Semantic Segmentation with Multi-path Refinement and Pyramid Pooling Dilated-Resnet

Zhipeng Cui, Qiao Zhang, Shijie Geng, Xiaoguang Niu, Yu Qiao
Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China

Motivation:

- **Resnet** which has best performance on object recognition could also lose fine structure due to repeated 2-step striding convolution .
- It has been verified in classification network that if we put batch normalization and relu before the convolution unit, we can get better results through **pre-activation**, which could make **optimization more efficiently and raise regularization** in case of overfitting.
- **Low-level features** are very necessary for **accurate high resolution semantic segmentation on the boundaries**. **PSPNet** lacks in extracting intermediate layer information from lower blocks which can be solved by **RefineNet**.

Network Structures:

- A new segmentation framework based on **ResNet-101** with new **dilated residual unit** illustrated in Figure. 1.
- **Multiple resolution of feature maps** from intermediate layers are combined to refine the output precision.
- The feature map from the last dilated residual unit is used as the input of **pyramid pooling**.

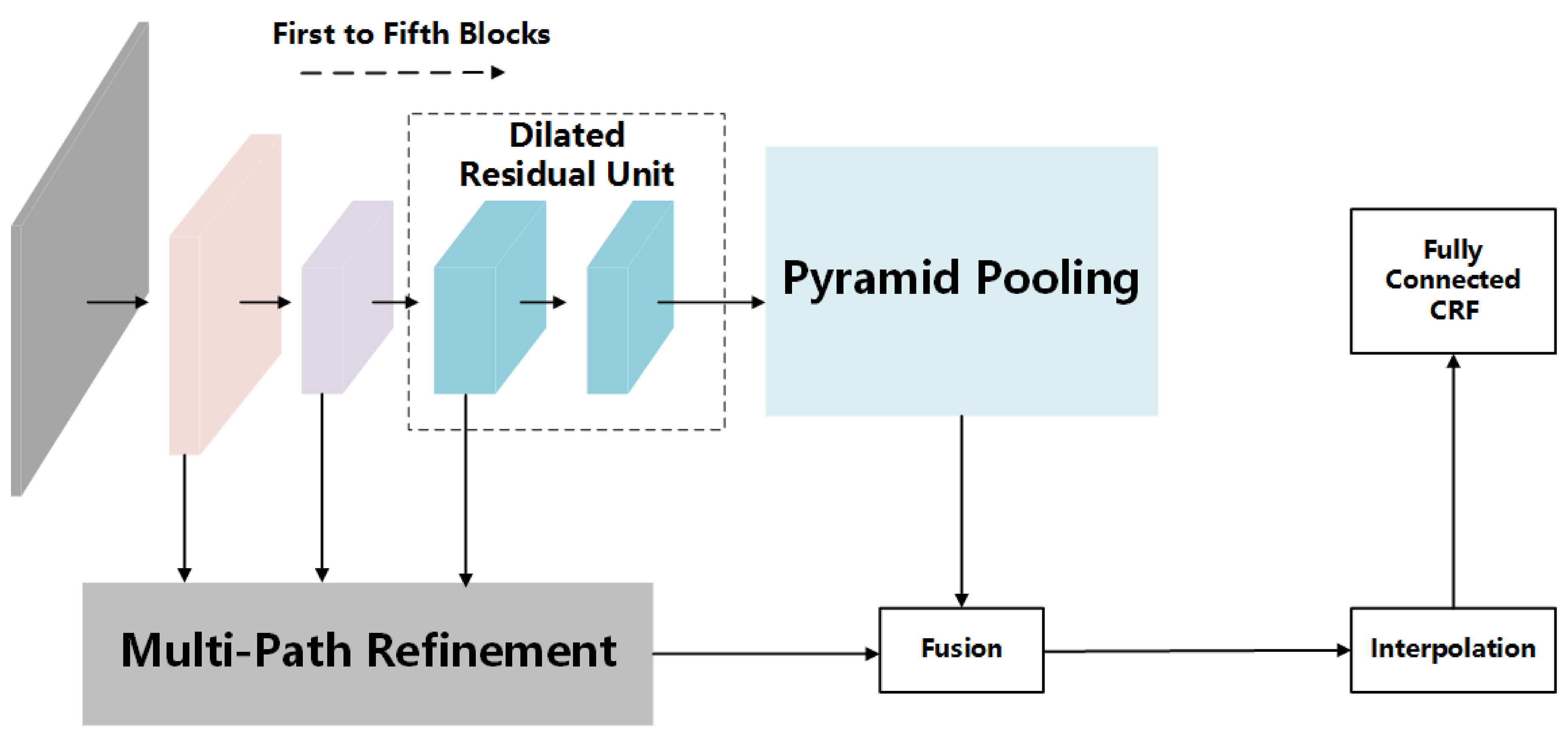

Figure 1. Network Architecture

BN-ReLU Dilated Residual Unit

- When using regular convolution operations to extract features, it would diminish view of field, which could not entirely rebuilt by upsampling. Therefore, we employ **dilated convolution** in 3 × 3 stage of residual unit.

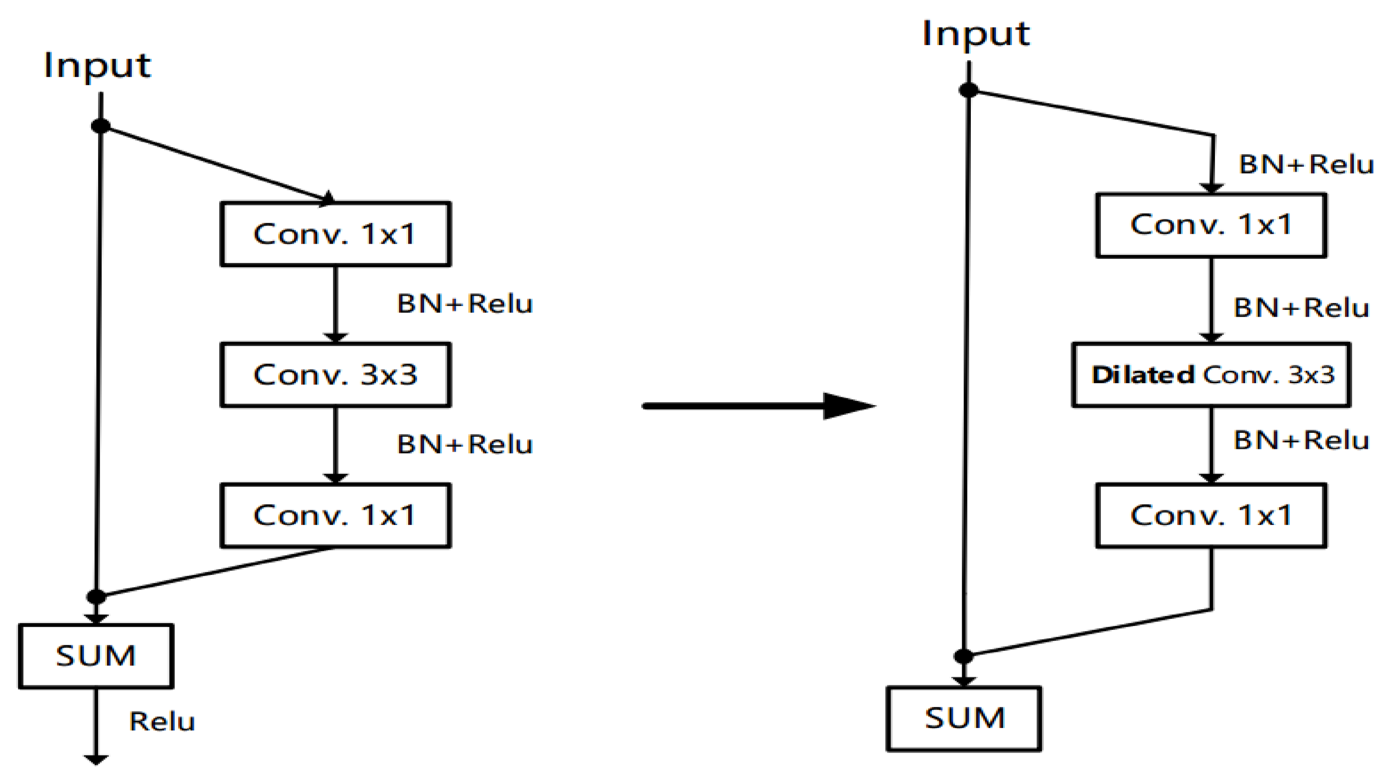

Figure 2. BN-ReLU Dilated ReLU Residual Unit

Pymraid Pooling Module

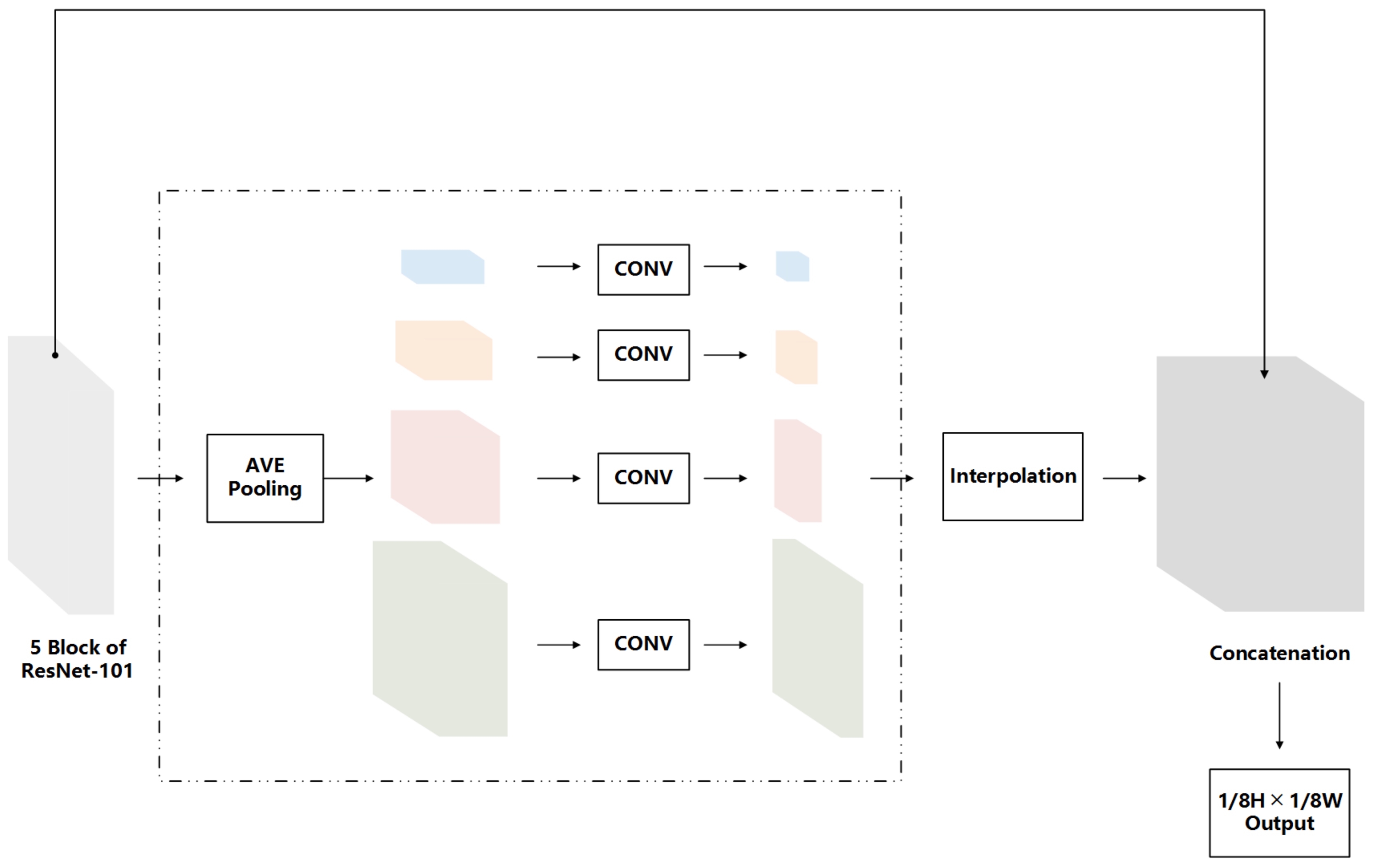

Figure 3. Pyramid Pooling

- **Global average pooling** can be implemented to extract pooled context features from any layer. The final block in our network consists of higher **semantic and global context information.** So a pyramid of global average pooling connected to the final output can **provide context information of multiple sub-regions**.

Results

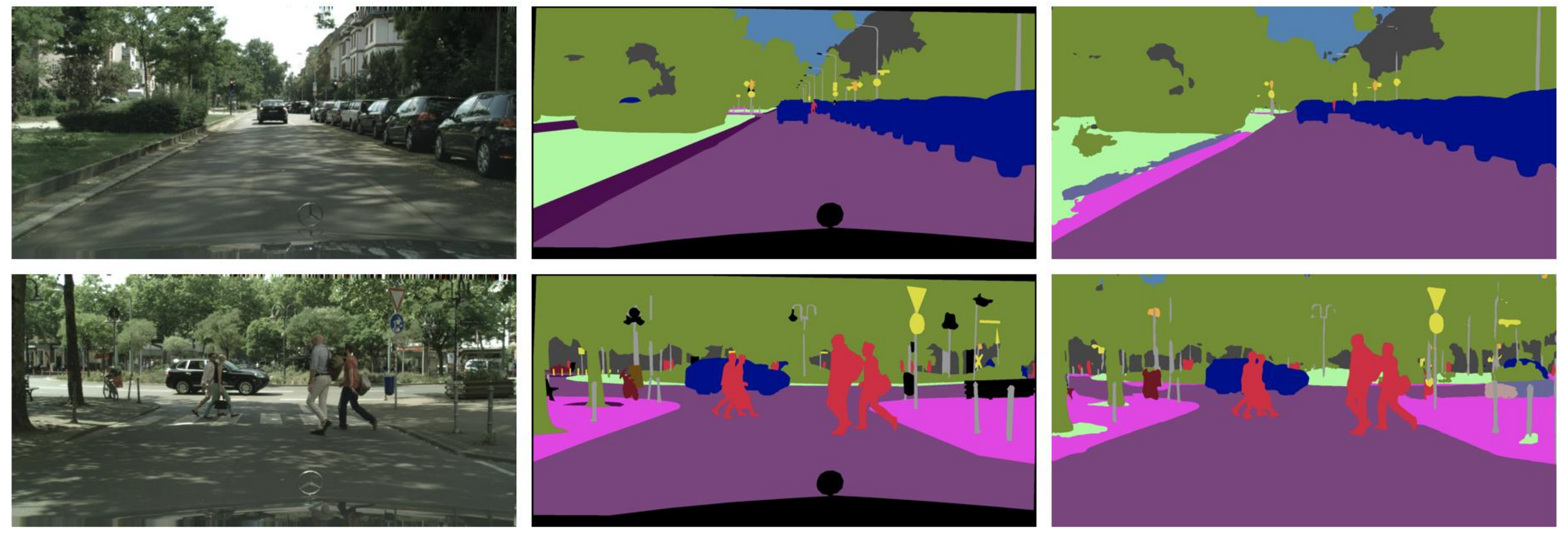

Figure 4. Results

	CRFasRNN	FCN	Dilation	LRR	Deeplab	Ours
road	96.3	97.4	97.6	97.7	**97.9**	**97.9**
swalk	73.9	78.4	79.2	79.9	81.3	**81.4**
build.	88.2	89.2	89.9	**90.7**	90.3	90.1
wall	47.6	34.9	37.3	44.4	**48.8**	48.7
fence	41.3	44.2	47.6	48.6	47.4	**59.0**
pole	35.2	47.4	53.2	58.6	49.6	**59.6**
tlight	49.5	60.1	58.6	**68.2**	57.9	**68.2**
sign	59.7	65.0	65.2	72.0	67.3	**75.2**
veg.	90.6	91.4	91.8	**92.5**	91.9	92.3
terrain	66.1	69.3	69.4	69.3	**69.4**	63.8
sky	93.5	93.9	93.7	**94.7**	94.2	86.5
person	70.4	77.1	78.9	81.6	79.8	**82.7**
rider	34.7	51.4	55.0	60.0	59.8	**63.5**
car	90.1	92.6	93.3	94.0	93.7	**95.2**
truck	39.2	35.3	45.5	43.6	56.5	**66.1**
bus	57.5	48.6	53.4	56.8	67.5	**83.7**
train	55.4	46.5	47.7	47.2	57.5	**67.8**
mbike	43.9	51.6	52.2	54.8	57.7	**65.0**
bike	54.6	66.8	66.0	69.7	68.8	**71.7**
Mean IoU	62.5	65.3	67.1	69.7	70.4	**74.7**

Table 1. Per-class and Mean IoU results on Cityscapes