Video 29

Tuesday, November 10, 2020 4:11 PM

Convolution

$$(A * B)_{ij} = \sum_s \sum_t A_{st} B_{i-s, j-t}$$

Convolution layer

Inputs : images or outputs of other convolution layers

Outputs :

$$(A * B)_{ij} = \sum_s \sum_t A_{st} B_{i+s, j+t}$$ (flipping B)

$(A * B)_{ij}$: weights

$\sum_s \sum_t A_{st} B_{i+s, j+t}$: (filtering) — weight matrix — local region

Inputs have multiple channels :

- image : channels are the # color encodings.
  - w/ RGB : 3 channels

Outputs have multiple channels :

- each channel represents a specific feature we're trying to recognize in the image.
  - ↳ feature map.

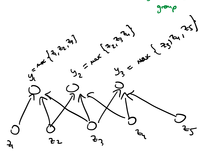
For each pair of (input, output) channels, we need to compute a 2D filtering operation

Each feature map is obtained with its own distinctive weight used in the 2D filtering operation.

Convolutional networks

Basic building block of a convolutional neural network (CNN)

1. Convolution layer
2. Non-linearity : ReLU $\max(0, z)$
   - ↳ we saw in previous weeks that the convolution is a linear operation
     - ⇒ multiple convolutions are equivalent to one single convolution
   - ↳ promote sparsity when using ReLU
3. Pooling layer
   - ↳ enable our predictions to be invariant to small input transformations
   - ↳ allows higher layers in the neural networks to consider larger regions of the input.
   - ↳ max pooling : $y_j = \max_{j \text{ in pooling group}} z_j$



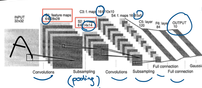
Object recognition :

task of identifying with objects is present in an image.

2 example datasets

MNIST (handwritten digits)	ImageNet (recognize objects in animals)
Classes : 10	Classes : 1000
Size : 6000 training images, 10 000 test images	Size : 1.2 million images
Input : $28 \times 28 \times 1$ images (grayscale)	Input : $224 \times 224 \times 3$ images (color)

LeNet architecture for MNIST (1998)



10 output neurons (10 classes)

(pooling) mean pooling over 2x2 regions

Layer	Type	# units	# connections	# weights
C1	convolution	4704	117,600	150
S2	pooling	1176	4704	0
C3	convolution	1600	240,000	2400
S4	pooling	400	1600	0
F5	fully connected	120	48,000	48,000
F6	fully connected	84	10,080	10,080
output	fully connected	10	840	840

↳ 10 classes in our tasks.

C1 : 6 feature maps

Filter size 5x5

Convolution uses :

- valid convolution (ignores locations where the filter lies partially outside of the input)
- same convolution (pad the input with zeros to apply the filter everywhere)

Each feature map is : 28×28

Units in C1 : $6 \times 28 \times 28 = 4704$

Connections in C1 : $28 \times 28 \times 5 \times 5 \times 6 = 117,600$

Weights in C1 : $5 \times 5 \times 6 = 150$ weights.

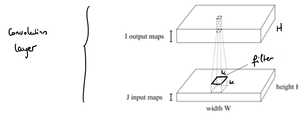
↳ shared

F5 : 120 units

400×120 connections

Take aways :

- most units are in the first convolution layer
- most of the connections are in the 1st convolution layer
- most of the weights in the fully connected layers



Output units : $W \times H \times I$

Weights : $K^2 \times I \times J$

Size of our filter

Connections : $WHIK^2 \times I \times J$

connections >> # weights

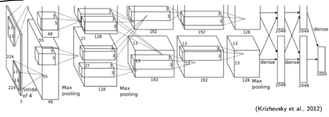
↳ consequence of weight sharing

↳ convolutions are memory efficient

ImageNet dataset & architectures

AlexNet

8 weight layers — 16.4% top 5 error (network is allowed to make 5 predictions)



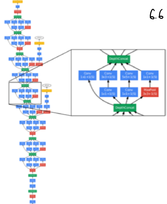
(Krizhevsky et al., 2012)

	LeNet (1989)	LeNet (1998)	AlexNet (2012)
classification task	digits	digits	objects
categories	10	10	1,000
image size	16 × 16	28 × 28	256 × 256 × 3
training examples	7,291	60,000	1.2 million
units	1,256	8,084	658,000
parameters	9,760	60,000	60 million
connections	65,000	344,000	652 million
total operations	11 billion	412 billion	200 quadrillion (est.)

GoogLeNet (2014)

22 weight layers

fully convolutional (no fully connected layers)



6.6 % top 5 test error on ImageNet

2 million parameters

↳ makes it more memory efficient (easier to deploy on mobile phones or devices w/ little capacity)

Year	Model	Top-5 error
2010	Hand-designed descriptors + SVM	28.2%
2011	Compressed Fisher Vectors + SVM	25.8%
2012	AlexNet	16.4%
2013	a variant of AlexNet	11.7%
2014	GoogLeNet	6.6%
2015	deep residual nets	4.5%
2020	Fix Efficient Net - L2	1.3%