DOI: 10.54254/2755-2721/100/2025.18292

An Enhanced U-Net Model for Segmenting CT Images of COVID-19 Patients

Zhengyao Shen[1,a,*]

[1]Jiangsu University of Science and Technology, No. 2 Mengxi Road, Zhenjiang City, Jiangsu Province, China

a. 232241807324@stu.just.edu.cn

**corresponding author*

Abstract: In the context of the global COVID-19 pandemic, medical imaging technology has played a crucial role in the diagnosis and treatment of the disease. Accurate segmentation of lesion areas in CT images is critical for assessing the condition and formulating treatment plans. This paper first outlines the importance of medical image segmentation. It then delves into the main challenges faced in image segmentation for COVID-19 diagnosis, including the diversity of lesions, inconsistencies in image quality, and the need for real-time processing. Following this discussion, the paper reviews existing medical image segmentation models, encompassing traditional methods such as watershed and threshold segmentation, as well as advanced deep learning models like U-Net and RCNN. Building on this foundation, the paper proposes an improved image segmentation framework aimed at enhancing the accuracy and processing speed of lesion area segmentation. The goal is to provide a more reliable decision support tool for clinical practice.

Keywords: COVID-19, CT image, medical image, segmentation, U-Net.

1. Introduction

In the context of the global COVID-19 pandemic, the disease has posed a severe threat to public health, particularly affecting the lungs [1]. As a highly contagious illness, the early diagnosis and timely treatment of COVID-19 are crucial for reducing mortality rates and mitigating disease progression. Therefore, medical imaging technology plays a key role in the diagnosis and treatment of this disease.

Computed tomography (CT) imaging, due to its high resolution and sensitivity to pulmonary lesions, has become an important tool for diagnosing COVID-19 [2]. CT images are not only used to confirm the presence of viral pneumonia but also to monitor disease progression and treatment efficacy, providing essential information for clinical decision-making. Through CT imaging, physicians can visually assess the extent and severity of lesions, which is vital for developing subsequent treatment plans. However, accurately segmenting lesion areas in CT images is critical for disease assessment and treatment planning. Image segmentation technology, as a core component of medical image analysis, effectively identifies and separates lesion areas, thereby enhancing diagnostic accuracy and treatment effectiveness. Furthermore, image segmentation provides important quantitative analysis tools for clinical research, aiding in early disease detection and long-term monitoring. Despite the strong support that CT imaging offers for diagnosing COVID-19, accurately

DOI: 10.54254/2755-2721/100/2025.18292

segmenting lesion areas from these images remains a technical challenge. Factors impacting segmentation accuracy include the ambiguity of boundaries between lesions and normal tissues, the complexity of lung structures, and individual differences among patients. These challenges render traditional image analysis methods inadequate for meeting the clinical demands for precise segmentation.

Beyond the segmentation of pulmonary lesions in COVID-19, the broader field of medical image segmentation encompasses various applications, such as tumor detection and segmentation, brain structure analysis, and cardiovascular assessment. Each of these applications faces its own challenges, including inconsistencies in image quality, differences across imaging modalities, and the diversity and complexity of lesions. For instance, in tumor segmentation, the shape and size of tumors vary due to individual differences, and they may not have distinct boundaries with surrounding tissues, complicating the segmentation process.

To address these challenges, researchers have developed various advanced image segmentation algorithms and techniques, such as deep learning-based methods, multimodal image fusion technologies, and segmentation models that incorporate clinical information. Notably, deep learning methods like U-Net [3] and the RCNN [4] series have made significant progress in the field of medical image segmentation. The U-Net model, with its encoder-decoder architecture, effectively combines contextual information with precise location data, achieving high segmentation accuracy.

The goal of this study is to propose an enhanced U-Net model to improve the accuracy and processing speed of lesion area segmentation in CT images of COVID-19 patients. We will discuss the design, implementation, and performance evaluation of the proposed model in detail. Through this research, we aim to provide new directions for the development of medical image segmentation technology and to offer a more reliable decision support tool for clinical practice.

2. Previous works

In the field of medical image analysis, segmentation techniques are essential for extracting lesion areas and analyzing image information. With the advancement of medical technology, research on image segmentation has become increasingly in-depth, resulting in a diverse array of methods. Traditional segmentation approaches primarily rely on classical image processing techniques and manual feature extraction. These include threshold-based methods, edge detection methods, region growing techniques, and watershed algorithms.

Threshold-based methods are among the simplest image segmentation techniques [5]. These methods classify image pixels into different categories based on their gray values. The threshold can either be fixed or dynamically calculated using various algorithms. Although fixed threshold methods are straightforward and computationally efficient, their performance often falls short when applied to images with complex gray distributions. This limitation arises from the strong dependency on the chosen threshold, which directly impacts the validity and effectiveness of the segmentation results. Edge detection methods operate by first identifying edge pixels within the image [6]. These pixels are then connected to form region boundaries. Commonly used edge detection operators include first-order differential operators such as Roberts, Prewitt, and Sobel, as well as second-order differential operators like Laplace and Kirsh. These operators detect edges by calculating the gradient of the image's gray values, making them suitable for images with well-defined edges and high contrast. However, the performance of edge detection methods significantly declines in the presence of noise or when edges are blurred. This limitation highlights the need for more robust techniques to effectively address these challenges. Region growing methods are based on the similarity of gray values among adjacent pixels [7]. This technique starts from a predefined seed point and iteratively adds neighboring pixels that meet a specified similarity criterion, ultimately forming the target region. While region growing can adapt to various segmentation tasks, it is highly sensitive to the selection of the initial seed point. Additionally, this method can be adversely affected by noise and uneven gray distributions, which may compromise segmentation quality. The computational cost of region growing is also relatively high, especially in images that require segmentation of large areas. The watershed algorithm, grounded in topological theory and mathematical morphology, conceptualizes the image as a topographic landscape where pixel gray values represent elevation [8]. This method segments the image by simulating the flow of water, filling the "valleys" of the image until it reaches the "ridges." Although the watershed algorithm is effective for segmenting images with pronounced gray differences, its performance deteriorates when dealing with images characterized by blurred boundaries or complex textures. This limitation is particularly pronounced in medical imaging, where overlapping structures and variable contrast can hinder accurate segmentation.

Despite achieving some success in specific scenarios, traditional image segmentation methods face significant limitations when handling complex scenes. These methods often exhibit low computational efficiency in high-resolution images or real-time applications, making them inadequate for practical needs. Furthermore, traditional techniques typically require adjustments for specific datasets, limiting their adaptability to new datasets or changing environments. Consequently, they often struggle to achieve satisfactory segmentation results in complex scenarios characterized by lighting variations, occlusions, and background interference. Additionally, traditional methods often focus on local pixel information while neglecting the overall contextual information of the image. This oversight can lead to segmentation errors in intricate situations. Due to these shortcomings, researchers have increasingly explored deep learning-based image segmentation methods in recent years. These new approaches effectively address the limitations of traditional methods, providing more accurate and reliable segmentation results.

3. Dataset and Preprocessing

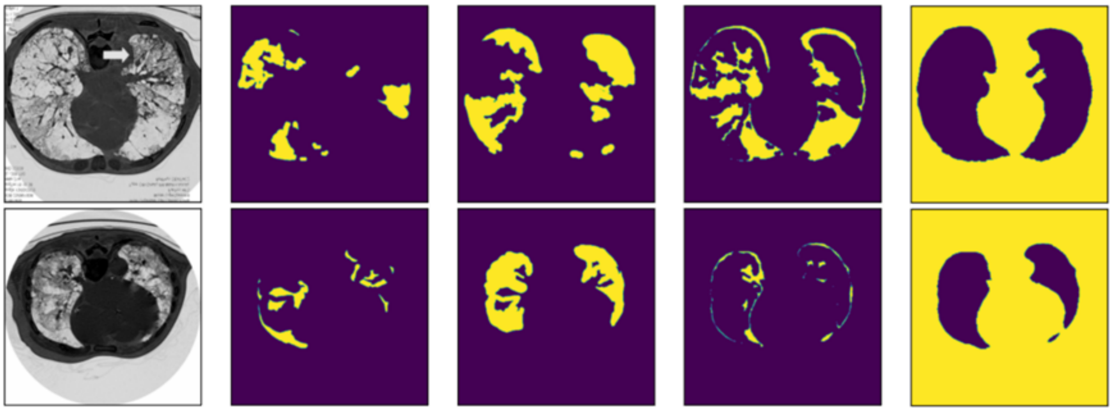

Figure 1: The display of the two original images and manual segmentation results. From left to right are the original image, ground glass, consolidations, lungs other, and background.

This study utilized two specialized medical image segmentation datasets, both provided by medicalsegmentation (Figure 1) [9]. The first dataset was created by two radiologists from Oslo and is sourced from publicly accessible resources. It focuses on the collection and segmentation of CT images, offering valuable resources for the field of medical image analysis. This dataset, named Medseg, consists of 100 axial CT images from over 40 COVID-19 patients. The images are publicly available in JPG format, with a resolution of 512x512 pixels, providing high-quality image resources. Specifically, the dataset includes four categories of lesion regions: “ground glass,” “consolidations,” “lungs other,” and “background.” The second dataset, sourced from Radiopaedia, contains nine axial volume CT images. These images cover the entire volume and include both positive and negative

DOI: 10.54254/2755-2721/100/2025.18292

slices. Out of a total of 829 slices, 373 have been evaluated by radiologists and marked for positivity and segmentation. All images also have a high resolution of 512x512 pixels and include the same four categories of lesion regions: "ground glass," "consolidations," "lungs other," and "background."

During the data preprocessing phase, we implemented several steps to ensure data quality and suitability. First, we plotted a histogram of the pixel value distribution of the image data to observe the distribution before normalization. This step helps to understand the pixel value distribution characteristics, providing a basis for subsequent normalization processes. Next, we performed normalization on the image data, which included truncating pixel values that exceeded a specific range. For instance, pixel values greater than 500 were set to 500, and those less than -1500 were set to -1500. Additionally, we calculated and removed outliers beyond the 5th and 95th percentiles to further reduce noise impact. Subsequently, we computed the mean and standard deviation of the valid pixel values. These statistics were then used to standardize the images by subtracting the mean and dividing by the standard deviation. To enhance model training effectiveness, the dataset was divided into a training set and a validation set, with the validation set comprising 10% of the total data. Finally, we binarized the prediction results, setting pixel values greater than 0.5 to 1, and those equal to or less than 0.5 to 0.

4. Model and Results

This model is based on the classic U-Net architecture and is designed for efficient image segmentation. Its overall structure comprises multiple encoders and decoders, facilitating feature extraction and reconstruction through a symmetrical up-sampling and down-sampling process. After the input layer, the model first goes through four down-sampling stages. In each stage, the input is processed through several convolutional layers that extract multi-level features while maintaining the input dimensions. Activation functions are applied to enhance nonlinear representations. Following each convolution, batch normalization is performed to accelerate training and improve the model's stability. At the end of each down-sampling stage, a pooling layer reduces the size of the feature maps by half, gradually decreasing spatial dimensions while increasing the number of channels, thereby enhancing the abstraction of features. After deep feature extraction, the model enters a bottleneck layer, which contains additional convolutional layers to learn higher-level feature representations and capture more complex information. Next is the up-sampling phase, which follows a similar structure with four stages. During this phase, the feature maps are first enlarged using transposed convolutional layers, and then the up-sampled features are concatenated with the corresponding features from the down-sampling stages. This skip connection mechanism effectively preserves low-level detail information, compensating for spatial information that may have been lost during down-sampling. Each up-sampling step is followed by convolutional processing to further blend the features and enhance model performance. Finally, the model produces the final segmentation results through a 1x1 convolutional layer, where the output channels correspond to the number of classes and facilitate classification. This structure not only makes optimal use of multi-scale features but also enables effective information transfer through skip connections, resulting in excellent performance when addressing complex image segmentation tasks. The final model is applicable to various real-world scenarios, including medical imaging and semantic segmentation, demonstrating superior performance and flexibility.

To enhance the model's performance and generalization capabilities for specific task requirements, we implemented several optimizations to the original U-Net architecture (Figure 2). Firstly, we increased the depth of the convolutional layers to three. This adjustment allows the model to learn more complex feature representations. By adding more convolutional layers, the model can capture finer image details, thereby improving segmentation accuracy. Additionally, a deeper network enhances the model's ability to model nonlinear relationships, which is particularly important for

DOI: 10.54254/2755-2721/100/2025.18292

tackling complex image segmentation tasks. Moreover, to further improve the model's performance, we optimized the structure of U-Net. We added a layer of padding around the feature maps before each convolution, ensuring that the output and input maintain the same spatial dimensions, thus eliminating the need for cropping operations. Secondly, we incorporated residual connections, which help alleviate the vanishing gradient problem. These enhancements enable the model to better recognize and segment lesion areas when faced with complex medical images.

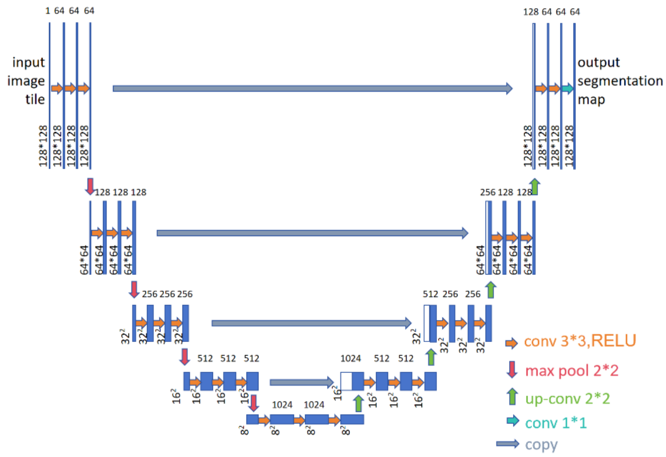

Figure 2: U-Net architecture

During the training process, the dataset was divided into a training set and a validation set, with the validation set accounting for 10% of the total data. In the model construction phase, we built the U-Net model based on the dimensions of the image data. We specified binary cross-entropy as the loss function and Adam as the optimizer, while also including accuracy and Intersection over Union (IoU) as evaluation metrics. During model training, we set the training data, validation data, and training epochs to 10, with a batch size of 1. After training, we evaluated the validation set using the predict method and plotted histograms of the predicted results. Following binarization of the predictions, we also plotted histograms of the processed results. Additionally, we utilized the Matplotlib library to create charts illustrating the changes in accuracy and IoU throughout the training epochs (see Figures 3 and 4). This visual representation intuitively displays the model's learning progress and performance.

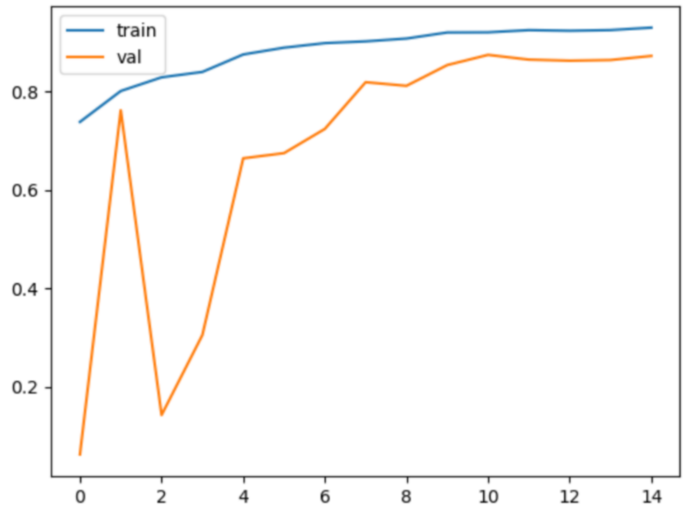

Figure 3: Accuracy curve.

DOI: 10.54254/2755-2721/100/2025.18292

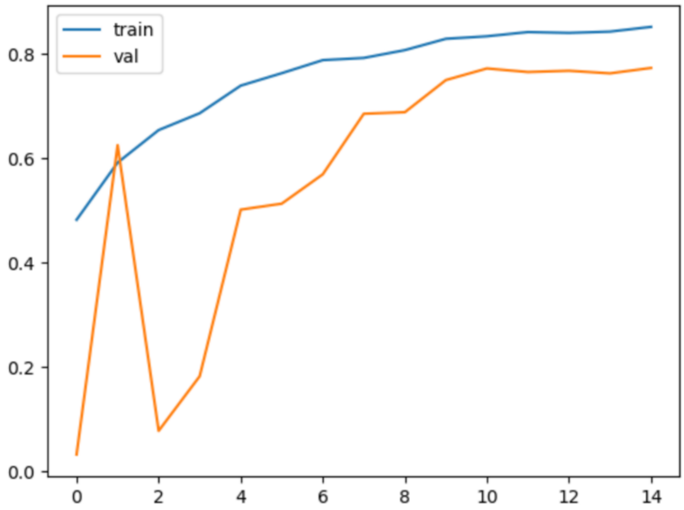

Figure 4: IoU curve.

Our model performed exceptionally well in the segmentation tasks, achieving an accuracy of 84%, significantly surpassing the CNN-based autoencoder model (70%) and the R-CNN model (74%). This result demonstrates the effectiveness of our improvements to the U-Net architecture. The performance boost is primarily attributed to the deeper U-Net design, where increasing the depth of the convolutional layers enabled the model to better capture complex features in the images. At the same time, we optimized the U-Net structure, particularly in the skip connections between the encoder and decoder paths. These optimizations ensured the efficient transfer of spatial information, preserving fine details in the images, which is critical for achieving precise segmentation. In addition, we conducted meticulous hyperparameter tuning, including adjustments to the learning rate, batch size, and activation functions. These refinements significantly enhanced the model's convergence speed and segmentation accuracy. The deeper convolutional layer design also allowed the model to extract features from a greater number of layers, thereby improving its ability to handle complex images and diverse structures.

Despite the good performance, there were still instances where the segmentation results were suboptimal. This limitation is primarily due to insufficient diversity in the training data, which restricted the model's generalization ability when faced with new scenarios. To address this issue, we plan to further expand the dataset and increase its diversity. Additionally, we aim to explore more advanced data augmentation techniques, such as elastic deformations, rotations, and contrast adjustments, to enhance the model's adaptability. Furthermore, refining the model architecture by introducing attention mechanisms or experimenting with hybrid models will be crucial for improving segmentation performance.

5. Conclusion

This study presents improvements to the U-Net model, demonstrating promising results in medical image segmentation tasks. By increasing the depth of the convolutional layers and optimizing the network structure, our model has significantly enhanced its ability to capture complex image features. These improvements enable the model to more accurately segment target areas, providing a more reliable tool for medical image analysis.

However, the limitations of the current work should not be overlooked. First, although the enhanced U-Net model shows improved performance, its computational resource requirements have also increased, potentially leading to longer training times. Additionally, the increase in the number of model parameters raises the risk of overfitting, especially when working with small datasets. Training deeper networks and implementing effective residual connections may complicate the training process, necessitating more precise hyperparameter tuning.

To address these issues, we believe future research can explore multiple directions for improvement. First, incorporating more data augmentation techniques—such as rotation, scaling, flipping, and color transformations—could enhance the model's generalization capabilities [10]. Additionally, adopting a multi-task learning framework could allow simultaneous training of the model on multiple related tasks, thereby improving overall performance. Furthermore, employing ensemble learning methods, such as voting, stacking, or averaging, could effectively increase the stability and accuracy of model predictions. Introducing regularization techniques, such as L1/L2 regularization [11], Dropout [12], or Batch Normalization [13], could also help mitigate overfitting. Looking ahead, we believe that larger models [14] and transfer learning [15] present new opportunities for model development. Larger models, with their robust generalization ability and versatility, can effectively lower the barriers to AI development while enhancing accuracy and adaptability. Transfer learning can leverage knowledge from pre-trained models on large datasets, applying it to specific tasks. This approach not only reduces the need for extensive labeled data but also accelerates the training process. Additionally, employing appropriate data augmentation techniques to increase training data diversity can further improve the model's generalization capabilities.

The improved U-Net model shows significant potential in the diagnosis and treatment of COVID-19. By accurately segmenting areas of lung infection, the model can assist physicians in quickly identifying and assessing conditions, thereby expediting the diagnostic process. Moreover, as the condition evolves, the model can monitor changes in the infected areas, providing critical insights for adjusting treatment plans. This capability enhances the specificity and effectiveness of treatment while potentially reducing the consumption of medical resources. Furthermore, the enhanced U-Net model has broad applications in lesion segmentation for other diseases. For instance, in cancer diagnostics, the model can help physicians identify tumors and their boundaries, supporting surgical planning and radiotherapy. This indicates that the improved U-Net model holds substantial promise in the field of medical image analysis, warranting further research and exploration.

References

[1] *T. P. Velavan and C. G. Meyer, "The COVID‐19 epidemic," Tropical Medicine & International Health, vol. 25, no. 3, p. 278, Feb. 2020, doi: 10.1111/tmi.13383.*

[2] *C. Bao, X. Liu, H. Zhang, Y. Li, and J. Liu, "Coronavirus disease 2019 (COVID-19) CT findings: A systematic review and meta-analysis," Journal of the American College of Radiology, vol. 17, no. 6, pp. 701–709, Jun. 2020, doi: 10.1016/j.jacr.2020.03.006.*

[3] *O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.*

[4] *B. Cheng, Y. Wei, H. Shi, R. Feris, J. Xiong, and T. Huang, "Revisiting RCNN: On awakening the classification power of faster RCNN," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 453–468. Accessed: Oct. 31, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Bowen_Cheng_Revisiting_RCNN_On_ECCV_2018_paper.html*

[5] *S. S. Al-amri, N. V. Kalyankar, and K. S. D, "Image segmentation by using threshold techniques," May 21, 2010, arXiv: arXiv:1005.4020. doi: 10.48550/arXiv.1005.4020.*

[6] *H. G. Kaganami and Z. Beiji, "Region-based segmentation versus edge detection," in 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sep. 2009, pp. 1217–1221. doi: 10.1109/IIH-MSP.2009.13.*

DOI: 10.54254/2755-2721/100/2025.18292

[7] M. Mary Synthuja Jain Preetha, L. Padma Suresh, and M. John Bosco, "Image segmentation using seeded region growing, " in 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), Mar. 2012, pp. 576–583. doi: 10.1109/ICCEET.2012.6203897.

[8] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical image segmentation using K-means clustering and improved watershed algorithm, " in 2006 IEEE Southwest Symposium on Image Analysis and Interpretation, Mar. 2006, pp. 61–65. doi: 10.1109/SSIAI.2006.1633722.

[9] "COVID-19 CT images segmentation." Accessed: Oct. 31, 2024. [Online]. Available: https://kaggle.com/competitions/covid-segmentation

[10] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey, " Nov. 05, 2023, arXiv: arXiv:2204.08610. doi: 10.48550/arXiv.2204.08610.

[11] O. Demir-Kavuk, M. Kamada, T. Akutsu, and E.-W. Knapp, "Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features, " BMC Bioinformatics, vol. 12, no. 1, p. 412, Oct. 2011, doi: 10.1186/1471-2105-12-412.

[12] P. Luo, X. Wang, W. Shao, and Z. Peng, "Towards Understanding Regularization in Batch Normalization, " Apr. 24, 2019, arXiv: arXiv:1809.00846. doi: 10.48550/arXiv.1809.00846.

[13] I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain, " Electronics, vol. 12, no. 14, Art. no. 14, Jan. 2023, doi: 10.3390/electronics12143106.

[14] J. Wang and L. Ke, "LLM-seg: Bridging image segmentation and large language model reasoning, " presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 1765–1774. Accessed: Oct. 31, 2024. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024W/MMFM/html/Wang_LLM-Seg_Bridging_Image_Segmentation_and_Large_Language_Model_Reasoning_CVPRW_2024_paper.html

[15] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols, " IEEE Transactions on Medical Imaging, vol. 34, no. 5, pp. 1018–1030, May 2015, doi: 10.1109/TMI.2014.2366792.