# Volumetric Adversarial Training for Ischemic Stroke Lesion Segmentation

Hao-Yu Yang[1,2(✉)]

[1] Cura Cloud Cooperation, Seattle, WA 98104, USA
haoyuy@curacloudcorp.com
[2] Yale University, New Haven, CT 06511, USA

**Abstract.** Ischemic stroke is one of the most common and yet deadly cerebrovascular diseases. Identifying lesion area is an essential step for stroke management and outcome assessment. Currently, manual delineation is the gold standard for clinical diagnosis. However, inter-annotator variances and labor-intensive nature of manual labeling can lead to observer bias or potential disagreement of between annotators. While incorporating a computer-aided diagnosis system may alleviate these issues, other challenges such as highly varying shapes and difficult boundaries in the lesion area make the designing of such system non-trivial. To address these issues, we propose a novel adversarial training paradigm for segmenting ischemic stroke lesion. The training procedure involves the main segmentation network and an auxiliary critique network. The segmentation network is a 3D residual U-net that produces a segmentation mask in each training iteration while critique network enforces high-level constraints on the segmentation network to produce predictions that mimic the ground truth distribution. We applied the proposed model on the 2018 ISLES stroke lesion segmentation challenge dataset and achieved competitive results on the training dataset.

**Keywords:** 3D convolution neural networks · Adversarial training · Ischemic Stroke Lesion Segmentation

## 1 Introduction

Stroke is one of the leading cause of death in developed countries. The disease is caused by either blockage (ischemic stroke) or rupture of a blood vessel (hemorrhagic stroke). Among the two types of stroke, ischemic stroke takes up roughly 80% [1]. The prevailing imaging modalities for diagnosing brain strokes are magnetic resonance imaging (MRI) and computed tomography (CT). Different MRI sequences such as T1 weighted, T2 weighted, Diffusion Weighted Imaging (DWI) and Fluid Attenuated Inversion Recovery (FLAIR) are utilized for specialized applications. DWIs are especially suitable for ischemic strokes since it is highly sensitive to lesion changes [8].

Segmentation of brain area affected by ischemic stroke lesion plays a crucial role in treatment assessment and prognosis. Producing accurate predictions is

challenging due to the variability in the shapes and sizes of the targets. Recent studies [2] have shown that perfusion Computed Tomography (CT) shows potential improvement in speed, availability and lack of contraindications compared to MRI. Computer-aided diagnosis (CAD) system using perfusion CT may help clinicians with faster and more accurate diagnosis. In previous works, models such as random forests, support vector machines and autoencoders [9] have been employed to segment ischemic stroke lesion and have shown successful results.

Computer vision tasks such as image recognition, detection, and segmentation have had significant advances in the past few years due to the rise of deep learning, specifically in Convolution Neural Networks (CNN). Medical applications of deep learning have also seen profound successes. As neural networks tend to get deeper as we harness more computational power, the problem of vanishing gradients problem ensues. Vanishing gradients occur when gradients become too small to change the weights of the neuron in back-propagation-trained neural networks. The residual learning networks [4] (ResNet) solves this problem by introducing stacked identity mappings in the form of residual blocks. These residual connections allow the neural network to collapse into a few layers during initialization and gradually expand in the feature space as training takes place. Recently, generative adversarial networks (GAN) [3] have been utilized extensively throughout image generation tasks. Recent studies [6,10] have shown that GANs can also be used in a critique framework for semantic segmentation tasks. The benefits of using such networks include comparing the higher level of inconsistencies between ground truth and predictions and enforcing spatial continuity. In this framework, generating pixel-wise segmentation masks are modeled as a generative procedure and the discriminator of the model attempts to distinguish between real and fake segmentation masks.

In this paper, we've developed a neural-network with adversarial training to segment irreversibly damaged brain area caused by ischemic stroke. The proposed model is trained and validated on the Ischemic Stroke Lesion Segmentation(ISLES) challenge dataset [7]. The ISLES challenge aims at providing a unified platform and high-quality data for training and evaluating models for automatic stroke lesion segmentation. In order to model the variability in the true distribution and improve prediction accuracy, adversarial training. For preprocessing, each modality is normalized and stacked as multi-channel inputs. The overall loss function consists of three terms: negative dice coefficient and binary cross-entropy between the ground truth mask and prediction plus the discriminator loss between real and generated segmentation masks. Our method produced promising results and achieved an average DICE coefficient of 0.87 on the ISLES training dataset.

## 2    Method

The detailed model architecture and training procedure of the proposed methods are described in this section. First, we address the necessary steps for preprocessing the data. Then we introduce the architecture of the segmentation network.

Finally, we illustrate two adversarial paradigms proposed for training the segmentation model.

## 2.1   Data

We performed training and validation on the 2018 ISLES challenge dataset. The training dataset contained a total of 63 patients each with 5 different perfusion maps: cerebral blood flow (CBF), Mean transit time (MTT), cerebral blood volume (CBV), time to peak of residue function (TMAX) and computed tomography angiography (CTP). An example of the training data can be found in Fig. 1. The training data also included gold standard diffusion-weighted imaging (DWI) maps that are not available in testing data. The ground truth segmentation masks were derived from the DWI. The data provided are in Neuroimaging Informatics Technology Initiative (NIfTI) format. We used Insight Segmentation and Registration Toolkit (ITK) [12] for data inspection and visualization.
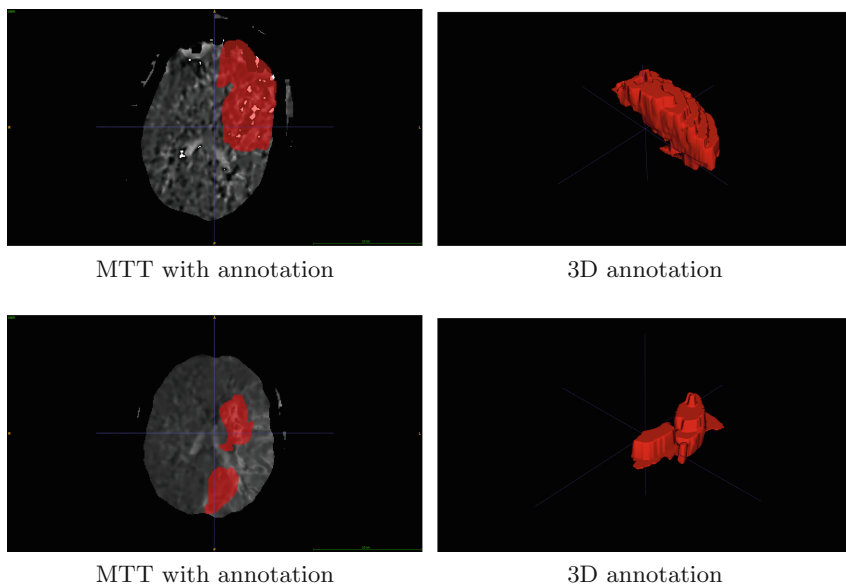


|                        |                |
|------------------------|----------------|
| MTT with annotation    | 3D annotation  |
| MTT with annotation    | 3D annotation  |

**Fig. 1.** Example of training data and corresponding 3-D annotation

## 2.2   Preprocess

Preprocessing is necessary due to the significant cross-modality variance. There are also substantial deviation in the spatial resolution as dimension of the z-axis ranges from 2 to 16 for different subjects. First, we conducted bicubic spline interpolation [5] to resize each volume to the same dimension. During training

and testing, each modality is then normalized respectively by subtracting the mean intensity and divide by the standard deviation as shown in the following equation:

$$x'_m(i,j,k) = \frac{x_m(i,j,k) - \mu_m}{\sigma_m} \tag{1}$$

Where m denotes the modality, $\mu_m$ denotes mean intensity and $\sigma_m$ denotes the standard deviation. i, j, k is the coordinates of the pixel to be normalized. Finally, the normalized whole volume are stacked as multi-channel inputs for the segmentation network.

### 2.3    3D Residual U-Net

The backbone of the segmentation network is a 3D U-net with residual connection [11]. Network structure and details of the residual block of the can be found in Fig. 2 The U-net consists of both down-sampling and up-sampling pathways. The down-sampling pathway is made up 4 residual blocks and the upsampling path contained 4 transposed convolution blocks. Each residual block contains three $3 \times 3 \times 3$ convolution layer, batch normalization and activation function with leaky reciftied liner unit in between. The up-sampling pathway contained 4 transpose convolution operation and concatenation with corresponding feature maps from the down-sampling pathway.

### 2.4    Adversarial Training

The adversarial pipeline is a two-player mini-max game between the segmentation network and the discriminator network. Figure 3 shows the high-level view of the training procedure. In each training iteration, the segmentation network will generate a pixel-wise probability map which is then fed to the discriminator network as inputs. The objective of the discriminator is to distinguish between ground truth segmentation mask and predicted mask. The discriminator is a 7-block network containing 3 residual blocks similar to the 3D U-net. Maxpooling was conducted after every residual block. The discriminator network is solely for auxiliary purposes and therefore removed during testing phase. The discriminator network enforces spatial continuity that is otherwise not obtainable by using only pixel-wise classification loss.

We denote the ground truth mask as $y$, image data as $x$, U-net as $U$ and discriminator as $D$, the solution to the mini-max game can, therefore, be written as:

$$\min_U \max_D E_{y \sim p(y)}[\log D(y)] + E_{x \sim p(x)}[\log(1 - D(U(x)))] \tag{2}$$

There are different ways that the adversarial training can be carried out. We proposed two training paradigm for the adversarial pipeline, namely:

– Integrated loss
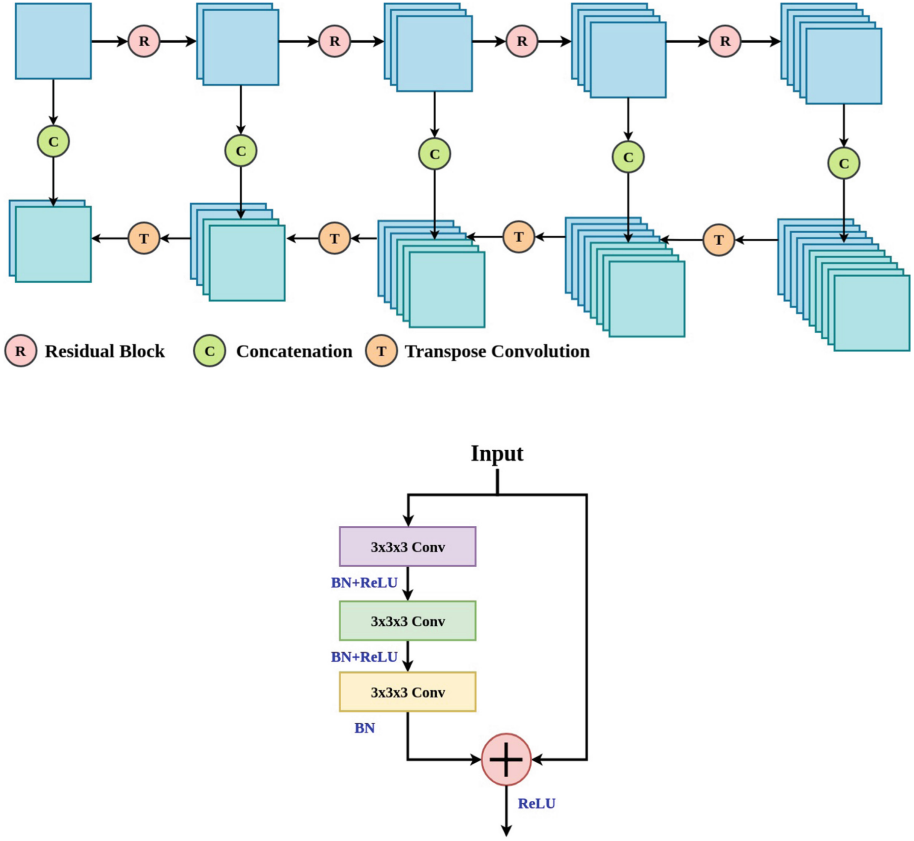– Second back propagation

**Fig. 2.** Top figure shows the architecture of the 3-D residual U-net. Bottom figure shows a single residual block

**Integrated Loss.** The integrated loss paradigm adds the adversarial loss to the traditional segmentation losses and forms a integrated loss term. Back-propagation are carried out based on the gradients of the integrated loss term. The discriminator network are back-propagated by the errors of not recognizing true label and misclassifying synthetic label as true. The integrated loss function for the segmentation network contained a total of three terms: binary cross entropy loss, negative dice score and adversarial loss as seen in the following equation:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{adver} + \beta\mathcal{L}_{BCE} + \gamma\mathcal{L}_{dice} \tag{3}$$

Where the $\alpha, \beta, \gamma$ are coefficients for each loss terms. We initialized all three coefficients as 1. The coefficients are adjusted by weight decay mechanism which we describe in implementation details section. Detailed algorithm can be found in Algorithm 1.
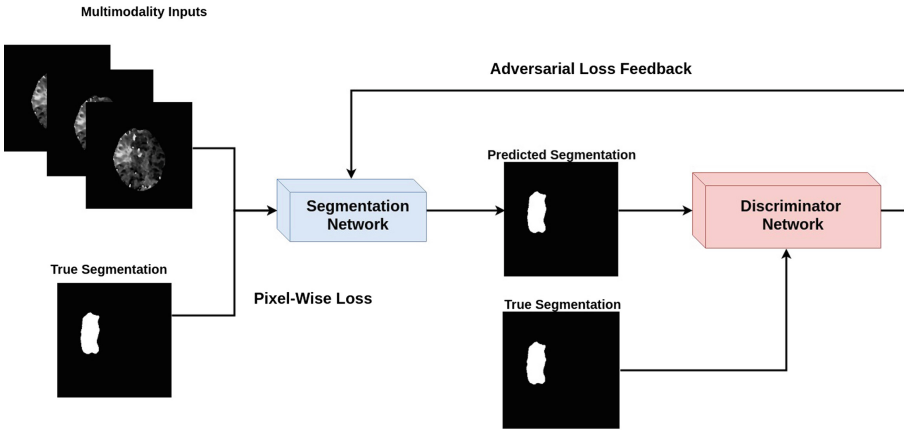
**Fig. 3.** Overview of the adversarial training paradigm

**Second Back-Propagation.** In the second back-propagation pardigram, the segmentation network is back-propagated twice. First, the weights are adjusted according to the gradients of the traditional segmentation loss. At the adversarial training phases of each iteration, gradients from the adversarial loss are then passed onto the segmentation network for a second back-propagation. The discrimination network is back-propagated only once. Detailed training algorithm can be found in Algorithm 2.

---

**Algorithm 1.** Training procedure with integrated loss function

---

**Require:** Initialize parameters of U-net $\theta_U$, Discriminator $\theta_D$
**Data:** Sample images $x_1 \ldots x_n$ and corresponding masks $y_1 \ldots y_n$ from data
Initialize parameters of U-net $\theta_U$, Discriminator $\theta_D$
**while** *While $\theta$ not converged* **do**
    **for** *Each $x_i$ and $y_i$ in mini-batch* **do**
        Compute mask $\hat{y}_i$ by $U(\hat{y}_i|x_i)$
        Compute $\mathcal{L}_{Dice}(\hat{y}_i, y_i)$, $\mathcal{L}_{BCE}(\hat{y}_i, y_i)$
        Compute $\mathcal{L}_{adver}$ by $\log(1 - D(\hat{y}_i))$
        Compute $\mathcal{L}_{total}$ by $\mathcal{L}_{adver} + \lambda\mathcal{L}_{BCE} + \mathcal{L}_{dice}$
        $\theta_D \leftarrow \nabla[\log D(y_i) + \log(1 - D(\hat{y}_i))]$
        $\theta_U \leftarrow \nabla\mathcal{L}_{total}$
    **end**
**end**

---

---

**Algorithm 2.** Training procedure with second back-propagation

---

**Require:** Initialize parameters of U-net $\theta_U$, Discriminator $\theta_D$
**Data:** Sample images $x_1 \ldots x_n$ and corresponding masks $y_1 \ldots y_n$ from data
**while** *While $\theta$ not converged* **do**
    **for** *Each $x_i$ and $y_i$ in mini-batch* **do**
        Compute mask $\hat{y_i}$ by $U(\hat{y_i}|x_i)$
        Compute $\mathcal{L}_{Dice}(\hat{y_i}, y_i)$
        $\theta_U \leftarrow \nabla \mathcal{L}_{Dice}$
    **end**
    **for** *Each $x_i$ and $y_i$ in mini-batch* **do**
        Compute mask $\hat{y_i}$ by $U(\hat{y_i}|x_i)$
        $\theta_D \leftarrow \nabla[\log D(y_i) + \log(1 - D(\hat{y_i}))]$
        $\theta_U \leftarrow \nabla \log(1 - D(\hat{y_i}))$
    **end**
**end**

---

### 2.5 Implementation Detail

The proposed model was established with python under the pytorch deep learning framework. The learning rate was set differently for the segmentation network and the critique network to avoid collapsing in early epochs, which is a common phenomenon in GANs. Learning rates were initialized at 0.001 for the segmentation model and 0.0005 for the discriminator network. Learning rate decay will take place if there were no improvements of the loss function 5 consecutive epochs. Each learning rate decay reduces the learning rate to 80% of the previous iteration. Early termination will take place if no improvements were seen for 20 consecutive epochs. The mini-batch size was set at 8. GPU training was conducted on 4 NVIDIA Tesla V100. The total training time for the entire pipeline that included segmentation network and discrimination network was approximately 24 h.

## 3 Results

In this section, we present quantitative results of the proposed model and qualitative comparison of the adversarial training effects. Several matrices including the mean of Dice score, the standard deviation of Dice score, mean of Hausdorff distance and standard deviation of Hausdorff distance were used for model evaluation. Figure 4 is a visualization of the adversarial training effects. As shown in the figure, models with adversarial training are able to capture subtle differences between ground truth and predictions. Table 1 shows that by incorporating adversarial training, dice score increased and Hausdorff distance reduced.
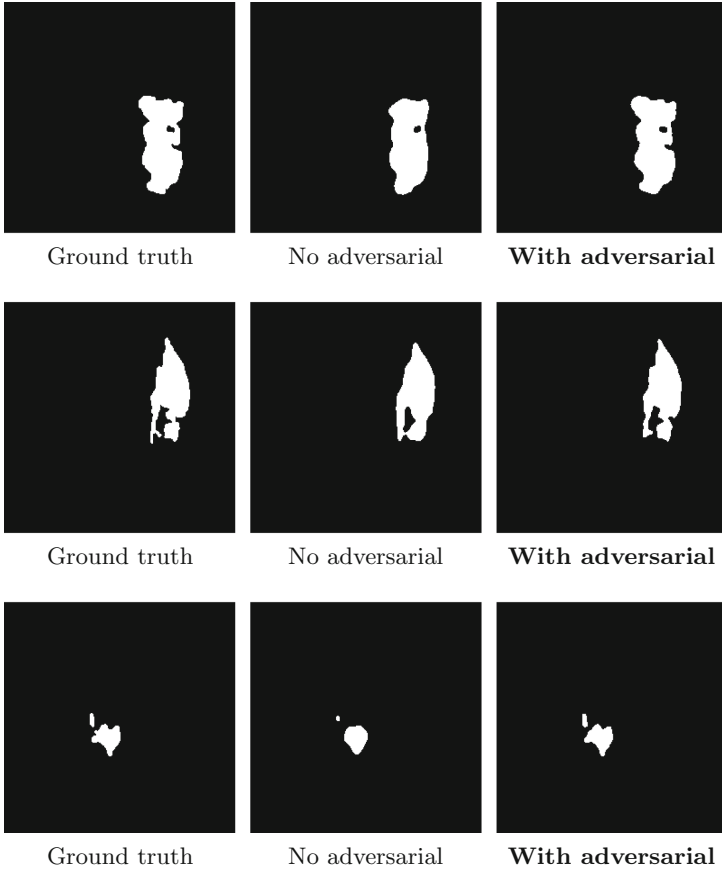
| Ground truth | No adversarial | **With adversarial** |

**Fig. 4.** Effects of adversarial training

**Table 1.** Dice and Hausdorff distance comparison between three training paradigm

|  | DICE mean | DICE std | Hausdorff mean | Hausdorff std |
|---|---|---|---|---|
| Without adversarial | 0.78 | 0.19 | 27.45 | **31.23** |
| With second back-propagation | 0.85 | **0.16** | 20.78 | 33.81 |
| **With integrated adversarial loss** | **0.87** | 0.21 | **18.89** | 37.32 |

## 4 Discussion

In this paper, we've presented an automatic ischemic stroke lesion segmentation model using multiple CT perfusion maps with varying dimensions as inputs. We proposed two adversarial training paradigm, namely integrated loss function and second back-propagation. We've demonstrated that by incorporating a discriminator network in the training procedure, the segmentation model is

able to mimic subtle inconsistencies between ground truth and prediction that cannot be corrected using only pixel-wise loss functions such as binary cross entropy and dice score. Quantitatively, employing adversarial training increases dice score and reduces Hausdorff distance.

# References

1. Feigin, V.L., Lawes, C.M., Bennett, D.A., Anderson, C.S.: Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurol. **2**(1), 43–53 (2003). https://doi.org/10.1016/S1474-4422(03)00266-7. http://www.sciencedirect.com/science/article/pii/S1474442203002667

2. Gillebert, C.R., Humphreys, G.W., Mantini, D.: Automated delineation of stroke lesions using brain ct images. NeuroImage: Clin. **4**, 540–548 (2014). https://doi.org/10.1016/j.nicl.2014.03.009. http://www.sciencedirect.com/science/article/pii/S2213158214000394

3. Goodfellow, I.J., et al.: Generative Adversarial Networks. arXiv e-prints, June 2014

4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, June 2016. https://doi.org/10.1109/CVPR.2016.90

5. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. **29**(6), 1153–1160 (1981). https://doi.org/10.1109/TASSP.1981.1163711

6. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. CoRR abs/1611.08408 (2016). http://arxiv.org/abs/1611.08408

7. Maier, O., et al.: ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. **35**, 250–269 (2017). https://doi.org/10.1016/j.media.2016.07.009. http://www.sciencedirect.com/science/article/pii/S1361841516301268

8. Moseley, M.E., et al.: Early detection of regional cerebral ischemia in cats: comparison of diffusion- and T2-weighted MRI and spectroscopy. Magn. Reson. Med. **14**(2), 330–346 (1990)

9. Praveen, G., Agrawal, A., Sundaram, P., Sardesai, S.: Ischemic stroke lesion segmentation using stacked sparse autoencoder. Comput. Biol. Med. **99**, 38–52 (2018). https://doi.org/10.1016/j.compbiomed.2018.05.027. http://www.sciencedirect.com/science/article/pii/S0010482518301409

10. Quan, T.M., Hildebrand, D.G.C., Jeong, W.: FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics. CoRR abs/1612.05360 (2016). http://arxiv.org/abs/1612.05360

11. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015). http://arxiv.org/abs/1505.04597

12. Yushkevich, P.A., et al.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)