

CSC 144V: Autonomous Vehicles

Lecture 3

Convolutional Neural Networks Exercises Solution

Department of Computer Science,
Hofstra University

Exercise: Layer shapes and parameters

Input $32 \times 32 \times 3 \rightarrow$ CONV1 (8 filters, 5×5 , stride 1, no padding) + ReLU $\rightarrow$ MAXPOOL (2×2 , stride 2) $\rightarrow$ CONV2 (16 filters, 3×3 , stride 1, padding 1) + ReLU $\rightarrow$ MAXPOOL (2×2 , stride 2) $\rightarrow$ flatten $\rightarrow$ FC (10 outputs). Each filter and each FC neuron has one bias.

Fill in the output shape and the number of learnable parameters of each layer.

Layer	Output shape ($H \times W \times C$)	# learnable parameters
CONV1		
MAXPOOL1		
CONV2		
MAXPOOL2		
FC		

Exercise Solution: Layer shapes and parameters

- Output size = $(N - F + 2P)/S + 1$: CONV1 $(32 - 5)/1 + 1 = 28$; CONV2 with padding 1 keeps 14×14 .
- Pooling layers have no parameters. **Total = 9,626 parameters.**

Layer	Output shape (H × W × C)	# learnable parameters
CONV1	$28 \times 28 \times 8$	$(5 \cdot 5 \cdot 3 + 1) \times 8 = 608$
MAXPOOL1	$14 \times 14 \times 8$	0
CONV2	$14 \times 14 \times 16$	$(3 \cdot 3 \cdot 8 + 1) \times 16 = 1,168$
MAXPOOL2	$7 \times 7 \times 16$	0
FC	10	$7 \cdot 7 \cdot 16 \times 10 + 10 = 7,850$

Exercise: FC vs convolution

Suppose the CONV1 output ($28 \times 28 \times 8$) were instead produced by a **fully connected layer** from the $32 \times 32 \times 3$ input (ignore biases).

- (i) How many weights would that need, and how many times more than the $5 \times 5 \times 3 \times 8$ weights of CONV1?
- (ii) Name the two properties of convolution that explain the difference.

Exercise Solution: FC vs convolution

- FC: 3,072 inputs $\times$ 6,272 outputs = **19,267,584 weights**.
- CONV1: $5 \cdot 5 \cdot 3 \cdot 8 =$ **600 weights**.
- Ratio: $19,267,584 / 600 \approx$ **32,000 $\times$** more.
- **Local connectivity**: each output sees only a $5 \times 5 \times 3$ patch.
- **Parameter sharing**: the same filter is reused at every spatial position.

Exercise: Convolution by hand

Convolve the 4×4 input with the 2×2 filter (stride 1, no padding, no flip), then apply **ReLU** and **2×2 max pooling** with stride 1.

Fill in the three tables.

Input (4×4)

1	2	0	1
0	1	3	1
2	1	0	2
1	0	1	3

Filter (2×2)

1	0
0	-1

stride 1, no padding, no flip

Conv (3×3)

After ReLU

Max-pool 2×2 , $s = 1$

Exercise Solution: Convolution by hand

- Each output = (top-left patch value) $\times$ 1 – (bottom-right patch value).
- First entry: $1 - 1 = 0$; last entry: $0 - 3 = -3$.
- ReLU zeroes the negative entries.
- Max-pooling (2×2 , stride 1) turns the 3×3 map into 2×2 .

Input (4×4)

1	2	0	1
0	1	3	1
2	1	0	2
1	0	1	3

Filter (2×2)

1	0
0	-1

stride 1, no padding, no flip

Conv (3×3)

0	-1	-1
-1	1	1
2	0	-3

After ReLU

0	0	0
0	1	1
2	0	0

Max-pool 2×2 , $s = 1$

1	1
2	1

Exercise: Separable convolution

Replace CONV2 (3×3 , 8 input channels $\rightarrow$ 16 output channels, 14×14 output, ignore biases) by a **depthwise separable** convolution (3×3 depthwise followed by 1×1 pointwise).

- (i) Number of parameters of the regular and the depthwise separable version.
- (ii) Number of multiplications for the whole 14×14 output of each version.
- (iii) By what factor is the cost reduced?

	Parameters	Multiplications
Regular		
Separable		
Reduction		

Exercise Solution: Separable convolution

- **Regular:** $3 \times 3 \times 8 \times 16 = 1,152$ parameters
- **Separable:** depthwise $3 \times 3 \times 8 = 72$, pointwise $1 \times 1 \times 8 \times 16 = 128$; total 200
- Multiplications: regular $14 \times 14 \times 1,152 = 225,792$; separable $14 \times 14 \times 200 = 39,200$
- $1,152 / 200 = 5.76\times$ fewer parameters and multiplications ($= 1/16 + 1/9 \approx 0.174$ of the regular cost)

	Parameters	Multiplications
Regular	1,152	225,792
Separable	200	39,200
Reduction	5.76×	5.76×