Enhancing Image Segmentation for ICH through Transfer Learning from Stroke MRI to ICH CT

Lianghan Dong

Computer Science, University of Waterloo, Waterloo, N2L 3E9, Canada
a7dong@uwaterloo.ca

Abstract. A serious brain condition with a high death rate, intracranial hemorrhage (ICH) requires a precise and timely diagnosis. While Computed Tomography (CT) is commonly used for its speed and accessibility, its diagnostic accuracy is limited compared to Magnetic Resonance Imaging (MRI). However, the latter is often less accessible and more time-consuming. This study addresses this challenge by leveraging transfer learning techniques to enhance ICH detection in CT images using the detailed features available in MRI scans. Specifically, this study employed the ATLAS v2.0 stroke MRI dataset as the source domain and adapt it to a target dataset of ICH CT images from Kaggle. Three models, including U-Net, Residual U-Net, and Attention U-Net, are trained and evaluated on both datasets to assess their performance and transfer learning capabilities. The models are pretrained on the MRI dataset and fine-tuned on the CT dataset, utilizing different learning rates for encoders and decoders to preserve and adapt features effectively. The results indicate that the Attention U-Net outperforms the other models, demonstrating superior performance in both training and testing metrics. This research demonstrates the potential of combining the rapid assessment capabilities of CT with the detailed visualization strengths of MRI, setting a new standard in the diagnosis of ICH. The findings suggest that transfer learning from stroke MRI to ICH CT can significantly improve segmentation accuracy, thereby enhancing diagnostic efficiency and patient outcomes.

Keywords: Computer Vision, Image Segmentation, Transfer Learning.

1 Introduction

The Intracranial Hemorrhage (ICH) is a severe neurological condition that requires a precise and timely diagnosis [1]. Research reveals that ICH patients has over 40% death rate in 30 days [2]. Therefore, a fast and accurate detection methodology is crucial for ICH diagnosis. Computed Tomography (CT) is usually a first choice for ICH detection because of its speed and availability. However, it is challenging to detect ICH via radiological assessment of intracranial blood, due to the variance of intracranial blood caused by age and location [1]. On Magnetic resonance imaging (MRI), the appearance of intracranial hemorrhage undergoes complex signal intensity changes [1], meaning an accurate detection could be established in this way. Since MRI is time consuming

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_56

and may not be available in some small hospitals, it is required to have a detection methodology that combines the speed of CT and the accuracy of MRI. Deep learning is being used extensively for medical picture segmentation these days, and numerous publications have been published demonstrating its effectiveness in the field [3]. Originally, a number of models were created for medical and biological picture segmentation, drawing inspiration from FCNs and encoder-decoder models [4]. For instance, U-Networks with very few training images and yields more precise segmentations [5]. Based on U-Net backbone, there are also several advanced models like Residual U-Net and Attention U-Net.

In general, previous research trained and validated the model on a dataset with a single distribution. For ICH CT/MRI datasets, the shortage of cases forces researchers to enhance model performance by reconstructing model itself. However, the disability of model generalization could not be solved in this way. Moreover, most of medical image datasets merely focus on one domain. For ICH, CT datasets are common, but MRI datasets are not because of the scarcity and collection difficulty of ICH MRI data. One simple and direct approach is to train model on MRI dataset in other domain, such as stroke or brain tumor, and then apply this pretrained model to ICH CT dataset. This approach is called transfer learning. Transfer learning improve learner from one domain by transferring information from a related domain [6]. Due to the relationship of Stroke MRI and ICH CT, the transfer learning method could be applicable.

This research aims to bridge this gap through the application of transfer learning techniques, leveraging the rich detail available in MRI data to enhance CT image segmentation. This study applies transfer learning method by utilizing the ATLAS v2.0 stroke MRI image segmentation dataset [7] as the source and adapting it to a target dataset of brain CT images with ICH masks from Kaggle. Three models will be trained and evaluated on both datasets to assess their transfer learning capabilities using several metrics. This approach promises to combine the rapid assessment capabilities of CT with the detailed visualization strengths of MRI, potentially setting a new standard in ICH diagnosis.

2 Method

2.1 Data Preparation

This study contains 2 domains of datasets. The source domain utilizes the ATLAS v2.0 dataset [7]. The ATLAS v2.0 dataset contains 955 MRI images (655 labeled, 300 non-labeled). MRIcron, an open-source program for visualizing brain imaging and designating volumes of interest, was used to manually draw masks on each individual brain in native space once brain lesions were discovered for each MRI. For every MRI, at least one lesion mask could be found [7]. Fig. 1 displays an example of the source dataset. The Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation v1.3.1 dataset is used as the target domain. This dataset with 82 CT images was gathered, 36 of which were for patients with an intracranial bleeding diagnosis [8]. Both source and target dataset are composed of gray scale images and

masks with 2 categories (background, focus). The formats of both source and target dataset are .nii.gz files.

The source and target dataset were processed by preprocessing, dateset generation and data augmentation. The preprocessing step includes loading, slicing, filtering, and saving the data. The .nii.gz data of images and masks was initially loaded by nibabel. Since each .nii.gz file contains multiple 2D images, it was sliced with interval = 10. The image, mask slices would be filtered by the ratio of focus in the mask with threshold = 0.005. Then the pipeline reshapes original images and masks into 224 × 224 × 1 images and save filtered slices into images and masks folders. For dataset generation, the tf.data.Dataset generated the dataset by reading and zipping images and masks. The images and masks were resized into 0 to 1 by dividing 255, then split into training, validation, and test set (60%, 20%, 20%). The training and validation applied data augmentations including random zooming, random rotating and random flipping. These augmentation process will be repeated for 5 times.

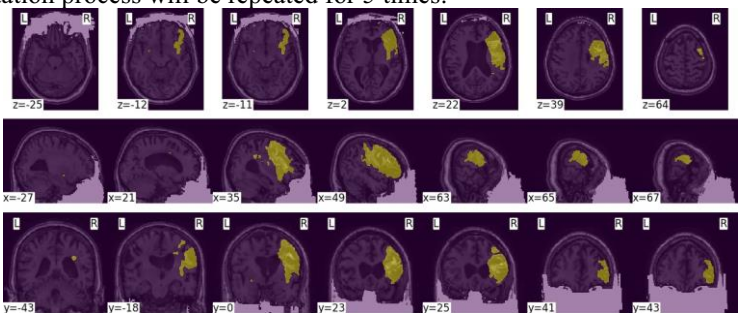

Fig. 1. A sample image with mask in ATLAS v2.0 dataset.

2.2 Transfer learning-based U-Nets

Transfer Learning. Transfer Learning can be categorized into heterogeneous and homogeneous. When the measurements (features) of the source domain project differ from those of the target domain project, this is known as heterogeneous transfer learning. On the other hand, homogenous transfer learning occurs when the domain metrics are same for projects in the source and destination domains.

This study is heterogeneous transfer learning of image segmentation from Stroke MRI to ICH CT. Three models are pretrained on source domain and fine-tuned on target domain by applying different learning rates in encoders and decoders. A lower learning rate for encoders preserves the pretrained features, preventing them from being drastically altered and preserves diversity of pretrained features. A higher learning rate for decoders improves the speed of adapting the target domain and utilizing the extracted features from encoders effectively.

Residual blocks. When deeper networks are able to begin converging, a degradation issue has been revealed: accuracy becomes saturated (which may not come as a surprise) and then rapidly declines as the depth of the network increases. [9]. The skip

connection in residual blocks helps mitigate the vanishing gradient problems and allows gradients directly flow through connections, making the training process easier and robust. The structure of the residual block is shown in Fig. 2.

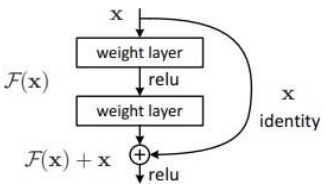

Fig. 2. A building block of Residual learning.

The structure of residual block in this work includes 1 input, 2 convolution layers with 3 × 3 filters (one stride=2, the other one stride=1), Batch Normalization, and activated by sigmoid. The input passes through a convolution layer with 1 × 1 filter and stride=2 to match the output shape, then skip connected with the activated result, and output with a Batch Normalization and Dropout.

Attention mechanisms. The human visual attention system, which selectively focuses on portions of the visual field to gather precise information while disregarding less relevant regions, is the model from which neural networks' attention mechanisms are derived. It learns to emphasize salient feature information that is essential for a given job in feature maps while automatically focusing on the target region and suppressing unnecessary feature responses [10, 11]. The basic architecture of Attention Gate is shown in Fig. 3.

The architecture of Attention Gate in this work contains 2 inputs x, g, which are convolved separately with filters shapes 1 × 1, then added together and activated by ReLU. This part is convolved by 1×1 filters again with sigmoid activation. After resampling, this part would multiply with input x and outputs xˆ.

Attention Gate enables the original model to have soft attention via multiplying extracted feature from CNN with a sigmoid generated attention ranging from 0 to 1. This technique allows model to focus on essential parts to acquire more details and prevent learning from noises and irrelevant features. This mechanism would lead to a better transfer learning result by preserving attention instead of just features of source.

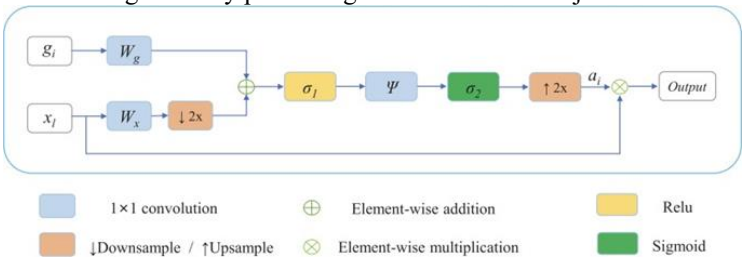

Fig. 3. The basic schematic of attention gate .

U-Net Architecture. For biomedical image segmentation, U-Net is one of the useful models with a Fully Convolutional Network (FCN) including encoder, bottleneck, and decoder. Because of its U-shaped structure with context information, short training time, and minimal amount of data required, the well-liked U-Net meets the needs of medical picture segmentation [12]. The basic U-Net architecture is shown in Fig. 4.

The U-Net architecture is used as a backbone for 3 models in this work. Those models are consisted of a set of downsampling and upsampling blocks with filters = [64, 128, 256, 512, 1024]. The downsampling block is consisted by 2 Convolution - Batch Normalization - ReLU blocks. The upsampling block is composed of Transpose Convolution - concatenate - Convolution blocks. 4 downsampling constitutes the encoder part, 1 for bottleneck, and 4 upsampling for decoder.

The Residual U-Net replaces all Double-Convolution Layers in the encoder of original U-Net architecture. The Attention U-Net keeps the same encoder, applying AG with inputs from encoders and upsampled result in decoders with same output shape.

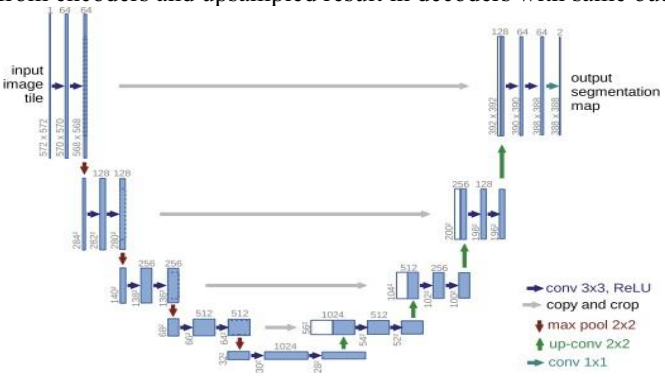

Fig. 4. The basic architecture of U-Net.

2.3 Implementation Details

Loss Function. This study applies the combined loss function of Binary Cross Entropy and Dice Loss. For each formula below, y represents the true value and p represents the predicted value.

For a given random variable or sequence of events, the Binary Cross Entropy (BCE) primarily analyzes the difference between two probability distributions. Since segmentation is pixel-level classification, it is frequently employed for classification objectives and performs well [13].

$$BCE = -ylogp - (1 - y)log(1 - p) \tag{1}$$

In the field of computer vision, the dice coefficient is a widely used statistic that evaluates the similarity between two images. Additionally, it was modified as the Dice Loss (DL) loss function later in 2016 [13]. It mainly focuses on the overlapping of true and predicted masks.

$$DL = 1 - \frac{2yp+1}{y+p+1} \tag{2}$$

The combined loss function is adding BCE and DL, which combines the difference of classes and the overlapping of masks.

$$L_{combined} = BCE + DL \tag{3}$$

Optimizer. The Adaptive Moment Estimation (Adam) optimizer, which combines AdaGrads with RMSProp, is used in this work. Adam has a number of benefits, including the ability to work with sparse gradients, the elimination of the need for a fixed objective, the ability to execute step-size annealing naturally, and step-sizes that are approximately limited by the step size hyperparameter [14].

Evaluation Metrics. The evaluation metrics for this study applies the Dice coefficient and the Intersection Over Union (IoU). The degree of overlap between two datasets is described by the Dice coefficient. It is computed by multiplying the overlap between the two datasets' total pixel counts by two. For an overlapping pair of reference and image objects of NR and NI pixels, respectively, the dice coefficient, Dc, can be expressed as follows:

$$D_c = \frac{2N_O}{N_R+N_I} \tag{4}$$

where NO is the amount of overlapped pixels [15]. The Jaccard index, usually referred to as the Jaccard similarity coefficient, is another name for IoU and is a commonly used metric for comparing the similarity between sample sets. In general, the intersection of sample sets A and B defines IoU (A, B).

$$IoU(A,B) = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{|A|+|B|-A \cap B} \tag{5}$$

Many works of computer vision like object detection and image segmentation at the pixel or instance level, have made extensive use of the IoU as an assessment metric [16].

3 Results and Discussion

This study evaluates the performance of U-Net, Residual U-Net and Attention U-Net on both source domain, target domain, and the performance of transfer learning, using dice score and IoU as evaluation metrics. Both evaluation and visualization are shown in the following sections.

3.1 The Performance on Source Domain

For the source domain, the Attention U-Net has the best result that is IoU=0.5619, which is shown in Table 1. Fig. 5 selects a sample from the testing set of Source dataset for visualization. All of the predictions are close to ground truth while the Attention U-Net performs better on details such as the size and shape of mask. For instance, the right boundary of prediction on Attention U-Net is flatter that is closer to the ground truth.

Table 1. The Performance of Different Models on Source Dataset

Model	Training Loss	Training IoU	Testing Dice Score	Testing IoU
U-Net	0.1777	0.7481	0.6503	0.5166
Residual U-Net	0.1752	0.7507	**0.6869**	0.5567
Attention U-Net	**0.1497**	**0.7821**	0.6863	**0.5619**

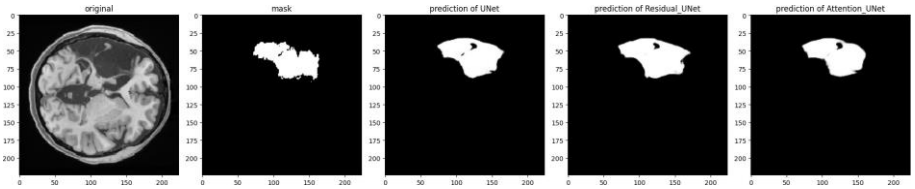

Fig. 5. Visualization of Results of Different Models on Source Dataset.

The reason for high testing evaluation metrics may be because the size of source dataset is large. With data augmentation, the test result would be boosted. The advantage of Attention U-Net is the soft attention mechanism, leading to more focus on key features and preventing disturbances.

3.2 The Performance on Target Domain

For the target domain, the Attention U-Net performs the best that is shown in Table 2 and Fig. 6. Both training and testing metrics of Attention U-Net are the best. All evaluation metrics for all models on target domain are lower than the result on source domain.

Table 2. The Performance of Different Models on Target Dataset

Model	Training Loss	Training IoU	Testing Dice Score	Testing IoU
U-Net	0.3887	0.6102	0.2757	0.2455
Residual U-Net	0.3468	0.5669	0.3250	0.2579
Attention U-Net	**0.2054**	**0.6881**	**0.4213**	**0.2714**

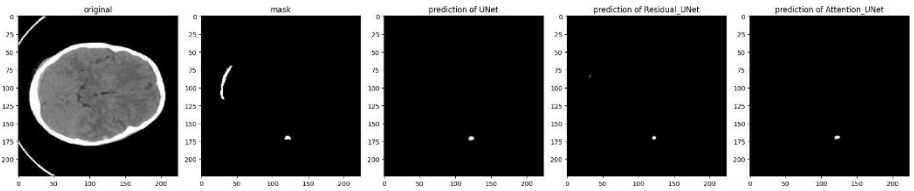

Fig. 6. Visualization of Results of Different Models on Target Dataset.

The decrease of evaluation metrics might be because the size of target dataset is small, leading to difficulty of feature extraction. Furthermore, the data quality of CT images is lower than MRI images, leading to abundant ambiguous features and noises. For Attention U-Net, it decreases the disturbance of noises in dataset by attention mechanism. Since the early-stopping is applied in training callbacks, it goes further than U-Net and Residual U-Net.

3.3 The Performance of Transfer Learning

For all these three models, transfer learning enhances most performances shown in Table 3 and Fig. 7. The Attention U-Net has the best performance, and U-Net has the best improvement.

Table 3. The Performance of Different Models of Transfer Learning

Model	Training Loss	Training IoU	Testing Dice Score	Testing IoU
U-Net	0.1743	0.7096	0.4203	0.2666
Residual U-Net	0.1871	0.6876	0.3811	0.2397
Attention U-Net	**0.1444**	**0.7365**	**0.4240**	**0.2704**

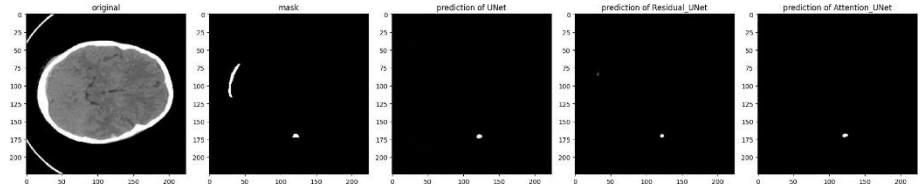

Fig. 7. Visualization of Results of Different Models of Transfer Learning.

This phenomenon might be because of the difference in feature reusing ability of different model. For U-Net, the model structure is simple, and the number of parameters is small, so that it is easier for this model to adjust parameters. Thus, the improvement of U-Net by using transfer learning is significant. Both training and testing metrics are improved for U-Net. For Attention U-Net, although the model structure is complex, it has attention mechanism, leading to a better feature selection capability. Therefore, the training performance and testing performance especially testing Dice Score can be improved. For Residual U-Net, the skip connection is capable for handling vanishing gradient, but it is not regnant for reusing features learnt from source domain. Unlike the features, the residuals might differ for different datasets, that is not beneficial for

Residual U-Net to adapt the new features. Also, both source and target domain contain non-labeled data which cannot be used directly in transfer learning. This limitation can be reached by domain adaptation in future research.

4 Conclusion

This study applies transfer learning from stroke MRI to ICH CT that enhances the performance of image segmentation for detecting and locating ICH. The study compares the performance of three models including U-Net, Residual U-Net and Attention U-Net, using Dice Score and IoU as testing evaluation metrics. The result suggests that Attention U-Net performs better in transfer learning than both Residual U-Net and U-Net. For future studies, the variety and non-labeled data of source and target domain could be considered, and domain adaptation techniques could be applied to further generalize this task.

References

1. Parizel, P., Makkat, S., Van Miert, E., Van Goethem, J., van den Hauwe, L., De Schepper, A.: Intracranial hemorrhage: principles of CT and MRI interpretation. European Radiology 11(9), 1770-1783 (2001).
2. Woo, D., et al: Risk Factors Associated with Mortality and Neurologic Disability After Intracerebral Hemorrhage in a Racially and Ethnically Diverse Cohort. JAMA Network Open 5(3), e221103 (2022).
3. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET Image Processing 16(5), 1243-1267 (2022).
4. Minaee, S., Boykov, Y., Plaza, A., Porikli, F., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: A survey. arXiv:2001.05566 (2020).
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015).
6. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. Journal of Big Data 3, 1-40 (2016).
7. Liew, S.-L., et al.: A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Scientific Data (2022).
8. Hssayeni, M.: Computed tomography images for intracranial hemorrhage detection and segmentation. PhysioNet (2020).
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015).
10. Zhang, J., Jiang, Z., Dong, J., Hou, Y., Liu, B.: Attention gate resu-net for automatic MRI brain tumor segmentation. IEEE Access 8, 58533-58545 (2020).
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. Yin, X.X., Sun, L., Fu, Y., Lu, R., Zhang, Y.: U-net-based medical image segmentation. Journal of Healthcare Engineering (2022).
13. Jadon, S.: A survey of loss functions for semantic segmentation. In 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE, (2020).

14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017).
15. Guindon, B., Zhang, Y.: Application of the dice coefficient to accuracy assessment of object-based image classification. Canadian Journal of Remote Sensing 43(1), 48-61 (2017).
16. Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., Yang, R.: Iou loss for 2D/3D object detection. In 2019 International Conference on 3D Vision (3DV), pages 85-94, (2019)