Brain MRI Segmentation

Table of Contents

1. INTRODUCTION

Semantic segmentation of brain MRI images is a crucial task in medical imaging, as it can help doctors diagnose various neurological disorders. In this project, different encoders for semantic segmentation of brain MRI images using the U-Net architecture were explored. The performance of different encoders was evaluated and compared to determine the most effective encoder for the task.

2. DATA ACQUISITION

Data source:
https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation

This dataset contains brain MR images together with manual FLAIR abnormality segmentation masks. The images were obtained from The Cancer Imaging Archive (TCIA).
They correspond to 110 patients included in The Cancer Genome Atlas (TCGA) lower-grade glioma collection with at least fluid-attenuated inversion recovery (FLAIR) sequence and genomic cluster data available.

3. DATA PREPROCESSING

3.1 Loading the image and mask paths

A list of all files containing the word 'mask' was created for later use.

3.2 Create data frame and train test split

There are 2828 training examples, 708 validation samples and 393 test examples.

3.3 Data generator, data augmentation and adjust data

Defined a function that can generate image and mask at the same time by referring a code that's widely used when doing semantic segmentation using Unet. Since the dataset has only 2828 training examples, which is small for a deep neural network, so data augmentation was implemented. Data augmentation includes rotation, width and heigh shift, random zoom, and flipping.

3.4 Normalize and Diagnose

Defined a function that labels mask. After mask normalization if the value is <= 0.5 then that mask will be considered a complete black one and does not have any tumor

4. DEFINE MERICS & LOSS

4.1 Metrics

The Dice coefficient, also known as the F1-score, is a measure of the overlap between the predicted segmentation mask and the ground truth mask. It is defined as twice the intersection between the two masks divided by the sum of the sizes of the masks:

Dice = 2 * Intersection / (Size_Predicted + Size_GroundTruth)

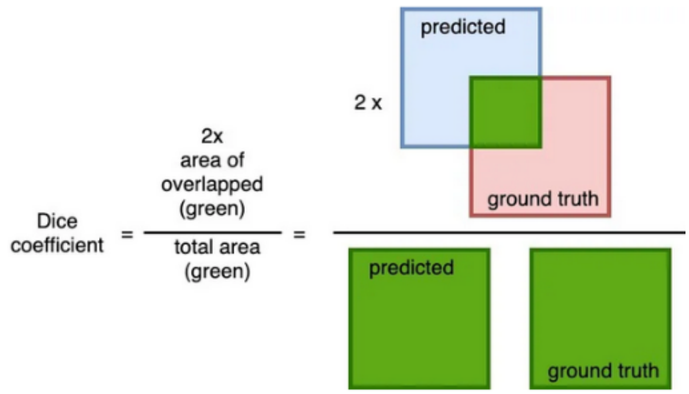

Figure 1: Dice Coefficient

The IoU, also known as the Jaccard index, is another measure of the overlap between the predicted and ground truth masks. It is defined as the ratio of the intersection over the union of the two masks:

IoU = Intersection / Union

Figure 2: IoU

Both metrics have values between 0 and 1, where a score of 1 indicates a perfect match between the predicted and ground truth masks.

The subtle difference between them is that the dice score tends to veer towards the average performance. Whereas the IOU helps one understand worst case performance. In practice, they're often both used.

4.2 Loss

Define loss as 1- dice score.

5. UNET FOR SEMANTIC SEGMENTATION

U-Net is a convolutional neural network that was developed for biomedical image segmentation at the Computer Science Department of the University of Freiburg.

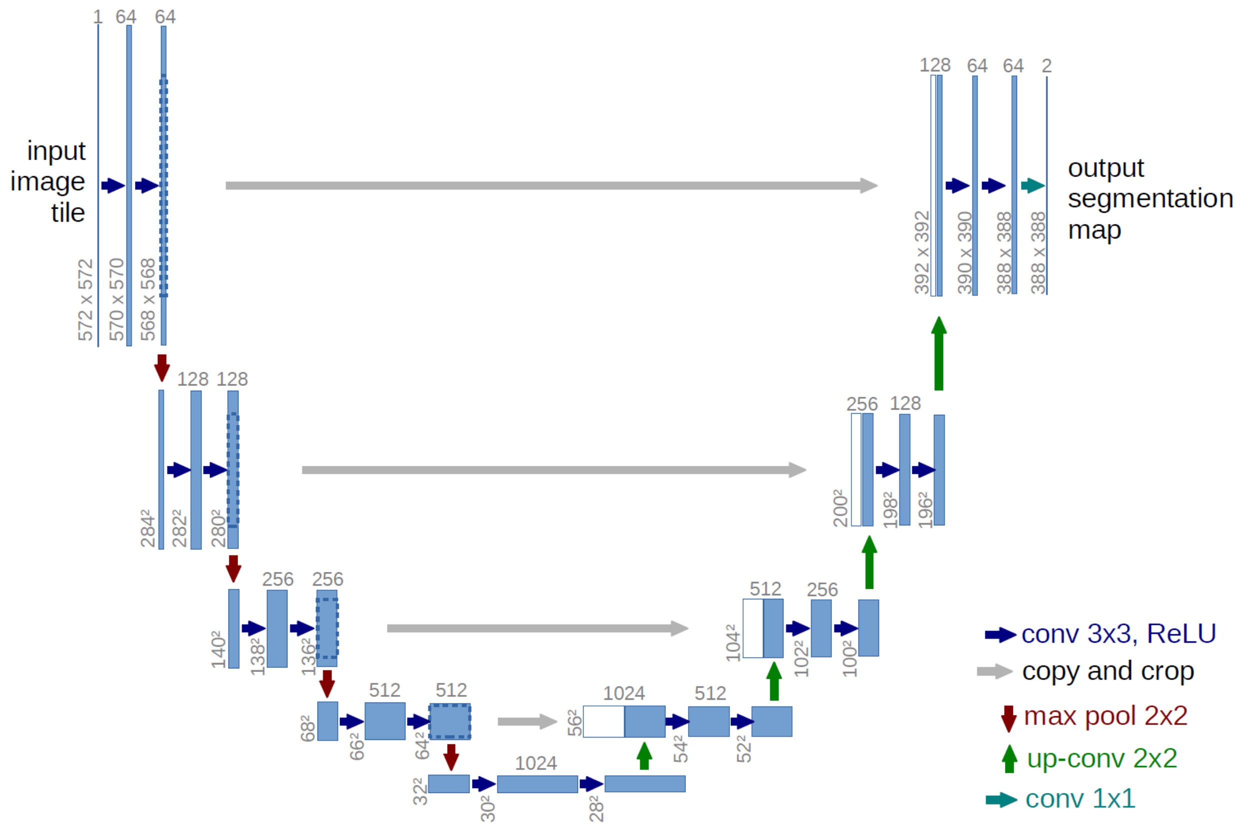

Figure 3: Unet

The U-Net architecture gets its name from its U-shape, which consists of a contracting path and an expanding path. The contracting path consists of several layers of convolution, pooling, and dropout operations, which gradually reduce the spatial resolution of the input and increase the number of feature maps. This is similar to the encoder of a typical convolutional neural network. The expanding path of the U-Net consists of a series of up-convolutional layers, which gradually increase the spatial resolution of the features while decreasing the number of feature maps. The up-convolutional layers are followed by concatenation with the corresponding feature maps from the contracting path at the same spatial resolution, which provides high-resolution context information to the decoder.

At the end of the expanding path, a final convolutional layer is applied to produce a pixel-wise segmentation mask. The output of this layer has the same spatial resolution as the input image and contains a probability map of the segmentation class for each pixel.

One key feature of U-Net is its use of skip connections, which provide a direct connection between the contracting and expanding paths. These connections help to preserve spatial information and improve the localization accuracy of the segmentation mask.

6. MODELING WITH DIFFERENT ENCODER

6.1 Generic Unet (no pretrained encoders)

U-Net is a specialized convolutional neural network architecture designed for biomedical image segmentation tasks. It features a U-shaped structure with a series of contracting (encoder) and expanding (decoder) layers connected by a bottleneck. The architecture leverages skip connections between corresponding layers of the encoder and decoder, which helps to retain spatial information and improve localization accuracy. U-Net has been widely adopted for various segmentation tasks due to its efficient training and precise segmentation capabilities. A Unet model with no pretrained encoder was trained 120 epochs. The model achieved great results:

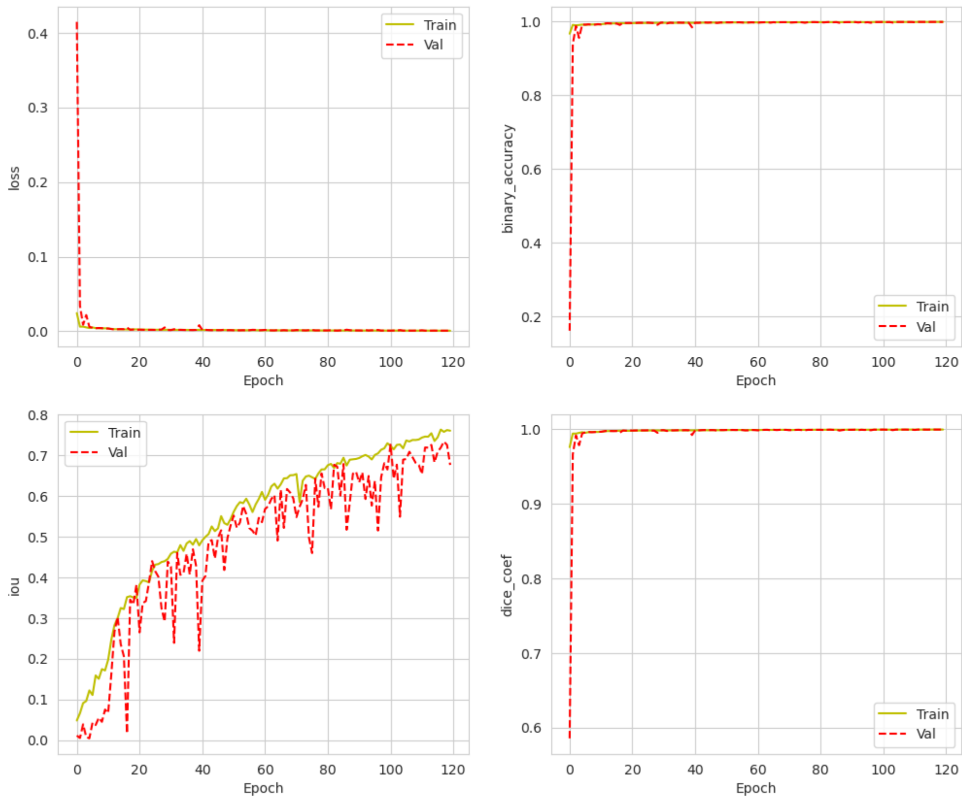

Figure 4: Unet training curve

6.2 VGG16 (retrain all layer)

VGG16 is a deep convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford for image recognition tasks. It consists of 16 weight layers, including 13 convolutional layers with small 3x3 filters, followed by max-pooling layers and three fully connected layers. VGG16 is known for its simplicity, yet effective performance, setting a benchmark in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). However, it has a large number of parameters, resulting in high computational and memory requirements.

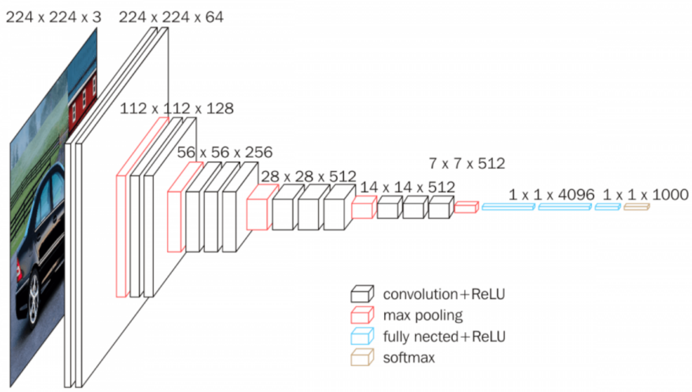

Figure 5: VGG16 architecture

Unet architecture with encoder of VGG16 pretrained weights that was trained on ImageNet give slightly worse results. But the validation dataset curve oscillates a lot more.

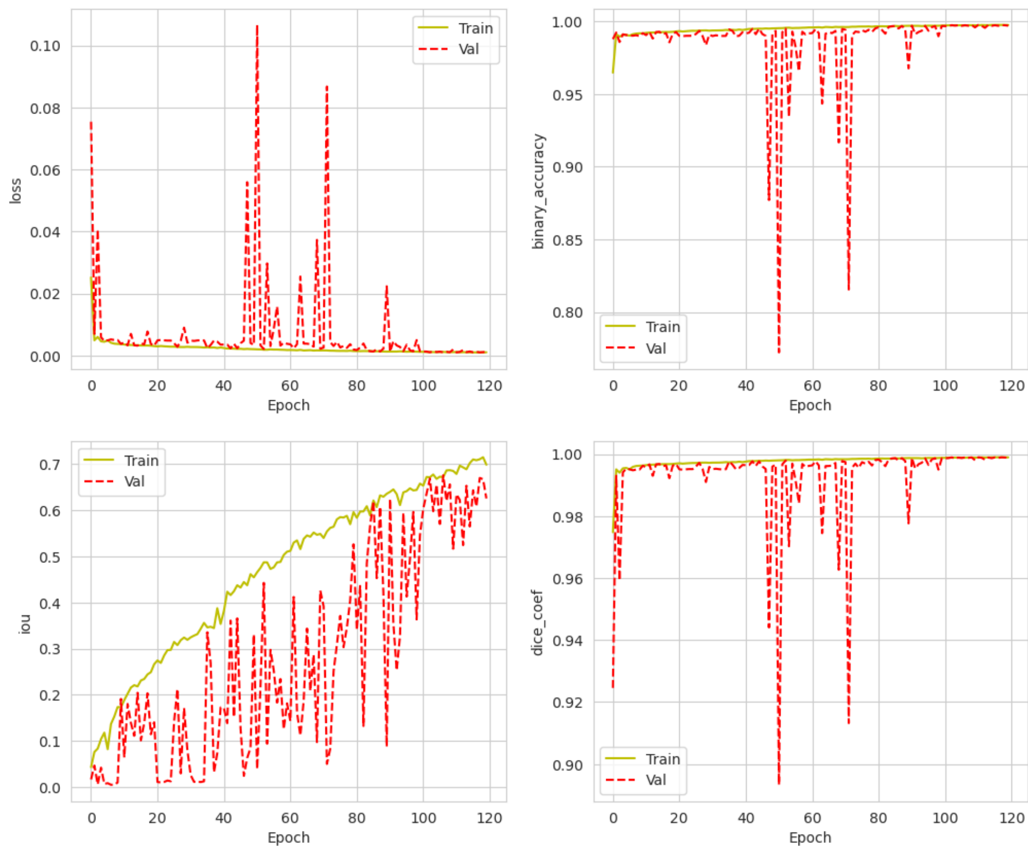

Figure 6: VGG16 backbone training curve

6.3 ResNet50 (retrain all layer)

ResNet, introduced by Microsoft Research in 2015, aimed to tackle the vanishing gradient problem associated with training deep neural networks. The main innovation of ResNet is the introduction of residual connections (also known as skip connections) that allow the network to learn residual mappings, bypassing certain layers and enabling the flow of gradients throughout the network. This architecture allows for much deeper networks without a decrease in accuracy, and in some cases, it even improves accuracy with increasing depth. ResNet has set new performance benchmarks in various computer vision tasks and has been widely adopted in the research community.

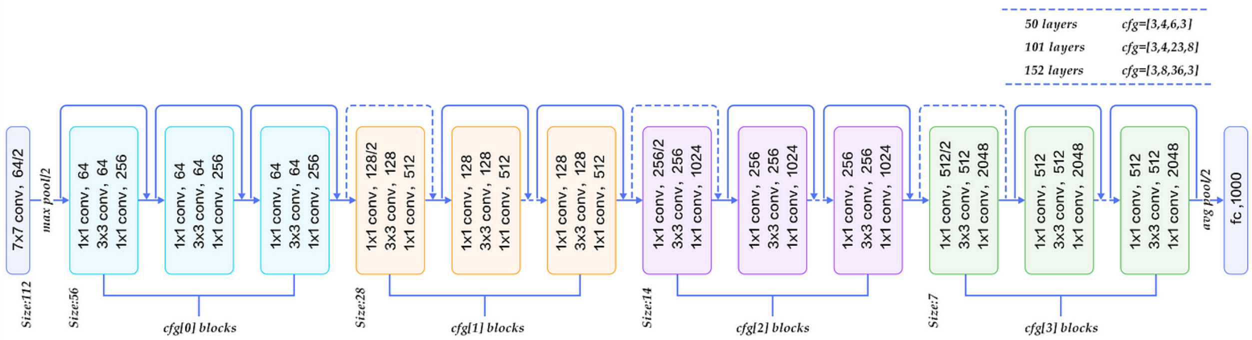

Figure 7: ResNet50 Architecture

Unet with ResNet50 as backbone retrained all layers gave the best results. IoU score improves faster compared to other models.

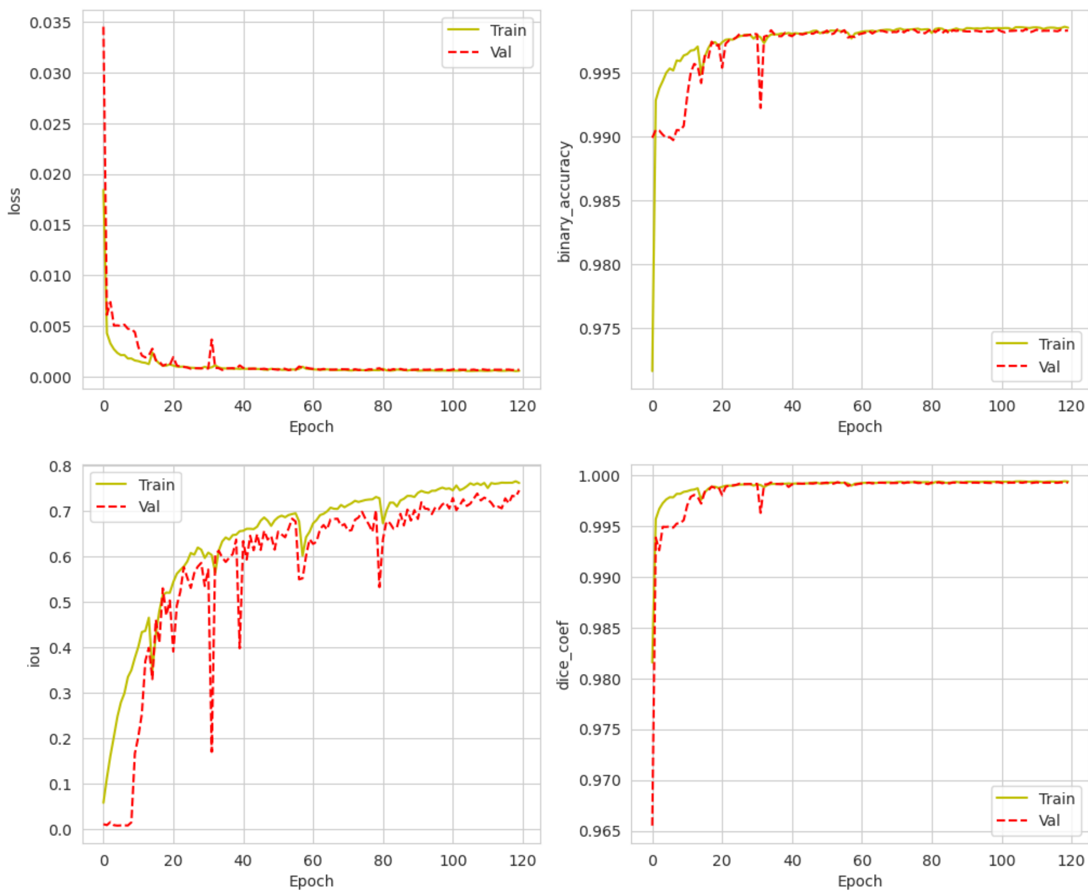

Figure 8: ResNet50 (retrain all layer) learning curve

At last, Unet with ResNet50 as backbone by freezing first two conv blocks were trained, it gave the worst results.

6.4 ResNet50 (freeze first two conv blocks)

Unet with ResNet50 as backbone by freeze first two conv blocks gave the worst results. IoU score improves fasted compared to other models.

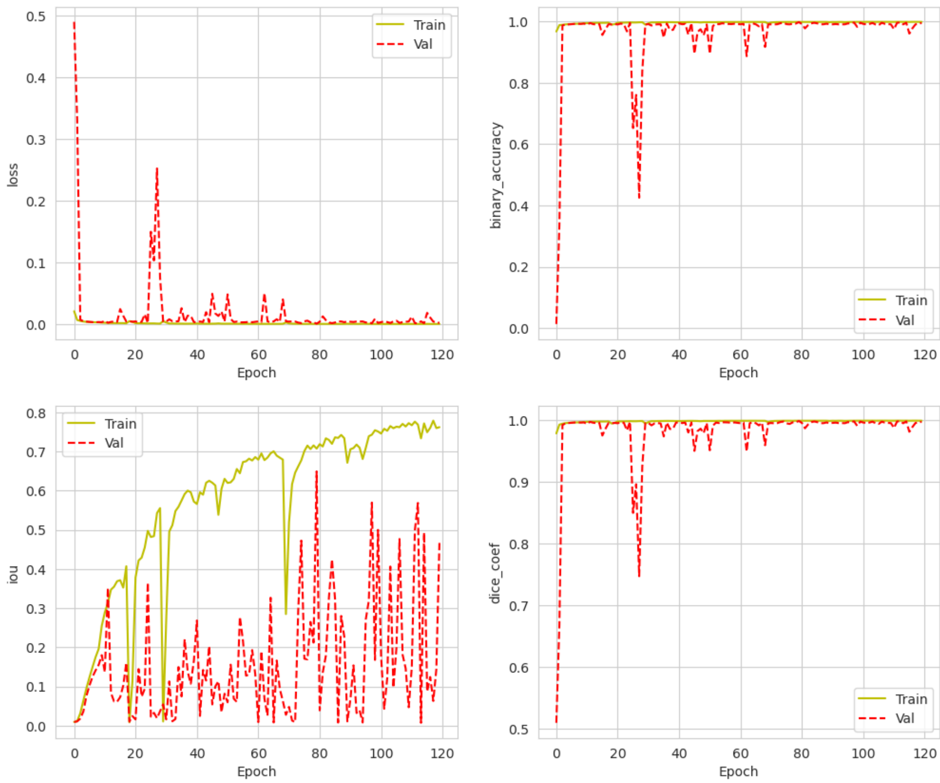

Figure 9: ResNet50 (freeze first two conv blocks) learning curve

When the segmentation results were plotted, the first three models gave similar masks:

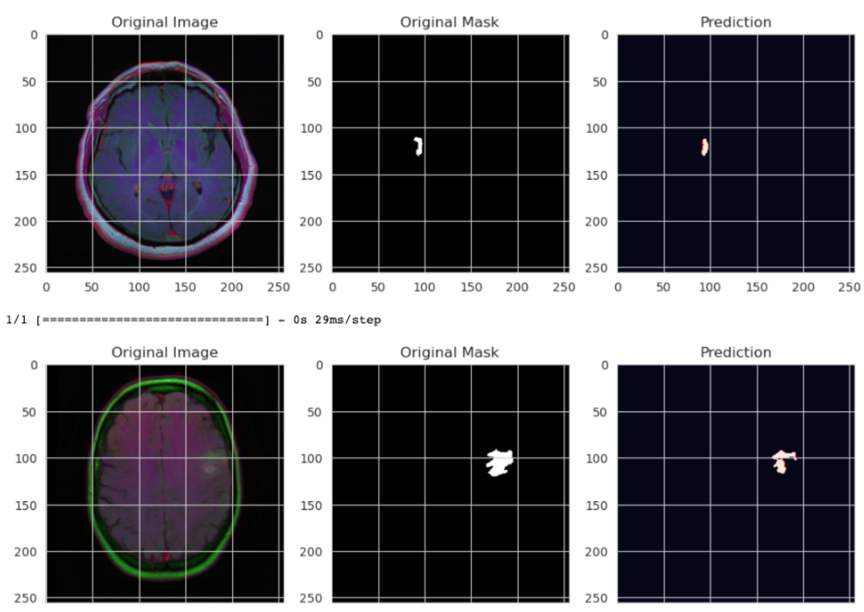

Figure 10: ResNet50 (retrain all layer) segmentation masks

Although the algorithm partially preserved edges, the boundaries between adjacent classes were occasionally blurry or imprecise.

7. CONCLUSION

	unet	vgg16	resnet50	partial_resnet50
Loss	0.000714	0.001196	0.000690	0.002926
Binary_Accuracy	0.998279	0.997081	0.998334	0.993751
IoU	0.703022	0.639999	0.732958	0.468088
Dice_Coefficient	0.999286	0.998804	0.999310	0.997074

Figure 11: Performance Metrics Comparison

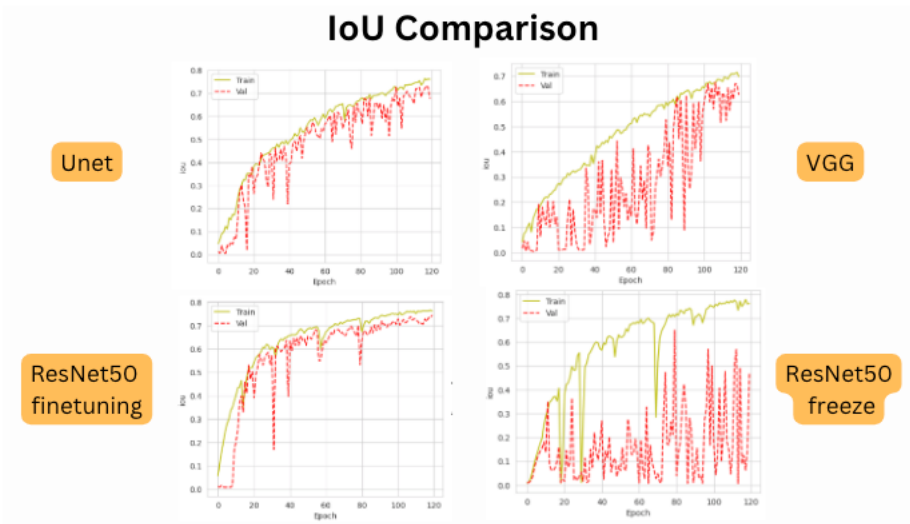

Figure 12: IoU Comparison

- Unet and resnet50 (retrain all layer) gave better results.
- Resnet50 (freeze first two conv blocks) gave the worst results, so the pretrained weights that was trained on ImageNet classification task didn't work well on medical image segmentation.
- Resnet50 (retrain all layer) converge faster then Unet, so the pertained weights helped

In the future, the following could be done to improve the performance:

- Increase training epochs of Unet and ResNet50 (retrain all layer) to improve the results
- Find pretrained models that were trained on medical images, then use the pretrained weights to do transfer learning.
- Try cyclical learning rate to fasten training and improve model performance.