RefineNet: Multi-path refinement networks for high-resolution semantic segmentation
(2017)

Lin, Guosheng, Milan, Anton, Shen, Chunhua Reid, Ian

Resume

March 18, 2019

---

1 Introduction

The authors present RefineNet, a generic multi-path refinement network that explicitly exploits all the information available along the downsampling process to enable high-resolution prediction using long-range residual connections.

2 Method

The proposed method, is to use the outputs of different stages in resnet, and pass them through RegineNet modules, each time fusing the outputs of smaller outputs of a refine module with the next bigger one, unitil we get a feature map of size 1/4 containing the predictions, this is another way to preserve the spatial information without using dilations that are computationnaly expensive.

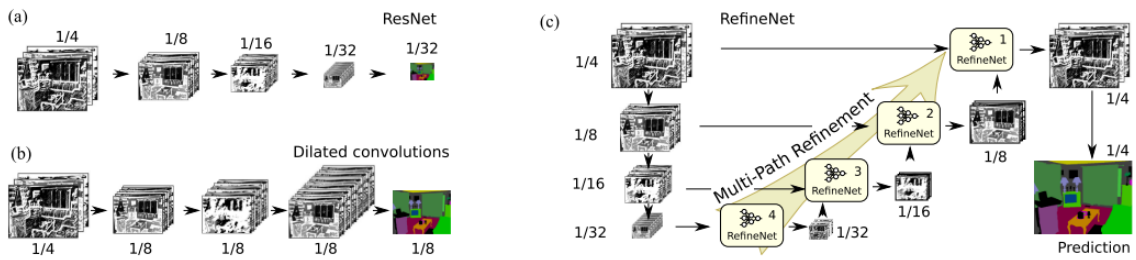

Figure 2. Comparison of fully convolutional approaches for dense classification. Standard multi-layer CNNs, such as ResNet *(a)* suffer from downscaling of the feature maps, thereby losing fine structures along the way. Dilated convolutions *(b)* remedy this shortcoming by introducing atrous filters, but are computationally expensive to train and quickly reach memory limits even on modern GPUs. Our proposed architecture that we call RefineNet *(c)* exploits various levels of detail at different stages of convolutions and fuses them to obtain a high-resolution prediction without the need to maintain large intermediate feature maps. The details of the RefineNet block are outlined in Sec. 3 and illustrated in Fig 3.

3 The model

We have four refine net blocks, each one takes as input a specific output of the resnet stages (with spatial dimensions equal to 1/4 - 1/8 - 1/16 - 1/32), :

1. Residual convolution unit. The first part of each RefineNet block consists of an adaptive convolution set that mainly fine-tunes the pretrained ResNet weights for our task. The filter number for each input path is set to 512 for RefineNet-4 and 256 for the remaining ones in our experiments.

2. Multi-resolution fusion. All path inputs are then fused into a high-resolution feature map by the multi-resolution fusion block, with a conv to adapth the number of channels to the smaller one of the two, and upsample the feaetures to the bigger spatial dimensions.

3. Chained residual pooling. Aims to capture background context from a large image region. It is able to efficiently apply a successive pooling of the input features with multiple window sizes and fuse them together using learnable weight (the convolutions before the sum is regarded the weights for the summation).

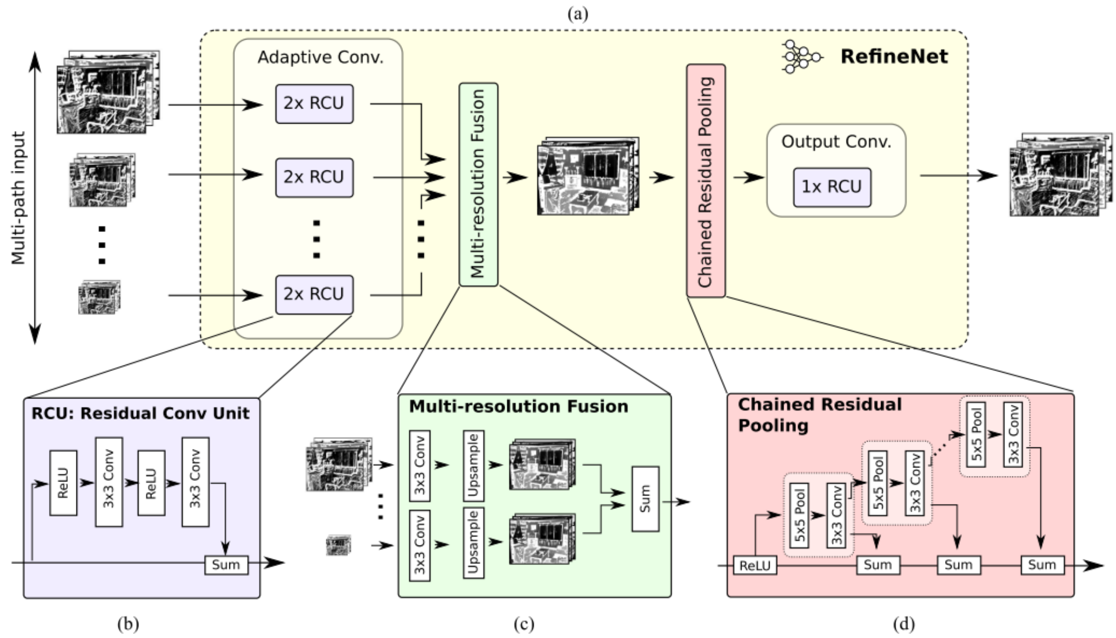

Figure 3. The individual components of our multi-path refinement network architecture RefineNet. Components in RefineNet employ residual connections with identity mappings. In this way, gradients can be directly propagated within RefineNet via local residual connections, and also directly propagate to the input paths via long-range residual connections, and thus we achieve effective end-to-end training of the whole system.

4 Results

Table 1. Object parsing results on the Person-Part dataset. Our method achieves the best performance (bold).

method	IoU
Attention [7]	56.4
HAZN [45]	57.5
LG-LSTM [29]	58.0
Graph-LSTM [28]	60.2
DeepLab [5]	62.8
DeepLab-v2 (Res101) [6]	64.9
RefineNet-Res101 (ours)	**68.6**

Table 2. Ablation experiments on NYUDv2 and Person-Part.

Initialization	Chained pool.	Msc Eva	NYUDv2	Person-Parts
ResNet-50	no	no	40.4	64.1
ResNet-50	yes	no	42.5	65.7
ResNet-50	yes	yes	43.8	67.1
ResNet-101	yes	no	43.6	67.6
ResNet-101	yes	yes	44.7	68.6
ResNet-152	yes	yes	**46.5**	**68.8**

Table 3. Segmentation results on NYUDv2 (40 classes).

method	training data	pixel acc.	mean acc.	IoU
Gupta et al. [20]	RGB-D	60.3	-	28.6
FCN-32s [36]	RGB	60.0	42.2	29.2
FCN-HHA [36]	RGB-D	65.4	46.1	34.0
Context [30]	RGB	70.0	53.6	40.6
RefineNet-Res152	RGB	**73.6**	**58.9**	**46.5**

Table 4. Segmentation results on the Cityscapes *test* set. our method achieves the best performance.

Method	IoU
FCN-8s [36]	65.3
DPN [35]	66.8
Dilation10 [46]	67.1
Context [30]	71.6
LRR-4x [17]	71.8
DeepLab [5]	63.1
DeepLab-v2(Res101) [6]	70.4
RefineNet-Res101 (ours)	**73.6**