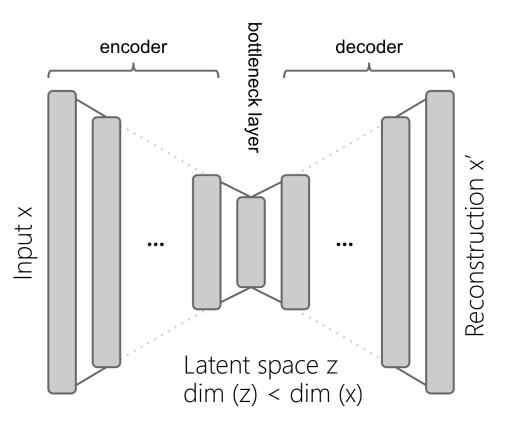


# Lecture 10 Recap

## **Training Autoencoders**

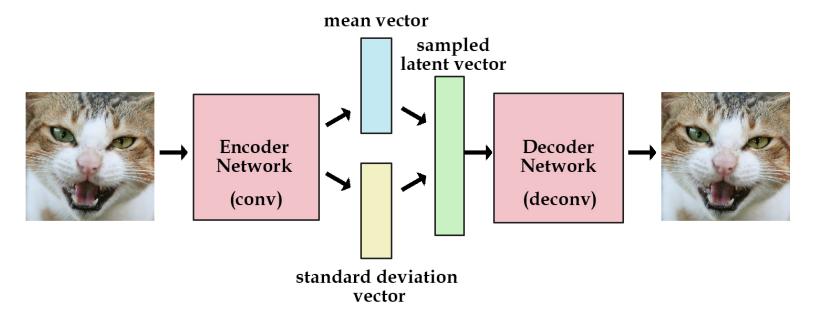




#### Reconstructed images



## Variational Autoencoders (VAE)

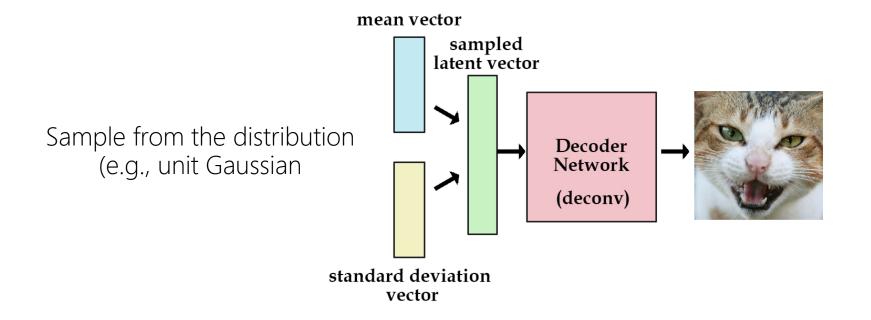


KL-Div Loss in latent space, forcing a unit Gaussian distribution -> now the latent vector becomes a distribution

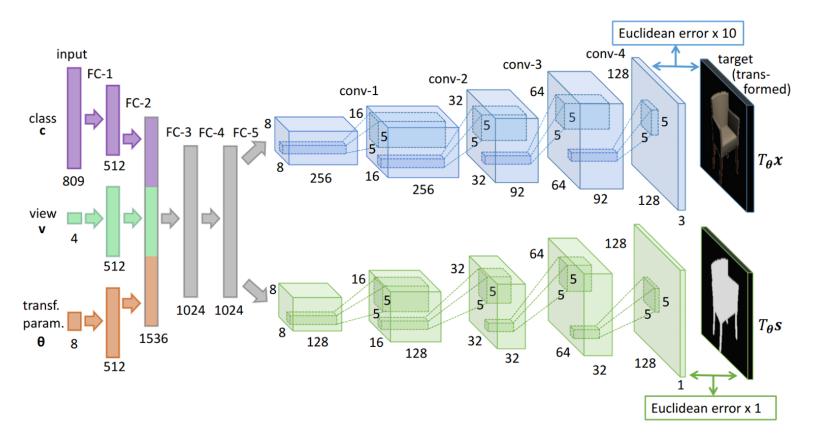
http://kvfrans.com/variational-autoencoders-explained/

## Variational Autoencoders (VAE)

• After training, generate random samples

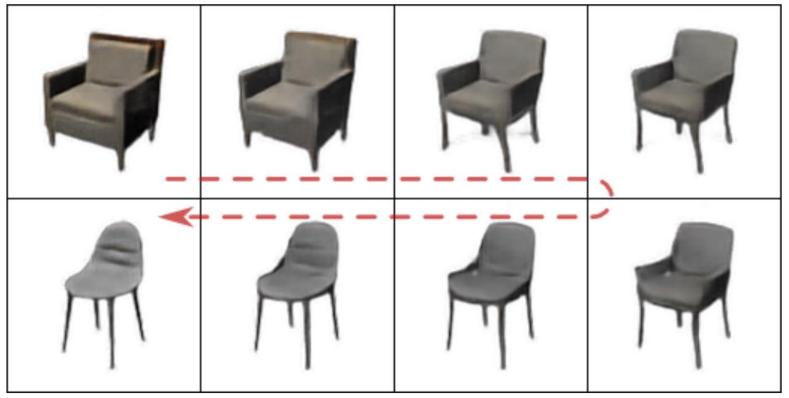


#### **Generative Models**



[Dosovitsky et al. 14] Learning to Generate Chairs

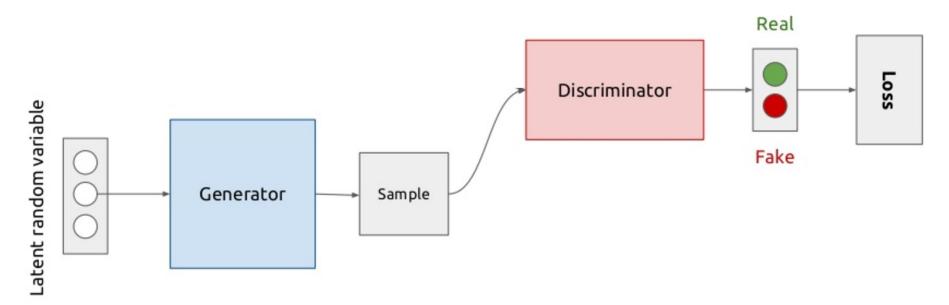
#### **Generative Models**



Interpolation between two chair models

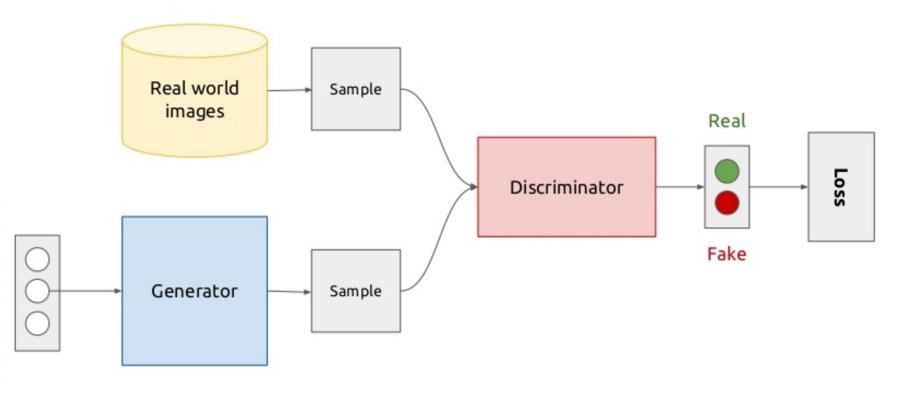
[Dosovitsky et al. 14] Learning to Generate Chairs

#### Generative Adversarial Networks (GANs)



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

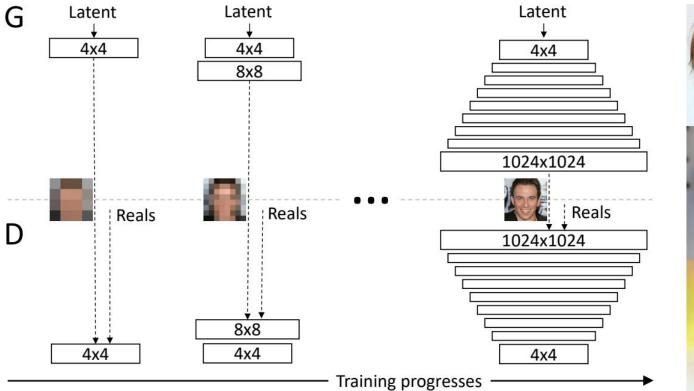
#### Generative Adversarial Networks (GANs)



-atent random variable

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

**Progressive GANs** 





https://github.com/tkarras/progressive\_growing\_of\_gans [Karras et al. 17]

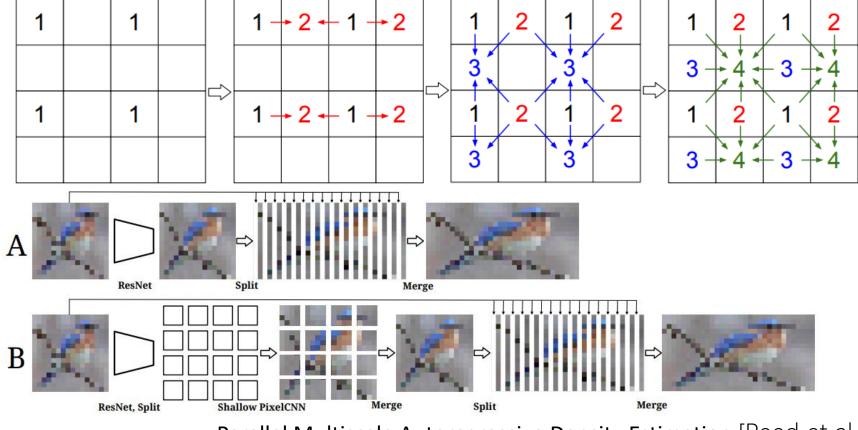
#### **Progressive GANs**

#### PROGRESSIVE GROWING OF GANS FOR IMPROVED QUALITY, STABILITY, AND VARIATION

Submitted to ICLR 2018

https://github.com/tkarras/progressive\_growing\_of\_gans [Karras et al. 17]

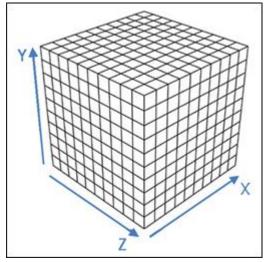
#### Other Generative Models: Autoregressive Models



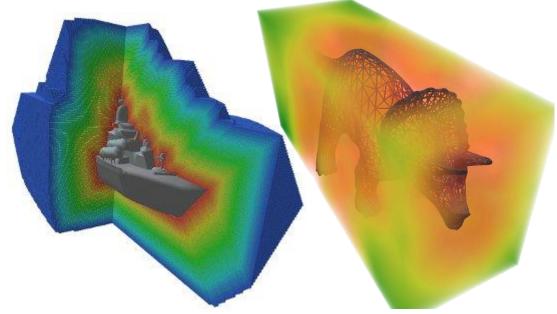
**Parallel Multiscale Autoregressive Density Estimation** [Reed et al. 17]

## **3D** Convolutions

#### On volumetric data structures

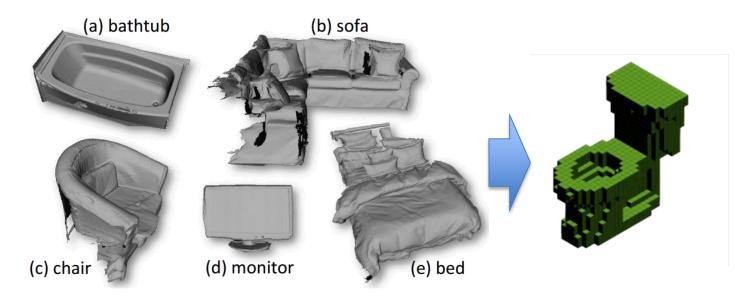


(binary) Voxel Grid



Implicit functions: e.g., signed distance field

### **3D** Classification

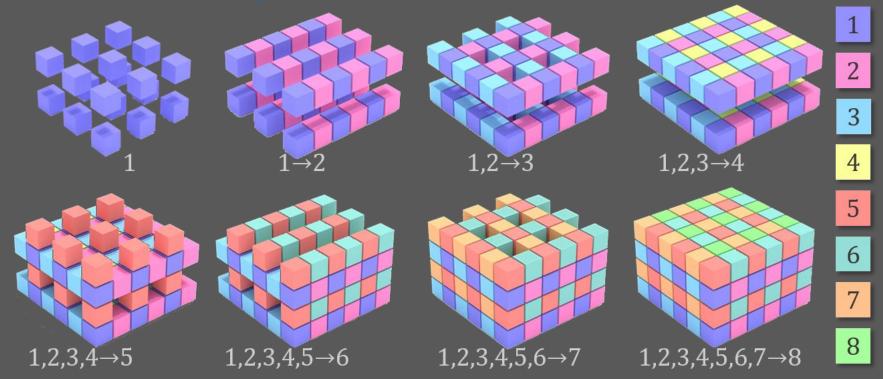


Instance: 010.toilet\_000000079.001 Predicted label: toilet True label: toilet

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

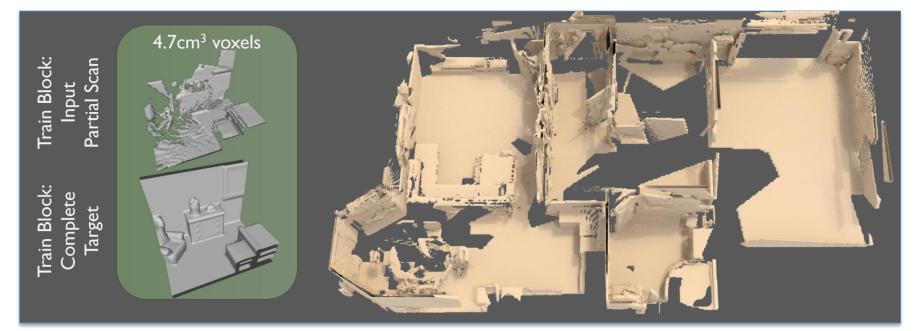


[Dai et al.]: ScanComplete

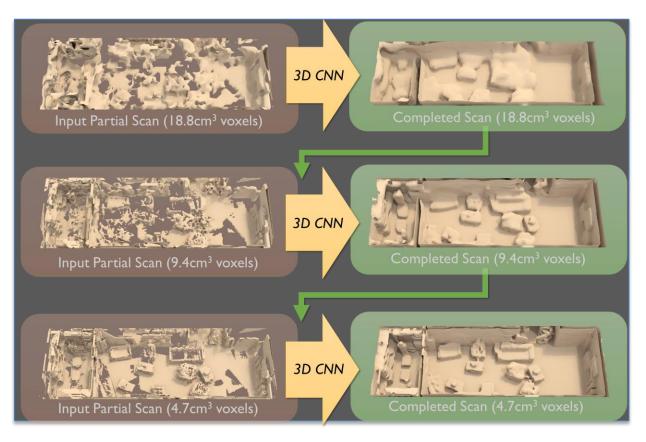
#### Spatial Extent: Fully Convolutional for Arbitrary Sizes

#### Train on crops of scenes

#### Test on entire scenes



#### Spatial Extent: Coarse-to-Fine Predictions



[Dai et al.]: ScanComplete

#### ScanComplete: Completing 3D Scans

Input



Scans from SUNCG [Song et al. 2017]

Completion



**Ground Truth** 



#### ScanComplete: Completing 3D Scans Completion

Input









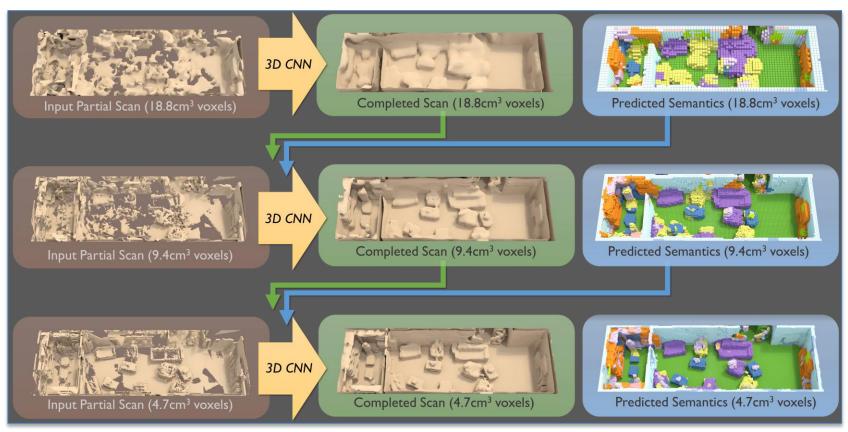
**Ground Truth** 





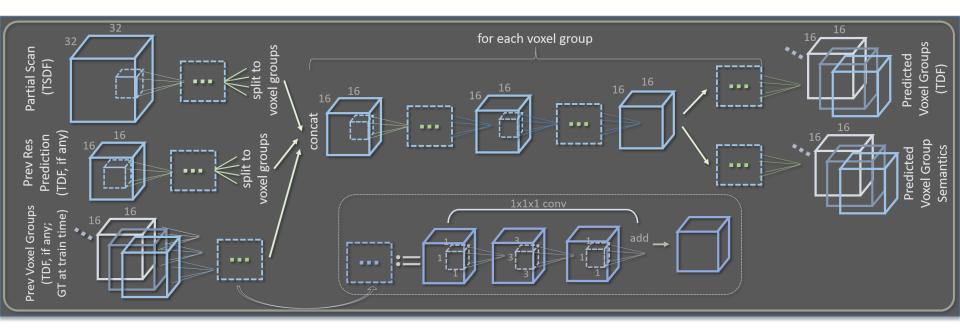
[Dai et al.]: ScanComplete

#### Jointly Prediction Completion and Semantics



#### [Dai et al.]: ScanComplete

#### Model Architecture

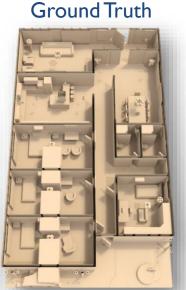


#### ScanComplete: Completing 3D Scans



Scans from SUNCG [Song et al. 2017]





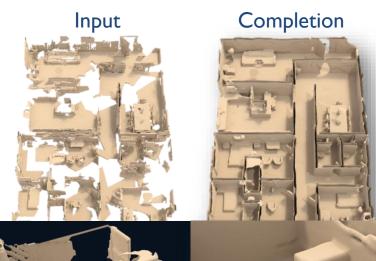


**Semantics** 

**Ground Truth** 



#### ScanComplete: Completing 3D Scans







**Ground Truth** 



[Dai et al.]: ScanComplete

## **Multi-Dimensional CNNs**

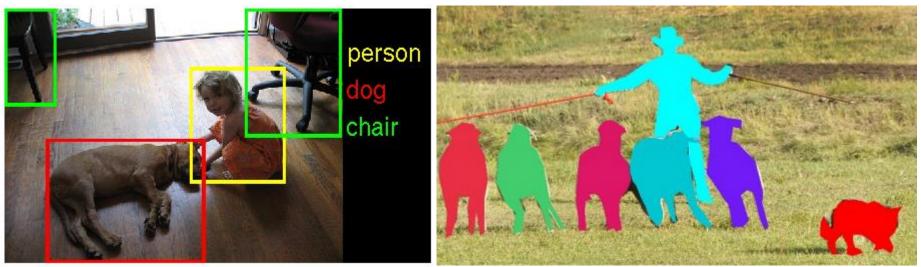
- 1D: e.g., WaveNet (audio)
- PointNet
- 2D Images
- 3D Scenes
- 3D Videos? <- we' II see more here next lecture (RNN/LSTM)
- 4D Lightfields
- Etc...



# 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

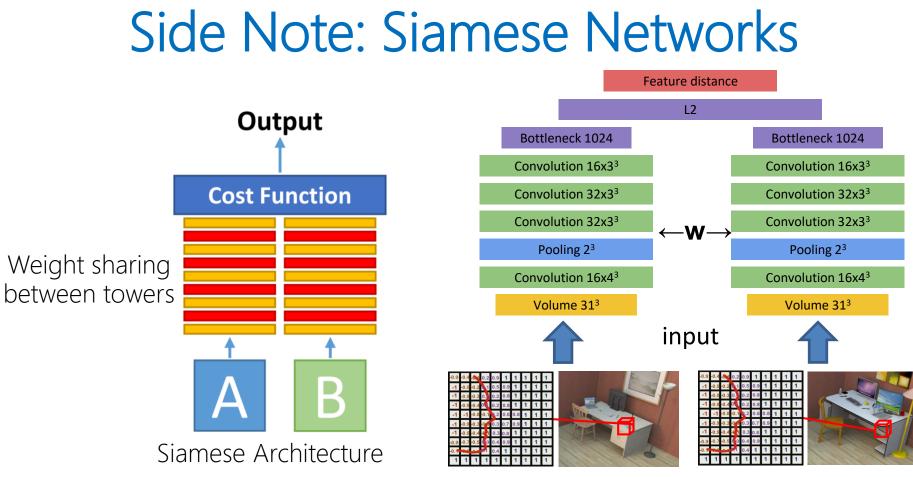
## How do we get the training data?

- 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

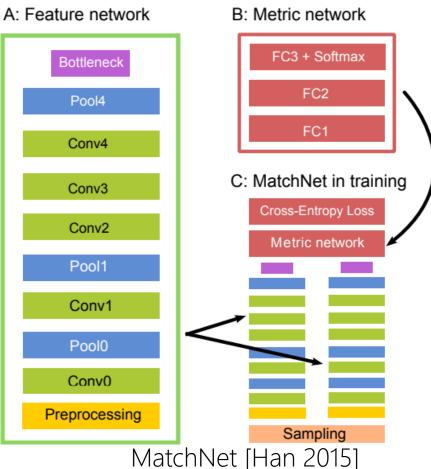
- Weakly-supervised vs Self-supervised
  - Good labeled data is \*always\* an issue



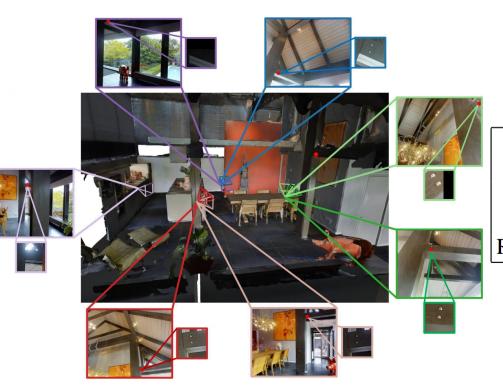
3D Siamese Network

### Side Note: Siamese Networks

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



#### 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

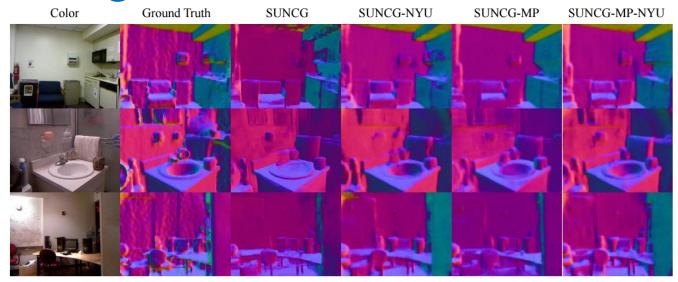


Example overlap views ranked by their overlap ration: Matterport3D provides a larger variety of camera points and wide baseline correspondences.

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!!!



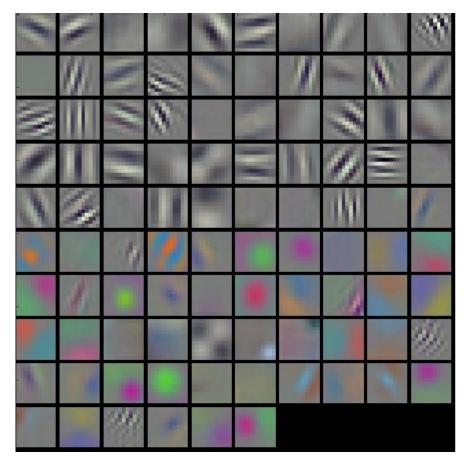
# Visualization of ConvNets

### 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**



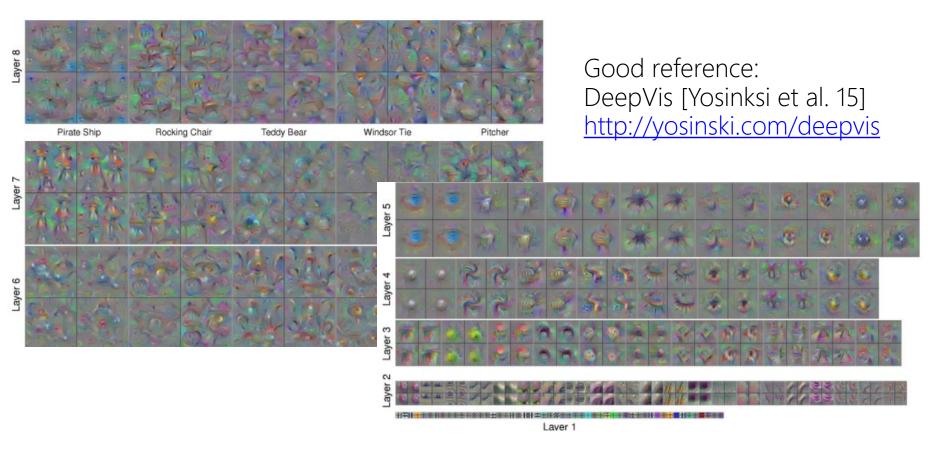
Visualization of AlexNet Features first Conv Layer (weights visualized)

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



#### **Deep Visualization Toolbox**

#### yosinski.com/deepvis

#### #deepvis



Jason Yosinski



Jeff Clune



Anh Nguyen



Thomas Fuchs



Hod Lipson







Jet Propulsion Laboratory California Institute of Technology

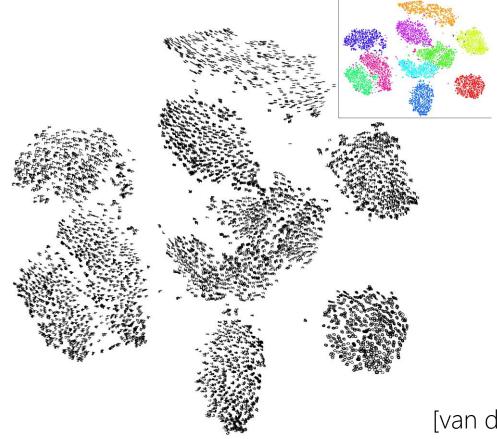
### t-SNE Visualization

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

福佳县

NE TERFET AGE O AD AS AS AS

44.1111

2-----

69.19

18

1/ 1/ 1/ 1/

191819

REENCE VIEWER IT VIEW

STATISTICS FALLET ATATAT

HE HE H H. HUMMANN

周期 開業

16.3

her

14 15

1

31.68 48

1 MAL

MALL (MF

15

LA PATE STAT AL

17 15 18 . 11 19 17 13

1184

Plane Share TALANDER SHARE

DEBRINATION VIRGAL AL DE

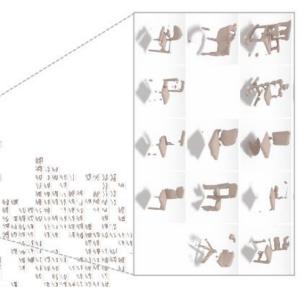
MOIDIDI VOIN, 846141

41 41

AT RIVER PI HAVE N

FILLIN VIEW H

le (ri





1 三郎上も that for NO REAL AND ADADADE OF THE SECTION deTa.d 60 60 3 B 8.0 19/23 THERE AN A REAL PROPERTY. 10044 14/48 6, 21.50 646.6 12 10 10 10 10100-9月5年4月18月4月1日7日 1000 10,00 54 SOUD, M a' at PONPOPER THE THE MAKE 1.00 0.79 6 62 8 67 002 10 - 10 12 647091093 en Pinte Presidente 198 ANY 448 a unio unicali. 120 100 7914197 35 CHINGS CONTROLS FT'90-17 100 RAD NA SH 80.80 THEFT THEFT 煙 and the address of the second 10 10 1.000 0H 050899 (0) (100,0.77) (100,0.07) (0.17)10-01100 ATTACK CONSTANCE AND A CAMPAN

3600 4.0 把目標門門 LAF-H-PHENFIN-THE-INFTHE-INFTHE-INFTHE-IN-THE-INFTH 200 (A 747 100 400 Bigto Perris erelagt 司信 11.40

15 网络柳貂貂 18 65 / P 儿生 WENTER MILL 3 Bee 4.5 1000-0 B/ P/ B/ /mg かべらる おおとうかうかぶかりかうち HER JENE 人相 作業委員会から かかくちょうなから パーク・ハー North And And And And And And And 61 - A1 14 17899114 68 11-44 NULLEN P. s-no no solar - 200 10000 1 12 - 1 FL AND H VICENT 

17. 71 11 11 14 11 11 M HEIDERING WOMACHARM AND 植开开村田 10 - Jun 10-18.15 mer Har --- B (171 B.) 1-1-12 14 MUNICAL REPORT A MAIN AND A MAINING A 11 12 12 11 18 181 14

18 11.11 新·特·特 80 M-14 24 11 111 11.03 15 23



21.00 11 24 10 12/10/14 18-342.15 19 1811 8018-840 11 11 11 11 12 14

1 12 13.14 14:15 18 11/11/17 11.11 11 15.11 1111101111 18 13 11 64 14 12.02. 19 12 12 13 11 13 12 12 12 14 18 11 83.00 14 18

12 14 10 10 12 12 12 18 15 11.13 11 1 13 18 15 11.11 19.11 33 14.14.15 7E 71 1311 11 12 12.14 11 181218 11 12 13 理 17 12.18 12.18 10.15 13 13 11:00 11 11/18 11 F.M. 12 1118 1011 16.18 12 16 11.11 161211-12 13 12 12 11 14 11 18 11 11 15 14 10 18 14: 15

12 11 1811.11 15 11 10 10 10 10

## DeepDream



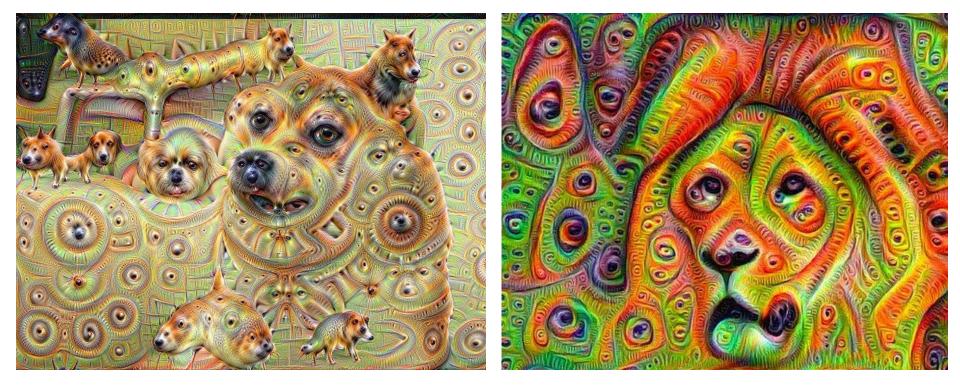
#### [Mordvintsev et al. 15] DeepDream

## 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





## Administrative Things

- How are projects going?
- Next week:
  - RNNs and LSTMs
  - Hints to exam

• Jan 30<sup>th</sup>: Invited talk by Scott Reed (DeepMind)