Case Studies

Why look at case studies?
Outline

Classic networks:

- LeNet-5
- AlexNet
- VGG

ResNet (152)

Inception
LeNet - 5

60K parameters.

\[ n_h, n_w \downarrow \quad n_c \uparrow \]

conv pool \rightarrow conv pool \rightarrow \text{fc} \rightarrow \text{fc} \rightarrow \text{output}

Advanced: sigmoid/tanh \quad \text{ReLU}

Andrew Ng

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

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

[Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks]
VGG - 16

CONV = \(3 \times 3\) filter, \(s = 1\), same

MAX-POOL = \(2 \times 2\), \(s = 2\)

[CONV 64] \(\times 2\)

[CONV 256] \(\times 3\)

[CONV 512] \(\times 3\)


Andrew Ng
Case Studies

Residual Networks (ResNets)
Residual block

\[ a^{[l]} \rightarrow a^{[l+1]} \rightarrow a^{[l+2]} \]

"Short cut" /skip connection

\[
\begin{align*}
    z^{[l+1]} &= W^{[l+1]} a^{[l]} + b^{[l+1]} \\
    a^{[l+1]} &= g(z^{[l+1]}) \\
    z^{[l+2]} &= W^{[l+2]} a^{[l+1]} + b^{[l+2]} \\
    a^{[l+2]} &= g(z^{[l+2]})
\end{align*}
\]

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

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

![Diagram of ResNet]

Plain

ResNet

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

Why ResNets work
Why do residual networks work?

Identify features for residual blocks to learn!
Plain

ResNet

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

Network in Network and $1 \times 1$ convolutions
Why does a $1 \times 1$ convolution do?

$1 \times 1$ convolution can be used to:
- Reduce the number of channels/feature maps
- Reduce the number of parameters

[Lin et al., 2013. Network in network]
Using 1x1 convolutions

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

Inception network motivation
Motivation for inception network

[Szegedy et al. 2014. Going deeper with convolutions]
The problem of computational cost

CONV
5 × 5,
same,
32

28 × 28 × 192

↓

28 × 28 × 32

32 filters. filters are 5 × 5 × 192.

28 × 28 × 32 × 5 × 5 × 192 = 120M.
Using 1x1 convolution

CONV 1x1, 16, 1 x 1 x 192

CONV 5x5, 32, 5 x 5 x 16

"bottleneck layer"
Case Studies

Inception network
Inception module

- Previous Activation: $28 \times 28 \times 92$
- $1 \times 1$ CONV
- $1 \times 1$ CONV
- $1 \times 1$ CONV
- MAXPOOL $3 \times 3, s = 1$
- $1 \times 1$ CONV
- $1 \times 1$ CONV
- $3 \times 3$ CONV
- $5 \times 5$ CONV
- Channel Concat

Dimensions:
- $28 \times 28 \times 64$
- $28 \times 28 \times 128$
- $28 \times 28 \times 32$
- $28 \times 28 \times 256$
- $28 \times 28 \times 192$

Andrew Ng
Inception network

[Andrew Ng]

[Szegedy et al., 2014, Going Deeper with Convolutions]
WE NEED TO GO
DEEPER

http://knowyourmeme.com/memes/we-need-to-go-deeper
Practical advice for using ConvNets

Transfer Learning
Transfer Learning

freeze trainableParams = 0, freeze = 1

Andrew Ng
Practical advice for using ConvNets

Data augmentation
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

Advanced:
PCA
ml-class.org

AlexNet paper
“PCA color augmentation”
R G B

Andrew Ng
Implementing distortions during training

CPU thread

Disk

Load

Distortion

Mini-batch

CPU/GPU

Training
Practical advice for using ConvNets

The state of computer vision
Data vs. hand-engineering

Two sources of knowledge

- Labeled data
- 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

**Multi-crop at test time**
- Run classifier on multiple versions of test images and average results
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