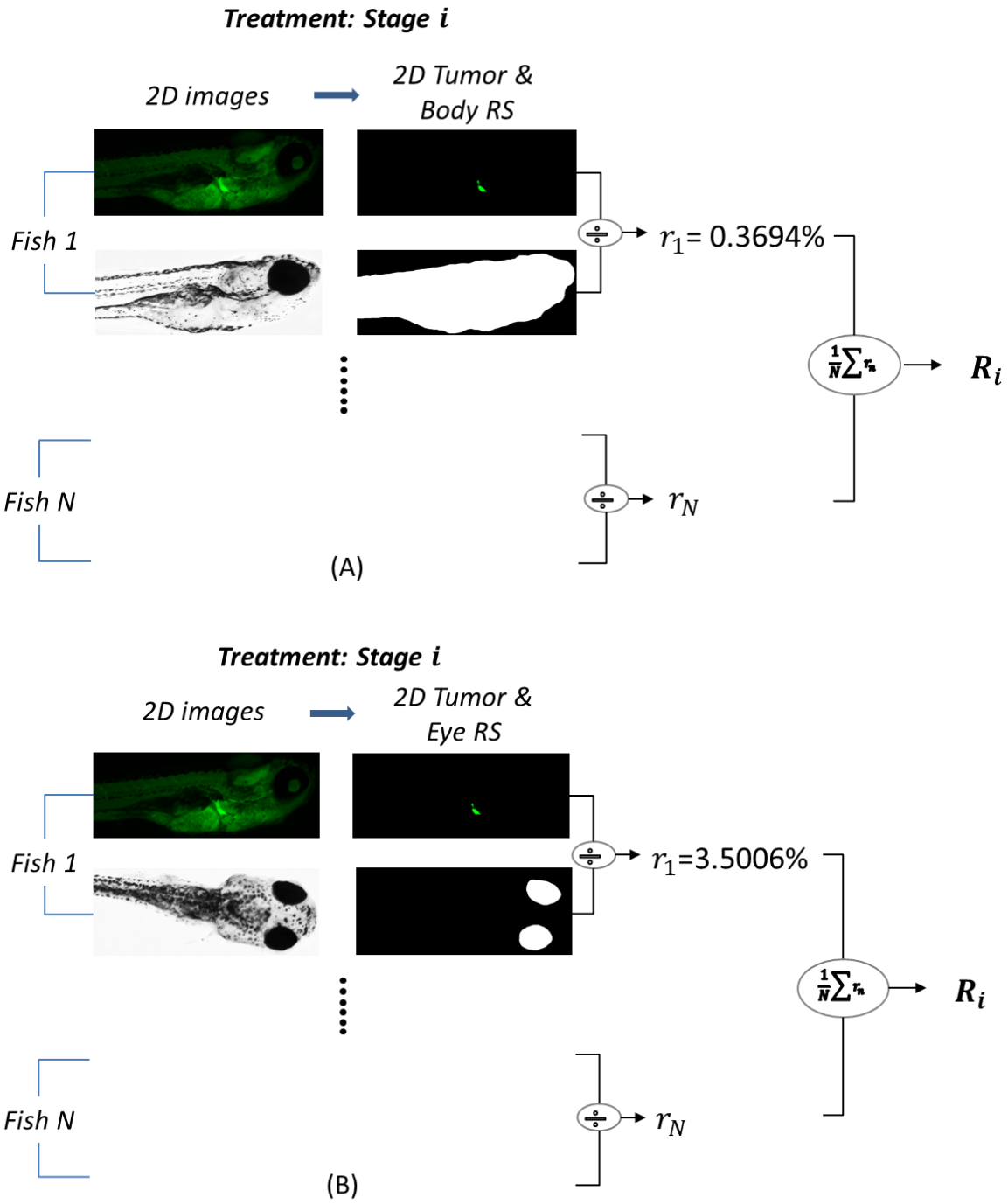

Figure 5.5. 2D relative quantification of tumour at a specific treatment stage *i* using *Body* (A) and *Eye* (B) RS labelled from the OPT tomogram. Each zebrafish sample is represented as a 2D image in the two channels (tumour in fluorescence channel and zebrafish structure in bright-field channel). The segmentation of tumour and the manual labelling of RS (*Body* or *Eye*) are used for the quantification. In the example of Fish 1 with $r_n = t_n/f_n$, the 2D relative quantification results r_n for the *Body* and *Eye* are separately $r_1 = 0.3694\%$ and $r_1 = 3.5006\%$. By averaging the 2D relative quantification r_n of N samples, the performance of treatment at time-point *i*, i.e. R_i is achieved.

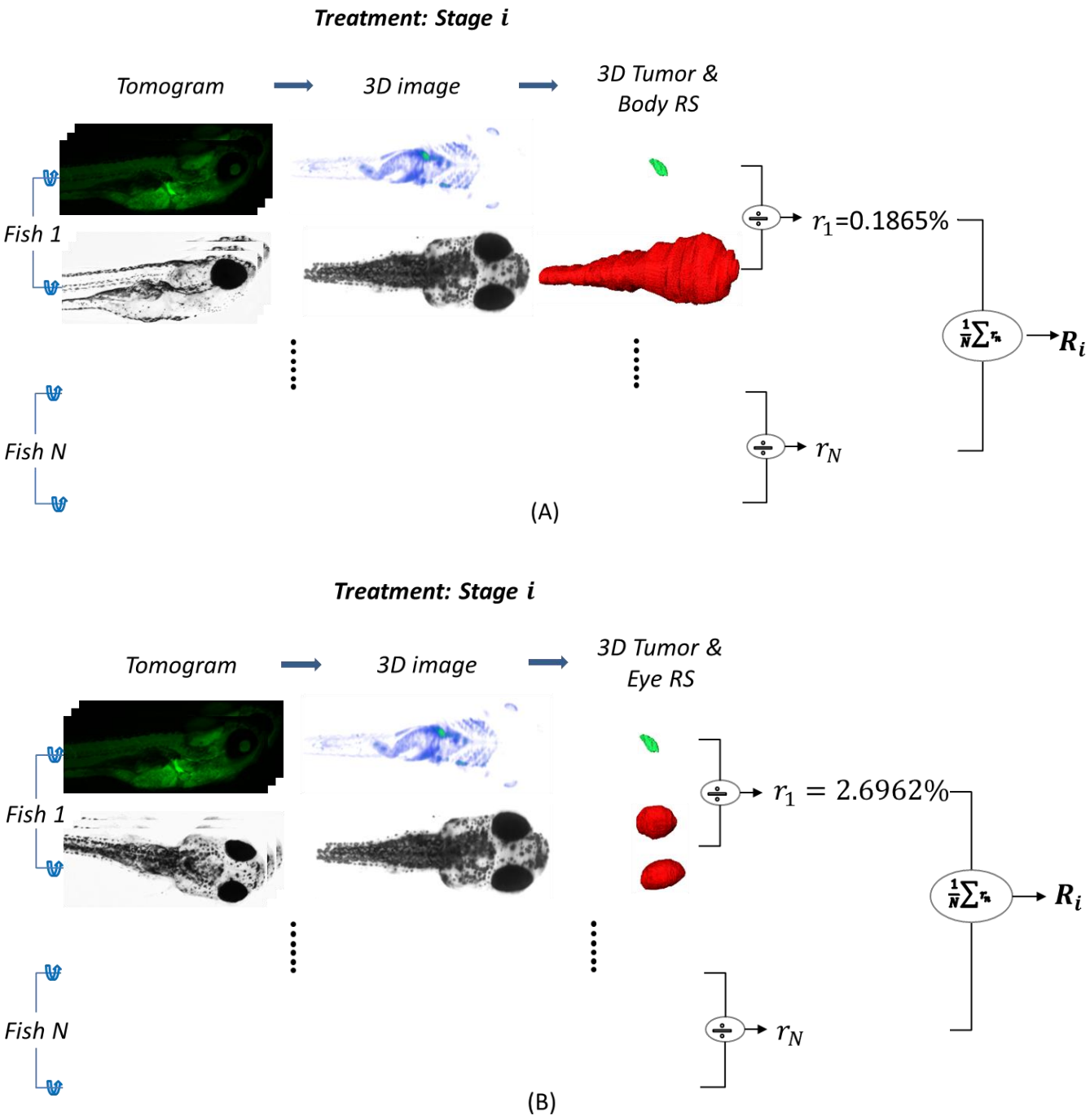

Figure 5.6. 3D relative quantification of tumour using two different 3D RSs. Each zebrafish sample is represented in two different channels, with the fluorescence channel for the tumour and the bright-field channel for the RS. (A) The 3D quantification based on the *Body*. (B) The 3D quantification based on the *Eye*. The 3D quantification is obtained by calculating the volume ratio between the tumour and RS. Specifically, in this example the 3D quantification results for Fish 1 are $r_1 = 0.1865\%$ and $r_1 = 2.6962\%$, calculated based on the two different RSs. As with 2D quantification the performance of treatment at time-point *i*, i.e. R_i is achievable by averaging the r_n of *N* different samples.

3) Comparisons of automated detection and manual labelling of RS for 3D quantification

Studies of tumour growth in zebrafish require measurements of the size and shape of the tumour. In our work the relative quantification of tumour is based on images in the fluorescent and bright-field channel (cf. § 5.2.4). Here we use the output of experiments on the automated detection method to automatically detect the RSs, i.e. *Body* or *Eye*, for a case study in tumour growth. Manual labelling a volume such as the *Body* normally takes at least 1-2 hours for one sample, while automated detection takes a few seconds. Figure 5.7 and 5.8 provide the comparisons of tumour quantification performance based on the RSs from automated detection and manual labelling. Because of its good performance for both RSs, we adopt 3D U-net to the tumour quantification for this case study. The sample compared in this section is a 25 *dpf* zebrafish and it is not included in the 38 samples mentioned before. The tail was not included in the imaging process because as a higher magnification is used, this results in an incomplete zebrafish. However, the segmentation method succeeds to detect the incomplete zebrafish even though it is trained and validated on the 35 complete zebrafishes.

For the *Body* RS, the automated detection using 3D U-net with Dice loss, achieves best overall performance when referring to the manual labelling result as the ground truth. Therefore, we use this segmentation network to automatically detect the *Body* in this case study. The relative ratios of the tumour referring to the *Body* from automated detection r and manual labelling r_f (§ 5.2.4), are separately 0.1883% and 0.1865%, with a quantification error of 0.9651%. The quantification error E_r explains the relationship between the FP and the FN error. A positive error in this case, i.e. the volume from automated detection is larger than the ground truth, means that the FN error is more than the FP error. With respect to *Eye*, the automated method has the best performance of 95.04% for DSC score, 93.77% for sensitivity, 99.98% for specificity, 94.27% for F2 score and 90.37% for APR, using Dice loss function. Dividing the number of tumour voxels by *Eye* volume, we obtain the relative ratio as r = 2.7679%, larger than r_f = 2.6962% from manual labelling as the ground truth, resulting in the positive ratio error E_r = 2.7261%. This is consistent with the fact that the FN error exceeds the FP error shown on the right of Figure 5.8.

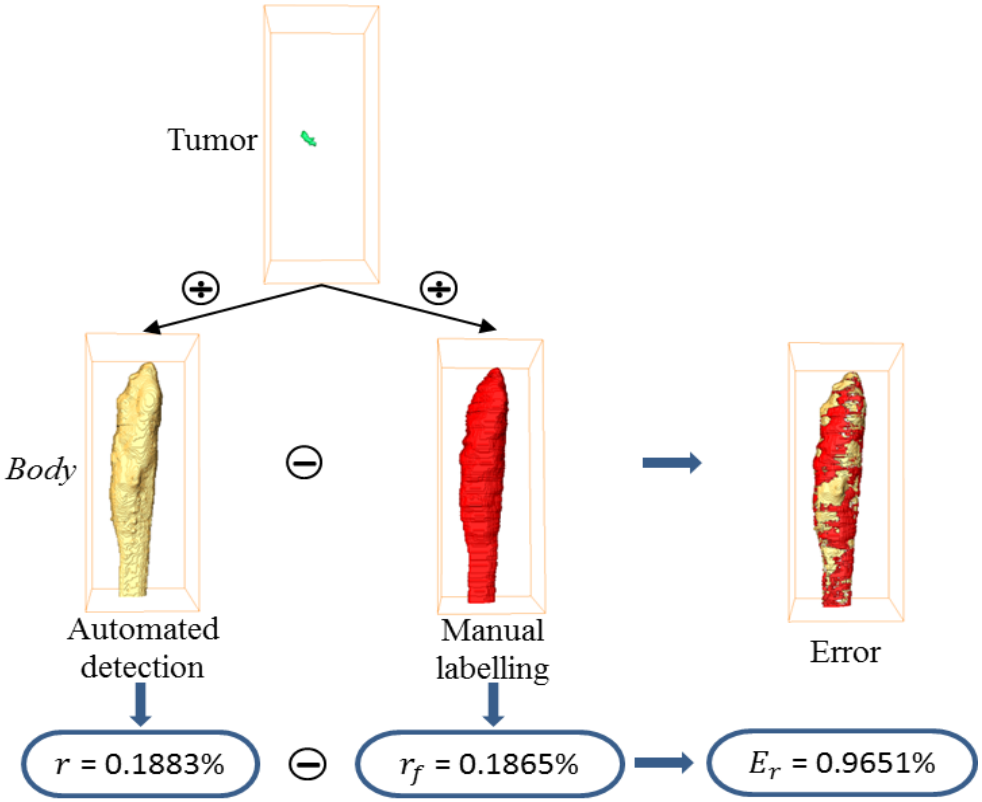

Figure 5.7. Comparison of tumour quantification based on the volumetric *Body* RS obtained from automated detection and manual labelling.

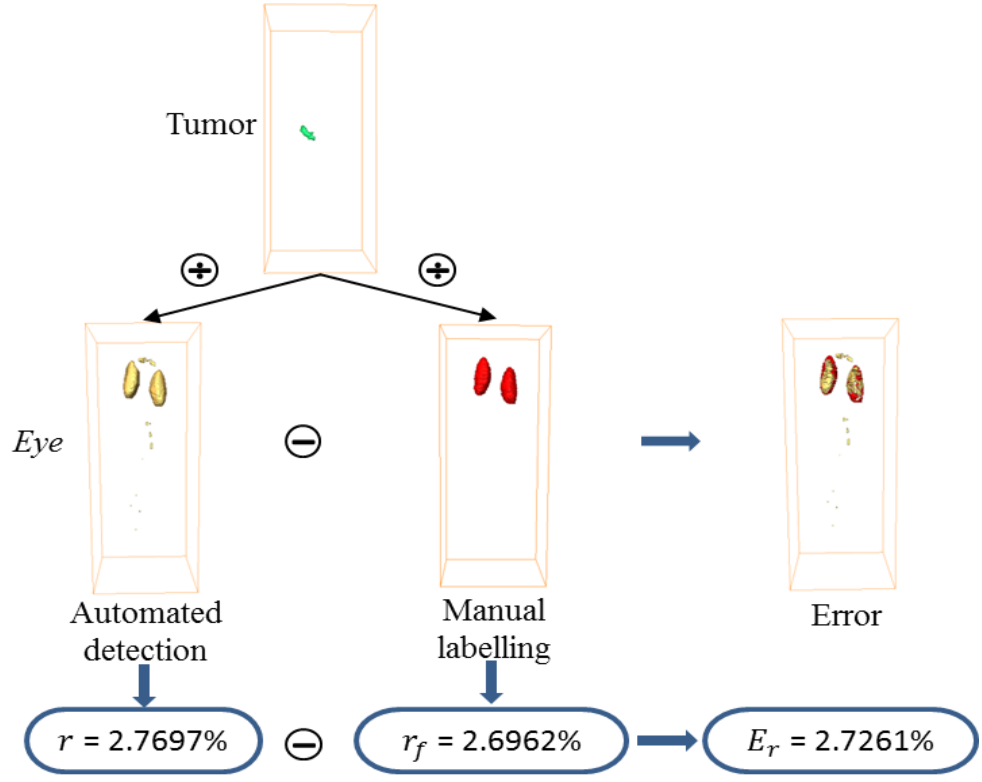

Figure 5.8. Example of tumour quantification based on the volumetric *Eye* RS obtained from automated detection and manual labelling.

5.5 Conclusions and discussion

In the previous sections, we introduced the concept of 2D and 3D relative quantification for fluorescent signals in zebrafish and focused on the technical solution of automated reference structure detection for the 3D quantification. In § 5.4, we compared the detection performances of different segmentation methods for both RSs and conclude that the 3D U-net segmentation network with Dice loss function performs best for automated detection of both RSs on the 38 samples. An overall promising accuracy of over 90% is achieved with respect to five evaluation metrics for both RSs. Subsequently, we compared the relative quantification of tumour between the automatically detected RSs and manually labelled ones. We further investigated how the segmentation errors influence the relative ratio r, compared to the ground truth ratio r_f. From our experiments it is shown that when FN exceeds FP this results in positive quantification error. Whereas, an FN smaller compared to FP results in a negative quantification error. The overall quantification error that we have established is 0.9651% for the zebrafish body RS and 2.7261% for the zebrafish eye RS. Given the experimental setting this is acceptable and reasonable. Nevertheless, given these acceptable outcomes we still can do the effort of further automation of the laborious manual labelling. Moreover, based on the results of the automated detection, further improvement can be accomplished by a careful manual error correction. In this case, the hybrid of artificial intelligence (AI) and human intelligence (HI) gains the best performance.

In this research project we focused on quantification of tumour growth, but the approach can be generalized to the quantification of fluorescent signals in zebrafish. Ideally, they are labelled with fluorescent markers for OPT imaging and reconstruction. With the promising results of automated detection on the limited dataset, better results can be achieved when training the network with a larger dataset. This way we can further improve the accuracy of automated detection of the RSs. Additionally, another contribution of this research is the introduction of a pipeline for relative quantification using automatically detected RSs. This pipeline can be transferred to high-throughput analysis of zebrafish. In the case study for tumour quantification, we presented and evaluated the pipeline for just one sample. Once more samples are available, we would continue with a statistical analysis of the performance of treatment using the proposed pipeline. This is motivated by the fact that statistical analysis of samples at either the same stage or different stages is getting increasingly important for drug discovery.

5.6 Acknowledgment

The work is partially funded by China Scholarship Council (Xiaoqin Tang). We further would like express our gratitude to Merel van’t Hoff and Hermes Spaink (LIACS, Leiden, Netherlands) for their contributions to sample preparation and imaging.