This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

Automatic Segmentation of Acute Ischemic Stroke from DWI using 3D Fully Convolutional DenseNets

Rongzhao Zhang, Lei Zhao, Wutao Lou, Jill M Abrigo, Vincent CT Mok, Winnie CW Chu, Defeng Wang and Lin Shi

***Abstract*—Acute ischemic stroke is recognized as a common cerebral vascular disease in aging people. Accurate diagnosis and timely treatment can effectively improve the blood supply of the ischemic area and reduce the risk of disability or even death. Understanding the location and size of infarcts plays a critical role in the diagnosis decision. However, manual localization and quantification of stroke lesions are laborious and time-consuming. In this paper, we propose a novel automatic method to segment acute ischemic stroke from diffusion weighted images (DWI) using deep 3D convolutional neural networks (CNNs). Our method can efficiently utilize 3D contextual information and automatically learn very discriminative features in an end-to-end and data-driven way. To relieve the difficulty of training very deep 3D CNN, we equip our network with dense connectivity to enable the unimpeded propagation of information and gradients throughout the network. We train our model with Dice objective function to combat the severe class imbalance problem in data. A DWI dataset containing 242 subjects (90 for training, 62 for validation and 90 for testing) with various types of acute ischemic stroke was constructed to evaluate our method. Our model achieved high performance on various metrics (Dice similarity coefficient: 79.13%, lesion-wise precision: 92.67%, lesion-wise F1 score: 89.25%), outperforming other state-of-the-art CNN methods by a large margin. We also evaluated the model on ISLES2015-SSIS dataset and achieved very competitive performance, which further demonstrated its generalization capacity. The proposed method is fast and accurate, demonstrating a good potential in clinical routines.**

***Index Terms*—Acute ischemic stroke segmentation, DWI, 3D convolutional neural networks, deep learning.**

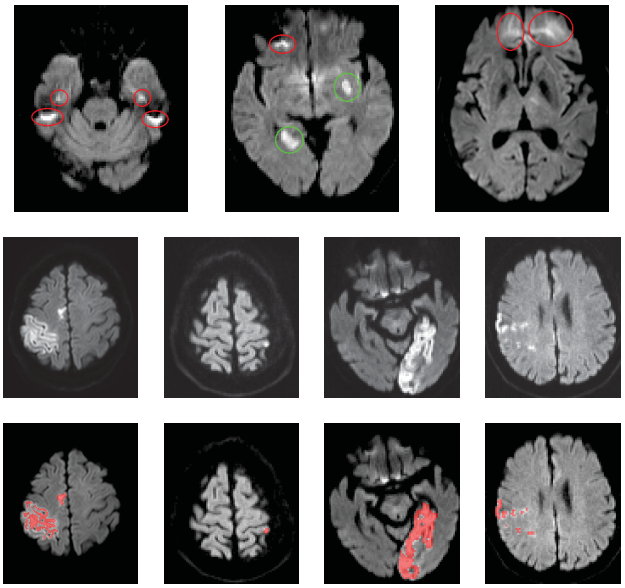


Fig. 1. Challenging examples in acute ischemic stroke segmentation. The first row shows three DWI slices with artifacts. Red and green circles denote artifacts and stroke lesions respectively. The second row shows large, small, ill-defined boundaries and multi-focal lesions respectively. The third row shows corresponding DWI slices overlaid with manual segmentation. Best viewed in color.

I. Introduction

STROKE is the second leading cause of death worldwide, accounting for 6.24 million deaths globally in 2015 [1]. The typical symptom of acute stroke is the sudden onset of a focal neurologic deficit, such as dysphasia, hemianopia, sensory loss, etc. [2]. These symptoms may develop into chronic diseases (e.g. dementia, hemiplegia, etc.), which can profoundly affect patients' life and consume a large part of social healthcare cost [3].

R. Zhang, L. Zhao, W. Lou and V. Mok are with the Department of Medicine and Therapeutics, The Chinese University of Hong Kong, HK, China. J. Abrigo, W. Chu, D. Wang and L. Shi are with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, HK, China. V. Mok and L. Shi are also with Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, HK, China. D. Wang is also with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University and School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing, China. L. Shi is also with BrainNow Medical Techonology Limited, Hong Kong Science and Technology Park, HK, China.

Corresponding author: L. Shi (email: shilin@cuhk.edu.hk).

Ischemic cases account for 87% of all stroke [4]. Prompt diagnosis and treatment in its acute stage are of vital importance for the recovery and outcome of stroke patients. Diffusion weighted imaging (DWI) is a commonly performed magnetic resonance imaging (MRI) sequence that has short acquisition time and high sensitivity in detecting acute ischemic stroke. Apparent diffusion coefficient (ADC) maps can be derived from multiple DWI images with different b-values, which are not affected by T2 shine-through artifacts. Acute ischemic lesions appear hyperintense on DWI and hypointense on ADC [5]. Since the time window for the treatment of acute ischemic stroke is very short (e.g. intravenous thrombolysis should be performed within 3 hours after stroke onset), fast localization and quantification of the stroke lesions in DWI scans are vitally important. However, due to the low signal-to-noise ratio (SNR) and artifacts present in DWI, manual delineation of acute stroke is laborious and time-consuming. Thus, automatic segmentation methods are highly demanded.

Though DWI is highly sensitive to acute ischemic stroke, the accurate automatic segmentation is very challenging for

Copyright ©2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

several reasons. First, there exist many artifacts in DWI scans that mimic the intensity and shapes of stroke lesions, and the high noise level and the very low image resolution of DWI images make it even more difficult to recognize small lesions. Second, there are various stroke subtypes, which leads to large variations in lesion size and location. Stroke lesion volume can vary from hundreds of to tens of thousands of cubic millimeters, and cerebral infarctions may occur in any brain area, such as brain lobes, cerebellum and brainstem. Third, the multi-focal distribution and the ill-defined boundaries of some acute stroke lesions further aggravate the situation, as those ambiguous voxels on the boundaries may confuse the algorithm. We show some examples of the above challenges in Fig. 1.

Many works have been dedicated to the automatic segmentation of acute ischemic stroke. Prakash *et al.* [6] applied a shallow probabilistic neural network to select candidate slices and then segmented stroke lesions by adaptive gaussian mixture models. To improve the class boundaries, Hevia-Montiel *et al.* [7] combined weighted mean shift procedures with an edge confidence map. Mujumdar *et al.* [8] proposed an elaborate framework to combine features from multiple b-value DWI images, and employed an active contour algorithm for segmentation. However, these low-level features (mainly intensity and edge information) are not robust enough to the large variations in lesion patterns, especially when numerous artifacts are present. Recently, Chen *et al.* [9] proposed a framework based on 2D convolutional neural networks (CNNs) for acute ischemic stroke segmentation, and achieved state-of-the-art performance on a large DWI dataset. But this method was flawed by its 2D nature, which ignored the important 3D contextual information in volumetric data.

More recently, some attempts have revealed the potential of 3D CNNs on stroke lesion segmentation tasks. For example, in the well-known Ischemic Stroke Lesion Segmentation (ISLES) [10] challenge, DeepMedic [11] ranked first on the task of sub-acute ischemic stroke segmentation (SSIS), which integrated a multi-scale 3D CNN model and a fully connected conditional random field (CRF) post-processing step. Nevertheless, the power of 3D CNNs has not been fully studied for the segmentation of *acute* ischemic stroke, whose imaging modality and lesion features are different from *sub-acute* lesions.

Convolutional neural networks are a kind of neural network specializing in processing data with grid-like topologies, e.g. images [12]. In recent years, CNNs have been bringing breakthroughs in many vision tasks in both natural and medical images, such as image classification [13]–[15], object detection [16]–[18], semantic segmentation [11], [19]–[21], etc. Though CNNs have presented outstanding performance significantly superior to previous methods based on hand-crafted features, most of these works focused on 2D images or only developed shallow 3D architectures. Nevertheless, volumetric data is very common in medical imaging, such as 3D MRI, 3D computed tomography (CT), etc. Considering that higher-dimensional information are usually more complicated, shallow CNNs may be not expressive enough to capture sufficiently high-level features from 3D images. However, how to apply very deep 3D CNNs to medical images is still an open problem. Firstly, 3D CNNs require much more parameters than their 2D counterparts, thus are faced with severer vanishing-gradient problem in deep models; secondly, medical image datasets are usually far smaller than those in natural image domain, which may hinder the training of complex models with a large amount of parameters.

Some investigations have been devoted to developing sufficiently deep 3D CNNs, aiming at more accurate segmentation of volumetric medical images. Çiçek *et al.* proposed 3D U-Net [21] that employed an encoder-decoder framework with skip connections for semantic segmentation. Milletari *et al.* [22] combined 3D U-Net and residual functions [14], yielding a V-Net. Yu *et al.* [23] also built upon the encoder-decoder framework, and introduced mixed long and short residual connections to boost information flow within the network. Besides, Chen *et al.* proposed VoxResNet [24] that exploited residual learning to achieve performance gains by increasing model depth. These methods successfully trained deep 3D CNNs with limited volumetric medical images, thanks to the skip connection scheme and residual learning technique. However, in these works, skip or residual connections are either sparse or fused into the network by summation operations, which may still hinder the information and gradient flow in the network.

In this paper, we propose a novel 3D fully convolutional and densely connected convolutional network (3D FC-DenseNet) for the accurate automatic segmentation of acute ischemic stroke. Basically, our model is a 3D fully convolutional network (FCN) [19] for semantic segmentation. It is very deep, with 20 3D convolutional layers in the main pathway. We build our network based on the idea of densely connected convolutional networks (DenseNets) [25], which allows each layer takes as input all its preceding feature maps, so that information and gradients can propagate unimpededly throughout the network. To recover the image details lost in down-sampling operations and enable the network to exploit multi-scale information, we incorporate multi-scale features into our model via auxiliary classification pathways. Moreover, in order to eliminate the influence of class imbalance in data, we utilize Dice objective function to optimize our model. To validate the efficacy of our proposed method, we built a relatively large dataset of DWI and ADC containing 242 subjects (where 90 for training, 62 for validation and 90 for testing) with various types of acute ischemic stroke. Extensive comparative experiments were conducted on this dataset, and the results corroborated the superiority of our network architecture and training strategy. We also evaluated the proposed method on a public dataset (i.e. ISLES2015-SSIS), where it presented very competitive performance among 28 entries.

The main contributions of our work are four-fold:

1) We extend DenseNets to 3D and tap their potential on acute ischemic stroke segmentation from DWI. Dense connectivity effectively boosts the information and gradients flow in the network. Experiment results corroborate the superiority of our proposed model compared to other state-of-the-art CNN models. Our study may inspire more works to utilize DenseNets to tackle other challenges in medical image analysis (MIA) field.

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

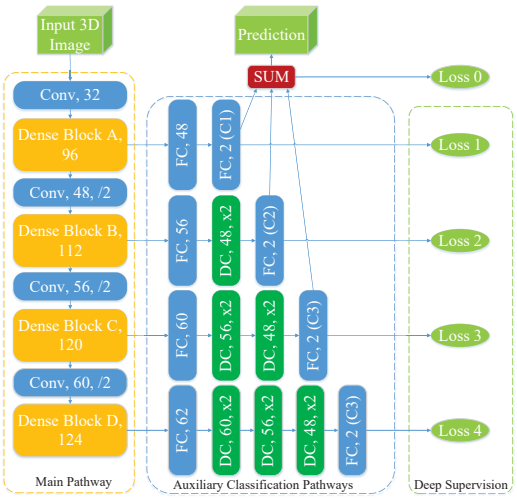


Fig. 2. Architecture of the proposed 3D FC-DenseNet. Each dense block contains 4 layers and has a growth rate of $k = 16$. The structure of dense blocks can be found in Fig. 4(b). 'FC' denotes convolutional layers with $1 \times 1 \times 1$ kernels. The numbers following layer names indicates the number of output feature volumes. 'DC' represents deconvolutional layers. '/2' and '×2' indicate down-sampling and up-sampling respectively. The first down-sampling layer and corresponding up-sampling layers are performed with an anisotropic stride of $1 \times 2 \times 2$.

2) We develop a novel 3D FC-DenseNet to meet the challenge of acute stroke segmentation, and propose to train it with Dice objective function (similar to the idea in [26]) to tackle the severe class imbalance problem in data. Extensive experiments evidence the effectiveness of the proposed 3D FC-DenseNet and Dice loss function. To the best of our knowledge, this is the first work that employs 3D CNNs to address the challenge of acute ischemic stroke segmentation from DWI, and achieves better performance than 2D CNN models, especially in terms of reducing false positive artifacts.
3) We extensively validated the proposed method on a DWI&ADC dataset with various types of acute ischemic stroke. Experiment results show that our method outperforms several state-of-the-art 2D or 3D CNN models on our application. In particular, compared with the 2D counterparts, the proposed 3D method is more robust in distinguishing stroke lesions from artifacts.
4) We also evaluated the proposed model on a MICCAI challenge dataset for sub-acute ischemic stroke segmentation from multi-modal MRIs (i.e. ISLES2015-SSIS). On this dataset, our method also achieved very competitive performance, which further evidenced its effectiveness and generalization capability.

The remainder of this paper is organized as follows. In Section II, we describe the detail of the proposed 3D FC-DenseNet, Dice objective function and other training techniques. Experiment setups and results are reported in Section III. We further discuss some key issues and limitations of the proposed method in Section IV, and finally draw conclusions in Section V.

(a) (b) (c)

Fig. 3. Importance of 3D contextual information. (a) is a 2D axial slice with a stroke lesion (in green circle) and a hard mimic (in red circle). (b) is its adjacent slice, with a red circle denoting the corresponding area of the hard mimic in first slice. (c) shows the first slice overlaid by manual segmentation. Best viewed in color.

II. Methods

In Fig. 2 we illustrate the architecture of our proposed 3D FC-DenseNet. Our model is basically a 3D FCN, equipped with dense connectivity and integrated with multi-scale contextual information. We employ deep supervision technique and Dice objective function to improve its optimization. In the remainder of this section, we first outline the 3D CNN and its fully convolutional variant (Section II-A), then provide a detailed description of DenseNets (Section II-B), and elaborate the proposed framework (Section II-C) as well as the training techniques (Section II-D). We describe the evaluation metrics in Section II-E.

A. 3D Convolutional Neural Network

For volumetric medical data, 3D patches carry much more information than independent 2D slices. Some key issues such as anatomical locations and pathological features can be easily identified from 3D context but are non-trivial for 2D slices. As a concrete example, on the 2D slice shown in Fig. 3(a), the artifact in the red circle resembles stroke lesions very much, but when we inspect its 3D context (Fig. 3(b)), it is much easier to decide that this is a common susceptibility artifact caused by air, because this area is on the brain edge and next to an unusually bright region also along the brain edge. In order to fully leverage 3D contextual information in volumetric medical data, we implement a 3D CNN.

To adapt CNN models to 3D data, all layers should perform in a 3D manner or be compatible with 3D operations. As the main component of 3D CNNs, 3D convolutional layers perform the following operation:

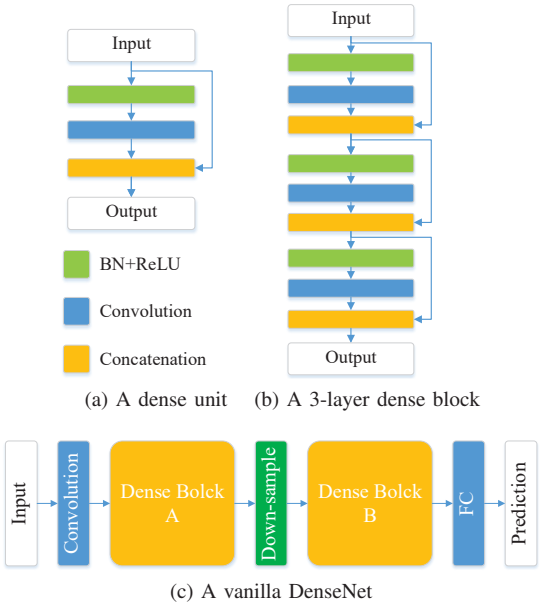
$$\boldsymbol{h}_i^{l+1} = \sigma\left(\sum_j \boldsymbol{u}_{ji}^{l+1} + \boldsymbol{b}_i^{l+1}\right) \quad (1)$$

$$\boldsymbol{u}_{ji}^{l+1}(x,y,z) = \sum_{m,n,t} \boldsymbol{h}_j^l(x+m, y+n, z+t) \cdot \boldsymbol{W}_{ji}^{l+1}(m,n,t) \quad (2)$$

where $\boldsymbol{u}_{ji}^{l+1}(x,y,z)$ is a 3D convolution with flipped kernel $\boldsymbol{W}_{ji}^{l+1}$, and $\boldsymbol{h}_i^l$ is the i-th channel of the l-th layer, $\boldsymbol{b}_i^{l+1}$ is a bias term. By hierarchically stacking 3D convolutional layers and down-sampling layers (e.g. pooling layers or convolutional layers with strides), a deep 3D CNN can efficiently extract high-level 3D features, which are necessary for solving

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging



(a) A dense unit (b) A 3-layer dense block

(c) A vanilla DenseNet

Fig. 4. Key concepts in a DenseNet. Best viewed in color.

complicated problems on volumetric data, e.g. acute stroke segmentation from DWI.

Another concern is that CNNs, which are originally designed for image recognition, do not allow end-to-end training and inference needed by segmentation tasks. To address this problem, we build our 3D CNN model based on the FCN framework [19], which enables the network to take arbitrary-sized inputs and produce equal-sized outputs. Specifically, we first replace the fully connected (FC) layers in the 3D CNN with $1\times1\times1$ convolutional layers, so that the network can take arbitrary-sized input. Then, to enable the network to produce an equal-sized prediction map, we add deconvolutional (DC) layers to restore the spatial size of internal feature volumes.

B. Densely Connected Convolutional Network

Convolutional layers are a main contributor to the strong expressive power of CNNs. Theoretical investigations have evidenced that, with reasonable size, a deeper computation network is significantly more expressive than its shallower counterparts [27], [28]. It is widely assumed that this theory also holds for CNNs. However, constructing very deep CNNs by simply stacking convolutional layers is not feasible, since too many layers can significantly impede the propagation of gradients, which is known as the *vanishing-gradient* problem [29]. Deep residual learning [14], [30] partially addresses this problem by introducing skip connections bypassing each residual blocks. Nevertheless, there exist a large amount of redundant features in a residual network (ResNet) since most residual blocks only slightly change the input signal [31]. Besides, as these skip connections are incorporated into residual networks by summation, the information flow can still be impeded.

To further boost the information flow and efficiently reuse features, we adopt a DenseNet [25] as the main pathway of our model (Fig. 2), which is built upon the idea of feature reuse and dense connectivity. In a DenseNet, each layer takes as input the feature maps from all its preceding layers. Fig. 4(b) illustrates the basic implementation of a DenseNet, where each dense unit is regarded as one layer. Formally, the output of the l-th layer (i.e. the l-th dense unit) is given by:

$$\boldsymbol{h}_l = [\mathcal{F}(\boldsymbol{h}_{l-1}), \boldsymbol{h}_{l-1}] \tag{3}$$

where $\boldsymbol{h}_l$ is the output of the l-th layer, $\mathcal{F}_l(\cdot)$ denotes the nonlinear mapping performed by this layer, and $[\cdots]$ denotes the concatenation of all inside tensors. Note that all these concatenated tensors should have the same size (except for the channel dimension). The formulation in (3) is equivalent to

$$\boldsymbol{x}_l = \mathcal{F}([\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{l-1}]) \tag{4}$$

where $\boldsymbol{x}_l$ is the newly produced feature maps of the l-th layer. The relation between $\boldsymbol{x}_l$ and $\boldsymbol{h}_l$ is

$$\boldsymbol{h}_l = [\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_l]. \tag{5}$$

From (4), it can be observed that any two layers in a DenseNet are directly connected, thus information and gradient can propagate throughout the network with no impediment. Moreover, since the growth rate (see the definition below) is usually small (only 16 in our experiment), the total number of feature maps in a DenseNet is far less than that in a residual network.

Before constructing a DenseNet, we clarify some key concepts as follows.

1) Dense Unit: A dense unit, as shown in Fig. 4(a), performs the operation given by (3), where $\mathcal{F}(\cdot)$ is a composite function consisting of batch normalization (BN) [32], nonlinear activation (ReLU [33] in our experiment) and a convolution.

2) Growth Rate: Supposing each composite function $\mathcal{F}(\cdot)$ produces k feature maps, we refer to k as *growth rate*, as the number of feature maps grows k per layer. For example, if the input has k_0 channels, the l-th layer will have $k \cdot (l-1) + k_0$ input feature maps.

3) Dense Block: Since the concatenation operation requires equal-sized feature maps, a DenseNet (Fig. 4(c)) is divided into several dense blocks (Fig. 4(b)) by down-sampling layers, where each dense block consists of multiple dense units.

C. 3D FC-DenseNet for Acute Stroke Segmentation

In view of the success of DenseNets on natural image datasets (e.g. CIFAR [34], SVHN [35] and ImageNet [36]), we extend them to 3D, and tap their potential on the challenging acute ischemic stroke segmentation task. The architecture of our neural network model is shown in Fig. 2.

In our model, there are four stacked dense blocks, each consisting of four dense units with a growth rate of 16. We employ small kernels (i.e. $3\times3\times3$) in convolutional layers as suggested by [11] and [13], which are demonstrated to be more efficient in computation and number of parameters. In between dense blocks, three convolutional layers with an stride of 2 (for the first down-sampling layer, $1\times2\times2$) are used to reduce the resolution size of input volumes, and to compress the number of feature volumes with a factor of 0.5. This enlarges the receptive field of our model and lowers memory

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

footprint. We use the anisotropic stride (i.e. 1×2×2) in the first down-sampling layer in order to reserve more details along the vertical axis since the slice spacing of the input DWI scans is very large. Note that we do not employ the more common pooling layers for down-sampling because their shift-invariance property is not desired in semantic segmentation.

Because of the large variations in lesion appearance and size, both fine details and large-scale context are necessary for the accurate segmentation of acute stroke. Therefore, we employ a multi-scale feature fusion scheme following [19] and [24]. Specifically, we add an auxiliary classification pathway (the second dashed box in Fig. 2) on the top of each dense block, so that we get four prediction volumes based on different image scales. Then, we fuse these four prediction volumes by summation, yielding a fine prediction that respects large context. In an auxiliary classification pathway, the starting FC layer is used to decrease the number of feature volumes for improving model compactness; the following deconvolutional layers are responsible for restoring resolution sizes and further reducing feature volumes; and the final FC layer with softmax nonlinearity calculates the prediction volume. Note that all these FC layers have been converted to $1\times1\times1$ convolutional layers, but we still refer to them as FC layers just for convenience.

D. Training Techniques

As depicted in Fig. 2, our model contains 20 volumetric convolutional layers in the main pathway. Such a deep 3D CNN poses a severe challenge to the training approach, especially when only limited training data are available. To address this challenge, we exploit deep supervision (DS) technique [37] to ease training and accelerate convergence, and utilize Dice objective function to bypass the severe class imbalance problem.

1) Deep Supervision: DS is motivated by the observation that a classifier trained on highly discriminative features performs better than that trained on less discriminative features [37]. For small dataset DS serves as a strong regularization, and in very deep CNNs the technique boosts gradient back-propagation, hence improves the convergence behavior.

To integrate DS into our 3D FC-DenseNet, we add a companion loss on the top of each auxiliary classification pathway (the third dashed box in Fig. 2). Thus, the model has four companion losses (Loss 1-4) and one main loss (Loss 0). We combine them and then the total objective function is

$$\begin{aligned}\mathcal{L}(I,G;\theta) = {} & C(\mathcal{H}(I;\theta),G) \\ & + \sum_{\alpha} w_\alpha C(\mathcal{H}_\alpha(I;\theta),G) + \lambda\psi(\theta)\end{aligned} \tag{6}$$

where I is the input 3D image, G is the corresponding ground truth, θ denotes the parameters of the model, $\psi(\theta)$ calculates the L2-norm of model parameters, α indicates the index of companion loss or auxiliary classification pathway, $\mathcal{H}(I;\theta)$ or $\mathcal{H}_\alpha(I;\theta)$ denotes the prediction volume for input image I, and $C(X,Y)$ represents the cost function, e.g. cross-entropy (CE) loss or Dice loss, which measures the similarity between model output X and ground truth Y. The first two terms on the right side of (6) constitutes the fidelity term, in which companion loss functions are weighted by w_α (we set $w_1, w_2, w_3, w_4 = 1$ and gradually decay them to 10^{-3}). The last term is the regularization term whose relative importance is controlled by the hyperparameter λ (which is set to 0.0005 in our experiment).

2) Dice Loss Function: Aside from the quantitative limitation of training data, another challenge posed by data is class imbalance. The class imbalance challenge is two-fold. First, there are much more negative voxels (whose ground truth label is 0, i.e. background) than positive voxels (whose label is 1, i.e. stroke lesion). Second, the ratio between the two classes (class imbalance ratio) can vary a lot among different subjects.

For semantic segmentation, the conventional CE cost function is defined as:

$$C_{CE} = \frac{1}{N}\sum_{i=1}^{N} \log p_{i,o_i} \tag{7}$$

where N is the voxel count, p_{i,o_i} is the predicted probability that the i-th voxel is assigned with the correct label o_i. From (7) it can be observed that each voxel contributes equally to the CE loss, thus the class imbalance problem can easily bias the model optimization. Moreover, we believe that a good cost function for semantic segmentation should consider each image as a whole, rather than regard an image as a set of independent voxels.

To avoid the disadvantages of CE loss, we propose to employ a subject-wise soft Dice objective function to train our model. Dice loss measures the relative overlap between the predicted probability map and the ground truth mask. It does not take into account true negative voxels so that the class imbalance problem in binary segmentation cannot affect Dice loss. Besides, Dice loss is a comprehensive measurement of the whole predicted probability map, instead of a simple average value across all voxels. Similar to [26], we define Dice cost function as:

$$C_{Dice} = 1 - \frac{2\sum_i o_i y_i + \epsilon}{\sum_i o_i + \sum_i y_i + \epsilon} \tag{8}$$

where $o_i \in [0,1]$ is the predicted value (softmax output of class 1) at location i, $y_i \in \{0,1\}$ is the corresponding ground truth label, and ϵ is a very small constant that is used to keep numerical stability and let the function has correct behavior when $\forall i, o_i, y_i = 0$. The derivative of Dice loss is then:

$$\frac{\partial C_{Dice}}{\partial y_i} = \begin{cases} \frac{2\sum_k o_k y_k}{(\sum_k o_k + \sum_k y_k)^2}, & o_i = 0, \\ \frac{2[\sum_k o_k y_k - (\sum_k o_k + \sum_k y_k)]}{(\sum_k o_k + \sum_k y_k)^2}, & o_i = 1. \end{cases} \tag{9}$$

For a mini-batch of size M, the Dice cost function is $C^B_{Dice} = \frac{1}{M}\sum_k C^{(k)}_{Dice}$ with k indexing subjects. There is a notable difference between our definition and that in [26]: our Dice loss is calculated on each 3D image, whereas [26] calculates it per batch of 2D slices. By our definition, the minimization of the objective function is basically equivalent to the maximization of average DSC across the training set.

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

E. Evaluation Metrics

To thoroughly evaluate the proposed model, we employ three kinds of criteria to evaluate our method, i.e. segmentation metrics, lesion-wise analysis and a quantification metric. The segmentation metrics include Dice similarity coefficient (DSC) and sensitivity (SE), which are defined as follows:

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}, \tag{10}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{11}$$

where TP, FP, TN and FN are the number of true positive, false positive, true negative and false negative voxels respectively.

Since the segmentation of large and small stroke lesions are equally important in clinical practices, we introduce lesion-wise criteria as additional metrics. We calculate the number of false positive lesions (#FPL) and false negative lesions (#FNL) per subject, as well as lesion-wise precision (Precision-L), recall (Recall-L) and F1 score (F1-L). These metrics are obtained using 3D connected component analysis. A false positive lesion is defined as a 3D connected component in the prediction volume that has no overlap with any connected component in the ground truth volume. False negative lesions are defined in a similar way by exchanging the position of prediction and ground truth masks.

We also utilize absolute relative volume difference (ΔV) to assess the consistency between true and predicted lesion volumes.

III. Experiments and Results

A. Dataset

We built a relatively large dataset of DWI and ADC images for acute ischemic stroke segmentation, containing 242 patients diagnosed with various types of acute ischemic stroke. Brain MRI was performed within 1 week of hospital admission. The mean (±std) age of included patients is 67.89 (±10.58), ranging from 35 to 90. There are 146 males and 96 females in the dataset. Among all MRI scans, 128 were obtained from a 1.5-T scanner (Sonata; Siemens Medical, Erlangen, Germany), and 114 from a 3.0-T scanner (Achieva 3.0 T TX Series; Philips Medical System, Best, the Netherlands), both with standard protocol. Some acquisition parameters are as follows: b-value: $1000\,\mathrm{s/mm^2}$; slice spacing: 5.5 or 6.5 mm; repetition time: 2300–2600 ms; echo time: 66–73 ms; flip angel: 90°; in-plane pixel spacing: 1.80×1.80 or 0.90×0.90 mm; dimension: $(19-25) \times 128 \times 128$ or $(19-25) \times 256 \times 256$. ADC maps were automatically calculated with a baseline T2 acquisition ($b = 0\,\mathrm{s/mm^2}$) and a DWI acquisition ($b = 1000\,\mathrm{s/mm^2}$) using the following equation

$$ADC(x, y, z) = \frac{\ln I(x, y, z) - \ln I_0(x, y, z)}{b} \tag{12}$$

where I and I_0 are the intensity of DWI and corresponding T2 scans respectively, $b = 1000\,\mathrm{s/mm^2}$ is the b-value of the DWI acquisition. The dataset was labeled by an experienced neurologist (Dr. Wenyan Liu), and was verified by another experienced rater (Dr. Lei Zhao). The label criterion is the presence of hyperintense DWI lesion with corresponding hypointensity in the ADC map [5].

TABLE I
Stroke Type Statistics

Patient Count		Stroke Subtype				Total
		L	S	C	Others	
Set	Training	47	28	8	7	90
	Validation	32	22	5	3	62
	Testing	47	28	9	6	90
Total		126	78	22	16	242

Note: L—Large-artery atherosclerosis, S—Small-artery occlusion, C—Cardioembolism

We randomly split the dataset into three subsets, with 90, 62, and 90 subjects for training, validation and testing respectively, and the proportion of each stroke type was balanced in different subsets. Details of stroke types in these three sets can be referred to in Table I, where stroke is classified according to Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria [38]. As preprocessing, we normalized each type of image to zero-mean and unit-variance by the statistics calculated from the training set. Each 3D image was first zero-padded or cropped to 24 slices, and was then resampled to $24 \times 128 \times 128$ for convenience. During training, each image was randomly cropped to $24 \times 80 \times 80$ and flipped along the three axes on the fly, so as to augment the dataset and reduce memory footprint. DWI and ADC volumes are concatenated along the fourth axis and fed into CNN models as 2-channel images. All the models evaluated in this section (except for the one in Section III-E) take both DWI and ADC volumes as input.

B. Efficacy of 3D FC-DenseNet

To investigate whether dense connectivity benefits CNN models in acute stroke segmentation, we compared the proposed 3D FC-DenseNet with two kinds of current state-of-the-art 3D CNN models in MIA domain.

1) VoxResNet and 3D U-ResNet: We chose two types of 3D CNN models for comparison: VoxResNet [24] and 3D U-Net with residual connections [23] (we call it 3D U-ResNet in this paper). Both models have achieved state-of-the-art performance on volumetric medical image segmentation tasks such as brain segmentation [24], prostate segmentation [23], etc. We first applied the original models proposed by Chen *et al.* [24] (VoxResNet) and by Yu *et al.* [23] (3D U-ResNet) to our dataset. And then we adapted these models based on the characteristics of our stroke lesion segmentation application. For VoxResNet, we inserted a group of residual blocks before the first downsampling layer, in order to retrieve more details in final prediction. For 3D U-ResNet, we first replaced pooling layers with downsampling convolutions, and then added one more residual blocks to each residual group and lowered the number of feature volumes in each layer, yielding a deeper but thinner model. To reserve more details along the low-resolution vertical axis, we utilized anisotropic stride (i.e.

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

(a) VoxResNet (b) 3D U-ResNet

Fig. 5. Architecture of adapted VoxResNet and 3D U-ResNet. Each residual group contains two residual blocks. The number following the name of each layer/group indicates its number of feature volumes.

$1\times2\times2$) in the first downsampling layer[1] for the two adapted models, the same as our proposed 3D FC-DenseNet. Fig. 5 schematically illustrates the architectures of these two adapted models.

2) Comparison Results: The performance of VoxResNet, 3D U-ResNet and the proposed 3D FC-DenseNet is listed in Table II. All these models were trained for 2000 epochs with Dice objective function, though in the literature VoxResNet and 3D U-ResNet were optimized with conventional CE loss. The application-specific adaption for VoxResNet and 3D U-ResNet effectively improves their performance by about 3%–4% in DSC, while the proposed 3D FC-DenseNet outperforms the adapted models by nearly 2% in DSC, and achieves the best sensitivity, volume difference ΔV as well as lesion-wise precision and F1 score. The comparison results evidence the efficacy of the proposed 3D FC-DenseNet model, indicating that dense connectivity is beneficial to CNN models in acute stroke segmentation. It is also noted that the 3D FC-DenseNet has a much smaller number of parameters, which indicates that the feature reuse scheme of DenseNets can effectively reduce the number of parameters without compromise on model expressiveness.

C. Dice vs Cross-Entropy Objective Function

We employed Dice objective function to address the class imbalance problem in data. To verify the efficacy of Dice loss, we compared the performance of several 3D CNN models

[1]The stride size of corresponding upsampling layers was modified accordingly. Since the anisotropic stride lowers the kernel size of DC layers (e.g. $2\times2\times2$ to $1\times2\times2$), the number of parameters in adapted models (e.g. VoxResNet-apt) is also reduced.

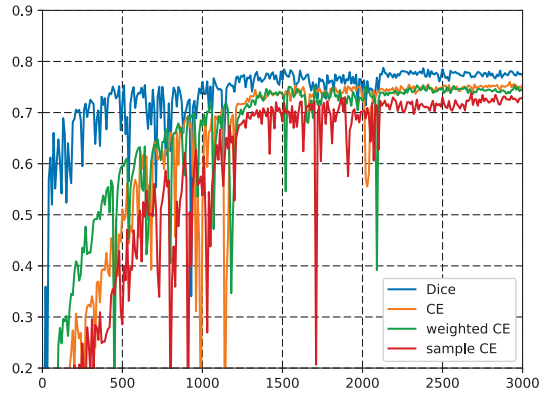


Fig. 6. Mean DSC on validation set throughout the training progress. Learning rate was decayed by 0.3 after 1200, 2100 and 2700 epochs. Best viewed in color.

trained with Dice or CE loss functions. Table III shows the performance of adapted VoxResNet and 3D U-ResNet trained with CE loss, as well as the performance of 3D FC-DenseNet optimized by several variants of CE loss (e.g. weighted and sample CE) and by Dice loss. The class weights of weighted CE were set to 0.3 and 0.7 for background and lesion classes respectively, which enforced the model to lay more stress on lesion voxels whose quantity was far smaller than background's. The sample scheme for sample CE loss was similar to the scheme used in [11], which equally extracted large image patches ($24\times80\times80$ in our experiment) centered on lesion or background voxels. Since we observed that the training with CE loss converged slower, we lengthened the training phase from 2000 to 3000 epochs.

From Table III, it can be observed that the performance of the proposed 3D FC-DenseNet degrades by 2%–4% in DSC when trained with CE loss or its variants. Among the three 3D FC-DenseNets optimized by different CE loss functions, the one trained with weighted CE achieves the highest sensitivity but has a lower DSC and worse volume difference ΔV, indicating that this model may be biased to the foreground (lesion) class. The model trained with sample CE has even worse performance in terms of DSC. These observations suggest that neither attaching more importance to one class nor sampling more positive patches is a good solution to the class imbalance problem in data. In contrast, when trained with Dice loss, our model not only achieves higher performance, but also has a better convergence behavior (Fig. 6). This demonstrates that Dice loss is an effective method for addressing the class imbalance problem in our application.

In Fig. 6 we plot the mean DSC over the validation set throughout the training progress. It is observed that the model trained with CE loss (or its variants) converges much slower, and their training curves are less stable. This phenomenon can be explained by the fact that CE loss is more susceptible to class imbalance problem, which can bias each single update step of model parameters towards either positive or negative, thus disturbs and slows down the convergence.

Moreover, we also notice that, when trained with CE

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

TABLE II
PERFORMANCE OF DIFFERENT 3D CNN MODELS TRAINED WITH DICE LOSS

Model	#param	DSC	SE	ΔV	#FPL	#FNL	Precision-L	Recall-L	F1-L
VoxResNet	3.6M	74.62%	72.49%	29.84%	0.61	1.11	87.86%	86.86%	87.34%
3D U-ResNet	15.2M	73.03%	70.74%	43.16%	0.71	1.22	83.67%	83.51%	83.59%
VoxResNet-apt	2.7M	77.52%	77.33%	26.74%	0.70	**1.09**	85.91%	**87.12%**	86.51%
3D U-ResNet-apt	13.2M	77.32%	75.66%	24.63%	**0.31**	1.16	91.95%	85.04%	88.36%
3D FCDenseNet	1.1M	**79.13%**	**78.15%**	**23.75%**	0.40	1.16	**92.67%**	86.08%	**89.25%**

TABLE III
MODEL PERFORMANCE WITH CE LOSS

Model	Loss Type	DSC	SE	ΔV	#FPL	#FNL	Precision-L	Recall-L	F1-L
VoxResNet-apt	CE	71.82%	67.92%	34.53%	0.32	1.39	90.83%	80.97%	86.52%
3D U-ResNet-apt	CE	70.90%	66.78%	34.21%	0.37	1.51	89.34%	78.37%	83.50%
3D FCDenseNet	CE	76.91%	75.13%	24.87%	0.43	1.16	89.93%	87.08%	88.48%
3D FCDenseNet	weighted CE	76.13%	**84.72%**	38.53%	0.37	**1.14**	91.41%	**87.14%**	89.22%
3D FCDenseNet	sample CE	75.12%	76.67%	32.93%	**0.31**	1.31	90.51%	83.71%	86.97%
3D FCDenseNet	Dice	**79.13%**	78.15%	**23.75%**	0.40	1.16	**92.67%**	86.08%	**89.25%**

loss, the performance degradation (compared to Dice loss) of adapted VoxResNet and 3D U-ResNet is severer than that of 3D FC-DenseNet. Without Dice loss, 3D FC-DenseNet's performance degrades to a DSC of 76.91% (decreasing about 2% relative to the one trained with Dice loss), whereas the performance of other two kinds of model largely decreases by 6% to about 71% in DSC. This observation not only evidences the effectiveness of Dice loss, but also indicates that dense connectivity can make the model more robust to the class imbalance problem.

D. 3D vs 2D Models

To investigate whether the proposed model can get benefits from its 3D nature and volumetric input data, we compared it with its 2D version, i.e. 2D FC-DenseNet, which had the same architecture as 3D FC-DenseNet but all its layers performed in a 2D manner. Besides, as simply converting 3D layers to 2D drastically reduced the model's complexity, which might affect its expressive power, we also built and evaluated another 2D model with more layers and parameters, i.e. 2D FC-DenseNet-Large, which had 8 dense units in each dense block. In addition, we implemented EDD Net [9] on our dataset for a direct comparison[2], which was a state-of-the-art acute ischemic stroke segmentation model based on 2D CNN. Since the subject-wise definition of Dice loss is not valid for independent 2D slices, and simply transforming it to a slice-wise manner led to bad performance, we trained the 2D models with sample CE loss, which was found to be better than conventional CE loss for 2D models in our preliminary experiments.

The evaluation results of 2D models are shown in Table IV. It is observed that 2D models can also achieve good performance in terms of DSC, but they have a noticeably lower lesion-wise precision (Precision-L) and larger number of false positive lesions (#FPL). The best 2D model (i.e. 2D FC-DenseNet-Large) has a lesion-wise precision of about 81.06%, and a #FPL of 1.07. By comparison, the proposed 3D method achieves a high lesion-wise precision of 92.67% and a low #FPL of 0.40. This suggests that the 3D model performs significantly better in avoiding false positive artifacts, which is one of the key obstacles in the task of acute stroke segmentation from DWI. This observation is in accordance with our assumption in Section II-A, which assumes that 3D contextual information is important in discriminating artifacts from ischemic lesions. Besides, though the 3D model introduces a few more false negatives (FNs), it shows a considerable gain on lesion-wise F1 score (89.25% vs 85.25%), which indicates that the 3D model reaches a better trade-off between false positives (FPs) and FNs. In conclusion, although 3D CNN models are more difficult to train than 2D ones, by leveraging the powerful dense connectivity structure and Dice loss, the proposed 3D FC-DenseNet can efficiently exploit the 3D contextual information in volumetric input data, which enables the 3D model to reject false positive artifacts more robustly and to outperform the 2D rivals on a variety of metrics (e.g. DSC, #FPL, lesion-wise precision, F1 score, etc.).

E. Evaluation on a Public Dataset

To further demonstrate the efficacy of the proposed segmentation algorithm, we evaluated our model on a public dataset, i.e. ISLES2015-SSIS [10], a MICCAI challenge dataset for the segmentation of sub-acute ischemic stroke lesions from multi-modal MR images. ISLES2015-SSIS dataset contains 28 subjects for training and 36 for testing. Each subject has four co-registered MRI sequences, i.e. Flair, DWI, T1 and T2, which are skull-stripped and re-sampled to an isotropic spacing of $1 \times 1 \times 1$ mm. The evaluation metrics of the ISLES challenge include average symmetric surface distance (ASSD), DSC, precision and recall. The evaluation results can be automatically calculated by the challenge website[3] after uploading the segmentation results of testing data.

[2]We have also attempted to implement MUSCLE Net, which was employed to further refine the output of EDD Net in [9]. However, in our experiments the performance of EDD+MUSCLE Net was similar to or even worse than that of single EDD Net, which might be caused by the overfitting problem in the training of MUSCLE Net. Therefore, we only report the performance of EDD Net.

[3]http://www.isles-challenge.org/ISLES2015/

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

TABLE IV
PERFORMANCE OF 3D AND 2D MODELS

Model	#param	DSC	SE	ΔV	#FPL	#FNL	Precision-L	Recall-L	F1-L
EDD Net	3.2M	77.03%	**80.07%**	31.06%	1.07	1.00	77.51%	88.39%	82.59%
2D FC-DenseNet	0.43M	76.77%	75.06%	26.05%	1.24	0.96	77.10%	88.99%	82.62%
2D FC-DenseNet-Large	1.5M	78.03%	80.06%	26.53%	1.07	**0.83**	81.06%	**89.90%**	85.25%
3D FC-DenseNet	1.1M	**79.13%**	78.15%	**23.75%**	**0.40**	1.16	**92.67%**	86.08%	**89.25%**

Since the data specification (e.g. resolution, voxel spacing, etc.) of SSIS is different from our dataset's and the feature of sub-acute stroke lesions is more complicated than acute ones, we performed some adjustments before evaluating our model on SSIS data. First, all downsampling strides were set to isotropic, i.e. 2×2×2. Second, since the data resolution size was high, the first dense block was replaced by 4 convolutional layers (with BN and ReLU) each having 32 feature volumes, in order to reduce GPU memory requirement. Third, the depth of each dense block was enlarged to 10 so as to account for the more complicated features of sub-acute stroke lesions, and the last dense block (as well as the associated auxiliary pathway and downsampling layer) was removed to reduce computational complexity. Other model settings are the same as the original 3D FC-DenseNet (Fig. 2). As for objective function, we utilized a hybrid loss which added Dice loss and CE loss together, because we found that training with single Dice loss was not very stable[4] on this dataset, while the performance degraded if using CE loss only. During training, image patches of size $64 \times 64 \times 64$ centered on positive or negative voxels were equally extracted, and the image was randomly flipped along the three axes. The model was trained for 3000 epochs with an initial learning rate of 0.1, decayed by 0.2 after 1200, 2000 and 2500 epochs.

We evaluated our 3D FC-DenseNet on ISLES2015-SSIS dataset. The performance of the top 5 teams (according to DSC ranking) on the running leaderboard[5] (until Mar 7, 2018) is listed in Table V. Among all 28 entries, our team achieves very competitive results, having the best ASSD metric and ranking top 3 on each single measurement, and in terms of DSC we only slightly trail the first-rank team kamnk1 [11] (58% vs 59%). Note that our submission is the output of an ensemble of two 3D FC-DenseNet models without post-processing, while the result of kamnk1 is based on an ensemble of three DeepMedic models as well as a CRF post-processing step [11]. The highly competitive performance of our method on SSIS dataset further evidences the effectiveness of 3D FC-DenseNet and Dice loss, and demonstrates the good generalization capability of our method on different kinds of stroke segmentation tasks.

F. Qualitative Evaluation

Fig. 7 presents several challenging cases. The first row presents an example of multi-focal lesions with ill-defined boundaries; the second row shows a small lesion appeared on the edge of brain; the third row shows a large lesion with complex texture; and the last row presents artifact areas that mimic stroke lesions very much. Our 3D FC-DenseNet model achieves satisfactory segmentation results on all these cases, and can robustly reject hyper-intensity artifacts that mimic stroke lesions. In contrast, the 2D models (the fourth and fifth columns) do not perform very well on the last two cases.

TABLE V
EVALUATION RESULTS ON ISLES2015-SSIS

Team	ASSD (mm)	Dice	HD (mm)	Precision	Recall
kamnk1	7.87	**59%**	39.61	**68%**	60%
zhanr6 (ours)	**7.52**	58%	38.98	60%	**68%**
fengc1	8.13	55%	**25.02**	64%	57%
shenh1	10.96	51%	56.2	50%	66%
martc2	14.69	50%	80.06	55%	53%

G. System Implementation

We implemented the evaluated methods with PyTorch[6] framework in Python 3.6 on a PC with 16GB RAM, an Intel Core i7-4790 3.60GHz CPU, and an NVIDIA TITAN X GPU. We initialized these networks by the method proposed in [39], and trained them by stochastic gradient descent (SGD) method with initial learning rate $\eta_0 = 0.05$ to 0.1 (for the proposed model $\eta_0 = 0.05$), momentum 0.9 and weight decay $\lambda = 0.0005$. The learning rate was decayed by 0.3 after 800, 1400 and 1800 epochs (for the models trained for 2000 epochs), or after 1200, 2100 and 2700 epochs (for the models trained for 3000 epochs). The time and GPU memory requirement of training and inference for all evaluated models are listed in Table VI, where the time and memory requirement for inference are associated with the processing of a 2-channel input image of size $2 \times 24 \times 128 \times 128$.

IV. DISCUSSION

DenseNets were originally proposed for the classification of 2D natural images, and achieved significantly better performance than ResNets due to their novel dense connectivity paradigm. Dense connectivity not only improves the information and gradients propagation throughout the network, but also enables the network to efficiently reuse existing features, which can effectively reduce the number of parameters required by the model. With boosted information flow and less parameters, the network becomes easier to train and less prone to over-fitting problem. These properties are especially desired in medical applications, where training datasets are usually smaller while data dimensions can be higher. Therefore, we extend DenseNets to 3D and tap their potential on acute stroke

[4]The instability may be caused by the zero-derivative problem of Dice loss, which will be further discussed in Section IV.

[5]https://www.virtualskeleton.ch/ISLES/Start2015

[6]http://pytorch.org/

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

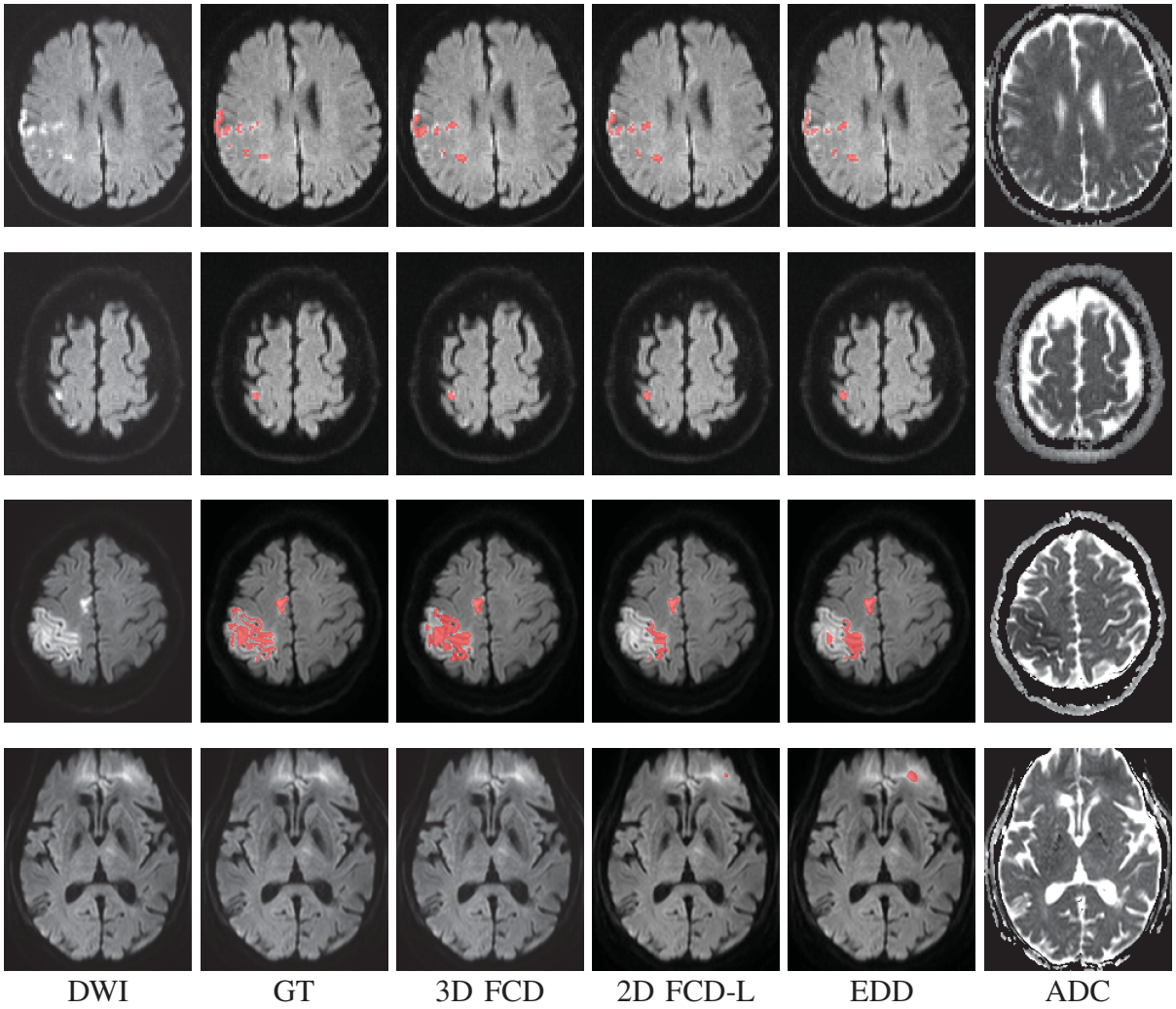


Fig. 7. Challenging cases in the test set and the corresponding segmentation. From left to right, the six columns show DWI slices, ground truth segmentation (overlaying on the corresponding DWI slice, similarly for the subsequent three columns), segmentation of 3D FC-DenseNet, segmentation of 2D FC-DenseNet-Large, segmentation of EDDNet, and ADC slices respectively. Best viewed in color.

TABLE VI
TRAINING AND INFERENCE DETAILS

Model	Dimension	Training Epochs	Batch Size	Training GPU Mem	Training Time	Inference Time	Inference GPU Mem
VoxResNet-Adapted	3D	2000	9	7078MB	6h24m	0.13s	844MB
3D-U-ResNet-Adapted	3D	2000	9	6860MB	8h5m	0.11s	898MB
3D-FCDenseNet	3D	2000	9	10612MB	6h23m	0.095s	1438MB
EDDNet	2D	3000	192	8074MB	8h10m	0.045s	654MB
2D-FCDenseNet	2D	3000	192	9654MB	6h34m	0.043s	1430MB
2D-FCDenseNet-Large	2D	2000	96	10722MB	8h35m	0.095s	3068MB

segmentation. Results of comparative experiments demonstrate the superiority of our method to other state-of-the-art 2D or 3D CNN models. Moreover, as the class imbalance problem can seriously bias the optimization process of 3D CNN models, we propose to train the model with soft Dice loss, which measures the relative overlap between the predicted probabilistic volume and ground truth mask. Our work demonstrates that 3D CNNs are a promising tool for the segmentation of lesions from volumetric images, e.g. acute stroke segmentation, and also corroborates that dense connectivity and Dice objective function are valuable components that can effectively boost the optimization and performance of 3D CNNs.

TABLE VII
MODEL PERFORMANCE WITH DIFFERENT INPUT

Input	DSC	SE	ΔV	#FP	#FN	Prec	Recl
DWI	77.14%	75.35%	22.37%	0.48	1.27	89%	83%
DWI,ADC	79.13%	78.15%	23.75%	0.40	1.16	93%	86%

One important issue of automatic acute stroke segmentation from DWI images is how to accurately reject the numerous artifacts present on DWI scans while retaining true stroke lesions. Most of previous works (e.g. [7], [8]) resorted to ADC maps by rejecting candidate voxels whose ADC value

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

exceeded a preset threshold. However, this kind of methods can be inaccurate and may lead to a severe decrease in sensitivity, since the ADC value of lesion voxels is dispersively distributed in a large range. Alternatively, CNN-based models can automatically learn how to discriminate artifacts according to their appearance, so their performance does not rely much on ADC maps. For verification, we trained and evaluated our model without the presence of ADC maps, and the results are shown in Table VII (we shorten some metric names and decimal digits to control the width of the table). In this case, our model still achieved a DSC of 77.14% and a lesion-wise precision of 89%, clearly corroborating its high expressive power.

In our experiment on ISLES2015-SSIS dataset, we noticed that the training with Dice loss only could be unstable sometimes, but the instability decreased when increasing the training patch size. We speculate that the instability is caused by the backward propagation of Dice objective function when the input patch has no positive voxels. As indicated by (9), when $\forall i, o_i = 0$ (i.e. no positive voxels in the image patch, a *totally negative patch*), the derivative $\frac{\partial \mathcal{C}}{\partial o_i}$ is always 0 whatever the value of y_i (i.e. the prediction of the model) is. The zero-derivative problem of Dice loss will let the model parameters update along the momentum rather than stay unchanged, which, however, might be a wrong direction. When the crop size of training samples is relatively small (e.g. $64 \times 64 \times 64$ from an around $160 \times 230 \times 230$ volume for SSIS data), the number of totally negative training patches can be larger, thus there are more chances that the model parameters are updated towards a wrong direction, leading to instability of the training. Therefore, we utilized additional CE loss to assist the training on SSIS dataset. The instability is a drawback of Dice loss. We will investigate new patch extraction strategies and better forms of Dice loss that can combat this problem in the future.

The multi-scale feature integration scheme employed by the proposed 3D FC-DenseNet is similar to the ones utilized by FCN [19] and by VoxResNet [24], which makes finer-grain intermediate predictions based on features from hidden layers, and then fuses intermediate and final predictions by summation. Compared to DeepMedic model [11], which utilizes two independent parallel pathways to tackle multi-scale information, our employed scheme can efficiently make use of the multi-scale features generated within the mainstream network, and can be integrated with deep supervision seamlessly.

Our method can be easily extended to multi-modal MRI sequences by combining multiple 3D images along the fourth axis. With extra information from other modalities, the segmentation of stroke lesions could be easier, but the resampling and co-registration among different modalities can also introduce inaccuracies. In particular, perfusion weighted imaging (PWI) can provide more information about blood flows, which is useful for the identification of penumbra and outcome prediction. However, PWI acquisition is slower and requires advanced MRI equipment, and its clinical significance in acute ischemic stroke is still controversial [40], so currently the acute stroke segmentation algorithms based on DWI are more applicable in clinical practices.

When implementing EDD Net, we noted that its performance on our dataset was higher than the reported performance in [9]. We speculate that this is because their employed DWI dataset has a lower SNR (as observed from the qualitative examples illustrated in [9]) and more artifacts, which may be caused by the higher b-values and different protocols in their DWI acquisition. In particular, for DWIs with lower SNR (e.g. resulted by the higher b-value in data acquisition), it would be even harder to discriminate artifacts from true lesions solely based on their in-plane (2D) patterns, and thus their 3D features could be even more critical, which, however, were ignored by those 2D models such as EDD Net.

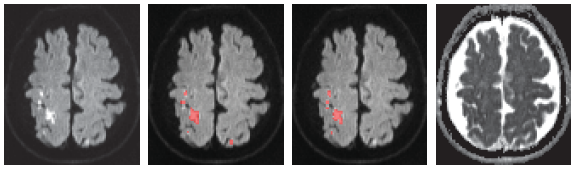


Fig. 8. An example of a stroke lesion missed by our method. Best viewed in color.

Although our method is highly accurate in acute stroke segmentation, there are still several limitations. First, since the employed dataset does not contain abnormal scans that have slices damaged by macroscopic patient movement, our 3D FCN model may not be robust to MRI images corrupted by severe motion artifacts. Second, in some very challenging cases, our method may miss several true lesions. Fig. 8 shows an example where the lesion in bottom-right is ignored by our method. The missed lesion appears along the edge of the brain (a feature of artifacts), and its intensity is relatively low compared with other lesions, so it is hard for our model to tell whether this region is a lesion or an artifact, especially when similar samples are rare in the training set. In the future, we will try to build a DWI dataset with motion artifacts, so as to evaluate and to improve the model's robustness to scans corrupted by patient motion. On the other hand, we will investigate how to borrow power from other medical or natural data, e.g. by transfer learning, to reduce our model's dependency on large annotated datasets and to improve its performance on infrequent lesion types.

V. Conclusion

In this paper, we present a novel deep 3D FCN model to meet the challenge of acute ischemic stroke segmentation from DWI. We integrate dense connectivity to boost information and gradient propagation within the deep 3D model, and optimize it with Dice objective function to tackle the severe class imbalance problem. By incorporating dense connectivity and Dice loss, the difficulty in training deep 3D CNN is effectively relieved, thus our model is able to outperform state-of-the-art 2D and 3D CNN methods on a variety of metrics. Extensive comparative experiments on our dataset corroborate the superiority of the proposed method. Moreover, we also evaluate the method on a public challenge dataset, where its effectiveness and generalization capability are further demonstrated. Because of its fast, accurate and robust performance, our method has a good potential in clinical practices, and

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2821244, IEEE Transactions on Medical Imaging

can serve as a preliminary step in a computer-aided diagnosis (CADx) system.

ACKNOWLEDGMENT

The authors sincerely thank Dr. Wenyan Liu for her efforts to carefully label the dataset.

The work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Poject No.: CUHK 14113214 and CUHK 14204117), and grants from the Innovation and Technology Commission (Project No: GHP-025-17SZ and GHP-028-14SZ).

REFERENCES

[1] "The top 10 causes of death." http://www.who.int/mediacentre/factsheets/fs310/en/, 2017. [Online; accessed 30-June-2017].

[2] H. B. Van der Worp and J. van Gijn, "Acute ischemic stroke," *New England Journal of Medicine*, vol. 357, no. 6, pp. 572–579, 2007.

[3] Ö. Saka, A. McGuire, and C. Wolfe, "Cost of stroke in the united kingdom," *Age and ageing*, vol. 38, no. 1, pp. 27–32, 2009.

[4] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, S. de Ferranti, J.-P. Després, H. J. Fullerton, *et al.*, "Heart disease and stroke statistics—2016 update," *Circulation*, vol. 133, no. 4, pp. e38–e360, 2016.

[5] J. Yang, A. Wong, Z. Wang, W. Liu, L. Au, Y. Xiong, W. W. Chu, E. Y. Leung, S. Chen, C. Lau, *et al.*, "Risk factors for incident dementia after stroke and transient ischemic attack," *Alzheimer's & Dementia*, vol. 11, no. 1, pp. 16–23, 2015.

[6] K. B. Prakash, V. Gupta, M. Bilello, N. J. Beauchamp, and W. L. Nowinski, "Identification, segmentation, and image property study of acute infarcts in diffusion-weighted images by using a probabilistic neural network and adaptive gaussian mixture model," *Academic radiology*, vol. 13, no. 12, pp. 1474–1484, 2006.

[7] N. Hevia-Montiel, J. R. Jimenez-Alaniz, V. Medina-Banuelos, O. Yanez-Suarez, C. Rosso, Y. Samson, and S. Baillet, "Robust nonparametric segmentation of infarct lesion from diffusion-weighted mr images," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 2102–2105, IEEE, 2007.

[8] S. Mujumdar, R. Varma, and L. T. Kishore, "A novel framework for segmentation of stroke lesions in diffusion weighted mri using multiple b-value data," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 3762–3765, Nov 2012.

[9] L. Chen, P. Bentley, and D. Rueckert, "Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks," *NeuroImage: Clinical*, 2017.

[10] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, *et al.*, "Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri," *Medical image analysis*, vol. 35, pp. 250–269, 2017.

[11] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[15] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[16] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[17] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.

[18] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[19] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432, Springer, 2016.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, pp. 565–571, IEEE, 2016.

[23] L. Yu, X. Yang, H. Chen, J. Qin, and P.-A. Heng, "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images.," in *AAAI*, pp. 66–72, 2017.

[24] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images," *NeuroImage*, 2017.

[25] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[26] M. Drozdzal, G. Chartrand, E. Vorontsov, L. Di Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, "Learning normalized inputs for iterative estimation in medical image segmentation," *arXiv preprint arXiv:1702.05174*, 2017.

[27] J. Håstad, "Computational limitations of small-depth circuits," 1987.

[28] O. Delalleau and Y. Bengio, "Shallow vs. deep sum-product networks," in *Advances in Neural Information Processing Systems*, pp. 666–674, 2011.

[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*, pp. 630–645, Springer, 2016.

[31] A. Veit, M. Wilber, and S. Belongie, "Residual networks are exponential ensembles of relatively shallow networks," *arXiv preprint arXiv:1605.06431*, vol. 1, 2016.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, pp. 448–456, 2015.

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

[34] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.

[35] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *NIPS workshop on deep learning and unsupervised feature learning*, vol. 2011, p. 5, 2011.

[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, IEEE, 2009.

[37] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial Intelligence and Statistics*, pp. 562–570, 2015.

[38] H. P. Adams, B. H. Bendixen, L. J. Kappelle, J. Biller, B. B. Love, D. L. Gordon, and E. Marsh, "Classification of subtype of acute ischemic stroke. definitions for use in a multicenter clinical trial. toast. trial of org 10172 in acute stroke treatment.," *Stroke*, vol. 24, no. 1, pp. 35–41, 1993.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[40] M. Goyal, B. K. Menon, and C. P. Derdeyn, "Perfusion imaging in acute ischemic stroke: let us improve the science before changing clinical practice," *Radiology*, vol. 266, no. 1, pp. 16–21, 2013.

0278-0062 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.