Eötvös Loránd University

Faculty of Informatics

Dept. of Artificial Intelligence.

Myocardial Perfusion Imaging using Vision Transformers

Supervisor:
Szűcs, Ádám István
PhD Candidate

Author:
Haris Ali
Computer Science MSc

Budapest, 2025

Declaration

I hereby declare that this thesis titled "Myocardial Perfusion Imaging with Vision Transformers" and the work presented in it is my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at Eötvös Loránd University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given.
- I have acknowledged all main sources of help.
- This thesis has not been submitted for any other degree or professional qualification.

Signed: Haris Ali
Date: 13th April, 2025

Acknowledgements

I would like to express my sincere gratitude to my supervisor, **Szűcs, Ádám István**, for his continous support and guidance throughout this research. His expertise were extremely valuable for the development of the project and reach it to completion.

I am also thankful to the faculty and staff of the Department of Computer Science at Eötvös Loránd University, whose valuable support were crucial for the carrying out of the research.

Lastly, I appreciate the open-source community and all the developers whose tools and libraries played a significant role in the implementation of this project.

Thank you all.

Contents

Abstract

The manual delineating the left ventricle (LV) in Myocardial Perfusion Imaging (MPI) is one of the most labour-intensive and time consuming tasks in nuclear cardiology and radiology. The outcome of the diagnosis of the MPI is extremely dependent on the accuracyand the consistency of the segmentation of the ventricles, hence the process is done under extreme caution in roder to minimize the risks of any possible error. However, the process of turning this task into an automated one present a number of challenges that need to bbe mitigated. Fist of all, the signal-to-noise ration (SNR) is mostly low and the resolution of the image is limited, complicating the process of detecting the boundaries. Secondly, the high disparity in both the cardiac traces uptake ad the differences in the hardware used for the imaging introduces inconsistencies. Finally, their is a lack of a standardized definition of the shape of the LV and the there is no standard shape that can be traced based purely on image data, which introduces a lot more ambiguity in the task.

This thesis proposes a novel method built to address the limitations mentioned above by using a Transformer-based architecture, integrating statistical shape prior (SSP) technique. This approach is specifically used to mitigate the data-hungry nature of the transformers in case of limited deta. The proposed architecture achieves over 4% improvementover a number of metrics in segmentation and classification against the benchmarked state-of-the-art (SOTA) approaches use for LV segmentation, both on the synthetic data and the real-world clinical scans.

In addition to the improvements in the quantitative metrics, the incorporation of the prior shape information enabled the model to learn insights into the variability and the structural patterns of the LV anatomy in MPI single-photon emission tomography (SPECT) imaging. This deeper understanding of the LV enhances the reliability of the AI-powered automatic segmentation of the LV and also the general comprehension of the morphology of the LV in clinical practice.

Chapter 1

Introduction

Myocardial Perfusion Imaging (MPI) using single-photon emission computed tomography (SPECT) pllays an important role in the process of non-invasive assessment of the coronary artery disease (CAD). Considerig cardiovascular diseases being one of the leading causes of mortality all across the world, the need for an efficient, accurate and accessible tool for diagnosis is at a high demand. MPI SPECT provides critial information about the perfusion status of the heart, which helps in the early detection and planning the treatment wich improves the outcomes of the patients.

Radionuclide MPI under a specific condition, such as stress, is majorly regarded as one of the most effective diagnosis technique, which ia also non-invasive, in order to identify the or detect the coronary artery disease (CAD). Using the application of MPI SPECT, clinicians become equipped to diagnose and detect the functionally relevant coronary stenoses with a relatively high level of specificity This actually enables them to make decisions that are informed and possibly the right ones recarding the pathways of the patients' treatment [1]. By visualizing the perfusion process of the heart muscles, clinicians can detect the areas of the heart where there is a presence of coronary stenoses or obstructions which may be the causing issue for inducible perfusion deficits under the conditions of stress or rest.This ability of diagnosis is not only essential to identify the patients with CAD but aso functions as an important tool for mitigaing patient risk and guiding the decision making process of the clinicians.

MPI using SPECT has emerged as both an effective and economically viable modality for the purpose of diagnosis. MPI based SPECT offers both the advantages of being accessible and having establined standard clinical protocools hence it is the

preferres choice of a number of diagnostic processes. One of the major strengths of MPI is the adaptability of the technique, as it can incorporate a number of radiopharmaceutical agents, such as 201Tl Chloride, 99mTc Tetrofosmin, and 99mTc Sestamibi, which basically is dependant upon the imaging protocols and imaging needs. The mentioned agents are typically administered intravenously before the image acquisition part, and then the collected image data are later reconstructed using techniques which are dedicatedly designed for cardiac imagery. The last, and possibly the most crucial, stages in the diagnostic process involves the segmentation of the anatomical structures relevant to the diseases and then the reorientation of this segmented volumetric data. This part of the diagnostic is usually performed by trained clinical professionals in order to ensure precision, better reliability and to mitigate the risks of errors.

Beyond the usage of the perfusion imaging alone, there are additional functional parameters, which are valuable, that can be derived when gated acquisition techniques are applied. These marameters include end-systolic volume (ESV), end-diastolic volume (EDV), and the left ventricular ejection fraction (LVEF). All of the mentioned parameters are essential in order to indicate the performance of the heart. The values of these parameters are computed through the precise delineation of the myocardial boundaries of the LV,, which makes the task of segmentation even more crucial in the whole pipeline. The perfusion and the functional analysis collectively provide a detaled understanding of not only the vascular but also the mechanical health of the heart.

Efficient and accurate quantitative analysis of the 3D MPI SPECT data is extremely sensitive to a number of factors that are involved in the full end-to-end imaging and reconstruction pipeline, as mentioned above. All of these factors together contribute not only to the reliability of the evaluation of the data, but also to the detection of a range of cardiac abnormalities [2]. The important step in this process is the segmentation and reorientation of the LV, which basically refers to the determination of te spatial alignment of the LV and its segmentation based on the anatomical midline. The tasks of both reorientation and the segmentation within MPI SPECT imaging have been acknowledged, for a long time, as one of the central yet difficult challenges. Over the course of years, mutltiple commercial systems have been developed in order to counter these issues, but more often relying on very extensive and curated datasets in order to ensure reliable performance in clinical

environments [3], [4], [5].However, the eixisting solutions fall short when they are applied to the newer reconstruction paradigms, especially in the situations where there are only a limited number of labled patients datasets. In order to mitigate these imitations faced by the current solutions and to increase the generalization capabilities of the models under llimited data conditions, approaches incorporating self-supervised learning and few-shot learning have gained popularity. Nevertheless, the effectiveness of these strategies is most of the times overshadowed by the high costs associated with the expert annotations. In addition to this the lack of consensus regarding a standard segmentation protocol also complicate the practical application of the processes.

In the recent years, within the field of MPI SPECT imaging, the adoption of Deep Learning (DL) techniques are looking at a significant revival [6]. This renewal is basically driven in part by the development of the novel radiotracers and also the growing clinical demand to minimize the amount of administered radiation dose and also the image acquisition time of the performed procedures [7]. As a consequence, the modern methods of reconstruction have been focusing on configurations that are based on low photon count data, sparse acquisition views and reduced amount of injected doses [8], [9], [10]. But inspite all that, the advancements do not fully rosolve the challenges which are inherent to the segmentation tasks of MPI SPECT. Despite using state-of-the-art neural network based reconstruction strategies, the segmentation accuracy is still heavily relied on the underlying reconstructed images. When working with lower-dose inputs, the images mostly lack proper structural clarity, which diminishes the benefits which are offered by the DL based reconstruction methods. Even in situations where the image reconstruction achieves are visual equivalence to a full-dose filtered back projection methods, the issues of low Signal-to-Noise Ratio (SNR), Poisson noise characteristics, and the impact of partial volume effect (PVE) continue to affect the generalization capabilities and hence the reliability of automated segmentation models.

In this work, is proposed a novel approach in order to solve the aforementioned bottlenecks of the segmentation task, all the while also constributing further detailed insights into the anatomical characteristics of the MPI SPECT LV. Contrary to the previous approaches employed for the task, where the use of isolated pre-processed regions, or usage of cropped volumes, is common, the method in this study makes use of the entire reconstructed image volumes, hence incorporating all of the contextual

spatial cues which are available within the full field-of-view (FOV). The choice of this design makes sure that no information that is diagnostically relevant is discarded, hence allowing the moel to infer the left ventricular structure in relations to the surrounding regions of the anatomy. This holistic approach increases the robustness of the model, specifilcally in cases where abnormalities in the patterns could possibly interfere with the more localized analysis.

In order to mitigate the limitations hat are associated with singg convolutional neural networks (CNN), specifilcally their receptive field being restricted which hinders them from learning long-range dependencies, the proposed method employes a fully transformer based architecture called nnFormer [11]. This architecture is specifilcally developed for tasks pertaining to volumetric medical imaging. It allows the network to learn global reasoning over the 3D structures which offeres a significant advantage over the traditional CNNs in situations where the boundaries of the organs are not sharply defined such as SPECT. But there is a limitation to using transformer architectures, which is that they require a huge amount of data in order to learn acceptable global representations and have a good generalization ability. Hence, in order to overcome such a limitation, the proposed method incorporates Statistical Shape Priors (SSP) as a regularization technique. Such shape priors introduce an anatomical consistency into the DL model which acts as a guidance signal during the training of the model. This approach helps the models in situations where the amount of avaiable data is limited. Using the shape prios, the model is made to learn meaningfull and spatially coherent segmentation outputs even with minimal amount of supervision. This whole process bridges the gap between the traditional rule-based segmentation models and the fully data driven DL approaches.

Chapter 2

Related Work

The automatic reorientation and segmntation process of MPI SPECT represent steps that are essential for the accurate and and efficient diagnosis andd the quantitative analysis of the heart. A number of commercial softwares have been developed in order to perform these tasks and are widely adopted in clinical practices. The Corridor4DM [12], which was developed at the university of Michigan ,provides a platform for the comprehensive quantitative analysis for Myocardial Perfusion and the functional assesment from SPECT. This extremely integrated system gives access to an automated processing tool for analysis and reporting which si specifically developed to meet the increasing demands of such tools. In a similar way, the Emory Cardiac Toolbox (ECTb) [13] implements an extensive pipeline of quantitative tools which are developed as a result of very extensive research and its validation. It features a database of normal perfusion with more than 150 patients each, the fourier analysis of regional thikening, used for functional assesment, and a very advanced display function that allows to display 3D volumes for image fusion. The Cedars-Sinai approach [14] fouses on an end-to-end automatic expert system which is based on mathematical algorithms and rules based on logic reasoning. The presented QGS software is been used at more than 20 thousand locations all across the globe. The yale method [15] is focused on the quantification process of both the MPI and specifically the LV funcctional abnormalities, which adress the challenges faced by the process by muultiple factors such as background of images and the defects of perfusion using specialized processing techniques.

Other than the mentioned commercial solutions for the tasks, there is significant amount of research efforts that have been devotedly carried out to develop more ad-

vanced approaches and algorithms for MPI SPECT segmentation and reorientation. The Level-Set Methods (LSMs) have been proved to me one of the most developed methods in the field. The research by [16] presented an automatic method for the segmentation of the LV based on variational level sets in volumetri SPECT. This methos integrates adaptive thresholding for initialization with the evolution of variational level set for the determination of the finall contour. Very efective performance has been demonstrated using this approach as compared to the manual delineation through ROC analysis. More advanced LSM techniques [17] developeda model for implicit level sets representations which is based on 4D statistical shape analysis that combined the temporal information gotten from gated SPECT sequences. This eliminatedd the need for challenging point correspondences, at the same time outperforming 3D models with a better characterization of the evolution of the temporal shape.

Multiple hybrid approaches have also been developed in order to address some specific challenges in MPI SPECT analysis. The charged contour model presented in [18] is designed specifically to handlle the concavities in the segmentation label volumes. Later, [19] proposed a novel approach that combined the shape and appearance priors using a constraint with levels set deformable models, hence implementing a soft-to-hard parobabiitic constraint that provided a lot more flexibility as compared to rigid shape constraints alone. This approach proved to be particularly efective for LV segmentation in 4 dimensional gated SPECT, even if there were perfusion defects present. [20] developed hybrid active contour model for Myocardial D-SPECT volumes thar combined local image fitting models with the region-scalable fitting energy functions in order to mitigate the inhomogeneity issues, all the while maintaining the computational effciency. More earlier work by [21] developed a statistical model-based approach with the usage of 3D Active Shape Models (ASM) which combined both the geometric shape and the information depicted by the grey-level appearance from training data for the purpose of achieving robust segmentation of gated SPECT MPI.

Despite all the advances in the research, the traditional approaches stil continue to face great challenges in order to achieve the globally optimal point especially when processing the complete field-of-view volumes with a varying amount of image quality and variability in the anatomy. Such limitations has drived the more recent exploration of the machine and deep learning techniques in the field of nuclear car-

diology. Early machine larning applications in the domain were specifically focused on sub-tasks of the whole segmentation pipeline. [22] showed the effectiveness of using support vector machines (SVMs) for predicting the optimal valve positioning, demonstraing that the approach can be comparable to expert performance in SPECT alignment, at thesame time reducing the dependance on users for quantification. This study highlighted how even the conventional machine learning approaches can improve some specific aspects of the workflow of cardiac analysis. The study done in [23] presents a comprehensive review showing that deep learning solutions have shown remarkable promise across multiple aspects of PET and SPECT imaging, from the quantitative analysis to the instrumentation part of it all. [24] presentes a specific discussion about the transformative impace of using convolutional neural networks (CNNs) on the task of LV segmentaion, showing their ability to learn and interpret complex features directly from images.

The dawn of deep learning in the world brought forward even more comprehensive solutions to the task of LV segmentation. [25] proposes an end-to-end fully CNN based architecture that directly learns the segmentation mapping from the SPECT images taken as input. This approach eliminated the need to process the volumes in multiple stages for the segmentation map. This way the approaches using deep learning in order to handle the entire segmentation task could be developed, replacing the traditional way, with a unified framework. Buiding on the same idea [26] implemented a U-Net [27] based CNN architecture that outperformed significantly their own previous dynamic programming solution presented in [28], and it particularly improves handling the complex variations of shape of the Lv myocardim. More recently, there have been an increasing number of studies in even more sophisticated network architectures that are tailored to some of the unique challenges that are prevelant in the analaysis of cardiac SPECT. [29] introduced convolutional long-short term memory (LSTM) units in the skip connections of a V-Net architecture [30], which enabled an effective way of extracting temporal features from gated SPECT sequences. This novel approach addressed an essential need to leverage the presence of temporal information in higher dimensional volumes of SPECT. In a similar way, [31] enhances the usual 3D U-Net arhitecture with a self-attention mechanism which is amployed at the bottleneck. This allows for better inclusion of the global contexual information throughout the volumetric data.

The usage of shape priors have been identified as one o the efficient strategies,

valuable for improving the accuracy of segmentation models. [32] proposes a method that includes the shape priors, which are generated using a dynamic programming algorithm into a 3D V-Net network using a spatial transformer network (STN) which proved to give extremely good resuts al the whhile maintaining anatomical consistency. Despite the research, there is still a lot of unexplored room when it comes to shape priors and their use in the segmentation process [25]. The existing solutions rely heavily on a number of factors. The first one being the availability of a huge dataset size available with labels in order to train a model. Secondly, the focus within the architecture have been convolutional neural networks and thirdly, even the use of hybrid mechanisms which include attension systems have at least one convolutional layer component. While being effective, these approaches most of the time require substantial training data or very complex model architecture that might limit clinical applicability.

As opposed to the previous works in the field, the method proposed in this study introduces a number of innovative ideas. Firstly, the DL model used in the study is a fully transformer based architecture, which diverts from the traditional approaches of using convolutional blocks. Using this approach allows to capture long-range global dependancies in the input 3D volumes. Moreover, and more importantly, our approach incorporates the statistical shape priors in a way that compliments the strength of the transformer model for better performance. This approach achieves performance that is comparable to the state-of-the-art methods, all the while requiring half the amount of training, with limited data, addressing a probllem that is critical in the clinical deployement where there is not a huge amount of annotated dataset available. The novelty of this research is built upon, with substantial etension of the previously present work. Even though [33] demonstrates the value of using spatial transformers and and [32] discuss te usage of shape priors, this method unifies the approaches within a single transformer framework. Compared to the attention methhod presented in [31], the approach presented in this study is more comprehensive self-attention based paradigm throughout the whole network. The gains in the efficiency, relative to [34] are very noteworthy, as a strong performance is achieved her without the need of a self-supervised pretraining phase.

Chapter 3

Methodology

3.1 Overview

The main goal of the research is to propose a comprehensive and strong method which is desiggned specifically to significantly improve the accuracy of the segmentation of MPI using SPECT for the Left ventricle. The precise segmentation of MPI SPECT images is extremely critical for the detection and the assessment of CAD. However, the task of achieving a high accuracy in segmentation poses a number of challenges due to a multitude of inherent limitations of MPI SPECT data, such as low signal-to-noise ratio (SNR) partial volume affects, substantial noise because of the poisson statistics, motion artifacts and, obviously, the anatomical variability among different patints.

In order to address these challlenges, the proposed research amalgamates advanced approaches in the field of deep learning, more specifically utilizing the transformer based architecture known as nnFormer [35], which is combined with an innovative idea of using statistical shape prios. The nnFormer architecture is chosen because of the ability of it of capturing both the ocal and the global information or contextual relationships in volumetric data as opposed to traditional CNNs which only capture local information. nnFormer leverages the local volume-based (LV-MSA) and the global volume-based (GV-MSA) multi-head self-attention mechanisms very efficiently in a unified method. These modules of the transformer architecture very effectively encode the longg-range dependencies which are extremely important for better segmentation accuracy specifically in medical imaging where they are characterized by indistinct boundaries.

Simultanously, the SSP are merged into the segmentation pipeline in order to improve the anatomical consistency of the model. SSPs provide the DL model with a mathematical model which captures the probabilistic variability off the LV that are derived from the data annotated by experts. This SSP methos employes an advanced technique of optimization such as Mahalanobis distance based regularization and the Kullback-Liebler (KL) divergence in order to refine the segmentation boundaries. This way the outputs of the segmentation model maintain plausibile anatomical outputs, which improves the segmentation accuracy even when the input data is incomplete or ambiguous. The combination of nnFormer and SSP provides us with a novel hybrid architecture.

This hybrid approach leverages not only the strengths of DL models in extracting complex and heirarchical feature representations from volumetric data, but also the advantages of SSPs in mantaining consistent anatomies. This approach also adressess the limitations of the existing methods, which include inefficient generalization capability and dependencies of large, precisely annotated data for training. It also mitigates the impact of a number of different imaging artifacts and the noise, which enhances the overall relaibility on the segmentations.

Extensive procedures for training involving efficient optimization strategies such as Adam, and specifically segmentation tailored loss such as the DiceCE loss, are implemented in order to ensure stable performance across a diverse set of paatients. The training and the validation aspects are conducted rigorously using an extensive dataset of MPI SPECT which consists of diverse collimation methods and demograhics of the patients which improves the generalization capability of the method. Extensive setups of computation which leverage high performance GPU computing environments ensure efficient training and inference of the model. Comprehensive evalaution metrics are utilized in order to quantitatively validate the performance of the segmentation. These metrics include precision, recall, intersection over union (IoU) and Dice coefficient. These metrics provide an extremely in-depth insights into the capability of the method to handle real-world variability and complex scenarios.

In a summary, the proposed methodology contibutes to not only a significant advancement in the division of cardiac image segmentation but also provides a practical and robust solution which is applicable in a clinical environment. The combination of nnFormer and SSP ensures reliable and precise segmentation having clinically mean-

ingful results, which paves the way for better and improved diagnosis and patient outcoms in CAD management.

3.2 Data Acquisition and Preprocessing

The MPI dataset which is utilized in this research was acquired using SPECT. This acquired dataset consists of volumes from a total of 74 patients, which are carefully selected in order to represent the diverse demographic and the characteristics of the clinic. The popultion of the patients inclued individuals with varying age groups, physiological conditions and gender distributions in order to ensure the relaibility, robustness and generalizability of the model. Multiple different radiotracers were employed in the process of acquiring the MPI SPECT, specifically agents labled by technetium-99m(Tc) such as Tc Tetrofosmin and TC Sestamibi, and also the thallium 201 chloride (T1 Chloride).

Each of these radiotraces offer unique properties in imaging thereby providing a comprehensive coverage of all the possible clinical scenarios which are encountered in everyday diagnostice. The dataset was recorded using multiple different collimation methods and also different imaging configurations, including muti-pinhole (MPH), low-energy high-resolution (LEHR), CardioC, CardioD collimators. These various imaging techniues simulate the real-world variability in the clinical practices and pose very distinct challenges for segmentation models because of the variations in spatial resolution, noise characteristics and the sensitivity. Specifically, the dataset consists of 40 patients who are imaged with mixed black-box collimators, 8 patients each images using MPH, CardioC and CardioD, and 10 patients images using LEHR collimators. This comprehensive approach ensures that the data consists of a wide rane of both the quality of the image and also the imaging artifacts which are typically observed in a clincal setting.

Each of the patient went through very rigorous imaging procedures which are adhering strictly to standard protocals of acquisition in clinics. The patients were administered the mentioned radiopharmaceuticals intravenously which was followed by image acquisition after the standardized waiting period that allows sufficient tracer uptake in the myocardial tissue. The image acquisition protocols were varying based on the collimation method which was emploed. For example imaging with the MPH collimator a very specific step-and-shoot helical trajectories, on the other hand the

stationary collimator positions were employed for the other collimators which creates different spatial sampling patterns and different challenges to image reconstruction. AFter the acquisition of the raw data, a number of precprocessing techniques were employed in ordder to prepare the data for the subsequent segmentation analysis. The preprocessing pipeline was developed in order to address a number of inherent issues with the imaging and to optimize the data quality for better segmentation outcomes.

The preprocessing steps began with the correction of the attenuation utilizing the TeraTomo reconstruction algorithm [36], which majorly removed the attenuation artifacts which are caused by the soft bone and tissue structures. This step is very essential in order to ensure the uniformity in the ditribution representation of the tracer across the myocardial tissue, hence improving the segmentation accuracy. In some cases where the attenuation correction data was not available, an Ordered Subset Expectation Maximization (OSEM) algorithm [37] was used in order to reconstruct the full field-of-view (FOV) volumes, providing us with data with robust handling of poisson noise and preserving essential details of the image.

In order to improve the image quality even more, noise reduction techniques are employed, specifically targeting the reduction of the poisson noise which is the most prominent in MPI SPECT imaging due to the low cound of the photons. More advanced filtering methods were also employed such as the Gaussian smoothing and the adaptive median filtering in order to balance the noise reduction with the preservation of important boundaries anatomically and the structural details. The partial volume effects (PVE), which majorly has an impact on the accuracy of the segmentation because of blurring tissue boundaries, were adressed systematically using dedicated PV correction tecnhiques and deconvolution techniques. These methods restored the sharpness in the images and enhanced the delineation of th myocardial boundaries, especially in the regions which have complex anatomical structures.

The images that are the result of the above preprocessing are made to go through further normalization procedures in order to ensure consistency in the scales of intensity across all the datasets which helps in haveing more robust training of the final segmentation models. Standardization of the intensities of the voxels involved scaling the pixel intensity distribution in order to have a mean of zero and a unit variance, which significantly improves the numerical stability, which in-turn helps the convergence of the DL models.

All the data preprocessing steps are performed in a very structured and repeatable framewor, usin scripts that are custom developed in Python and specilized libraries for medical imaging and DL such as PyTorch, TorchIO and Scikit-learn. The comprehensive documentation of the preprocessing parameters and the configurations was nicely maintained so as to ensure the transparancy and the reproducibilit of the methodology. The final dataset after the preprocessing provided us with a very high-quality and standardized input data for the training, validation and the testing of the DL models. The careful handling of the whole data acquisition pipeline ensuring the variability and the rigorous preprocessing ensured optimal MPI SPECT image preparation which maorly enhanced the accuraacy and the relaiability of the segmentation which was obtained from the hybrid model.

3.3 Detailed Description of nnFormer Architecture

The nnFormer architecture introduces an innovative advancements in the field of medical image segmentation, which is specifically designed in order to address the limiations that are faces by the traditional CNNs. nnFormer does it by efficiently capturing both the local and the global spatial relationships in the volumetric medical data. This section gives a very detailed description of the architecture explaining the components, their integration, and the rationale behind the usage of the components in the spicified manner.

3.3.1 Overall Architecture

The nnFormer architecture follows a structure that is very much used in the image segmentation field. It follows a U-shaped encoder decoder architecture which is inspired by the widely used U-Net. This choise of the structure helps in efficient learning of the detailed local features, all the while preserving and leveraging the global contexutual information across multiple scales of resolution. The nnFormer architecture comprises of three main components: an encoder, a bottleneck and a decoder which are all interconnected via skip connections.

3.3.2 Encoder

The encoder of nnFormer starts with an embedding layer, which consists of multiple convolutional layers with very small kernel sizes, typically using 3x3x3. This is followed by Gaussan Error Linear Units (GELU) activation and then ends with a normalization layer. This initial convolutional based embedding transforms the input volume into a higher dimentional featurespace which encodes the low-level spatial detailed, which are necessary for subsequent processing, very efficiently. After the embedding layer, the encoder uses a Local Volume-based Multi-hea Self-attention (LV-MSA) blocks. These LV-MSA blocks are designed so that they can capture the local spatial dependencies within the segmented volumes, which significantly reduces the computaitonal compllexity of the model as compared to the conventional global self-attention mechanisms.

Each of the LV-MSA blocks is further comprised of successive layers of transformer modules in order to use attention mechanisms to effectively model the ver intricate local intractions contexxually. The encoder also combines very strategically placed downsampling convolutional layers which reduces the spatial dimensions of the feature maps while at the same time progressively increasing the depth of the feature map. This downsampling process helps the extraction of the heirarchical features in a more abstract way and on a global scale based representations at lower resolutions, which are essential for capturing the borader variations and anatomical structures.

3.3.3 Bottleneck

In the center of the nnFormer there is a bottleneck. This bottleneck has global volume-based multi-head self-attention (GV-MSA) mechanisms. Contrary to the LV-MSA, the GV-MSA provides with a significantly bigger receptive field which captures the long-range dependencies across the whole global context of the volumetric feature map. This increased area of the receptive field is essential at this stage in order to allow the network to achieve a comprehensive understanding of the global representation and the anatomical structures, improving the overall segmentation accuracy. The bottleneck very effectively combines the compllex spatial dependencies and also the high level features which are exracted by the encoder.

Thi serves as a robust foundation for the decoder for accurate and consistent output during the decoding.

3.3.4 Decoder

The decoder is, in a way, mirrored version of the encoder it also employs LV-MSA blocks but coupled with convolutional upsampling, instead of the downsampling, or so-called transposed convolution. This restores the spatial resolution of the feature maps gradually to the original dimenstions of the input. Each step of the upsampling process in the decoder is designed so as to reconstruct the detailed anatomical information by combining the high resolution detail captured sptially from the corresponding encoding stages via skip connections.

A prime innovation of the nnFormer is the use of the skip attention mechanisms, in place of he traditional concatenation or summation which is typically used in skip connection mechanisms. These skip connections very selectively integrates the features of the encoder with the corresponding features of the decoder, which are guided by the attention weights that highlight relevant spatial features dynamically and hence suppress the irrelevant features. This selective combinaton majorly improves the precision of the segmentation, specifically in the areas where the clear anatomical delineation is very challenging due to the noise and artifacts.

3.3.5 Attention Mechanisms

nnFormerutilizes two different types of attention mechanisms: the Local Volume-based Multi-head Self-attention (LV-MSA) and Global Volume-based Multi-head Self-attention (GV-MSA). LV-MSA very effficiently models the local spatial dependencies by partitioning the feature maps into manageable patches of volumes, which reduces the computational complexity without sacrificing a lot of the performance of the attention mechanism. In contrast to this, GV-MSA models the global interactions spatially across the entire volumetric feature maps, which are essential for capturing the large scale anatomical structural integrity and the contextual relationships. Both of thse attention mechanisms use multi-head configurations which enable parallel computations of attention across multiple representational subspaces. This multi-ha design greatly improves the capability of the network in order to con-

currently capture very diverse spatiall relationships and the interactions at multiple scales hence thereby improving the segmentation accuracy.

3.3.6 Integration and Optimization

The amalgamation of LV-MSA, GV-MSA and the convolutional operations in the nnFormer is very carefully optimized in order to leverage the strength of each of these methods. The convolutional layers provide the efficient encoding of the low level spatual features, while the LV-MSA and the GV-MSA collectively capture both the complex spatial contexts and the long-range dependencies which are crucial for the precise and robust segmentaiton. Optimization of the nnFormer involves a specialized loss function called the Dice cross entropy loss (DiceCELoss). This loss function objectivises a high segmentation accuracy while maintaining a reliable realisic plausibility, which effectively guides the learning process towards more generalizable models acrosss diverse imaging conditions.

In conclusion, the nnFormer architecture represents a very sophisticated segmentation framework which is specifically tailored for medical imaging systems. The innovative design of the architecture and the advaned methods of attention overcomes the traditional limiations of segmentation models by effectively using optimization techniques across components ensuring accurate, robust and clinically meaningfull segmentation outputs.

Chapter 4

Methodology

4.1 Theorem-like environments

Definition 1.

Theorem 1.

Proof. □

Remark.

Note.

Chapter 5

Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. In eu egestas mauris. Quisque nisl elit, varius in erat eu, dictum commodo lorem. Sed commodo libero et sem laoreet consectetur. Fusce ligula arcu, vestibulum et sodales vel, venenatis at velit. Aliquam erat volutpat. Proin condimentum accumsan velit id hendrerit. Cras egestas arcu quis felis placerat, ut sodales velit malesuada. Maecenas et turpis eu turpis placerat euismod. Maecenas a urna viverra, scelerisque nibh ut, malesuada ex.

Aliquam suscipit dignissim tempor. Praesent tortor libero, feugiat et tellus porttitor, malesuada eleifend felis. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nullam eleifend imperdiet lorem, sit amet imperdiet metus pellentesque vitae. Donec nec ligula urna. Aliquam bibendum tempor diam, sed lacinia eros dapibus id. Donec sed vehicula turpis. Aliquam hendrerit sed nulla vitae convallis. Etiam libero quam, pharetra ac est nec, sodales placerat augue. Praesent eu consequat purus.

Acknowledgements

In case your thesis received financial support from a project or the university, it is usually required to indicate the proper attribution in the thesis itself. Special thanks can also be expressed towards teachers, fellow students and colleagues who helped you in the process of creating your thesis.

Appendix A

Simulation results

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque facilisis in nibh auctor molestie. Donec porta tortor mauris. Cras in lacus in purus ultricies blandit. Proin dolor erat, pulvinar posuere orci ac, eleifend ultrices libero. Donec elementum et elit a ullamcorper. Nunc tincidunt, lorem et consectetur tincidunt, ante sapien scelerisque neque, eu bibendum felis augue non est. Maecenas nibh arcu, ultrices et libero id, egestas tempus mauris. Etiam iaculis dui nec augue venenatis, fermentum posuere justo congue. Nullam sit amet porttitor sem, at porttitor augue. Proin bibendum justo at ornare efficitur. Donec tempor turpis ligula, vitae viverra felis finibus eu. Curabitur sed libero ac urna condimentum gravida. Donec tincidunt neque sit amet neque luctus auctor vel eget tortor. Integer dignissim, urna ut lobortis volutpat, justo nunc convallis diam, sit amet vulputate erat eros eu velit. Mauris porttitor dictum ante, commodo facilisis ex suscipit sed.

Sed egestas dapibus nisl, vitae fringilla justo. Donec eget condimentum lectus, molestie mattis nunc. Nulla ac faucibus dui. Nullam a congue erat. Ut accumsan sed sapien quis porttitor. Ut pellentesque, est ac posuere pulvinar, tortor mauris fermentum nulla, sit amet fringilla sapien sapien quis velit. Integer accumsan placerat lorem, eu aliquam urna consectetur eget. In ligula orci, dignissim sed consequat ac, porta at metus. Phasellus ipsum tellus, molestie ut lacus tempus, rutrum convallis elit. Suspendisse arcu orci, luctus vitae ultricies quis, bibendum sed elit. Vivamus at sem maximus leo placerat gravida semper vel mi. Etiam hendrerit sed massa ut lacinia. Morbi varius libero odio, sit amet auctor nunc interdum sit amet.

Aenean non mauris accumsan, rutrum nisi non, porttitor enim. Maecenas vel tortor ex. Proin vulputate tellus luctus egestas fermentum. In nec lobortis risus,

sit amet tincidunt purus. Nam id turpis venenatis, vehicula nisl sed, ultricies nibh. Suspendisse in libero nec nisi tempor vestibulum. Integer eu dui congue enim venenatis lobortis. Donec sed elementum nunc. Nulla facilisi. Maecenas cursus id lorem et finibus. Sed fermentum molestie erat, nec tempor lorem facilisis cursus. In vel nulla id orci fringilla facilisis. Cras non bibendum odio, ac vestibulum ex. Donec turpis urna, tincidunt ut mi eu, finibus facilisis lorem. Praesent posuere nisl nec dui accumsan, sed interdum odio malesuada.

Bibliography

[1] Ibrahim Danad et al. "Comparison of Coronary CT Angiography, SPECT, PET, and Hybrid Imaging for Diagnosis of Ischemic Heart Disease Determined by Fractional Flow Reserve". In: *JAMA Cardiology* 2.10 (Oct. 2017), pp. 1100–1107. ISSN: 2380-6583. DOI: 10.1001/jamacardio.2017.2471. eprint: https://jamanetwork.com/journals/jamacardiology/articlepdf/2648688/jamacardiology_danad_2017_oi_170038.pdf. URL: https://doi.org/10.1001/jamacardio.2017.2471.

[2] Piotr Slomka et al. "Quantitative analysis of perfusion studies: Strengths and pitfalls". In: *Journal of Nuclear Cardiology* 19.2 (2012), pp. 338–346. ISSN: 1071-3581. DOI: https://doi.org/10.1007/s12350-011-9509-2. URL: https://www.sciencedirect.com/science/article/pii/S1071358123030908.

[3] Ernest V. Garcia et al. "The increasing role of quantification in clinical nuclear cardiology: The Emory approach". In: *Journal of Nuclear Cardiology* 14.4 SPEC. ISS. (2007), pp. 420–432. ISSN: 10713581. DOI: 10.1016/j.nuclcard.2007.06.009.

[4] Yi Hwa Liu. "Quantification of nuclear cardiac images: The Yale approach". In: *Journal of Nuclear Cardiology* 14.4 SPEC. ISS. (2007), pp. 483–491. ISSN: 10713581. DOI: 10.1016/j.nuclcard.2007.06.005.

[5] Edward P. Ficaro et al. "Corridor4DM: The Michigan method for quantitative nuclear cardiology". In: *Journal of Nuclear Cardiology* 14.4 SPEC. ISS. (2007), pp. 455–465. ISSN: 10713581. DOI: 10.1016/j.nuclcard.2007.06.006.

[6] Oluwaremilekun Zeth Tolu-Akinnawo et al. "Advancements in Artificial Intelligence in Noninvasive Cardiac Imaging: A Comprehensive Review". In: *Clinical Cardiology* 48.1 (2025), e70087.

[7] Milena J Henzlova and W Lane Duvall. "The future of SPECT MPI: time and dose reduction". In: *Journal of nuclear cardiology* 18.4 (2011), pp. 580–587.

[8] Huidong Xie et al. "Transformer-Based Dual-Domain Network for Few-View Dedicated Cardiac SPECT Image Reconstructions". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 163–172.

[9] Huidong Xie et al. "A Generalizable 3D Diffusion Framework for Low-Dose and Few-View Cardiac SPECT". In: *arXiv preprint arXiv:2412.16573* (2024).

[10] Xiongchao Chen et al. "DuDoCFNet: Dual-Domain Coarse-to-Fine Progressive Network for Simultaneous Denoising, Limited-View Reconstruction, and attenuation correction of Cardiac SPECT". In: *IEEE transactions on medical imaging* (2024).

[11] Hong-Yu Zhou et al. "nnFormer: volumetric medical image segmentation via a 3D transformer". In: *IEEE transactions on image processing* 32 (2023), pp. 4036–4045.

[12] Edward P. Ficaro et al. "Corridor4DM: The Michigan method for quantitative nuclear cardiology". In: *Journal of Nuclear Cardiology* 14.4 (2007). Abstracts of Original Contributions, ASNC 2007, 12th Annual Scientific Session, pp. 455–465. ISSN: 1071-3581. DOI: https://doi.org/10.1016/j.nuclcard.2007.06.006. URL: https://www.sciencedirect.com/science/article/pii/S1071358107002942.

[13] Ernest V. Garcia et al. "The increasing role of quantification in clinical nuclear cardiology: The Emory approach". In: *Journal of Nuclear Cardiology* 14.4 (2007). Abstracts of Original Contributions, ASNC 2007, 12th Annual Scientific Session, pp. 420–432. ISSN: 1071-3581. DOI: https://doi.org/10.1016/j.nuclcard.2007.06.009. URL: https://www.sciencedirect.com/science/article/pii/S1071358107002978.

[14] Guido Germano et al. "Quantitation in gated perfusion SPECT imaging: The Cedars-Sinai approach". In: *Journal of Nuclear Cardiology* 14.4 (2007). Abstracts of Original Contributions, ASNC 2007, 12th Annual Scientific Session, pp. 433–454. ISSN: 1071-3581. DOI: https://doi.org/10.1016/j.

nuclcard.2007.06.008. URL: https://www.sciencedirect.com/science/article/pii/S1071358107002966.

[15] Yi-Hwa Liu. "Quantification of nuclear cardiac images: The Yale approach". In: *Journal of Nuclear Cardiology* 14.4 (2007). Abstracts of Original Contributions, ASNC 2007, 12th Annual Scientific Session, pp. 483–491. ISSN: 1071-3581. DOI: https://doi.org/10.1016/j.nuclcard.2007.06.005. URL: https://www.sciencedirect.com/science/article/pii/S1071358107002930.

[16] Mohammad Hosntalab et al. "Automatic left ventricle segmentation in volumetric SPECT data set by variational level set". In: *Int. J. Comput. Assist. Radiol. Surg.* 7.6 (2012), pp. 837–843. DOI: 10.1007/S11548-012-0770-X. URL: https://doi.org/10.1007/s11548-012-0770-x.

[17] Timo Kohlberger et al. "4D Shape Priors for a Level Set Segmentation of the Left Myocardium in SPECT Sequences". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. Ed. by Rasmus Larsen, Mads Nielsen, and Jon Sporring. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 92–100. ISBN: 978-3-540-44708-5.

[18] Ronghua Yang, Majid Mirmehdi, and David Hall. "A Charged Contour Model for Cardiac SPECT Segmentation". In: 2006. URL: https://api.semanticscholar.org/CorpusID:2756798.

[19] Ronghua Yang et al. "Shape and appearance priors for level set-based left ventricle segmentation". In: *IET Computer Vision* 7.3 (2013), pp. 170–183. DOI: https://doi.org/10.1049/iet-cvi.2012.0081. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-cvi.2012.0081. URL: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cvi.2012.0081.

[20] Chenxi Huang et al. "A Hybrid Active Contour Segmentation Method for Myocardial D-SPECT Images". In: *IEEE Access* 6 (2018), pp. 39334–39343. DOI: 10.1109/ACCESS.2018.2855060.

[21] Xinxin Liu et al. "Materials Integration for Printed Zinc Oxide Thin-Film Transistors: Engineering of a Fully-Printed Semiconductor/Contact Scheme".

In: *Journal of Display Technology* 12.3 (2016), pp. 214–218. DOI: 10.1109/JDT.2015.2445378.

[22] Julian Betancur and et al. "Automatic Valve Plane Localization in Myocardial Perfusion SPECT/CT by Machine Learning: Anatomic and Clinical Validation". In: *Journal of Nuclear Medicine* 58.6 (2017). ISSN: 0161-5505. DOI: 10.2967/jnumed.116.179911.

[23] Hossein Arabi et al. "The promise of artificial intelligence and deep learning in PET and SPECT imaging". In: *Physica Medica* 83 (2021), pp. 122–137. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2021.03.008. URL: https://www.sciencedirect.com/science/article/pii/S1120179721001241.

[24] Jelmer M. Wolterink. "Left ventricle segmentation in the era of deep learning". In: *Journal of Nuclear Cardiology* 27.3 (2020), pp. 988–991. ISSN: 1071-3581. DOI: https://doi.org/10.1007/s12350-019-01674-3. URL: https://www.sciencedirect.com/science/article/pii/S1071358123018809.

[25] Tonghe Wang and et al. "A learning-based automatic segmentation and quantification method on left ventricle in gated myocardial perfusion SPECT imaging: A feasibility study". In: *Journal of Nuclear Cardiology* 27 (3 June 2020), pp. 976–987. ISSN: 15326551. DOI: 10.1007/s12350-019-01594-2.

[26] Haixing Wen et al. "Analysis on SPECT myocardial perfusion imaging with a tool derived from dynamic programming to deep learning". In: *Optik* 240 (2021), p. 166842. ISSN: 0030-4026. DOI: https://doi.org/10.1016/j.ijleo.2021.166842. URL: https://www.sciencedirect.com/science/article/pii/S0030402621005477.

[27] O. Ronneberger, P.Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, pp. 234–241. URL: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a.

[28] Shaojie Tang et al. "Dynamic programming-based automatic myocardial quantification from the gated SPECT myocardial perfusion imaging". In: *The*

International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Xi'an, China. 2017, pp. 462–467.

[29] Chen Zhao et al. "Spatial-temporal V-Net for automatic segmentation and quantification of right ventricle on gated myocardial perfusion SPECT images". In: *Medical Physics* 50.12 (2023), pp. 7415–7426.

[30] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *2016 Fourth International Conference on 3D Vision (3DV)*. 2016, pp. 565–571. DOI: 10.1109/3DV.2016.79.

[31] Yangmei Zhang et al. "An automatic segmentation method with self-attention mechanism on left ventricle in gated PET/CT myocardial perfusion imaging". In: *Computer Methods and Programs in Biomedicine* 229 (2023), p. 107267. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2022.107267. URL: https://www.sciencedirect.com/science/article/pii/S0169260722006484.

[32] Fubao Zhu et al. "A new method incorporating deep learning with shape priors for left ventricular segmentation in myocardial perfusion SPECT images". In: *Computers in Biology and Medicine* 160 (2023), p. 106954. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2023.106954. URL: https://www.sciencedirect.com/science/article/pii/S0010482523004195.

[33] Yangfan Ni et al. "A Multiscale Spatial Transformer U-Net for Simultaneously Automatic Reorientation and Segmentation of 3-D Nuclear Cardiac Images". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 8.6 (2024), pp. 632–645. DOI: 10.1109/TRPMS.2024.3382318.

[34] Ádám István Szundefinedcs et al. "Self-supervised segmentation of myocardial perfusion imaging SPECT left ventricles". In: *Proceedings of the 2023 10th International Conference on Bioinformatics Research and Applications*. ICBRA '23. Barcelona, Spain: Association for Computing Machinery, 2024, pp. 206–211. ISBN: 9798400708152. DOI: 10.1145/3632047.3632078. URL: https://doi.org/10.1145/3632047.3632078.

[35] Hong-Yu Zhou et al. "nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer". In: *Trans. Img. Proc.* 32 (Jan. 2023), pp. 4036–4045. ISSN:

1057-7149. DOI: 10.1109/TIP.2023.3293771. URL: https://doi.org/10.1109/TIP.2023.3293771.

[36] K. Nagy et al. "Performance Evaluation of the Small-Animal nanoScan PET/MRI System". In: *Journal of Nuclear Medicine* 54.10 (2013), pp. 1825–1832. DOI: 10.2967/jnumed.112.114785.

[37] H. M. Hudson and R. S. Larkin. "Accelerated image reconstruction using ordered subsets of projection data". In: *IEEE Transactions on Medical Imaging* 13.4 (1994), pp. 601–609. DOI: 10.1109/42.363108.

List of Figures

List of Tables

List of Algorithms

List of Codes