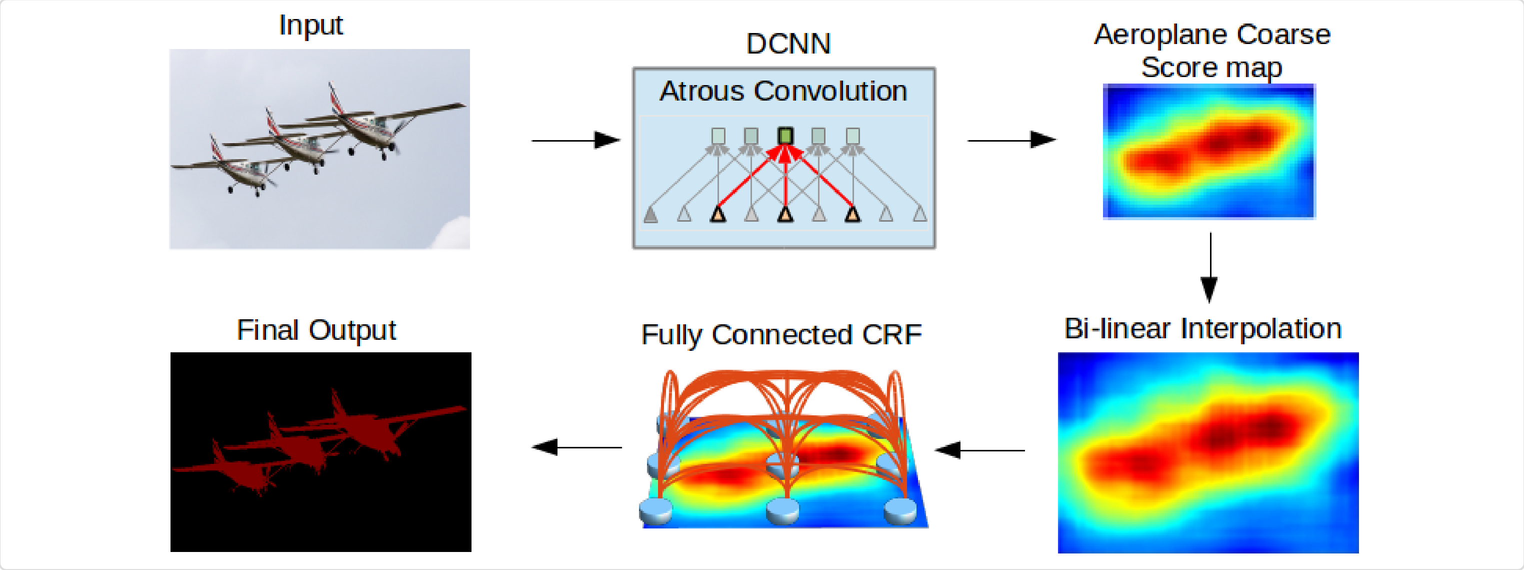
DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs



We address the task of semantic image segmentation with Deep Learning and make three main contributions that are experimentally shown to have substantial practical merit. First, we highlight convolution with upsampled filters, or ‘atrous convolution’, as a powerful tool in dense prediction tasks. Atrous convolution allows us to explicitly control the resolution at which feature responses are computed within Deep Convolutional Neural Networks. It also allows us to effectively enlarge the field of view of filters to incorporate larger context without increasing the number of parameters or the amount of computation. Second, we propose atrous spatial pyramid pooling (ASPP) to robustly segment objects at multiple scales. ASPP probes an incoming convolutional feature layer with filters at multiple sampling rates and effective fields-of-views, thus capturing objects as well as image context at multiple scales. Third, we improve the localization of object boundaries by combining methods from DCNNs and probabilistic graphical models. The commonly deployed combination of max-pooling and downsampling in DCNNs achieves invariance but has a toll on localization accuracy. We overcome this by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF), which is shown both qualitatively and quantitatively to improve localization performance.

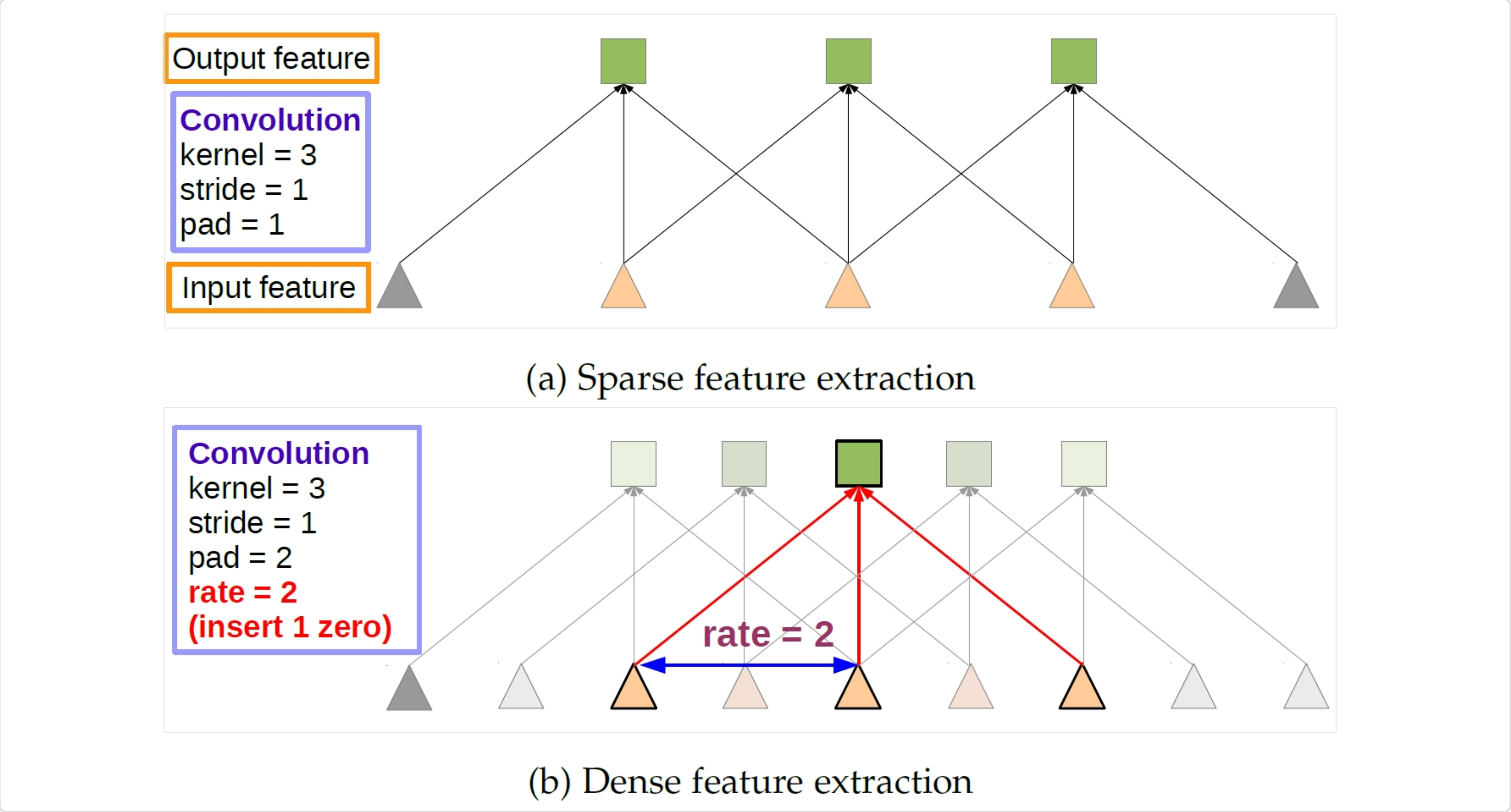
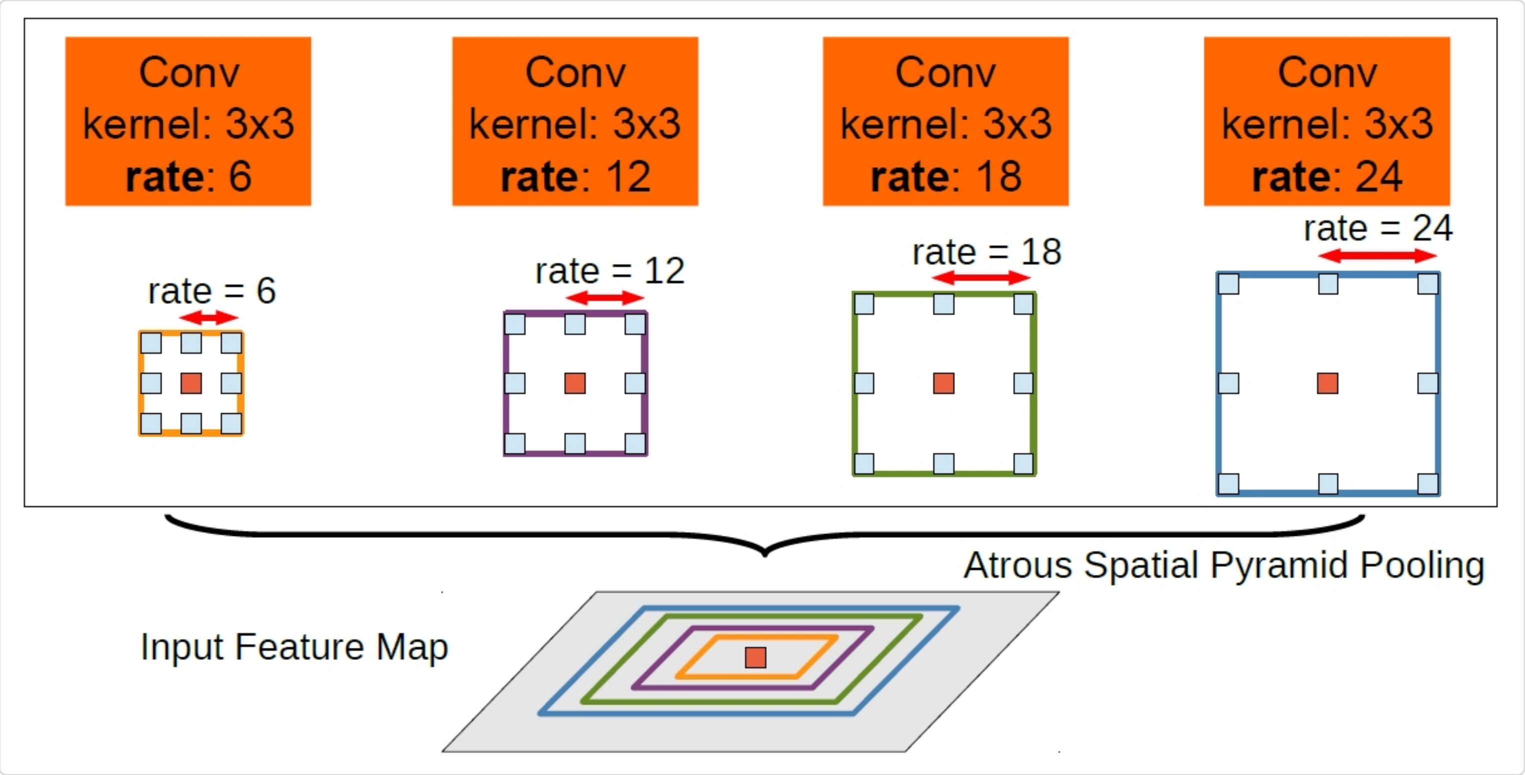
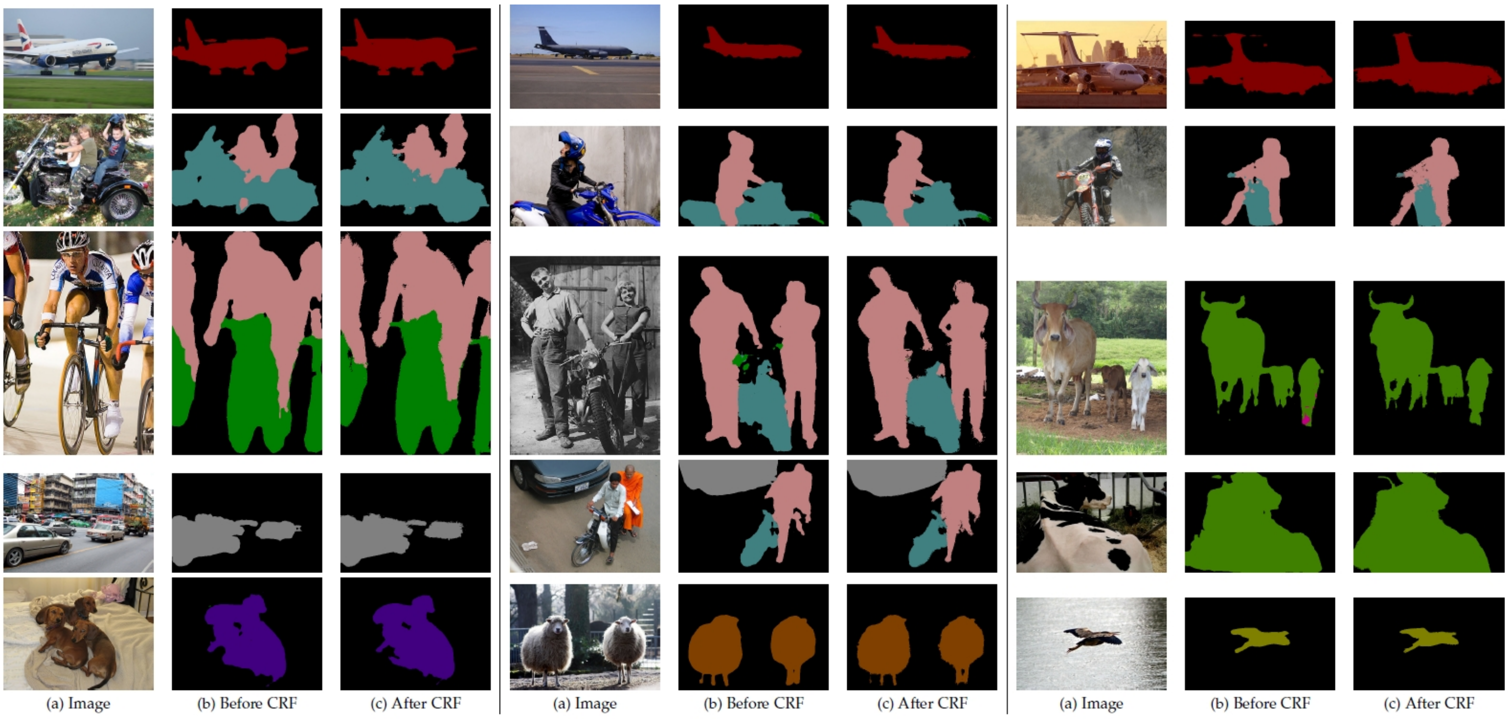


Illustration of dense feature extraction in 1-D. (a) Sparse feature extraction with standard convolution on a low resolution input feature map. (b) Dense feature extraction with atrous convolution with rate r = 2, applied on a high resolution input feature map.



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



(a) Image (b) Before CRF (c) After CRF (a) Image (b) Before CRF (c) After CRF (a) Image (b) Before CRF (c) After CRF

In this figure, we show some PASCAL VOC 2012 segmentation results by our DeepLab before and after CRF.

Download

Code and Dataset

(1) DeepLab v2 Codes used for the latest experiments is available (https://bitbucket.org/aquariusjay/deeplab-public-ver2) now! Note that this version also supports the experiments (DeepLab v1) in our ICLR'15. You only need to modify the old prototxt files. For example, our proposed atrous convolution is called dilated convolution in CAFFE framework, and you need to change the convolution parameter "hole" to "dilation" (the usage is exactly the same). For the experiments in ICCV'15, there are some minor differences between the argmax and softmax_loss layers for DeepLabv1 and DeepLabv2. Please refer to DeepLabv1 (https://bitbucket.org/deeplab/deeplab-public) for details.

(2) All trained models and corresponding prototxt files can be downloaded from this link (http://liangchiehchen.com/projects/DeepLab_Models.html).

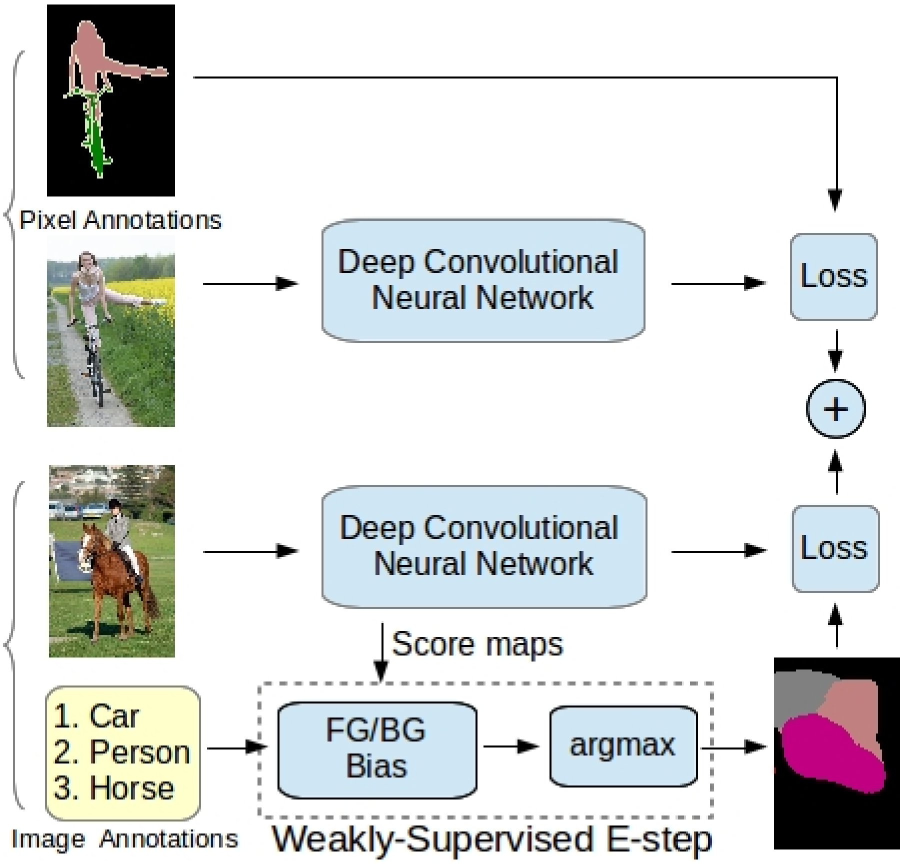
(3) DeepLab v1 Codes used for the experiments (ICLR'15 and ICCV'15) can be downloaded from this link (https://bitbucket.org/deeplab/deeplab-public).

(4) PASCAL-Person-Part dataset can be downloaded from this link (../data/pascal_person_part.zip). Please also refer to the original project website (http://www.stat.ucla.edu/~xianjie.chen/pascal_part_dataset/pascal_part.html).

🡅 back to top

DeepLab has been further extended to several projects, listed below:

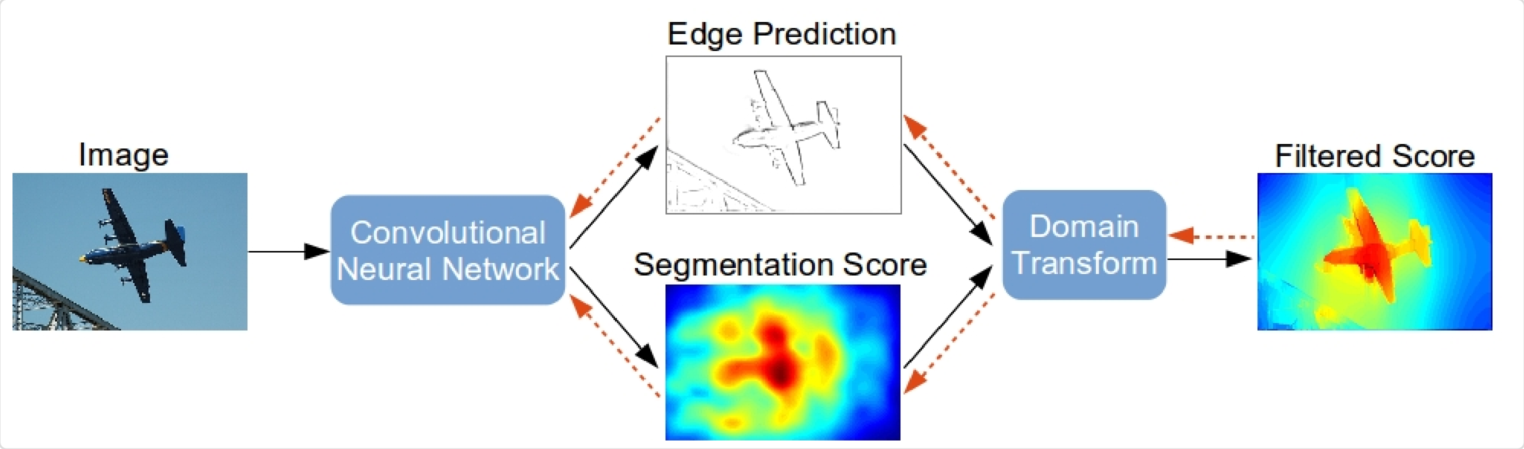
1. Weakly- and Semi-Supervised Learning of a Deep Convolutional Network for Semantic Image Segmentation



Deep convolutional neural networks (DCNNs) trained on a large number of images with strong pixel-level annotations have recently significantly pushed the state-of-art in semantic image segmentation. We study the more challenging problem of learning DCNNs for semantic image segmentation from either (1) weakly annotated training data such as bounding boxes or image-level labels or (2) a combination of few strongly labeled and many weakly labeled images, sourced from one or multiple datasets. We develop Expectation-Maximization (EM) methods for semantic image segmentation model training under these weakly supervised and semi-supervised settings. Extensive experimental evaluation shows that the proposed techniques can learn models delivering competitive results on the challenging PASCAL VOC 2012 image segmentation benchmark, while requiring significantly less annotation effort.

🡅 back to top

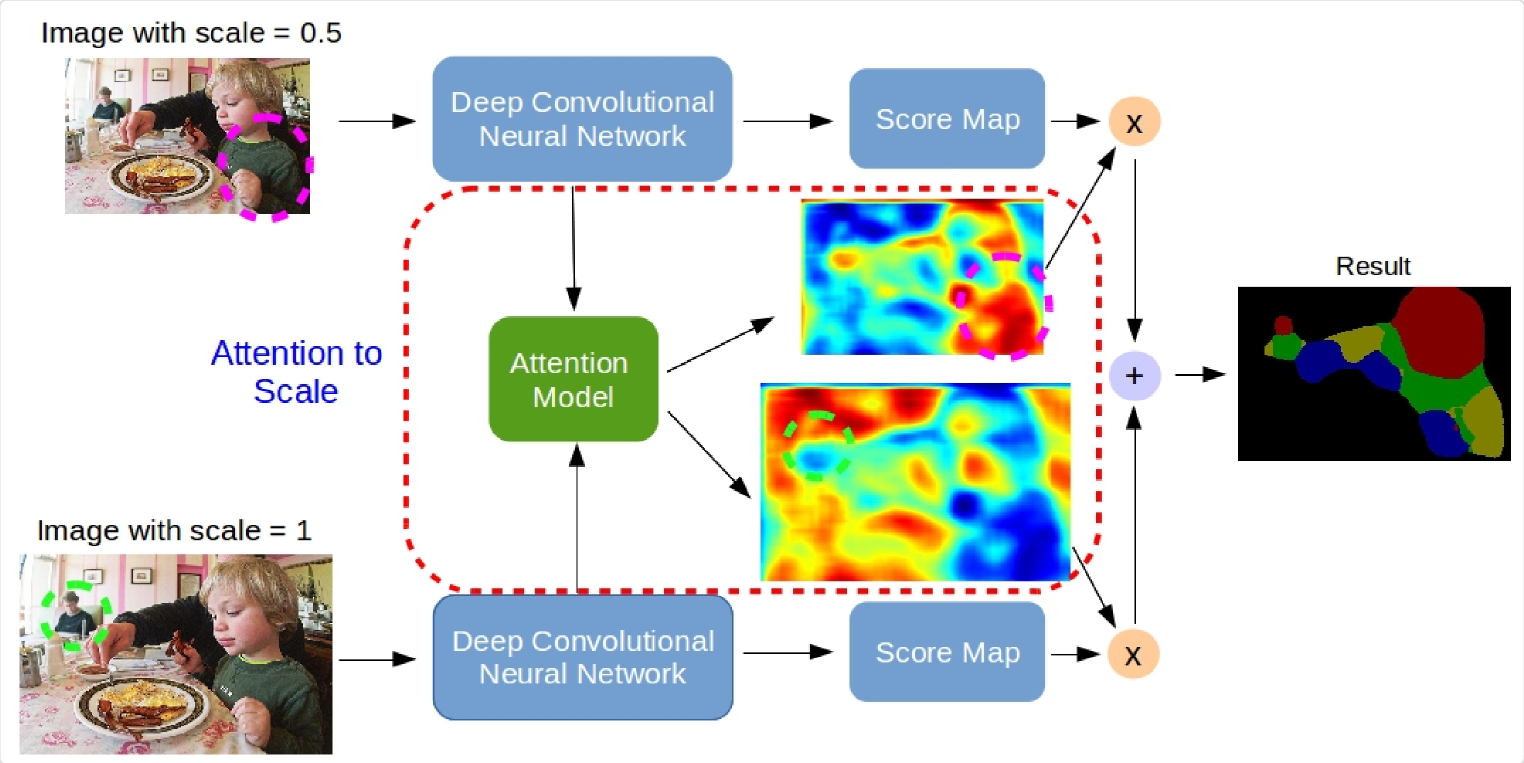
2. Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform



Deep convolutional neural networks (CNNs) are the backbone of state-of-art semantic image segmentation systems. Recent work has shown that complementing CNNs with fully-connected conditional random fields (CRFs) can significantly enhance their object localization accuracy, yet dense CRF inference is computationally expensive. We propose replacing the fully-connected CRF with domain transform (DT), a modern edge-preserving filtering method in which the amount of smoothing is controlled by a reference edge map. Domain transform filtering is several times faster than dense CRF inference and we show that it yields comparable semantic segmentation results, accurately capturing object boundaries. Importantly, our formulation allows learning the reference edge map from intermediate CNN features instead of using the image gradient magnitude as in standard DT filtering. This produces task-specific edges in an end-to-end trainable system optimizing the target semantic segmentation quality.

🠕 back to top

3. Attention to Scale: Scale-aware Semantic Image Segmentation



Incorporating multi-scale features to deep convolutional neural networks (DCNNs) has been a key element to achieve state-of-art performance on semantic image segmentation benchmarks. One way to extract multi-scale features is by feeding several resized input images to a shared deep network and then merge the resulting multi-scale features for pixel-wise classification. In this work, we adapt a state-of-art semantic image segmentation model with multi-scale input images. We jointly train the network and an attention model which learns to softly weight the multi-scale features, and show that it outperforms average- or max-pooling over scales. The proposed attention model allows us to diagnostically visualize the importance of features at different positions and scales. Moreover, we show that adding extra supervision to the output of DCNN for each scale is essential to achieve excellent performance when merging multi-scale features.

↑ back to top

Citation

If you use our code, please consider citing relevant papers:

PDF (http://arxiv.org/pdf/1606.00915)

"DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs"

Liang-Chieh Chen*, George Papandreou*, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille (*equal contribution)

arXiv preprint, 2016

PDF (http://arxiv.org/pdf/1412.7062)

"Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs"

Liang-Chieh Chen*, George Papandreou*, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille (*equal contribution)

International Conference on Learning Representations (ICLR), 2015

If you use the implementation related to **weakly** or **semi- supervised training**, please consider citing:

PDF (http://arxiv.org/pdf/1502.02734)

"Weakly- and Semi-Supervised Learning of a Deep Convolutional Network for Semantic Image Segmentation"

George Papandreou*, Liang-Chieh Chen*, Kevin Murphy, and Alan L. Yuille (*equal contribution)

International Conference on Computer Vision (ICCV), 2015

If you use the implementation related to **discriminatively trained domain transform**, please consider citing:

PDF (http://arxiv.org/pdf/1511.03328)

"Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform"

Liang-Chieh Chen, Jonathan T. Barron, George Papandreou, Kevin Murphy, and Alan L. Yuille

In Conference on Computer Vision and Pattern Recognition (CVPR), 2016

If you use the implementation related to **attention model** or **multi-scaled inputs**, please also consider citing:

PDF (http://arxiv.org/pdf/1511.03339)

"Attention to Scale: Scale-aware Semantic Image Segmentation"

Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L. Yuille

In Conference on Computer Vision and Pattern Recognition (CVPR), 2016

↑ back to top

Acknowledgements

(1) We gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs used for this research.

(2) This work was partly supported by ARO 62250-CS, NIH Grant 5R01EY022247-03, EU Project RECONFIG FP7-ICT-600825 and EU Project MOBOT FP7-ICT-2011-600796.

↑ back to top