

## Article

# Deep Learning Based Real-Time Semantic Segmentation of Cerebral Vessels and Cranial Nerves in Microvascular Decompression Scenes

Ruifeng Bai <sup>1,2</sup>, Xinrui Liu <sup>2,3</sup>, Shan Jiang <sup>1</sup> and Haijiang Sun <sup>1,\*</sup>

<sup>1</sup> Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; bairuifeng18@mails.ucas.ac.cn (R.B.); jiangs@ciomp.ac.cn (S.J.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China; liuxinr@jlu.edu.cn

<sup>3</sup> Department of Neurosurgery, The First Hospital of Jilin University, Changchun 130021, China

\* Correspondence: sunhj@ciomp.ac.cn; Tel.: +86-135-7868-7727

**Abstract:** Automatic extraction of cerebral vessels and cranial nerves has important clinical value in the treatment of trigeminal neuralgia (TGN) and hemifacial spasm (HFS). However, because of the great similarity between different cerebral vessels and between different cranial nerves, it is challenging to segment cerebral vessels and cranial nerves in real time on the basis of true-color microvascular decompression (MVD) images. In this paper, we propose a lightweight, fast semantic segmentation Microvascular Decompression Network (MVDNet) for MVD scenarios which achieves a good trade-off between segmentation accuracy and speed. Specifically, we designed a Light Asymmetric Bottleneck (LAB) module in the encoder to encode context features. A Feature Fusion Module (FFM) was introduced into the decoder to effectively combine high-level semantic features and underlying spatial details. The proposed network has no pretrained model, fewer parameters, and a fast inference speed. Specifically, MVDNet achieved 76.59% mIoU on the MVD test set, has 0.72 M parameters, and has a 137 FPS speed using a single GTX 2080Ti card.

**Keywords:** microvascular decompression; real-time semantic segmentation; encoder–decoder



**Citation:** Bai, R.; Liu, X.; Jiang, S.; Sun, H. Deep Learning Based Real-Time Semantic Segmentation of Cerebral Vessels and Cranial Nerves in Microvascular Decompression Scenes. *Cells* **2022**, *11*, 1830. <https://doi.org/10.3390/cells11111830>

Academic Editors: An Pan, Baoli Yao, Chao Zuo, Fei Liu, Jiamiao Yang, Liangcai Cao and Alexander E. Kalyuzhny

Received: 7 April 2022

Accepted: 25 May 2022

Published: 2 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

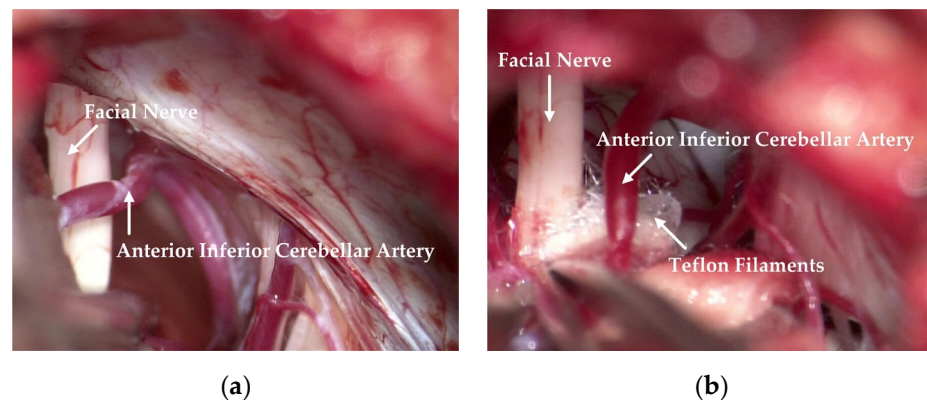
## 1. Introduction

Trigeminal neuralgia (TGN) and hemifacial spasm (HFS) are the most common brain diseases. TGN is mainly manifested as short-term pain similar to an electric shock in the trigeminal nerve distribution area. A slight touch can induce TGN, seriously affecting the patient's life quality [1]. HFS is characterized by unilateral facial muscle painless and repetitive involuntary convulsions [2], there are increasingly younger trend [3]. Microvascular decompression (MVD) is the most commonly used surgical method for the treatment of TGN [4,5] and HFS [6,7] because of its minor operative injury and obvious therapeutic effect. In 1959, Gardner and Miklo [8] first reported the successful treatment of TGN by separating the arteries that compress the trigeminal nerve during surgery. In the 1960s, Jannetta [9] reported for the first time that MVD was performed by microscope and proposed the concept of MVD. MVD is used to remove cerebral vessels from the compressed nerve by a microscopic operation and relieves the pressure of blood vessels on the nerve to achieve therapeutic purposes. When the conservative treatment effect is poor, MVD is the preferred surgical method as long as conditions permit [10]. With the progress of surgery and anesthesia technology, MVD has no clear age limit. As long as the general situation is fine, elderly patients can tolerate general anesthesia and undergo MVD treatment [11].

With the development of endoscopic technology, there is no significant difference in the efficacy of endoscopic MVD for TGN compared with MVD under a traditional microscope, but it can shorten the operation time and effectively reduce the incidence of postoperative complications [12]. Broggi et al. [13] reported that in the case of unclear neurovascular

presentation under the microscope, the problem was solved by endoscopy. Endoscopy has a good light source, its wide-angle imaging can basically cover dead angles of vision, and the local amplification of endoscopy can clearly identify the responsible vessels. Endoscopy can extend the lens into the microvascular decompression area, which is not limited by the surrounding anatomical structure, and can provide a broader and clearer surgical vision image [14]. At the same time, the lens can choose multiple different angles ( $0^\circ$ ,  $30^\circ$ ,  $45^\circ$ ) [15]. In the operation, the lens selects the appropriate angle and the surgeons can fully explore the responsible vessels in the case of zero damage [16]. Under the condition of visualization, the location of nerve and vessel interactions can be determined and it can be ensured that the responsible vessels are sufficiently identified without stretching the brain and nerves. Then, the microscopic operation is performed to effectively reduce the incidence of tissue damage and complications. However, after successful insertion of Teflon filaments, endoscopy can be used to further observe the position between the filaments, nerves, and responsible vessels in multiple directions and without dead angles and effectively identify whether the decompression is sufficient [12] so as to accurately evaluate the surgical effect. Based on the above advantages of endoscopy, endoscopy can be used throughout MVD [17].

The responsible vessels are mostly the superior cerebellar artery, the anterior inferior cerebellar artery and its branches, the basilar artery, etc. Simple venous compression, arachnoid adhesion, and hypertrophy are also important factors in the pathogenesis. Attention should be paid to protecting the petrosal vein and its branches during the operation to prevent injury or tear-bleeding caused by excessive traction. The corresponding decompression method can be selected according to the thickness, elasticity, and length of the blood vessels. As shown in Figure 1, after clarifying the distribution and compression of the responsible vessels, they can then be explored with a microscope. The filler consists of Teflon filaments, which isolate the facial nerve from the responsible blood vessels so that the compression of the facial nerve is relieved. Subsequently, the filaments are observed by means of a neuro endoscope and sutured layer by layer after it has been checked that there are no actively bleeding or missing blood vessels.



**Figure 1.** Relationship observed between the facial nerve and the responsible vessel during an operation. (a) The anterior inferior cerebellar artery crosses the facial nerve, resulting in compression. (b) Teflon filaments are inserted between the facial nerve and the anterior inferior cerebellar artery to relieve compression.

Computer tomography angiography (CTA) is insufficient and there are risks of radiation, contrast-induced nephropathy, and life-threatening allergic reactions. The potential benefits of emergency CTA must be carefully weighed [18]. It should be noted that the soft-tissue resolution of CTA images is not high, which affects the diagnostic accuracy of some cerebrovascular diseases. Digital subtraction angiography (DSA) is the gold standard for the diagnosis of vascular abnormalities. However, DSA is expensive and traumatic [19] and there are various defects, such as radiation exposure, complex operation, and its

time-consuming nature, which render it unsuitable for primary screening and census purposes [20]. Magnetic resonance angiography (MRA) is a non-invasive vascular imaging technique which does not require intubation or a contrast medium. With the continuous progress and development of inspection equipment and post-processing methods, the accuracy and sensitivity of detection are also increasing [19]. During and after the operation, MRA has the advantages of good soft-tissue resolution, multi-parameter imaging, multi-planar scanning, and multi-mode reconstruction. MRA can clearly show the anatomical relation between cranial nerves and responsible vessels, which is an important method for screening etiology. MRA is of great significance in the etiological diagnosis of TGN. It can accurately display the relationship between TGN and peripheral blood vessels in patients with primary trigeminal neuralgia and provide detailed anatomical details for MVD surgery. Compared with traditional CTA, DSA, MRA, and other methods used to obtain cerebral vessel images, it is simpler and more convenient to obtain true-color medical images through endoscopy, and the processing of true-color MVD images has greater development prospects.

In this paper, we propose a new real-time semantic segmentation network based on deep learning, MVDNet, which could directly identify the location of cerebral vessels and cranial nerves in MVD images quickly and accurately. MVDNet can rapidly identify cerebral vessels and cranial nerves during operations, which saves a lot of time and reduces the workloads of surgeons. Because the intra-class of cerebral vessels is too similar and the boundary between cerebral vessels and other brain tissues is not obvious, these make cerebral vessels segmentation difficult. So is the cranial nerves. These make the segmentation of cerebral vessels and nerves difficult. Therefore, we propose a more efficient encoder and a more sophisticated decoder. To train and evaluate our proposed MVDNet network, we collaborated with the First Hospital of Jilin University. We built a dataset containing 3,087 MVD images with well-annotated ground truth. MVDNet can provide real-time inferences for the segmentation and location of cerebral vessels and cranial nerves. For less experienced doctors, it helps to simplify the complexity of intraoperative operations and reduce the number of intraoperative errors. This real-time and accurate segmentation further helps doctors to obtain the best position of cerebral vessels and cranial nerves, which can assist doctors to quickly diagnose and reach the level of professional doctors even beyond professional doctors. The proposed method is compared with several other advanced methods. Experimental results show that our method has achieved the best performance.

In summary, the main contributions of this paper are as follows:

- We created a dataset consisting of 3,087 true-color MVD images and this dataset was used for the segmentation and localization of cerebral vessels and cranial nerves during MVD.
- We propose a new end-to-end network, MVDNet—a lightweight U-shaped model composed of an encoder and a decoder.
- We designed a Light Asymmetric Bottleneck (LAB) module and Feature Fusion Module (FFM) to further improve accuracy at an acceptable cost.
- Extensive experiments were conducted by comparing several of the latest methods with our proposed MVDNet and the experimental results showed that MVDNet was superior to other networks.

## 2. Related Work

**Traditional Methods:** Due to the complexity of the vascular structure, the early vessel segmentation algorithm is mainly based on intensity pattern recognition. For example, the threshold segmentation method mainly depends on the threshold value and the image is divided into different regions by a specific threshold value. Wang et al. [21] extracted foreground and background areas from MRA images through the Otsu threshold method, and divided the foreground area. This method is simple in structure and fast. It may be difficult to obtain robust and accurate segmentation results due to factors such as image

noise and uneven contrast of the image. Region-growing segmentation is another commonly used vessel segmentation method. Starting from the seed points in the vessels, this method gradually adds the pixels satisfying the growth conditions in the field to a set and then realizes the vessel segmentation [22]. This method is highly dependent on the correctness of seed points and is difficult to use in automatic diagnosis systems. Some studies have used different filters for vessel enhancement, including Hessian enhancement [23], Frangi enhancement [24], Satori [25], Erdt [26], Jerman [27], matched filter [28], Gabor wavelet [29], etc. These filter methods are sensitive to noise. In addition, methods based on active contour models are also commonly used vessel segmentation methods. Cheng et al. [30] used the Snake model to complete the fundus vessel segmentation task, which improved the robustness of processing lesion data to a certain extent. Lee et al. [31] proposed a parameter active contour segmentation method based on the Kalman filter, which automatically initialized the contour curve and greatly reduced computational cost. Wang et al. [32] proposed a level set method based on adaptive thresholding by combining global and local threshold information to extract cerebral vessels from MRA data, which enhanced the extraction of small vessels.

**Deep Learning-Based Methods:** Most start-of-the-art semantic segmentation networks are based on fully convolutional neural networks (FCNs) [33]. This type of network builds a fully convolutional neural network model using convolutional layers instead of final fully connected layers, without limiting the size of the input image. The output is fused with shallow feature maps by skip connection to compensate for the detail information of the feature maps. Commonly used cerebrovascular segmentation networks include U-Net [34], ResNet [35], FCN [33], DenseNet [36], etc. Livne et al. [37] segmented the cerebral artery in TOF-MRA images by half of the number of channels in each layer of U-Net. However, this method of reducing the number of channels has limitations in detecting small blood vessels around the skull. Hilbert et al. [38] identified small vessels in TOF-MRA images by integrating 3D U-Net, multi-scale, and depth-supervised methods. This method guides the network middle layer to better generate discriminative features, avoids the problem of gradient explosion and gradient disappearance, and improves the convergence of the model. Wang et al. [39] proposed the MCANet network for the segmentation of the fetal middle cerebral artery (MCA) in ultrasound images. The network removes artifacts caused by dilated convolution through residual connection and improves the accuracy of segmentation. Wang et al. [40,41] extracted and visualized 3D cerebrovascular structures from highly sparse and noisy MRA images based on deep learning. The learned 2D multi-view slice feature vector is projected into 3D space to extract small blood vessels and improve vascular connectivity. Zhang et al. [42] introduced the reverse edge attention network to find missing cerebrovascular edge features and details, and, furthermore, improved the segmentation effect of small blood vessels. Nazir et al. [43] proposed an efficient fusion network for automatic segmentation of cerebral vessels from CTA images and used residual mapping to solve the problems of network convergence.

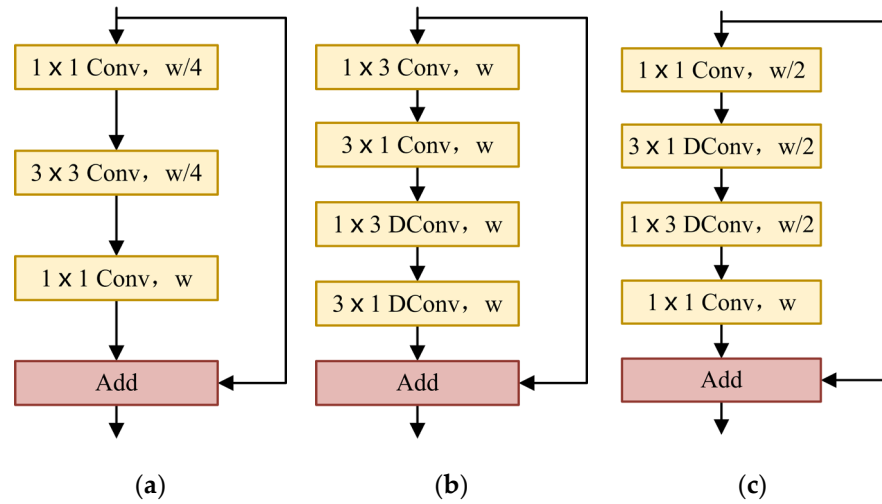
### 3. Method

In this section, we first introduce the light asymmetric bottleneck (LAB) module and the feature fusion module (FFM) in detail. In addition, we describe the effectiveness of these two modules. Finally, we describe the complete network architecture.

#### 3.1. Light Asymmetric Bottleneck Module

In semantic segmentation tasks, some methods are implemented by maintaining the resolution of the input images to ensure sufficient spatial information [44–47]. The receptive field is also an important factor affecting segmentation. The existing methods include pyramid pooling or atrous spatial pyramid pooling to obtain large receptive fields [44,45,47,48]. Especially for real-time semantic segmentation, in order to improve the speed of the network while ensuring segmentation accuracy, it is necessary to use small input images or lightweight basic models. In this paper, we designed the light asymmetric

bottleneck module (see Figure 2c) by observing the bottleneck design in ResNet [35] (see Figure 2a) and the factorized convolutions in ERFNet [49] (see Figure 2b). The LAB module has the advantages of both the bottleneck design and the factorized convolutions.



**Figure 2.** (a) ResNet bottleneck design. (b) ERFNet non-bottleneck-1D module. (c) Our LAB module. w: the number of input channels; DConv: depthwise dilated convolution.

In the LAB module, we also used the bottleneck module. The LAB module is composed of  $1 \times 1$  convolution,  $3 \times 1$  depthwise convolution,  $1 \times 3$  depthwise convolution, and the final  $1 \times 1$  pointwise convolution. The residual connection is used where the number of input channels is the same as that of the output channels. The number of channels is reduced to half of the original when passing the first  $1 \times 1$  convolution and the number of channels remains unchanged through  $3 \times 1$  and  $1 \times 3$  depthwise convolution. Finally, the original channel is restored by  $1 \times 1$  pointwise convolution.

We used a  $1 \times 1$  convolution at the beginning of each LAB module. For the  $1 \times 3$  convolution used at the beginning of the ERFNet [49] non-bottleneck-1D module, the parameters of  $1 \times 1$  convolution are far less than  $1 \times 3$  convolution, which reduces the runtime and memory requirements of the network model and improves the inference speed. After the first convolution, we reduced the number of channels by half. Compared with the thousands of channels in ResNet [35], the maximum number of channels in our model is only 128, which effectively saves a lot of spatial information.

In order to further reduce the number of parameters, we referred to the non-bottleneck-1D module of ERFNet [49]. Convolutional decomposition is applied to depthwise convolution to obtain a more lightweight structure. Alvarez et al. [50] proposed that the standard convolution layer can be decomposed by 1D filters. Let  $W \in \mathbb{R}^{C \times d^h \times d^v \times F}$  denote the weights of the typical 2D convolution layer,  $C$  be the number of input feature maps,  $F$  the number of output feature maps, and  $d^h \times d^v$  the kernel size of each feature map (usually  $d^h = d^v \equiv d$ ). Let  $b \in \mathbb{R}^F$  be the vector representing the bias term for each filter. The  $i$ -th output of a decomposition layer  $a_i^1$  can be expressed as a function of its input  $a_*^0$  in the following way:

$$a_i^1 = \varphi \left( b_i^h + \sum_{l=1}^L \bar{h}_{il}^T * \left[ \varphi(b_l^v) + \sum_{c=1}^C \bar{v}_{lc} * a_c^0 \right] \right) \tag{1}$$

where  $L$  denotes the number of filters in the middle layer and  $\varphi(\cdot)$  is implemented by PReLU [51]. That is to say, the standard  $n \times n$  depthwise convolution is replaced by  $n \times 1$  depthwise convolution and  $1 \times n$  depthwise convolution. For a kernel of  $N \times N$ , asymmetric convolution reduces the computational complexity of each pixel from  $O(N^2)$  to  $O(N)$ . At the same time, in order to extract more abundant contextual information, we introduce the dilation rates in depthwise convolution. The depthwise convolution of  $3 \times 1$

and  $1 \times 3$  is improved to depthwise dilated convolution, which increases the receptive field and captures more complex features without reducing the resolution of the feature maps.

Considering the shallow network models, PReLU [51] performance is slightly better than ReLU [52]. Therefore, PReLU [51] was selected as a nonlinear function in our LAB module and the pre-activation scheme was adopted [53]. Normalization is used before each nonlinear function [54]. It should be noted that the use of nonlinear layers in bottlenecks affects performance [55]. Accordingly, the nonlinear layer is eliminated after the final  $1 \times 1$  pointwise convolution.

### 3.2. Feature Fusion Module

A general semantic segmentation model can be considered as a synthesis of a front-end encoder and a back-end decoder network [56]. The encoder is composed of multiple convolutional layers to obtain the overall and local features of the image. The convolutional layers and pooling layers can gradually reduce the spatial dimensions of inputted data and the feature dimensions. The decoder is composed of multiple deconvolution layers or unpooling layers and gradually restores the details and spatial dimensions of the target. The discriminable features, which are lower resolution and learned by the encoder, are semantically mapped to pixel spaces of higher resolution for pixel classification. Some decoders are simply composed of bilinear upsampling or several simple convolutions [47,57,58]. These decoders ignore low-level information, resulting in rough segmentation and low segmentation accuracy. Some decoders aggregate different stage features through complex modules and use low-level features to refine boundaries [48,59,60]. However, these decoders inevitably have the disadvantages of large calculation, high memory consumption, and slow rate.

High-level features contain semantic information. Low-level features contain rich spatial details. Due to the differences in semantic levels and spatial details, simple low-level features and high-level features are difficult to effectively integrate. Zhang et al. [58] found that introducing semantic information into low-level features and introducing spatial details into high-level features can enhance feature fusion. In this paper, we embed low-level features with rich spatial information into high-level features.

The FFM uses a U-shape style to fuse low-level features with spatial information and high-level features with semantic information, as shown in Figure 3. Firstly, the low-level features are processed by  $1 \times 1$  convolution and batch normalization to balance the scale of features. Then, low-level features are squeezed by an average pooling operation along the channel axis. Next, a sigmoid activation function is applied to generate a single-channel attention map. Afterwards, it is multiplied by the high-level features after  $3 \times 3$  convolution. In the initial FFM, it should be noted that the high-level features are not upsampled but multiplied directly by a single-channel attention map. The subsequent two Feature Fusion Modules (FFMs) need to upsample the high-level features and multiply them with the single-channel attention map. Ultimately, the high-level features are fused with the weighted features by element addition. In short, when  $\mathbf{F}_L \in \mathbb{R}^{C_L \times H_L \times W_L}$  and  $\mathbf{F}_H \in \mathbb{R}^{C_H \times H_H \times W_H}$  are inputted as low-level and high-level features into the FFM, the final output of the FFM is computed as:

$$\mathbf{F} = \varphi \left( \text{BN} \left( \sigma \left( \left( \text{AvgPool} \left( \text{BN} \left( f^{1 \times 1}(\mathbf{F}_L) \right) \right) \right) \right) \otimes f^{3 \times 3}(\mathbf{F}_H) + f^{3 \times 3}(\mathbf{F}_H) \right) \right) \quad (2)$$

where  $\varphi(\cdot)$  is realized by PReLU [51],  $\sigma$  represents the sigmoid function, *AvgPool* implies the average pooling operation, *BN* is the batch normalization,  $f^{1 \times 1}$  denotes a convolution operation with the filter size of  $1 \times 1$ ,  $f^{3 \times 3}$  indicates a convolution operation with a filter size of  $3 \times 3$ ,  $\otimes$  implies element-wise multiplication,  $+$  is element-wise addition, and  $\mathbf{F} \in \mathbb{R}^{C_L \times H_L \times W_L}$  represents the final output feature maps.

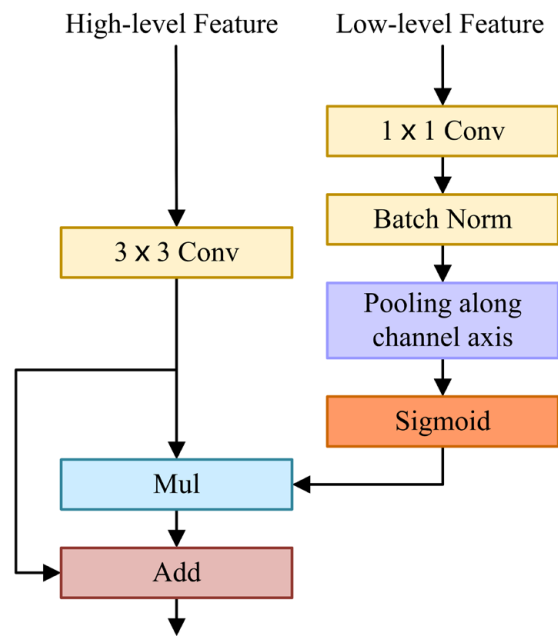


Figure 3. Feature Fusion Module.

The spatial attention map generated by low-level features reflects the importance of each pixel, which contains abundant spatial information, guides feature learning, and uses spatial details to refine boundaries. FFM extracts a spatial attention map and fuses it with high-level features containing rich semantic information effectively.

### 3.3. Network Architecture

Based on the LAB module and FFM, we have designed the NVDNet architecture of the encoder–decoder model, as shown in Figure 4. In this section, we discuss our aim to produce a network model with fast inference speed and a high Intersection-over-Union (mIoU) metric and that is also lightweight. We analyze the optimal design of MVDNet. The detailed architecture of MVDNet is set out in Table 1.

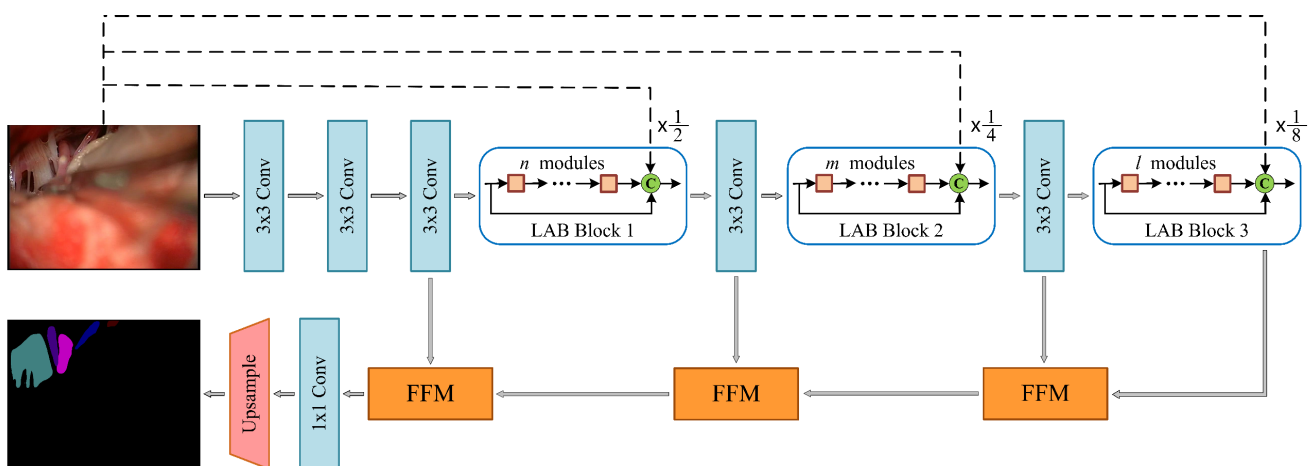


Figure 4. Architecture of the proposed MVDNet. C: concatenation; dashed lines indicate average pooling operations.

**Table 1.** Architecture details of the proposed MVDNet.

Layer	Operator	Mode	Channel	Output size
1	$3 \times 3$ Conv	Stride 2	32	$256 \times 256$
2	$3 \times 3$ Conv	Stride 1	32	$256 \times 256$
3	$3 \times 3$ Conv	Stride 1	32	$256 \times 256$
4–5	$n \times$ LAB module	Dilated 2	32	$256 \times 256$
6	$3 \times 3$ Conv	Stride 2	64	$128 \times 128$
7–8	$m \times$ LAB module	Dilated 4	64	$128 \times 128$
9	$3 \times 3$ Conv	Stride 2	128	$64 \times 64$
10–12	$l \times$ LAB module	Dilated 8	128	$64 \times 64$
13	$1 \times$ FFM module	-	128	$64 \times 64$
14	$1 \times$ FFM module	-	64	$128 \times 128$
15	$1 \times$ FFM module	-	32	$256 \times 256$
16	$1 \times 1$ Conv	Stride 1	10	$256 \times 256$
17	Bilinear interpolation	$\times 2$	10	$512 \times 512$

**Encoder:** In the MVDNet, the encoder is composed of  $3 \times 3$  convolution layers and LAB blocks. In the encoder, the  $3 \times 3$  convolution with a step size of 2 constitutes the downsampling block. The downsampling operation reduces the size of the feature maps, expands the receptive field, and extracts more contextual information. However, it is difficult to obtain accurate segmentation results because the low resolution of feature maps leads to information loss. Therefore, in order to retain sufficient spatial information, we performed three downsampling operations on the original image to obtain  $1/2$ ,  $1/4$ , and  $1/8$  feature map resolutions, respectively. Subsequently, we built a long-range shortcut connection between the input image and each LAB block, which facilitates feature reuse and compensates information loss.

We designed three LAB blocks in the MVDNet, which included several consecutive LAB modules for dense feature extraction. The first, the second, and the third LAB blocks consist of  $n$ ,  $m$ , and  $l$  LAB modules, respectively. In order to strengthen the spatial relationship and feature propagation, inter-module concatenation was introduced, which realizes the fusion of high-level features and low-level features. As mentioned in Section 3.1, we applied dilated convolution in LAB modules to obtain larger receptive fields and more complex features. Finally, the input of the LAB block and the output of the LAB block are fused with the original image after downsampling, which effectively improves the feature extraction.

**Decoder:** For the decoders, we applied three feature fusion modules to aggregate low-level features and high-level features and gradually restore resolution. Next,  $1 \times 1$  convolution and two-times upsampling were used to complete the segmentation. Compared to the decoder used in most semantic segmentation networks, the segmentation prediction is generally obtained by four-times [57] or eight-times [61,62] upsampling. The two-times upsampling, which we adopted, can retain more feature information as well as make the boundary information more complete and the semantic information clearer.

Our model belongs to an end-to-end deep learning architecture and does not depend on any backbone. It is noteworthy that the capacity of MVDNet is seriously low and that we use less than 0.72 million parameters.

#### 4. Experiments

In this section, we evaluate our proposed network on the MVD dataset, which was provided by the First Hospital of Jilin University. Firstly, we introduce the MVD dataset and the preprocessing and implementation protocol. Then, we describe the experiments conducted on the validation set of the MVD dataset to prove the effectiveness of our



network. Finally, we report the accuracy and speed results for the MVD dataset and compare them with other real-time semantic segmentation networks.

#### 4.1. Data

In the medical field, medical images generally have the characteristics of simple semantics, fewer data, and being difficult to obtain. Moreover, medical image segmentation based on deep learning requires professional annotation by doctors. Here, we cooperated with the First Hospital of Jilin University. The study involved 60 patients, 23 males and 37 females, aged 40–70. MVD data were collected with an OPMI@VARIO 700 operation microscope produced by the manufacturer ZEISS. During the period from cerebrospinal fluid release to dura suture, 3,087 MVD images were obtained and manually labeled by experts. Then, we referred to the PASVOL VOC 2012 dataset format. Finally, the MVD dataset for semantic segmentation network training was obtained. The dataset has 9 categories (10 categories when adding background). The category names and their corresponding colors are presented in Figure 5. Abbreviations of relevant medical terms are listed in Table 2.

cn5	cn7	cn9
cn10	pica	aica
aica+cn7	pica+cn7	pv

Figure 5. Colormap.

Table 2. Abbreviations of medical terms.

Abbreviation	Full Name in English
cn5	Trigeminal nerve
cn7	Facial nerve
cn9	Glossopharyngeal nerve
cn10	Vagus nerve
aica	Anterior inferior cerebellar artery
pica	Posterior inferior cerebellar artery
aica + cn7	Anterior inferior cerebellar artery and facial nerve
pica + cn7	Posterior inferior cerebellar artery and facial nerve
pv	Petrosal vein

#### 4.2. Experimental Settings

**MVD dataset:** The MVD dataset was provided by the First Hospital of Jilin University. It is an intraoperative scene dataset of microvascular decompression, including 3,087 finely annotated MVD images. We randomly selected 1,806 for training, 973 for verification, and 308 for testing. The resolution of these images was  $768 \times 576$  and there were 9 categories.

**Implementation protocol:** All experiments were performed with two 2080Ti GPU cards and CUDA 10.1 and cuDNN 7.6 on the Pytorch platform. The evaluation of runtime was performed on a single 2080Ti card. Mini-batch stochastic gradient descent (SGD), with a batch size of 8, a momentum of 0.9, and a weight decay of 0.0001, was used to train the networks. We applied the “poly” learning rate policy [31], and the initial learning rate was 0.16 with power 0.9. For data augmentation, we employed random horizontal flip, random Gaussian blur, and standardization strategies. During training, we randomly cropped the

input image to  $512 \times 512$  and set the number of epochs to 100. The mIoU metric was used to measure accuracy. The mean of cross-entropy error over all pixels was applied as the loss.

#### 4.3. Ablation Studies

In this section, we describe a series of experiments designed to prove the effectiveness of our network. These ablation studies were based on the training set of the MVD dataset and intended to evaluate our network on the validation set of the MVD dataset to observe the influence of each component in MVDNet.

**Ablation on dilation rates:** We adopted three LAB blocks with different dilation rates, LAB block 1, LAB block 2, and LAB block 3. The encoder of our network is composed of these three LAB blocks. As shown in Table 3, we set different dilation rates, 2, 4, and 8 and 4, 8, and 16. Selecting the appropriate receptive field can learn better multi-scale features. If the receptive field is too large, this will lead to the loss of small targets. For MVD images, it is more effective when the dilation rate is 2, 4, and 8.

**Table 3.** Results of the LAB encoder with different combinations of dilation rates.

Name	Dilation Rates	mIoU (%)
LAB_N2M2L4	2,4,8	73.49
LAB_N2M2L4	4,8,16	73.07

**Ablation on downsampling of the original images:** In the encoder, we downsampled the original images by 1/2, 1/4, and 1/8 and fused them into the encoder. In Table 4, after adding the downsampling operation, the accuracy was increased from 73.49% to 73.71%. This improvement effectively preserves spatial information and details and also extracts more contextual information.

**Table 4.** Results of the LAB encoder with different settings.  $n = 2, m = 2, l = 4$ .

Downsampling	Concatenation	mIoU (%)
		73.49
✓		73.71
	✓	73.76
✓	✓	73.83

**Ablation on concatenation:** We concatenated the input features and output features of the LAB blocks. As listed in Table 4, the accuracy increased by 0.27% after adding a concatenation operation in the LAB blocks. If both downsampling and concatenation operations are introduced into the network, the accuracy reaches 73.83%. Concatenation operation is applied to the encoder, effectively increasing information flow.

**Ablation on encoder depth:** We used different numbers of LAB modules for LAB block 1, LAB block 2, and LAB block 3 to change the depth of the encoder. The parameters, FLOPs, and mIoU values of different configurations are shown in Table 5. We can see that the values of  $m$  and  $l$  have a greater impact on accuracy than  $n$ , and fine segmentation results can be obtained when  $m$  and  $l$  are superimposed with four LAB modules, respectively. When  $m$  and  $l$  increase to 8, the parameters and FLOPs increase significantly and the improvement in accuracy is minor. We made a trade-off between accuracy and computational complexity, eventually setting  $n$  to 2,  $m$  to 4, and  $l$  to 4.

**Table 5.** Results of MVDNet with different depths; the number of parameters and FLOPs are estimated for a  $512 \times 512$  input.

$n$	$m$	$l$	Params(M)	FLOPs (G)	mIoU (%)
2	2	2	0.56	4.24	72.89
2	2	4	0.59	4.38	73.83
2	4	4	0.60	4.54	74.59
4	4	4	0.60	4.71	74.63
2	8	8	0.69	5.13	74.76

**Ablation on the decoder:** In MVDNet, we used a LAB block to extract features and chose a FFM to aggregate features. We applied the average pooling operation along the channel axis in the FFM to test the performance, as shown in Table 6. The average pooling operation along the channel axis can improve accuracy and ensure effective access to spatial details. This shows that embedding spatial details into high-level features through FFMs can effectively improve accuracy and obtain better pixel-level prediction.

**Table 6.** Results of the FFM module with different components.  $n = 2, m = 4, l = 4$ .

FFM	Average Pooling	mIoU (%)
w/o	-	74.59
w		77.02
w	✓	77.45

**mIoU performance:** In order to explore the influence of dilated convolution on mIoU performance, we designed two comparative experiments. In the first experiment, we removed all the dilated convolutions in MVDNet. In the other experiment, we set the first  $3 \times 3$  standard convolution of the LAB block to dilated convolutions with a dilation rate of 2. As shown in Table 7, we removed all dilated convolutions and there was a significant decrease in mIoU, from 77.45% to 75.59%. When we applied dilated convolutions with a dilation rate of 2 for standard convolution, the values of mIoU also decreased (range: 77.45% to 76.76%). The experimental results show that the dilated convolution has a significant effect on mIoU performance.

**Table 7.** Dilation of MVDNet effect on mIoU.

Model	mIoU (%)	Params (M)
MVDNet	77.45	0.72
MVDNet_w/o dilation	75.59	0.72
MVDNet_First $3 \times 3$ conv ( $r = 2$ )	76.76	0.72

#### 4.4. Comparison with the State of the Art

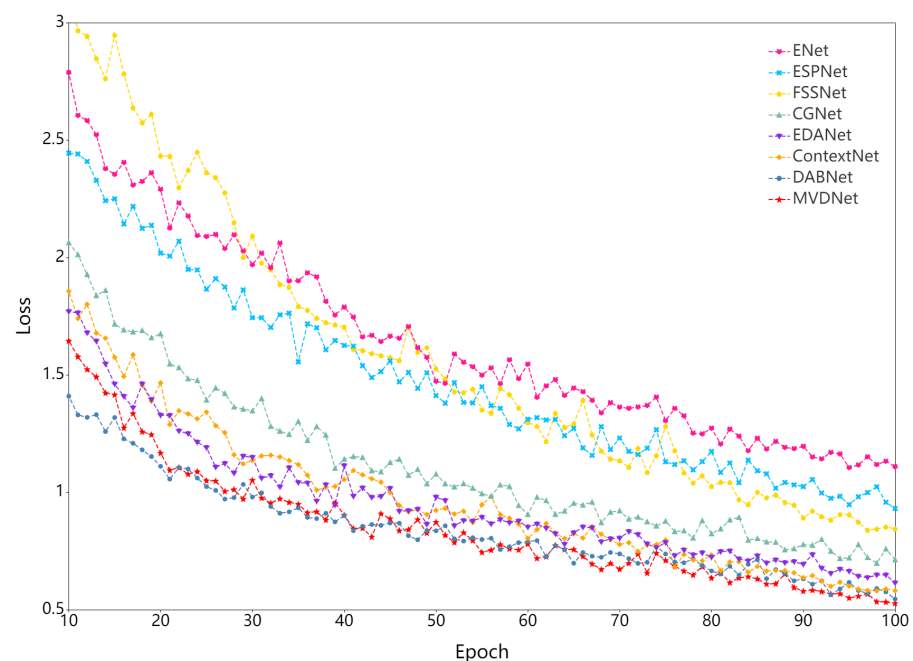
In this section, based on the study of ablation, we combined the LAB blocks and the FFMs to build a complete network and experimented with it on the MVD dataset. Firstly, all networks completed 100 epochs of training under the MVD dataset, the cross-entropy loss function, and mini-batch SGD (batch size: 8, momentum: 0.9, weight decay: 0.0001). Then, we conducted experiments to estimate the inference speed at a resolution of  $768 \times 576$  and compared the results with those of other methods. For fair comparison, we did not adopt multi-scale or multi-crop tests.

As shown in Table 8, MVDNet has 0.72 million parameters and the number of parameters is close to those of EDANet [63] and DABNet [62]. Nevertheless, the accuracy is 2.1% and 0.3% higher at the same input size. MVDNet is significantly superior to most real-time semantic segmentation methods in terms of accuracy. ESPNet [64], one of the fastest real-time networks and slightly faster than our network, only achieves 57.71% mIoU, which is 18.8% less than MVDNet. Speed comparison shows that MVDNet has a fast

inference speed given the condition of ensuring accuracy. At the same time, to facilitate observation, we intercept the training loss of all networks in 10–100 epochs, as illustrated in Figure 6. From the loss curve, it can be seen that MVDNet has faster and smoother loss attenuation than the other networks after 66 epochs.

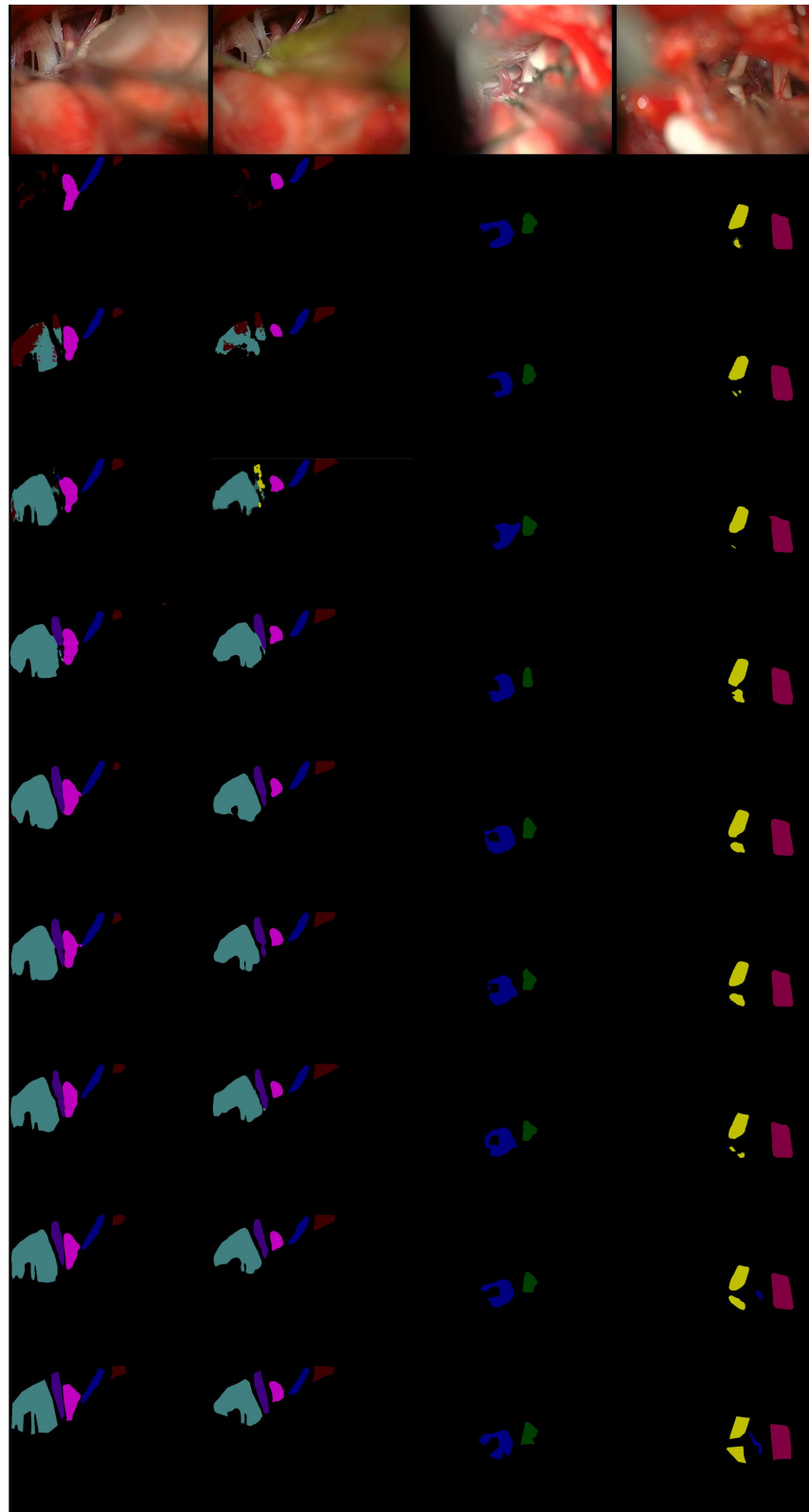
**Table 8.** Speed and accuracy comparison of MVDNet on the MVD test set.

Method	Params (M)	Time (ms)	Speed (fps)	mIoU (%)
ENet [65]	0.36	12.7	78.5	51.69
ESPNet [64]	0.19	6.4	156.4	57.71
FSSNet [66]	0.17	7.6	131.7	61.27
CGNet [67]	0.49	11.4	87.4	71.35
EDANet [63]	0.69	8	125	74.49
ContextNet [68]	0.88	6.1	163.3	75.62
DABNet [62]	0.75	7.7	129.1	76.29
MVDNet (Ours)	0.72	7.3	137.6	76.59



**Figure 6.** Training loss curve.

It can be seen from Figure 7, in the first row, that only our MVDNet accurately could accurately locate the segmentation boundary of “aica” and that the object contour is clear. In the second and third rows, ENet [65] has an error segmentation and fails to segment “cn10”. ESPNet [64] and FSSNet [66] have obvious multi-pixel mixing problems. CGNet [67], EDANet [63], ContextNet [68], and DABNet [62] do not accurately locate the segmentation boundary of “cn10”, showing obvious loss of target contour segmentation. In the third row, only our MVDNet segments “aica” and the segmentation of “cn5” are complete. The proposed method has higher segmentation performance and contains more feature information.



**Figure 7.** Visual comparison on MVD validation set. From top to bottom: input images, segmentation outputs from ENet [65], ESPNet [64], FSSNet [66], CGNet [67], EDANet [63], ContextNet [68], DABNet [62], our MVDNet, and ground truth.

With the test set, results were compared for patients with different background information (age and gender). According to the different age groups, there were 107 images

for the 40–50-year-old group, 117 images for the 50–60 group, and 84 images for the 60–70 group. By gender, there were 124 images for males and 184 images for females.

It can be seen from Tables 9 and 10 that the operation of MVD is mostly concentrated in “pica + cn7”, that the main responsible vessel is “pica”, and that the segmentation accuracy of cerebral vessels is less than that of cranial nerves. In Table 9, with increasing age, the segmentation accuracy of cerebral vessels is significantly reduced due to the more tortuous cerebral vessel arrangements characteristic of the elderly. In Table 10, there is a small gap between male and female mIoU values. Comparing Table 9 with Table 10, it can be found that the segmentation accuracy of cerebral vessels is mainly affected by age. Age has a greater impact on final segmentation results than gender.

**Table 9.** Accuracy comparison for different age groups on the MVD test set.

Age	mIoU (%)	cn5	cn7	cn9	cn10	aica + cn7	pica + cn7	pica	aica	pv
40–50	76.87	82.51	82.52	74.12	76.93	77.46	88.43	74.08	71.43	64.35
50–60	77.11	82.62	84.41	79.39	81.37	73.55	87.89	73.69	70.4	60.64
60–70	76.25	84.05	87.29	75.08	81.29	79.41	88.12	70.33	68.43	52.37

**Table 10.** Accuracy comparison for gender on the MVD test set.

Gender	mIoU(%)	cn5	cn7	cn9	cn10	aica + cn7	pica + cn7	pica	aica	pv
Male	76.4	83.18	84.23	77.22	80.12	71.6	88.71	73.68	68.62	60.21
Female	76.29	85.19	85.19	73.06	74.41	78.88	87.93	72.38	71.74	60.14

## 5. Discussion

In previous studies, there have been few real-time and accurate segmentations of cerebral vessels and cranial nerves in microvascular decompression. In this study, a new encoder–decoder structure was adopted which has the characteristics of accuracy and speed and can accurately and quickly complete the segmentation of cerebral vessels and cranial nerves in microvascular decompression. Compared with previous studies, the encoder has a simpler structure and fewer convolutional layers, yet it can obtain more contextual information. The decoder more effectively integrates high-level and low-level features.

During microvascular decompression, the surgeon usually judges the cerebral vessels and cranial nerves according to the treatment group and experience. However, the bone flap diameter is only 2.5 cm, and the operation space is small. There will also be a small amount of bleeding and cerebrospinal fluid present at any time during the operation, which will affect the operation of the surgeon and cause a lot of inconvenience, leading to many uncertainties and risks during the operation. The proposed method realizes the rapid and accurate segmentation of cerebral vessels and cranial nerves, reduces mental pressure on the surgeon, and provides a basis for rapid decision-making and judgment on the part of the surgeon. It is beneficial to reduce the release of cerebrospinal fluid during the operation, avoid excessive traction of nerves and blood vessels, effectively reduce surgical trauma, and reduce the occurrence of postoperative complications. However, the limitation of this paper is that we only used data obtained by a single medical institution (the First Hospital of Jilin University) to train, validate, and test the proposed network. The generalization performance of our approach needs to be further improved in the future.

## 6. Conclusions

For MVD images, in order to improve the speed and accuracy of real-time semantic segmentation, this paper presents the Microvascular Decompression Network (MVDNet). We have proposed a new Light Asymmetric Bottleneck (LAB) module to extract contextual information and designed an encoder based on this module. The decoder applied a

Feature Fusion Module (FFM) to aggregate different features. The ablation experiments showed that the LAB blocks effectively extracted the contextual features and that the Feature Fusion Modules (FFMs) integrated the deep contextual features and shallow spatial features efficiently. We achieved a result of 76.59 % mIoU on the MVD test set at 137 FPS. Compared with other real-time methods, our network has significant improvements in terms of accuracy and speed.

**Author Contributions:** Designation, methodology: R.B. and H.S.; data analysis, writing—original draft preparation: R.B.; funding acquisition: H.S.; data curation: X.L.; writing—review and editing: R.B., X.L., and S.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Jilin Scientific and Technological Development Program (no. 20200404155YY), the Jilin Scientific and Technological Development Program (no. 20200401091GX), and the foundation of Bethune Center for Medical Engineering and Instrumentation (Changchun) (no. BQEGCZX2019047).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the First Hospital of Jilin University (protocol code: 2019-194; date of approval: 9 April 2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from the First Hospital of Jilin University and are available from the authors with the permission of the First Hospital of Jilin University.

**Acknowledgments:** Thanks to the First Hospital of Jilin University for providing the microvascular decompression image dataset.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Bennetto, L.; Patel, N.K.; Fuller, G.J.B. Trigeminal neuralgia and its management. *BMJ* **2007**, *334*, 201–205. [[CrossRef](#)] [[PubMed](#)]
2. Kıziltan, M.E.; Gunduz, A.J.N.S. Reorganization of sensory input at brainstem in hemifacial spasm and postparalytic facial syndrome. *Neurol. Sci.* **2018**, *39*, 313–319. [[CrossRef](#)] [[PubMed](#)]
3. Feng, B.; Zheng, X.; Zhang, W.; Yang, M.; Tang, Y.; Zhong, J.; Hua, X.; Ying, T.; Li, S.J. Surgical treatment of pediatric hemifacial spasm patients. *Acta Neurochir.* **2011**, *153*, 1031–1035. [[CrossRef](#)] [[PubMed](#)]
4. Apra, C.; Lefaucheur, J.-P.; Le Guérinel, C.J.N.R. Microvascular decompression is an effective therapy for trigeminal neuralgia due to dolichoectatic basilar artery compression: Case reports and literature review. *Neurosurg. Rev.* **2017**, *40*, 577–582. [[CrossRef](#)]
5. Jannetta, P.J.; Mclaughlin, M.R.; Casey, K.F. Technique of microvascular decompression. *Neurosurg. Focus* **2005**, *18*, 1–5. [[CrossRef](#)]
6. Miller, L.E.; Miller, V.M. Safety and effectiveness of microvascular decompression for treatment of hemifacial spasm: A systematic review. *Br. J. Neurosurg.* **2012**, *26*, 438–444. [[CrossRef](#)]
7. Sindou, M.; Mercier, P.J.N. Microvascular decompression for hemifacial spasm: Outcome on spasm and complications. A review. *Neurochirurgie* **2018**, *64*, 106–116. [[CrossRef](#)]
8. Gardner, W.J. Concerning the mechanism of trigeminal neuralgia and hemifacial spasm. *J. Neurosurg.* **1962**, *19*, 947–958. [[CrossRef](#)]
9. Jannetta, P.J. Observations on the etiology of trigeminal neuralgia, hemifacial spasm, acoustic nerve dysfunction and glossopharyngeal neuralgia. Definitive microsurgical treatment and results in 117 patients. *Minim. Invasive Neurosurg.* **1977**, *20*, 145–154. [[CrossRef](#)]
10. Hyun, S.-J.; Kong, D.-S.; Park, K.J. Microvascular decompression for treating hemifacial spasm: Lessons learned from a prospective study of 1174 operations. *Neurosurg. Rev.* **2010**, *33*, 325–334. [[CrossRef](#)]
11. Ogungbo, B.I.; Kelly, P.; Kane, P.J.; Nath, F.P. Microvascular decompression for trigeminal neuralgia: Report of outcome in patients over 65 years of age. *Br. J. Neurosurg.* **2000**, *14*, 23–27. [[PubMed](#)]
12. Miyazaki, H.; Deveze, A.; Magnan, J.J. Neuro-otologic surgery through minimally invasive retrosigmoid approach: Endoscope assisted microvascular decompression, vestibular neurectomy, and tumor removal. *Laryngoscope* **2005**, *115*, 1612–1617. [[CrossRef](#)] [[PubMed](#)]
13. Broggi, M.; Acerbi, F.; Ferroli, P.; Tringali, G.; Schiariti, M.; Broggi, G.J. Microvascular decompression for neurovascular conflicts in the cerebello-pontine angle: Which role for endoscopy? *Acta Neurochir.* **2013**, *155*, 1709–1716. [[CrossRef](#)]
14. Wick, C.C.; Arnaoutakis, D.; Barnett, S.L.; Rivas, A.; Isaacson, B.J. Endoscopic transcanal transpromontorial approach for vestibular schwannoma resection: A case series. *Otol. Neurotol.* **2017**, *38*, e490–e494. [[CrossRef](#)] [[PubMed](#)]

15. El Refaee, E.; Langner, S.; Marx, S.; Rosenstengel, C.; Baldauf, J.; Schroeder, H.W. Endoscope-assisted microvascular decompression for the management of hemifacial spasm caused by vertebrobasilar dolichoectasia. *World Neurosurg.* **2019**, *121*, e566–e575. [[CrossRef](#)]
16. El Refaee, E.; Langner, S.; Baldauf, J.; Matthes, M.; Kirsch, M.; Schroeder, H.W. Value of 3-dimensional high-resolution magnetic resonance imaging in detecting the offending vessel in hemifacial spasm: Comparison with intraoperative high definition endoscopic visualization. *Neurosurgery* **2013**, *73*, 58–67. [[CrossRef](#)]
17. Cruccu, G.; Gronseth, G.; Alksne, J.; Argoff, C.; Brainin, M.; Burchiel, K.; Nurmikko, T.; Zakrzewska, J. AAN-EFNS guidelines on trigeminal neuralgia management. *Eur. J. Neurol.* **2008**, *15*, 1013–1028. [[CrossRef](#)]
18. Macellari, F.; Paciaroni, M.; Agnelli, G.; Caso, V. Neuroimaging in intracerebral hemorrhage. *Stroke* **2014**, *45*, 903–908. [[CrossRef](#)]
19. Takahashi, M.; Uchino, A.; Suzuki, C.J.S.; Anatomy, R. Anastomosis between accessory middle cerebral artery and middle cerebral artery diagnosed by magnetic resonance angiography. *Surg. Radiol. Anat.* **2016**, *39*, 685–687. [[CrossRef](#)]
20. Deng, X.; Zhang, Z.; Zhang, Y.; Zhang, D.; Wang, R.; Ye, X.; Xu, L.; Wang, B.; Wang, K.; Zhao, J.J. Comparison of 7.0- and 3.0-T MRI and MRA in ischemic-type moyamoya disease: Preliminary experience. *J. Neurosurg.* **2016**, *124*, 1716–1725. [[CrossRef](#)]
21. Wang, R.; Li, C.; Wang, J.; Wei, X.; Li, Y.; Zhu, Y.; Zhang, S.J. Threshold segmentation algorithm for automatic extraction of cerebral vessels from brain magnetic resonance angiography images. *J. Neurosci. Methods* **2015**, *241*, 30–36. [[CrossRef](#)] [[PubMed](#)]
22. Bhuiyan, A.; Nath, B.; Chua, J.J. An adaptive region growing segmentation for blood vessel detection from retinal images. In Proceedings of the Second International Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, Spain, 8–11 March 2007; pp. 404–409.
23. Huang, Z.; Li, Q.; Hong, H.; Zhang, T.; Sang, N. Non-local Hessian-based weighted filter for 2D noisy angiogram image enhancement with cardiovascular tree structure preservation. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 2129–2134.
24. Frangi, A.F.; Niessen, W.J.; Vincken, K.L.; Viergever, M.A. Multiscale vessel enhancement filtering. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Cambridge, MA, USA, 11–13 October 1998; pp. 130–137.
25. Sato, Y.; Nakajima, S.; Atsumi, H.; Koller, T.; Gerig, G.; Yoshida, S.; Kikinis, R. 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In Proceedings of the CVRMed-MRCAS'97, Grenoble, France, 19–22 March 1997; pp. 213–222.
26. Erdt, M.; Raspe, M.; Suehling, M. Automatic hepatic vessel segmentation using graphics hardware. In Proceedings of the International Workshop on Medical Imaging and Virtual Reality, Tokyo, Japan, 1–2 August 2008; pp. 403–412.
27. Jerman, T.; Pernuš, F.; Likar, B.; Špiclin, Ž. Enhancement of vascular structures in 3D and 2D angiographic images. *IEEE Trans. Med. Imaging* **2016**, *35*, 2107–2118. [[CrossRef](#)] [[PubMed](#)]
28. Youssif, A.A.-H.A.-R.; Ghalwash, A.Z.; Ghoneim, A.A. Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *IEEE Trans. Med. Imaging* **2007**, *27*, 11–18. [[CrossRef](#)] [[PubMed](#)]
29. Hu, Z.; Niemeijer, M.; Abramoff, M.D.; Lee, K.; Garvin, M.K. Automated segmentation of 3-D spectral OCT retinal blood vessels by neural canal opening false positive suppression. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Beijing, China, 20–24 September 2010; pp. 33–40.
30. Cheng, Y.; Hu, X.; Wang, J.; Wang, Y.; Tamura, S.J. Accurate vessel segmentation with constrained B-snake. *IEEE Trans. Image Process.* **2015**, *24*, 2440–2455. [[CrossRef](#)] [[PubMed](#)]
31. Lee, S.-H.; Lee, S. Adaptive Kalman snake for semi-autonomous 3D vessel tracking. *Comput. Methods Programs Biomed.* **2015**, *122*, 56–75. [[CrossRef](#)] [[PubMed](#)]
32. Wang, J.; Zhao, S.; Liu, Z.; Tian, Y.; Duan, F.; Pan, Y. An active contour model based on adaptive threshold for extraction of cerebral vascular structures. *Comput. Math. Methods Med.* **2016**, *2016*, 6472397. [[CrossRef](#)]
33. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
34. Shvets, A.A.; Iglovikov, V.I.; Rakhlin, A.; Kalinin, A.A. Angiodysplasia detection and localization using deep convolutional neural networks. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 612–617.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
37. Livne, M.; Rieger, J.; Aydin, O.U.; Taha, A.A.; Akay, E.M.; Kossen, T.; Sobesky, J.; Kelleher, J.D.; Hildebrand, K.; Frey, D.J. A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Front. Neurosci.* **2019**, *13*, 97. [[CrossRef](#)]
38. Hilbert, A.; Madai, V.I.; Akay, E.M.; Aydin, O.U.; Behland, J.; Sobesky, J.; Galinovic, I.; Khalil, A.A.; Taha, A.A.; Wuerfel, J. BRAVE-NET: Fully automated arterial brain vessel segmentation in patients with cerebrovascular disease. *Front. Artif. Intell.* **2020**, *3*, 78. [[CrossRef](#)]



39. Wang, S.; Hua, Y.; Cao, Y.; Song, T.; Xue, Z.; Gong, X.; Wang, G.; Ma, R.; Guan, H. Deep learning based fetal middle cerebral artery segmentation in large-scale ultrasound images. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 532–539.
40. Wang, Y.; Yan, G.; Zhu, H.; Buch, S.; Wang, Y.; Haacke, E.M.; Hua, J.; Zhong, Z. JointVesselNet: Joint volume-projection convolutional embedding networks for 3D cerebrovascular segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Virtual, 4–8 October 2020; pp. 106–116.
41. Wang, Y.; Yan, G.; Zhu, H.; Buch, S.; Wang, Y.; Haacke, E.M.; Hua, J.; Zhong, Z.J. VC-Net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 1301–1311. [[CrossRef](#)]
42. Zhang, H.; Xia, L.; Song, R.; Yang, J.; Hao, H.; Liu, J.; Zhao, Y. Cerebrovascular segmentation in MRA via reverse edge attention network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Virtual, 4–8 October 2020; pp. 66–75.
43. Nazir, A.; Cheema, M.N.; Sheng, B.; Li, H.; Li, P.; Yang, P.; Jung, Y.; Qin, J.; Kim, J.; Feng, D. OFF-eNET: An optimally fused fully end-to-end network for automatic dense volumetric 3D intracranial blood vessels segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 7192–7202. [[CrossRef](#)]
44. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
45. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
46. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
47. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
48. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—improve semantic segmentation by global convolutional network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
49. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R.J. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 263–272. [[CrossRef](#)]
50. Alvarez, J.; Petersson, L. Decomposeme: Simplifying convnets for end-to-end learning. *arXiv* **2016**, arXiv:1606.05426.
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
52. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.
54. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
55. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
56. Lateef, F.; Ruichek, Y.J.N. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. [[CrossRef](#)]
57. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
58. Zhang, Z.; Zhang, X.; Peng, C.; Xue, X.; Sun, J. Exfuse: Enhancing feature fusion for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 269–284.
59. Jégou, S.; Drozdal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisù: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
60. Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
61. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
62. Li, G.; Yun, I.; Kim, J.; Kim, J. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation. *arXiv* **2019**, arXiv:1907.11357.

63. Lo, S.-Y.; Hang, H.-M.; Chan, S.-W.; Lin, J.-J. Efficient dense modules of asymmetric convolution for real-time semantic segmentation. In Proceedings of the 2019 ACM Multimedia Asia, Beijing, China, 15–18 December 2019; pp. 1–6.
64. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 552–568.
65. Lee, J.; Lee, Y. A Deep Neural Network Architecture for Real-Time Semantic Segmentation on Embedded Board. *J. Korean Inst. Inf. Sci. Eng.* **2018**, *45*, 94–98.
66. Zhang, X.; Chen, Z.; Wu, Q.J.; Cai, L.; Lu, D.; Li, X. Fast semantic segmentation for scene perception. *IEEE Trans. Ind. Inform.* **2018**, *15*, 1183–1192. [[CrossRef](#)]
67. Wu, T.; Tang, S.; Zhang, R.; Cao, J.; Processing, Y. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 1169–1179. [[CrossRef](#)]
68. Poudel, R.P.; Bonde, U.; Liwicki, S.; Zach, C. Contextnet: Exploring context and detail for semantic segmentation in real-time. *arXiv* **2018**, arXiv:1805.04554.