This is the Pre-Published Version.

The following publication A. Nazir et al., "OFF-eNET: An Optimally Fused Fully End-to-End Network for Automatic Dense Volumetric 3D Intracranial Blood Vessels Segmentation," in IEEE Transactions on Image Processing, vol. 29, pp. 7192-7202, 2020 is available at https://doi.org/10.1109/TIP.2020.2999854.

OFF-eNET: An Optimally Fused Fully End-to-End Network for Automatic Dense Volumetric 3D Intracranial Blood Vessels Segmentation

Anam Nazir, Muhammad Nadeem Cheema, Bin Sheng, *Member, IEEE*, Huating Li, Ping Li, *Member, IEEE*, Po Yang, *Senior Member, IEEE*, Younhyun Jung, Jing Qin, *Member, IEEE*, Jinman Kim, *Member, IEEE*, and David Dagan Feng, *Fellow, IEEE*

***Abstract*—Intracranial blood vessels segmentation from computed tomography angiography (CTA) volumes is a promising biomarker for diagnosis and therapeutic treatment in cerebrovascular diseases. These segmentation outputs are a fundamental requirement in the development of automated decision support systems for preoperative assessment or intraoperative guidance in neuropathology. The state-of-the-art in medical image segmentation methods are reliant on deep learning architectures based on convolutional neural networks. However, despite their popularity, there is a research gap in the current deep learning architectures optimized to address the technical challenges in blood vessel segmentation. These challenges include: (i) the extraction of concrete brain vessels close to the skull; and (ii) the precise marking of the vessel locations. We propose an Optimally Fused Fully end-to-end Network (OFF-eNET) for automatic segmentation of the volumetric 3D intracranial vascular structures. OFF-eNET comprises of three modules. In the first module, we exploit the up-skip connections to enhance information flow, and dilated convolution for detailed preservation of spatial feature map that are designed for thin blood vessels. In the second module, we employ residual mapping along with inception module for speedy network convergence and richer visual representation. For the third module, we make use of the transferred knowledge in the form of cascaded training strategy to gradually optimize the three segmentation stages (basic, complete, and enhanced) to segment thin vessels located close to the skull. All these modules are designed to be computationally efficient. Our OFF-eNET, evaluated using 70 CTA image volumes, resulted in 90.75% performance in the segmentation of intracranial blood vessels and outperformed the state-of-the-art counterparts.**

Manuscript received September 25, 2019; revised March 7, 2020 and April 30, 2020; accepted June 1, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61872241 and Grant 61572316, in part by the Science and Technology Commission of Shanghai Municipality under Grant 18410750700, Grant 17411952600, and Grant 16DZ0501100, in part by the Hong Kong Research Grants Council under Grant PolyU 152035/17E, and in part by The Hong Kong Polytechnic University under Grant P0030419 and Grant P0030929.

Anam Nazir, Muhammad Nadeem Cheema, and Bin Sheng are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: shengbin@sjtu.edu.cn).

Huating Li is with the Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, China (e-mail: huarting99@sjtu.edu.cn).

Ping Li is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: p.li@polyu.edu.hk).

Po Yang is with the Department of Computer Science, The University of Sheffield, Sheffield S1 4DP, U.K. (e-mail: po.yang@sheffield.ac.uk).

Younhyun Jung, Jinman Kim, and David Dagan Feng are with the Biomedical and Multimedia Information Technology Research Group, School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: dagan.feng@sydney.edu.au).

Jing Qin is with the Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong (e-mail: harry.qin@polyu.edu.hk).

***Index Terms*—Convolution neural network, computed tomography angiography, dilated convolution, inception module, up-skip connection, intracranial vessels segmentation.**

I. Introduction

INTRACRANIAL segmentation of blood vessels from computed tomography angiography (CTA) in neuropathology is of significant importance in diagnostic and clinical applications such as, stroke prevention and treatment, embolization of cerebral aneurysms, and arteriovenous malformations (AVMs) [1]. For example, to carry out improved embolization of the AVM, the structural and geometric information of blood vessels from 3D images is of supreme importance. Likewise, detailed information about arterial vessel status play a crucial role for both the prevention and the improvement of stroke therapy. For this reason, the in-depth segmentation of cerebral blood vessels [2] is required for guiding physicians in planning better pre-operative strategies, monitoring intraoperative surgical progression, and predicting postoperative patient outcome in neurosurgical analysis.

Considering the crucial importance of brain vessel status in a routine procedure, vessel information could be easily integrated in the clinical workflow, if in-depth segmentation methods are available and practically applicable [3], [4]. Possibly an ideal segmentation results can be obtained through manual delineation by a medical expert; however this is tiresome, impractical in larger studies, and subject to inter-observer inconsistencies. Despite technology advances in image segmentation [5], existing methods still suffer from low accuracy for in-depth vessel segmentation particularly to deliver segmentation of vessels close to skull. An analytic and diagnostic medical data often comprises 3D images, volumetric segmentation of such large volumes using slice-by-slice approach is very cumbersome. Additionally, aiming to learn 3D complex visual patterns with a low computational burden from volumetric images is an ongoing research challenge. Therefore, designing efficient and automatic 3D technique for learning visual representations for segmentation of fine structures in volumetric images is of significant interest.

Consequently, various methods have been proposed such as rule-based, which are implemented using vessel intensity distributions, geometric models to extract vessels for developing more robust and accurate automatic vessel segmentation methods [6], [7]. However, these approaches are based on

©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

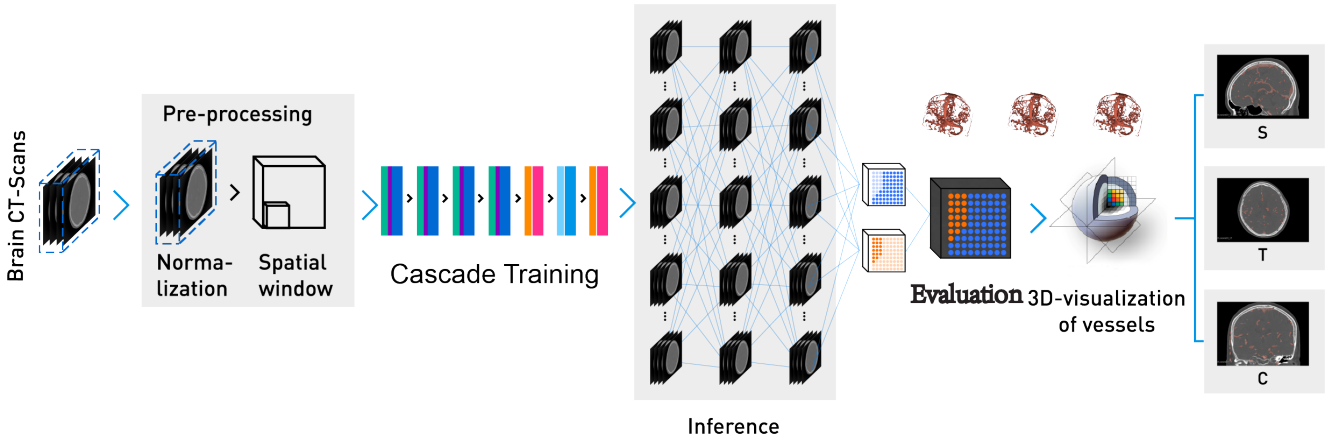


Fig. 1: Workflow of our proposed OFF-eNET for automatic intracranial blood vessels segmentation. First step is to collect brain CTA volumes and apply various preprocessing techniques such as normalization and re-sampling the dataset with spatial window using spatially balanced sampling. Secondly, end-to-end cascade training of CNN with a 22-layered architecture is carried out. Thirdly, our proposed CNN automatically segment and label brain vessels using inference. Finally, 3D slicer is used for generating and enhancing 3D representation of segmented blood vessels with respect to multiple 3D views with respect to sagittal (S), coronal (C), and transverse (T) planes.

manually engineered hand-crafted features and may prove to be inadequately validated. Convolutional neural networks (CNN) has shown to be promising tools for learning visual representations for image classification and segmentation tasks in various fields [8]–[11], and particularly in neuroimaging [12]. The distinguishing ability of CNN to learn hierarchical feature representation [13] without relying on manually detected features makes them suitable for automatic delineation of volumetric images. Multiple CNN implementations have been proposed, Zikic et al. [14] presented a pioneer work based on 2D CNN for segmentation of neural membranes. Further studies on CNN for automatic segmentation of brain images were introduced by [15]–[17]. Although these studies contributed reasonably in tailoring CNN to analyze volumetric images, the majority of the existing work studies image representations in 2D. Here, to carry out brain segmentation task from 3D CT, each 2D slice has to be independently processed, which is perhaps an impractical utilization of the volumetric representations [18]. Phellan et al. [19] showed promising preliminary results while exploring the small sample size with shallow net lead to limited performance. One of the most promising deep learning frameworks is the U-Net [20] specifically designed for segmentation tasks and has shown high performance for biomedical images [21]–[23]. Livne et al. [24] presented a revised U-Net for high-performance brain vessels segmentation [25]. In [26], Vesal et al. presented global context through the use of dilated convolutions in fully 3D CNN which helps in domain adaptation, and the overall segmentation accuracy for Left atrial segmentation in 3D gadolinium enhanced-MRI.

Another paradigm of fully 3D CNN architecture follows a pathway of convolution and deconvolution combination to accomplish high-resolution segmentation [27], [28]. S-abokrou et al. [29] proposed a two-stage cascade strategy for anomaly detection and localization in video data showing crowded scenes. A semi-supervised method for brain vessels segmentation using hierarchical CNN was proposed in [30] which showed better performance than classical rule-based segmentation models. In another study [31], DeepVesselNet was introduced that optimized for segmenting and analyzing vessels (centerline prediction, and bifurcation detection) using angiographic volumes. A voxel wise analysis for brain vascular segmentation using time-of-flight magnetic resonance angiography (TOF MRA) images based on trained CNN [19] was proposed for bi-dimensional manually annotated image patches. Another recent method [32] have utilized variability of blood flow signals to segment brain vessel from TOF MRA Images.

A 3D FCNN for subcortical segmentation of brain MRI [33] and for segmenting infant brain MRI [34] have proved remarkable performance. It uses small kernels and intermediates layers for local and global contextual information with efficient processing. It involves post processing to achieve concrete segmentation. A deep voxel wise residual network [35] is proposed for volumetric brain segmentation with greater efficiency. Despite the popularity of 3D CNN architectures, their computational cost and memory requirements during inference to deliver in-depth and robust segmentation is still an open challenge. To date, no 3D architecture is presented that is specifically designed for intracranial vascular segmentation to clearly separate blood vessels in areas closely adjacent to the skull. In this study, we have proposed an Optimally Fused Fully end-to-end network (OFF-eNET) for automatic dense volumetric 3D intracranial vascular segmentation from CTA volumes in areas closely adjacent to the skull. Fig. 1 depicts workflow of the proposed method. The main contributions of this study are summarized as three modules:

- For the segmentation of intracranial vessels near the skull, we propose a dilated convolution with cross-layer architecture as the up-skip connections. This combination enhances network connectivity between the convolution and dilated convolution path to facilitate the model's capacity to learn multilevel features as well as detailed

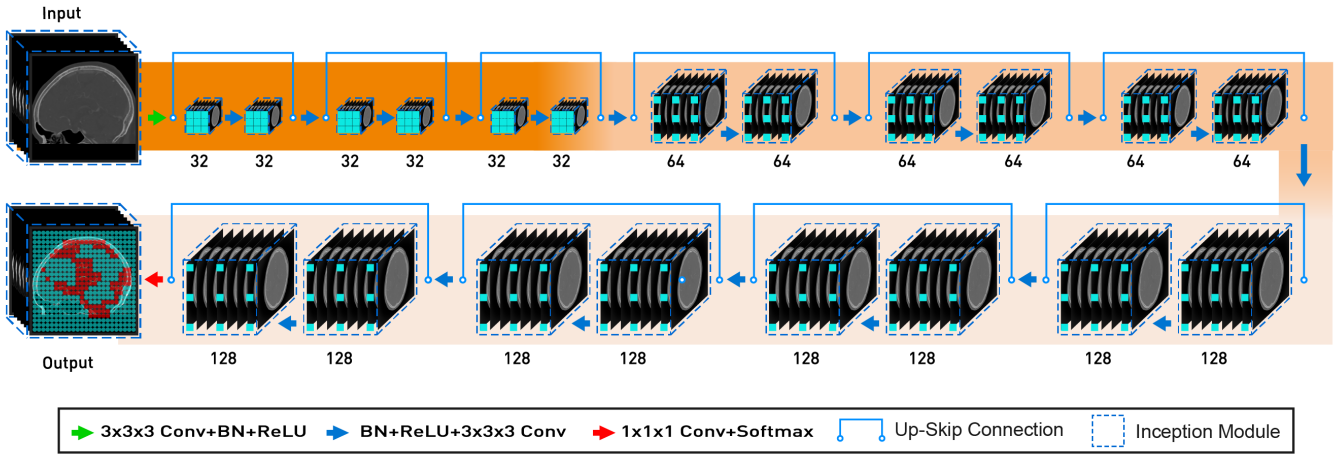


Fig. 2: The schematic representation of our proposed architecture. It consists of 22 layers with 10 residual connection blocks. Each block with an inception module is composed of two 3×3×3 paired convolution layers followed by batch normalization layer and element-wise rectified linear unit (ReLU) layer with up-skip connections. The first three blocks with 32×32 feature map shows the normal convolution layer designed to extract low-level volume features such as corners and edges, the three blocks in the middle with feature map of 64×64 convoluted layers have dilation factor of 2, and the last 4 blocks consist of 128×128 feature map comprising dilation factor of 4. The inputs are 128×128×128 voxel slices and fully connected (FC) last layer gives classification results for each respective input voxel.

preservation of spatial resolution feature map.

- To increase the network's capacity for learning richer representations, in addition to accelerate network convergence, we introduce an extension to the residual connection by adopting inception modules fusion of residual connection which enabled the proposed method to capture advanced visual information under controlled computational complexity.
- An optimized way of employing cascade training strategy is employed for using transferred knowledge to achieve concrete segmentation into three sub-stages (basic, complete, and enhanced).

Related to our problem scenario, the scene image segmentation approaches such as CCL [36], SVCNet [37], and BFP [38] tailored skip connections, dilated convolution and dilated FCN for contextual and sematic segmentation where to be detected objects are specified determined objects. OFF-eNET mainly deals with medical image segmentation where the to be segmented objects i.e. intracranial vessels are not clear and specified. Our problem scenario is significantly different from CCL [36], SVCNet [37], and BFP [38] in the way that we have personalized the use of skip connection and dilated convolution by introducing a fusing strategy to preserve very low level details from the image to segment vessels close to skull which are very thin and need segmentation of detailed features. In this context our central contribution is an improved and compact fully 3D end-to-end CNN architecture for automatic intracranial vascular segmentation from CTA volumes in areas closely adjacent to the skull. To achieve better performance, we have used small kernels which exploit the capability of dilated convolution for preserving fine structural details along with up-skip connections to accelerate information flow to ensure more low-level features. We further extend our architecture by an innovative fusion of residual mapping with inception modules to increase the network's capacity for learning richer representations in addition to accelerate network convergence under controlled computational complexity. The motivation of modifying residual connection and dilated convolution is that they both show good result in natural 2D image segmentation such as DeepLab [39], so we try them for 3D volume segmentation.

II. Approach

Fig. 1 shows our workflow, at first, we collect the brain CTA volumes, apply various preprocessing techniques such as normalization and re-sampling. The input volumes are sampled by a [128 × 128 × 128] spatial window from the CTA volumes. After a cascaded training strategy of network, we have obtained intracranial labeled blood vessels in three sub-stages (basic, complete, and enhanced) through inference and verified the automatic segmentation results by comparing them with manually labeled gold standard CTA volumes as well as using quantitative metrics analysis. We enhance the visualization of labeled blood vessels using 3D slicer version 4.7.0 [40], which reconstruct the vessels from different viewpoints. Further explanation for each component of the proposed network is given below.

A. Network Architecture for Fully 3D End-to-End CNN

Fig. 2 shows architectural design of our network. It consists of 22 layers with 10 residual connection blocks having an inception module associated to all 10 blocks. Each block is composed of two 3×3×3 paired 3D convolution layers [41] followed by batch normalization layer and elementwise rectified linear unit (ReLU) layer. The first three blocks

are normal convolution layers designed to extract low-level volume features such as corners and edges, the three blocks in the middle are convoluted layers with dilation factor of 2, and the last 4 blocks consist of convolution layers comprising dilation factor of 4. The inner layers with larger dilated factors along with cross layer up-skip connections to learn multi and high level features abstraction from input volumes. The network is designed to train end-to-end in a cascade way to exploit transferred knowledge learning and corresponding convolutional layers along with rectified element and residual blocks are arranged in a pre-activation manner to maintain symmetry [42]. We have considered single class at output (fore-ground) and employed dice loss to measure the overlap between foreground prediction (i.e., segmented vessels) and foreground ground truth.

B. Dilated Convolution with Up-Skip Connections

Generally, to recover feature spatial details, deconvolution layers are usually integrated in the network design. In order to efficiently produce dense feature maps, instead of using down-sampling layer, we have employed dilated convolution with up-skip connections [43]. The core idea behind employing dilated convolution with up-skip connection is that, it enhances information flow between the encoding portion and decoding portion to ensure more low level features are used for optimizing the segmentation results. The up-skip connection provides a new pathway between the convolution layers and dilated convolution during the forward propagation process, which allows the dilated convolution layers to extract more low-level features and thus helps recover spatial information. This concept up-sample the filters in successive convolutional layers resulting in feature maps of enhanced spatial resolution computed at a higher sampling rate with increased network connectivity [41]. This policy offers a straightforward yet prevailing substitute of deconvolutional layers for detailed segmentation maps along delineated object boundaries to learn multilevel features [41], [44]. Moreover, deconvolution layers also introduce additional computational costs thus dilated convolution is a way to reduce network computational cost efficiently. Inspired by the work of [42], we have incorporated dilated convolution for volumetric segmentation task. We have up-sampled 3D convolution kernels with a dilation factor k. Considering an input feature map of size X with N channels, the output feature map Z is generated for our experimental setup as:

$$Z_{i,j,l} = \sum_{n=0}^{N=1} \sum_{a=0}^{2} \sum_{b=0}^{2} \sum_{c=0}^{2} R_{a,b,c,n} X_{(i+ak),(j+bk),(l+ck),n} \quad (1)$$

where, the variables (i, j, l) go through every spatial position in the input volumes. Hence, we can set receptive field up to a $(2kl + 1)3$ voxel with kernels R. For minimal use of parameters, we have selected the kernel size of 3×3×3 for 3D convolution layer which can cover 3D features relative to a central voxel [42] in all directions.

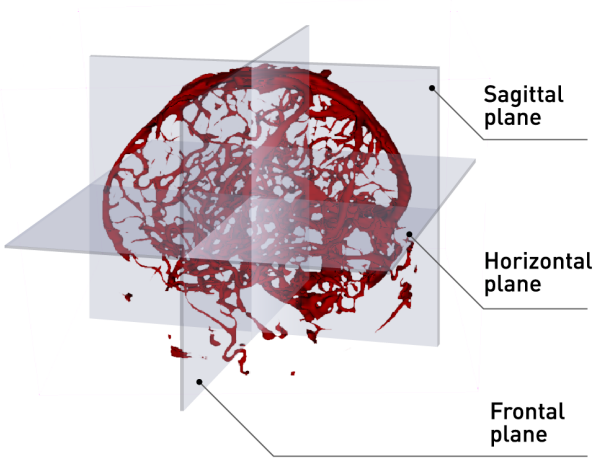


Fig. 3: The segmentation result of our approach. Visualized using volume rendering of the blood vessels, together with the trans-axial multi-planar views.

C. Residual Mapping with Inception Module

The depth of fully 3D CNN plays a fundamental role in achieving finer performance [45]. However, with an increase in network depth, the notorious problem of network degradation enhances due to the gradient diffusion which makes network training process difficult. Hence, CNN generates unsatisfying segmentation results. To address this issue, He et al. [46] proposed and refined residual connections for effective training and optimization. In our setup, instead of direct mapping between input and output, we have added a residual block after two layers with an inception module for deep analysis.

To advance the richer representation capacity of the segmentation network for blood vessels close to skull and to optimize the segmentation performance, we have adopted an inception block within convolution layers, which has been experimentally verified to boost the capturing of advanced visual information under controlled computational complexity. The inception module implemented in our end-to-end FCN is redesigned, where a 3×3 convolution layer to enlarge the receptive field and the max pooling layer is replaced by a short path to directly incorporate input filters. The output filters generated from the 1×1 and 3×3 convolutional layers are concatenated with input feature map to achieve feature fusion.

According to our 3D CNN design, x represents the input and $F(x)$ denotes the original function of the network. By adding a direct bypass from the input layer to the output layer, we reshape the mapping function F to newer version $O : O(x) = F(X) + X$ to make information propagation smooth via network and for speedy training. The network is set to be trained to estimate a residual function based on new modification $O(x)-x$, which is cost effective and does not required new parameters during training. Moreover, residual mapping along with the inception module improves the gradient flow in the backward propagation and helps to avoid the gradient vanishing during training, offering additional guidance for the

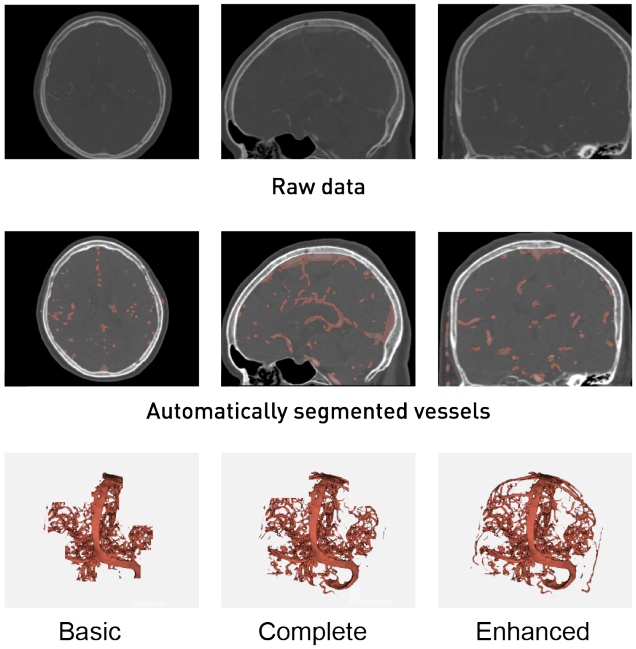


Fig. 4: Our segmentation results with respect to multiple views on three sub-stages of basic, complete, and enhanced. The first row shows an original CTA volume, and the second row shows the segmentation results. The columns represent the axial (left), sagittal (middle), and coronal (right) views. The third row illustrates the corresponding 3D visualizations of the segmentation results.

learning of earlier layers.

D. Optimization by Maximizing Mean Dice Coefficient

CNN can suffer from biased classification towards majority class problem when the training data is severely unbalanced for medical image segmentation task. To mitigate the effects of biases and resultant misclassification, one optimal solution is to maximize the mean dice coefficient directly, a solution proposed by [47]. End-to-end trained fully 3D CNN reduce the entropy of total error function L by utilizing K voxel of volume $\{a_k\}_{k=1}^{K}$ and the training data of S class segmentation $map\{b_k\}_{k=1}^{K}$, $b_k 1, ...S$ as:

$$L(a_k, b_k) = -\frac{1}{K}\sum_{k=1}^{K}\sum_{k=1}^{S}\lambda(b_k = S)\log E_S(a_k) \qquad (2)$$

where λ shows Dirac delta function, $E_S(a_k)$ is the final score of a_k over the s^{th} class. To overcome the problem of imbalance training data we have maximized the mean dice coefficient by employing the following, where C is dice coefficient.

$$C(a_k, b_k) = \frac{1}{S}\sum_{s=1}^{S}\frac{2\sum_{k=1}^{K}\lambda(b_k = S)E_s(a_k)}{\sum[\lambda(b_k = S)^2 + \sum_{k=1}^{k}[E_s(a_k)]^2} \qquad (3)$$

TABLE I: Quantitative analysis of integrating variations of activation function with a different combination of architectural designs using 20 and 22 layers of our proposed network (i.e., ReLU, elastic ReLU, and parametric ReLU). The second section in the table demonstrates the proposed method variants with incremental usage of different architectural modules.

Layers: Metrics	DSC (%)	PPV (%)	AVD (%)	Hausdorff (mm)
20L:ReLUs	89.56	90.47	89.28	5.32±1.10
20L:ELUs	88.32	89.17	87.93	5.71±1.52
20L:PReLUs	88.85	89.45	87.78	5.57±1.34
22L:ReLUs	90.75	91.56	89.86	5.01±1.05
22L:ELUs	89.23	90.23	88.95	5.42±1.22
22L:PReLUs	89.89	90.79	89.25	5.35±1.08

E. Cascaded Training Strategy

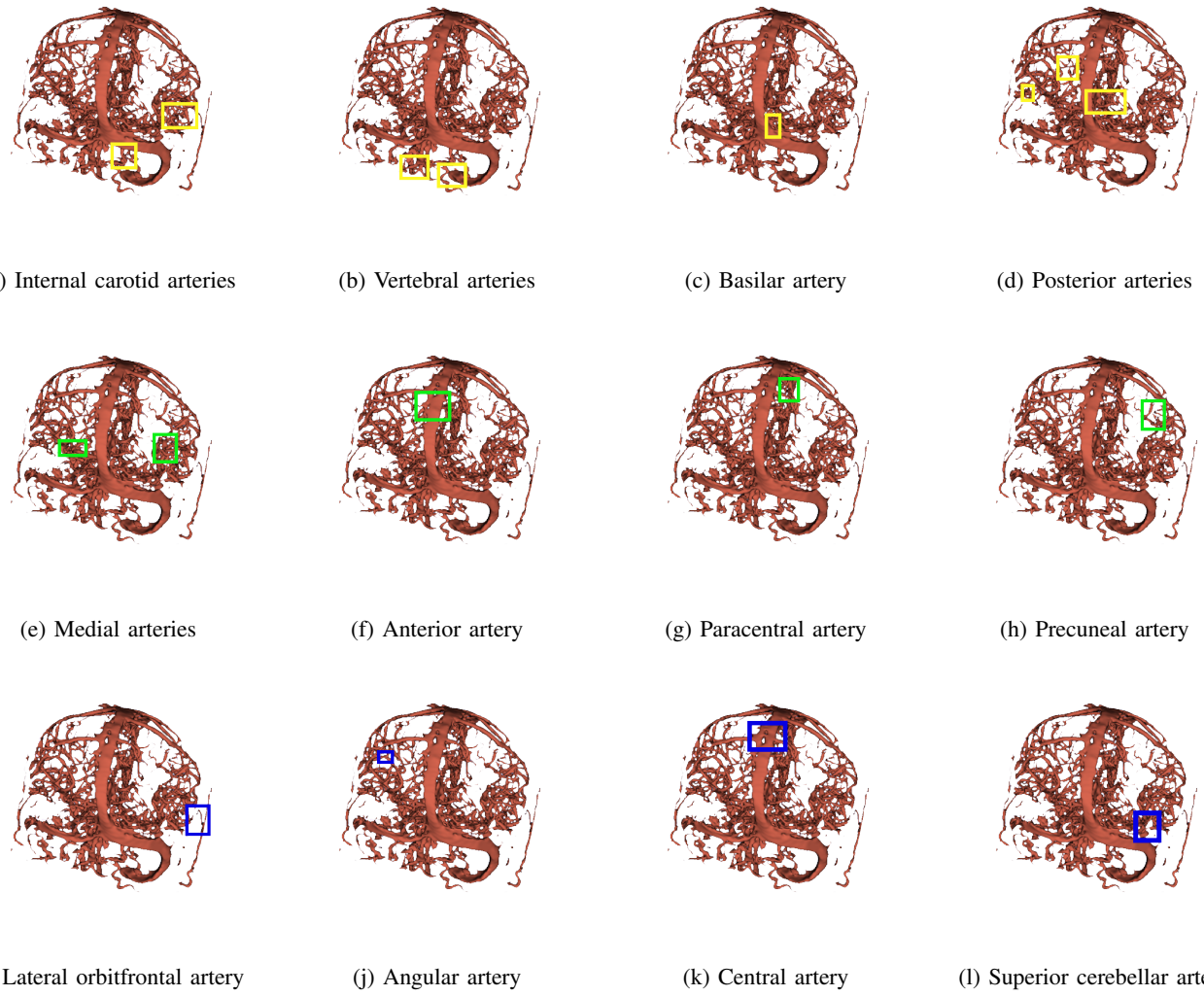
Inspired by the success of knowledge transfer in boosting performance [29], this research employed a three-stage cascaded training approach for concrete segmentation of brain vessels close to skull (basic, complete, and enhanced sub-stages) sequentially. During cascade training of the FCN, the knowledge learned from the basic segmentation sub-stage is transferred to the complete segmentation sub-stage, and the information learned from the complete segmentation sub-stage is shared with the enhanced segmentation sub-stage. Cascaded training approach gradually optimizes the segmentation results for blood vessels near to skull considering the inclusion associations of topologies between basic, complete and enhanced sub-stages. To summarize the whole concept, the well trained basic sub-stage is utilized as pre-trained model to initialize the complete segmentation training sub-stage, and the enhanced sub-stage is fine-tuned via well-trained complete segmentation stage at both image and sub-volume level.

F. Data Pre-Processing

The CTA volumes from our original dataset have size up to 300×256×328, which increases computation cost (mainly the GPU memory). We have incorporated a concept of sliding windows to split the whole volume into slices with a fixed size of 128×128×128. Additionally, we have employed re-sampling method with bilinear interpolation to make the image spacing of all the CTA volumes as ($1mm$, $1mm$, $1mm$) and divide the dataset according to the width of the window. To augment the dataset for avoiding over-fitting segmentation results, we have applied augmented operation such as rotating (10 degrees left, 10 degrees-right), scaling by (1.2 times), reducing (up to 0.8 times) and symmetrical transformation (upper, lower, left, and right). We have utilized method described in [42] to initialize the hyper-parameters for training, such as Adam as training optimizer, the learning rate is 0.00005 coefficient of weight decay regularization is 0.000006, the batch size is 4, the regularization type is L2, and dice similarity coefficient (DSC) as the loss function. It costs around 15 hours to complete the training procedure on a Nvidia Telsa V100 GPU.

III. Experimental Results

This section discusses the dataset preparation, implementation details, experimental setup, qualitative and quantitative



(a) Internal carotid arteries (b) Vertebral arteries (c) Basilar artery (d) Posterior arteries

(e) Medial arteries (f) Anterior artery (g) Paracentral artery (h) Precuneal artery

(i) Lateral orbitfrontal artery (j) Angular artery (k) Central artery (l) Superior cerebellar artery

Fig. 5: Qualitative results of our method for the blood vessels located close to the skull. From (a)-(l) yellow, green, and blue boxes show the labeled vessels. From (a)-(l), 12 major blood vessels are internal carotid vertebral, vertebral, basilar, posterior, medial, anterior, paracentral, precuneal, lateral orbitfrontal, angular, central, and superior cerebellar arteries.

TABLE II: Quantitative results of our proposed approach with respect to different batch sizes and learning rates for three sub-stages of segmentation process using DSC metric.

DSC w.r.t Batch Size					DSC w.r.t Learning Rate				
Batch Size	@8	@16	@32	@64	Learning Rate	@0.00002	@0.00003	@0.00004	@0.00005
Basic	75.86	78.45	80.25	82.15	Basic	77.44	78.56	79.38	80.99
Complete	83.15	85.75	87.66	88.12	Complete	81.75	84.63	86.53	87.67
Enhanced	85.75	81.35	89.11	90.75	Enhanced	87.83	88.11	89.09	90.70

evaluation of results along with a comparison of proposed method with state-of-the-art methods.

A. Dataset Preparation and Implementation Details

This study has been evaluated on clinical dataset of 70 CTA volumes and their corresponding labeled ground truth collected from the cooperative Shanghai hospital, China. The manual labeling of the dataset for validation purposes is done by an expert medical physician, for 20 CTA it cost around one month to complete manual labeling, five to six hours for each CTA. The implementation was conducted on P3xlarge server instance provided by AWS (Amazon Web Service). It is set to a Tesla V100 GPU, 16GB GPU memory, 8 virtual CPUs. The operating system was 64-bit version of the 16.04LTS, the deep learning library was TensorFlow 1.3.0 version with cuda9.0 and cuDNN7.1 of the NVIDIA. For 3D visualization of results, we have used 3D slicer version 4.7.0 [40] which reconstruct blood vessels of the brain in 3D with multiple viewpoints (sagittal, horizontal, and frontal plane) using iso-surface extraction. 3D slicer axis with respect to the segmented brain vessels is shown in Fig. 3.

TABLE III: Ablation study of the proposed method. Here, 22L is used for base net (22 Layers), D is for dilated convolutions, I for inception modules, U for up-skip connections, and C for cascaded training strategy.

Layers: Metrics	DSC (%)	PPV (%)	AVD (%)	Hausdorff (mm)	Training Time (sec)
22L:D	86.75	91.56	89.86	5.01±1.05	35-40
22L:U	89.23	90.23	88.95	5.42±1.22	30-35
22L:D+U	89.89	90.79	89.25	5.35±1.08	25-30
22L:I+U+C	90.75	89.56	87.86	5.01±1.05	15-20

B. Quantitative Evaluation

For quantitative assessment of this research, we have used four types of metrics named as spatial overlap, distance based, accuracy measure, and volumetric metric. The training process of proposed OFF-eNET is based on backpropagation. The dice coefficient quantifies extent of spatial overlapping between two binary images [48]. A dice value varies from 0 to 1 (0 means no overlap and 1 means perfect agreement). Additionally, we have reported three other metrics for showing the quantitative evaluation of this study, i.e., Hausdorff distance (HD), positive predictive value (PPV), and absolute volumetric difference (AVD) [48]. The Hausdorff distance H_{xy}, across two sets of points x and y is the maximum value of the distance for all volume voxels defined as $H = \max\ (h_{xy}, h_{vx})$. The PPV or precision measure is the relation between true positives output divided by all elements classified as positives. For our experimental setup, it is defined as the percentage of how many truly segmented vessels are actually taking part in measuring the accuracy of segmentation result. AVD represents percentage ratio of the absolute difference between input volume and the segmented output volume, to the input volume.

C. Segmentation Results with Respect to Different Combination of Proposed Method Architectural Variants

Table I implies that integrating residual connections along with a different variation of activation function with a various combination of architectural designs (i.e. ReLU, elastic ReLU, and parametric ReLU) improved the segmentation results. We have achieved best results in terms of DSC which are 90.75% for deploying 22 layers + ReLU combination. Moreover, it is inferred from the Table I that PPV and AVD values are increased and Hausdorff distance got decreased with layer setting of 22 Layers + ReLUs comparing with other combination of layers. Additionally, We present an ablation experiment of the gated sum in Table III. we have demonstrated the incremental performance gained from various combination of modules like dilated convolution, inception module, up-skip connections, and cascade training. Where the last row in Table III shows that the novel fusion strategy improves the performance visibly. Comparing the following two combinations i.e. (base net (22 layers) + dilated), (base net (22 layers) + up-skip), (base net (22 layers) + dilated, up-skip), and (base net(22 layers) + inception, up-skip, cascade training), we can see that the best results are achieved by utilizing the advantages of all the above-mentioned extensions in the proposed network. We have carried out another set of experiments for the proposed method to validate the results with various batch sizes and learning rates to investigate the impact of dice loss with respect to three sub-stages of segmentation, see Table II. The comparison results indicate that with a larger batch size (i.e., 64) and higher segmentation sub-stage leads to better segmentation performance. Similarly, with a higher learning rate up to 0.00005 our method achieves the best performance for enhanced segmentation stage.

TABLE IV: Quantitative analysis of segmentation results for three sub-stages by comparing dice loss and cross entropy loss.

Segmentations Stages	Dice Loss	Cross Entropy Loss
Basic	78.37	73.34
Complete	85.43	76.34
Enhanced	90.03	80.45

D. Analysis of Segmentation Results for Three Sub-Stages Using Cross Entropy Loss and Dice Loss

For this study, we have calculated two losses, i.e. dice loss and cross entropy loss (CEL), to compare and reveal the effectiveness of dice loss over CEL for mitigating class imbalance problem. CEL is calculated as the log loss, summed over all possible class values to get final score for all pixels and then averaged. Considering equal learning to each pixel in the input image, CEL evaluates the class predictions for each pixel vector individually which may can cause a problem when output classes have unbalanced representation in the image. Hence dice loss as an optimum choice to overcome the class imbalance problem. Table IV demonstrates the results of proposed method with CEL and dice loss functions for three segmentation sub-stages.

E. Qualitative Evaluation

In this subsection qualitative results of proposed method have been demonstrated by vessel segmentation outputs and their visualizations are provided. Fig. 4 shows the output of three stages of our segmentation process (i.e. basic, complete, and enhanced) with respect to the sagittal, horizontal, and frontal plane in a 3D slicer. The top layer is the representation of raw data with anterior, superior and right 3D viewpoints of the brain. The second layer is automatic segmentation results of respective 3D views and last layer is 3D visualization of labeled and segmented blood vessels with clear boundaries. Fig. 5 demonstrates the detailed analysis of achieved results especially for the blood vessels close to skull, each part in the illustration from (a-l) is showing a labeled vessel marked with colored boxes. We have marked and identified 12 major intracranial blood vessels (i.e., internal carotid vertebral, vertebral, basilar, posterior, medial, anterior, paracentral, precuneal, lateral orbitfrontal, angular, central, and uperior cerebellar arteries) for brain in areas close to skull [51]. The results were verified by medical physician with satisfied performance. Fig. 6 compares the results of automatic blood vessel segmentation with that of manually labeled CTA's by medical experts. It is shown from the illustration that the proposed method delivered comparable segmentation results with the ground truth, where the major blood vessels close to

TABLE V: Comparison of our approach with five state-of-the-art methods in terms of mean DSC, PPV, and AVD.

Method	DSC (%)	PPV (%)	AVD (%)	Hausdorff Distance (mm)
Noori et al. [49]	83.44±0.79	82.23±0.64	81.75±0.98*	6.24±1.78*
Shi et al. [50]	87.99±0.34*	85.32±0.73*	83.45±0.67	6.89±1.57*
Li et al. [42]	84.03±0.75	87.11±0.77*	86.49±0.16*	6.25±1.96*
Chen et al. [35]	86.99±0.41*	86.91±0.25	85.43±0.54*	6.01±1.25
Livne et al. [24]	87.44±0.32	86.99±0.94*	86.11±0.44*	6.51±1.05
Our Method	**90.75±0.30**	**89.56±0.20**	**87.86±0.10**	**5.01±1.05**

$P < 0.05$, * derived for a paired t-test shows statistically significance difference.

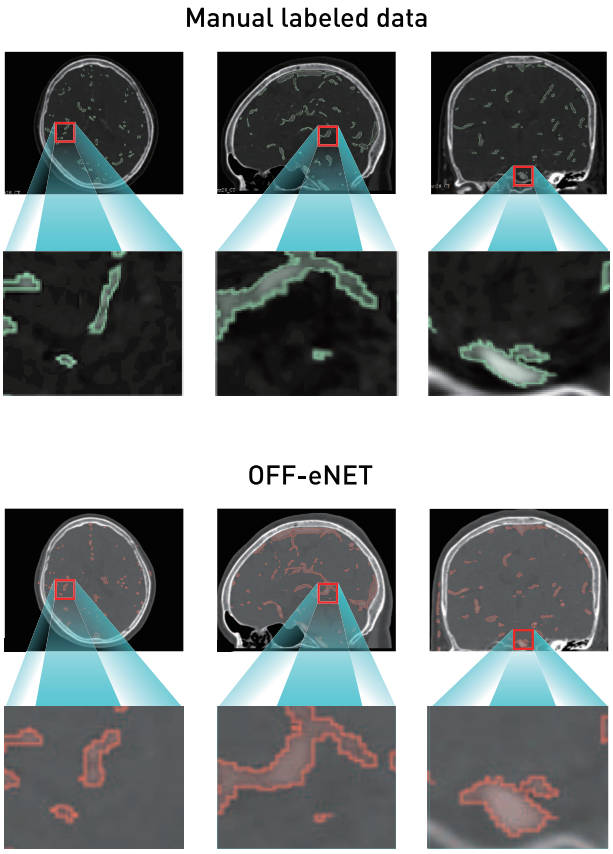


Fig. 6: This figure compares the results of our proposed OFF-eNET with manually labeled CTA volumes by medical experts. The second and fourth rows are amplified areas of the first and third rows respectively for the marked critical arteries close to skull, i.e. anterior, posterior, and medial brain arteries.

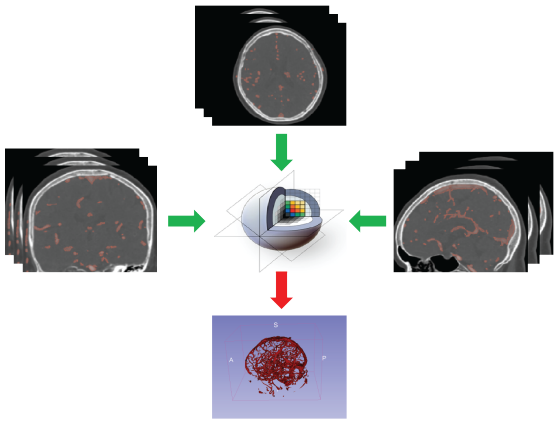


Fig. 7: Qualitative results of another patient.

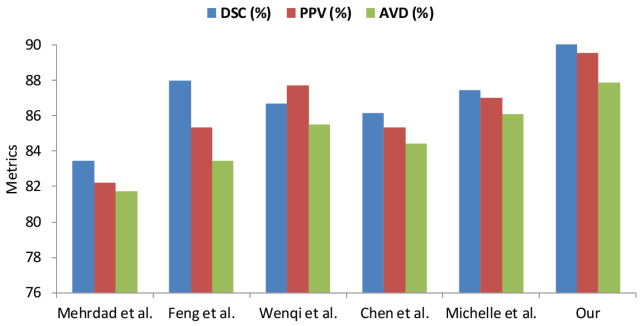


Fig. 8: Comparison of proposed method with five state-of-the-art Noori et al. [49], Shi et al. [50], Li et al. [42], Chen et al. [35], and Livne et al. [24] methods.

skull can be clearly seen with an enlarged view on automatic segmented results. More qualitative results are represented in Fig. 7.

F. Comparison with State-of-the-Art

We have carried out comparisons with five recent state-of-the-art volumetric segmentation techniques including Noori et al. [49], Shi et al. [50], Li et al. [42], Chen et al. [35], and Livne et al. [24]. We have utilized mean DSC, AVD, PPV, and Hausdorff distance (mm) as the performance metrics for the comparisons. For testing the significance of the differences between the results of comparison methods, we computed the p-value using the paired t-test (two-sample t-test) with significance level of p at 0.05. Table V and Fig. 8 demonstrate the comparison results. With reasonably fewer parameters than Noori et al. [49] and Shi et al. [50], the proposed method outperforms the competing method in terms of mean DSC of 90.75%, PPV up to 89.56%, AVD value increases up to 87.86% and Hausdorff distance (mm) decreased to 5.01±1.05mm. These results show that we have achieved significantly higher results, i.e., approximately 3-4% accuracy compared to Livne et al. [24] and Chen et al. [35] in terms of DSC and PPV.

The obtained dice coefficient with error value of approximately 10% shows that the OFF-eNET is an accurate solution to optimally detect thin vessels, and outperforming the existing models such as Livne et al. [24], and Li et al. [42]. Qualitatively, we can see that the OFF-eNET resulted in slightly greater amount of depiction of the thick blood vessels when compared to the manual labelling counterpart. We assessed our model's

performance using four different measures: First, the dice coefficient which is a well-known for quantitative measure of segmentation accuracy [24]; Second and third are the distance-based measures, the HD and the AVG, respectively; Fourth one is the accuracy measure, i.e., PPV. These differences suggest that the use of hybrid combination of dilation factor wit up skip connection and inception module with cascade training strategy can actually improve the performance of segmentation networks especially for thin intracranial blood vessels.

Currently, our technique takes approximately 120 seconds to complete the segmentation process on an image volume of size 128×128×128 voxel on Nvidia GTX 1060. The time is reduced to 15-20 seconds when using Tesla V100. We have used 70 CTA volumes out of which we have used 50 for training and 20 for testing. We divided the image volume in to 128×128×128 size sub-volumes, thereby dividing the testing set of 20 CTA volumes approximately 5 million sub-volumes. We have conducted 5-fold cross validation to estimate how accurately our automatic segmentation results will perform in clinical practice. We divided the 70 volumes dataset into five equal folds, with each of the first four folds used exactly once as the validation data. The 5^{th} fold was solely used for validation to test the dataset. Table VI shows the results for five folds with averaged results in terms of four metrics.

In summary, this study provides a starting point to develop a brain vessels segmentation solution which can be applicable in the neuropathology settings. For qualitative evaluation, two experienced medical experts visually assessed the superiority of our segmentation. We found an excellent performance for thick, as well as thin blood vessels close to the skull area which showed higher performance of OFF-eNET as compared to state-of-the-art segmentation methods.

TABLE VI: Cross validation.

Validation	DSC (%)	PPV (%)	AVD (%)	Hausdorff (mm)
Fold_1	90.45	86.34	86.11	6.0
Fold_2	89.95	88.46	87.49	5.4
Fold_3	88.36	87.56	88.56	5.0
Fold_4	87.49	86.23	89.67	4.5
Fold_5	90.75	89.11	86.89	4.8
Fold_{avg}	89.46	87.8	87.2	5.04

IV. Conclusion

We presented an efficient OFF-eNET architecture for automatic intracranial vascular segmentation from CTA volumes. We focused on enhancing our work compared to the existing methods by segmenting thin blood vessels near the skull for improving the accuracy of dense volumetric segmentation. We have applied various preprocessing techniques to resample 3D volumes for effective training. To handle network convergence, this research mainly employed residual mapping along with inception module. To deliver richer representation for preserving spatial feature map, we have employed dilated convolution along with innovative use of up-skip connections. The proposed method is evaluated by verifying the automatic segmentation result against manually labeled gold standard CTA volumes. This type of segmentation has remained ignored in previous state-of-the-art due to requirement of preserving concrete details. Our proposed architecture proved to be a promising contribution having clinical importance both during diagnosis, where the volume of brain CTA needs to be analyzed, and during treatment planning when the estimation of the anatomical structure of blood vessels needs to be accurate.

References

[1] H. Ali, L. Rada, and N. Badshah, "Image segmentation for intensity inhomogeneity in presence of high noise," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3729–3738, 2018.

[2] Z. Tang, P.-T. Yap, and D. Shen, "A new multi-atlas registration framework for multimodal pathological images using conventional monomodal normal atlases," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2293–2304, 2019.

[3] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen, "High-resolution encoder-decoder networks for low-contrast medical image segmentation," *IEEE Transactions on Image Processing*, pp. 1–14, 2019.

[4] H. Li, A. Li, and M. Wang, "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks," *Computers in Biology and Medicine*, vol. 108, pp. 150–160, 2019.

[5] S. Mishra, P. Liang, A. Czajka, D. Z. Chen, and X. S. Hu, "CC-NET: Image complexity guided network compression for biomedical image segmentation," in *IEEE International Symposium on Biomedical Imaging*, 2019, pp. 57–60.

[6] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Medical Image Analysis*, vol. 13, no. 6, pp. 819–845, 2009.

[7] F. Zhao, Y. Chen, Y. Hou, and X. He, "Segmentation of blood vessels using rule-based and machine-learning-based methods: A review," *Multimedia Systems*, vol. 25, no. 2, pp. 109–118, 2019.

[8] Z. Han, Z. Liu, C.-M. Vong, Y.-S. Liu, S. Bu, J. Han, and C. L. P. Chen, "Deep spatiality: Unsupervised learning of spatially-enhanced global and local 3D features by deep neural network with coupled softmax," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3049–3063, 2018.

[9] X. Chen and L. Pan, "A survey of graph cuts/graph search based medical image segmentation," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 112–124, 2018.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[12] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, and C. P. Langlotz, "Deep learning in neuroradiology," *American Journal of Neuroradiology*, vol. 39, no. 10, pp. 1776–1784, 2018.

[13] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. Todd Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "On the necessity of fine-tuned convolutional neural networks for medical imaging," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing: Precision Medicine, High Performance and Large-Scale Datasets*, L. Lu, Y. Zheng, G. Carneiro, and L. Yang, Eds., 2017, pp. 181–193.

[14] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," in *MICCAI BraTS*, 2014, pp. 36–39.

[15] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[16] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI," in *International MICCAI Brainlesion Workshop*, 2016, pp. 131–143.

[17] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *MICCAI*, 2014, pp. 520–527.

[18] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, "An ensemble of 2D convolutional neural networks for tumor segmentation," in *Scandinavian Conference on Image Analysis*, 2015, pp. 201–211.

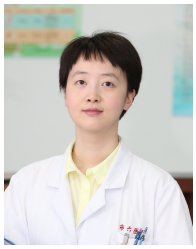
[19] R. Phellan, A. Peixinho, A. Falcão, and N. D. Forkert, "Vascular segmentation in TOF MRA images of the brain using a deep convolutional neural network," in *Intravascular Imaging and Computer Assisted*

Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, 2017, pp. 39–46.
[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.
[21] A. Fabijaska, "Segmentation of corneal endothelium images using a U-Net-based convolutional neural network," *Artificial Intelligence in Medicine*, vol. 88, pp. 1–13, 2018.
[22] Q. Huang, J. Sun, H. Ding, X. Wang, and G. Wang, "Robust liver vessel extraction using 3D U-Net with variant dice loss function," *Computers in Biology and Medicine*, vol. 101, pp. 153–162, 2018.
[23] B. Norman, V. Pedoia, and S. Majumdar, "Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry," *Radiology*, vol. 288, no. 1, pp. 177–185, 2018.
[24] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey, and V. I. Madai, "A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," *Frontiers in Neuroscience*, vol. 13, pp. 97:1–97:13, 2019.
[25] D. E. Cahall, G. Rasool, N. C. Bouaynaya, and H. M. Fathallah-Shaykh, "Inception modules enhance brain tumor segmentation," *Frontiers in Computational Neuroscience*, vol. 13, pp. 44:1–44:8, 2019.
[26] S. Vesal, N. Ravikumar, and A. Maier, "Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2019, pp. 319–328.
[27] 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.
[28] J. Merkow, A. Marsden, D. Kriegman, and Z. Tu, "Dense volume-to-volume vascular boundary detection," in *MICCAI*, 2016, pp. 371–379.
[29] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-Cascade: Cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1992–2004, 2017.
[30] F. Zhao, Y. Chen, F. Chen, X. He, X. Cao, Y. Hou, H. Yi, X. He, and J. Liang, "Semi-supervised cerebrovascular segmentation by hierarchical convolutional neural network," *IEEE Access*, vol. 6, pp. 67 841–67 852, 2018.
[31] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "DeepVesselNet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes," *CoRR*, vol. abs/1803.09340, pp. 1–13, 2018.
[32] H. Kandil, A. Soliman, F. Taher, A. Mahmoud, A. Elmaghraby, and A. El-Baz, "Using 3-D CNNs and local blood flow information to segment cerebral vasculature," in *IEEE International Symposium on Signal Processing and Information Technology*, 2018, pp. 701–705.
[33] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, 2018.
[34] J. Dolz, C. Desrosiers, L. Wang, J. Yuan, D. Shen, and I. Ben Ayed, "Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation," *Computerized Medical Imaging and Graphics*, vol. 79, pp. 101 660:1–101 660:12, 2020.
[35] 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*, vol. 170, pp. 446–455, 2018.
[36] H. Ding, X. Jiang, B. Shuai, A. Q. Liu, and G. Wang, "Context contrasted feature and gated multi-scale aggregation for scene segmentation," in *IEEE CVPR*, 2018, pp. 2393–2402.
[37] H. Ding, X. Jiang, B. Shuai, A. Q. Liu, and G. Wang, "Semantic correlation promoted shape-variant context for segmentation," in *IEEE CVPR*, 2019, pp. 8877–8886.
[38] H. Ding, X. Jiang, A. Q. Liu, N. M. Thalmann, and G. Wang, "Boundary-aware feature propagation for scene segmentation," in *IEEE ICCV*, 2019, pp. 6818–6828.
[39] 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*, vol. 40, no. 4, pp. 834–848, 2018.
[40] S. S. F. Yip, C. Parmar, D. Blezek, R. S. J. Estepar, S. Pieper, J. Kim, and H. J. W. L. Aerts, "Application of the 3D slicer chest imaging platform segmentation algorithm for large lung nodule delineation," *PLOS ONE*, vol. 12, no. 6, pp. e0 178 944:1–e0 178 944:17, 2017.
[41] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *MICCAI*, 2016, pp. 424–432.
[42] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task," in *International Conference on Information Processing in Medical Imaging*, 2017, pp. 348–360.
[43] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, 2014, pp. 818–833.
[44] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
[45] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*, 2016, pp. 630–645.
[47] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *International Conference on 3D Vision*, 2016, pp. 565–571.
[48] K. O. Babalola, B. Patenaude, P. Aljabar, J. Schnabel, D. Kennedy, W. Crum, S. Smith, T. F. Cootes, M. Jenkinson, and D. Rueckert, "Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI," in *MICCAI*, 2008, pp. 409–416.
[49] M. Noori, A. Bahri, and K. Mohammadi, "Attention-guided version of 2D UNet for automatic brain tumor segmentation," in *International Conference on Computer and Knowledge Engineering*, 2019, pp. 269–275.
[50] F. Shi, Q. Yang, X. Guo, T. A. Qureshi, Z. Tian, H. Miao, D. Dey, D. Li, and Z. Fan, "Intracranial vessel wall segmentation using convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2840–2847, 2019.
[51] L. D. J. Fiederer, J. Vorwerk, F. Lucka, M. Dannhauer, S. Yang, M. Dümpelmann, A. Schulze-Bonhage, A. Aertsen, O. Speck, C. H. Wolters, and T. Ball, "The role of blood vessels in high-resolution volume conductor head modeling of EEG," *NeuroImage*, vol. 128, pp. 193–208, 2016.

Anam Nazir received the B.Sc. (Honors) degree in telecommunication and networking from the COMSATS University Islamabad, Wah Cantt, Pakistan, and the M.Sc. degree in computer science from the COMSATS University Islamabad, Islamabad, Pakistan. She is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. She is also a Lecturer with the Department of Computer Science, COMSATS University Islamabad, Wah Cantt, Pakistan. Her current research interests include medical image processing, computer vision, medical image analysis, and deep learning techniques.

Muhammad Nadeem Cheema received the B.Sc. (Honors) degree in computer science from the Capital University of Science & Technology, Islamabad, Pakistan, and the M.Sc. degree in computer science from the COMSATS University Islamabad, Attock, Pakistan. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. He is also a Lecturer with the Department of Computer Science, COMSATS University Islamabad, Attock, Pakistan. His current research interests include medical image processing, computer vision, medical image analysis, and deep learning techniques.

Bin Sheng (Member, IEEE) received the B.A. degree in English and the B.Eng. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, and the M.Sc. degree in software engineering from the University of Macau, Macau, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong. He is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include virtual reality and graphics. He is an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology.



Huating Li received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, and Pennington Biomedical Research Center, Baton Rouge, LA, USA. She is currently an Associate Professor with the Shanghai Jiao Tong University Affiliated Sixth People's Hospital and the Shanghai Diabetes Institute. Her current research interests include the role of cytokines in the development of fatty liver disease, diabetes, and other obesity-related diseases. Dr. Li was a recipient of the Shanghai Science and Technology Progress Award.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong. He is currently a Research Assistant Professor with The Hong Kong Polytechnic University, Hong Kong. His current research interests include image/video stylization, artistic rendering and synthesis, and creative media. He has one image/video processing national invention patent, and has excellent research project reported worldwide by *ACM TechNews*.

Po Yang (Senior Member, IEEE) received the B.Sc. degree in computer science from Wuhan University, Wuhan, China, the M.Sc. degree in computer science from the University of Bristol, Bristol, U.K., and the Ph.D. degree in electronic engineering from the Staffordshire University, Stoke-on-Trent, U.K. He is currently a Senior Lecturer in Large Scale Data Fusion with the Department of Computer Science, The University of Sheffield, Sheffield, U.K. He also holds a strong tracking of high-quality publications and research experiences. He has published over 40 papers. His current research interests include IoT, RFID and indoor localization, pervasive health, image processing, GPU, and parallel computing.

Younhyun Jung received the B.Sc. degree in computer science from Inha University, Incheon, Korea, in 2008, and the Ph.D. degree in computer science from The University of Sydney, Sydney, Australia, in 2016. He is currently a Postdoctoral Research Fellow in computer science with The University of Sydney, Sydney, Australia. He worked at Samsung Electronics as a software engineer from 2007 to 2010. His current research interests include volume rendering and multi-modal medical image visualization.

Jing Qin (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, China, in 2009. He is currently an Assistant Professor with the School of Nursing, The Hong Kong Polytechnic University, Hong Kong, where he is also a Key Member with the Centre for Smart Health. His current research interests include virtual/augmented reality for healthcare, medical image processing, deep learning, and health informatics.

Jinman Kim (Member, IEEE) received the B.S. (Hons.) and Ph.D. degrees in computer science both from The University of Sydney, Sydney, Australia. Since 2006, he has been a Research Associate with the leading teaching hospital, the Royal Prince Alfred. From 2008 to 2012, he was an ARC Post-Doctoral Research Fellow, one year leave from 2009 to 2010 to join the MIRALab Research Group, Geneva, Switzerland, as a Marie Curie Senior Research Fellow. Since 2013, he has been with the School of Information Technologies, The University of Sydney, where he was a Senior Lecturer, and became an Associate Professor in 2016. His current research interests include medical image analysis and visualization, computer aided diagnosis, and telehealth technologies.

David Dagan Feng (Life Fellow, IEEE) received the M.Eng. degree in electrical engineering and computer science (EECS) from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.Sc. degree in biocybernetics and the Ph.D. degree in computer science from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 1985 and 1988, respectively. He received the Crump Prize for Excellence in medical engineering from UCLA. He is currently the Head of the School of Information Technologies, the Director of the Biomedical and Multimedia Information Technology Research Group, and the Research Director of the Institute of Biomedical Engineering and Technology, The University of Sydney, Sydney, NSW, Australia. He has published more than 700 scholarly research articles, pioneered several new research directions, and made a number of landmark contributions in his field. More importantly, however, is that many of his research results have been translated into solutions to real-life problems and have made tremendous improvements to the quality of life for those concerned. He has served as the Chair of the International Federation of Automatic Control (IFAC) Technical Committee on Biological and Medical Systems, has organized/chaired more than 100 major international conferences/symposia/workshops, and has been invited to give more than 100 keynote presentations in 23 countries and regions. He is a fellow of the Australian Academy of Technological Sciences and Engineering.