



Enhancing Medical Image Segmentation Based on Loss Functions Integration

Amaal Apojila Oshah¹, Ahmed Elajeli Rgibi¹, and Hesham H. Amin^{*2}

¹ Department of Computer Engineering and Information Technology, Sabratha University, Faculty of Engineering, Sabratha, Libya
Amal.adoude@sabu.edu.ly, ahmed.rgibi@sabu.edu.ly

² Department of Electrical Engineering, Faculty of Engineering, Sohag University, Sohag, Egypt hhamin@eng.sohag.edu.eg

Abstract

Medical image segmentation is an essential field of image analysis that processes and extracts information using state-of-the-art deep learning techniques. However, there are various challenges to overcome. One of these challenges is the class imbalance for medical image datasets in which lesions often occupy a much smaller volume than the background. Thus, deep learning algorithms vary in robustness to class imbalance in medical images. Moreover, most training for standard medical datasets uses loss functions for segmentation based on cross-entropy loss, dice loss, or a combination of both. Selecting an optimal loss function affects the performance of the segmentation results. To address these topics, this research has proposed integrating focal loss into a hierarchical framework to improve these traditional loss functions. The proposed method is evaluated on a medical imaging dataset related to the abdominal cavity, known for its imbalances. A comparative analysis is conducted between the original LeViT-UNet model and its modified version using the new model. Results show that the modified model significantly outperforms the original one. It indicates the potential of focal loss integration as an effective solution for improving segmentation performance in medical imaging.

© 2025 Published by Faculty of Engineering – Sohag University. DOI: 10.21608/sej.2025.357874.1073.

Keywords: Image Segmentation, Focal loss function, deep neural networks, LeViT-UNet model.

1. INTRODUCTION

Image segmentation is the process of partitioning pixels of an image into separate regions corresponding to an object or a class. Specifically, in medical imaging, image segmentation involves separating and detecting organs, lesion areas, tumors, anomalies, pathological issues, and monitoring progressing diseases [1]. Thus, medical image segmentation is a challenging task because small object segmentation is difficult to detect because of its low contrast and strong misleading appearance in the images. The object boundaries for medical images are ambiguous because of the influence of image acquisition [2]. Medical image datasets used for segmentation are created from unimodal or multimodal pictures obtained by advanced medical equipment such as magnetic resonance imaging (MRI), Computed Tomography (CT), and ultrasonography (US). The two used imaging technologies are CT and MRI. However, CT is often a preferred option because of its greater accessibility and cost-effectiveness [3]. In recent years, Artificial Neural Network (ANN) models have shown their importance for most medical image segmentation applications. Specifically, Convolutional Neural Networks (CNNs) have made substantial progress in medical image segmentation beginning with the fully convolutional network (FCN) [4] and its variants (e.g., U-Net[4], SegNet [5], Deep Lab [6], CCNet [7]). They have been applied in cardiac segmentation from MRI [8], liver and tumor segmentation from CT [9], and abnormal lymph node segmentation from PET/CT [10], amongst others. Nevertheless, these models exhibit inherent limitations, as they are built based on the convolutional network architecture that faces several challenges, including their inability to effectively capture long-term dependencies in data, which can limit performance in tasks that require understanding context. Additionally, CNNs can struggle to recognize small objects due to pooling layers that can reduce the size of essential features. They are also prone to overfitting, especially when trained on small datasets, as they can learn

* Corresponding author: hhamin@eng.sohag.edu.eg

noise rather than generalizable patterns. The CNN architecture can be computationally demanding, making them less accessible for real-time applications or deployment in resource-constrained devices. Furthermore, CNNs can exhibit bias toward classes with more abundant training samples, leading to unbalanced performance across classes. Finally, the lack of explainability and understanding of how features are learned poses challenges in applications that require transparency, such as medical diagnosis [11].

For years, the U-Net model has been state-of-the-art in medical image segmentation technology and has become standard in medical image segmentation tasks. However, it remains constrained by restrictions in explicitly modeling long-term dependencies [12]. This is because of the ignorance of global context information at different scales. Also, it cannot use global semantic information interaction, which can be addressed by combining the Transformer and U-net in an advanced model such as LeViT [13]. LeViT is used for fast inference image classification with hybrid transformers and convolutional blocks. However, its architecture has not fully exploited various scales of feature maps from the transformer and convolutional blocks, which facilitate image segmentation. Later, LeViT-UNet [14] for 2D medical image segmentation was proposed to make faster encoding using the transformer technique and to improve the segmentation performance by obtaining multi-scale feature maps (local and global). It was also the first work that studied the speed and accuracy using transformer-based architecture for the medical image segmentation task [15]. Not only does the model make architectural improvements, but also it relies on effective loss functions to achieve optimal performance. On the other hand, in most research related to medical image segmentation, cross-entropy and dice equations are used to compute the loss function [15]. Using the cross-entropy loss function, all pixels in the image are treated equally [16]. However, the network is dominated by classes with a larger number of pixels. Thus, it is difficult for a neural model to recognize the features of small objects [17]. Using dice loss, that directly ignore background areas can result in important loss of information. Therefore, the network segmentation performance for small objects is very poor [17]. Unbalanced data refers to a scenario where the number of samples associated with each category is highly variable. This causes the model to be biased in predicting the majority class, even if identifying the minority class is equally or more important. Focal loss uses the confidence of each sample to generate an energetic weight to maximize the loss value of negative samples (small regions) and minimize the loss value of positive samples (background region) to make sure that the network can optimize the process taking into account the lesion areas [16].

In this research, the effect of the choice of loss function on convergence behavior and segmentation performance is analyzed using focal loss and dice loss that can guide the optimization process toward better feature learning and improved accuracy. The proposed model is a modified LeViT-UNet model adding a focal loss function to the other loss functions that were used in the basic model. The proposed model can provide good results, as the focal loss function is a framework that generalizes dice and entropy-based loss to address imbalance issues, which gives outstanding performance and improves the medical image segmentation. Several performance metrics have been examined to verify the performance of the proposed model. These metrics included accuracy, the Jaccard index, and the dice coefficient to evaluate the segmentation results. The major contributions of this paper can be summarized as follows:

- Exploring the effect of changing loss function types on the performance of the modified LeViT-UNet model and segmentation results.
- Experiments are conducted using two different standard public databases to validate the model modification, and a comparison of the results with other state-of-the-art methods is performed.

The rest of the paper is organized as follows: the related works are reviewed in section Our proposed methodology is presented in section 3. Then, we come to the experimental results and discussion in section 4, followed by the analysis of the obtained results. Finally, the conclusion is given in section 5.

2. RELATED WORK

Neural networks have made major progress, with the latest architectures focusing on achieving efficiency and performance. Loss functions have played a key role in these advancements. The following are important studies investigating loss functions as tools for optimizing model performance. In [18] a generalized dice overlap was introduced to improve class rebalancing, aiming at balanced learning by adjusting the convergence rate of learning errors for each class. However, their approach requires further optimization to handle extreme imbalance cases and to achieve an optimal balance between capturing anatomical variation across classes and effectively managing class imbalance. In [19] a focal dice loss was proposed to address class imbalance for multimodal brain tumor segmentation, using a structural element that gradually shrinks to expand, leading to a coarse-to-fine and incremental learning process without changing the network structure. In [20] a generalized focal loss function based on the Tversky similarity index was proposed, to address the problem of data imbalance, and to achieve recall when training on small structures such as lesions. In [21] a focal dice loss, a loss function with balanced sampling that dynamically focuses on difficult examples to address imbalance, was proposed. It has been tested on 2D and 3D convolutional networks across medical datasets, and it has effectively reduced false positives and mitigated overfitting. In [16] dice loss was improved on weighted soft dice loss, and a successive focal loss and

WSDice loss were proposed, to address the problem of unbalanced sample distribution, and it can extract information in depth in both positive and negative samples. However, this approach has some defects as it introduces a couple of hyperparameters that need to be carefully adjusted during use, and weighted soft dice loss is more inconvenient to implement and does not solve the problem of the dice loss error being sensitive to the calculation of the loss function when it is used to segment small objects. In [22] an evaluation of 12 different loss functions applied to medical image segmentation using a 3D U-Net model was presented. To address the imbalance in labels, the study used oversampling techniques with a focus on foreground regions. The results indicated that the composite loss functions associated with Dice were the most effective choice, providing superior performance for this application. In [23] *clDice* (center-lineDice) was defined to enhance topology preservation up to parity for 2D and 3D binary segmentation. *Soft-clDice* is proposed to give more accurate connectivity information and higher graph similarity. In [24] an accelerated Tversky loss (ATL) function was proposed, which uses the log cosh function to optimize the gradients. The No-new U-Net (nn-Unet) model was adopted as the base model to validate the behavior of the loss functions using standard segmentation performance metrics. It provided faster convergence and better mask generation. In [15] a uniform focus loss is proposed to deal with class imbalance. It is evaluated on five medical imaging datasets. It is compared with six Dice-based or cross-entropy-based loss functions, across 2D, 3D, and multi-class 3D binary segmentation tasks. In [25] a comprehensive evaluation of 25 dedicated semantic segmentation loss functions, organized in a hierarchical format, is performed. They conducted comparative experiments using UNet and TransUNet models on two datasets characterized by natural and medical image segmentation. This research has led to the choice of loss functions is more affected by the data than by the network used. In [26] two new loss functions were introduced: *t-vMF Dice loss*, a compact similarity-based alternative to Dice loss, and *adaptive t-vMF Dice loss*, which adjust the similarity levels for easier and harder classes using cosine similarity. In [27] *T-Loss*, a single-parameter loss function, was introduced. This function learns how to adaptively manage tolerance to label noise during the backpropagation process. It eliminates the need for additional computations such as EM and reduces label noise retention.

3. METHODOLOGY

This study's methodology consists of three main steps: (1) selecting an appropriate model architecture specifically tailored for medical image segmentation, (2) choosing relevant datasets for robust evaluation of the model's segmentation capabilities, and (3) implementing a loss function that addresses class imbalance in medical images, which improves segmentation accuracy by effectively capturing details across different anatomical structures.

3.1. MODEL ARCHITECTURE

In this paper, the LeViT-UNet [14] hybrid model has been chosen as a platform model because of its advantages in combining the UNet with the transformer blocks. This combination allows effective capturing of global context features, using transformers and high-resolution spatial information through the UNet. Furthermore, the model's skip connections contribute to segmentation accuracy by integrating low-level features with global context information, making it suitable for medical image segmentation tasks. The focal loss function is employed to enhance the model's performance because it is designed to address the class imbalance by down-weighting the loss contribution of easy-to-classify examples. Moreover, it focuses on complicated and misclassified examples. This approach helps improve segmentation quality, especially for small or underrepresented structures in medical images.

3.2. Dataset

The two employed datasets in this paper will be described. The first is the Synapse multi-organ segmentation dataset (Synapse) [28]. It comprises 30 abdominal CT scans, which include 3,779 axial contrast-enhanced abdominal clinical CT images. It is divided into 18 cases for training and 12 cases for validation, covering 8 abdominal organs: the aorta, gallbladder, spleen, left kidney, right kidney, liver, pancreas, and stomach. The second is the Automated Cardiac Diagnosis Challenge dataset (ACDC) [28]. It has been collected from 150 patients using cine-MR scanners. It comprises 100 volumes with human annotations and 50 private volumes intended for evaluation. The 100 annotated volumes are divided into 80 training samples and 20 validation samples. In this dataset, there is a considerable amount of overlap between the stomach, large intestine, and small intestine classes, resulting in a task that involves multi-label segmentation. These overlaps create a challenge for class imbalance, as the presence of multiple regions of interest in the same area can lead to an uneven distribution of positive and negative samples. This imbalance makes the segmentation process more complex, as the model may struggle to differentiate between the foreground (target regions) and the background.

3.3. Loss Functions

Loss functions play a critical role in optimizing deep learning models and affect their convergence during training. As shown in Table 1, loss functions are categorized into four major groups based on their specific focus and objectives [26] [29].

Table 1. Segmentation loss function types

Groups of loss function	Type of loss function	Sub type
Region-based Loss	Dice Loss	
	Sensitivity-Specificity Loss	
	Tversky Loss	
	Focal Tversky Loss	
	Log-Cosh Dice Loss	
	Region Mutual loss	
	Robust T-Loss	
	IOU(Jaccard) Loss	
	Lovasz-softmax loss	
	Generalised Wasserstein Dice Loss	
Distribution-based Loss	Cross-Entropy Loss	Binary Cross-Entropy Weighted Cross-Entropy Balanced Cross-Entropy
	Focal Loss	
	Distance map derived Loss	
Boundary-based Loss	Topk Loss	
	Hausdorff Distance Loss	
	Shape aware loss	
	Active Boundary	
	Boundary-aware Loss	
	Conditional Boundary Loss	
	InverseForm Loss	
	Boundary Loss	
	Boundary Difference Over Union Loss	
	Combo Loss	Cross-Entropy Dice
Compounded Loss	Exponential Logarithmic Loss	Cross-Entropy Dice
	Unified Focal Loss	Focal Focal Tversky

In imbalanced datasets, positive samples contribute much more gradients than negative samples. Therefore, the optimization process can reduce the gradient component of the dominant class after multiple training iterations because of the loss functions effect. The following are the most important loss functions that affect the model performance.

1-Cross-Entropy: Cross-entropy [28] is defined as a measure of the difference between two probability distributions for a given random variable or set of events. It is widely used for classification objective, and as segmentation is pixel level classification it works well. Cross-Entropy is defined as:

$$LBCE(y, y') = -(y \log(y') + (1 - y) \log(1 - y'))$$

Here, y' is the predicted value by the prediction model

2- Dice Loss: is widely used in medical image segmentation for determining similarity. It is typically applied to determine how identical or overlapped two samples are. Its possible values are 0 to 1. The segmentation impact improves as the value approaches one [14].

$$Dice(A, B) = 2 \frac{|A \cap B|}{|A + B|}$$

3- Focal loss: is a form of binary cross-entropy loss that addresses the class imbalance problem with standard cross-entropy loss by reducing the contribution of positive samples [25].

$$FL(pt) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

where, $\gamma > 0$ and when $\gamma = 1$ Focal Loss works like Cross-Entropy loss function, and α ranges from [0,1] that can be treated as a hyperparameter.

4- Proposed Combined Loss Function:

To leverage the strengths of both Dice Loss and Focal Loss, we propose a weighted combination:

$$\text{Loss total} = \alpha \text{Dice} + \beta \text{Focal}$$

where α and β are hyperparameters that balance the contributions of each loss function. Empirical experiments showed that setting $\alpha = 0.5$ and $\beta = 0.5$ provides an effective trade-off between segmentation accuracy and robustness to class imbalance. Compared to using Dice or Focal Loss individually, this combination demonstrated superior performance, particularly in highly imbalanced datasets.

4. EXPERIMENTAL RESULT AND DISCUSSION

4.1. Implementation Details:

The experiments are conducted using Python 3.7.10, PyTorch 1.9.1, and Linux 5.15.154+-x86_64. The optimizer used is Adam, who has a learning rate of $2e-3$. All models are trained on a Tesla P100-PCIE GPU with 16GB memory. The input resolution of images is 224x224, with a batch size of 8 for training and 16 for validation. The models used transformer backbones pre-trained on ImageNet-1k, and training is conducted for 30 epochs on the Synapse and ACDC datasets.

4.2. Results:

To emphasize the pivotal role of loss functions in enhancing medical image segmentation tasks, the performance of the proposed model was evaluated on two benchmark datasets: the Synapse and ACDC Datasets. The evaluation of model accuracy was conducted using two primary metrics: the Jaccard Index and the dice coefficient, which serve as key indicators of segmentation precision. Two scenarios were assessed, one employing the focal loss function and the other operating without it. The results of this comparison, summarized in Table 2, present the segmentation accuracy and model quality in the context of imbalanced datasets.

Table 2. Performance Metrics of LiVeT-Unet modified model.

Metric	LiVeT-UNet model	modified LiVeT-UNet model
Best of Epoch	29	23
Best Dice	0.7853	0.79954
Best Jaccard	0.7310	0.73653
Training Loss	0.0705	0.06794
Validation Loss	0.02467	0.14535
Validation Dice	0.78214	0.77675
Validation Jaccard	0.71328	0.71341
Trainable Parameters (Million)	52,17	52,156,519

The relationship between the training loss and the validation loss is shown in Fig. 1, and Fig. 2 which illustrate both validation Dice and validation Jaccard.

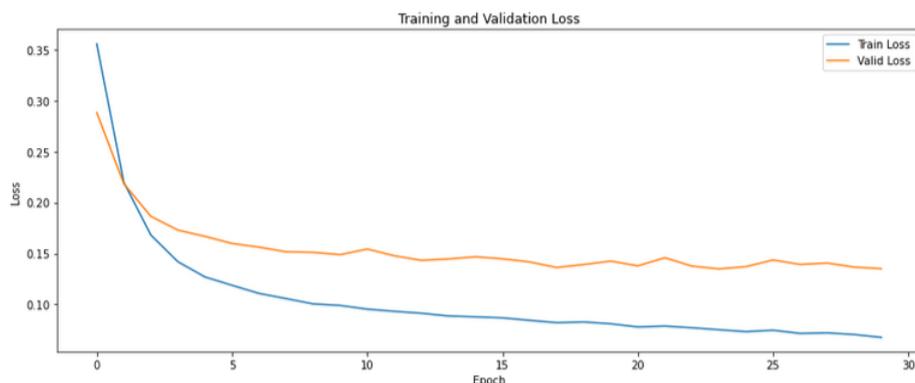


Fig. 1. Training and validation loss

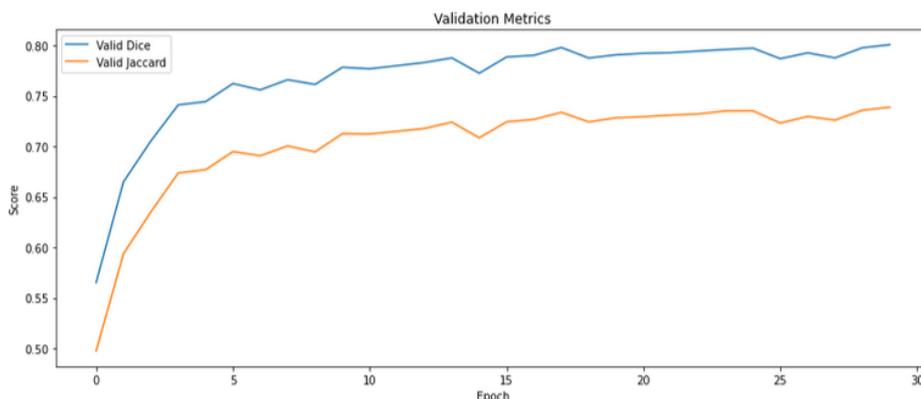


Fig. 2. Validation Metrics

TABLE 3 COMPARE THE MODIFIED LiVET-UNET WITH FAST SEGMENTATION METHODS.

Model Name	Trainable Parameters (Million)	Inference Time FLOPs (Million)	Dice Coefficient	Jaccard Index	Dataset	Memory Usage
Modified LiVet-UNet	52,156,519	150	0.79954	0.7365	Synapse, ACDC	198.96 MB
Swin-Net	61.98	394.84	0.81	0.76	Synapse	29.87 GB
TransUNet	105	1186.9	0.77	0.71	Synapse	-----
Med Transformer	70	290	0.85	0.79	Synapse, ACDC	24.66 GB
UNETR	92.5	41.1	0.89	0.78	Synapse, ACDC	19.52 GB
nnUnet	87.98	8712.56	0.91	0.84	ACDC	15.06 GB

4.3. Discussion

Integrating Focal Loss demonstrates measurable improvements in LiVeT-UNet's segmentation performance as shown in Table 2. The Dice Coefficient improved to 0.79954, and the Jaccard Index increased to 0.73653, showing enhanced precision in handling challenging segmentation tasks. The model achieved its optimal performance within 23 epochs, highlighting faster convergence and reduced training time. While training loss decreased (0.06794 vs. 0.0705), validation loss increased (0.14535 vs. 0.02467), suggesting Focal Loss prioritizes complex samples, potentially at the expense of overall loss stability, but Validation performance remained stable (Dice: 0.77675, Jaccard: 0.71341), showing consistent generalization despite focusing more on challenging cases. Thus, the Focal Loss effectively enhances accuracy, benefiting datasets with class imbalance, while maintaining robust validation performance. The training curve in Fig. 1 shows that training loss decreases steadily over the epochs, reflecting the model's ability to learn patterns from the training data. While the validation loss stabilizes or increases slightly after several epochs, showing potential overfitting. However, mitigation strategies like data augmentation, regularization (as dropout), or Early Stopping can improve generalization. Fig. 2 highlights the model's powerful performance, achieving a Dice score of 0.79954 and a Jaccard index of 0.73653 on imbalanced data. Table 3 illustrates that, while some other models achieve higher performance, in terms of accuracy, the

proposed model balances accuracy and efficiency, making it an ideal choice for real-world applications. The Swin-Net model achieved higher performance in accuracy (Dice: 0.81, Jaccard: 0.76), but suffers from a very high computational cost (394.84G FLOPs) and large memory consumption (29.87 GB), making it less efficient in resource-constrained environments compared to LiVeT-UNet (150G FLOPs, 198.96 MB). But TransUNet shows lower performance than LiVeT-UNet (Dice: 0.77, Jaccard: 0.71) with higher computational complexity (1186.9G FLOPs), making it less practical in systems that require high efficiency. The MedTransformer outperforms in accuracy (Dice: 0.85, Jaccard: 0.79) but consumes almost twice as much resources (290G FLOPs, 24.66 GB) as LiVeT-UNet, which limits its usability in resource-constrained systems. The UNETR is Achieve substantial accuracy (Dice: 0.89, Jaccard: 0.78) with lower computational cost (41.1G FLOPs) but consumes more memory (19.52 GB), which limits its efficiency compared to the lightness of LiVeT-UNet. Finally, the nnU-Net delivers exceptional performance (Dice: 0.91, Jaccard: 0.84) but requires tremendous resources (8712.56G FLOPs, 15.06 GB), making it unsuitable for practical environments and is limited to high-performance systems.

5. CONCLUSION

This research enhances the LiVeT-UNet segmentation model by integrating focal and dice loss functions, effectively addressing sample imbalance in medical image segmentation. Unlike traditional dice loss, which primarily focuses on positive samples and overlooks negative ones, the proposed approach generates adaptive weights from the labels, ensuring that both positive and negative samples contribute to the loss calculation. This integration retains the robustness of dice loss while enhancing the model's ability to learn from underrepresented regions. Experimental results demonstrate that combining focal and dice loss functions improves segmentation performance, particularly in highly imbalanced datasets. While the proposed approach may not achieve the highest accuracy compared with other models, it offers notable computational efficiency and optimized resource utilization, making it a practical choice for real-world medical imaging applications, especially in resource-constrained environments. Future research could explore further optimizations, such as dynamically adjusting the contribution of the dice loss and the focal loss during training using adaptive weighting strategies or reinforcement learning. Moreover, incorporating self-supervised learning techniques, such as contrastive learning, could enhance feature extraction and model generalization. Evaluating the approach to larger medical imaging datasets would further validate their robustness and clinical applicability. Finally, an extensive hyperparameter search for focal loss could provide deeper insights into their impact on different segmentation tasks, improving model stability and performance.

References

- [1] L. Dao and N. Q. Ly, "A Comprehensive Study on Medical Image Segmentation using Deep Neural Networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 167–184, 2023, doi: 10.14569/IJACSA.2023.0140319.
- [2] F. Tang *et al.*, "DuAT: Dual-Aggregation Transformer Network for Medical Image Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14429 LNCS, pp. 343–356, 2024, doi: 10.1007/978-981-99-8469-5_27.
- [3] J. Zhuang, J. Yang, L. Gu, and N. Dvornek, "Shelfnet for fast semantic segmentation," *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, pp. 847–856, 2019, doi: 10.1109/ICCVW.2019.00113.
- [4] N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, "Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference Munich, Germany, October 5-9, 2015 proceedings, part III," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, no. Cvd, pp. 12–20, 2015, doi: 10.1007/978-3-319-24574-4.
- [5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017, [Online]. Available: <https://arxiv.org/pdf/1511.00561.pdf>
- [6] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018, doi: 10.1109/TPAMI.2017.2699184.
- [7] Z. Huang *et al.*, "CCNet: Criss-Cross Attention for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 6896–6908, 2023, doi: 10.1109/TPAMI.2020.3007032.
- [8] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans," *Front. Bioeng. Biotechnol.*, vol. 8, no. December, pp. 1–15, 2020, doi: 10.3389/fbioe.2020.605132.
- [9] G. Xu, J. K. Udupa, Y. Tong, D. Odhner, H. Cao, and D. A. Torigian, "AAR-LN-DQ: Automatic anatomy recognition based disease quantification in thoracic lymph node zones via FDG PET/CT images without Nodal Delineation," *Med. Phys.*, vol. 47, no. 8, pp. 3467–3484, 2020, doi: 10.1002/mp.14240.
- [10] C. Li, L. Wang, and Y. Li, "Transformer and group parallel axial attention co-encoder for medical image segmentation," *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.1038/s41598-022-20440-z.
- [11] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: Advances, challenges and applications," *BMC Bioinformatics*, vol. 20, no. Suppl 11, 2019, doi: 10.1186/s12859-019-2823-4.

- [12] R. Ramadan and S. Aly, "CU-Net: A New Improved Multi-Input Color U-Net Model for Skin Lesion Semantic Segmentation," *IEEE Access*, vol. 10, pp. 15539–15564, 2022, doi: 10.1109/ACCESS.2022.3148402.
- [13] B. Graham *et al.*, "LeViT: a Vision Transformer in ConvNet's Clothing for Faster Inference," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 12239–12249, 2021, doi: 10.1109/ICCV48922.2021.01204.
- [14] [14] G. Xu, X. Zhang, X. He, and X. Wu, "LeViT-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14432 LNCS, no. December 2023, pp. 42–53, 2024, doi: 10.1007/978-981-99-8543-2_4.
- [15] M. Yeung, E. Sala, C. B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, no. November 2021, 2022, doi: 10.1016/j.compmedimag.2021.102026.
- [16] L. Wang, C. Wang, Z. Sun, and S. Chen, "An improved dice loss for pneumothorax segmentation by mining the information of negative areas," *IEEE Access*, vol. 8, pp. 167939–167949, 2020, doi: 10.1109/ACCESS.2020.3020475.
- [17] A. Hatamizadeh *et al.*, "UNETR: Transformers for 3D Medical Image Segmentation," *Proc. - 2022 IEEE/CVF Winter Conf. Appl. Comput. Vision, WACV 2022*, pp. 1748–1758, 2022, doi: 10.1109/WACV51458.2022.00181.
- [18] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10553 LNCS, no. 2017, pp. 240–248, 2017, doi: 10.1007/978-3-319-67558-9_28.
- [19] M. Cheng, A. Galimzianova, Ž. Lesjak, Ž. Špiclin, C. B. Lock, and D. L. Rubin, *A multi-scale multiple sclerosis lesion change detection in a multi-sequence mri*, vol. 11045 LNCS. 2018. doi: 10.1007/978-3-030-00889-5_40.
- [20] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2019-April, pp. 683–687, 2019, doi: 10.1109/ISBI.2019.8759329.
- [21] R. Zhao *et al.*, "Rethinking dice loss for medical image segmentation," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2020-Novem, no. Icdm, pp. 851–860, 2020, doi: 10.1109/ICDM50108.2020.00094.
- [22] J. Ma *et al.*, "Loss odyssey in medical image segmentation," *Med. Image Anal.*, vol. 71, 2021, doi: 10.1016/j.media.2021.102035.
- [23] S. Shit *et al.*, "clDice-a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation."
- [24] N. Nasalwai, N. S. Pun, S. K. Sonbhadra, and S. Agarwal, "Addressing the Class Imbalance Problem in Medical Image Segmentation via Accelerated Tversky Loss Function," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 390–402. doi: 10.1007/978-3-030-75768-7_31.
- [25] R. Azad *et al.*, "Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook," pp. 1–24, 2023, [Online]. Available: <http://arxiv.org/abs/2312.05391>
- [26] S. Kato and K. Hotta, "Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation," *Comput. Biol. Med.*, vol. 168, Jan. 2024, doi: 10.1016/j.compbiomed.2023.107695.
- [27] A. Gonzalez-Jimenez *et al.*, "Hyperbolic Metric Learning for Visual Outlier Detection," pp. 1–22, 2024, [Online]. Available: <http://arxiv.org/abs/2403.15260>
- [28] S. Jadon, "A survey of loss functions for semantic segmentation," *2020 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2020*, 2020, doi: 10.1109/CIBCB48159.2020.9277638.