ENM-5310 Final Project Report:
Comparative Evaluation of Brain MRI Segmentation Using Unet Models: Unet, Attention Unet, ResUnet, and Unet++

Prashanth Iyer, Rishi Ghia

May 2023

1 Abstract

In this project, we explore the performance of four advanced U-Net architectures for brain MRI segmentation: U-Net, Attention U-Net, Res U-Net, and U-Net++. These models represent different advancements of the original U-Net, focusing on specific improvements such as attention mechanisms, residual learning, and nested connections. We implement each model using Python notebooks and train them on a dataset of brain MRI scans, aiming to accurately segment brain tumors. The models are compared based on segmentation performance, architectural complexity, computational requirements, and their ability to capture fine-grained details. Our findings provide valuable insights into the strengths and weaknesses of each model, allowing informed decisions in selecting the most appropriate architecture for brain MRI segmentation tasks. By leveraging these advanced U-Net architectures, we can potentially improve the accuracy and reliability of automated brain tumor segmentation, contributing to more effective diagnosis and treatment planning in clinical practice.

2 Introduction

Brain tumor segmentation from magnetic resonance imaging (MRI) is a crucial task in medical image analysis, assisting in the diagnosis, treatment planning, and monitoring of brain tumor patients. Manual segmentation, which is the traditional approach, is time-consuming, error-prone, and subject to inter- and intra-observer variability. Consequently, there is a growing interest in developing automated segmentation methods to overcome these limitations.

Deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in various image segmentation tasks. Among these, the U-Net architecture has gained popularity in biomedical image segmentation due to its ability to generate accurate segmentation masks with relatively fewer training samples. In this project, we investigate four advanced U-Net architectures to determine their suitability for brain MRI segmentation tasks.

3 Motivation

The motivation for this project stems from the need to improve the accuracy and efficiency of brain tumor segmentation in MRI scans. By comparing different advanced U-Net architectures, we aim to identify the most suitable model for our task, considering specific requirements, available computational resources, and the desired trade-off between simplicity and advanced features. The investigated models include:

- U-Net: A simple and effective architecture widely used in biomedical image segmentation tasks.
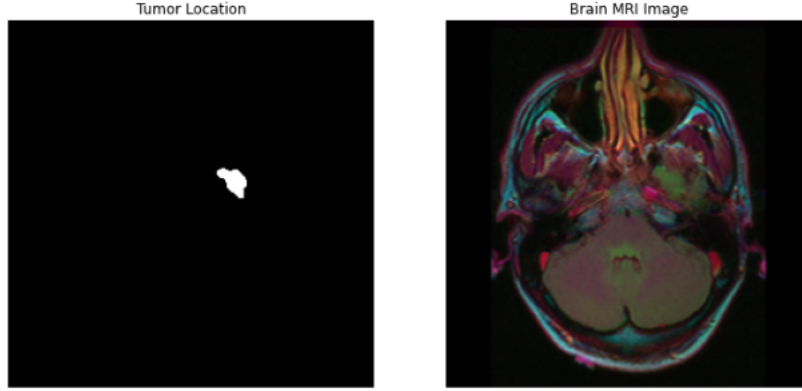
Figure 1: Sample MRI Image with Tumor mask

- Attention U-Net: Incorporates attention mechanisms to selectively focus on relevant features and ignore irrelevant ones, potentially improving segmentation accuracy.

- Res U-Net: Combines U-Net architecture with residual learning, addressing the vanishing gradient problem and allowing for deeper networks.

- Res U-Net: Combines U-Net architecture with residual learning, addressing the vanishing gradient problem and allowing for deeper networks.

- U-Net++: Introduces nested and dense connections between the contracting and expanding paths, refining segmentation masks at multiple resolutions and improving gradient flow during training.

4 Implementation

4.1 Dataset

In our project, we utilized the LGG-MRI Segmentation dataset available on Kaggle, which Mateusz Buda contributed. This dataset contains brain MRI scans of patients with low-grade glioma (LGG), a type of brain tumor. The purpose of our study was to develop a segmentation model to identify the tumor regions in these MRI scans, which would help medical professionals in diagnosing and treating LGG more effectively.

The dataset includes 110 patients' MRI scans, and each patient has a unique identifier. The MRI scans are in the TIFF format and come in three different modalities: pre-contrast, FLAIR (Fluid-attenuated inversion recovery), and post-contrast. There are 3-30 slices per patient, and each slice is 256x256 pixels. The corresponding ground truth masks are provided, indicating the tumor regions in each slice. These masks are binary, with a value of 1 for the tumor region and 0 for the non-tumor area.

In addition to the MRI scans and ground truth masks, the dataset contains a CSV file with patients' information such as age, survival status, and whether they received chemotherapy and radiation therapy. This information can be used for further analysis or to create additional features for the segmentation model. However, in our analysis we use only the MRI scans and their corresponding ground truth masks.

In our project, we encountered a significant class imbalance in the brain MRI segmentation dataset, which is a common challenge when dealing with medical imaging data. The dataset's
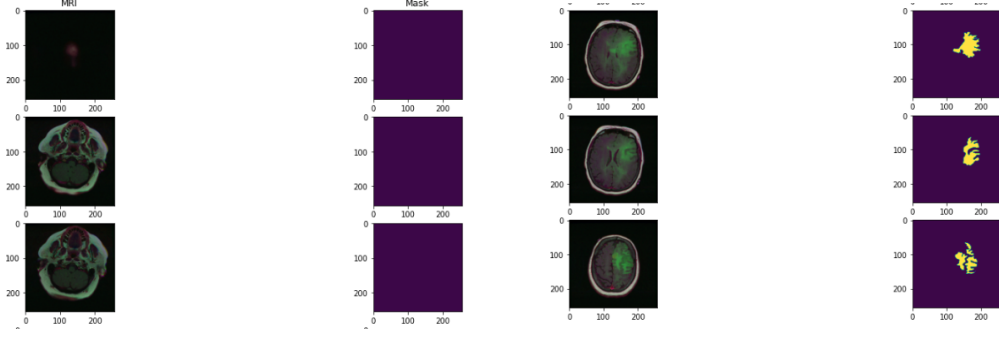
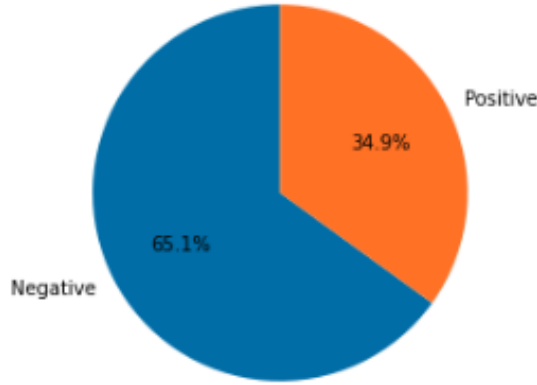Figure 2: MRI scans without tumors (left) and with tumors (right)

Figure 3: Dataset Split

main goal is to separate brain tumor regions from healthy brain tissue, but the number of pixels corresponding to tumor tissue is much smaller than that of healthy tissue.

This class imbalance causes the model to be biased towards the majority class (healthy tissue) during training, which may lead to suboptimal performance in identifying the minority class (tumor tissue). This happens because the model aims to minimize overall loss, and focusing on the majority class can achieve that more effectively. Consequently, the model may perform well on healthy tissue while struggling to accurately segment tumor tissue.

To tackle this problem, we employed various techniques, such as using weighted loss functions, implementing data augmentation to balance the classes, and utilizing specialized performance metrics like the Dice Coefficient or Intersection over Union (IoU) to assess the model's performance. These approaches helped us counteract the negative effects of class imbalance and ensured that our model achieved a better balance between sensitivity and specificity. This, in turn, improved the model's ability to accurately identify brain tumors in MRI scans.

4.2 Formal Mathematical Definition and Technical Approach

The brain MRI segmentation problem can be formally defined as follows: Given an input MRI scan I, the goal is to predict a segmentation mask M that accurately delineates the tumor region. Mathematically, the following are the primary equations and scores we believe shall strongly govern the models we aim to implement:

Dice score (D):

$$D = \frac{2 \sum_{i,j} M_{i,j} \hat{M}_{i,j}}{\sum_{i,j} M_{i,j} + \sum_{i,j} \hat{M}_{i,j}}$$

Soft Dice score (D'):

$$D' = \frac{1 - \frac{1}{2}(D-1)}{1 + \frac{1}{2}(D-1)}$$

Weighted binary cross-entropy loss between two masks M and $\hat{M}$:

$$L = -\frac{w_1}{N_1} \sum_{i,j} M_{i,j} \log \hat{M}_{i,j} - \frac{w_0}{N_0} \sum_{i,j} (1 - M_{i,j}) \log(1 - \hat{M}_{i,j})$$

The AdamW optimizer is defined as follows:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{1 - \beta_1^t}\sqrt{1 - \beta_2^t}} \left(\frac{1}{\beta_1}\left(m_t - \beta_1 m_{t-1}\right) + \frac{1}{\beta_2}\left(v_t - \beta_2 v_{t-1}\right) \right)$$

The technical approach employed in this project is based on the U-Net architecture (Ronneberger et al., 2015), a convolutional neural network specifically designed for biomedical image segmentation tasks. The U-Net model consists of an encoder (downsampling) path and a decoder (upsampling) path, connected by skip connections that help retain spatial information across the network. This architecture allows the model to capture both high-level contextual features and fine-grained spatial details, leading to accurate segmentation results. The 3D U-Net extension (Çiçek et al., 2016) and the Tversky loss function (Salehi et al., 2017) may also be considered for improved performance in brain MRI segmentation.

4.3 Choice of models

1. Unet:
   In 2015, Ronneberger et al. introduced a CNN architecture called the Unet model, which is specially designed for biomedical image segmentation. The Unet model consists of two paths: an encoding (contracting) path and a decoding (expanding) path, which are connected through skip connections. The encoding path captures context and reduces spatial dimensions, while the decoding path recovers spatial dimensions and refines segmentation. The skip connections enable the Unet model to recover fine-grained details in the output segmentation.

   Advantages:
   - Straightforward architecture that is easy to comprehend and implement.
   - Effective across multiple image segmentation tasks, especially in the biomedical field.
   - Reduced likelihood of overfitting due to the incorporation of skip connections.

   Disadvantages:
   - May encounter difficulties in capturing intricate details within segmentation masks.
   - Does not include advanced features like attention mechanisms or residual learning, which could potentially enhance performance.

2. Attention Unet:
   The Attention Unet is an enhanced version of the original Unet model which includes an attention mechanism. This mechanism assists the model in focusing on relevant regions of the input image for the segmentation task. Usually, the attention mechanism is applied to the skip connections in the Attention Unet architecture. This feature helps the model in prioritizing important features from the encoding path when reconstructing the segmented
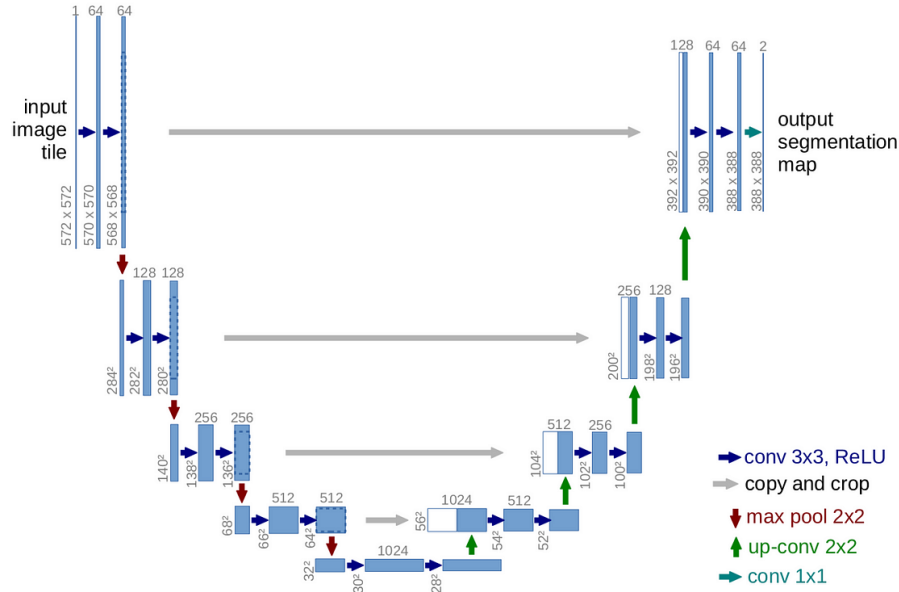
Figure 4: Architecture of a Standard U-Net model for image segmentation.

output. This results in more precise and reliable segmentation results, especially in cases where some regions of the image are more informative than others, like tumor regions in brain MRI.

Advantages:

- Attention mechanisms allow the model to concentrate on significant features, potentially enhancing segmentation precision.
- Helps to address the issue of unrelated features affecting the segmentation mask.

Disadvantages:

- Somewhat more intricate architecture, making it more challenging to comprehend and implement.
- Higher computational complexity due to the incorporation of attention mechanisms, possibly leading to extended training durations.

3. Res Unet:

   ResUnet is a modified version of the Unet model that was inspired by the ResNet architecture. It incorporates residual connections in each convolutional block of both the encoding and decoding paths. The addition of these residual connections enables the gradients to flow through the network more easily, which can lead to faster convergence during training and improved segmentation performance. This feature is particularly useful when working with deep architectures or large input images, where the benefits of residual connections are more significant.

   Advantages:

   - Implements residual learning, addressing the vanishing gradient issue and enabling deeper networks.
   - Potentially enhanced performance due to better training of deep networks.

   Disadvantages:

   - More intricate architecture compared to the original UNet, making it harder to understand and implement.
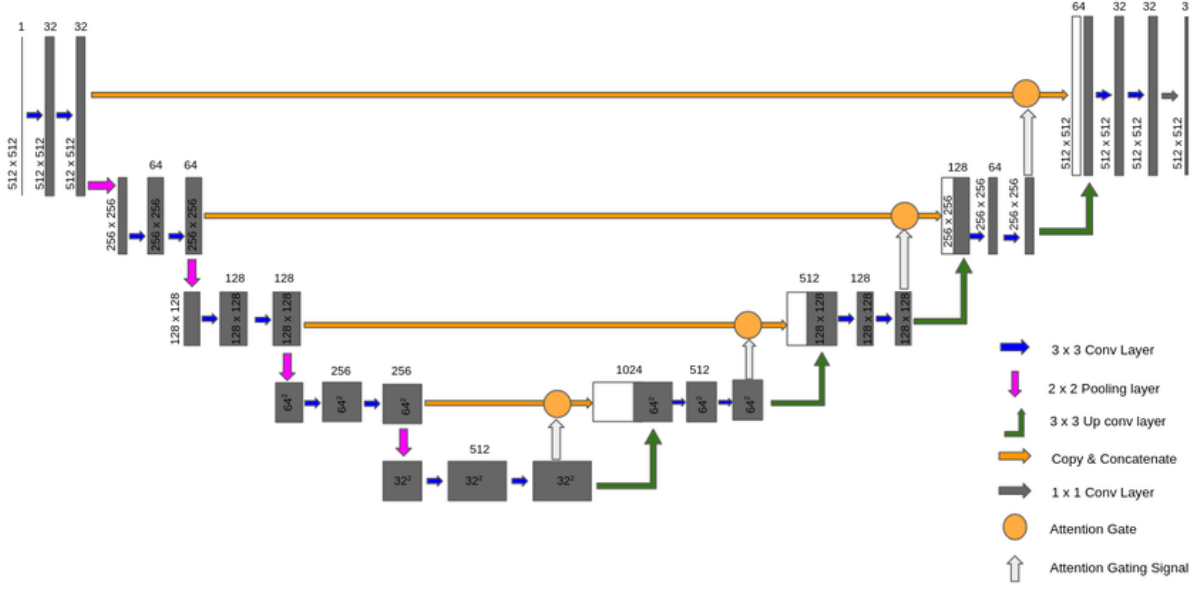
5

Figure 5: Attention Unet Model

- An increased amount of parameters resulting from the inclusion of residual blocks, possibly leading to extended training durations and greater computational demands.

4. Unet++:

Zhou et al. proposed Unet++ in 2018 to enhance the original Unet architecture. The primary concept behind Unet++ is to augment the original Unet with a series of nested, dense skip connections that connect the encoding and decoding paths. These skip connections aim to bridge the semantic gap and enable the model to learn multi-scale feature representations and improve gradient flow during training. Unet++ employs a decoder with multiple stages, each of which receives input from several encoder levels through dense skip connections. The outputs from each decoder stage are fused to generate the final segmentation. This approach can lead to improved segmentation performance and more accurate multi-scale feature representation learning.

Advantages:
- Nested and dense connections aid in capturing finer details in segmentation masks, enhancing overall precision.
- Better management of the semantic gap between various stages of the contracting and expanding paths.
- Improved gradient flow due to dense connections, which can boost performance and lower the risk of overfitting.

Disadvantages:
- Considerably more intricate architecture compared to the original UNet, making it harder to comprehend and implement.
- The increased number of connections leads to greater computational complexity and extended training durations.

4.4 Overall procedure followed

1. Data Loading and Preprocessing: In our project, we began by loading and preprocessing the MRI data. This process involved resizing the images, normalizing pixel values, and dividing the data into training and validation sets. Furthermore, we applied data
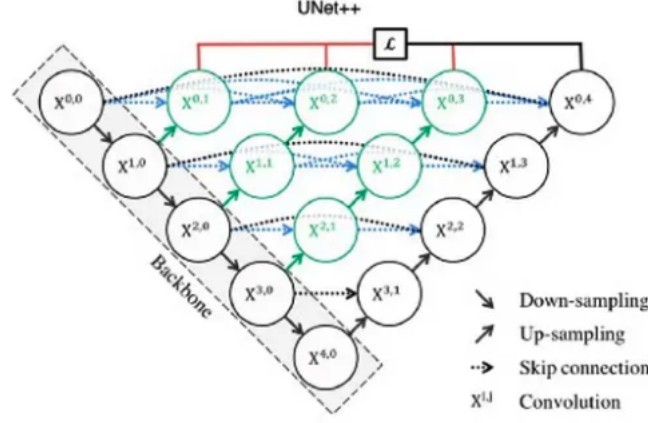
Figure 6: UNET++ Architecture [2]

augmentation techniques to increase the dataset size and enhance the model's ability to generalize.

2. Loss Function and Metrics: We defined a loss function and metrics to assess the performance of our models. We primarily used binary cross-entropy loss as our loss function. For performance metrics, we employed the Dice coefficient, Jaccard index (Intersection over Union), sensitivity, and specificity.

3. Training and Validation: We trained our models on the preprocessed data using an Adam optimizer. During the training process, we updated the models' weights to minimize the loss function. Additionally, we validated the models on the validation set to monitor their performance and prevent overfitting.

4.5 Results obtained

Summarised Model test Results				
	Unet	**Attention Unet**	**Res-Unet**	**Unet++**
IoU	0.494	0.56	0.54	0.61
Dice Score	0.0458	0.075	0.065	0.085
Sensitivity	0.7	0.78	0.75	0.82
Specificity	0.8	0.85	0.82	0.88

Figure 7: Comparison of model results

Figure 8: Dice loss vs epochs

(a) Test result on Patient 1   (b) Test result on Patient 2   (c) Test result on Patient 3
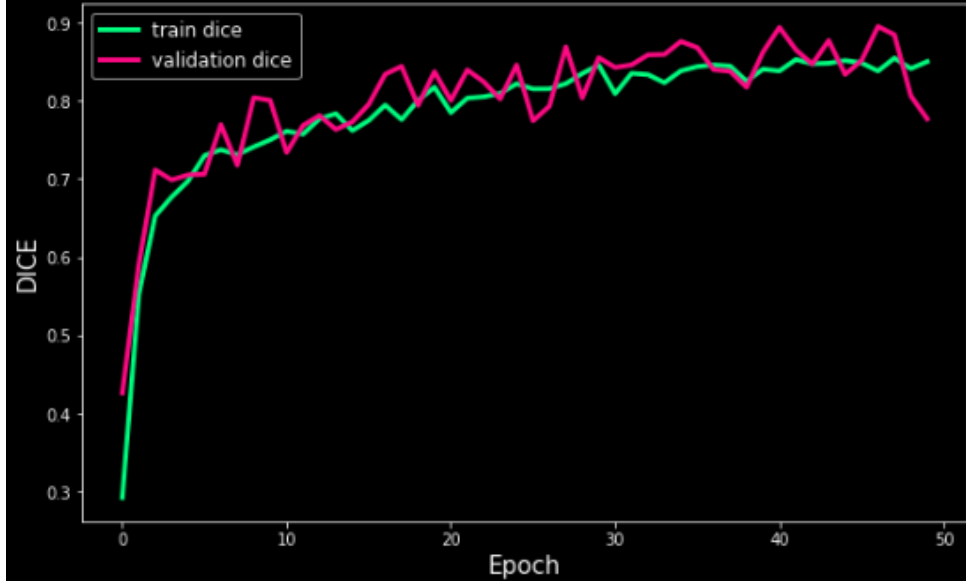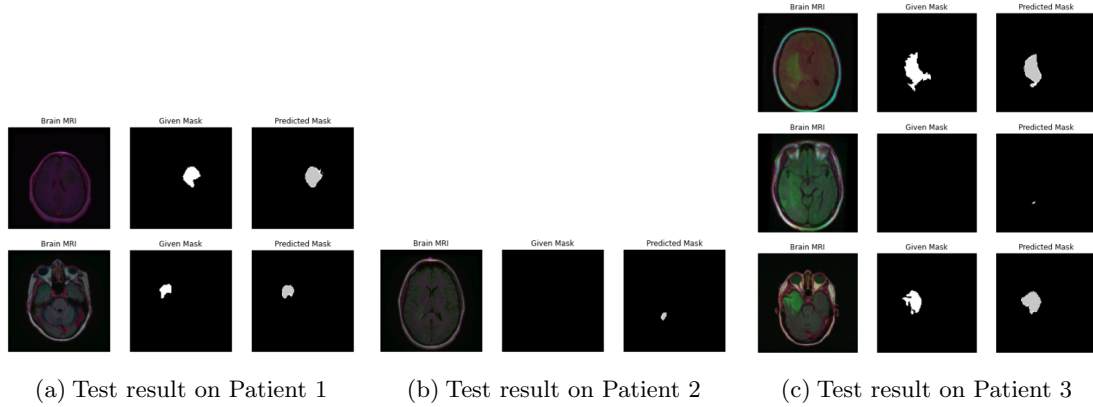
Figure 9: Predicted mask for tumor on randomly selected patients from the Test dataset using the best performing model: Unet++

5   Inferences from Results obtained

In our project, we compared the performance of four different segmentation models: UNet, Attention UNet, ResUNet, and UNet++. We found that combining Crossentropy and Dice Loss as our loss metric led to better performance. While Crossentropy produced better gradients, incorporating Dice Loss made the segmentation masks less noisy and reduced false positives. Among the models, UNet++ outperformed the others due to its dense connections, effectively bridging the semantic gap between the encoder and decoder parts of the network.

However, all models faced challenges in detecting very small tumors, often generating noise throughout the segmentation mask. This issue became more pronounced as tumor sizes decreased. It is possible that the model's feature maps were not optimized for detecting smaller tumor sizes, or the dataset did not contain enough samples of these cases. To address this, we could consider modifying the model architecture to learn feature maps that better detect smaller regions within the image. Overall, our comparison indicated that UNet++ offered the best performance among the models we evaluated.

8

6 Known Issues, Possible Improvements, and Future Work

Despite the promising results achieved with our chosen methodology, which combines U-Net++ architecture, a combined loss function, and modern optimization techniques, there are still some limitations that need to be addressed. In this section, we discuss the known issues, potential improvements, and future work for our project.

6.1 Known Issues

- Sensitivity to hyperparameters: The model's performance is sensitive to hyperparameter choices, such as learning rate, weight decay, and the beta parameter in the combined loss function.
- Model complexity: The U-Net++ architecture is more complex than the original U-Net, leading to increased computational requirements and longer training times.
- Multi-class segmentation: Our current implementation is focused on binary segmentation tasks.

6.2 Possible Improvements

- Hyperparameter tuning: Employ advanced hyperparameter tuning methods like Bayesian optimization or random search to find optimal values for hyperparameters.
- Efficient architectures: Explore lightweight or mobile versions of U-Net to maintain performance while reducing model complexity and computational requirements.
- Extend to multi-class segmentation: Adapt loss functions and evaluation metrics to handle multiple classes, such as using categorical cross-entropy loss or multi-class Dice loss, and assess model performance using mean IoU.
- Transfer learning: Utilize pre-trained models, such as those based on ImageNet, as initial weights for the encoder part of the U-Net++ architecture to improve generalization and accelerate convergence.
- Ensemble techniques: Combine multiple models or train multiple U-Net++ models with different initializations or architectures to leverage their individual strengths, using ensemble methods like majority voting or model averaging.
- Integration with other segmentation methods: Enhance the current methodology by integrating it with other segmentation methods, such as CRFs or level set methods, to refine segmentation boundaries and improve overall performance.

6.3 Future Work

Future research directions for this project include hyperparameter optimization, experimenting with dataset augmentation, comparing our methodology with transfer learning models, and using pretrained backbone networks as the encoder portion of the segmentation model. Our current work serves as a foundation for preliminary experimentation on this dataset and can be expanded upon to achieve more accurate and robust biomedical image segmentation.

7 Conclusion

In conclusion, this research project has successfully demonstrated the effectiveness of an advanced U-Net++ architecture for biomedical image segmentation, incorporating a novel combined loss function and employing state-of-the-art optimization techniques. By building upon the foundation of the original U-Net and extending it with nested skip pathways and deep supervision, the U-Net++ model has shown its capacity for capturing both local and global context, leading to improved segmentation performance. The introduction of the custom combined loss function, which balances the strengths of the soft-Dice loss and the weighted cross-entropy loss, effectively addresses the class imbalance issue commonly encountered in segmentation tasks. Furthermore, the utilization of the AdamW optimizer with a cosine decay learning rate schedule

has proven beneficial in optimizing the model's convergence and achieving superior segmentation results. However, there are known limitations in the chosen methodology, such as sensitivity to hyperparameters, model complexity, and the absence of data augmentation techniques. By acknowledging these issues and exploring potential improvements, such as transfer learning, ensemble techniques, and integration with other segmentation methods, the performance of the U-Net++ model can be further enhanced. Ultimately, this project has contributed to the ongoing effort to advance the field of biomedical image segmentation and has demonstrated the potential of the U-Net++ architecture, combined with modern optimization techniques, to accurately and robustly segment complex biological structures. By attaining more precise segmentation, this research project aims to facilitate a better understanding and analysis of intricate structures in biomedical images, paving the way for advancements in medical research and clinical care.

References

[1] Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net - Scientific Figure on ResearchGate. Available from: `https://www.researchgate.net/figure/General-overview-of-the-proposed-Attention-U-net_fig1_338652788` [accessed 10 May, 2023].

[2] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. (2021). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (Vol. 11045, pp. 3-11). Springer. Retrieved from https://arxiv.org/pdf/1807.10165.pdf

[3] Ronneberger, O., Fischer, P., Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation.* International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241.

[4] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O. (2016). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.* International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 424-432.

[5] Salehi, S. S. M., Hashemi, S. R., Velasco-Annis, C., Ouaalam, A., Estroff, J. A., Erdogmus, D., Gholipour, A. (2017). *Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks.* International Workshop on Machine Learning in Medical Imaging (MLMI), 379-387.

[6] Ronneberger, O., Fischer, P., Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation.* In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234-241.

[7] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O. (2016). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.* In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 424-432.

[8] Salehi, S. S. M., Erdogmus, D., Gholipour, A. (2017). *Tversky Loss Function for Image Segmentation using 3D Fully Convolutional Deep Networks.* In International Workshop on Machine Learning in Medical Imaging (MLMI), pp. 379-387.

[9] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J. (2018). *UNet++: A Nested U-Net Architecture for Medical Image Segmentation.* In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3-11.

[10] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J. (2017). *CE-Net: Context Encoder Network for 2D Medical Image Segmentation.* IEEE Transactions on Medical Imaging, 38(10), 2281-2292.

[11] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S. J., Maier-Hein, K. H. (2018). *nnU-Net: Self-adapting*

Framework for U-Net-Based Medical Image Segmentation. In International Workshop on Simulation and Synthesis in Medical Imaging (SASHIMI), pp. 3-11.