DeepLab V1

SEMANTIC IMAGE SEGMENTATION WITH DEEP CONVOLUTIONAL NETS AND FULLY CONNECTED CRFS

2014 CVPR

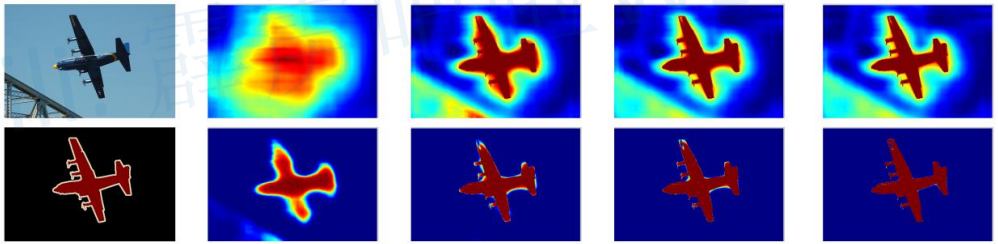


Figure 2: Score map (input before softmax function) and belief map (output of softmax function) for Aeroplane. We show the score (1st row) and belief (2nd row) maps after each mean field iteration. The output of last DCNN layer is used as input to the mean field inference. Best viewed in color.

论文下载地址：https://arxiv.org/abs/1412.7062

博文推荐：https://blog.csdn.net/qq_37541097/article/details/121692445

DeepLab V1

语义分割任务中存在的问题

There are two technical hurdles in the application of DCNNs to image labeling tasks: ***signal downsampling****, and* ***spatial 'insensitivity' (invariance)****.*

解决方案：

- 'atrous'(with holes) algorithm（空洞卷积 / 膨胀卷积 / 扩张卷积）
- fully-connected CRF(Conditional Random Field)

DeepLab V1

网络优势

- ✓ 速度更快，论文中说是因为采用了膨胀卷积的原因，但fully-connected CRFs很耗时
- ✓ 准确率更高，相比之前最好的网络提升了7.2个点
- ✓ 模型结构简单，主要由DCNNs和CRFs联级构成

(Deep Covolutional Neural Netsworks)

Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.3
DeepLab-MSc-CRF-LargeFOV	71.6

DeepLab V1

LargeFOV(Field of View)

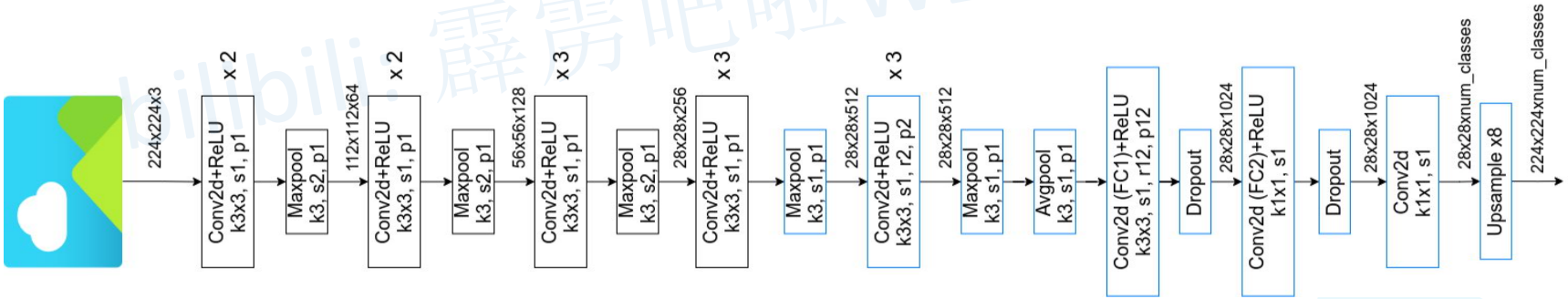
After converting the network to a fully convolutional one, the first fully connected layer has 4,096 filters of large 7 × 7 spatial size and ***becomes the computational bottleneck*** *in our dense score map computation. We have addressed this practical problem by* ***spatially subsampling (by simple decimation) the first FC layer to 4×4 (or 3×3) spatial size****.*

Method	kernel size	input stride	receptive field	# parameters	mean IOU (%)	Training speed (img/sec)
DeepLab-CRF-7x7	7×7	4	224	134.3M	67.64	1.44
DeepLab-CRF	4×4	4	128	65.1M	63.74	2.90
DeepLab-CRF-4x4	4×4	8	224	65.1M	67.14	2.90
DeepLab-CRF-LargeFOV	3×3	12	224	20.5M	67.64	4.84

Table 2: Effect of Field-Of-View. We show the performance (after CRF) and training speed on the PASCAL VOC 2012 'val' set as the function of (1) the kernel size of first fully connected layer, (2) the input stride value employed in the atrous algorithm.

DeepLab V1

DeepLab-LargeFOV



https://www.cs.jhu.edu/~alanlab/ccvl/DeepLab-LargeFOV/train.prototxt

DeepLab V1

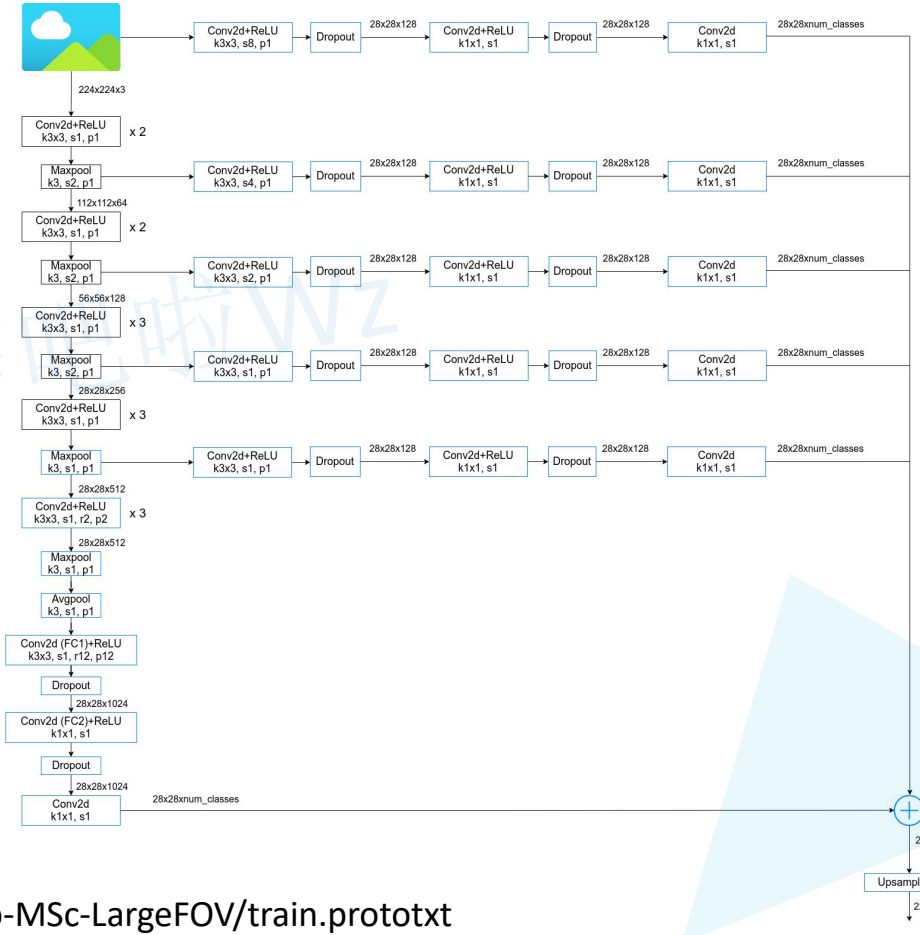
MSc(Multi-Scale)

*Specifically, we attach to the **input image** and the **output of each of the first four max pooling layers** a two-layer MLP (first layer: 128 3x3 convolutional filters, second layer: 128 1x1 convolutional filters) whose feature map is concatenated to the main network's last layer feature map. The aggregate feature map fed into the softmax layer is thus enhanced by 5 * 128 = 640 channels.*

Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.3
DeepLab-MSc-CRF-LargeFOV	71.6

DeepLab V1

DeepLab-MSc-LargeFOV



https://www.cs.jhu.edu/~alanlab/ccvl/DeepLab-MSc-LargeFOV/train.prototxt