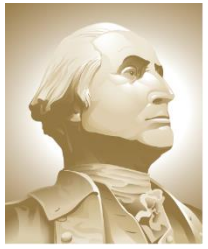


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

GW | SEAS

# CFPNet: Channel-wise Feature Pyramid for Real-Time Semantic Segmentation


Ange Lou, Murray H. Loew

The Department of Biomedical Engineering, The George Washington University, Washington DC



IEEE International Conference on Image Processing (ICIP)

September, 2021

# Background

- **Real-time** semantic segmentation is playing a more important role in computer vision, due to the growing demand for **mobile devices** and **autonomous driving**.
- The goal in our study is to design an algorithm to **balance** the **performance**, inference **speed** and **parameters**.
- Main contributions:

  (1) Proposed an efficient CNN module – Channel-wise Feature Pyramid (CFP) module;

  (2) Designed the CFPNet based on CFP module.

# Previous work

- In the DC-UNet, we proposed a dual-channel U-Net that achieve better performance in medical segmentation.
- DC-UNet only has **1/3** parameters compared with U-Net.

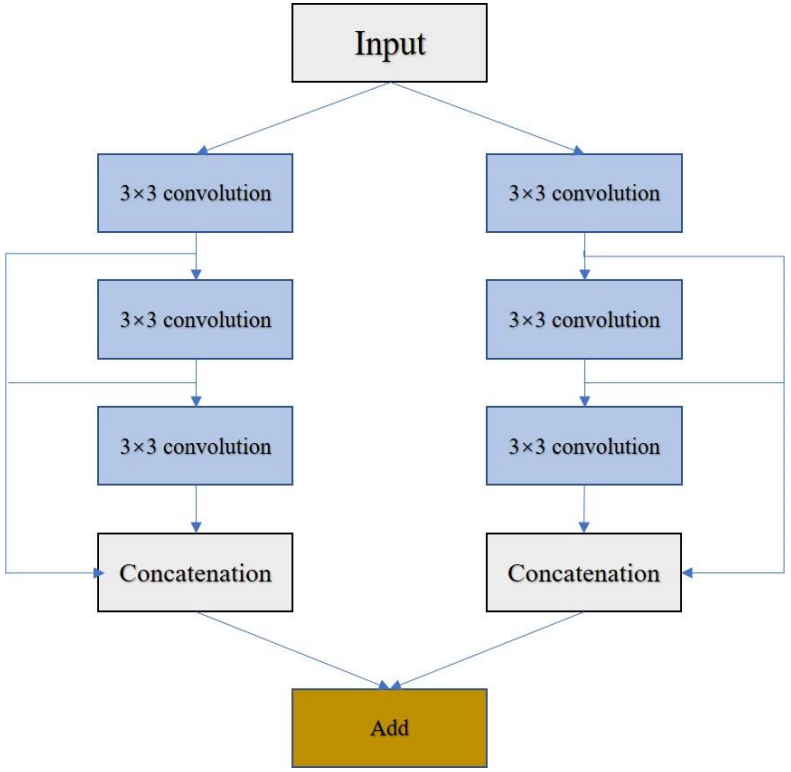



**Figure 1**. Dual-channel block.

# Method

- CFP module is based on Feature Pyramid (FP) channel.
- Each channel is designed by using asymmetric convolution.

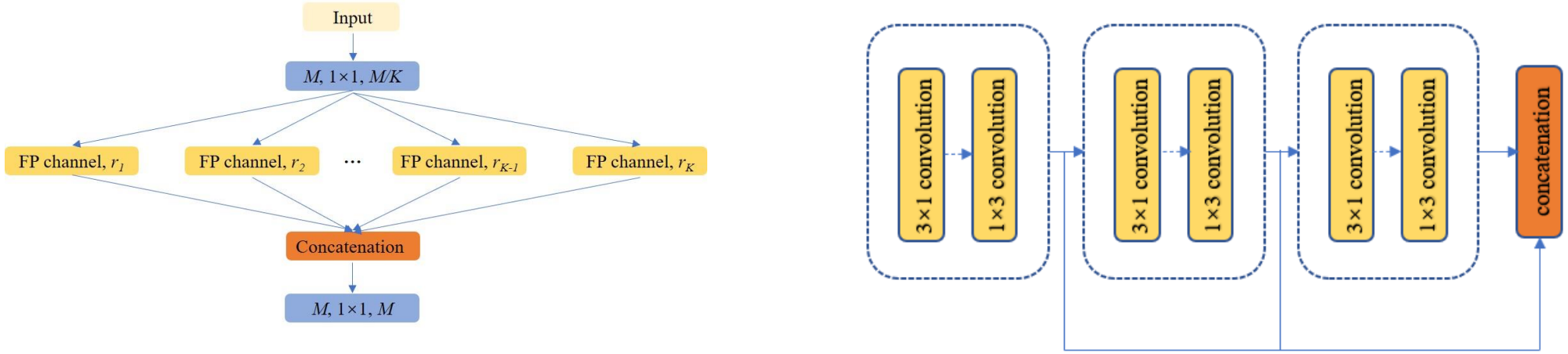
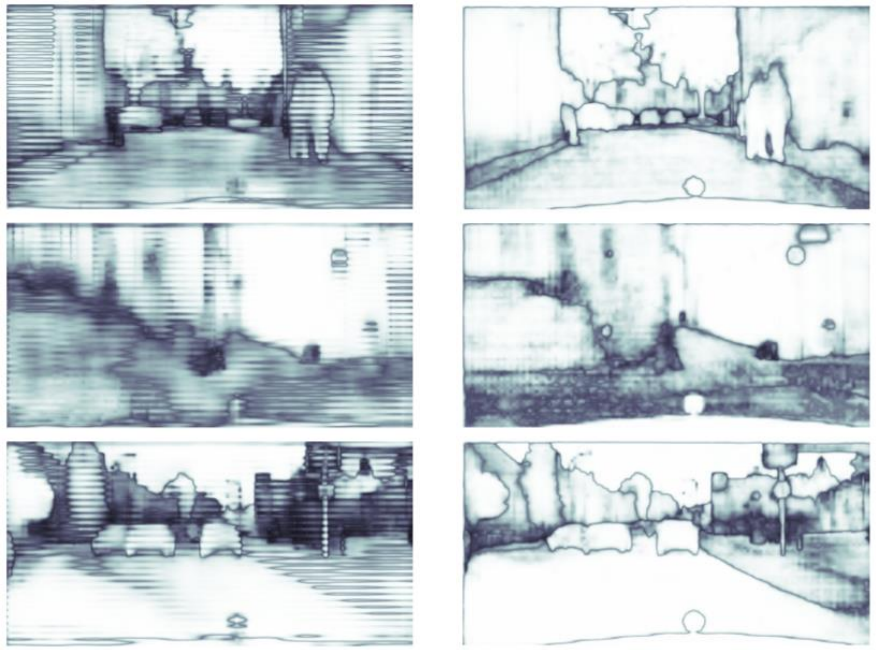


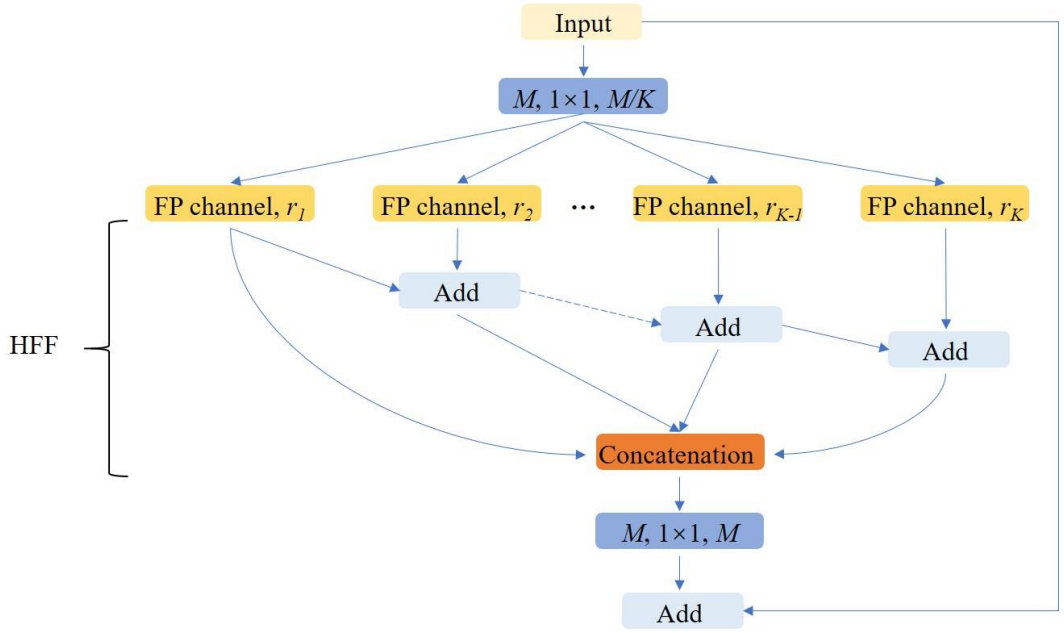

**Figure 2**. Naïve CFP module (left); Feature Pyramid (FP) channel (right).

# Method

- In order to eliminate the gridding artifacts, we applied hierarchical feature fusion (HFF) in naïve CFP module.



(a) (b)




(c)

**Figure 3**. (a) Feature without HFF; (b) Feature with HFF; (c) CFP module.

# Method

- We build CFPNet based on CFP module.

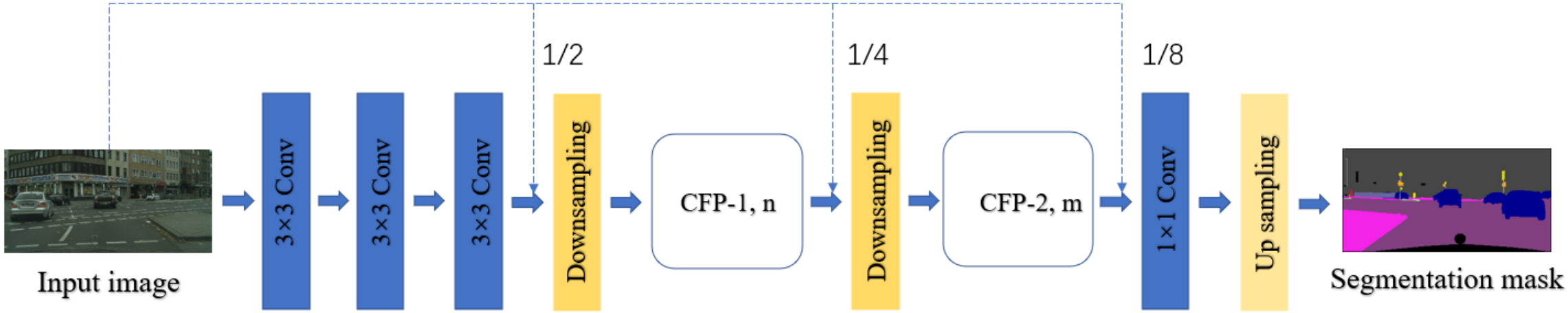



**Figure 4**. Architecture of CFPNet. (n, m) are repeat times of CFP-1 and CFP-2

# Datasets

- Cityscapes: It contains 5000 fine annotation and 20000 coarser annotation images. The whole datasets contain **19** classes that belong to 7 categories.
- CamVid: 101 for validation and 233 for testing. It contains **11** classes.

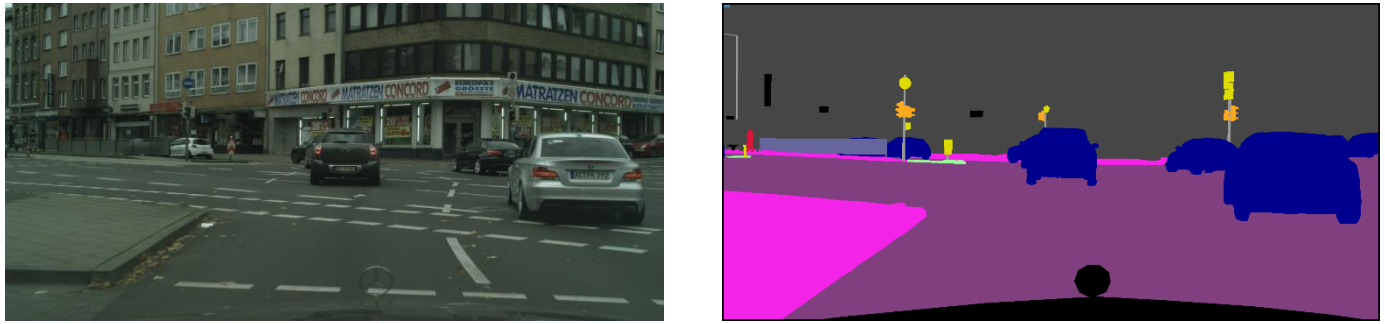


**Figure 5**. **Cityscapes** dataset. Images (left); Label (right).

**Figure 6**. **CamVid** dataset. Images (left); Label (right).

# Experiments

- We test different repeat times (n, m) = {(1, 2), (1, 3) and (2, 6)} on Cityscapes dataset.

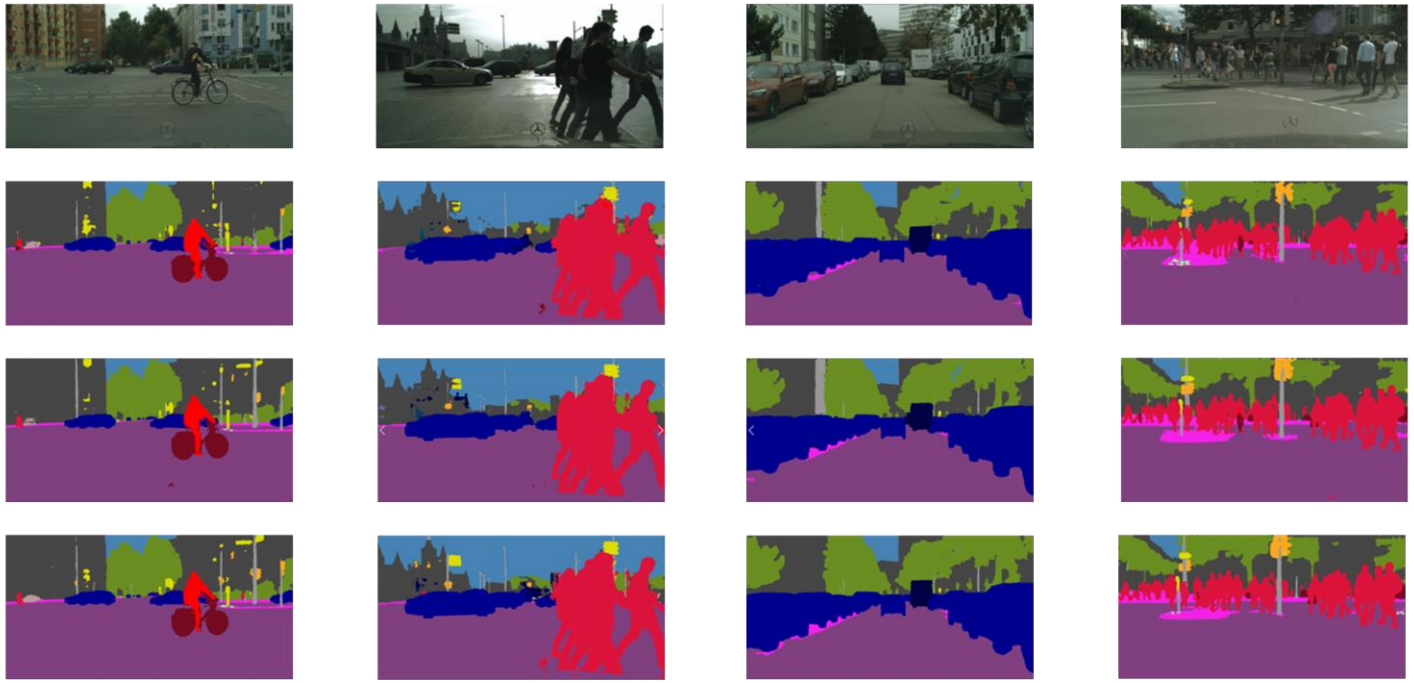


**Figure 7**. Results. From top to bottom is original image, CFPNet-V1, CFPNet-V2 and CFPNet-V3

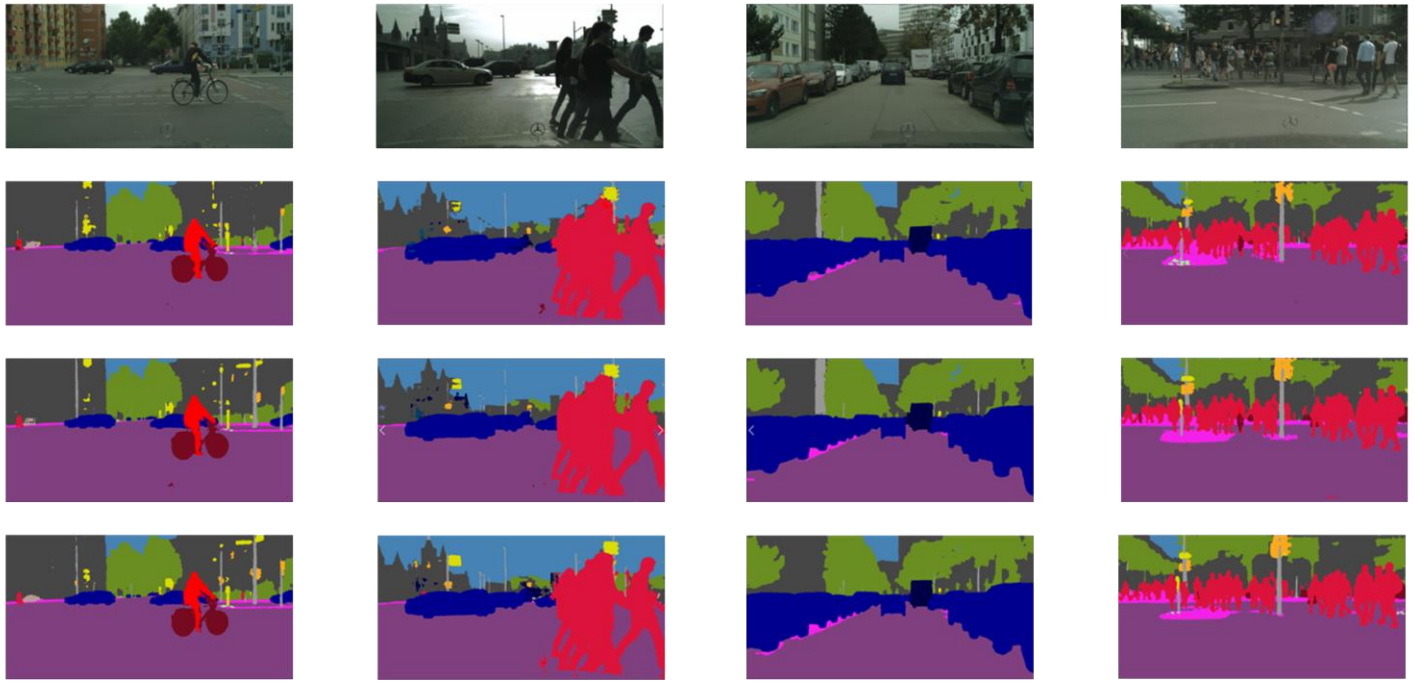


**Figure 7**. Results. From top to bottom is original image, CFPNet-V1, CFPNet-V2 and CFPNet-V3

| Net | Roa | Sid | Bui | Wal | Fen | Pol | TLi | TSi | Veg | Ter | Sky | Per | Rid | Car | Tru | Bus | Tra | Mot | Bic | mIoU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **V1** | 96.7 | 76.3 | 88.0 | 36.0 | 46.9 | 50.2 | 43.9 | 58.5 | 89.8 | 55.1 | 91.2 | 69.0 | 41.5 | 89.7 | 39.5 | 54.4 | 23.7 | 26.6 | 63.2 | 60.4 |
| **V2** | 97.4 | 79.8 | 89.9 | 47.4 | 50.7 | 55.3 | 53.9 | 67.2 | 90.9 | 59.2 | 93.4 | 74.5 | 48.6 | 92.3 | 49.9 | 63.1 | 31.1 | 49.3 | 69.9 | 66.5 |
| **V3** | **97.8** | **81.4** | **90.5** | 46.4 | 50.6 | **56.4** | **61.5** | **67.7** | **92.1** | **68.9** | **94.3** | **80.4** | **60.7** | **93.9** | **51.4** | **68.0** | **50.8** | **51.2** | 67.7 | **70.1** |

# Results

| Network | Pretrain | InputSize | mIoU (%) | FPS | Parameters | GPU |
|---|---|---|---|---|---|---|
| DeepLab-v2 | ImageNet | 512 × 1024 | 70.4 | < 1 | 44 | TitanX |
| PSPNet | ImageNet | 713 × 713 | 78.4 | < 1 | 65.7 | TitanX |
| SegNet | ImageNet | 360 × 640 | 56.1 | 14.6 | 29.5 | TitanX |
| ENet | None | 512 × 1024 | 58.3 | 76.9 | 0.4 | TitanX |
| SQ | ImageNet | 1024 × 2048 | 59.8 | 16.7 | - | TitanX |
| ESPNet | None | 512 × 1024 | 60.3 | 112 | 0.4 | TitanX-P |
| ContextNet | None | 1024 × 2048 | 66.1 | 18.3 | 0.85 | TitanX |
| ERFNet | None | 512 × 1024 | 68.0 | 41.7 | 2.1 | TitanX |
| BiSeNet | ImageNet | 768 × 1536 | 68.4 | 105.8 | 5.8 | TitanXp |
| ICNet | ImageNet | 1024 × 2048 | 69.5 | 30.3 | 7.8 | TitanX |
| CGNet | None | 360 × 640 | 64.8 | 50 | 0.5 | 2×V100 |
| LEDNet | None | 512 × 1024 | 70.6 | 71 | 0.94 | 1080Ti |
| DABNet | None | 1024 × 2048 | 70.1 | 27.7 | 0.76 | 1080Ti |
| **CFPNet** | **None** | **1024 × 2048** | **70.1** | **30** | **0.55** | **2080Ti** |

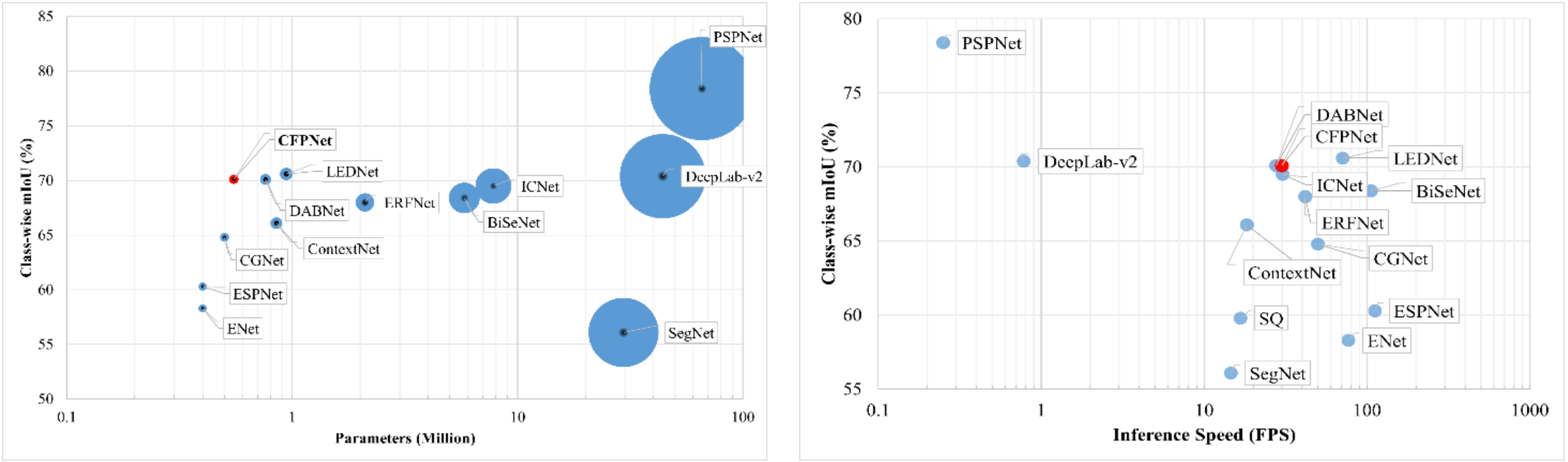
Table 1. Evaluation results on the **Cityscapes** test set.
TitanX represents the TitanX Maxwell, TitanX-P represents the TitanX Pascal and 2080Ti represents RTX 2080Ti

# Results

| Network | mIoU (%) | Parameters (M) |
|---|---|---|
| **ENet** | 51.3 | 0.36 |
| **SegNet** | 55.6 | 29.5 |
| **FCN-8s [21]** | 57.0 | 134.5 |
| **Dilation8 [33]** | 65.3 | 140.8 |
| **BiSeNet [32]** | 65.6 | 5.8 |
| **ESPNet [5]** | 55.6 | 0.36 |
| **CFPNet** | 64.7 | 0.55 |

Table 2. Performance on **CamVid** test set

# Results




(a) (b)

Figure 8. (a) Acc vs Parameters. (Note: the areas of circle are **proportional** to the parameters);
(b) Acc vs Speed. (Note: the inference of **GPU** and **input size**).

# Conclusion

Both the analysis and experimental results on Cityscapes and CamVid dataset show the potential of CFPNet as a tiny, fast and efficient network.

We continue to lead it into medical image segmentation.

Thank you for your attention.

# Questions & Comments