Graph-guided Architecture Search for Real-time Semantic Segmentation Supplemental Material

1. Network Visualization

As shown in Section 4.4.1 of the paper, the network searched by our GAS with GGM has smaller parameter size while achieving much higher performance. The visualization result can effectively help to analyze which component brings in the performance improvement. We thus visualize the networks searched by the three methods: 1) GAS with GGM; 2) GAS with fully connected layer; and 3) random search in Figure 1, Figure 2 and Figure 3, respectively.

Compared to the other methods, the network searched by our GAS with GGM shows the following three advantages:

1) The cells in the low stage tend to choose light-weight operations (i.e. none, max pooling, skip connection) and the cells in the high stage enjoy the complex ones, which is the goal of pursuing high speed as described in the introduction of our paper. Specifically, under the same latency loss weight, the network searched by our GAS with GGM contains thirty light-weight operations (dashed-line arrow in the picture) with lower latency, while the other two methods use twenty-one and twenty-three light-weight operations, respectively. However, our GAS with GGM achieves higher performance, which exhibits the emergence of the concept of burden sharing in a group of cells when they know how much others are willing to contribute.

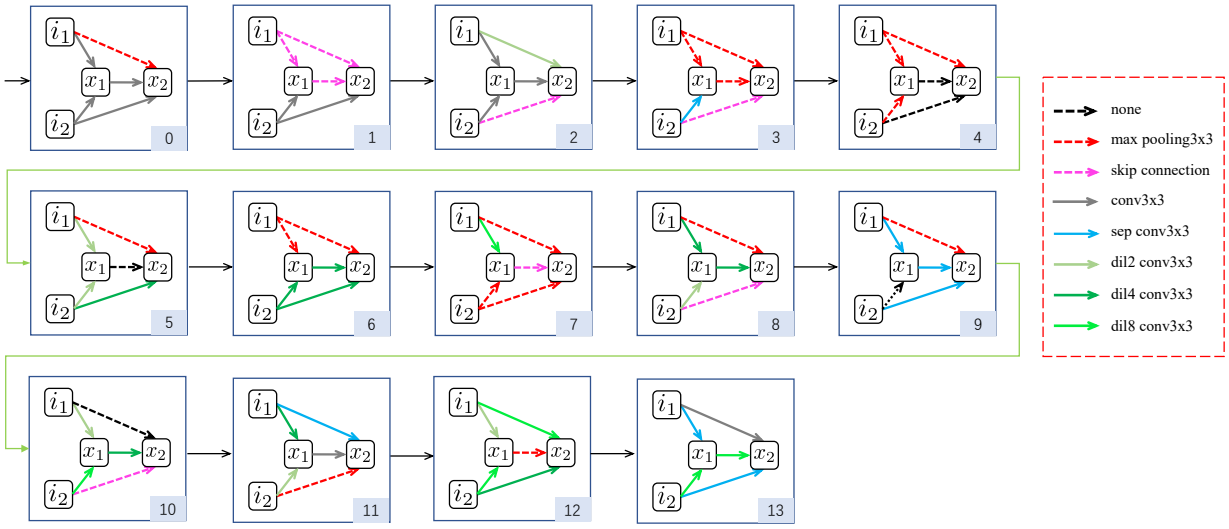

Figure 1. The network searched by our GAS with GGM exhibits the benefit property (e.g. more dilation convolution operations in deep layers and more low computational operations for fast speed) for real-time semantic segmentation.

2) The deeper layers tend to utilize larger receptive field operations (e.g. conv with dilation = 4 or 8), which plays a key role to improve performance in semantic segmentation [1, 2, 3]. Specifically, the network searched by our GAS with GGM uses 11 large receptive field operations (denoted by green arrow) in the last four cells and the other methods only use 4 or 8 operations, respectively.

3) The final structure has sufficient cell-level diversity as we expected. On the contrary, the network search by GAS with fully connected layer tends to use similar structures, for example, cell 7 is similar to cell 8 and 9, and cell 1 is similar to the cell 2, 3 and 4.

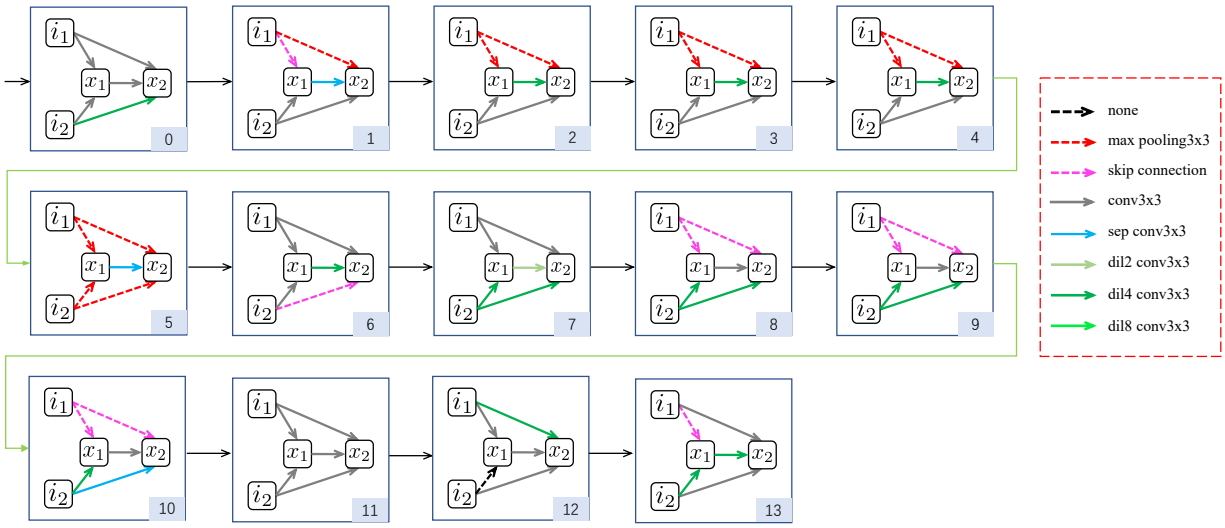

Figure 2. The network searched by our GAS with fully connected layer.

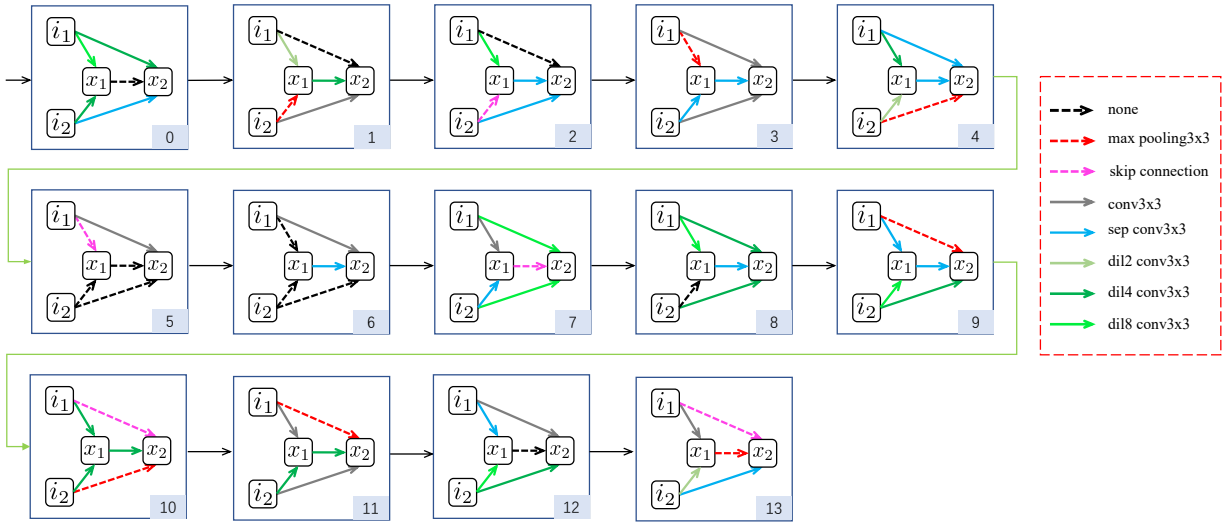

Figure 3. The random searched network.

2. Multi-Scale Module Exploration

Methods	mIoU (%)	FPS
ASPP	72.4	108.4
PPM	72.5	114.1

Table 1. The performance for different multi-scale modules on the Cityscapes validation set.

When considering multi-scale features, we also try the PPM module in PSPNet [4], and our GAS achieves the similar performance with faster speed on the Cityscapes validation set in Table 1.

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE trans. PAMI*, 40(4):834–848, 2018. 1

[2] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 1

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1

[4] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 2