Combining Bottom-Up, Top-Down, and Smoothness Cues for Weakly Supervised Image Segmentation

Anirban Roy and Sinisa Todorovic

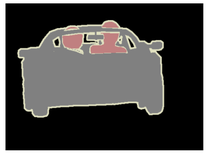
Oregon State University

CVPR July 21-26 2017 HONOLULU

Problem: Weakly Supervised Segmentation

❖ Goal: Pixel-wise prediction from image-level tags

Input Image



Fully supervised ground truth Expensive!

1. Car
2. Person

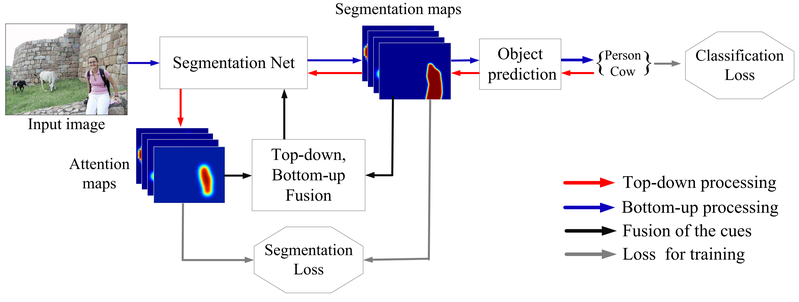
Weakly supervised ground truth Free!

Prediction

Key Idea: Attention Based Localization

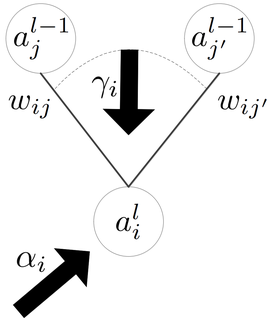
❖ Attention maps can be used as approximate ground-truth

Our Framework

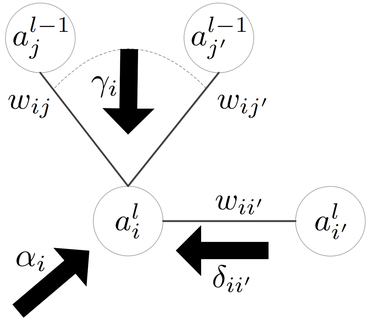


Combining top-down attention, bottom-up segmentation maps and smoothness cues for weakly supervised segmentation

Attention Maps Considering Smoothness



Prior work: Top-down attention computation



Our Approach: Top-down attention computation using additional smoothness cues

α_i : Bottom-up activation of i $\quad$ γ_i : Top-down attention to i

$\delta_{ii'}$: Lateral influence on i

Rectified-Gaussian Formulation of Attention

$$\max_{\boldsymbol{p} \succeq \mathbf{0}} \; \frac{1}{2}\boldsymbol{p}^T D\boldsymbol{p} + \boldsymbol{b}^\top \boldsymbol{p}$$

Attention — Pair-wise smoothness cues activation of i — Top-down attention

$$D = D^l = [\delta_{ii'}^l] \qquad \boldsymbol{b}^l = \gamma_i^l + \alpha_i^l$$

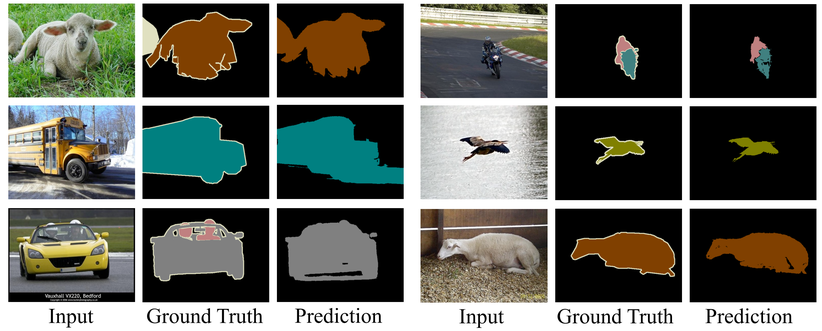
Results on the PASCAL VOC 2012

Method	Pascal validation	Pascal test
w/o attention cues	30.5	31.6
w/o smoothness cues	51.3	52.1
Fully supervised	73.0	75.0
Our full approach	**52.8**	**53.7**

Comparison with the baseline approaches in terms of mIOU (%)

Method	Pascal validation	Pascal test
MIL+ILP [Pinheiro et al., 15]	36.6	35.8
EM [Papandreou et al., 15]	33.8	-
CCNN [Pathak et al., 15]	35.6	35.6
DSCM [Shimoda et al., 16]	44.1	45.1
F-B [Saleh et al., 16]	46.6	48.0
SEC [Kolesnikov et al., 16]	50.7	51.7
Ours	**52.8**	**53.7**

Comparison with the state-of-the-art approaches in terms of mIOU (%)



Acknowledgment: DARPA XAI and NSF RI1302700.