

1. Convolutional layer volume W_2, H_2, D_2 :

$$W_2 = (W_1 - F + 2P)/S + 1$$

$$H_2 = (H_1 - F + 2P)/S + 1$$

$$D_2 = K$$

From the reading material from Stanford, to calculate how many neurons “fit” is given by $(W-F+2P)/S+1$, and the width(W) = height(H) and the depth(D) will stay the same, therefore we can get the formulas for W_2 and H_2 .

2. Volume calculation example

Input size 32x32, 100 filters, filter size 3x3, stride =1, depth = 1

A. The size of feature map (including depth):

$$(32 - 3)/1 + 1 = 30$$

Therefore, the feature map: $30 \times 30 \times 100 = 90000$

B. The total number of parameters

$$(3 \times 3 \times 1 + 1) \times 100 = 1000$$

3. Convolution filter application

A.

$$1. 0*1+0*1+0*-1+0*-1+0*1+2*1+2*-1+2*-1+0*1+2*1+2*-1+2*-1+0*1+2*1+2*-1+2*-1 = -6$$

$$2. 0*1+0*1+0*-1+0*-1+2*1+2*1+2*-1+0*-1+2*1+2*1+2*-1+0*-1+2*1+2*1+2*-1+0*-1 = 6$$

$$3. 0*1+0*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1 = 12$$

$$4. 0*1+2*1+2*-1+2*-1+0*1+2*1+2*-1+2*-1+0*1+2*1+2*-1+2*-1+0*1+0*1+0*-1+0*-1 = -6$$

$$5. 2*1+2*1+2*-1+0*-1+2*1+2*1+2*-1+0*-1+2*1+2*1+2*-1+0*-1+0*1+0*1+0*-1+0*-1 = 6$$

$$6. 2*1+2*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1+0*1+0*1+0*-1+0*-1 = 12$$

$$7. 0*1+2*1+2*-1+2*-1+0*1+2*1+2*-1+2*-1+0*1+0*1+0*-1+0*-1+0*1+0*1+0*-1+0*-1 = -4$$

$$8. 2*1+2*1+2*-1+0*-1+2*1+2*1+2*-1+0*-1+0*1+0*1+0*-1+0*-1+0*1+0*1+0*-1+0*-1 = 4$$

$$9. 2*1+2*1+0*-1+0*-1+2*1+2*1+0*-1+0*-1+0*1+0*1+0*-1+0*-1+0*1+0*1+0*-1+0*-1 = 8$$

-6	6	12
-6	6	12

-4	4	8
----	---	---

B.

With the stride = 1, the image focuses mainly on the top left corner since the 2's are present mainly on the top left corner and show brighter color.

4. Padding

Input = 7x7, filter 3x3, stride = 2, output 3x3, depth 1

$(7 + 2P - 3)/2 + 1 = 7$, therefore, $P = 4$

So the padding size is 4x4.

5. CNN Parameters

Input: 32x32x3: 32 wide, 32 high, 3 color channels

[1st layer] Convolution – 32 5x5 filters, stride (1,1), activation ReLU

conv2d_1 (Conv2D)	(None, 28, 28, 32)	2432
-------------------	--------------------	------

param # for each filter: $(5*5*3)+1 = 76$

param # for all: $32*76 = 2432 = \text{param \#}$

[2nd layer] Max pooling – size 2x2, stride (2,2)

max_pooling2d_1 (MaxPooling2)	(None, 14, 14, 32)	0
-------------------------------	--------------------	---

The max_pooling layer will reduce the number of parameters, operate independently on every depth slice of the input, and resizes it spatially, using the MAX operation; therefore, no param #.

[3rd layer] Convolution – 32 5x5 filters, stride (1,1), activation ReLU

conv2d_2 (Conv2D)	(None, 10, 10, 32)	25632
-------------------	--------------------	-------

param # for each filter: $(5*5*32)+1 = 801$

param # for all: $801*32 = 25632 = \text{param \#}$

[4th layer] Max pooling – size 2x2, stride (2,2)

max_pooling2d_2 (MaxPooling2 (None, 5, 5, 32)	0
-----------------------------------------------	---

The max_pooling layer will reduce the number of parameters, operate independently on every depth slice of the input, and resizes it spatially, using the MAX operation; therefore, no param #.

[5th layer] Fully connected (Dense) – 512 nodes, activation ReLU

dense_1 (Dense)	(None, 512)
410112	

$$(800+1) * 512 = 410112 = \text{param \#}$$

[6th layer] Fully connected (Dense) – 10 nodes, activation Softmax

dense_2 (Dense)	(None, 10)	5130
-----------------	------------	------

$$(512+1) * 10 = 5130$$

Total params: 443,306

$$2432+0+25632+0+0+410112+5130 = 443306 = \text{total params}$$