CS540 Introduction to Artificial Intelligence

Lecture 12

Young Wu

Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

July 8, 2022

Discussion

Admin

- *M*3 bug to be fixed, no need to resubmit.
- *D*1 grades still not fixed.
- Please do not sign up for homework not assigned. The first two (correct) posts will get the points (regardless of the sign up).

add group discussion $\{ 0,5 + 0,5 = 1$

two share solution $0,5 + 0,5 = 1$

SIFT and HOG Features

Motivation

- SIFT and HOG features are expensive to compute.
- Simpler features should be used for real-time face detection tasks.

Real-Time Face Detection

Motivation

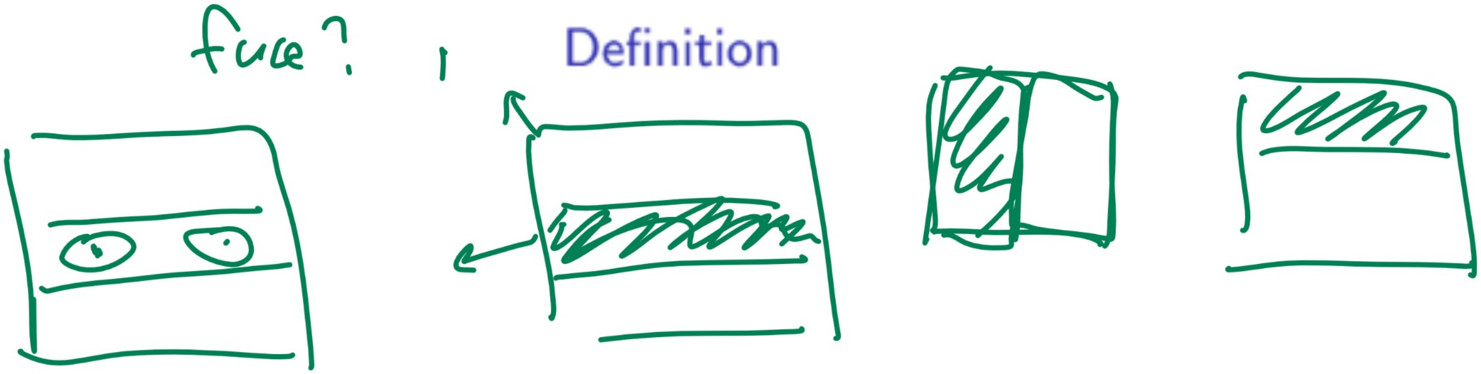
- Each image contains 10000 to 500000 locations and scales.
- Faces occur in 0 to 50 per image.
- Want a very small number of false positives.

Haar Features Diagram

Motivation

Haar Features

Definition



- Haar features are differences between sums of pixel intensities in rectangular regions. Some examples include convolution with the following filters.

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \cdots$$

Weak Classifiers

Definition

P2

- Each weak classifier is a decision stump (decision tree with only one split) using one Haar feature x.

$$f(x) = \mathbb{I}_{\{x>\theta\}}$$

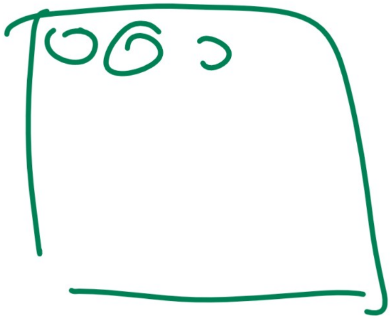
which $\theta \rightarrow$ max IG

- Finding the threshold by comparing the information gain from all possible splits is too expensive, so θ is usually computed as the average of the mean values of the feature for each class.

$$\theta = \frac{1}{2}\left(\frac{1}{n_0}\sum_{i:y_i=0} x_i + \frac{1}{n_1}\sum_{i:y_i=1} x_i\right)$$

Strong Classifiers

Definition



1 ⟶ not face

2 ⟶ 1 maybe face ⟶ not face

3 ⟶ 2 maybe face ⟶ not face

- The weak classifiers are trained sequentially using ensemble methods such as AdaBoost.
- A sequence of T weak classifiers is called aT -strong classifier.
- Multiple T -strong classifiers can be trained for different values of T and combined into a cascaded classifier.

Cascaded Classifiers

Definition

- Start with aT -strong classifier with small T, and use it reject obviously negative regions (regions with no faces).
- Train and use aT -strong classifier with larger T on only the regions that are not rejected.
- Repeat this process with stronger classifiers.

Cascading

Definition

not face → not-face
→ face
face → face

- For example, at $T = 1$, the classifier achieves a100 percent detection rate and a50 percent false-positive rate.
- At $T = 5$, the classifier achieves a100 percent detection rate and a40 percent false-positive rate.
- At $T = 20$, the classifier achieves a100 percent detection rate and a10 percent false-positive rate.
- The result is a cascaded classifier with 100 percent detection rate and $0.5 \cdot 0.4 \cdot 0.1 = 2$ percent false positive rate.

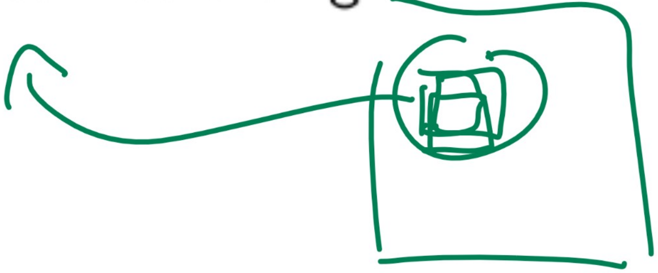
Viola-Jones

Discussion

boosting

Haar + Decision Stump.

- Each classifier operates on a 24 by 24 region of the image.
- Multiple scales of the image with a scaling factor of 1.25 are used. The classifiers can be scaled instead in practice so that the integral image only needs to be calculated once.
- The detector is moved around the image with stride 1.
- Nearby detections of faces are combined into a single detection.



Viola-Jones Diagram

Discussion

Learning Convolution

Motivation

weights

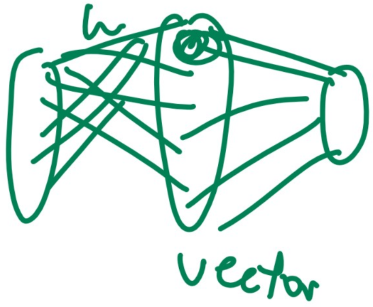
- The convolution filters used to obtain the features can be learned in a neural network. Such networks are called convolutional neural networks and they usually contain multiple convolutional layers with fully connected and softmax layers near the end.

Convolutional Layers

Definition

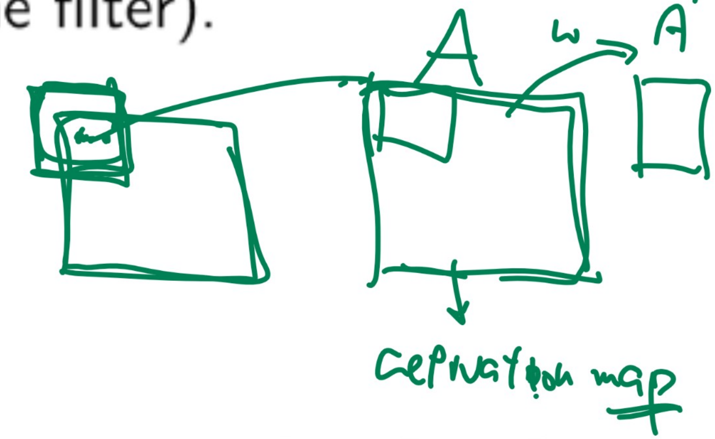
- In the (fully connected) neural networks discussed previously, each input unit is associated with a different weight.

$$\underset{\substack{a_1\\a_2}}{a} = g\left(\underbrace{w^T x}_{\text{dot product}} + b\right)$$



- In the convolutional layers, one single filter (a multi-dimensional array of weights) is used for all units (arranged in an array the same size as the filter).

$$A = g\left(\underbrace{W * X}_{\text{Convolution}} + b\right)$$



$2D$ Convolutional Layer Diagram

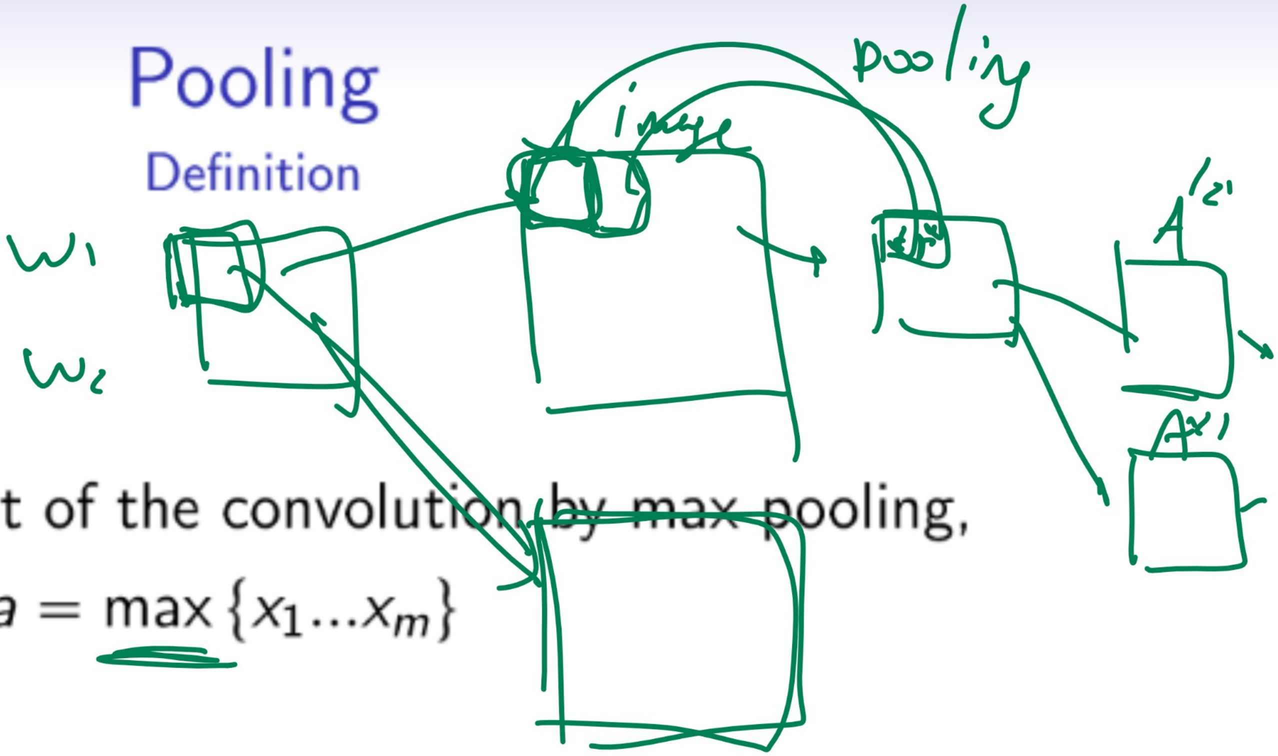
Definition

$3D$ Convolutional Layer Diagram

Definition

Pooling

Definition



- Combine the output of the convolution by max pooling,

$$a = \underline{\max}\left\{x_1 \ldots x_m\right\}$$

- Combine the output of the convolution by average pooling,

$$a = \frac{1}{m}\sum_{j=1}^{m} x_j$$

Pooling Diagram

Definition

Training Convolutional Neural Networks, Part I

Discussion

- The training is done by gradient descent.
- The gradient for the convolutional layers with respect to the filter weights is the convolution between the inputs to that layer and the output gradient from the next layer.

$$\longrightarrow \quad \frac{\partial C}{\partial W} = \underline{X * \frac{\partial C}{\partial O}}$$

$$a = g(w^T x + b)$$

$$A = g(\underline{w * X} + b)$$

- The gradient for the convolutional layers with respect to the inputs is the convolution between the 180 degrees rotated filter and the output gradient from the next layer.

$$\underline{\frac{\partial C}{\partial X}} = \underline{\text{rot } W} * \underline{\frac{\partial C}{\partial O}}$$

Training Convolutional Neural Networks, Part II

Discussion

- There are usually no weights in the pooling layers.
- The gradient for the max-pooling layers is 1 for the maximum input unit and 0 for all other units.
- The gradient for the average pooling layers is $\frac{1}{m}$ for each of the m units.

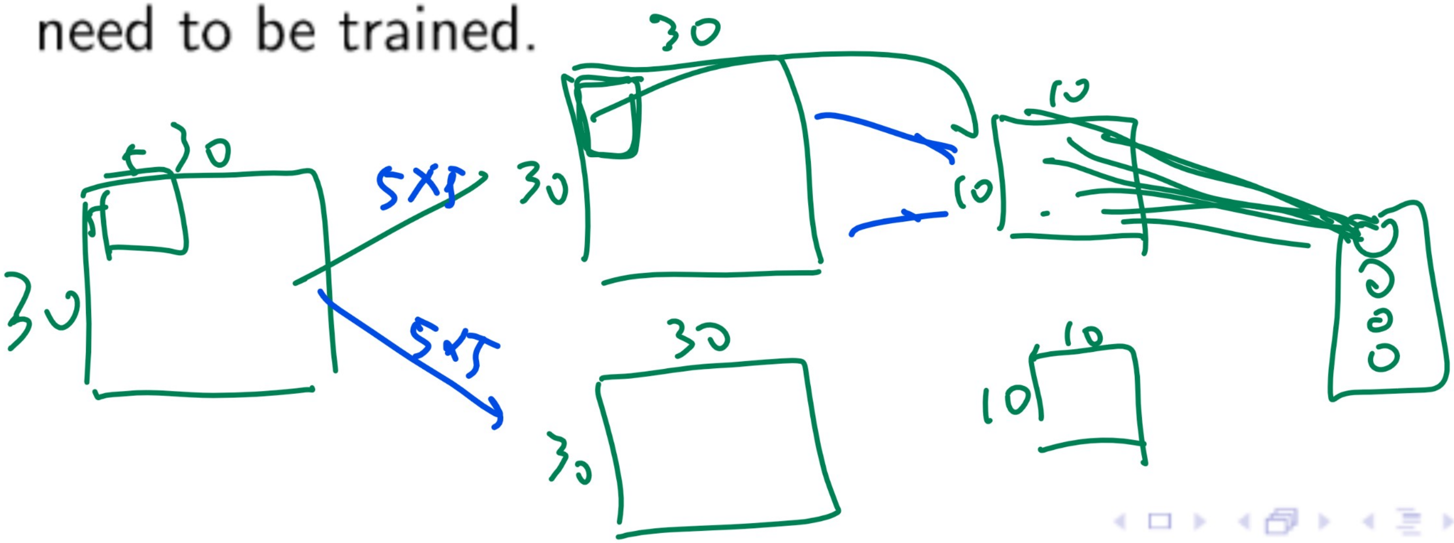
LeNet Diagram and Demo

Discussion

Convolutional Neural Network Weights 1

Quiz

- Given a CNN with 30 × 30 input images, with 5 × 5 filters, zero-padding, stride 1, and two activation maps in the first layer, then 3 × 3 max pooling, no padding, stride 3 in the second layer, 4 output units in the last fully-connected layer. What is the number of weights (not including biases) that need to be trained.



$2 \cdot 5 \cdot 5 + 0 + \underline{2 \cdot 10 \cdot 10 \cdot 4} =$

Convolutional Neural Network Weights 2

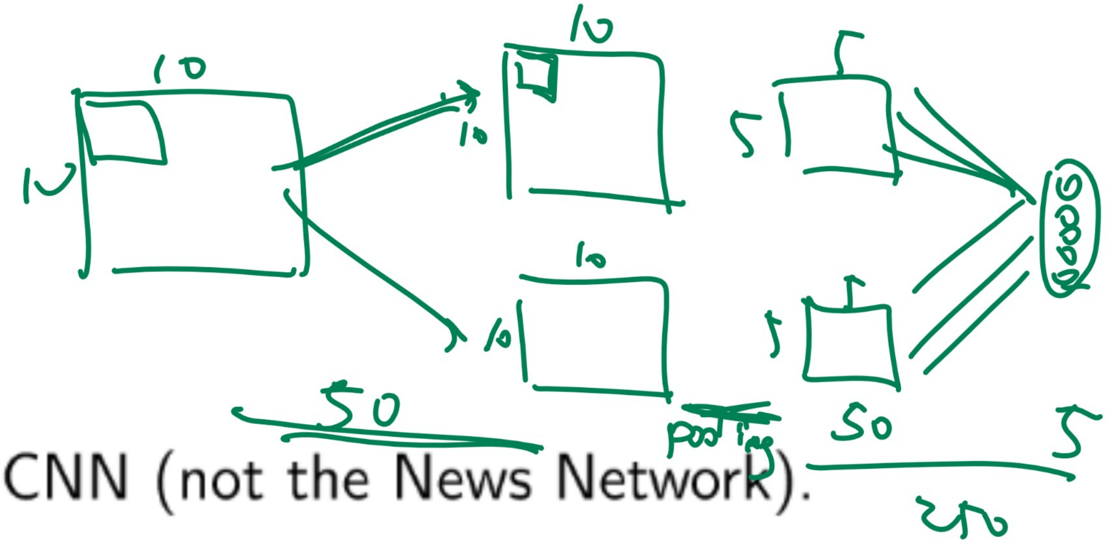
Quiz

- Given a CNN with 10 × 10 input images, with 5 × 5 filters, zero-padding, stride 1, and two activation maps in the first layer, then 2 × 2 max pooling, no padding, stride 2 in the second layer, 5 output units in the last fully-connected layer. What is the number of weights (not including biases) that need to be trained.

avg

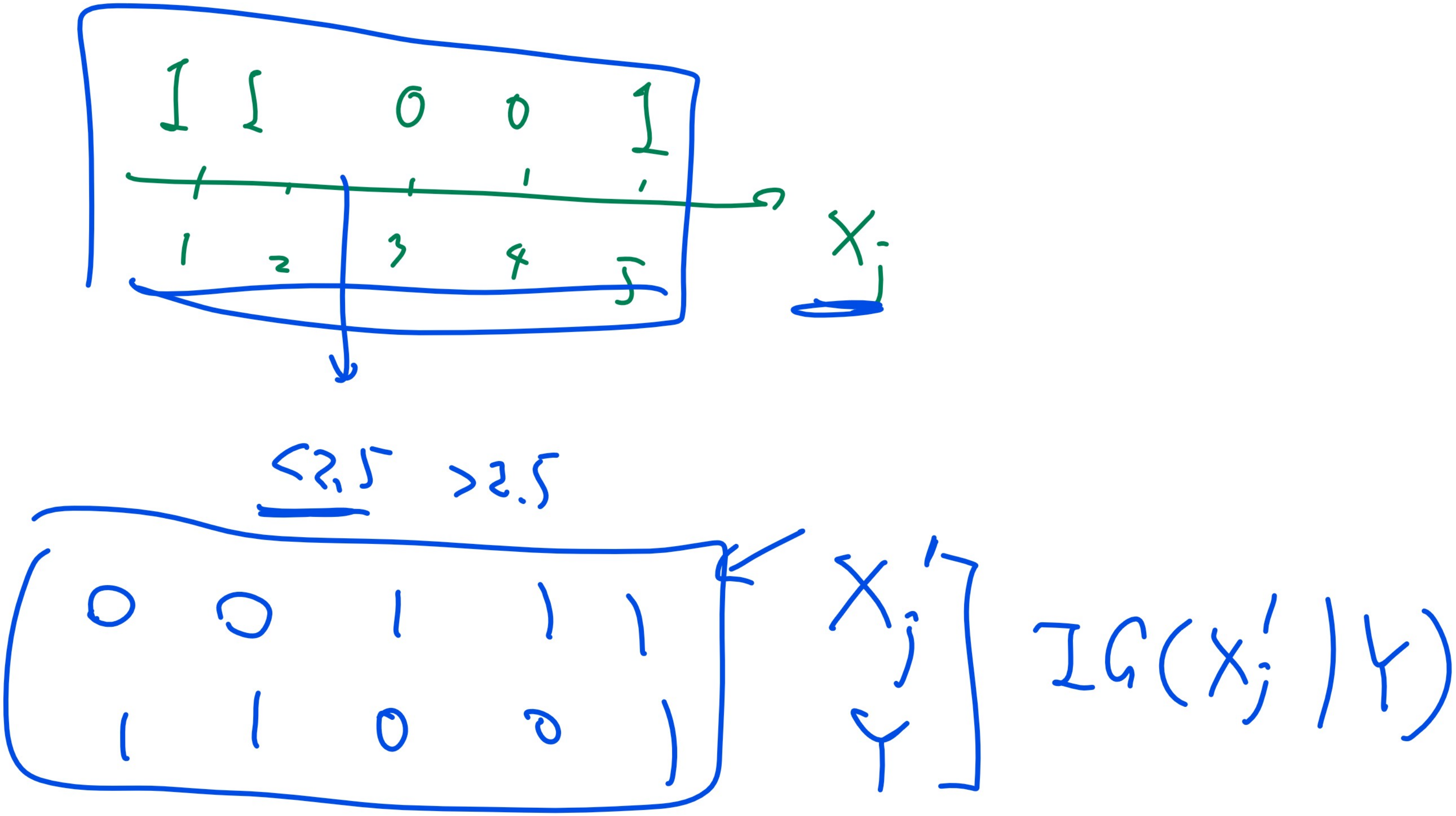
Q3

- $A : 25 + 0 + 125$
- $B : 50 + 0 + 250$
- $C : 25 + 4 + 125$
- $D : 50 + 8 + 250$
- E : *I* don't understand CNN (not the News Network).



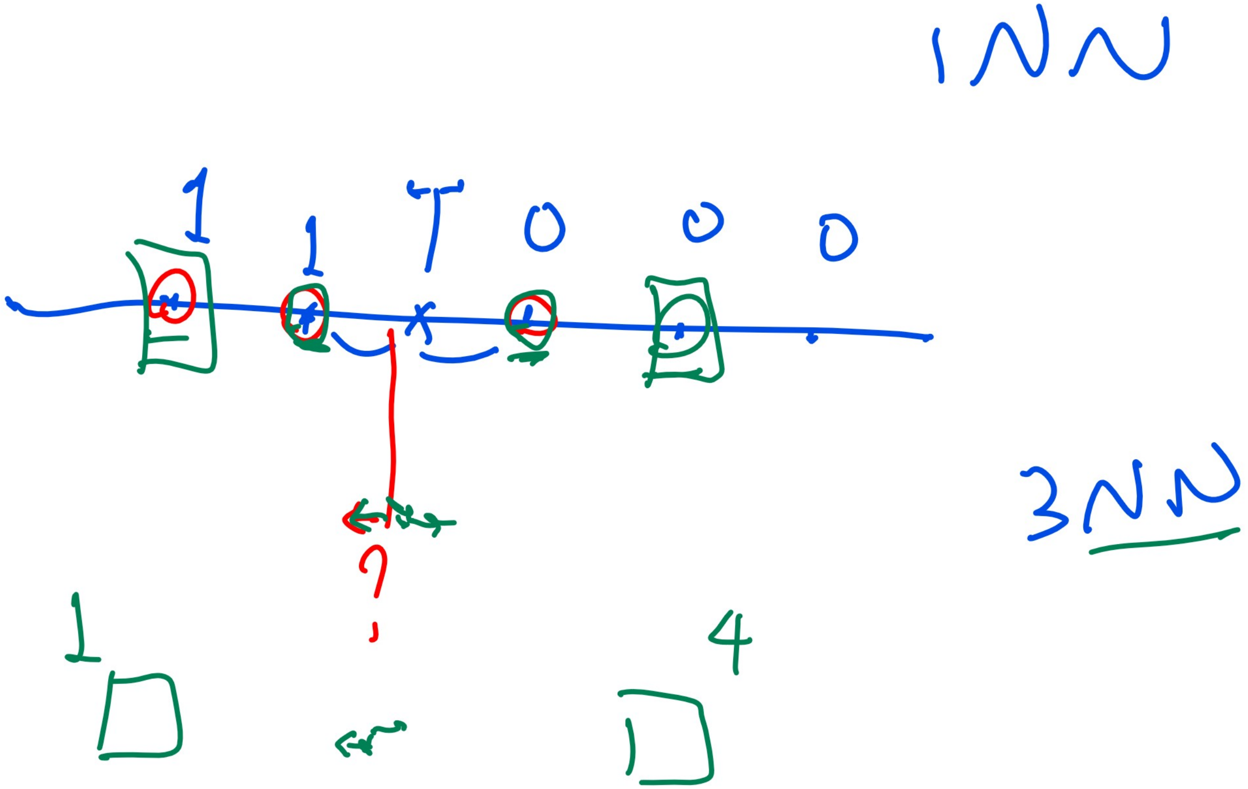
AlexNet Diagram

Discussion



VGG, GoogleNet, ResNet

Discussion


1NN
1
1
T
0
0
0
3NN
?
1
4