4-3 Convolution Neural Network II

Zhonglei Wang

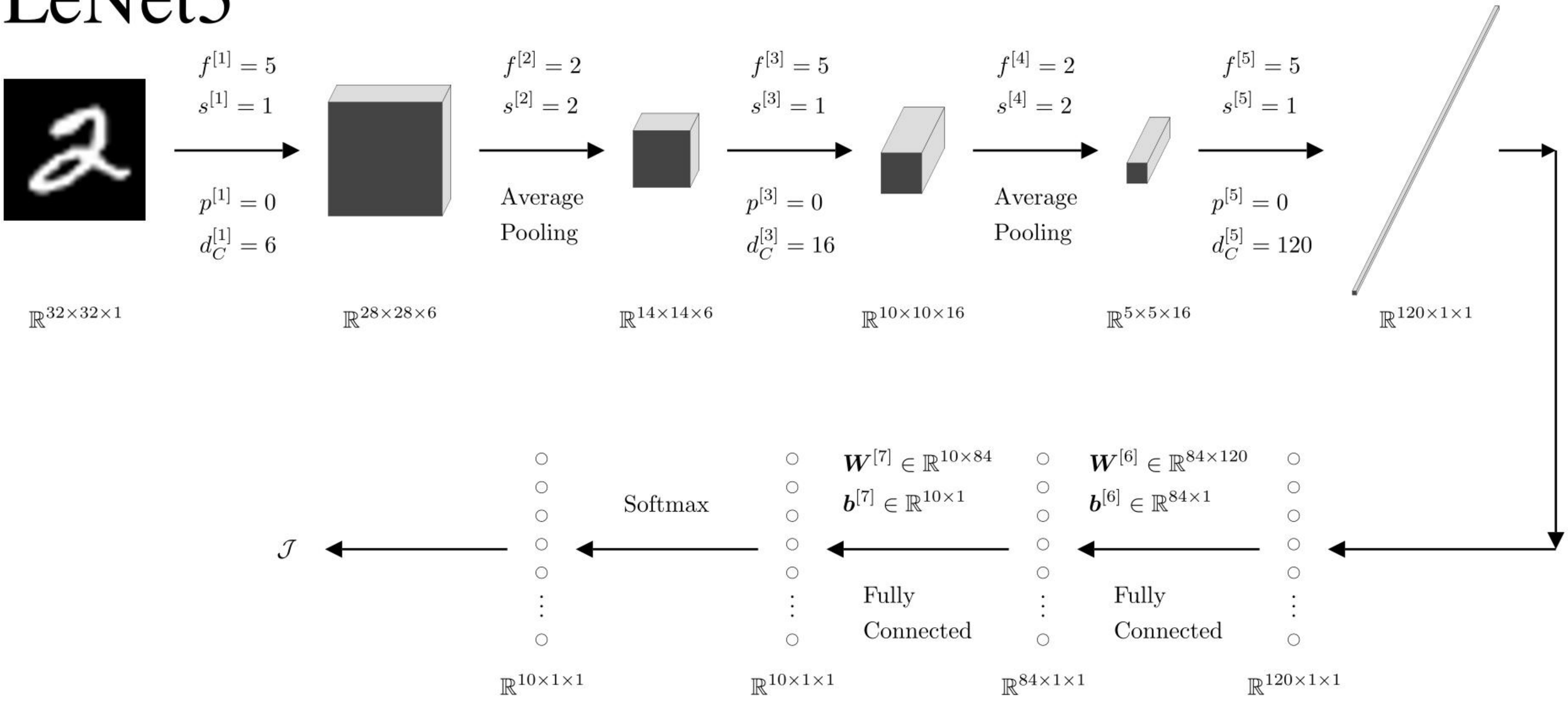
WISE and SOE, XMU, 2025

Contents

LeNet5

1. LeNet5 was proposed by LeCun in 1998
   - It is used to classify a 32×32 gray-scale handwritten digits (10 classes)
   - tanh is used as the activation function for the hidden layers
   - It is one of the earliest CNN
   - It is a landmark, showing success of CNN in CV
   - "5" means there exist 5 (or 7) layers

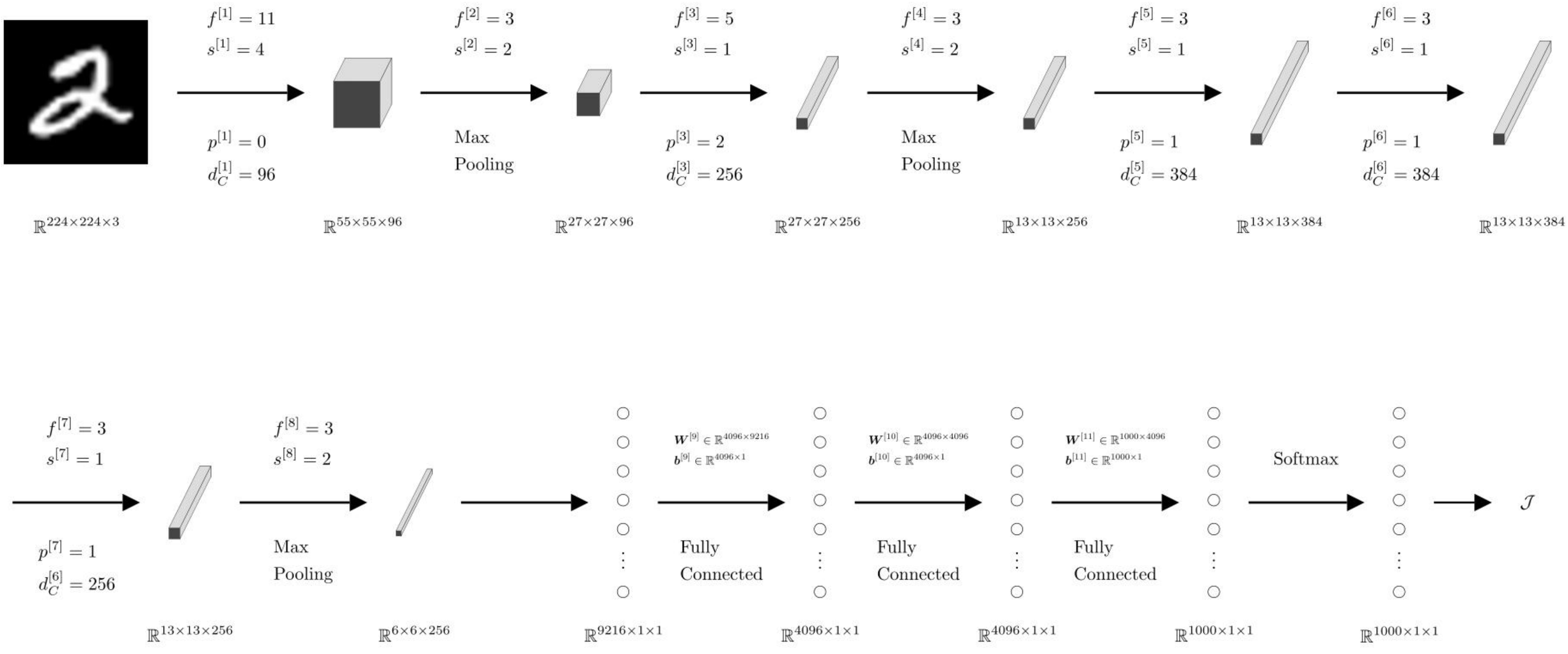
LeNet5


$f^{[1]} = 5$
$s^{[1]} = 1$
$p^{[1]} = 0$
$d_C^{[1]} = 6$
$f^{[2]} = 2$
$s^{[2]} = 2$
Average Pooling
$f^{[3]} = 5$
$s^{[3]} = 1$
$p^{[3]} = 0$
$d_C^{[3]} = 16$
$f^{[4]} = 2$
$s^{[4]} = 2$
Average Pooling
$f^{[5]} = 5$
$s^{[5]} = 1$
$p^{[5]} = 0$
$d_C^{[5]} = 120$
$\mathbb{R}^{32\times32\times1}$
$\mathbb{R}^{28\times28\times6}$
$\mathbb{R}^{14\times14\times6}$
$\mathbb{R}^{10\times10\times16}$
$\mathbb{R}^{5\times5\times16}$
$\mathbb{R}^{120\times1\times1}$
$\boldsymbol{W}^{[6]} \in \mathbb{R}^{84\times120}$
$\boldsymbol{b}^{[6]} \in \mathbb{R}^{84\times1}$
$\boldsymbol{W}^{[7]} \in \mathbb{R}^{10\times84}$
$\boldsymbol{b}^{[7]} \in \mathbb{R}^{10\times1}$
Fully Connected
Fully Connected
Softmax
$\mathcal{J}$
$\mathbb{R}^{120\times1\times1}$
$\mathbb{R}^{84\times1\times1}$
$\mathbb{R}^{10\times1\times1}$
$\mathbb{R}^{10\times1\times1}$

AlexNet

1. It is also a landmark in CV
   - It won the ImageNet Large Scale Visual Recognition Challenge 2012
   - It is used to classify a $224 \times 224 \times 3$ colorful images among 1000 classes
   - Its performance surpassed others significantly
   - It promoted the development of deep neural networks in CV greatly

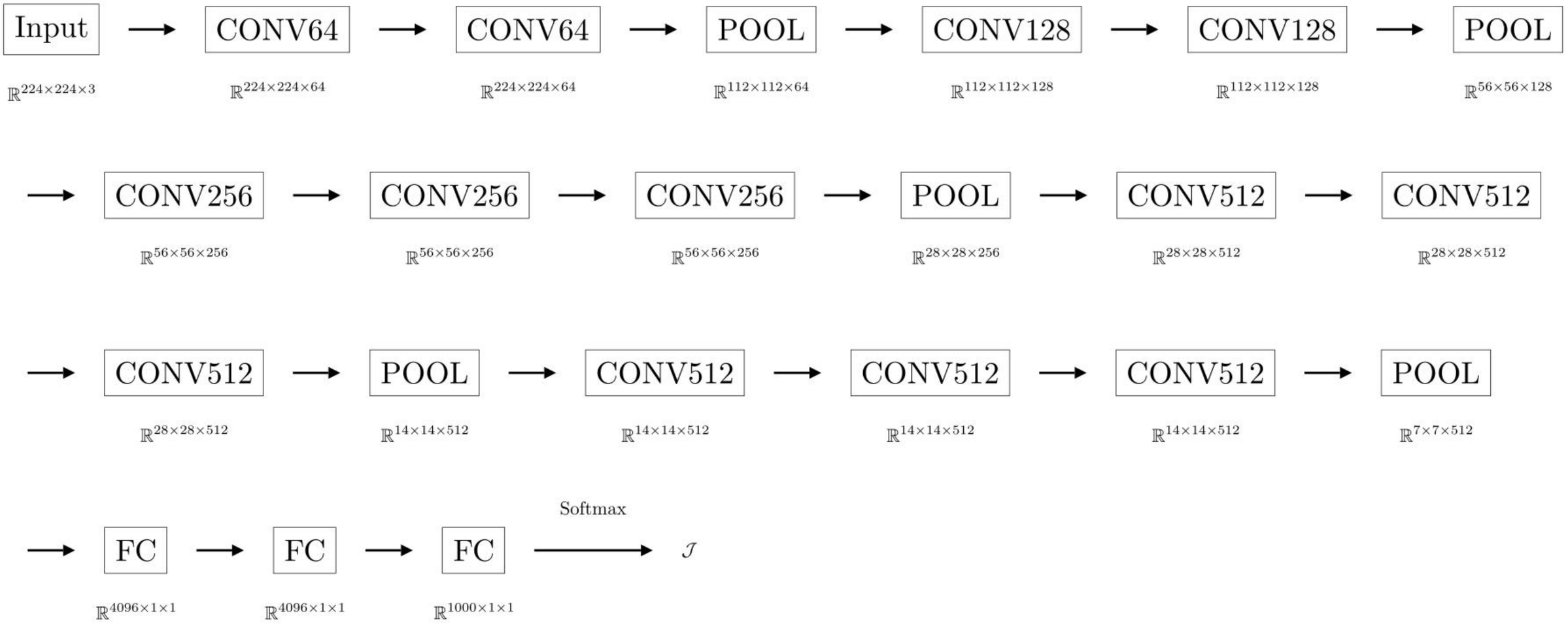
AlexNet


$f^{[1]} = 11$
$s^{[1]} = 4$
$p^{[1]} = 0$
$d_C^{[1]} = 96$
$f^{[2]} = 3$
$s^{[2]} = 2$
Max Pooling
$f^{[3]} = 5$
$s^{[3]} = 1$
$p^{[3]} = 2$
$d_C^{[3]} = 256$
$f^{[4]} = 3$
$s^{[4]} = 2$
Max Pooling
$f^{[5]} = 3$
$s^{[5]} = 1$
$p^{[5]} = 1$
$d_C^{[5]} = 384$
$f^{[6]} = 3$
$s^{[6]} = 1$
$p^{[6]} = 1$
$d_C^{[6]} = 384$
$\mathbb{R}^{224\times224\times3}$
$\mathbb{R}^{55\times55\times96}$
$\mathbb{R}^{27\times27\times96}$
$\mathbb{R}^{27\times27\times256}$
$\mathbb{R}^{13\times13\times256}$
$\mathbb{R}^{13\times13\times384}$
$\mathbb{R}^{13\times13\times384}$
$f^{[7]} = 3$
$s^{[7]} = 1$
$p^{[7]} = 1$
$d_C^{[6]} = 256$
$f^{[8]} = 3$
$s^{[8]} = 2$
Max Pooling
$\boldsymbol{W}^{[9]} \in \mathbb{R}^{4096\times9216}$
$\boldsymbol{b}^{[9]} \in \mathbb{R}^{4096\times1}$
Fully Connected
$\boldsymbol{W}^{[10]} \in \mathbb{R}^{4096\times4096}$
$\boldsymbol{b}^{[10]} \in \mathbb{R}^{4096\times1}$
Fully Connected
$\boldsymbol{W}^{[11]} \in \mathbb{R}^{1000\times4096}$
$\boldsymbol{b}^{[11]} \in \mathbb{R}^{1000\times1}$
Fully Connected
Softmax
$\mathcal{J}$
$\mathbb{R}^{13\times13\times256}$
$\mathbb{R}^{6\times6\times256}$
$\mathbb{R}^{9216\times1\times1}$
$\mathbb{R}^{4096\times1\times1}$
$\mathbb{R}^{4096\times1\times1}$
$\mathbb{R}^{1000\times1\times1}$
$\mathbb{R}^{1000\times1\times1}$

VGG16

1. It is an important CNN in CV
   - It won the ImageNet Large Scale Visual Recognition Challenge 2014
   - It is used to classify a $224 \times 224 \times 3$ colorful images among 1000 classes
2. Some notations
   - CONV[num]: conduct "num" convolution operations with size 3×3 stride 1 and padding 1
   - POOL: conduct Max pooling with size 2×2 stride 2
3. Remark
   - Convolution operations do not change the image size
   - Max pooling operations decrease half of the image size

VGG16


Input
CONV64
CONV64
POOL
CONV128
CONV128
POOL
$\mathbb{R}^{224\times224\times3}$
$\mathbb{R}^{224\times224\times64}$
$\mathbb{R}^{224\times224\times64}$
$\mathbb{R}^{112\times112\times64}$
$\mathbb{R}^{112\times112\times128}$
$\mathbb{R}^{112\times112\times128}$
$\mathbb{R}^{56\times56\times128}$
CONV256
CONV256
CONV256
POOL
CONV512
CONV512
$\mathbb{R}^{56\times56\times256}$
$\mathbb{R}^{56\times56\times256}$
$\mathbb{R}^{56\times56\times256}$
$\mathbb{R}^{28\times28\times256}$
$\mathbb{R}^{28\times28\times512}$
$\mathbb{R}^{28\times28\times512}$
CONV512
POOL
CONV512
CONV512
CONV512
POOL
$\mathbb{R}^{28\times28\times512}$
$\mathbb{R}^{14\times14\times512}$
$\mathbb{R}^{14\times14\times512}$
$\mathbb{R}^{14\times14\times512}$
$\mathbb{R}^{14\times14\times512}$
$\mathbb{R}^{7\times7\times512}$
Softmax
FC
FC
FC
$\mathcal{J}$
$\mathbb{R}^{4096\times1\times1}$
$\mathbb{R}^{4096\times1\times1}$
$\mathbb{R}^{1000\times1\times1}$