MAIA

ERASMUS MUNDUS

JOINT MASTER
IN MEDICAL IMAGING
AND APPLICATIONS

Joint Master in Medical Imaging and Applications

Master Thesis Proceedings

Promotion 2019-21

www.maiamaster.org

MAIA

An international programme by the University of Girona (Spain), the University of Bourgogne (France) and the University of Cassino (Italy) funded by Erasmus + Programme.

Erasmus+ Universitat de Girona uB

Copyright © 2021 MAIA

Published by the MAIA Master

www.maiamaster.org

This document is a compendium of the master thesis works developed by the students of the Joint Master Degree in Medical Imaging and Applications. Therefore, each work is independent on the other, and you should cite it individually as the final master degree report of the first author of each paper (Student name; title of the report; MAIA MSc Thesis; 2021).

Editorial

Computer aided applications for early detection and diagnosis, histopathological image analysis, treatment planning and monitoring, as well as robotised and guided surgery will positively impact health care during the new few years. The scientific community needs of prepared entrepreneurships with a proper ground to tackle these topics. The Joint Master Degree in Medical Imaging and Applications (MAIA) was born with the aim to fill this gap, offering highly skilled professionals with a depth knowledge on computer science, artificial intelligence, computer vision, medical robotics, and transversal topics.

The MAIA master is a two-years joint master degree (120 ECTS) between the Université de Bourgogne (uB, France), the Università degli studi di Cassino e del Lazio Meridionale (UNICLAM, Italy), and the Universitat de Girona (UdG, Spain), being the latter the coordinating institution. The program is supported by associate partners, that help in the sustainability of the program, not necessarily in economical terms, but in contributing in the design of the master, offering master thesis or internships, and expanding the visibility of the master. Moreover, the program is recognised by the European Commission for its academic excellence and is included in the list of Erasmus Mundus Joint Master Degrees under the Erasmus+ programme.

This document shows the outcome of the master tesis research developed by the MAIA students during the last semester, where they put their learnt knowledge in practice for solving different problems related with medical imaging. This include fully automatic anatomical structures segmentation, abnormality detection algorithms in different imaging modalities, biomechanical modelling, development of applications to be clinically usable, or practical components for integration into clinical workflows. We sincerely think that this document aims at further enhancing the dissemination of information about the quality of the master and may be of interest to the scientific community and foster networking opportunities amongst MAIA partners.

We finally want to thank and congratulate all the students for their effort done during this last semester of the Joint Master Degree in Medical Imaging and Applications.

MAIA Master Academic and Administrative Board

Contents

Optimizing stroke segmentation using acute brain CTA

Uma Maria Lal-Trehan Estrada, Luca Giancardo

School of Biomedical Informatics, UTHealth, Houston, TX, USA

Abstract

Background. Stroke is the second leading cause of death and the third leading cause of disability worldwide. Acute ischemic stroke (AIS) accounts for 87% of all strokes. Prompt diagnosis and treatment of AIS patients is of vital importance for the recovery and outcome of these patients. Effective treatment of AIS is predicted on neuroimaging. Some models have been proposed to automatically segment stroke lesions from neuroimaging. Different from other studies proposed in the literature, this work focuses on optimizing stroke lesion segmentation using acute brain CTA, a widely available imaging modality. The hypothesis of this project is that stroke segmentation can be optimized using as input a CTA scan from which there is a high difference between the stroke region and the rest of the brain.

Material and Methods. This study uses whole brain CTP data from AIS patients and treats them as sequential brain CTA scans. For each patient, the ground truth stroke lesion segmentation is provided in MRI DWI images acquired after successful recanalization. Firstly, the difference between the stroke region and the symmetrical region on the contralateral hemisphere is analyzed using the Jensen Shannon Distance. The same analysis is made between the hemispheres without considering the stroke area. This distance is statistically compared between different CTP time scans (equivalent to CTA scans acquired at different times). Then, an adaptation of a previously proposed Deep Learning based model for stroke detection (DeepSymNet) is implemented to automatically segment the stroke infarct core from the CTAs. A patch based strategy is used to train the model, which takes as input the main patch to be classified and the symmetric patch on the contralateral hemisphere. The segmentation results are evaluated using F1 score and are also statistically compared between different CTP time scans.

Results. Regarding the analysis of difference, the results show that a CTA scan acquired approximately 3 seconds after the peak arterial enhancement phase provides significantly higher differences in the stroke area compared to other scan start times. Concerning the segmentation, due to a high number of false positives mainly appearing in the area of the skull, poor results (F1 score around 0.2) are obtained.

Conclusions. This works shows that a brain CTA scan acquired approximately 3 seconds after the peak arterial enhancement phase, target phase for current CTA scan protocols, can provide higher information related to stroke compared to acquisitions performed at any other time. Further work is required in order to improve the segmentation results and being able to robustly compare the segmentation performance between CTAs acquired at different instances of time so as to test the hypothesis of the project.

Keywords: Acute Ischemic Stroke, CT Angiography, Stroke Segmentation, Optimal CTA Scan Start Time, Peak Arterial Enhancement Phase

1. Introduction

A stroke is a lesion characterized by the sudden death of nerve cells due to restricted blood flow to a region of the brain, resulting in a corresponding loss of neurological function. This condition can be caused by a blockage or a rupture of a brain artery. Depending on the cause, blockage or rupture, stroke is referred to as ischemic stroke or hemorrhagic stroke, respectively (Johnson et al., 2016).

The effects of stroke are diverse and depend on the brain region that is affected. It can cause partial or complete disability and it is a major cause of human death.

In fact, stroke is the second leading cause of death and the third leading cause of disability worldwide (Johnson et al., 2016).

1.1. Acute Ischemic Stroke

Acute ischemic stroke (AIS) accounts for 87% of all strokes and its main cause is atherosclerosis, which is a narrowing of the arteries caused by accumulation of plaque. Plaque, which consists of fatty deposits that line the vessel walls, can provoke two types of obstruction: cerebral thrombosis, which is a blood clot that grows at the fatty plaque within the blood vessel, or cerebral embolism, which is a blood clot developed at a different location that travels through the circulatory system until reaching vessels that are too narrow to allow its circulation (der Worp and van Gijn, 2007).

For the patients with AIS, time is brain. With the onset of acute ischemia, cerebral autoregulation (a manifestation of local blood flow regulation) is able to maintain the cell viability. However, when the processes involved in this autoregulation are exhausted, irreversible ischemic damage (cell death) occurs in a progressive, time-dependent manner (Brouns and De Deyn, 2009). Therefore, prompt diagnosis and treatment of AIS are of vital importance for the recovery and outcome of stroke patients.

1.2. Neuroimaging

Effective treatment of AIS is predicted on neuroimaging. While non-contrast head computed tomography (NCCT) has played a role in excluding hemorrhage, the introduction of intravenous thrombolysis with alteplase (a thrombolytic agent aimed at dispersing the blood clot) (Hacke et al., 2008) has made advanced imaging to be placed at the center of the immediate patient evaluation. Since the proven efficacy of mechanical thrombectomy (Wahlgren et al., 2016), the main goal of advanced imaging is to identify patients with salvageable brain tissue who may benefit from vessel recanalization. Potentially salvageable tissue is called the penumbra, while irreversibly injured tissue is referred to as the ischemic core.

1.2.1. Non-contrast Computed Tomography

The initial imaging evaluation of an AIS patient usually begins with non-contrast head CT (NCCT). NCCT is the fastest, easiest and most widely utilized imaging modality in stroke. Its main purpose is excluding hemorrhage. Ischemic changes, which are defined in NCCT as hypoattenuation or loss of gray-white differentiation, begin to appear on NCCT scan within several hours. The Alberta Stroke Program Early CT Score (ASPECTS) is used to recognize the importance of these changes and to subjectively asses the extent of early infarct on NCCT. ASPECTS is a 10-points segmental assessment of the middle cerebral artery (MCA) vascular territory. One point is deducted from the initial score of 10 points if a region is involved (Pexman et al., 2001). Despite its simplicity, ASPECTS can be used as a measure of ischemic core. However, as a semi-quantitative measure, it requires expertise and is observer dependent. Another weakness of ASPECTS includes the low resolution of CT in the detection of early changes after AIS. Therefore, ASPECTS is commonly used, but its relevance in patient triage for treatment has been limited due to the existence of more advanced imaging modalities.

1.2.2. Magnetic Resonance Imaging

As an initial evaluation, MRI is both more sensitive and more specific for AIS than CT. Diffusion-weighted imaging (DWI) shows changes as early as within a few minutes of cell death. The ischemic core is identified as a DWI lesion, characterized by its restricted diffusion (Pavlina et al., 2018). Resource constraints limit the application of non-contrast MRI to stroke triage at many centers. Moreover, MRI imaging has been shown to delay time to treatment, and the benefit of MRI over CT for patient triage for thrombectomy is still unclear.

Ischemic changes can also be identified on fluid-attenuated inversion recovery (FLAIR). However, in this case, the changes appear over several hours (Wouters et al., 2016).

1.2.3. Computed Tomography Angiography

CT angiography (CTA) of the brain is a noninvasive technique which allows the visualization of the cerebral arteries via contrast enhancement. CTA is standardly used in the detection of large vessel occlusion (LVO) (Reidler et al., 2020). After identifying an occlusion, the presence of penumbral tissue guides the decision to perform endovsacular thrombectomy (EVT). The penumbra can be determined by qunatifying the collateral supply. Collateral circulation, which can be assessed by CTA, is a network of arterial anastomoses supplying the brain tissue when the main blood flow providers fail to meet demands (Liebeskind, 2003). Poor collateral circulation is associated to early and more extensive ischemic damage. Several studies have shown that the degree of collateralization is correlated with the benefit to thrombectomy (Bonney et al., 2019). However, the quantification of collaterals has not been adopted in the patient triage for EVT.

Although collaterals have also been assessed with MR angiography (MRA), the techniques are less validated than with CTA (Bonney et al., 2019).

In CTA, the optimal shape of the contrast bolus passage corresponds to a rapid rise, followed by a peak enhancement plateau and a rapid fall (Hinzpeter et al., 2019). The scan acquisition is aimed at capturing the peak arterial enhancement phase, also referred to as mid arterial phase. As opposite to single-phase CT angiography (sCTA), multiphase CT angiography (mCTA) was developed in order to obtain temporal information by scanning beyond the peak arterial phase (Menon et al., 2015). With mCTA, images in the late arterial and venous phases can be obtained, by repeating the CT acquisition following a certain delay. Hypoperfused regions will show low vessel opacification in the peak arterial phase, while regions with delayed perfusion supplied by collateral vessels will show vessel enhancement only in the later phases. Therefore, collateral flow can be more accurately analyzed by mCTA than by sCTA.

1.2.4. Computed Tomography Perfusion

CT perfusion (CTP) is a dynamic imaging modality where the first pass of a bolus of iodinated contrast agent is traced through the brain parenchyma by repeated rapid scanning of the volume of interest immediately after injection. The passage of the contrast is depicted with a time-density curve (also called time attenuation curve), which indicates the change in CT density of the imaged tissues as the agent passes through the vasculature. This imaging modality provides physiologycal and pathophysiological information of the cerebral hemodynamics.

CTP, as well as MR perfusion, allow the calculation of different parameters that reflect different aspects of the hemodynamic state, such as cerebral blood flow (CBF), time to peak (TTP), cerebral blood volume (CVB) and mean transit time (MTT). CBF is the flow rate of the blood within a brain region. When the blood supply is compromised in a region, the CBF in that region is decreased. When there is no or very small CBF, the brain tissue becomes irreversibly injured. TTP for a given region is the time required to reach the maximal attenuation. It is used to measure the time to perfuse a territory. This time is increased in a region with compromised blood supply. Compared to CBF and TTP, the effect of AIS on CBV is less intuitive. CBV is the total blood volume present in a region of the brain. In AIS, CBV may be "normal" or even increased as a consequence of the autoregulatory vasodilation. However, when autoregulatory processes are exhausted, which leads to irreversibly injured tissue, CBV is decreased. CBV is divided by CBF to obtain MTT, which is the time spent by the blood in the capillary phase.

As the contrast bolus flows through the vasculature, the enhancement profile in Hounsfield units (HU) is plotted against time and two attenuation curves are obtained: a tissue enhancement curve and an arterial enhancement curve. These curves undergo mathematical postprocessing (deconvolution), essentially, pixel scaling, to create a single residue function enhancement curve. All perfusion variables (CBF, TTP, CVB, MTT) and final perfusion maps (maps representing the values of the perfusion variables for each voxel of the brain) are derived from these curves (Catanese et al., 2017).

Visual inspection of perfusion maps yields different results between readers (Bivard et al., 2015) . To obtain more consistent and objective results, perfusion maps are commonly subjected to thresholding to identify ischemic core and differentiate it from penumbral tissue. These thresholds held true when the pathological hemisphere is compared with the healthy contralateral hemisphere. The optimal thresholds to assess the core and the penumbra on CTP have been extensively evaluated. However, there is no conclusive definition of ischemic core using CTP maps. Currently, the volume of tissue with a relative reduction of CBF <30% as compared with the contralateral hemisphere is commonly used to estimate core on CTP (Garcia-Tornel et al., 2021).

Complete multimodal CT has been used to extend the EVT window to 24 h (Nogueira et al., 2018).

However, CTP has some negative aspects. These negative aspects include a higher radiation dose compared to sCTA and mCTA; a higher bolus of contrast material, which could cause renal complications; an increased acquisition time and data processing time; the requirement of expensive software and licenses; the lack of standardization across commercialized platforms; and the limited availability of technicians trained to perform such acquisition. Therefore, there is a pressing need to screen patients that may be candidates for treatment using imaging data and expertise that are available while maintaining the diagnostic and prognostic accuracy.

1.3. Goal of this project

Given the fact that CTA is available in most medical centers and compared to NCCT, it performs better for the detection of AIS and estimation of infarct volume (Camargo et al., 2007), this work is focused on this imaging modality, in particular, on optimizing stroke segmentation using acute brain CTA.

Commonly, CTA acquisition protocols are designed to maximize arterial-phase contrast enhancement, that

is, they aim to scan the brain at the mid arterial phase. When there is ischemia, differences in the vasculature and the brain tissue can be observed in the two hemispheres of the brain. Depending on when the CTA scan is started, there may not be sufficient delay time for a steady state between arterial and tissue contrast material to be reached, leading to low tissue enhancement and a consequent low difference between hemispheres. In particular, a non-considerable enhancement difference between the stroke region and its symmetric contralateral region hinders the detection of stroke.

In order to optimize stroke segmentation using acute brain CTA, this work is aimed at finding the CTA scan start time that maximizes the amount of stroke information that can be extracted so as to use it as input for the segmentation process.

This is achieved by comparing CTA scans acquired at different instances of time after contrast material injection. Specifically, based on what experts do when visually inspecting, the difference between hemispheres is analyzed. Although the brain is not perfectly symmetric, in healthy conditions, there are not substantial differences in the intensity between one hemisphere and the other (both hemispheres are equivalently perfused). However, in an ischemic scenario, the intensity distribution may well differ between the affected region and the contralateral symmetric region.

Furthermore, a representation learning model for ischemic core segmentation is trained and validated on CTA scans acquired at different instances of time and the segmentation results are compared with respect to the CTA scan start time.

2. State of the art

Computed tomography perfusion (CTP) is a dynamic imaging modality which gives functional information useful to differentiate salvageable ischemic brain tissue (the penumbra) from the irrevocable damaged tissue (the infarct core), thereby guiding acute ischemic stroke (AIS) therapy. However, perfusion imaging and interpretation may not be available in a substantial number of institutions. On the other hand, computed tomography angiography (CTA) is available in most centers.

Considering this, a potential research path involves the analysis of CTA as a proxy for perfusion imaging in the detection of blood flow abnormalities.

Perfusion source data are a 4-dimensional data set, 3-dimensional volumes captured over time. This 4-dimensional data set (CTP) can be seen as a sequence of CTA scans performed at different time delays. For each voxel in CTP, time-attenuation curves (TAC) can be obtained. A CTA scan corresponds to one time point of these curves for each voxel. Multi-phase CTA (mCTA) corresponds to several (normally, 3) time points of these curves.

Being able to automatically reconstruct the CTP TACs from single-phase CTA (or mCTA), that is, for each voxel, going from 1 (or 3) time points of the TAC to several time points, would allow to obtain a synthetic CTP and compute from that the perfusion parameters necessary to assess AIS patient for a treatment. This would be obtained using a faster, more widely available, less radiation-requiring and less contrast-requiring imaging modality (CTA), compared to CTP. In addition, unlike estimating directly stroke related characteristics, such as the infarct core volume, location and delineation, which do not have a physical basis, reconstructing the time-attenuation curves of CTP consists of learning a real signal.

The ability to reconstruct the CTP TACs may vary depending on the time at which the CTA scan given as input is acquired. In particular, it is likely that inputting a CTA scan acquired at a time instance at which higher stroke information amount can be extracted will contribute to a better TAC reconstruction and a more accurate computation of the perfusion parameters.

This work analyzes CTA scanning start time for higher stroke information amount extraction with the final goal to optimize stroke segmentation. Therefore, a literature review related to this topic is presented in this section.

2.1. CTA protocol

In the literature, there are works that have analyzed the CTA scan protocol focused on providing diagnostically useful information about both the vascular and parenchymal phases of brain enhancement. Some studies have analyzed the CTA scan protocol with the specific aim of obtaining more stroke information. However, as far as we are concerned, there is no work that studies this by analyzing the difference between intensity distributions of the stroke region and the symmetric contralateral region.

Cademartiri et al. (2002) reviewed the parameters that affect the geometry of the bolus in CTA and gave recommendations for an optimal and robust scan protocol for CTA. They define as optimal bolus geometry for CTA an ”immediate increase in the enhancement of the studied artery to a high maximum value of enhancement just before the start of the acquisition of CT data and a steady state in which the enhancement does not alter during the acquisition”. In a real scenario, though, the

optimal bolus geometry cannot be not reached. CTA is normally acquired during the rise and the fall of the time attenuation curve.

The time required for a contrast material to reach the target vasculature varies among patients. For this reason, ACR et al. (2020) recommend to perform an assessment of patient-specific circulation time for optimal CTA scanning.

Circulation timing can be performed using one of the following techniques: bolus tracking or test bolus technique.

Bolus tracking is the most common technique for timing the scan start of CTA. This technique involves monitoring the contrast enhancement in a predefined vessel. After a threshold is reached, the scan is automatically started after a predefined and fixed trigger delay (Cademartiri et al., 2002).

For the test bolus technique, a low dose of contrast material is injected and low radiation dose, single-slice, sequential scans are acquired at a slice containing the vessel of interest, allowing the analysis of contrast enhancement over time and obtaining a time attenuation curve. The peak of the obtained time attenuation curve is used to compute the scanning delay after contrast agent injection (Cademartiri et al., 2002).

Compared to the test bolus technique, the main benefits of bolus tracking include its time efficiency, the lower contrast material requirement and lower radiation dose exposure. On the other hand, test bolus is a patient-specific technique and it is able to remove the variability of the bolus tracking technique. These two techniques, though, are not specifically designed to obtain more information related to stroke, but for global optimal arterial enhancement.

2.2. CTA optimization for stroke information

Focused on stroke, in the work published by Pulli et al. (2012), they test whether the parameters of CTA acquisition protocol affect the relationship between acute ischemic infarct size on CTA source images and concurrent DWI MRI images, being the measured areas of hyperintensity at acute DWI used as the standard of reference for infarct size (outlined manually by two neuroradiologists). They observe that CTA protocol modifications aimed at increasing imaging speed and optimizing arterial enhancement are associated with significant overestimation of infarct size on CTA, which could lead to inappropriate exclusion of patients who may benefit from treatment. These protocol changes may prevent sufficient delay time for a steady state between arterial and tissue contrast material to be reached. In this work, they also report that a time of 38 seconds to imaging (with respect to contrast injection) of the middle of the anterior circulation territory shows good discrimination between good and poor agreement between the infarct volume on CTA and the infarct volume on DWI.

Based on the fact that some studies have shown that CTA ASPECTS can predict final infarct size and clinical outcome better than NCCT ASPECTS, in the work done by Lee et al. (2020), they assess the value of each phase of a mCTA comprised of NCCT, arterial phase CTA (CTA-AP) and delayed phase CTA (CTA-DP) in predicting final infarct core and clinical outcome in AIS patients undergoing endovascular treatment (EVT). The final ASPECTS (ground truth) is obtained from DWI performed 36 h after treatment. By computing the correlation between pretreatment ASPECTS (NCCT, CTA-AP and CTA-DP) and final DWI ASPECTS, they observe that CTA-DP ASPECTS is a reliable tool to predict final infarct score in patients undergoing EVT of AIS, better than CTA-AP and NCCT (the correlation between CTA-DP ASPECTS and DWI ASPECTS is greater).

These studies show that a CTA acquired some seconds after the peak of arterial enhancement may well increase the amount of stroke information that can be extracted.

2.3. Metrics for the calculation of the difference between distributions

In a lot of situations, it is commonly necessary to compute the difference between two probability distributions for a given variable. For instance, in machine learning, it is often required to calculate the difference between an observed probability distribution and an actual one.

This can be performed using methods from information theory, like the Kullback-Leibler Divergence (KL Divergence) and the Jensen-Shannon Divergence (JS Divergence), which is a normalized and symmetrical version of the KL divergence.

KL divergence. The Kullback-Leibler divergence (Kullback, 1959), also referred to as relative entropy, between two distributions P1 and P2 is calculated as the negative sum of the probability of each event in P1 multiplied by the logarithm of the probability of the same event in P2 divided by the probability of the event in P1 (Equation 1).

$$KL(P1 \parallel P2) = \sum_{x \in \mathcal{X}} P1(x) \log\left(\frac{P1(x)}{P2(x)}\right) \qquad (1)$$

In the case of intensity probability distributions, an event corresponds to an intensity value. Therefore,

when the probability for an intensity from P1 is high, but the probability for the same intensity in P2 is small, the divergence is high. The logarithm can be base-2 or base-e giving "bit" or "nats" units. A score of 0 indicates identical distributions. A positive divergence indicates different distributions.

JS divergence. The Jensen-Shannon divergence (Lin, 1991) uses the KL divergence to compute a normalized, symmetrical score (Equation 2). Since it is symmetrical, the divergence from one probability P1 to another probability P2 is the same as the divergence of P2 from P1.

$$JS(P1 \parallel P2) = \frac{1}{2}KL(P1 \parallel M) + \frac{1}{2}KL(P2 \parallel M) \quad (2)$$

where M is $\frac{1}{2}(P1 + P2)$ and KL is the KL divergence.

JS distance. The Jensen-Shannon distance (JS distance) is calculated as the square root of the Jensen-Shannon divergence.

These metrics can be used for providing a score that represents the difference between the intensity probability distributions of the stroke region and the symmetric contralateral region in a brain CTA of an AIS patient.

2.4. Stroke segmentation / detection from CTA

Stroke lesion segmentation from neuroimaging has been a research area in the medical image analysis field. Different machine-learning based models have been proposed to segment stroke lesions in an automated way (Lucas et al., 2018), (Zhang et al., 2018), (Kuang et al., 2021). However, considering the scope of the literature searched, no model has been found to automatically segment stroke from brain CTA scans.

Deep Symmetry-Sensitive Convolutional Neural Network (DeepSymNet) is a model designed for predicting if a patient has suffered from AIS (Barman et al., 2019), (Sheth et al., 2019). Based on the fact that the two hemispheres of the brain are visibly different in CTA images in an ischemic scenario, and inspired from the paradigm of Siamese networks, the model compares the two hemispheres of the brain, which are processed in parallel, to detect ischemic stroke from CTA brain images.

A Siamese network, originally proposed in Bromley et al. (1993), uses identical neural networks with the same weights to learn the differences and the similarities between two or more inputs.

Instead of using a cost function to compare the outputs of the identical networks as most proposed models do, DeepSymNet employs convolutional layers to learn the differences contained in the output of the identical networks.

The model takes as inputs the CTA images of the two hemispheres of the brain. Two identical convolutional neural networks (CNNs) are used to learn low and high level features of the 3D volumes. Then, a merge layer calculates the difference in absolute value between the outputs of the CNNs. By adding further convolutional layers, the model is aimed at learning the asymmetry information contained in the difference obtained by the merge layer. Finally, after max-pooling, the output predictions (stroke or no stroke) are obtained through a fully connected layer.

Since DeepSymNet provided good results (AUC 0.914) in recognizing stroke from CTA images, an adaptation of this model is used in the current work for segmenting the stroke core from specific time scans of CTP images (equivalent to CTA scans).

3. Material and methods

In this section, the data used in this study is described, followed by the description of the procedures that have been implemented, which include:

- Data preparation
- Extraction of the infarct core region and the contralateral region (and corresponding hemispheres)
- Difference between intensity distributions
- Core segmentation
- Core segmentation evaluation

3.1. Data

The data used in this study was acquired in the University of Iowa Hospitals & Clinics, Iowa City, United States. 112 acute ischemic stroke patients who underwent whole brain CT perfusion acquired at acute phase and MRI diffusion-weighted imaging acquired after successful recanalization were included in this study. Data from one patient of the IOWA dataset can be seen in Figure 1.

As for the CTP, all scans were done with 40 ml of nonionic iodinated contrast (Isovue-370) followed by 50 ml of saline. The CTP protocol included a rapid sequential scanning, with 4 scans each 3 seconds (s) apart followed by 15 scans 1.5s apart and another 9 scans 3s apart, totaling 28 scans over approximately 60s (Limaye et al., 2019).

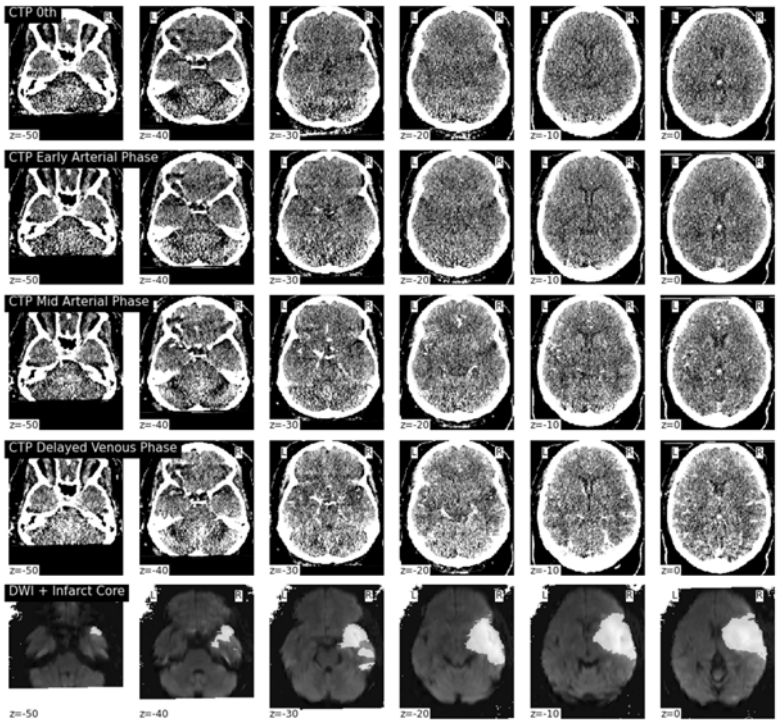


Figure 1: Axial slices of 4 different instances of time of the CTP (4 first rows: 0th time, early arterial phase, mid arterial phase and delayed venous phase) and axial slices of the MRI DWI with the manual segmentation overlayed in white (last 5th row) of one patient of the dataset. Images are aligned to the MNI space.

Even though the same protocol was aimed to be followed, the number of scans varied from subject to subject depending mainly on the tolerance of the patient. In the used dataset, this number ranges from 8 to 30 scans.

Each CTP scan consisted of 512x512x158 voxels with a slice thickness of 1 mm and a resolution of 0.39 mm on the x-y plane.

RAPID charts (iSchemaView Inc, Menlo Park, CA) were also obtained from the CTP data. Based on arterial and venous contrast vs. time concentration curves obtained by RAPID, three time points, simulating a three-phase CTA (early arterial phase, mid arterial phase and delayed venous phase), were selected from the CTP of each patient.

The gold standard, manual infarct core segmentation, was delineated by a medical doctorate student on MRI DW imaging.

The DWI data consists of 128x128x22 voxels with a slice thickness of 7 mm and a resolution of 1.875 mm on the x-y plane.

3.2. Data preparation

3.2.1. Image conversion

MRI DWI scans and the corresponding manual infarct core segmentations were given in Nifti format, while CTP images were in DICOM format.

CTP data was converted from DICOM format (.dcm) to compressed Nifti format (.nii.gz). This conversion was performed using "dcmstack", a package for stacking DICOM images into multi dimensional arrays.

3.2.2. Image registration

Image registration was performed in order to align the CTP and the DWI data.

"SimpleElastix", which is an extension of "SimpleITK" that allows to configure and run "Elastix" program, was used to perform the automatic intensity-based registrations.

In all cases, a rigid affine registration with multi resolution strategy was performed. Advanced normalized correlation was used as the metric for the transformation optimization.

DWI to CTP. For each patient, DWI was registered to CTP. Since CTP consists of various scans, a processed

version of the median across time of the CTP scans (median CTP) was used as the fixed image for the registration process. The processing performed to the median CTP consisted in setting to 0 intensity values lower than 0 or higher than 95 Hounsfield Units (HU) and smoothing using a Gaussian filter.

This processing steps were aimed at obtaining a fixed image on the CTP space that is more similar to a DWI in terms of intensity values. In DW imaging the skull and the vessels are not enhanced, while in contrast-enhanced CT imaging, the skull and the vessels have very high intensity values.

The transformation parameters found in the DWI-to-CTP registration were used to align the manual infarct core segmentation to the CTP space.

After this process, all data belonging to the same patient was aligned in the same space.

CTP to MNI. In order to have all the data in the same space, data was registered to an image previously aligned to the MNI space, a standard space for the human brain.

Specifically, the fixed image corresponded to the median taken from a number of CTA scans previously registered to the MNI space.

For each patient, the moving image used in this registration process was a processed version of the median across time of the CTP scans (median CTP). The processing consisted of setting to 0 all negative intensity values and smoothing using a Gaussian filter. This processing was aimed at increasing the similarity between the moving and the fixed image. Compared to the previous registration process, in this case, in both fixed and moving images, the skull was represented by very high intensity values. Therefore, high intensity values were not removed in this case.

The transformation parameters found in the CTP-to-MNI registration for each patient were used to transform all the CTP acquisitions, the DWI and the manual infarct core segmentation to the MNI space.

After this process, all data was aligned to the MNI space.

CTP registration across time. In order to correct possible motion artifacts, CTP scans already registered to the MNI space were registered across time.

The median of the CTP images across time was used as the fixed image to which all CTP scans were registered.

Registration corrections. After the automatic registration process, the resulting transformed images were visual inspected and registration corrections were performed if considered necessary.

Registration corrections were made using 3D Slicer software (Fedorov et al., 2012).

3.3. Extraction of the infarct core region and the contralateral region (and corresponding hemispheres)

The amount of stroke information that can be extracted for each CTP scan was quantified based on what experts do when visually inspecting CT brain scans for stroke detection, by comparing the two hemispheres of the brain. Since the CTP images are equivalent to CTA scans acquired at different times, with lower quality, though, they will be referred to as CTA scans.

In particular, for each CTA, the intensity distribution of the following region pairs was compared:

3.3.1. Infarct core region and contralateral region

Core region. For each CTA scan, all the voxels manually segmented as infarct core in the corresponding DWI image were extracted.

Contralateral region. The contralateral region was obtained by taking the symmetric region on the contralateral hemisphere.

3.3.2. Stroke hemisphere without the stroke region and corresponding contralateral region

The intensity difference was also analyzed in the healthy tissue of the brain.

The healthy tissue extraction was made on a preprocessed brain without the skull and the backgorund. After removing the voxels corresponding to skull and background based on intensity thresholding, grayscale morphology was used to fill the possible holes (black voxels) that could have appeared. Holes are local minima in the grayscale topography that are not connected to boundaries of the image. Gray level values adjacent to a hole were extrapolated across the hole.

Core hemisphere. The voxels on the affected hemisphere were extracted without considering the affected region, that is, the infarct core region.

Contralateral hemisphere. The voxels on the contralateral healthy hemisphere were extracted also without considering the contralateral region symmetric to the core region.

3.4. Difference between intensity distributions

With the aim of quantifying the amount of stroke information that can be obtained from CTAs acquired at different times, the difference between the "core region" and the "contralateral region" was analyzed for each CTA scan. This difference was also obtained for the "core hemisphere" with respect to the "contralateral hemisphere".

Jensen-Shannon distance (JS distance) was used to calculate the difference between intensity distributions. For each patient, the JS distance was calculated for each CTA scan, and the instance of time corresponding to the maximum value was obtained.

For each patient, the following variables calculated from the CTA were plotted against the scan start time:

- Median intensity of each of the four regions ("core region", "contralateral region", "core hemisphere" and "contralateral hemisphere").
- Median intensity difference between the two region pairs ("core region" vs "contralateral region" and "core hemisphere" vs "contralateral hemisphere").
- JS distance of "core region" vs "contralateral region" and "core hemisphere" vs "contralateral hemisphere".

The obtained curves were analyzed visually. Based on this visual inspection, motion artifacts present on the curves were examined and patient discard was performed if considered appropriate.

The time corresponding to the maximum JS distance was obtained for each patient. Also, the median along the patient dimension of this time, calculated with respect to the mid arterial phase time, was computed. This was used to have an equivalent time to maximum JS distance for all the patients, and it will be referred to as the relative time of maximum JS distance.

Then, the JS distance values were compared among CTAs corresponding to 6 particular time points:

1. 0th time point. Equivalent to a NNCT.
2. Early arterial phase.
3. Mid arterial phase.
4. Delay venous phase.
5. Time of maximum JS distance.
6. Relative time of maximum JS distance.

The comparison was done on these selected time points since they correspond to enhancement phases equivalent among patients, except for point 5. The time corresponding to CTA providing the maximum JS distance represents the "optimal" CTA that can be obtained for each patient, thus, specific to each patient. The relative time of maximum JS distance (point 6) is the patient-equivalent version of the previous one. The 0th time point was used to have an equivalent non-contrast CT scan. The three phases, early arterial, mid arterial and delayed venous, correspond to the time points provided by the RAPID charts.

For the 6 CTAs of each patient, the JS distance between the two region pairs was obtained.

Statistical difference. Mann–Whitney U test was used to compute the statistical significance of the difference between the JS distance values. It consists of a nonparametric test of the null hypothesis used to test whether two samples are likely to derive from the same population.

Using this test, for both cases ("core region" vs " contralateral region" and "core hemisphere" vs "contralateral hemisphere"), the JS distance was compared between all the possible pairs of the 6 above-mentioned time points.

Also, for each of the 6 time points, the statistical difference between the JS distance "core vs contralateral" and the JS distance "core hemisphere vs contralateral hemisphere" was analyzed.

3.5. Core segmentation

Automatic segmentation of the infarct core was performed from each of the 6 CTA scans listed in 3.4. Thus, in total 6 experiments were performed:

1. Segmentation from 0th time CTA.
2. Segmentation from early arterial phase CTA.
3. Segmentation from mid arterial phase CTA.
4. Segmentation from delay venous phase CTA.
5. Segmentation from CTA providing the maximum JS distance.
6. Segmentation from CTA corresponding to the relative time of maximum JS distance.

DeepSymNetV2. The segmentation was based on an adapted version of DeepSymNet classification model, referred to as DeepSymNetV2 (Figure 2):

- Input: CTA patch (3D) of certain window size and symmetric contralateral patch (flipped).
- Output: classification of the central voxel of the patch into infarct core or not infarct core.

In this architecture (Figure 2), a VGG module (VM) consists of two consecutive convolutional layers with 3x3x3 filters (24 filters are used in total) followed by a max pooling layer of 2x2x2 size. The difference layer

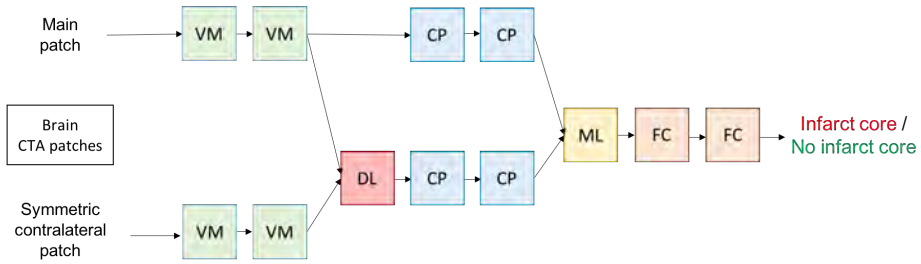


Figure 2: DeepSymNetV2. (VM: VGG module, DL: difference layer, CP: convolution followed by max pooling, ML: merge layer, FC: fully connected layer).

(DL) calculates the difference between VM outputs. A CP module is composed by a convolutional layer with 3x3x3 filters (8 filters are used in total) followed by max pooling (2x2x2). ML stands for merge layer and consists of a concatenation operation. Two consecutive fully connected layers (FC) preceed the final output layer with the prediction of the class (infarct core / not infarct core). The convolution outputs are activated using ReLU. The final FC that leads to the output probabilities uses softmax activation.

DeepSymNetV2 was used to classify the central voxel of 3D patches of brain CTAs into infarct core or not infarct core. In order to balance the computational cost with an appropriate size to capture the stroke region, a window size of 31x31x31 was used to extract patches from CTA data.

The dataset was split into training (70% of the data) and test sets (30% of the data). 20% of the data of the training set was used for validation. The same train-validation-test split was used in all experiments.

For training the model, not all the patches from the training set, but a selected subset, were used. Specifically, both classes (infarct core and not infarct core) were equally represented in the training subset. This way, class imbalance present in the whole training set was avoided.

The input of the model consists of the main patch, from which the central voxel is aimed to be classified, along with its symmetric patch on the contralateral hemisphere. The contralateral patch is flipped in order to facilitate the network learning process.

Two identical convolutional neural networks (CNNs), with identical weights, are used to learn the low and high level 3D patch representations. This CNNs architecture employs 2 VMs one after the other. Then, a difference layer (DL) calculates the difference between the high-level convolution filter outputs common to the two patches. Afterwards, two modules composed by a convolutional layer followed by max pooling layer (CP) are used to learn information about the asymmetry of the two patches. Two other CP modules are used to learn further information from the VM outputs of the main patch. The outputs from the CP modules are then merged (ML), simply concatenated, and the merged result is then connected to two fully connected layers (FC) leading to the final output prediction layer, outputting logits with values from 0 to 1.

For all experiments, Adam optimizer with a learning rate of 1E-4 was used to minimize the classification loss function, binary cross-entropy. The model was trained for 100 epochs. Early stopping was implemented by monitoring the validation average precision (AP), a metric that summarizes the precision-recall curve as the weighted mean of precision values achieved at each threshold (Equation 3).

$$AP = \sum_n (R_n - R_{n-1}) P_n \tag{3}$$

where P_n and R_n are the precision and recall at the nth threshold.

Therefore, for each experiment, the best model was selected as the one providing the highest AP score in the validation set.

3.6. Core segmentation evaluation

For each experiment, the test set was used to evaluate the segmentation model performance.

Core segmentation from patch classification. Due to computation related issues, not all the brain voxels of each test patient were inferred. Only the central voxel of windows covering the brain with a 50% of overlap were inferred. Afterwards, in order to get segmentation probabilities for the whole brain, not inferred voxels were assigned the classification prediction of their nearest neighbor classified voxel.

Once the probabilities were obtained for each test brain, F1 score (Equation 4) was calculated between

the ground truth segmentation and different thresholded segmentation results. Since the network outputs probabilities, to obtain a binary segmentation result, the probabilities need to be thresholded. F1 score is equivalent to a metric commonly used to evaluate segmentation results, the Dice Coefficient:

$$F1score = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (4)$$

where an F1 score reaches its best value at 1 and worst score at 0.

"F1 score vs threshold" curves were obtained for each patient in the test set. A median curve along the patient dimension was calculated and from that, the maximum F1 score and corresponding threshold, referred to as "best threshold", was obtained.

Statistical difference. For each test patient, an F1 score value was obtained using the specific "best threshold", for each experiment. The F1 score distribution on the test set was statistically compared between experiments using the Mann-Whitney U test.

4. Results

4.1. Image registration

Some misregistration results were observed after the automatic alignment processes, thus, manual registrations using 3D slicer were performed on those cases. Regarding the alignment to the CTP space, the registration was reperformed for 10 cases. Concerning the alignment to the MNI space, the registration was reperformed for 5 cases.

Due to high motion observed between the CTP scans and because of bad registration results difficult to recover, 2 cases were discarded from the study. This led to a dataset consisting of 110 subjects.

4.2. JS distance

The JS distance between the "core region" and the "contralateral region" and the JS distance between the "core hemisphere" and the "contralateral hemisphere" was obtained for the 110 subjects of the registered dataset.

Due to lack of information regarding the three phases (early arterial, mid arterial and delayed venous), 18 cases were further discarded for the subsequent experiments, leading 92 patients in total.

The plots obtained for some of the patients can be seen in Figures 3 and 4.

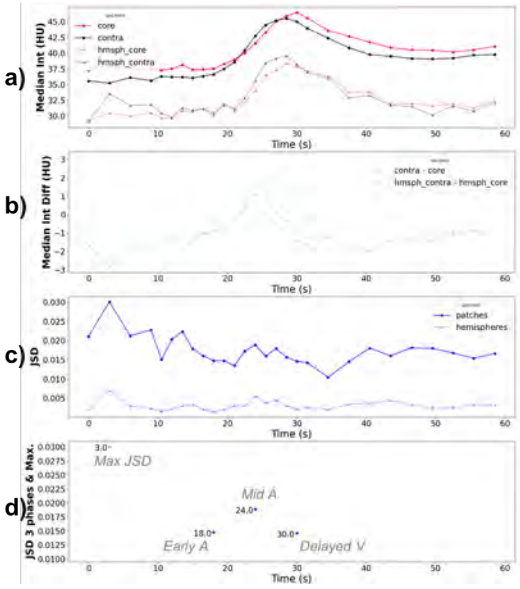


Figure 3: Patient SAS-0905: a) Median intensity (HU) of the core region (red), the contralateral region (black), the core hemisphere (light red) and the contralateral hemisphere (gray). b) Difference of the median intensity between the contralateral and the core regions (light blue) and between hemispheres (lighter blue). c) JS distance between the core and contralateral regions (patches in the plot legend) and hemispheres. d) JS distance between the core and contralateral regions corresponding to the three phases (early arterial, mid arterial and delayed venous phase) (blue points), and the green point, the value corresponding to the time of maximum JS distance. The number next to the colored points is the corresponding time in seconds.

Inspecting the curves from patient "SAS-0905" (Figure 3), it can be observed that, at the beginning (first time points), the curves are not smoothed, rather, they change abruptly from one time point to another. By further inspecting the CTP registered data using 3D slicer software, motion was observed in the first CTP scans of this patient, which led to the decision of discarding this patient for the subsequent experiments (segmentation task).

Other similar cases (5 in total) were observed and also discarded. This led to a final dataset of 87 patients.

In most of the cases, 49 out of 87, the CTA providing the highest JS distance between the "core region" and the "contralateral region" corresponds to a scan start time between the mid arterial phase and the delay venous phase, as in Figure 4 (green point in the 4th row graph) (see Table 1).

4.3. Relative time to maximum JS distance

The time corresponding to the mid arterial phase was used as a reference to compute the relative time of maximum JS distance.

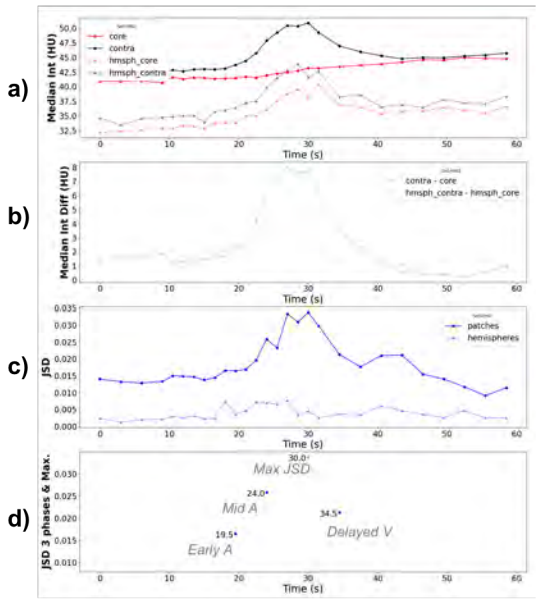


Figure 4: Patient SAS-0902: a) Median intensity (HU) of the core region (red), the contralateral region (black), the core hemisphere (light red) and the contralateral hemisphere (gray). b) Difference of the median intensity between the contralateral and the core regions (light blue) and between hemispheres (lighter blue). c) JS distance between the core and contralateral regions (patches in the plot legend) and hemispheres. d) JS distance between the core and contralateral regions corresponding to the three phases (early arterial, mid arterial and delayed venous phase) (blue points), and the green point, the value corresponding to the time of maximum JS distance. The number next to the colored points is the corresponding time in seconds.

Table 1: Time of acquisition of the CTA providing the highest JS distance between the "core region" and the "contralateral region".

CTA scan time	**Cases**
Before early art. phase	1
= Early art. phase	1
Between early and mid art. phases	3
= Mid art. phase	10
Between mid art. and delayed ven. phases	49
= Delayed ven. phase	5
After delayed ven. phase	18

The median, along the patient dimension, of the relative time of maximum JS distance is 3 seconds with respect to the mid arterial phase (Table 2). Therefore, this time is used to take for each patient an equivalent acquisition that approximates the optimum one (the one that maximizes the JS distance).

Table 2: Quantile values of the relative time to maximum JS distance (time of maximum JS distance minus time of mid arterial phase).

Quantile [0 - 1]	0	0.25	0.5	0.75	1
Time [s]	-3.0	.5	**3.0**	6.0	19.5

4.4. Statistical difference of JS distance

The statistical analysis of the JS distance is presented in this section by box plots for each distribution and the statistic test result above each plot. The following p-value annotation is used:

- ns: 5E-02 <p <= 1
- *: 1E-02 <p <= 5E-02
- **: 1E-03 <p <= 1E-02
- ***: 1E-04 <p <= 1E-03
- ****: p <= 1E-04

4.4.1. Stroke region vs. Contralateral region

The statistical difference between each pair of CTAs of the JS distance (JSD) between "stroke region" and "contralateral region" can be seen in Figure 5.

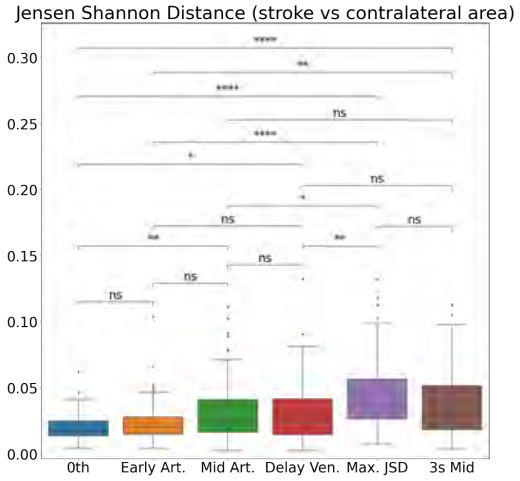


Figure 5: Box plots of the distribution of the JS distance between regions for each of the 6 different CTAs.

CTA providing the maximum JSD vs Others. It can be observed that the JSD distribution from the CTAs acquired at the instance of time of maximum JS (Max. JSD, purple in Figure 5) is significantly higher than all the others except for 3s Mid, which is its equivalent version (CTAs of the relative time of maximum JS, brown in Figure 5).

NCCT (0th) vs Others. The JSD values provided by the equivalence of a NCCT (0th time CTA) are significantly lower than the values obtained by all the other CTAs, except for the early arterial phase CTA. Both 0th and early arterial phase CTAs provide lower JSD compared to the other CTAs.

Mid arterial phase CTA vs Others. Mid arterial JSD distribution (in green) is not significantly different from the ones corresponding to the delayed venous CTA, early phase CTA and NCCT. However, it is significantly lower than the maximum JSD distribution (Max. JSD, in purple).

CTA acquired 3s after the mid arterial phase vs Others. The JSD values obtained by the CTA scan corresponding to the relative time of maximum JSD (3s Mid) are significantly higher than the ones obtained with a NCCT and an early phase CTA scan. However, they are not significantly higher than the ones obtained with a mid arterial phase CTA, which is the currently used target CTA scan.

The highest median JSD value corresponds to the Max. JSD distribution (Figure 5).

4.4.2. Stroke hemisphere vs. Contralateral hemisphere

The statistical difference between each pair of CTAs of the JS distance (JSD) between "stroke hemisphere" and "contralateral hemisphere" can be seen in Figure 6.

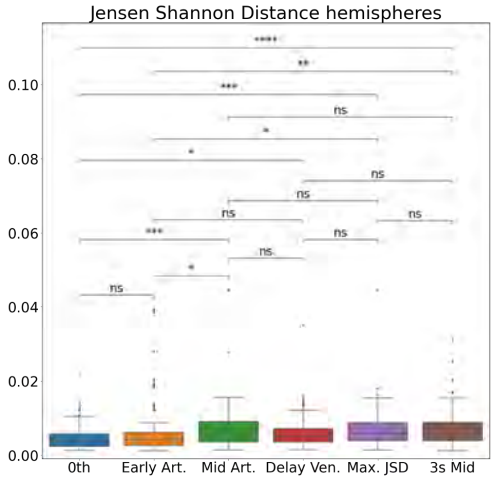


Figure 6: Box plots of the distribution of the JS distance between hemispheres for each of the 6 different CTAs.

NCCT (0th) vs Others. As it happened in the difference between regions, between hemispheres, the JSD values provided by the equivalence of a NCCT (0th time CTA) are significantly lower than the values obtained by all the other CTAs, except for the early arterial phase CTA.

Mid arterial phase CTA vs Others. The JSD distribution obtained from the mid arterial phase CTA is only significantly different (higher) with respect to the NCCT and the early arterial phase CTA.

CTA providing the maximum JSD vs Others. The Max. JSD distribution is only significantly different from the early arterial phase JSD and the NCCT JSD and it is very similar to the CTA corresponding to 3s after the mid arterial phase (3s Mid).

The highest median JSD value corresponds to the mid arterial phase distribution (Figure 6).

4.4.3. Stroke regions vs. Stroke hemispheres

As it can be noticed in Figure 7, for all CTAs, the JS distance between the "core region" and the "contralateral region" is significantly higher than the JS distance computed between the "core hemisphere" and the "contralateral hemisphere".

4.5. Stroke segmentation

The results of the segmentation process can be seen in Table 3.

For all 6 experiments, the best validation average precision achieved was higher or equal than 0.75. The highest validation average precision was achieved when training from mid arterial phase CTAs and CTAs corresponding to the relative maximum JSD (0.84 in both cases).

As for the AUC achieved on the validation set, in all cases, it was higher than 0.8, being the highest 0.87 achieved in the mid arterial phase CTA experiment.

Regarding the evaluation of the segmentation on the test set, the maximum F1 score obtained from the median "F1 score vs threshold" curve across the test set can be seen in Table 3. Although high classification performance was achieved in the training and validation sets, noticeably, F1 score values obtained for the reconstructed test segmentations are very low. The highest F1 score is 0.24, for the relative time of maximum JSD experiment.

The results obtained for one of the test examples for the mid arterial phase experiment are shown in Figure 8.

As an issue observed in many other patients for all 6 experiments, from the axial slices in Figure 8, it can be seen that the network is able to predict quite well the location of the stroke, since high probabilities are

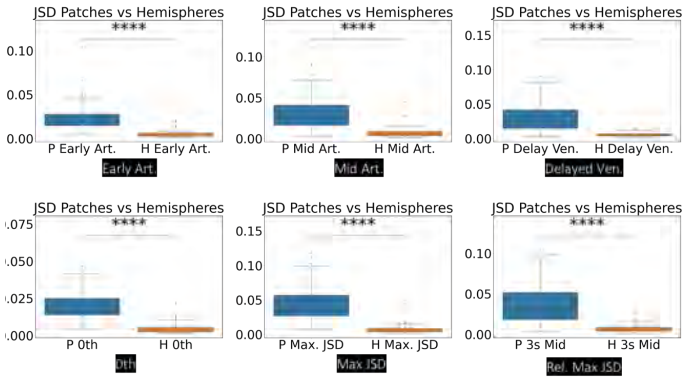


Figure 7: Statistical comparison of the JS distances between "regions" and "hemispheres" for each of the 6 analyzed CTAs.

Table 3: Best validation average precision and corresponding validation AUC (rows 2 and 3) of the validation brain patches classification results. Maximum F1 score (row 3) using the "best threshold" specific to each experiment (row 4) of the test segmentation results.

CTA	0TH	EARLY ART.	MID ART.	DELAYED VEN.	MAX. JSD	REL. MAX. JSD
Validation Average Precision [0-1]	0.75	0.80	**0.84**	0.82	0.82	**0.84**
Validation AUC [0-1]	0.80	0.81	**0.87**	0.83	0.82	0.85
Max. F1 score of the median curve [0-1]	0.1	0.2	0.19	0.2	0.15	**0.24**
"Best" threshold [0.000 - 1.000]	0.715	0.756	0.690	0.884	0.737	**0.833**

observed in the core region. However, a lot of false positives can be observed around the skull area. Some false negatives are observed too.

The poor F1 score values are mainly due to the high quantity of false positives present around the brain skull observed in many cases.

A statistical comparison of the F1 score values for the test patients obtained for each experiment can be seen in Figure 9, where it can be seen that there is no significant difference in any case except between the F1 score distributions obtained from the "mid arterial phase" and the "0th" time. That is, the distribution of F1 scores of the test set computed from the mid arterial phase CTA is significantly higher than the one obtained when using NCCT scans. However, the values of this metric, as commented before, are very low in all cases.

5. Discussion

This study focuses on optimizing stroke segmentation using acute brain CTA. To our knowledge, there is no method described in the literature that analyzes the CTA scan protocol with the specific aim of extracting higher stroke information and thus helping the segmentation process.

In this work, this is tackled by examining the difference in intensity between the stroke region in the affected brain hemisphere and the symmetric region in the contralateral hemisphere. This analysis of difference is based on what experts do when visually inspecting CT-based scans for detecting stroke, which consists of comparing visually both hemispheres of the brain to detect the subtle changes that are caused by ischemia. In an ischemic scenario, the blood flow to a region of the brain is restricted. Therefore, during the injection of contrast enhancement material, which travels through the blood vessels, less enhancement will be observed in the stroke area compared to the rest of the brain, which is normally perfused, as in healthy conditions.

Jensen Shannon distance, a metric based on Kullabck-Leibler divergence, is used in this work to quantify the difference between intensity distributions. This metric has been chosen as it is a robust metric to compare two distributions that was originated in information theory and used in many other applications, such as in the machine and deep learning fields.

In most of the patients analyzed in this work, a higher Jensen Shannon distance is obtained in CTAs acquired between the mid arterial phase and the delayed venous phase. It is worth to notice that this difference is significantly higher between the "stroke region" and the "contralateral region" than between the "stroke hemisphere" and the "contralateral hemisphere". This is important considering that the proposed classification model bases its decision by comparing each patch to its symmetrical contralateral patch. Therefore, if the JS distance

Figure 8: Test example. Axial slices of the mid arterial pahse CTA (in the background), ground truth core segmentation (overlayed in pale orange) and the reconstructed network probabilities (overlayed in gray scale, from black to white). Examples of false positives in red frame; examples of true positives in green frame.

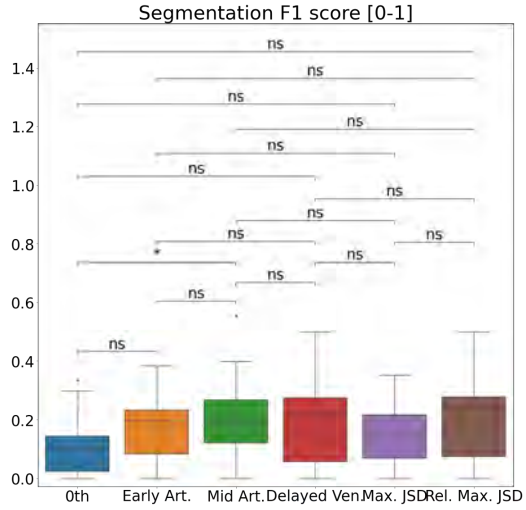


Figure 9: F1 score distributions on the test set for each of the 6 experiments and corresponding statistical difference.

between the ”core region” and the ”contralateral region” was not significantly higher than the JS distance between the ”core hemisphere” and the ”contralateral hemisphere”, the network would not be able to detect the stroke area and differentiate it from the rest of the brain.

Then, if using a CTA scan acquired between the mid arterial and delayed venous phases is associated to a better automatic stroke segmentation, something not proved in this study, then, the CTA protocol should be changed according to that. Instead of targetting the peak arterial enhancement phase, CTAs should be acquired some seconds after this phase. However, it should also be pointed out that the time required for contrast material to reach the target vasculature varies among patients. Therefore, instead of providing the same time to scanning for all subjects, a patient-specific variable should be calculated. However, this becomes more complex since it requires the prior knowledge or prior analysis of the patient-specific circulation timing.

Deep learning based methods have yield very good results in many image medical problems. Stroke lesion segmentation has been a research of study and different methods have been proposed to automatically segment stroke core from neuroimaging. However, these methods have designed their models using NCCT, CTP or MRI DWI images. On the contrary, in this study we focus on CTA, because it is an imaging modality used in many medical centers, more feasible than CTP and MRI for evaluating AIS patients at acute phase and better than NCCT in the detection of stroke, as reported in previous studies.

CTA protocol is aimed at capturing the peak arterial enhancement phase. The CTA scan start time is an essential variable in the context of this work, since depending on this time, the enhancement of the brain tissue varies and so does the difference intensity between the stroke region and its corresponding symmet-

rical area. Although not proved in this study due to the poor segmentation results obtained, it is likely that a model designed for segmenting infarct core from CTA will benefit from the difference that can be extracted from the stroke region and its corresponding contralateral region. In particular, providing a CTA with a higher difference is expected to be associated to a higher segmentation performance.

In this work, a segmentation model has been trained, validated and tested with the aim to test this hypothesis. However, poor segmentation results have been obtained, limiting the comparison among the different CTA scans (with different scan start times).

The low values of the F1 score metric used to evaluate the segmentation results are mainly caused by the false positives observed around the area of the brain skull. Since the brain is not symmetric, compared to patches inside the brain, patches around the area of the skull are likely to be more different than their symmetric patches on the contralateral hemisphere. A patch from the skull may take more voxels, or less, corresponding to background compared to its corresponding symmetrical patch. Background intensities are different from the ones shown inside the brain. This may well lead to a misclassification of the patch as infarct core, since the proposed network is focused on detecting differences between patches. Furthermore, since the proposed classification network (DeepSymNetV2) only sees patches but not the whole brain, it is not able to learn where the skull is and thus it is not able to tackle this problem.

Future work includes trying to reduce this false positives so as to obtain robust results and being able to perform a comparison among CTAs acquired at different instances of time. A potential way to approach this issue is to add at a second stage of the segmentation process a U-Net model. The input of the U-Net would be slices of the brain CTAs along with the inferred probabilities of those slices. The aim of the U-Net would be to learn the skull location and learn not to classify voxels from that area as infarct core.

Despite the obtained low F1 scores, by visual examination of the segmentation results, it has been observed that in many cases the location of the stroke is well predicted by the network. This demonstrates the power of the adapted version of the DeepSymNet (inspired by the paradigm of siamese networks) in the detection of stroke even from low-quality images, as the ones used in the current study.

Another aspect to point out is that the proposed model has been trained, validated and tested using the same data split. Instead, in order to utilize the data better, cross-validation technique should be considered in future work.

It is also worth mentioning that the CTP data used in this study has followed the same acquisition protocol in all cases. Then, image quality, which is quite low (lower signal-to-noise ratio compared to other imaging modalities), is the same in all CTAs and NCCT since these scans come from the same CTP acquisition for each patient. Therefore, a fair comparison (not misled by external aspects such as image quality, protocol and machine used) between variables extracted from these scans can be performed.

6. Conclusions

This study proposes a way to optimize stroke segmentation using acute brain CTA. Brain CTP data obtained for 87 AIS patients is used and treated as sequential brain CTA scans.

In particular, the amount of stroke information that can be extracted from acute brain CTA is analyzed with respect to the CTA scan start time.

Based on the Jensen Shannon Distance, the results show that the best CTA for higher stroke information extraction corresponds to a scan start time that is approximately 3 seconds after the mid arterial phase. This best CTA maximizes the difference between the stroke region and its corresponding symmetrical region on the contralateral hemisphere.

An automatic segmentation model is designed to automatically segment the infarct core from CTA. The results obtained, in terms of F1 score, are very low though. Further improvement of the segmentation performance is required for being able to robustly compare the stroke segmentation from different CTA scan start times and test the main hypothesis of this project: whether stroke segmentation can be optimized using a CTA providing higher difference between the stroke region with respect to the rest of the brain.

7. Acknowledgments

I would like to thank my supervisor Dr. Luca Giancardo for his passion in sharing his knowledge and experience and for the guidance throughout this work. To Dr. Sunil Sheth for his help in the clinical part of the project and to the medical doctorate student, Grant Meeks, for his work in the manual segmentation of the images. Also, to Dr. Santiago Ortega of the University of Iowa Hospitals & Clinics for sharing the data used in this project.

References

ACR, ASNR, SPR, 2020. ACR–ASNR–SPR practice parameter for the performance and interpretation of cervicocerebral computed tomography angiography (CTA).

Barman, A., Inam, M., Lee, S., et al, 2019. Determining Ischemic Stroke From CT-Angiography Imaging Using Symmetry-Sensitive Convolutional Networks. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) , 1873–1877doi:10.1109/ISBI.2019.8759475.

Bivard, A., Levi, C., Krishnamurthy, V., et al, 2015. Perfusion computed tomography to assist decision making for stroke thrombolysis. Brain 138, 1919–1931. doi:10.1093/brain/awv071.

Bonney, P., Walcott, B., Singh, P., et al, 2019. The continued role and value of imaging for acute ischemic stroke. Neurosurgery 85, S23–S30. doi:10.1093/neuros/nyz068.

Bromley, J., Guyon, I., LeCun, Y., et al, 1993. Signature verification using a 'siamese' time delay neural network. Advances in neural information processing systems 6, 737–744. doi:10.5555/2987189.2987282.

Brouns, R., De Deyn, P., 2009. The complexity of neurobiological processes in acute ischemic stroke. Clinical Neurology and Neurosurgery 111, 483–495. doi:10.1016/j.clineuro.2009.04.001.

Cademartiri, F., van der Lugt, A., Luccichenti, G., et al, 2002. Parameters affecting bolus geometry in CTA: a review. Journal of computer assisted tomography 26, 598–607. doi:10.1097/00004728-200207000-00022 6.

Camargo, E., Furie, K., Singhal, A., et al, 2007. Acute brain infarct: detection and delineation with CT angiographic source images versus nonenhanced CT scans. Radiology 244, 541–548. doi:10.1056/NEJMoa0804656.

Catanese, L., Tarsia, J., Fisher, M., 2017. Acute ischemic stroke therapy overview. Circulation research 120, 541–558. doi:10.1161/CIRCRESAHA.116.309278.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., et al, 2012. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging 30, 1323–1341. doi:10.1016/j.mri.2012.05.001.

Garcia-Tornel, A., Campos, D., Rubiera, M., et al, 2021. Ischemic Core Overestimation on Computed Tomography Perfusion. Stroke 52, 1751–1760. doi:10.1371/journal.pone.0239510.

Hacke, W., Kaste, M., Bluhmki, E., et al, 2008. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. New England journal of medicine 359, 1317–1329. doi:10.1056/NEJMoa0804656.

Hinzpeter, R., Eberhard, M., Gutjahr, R., et al, 2019. CT angiography of the aorta: contrast timing by using a fixed versus a patient-specific trigger delay. Radiology 291, 531–538. doi:10.1148/radiol.2019182223.

Johnson, W., Onuma, O., Owolabi, M., Sachdev, S., 2016. Stroke: a global response is needed. Bulletin of the World Health Organization 94, 634. doi:10.2471/BLT.16.181636.

Kuang, H., Menon, B., Sohn, S., et al, 2021. EIS-Net: Segmenting early infarct and scoring ASPECTS simultaneously on non-contrast CT of patients with acute ischemic stroke. Medical Image Analysis 70, 1361–8415. doi:10.1016/j.media.2021.101984.

Kullback, S., 1959. Statistics and Information theory .

Lee, B., Hwang, Y., Kim, J., 2020. Delayed phase computed tomography angiography ASPECTS predicts clinical outcome and final infarct volume. PLoS ONE 15, e0239510. doi:10.1371/journal.pone.0239510.

Liebeskind, D., 2003. Collateral circulation. Stroke 34, 2279–2284. doi:10.1161/01.STR.0000086465.41263.06.

Limaye, K., Bryant, A., Bathla, G., et al, 2019. Computed tomography angiogram derived from computed tomography perfusion done with low iodine volume protocol preserves diagnostic yield for middle cerebral artery-m2 occlusions. Journal of Stroke and Cerebrovascular Diseases 28, 104458. doi:10.1016/j.jstrokecerebrovasdis.2019.104458.

Lin, J., 1991. Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory 37, 145–151.

Lucas, C., Kemmling, A., Mamlouk, A., et al, 2018. Multiscale neural network for automatic segmentation of ischemic strokes on acute perfusion images. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) , 1118–1121doi:10.1109/ISBI.2018.8363767.

Menon, B., d'Esterre, C., Qazi, E., et al, 2015. Multiphase CT angiography: a new tool for the imaging triage of patients with acute ischemic stroke. Radiology 275, 510–520. doi:10.1148/radiol.15142256.

Nogueira, R., Jadhav, A., Haussen, D., et al, 2018. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. New England Journal of Medicine 378, 11–21. doi:10.1056/NEJMoa17064426.

Pavlina, A., Radhakrishnan, R., Vagal, A., 2018. Role of imaging in acute ischemic stroke. Seminars in Ultrasound, CT and MRI 39, 412–424. doi:10.1053/j.sult.2018.01.002.

Pexman, J., Barber, P., Hill, M., et al, 2001. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. American Journal of Neuroradiology 22, 1534–1542.

Pulli, B., Schaefer, P., Hakimelahi, R., et al, 2012. Acute ischemic stroke: infarct core estimation on CT angiography source images depends on CT angiography protocol. Radiology 262, 593–604. doi:10.1148/radiol.11110896.

Reidler, P., Stueckelschweiger, L., Puhr-Westerheide, D., et al, 2020. Performance of Automated Attenuation Measurements at Identifying Large Vessel Occlusion Stroke on CT Angiography. Clinical neuroradiology , 1–10doi:10.1007/s00062-020-00956-5.

Sheth, S., Lopez-Rivera, V., Barman, A., et al, 2019. Machine learning–enabled automated determination of acute ischemic core from computed tomography angiography. Stroke 50, 3093–3100. doi:10.1161/STROKEAHA.119.026189.

Wahlgren, N., Moreira, T., Michel, P., et al, 2016. Mechanical thrombectomy in acute ischemic stroke: Consensus statement by ESO-Karolinska Stroke Update 2014/2015, supported by ESO, ESMINT, ESNR and EAN. International Journal of Stroke 11, 134–147. doi:10.1056/NEJMoa0804656.

der Worp, H.V., van Gijn, J., 2007. Acute ischemic stroke. New England Journal of Medicine 357, 572–579. doi:10.1056/NEJMcp072057.

Wouters, A., Dupont, P., Christensen, S., et al, 2016. Association between time from stroke onset and fluid-attenuated inversion recovery lesion intensity is modified by status of collateral circulation. Stroke 47, 1018–1022. doi:10.1016/j.clineuro.2009.04.001.

Zhang, R., Zhao, L., Lou, W., et al, 2018. Automatic Segmentation of Acute Ischemic Stroke From DWI Using 3-D Fully Convolutional DenseNets. IEEE Transactions on Medical Imaging 37, 2149–2160. doi:10.1109/TMI.2018.2821244.

Automatic Registration of Pre-Operative CT-Scan and Surgical Video of Augmented Reality System for Ear Surgery

Ali Taleb, Alain Lalande

Dijon, France

Abstract

Transtympanic surgical procedures are a form of minimally invasive surgery and involve penetration of micro-instruments through a small opening in the tympanic membrane to access middle ear cleft structures. Augmented Reality system for handling Transtympanic procedure, requires registration between pre-operative CT-scan and surgical video as a preliminary step. In our work we propose a fully automatic free registration pipeline. The algorithm works by detecting contours then registering through Iterative closet point algorithm. Results were within the specified surgical tolerance with mean overlay error of 0.565 mm and a maximum duration of 18 seconds. The proposed registration method does not rely on any external fducial markers attached to the patient and performs automatically with an acceptable processing duration. In parallel we studied the contribution of persistent homology in the same problem. Results were interesting with more work currently in progress.

Keywords: Registration, Augmented Reality, Contour Detection, Iterative closest point, Persistent homology, Transtympanic

1. Introduction

Augmented reality (AR) is about enriching the sensory perception of the user and enhancing the reality with convenient information. The enhancement can take a form of visual, audio, haptic, etc. Such technology is immersing in real time surgical video, usually in the form of a combination of pre-operative data with real time one. AR is under development to be used in surgical room. It can be delivered via screens, speakers, gloves or co-manipulated robots, etc. The main contribution of AR in surgery is the ability to see through structures and access hidden information without interfering with the surgical process (Hussain et al., 2020b). It could significantly improve critical tasks during surgical operation such as localization, efficiency and safety. It also has the potential to facilitate minimally invasive procedures by allowing surgeons to visualize structures without exposing them (Bernhardt et al., 2017; Marroquin et al., 2018). The main advantage that AR based procedures offer over traditional image guided procedures is the significant improvement in ergonomics of the system. With AR, everything can be available on a single view thus eliminating the need for the surgeon to go back and forth between different sensorial systems.

AR has been successfully applied in different surgical domains (Hussain et al., 2020b; Vávra et al., 2017; Yoon et al., 2018). However, very few applications have been designed for otological procedures. Despite offering many advantages such as rigid bony structures and limited inter-structural movements, limitations in workspace and maneuverability coupled with high precision requirements have been the main hindrance in developing AR solutions for otological procedures.
Transtympanic surgical procedures are a form of minimally invasive surgery and involve penetration of micro-instruments through a small opening in the tympanic membrane to access middle ear cleft structures, such as ossicles and labyrinthine windows 1 (Hussain et al., 2018). These procedures have been mainly designed for diagnostic and therapeutic purposes and have been employed for ossicular chain repair, drug administration, labyrinthine fistula diagnosis, cholesteatoma removal, vertigo treatment, and electrocochleography. This approach offers many advantages over traditional approaches such as preservation of tympanic membrane, reduced bleeding, faster procedure and less painful post-operative course as the tympanomeatal flap is not elevated and remains at its position throughout the process (Marroquin et al., 2018).

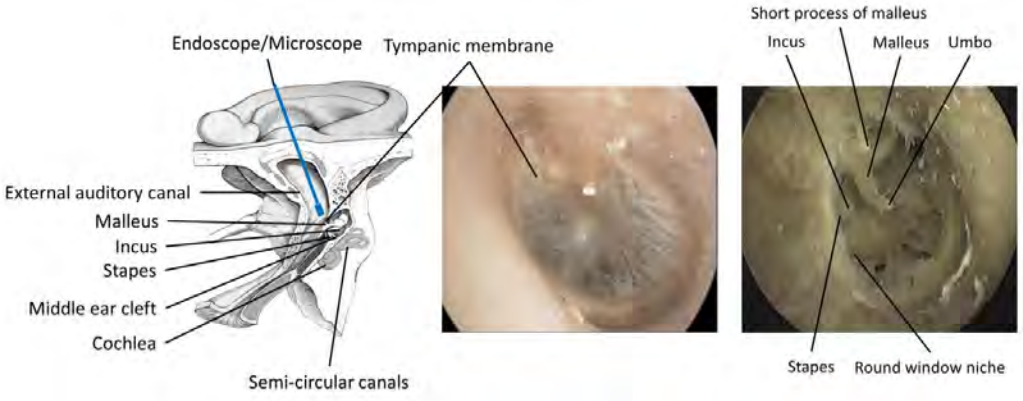


Figure 1: Schema of Transtympanic procedures

Until today, these procedures have been achieved by introducing both an endoscope and the surgical instruments through a puncture hole in the tympanic membrane. Such procedures introduce complications into the surgical process by reducing the available operative space and field of view and limiting the instrument movements (Hussain et al., 2020b). Indeed, the instruments and the endoscope must be very carefully maneuvered in order to avoid tympanic membrane tear at the penetration point. Alternatively, visualization of middle ear structures through an intact tympanic membrane by optical instruments outside the middle ear is very limited since the eardrum has a poor translucency and may be even totally opaque in pathological cases. Recently, a vision-based AR system, was proposed for transtympanic procedures without any external tracking system (Transtympanic Augmented Reality System) (Kakehata, 2013). The system was designed to visualize middle ear cleft structures behind the tympanic membrane by overlapping a virtual endoscopy image derived from preoperative computed tomography (CT) over the tympanic membrane. The system employed a semi-automatic fiducial marker-based endoscope CT registration followed by movement tracking based on image feature processing algorithms.

Problem Analysis. Our problem constitutes in registering a preoperatively acquired CT scan to an endoscopic image. The structure of the ear behaves in a rigid manner, however considering orientation of both source and target image, this will introduce additional skewing parameters thus ending with a projective transformation with 8 degrees of freedom.
The fact that we aim to automate the procedure, our first goal is to integrate information from both modalities without the use of any fiducial markers. In previous work there has been several algorithms applied between multi modal medical image acquisition (MRI – CT – ultra . . .) which are usually more robust to feature extraction. However, in cranial base regions, which is our case, anatomical landmarks are not so apparent, and complications may be introduced in selecting them. Feature may vary within subject or cross subject. Also, accuracy poses another critical issue where acceptable margin error in other work would not be acceptable in ours due to the size and sensitivity of the proposed tissue. Upon our research only few approaches were found to be fully automatic, marker less, registering to an endoscopic image without the aid of any extra hardware and non were applied on the temporal bone area.

Target. After analyzing the problem and exploring the possibility of using features and similarities between images, we set our target to first semi automate the procedure through Iterative Closest Point (ICP) algorithm. Next step would be fully automating it through contour detection neural network. At each step we will explore the possibility of using Persistent Homology (PH) within our process.

2. Background

2.1. Registration

Registration presents an essential role in AR. It is an early crucial process as any error incurred during this step will propagate throughout the whole procedure. Registration is about transforming an initial object (image or set of points) to match another. It is usually represented by a transformation matrix.
Navigation data could be obtained intraoperatively or

preoperatively. The principle of the first is a registration is to obtain a rigid transformation between different imaging modalities acquired at the same time. This approach needs intraoperative imaging equipment such as CBCT and external tracking devices with markers (Chaudhury, 2020). In contrast, preoperatively acquired data could introduce more complex deformations that could be general or local.

Depending on the tissue under study and the nature of acquisition registration could be rigid or non-rigid (aka deformable, local). Soft tissues require a nonrigid registration that considers any local deformation that could occur due to tissue movement for (Wang et al., 2019), while rigid registration considers only global transformation and normally requires less degrees of freedom which result in less computational costs. Such deformations could be even resolved within a fully automatic technique as long as features and mutual information modalities could be clearly extracted (Tan et al., 2020).

In both cases of rigid or non-rigid, registration could be manual, semi-automatic or automatic. While the future focuses on removing the man from the loop, leading to better performance and less error, the emerge towards automating the whole procedures is more under study. A manual registration is performed by a surgeon or expert by selecting or aligning a set of corresponding points in the fixed and the target image based on a qualitative analysis. A semi-automatic introduces a step that can assist the process by means of better performance, accuracy, or computational costs but still requires the intervention of an expert. The step could be followed or preceded by well-known algorithms such as ICP or NICP (Lee et al., 2019) or even learning based algorithms (Alam et al., 2017). While automatic registration does not require any intervention, still in some cases it is required to use invasive external marker fixation which may also require external tracking devices with bulky tracking sensors.

Approaches for registration could be segmentation based (Zhang et al., 2019), feature based (aka sparse) (Schneider et al., 2020), intensity based (aka dense) (Liu et al., 2017). The main steps for any registration procedure are: feature detection, feature matching, computing mapping functions and resampling.

2.2. Persistent Homology

Persistent homology is a method for computing topological features of a space at different spatial resolutions. More persistent features are detected over a wide range of spatial scales and are deemed more likely to represent true features of the underlying space rather than artifacts of sampling, noise, or particular choice

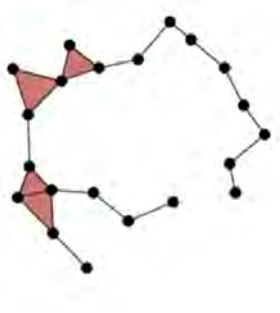


Figure 2: building a complex by indicating all connected components. 1-simplex are connected by an edge and 2-simplex are colored

of parameters. The study of this concept started in late 1980's and till now is an active research area. In a simple illustration of the theory, a space is composed of vertices, edges, triangles and tetrahedrons... Each named as k-simplex. Simplices to be connected if the distance between the components is greater than a threshold. A complex is a set of simplices which we can build by connecting the vertices and then mark or fill complete graphs 2.

The term homology represents the count of holes in the space aka Betti number, in a certain dimension. For instance, a H_1 homology represents the number of 1-dimensional holes (circular holes), H_2 represents the 2-dimensional holes (cavity in a tetrahedron)... Persistent stands for how long this hole persist in function of the variability of the threshold distance. Usually, the output could be a barcode diagram or a persistent diagram (PD) which later could be endowed with metrics.

For simplicity, we will demonstrate with a set of four points 3. Filtration is applied with an euclidean distance d starting from 0 we observe four 0-simplices. While increasing the value of d we keep the same observation until we reach a value of $d = 2$ of which then we observe two 1-simplices implying 2 edges. For $d = 3$ the graph is formed of 1-simplex and we witness the birth of the first hole. Upon reaching a value of $d = \sqrt{1}3$ the previous hole vanishes.
Notice that at $d = 3$, a Betti value equals to 1 is produced, and this hole was persistent between the interval $[3, \sqrt{1}3]$.
Representing those information could take different forms, one of which is the PD 4, where for each point we can associate a corresponding death and birth. Birth is the value where the hole was created and death when it was destroyed. Hence none of the points in a PD could appear below the identity line $y = x$.

Note that a 0-dimensional homology is defined as the number of connected components a shape has. In prac-

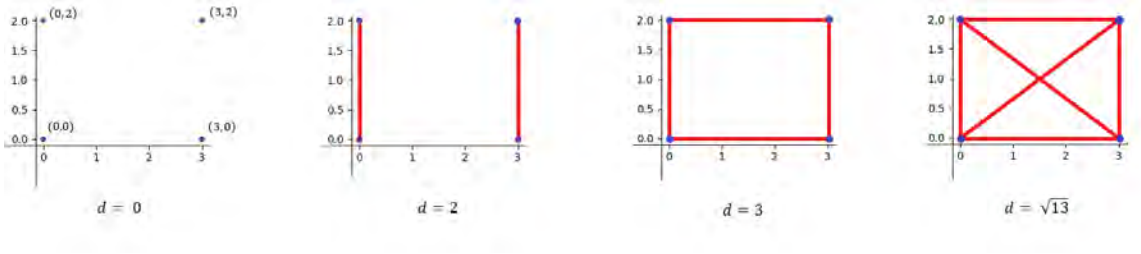


Figure 3: Demonstration of persistence homology on four points. Filtration with euclidean distance in a coordinate system. Red segments represent connected components.

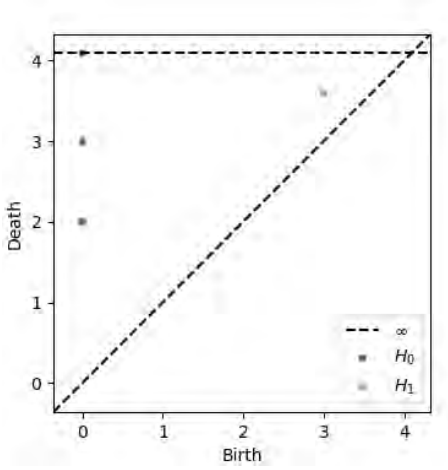


Figure 4: Persistence Diagram of set of four points

tice, we will be focusing on two main concepts from the theory. The 0-dimensional homology representing the connectivity of the shape and the 1-dimensional homology representing how many circular holes defined by it. A PD could be endowed with a metric distance (bottleneck for instance). Another option is to transform for a more stable vector representation. By giving the space of persistence diagrams a metric structure, a class of effective machine learning technique can be applied. Several representations have been proposed, one of which we rearrange entries of the distance matrix between points in a PD which yields to more stable representation in terms of slight variations (Adams et al., 2016; Reininghaus et al., 2015). Another form is superimposing a grid over a PD and counting number of points in each bin (Chepushtanova et al., 2015).

Applications of PH in medical applications. The contribution of PH in medical applications has been under study. One of which PH was used to better describe brain white matter in Alzheimer disease (Kuang et al., 2020). Gamble and Heo (2010) proposed PH as a method for statistical analysis of human Jaws and distinguish clinically relevant treatment effects in orthodontistry.

In Image processing, PH has been used with segmentation tasks. In an application of tumor segmentation in histological images, it was proposed to classify patches as tumor or non-tumor, and this by exploring the connectivity between cells (Qaiser et al., 2016). A homogenous connectivity is associated with a more homogenous barcode or PD.

Another approach was embedding the PH in the loss function (Clough et al., 2021). Importantly this process might be used with no requirement of truth labels, however a prior topological information of the object being segmented is required.

3. State of the art

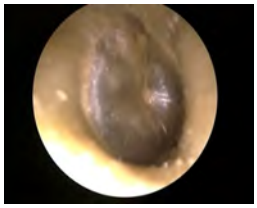
To the best of our knowledge, no other work has been published for transtympanic procedures. Other similar studies in otology have been restricted by the use of either fiducial markers or manual identification of anatomical landmarks (Hussain et al., 2020a). In our previous work, manual registration yielded a mean registration overlay error of 0.31 mm (Hussain et al., 2020c). In otology, owing to minuscule structure and technological limitations, an error of 1-2 mm with a registration duration less than 1 minutes have been suggested as clinically acceptable (Hussain et al., 2020b).

4. Material and methods

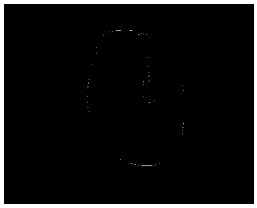
4.1. Data

Dataset formed of two image modalities, CT-scans and endoscopic video frames. CT-scans of spatial resolution range 0.14x0.14x0.15 to 0.44x0.44x5 mm^3.

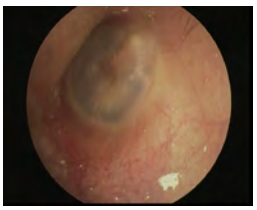
For a total of 41 patients, 58 ears were selected (29 left



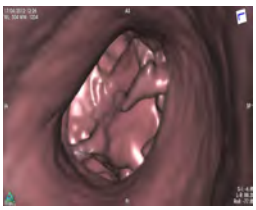
(a) Cadaver Endoscopic Frame



(b) Ground Truth



(c) Patient Endoscopic Frame



(d) Patient CT-Scan

Figure 5: Data Sample Images and annotations

	CT-Scans	Endoscopic Frames
Cadavers	17	167
Real Patients	48	186
Total	65	353

Table 1: Count of Images in training Data

and 29 right). Each ear is associated with one virtual endoscopy (VE) and multiple endoscopic frames chosen randomly from a surgical video.
VE is selected by an expert from a 3D model, reconstructed from its corresponding CT scan using Osirix software. According to its projection angle the appropriate VE is selected to match the endoscopic video frames.
Patients ages range from 5 to 83 and average 51.57. An expert surgeon annotated corresponding contours of tympanic membrane and malleus handle in both the CT image and video frames. Ground truths are used as contours rather than masks (segmented areas). Utilizing masks will change the problem into a segmentation with higher priority on the edges. Contours as labels might present a high misbalance problem, but it was found interesting to approach the problem in this manner.
In the training data a set of cadavers images were added, thus constituting 47 % of the training dataset. For validation and testing, only endoscopic images were used therefore simulating the real scenario. 4 patients were used for validation and 6 for testing.

4.2. Methods

Process pipeline 6 starts with a neural network for detecting the contours of the tympanic membrane and malleus. The output is processed for enhancement, noise removal and point sampling. ICP algorithm registers both modalities and then outputs a registration matrix.

Note that the below order does not reflect the pipeline flow, it represents our work order. We started with semi-automatic registration, U-Net and then post processing.

4.2.1. Semi Automatic Registration

The algorithm of ICP is the process of matching each point in a fixed data cloud (F) into its corresponding nearest distance point in a moving point cloud (M) (Chen and Medioni, 1992). The algorithm iteratively revises the transformation needed to minimize the cost function. We implement an algorithm variant from scratch since none of the libraries did include a solution for 2-dimensional and projective transformation. The cost function used is the Euclidean distance.
Implementing and testing semi registration was performed on ground truth contours.

Algorithm. We modify the search algorithm of closet point to seek the least distance in both directions from F to M and M to F 1. For each point cloud we iterate through its elements and associate its corresponding closet point from the other cloud.
If one point was associated with multiple matches, the one with the highest distance was eliminated. This step will produce a stable correspondence and clear convergence between points over iterations.

We estimate the unknowns in the registration matrix with least squares and outliers are removed with M-estimator sample consensus (MSAC) algorithm (Torr and Zisserman, 2000). Parameters in MSAC (confidence and distance) are set to constant values .

Listing 1: ICP algorithm

```
for iter in range(MaxIterations):
    for point in FixedCloud:
        find nearest point
    for point in MovingCloud:
        find nearest point
    Remove points with same matching
    Remove outliers
    Estimate transformation
```

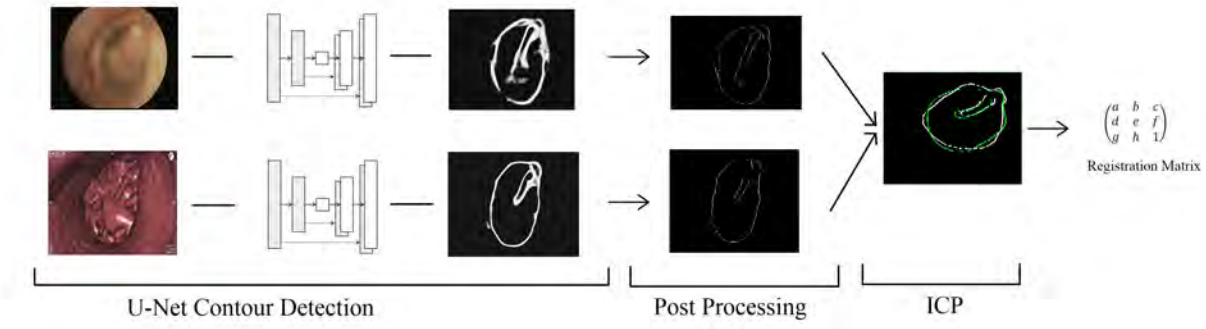


Figure 6: Process Pipeline

Persistence and Initialization. Results may not be identical between runs due to the randomized nature of the MSAC algorithm. In 10 test runs, it is observed one run that diverges. To overcome this issue, an initialization was required to set the algorithm on the right track.
Translation initialization was sufficient in our application. For this purpose, M is transformed by the following matrix 1

$$\begin{pmatrix} 1 & 0 & t_a \\ 0 & 1 & t_b \\ 0 & 0 & 1 \end{pmatrix} \begin{cases} t_a = \frac{\sum x_i}{n} - \frac{\sum X_i}{m} & (x,y) \in n\ fixedpoints \\ t_a = \frac{\sum y_i}{n} - \frac{\sum Y_i}{m} & (X,Y) \in m\ movingpoints \end{cases} \tag{1}$$

Thus, translation is defined by the vector formed by centroids of both clouds.

Selection of the best transformation. It was realized that the algorithm reaches its stable phase after an average of 25 loops. We then propose 70 iterations as the maximum number of iterations.

The iteration with least variance is selected. Variance was preferred over mean of distances, since we seek to minimize distances between all points. A distribution of error over all set is preferred. Such approach did have the impact on hausdorff metric with a little trade off on the mean distance 2 .

	Hausdorf in mm	Mean Distance in mm
Variance	1.01 ± 0.44	0.46 ± 0.15
Sum Distance	1.18 ± 0.43	0.4 ± 0.1

Table 2: Accuracy results of ICP on input ground truth contours

4.2.2. *Contour Detection with Deep Learning*

For the purpose of contour detection, U-Net network was adopted. U-Net is popular in biomedical field and usually associated with segmentation tasks(Ronneberger et al., 2015). The fact that it performs good even with small data sets makes it efficient in medical applications.

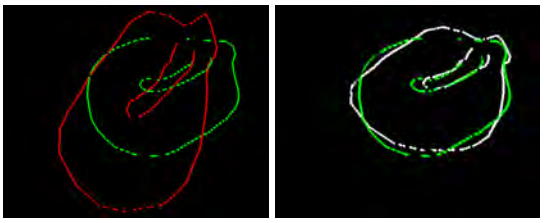


Figure 7: Red: CT initial moving point cloud position.
Green: Endoscopic fixed point cloud position.
White: ICP result transformed CT point cloud

Specifications.

Architecture The network architecture is illustrated in 8. It consists of a contracting path (left side) and an expansion path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3×3 convolutions, each followed by a batch normalization and rectified linear unit (ReLU) and a 3×3 max pooling.

Pre-processing In our work we did not undergo any type of pre-processing except for RGB channels normalizing.

Data Augmentation Several techniques of data augmentation were used. Techniques that mimic the real scenario of data acquisition or sequence of frame in endoscopic video.
Affine transformation with free rotation but limited

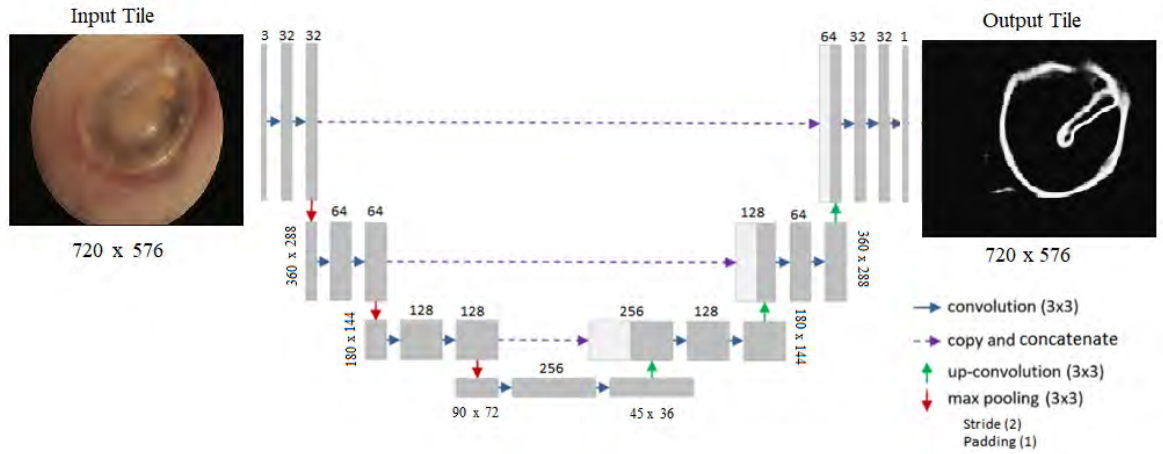


Figure 8: U-Net Architecture for our proposed method

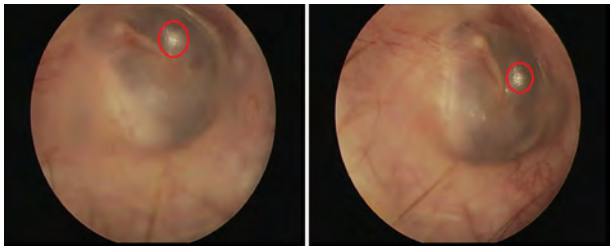
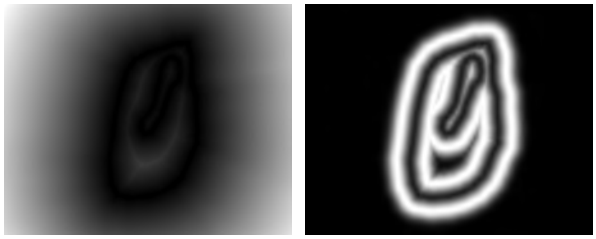


Figure 9: Endoscopic Frames with high contrast spots



(a) Mean distance (b) Gaussian mean distance

Figure 10: Euclidean distance images.

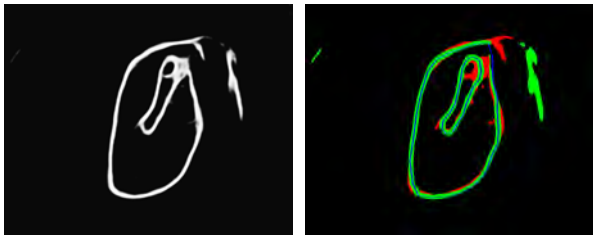
translation so we would not crop the ROI under study, and shear limited to 8°.
Random erasing which imitates the high contrast areas present in some endoscopic images 9

Additional techniques: Salt and pepper noise, Gaussian blur, randomly changing brightness-saturation-contrast, horizontal and vertical flipping.

Metric Depending on the way the distance between predictions and ground truths is defined, evaluation can take many forms. As the output of the U-Net will be processed and inputted to the ICP, we shall analyze ICP requirements for a good performance.
ICP performs the best when the input cloud points are well connected and free of noise in the cloud's moderate neighberhood 11. The moderate neighberhood is defined by the set of points not too near from ground truth nor too far.
The algorithm is robust to noise far from the ground truth, as they are considered as outliers and excluded by MSAC. Also, false detections in the very near neighborhood will be removed as post-processing will include an edge thinning.
A high noise is considered at its peek when miss-classified pixels appear at a distance of 25 pixels from the ground truth.

We started by finding the mean distance between output contours and ground truth. Then it was improved to meet our goal. The mean distance is filtered within a gaussian 10 . Thus, a very low or high mean distance will be minimized. At the end among the best 10 values of gaussian mean distance, the best with true positive detections is selected to enhance the connectivity of the contour. All distances are normalized to the size of the image.

Loss Function Three loss functions were under study and their performances evaluation adapted to our



(a) Network Contour output (b) Noise map

Figure 11: Analysis of the noise with respect to distance from contour. Blue: Ground truth . Green : noise with low impact . Red: noise with high impact.

application is compared.

The standard binary cross entropy BCE with a weight ratio 0.003 for tackling the class imbalance problem.

The focal loss FL which is an improved version of cross entropy. FL tries to handle the class imbalance by assigning more weights to hard or easily misclassified examples (Lin et al., 2018). FL is defined by

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma_t}\log(p_t)$$

Prior value for class imbalance was maintained at 0.01 as proposed in the paper and same values for the parameters α and γ were adopted.

Active Contour loss AC is also experimented. In contrast to BCE and FL, AC considers the region energy minimization rather than pixelwise loss (Chen et al., 2019). The loss is defined by :

$$loss_{AC} = Length \; + \; \lambda \cdot Region$$

in which,

$$Length = \sum_{\Omega}^{i=1,j=1} \sqrt{\left|(\nabla_{u_{x_{i,j}}})^2 + (\nabla_{u_{y_{i,j}}})^2\right| + \epsilon}$$

and

$$Region = \left|\sum_{\Omega}^{i=1,j=1} u_{i,j}(c_1 - v_{i,j})^2\right| + \left|\sum_{\Omega}^{i=1,j=1} (1-u_{i,j})(c_2 - v_{i,j})^2\right|$$

with $c_1 = 1$ and $c_2 = 0$ as the proposed problem is a binary classification

Training. Endoscopic frames and CT-scans with their corresponding contours are used to train the network with Adam optimizer. The learning rate was maintained at a constant value of 5×10^{-3} for endoscopic and 1×10^{-4} for CTs. Using the same size of raw data without resizing, limited the option of a batch size to 16 so we would not bypass allowed RAM size. Loss functions were compared on same parameters. Hardware used titan X GPU.

4.2.3. Post Processing

We aim in post processing to remove noise and then sample the contour into a point cloud. The following techniques are done in the order.

Top-hat 12b : the algorithm computes the morphological opening and then subtracts the result from the original image. This will remove uneven background illumination, particularly the inside area of the malleus.

Thresholding 12c : a constant threshold is applied. The step binarizes the image and removes the easy noise (pixels with low intensity).

Remove non-complete 12d : an operation is performed by sliding a batch of size ($k \times k$) over the image. If it contained any 0 pixel, the whole batch is deleted . In other terms if the sum of pixels in the batch $< k^2$ the pixels are all set to 0.

Thinning 12e : we reproduce the skeleton of the contour through an iterative algorithm (Lam et al., 1992). The result is a thin line but with some small leaf branches.

Remove non-dense 12f : similarly to *remove non-complete* in previous step, we slide a batch over the image. The condition of erasing applies if the batch is not dense. We define a dense batch as the one with minimum number of pixels m. { delete if sum batch $< m$ where $m < k^2$ }.

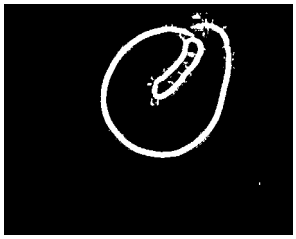
Point Sampling 12g: the result was sampled into point cloud by projecting a grid on the image and uniformly selecting 20 % of the contour.

4.3. Persistent Homology

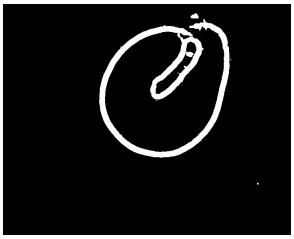
The first attempts were to use PH applied directly to the input image. This has been applied within previous applications but preceded by preprocessing step to binarize the image and transforming into a set of point cloud. In our case this was impossible with high intensity and texture variation in endoscopic images. Several attempts were performed applying edge detection. The anatomical structures under study were not clearly visible. Thus, making it not possible to detect the persistent shape of the malleus or the tympanic membrane. So, PH requires landmarks or a set well described input image. Embedding PH in the contour detection could be achieved if used in the loss function. A differentiable topological layer has been implemented by Brüel-Gabrielsson et al. (2020). The library maps each point

(a) Contour output

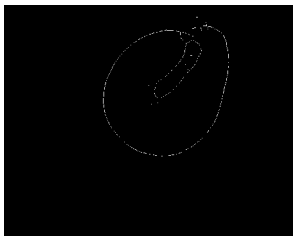
(b) Top-hat



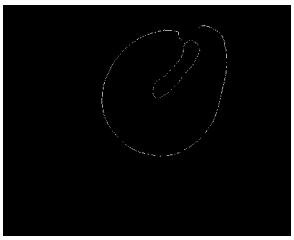
(c) Thresholding



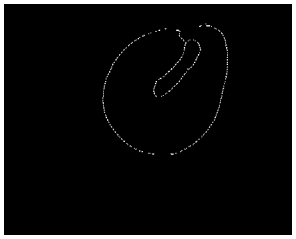
(d) Remove non-complete



(e) Thining

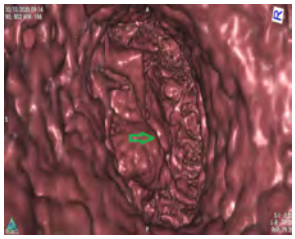


(f) Remove non-dense

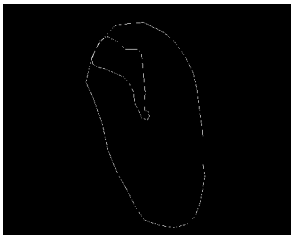


(g) Point sampling

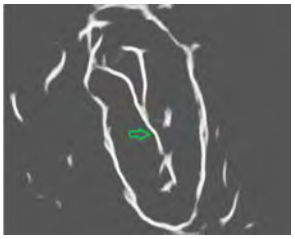
Figure 12: Post-processing steps



(a) CT scan



(b) Ground truth



(c) Contour detection - BCE loss

Figure 13: Contour detection (BCE) with a false detection of the area below the malleus due to the similarity in the edge

in the PD to its corresponding set of points in the point cloud. The library has not been implemented to run on GPUs, which resulted in a long processing time (up to 3 days for a single run). We ran tests only on CT scans being smaller and require less duration of processing. PH was tested in two manners.
First as a regularization to the BCE 2 .

$$loss_{T1} = BCE + \lambda.Topological \tag{2}$$

Recall that 0-dimensional homology represents the connectivity of the elements and 1-dimensional homology represents the holes in a set of points. As pre-implemented by the library, the topological loss function penalizes all elements in the PD except for the most persistent one by returning their sum (the most persistent is the fully connected component representing the contour). The possibility of using PH initially from the start of the training or a late regularization after BCE reaches a stable phase were tried.
The other manner, was to use persistent diagrams as ground truth. The loss function returns the sum of difference square for each element in the PD 3 .

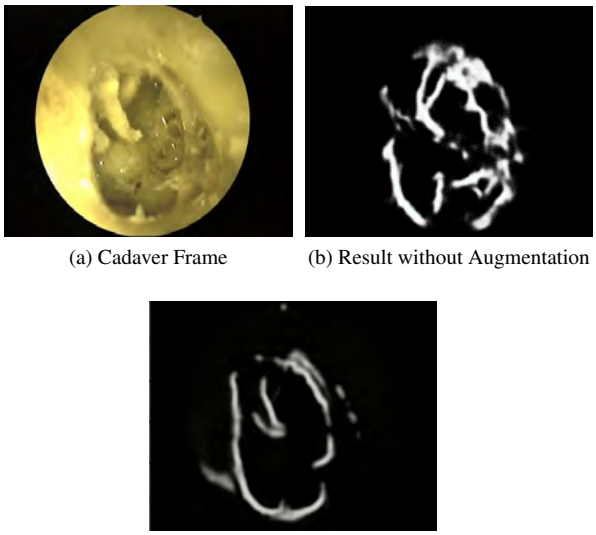
$$loss_{T2} = (x - X)^2 + (y - Y)^2 \tag{3}$$

(x,y) the coordinate of point in persistent diagram associated to its corresponding pixel in output.
(X,Y) the coordinate of point in persistent diagram associated to its corresponding pixel in ground truth.

The model would have the same input and output labels, but at the level of loss function we calculate the persistent diagrams of both and then compare. Class-imbalance was not considered in this approach.
Since the loss relies on the shape of the input, we then expect it to solve one of the issues which is multiple or false detection of the edge of the malleus 13

5. Results

ICP execution process required a duration of 18 seconds. The average count of iterations to reach preferred result was 32 iterations.
Network output results were good enough to be used with ICP. Some endoscopic images introduced a miss-detection of the malleus part 18 . Contours were well connected and wide compared to the ground truth. The effect of data augmentation on the results was significant 14.



(a) Cadaver Frame (b) Result without Augmentation

(c) Result with Augmentation

Figure 14: Data augmentation effect on results

BCE and FL outperformed AC with a slight advantage to BCE . True positive detections reached 94% with CT scans and 86% with endoscopic frames.

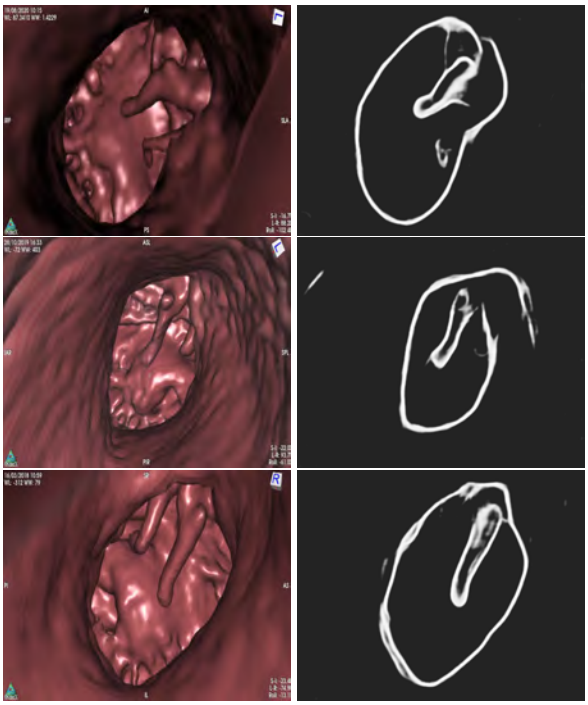
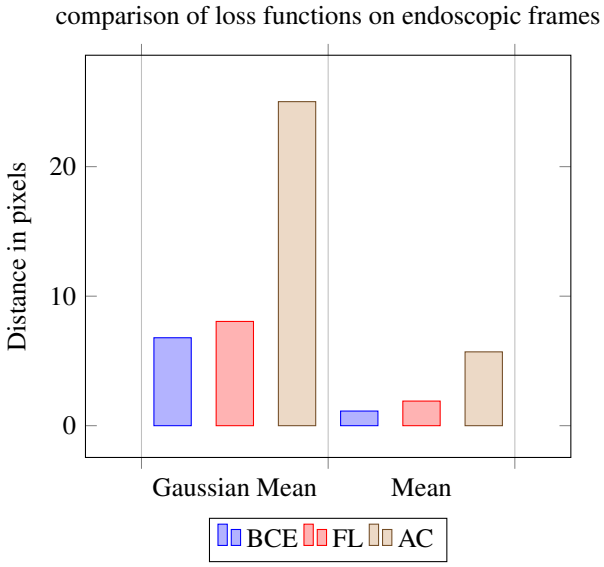


Figure 15: CT scans Contour Detection Results - BCE Loss



For final evaluation, an expert surgeon selected the test patients and annotated two visible anatomies (Round Window and Incus) in CT scans and endoscopic frames. Hausdorff and mean distance were measured. All results were within the specified surgical tolerance 3. Moreover all patients had distance values less than 1mm. except for one patient where hausdorff distance reached 1.8mm.

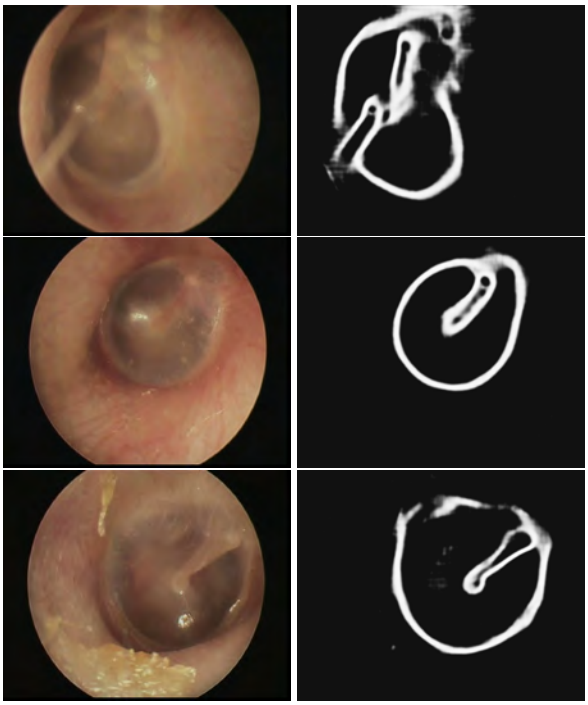


Figure 16: Endoscopic Contour Detection Results - BCE Loss

Hausdorf in mm	Mean Distance in mm
0.88 ± 0.5	0.565 ± 0.495

Table 3: Final Registration error accuracies

For PH work. No quantitative evaluation was done on the results, since we were in the early stages of implementation, but they were visually evaluated.
The test failed for the first loss approach 2.The loss output being the sum could not be normalized and modifying is not permitted by the library. The high values from the topological loss dominated BCE leading to an output converging to null (black image).

For the second approach 3, the output showed a clear and visible shape of the membrane and the malleus but low accuracy on edges. The results can be built on for a further study, yet they were not good enough to be used within the application.
As for the false detection in the area of the malleus, we observed that the edge did not show in the result, but the quality was bad and not to be relied on.

6. Discussion

With some patients it was found that the final pipeline results were better than when using the contours ground truth (instead of neural network) with ICP.

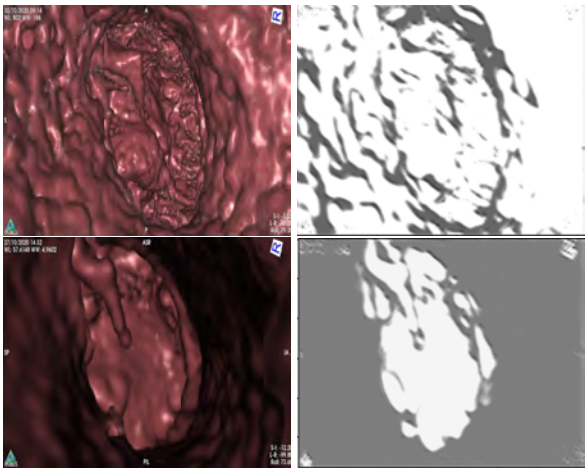
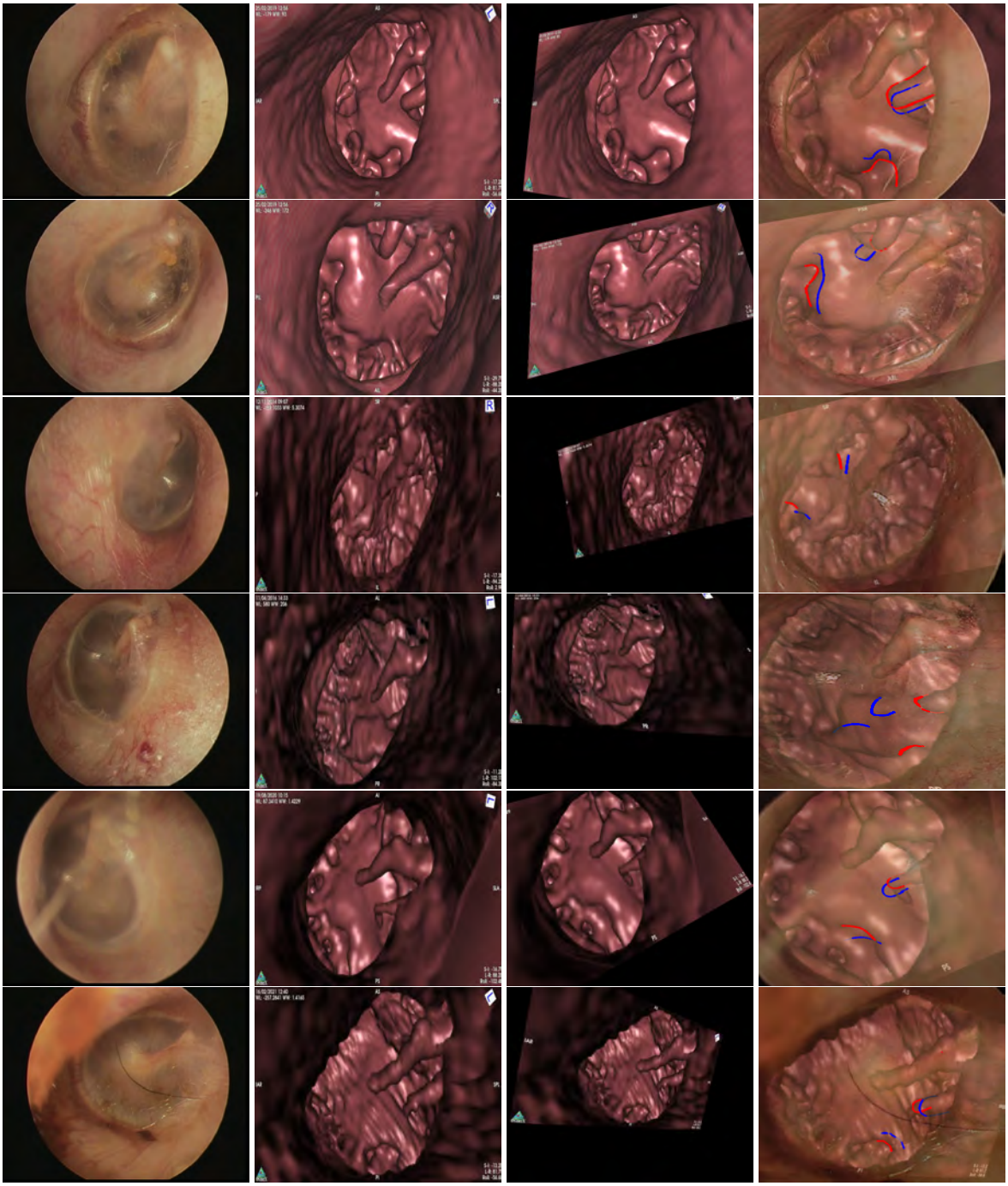


Figure 17: Network Contour Detection with PH loss function. (left: input ; right: output)

The hypothesis behind the results of the patient with maximum error is the choice of virtual endoscopy where its projection angle was distinctive from the endoscopic frame.
The approach of active contour loss could perform better in case of masks as labels, then modifying the contour detection to a segmentation task.

Several steps could be improved to make the registration more robust. Starting with ICP, instead of using the variance in the evaluation of the best iteration another metric or a combination of multiple metrics could be used to improve the results. A rotation initialization should be added as this is essential for persistence of results.
For the neural network used for contour detection, we shall start by trying to merge the information from both CT and endoscopic frame as one input. Another non-pixelwise losses could be used to improve the quality of contour detection results.
Selection of the best virtual endoscopy from CT-scan used for registration should be automated after 3D reconstruction, to better choice compatible with corresponding endoscopic frame. Also, more data could be added, then making the validation set bigger for better tuning of parameters.

PH discussion. One of the causes of the poor results could be the distance metric used on the PD. Add to that the fact we are using the loss without any consideration of the imbalanced class thus making it converging fast on the background and the inside of the shape. In future work it is to be considered to focus on the edge neighborhood.



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

Figure 18: Examples of automatic Registration (a) original frame (b) original CT-Scan (c) Transformed Ct-Scan (d) cropped blended image with Test annotations. Blue: annotations on endoscopic frame. Red: annotations on CT-Scan.

7. Conclusions

In this work we implemented a pipeline of process for automatic registration of pre-operative CT-Scan and surgical video of AR system for ear surgery. The results (average Hausdorff distance 0.88*mm* and mean error of 0.565*mm*) were within the specified surgical tolerance.

8. Acknowledgments

I would like to thank my supervisor, Prof. Alain Lalande, for the patient guidance, encouragement and advice he has provided throughout my internship. I sincerely appreciate Prof. Bozorg Grayeli for the help in thinking up and providing the necessary material required to accomplish this work. Also I would like to thank Prof. Sara Leclerc for the valuable discussions and follow-up.
Many thanks for Dr. Raabid Hussain for the continuous help and guidance throughout the whole period.

References

Adams, H., Chepushtanova, S., Emerson, T., Hanson, E., Kirby, M., Motta, F., Neville, R., Peterson, C., Shipman, P., Ziegelmeier, L., 2016. Persistence images: A stable vector representation of persistent homology. arXiv:1507.06217.

Alam, F., Ur Rahman, S., Hassan, M., Khalil, A., 2017. An investigation towards issues and challenges in medical image registration. Journal of Postgraduate Medical Institute 31.

Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C., 2017. The status of augmented reality in laparoscopic surgery as of 2016. Medical Image Analysis 37, 66–90. URL: https://www.sciencedirect.com/science/article/pii/S1361841517300178, doi:https://doi.org/10.1016/j.media.2017.01.007.

Brüel-Gabrielsson, R., Nelson, B.J., Dwaraknath, A., Skraba, P., Guibas, L.J., Carlsson, G., 2020. A topology layer for machine learning. arXiv:1905.12200.

Chaudhury, A., 2020. Multilevel optimization for registration of deformable point clouds. IEEE Transactions on Image Processing 29, 8735–8746. doi:10.1109/TIP.2020.3019649.

Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y., 2019. Learning active contour models for medical image segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11624–11632. doi:10.1109/CVPR.2019.01190.

Chen, Y., Medioni, G., 1992. Object modelling by registration of multiple range images. Image and Vision Computing 10, 145–155. URL: https://www.sciencedirect.com/science/article/pii/026288569290066C, doi:https://doi.org/10.1016/0262-8856(92)90066-C. range Image Understanding.

Chepushtanova, S., Emerson, T., Hanson, E., Kirby, M., Motta, F., Neville, R., Peterson, C., Shipman, P., Ziegelmeier, L., 2015. Persistence images: An alternative persistent homology representation .

Clough, J., Byrne, N., Oksuz, I., Zimmer, V.A., Schnabel, J.A., King, A., 2021. A topological loss function for deep-learning based image segmentation using persistent homology. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1–1URL: http://dx.doi.org/10.1109/TPAMI.2020.3013679, doi:10.1109/tpami.2020.3013679.

Gamble, J., Heo, G., 2010. Exploring uses of persistent homology for statistical analysis of landmark-based shape data. Journal of Multivariate Analysis 101, 2184–2199. URL: https://www.sciencedirect.com/science/article/pii/S0047259X1000117X, doi:https://doi.org/10.1016/j.jmva.2010.04.016.

Hussain, R., Lalande, A., Girum, K., Guigou, C., Bozorg Grayeli, A., 2020a. Augmented reality for inner ear procedures: visualization of the cochlear central axis in microscopic videos. International Journal of Computer Assisted Radiology and Surgery 15. doi:10.1007/s11548-020-02240-w.

Hussain, R., Lalande, A., Guigou, C., Bozorg-Grayeli, A., 2020b. Contribution of augmented reality to minimally invasive computer-assisted cranial base surgery. IEEE Journal of Biomedical and Health Informatics 24, 2093–2106. doi:10.1109/JBHI.2019.2954003.

Hussain, R., Lalande, A., Marroquin, R., Girum, K.B., Guigou, C., Grayeli, A.B., 2018. Real-Time Augmented Reality for Ear Surgery, in: MICCAI 2018: Medical Image Computing and Computer Assisted Intervention, Granada, Spain. pp. 324–331. URL: https://hal.archives-ouvertes.fr/hal-02308091, doi:10.1007/978-3-030-00937-3_38.

Hussain, R., Lalande, A., Marroquin, R., Guigou, C., Bozorg Grayeli, A., 2020c. Video-based augmented reality combining ct-scan and instrument position data to microscope view in middle ear surgery. Scientific Reports 10. doi:10.1038/s41598-020-63839-2.

Kakehata, S., 2013. Transtympanic endoscopy for diagnosis of middle ear pathology. Otolaryngologic Clinics of North America 46, 227–232. URL: https://www.sciencedirect.com/science/article/pii/S0030666512001715, doi:https://doi.org/10.1016/j.otc.2012.10.006. endoscopic Ear Surgery.

Kuang, L., Gao, Y., Chen, Z., Xing, J., Xiong, F., Han, X., 2020. White matter brain network research in alzheimer's disease using persistent features. Molecules 25. URL: https://www.mdpi.com/1420-3049/25/11/2472, doi:10.3390/molecules25112472.

Lam, L., Lee, S.W., Suen, C., 1992. Thinning methodologies-a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, 869–885. doi:10.1109/34.161346.

Lee, J., Thornhill, R.E., Nery, P., Robert deKemp, Peña, E., Birnie, D., Adler, A., Ukwatta, E., 2019. Left atrial imaging and registration of fibrosis with conduction voltages using lge-mri and electroanatomical mapping. Computers in Biology and Medicine 111, 103341. URL: https://www.sciencedirect.com/science/article/pii/S0010482519302100, doi:https://doi.org/10.1016/j.compbiomed.2019.103341.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2018. Focal loss for dense object detection. arXiv:1708.02002.

Liu, Y., Song, Z., Wang, M., 2017. A new robust markerless method for automatic image-to-patient registration in image-guided neurosurgery system. Computer Assisted Surgery 22, 319–325. URL: https://doi.org/10.1080/24699322.2017.1389411, doi:10.1080/24699322.2017.1389411, arXiv:https://doi.org/10.1080/24699322.2017.1389411. pMID: 29094615.

Marroquin, R., Lalande, A., Hussain, R., Guigou, C., Grayeli, A.B., 2018. Augmented reality of the middle ear combining otoendoscopy and temporal bone computed tomography. Otology & neurotology : official publication of the Ameri-

can Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 39, 931—939. URL: https://doi.org/10.1097/MAO.0000000000001922, doi:10.1097/mao.0000000000001922.

Qaiser, T., Sirinukunwattana, K., Nakane, K., Tsang, Y.W., Epstein, D., Rajpoot, N., 2016. Persistent homology for fast tumor segmentation in whole slide histology images. Procedia Computer Science 90, 119–124. URL: https://www.sciencedirect.com/science/article/pii/S1877050916312133, doi:https://doi.org/10.1016/j.procs.2016.07.033. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016).

Reininghaus, J., Huber, S., Bauer, U., Kwitt, R., 2015. A stable multi-scale kernel for topological machine learning, pp. 4741–4748. doi:10.1109/CVPR.2015.7299106.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597.

Schneider, C., Thompson, S., Totz, J., Song, Y., Allam, M., Sodergren, M., Desjardins, A., Barratt, D., Ourselin, S., Gurusamy, K., Stoyanov, D., Clarkson, M., Hawkes, D., Davidson, B., 2020. Comparison of manual and semi-automatic registration in augmented reality image-guided liver surgery: a clinical feasibility study. Surgical Endoscopy 34. doi:10.1007/s00464-020-07807-x.

Tan, W., Alsadoon, A., Prasad, P., Al-Janabi, S., Haddad, S., Venkata, H.S., Alrubaie, A., 2020. A novel enhanced intensity-based automatic registration: Augmented reality for visualization and localization cancer tumors. The International Journal of Medical Robotics and Computer Assisted Surgery 16, e2043. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.2043, doi:https://doi.org/10.1002/rcs.2043, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcs.2043. e2043 RCS-19-0099.R2.

Torr, P., Zisserman, A., 2000. Mlesac: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding 78, 138–156. doi:10.1006/cviu.1999.0832.

Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., Habib, N., El-Gendi, A., 2017. Recent development of augmented reality in surgery: A review. Journal of Healthcare Engineering 2017, 1–9. doi:10.1155/2017/4574172.

Wang, J., Shen, Y., Yang, S., 2019. A practical marker-less image registration method for augmented reality oral and maxillofacial surgery. International Journal of Computer Assisted Radiology and Surgery 14. doi:10.1007/s11548-019-01921-5.

Yoon, J.W., Chen, R.E., Kim, E.J., Akinduro, O.O., Kerezoudis, P., Han, P.K., Si, P., Freeman, W.D., Diaz, R.J., Komotar, R.J., Pirris, S.M., Brown, B.L., Bydon, M., Wang, M.Y., Wharen Jr, R.E., Quinones-Hinojosa, A., 2018. Augmented reality for the surgeon: Systematic review. The International Journal of Medical Robotics and Computer Assisted Surgery 14, e1914. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1914, doi:https://doi.org/10.1002/rcs.1914, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcs.1914. e1914 RCS-17-0081.R4.

Zhang, X., Wang, J., Wang, T., Xuquan, J., Shen, Y., Sun, Z., Zhang, X., 2019. A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy. International Journal of Computer Assisted Radiology and Surgery 14. doi:10.1007/s11548-019-01974-6.