



PASCAL VOC Challenge performance evaluation and download server

Home | Leaderboard

Segmentation Results: VOC2012 BETA

Competition "comp6" (train on own data)

This leaderboard shows only those submissions that have been marked as public, and so the displayed rankings should not be considered as definitive.

- The highest scoring entry in each column is shown in **bold**.
- Clicking on the **blue** arrow symbol (▼) at the top of a column will order the submissions from high to low wrt performance on that column.
- Clicking on the **orange** arrow symbol (▶) at the left hand side of a table will identify entries with performance equivalent to the selected submission. For each object class all entries equivalent to the selected submission are marked by orange highlighting.
- gray highlighting with double stars (**) indicate that equivalent entries are not available for this submission yet.

Entries equivalent to a selected submission are determined by bootstrapping the performance measure, and assessing if the differences between the selected submission and the others are not statistically significant (see sec 3.5 in VOC 2014 paper).

Average Precision (AP %)

	mean	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/monitor	submission date
	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
▶ IDW-CNN [?]	86.3	94.8	67.3	93.4	74.8	84.6	95.3	89.6	93.6	54.1	94.9	79.0	93.3	95.5	91.7	89.2	77.5	93.7	79.2	94.0	80.8	30-Jun-2017
▶ CASIA_IVA_SDN [?]	86.0	96.6	76.4	95.6	79.3	81.4	97.0	91.7	96.4	45.5	94.1	78.7	92.4	95.4	92.5	90.5	73.7	93.8	64.9	90.2	83.0	15-Jun-2017
▶ DeepLabv3 [?]	85.7	96.4	76.6	92.7	77.8	87.6	96.7	90.2	95.4	47.5	93.4	76.3	91.4	97.2	91.0	92.1	71.3	90.9	68.9	90.8	79.3	20-Jun-2017
▶ PSPNet [?]	85.4	95.8	72.7	95.0	78.9	84.4	94.7	92.0	95.7	43.1	91.0	80.3	91.3	96.3	92.3	90.1	71.5	94.4	66.9	88.8	82.0	06-Dec-2016
▶ ** ResNet-38_COCO ** [?]	84.9	96.2	75.2	95.4	74.4	81.7	93.7	89.9	92.5	48.2	92.0	79.9	90.1	95.5	91.8	91.2	73.0	90.5	65.4	88.7	80.6	22-Jan-2017
▶ Multipath-RefineNet [?]	84.2	95.0	73.2	93.5	78.1	84.8	95.6	89.8	94.1	43.7	92.0	77.2	90.8	93.4	88.6	88.1	70.1	92.9	64.3	87.7	78.8	17-Jan-2017
▶ Large_Kernel_Matters [?]	83.6	95.3	68.7	94.1	72.6	82.4	96.0	89.3	93.0	47.8	89.6	70.8	89.2	93.3	90.1	91.2	72.0	89.8	67.8	88.9	76.9	16-Mar-2017
▶ TuSimple [?]	83.1	92.1	64.6	94.7	71.0	81.0	94.6	89.7	94.9	45.6	93.7	74.4	92.0	95.1	90.0	88.7	69.1	90.4	62.7	86.4	78.2	09-Nov-2016
▶ ResNet-38_MS [?]	83.1	95.2	72.5	95.1	70.8	78.5	91.7	90.0	92.4	41.9	90.8	73.9	90.6	93.8	90.5	89.5	72.6	89.8	63.2	87.8	79.1	09-Dec-2016
▶ Deep Layer Cascade (LC) [?]	82.7	85.5	66.7	94.5	67.2	84.0	96.1	89.8	93.5	47.2	90.4	71.5	88.9	91.7	89.2	89.1	70.4	89.4	70.7	84.2	79.6	06-Apr-2017
▶ SegModel [?]	81.8	93.6	60.2	93.6	69.1	76.4	96.3	88.2	95.5	37.9	90.8	73.3	91.1	94.3	88.6	88.6	64.8	90.1	63.7	87.3	78.2	23-Aug-2016
▶ HikSeg_COCO [?]	81.4	95.0	64.2	91.5	79.0	78.7	93.4	88.4	94.3	45.8	89.6	65.2	90.6	92.8	88.7	87.5	62.4	88.4	56.4	86.2	75.3	02-Oct-2016
▶ DP_ResNet_CRF [?]	81.0	94.0	59.5	91.8	68.1	75.9	95.2	88.9	93.2	37.7	90.8	70.8	89.2	92.7	87.7	87.9	65.5	90.3	62.6	87.2	75.5	10-Nov-2016
▶ OBP-HJLCN [?]	80.4	92.7	54.8	91.6	68.0	76.9	95.7	89.3	92.6	35.2	89.0	69.3	89.4	92.7	87.9	87.5	66.8	88.5	62.2	86.1	76.2	13-Sep-2016
▶ CentraleSupélec Deep G-CRF [?]	80.2	92.9	61.2	91.0	66.3	77.7	95.3	88.9	92.4	33.8	88.4	69.1	89.8	92.9	87.7	87.5	62.6	89.9	59.2	87.1	74.2	12-Aug-2016
▶ CMT-FCN-ResNet-CRF [?]	80.0	92.5	55.3	92.2	66.0	76.9	95.1	88.6	93.9	35.1	87.6	71.6	89.3	92.8	87.9	88.0	62.0	88.0	59.7	86.1	75.7	02-Aug-2016
▶ DeepLabv2-CRF [?]	79.7	92.6	60.4	91.6	63.4	76.3	95.0	88.4	92.6	32.7	88.5	67.6	89.6	92.1	87.0	87.4	63.3	88.3	60.0	86.8	74.5	06-Jun-2016
▶ LRR_4x_ResNet_COCO [?]	79.3	92.4	45.1	94.6	65.2	75.8	95.1	89.1	92.3	39.0	85.7	70.4	88.6	89.4	88.6	86.6	65.8	86.2	57.4	85.7	77.3	18-Jul-2016
▶ CASIA_SegResNet_CRF_COCO [?]	79.3	93.8	42.2	93.1	68.6	75.3	95.3	88.8	92.5	36.5	84.3	64.2	86.8	87.8	87.5	88.5	69.2	89.7	64.1	86.8	74.6	03-Jun-2016
▶ Adelaide_VeryDeep_FCN_VOC [?]	79.1	91.9	48.1	93.4	69.3	75.5	94.2	87.5	92.8	36.7	86.9	65.2	89.1	90.2	86.5	87.2	64.6	90.1	59.7	85.5	72.7	13-May-2016
▶ LRR_4x_COCO [?]	78.7	93.2	44.2	89.4	65.4	74.9	93.9	87.0	92.0	42.9	83.7	68.9	86.5	88.0	89.0	87.2	67.3	85.6	64.0	84.1	71.5	16-Jun-2016
▶ CASIA_IVA_OASeg [?]	78.3	93.8	41.9	89.4	67.5	71.5	94.6	85.3	89.5	38.1	88.4	64.8	87.0	90.5	84.9	83.3	67.5	86.9	68.1	83.4	74.0	21-May-2016
▶ Oxford_TVG_HO_CRF [?]	77.9	92.5	59.1	90.3	70.6	74.4	92.4	84.1	88.3	36.8	85.6	67.1	85.1	86.9	88.2	82.6	62.6	85.0	56.3	81.9	72.5	16-Mar-2016
▶ Adelaide_Context_CNN_CRF_COCO [?]	77.8	92.9	39.6	84.0	67.9	75.3	92.7	83.8	90.1	44.3	85.5	64.9	87.3	88.8	84.5	85.5	68.1	89.0	62.8	81.2	71.4	06-Nov-2015

▷ CUHK_DPN_COCO [?]	77.5	89.0	61.6	87.7	66.8	74.7	91.2	84.3	87.6	36.5	86.3	66.1	84.4	87.8	85.6	85.4	63.6	87.3	61.3	79.4	66.4	22-Sep-2015
▷ Adelaide_Context_CNN_CRF_COCO [?]	77.2	92.3	38.8	82.9	66.1	75.1	92.4	83.1	88.6	41.8	85.9	62.8	86.7	88.4	84.0	85.4	67.4	88.8	61.9	81.9	71.7	13-Aug-2015
▷ Ladder_DenseNet [?]	77.0	91.4	44.6	86.5	64.6	74.6	92.1	86.2	91.3	37.0	86.0	62.6	85.2	83.5	84.1	85.4	61.5	91.9	59.5	83.6	71.4	26-May-2017
▷ DeepLab-CRF-Attention-DT [?]	76.3	93.2	41.7	88.0	61.7	74.9	92.9	84.5	90.4	33.0	82.8	63.2	84.5	85.0	87.2	85.7	60.5	87.7	57.8	84.3	68.2	03-Feb-2016
▷ CentraleSuperBoundaries++ [?]	76.0	91.1	38.5	90.9	68.7	74.2	89.9	85.3	89.1	34.4	82.5	65.6	83.1	82.9	85.7	85.4	60.6	84.5	59.9	80.2	69.9	13-Jan-2016
▷ LRR_4x_de_pyramid_VOC [?]	75.9	91.8	41.0	83.0	62.3	74.3	93.0	86.8	88.7	36.6	81.8	63.4	84.7	85.9	85.1	83.1	62.0	84.6	55.6	84.9	70.0	07-Jun-2016
▷ DeepLab-CRF-Attention [?]	75.7	91.1	40.9	86.9	62.1	74.2	92.3	84.4	90.1	34.0	81.7	66.0	83.5	83.9	86.5	84.6	59.1	87.2	59.6	81.0	66.2	03-Feb-2016
▷ Adelaide_Context_CNN_CRF_VOC [?]	75.3	90.6	37.6	80.0	67.8	74.4	92.0	85.2	86.2	39.1	81.2	58.9	83.8	83.9	84.3	84.8	62.1	83.2	58.2	80.8	72.3	30-Aug-2015
▷ MSRA_BoxSup [?]	75.2	89.8	38.0	89.2	68.9	68.0	89.6	83.0	87.7	34.4	83.6	67.1	81.5	83.7	85.2	83.5	58.6	84.9	55.8	81.2	70.7	18-May-2015
▷ POSTECH_DeconvNet_CRF_VOC [?]	74.8	90.0	40.8	84.2	67.3	70.7	90.9	84.8	87.4	34.8	83.0	58.7	82.3	87.1	86.9	82.4	64.5	84.6	54.9	77.5	64.1	18-Aug-2015
▷ MERL_UMD_Deep_GCRF_COCO [?]	74.8	89.9	42.6	90.0	65.0	69.2	89.9	83.9	88.2	31.3	81.8	66.4	82.9	81.1	85.7	83.4	58.4	88.4	56.7	77.7	64.3	15-Jan-2016
▷ Oxford_TVG_CRF_RNN_COCO [?]	74.7	90.4	55.3	88.7	68.4	69.8	88.3	82.4	85.1	32.6	78.5	64.4	79.6	81.9	86.4	81.8	58.6	82.4	53.5	77.4	70.1	22-Apr-2015
▷ UNIST_GDN_CRF_ENS [?]	74.0	88.6	48.6	88.8	64.7	70.4	87.2	81.8	86.4	32.0	77.1	64.1	80.5	78.0	84.0	83.3	59.2	85.9	56.8	77.9	65.0	29-Jul-2016
▷ DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint [?]	73.9	89.2	46.7	88.5	63.5	68.4	87.0	81.2	86.3	32.6	80.7	62.4	81.0	81.3	84.3	82.1	56.2	84.6	58.3	76.2	67.2	26-Apr-2015
▷ UNIST_GDN_CRF [?]	73.2	87.9	37.8	88.8	64.5	70.7	87.7	81.3	87.1	32.5	76.7	66.6	80.3	76.6	82.2	82.3	57.9	84.5	55.9	78.5	64.2	29-Jul-2016
▷ MERL_DEEP_GCRF [?]	73.2	85.2	43.9	83.3	65.2	68.3	89.0	82.7	85.3	31.1	79.5	63.3	80.5	79.3	85.5	81.0	60.5	85.5	52.0	77.3	65.1	17-Oct-2015
▷ Bayesian Dilation Network [?]	73.1	88.6	39.0	86.2	63.3	67.1	88.1	81.9	86.8	34.7	81.1	57.1	81.3	86.5	83.4	83.4	53.7	84.0	53.3	80.5	62.5	07-Jun-2016
▷ DeepLab-CRF-COCO-LargeFOV [?]	72.7	89.1	38.3	88.1	63.3	69.7	87.1	83.1	85.0	29.3	76.5	56.5	79.8	77.9	85.8	82.4	57.4	84.3	54.9	80.5	64.1	18-Mar-2015
▷ POSTECH_EDeconvNet_CRF_VOC [?]	72.5	89.9	39.3	79.7	63.9	68.2	87.4	81.2	86.1	28.5	77.0	62.0	79.0	80.3	83.6	80.2	58.8	83.4	54.3	80.7	65.0	22-Apr-2015
▷ Dual-Multi-Reso-MR [?]	72.4	87.6	40.3	80.6	62.9	71.3	88.1	84.4	84.7	29.6	77.8	58.5	80.0	81.0	85.4	82.1	55.0	83.8	48.2	80.3	65.3	03-Nov-2016
▷ CCBM [?]	72.3	87.8	46.7	79.0	63.6	70.5	83.7	75.5	86.9	31.0	81.9	61.3	81.5	85.9	81.1	76.5	58.7	77.7	50.4	76.6	69.8	29-Nov-2015
▷ Oxford_TVG_CRF_RNN_VOC [?]	72.0	87.5	39.0	79.7	64.2	68.3	87.6	80.8	84.4	30.4	78.2	60.4	80.5	77.8	83.1	80.6	59.5	82.8	47.8	78.3	67.1	22-Apr-2015
▷ DeepLab-MSc-CRF-LargeFOV [?]	71.6	84.4	54.5	81.5	63.6	65.9	85.1	79.1	83.4	30.7	74.1	59.8	79.0	76.1	83.2	80.8	59.7	82.2	50.4	73.1	63.7	02-Apr-2015
▷ MSRA_BoxSup [?]	71.0	86.4	35.5	79.7	65.2	65.2	84.3	78.5	83.7	30.5	76.2	62.6	79.3	76.1	82.1	81.3	57.0	78.2	55.0	72.5	68.1	10-Feb-2015
▷ FCN_CLC_MSP [?]	70.8	86.2	40.1	83.9	57.8	64.7	87.9	81.3	85.9	28.3	80.0	61.9	80.7	82.5	79.7	80.2	54.7	81.3	39.3	78.9	59.2	01-Jul-2016
▷ DeepLab-CRF-COCO-Strong [?]	70.4	85.3	36.2	84.8	61.2	67.5	84.6	81.4	81.0	30.8	73.8	53.8	77.5	76.5	82.3	81.6	56.3	78.9	52.3	76.6	63.3	11-Feb-2015
▷ DeepLab-CRF-LargeFOV [?]	70.3	83.5	36.6	82.5	62.3	66.5	85.4	78.5	83.7	30.4	72.9	60.4	78.5	75.5	82.1	79.7	58.2	82.0	48.8	73.7	63.3	28-Mar-2015
▷ DeepSqueeNet_CRF [?]	70.1	85.7	37.4	83.4	59.7	67.8	85.2	79.8	81.4	27.9	72.3	60.4	76.5	78.2	82.7	78.8	57.3	78.6	49.0	77.6	61.0	21-Jul-2016
▷ TTI_zoomout_v2 [?]	69.6	85.6	37.3	83.2	62.5	66.0	85.1	80.7	84.9	27.2	73.2	57.5	78.1	79.2	81.1	77.1	53.6	74.0	49.2	71.7	63.3	30-Mar-2015
▷ RRF-4s [?]	69.4	79.5	57.3	78.7	61.8	64.1	83.9	78.1	80.4	30.0	73.0	59.4	74.3	73.9	80.8	77.9	53.9	76.4	46.1	71.7	63.9	30-Nov-2016
▷ FCN-2s_Dilated_VGG19 [?]	69.0	81.8	37.0	79.5	57.2	67.5	83.8	79.3	83.0	28.5	74.5	57.5	76.0	75.9	79.5	78.6	57.0	77.8	45.3	73.7	63.2	11-Jul-2017
▷ VGG19_FCN [?]	68.1	81.7	35.9	79.8	57.5	66.9	84.1	79.6	80.8	28.2	72.1	53.3	74.0	72.1	78.5	78.2	55.5	76.7	43.4	73.8	65.1	06-Apr-2017
▷ FCN-8s-heavy [?]	67.2	82.4	36.1	75.6	61.5	65.4	83.4	77.2	80.1	27.9	66.8	51.5	73.6	71.9	78.9	77.1	55.3	73.4	44.3	74.0	63.2	06-Feb-2016
▷ DeepLab-CRF-MSc [?]	67.1	80.4	36.8	77.4	55.2	66.4	81.5	77.5	78.9	27.1	68.2	52.7	74.3	69.6	79.4	79.0	56.9	78.8	45.2	72.7	59.3	30-Dec-2014
▷ DeepLab-CRF [?]	66.4	78.4	33.1	78.2	55.6	65.3	81.3	75.5	78.6	25.3	69.2	52.7	75.2	69.0	79.1	77.6	54.7	78.3	45.1	73.3	56.2	23-Dec-2014
▷ DeepSqueeNet [?]	65.7	76.1	34.3	76.4	56.0	62.0	82.7	75.4	78.3	25.6	64.3	58.8	73.3	69.3	79.3	76.7	53.2	72.1	46.2	69.3	59.1	20-Jul-2016
▷ Bayesian FCN [?]	65.4	80.8	34.9	75.2	57.0	64.1	80.9	77.2	78.0	26.4	65.6	44.0	72.6	70.8	78.7	76.8	52.4	71.0	40.4	73.8	61.8	07-Jun-2016
▷ Weak_manifold_CNN [?]	65.3	80.9	32.9	73.2	57.7	63.0	83.9	73.5	76.6	27.0	65.9	52.6	70.9	69.8	73.0	74.9	53.3	70.1	45.4	72.4	62.7	11-Nov-2016
▷ CRF_RNN [?]	65.2	80.9	34.0	72.9	52.6	62.5	79.8	76.3	79.9	23.6	67.7	51.8	74.8	69.9	76.9	76.9	49.0	74.7	42.7	72.1	59.6	10-Feb-2015
▷ UNIST_GDN_FCN_FC [?]	64.4	75.6	31.5	69.2	51.6	62.9	78.8	76.7	78.7	24.6	61.7	60.3	74.5	62.6	76.1	74.3	51.5	70.6	47.3	74.0	58.4	27-Jul-2016
	64.4	81.9	35.1	78.2	57.4	56.5	80.5	74.0	79.8	22.4	69.6	53.7	74.0	76.0	76.6	68.8	44.3	70.2	40.2	68.9	55.3	24-Nov-2014

		University of Coimbra	Jorge Batista (2) Cristian Sminchisescu (1)	extracted on the whole image, then pooled over each object segmentation candidate to generate global region descriptors. We employ a novel second-order pooling procedure, O2P, with two non-linearities: a tangent space mapping and power normalization. The global region descriptors are passed through linear regressors for each category, then labeled segments in each image having scores above some threshold are pasted onto the image in the order of these scores. Learning is performed using an epsilon-insensitive loss function on overlap with ground truth, similar to [2], but within a linear formulation (using LIBLINEAR). comp6: learning uses all images in the segmentation+etection trainval sets, and external ground truth annotations provided by courtesy of the Berkeley vision group. comp5: one model is trained for each category using the available ground truth segmentations from the 2012 trainval set. Then, on each image having no associated ground truth segmentations, the learned models are used together with bounding box constraints, low-level cues and region competition to generate predicted object segmentations inside all bounding boxes. Afterwards, learning proceeds similarly to the fully annotated case. 1. "Semantic Segmentation with Second-Order Pooling", Carreira, Caseiro, Batista, Sminchisescu. ECCV 2012. 2. "Object Recognition by Ranking Figure-Ground Hypotheses", Li, Carreira, Sminchisescu. CVPR 2010.	
BONN_O2PCPMC_FGT_SEG	BONN_O2PCPMC_FGT_SEG	(1) University of Bonn, (2) University of Coimbra, (3) Georgia Institute of Technology, (4) Vienna University of Technology	Joao Carreira(1,2), Adrian Ion(4), Fuxin Li(3), Cristian Sminchisescu(1)	We present a joint image segmentation and labeling model which, given a bag of figure-ground segment hypotheses extracted at multiple image locations and scales using CPMC (Carreira and Sminchisescu, PAMI 2012), constructs a joint probability distribution over both the compatible image interpretations (tilings or image segmentations) composed from those segments, and over their labeling into categories. The process of drawing samples from the joint distribution can be interpreted as first sampling tilings, modeled as maximal cliques, from a graph connecting spatially non-overlapping segments in the bag (Ion, Carreira, Sminchisescu, ICCV2011), followed by sampling labels for those segments, conditioned on the choice of a particular tiling. We learn the segmentation and labeling parameters jointly, based on Maximum Likelihood with a novel Incremental Saddle Point estimation procedure (Ion, Carreira, Sminchisescu, NIPS2011). As meta-features we combine outputs from linear SVRs using novel second order O2P features to predict the overlap between segments and ground-truth objects of each class (Carreira, Caseiro, Batista, Sminchisescu, ECCV2012; see VOC12 entry BONNCMBR_O2PCPMC_LINEAR), bounding box object detectors, and kernel SVR outputs trained to predict the overlap between segments and ground-truth objects of each class (Carreira, Li, Sminchisescu, IJCV 2012). comp6: the O2P SVR learning uses all images in the segmentation+etection trainval sets, and external ground truth annotations provided by courtesy of the Berkeley vision group.	2012-09-23 21:39:35
BONN_O2PCPMC_FGT_SEG	BONN_O2PCPMC_FGT_SEG	(1) University of Bonn, (2) University of Coimbra, (3) Georgia Institute of Technology, (4) Vienna University of Technology	Joao Carreira(1,2), Adrian Ion(4), Fuxin Li(3), Cristian Sminchisescu(1)	Same as before, except tilings non-maximal	2013-08-08 05:54:53
Bayesian Dilation Network	Bayesian Dilation Network	University of Cambridge	Alex Kendall	http://arxiv.org/abs/1511.02680	2016-06-07 08:28:00
Bayesian FCN	Bayesian FCN	University of Cambridge	Alex Kendall	http://mi.eng.cam.ac.uk/projects/segnet/	2016-06-07 08:36:38
Objectness-aware Semantic Segmentation	CASIA_IVA_OASeg	Institute of Automation, Chinese Academy of Sciences	Yuhang Wang, Jing Liu, Yong Li, Jun Fu, Hang Song, Hanqing Lu	We propose an objectness-aware semantic segmentation framework (OA-Seg) consisting of two deep networks. One is a lightweight deconvolutional neural network (Light-DCNN) which obviously decreases model size and convergence time with impressive segmentation performance. The other one is an object proposal network (OPN) used to roughly locate object regions. MSCOCO is used to extend training data and CRF is used as post-processing.	2016-05-21 01:52:15
CASIA_IVA_SDN	CASIA_IVA_SDN	National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences	Jun Fu, Jing Liu, Yuhang Wang, Zhenwei Shen, Zhiwei Fang, Hanqing Lu	We propose a stacked deconvolutional network (SDN) for semantic segmentation. SDN can effectively encode input images and recover the spatial resolution for accurate boundary localization. Dense connections and hierarchical supervision are used. CRF is not employed!	2017-06-15 09:45:39
CASIA_SegResNet_CRF_COCO	CASIA_SegResNet_CRF_COCO	Institute of Automation, Chinese Academy of Sciences	Xinze Chen, Guangliang Cheng, Yinghao Cai	We propose a novel semantic segmentation method, which consists of three parts: a SAR-based data augmentation method, a deeper residual network including three effective techniques and an online hard pixels mining. We combine these three parts to train an end-to-end network.	2016-06-03 09:20:50
CCBM	CCBM	University of Tsinghua	Qiuwei Wang, Chun Yuan, Zhihui Lin, Zhicheng Wang, Xin Qiu	We propose a method combined with convolutional neural network and Conditional Boltzmann Machines for object segmentation, called CCBM, which further utilizes human visual border detection method. We use CNNs to extract features and segment them by improved Conditional Boltzmann Machines. We also use Structured Random Forests based method to detect object border for a better effort. Finally, each superpixel is labelled as output. The proposed method for this submission was trained on VOC 2012 Segmentation training data and a subset of COCO 2014 training data.	2015-11-29 07:26:11
CMT-FCN-ResNet-CRF	CMT-FCN-ResNet-CRF	Intel labs China and Tsinghua University	Libin Wang, Anbang Yao, Jianguo Li, Yurong Chen, Li Zhang?	We propose a novel coupled multi-task FCN. Both VOC 2012 and COCO dataset are used for training, and CRF is applied as post-processing step.	2016-08-02 09:57:05
CRF as RNN	CRF_RNN	University of Oxford	Shuai Zheng; Sadeep Jayasumana; Bernardino Romera-Paredes; Philip Torr	We introduce a new form of convolutional neural network, called CRF-RNN, which expresses a conditional random field (CRF) as a Recurrent Neural Network (RNN). We plug this CRF-RNN network into an existing deep CNN to obtain a system that combines the desirable properties of both CNNs and CRFs. Importantly, our system fully integrates CRF inference with CNNs, making it possible to train the whole system end-to-end with the usual back-propagation algorithm. See the paper: "Conditional Random Fields as Recurrent Neural Networks".	2015-02-10 11:03:16
Deep Parsing Network	CUHK_DPN_COCO	The Chinese University of Hong	Ziwei Liu*, Xiaoxiao Li*,	This work addresses semantic image segmentation by incorporating rich information into Markov Random Field (MRF), including high-order relations and mixture of label contexts. Unlike previous works that optimized MRFs using iterative algorithm, we solve	2015-09-22

		Kong	Ping Luo, Chen Change Loy, Xiaoou Tang	MRF by proposing a Convolutional Neural Network (CNN), namely Deep Parsing Network (DPN), which enables deterministic end-to-end computation in a single forward pass. Specifically, DPN extends a contemporary CNN architecture to model unary terms and additional layers are carefully devised to approximate the mean field algorithm (MF) for pairwise terms. It has several appealing properties. First, different from the recent works that combined CNN and MRF, where many iterations of MF were required for each training image during back-propagation, DPN is able to achieve high performance by approximating one iteration of MF. Second, DPN represents various types of pairwise terms, making many existing works as its special cases. Third, DPN makes MF easier to be parallelized and speeded up in Graphical Processing Unit (GPU). The system used for this submission was trained on augmented VOC 2012 and MS-COCO 2014 training set. Please refer to the paper "Semantic Image Segmentation via Deep Parsing Network" (http://arxiv.org/abs/1509.02634) for further information.	16:52:27
Deep G-CRF (QO) combined with Deeplab-v2	CentraleSupelec Deep G-CRF	CentraleSupelec / INRIA	Siddhartha Chandra & Iasonas Kokkinos	We employ the deep Gaussian CRF Quadratic Optimization formulation to learn pairwise terms for semantic segmentation using the Deeplab-v2-resnet-101 network. Additionally, we use the dense-CRF post-processing to refine object boundaries. This work is an accepted paper at ECCV 2016 and will be presented at the conference. Please refer to our arXiv report here: http://arxiv.org/abs/1603.08358 We will update the report with more details soon.	2016-08-12 11:21:28
"Super-Human" boundaries combined with Deeplab	CentraleSuperBoundaries++	CentraleSupelec / INRIA	Iasonas Kokkinos	We exploit our "super-human" boundary detector with a multi-resolution variant of the Deeplab system (LargeFOV, pre-trained on MSCOCO). The boundary information comes in the form of Normalized Cut eigenvectors used in DenseCRF inference and boundary-dependent pairwise terms, used in Graph-Cut inference. This is an updated version of our earlier submission, using more training rounds and a single-shot training algorithm. Details on the system and our "super human" boundary detector are provided in http://arxiv.org/abs/1511.07386	2016-01-13 16:00:02
DP_ResNet_CRF	DP_ResNet_CRF	(1) Beijing University of Posts and Telecommunications (BUPT); (2) Beijing Moshanghua Tech (DressPlus)	Lu Yang(1, 2), Qing Song(1), Bin Liu(2), Yuhang He(2), Zuoxin Li(2), Xiongwei Xia(2)	Our network is based on ResNet-152, dilation convolution \ data augmentation \ pre-train on coco \ multi scale test are used for this submission. We also use densecrf as post-processing to refine object boundaries.	2016-11-10 12:05:10
Deep Layer Cascade (LC)	Deep Layer Cascade (LC)	The Chinese University of Hong Kong	Xiaoxiao Li, Ziwei Liu, Ping Luo, Chen Change Loy, Xiaoou Tang	We propose a novel deep layer cascade (LC) method to improve the accuracy and speed of semantic segmentation. Unlike the conventional model cascade (MC) that is composed of multiple independent models, LC treats a single deep model as a cascade of several sub-models. Earlier sub-models are trained to handle easy and confident regions, and they progressively feed-forward harder regions to the next sub-model for processing. Convolutions are only calculated on these regions to reduce computations. The proposed method possesses several advantages. First, LC classifies most of the easy regions in the shallow stage and makes deeper stage focuses on a few hard regions. Such an adaptive and 'difficulty-aware' learning improves segmentation performance. Second, LC accelerates both training and testing of deep network thanks to early decisions in the shallow stage. Third, in comparison to MC, LC is an end-to-end trainable framework, allowing joint learning of all sub-models. We evaluate our method on PASCAL VOC and Cityscapes datasets, achieving state-of-the-art performance and fast speed. Please refer to the paper "Not All Pixels Are Equal: Difficulty-aware Semantic Segmentation via Deep Layer Cascade" (https://arxiv.org/abs/1704.01344) for further information.	2017-04-06 14:46:45
DeepLab-CRF	DeepLab-CRF	(1) UCLA (2) Google (3) TTIC (4) ECP / INRIA	Liang-Chieh Chen (1) and George Papandreou (2,3) and Iasonas Kokkinos (4) and Kevin Murphy (2) and Alan L. Yuille (1)	This work brings together methods from Deep Convolutional Neural Networks (DCNNs) and probabilistic graphical models for the task of semantic image segmentation. We show that responses at the final layer of DCNNs are not sufficiently localized for accurate object segmentation. This is due to the very invariance properties that make DCNNs good for high level tasks. We overcome this poor localization property of deep networks by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF). Efficient computation is achieved by (i) careful network re-purposing and (ii) a novel application of the 'hole' algorithm from the wavelet community, allowing dense computation of neural net responses at 8 frames per second on a modern GPU. See http://arxiv.org/abs/1412.7062 for further information.	2014-12-23 02:29:44
DeepLab-CRF-Attention	DeepLab-CRF-Attention	(1) UCLA (2) Baidu	Liang-Chieh Chen (1) and Yi Yang (2) and Jiang Wang (2) and Wei Xu (2) and Alan L. Yuille (1)	This work is the extension of DeepLab-CRF-COCO-LargeFOV (pretrained on MS-COCO) by further incorporating (1) multi-scale inputs (2) extra supervision and (3) attention model. Further information will be provided in an *updated* version of http://arxiv.org/abs/1511.03339 .	2016-02-03 23:10:45
DeepLab-CRF-Attention-DT	DeepLab-CRF-Attention-DT	(1) UCLA (2) Google	Liang-Chieh Chen (1) and Jonathan T. Barron (2) and George Papandreou (2) and Kevin Murphy (2) and Alan L. Yuille (1)	This work is the extension of DeepLab-CRF-Attention by further incorporating a discriminatively trained Domain Transform. Further information will be provided in an *updated* version of http://arxiv.org/abs/1511.03328 .	2016-02-03 23:13:01
DeepLab-CRF-COCO-LargeFOV	DeepLab-CRF-COCO-LargeFOV	(1) Google (2) UCLA	George Papandreou (1) and Liang-Chieh Chen (2) and Kevin Murphy (1) and Alan L. Yuille (2)	Similar to DeepLab-CRF-COCO-Strong, but the network has a larger field-of-view on the image. Further information will be provided in an updated version of http://arxiv.org/abs/1502.02734 .	2015-03-18 04:09:39
DeepLab-CRF-COCO-Strong	DeepLab-CRF-COCO-Strong	(1) Google (2) UCLA	George Papandreou (1) and Liang-Chieh Chen (2) and Kevin Murphy (1) and Alan L. Yuille (2)	Similar to DeepLab-CRF, but network training also included the pixel-level semantic segmentation annotations of the MS-COCO (v. 2014) dataset. See http://arxiv.org/abs/1502.02734 for further information.	2015-02-11 01:44:22

DeepLab-CRF-LargeFOV	DeepLab-CRF-LargeFOV	(1) Google (2) UCLA	George Papandreou (1) and Liang-Chieh Chen (2) and Kevin Murphy (1) and Alan L. Yuille (2)	Similar to DeepLab-CRF, but the network has a larger field-of-view on the image. Note that the model has NOT been fine-tuned on MS-COCO dataset. Further information will be provided in an updated version of http://arxiv.org/abs/1412.7062 .	2015-03-28 17:22:26
DeepLab-CRF-MSc	DeepLab-CRF-MSc	(1) UCLA (2) Google (3) TTIC (4) ECP / INRIA	Liang-Chieh Chen (1) and George Papandreou (2,3) and Iasonas Kokkinos (4) and Kevin Murphy (2) and Alan L. Yuille (1)	Similar to DeepLab-CRF, except that multiscale features (direct connections from intermediate layers to the classifier) are also exploited. Specifically, we attach to the input image and 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 score map is concatenated to the VGG final layer score map. The final score map fed into the softmax layer thus consists of $4,096 + 5 * 128 = 4,736$ channels.	2014-12-30 02:52:40
DeepLab-MSc-CRF-LargeFOV	DeepLab-MSc-CRF-LargeFOV	(1) Google (2) UCLA	George Papandreou (1) and Liang-Chieh Chen (2) and Kevin Murphy (1) and Alan L. Yuille (2)	Similar to DeepLab-MSc-CRF, but the network has a larger field-of-view on the image. Note that the model has NOT been fine-tuned on MS-COCO dataset. Further information will be provided in an updated version of http://arxiv.org/abs/1412.7062 .	2015-04-02 06:57:21
DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint	DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint	(1) Google (2) UCLA	George Papandreou (1) and Liang-Chieh Chen (2) and Kevin Murphy (1) and Alan L. Yuille (2)	Similar to DeepLab-CRF model, but with feature extraction at multiple network levels and large field of view. We jointly train DeepLab on Pascal VOC 2012 and MS-COCO, sharing the top-level network weights for the common classes, using pixel-level annotation in both datasets. Further information will be provided in an updated version of http://arxiv.org/abs/1412.7062 and http://arxiv.org/abs/1502.02734 .	2015-04-26 17:48:09
DeepLabv2-CRF	DeepLabv2-CRF	(1) UCLA (2) Google (3) ECP / INRIA	Liang-Chieh Chen (1,2) and George Papandreou (2) and Iasonas Kokkinos (3) and Kevin Murphy (2) and Alan L. Yuille (1)	DeepLabv2-CRF is based on three main methods. First, we employ convolution with upsampled filters, or 'atrous convolution', as a powerful tool to repurpose ResNet-101 (trained on image classification task) in dense prediction tasks. Atrous convolution allows us to explicitly control the resolution at which feature responses are computed within DCNNs. 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 fully connected Conditional Random Fields (CRFs). See http://arxiv.org/abs/1606.00915 for further information.	2016-06-06 01:59:20
DeepLabv3	DeepLabv3	Google Inc.	Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam	In this work, we revisit atrous convolution, a powerful tool to explicitly adjust filter's field-of-view as well as control the resolution of feature responses computed by Deep Convolutional Neural Networks. We propose to augment our previously proposed Atrous Spatial Pyramid Pooling module, which probes convolutional features at multiple scales, with image-level features encoding global context and further boost performance. See http://arxiv.org/abs/1706.05587 for further information.	2017-06-20 01:59:26
DeepSqueeNet	DeepSqueeNet	Sun Yat-sen University, SYSU	HongPeng Wu, Long Chen, Kai Huang	We propose a method for semantic image segmentation. With equivalent accuracy, smaller DNN architectures offer at least three advantages: (1) Smaller DNNs require less communication across servers during distributed training. (2) Smaller DNNs require less bandwidth to export a new model from the cloud to an au-tonomous car. (3) Smaller DNNs are more feasible to deploy on FPGAs and other hardware with limited memory. To provide all of these advantages, we propose a CNN architecture called DeepSqueeNet to semantic image segmentation. It based on SqueezeNet and VGG16. DeepSqueeNet achieves DeepLab (Based on VGG16) accuracy on semantic image segmentation with 10x fewer parameters.	2016-07-20 13:16:16
DeepSqueeNet_CRF	DeepSqueeNet_CRF	Sun Yat-sen University, SYSU	HongPeng Wu, Long Chen, Kai Huang	We propose a method for semantic image segmentation. With equivalent accuracy, smaller DNN architectures offer at least three advantages: (1) Smaller DNNs require less communication across servers during distributed training. (2) Smaller DNNs require less bandwidth to export a new model from the cloud to an au-tonomous car. (3) Smaller DNNs are more feasible to deploy on FPGAs and other hardware with limited memory. To provide all of these advantages, we propose a CNN architecture called DeepSqueeNet to semantic image segmentation. It based on SqueezeNet and VGG16. DeepSqueeNet achieves DeepLab (Based on VGG16) accuracy on semantic image segmentation with 10x fewer parameters. we add CRF	2016-07-21 12:47:19
Dual Multi-Scale Manifold Ranking Network	Dual-Multi-Reso-MR	Wuhan University	Mi Zhang, Ye Lv, Min Luo, Jiasi Yi	We proposed a multi-scale network which utilize the dilated and non-dilated convolutional network as a dual. In both networks, a manifold ranking optimization method is embedded to optimize in a single stream jointly, i.e. no need to train the unary and pairwise network separately. And such a feedforward network makes it possible to train in an end-to-end fashion and guarantee a global optimal.	2016-11-03 12:27:49
Dilated FCN using VGG19 and Skip Architectures	FCN-2s_Dilated_VGG19	Center for Cognitive Skill Enhancement, Independent University Bangladesh	Sharif Amit Kamran, Ali Shihab Sabbir	The weights were transferred from VGG19 and then the fully connected layers were converted to convolutional layers. Dilated convolution was used instead of vanilla convolution in fc6 layer. The upsampling was done with Stride 2 and the upsampled layers were concatenated in steps using four skip architectures. Pascal VOC2012 training data and SBD training and validation data was used for training in two stages.	2017-07-11 16:57:52
Fully convolutional net	FCN-8s	UC Berkeley	Jonathan Long, Evan Shelhamer, Trevor Darrell	We apply fully convolutional nets end-to-end, pixels-to-pixels for segmentation, rearchitecting nets that have been highly successful in classification. We achieve pixelwise prediction and learning in nets with extensive pooling and subsampling using in-network upsampling layers. Inference and learning are both performed on whole images by dense feedforward computation and backpropagation. With skip layers that combine deep, coarse, semantic information and shallow, fine, appearance information, we produce refined, detailed segmentations. We train our fully convolutional net, FCN-8s, end-to-end for segmentation while taking advantage of recent successes in classification by initializing from parameters adapted from the VGG 16-layer net.	2014-11-12 09:08:39
Fully convolutional net	FCN-8s-heavy	UC Berkeley	Jonathan Long,	We apply fully convolutional nets end-to-end, pixels-to-pixels for segmentation, rearchitecting nets that have been highly	2016-02-

			Evan Shelhamer, Trevor Darrell	successful in classification. We achieve pixelwise prediction and learning in nets with extensive pooling and subsampling using in-network upsampling layers. Inference and learning are both performed on whole images by dense feedforward computation and backpropagation. With skip layers that combine deep, coarse, semantic information and shallow, fine, appearance information, we produce refined, detailed segmentations. We train our fully convolutional net, FCN-8s, end-to-end for segmentation while taking advantage of recent successes in classification by initializing from parameters adapted from the VGG 16-layer net. The network is learned online with high momentum for better optimization.	06 09:57:31
FCN with Cross-layer Concat and Multi-scale Pred	FCN_CLC_MSP	National Tsing Hua University, Taiwan	Tun-Huai Shih, Chiou-Ting Hsu	We replace the original fc layers in VGG-16 with several conv and pool layers to extract hierarchical features (Pool3-5 and additional pool6-8). We then use pool3-8 to generate multi-scale predictions, and aggregate them to derive the dense prediction result. To jointly exploit the information from lower- and higher-level layers when making prediction, we adopt cross-layer concatenation to combine pool features (lower-level) with prediction result of coarser stream (high-level). This makes the predictions of finer streams more robust. We do not adopt any pre- or post- processing steps. The number of parameters is about 36M, while the original FCN is 134M. We train all prediction streams at the same time using VOC additional annotated images (10582 in total), and it takes less than one day to train our FCN model on a single GTX Titan X GPU.	2016-07-01 04:27:14
Weakly sup. segmentation by region scores' pooling	FER_WSSS_REGION_SCORE_POOL	University of Zagreb	Josip Krapac Sinisa Segvic	We address the problem of semantic segmentation of objects in weakly supervised setting, when only image-wide labels are available. We describe an image with a set of pre-trained convolutional features (from layer conv5.4 of 19-layer VGG-E network) and embed this set into a Fisher vector (64 component GMM, diagonal covariance for components, normalization only with inverse of Fisher matrix). We learn a linear classifier (logistic regression), apply the learned classifier on the set of all image regions (efficiently, using integral images), and propagate region scores back to the pixels. Compared to the alternatives the proposed method is simple, fast in inference, and especially in training. The details are described in the conference paper Krapac, Segvic: "Weakly-supervised semantic segmentation by redistributing region scores back to the pixels", GPCR 2016	2016-06-14 15:02:23
HikSeg_COCO	HikSeg_COCO	Hikvision Research Institute	Haiming Sun, Di Xie, Shiliang Pu	We begin with DilatedNet, and add a module which multi-scale features are combined step-wise. The network is able to learn to put different weights to features of different scales. This submission is first trained on COCO training set and validation set, then fine-tuned on PASCAL training set.	2016-10-02 09:16:41
Hypercolumn	Hypercolumn	UC Berkeley	Bharath Hariharan, Pablo Arbelaez, Ross Girshick, Jitendra Malik	Recognition algorithms based on convolutional networks (CNNs) typically use the output of the last layer as feature representation. However, the information in this layer may be too coarse to allow precise localization. On the contrary, earlier layers may be precise in localization but will not capture semantics. To get the best of both worlds, we define the hypercolumn at a pixel as the vector of activations of all CNN units above that pixel. Using hypercolumns as pixel descriptors, we show results on three fine-grained localization tasks: simultaneous detection and segmentation, where we improve state-of-the-art from 49.7 mean AP _r to 60.0, keypoint localization, where we get a 3.3 point boost over and part labeling, where we show a 6.6 point gain over a strong baseline.	2015-04-09 02:01:36
Learning Object Interactions and Descriptions for	IDW-CNN	Sun Yat-sen University; The Chinese University of Hong Kong	Guangrun Wang*, Ping Luo*, Liang Lin, Xiaogang Wang	This work increases segmentation accuracy of CNNs by learning from an Image Descriptions in the Wild (IDW) dataset. Unlike previous image captioning datasets, where captions were manually and densely annotated, images and their descriptions in IDW are automatically downloaded from Internet without any manual cleaning and refinement. An IDW-CNN is proposed to jointly train IDW and existing image segmentation dataset such as Pascal VOC 2012 (VOC).	2017-06-30 00:11:24
Laplacian reconstruction and refinement	LRR_4x_COCO	University of California Irvine	Golnaz Ghiasi, Charless C. Fowlkes	We introduce a CNN architecture that reconstructs high-resolution class label predictions from low-resolution feature maps using class-specific basis functions. Our multi-resolution architecture also uses skip connections from higher resolution feature maps to successively refine segment boundaries reconstructed from lower resolution maps. The model used for this submission is based on VGG-16 and it was trained on augmented PASCAL VOC and MS-COCO data. Please refer to our technical report: Laplacian Reconstruction and Refinement for Semantic Segmentation (http://arxiv.org/abs/1605.02264).	2016-06-16 06:19:08
Laplacian reconstruction and refinement	LRR_4x_ResNet_COCO	University of California Irvine	Golnaz Ghiasi Charless C. Fowlkes	We introduce a CNN architecture that reconstructs high-resolution class label predictions from low-resolution feature maps using class-specific basis functions. Our multi-resolution architecture also uses skip connections from higher resolution feature maps to successively refine segment boundaries reconstructed from lower resolution maps. The model used for this submission is based on ResNet-101 and it was trained on augmented PASCAL VOC and MS-COCO data. Please refer to our technical report: Laplacian Reconstruction and Refinement for Semantic Segmentation (http://arxiv.org/abs/1605.02264).	2016-07-18 19:07:32
Laplacian reconstruction and refinement	LRR_4x_de_pyramid_VOC	University of California Irvine	Charless C. Fowlkes Golnaz Ghiasi	We introduce a CNN architecture that reconstructs high-resolution class label predictions from low-resolution feature maps using class-specific basis functions. Our multi-resolution architecture also uses skip connections from higher resolution feature maps to successively refine segment boundaries reconstructed from lower resolution maps. The model used for this submission was trained on augmented PASCAL VOC. Please refer to our technical report: Laplacian Reconstruction and Refinement for Semantic Segmentation	2016-06-07 03:55:11
ICCV submission, paper ID 3205	Ladder_DenseNet	anonymous	anonymous	ICCV submission paper ID 3205: Ladder-style DenseNets for Semantic Segmentation of Large Natural Images	2017-05-26 13:58:49
Large_Kernel_Matters	Large_Kernel_Matters	Tsinghua University	Peng Chao, Yu Gang, Zhang Xiangyu	We use the large kernel to generate the feature map and score map, resnet101 is applied with COCO, SBD datasets. No CRF or similar post processing methods are employed! No Multiscale	2017-03-16 01:58:16
Deep Gaussian CRF	MERL_DEEP_GCRF	Mitsubishi Electric Research Laboratories	Raviteja Vemulapalli Oncel Tuzel	We use two deep networks, one for generating unary potentials and the other for generating pairwise potentials. Then we use Gaussian CRF model for structured prediction.	2015-10-17 14:55:31
Gaussian CRF on top of Deeplab CNN	MERL_UMD_Deep_GCRF_COCO	University of Maryland, College Park	Raviteja Vemulapalli (UMD) Oncel Tuzel (MERL) Ming-Yu Liu (MERL) Rama Chellappa (UMD)	We use two deep networks, one for generating unary potentials and the other for generating pairwise potentials. Then we use a Gaussian CRF model for structured prediction. The entire model is trained end-to-end.	2016-01-15 05:23:48
Box-Supervision	MSRA_BoxSup	Microsoft Research Asia	Jifeng Dai, Kaiming He, Jian Sun	BoxSup makes use of bounding box annotations to supervise convolutional networks for semantic segmentation. From these boxes, we estimate segmentation masks with the help of region proposals. These masks are used to update the convolutional network, which is in turn fed back to mask estimation. This procedure is iterated. This result is achieved by semi-supervised training on the segmentation masks from PASCAL VOC and a large amount of bounding boxes from Microsoft COCO. See http://arxiv.org/abs/1503.01640 for details.	2015-02-10 09:35:40
MSRA_BoxSup	MSRA_BoxSup	Microsoft Research	Jifeng Dai,	This is an implementation of "BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic	2015-05-

		Asia	Kaiming He, Jian Sun	Segmentation". We train a BoxSup model using the union set of VOC 2007 boxes, COCO boxes, and the augmented VOC 2012 training set. See http://arxiv.org/abs/1503.01640 for details.	18 09:42:54
Convolutional Feature Masking	MSRA_CFM	Microsoft Research Asia	Jifeng Dai, Kaiming He, Jian Sun	The method exploits shape information via "masking" convolutional features. The proposal segments (e.g., super-pixels) are treated as masks on the convolutional feature maps. The CNN features of segments are directly masked out from these maps and used to train classifiers for recognition. Competitive accuracy and compelling computational speed are demonstrated by the proposed method. We achieve this result by utilizing segment proposal generated by Multi-scale Combinatorial Grouping (MCG), and initializing network parameters from the VGG 16-layer net. See http://arxiv.org/abs/1412.1283 for details.	2014-12-17 02:56:52
Multipath-RefineNet	Multipath-RefineNet	The University of Adelaide; ACRV;	Guosheng Lin; Anton Milan; Chunhua Shen; Ian Reid;	Please refer to our technical report for details: "RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation" (https://arxiv.org/abs/1611.06612). Our source code is available at: https://github.com/guosheng/refinenet	2017-01-17 18:03:57
Unified Object Detection and Semantic Segmentation	NUS_UDS	NUS	Jian Dong, Qiang Chen, Shuicheng Yan, Alan Yuille	Motivated by the complementary effect observed from the typical failure cases of object detection and semantic segmentation, we propose a unified framework for joint object detection and semantic segmentation [1]. By enforcing the consistency between final detection and segmentation results, our unified framework can effectively leverage the advantages of leading techniques for these two tasks. Furthermore, both local and global context information are integrated into the framework to better distinguish the ambiguous samples. By jointly optimizing the model parameters for all the components, the relative importance of different component is automatically learned for each category to guarantee the overall performance. [1] Jian Dong, Qiang Chen, Shuicheng Yan, Alan Yuille: Towards Unified Object Detection and Semantic Segmentation. ECCV 2014	2014-10-29 16:07:10
Joint a network to guided and masking	OBP-HJLCN	national central university	Jia-Ching Wang, Chien-Yao Wang, Jyun-Hong Li	We proposed a hierarchical joint guided networks which has ability to predict objects greater and finer. We also proposed a novel way to guided segmentation by object and boundary.	2016-09-13 15:21:45
Oxford_TVG_CRF_RNN_COCO	Oxford_TVG_CRF_RNN_COCO	[1] University of Oxford / [2] Baidu IDL	Shuai Zheng [1]; Sadeep Jayasumana [1]; Bernardino Romera-Paredes [1]; Chang Huang [2]; Philip Torr [1]	We introduce a new form of convolutional neural network, called CRF-RNN, which expresses Dense Conditional Random Fields (Dense CRF) as a Recurrent Neural Network (RNN). We plug this CRF-RNN network into an existing deep CNN to obtain a system that combines the desirable properties of both CNNs and CRFs. Importantly, our system fully integrates CRF inference with CNNs, making it possible to train the whole system end-to-end with the usual back-propagation algorithm. The system used for this submission was trained on VOC 2012 Segmentation challenge train data, Berkeley augmented data and a subset of COCO 2014 train data. More details will be available in the paper http://arxiv.org/abs/1502.03240 .	2015-04-22 11:26:57
Oxford_TVG_CRF_RNN_VOC	Oxford_TVG_CRF_RNN_VOC	[1] University of Oxford / [2] Baidu IDL	Shuai Zheng [1]; Sadeep Jayasumana [1]; Bernardino Romera-Paredes [1]; Chang Huang [2]; Philip Torr [1]	We introduce a new form of convolutional neural network, called CRF-RNN, which expresses Dense Conditional Random Fields (Dense CRF) as a Recurrent Neural Network (RNN). We plug this CRF-RNN network into an existing deep CNN to obtain a system that combines the desirable properties of both CNNs and CRFs. Importantly, our system fully integrates CRF with CNNs, making it possible to train the whole system end-to-end with the usual back-propagation algorithm. The system used for this submission was trained on VOC 2012 Segmentation challenge train data, and Berkeley augmented data (COCO dataset was not used). More details will be available in the paper http://arxiv.org/abs/1502.03240 .	2015-04-22 10:24:43
Higher Order CRF in CNN	Oxford_TVG_HO_CRF	University of Oxford	Anurag Arnab Sadeep Jayasumana Shuai Zheng Philip Torr	We integrate a conditional random field with higher order potentials into a deep neural network. Our higher order potentials are based on object detector outputs and superpixel oversegmentation, and formulated such that their corresponding mean-field updates are differentiable. For further details, please refer to http://arxiv.org/abs/1511.08119	2016-03-16 21:12:47
POSTECH_DeconvNet_CRF_VOC	POSTECH_DeconvNet_CRF_VOC	POSTECH (Pohang University of Science and Technology)	Hyeonwoo Noh, Seunghoon Hong, Bohyung Han.	We propose a novel semantic segmentation algorithm by learning a deconvolution network. Our deconvolution network is composed of deconvolution and unpooling layers, which identify pixel-wise class labels and predict segmentation masks. The trained network is applied to each proposal in an input image, and the final semantic segmentation map is constructed by combining the results from all proposals in a simple manner. The proposed algorithm mitigates the limitations of the existing methods based on fully convolutional networks; our segmentation method typically identifies more detailed structures and handles objects in multiple scales more naturally. Our network demonstrates outstanding performance in PASCAL VOC 2012 dataset without external training data. See http://arxiv.org/abs/1505.04366 for details.	2015-08-18 18:42:18
POSTECH_EDeconvNet_CRF_VOC	POSTECH_EDeconvNet_CRF_VOC	POSTECH(Pohang University of Science and Technology)	Hyeonwoo Noh, Seunghoon Hong, Bohyung Han	We propose a novel semantic segmentation algorithm by learning a deconvolution network. Our deconvolution network is composed of deconvolution and unpooling layers, which identify pixel-wise class labels and predict segmentation masks. The trained network is applied to each proposal in an input image, and the final semantic segmentation map is constructed by combining the results from all proposals in a simple manner. The proposed algorithm mitigates the limitations of the existing methods based on fully convolutional networks; our segmentation method typically identifies more detailed structures and handles objects in multiple scales more naturally. Our network demonstrates outstanding performance in PASCAL VOC 2012 dataset without external training data.	2015-04-22 21:33:03
PSPNet	PSPNet	CUHK, SenseTime	Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia	Scene parsing is challenging for unrestricted open vocabulary and diverse scenes. In this paper, we exploit the capability of global context information by different-region-based context aggregation through our pyramid pooling module together with the proposed pyramid scene parsing network (PSPNet). Our global prior representation is effective to produce good quality results on the scene parsing task, while PSPNet provides a superior framework for pixel-level prediction tasks. The proposed approach achieves state-of-the-art performance on various datasets. It came first in ImageNet scene parsing challenge 2016, PASCAL VOC 2012 benchmark and Cityscapes benchmark. A single PSPNet yields new record of mIoU score as 85.4% on PASCAL VOC 2012 and 80.2% on Cityscapes. https://arxiv.org/abs/1612.01105	2016-12-06 02:22:13
Residual Forest classifier with FCN features	RRF-4s	Monash University	Yan Zuo, Tom Drummond	We replace the solver component of FCN with a Random Residual Forest (RRF) Classifier and treat FCN as a generic feature extractor to train the RRF classifier	2016-11-30 23:31:43
ResNet-38 with COCO	ResNet-38_COCO	The University of Adelaide	Zifeng Wu, Chunhua Shen, Anton van den Hengel	Pre-trained with COCO, and tested with multiple scales. See our report https://arxiv.org/abs/1611.10080 for details.	2017-01-22 04:44:14
ResNet-38 Multi-scale	ResNet-38_MS	The University of Adelaide	Zifeng Wu, Chunhua Shen,	Single model; multi-scale testing; NO COCO; NO CRF-based post-processing. For more details, refer to our report https://arxiv.org/abs/1611.10080 and code https://github.com/itijoy/ademxapp .	2016-12-09

Anton van den Hengel

12:19:24

SDS	SDS	UC Berkeley	Bharath Hariharan Pablo Arbelaez Ross Girshick Jitendra Malik	We aim to detect all instances of a category in an image and, for each instance, mark the pixels that belong to it. We call this task Simultaneous Detection and Segmentation (SDS). Unlike classical bounding box detection, SDS requires a segmentation and not just a box. Unlike classical semantic segmentation, we require individual object instances. We build on recent work that uses convolutional neural networks to classify category-independent region proposals (R-CNN [1]), introducing a novel architecture tailored for SDS. We then use category-specific, top-down figure-ground predictions to refine our bottom-up proposals. We show a 7 point boost (16% relative) over our baselines on SDS, a 4 point boost (8% relative) over state-of-the-art on semantic segmentation, and state-of-the-art performance in object detection. Finally, we provide diagnostic tools that unpack performance and provide directions for future work.	2014-07-21 22:46:22
SegModel	SegModel	Peking Univerisity	Falong Shen, Peking University	Deep fully convolutional networks with conditional random field. Trained on MSCOCO trainval set and Pascal VOC 12 train set.	2016-08-23 04:04:21
SegNet	SegNet	University of Cambridge	Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla	SegNet is a memory efficient real time deep convolutional encoder-decoder architecture. For more information, please see our publications and web demo at: http://mi.eng.cam.ac.uk/projects/segnet/	2015-11-10 09:48:12
Diverse M-Best with discriminative reranking	TTIC-divmbest-rerank	(1) Toyota Technological Institute at Chicago, (2) Virginia Tech	Payman Yadollahpour (1), Dhruv Batra (1,2), Greg Shakhnarovich (1)	We generate a set of M=10 full image segmentations using Diverse M-Best algorithm from [BYGS'12], applied to inference in the O2P model (Carreira et al., 2012). Then we discriminatively train a reranker based on a novel set of features. The learning of the reranker uses relative loss, with the objective to minimize gap with the oracle (the hindsight-best of the M segmentations), and relies on slack-rescaling structural SVM. The details are described in [YBS'13]. References: [BYGS'12] Batra, Yadollahpour, Guzman, Shakhnarovich, ECCV 2012. [YBS'13] Yadollahpour, Batra, Shakhnarovich, CVPR 2013.	2012-11-15 04:03:01
Feedforward segmentation with zoom-out features	TTI_zoomout	TTI-Chicago	Mohammadreza Mostajabi, Payman Yadollahpour, Gregory Shakhnarovich	Our method uses a feedforward network to directly label superpixels. For each superpixel we use features extracted from a nested set of "zoom-out" regions, from purely local to image-level.	2014-11-17 04:57:49
Feedforward segmentation with zoom-out features	TTI_zoomout_16	TTI-Chicago	Mohammadreza Mostajabi, Payman Yadollahpour, Gregory Shakhnarovich	Same as before, except using VGG 16-layer network instead of VGG CNN-S network. Fine-tuning on VOC-2012 was not performed. See http://arxiv.org/abs/1412.0774 for details.	2014-11-24 08:54:05
Feedforward semantic segmentation with zoom-out features	TTI_zoomout_v2	TTI-Chicago	Mohammadreza Mostajabi, Payman Yadollahpour, Gregory Shakhnarovich	Similar to TTI_zoomout_16, except the way that we set the number and scope of zoom-out levels. In this version, zoom-out levels correspond to receptive field sizes of different layers in a convolutional neural network. Our model is trained only on VOC-2012. Details are provided in our CVPR 2015 paper available at http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Mostajabi_Feedforward_Semantic_Segmentation_2015_CVPR_paper.pdf .	2015-03-30 18:40:04
TuSimple	TuSimple	UC San Diego, CMU, UIUC, TuSimple	Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, Garrison Cottrell	We use ResNet 152, COCO Data, and denseCRF	2016-11-09 01:10:04
Global Deconvolutional Network with CRF	UNIST_GDN_CRF	Ulsan National Institute of Science and Technology (UNIST)	Vladimir Nekrasov, Janghoon Ju, Jaesik Choi	We propose a novel architecture to conduct the deconvolution operation and acquire dense predictions, and an additional loss function, which forces the network to boost the recognition accuracy. Our model shows superior performance over baseline DeepLab-CRF. Further details are provided in "Global Deconvolutional Networks", http://arxiv.org/abs/1602.03930 .	2016-07-29 07:23:03
Global Deconvolutional Network with CRF	UNIST_GDN_CRF_ENS	Ulsan National Institute of Science and Technology (UNIST)	Vladimir Nekrasov, Janghoon Ju, Jaesik Choi	We propose a novel architecture to conduct the deconvolution operation and acquire dense predictions, and an additional loss function, which forces the network to boost the recognition accuracy. Our model shows superior performance over baseline DeepLab-CRF. Further details are provided in "Global Deconvolutional Networks", http://arxiv.org/abs/1602.03930 .	2016-07-29 07:25:56
Global Deconvolutional Network	UNIST_GDN_FCN	Ulsan National Institute of Science and Technology (UNIST)	Vladimir Nekrasov, Janghoon Ju, Jaesik Choi	We propose a novel architecture to conduct the deconvolution operation and acquire dense predictions, and an additional loss function, which forces the network to boost the recognition accuracy. Our model shows superior performance over baseline FCN-32s. Further details are provided in "Global Deconvolutional Networks", http://arxiv.org/abs/1602.03930 .	2016-07-27 01:39:17
Global Deconvolutional Network	UNIST_GDN_FCN_FC	Ulsan National Institute of Science and Technology (UNIST)	Vladimir Nekrasov, Janghoon Ju, Jaesik Choi	We propose a novel architecture to conduct the deconvolution operation and acquire dense predictions, and an additional loss function, which forces the network to boost the recognition accuracy. Besides that, we append a fully-connected layer after the down-sampled image to refine current predictions. Our model shows superior performance over baseline FCN-32s and even outperforms more powerful multi-scale variant. Further details are provided in "Global Deconvolutional Networks", http://arxiv.org/abs/1602.03930 .	2016-07-27 01:49:02
Fully convolutional neural net using VGG19	VGG19_FCN	-	Sharif Amit Kamran , Md. Asif Bin Khaled , Sabit Bin Kabir , Dr. Hasan Muhammad , Moin Mostakim	We use VGG-19 classification neural net and then make it fully convolutional. Moreover, we use skip architectures by concatenating upsampled pool 1 to 4 with the score layer to get finer features. Training was done on two stages, first on Pascal VOC training dataset , secondly on both SBD training plus validation datasets.	2017-04-06 23:22:53

CNN segmentation based on manifold learning	Weak_manifold_CNN	University of Central Florida	Marzieh Edraki	CNN manifold learning for segmentation	2016-11-11 23:34:20
---	-------------------	-------------------------------	----------------	--	---------------------
