Exploring Generalization Capability of U-net Architecture through Domain Adaptation

Yuhao Shen

Department of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou Zhejiang, 310018, China
ssyee@hdu.edu.cn

Abstract. In today's practice of medicine, imaging has significantly transformed the process of disease diagnosis, especially in neurology with brain tumor identification. Image segmentation, crucial for accuracy and efficiency, has been enhanced by U-net architecture. This neural network effectively segments brain tumors by extracting features and precisely localizing boundaries. Assessing its generalization abilities opens avenues for improved diagnostic methods and treatment plans, showcasing the potential of deep learning in advancing medical image analysis. The study utilized BRATS2020 from the source domain and BRATS2018 from the target domain, benchmark datasets for brain tumor identification. Data preprocessing aligns 3D NIfTI images with masks, extracts 2D slices, augments data and Volumetric data is reshaped for segmentation, making labels adjusted for accurate tumor localization. Domain adaptation for the U-net model involved retraining convolutional layers on a new dataset. Performance metrics were compared between source and target domains, showcasing adaptability and generalization. In the source domain, dice coefficient and Intersection over Union (IoU) were 0.6649 and 0.6366, respectively, demonstrating strong segmentation accuracy. Transfer to the target domain showed slightly lower metrics at 0.5788 and 0.5738. This highlights the model's generalization capability across different domains. The study underscores U-net's versatility, suggesting potential breakthroughs in domain adaptation. Further research on enhancing generalizability and internal parameters of the U-net model could advance segmentation tasks and optimize deep learning applications.

Keywords: Machine Learning, Deep Learning, U-net, CNN, Medical Image Segmentation, Brain Tumor

1 Introduction

In the context of contemporary medicine, the revolutions in diagnostics have played a crucial role in the advancement of knowledge of diseases. Medical imaging has brought about a drastic change in the health care sector by allowing diagnosis of internal body structures without necessarily performing surgery. However, the process of drawing and analyzing these detailed images is not an easy task that can be accomplished effortlessly. In this domain, image segmentation of medical images is

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_42

found to be a challenging task. It is very important for the regions of interest to be segmented accurately so as to minimize errors and increase efficiency. When the emphasis is placed on neurology, one of the most significant issues to address is the identification and description of brain tumors. Tumours of the brain refer to abnormal growths that occur within the brain and are a difficult subject to address because they can be in many forms and they can affect the prognosis of the patient. These tumors must be correctly identified and characterized for the best treatment plan to be developed and for the patients' benefit. Traditionally, the diagnosis of brain tumors and prediction of their growth has relied on the conventional imaging and analysis which is time consuming and requires manual intervention, the need for a more sophisticated computational tool to enhance accuracy and efficiency cannot be overemphasized.

Recently, the integration of machine learning and deep learning has been developed as potential solutions to segmentation and prediction problems in the context of medical imaging. In the classification and segmentation of medical images, Deep Learning models such as CNNs are highly effective in that they train on the dataset in order to recognize intricate patterns. As it has been noted, deep learning models, especially the models that are capable of learning hierarchical features are capable of modeling complex structures in medical images and have produced high accuracy segmentations. For instance, Kamnitsas et al. [1] have performed the current research that used CNN for segmenting the brain tumor images and outcompeting other methods by the learning ability of CNN. Another interesting study is by Isensee et al. [2] where the authors developed a deep learning model for automatic brain tumour segmentation where the multi-scale architecture was employed to enhance the accuracy of the segmentation model and its ability to generalize across different datasets.

U-net architecture that is widely used for medical image analysis has been most helpful in producing better segmentation and prediction of brain tumors. The U-net has a contracting path used to extract features and an expansive path used to precisely localize the tumor boundaries and thus improve the prediction of the model. Examining the generalization ability of the U-net model in the prediction of brain tumors opens up new possibilities for developing new approaches in the analysis of medical images, which can be useful in improving existing diagnostic methods and the development of effective treatment plans.

While conducting this study that focused on understanding the ability of the U-net model in generalizing to other datasets, it was possible to develop a technique for pre-training the U-net architecture on the source domain dataset, which is BRATS2020, a large database for medical imaging of brain tumor segmentation [3]. The U-net model for instance, with its encoder-decoder architecture accompanied by intelligent skip connections to enhance feature reuse is quite valuable in application areas that require precise segmentation due to its ability to capture high level of details and spatial dependencies of images. The U-net model was trained on the source domain dataset where it became efficient in identifying relevant features that are essential when defining the boundaries of a brain tumor accurately. As a result of this initial step, an evaluation of the pre-trained model was made with reference to the target domain dataset, BRATS2018, to assess its performance in an unfamiliar data environment, and hence test its portability to an alien data distribution. This utilisation of transfer learning was expected to strengthen the generalisability of the model and its robustness specifically within the framework of brain tumor segmentation across various datasets in order to facilitate progress in medical image analysis [4].

2 Method

2.1 Datasets Preparation

In this study the datasets for the source domain were from BRATS2020 while BRATS2018 was used for the target domain. BRATS2020 [5] is a recently proposed benchmark dataset that focuses on the identification of brain tumors and is aimed at assessing the efficiency of advanced approaches. This dataset involves utilizing pre-operative Nuclear Magnetic Resonance Imaging (MRI) scans from multiple institutions; however, the emphasis is on segmenting brain tumors, particularly gliomas due to inherent heterogeneity. These tumors have variations in appearance in the MR images, their shape and their histology which makes accurate segmentation difficult. The main focus of BRATS2020 [6] is the segmentation of the intrinsically highly heterogeneous brain tumors, which are gliomas, and may look different, have different shapes, and exhibit different histological characteristics. This means that through combining multiple MRI scans across multiple institutions, this dataset can offer a comprehensive look at these complex tumors so that researchers can explore tumor morphology and growth in detail. BRATS2018 dataset arises in the form of a sequence of NIfTI files (. nii. gz) containing native T1 and T1 post contrast and T2 and T2 FLAIR. These scans have originated from different clinical protocols and scanners of 19 centres and present a broad spectrum of routine clinical cases. Each imaging dataset included in the current study has been manually segmented by one to four raters following a specific annotation guideline that has been reviewed and approved by expert neuro-radiologists.

The data preprocessing begins with loading and organizing the dataset meticulously to align volumetric NIfTI format images with their masks. 2D slices are extracted from the 3D NIfTI images along axial, sagittal, and coronal axes to create input-output pairs for the segmentation model. Data augmentation, including techniques like horizontal and vertical flipping, enhances dataset variability and model generalization. Images are converted to grayscale for simplicity and computational efficiency, followed by grayscale normalization to standardize pixel intensity values. Volumetric data is reshaped into 2D slices for segmentation, ensuring relevant information for accurate tumor localization. Label processing involves adjusting pixel values in masks to align with the model's output requirements, such as converting to binary representations or normalizing values for the model's activation function. This standardization enables the model to differentiate between tumor and non-tumor regions effectively, leading to precise segmentation results. After the preprocessing, the MRI images of both two domains dataset have been split up into 154 slides, while BRATS2020 includes 369 images and BRATS2018 includes 240 images in total. Both of them will be divided further serving as the training and testing group.

2.2 U-net

Convolutional Neural Networks have been the key breakthrough in the field of computer vision since they allow machines to learn and analyze the images and video with good accuracy. CNNs foundation consists of convolutional layers, which involve applying filters over images to obtain features via convolution. These filters assist the network in establishing the different levels of abstraction, thus supporting the hierarchical feature extraction. Some strides or pooling layers that are utilized along with the convolutional layers help in subsampling of the feature maps, which in turn helps in reduction of the computational load and increasing the invariance to the translation. By summing up the information from the local regions, pooling layers facilitate the extraction of the most important features while rejecting unimportant features. As it was introduced by Ronneberger et al. in 2015, U-Net [7] can be classified as a Fully Convolutional Neural Network Model that has been designed for semantic segmentation. It is originally a seminal work in semantic segmentation, has a novel encoder-decoder architecture that is fine tuned for image processing, especially in medical applications. The encoder component records the context of the input image through convolution and pooling to provide the image's feature map, which is smaller than the original image. On the other hand, the decoder part helps to locate the object accurately with the help of transposed convolution structure that helps to produce a high-resolution segmentation map. This is because unlike many other networks used for semantic segmentation, the U-net shown in Fig. 1 uses the encoder's contextual understanding of the image alongside the decoder's ability to localize the pixel-wise output.

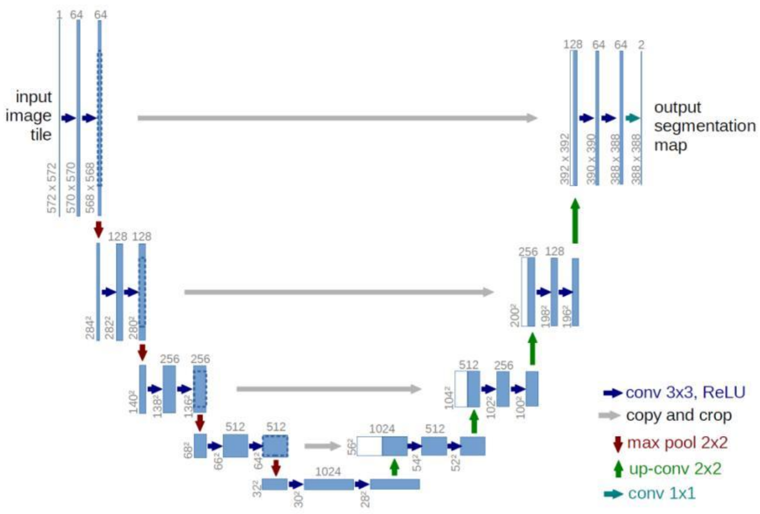

Fig. 1. The architecture of U-net [7].

The architecture can be best described as U-shaped where the encoder makes up the left side or bar of the 'U' while the decoder makes up the right side of the 'U'. The encoder pathway gradually subsamples the input image and concurrently extracts features of increasing levels of abstraction. In contrast, the decoder pathway gradually increases the size of feature maps and produces accurate segmentation masks while keeping the spatial context of the image. The skip connections are particularly crucial to retain the high frequency details while up-scaling which are the pathways between the corresponding encoder and decoder layers. Such connections help the decoder to obtain both local and global information which will enhance the possibility of segmenting the image accurately and minimize the loss of information. This kind of symmetry is not only effective in enhancing the dispersal of the information, but it also helps in achieving a balance between the general information and the specifics of the localization segmenting process. Its design is unique and highly innovative: It incorporates an encoder-decoder framework, which is a two-stream pattern that makes it possible to capture long-range dependencies in addition to the specific positions. On the other hand, the expansion path that is symmetric, the decoder, is useful in attaining accurate localization since there are transposed convolutions involved. Based on the spatial information of the corresponding feature maps of the encoder, the decoder reconstructs the high-resolution feature representations and learns to generate the segmentation maps and the pixel-wise segmentation masks for the input image. In this respect, the interconnection of the encoder and the decoder in U-net also points to the extent to which the network can maintain global context knowledge while at the same time providing the precise object boundaries in the semantic segmentation. As seen in the concise structure map below this particular architecture has placed U-net on the map for image segmentation tasks to help researchers and practitioners decode complex patterns of medical images, satellite images, self-driving cars, and many more.

2.3 Implementation Details

In the training approach of the U-net model for the identification of the BRATS set, particularly for the challenging task of brain tumor segmentation, this paper assumed that the input images are in the grayscale. In this case the learning rate is set to a small value of 0. The focus of this paper is to achieve a balance between fast convergence and high generalization of the model by using the RMSprop optimizer with selected parameters such as weight decay and momentum. Deciding to use the Binary Cross Entropy with Logits Loss function underscores the significance of the accurate outlining of the tumor margins at the time of segmentation. To ensure the efficient convergence of the model, learning rate scheduling techniques that has been proposed by Smith et al. [8] has been incorporated in the model and for the model checkpointing, the strategies recommended by He et al. [9] has been implemented. During the training process as the process iterates over the data in batches and computes the loss, this study uses tensorboardX to display the training progress in real-time based on the visualization methods described in a work by Paszke et al [10]. It does not only help to visualize the model's performance in real-time, but it also helps to fine-tune the training parameters in real-time. In this study, PyTorch and GPU acceleration are utilized to enhance the training process and at the same time, follow standard procedures recommended by PyTorch documentation and GPU acceleration of NVIDIA.

3 Results and Discussion

When domain adaptation [11] is applied for U-net model, the convolutional layers are retrained with a new dataset. The performance metrics of the model are measured in both the source domain, where the model was initially trained, and the target domain, which has a completely different channel environment. Based on the experimental process performed on the given dataset for 500 epochs with a batch size of 32, learning rate was kept to 0.1, the model also revealed a reasonable level of adaptability as well as the aptitude to generalize. The two performances – Dice coefficient and Intersection over Union in segmentation were used to assess its performance shown in Table 1.

Table 1. The performance of U-net model in different domains.

Heading level	BRATS2020 (Source)	BRATS2018 (Target)
IoU	0.636603	0.573829
Dice	0.664878	0.578848

$$Dice(A, B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{1}$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

The outcome as seen in the source domain revealed reduce a Dice number 0. 6649 and an intersect over union of 0.6366 highlights a strong correspondence to the actual degree of segmentation across all images in study while further pointing to a high degree of congruency with the ground truths. Dealing with the same pre-trained model and translating it to the target domain provided a slightly worse performance of the metrics, where including the Dice of 0.5788 and an IoU of 0.5738. This comparative analysis clearly proves the power of the model to generalize well on related but different domains, indicating that the model is flexible and versatile and performs well in unknown domains. Some samples are provided in Fig. 2 and Fig. 3.

The observed phenomenon adds to the existing evidence about how indispensable the U-net model is for generalization and implies that there's an opportunity for the network to break free from certain domain constraints and perform well in different conditions. Such possibilities of enhancing the ability of domain adaptation and generalization, which will be worthwhile to the wider scientific community and the question of what may underlie such adaptability arises [12].

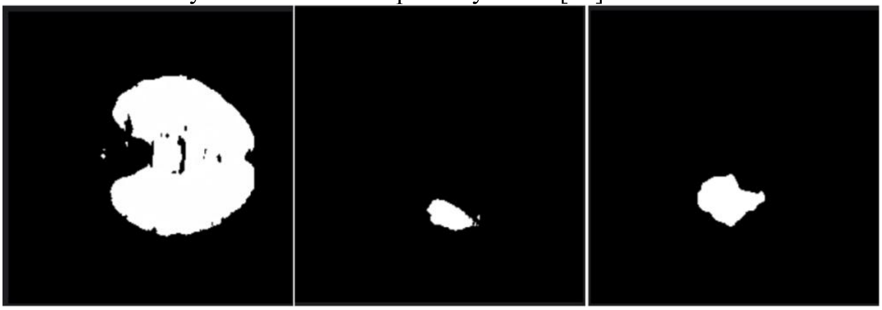

Fig. 2. Results of Source Domain (BRATS2020) (Photo/Picture credit : Original).

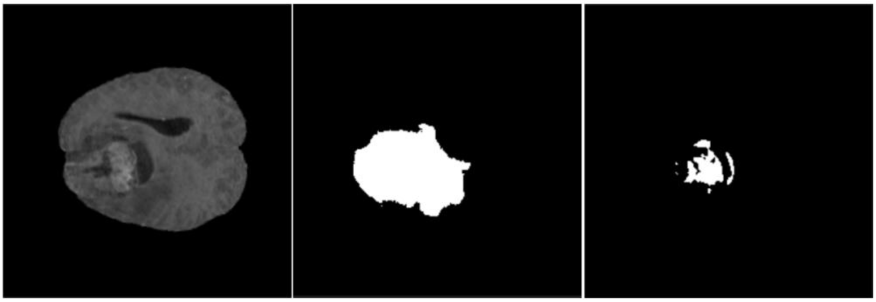

Fig. 3. Results of Target Domain (BRATS2018) (Photo/Picture credit : Original).

By studying the various factors that directly relate to the performance of particular models varying across domains, researchers stand a chance to open the door to a variety of segmentation issues and arrive at an ultimate solution for improving overall organization functionality and achieving optimal success rates in image analyzing work. In light of these positive effects, it is essential to conduct additional research efforts to explore the internal parameters that affect the generalizability of the U-net model. This investigation into these pertinent factors not only has the potential for improving the segmentation, but also the improvement of variability of parametric operations and constant optimization of deep learning image recognition methods in varied applications.

4 Conclusion

In this study, the model based on U-net architecture was firstly trained in the source domain and then applied in the target domain, concentrating on the medical image segmentation tasks. The datasets including hundreds of MRI images of tumor brain are related and preprocessed, facilitating the training section. The RMSprop algorithm is developed serving as the optimizer and single tunnel and class is applied during the training process. Experimental results showed the remarkable generalization ability of

U-net architecture compared with other neural network models. In the future, further study expects to promote the model ability of recognition and segmentation and enhance the generalization ability, trying to figure out the key factors of further success. By strengthening the weakness and improving comprehensive ability, some practical contribution can be made in the medical image segmentation especially the brain tumor recognition in the near future.

References

1. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61–78 (2017).
2. Isensee, F., Jaeger, P.F., Vessaz, M., Wengert, G., Roysam, B., Maier-Hein, K.H.: Automating medical-image segmentation with u-net deep learning. arXiv preprint arXiv:2101.02154 (2021).
3. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, Cham (2015).
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Menze, B.H.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629 (2018).
5. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (2015).
6. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data 4:170117 (2017).
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241. Springer, Cham (2015).
8. Smith, L.N.: Cyclical learning rates for training neural networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 464-472 (2017).
9. He, K., et al.: Bag of Tricks for Image Classification with Convolutional Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 558-567 (2019).
10. Paszke, A., et al.: Automatic differentiation in PyTorch. NIPS-W 8024, 8025 (2017).
11. Qiu, Y., Hui, Y., Zhao, P., Wang, M., Guo, S., Dai, B., Dou, J., Bhattacharya, S., Yu, J.: The employment of domain adaptation strategy for improving the applicability of neural network-based coke quality prediction for smart cokemaking process. Fuel 372:132162 (2024).
12. Bengio, Y., Goodfellow, I.J., Courville, A.: Deep Learning. MIT Press (2016).