Executive Summary of the Thesis

Automated Left Ventricle Segmentation in Cardiac CT Images using Deep Learning Techniques

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Juan F. Calderon

Advisor: Prof. Valentina Corino

Co-advisor: Francesca Lo Iacono

Academic year: 2022-2023

1. Introduction

Image segmentation is a fundamental yet challenging task in the field of computer vision and image processing. It aims to partition a digital image into multiple segments, identifying and separating distinct regions or objects within the image. This process is essential for numerous applications including medical imaging.

In the field of image processing, various techniques have been developed over the years, which involves partitioning an image into meaningful objects. These techniques include traditional methods such as thresholding, edge detection, and clustering, among others.

However, with the advent of deep learning(DL), particularly convolutional neural networks (CNNs), a new era has emerged in the field of image segmentation. DL models have been shown to outperform traditional techniques in terms of accuracy and robustness, particularly for complex and large-scale datasets.

One crucial aspect of cardiac computed tomography (CT) imaging is the segmentation of the left ventricle (LV), as it yields valuable insights into cardiac function and morphology. However, manual segmentation of the LV is a labor-intensive and time-consuming process, prone to inter- and intra-observer variability. Therefore, the development of automated segmentation techniques, is necessary to enhance efficiency and accuracy in this domain.

As such, this thesis aims to explore the effectiveness of DL models for image segmentation tasks, comparing their performance with traditional methods. This work investigates the potential of DL models for the segmentation of the LV in cardiac CT images. The primary objective is to compare the performance of three DL models—U-Net, U-Net 2.5D, and U-Net with LSTM—for the LV segmentation task. Additionally, the study examines the strengths and weaknesses of each model in terms of accuracy, efficiency, and generalization. By doing so, this research aims to offer valuable insights into the suitability of these models for clinical applications and contribute to the advancement of more precise and efficient methods for LV segmentation in cardiac CT images.

2. Materials and methods

2.1. Dataset

The dataset contains patients having 3 different pathologies : amyloidosis, hypertrophy, and

stenosis. The whole datasets consisted of cardiac CT scans collected from 85 patients. Each patient scan contained between 30 and 115 of slices provided with ground truth LV segmentation masks manually delineated by expert radiologists. The dataset was randomly shuffled (leaving each patient in one set only) , and training, validation, and test sets were distributed as 75%, 15%, and 10%, respectively.

2.2. CT Scan Protocol

Cardiac CT images used in this study were acquired using two different scanners: a 256-slice scanner and a 320-slice wide volume coverage scanner, without any premedication. The Revolution CT scans were obtained with a peak tube voltage of 100 kV, detector collimation of 160 mm using 256 rows by 0.625 mm on the Z-axis, and a fixed dose of 50 ml bolus of contrast medium via an antecubital vein. The Aquilion ONE Vision scans were acquired with a peak tube voltage of 120 kV, tube current-time product of 160 mAs, and a section collimation of 320 detector rows and 1.2-mm section thickness.

2.3. Image Preprocessing and Augmentation

Before training the models, the CT scans were preprocessed to ensure consistent input dimensions and intensity values. The HU transformation involved converting the raw pixel values in the CT scans to HU values, which provided a standardized representation of tissue densities. Windowing was applied to focus on a specific range of HU values relevant to the left ventricle, effectively enhancing the contrast between the region of interest and the surrounding tissues. Data augmentation techniques, such as rotation, scaling, and flipping, were applied to the training set to increase its diversity and enhance the models' generalization capabilities.

2.4. Segmentation Models

Three DL-based segmentation models were implemented and evaluated in this study:
Baseline-UNet: a variant of the original UNet architecture with modifications that include dropout, L2 regularization, and batch normalization to improve training stability and prevent overfitting.
UNet 2.5D: a variant of the UNet architec-ture that takes advantage of the temporal dependency between consecutive CT slices. This is accomplished by using three or more consecutive CT slices as input channels and predicting the corresponding segmentation mask using the middle slice. This is called the 2.5D approach and has been shown to improve segmentation performance compared to using a single slice as input. The input to the model is a N-channel input, where each channel represents a consecutive CT slice, and the model outputs a single-channel segmentation mask corresponding to the middle slice of the input.
UNet-LSTM: UNet architecture integrated with LSTM layers for modeling temporal dependencies in the input data, which is particularly useful for sequential input data, such as a series of images or volumes.

2.5. Model Evaluation

The performance of the segmentation models was assessed using the Intersection over Union (IoU) and Dice Coefficient metrics, which measure the models' ability to accurately segment the left ventricle.

$$IoU = \frac{TP}{TP + FP + FN} \tag{1}$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{2}$$

The models were trained using different loss functions, such as Binary Cross Entropy and Soft Dice Loss, to investigate their impact on the segmentation performance.

3. Results

The 3 models, UNet, UNet 2.5D, and LSTM, were evaluated on the validation set using the Dice coefficient and the IoU.

3.1. Validation results

3.1.1 UNet

Table 1 shows the results of the U-Net model on a per-patient basis. The table includes the number of images for each patient, as well as the mean and standard deviation of the Dice coefficient and the IoU. The results show that the U-Net model achieved varying levels of performance across the different patients. Patients HCM29 and AS043 had the highest mean Dice

coefficient and IoU, while patient HCM05 had the lowest mean Dice coefficient and IoU. Overall, the U-Net model achieved a mean Dice coefficient of 0.864 and a mean IoU of 0.763 across all patients. The standard deviations of these metrics indicate that the model's performance did not vary considerably across different images within each patient, with some images achieving high scores while others achieved low scores.

Patient	N images	Dice	IoU
HCM29	56	0.895±0.053	0.814±0.080
HCM27	68	0.859±0.038	0.755±0.058
HCM05	73	0.767±0.063	0.627±0.084
HCM10	70	0.871±0.032	0.772±0.050
HCM30	47	0.883±0.031	0.792±0.049
AM08	90	0.868±0.037	0.769±0.059
AM01	39	0.852±0.034	0.743±0.051
AM37	58	0.852±0.036	0.744±0.055
AS030	60	0.896±0.030	0.812±0.050
AS031	58	0.857±0.026	0.751±0.040
AS043	67	0.899±0.026	0.818±0.041
Mean	62.364	0.864±0.037	0.763±0.056

Table 1: Metric results for UNet on the validation set.

3.1.2 UNet 2.5D

Table 2 summarizes the performance of the UNet 2.5D model in the validation set. Overall, the UNet 2.5D model achieved a mean Dice coefficient of 0.874 and a mean IoU coefficient of 0.780, with standard deviations of 0.031 and 0.048, respectively. The model performed well on most patients, with mean Dice and IoU coefficients ranging from 0.849 to 0.905 and from 0.739 to 0.828, respectively. However, some patients, such as HCM05, had lower performance, with mean Dice and IoU coefficients of 0.791 and 0.658, respectively. These results suggest that the UNet 2.5D model may be a suitable option for segmentation.

3.1.3 UNet LSTM

Table 3 reports the evaluation metrics of the third model, UNet LSTM, on the validation set. The average Dice coefficient over all patients is 0.864, with a standard deviation of 0.040. The IoU metric has a mean value of 0.765, with a standard deviation of 0.060. Looking at individual patients, the model achieved a Dice coeffi-

Patient	N images	Dice	IoU
HCM29	56	0.914±0.032	0.843±0.052
HCM27	68	0.870±0.034	0.771±0.054
HCM05	73	0.791±0.053	0.658±0.075
HCM10	70	0.880±0.030	0.788±0.048
HCM30	47	0.886±0.029	0.796±0.047
AM08	90	0.879±0.027	0.786±0.043
AM01	39	0.861±0.023	0.757±0.036
AM37	58	0.883±0.027	0.791±0.043
AS030	60	0.899±0.029	0.818±0.049
AS031	58	0.850±0.029	0.740±0.043
AS043	67	0.905±0.027	0.828±0.043
Mean	62.364	0.874±0.031	0.780±0.048

Table 2: Metric results for UNet 2.5D on the validation set.

cient ranging from 0.766 to 0.901, and an IoU ranging from 0.624 to 0.820. The highest performance was obtained on patient AS043, with a Dice coefficient of 0.901 and an IoU of 0.820. Compared to the previous models, UNet LSTM generally performed slightly worse in terms of the mean Dice coefficient, but still achieved satisfactory results. The standard deviation of the Dice coefficient was also slightly higher than in the other models, indicating a larger variability in performance across patients. However, the IoU metric showed a similar level of performance compared to the other models, with a mean value around 0.76-0.82.

Patient	N images	Dice	IoU
HCM29	56	0.894±0.064	0.814±0.094
HCM27	68	0.863±0.037	0.761±0.056
HCM05	73	0.765±0.059	0.624±0.079
HCM10	70	0.876±0.031	0.780±0.048
HCM30	47	0.883±0.032	0.792±0.051
AM08	90	0.864±0.038	0.762±0.059
AM01	39	0.846±0.036	0.734±0.054
AM37	58	0.846±0.066	0.739±0.089
AS030	60	0.899±0.027	0.818±0.045
AS031	58	0.870±0.026	0.771±0.040
AS043	67	0.900±0.025	0.820±0.041
Mean	62.364	0.864±0.040	0.765±0.060

Table 3: Metric results for UNet LSTM on the validation set.

3.1.4 Comparison

As seen in Figure 1, the global performances of the three models in the validation set can be compared using the Dice coefficient. The
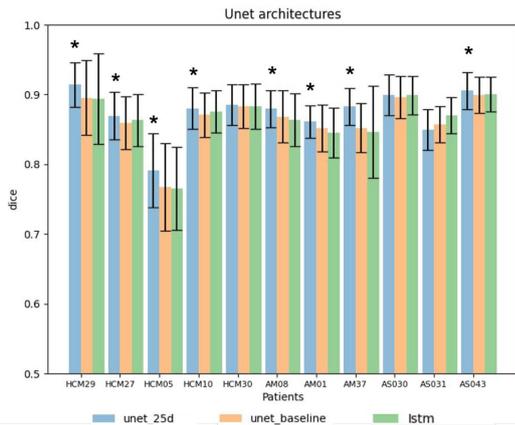
Figure 1: Global performances of the 3 models in the validation set. The asterisk (*) represents the statistical significant model tested by Wilcoxon test ($p < 0.05$), in which the UNet 2.5D was tested against the other 2 models.

UNet 2.5D model achieved the highest average Dice score of 0.874, followed by the UNet LSTM model with an average Dice score of 0.864, and the original UNet model with an average Dice score of 0.864. These results are consistent with the patient-wise performance metrics shown in Tables 1, 2, and 3, where the UNet 2.5D model achieved the highest mean Dice scores for all patients except for AS031.

3.2. Test results

We use the best performing model from the validation set, which is the UNet 2.5D, to evaluate the segmentation performance on the test set.

The segmentation results for the studied patients are summarized in Table 4. We can see that the models achieved high performance, with mean Dice coefficients ranging from 0.877 to 0.912 and mean IoU coefficients ranging from 0.782 to 0.839.

Overall, the results demonstrate the effectiveness of the segmentation models in accurately segmenting the studied patients' images. The high mean Dice and IoU coefficients indicate that the models were able to accurately segment the left ventricle, which is critical for diagnosing and treating heart disease. The additional visualizations of the masks and contour provide a better understanding of the models' performance and can help guide future improvements in segmentation algorithms

Patient	N images	Dice	IoU
HCM13	75	0.869 ± 0.111	0.776 ± 0.149
HCM01	64	0.907 ± 0.021	0.831 ± 0.035
AM29	115	0.900 ± 0.031	0.819 ± 0.050
AS054	54	0.890 ± 0.020	0.802 ± 0.032
AS002	71	0.912 ± 0.012	0.839 ± 0.021
AS021	40	0.885 ± 0.017	0.794 ± 0.027
AS024	75	0.877 ± 0.031	0.782 ± 0.048
Mean	70.571	0.891 ± 0.0345	0.806 ± 0.051

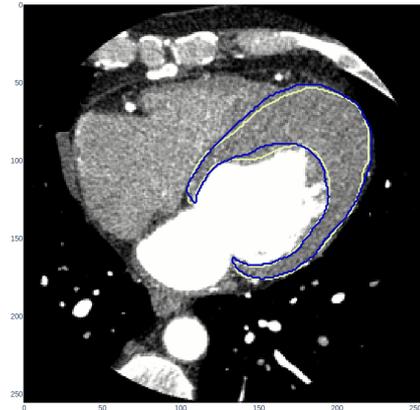Table 4: Metric results for UNet on the test set.

Figure 2: Visualization of the mask prediction and ground truth of a slice of the patient HCM13.

3.3. Visual inspection

In addition to the quantitative metrics obtained from the model, a manual inspection of the images and the corresponding masks was performed to gain a better understanding of the model's performance on the test set. This qualitative evaluation offers a complementary insight into the models' performance and can identify areas for improvement.

We created a comparison between the original CT scan, the ground truth, and the predicted contours overlaid on top for patient HCM13, as shown in Figure 2. The ground truth contours are displayed in blue, while the predicted contours are in yellow. This image provides an even clearer comparison between the ground truth and predicted masks and highlights the high accuracy achieved by the segmentation model for this patient.

4. Discussion

Image segmentation plays a crucial role in the field of medical imaging, particularly in the analysis of cardiac structures like the left ventricle (LV). Accurate segmentation of the LV is essential for diagnosing and monitoring various cardiovascular diseases. Over the years, numerous approaches have been proposed for the segmentation of the LV in CT scans. Machine learning-based methods have gained popularity in recent years due to their ability to learn complex patterns in the data. In particular some models have direclty been used for LV segmentation. Milletari et al. [1] proposed the V-Net, a 3D CNN architecture designed for volumetric medical image segmentation. The authors reported a Dice coefficient of 0.91 on the Prostate dataset. The V-Net has been adapted for LV segmentation tasks in CT scans, showing promising results. Chen et al. [2] introduced DeepLabv3+, an encoder-decoder architecture that combines atrous spatial pyramid pooling (ASPP) and a fully connected conditional random field (CRF). They achieved a mean Dice coefficient of 0.80 on the PASCAL VOC 2012 dataset. Although this study focused on natural images, DeepLabv3+ has been adapted for LV segmentation in CT scans, yielding competitive results.

It was observed that the UNet 2.5D model exhibited the best performance, with the highest values of Dice coefficient and Intersection over Union (IOU). The models achieved high performance, with mean Dice coefficients ranging from 0.877 to 0.912 and mean IoU coefficients ranging from 0.782 to 0.839 with an average Dice score of 0.891. Strengths and limitations of each model were evaluated based on manual examination of the images and masks. The U-Net model combined with LSTM produced masks with thicker boundaries, whereas the 2.5D U-Net model generated masks with precise boundaries and accurately captured the anatomical characteristics of the left ventricle. To the best of our knowledge the UNet 2.5D has never been tested on the LV segmentation task. This is the first attempt to use the architecture for LV segmentation.

5. Conclusions and Future Developments

This study has demonstrated the potential of DL models for accurate segmentation of the left ventricle in cardiac CT images. The results provide important insights into the strengths and weaknesses of different models for medical image segmentation tasks, which can help researchers and practitioners choose the most suitable model for their specific application. Furthermore, accurate segmentation of the left ventricle is essential for diagnosing and monitoring cardiovascular diseases, making the potential of these models particularly promising for clinical applications.

UNet 2.5D provides a better understanding of the spatial relationships between slices, leading to more accurate segmentation.

Compared to 3D approaches and LSTM, UNet 2.5D has reduced computational requirements. The use of simple convolutions allows for faster processing and less memory consumption, making it more efficient and suitable for real-time applications [3].

UNet 2.5D has shown improved robustness against variations in imaging protocols, slice thickness, and image resolutions [4]. This is particularly important for left ventricle segmentation, as images can be acquired from different medical imaging modalities and conditions.

The UNet 2.5D architecture allows for easy implementation of transfer learning, enabling the network to leverage pre-trained models from related tasks for improved performance [5]. This can help in reducing training time and improving generalization across different datasets.

5.1. Limitations

The model have demonstrated promising results but some limits have to be considered when interpreting its performance and generalizability.

The model was not tested on publicly available datasets, which makes it difficult to directly compare its performance with other models in the literature that use the same data.

Also as with most studies on CT scan segmentation, the images typically come from the same scanners. In real-world situations, we may encounter cases where we need to test the model with new scanners, which could potentially decrease its performance. This indicates that our model may not be as robust to generalization as

desired.

Finally the manual annotations on the datasets can vary in terms of the criteria used for segmenting the LV, as different expert radiologists were involved in the annotation process. This inconsistency may affect the quality of the ground truth labels and, in turn, the performance of the segmentation models.

5.2. Future works

Future works might focus on improving the model in terms of performances and generalization. Ensemble models, for example, might enhance U-Net models accuracy as they combine the strengths of multiple U-Nets, which reduces the impact of individual model weaknesses and leads to better overall performance.In addition, ensemble models provide diversity by combining multiple base models. This could lead to better generalization, as each model can learn different aspects of the data [6]. Furthemore exploring additional post processing techniques such as Markov Random Fields(MRF) and active contours snake might increase model segmentation precision. MRF can be used to model the spatial dependencies between neighboring pixels, while active contours snake can be used to iteratively adjust the contour shape to minimize an energy function [7].Additionally, it could be important to explore the transferability of the models to different imaging modalities and patient populations. This might expand the applicability of these models to a wider range of clinical settings. Finally, it might be useful to assess the clinical feasibility and impact of using DL models for medical image segmentation tasks. This could involve conducting clinical trials to evaluate the impact of using these models on patient outcomes such as diagnosis, treatment planning, and prognosis. Ensuring the safety and efficacy of these models in clinical settings will be essential to their successful implementation.

5.3. Conclusion

In conclusion, this study has demonstrated the potential of DL models for accurate medical image segmentation and provided important insights into future directions for research in this field. Further research in this area could lead to more accurate and efficient medical image segmentation, ultimately improving patient out-

come

References

[1] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[2] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[3] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:2001.06268*, 2020.

[4] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[5] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[6] Z.-H. Zhou, "Ensemble methods: Foundations and algorithms," 2012.

[7] B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," 2021.