deeplearning.ai

Case Studies

Why look at case studies?

Outline

Classic networks:

- LeNet-5 ←
- AlexNet ←
- VGG ←

ResNet (152)

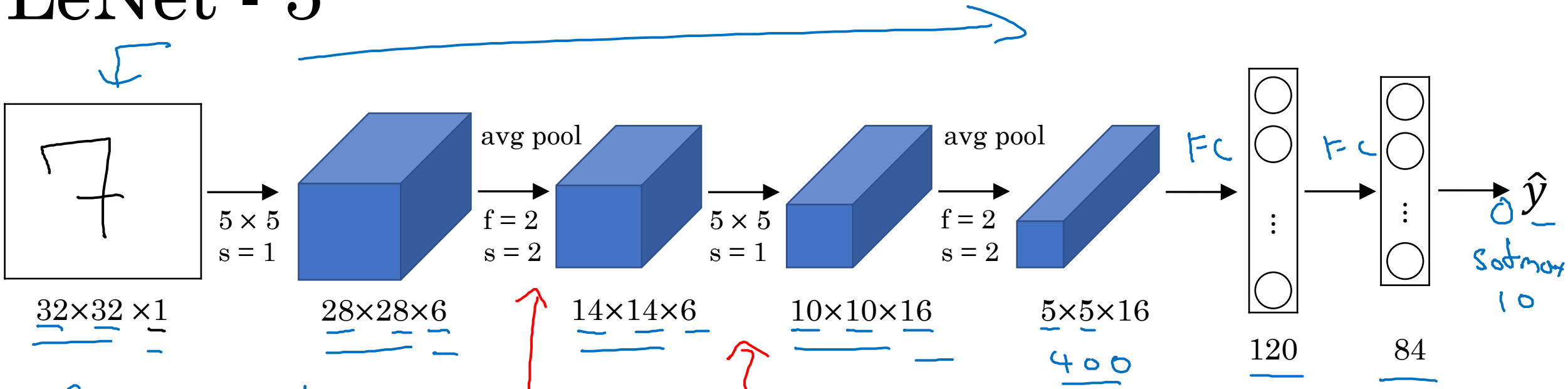
Inception

deeplearning.ai

Case Studies

Classic networks

LeNet - 5



60k parameters.

$n_H, n_w \downarrow$ $n_c \uparrow$

Non-linearity after pooling

$n_{H} \times n_{w} \times n_c$ $f \times f \times n_c$

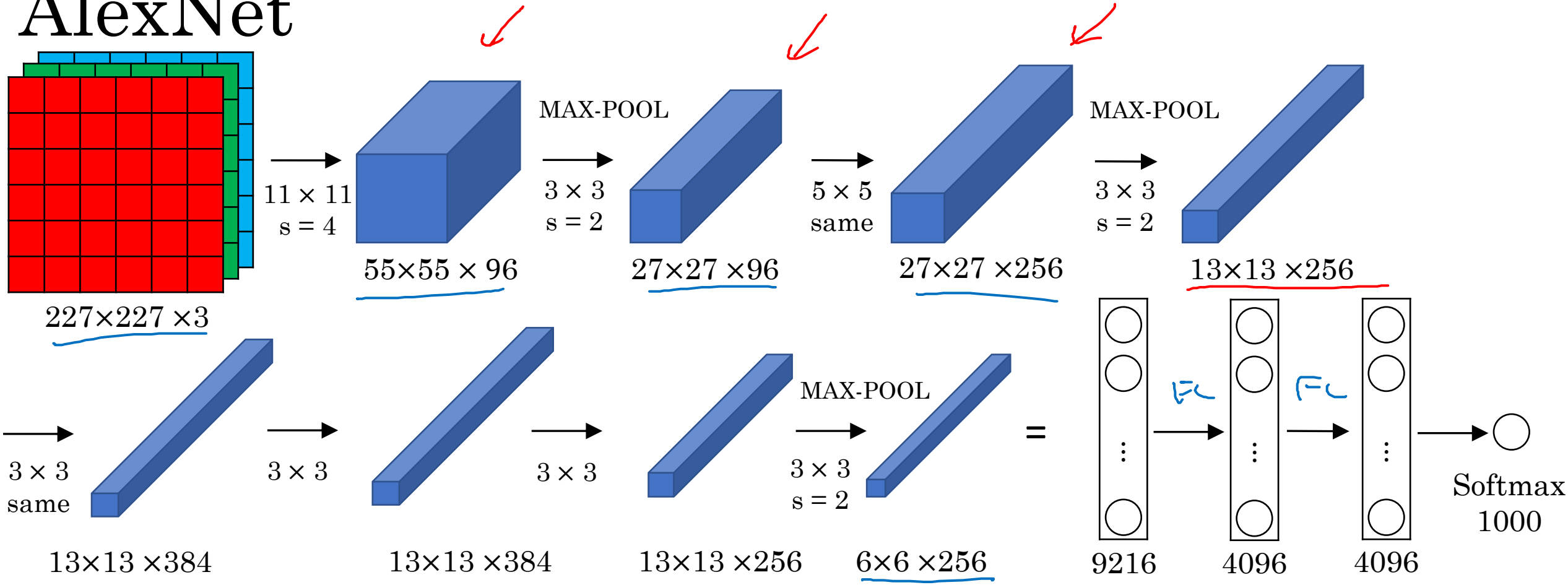
conv pool conv pool fc fc output

Activation: sigmoid/tanh ReLU

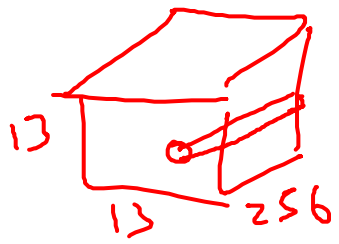
II III

[LeCun et al., 1998. Gradient-based learning applied to document recognition]

AlexNet



- Similar to LeNet, but much bigger. 9216 60M parameters
- ReLU
- Multiple GPUs.
- Local Response Normalization (LRN)

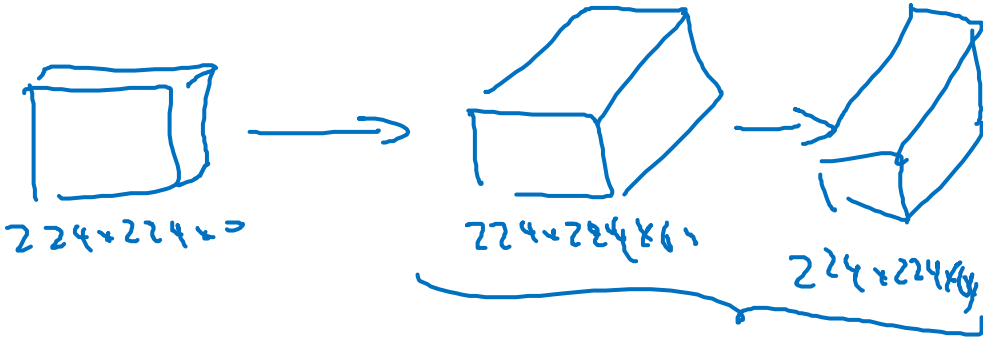


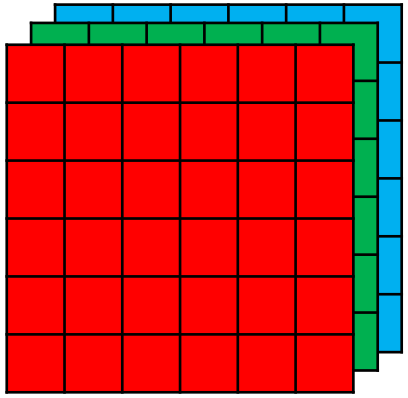
[Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks]

VGG - 16

VGG-19

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





224×224 ×3

→ [CONV 64] ×2 → 224×224×64 → POOL → 112×112 ×64 → [CONV 128] ×2 → 112×112 ×128 → POOL → 56×56 ×128

“ x2 ”指两个SAME的ConV层

→ [CONV 256] ×3 → 56×56 ×256 → POOL → 28×28 ×256 → [CONV 512] × 3 → 28×28 ×512 → POOL → 14×14×512

→ [CONV 512] ×3 → 14×14 ×512 → POOL → 7×7×512 → FC 4096 → FC 4096 → Softmax 1000

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

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

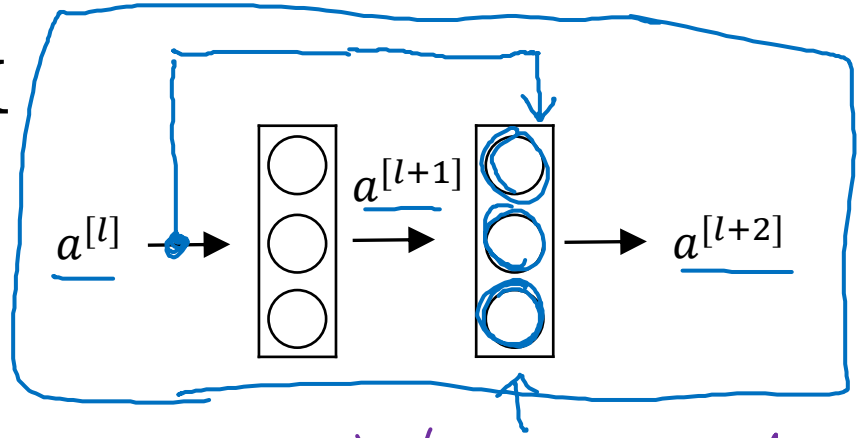
Andrew Ng

deeplearning.ai

Case Studies

Residual Networks (ResNets)

Residual block



"short cut" / skip connection

$a^{[l]}$ → Linear → ReLU → $a^{[l+1]}$ → Linear → (+) → ReLU → $a^{[l+2]}$

"main path"

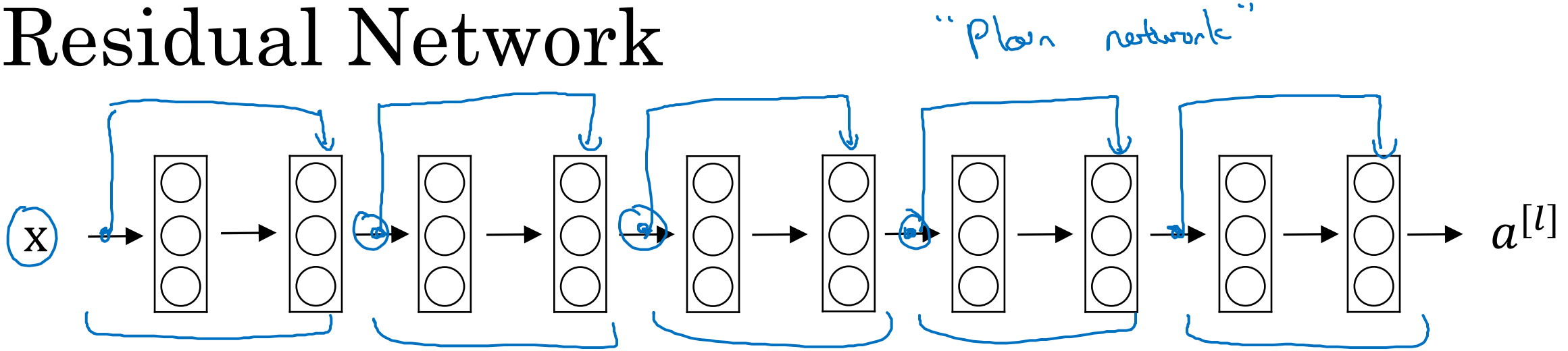
$$z^{[l+1]} = W^{[l+1]} a^{[l]} + b^{[l+1]} \qquad a^{[l+1]} = g(z^{[l+1]}) \qquad z^{[l+2]} = W^{[l+2]} a^{[l+1]} + b^{[l+2]} \qquad a^{[l+2]} = g(z^{[l+2]})$$

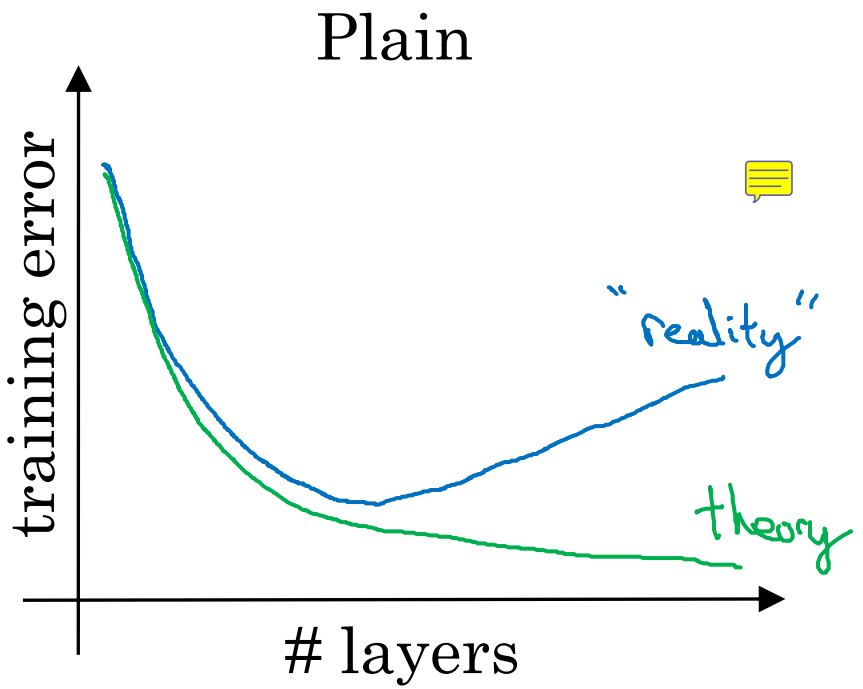
$$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$$

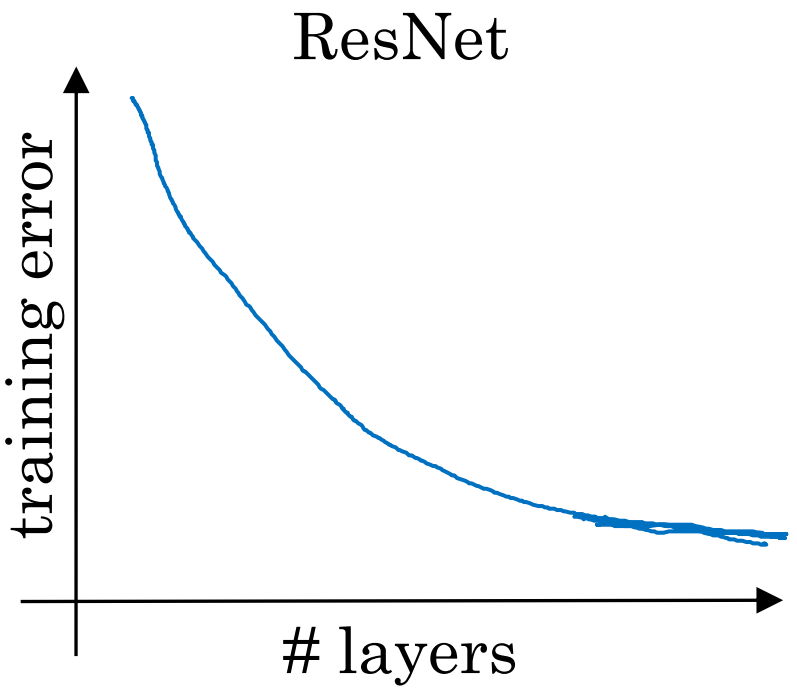
[He et al., 2015. Deep residual networks for image recognition]

Andrew Ng

Residual Network







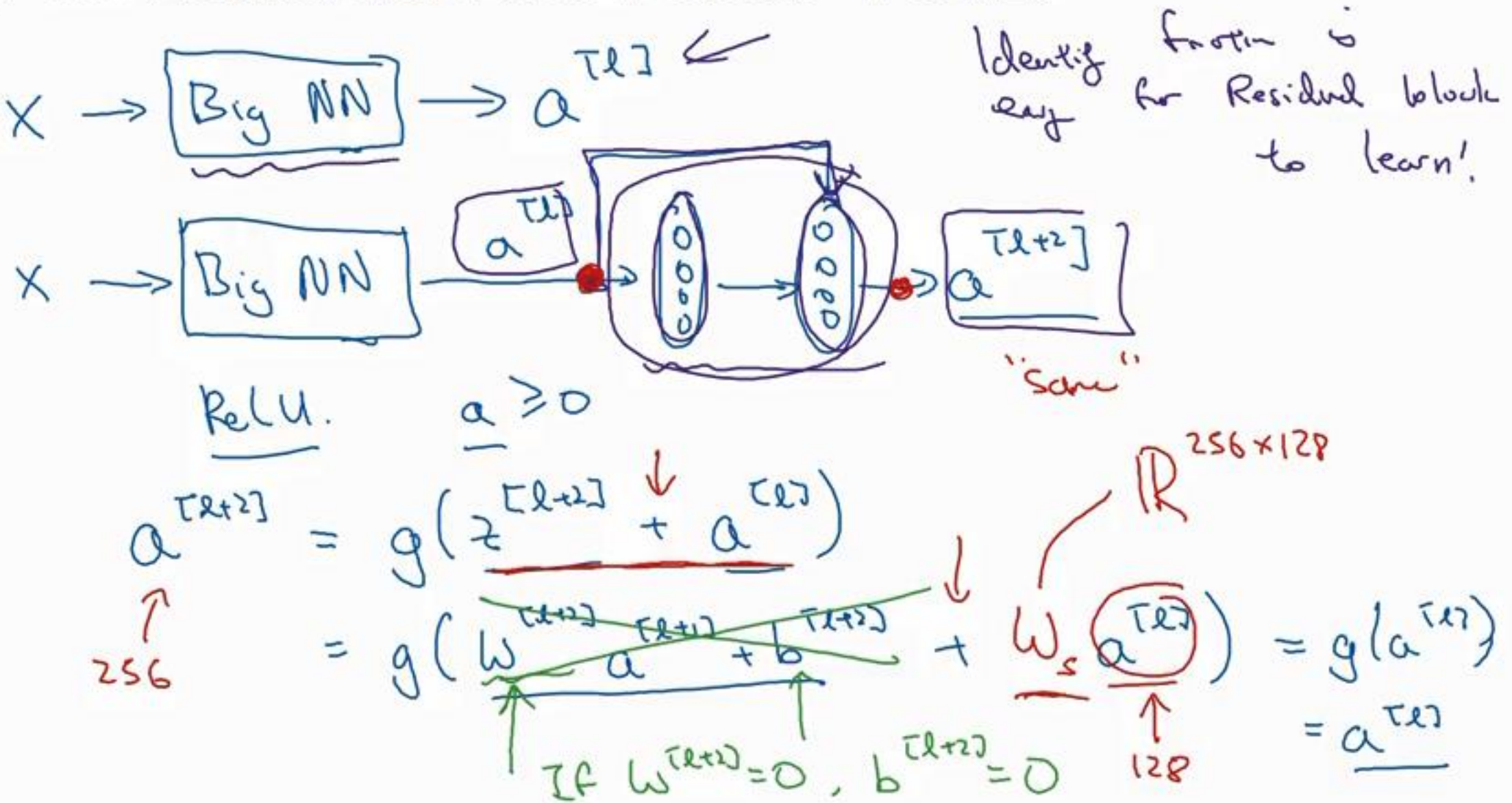
[He et al., 2015. Deep residual networks for image recognition]

deeplearning.ai

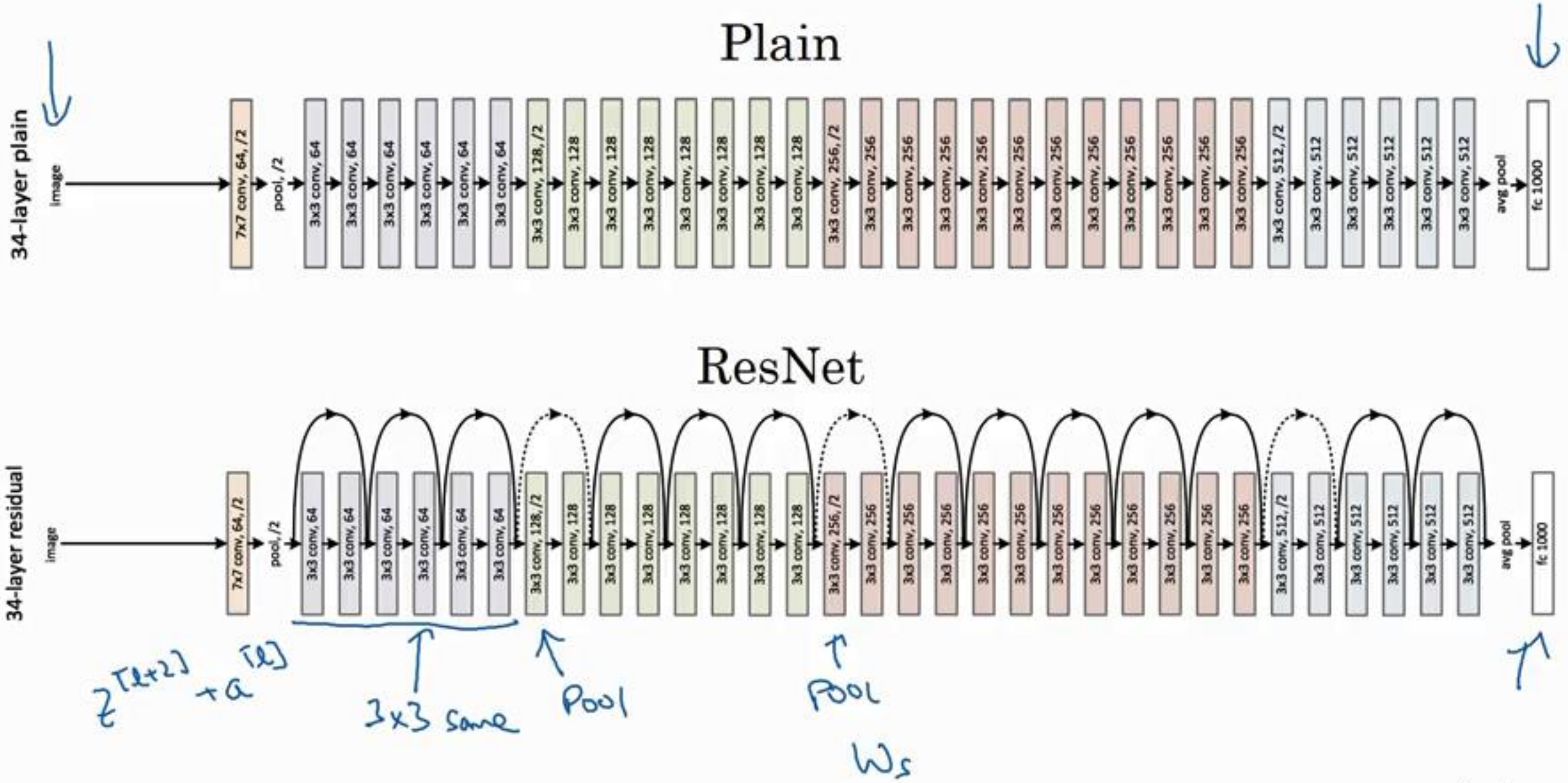
Case Studies

Why ResNets work

Why do residual networks work?



ResNet



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

Andrew Ng

deeplearning.ai

Case Studies

Network in Network and 1×1 convolutions

Why does a 1 × 1 convolution do?

1	2	3	6	5	8
3	5	5	1	3	4
2	1	3	4	9	3
4	7	8	5	7	9
1	5	3	7	4	8
5	4	9	8	3	5

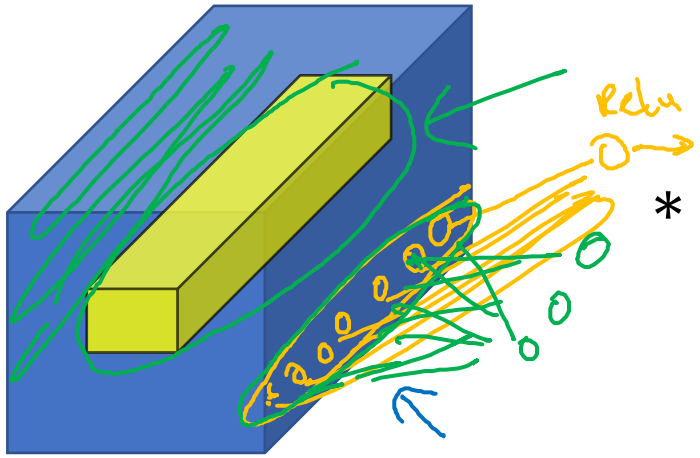
6 × 6 × 1

*

2

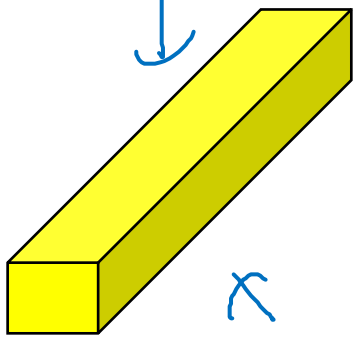
=

2	4	6	...		



6 × 6 × 32

*

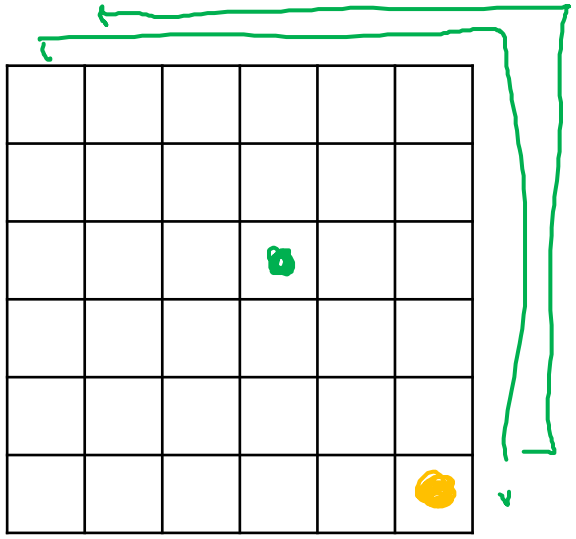


1 × 1 × 32

32 → # filters.

$n_c^{[l+1]}$

=



6 × 6 × # filters

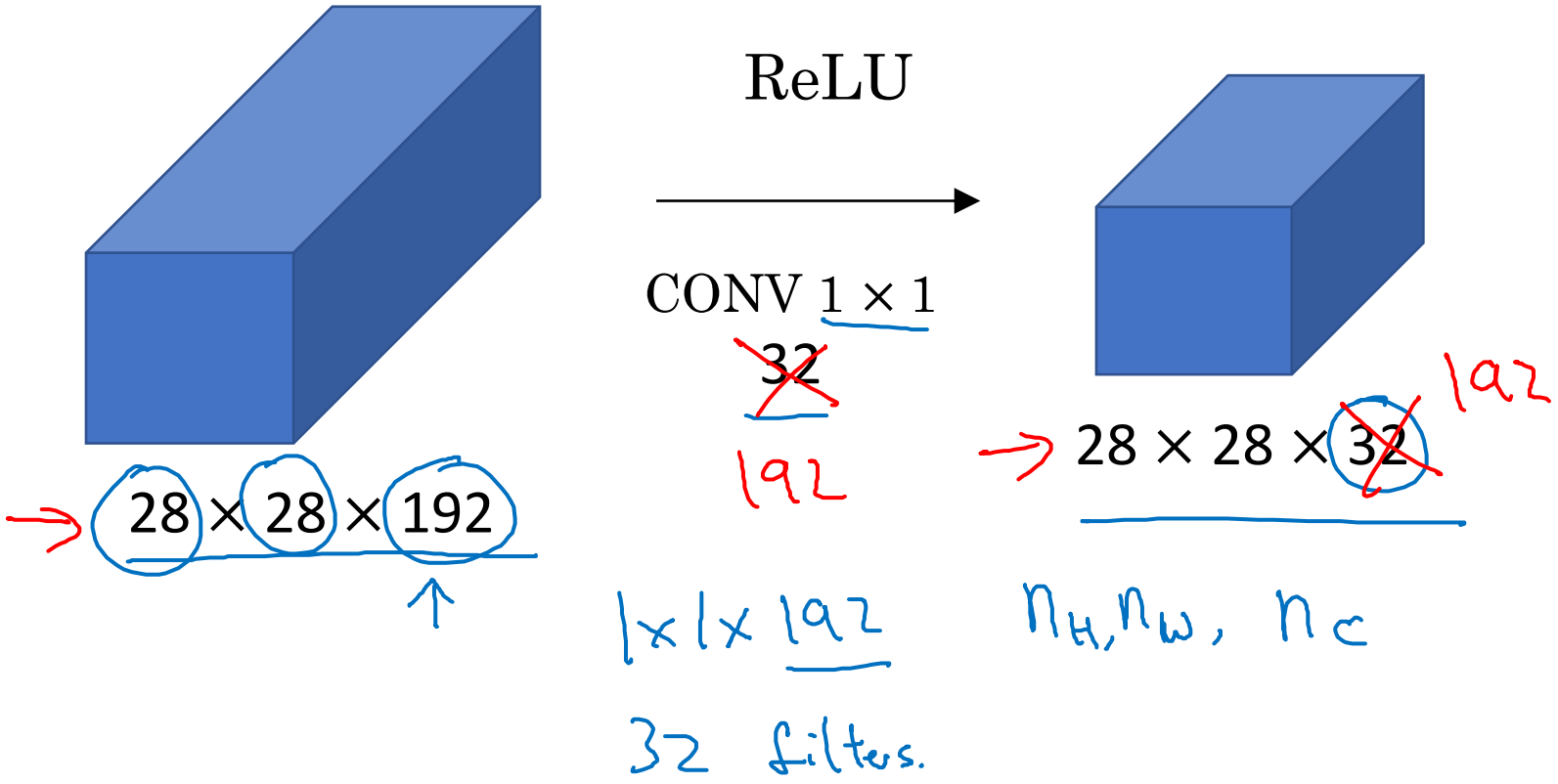
ReLU

Network in Network

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

Andrew Ng

Using 1×1 convolutions



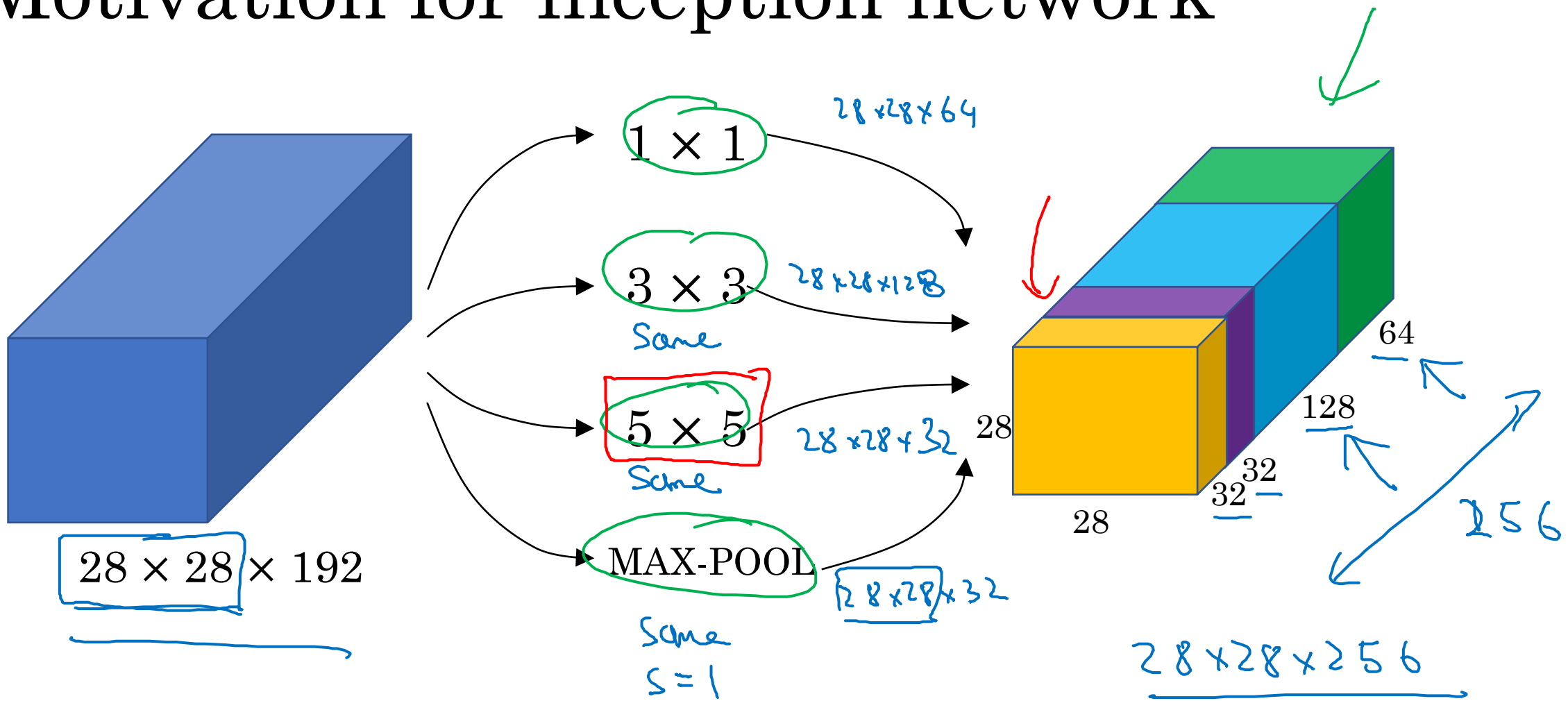
[Lin et al., 2013. Network in network]

deeplearning.ai

Case Studies

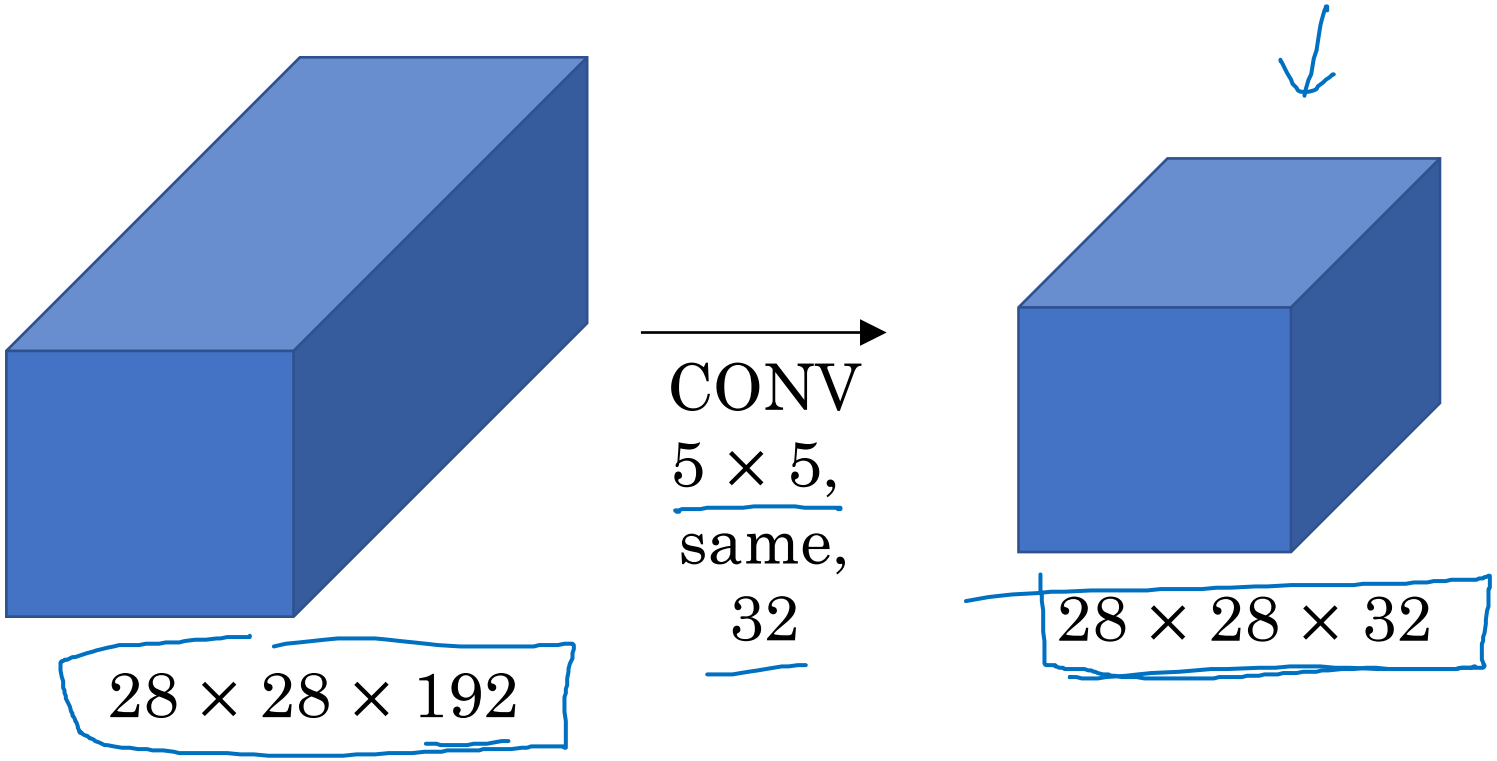
Inception network motivation

Motivation for inception network



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

The problem of computational cost

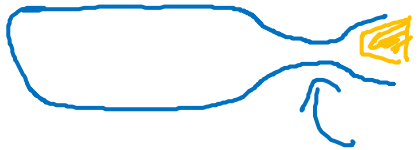


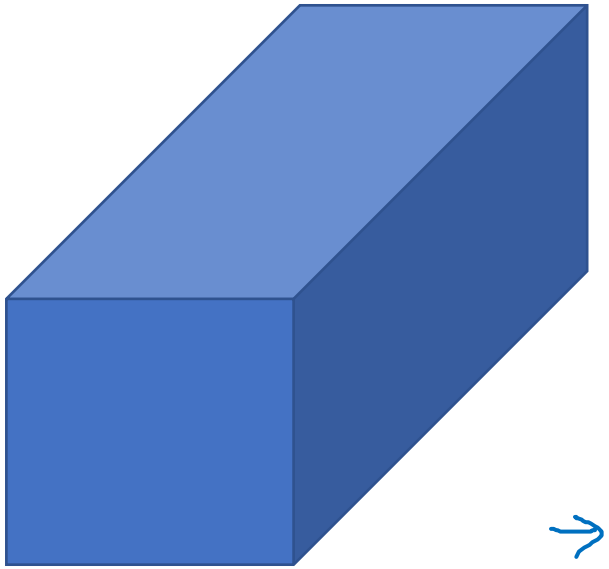
32 filters. filter one 5×5×192.

28×28×32 × 5×5×192 = 120M.

Using 1×1 convolution

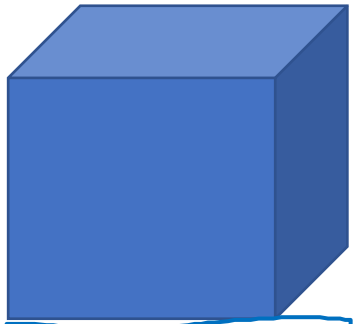
"bottleneck layer"





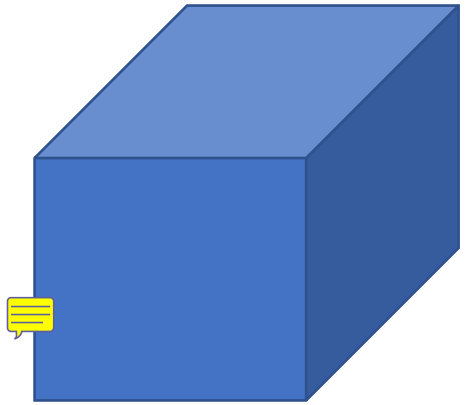
28 × 28 × 192

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

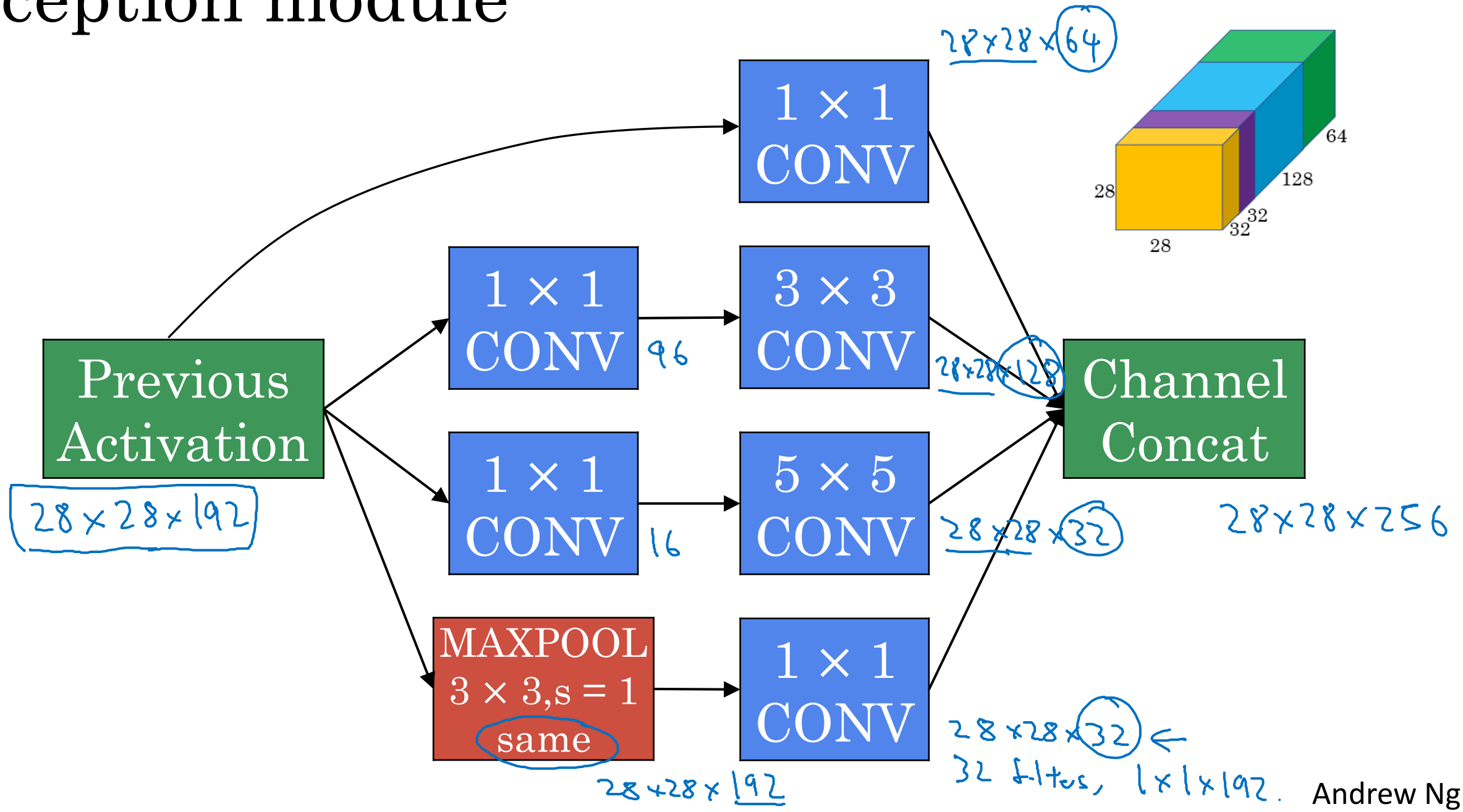
120M

deeplearning.ai

Case Studies

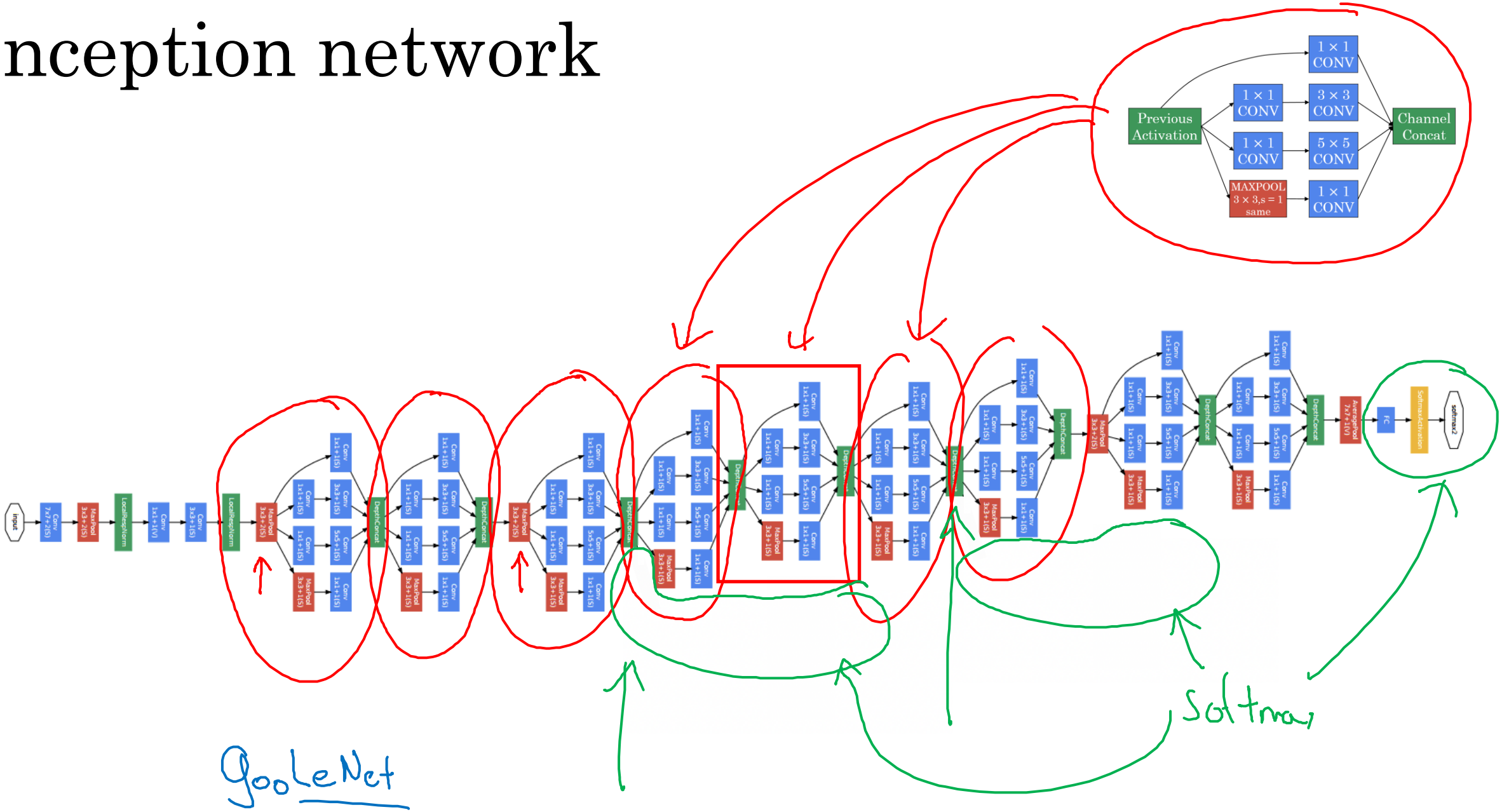
Inception network

Inception module



Andrew Ng

Inception network



[Szegedy et al., 2014, Going Deeper with Convolutions]

http://knowyourmeme.com/memes/we-need-to-go-deeper

deeplearning.ai

Practical advice for using ConvNets

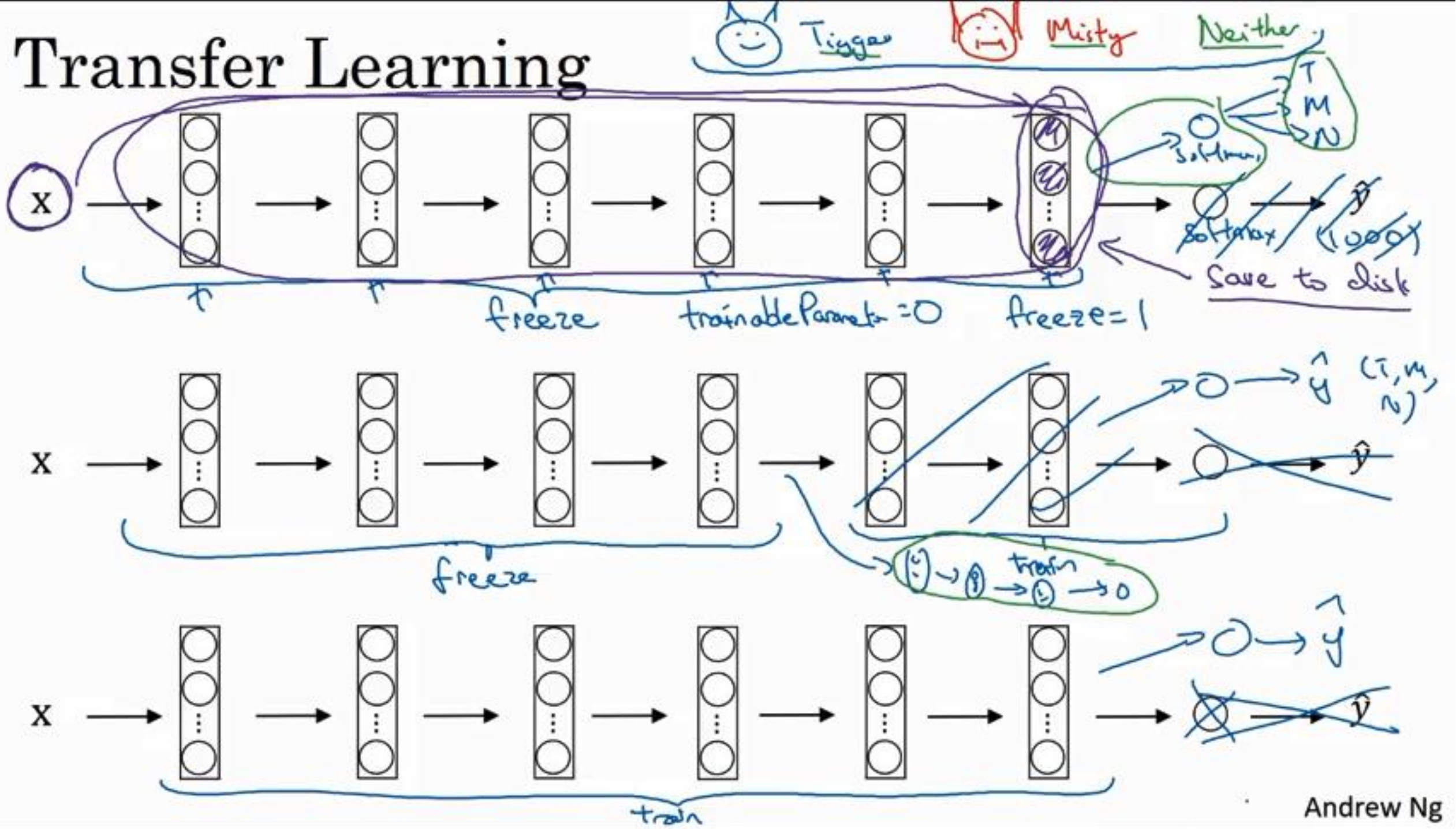
Using open-source implementations

deeplearning.ai

Practical advice for using ConvNets

Transfer Learning

Transfer Learning



Andrew Ng

deeplearning.ai

Practical advice for using ConvNets

Data augmentation

Common augmentation method

Mirroring

Random Cropping

- Rotation
- Shearing
- Local warping
- ...

Color shifting

R G B

+20,-20,+20

-20,+20,+20

+5,0,+50

y

Advanced:

PCA

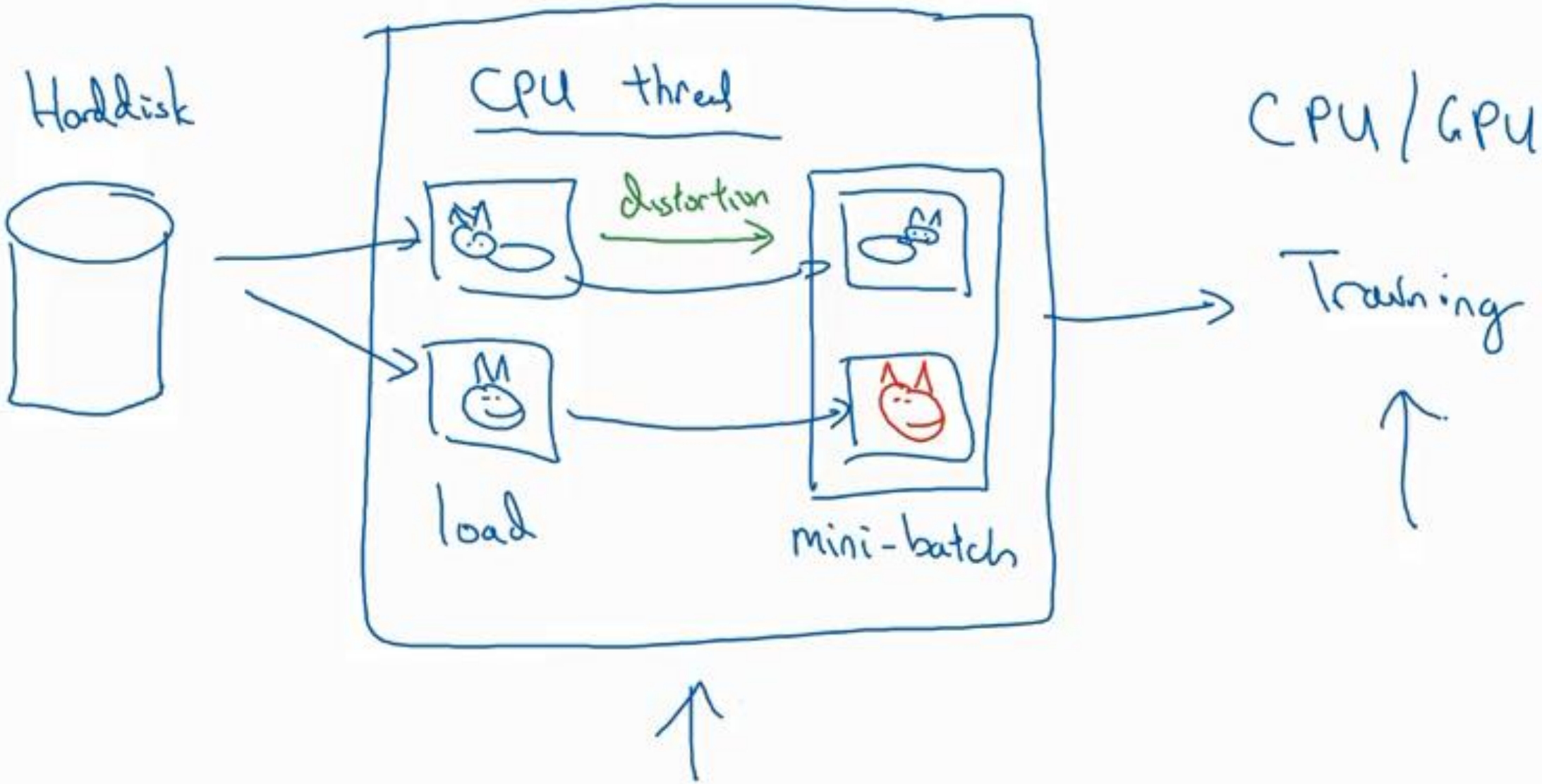
ml-class.org

[AlexNet paper

["PCA color augmentation."

R B G

Implementing distortions during training

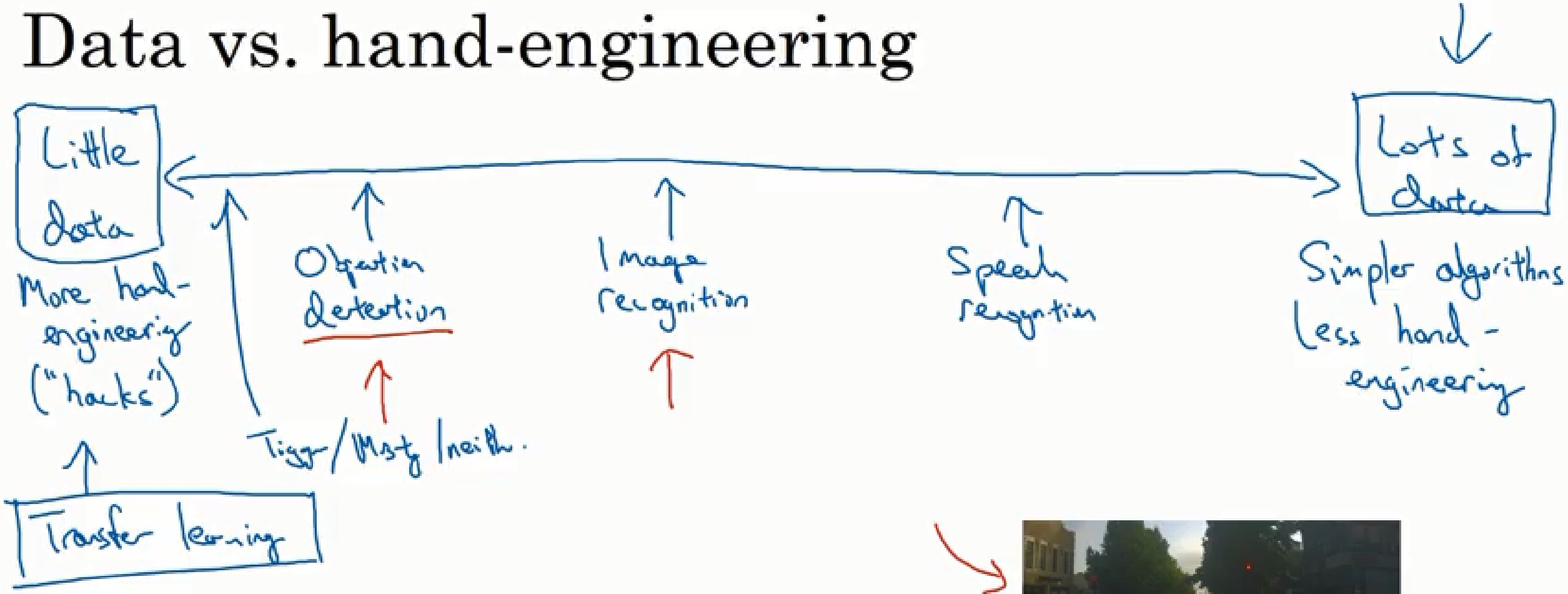


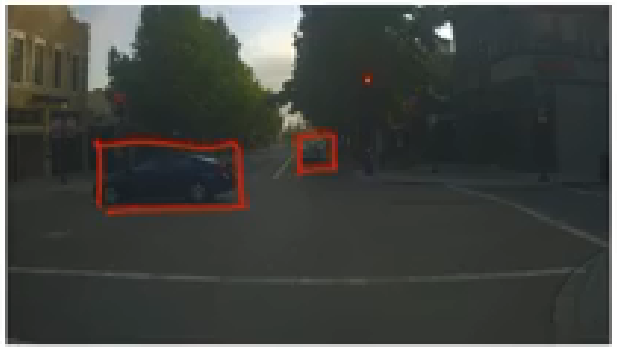
deeplearning.ai

Practical advice for using ConvNets

The state of computer vision

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

3-15 networks

- Train several networks independently and average their outputs

→ $\hat{y}$

Multi-crop at test time

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

10-crop

1 + 4 + 1 + 4

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