IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 3, 2020, accepted December 12, 2020, date of publication December 16, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045280

Fully Convolutional Pyramidal Networks for Semantic Segmentation

FENGXIAO LI[1], ZOURONG LONG[1], PENG HE[2], PENG FENG[2], XIAODONG GUO[2], XUEZHI REN[2], BIAO WEI[2], MINGFU ZHAO[1], AND BIN TANG[1]

[1]Intelligent optical fiber sensing technology Chongqing University Engineering Research Center, Chongqing University of Technology, Chongqing 400054, China

[2]Key Lab of Optoelectronic Technology and Systems of the Education Ministry of China, Chongqing University, Chongqing 400044, China

Corresponding authors: Zourong Long (longzourong@cqut.edu.cn) and Bin Tang (tangbin@cqut.edu.cn)

This work was supported in part by National Natural Science Foundation of China 61805029, in part by the National Key Research and Development Program of China under Grant 2016YFC0104609 and Grant 2019YFC0605203, in part by the Fundamental Research Funds for the Central Universities under Grant 2019CDYGYB019, and in part by the Scientific and Technological Research Program of Chongqing Municipal Educational Commission under Grant KJQN201904007.

ABSTRACT Semantic segmentation networks focus on the scene parsing of an unrestricted open scene. The typical segmentation architectures are stacks consisting of convolutional layers, which are used to extract semantic features. The feature map dimension is sharply changed at sampling units for most of networks, which ensure effective propagation of the gradient in deep nets. In this article, we proposed a state-of-the-art network model named Fully Convolutional Pyramidal Networks (FC-PRNet), which employs pyramidal residual structure to change the feature map dimension at all convolutional layers. This design is an effective way of improving generalization ability and optimizing parameters, and FC-PRNet could achieve excellent capability of semantic extraction. We used urban scene benchmark CamVid and KITTI dataset to test our network, the experimental results show that FC-PRNet achieves better results without any pre-training or post-treatment module. Moreover, due to smart construction of pyramidal residual structures, FC-PRNet has less parameters than other existing networks trained on these datasets.

INDEX TERMS Semantic segmentation, artificial intelligence, lightweight model, KIITI data sets.

I. INTRODUCTION

In 2012, Hinton proposed AlexNet [1], which occupies an important historical position in the field of convolutional neural networks (CNNs). Nowadays CNNs are driving advances in different vision tasks such as: image classification, style transfer, object detection, and local recognition. Scene parsing is a fundamental topic in local recognition tasks. Its goal is to assign each pixel in the image a category label. Scene parsing frameworks are mostly based on Fully Convolutional Networks (FCNs) [2], which is one of the natural extensions of CNNs tackling per pixel prediction problems of semantic segmentation. FCNs design an up-sampling path after CNNs and introduces skip connections compensating for the feature loss in pooling layers. Due to the up-sampling path, FCNs can process the input images at any resolution and meet the requirements of most images take out pixel.

A large number of CNNs networks have been extended to FCNs. For the more traditional especially Deep Residual Networks (ResNet) [3] implements hundreds of convolutional layers by introducing a new building block, which consists of two convolutional layers and a shortcut. The block dose the sum of the input and non-linear transformation of input. ResNet has a problem of diminishing feature reuse, which is that gradients are not forced through the convolution layers in deep networks [6], [7]. Many scholars have studied this problem from the network structure and training process [6]–[8]. FCNs extended from ResNet have achieved very good results [4], [5].

Recently, Dongyoon Han proposed Deep Pyramidal Residual Networks (PyramidNet) [9], which utilizes a new method of dimension growth. It is a strictly linear relationship between the dimensions of the network and the number of

The associate editor coordinating the review of this manuscript and approving it for publication was Yilun Shang.

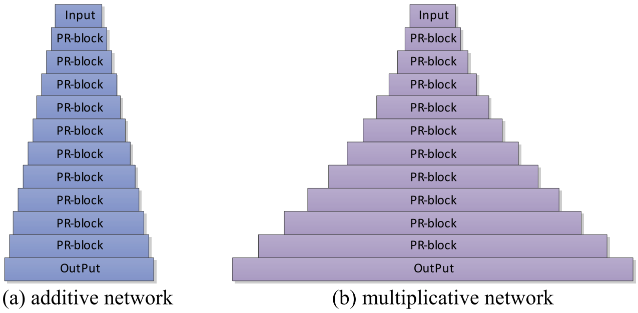
convolutional layers. PyramidNet shows good performance in solving the diminishing feature reuse problem. The linear increase of dimension leads to fewer parameters of deep convolutional layers and this structure has a high utilization rate of parameters to improve accuracy [9].

In this article, we introduce the PyramidNet architecture to FCNs for semantic segmentation and proposed a new network named Fully Convolutional Pyramidal Residual Networks (FC-PRNet). We designed residual blocks in up/down-sampling paths, and the up/down-sampling paths form a complete semantic segmentation network by connecting sampling layers with several skip connections [10]. The FC-PRNet achieves good segmentation results in CamVid and KITTI datasets. In part II, we will introduce pyramidal residual blocks and the constructions of FC-PRNet in detail. After introducing different kinds of FC-PRNet in part III, we will show the results of two urban scene benchmark datasets. Part IV is the summary of this article and the arrangement of future work.

II. MATERIALS AND METHODS

A. PYRAMIDAL RESIDUAL NETWORKS

Most CNNs [7], [10]–[13] employ an approach whereby feature maps dimensions and feature maps sizes change at down-sampling layers. In the case of the original ResNet, the feature size is down-sampled by half and the number of dimensions is doubled. PyramidNet is a derivative network of ResNet and it proposes a new method of dimensions growth: the dimension is increased by a value during each extracting layer and the feature size still decreases during down-sampling layers. To achieve this dimensional variation, PyramidNet designs a unique feature extraction unit referred to as a pyramidal residual block (PR-block). There are two kinds of PyramidNet, which are additive mode and multiplicative mode as shown in Figure 1.



(a) additive network (b) multiplicative network

FIGURE 1. Dimensional schematic. The width of each block denotes the output dimension of the PR-block. The wider the block is, the more parameters the network has. Even when the number of layers are the same,(b) has many more parameters than (a).

PR-blocks, as shown in Figure 2, are the basic building bolcks of PyramidNet. Y denotes the output, whose dimension is n bigger than that of input x. Due to different dimensions among individual PR-blocks, an identity-mapping shortcut is unusable. Therefore, only a zero-padded shortcut or a projection shortcut is available. In view that a projection shortcut will hamper feature propagation and lead to a problem of optimization [14], PyramidNet adopted a zero-padded shortcut, which does not introduce additional non-zero parameters. Moreover, each zero-padded shortcut can provide a mixture of the residual network and the plain network. With the dimension increasing at each unit, the mixture effect is more marked.

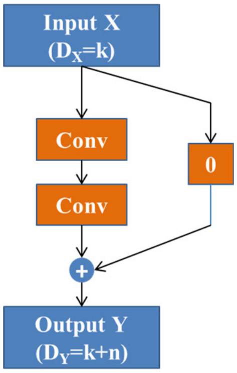


FIGURE 2. PR-block: D_X denotes the dimension of X and D_Y denotes the dimension of Y.

The variation of dimensions between adjacent PR-block is called growth-rate. It can be constructed in two different ways: additive mode expressed as (1) and multiplicative mode expressed as (2):

$$D_k = \begin{cases} D_{in} & k = 1 \\ D_{k-1} + \alpha & k \geq 2 \end{cases} \tag{1}$$

$$D_k = \begin{cases} D_{in} & k = 1 \\ D_{k-1} \times \beta^{\frac{1}{k}} & k \geq 2 \end{cases} \tag{2}$$

where α and β are both growth-rate and k is the number of PR-blocks. The main difference between additive networks and multiplicative networks is that the dimension of additive networks increases linearly, whereas the dimension of multiplicative networks increases geometrically. The process of multiplication network is similar to that of original deep network architectures, whose dimension of input-site increases slowly and the dimension of output-site increases sharply. It means that multiplicative networks have more parameters than additive networks as the network gets deeper.

Two kinds of PyramidNet show similar performance due to unobvious difference in their significant structures when they are shallow. As the nets get deeper, they show some differences in capabilities of feature extraction. The feature map dimensions of multiplicative networks tend to be much larger at the output-side than that of additive networks, and redundant parameters will make the network harder to train and affect network performance. Comparative experiments show that additive network has better performance than multiplicative the network.

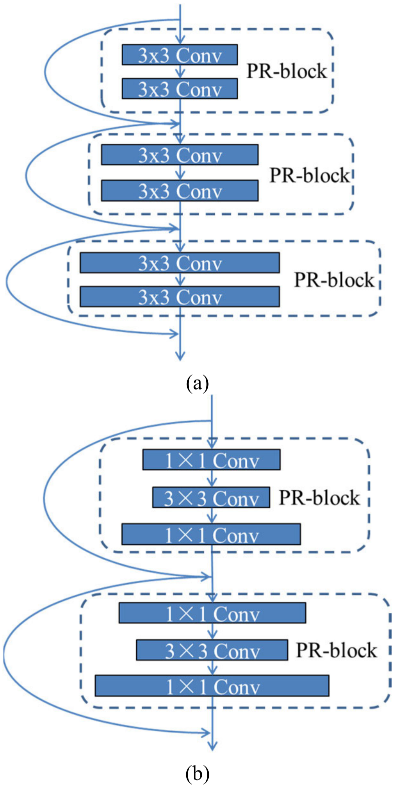


FIGURE 3. Schematic illustration of PR-block: (a) basic PR-block; (b) bottleneck PR-block (The size of the block indicates the number of network parameters).

B. FULLY CONVOLUTIONAL PYRAMIDAL RESIDUAL NETWORKS

The PyramidNet architecture described in section 2.1 builds the down-sampling path of our FC-PRNet. In order to recover the high-dimensional feature, FC-PRNet introduces a corresponding up-sampling path, which is composed of PR-block, up-sampling layers and skips connections. We design basic PR-block and bottleneck PR-block as shown in Figure 3. Basic PR-block consists of two 3 × 3 Conv. While, bottleneck PR-block uses a combination of 1 × 1 Conv, 3 × 3 Conv and 1 × 1 Conv, which can reduce the parameters effectively.

FC-PRNet adopted two sampling layers to change the feature size. In the down-sampling path, we introduced a transition down layer (TD) to reduce the feature size. In the up-sampling path, we introduced a transition up layer (TU) to recover the feature size. Note that TD has an operation of pooling, which will lead to some losses of information from earlier PR-blocks. Nevertheless, this information is available in the down-sampling path of the network and can be passed via skip connections. Besides two kinds of transition layers, different structures of PR-blocks are used in two paths. Different from the PR-blocks in the down-sampling path, the PR-blocks in the up-sampling path gradually reduces the feature map dimensions. Figure 4 shows the schematic diagram of FC-PRNet. The dimension increases gradually when feature maps extracting in the down-sampling path and decreases gradually in up-sampling path. Several skip connections connect TD and TU.

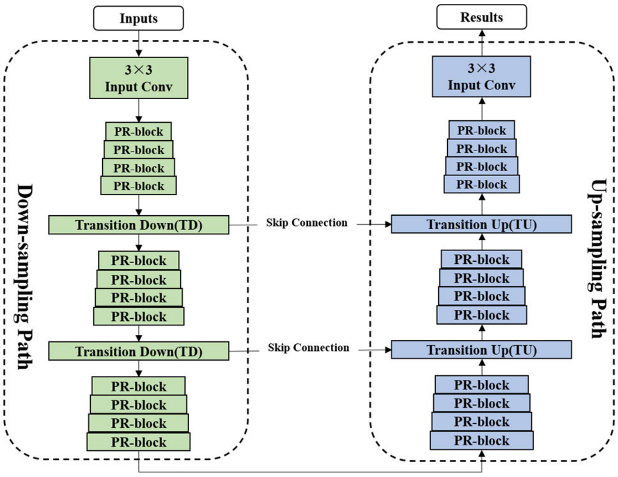


FIGURE 4. Schematic diagram of FC-PRNet.

III. EXPERIMENTS

A. ARCHITECTURE

In this section, we will introduce architectures of FC-PRNet with additive and multiplicative mode used in the subsequent experiments. Firstly, in Figure 5, we define 6 kinds of building blocks used in the network. Differ from PR-block in the down-sampling path, PR-block in up-sampling path uses 1 × 1 convolutional layers for adjustment of dimensions. TD consists of BN, ReLU, 1 × 1 Conv and Maxpooling. TU only contains one 3 × 3 Transposed Conv.

Secondly, we define additive FC-PRNets and multiplicative FC-PRNets. Both kinds of networks were modeled in basic PR-blocks and bottleneck PR-Block. We summarize all kinds of FC-PRNet in Table 3 and take FC-PRNet94 with basic PR-Blocks as an example to introduce the networks. FC-PRNet94 with basic PR-Block is built from 94 convolutional layers: a layer with 48 convolution cores to process RGB images, 48 layers in the down-sampling path, 44 layers in the up-sampling path and a final layer to process output data followed by a SoftMax non-linearity to predict each pixel. If the growth-rate α is 4, the biggest dimension is 128. Compared with additive FC-PRNets, multiplicative FC-PRNets with basic PR-Block are much wider and the biggest dimension is 1840 when the growth-rate β is 1.2.

Thirdly, we test out models using a desktop computer with an Intel I7 4790k CPU and a TITAN XP GPU. We use minimum pixel cross-entropy and Adam (Adaptive Moment Estimation) in training. The learning rate was set to 0.001, reduced by 5% per epoch and the batch_size is 4. We monitor the mean intersection over union (MIoU) and the global accuracy.

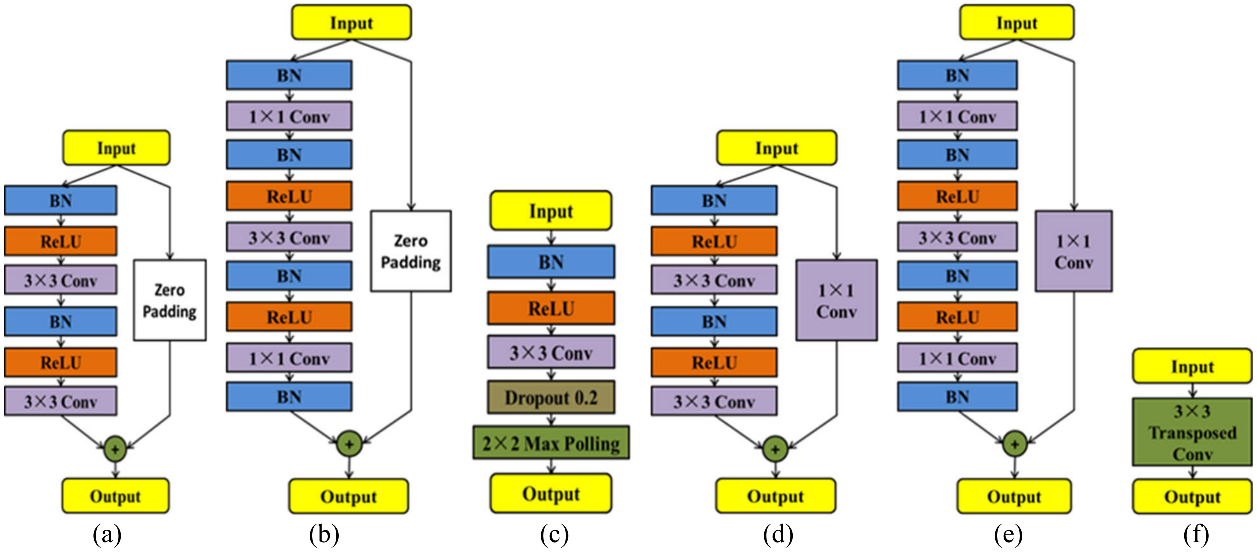


FIGURE 5. Layer structure: (a) basic PR-Block in down-sampling path; (b) Bottleneck PR-Block in down-sampling path; (c) transition down layers; (d) basic PR-Block in up-sampling path; (e) Bottleneck PR-Block in up-sampling path; (f) transition up layers.

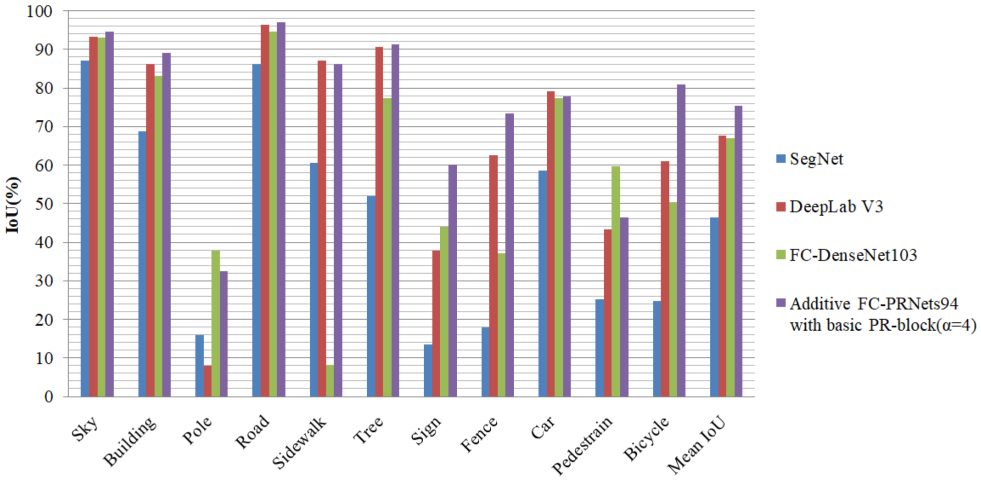


FIGURE 6. IoU on CamVid.

B. CAMVID DATASET

CamVid [15] is the first video collection with semantic tags, providing each pixel with an associated tag in 11 semantic classes. Differ from the fixed-position mode of other videos, this data set is taken from the perspective of a driving automobile. We use data of CamVid Group III, including 367 frames for training, 101 frames for validation and 233 frames for testing, with a resolution of 360 × 480 per frame. We trained and predicted with full-size images without any post-treatment or pre-training module.

Table 1 and Figure 6 show the comparison of FC-PRNet with other networks. The experimental results show that additive FC-PRNets94 with basic PR-block ($\alpha = 4$) gets the best results. The pyramid residual structure has a maximum result, and can effectively improve the MIoU of all kinds by 15%-20%, especially trees, bicycles and road signs. It is noteworthy that the image in the camera corresponds to the video frame, so the data set contains temporal information. If we introduced advanced video timing processing methods, the performance of the network can be improved.

Figure 7 shows some segmentation results of FC-PRNet94 with basic PR-block ($\alpha = 4$) on CamVid datasets. The qualitative results are in good agreement with the quantitative results, showing clear segments explaining many details, such as cars, pedestrians, trees and the rest of the labels of the dataset. In the category of poor segmentation, we can see that there are some misidentifications of road signs, columns, buildings and cars.

C. KIITI DATASET

The KITTI [18], [19] road benchmark is a comprehensive dataset and it is very popular with road detection researchers.

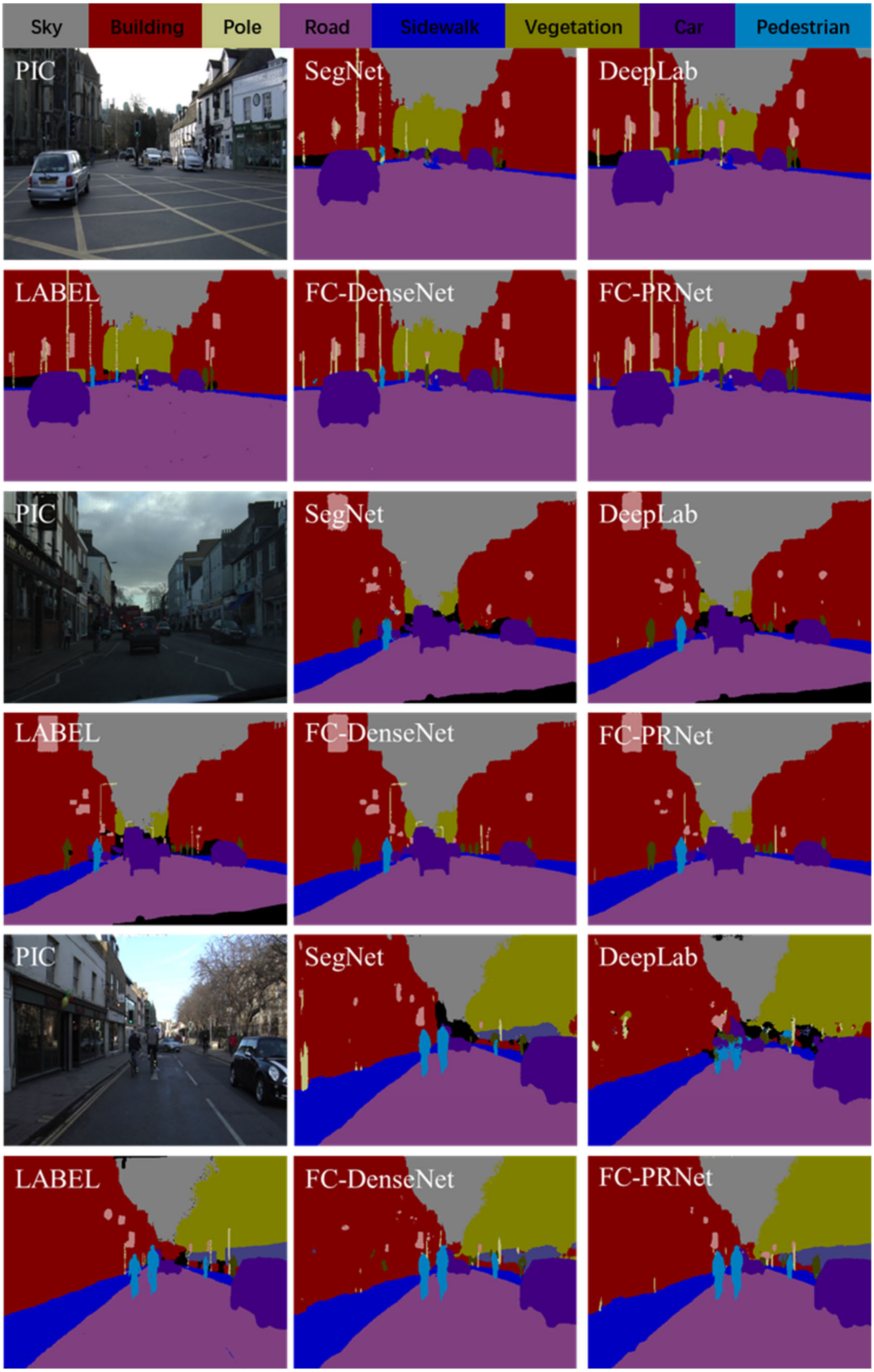


FIGURE 7. CamVid segmentation results.

TABLE 1. Results of Various Networks in CamVid:FC-PRNets Perform Better Than Multiplicative FC-PRNets;the Network Works Best When the Growth-Rate α is 4.

Model	Pretrained	Parameters (M)	Sky (%)	Building (%)	Pole (%)	Road (%)	Sidewalk (%)	Tree (%)	Sign (%)	Fence (%)	Car (%)	Pedestrain (%)	Bicycle (%)	Mean IoU (%)	Global Accuracy (%)
Additive FC-PRNets94 with basic PR-block(α=4)	×	**6.5**	**94.6**	**89.0**	32.6	**97.0**	86.3	**91.2**	**60.1**	**73.3**	**77.8**	46.5	**80.9**	**75.4**	**93.6**
Additive FC-PRNets95 with bottlenneck PR-block(α=6)	×	6.1	92.7	86.6	31.8	95.2	85.4	90.7	59.5	71.9	75.1	45.1	78.8	73.9	93.2
Additive FC-PRNets94 with basic PR-block(α=8)	×	14.4	93.4	87.1	29.9	93.7	85.7	89.4	57.5	72.6	70.1	42.9	77.4	72.7	92.1
DeepLab V3 [17]	×	15.3	93.2	86.1	8.0	96.3	**87.0**	90.6	37.9	62.6	79.1	43.4	60.9	67.7	92.4
FC-DenseNet103 [11]	×	9.4	93.0	83.0	**37.8**	94.5	8.2	77.3	43.9	37.1	77.3	**59.6**	50.5	66.9	91.5
Additive FC-PRNets94 with basic PR-block(α=16)	×	39.9	89.2	78.2	22.0	84.4	79.1	87.7	53.7	64.8	68.1	33.2	68.6	66.3	88.4
Multiplicative FC-PRNets94 with basic PR-block (β=1.2)	×	21.4	87.4	80.8	25.9	84.1	68.3	82.6	39.1	47.4	69.2	38.0	74.5	63.4	89.7
Multiplicative FC-PRNets95 with bottlenneck PR-block (β=1.3)	×	20.2	84.0	78.5	25.8	81.0	66.5	82.3	35.9	46.3	67.0	35.1	72.2	61.3	83.6
SegNet [16]	√	29.5	87.0	68.7	16.0	86.2	60.5	52.0	13.4	17.9	58.5	25.3	24.8	46.4	62.5

TABLE 2. Average Results on "UM", "UMM" and "UU" Test Sets. Best Scores are Presented in Bold.

Methods	MaxF	AP	PRE	REC	FPR	FNR
StixelNet	89.12%	81.23%	85.80%	92.71%`	8.45%	7.29%
SPRAY	87.09%	91.12%	87.10%	87.08%	7.10%	12.92%
Up-Conv-Poly	93.83%	90.47%	94.00%	93.67%	3.29%	6.33%
RBNet	94.97%	91.49%	94.94%	95.01%	2.79%	4.99%
Additive FC-PRNets94 with basic PR-block(α=4)	**95.38%**	**91.86%**	**95.63%**	**95.17%**	**2.93%**	**4.86%**

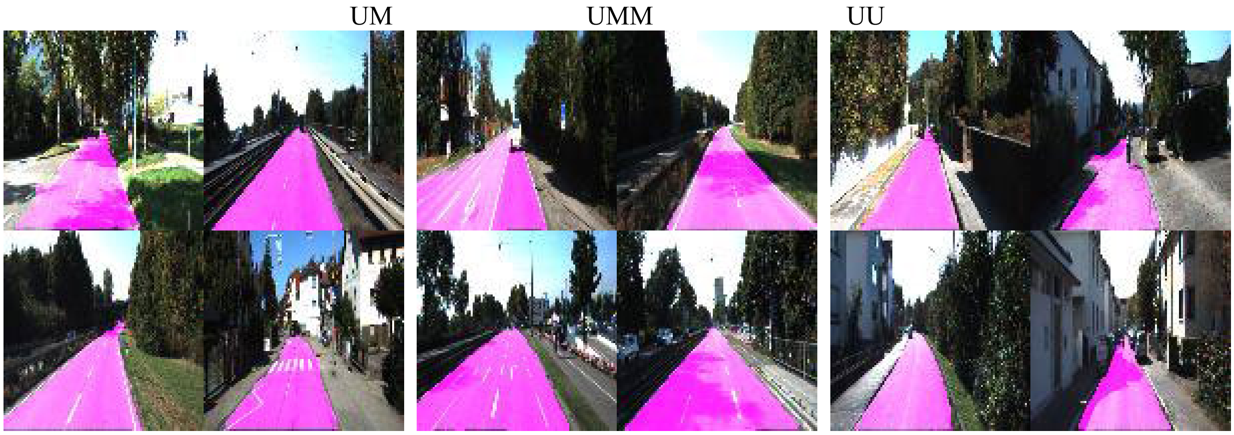


FIGURE 8. Segmentation results of FC-PRNet on KITTI road benchmark.

KITTI uses a wide variety of evaluation metrics to assess algorithm performance and also provides information captured by various sensors including visual cameras, LiDAR sensor, and GPS. KITTI estimation benchmark consists of 289 training and 290 test images, both containing three different road scene categories including urban unmarked roads (UU, 98/100), urban marked roads (UM, 95/96) and urban multiple marked lanes (UMM, 96/94). For evaluation, ground truth is provided for training images only and the number of submissions for online evaluation is limited. KITTI provides some established measures of Maximum F1-measure (MaxF) [20], Average precision as

TABLE 3. Four Structures of FC-PRNets.

Additive FC-PRNets with basic PR-block94(α)	
Down-sampling path	Input,m=1
	3×3 Conv m=48
	2 PR-block + TD m=48+2×α
	4 PR-block + TD m=48+6×α
	4 PR-block + TD m=48+10×α
	5 PR-block + TD m=48+15×α
	5 PR-block m=48+20×α
Up-sampling path	5 PR-block m=48+15×α
	5 PR-block + TU m=48+10×α
	4 PR-block + TU m=48+6×α
	4 PR-block + TU m=48+2×α
	2 PR-block + TU m=48
	3×3 Conv,m=c
	SoftMax

Additive FC-PRNets with bottlenneck PR-block95(α)	
Down-sampling path	Input,m=1
	3×3 Cov m=48
	2 PR-block + TD m=48+2×α
	3 PR-block + TD m=48+5×α
	4 PR-block + TD m=48+9×α
	5 PR-block m=48+14×α
Up-sampling path	5 PR-block m=48+9×α
	4 PR-block + TU m=48+5×α
	3 PR-block + TU m=48+2×α
	2 PR-block + TU m=48
	3×3 Conv m=c
	SoftMax

Multiplicative FC-PRNets with basic PR-block94(β)	
Down-sampling path	Input,m=1
	3×3 Conv m=48
	2 PR-block + TD $m=48\times\beta^{2}$
	4 PR-block + TD $m=48\times\beta^{6}$
	4 PR-block + TD $m=48\times\beta^{10}$
	5 PR-block + TD $m=48\times\beta^{15}$
	5 PR-block $m=48\times\beta^{20}$
Up-sampling path	5 PR-block $m=48\times\beta^{15}$
	5 PR-block + TU $m=48\times\beta^{10}$
	4 PR-block + TU $m=\times\beta^{6}$
	4 PR-block + TU $m=\times\beta^{2}$
	2 PR-block + TU m=48
	3×3 Conv,m=c
	SoftMax

Additive FC-PRNets with bottlenneck PR-block95(β)	
Down-sampling path	Input,m=1
	3×3 Conv m=48
	2 PR-block + TD $m=48\times\beta^{2}$
	3 PR-block + TD $m=48\times\beta^{5}$
	4 PR-block + TD $m=48\times\beta^{9}$
	4 PR-block $m=48\times\beta^{13}$
Up-sampling path	4 PR-block $m=48\times\beta^{9}$
	4 PR-block + TU $m=48\times\beta^{5}$
	3 PR-block + TU $m=48\times\beta^{2}$
	2 PR-block + TU m=48
	3×3 Conv,m=c
	SoftMax

used in PASCAL VOC challenges (AP), Precision (PRE), Recall (REC), False Positive Rate (FPR) and False Negative Rate (FNR).

We compare the performance of FC-PRNet with other state-of-the-art methods on the KITTI road benchmark. The compared algorithms include StixelNet [21], SPRAY [22], Up_Conv_Poly [23] and RBNet [24]. Table 2 shows the results of different algorithms on the evaluation. It is worth that KITTI road benchmark includes LiDAR data, while our algorithm only uses image data from visual cameras. Therefore, the comparison experiment only contains algorithms that use the same image data. It can be seen that in all these metrics the FC-PRNET algorithm outperformed its competitors, neral road detection. Some qualitative results of FC-PRNet algorithm are shown in Figure 8.

IV. DISCUSSION AND CONCLUSION

This article focuses on segmenting objects at scene parsing. By introducing pyramid residual blocks, FC-PRNet can avoid the diminishing feature reuse problem. The dimensions are forced to grow gradually in order to reduce the parameters. To analyze the effectiveness of the proposed algorithm, well-known datasets of CamVid and KITTI were tested. FC-PRNet achieves good semantic recognition results for segmenting objects at scene parsing without additional post-processing and pre-training.

At present, FC-PRNet just processes one piece of the color image. However, the datasets contain information about time series and LiDAR, which are important to improve the results of scene parsing. In the follow-up work, we managed to incorporate information about time series and LiDAR in the training process to obtain better results. Meanwhile, we will design optimal models with more layers to improve segmentation performance.

In conclusion, we study a new semantic segmentation network named Fully Convolutional Pyramidal Residual Networks (FC-PRNet). By designing pyramid residual blocks and sampling modules in down/up-sampling paths, the network achieves excellent capability of semantic recognition with few parameters. In the CamVid dataset, FC-PRNets obtained 75.4% of MIoU and 93.6% of Pacc, higher than SegNet, DeepLab V3 and FC-Densenet. In KITTI road benchmark, FC-PRNets obtained 95.38% of MaxF, 91.86% of AP, 95.63% of PRE, 95.17% of REC, 2.93% of FPR and 4.86% of FNR. FC-PRNets made better performance than StixelNet, SPRAY, Up_Conv_Poly and RBNet.

APPENDIX

See Table 3.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[2] E. Shelhamer, J. Long, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun, ''Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[4] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[5] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, ''The importance of skip connections in biomedical image segmentation,'' in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 179–187.

[6] S. Zagoruyko and N. Komodakis, ''Wide residual networks,'' 2016, *arXiv:1605.07146*. [Online]. Available: https://arxiv.org/abs/1605.07146

[7] G. Huang and Z. K. Liu Weinberger, ''Densely connected convolutional networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 4700–4708.

[8] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, ''Aggregated residual transformations for deep neural networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5987–5995.

[9] D. Han, J. Kim, and J. Kim, ''Deep pyramidal residual networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 5927–5935.

[10] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[11] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, ''The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 1175–1183.

[12] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2016, pp. 1440–1448.

[13] J. Schmidhuber, ''Deep learning in neural networks: An overview,'' *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[14] K. He, X. Zhang, S. Ren, and J. Sun, ''Identity mappings in deep residual networks,'' 2016, *arXiv:1603.05027*. [Online]. Available: http://arxiv.org/abs/1603.05027

[15] L. Castrejon, Y. Aytar, C. Vondrick, H. Pirsiavash, and A. Torralba, ''Learning aligned cross-modal representations from weakly aligned data,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2940–2949.

[16] V. Badrinarayanan and A. R. S. Kendall Cipolla, ''SegNet: A deep convolutional encoder-decoder architecture for image segmentation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking atrous convolution for semantic image segmentation,'' 2017, *arXiv:1706.05587*. [Online]. Available: https://arxiv.org/abs/1706.05587

[18] A. Geiger, P. Lenz, and R. Urtasun, ''Are we ready for autonomous driving? The KITTI vision benchmark suite,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[19] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, ''Vision meets robotics: The KITTI dataset,'' *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.

[20] J. Fritsch, T. Kuhnl, and A. Geiger, ''A new performance measure and evaluation benchmark for road detection algorithms,'' in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1693–1700.

[21] D. Levi and N. E. Garnett Fetaya, ''StixelNet: A deep convolutional network for obstacle detection and road segmentation,'' in *Proc. BMVC*, 2015, pp. 109.1–109.12.

[22] T. Kuhnl and F. J. Kummert Fritsch, ''Spatial ray features for real-time ego-lane extraction,'' in *Proc. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 288–293.

[23] G. L. Oliveira, W. Burgard, and T. Brox, ''Efficient deep models for monocular road segmentation,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4885–4891.

[24] C. Zhe and Z. Chen, ''Rbnet: A deep neural network for unified road and road boundary detection,'' in *Proc. Int. Conf. Neural Inf. Process.*, Nov. 2017, pp. 677–687.

FENGXIAO LI received the B.E. degree from the Chongqing University of Technology, China, in 2018, where he is currently pursuing the master's degree in communication and information systems. His research interest includes the application of deep learning in water quality spectrum processing.

ZOURONG LONG received the B.S. degree from the Shaanxi University of Science & Technology, Shannxi, China, in 2013, and the Ph.D. degree in optical engineering from Chongqing University, Chongqing, China, in 2020. He is currently a Professional Teacher with the Department of Optoelectronics Engineering, Chongqing University of Technology. His research interest includes the application of deep learning in image processing.

PENG HE received the B.S. degree from Nanchang Hangkong University, Nanchang, Jiangxi, China, in 2007, and the Ph.D. degree in optical engineering from Chongqing University, Chongqing, China, in 2013. He is currently pursuing the joint Ph.D. degree in biomedical engineering with the Virginia Polytechnic Institute and State University. From July 2013 to June 2015, he held a postdoctoral position with Chongqing University, where he was a Lecturer of instrument science and technology from July 2015 to December 2016 and is currently an Associate Professor with the Department of Optoelectronics Engineering. His research interests include X-ray spectral CT imaging, digital image processing, and big data artificial intelligence. He has published more than 30 papers in his research areas.

PENG FENG received the B.S. degree in mechanical and electronics engineering and the Ph.D. degree in optics engineering from Chongqing University, in 2002 and 2007, respectively, which is one of the most prestigious universities in China. From June 2008 to January 2012, he was a Postdoctoral Fellow with the School of Biomedical Engineering, Chongqing University, where he is currently an Associate Professor with the Department of Optoelectronics Engineering. His research interests include computed tomography, compressive sensing, image representation, and biomedical image processing. He has published more than 30 peer-reviewed journal articles.

XIAODONG GUO received the B.S. degree from the University of Electronic Science and Technology of China, Sichuan, China, in 2014, and the Ph.D. degree from Southwest Jiaotong University, in 2018. He is currently pursuing the Ph.D. degree in optical engineering from Chongqing University, Chongqing, China, under the supervision of Prof. Xiaohua Lei. His research interests include the application of deep learning in medical CT images and image processing.

XUEZHI REN received the B.S. degree in applied physic from the China University of Petroleum (East China), Qingdao, China, in 2017. He is currently pursuing the M.Sc. degree in instrumentation science and technology with Chongqing University, Chongqing, China, under the supervision of Prof. P. He. His research interests include the application of machine learning in image processing and multi-material decomposition of spectral CT.

BIAO WEI received the B.S. degree from the College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, Sichuan, China, in 1988, and the Ph.D. degree from the College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, in 1996. From October 1996 to May 1999, he held a postdoctoral position with Chongqing University, Chongqing, China, where he was a Lecturer from June 1999 to November 2004 and currently a Professor and the Vice Dean of the College of Optoelectronics Engineering. His research interests include optoelectronic imaging and light energy detection of X-rays, neutrons and visible light, high-resolution optoelectronic imaging detection (sensing) technology with scientific-grade CCD, EMCCD, and ICCD digital camera, and optoelectronics image compression coding technology. He has published more than 60 articles in his research areas.

MINGFU ZHAO is professor, doctor of engineering, doctoral supervisor. Vice Dean of the School of Electrical and Electronic Engineering, Chongqing University of Technology. His current research interests include optical fiber biochemical sensing theory and application, intelligent optical fiber sensing theory and technology, bionic multi-sensing fusion technology, information acquisition and processing, and intelligent information processing.

BIN TANG is mainly engaged in the research work of spectroscopic water quality detection, pesticide residue analysis, environmental spectroscopy analysis, and digital signal processing. He is also mainly responsible for the design of water quality parameter detection schemes by spectroscopy, the establishment of suspended solids scattering models in water, the research of spectral information processing algorithms, and the design and implementation of key algorithm modules.