The progress of the reading plan:

Index	Semantic Seg	All
14	13	50

Paper Information

Paper Title :

[DenseASPP for Semantic Segmentation in Street Scenes]

Conference :

CVPR 2018

Authors and Institutions

Authors

- Maoke Yang
- Kun Yu
- Chi Zhang
- Zhiwei Li
- Kuiyuan Yang

Institutions

DeepMotion Inc.

Official Codes

https://github.com/DeepMotionAIResearch/DenseASPP

Some articles to comprehend this paper

语义分割–(DenseASPP)DenseASPP for Semantic Segmentation in Street Scenes

Network Structure

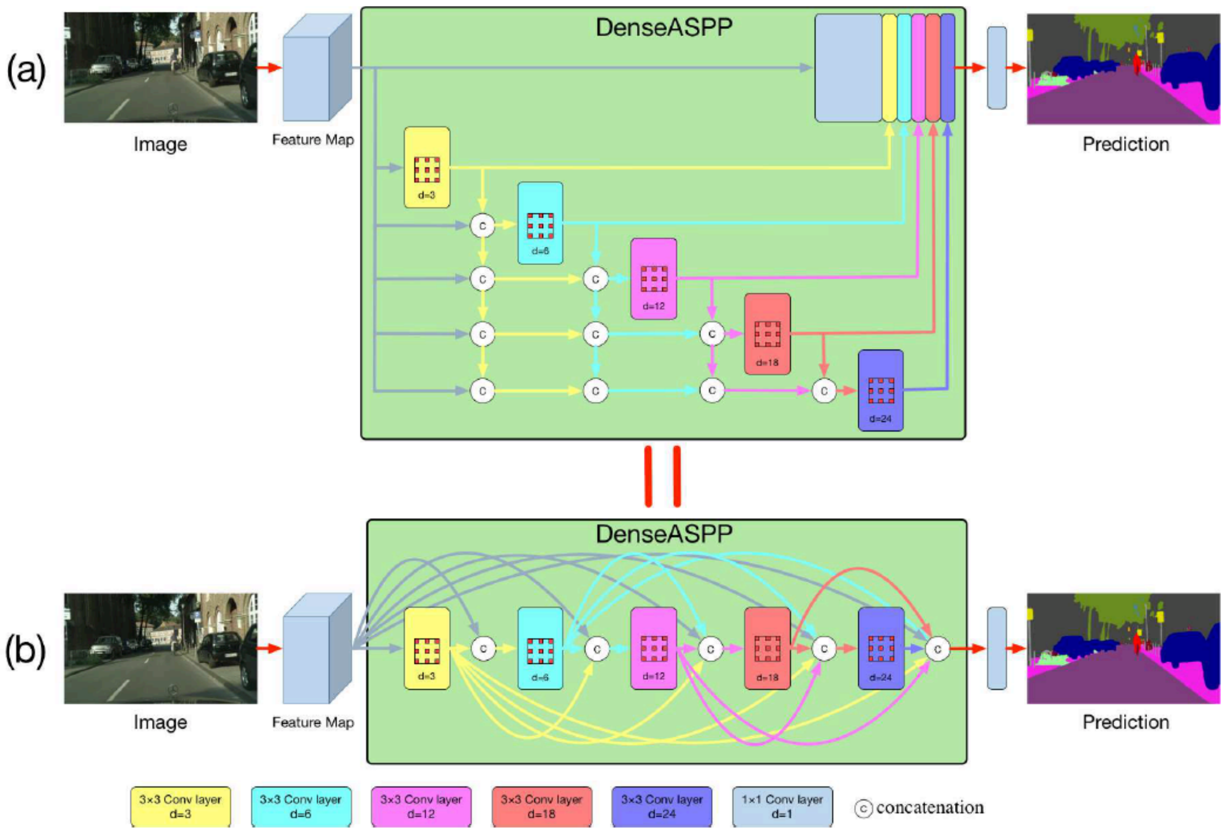

Figure 2. The structure of DenseASPP, (a) illustrate DenseASPP in detail, the output of each dilated convolutional layer is concatenated with input feature map, and then feed into the next dilated layer. Each path of DenseASPP compose a feature representation of correspond scale. (b) illustrate this structure in a more concrete version

Note

This paper belongs to the branch of enrich the information of features. Though the struture seems easy, but this kind of method is really useful.

Key Words

Five questions about this paper:

1. [Problem Definition / Motivation] What problem is this paper trying to solve?

- Atrous convolution is proposed to enlarge the size of reception field and dicrease the cost of computing simultaneously.
- ASPP is proposed to concatenate feature maps generated by atrous convolution with different dilation rates.

But when high resolution images are used as the input of segmentation methods, the size of reception field never becomes big enough. And large enough dilation ratio will decrease the power of model.

Therefore, it is important to design a network structure that is able to encode multi-scale information, and simultaneously achieves a large enough receptive field size.

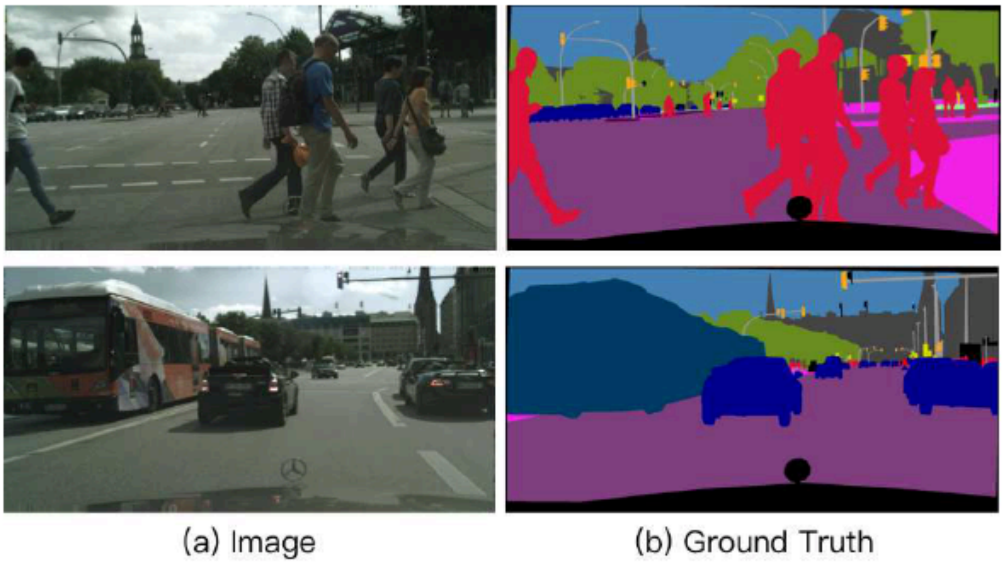

Figure 1. Illustration of challenging scale variations of street scenes from Cityscapes [4]. In the first exemplar image, the same category such as person varies largely in scale caused by distance to the camera. The second exemplar image illustrates an even challenging case where a large bus is close while several small traffic lights are far away.

PS: deeper network will have bigger receptive field, so actually I'dont think it’s a big problem.

2. [Contribution / Method] What’s new in this paper? / How does this paper solve the above problems?

The network structure is neat and articulate.

Through cascade, applying d=6 conv on the feature map of d=3 conv, we can actually get the bigger receptive field than a normal d=6 conv can get.

This undirect method is wonderful!!!

Multi-grid features are merged together to get bigger receptive field and in the same time remaining the rich information.

3. Details about the experiment

3.1 Which Datasets are used?

- CityScapes

3.2 How is the experiment set up?

3.3 What's the evaluation metric?

mIoU

3.4 Ablation Study

3.5 What is the ranking of the experiment results?

Table 1. Category-wise comparison on the Cityscapes test set.

Methods	mIoU	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle
FCN-8s [16]	65.3	97.4	78.4	89.2	34.9	44.2	47.4	60.1	65	91.4	69.3	93.9	77.1	51.4	92.6	35.3	48.6	46.5	51.6	66.8
DeepLabv2-CRF[2]	70.4	97.9	81.3	90.3	48.8	47.4	49.6	57.9	67.3	91.9	69.4	94.2	79.8	59.8	93.7	56.5	67.5	57.5	57.7	68.8
FRRN[19]	71.8	98.2	83.3	91.6	45.8	51.1	62.2	69.4	72.4	92.6	70	94.9	81.6	62.7	94.6	49.1	67.1	55.3	53.5	69.5
RefineNet[15]	73.6	98.2	83.3	91.3	47.8	50.4	56.1	66.9	71.3	92.3	70.3	94.8	80.9	63.3	94.5	64.6	76.1	64.3	62.2	70
PEARL[11]	75.4	98.4	84.5	92.1	54.1	56.6	60.4	69	74	92.9	70.9	95.2	83.5	65.7	95	61.8	72.2	69.6	64.8	72.8
GCN[18]	76.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DUC[21]	77.6	98.5	85.5	92.8	58.6	55.5	65	73.5	77.9	93.3	72	95.2	84.8	68.5	95.4	70.9	78.8	68.7	65.9	73.8
PSPNet[24]	78.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ResNet-38[22]	78.4	98.5	85.7	93.1	55.5	59.1	67.1	74.8	78.7	93.7	72.6	95.5	86.6	69.2	95.7	64.5	78.8	74.1	69	76.7
DenseASPP(Ours)	80.6	98.7	87.1	93.4	60.7	62.7	65.6	74.6	78.5	93.6	72.5	95.4	86.2	71.9	96.0	78.0	90.3	80.7	69.7	76.8

4. Advantages (self-summary rather than the author's)

1. This idea is very simple but effective. It's good for us to read this papaer carefully for many times.
2. The experiment contains in this paper is really comprehensive.
3. The charts are beautiful!

5. Disadvantages (self-summary rather than the author's)

1. Different level of layers contain different amount of information. So may be some layer is more important than the other layers. If we add the attention module here, the better performance will be reached.
2. Further more, for faster computation, we can take apart the kernel of atrous convolution from kxk to kx1 and 1xk.