CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2024.326

Using U-Net models in deep learning for brain tumor detection from MRI scans

Nguyen Minh Khiem[1*], Tran Phuoc Huy[2], and Phan Tan Tai[1]

[1]College of Information and Communication Technology, Can Tho University, Viet Nam

[2]VNPT Soc Trang, Viet Nam

*Corresponding author (nmkhiem@cit.ctu.edu.vn)

Article info.

Received 30 Jun 2024
Revised 4 Sep 2024
Accepted 10 Oct 2024

Keywords

EfficiencyNet FPN, ResneXt-50, Tumor disease

ABSTRACT

Tumor diseases in the nervous system are both dangerous and complex. Magnetic Resonance Imaging (MRI) is crucial for detecting brain disease; however, identifying the presence of tumors from these is time-consuming and requires a professional doctor. Utilizing deep learning for tumor detection in MRI images can reduce waiting times and enhance detection accuracy. We propose a method employing two U-Net models: ResNeXt-50 and EfficientNet architectures, integrated with a Feature Pyramid Network (FPN) for segmenting brain tumor. The models were trained on the BraTS 2021 dataset, consisting of 3,929 MRI scan images with 3,929 corresponding masks, divided into training, testing, and evaluation sets in a 70:15:15 ratio. The results indicate that the hybrid model, which combines EfficientNet and FPN, delivers superior performance, with an average Intersection over Union (IoU) accuracy of 0.90 on the test set compared to 0.50 for ResNeXt-50, and Dice accuracy of 0.92 compared to 0.66 for ResNeXt-50. Furthermore, we developed a web application that implements the EfficientNet with FPN model, facilitating convenient tumor detection from uploaded MRI images for doctors.

1. INTRODUCTION

Tumors are one of the most dangerous diseases of the human nervous system, and accurately identifying the brain tumor region is crucial for determining whether the tumor is benign or malignant, which in turn informs appropriate and timely treatment methods. This is highly important and significant for patient health as well as the advancement of medicine.

Magnetic Resonance Imaging (MRI) is a method that captures images of the brain by observing the water content within it. This method can clearly detect and describe abnormalities in the brain parenchyma as well as vascular tumors, arterial occlusions, venous sinus invasions, and the relationship between the tumor and surrounding structures. MRI has proven to be superior in visualizing tumors and their relationships with surrounding structures (Grover et al., 2015). To study brain tumors, the BraTS dataset (2021) (Baid et al., 2021) is a publicly available dataset consisting of MRI images of brain tumors from 2,000 patients. It was developed by the University of Pennsylvania, the U.S. Department of Health and Human Services, and the Slovak Academy of Sciences and is widely used in research.

Deep learning is a specialization within Artificial Intelligence (AI) used to solve complex and extensive problems related to image detection, speech, text, and natural language processing, and it is also applied in other areas including computer vision and natural language processing (Chen et al., 2024). Among these, U-Net is a deep learning model designed for image processing applications and is

commonly used in the medical field for medical image segmentation (Ronneberger, 2015; Yin et al., 2022). The U-Net model is designed based on the encoder-decoder architecture, where the encoder extracts important features from the image and the decoder reconstructs the image dimensions. Simultaneously, information layers are combined between the encoder and decoder to produce the final segmentation result.

The EfficientNet variant of the U-Net model is more efficient and accurate, combining three techniques: scaling, compound scaling, and neural architecture Search (Mingxing, 2019). EfficientNet has demonstrated superior performance in many image recognition tasks (Krizhevsky, 2012). A standout feature of EfficientNet is its use of neural architecture search to identify the optimal network structure (Zoph, 2017). This process automatically adjusts the network's parameters and configurations to optimize performance, saving time and effort compared to traditional manual methods.

The ResNeXt-50 backbone is an architecture of the U-Net network used for machine learning tasks, particularly in image recognition. ResNeXt-50 utilizes a residual block architecture, allowing the model to learn features from various aspects [9]. ResNeXt-50 is a powerful and flexible model for image recognition with high accuracy, especially in medical diagnosis (Zhao et al., 2021; Tanwar & Singh, 2023). However, the computational resource requirements and the complexity of optimization remain challenges that need to be addressed to fully exploit the model's potential (Mingxing, 2019; Ahmed & Sabab, 2022).

The Feature Pyramid Network (FPN) is an intelligent architecture used for object detection at different scales (Lin et al., 2017), and it combines these into a final prediction (Quyen & Min, 2023). The FPN model is built upon a pretrained network such as ResNeXt or EfficientNet (Wu et al., 2021) and is very effective in medical image recognition (Zhenghua et al., 2023). FPN uses bottom-up layers to extract features from larger feature maps. Then, the generated feature maps from top-down layers create higher-resolution feature maps. These feature maps are combined and fed into a fully connected layer to produce the final prediction for the image segmentation task (Chavan et al., 2022). Our results revealed that EfficientNet combined with FPN produced better outcomes on MRI image data.

Our contributions include:

Firstly, we utilize two deep learning methods, ResNeXt-50 and EfficientNet combined with FPN, to identify the model that detects tumor diseases from the Brain Tumor Segmentation dataset, which consists of 3,929 MRI images with 3,929 corresponding mask labels. Secondly, we compare these two methods to determine the best model for detection. Finally, we develop an application that allows users to detect tumor region from input images.

2. MATERIALS AND METHOD

2.1. Dataset

The dataset used to train and test the segmentation model was the Brain Tumor Segmentation (BraTS2021) dataset. The dataset utilized in this research consists of 3,929 MRI images and 3,929 corresponding mask images. Each image has a size of 240x240x155.

The shapes of the tumor regions in MRI images and their corresponding mask images in the dataset are illustrated in Fig 1.

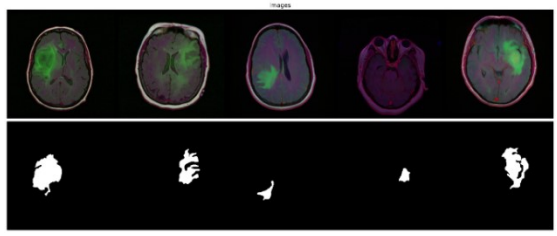

Figure 1. MRI images and their corresponding tumor masks in the dataset.

2.2. Validation

To prepare for model training, the team divided the dataset into three portions: training, validation, and testing, with the ratios: 70% for training, 15% for validation, and 15% for testing. Here, the number of images in each part was 2,750 for training, 590 for validation, and 589 for testing. The data was split using the train_test_split() function from Scikit-learn (Pedregosa et al., 2011) to create the training, validation, and testing sets.

The data ratios are shown in Figure 2 below, with the training set consisting of 2,750 images, including 1,840 images of the Positive class and 910 images of the Negative class; the validation set consisting of 590 images, including 384 images of the Positive class and 206 images of the Negative class; and the testing set consisting of 589 images; in that, 372 images of the Positive class and 217 images of the Negative class. Note that, Positive

class is tumor disease while negative is non-tumor disease.

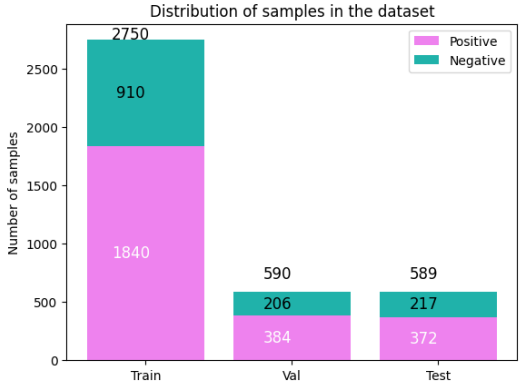

Figure 2. Ratio of training, validation and testing set

The datasets were randomly selected before training. For the U-Net model with a ResNeXt-50 backbone, the input images were resized to 256x256x4. For the EfficientNet and FPN models, the image size is 256x256x3.

2.3. Evaluation model

To evaluate the accuracy of the model, common metrics used in segmentation tasks include:

– Intersection over Union (IoU): Measures the overlap between the predicted region and the actual region. This score is also known as Jaccard index.

– Dice Similarity Coefficient (Dice): Measures the similarity between two sets, often used in medical segmentation to assess the agreement between the predicted region and the actual region.

2.3.1. IoU score

This type of metrics is used to measure the similarity between the ground truth and the prediction. The IoU formula is shown as IoU = TP / (TP + FP + FN), where TP represents the count of accurately classified pixels, while FP and FN indicate the counts of inaccurately classified pixels. (Everingham et al., 2015).

The IoU value ranges from 0 to 1, with values close to 1 indicating that the predicted pixels closely match the ground truth, signifying accurate detection or localization. On the other hand, values near 0 represent poor overlap, indicating less precise detection.

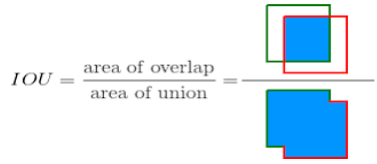

Figure 3. IoU score

2.3.2. Dice coefficient

Similar to IoU, the Dice coefficient is used to assess the similarity between the predicted object and the ground truth, and it is widely utilized in image segmentation. This measurement is computed as Dice = 2 * TP / (2 * TP + FP + FN) (Dice, 1945). Both metrics have values ranging from 0 to 1, with higher values indicating better performance of the model in image segmentation.

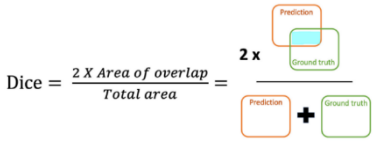

Figure 4. Dice coefficient

The Dice coefficient has the notable advantage of emphasizing the similarity between two sets, making it particularly effective when dealing with datasets that have significant overlap or when the objects to be segmented are small. This makes Dice an ideal choice for medical applications, where high accuracy in segmenting small structures such as tumors is crucial (Michal et al., 2016).

2.4. Deep learning

2.4.1. ResNeXt-50

When training the model, ResNeXt-50 is utilized to substitute the convolutional layers in the U-Net architecture. This allows the U-Net model to learn more complex and effective features, leveraging ResNeXt's ability to capture features at various levels. This can enhance the accuracy of image segmentation, especially with large and complex datasets. The architecture of ResNeXt-50 (Fatih et al., 2021) is displayed in Figure 5.

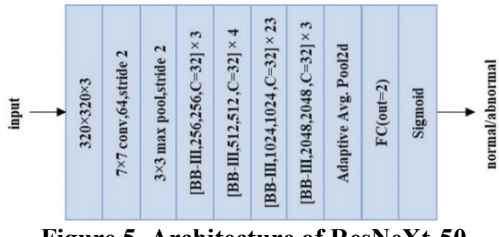

Figure 5. Architecture of ResNeXt-50

Additionally, using ResNeXt in training The U-Net model enhances the training process's efficiency by addressing the vanishing gradient issue, which happens when the gradient of the loss function nears zero, leading to a halt in the model's learning.

2.4.2. EfficientNet

EfficientNet is a deep neural network architecture designed for improved performance with lower resource consumption compared to traditional networks. The architecture of EfficientNet (Ahmed & Sabab, 2022) is shown in Figure 6.

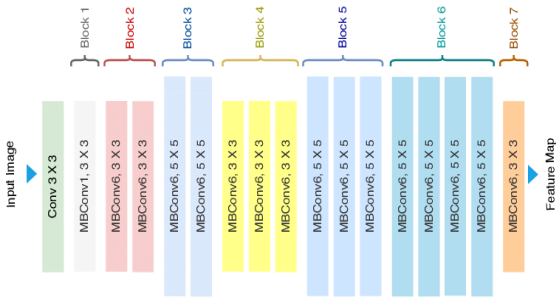

Figure 6. Architecture of EfficientNet

In this study, we use EfficientNet-B7 (Tan & Le, 2019) as the encoder to extract features from the input images, combined with a Feature Pyramid Network (FPN) to combine information from multiple levels of features to make the final prediction. To combine EfficientNet with a Feature Pyramid Network (FPN), we first extracted feature maps from multiple EfficientNet layers at different scales. Then, we used the FPN's top-down pathway to upsample the deeper feature maps and combine them with higher-resolution maps via lateral connections. This combination enhances image segmentation capabilities by combining data from different scales of the image, leading to improved and more precise segmentation. This approach produces high-quality image segmentation models.

Specifically, we used the "segmentation-models-pytorch" library to initialize a neural network model for semantic segmentation. The FPN architecture was built upon an EfficientNet-B7 encoder. Here, the "classes" parameter was set to 1 therefore, the model produced a single mask for each input image, with pixel predictions spanning from 0 to 1, and the "activation" parameter was set to "sigmoid".

3. RESULTS AND DISCUSSION

3.1. ResneXt-50

The ResNeXt-50 model was trained and stopped at epoch 60. The parameters used were: the optimizer is set to "Adam" with a learning rate of 0.001, and the activation function was "sigmoid". Notably, the model was configured with an "early stop criterion" of 6 epochs; this means that if there is no improvement in the training process for 6 consecutive epochs, the training will be halted. The findings suggest that the model converged, with a testing IoU of 0.50 and a training Mean IoU of 0.55.

As shown in Figure 7, the Dice score on the validation and test sets is 0.71 and 0.66, respectively, which is higher than the corresponding IoU values.

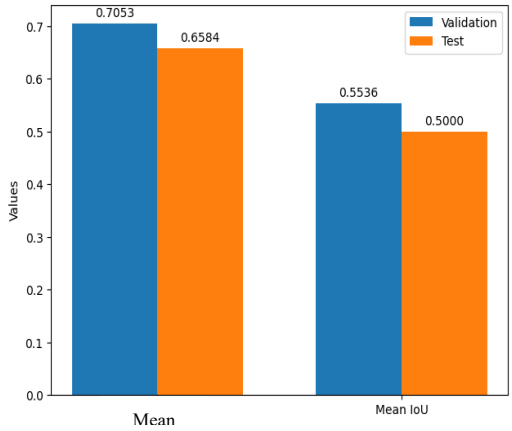

Figure 7. Comparison between IoU and Dice on the Validation/Test Sets of ResNeXt-50

3.2. EfficientNet and FPN

This model was trained with an "early stop loss" value of 6 and monitored using the loss function on the validation set. A total of 13 training runs were conducted, using the "Adam" optimizer for gradient-based optimization of the loss function during training. Figure 8 presents that the model has high results on the validation set, with a mean IoU of 0.91 and Dice coefficient of 0.93. Figure 9 shows that the accuracy between the IoU and Dice score on the validation set and testing showed no significant difference in accuracy. Specifically, the IoU values on the validation set and test set were 0.91 and 0.89, respectively, while the Dice score was 0.93 for both.

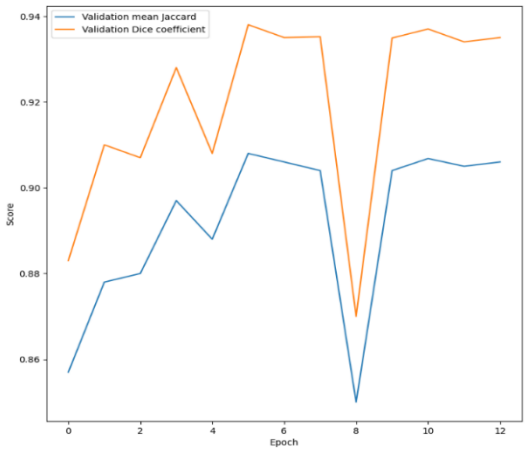

Figure 8. IoU and Dice Scores on the Validation Set for EfficientNet and FPN

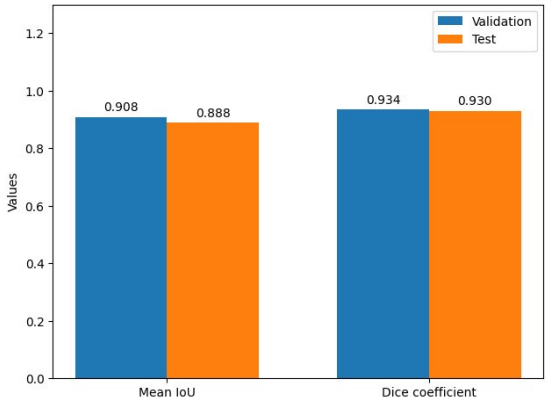

Figure 9. Comparison of IoU and Dice Scores for EfficientNet and FPN on the Validation/Test set

To compare the tumor detection performance on the same MRI image, Figure 10 displays the MRI and its mask image. Using this mask, Figure 11 presents the brain tumor detection results with the ResNeXt-50 model, achieving an IoU of 0.57 and a Dice score of 0.64.

Meanwhile, Figure 12 presents the prediction results using the EfficientNet model combined with FPN, showing an accuracy with an IoU of 0.92 and a Dice score of 0.96.

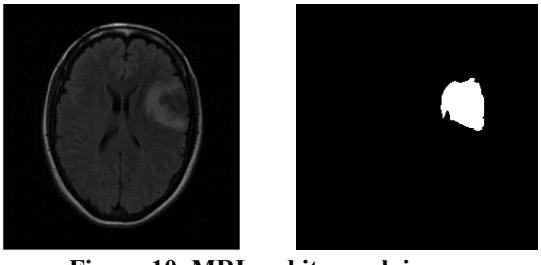

Figure 10. MRI and its mask image

Figure 13 illustrates that the combined EfficientNet and FPN model outperformed the ResneXt-50 model considerably, with Dice coefficient values of 0.92 versus 0.66 and IoU values of 0.90 versus 0.50, respectively. These metrics highlight the superior performance of the EfficientNet and FPN model in accurately delineating brain tumor regions.

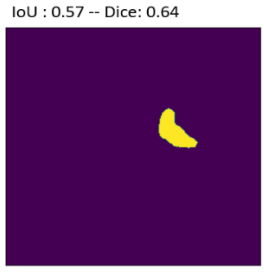

Figure 11. Predicted mask image with U-Net backbone ResNeXt-50

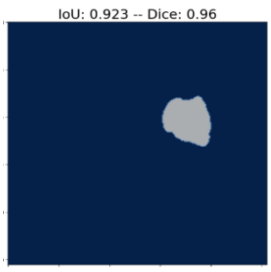

Figure 12. Predicted mask image of EfficientNet with FPN

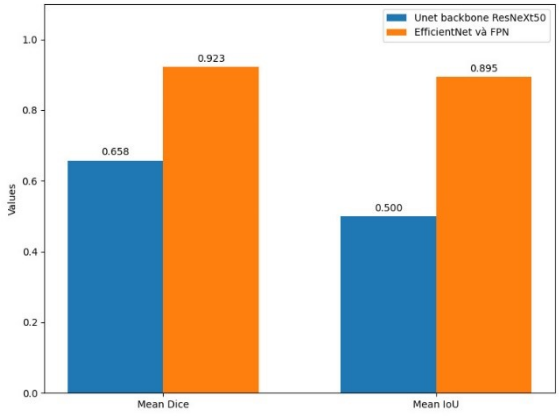

Figure 13. Comparison of IoU and Dice scores between U-Net with ResneXt-50 Backbone and EfficientNet FPN

Besides research on combined models to improve accuracy in brain tumor detection (Saddam et al., 2018; Daimary et al., 2020; Ghosh & Santosh, 2021), this study reinforces the role of FPN in enhancing accuracy when integrated with the U-Net model, specifically EfficientNet, for brain tumor detection. Tumors vary in size, shape, and location, with smaller tumors or abnormalities being critical for early intervention and patient survival. FPN enhances detection by merging high-resolution features from shallow layers with coarse, context-rich features from deeper layers, improving segmentation and classification across varying tumor sizes. EfficientNet provides a robust backbone for generating high-quality feature maps,

while FPN optimizes their utilization across multiple scales. This synergy is crucial in medical diagnosis, capturing both fine details like tumor boundaries and overall organ structure.

Despite its higher accuracy, The EfficientNet model is linked to higher computational complexity, making it more resource-intensive compared to simpler models like U-Net (Tan & Le, 2019). This complexity is due to the sophisticated architecture and extensive parameter tuning required for EfficientNet, which can pose challenges in terms of computational resources and processing time. As a result, selecting a model requires a thoughtful consideration of both accuracy and computational efficiency. In clinical settings where accurate diagnosis is critical, it is important to balance model performance with computational requirements to achieve the best results. This study underscores the importance of evaluating both the predictive performance and practical feasibility of deep learning models in medical image analysis.

To support the medical team, we propose developing a web application that implements the EfficientNet and FPN models, which run on desktops to facilitate diagnosis. Users can upload medical images, view prediction results, and store prediction history. The main screen features a button for users to select and upload images from their devices. Once an image is uploaded, the system processes it and displays the prediction results alongside the original image and the predicted mask, as shown in Figure 14.

After the user selects an image for prediction, the interface will display the original uploaded image on the left. On the right side of the interface, two result images will be displayed: the prediction image, showing the predicted brain tumor region, and the predicted mask image, indicating the precise location of the brain tumor on the medical image.

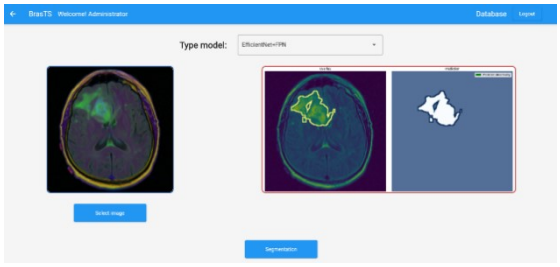

Figure 14. Result of detection

In future work, we intend to enhance the model's accuracy by optimizing hyperparameters such as learning rate, batch size, and the number of training epochs. Additionally, we plan to implement data augmentation techniques to strengthen the model's learning capacity and improve its generalization performance across varied datasets.

4. CONCLUSION

The study identified brain tumor regions using two advanced deep learning models: ResNeXt-50 and EfficientNet combined with FPN. These models were selected for their robustness in handling complex image classification tasks. After rigorous testing on the same dataset, it was observed that the EfficientNet with FPN model achieved significantly higher accuracy compared to the ResNeXt-50 model. This model can be employed in an application to assist doctors in predicting the presence of tumors in MRI images.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Ahmed, T., & Sabab, N. H. N.(2022).Classification and Understanding of Cloud Structures via Satellite Images with EfficientUNet. *SN Computer Science*, *3*(99). https://doi.org/10.1007/s42979-021-00981-2

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., ... & Bakas, S. (2021). *The RSNA-ASNR-MICCAI BRATS 2021 benchmark on brain tumor segmentation and radiogenomic classification*. https://arxiv.org/abs/2107.02314

Chen, Y., Wang, H., Yu, K., & Zhou, R. (2024). Artificial intelligence methods in natural language processing: A comprehensive review. Highlights in Science. *Engineering and Technology*, *85*, 545-550. https://doi.org/10.54097/vfwgas09

Chavan, T. A.; Lokhande, D., & Patil, D. P. (2022). Use of a Feature Pyramid Network for object detection-a deep learning technology. *Journal of Emerging Technologies and Innovative Research*, *9*(12), 2022.

Daimary, D., Bora, M., Amitab, K., & Kandar, D. (2020). Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks. *Procedia Computer Science*, *167*, 2419-2428. https://doi.org/10.1016/j.procs.2020.03.295

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297-302 (1945). https://doi.org/10.2307/1932409

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015).The PASCAL Visual Object Classes

Challenge: A Retrospective *International Journal of Computer Vision*, *111*(1), 98-136. https://doi.org/10.1007/s11263-014-0733-5

Fatih, U., Firat, H., Ozan, P., Tolga, T., & Nil, T. (2021). Classification of Shoulder X-ray Images with Deep Learning Ensemble Models. *Applied Sciences*, *11*(6). https://doi.org/10.3390/app11062723

Grover, V. P., Tognarelli, J. M., Crossey, M. M., Cox, I. J., Taylor-Robinson, S. D., & McPhail, M. J. (2015). Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. *Journal of Clinical and Experimental Hepatology*, *5*(3), 246-55. https://doi.org/10.1016/j.jceh.2015.08.001

Ghosh, S., & Santosh, K. C. (2021). Tumor Segmentation in Brain MRI: U-Nets versus Feature Pyramid Network. In IEEE 34th International Symposium on *Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 202* (pp. 31-36). https://doi.org/10.1109/CBMS52027.2021.00013

Krizhevsky, I. S. G. E. H. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

Lin, T. -Y. P., Dollár, R., Girshick, K. He, Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA* (pp. 936-944). https://doi.org/10.1109/CVPR.2017.106

Mingxing, Q. V. L. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on *Machine Learning (ICML)* (pp. 5-46). https://doi.org/10.48550/arXiv.1905.11946

Michal, D., Ethan, V., Gabriel, C., Samuel, K., & Chris, P. (2016). The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*. Springer.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Quyen, T. V., & Min, Y. K. (2023). Feature pyramid network with multi-scale prediction fusion for real-time semantic segmentation. *Neurocomputing*, *519*, 104-113. https://doi.org/10.1016/j.neucom.2022.11.062

Ronneberger, P. F. T. B (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv (Computer Science)*. https://doi.org/10.48550/arXiv.1505.04597

Saddam, H., Syed, M., & Anwar, M. M. (2018). Segmentation of glioma tumors in brain using deep convolutional neural network. *Neurocomputing*, *282*, 248-261.

Tanwar, S., & Singh, J. (2023). ResNext50 based convolution neural network-long short term memory model for plant disease classification. *Multimedia Tools and Applications*, *82*(1), 1-19. https://doi.org/10.1007/s11042-023-14851-x

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, ICML 2019, Long Beach (pp. 9-15). http://proceedings.mlr.press/v97/tan19a.html

Wu, T., Zhu, H., Fan, H., & Zhou, H. (2021). An improved target detection algorithm based on EfficientNet. *Journal of Physics: Conference Series*, *1983*(1). https://doi.org/10.1088/1742-6596/1983/1/012017

Yin, X. X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-Based Medical Image Segmentation. *Journal of Healthcare Engineering*, *15*(1). https://doi.org/10.1155/2022/4189781

Zoph, Q. V. L. B. (2017). Neural architecture search with *reinforcement learning. In Proceedings of the International Conference on Learning Representations (ICLR)*.

Zhao, X., Zhang, P., Song, F., Fan, G., Sun, Y., Wang, Y., Tian, Z., Zhang, L., & Zhang, G. (2021). D2A U-Net: Automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution. *Computers in Biology and Medicine*, *135*(104526). https://doi.org/10.1016/j.compbiomed.2021.104526

Zhenghua, X., Xudong, Z., Hexiang, Z., Yunxin, L., Yuefu, Z., & Thomas, L. (2023) EFPN: Effective medical image detection using feature pyramid fusion enhancement. *Computers in Biology and Medicine*, *163*(1). https://doi.org/10.1016/j.compbiomed.2023.107149