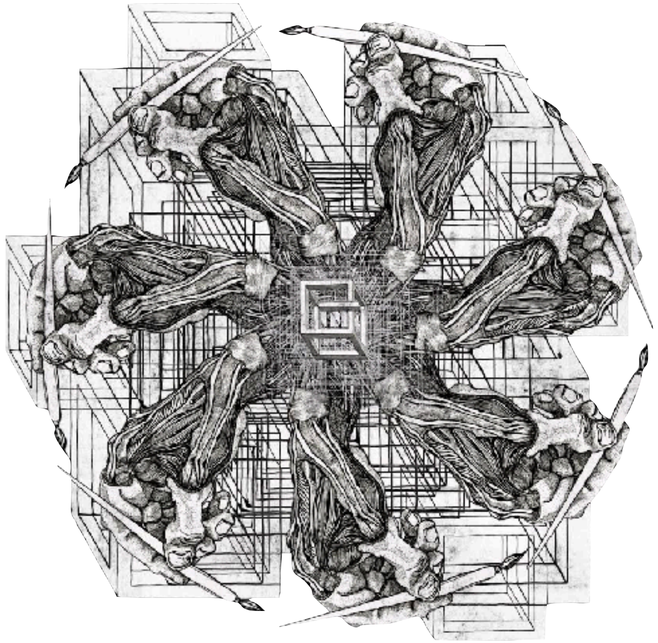


Learning Visual Representations

Amir R. Zamir

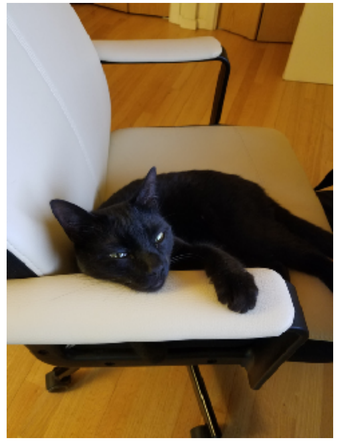
CS231A: Winter 2018

“Solving a problem simply means representing it so as to make the solution transparent.”

–- Herbert Simon, Sciences of the Artificial

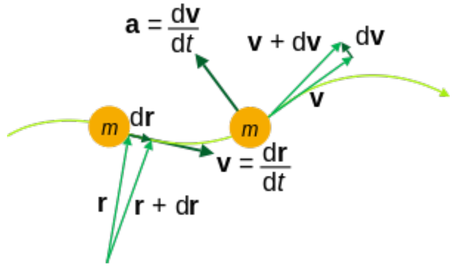
“Representation?”

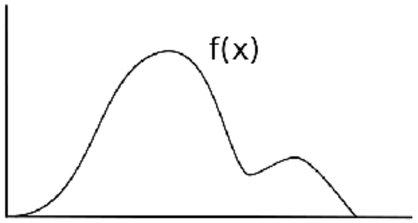
Things...

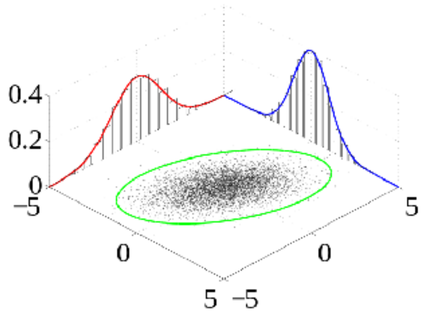


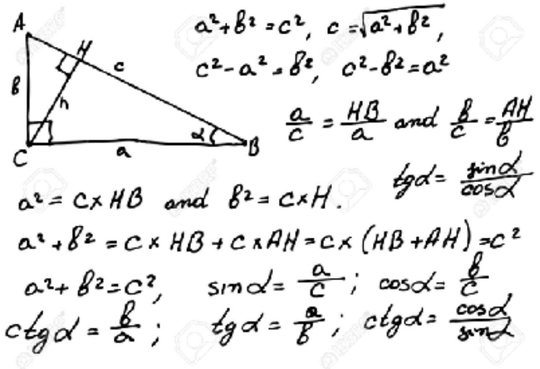
My heart beats as if the world is dropping,
you may not feel the love but i do its a heart
breaking moment of your life. enjoy the times
that we have, it might not sound good but
one thing it rhymes it might not be romantic
but i think it is great,the best rhyme i've ever
heard.

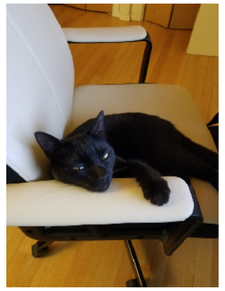
Our Knowledge...











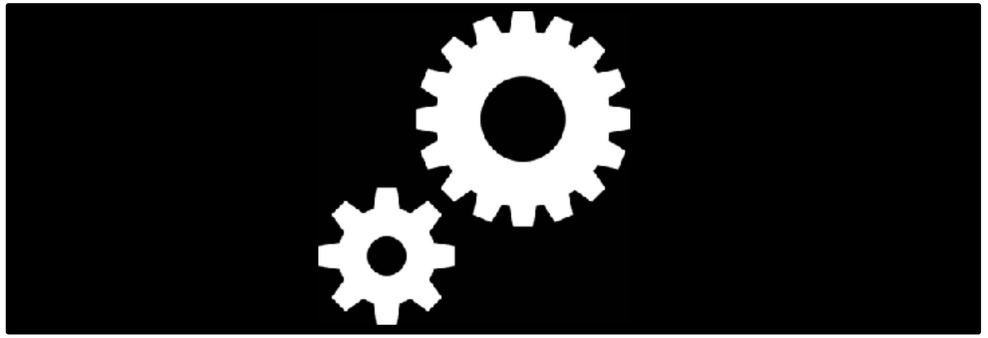
I dare not speak of what I have done. Such twisted thoughts overtook my mind and now I am sorry to say that *I have done the deed.* I have murdered the King of Scotland, King Duncan.

After naming me the *Worthy Thane of Cawdor*, this is how I repay him. I have betrayed him in the most unimaginable way a person possibly could, and I've been disloyal to him, just like I have to Banquo, whom I have lost as a dear friend. I wish that I had never done such a treacherous thing, as I am afraid that *I shall sleep no more.*

I was waiting anxiously for my Lady to sound the bell that called me to do the deed. But before she did, a symbol of the supernatural appeared before my eyes. *The dagger of the mind* captured me and the handle was to my hand yet I couldn't grasp it, but *I could see thee still.* I knew not whether to follow or to discard it from my eyes, but the *false creation* remained.

As I stepped closer to Duncan's room, I thought that I would panic and freeze, but when I got nearer, a sickening thought made me feel like I was doing the right thing! As soon as I heard the bell I knew that it was the bell summoning me.

I heard him pleading as the dagger pierced through his skin,



I dare not speak of what I have done. Such twisted thoughts overtook my mind and now I am sorry to say that *I have done the deed.* I have murdered the King of Scotland, King Duncan.

After naming me the *Worthy Thane of Cawdor*, this is how I repay him. I have betrayed him in the most unimaginable way a person possibly could, and I've been disloyal to him, just like I have to Banquo, whom I have lost as a dear friend. I wish that I had never done such a treacherous thing, as I am afraid that *I shall sleep no more.*

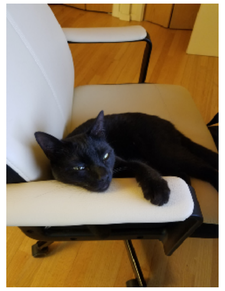
I was waiting anxiously for my Lady to sound the bell that called

handle was to my hand yet I couldn't grasp it, but *I could see thee still.* I knew not whether to follow or to discard it from my eyes, but the *false creation* remained.

As I stepped closer to Duncan's room, I thought that I would panic and freeze, but when I got nearer, a sickening thought made me feel like I was doing the right thing! As soon as I heard the bell I knew that it was the bell summoning me.

I heard him pleading as the dagger pierced through his skin,

Macbeth was guilty.



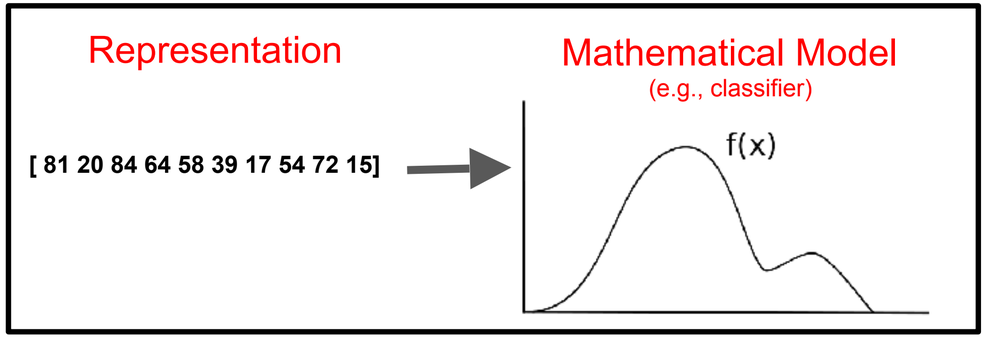
I dare not speak of what I have done. Such twisted thoughts overtook my mind and now I am sorry to say that *I have done the deed.* I have murdered the King of Scotland, King Duncan.

After naming me the *Worthy Thane of Cawdor*, this is how I repay him. I have betrayed him in the most unimaginable way a person possibly could, and I've been disloyal to him, just like I have to Banquo, whom I have lost as a dear friend. I wish that I had never done such a treacherous thing, as I am afraid that *I shall sleep no more.*

I was waiting anxiously for my Lady to sound the bell that called me to do the deed. But before she did, a symbol of the supernatural appeared before my eyes. *The dagger of the mind* captured me and the handle was to my hand yet I couldn't grasp it, but *I could see thee still.* I knew not whether to follow or to discard it from my eyes, but the *false creation* remained.

As I stepped closer to Duncan's room, I thought that I would panic and freeze, but when I got nearer, a sickening thought made me feel like I was doing the right thing! As soon as I heard the bell I knew that it was the bell summoning me.

I heard him pleading as the dagger pierced through his skin,



Macbeth was guilty.

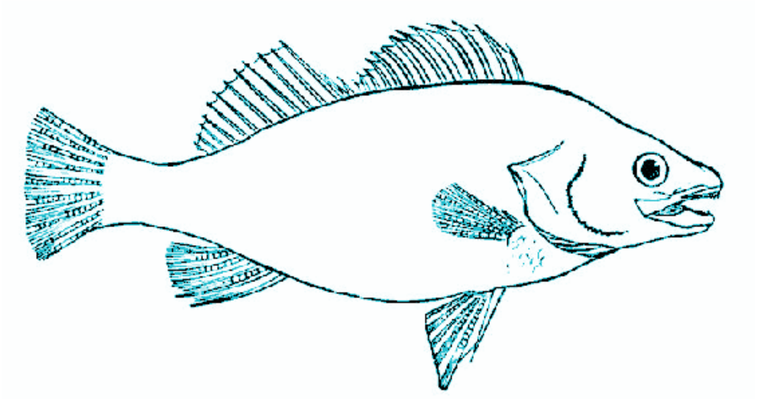
I dare not speak of what I have done. Such twisted thoughts overtook my mind and now I am sorry to say that *I have done the deed.* I have murdered the King of Scotland, King Duncan.

After naming me the *Worthy Thane of Cawdor*, this is how I repay him. I have betrayed him in the most unimaginable way a person possibly could, and I've been disloyal to him, just like I have to Banquo, whom I have lost as a dear friend. I wish that I had never done such a treacherous thing, as I am afraid that *I shall sleep no more.*

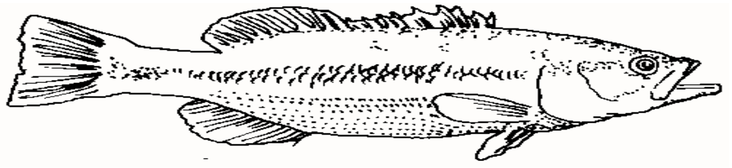
I was waiting anxiously for my Lady to sound the bell that called me to do the deed. But before she did, a symbol of the supernatural appeared before my eyes. *The dagger of the mind* captured me and the handle was to my hand yet I couldn't grasp it, but *I could see thee still.* I knew not whether to follow or to discard it from my eyes, but the *false creation* remained.

As I stepped closer to Duncan's room, I thought that I would panic and freeze, but when I got nearer, a sickening thought made me feel like I was doing the right thing! As soon as I heard the bell I knew that it was the bell summoning me.

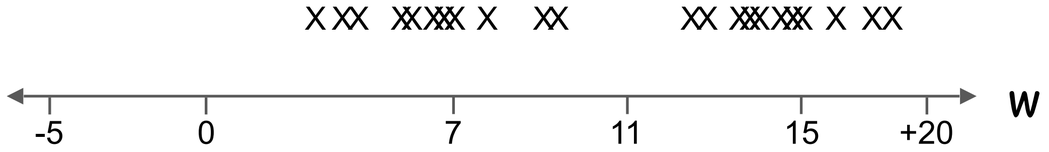
I heard him pleading as the dagger pierced through his skin,

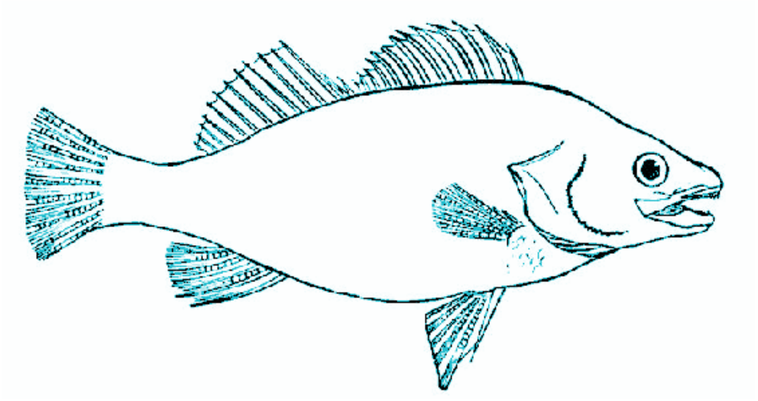


~12 lbs

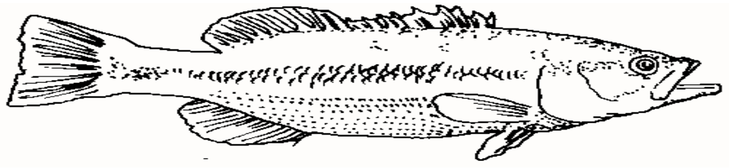


~8 lbs

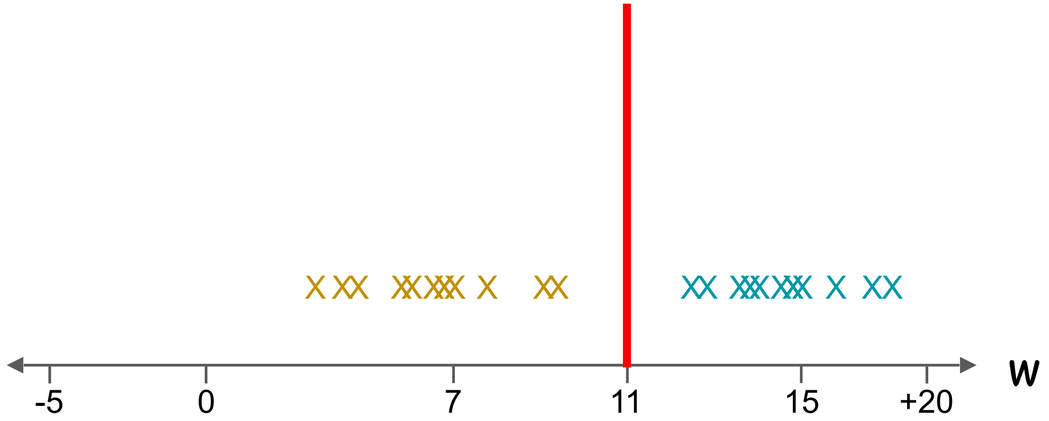




~12 lbs

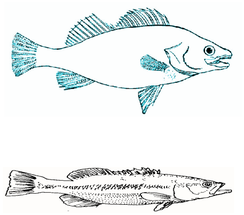


~8 lbs



~12 lbs

~8 lbs



Representation

Weight (w)

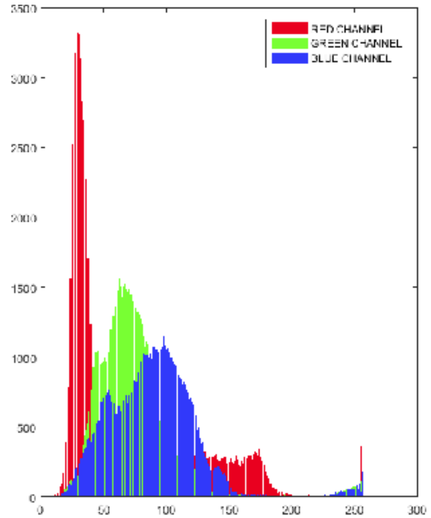
Mathematical Model (Classifier)

w>11

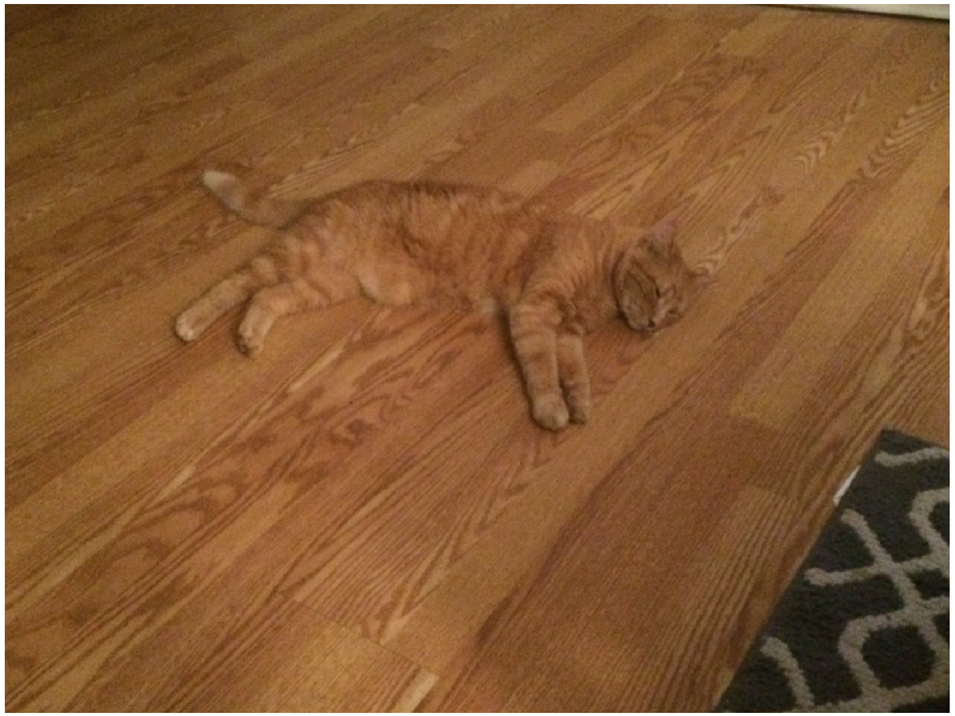
Type A

Type B

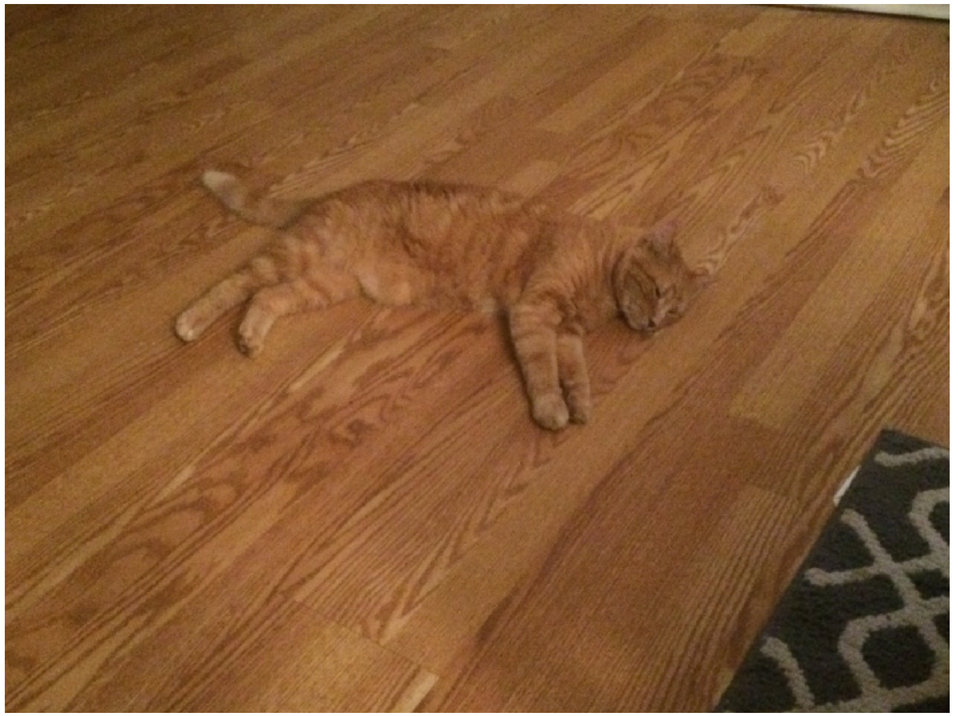
Represent these cats for a cat detector!

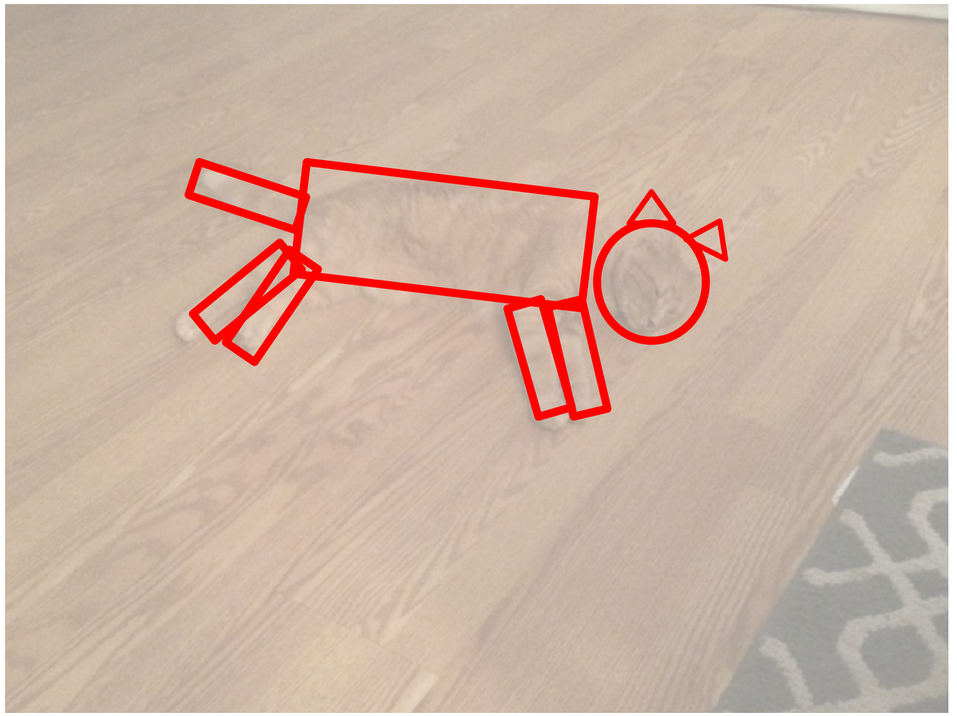


Represent these cats for a cat detector! (II)

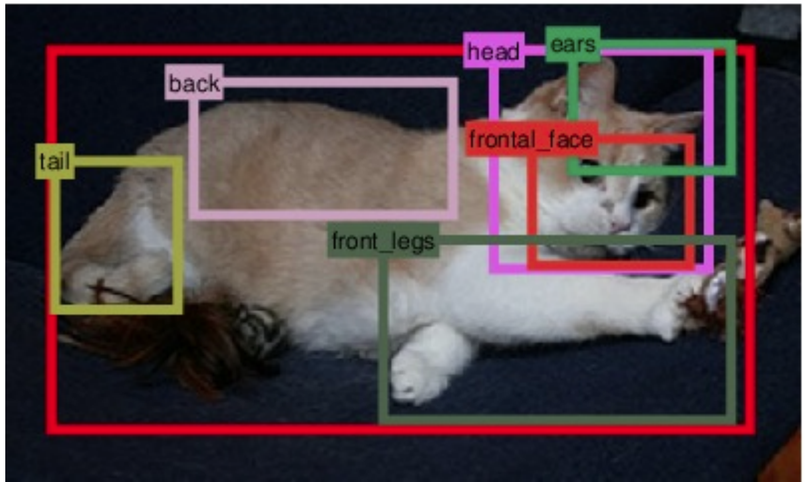


Represent these cats for a cat detector! (II)

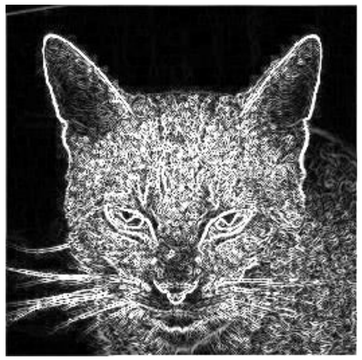


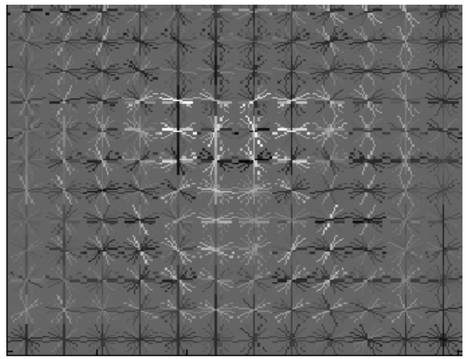


Represent these cats for a cat detector! (III)

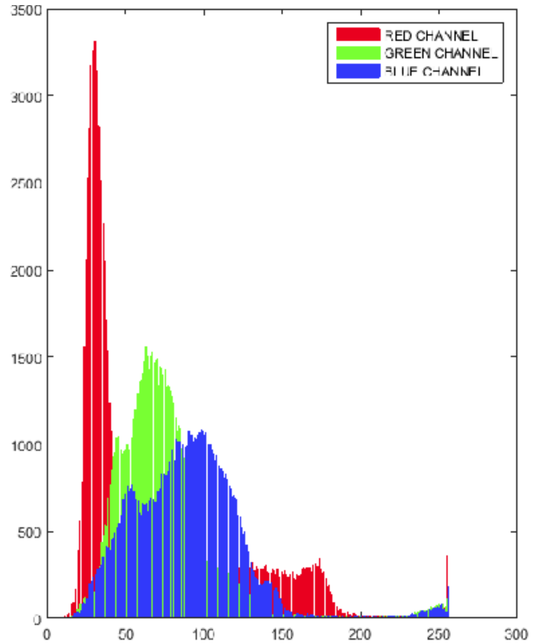


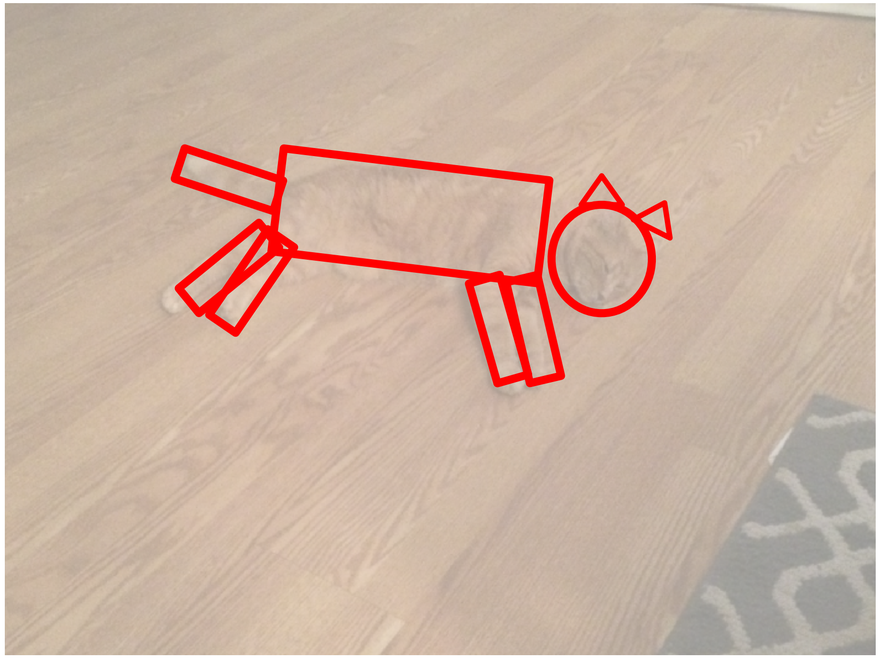
Represent these cats for a cat detector! (IV)





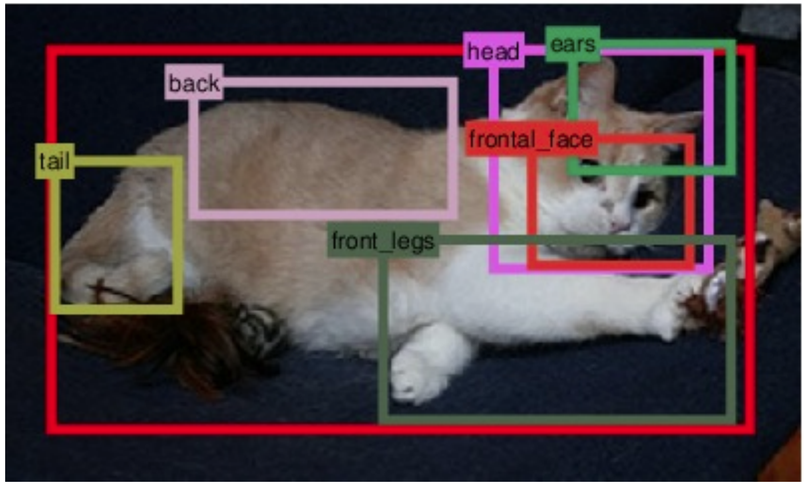
Color Histograms

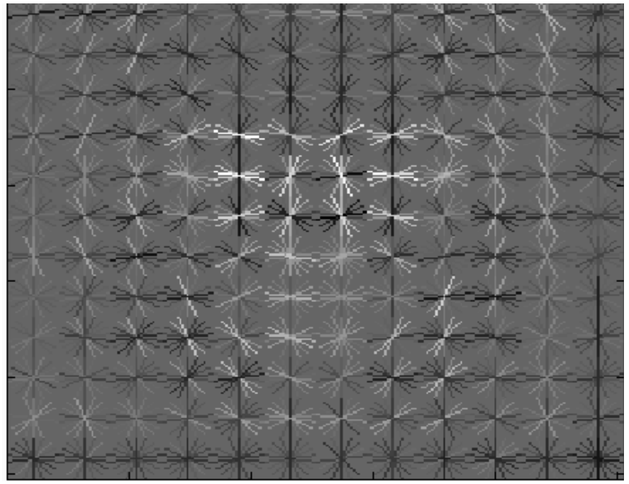




Models based Shapes

Deformable Part based Models (DPM)





Histogram of Gradients (HOG)

Felzenszwalb et al. 2010.
Dalal and Triggs, 2005.
Beis and Lowe, 1997.

Not always as easy (Happy vs Sad)

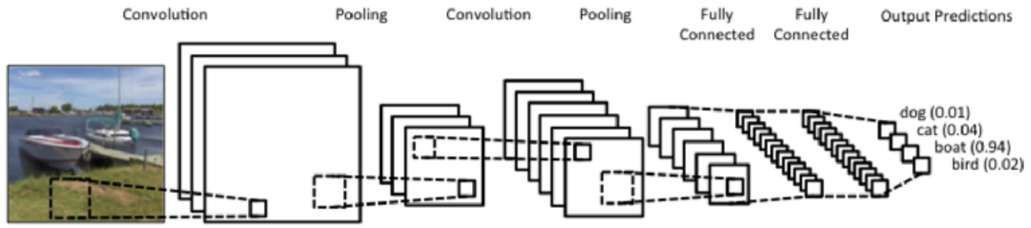
Not always as easy (Sad)

Learning Representations

Two approaches to representation learning

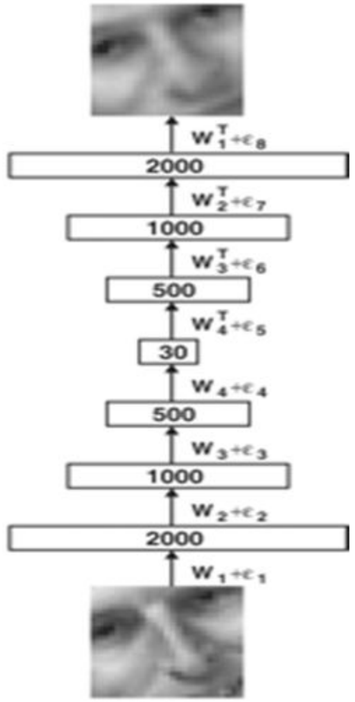
Supervised

Representation constrained on observed output (task(s)).



Unsupervised

Representation constrained on inherent data regularities.



LeCun et al. 1998.
Hinton et al. 2006.

Two approaches to representation learning

Supervised

Representation constrained on observed output (task(s)).

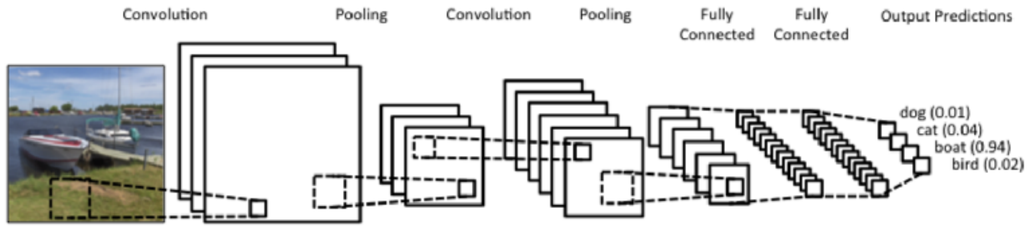
Unsupervised

Representation constrained on inherent data regularities.

Two approaches to representation learning

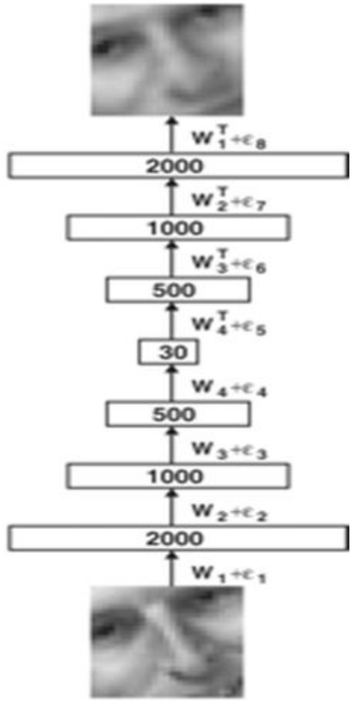
Supervised

Representation constrained on observed output (task(s)).



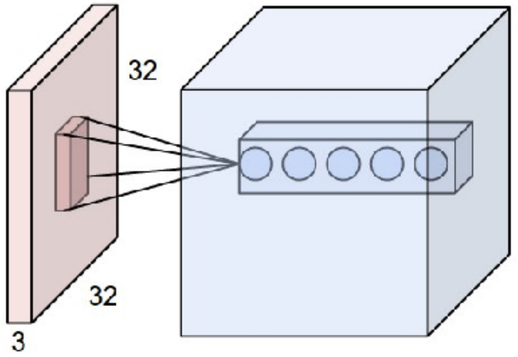
Unsupervised

Representation constrained on inherent data regularities.

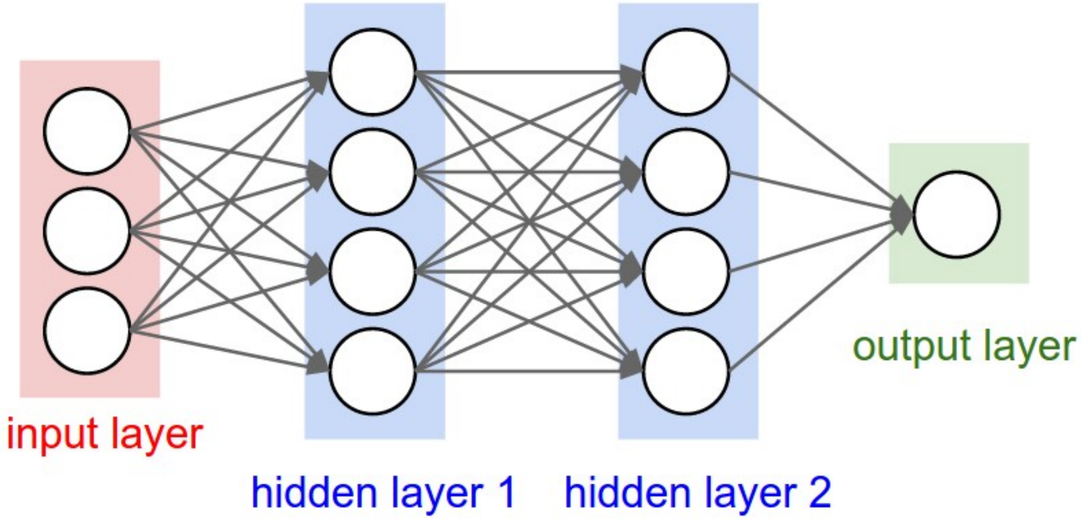


LeCun et al. 1998.
Hinton et al. 2006.

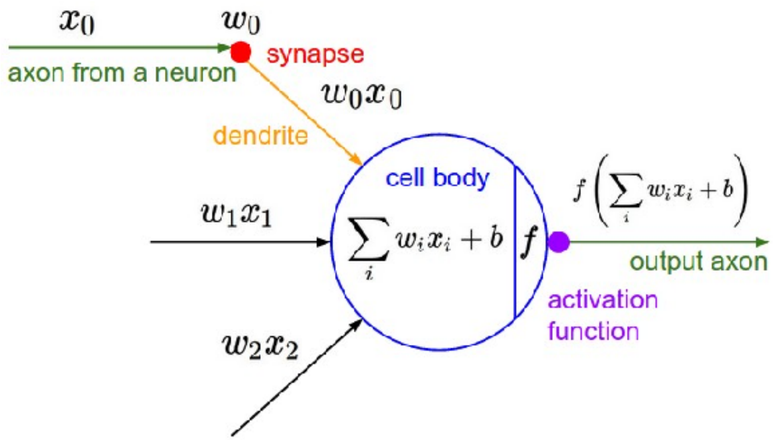
Neural Networks



Convolutional *X*

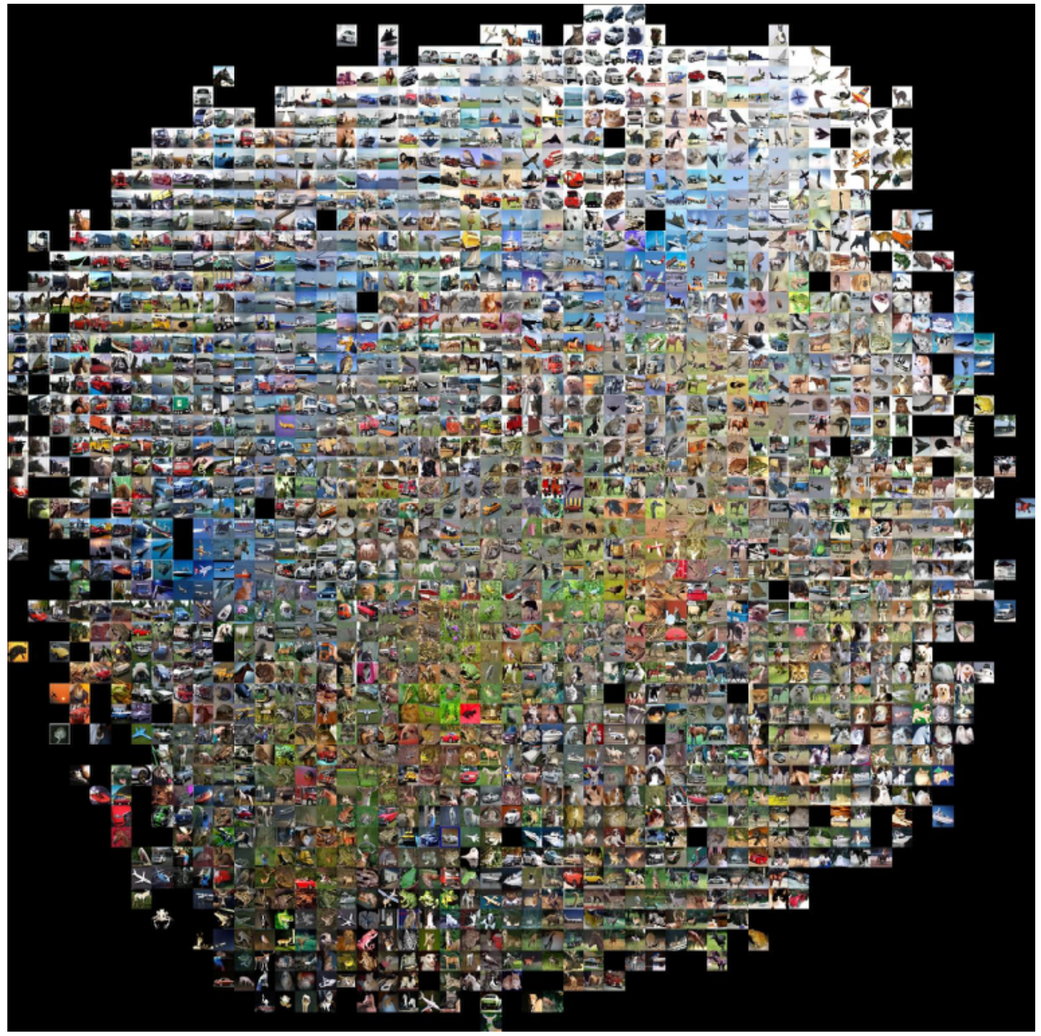


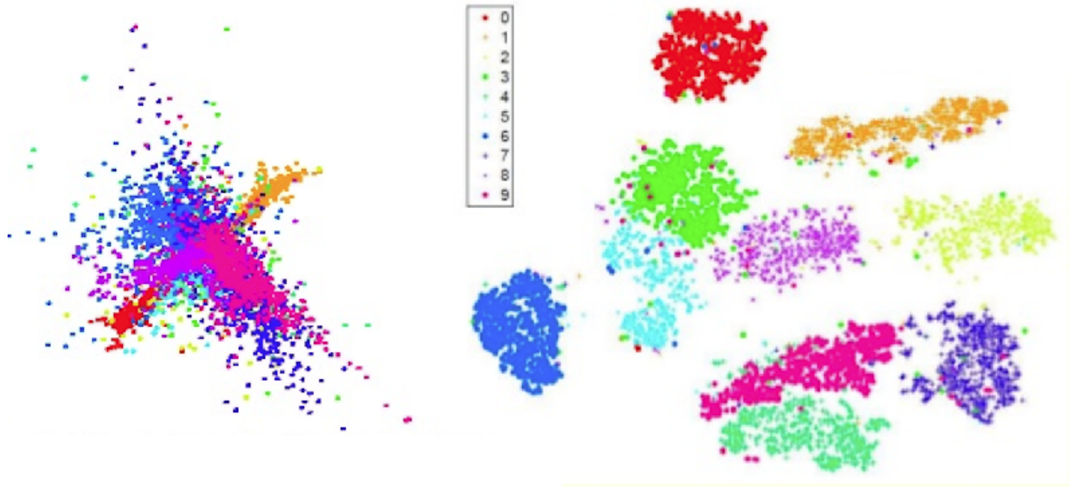
Neural Net



A Neuron

Understanding Representations





Embedding

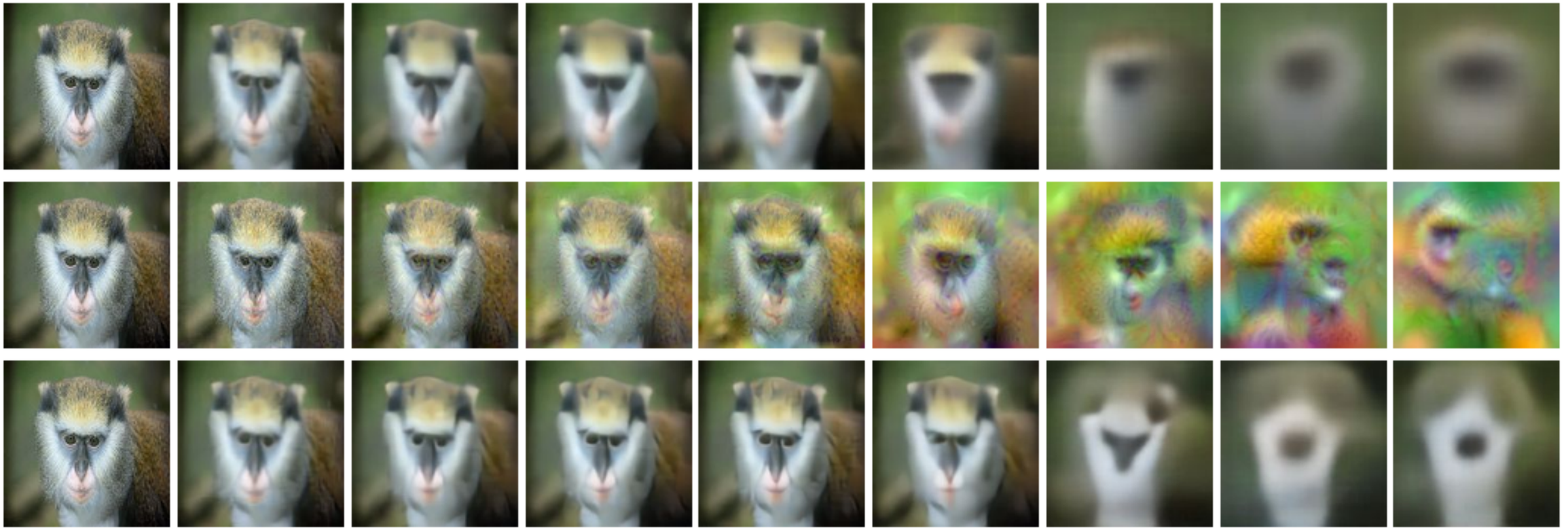
Maaten and Hinton, 2008.

Understanding Representations

[81 20 84 64 58 39 17 54 72 15]

Inverting a representation

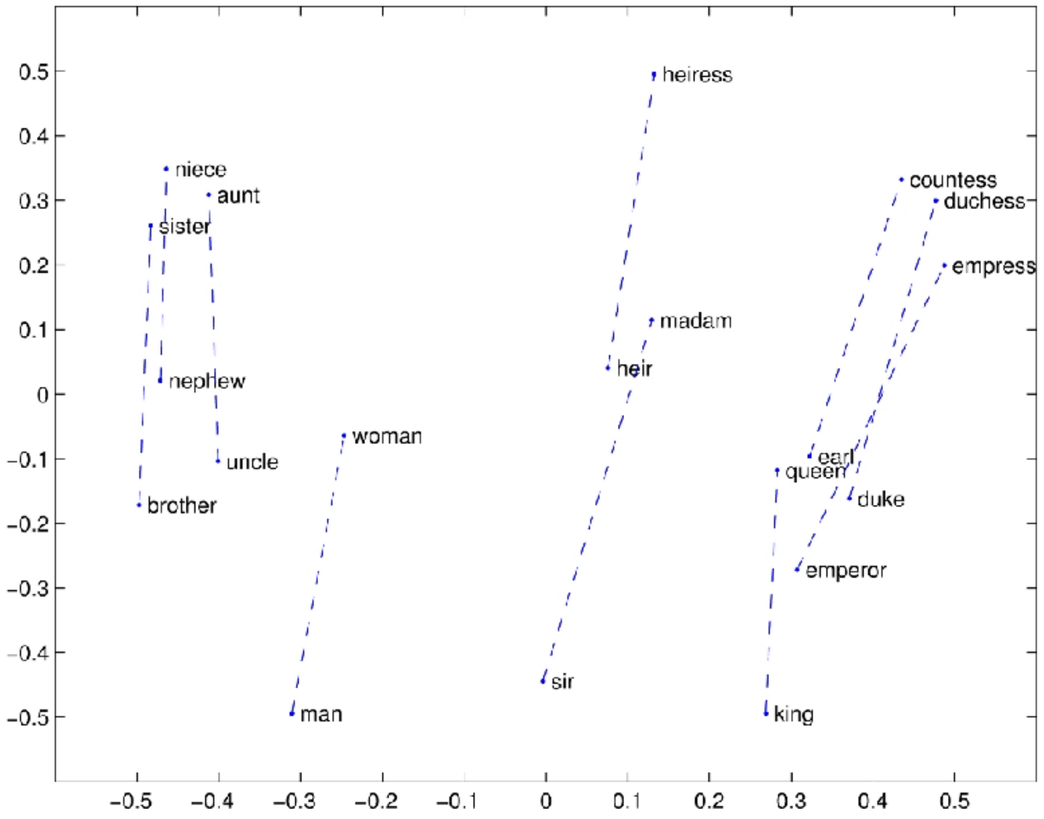
Understanding Representations



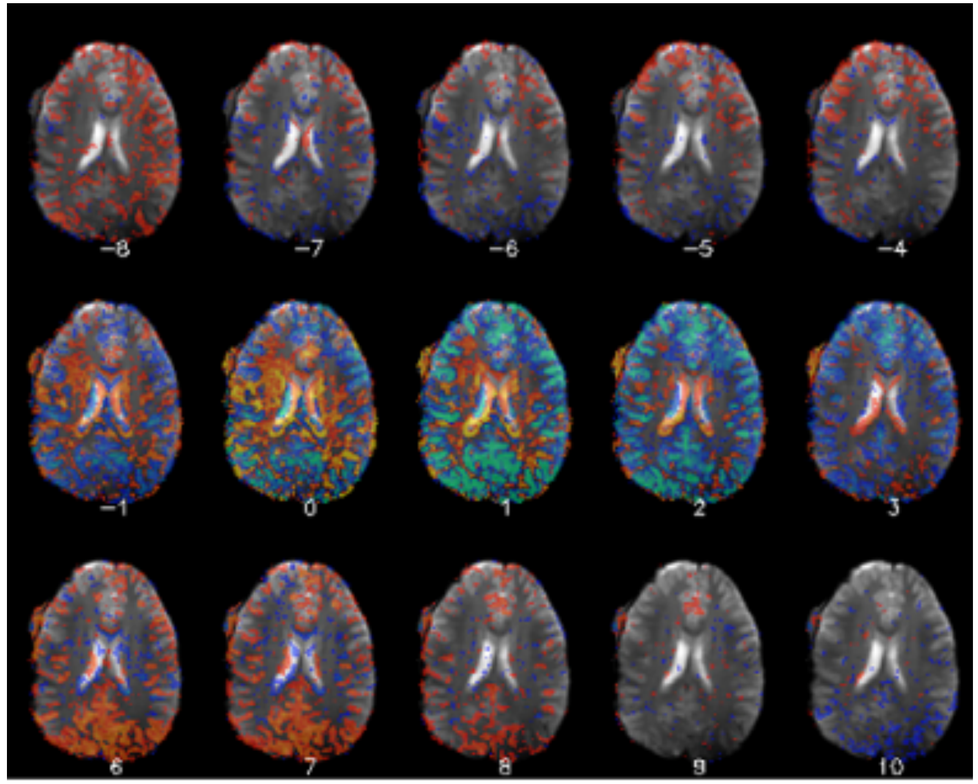
Inverting a representation

Dosovitskiy and Brox, 2015.

Representations in NLP, Brain, Speech, etc.

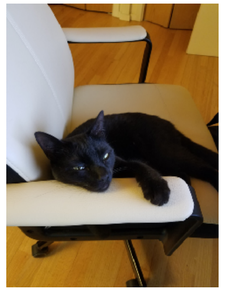


Word2Vec (NLP)



FMRI Scan (brain)

Mikolov et al. 2013



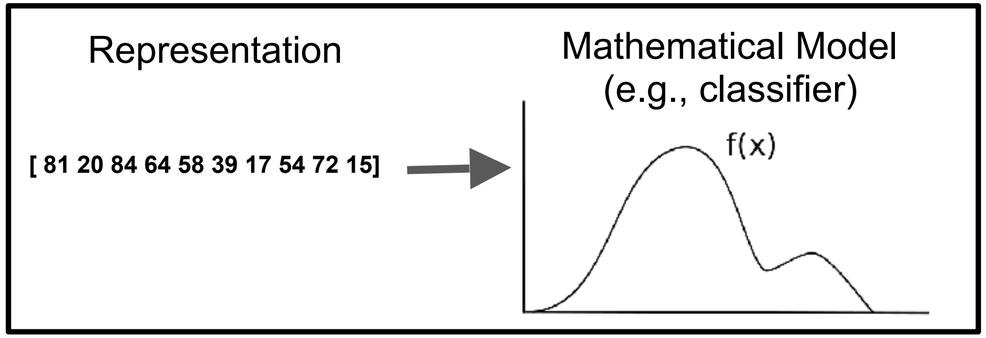
I dare not speak of what I have done. Such twisted thoughts overtook my mind and now I am sorry to say that *I have done the deed.* I have murdered the King of Scotland, King Duncan.

After naming me the *Worthy Thane of Cawdor*, this is how I repay him. I have betrayed him in the most unimaginable way a person possibly could, and I've been disloyal to him, just like I have to Banquo, whom I have lost as a dear friend. I wish that I had never done such a treacherous thing, as I am afraid that *I shall sleep no more.*

I was waiting anxiously for my Lady to sound the bell that called me to do the deed. But before she did, a symbol of the supernatural appeared before my eyes. *The dagger of the mind* captured me and the handle was to my hand yet I couldn't grasp it, but *I could see thee still.* I knew not whether to follow or to discard it from my eyes, but the *false creation* remained.

As I stepped closer to Duncan's room, I thought that I would panic and freeze, but when I got nearer, a sickening thought made me feel like I was doing the right thing! As soon as I heard the bell I knew that it was the bell summoning me.

I heard him pleading as the dagger pierced through his skin,



Macbeth was guilty.

CS331b **CS229**

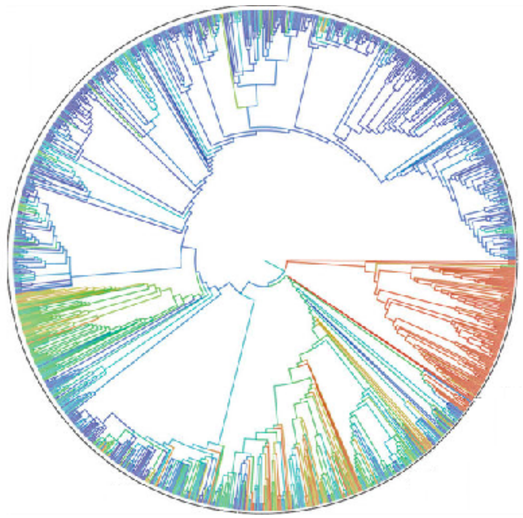
CS231n

Now that we’re done with building some background…
Some recent directions

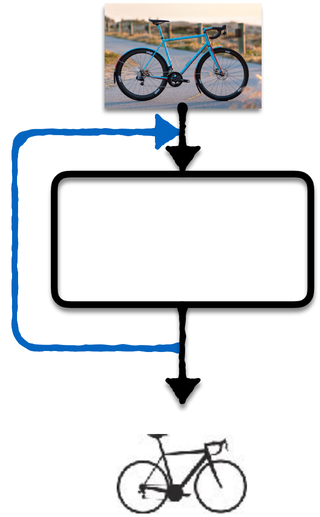
Intelligence as Efficiency

- Yielding higher value for less resources.

Efficient at learning



Efficient at testing

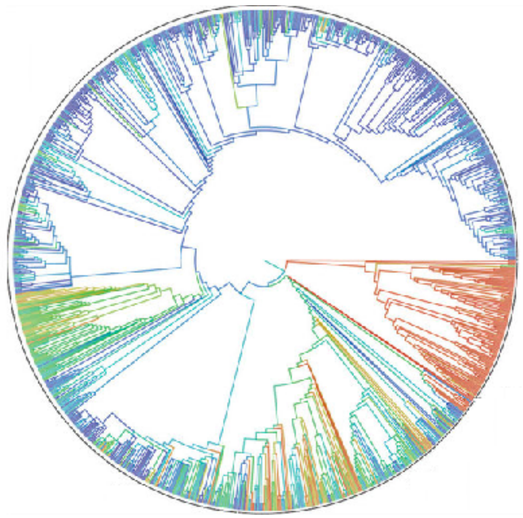


Yudkowsky, Efficient Cross-Domain Optimization, 2008.

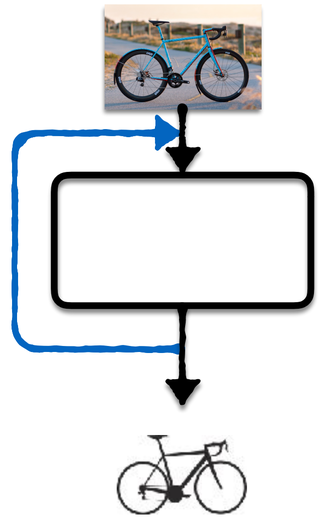
Intelligence as Efficiency

- Yielding higher value for less resources.

Efficient at learning



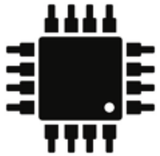
Efficient at testing



Yudkowsky, Efficient Cross-Domain Optimization, 2008.

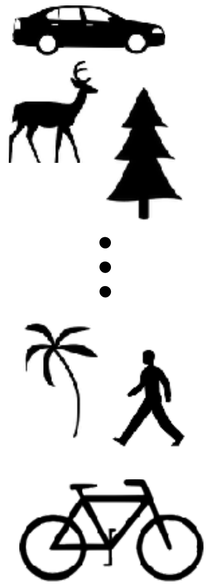
Practice: Prediction on a Budget

Time



Resources

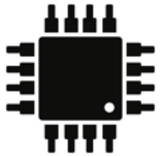
65 mph



1 second
40 meters

Practice: Prediction on a Budget

Time



Resources

A tradeoff game

65 mph

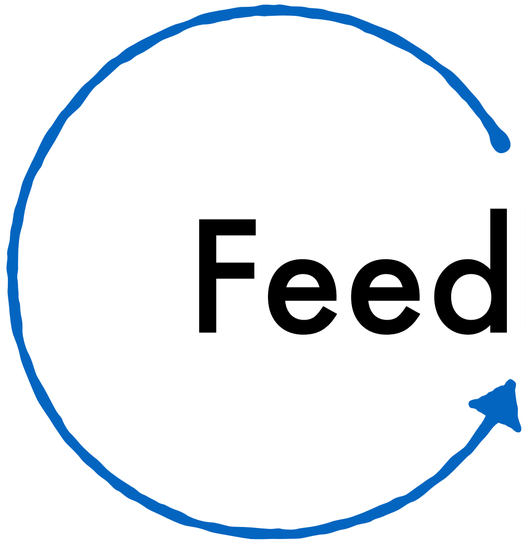
✗ foliage

tandem bike

✓ wheeled vehicle

road bike

flat bar bike



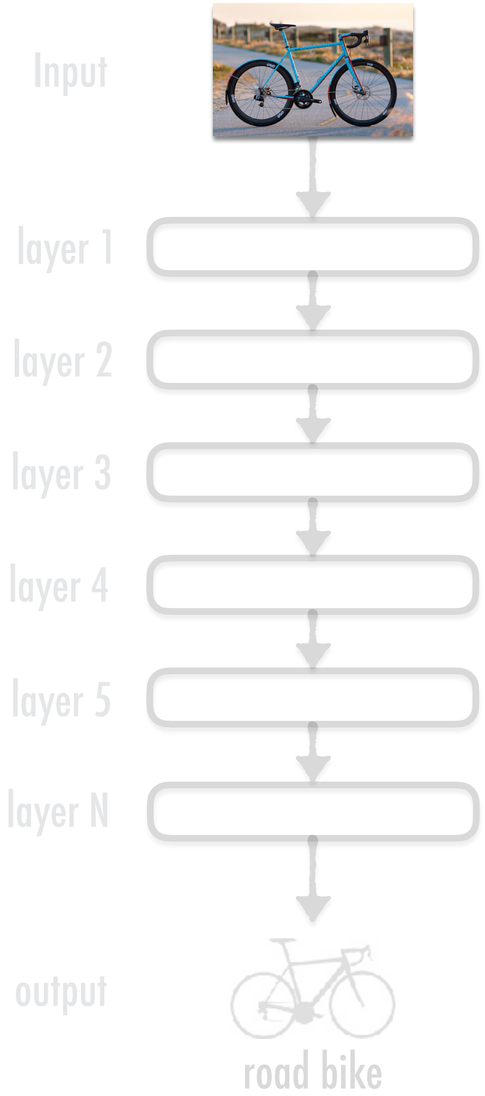
Feedback networks

Amir Zamir, Te-Lin Wu*, Lin Sun, William Shen, Bertram Shi,*

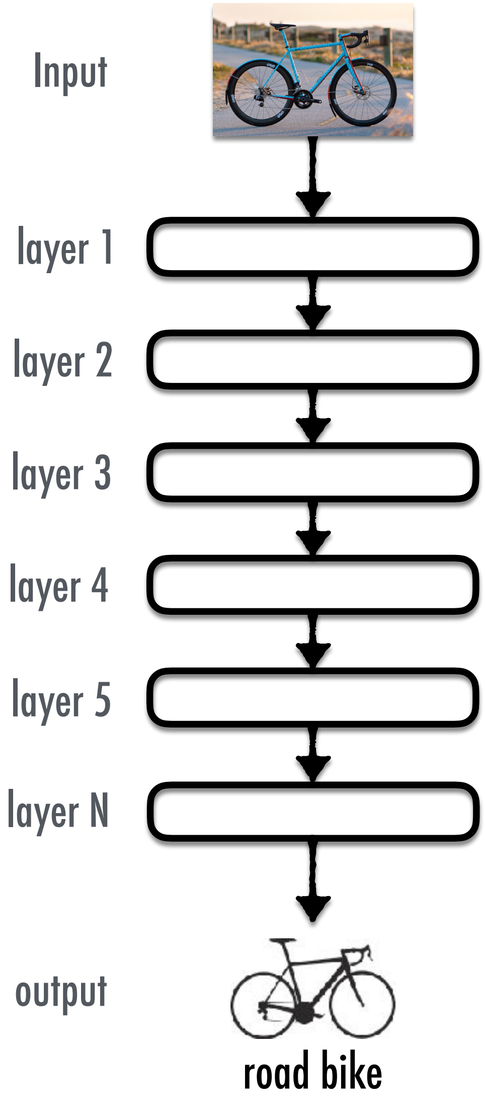
Jitendra Malik, Silvio Savarese

CVPR 2017

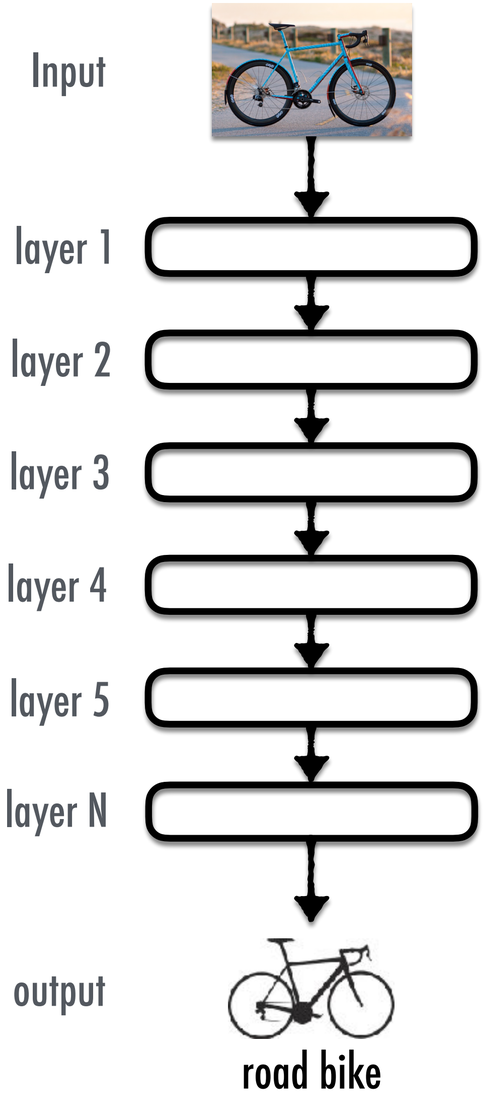
http://feedbacknet.stanford.edu



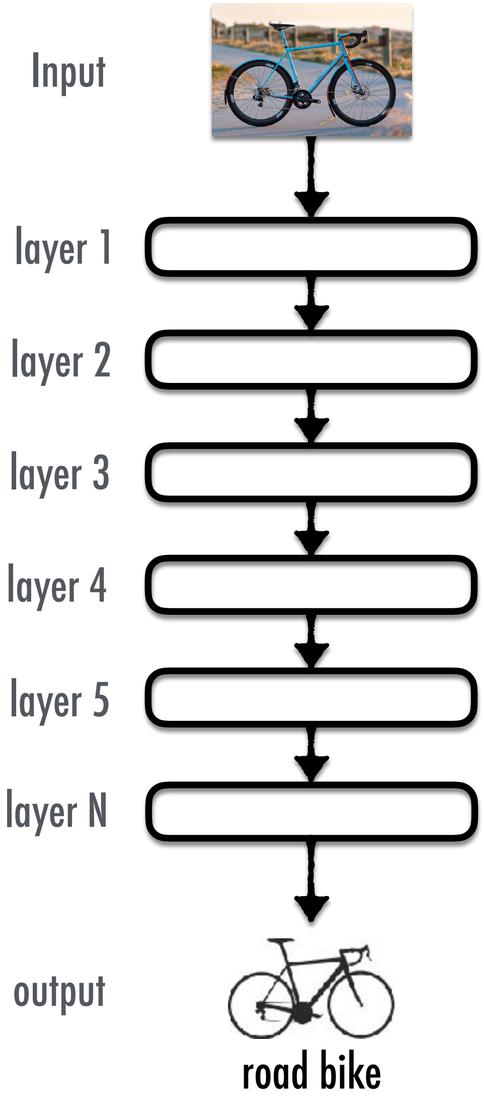
Feedforward model.

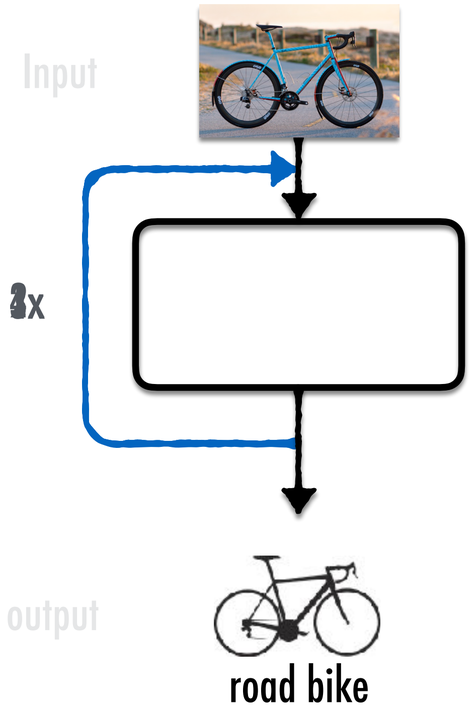


Feedforward model.



→ *Feedforward model.*

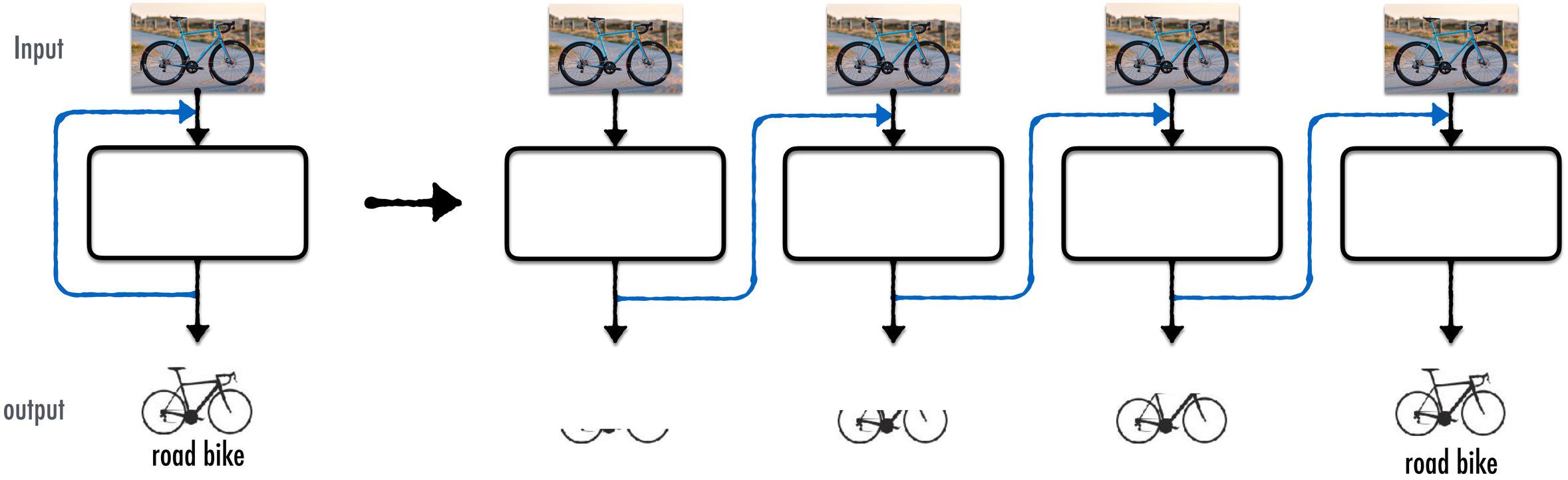




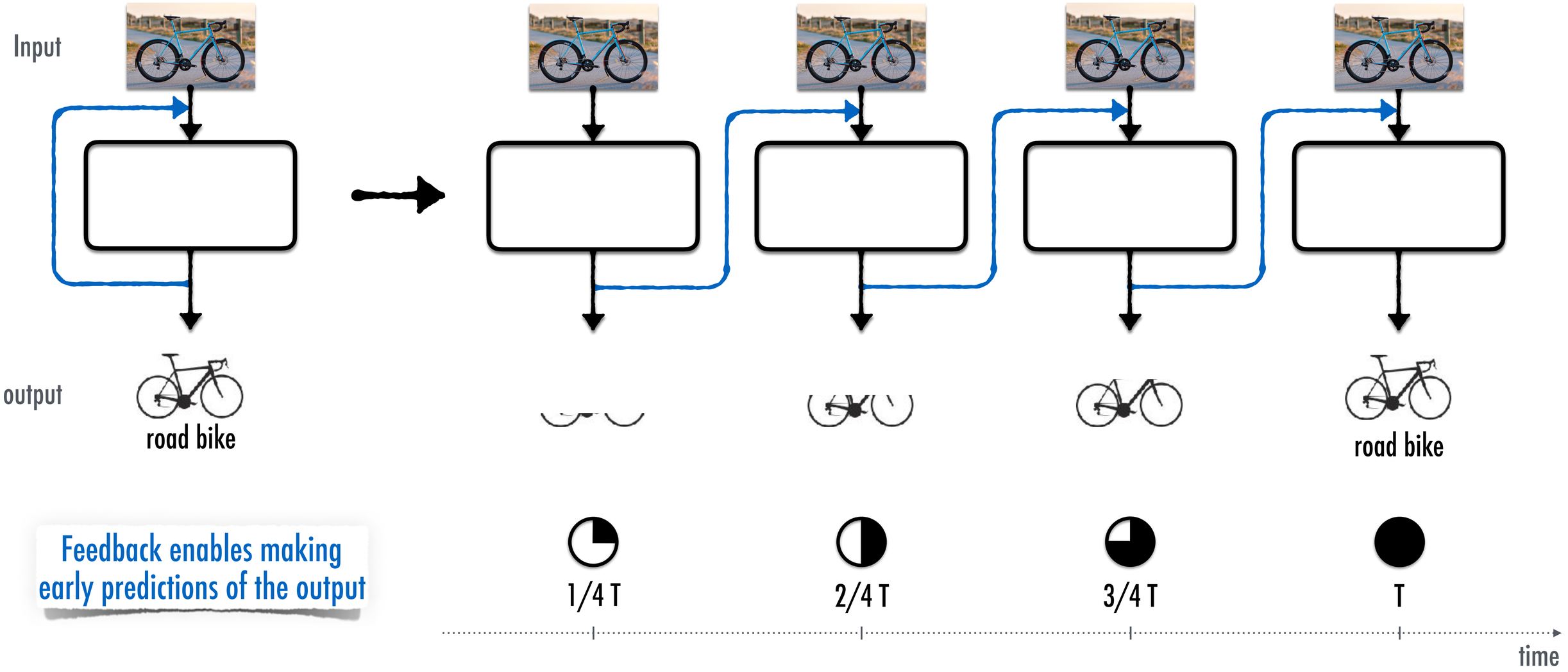
an alternative with several advantages

Feedforward model.

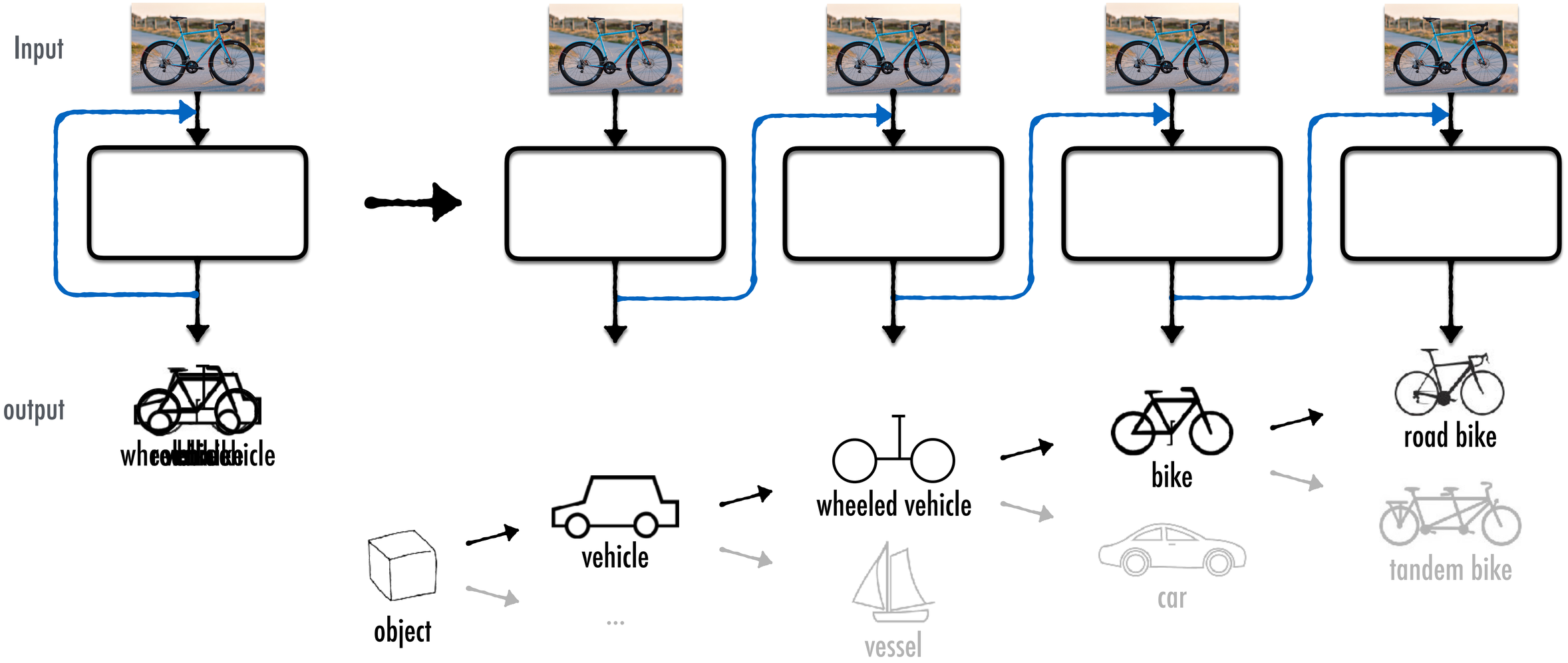
Feedback model.



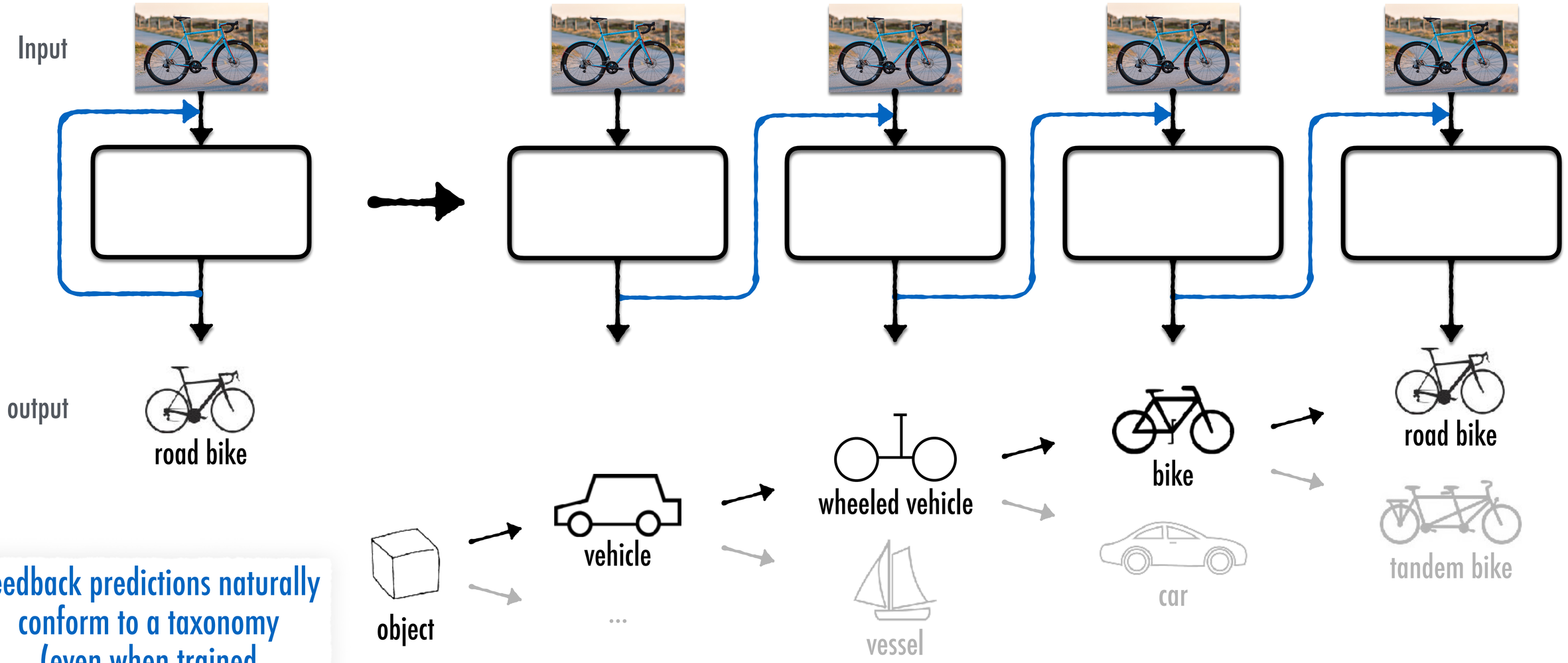
Feedback model unrolled.



Advantage I: Early Prediction



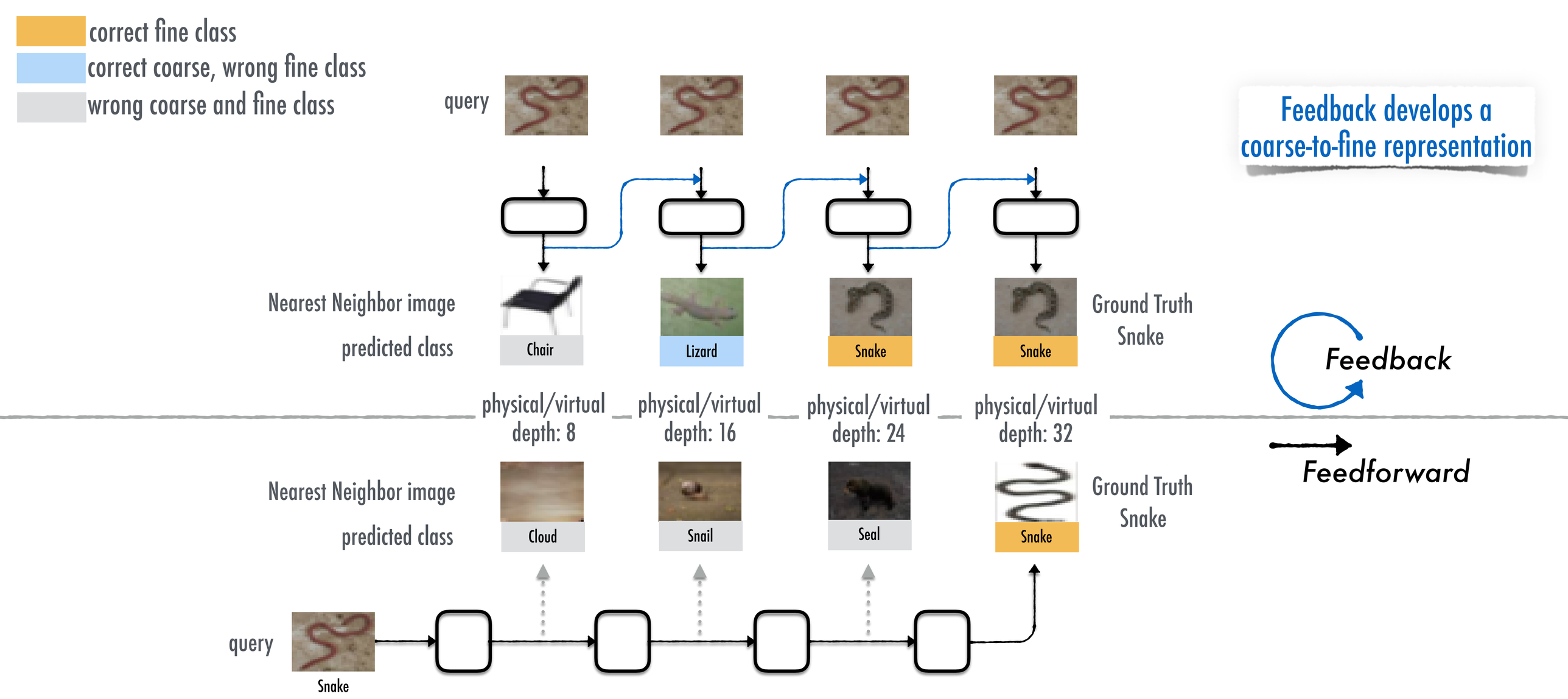
Advantage II: Taxonomic Prediction



Feedback predictions naturally conform to a taxonomy (even when trained without a taxonomy)

Advantage II: Taxonomic Prediction

Experimental Results



Qualitative results on CIFAR100 test set

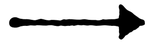
Feedback's representation is rather disentangled from early on.

Iteratio
Sugge

A coarse-to-fine representation

Feedforward's representation becomes disentangled only towards the end.

Feedback



Feedforward

"Timed-TSNE":

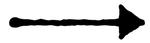
How the representation changes through the network

Feedback's representation is rather disentangled early on.
Iterations mainly form fine-grained borders.
Suggesting a coarse-to-fine representation.

A coarse-to-fine representation

Feedforward's representation becomes disentangled only towards the end.
Suggesting a low-level-to-abstract representation.

Feedback



Feedforward

"Timed-TSNE":

How the representation changes through the network

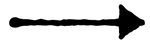
Feedback's representation is rather disentangled from early on.

Iteratio
Sugge

A coarse-to-fine representation

Feedforward's representation becomes disentangled only towards the end.

Feedback



Feedforward

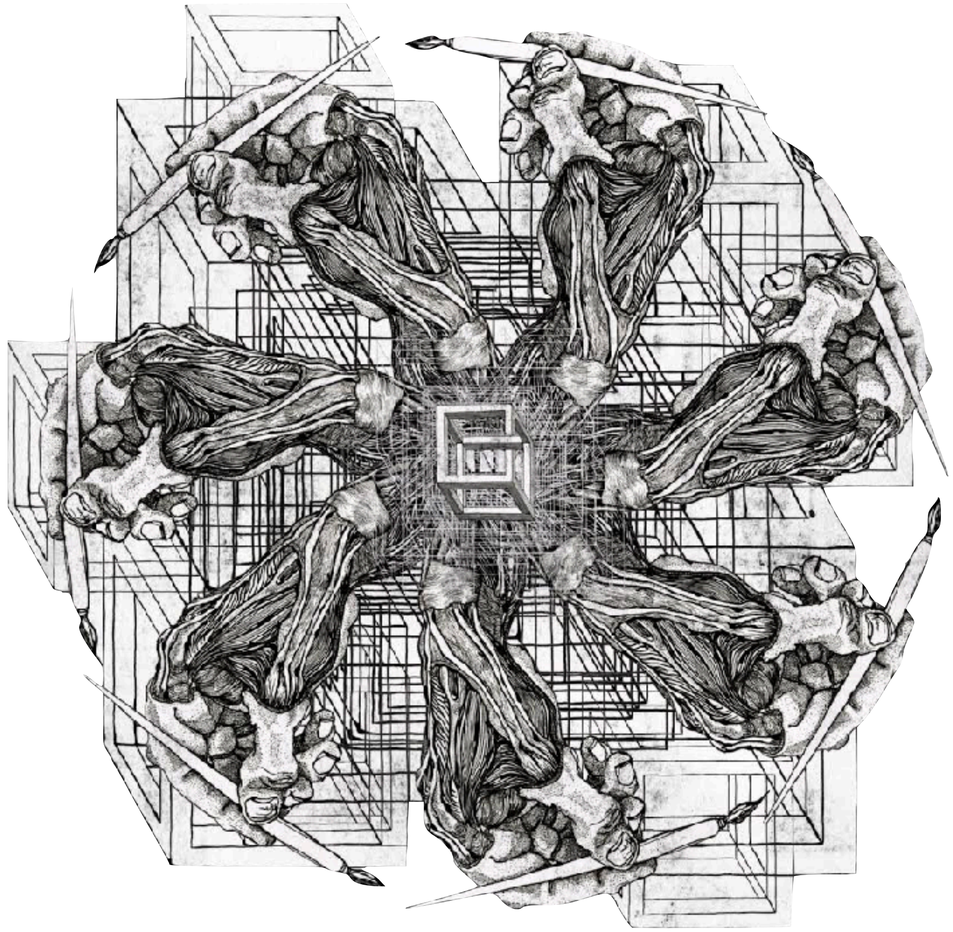
"Timed-TSNE":

How the representation changes through the network

Activations of Feedback and Feedforward networks
are largely dissimilar.
Suggesting a considerably different representation.

CIFAR100 Query Image
(snake)

Activation maps of different layers/iterations



Visual Polymath

a.k.a generic/comprehensive/broad representation/perception

The main lesson of thirty-five years of AI research is that **the hard problems are easy and the easy problems are hard**. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived...

As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers who are in danger of being displaced by robots. The gardeners, receptionists, and cooks are secure in their jobs for decades to come. [2]

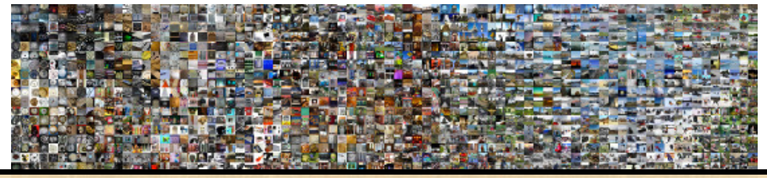
–Steven Pinker

The main lesson of thirty-five years of AI research is that **the hard problems are easy and the easy problems are hard**. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived...

As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.

–Steven Pinker

An Exciting Time!



Fully Supervised Learning

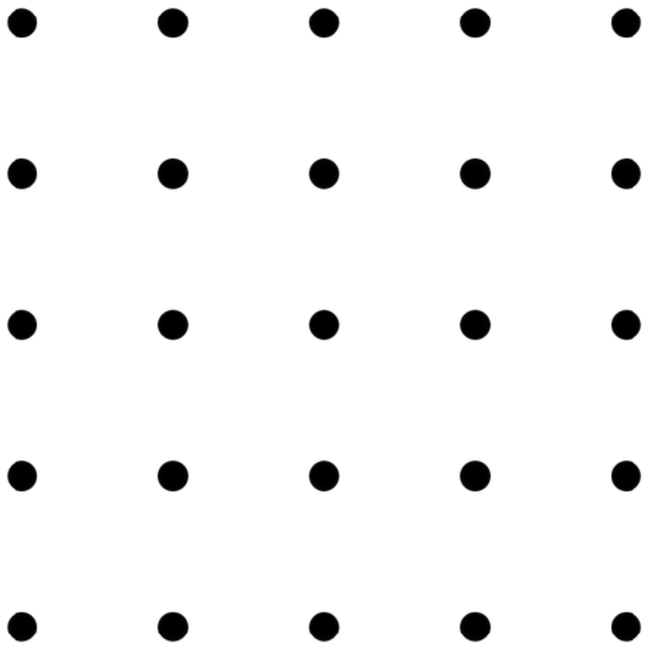
AlphaGo

IMAGENET

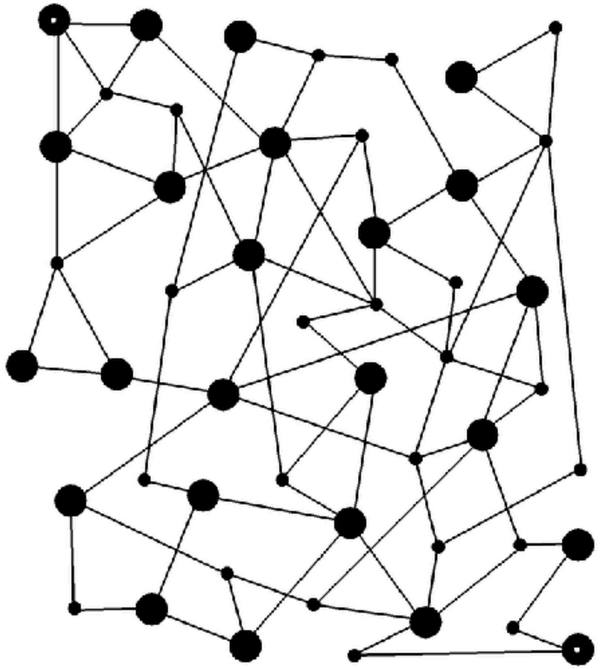
DEEPMIND

Fully Supervised Learning

Isolation ~ “Idiot Savant”

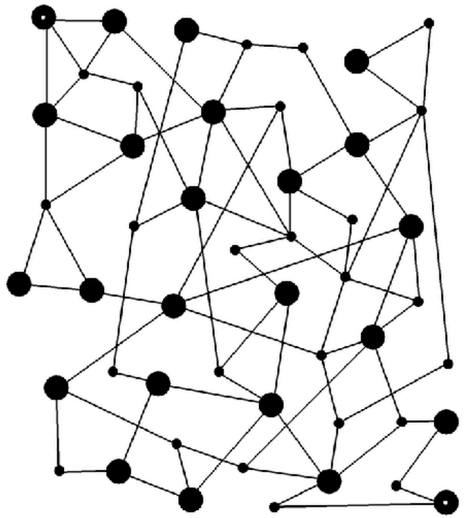


Task Interplay



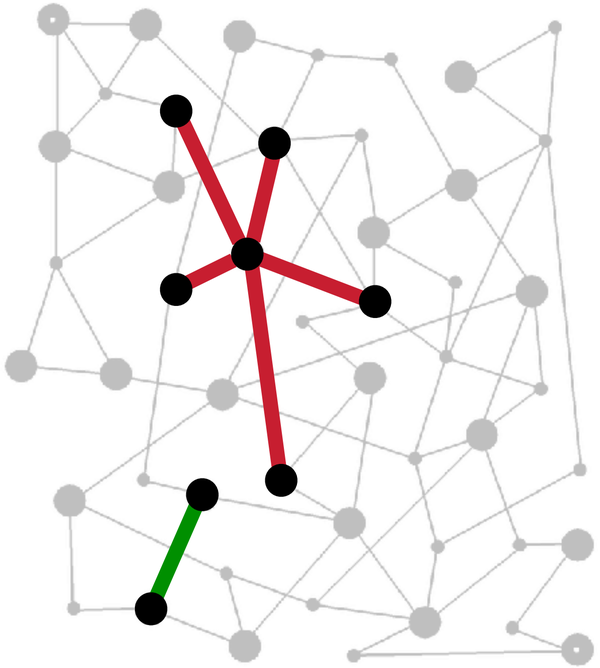
Task Interplay

Colorization → Object Detection



Zhang et al., Colorful Image Colorization, 2016.

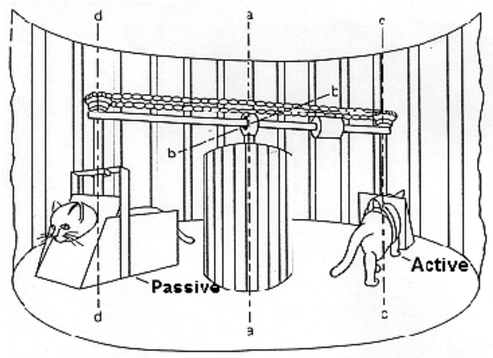
Task Interplay



- **Proposition**:
  - (Instead of providing supervision over the desired tasks)
  - Provide supervision over a **<u>set of selected foundational tasks</u>** ⇒

    generalization to novel tasks and abstraction capabilities.

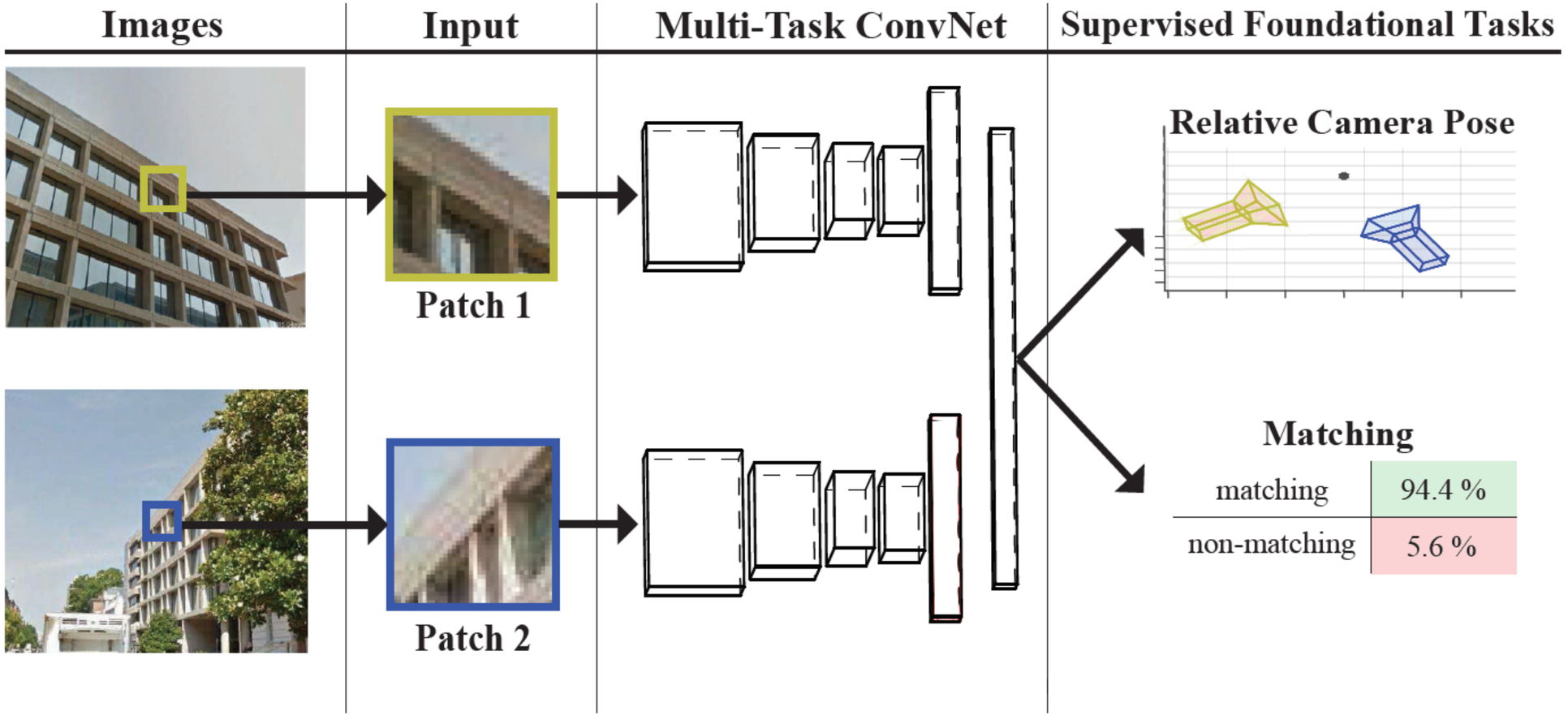
- **But how to pick the foundational tasks?**
  - Biology
  - Inspirations from **<u>developmental stages of visual skills in brain</u>**



Ostrovsky et al., 2009. Held & Hein. 1963.

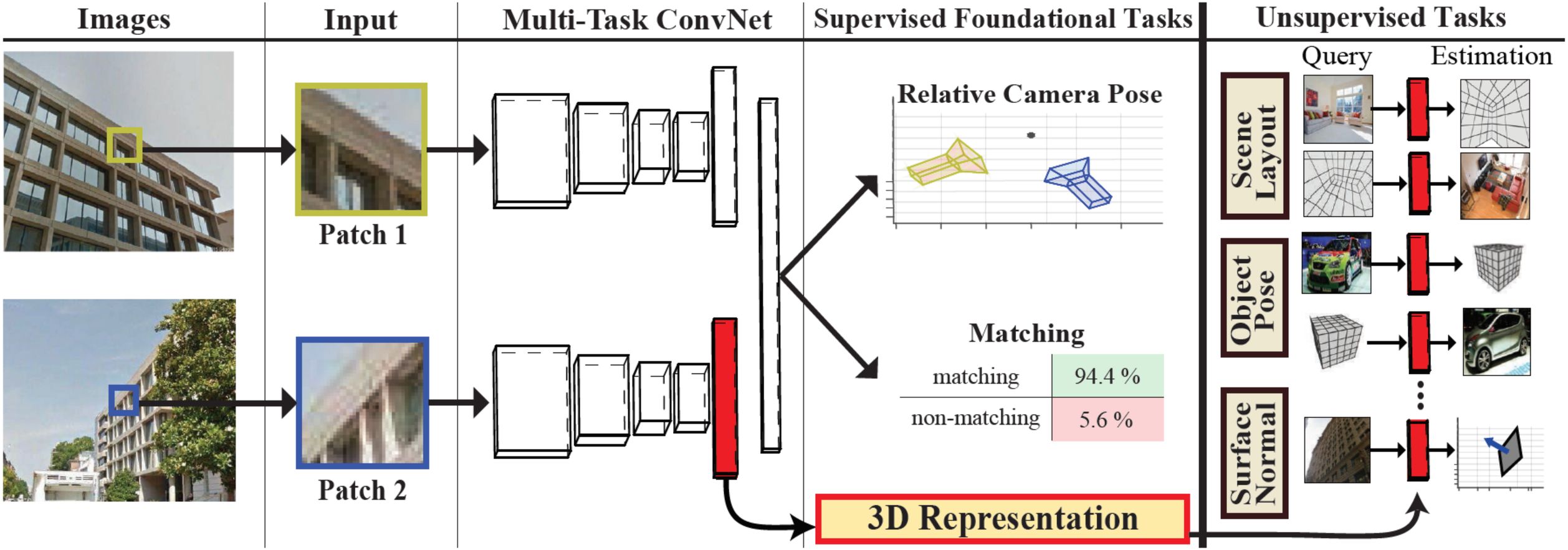
Generic 3D Representation via Pose Estimation and Matching. ECCV 2016.
Amir Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, Silvio Savarese.

Generic 3D Representation Learning



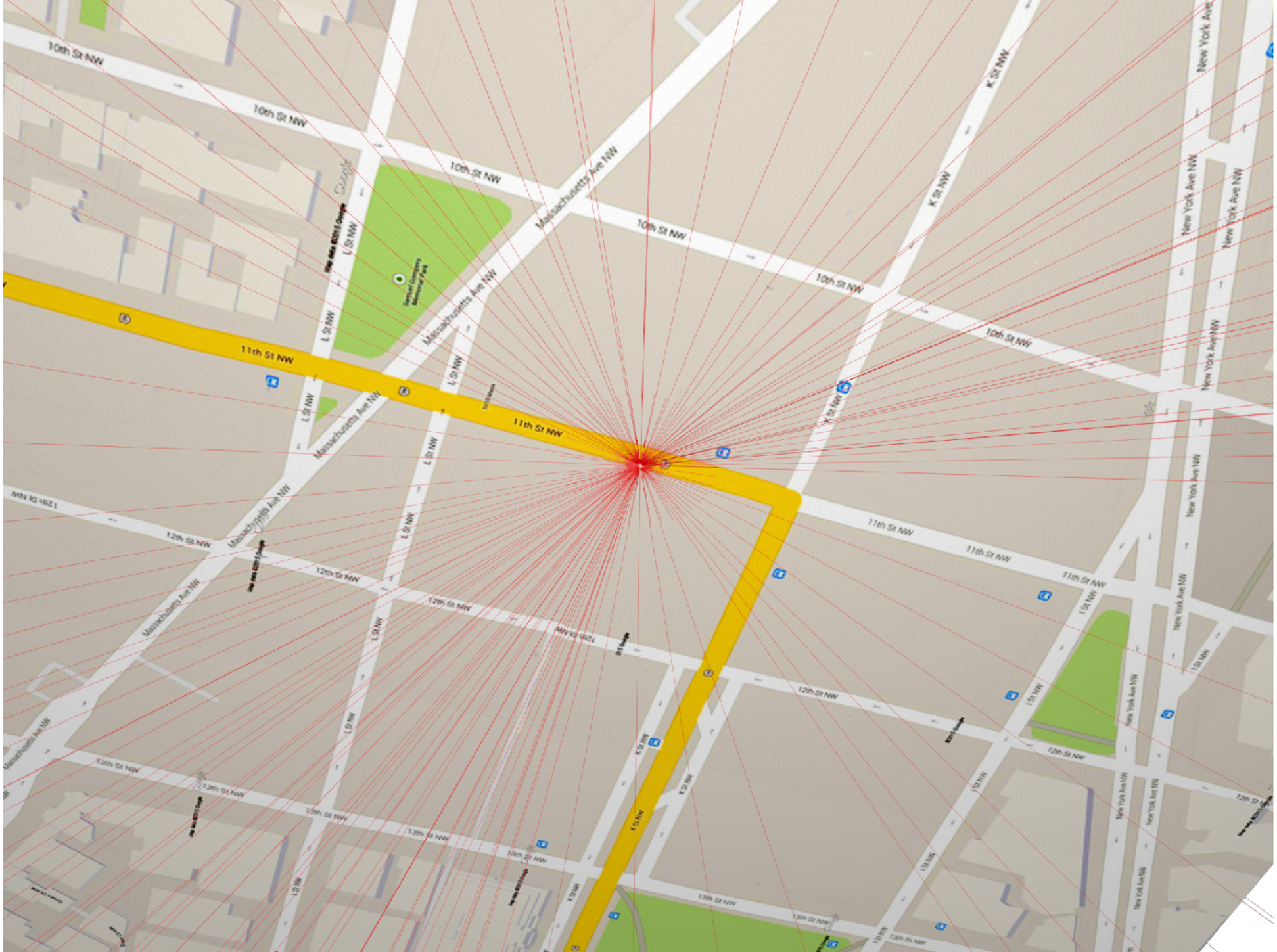
Generic 3D Representation via Pose Estimation and Matching.
Amir Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, Silvio Savarese.
ECCV 2016.

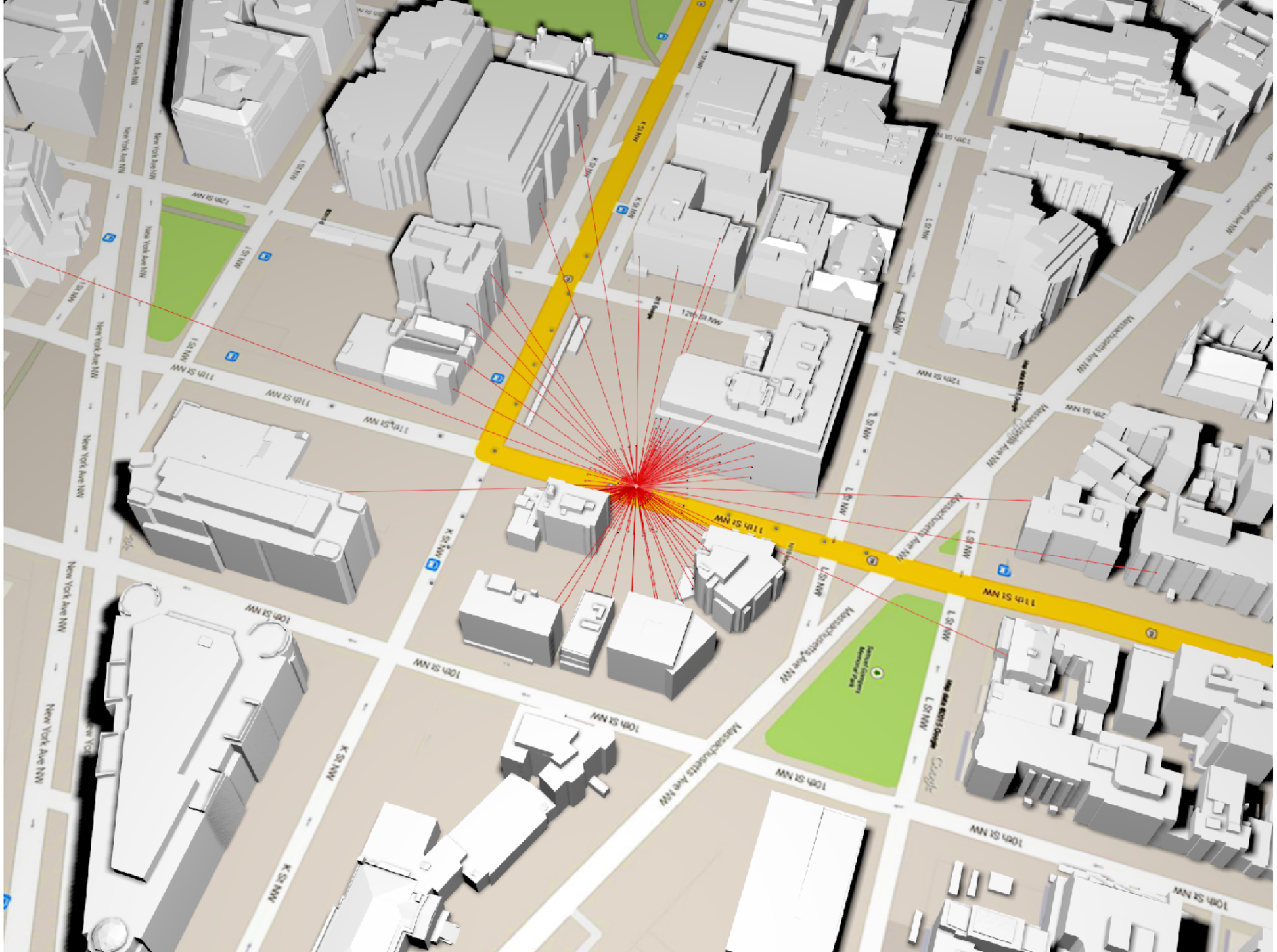
Generic 3D Representation Learning

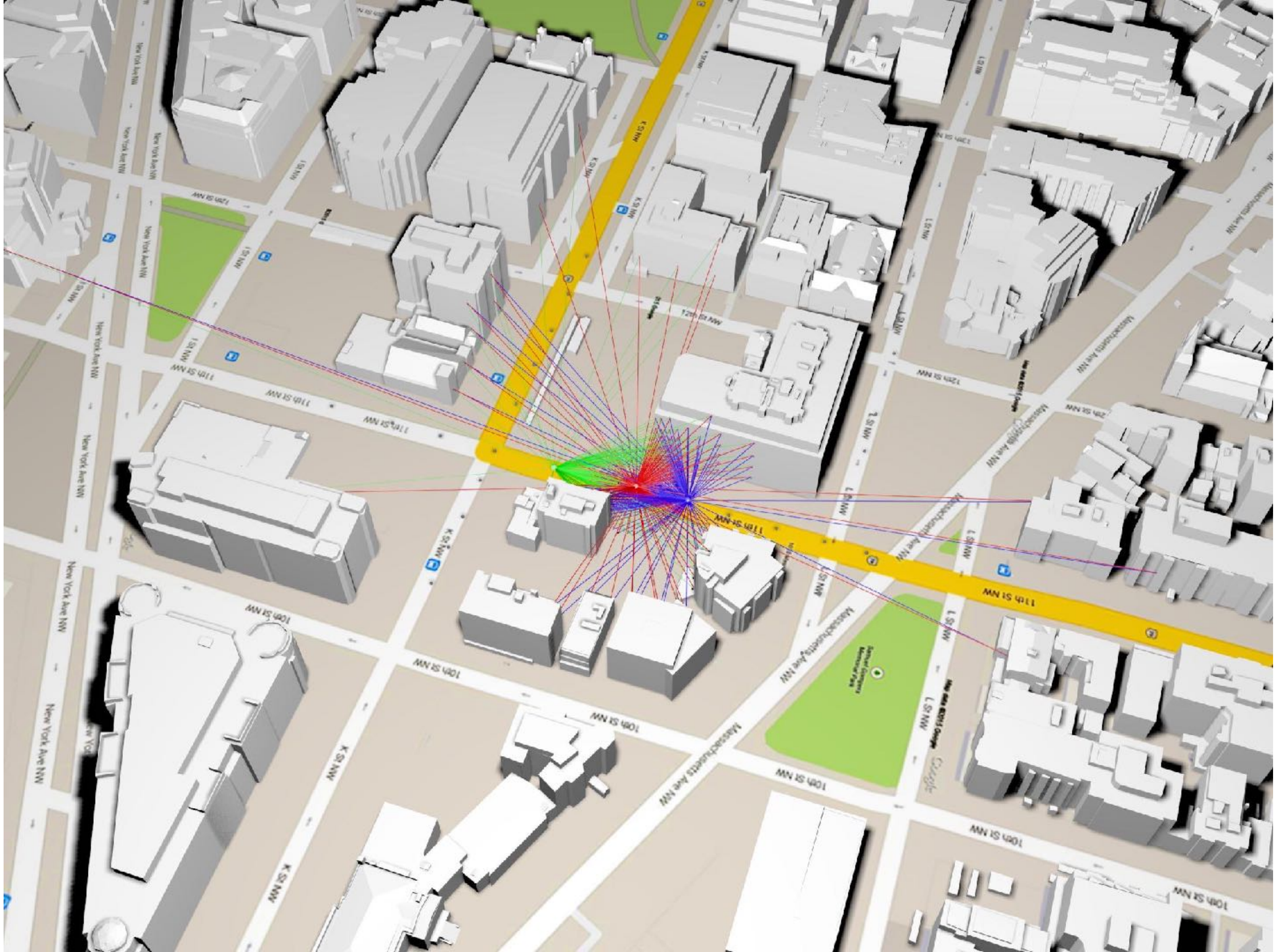


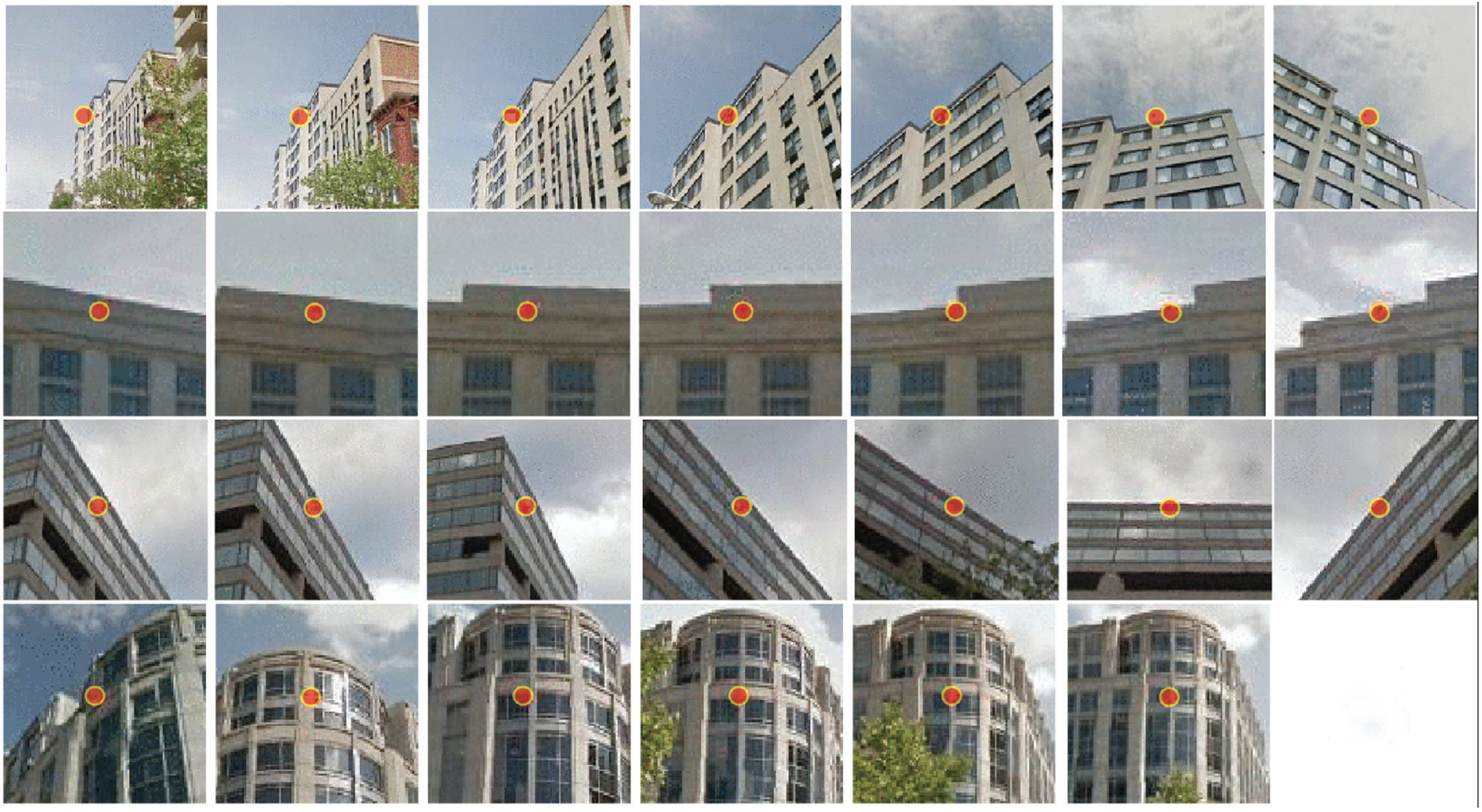
Generic 3D Representation via Pose Estimation and Matching.
Amir Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, Silvio Savarese.
ECCV 2016.

Learn it from the world!


10th St NW
10th St NW
10th St NW
10th St NW
10th St NW
11th St NW
11th St NW
11th St NW
11th St NW
11th St NW
12th St NW
12th St NW
12th St NW
12th St NW
12th St NW
13th St NW
13th St NW
13th St NW
13th St NW
Massachusetts Ave NW
Massachusetts Ave NW
Massachusetts Ave NW
Massachusetts Ave NW
L St NW
L St NW
L St NW
L St NW
K St NW
K St NW
K St NW
K St NW
New York Ave NW
New York Ave NW
New York Ave NW
New York Ave NW
Samuel Gompers Memorial Park


New York Ave NW
12th St NW
11th St NW
10th St NW
K St NW
L St NW
Massachusetts Ave NW

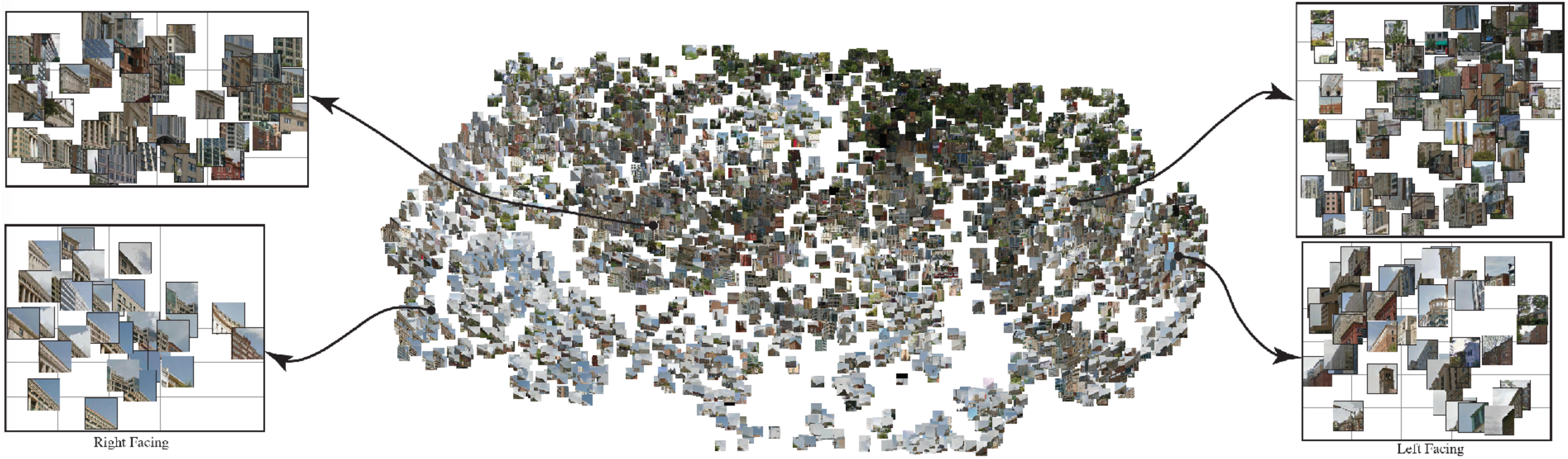

New York Ave NW
12th St NW
11th St NW
10th St NW
K St NW
L St NW
Massachusetts Ave NW

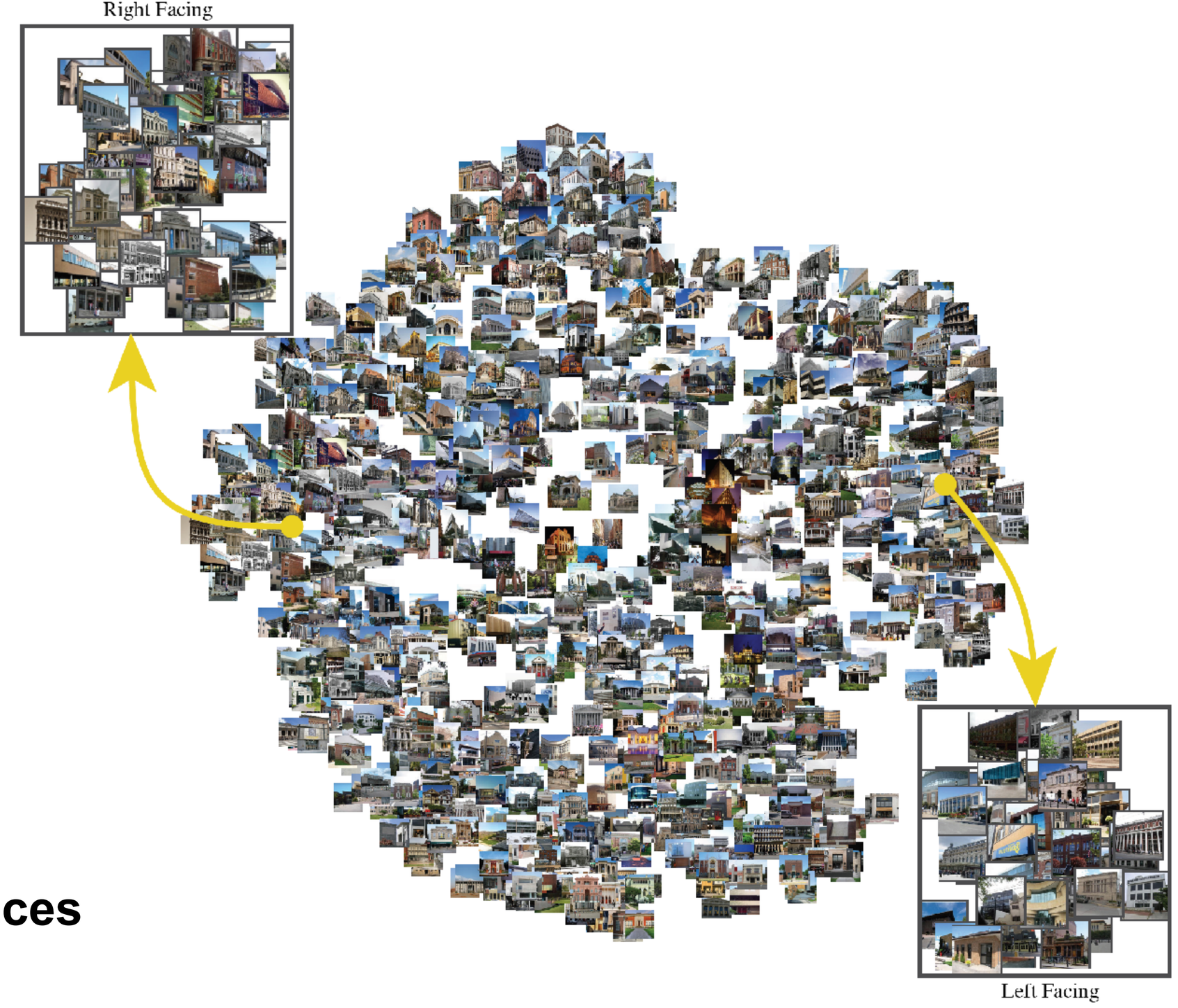


Evaluations

Supervised Tasks	Unsupervised Tasks
• Camera pose estimation • Matching (wide-baseline)	• Surface Normal • 3D Object Pose • 3D Scene Layout • Visual Abstraction
State-of-the-art **Human-level**	**State-of-the-art** **unsupervised**

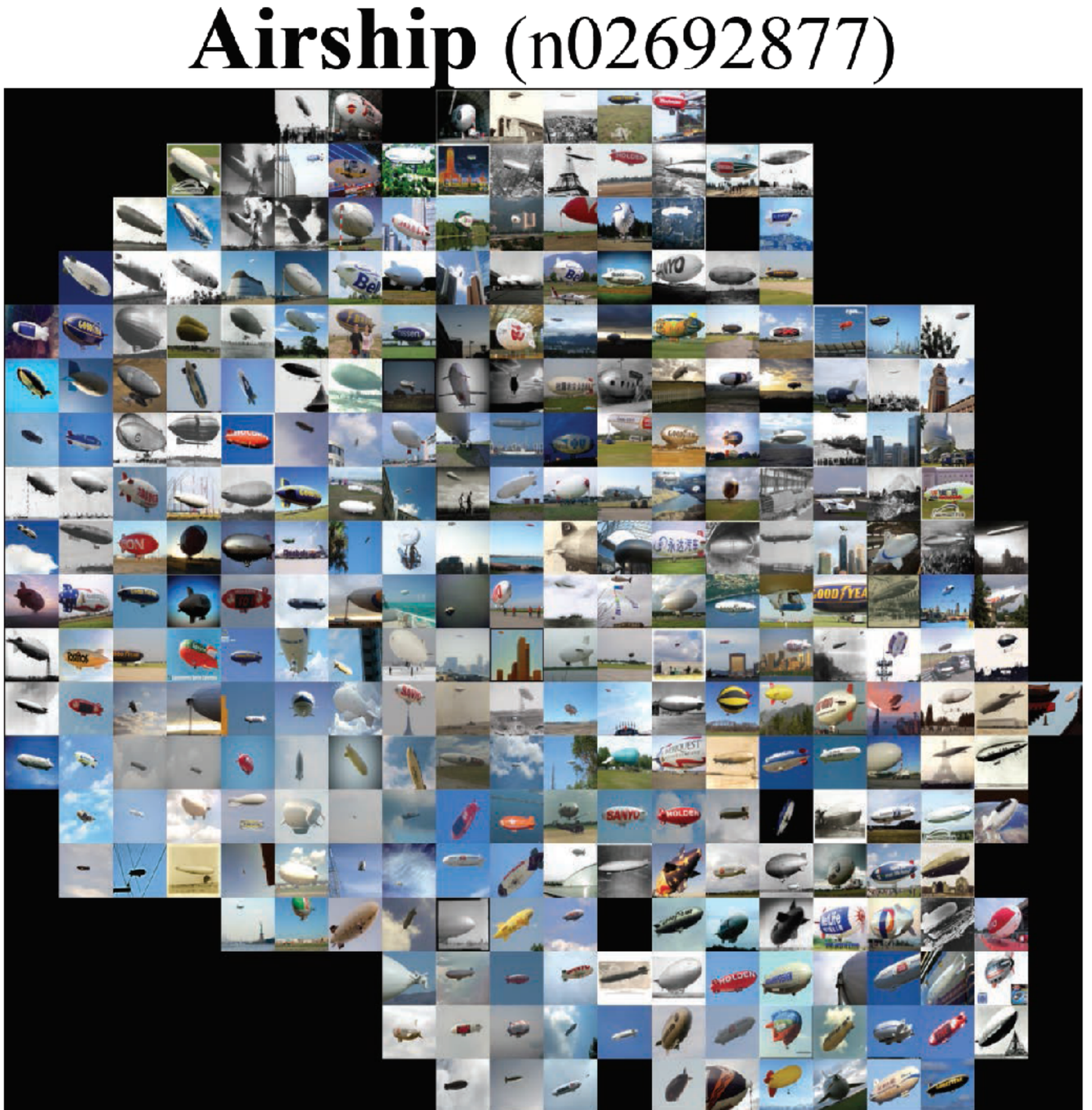
Pose (surface normal) Embedding



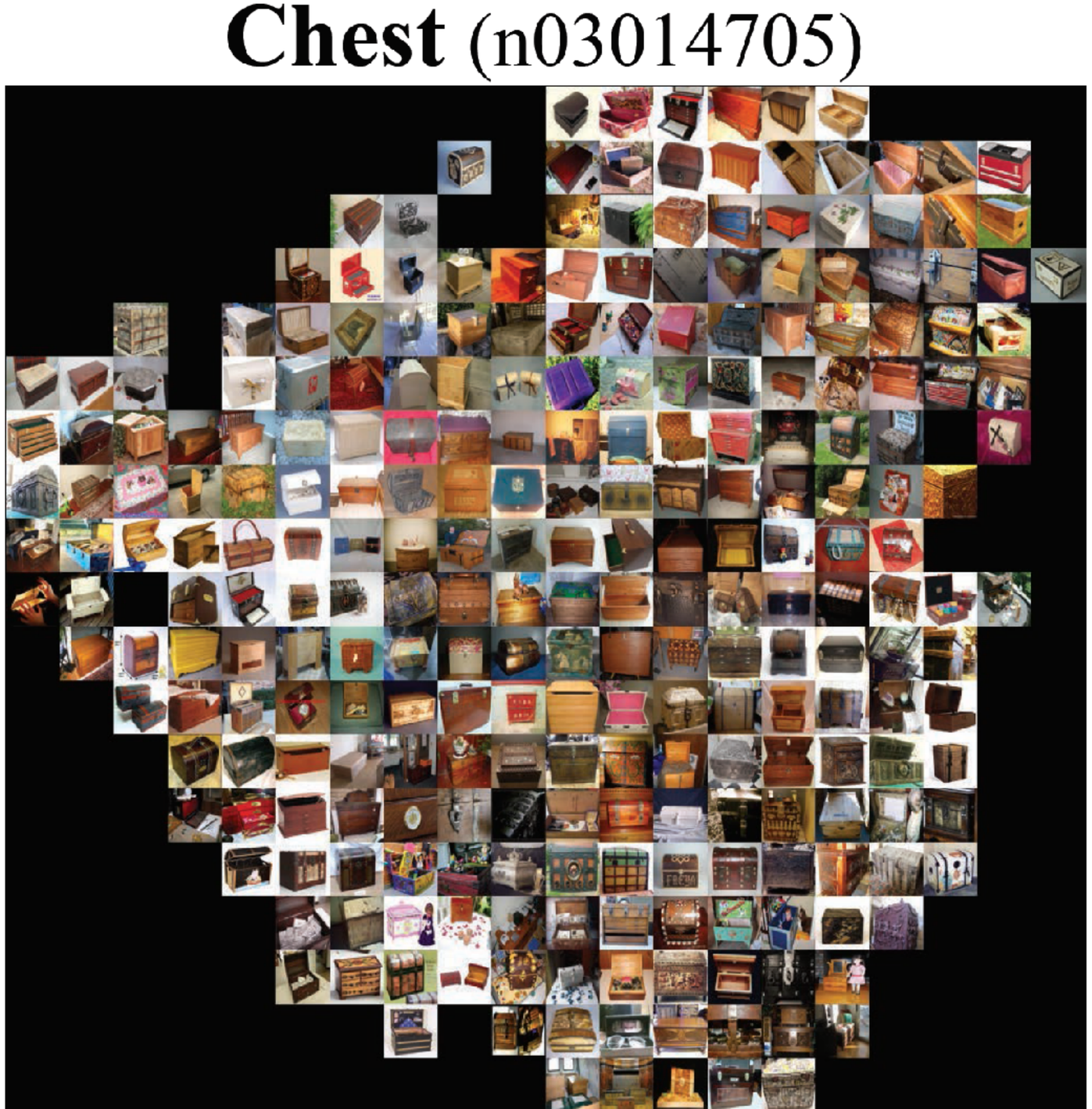


MIT Places

3D Object Pose - ImageNet



3D Object Pose - ImageNet



Demo

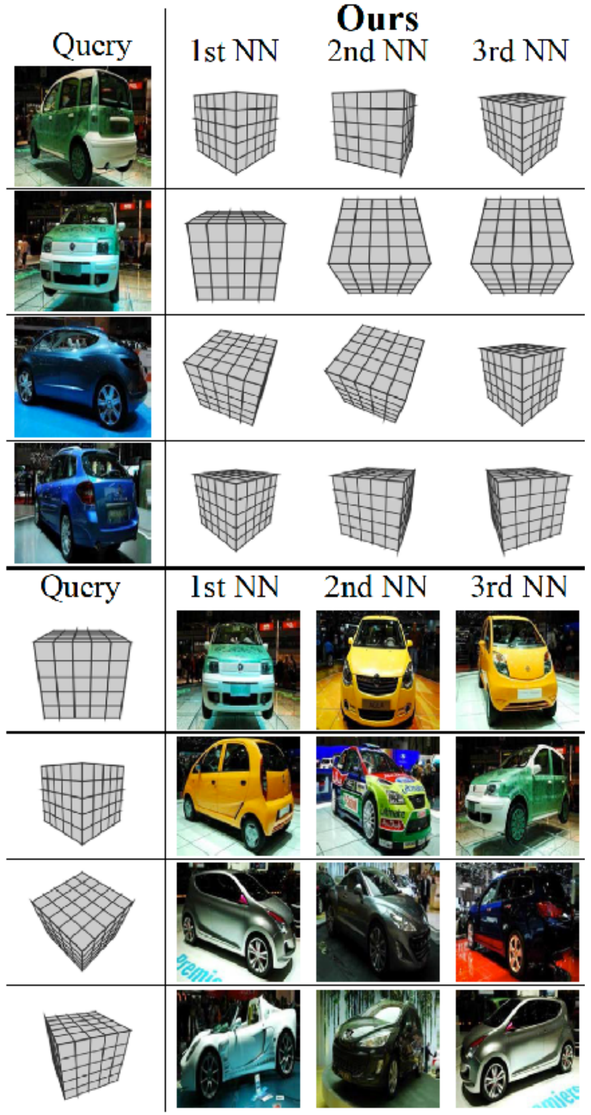
http://3drepresentation.stanford.edu/

Query Image

Generic 3D Representation

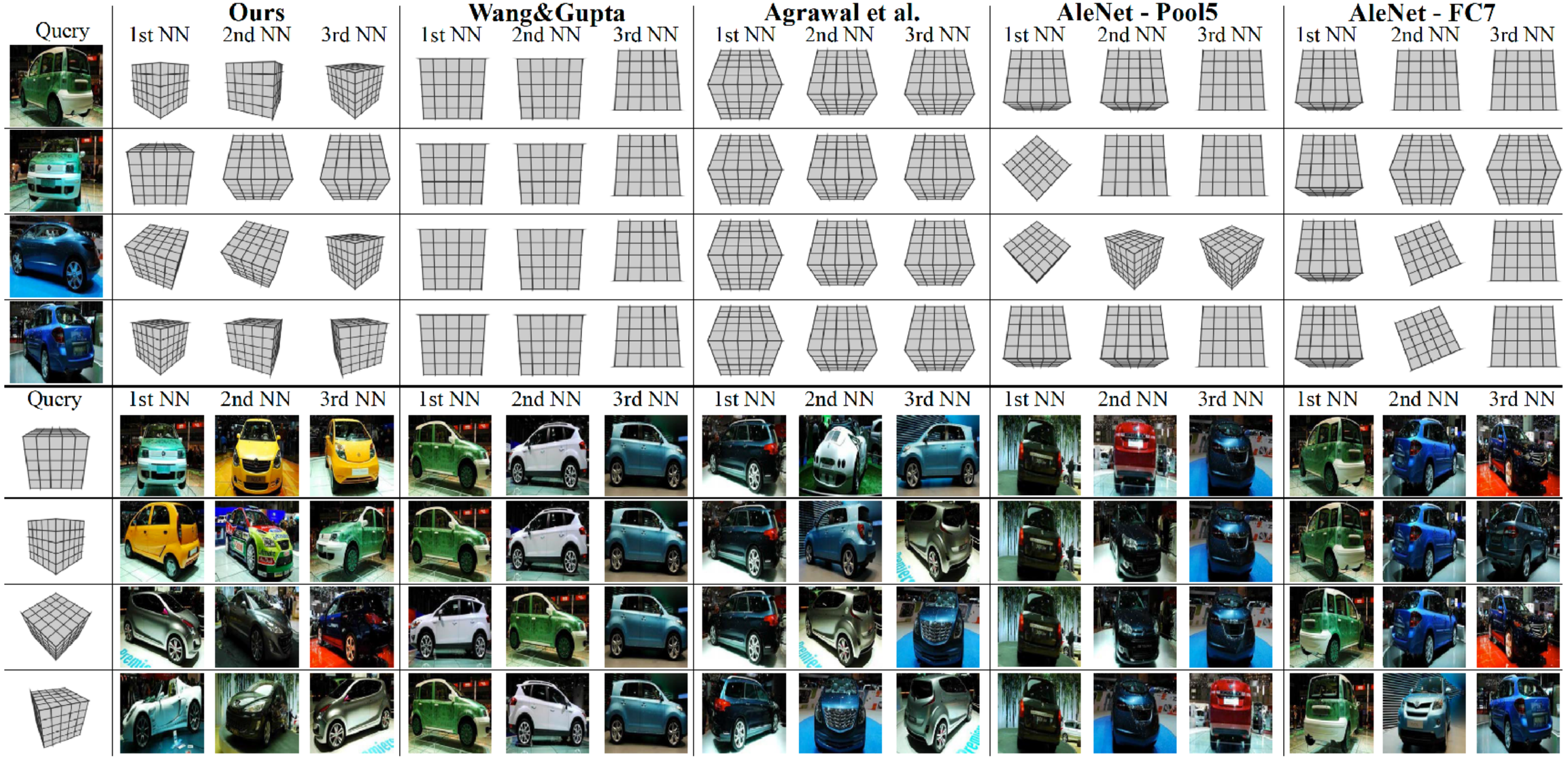
ImageNet (AlexNet)

3D Object Pose Estimation - Abstraction



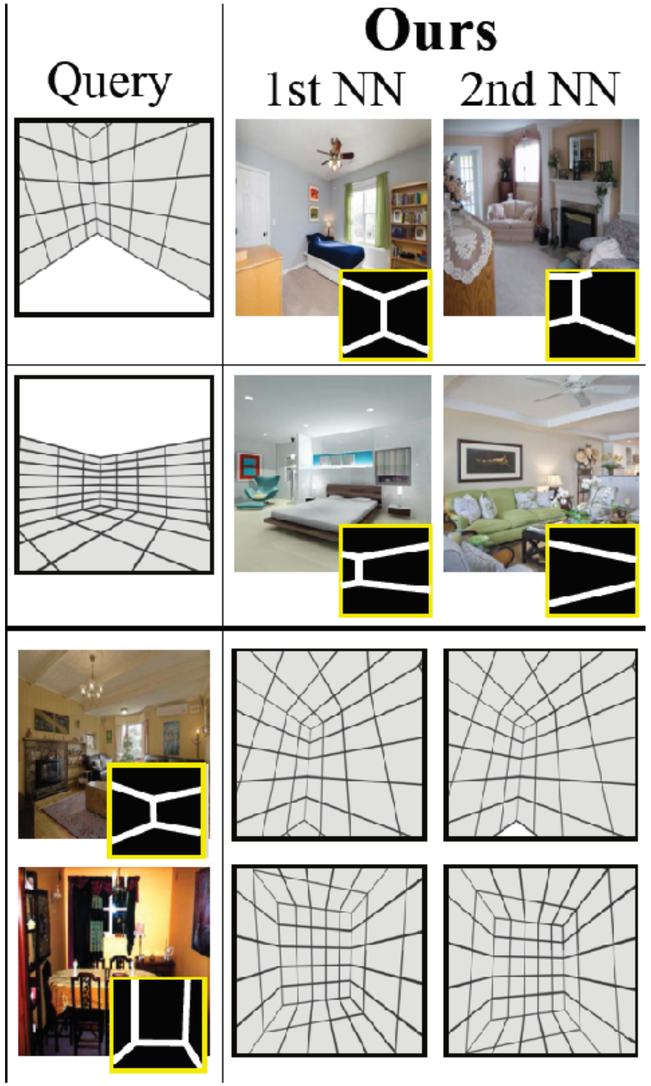
Wang & Gupta. 2015
Agrawal et al. 2015.
Russakovsky et al. 2015
Ozuysal et al. 2009.

3D Object Pose Estimation - Abstraction



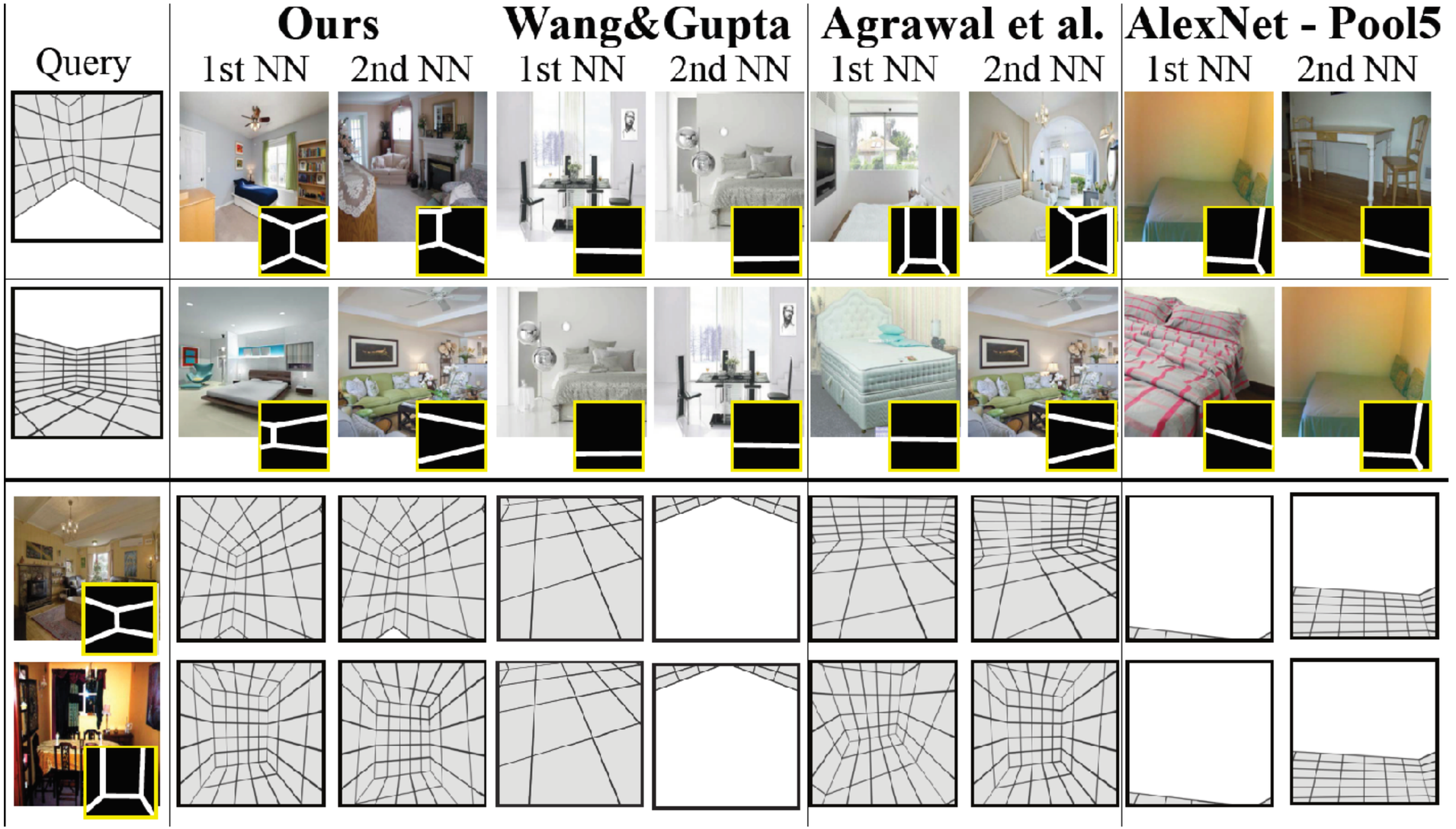
Wang & Gupta. 2015
Agrawal et al. 2015.
Russakovsky et al. 2015
Ozuysal et al. 2009.

3D Layout Estimation - Abstraction



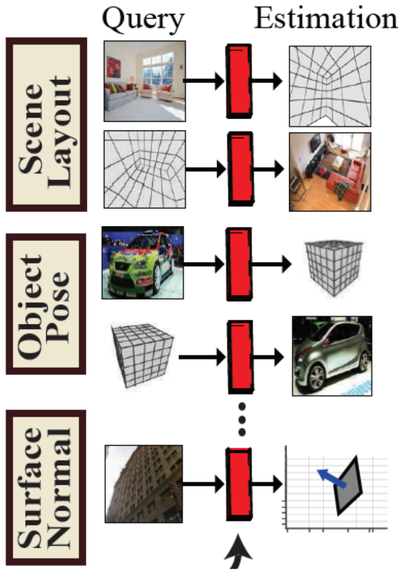
Wang & Gupta. 2015
Agrawal et al. 2015.
Russakovsky et al. 2015
Zhang et al. 2016.

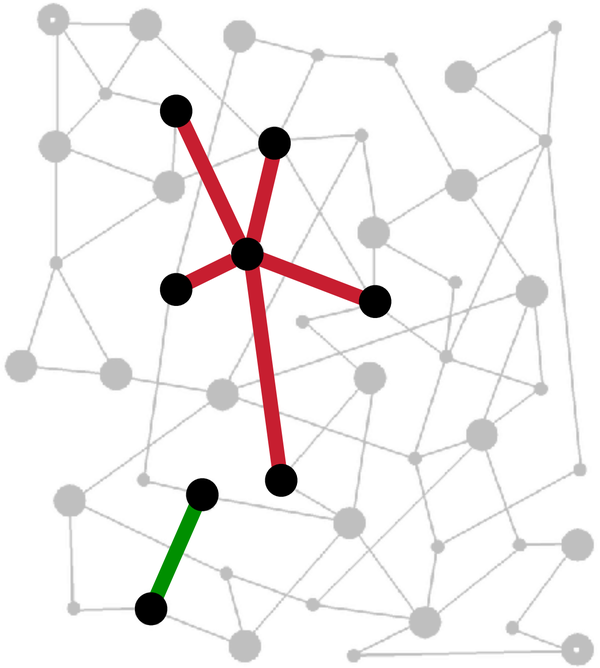
3D Layout Estimation - Abstraction



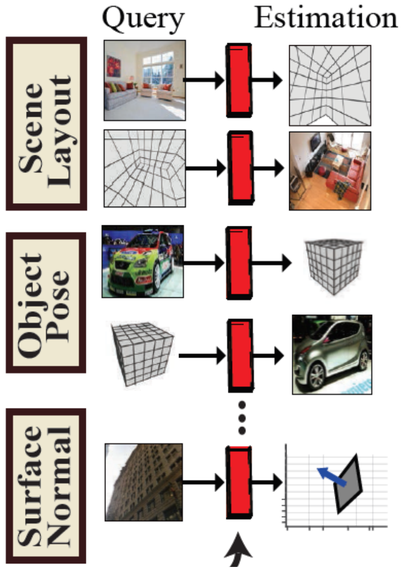
Wang & Gupta. 2015
Agrawal et al. 2015.
Russakovsky et al. 2015
Zhang et al. 2016.

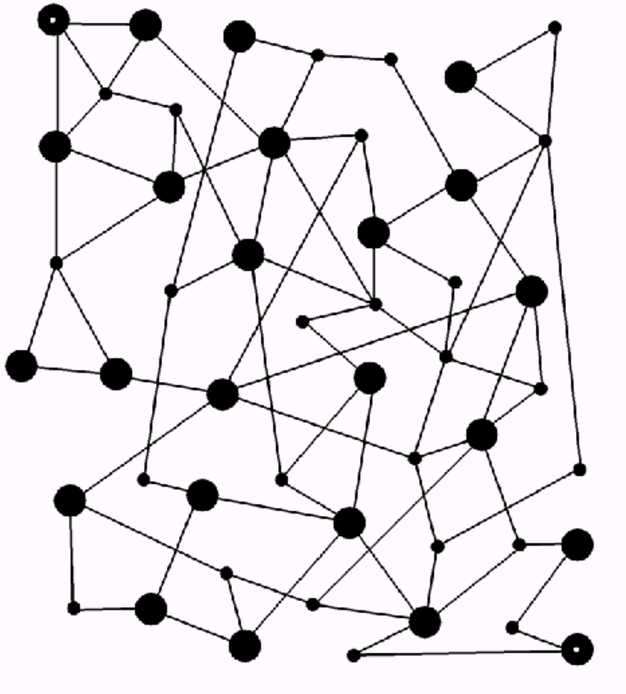
Task Interplay

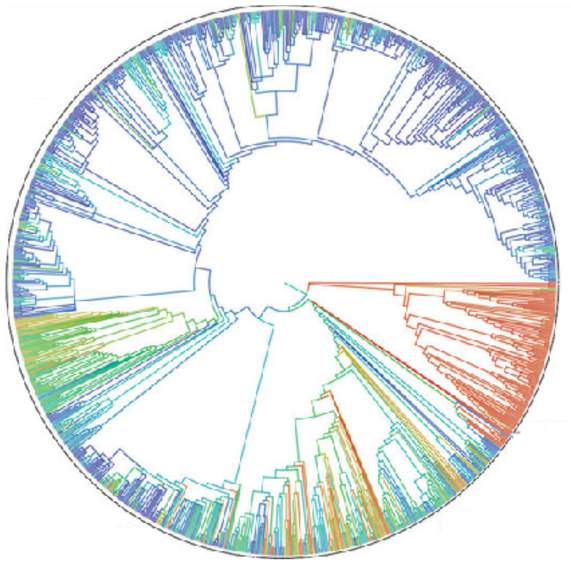




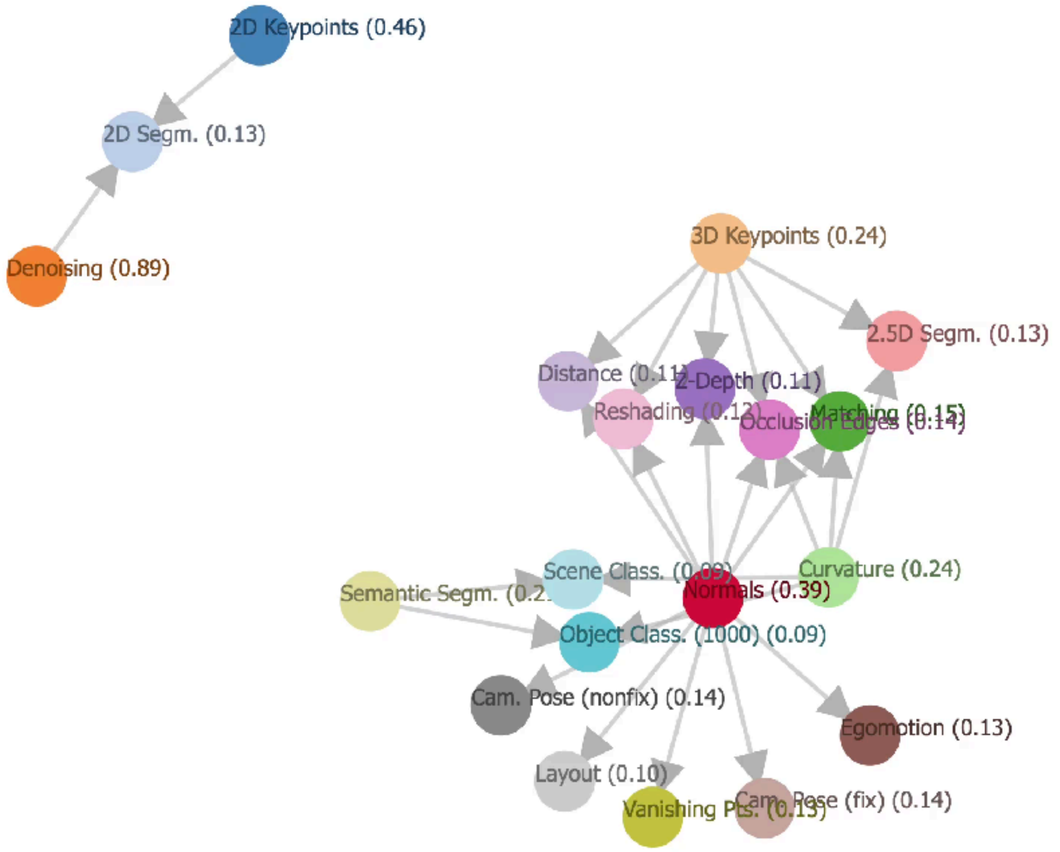
Task Interplay







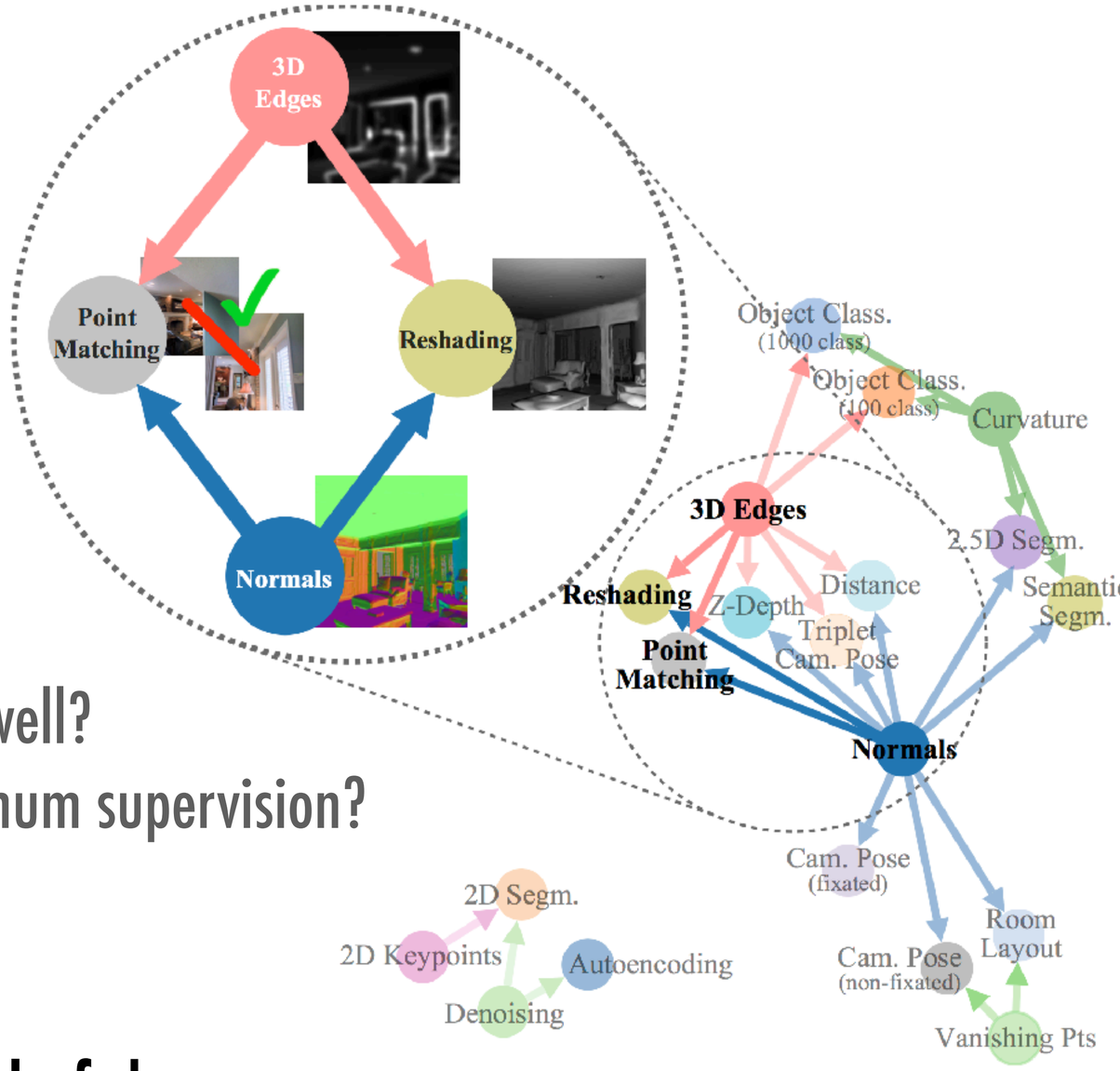
Taskonomy



Taskonomy:
Disentangling Task Transfer Learning

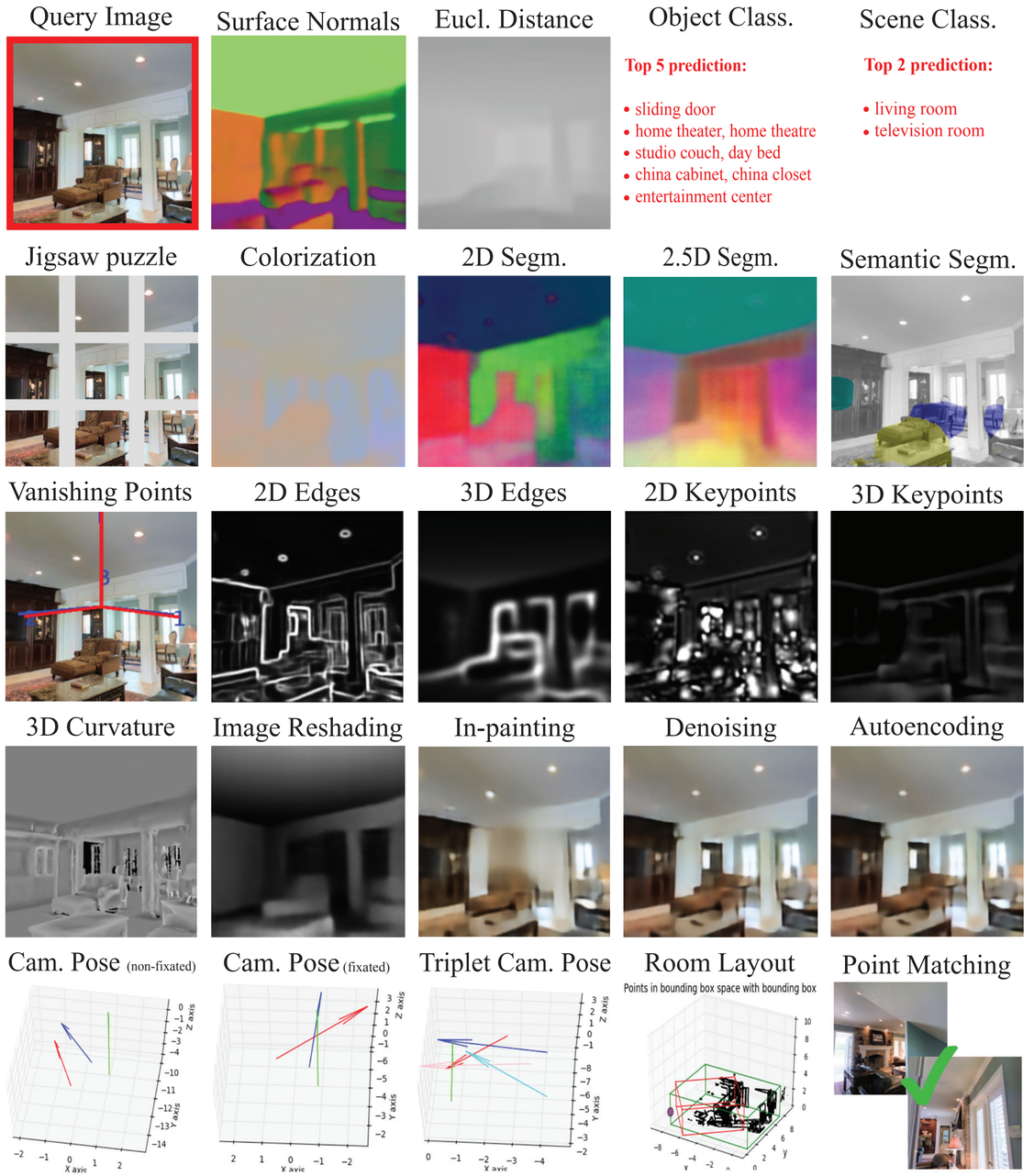
Zamir, Sax*, Shen*, Guibas, Malik, Savarese

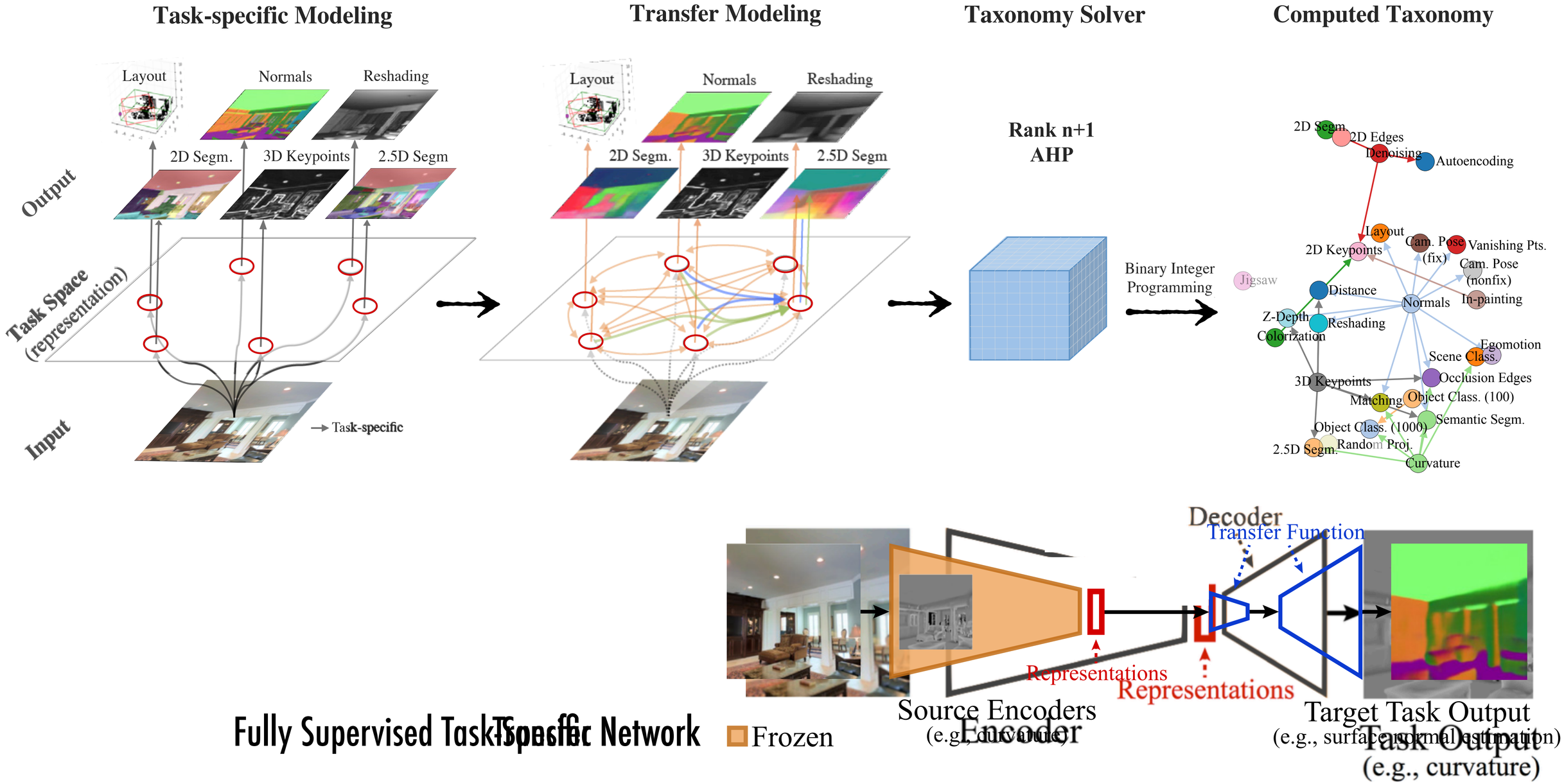
- How visual tasks are related to each other?
- What task(s) transfer to arbitrary task(s)? And how well?
- How to find a curriculum for solving tasks with minimum supervision?
- How to flow the supervision across tasks?



Taskonomy: A fully computational model of the structure in the space of visual tasks.

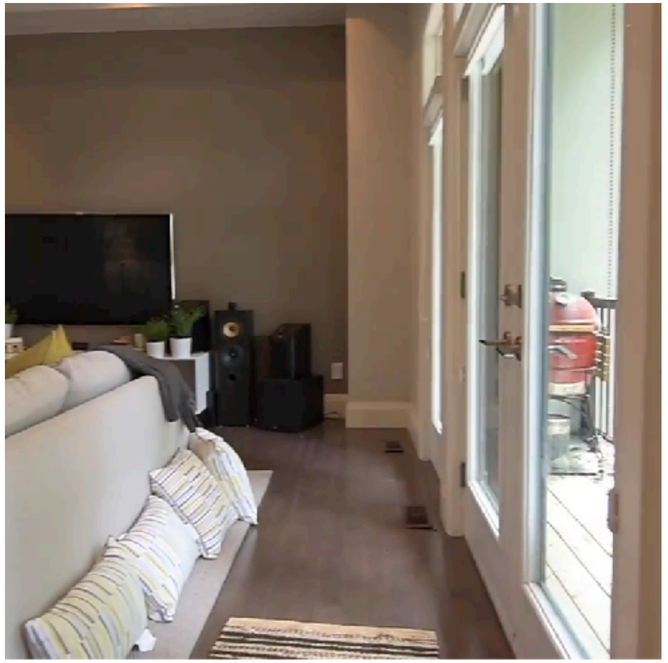
- Dictionary of Tasks (a subset shown)
- A dataset of 3 million images where each image has the GT label for all tasks





Fully Computational pipeline to modeling the task space

Query:

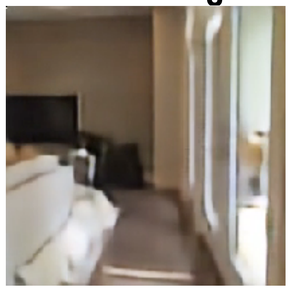


Prediction Results of our Task-Specific Networks on a YouTube Video

(Frame-by-frame. 16 tasks out of 26)

Normals

Denoising



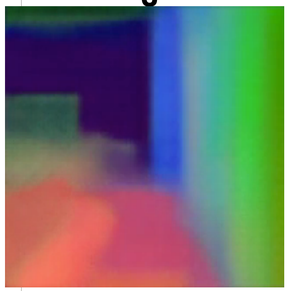
Object Class.

Top 5 prediction:
sliding door
studio couch, day bed
shoji
quilt, comforter, comfort, puff
four-poster

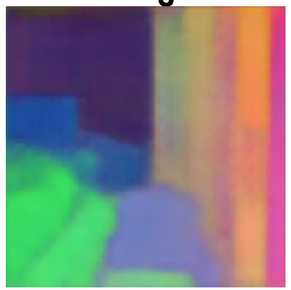
Scene Class.

Top 5 prediction:
dorm room
bedroom
hotel room
youth hostel
television room

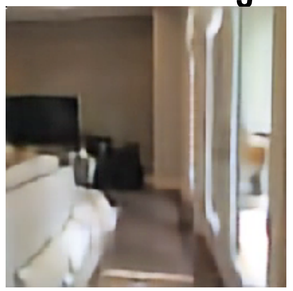
2D Segment



2.5D Segment



Autoencoding



Semantic Seg.

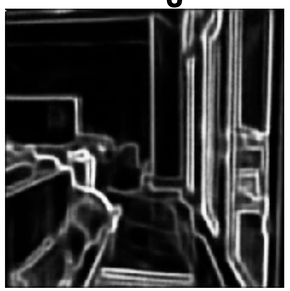
Curvature

Reshading

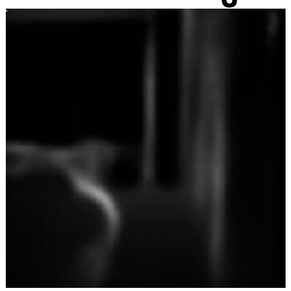
Depth

Eucl. Distance

2D Edges

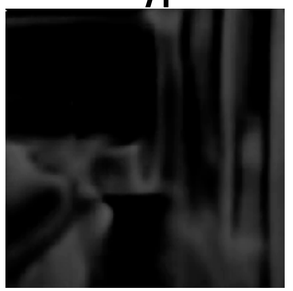


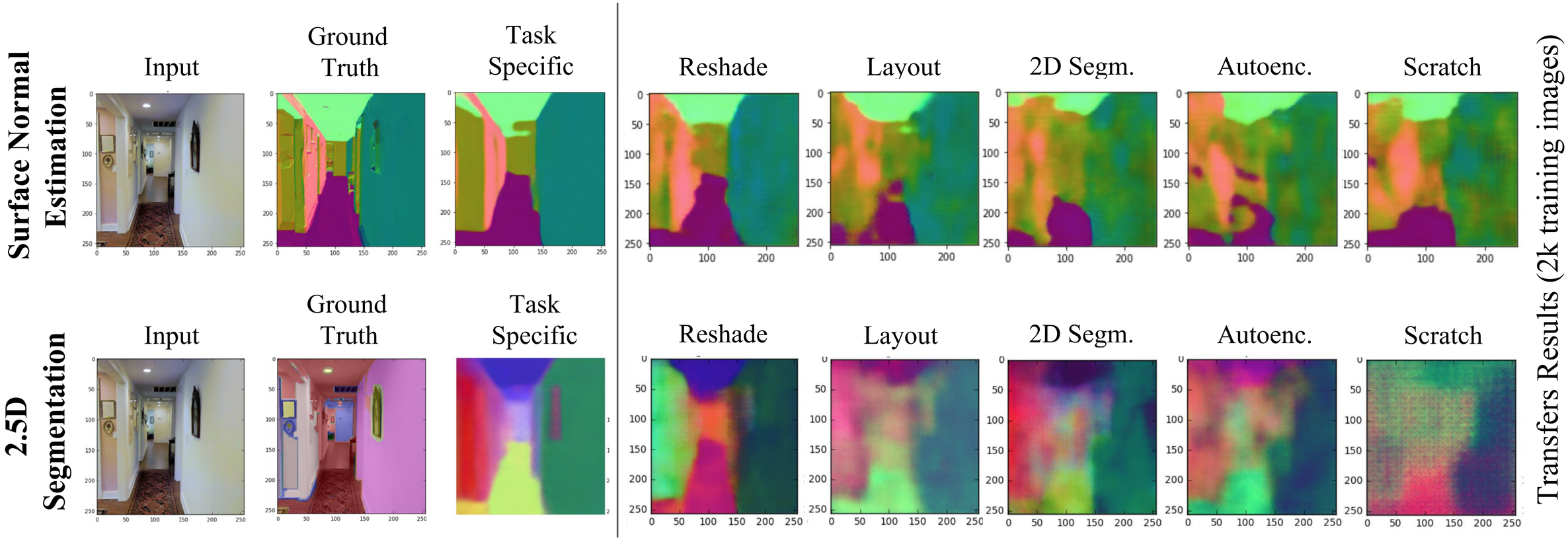
Occlusion Edge.



2D Keypoint

3D Keypoint





Qualitative Transfer Learning Results

Note the large difference in the quality of transfer results based on the source task

File Edit View Insert Cell Kernel Widgets Help Trusted Python (analysis)

Table of contents

To interact with the API, please run the initialization and then sections 1, 2, and 3. After that, you can toggle the tabs in order to find new taxonomies

Initialization: Run this once:

In [1]:
```
%load_ext autoreload
%autoreload 2
import json
import taxonomy_api

# Load values so here so that JavaScript can access them
dists_and_wins, all_tasks = taxonomy_api.load_affinities()
task_to_set, minimal_tasks_to_set = taxonomy_api.get_default_sets(all_tasks, dists_and_wins)
map_to_display_names = taxonomy_api.map_to_display_names
```

API

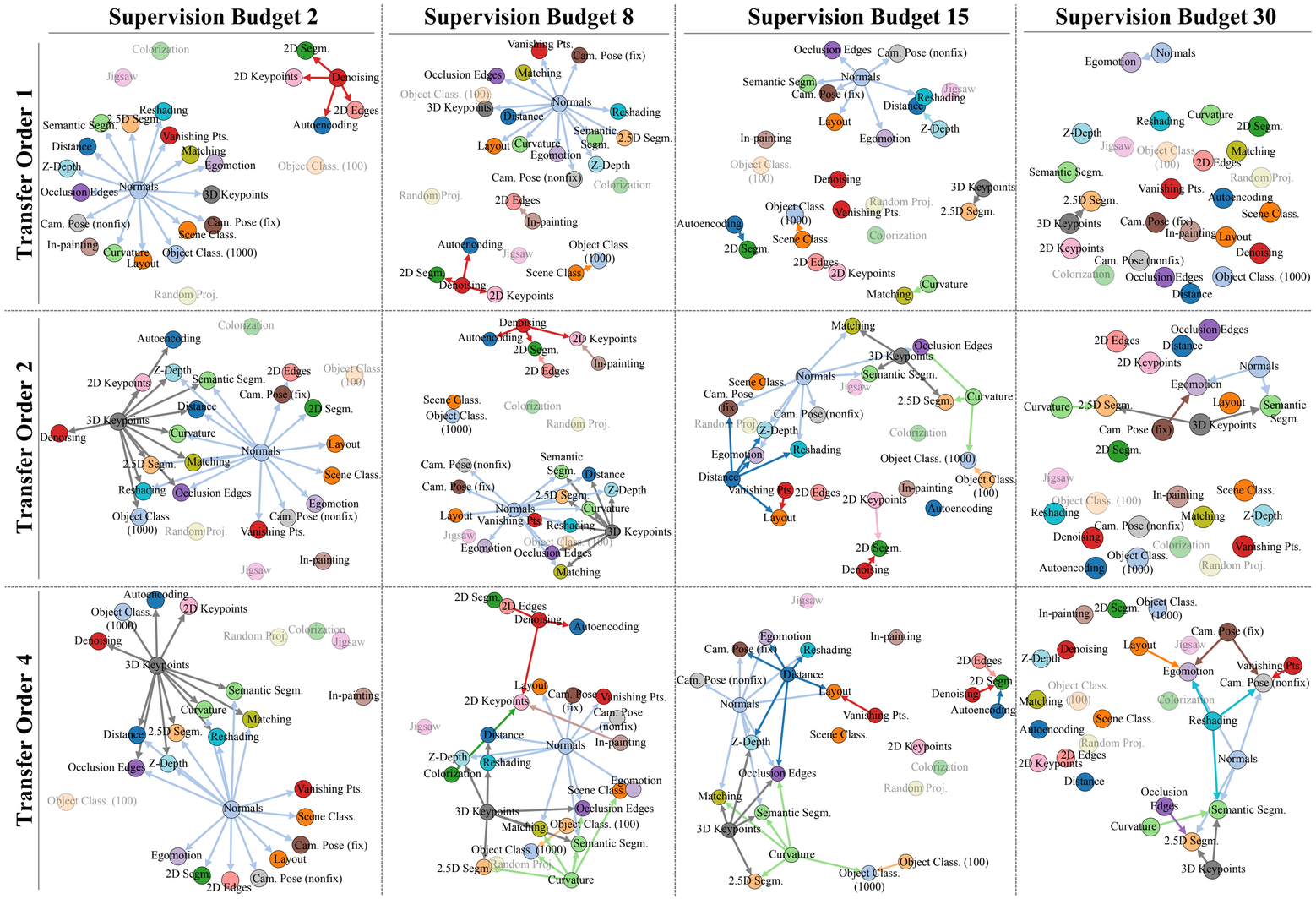
1. Choose our dictionary

In [2]:
```
# Click 'Lock in!' after changing the dictionary
taxonomy_api.HTML(filename='viz/task_selector.html')
```

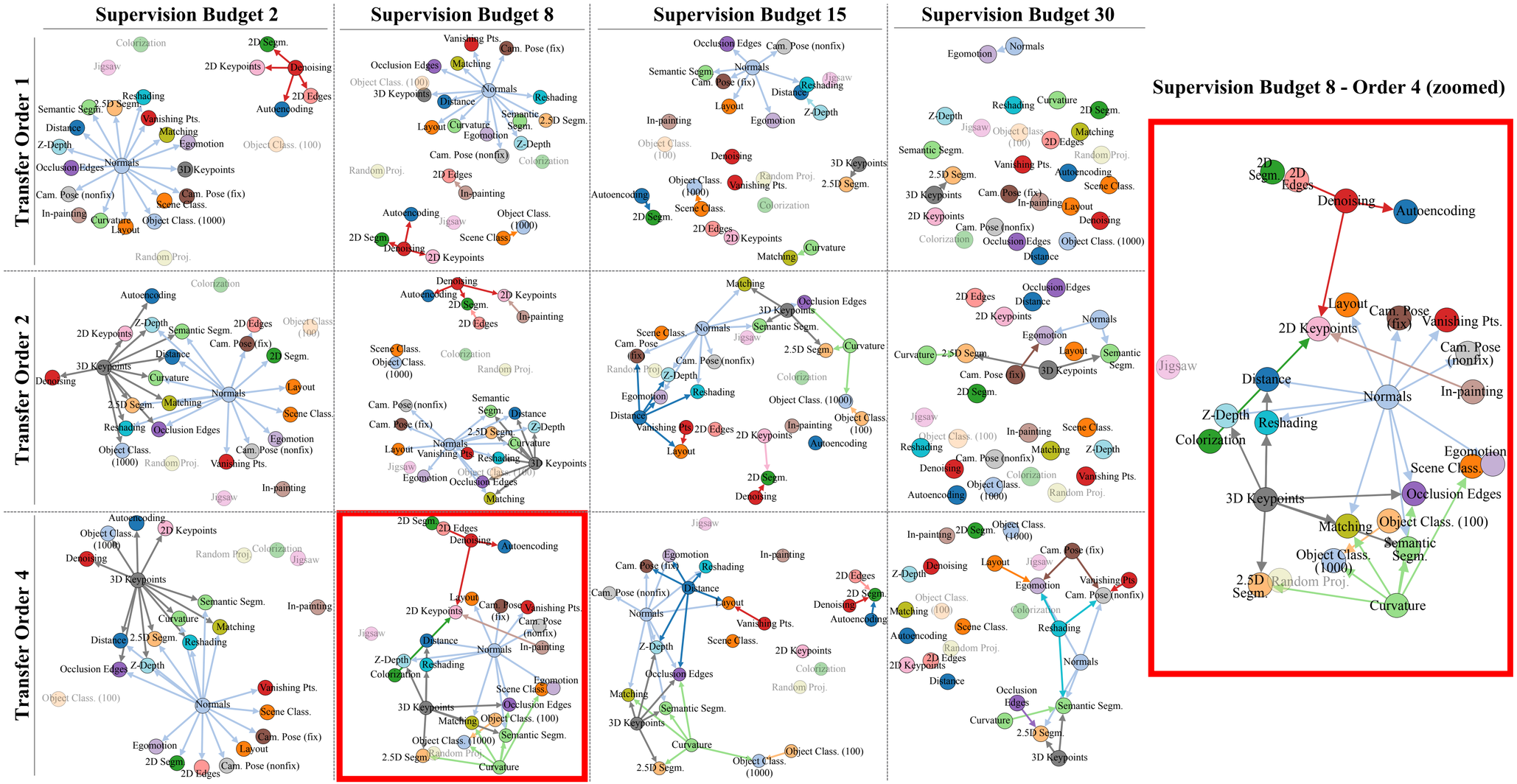
Out[2]:

Transfer Learning API. Based on Taskonomy BIP solver.

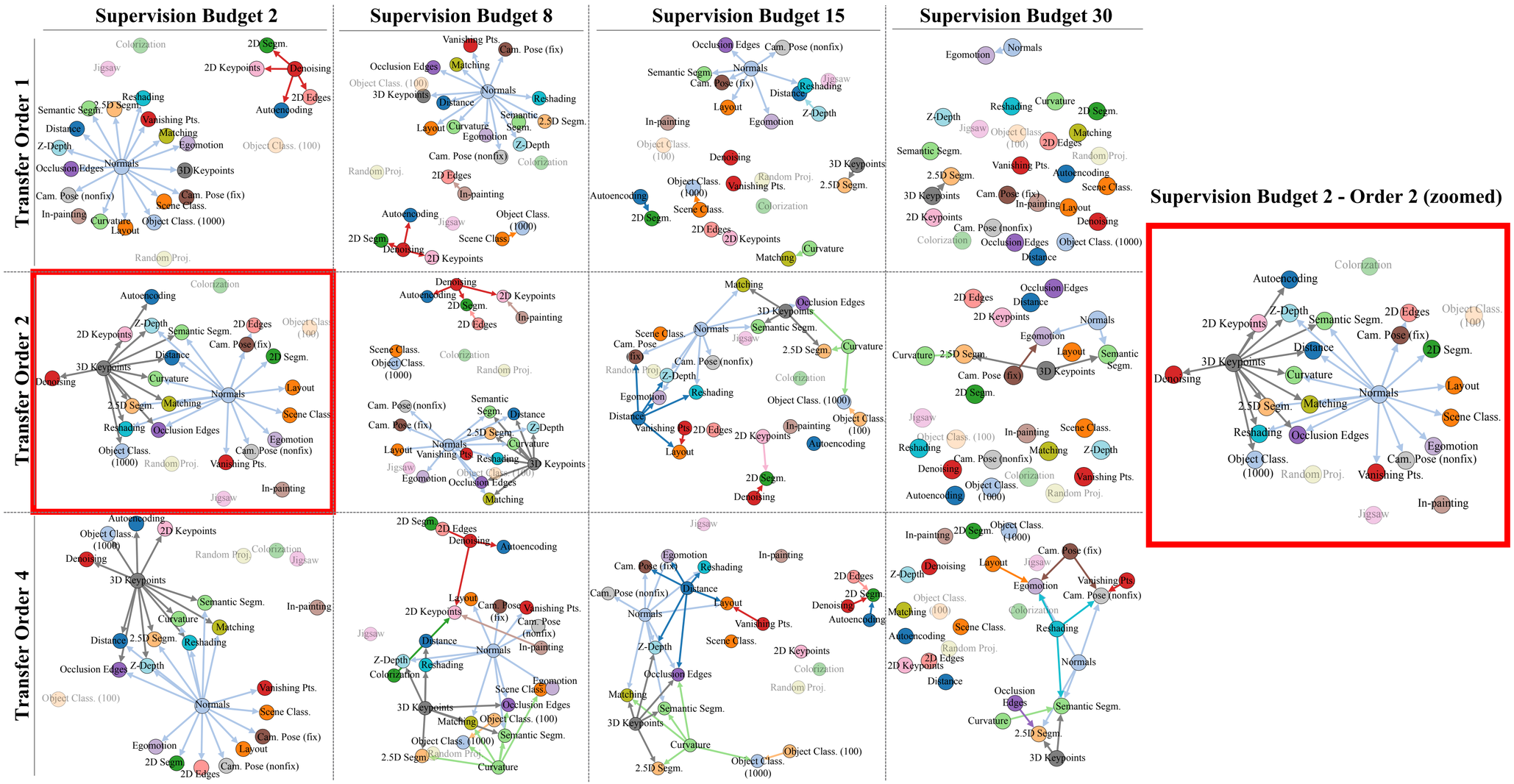
Can be employed by users to devise efficient supervision policies for desired tasks with arbitrary arguments



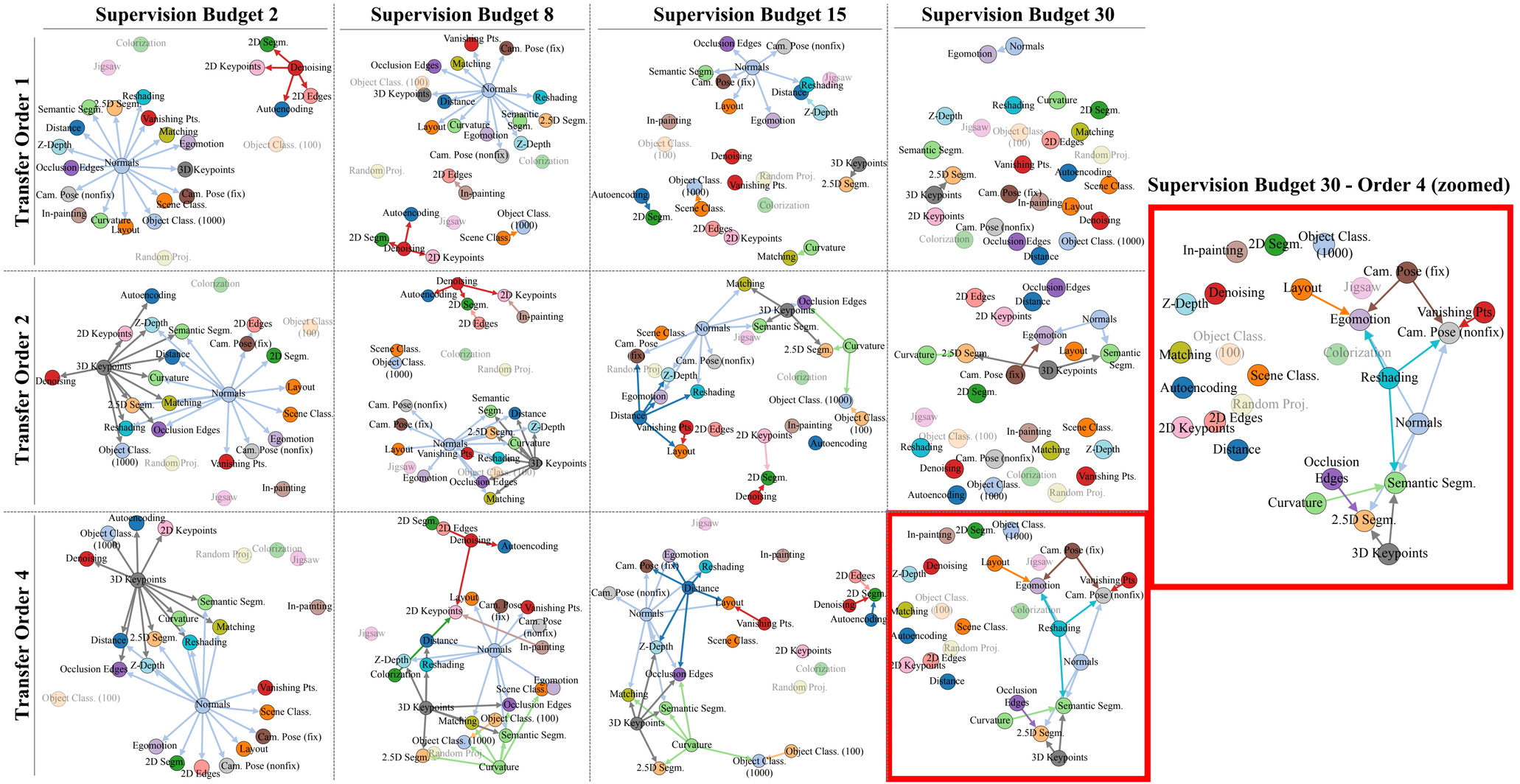
Computed Taxonomies using BIP for various order and supervision budget arguments



Computed Taxonomies using BIP for various order and supervision budget arguments



Computed Taxonomies using BIP for various order and supervision budget arguments



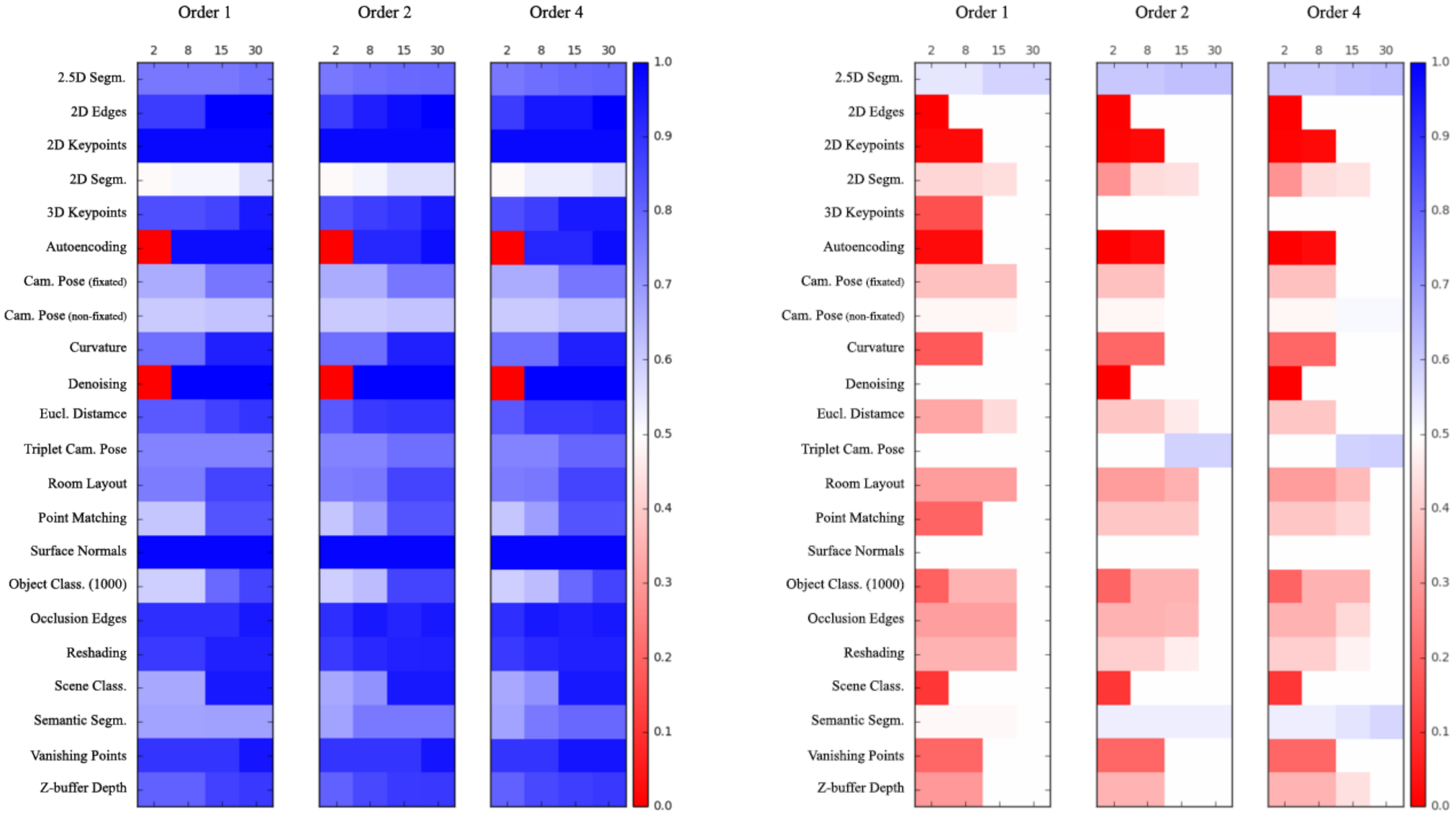
Computed Taxonomies using BIP for various order and supervision budget arguments

Metrics

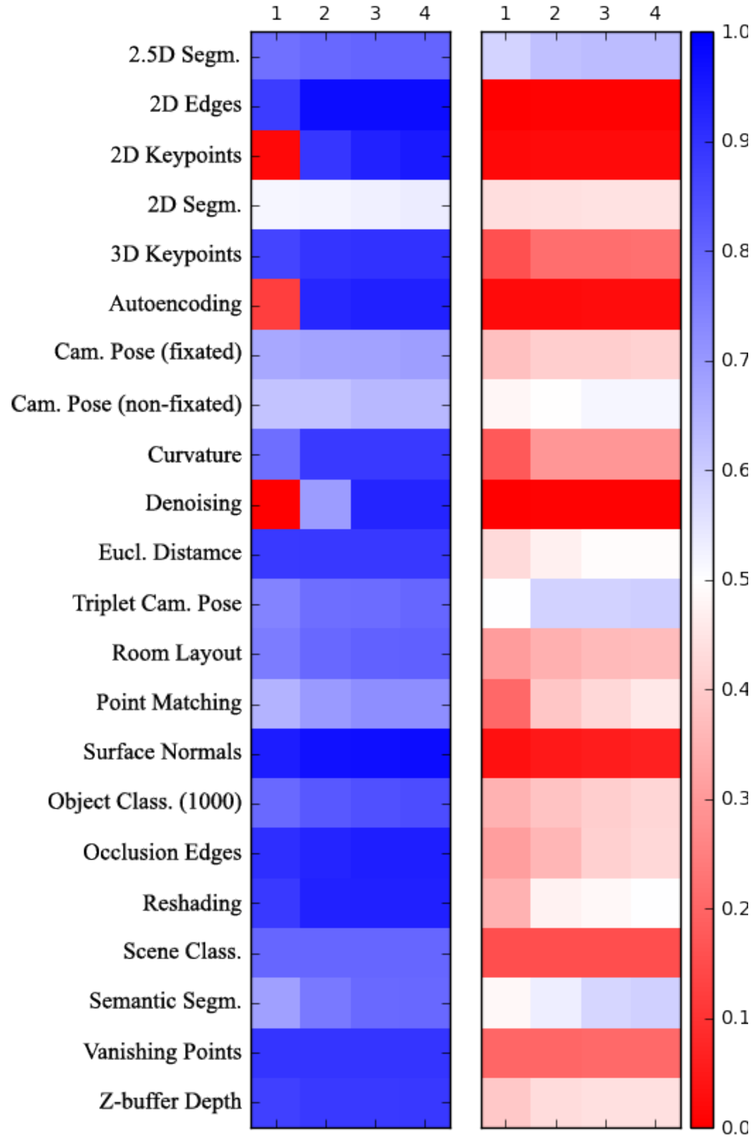
Gain: Expected win rate of a transfer network over a network trained on the same number of images. That is, the best that could be done if transfer learning was not utilized.

Quality: Expected win rate of a transfer network over a fully supervised network with full 120k images (gold standard).

Significance: How strong is the structure in the space, i.e. how better is the found taxonomical structure compared to average connectivity/taxonomy.



Gain (left) and Quality (right) metrics on Taxonomy results for full set of tasks as target



Gain (left) and Quality (right) metrics on Taxonomy results for novel tasks

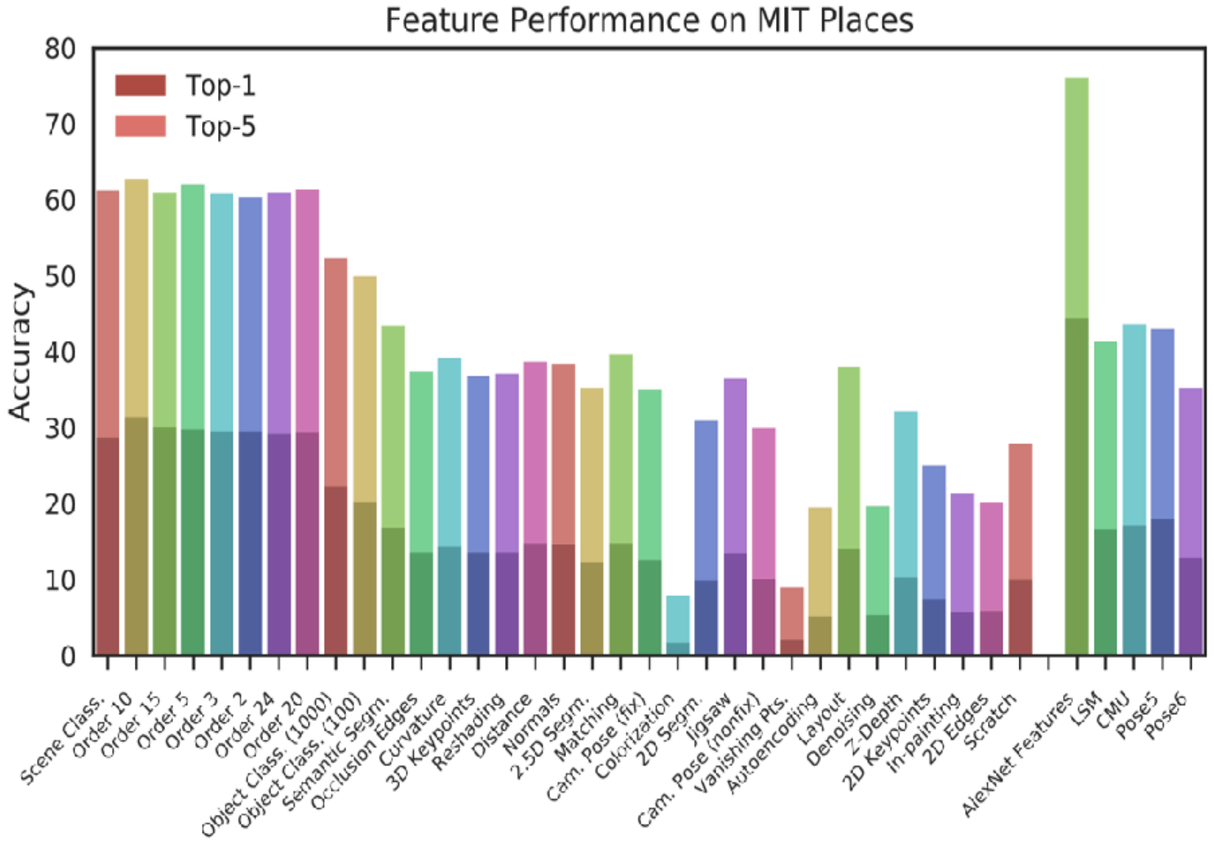
Computed Transfer Policies for individual novel tasks

- On our dataset
- Metric: win-rate (%) of taxonomy. >50, <50.
- Note the strong margin of win by Taxonomy Transfer Policies
- Suggesting a both strong and fluid structure

Task	scratch	ImageNet fc7[39]	Wang.[72]	Agrawal.[1]	Zamir.[76]	Zhang.[77]	Noroozi.[51]	full sup.	Ours
Depth	88	88	93	89	88	84	86	43	-
	.03	.04	.04	.03	.04	.03	.03	**.02**	.02
Scene Cls.	80	52	83	74	74	71	75	15	-
	3.30	2.76	3.56	3.15	3.17	3.09	3.19	**2.23**	2.63
Sem. Segm.	78	79	82	85	76	78	84	21	-
	1.74	1.88	1.92	1.80	1.85	1.74	1.71	**1.42**	1.53
Object Cls.	79	54	82	76	75	76	76	34	-
	4.08	3.57	4.27	3.99	3.98	4.00	3.97	**3.26**	3.46
Normals	97	98	98	98	98	97	97	6	-
	.22	.30	.34	.28	.28	.23	.24	**.12**	.15
2.5D Segm.	80	93	92	89	90	84	87	40	-
	.21	.34	.34	.26	.29	.22	.24	**.16**	.17
Occ. Edges	93	96	95	93	94	93	94	42	-
	.16	.19	.18	.17	.18	.16	.17	**.12**	.13
Curvature	88	94	89	85	88	92	88	29	-
	.25	.28	.26	.25	.26	.26	.25	**.21**	.22
Egomotion	79	78	83	77	76	74	71	59	-
	8.60	8.58	9.26	8.41	8.34	8.15	7.94	7.32	**6.85**
Layout	80	76	85	79	77	78	70	36	-
	.66	.66	.85	.65	.65	.62	.54	**.37**	.41

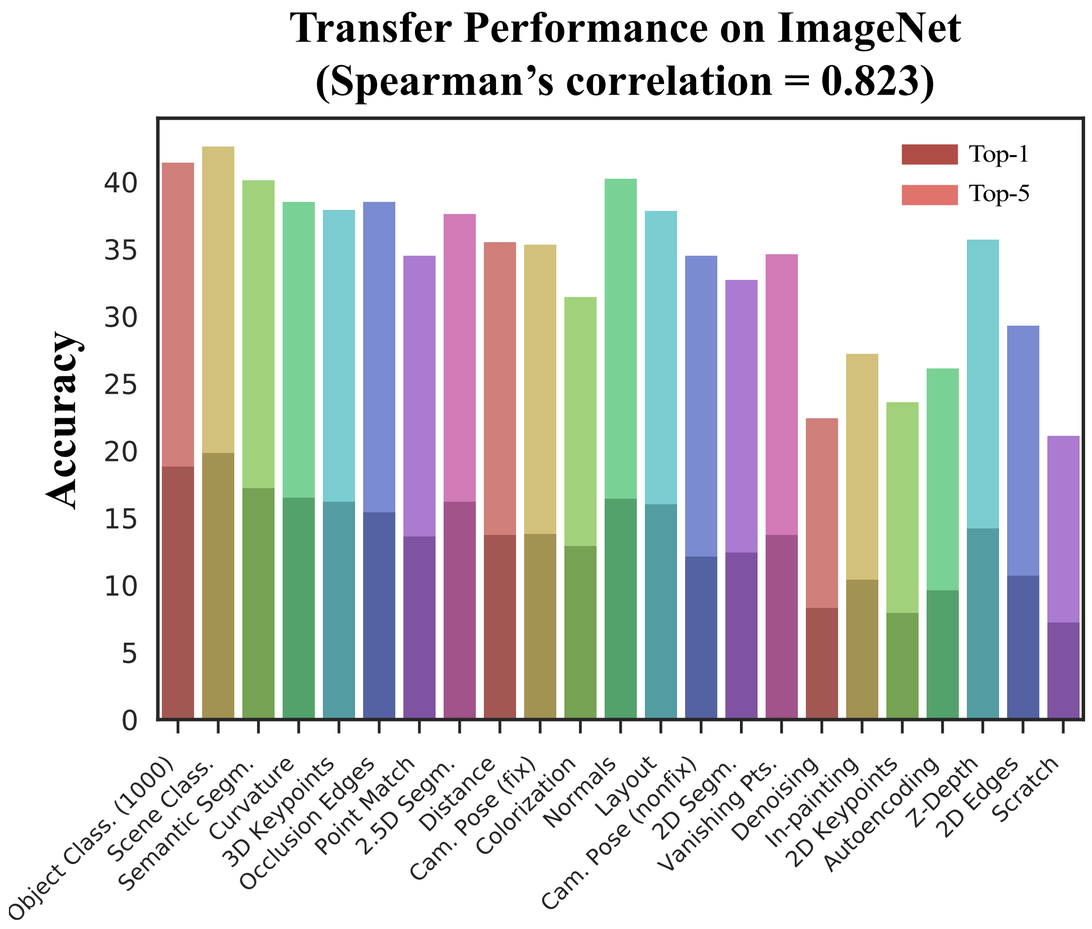
Comparison of the results of Taxonomy Transfer Policies with various Self-Supervised Baselines

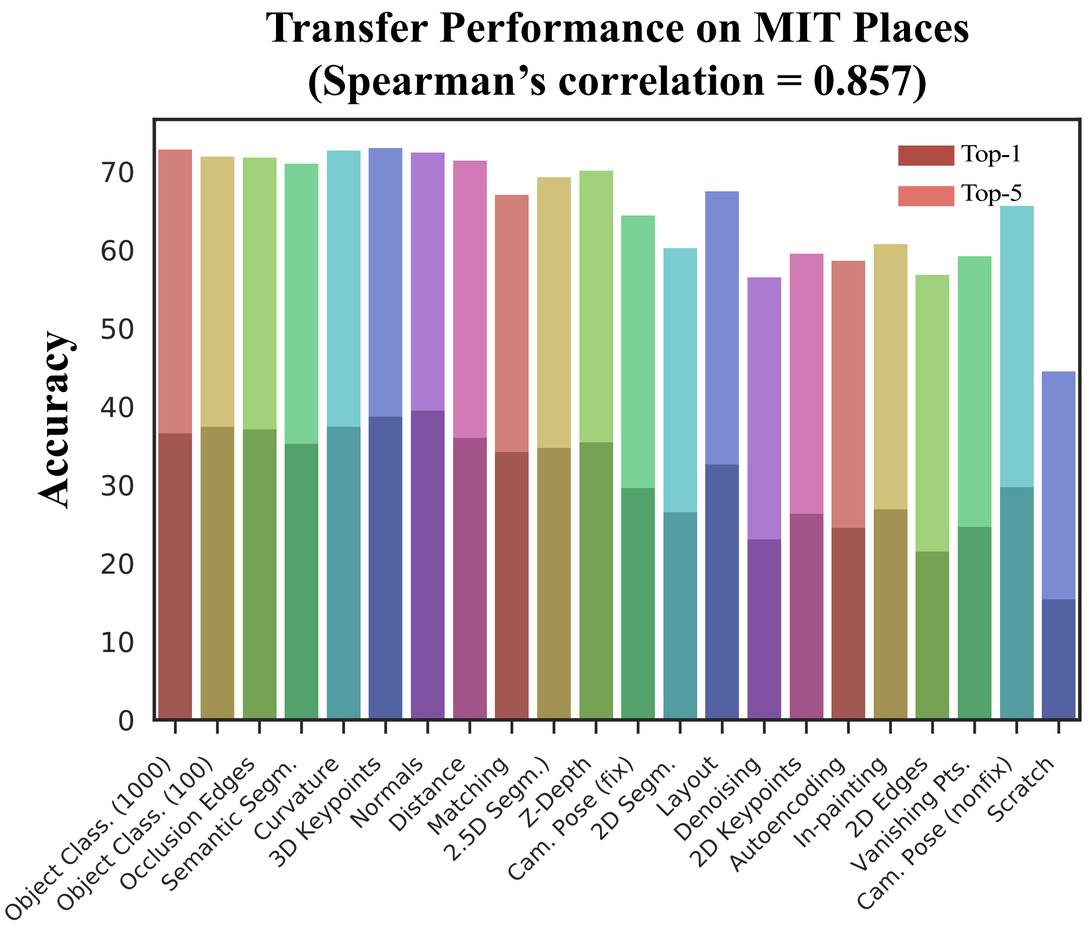
- On MIT Places semantic scene categorization dataset
- Metric: MIT Places standard metric
- Note the strong margin of win by Taxonomy Transfer Policies over self-supervised methods.
- Taxonomy is beaten by only ImageNet (AlexNet) features, which are trained on 8x more data, are directly based on semantics, and are manually annotated.



Comparison of the results of Taxonomy Transfer Policies with various Self-Supervised Baselines

Are the findings dataset dependent?

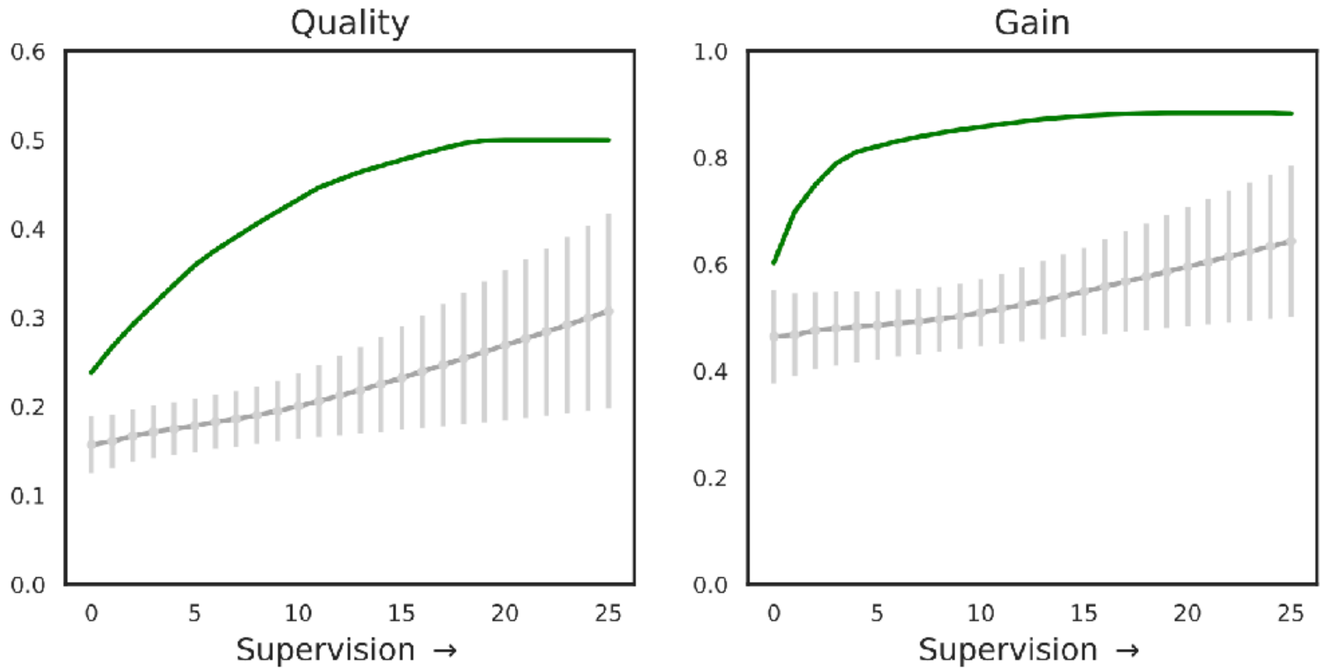




- Verification of the found structure on external datasets
- Showing a strong correlation suggesting dataset independence

Significance Test

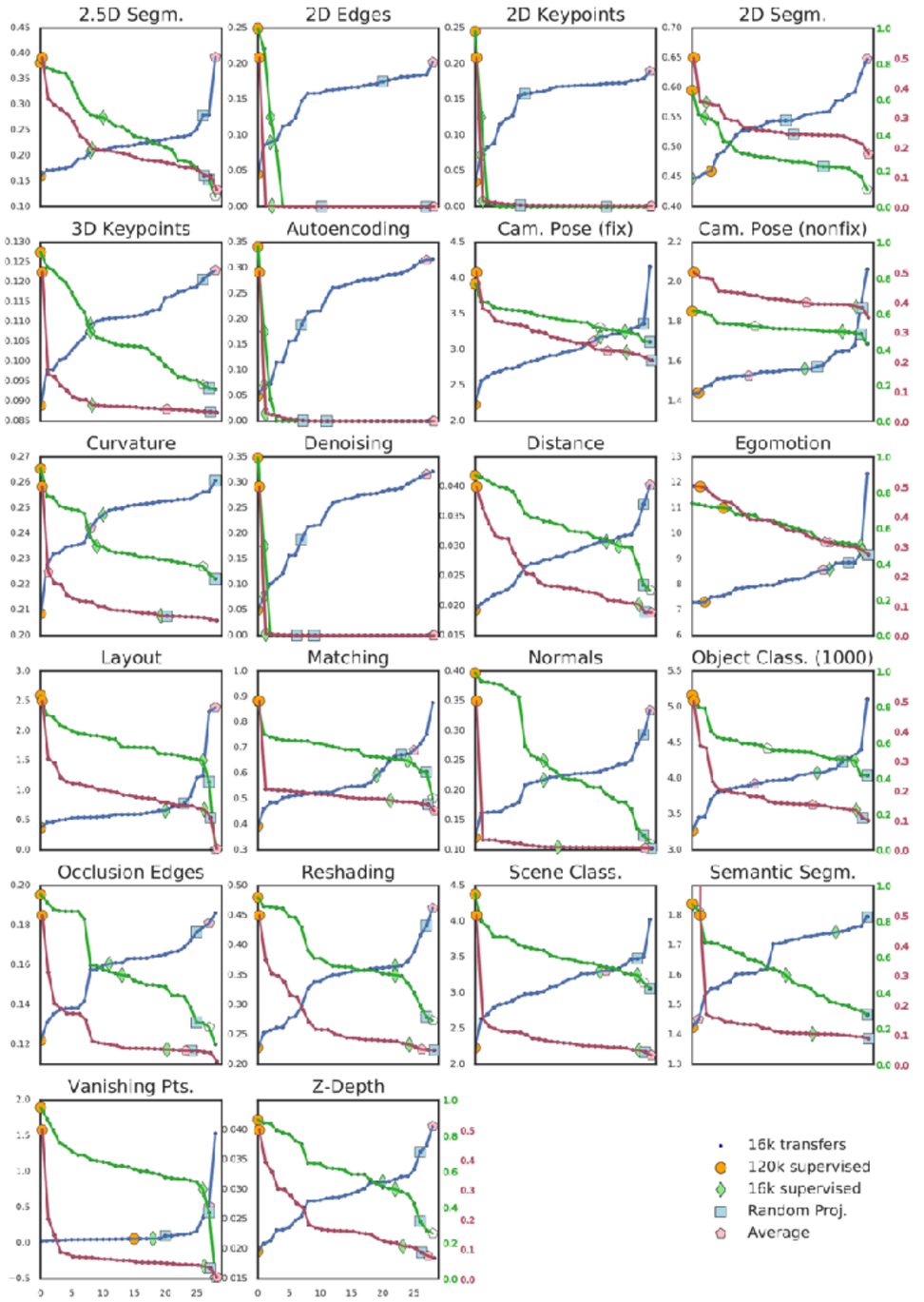
Taxonomy vs Random Connectivity



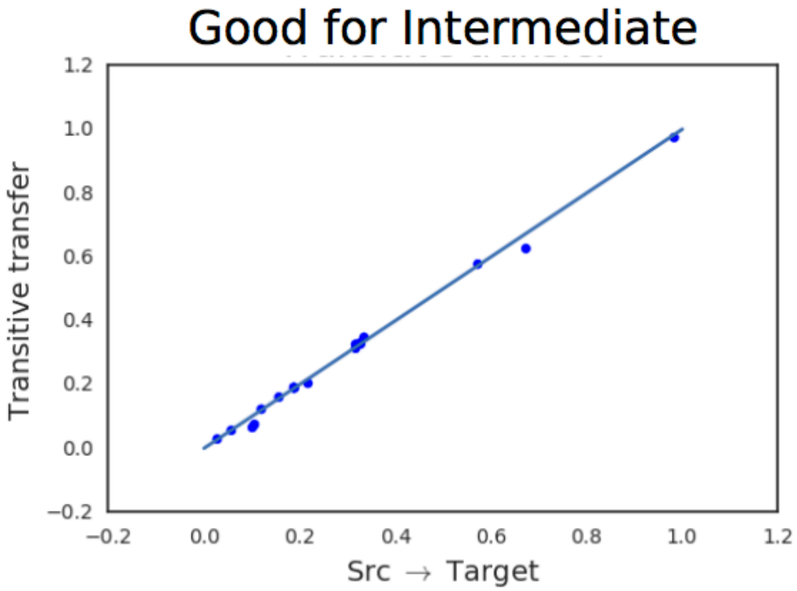
Significance: How strong is the structure in the space, i.e. how better is the found taxonomical structure compared to average connectivity/taxonomy.
Evaluation suggests a significantly structured space and good modeling of it by Taskonomy.

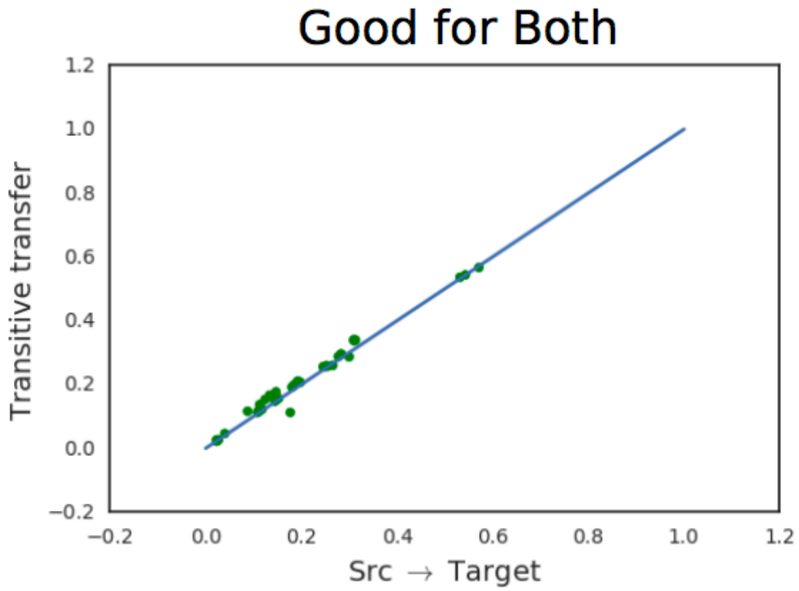
Significance test of structure per task:

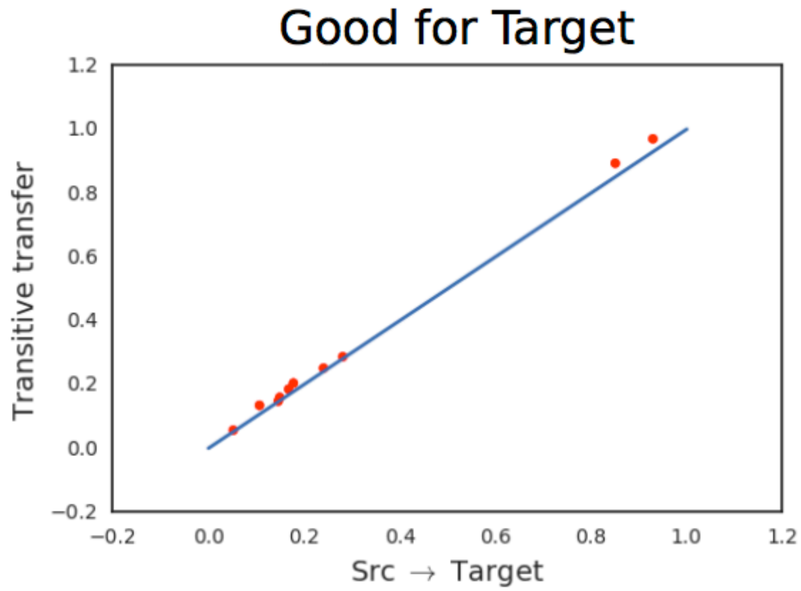
- Sorted Quality, Gain, and Loss of 1st order transfers for each task.
- Evaluation suggests a significantly structured space and good modeling of it by Taskonomy.



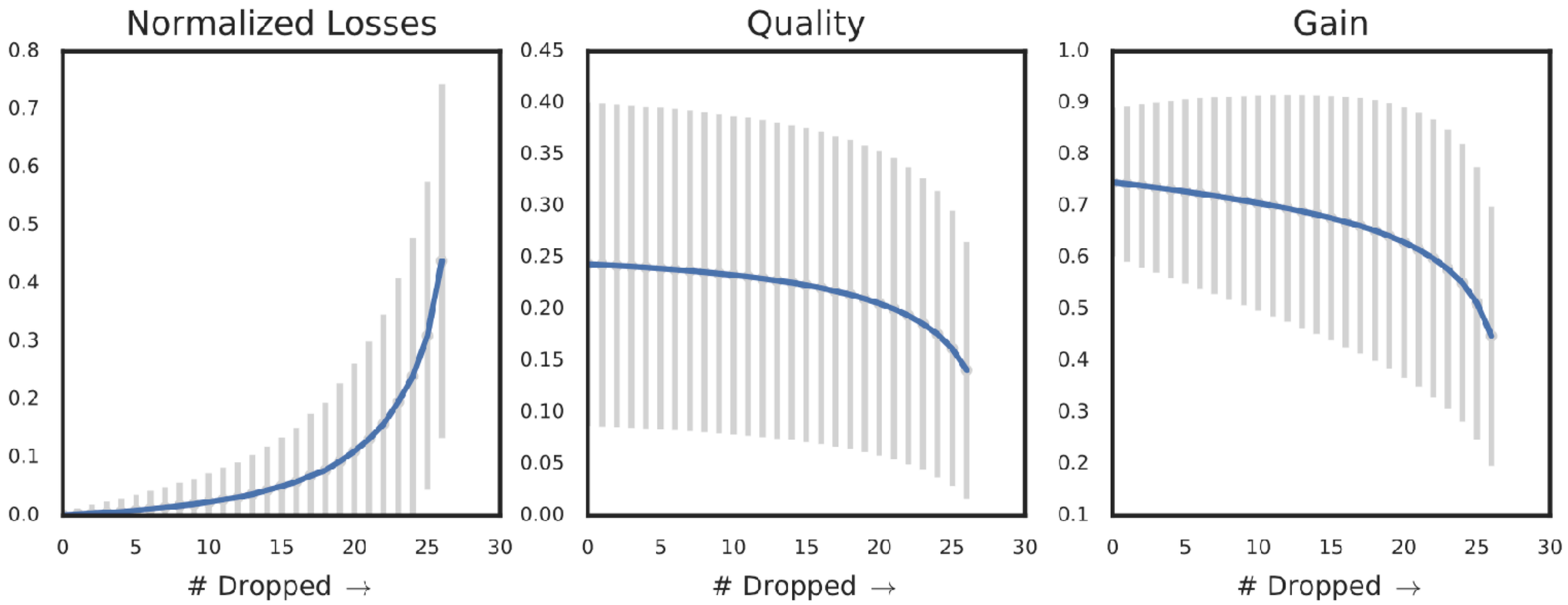
Transitive Transfer Test





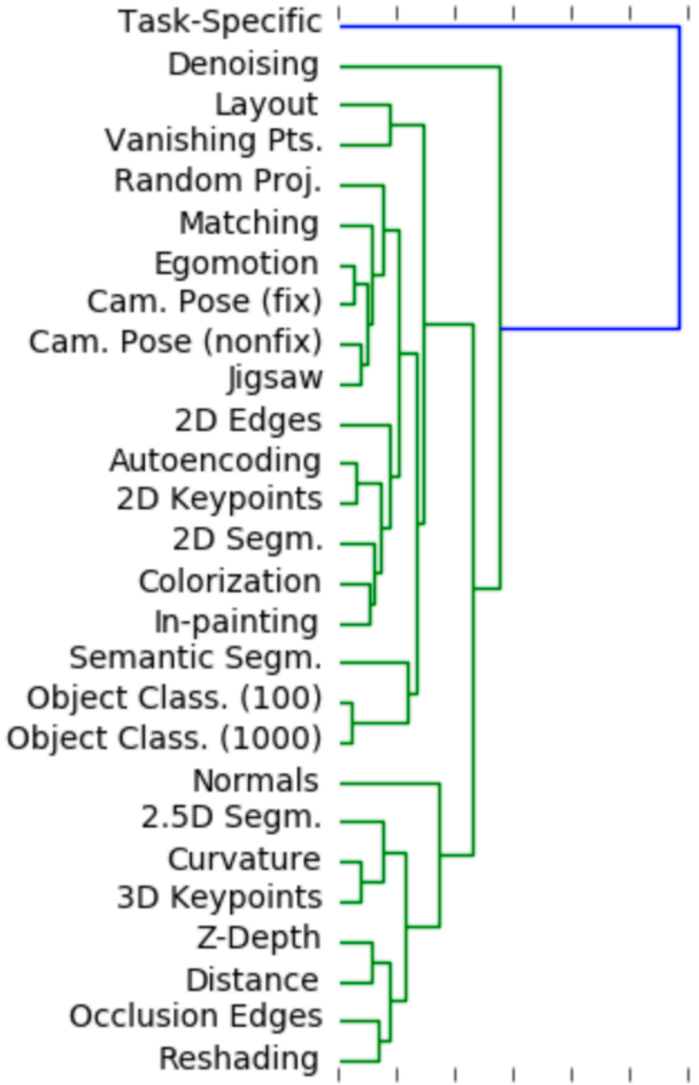


- Can transfer through long transitivity be utilized by Neural Net transfer functions?
- The test suggests no.



How much the choice of sampled tasks in dictionary matters?

Showing robustness as long as dictionary size doesn't drop significantly



Task Similarity Tree: Agglomerative Clustering of Tasks based on their Transfer Pattern

Down-Stream vs Middle Goal

How to let perception help a robot do something useful?

Gibson Environment

for Embodied Real-World Active Perception


Fei Xia


Zhi-Yang He

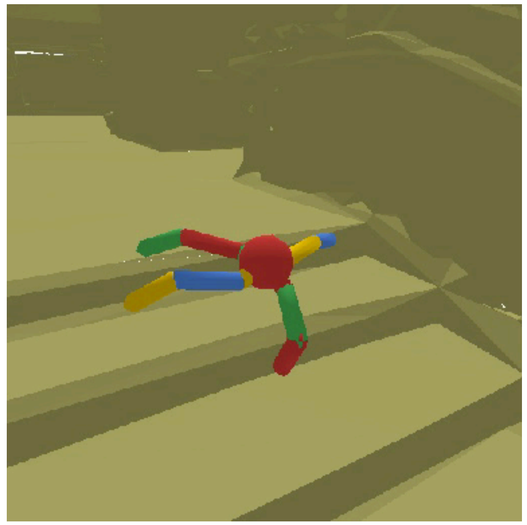

Sasha Sax


Jitendra Malik


Silvio Savarese

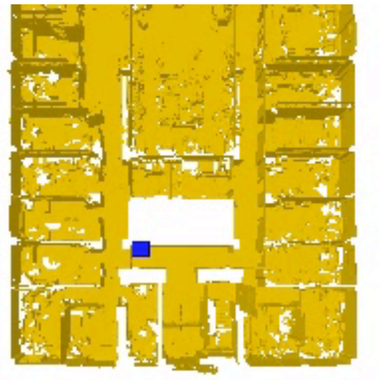
Active Agent

Active Agent



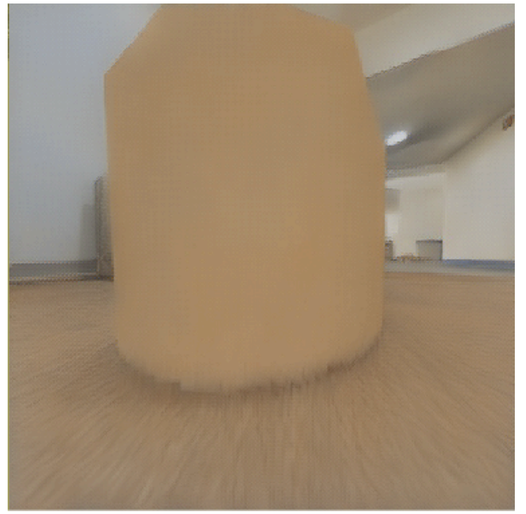
Gibson at a glance

Gibson at a glance

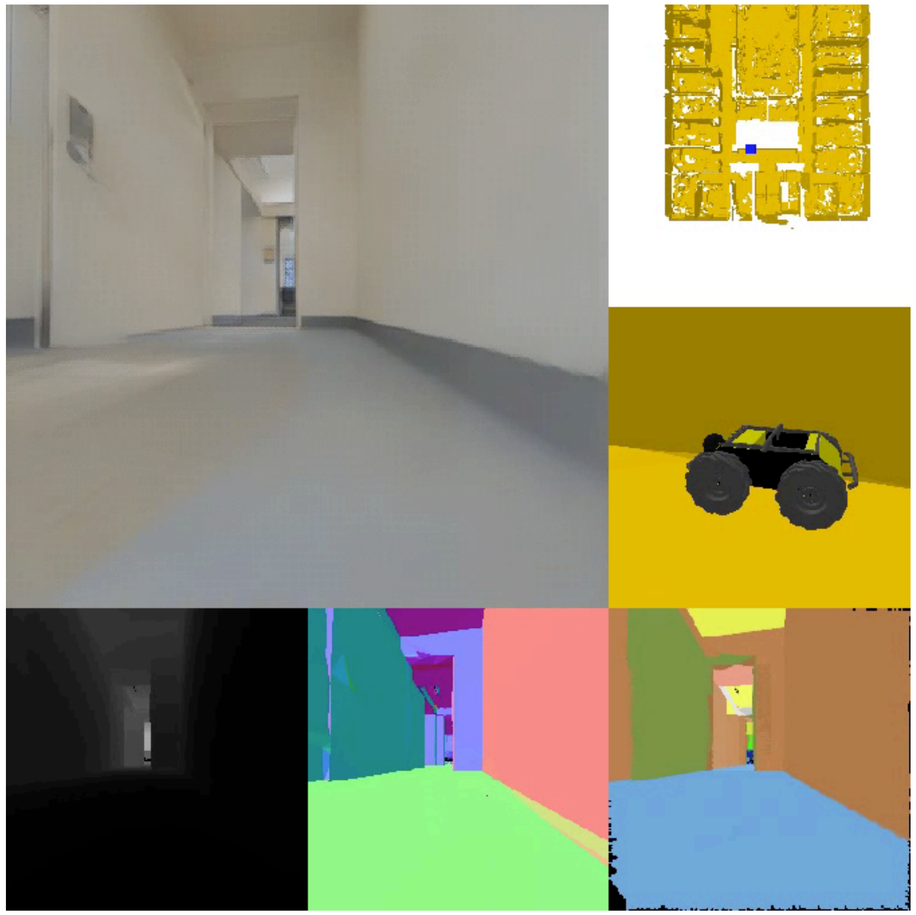


Active in a Large Real Space

Embodied & Subject to Physics



Constant Visual Observation from camera



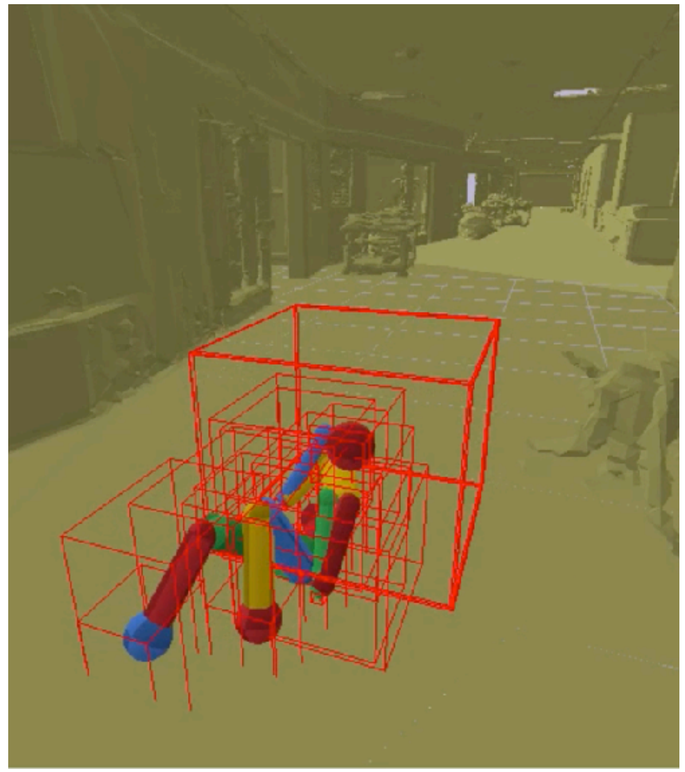
Environment

Perceptual Reinforcement Learning Results

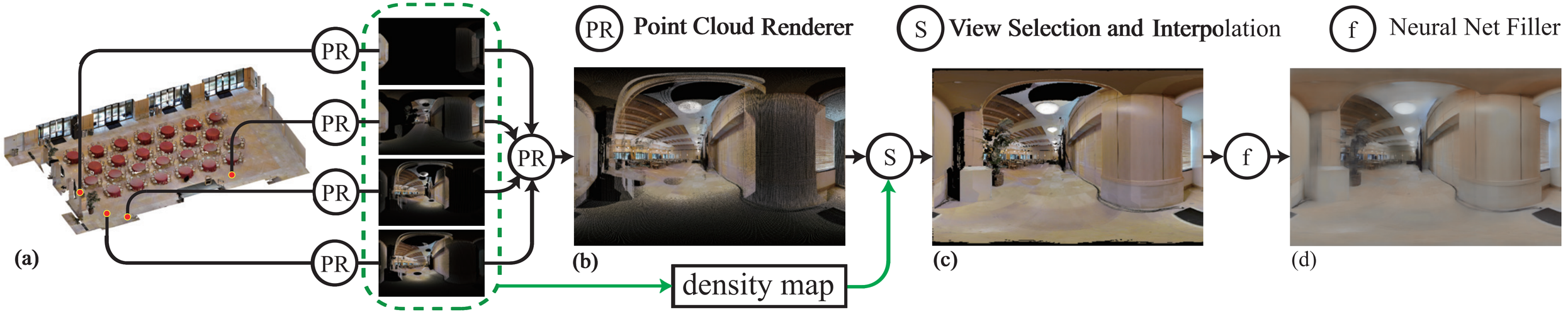
Components of Gibson

I. View Synthesis

II. Agent



III. Physics Integration



- Stochastic Identity Initialization
- Multi-Layer Perceptual Loss
- Color Moment Matching in Loss
- Fast Bilinear Interpolation and Dilated Convolution

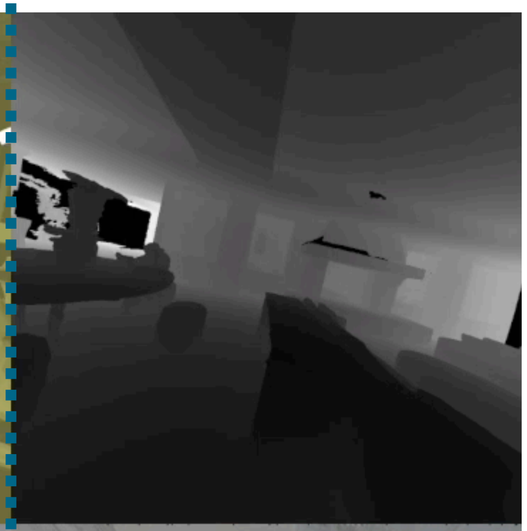
I. Neural Network based View Synthesis

Post neural network filling

Pre neural network filling

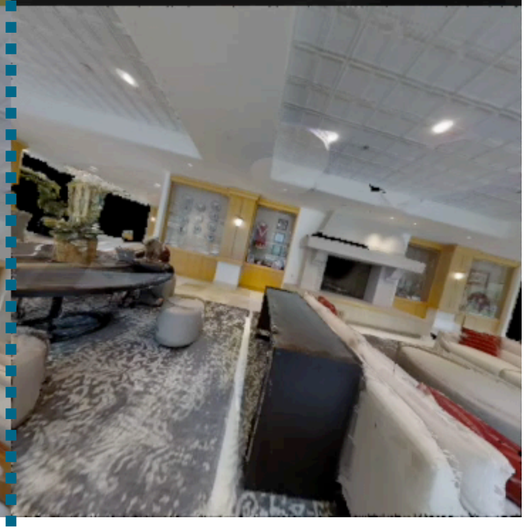
Physics

Depth



Post NN

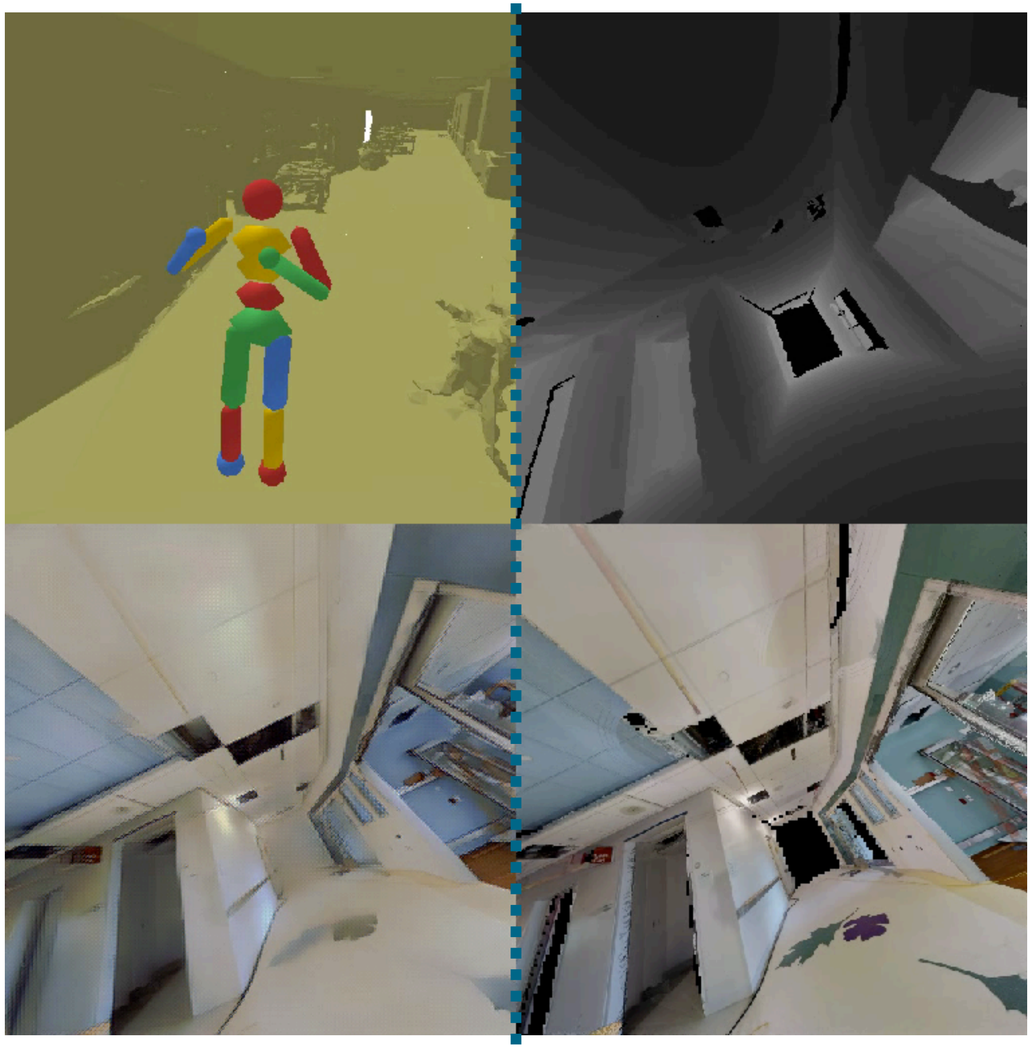
Pre NN



Neural Network filler fixes some of pathological geometric and occlusion errors

Post neural network filling

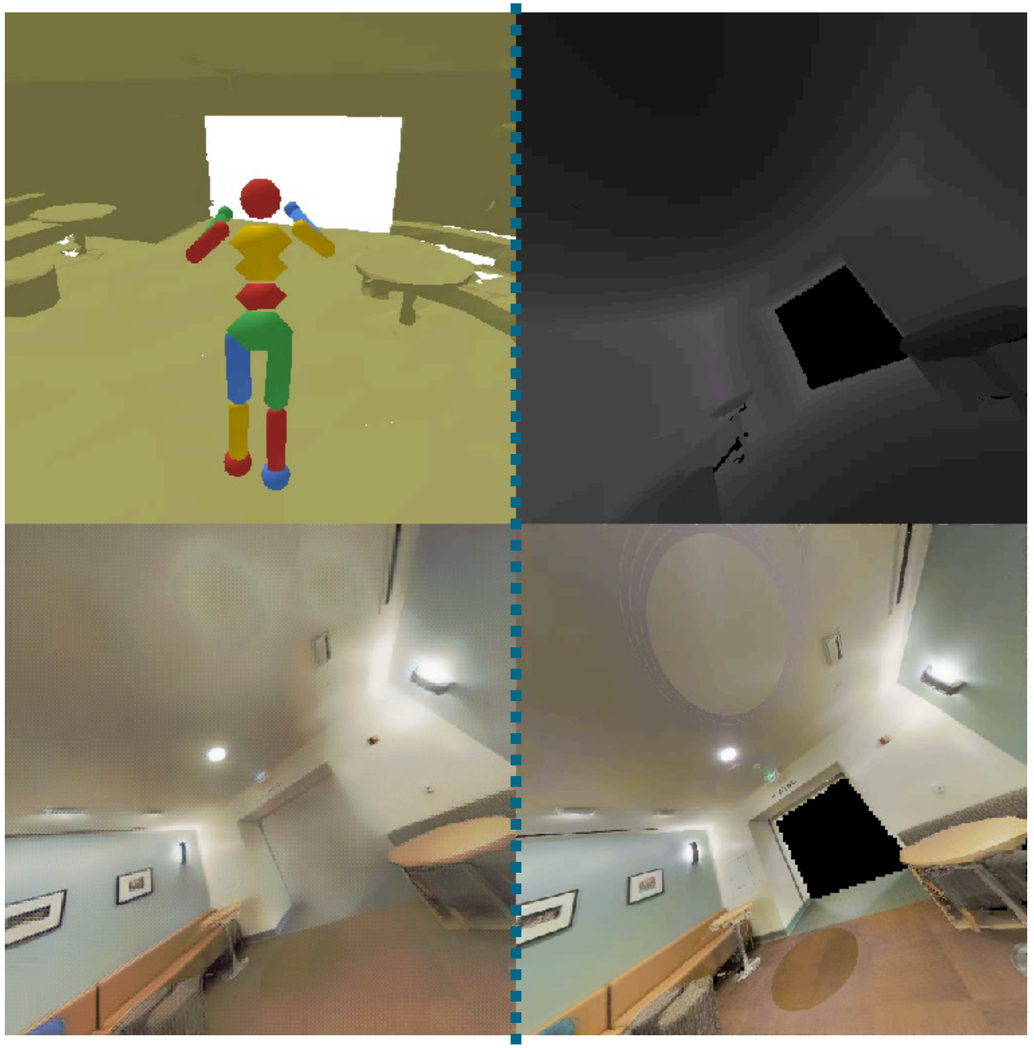
Pre neural network filling



Neural Network filler fixes some of pathological geometric and occlusion errors

Post neural network filling

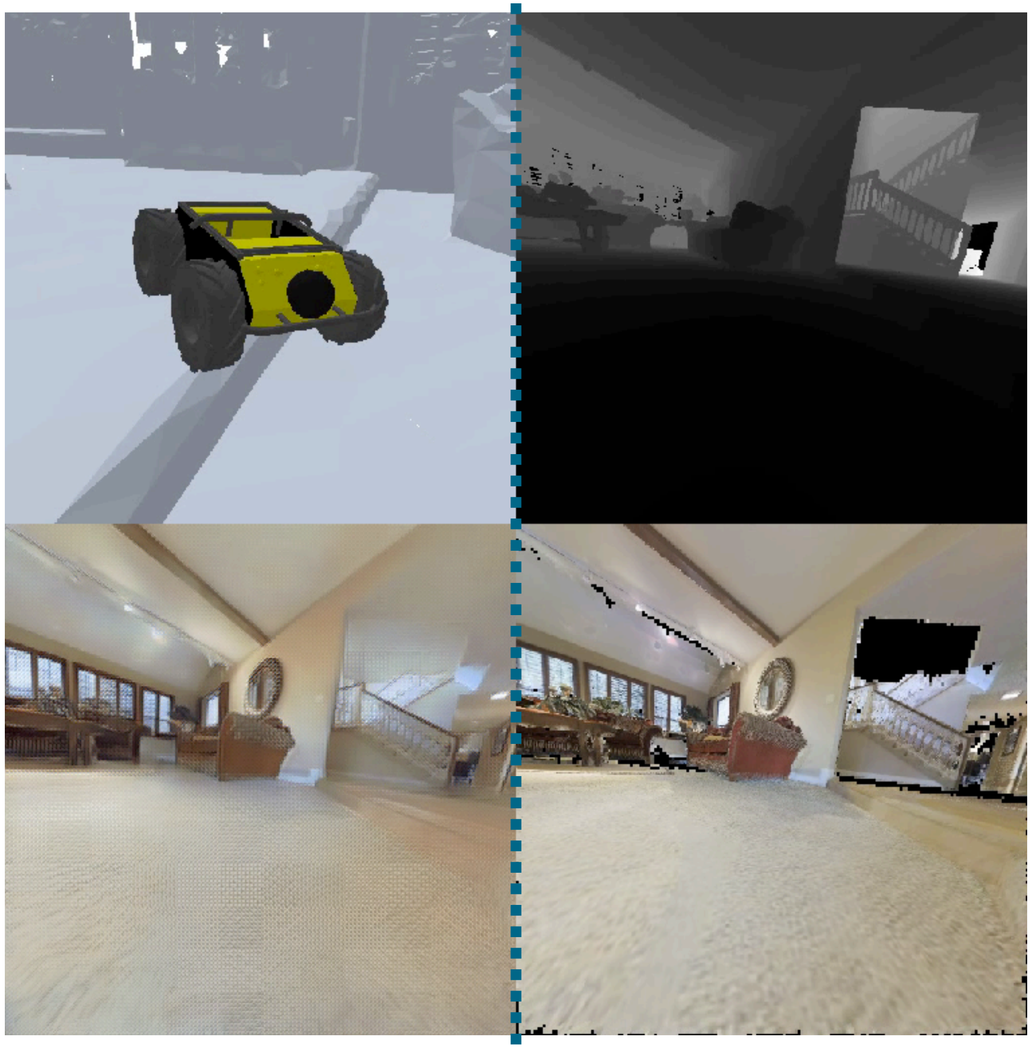
Pre neural network filling



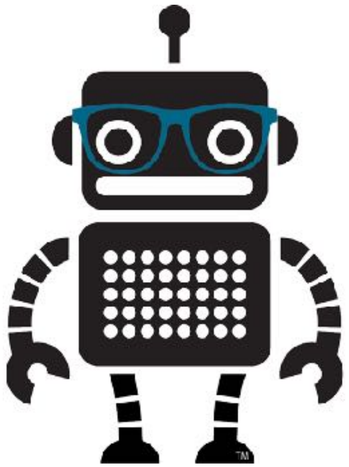
Neural Network filler fixes some of pathological geometric and occlusion errors

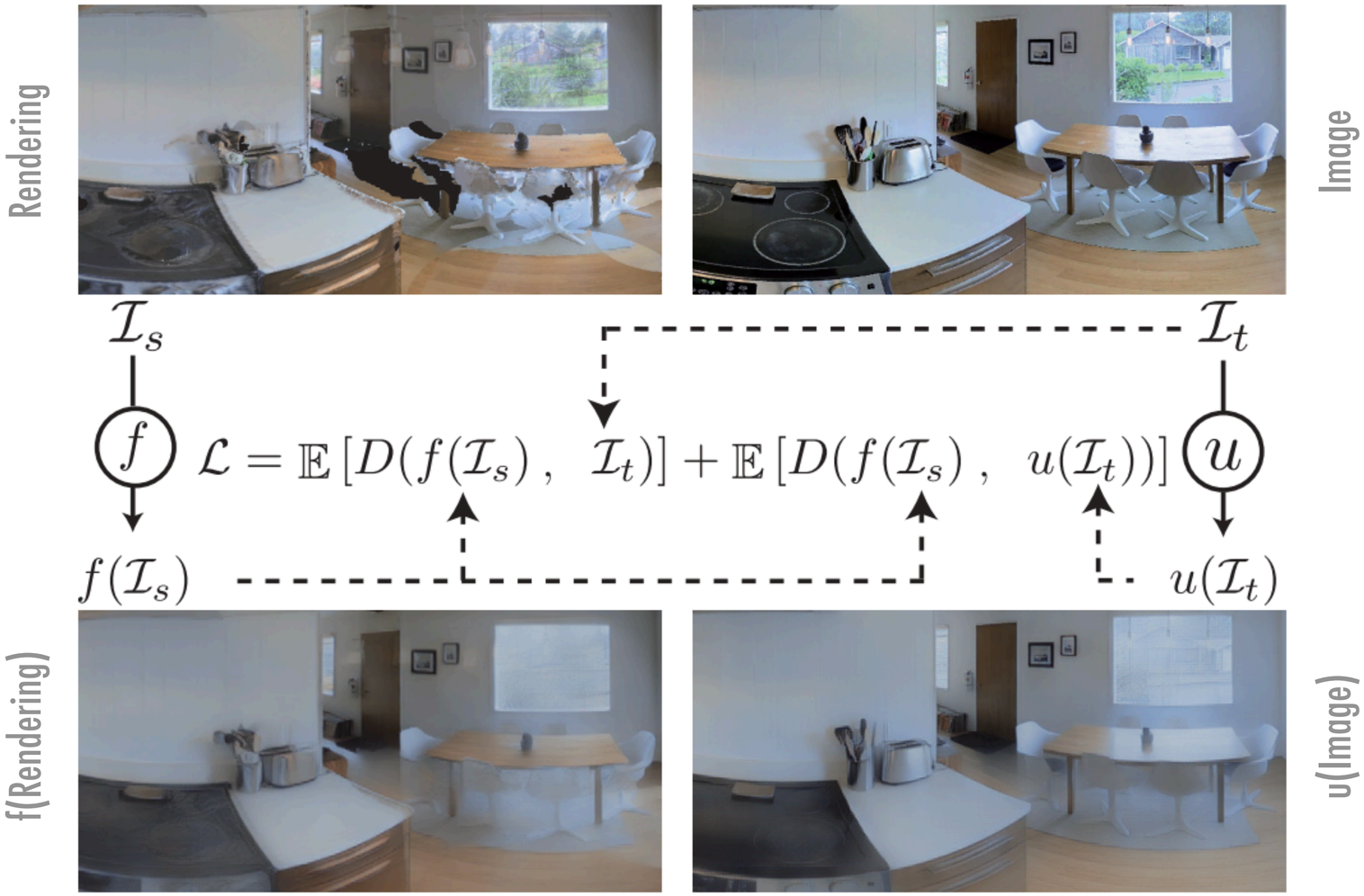
Post neural network filling

Pre neural network filling



Neural Network filler fixes some of pathological geometric and occlusion errors

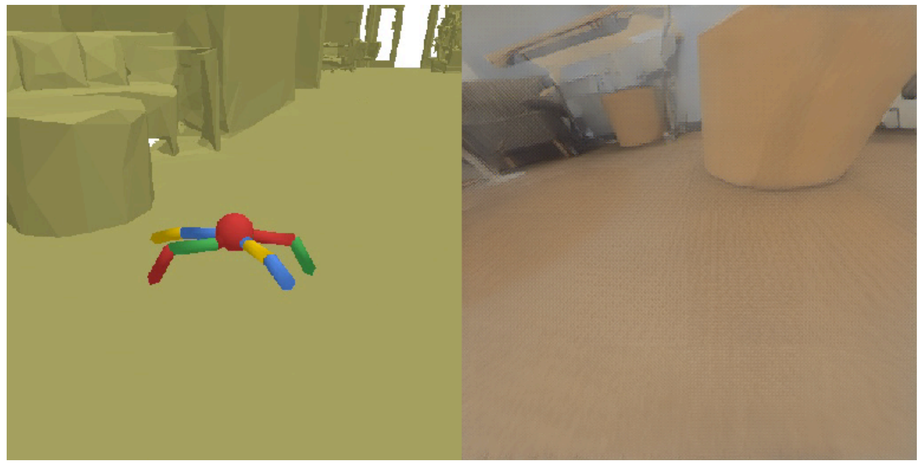




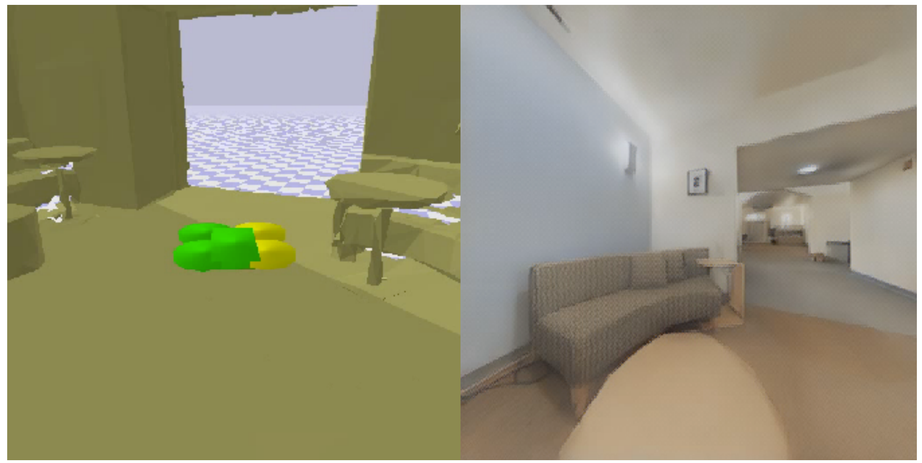
Transfer to Real World: goggles

II. Agent

Ant



Quadrotor

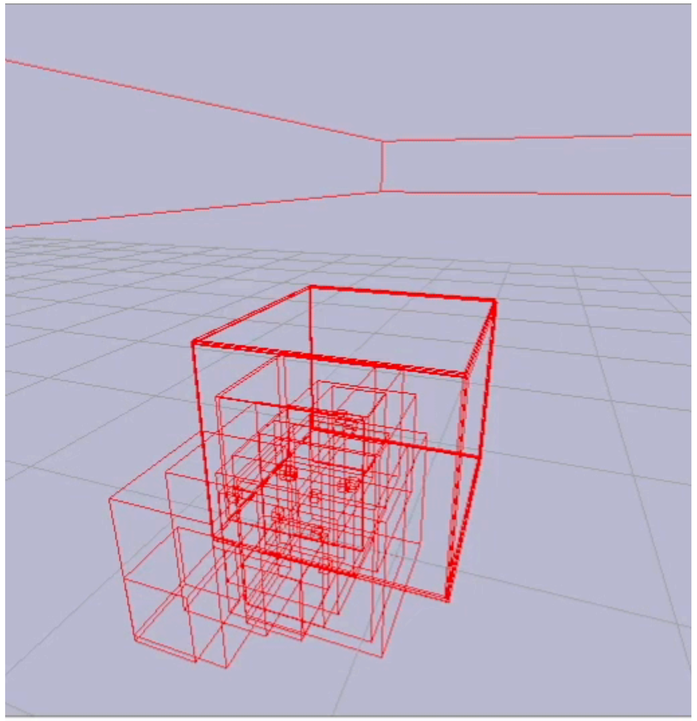


Humanoid

Husky

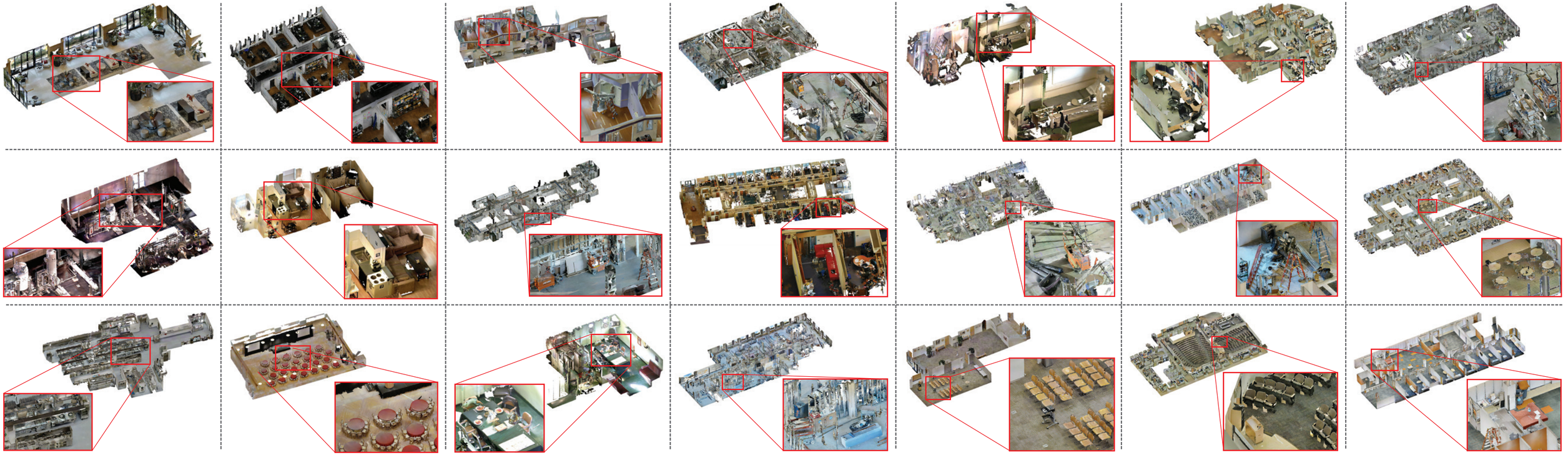
Convenient import of new agents

III. Physics Integration



Bullet Physical Simulator [1]

[1] http://bulletphysics.org/wordpress/

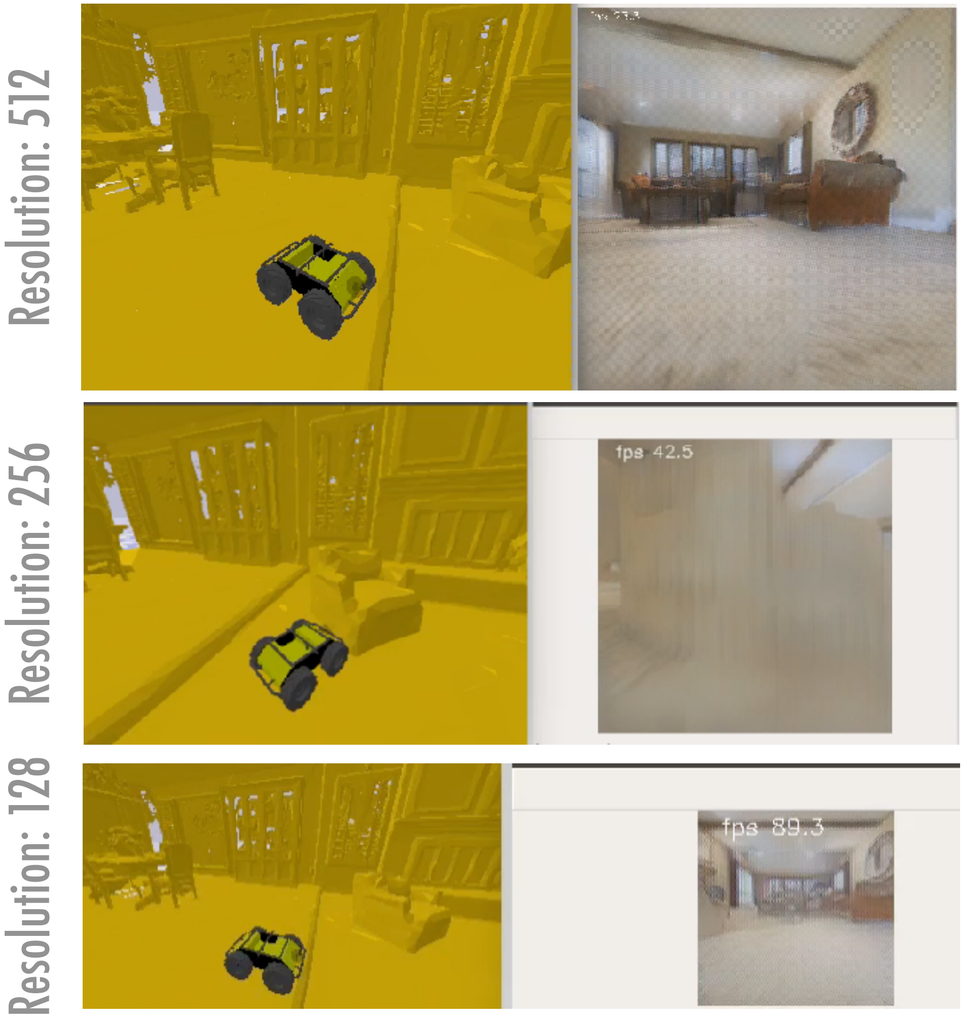


Database of Spaces: 1145 floor-spaces.

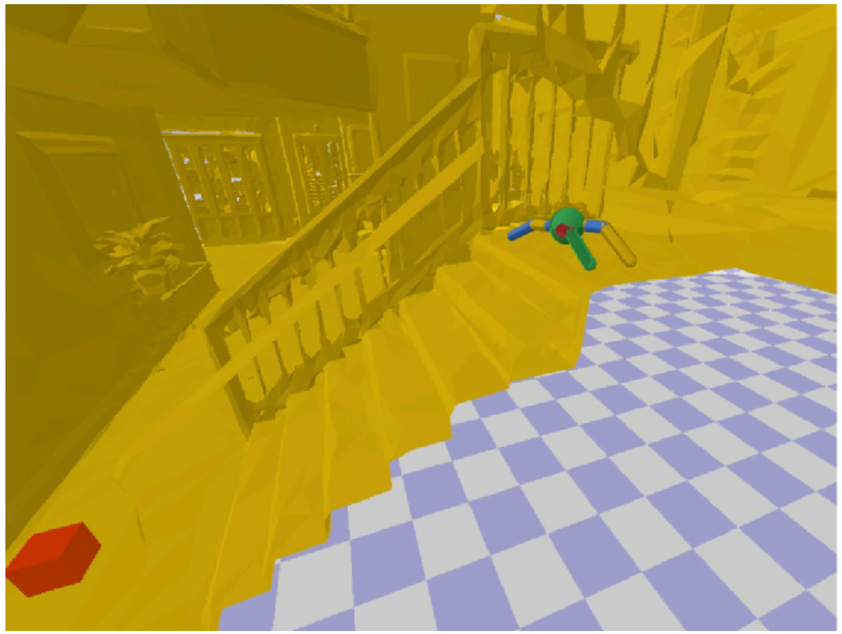
Stanford 2D3DS and Matterport3D Semantics integrated.

Speed
(reported on a single GPU)

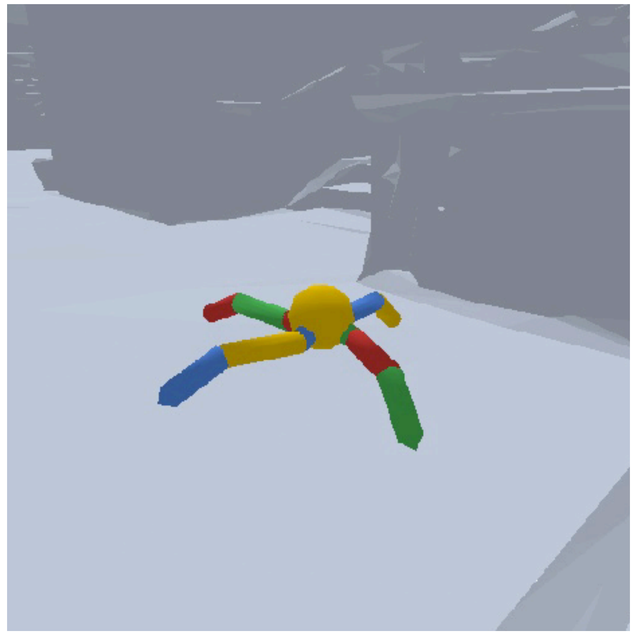
Resolution	128x128	256x256	512x512
Sensor (fps)	427.9	427.9	427.9
Depth Only (fps)	159.4	113.3	79.2
RGBD Pre-filled (fps)	81.5	50.9	33.3
RGBD Filled (fps)	73.6	42.7	18.3
Semantic Only (fps)	93.1	79.5	50.9
Surface Normal (fps)	89.3	73.7	45.4



training & testing (reinforcement) learning in different spaces



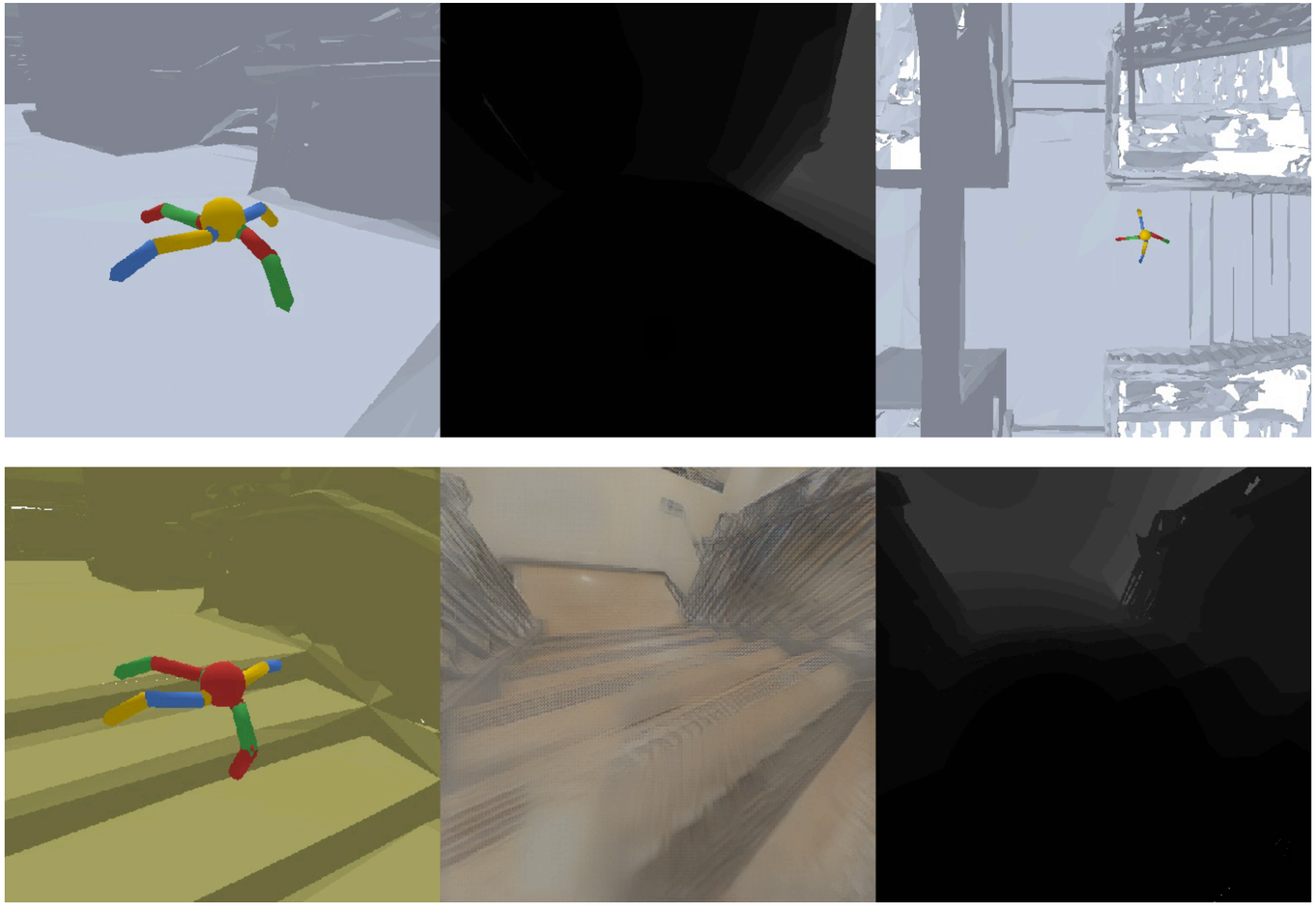
Training on one stairway



Testing on an unseen stairway

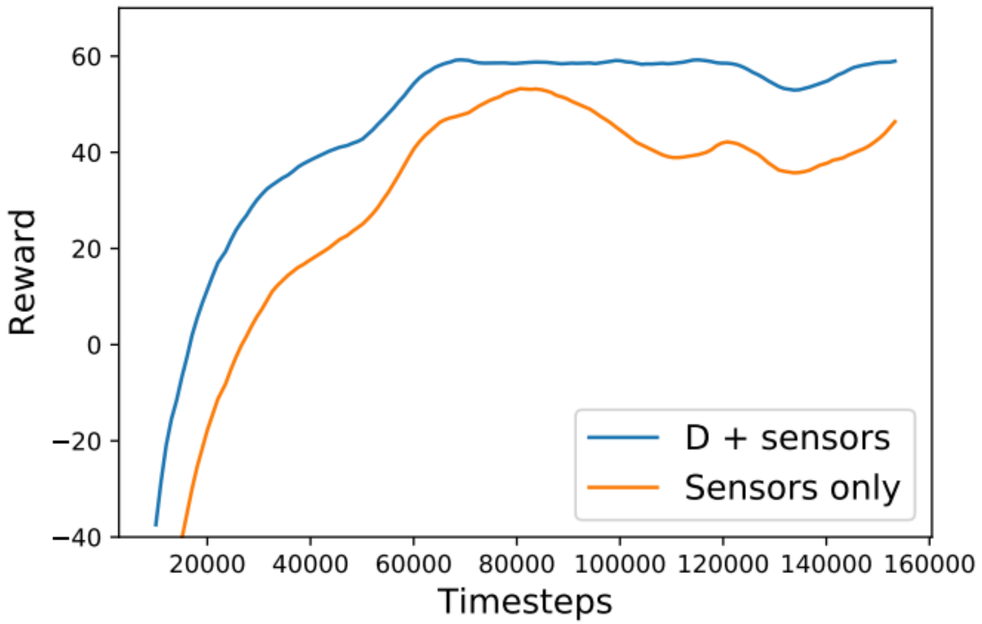
Climbing Stairways

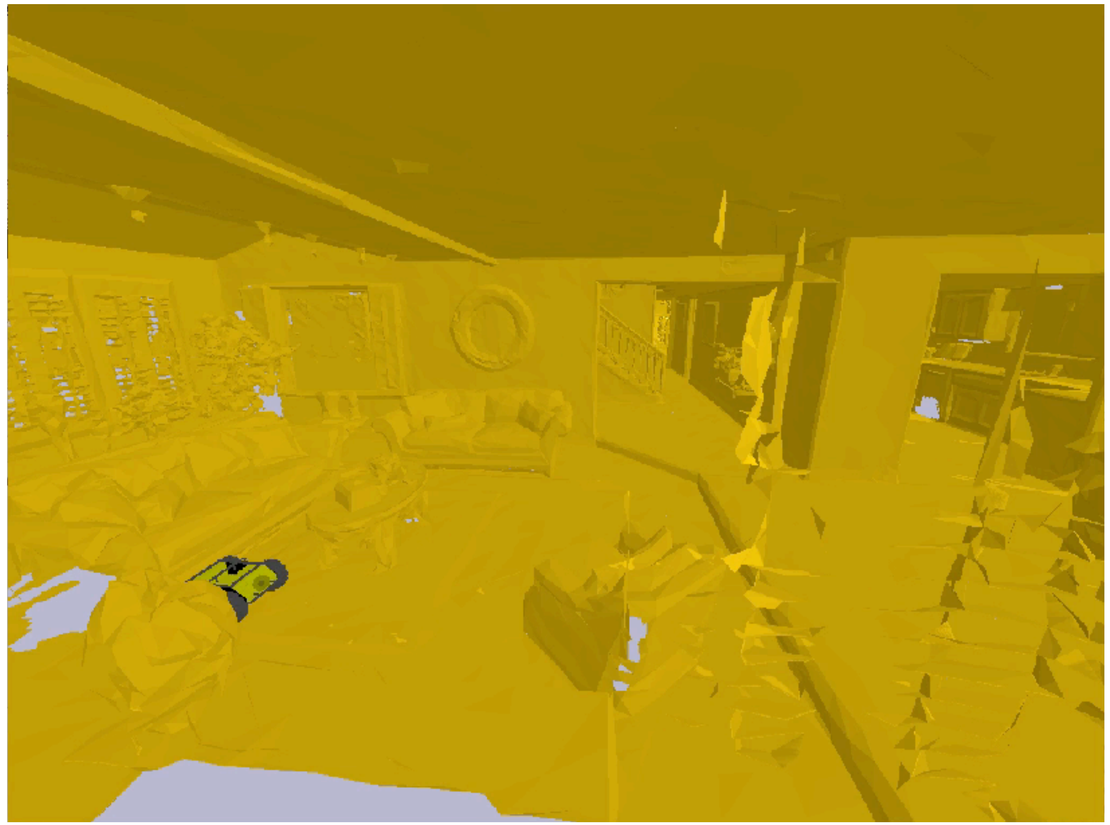
Reinforcement Learning on Perceptual Agents



Locomotion: sensorimotor control for climbing stairs

Reinforcement Learning on Perceptual Agents





Local planner: Learning obstacle avoidance using visual signal

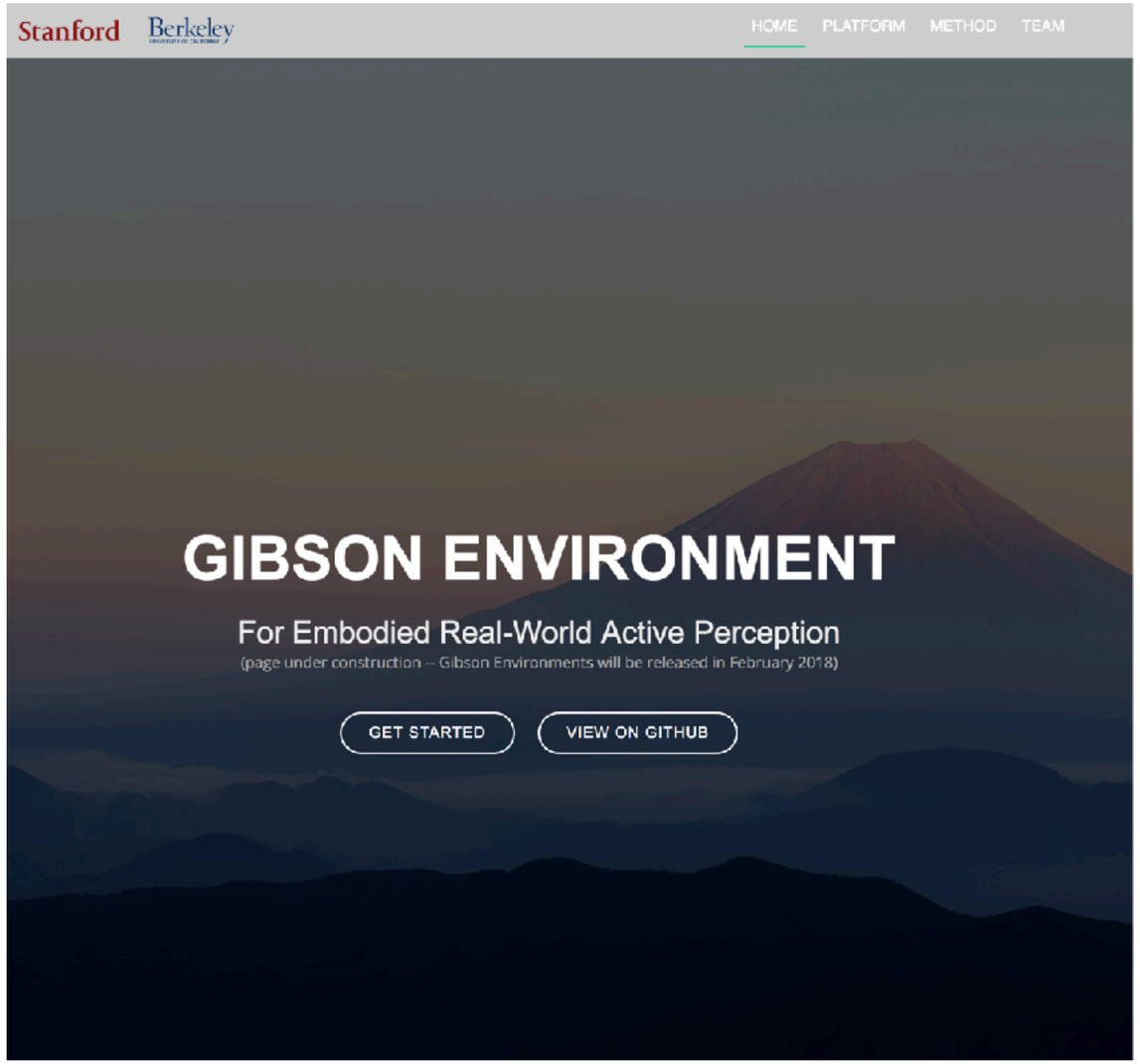
Reinforcement Learning on Perceptual Agents

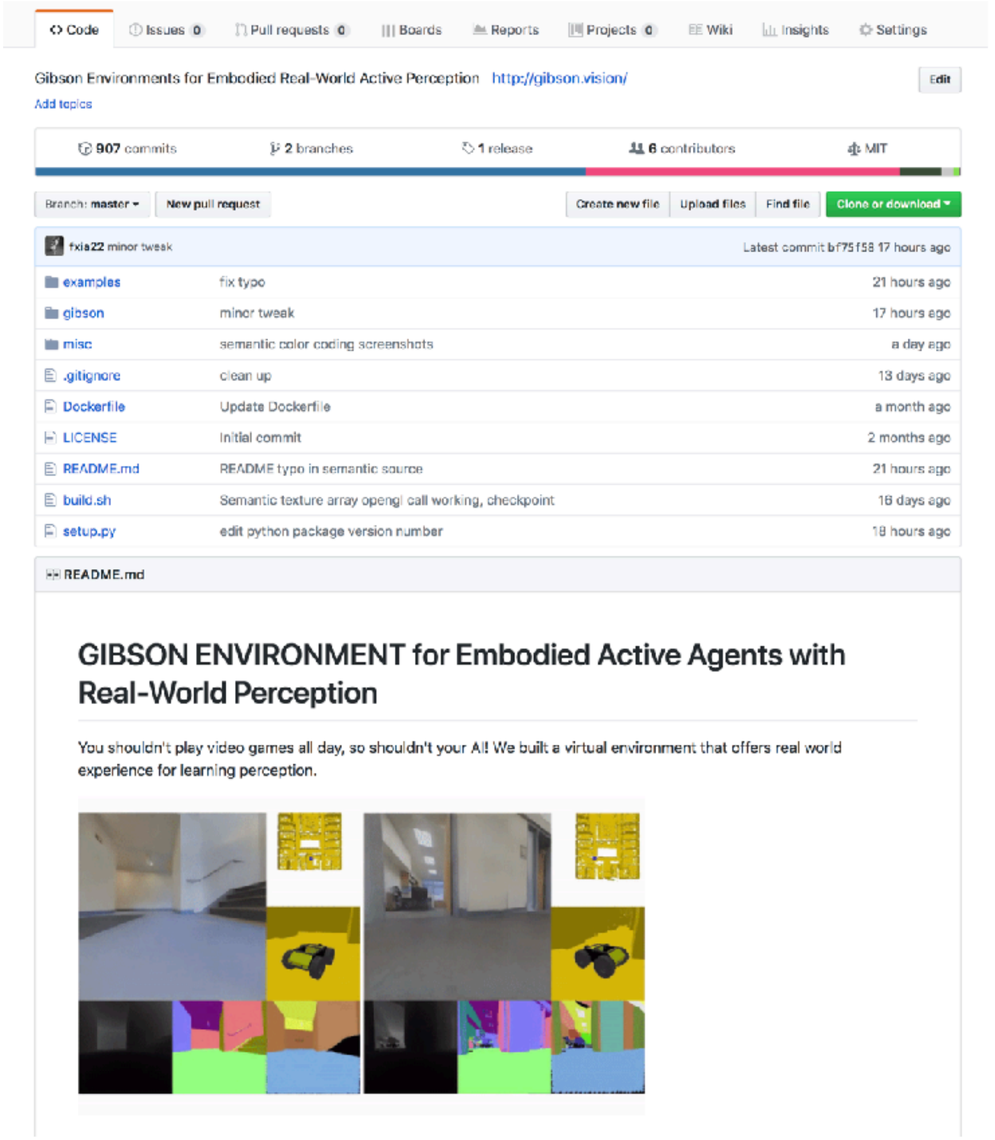
Episode 10

Episode 150

Long navigation: travel to a distant destination

http://gibson.vision/





Thank you!

Amir R. Zamir
zamir@cs.stanford.edu