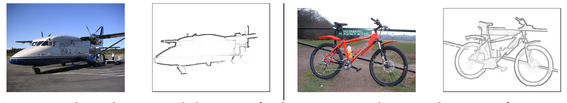
USC Viterbi School of Engineering

Semantic boundary refinement by Joint Inference from edges and regions

Chao "Harry" Yang University of Southern California

Detecting Boundaries for Specific Objects



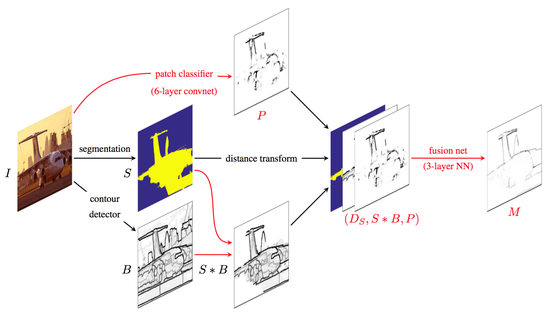
We study the problem of detecting boundaries for specific classes of objects. Our approach leverages recent advances in semantic segmentation and bottom-up boundary detection. We propose a mechanism for combining multiple sources of information: predicted segmentation masks, bottom-up contours, and a novel local class-specific boundary detector. These are jointly mapped to final category-specific boundary strength estimate by a trained classifier. In experiments on VOC2010 and Microsoft COCO data sets, our method dramatically outperforms recent prior work, for some classes doubling the accuracy of boundary prediction.

The Pipeline

Our system of category-specific boundary detection consists of the following components:

- Semantic segmentation (DeepLab).
- Contour detector (UCM).
- Local boundary detector (ConvNets)

A fusion component is used at the end to combine the outputs of the three components. The pipeline is illustrated in the following graph:



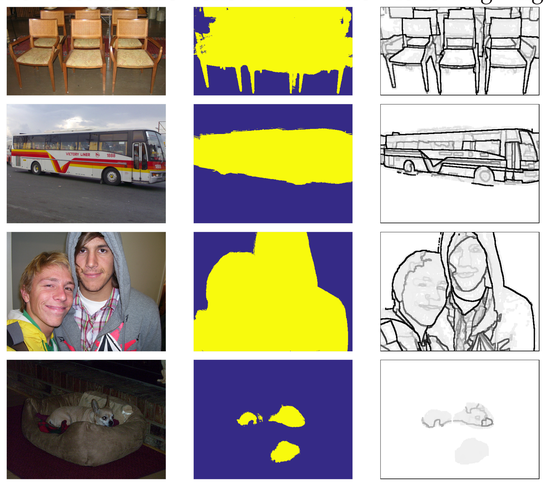
Our experiments use two datasets: VOC 2012 and COCO.

Segmentation and Masked Boundary Baseline

Segmentation and masked boundary are the first two components of our system and are also used as the baseline.

- Segmentation: we use the semantic segmentation contour (DeepLab) as the boundary.
- Masked boundary: We intersect class agnostic boundary detection (UCM) with semantic segmentation mask (original segmentation dilated by 5 pixels).

The results are shown in the following figure:



input seg masked bdry

Local Class-Specific Boundary Detector

We can further improve the results by training a class-specific boundary predictor from scratch. We implement it as a multi-layer convnet, trained to classify image patches as boundary or non-boundary for the class at hand:

	layer type	RF size	#units	stride
1	conv	4	96	1
2	max-pool	2		2
3	conv	3	256	1
4	max-pool	2		2
5	conv	4	64	1
7	ip		256	
8	ip		256	
9	ip		2	

Information Fusion

The three components described above each capture a different aspect of visual information.

- Segmentation: may offer poor localization of the boundaries.
- Masked boundary: low precision.
- local detector: based on limited view of the image due to its local nature.

We use a non-linear classifier to combine them:

$$\log \Pr(m = 1|\mathbf{x}) \propto \mathbf{w}_3^T \sigma(\mathbf{w}_2^T \sigma(\mathbf{w}_1^T \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + b_3,$$

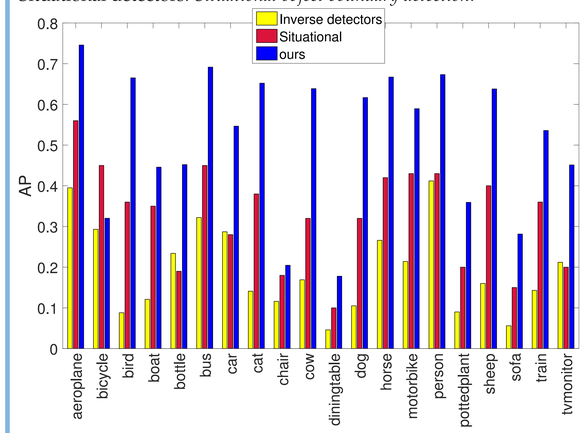
It can be viewed as a three-layer feedforward network with fully connected layers and sigmoid activation functions!

Numerical Results on VOC2012

1. Per-class accuracy comparing with STOA:

Inverse detectors: *Semantic contours from inverse detectors.*

Situational detectors: *Situational object boundary detection.*



2. Mean average precision:

Method	Mean precision
Inverse detectors	19.9
Situational detectors	31.6
Segmentation only S	45.8
Masked boundary $S * B$	46.0
Local patch classifier P	26.9
Fusion $(D_s, S * B, P)$	51.8

We can see that our simple baseline already out-performs the inverse detectors and situational detectors by a large margin.

Visual Results on VOC2012

Numerical Results on Microsoft COCO

we used the first 5,000 images from the `val` set as the test set for our experiments.

Method	Mean precision
Segmentation only S	36.9
Masked boundary $S * B$	38.5
Local patch classifier P	22.4
Fusion $(D_s, S * B, P)$	45.0

The relative standing reflects that in VOC results, but all the numbers are lower, reflecting the increased difficulty of COCO compared to VOC.

Visual Results on Microsoft COCO

Class Agnostic Results

We can also max-pooling the predictions of our category-specific detectors for each pixel, forming a class-agnostic boundary detection map: