Seeing 3D chairs:
Exemplar part-based 2D-3D alignment using a large dataset of CAD models

16-824 Learning Based Method in Vision

Namhoon Lee

The figures in the slides are from the original paper as well as the author's presentation at CVPR14.
http://www.di.ens.fr/willow/research/seeing3Dchairs/

Motivation

Is there a chair?

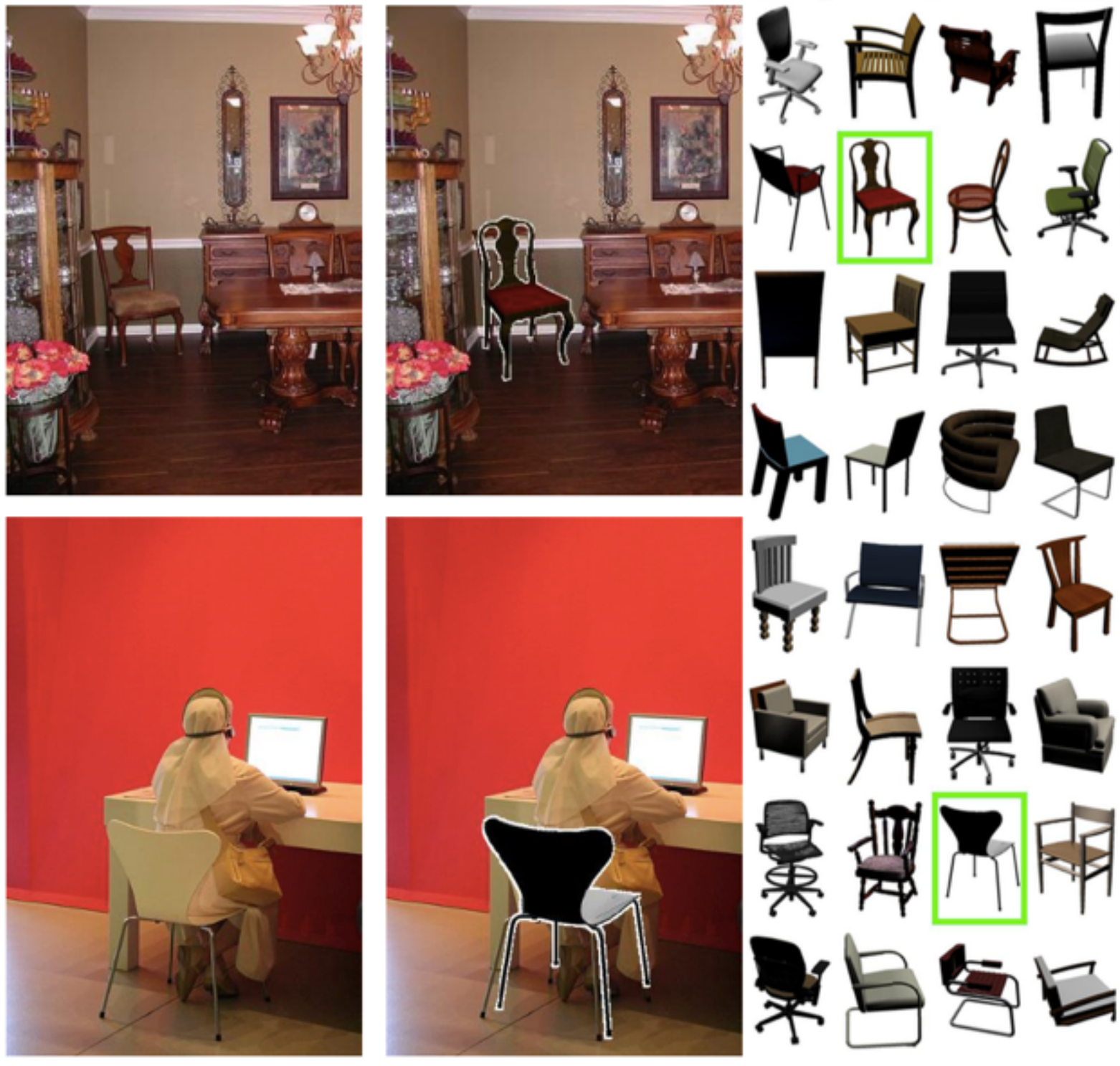
There is a chair!

Can we say more about it?

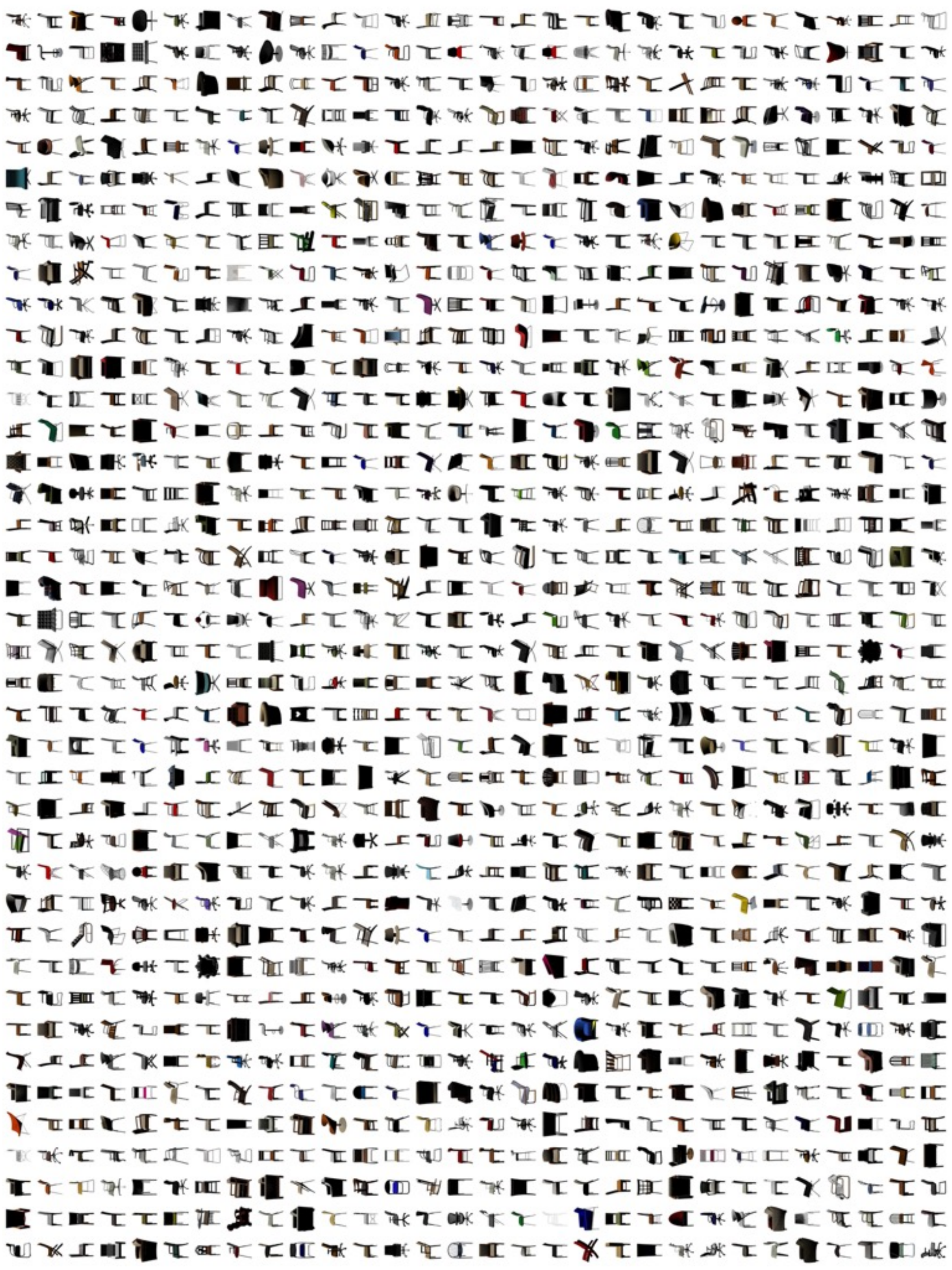
Can we sit on it?

Style and pose

Goal



Dataset



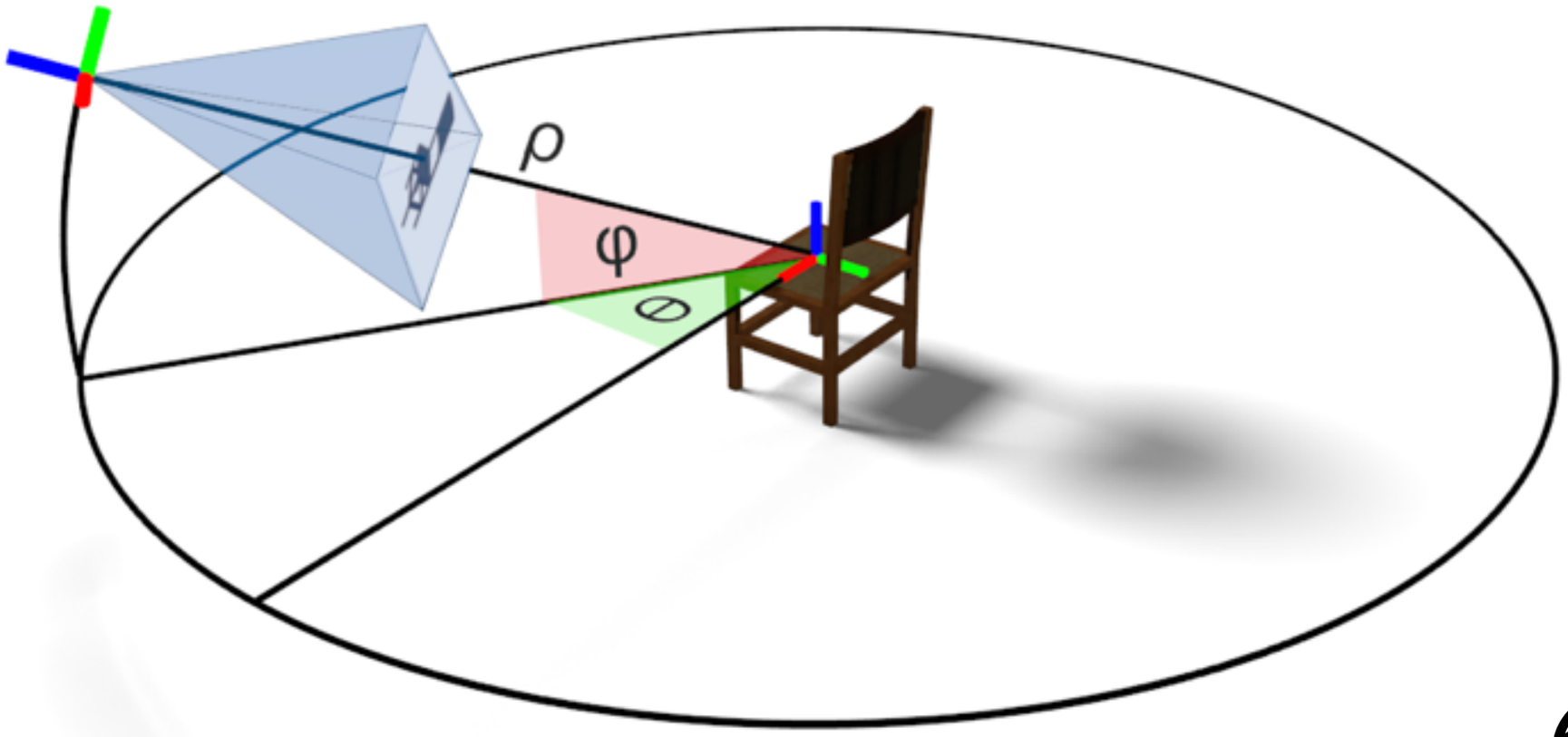
Large dataset of 3D Chairs





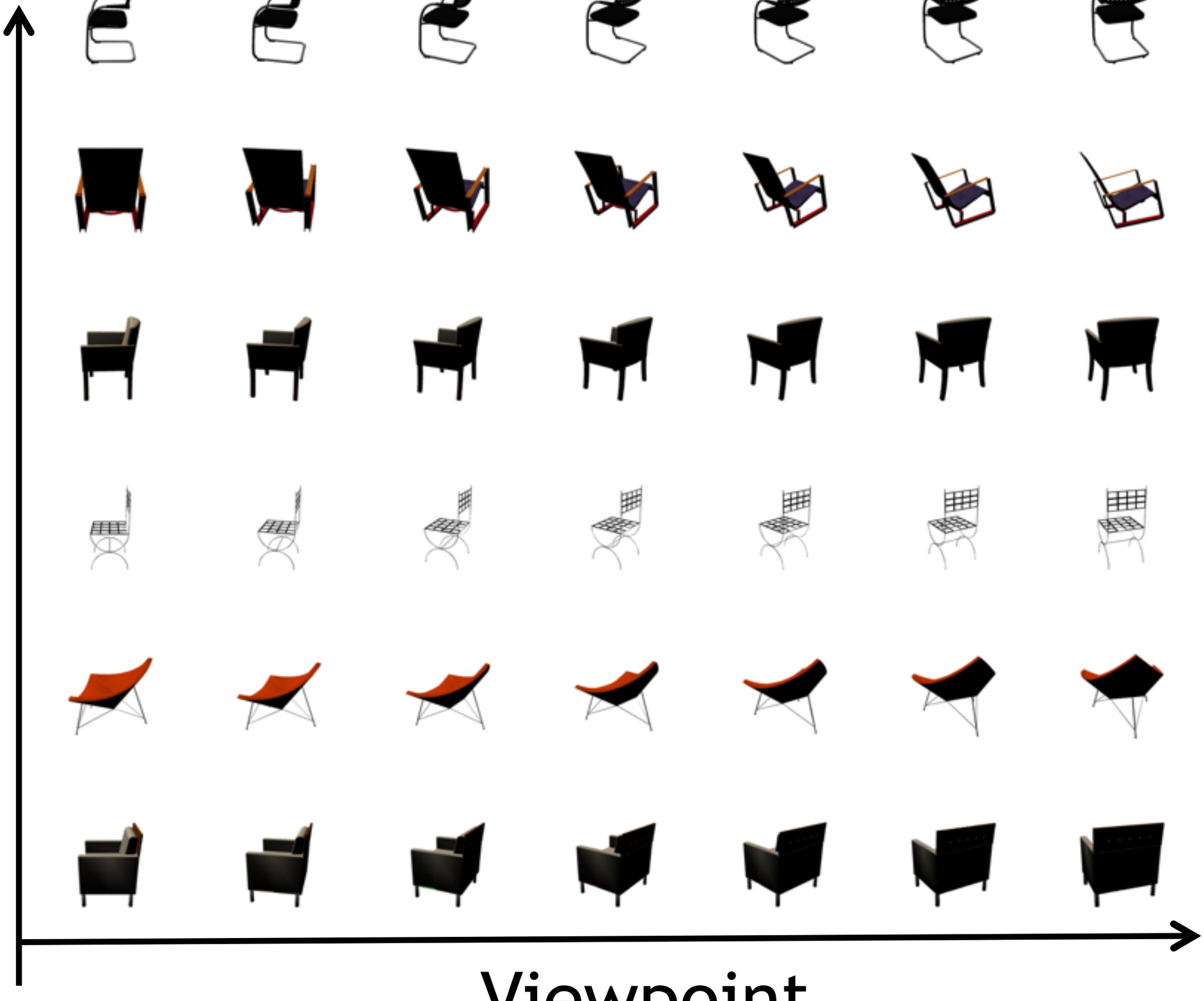

- 1,393 3D chair models from 3D Warehouse
- 62 different viewpoints => 1,393*62 = 86,366 synthesized images
- 10 discriminative patches for each view ~ 800K patches

Approach: use 3D models



62 views

Style



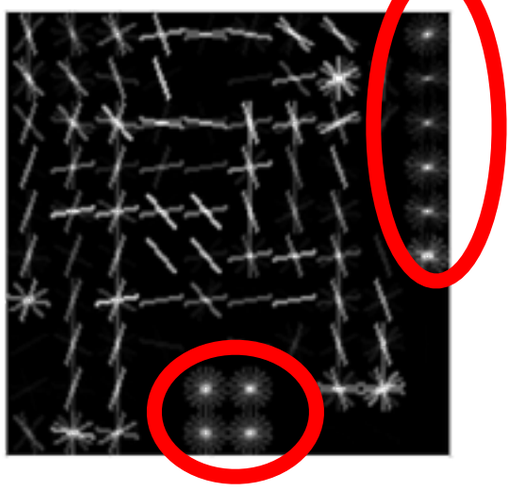
Viewpoint

Approach

How to match CG to photograph?

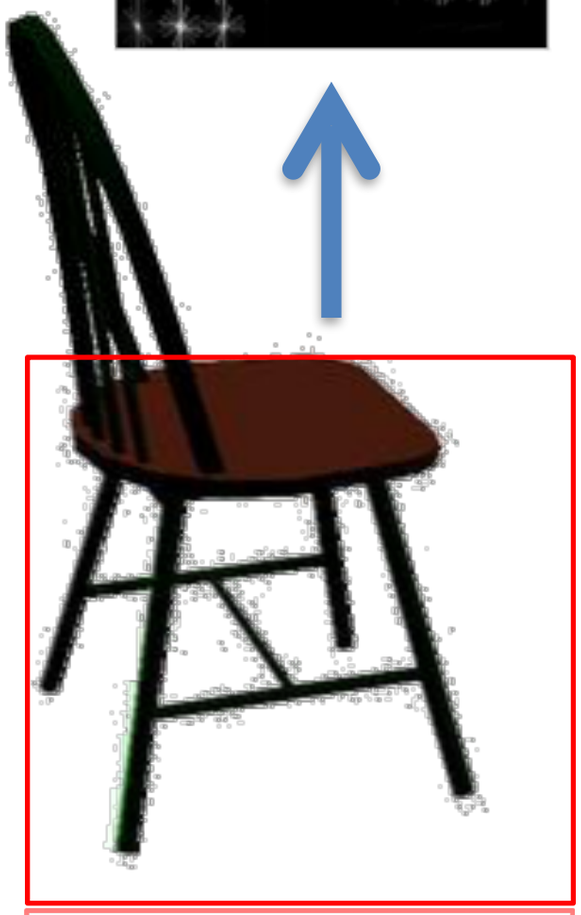
HOG

Classifier



Masked classifier

$$w^\top x + b$$



HOG mask

Implementation: exemplar-LDA

See also:
[Shrivastava et al. 2011]

Approach: CG-to-photograph

Implementation: exemplar-LDA

Discriminative visual element detectors using exemplar-LDA based training

q : A square image patch (represented as a HOG descriptor)

x : patch in the input image

similarity: $s_q(x) = w_q^T x$

- exemplar-classifier: q VS all the negatives

- LDA

$$w_q = \Sigma^{-1}(q - \mu_n) \qquad \mu_n = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_n)(x_i - \mu_n)^T$$

$$S_q(x) = \Phi(q)^T\Phi(x) - \Phi(q)^T\Phi(0) \qquad \Phi(x) = \Sigma^{-\frac{1}{2}}(x - \mu_n)$$

Technical details

- select 10 most discriminative patches per 3D view based on $\|\Phi(q)\|^2$
- Non-max suppression intersection area to union ratio of 0.25
- filter out patches less than 100x100 pixels
- 10x10 HOG cells —> 900 dimensional feature
- set to 0 for the component in the weight, which has the sum of the absolute value across the HOG channels < 0.01

Further details

- Calibrating visual element detectors (800K)

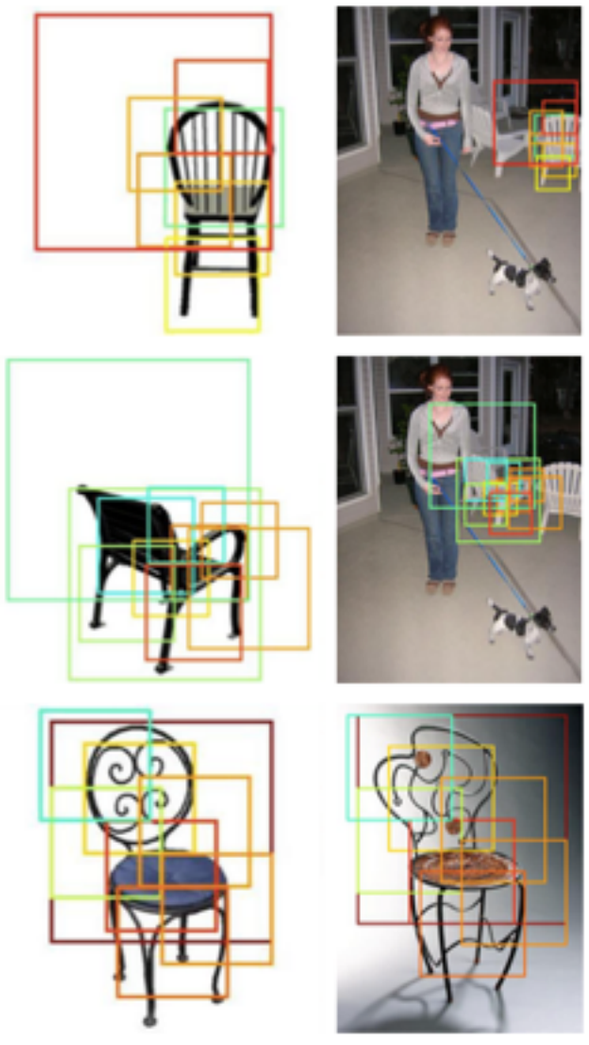
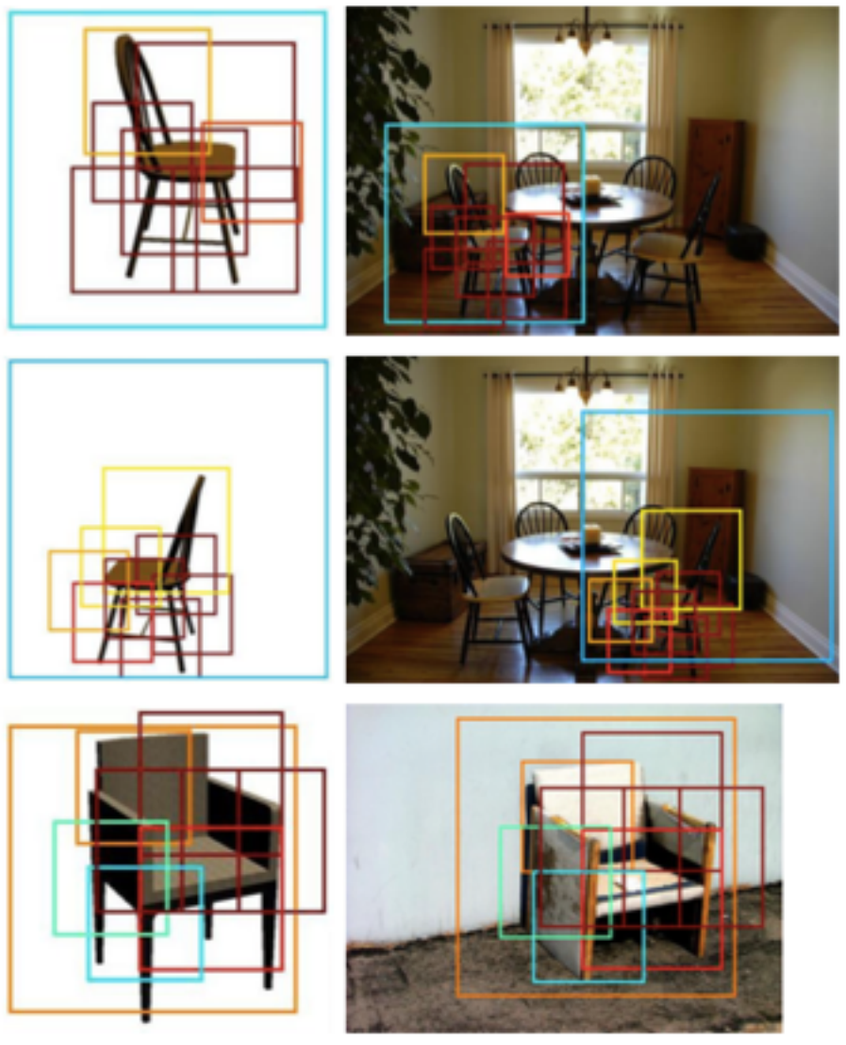
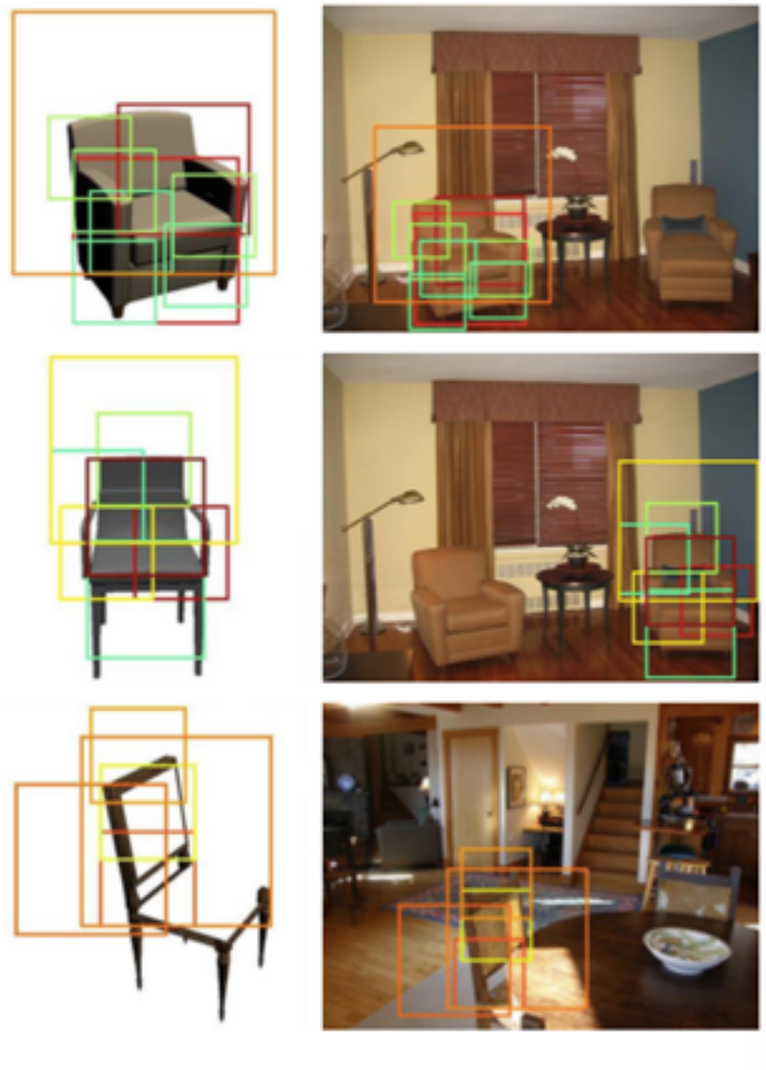
$$S_q^{'}(x) = a_q S_q(x) + b_q$$

1. run on a 200K negative patches
2. select x_n (false positive rate: 0.01%)
3. μ_n: mean HOG feature
4. set $S_q^{'}(x_n) = 0 \quad S_q^{'}(\mu_n) = -1$

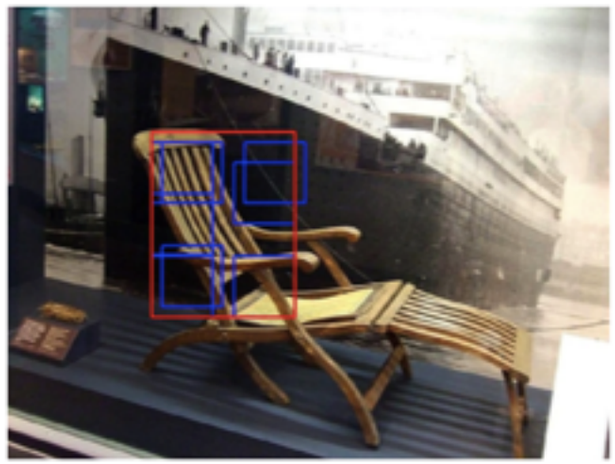
- Part-based detection
  1. find a patch whose response > 0.2 —> root
  2. local search around the expected location
  3. final score:= sum of the visual element detection scores

Results

Results

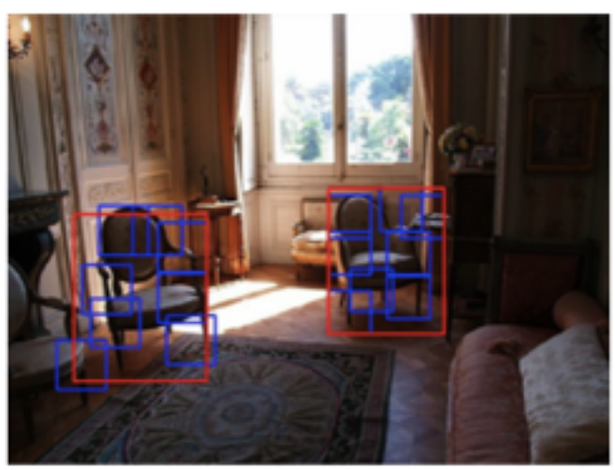




Results

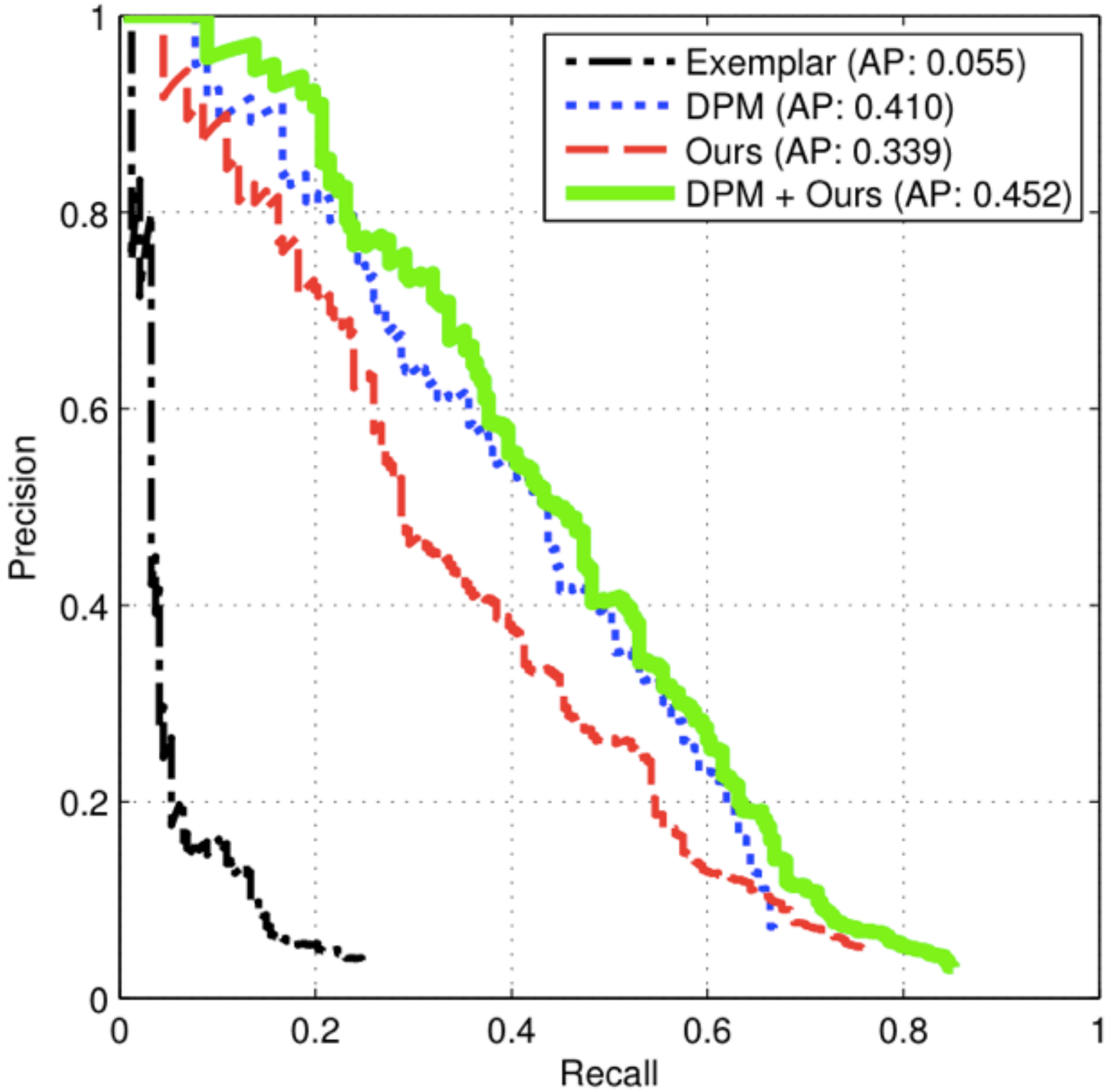






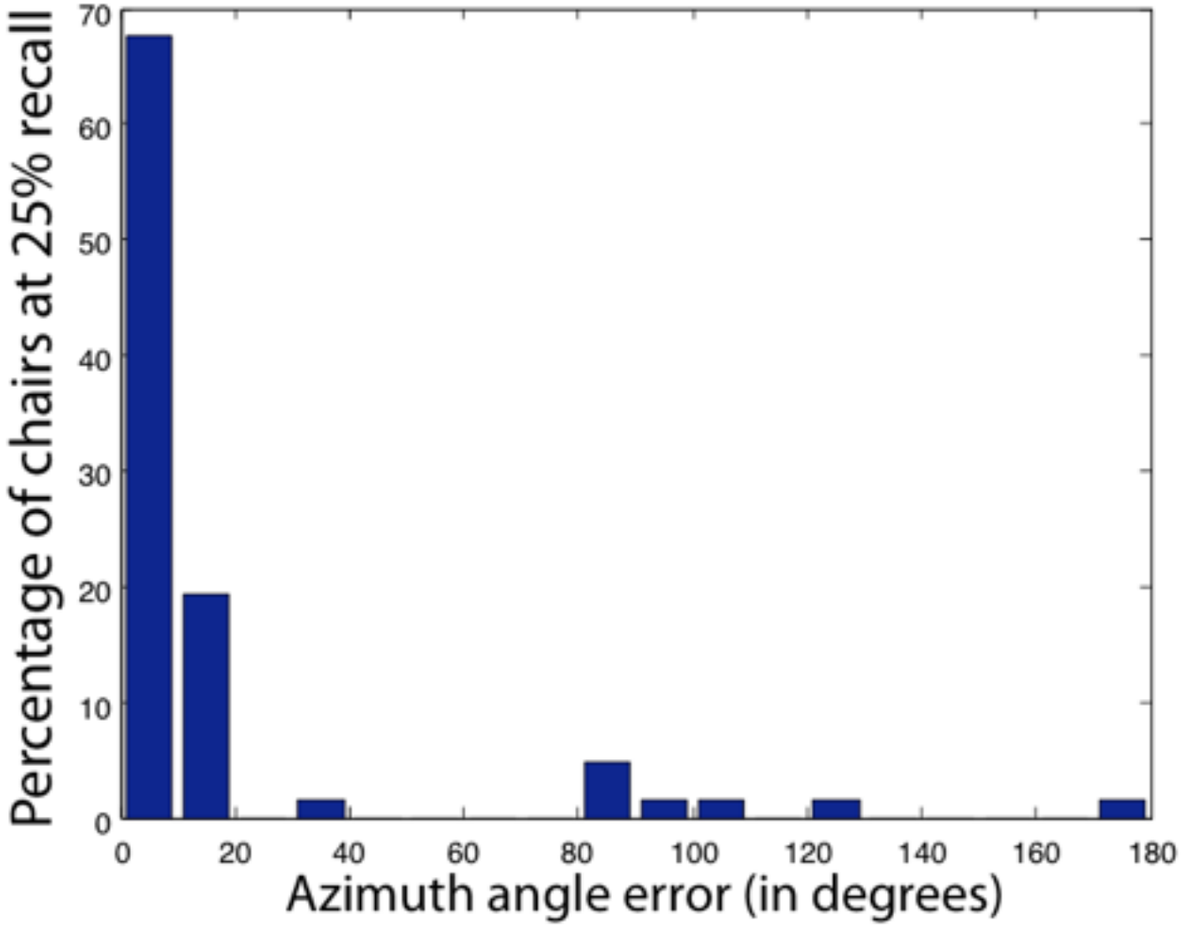

Results

Results


Exemplar (AP: 0.055)
DPM (AP: 0.410)
Ours (AP: 0.339)
DPM + Ours (AP: 0.452)
Precision
Recall
0
0.2
0.4
0.6
0.8
1

Results

	Alignment		Style		
	Good	Bad	Good	Ok	Bad
Exemplar-LDA	52%	48%	3%	31%	66%
Ours	90%	10%	21%	64%	15%



Thank you