**Computer Vision 2016 Spring HW#5 theory**

2013-11415 Sanha Lee

**Q1.**

Input layer: 224 x 224 x 3

1. Convolution layer with 11 x 11 size filters, 96 neurons, stride 4 -> 55 x 55 x 96

2. Convolution layer with 5 x 5 size filters, 256 neurons, stride 1, pad 2 -> 55 x 55 x 256

3. Pooling layer with 2 x 2 -> 27 x 27 x 256

4. Convolution layer with 3 x 3 size filters, 384 neurons, stride 1, pad 1 -> 27 x 27 x 384

5. Pooling layer with 2 x 2 -> 13 x 13 x 384

6. Convolution layer with 3 x 3 size filters, 384 neurons, stride 1, pad 1 -> 13 x 13 x 384

7. Convolution layer with 3 x 3 size filters, 256 neurons, stride 1, pad 1 -> 13 x 13 x 256

8. Pooling layer with 3 x 3 -> 4 x 4 x 256 (= 4096)

=> 4096 dimension vector connected!

**Q2.**

During pooling operation, max pooling picks maximum value, mean pooling picks mean value, and min pooling picks minimum value of selected region.

Usually, max pooling shows better performance.