LEARNING TO GENERATE CHAIRS WITH CONVOLUTIONAL NEURAL NETWORKS

Alexey Dosovitskiy, Jost Tobias Springenberg and Thomas Brox

University of Freiburg

Presented by: Shreyansh Daftry

Visual Learning and Recognition [16-824], Spring 2015

MOTIVATION

- Generate images of object given high-level inputs.

Object Type

Viewpoint

Color

Brightness

⋮

APPROACH

- Discriminative CNN

CNN

Chair,
Style,
3D Pose,
etc …

APPROACH

- Discriminative CNN

CNN

Chair,
Style,
3D Pose,
etc …

- Generative CNN

Chair,
Style,
3D Pose,
etc …

CNN

APPROACH

- Discriminative CNN

CNN

Chair,
Style,
3D Pose,
etc ...

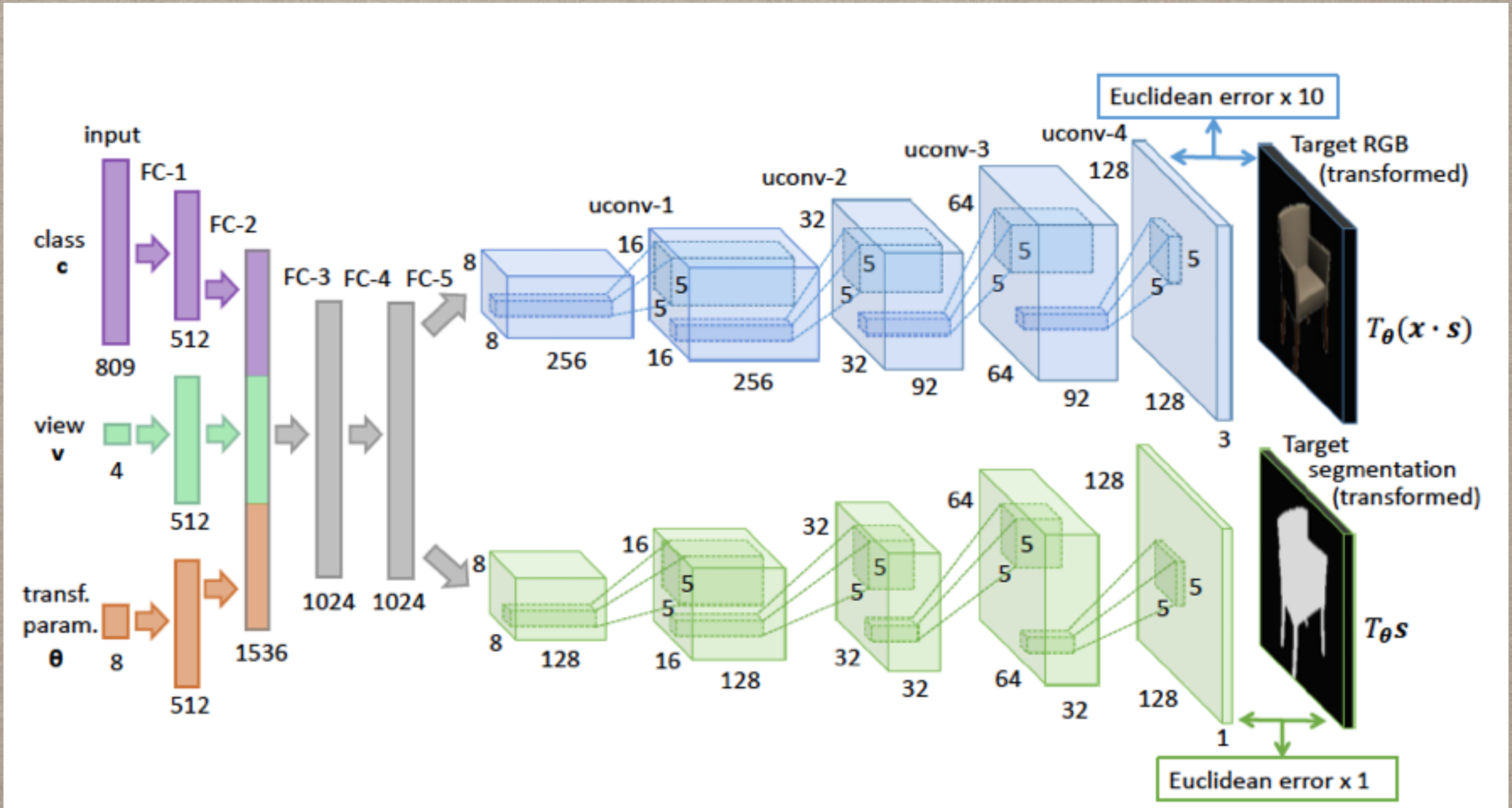
- Generative CNN

Chair,
Style,
3D Pose,
etc ...

$(\text{CNN})^{-1}$

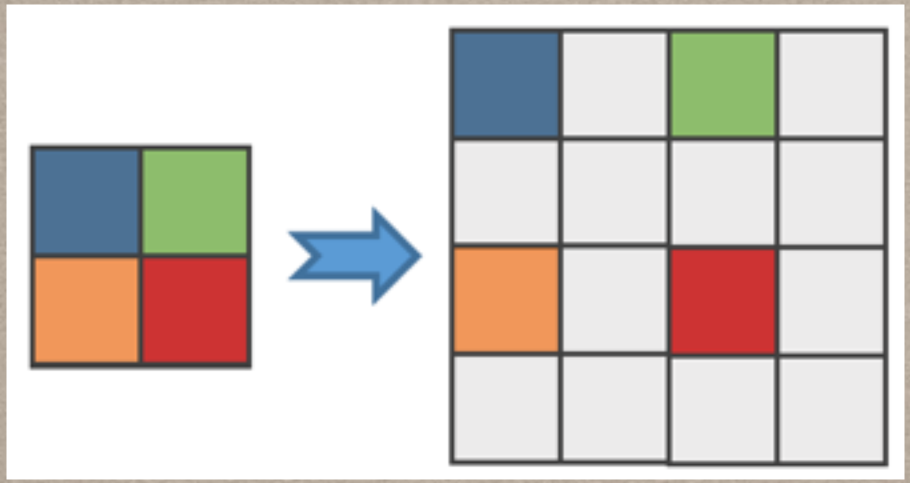
APPROACH

- CNN Architecture

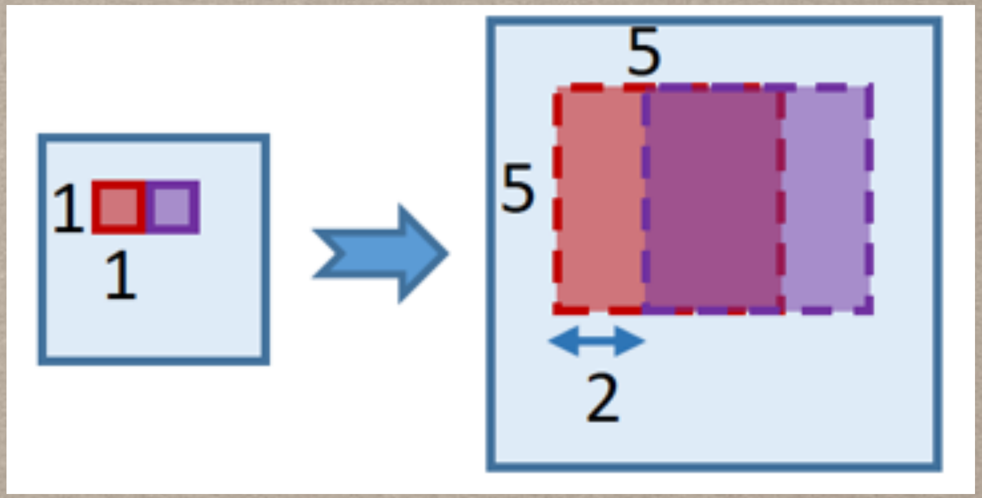


APPROACH

- Un-Pooling + Convolution Layer



2x2 Un-Pooling

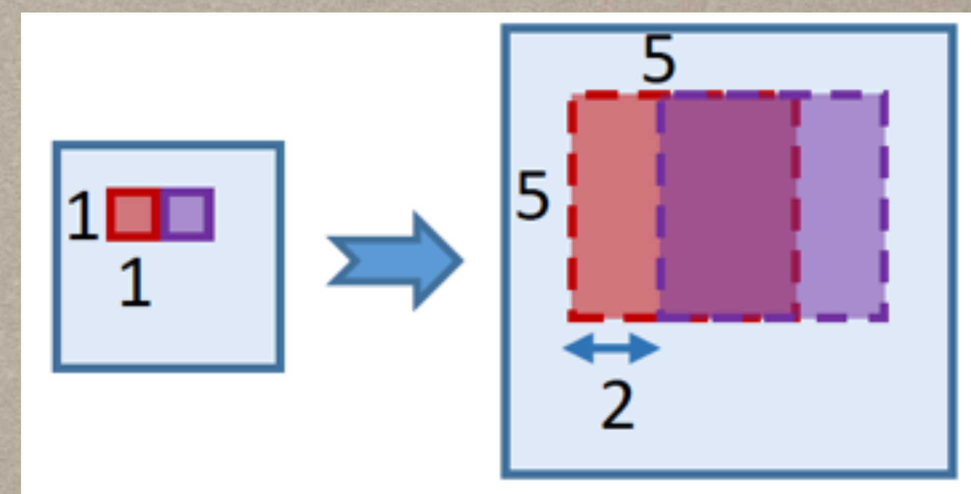


2x2 Un-Pooling + 5x5 Convolution

APPROACH

- Un-Pooling + Convolution Layer

2x2 Un-Pooling



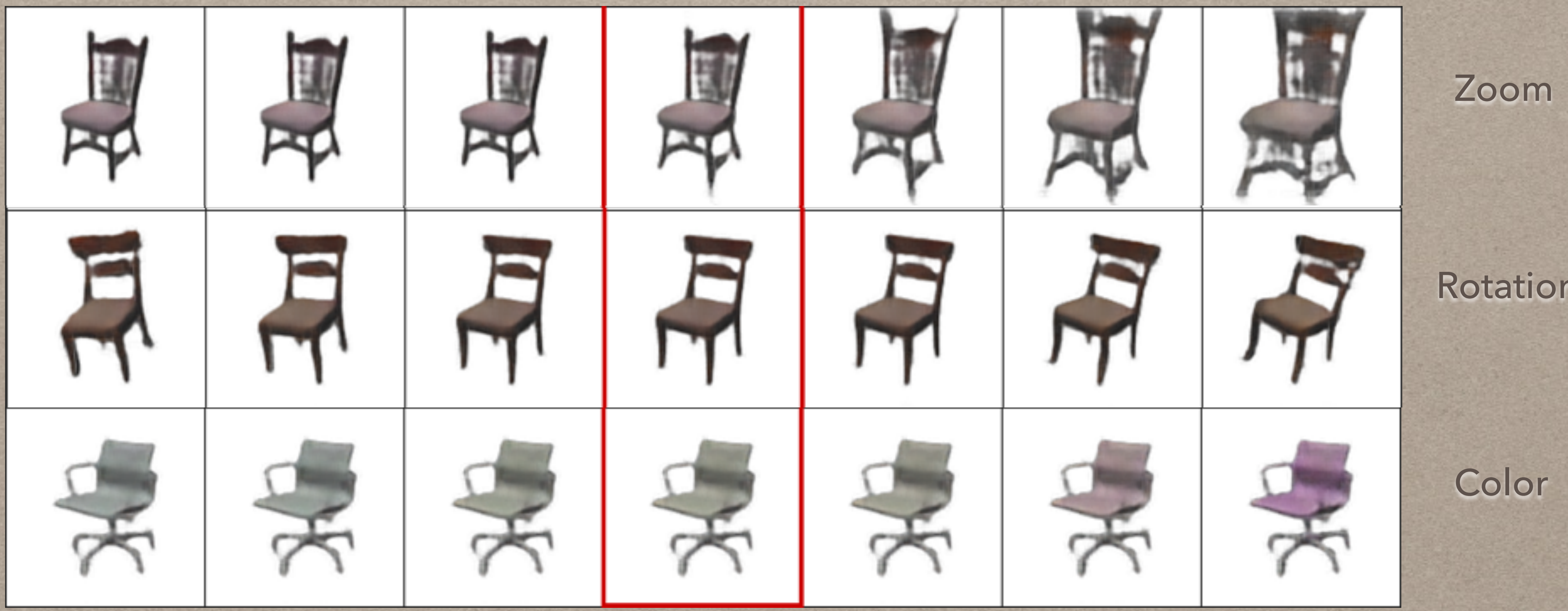
2x2 Un-Pooling + 5x5 Convolution

- Discriminative Vs. Generative

Conv + Pooling $\xrightarrow{\text{Inverse}}$ Un-Pooling + Conv

ANALYSIS OF NETWORK

- Network Capacity



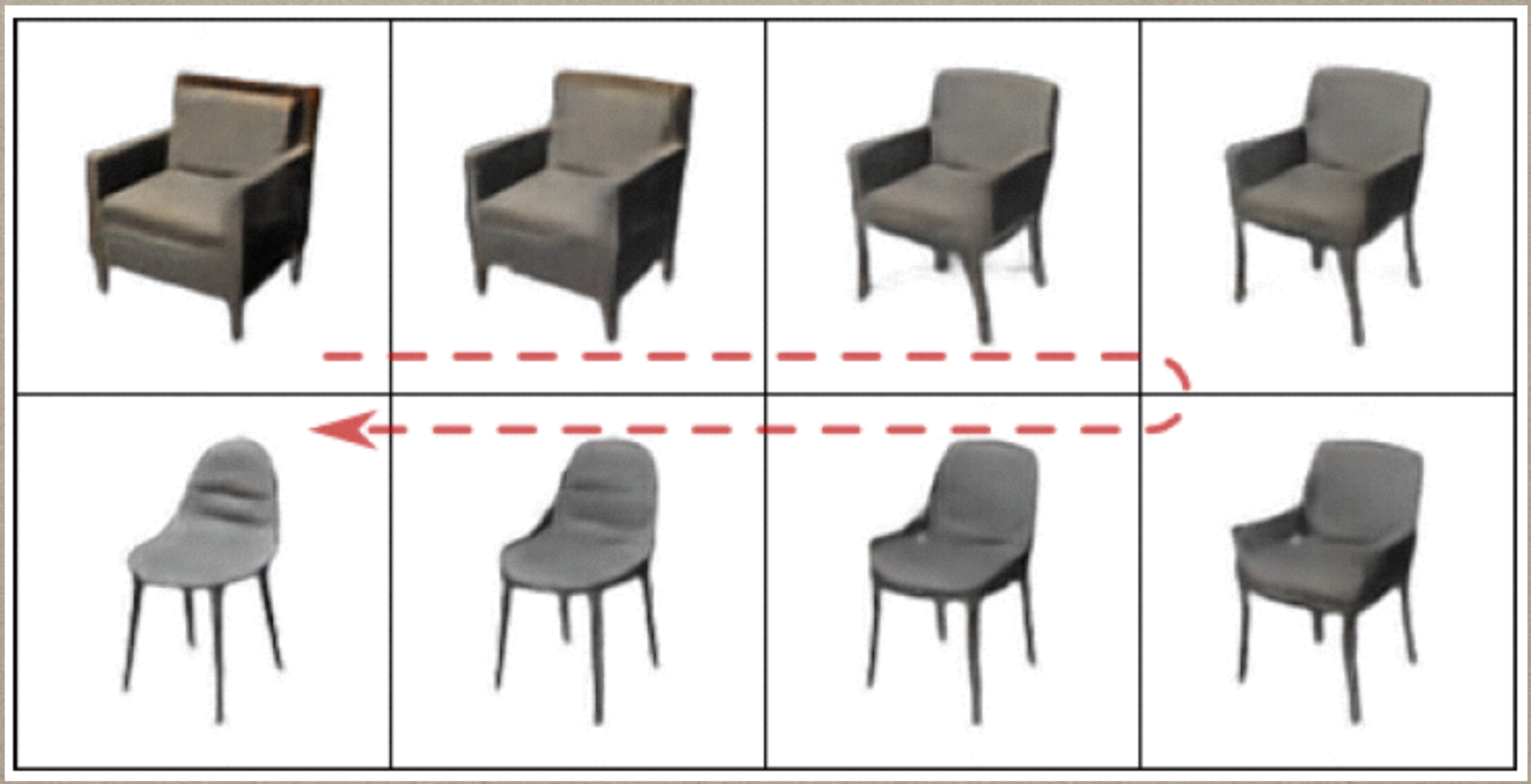
ANALYSIS OF NETWORK

- Activating Single Units

"Zoom" Neuron

EXPERIMENTS

- Knowledge Transfer



EXPERIMENTS

- **Knowledge Transfer between Views**

Knowledge Transfer

No Transfer

15 Views

Knowledge Transfer

No Transfer

1 View

EXPERIMENTS

- Knowledge Transfer between Classes

CONCLUSION

- Supervised Training of CNN can also be used to generate images.
- Generative network does not merely learn, but also generalizes well.
- The proposed network is capable of processing very different inputs using the same standard layers

FPM: FINE POSE PARTS-BASED MODEL WITH 3D CAD MODELS

Joseph Lim, Aditya Khosla and Antonio Torralba

Massachusetts Institute of Technology

Presented by: Shreyansh Daftry

Visual Learning and Recognition [16-824], Spring 2015

MOTIVATION

- Why do we need Fine Pose Estimation?
- Why is it a hard problem?

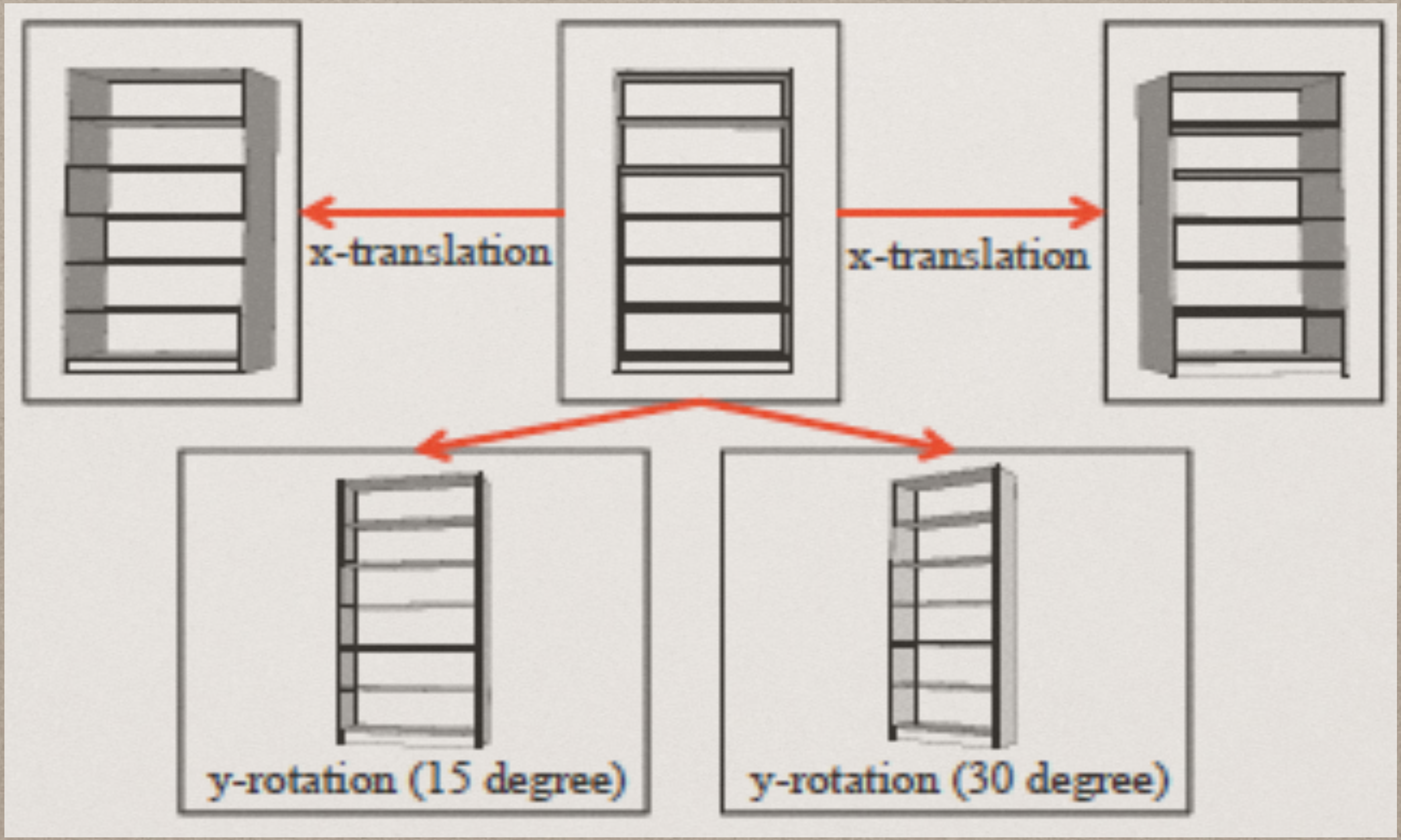
(a) standard object detection (b) fine pose estimation

APPROACH

- Goal: Given a set of CAD models, accurately detect and pose align them in RGB images if they contain instance of that object.

APPROACH

- Advantages of using CAD Models



APPROACH

- Disadvantages of using CAD Models

Image Statistics are significantly different

- No Texture
- No Occlusion
- No Illumination Artifacts

APPROACH

- Model:
  - Advantages of CAD Models + Difference in modalities.
  - We define a function $F_{\ominus}$ which measures how well a pose $\ominus$, fits a rectangular image window:

$$F_{\ominus}(x) = \alpha^{T} S_{\ominus}(x) + \beta^{T} O_{\ominus}(x) + \Upsilon^{T} Q_{\ominus}(x)$$

  - Goal is to maximize $F_{\ominus}$ for positive poses and minimize for negative poses.

DPM WITH 3D SHARED PARTS ($S_\ominus$)

- Training on simple parts based model using rendered images:

$$S_\ominus = \max [\ S^r_\ominus(x) + \sum S^p_\ominus(P_i, x)\ ]$$

where $S^r_\ominus(x) = w_\ominus . x_{hog}$ and $S^p_\ominus(P_i, x) = (w^p_\ominus . x_{hog} - \Phi)$

- Each pose $\ominus$ is considered a mixture of possible discretized poses

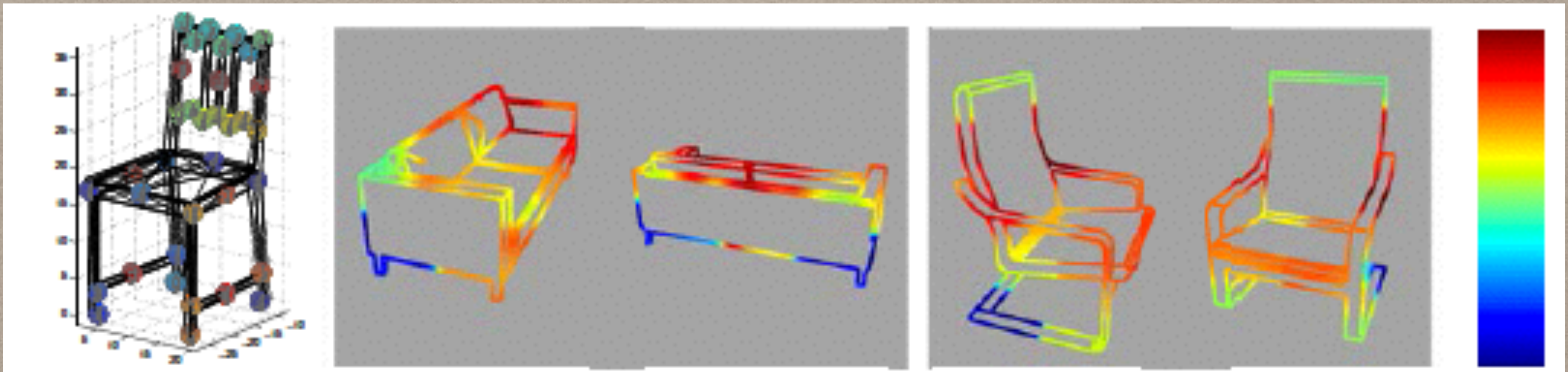
DPM WITH 3D SHARED PARTS (S_Θ)

- Obtaining Parts:
  - Unlike DPM, parts are not treated as Latent variables.
  - They are explicitly found using joints in 3D models
- Learning Mixture components:
  - It is computationally expensive to learn weights using SVM.
  - Exemplar LDA is used on root and part templates.

$$w_\Theta = \Sigma^{-1}_{real} \left(u^{+}_{\Theta} - u^{-}_{real} \right)$$

DPM WITH 3D SHARED PARTS (S_Θ)

- Part Importance: Goal is to learn which part is frequently occluded or does not contain discriminative shapes from real data.



(a) Object parts (b) Part importance (increases from blue to red)

OBJECTNESS SCORE ($O_{\ominus}$)

- Detect whether image window contains an object or not.
- Learnt using objectness classifier [Alexe *et. al*, CVPR 2010].
- Deep features and selective search used.

POSE QUALITY ($Q_\ominus$)

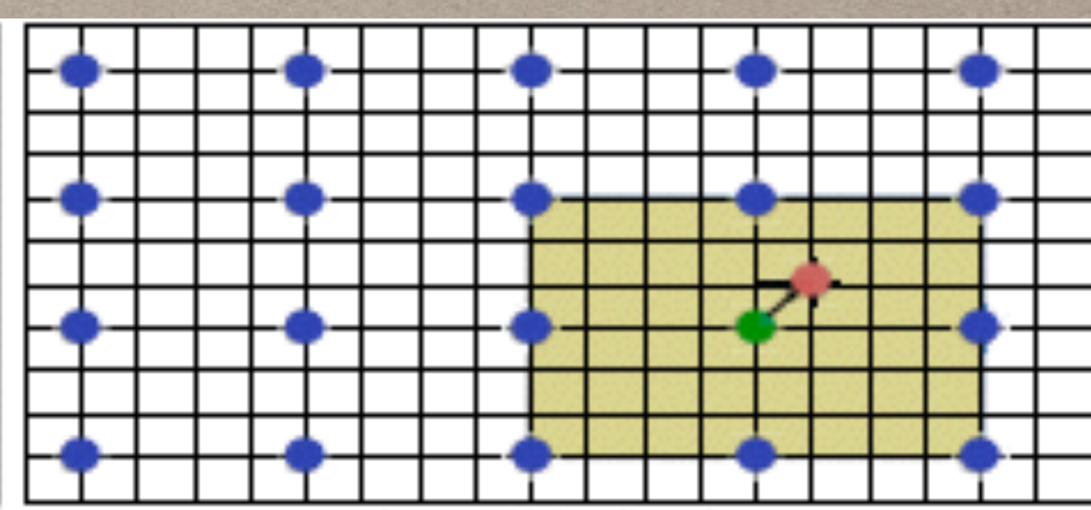
- Too many false positives when training with rendered images.
- More non-empty cells a view of a model has, more it suffers from false positives.
- To address this, they model the emptiness using 2 terms:

$$Q_\ominus = [\ \|w_\ominus\|\ ,\ n_\ominus\ ]^T$$

LEARNING AND INFERENCE

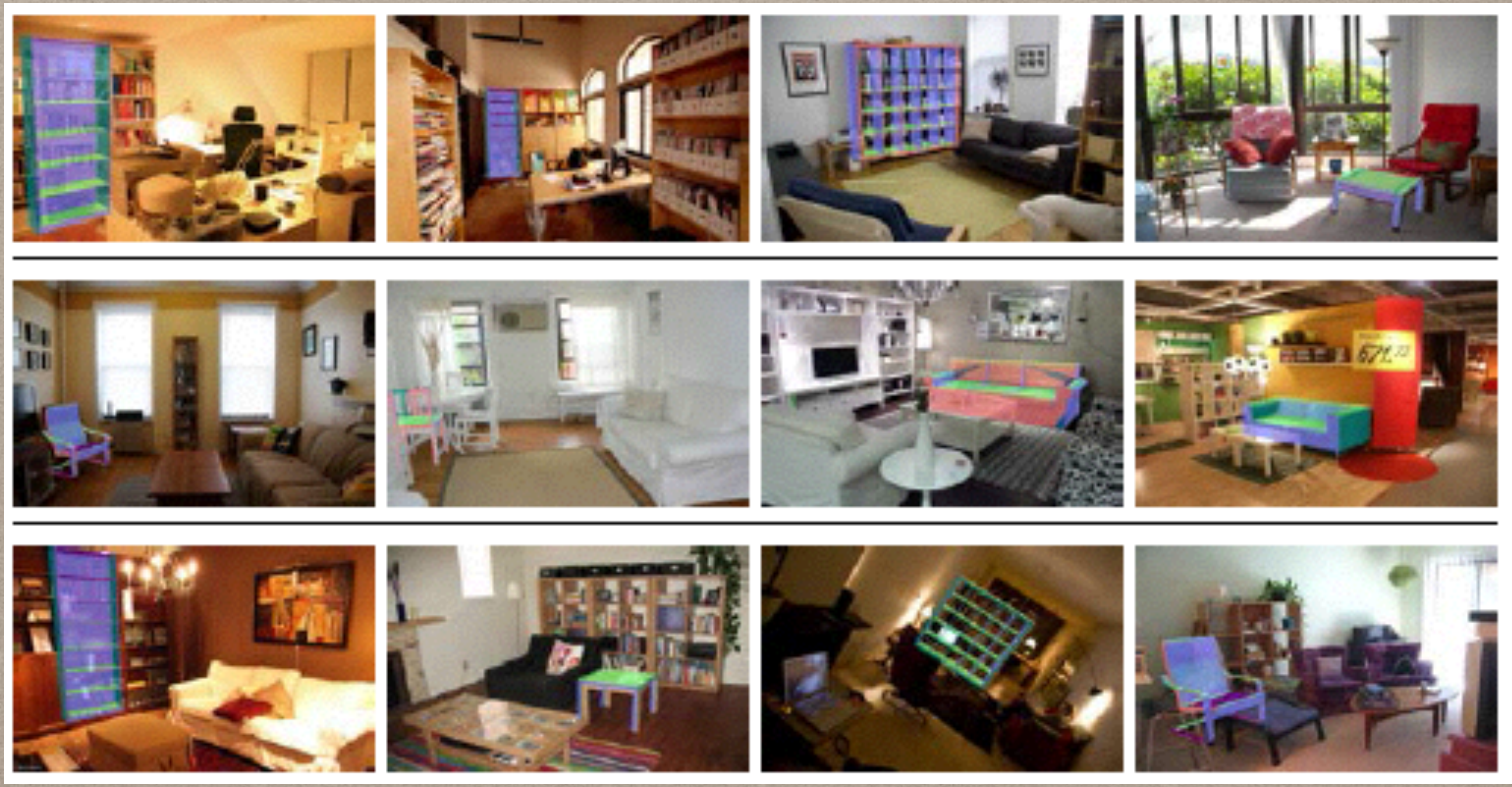
- S_Θ, O_Θ and Q_Θ are well defined. Thus we have a Linear System!
- Solved using a linear SVM in a max margin framework.
- Weights refined using hard negative mining.
- During inference, we find the nearest neighbor Θ'_i and borrow its weight.

(a) Pose to test (b) Nearest pretrained pose (c) Pose (a) made from (b)



EXPERIMENTS

- Top Detections



EXPERIMENTS

- Fine Pose Estimation

	bookcase *billy1*	chair *poang*	bookcase *expedit*	table *lack2*	sofa *karlstad*	bookcase *billy5*	chair *stefan*	sofa *ektorp*	mean
IKEA [27]	7.53	9.43	3.28	9.14	3.22	4.21	14.87	0.48	6.52
R	1.45	1.90	0.24	5.86	0.19	1.19	4.10	0.00	1.87
R+P	3.11	7.24	3.21	13.90	2.84	4.05	7.61	0.36	5.29
R+P+S	6.37	9.11	6.78	14.00	6.23	5.34	9.66	1.80	7.41
Full FPM	**10.52**	**14.02**	**9.30**	**15.32**	**7.15**	**6.10**	**16.00**	**5.66**	**10.51**

bookcase *billy1*	chair *poang*	bookcase *expedit*	table *lack2*	sofa *karlstad*	bookcase *billy5*	chair *stefan*	sofa *ektorp*	mean
0.85	0.93	0.96	0.90	0.82	0.75	0.90	0.63	0.84

EXPERIMENTS

- Pose Proposal

	bookcase *billy1*	chair *poang*	bookcase *expedit*	table *lack2*	sofa *karlstad*	bookcase *billy5*	chair *stefan*	sofa *ektorp*	mean
[27]	0.83	0.86	0.89	0.83	0.37	0.88	0.77	0.56	0.75
FPM	**0.87**	**0.93**	**0.94**	**0.88**	**0.83**	**0.91**	**0.99**	**0.91**	**0.91**

- Bounding Box Detection

	bookcase *billy1*	chair *poang*	bookcase *expedit*	table *lack2*	sofa *karlstad*	bookcase *billy5*	chair *stefan*	sofa *ektorp*	mean
(a) Intersection over Union ≥ 0.5									
IKEA [27]	24.41	28.32	21.73	11.12	22.65	11.22	28.57	2.37	18.80
DPM [8]	**49.89**	**51.63**	**71.87**	**48.85**	34.01	**42.11**	**45.34**	**28.80**	**46.56**
FPM	23.51	29.83	37.26	38.16	**35.85**	33.00	30.52	27.13	31.91
(b) Intersection over Union ≥ 0.8									
IKEA [27]	**20.34**	14.43	15.74	9.14	15.32	7.73	20.45	1.58	13.09
DPM [8]	9.41	15.58	15.47	10.02	20.12	3.05	20.44	11.59	13.21
FPM	17.37	**22.36**	**22.89**	**29.88**	**22.26**	**8.71**	**24.31**	**12.64**	**20.05**