The Influence of Multiple Loss Functions on MRI Stroke Lesion Area Segmentation

Ruihui Cao

Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo NSW, 2007, Australia
email:Ruihui.Cao@student.uts.edu.au

Abstract. The study solved the imbalanced Magnetic resonance imaging (MRI) dataset problem by choosing different loss functions to achieve a higher stroke lesion area segmentation accuracy. It is helpful for doctors to treat patients efficiently by segmenting the stroke areas quickly with the machine learning model. The study compared focal loss and dice loss based on the same dataset, keeping the model structure and parameters the same, using different evaluation metrics to evaluate the performance of the two loss functions. Then, the study also demonstrated some sample segmentation images to specify the segmentation in detail, helping to understand the result. The study found that the model using focal loss had a better segmentation performance on the imbalanced dataset than the model using dice loss. It was noticed that the focal loss model had clearer boundaries and more precise segmented lesion areas than the dice loss model. That means the focal loss is more suitable for doing the segmentation with small pixels than the dice loss, which would be more useful in medical images, as most of the medical images contain small positive areas and large negative areas. The study could be supportive evidence for future research by providing a strong reason for choosing focal loss as the loss function when training medical image models.

Keywords: Machine Learning, Loss Function, Stroke Lesion Area Segmentation

1 Introduction

Stroke has a significant impact on human health. It is listed as a global primary health issue by the World Health Organization (WHO). Reducing the number of stroke-related deaths will increase the global health level [1]. In 2019, there were about 12.2 million new stroke cases and 6.55 million deaths from stroke only in 2019 [2]. It is essential to be diagnosed and receive treatment as soon as possible to reduce the chance of permanent brain damage. Magnetic resonance imaging (MRI) is an effective and primary way to guide acute treatments due to its high resolution and sensitivity to the contrast of brain tissues [3]. However, traditional manual segmentation methods are time-consuming and susceptible to operator experience [4]. Therefore, utilizing artificial intelligence (AI) to support analyzing MRI images helps doctors gain more time for treatment and

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_88

reduces the workload of doctors due to their excellent performance in many tasks [5-10].

In the last decade, numerous studies have demonstrated the effectiveness of machine learning methods [11-13]. For instance, Soleimani et al. [4] show that by using a 3D U-Net to directly process 3D data input, the model performs with a distinguishing accuracy in segmentation. The F1 Score has reached 0.94226, and the IOU reached 0.88342, which shows the model's accuracy and reliability in stroke lesion segmentation. This study has shown that using U-Net is an effective approach to delineating brain stroke lesions in MRI images. Another study conducted by Hui et al. [14] described that by using an enhanced U-Net model equipped with a dual-path attention compensation mechanism, the enhanced U-Net model has better performance and accuracy when lesions boundaries are unclear, or lesions are small. This result also indicates that using deep learning models such as U-Net is a useful approach to help in MRI stroke lesions segmentation medical practices. However, the dataset used by them, such as the dataset Sub-Acute Ischemic Stroke Lesion Segmentation (SISS), has a significant problem of imbalance. Specifically, most MRIs have way smaller stroke lesion areas than the whole images. When the lesion areas are labeled as 1 and other parts are labeled as 0, it is easy to make the datasets have way more 0s than 1s.

More relevant studies described that their datasets are also imbalanced. The study by Kim et al. [15] used a training dataset consisting of 303 acute ischemic stroke (AIS) cases, which is imbalanced. Another project by An et al. [16] reanalyzed data from previous stroke studies involving 331 C57/BL6 J mice with detectable lesions on T2w images, which is imbalanced as well.

The above studies clearly indicate that it is common to have imbalanced datasets to process while training automated stroke lesion segmentation models. However, few studies have focused on this problem, which may have significant impacts on the models' performance. Therefore, this study aims to solve this problem by finding a suitable loss function to let the model pay more attention to the important data (label 1 area) on the dataset the project uses and improve the segmentation performance in the end.

This study uses the imbalanced Anatomical Tracings of Lesions After Stroke (ATLAS) 2.0 dataset, processes the data from the dataset to fit the input size of a U-Net built, compares different loss functions and uses different evaluation functions to evaluate them. This paper aims to find a loss function with the best segmentation performance among the loss functions tested to train the model based on the imbalanced dataset ATLAS 2.0.

2 Method

2.1 Dataset Preparation

The study used the dataset called ATLAS R2.0 from the University of Southern California, Stevens Neuroimaging and Informatics Institute [17]. The dataset has 955 T1 weighted (T1w) MRI images from different centers. The dataset contains 655 training images with labeled lesion masks and 300 test images without masks. It involves annotations of ischemic and chronic stroke lesions. The size of the images is 197x233x189.

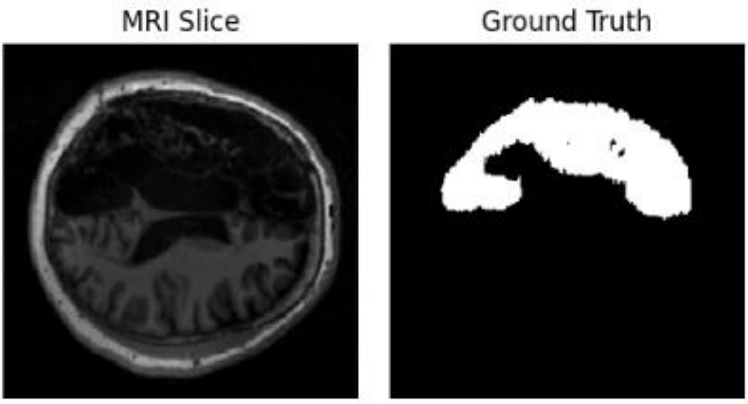

Fig. 1. Sample MRI slice and labeled ground truth from the dataset ATLAS R2.0.

Fig. 1 shows an MRI image with stroke lesions. The stroke lesion areas usually appear to be darker gray areas in the MRI images. It can be observed in the image that the labeled lesion areas are relatively smaller than the normal areas. This commonly happens on all the images in this dataset, which makes the dataset imbalanced.

The study also preprocesses the data from the dataset. It uses data normalization, data augmentation and data resizing to make sure the data inputted into the model can make the model learn efficiently. In terms of data normalization, the study scales the voxel values to the range of 0 to 1 to improve convergence in learning. Data augmentation, specifically in this study, filters out slices that contain at least 100 pixels of stroke lesions to help address data imbalance issues, making the model able to learn more useful information. The study uses data resizing to map 3D MRI images into 2D space. The dimensions are resized to 256×256.

2.2 U-Net Model

Due to the cost and time of acquiring and labeling medical images, medical datasets commonly have only a limited amount of labeled data. In order to achieve optimal performance, a CNN model often requires a substantial volume of training data. The U-Net is a CNN model created by Ronneberger et al. that demonstrates strong performance in segmenting biological images, even when trained with a small dataset [18]. It can accurately segment lesion areas with relatively less data.

The reason why this model is called U-Net is that the architecture of the model is a U shape. The U-Net has a symmetric structure, with contracting (encoder), latent spaces and expanding (decoder) paths. The encoder captures contextual information by downsampling, and the decoder enables precise localization with transposed convolutions by upsampling. Between the contracting and expanding paths, there is a latent space. The latent space is in the deepest part of the U-Net model, which allows it to capture the high-level, global, and abstract representation of the input image. This model is suitable for learning through a limited number of images while performing well in capturing contextual and detailed information.

The model in this study uses a series of downsampling blocks and upsamling blocks. Each downsampling block receives the input and processes it through the convolutional layers. The main part of the block is the 2D convolutional layers. Each 2D convolution layer has 32 kernels, each with a size of 3x3. Then, it is followed by a batch normalization to help avoid overfitting, improve gradient flow, and speed up training. In the end, there is a Rectified Linear Unit (ReLU) to introduce non-linearity, making the model able to learn complex patterns, address the vanishing gradient problem and increase computational efficiency. Each downsampling block contains two rounds of the processes above. After that, the block uses a 2x2 max pooling with a stride of 2 to reduce the dimension and suppress the noise at the end of the block. Each upsampling block adds an upsampling layer with an upsampling size of 2 to recover the spatial resolution that was reduced during the max pooling and a concatenation layer to concatenate with skip connection from corresponding encoder feature maps. The concatenation layer plays an important role in helping preserve spatial hierarchies and fine-grained details for accurate pixel-wise prediction, which is necessary for having a precise boundary when doing segmentation.

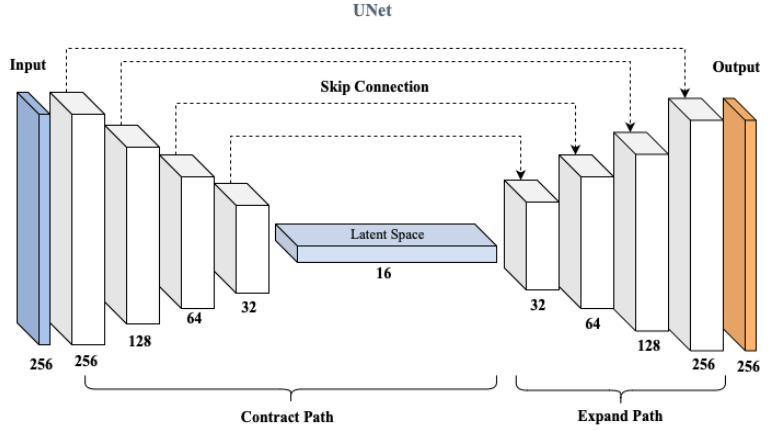

Fig. 2. The U-Net Architecture.

The U-Net model shown in Fig. 2 is built with 4 downsampling blocks, 2 latent space bridges, 4 upsampling blocks and 3 output blocks. The 3 output blocks contain 2 pre-output blocks and an output block. The pre-output blocks are used to further extract features, and the output block is to combine multiple feature channels into a single output channel. The model also makes the filter number able to be customized by providing a parameter called kernel base, which allows the model to change the width easily based on the need.

2.3 Implementation Details

The study uses the U-Net model above with Adam optimizer for updating model parameters, with a default learning rate set to 0.001, built with TensorFlow python library. The Adam optimizer is an adaptive optimization algorithm for learning rates that

assigns distinct rates to individual parameter dimensions, hence enhancing the speed of convergence.

The study also chooses a batch size of 128. This batch size balances the hardware limitation and the efficiency of the model training. The size of 128 could not only improve memory utilization and training speed but also ensure the ability to generalize. The training dataset is set to be 75% for training and 25% for validation. This could help prevent the model from overfitting. The training epochs are set to be 60 with early stopping. The study sets the early stopping to have a patience of 3, monitoring the validation loss. If the validation loss has no improvement in 3 epochs, the model will stop training and restore the best weight. The reason to set the epochs to 60 is that with early stopping, the model usually stops training before its 40^{th} epoch. There is no need to have more epochs as the training processes seldom reach there.

The study aims to find out how the same U-Net model performs on the same dataset with two different loss functions: Dice Loss and Focal Loss. Both of these loss functions are developed to deal with imbalance datasets and optimized for segmentation, so they are suitable for this study. Dice Loss is defined as:

$$\text{Dice Loss} = 1 - \frac{2|AB|}{|A| + |B|} \tag{1}$$

A represents the expected pixels, while B represents the ground truth pixels. |A∩B| represents the count of pixels that are common to both the predicted image and the ground truth image. |A| and |B| denote the aggregate number of pixels in the anticipated region and the actual region, respectively.

Focal Loss is defined as:

$$\text{Binary Focal Loss}(p, g) = \begin{cases} \sum_{i=1}^{N_f} \alpha(1-p)^{\gamma}, & \text{if} \quad g = 1 \\ \sum_{i=1}^{N_b} (1-\alpha)p^{\gamma}, & \text{if} \quad g = 0 \end{cases} \tag{2}$$

Where $g \in 0, 1$ is the ground truth, $p \in [0, 1]$ is the model's prediction probability, N_f and N_b are the number of pixels for class 0 and class 1, and $\alpha \in [0,1]$ and γ are modulation factors that can be adjusted.

The study uses a series of evaluation metrics to evaluate the performance of the models. The Coefficient of Dice Similarity (DSC) is the primary one. The intersection of the expected and actual segmentation is a statistic used to assess a model's anticipated performance, and it is referred to as the F1 score. By definition, the DSC is:

$$\text{DSC} = \frac{2|AB|}{|A| + |B|} \tag{3}$$

The A and B definitions are the same as the dice loss definition.

The study also uses recall, precision, Area Under Curve (AUC) and Intersection Over Union (IOU). These evaluation metrics are used to assist in adjusting the model parameters.

3 Results and Discussion

The study trained two models with the same training parameters to ensure the comparison of the results was reasonable.

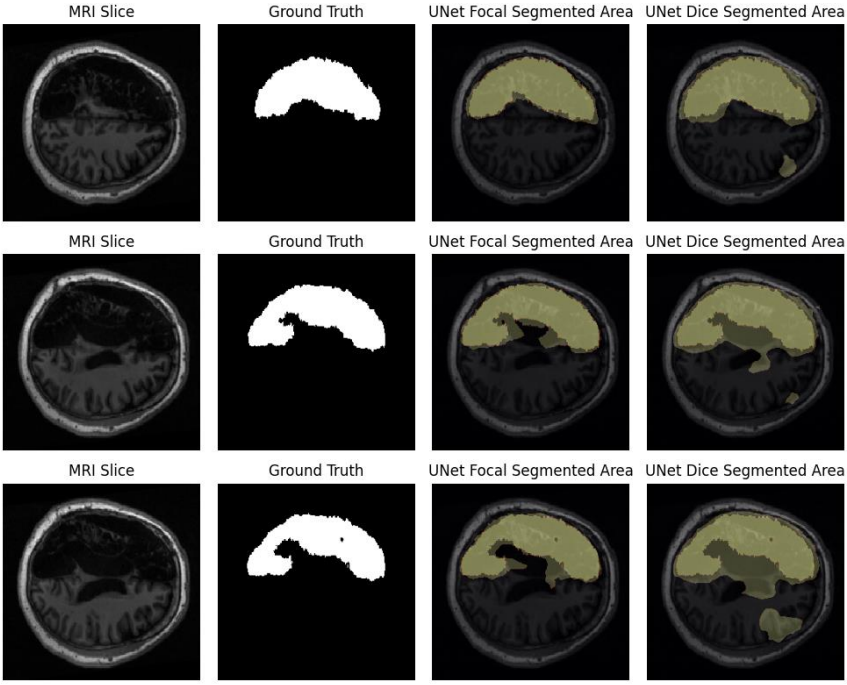

Fig. 3. The Segmentation Result based on different configurations.

Fig. 3 shows the segmentation result of the two models using different loss functions. The right two columns show images with overlapping model segmentation results and the ground truth. The brighter part is the ground truth, and the darker part is the model segmentation result. This result demonstrated that the U-Net model using the focal loss clearly has a better performance in segmenting the lesion areas with clearer boundaries, while the U-Net model using dice loss usually segments larger areas. Sometimes, the model with dice loss even segments separated areas with no lesion.

To understand the detailed segmentation accuracy, the study used the following metrics to evaluate the model segmentation performance.

Table 1. The Evaluation Metrics of Dice Loss and Focal Loss.

Loss Evaluation	Precision	Recall	DSC	AUC	IOU
Focal Loss	0.6154	0.8875	0.7268	0.9393	0.5709
Dice Loss	0.2656	0.9686	0.4196	0.9628	0.2633

From Table 1, the evaluation metrics show the performance of the models from different aspects. The study mainly uses DSC to evaluate the segmentation performance. The focal loss model has a DSC of 0.7268, while the dice loss model only has a DSC of 0.4196. This suggests that the focal loss model exhibits a notable benefit in terms of the resemblance between the model's segmentation and the ground truth. Furthermore, the accuracy of the focal loss model demonstrates its significant advantage in avoiding the missegmentation of areas, hence eliminating the need for costly human correction.

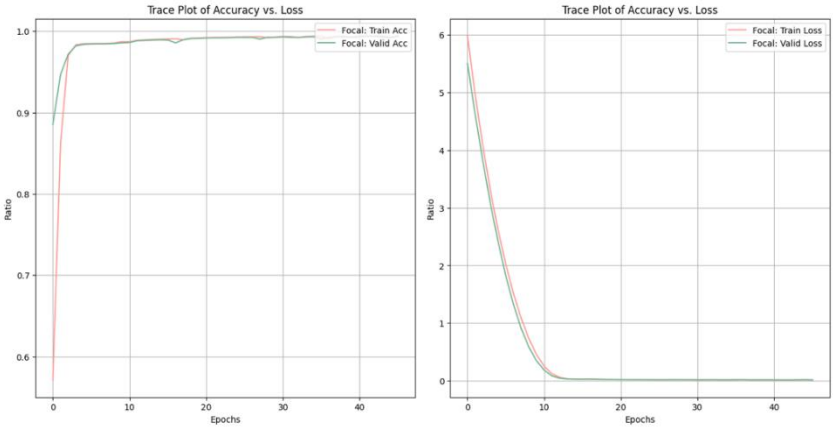

Fig. 4. Training and Validation Accuracy and Loss for the Focal Loss Model.

Fig. 4 displays the training and validation accuracy diagram on the left, while the right diagram represents the training and validation loss of the focal loss model. The accuracy diagram illustrates that the initial training accuracy and validation accuracy were very low, but both steadily improved until reaching a value close to 1. Although the validation accuracy had some oscillations while training, the accuracy between training and validation was still close, which means the generalization ability was good on this model, and there was no evidence showing the model was overfitting. The loss diagram shows that the training loss and validation loss both gradually decreased to less than 1. The diagram shows that the model was learning and had a similar performance on the training dataset and the validation dataset.

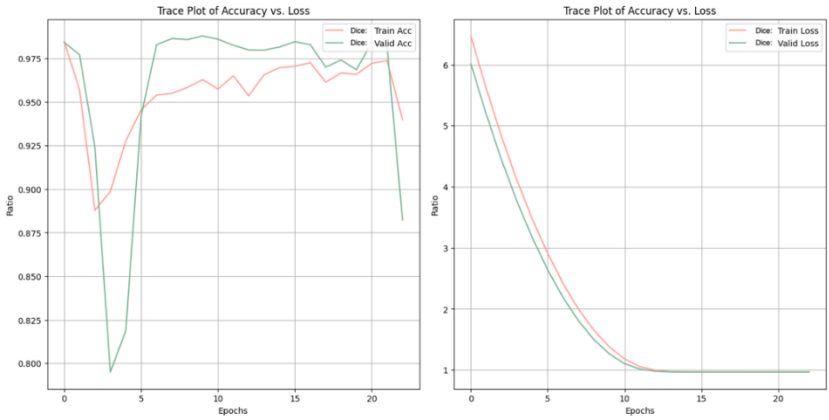

Fig. 5. Training and Validation Accuracy and Loss for the Dice Loss Model.

The dice loss model displayed the training and validation accuracy diagram on the left side of Fig. 5, while the right diagram depicted the training and validation loss of the model. The accuracy plot exhibits significant oscillations, suggesting that there is heterogeneity in the learning process throughout the epochs. However, the diagram shows that both the training accuracy and validation accuracy were gradually increasing. Although there were oscillations, the training accuracy was always higher than 90%, reaching the top of 98%. Moreover, the training and validation accuracy had significant overlaps, which means the model also performed well in generalizing unseen data. The oscillations in the validation accuracy might indicate the model's sensitivity to certain batch variations or learning rate adjustments. The loss diagram displays similar information to the loss information of the focal loss model. This diagram also shows that the model soon became convergent, which means the model was learning fast.

A study by G Dollár et al. [19] shows that focal loss can naturally handle class imbalance in one-stage detectors, promoting efficient training without selective sampling, which can be learned from images with small lesion areas. While dice loss is not that sensitive for images with large areas of background information, it is easier to ignore the minority but important class. This is the reason why the focal loss model of this study performed better than the dice loss model.

The limitation this study faces is that some segmentation results of the focal loss model are still larger than the ground truth. This is probably because of the lack of training data and the number of epochs limited by the computational recourses the model had run on. In the future, it is possible to improve the performance of this model by using a training dataset with more images that contain at least 100 pixels of lesion areas and training more epochs on a better hardware environment.

4 Conclusion

This study uses the same U-Net models with different loss functions to segment stroke lesion areas with the same imbalanced MRI dataset. The study used the same

parameters except for the loss functions to train the model in order to acquire a convincing result. The segmentation result is primarily evaluated by the DSC. This demonstrates how similar the overlapping part of the model's segmentation areas and the ground truth. As a result, the focal loss model had a DSC of 0.7628, which is significantly higher than the dice loss model. It shows that by using focal loss in this imbalanced MRI dataset, the model would be able to learn more details than using dice loss as the loss function. Nevertheless, the main constraint of the work is its exclusive reliance on a single dataset for MRI segmentation. It is hard to come to a conclusion as to whether the focal loss will have the same kind of effect on other imbalanced datasets. In the future, it is possible to find a better loss function to be used on different imbalanced datasets.

References

1. Kyu, H. H., Addo, I. Y., Adibi, A., et al.: Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. The Lancet (British Edition) 404(10157), 1173-1211 (2024).
2. Zhang, L., Lu, H., Yang, C.: Global, regional, and national burden of stroke from 1990 to 2019: a temporal trend analysis based on the Global Burden of Disease Study 2019. International Journal of Stroke 20(2), 17474930241246955 (2024).
3. Althaus, K., Dreyhaupt, J., Hyrenbach, S., Pinkhardt, E. H., Kassubek, J., Ludolph, A. C.: MRI as a first-line imaging modality in acute ischemic stroke: a sustainable concept. Therapeutic Advances in Neurological Disorders 14, 175628642110303 (2021).
4. Soleimani, P., Farezi, N.: Utilizing deep learning via the 3D U-net neural network for the delineation of brain stroke lesions in MRI image. Scientific Reports 13(1), 1-7 (2023).
5. Qiu, Y., Hui, Y., Zhao, P., Cai, C. H., Dai, B., Dou, J., Bhattacharya, S., Yu, J.: A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process. Energy 294, 130866 (2024).
6. Liu, Y., Liu, L., Yang, L., Hao, L., Bao, Y.: Measuring distance using ultra-wideband radio technology enhanced by extreme gradient boosting decision tree (XGBoost). Automation in Construction 126, 103678 (2021).
7. Liu, Y., Bao, Y.: Real-time remote measurement of distance using ultra-wideband (UWB) sensors. Automation in Construction 150, 104849 (2023).
8. Zhao, F., Yu, F., Trull, T., Shang, Y.: A new method using LLMs for keypoints generation in qualitative data analysis. In: 2023 IEEE Conference on Artificial Intelligence (CAI), pp. 333-334. IEEE (2023).
9. Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., Guo, L.: Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control 72, 103323 (2022).
10. Zhao, Y., Moayedi, H., Foong, L. K., Thi, Q. T.: Slime mold and four other nature-inspired optimization algorithms in analyzing the concrete compressive strength. Smart Structures and Systems 33(1), 65 (2024).
11. Liu, Y., Yang, H., Wu, C.: Unveiling patterns: A study on semi-supervised classification of strip surface defects. IEEE Access 11, 119933-119946 (2023).

12. Li, S., Kou, P., Ma, M., Yang, H., Huang, S., Yang, Z.: Application of Semi-supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. IEEE Access. (2024).
13. Zhao, Y., Dai, W., Wang, Z., Ragab, A. E.: Application of computer simulation to model transient vibration responses of GPLs reinforced doubly curved concrete panel under instantaneous heating. Materials Today Communications 38, 107949 (2024).
14. Hui, H., Zhang, X., Wu, Z., Li, F.: Dual-path attention compensation U-Net for stroke lesion segmentation. Computational Intelligence and Neuroscience 2021, Article 7552185 (2021).
15. Kim, C., Zhu, V., Obeid, J., Lenert, L.: Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. PloS One 14(3), e0212778 (2019).
16. An, J., Wendt, L., Wiese, G., Herold, T., Rzepka, N., Koch, S.: Deep learning-based automated lesion segmentation on mouse stroke magnetic resonance images. Scientific Reports 13, 39826 (2023).
17. Nitrc: ATLAS: Anatomical Tracings of Lesions After Stroke. http://fcon_1000.projects.nitrc.org/indi/retro/atlas.html, last accessed 2024/04/18.
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, LNCS, vol. 9351, pp. 234-241. Springer, Cham (2015).
19. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988. IEEE, Venice (2017).