

Deep Learning for Hemorrhagic Lesion Detection and Segmentation on Brain CT Images

Lu Li¹, Meng Wei¹, Bo Liu¹, Kunakorn Atchaneyasakul, Fugen Zhou¹, Zehao Pan, Shimran A.Kumar¹, Jason Y. Zhang, Yuehua Pu, David S. Liebeskind, and Fabien Scalzo¹

Abstract—Stroke is an acute cerebral vascular disease that is likely to cause long-term disabilities and death. Immediate emergency care with accurate diagnosis of computed tomographic (CT) images is crucial for dealing with a hemorrhagic stroke. However, due to the high variability of a stroke's location, contrast, and shape, it is challenging and time-consuming even for experienced radiologists to locate them. In this paper, we propose a U-net based deep learning framework to automatically detect and segment hemorrhage strokes in CT brain images. The input of the network is built by concatenating the flipped image with the original CT slice which introduces symmetry constraints of the brain images into the proposed model. This enhances the contrast between hemorrhagic area and normal brain tissue. Various Deep Learning topologies are compared by varying the layers, batch normalization, dilation rates, and pre-train models. This could increase the respective field and preserves more information on lesion characteristics. Besides, the adversarial training is also adopted in the proposed network to improve the accuracy of the segmentation. The proposed model is trained and evaluated

Manuscript received April 1, 2020; revised August 18, 2020; accepted September 26, 2020. Date of publication October 1, 2020; date of current version May 11, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61806015, 61772054, National Key R&D Program of China under Grant 2018YFA0704101. (*Corresponding authors:* Lu Li; Bo Liu.)

Lu Li is with the Image processing center, Beihang University, Beijing 100083, China (e-mail: liliu@buaa.edu.cn).

Meng Wei is with the Department of electrical and electronic engineering, University of Liverpool, Liverpool, Merseyside L3 5TR, U.K. (e-mail: m.wei14@student.liverpool.ac.uk).

Bo Liu and Fugen Zhou are with the Image processing center, Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China (e-mail: bo.liu@buaa.edu.cn; zhfugen@buaa.edu.cn).

Kunakorn Atchaneyasakul, Shimran A.Kumar, and Jason Y. Zhang are with the UCLA Department of Neurology, CA 90024 USA (e-mail: k.atchanee@gmail.com; shimrankumar60@gmail.com; jasonyzhang@gmail.com).

Zehao Pan is with the University of the Pacific, Stockton, CA 95211 USA (e-mail: r_pan@u.pacific.edu).

Yuehua Pu is with the Neurointensive Care Unit, Department of Neurology, Beijing Tiantan Hospital of Capital Medical University, Beijing 100050, China (e-mail: sophiepu@me.com).

David S. Liebeskind is with the Director of UCLA Department of Neurology, the Director of Neurovascular Imaging Research Core, CA 90024 USA (e-mail: davidliebeskind@yahoo.com).

Fabien Scalzo is with the UCLA Department of Computer Science, UCLA Department of Neurology, Neurovascular Imaging Research Core, CA 90024 USA (e-mail: fabien.scalzo@gmail.com).

Digital Object Identifier 10.1109/JBHI.2020.3028243

on two different datasets, which achieve the competitive performance with human experts with the highest location accuracy 0.9859 for detection, 0.8033 Dice score, and 0.6919 IoU for segmentation. The results demonstrate the effectiveness, robustness, and advantages of the proposed deep learning model in automatically hemorrhage lesion diagnosis, which make it possible to be a clinical decision support tool in stroke diagnosis.

Index Terms—Deep learning, hemorrhage, stroke, automatic diagnosis, segmentation.

I. INTRODUCTION

STROKE is a disease caused by the decrease of blood flow to the brain. It is the second most common cause of death and a leading cause of disability in the world [1]. A stroke occurs when the blood supply to part of the brain is suddenly interrupted (ischemic stroke) or when a blood vessel in the brain bursts and spills blood into the spaces surrounding brain cells (hemorrhage stroke). Even though hemorrhagic strokes make up about 10-30% of all stroke cases, it is much more serious because of the high mortality rate at the range of 35-52%. Computed Tomography (CT) images are widely used to diagnose brain strokes due to wider availability, lower cost, and sensitivity to early strokes signs. An accurate and immediate diagnosis by an experienced neuro-radiologist is crucial for dealing with a hemorrhagic stroke. The treatment is significantly different between ischemic and hemorrhagic strokes, but the initial clinical findings can be very similar. It is imperative to have the highest accuracy diagnosis via advanced imaging, such as CT or MRI scans, to help identify if a hemorrhagic area exists and where the hemorrhaging regions are. Accurate segmentation of the hemorrhagic region is also important to plan clinical treatment steps. It can offer better interpretability and quantifiable metrics for disease prognosis instead of classification only [2], [3]. However, skilled and experienced neuro-radiologists are in short supply, especially in resource-limited countries. Apart from the large scale of spatial and structural variability in each scan, there are also noise and complexity of spatial features in the low contrast between hemorrhage and normal regions. The hemorrhage region in CT brain images can be difficult to identify even for an experienced radiologist.

In the last decades, lots of researchers have aimed to develop image processing algorithms and methodologies to enable an easy and faultless identification of abnormalities present in the scanned region [4]. T. Gong *et al.* [5] proposed a method to find potential hemorrhage regions in CT images by using ellipse fitting, background removal, and wavelet decomposition techniques. T. Chan [6] used thresholding and morphological operations to segment candidate regions and then extract candidate regions based on top-hat transformation and left-right asymmetry. R. Liu *et al.* [7] adopted wavelet-based texture analysis to eradicate all the nasal cavity slices followed by intensity-based thresholding to identify the stroke-affected regions. Kyaw *et al.* [4] introduced an automated method for the detection and classification of an abnormality (hemorrhage) or stroke in brain CT images, in which pre-processing was used to remove film artifacts and skull region. Ali *et al.* [9] propose Segmentation of Hemorrhage Stroke using Textural Analysis on Brain CT Images. A more comprehensive review of other methods for hemorrhage analysis can be found in [8].

Recently, as the developing of Deep Learning based methods becoming popular in many field, Deep CNNs have been successfully used in medical research for image segmentation and computer-aided diagnosis [10]. In contrast to previous approaches of segmentation of hemorrhage regions in CT images that rely on the development of hand-crafted features, deep CNNs learn increasingly complex features from data automatically. Thus, many CNN models, both 2D and 3D, were introduced into automatic stroke diagnosis. For the brain CT image, there are many different slices for each patient. Thus, many researchers introduced 3D networks into CT image segmentation and tried to use the spatial information of the different slices. Grewal *et al.* [11] presented a 3D deep learning-based approach RADnet that emulates radiologists' method for diagnosis of the brain hemorrhage from CT scans. Jnawali *et al.* [12] implemented three different 3D CNN architectures for the detection of brain hemorrhages in head CT images. However, due to the high computational cost and GPU memory consumption, the depth of the 3D network is limited compared to that of a 2D network. This makes the 3D network impractical in stroke diagnosis [13]. Besides, due the high complexity of 3D networks, it is even more difficult to get an adjustable parameter for the ideal model to get the best performance. As a result, the 3D network might be not as efficient and flexible as the corresponding 2D networks, which is also verified in [14].

As for 2D networks, Liu *et al.* [15] proposed a residual-structured, fully convolutional network (Res-FCN) to get automated Stroke Lesion Segmentation based on Multi-Spectral MR Images. Kuang *et al.* [16] proposed a joint segmentation approach to segment ischemic and hemorrhage infarct simultaneously, in which the semantic information, local image context, and high-level user initialized prior are integrated into a multi-region time-implicit contour evolution scheme. Abulnaga *et al.* [14] used PSPNet, a network architecture that makes use of pyramid pooling to provide global and local contextual information to get ischemic stroke segmentation. However, most of the recent deep learning-based methods are aiming to detect or segment the ischemic stroke. There is less work done on

hemorrhage stroke semantic segmentation, and, to our knowledge, no semantic segmentation approach has shown competitive performance compared with human experts. There is still a lot of work to be done to get the accurate hemorrhage segmentation based on CT brain images.

In this paper, we proposed a U-net based deep learning framework to automatically detect and segment hemorrhage strokes in CT brain images. Then we compare its performance against the reference standard of expert manual segmentation. Different U-net based CNN architectures were implemented by changing the input of the network, adjusting the structure of different models, and trying different training processes. The main contributions of this paper are as follows:

- 1) Input of the network: Instead of using the original CT images, this paper concatenates the flipped image with the original CT slice as input of the network, which validate the effective of symmetric structure in contrast enhancement. Aside from data argumentation, data balance was also discussed in the proposed model.
- 2) Structure of the model: Based on the most state-of-the-art UNet model, various kinds of modified architectures with different layers, batch normalization, dilation rates and pre-train models were compared and evaluated, which provide a solution on how to increase the respective filed and preserves more information of lesion characteristic.
- 3) Training process: To reduce the effect of the normal area in the brain image, the conditional generative adversarial nets (GAN) were modified by using UNet as a generator. Then the discriminator can be used to exploit higher-order inconsistencies in the samples synthesized by the generator. Patch training and the activation layer of LeakyReLU are used in the training process to improve the performance in the small lesion part of the image.

II. METHODOLOGY

In this section we present our proposed framework in detail. We first introduce the input of the neural network that will be applied in our model, and it is followed by introducing the data symmetry and data augmentation. Then we elaborate on the structure of the model by the UNet and how the pre-trained model and dilated convolution are utilized for more discriminative and robust representations. Finally, the adversarial training is presented to enhance the performance of the model. The flowchart of the whole framework is illustrated in Figure. 1

A. Input of the Network

Data Symmetry: As the contents of the cranium are normally remarkably symmetrical on the axial CT, brain Quasi-symmetry property is the one of critical priors for doctors to identify lesions [17]. The traditional method did not consider this special symmetrical character processing segmentation with the contents of a small neighborhood patch (unilateral descriptor), which is not consistent with the actual gold standard labeling. Lesions rarely occur simultaneously in symmetrical areas, so it is natural to compare the symmetrical areas of the image to find the difference. When the stroke lesion occurs, the symmetrical

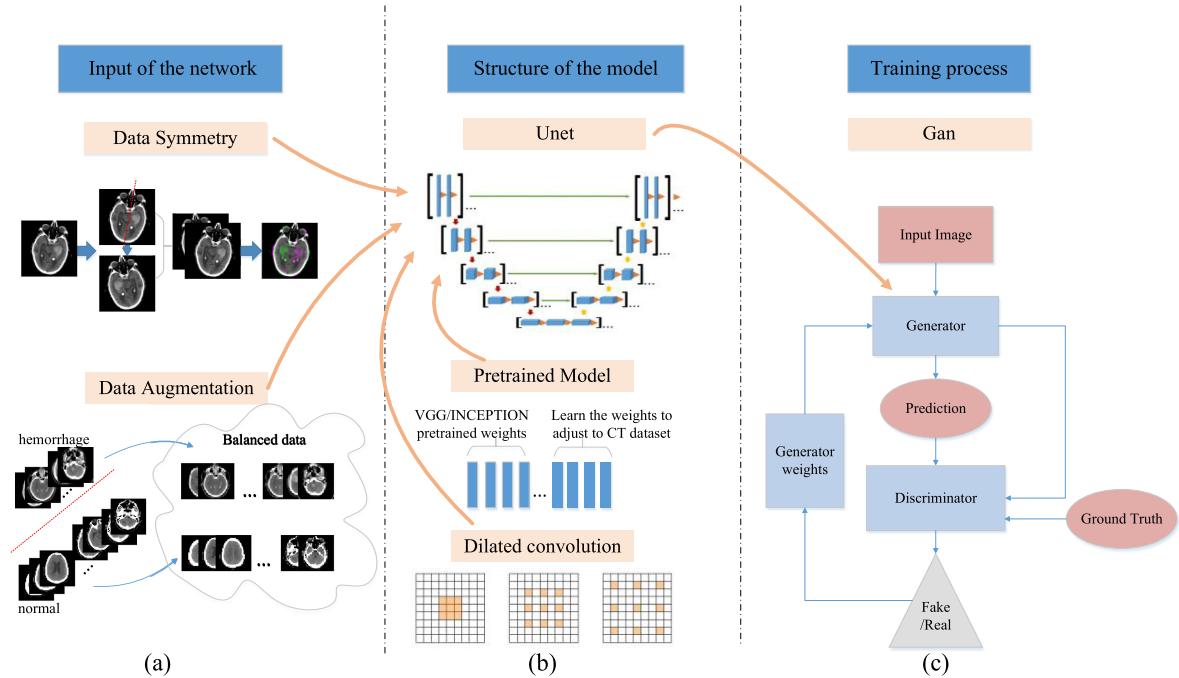


Fig. 1. Pipeline of proposed deep learning-based model for CT brain hemorrhage detection and segmentation: (a) Input of the network: Instead of using the original CT images, we concatenate the flipped image with the original CT slice as input of the network. Also, data argumentation was used to obtain balanced training data. (b) Structure of the model: to increase the respective field and preserves more information of lesion characteristics, various kind of modified UNet architectures with different layers, batch normalization, dilation convolution and pre-train models were adopted and discussed. (c) Training process: The Gan network structure was introduced into the proposed model, which train a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation results either come from the ground truth or the segmentation network.

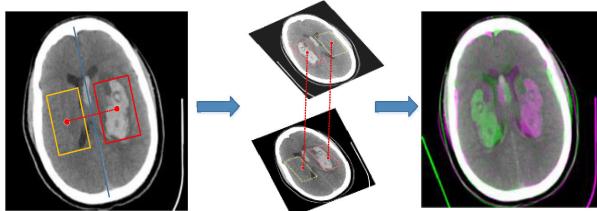


Fig. 2. Symmetrical property of CT brain image: Lesions rarely occur simultaneously in symmetrical areas, by flipping the original image with the axis of symmetry, we can combine the flipped image with original one as input, so that to compare the symmetrical areas of the image to find the difference).

property of CT brain image will decrease. As shown in Figure 2, the blue line is the axis of symmetry for the brain CT image. The hemorrhage region (areas in the red bounding box) presents high contrast not only with the surrounding areas but also with its symmetrical areas (areas in the yellow dotted bounding box). That's the rule the doctor uses to identify the hemorrhage lesions.

Different from [17], which extracts two symmetrical patches to describe one voxel and feeds them into the network at the same time. Here we flip the original image with left-right direction and concatenate them together as the input of the network, as can be seen in Figure 1. Thus, for each patch, the convolution is based on the original region and its symmetrical part at the same time. Through encoder-decoder structure and iteratively learning, the

network can learn how to express the difference for each pair of images.

However, the symmetrical line of CT images are not always vertical and the image is not axisymmetric. It is necessary to make image alignment before the flip rotation and concatenate to make sure the symmetrical areas can be considered at the same time. In this paper, we encode brain symmetry information to the UNet-based deep learning model by adding extra flipped and aligned symmetry maps. Our symmetry maps are computed as follows: 1) Produce the flipped image for each original brain image; 2) Find the corresponding matching pixel pairs for the flipped image and the original image; 3) Align the flipped image with rigid transformation (translation and rotation) in corresponding original image basically; 4) Concatenate the flipped image with the original one as 2 channel images as input.

Data Augmentation: Due to the various sizes and irregular shapes of the lesion areas, it's a challenging and time-consuming task to segment hemorrhage regions from CT images accurately. Lacking enough training data is one of the critical reasons for the poor effect of network training. Data augmentation is an essential part of training Deep Learning models, especially for medical images. A variety of augmentation strategies, including horizontal flips, random crops, and principal component analysis (PCA), have been proposed and shown to capture important characteristics of images.

The data imbalance is also a problem that influences the training efficiency of the network. In medical images, the lesion

areas and the slices with lesions are only a small part of the data. For example, for a patient with 24 slice brain CT images, there might be hemorrhage regions existing in 4 slices. So the amount of positive and negative samples are extremely imbalanced, which might introduce bias towards the majority (potentially uninteresting) class while developing a classification model. Based on that, we divide the training data as hemorrhage and normal, then use horizontal flip, random crop, and rotation to augment the two datasets differently. This ensures that there are equivalent amounts of training data for hemorrhage and normal.

B. Network Architectures

U-Net: U-Net is a type of CNN architecture which was first proposed in 2015 [18]. From then on, it has been successfully applied in many different kinds of biomedical image segmentation problems [11], [13], [19]. Its competitive performance was based on the two distinguished characters of U-Net: the U-shape structure and skip connection. U-shape structure refers to its symmetric shape, which is comprised of two parts: the encoding part, where feature maps are down-sampled by max-pooling layers, and the decoding part, where the reduced size of feature maps are up-sampled to the original size [20]. The skip connection between the down-sampling path and the up-sampling path apply a concatenation operator instead of sum, which makes the spatial information directly applied to much deeper layers and get a more accurate segmentation result.

In this paper, we adopt U-Net as our basic segmentation model, but instead of using 10 total convolutional blocks (one convolutional block contains two or three convolutional layers with 3×3 kernels and 1 max-pooling layer with 2×2 strides), we use 6 convolutional blocks based on the following considerations. First, medical images are mostly based on low-level features, which can be abstracted on the shallow layer of the network. Second, there are both big hemorrhage lesions and very small hemorrhage lesions in CT brain images. If the network goes deeper, it may cause some tiny lesion to vanish in the later convolutional blocks [21]. Third, fewer convolutional blocks will reduce the computational burden of the model and make it easier convergence.

Batch Normalization [22] is employed to accelerate the training process, and Rectified Linear Unit (ReLU) [23] activation function is used after each convolutional layer to improve the nonlinearity.

Dilated convolution: Dilated convolutions [24] are a special form of standard convolutions in which the effective receptive field of kernels is increased by inserting zeros (or holes) between each pixel in the convolutional kernel. Dilated CNNs are therefore an effective approach to increasing the accuracy of segmentation through a large receptive field with a limited number of trainable weights and a limited number of convolutional layers [18].

As described in [24], let $F : \mathbb{Z}^2 \rightarrow \mathbb{R}$ be a discrete function, and $\Omega_r = [-r, r]^2 \cap \mathbb{Z}^2$. Let $k : \Omega_r \rightarrow \mathbb{R}$ be a discrete filter of size $(2r+1)^2$. The discrete convolution operator $*$ can be defined as

$$(F^*k)(p) = \sum_{s+lt=p} F(s)k(t)$$

Let l be a dilated factor, $*^l$ can be defined as

$$(F^{*l}k)(p) = \sum_{s+lt=p} F(s)k(t)$$

It is traditional convolution when $l = 1$, and dilated convolution when $l > 1$.

There are two advantages for using dilated convolution instead of traditional convolution in medical image segmentation. First, though large receptive fields learn the global characteristic of lesions by covering a wider area over the images, it might extend the cost of computational memory and time consumption for a growing number of parameters. While dilation can expand the receptive field of the convolution layer as much as skipped pixels without extra parameters. Second, when using dilated convolution, no down-sampling layers are required to obtain large receptive fields. So high-resolution segmentation results can be directly predicted by the network, which satisfies the needs for the substantial accurate disease diagnosis based on the results.

Pre-trained Model: Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned [25]. It can be implemented by applying knowledge from one problem to another related problem, or used in a situation where the training data is not big enough, or to prevent over fitting [26]. Since data annotation in medical areas is difficult and time-consuming, it's always hard to get enough training data for Deep Learning models. Transfer learning has been widely used in medical image problems [27], [28] by using a pre-trained model on large-size datasets, like ImageNet, and then fine-tuned with a small-sized dataset.

Here, we adopted the pre-trained UNet++ [29] model to the CT brain images for hemorrhage segmentation. UNet++ has a similar encoder-decoder structure with UNet where the encoder and decoder sub-networks are connected through a series of nested, dense skip pathways. The re-designed skip pathways aim at reducing the semantic gap between the feature maps of the encoder and decoder sub-networks.

To use the pre-trained weights on ImageNet, we use pre-trained Inception/VGG as an encoder instead of the original encoder. To construct an encoder we remove the last fully-connected layers and replace them with a single convolution layer that separates encoder from the decoder. The decoder is the same as the original UNet++ and then fine-tuned on CT brain images by fixing the weight of encoder so we can get a well-trained model that is specifically for hemorrhage segmentation.

C. Adversarial Training

Adversarial training was first proposed in Generative Adversarial Network (GAN) [30] in the context of generative modeling. GAN is composed of two modules: a generator and a discriminator. The generator synthesizes samples whose distribution closely matches that of real data, while the discriminator estimates the probability that a sample comes from the training data instead of data that is synthesized by the generator [31]. The

generator's objective is to maximize the probability that the discriminator makes a mistake, while the discriminator is optimized to minimize the chance of mistake. It has been demonstrated that the generator produces samples (e.g., images) that are highly realistic [32]. A key insight in this adversarial process is that the discriminator, which itself can be a complex neural network, can learn to exploit higher-order inconsistencies in the samples synthesized by the generator [33].

We propose to use adversarial training on the UNet model for segmenting CT images. Figure.1 shows the pipeline in incorporating the adversarial process to the UNet segmentation model. We train a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation results that either come from the ground truth or the segmentation network. The UNet segmentation network, which plays the role of the generator in GAN, makes pixel-level predictions of the target classes. The discriminator network takes the ground truth and the segmentation result of UNet as input, while outputting the probability of prediction results.

According to the chart for the generator of GAN UNet, as can be seen in Figure.3(a), the depth of the down-sampling procedure is 7, and each step has 1 convolutional layer, 1 activation layer, and 1 batch normalization layer. Comparing to the traditional UNet, patch training and the activation layer of LeakyReLu were used here to preserve more features in the neurons in the convolutional layer. The up-sampling procedure has the same depth as the down-sampling encoder. In this decoder particularly, each down-sampling was connected to the corresponding up-sampling step instead of the skip connection in the original UNet. This can provide more training parameters for the discriminator and increases the accuracy of the discriminator's judgment. The discriminator contains 4 steps, as can be seen in Figure.3(b). The last three steps consist of 1 convolutional layer, 1 activation layer, and 1 batch normalization layer. While the first step does not have the batch normalization layer, which was designed to prevent overfitting issues.

As shown in Figure 1, the proposed model considers the input of the network, the structure of the network and the training process. Through analysis on the characteristic of medical image and through exhaustive experiments on different model, the pseudocode of the proposed model is as shown in Figure 4.

III. EXPERIMENTAL SETTING AND RESULT

In the experiment, we compare the different performances of the proposed UNet based model with different configuration network structures: original UNet(UNet10), streamlined UNet(UNet6), UNet with adversarial training(UNet10Gan), streamlined UNet with adversarial training (UNet6Gan), UNet with dilated convolution(UNet6Dil), UNet with symmetric information(UNet6Flip), UNet with symmetric and dilated convolution(UNet6FlipD), UNet++ with Inception pre-trained model(UNet6Incep) and UNet++ with VGG pre-trained model(UNet6Vgg). Besides, several state-of-the-art deep learning segmentation models such as DeepLabV3+[40], FCN[41], RefineNet[42], SegNet[43] and PspNet[44] were also included

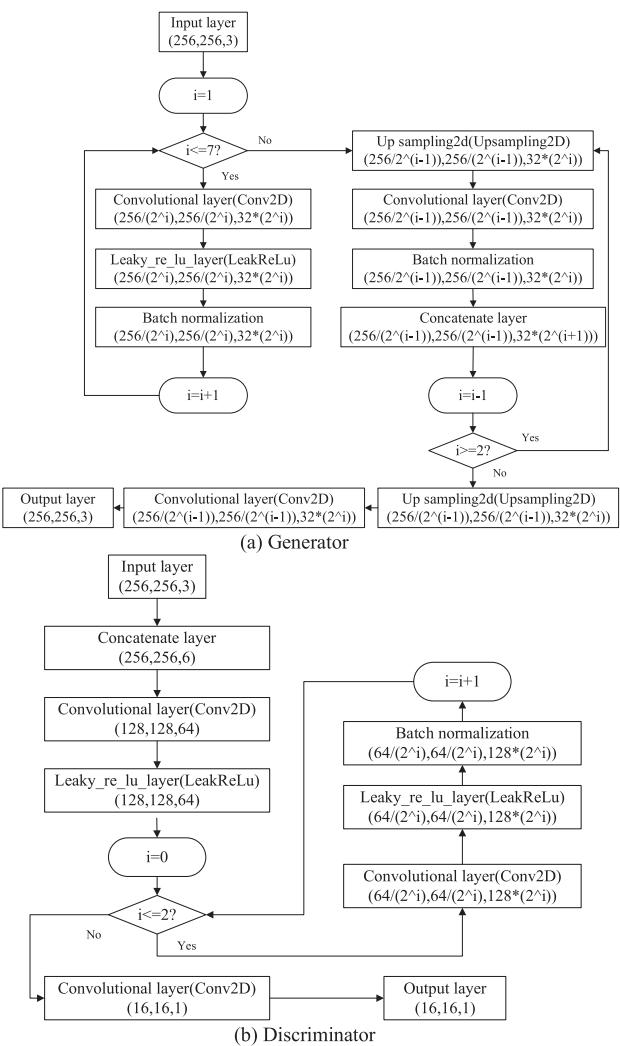


Fig. 3. the structure of GAN.

in the performance comparisons to validate the effectiveness of the proposed model.

A. Experimental Setting

1) Datasets: The experiments are performed based on two datasets:

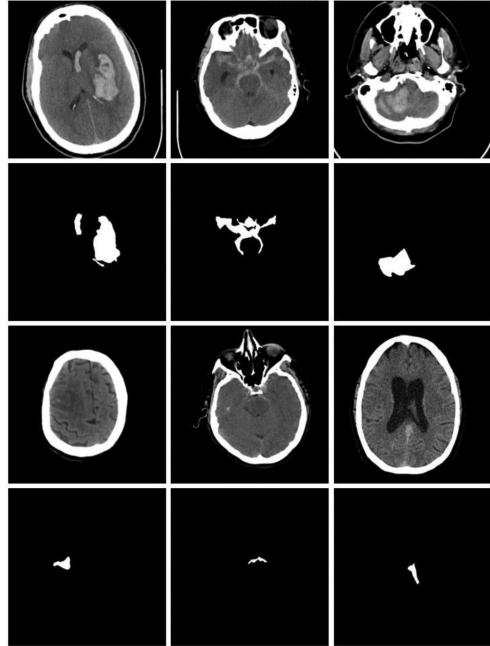
Dataset1: A single comprehensive stroke center dataset is collected, which includes the brain CT images of 190 patients presenting acute stroke symptoms to the emergency department (58 with hemorrhage, 101 normal). Strokes occur in different locations, with large differences in shape, size, contrast, and unclear boundaries compared with normal parts of the brain across different patients. Interpreting a CT image for hemorrhagic stroke requires thorough evaluation, especially to identify the small area of hemorrhage with low contrast from normal brain tissue. Figure. 5 shows a large size hemorrhagic stroke easily diagnosed visually. But for lesions that do not have distinguished edges or contrast with normal parts, it is very difficult to confirm even by an expert. Based on that, stroke neurologists divided the dataset into two different subsets: one containing subjects

Pseudocode of the proposed model:

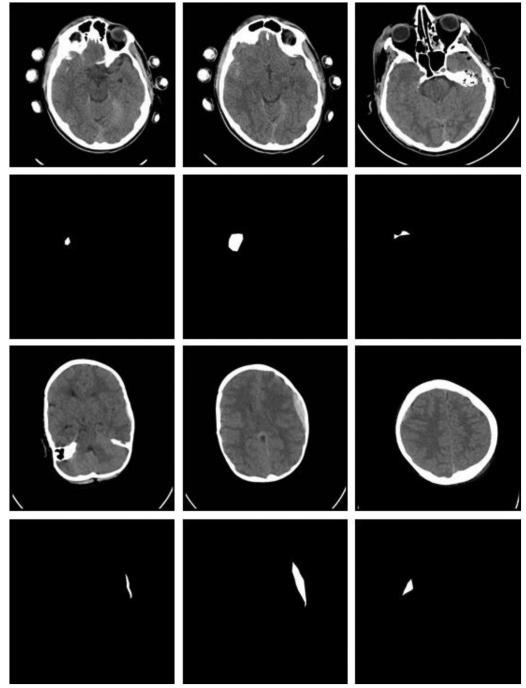
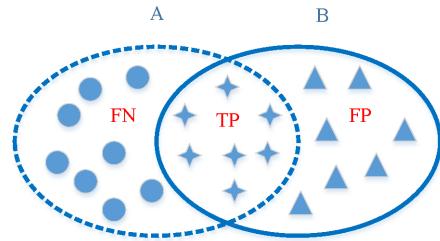
```

Input: CT brain image  $x$  from dataset, segmenting ground truth  $y$ , random noise  $z$ 
Output: predicting segmentation image  $p$ 
Pre-train the generator  $G$ :
    Sample image  $x$  with ground truth  $y$  from dataset
    Minimum  $L_{dice}$ 
     $minL_{dice} = 1 - \frac{2\omega \sum_{i=1}^N x(i)y(i)}{\sum_{i=1}^N x^2(i) + \sum_{i=1}^N y^2(i)}$ 
    % $x(i)$ -the value of pixel  $i$  for input image  $x$ 
    % $y(i)$ -the value of pixel  $i$  for input image  $y$ 
    % $N$ -image size (256)
    Gradient decrement till convergence
     $\theta_j := \theta_j - \alpha \frac{\partial L_{dice}}{\partial p(i,k)}$ 
For each training epochs do:
    For each batch do:
        Sample image  $x$  with ground truth  $y$  from dataset
        Generator  $G$  predict the fake image  $G(x, z)$  by original image  $x$  and noise  $z$ 
        Test  $G(x, z)$  by discriminator  $D$ 
         $D(x, G(x, z))$  %Possibility of  $G(x, z)$  is real
        Train discriminator  $D$  % Maximum  $D(x, G(x, z))$ 
    Re-train generator  $G$ :
         $G^* = argmin_G max_D L_{GAN}(G, D) + L_{dice}(G)$ 
        While not convergent do:
             $\theta_j^* = \theta_j - \alpha \frac{\partial L_g}{\partial y(i)} = \theta_j - \alpha \frac{\partial L_{dice}(G)}{\partial y(i)} + \alpha \beta \frac{\partial L_{GAN}(G,D)}{\partial y(i)}$  % Gradient decrement
        end
    end
end

```

Fig. 4. pseudocode of the proposed model.**Fig. 5.** Hemorrhage stroke image from Dataset1. First row: image with big lesions (from subset1). Third row: image with small lesions (from subset2). The first and third row is the raw data, the second and fourth row is the gold standard from the hand-marked lesions by the doctor.

with large lesions that are easily diagnosed and one containing subjects with small lesions that are hard to diagnose, as can be seen in Figure 4. The former had 71 images and the latter had 49 images. Manual segmentation of the hemorrhage CT scans was used as the ground truth to evaluate different UNet structures.

**Fig. 6.** Hemorrhage stroke images from Dataset2. The first and third row is the raw data, the second and fourth row is the gold standard from the hand-marked lesions by the doctor.**Fig. 7.** FN, TP, FP with ground truth and automated result. A: Ground truth segmentation, B: Automated segmentation.

Dataset2: A benchmark dataset, which collected by Hssayeni[45], [46], includes 36 scans for patients diagnosed with intracranial hemorrhage with the following types: Intraventricular, Intraparenchymal, Subarachnoid, Epidural and Subdural. Each CT scan for each patient includes about 30 slices with 5 mm slice-thickness. The mean and std of patients' age were 27.8 and 19.5, respectively. 46 of the patients were males and 36 of them were females. Each slice of the non-contrast CT scans was examined by two radiologists who recorded hemorrhage types if hemorrhage occurred or if a fracture occurred. Typical images of dataset2 are as shown in Figure 5.

2) Evaluation Criteria: We evaluate the performance of the proposed model using Location Accuracy (LA), Dice similarity coefficient (DSC), Intersection over Union (IoU), and Precision and Recall.

Location Accuracy (LA). Location accuracy (LA) is used to measure the accuracy of detection of lesions. One of the essential functions in the automatic lesion diagnosis system is to find the

position of the lesion region, which can give the doctor a quick view to guide the area to be carefully distinguished. It can be defined as below:

$$LA = \sum_{i=1}^n \alpha / n$$

$$\alpha = \begin{cases} 1, & \text{if } OP \cap GT \neq 0 \\ 0, & \text{if } OP \cap GT = 0 \end{cases}$$

In which, α indicates whether the output can locate the correct lesion region in the image. n is the number of images. OP is the image of output generated through the proposed model, GT is the image of the ground truth.

Dice similarity coefficient (DSC). The Dice similarity coefficient (DSC), also known as the Dice score, is the most widely used metric to measure the similarity/overlap between manual and automatic segmentation [34] [35]. The quantitative indicator formulae are shown below:

$$Dice = \frac{2TP}{2TP + FP + FN}$$

In which, true positive (TP) indicates that the model correctly predicted voxel. False-positive (FP) indicates the voxel that the model classifies negative as positive. False-negative (FN) indicates that the positive voxel is mistakenly classified as negative by the model. Dice ranges from 0 to 1 (where 0 means there were no overlapping voxels and 1 means that the segmentations were completely the same).

Intersection over Union (IoU). Intersection Over Union (IoU), also known as the Jaccard index, is another evaluation metric for segmentation. It can be interpreted as a similarity measure between a finite number of sets [36]. Intersection over union for similarity measure between two sets A and B can be defined as follows:

$$IoU = \frac{TP}{TP + FP + FN}$$

If the prediction is perfect, $IoU = 1$, and if it completely misses, $IoU = 0$. A degree of overlap will produce an IoU value between those two. Although both Dice and IoU are varied from 0 to 1, IoU are usually lower than Dice. That's because the two of them have the following relationship:

$$\frac{Dice}{IoU} = 2 - \frac{2TP}{2TP + FP + FN} \geq 1$$

Compared with DSC, IoU provides a wider range of numerical assessments, especially in the case of poor segmentation. However, with the increase of overlapping areas, the Dice coefficient changed linearly, while the IoU changed non-linear. So DSC is more intuitive to evaluate the improvement of the algorithm.

Precision and Recall. Precision is the correct segmentation that refers to the percentage of true positive. In other words, it describes the purity of the positive detections relative to the ground truth.

$$Precision = \frac{TP}{TP + FP}$$

Recall, or sensitivity, is defined as the number of true positives

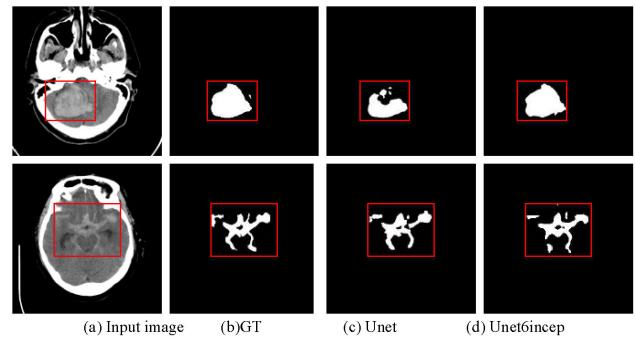


Fig. 8. The performance of detection on big hemorrhage lesion region.

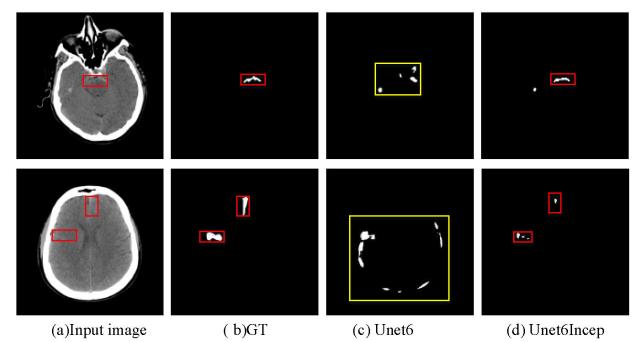


Fig. 9. The performance of detection on small hemorrhage lesion region.

divided by the total number of elements that belong to the positive cluster [37]. In other words, it describes the completeness of the positive predictions relative to the ground truth.

$$Recall = \frac{TP}{TP + FN}$$

Precision and recall are attractive as measures of segmentation quality because they are sensitive to over and under-segmentation. Over-segmentation leads to low precision scores, while under-segmentation leads to low recall scores. It is always a trade-off between precision and recall. Usually, we use average precision to measure the performance of the prediction. It is defined as the area under the precision-recall curve (PR curve). The x-axis is recall and the y-axis is precision. By changing the threshold cutoff (any detection with a score lower than the cutoff is treated as a false positive), we can get multiple precision-recall value pairs. After going through all the precision-recall value pairs corresponding to each unique threshold cutoff, then we have a precision-recall curve.

B. Experiments on Detection of the Hemorrhage Region

Figure 8 and 9 show some of the results for the detection of hemorrhage regions. As can be seen from the results, different models can get different segmentation results when used to detect the legions. It is effective to give the doctor enough information about whether the slice of the image contain a hemorrhage region and where the lesion is accurately. The first step of automatically diagnosis for hemorrhage stroke is

TABLE 1
LOCATION ACCURACY OF DIFFERENT MODELS FOR CT BRAIN IMAGES ON DATASET1

Models	UNet10	UNet10Gan	UNet6	UNet6Gan	UNet6Dil	UNet6Flip	UNet6FlipD	UNet6Incep	UNet6Vgg
Big hemorrhage dataset	0.9577	0.9577	0.9577	0.9155	0.9718	0.9155	0.9014	0.9859	0.9437
Small hemorrhage dataset	0.5510	0.5102	0.5510	0.5102	0.5306	0.4898	0.3061	0.6735	0.4498

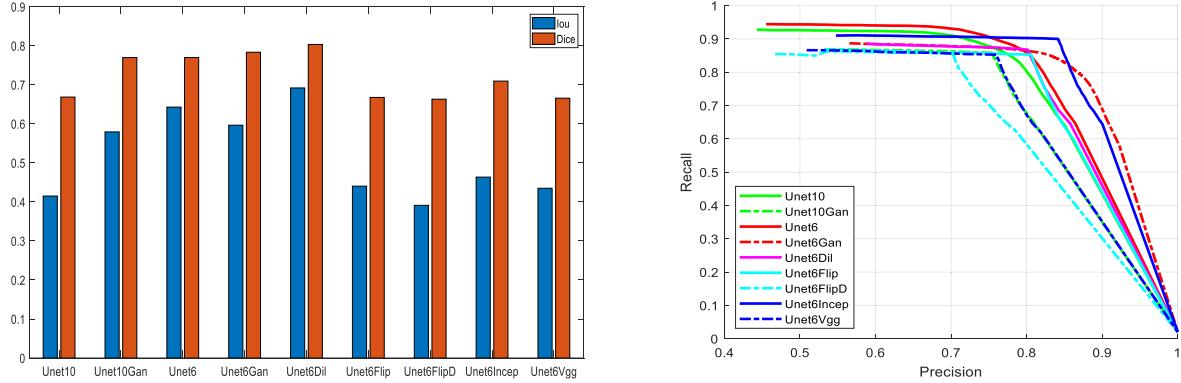


Fig. 10. Quantitative comparison for big hemorrhage region segmentation of Dataset1. Left: IoU and Dice. Right: PR curve.

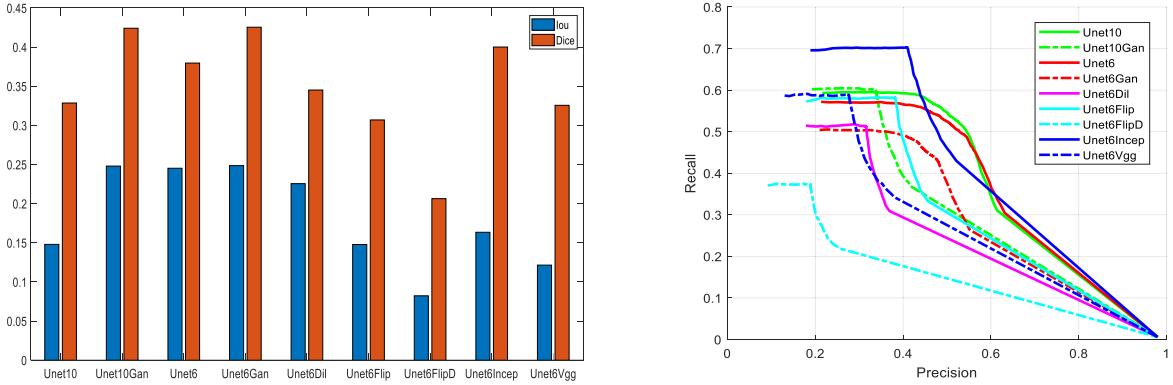


Fig. 11. Quantitative comparison for Small hemorrhage region segmentation of Dataset1. Left: IoU and Dice. Right: PR curve.

complete once the model can locate the accurate position of the lesion. Then it has detected the lesion successfully.

Table 1 shows the location accuracy for the two datasets. As can be seen, the location accuracy is all above 0.9 for the first dataset, which demonstrates that it is very effective at detecting hemorrhage lesions with a deep learning model. When using UNet6Incep, it can achieve the highest location accuracy 0.9859 on subset1 and 0.6735 on subset2. As can be seen in Figure.8, the lesion region is very difficult to distinguish for the input image. However, UNet6Incep can still locate the hemorrhage region accurately, which also shows significant advantages of the deep learning-based model.

C. Performance Comparisons on Proposed Unet-Based Models

Figure 10-11 presents the quantitative performance of the proposed UNet based model on the two subsets described above,

with a comparison of different data inputs, network structures, and different training processes. As can be seen, on the first dataset, with big hemorrhage lesions, the Dice on different models is above 0.67 and the IoU is above 0.4. While on the second dataset with small hemorrhage lesions, the Dice and IoU is much smaller. It is not surprising that the deep learning model shows considerable ability when compared to doctors in diagnosing.

In the big hemorrhage lesions subset of Dataset1, as can be seen in Figure.9, the IoU for all the different models is always above 0.67. This means that most of the lesions can be found by UNet based model. However, the accuracy of different models varied and the UNet6Dil obtained the best performance, with the highest Dice 0.8033 and IoU 0.6919. There are two reasons for that: First, we use 6 layers instead of 10 layers for the UNet model to avoid information loss due to the network being too deep; second, the dilated convolution, which has a larger receptive field compared to normal convolution, can obtain much

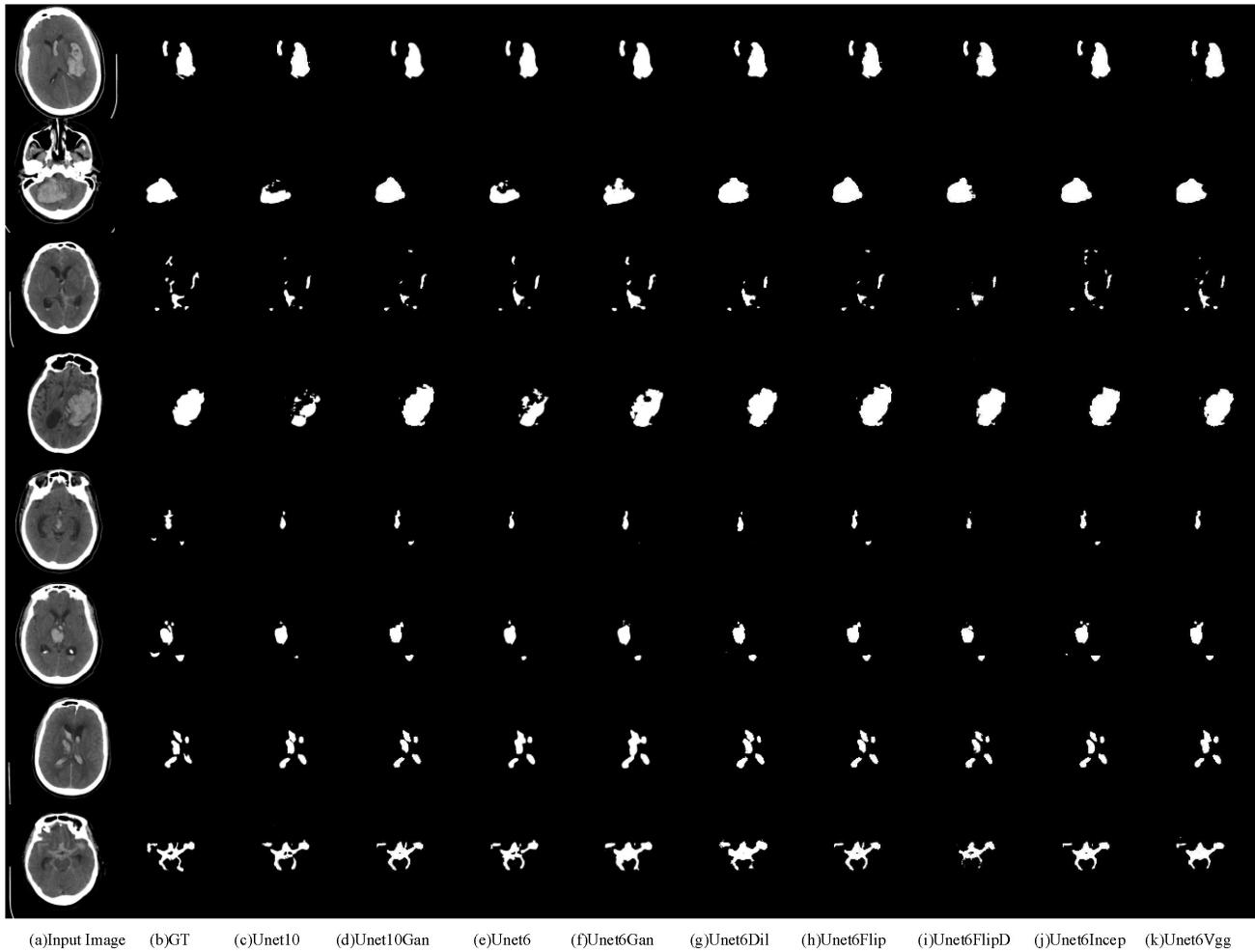


Fig. 12. Qualitative performance for big hemorrhage region segmentation on Dataset1.

more information for lesions. On the other hand, UNet6Flip, UNet6FlipD, UNet6Incep, and UNet6Vgg don't work well for IoU and Dice. But when using a PR curve to evaluate the performance, UNet6Incep gets the best precision and recalls at different thresholds. That's because, for model UNet6Incep, we use Inception as the encoder of the network, which can have a better description to capture the different characteristics of edge and contrast in lesion regions. Thus, even though UNet6Dil gets the highest IoU and Dice for the accuracy of segmentation, UNet6Incep has its advantages for accurately positioning the hemorrhage region.

In the subset of Dataset1 with small hemorrhage lesions, as can be seen in [Figure.10](#), the mean IoU for all the different models is around $0.3 \sim 0.45$, while IoU is lower than 0.3, which means that the UNet based model does not show the expected results on this dataset. However, we can still observe that when we adopt Gan as the training process for UNet, the performance of the network gets improved both on UNet10 and UNet6. Due to the small lesions that appear in this dataset, dilated convolution did not work as well as it did in the first dataset. That's because enlarging the receptive field on small lesions does not help capture the low contrast and weak characteristic of hemorrhage

regions. Also, in the PR curve, UNet6Incep gets very high recall compared to other models.

[Figure.12–13](#) presents the quantitative performance of the proposed UNet based model on the two subsets. As can be seen in [Figure.11](#), almost all the models can get the accurate location of the hemorrhage region in the CT images with big lesions. However, if we want to calculate the accurate volume of hemorrhage region, besides the location information, we need to get accurate segmentation with clear boundaries. For the third and fifth images, it is easy to see that UNet10Gan and UNet6Gan can get more lesion regions and clear boundaries than UNet. However, it might misclassify the normal region as hemorrhages when the normal one is surrounded by lesions. For the lesions that have small regions and are separated from each other, it is even harder for an expert to get all the accurate contours. Inspecting the fourth, sixth, and seventh rows of images, it can be seen that it's hard to get the accurate segmentation results as ground truth. During the different UNet based model, UNet6Incep generally does a better job at differentiating the lesion and the normal parts of the image. It can get the maximum hemorrhage region consistent with the ground truth set by experts. In the seventh image especially, it can be observed that



Fig. 13. Qualitative performance for small hemorrhage region segmentation on Dataset1.

only UNet6Incep detected the left-bottom small lesion, while all the other models missed it.

The qualitative comparison results for small hemorrhage region of dataset1 are shown in **Figure 13**. Inspecting all the images in the first column, it is challenging to find the hemorrhage lesion part from the CT image. As can be seen, UNet10, UNet6, and Gan-based training processes cannot locate the lesion region correctly, not to mention to segmentation. Dilated convolution and symmetry information doesn't work because of the weak characters in the original image might be lost. Even though all the UNet based models cannot produce as good results as showed in the first dataset, there are still some encouraging results. For the second image, the hemorrhage region, which was marked by experts as the ground truth, was very difficult to distinguish because there are also some high contrast normal regions near the hemorrhage region. All the UNet based models mistake the high contrast normal region as hemorrhage results, except for UNet6Incep, which can locate the position of the lesion region accurately. It also demonstrates that the pre-trained model shows great help for detection and segmentation, especially when it comes to small datasets and tricky problems.

D. Performance Comparisons on State-of-the Art Models

Since UNet6Gan achieves the best performance in Dataset1, to better validate the effectiveness of the model, this section compares the performance of the proposed UNet6Gan model, with six state-of-the-art deep learning models, including DeepLabV3+ [40], FCN [41], RefineNet [42], SegNet [43], PspNet [44]

and UNet [18]. **Figure 14** show the quantitative comparison of UNet6Gan model with state-of-the-art models on Dataset1. As can be seen, DeepLabv3+, FCN, RefineNet, SegNet and UNet get relatively low IOU and Dice scores, which means that these models work not so well on Hemorrhage lesion segmentation. UNet6Gan and PSPNet achieve the higher Dice scores which are above 0.6. Besides, UNet6Gan gets higher IOU than PSPNet. It also can be seen in the P-R curve that UNet6Gan got the highest recall than the other models.

Figure 15 show the typical results of the qualitative comparison of UNet6Gan model with state-of-the-art models on Dataset2. For the first image, the hemorrhage regions have clearly boundaries and high contrast with the background. Thus, all the models are able to get the segmentation results very well. For the second image, only DeepLabv3+ and UNet6Gan get the accurate segmentation results compared with the other models. For other images with low contrast lesion part and fuzzy edges, it is very difficult to achieve accurate segmentation result. However, the UNet6Gan model can get a better location and segmentation results of the lesion parts.

In order to further verify the superiority of the UNet6Gan model, quantitative and qualitative comparisons are also performed on a benchmark dataset, which is described briefly in Section III.A. **Figure 16** show the quantitative comparison of UNet6Gan model with state-of-the-art models on Dataset2. Since most of the hemorrhage lesion parts in Dataset2 are relatively small, fuzzy, and with low contrast, it is very difficult to diagnosis the accurate lesion areas even for an experienced doctor. So, the performance of the models including IOU and Dice are relatively low, most of IOU scores are under 0.3 and most of dice scores

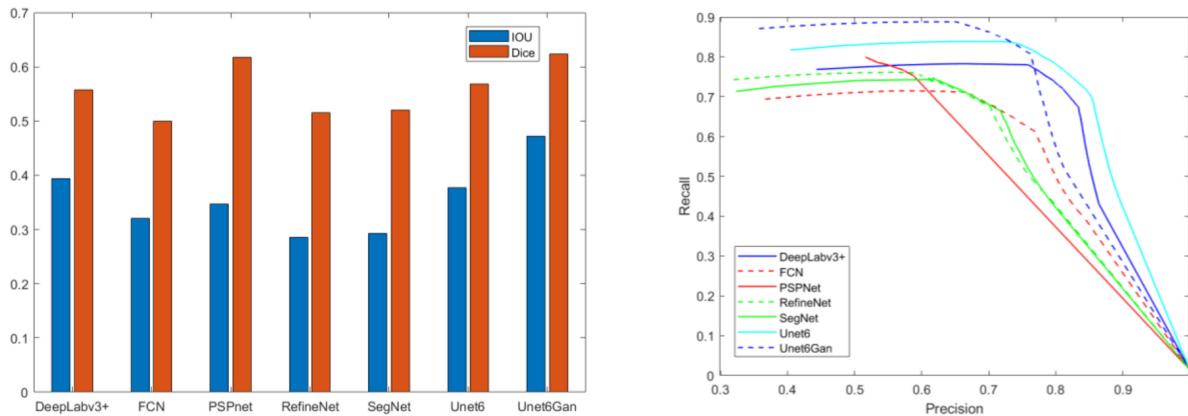


Fig. 14. Quantitative comparison of Unet6Gan with state-of-the-art models on Dataset1. Left: IoU and Dice. Right: PR curve

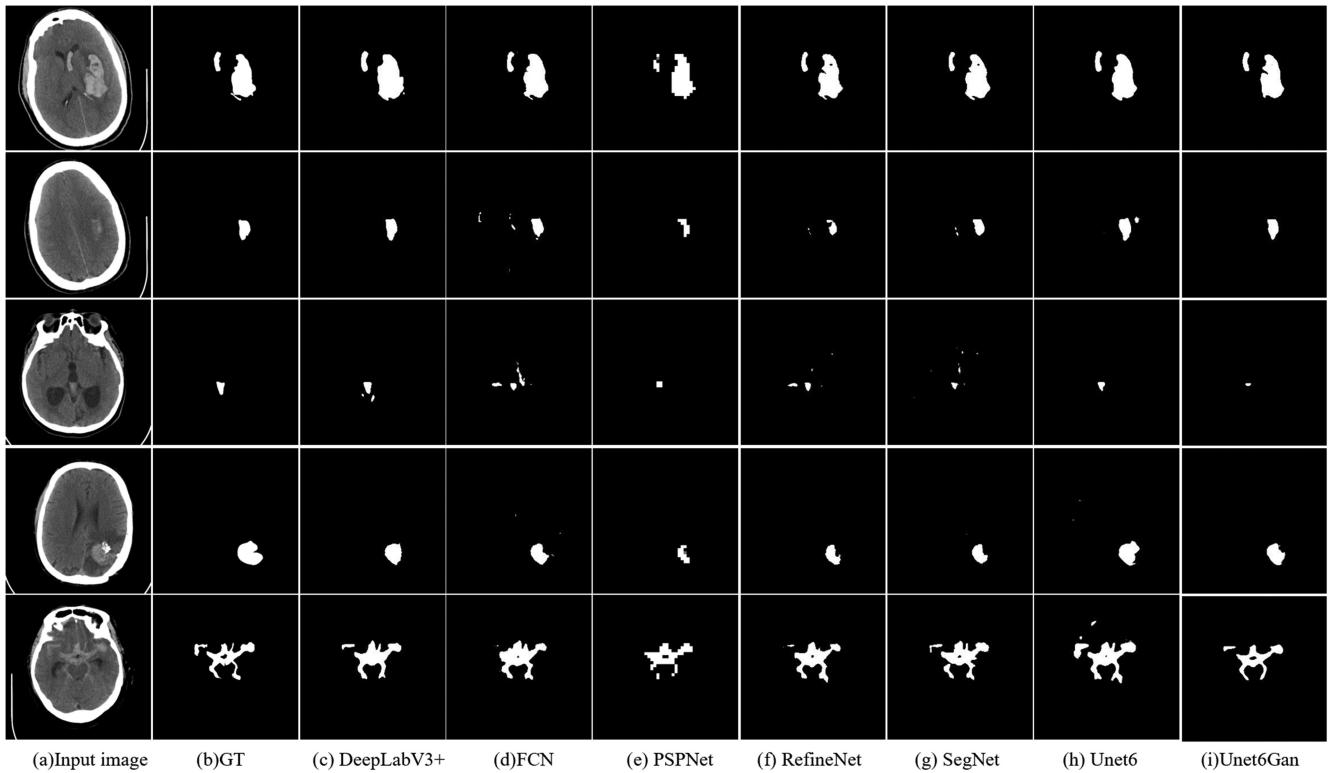


Fig. 15. Qualitative comparison with state-of-the-art models on Dataset1.

are under 0.4. However, it still can be seen that, among all the state-of-the-art models, the proposed Unet6Gan model achieves the highest Dice score which is above 0.4. Besides, the P-R curve also demonstrate the effectiveness of proposed model. We can observe that the proposed model significantly outperforms the other methods, which have higher precision and recall.

Figure 17 show the typical results of the qualitative comparison. Also, the proposed model yields more visually informative results compared with other state-of-the-art models. As can be seen, for the first, second, fourth images, even though other models such as DeepLabV3+, FCN, PSPNet and RefineNet could get the lesion parts, the segmentation results of Unet6Gan

are more complete and consistent with the ground truth. As for the third and fifth images, most of other state-of-the-art models can not achieve sufficient result. However, it is clear from visual inspection that only Unet6Gan model has identified and segmented the lesion area accurately. To better evaluate the performance of the proposed model, quantitative performance comparison with standard deviations and the running time of state-of-the-art models on Dataset1 and Dataset2 are shown in Table 2. Unet6Gan achieve the highest IOU scores on Dataset1 0.4723, and on Dataset2 0.2754. Also, Unet6Gan achieve the highest Dice scores on Dataset1 0.6242, on Dataset2 0.4348. We can observe that, not only on our Dataset1, but also

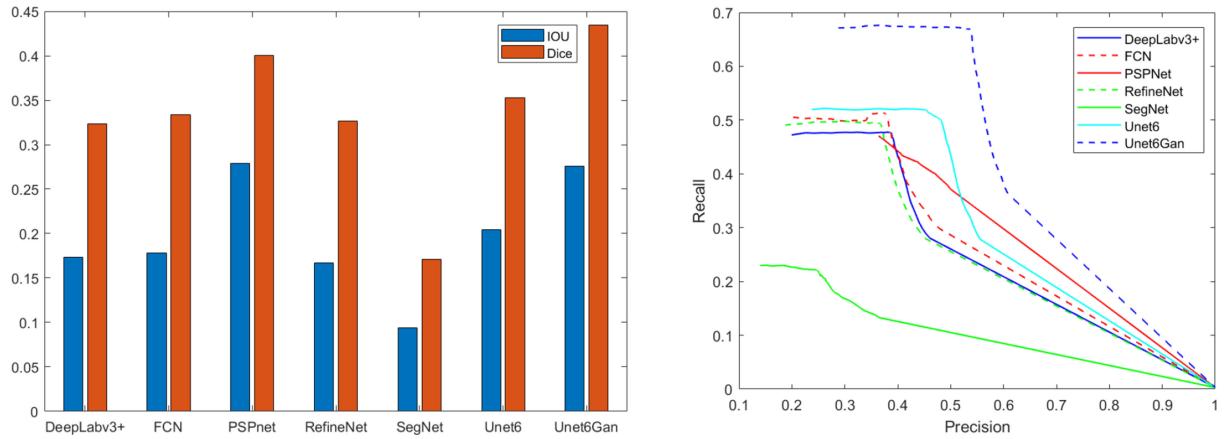


Fig. 16. Quantitative comparison of Unet6Gan with state-of-the-art models on Dataset2. Left: IoU and Dice. Right: PR curve

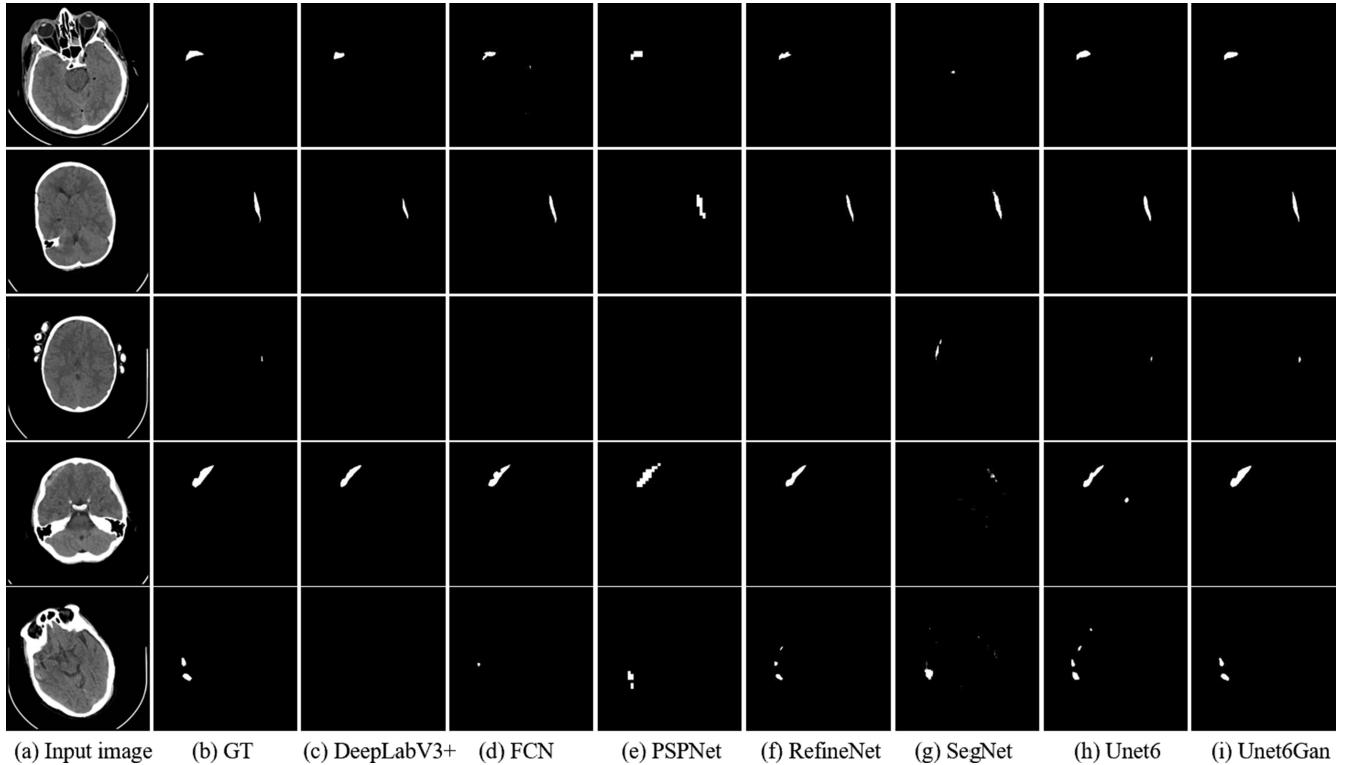


Fig. 17. Qualitative comparison with state-of-the-art models on Dataset2.

on benchmark Dataset2, Unet6Gan model always achieve the best performance. Besides, also we evaluate the computational complexity of different models. Our method is implemented in PyTorch framework with SimpleITK on Ubuntu 16.4 (64 Bit) operating system. All of our experimental networks are fitted into a powerful workstation equipped with two Intel XeonTM E5 2630V4 CPUs (2.20 GHz) and four NVIDIA GTX 1080 Ti graphic cards each with the memory of 11 GB. DeepLabv3+ is the most time-consuming and Unet is the most time-efficient model. Compared with Unet, Unet6Gan introduce discriminator structure to exploit higher-order inconsistencies in the samples

synthesized by the generator, so it needs more time than Unet. Compared with the performance, the extra time is worth for it to have more accurate segmentation results.

IV. DISCUSSION

Our study verifies that using U-net-based deep learning frameworks to detect and segment hemorrhagic strokes automatically is feasible as it is able to localize the hemorrhage area in the brain with high accuracy. By comparing Unet6Gan model with six state-of-the-art segmentation models, the superiority of the

TABLE II
RUNNING TIME OF STATE-OF-THE-ART MODELS ON DATASET1 AND DATASET2

Running time	DeepLabv3+	FCN	PSPnet	RefineNet	SegNet	Unet	Unet6GAN
On Dataset1 (ms)	311.6	201.3	188.2	298.5	279.2	156.4	172.1
On Dataset2 (ms)	302.1	197.9	179.2	287.5	257.3	158.2	163.7

proposed Unet6Gan model was demonstrated in the segmentation of hemorrhage lesion on CT brain images. A hemorrhagic stroke is a common and deadly disease in which the presenting symptoms may mimic the more common diagnosis of ischemic stroke. The treatments for ischemic strokes and hemorrhagic strokes are substantially different, in which ischemic stroke treatment is focused on busting the clot blockage in the cerebral vessel that is causing ischemia, whereas hemorrhagic stroke treatment is focused on controlling the bleeding from ruptures of small cerebral vasculature. Misdiagnosis leading to utilizing clot-busting medication in a hemorrhagic stroke mimicking an ischemic stroke can cause fatal bleeding. CT brain without contrast is most often the recommended test of choice to identify the type of stroke. However, due to the time-sensitivity of ischemic stroke treatment, the CT brain is often interpreted in a rushed manner, leading to potential missed diagnosis of hemorrhagic areas.

In a published study by Schriger *et al.* [38], the image interpretation accuracy of emergency physicians taking care of stroke patients without the support of a radiologist is 67%. Due to the low accuracy, interpretation by a radiologist can be crucial decision support, but the turnaround time is often delayed in many centers. Utilizing an accurate artificial intelligence as a clinical decision support tool to assist in radiologist interpretation turnaround time and for support of non-radiologist physicians in resource-limited settings can be helpful. The proposed machine learning-based U-net framework in our study has shown high accuracy (0.9859) in locating the hemorrhage area, which is the most useful feature in determining the presence of hemorrhage in the setting of differentiating the cause of acute stroke symptoms. The highest accuracy for location can be acquired by the UNet6Incep model, which uses 6 dilated convolutional layers for UNet with pre-trained Inception on ImageNet. This kind of structure can abstract the low-level features of lesions in medical images on the shallow layer of the network. Besides, the pre-trained Inception makes the network have better representation to distinguish the different characteristic between the lesion region and the normal part. Thus, it is easy for the model to locate the hemorrhage region in the CT brain image accurately and make it possible to be a useful support tool for the doctor to make a clinical decision.

Apart from locating the area of a hemorrhagic stroke, accurate segmentation is also a useful feature. The size and location of the hemorrhagic stroke are well-known indicators utilized in the Intracerebral Hemorrhage scoring system (ICH score) [39] for prognostication. Automation of hemorrhagic stroke segmentation and volumetric calculation can be useful in this case. Automated localization and segmentation can also be useful in the setting of identifying surgical candidate, rather than the crude manual calculation. In our study, the machine learning

algorithm demonstrated an accuracy of 0.8033 for segmentation, which can be realized through the UNet6Dil model. Since UNet is designed for medical image segmentation, it has the most advantageous encoder-decoder structure and skip connection to make the model capture the spatial information of the lesion regions. Also, UNet6Dil uses dilated convolution to increases the accuracy of segmentation through a large receptive field.

However, even though the competitive location accuracy and segmentation performance can be achieved on dataset1 with large lesions, it is hard to get satisfactory performance on dataset2, which contains small lesions as of now. For the images with very small lesions and low contrast areas, it is very difficult even for an experienced radiologist to make a decision. Therefore, improving the accuracy of location accuracy and segmentation to make the model achieve and surpass the capability of experienced experts is still the most important thing we need to do. In our following research work, annotating more data for the model to learn should be useful to improve the performance of the model in the future. We will also develop advanced networks, such as unsupervised learning, to let the model learn from the data itself, and saliency model, which introduces the basic principles of human visual attention.

V. CONCLUSIONS

In this paper, we proposed a U-net based deep learning framework to get accurate hemorrhage segmentation in CT brain images. First, for the input of the network, data symmetry, and data augmentation are considered in the proposed model to abstract the structural symmetry of the brain image and prepare enough training data. Second, for the network structure, a streamlined UNet model with dilated convolutions was adopted, aiming to get a better description of the low contrast character for the hemorrhage lesions. During which, a pre-trained model was also compared with the proposed UNet model. Finally, adversarial training is also used in the proposed UNet model to improve the accuracy of the segmentation. Comparison experiments based on CT brain images demonstrated that the proposed UNet based model shows great advantages compared with human experts on hemorrhage lesion diagnosis.

REFERENCES

- [1] M. Vijayan and P. H. Reddy, "Peripheral biomarkers of stroke: Focus on circulatory microRNAs," *Biochimica et Biophysica Acta -Mol. Basis Dis.*, vol. 1862, no. 10, pp. 1984–1993, 2016.
- [2] P. Chang *et al.*, "Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT," *Amer. J. Neuroradiol.*, vol. 39, no. 9, pp. 1609–1616, 2018.
- [3] H. Lee *et al.*, "An explainable deep-learning algorithm for the detection of acute intracranial hemorrhage from small datasets," *Nat. Biomed. Eng.*, vol. 3, no. 3, p. 173, 2019.

- [4] M. M. Kyaw, "Computer-aided detection system for Hemorrhage contained region," *Int. J. Comput. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 11–16, 2013.
- [5] T. Gong *et al.*, "Classification of CT brain images of head trauma," in *Proc. IAPR Int. Workshop Pattern Recognit. Bioinf.*, 2007, pp. 401–408, Berlin, Heidelberg: Springer.
- [6] T. Chan, "Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain," *Computerized Med. Imag. Graph.*, vol. 31, no. 4-5, pp. 285–298, 2007.
- [7] R. Liu *et al.*, "Hemorrhage slices detection in brain ct images," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.
- [8] N. Pérez *et al.*, "Spontaneous intracerebral hemorrhage image analysis methods: A survey," in *Advances in Computational Vision and Medical Image Processing*. Berlin, Germany: Springer, 2009, pp. 235–251.
- [9] A. H. Ali, S. I. Abdulsalam, and I. S. Nema, "Detection and segmentation of hemorrhage stroke using textural analysis on brain CT images," *Int. J. Soft Comput. Eng.*, vol. 5, pp. 11–14, 2015.
- [10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [11] M. Grewal *et al.*, "Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 281–284.
- [12] K. Jnawali *et al.*, "Deep 3D convolution neural network for CT brain hemorrhage classification," in *Medical Imaging 2018: Computer-Aided Diagnosis*. Bellingham, Washington USA: International Society for Optics and Photonics, 2018, Art. no. 105751C.
- [13] Q. Jin *et al.*, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans" 2018, *arXiv:1811.01328*.
- [14] S. M. Abulnaga and J. Rubin, "Ischemic stroke lesion segmentation in CT perfusion scans using pyramid pooling and focal loss," in *Proc. Int. MICCAI Brainlesion Workshop*, Berlin, Germany: Springer, 2018, pp. 352–363.
- [15] Z. Liu *et al.*, "Towards clinical diagnosis: Automated stroke lesion segmentation on multi-spectral MR image using convolutional neural network," *IEEE Access*, vol. 6, pp. 57006–57016, 2018.
- [16] H. Kuang, B. K. Menon, and W. Qiu, "Segmenting hemorrhagic and ischemic infarct simultaneously from follow-up non-contrast CT images in patients with acute ischemic stroke," *IEEE Access*, vol. 7, pp. 39842–39851, 2019.
- [17] Y. Wang *et al.*, "A deep symmetry convnet for stroke lesion segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 111–115.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [19] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: A dimension-fusion U shape network for chronic stroke lesion segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published.
- [20] Y. Jeong *et al.*, "Dilated saliency U-Net for white matter hyperintensities segmentation using irregularity Age Map," *Front. Aging Neurosci.*, vol. 11, p. 150, 2019.
- [21] H. Shen, J. Zhang, and W. Zheng, "Efficient symmetry-driven fully convolutional network for multimodal brain tumor segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3864–3868, IEEE.
- [22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
- [23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learning*. 2010, pp. 807–814.
- [24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
- [25] L. Torrey and J. Shavlik, "Transfer learning, in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Pennsylvania, USA: IGI Global, 2010, pp. 242–264.
- [26] Y. Seo and K.-S. Shin, "Image classification of fine-grained fashion image based on style using pre-trained convolutional neural network," in *Proc. IEEE 3rd Int. Conf. Big Data Anal.*, 2018, pp. 387–390.
- [27] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *J. Med. Imag.*, vol. 3, no. 3, Art. no. 034501, 2016.
- [28] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018, Paper e9.
- [29] Z. Zhou *et al.*, "UNet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2018, Berlin, Germany: Springer, pp. 3–11.
- [30] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*. Montreal, Quebec, Canada, 2014, pp. 2672–2680.
- [31] P. Russo *et al.*, "From source to target and back: symmetric bi-directional adaptive gan," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8099–8108.
- [32] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv:1511.06434*, 2015.
- [33] W. Dai *et al.*, "Scan: Structure correcting adversarial network for organ segmentation in chest x-rays," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2018, Berlin, Germany: Springer, pp. 263–273.
- [34] L. R. Dice, "Measures of the amount of ecological association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [35] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, p. 29, 2015.
- [36] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," 2018, *arXiv:1801.05746*.
- [37] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," in *Proc. 7th Australasian Data Mining Conf.*, vol. 87, pp. 27–32, 2008.
- [38] D. L. Schriger *et al.*, "Cranial computed tomography interpretation in acute stroke: Physician accuracy in determining eligibility for thrombolytic therapy," *JAMA*, vol. 279, no. 16, pp. 1293–1297, 1998.
- [39] J. C. Hemphill *et al.*, "The ICH score," *Stroke*, vol. 32, no. 4, pp. 891–897, 2001.
- [40] L. C. Chen *et al.*, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.
- [41] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [42] G. Lin *et al.*, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1925–1934.
- [43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [44] H. Zhao *et al.*, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2017, pp. 2881–2890.
- [45] M. Hssayeni, "Computed tomography images for intracranial hemorrhage detection and segmentation (version 1.3.1)," *Physio Net.*, 2020. [Online]. Available: <https://physionet.org/content/ct-ich/1.3.1/>
- [46] M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial hemorrhage segmentation using a deep convolutional model," *Data*, vol. 5, no. 1, p. 14, 2020.