Case Studies

Why look at case studies?
Outline

Classic networks:

- LeNet-5
- AlexNet
- VGG

ResNet (152)

Inception
Case Studies

Classic networks
LeNet - 5

[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 = 3x3 filter, s = 1, same

MAX-POOL = 2x2, s = 2

CONV = 64

CONV = 128

CONV = 256

CONV = 512

CONV = 512

CONV = 512

CONV = 256

CONV = 128

CONV = 64

CONV = 3x3 filter, s = 1, same

MAX-POOL = 2x2, s = 2

CONV = 64

CONV = 128

CONV = 256

CONV = 512

CONV = 512

CONV = 512

CONV = 256

CONV = 128

CONV = 64

Case Studies

Residual Networks (ResNets)
Residual block

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

"Short cut" / skip connection

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

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

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

Why ResNets work
Why do residual networks work?

Identifying a factor is easy for Residual block to learn!

ReLU : 

\[ a \geq 0 \]

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

\[ 256 = g(W^{[t+2]}a + b^{[t+2]}) + \]

\[ W^{[t]}a^{[t]}) = g(a^{[t]}) + \]

If \( W^{[t+2]} = 0 \), \( b^{[t+2]} = 0 \)
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?

$$\begin{array}{cccc} 1 & 2 & 3 & 6 \\ 3 & 5 & 5 & 1 \\ 2 & 1 & 3 & 4 \\ 4 & 7 & 8 & 5 \\ 1 & 5 & 3 & 7 \\ 5 & 4 & 9 & 8 \\ \end{array} \quad \times \quad \begin{array}{c} 2 \end{array} \quad = \quad \begin{array}{c} 6 \\ 6 \end{array} \times \begin{array}{c} 6 \end{array}$$

[Lin et al., 2013. Network in network]
Using $1 \times 1$ convolutions

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

Inception network motivation

deeplearning.ai
Motivation for inception network

[Andrew Ng]

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

**CONV**

$5 \times 5$, same, $32$

$28 \times 28 \times 192 \rightarrow 28 \times 28 \times 32$

32 filters.

$28 \times 28 \times 32 \times 5 \times 5 \times 192 = 120 \text{M}$. 

Andrew Ng
Using 1x1 convolution

```
28 x 28 x 192

CONV 1x1, 16
28 x 28 x 16

CONV 5x5, 32
28 x 28 x 32
```

"bottleneck layer"
Case Studies
Inception network

deeplearning.ai
Inception module

Previous Activation

MAXPOOL $3 \times 3, s = 1$

same

$28 \times 28 \times 192$

$28 \times 28 \times 92$

$28 \times 28 \times 32$

$28 \times 28 \times 192$

$28 \times 28 \times 92$

$28 \times 28 \times 32$

$28 \times 28 \times 128$

$28 \times 28 \times 756$

$28 \times 28 \times 64$

$28 \times 28 \times 32$

$28 \times 28 \times 32$

$1 \times 1 \text{ CONV}$

$3 \times 3 \text{ CONV}$

$5 \times 5 \text{ CONV}$

$1 \times 1 \text{ CONV}$

$1 \times 1 \text{ CONV}$

Channel Concat
Inception network

[Andrew Ng, 2014, Going Deeper with Convolutions]
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

…
Color shifting

- RGB
- +20, -20, +20
- -20, +20, +20
- +5, 0, +50

Advanced:
PAC
md-class.org
AlexNet paper
"PCA color augmentation"

Andrew Ng
Implementing distortions during training
Practical advice for using ConvNets

The state of computer vision
Two sources of knowledge

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