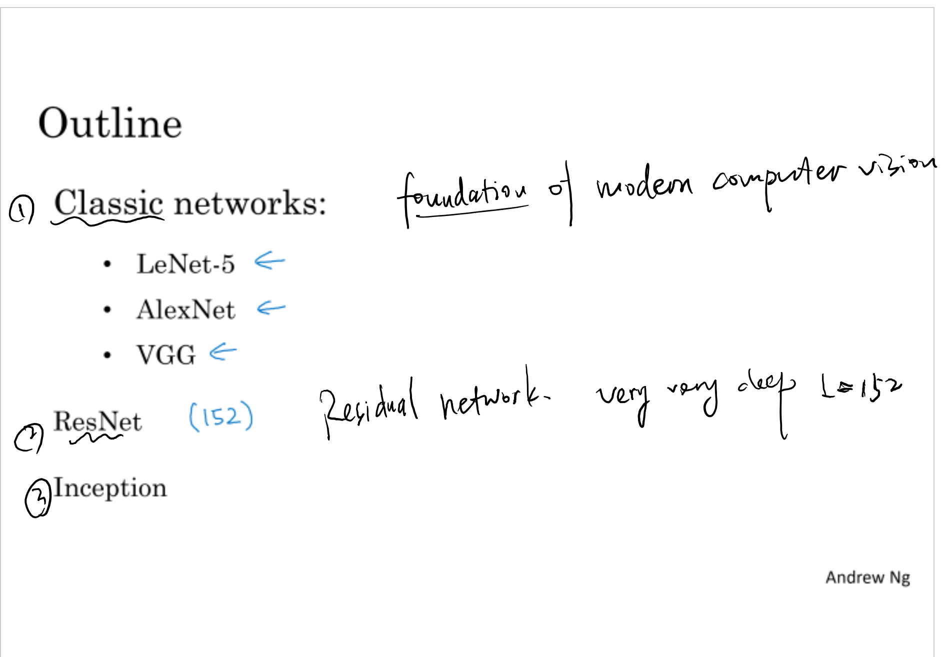
Case studies of CNN

Monday, June 11, 2018 7:59 AM

why case studies

CNN examples

transferable among tasks



Ideas from these examples can be applied in disciplines outside CV.

classic cnn

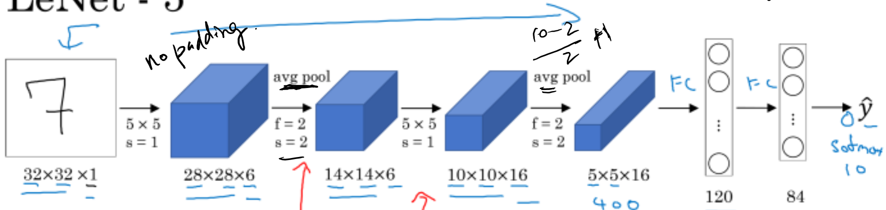
deeplearning.ai

Case Studies

Classic networks

LeNet - 5

can change avg to max pooling in own application



* when this is developed, people don't pad. So size keep shrinking

60K parameters.

$n_H, n_W \downarrow$ $n_C \uparrow$

conv pool conv pool fc fc output

Advanced: sigmoid/tanh

nonlinearity after pooling

didn't use ReLU

$n_H \times n_W \times n_C$

$f \times f \times n_C$ ← now, every filter looks at all channels

Back then one filter focus on one channel. (due to computational limit).

What's in red: technical details of the paper.

II III

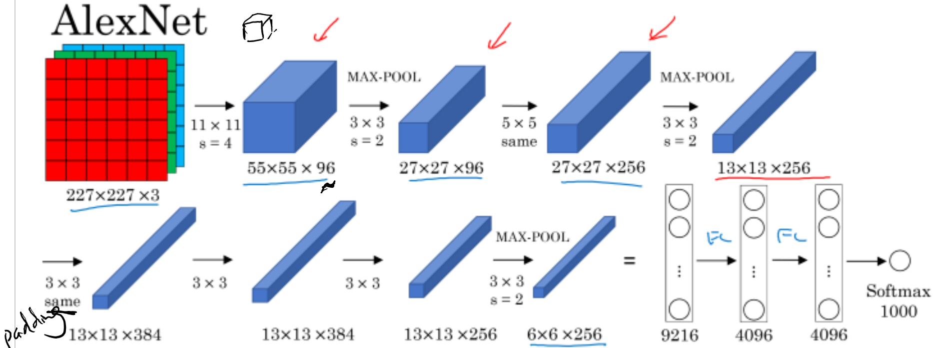
[LeCun et al., 1998. Gradient-based learning applied to document recognition] ← hard-to-read paper

Andrew Ng

Read Section 2 and (maybe) Section 3

AlexNet

← this paper convinces the computer vision community that DL works in CV task

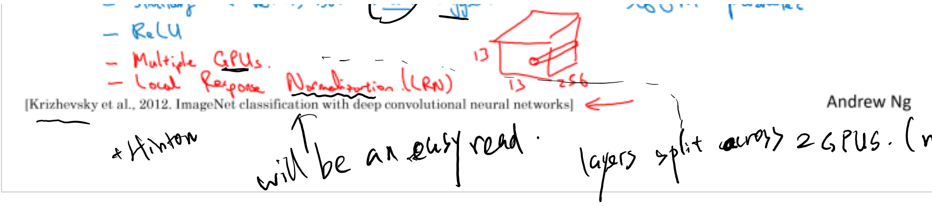


red text: details in the paper.

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

60 million parameters

Andrew Ng

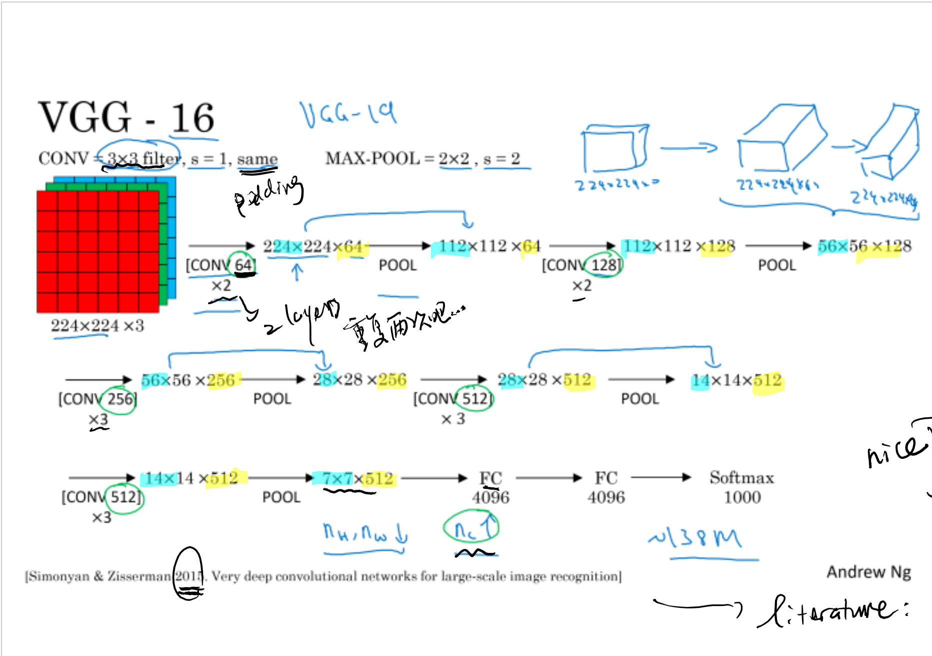


60 million parameters

+Hinton

will be an easy read.

layers split across 2 GPUs. (model developed when GPU is slow)



所以第一层到后是 3×3×3 呢 不是是3吗

Why Ng thinks this model is nice: Depth ↑ . H, W ↓

→ literature: VGG19 more complicated

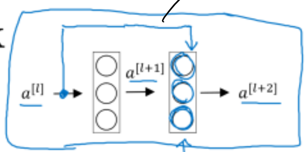
VGG16 ≈ VGG19 performancing

resnet

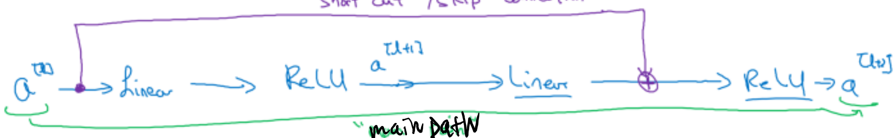
Very deep NN has vanishing or exploding gradient problem.

Residual block

an extra short cut added before the ReLU non-linearity



"short cut" / skip connection



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

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

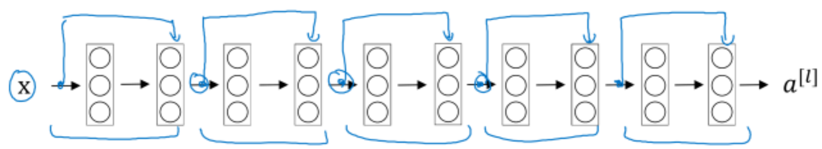
Using Residual Blocks allow you to training much deeper neural network

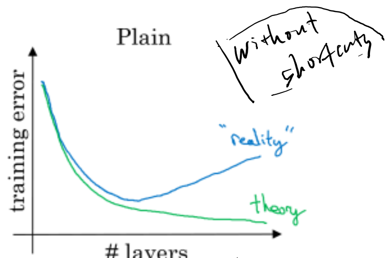
[He et al., 2015. Deep residual networks for image recognition]

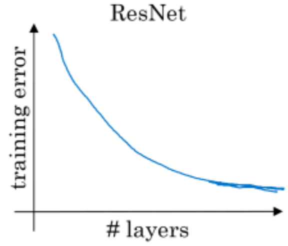
Andrew Ng

Residual Network

"Plain network"







Deeper network makes performance worse

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

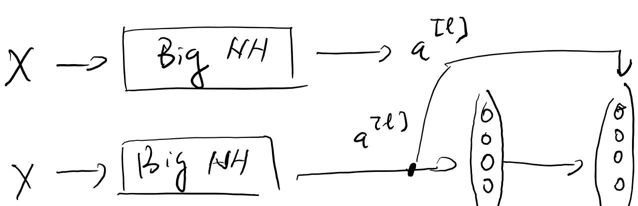
Andrew Ng

Why ResNet work

Doing well in the training is a good first step.

Does it?

ResNet "same" convolution preserve dimension ⇒ easy to operate



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

- Identity function is easy for residual block to learn.

⇒ adding residual block somewhere in the NN does not hurt performance

$R^{256 \times 128}$ ← adjust for dimension, in case different

$$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$$
256

$$= g(\underline{W^{[l+2]}} a^{[l+1]} + b^{[l+2]} + W_s a^{[l]})$$
(128)

$\mathbb{R}^{256 \times 128}$ ← adjust for dimension, in case different

$$\underline{\text{if } W^{[l+2]} = 0 \quad b^{[l+2]} = 0} \quad g(a^{[l]}) = a^{[l]}$$

When you make NN deeper => makes it harder and harder to learn parameters

Adding residual ——> Never hurt, Sometimes even help.

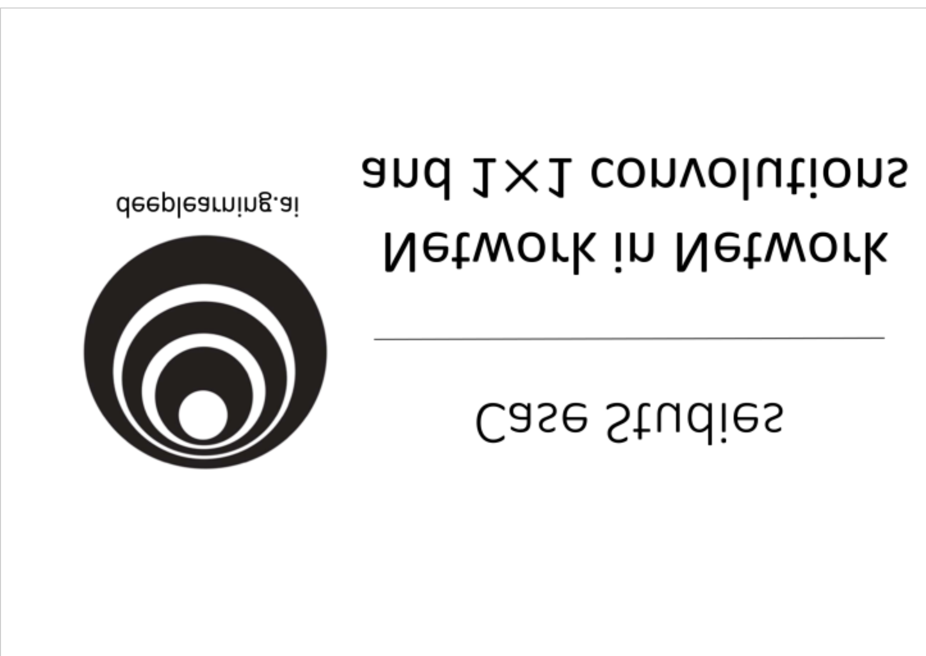
<u>Resnet on images</u> "same" convolution.

Input : image

Add skip connections , <u>pooling layers too</u>.

Use matrix W_s to adjust dimensions

conv1by1



Used to change n_c — # of channels

=> can be used to build inception network

Fully connected network that applies for each of the $n_H \times n_W$ positions.

压缩第3个 dimension 的作用么? — 是为了 capture channel 的特征

technically, each filter $1 \times 1 \times 192$ match # of ~~chanel~~ channel

1x1 convolution allows you to change the # of channels in your volumes.

helps to shrink the # of channels

inception

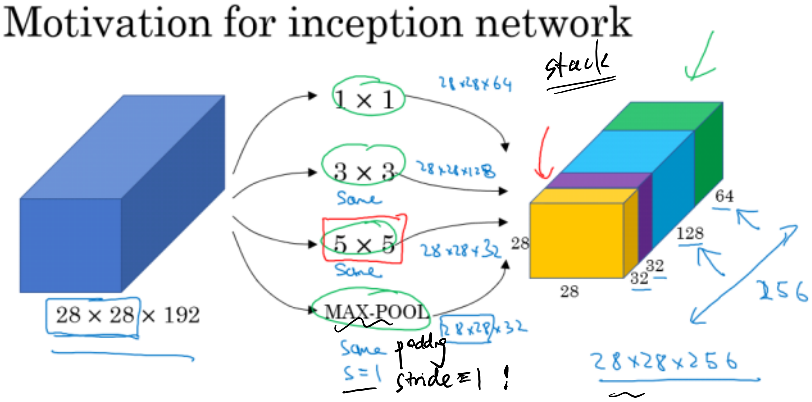
Case Studies

deeplearning.ai

Inception network motivation

Inception layer

Motivation for inception network



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

Andrew Ng

stack output that apply the input with filters of different sizes.

- use "same" ~~cost~~ convolution to keep dimension the same

the heart of the inception network

problem with the inception layer.

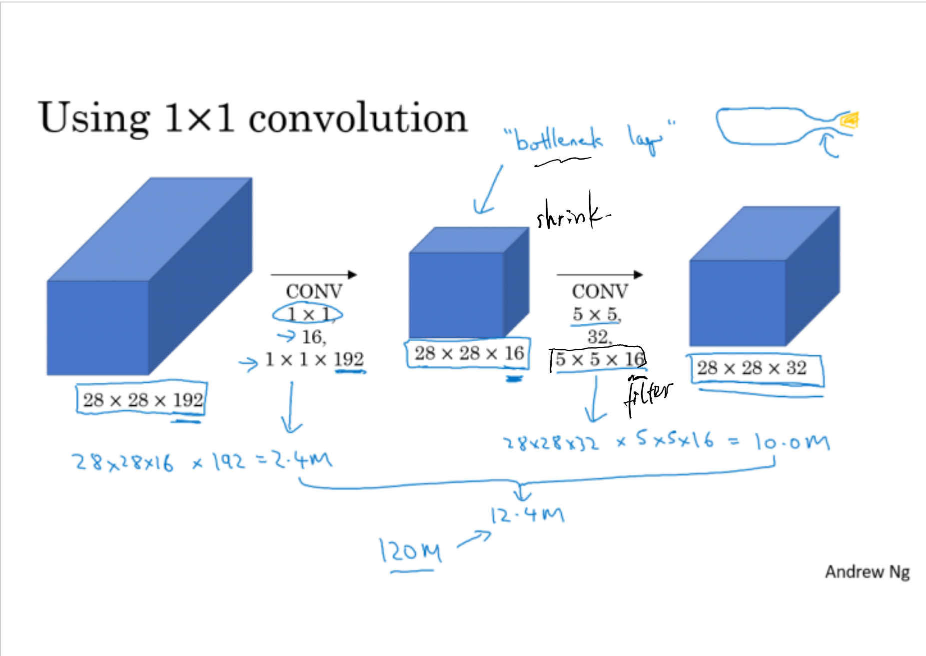
Computational cost.

Just focus on the 5×5 same conv.

120 million parameters!

use 1×1 convolution to reduce

use 1x1 convolution to reduce # of parameters.



Bottleneck layer

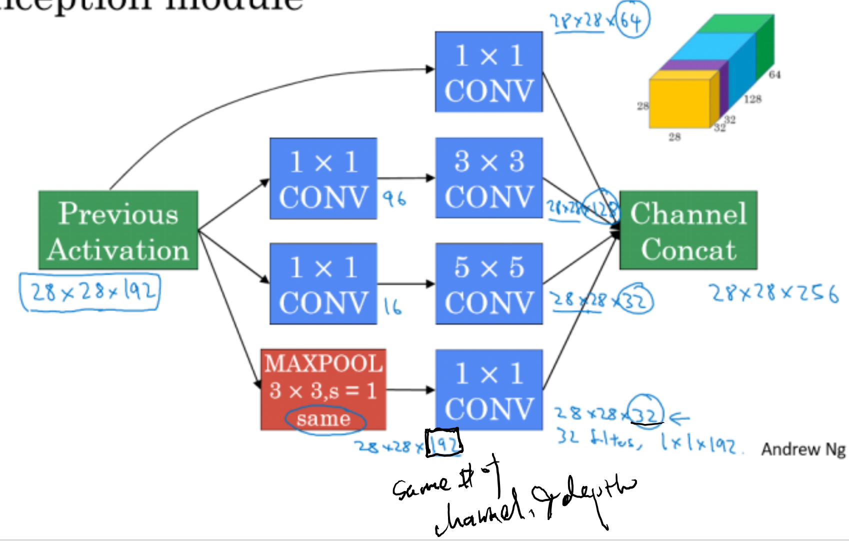
Does shrinking hurt performance?

No as long as the size of the bottleneck layer is reasonable

inception2

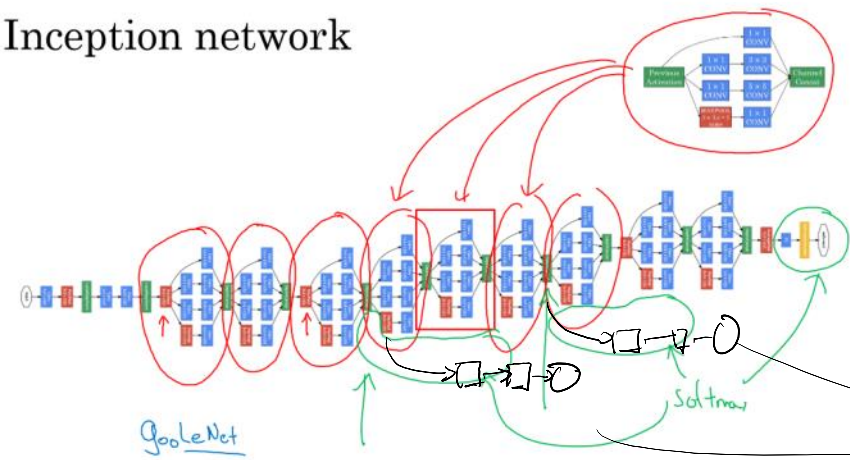
Humm... 这么拼起来有什么道理吗?

Inception module



Andrew Ng

Inception network



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

Andrew Ng

GoogLeNet

Reader friendly.

A picture of the paper

- Just lots of repeated blocks as seen in the previous slide.
- Some max-pooling layers to change height & width.

★ Additional side-branches.
~~to~~ make prediction half-way.
Make sure even half-way in the model the predictive power is good.
∆ Have regularization effect.

Fun fact: why "Inception network"

The author ~~st~~ cite this meme from the movie inception.

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

Andrew Ng

1 point

9. Which ones of the following statements on Inception Networks are true? (Check all that apply.)

- [x] Making an inception network deeper (by stacking more inception blocks together) should not hurt training set performance.
- [x] A single inception block allows the network to use a combination of 1x1, 3x3, 5x5 convolutions and pooling.
- [ ] Inception networks incorporates a variety of network architectures (similar to dropout, which randomly chooses a network architecture on each step) and thus has a similar regularizing effect as dropout.
- [x] Inception blocks usually use 1x1 convolutions to reduce the input data volume's size before applying 3x3 and 5x5 convolutions.

旧的会通过也行

1 point

10. Which of the following are common reasons for using open-source implementations of ConvNets (both the model and/or weights)? Check all that apply.

- [x] Parameters trained for one computer vision task are often useful as pretraining for other computer vision tasks.
- [x] It is a convenient way to get working an implementation of a complex ConvNet architecture.
- [ ] A model trained for one computer vision task can usually be used to perform data augmentation even for a different computer vision task.
- [ ] The same techniques for winning computer vision competitions, such as using multiple crops at test time, are widely used in practical deployments (or production system deployments) of ConvNets.

6. Which ones of the following statements on Residual Networks are true? (Check all that apply.)

- [x] A ResNet with L layers would have on the order of L^2 skip connections in total.
- [x] The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.
- [x] Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks
- [ ] The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.

8. Suppose you have an input volume of dimension n_H x n_W x n_C. Which of the following statements you agree with? (Assume that "1x1 convolutional layer" below always uses a stride of 1 and no padding.)

- [ ] ✓ You can use a pooling layer to reduce n_H, n_W, and n_C.
- [x] You can use a 1x1 convolutional layer to reduce n_C but not n_H, n_W.
- [x] ~~✗~~ You can use a pooling layer to reduce n_H, n_W, but not n_C. can! 和conv一样
- [ ] You can use a 1x1 convolutional layer to reduce n_H, n_W, and n_C.