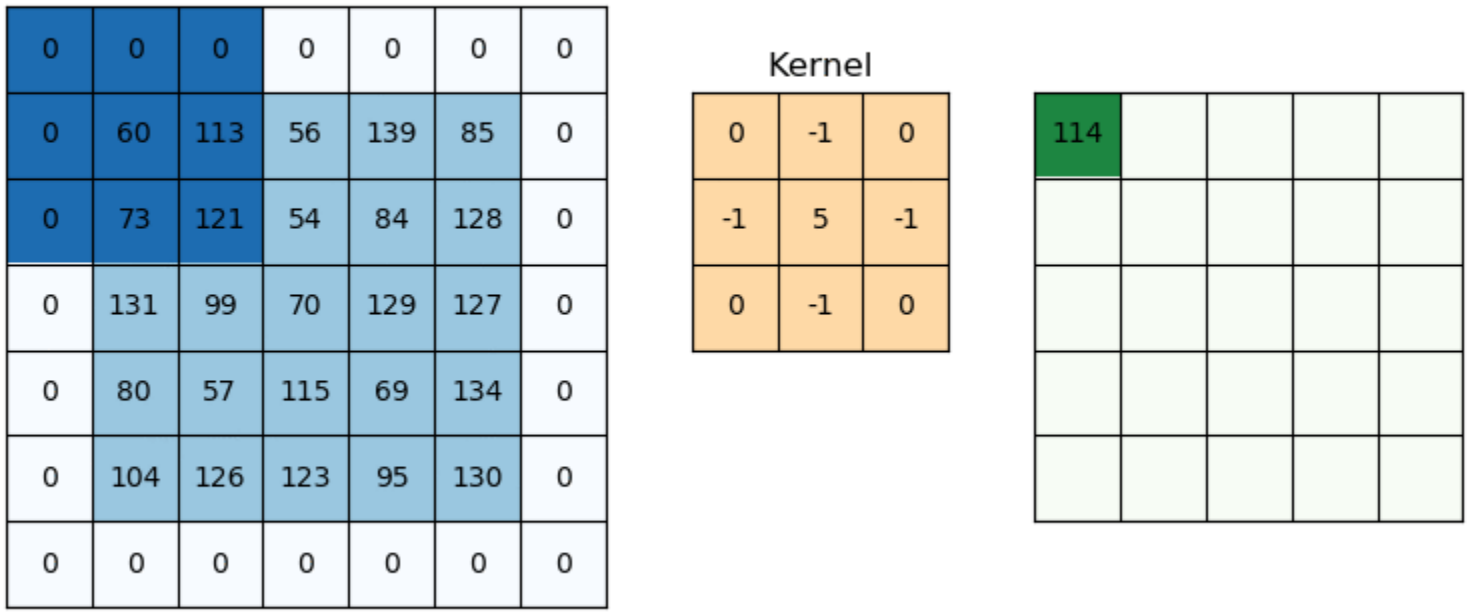
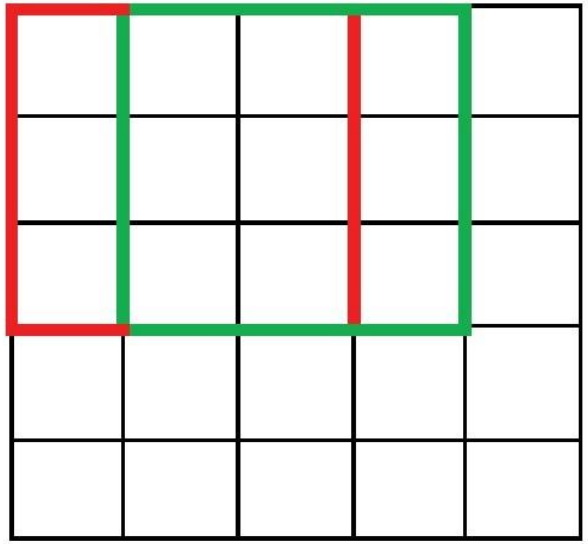
Convolution Operation

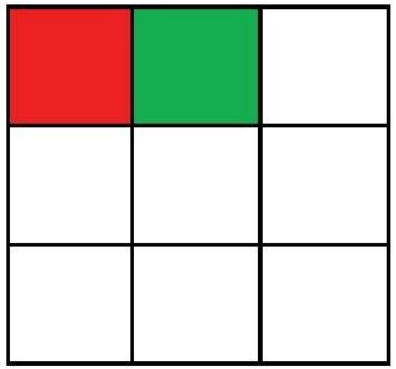


Strides

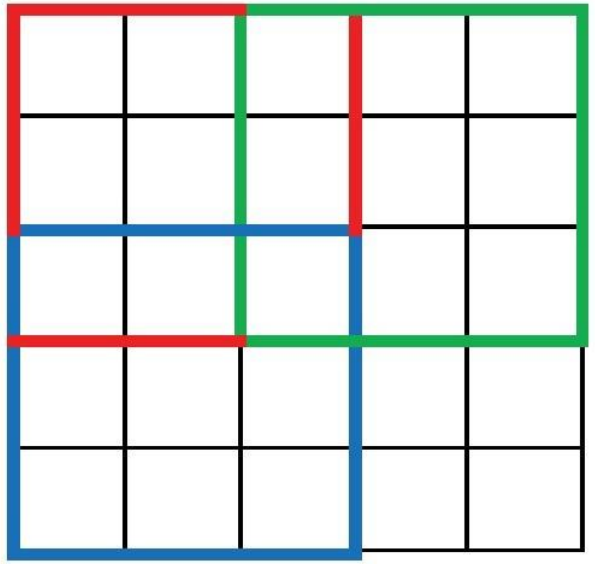
Convolution with Stride=1



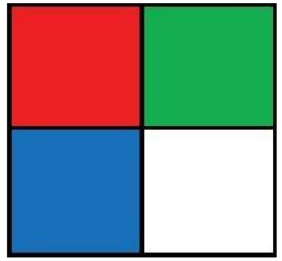
Output



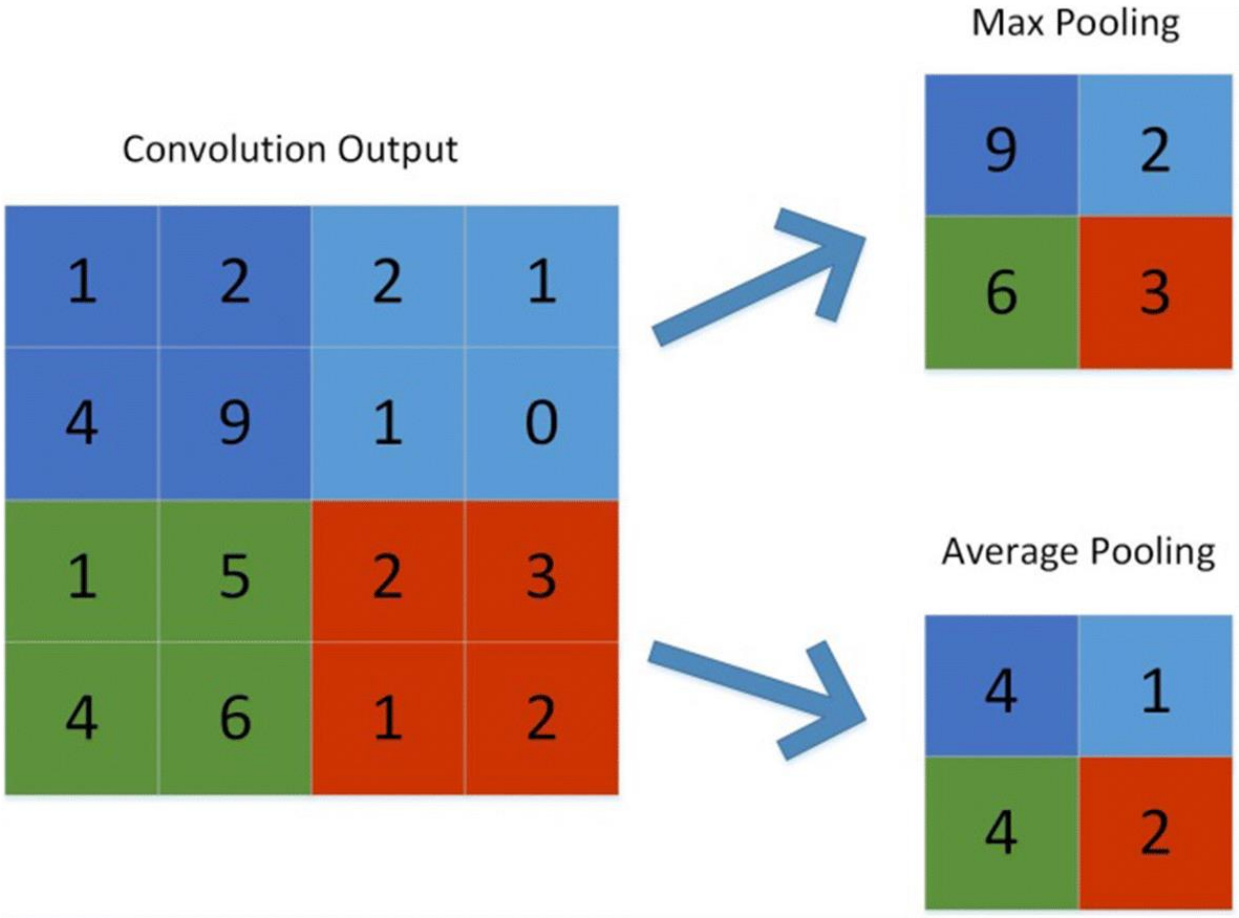
Convolution with Stride=2



Output



Pooling


Convolution Output
1 2 2 1
4 9 1 0
1 5 2 3
4 6 1 2
Max Pooling
9 2
6 3
Average Pooling
4 1
4 2

Loss Function

- Entropy of a random variable X:

$$H(X) = -\sum_{x \in X} P(x) * \log(P(x))$$

- Entropy is a measure of the disorder or randomness of a system.
- More randomness in system, higher the entropy.
  Less randomness, lower the entropy
- Cross-entropy: a measure of “distance” between two distributions.
- In classification problems, you have a true distribution: image belongs to a class with Pr=1 and Pr=0 for all other classes.
- CNN produces an estimate of the probability that the image belongs to each class.
- Cross-entropy : Difference between the true distribution & the CNN predictions.

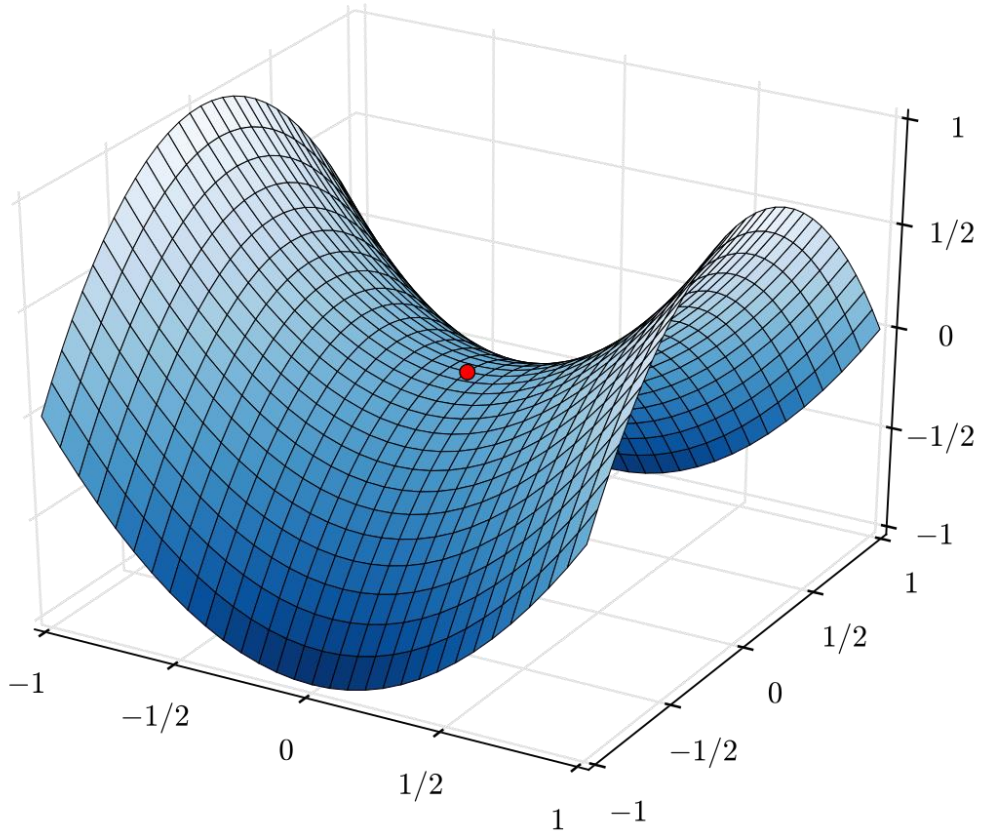
$$H(P, Q) = -\sum_{x \in X} P(X = x) * \log(Q(X = x))$$

- Averaging the cross-entropy loss across the test data give a measure of how well CNN model can

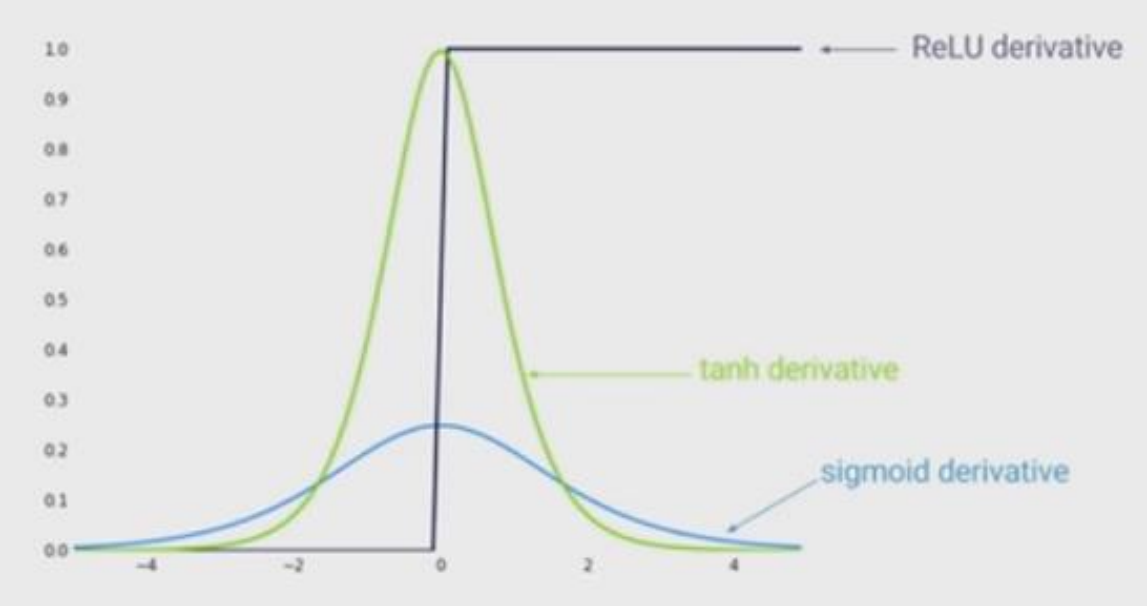
Goal: Minimize the loss

- The goal of training CNN model is to learn a set of weights of the network that can minimizes the loss function.
- Gradient descent is a classic way to solve the above optimization problem.
- In neural networks, backpropagation is used to compute the gradient of loss function w.r.t network weights of all layers.

Saddle Point



Vanishing Gradient Problem



- If the ReLU function is used for activation in a neural network in place of a sigmoid function, the value of the partial derivative of the loss function will be having values of 0 or 1 which prevents the gradient from vanishing.
- The use of ReLU function thus prevents the gradient from vanishing