CTPA segmentation to calculate biomarkers for pulmonary embolism risk stratification

Valentine Tcheou
(supervisors: Élodie Puybareau and Odyssée Merveille)

June 2024

Pulmonary embolism (PE) is the third leading cause of death in Europe. It is characterized by the obstruction of a pulmonary artery by a blood clot. Current PE risk stratification models classify a patient's risk of death after 30 days into three categories: low, intermediate and high. The result determines the patient's management protocol, and therefore their treatment. Today, these models are based on functional biomarkers, as well as a morphological biomarker, the right-to-left ventricle (RV/LV) size ratio. This ratio is measured manually by experts on CTPA (Computed Tomography Pulmonary Angiography) exams, which are performed in over 90% of cases. This report presents different methods of CTPA segmentation of ventricles to compute automatically the volume ratio of the ventricles and see if it better correlates with the risk of patient death compared to the size ratio.

L'embolie pulmonaire (EP) est la troisième cause de mortalité en Europe. Elle se caractérise par l'obstruction d'une artère pulmonaire par un caillot sanguin. Les modèles de stratification du risque d'EP actuels classifient le risque de décès de patient après 30 jours en trois catégories : faible, intermédiaire et haut. Le résultat détermine le protocole de gestion du patient et donc son traitement. Ces modèles sont aujourd'hui basés sur des biomarqueurs fonctionnels, ainsi qu'un biomarqueur morphologique, le ratio de la taille entre le ventricule droit et le ventricule gauche (VD/VG). Ce ratio est mesuré manuellement par des experts sur des examen CTPA (Computed Tomography Pulmonary Angiography) performés dans 90% des cas. Ce rapport présente différentes méthodes de segmentation CTPA de ventricules pour calculer automatiquement le ratio volumique des ventricules et voir s'il corrèle mieux au risque de décès des patients que le ratio de taille.

Keywords
pulmonary embolism, risk stratification, segmentation, CTPA, ventricle

Thanks
I would like to thank Élodie Puybareau and Odyssée Merveille for their supervision. I also extend my gratitude to Allan Serva and Morgane des Ligneris for giving me advice on my segmentation process, and for taking the time to correct and analyze them. And I am very grateful to Odyssée Merveille and Cyril Barbel for taking the time to review this report.

Laboratoire de Recherche de l'EPITA
14-16, rue Voltaire
94270 Le Kremlin-Bicêtre CEDEX
France

Copying this document

Contents

1 Introduction **4**

2 State of the Art **7**
2.1 CNN, FCN, and U-Net . . . 7
2.2 Loss functions . . . 9
2.3 Segmentation Methods . . . 10

3 Configurable Vascular Segmentation **12**
3.1 Pre-existing code . . . 12
3.2 Our contribution . . . 13
3.2.1 Refactoring with PyTorch Lightning and other additions . . . 13
3.2.2 Configuration management with Hydra . . . 13
3.2.3 Documentation with Sphinx . . . 13

4 Dataset **14**
4.1 Project Dataset . . . 14
4.2 Resampling . . . 14
4.3 Cropping . . . 16

5 Segmentations **17**
5.1 Gain in expertise . . . 17
5.2 Ventricles' volume limits . . . 18
5.3 Semi-manual segmentation . . . 20
5.3.1 Abdominal window . . . 20
5.3.2 Segmentation tools . . . 21
5.3.3 Manual segmentation issues . . . 22
5.4 Segmentation Models . . . 22
5.4.1 TotalSegmentator . . . 22
5.4.2 U-Net3D . . . 22

6 Evaluation **24**
6.1 Inference duration . . . 24
6.2 Results . . . 24
6.3 Statistical Analysis . . . 26

7 Conclusion and Future Work **29**
7.1 Conclusion . . . 29
7.2 Future Work . . . 30

Chapter 1

Introduction

Pulmonary embolism (PE) is the third leading cause of death in Europe. It is characterized by a blood clot called thrombus that blocks the pulmonary arteries and decreases blood flow which can cause systemic hypotension, as shown in Fig. 1.1. It usually takes a large thrombus to cause hypotension, but if a patient already has pre-existing cardiopulmonary disease, a small thrombus can also cause hypotension.

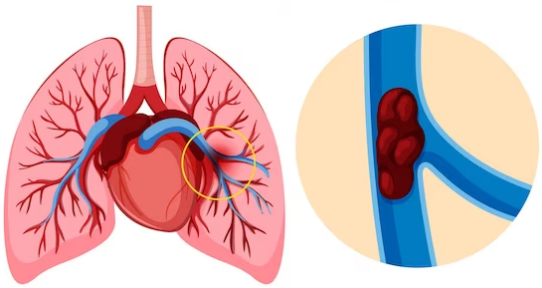


Figure 1.1: Pulmonary embolism (https://fr.freepik.com/photos-vecteurs-libre/embolie)

To quantify the severity of patients' PE, current risk stratification models classify a patient's risk of death after 30 days into three categories defined by the European Society of Cardiology: low, intermediate and high. The result determines the patient's management protocol, and therefore their treatment, such as anticoagulation, thrombolysis, thrombectomy, or mechanical cardiopulmonary support [1].

Today, these models are based on functional biomarkers, which correspond to the protein levels in the blood linked with heart failure; as well as a morphological biomarker, the right-to-left ventricle (RV/LV) diameter ratio.

Our work is part of a project funded by the ANR Young Researchers program (ANR JCJC) called PERSEVERE, which is conducted by INSA Lyon and coordinated by Odyssée Merveille. Furthermore, Élodie Puybareau is a collaborator on this project. The goal of this project is to propose novel risk stratification models for pulmonary embolism directly related to the routinely performed CTPA (Computed Tomography Pulmonary Angiography), as illustrated in Fig. 1.2 [2].

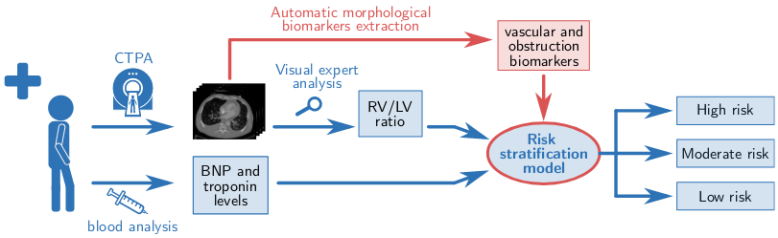


Figure 1.2: Current PE patient prognosis evaluation procedure in blue, and the modification proposed in the PERSEVERE project in red. [2]

Our research scope will be focused on the RV/LV ratio.

An early study showed that right ventricular enlargement on chest CT predicted early death in patients with acute PE. They used a RV/LV ratio of 0.9 to define right ventricle enlargement [3]. A following study showed that an increased RV/LV ratio was associated with an increased risk of mortality of approximately 2.5 [4]. By using the smaller ratio of 0.9 as threshold, the sensibility is going to be higher to minimize false negatives. And by using the higher ratio of 1.0, the specificity is going to be higher to minimize false positives.

If the RV is dilated, it means that the heart must have pumped harder to oxygenate the same amount of blood, resulting in the RV becoming more muscular. The cause could be a PE that prevents part of the lungs from oxygenating the blood.

The gold standard to assess the RV/LV ratio is the transthoracic echocardiography as it is synchronised with the patients' cardiac cycles. However, its access in the initial stage of PE is limited [5]. Therefore, the RV/LV diameter ratio is measured on CTPA exams in over 90% of PE cases [2].

CTPA is a CT scan which consists of injecting a contrast dye to enhance the visibility of the pulmonary arteries using X-rays. It provides a more comprehensive assessment of the severity of the pulmonary artery's obstruction compared to echocardiography [6]. Other advantages of this modality are the ability to assess other cardiopulmonary conditions, its availability in the vast majority of small and large hospitals and its non-invasiveness. However, it has contraindications due to the dangerous radiation, especially for pregnant and younger patients, but also due to the risk of allergy to the contrast product [5]. It is important to mention that for very high risk PE patients, CTPA exams are too dangerous to perform. Therefore, datasets for PE research are limited to the scope of low, intermediate and high risk patients.

To measure the ventricles' diameters, radiologists search for the slices where each ventricle has the largest diameter, delimited by the endocardium, as shown in Fig. 1.3. However, recent studies showed that since CTPA is not synchronised with the heart rate, it blurs the images and leads to unreliable measurements that are poorly correlated with the patient's prognosis [2]. Therefore, it is ideal to measure both ventricles' diameters on the same slice because different slices could correspond to a different phase of the cardiac cycle.

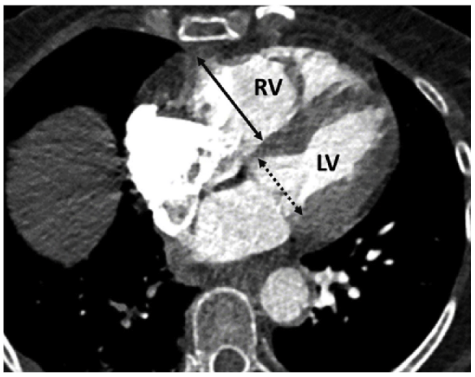


Figure 1.3: RV/LV diameter measurements on CTPA [6]

To respond to current issues regarding the ventricles' diameter measurements, we propose to use the volumetric ventricles ratio.

Our strong hypothesis is that the volumetric ratio of ventricle correlates more to the risk of death of PE patients, since it removes bias caused by CTPA's lack of synchronisation with the heart rate. To accurately measure the volumetric ratio, we need segmentations of PE patients' ventricles. However, there is currently no annotated CTPA dataset of ventricles because most of the research on PE is focused on segmentations of thrombus and pulmonary vascular tree (PVT). Some automatic model to segment ventricles exists, however they are mainly trained on other acquisition modalities and do not perform as well on CTPA [7] [8].

Our objective is to develop an automatic segmentation method to compute the volumetric ratio of ventricles, and verify our hypothesis on PERSEVERE's dataset. Therefore, segmentations will be needed to supervise the training of our models.

Our main contributions to the PERSEVERE projects are **1.** the ventricles' manual segmentations and automatic segmentations, and **2.** the statistical analysis of the correlation between the volumetric ventricles ratio and severity of patients' PE. We have also worked with other collaborators of the project on the vascular segmentation task and PERSEVERE's dataset. Therefore, we **3.** analysed and preprocessed the dataset, and **4.** refactored a code base aimed at running configurable training and inferences for vascular segmentation.

With our segmentations, we observed that PE patients at high risk of death have a significantly higher volumetric ventricles ratio compared to PE patients with intermediate or low risk of death.

In the first part, this report presents our bibliographic work. Then, the report explains the refactoring of a repository used to easily configure vascular segmentation training and inference of different models and datasets, as well as the addition of multi-class segmentation and a script to optimise hyperparameters. After that, the analysis and preprocessing of the dataset we used is explained. Then, the report focuses on the pipeline to manually segment CTPA ventricles with 3DSlicer, as well as the segmentation models we tried. Finally, the evaluation of the segmentation models will be discussed, as well as the analysis of the distribution of ventricles volumes and their ratio depending on the severity of the embolism.

Chapter 2

State of the Art

A segmentation consists of classifying every voxel of an image as either belonging to the background or a region of interest. Segmentations of medical images are very important for clinical practice since they facilitate accurate diagnosis, treatment planning, and disease monitoring, but they are often performed manually by experts. This process is a very time-consuming task, requires anatomical knowledge and is often error-prone due to the lack of experience and eye fatigue [2].

For cardiac image segmentation, traditional machine learning and atlas-based methods were used, but they often required important feature engineering or prior knowledge to achieve satisfactory accuracy [9] [10]. Whereas deep learning based algorithms automatically learn those features from data. Additionally, advancement in computer hardware and increased available data for training encouraged its development and explains deep learning-based method out performance over previous state-of-the-art traditional methods [11]. Automatic methods have become widely used for medical image segmentation but validation of experts is still necessary.

2.1 CNN, FCN, and U-Net

Convolutional Neural Networks (CNN) are the most common type of deep neural networks for image segmentation, classification and detection tasks. They are composed of an input layer, an output layer and a stack of convolutional layers, pooling layers and / or fully-connected layers in between. The main application of CNNs with connected layers for cardiac segmentation is object localization to estimate the bounding box of the object of interest in an image. The bounding box is then used to crop the image to reduce the computational cost for segmentation, as shown in Fig. 2.1 [11].

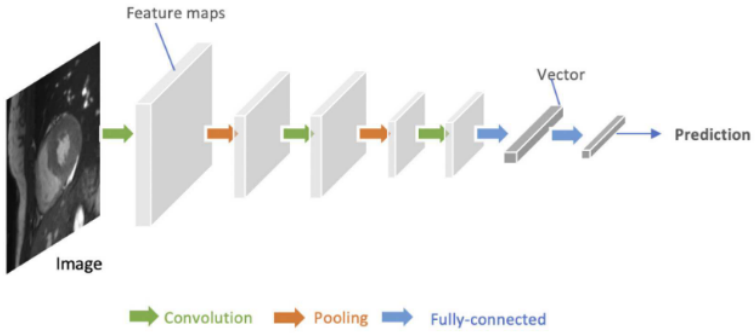


Figure 2.1: Generic architecture of convolutional neural networks (CNN) [11]

Fully convolutional neural networks (FCN) [12] are a type of CNN composed of an encoder and a decoder, they can take an input of arbitrary size and produce the output with the same size. The encoder transforms the input into a high-level feature representation, and the decoder interprets the feature maps and recovers spatial detail in the image-space for pixel-wise prediction using up-sampling and convolutions. FCN can be limited to capturing detailed context information in an image for precise segmentation because features may be removed by the pooling layers in the encoder, as illustrated in Fig. 2.2 [11].

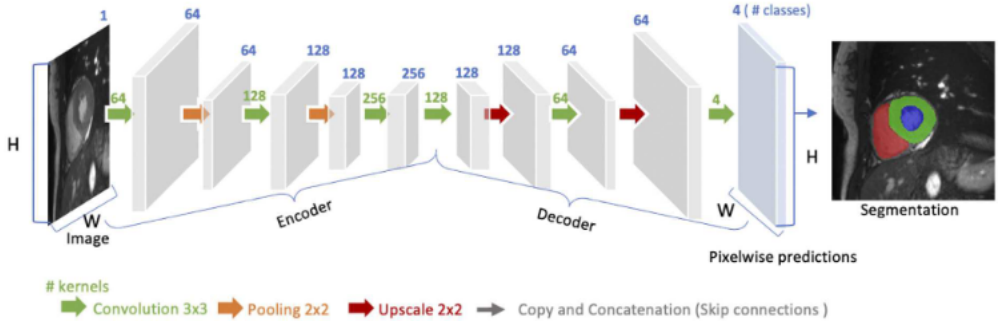


Figure 2.2: Generic architecture of fully convolutional neural networks (FCN) [11]

U-Net [13] is a variant of FCNs, that uses skip connections between the encoder and decoder to recover spatial context lost during the down-sampling to increase the segmentation accuracy, as shown in Fig. 2.3. It is the reference model for medical image segmentation and often serves as a baseline for benchmarks. There are numerous variants, such as UNet 3D using 3D convolutional layers for encoding and decoding to keep the spatial context of each patch across the volume [14]; or the nn-UNet which automatically optimizes the architecture and hyperparameters of the model for the given dataset [15].

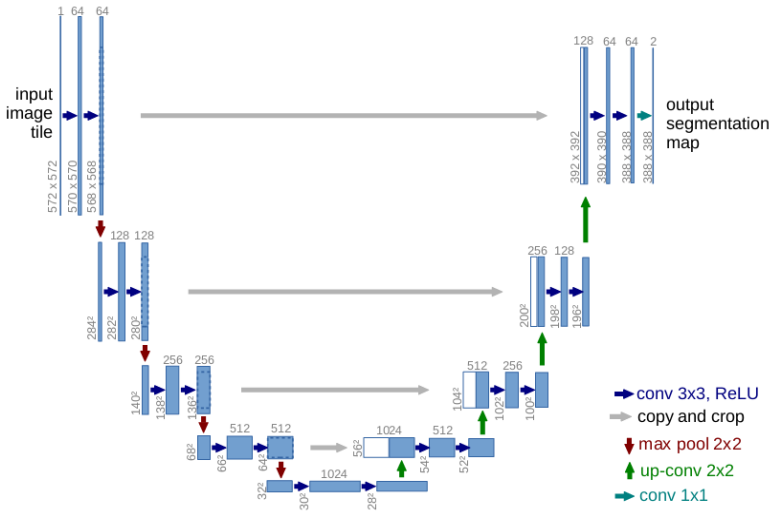


Figure 2.3: Architecture of U-Net [13]

2.2 Loss functions

Loss functions are used to measure the error between the predicted output of a model and the ground truth. Then, the gradient of the loss for each parameter of the model is computed to update its weights and biases using a optimisation algorithm.

Cross-entropy is the most common loss for image segmentation tasks. It summarizes pixel-wise probability errors between a predicted probabilistic output p_i^c and its corresponding target segmentation map y_i^c for each class c [11]:

$$\mathcal{L}_{CE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_i^c \log\left(p_i^c\right)$$

where C is the number of all classes.

A specific problem in medical image segmentation is the class imbalance. The number of pixels belonging to the object of interest is extremely small compared to the number of background pixels. This causes bias for the loss function because the accuracy score can be high just by predicting background pixels correctly. There are loss functions that ignore the false negatives, such as the Dice or its continuous variant, the soft-Dice.

The Dice loss function penalizes the mismatch between a predicted segmentation map and its target map at pixel-level [11]:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i=1}^{n}\sum_{c=1}^{C} y_i^c p_i^c}{\sum_{i=1}^{n}\sum_{c=1}^{C}(y_i^c + p_i^c)}$$

Other loss functions are designed to focus on the classes of interest, such as the weighted cross-entropy loss and weighted soft-Dice loss.

2.3 Segmentation Methods

Deep learning-based methods are often specialized to segment specific organs or tissues in particular modalities, so they lack generalisability, and there is currently no specialized method to segment ventricles on CTPA images. This section presents different types of deep learning methods to segment ventricles of CT images.

Two-step segmentation methods first extract a bounding box delimiting the regions of interest and then feed it into a CNN for segmentation, as illustrated in Fig. 2.4 [16]. This reduces the class imbalance problem because there are less background pixels, and facilitate the segmentation of regions of interest.

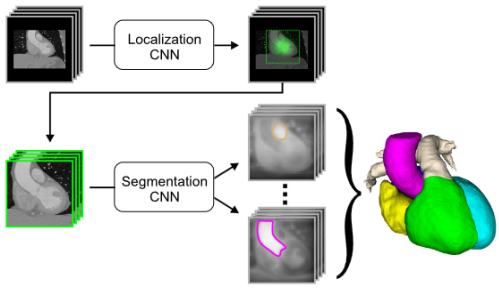


Figure 2.4: Two-step multi-label segmentation pipeline [16]

The main problem of most 2D segmentation methods is that they segment each slice independently without using its context in the image. This can create artefacts or incoherences when they are merged together during the volumetric reconstruction. Multi-view CNNs palliate this issue by training multi-planar CNNs using axial, sagittal and coronal views separately and then combining them, as shown in Fig. 2.5. This allows the model to use the context of each pixel during the segmentation [17] [18] [19].

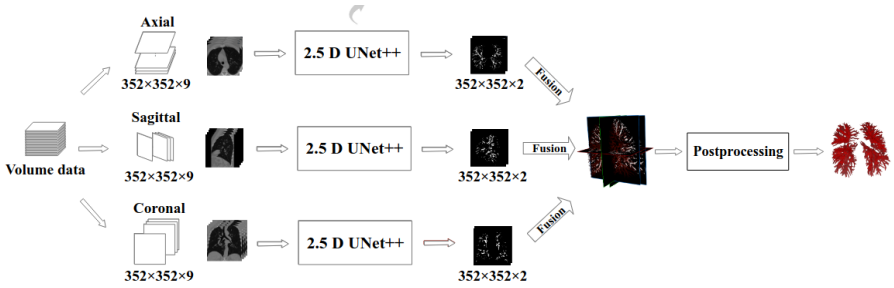


Figure 2.5: Multi-view segmentation pipeline [19]

Another way to address the class imbalance issue is to use hybrid loss, meaning combining different loss functions to improve the segmentation performance, as shown in Fig. 2.6 [20] [21].

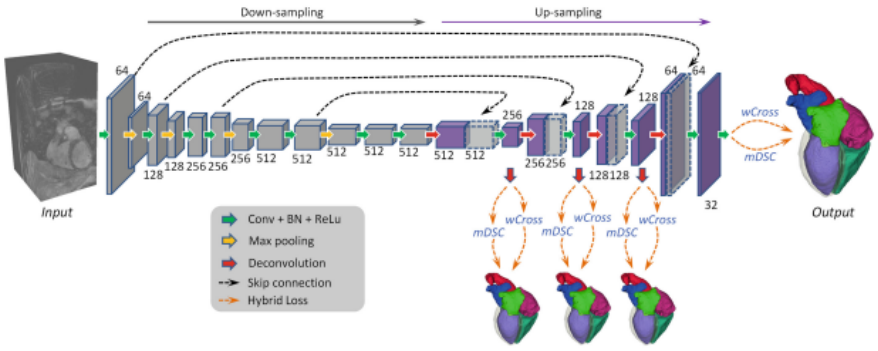
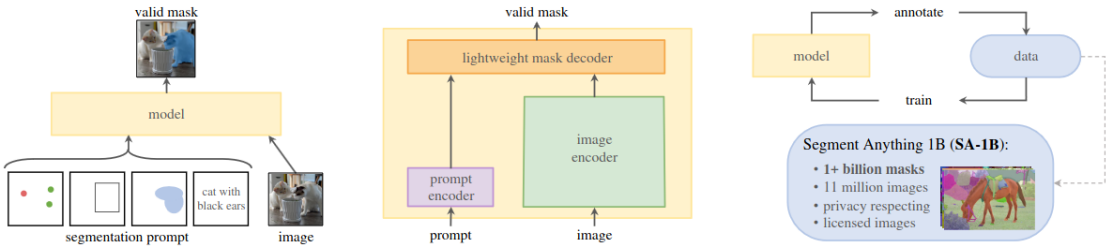


Figure 2.6: Hybrid loss segmentation pipeline [20]

Most models are trained for a specific task, to classify a limited number of classes, or to segmented certain anatomic structure, and on specific acquisition modalities. As such, most models lack generalisation and need to be fine-tuned for a new task. Foundation models are pre-trained on massive datasets and are adapted to an extensive number of tasks. They improve contextual reasoning, generalisation, and prompt performances at test time [22]. They can also be fine-tuned for specific tasks by feeding more data specific to the task. Following the same principle as two-step segmentation methods, which use a first network to localise the region of interest and then do the segmentation, foundation models can be guided by user-prompt as seed point or bounding-box, as illustrated in Fig. 2.7 [23]. Some foundations models are fine-tuned to enable universal medical images segmentation to improve generalisability across imaging modalities, organs and diseases [7].



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Figure 2.7: Segment Anything Model (SAM) Architecture [23]

Chapter 3

Configurable Vascular Segmentation

The first task was to refactor a code repository belonging to the laboratory of the INSA Lyon called CREATIS Laboratory (Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé), which aims to facilitate vascular segmentation experiments using different model architectures, hyperparameters and datasets. This code repository is notably used for the experiments of automatic vascular and thrombus segmentation models in the context of the PERSEVERE project.

3.1 Pre-existing code

The code uses **PyTorch**, a popular open-source framework to build and train deep learning segmentation models. It also utilises **MONAI** (Medical Open Network for Artificial Intelligence), an open-source project built on top of PyTorch that standardises AI development for healthcare research with a focus on medical imaging. The code uses MONAI's convolutions layers, transforms and dataset classes.

For data augmentation, the **batchgenerators** library was employed. Some of the augmentations used include brightness multiplicative transform, contrast augmentation transforms, gamma transformation, Gaussian noise transform, Gaussian blur transform, spatial transform, and mirror transform.

To track experiment performance, the code used **Weights and Biases**, a MLOps platform for AI developers that provides tools to train and fine-tune models. It can be used to keep track of each experiment's model architecture, hyperparameters configuration, logs of training, validation, testing loss, and accuracy.

The repository contained multiple configurable PyTorch modules such as encoder, decoder, auto encoder and U-Net [13]. There were Python scripts that had the same pipeline but with different model architecture, hyperparameters, data pre-processing, data augmentation, datasets, loss functions, or evaluation metrics. There were also scripts to automatically execute the inference of a dataset using a pre-trained model checkpoint. The inference was done using MONAI's sliding windows function and transforms. Additionally, there was an option to flip the input, compute the inference for each image configuration and to combine them by calculating the mean of the prediction to have a more accurate segmentation.

The main problems we aimed to fix were the code duplication, the non-intuitive and impractical configuration of the parameters for training and inference, the lack of documentation, and the lack of standardisation of the different modules.

3.2 Our contribution

3.2.1 Refactoring with PyTorch Lightning and other additions

PyTorch Lighting is an open-source deep learning framework that aims to standardise the structure of PyTorch projects and make the codebase cleaner, more readable, and maintainable by reducing the boilerplate code and removing unnecessary abstractions on top of PyTorch. It also simplifies the training loop and other aspects of deep learning model training, such as the automatic handling of GPU setup and optimization. The library is used in other projects developed in CREATIS. The first task was to refactor the current code base using PyTorch Lightning. It implied creating LightningModule wrappers for existing PyTorch modules and replacing the training loop with a PyTorch Lightning Trainer that contains the boilerplate training, validation, and testing logic. PyTorch Lightning Logger, which uses the Weights and Biases wrapper, has been implemented to log the experiments.

Since we want two different labels for the ventricles, we added an option for multi-class segmentation model. This addition did not require changes of previous configuration files since it is activated based on the number of output channels of the model used. Furthermore, we added a script to automatically optimize hyperparameters given a variable to minimize or maximize, as well as a range or list of options for the concerned hyperparameters. This script uses the open source framework **Optuna** that we used in a practical work this semester.

3.2.2 Configuration management with Hydra

O. Merveille proposed to use **Hydra** to manage the different script configurations. It is an open-source framework that enables hierarchical configuration composed of multiple files for each module, removes boilerplate for handling configuration or command-line flags, and facilitates running jobs with different parameters with a single command. It clearly exposes the configuration options and allows dynamically overriding an existing value, appending a value or removing an existing value.

Configurations for training and inference were created and implemented in the scripts. The configuration modules are written in YAML. The final configuration hierarchy of the project includes data splitting rules, dataset information, inference parameters, Weights and Biases loggers information, model architecture, transforms and data augmentation for training and inference, and inference and training parameters.

3.2.3 Documentation with Sphinx

Documentation, written with **Sphinx**, was added to the repository. It includes a project presentation, setup guide, configuration guide for Hydra, and training and inference guides.

Chapter 4

Dataset

Since the PERSEVERE dataset does not have ground truth of ventricles, we searched for a CTPA dataset with ventricles annotations. However, the research on PE risk stratification is focused on the segmentation of PE and PVT, and there is no public dataset of segmented ventricles of CTPA. We found a dataset of CTPA, containing classification labels for presence or absence of a thrombus, and the RV/LV size ratio for each patient. However, the objective is to automatically segment the ventricles of CTPA but there was no ground truth for ventricles segmentation. Therefore, this dataset could not be used for supervised training, but it is still an interesting dataset to keep [24].

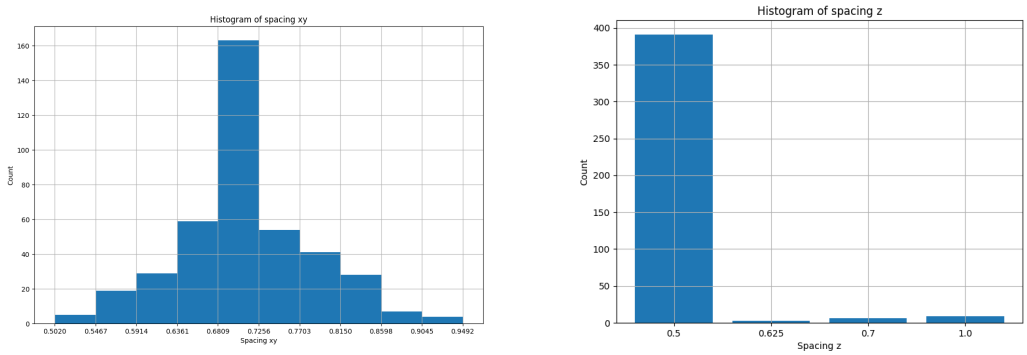
4.1 Project Dataset

Eventually, the dataset of PERSEVERE was chosen to test the different segmentation models. It is stored in a **Girder**, an open-source web-based data management platform developed by **Kitware**. It contains the CTPA of 431 patients of multiple PE risk categories that are cropped using a mask of the lungs and padded by ten voxels on all sides. And they are stored in **NIfTI** (Neuroimaging Informatics Technology Initiative) format.

4.2 Resampling

We wrote a script that retrieved, and stored the dimension and voxel's spacing of every image in the dataset from their headers. It revealed that the images have different sizes, resolutions and spacing. This was expected since the data are acquired using multiple scanners with different parameters.

The images had different voxel spacing, therefore a voxel represented different volumes, so we needed to homogenise the dataset. To find the most represented spacing in dataset and resample the other images, we wrote a script to compute the histogram of the spacing of the image in the dataset with the number of bins as the parameter, as illustrated in **Fig. 4.1**. The figure was displayed and saved in order to document the Girder.



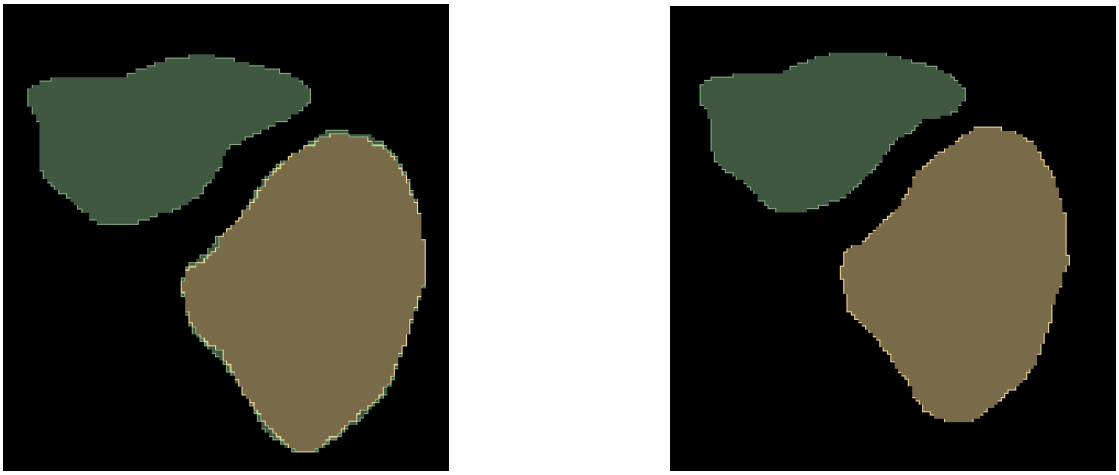
(a) Spacing XY

(b) Spacing Z

Figure 4.1: Histogram of the dataset spacing

The resampling used the **SimpleITK** library.

For the ground truth, we first tried to use trilinear interpolation, which computes the weighted average of the six nearest neighbour pixel values at non-discrete coordinates, as shown in **Fig. 4.2a**. However, there is a different class for each ventricle represented as a discrete value. Therefore, the resampled ground truth of the LV had its border incorrectly labeled as RV. To solve this issue, nearest neighbour interpolation was used for the segmentation, as shown in **Fig. 4.2b**.



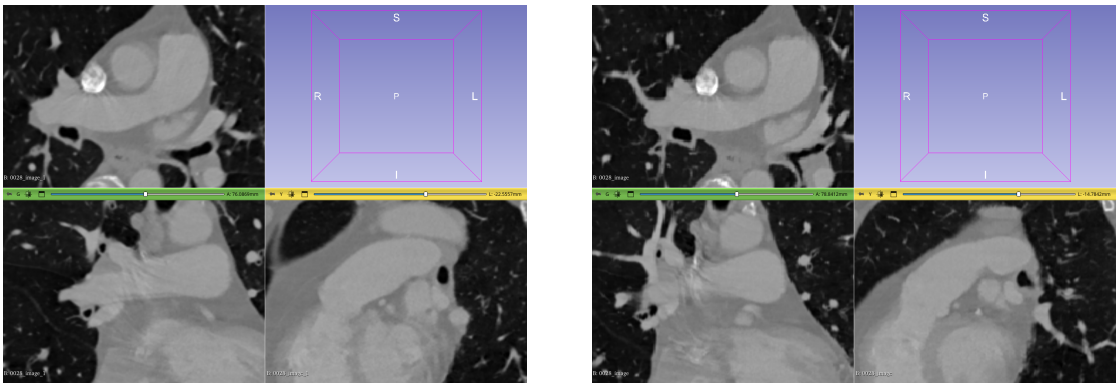
(a) Trilinear interpolation

(b) Nearest neighbour interpolation

Figure 4.2: Results of the ventricles' segmentation resampling

It was important to verify that the interpolation did not alter the images or the ground truth. Specifically, the vascular structures connectivity were carefully verified since they are the thinnest elements, and it is important to keep their connectivity. Using nearest neighbour interpolation degraded the images, so they were resampled using trilinear interpolation , as

shown in Fig. 4.3b and Fig. 4.3a.



(a) Trilinear interpolation

(b) Nearest neighbour interpolation

Figure 4.3: Results of the image resampling

4.3 Cropping

Since our regions of interest are the ventricles, the arteries, and the embolisms, the images were cropped using lungs mask computed using TotalSegmentator [25] Fig. 4.4. However, our segmentations were made on already cropped data using other coordinates. Thankfully, the headers of the new resampled and cropped images was still there, which allowed us to retrieve valid segmentations with the right shape and voxel spacings. This showed that it is very useful to keep the header information of Nifti images.

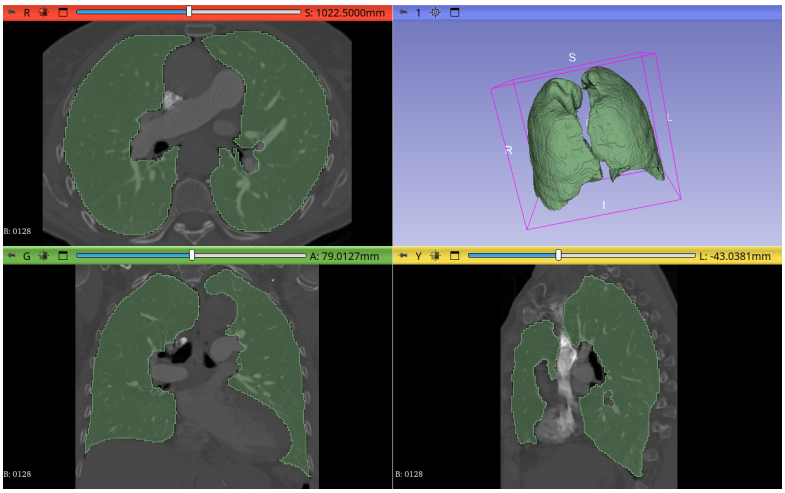


Figure 4.4: TotalSegmentator's [25] lung mask

Chapter 5

Segmentations

In order to start supervised training of our model, ground truths of the ventricles are needed. Allan Serva, the project's radiologist, did not have time to do the segmentation himself. Therefore, the segmentation task had to be done ourselves, and they will then be reviewed and corrected. This is important as those segmentations will be used to train automatic ventricles segmentation model that will serve for medical diagnosis.

Since the segmentation software used by A. Serva is licensed, an open-source replacement had to be found. The standard open-source software for visualization, segmentation, and analysis of medical images, **3DSlicer**, was chosen.

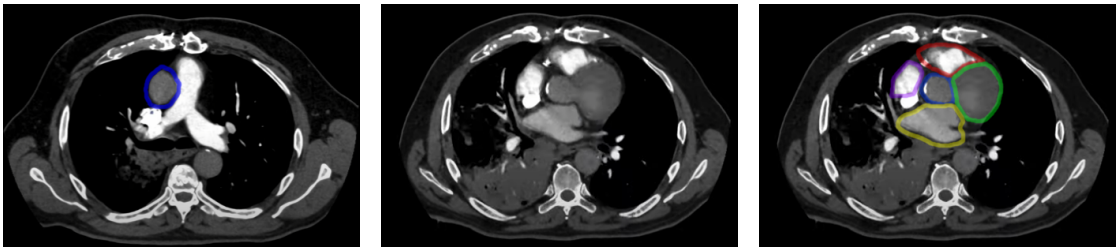
However, the lack of anatomical and CTPA knowledge was problematic, as we needed to know the position of the ventricles in the volume. Furthermore, it was important that we could evaluate the segmentations we had done and later the automatic segmentations.

5.1 Gain in expertise

To gain the necessary anatomical and CTPA knowledge needed to localise the ventricles' volumes, RV/LV ratio related articles were read **[6]** **[3]** **[4]**, and CT, CTPA and PE related videos were watched [1] [2] [3] [4] [5].

We were then able to localise the ventricles in the volume. The technique used was to start from the top of the patient, follow the ascending aorta and search for the connection to the left ventricle. From there, it is simple to recognize the ventricles and the atria, as illustrated in **Fig. 5.1**.

[1] https://youtu.be/8WUgH4WHlLE?si=uk5pXg1hqpcg8VJl&t=1111
[2] https://youtu.be/yeNd_nhiQ6U?si=xgSMQvEfLg20B-XV
[3] https://youtu.be/Ibc_ryTwX40?si=S7OxcuyMcsNtlaFv
[4] https://youtu.be/ciCYQWm0VTE?si=FpOoV28lYBOjZZ0i
[5] https://youtu.be/QHYOuudasF0?si=EAIzQa2UsBCeSvEH



(a) blue: ascending aorta

(b) Slice where the ascending aorta is connected to the LV

(c) blue: ascending aorta ; green: LV ; red: RV; yellow: left atrium; purple: right atrium

Figure 5.1: Localisation of the ventricles (https://youtu.be/8WUgH4WHlLE?si=uk5pXg1hqpcg8VJl&t=1111)

However, a crucial information was missing from those resources, the ventricles' volume limits.

5.2 Ventricles' volume limits

This section defines the limits instructed by A. Serva, and the characteristics of the ventricles.

The ventricles consist of two distinct layers : the myocardium, which is the middle layer, and the endocardium, which is the innermost layer, as illustrated in Fig. 5.2. However, the RV/LV ratio is measured based on the ventricles diameters delimited by the endocardium.

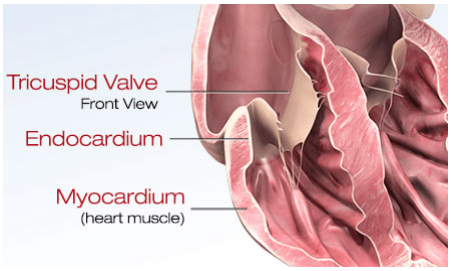
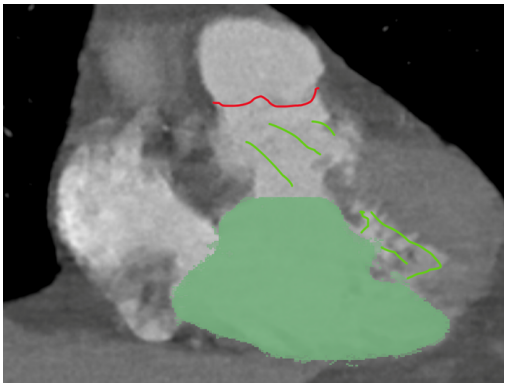


Figure 5.2: Heart wall layers (https://www.heart.org/en/health-topics/myocarditis)

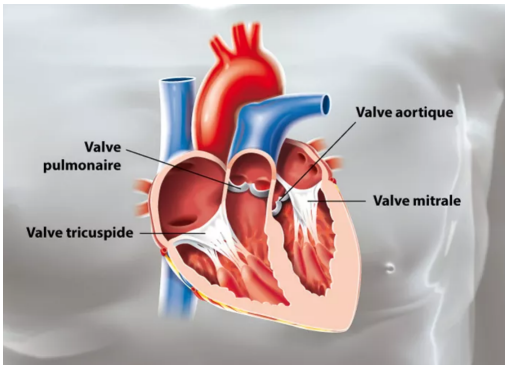
The thickness of the myocardium is correlated with the strength of the heart's ability to pump blood. It is thickest in the LV as it pumps oxygenated blood to the entire body, whereas the RV only pumps blood to the lungs. Therefore, the limits between the myocardium and the endocardium are easier to recognise on the LV, as the border between the two is more visible. In contrast, the separation between the myocardium and endocardium in the RV is more difficult to distinguish, as the myocardium is thinner.

An important point to consider for the volumetric segmentation is the upper limits of the ventricle, which were not taken into account for the diameter measurement. After the first segmentation review, the radiologist explained that the upper limits were the pulmonary valves,

as shown in Fig. 5.3.



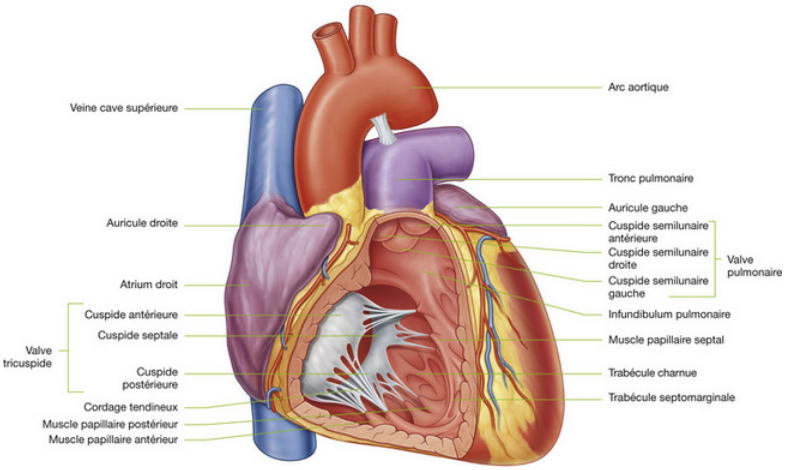
(a) LV segmentation and markers for upper limit



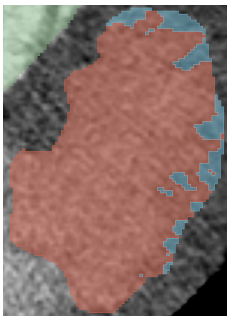
(b) Pulmonary valves (https://www.ramsaysante.fr/vous-etes-patient-en-savoir-plus-sur-ma-pathologie/valves-cardiaques)

Figure 5.3: Segmentation correction : LV upper limit

During a later review, A. Serva explained that we should include papillary muscles, represented in Fig. 5.4a, in the ventricles' segmentations even if they are part of the wall and not of the ventricular cavity itself. In methods considered as the gold standard (TTE MRI), the papillary muscles are often included in the volume calculation, primarily for practical reasons and standardization. Not including the papillary muscles would reduce the LV volume, potentially introducing a systematic bias in the measurement of the RV/LV ratio, as suggested by Fig. 5.4b.



(a) Diagram of the anterior surface of the heart (https://clemedicine.com/3-thorax/)



(b) Ventricle segmentation - red: LV ; blue: papillary muscles

Figure 5.4: Addition of papillary muscle to ventricles segmentations

For the segmentations, the instructions were to overestimate rather than underestimate the

ventricles' limits in order to measure the largest volume.

And now that the anatomical and CTPA knowlegde were acquired, and that the ventricles' volume limits defined, the segmentation process could begin.

5.3 Semi-manual segmentation

A. Serva made a list of ten PE patient with different severity. We performed their segmentation in order to use them as ground truth to train our deep learning model. Since the patients on the list were changed, we had thirteen segmentations of ventricles.

To learn how to segment ventricles on 3DSlicer, tutorial videos were watched [6] [7] [8] [9]. This section describes the pipeline we followed for each segmentation, as well as advice given by A. Serva.

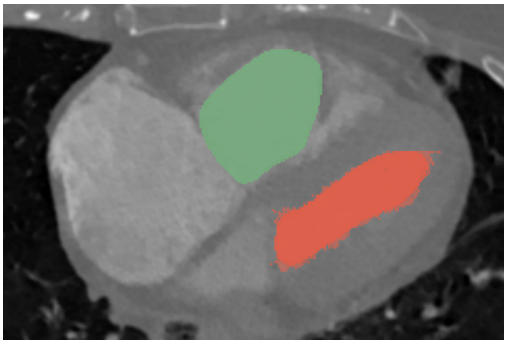
5.3.1 Abdominal window

Windows are used to highlight different tissues or organs by adjusting the contrast and brightness of the image. There are different types of windows for most anatomical structures or tissues based on typical Hounsfield units (HU) for specific tissues.

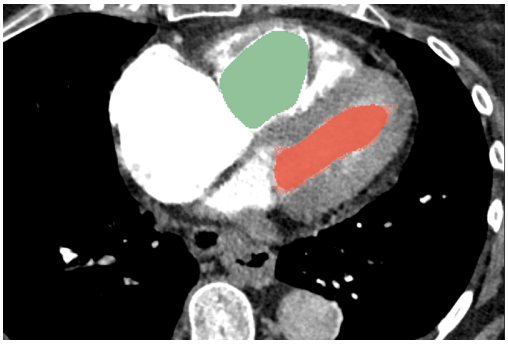
HU are the units for the CT scans, they measure the radio intensity of different tissues in the body based on their density.

The CTPA contrast is not easily noticeable with the default volume rendering, which made the limits of the ventricles difficult to see for some patients. After the first segmentation review, A. Serva advised us to use the media intestinal window, as the contrast product exceeds the segmentations of both ventricles. The limits of the ventricles were underestimated, as shown in Fig. 5.5.

[6] https://youtu.be/BJoIexIvtGo?si=iwHpgTEqLTli3bRh
[7] https://youtu.be/xIts_5fctYg?si=mmLgStvT9YjwmtkL
[8] https://youtu.be/9iiOBmaP8bA?si=9hkYXOZRcGk486Bc
[9] https://youtu.be/aeOFl19fh_c?si=TkfNvhT1Rc6ZrwsS



(a) Segmentation on default window

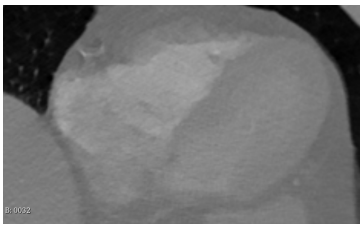


(b) Same segmentation on media intestinal window

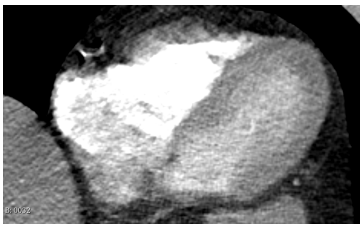
Figure 5.5: Same segmentation on different windows

However, the media intestinal window is not a preset available in 3DSlicer. A. Serva advised to used, the abdominal window preset instead, and it greatly facilitated the segmentation process.

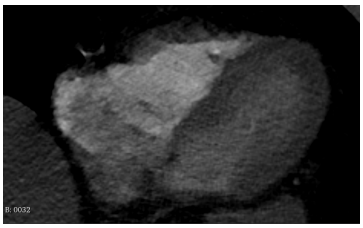
For most patients, the preset abdominal window did not help to see the limits of the ventricles, so for those, the window was manually adapted, as illustrated in Fig. 5.6.



(a) Default window



(b) Abdominal window preset



(c) Manually adapted window

Figure 5.6: Different 3DSlicer windows used for segmentation

5.3.2 Segmentation tools

In the segment editor of 3DSlicer, there is a paint tool to directly draw the segmentation on either the axial, sagittal, or coronal view. However, it is easy to draw outside the limit of the ventricles or to not paint enough, even with adapted window. To facilitate the painting task, we used threshold tool to limit the range of HU that can be painted on.

To avoid painting both ventricles on each view manually, the growing regions tool was used. Seeds were drawn for each ventricle and more seeds were added to delimit the background. The algorithm propagates the seeds in the neighbouring voxels and adds them to the corresponding class based on their intensity. However, it is sensitive to small intensity variations.

In post-processing, smoothing was applied to reduce the noise and have cleaner boundaries for the segmentations, holes in the segmentations were filled, and isolated voxels were removed.

5.3.3 Manual segmentation issues

This section describes manual segmentation known issues.

Manually annotating medical images is a very time-consuming task. It requires anatomical knowledge and is often error-prone due to the lack of experience and eye fatigue [2]. Additionally, there is also the problems of inter-observator variability and intra-observator variability. Two experts will not do the exact same segmentations and even one expert will not do the same segmentation twice.

In contrast, deep learning models perform fast and consistent segmentations that can be used to have a basis that can be refined manually.

5.4 Segmentation Models

Automatic segmentation models are often used to facilitate the annotating process. This section describes the different models tested to segment cardiac ventricles.

5.4.1 TotalSegmentator

TotalSegmentator is a robust deep learning model publicly available, easy to use, and is able to segment most anatomically relevant structures throughout the body. It was trained on a dataset composed of 1204 CT from the University Hospital Basel [25].

We used their Python API with the task 'heartchamber_highres' since we need the segmentations of the cardiac ventricles. The method used is a two-step multi-label segmentation pipeline. A first model computes a rough segmentation of the lungs in order to crop the original image. Then a second specialised model segments the different parts of the heart. Both models are nnU-nets [15].

The results are saved in different files, so we combined the two masks corresponding to the ventricles, while paying attention to using different labels for each ventricle.

5.4.2 U-Net3D

The U-Net is the model of reference for biomedical image segmentation tasks, and is often used as a benchmark's baseline [13].

This model has a lot of advantages. It learns low-level features using the high-level global context, using the contracting path to capture the context and the expansive path to localize the regions of interest precisely. The skip connections retrieve the spatial information lost during the down-sampling. And it is efficient in all medical image modalities [13].

We used the U-Net3D version because we want to use 3D contextual information to have a more coherent segmentation.
We split our data such that 8 patients were on the train set, 2 patients were on the validation set, and 3 patients were on the test set.

For the hyperparameters, since we use a U-Net3D, we used a batch size of 2, and a patch size of [192,192,64]. For the optimiser, we used Adam with a learning rate of 0.01, and a linear scheduler. For the loss function, we selected the sum of the Dice Loss and the Cross Entropy Loss [11]. Compared to the training of a U-Net2D, the number of epochs necessary to have satisfactory results is far greater. We trained our model for six-hundred epochs, which took approximately five days to finish.

For the inference, we used sliding windows with ROI size of [192, 192, 64], an overlap between scans along each spatial dimension of 0.25, and a Gaussian blending.

Chapter 6

Evaluation

In total, we performed thirteen segmentations of ventricles' volume semi-manually, and we used TotalSegmentator to compute ventricles segmentations of all the patients in the dataset.

During our latest review, A. Serva said that the contours of our segmentations were better than those generated by TotalSegmentator for most of our segmentations, but there were still mistakes. As of the writing this report, no segmentations have been corrected. We have thus chosen to keep our manual segmentations as the ground truth to evaluate our automatic segmentations.

We first compare the segmentation duration of each model, then we evaluate the inference results quantitatively and qualitatively. Finally, we look at the statistical analysis of the ventricles intensities, and volumes, as well as their ratio.

6.1 Inference duration

The average manual segmentation duration is 6 hours. In comparison, the average segmentation duration of TotalSegmentator is 1 minute and 46 seconds. And the average segmentation duration of the U-Net3D we trained is 1 minute and 21 seconds.

6.2 Results

In this section, we first introduce the different metrics we used to evaluate the segmentations of our models quantitatively and qualitatively.

The Dice Similarity Coefficient (DSC) is one of the most common overlap-based metrics. It calculates the spatial similarity between the segmented result and the ground truth based on their overlapping regions [26].

$$DSC = \frac{2 \cdot TP}{FP + FN + (2 \cdot TP)}$$

with TP: true positive; FP: false positive; FN: false negative.

The surface distance measures the shortest distance of each point on the surface of the segmentation result to the surface of the ground truth [26]. Whereas, the Average Symmetric Surface Distance (ASSD) is the average of all distances for every point from one object to the other and vice versa.

Given two finite point sets $A = \{a_1, ..., a_p\}$ and $B = \{b_1, ..., b_q\}$:

$$ASSD(A, B) = \frac{\sum_{a \in A} d(a, B) + \sum_{b \in B} d(A, b)}{|A| + |B|}$$

$$d(a, B) = \min_{b \in B}(a, b)$$

with $d(a, B)$ is the Euclidean distance between the point a and the points of B.

The Hausdorff distance (HD) is a dissimilarity measure that evaluates the maximum distance to which each point of the segmented result lies near some point of the ground truth [27] [28] and vice versa.

Given two finite point sets $A = \{a_1, ..., a_p\}$ and $B = \{b_1, ..., b_q\}$:

$$HD(A, B) = \max(h(A, B), h(B, A))$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||$$

and $||.||$ is a norm on the points of A and B.

HD is generally sensitive to outliers commonly present in medical segmentations. Thus, the 95^{th} percentile HD (HD95) is generally used [29].

$$95HD(A, B) = percentile(h(A, B), 95)$$

where $h(A, B)$ represents the set of all distances from points in set A to their nearest points in set B [26].

To verify the topology of the segmentations, we use Betti numbers. A Betti number β_k is determined for a dimension k, and refers to the number of k-dimensional holes on a topological surface. A k-dimensional hole is a k-dimensional cycle that is not a boundary of a $(k + 1)$-dimensional object. As such, β_0 represents the number of connected components, β_1 the number one-dimensional of holes, and β_2 the number of two-dimensional cavities [30]. The expected Betti numbers for each ventricle are : $\beta_0 = 1$, $\beta_1 = 0$, and $\beta_2 = 0$.

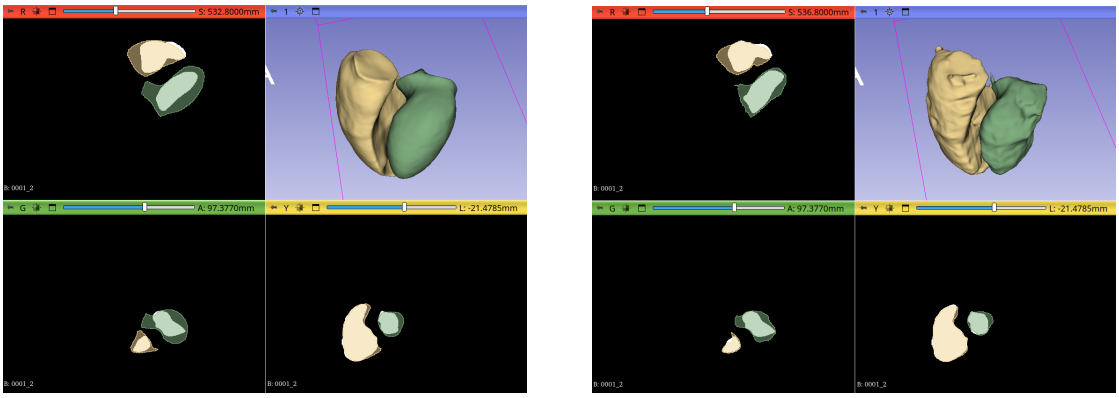
We computed the metrics on each ventricle separately to see if one had a better segmentation result than the other. The results are presented in Fig. 6.1.

		TotalSegmentator [25]	U-Net3D [14]
DSC ↑ [26]	RV	0.625	**0.853**
	LV	0.598	**0.775**
ASSD ↓ [26]	RV	13.448	**3.562**
	LV	15.065	**4.972**
95HD ↓ [29]	RV	34.443	**14.362**
	LV	42.355	**15.273**
β_0 [30]	RV	**1.0**	1.222
	LV	**1.154**	1.222
β_1 [30]	RV	**0.0**	7.556
	LV	**0.154**	4.889
β_2 [30]	RV	**0.0**	5.222
	LV	**0.0**	3.111

Table 6.1: Table of automatic segmentations results based on the 13 semi-manual segmentations

We observe that, on average, the RV segmentation are better than the LV segmentation for both models. This result is expected since the limits between the myocardium and the endocardium are easier to recognise on the LV.

The U-Net3D has a better DSC, ASSD, and 95HD than TotalSegmentator since it was trained on those images. Thus, the contours are closer to the ground truths. However, the U-Net3D has higher Betti numbers for both ventricles, showing that the segmentation is not smooth or full, so post-processing needs to be applied as can be seen in Fig. 6.1. We could use closing to fill the holes in the segmentations, joint smoothing to improve the smoothness, and we could remove the smaller connected components for each label.



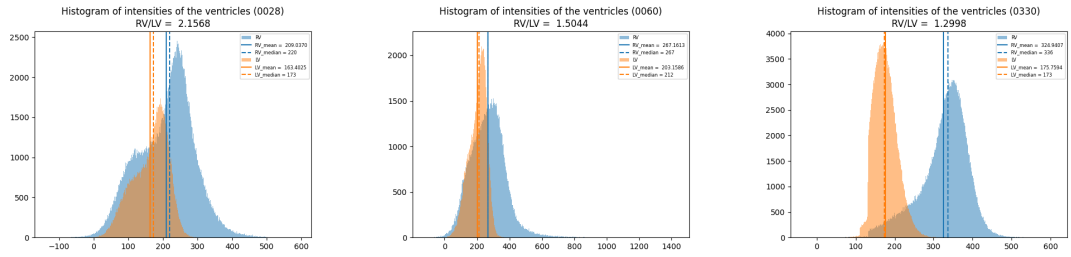
(a) TotalSegmentator [25] (b) U-Net3D [14]

Figure 6.1: Automatic 3D segmentations of cardiac ventricles

6.3 Statistical Analysis

During a meeting, A. Serva told us that since patients with severe PE's heart must pump harder to oxygenate the same amount of blood, the quantity of contrast product in the RV should be

higher than in the LV. So we analysed the histogram of intensities of each ventricle for the patients with a semi-manual segmentation. Some results are shown in Fig. 6.2. We observed that the intensity ranges vary for each patient independently of the sPESI, especially for the RV.



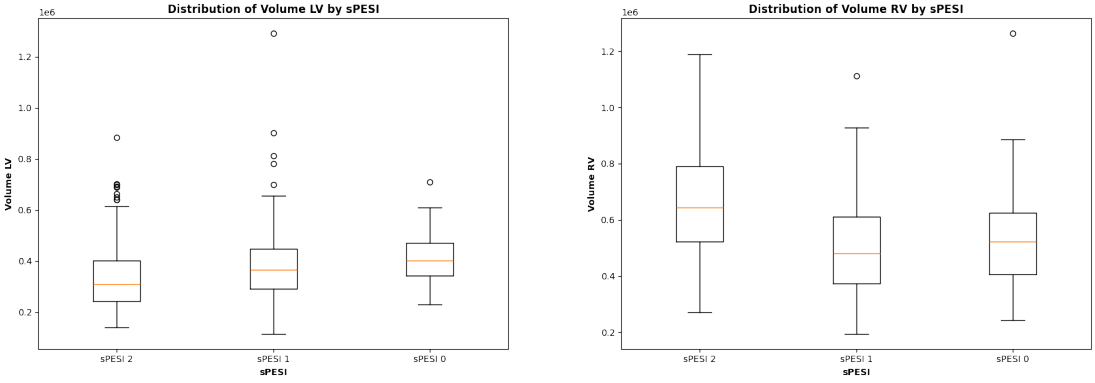
(a) sPESI 2 (b) sPESI 1 (c) sPESI 0

Figure 6.2: Histograms of intensities of the ventricles

We concluded that the intensity histograms depend too much on external factors, such as contrast agent washout, variation in the waiting time between the injection of the contrast agent and image acquisition, or partially obstructed veins slowing down the propagation of the contrast agent.

We also analysed the volume of the RV and the LV, as well as their ratio. Since we only have semi-manual segmentations of thirteen patients, we will use the segmentation made by TotalSegmentator for the remaining patients of the dataset.

We can observe the distribution of LV and RV volumes in Fig. 6.3, and the distribution of volumetric RV/LV ratio Fig. 6.4 based on the Simplified Pulmonary Embolism Severity Index (sPESI) of the patients.



(a) Left Ventricle (b) Right Ventricle

Figure 6.3: Distribution of Ventricles Volume by sPESI

We can see that patients with higher sPESI have lower LV volume on average. However, their variance is higher compared to patients with a sPESI of 0. Additionally, patients with a sPESI

of 2 have larger RV volume on average. However, their variance is higher compared to patients with a sPESI of 1 and 0. We can also observe that the average RV volume for patients with a sPESI of 0 is greater than the average RV volume for patients with a sPESI of 1.

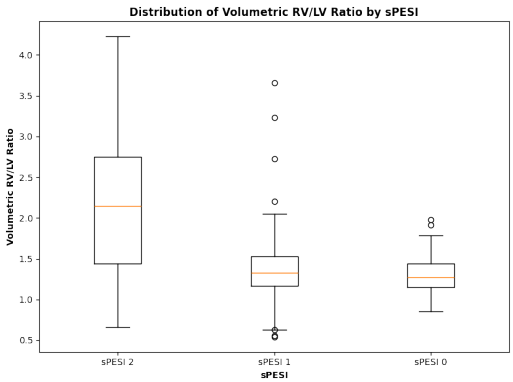


Figure 6.4: Distribution of Volumetric RV/LV Ratio by sPESI

For the patients of the PERSEVERE dataset, the average volumetric RV/LV ratio is higher for patients with a sPESI of 2. However, the volumetric RV/LV ratios for patients with a sPESI of 1 and 0 are similar.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this report we have presented our work related to the PERSEVERE project, whose goal is to propose novel risk stratification models for PE directly related to the routinely performed CTPA.

Our strong hypothesis is that the volumetric ratio of ventricles correlates more with the risk of death of PE patients than current RV/LV diameter ratio. Our objective is to develop an automatic segmentation method to compute the volumetric ratio of ventricles, and verify our hypothesis on PERSEVERE's dataset.

Our main contributions to the PERSEVERE projects are **1.** the ventricles' manual segmentations and automatic segmentations, and **2.** the statistical analysis of the correlation between the volumetric ventricles ratio and severity of patients' PE. We have also worked with other collaborators of the project on the vascular segmentation task and PERSEVERE's dataset. Therefore, we **3.** analysed and preprocessed the dataset, and **4.** refactored a code base aimed at running configurable training and inferences for vascular segmentation.

From our analysis on the PERSEVERE dataset with the segmentations of TotalSegmentator, the volumetric ratio for patients with a sPESI of 1 and 0 are very similar, and the volumetric ratio for patients with sPESI 2 is significantly higher in average. So it does not seem usable by itself to precisely diagnose PE patients. However, it could be used with other biomarkers. It would also be interesting to repeat this analysis after correcting the segmentations.

The limitations of our work are caused by the segmentations we used are ground truth for the evaluation. They have not been corrected by an expert which reduces the reliability of the evaluation. Additionally, since the manual segmentations were performed over different time periods, they lack consistency, which can hinder the model's training and its understanding of the concepts of ventricles.

7.2 Future Work

In the future, those ground truths will be corrected, and then used in active learning to perform the segmentation and correction of the cardiac ventricles for all the patients in the PERSEVERE's dataset, with the validation of an expert. Then, the volumetric ventricles ratio will be compared to the classic RV/LV ratio. Furthermore, the ventricle ratio will be integrated to a new risk PE stratification model.

Bibliography

[1] Connor Tice et al. "Management of Acute Pulmonary Embolism". In: *Current Cardiovascular Risk Reports* (2020). DOI: 10.1007/s12170-020-00659-z.

[2] Odyssée MERVEILLE. "Proposal PERSEVERE". ANR JCJC Proposal - PERSEVERE.

[3] U. Joseph Schoepf et al. "Right Ventricular Enlargement on Chest Computed Tomography A Predictor of Early Death in Acute Pulmonary Embolism". In: *Circulation. Cardiovascular imaging* (2004). DOI: 10.1161/01.CIR.0000147612.59751.4C.

[4] Felix G. Meinel et al. "Predictive Value of Computed Tomography in Acute Pulmonary Embolism: Systematic Review and Meta-analysis". In: *The American Journal of Medicine* (2015). DOI: 10.1016/j.amjmed.2015.01.023.

[5] Ghazaleh et al. Mehdipoor. "Patient-Level, Institutional, and Temporal Variations in Use of Imaging Modalities to Confirm Pulmonary Embolism". In: *Circulation. Cardiovascular imaging* (2020). DOI: 10.1007/s12170-020-00659-z.

[6] Narumol Chaosuwannakit et al. "Importance of computed tomography pulmonary angiography for predict 30-day mortality in acute pulmonary embolism patients". In: *European Journal of Radiology Open* (2021). DOI: 10.1016/j.ejro.2021.100340.

[7] Jun Ma et al. "Segment anything in medical images". In: *Nature Communications* (Jan. 2024). DOI: 10.1038/s41467-024-44824-z.

[8] Andriy Myronenko. "3D MRI brain tumor segmentation using autoencoder regularization". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. 2019. DOI: 10.1007/978-3-030-11726-9_28.

[9] Vahid Tavakoli and Amir A. Amini. "A survey of shaped-based registration and segmentation techniques for cardiac images". In: *Computer Vision and Image Understanding* (2013). DOI: https://doi.org/10.1016/j.cviu.2012.11.017.

[10] Peng Peng et al. "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* (Jan. 2016). DOI: 10.1007/s10334-015-0521-4.

[11] Chen Chen et al. "Deep Learning for Cardiac Image Segmentation: A Review". In: *Frontiers in Cardiovascular Medicine* (2020). DOI: 10.3389/fcvm.2020.00025.

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. DOI: 10.1109/CVPR.2015.7298965.

[13] Ronneberger Olaf et al. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. (2015). DOI: 10.48550/arXiv.1505.04597.

[14] Özgün Çiçek et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (June 2016). DOI: 10.1007/978-3-319-46723-8_49.

[15] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* (2021). DOI: 10.1038/s41592-020-01008-z.

[16] Christian Payer et al. "Multi-label Whole Heart Segmentation Using CNNs and Anatomical Label Configurations". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. 2018. DOI: 10.1007/978-3-319-75541-0_20.

[17] Chunliang Wang and Orjan Smedby. "Automatic Whole Heart Segmentation Using Deep Learning and Shape Context". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Mar. 2018. DOI: 10.1007/978-3-319-75541-0_26.

[18] Aliasghar Mortazi et al. "Multi-Planar Deep Segmentation Networks for Cardiac Substructures from MRI and CT". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. 2018. DOI: 10.1007/978-3-319-75541-0_21.

[19] Hejie Cui, Xinglong Liu, and Ning Huang. "Pulmonary Vessel Segmentation Based on Orthogonal Fused U-Net++ of Chest CT Images". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Cham, 2019. DOI: 10.1007/978-3-030-32226-7_33.

[20] Xin Yang et al. "Hybrid Loss Guided Convolutional Networks for Whole Heart Parsing". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Jan. 2018. DOI: 10.1007/978-3-319-75541-0_23.

[21] Chengqin Ye et al. "Multi-Depth Fusion Network for Whole-Heart CT Image Segmentation". In: *IEEE Access* (Feb. 2019). DOI: 10.1109/ACCESS.2019.2899635.

[22] Bobby Azad et al. "Foundational Models in Medical Imaging: A Comprehensive Survey and Future Vision". In: 2023. DOI: 10.48550/arXiv.2310.18689.

[23] Alexander Kirillov et al. "Segment Anything". In: *IEEE International Conference on Computer Vision* (2023). DOI: 10.48550/arXiv.2304.02643.

[24] João Andrade et al. "Pixel-level annotated dataset of computed tomography angiography images of acute pulmonary embolism". In: *Scientific Data* (Aug. 2023). DOI: 10.1038/s41597-023-02374-x.

[25] Jakob Wasserthal et al. "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images". In: *Radiology: Artificial Intelligence* 5.5 (2023), e230024. DOI: 10.1148/ryai.230024. URL: https://doi.org/10.1148/ryai.230024.

[26] Ekin Yagis et al. "Deep Learning for Vascular Segmentation and Applications in Phase Contrast Tomography Imaging". In: *Electrical Engineering and Systems Science* (Nov. 2023). DOI: 10.48550/arXiv.2311.13319.

[27] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. "Comparing images using the Hausdorff distance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1993). DOI: 10.1109/34.232073.

[28] Abdel Aziz Taha and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool". In: *BMC Medical Imaging* 15 (Aug. 2015). DOI: 10.1186/s12880-015-0068-x.

[29] Dennis Bontempi et al. “CEREBRUM: a fast and fully-volumetric Convolutional Encoder-decodeR for weakly-supervised sEgmentation of BRain strUctures from out-of-the-scanner MRI”. In: *Medical Image Analysis* (Mar. 2020). DOI: 10.1016/j.media.2020.101688.

[30] Rosana El Jurdi et al. “High-level prior-based loss functions for medical image segmentation: A survey”. In: *Computer Vision and Image Understanding* (2021). DOI: https://doi.org/10.1016/j.cviu.2021.103248.