Learning Informative Edge Maps for Indoor Scene Layout Prediction

Arun Mallya and Svetlana Lazebnik

Large-scale Scene Understanding Challenge

The submission site will be open for online evaluation for regular submissions after the challenge.
We will organize LSUN 2016 challenge.
Never be too late, and start preparing your submission today!
Add yourself to our Google Group to receive the news.

Introduction

PASCAL VOC and ImageNet ILSVRC challenges have enabled significant progress for object recognition in the past decade. We plan to borrow this mechanism to speed up the progress for scene understanding as well. Complementary to the object-centric ImageNet ILSVRC Challenge hosted at ICCV/ECCV every year, we are hosting a scene-centric challenge at CVPR every year. Our challenge focuses on four major tasks in scene understanding, including scene classification, saliency prediction, room layout estimation, and caption generation (hosted by MS COCO). Inspired by recent success using big data, such as deep learning, we will focus on providing benchmarks that are at least several times bigger than the existing ones, to support training these data-hungry algorithms. By providing a set of large-scale benchmarks in an annual challenge format, we expect significant progress to be made for scene understanding in the coming years.

Layout Prediction

Task:

Predict layout of inside room (with clutter) given 2-D image.

Figure 2. Examples of training images with their respective groundtruth informative edge maps. Orange, purple, and yellow lines indicate wall/wall, wall/ceiling, and wall/floor edges respectively. Blue indicates background containing uninformative edges. Note that these edges are marked by ignoring the occluding clutter, as if the room were empty.

4 steps

1. Predict informative edge maps
2. Detect vanishing points
3. Sample box layouts
4. Rank the layout by ranking-svm

Informative Edge Prediction

Structured Forest Edge Maps

Structured Forests were introduced by Dollar [1] for generating contour maps. It achieved remarkable performance on BSDS500 segmentation dataset.

Classifier: random forest - 8 trees with depth 64

Features: color, gradient features, location features within 32x32 patches

[1] P. Dollar and C. L. Zitnick. Structured forests for fast edge detection. In ICCV, 2013

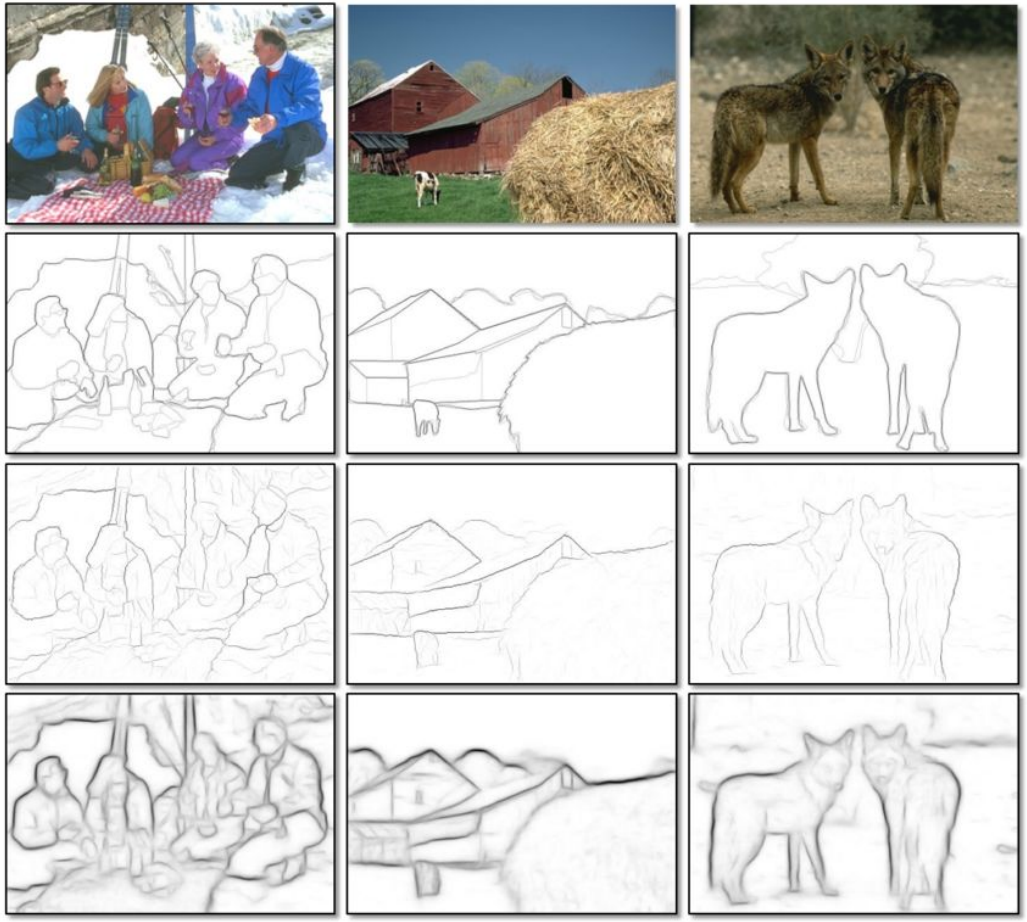


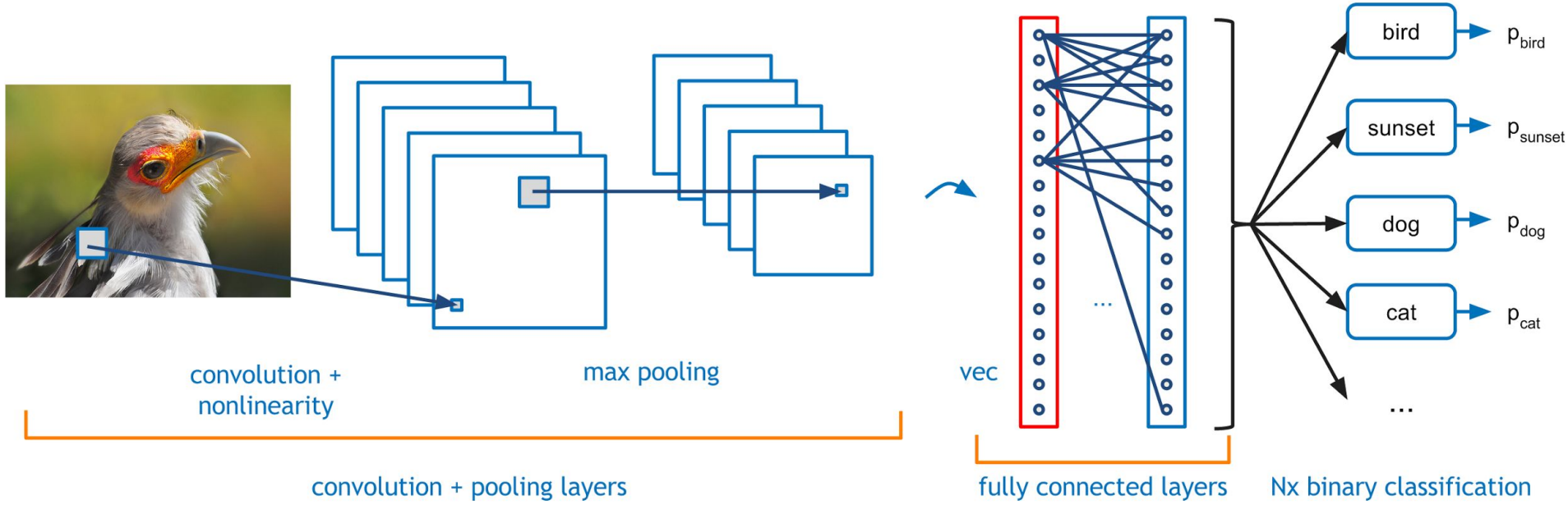
Figure 2. Illustration of edge detection results on the BSDS500 dataset: (top row) original image, (second row) ground truth, (third row) results of SCG [31], and (last row) our results for SE-MS.

Fully Convolutional Network Edge Maps

Use FCN to predict:

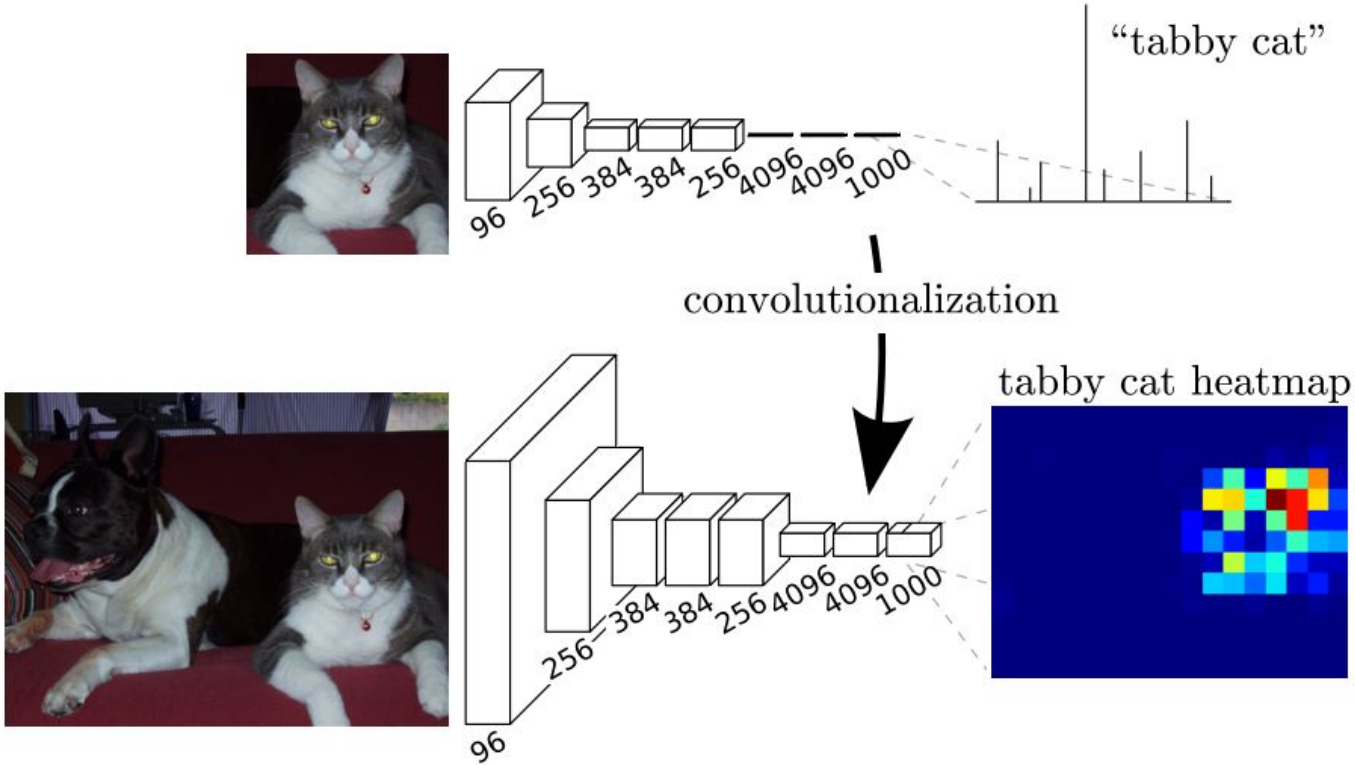
1) Informative edge map: edges of the projected 3D box fitting the room
2) Geometric context labels: five faces and clutter

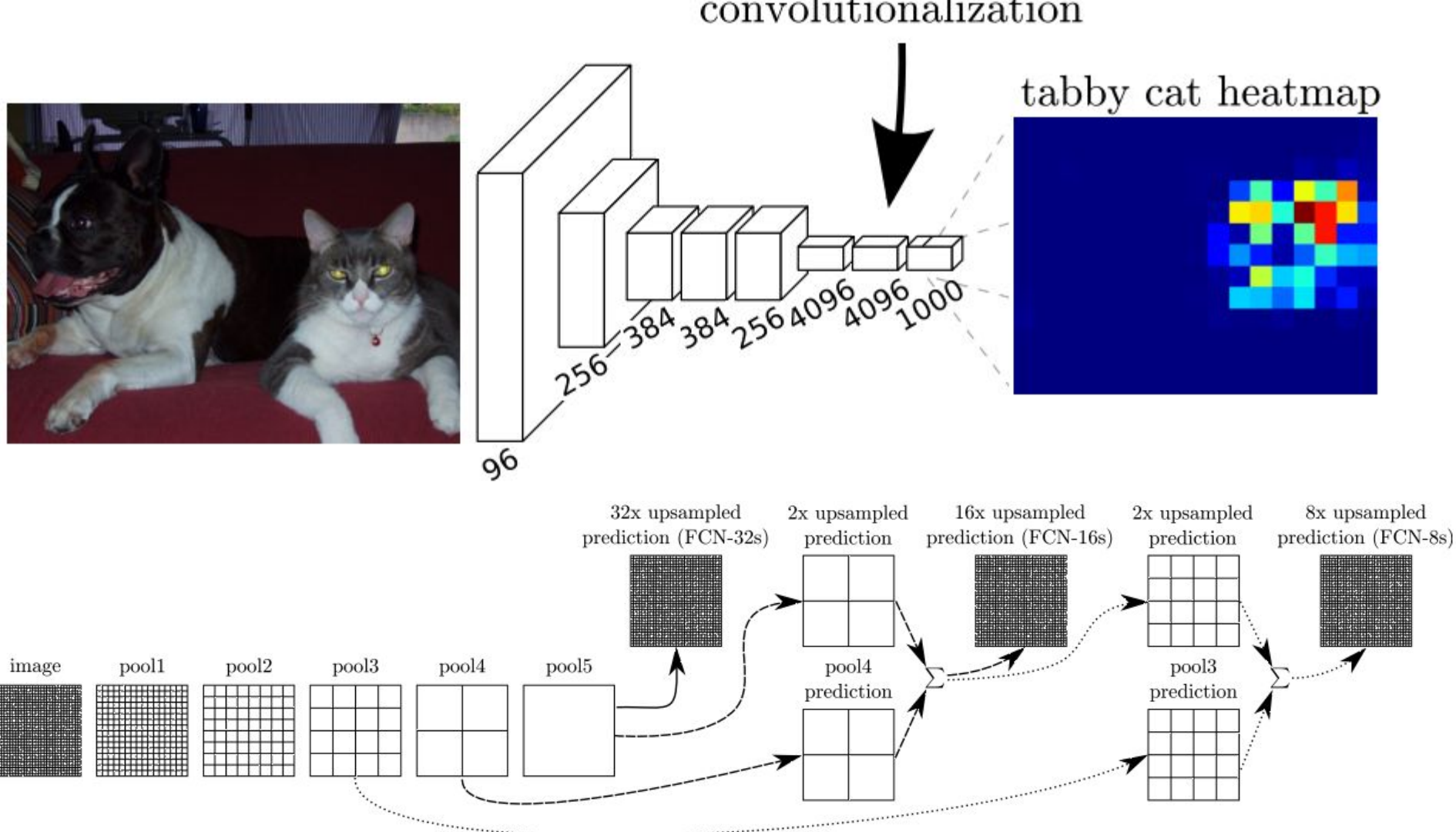
Fully connected layer → Fully convolutional layer



Fully connected layer: Linear Weight Matrix = (256*7*7) x 4096

Fully convolutional layer: Convolution Kernel Matrix = (256 x 7 x 7) x 4096

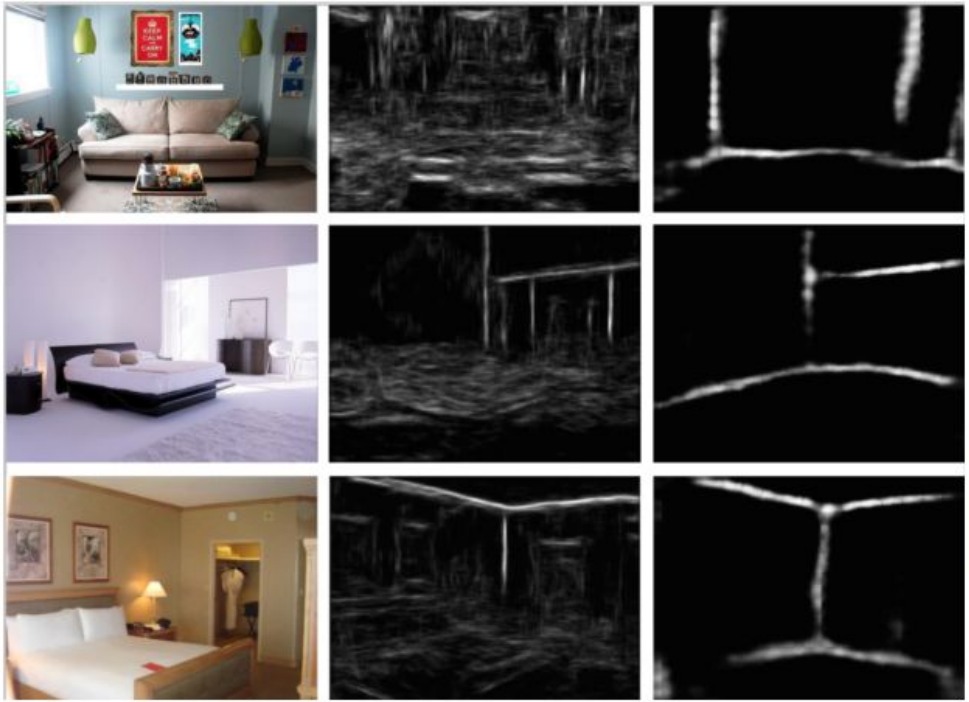



convolutionalization
tabby cat heatmap
96
256
384
384
256
4096
4096
1000
image
pool1
pool2
pool3
pool4
pool5
32x upsampled prediction (FCN-32s)
2x upsampled prediction
pool4 prediction
16x upsampled prediction (FCN-16s)
2x upsampled prediction
pool3 prediction
8x upsampled prediction (FCN-8s)

Fully Convolutional Network Edge Maps

1) Use Pre-trained model on NYUDv2 RGBD dataset on segmentation task
2) Discard the layers related to depth feature inputs
3) Labels: five faces and clutter
4) Finetune

Structured Forest vs. FCN



Inference Model

Estimating the Box Orientation

Assumption:

1) room can be modelled by a box layout
2) most surfaces inside room are aligned with the room directions.

Goal: Estimate a triplet of vanishing points

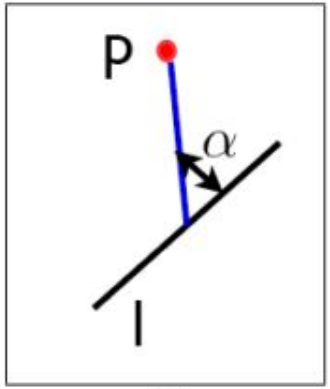
Estimating the Box Orientation

Approaches: Rother's algorithm

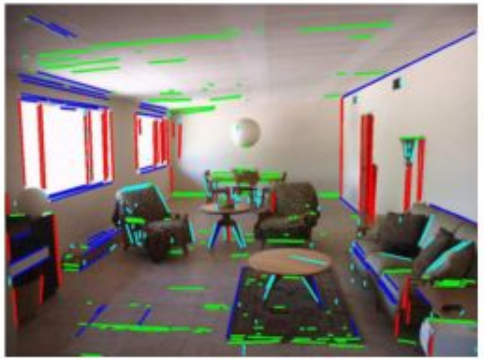
1) Candidate points are chosen as intersection points of all detected lines.
2) Rank all triplets using voting strategy:

$$\boldsymbol{v}(l,p) = |l| * \exp -(\frac{\boldsymbol{\alpha}}{2\sigma^2})$$

3) Assign each detected line in the image to one of the vanishing points - line membership



(a)

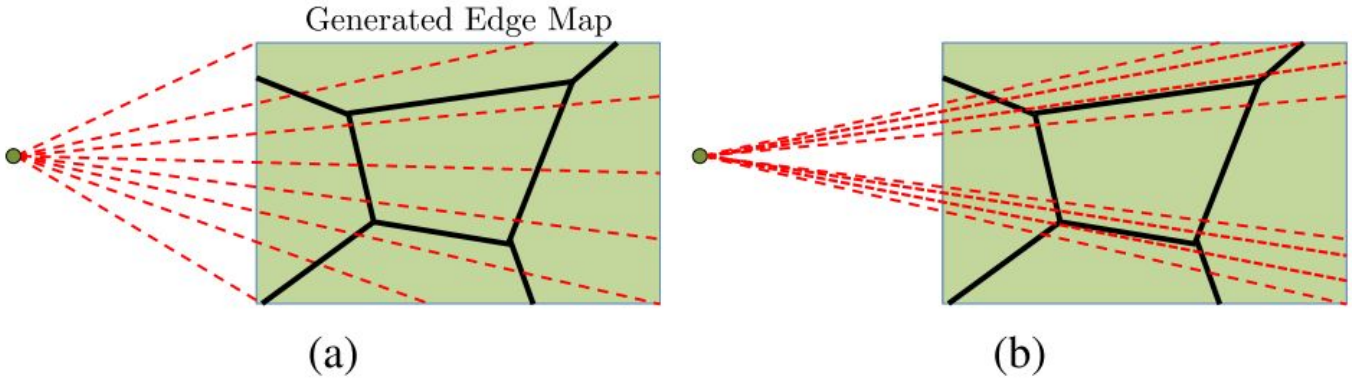


(b)

Figure 3. **(a) Angular distance** of a line segment to a vanishing point, computed as the angle between the line segment and the line joining the mid point of the segment to the vanishing point. **(b) Line memberships**: red, green and blue lines correspond to 3 vanishing points, and the outlier lines are shown in cyan.

Adaptive Layout Generation

1) Take informative edge maps as rough estimation
2) Draw uniformly spaced sectors originating from the vanishing points.
3) Rank all resulting sectors by the average informative edge strength and retain top K sectors.



Ranking Box Layouts

Among all possible layouts consisting of the sectors, rank the box layouts according to how well they fit the ground truth layout.

Given {x1, x2, …, xn} training images, and their layouts {y1, y2, …, yn}, we wish to learn a mapping $f(x, y) = w^T \psi(x, y)$ that can be used to assign a score to the automatically generated candidate layouts for an image, i.e.,

$$\min_{w,\xi} \frac{1}{2}||w||^2 + C \sum_i \xi_i$$
$$\text{s.t.} \quad \xi_i \geq 0 \; \forall i, \quad \text{and}$$
$$w^T \psi(x_i, y_i) - w^T \psi(x_i, y) \geq \Delta(y_i, y) - \xi_i,$$
$$\forall i, \forall y \in \mathbb{Y} / \; y_i$$

Ranking Box Layouts

Score function: $$f(x, y) = w^T \psi(x, y)$$

Features:

1) Informative Edge (IE): <mask of layout y, mask of informative edge maps>
2) Line Membership (LM):

$$f_l(F_k) = \frac{\sum_{l_j \in C_k} |l_j|}{\sum_{l_j \in L_k} |l_j|}$$

Ranking Box Layouts

Score function: $$f(x, y) = w^T \psi(x, y)$$

Features:

1) Informative Edge (IE) edges: <mask of layout y, mask of informative edge maps>
2) Line Membership (LM):
$$f_l(F_k) = \frac{\sum_{l_j \in C_k} |l_j|}{\sum_{l_j \in L_k} |l_j|}$$
3) Geometric Context (GC): Each line is weighted by the confidence that a line is inside or outside an object region

Evaluation

Datasets:

Dataset	# Train	# Val.	# Test
Hedau	209	–	105
Hedau+	284	53	–
SUNbox	543	53	–
LSUN	–	–	1000

Evaluation

Informative edge prediction:

	Forest			FCN		
Setting	**ODS**	**OIS**	**AP**	**ODS**	**OIS**	**AP**
BSDS [7]	0.159	0.165	0.052	–	–	–
3 Class - Hedau+	**0.178**	0.176	**0.104**	0.235	0.237	0.084
3 Class - SUNbox	0.177	0.169	0.103	0.227	0.232	0.086
1 Class - Hedau+	0.174	**0.177**	0.094	0.226	0.227	0.080
1 Class - Hedau+ (Joint)	–	–	–	**0.255**	**0.263**	**0.130**
1 Class - SUNbox	**0.178**	0.172	0.103	0.179	0.180	0.056
1 Class - SUNbox (Joint)	–	–	–	0.206	0.209	0.069

Evaluation

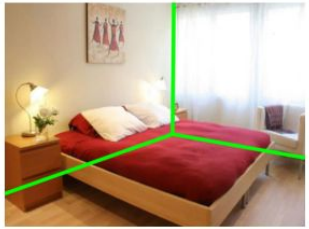
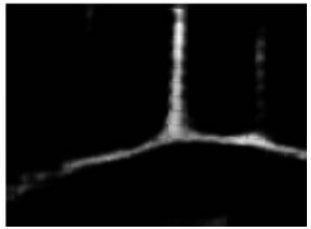
Layout Estimation Performance:

Error = percentage of pixels whose face identity disagrees with ground-truth.

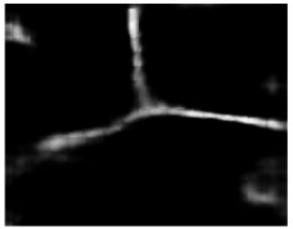
Setting	Forest		FCN	
	Uniform Layout Error (%)	**Adaptive Layout Error (%)**	**Uniform Layout Error (%)**	**Adaptive Layout Error (%)**
3 Class - Hedau+	23.66	20.59	26.19	16.05
3 Class - SUNbox	**19.97**	**16.89**	20.90	16.20
1 Class - Hedau+	23.15	21.71	20.62	13.89
1 Class - Hedau+ (Joint)	–	–	**18.30**	**12.83**
1 Class - SUNbox	20.81	18.03	23.64	18.43
1 Class - SUNbox (Joint)	–	–	18.95	15.09

Results

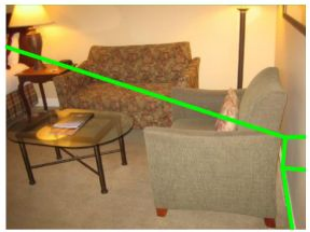
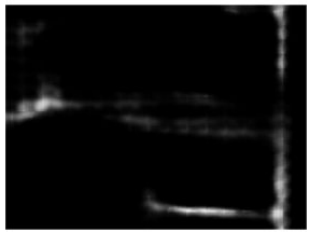

0.45%

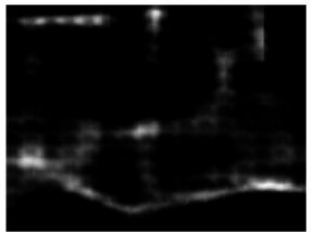
1.41%

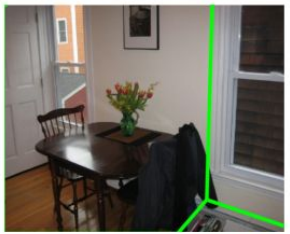
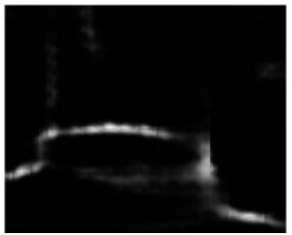
1.77%

90.38%

68.12%

61.05%

Future work

1. FCN trained on LSUN
2. Clutter information
3. Instead of post-processing, incorporate the vanishing points and layout candidates into FCN training.