CS GY 6643 - Computer Vision, Fall 2024

Homework 3

Due: 2024/11/21 10:59 AM

Note: Delays will incur 1 point deduction for every hour of delay in submissions(rounded down). Discussions are allowed on homework, but solutions must be written independently.

---

1. Design a CNN architecture with the following specifications: **(Points 5)**
   - Input: 300×300 image with 3 channels
   - 6 convolutional layers, each with 3×3 filters and stride 2
   - 16 filters in each convolutional layer
   - 2×2 max pooling after every 2 convolutional layers
   - 2 fully connected layers of size 512 and 128 at the end before output layer
   - Output: 10 classes

   Construct and specify the shape of each layer in the final architecture. Provide a summary of the network structure. Convolution operations should be without any padding, and Max Pooling layers should have a stride of 2. Refer to the image shown below as an example, and create a similar illustration for the given scenario.

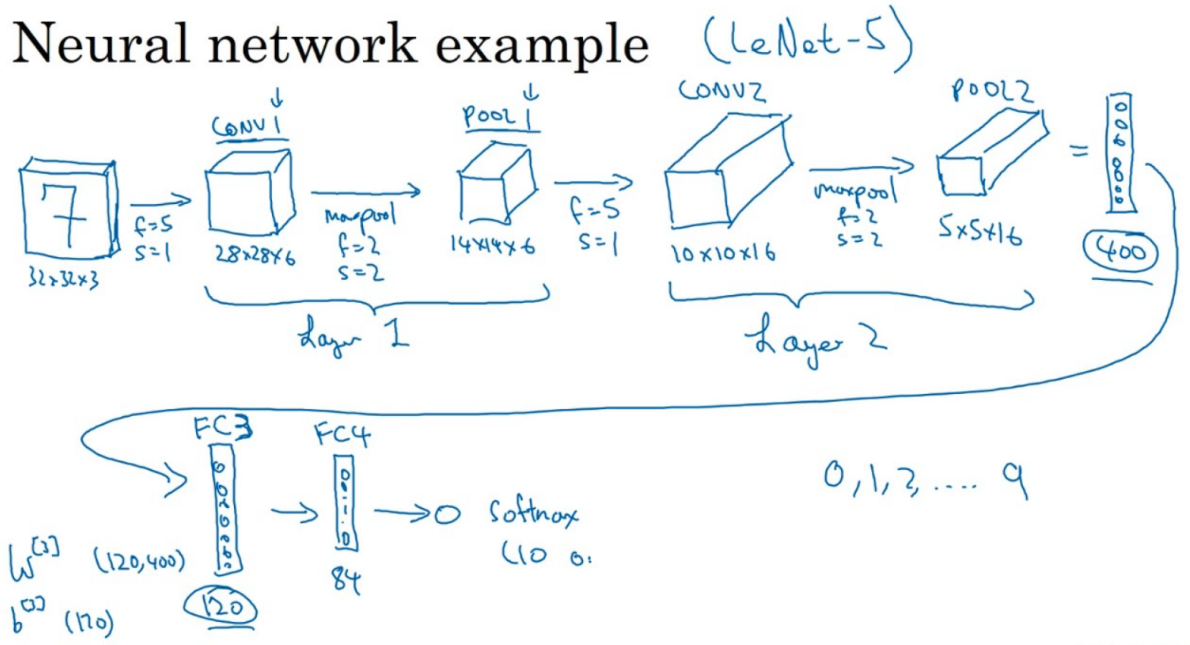

Figure 1: Example Architecture we expect

2. Consider the following CNN architecture: **(Points 5)**
   - Input: 64×64 image with 1 channel
   - Conv1: 32 filters of size 5×5, stride 1, padding 2
   - MaxPool1: 2×2, stride 2
   - Conv2: 64 filters of size 3×3, stride 1, padding 1
   - MaxPool2: 2×2, stride 2
   - Fully Connected 1: 1024 neurons
   - Fully Connected 2: 10 neurons (output)

   Calculate the total number of trainable parameters in this network. Show your work for each layer.

3. Answer the following questions: **(Points 2.5+2.5)**
   (a) Show why a fully connected neural network having multiple layers with no activation functions behaves just like a single linear equation applied to the input features.
   (b) Prove that a CNN with multiple convolutional layers, without max pooling and activation functions, is equivalent to a single convolutional layer.

4. Consider a 5×5 input image: **Bonus Question (Points 2.5)**

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

   And a 3×3 convolutional filter:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

   (a) Apply this filter to the image with stride 1 and no padding. Show the resulting feature map after one convolution step.
   (b) Apply the same filter to the result from part (a). Show that after just 2 convolution steps, you have increased the receptive field. Explain what this means in terms of the original input pixels.