Faster Segmentation of Cardiac MRI using U-Net

by Ankita Samanta

Submission date: 28-Feb-2022 01:07PM (UTC+0530)
Submission ID: 1772759152
File name: Final_Second_review_report.pdf (462.36K)
Word count: 2421
Character count: 13487

Faster Segmentation of Cardiac MRI using U-Net

Ankita Samanta
RA1811003010172
Department of Computer Science and Engineering,
School of Computing
SRM Institute of Science and Technology
Kattankulathur, Chennai
Email:as6480@srmist.edu.in

P. Saksham
RA1811003010142
Department of Computer Science and Engineering,
School of Computing
SRM Institute of Science and Technology
Kattankulathur, Chennai
Email:ps9469@srmist.edu.in

Guide Name: Dr. R. I. Minu
Department of Computer Science and Engineering,
School of Computing
SRM Institute of Science and Technology
Kattankulathur, Chennai
Email:minur@srmist.edu.in

***Abstract*—**The project deals with multiclass medical image segmentation to make a pixel level classification of the left ventricle endocardium (LV), left ventricle myocardium (Myo), right ventricle endocardium (RV), and background from MRI scans. Such techniques are useful for automating the process of diagnosing and differentiating between normal cases, heart failure with infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, abnormal right ventricle by cardiologists. We use U-Net architecture which is essentially used for medical image segmentation and hope to devise a technique for automation of MRI diagnosis which requires less data for training and still produces sufficient accuracy(IOU or DICE are the metrics used in case of segmentation).

***Keywords*—** LV- left ventricle endocardium, Myo-left ventricle myocardium, RV- right ventricle endocardium, AI- Artificial Intelligence,

I. INTRODUCTION

Recent years have seen an unprecedented improvement in medical imaging technology. With rapid development in the field of medical imaging, the quality and quantity of these images are significantly rising. Medical imagery finds its use in various cases. Various kinds of medical images like MRIs, CAT scans, X-Rays are used in order to identify fractured bones, blood clots, tumors, etc. Medical imaging, also known as radiology, is the field of medicine in which medical professionals recreate various images of parts of the body for diagnostic or treatment purposes. Medical imaging procedures include non-invasive tests that allow doctors to diagnose injuries and diseases without being intrusive. In a number of medical imaging technologies artificial intelligence (AI) is enhancing the ability to interpret and analyze results. Computer vision is being used to visually diagnose conditions not yet visible to the human eye. A radiographer -- also known as a medical imaging technologist or radiology technologist -- is responsible for administering medical imaging procedures. Radiographers are university-trained with thorough knowledge of the body's structure and how it is affected by different diseases and injuries.

II. STATE OF THE ART (LITERATURE SURVEY)

A. *General*

This chapter offers a complete account of the various methodologies and techniques we researched and went through to better understand the approach required for our project. The reference papers we studied helped us to explore efficient solutions to some of the issues that were present in detection of cell nuclei and segmentation and helped us to improve our algorithm to maximise the speed , efficiency and quality to perform semantic and instance detection in our system.

A. *Literature Survey Inference*

Satellite image analysis plays an important role. Deep Learning in the field of computer vision has developed a lot since the first CNN and now CNNs are very powerful thanks to new and improved techniques. This poses a problem as now CNNs tend to over fit data and thus more focus is required on building better datasets too. U-Net is a robust model in terms of segmentation techniques. Gives accurate results due to multiple modules being end-to-end trainable. Similar variants of segmentation models have been used extensively now even in the detection of novel diseases. A combination of a good model and dataset can help detect tumor cells before they mature and even help in expediting critical cures.

III.PROPOSED WORK

We hope that feeding input image to the model along with mask locations would help in faster convergence and if lucky, might even improve the benchmark accuracy.Instead of using a 3 channel input, we can use a 4 channel input with the last channel indicating all the points of interest.

Our model consists of two modules, first is a regressive model which produces coordinates of the points of interest using landmark detection.

Second module consists of the traditional U-Net model which would be used to segment buildings from the images given to it along with an extra landmark channel.

A. *Network Architecture*

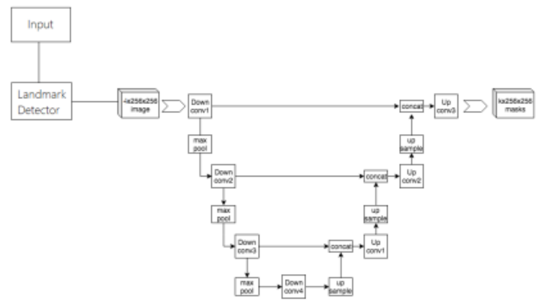

Fig 1. Network Architecture

B. *U-Net Module For Segmentation*

U-Net module consists essentially of two parts -

1. Successive convolutions
2. Successive deconvolutions

Successive convolutions are applied to obtain a feature with 30x30x1024 which is further upscaled to the original size.

The U-Net module consists essentially of two parts - Successive convolutions Successive deconvolutions Successive convolutions are applied to obtain a feature with 30x30x1024 which is further up scaled to the original size.

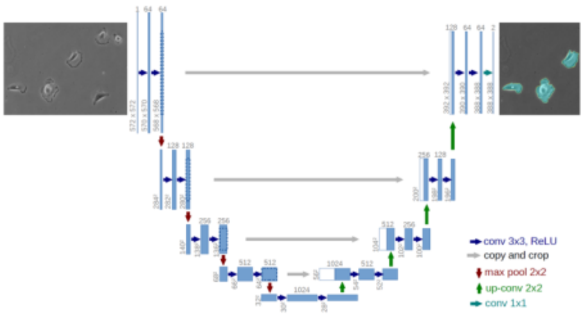

Fig 2. Traditional U-Net Model

C. *Elaborated U-Net Architecture*

Encoder (left side): It consists of the repeated application of **two 3x3 convolutions**. Each conv is followed by a ReLU and batch normalization. Then a 2x2 max pooling operation is applied to reduce the spatial dimensions. Again, at each downsampling step, we double the number of feature channels, while we cut in half the spatial dimensions.

Decoder path (right side): Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 transpose convolution, which halves the number of feature channels. We also have a concatenation with the corresponding feature map from the contracting path, and usually a 3x3 convolution (each followed by a ReLU). At the final layer, a 1x1 convolution is used to map the channels to the desired number of classes.

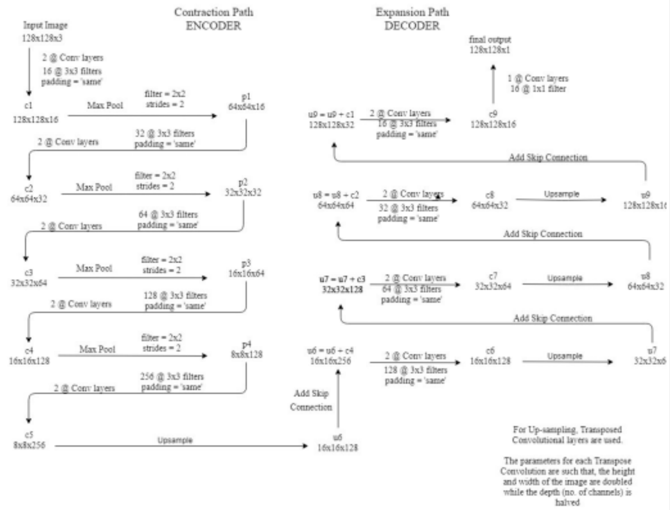

Fig 3. Elaborate U-Net Architecture

D. *Description of the U-net Architecture*

- c1, c2, …. c9 are the output tensors of Convolutional Layers
- p1, p2, p3 and p4 are the output tensors of Max Pooling Layers
- u6, u7, u8 and u9 are the output tensors of up-sampling (transposed convolutional) layers
- The left hand side is the contraction path (Encoder) where we apply regular convolutions and max pooling layers.
- In the Encoder, the size of the image gradually reduces while the depth gradually increases. Starting from 128x128x3 to 8x8x256
- This basically means the network learns the "WHAT" information in the image, however it has lost the "WHERE" information
- The right hand side is the expansion path (Decoder) where we apply transposed convolutions along with regular convolutions
- In the decoder, the size of the image gradually increases and the depth gradually decreases. Starting from 8x8x256 to 128x128x1

- Intuitively, the Decoder recovers the "WHERE" information (precise localization) by gradually applying up-sampling
- To get better precise locations, at every step of the decoder we use skip connections by concatenating the output of the transposed convolution layers with the feature maps from the Encoder at the same level:
  u6 = u6 + c4
  u7 = u7 + c3
  u8 = u8 + c2
  u9 = u9 + c1
  After every concatenation we again apply two consecutive regular convolutions so that the model can learn to assemble a more precise output
- This is what gives the architecture a symmetric U-shape, hence the name UNET
- On a high level, we have the following relationship:
  Input (128x128x1) => Encoder =>(8x8x256) => Decoder => Ouput (128x128x1)

E. *Hyperparameter Tuning*

Consider a set of the most influential hyperparameters that need to be evaluated to the right values so that the model can optimize itself to the best results while training.

- Learning Rate- Controls the depth of gradient descent related model weight updates.
- Epochs- Number of iterations to train the model for. Too much leads to overfitting and too little under fits the model .
- Minimum Support Threshold- Determines the pruning of false positive results while training .
- Data Augmentation- Randomly distorts and transforms the training data so that overfitting is prevented .
- Optimizer- Choice between Stochastic Gradient Descent and Adaptive Optimizers like ADAM.

F. Inference Module

Visualizing the results of training by testing on separate set and generating metrics to compare efficiency

- <u>IOU Score</u>- Calculates an intersection of the predicted mask over the ground truth mask to check the ration of intersection to that of union between them.
- <u>MaP Score</u>- Generates a class vice precision score for each successful model recall on that class.
- <u>Result Visualization</u>- Use alpha-blending to draw a mask over the detected region which can be further used to segment buildings from aerial images.

IV. SYSTEM REQUIREMENTS

A. *Hardware Requirements*

GPUs for model graph computations - NVIDIA GTX 1050ti

B. *Software Requirements*

Frameworks: Pytorch, Numpy, OpenCV, Matplotlib
IDE: Pycharm
Language: Python

C. *Frameworks*

- Pytorch

PyTorch is an open source machine learning library based on the Torch library, used for applications such as computer vision and natural language processing, primarily developed by Facebook's AI Research lab. It is free and open-source software released under the Modified BSD license

- Numpy

NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python. In Python we have lists that serve the purpose of arrays, but they are slow to process. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists. The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy. Arrays are very frequently used in data science, where speed and resources are very important.

- OpenCV

OpenCV is an open-source image processing library which is written in convenient languages such as C++ and Python. It makes image manipulation extremely easy with it's built in mathematical image morphing algorithms and is thus used by developers as a image pre-processor in many computer vision tasks.

- Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

D. Languages

- Python

Python is a high level programming language that was created in 1991. It is a very easy language to use and it's almost similar to English. It is used for many purposes like web development, system scripting and machine learning. It can also be used for servers and different software. It can be object oriented or functional. Python is slower compared to C++.

- Pycharm

PyCharm is an integrated development environment used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.

V. IMPLEMENTATION

The UNet module was constructed using pytorch. The UNet essentially consists of 2 modules - encoder and decoder. **Encoder** (left side): It consists of the repeated application of **two 3x3 convolutions**. Each conv is followed by a ReLU and batch normalization. Then a 2x2 max pooling operation is applied to reduce the spatial dimensions. Again, at each downsampling step, we double the number of feature channels, while we cut in half the spatial dimensions. **Decoder path** (right side): Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 transpose convolution, which halves the number of feature channels. We also have a concatenation with the corresponding feature map from the contracting path, and usually a 3x3 convolutional (each followed by a ReLU). At the final layer, a 1x1 convolution is used to map the channels to the desired number of classes.

To tackle the overfitting situation, different data augmentation techniques like rotation, cropping, color jitter and skewing were used. In addition to that dropout regularization was also used to take care of possible overfit by dropping some connections in the neural network used to train the landmark detector model's dense part.

	1 scan	4 scans	8 scans	16 scans	32 scans	All scans
UNet	0.624	0.782	0.883	0.886	0.900	0.913
UNet++	0.625	0.785	0.876	0.889	0.911	0.915

Fig 4. Results

CONCLUSION

So far we observe that U-Net++ performs better on the ACDC dataset.

We hope to achieve a better DICE score using our method. Our next step includes modification of the existing architecture combined with a unique set of dataset augmentations and hyperparameter tuning that will provide us with the most optimal results for this dataset. We tested a new idea of modifying the input image before it is fed to the model during our minor project. Therefore we will also test this new idea of an added channel with mask landmarks with this dataset.

REFERENCES

1. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
2. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
3. Poster, D., Hu, S., Nasrabadi, N., & Riggan, B. (2019). An examination of deep-learning based landmark detection methods on thermal face imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 0-0).
4. Soni, A., Koner, R., & Villuri, V. G. K. (2020). M-unet: Modified u-net segmentation framework with satellite imagery. In *Proceedings of the Global AI Congress 2019* (pp. 47-59). Springer, Singapore.
5. Piao, S., & Liu, J. (2019, November). Accuracy improvement of UNet based on dilated convolution. In *Journal of Physics: Conference Series* (Vol. 1345, No. 5, p. 052066). IOP Publishing.
6. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
7. Zhang, L., Shen, J., & Zhu, B. (2021). A research on an improved Unet-based concrete crack detection algorithm. *Structural Health Monitoring*, *20*(4), 1864-1879.
8. Cao, K., & Zhang, X. (2020). An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sensing*, *12*(7), 1128.
9. Wang, F., & Xie, J. (2020, August). A context and semantic enhanced UNet for semantic segmentation of high-resolution aerial imagery. In *Journal of Physics: Conference Series* (Vol. 1607, No. 1, p. 012083). IOP Publishing.
10. Kazerouni, I. A., Dooly, G., & Toal, D. (2021). Ghost-UNet: An asymmetric encoder-decoder architecture for semantic segmentation from scratch. *IEEE Access*, *9*, 97457-97465.
11. Soulami, K. B., Kaabouch, N., Saidi, M. N., & Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomedical Signal Processing and Control*, *66*, 102481.
12. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of*

the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

13. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*, 91-99.

14. Ma, B., Zhang, J., Cao, F., & He, Y. (2020). MACD R-CNN: An Abnormal Cell Nucleus Detection Method. *IEEE Access*, *8*, 166658-166669.

Faster Segmentation of Cardiac MRI using U-Net

ORIGINALITY REPORT

51%	47%	26%	37%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	theaisummer.com Internet Source	13%
2	towardsdatascience.com Internet Source	12%
3	Submitted to The NorthCap University, Gurugram Student Paper	6%
4	whatis.techtarget.com Internet Source	6%
5	Submitted to SRM University Student Paper	4%
6	A.Helen Victoria, G. Maragatham. "Human Activity Classification for Radar Signatures using Widened Convolutional Neural Networks", 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021 Publication	3%
7	enoumen.com Internet Source	2%

8	Submitted to muetjamshoro Student Paper	2%
9	cloudstor.aarnet.edu.au Internet Source	1%
10	www.journaltocs.ac.uk Internet Source	1%
11	export.arxiv.org Internet Source	1%
12	scholars.cityu.edu.hk Internet Source	1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		