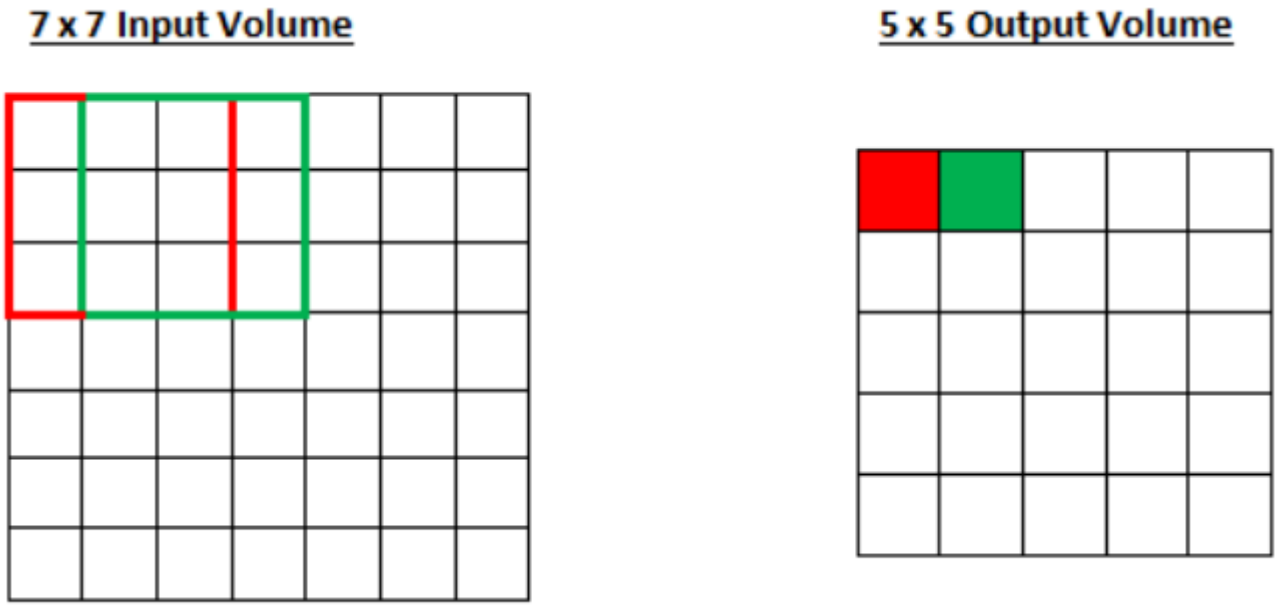
Convolutional Neural Networks

CNN Steps

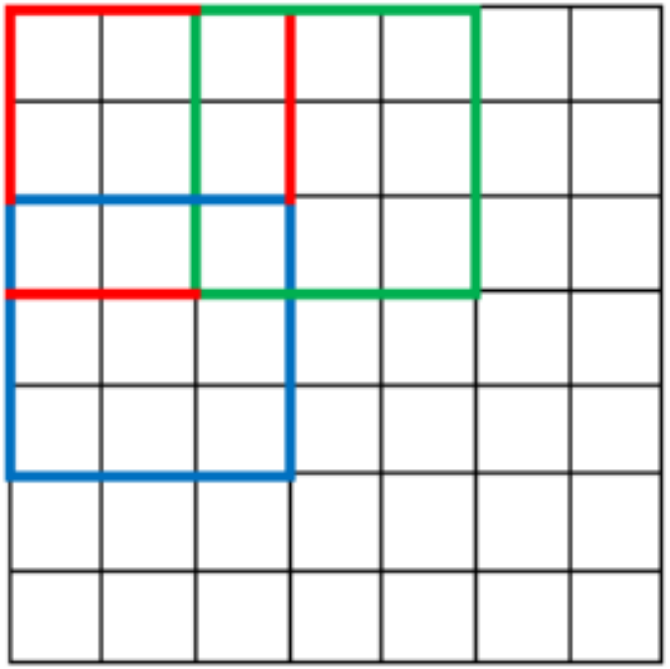
- Provide input image into convolution layer.
- Choose parameters, apply filters with strides, padding if requires. ...
- Perform pooling to reduce dimensionality size.
- Add as many convolutional layers until satisfied.
- Flatten the output and feed into a fully connected layer (FC Layer)

Stride 1

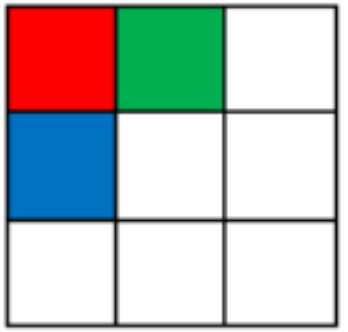


Stride 2

7 x 7 Input Volume



3 x 3 Output Volume

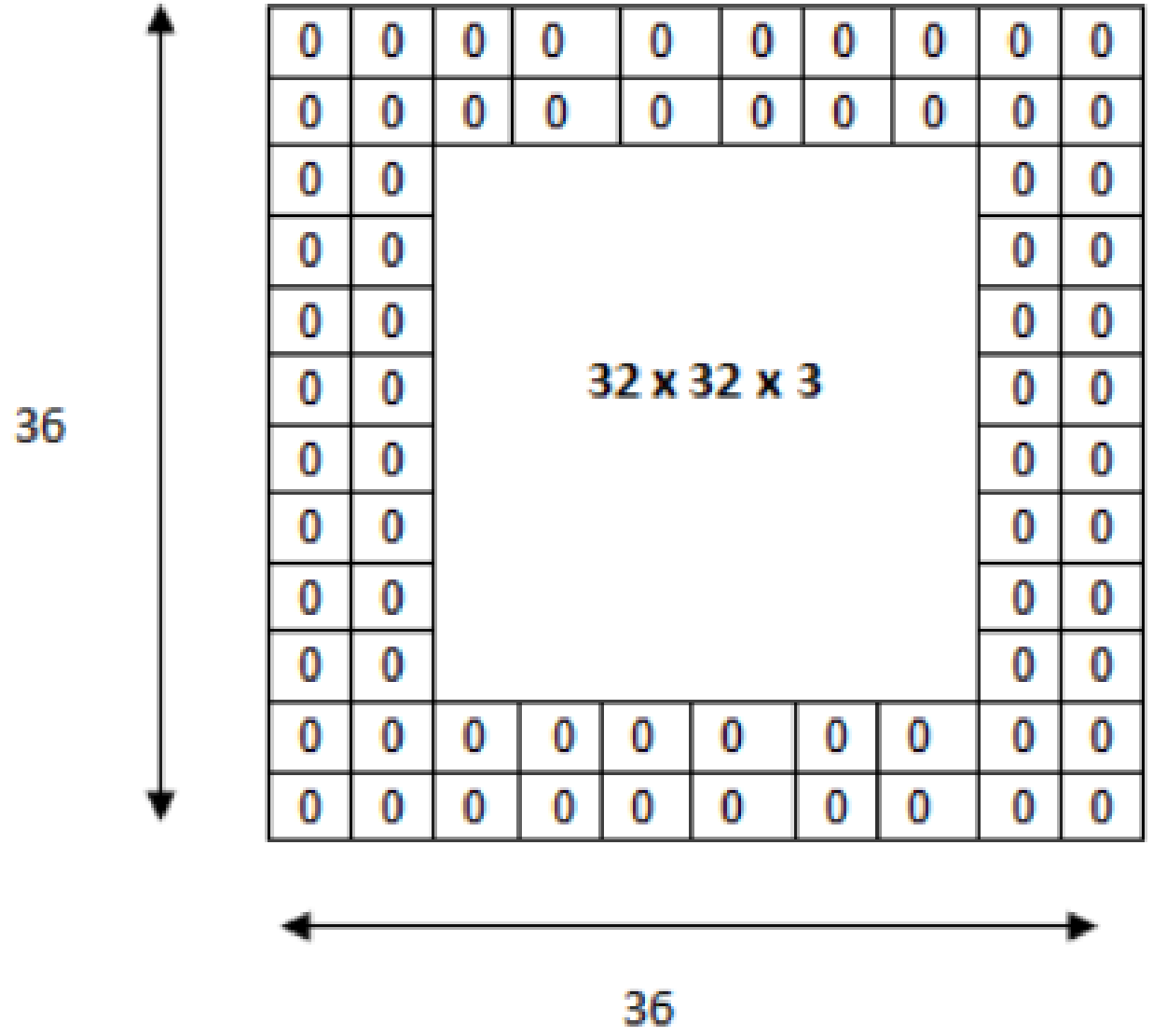


Padding

When you apply three 5 x 5 x 3 filters to a 32 x 32 x 3 input volume?
The output volume would be 28 x 28 x 3.
The spatial dimensions decrease. As we keep on applying conv layers, the size of the volume will decrease faster than we would like.

In the early layers of our network, we want to preserve as much information about the original input volume so that we can extract those low level features.
We want to apply the same conv layer but we want the output volume to remain 32 x 32 x 3.
To do this, we can apply a zero padding of size 2 to that layer. Zero padding pads the input volume with zeros around the border. If we think about a zero padding of two, then this would result in a 36 x 36 x 3 input volume.

Padding



The input volume is 32 x 32 x 3. If we imagine two borders of zeros around the volume, this gives us a 36 x 36 x 3 volume. Then, when we apply our conv layer with our three 5 x 5 x 3 filters and a stride of 1, then we will also get a 32 x 32 x3 output volume.

Padding

If you have a stride of 1 and if you set the size of zero padding to

$$Zero\ Padding = \frac{(K-1)}{2}$$

where K is the filter size, then the input and output volume will always have the same spatial dimensions.

Input and Output Size

The formula for calculating the output size for any given conv layer is

$$O = \frac{(W - K + 2P)}{S} + 1$$

O is the output height/length

W is the input height/length

K is the filter size

P is the padding &

S is the stride.

Ip image size is 96 x 96. It is convolved with a filter size of 7 x 7. Stride size is (1, 1). Padding size is (2, 2). What is the output size?

94 x 94

$$O = \frac{(W - K + 2P)}{S} + 1$$

Ip image size is 96 x 96. It is convolved with a filter size of 7 x 7. Stride size is (1, 1). Padding size is (2, 2). What is the output size?

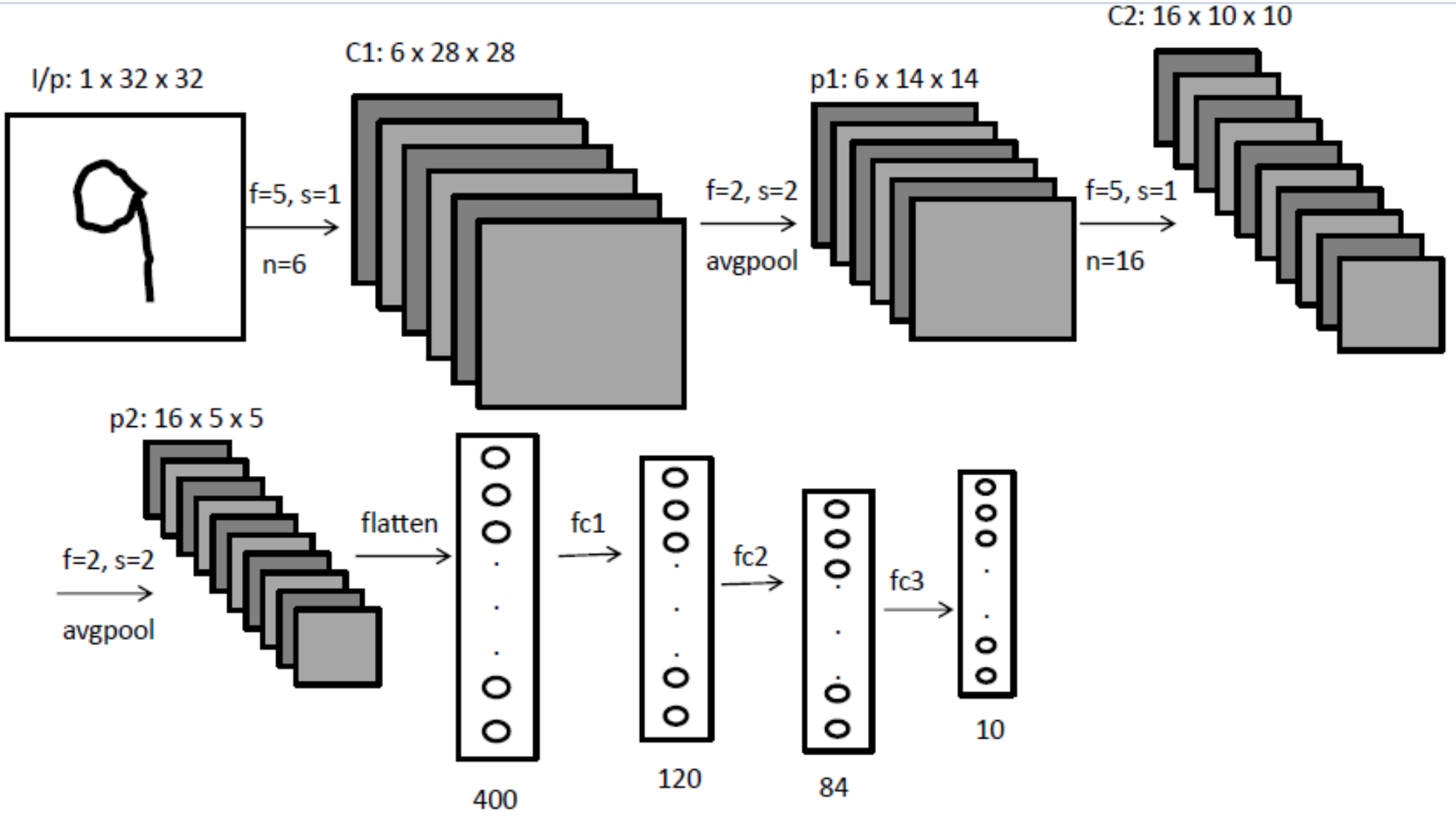
94 x 94

Suppose incoming batch dimension is 128 x 3 x 28 x 28. There are two filter of dimension 3 x 3 each. Stride size is (2, 2). No padding. What is output size?

128 x 2 x 13 x 13

$$O = \frac{(W - K + 2P)}{S} + 1$$

A simple CNN


l/p: 1 x 32 x 32
C1: 6 x 28 x 28
f=5, s=1
n=6
p1: 6 x 14 x 14
f=2, s=2
avgpool
C2: 16 x 10 x 10
f=5, s=1
n=16
p2: 16 x 5 x 5
f=2, s=2
avgpool
flatten
400
fc1
120
fc2
84
fc3
10

Links

http://machinelearninguru.com/computer_vision/basics/convolution/convolution_layer.html