Learning Deep Neural Networks (Deep Learning)

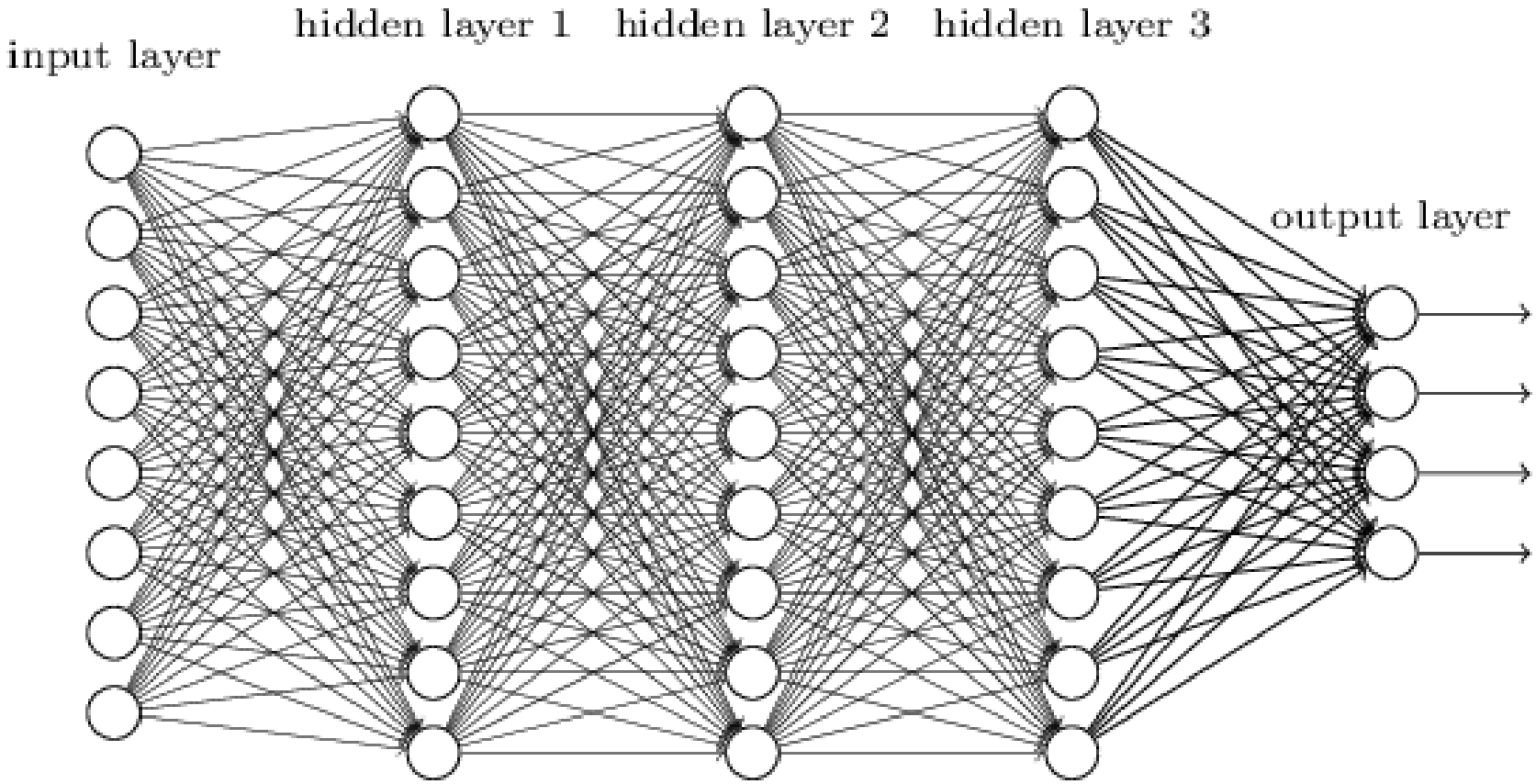
CONVOLUTIONAL NEURAL NETWORK

RECURRENT NEURAL NETWORK

What is deep learning? Why to use it?

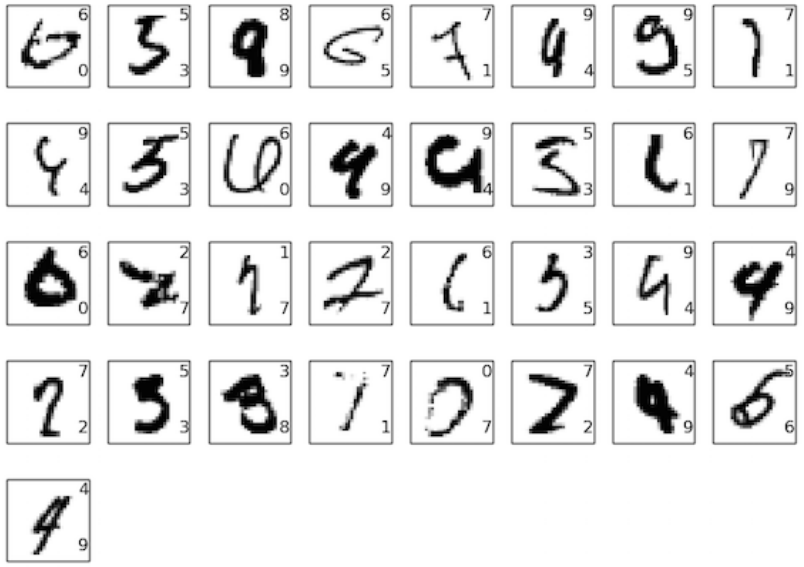
- Deep Neural Networks
  - Having lots of hidden layer neurons
  - Hard to train
  - but much more powerful!!!
- Deep Learning
  - Techniques to use/train a deep neural net efficiently
  - CNN, RNN, GPU processing, …

How deep are the deep neural nets



MNIST

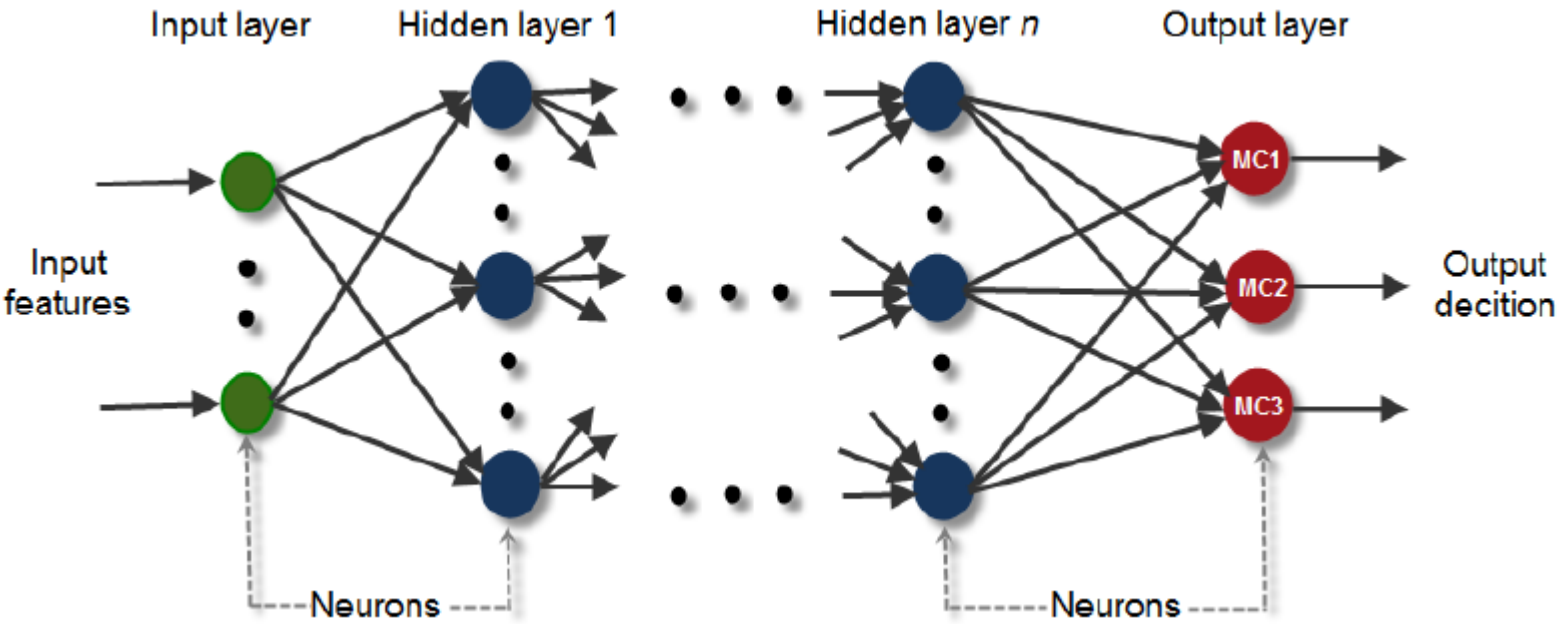
- A famous database for image processing purposes
- MNIST database of handwritten digits
  - 28*28 pixel images



http://yann.lecun.com/exdb/mnist/

Classic MLP Networks; a review

- Fully connected
- Input – hidden – output

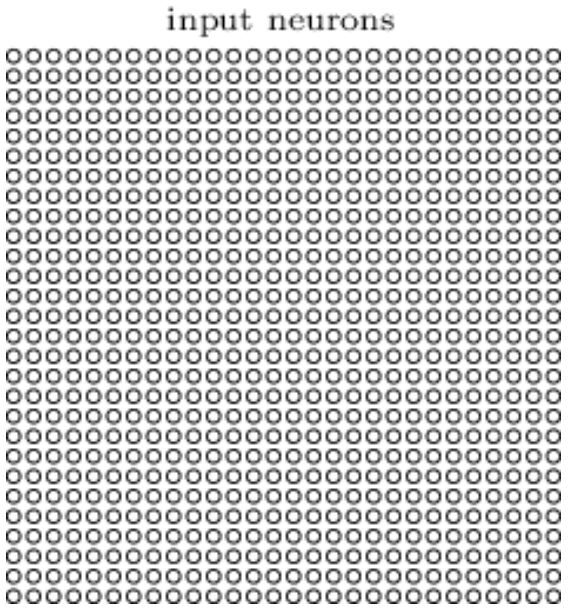


Convolutional Neural Network (CNN)

- A powerful classifier
  - Recognition
  - Detection
  - Extracts useful features automatically!
- 3 Basic feature (To be discussed later)
  - Local Receptive Fields
  - Shared Weight and Biases
  - Pooling Layer

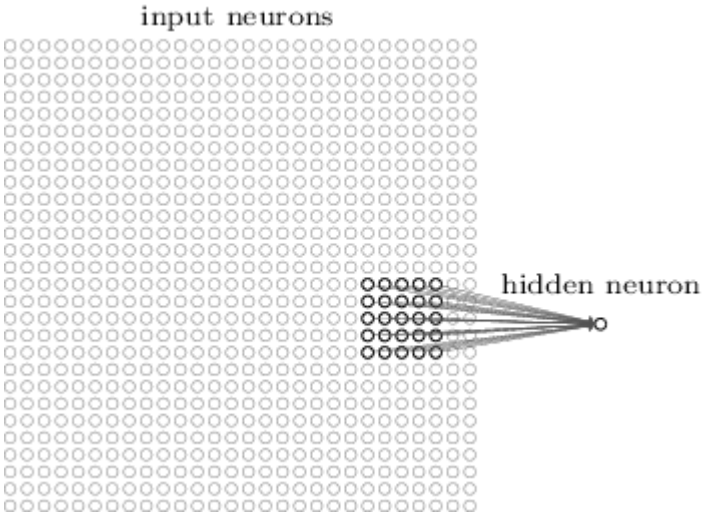
CNN structure

- 4 group of layers
  I. Input layer
     - Neurons are ordered in a matrix form not a vector
  II. Convolutional layer(s)
  III. Pooling layer
  IV. Output layer



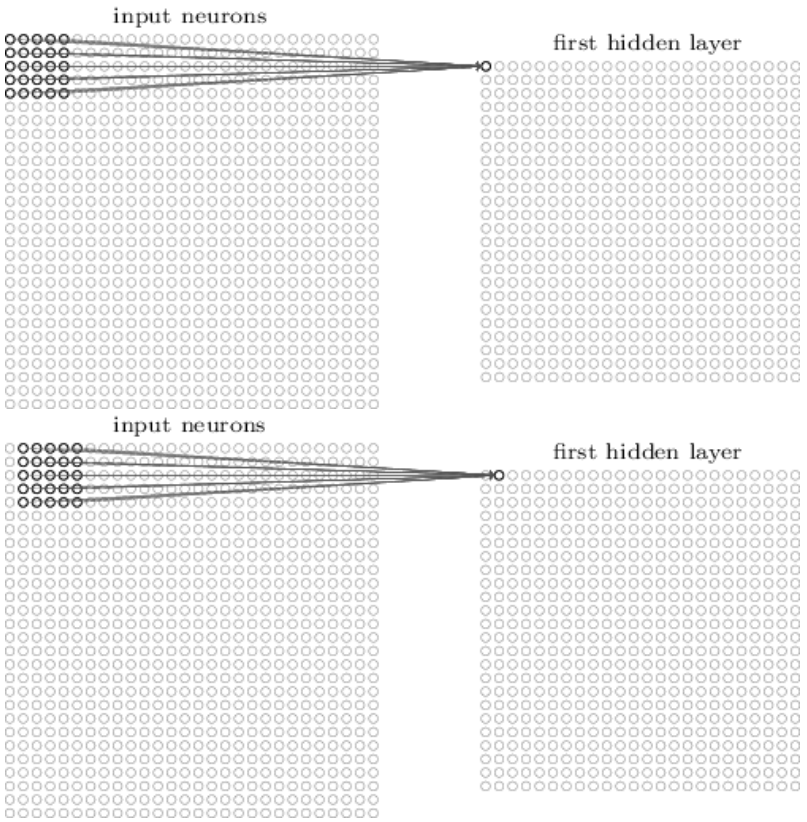
Local Receptive Field

- Unlike classic nets
  - Each hidden layer neuron is connected only to a small region of the input/previous layer
  - We call this region "Local Receptive Field" of that specific neuron



Local Receptive Field (cont.)

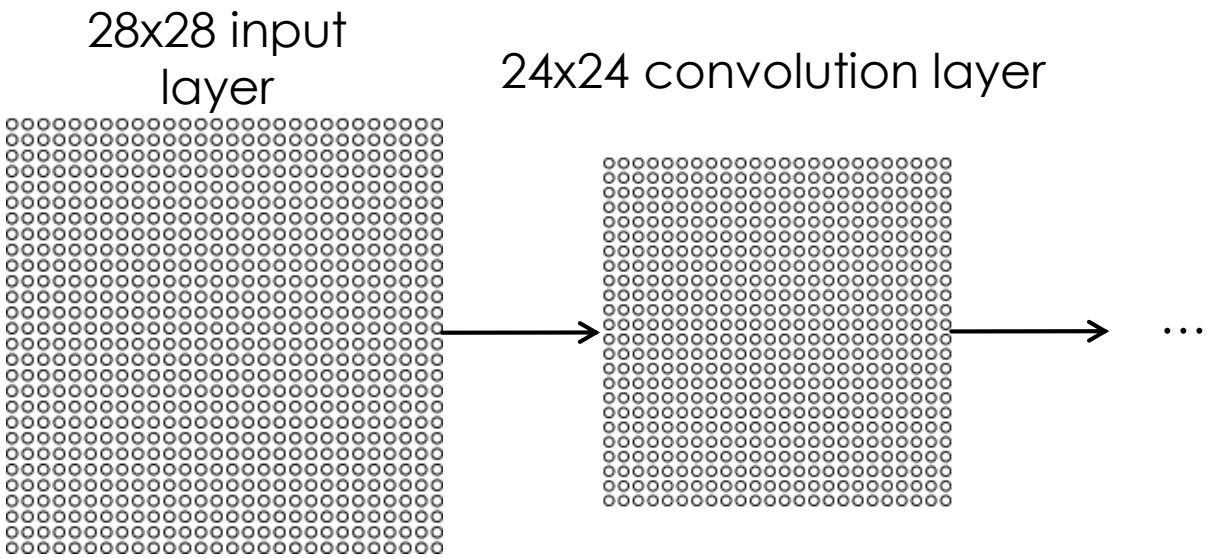
- By sliding this field (region) over the input layer we form the first layer of hidden layer.
- This is a convolutional layer. Guess why?!
- Convolution operation
- Stride length



Local Receptive Field (cont.)

- Consider the followings
  - input as 28x28 pixel image
  - Local receptive field as 5x5 window
- What is the size for the convolutional layer?
  - 24x24

CNN structure so far



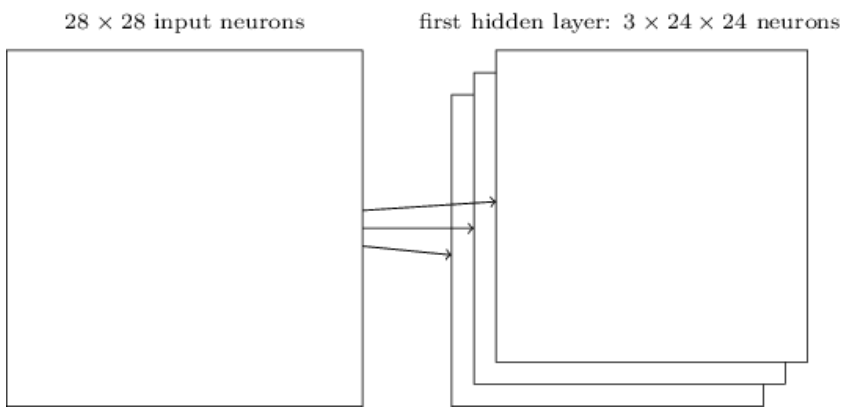
Shared Weight and Biases

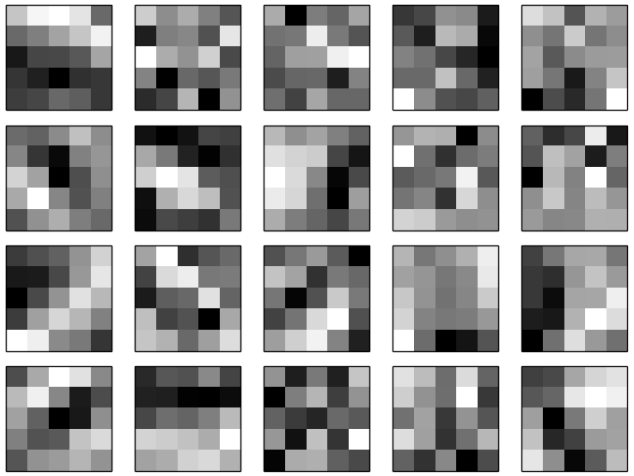
- hidden layer neurons
  - 1 Bias
  - 5x5 = 25 Weights
- Interesting that
  - All the hidden neuron in a single layer shares the weights and bias!
- How this helps?!
- Output for a hidden layer neuron

$$a_{j,k} = f\left(b + \sum_{l=0}^{4}\sum_{m=0}^{4} w_{l,m} \times a_{j+l,k+m}\right)$$

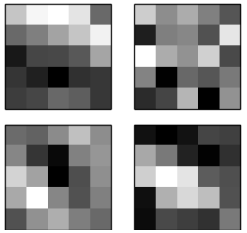
Shared Weight and Biases (cont.)

- So shared weights and biases means
  - All neurons in the first hidden layer detect exactly the same feature in different locations.
  - This feature can be a simple vertical edge
- Now if we have multiple convolutional layers
  - Detect different features





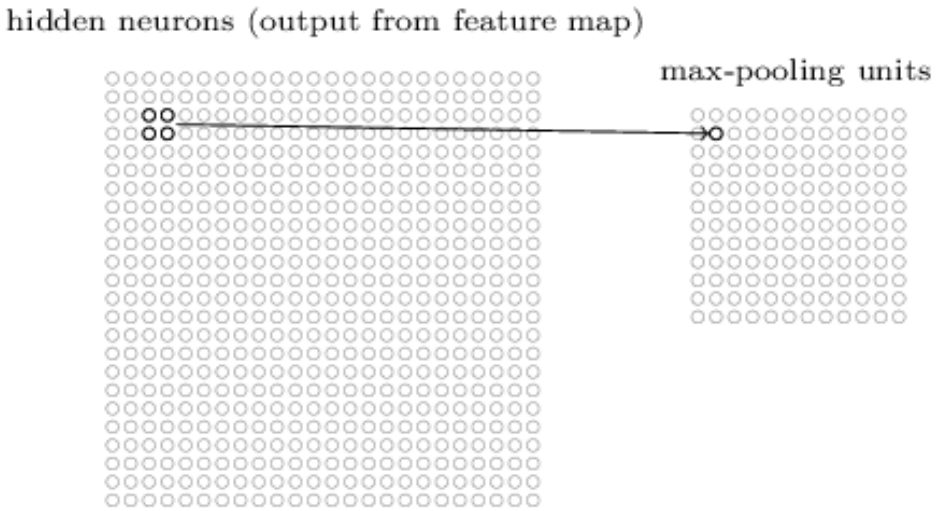
Shared Weight and Biases (cont.)



- Shared weights + bias defines a kernel/filter
- Feature map
  - Map from input layer to hidden layer
  - Containing shared weights and bias
  - Visualizing
  - Darker pixels: larger weights
  - Brighter pixels: smaller weights
- Considering all above so far we are able to
  - Detect multiple features at different locations

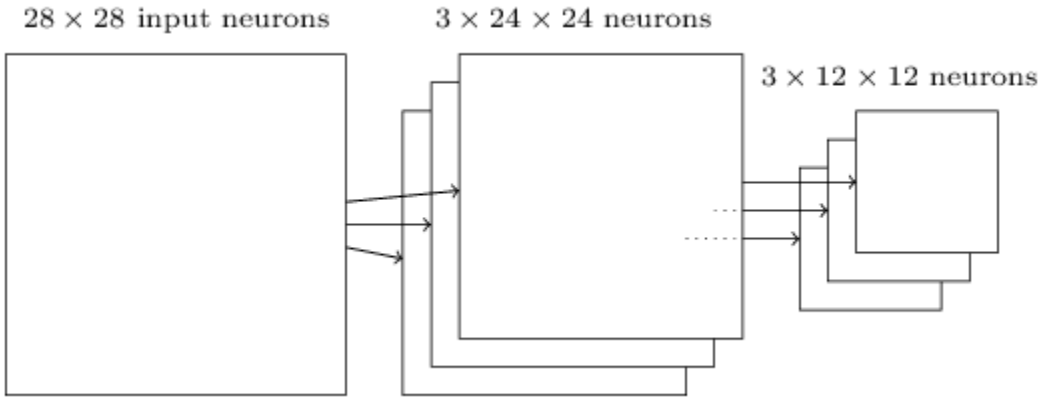
Pooling Layer

- Right after the convolutional layer
- Summarize results in convolution layer
  - Represents finding a feature in a region without any consideration of the exact location



Pooling Techniques

- Max Pooling
- L2 Pooling
- Each map (convolution layer) has a pooling layer



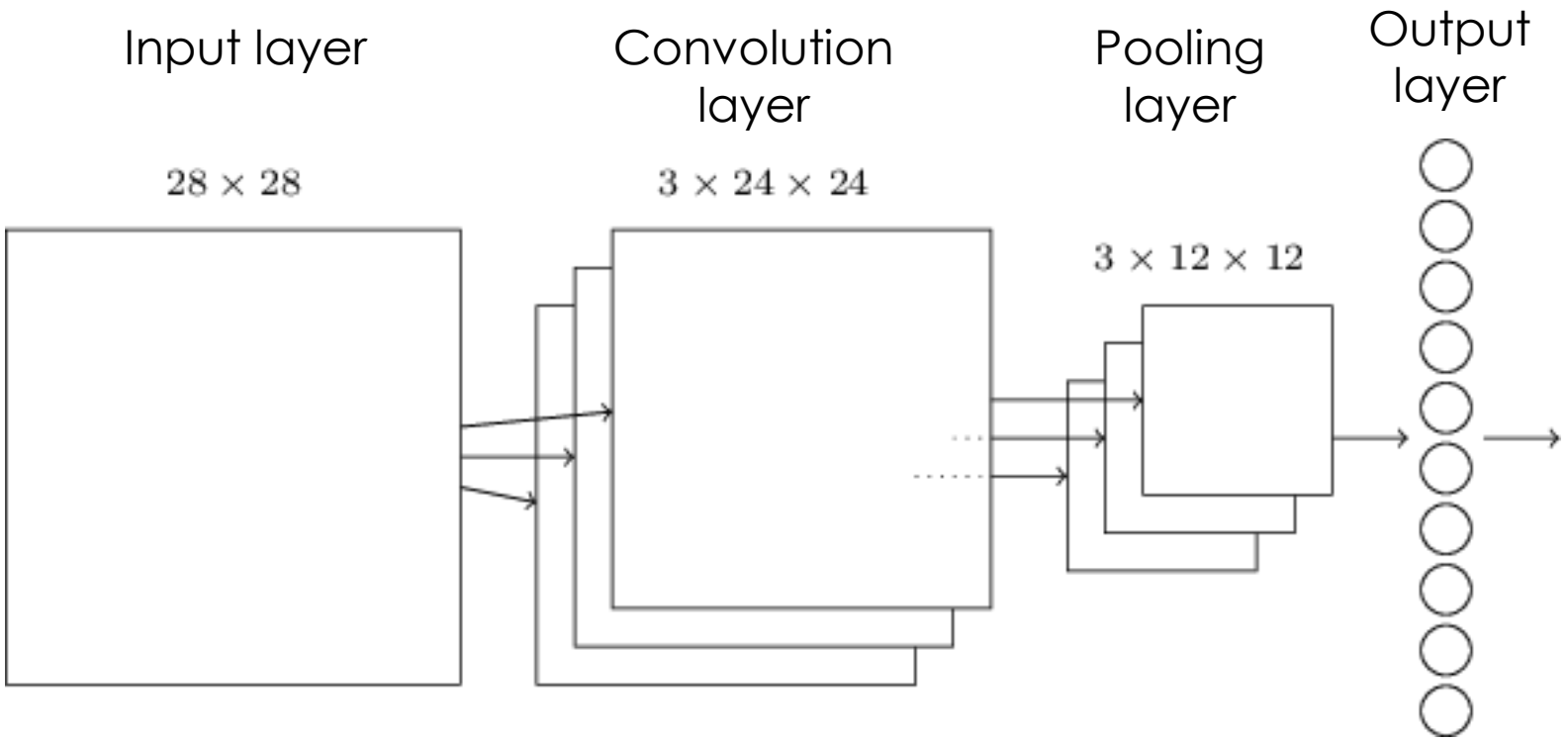
$$\max_{i=1..l} \max_{j=1..k} a_{i,j}$$

$$\sqrt[2]{\sum_{i=1}^{l} \sum_{j=1}^{k} (a_{i,j})^2}$$

Output Layer

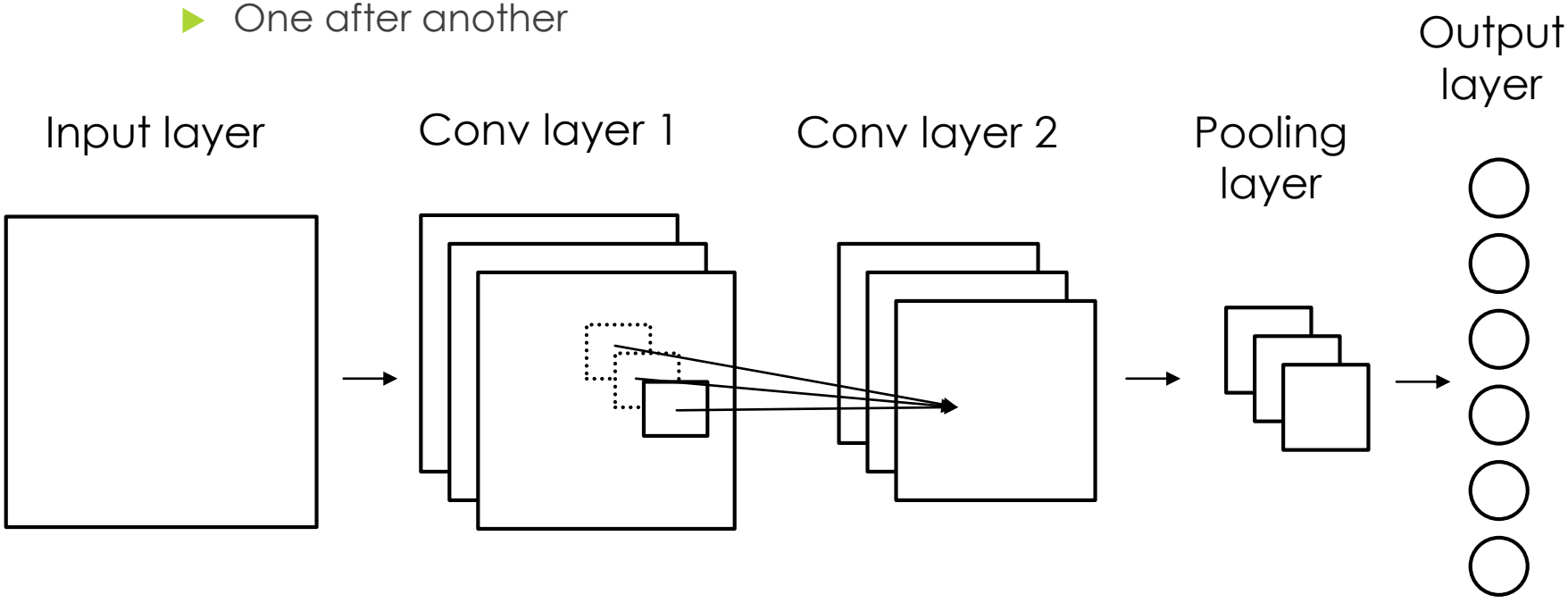
- Pulling all together
- Fully connected network
- Contains N neurons, number of all possible values

Final CNN Structure



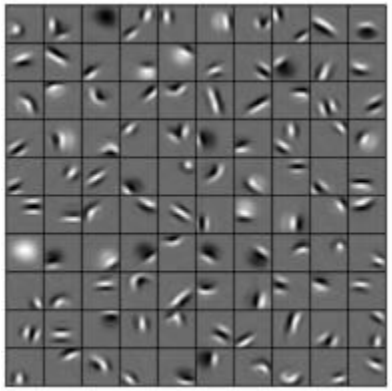
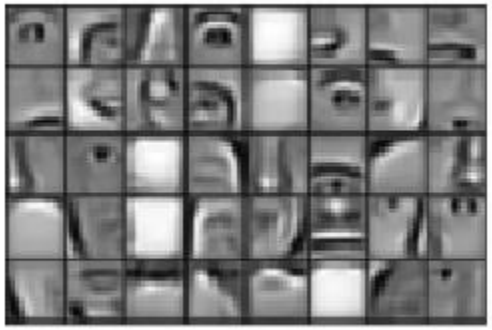
Supplementary Notes

- Training using a modified version of EBP
- We can add more convolutional layers to detect complicated features
  - One after another



Complicated Feature

- By visualizing the Weights matrix in each convolutional layer

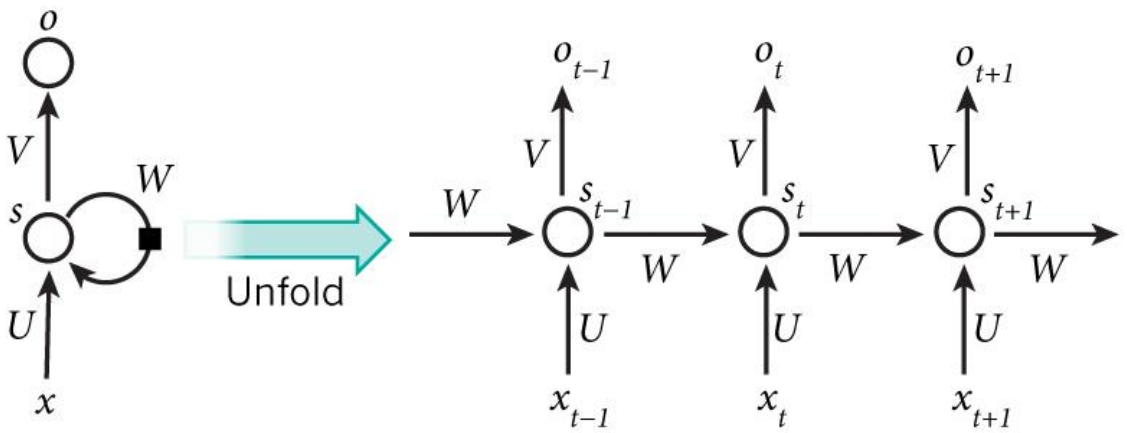




RNN; a short glance

- Recurrent Neural Networks
- Applications in
  - Character recognition
  - Speech recognition
  - NLP
- Use sequential information
- Inputs are not independent from each other
  - Predict next word in a sentence
  - Depends on previous words

RNN; a short glance

- X is the input for each step
- S is the memory unit at each step (Cumulative memory) or "hidden state"
- O is the output at each step
- They are all shared (W,U,V)
- To predict a 5 word sentence, 5 layer is needed.



Resources

- For more resources visit
  - http://neuralnetworksanddeeplearning.com/chap6.html
  - RNN, DL, NLP: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

Thank you for your time