Stroke detection by image segmentation with Residual U-Net

Anonymous CVPR submission

Paper ID *****

Abstract

Neural networks are commonly utilized for medical purposes. Since medical fields require precise pixel-wise segmentation, the models trained for medical tasks are needed to perfect their performance. U-Net, a common neural network architecture named after its symmetric u-shape, is well studied for integration with automatic diagnoses thanks to its performance in image feature extraction. Stroke, with one of the most devastating symptoms, is often studied along with U-Net for goals to improve the efficiency and precision of its auto diagnoses. In this work, the Anatomical Tracings of Lesions After Stroke Dataset Version 2.0 (ATLAS R2.0) is used as the dataset for both training and testing. The dataset contains annotated T1-weighted brain MRI scans. The main focus of the work is the possible improvements integrating residual blocks brings to the regular U-Net architecture in the task of segmenting lesion areas. Hence, a basic U-Net and a U-Net with residual blocks in its encoder are trained in this work. The performance of the two models is compared based on their scores on the selected evaluation metrics, Intersection over Union, and Dice Coefficient. The two metrics that compute overlap between datasets are selected because the goal of the work is to make the predicted lesion mask as close to the annotated one as possible. The final evaluation of the output of the two models showcases that the residual U-Net scores higher in both IoU and Dice. Last but not least, a visualized display consisting of the actual lesion mask along with sample lesion predictions of both models is presented in the work. The visualized result shows that the residual model's prediction is more similar to the actual lesion. This work suggests improved precision in the integration of residual blocks of U-Net, which spurs the utilization of residual blocks in other fields.

1. Introduction

Worldwide, cerebrovascular accidents (stroke) are the second leading cause of death and the third leading cause of disability [1]. Thus, a powerful detection method for stroke is extremely required for treating and preventing it. For medical usage, image segmentation is crucial for the localization of lesions, enabling people to diagnose the cases quantitatively and qualitatively. Therefore, a robust stroke segmentation model is highly demanded.

This context is aimed to introduce the advantage of Residual U-Net in medical image segmentation using ATLAS Dataset [2]. The ATLAS v2.0 [2] was released in 2018 and of which, 229 standardized subjects were available with T1-weighted MRI image and its corresponding lesion mask [3].

1.1. Background

Nowadays, there are a variety of image segmentation models based on deep learning. One of the most familiar models is the encoder-decoder-based model. The encoder part aims to extract features and the decoder part reconstructs the result (masks, new patterns) that we need. For medical image segmentation, we need to extract the features of tissues and reconstruct the mask of lesions for localization. However, a straightforward sequential encoder-decoder structure may not have a satisfying result since the boundary in tissues may not be obvious. Moreover, medical image data is too expensive to generate so the datasets are generally too small for common encoder-decoder structures. To solve this, several models were initially developed for medical/biomedical image segmentation, which is inspired by FCNs and encoder-decoder models [4].

1.2. U-Net

The U-Net is built upon a more elegant architecture, the so-called "fully convolutional network" [5]. It has been modified and extended such that it works with very few training images and yields more precise segmentations [6]. The downsampling and upsampling with skip connections form the network a U-shape. At each downsampling step, we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a

ReLU [6].

To make the input shape more compatible with the ATLAS v1.2 dataset, we reshaped some parts of U-Net. In this context, we use the result of the U-Net as a baseline to show how the Residual block enhances the common U-Net.

1.3. Residual U-Net

Deeper neural networks are more difficult to train [7]. The depth of U-Net is relatively high so applying residual block is a considerable approach. We have referenced the structure of Deep Residual U-Net [8] and choose to apply residual blocks into the encoder part that prevents our network goes too deep. We use BatchNormalization, ReLU Activation, Conv2D, and Add to construct the residual block.

1.4. Dataset

The ATLAS v2.0 dataset contains 955 MRI images (655 training, 300 tests) from 11 cohorts worldwide from research groups in the ENIGMA Stroke Recovery Working Group consortium. For each MRI, brain lesions were identified and masks were manually drawn on each individual brain in native space using MRIcron, an open-source tool for brain imaging visualization and defining volumes of interest (http://people.cas.sc.edu/rorden/mricron/index.html). At least one lesion mask was identified for each individual MRI [2].

1.5. Contribution

We confirm the improvement of model predictions by applying residual block in the U-Net encoder part on the ATLAS v2.0 dataset compared with the U-Net model.

2. Methodology

This section mainly discusses processes of data visualization, data preprocessing, model, and result to prove the advantage of Residual U-Net compared with U-Net.

2.1. Data Visualization

The dataset consists of 955 .nii.gz data, which requires nibabel and niblearn to read and visualize. We directly visualize this dataset by niblearn in Figure 1. We observe that each sample contains the original and mask shaped in 3 dimensions. To apply a 2D model, it's crucial to slice into 2D images. Moreover, some slices may not contain masks. Therefore, selecting a range of slices that contain most masks could improve model performance and reduce training time.

2.2. Data Preprocessing

For the ATLAS v2.0 dataset, 655 images were labeled with masks, and 300 images aimed to be test data. We first split the dataset into train, val, and test sets (60%, 20%, 20%). In the dataloader, we slice each sample from a 3D block into 100 2D images. We pick index 70 - 90 as a batch.

Then we apply data augmentation: Windowing with normalization, random flipping, random rotation, and random zooming. We repeated this process 3 times to generate a larger dataset.

The entire data pipeline is shown in Figure 2.

2.3. Model

In this work, the architectures of the models used are based on U-Net. Following is a detailed explanation of the models.

2.3.1 U-Net

The model includes an encoding path, a bottleneck, and a decoding path. The parts together facilitate the extraction and utilization of key information for the segmentation of the MRIs in ATLAS R2.0. The u-net architecture used in the work consists of a filter set of 32, 64, 128, 256, and 512. Noticeably, the basic convolutional block consists of 2 layers of conv2D functions with 3×3 kernels, Batch Normalizations, "ReLU" activations, and "same" padding. In the encoder, we use max pooling with a 2 by 2 kernel and a stride of 2. We first apply pooling to the output of a basic convolutional block with input images as the input and the first filter size in the above filter set. Then, the output is taken as the input, and the next filter size is utilized. After repeating this process four times, the output obtained is utilized to generate the bottleneck convolution layer by applying the last kernel size. The decoder recovers the spatial dimensions and as many features as possible from the input. The decoder mimics the repetitions. In each repetition, the Con2D transpose is applied to the bottleneck layer to generate a layer. The layer is concatenated with the convolutional layer right before the bottleneck. Then, conv2D is applied to the concatenated one to generate a refined layer with better segmentation. The process is repeated four times. In the end, the output of the decoder is used as the input of a convolutional layer with a 1 by 1 kernel and a sigmoid activation to map the features into a single channel. Eventually, the output showcases the segmentations.

2.3.2 Residual U-Net

The motive behind the integration of residual units is residual U-Net's proven performance enhancements in feature extraction for other tasks [8]. The residual block in this work identifies mapping and transformation using convolutions. It comprises two convolutional layers, followed by batch normalization and ReLU activation. The first convolution reduces the dimensions while improving the depth of the feature maps by applying a 3x3 kernel with the same

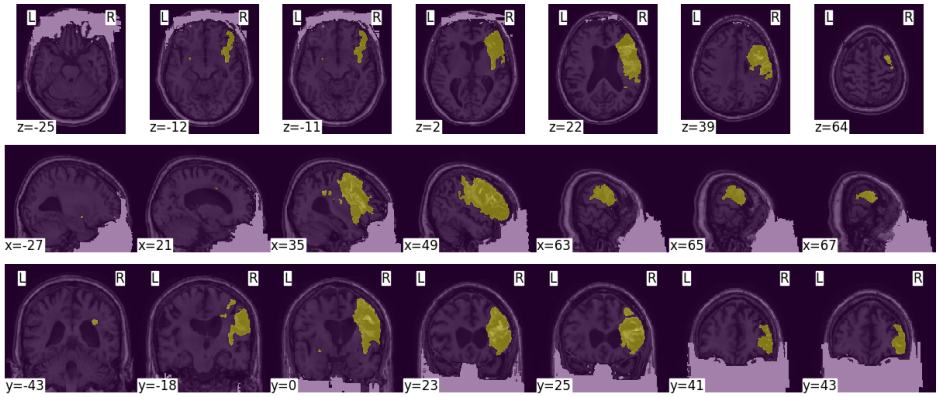


Figure 1. Image with mask of a sample in ATLAS dataset

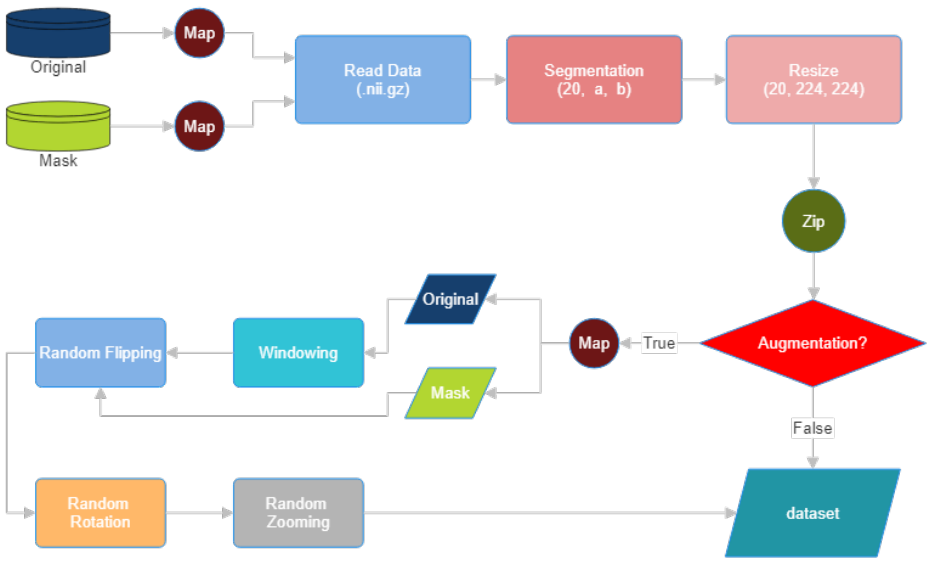


Figure 2. Data Pipeline

padding and strides of 2. The second one maintains the dimensions of the output of the first convolution layer with a 3x3 kernel and a stride of 1. After the second layer, batch normalization and relu activation are sequentially applied. The completion of the above process marks the output of the two convolutional layers. Simultaneously, an identity mapping with a 1×1 kernel is applied to the input data. It unifies the dimensions of the input and the output. The identity mapping reduces the dimensionality, which makes the residual block easier to train. The output of the previous two convolutional layers and the output of the identity mapping are combined using the Add() function. The combined output is eventually applied with batch normalization and a 0.5 dropout layer. The modification of this residual u-net architecture is that every repetition of the basic convolution block is replaced by the residual block described above. The bottleneck and the decoder remain similar to the regular U-Net, as explained in Section 2.3.1.

2.4. Evaluation Metrics

In this work, commonplace precision and sensitivity evaluation metrics such as the Area Under Curve (AUC) are not ideal since the main focus of this task is to determine how well the predicted segmentations overlap with the actual lesion segmentations. Hence, the limit of precision and recall metrics in this scenario is that they focus on how precisely a model finalizes the classification. Instead, other metrics should be considered.

2.4.1 Intersection Over Union (IoU)

In the task of stroke lesion detection, the criteria is to mark how precise a predicted lesion mask is. Mathematically, the

IoU computes the overlap between two datasets. Its equation is defined by the area of intersection over the area of union. IoU is also known as the Jaccard index (or the Jaccard similarity coefficient) which has been widely used to measure the similarity between finite sample sets. Generally, for two finite sample sets A and B, their IoU is defined as the intersection of A and B.

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

The IoU has been widely employed as an evaluation metric for many tasks in computer vision, e.g., pixel- or instance-level image segmentation, 2D/3D object detection etc. [9]. In the work, IoU is used to measure how much the predicted masks overlap with the manually segmented lesion areas. A 1 refers to a perfect overlap, which means the model makes perfect predictions, and a 0 means the segmented area fails.

2.4.2 Dice Coefficient

The Dice Coefficient is another evaluation metric utilized in the work. The dice coefficient, similarly, describes the degree of overlap between two datasets. It is calculated by 2 times the overlap over the total pixels of the two datasets. The Dice Coefficient, D_c, for an overlapping pair of reference and image objects of N_R and N_I pixels, respectively can be written as

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

where N_O is the number of overlap pixels [10]. In this work, the dice loss effectively handles class imbalance since it gives more weight to the correctly classified pixels. Hence, the model gains better accuracy in segmenting relevant features.

2.4.3 Evaluation Metrics Implementation

The functions of IoU and Dice Coefficient are implemented in the work. IoU is passed as one of the metrics during the compilation of both the U-Net and the residual U-Net models for monitoring purposes. Both the IoU and the dice coefficient are used in the final evaluation of the predicted results of the two models. Higher scores of the Dice and the IoU (more overlap between the predicted lesion mask and the actual lesion mask) refer to a more precise predicted lesion area, which indicates better performance of the corresponding model.

2.5. Result

In the work, the result of the two models could be interpreted by evaluative coefficients and visualized displays.

2.5.1 Predictions

For the images of predictions compared with the mask in Figure 3, we observe that the prediction of Residual U-Net captures more details, like the cross-like shape and the sharp bottom. However, the prediction of U-Net only predicts the approximate position and shape.

2.5.2 Training Metrics

From the loss function, we observe that both those 2 models tend to be overfitting. The training loss tends to converge to 0 but validation loss tends to remain at a higher place.

From the AUC, precision, and recall, we observe that both training and validation have a high auc. However, the validation precision is relatively low, meaning that although the model is good at predicting True Positives, it may over-predict some positives.

The visualization of training metrics can be observed in Figure 4.

2.5.3 Evaluation Metrics

For both Dice Coefficient and IoU, Residual U-Net got a higher score than U-Net, meaning that using residual block for feature extraction can boost the model performance.

The evaluation metrics is also shown in Figure 4

3. Conclusion

In conclusion, the integration of residual blocks within the U-Net architecture has demonstrated a marked improvement in the segmentation of lesions in MRI scans from ATLAS R2.0. The comparison based on IoU and Dice Coefficient metrics indicated that the residual U-Net outperformed the basic U-Net. These findings underscore the effectiveness of residual blocks in enhancing the neural network's ability to learn and retain features of complex images. The higher scores in IoU and Dice Coefficient suggest that the residual U-Net is better equipped to handle the intricate details and variability inherent in medical image segmentation, particularly in identifying stroke lesions. Moreover, the visual comparisons between the predicted lesion masks and the actual annotated masks further validate the superiority of the residual-enhanced model. These visual results highlight the practical implications of using residual-enhanced neural network architectures in daily medical diagnostics. The implications of this research extend beyond stroke lesion segmentation, suggesting potential benefits of residual blocks in other medical imaging tasks. On top of this work, future studies may explore the scalability of this approach to other types of medical conditions and imaging modalities, potentially broadening the scope of automatic medical diagnosis systems.

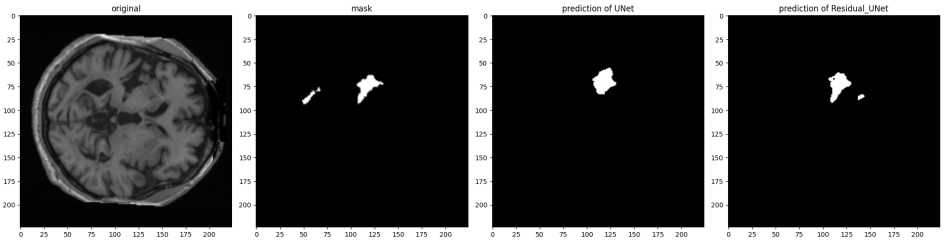
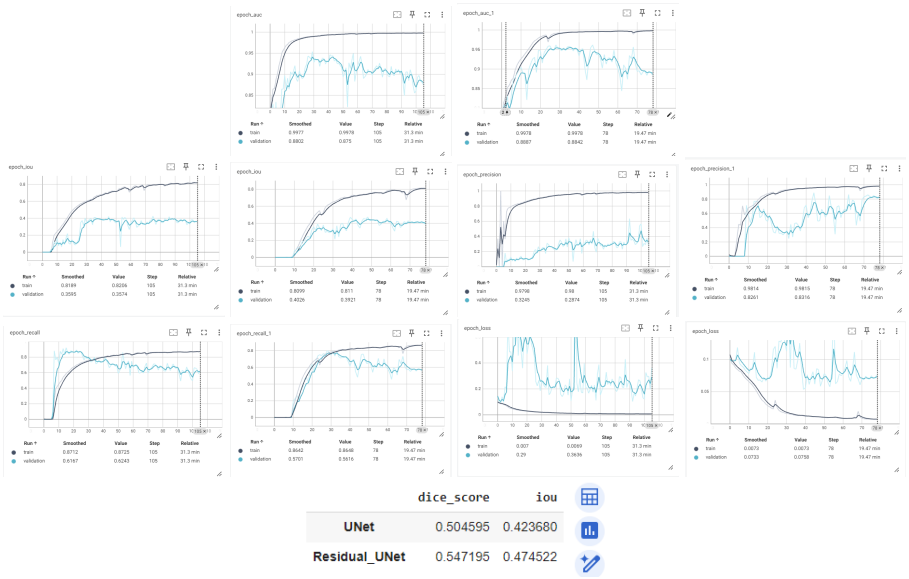


Figure 3. Prediction Output Sample



	dice_score	iou
UNet	0.504595	0.423680
Residual_UNet	0.547195	0.474522

Figure 4. Training (AUC, IoU, Precision, Recall, and loss) & Evaluation (Dice coefficient, IoU) Metrics of U-Net & Residual U-Net

References

[1] Johnson W, Onuma O, Owolabi M, and Sachdev S. Stroke: a global response is needed. *Bull World Health Organ*, 2016. 1

[2] et al. Liew, Sook-Lei. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Scientific data*, 2022. 1, 2

[3] Prantik Deb, Lalith Bharadwaj Baru, Kamalaker Dadi, and Bapi Raju S. Best-les: Benchmarking stroke lesion segmentation using deep supervision. 2023. 1

[4] Shervin Minaee, Yuri Boykov, Antonio Plaza Fatih Porikli, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *arXiv:2001.05566*, 2020. 1

[5] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *arXiv:1411.4038*, 2014. 1

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv:1505.04597v1*, 2015. 1, 2

[7] He K, Zhang X, Ren S, and et al. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. 2

[8] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 2018. 2

[9] Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. Iou loss for 2d/3d object detection. In *2019 International Conference on 3D Vision (3DV)*, pages 85–94, 2019. 4

[10] Bert Guindon and Ying Zhang. Application of the dice coefficient to accuracy assessment of object-based image classification. *Canadian Journal of Remote Sensing*, 43(1):48–61, 2017. 4