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 imaging in medicine, 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, X-Rays, CAT scans, are used in order to identify blood clots, fractured bones, tumors, etc. Non-invasive medical imaging technology enables professionals to analyze a wide range of diseases without being obtrusive. AI is improving the ability to comprehend and evaluate datas in a variety of medical imaging systems. Computer vision is used to detect problems that aren't evident to the naked eye yet. Radiographers are university-trained experts with a thorough understanding of the nature of the human body and how diseases and accidents influence it.

II. STATE OF THE ART (LITERATURE SURVEY)

A. *General*

This chapter provides a detailed explanation of the many processes and strategies that we investigated and used to better understand the approach needed for our project. The reference articles we looked at aided us in finding effective answers to some of the problems we were having with cell nuclei detection and segmentation, as well as improving our algorithm to increase the speed, efficiency, and quality of semantic and instance detection in our system.

A. *Literature Survey Inference*

The study of satellite images is essential. Deep Learning in the field of computer vision has come a long way since the initial CNN, and because to new and improved methodologies, CNNs are now extremely powerful. 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*

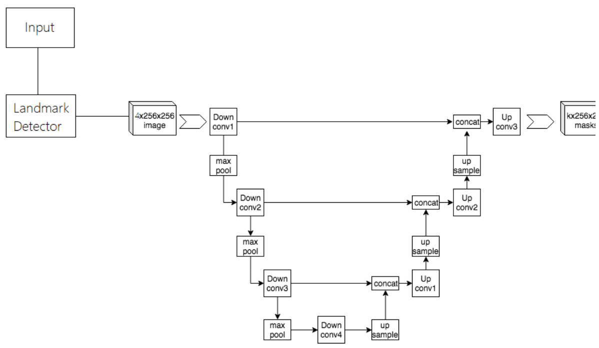

Fig 1. Network Architecture

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.

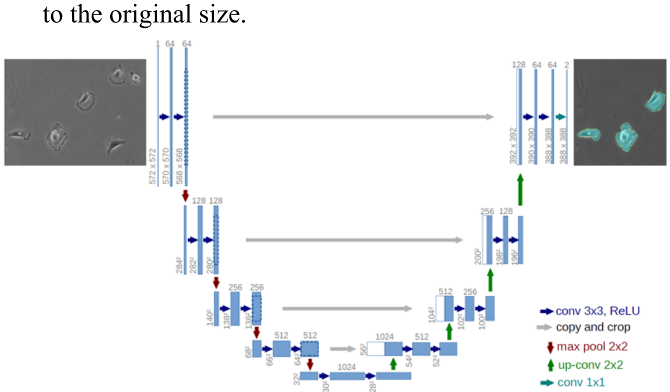

Fig 2. Traditional U-Net Model

C. *Elaborated U-Net Architecture*

The encoder (left side) is made up of two 3x3 convolutions that are applied repeatedly. After each convolution, there is a ReLU and batch normalisation. The spatial dimensions are then reduced using a 2x2 max pooling operation. The number of feature channels is doubled by halving the dimensions of spatial while downsampling step.

The extended path starts with an upsampling of the feature map, followed by a 2x2 transpose convolution, which reduces the number of feature channels by half. A 3x3 convolution and a concatenation with the matching feature map from the contracting path are also present. In the final layer, a 1x1 convolution is used to map the channels to the desired number of classes.

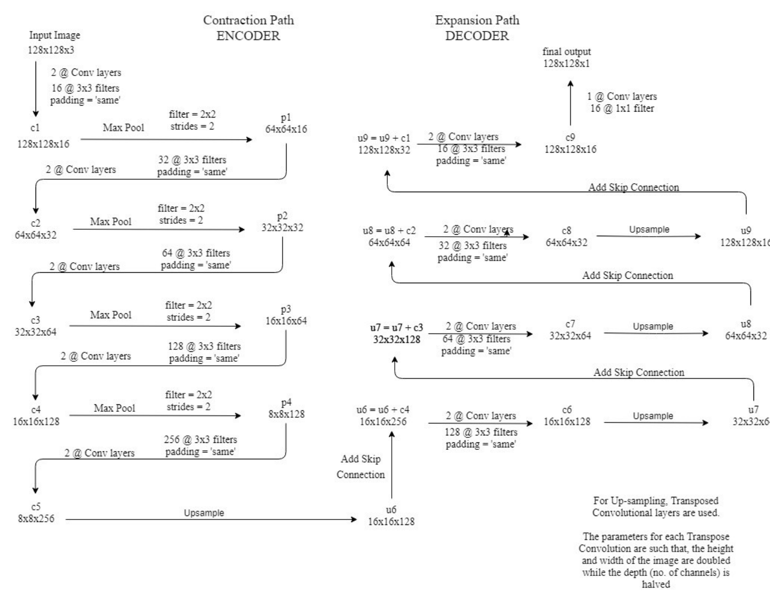

Fig 3. Elaborate U-Net Architecture

D. *Description of the U-net Architecture*

- The output tensors of Convolutional Layers are c1, c2,..., c9.
- The output tensors of Max Pooling Layers are p1, p2, p3, and p4.
- The output tensors of up-sampling (transposed convolutional) layers are u6, u7, u8, and u9.
- The contraction path (Encoder) on the left is where we apply standard convolutions and max pooling layers.
- The image size is gradually reduced as the depth is steadily increased in the Encoder. Starting at 128x128x3 and going up to 8x8x256
- This fundamentally means that the network has learned the "WHAT" information in the image, but not the "WHERE" information.
- The expansion path (Decoder) on the right hand side is where we apply transposed convolutions with normal convolutions.
- The image size steadily increases in the decoder, while the depth gradually decreases. Starting at 8x8x256 and going up to 128x128x1
- By gradually applying up-sampling, the Decoder recovers the "WHERE" information (precise localisation).
- We use skip connections at each step of the decoder to obtain more precise locations by concatenating the output of the transposed convolution layers with the feature maps from the Encoder at the same level:u6 + c4 = u6 ; u7 + c3 = u7; u8 + c2 = u8; u9 + c1 = u9; We use two consecutive regular convolutions after each concatenation so that the model can learn to create a more precise result.
- The symmetric U-shape of the design is what gives it the term UNET.
- We have the following relationship at a high level:
  "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

- IOU Score- Calculates an intersection of the predicted mask over the ground truth mask to check the ration of intersection to that of union between them.
- MaP Score- Generates a class vice precision score for each successful model recall on that class.
- Result Visualization- 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 - sourced software library based on the Torch library, largely created by Facebook's AI Research division for applications such as computer vision and natural language processing. It's free, open-source software licensed under the Modified BSD licence.

- Numpy

NumPy is a Programming framework that makes it possible to work with arrays. There are additionally matrices, fourier transforms, and linear algebra functions included. Travis Oliphant invented NumPy in 2005. We are free to use it being an open source platform. Numerical Python is referred to as NumPy. In Python, we have lists that act as arrays, but they are glitchy to process. NumPy intends to provide a 50-fold faster array object than standard Python lists. The array object in NumPy is named ndarray, and it comes with a slew of helper functions that make working with it a breeze. In data research, when speed and resources are critical, arrays are widely employed.

- 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 Python plotting library that synchronises with NumPy, the Python numerical mathematics extension. It offers an object-oriented API for embedding charts into applications that use 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 relatively simple language to learn and is nearly identical to English. It's used for a variety of tasks such as web development, scripting, and machine learning. It's also appropriate for servers and other applications.When compared to C++, Python is slower.

- Pycharm

PyCharm is an application framework for computer programming, with an emphasis on the Python programming language. JetBrains, a Czech firm, developed it.

V. IMPLEMENTATION

The UNet module was constructed using pytorch. The UNet essentially consists of 2 modules - encoder and decoder. The encoder (left side) is made up of two 3x3 convolutions that are applied repeatedly. After each conv, there is a ReLU and batch normalisation. The spatial dimensions are then reduced using a 2x2 max pooling operation. We double the number of feature channels while halving the spatial dimensions at each downsampling step. Path of the decoder (right side): Every step in the expanding route starts with an upsampling of the feature map and then a 2x2 transpose convolution, which reduces the number of feature channels in half. We also have a concatenation with the contracting path's matching feature map, as well as a 3x3 convolutional neural network (each followed by a ReLU). A 1x1 convolution is utilised to map the channels to the final layer.

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.