Faster Segmentation of Cardiac MRI using U-Net

Major Project - Second Review

By:
P. Saksham(RA1811003010142)
Ankita Samanta(RA1811003010172)

Guide Name: Dr. R. I. Minu

Contents

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 case, 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).

Motivation

- Time is a very important resource when it comes to diagnostics. A delay in the diagnostic process puts the patient at risk of the illness getting worse the longer it remains undiagnosed. However, trying to hurry up the process also makes it more likely for medical professionals to come up with inaccurate results.
- In the United States, 12 million people are affected by medical diagnostic errors each year. An estimated 40,000 to 80,000 people die annually from complications from these misdiagnoses.
- A 2015 Dutch study showed that AI diagnosis of prostate cancer using MRI (magnetic resonance imaging) was as good as that of human radiologists.
- Therefore automation could reduce the possible misdiagnosis or at the least, assist doctors in making better decisions.

Innovativeness

- Modification of the existing architecture combined with a unique set of dataset augmentations and hyperparameter tuning 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.

Scope of Project

- Computational Cardiology: - Using computers in areas of cell identification and disease prediction.
- Cardiovascular Research: - Identifying features and finding similarities between the various types of cardiovascular tissues to find remedies.
- Monitor Cardiac Diseases: Cardiac MRI helps your doctor detect or monitor cardiac disease by evaluating the anatomy and function of the heart chambers, heart valves, size of and blood flow through major vessels, and the surrounding structures such as the pericardium
- Heart Disease Treatment Planning: Planning a patient's treatment for cardiovascular disorders.

LITERATURE SURVEY

Ref No.	DATASET	OBJECTIVE	ALGORITHMS	RESULTS	REMARKS
[1]	Inria dataset	A modified U-Net architecture has designed and proposed. This new architecture is deep enough to extract contextual information from satellite imagery	k-means clustering algorithm	Successfully achieved better performance on very different biomedical segmentation application	In the proposed model convolution and ReLU layers are altered.
[2]	DeepGlobe road extraction dataset	Proposed a powerful dilated convolution module and successfully applied it into the improved UNet network structure	Semantic segmentation neural network – patch classification, dilated convolution model	Successfully improved the performance of the VGG16 model in the semantic segmentation of satellite images.	Our proposed network model the mIOU score of 5.43%
[3]	ImageNet dataset	To present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently	Convolution algorithms	Successfully achieved better performance on very different biomedical segmentation application	In the proposed model convolution and ReLU layers are altered.
[4]	COCO datasets	To introduce an alternative to bounding boxes as the intermediate object representation in the form of a set of points	Classification algorithms	Successfully changed the object representation from bounding boxes to RepPoints	N/A

[1] 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.
[2]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.
[3]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.
[4]Yang, Z., Liu, S., Hu, H., Wang, L., & Lin, S. (2019). Reppoints: Point set representation for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9657-9666).

Ref No.	DATASET	OBJECTIVE	ALGORITHMS	RESULTS	REMARKS
[5]	the original images are pre-processed, and a crack dataset is created	To proposed an improved Unet-based method to find crack, for evaluating the damage level of concrete structures	FFT-based algorithm	An improved Unet-based model was proposed for detecting surface cracks in pixel level	the size of the dataset and the depth of the model influence the performance of DL-based model is studied.
[6]	ImageNet and PASCAL VOC datasets	To proposed a method exhibits higher classification accuracy than the U-Net and ResNet networks	classification algorithm	Successfully proposed an improved Res-UNet network for tree classification using high-scoring remote sensing images	overall accuracy - 87.51%, average accuracy - 85.43%, and Kappa coefficients - 84.21%
[7]	ShapeNet-pairs and ShapeNet-triplets	To propose a method to detect and reconstruct multiple 3D objects from a single RGB image	Convolution algorithms	Successfully presented an end-to-end trainable model for realistic and joint 3D multi object reconstruction from a single input RGB image.	model encourages collision-free reconstructions and uses CAD models as shape representations to guarantee valid and realistic object shapes.
[8]	ISPRS Potsdam, Vaihingen	to proposed a novel end-toend semantic segmentation network for high-resolution aerial imagery, namely Context and Semantic Enhanced UNet (CSE-UNet),	Semantic segmentation neural network	The proposed architecture equips UNet with multi-level RFB-based skip pathways and a multi-kernel dual-path encoder to resolve the issues of intra-class heterogeneity and inter-class homogeneity separately.	N/A

[5]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.
[6]Cao, K., & Zhang, X. (2020). An improved res-unet model for tree species classification using airborne high-resolution images.
[7]Engelmann, F., Rematas, K., Leibe, B., & Ferrari, V. (2021). From points to multi-object 3D reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4588-4597).
[8]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.

Ref No.	DATASET	OBJECTIVE	ALGORITHMS	RESULTS	REMARKS
[9]	Cityscapes dataset	proposed a deep architecture for semantic segmentation based on an asymmetry encoder- decoder architecture using Ghost-Net and U-Net	Segmentation algorithm- Conditional Random Field (CRF) and Atros Spatial Pyramid Pooling (ASPP)	Successfully trained and validated process of the proposed Ghost-Unet model	The proposed model has good pixel accuracy and mean Intersection over Union (mIoU) compared with other valid literature.
[10]	DDSM and INbreast	proposed an end-to-end UNet model for the detection, segmentation, and classification of breast masses in one-stage	Segmentation algorithm	The proposed model detects, segments, and classifies in one-stage each pixel of the mammogram images into normal, benign, and malignant.	dice coefficient of 99.20% and 99.56% and a weighted F1-score of 99.19% and 99 or the DDSM and INbreast datasets, respectively.
[11]	SIFT Flow	Introduced the concept of a fully CNN	VGG CNN	Considerably reduces model footprint and makes feature extraction and combination easy.	Bounding Box regression is not possible
[12]	PASCAL VOC07	Improved the R-CNN speed by using an ROI pooling module	Fast R-CNN	ROI pooling culls the large number of regions	Network not end-to-end trainable

[9]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.
[10]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.
[11]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).
[12]Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
`

Ref No.	DATASET	OBJECTIVE	ALGORITHMS	RESULTS	REMARKS
[13]	COCO	Improved the Faster R-CNN by adding a convolutional branch which can generate a binary mask	Mask R-CNN	The network is capable of producing a binary mask to detect and localize irregular objects.	Complicates architecture further
[14]	ILSVRC and COCO	Further improved the R-CNN speed by introducing a RPN	Faster R-CNN	The network becomes end-to-end trainable and regions are further refined for better results.	Complex Architecture
[15]	COCO	Proposes a model for cervical cancer segmentation	Mask R-CNN	Uses attention mechanisms to combine fixed sized ROIs with original ROIs	N/A

[13]He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
[14]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*.
[15]Ma, B., Zhang, J., Cao, F., & He, Y. (2020). MACD R-CNN: an abnormal cell nucleus detection method. *IEEE Access*, *8*, 166658-166669.

Problem Statement

- To perform a 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.
- We also hope to improve the DICE score by finding an optimal set of hyperparameters and a modification to the traditionally used U-Net.

Project Objective

- To create an optimal training pipeline which requires less data to train and is faster at producing results to .
- This model can be used by cardiologists as it automates the process of medical image diagnosis and thus reduces the possibility of humans making an error.
- We also hope to reduce this time taken to identify the masks by tweaking the following parameters:
    - Data augmentation
    - Hyperparameters
    - Model architecture

Comparison of Existing Models

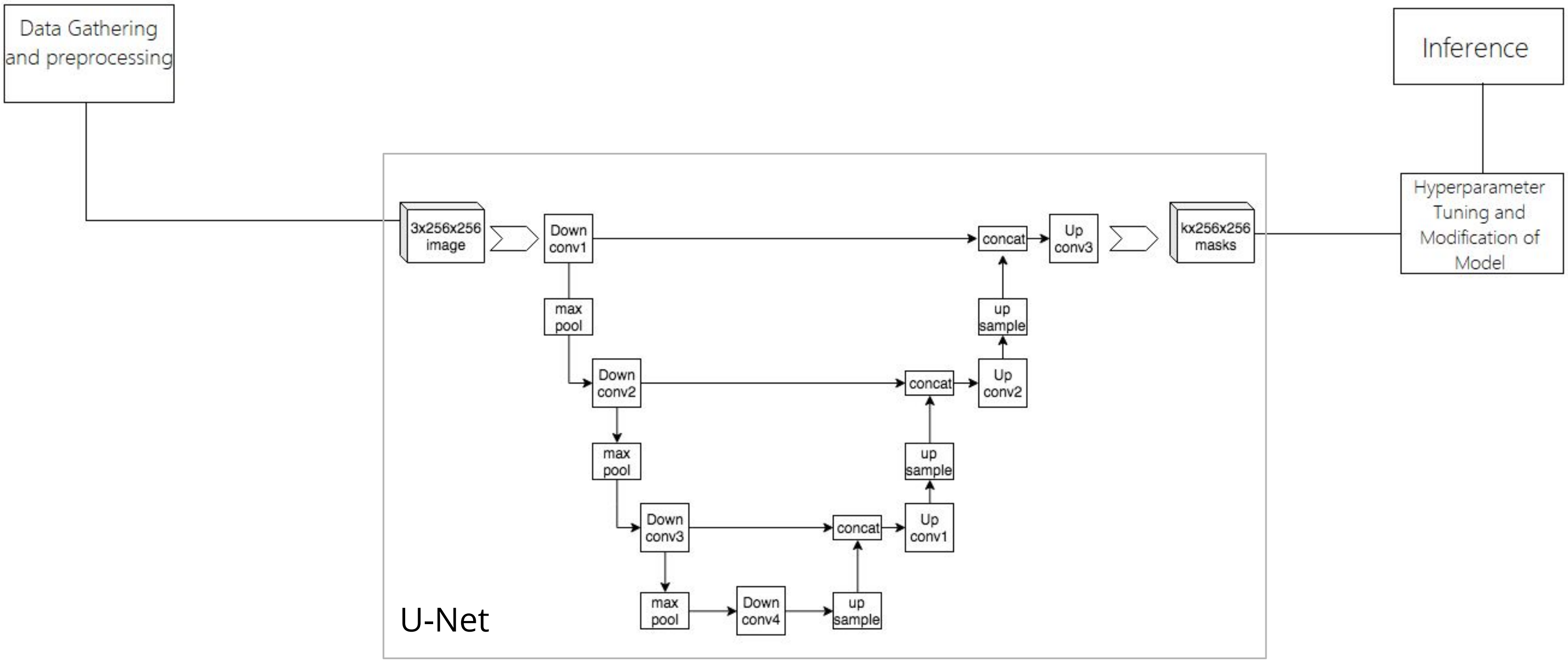
- Medical image analysis plays an important role.
- Deep Learning in the field of computer vision has developed a lot since the first CNN and now CNNs are very powerful thanks to new and improved techniques
- This poses a problem as now CNNs tend to overfit data and thus more focus is required on building 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 tumour cells before they mature and even help in expediting critical cures
- When MRI scans diagnosis is automated, error made because of humans is greatly reduced.

Challenges and Limitations in Existing System

1. Existing semantic segmentation techniques mostly use U-Net.
2. A drawback related to U-Net is because of the upsampling or the decoder half of the architecture.
3. While encoding the input, a lot of information is lost and therefore while upsampling the feature, a lot of noise is present in the features which produces an inferior quality of mask.
4. Also, traditional techniques use large amounts of training data which is hard to train.

Architectures/Block Diagram of the proposed model

Overall Architecture

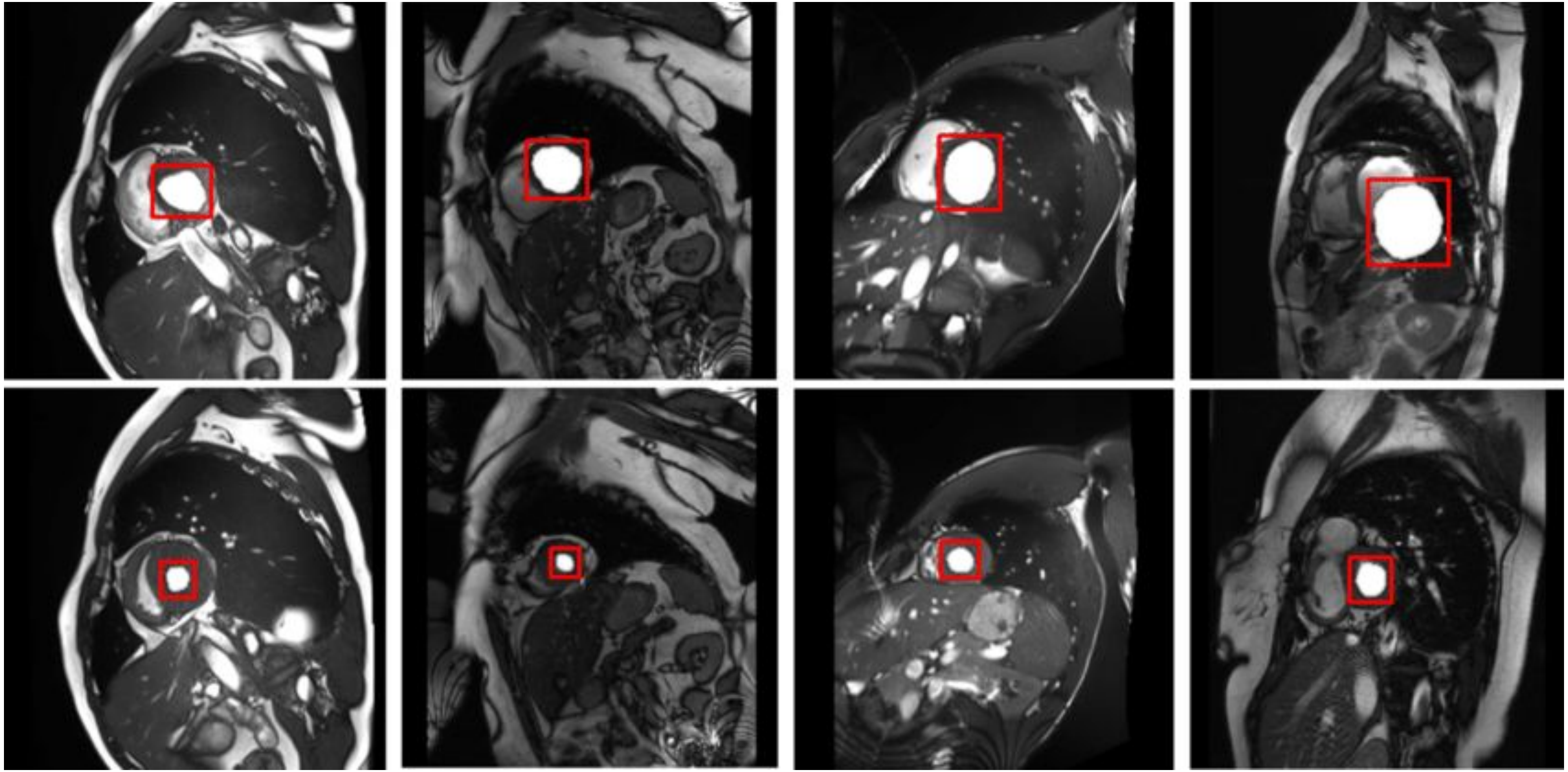


Module Description

- Dataset gathering and preprocessing: We are using a publicly available dataset called the Adverse Conditions Dataset(ACDC) which we obtained from https://acdc.vision.ee.ethz.ch/.

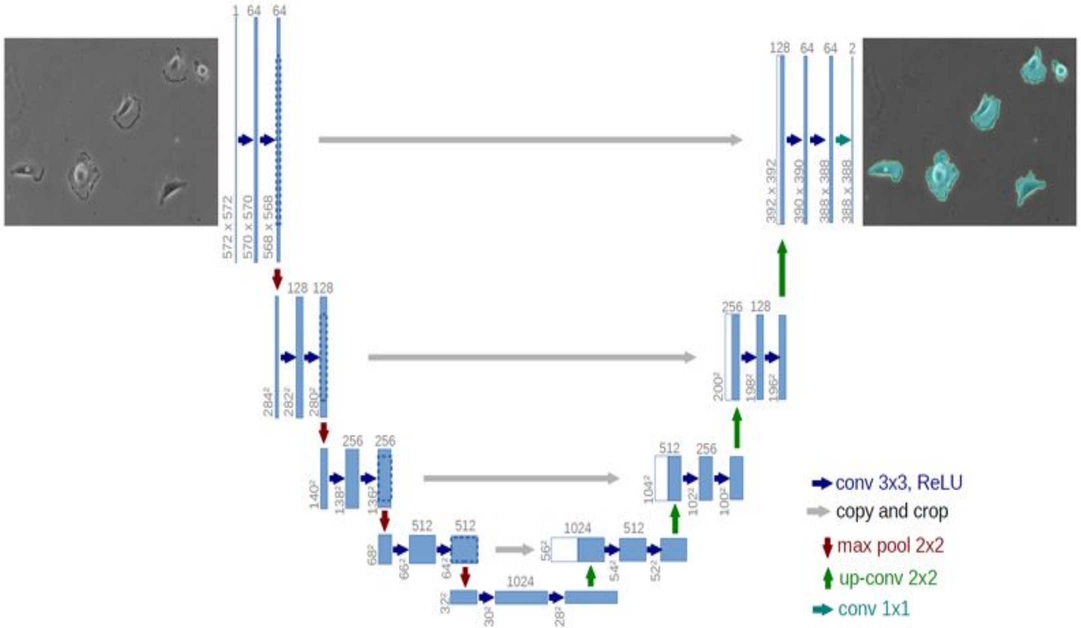
    a. The publicly-available ACDC dataset consists of 200 short-axis cine-MRI scans from 100 patients, evenly distributed in 5 subgroups: normal, myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, and abnormal right ventricles.
    b. Scans correspond to end-diastolic (ED) and end-systolic (ES) phases, and were acquired on 1.5T and 3T systems with resolutions ranging from 0.70 × 0.70 mm to 1.92 × 1.92 mm in-plane and 5 mm to 10 mm through-plane.
    c. Segmentation masks delineate 4 regions of interest: left ventricle endocardium (LV), left ventricle myocardium (Myo), right ventricle endocardium (RV), and background.
    d. Normalized slices are then cropped to 384 × 384 pixels.

Sample Dataset Scans



Module Description(Continued...)

- Model Building: We developed a U-Net model by using pytorch. The U-Net module basically consists of two segments as explained below:



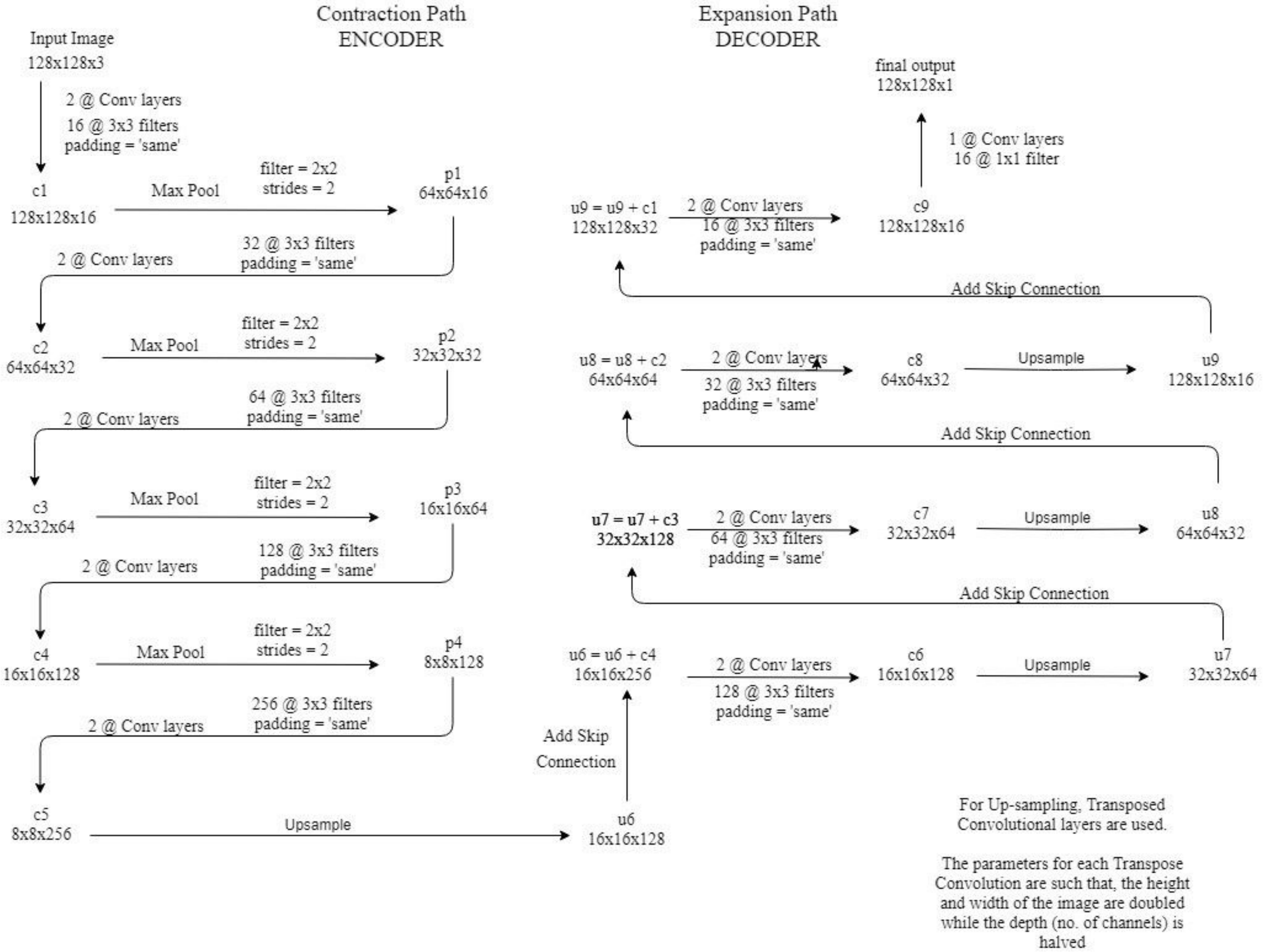
U-Net module consists essentially of two parts -

1. Successive convolutions
2. Successive deconvolutions

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

Module Description(Continued...)

- 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 (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.



Contraction Path
ENCODER
Expansion Path
DECODER
Input Image
128x128x3
2 @ Conv layers
16 @ 3x3 filters
padding = 'same'
c1
128x128x16
Max Pool
filter = 2x2
strides = 2
p1
64x64x16
2 @ Conv layers
32 @ 3x3 filters
padding = 'same'
c2
64x64x32
Max Pool
filter = 2x2
strides = 2
p2
32x32x32
2 @ Conv layers
64 @ 3x3 filters
padding = 'same'
c3
32x32x64
Max Pool
filter = 2x2
strides = 2
p3
16x16x64
2 @ Conv layers
128 @ 3x3 filters
padding = 'same'
c4
16x16x128
Max Pool
filter = 2x2
strides = 2
p4
8x8x128
2 @ Conv layers
256 @ 3x3 filters
padding = 'same'
c5
8x8x256
Upsample
u6
16x16x128
Add Skip
Connection
u6 = u6 + c4
16x16x256
2 @ Conv layers
128 @ 3x3 filters
padding = 'same'
c6
16x16x128
Upsample
u7
32x32x64
Add Skip Connection
u7 = u7 + c3
32x32x128
2 @ Conv layers
64 @ 3x3 filters
padding = 'same'
c7
32x32x64
Upsample
u8
64x64x32
Add Skip Connection
u8 = u8 + c2
64x64x64
2 @ Conv layers
32 @ 3x3 filters
padding = 'same'
c8
64x64x32
Upsample
u9
128x128x16
Add Skip Connection
u9 = u9 + c1
128x128x32
2 @ Conv layers
16 @ 3x3 filters
padding = 'same'
c9
128x128x16
1 @ Conv layers
16 @ 1x1 filter
final output
128x128x1
For Up-sampling, Transposed Convolutional layers are used.
The parameters for each Transpose Convolution are such that, the height and width of the image are doubled while the depth (no. of channels) is halved

Module Description(Continued...)

- Hyperparameter tuning and model modification: 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 underfits 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

Module Description(Continued...)

- Inference: Visualizing the results of training by testing on separate set and generating metrics to compare efficiency
  - Dice Score(Sørensen–Dice coefficient)- 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 the library matplotlib to create contours of masks produced which can be further used to segment cardiac MRIs.

U-Net++ for Benchmarking

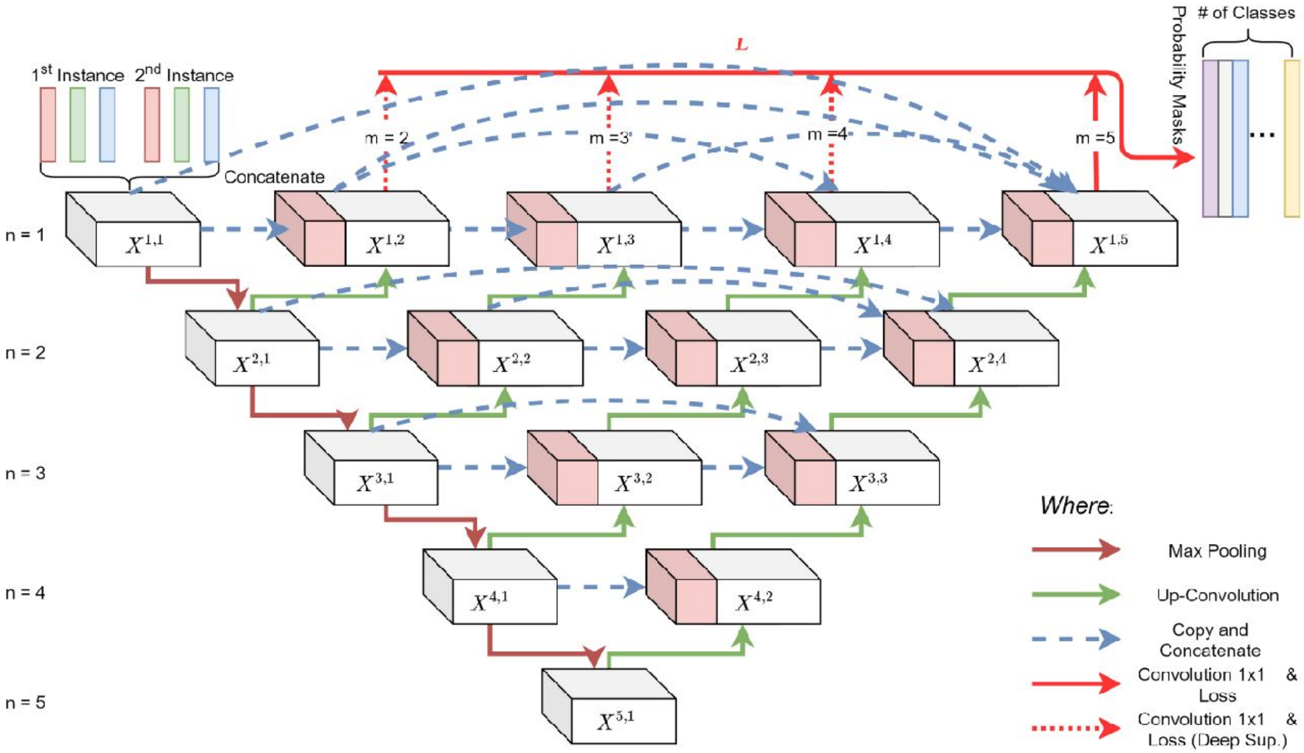
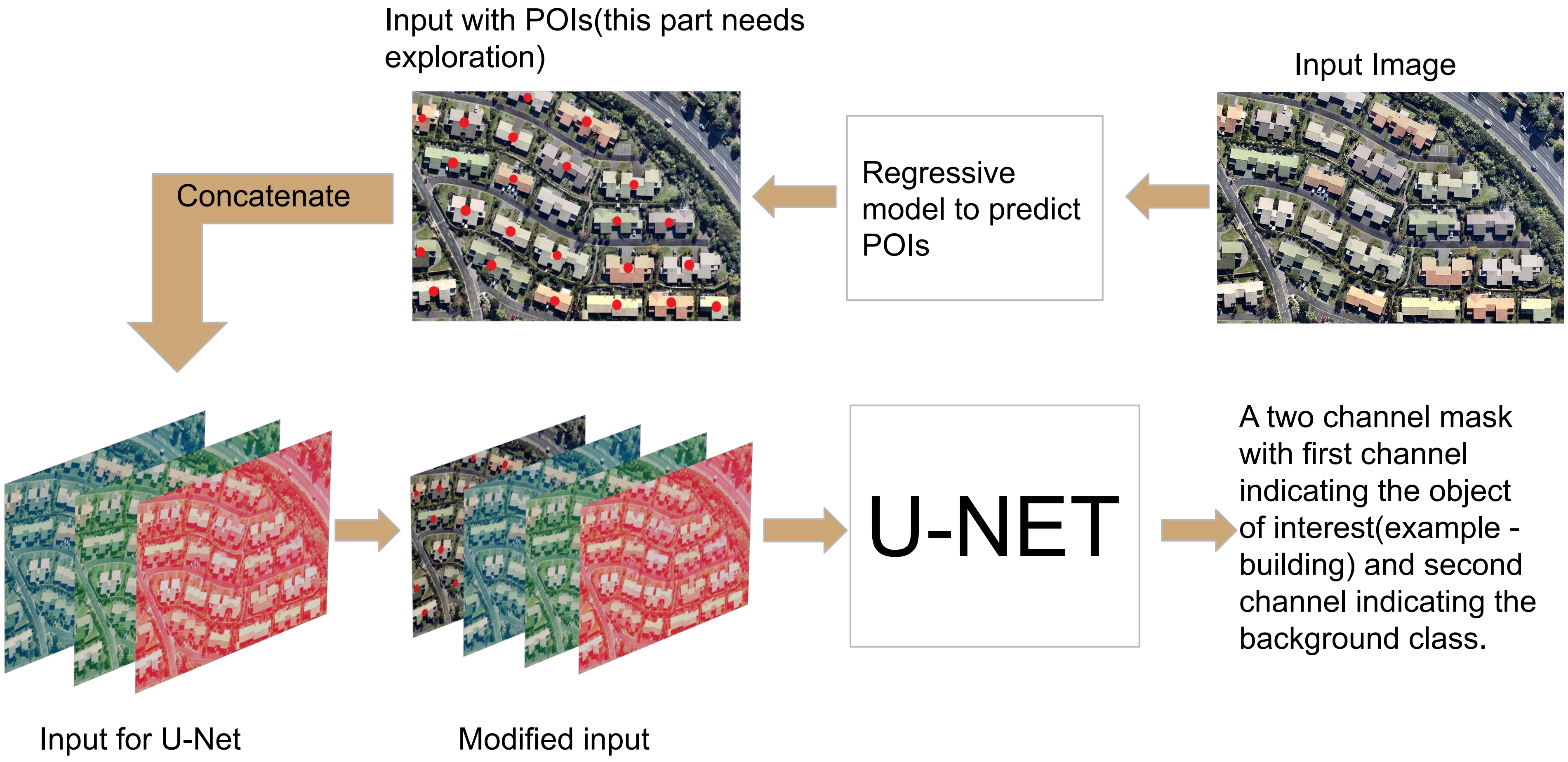


Figure 2: Nested UNet Architecture.

- Is just an updated version of U-Net.
- Designed like a spider's web for carrying forward information that is lost during downsampling to the decoder half of the architecture.

IDEA FROM MINOR PROJECT


Input with POIs(this part needs exploration)
Input Image
Regressive model to predict POIs
Concatenate
U-NET
A two channel mask with first channel indicating the object of interest(example - building) and second channel indicating the background class.
Input for U-Net
Modified input

Implementation

- So far we have successfully developed the training pipeline required to train the MRI scans
- We have also implemented the U-Net and U-Net++ architectures which will be used as backbone architecture and benchmarking architecture respectively.
- We also have training results from U-Net and U-Net++ which we will use for benchmarking.

Tools used for implementation

Hardware Requirements

- GPUs for model graph computations - NVIDIA GTX 1050ti

Software Requirements

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

Intermediate Results

	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

The following table represents the Dice score for various number of MRI scans when used to train these 2 models.

$$\text{Dice score} = \frac{2 \cdot |A \cap B|}{2 \cdot |A \cap B| + |B \setminus A| + |A \setminus B|} = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

Discussions

- 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

- 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).
- 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.
- 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).
- 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.
- 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.
- 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.
- 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.
- Cao, K., & Zhang, X. (2020). An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sensing*, *12*(7), 1128.
- 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.
- 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.
- 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.
- 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).
- 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.
- Ma, B., Zhang, J., Cao, F., & He, Y. (2020). MACD R-CNN: An Abnormal Cell Nucleus Detection Method. *IEEE Access*, *8*, 166658-166669.

THANK YOU