



# Real-Time Semantic Edge Segmentation Using Modified Channelwise Feature Pyramid

H. Harish<sup>1,2</sup> · A. Sreenivasa Murthy<sup>1</sup>

Received: 29 June 2023 / Accepted: 18 September 2023  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

## Abstract

In the forthcoming decades, real-time image processing will play a crucial role in computer vision. The rise in population has resulted in a higher usage of smart devices in various industries, including the automobile, medical, and consumer electronics sectors. Therefore, it is imperative for researchers to play a significant role in enhancing the performance, processing speed, and optimizing the model quantity. In this particular study, a modified channelwise feature pyramid (M-CFPNet) model is developed for real-time semantic edge segmentation to balance the aforementioned factors. The proposed system is implemented to extract valuable features of the edges in the image. Using the CFPNet module, M-CFPNet is constructed and evaluated on various BSD, CamVid, and Cityspace databases. The Cityscapes database achieves a classwise mIoU of 73.3% with 0.55 million parameters. The execution speed has reached 30 FPS on RTX GPU for 1024 × 2048 dimension images. When compared to CFPNet the M-CFPNet algorithm improves 3.2% of mIoU and 0.04% parameter.

**Keywords** Semantic edge segmentation · CFPNet · M-CFPNet · BSD · Cityspace · CamVid · Image edge segmentation

## Introduction

Real-time semantic edge segmentation is the most crucial topic in image processing and computer vision. There are several applications for this system or model, such as bio-medical image processing, autonomous self-driving system, and many more. All these applications have critical use cases and need more accuracy. Increasing the network quantity has reduced the inference, which may lag in real-time operating conditions. To mention some examples, such as DeepLab and PSPNet's are best in performance with semantic edge segmentation; however, they consist of millions of limiting factors and a minimal processing speed of one frame per second (FPS). In general, 30FPS speed is maintained typically for real-time processing.

To conclude, the larger the network lesser the processing speed. Along with processing speed memory, consumption is also considered the primary parameter. Building an optimistic and efficient semantic edge segmentation network with less memory and more processing speed is achieved in this study.

Current semantic segmentation models are high speed, mainly ESPNET [5] and ENET [4], which require high costs for execution with high inference speeds. Slightly modified models such as ICNET [6] and ContexNet [5] have achieved good results, but their models do not achieve with best size and speed. Hence, this study concentrated on maintaining or obtaining better inference, accuracy, and model size. The very old manuscripts [7, 8] had delivered the process of achieving the multi-scale convolution with various sizes and various fields. This phenomenon allows the models to process with multilevel feature extraction and contains multiple data scales. The dilated convolution was one of the best processing models to extract the large scale features within the maximum available number of features [10, 11]. Both the networks have certain limitations, whereas a channel wise model and the inception models contain more parameters, though they contain factorizations. The dilation convolution may miss local features but achieves the best global information with a single dilation rate. Fixed dilation makes the

---

This article is part of the topical collection “SWOT to AI-embraced Communication Systems (SWOT-AI)” guest edited by Somnath Mukhopadhyay, Debashis De, Sunita Sarkar and Celia Shahnaz.

---

✉ H. Harish  
hh.harish@gmail.com

<sup>1</sup> UVCE, Bangalore University, Bengaluru, India

<sup>2</sup> Maharani Lakshmi Ammanni College for Women, Bengaluru, Karnataka, India

model extract the large class of features in the cityscapes data set but avoids identification of minute ones.

The proposed methodology is a tiny and new CNN module that achieves the pros of dialed and inception convolution. This prone module is applied to construct superficial and practical encoder- and decoder-based models to pull out the dense features.

The modified channel wise feature pyramid network design (M-CFPNet) depends on the CFPNet model. Many new parameters have better results than the current semantic segmentation models. The proposed module can efficiently incorporate dilated and inception convolutions. Hence, they are called channelwise feature pyramid (CFP) networks. This framework can pull out numerous contingent information jointly and size feature maps and significantly reduce the size and number of parameters.

## Literature Survey

Recent works mainly specify semantic edge segmentation process that demonstrates either factorization, dilated convolution or low bit networks, or a mixture of these techniques to optimize the model size and speed of the CNN. The primary step is briefly describing the technique opted for and overviewing the decoder–encoder-based edge semantic segmentation.

The dilated convolution [17] develops a  $3 \times 3$  convolution, which is a standard and unique form by filling gaps between enhanced effective receptive field and convolution parameters without producing more elements. The rate of dilation  $r$  is denoted as  $[r(n-1)+1]^2$  for an  $n \times n$  dilated convolution in the kernel, where  $r$  represents pixel gap numbers between adjacent convolution element and  $n^2$  elements participate in module training. Several research have already experimented on dilated convolution to pick off multi scale features, which is demonstrated and developing a spatial feature pyramid, to name a few DenseASAPP [18, 19] and DeepLab [10, 12, 13]. The application patterns shows the stamina in pixel-level tasks. This study tried to implement dilated convolution for each CFP network channel.

The naïve Inspection architecture [7] demonstrated a jointing model which consists  $1 \times 1$ ,  $3 \times 3$  and  $5 \times 5$  convolutions which achieves multi scale feature maps of kernel. The large convoluted kernels, leads to more processing cost.

Hence, current versions of inception architecture initiate factorization convolutions to decrease the number of elements. This factorization has two parts: asymmetric convolution and smaller convolution by factorization. To state an example, where 34% of elements are saved with the same filter size is achieved, i.e., for  $5 \times 5$  convolution operator, it is replaced with  $3 \times 3$  convolutions and then factorized a convolution into  $3 \times 1$  convolution from  $1 \times 3$  convolutions.

The TesNext [18], MobileNets [14] and inception [16] are the two factorization modules that had been applied successfully for great stamina in decreasing the processing of the CNN models. The factorization module inspired in development of CFPNet. Each CNN channel is used with a small convolution approach to reduce the inception-like model in CFPNet. The asymmetric convolution technique reduced channel parameters. Factorization decreases the execution substantially, where the module is allowed to learn from the features of respective fields with a series of sizes [33–35] segmented image using watershed and particle swarm optimization techniques to obtain good results.

The encoder and decoder are two different parts of the encoder–decoder network. The encoder consists of down-sampled operators for extracting the high dimension features and sequence of convolution. The exact process is reverted for a decoder, like instead of down sampling, the upsampling are convoluted to create masks. Few of encoder–decoder-based designs available are, U\_Net [19], FCN [16], and Seg-Net [21] that demonstrated great results with pixel-level edge segmentation. The entire process in the study discusses, the architecture of MCFPNet, which is derived from CFPNet.

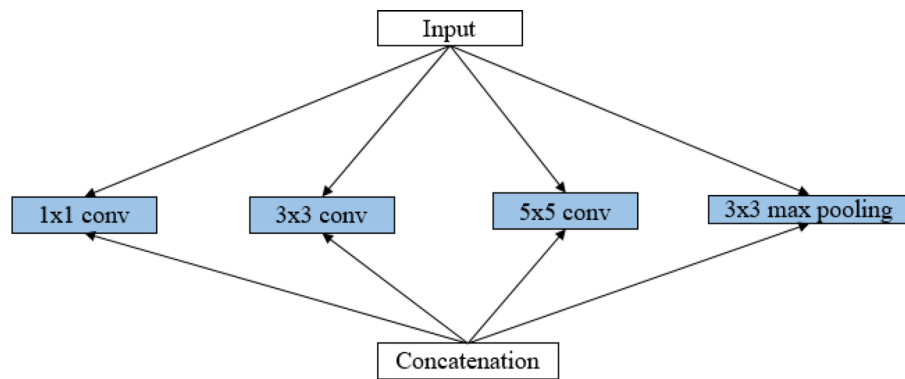
## Methodology

### Channelwise Feature Pyramid Channel

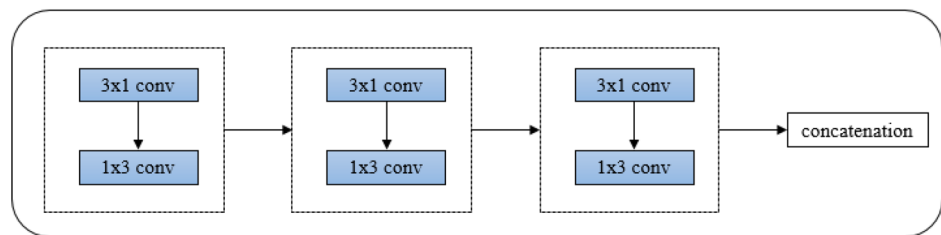
CFPNet derived from CFP module, which traces convolution network operation that decays a larger kernel into several minimal convolutions in Fig. 1a, b. The modified CFPNet has better performance in results than CFPNet. The traditional Inception model uses directly  $5 \times 5$  size kernel, and Inception-V2 as shown in Fig. 1c is deployed with dual  $3 \times 3$  convolution operators instead. Depending on the thought of factorization and multi-scale feature maps, the current system is designed with a  $7 \times 7$  kernel. Same way, Inception-V2 introduced three  $3 \times 3$  convolution kernels instead of  $5 \times 5$  and  $7 \times 7$  size kernels. Due to this functioning, 45% and 28% of parameter elements are saved, respectively. It is hard to achieve a real-time goals hence, combined both convoluted kernels into single channels with only  $3 \times 3$  sized kernels. Then, perish the convention convolution to unsymmetrical conventions to construct the feature pyramid channels. To generate a multi scale feature map the connection is skipped to concentrate parameters that are pulled out from an unsymmetrical convolution set. The feature pyramid is reduced to 69% of parameters when compared with CFPNet and Inception-v2. In addition, FP channel saves 69% of elements but has the capability to understand the attribute data and retains the original dimension.

Because of uniting features from asymmetric or skewed convolution blocks, it maintains the equal dimensions of

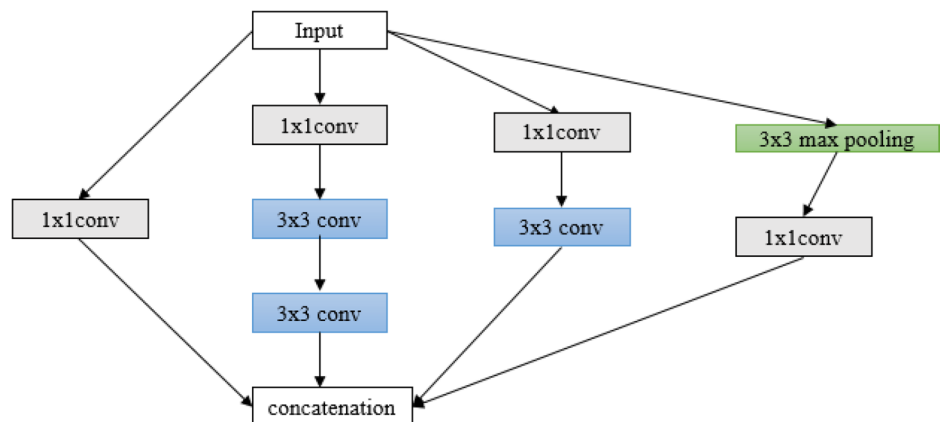
**Fig. 1** **a** Naive inception module. **b** Feature pyramid module. **c** Inception V2-module



(a) Naive Inception module



(b) Feature pyramid module



(c) Inception V2- module

output and inputs by reshuffling filter numbering per unsymmetrical set. If the input size is ‘ $N$ ’, then  $N/4$  is assigned as primary and secondary sets, which point to  $5 \times 5$  and  $3 \times 3$  convolution. To the tertiary set, the  $7 \times 7$  kernel is allocated with  $N/2$  filter, which pulls out a substantial symmetrical size feature.

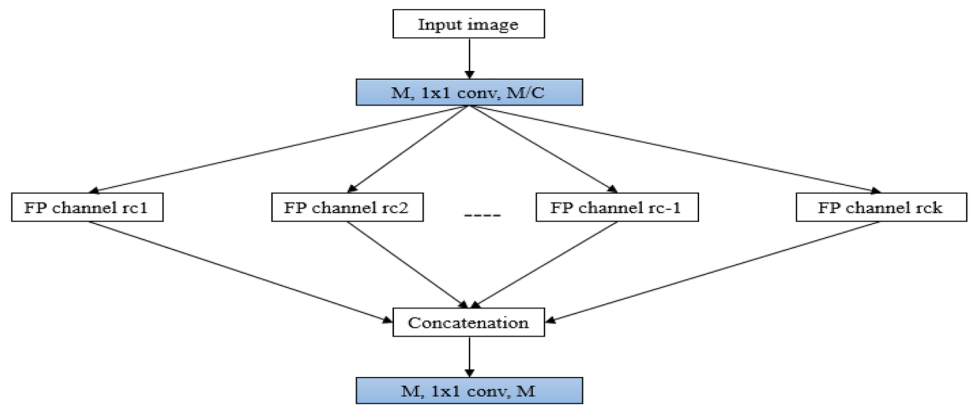
**Channelwise Feature Pyramid Module**

CFP module has  $L$  FP channels with multiple dilation rates  $\{r_1, r_2, \dots, L\}$ . Conventional CFP is primarily applied with  $1 \times 1$  complication to decrease the input facet from  $m$  to  $m/L$ .

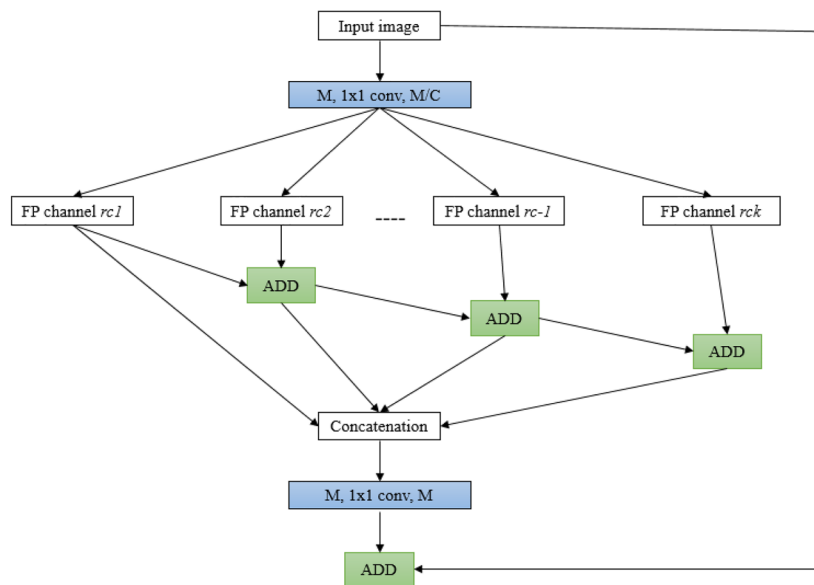
Later the dimension of the primary and tertiary irregular set are  $m/4L$ ,  $m/4L$ , and  $m/2L$ , respectively.

Figure 2 represents detailed information about the CFP module. There,  $1 \times 1$  convolution achieves high dimension feature maps with lesser measurement. Later, set multi FP channels to parallel arrangement along with multiple dimension values. Then, all feature maps are united into single dimension input and apply convolution of  $1 \times 1$  to trigger the output. This is the fundamental architecture of any conventional CFP module which is demonstrated in Fig. 2a. The enhancement in the depth of the module network is done using unsymmetrical convolution, which

**Fig. 2** **a** Conventional CFP network. **b** CFP network module



(a) Conventional CFP network



(b) CFP Network module

leads to harder for learning or training. In addition, an ordinary combination method initiated some griddling disturbance or unwanted checkboard that has more impact on accuracy and excellence in edge identification masks. To improve the struggle in the training, the remaining connection is made as the primary step as the deep network module is trainable and gives additional feature data [21]. To avoid the impact of griddling disturbance, need to apply HFF (hierarchical feature fusion) [5] to de-griddling.

From initiating the secondary channel, apply the addition operation to sum the feature maps stage by stage, later uniting the constructed final hierarchical feature map. The griddling disturbances are finally reduced. The final CFP module with less griddle disturbance is represented in Fig. 2b.

### MCFPNet Module

Primarily, mentioning the CFP module details which are used to construct MCFPNet. Selecting FP channels to  $C = 4$ . The dimension of input is  $D = 32$ , filter number with 8 as the channel size. opting filter number of primary and tertiary asymmetric convolution sets are 3,3, and 4, respectively, but in CFPNet [29] it is represented as 2,2 and 4, respectively. Later, the various dilation rates is set to individual FP channel. Perform the dilation rate equal to  $rC$ , set the 1st and 4th channel dilation rate to  $r1 = 1$  and  $rC$ , and want the MCFP module can pull global and local features. The secondary and tertiary channels are set with rates of dilation equal to  $r2 = rC/4$  and  $r3 = rC/2$ . Hence, the CFPNet module could train for

midsized features. If  $rC/4$  is lesser than 1 and if  $rC = 2$ , then unswervingly setting the channel to amplify rate equal to 1.

**MCFPNet architecture:** Though Agenda of the manuscript is to develop a lightweight module with the best performance. Hence, a shallow network is proposed in the manuscript, as shown in Fig. 3. The detailed architecture is represented in Table 1. Initially, three  $3 \times 3$  convolution is performed on the feature extractor and apply the down-sampled method as in ENet [4], which fuses a  $3 \times 3$  convolution with a stride two and  $2 \times 2$  maximum pollings. These down-sampling operation process outputs by three times, and output dimensions are 18<sup>th</sup> of the original input size. Skip the connection to insert and resize the input images before the first and second max polling final  $1 \times 1$  convolution, providing additional data for the segmentation network. In the CFP-2 and CFP-1 clusters, the CFP module is repeated with  $n = 2$  and  $m = 6$  times with rate of dilation  $rKCFP - 1 = [3, 3]$  and  $rKCFP - 2 = [4, 4, 8, 8, 16, 16]$ . As a last step, a  $1 \times 1$  convolution is applied to trigger the output feature map to obtain final edge segmentation masks using bilinear interpolation. Each convolution is performed by triggering PteLU [23] batch normalization. The study proved to attain better results in performance than TeLU in shallow networks. The CFP-2, m output if fed to up sampling with  $1 \times 3$  conv instead of  $3 \times 3$  conv has improved the results in MCPF net. As CFP-2, m and CFP-2, n has max co-ordinate values which may not need of  $3 \times 3$  conv [32].

The proposed neural network has been tested on BSDS500 and also on CamVid and cityscape data sets. These data sets are widely used in semantic edge segmentation. The whole work was validated on CamVid and cityscape data sets, with some selected images from BSDS500. The parameters such as repeat times, the number of channels, and dilations have been experimented. Finally, the networks are compared between the data sets.

**Table 1** Design details of MCFPNet

No	Layer	Mode	Dimension
1	$3 \times 3$ Conv	Stride 2	32
2	$3 \times 3$ Conv	Stride 1	32
3	$3 \times 3$ Conv	Stride 1	32
4	Downsampling	–	32
5–6	$2 \times$ CFP	$rC = 2$	32
7	Downsampling	–	128
8–9	$2 \times$ CFP	$rC = 4$	128
10–11	$2 \times$ CFP	$rC = 8$	128
12–13	$2 \times$ CFP	$rC = 16$	128
14	$1 \times 1$ Conv	Stride 1	18
15	Bilinear interpolation	$\times 8$	18

## Results and Discussion

### Data Set

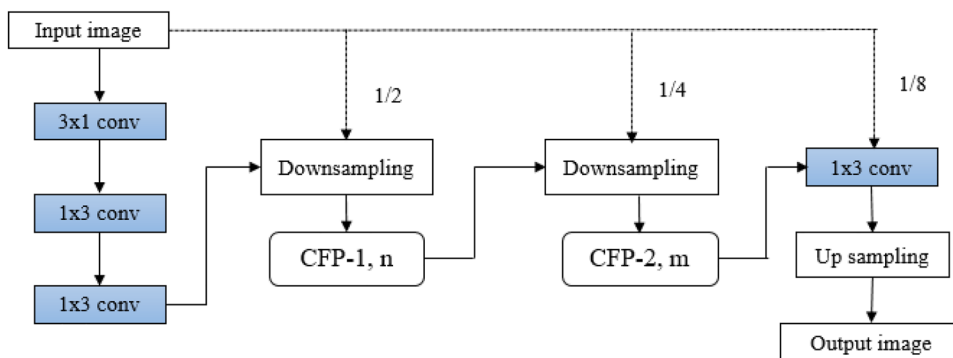
**Cityscapes** The cityscapes consist of 5000 fine annotations and 20,000 coarser annotation images. In addition, the data set has data from 50 different cities. The original input image resolution is  $1024 \times 2048$ . The data set consists of seven categories: cars, trucks, buses, and other vehicles.

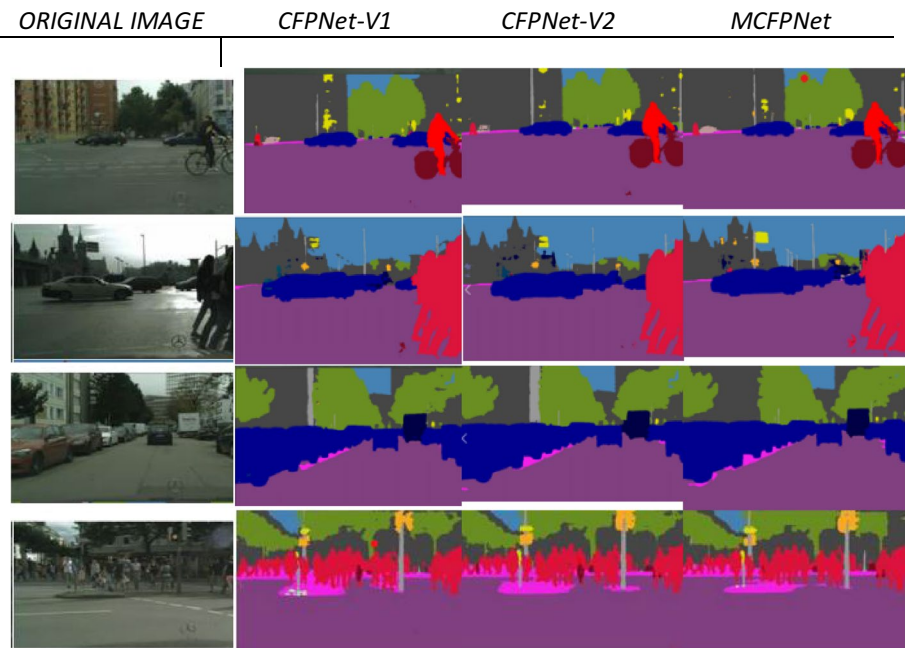
**CamVid** This data set includes an urban scene data set which can be used in automotive applications, such as self-driving. It has 701 images, with each image resolution of  $720 \times 790$ . 234 for training and 101 for validation are used. In the proposed work, the images are resized to  $360 \times 480$  before training the data set.

### Analysing MCFPNet Architecture

The proposed multiple CFPNet consist of repeat times, different channel numbers, and rate of dilation, to analyze their performance of CamVid test data set. MCFPNet is represented in Table 2. The multiple features of the edge

**Fig. 3** Proposed MCFPNet architecture



**Table 2** Edge segmented examples from cityscapes**Table 3** Evaluation of MCFPNet on CamVid

Network	Parameter	Size	mIoU (%)
CFPNet-V1	0.31 M	1.34 MB	61.6
CFPNet-V2	0.37 M	1.6 MB	62.9
CFPNet-V3	0.55 M	2.5 MB	64.3
MCFPNet	0.55 M	2.5 MB	65.94

**Table 4** Evaluation of MCFPNet on Cityscapes

Network	Parameter	Size	mIoU (%)
CFPNet-V1	0.31 M	1.34 MB	60.6
CFPNet-V2	0.37 M	1.6 MB	66.9
CFPNet-V3	0.55 M	2.5 MB	70.1
MCFPNet	0.55 M	2.5 MB	73.30

segmented images is extracted using region of interest (ROI) and classifiers.

As the study do not use the pre-trained models, the maximum epochs used is 1024. The ADAM [24] is used to train the network with a momentum 0.9 and weight decay  $4.5e-4$ . Applying the “poly” learning rate policy [25] and the initial learning rate is set with power 0.9. Later opting a multiple batch number for two data sets, eight for Cityscapes and sixteen for CamVid represented in Table 3. Data augmentation is also performed to create diversity in training. The measurement rates are as  $\{0.5, 0.75, 1.0, 1.25, 1.5, 1.75\}$ .

*CFPNet-V1* This is the superficial version, because it repeats  $n$  times, and  $m$  are 1 and 2. For primary MCFPNet, the dilation rates are set to  $rCCFP-1 = [4]$  and  $rCCFP-2 = [8, 16]$ .

*CFPNet-V2* When Compared to MCFPNet-V1, the network can be able to extract more local features. Hence, modified the repeat time from  $\{n,m\} = \{1,2\}$  to  $\{n,m\} = \{1,3\}$ , and their corresponding distend rates are changed to  $rCCFP-1 = [2]$  and  $rCCFP-2 = [4, 8, 16]$ .

*MCFPNet* The operation of the network has increased and controlled the model size. The continuity of the CPFNet-V2

**Table 5** Cityscapes data set results

Network	Class	Category
ENet [4]	58.3	80.4
ESPNet [5]	60.3	82.2
ETFNet [26]	68.0	86.5
SQNet [28]	59.8	84.3
SegNet [19]	57.0	79.1
FCN-8 s [20]	65.3	85.7
DeepLab-v2 [3]	70.4	86.4
PSPNet [2]	78.4	86.4
CFPNet [32]	70.1	87.4
MCFPNet	73.3	87.9

is doubled. The distend rates are modified per cluster to  $rCFP-1 = [2, 2]$  and  $rCCFP-2 = [4, 4, 8, 8, 16, 16]$ .

Table 4 represents the evaluation results of MCFPNet on Cityscapes which improves the results with 3.2% more accurate with same size of CFPNet-V3.

Although the dimensions of MCFPNet are diminutive, its mIoU accuracy is impressively competitive, as demonstrated in Table 5, both in terms of classwise and



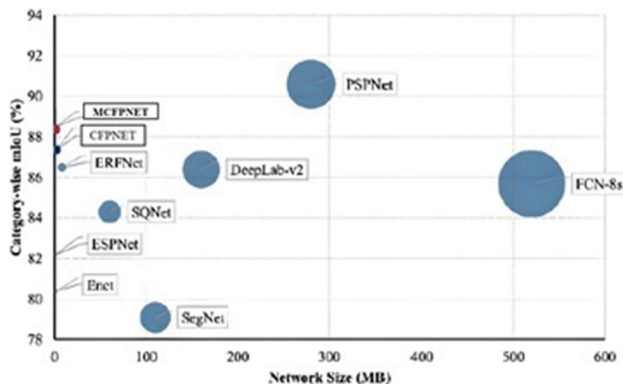


Fig. 4 Network size and classwise mIoU accuracy

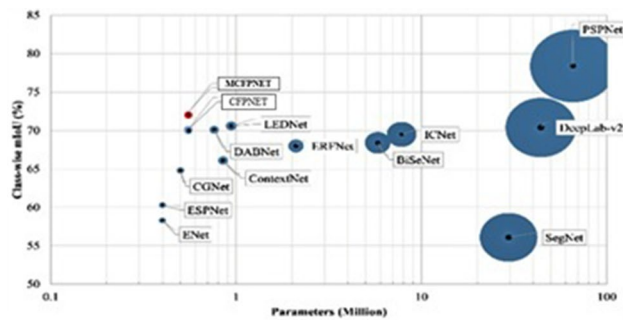


Fig. 5 Classwise mIoU and parameters

categorywise evaluations. A closer examination of the results reveals that MCFPNet exhibits a higher level of sensitivity and precision, particularly in the detection of small and low-frequency classes, such as traffic lights.

### Comparisons

The CamVid and Cityscapes, test data set results are compared with the proposed system to existing convention systems. Figure 4 represents the relationship between classwise mIoU accuracy and network size.

In Fig. 4, the blue circle represents the size of the model, i.e., the smaller circle, the smaller the model size. The size of the MCFPNet is small, as mIoU has very competitive accuracy, as represented in Table 5. CFPNet-V3 reported more sensitivity and accuracy for tiny and less frequency classes.

Figure 5 represents the test results plotted accuracy versus parameters for Cityscapes. The proposed MCFPNet has an excellent accuracy which achieves 71.0% and the best accuracy compared to LEDNet [29], ESPNet, and CGNet [27]. It also throws better performance than the existing CFPNet [32] and gives better training and segmentation results.

The different existing protocols are executed in various GPUs and their comparative results are placed in Table 6. The MCFPNet has a very similar processing speed as CFPNet [32], DABNet, and ICNet for the same input of  $1024 \times 2048$ . As CFPNet [32] saved 28.6%, but MCFPNet saved 28.9%, which is closer but improved in network performance. As the CamVid data set is tested, state-of-the-art is placed in Table 7. MCFPNet achieves slightly better results with a small network. When compared to CPFNet the M-CFPNet improves 3.2% of mIoU and 0.04% parameter.

Given the variability in input size and GPU specifications across networks, both of these factors are reported in Table 6. In terms of computational ability, the hierarchy of GPUs is as follows: Titan Maxwell < Titan X Pascal  $\approx$  GT 1080Ti < Titan p < TT 2080Ti < V100. Despite

Table 6 Evaluation results for Cityscapes data set for testing

Network	Pretrain	I/P size	FPS	MIoU%	GPU	Parameters
DeepLab-v2 [3]	ImageNet	$512 \times 1024$	< 1	70.4	44	Titan
PSPNet [2]	ImageNet	$713 \times 713$	< 1	78.4	65.7	Titan
SegNet [19]	ImageNet	$360 \times 640$	14.6	56.1	29.5	Titan
ENet [4]	None	$512 \times 1024$	76.9	58.3	0.4	Titan
SQ [28]	ImageNet	$1024 \times 2048$	16.7	59.8	–	Titan
ESPNet [5]	None	$512 \times 1024$	112	60.3	0.4	Titan-P
ContextNet [5]	None	$1024 \times 2048$	18.3	66.1	0.85	Titan
ETFNet [26]	None	$512 \times 1024$	41.7	68.0	2.1	Titan
BiSeNet [30]	ImageNet	$768 \times 1536$	105.8	68.4	5.8	Titan p
ICNet [6]	ImageNet	$1024 \times 2048$	30.3	69.5	7.8	Titan
CGNet [27]	None	$360 \times 640$	50	64.8	0.5	$2 \times$ V100
LEDNet [29]	None	$512 \times 1024$	71	70.6	0.94	1080Ti
DABNet [28]	None	$1024 \times 2048$	27.7	70.1	0.76	1080Ti
CFPNet [32]	None	$1024 \times 2048$	30	70.1	0.55	208Ti
MCFPNet	None	$1024 \times 2048$	30	73.4	0.59	AMD

**Table 7** CamVid testing set performance

Network	mIoU (%)	Parameter
ENet [4]	51.3	0.36
SegNet [19]	55.6	29.5
FCN-8 s [20]	57.0	134.5
Dilation8 [31]	65.3	140.8
BiSeNET [30]	65.6	5.8
ESPNet [5]	55.6	0.36
CPFNet [32]	64.7	0.55
MCFPNet	65.1	0.59

the differences in input size and GPU devices, Fig. 7 is included to facilitate comparison of the results presented in Table 6.

In addition, we conduct performance assessments on the CamVid test data set and undertake comparisons with several other current techniques. Our findings, as outlined in Table 7, indicate that MCFPNet also delivers exceptional results despite its compact size. For instance, when comparing with ENet and ESPNet, it is evident that their minimal parameter count has a significant impact on their overall performance, as they rank the lowest in Table 7. In comparison with other methods boasting high performance, MCFPNet demonstrates a competitive level of accuracy despite having fewer than 3.4% of their parameters.

## Conclusion

In this paper, a small real-time semantic edge segmentation network is developed. MCFPNet or modified CFPNet is mainly deployed on the basis of the Feature Pyramid channel and modified version of CFPNet. The analysis and results of CamVid and cityscapes data sets are running on MCFPNet module. The MCFPNet is overall deployed to enhance the accuracy to 71% with the best in inference speed, parameters, and model size. To conclude, the overall module efficiently finds semantic edge segmentation. When compared to CFPNet the M-CFPNet improves 3.2% of mIoU and 0.04% parameter.

**Funding** The study is not a funded.

## Declarations

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Ethical Approval** This article does not contain any studies with human participants performed by any of the authors.

## References

- Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J. A survey on deep learning techniques for image and video semantic segmentation. *Appl Soft Comput.* 2018;70:41–65. <https://doi.org/10.1016/j.asoc.2018.05.018>.
- Zhao H, Shi J, Wang Q, Jia J. Pyramid scene parsing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017. pp 2881–2890. <https://doi.org/10.1109/CVPR.2017.660>
- Brostow GJ, Shotton J, Fauqueur J, Cipolla R. Segmentation and recognition using structure from motion point clouds. *ECCV.* 2008;1:44–57. [https://doi.org/10.1007/978-3-540-88682-2\\_5](https://doi.org/10.1007/978-3-540-88682-2_5).
- Paszke A, Chaurasia A, Kim S, Culurciello E. Enet: A deep neural network architecture for real-time semantic segmentation. *ariv preprint ar iv:1606.02147.* 2016. <https://doi.org/10.48550/arXiv.1606.02147>.
- 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).* 2018. pp 552–568. <https://doi.org/10.48550/arXiv.1803.06815>
- Poudel TP, Bonde U, Liwicki S, Zach C. Contextnet: Exploring context and detail for semantic segmentation in real-time. *ariv preprint ar iv:1805.04554.* 2018. <https://doi.org/10.48550/arXiv.1805.04554>
- Zhao H, Qi S, Shi J, Jia J. Icnets for real-time semantic segmentation on high-resolution images. In: *Proceedings of the European Conference on Computer Vision (ECCV).* 2018. pp. 405–420. <https://doi.org/10.48550/arXiv.1704.08545>
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015. pp 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016. pp 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017. pp 1492–1500. <https://doi.org/10.1109/CVPR.2017.634>
- 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).* 2018. pp. 801–818. <https://doi.org/10.48550/arXiv.1802.02611>
- Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation, *ariv preprint ariv:1706.05587.* 2017.
- Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell.* 2017;40(4):834–48. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson T, Franke U, Thoth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016. pp 3213–3223. <https://doi.org/10.1109/CVPR.2016.350>
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient



- convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017.
16. Tonneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. 2015. pp. 234–241, Springer. <https://doi.org/10.48550/arXiv.1505.04597>
  17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
  18. Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P. A realtime algorithm for signal analysis with the help of the wavelet transform. In: Wavelets. 1990. pp. 286–297, Springer. [https://doi.org/10.1007/978-3-642-75988-8\\_28](https://doi.org/10.1007/978-3-642-75988-8_28)
  19. Yang M, Yu K, Zhang C, Li Z, Yang K. Denseaspp for semantic segmentation in street scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 3684–3692. <https://doi.org/10.1109/CVPR.2018.00388>
  20. He K, Zhang, Ten S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
  21. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. <https://doi.org/10.1109/CVPR.2017.195>.
  22. Badrinarayanan V, Kendall A, Cipolla T. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017. <https://doi.org/10.1109/TPAMI.2016.2644615>.
  23. He K, Zhang, Ten S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. 2015. pp. 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>
  24. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. Automatic differentiation in pytorch. 2017.
  25. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
  26. You Y, Zhang Z, Hsieh C-J, Demmel J, Keutzer K. Imagenet training in minutes. In: Proceedings of the 47th international conference on parallel processing. 2018. pp. 1–10. <https://doi.org/10.48550/arXiv.1709.05011>
  27. Romera E, Alvarez JM, Bergasa LM, Arroyo R. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Trans Intell Transp Syst. 2017;19(1):263–72. <https://doi.org/10.1109/TITS.2017.2750080>.
  28. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, Cottrell G. Understanding convolution for semantic segmentation. In: IEEE winter conference on applications of computer vision (WACV). IEEE; 2018. pp. 1451–1460. <https://doi.org/10.1109/WACV.2018.00163>
  29. Wu T, Tang S, Zhang T, Zhang Y. Cgnet: A light-weight context guided network for semantic segmentation. 2018. arXiv preprint [arXiv:1811.08201](https://arxiv.org/abs/1811.08201)
  30. Li G, Yun I, Kim J, Kim J. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation. 2019. arXiv preprint [arXiv:1907.11357](https://arxiv.org/abs/1907.11357)
  31. Wang Y, Zhou Q, Liu J, Xiong J, Gao G, Wu X, Latecki LJ. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation. In: 2019 IEEE international conference on image processing (ICIP). 2019. <https://doi.org/10.1109/ICIP.2019.8803154>
  32. 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). 2018. <https://doi.org/10.48550/arXiv.1808.00897>
  33. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 2015. [arXiv:1511.07122](https://arxiv.org/abs/1511.07122)
  34. Lou A, Loew M. CFPNet: channel-wise feature pyramid for real-time semantic segmentation. Comput Sci Comput Vis Pattern Recognit. 2021. <https://doi.org/10.1109/ICIP42928.2021.9506485>.
  35. Harish H, Sreenivasa Murthy A. Identification of Lane Lines Using Advanced Machine Learning. In: 2022 8th international conference on advanced computing and communication systems (ICACCS). 2022. Vol. 1. <https://doi.org/10.1109/ICACCS54159.2022.9785221>.
  36. Harish H, Sreenivasa Murthy A. Identification of Lane Line Using PSO Segmentation. In: 2022 IEEE international conference on distributed computing and electrical circuits and electronics (ICDCECE). IEEE. 2022. <https://doi.org/10.1109/ICDCECE53908.2022.9793266>
  37. Harish H, Sreenivasa Murthy A. Edge Discerning Using Improved PSO and Canny Algorithm Communication, Network and Computing (CNC-2022) Part1, CCIS 1893. 2023. [https://doi.org/10.1007/978-3-031-43140-1\\_17](https://doi.org/10.1007/978-3-031-43140-1_17).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.