Ischemic Stroke Lesion Segmentation with AttentionUnet and MultiResUnet

Siqi Huang sh5688@nyu.edu
Center for Urban Science and Progress
New York University
New York, NY 98195-4322, USA

Kaifu Ren kr2516@nyu.edu
Center for Urban Science and Progress
New York University
New York, NY 98195-4322, USA

Editor:

Abstract

We present two Unet variant deep learning networks for segmenting ischemic stroke lesions in CT perfusion images for ISLES 2018 challenge. It is challenging to identify stroke legions and it requires manual segmentation from MRI DWI images. Automatic segmentation methods using CTP images present the possibility of accurately measuring legions. Our models are based on the Unet, the major architecture for biomedical segmentation tasks. We improved the structure by adding Attention Gate, MultiRes Block and ResPath to learn the varying shapes of the lesions. We trained our networks with weighted cross entropy loss and evaluated with Dice Similarity Coefficient. We compared our models with the baseline Unet. Our approach demonstrates effective performance in the legion segmentation task.

Keywords: Legion segmentation, Unet, MultiResUnet, AttentionUnet

1. Introduction

For the clinicians' decision-making process in acute stroke, identifying locations and extent of irreversibly damaged brain tissue is the critical task. Currently, CT scans have been carried out to classify stroke patients, with their merits in speed, availability, inexpensive, and lack of contraindications (Lev et al., 1999). However, it is challenging to distinguish infarcted tissue and hypoperfused lesion tissue accurately (Jóźwiak et al., 2011). Thus, MRI using diffusion and perfusion imaging techniques were acquired immediately after the CT scans. Patients have to bear the cost and after-effect. Recently, there is a great demand for carrying out deep learning methods over CT perfusion (CTP) images with high sensitivity and specificity (Biesbroek et al., 2013) to perform the accurate measurement of stroke lesions and automate the process to skip the procedure of MRI examinations. To acquire CTP images, the intravenous contrast agent is injected and then repeated scans are made to capture its dynamic flow through the brain (Gillebert et al., 2014). The hypothesis is that through various deep learning experiments over the CTP images, we can achieve high lesion segmentation accuracy compared to the ground truth labelled from MRI images.

2. Background and Related Work

The Ischemic Stroke Lesion Challenge (ISLES) 2018 was a major effort to address the segmentation of stroke lesion based on CT perfusion data. The top five performing participants all used the U-Net variant networks. So our project aims to apply some new major advances in U-Net based architectures since 2018 to see if better results can be achieved. U-Net(Ronneberger et al., 2015) is an encoder decoder architecture with skip connections in between to concatenate high level and low level features. It led to a major improvement in biomedical segmentation tasks where data is limited. Among many advances, Attention U-Net(Oktay et al., 2018) successfully incorporated attention gates as part of the feature concatenation process. They can be trained to suppress irrelevant regions and to highlight salient features. This network consistently leads to better model sensitivity while preserving computational efficiency. Another contribution to the U-Net family is MultiResUNet(Ibtehaz and Rahman, 2019). It replaces the standard U-Net convolutional block with inception-like module that allows learning of features from different scales; it also adds a chain of convolution layers with residual connections to facilitate the fusion of encoder and decoder features. It achieved better performance than U-Net when it came to segmenting images with background visually similar to the regions of interest. In this project we build and test all three architectures mentioned above and evaluated them on CT perfusion dataset.

3. Data

The ISLES challenge 2018 data include CT scan, the source for perfusion images, cerebral blood flow (CBF), cerebral blood volume (CBV), time to peak (TTP) and mean transit time (MTT) of the contrast agent and the label mask drawn based on the corresponding MRI DWI scans (Figure 1). CT perfusion images were taken from acute stroke patients who presented within 8 hours of stroke onset and MRI DWI were taken within 3 hours after CTP. The training set contained 63 patients and 94 cases, and the test set contained 40 patients with 62 cases. Some patients might have two slabs. For each case, the data were three-dimensional volumetric images with varying depth in the axial dimension, ranging from 2 to 22 slices. Each slide was a 256 x 256 image.

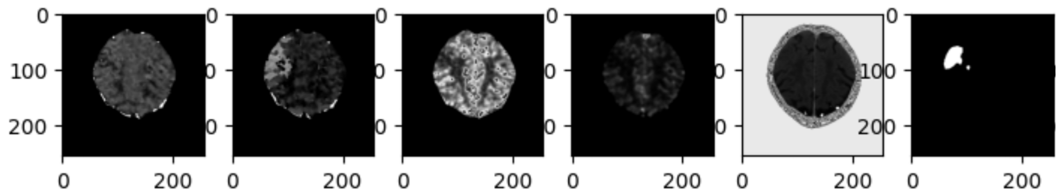


Figure 1: Example CTP data from one case in the study. From left to right: MTT, Tmax, CBV, CBF, CT and label.

4. Materials and Methods

4.1 Data processing

We created the image reader to convert the 3D image to 2D by expanding the axial dimension. Based on the visual effect of Figure 1, the layer of Tmax tends to perform high relevance compared to the ground truth. We decided to use this single source and label mask as inputs for the 2D deep learning models. Then, Tmax will be stacked together with its four corresponding perfusion map slices (MTT, CBF, CBV, CT) for further experiment.

Training set has been randomly splitted into two parts. 70 percent of total were used for training and 30 percent were used for validation. The separation was based on the subjects' number in the original dataset. We ensured that there was no overlap among cases.

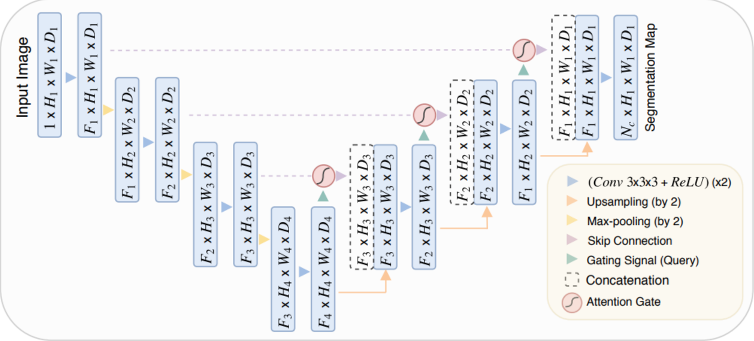
Data augmentation was used to artificially increase the size of the limited training set. The augmentation would randomly flip the images vertically and horizontally, randomly rotate by [-30, 30]. The images also have been normalized to make their pixels within the range of [0, 1].

5. Model

In this project we built and evaluated 3 networks: Unet, Attention-Unet and MultiRes-Unet. The most genuine feature of Unet is probably the skip connection that passes information lost during the max pooling process directly to the decoder side. Attention-Unet builds on this critical step.

5.1 Attention Unet

In image segmentation tasks, when target regions show large inter-patient variation in terms of shape and size, cascaded frameworks are used to extract a region of interest. Attention gate proposed in the paper(Oktay et al., 2018) can learn to automatically focus on the target region and replaces the preceding localization network. This gate can be easily integrated into Unet architecture like in Figure 2.



Input Image
$1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_4 \times W_4 \times D_4$
$F_4 \times H_4 \times W_4 \times D_4$
$F_3 \times H_3 \times W_3 \times D_3$
$F_3 \times H_3 \times W_3 \times D_3$
$F_2 \times H_3 \times W_3 \times D_3$
$F_2 \times H_2 \times W_2 \times D_2$
$F_2 \times H_2 \times W_2 \times D_2$
$F_1 \times H_2 \times W_2 \times D_2$
$F_1 \times H_1 \times W_1 \times D_1$
$F_1 \times H_1 \times W_1 \times D_1$
$N_c \times H_1 \times W_1 \times D_1$
Segmentation Map
($Conv$ 3x3x3 + $ReLU$) (x2)
Upsampling (by 2)
Max-pooling (by 2)
Skip Connection
Gating Signal (Query)
Concatenation
Attention Gate

Figure 2: Attentionn Unet architecture

5.2 Attention gate

Gating signal g passed from coarser level contains contextual information. It together with feature map input x^l to learn an attention coefficient for each pixel. The attention coefficient then multiplies with the feature maps to highlight the salient image region. This "filtering" is done before encoder features are concatenated with decoder features.

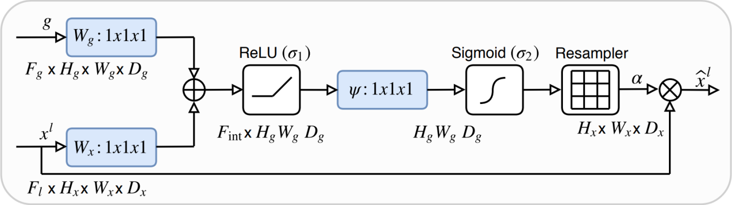


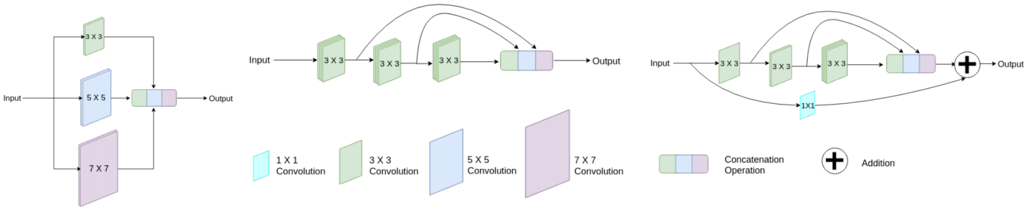
Figure 3: Attention Gate

5.3 MultiRes Unet

Another two problems faced in our dataset as well as in many other segmentation tasks are (1) lack of clear boundary between regions of interest and background, (2) visual similarity between foreground and background due to irregularity and perturbations. Traditional Unet struggle in these two cases. Ibtehaz experimentally showed that two modifications upon Unet architecture: MultiRes Block and ResPath helped to mitigate above two problems(Ibtehaz and Rahman, 2019).

5.4 MultiRes Block

MultiRes Block is a variant of inception block structure as shown in Figure 4. A series of 3x3 convolutions is used to approximate the parallel 3x3, 5x5, 7x7 convolution. The different kernel size allows the network to learn from images at different scales at each level of Unet. This design may have led to the more pixel perfect segmentation result.



(a) (b) (c)

Figure 4: MultiRes Block

5.5 ResPath

ResPath shown in Figure 5 is simply some additional convolution and nonlinearities applied to features passed through skip connections. It is hypothesized that these convolution blocks can reduce the semantic gaps between encoder and decoder feature maps to be merged. The number of convolution layers on the path gradually decreases as we move down the level of Multires-unet because the semantic gap between encoder and decoder feature maps also gradually decreases. From the experiment result, this architecture helps to distinguish the confusing foreground and background.

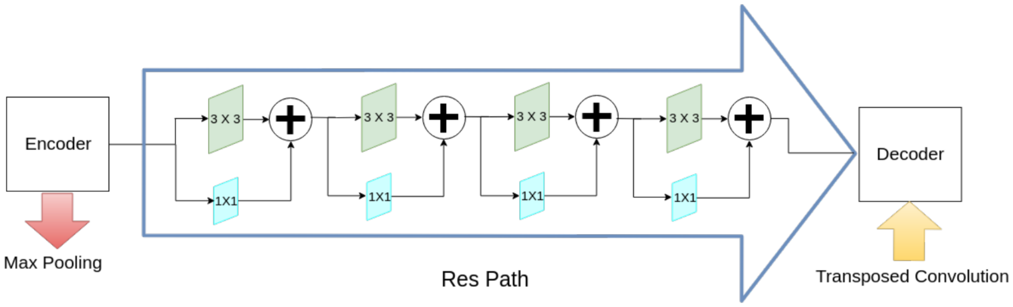


Figure 5: Res Path

6. Training

6.1 Loss Function

Since our labels are imbalanced meaning the size of legion only holding less than 5 percent of the images in most cases, we trained the networks using the weighted cross entropy loss. Given the true label for pixel i, $y_i \in \{0, 1\}$, and a predicted pixel probability (after sigmoid function) $p_i \in \{0, 1\}$, the weighted cross entropy is formulated as

$$WCEL_n = -w_n[y_n log(p_n) + (1 - y_n) log(1 - p_n)]$$

where n is the batch size, and w is the empirical measure of lesion size in the training dataset. After rounds of test and validation, we found the optimal w setting as 20 for this dataset.

6.2 Implementation details

We used Adam as optimizer and StepLR as scheduler to decay the learning rate after several training epochs. We evaluated the networks using the Dice Similarity Coefficient (DSC). Given a predicted image X and the ground truth label Y, the DSC is defined as

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

where $|X|$ and $|Y|$ are the cardinalities of the two sets. The dice coefficient and weighted cross entropy loss on the validation set were monitored for improvement after every training epoch. We trained each model with 30 epochs and performed hyper-parameter tuning to find the optimal setting of parameters including initial learning rate, extent of decays for learning rate. The batch size was set to 4 for all the 2D networks. All models were trained using the PyTorch library.

7. Results

After experiments, we found the practical hyper-parameter for each model in Table 1.

Model	Initial learning rate	Epochs to decay	Decay factor
Unet	0.01	Every 5	0.8
Attention Unet	0.02	Every 10	0.6
MultiRes Unet	0.05	[3,10,15,25]	0.5

Table 1: Optimal Hyper-parameter for All Three Models

Using the above parameters and single channel input (Tmax), we performed a 5-fold cross-validation to determine model performance. The results are shown on Figure 6. MultiRes Unet has overall best performance with highest average DSC (0.467) and lowest average loss (0.000408). Attention Unet doesn't outperform the baseline model (Unet) with holding the lowest average DSC (0.382), but it has the second best average loss. Additionally, we tried to examine all five layers by simply concatenating them as five channels with uniform weighting. The results are shown on Figure 7. There is a slight increase for MultiRes Unet by approximately 3 dice points and a slight decrease for Unet by around 2 dice points. The performance of Attention Unet tends to remain.

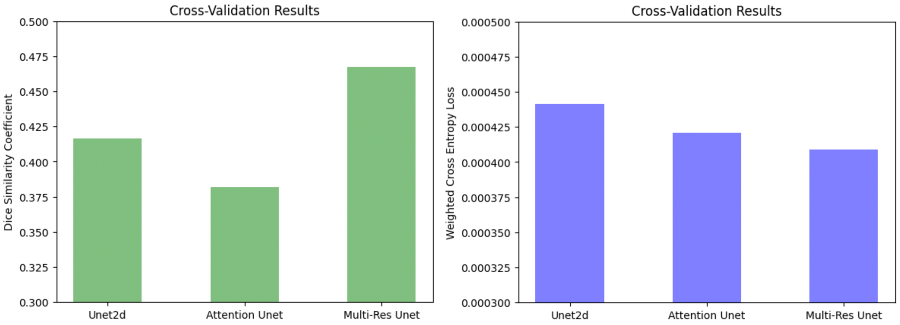


Figure 6: 5-fold cross validation results on each model, using DSC and WCEL and Tmax as input.

8. Discussion

We printed out samples of learning outcomes for each model and did a comparison study (Figure 8). Clearly, the result of Unet tends to be light and shallow. MultiRes Unet can capture the ground truth pretty well, while Attention Unet tends to miss some information during learning. From visual effect, Attention Unet has better results compared to Unet, however, the DSC doesn't prove this. We hypothesize this is due to the fact that during the training stage we didn't find the optimal hyper-parameter for Attention Unet. On other hand, there is no leap in performance when simply increasing model inputs from 1 layer to 5 layers. Our hypothesis is that assigning different weights towards different channels might produce better results, which can be future works. An improvement for the deep

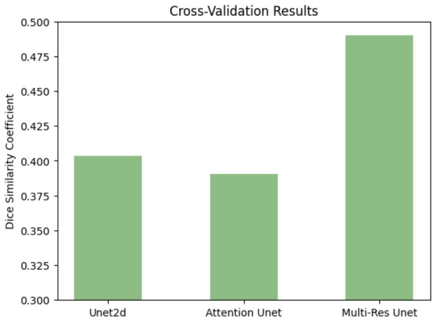


Figure 7: Validation results on each model, using DSC and all five layers as input.

learning models can also be discussed further. It is potential to fuse the two structures - Attention Unet and MultiRes Unet together. Furthermore, 3D architecture can be applied to the dataset and splitted by case numbers directly.

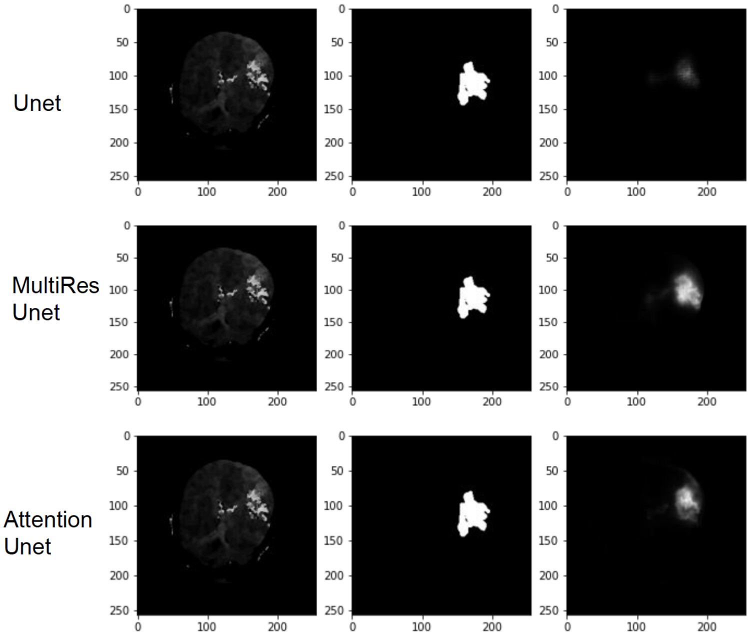


Figure 8: Samples of results. From left to right, Tmax, label and results of respective model.

9. Team Contribution

Kaifu Ren wrote dataloader and Attention-Unet.
Siqi Huang built Unet, MultiRes-Unet and conducted validation.

References

JM Biesbroek, JM Niesten, JW Dankbaar, GJ Biessels, BK Velthuis, JB Reitsma, and IC Van Der Schaaf. Diagnostic accuracy of ct perfusion imaging for detecting acute ischemic stroke: a systematic review and meta-analysis. *Cerebrovascular diseases*, 35(6): 493–501, 2013.

Celine R Gillebert, Glyn W Humphreys, and Dante Mantini. Automated delineation of stroke lesions using brain ct images. *NeuroImage: Clinical*, 4:540–548, 2014.

Nabil Ibtehaz and M. Sohel Rahman. Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation. *CoRR*, abs/1902.04049, 2019. URL `http://arxiv.org/abs/1902.04049`.

Rafał Jóźwiak, Artur Przelaskowski, and Grzegorz Ostrek. Conceptual improvements in computer-aided diagnosis of acute stroke. *Journal of Medical Informatics & Technologies*, 17, 2011.

Michael H Lev, Jeffrey Farkas, Joseph J Gemmete, Syeda T Hossain, George J Hunter, Walter J Koroshetz, and R Gilberto Gonzalez. Acute stroke: improved nonenhanced ct detection—benefits of soft-copy interpretation by using variable window width and center level settings. *Radiology*, 213(1):150–155, 1999.

Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018. URL `http://arxiv.org/abs/1804.03999`.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL `http://arxiv.org/abs/1505.04597`.