CMPE258 April 5, '22

X_4: A Class of Features for the size of Bounding Boxes.

Too Small → Noise, filter out

Too Big → UnLikely the Right ROI.

$X_1 \; X_2 \; X_3 \; X_4 \quad f$

✓ (under X_1) ✓ (under X_3)

$f = X_1 X_3$

Note: Use OpenCV Function to Construct A Square Image.

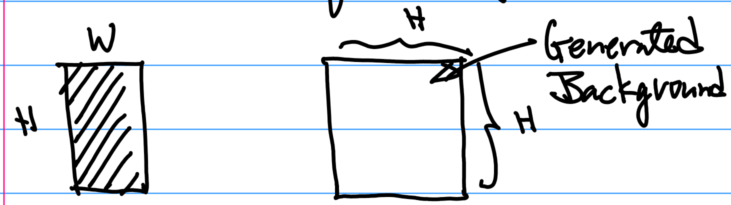
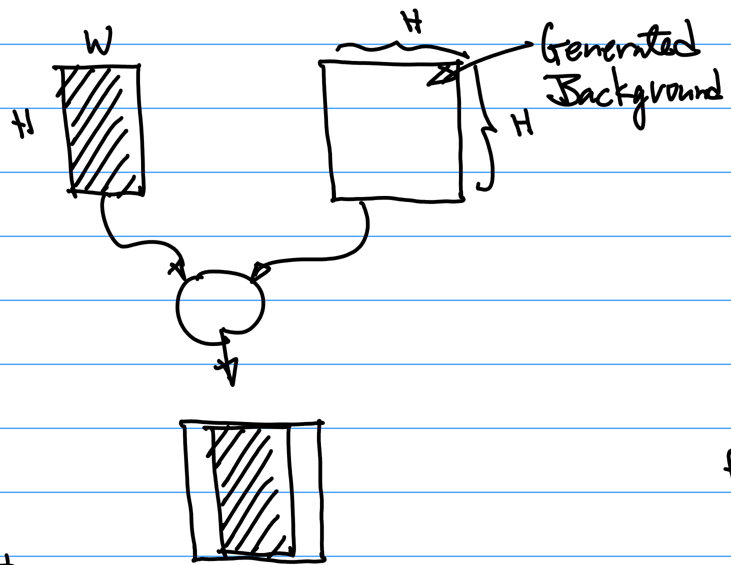


Fig 2.

$\max\{H, W\} = H$, use H as dimension for Square Image.



Then, Reduce/Resize the Image to the right size, 28x28

c. Ref: Reshape the image Format

20-2021S-3-load-deployment.py

```python
from keras.models import load_model
import cv2
import numpy as np
from PIL import Image

model = load_model('mnist.h5')
model.compile(optimizer='rmsprop',
              loss='categorical_crossentropy',
              metrics=['accuracy'])

img = Image.open('1.jpg').convert('L')
img = np.resize(img, (28,28,1))
im2arr = np.array(img)
im2arr = im2arr.reshape(1,28*28)
y_pred = model.predict_classes(im2arr)
print(y_pred)
```

d. Save Detection Result (Video Clips, 5~10 sec), Be Sure to use your personal SID (Just Last 4 Digits)

Topic on the 2nd Part of this Class Yolo (You Only Look Once)

Source Code distribution for Yolo

Reference Paper On Yolo

2022S-112-yolo-paper.pdf

Introduction to Yolo Algorithm.

Example:

CMPE258 April 5, 22

Note: a. Input Image: Square Image, Resolution of the image: 448x448

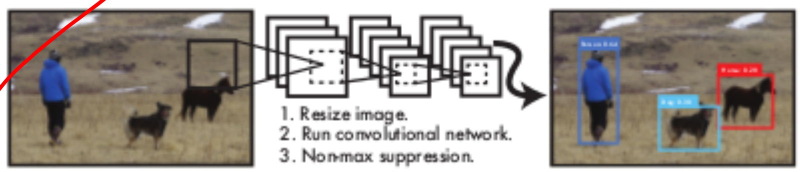
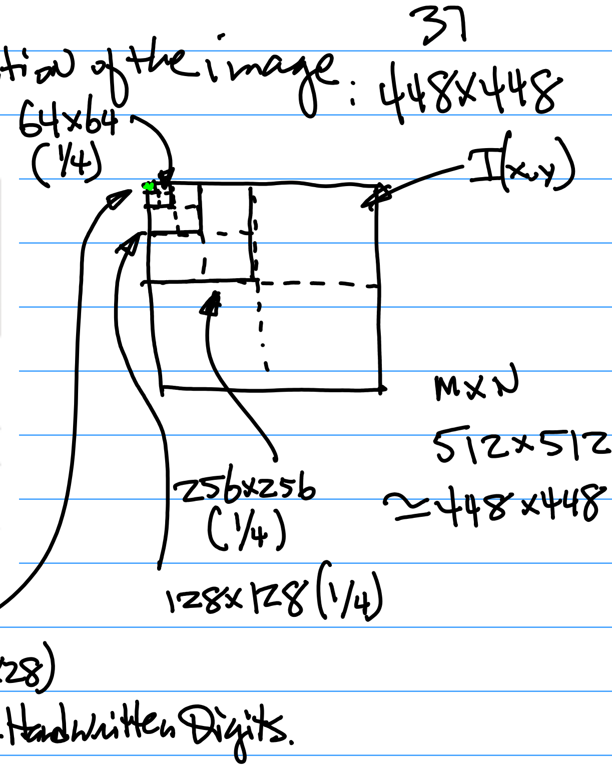


Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.



Note: Architecture, 24 Convolutional Layers. 2 fully Connected layers.

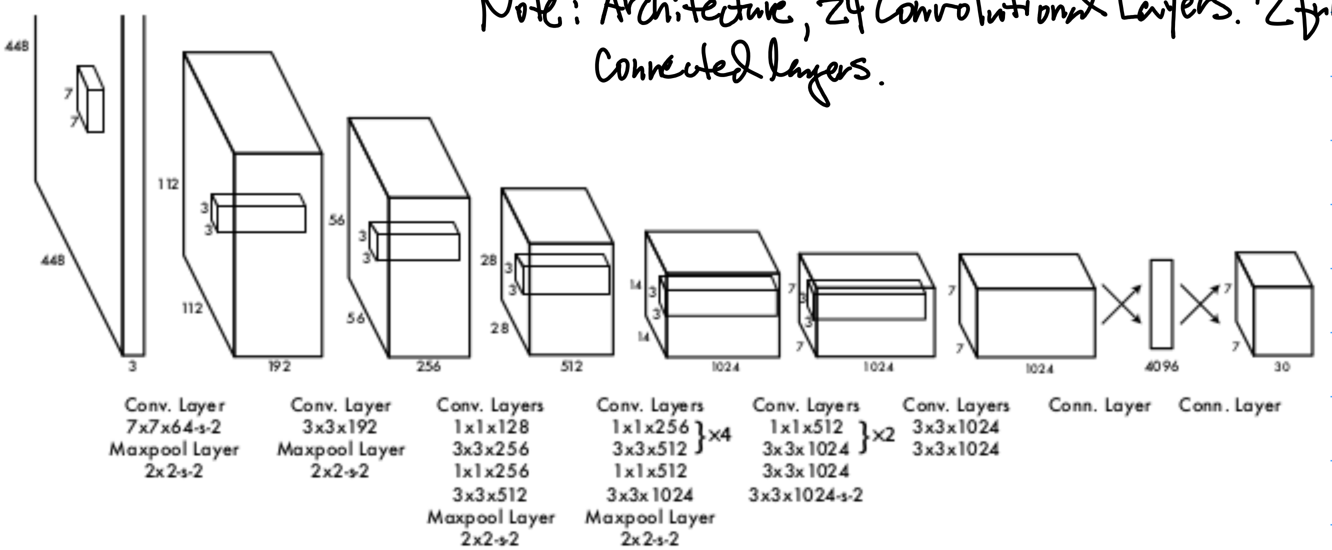


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

Source Distribution, See Handout to Be Posted online (github).

1. Readme Document.

   a. URL for Yolo Source distribution Down Load

   b. anaConda to Configure the environment

   c. Activate the Conda environment

   d. 3 Samples to test

CmpE258 April 5, 22

First Example, to perform Training

Save the training Result;

2nd works on input image (Single Frame).

3rd program works on input video file.

Homework: Due 2 weeks from today. (April 19th).

1° Down Load, Install yolo program. (yolo4).

2° Use cellphone to record 5-10 Seconds of video clip for testing Purpose.

3° Run Yolo Code to Perform default Detection Task

4° Submission on CANVAS.

a. Screen Capture of yolo program

b. Processed Video Clip (Be Sure to use your own Video Clip).

Example: Loss Function (Brief)

Note:

loss function:

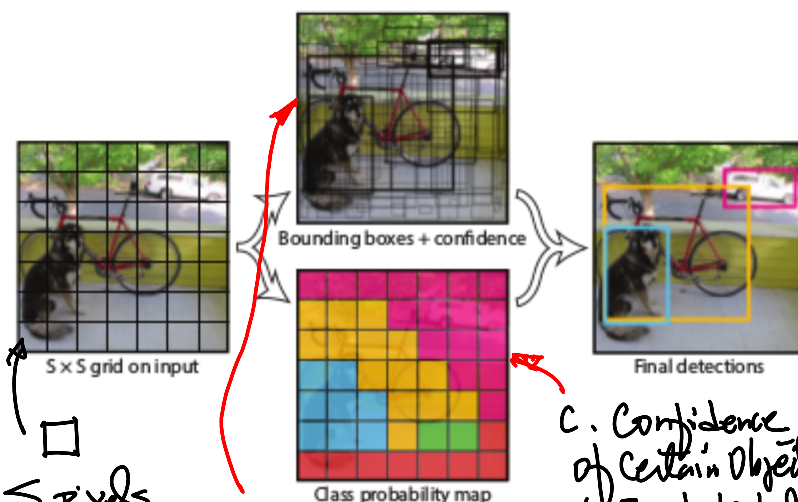
$$
\begin{aligned}
&\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
&+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2
\end{aligned} \tag{3}
$$

Example: On Probability Classification.

$$
\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}} \tag{1}
$$

Definition of Conditional Probability.

Example: On Partition of a Test Image



a. S pixels Both w, h.

b. Bounding Box(es)

c. Confidence of Certain Objects to Be detected

Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

d. The Classification of Object Detected is Based on its Confidence Level.

April 12, 22

Topics: 1° DNN Yolo Formulation

2° Back-propagation (BackProp) Algorithm.

2022S-112-yolo-paper.pdf

**You Only Look Once:
Unified, Real-Time Object Detection**

Joseph Redmon*, Santosh Divvala*†, Ross Girshick¶, Ali Farhadi*†
University of Washington*, Allen Institute for AI†, Facebook AI Research¶
http://pjreddie.com/yolo/

$$\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}^{\text{truth}}_{\text{pred}} = \Pr(\text{Class}_i) * \text{IOU}^{\text{truth}}_{\text{pred}} \quad (1)$$

Theoretical Analysis of Yolo.

Simple Example As A Starting Point of Our Discussion, from Eqn (1)

Note: 1. Prob(Class$_i$)

Probability of a given Object(s) which belongs to a class i

$i = 1, 2, \cdots, K$

Let $i=1$ for Simplicity purpose

Prob(Class$_i$) = Prob(Class)

2. IOU (Intersection of Union)

η_{IOU} Coefficient

$$\text{Prob(Class)}\,\eta_{IOU} \xrightarrow{\text{Simplify}} \text{Prob(Class)}$$

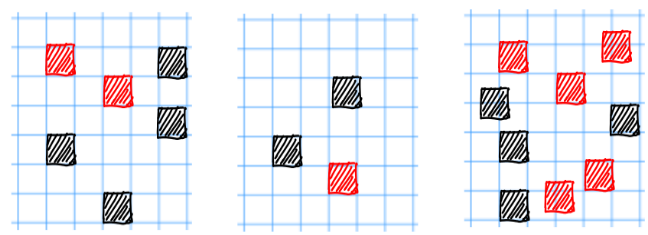
Prob(C), Denote Class as "C".

Example: Consider Object Identification, In particular, Detect Pedestrain.

Ref:

2022S-114-yolo-designGuide-2022-4-9.pdf

Red, R: Pedestrains;
Black, B: Vehicle;



CAM$_1$ $I_1(x,y)$ CAM$_2$ $I_2(x,y)$ CAM$_3$ $I_3(x,y)$

Fig 1.

1. Define An Event R
Detection of Pedestrain.

Set Relations.

$$R = RI_1 + RI_2 + RI_3 \quad \ldots (1^*)$$

$R \cap R_1$, $I_1 \cap I_2 \cap I_3 = \phi$ Empty

2. Formulate the Likelihood of the event R

$$\text{Prob}(R) = \text{Prob}(RI_1) + \text{Prob}(RI_2) + \text{Prob}(RI_3) \quad \ldots (2)$$

3. Conditional Probability, Bayesian Theory

$$\text{Prob}(RI_1) = \text{Prob}(R|I_1)\,\text{Prob}(I_1) \quad \ldots (3)$$

$$\text{Prob}(RI_2) = \text{Prob}(R|I_2)\,\text{Prob}(I_2)$$

$$\text{Prob}(RI_3) = \text{Prob}(R|I_3)\,\text{Prob}(I_3)$$

4. $$\mathrm{Prob}(R) = \mathrm{Prob}(R|I_1)\mathrm{Prob}(I_1) + \mathrm{Prob}(R|I_2)\mathrm{Prob}(I_2) + \mathrm{Prob}(R|I_3)\mathrm{Prob}(I_3)$$
$$= \sum_{i=1}^{3} \mathrm{Prob}(R|I_i)\mathrm{Prob}(I_i) \quad \cdots (4)$$

5. Notations for Yolo

①

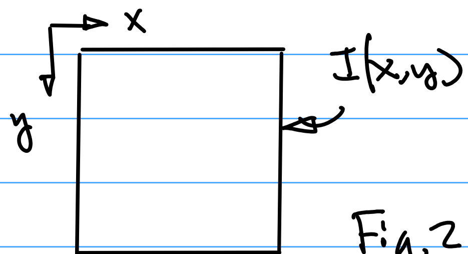


Fig.2

② Define Grids $\underset{S\times S}{G(x,y)}$

↑

S pixel width, S pixel Height

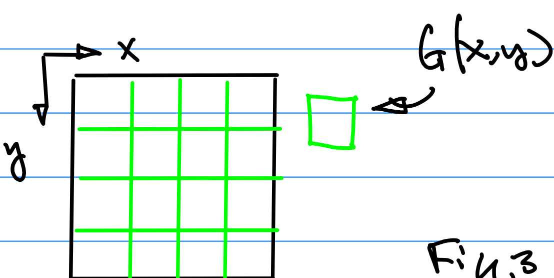


Fig.3

we have multiple grids, denoted as

$$G_p(x,y), \quad p=1,2,\cdots \quad \cdots (5)$$

③ Define Bounding Box for ROI. as $B(x,y)$

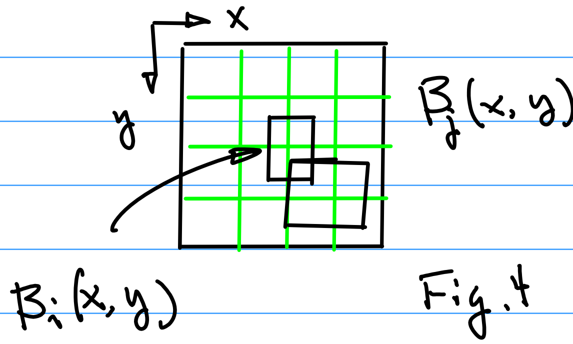


Fig.4

Note:

$$B_i(\underbrace{x,y}_{\text{Location}};\underbrace{W,H}_{\text{shape}};\underbrace{f}_{\text{Confidence}}), \text{ for } i=1,2,\cdots,\underbrace{M}_{\text{No. of B. Boxes}}$$

$$B_j(x,y;W_j,H_j;f_j) \quad \cdots (6)$$

Note: Among them, we have one ground Truth Bounding Box

$B_{GT}(x,y;W,H,f)$, $f=1$ for the Ground Truth.

④ IOU Definition.

Example. Illustration of IOU

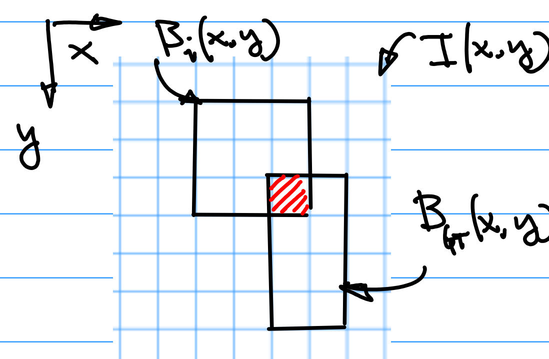


Fig.5

Intersection of Union

$$\mathrm{IOU} = \frac{\text{Intersection}}{\text{Union}} \quad \cdots (7)$$

$$\text{IOU} = \frac{\text{Intersection}}{\text{Union}}$$

$$= \frac{B_i(x,y) B_{GT}(x,y)}{B_i(x,y) + B_{GT}(x,y) - B_i(x,y) B_{GT}(x,y)}$$

$$= \frac{1}{9+8-1} = 1/16 \quad //$$

Physical meaning of IOU.

If $B_i(x,y)$ & $B_{GT}(x,y)$ Are Identical then

$$\text{IOU} = \frac{B_i(x,y) B_{GT}(x,y)}{B_i(x,y) + B_{GT}(x,y) - B_i(x,y) B_{GT}(x,y)}$$

$$= 1 \quad //$$

Denote IOU as η_{IOU}

From Eqn (1)

$$\text{Prob}(C)\eta_{IOU} \quad \ldots (8)$$

Note if more than One class,

$$\text{Prob}(C_i) = \text{Prob}(C_i)\eta_{IOU_i}$$

6. Consider Multiple Classes.

One class, from one Camera

$$\text{Prob}(C)\eta_{IOU}$$

One class, from multiple Cameras, from Eqn

$$\text{Prob}(C) = \sum_i \text{Prob}(C/I_i)\text{Prob}(I_i)$$

Hence

$$\text{Prob}(C_i)\eta_{IOU,i} = \sum_{i=1}^{K} \text{Prob}(C/I_i)\text{Prob}(I_i)\cdot\eta_{IOU_i} \quad \ldots (9)$$

Note: Now Expand the Above Analysis Beyond Multiple Cameras, treat Each Camera as input Grid $\underset{\sim}{G}(x,y)$ from a given image, the Above Analysis holds good.

Homework: Due 1 week from today. (April 19th).

1° Down Load, Install yolo program. (yolo4).

2° Use cellphone to record 5-10 Seconds of video clip for testing Purpose.

3° Run Yolo Code to Perform default Detection Task

Team Project:

1° Presentation PPT. 7 slides 10 minutes for each team entire entire presentation. PPT.; Demo; Q&A Session

2° Topics, Deep Learning CNN Applications.

Consider Back Propagation Algorithm:

Background

1. Multiple Layer Feedforward Neural Networks. MNIST

CM, ... F.D.D

$\underbrace{\qquad\qquad\qquad}$

Layers

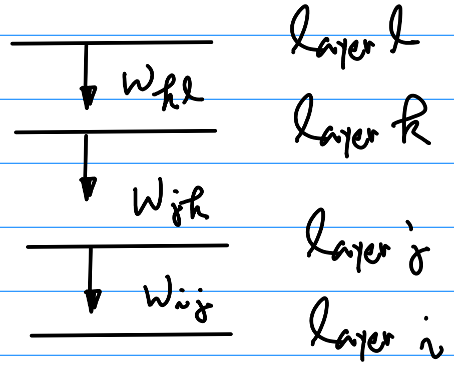


Fig 6

2. For Each Layer.

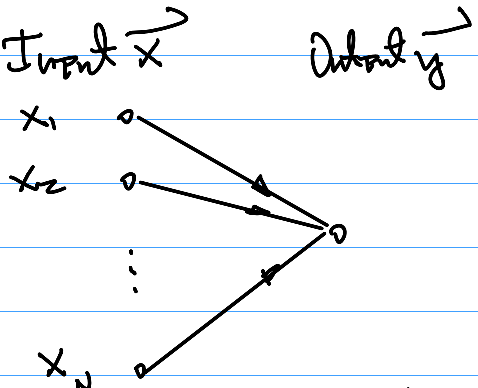
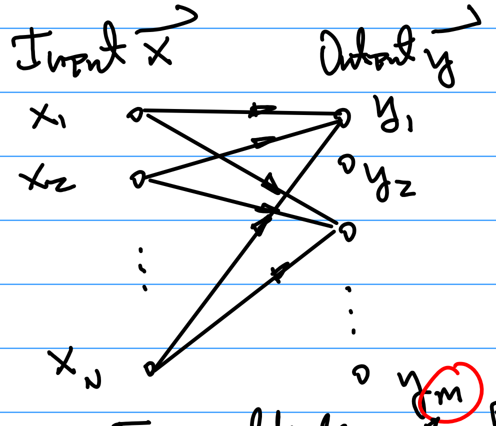


Fig. 7

$$y = f\left(\sum_{i=1}^{N} w_i x_i + b\right) = f(h(w_i, x_i, b)) \quad \ldots (1)$$



For multiple Output Neurons.

Fig. 8

3. Notation for Weights at Each Layer, From the Output layer i towards the input j, k, and l

W_{ij}

The Index of the Output @ the Current Output layer i

The Index of the Input @ layer j

So, we have

W_{ij} W_{jk} W_{kl}

3. Output

$$y_j = f\left(\sum_{i=1}^{N} w_i x_i + b\right) \text{ for one Output Neuron.} \quad \ldots (2)$$

Output for multiple Neurons per Each Layer, and multiple Layers as well.

In Addition, Let's Denote Experiment (CNN output y) as $\tilde{y}$ ("tilde"),
Let ground Truth denoted as y.

Error for one experiment then is defined as

$$\tilde{y} - y \quad \text{or} \quad y - \tilde{y} \quad \dots (4)$$

Now. Loss function:

4. Loss function.

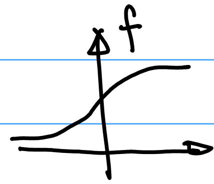
$$L = \frac{1}{2} \sum_{M=1}^{M} \sum_{i=1}^{N} (\tilde{y}_i^M - y_i^M)^2 \quad \dots (5)$$

$\uparrow$ No. of Experiments; $\nwarrow$ No. of Output

5. Now, Consider Eqn (1) & (2)

Eqn (1): has Activation function f( );
has Transfer function h( );

$$y = f\left(\sum_{i=1}^{N} w_i x_i + b\right) = f(h(w_i; x_i; b)) \quad \dots (1)$$



$$\sum_{i=1}^{N} w_i x_i + b \stackrel{\Delta}{=} h(w_i; x_i; b) = h(\cdot)$$

Based on the Notation for the Experiments and Ground Truth, Rewrite Eqn (1) as

$$\tilde{y} = f\left(\sum_{i=1}^{N} w_i x_i + b\right) = f(h(w_i; x_i; b))$$

OR

$$\tilde{y} = f(h)$$

5. The weights update equation

$$W_{ij}(t+1) = W_{ij}(t) + \delta W_{ij} \quad \dots (6)$$

$$\frac{\partial L}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \frac{1}{2} \left[\sum_{M=1}^{M} \sum_{i=1}^{N} (\tilde{y}_i^M - y_i^M)^2 \right]$$

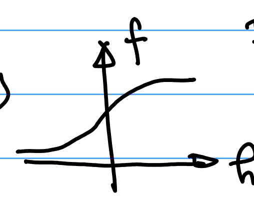
From Here, Apply "Chain-Rule" for multiple Layers. i, j, k, l.

April 19.

Topics: 1. Bounding Boxes Selection.
2. K-mean Cluster Algorithm.

2022S-114a-BoundingBoxSelection-Yolo4-2022-4-19.pdf

2022S-114b-boundingBox-Selection-yolo-tutorial-2022-...



Example: NMS Algorithm to select Bounding Boxes

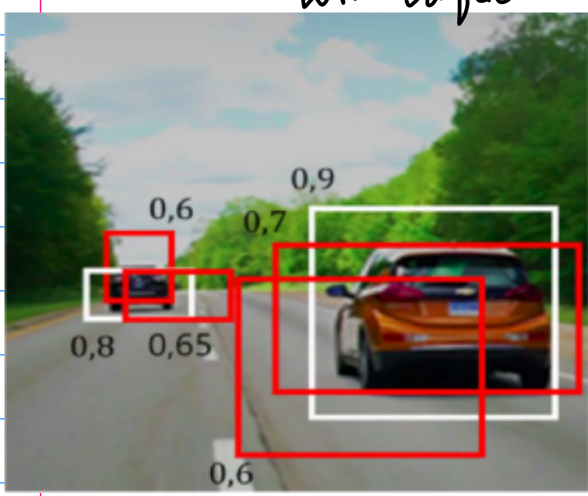
Note: the Notation for Our discussion

Image Partition into
$S \times S$ Grids $G_i(x, y)$.
$B_i(x, y)$, $i = 1, 2, \cdots, K$

Bounding Boxes.

B_i, $\underline{x}, \underline{y}$, w, h, C_i ← Confidence
Center point of B_i — Shape of B_i

(B_i, C_i) Pair Bounding Box with Confidence



Note: Goal is to Select/Decide one Bounding Box for Each Object with the highest Confidence Level.

Calculation of Bounding Box Selection.
See Handout Example.
IOU Computation (see PP40)

Note this Process selected the Bounding Box with the highest Confidence Level By Elimination process.

Now, consider the K-mean algorithm
A Technique that allows us to define Probability of Classes/Objects.

Ref:

2022S-114c-Kmean-handCalculation1.jpg
2022S-114c-Kmean-handCalculation2.jpg
2022S-114c-KmeanCluster-v3-2022-4-19.pdf

Example:
Feature Vectors
$$\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n \quad \cdots (1)$$
where
$$\vec{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \quad \cdots (1\text{-}b)$$

K-mean Clustering Algorithm is to Partition these feature vectors into K Classes.
$$S_1, S_2, \cdots, S_K \quad \cdots (2)$$
in such a way to minize the Within-Class Variation, e.g.
$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2 \quad \cdots (3)$$

$$\|\vec{x} - \vec{m}_i\|^2 \quad \dots (4)$$

m_i Cluster for S_i

Class i.

$$\vec{m}_i = \begin{pmatrix} m_{i1} \\ m_{i2} \end{pmatrix}, \quad \vec{x} - \vec{m}_i = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} m_{i1} \\ m_{i2} \end{pmatrix}$$

$$= \begin{pmatrix} x_1 - m_{i1} \\ x_2 - m_{i2} \end{pmatrix}, \quad \dots (5)$$

$$\|\vec{x} - \vec{m}_i\| = \sqrt{(x_1 - m_{i1})^2 + (x_2 - m_{i2})^2} \quad \dots (6)$$

minimization of the difference, e.g. Variation from its Class' Cluster $\vec{m}_i$

Arg min

Minimization for all feature Vectors in the Class S_i

$$\sum_{\vec{x} \in S_i} \quad \dots (7)$$

The minimization will have to Be carried out for all the Classes, therefore

$$\sum_{i=1}^{K} \quad \dots (8)$$

K Classes

The Algorithm:

"for Any" j

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \le \|x_p - m_j^{(t)}\|^2 \; \forall j, 1 \le j \le k\}, \quad \dots (a)$$

mean given set $x_1, x_2, \dots, x_N$

$$m = \frac{1}{N} \sum_{i=1}^{N} x_i \text{ for scalar} \quad \dots (10)$$

for vector, we have feature vectors.

$$\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$$

$$\vec{m}_i = \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} \quad \dots (11)$$

Class i

Note: Eqn(a) defines the Algorithm which make selection of cluster $\vec{m}_i$ (for Class i) satisfy the condition, such that each of Every $\vec{x}_p$ in the Class will have

$$\|\vec{x}_p - \vec{m}_i\| \le \|\vec{x}_p - \vec{m}_j\| \quad \dots (12)$$

We conduct the Computation Based on Eqn(a).

First, Select Number of Class K Based on Heuristics

Then, the initial process,

Step 2 Pick mean value for each class,

$$\vec{m}_i = \frac{1}{N} \sum_{j=1}^{N} \vec{x}_{ij} \quad \cdots (13)$$

Note: the class i is arbitrarily assigned. $(i = 1, 2, \cdots, K)$.

Step 3. Use Equation Below to check if the Equation satisfied, if yes, then keep the vector in the class i,

$$\| \vec{x}_p - \vec{m}_i \| \leq \| \vec{x}_p - \vec{m}_j \| \quad \cdots (14)$$

O/W, reassign the vector to its smaller distance group.

Step 4. Update the cluster By Re-evaluating mean.

$$\vec{m}_i = \frac{1}{N} \sum_{j=1}^{N} \vec{x}_{ij}$$

for all i classes.

Step 5. Check and verify the termination condition. by Comparing the current means to their previous mean value.

e.g.

$$\vec{m}_i^{(t)} = \vec{m}_i^{(t-1)} \quad \cdots (15)$$

Step t and step $t-1$ for mean i, $i = 1, 2, \cdots, K$

If Eqn (15) holds good, then, the process is done,

Clusters are given as $\vec{m}_i$, for $i = 1, 2, \cdots, K$.

Otherwise, Continue the Process as in Step 3. Continue till the termination condition satisfied.

Homework (Due A week from Today, April 26).

1. Write a python code to implement K-mean cluster Algorithm. (No OpenCV necessary)
2. Test Data Pattern is given in the previous gender prediction Python Code; (Note the height,

Weight of Each person from 2 Classes are given)

3. Submission of the Source code

4. Screen Capture of the Success execution of the program. (Submission to CANVAS)

5. Please make one zip file for your submission.

April 26

Note: Final Exam. May 18 (Wed).

Team Presentation: May 10.

a. 7~8 min. for Presentation & Demo.

b. 5~6 Slides

c. Demo, Show & Tell

d. $1\sim 1\frac{1}{2}$ min Q&A

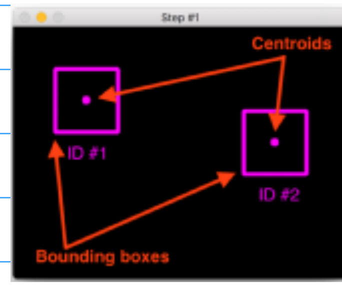
Topics: 1. Tracking Algorithm.

2. Semantic Segmentation

2022S-116a-semantic-segmentation-rcnn-2022-4-26.pdf

2022S-115-object-tracker-hand-calculation-HL-2020-8-3…

Objective: To keep Track object Movement/Position from time t_1 to t_2, to establish one-to-one relation Based on the shortest distance principle.



1. Denote Objects as $\vec{P_i}(x_i, y_i)$, and $\vec{P_{i+1}}(x_{i+1}, y_{i+1})$

$\vec{P_i} \to \vec{P_i}(x_i, y_i), (x_i, y_i)$

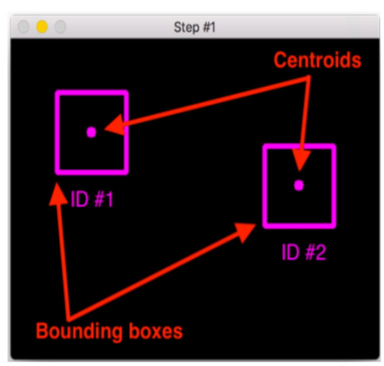
2. distance

$$d(\vec{P_i}(x_i, y_i), \vec{P_{i+1}}(x_{i+1}, y_{i+1}))$$

$$= d(\vec{P_i}, \vec{P_{i+1}}) = \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \quad \cdots (1)$$

Step 1. $\bar{x}, \bar{y}$ Computation.

Step 1. Compute Centroid



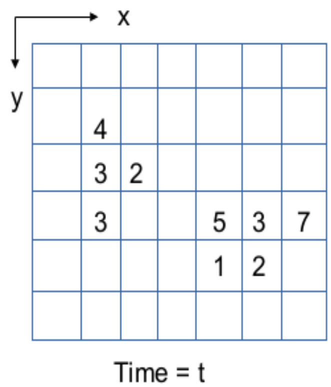
The algorithm:

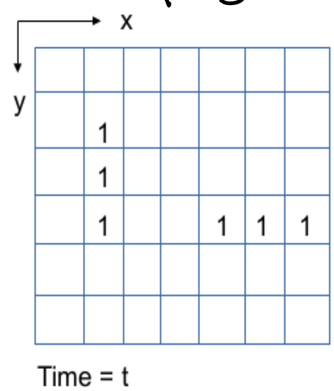
1. Compute centroid, is given in my lecture image processing an computation

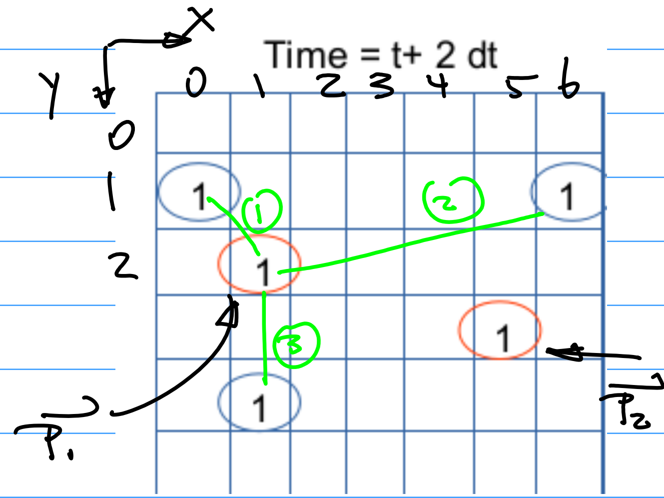
2. Assign ID to each

Centroid computation re 2019S-24-2018S-114-C Inference-final-2018-4-

Gray Scale Image → Binarization
T = 3





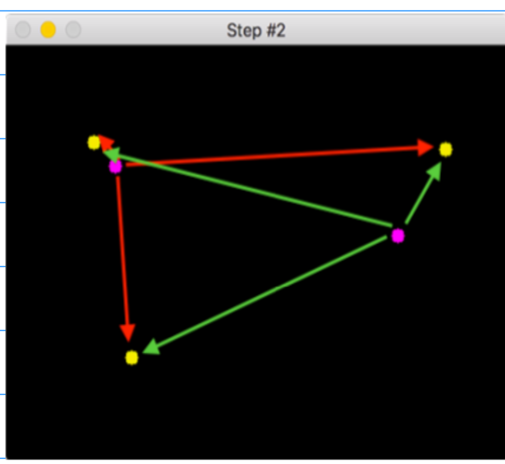


Compute $\bar{x}, \bar{y}$ for Each object And Create A Registration Table & place the Object ID, and its $\bar{x}, \bar{y}$ in the table

Obj. No.	ID	x-bar, y-bar
Object 1.	1	1, 2
Object 2.	2	5, 3

Next, Compute All Distances

Note: Pink Dots (t_1), Yellow Dots (t_2) the Distance should be Computed for Time t_1 & t_2.



$$dist_1 = \sqrt{(x_1-0)^2+(y_1-1)^2}$$
$$= \sqrt{1^2+1^2} = \sqrt{2} = 1.414$$

✓ D(o1, o1_new) = sqrt(2); D(o1, o2_new) = sqrt(26); D(o1, o3_new) = sqrt(4); D(o2, o1_new) = sqrt(29); D(o2, o2_new) = sqrt(5); D(o1, o3_new) = sqrt(17);

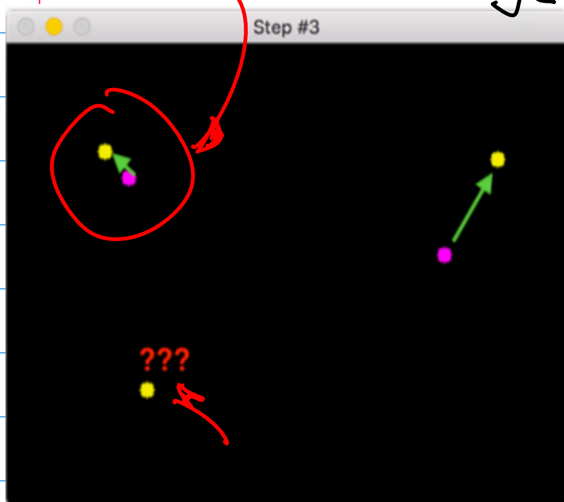
$$dist_2 = \sqrt{(1-6)^2+(2-1)^2}$$
$$= \sqrt{26}$$

$$dist_3 = \sqrt{(1-1)^2+(2-4)^2}$$
$$= 2$$

$$\min\{dist_1, dist_2, dist_3\} = dist_1$$

Shortest Distance, therefore Object $\vec{P_1}(x_i, y_i; t)$ Now is at Location $\vec{P_1}(x'_i, y'_i; t+1)$,
$(x'_i, y'_i) = (0, 1)$

Verified, Shortest Distance $=\sqrt{2}$



New Point Appeared

Note 1: at time t_2, Generally Speaking this new point does not provide shortest distance (!) So place this new point in the Registration Table.

Note 2: Once the matching is established, then Update the Registration Table.

Homework (1st, Optional)

Implement the tracking Algorithm in Python or C++ to Verify the Example given in the class (PPT).

Due A week from today May 3rd.

Example: Semantic Segmentation. 4 Steps

Fully Connected

1. Replace FC layers with convolutional layers.
2. Convert the last layer output to the original resolution.
3. Do softmax-cross entropy between the pixelwise predictions and segmentaion ground truth.
4. Backprop and SGD

a.

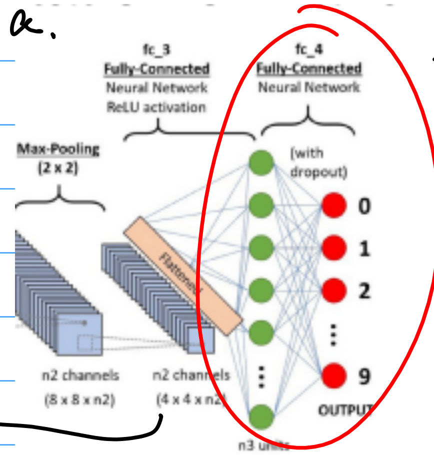


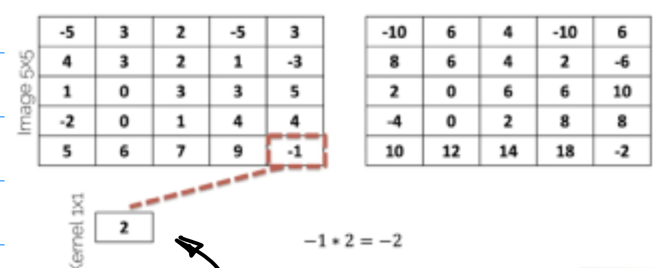
Fig. 1

First, Convolutionalization of F.C. Layer.

1x1 Convolution

Same Process as KxK Convolution.

1x1 Convolution



-5	3	2	-5	3
4	3	2	1	-3
1	0	3	3	5
-2	0	1	4	4
5	6	7	9	-1

-10	6	4	-10	6
8	6	4	2	-6
2	0	6	6	10
-4	0	2	8	8
10	12	14	18	-2

Kernel 1x1 Scaling Operation.

Fig. 2

Note: the Number of Kernels for Convolution And the matching Output Layers.

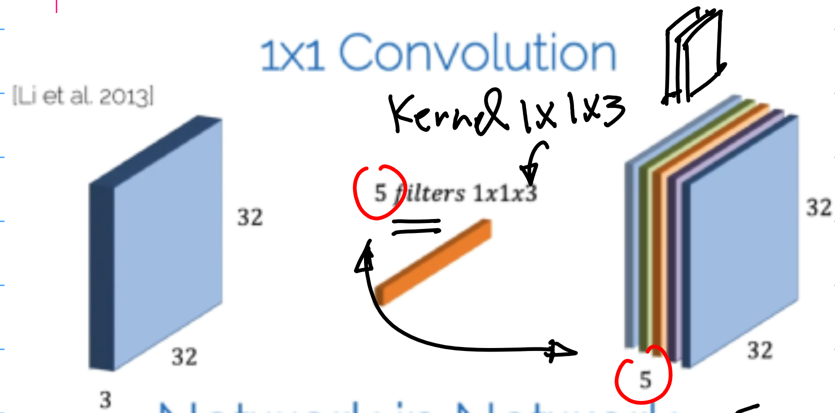


Fig 3.

Interpolation.
Un-pooling.

New layer

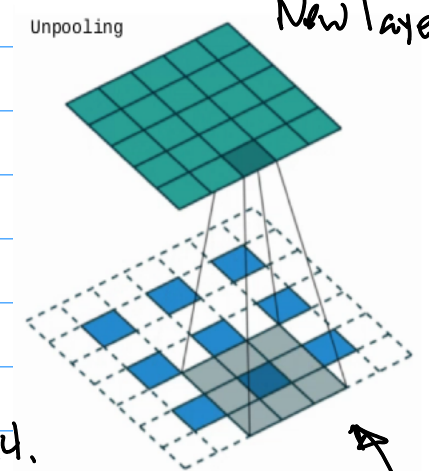


Fig 4.

Convolve Feature map 32x32x3 once by 1x1x3. → Output 32x32x1

Pad "0"s Around each element (pixel)

Convolution

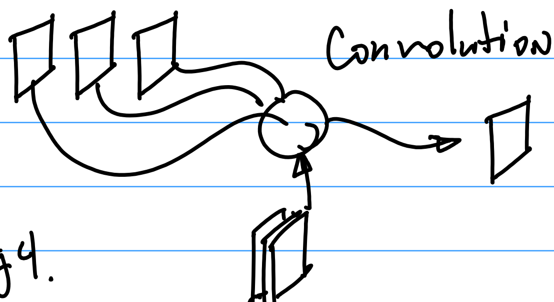


Fig 4.

1x1x3 Kernel

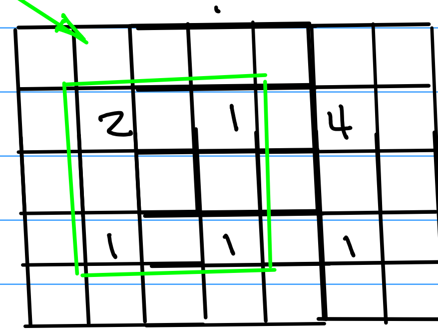
Observation: 1x1xK Convolution allows us to convert a Single Neuron to 2D Convolutional Layer (point).

1x1xK Convolution allows us to implement filtering operation as illustrated Above.

Suppose

1	-2	3
2	1	4
1	1	1

$\frac{1}{9}(2+1+1+1)$



Consider Up Sampling (Super-Sampling) Techniques.

Then, choose un-pooling Operation, such as Average if more than 1 Non-Zero elements under the position of 3x3 Kernel.

May 3rd, 22

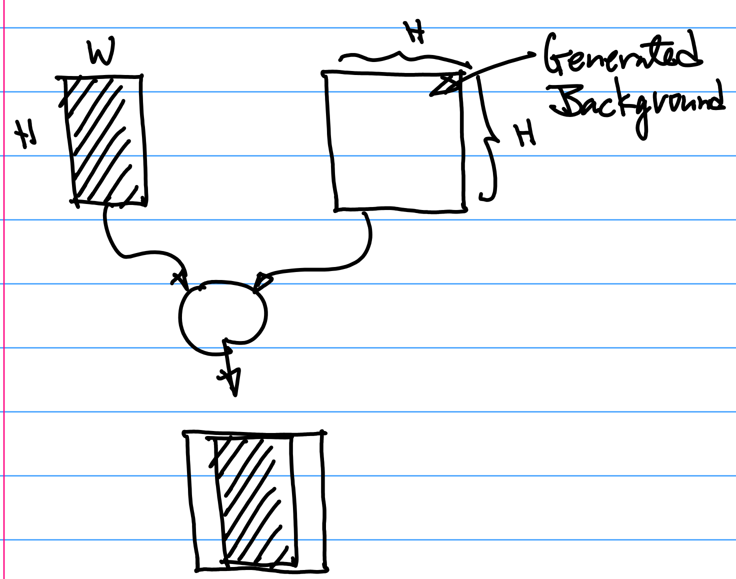
1. Final Schedule

Group II Classes

Group II classes are those classes which meet TR, T, R, TWR, MTR, TRF, MTRF, MTWR, TWRF, RF, TF.

Regular Class Start Times	Final Examination Days	Final Examination Times
7:00 through 8:25 AM	Thursday, May 19	7:15-9:30 AM
8:30 through 9:25 AM	Monday, May 23	7:15-9:30 AM
9:30 through 10:25 AM	Wednesday, May 18	9:45 AM-12:00 PM
10:30 through 11:25 AM	Friday, May 20	9:45 AM-12:00 PM
11:30 AM through 12:25 PM	Tuesday, May 24	9:45 AM-12:00 PM
12:30 through 1:25 PM	Thursday, May 19	12:15-2:30 PM
1:30 through 2:25 PM	Monday, May 23	12:15-2:30 PM
2:30 through 3:25 PM	Wednesday, May 18	2:45-5:00 PM
3:30 through 4:25 PM*	Friday, May 20	2:45-5:00 PM
4:30* through 5:25 PM*	Tuesday, May 24	2:45-5:00 PM

2. Aspect Ratio for Preprocessing, please check your source code and Preprocessing Algorithm.



3. Team Project Requirements including Responsibility & Amount Implementation.

Ref

2022S-116b-#190I-1-README-Setup-Mask-RCNN-v1-Y...

2022S-116c-entropy.pdf

Example: From pp.50. 1x1xK Convolution.

1. Calculation. Given A Kernel = 2 (1x1) And feature map (below), find the Convolution Result.

Image 5x5

-5	3	2	-5	3
4	3	2	1	-3
1	0	3	3	5
-2	0	1	4	4
5	6	7	9	-1

-10	6	4	-10	6
8	6	4	2	-6
2	0	6	6	10
-4	0	2	8	8
10	12	14	18	-2

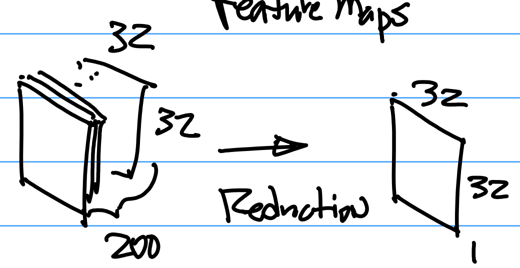
Kernel 1x1

2

$-1 * 2 = -2$

2. Suppose the given condition changed to 32x32x200

Feature maps



Note: Inspired by Human Visual Perception System, Retina (Early Vision)

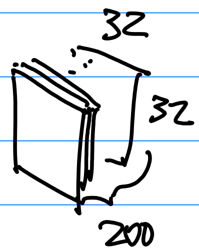
Spatial Scales. → Convolutional Kernels Designed to Capture Different Scales, Orientation

Examples such as Canny Edge Detection.

Contours Analysis.

a. Design a convolutional Kernel to Realize the Above mentioned Task.

b. Carry out the hand Calculation.

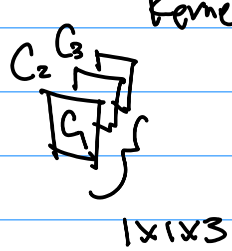


Use 3x3x3 Feature Map Below to Demonstrate the process

$$F_1(x,y)=\begin{bmatrix}1&1&3\\2&0&-5\\1&1&0\end{bmatrix}\quad F_2(x,y)=\begin{bmatrix}2&0&-1\\1&0&2\\0&0&1\end{bmatrix}\quad F_3(x,y)=\begin{bmatrix}2&3&2\\1&0&2\\6&0&1\end{bmatrix}$$

Design a Kernel to Produce 1 Feature By Combing All the given feature maps.

Sol Pick 1x1x3 Convolutional Kernel



(1) $C_1 = C_2 = C_3$ Treat them Equal

(2) Scaling factor to keep the Output feature map in the Same Range.

$$C_1 = C_2 = C_3 = 1/3$$

For Step 1.

$$\begin{bmatrix}1&1&3\\2&0&-5\\1&1&0\end{bmatrix}\quad \begin{bmatrix}2&0&-1\\1&0&2\\0&0&1\end{bmatrix}\quad \begin{bmatrix}2&3&2\\1&0&2\\6&0&1\end{bmatrix}$$

$$1*C_1 + 2*C_2 + 2*C_3 = \frac{1}{3}(1+2+2)$$
$$= 5/3$$

Step 2.

$$\begin{bmatrix}1&1&3\\2&0&-5\\1&1&0\end{bmatrix}\quad \begin{bmatrix}2&0&-1\\1&0&2\\0&0&1\end{bmatrix}\quad \begin{bmatrix}2&3&2\\1&0&2\\6&0&1\end{bmatrix}$$

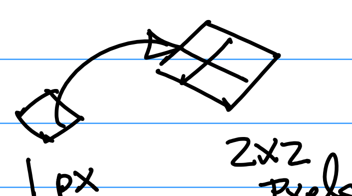
$$1*C_1 + 0*C_2 + 3*C_3$$
$$= \frac{1}{3}(1+0+3) = 4/3$$

And for the rest of the Steps, just to Continue the convolution.

Note: By Adjusting the Kernel Coefficient C_x we can emphasize Certain Layers.

Now, Consider the Design Technique to Increase the Resolution of Feature Maps.

The Simplest Way:



By Copying 1 pixel to 2x2 pixels.

May 3rd, 22

Example: Loss function for Mask RCNN.

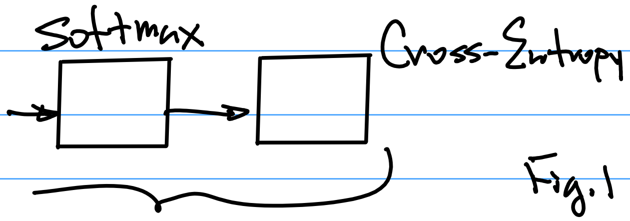


Fig. 1

To Define Loss function

Note: 1. Define Entropy.

2022S-116c-entropy.pdf

a. Define a set of Experiments

$S=\{s_1, s_2, \ldots, s_n\}$

$\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n\}$

Prediction, Output from Experiments

...e associated probabilities of

$\{p(s_1), p(s_2), \ldots, p(s_n)\}$

...nformation content is defined ...unit "bit,"

$$I(s_i) = \log_2 \frac{1}{p(s_i)} \quad \cdots (1)$$

b. $Prob(S_i)$ or. $Prob(y_i)$, $Prob(\hat{y}_i)$

Ground Truth   Predicted

c. Information Content Shannon Theory.

$\hat{y} - y \quad \cdots (1)$

Predicted Output By CNN   Ground Truth

S_i or $y_i, \hat{y}_i \to Prob(S_i), Prob(y_i), Prob(\hat{y}_i)$

Event

$$\log_2 \frac{1}{Prob(S_i)}, \quad \leftarrow \quad \frac{1}{Prob(S_i)}, \frac{1}{Prob(y_i)}, \frac{1}{Prob(\hat{y}_i)} \quad \cdots (2a)$$

$\log_2 \frac{1}{Prob(y_i)}, \log_2 \frac{1}{Prob(\hat{y}_i)}$ unit "bit" ($\log$ Base $b=2$) $\cdots (2b)$

$$H(S) = \sum_{i=1}^{n} p(s_i) I(s_i) = -\sum_{i=1}^{n} p(s_i) \log_2 p(s_i) \text{ (bit/symbol)}$$

$\log_2 \frac{1}{Prob(S_i)} = -\log_2 Prob(S_i)$

Entropy:

$$\boxed{H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)} \quad \cdots (3)$$

Example: Suppose 2 experiments from CNN Output.

$\hat{y}_1$ Stop Sign

$Prob(\hat{y}_1) = 3/4$

$\hat{y}_2$ Right-Turn Sign

$Prob(\hat{y}_2) = 1/4$

Find Entropy:

Sol: From Eqn (3),

$$H(\hat{y}) = \sum_{i=1}^{2} Prob(\hat{y}_i) \log_2 \frac{1}{Prob(\hat{y}_i)}$$

$$= Prob(\hat{y}_1) \log_2 \frac{1}{Prob(\hat{y}_1)} + Prob(\hat{y}_2) \log_2 \frac{1}{Prob(\hat{y}_2)}$$

$$= \frac{3}{4}\left(-\log_2 \frac{3}{4}\right) + \frac{1}{4}\left(-\log_2 \frac{1}{4}\right)$$

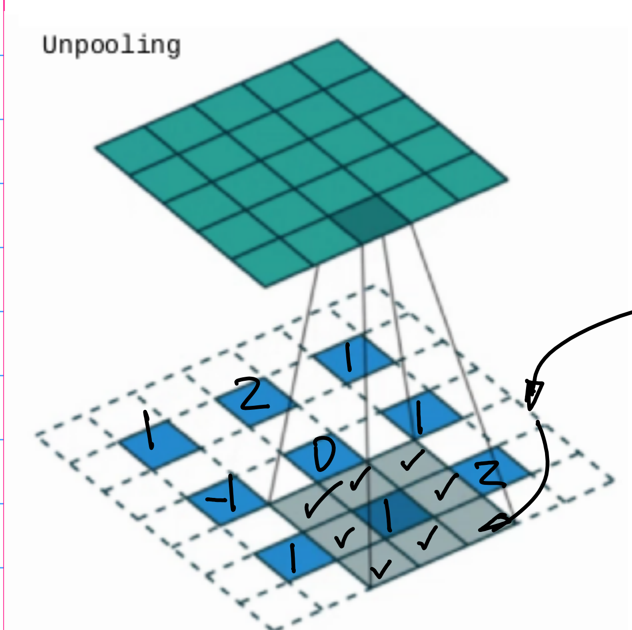
$$= \frac{3}{4}(-\log_2 3 + 2) + \frac{1}{4}(-0 + 2)$$

$$= \frac{3}{4}(2 - 1.6) + \frac{1}{2} = \frac{3 \times 0.4}{4} + 0.5$$

$$\simeq 1.2/4 + 0.5 = 0.8$$

A Better Way for Un-Polling, e.g. up-Sampling.

Example: Given a 3x3 feature Map, up-Sampling to generate 5x5 feature Map.



Sol

Step 1. Pad "0"s Around Each Non-Zero element.

8-Connected Neighbours.

Step 2. Design Your Kernel $K \times K$ ($K=3$), use the following Kernel

-1	0	1
-2	0	2
-1	0	1

Orientation Selective Edge Detector (Vertical Features Being Selected / Emphasized)

Step 3. Conduct the convolution. (Take Kernel Position As illustrated here)

$$
\begin{aligned}
& 0*(-1)+0*(0)+0*(1) \\
+\ & 0*(-2)+1*(0)+0*(2) \\
+\ & 0*(-1)+0*(0)+0*(1) \\
=\ & 0
\end{aligned}
$$

Continuous this process till all the feature points on the feature map are taken Care of, Hence 5x5 feature map is generated.

Cross Entropy.

Cross Entropy: Measure of difference between 2 probabilities

$$\boxed{H(p,q) = -\sum_{x\in\mathcal{X}} p(x)\,\log q(x)} \quad \ldots(4)$$

Prob(y_i) Ground Truth

Prob($\hat{y}_i$) Predict (Output from Softmax for Example).

Measure the difference Between the probability of the ground Truth Prob(y_i) and the probability of the Prediction Prob($\hat{y}_i$) → then Build Loss

↓

As an Objective Function → Gradient Descent

GD technique to train CNN.

Example: Suppose for traffic Sign Classification.

Prob(y_1) = 1/4, Prob(y_2) = 3/4 (Ground Truth)

Prob($\hat{y}_1$) = 3/4, Prob($\hat{y}_2$) = 1/4 (CNN output)

Find Cross-Entropy.

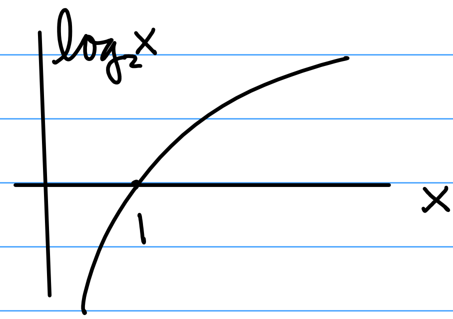
$$\sum_{i=1}^{2} \text{Prob}(y_i)\log_2 \frac{1}{\text{Prob}(\hat{y}_i)}$$

$$= -\text{Prob}(y_1)\log_2 \text{Prob}(\hat{y}_1) - \text{Prob}(y_2)\log_2 \text{Prob}(\hat{y}_2)$$

$$= -\frac{1}{4}\log_2\frac{3}{4} - \frac{3}{4}\log_2\frac{1}{4}$$

$$= -\frac{1}{4}(\log_2 3 - 2) - \frac{3}{4}(-2)$$

$$= \frac{1}{4}\times 0.4 + 3/2 = 0.1 + 1.5 = 1.6 \text{ bit}$$



Example: Now, Loss function can be defined

Define loss function (Cross entropy)

$$\boxed{D(\hat{Y}, Y) = -\frac{1}{N}\cdot\sum Y_i \cdot \log(\hat{Y}_i)} \quad \cancel{(3)} \ldots(5)$$

$$\boxed{Q(w) = \sum_{i=1}^{n} Q_i(w)} \quad \cancel{(4)} \text{ Objective function} \ldots(6)$$

Difference Between the probability of the ground Truth Prob(y_i) and the probability of the Prediction Prob($\hat{y}_i$)

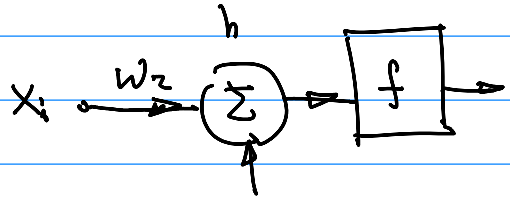
a. $(\hat{y}_i - y_i)^2 = (y_i - \hat{y}_i)^2$

Compare the experiment i with the ground truth i

$$Q(w) = \sum_{i=1}^{n} Q_i(w) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (w_1 + w_2 x_i - y_i)^2. \quad \cdots (7)$$

b. All Experiments v.s. All ground Truth

c. Simple Example



Activation Function.

Now, Let's gradient Descent Technique

Gradient

$$\boxed{w := w - \eta \nabla Q(w) = w - \frac{\eta}{n} \sum_{i=1}^{n} \nabla Q_i(w)} \quad \cdots (8)$$

Weight at $k+1$ | Weight at k | η: controls the update Step Size

d.

Now, from Eqn (8), find solution (Calculation) of Eqn (7).

at k+1   at k $\frac{\partial}{\partial w_1}$ Gradient Descent $2\cdot(w_1+w_2x_i-y_i)$

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} := \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \eta \begin{bmatrix} \frac{\partial}{\partial w_1}(w_1 + w_2 x_i - y_i)^2 \\ \frac{\partial}{\partial w_2}\underbrace{(w_1 + w_2 x_i - y_i)^2}_{Q\text{ : Objective function.}} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \eta \begin{bmatrix} 2(w_1 + w_2 x_i - y_i) \\ 2x_i(w_1 + w_2 x_i - y_i) \end{bmatrix}. \quad \cdots (a)$$

$\frac{\partial}{\partial w_2}$

$$o \xrightarrow{w_1} h_1 \to h_2 \to h_3 \to \text{Output}$$

$$\frac{\partial Q}{\partial w_1} = \frac{\partial Q}{\partial h_3} \cdot \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$ Back Prop.