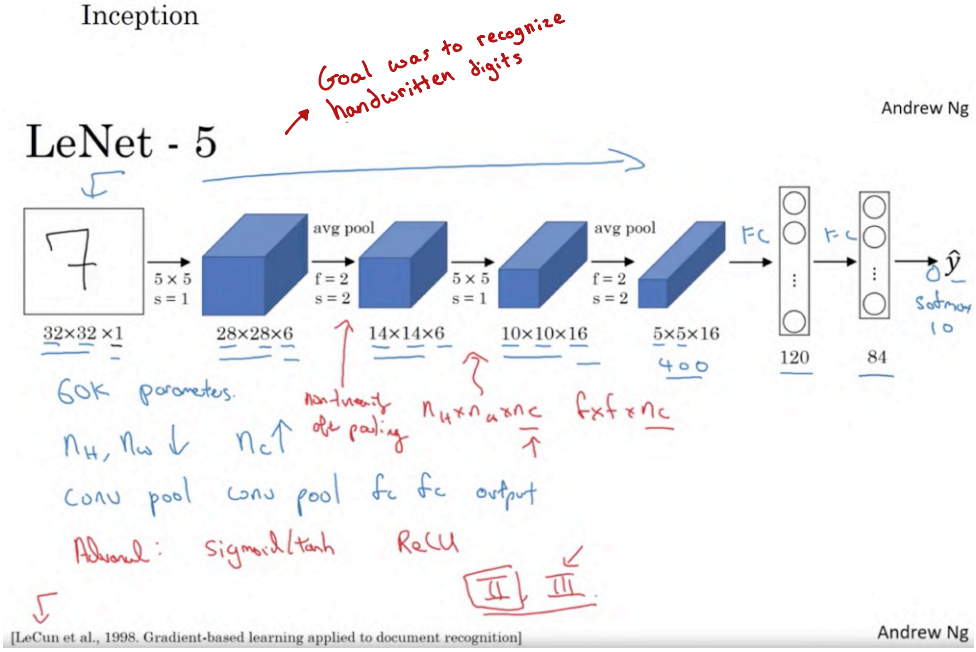
Outline

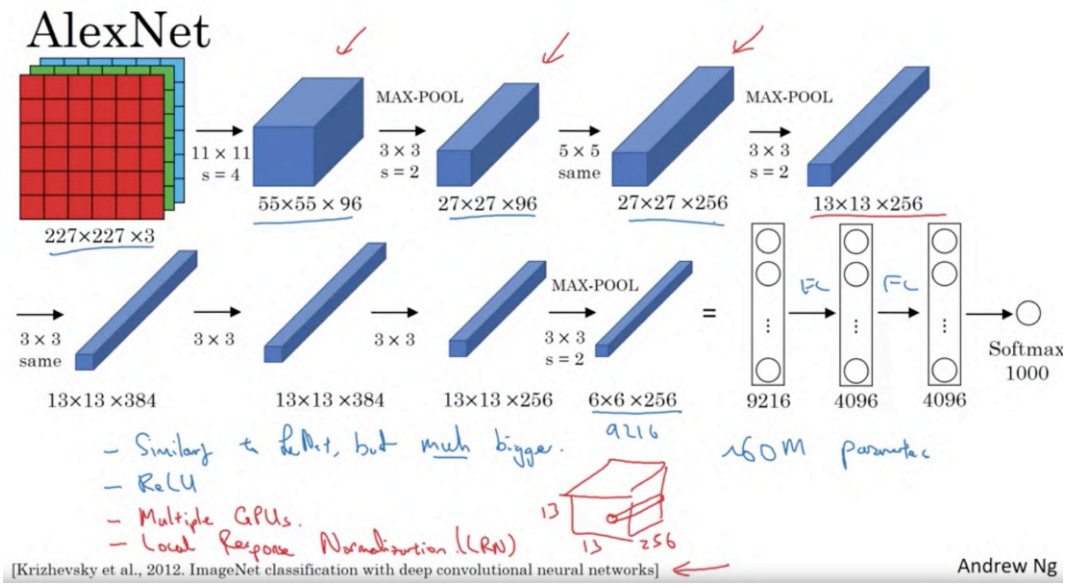
Classic networks:

- LeNet-5
- AlexNet
- VGG

ResNet (152)

Inception

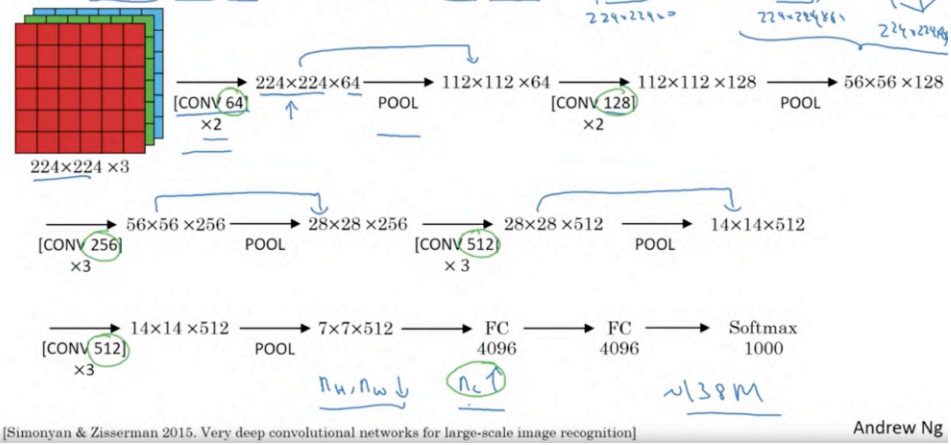




VGG - 16

VGG-16

CONV = 3×3 filter, s = 1, same MAX-POOL = 2×2 , s = 2

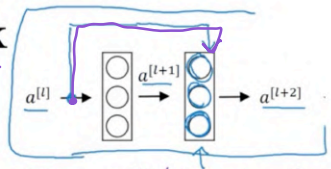


$n_H, n_W \downarrow$ $n_c \uparrow$ ~138M

[Simonyan & Zisserman 2015. Very deep convolutional networks for large-scale image recognition]

Andrew Ng

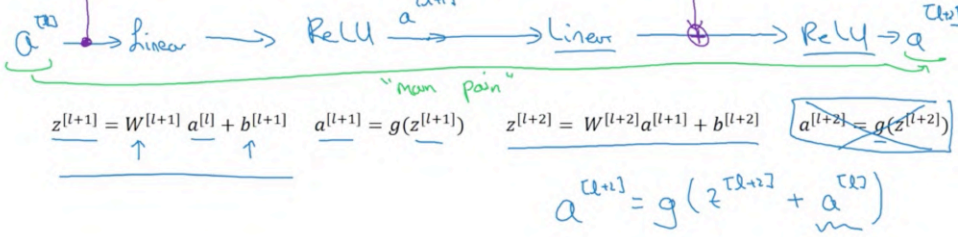
Residual block



"Short cut" / skip connection

"skip over layer"

Using residual blocks allows you to train much deeper neural networks. And the way you build a ResNet is by taking many of these residual blocks, blocks like these, and stacking them together to form a deep network



$z^{[l+1]} = W^{[l+1]} a^{[l]} + b^{[l+1]}$ $a^{[l+1]} = g(z^{[l+1]})$ $z^{[l+2]} = W^{[l+2]} a^{[l+1]} + b^{[l+2]}$ ~~$a^{[l+2]} = g(z^{[l+2]})$~~

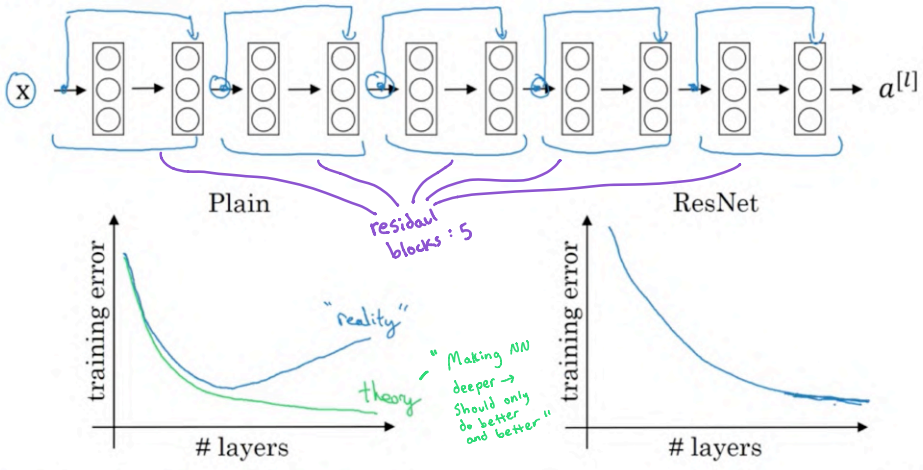
$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$

[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng

Residual Network

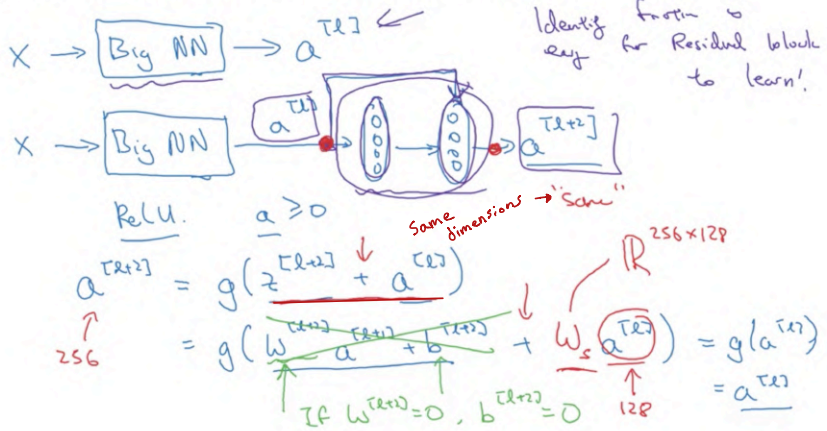
"Plain network"



[He et al., 2015. Deep residual networks for image recognition]

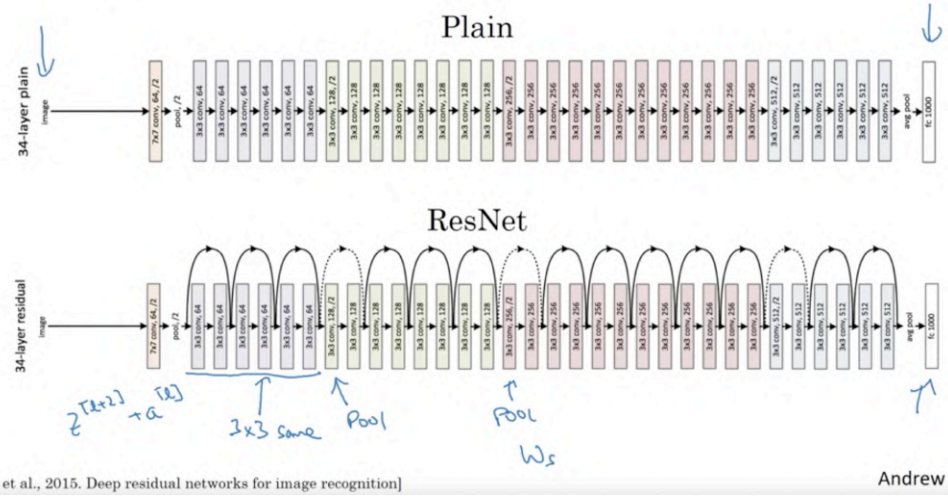
Andrew Ng

Why do residual networks work?



Andrew Ng

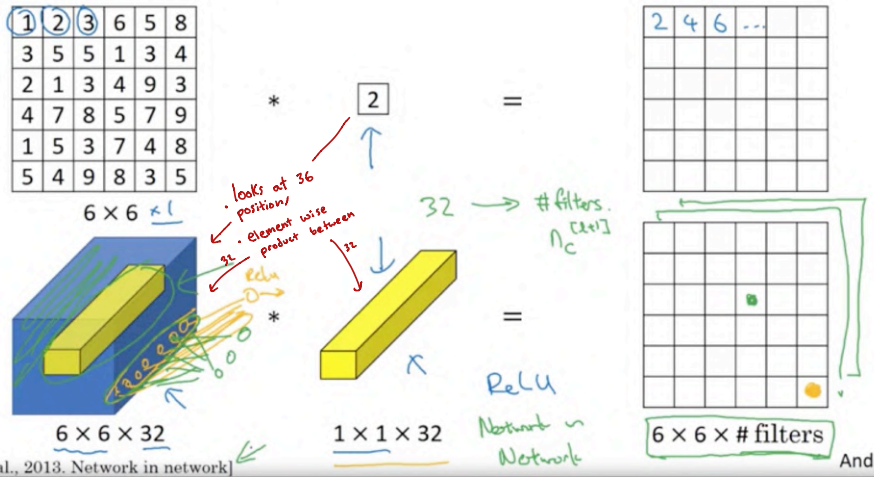
ResNet



[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng

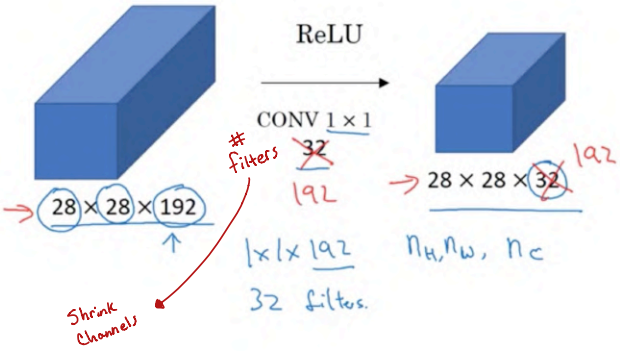
Why does a 1 × 1 convolution do?



[Lin et al., 2013. Network in network]

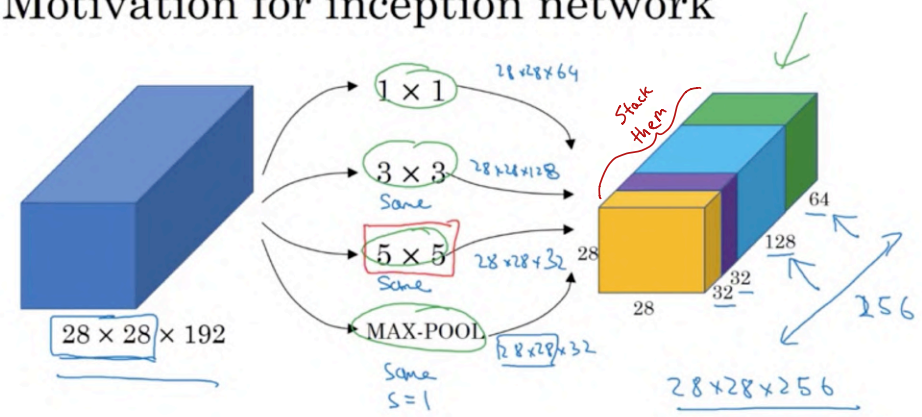
Andrew Ng

Using 1×1 convolutions



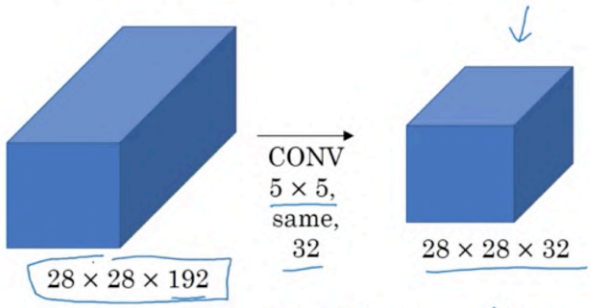
[Lin et al., 2013. Network in network]

Motivation for inception network



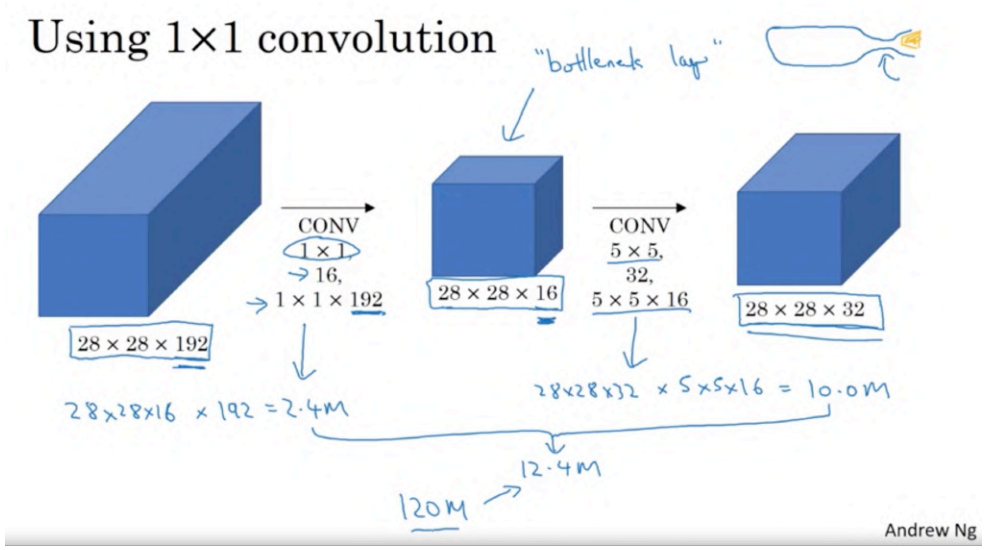
[Szegedy et al. 2014. Going deeper with convolutions]

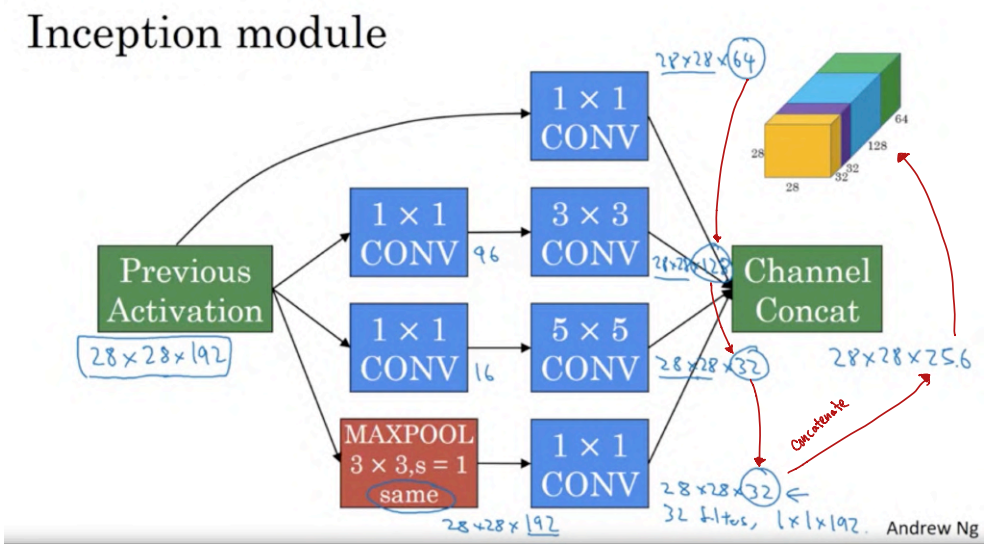
The problem of computational cost

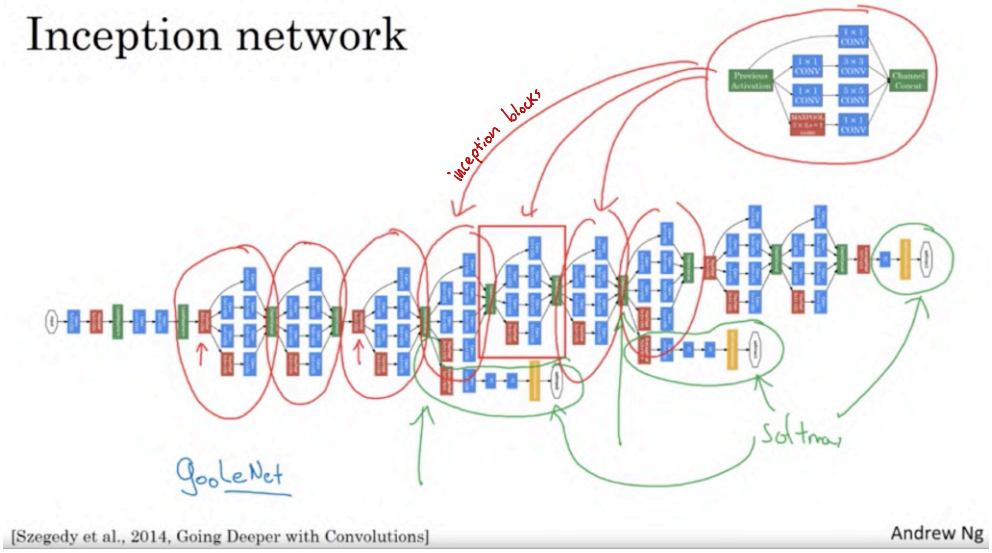


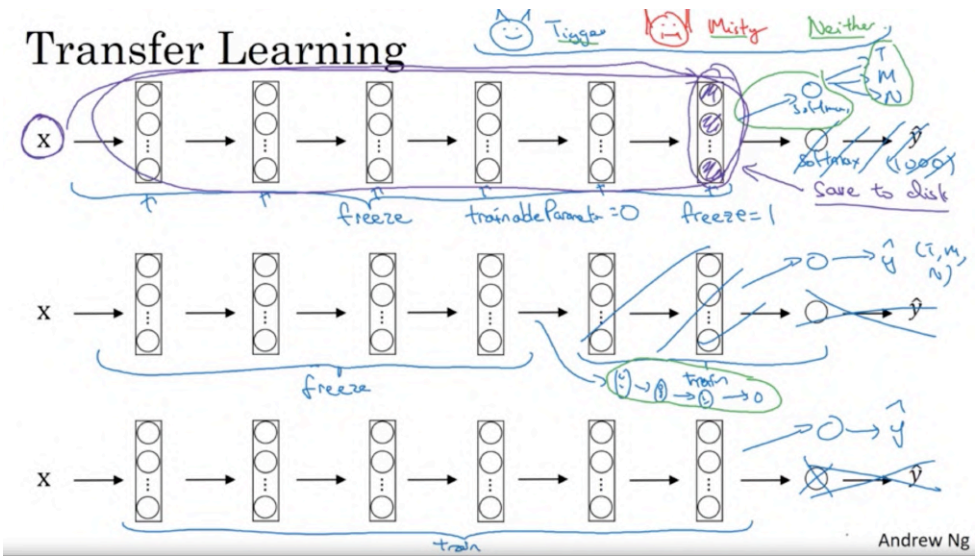
32 filters. filter one 5×5×192.

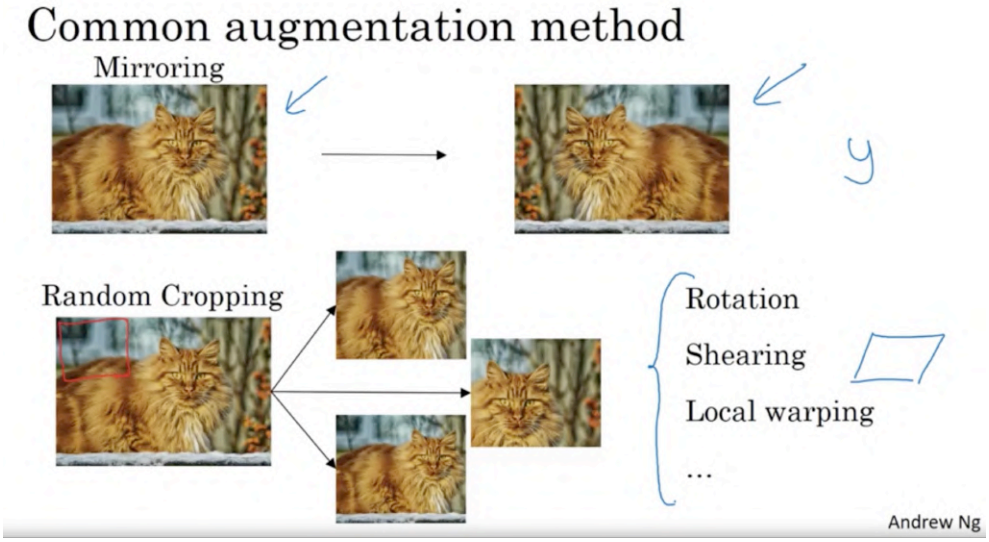
28×28×32 × 5×5×192 = 120M.


Using 1×1 convolution
"bottleneck layer"
28 × 28 × 192
CONV
1 × 1,
16,
1 × 1 × 192
28 × 28 × 16
CONV
5 × 5,
32,
5 × 5 × 16
28 × 28 × 32
28×28×16 × 192 = 2.4M
28×28×32 × 5×5×16 = 10.0M
12.4M
120M
Andrew Ng


Inception module
Previous Activation
28×28×192
1 × 1 CONV
1 × 1 CONV
96
3 × 3 CONV
1 × 1 CONV
16
5 × 5 CONV
MAXPOOL 3 × 3,s = 1 same
1 × 1 CONV
28×28×192
28×28×64
28×28×128
28×28×32
28×28×32
32 filters, 1×1×192.
28×28×256
Channel Concat
concatenate
Andrew Ng

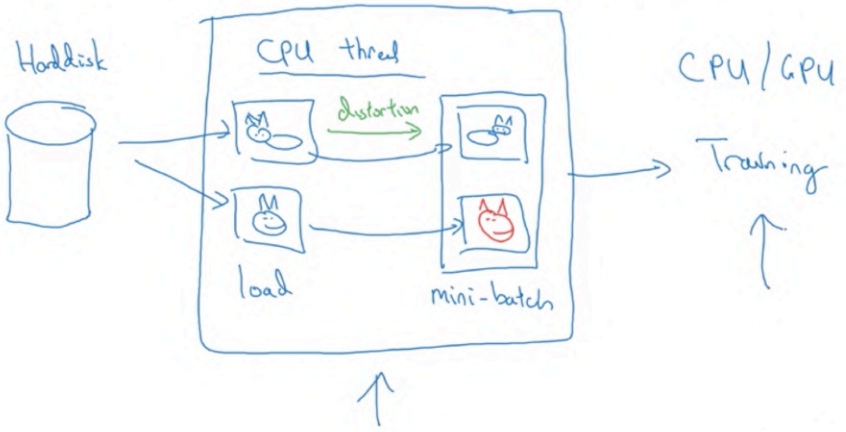

Inception network
inception blocks
GoogLeNet
Softmax
[Szegedy et al., 2014, Going Deeper with Convolutions]
Andrew Ng


Transfer Learning
Tigger
Misty
Neither
x
Softmax
Save to disk
freeze
trainableParameters = 0
freeze = 1
x
freeze
x
train
Andrew Ng


Common augmentation method
Mirroring
Random Cropping
Rotation
Shearing
Local warping
...
Andrew Ng

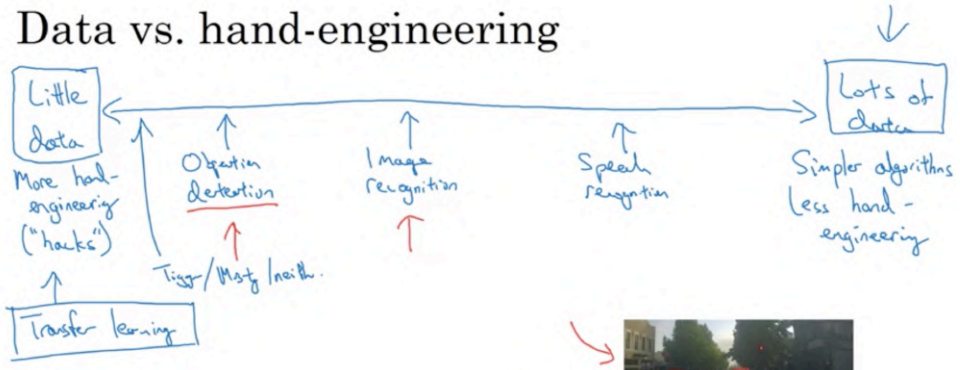

Color shifting
R G B
+20,-20,+20
-20,+20,+20
+5,0,+50
More robust to colors
Advanced:
PCA
ml-class.org
AlexNet paper
"PCA color augmentation."
R B G
Andrew Ng

Implementing distortions during training



Andrew Ng

Data vs. hand-engineering



Two sources of knowledge

- Labeled data (x, y)
- Hand engineered features/network architecture/other components

Andrew Ng

Tips for doing well on benchmarks/winning competitions

Ensembling

- Train several networks independently and average their outputs

3-15 networks → y (Average y, not weights)

Multi-crop at test time

- Run classifier on multiple versions of test images and average results

10-crop

Andrew Ng

Use open source code

- Use architectures of networks published in the literature
- Use open source implementations if possible
- Use pretrained models and fine-tune on your dataset