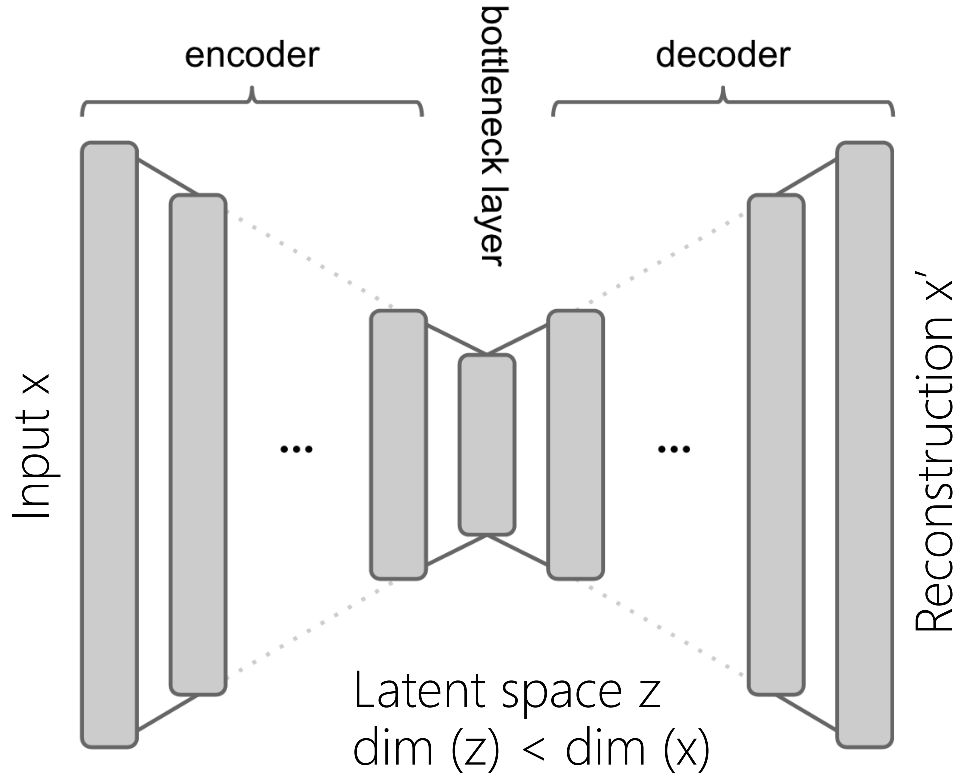
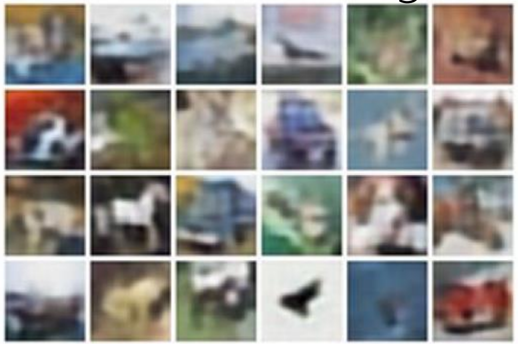
Lecture 10 Recap

Training Autoencoders

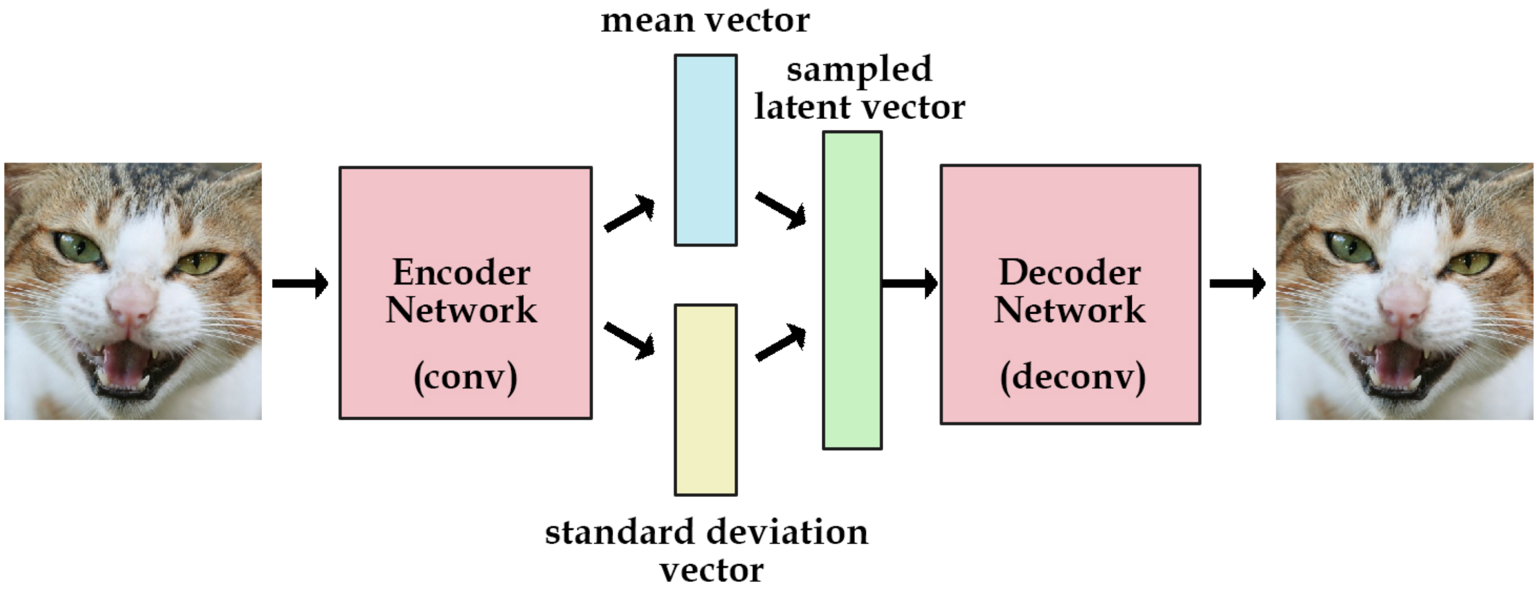


Input images

Reconstructed images



Variational Autoencoders (VAE)



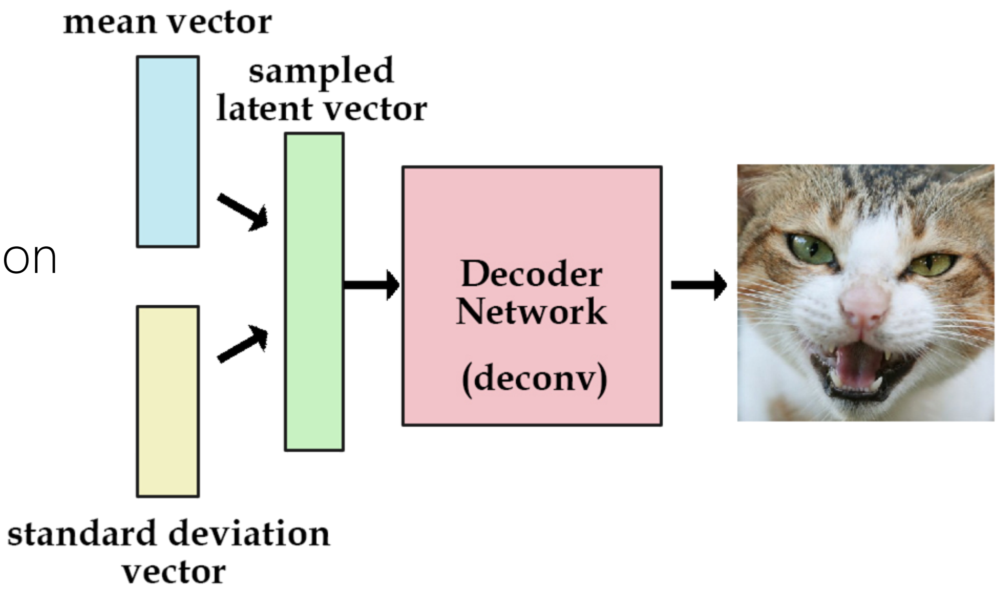
KL-Div Loss in latent space, forcing a unit Gaussian distribution
-> now the latent vector becomes a distribution

http://kvfrans.com/variational-autoencoders-explained/

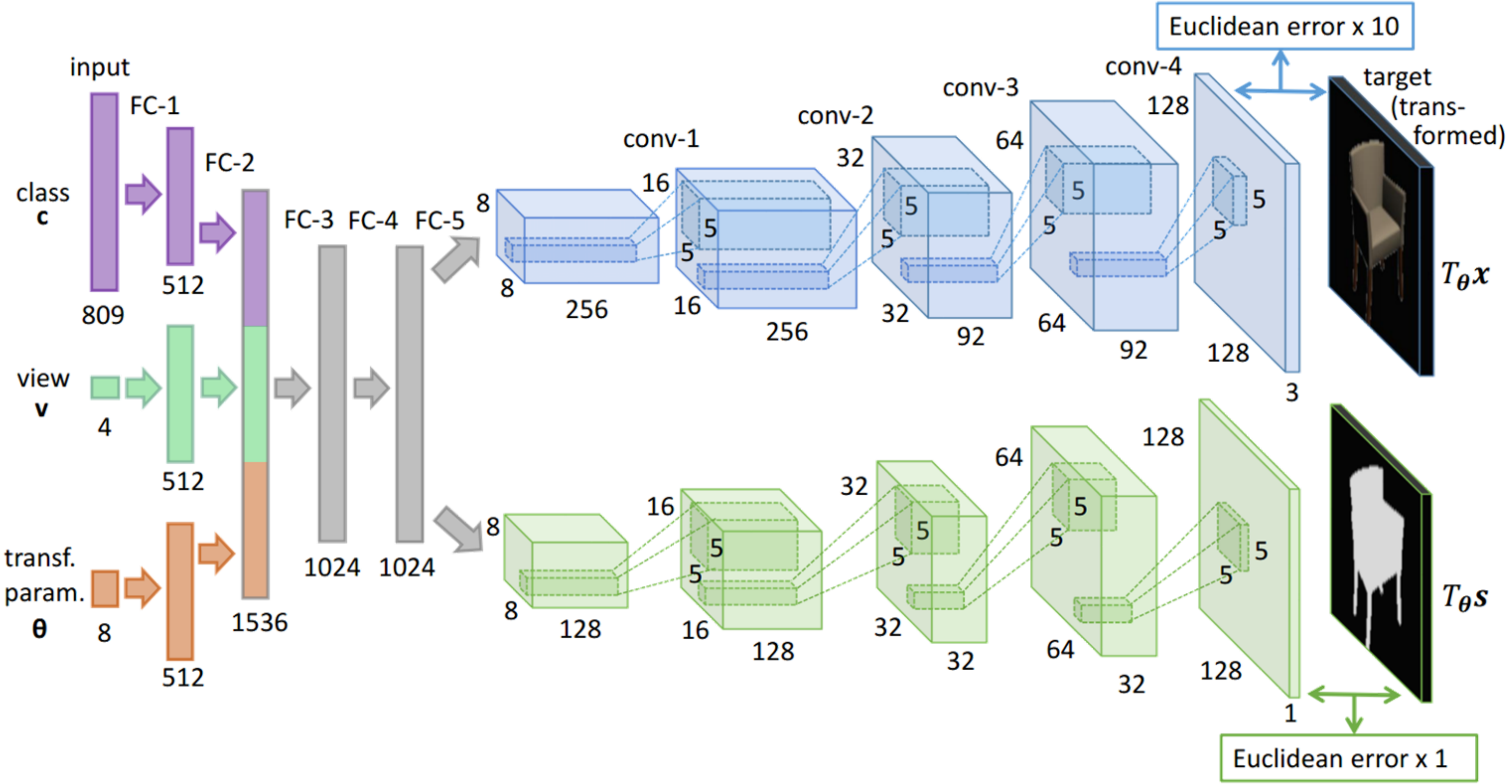
Variational Autoencoders (VAE)

- After training, generate random samples

Sample from the distribution
(e.g., unit Gaussian

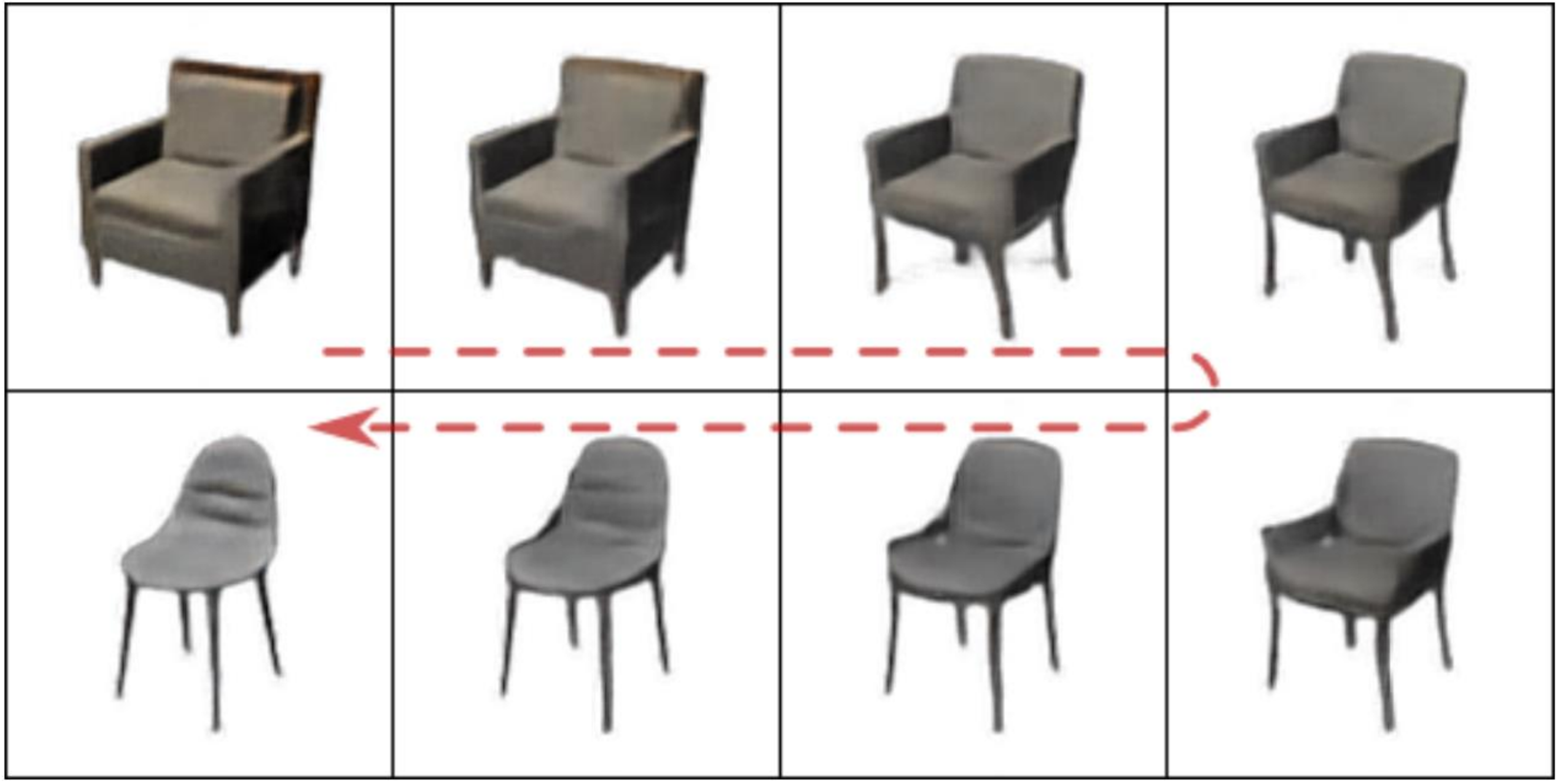


Generative Models



[Dosovitsky et al. 14] Learning to Generate Chairs

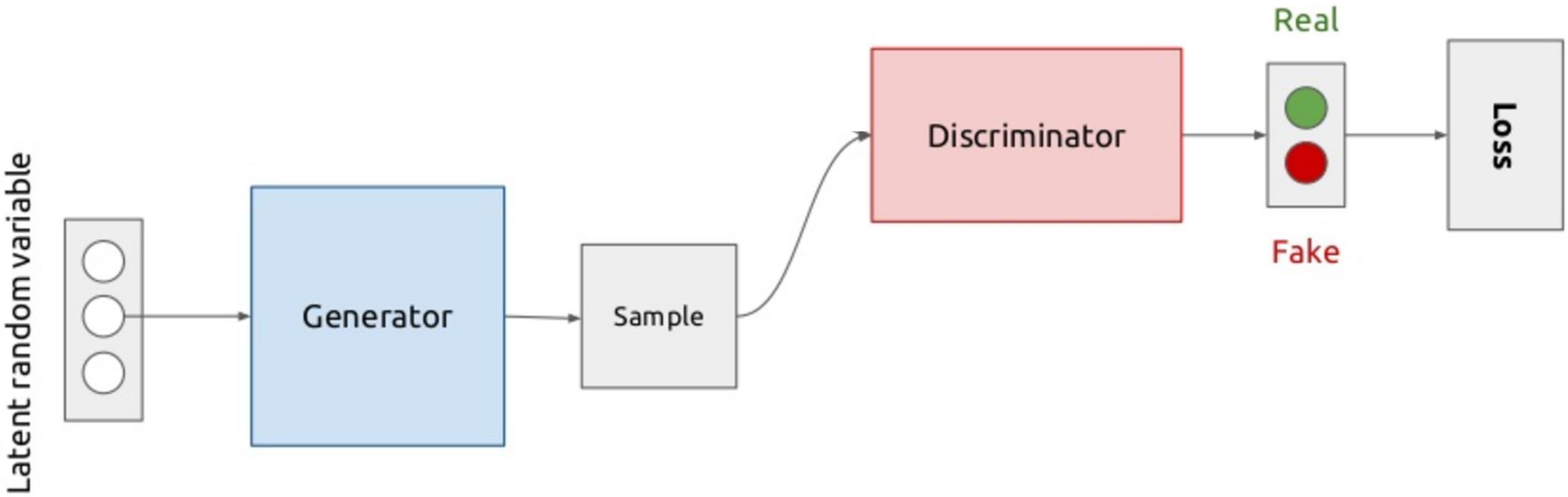
Generative Models



Interpolation between two chair models

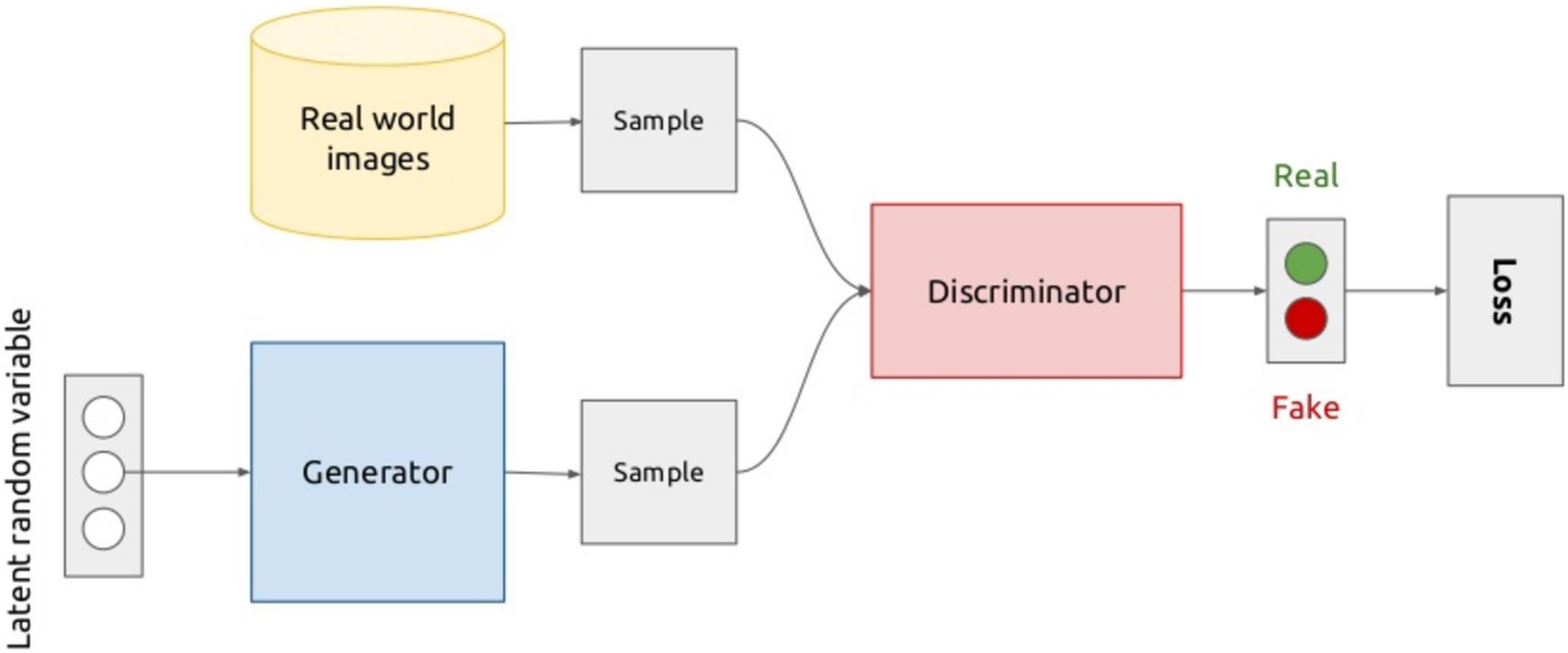
[Dosovitsky et al. 14] Learning to Generate Chairs

Generative Adversarial Networks (GANs)



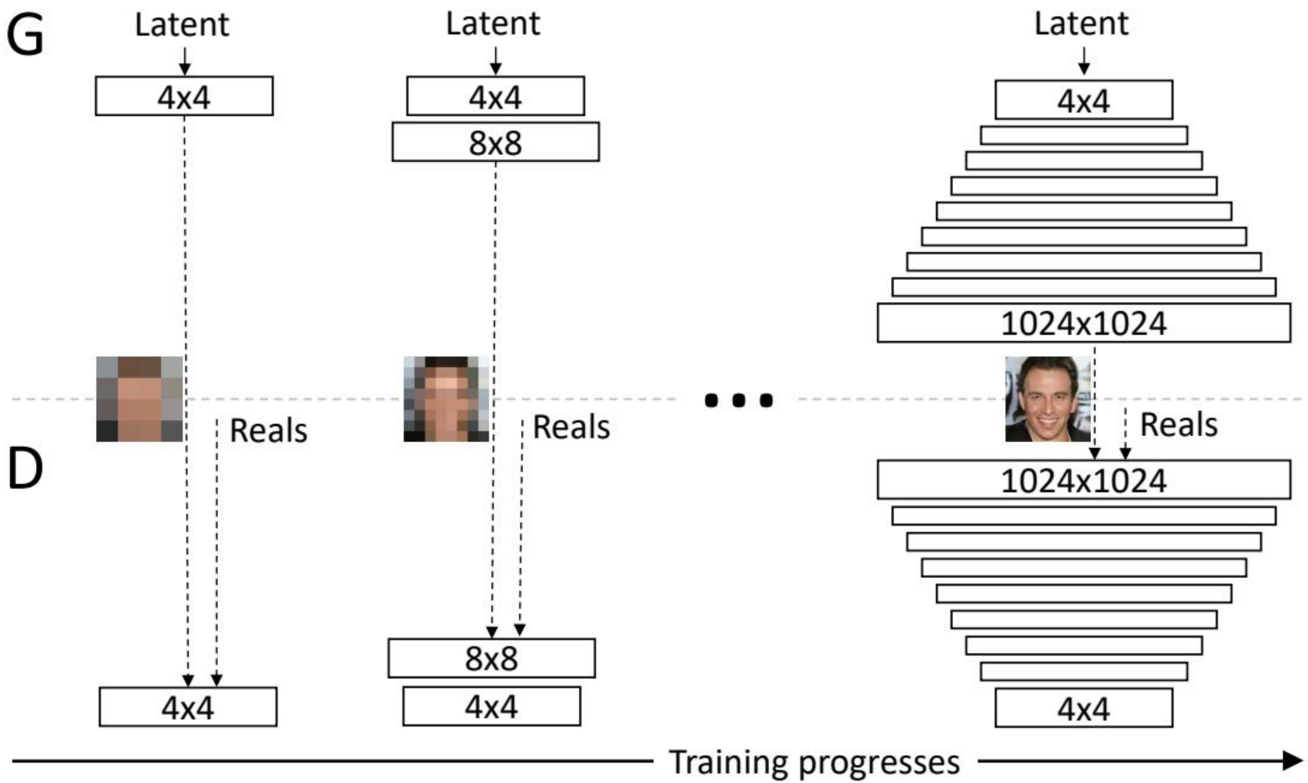
[Goodfellow et al. 14] GANs (slide McGuinness)

Generative Adversarial Networks (GANs)



[Goodfellow et al. 14] GANs (slide McGuinness)

Progressive GANs

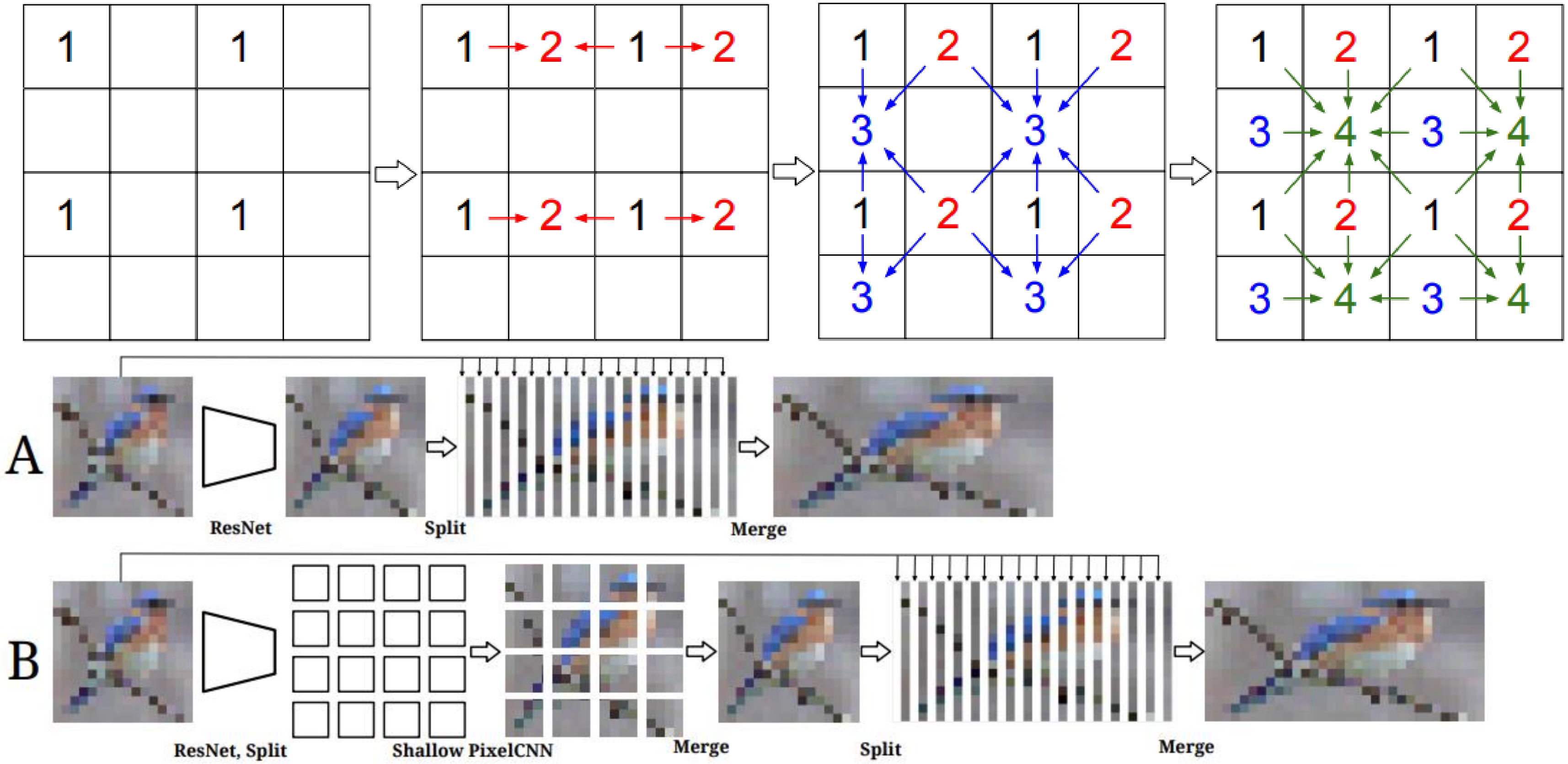


https://github.com/tkarras/progressive_growing_of_gans [Karras et al. 17]

Progressive GANs

https://github.com/tkarras/progressive_growing_of_gans [Karras et al. 17]

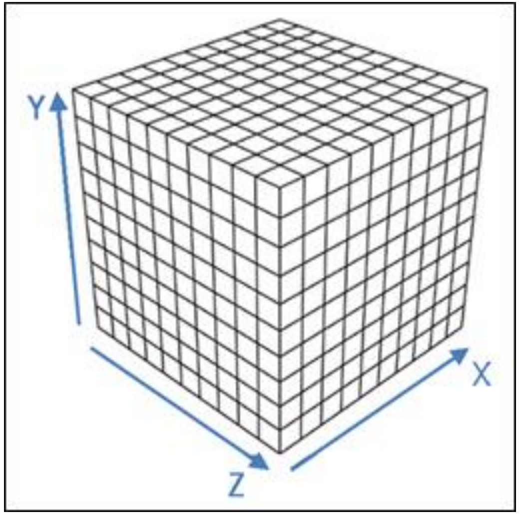
Other Generative Models: Autoregressive Models



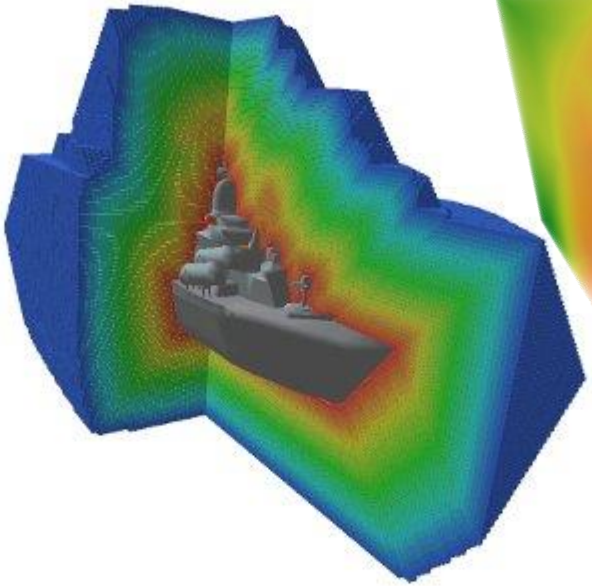
Parallel Multiscale Autoregressive Density Estimation [Reed et al. 17]

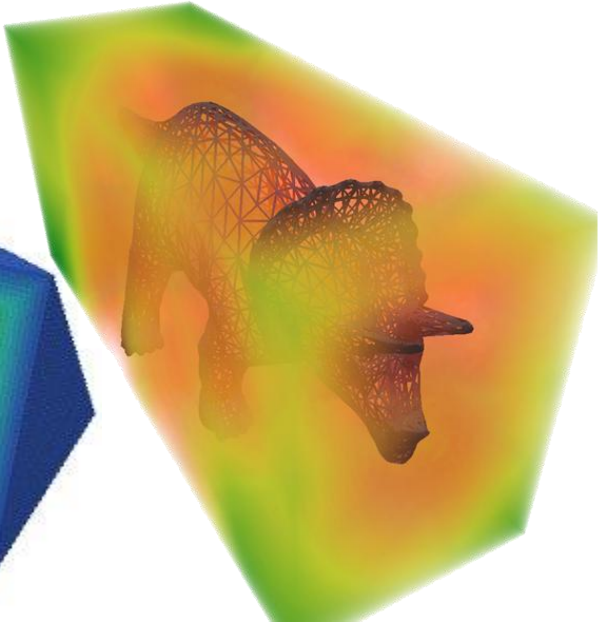
3D Convolutions

On volumetric data structures



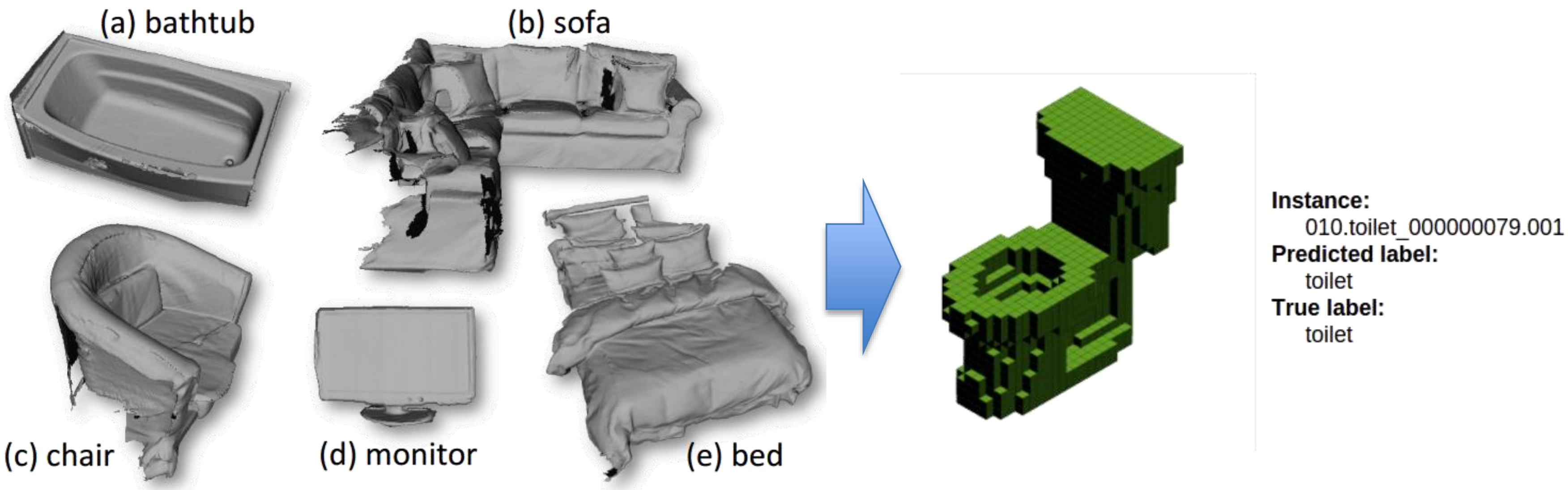
(binary) Voxel Grid





Implicit functions: e.g., signed distance field

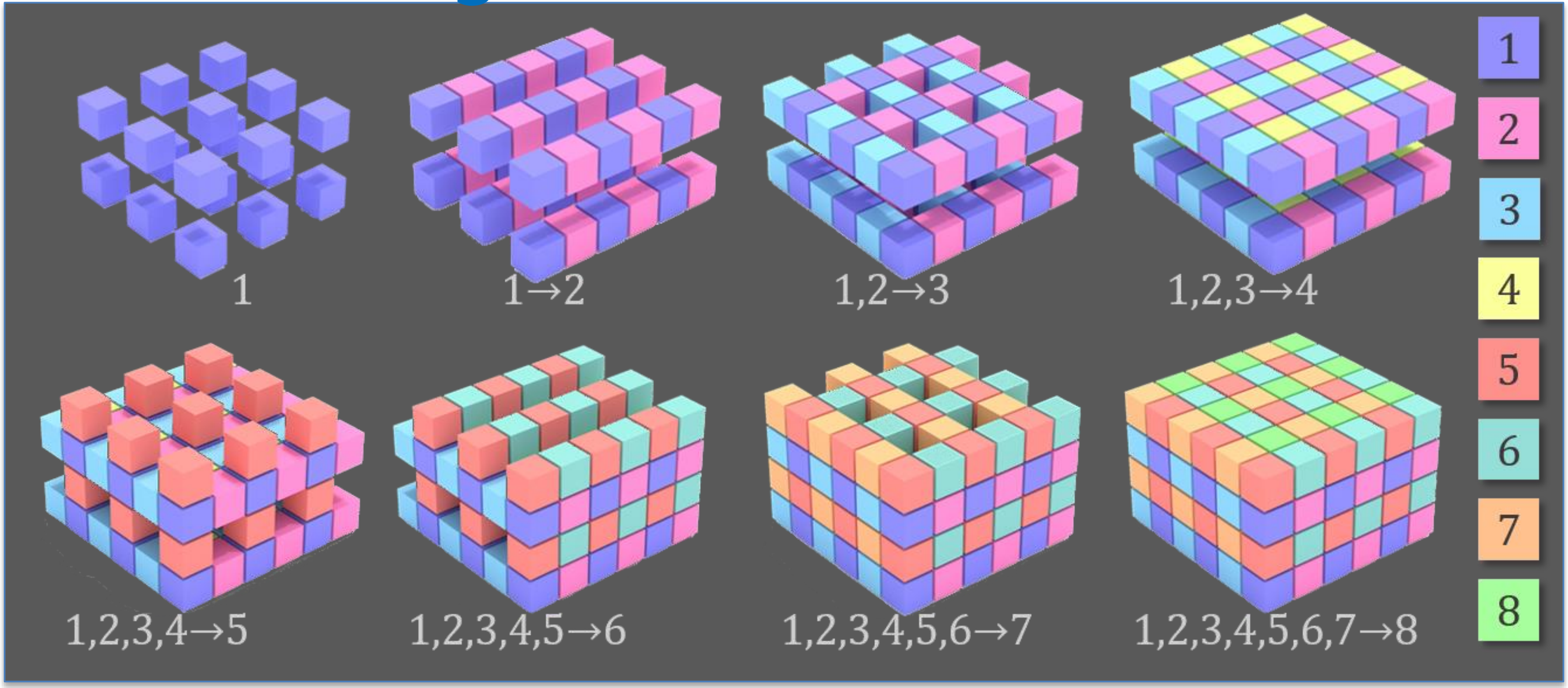
3D Classification



Class from 3D model (e.g., obtained with Kinect Scan)

[Maturana et al. 15] & [Qi et al. 16] 3D vs Multi-view

Dependent Predictions: Autoregressive Neural Networks

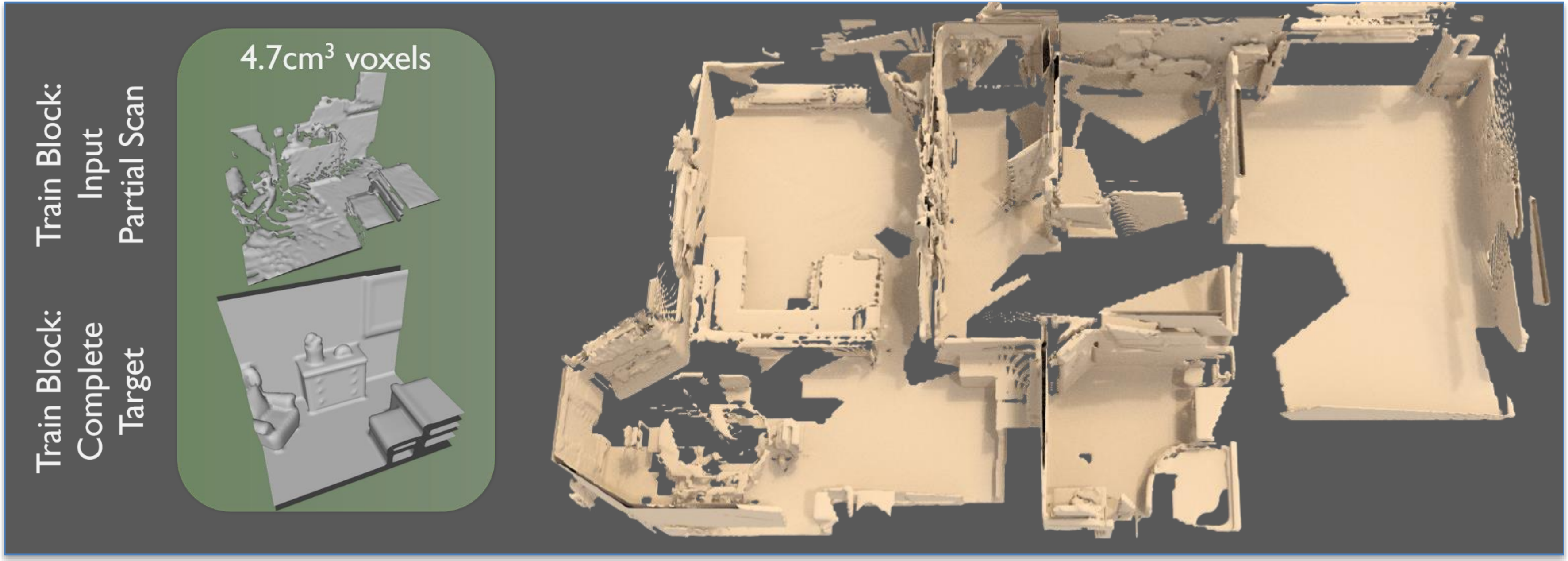


[Dai et al.]: ScanComplete

Spatial Extent: Fully Convolutional for Arbitrary Sizes

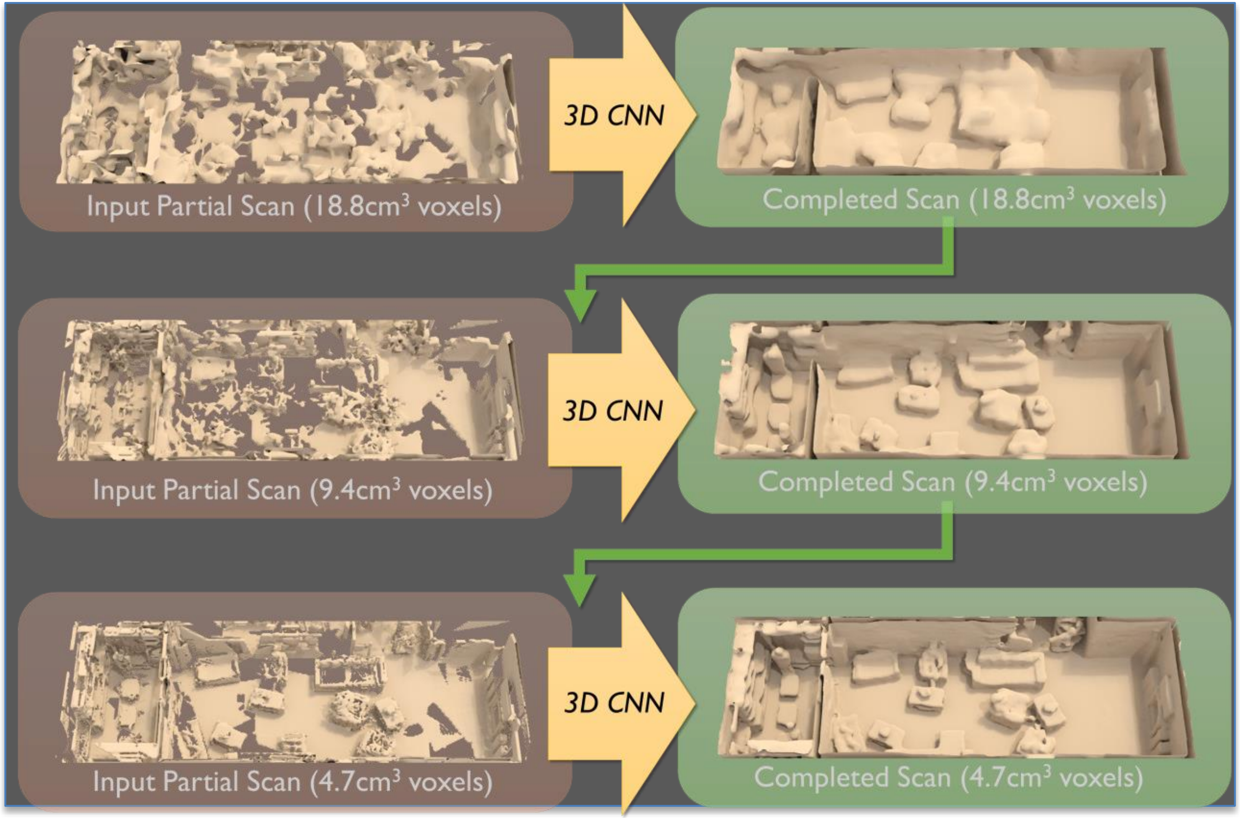
Train on crops of scenes

Test on entire scenes



[Dai et al.]: ScanComplete

Spatial Extent: Coarse-to-Fine Predictions

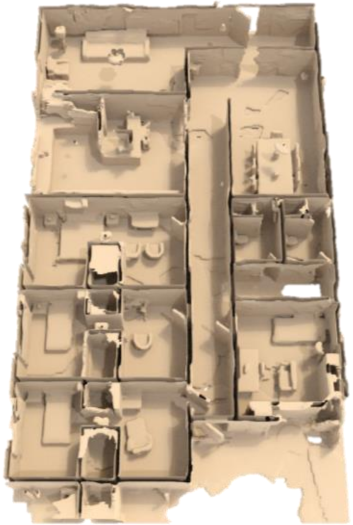


[Dai et al.]: ScanComplete

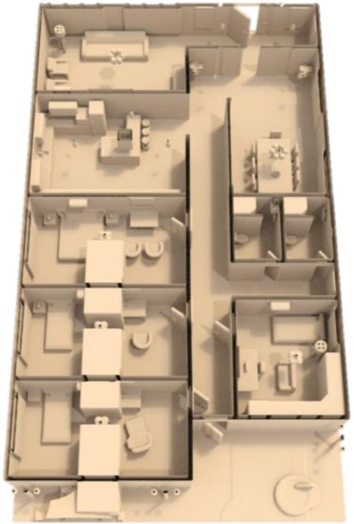
ScanComplete: Completing 3D Scans

Input

Completion



Ground Truth

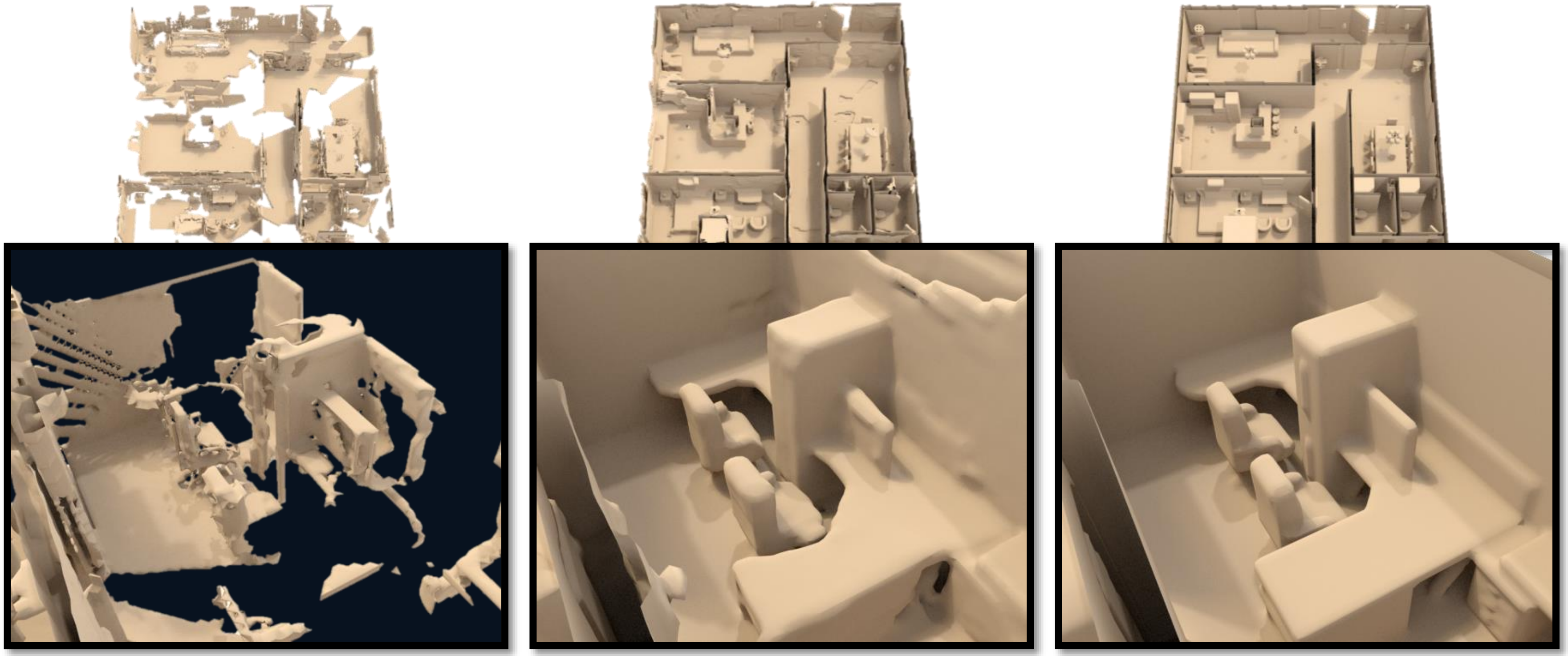


Scans from SUNCG [Song et al. 2017]

[Dai et al.]: ScanComplete

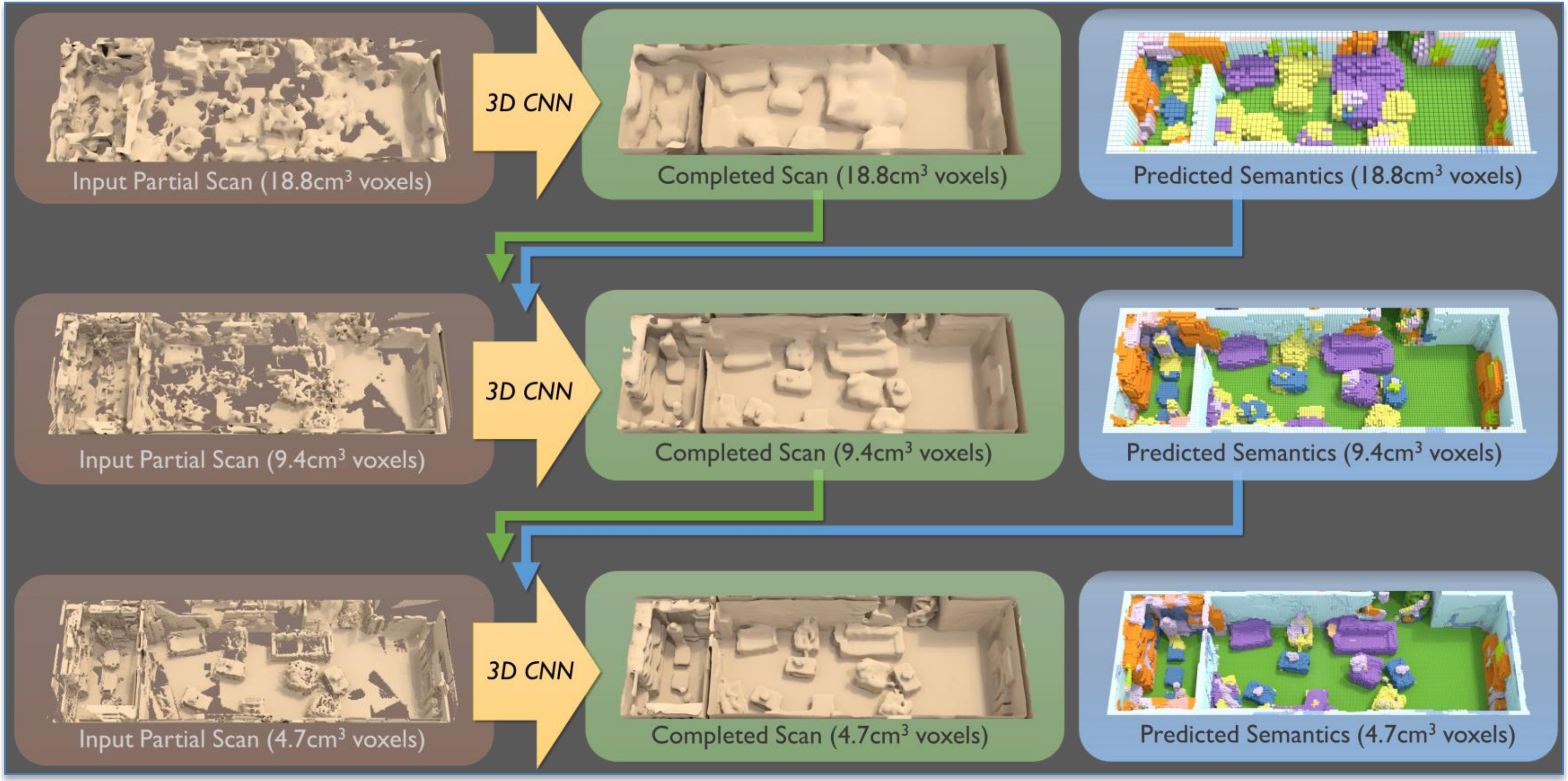
ScanComplete: Completing 3D Scans

Input | Completion | Ground Truth



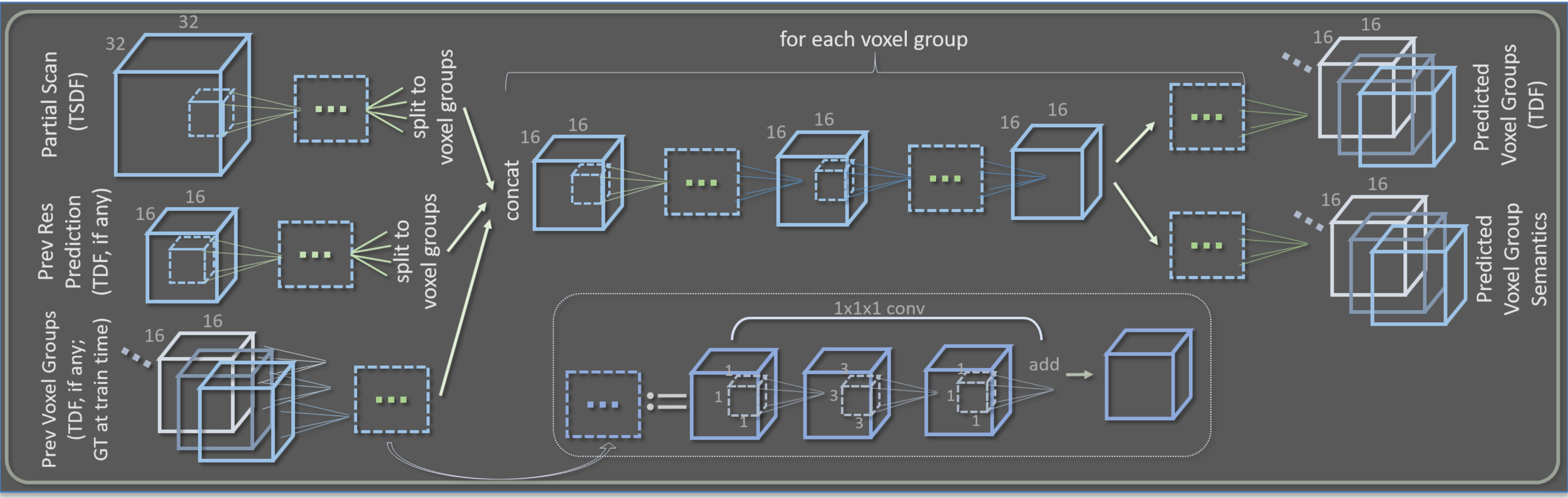
[Dai et al.]: ScanComplete

Jointly Prediction Completion and Semantics



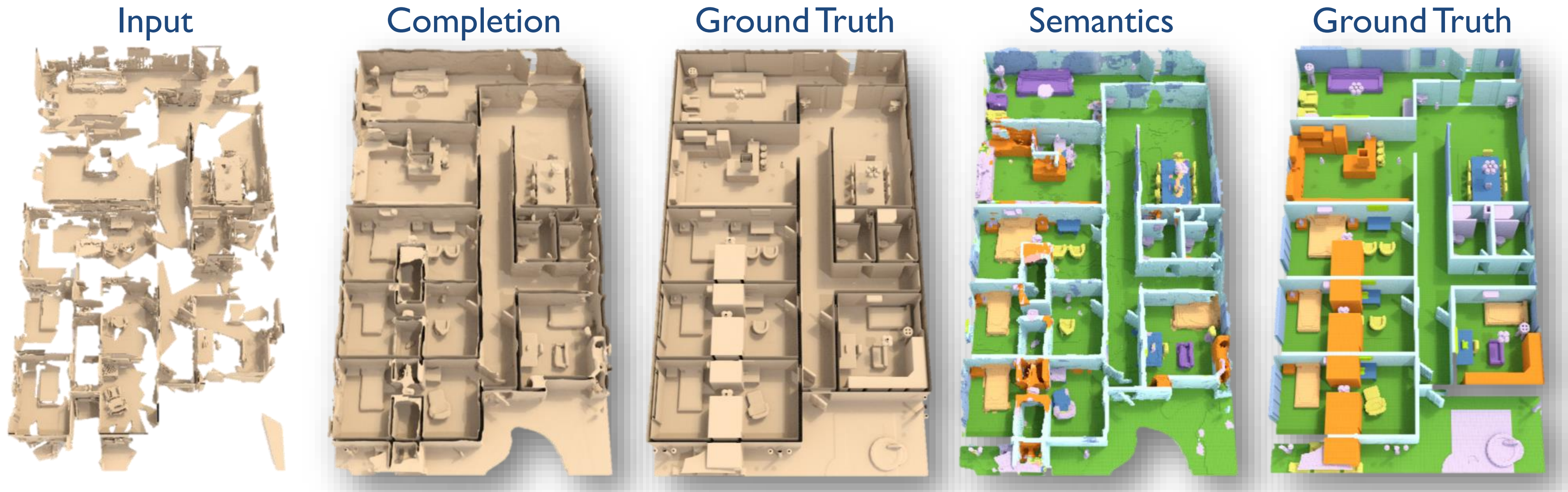
[Dai et al.]: ScanComplete

Model Architecture



[Dai et al.]: ScanComplete

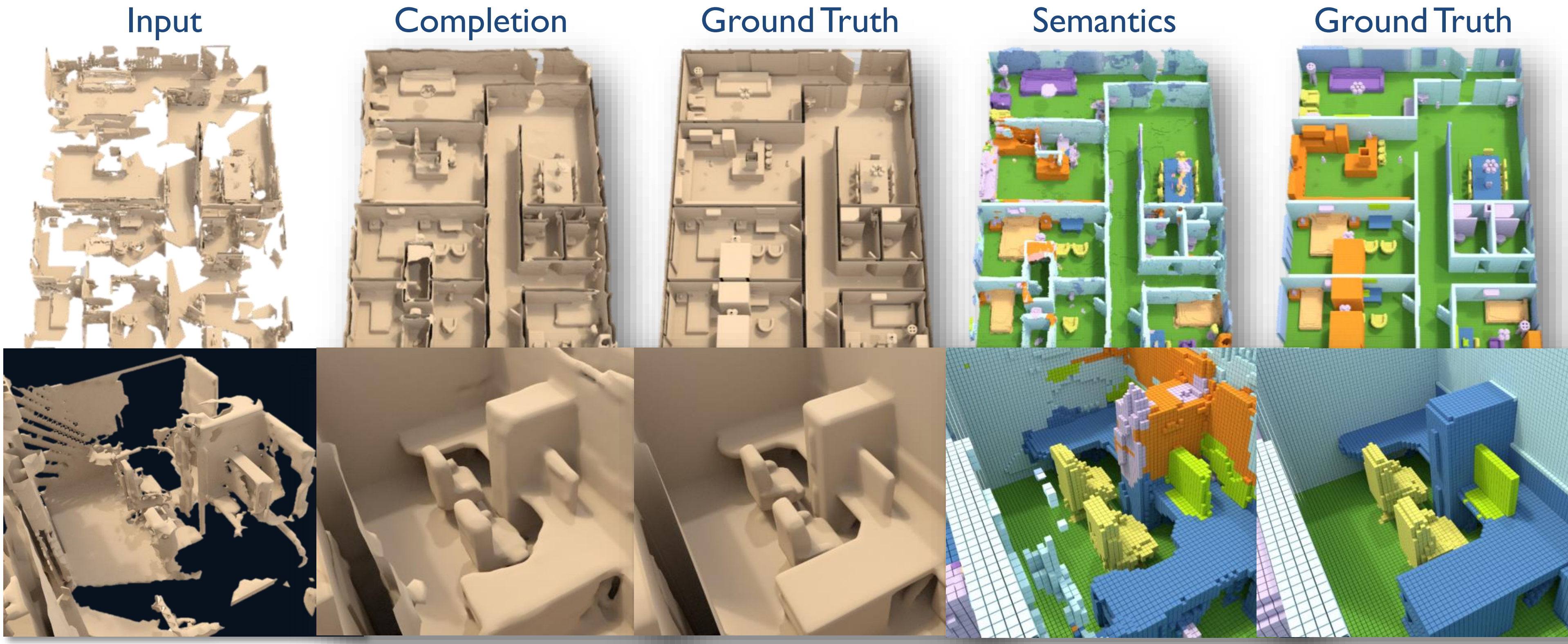
ScanComplete: Completing 3D Scans



Scans from SUNCG [Song et al. 2017]

[Dai et al.]: ScanComplete

ScanComplete: Completing 3D Scans



[Dai et al.]: ScanComplete

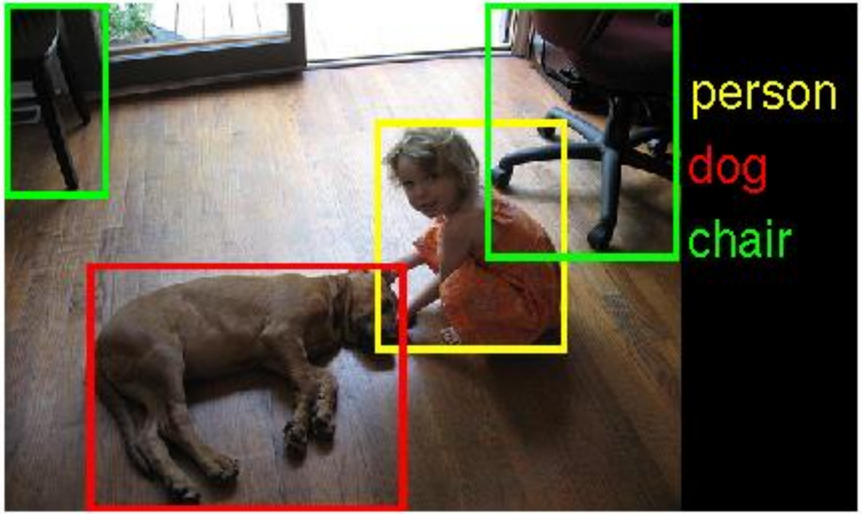
Multi-Dimensional CNNs

- 1D: e.g., WaveNet (audio)
- PointNet
- 2D Images
- 3D Scenes
- 3D Videos? <- we’ ll see more here next lecture (RNN/LSTM)
- 4D Lightfields
- Etc…

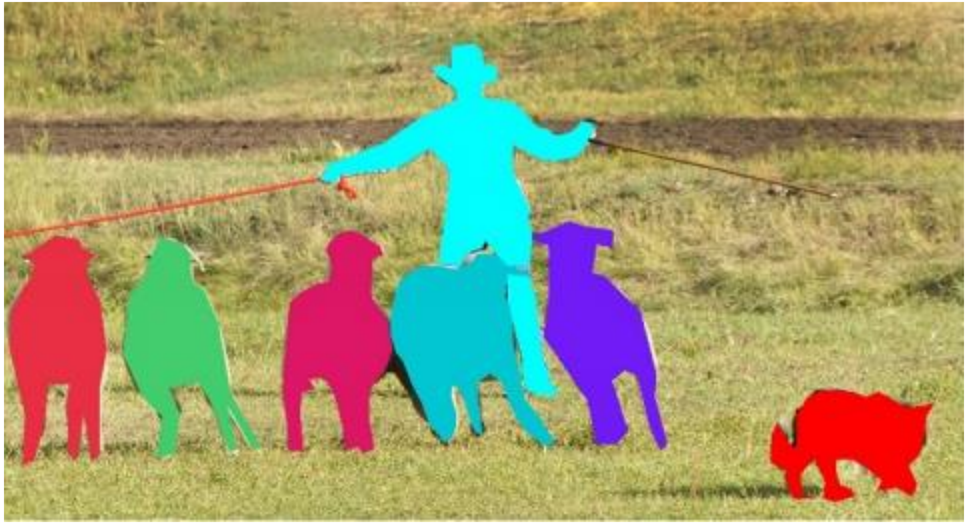
Training Data for CNNs

How do we get the training data?

- Labeled data from crowd sourcing
  - E.g., Amazon Mechanical Turk (AMT)
  - Or various other labeling platforms / services



ImageNet



MS COCO

How do we get the training data?

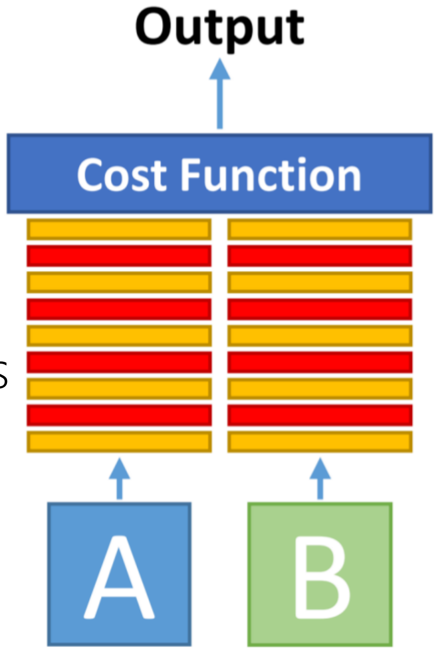
- What if we have limited data? (which is always the case)
  - Don’ t train from scratch
  - Use transfer learning when possible
  - Think about smart ways for data augmentation
  - Pre-train with auto-encoder if only small labeled dataset
  - Check training progress early on!

Self-Supervised Learning

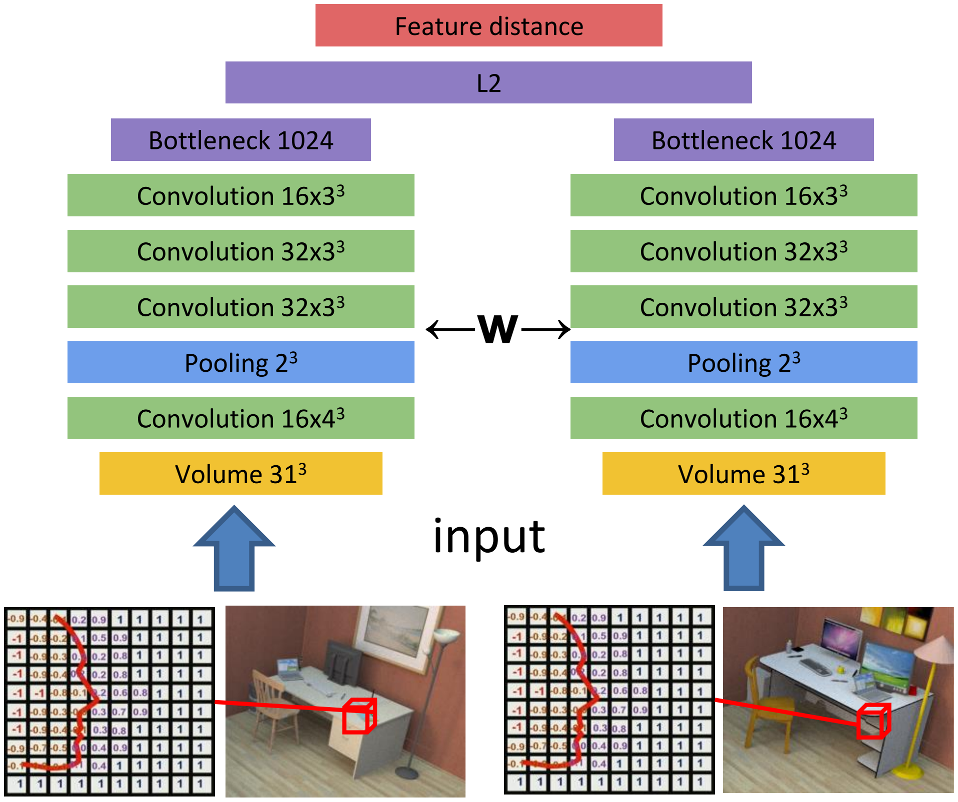
- Supervised vs Self-supervised
- Weakly-supervised vs Self-supervised
  - Good labeled data is *always* an issue

Side Note: Siamese Networks

Weight sharing between towers



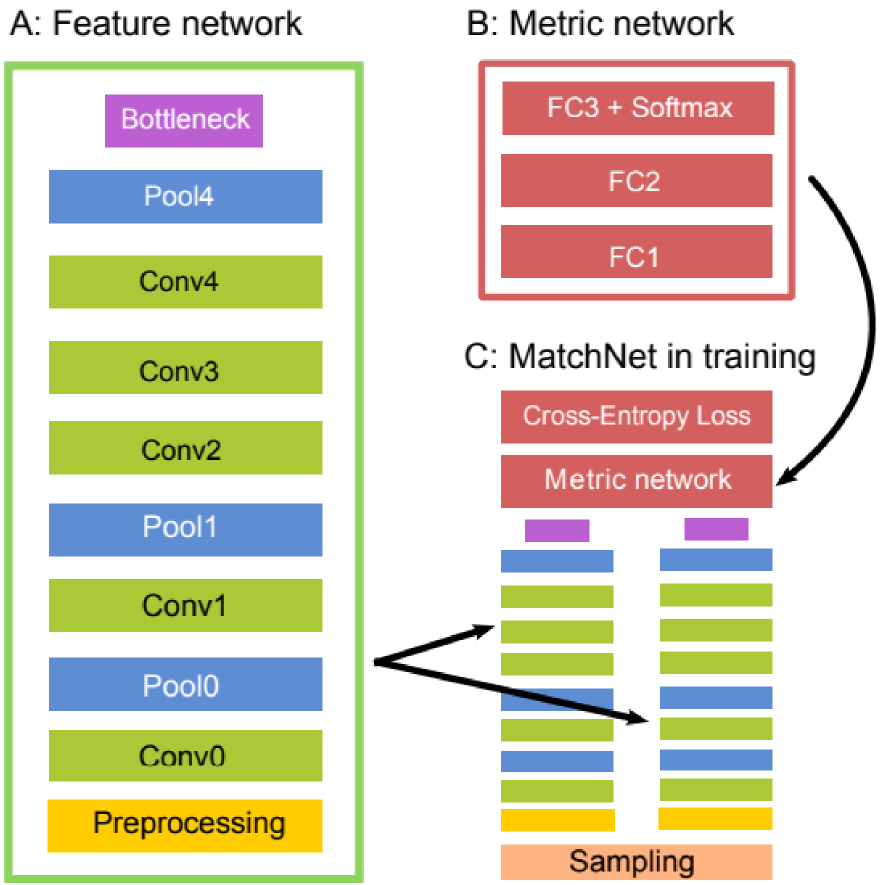
Siamese Architecture



3D Siamese Network

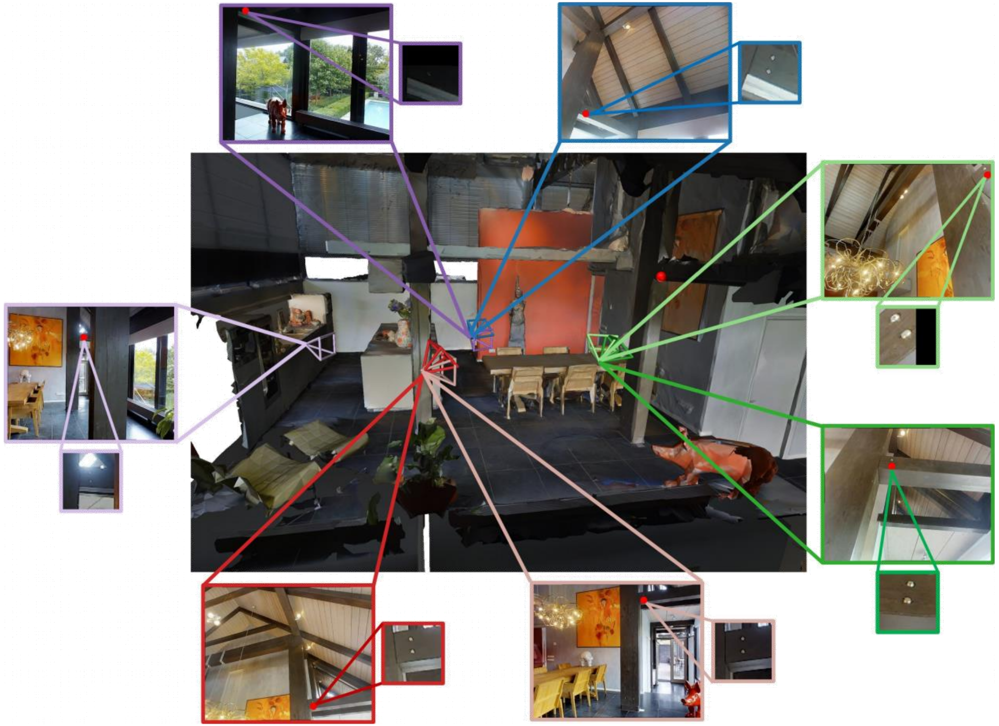
Side Note: Siamese Networks

- Many variations
  - Metric network vs L1/2 loss
  - Where to "merge"
  - How to pre-train
  - Etc.



MatchNet [Han 2015]

Self-Supervised Learning: Learning to Match Keypoints



SURF	46.8%
SIFT	37.8%
ResNet-50 w/ Matterport3D	10.6%
ResNet-50 w/ SUN3D	10.5%
ResNet-50 w/ Matterport3D + SUN3D	9.2%

Error (%) at 95% recall tested on SUN3D

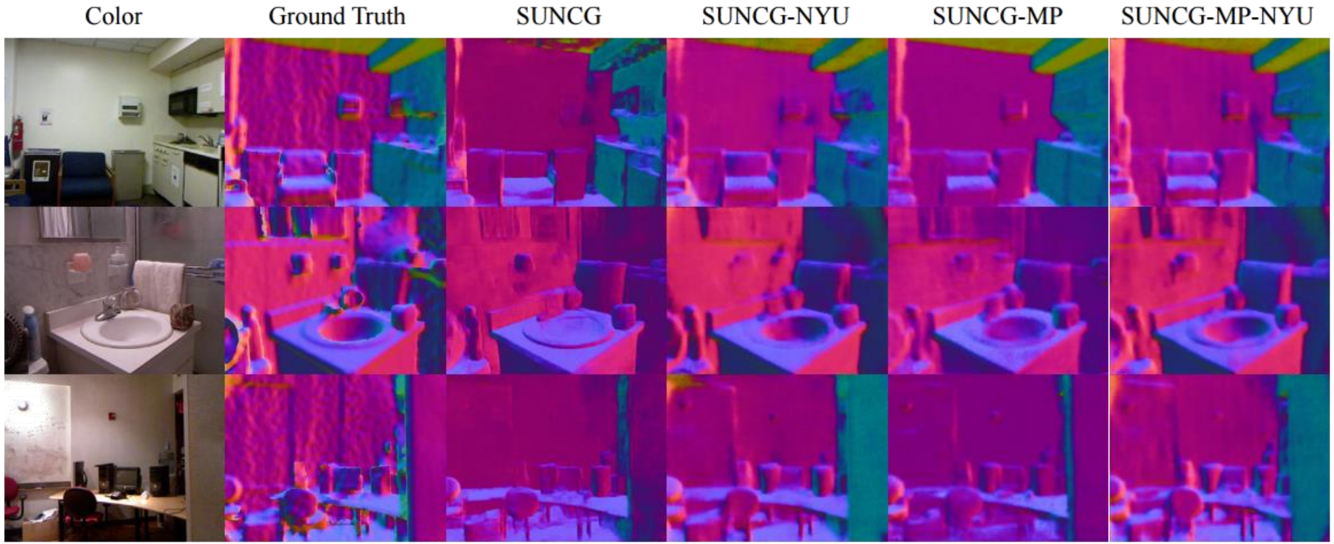
Self-Supervised Learning: Learning to Predict View Overlap

Example overlap views ranked by their overlap ration: Matterport3D provides a larger variety of camera points and wide baseline correspondences.

Training	Testing	triplet	triplet + regression
Matterport3D	SUN3D	74.41	81.97
SUN3D	SUN3D	79.91	83.34
Matterport3D + SUN3D	SUN3D	84.10	85.45
Matterport3D	Matterport3D	48.8	53.6

View overlap prediction measured by normalized discounted cumulative gain.

Self-Supervised Learning: Learning to Estimate Surface Normals



Train Set 1	Train Set 2	Train Set 3	Mean(°)↓	Median(°)↓	11.25(%)↑	22.5(%)↑	30(%)↑
SUNCG	-	-	28.18	21.75	26.45	51.34	62.92
SUNCG	NYUv2	-	22.07	14.79	39.61	65.63	75.25
MP	-	-	31.23	25.95	18.17	43.61	56.69
MP	NYUv2	-	24.34	16.94	35.09	60.72	71.13
SUNCG	MP	-	26.34	21.08	23.04	53.36	67.45
SUNCG	MP	NYUv2	**20.89**	**13.79**	**42.29**	**67.82**	**77.16**

Impact of training with Matterport3D (MP) on performance in the NYUv2 dataset

Self-Supervised Learning

- Feature matching
- Normal predictions
- Novel view prediction
- Camera pose between two images
- Depth map prediction / in-painting depth
- Optical flow / Scene flow
- Generate color for depth geometry
- …

*Always think if there are *free* training labels!!!*

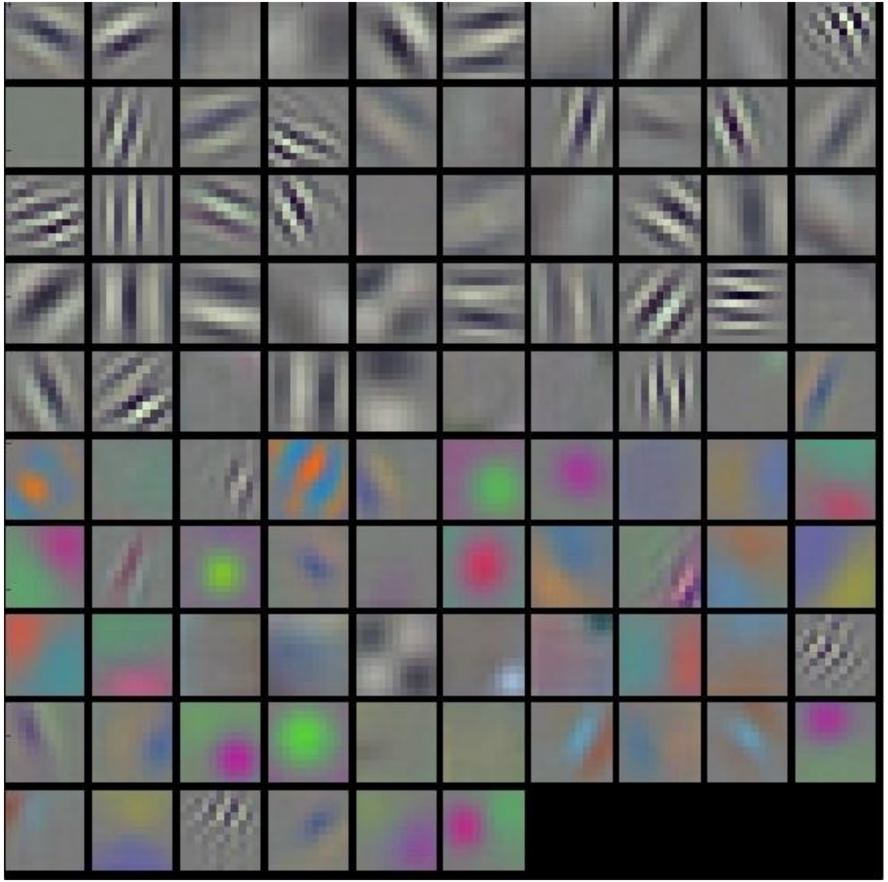
Visualization of ConvNets

Visualization of ConvNets

- Visualization of Features
- Visualization of Activations
- Visualization of Gradients
- T-SNE Visualization
- DeepDream
- ….

Visualization is a great way for debugging!

Visualization of Features



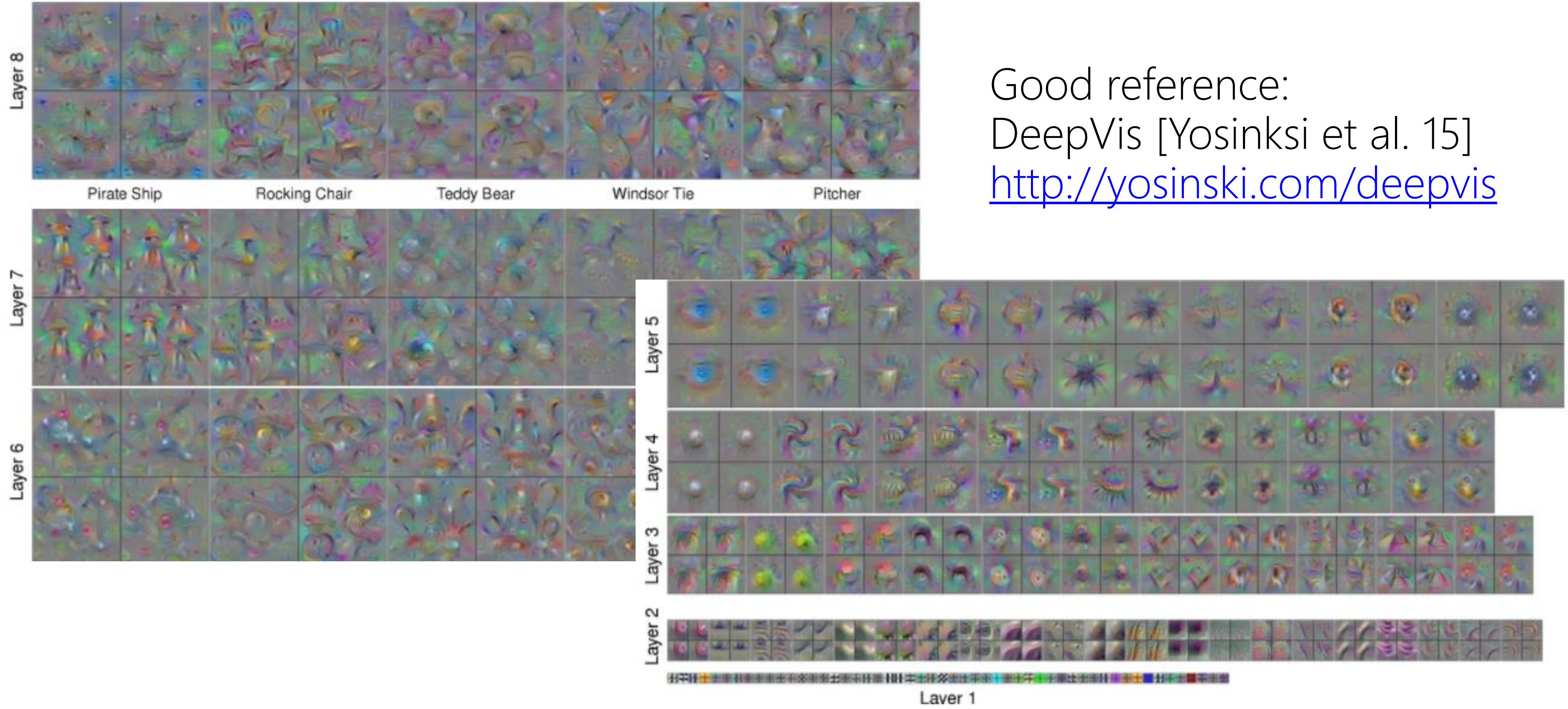
Visualization of AlexNet Features
first Conv Layer (weights visualized)

Color clusters are due to AlexNet streams

Other layers are not so easy to visualize typically need projection first

Credit: Li/Karpathy/Johnson

Visualization of Gradients



Good reference:
DeepVis [Yosinksi et al. 15]
http://yosinski.com/deepvis

Deep Visualization Toolbox

yosinski.com/deepvis

#deepvis


Jason Yosinski


Jeff Clune


Anh Nguyen


Thomas Fuchs


Hod Lipson

Cornell University

University of Wyoming

NASA Jet Propulsion Laboratory
California Institute of Technology

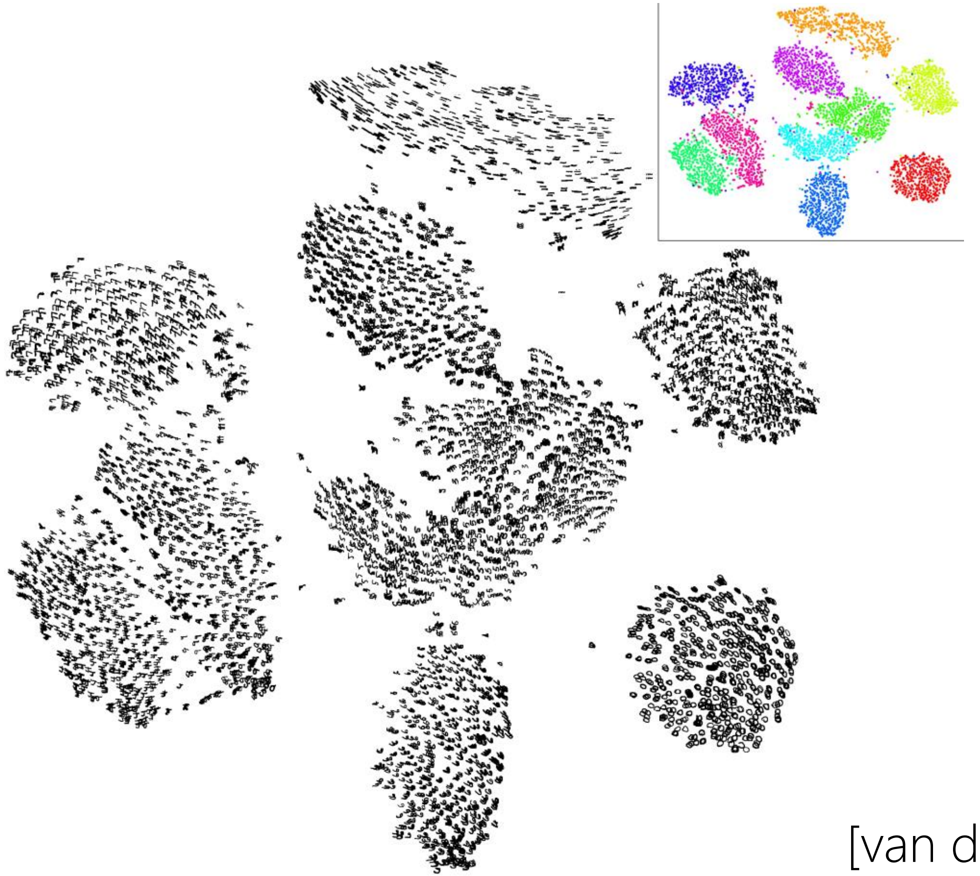
t-SNE Visualization

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Map high-dimensional embedding to 2D map
- Add samples from dataset according to their features to large image
- Very useful to spot clusters and debug embedding

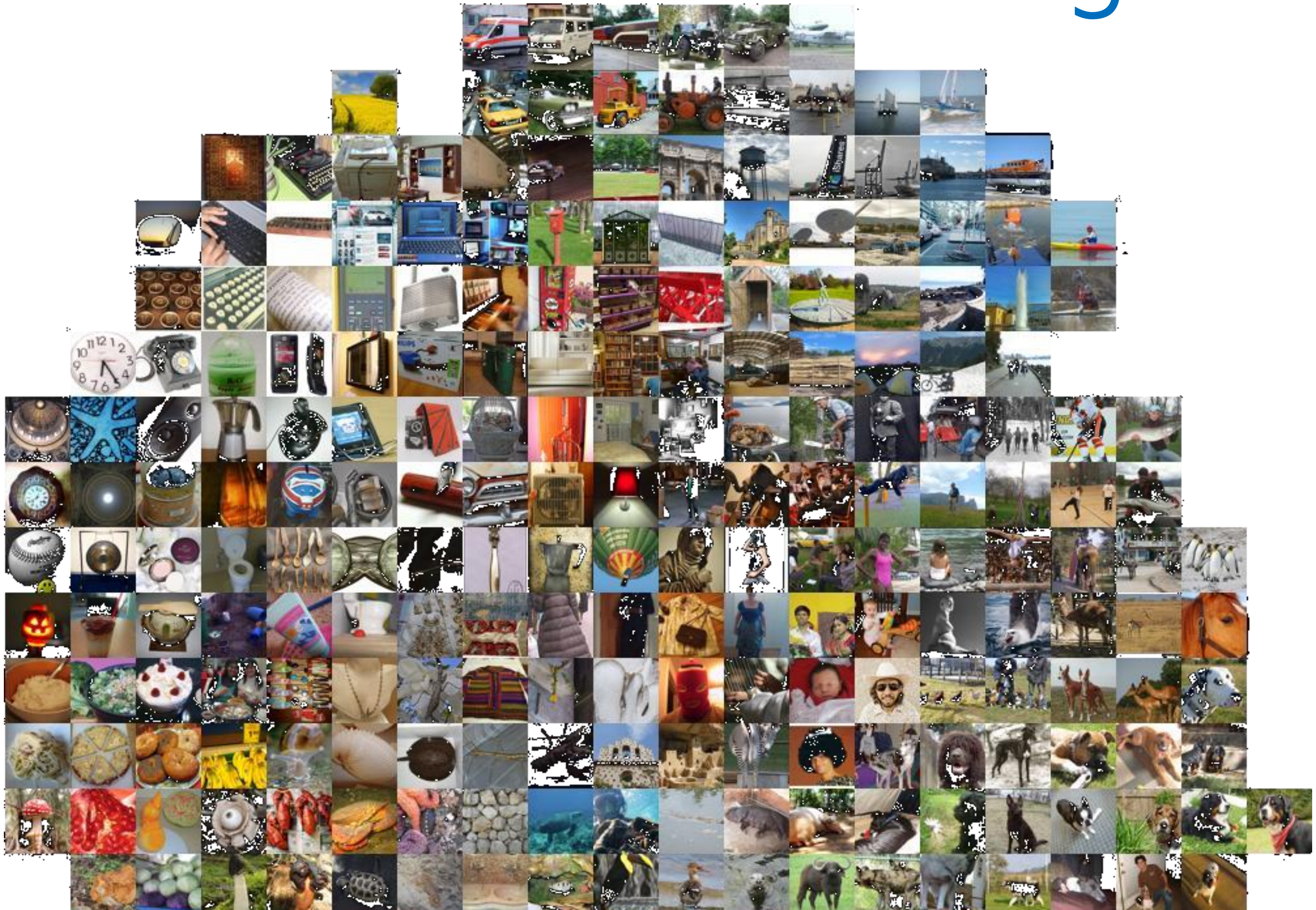
[van der Maaten et al.] t-SNE

t-SNE Visualization: MNIST



[van der Maaten et al.] t-SNE

t-SNE Visualization: ImageNet



Karpathy

t-SNE Visualization: ShapeNet

DeepDream

[Mordvintsev et al. 15] DeepDream

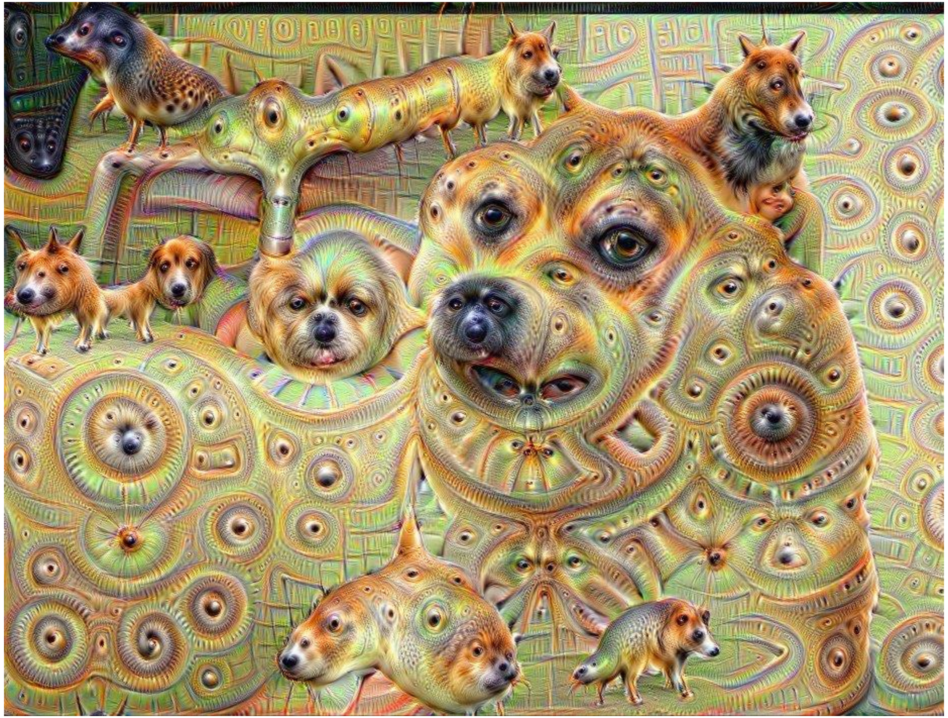
DeepDream

- Forward pass to the layer where you want to dream
- Make the gradients = raw activations
- Backprop to the image

- Amplifying the features that were activated at that layer
- Repeat

[Mordvintsev et al. 15] DeepDream

DeepDream

[Mordvintsev et al. 15] DeepDream

DeepDream

Administrative Things

- How are projects going?
- Next week:
  - RNNs and LSTMs
  - Hints to exam
- Jan 30th: Invited talk by Scott Reed (DeepMind)