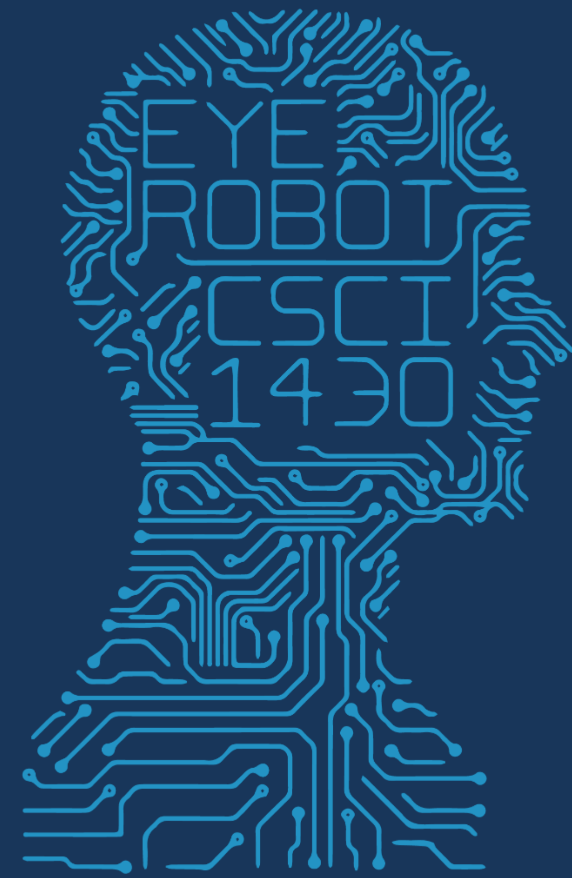
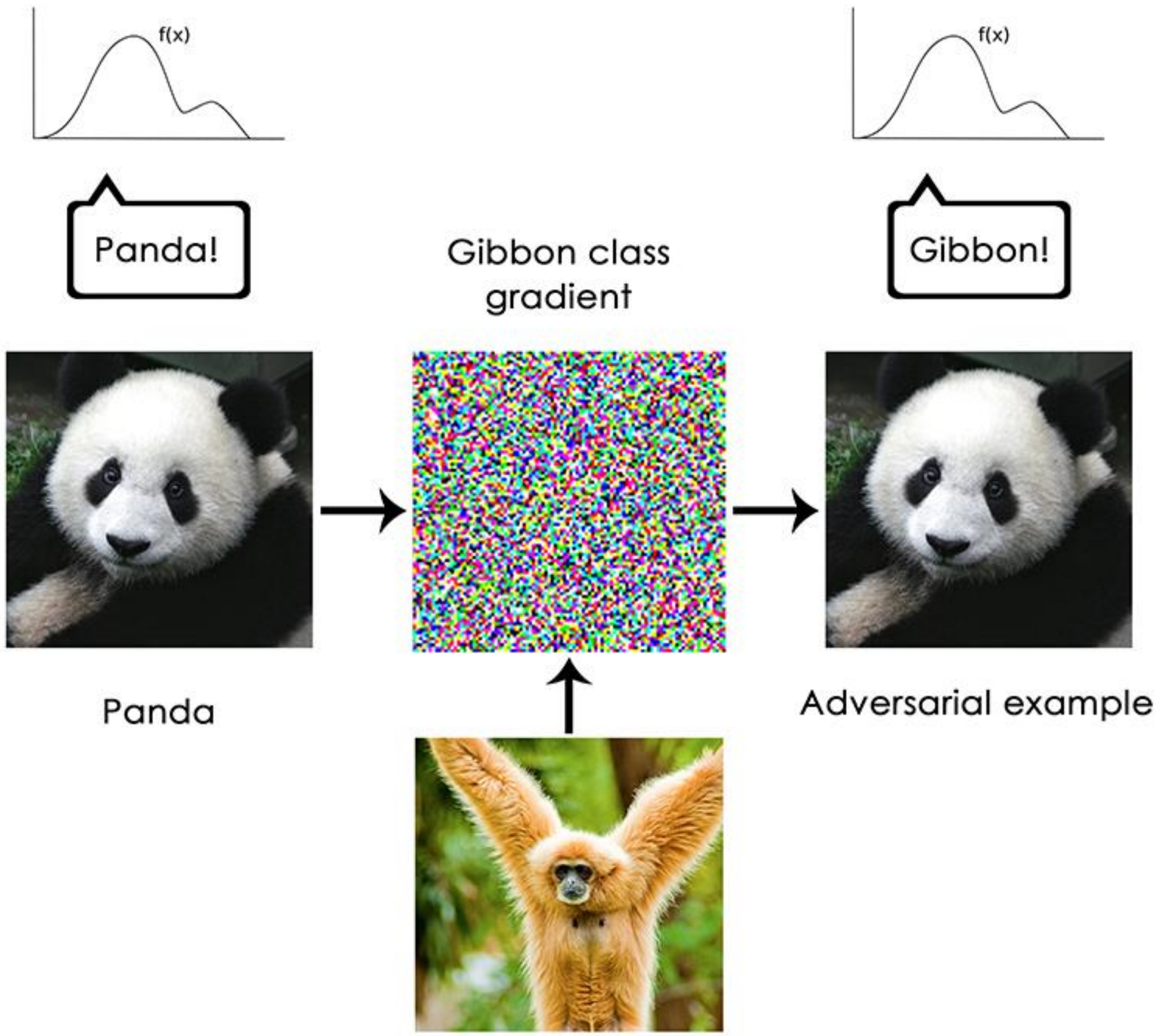
1950

Future Vision



2017 MWF 1pm

Computer Vision



Francois Chollet - https://blog.keras.io/the-limitations-of-deep-learning.html

3rd November 2017

BBC | Your account | News | Sport | Weather | Shop | Earth | Travel | More | Search

NEWS

Home | Video | World | US & Canada | UK | Business | Tech | Science | Stories | Entertainment & Arts | Health | More

Technology

Single pixel change fools AI programs

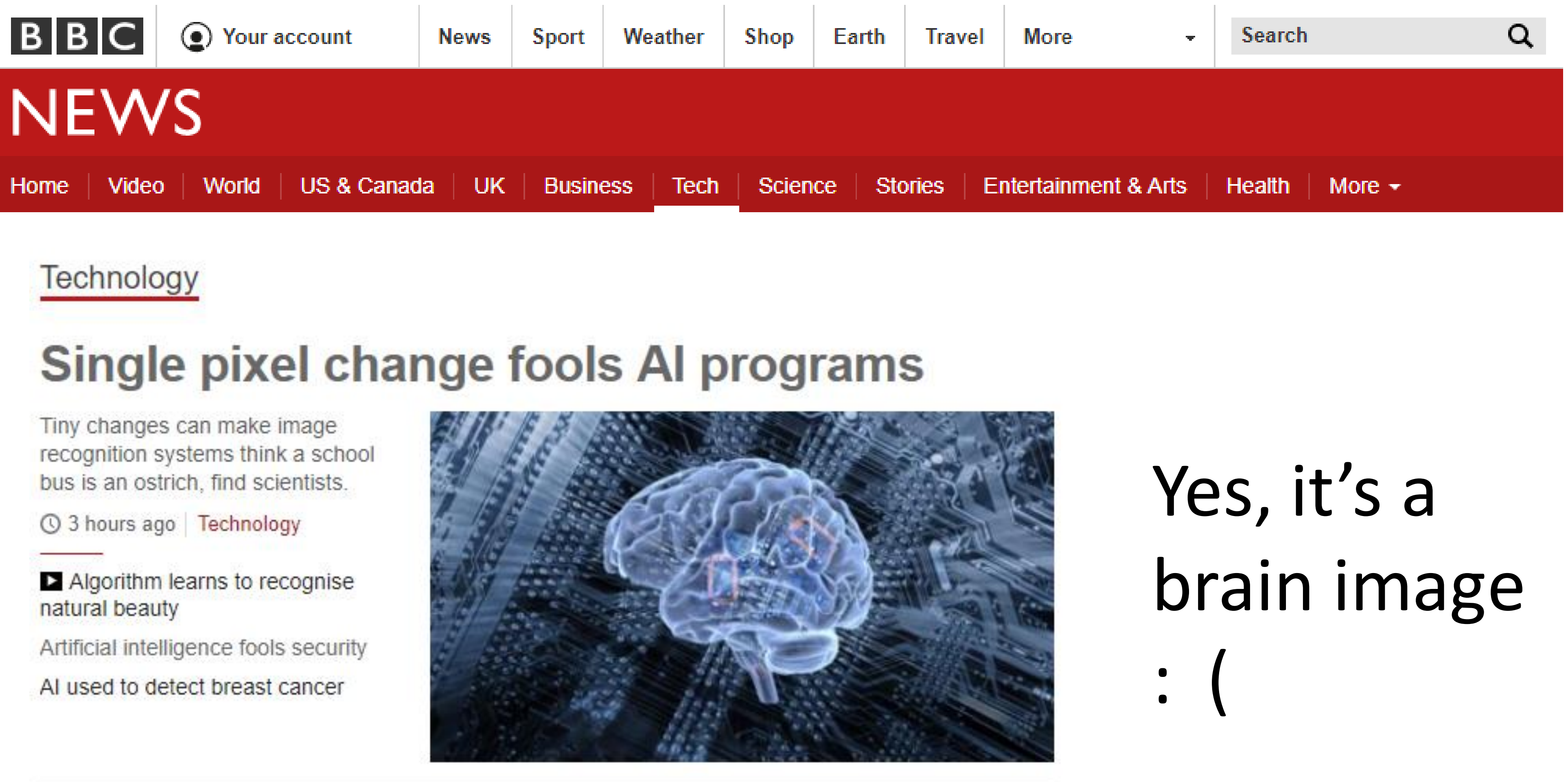
Tiny changes can make image recognition systems think a school bus is an ostrich, find scientists.

3 hours ago | Technology

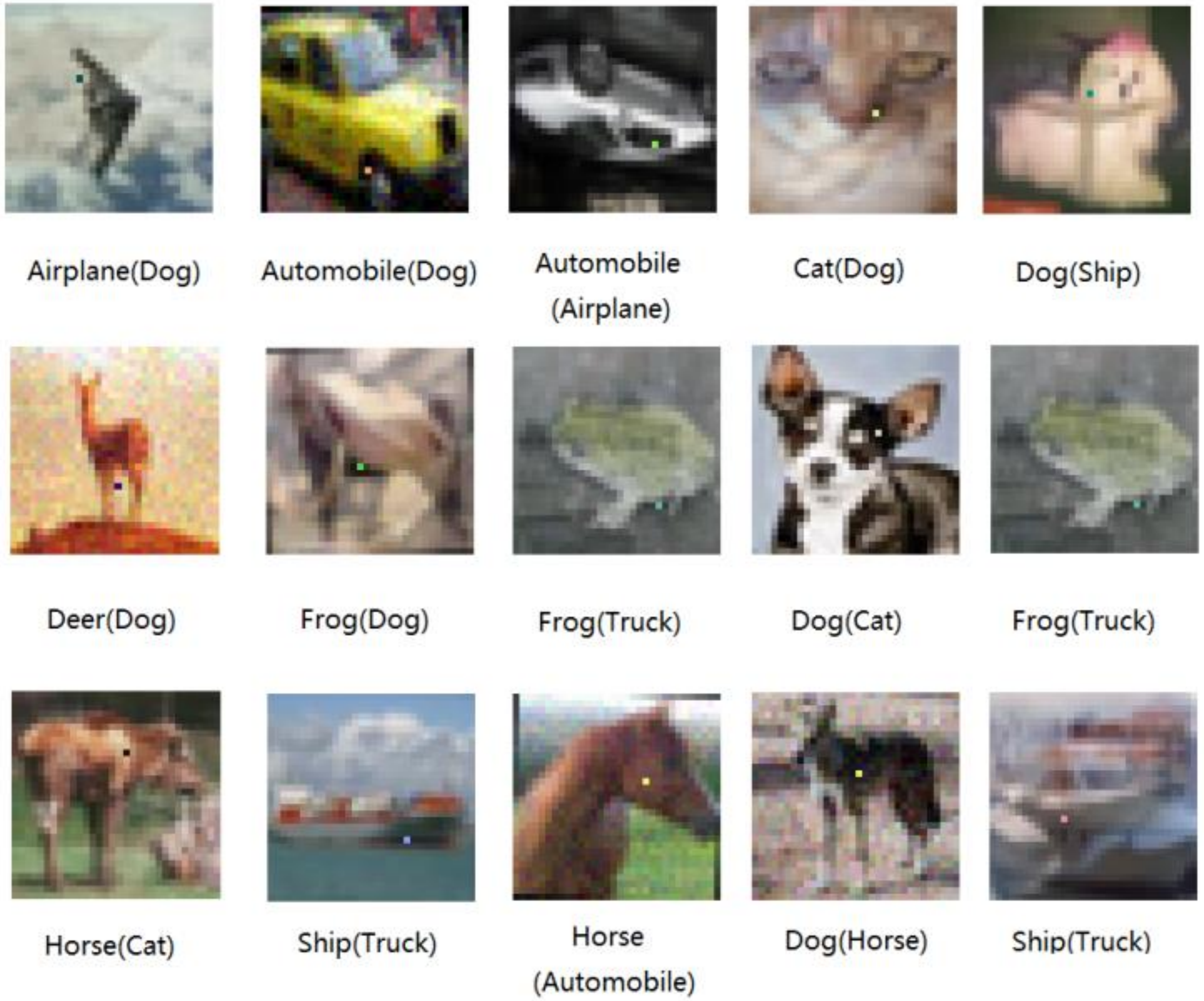
Algorithm learns to recognise natural beauty

Artificial intelligence fools security

AI used to detect breast cancer



Yes, it’s a brain image : (

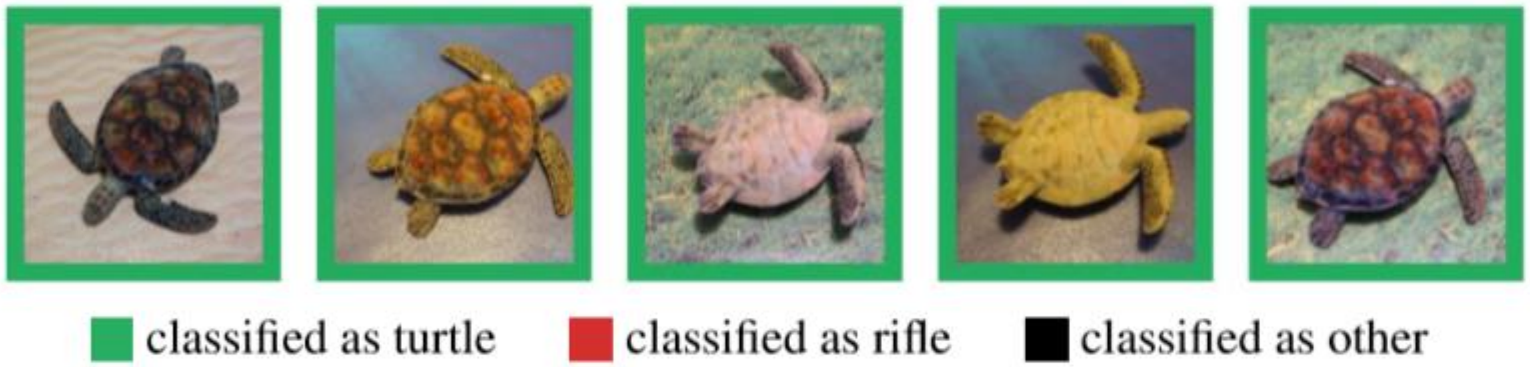


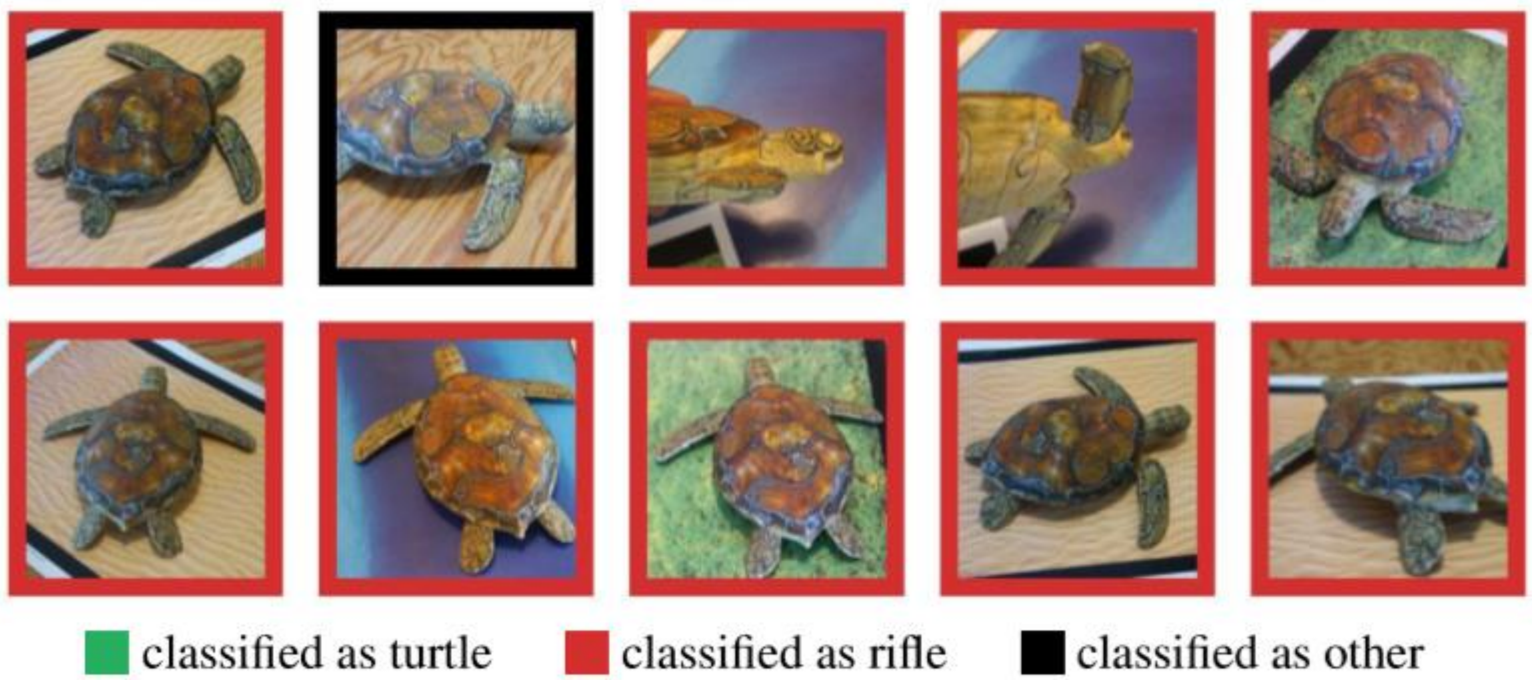
Su et al., One pixel attack for fooling deep neural networks
https://arxiv.org/abs/1710.08864

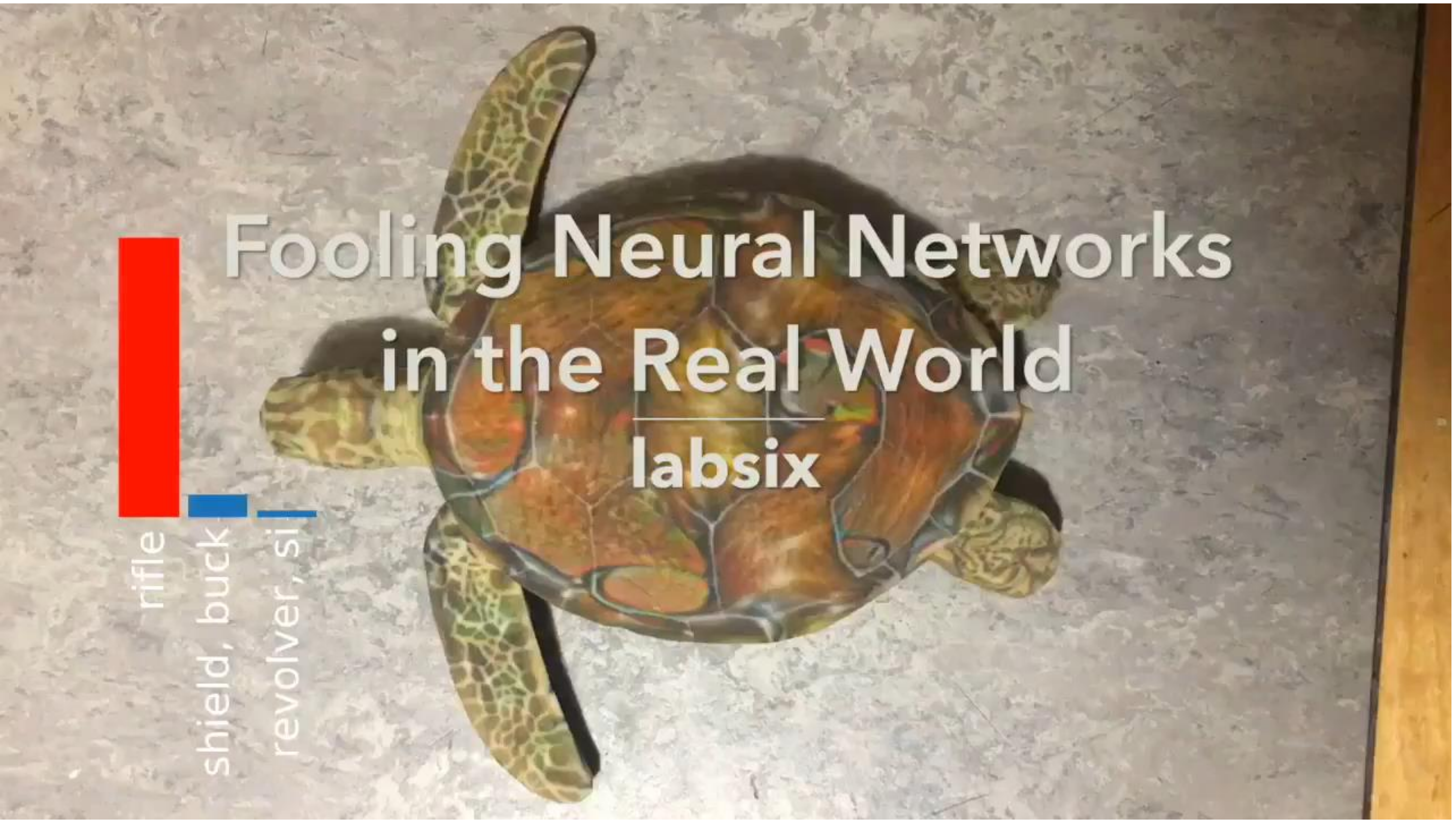
Synthesizing Robust Adversarial Examples

https://arxiv.org/pdf/1707.07397.pdf

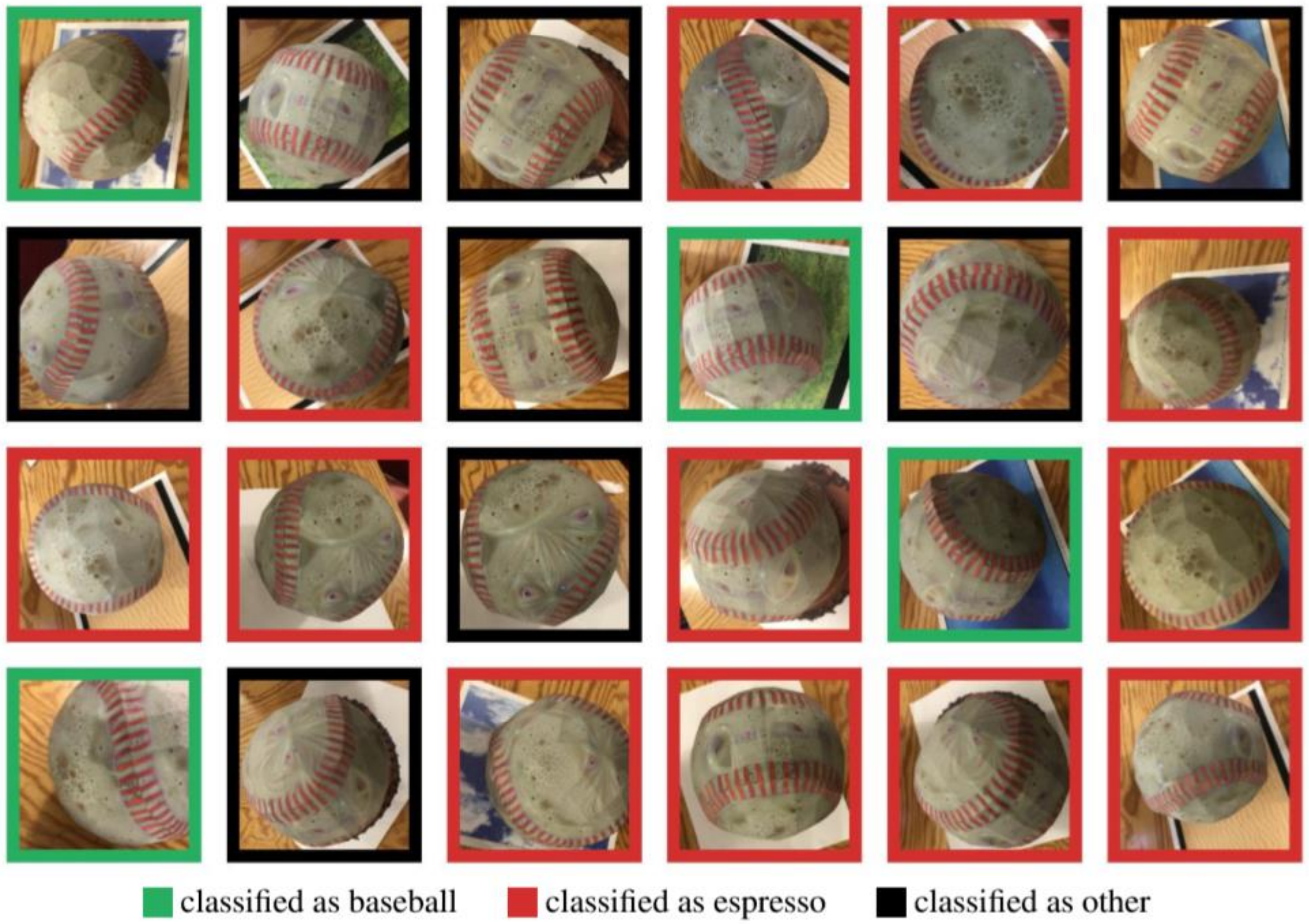
Anish Athalye[*1,2], **Logan Engstrom**[*1,2], **Andrew Ilyas**[*1,2], **Kevin Kwok**[2]
[1]Massachusetts Institute of Technology, [2]LabSix
{aathalye,engstrom,ailyas}@mit.edu, kevin@labsix.org

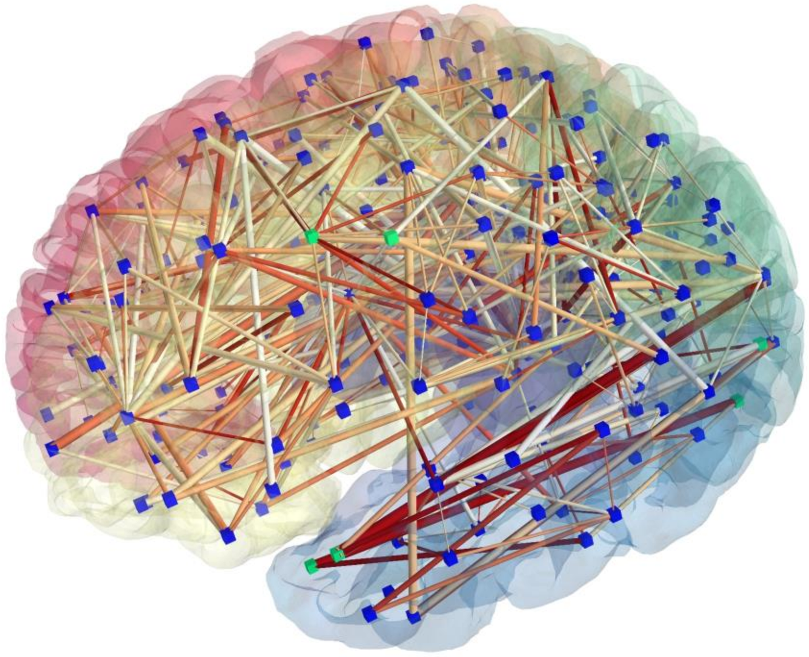





Fooling Neural Networks
in the Real World
labsix
rifle
shield, buck
revolver, si


classified as baseball
classified as espresso
classified as other

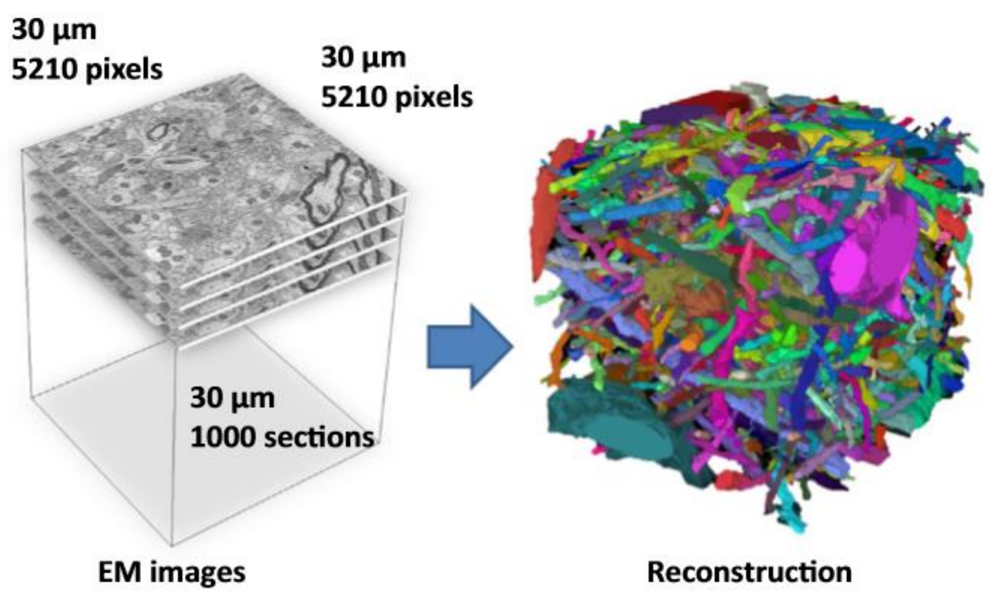

classified as baseball
classified as espresso
classified as other



[Patric Hagmann]

Connectomics: Neural nets for neural nets

Vision for understanding the brain



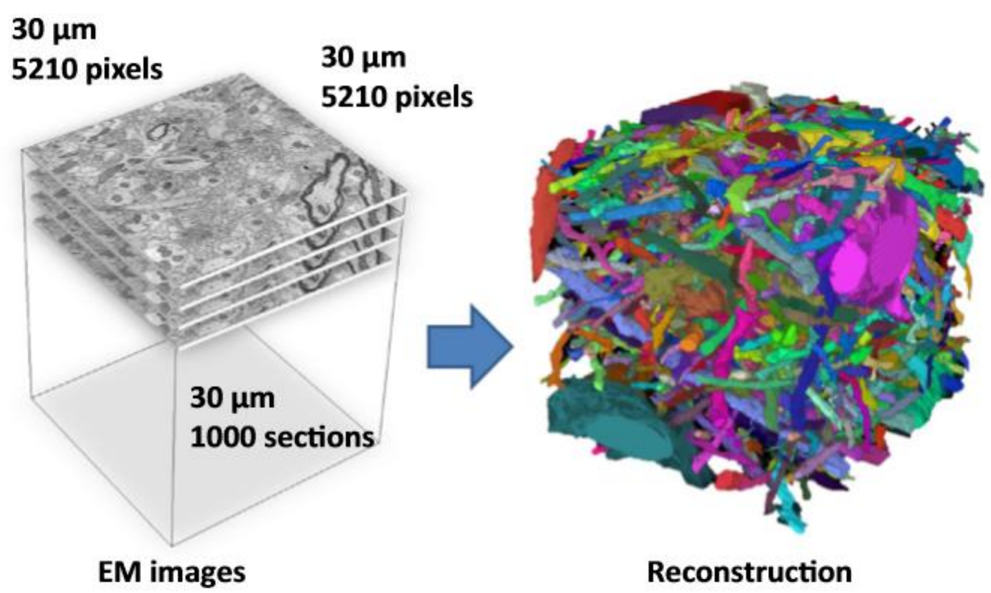
1mm cubed of brain

Image at 5-30 nanometers

How much data?

[Kaynig-Fittkau et al.]

Vision for understanding the brain



1mm cubed of brain

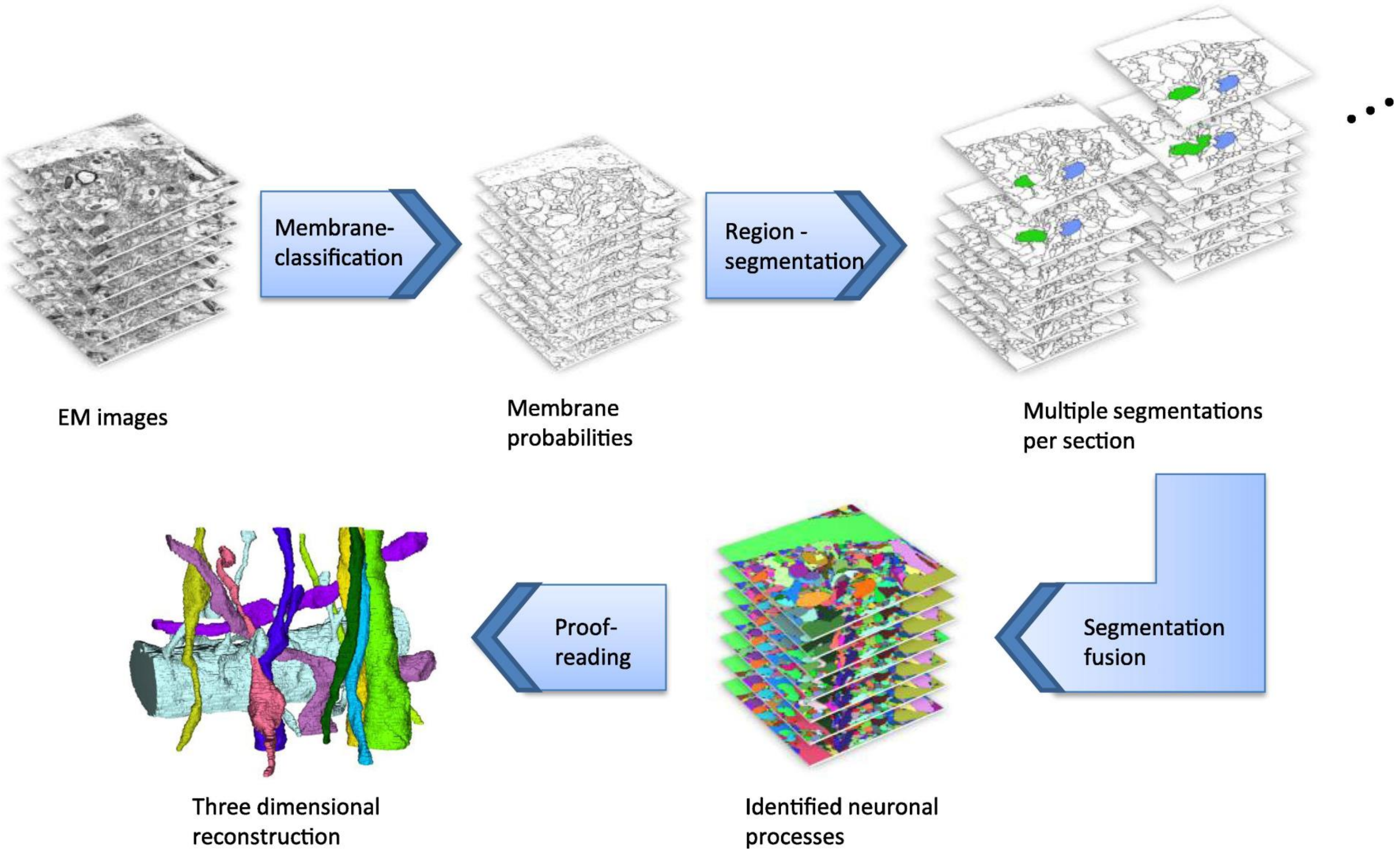
Image at 5-30 nanometers

How much data?

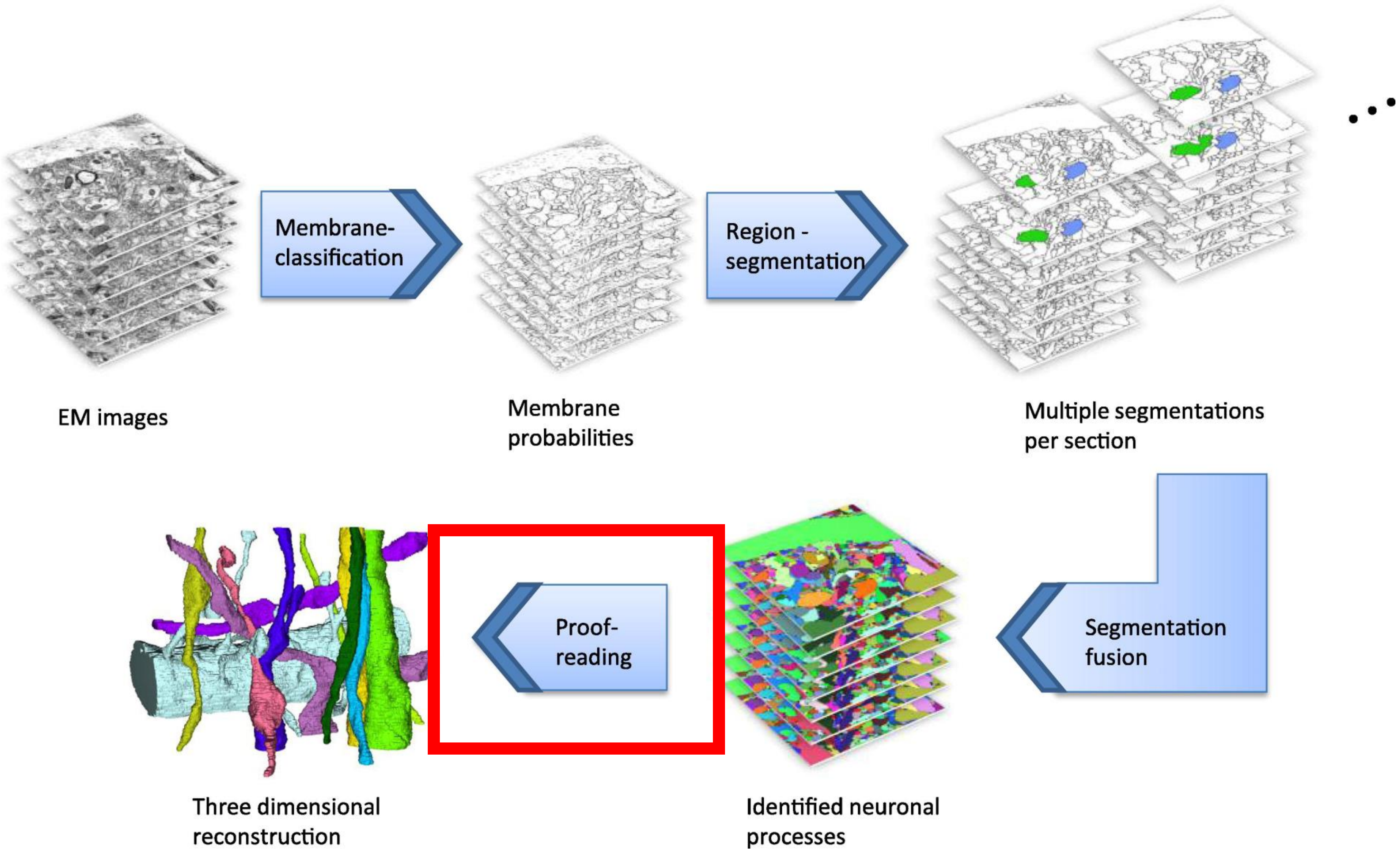
1 Petabyte –
1,000,000,000,000,000

~ All photos uploaded to Facebook per day

[Kaynig-Fittkau et al.]

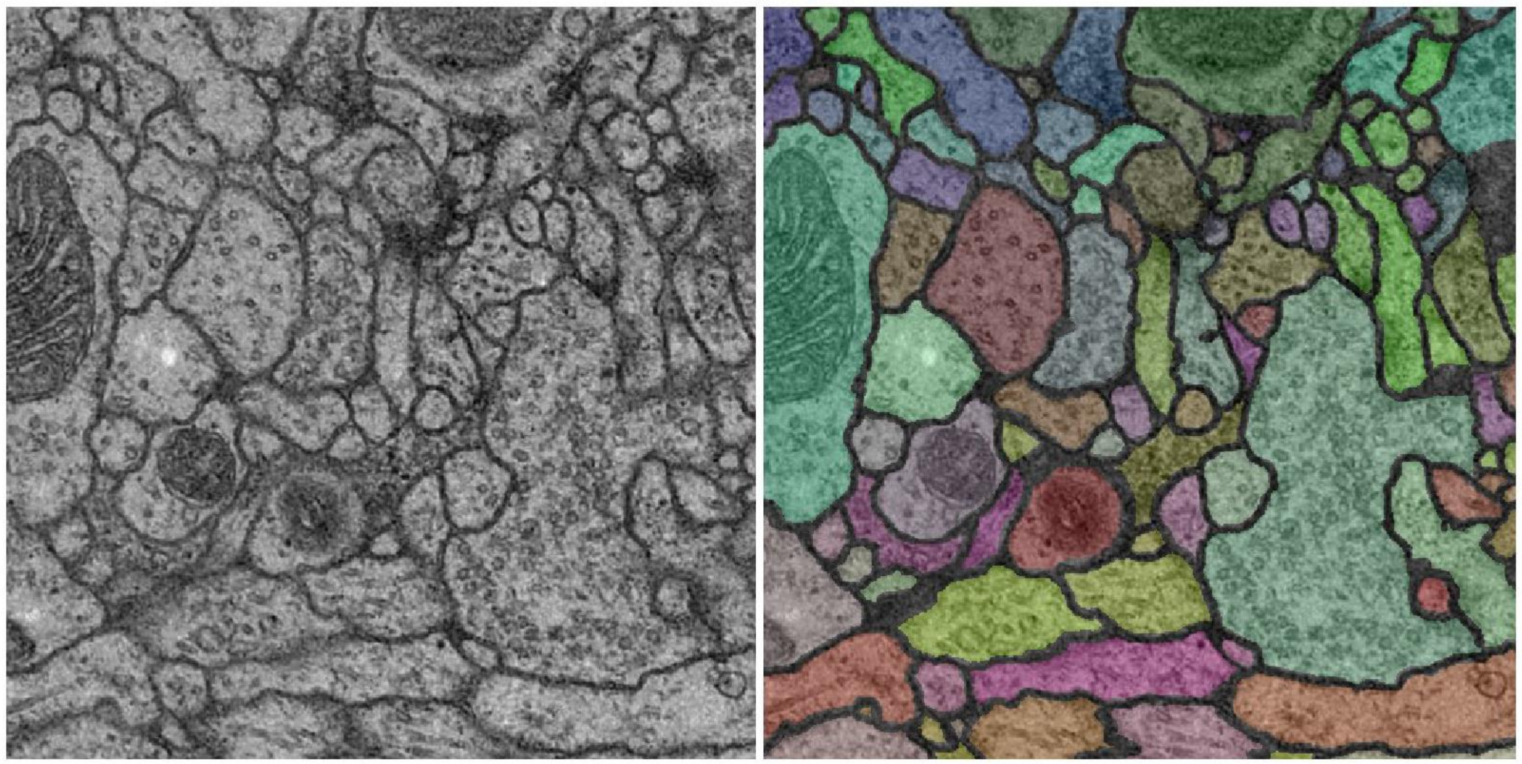


[Kaynig-Fittkau et al.]

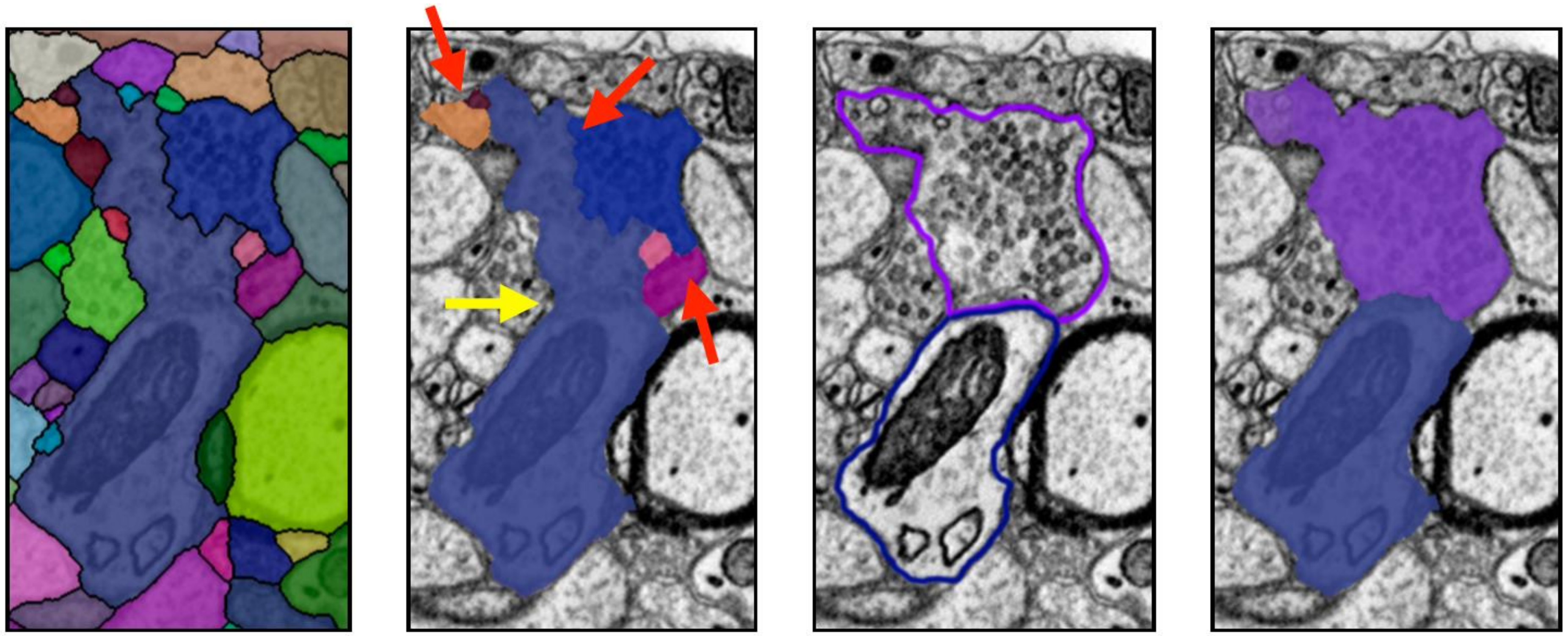


[Kaynig-Fittkau et al.]

Vision for understanding the brain

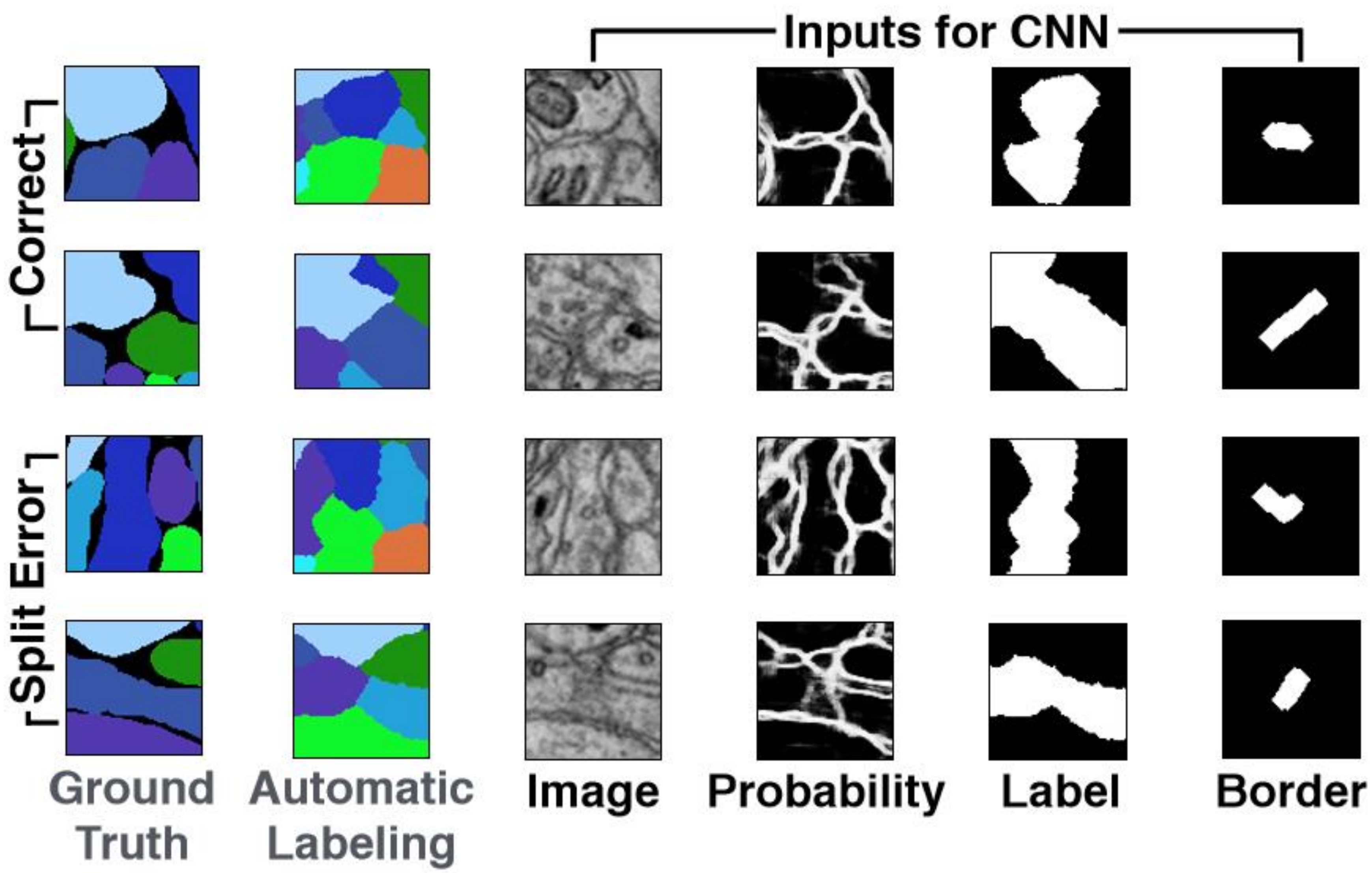


Ranzato



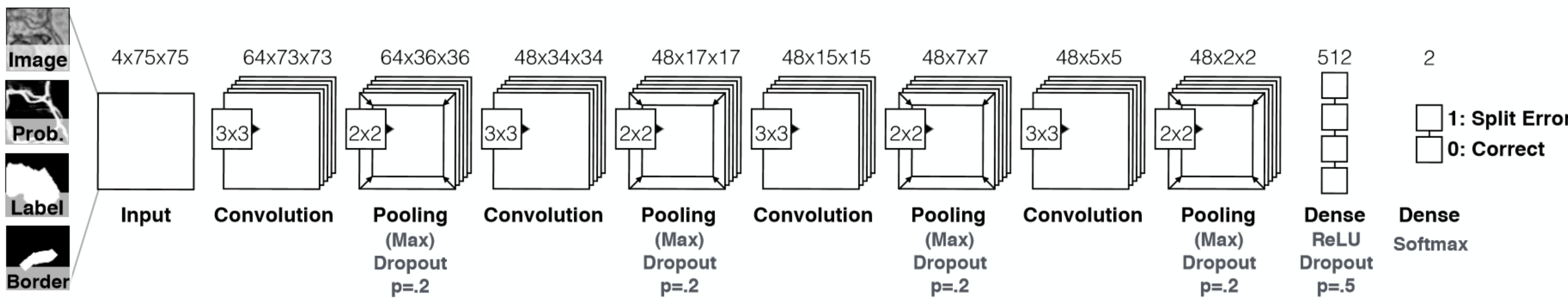
Initial Segmentation | Merge- and Split Errors | Correct Borders | Fixed Segmentation

[Haehn et al.]

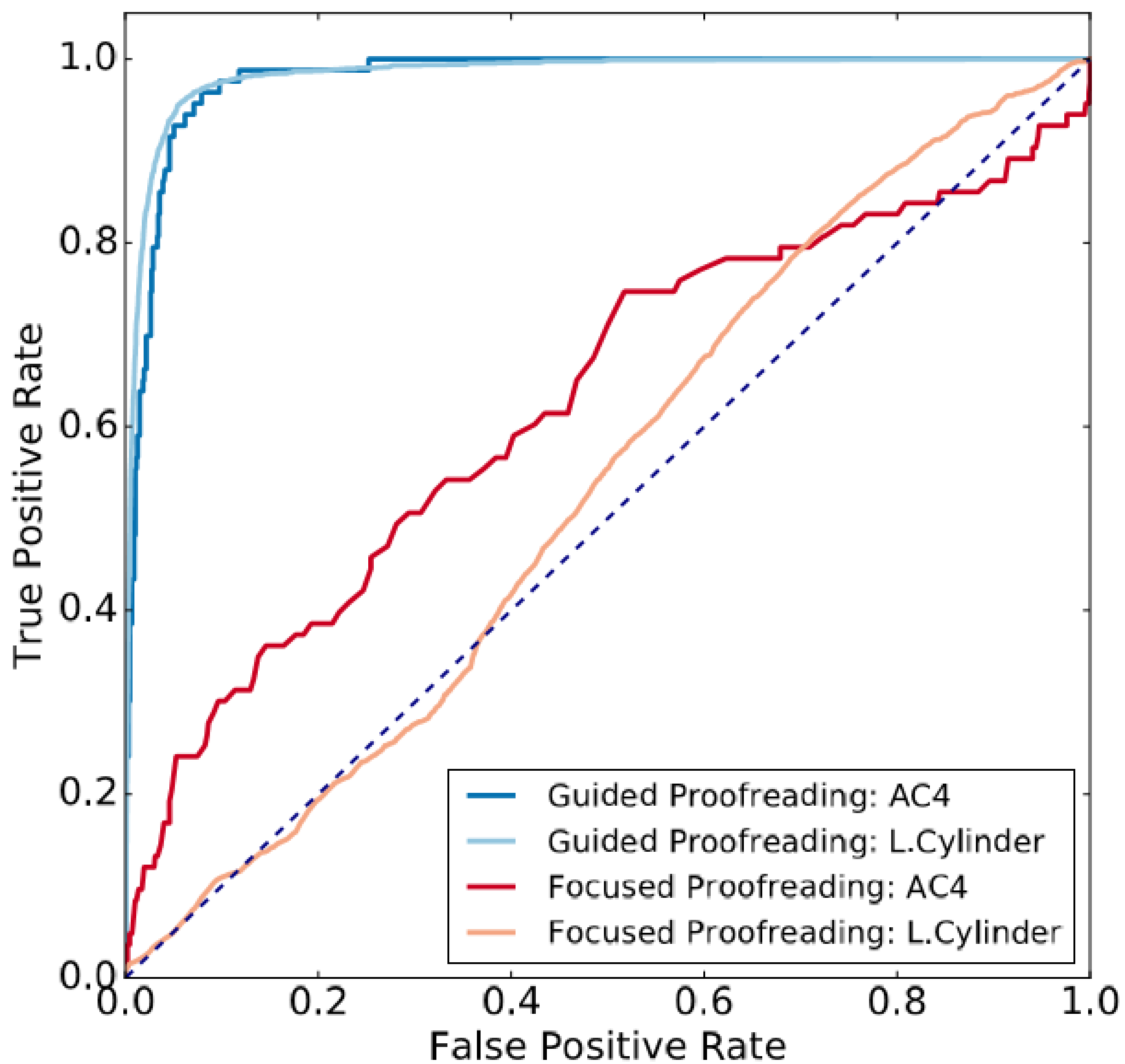

Inputs for CNN
Correct
Split Error
Ground Truth
Automatic Labeling
Image
Probability
Label
Border

[Haehn et al.]

Network Architecture

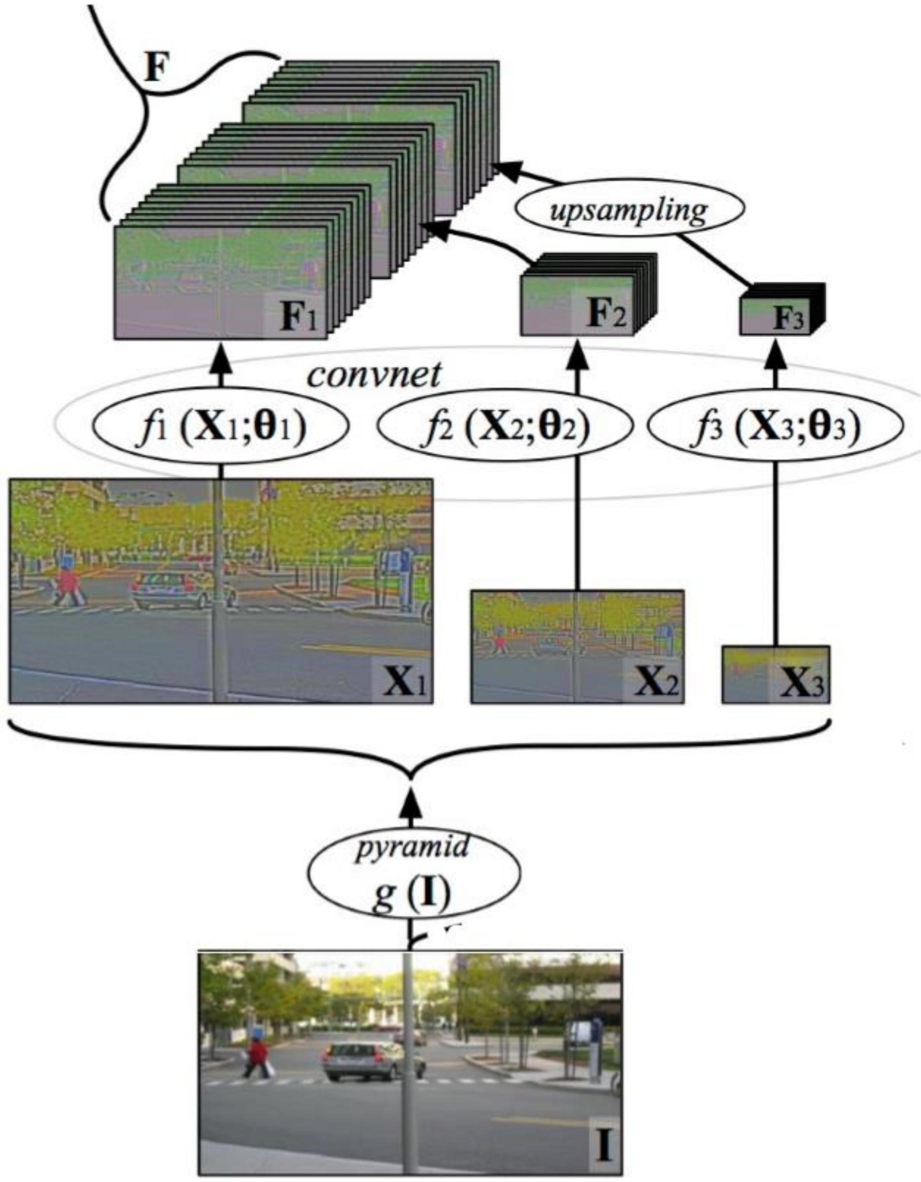


[Haehn et al.]


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Guided Proofreading: AC4
Guided Proofreading: L.Cylinder
Focused Proofreading: AC4
Focused Proofreading: L.Cylinder

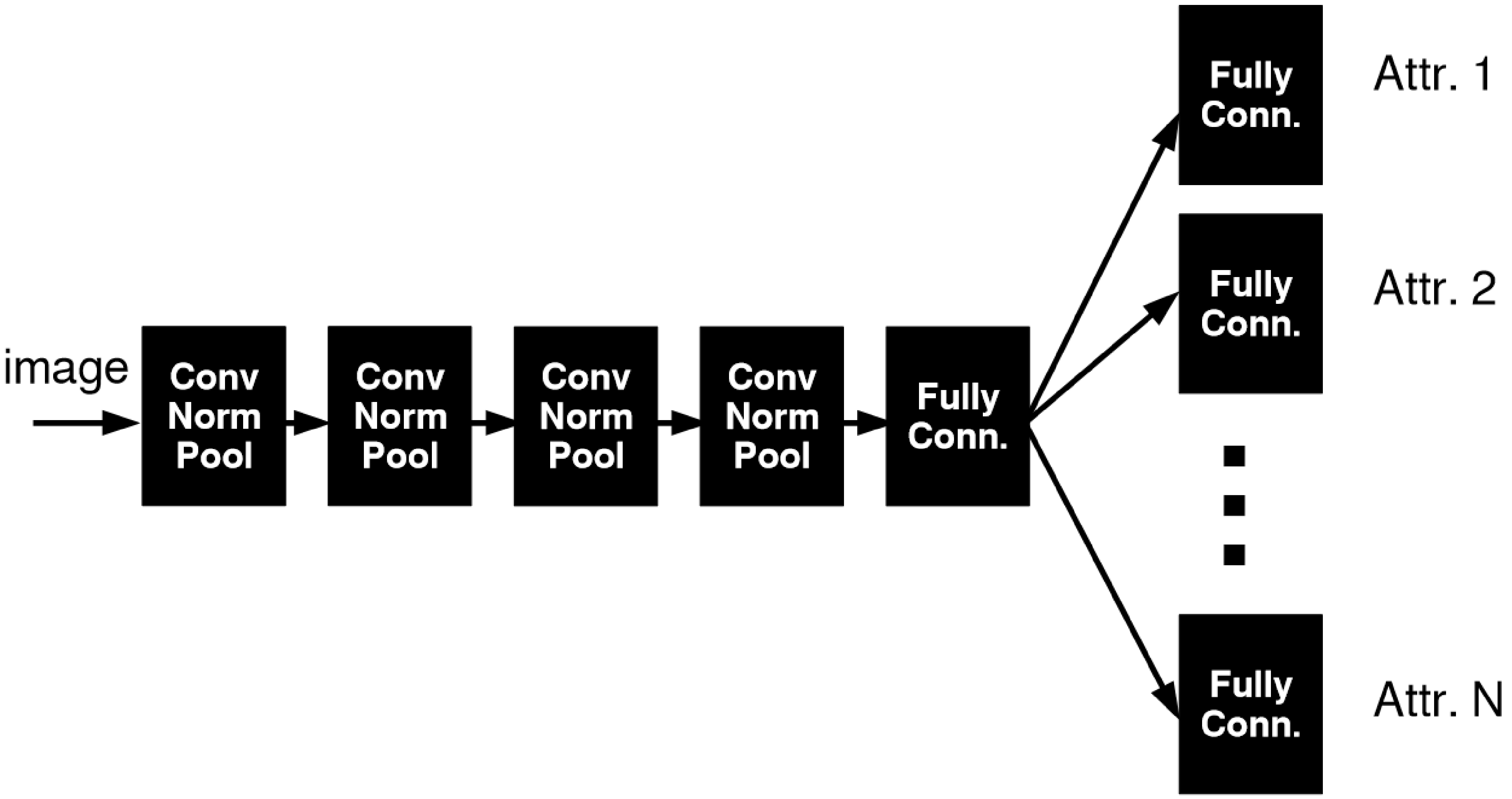
[Haehn et al.]

Fancier Architectures: Multi-Scale



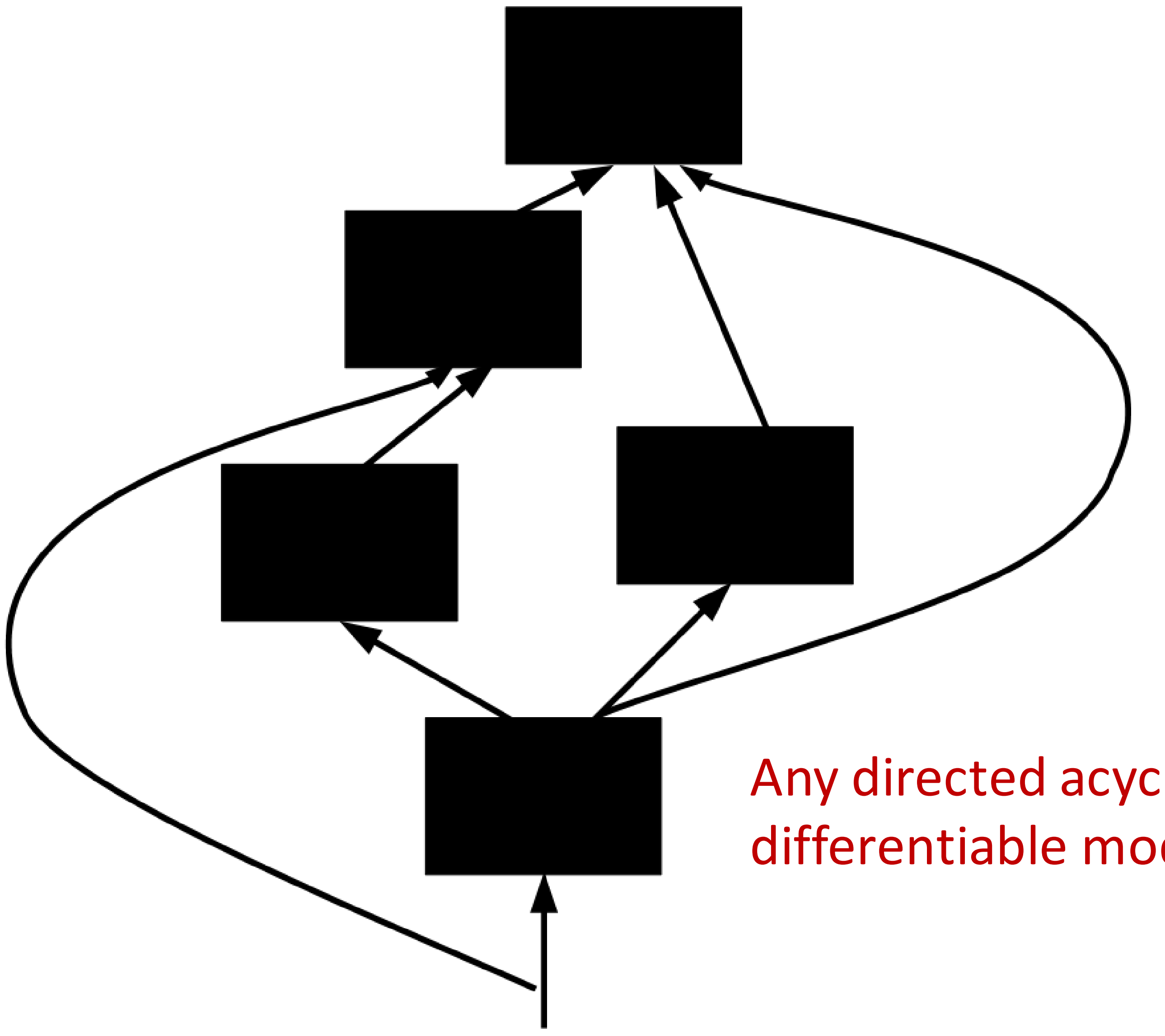
Farabet et al. “Learning hierarchical features for scene labeling” PAMI 2013

Fancier Architectures: Multi-Task



Zhang et al. “PANDA..” CVPR 2014

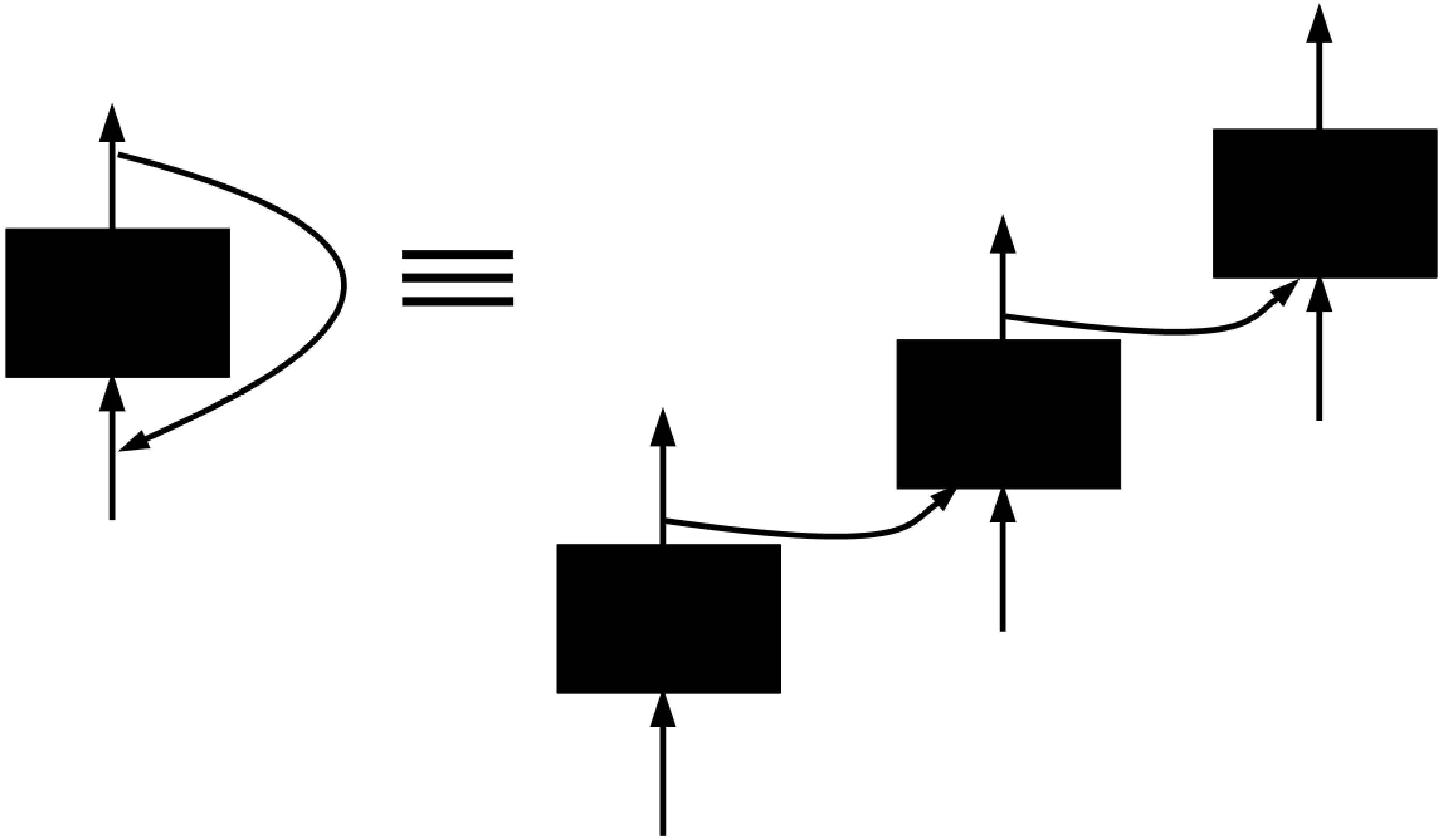
Fancier Architectures: Generic DAG



Any directed acyclic graph (DAG) of differentiable modules is allowed.

Fancier Architectures: Generic DAG

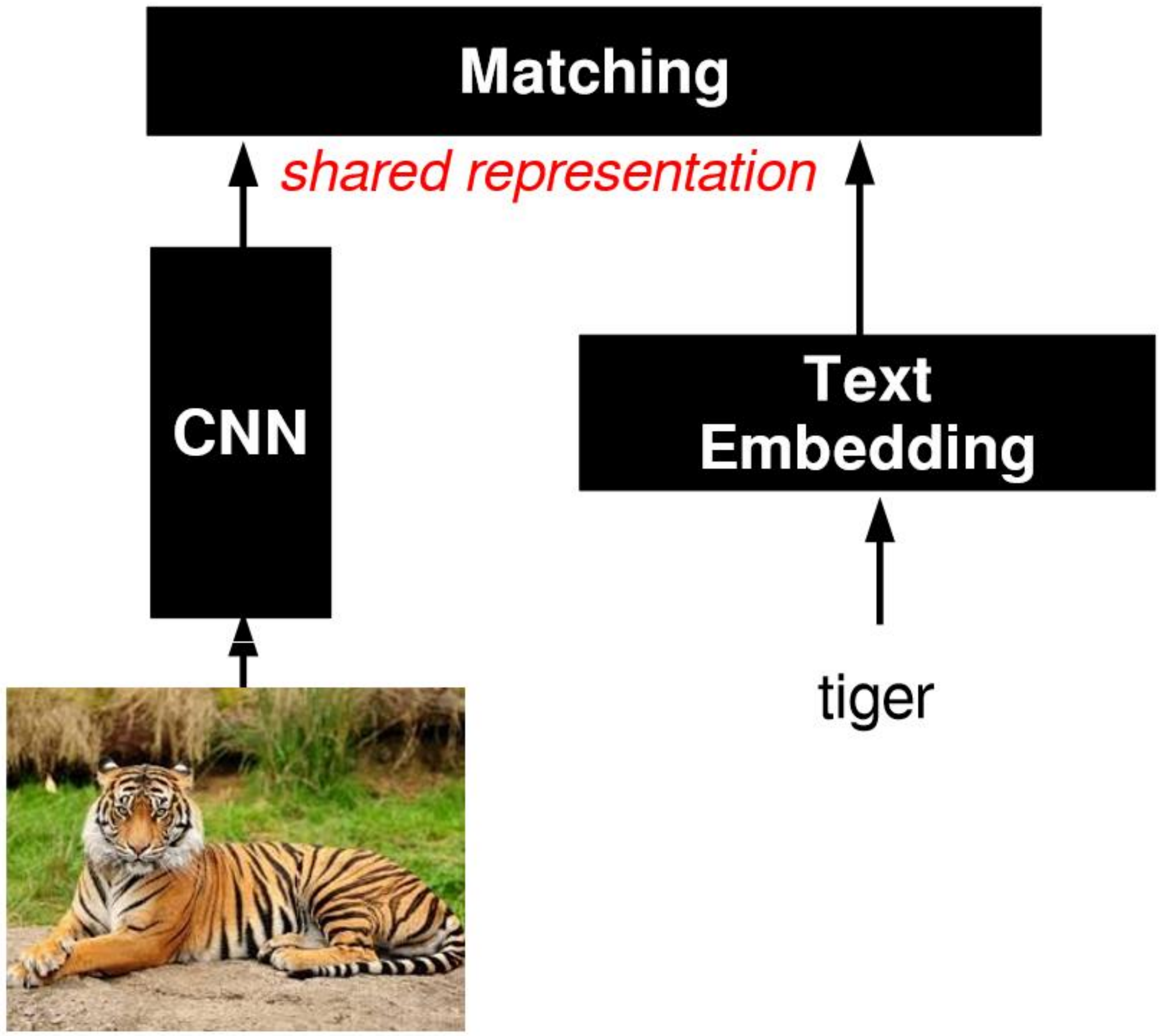
If there are cycles (RNN), one needs to un-roll it.



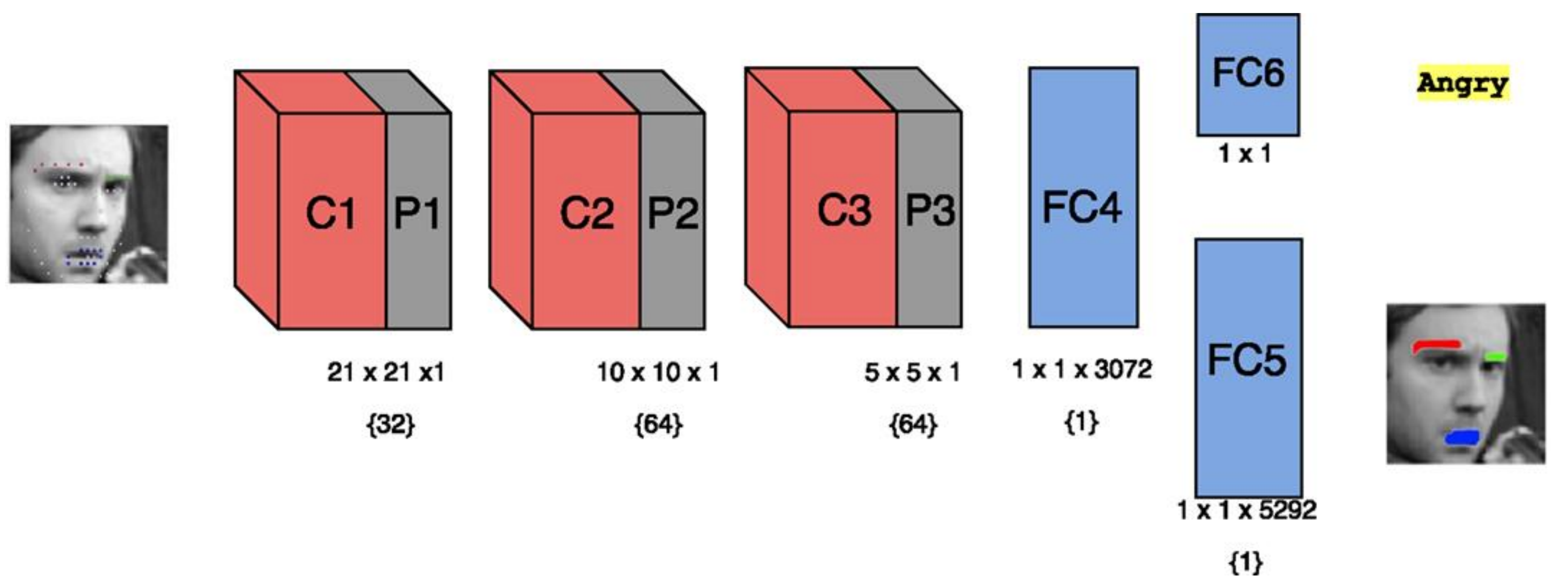
Pinheiro, Collobert "Recurrent CNN for scene labeling" ICML 2014
Graves "Offline Arabic handwriting recognition.." Springer 2012

What about learning across ‘domains’?

Fancier Architectures: Multi-Modal

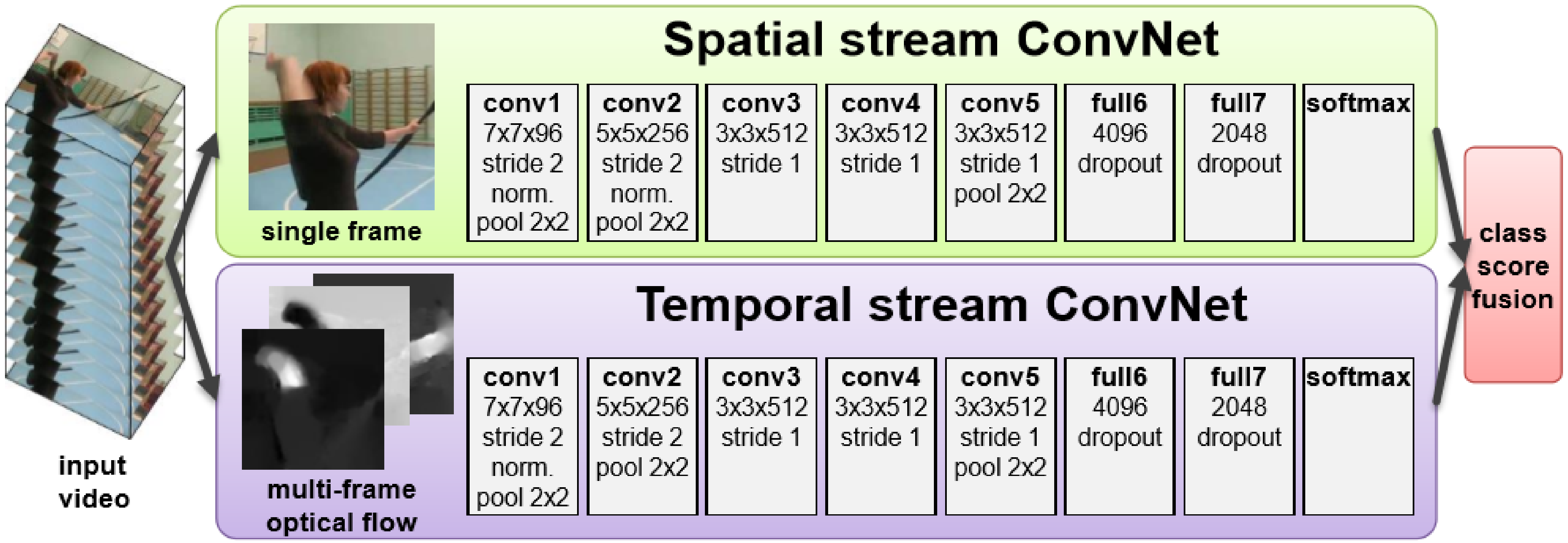


Frome et al. “Devise: a deep visual semantic embedding model” NIPS 2013

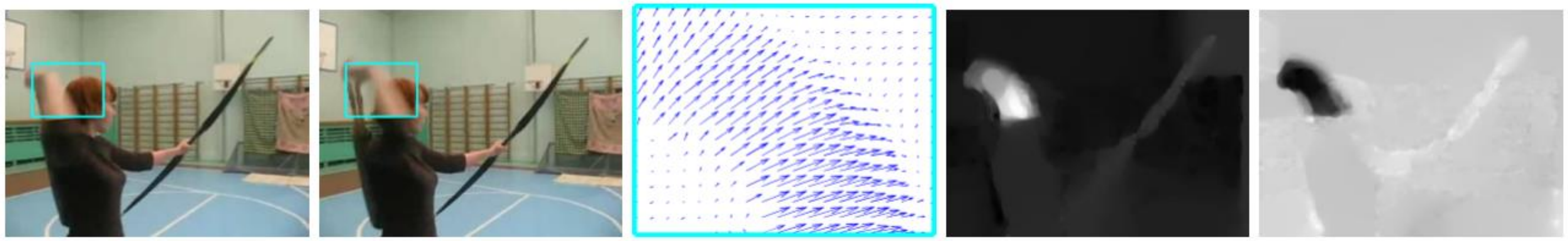


C1
P1
C2
P2
C3
P3
FC4
FC6
FC5
Angry
21 x 21 x1
{32}
10 x 10 x 1
{64}
5 x 5 x 1
{64}
1 x 1 x 3072
{1}
1 x 1
1 x 1 x 5292
{1}

Two-stream networks – *action recognition*



[Simonyan et al. 2014]



(a) (b) (c) (d) (e)

[Simonyan et al. 2014]

Learning Deep Representations For Ground-to-Aerial Geolocalization

Tsung-Yi Lin, Yin Cui, Serge Belongie, James Hays

CORNELL NYC TECH

Georgia Tech

CVPR 2015

View From Your Window Contest

June 9, 2010 – Feb. 4, 2015

Where was the photo taken?

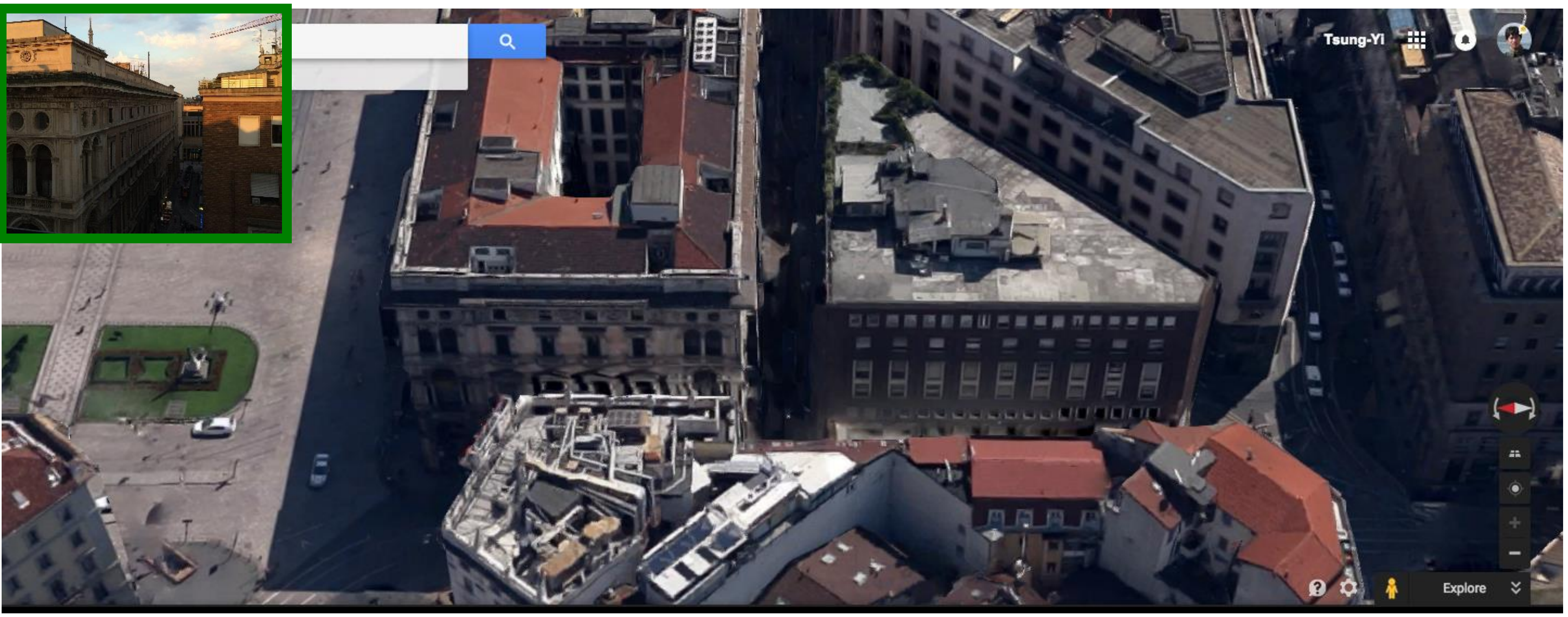
Ans:
Milano, Italy

To Geolocalize a Photo

- One can capture every corner on the earth

    …

To Geolocalize a Photo




Tsung-Yi
Explore


Tsung-Yi
Explore

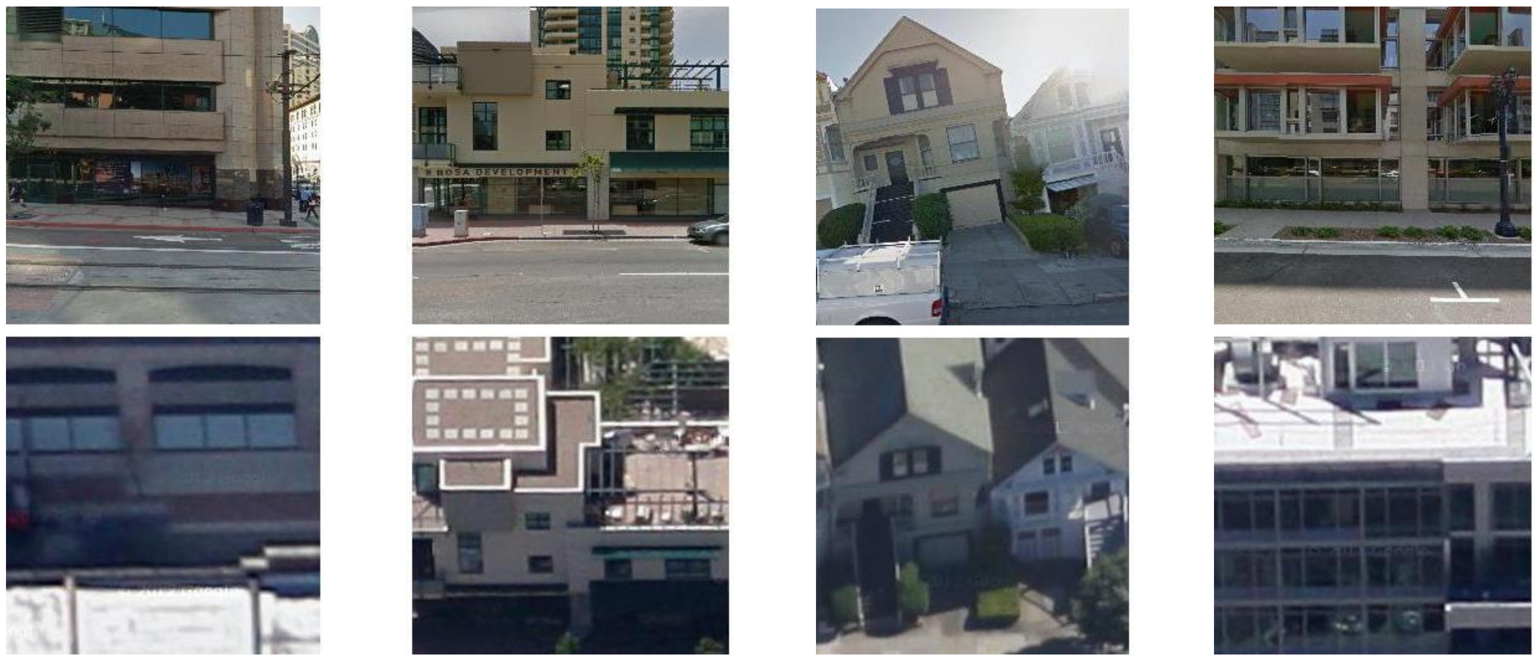
How To Match Ground-to-Aerial?

Shan et al., Accurate Geo-registration by Ground-to-Aerial Image Matching, 3DV’14
Bansal et al., Ultra-wide baseline façade matching for geo-localization, ECCV workshop’12

Are these the same location?

Ground

Aerial

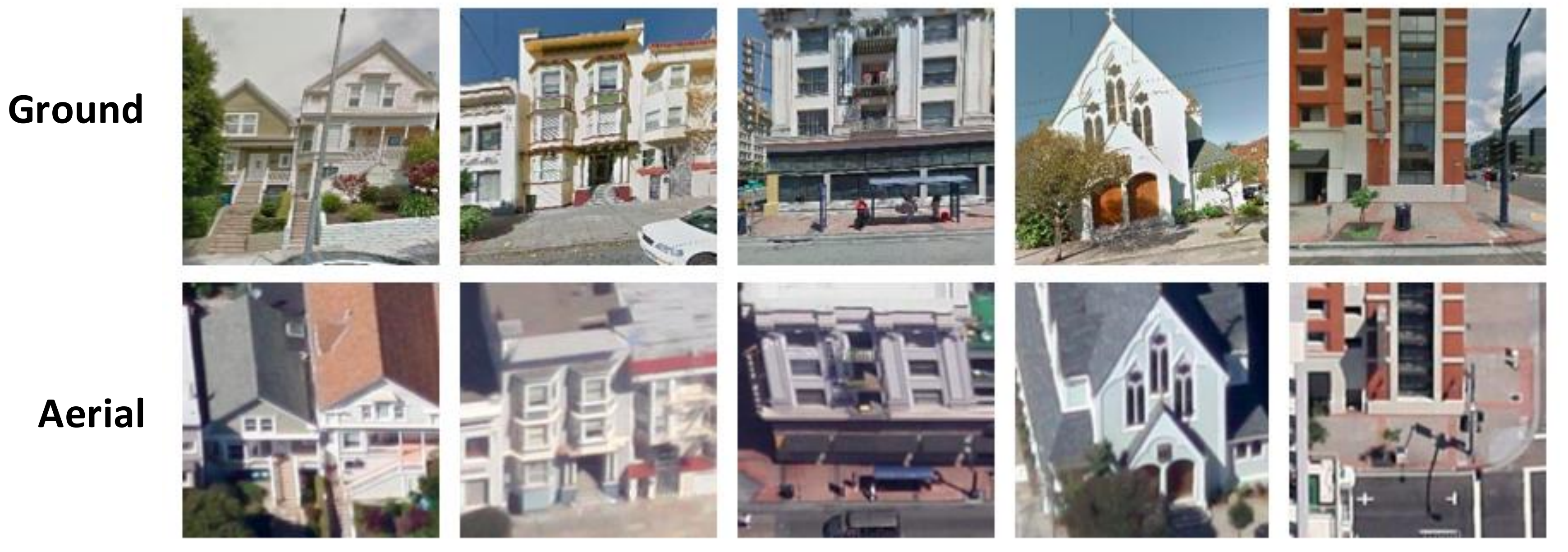


Are these the same location?

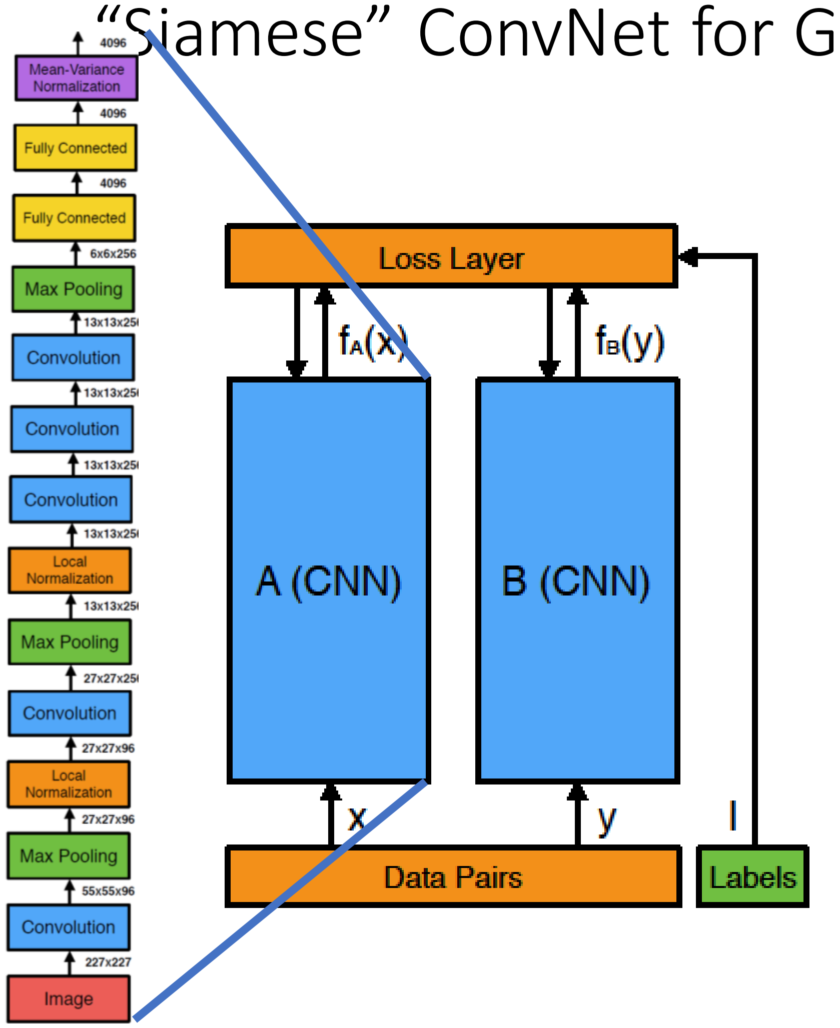
Ground

Aerial

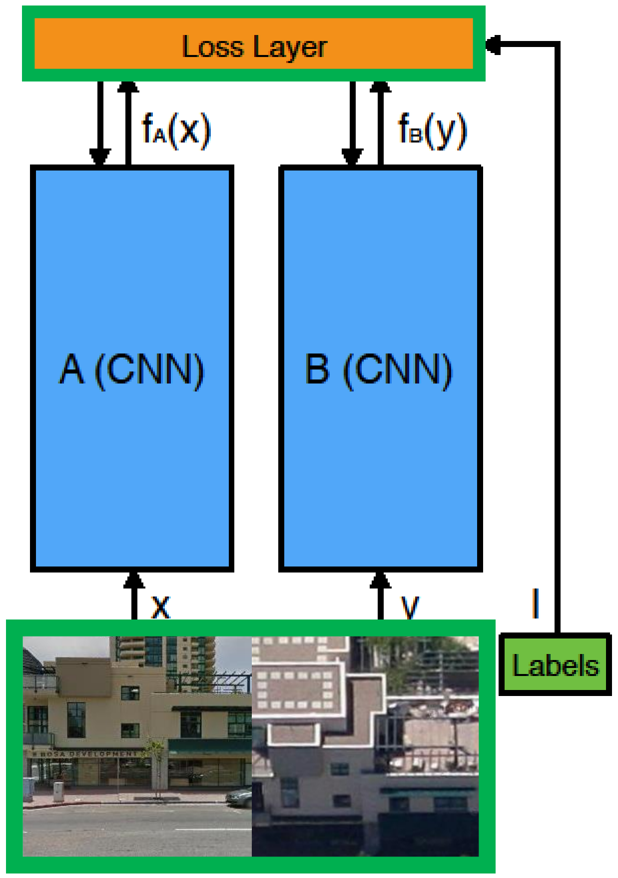
Cross-view Pairs



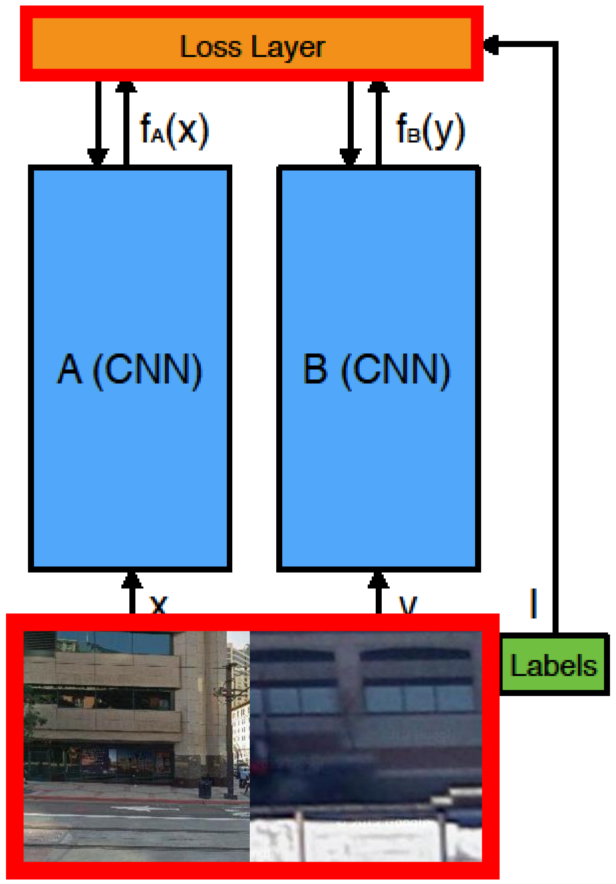
“Siamese” ConvNet for Ground-to-Aerial Matching

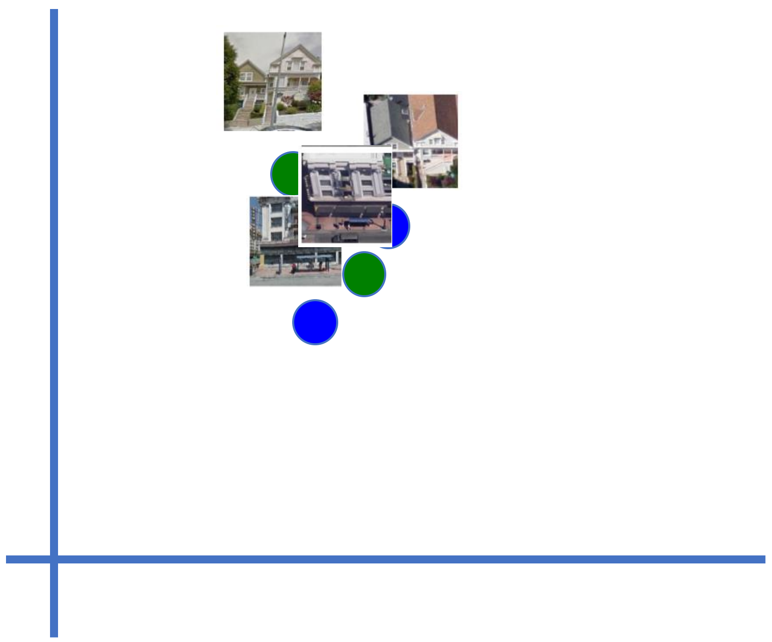


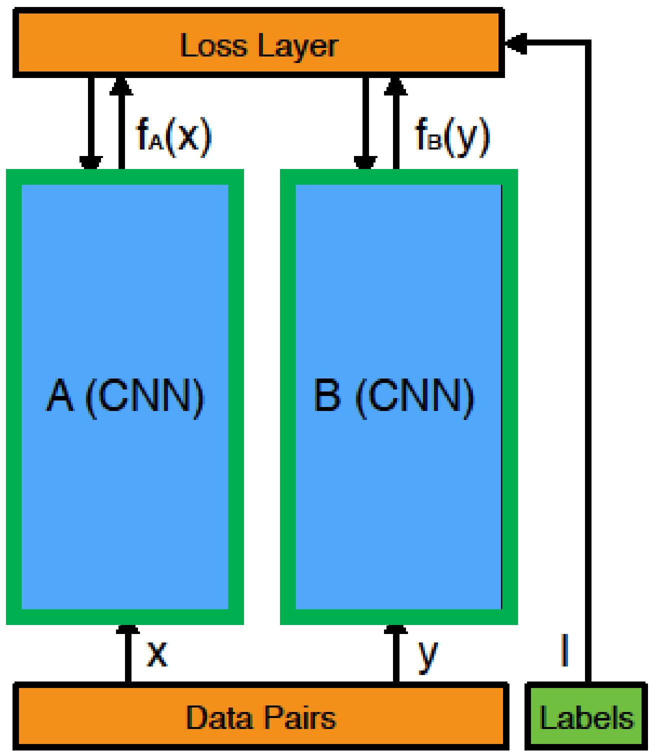
“Siamese” ConvNet for Ground-to-Aerial Matching



“Siamese” ConvNet for Ground-to-Aerial Matching

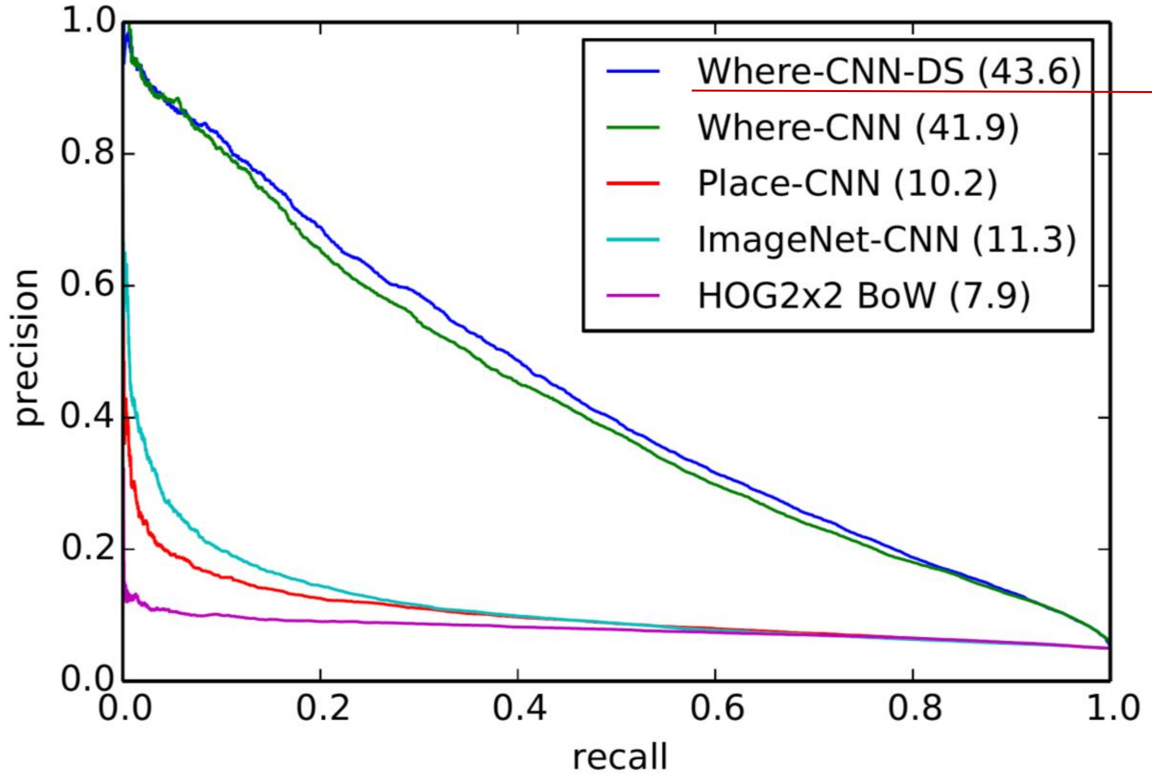






For ground-aerial image pairs, should A, B networks share the same weights?

Quantitative Evaluation



Minor improvement with 'domain-specific' weights.

Big space of designs!

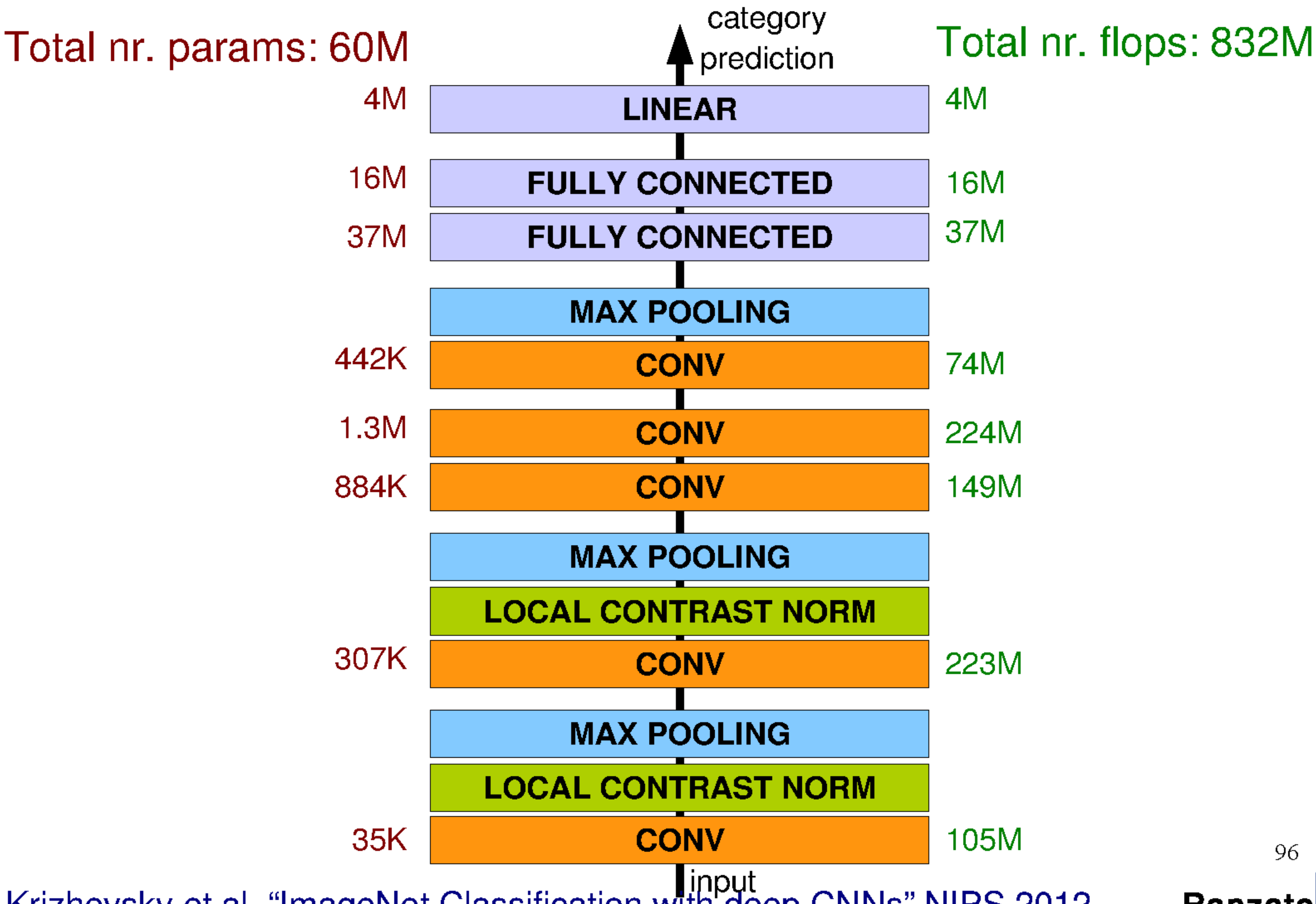
But we still don’t even know how many layers we need.


WE NEED TO GO
DEEPER


NOT SURE IF
WE NEED TO GO DEEPER
imgflip.com

WE NEED TO GO
DEEPER

Architecture for Classification



Krizhevsky et al. "ImageNet Classification with deep CNNs" NIPS 2012

Ranzato

Beyond AlexNet

VERY DEEP CONVOLUTIONAL NETWORKS
FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan & Andrew Zisserman 2015

These are the pre-trained “VGG” networks that you use in project 5

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

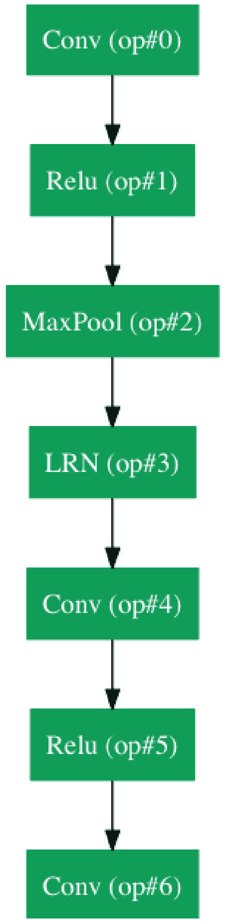
Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**

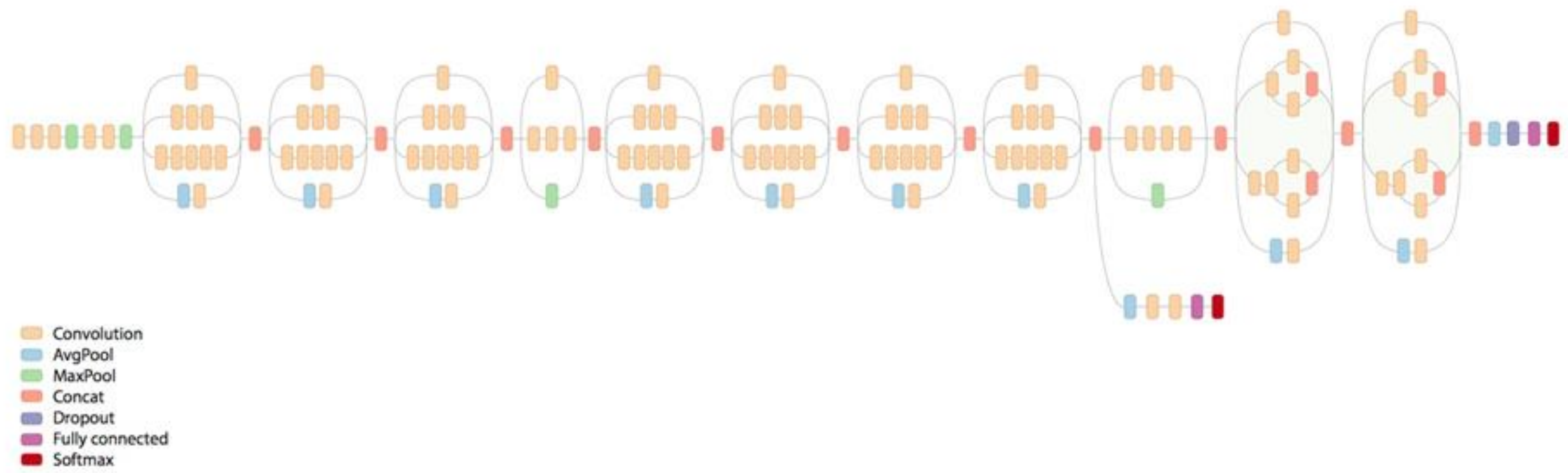
Google LeNet (2014)



22 layers

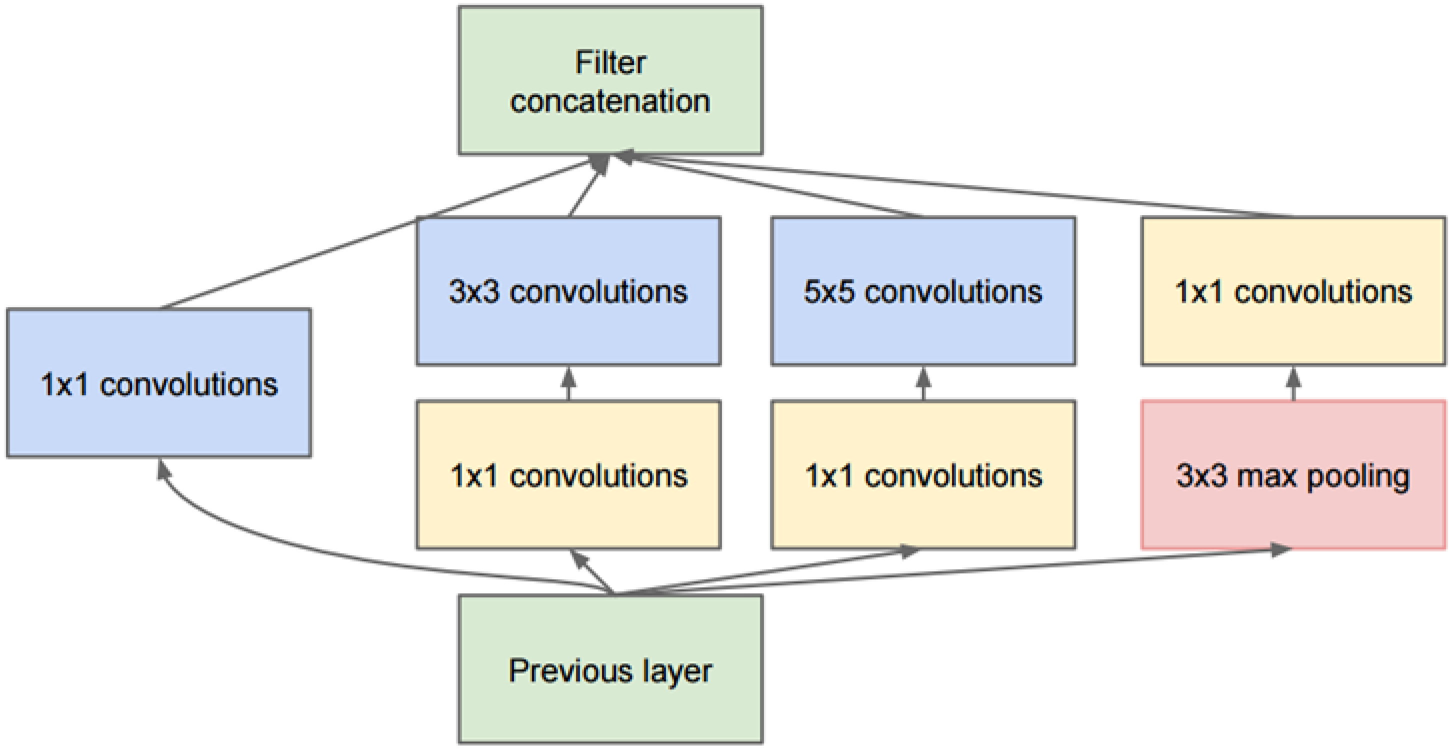
6.67% error
ImageNet top 5

Inception!



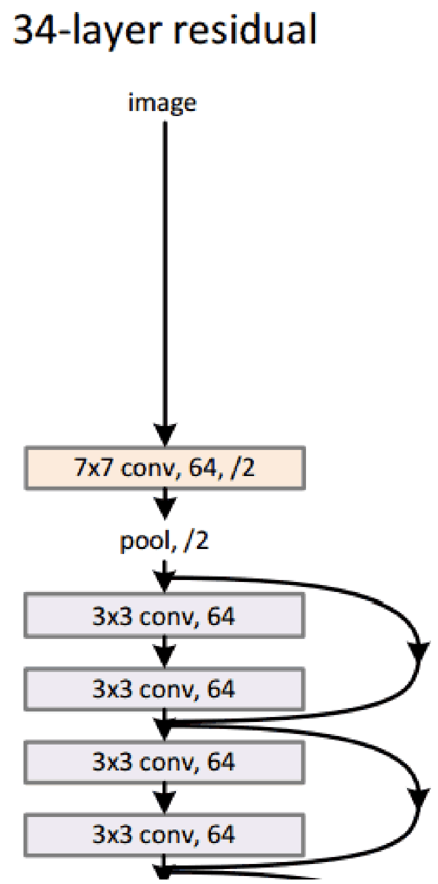
Another view of GoogLeNet's architecture.

Parallel layers



Full Inception module

ResNet (He et al., 2015)

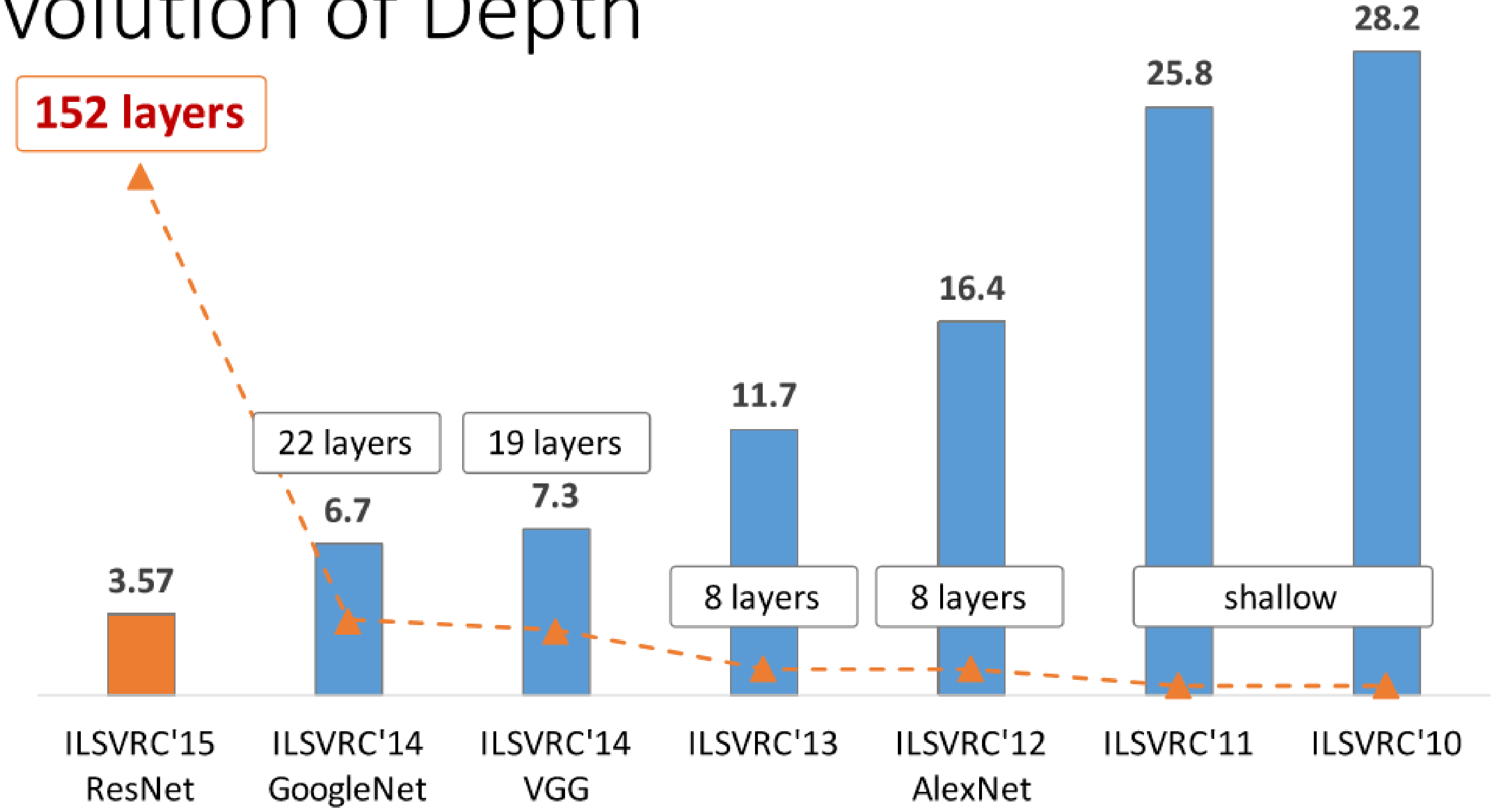


ResNet won ILSVRC 2015 with a top-5 error rate of 3.6%

Depending on their skill and expertise, humans generally hover around a 5-10% error.

But the task is arguably not well defined.

Revolution of Depth



ImageNet Classification top-5 error (%)

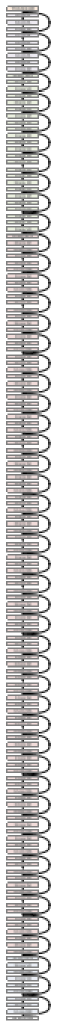
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)

VGG, 19 layers
(ILSVRC 2014)

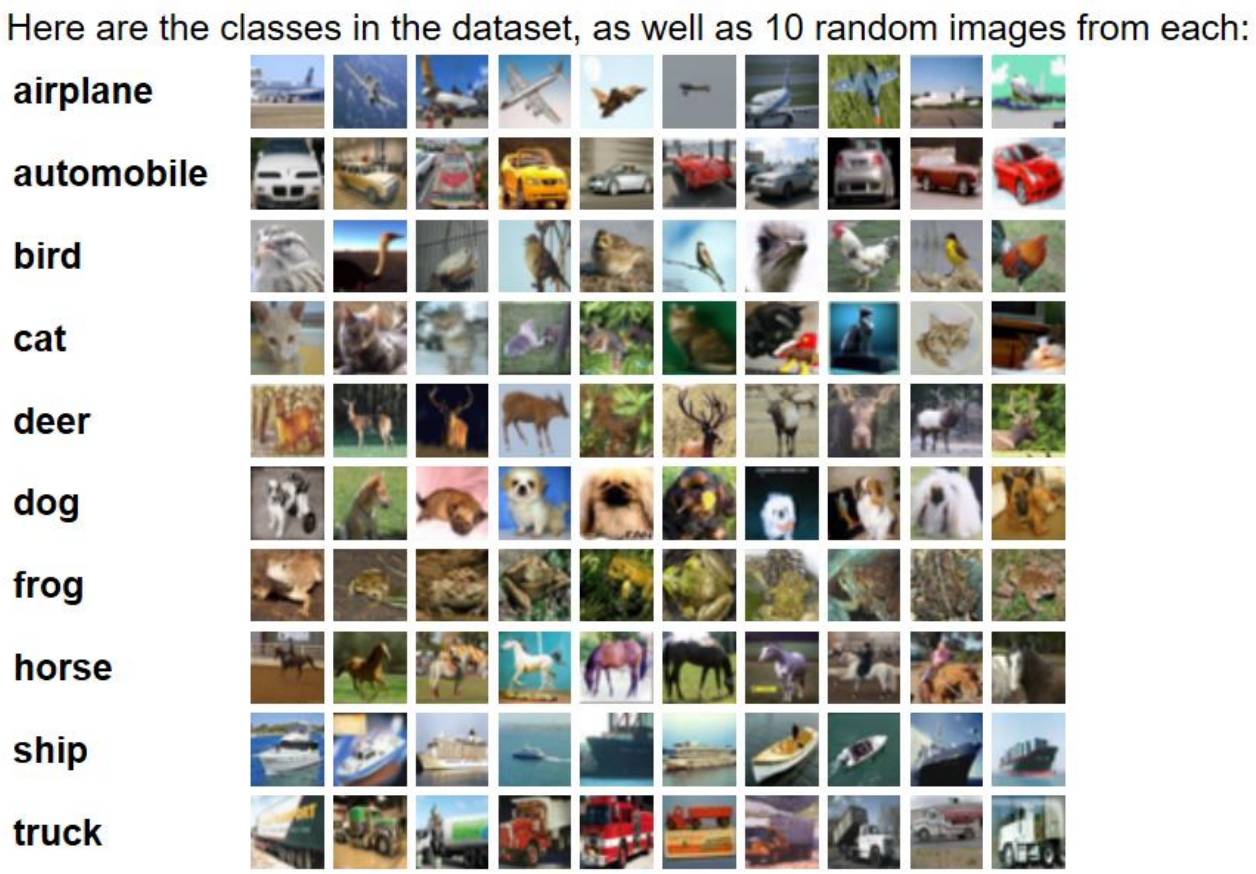
ResNet, 152 layers
(ILSVRC 2015)



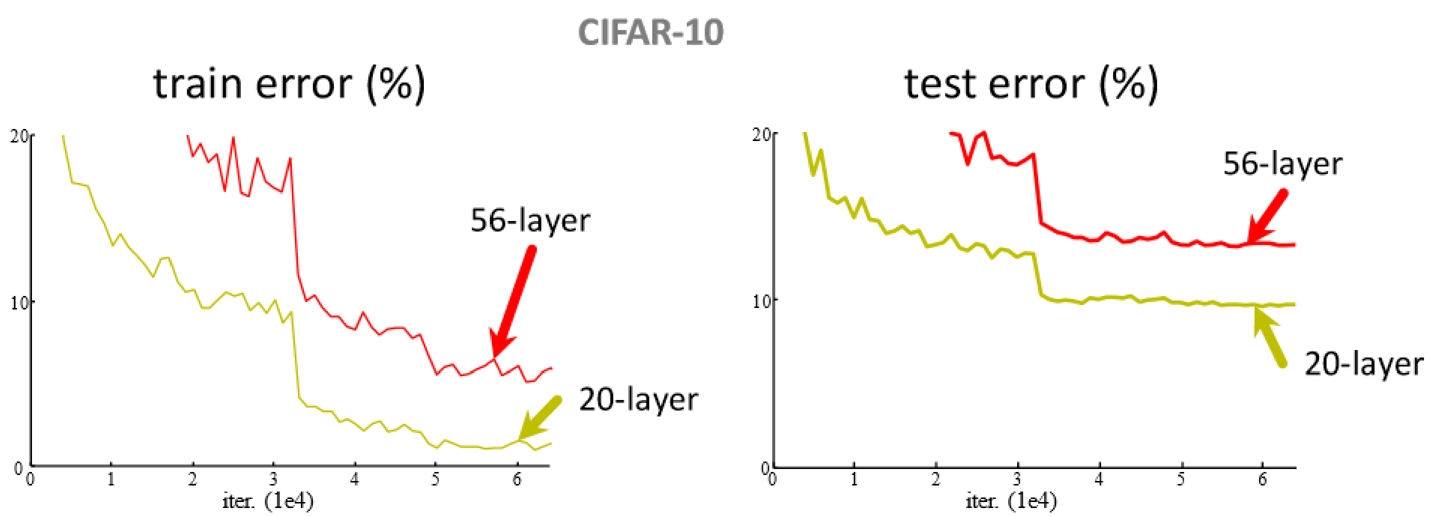
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. CVPR 2016.

CIFAR-10

- 60,000 32x32 color images, 10 classes



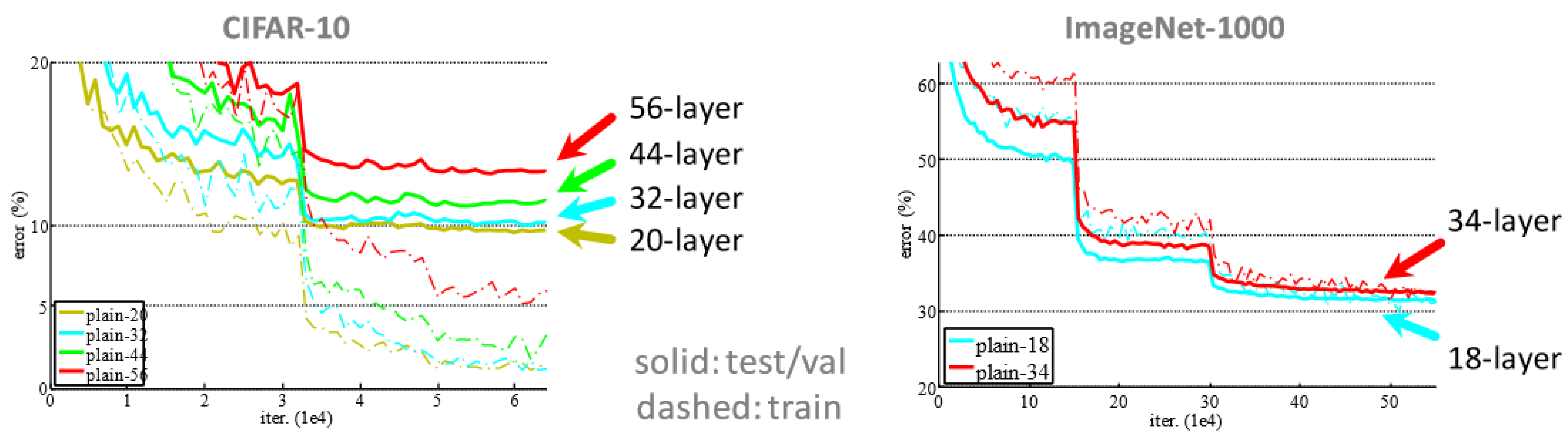
Simply stacking layers?



- *Plain* nets: stacking 3x3 conv layers…
- 56-layer net has **higher training error** and test error than 20-layer net

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

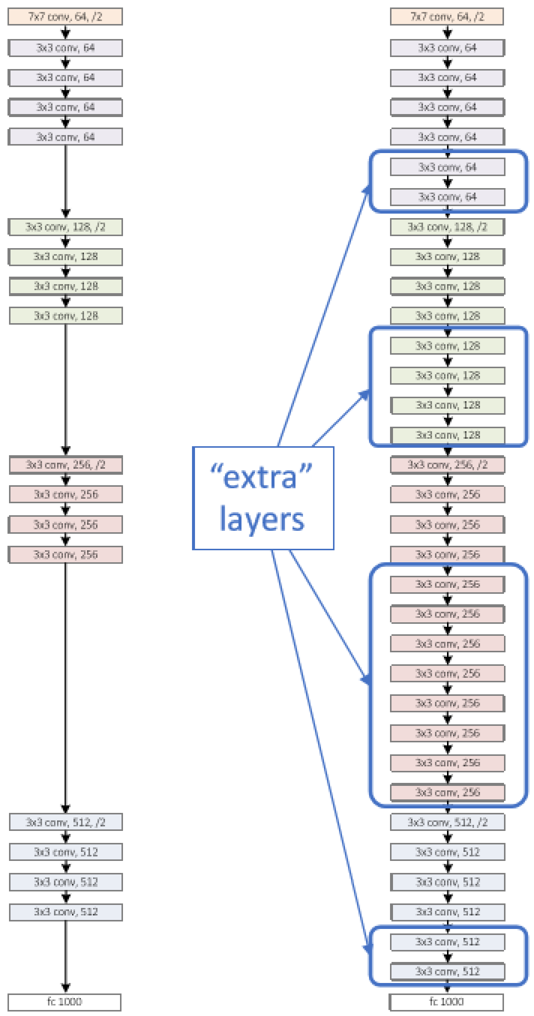
Simply stacking layers?



- “Overly deep” plain nets have **higher training error**
- A general phenomenon, observed in many datasets

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. CVPR 2016.

a shallower model (18 layers)



a deeper counterpart (34 layers)

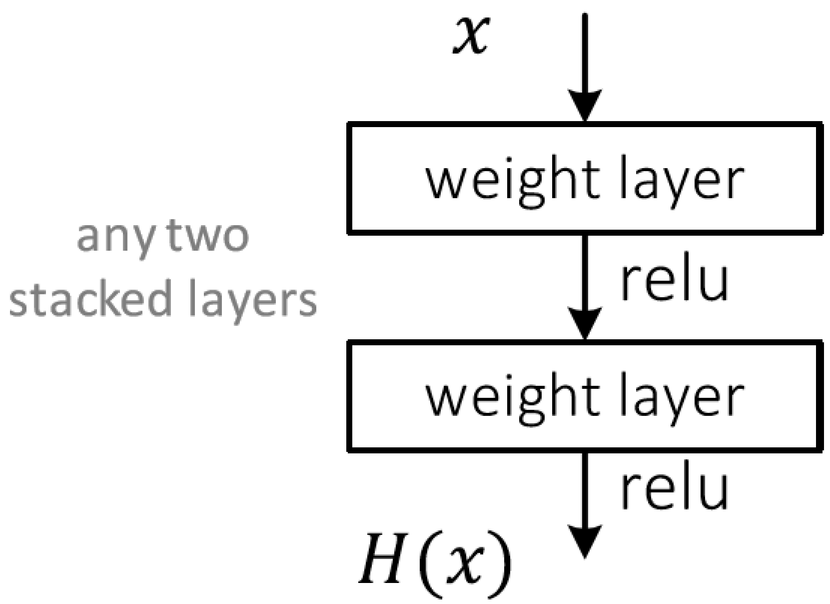
- Richer solution space
- A deeper model should not have **higher training error**

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. CVPR 2016.

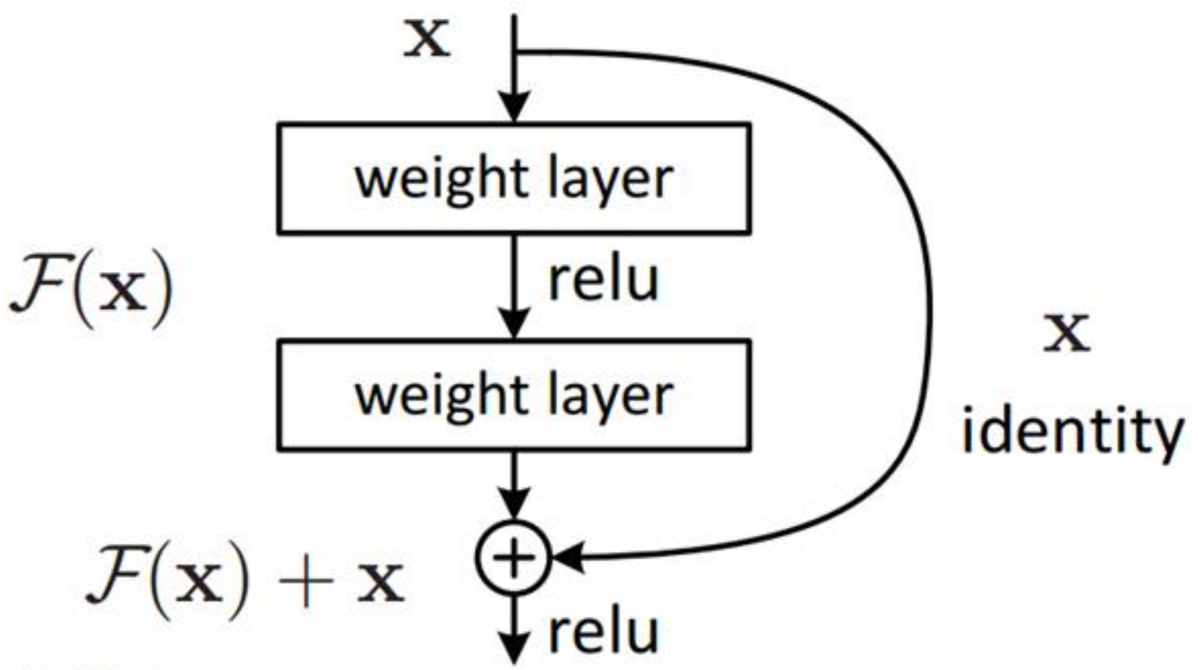
Regular net

$H(x)$ is any desired mapping,

hope the 2 weight layers fit $H(x)$



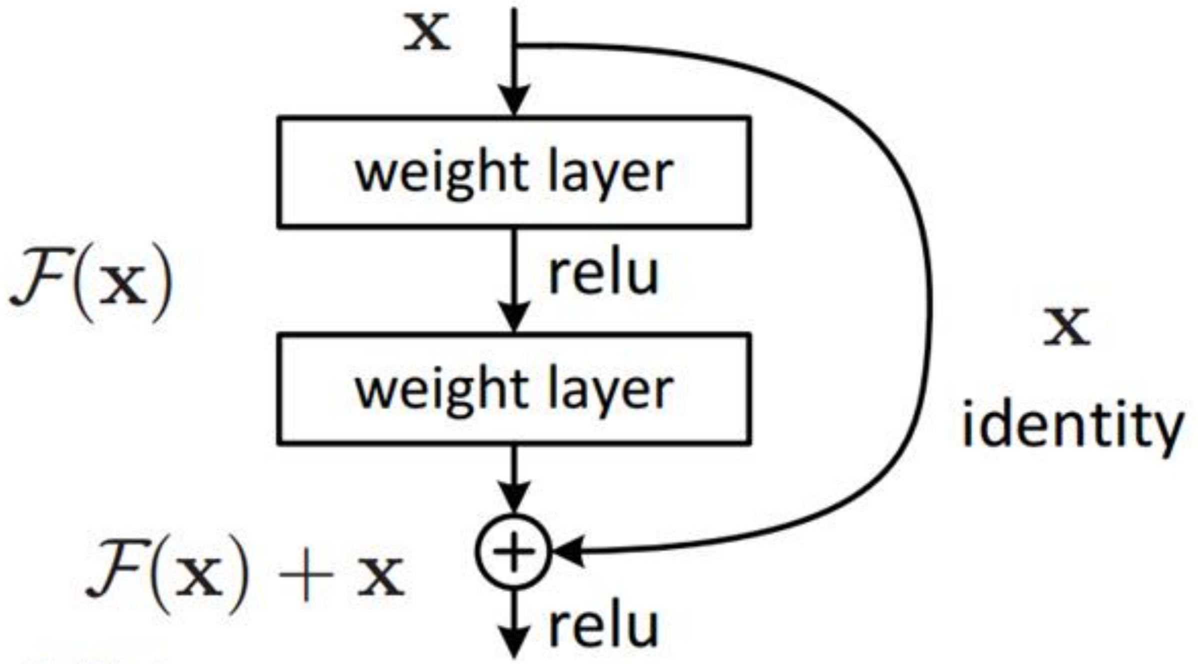
Residual Unit



A residual block

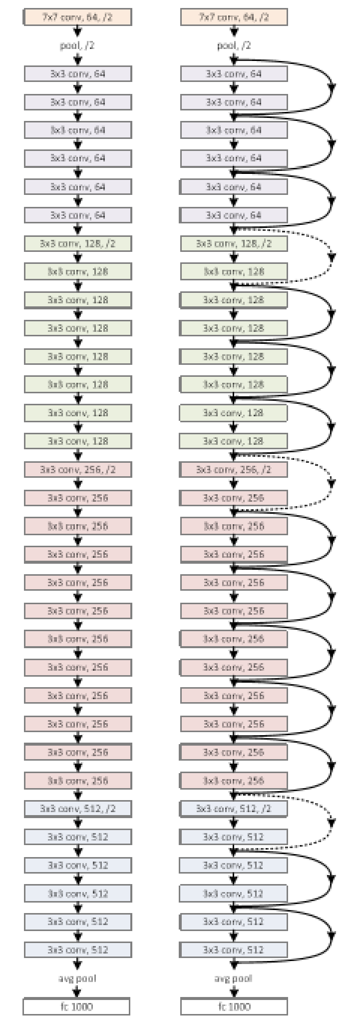
Residual Unit

The inputs of a lower layer is made available to a node in a higher layer.



A residual block

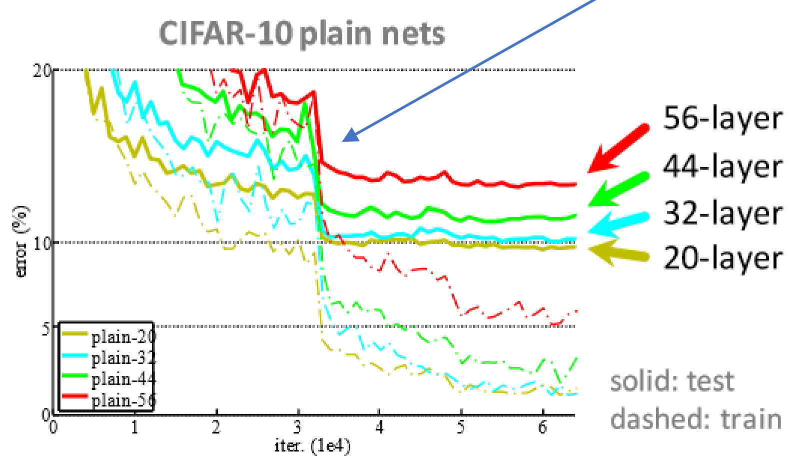
Network “Design”

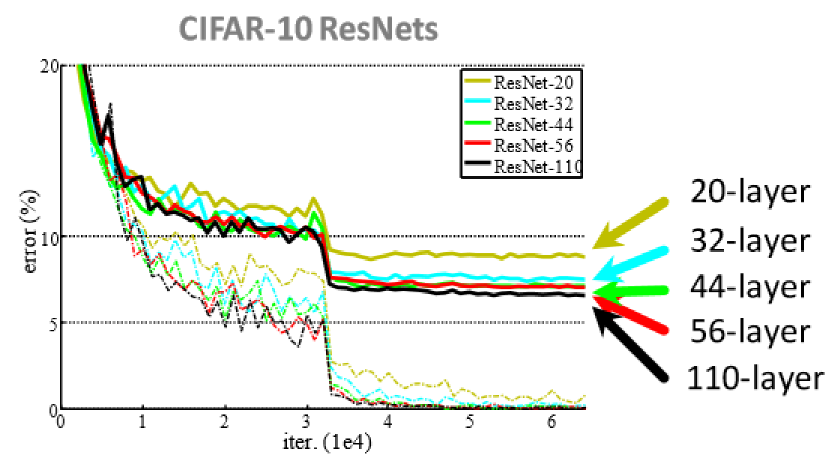


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. CVPR 2016.

Why so steep?

CIFAR-10 experiments

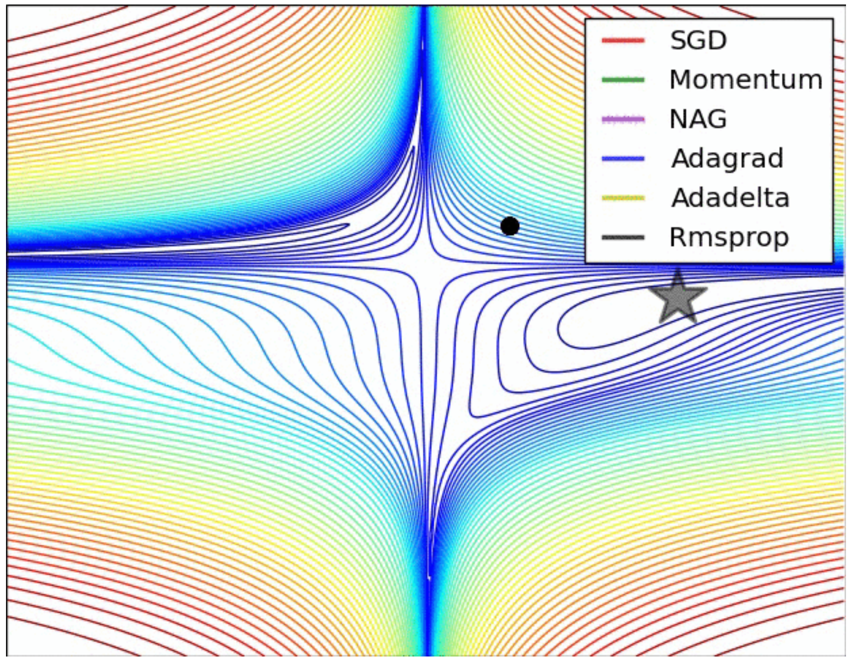


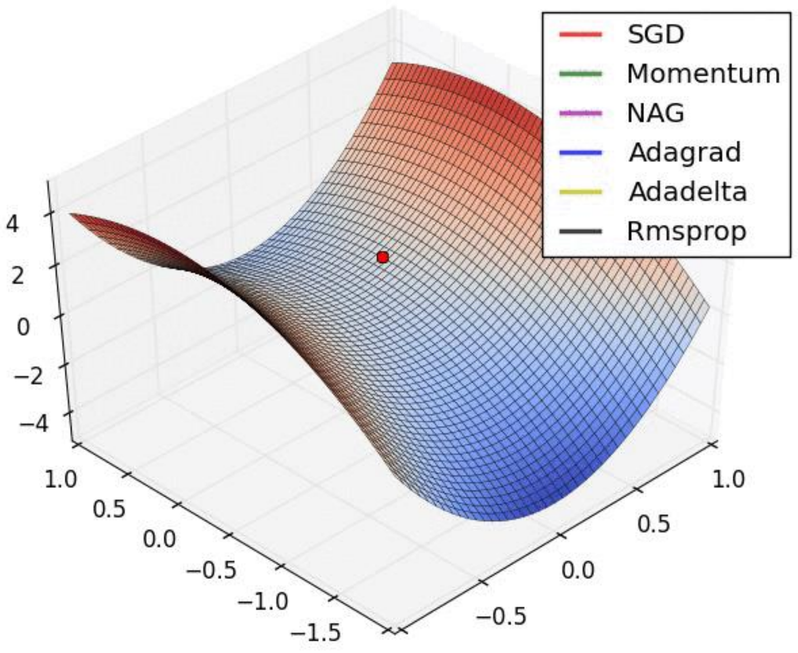


- Deep ResNets can be trained without difficulties
- Deeper ResNets have **lower training error**, and also lower test error

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. CVPR 2016.

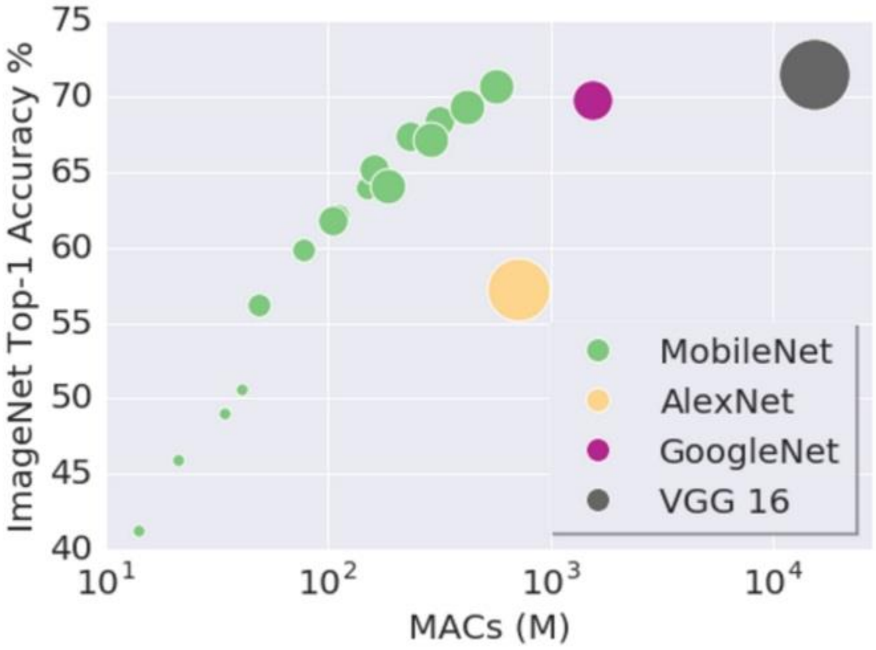
Flat regions in energy landscape





James, do we *have* to go deeper?

Compute vs. parameters / multiply-adds



Hmm…efficient nets…
might be useful for final project ???

https://www.infoq.com/news/2017/06/google-mobilenets-tensorflow
https://arxiv.org/abs/1704.04861