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, Cyril Barbel, Laura Martin, Darius Engler, Edgar Delaporte and Jules Diaz 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 . . . 9

3 Configurable Vascular Segmentation **12**
3.1 Pre-existing code . . . 12
3.2 Our contribution . . . 13
3.2.1 Refactoring with PyTorch Lightning . . . 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

5 Segmentations **16**
5.1 Gain in expertise . . . 16
5.2 Ventricles' volume limits . . . 17
5.3 Manual segmentation . . . 18
5.3.1 Abdominal window . . . 18
5.3.2 Segmentation tools . . . 19
5.3.3 Manual segmentation issues . . . 20
5.4 Segmentation Models . . . 20
5.4.1 MedSAM . . . 20
5.4.2 Monai Auto 3D Seg . . . 21

6 Evaluation **22**
6.1 Inference duration . . . 22
6.2 Quantitative results . . . 22
6.3 Qualitative results . . . 23

7 Conclusion and Future Work **24**
7.1 Conclusion . . . 24
7.2 Future Work . . . 24

Chapter 1

Introduction

Pulmonary embolism (PE) is the third leading cause of death in Europe. It is characterized by a blood cloth called thrombus that blocks the pulmonary arteries and decreases blood flow which could cause systemic hypotension (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.

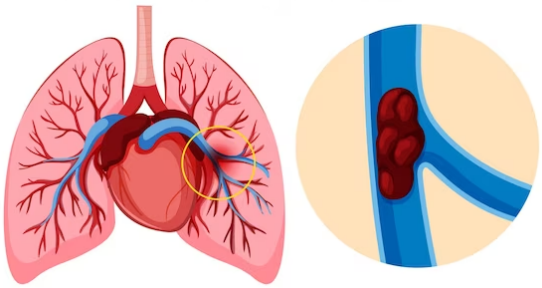


Figure 1.1: Pulmonary embolism (https://fr.freepik.com/photos-vecteurs-libre/embolie)

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) (Fig. 1.2) [2].

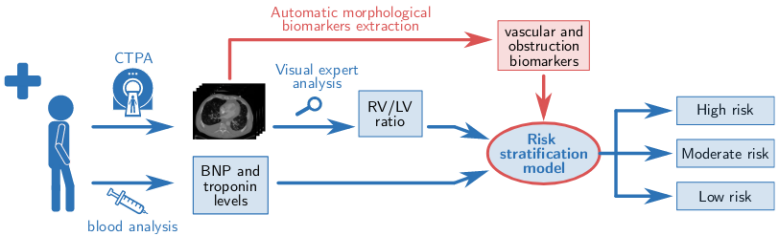


Figure 1.2: Current PE patient prognosis evaluation procedure in blue, and the modification proposed in the PERSEVERE project in red. [2]

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 synchronized with patients' cardiac cycle. 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. However, recent studies showed that since CTPA is not synchronised with the heart rate, it blurs the image 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 (Fig. 1.3).

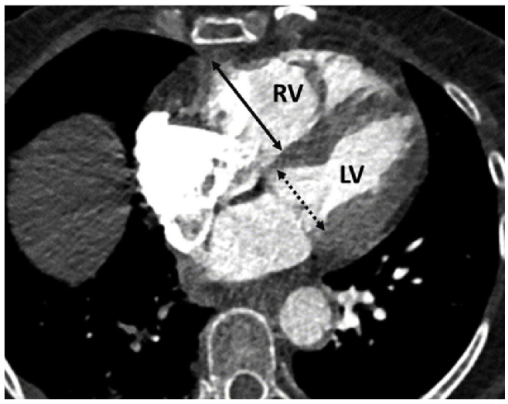


Figure 1.3: RV/LV diameter measurements on CTPA [6]

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 CTPA dataset of ventricles annotated because most of the research on PE is focused on segmentations of thrombus and PVT. And there is currently no automatic method to specifically segment ventricles on CTPA.

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. The resulting annotated dataset will be made publicly available to contribute to research on RV/LV ratio of ventricles on CTPA for PE risk stratification.

Our main contributions to the PERSEVERE projects are **1.**the ventricles' segmentations. However, we have also worked with other collaborators of the project on the vascular segmentation task and PERSEVERE's dataset. Therefore, we **2.** analysed and preprocessed the dataset, and **3.** refactored a code base aimed at running configurable training and inferences for vascular segmentation.

With our segmentations of PE patients with intermediate and high risk, we could easily see that the RV is larger than the LV. Therefore in real conditions, the automatic volumetric ratio would not be necessary for those patients. However, the segmentations will still be used to train the automatic segmentation model. The volumetric ratio of the ventricles is useful in more subtle cases, where the RV/LV diameter ratio would not be enough. Some segmentations fit this description, but require validation and analysis by an expert to draw conclusions.

In the first part, this report presents our bibliographic work. Following that, the report explains the refactoring of a repository used to easily configure vascular segmentation training and inference of different model and dataset. After that, the analysis and preprocessing of the dataset we used is explained. Then, the report focuses on the pipeline to segment CTPA ventricles manually with 3DSlicer, as well as some segmentation models tested. Finally, the evaluation of the segmentation models will be discussed.

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 [7] [8]. 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 [9]. 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) is the most common type of deep neural networks for image segmentation, classification and detection tasks. It is 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 (Fig. 2.1) [9].

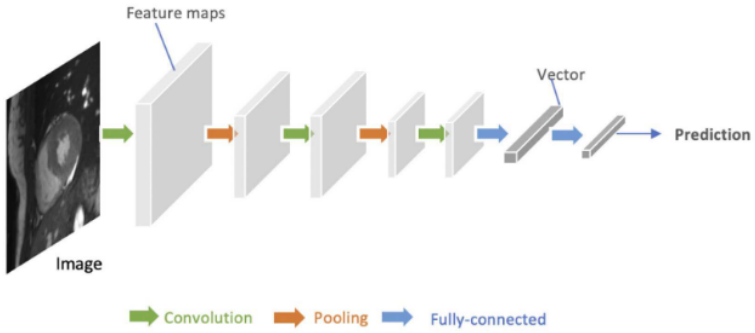


Figure 2.1: Generic architecture of convolutional neural networks (CNN) [9]

Fully convolutional neural network (FCN) [10] 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 capture detailed context information in an image for precise segmentation because features may be removed by the pooling layers in the encoder (Fig. 2.2) [9].

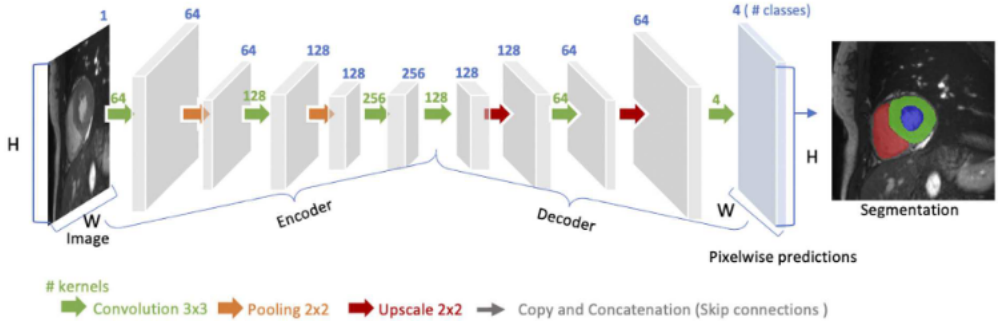


Figure 2.2: Generic architecture of fully convolutional neural networks (FCN) [9]

U-Net [11] is a variant of FCN, that uses skip connections between the encoder and decoder to recover spatial context lost during the down-sampling to increase the segmentation accuracy (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 [12]; or the nn-UNet which automatically optimizes the architecture and hyperparameters of the model for the given dataset [13].

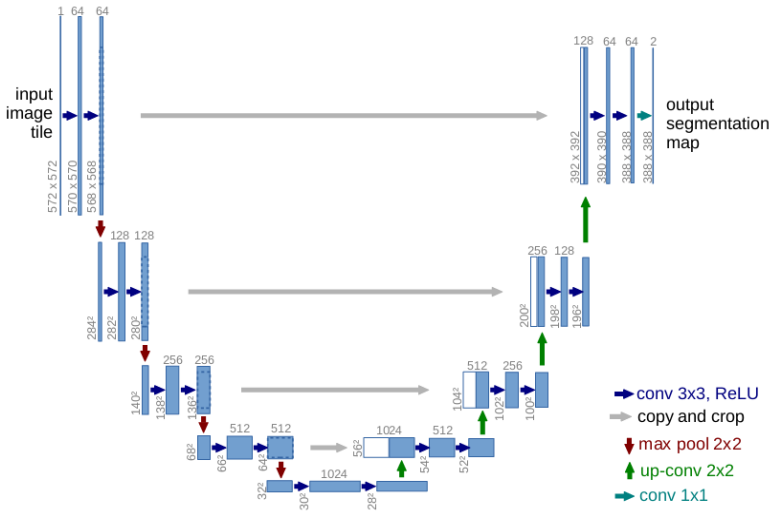


Figure 2.3: Architecture of U-Net [11]

2.2 Loss functions

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 :

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

where C is the number of all classes.

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

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

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 cause bias for the loss function because the accuracy score can be high just by predicting background pixels correctly. There are variants of loss functions that ignore the true positives, such as the soft-Dice; or that 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 generalizability, 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 (Fig. 2.4) [14]. This reduces the class imbalance problem because there are less background pixels, and facilitate the segmentation of regions of interest.

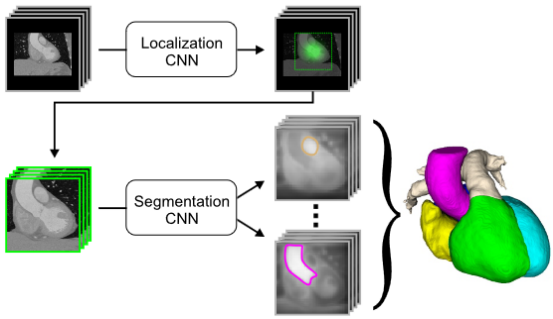


Figure 2.4: Two-step multi-label segmentation pipeline [14]

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 (Fig. 2.5). This allows the model to use the context of each pixel during the segmentation [15] [16] [17].

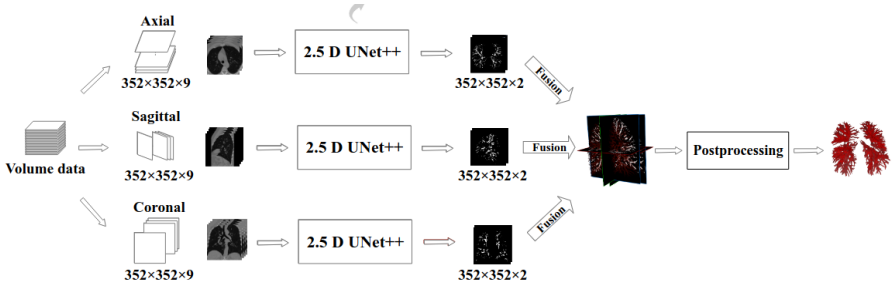


Figure 2.5: Multi-view segmentation pipeline [17]

Another way to address the class imbalance issue is to use hybrid loss, meaning combining different loss functions to improve the segmentation performance (Fig. 2.6) [18] [19].

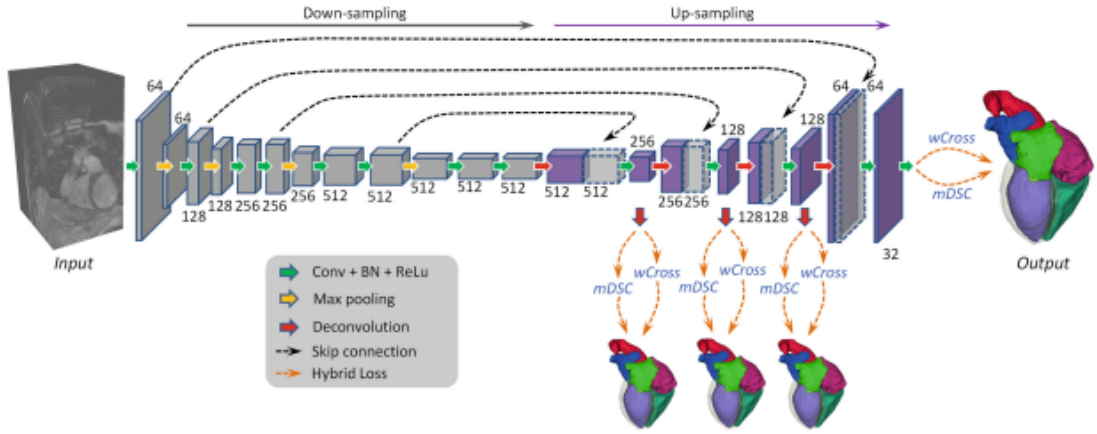
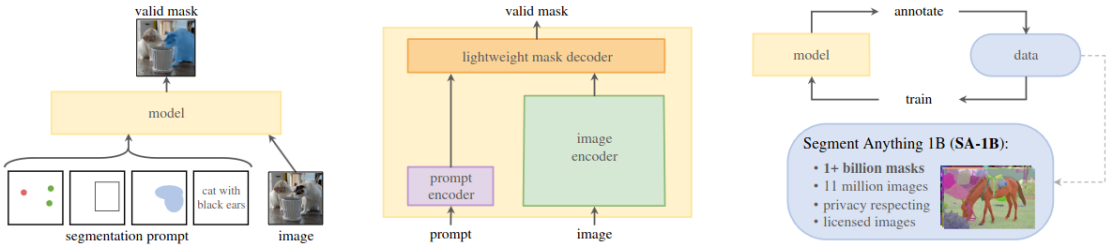


Figure 2.6: Hybrid loss segmentation pipeline [18]

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 [20]. 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 localize the region to segment and then do the segmentation, foundation models can be guided by user-prompt as seed point or bounding-box (Fig. 2.7) [21]. Some foundations models are fine-tuned to enable universal medical images segmentation to improve generalisability across imaging modalities, organs and diseases [22].



(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 [21]

Chapter 3

Configurable Vascular Segmentation

The first task was to refactor a code repository belonging to the laboratory of the INSA Lyon called CREATIS (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 [11]. There were Python scripts that had the same pipeline but with different model architecture, hyperparameters, data pre-processing, data augmentation, dataset, 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 inference 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.

3.2 Our contribution

3.2.1 Refactoring with PyTorch Lightning

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. A 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.

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 composable from multiple sources, 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. This allows to have one script with configurable parameters for training and another one for inference.

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

A 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 annotated. 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. There were a dataset of ventricles in CTPA for classification with labels indicating the presence of a PE in the image, and the corresponding RV/LV size ratio. However, the objective is to automatically segment the ventricles of CTPA but there were no ground truth for ventricles segmentation. Therefore, the dataset could not be used for supervised training, but it is still an interesting dataset to keep [23].

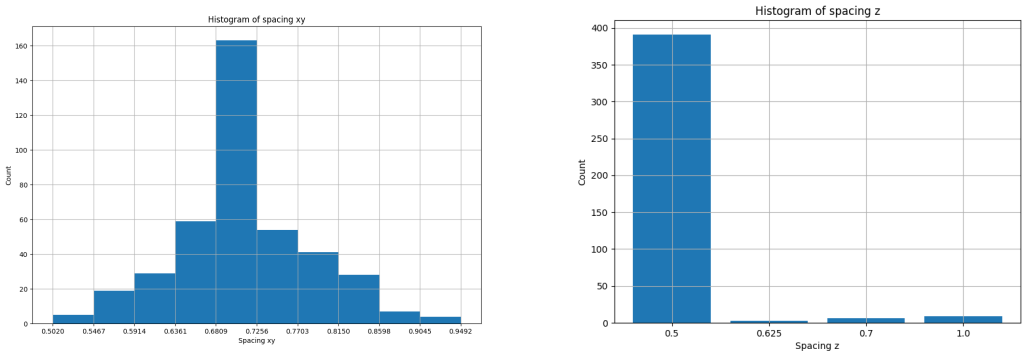
4.1 Project Dataset

Eventually, the dataset of PERSEVERE was chosen to test the different segmentation models. It is stored in a **Girder**, it is 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 padding. And they are stored in **NIfTI** (Neuroimaging Informatics Technology Initiative) format.

4.2 Resampling

An analysis of the dataset images header was done with a script that creates an Excel table of its dimension, voxel's dimensions (spacing) for every image in the dataset. 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.

To avoid pre-processing during the training, a separate dataset resampled was created. A script was written to compute the histogram of the spacing of the image in the dataset with the number of bins as parameter (**Fig. 4.1**). The figure is then displayed and saved in order to document the Girder.

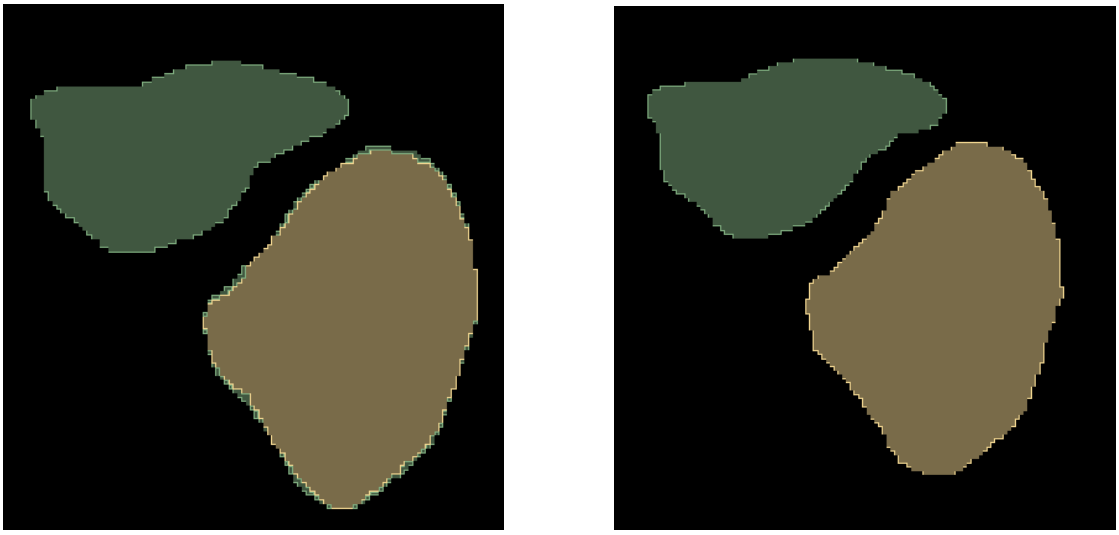


(a) Spacing XY

(b) Spacing Z

Figure 4.1: Histogram of the dataset spacing

The most represented spacing was selected to resample the other using a script with the **SimpleITK** library. We first tried to use linear interpolation, which computes the weighted average of the four nearest neighbour pixel values at non-discrete coordinates (**Fig. 4.2a**). However, there is a different class for each ventricle represented as a discrete value. Therefore, the resampled ground truth of the class represented by the label 2 had its border set has the label 1. To solve this issue, nearest neighbour interpolation was used (**Fig. 4.2b**).



(a) Linear interpolation

(b) Nearest neighbour interpolation

Figure 4.2: Results of the ventricles' resampling

It was important to verify that the interpolation did not deform the images or the ground truth. Specifically, the vascular structures connections were carefully verified since they are the thinnest elements, and it is important to keep their connectivity.

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 if needed. 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.

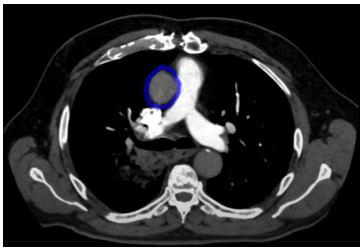
However, the lack of anatomical and CTPA knowledge was problematic, as we needed to know the position of the ventricles in the volume. Furthermore, it was important that we could evaluate the segmentations we had done and later the automatic segmentations.

5.1 Gain in expertise

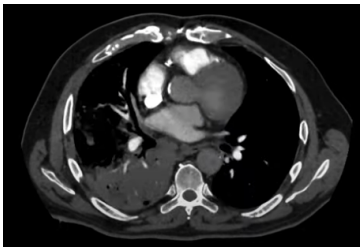
To gain the necessary anatomical and CTPA knowledge needed to localise the ventricles' volumes, RV/LV ratio related articles were read **[6]** **[3]** **[4]**, and CT, CTPA and PE related videos were watched [1] [2] [3] [4] [5].

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 (**Fig. 5.1**).

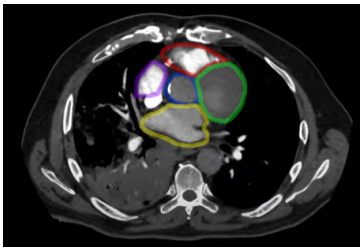
[1] **https://youtu.be/8WUgH4WHlLE?si=uk5pXg1hqpcg8VJl&t=1111**
[2] **https://youtu.be/yeNd_nhiQ6U?si=xgSMQvEfLg20B-XV**
[3] **https://youtu.be/Ibc_ryTwX40?si=S7OxcuyMcsNtlaFv**
[4] **https://youtu.be/ciCYQWm0VTE?si=FpOoV28lYBOjZZ0i**
[5] **https://youtu.be/QHYOuudasF0?si=EAIzQa2UsBCeSvEH**



(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 in those resources, the ventricles' volume limits.

5.2 Ventricles' volume limits

This section defines the limits instructed by A. Serva, and the specificity 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 (Fig. 5.2). However, the RV/LV ratio is measured based on the ventricles diameters delimited by the endocardium.

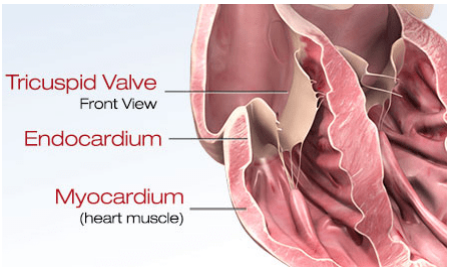
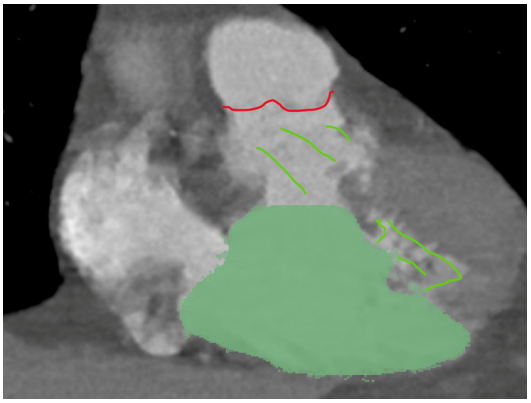


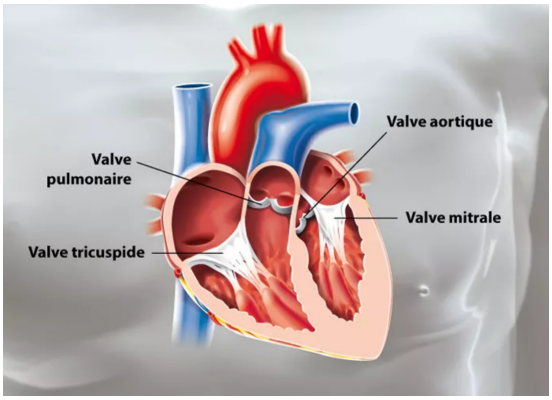
Figure 5.2: Heart wall layers (https://www.heart.org/en/health-topics/myocarditis)

The thickness of the myocardium is correlated to 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 (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

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 Manual segmentation

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.

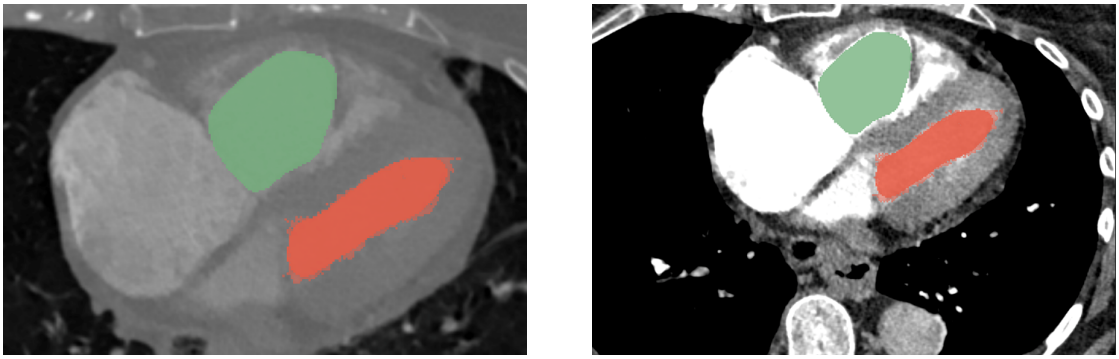
5.3.1 Abdominal window

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 (Fig. 5.4).

[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



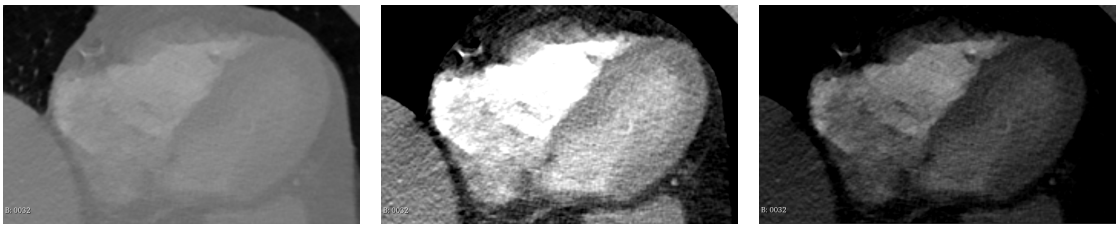
(a) Segmentation on default window

(b) Same segmentation on media intestinal window

Figure 5.4: 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 (Fig. 5.5).



(a) Default window

(b) Abdominal window preset

(c) Manually adapted window

Figure 5.5: 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.

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.

To accelerate the segmentation process, segmentation models can be used to have a basis that can be refined manually.

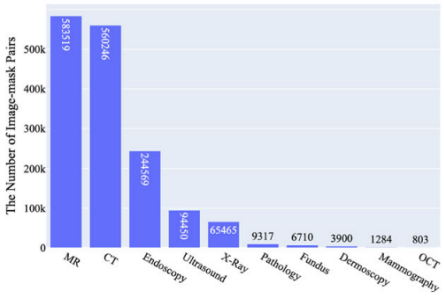
5.4 Segmentation Models

Automatic segmentation models are often used to facilitate the annotating process. However, there is currently no specialized model to segment ventricles on CTPA.

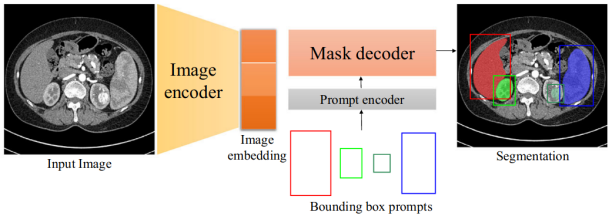
Since the segmentations were done on 3DSlicer, plugins of automatic segmentation model for ventricles were interesting as the inference result could be computed and visualized easily. Among those plugins were **MedSAM** and **MONAI Auto3DSeg**.

5.4.1 MedSAM

MedSAM is a recent foundation model based on the architecture of SAM [21] but fine-tuned for medical images (Fig. 5.6a). It is composed of an image encoder, a prompt encoder, and a mask decoder. The image encoder maps the input image into a high-dimensional image embedding space; the prompt encoder transforms the bounding boxes into feature representations via positional encoding [24]; the mask decoder fuses the image embedding and prompt features using cross-attention (Fig. 5.6b) [22].



(a) Modality distribution of MedSAM's dataset [22]



(b) Architecture of MedSAM [22]

MedSAM has shown better performance and less variance on segmentation tasks of medical images compared to SAM since there are significant differences between natural images and

medical images. Furthermore, MedSAM also shows performances that compare or exceed specialized models [22].

The main advantage of MedSAM is its ability to learn intricate image features and deliver accurate segmentation for different tasks. However, the training cost is consequent due to the large number of data used.

Their official 3DSlicer plugin, which is publicly available on the developer tools extensions, was utilised. Similarly to SAM, the plugin requires user interaction to input a segmentation box, in which the model will perform the segmentation. The model is already trained and no fine-tuning was needed.

5.4.2 Monai Auto 3D Seg

Monai Auto 3D Seg regroups a dozen ready-to-use segmentation models for multiple imaging modalities and regions of interest.

The whole body segmentation TS1 quick model was tested. It is based on the SegResNet architecture, an encoder-decoder based CNN with an asymmetrically larger encoder to extract image features and a smaller decoder to reconstruct the segmentation mask (Fig. 5.7) [25]. It segments 104 structures including ventricles on CT images.

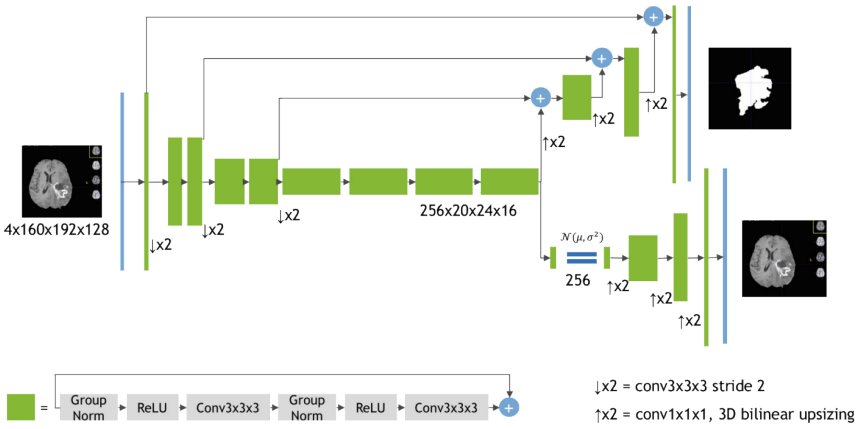


Figure 5.7: Architecture of the SegResNet [25]

Beyond the use of the given models, customized models can be added to the plugin. This will be useful to test different models and analyse the result easily.

Chapter 6

Evaluation

In total, we did eight segmentations of ventricles' volume manually. They are not reviewed but they will be used as ground truth to evaluate and compare the segmentation of MedSAM and Monai Auto 3D Seg.

We first compare the segmentation duration of each model, then we evaluate the inference result quantitatively. And finally, we look at the qualitative results to select a segmentation model.

The selected model will be used to accelerate the segmentation process.

6.1 Inference duration

The average manual segmentation duration is 5 hours.

In comparison, the segmentation duration of MedSAM is approximately 5 minutes, including the creation of the bounding box and the inference. And the segmentation duration of Monai Auto 3D Seg is approximately 2 minutes.

6.2 Quantitative results

To evaluate the segmentations, the Dice Similarity Coefficient (DSC) was used. It is one of the most common overlap-based metrics. It calculates the spacial similarity between the segmented result and the ground truth based on their overlapping regions [26].

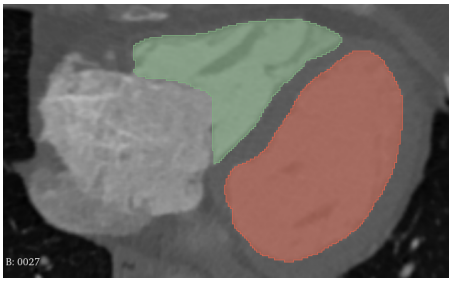
$$DSC = \frac{2 \cdot TP}{FP + FN + (2 \cdot TP)}$$

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

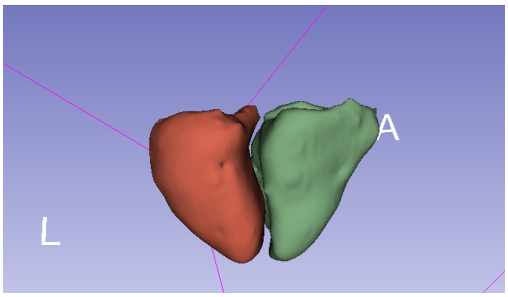
The DSC of the MedSAM segmentation for this patient is 0.697 for the RV and 0.925 for the LV.

The DSC of the Monai Auto 3D Seg segmentation for this patient is 0.773 for the RV and 0.820 for the LV.

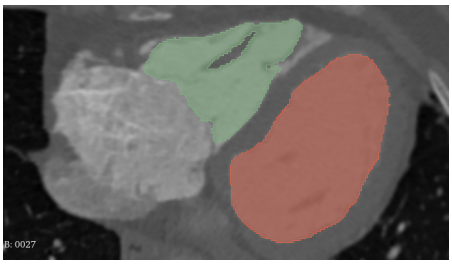
6.3 Qualitative results



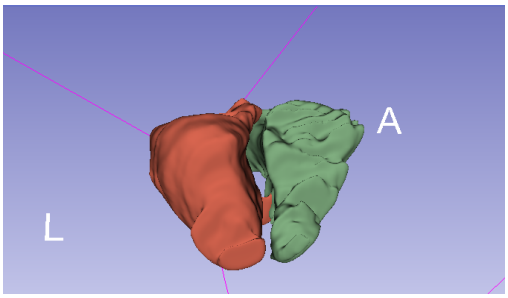
(a) Ground truth slice in the axial view



(b) Ground truth in 3D view

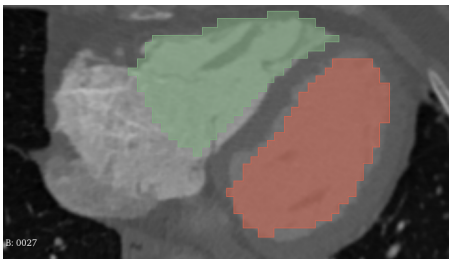


(a) MedSAM segmentation slice in the axial view

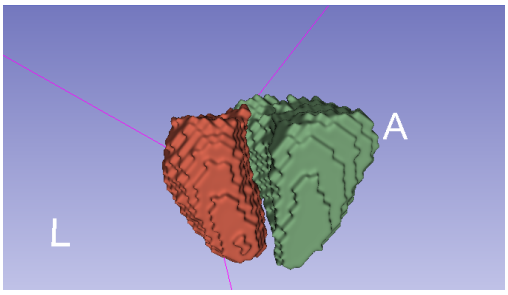


(b) MedSAM segmentation in 3D view

The model segmented the RV very accurately, however it made different decisions for segmentation of the LV, which might explain its low DSC. It also struggled with the pulmonary valves limits.



(a) Monai 3D Auto Seg segmentation slice in the axial view



(b) Monai 3D Auto Seg segmentation in 3D view

The low resolution of the segmentation can explain the low DSC despite the seemingly correct boundary decisions, except for the pulmonary valves.

MedSAM generates more precise segmentations, especially around the pulmonary valves. Therefore, it will be used to accelerate our segmentation process.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this report we have presented our works 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 ventricle correlates more to 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 major contribution was ventricles segmentations of multiple patients.

With our segmentations of PE patients with intermediate and high risk, we could easily see that the RV is larger than the LV. Therefore in real conditions, the automatic volumetric ratio would not be necessary for those patients. However, the segmentations will still be used to train the automatic segmentation model. The volumetric ratio of the ventricles is useful in more subtle cases, where the RV/LV diameter ratio would not be enough. Some segmentations fit this description, but require validation and analysis by an expert to draw conclusions.

In the future, we would like to develop a automatic segmentation method to have a fully annotated dataset. The goal being to make it publicly available to contribute to research on RV/LV ratio of ventricles on CTPA for PE risk stratification.

7.2 Future Work

Our next step will be to finish the manual segmentations needed by the radiologist.

After the review and corrections of the segmentations, we will start to experiment multiple methods and model architectures to automatically segment ventricles. Then after verification, we will analyze the RV/LV ratio of each patient.

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] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

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

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

[23] 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.

[24] Matthew Tancik et al. "Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains". In: *Advances in Neural Information Processing Systems* (2020). DOI: 10.48550/arXiv.2006.10739.

[25] 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.

[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.