1. DeepLab v1

Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs

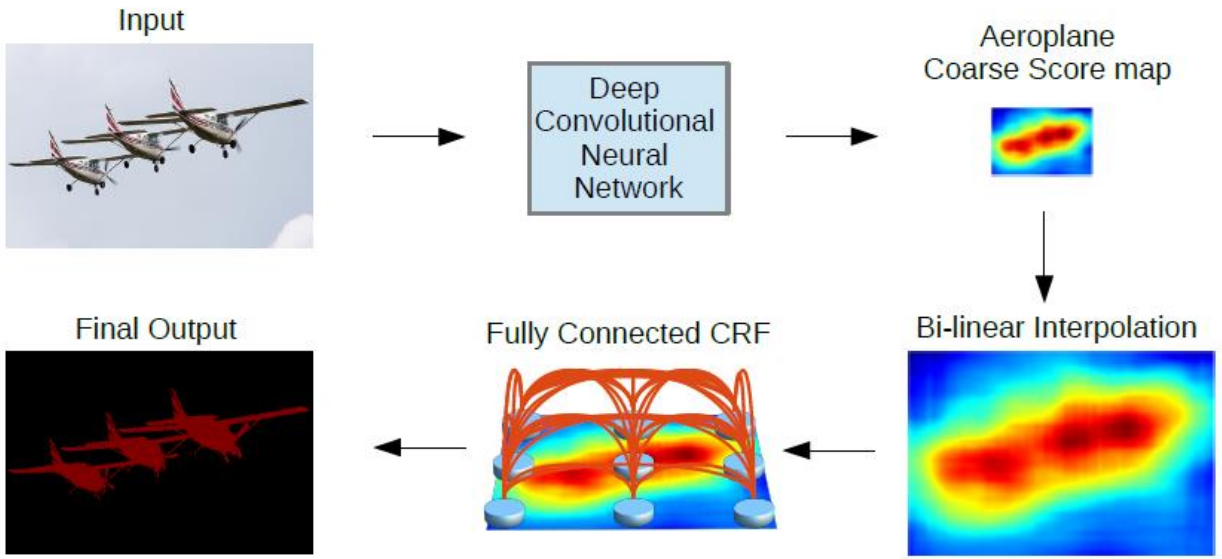
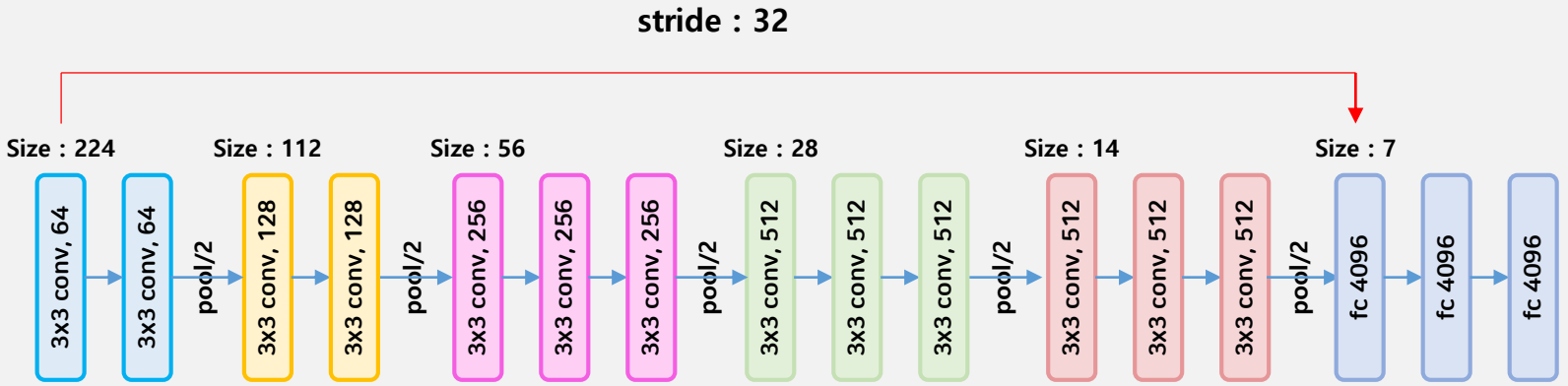


Figure 3: Model Illustration. The coarse score map from Deep Convolutional Neural Network (with fully convolutional layers) is upsampled by bi-linear interpolation. A fully connected CRF is applied to refine the segmentation result. Best viewed in color.

1. DeepLab v1


stride : 32
Size : 224
Size : 112
Size : 56
Size : 28
Size : 14
Size : 7
3x3 conv, 64
3x3 conv, 64
pool/2
3x3 conv, 128
3x3 conv, 128
pool/2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
pool/2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool/2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool/2
fc 4096
fc 4096
fc 4096

1. DeepLab v1

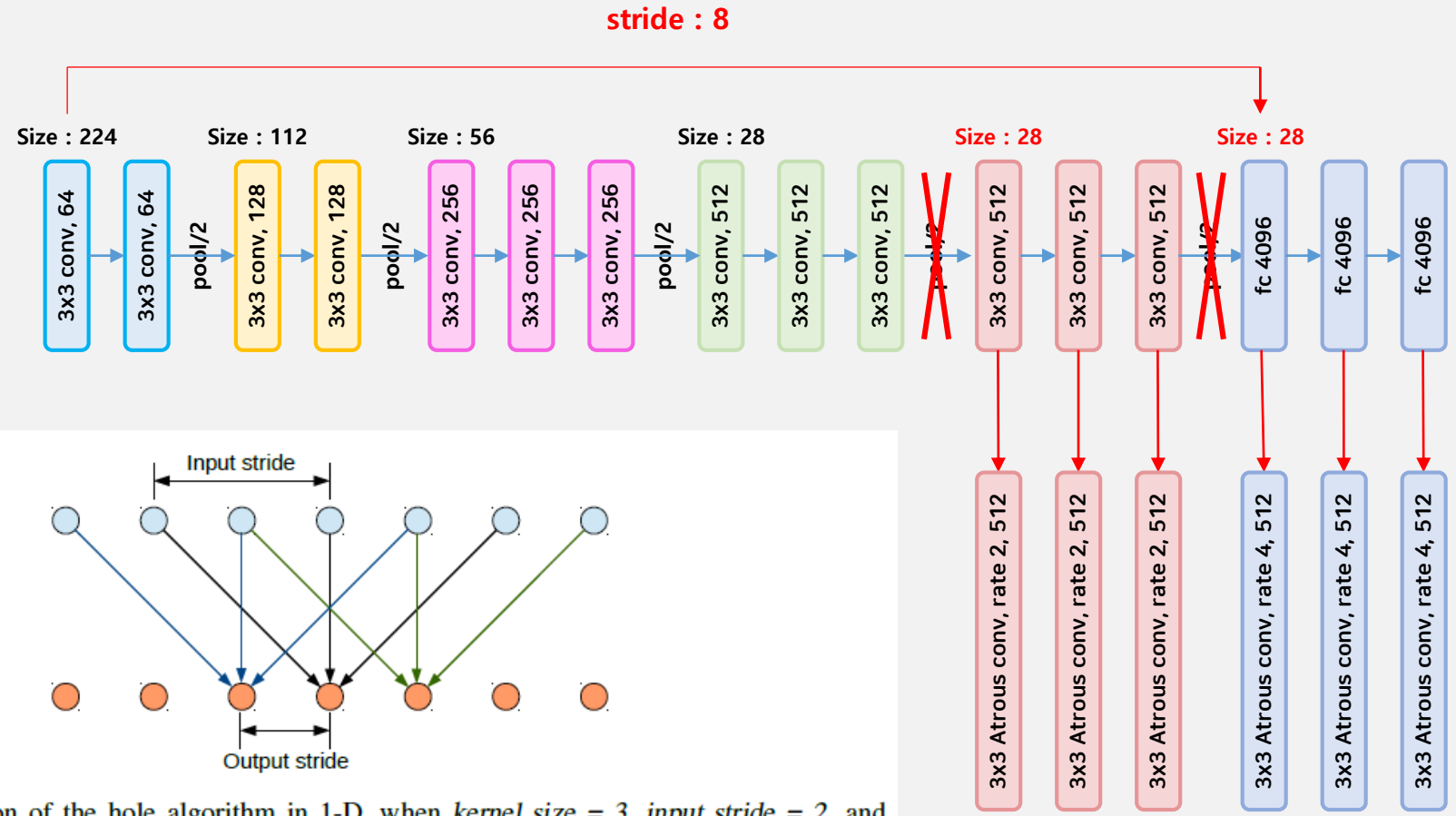
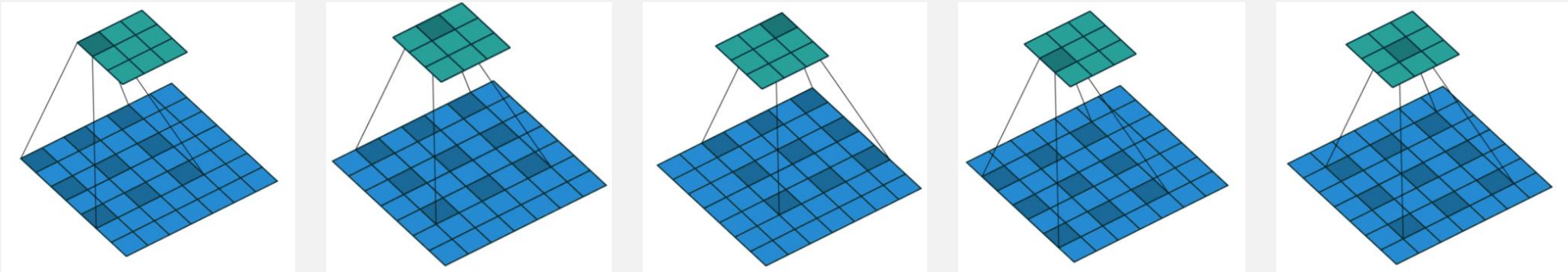


Figure 1: Illustration of the hole algorithm in 1-D, when *kernel_size* = 3, *input_stride* = 2, and *output_stride* = 1.

1. DeepLab v1



1. DeepLab v1

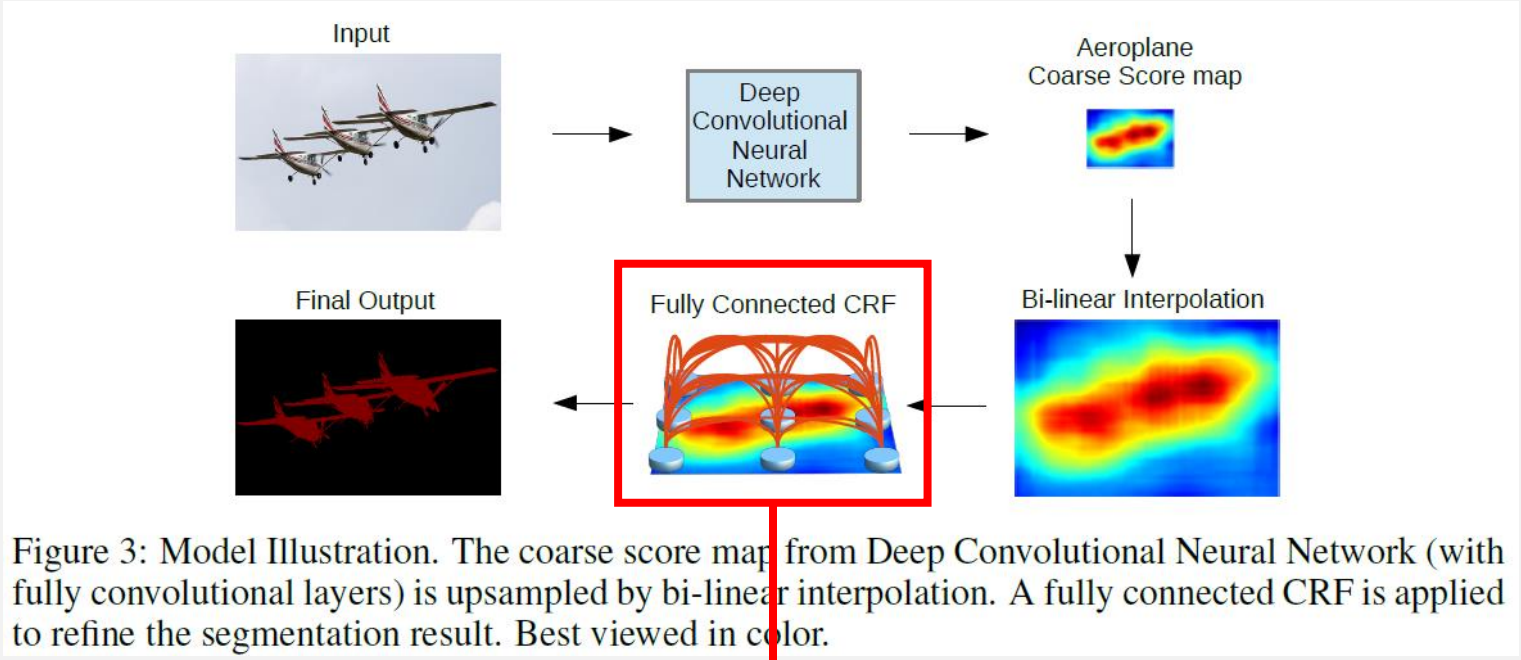
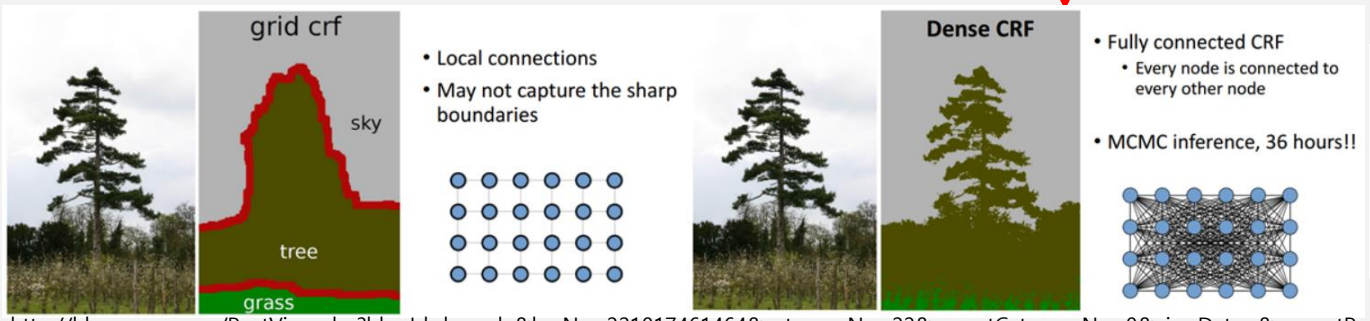


Figure 3: Model Illustration. The coarse score map from Deep Convolutional Neural Network (with fully convolutional layers) is upsampled by bi-linear interpolation. A fully connected CRF is applied to refine the segmentation result. Best viewed in color.



$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$w_1 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\alpha^2} - \frac{||I_i - I_j||^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{||p_i - p_j||^2}{2\sigma_\gamma^2}\right)$$

http://blog.naver.com/PostView.nhn?blogId=laonple&logNo=221017461464&categoryNo=22&parentCategoryNo=0&viewDate=&currentPage=2&postListTopCurrentPage=&from=postList&userTopListOpen=true&userTopListCount=10&userTopListManageOpen=false&userTopListCurrentPage=2

1. DeepLab v1

Method	mean IOU (%)
DeepLab	59.80
DeepLab-CRF	63.74
DeepLab-MSc	61.30
DeepLab-MSc-CRF	65.21
DeepLab-7x7	64.38
DeepLab-CRF-7x7	67.64
DeepLab-LargeFOV	62.25
DeepLab-CRF-LargeFOV	67.64
DeepLab-MSc-LargeFOV	64.21
DeepLab-MSc-CRF-LargeFOV	68.70

(a)

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

(b)

Table 1: (a) Performance of our proposed models on the PASCAL VOC 2012 'val' set (with training in the augmented 'train' set). The best performance is achieved by exploiting both multi-scale features and large field-of-view. (b) Performance of our proposed models (with training in the augmented 'trainval' set) compared to other state-of-art methods on the PASCAL VOC 2012 'test' set.

2. DeepLab v2

DeepLab v1 vs. DeepLab v2 (DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs)

➢ **공통점**

- Atrous Convolution
- Fully Connected CRF

➢ **차이점**

- Multiple Scale 처리방법 (ASPP)
- Back bone network (VGG-16 / ResNet-101)

2. DeepLab v2

Multiple Scale → Atrous Spatial Pyramid Pooling (ASPP)

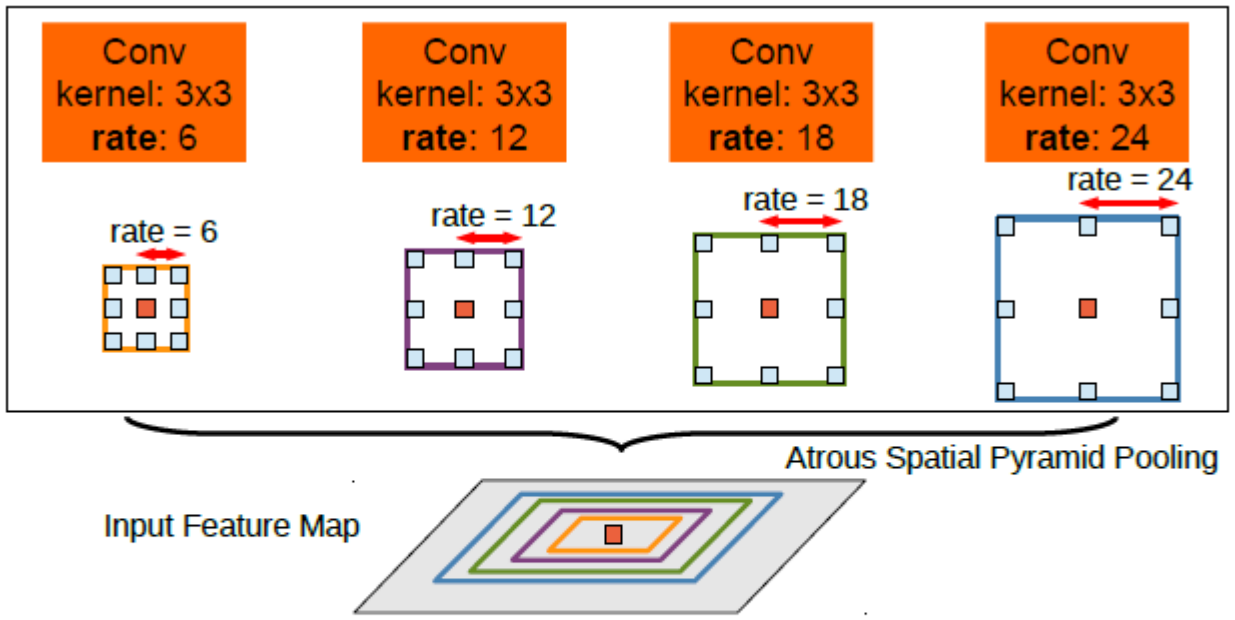
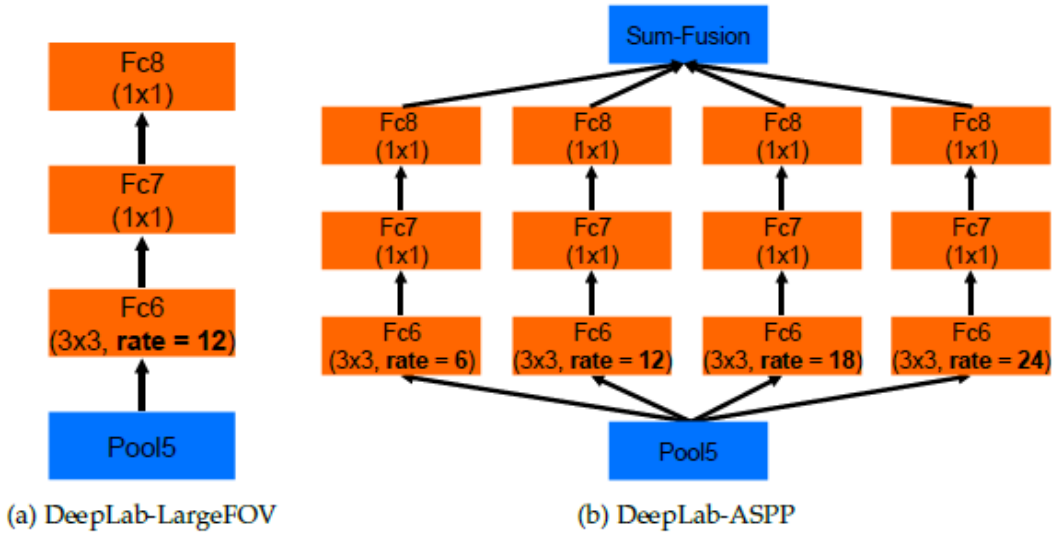


Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

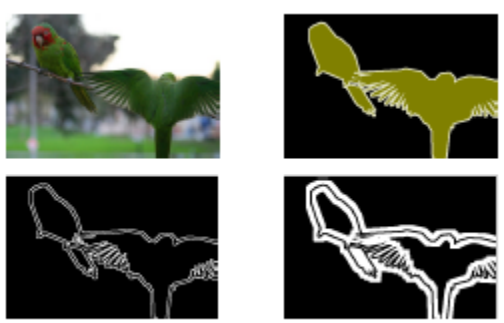


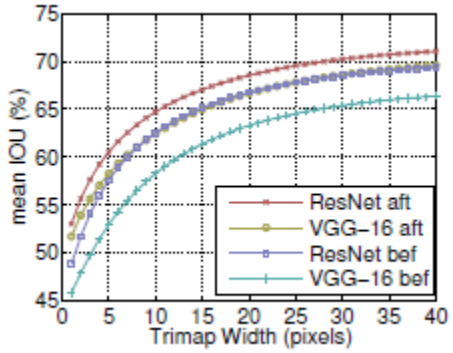
(a) DeepLab-LargeFOV (b) DeepLab-ASPP

Fig. 7: DeepLab-ASPP employs multiple filters with different rates to capture objects and context at multiple scales.

2. DeepLab v2

Back bone network → VGG-16 / ResNet-101





(a) (b)

Fig. 10: (a) Trimap examples (top-left: image. top-right: ground-truth. bottom-left: trimap of 2 pixels. bottom-right: trimap of 10 pixels). (b) Pixel mean IOU as a function of the band width around the object boundaries when employing VGG-16 or ResNet-101 before and after CRF.

2. DeepLab v2

Method	mIOU
DeepLab-CRF-LargeFOV-COCO [58]	72.7
MERL_DEEP_GCRF [88]	73.2
CRF-RNN [59]	74.7
POSTECH_DeconvNet_CRF_VOC [61]	74.8
BoxSup [60]	75.2
Context + CRF-RNN [76]	75.3
QO_4^{mres} [66]	75.5
DeepLab-CRF-Attention [17]	75.7
CentraleSuperBoundaries++ [18]	76.0
DeepLab-CRF-Attention-DT [63]	76.3
H-ReNet + DenseCRF [89]	76.8
LRR_4x_COCO [90]	76.8
DPN [62]	77.5
Adelaide_Context [40]	77.8
Oxford_TVG_HO_CRF [91]	77.9
Context CRF + Guidance CRF [92]	78.1
Adelaide_VeryDeep_FCN_VOC [93]	79.1
DeepLab-CRF (ResNet-101)	79.7

TABLE 5: Performance on PASCAL VOC 2012 *test* set. We have added some results from recent arXiv papers on top of the official leadearboard results.