Accurate Segmentation of Ischemic Stroke Lesion Areas Based on Pre-trained UNets

Zhewen Guo

Information Engineering, Beijing Jiaotong University, Beijing, 100091, China
20211249@bjtu.edu.cn

Abstract. Due to the mortality and disabilities caused by ischemic stroke, it is of great significance to provide accurate segmentation during the treatment of ischemic stroke. In this study, pre-trained UNets were utilized to save the computational resource and provide accurate prediction of lesion area caused by ischemic stroke. More specifically, this study proposed four ImageNet-based pre-trained models as encoders for constructing UNets, which aim at applying non-medical priori knowledge to improve the efficiency and performance of each neural network. Additionally, a self-defined UNet was built as a baseline. All five models were trained on the ATLAS 2.0 dataset after a data filter and binary focal loss were used to mitigate the data imbalance. Finally, trainable parameters, training time and segmentation results from all five models were used for comparison. Experimental results indicate that pre-trained models achieve a recall rate of approximately 0.95 on average and consume only half of the time that self-defined UNet costs. Briefly, pre-trained models achieve a more competitive performance than that of the self-defined UNet and can deliver accurate segmentation results for patients suffering from ischemic stroke.

Keywords: Machine Learning, Transfer Learning, Image Processing.

1 Introduction

Stroke is a detrimental and potentially lethal disease, causing significant mortality and disabilities worldwide [1]. Additionally, approximately 60–80% of strokes are caused by obstructions disrupting blood flow in brain tissue, also known as ischemic stroke [2]. Thus, timely diagnosis and accurate detection of the lesion area are crucial and urgent. Magnetic Resonance Imaging (MRI) is commonly utilized to help identify and quantify these infarcts for diagnosis, supporting personalized treatment plans. However, manual detection cost considerable time and effort, potentially leading to delays in treatment timing. Besides, the process of artificially identifying lesions from MRI slices is rather subjective. Such biased diagnoses can elevate the risk of irreversible and severe brain damage, potentially causing secondary damage to patients [3]. Thus, computer science can be considered to help with the early detection and recognition of brain lesion with MRI due to their superior processing performance in a short period of time [4].

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_90

In the early field of ischemic stroke segmentation, the previous segmentation was using digital image processing technology, such as thresholding, region-growing, clustering, etc [5]. Even though these algorithms notably improve the efficiency and accuracy compared to manual inspection at that time. However, the prediction performance still has a large space to be further improved. With the rapid development of Graphics Processing Unit (GPU) and Central Processing Unit (CPU) processing power, deep learning has become extraordinarily popular and impressive, allowing model training with millions of images and providing robustness to variations in image [6]. 1989 witnessed the birth of Convolutional Neural Network (CNN) [7] which is still one of the most significant and powerful machine learning algorithms widely used in the field of medical image processing [8-11]. The U-net architecture, composed of CNN, was then proposed in 2015 and showed awesome performance on a wide range of biomedical segmentation tasks [12]. However, deep learning models normally require advanced computational resources, such as high-level CPU or GPU and large amounts of Random Access Memory (RAM) [13], which is too expensive and unacceptable to the majority of small businesses and organizations. To mitigate such problems, transfer learning is an effective and efficient method by using pre-trained structures of deep neural networks and their weights. El Jurdi et al. suggested that a Fully Convolutional Networks (FCN) with indirectly related prior knowledge could benefit the medical segmentation task [14]. But most of the pre-trained models are based on ImageNet instead of a MRI dataset, which may cause potential negative impact on training process. To address this gap, this research aims to assess the feasibility of employing an FCN with indirectly related prior knowledge for Infarct Segmentation.

In this regard, this study selected several popular pre-trained models to do transfer learning on the ATLAS 2.0R dataset, including VGG16, MobileNet, ResNet50. And these models are utilized as encoders in each U-net. To evaluate and compare the segmenting ability of different U-net combinations, metrics such as Precision, Recall, Area Under the Curve (AUC), Dice Similarity Coefficient (DSC) were measured, and this paper took trainable parameter and training time into consideration. Through this analysis, this study aims to prove that non-medical pre-trained encoder is beneficial and valid when training a model with limited computational resources and memory space.

2 Method

2.1 Data Preprocessing

In this study, the ATLAS 2.0 dataset shown in Fig. 1 is utilized to segment ischemic stroke lesions. The original dataset consisted of 955 T1-weighted 3D brain scans, which are split into a training dataset (655) with corresponding masks and a testing dataset (300) without any labels. All grayscale images are pre-processed using the MNI152 template. To evaluate the performance of the model's predictions, only a portion of the data from the training dataset with their labels is selected for this research.

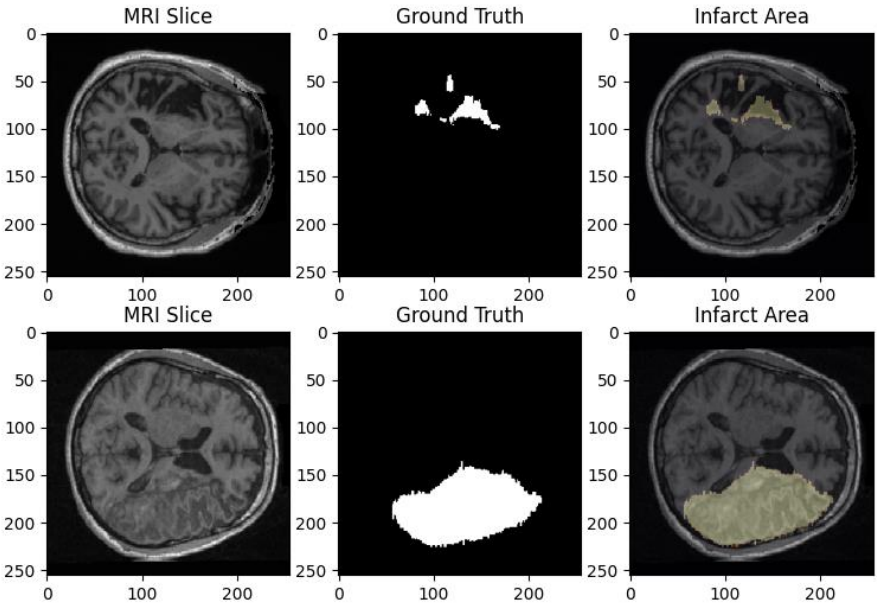

Fig. 1. The sample images and corresponding labels in the collected dataset used.

The shape of each brain scan is 197 × 233 × 189. Only 100 patients with 3D MRI scans from the training pictures with corresponding marks are selected, amounting to a total of 18,900 slices. As shown in Fig. 2, the ATLAS 2.0 dataset is extremely imbalanced with a large number of negative labels (1,229,615,012), while positive labels (9,015,388) are rarely presented. To be more specific, the ratio of the positive and negative labels is about 0.7 to 99.3 in the original dataset, which notably slows the convergence of the training process and deteriorates model's performance. To address this issue preliminarily, a data filter is made to retain those slices with at least 100 pixels labeled as positive during the data preprocessing. Following that, the proportion of positive labels sharply increases to 2.8%, which effectively enhances the quality of dataset and potentially contributes to a better model performance.

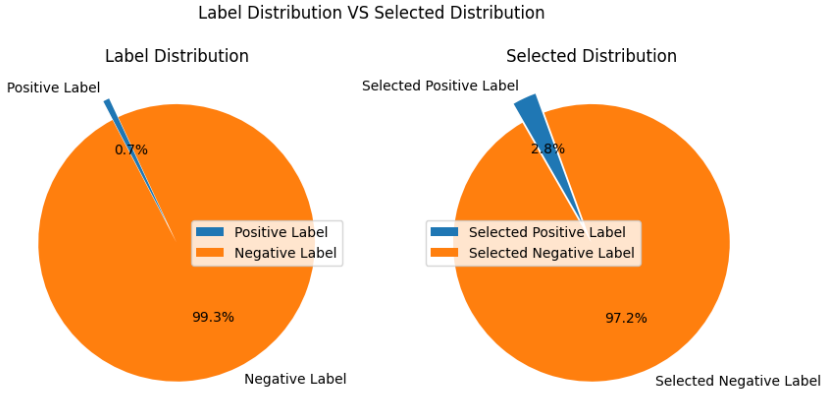

Fig. 2. The proportion of positive labels and negative labels

To ensure compatibility with various transfer learning models, the data will be standardized to a size of 255x255 through resizing. Afterward, this study splits this dataset in a ratio of 6:2:2 for the training, validation, and testing sets. To improve model convergence and reduce gradient vanishing risks, normalization was applied to scale voxel-based values within the range of 0 to 1.

2.2 Self-designed UNet

Firstly, a self-designed UNet is built as the baseline for the experiment. Then, four ImageNet-based pretrained models are used to build combined UNet, including VGG16-UNet, MobileNet-UNet, ResNet-UNet, DenseNet-UNet. As suggested by the authors [15], the FCN employs pixel-to-pixel mapping to determine pixel classes using the ground truth. FCN evolves from the classical CNN, retaining convolution and pooling layers but replacing fully connected layers with convolutional layers. The architecture of these unets commonly consists of contracting and expanding paths. A contracting path encodes pixel-based inputs into low-dimensional features (e.g., edges or lines) by down sampling, while an expanding path decodes these low-level characteristics through up sampling. The decoding process can restore the spatial resolution lost during the down sampling operations, aiding networks in generating accurate predictions.

UNet using Transfer Learning. Transfer Learning empowers a model to leverage prior knowledge to solve a related problem [16]. However, as the pre-trained model in the medical domain is rarely accessible, a model with indirectly related prior knowledge can potentially improve the performance of segmentation [17]. The architecture of the following 4 UNets using Transfer Learning utilizes pre-trained models as the encoders. These encoders all require input shapes of 256x256x3. To simulate RGB inputs, a concatenation operation is performed after the input layer stacks three inputs together.

VGG16. The architecture of VGG16 [18] is distinguished by its uniformity and depth, comprising a total of 16 convolutional and fully connected layers. The network architecture is meticulously organized into successive blocks, each containing multiple convolutional layers, interspersed with max-pooling layers for spatial down-sampling.Notably, VGG16 exhibits a substantial depth compared to its predecessors, with an extensive stack of convolutional layers, contributing to its expressive power in modeling intricate patterns within visual data. This depth facilitates the hierarchical extraction of abstract features at multiple levels of granularity, enabling the network to learn rich representations of visual semantics.

ResNet. The innovation of ResNet [19] architecture are residual blocks, which consist of a series of convolutional layers accompanied by identity mappings through skip connections. These connections allow the network to learn residual functions, capturing the difference between the input and output of each block. By learning residual mappings rather than raw mappings, ResNet facilitates smoother gradient flow during training, enabling the direct propagation of information from earlier layers to deeper layers.

MobileNet. MobileNet [20] emerges as a pioneering solution to the burgeoning demand for lightweight yet powerful CNN architectures tailored for mobile and embedded vision applications. MobileNet embodies a paradigm shift towards efficiency-driven design principles, enabling real-time inference on resource-constrained devices without compromising on performance.

DenseNet. The hallmark of DenseNet [21] lies in its dense connectivity pattern, wherein each layer receives feature maps from all preceding layers and passes its own feature maps to all subsequent layers within a dense block. This densely connected architecture promotes feature reuse and fosters deep supervision, enabling effective information flow throughout the network while mitigating the degradation issue associated with very deep networks.

2.3 Loss Function

The loss function, Binary Focal Loss, is primarily used in binary classification tasks, designed to address class imbalance and focus on hard-to-classify examples. It enhances the training of CNN by penalizing confident predictions more than uncertain ones, thereby improving the model's ability to learn from challenging samples.

$$L_{fl} = \begin{cases} -(1-p)^{\gamma}\log(p) & \text{if } y = \text{positive} \\ -(p)^{\gamma}\log(1-p) & \text{if } y = \text{negative} \end{cases} \quad (1)$$

$p \in [0,1]$ and γ are modulation factors that can be adjusted.

2.4 Evaluation Metrics

To have a comprehensive evaluation and comparison of models' performance, Area Under Curve, Dice Similarity Coefficient (DSC), Recall and Precision are used in this project. AUC is widely used metric in binary classification tasks. It represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate. And it indicates the model's ability to distinguish between positive and negative labels, with higher values indicating better performance. DSC, known as the F1 score, evaluates a model's predicted performance by calculating the intersection of the predicted and actual segmentation. A higher DSC indicates a desirable overlapping area segmented between the predicted and ground truth. Besides, Recall measures a model's capability in predicting positive instances from actual positive samples. A higher recall value illustrates that the model excels in capturing all positive instances. Similarly, precision reveals the precision of predicted positive predictions to all ground truth positive instances. A greater precision value signifies the model's proficiency in minimizing false positives.

3 Results and Discussion

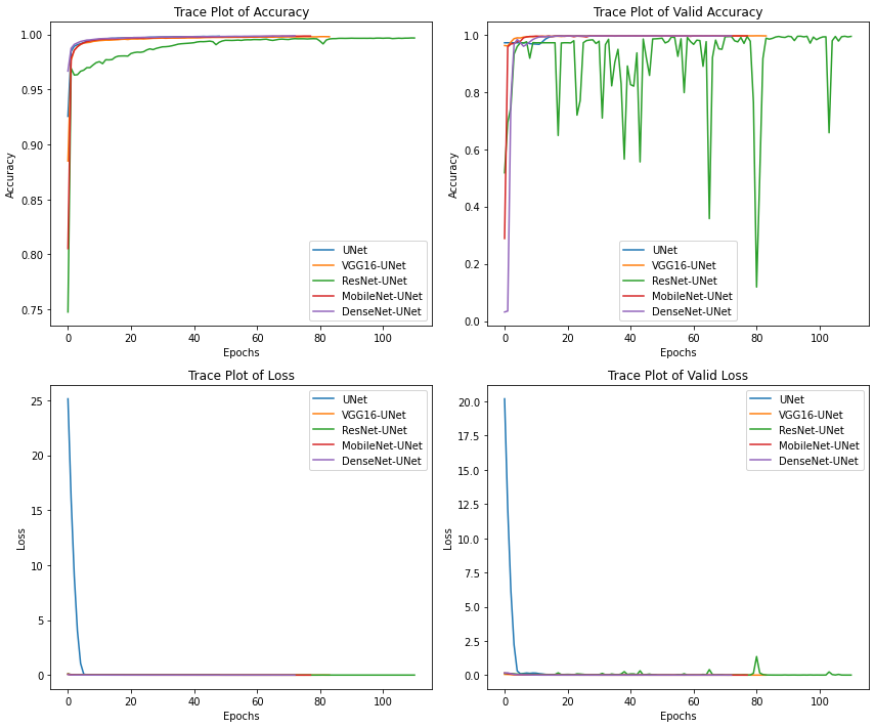

Fig. 3. UNet and combined UNet Curves in Accuracy and Loss

Fig. 3 presents four graphs that reveal improvements for the self-defined UNet and four pretrained UNet with increasing epochs. Noticeably, during the initial stage, the validation loss of self-defined UNet showed little drop, while other pretrained UNET exhibited a stable performance. After several epochs, all networks have reached saturation, achieving approximately 98% accuracy. This suggests that they have already undergone significant learning with a large number of 0 labels and have performed well on validation. However, due to the scarcity of positive labels, their predictions may not accurately reflect in accuracy. This viewpoint is validated by a continuously decreasing loss, indicating that both networks keep learning from the minority positive labels. After around 10 epochs, their loss values saturate at the same position, approximately around 0.1. However, the validation curve serves only as a reference for model training in the experiment and cannot be considered the final evaluation criterion. It is worth noting that ResNet-UNet exhibits significant fluctuations on both the accuracy and loss curves, since ResNet utilizes skip connections that directly bypass some layers, which helps the network learn more complex features. However, this architecture may make the network more susceptible to noise in the input data, leading to fluctuations in the loss curve during training. Besides, ResNet is typically deeper and more complex than the standard encoder in UNet. The increased number of layers and connections may introduce instability and variability into the training process.

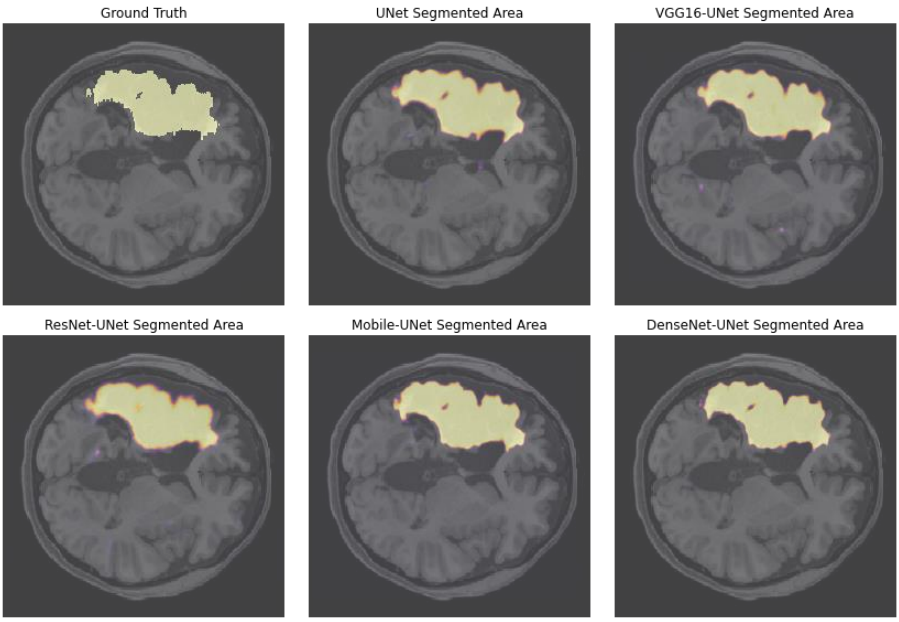

Fig. 4. Visualization of Results of each UNet

Fig. 4 showcases six sets of MR slices depicting delineated areas in the brain by self-defined UNet and four pretrained UNet, aimed at demonstrating their segmentation performance. Ischemic stroke, caused by blood flow blockages, manifests as a dark area on an MRI slice, distinctly discernible from surrounding brain cells. This lesson aligns with the ground truth, indicating the location of infarcts. In the prediction image, the

shape predicted by the model is obtained, and when overlaid on the previous image, it delineates the predicted infarct region. The highlights denote the predicted region or ground truth. It is evident that the model's predictions are relatively accurate.

The figure shows that the pretrained UNet model's predictions closely resemble the distribution of actual labels and produce fewer false positive and false negative results. On the other hand, the UNet model tends to predict more positive labels, thereby producing more False Positive results. This observation is consistent with UNet's high recall and low DSC results.

Table 1. A tabular representation of evaluation results (threshold = 0.5)

Model	AUC	Recall	Precision	DSC	Trainable Parameter	Non-Trainable Parameter	Time per epoch (s)
UNet	0.9993	0.9544	0.8969	0.9248	17,302,017	7,680	40
VGG16-UNet	0.9990	0.9622	0.8942	0.9269	11,143,809	14,718,528	20
ResNet-UNet	0.9987	0.9344	0.8847	0.9089	12,083,521	8,593,024	20
MobileNet-UNet	0.9983	0.9478	0.9077	0.9273	9,470,273	620,096	20
DenseNet-UNet	0.9986	0.9522	0.9035	0.9272	12,083,521	4,326,720	20

Table 1 presents the AUC metric that demonstrates each model equipped with a desirable ability to identify negative and positive samples, indicating the effectiveness of representation learning. It's obvious that all five models exhibit similar ability to predict positive instances, as indicated by their recall values of around 0.95 and precision values of about 0.9. DSC balances recall and precision, revealing that pretrained UNET excels in segmenting infarcts. The results demonstrate that a non-medical pre-trained encoder is beneficial and valid.

Additionally, the proposed self-defined UNet has significantly more trainable parameters than the other four pretrained models, as the encoders of the pretrained models are frozen during the training process. This also explains why the pretrained models require only half the training time per epoch compared to our self-defined UNet (40 seconds). The results illustrate that utilizing non-medical pre-trained encoder is an optional and effective way to save computational resources and memory space.

4 Conclusion

In this work, non-medical pre-trained encoders are proposed to train UNet models segmenting ischemic stroke. Data filter and focal loss function are used to mitigate the impact of imbalanced dataset and improve the performance of all five models. Additionally, all four ImageNet-based pretrained encoders models performed well on the ATLAS 2.0 dataset. Some experiments were conducted to evaluate the proposed method. Experimental results demonstrated that the models trained with this method are comparable to or even better than the centralized baseline with much fewer parameters and training time. In future research, some preprocessing methods such as Synthetic Minority Over-sampling Technique (SMOTE) will be adopted to increase the number of minority class samples for balancing class distribution.

References

1. Vidal, S. M., Chaudhry, F. S., Schneck, M.: Management of acute ischemic stroke. Hospital Practice 2(5), 99–110 (2013).
2. Chugh, C.: Acute ischemic stroke: management approach. Indian Journal of Critical Care Medicine 23(Suppl 2), S140 (2019).
3. Bao, J., Chole, R.: Combinations and Methods for Treating Non Age-Related Hearing Impairment In A Subject. US, US2013210784A1 (2013).
4. Khan, M. K. H., Guo, W., Liu, J., Dong, F., Li, Z., Patterson, T. A., Hong, H.: Machine learning and deep learning for brain tumor MRI image segmentation. Experimental Biology and Medicine 248(21), 1974-1992 (2023).
5. Deshmukh, R.D.: Study of Different Brain Tumor MRI Image Segmentation Techniques (2014).
6. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., Erickson, B. J.: Deep learning for brain MRI segmentation: state of the art and future directions. Journal of Digital Imaging 24(7), 971-987 (2017).
7. Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W.: Backpropagation applied to handwritten zip code recognition. Neural Computation 5(2), 555-568 (1989).
8. 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).
9. Ye, X., Wu, P., Liu, A., Zhan, X., Wang, Z., Zhao, Y.: A deep learning-based method for automatic abnormal data detection: Case study for bridge structural health monitoring. International Journal of Structural Stability and Dynamics 23(11), 2350131 (2023).
10. Liu, Y., Bao, Y.: Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning. Measurement 220, 113418 (2023).
11. 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).
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, LNCS, vol. 9351, pp. 234-241. Springer, Cham (2015).

13. Aghapour, Z., Sharifian, S., Taheri, H.: Task offloading and resource allocation algorithm based on deep reinforcement learning for distributed AI execution tasks in IoT edge computing environments. Computer Networks 223, 109577-109594 (2023).
14. El Jurdi, R., Petitjean, C., Honeine, P., Abdallah, F.: BB-UNet: U-Net with bounding box prior. IEEE Journal of Selected Topics in Signal Processing 14(6), 1189-1198 (2020).
15. Shelhamer, E., et al.: Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(4), 640–651 (2017).
16. Zhuang, F., et al.: A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE 109(1), 43–76 (2021).
17. Rosana, E. J., et al.: BB-UNet: U-Net With Bounding Box Prior. IEEE Journal of Selected Topics in Signal Processing 14(6), 1189-1198 (2020).
18. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science (2014).
19. He, K., Zhang, X., Ren, S., et al.: Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2016).
20. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv, abs/1704.04861 (2017).
21. Huang, G., Liu, Z., Laurens, V. D. M., et al.: Densely Connected Convolutional Networks. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE (2017).