

CMPE258

Fall 2023

✓

August 22 (Tues)

Organizational meeting.

1. Class material on github
github/huawili

alv100	Add files via upload
deep-learning-2020S	Add files via upload
deep-learning-2022s	Add files via upload
deep-learning-2023s	Add files via upload
facial-detect	Add files via upload
lec1Capture/CMakeFiles	Delete CMakeDirectoryInformation
lecOpenCV_GL	openGL and openCV sample commi
riscv	Create readme.txt
20-2021S-0-7-1convnets-NumericalD...	Add files via upload

Course and Contact Information

Instructor(s): Harry Li

Office Location: Engineering Building, Room 267A

Telephone: (650) 400-1116 for text messaging only

Email: hua.li@sjsu.edu

Office Hours: M.W. 3:00-4:00 pm

In-Person.

Class Days/Time: Tuesdays and Thursdays 4:30-5:45 pm.

Classroom: Engineering Building Room 337

Prerequisites: CMPE 255 or CMPE 257 or instructor consent. Computer Engineering majors only.

Course Description

2. Prerequisites Requirements

Bring your Proof to the next Class.

3. Emphasis on "Deep Neural Networks",
Semantic Segmentation

Course Description



Deep neural networks and their applications to various problems, e.g., speech recognition, image segmentation, detection and recognition of temporal and spatial patterns, and natural language processing. Covers underlying theory, the range of applications to which it has been applied, and learning from very large data sets.

Note: Definition (n.b.): (†) Human Intelligence
is Symbolic Representation of
Learn + Experience.

4. Projects. 2

Plks | team project
(Semester Long) } ~~30%~~ ~~25%~~
30 pts

5. In-Person Class; CANVAS is

utilized to Post Homework/Project Requirements, and to Collect the Submission.
of the Homework,
as well as for the Exams.

No

This course is an online course. The students must have Internet connectivity and Zoo his/her machine. The students must participate in the class activities and submit all assignments to SJSU CANVAS. The syllabus, faculty contact information on the syllabus, projects, and exam papers are all available on CANVAS. See [University Policy F13-2](#)

6.

Grading Information

Quiz, Homework, Projects	30%
Midterm Examination	30%
Final Examination	40%

7. Textbooks & References

Textbook

- Deep Learning with Python, 1st or 2nd Edition, by François Chollet, ISBN-10: 9781617294433, <https://github.com/fchollet/keras/blob/master/2018F-6-DeepLearningCh02.pdf>
- Robot Vision by B.K. P. Horn, the MIT press, ISBN 0-262-08159-8, (Hill).
- Reference textbook Learning OpenCV, Computer Vision with the OpenCV Library, O'Reilly Publisher, ISBN 978-0-596-51613-0, 2011.

8. Software Tools & Dev. Environment

Python.
Anaconda.
TensorFlow.

Rcharm.

Note: Rm268 Available per ON
Approved Basis.

August 24 (Thursday)

Note:

- CANVAS To Be up by the end of the day, Friday;
- Tools & Software To be installed (will provide "Readme" as Ref)

OpenCV

Python.

T.F. Version 2.0 or higher

Today's Topic:
Intro. to Deep Convolutional Neural Networks.

Ref: github.

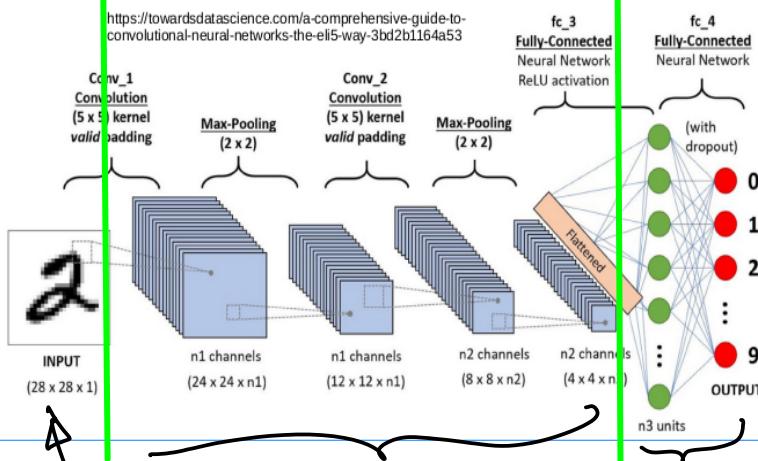
2022F-103-NN-Intro-Python-v5-2022-8-25

Note: Lab Space for the class
Rm268.

Example: Architecture Overview.

Note 1.

Illustration of A CNN for Digits Recognition



Preprocessing
Computer
Vision.

Convolution
Layers

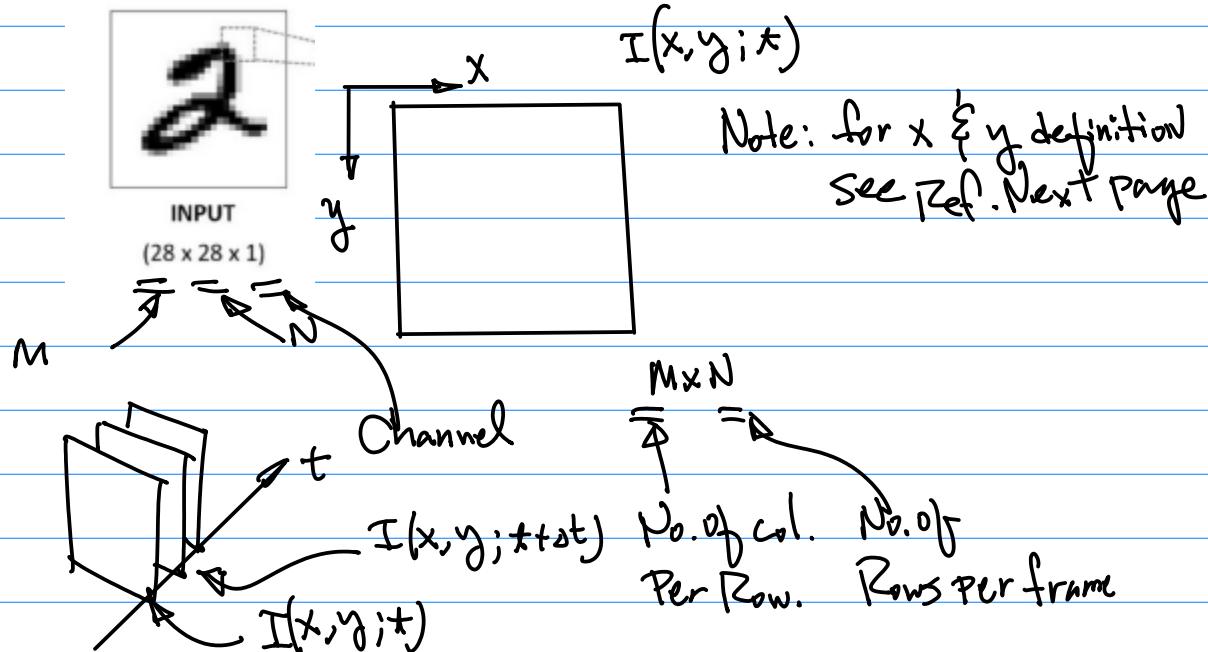
Feature Extraction
To produce Feature Vectors.

Feed Forward
Neural Networks.

Decision
Making

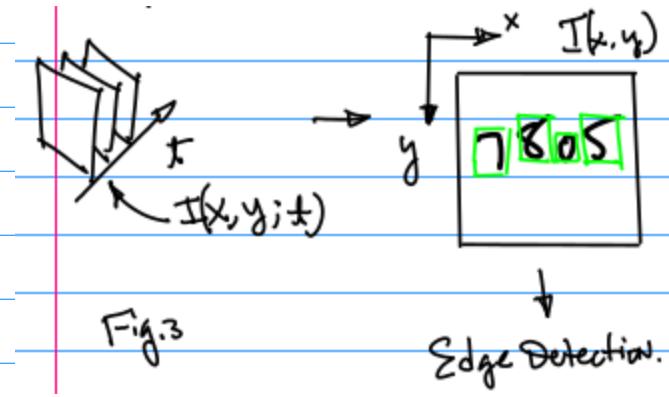
$$\vec{u} = (u_1, u_2, \dots, u_N)$$

2. Image Definition



Ref: on the github.

[2023S-101-Notes-cmpe258-2023-03-16.pdf](#)



Comparison to Web Cam.

Web Resolution: 1920x1080

$$\begin{aligned} & \downarrow \\ & \sim 2^{11} = 2 \cdot 2^9 \\ & = 2K \\ & \sim 2^{11} \\ \\ & 2^{22} = 2 \cdot 2^{20} \\ & = 4 \text{ Meg} \\ & \simeq 4 \text{ million.} \end{aligned}$$

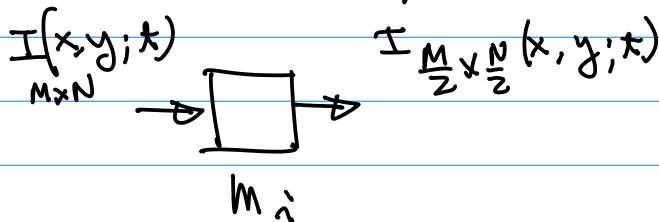
3. For the Convolutional Layer, we denote it as C_i , where $i=1, 2, \dots, N$

for Max Pooling Layer, we denote it as M_j , where $j=1, 2, \dots, K$

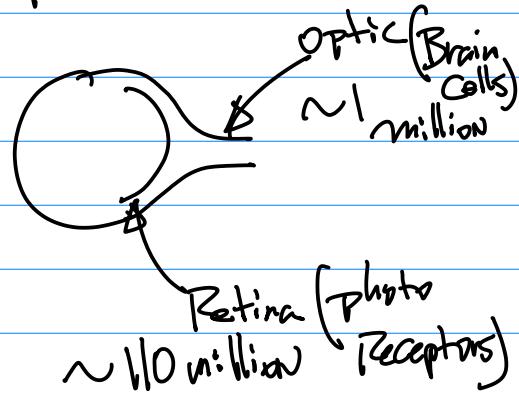
So, for the example, we have

$$C_1 M_1 \rightarrow C_2 M_2$$

Maxpooling for Resolution Reduction/Feature Reduction.



Note: Biologic Inspiration,
Human Visual Perception System,
Retina, $\sim 1/0$ million
photo Receptors



4. At the 3rd Segment (Blocks) of the Architecture, we denote Feed Forward Neural Networks as F_M

$$F_M$$

No. of Neurons / Nodes

for example, F_{10} (10 Nodes) for the output layer.

August 29 (Tue)

Note: 1^o CANVAS is up.2^o Homework Assignment

} a. Honesty Pledge to Be
Signed/Signed Copy
has to be uploaded
to CANVAS.
b.

Software Tools
Installation. (0 pt)

(1) OpenCV. By Friday

Next Tuesday. Bring Your
Laptop w/ OpenCV installed.

↓
Please use Smartphone
to take a photo, And upload
the photo to your laptop to
display.

Note: Sample code (Python)

was posted on the github.

(2) Anaconda Installed on
your Laptop.Note: Readme for Anaconda
installation was posted on
the github.

(3) Create ChatGPT Account.

Python Interface to ChatGPT
API (3.5 Version) is to
be utilized in your Team Project.3^o Form Team for the
Semester Long Project.Example: Consider 2D Convolution
Technique.Ref: [github/valihi/OpenCV](https://github.com/valihi/OpenCV)

2022

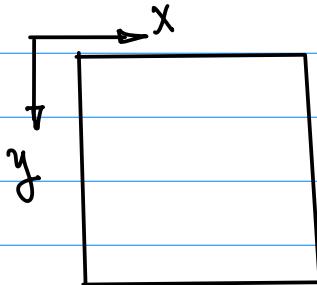
Note 1:

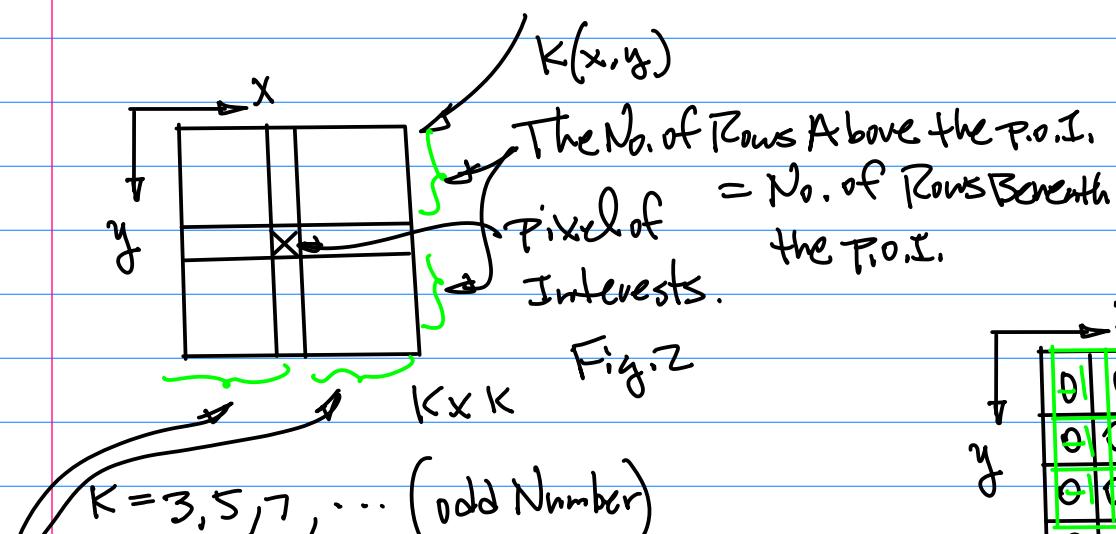
$$c(n_1, n_2) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} a(k_1, k_2) b(n_1 - k_1, n_2 - k_2)$$

Image Kernel

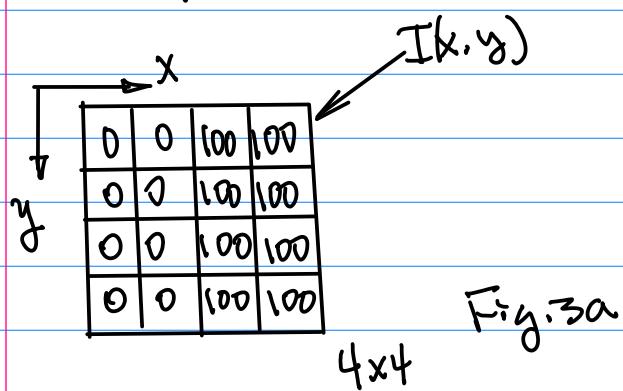
$n_1 = x, n_2 = y$

... (1)

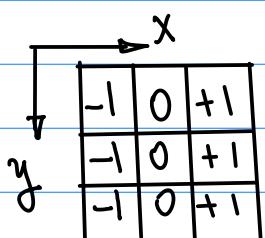
Summation
Index: k_1, k_2 are
for x and y .Note 2: $a(x, y)$ or $a(k_1, k_2)$ as
an Image. $I(x, y)$ $b(x, y)$: a Kernel for 2D
Convolution, $b(k_1, k_2)$ Note 3: Kernel $b(x, y)$ can be rewritten
 $K(x, y)$ Size of A kernel is denoted as
 $K \times K$



The No. of Col. Right to the P.O.I. = The No. of Col. Left to the P.O.I.

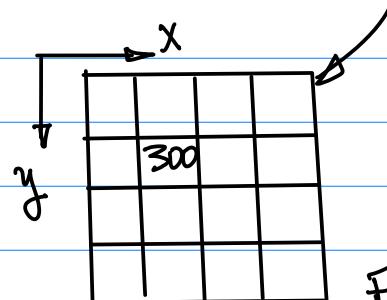


$$8 \text{ bit Gray scale: } 2^8 - 1 = 255$$

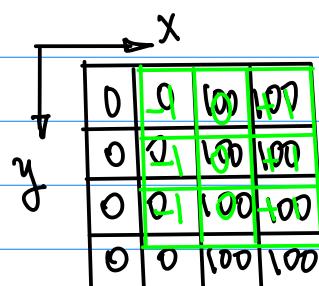


Step 1. Take the Kernel, and place it at the initial position.

$$\begin{aligned} -1 \times 0 + 0 \times 0 + 1 \times 100 &+ \\ -1 \times 0 + 0 \times 0 + 1 \times 100 &+ \\ -1 \times 0 + 0 \times 0 + 1 \times 100 &+ \\ \hline &\equiv 300 \end{aligned}$$



Step 2. Shift $K(x, y)$ to the Right by 1 pixel (on the same Row).



$$\begin{aligned} -1 \times 0 + 0 \times 100 + 1 \times 100 &+ \\ -1 \times 0 + 0 \times 100 + 1 \times 100 &+ \\ -1 \times 0 + 0 \times 100 + 1 \times 100 &= 300 \end{aligned}$$

$$I(x,y) * K(x,y)$$

August 31 (Th).

Note: 1° Honesty Pledge on CANVAS,
Signed pdf due this Friday.
(ON CANVAS). Homework ON CANVAS

2° Will post Anaconda Installation
& OpenCV Installation, display
an image.

(1) Screen Capture of the activated
CONDA Environment, with personal
identifier;

(2) Screen Capture of the OpenCV
Display with P.I.D.

Sample Code for the display to
Be Re-posted on github, 2023F

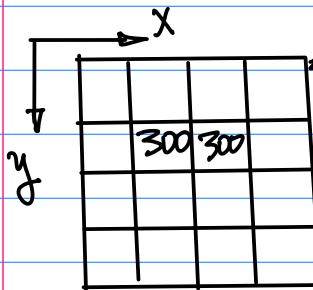


Fig. 3 F

Step 3.

$$I(x,y) * K(x,y)$$

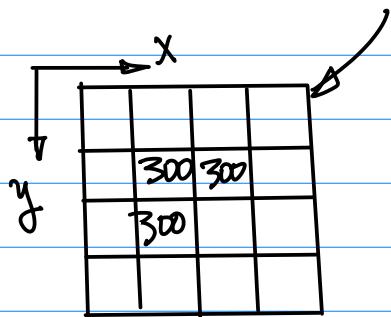


Fig. 3 G

Step 3.

$$I(x,y) * K(x,y)$$

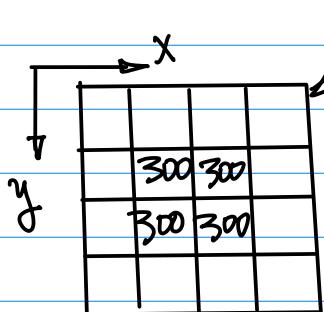


Fig. 3 H

Note: 2D Convolution (Continuous Case)

given Image $f(x,y)$
a Kernel $K(x,y)$

$$\iint f(u,v) K(x-u, y-v) du dv \dots [2]$$

Example: Theoretical Aspects
for 2D Convolution.

Ref: 1° PPT on the github.



1-2020S-#2019S-23-
2DConvolution-2019-2-4.
pdf

2° 2023S 1D-note-...

2023-03-21.pdf

$$\sum_{\Sigma} f(u,v) h(x-u, y-v) du dv \quad \dots (1)$$

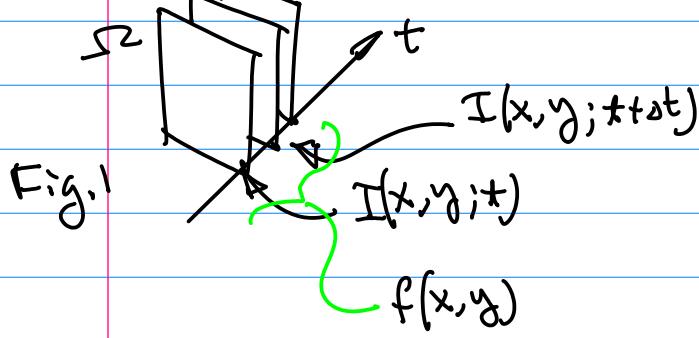
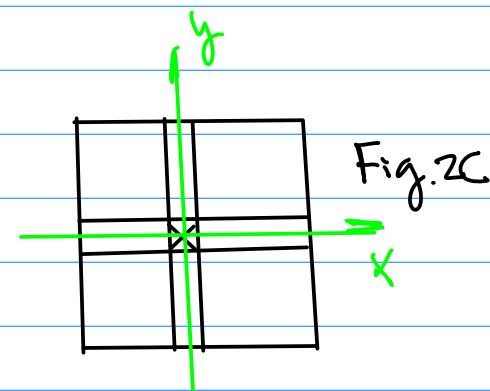
~~\sum~~ = 

Kernel

Note 1. Σ : Image Planes

→ (2) Flip the kernel w.r.t. X-axis.
(since $k(-x, y)$ changed to $k(-x, -y)$)

$I(x, y; t)$
 $f(x, y)$



Note: for 1D Case



Note 2: Kernel

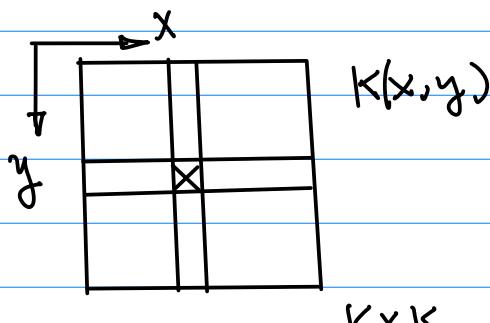
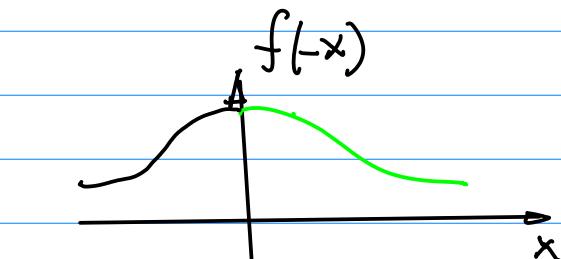
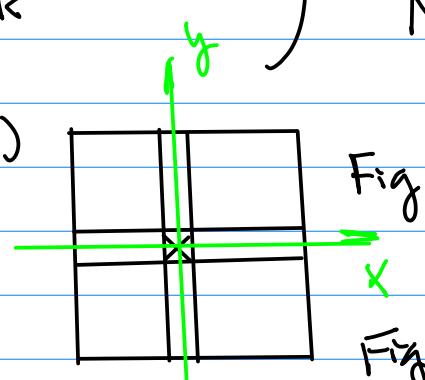


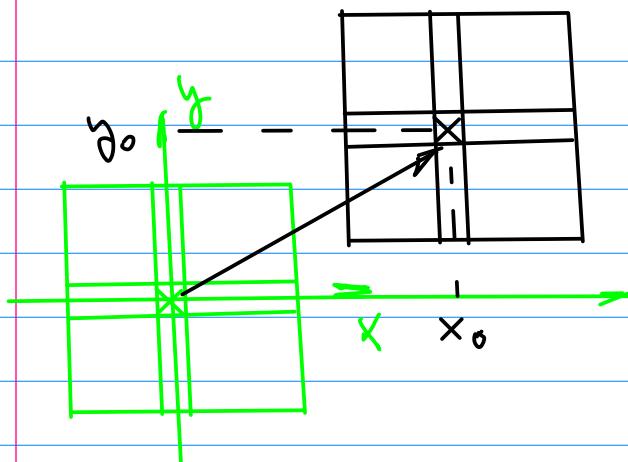
Fig. 2a
↓ (1)
 $K(-x, y)$

Flip the Kernel
w.r.t y-axis



Next, make $k(-x, -y)$ to
 $k(x-x, y-y)$





Note: $\frac{k-1}{2}$ for the top Rows
 $\frac{k-1}{2}$ for the Bottom Rows
 Lost
 $2 \times \frac{k-1}{2} = k-1$

Similarly for the Col.s.

Fig.2d

Now, for Discrete Image (OR
Digital
Feature plane)

Replace $\int \int$ by $\sum \sum$

$$c(n_1, n_2) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} a(k_1, k_2) b(n_1 - k_1, n_2 - k_2)$$

Image Kernel ... (z)

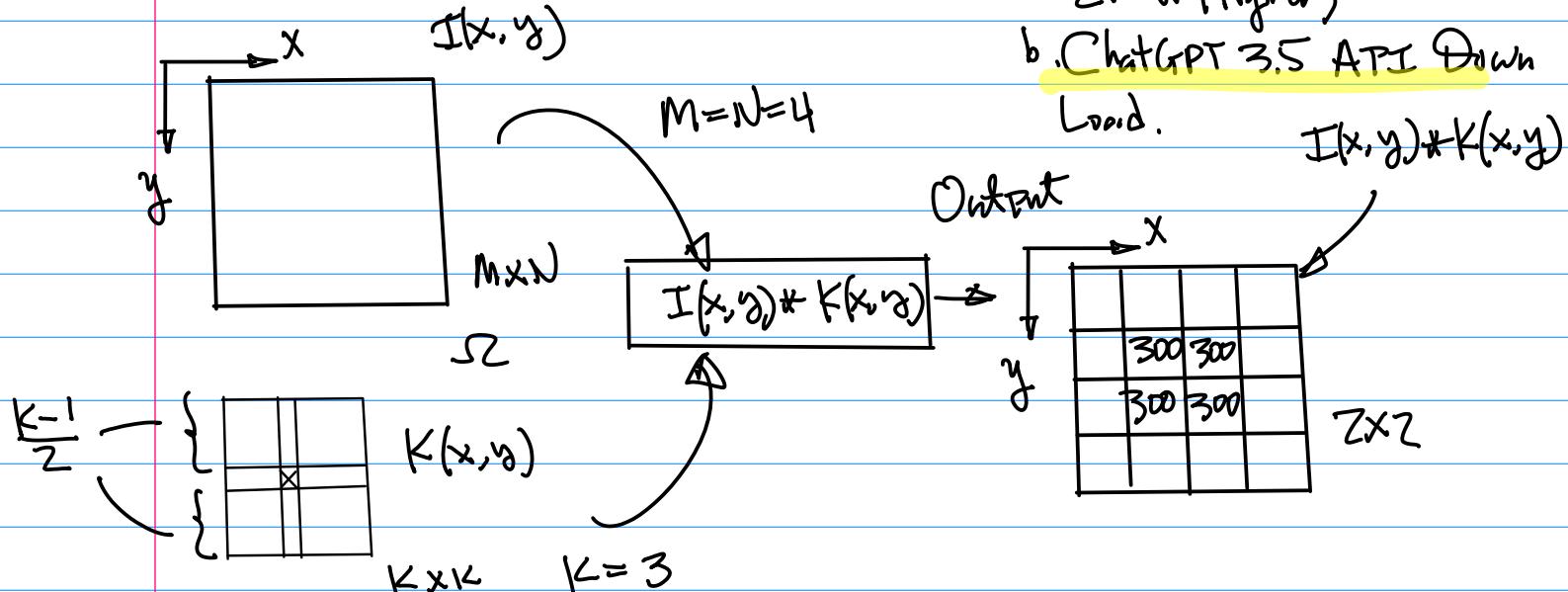
Dimension Change Due to the Convolution.

Therefore,
 Input Dimension $M \times N$ \rightarrow Output After Convolution
 $m - (k-1) \times N - (k-1)$
 ... (3)

Sept. 5 (Tue)

Note: 1^o CANVAS OpenCV, Anaconda Installation;
 2^o Homework Coming On Thursday.

- a. Installation of T.F. Version 2.0 or Higher;
- b. ChatGPT 3.5 API Down Load.



Example: Continuation on
Convolutional Layers and
Architecture for Handwritten
Digits Recognition.
Based on Eq(3), PP.9.

$M \times N$ Image $\rightarrow M - (K-1) \times N - (K-1)$ System.

$K \times K$
Convolution

From MNIST Architecture,
the Input Image $I(x, y)$ with
 $M \times N$ ($M = N = 28$) + the
Output feature layers Resolution
 24×24 . $\rightarrow K = (28-24) + 1$
 $\underbrace{}_{\text{Resolution difference.}}$

Now, the total No. of feature
layers = n

Since one Convolution (e.g.
using one $K \times K$ Kernel
to perform Convolution)
produces 1 feature layer,

Therefore
 n Convolutional feature
Layers is the
result of n Convolutions,
e.g. n different
Kernels.

Note: Multiple feature layers
are due to the inspiration
of Human Visual Perception
System, e.g. "Early" Vision

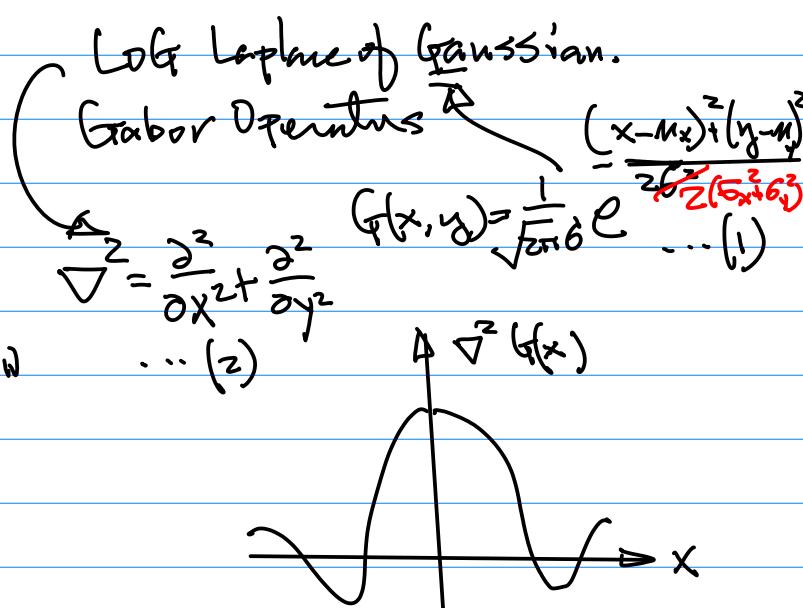


Fig.1.

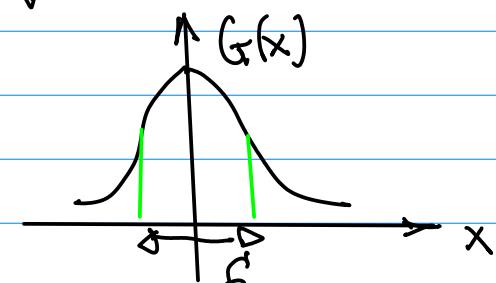


Fig.2

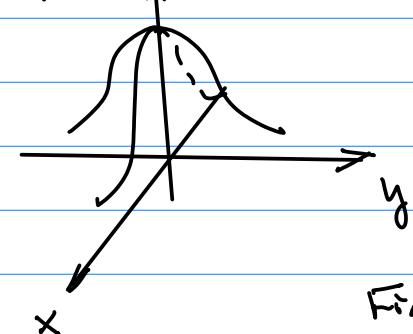
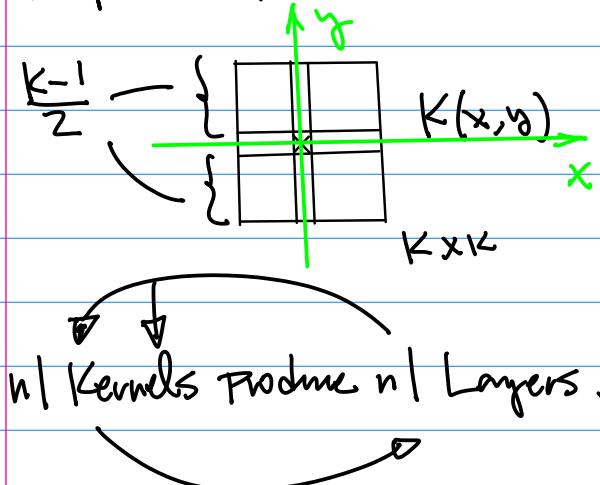


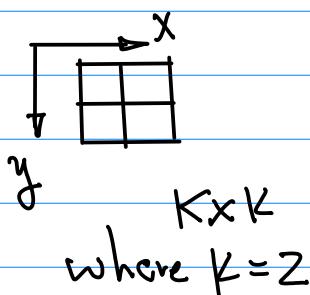
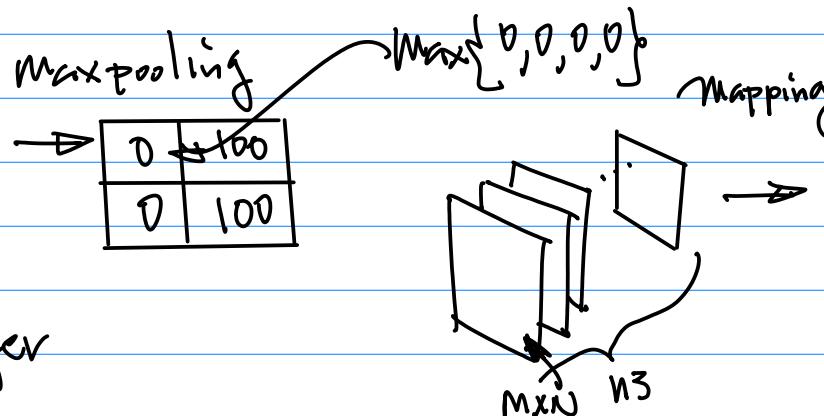
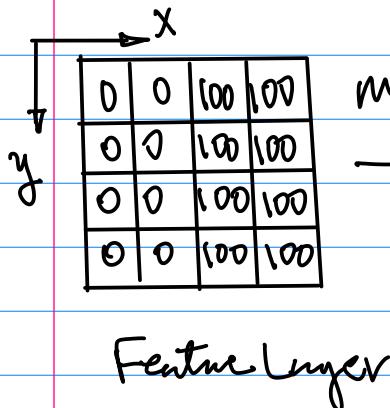
Fig.3

Take $G(x, y)$ in Fig. 3,
Map it to $K \times K$.



Now, consider a technique
for feature Reduction,

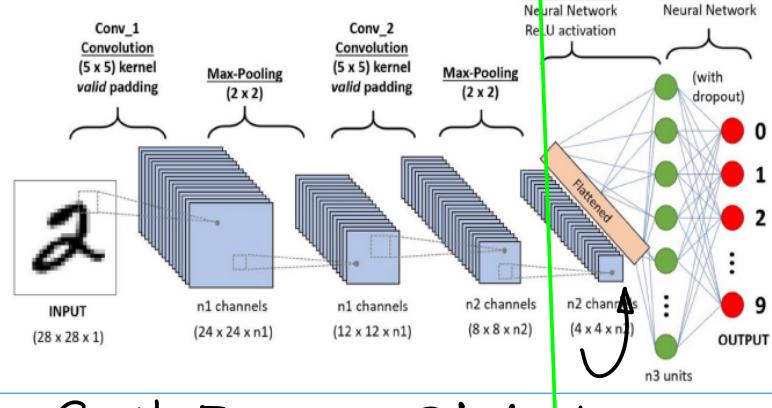
Given an image $I(x, y)$



Now, consider the Architecture of MNIST

Illustration of A CNN for Digits Recognition

<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>



Consider Re-arrange 2D feature layers
into 1D
Each
Neurons.

Example: Sample code
OpenCV for Displaying
an image

cv2.imread(imageName, cv2.IMREAD_COLOR)
cv2.imshow(window_name, src)

Sept. 7 (Th).

Note: 1^o About OpenCV
Reference/Sample
Code.

a) On github. for Video Capture

[opencv / deep-learning-2022s / 2022S-109c-v2-localization-contours.py](#)

hualili Add files via upload

to get A Single

Frame.

b) Sample for Single frame, e.g.
.jpg, .png etc Image Input.

2022S-104d ~ Python Code

c) Script for Conda Environment

On github :

2022S-104b ~

" - 104c ~

d) Readme for Creating Conda
Environment.

Note 2. Homework Due A week

from today (Sept. 17, Sun)

Requirements :

a) Installation of T.F. Version 2.0
or higher

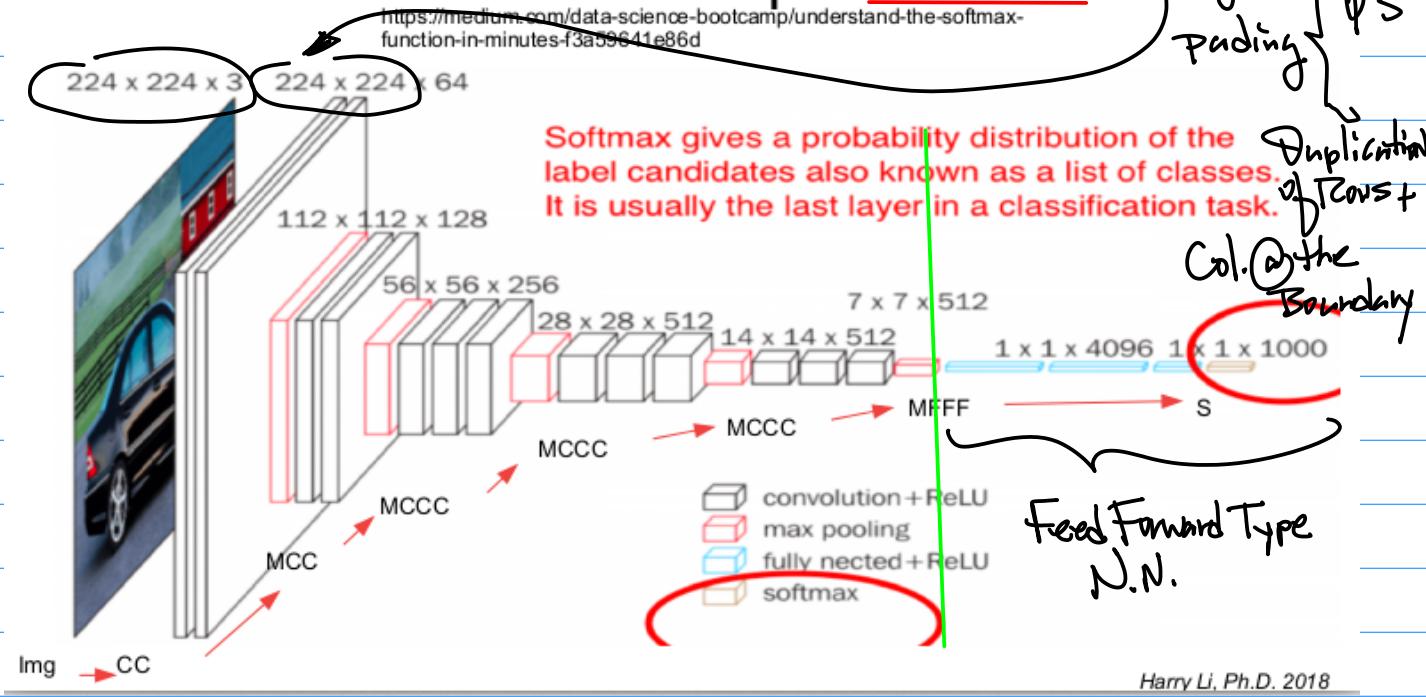
b) Screen Capture that Shows
T.F. installed Successfully.
(with Personal Identifier)

Submission on CANVAS.

Example: Convolutional Layer,
Architecture Analysis.

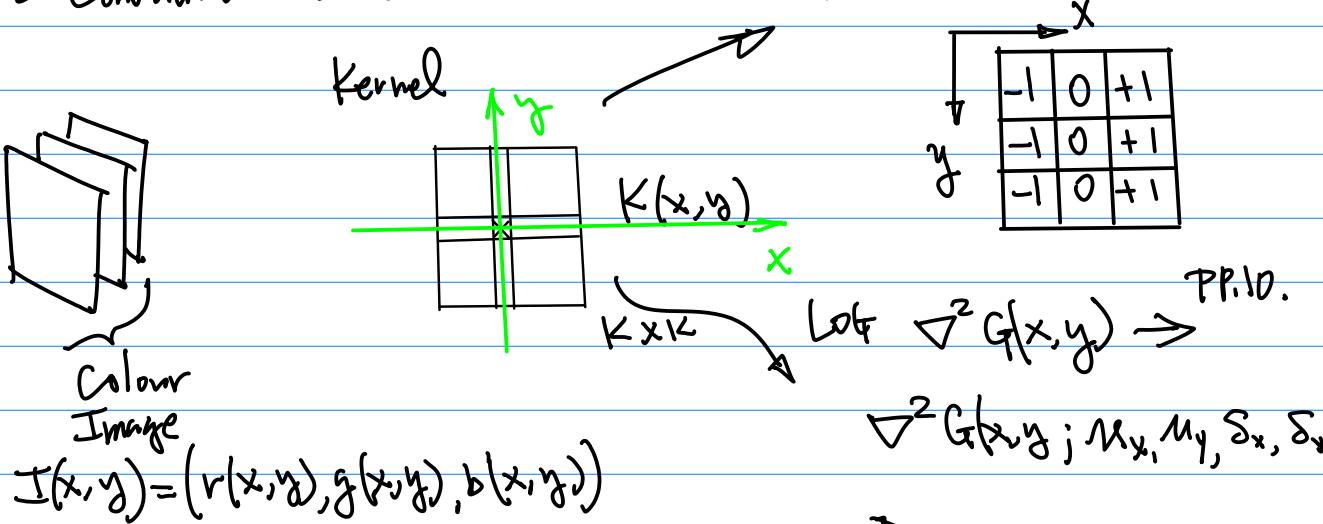
Note 1. Architecture; Note 2:
Padding v.s. No
Padding.
"ψ"s
Duplication
of Prev +
Col. @ the
Boundary

Architecture Example VGG16



Note 3: Convolution Discussion.

One Example

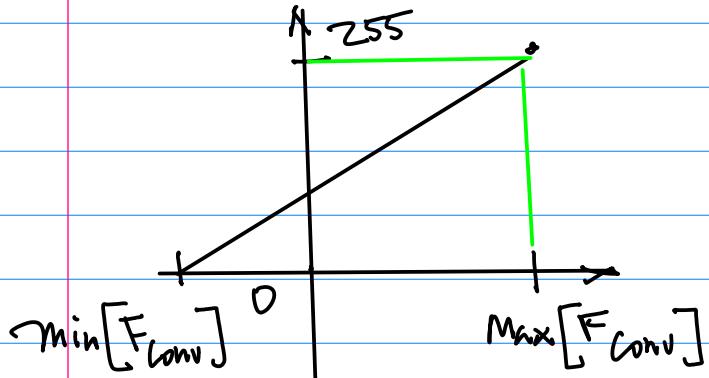


To Perform Convolution. $\rightarrow K(x, y)$

First, kernel is one channel in this case;

Then, convolutions. $r(x, y) * K(x, y)$, $g(x, y) * K(x, y)$, $b(x, y) * K(x, y)$

$$F_{\text{Conv}}(x, y) = \frac{1}{3} [r(x, y) * k(x, y) + g(x, y) * k(x, y) + b(x, y) * k(x, y)]$$



Consider A Design of S.I.D
Recognition System Using T.F &
MNIST CNN as a processing
Engine.

Requirements:

1° Live CAM Input,

JPG or JPEGP

$(1280 \times 720 \text{ (MxN)})$

Col.

Row.

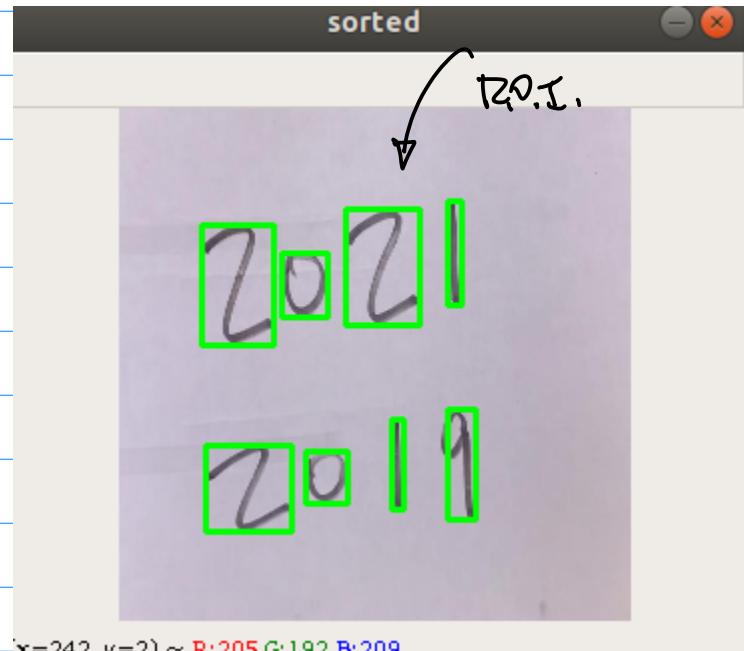
1920×1080
 $(M \times N)$

Col.

Row.

2° Printer Paper (Blank/White)
Black Mark to Write
4 Digits

3° Localize R.O.I. On Each
Digit



Region of Interest.

Sept. 12 (Tue).

Note 1. For Homework 1 Extended to
the Saturday, Requires the
Submission of Python Code.

Note 2. Team Project.

Ref:

→ 2023F-103-project-team-2023-9-12.pdf

ON CANVAS as well.

Due on Nov. 27 (Monday) 11:59 PM

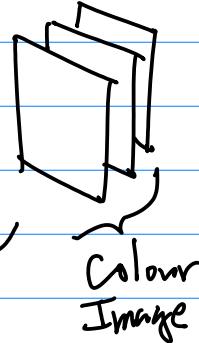
Example: Suppose a color image
 $I(x, y)$ is given below, and
a Convolution Kernel $K(x, y)$

is given as follows

$$\begin{array}{c} \downarrow y \\ \frac{1}{5} \end{array} \quad \begin{array}{c} \rightarrow x \\ \hline -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{array}$$

3×3 .

A Color Image $I(x, y)$



Find Convolution Result.

$$\begin{array}{c} \downarrow y \\ \frac{1}{5} \end{array} \quad \begin{array}{c} \rightarrow x \\ \hline 0 & 0 & 0 & 255 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}$$

$I_r(x, y)$

$$\begin{array}{c} \downarrow y \\ \frac{1}{5} \end{array} \quad \begin{array}{c} \rightarrow x \\ \hline 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 100 \\ 0 & 0 & 5 & 100 \\ 0 & 0 & 2 & 0 \end{array}$$

$I_g(x, y)$

$$\begin{array}{c} \downarrow y \\ \frac{1}{5} \end{array} \quad \begin{array}{c} \rightarrow x \\ \hline 0 & 0 & 255 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 255 & 0 \end{array}$$

$I_b(x, y)$

$$\begin{array}{c} \downarrow y \\ \frac{1}{5} \end{array} \quad \begin{array}{c} \rightarrow x \\ \hline -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{array}$$

$$\frac{1}{5} \quad \begin{array}{c} \rightarrow x \\ \hline -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{array}$$

$$\frac{1}{5} \quad \begin{array}{c} \rightarrow x \\ \hline -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{array}$$

$$\begin{array}{c} 0 & -255 \\ \hline 0 & 1 \end{array}$$

$$\begin{array}{c} -95 & 5 \\ \hline 2 & -98 \end{array}$$

$$\begin{array}{c} 255 & 255 \\ \hline 255 & 255 \end{array}$$

Note: In the future, we have need to reduce the Number of feature layers

By introducing Newer Kernel(s) for Convolution.

Homework: Due 1 week from Today.
(Sept. 19th).

1^o Installation of T.F., Screen
Capture the installation Result.
(W/Personal ID).

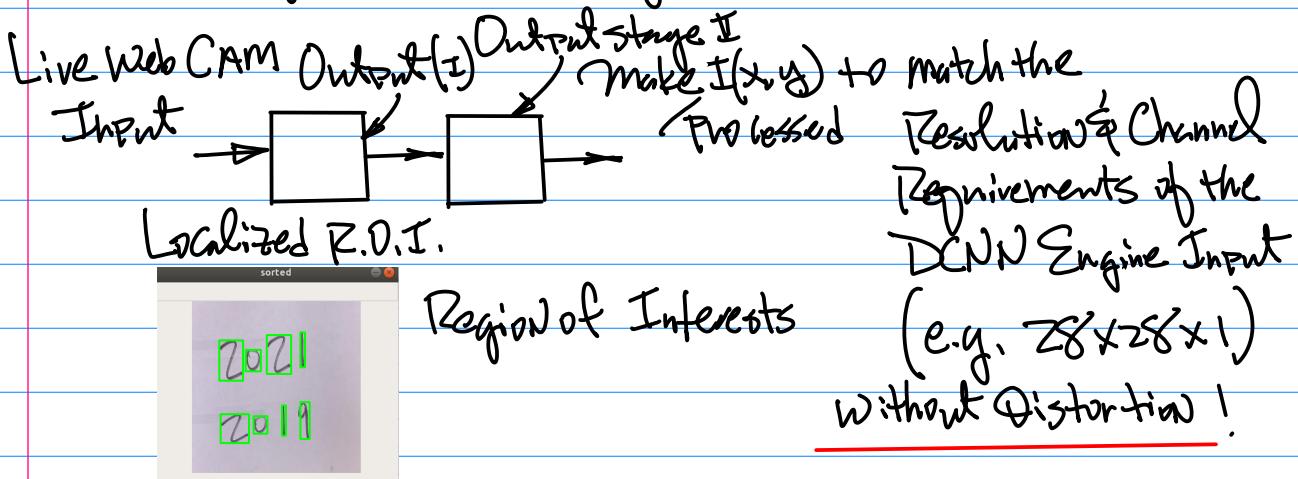
2^o OpenCV Code to Handle

- a. Live CAM Input Video,
Web
and Display it;
- b. file input, MP4G4 format
Video and display it.

Note: Please use A4 printer
paper, Write with a Black
marker, 4 Digits of your
SID.

Submission: the code &
Video Clips.

Preprocessing for Hand Written
SID Recognition System Design.



Ref:

Homework: The 1 week from today.
Use your OpenCV for the Sept 21st.
following Preprocessing functions.

1° Convert the color Image $I(x,y)$
to a gray Scale Image. $I_g(x,y)$

2° Binarize the gray scale image
to Obtain a Binary Image $I_b(x,y)$.

3° Perform Canny Edge Detection
on $I_g(x,y)$, and display the
edge map.

4° Perform Gaussian Blur on
the Gray Scale Image.

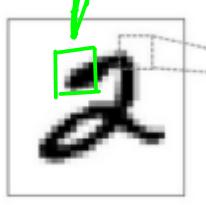
Submission:

5° Screen Capture of 1° to 4°.
with Personal Identifier.

Example: Continuation of project
Discussion. Preprocessing.

The First Objective: To get all
the Bounding Boxes

Color Video \rightarrow Gray Scale
Image.



INPUT
(28 x 28 x 1)

C. Filtering Operation, conducted
by Convolution w/ predefined
Kernel Coefficient.

→ B. To Be Continued.

Edge Detection Canny, Log,
Gabor

Note: Input Should be a
grayScale(8 bits), Output
Should be 8 bit (1 Channel)

A.

→ Binarization. To Obtain A

Representation of A Digit.

Note: Binarization May lead to the
Loss of Useful information.

Example: Image Binarization.

Given $I_g(x, y)$ or Simply $I(x, y)$

Binarization is defined as

$$B(x, y) = \begin{cases} 255 & \text{if } I(x, y) \geq T \\ 0 & \text{D/W} \end{cases} \quad \dots (1)$$

$$A = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} B(x, y) \quad \dots (3)$$

$$\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (x - \bar{x})^m (y - \bar{y})^n B(x, y) \quad \dots (4)$$

Note: T is set for the entire image for now.

Example from PP. 18, Lecture

Notes on GitHub, 2023S.

Note: Treat the top left corner

as (1,1) for the Binarized

Image Descriptors

Calculation (to Void

Skewed Result).

Most Important Descriptors
are those Built Based on
Moments.

Where

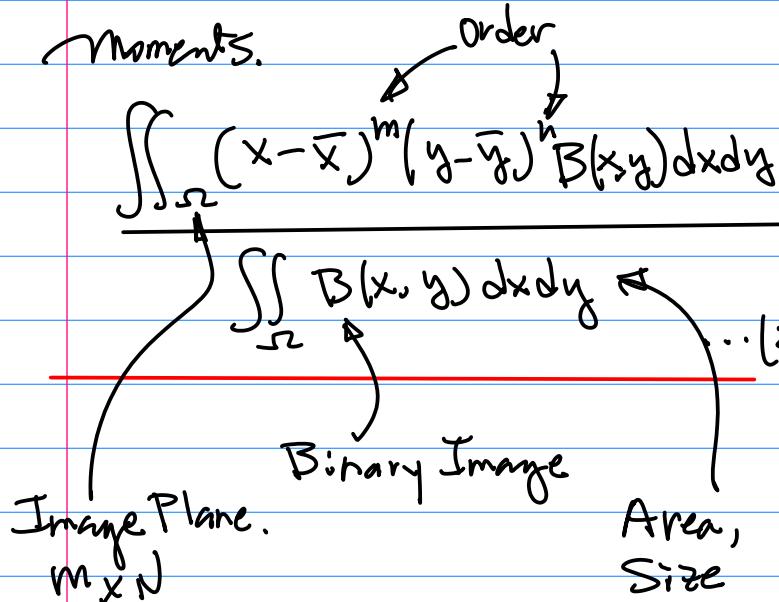
$$\bar{x} = \frac{\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x B(x, y)}{A} \quad \dots (4-a)$$

$$\bar{y} = \frac{\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} y B(x, y)}{A} \quad \dots (4-b)$$

Sept 19 (Tue).

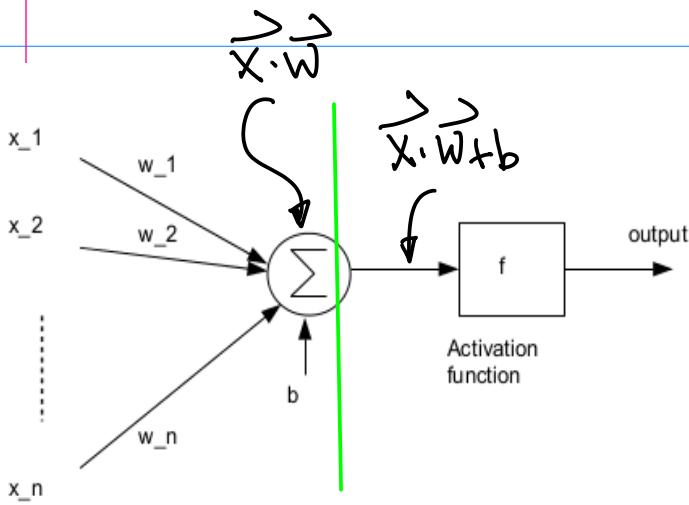
Example: Consider the Architecture
of MNIST CNN.

Roadmap:



MNIST \rightarrow Yolo. \rightarrow Yolact
CNN with
Preprocessing. To generate
Bounding Boxes
Semantic Segmentation
Creation
R.D.I.

Ref: On the github, 2022S-103C ~



$$\sum_{i=1}^n x_i w_i \text{ or } \sum_{i=1}^n w_i x_i$$

$$= \vec{x} \cdot \vec{w} \quad \dots (3)$$

Together with b (Bias), we have

$$\sum_{i=1}^n w_i x_i + b = \vec{x} \cdot \vec{w} + b \quad \dots (4)$$

Note 1. Feature vector, or Input, Excitation

$$\vec{x} = (x_1, x_2, \dots, x_n) \quad \dots (1)$$

$$\vec{w} = (w_1, w_2, \dots, w_n) \quad \dots (2)$$

Weights. $w_i \in [0, 1]$

where $i = 1, 2, \dots, n$;

b : Bias, e.g., offset

Note 2. $x_i w_i \rightarrow$ the Neuron
input 1 to

$x_2 w_2$: input 2 to
the Neuron.

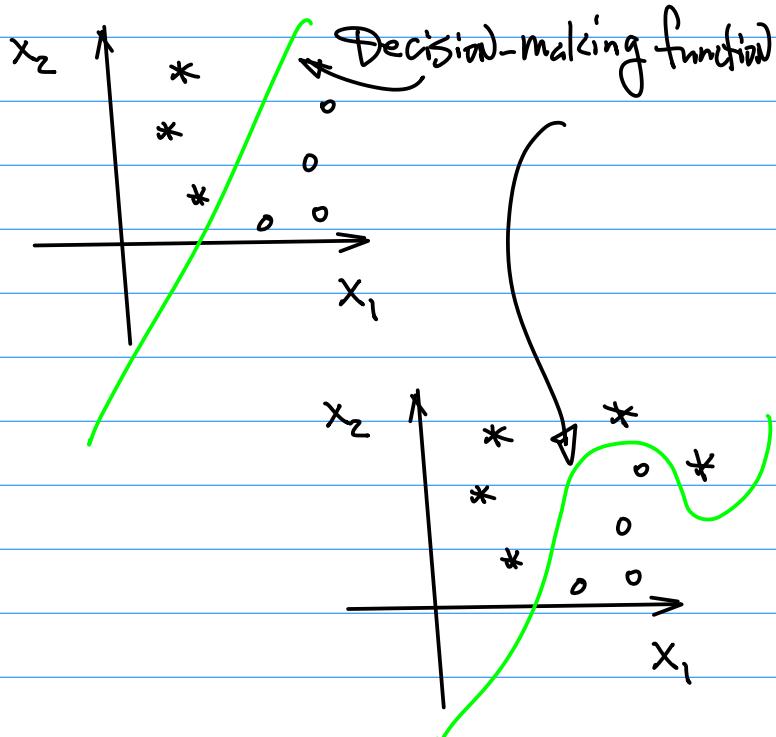
:

$x_i w_i$: .. i ..

$$x_1 w_1 + x_2 w_2 + \dots + x_i w_i + \dots + x_n w_n$$

Note 2: The "Bigger Picture" of this Part of the Architecture.

Feed forward NN.

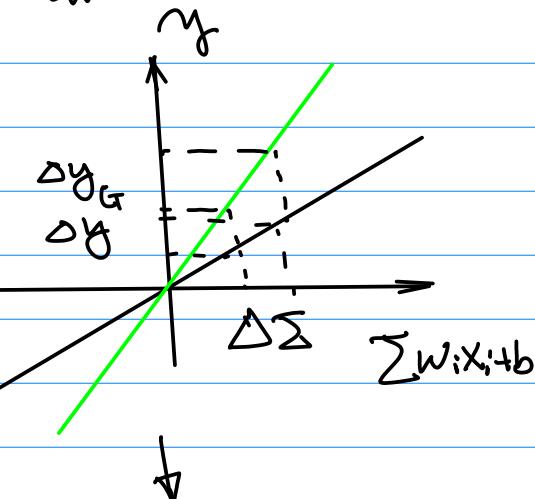
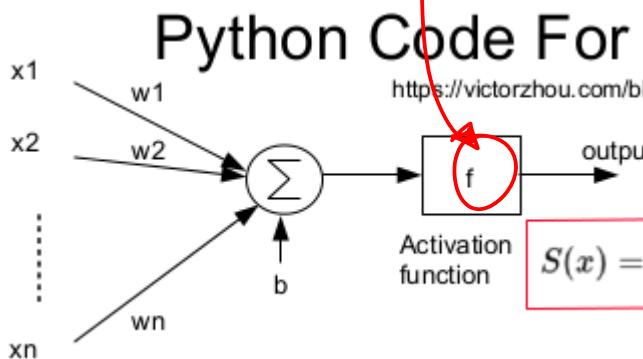


Note 3. Activation Function.

$$f\left(\sum_{i=1}^n w_i x_i + b\right)$$

... (5)

Note: RELU



$$o = f\left(\sum_{i=1}^n w_i x_i + b\right), \text{ OR}$$

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

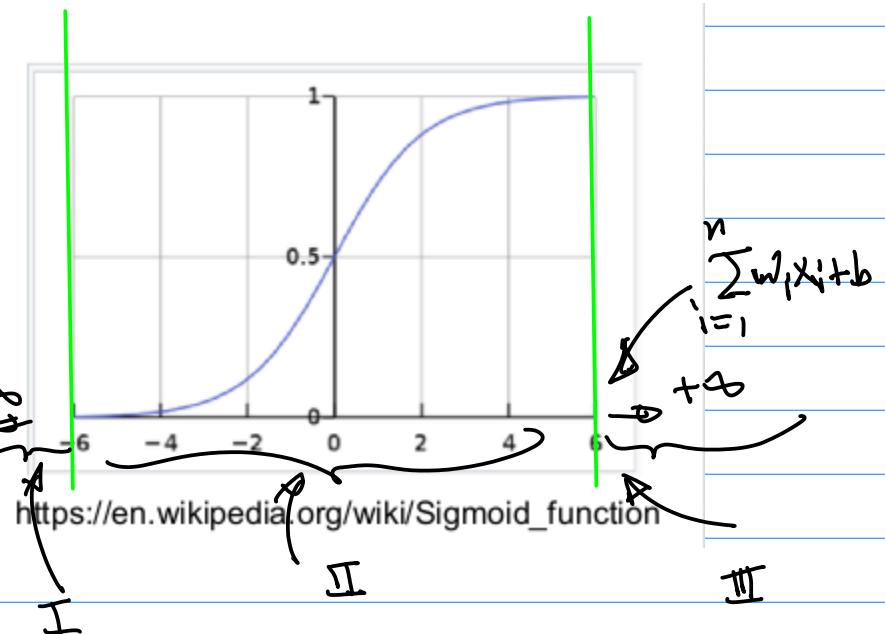
Output of the Neuron.

Note 4. Sigmoid function
victorzhou.com/blog/intro-to-neural-networks/

$\sum w_i x_i + b$

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1} \quad (1)$$

A sigmoid function is a mathematical function having a characteristic "S"-shaped curve



CMPE258

F2023

21/

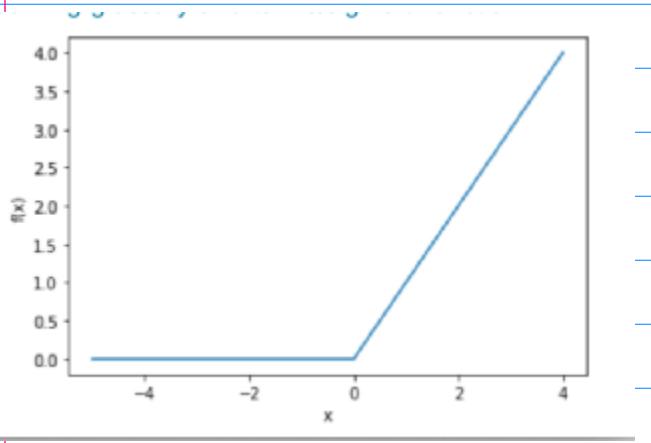
Sept. 21 (Th),

Continuation on the Formulation
of Feedforward NN.

Ref: github.



Fig. 1.



$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & \text{o/w} \end{cases} \dots (1)$$

Note: Comparison to Sigmoid
in terms of its Computation

$$S(x) = \frac{1}{1 + e^{-x}} \dots (2)$$

from Taylor Expansion.

$$f(x) = f(x_0) + \left. \frac{df}{dx} \right|_{x=x_0} (x-x_0) + \left. \frac{d^2f}{dx^2} \right|_{x=x_0} \frac{(x-x_0)^2}{2!} + \dots + \left. \frac{d^k f}{dx^k} \right|_{x=x_0} \frac{(x-x_0)^k}{k!} + \dots + R_n(x) \dots (3)$$

Sample plain python code on github.
Example.

Note: 1. define activation function

```
def sigmoid(x):  
    # Our activation function: f(x) = 1 / (1 + e^(-x))  
    return 1 / (1 + np.exp(-x))
```

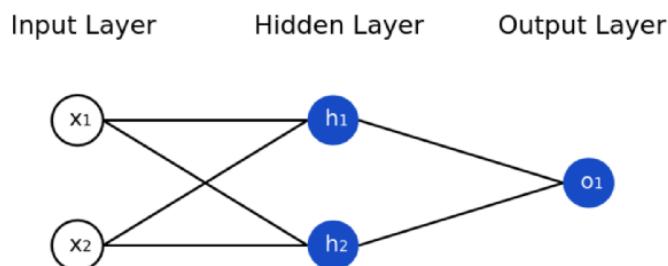
2. define a single neuron

```
class Neuron:  
    def __init__(self, weights, bias):  
        self.weights = weights  
        self.bias = bias
```

```
def feedforward(self, inputs):  
    # Weight inputs, add bias, then use the activation function  
    total = np.dot(self.weights, inputs) + self.bias  
    return sigmoid(total)
```

$$\text{Note: } y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

Can be applied to the Neurons
in the following illustration



A sigmoid
mathemat
characteris

Note 1. If output of a Neuron. { y_{pred} Compare the two, to
 y_{true} , evaluate the

Define Loss Function

<https://victorzhou.com/blog/intro-to-neural-networks/>

Performance of

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_{true} - y_{pred})^2$$

$$y_{true} - y_{pred}$$

$$(y_{true} - y_{pred})^2$$

$$\sum_{i=1}^n (y_{true} - y_{pred})^2$$

The Mean Square Error Function as a loss function very often it is also defined as objective function, e.g., minimize the loss function becomes objective function, as shown below step by step.

$$\begin{aligned} & \text{Note 2. } y_{pred} \neq y_{true} ? \\ & \quad \left\{ \begin{array}{l} y_{pred} - y_{true} = 0 \quad \text{the same} \\ y_{pred} - y_{true} > 0 \quad y_{pred} > y_{true} \\ y_{pred} - y_{true} < 0 \quad \text{o/w} \end{array} \right. \\ & \quad \rightarrow \text{Exp} \left[\sum_{i=1}^n (y_{true} - y_{pred})^2 \right] \end{aligned}$$

Note 3. Computation for a large dataset

$$\tilde{y}_1 - y_1, \tilde{y}_2 - y_2, \tilde{y}_3 - y_3, \dots$$

$$\tilde{y}_n - y_n.$$

Where \tilde{y}_i : Experiment Output.
 Actual Output.

y_i ground Truth.

$$\sum_{i=1}^n \tilde{y}_i - y_i$$

↓ To Avoid Cancellation
 of errors. Let's
 square them.

$$\sum_{i=1}^n (\tilde{y}_i - y_i)^2 \dots (4)$$

$$\overline{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)^2 \dots (4-b)$$

Average of the Accumulative
 Computation for the entire Data
 Set.

M.S.E (mean square error).

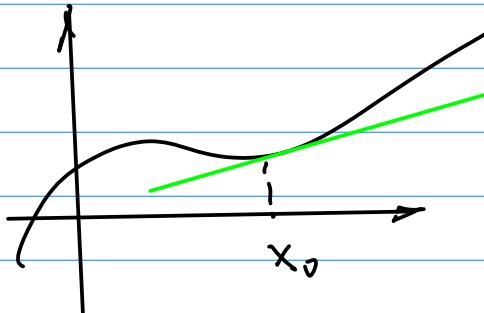
Example: Gradient Descent

Given $f(x_1, x_2, \dots, x_n)$, such as
 the objective function in
 Eqn.(4-b).

To Evaluate its performance
 for training purpose, we use
 derivative.

Consider One dimensional
Case first.

$$\frac{df(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \dots (5)$$



if Eqn(5) > 0

$$f(x+\Delta x) > f(x)$$

if Eqn(5) < 0, then

$$f(x+\Delta x) < f(x).$$

The goal is to make the objective function, e.g., loss function reduced to its minimum through the training.

For $f(x_1, x_2, \dots, x_n)$, we have partial derivatives.

$$\frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_1},$$

$$\frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_2}$$

⋮

$$\frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_n}$$

Now, to put each partial derivative together to get a whole view of the loss (e.g. Objective function), we define a gradient.

$$\left[\begin{array}{c} \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_1} \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_n} \end{array} \right]$$

... (b)

Sept. 26 (Tue).

Note 1. New Homework Posted
on the CANVAS.

k+1

$$w_1 = w_1^k + (-\eta) \frac{\partial f}{\partial w_1} \quad (1-c)$$

$$w_2 = w_2^k + (-\eta) \frac{\partial f}{\partial w_2}$$

⋮

$$w_i = w_i^k + (-\eta) \frac{\partial f}{\partial w_i}$$

⋮

$$w_n = w_n^k + (-\eta) \frac{\partial f}{\partial w_n}$$

Example: On the gradient descent
Technique.

Ref:

2022S-105c-#20-2021S-4gradie

$$(x_1^{k+1}, x_2^{k+1}) = (x_1^k, x_2^k) + [-\eta(\nabla f)^t] \quad (5)$$

Note 1. Weights, $\vec{w} = (w_1, w_2, \dots, w_n)$
where \vec{w} matches to
the feature vector $\vec{x} =$
 (x_1, x_2, \dots, x_n)

Superscript "k+1", index for
the training at Step k+1

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{pmatrix} \quad \text{... (1)}$$

" " transpose
→

$$\nabla f^t = \left(\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \dots, \frac{\partial f}{\partial w_n} \right) \quad \text{... (1b)}$$

Now, verify the equation.

From Eqn (b).

$$f(x_1, x_2) \approx f(a, b) + \frac{\partial f}{\partial x_1}(x_1 - a) + \frac{\partial f}{\partial x_2}(x_2 - b) \quad (6)$$

Consider 1D Case

$$f(x) = f(x_0) + \frac{f'(x)}{1!}(x-x_0) + \frac{f''(x)}{2!}(x-x_0)^2 + \frac{d^3 f(x)}{dx^3}(x-x_0)^3 + \dots + R_n(x). \quad \text{... (2)}$$

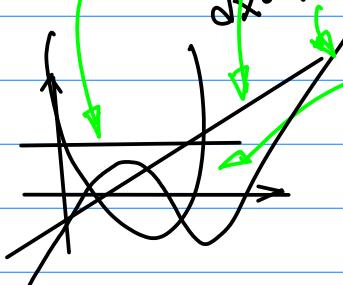


Fig. 1

The multidimensional Taylor
Expansion can be realized based
on the Similar Principle.

Simplify it, so just use the constant and the linear term \rightarrow Linear Term.

$$y = ax + b \rightarrow y' = a x + \underbrace{b + c}_{b'}$$

$$f(x_1, x_2) \approx f(a, b) + \frac{\partial f}{\partial x_1}(x_1 - a) + \frac{\partial f}{\partial x_2}(x_2 - b) \quad (6)$$

Start from here

Note: Denote $\frac{\partial f}{\partial x_1}$ as $f_{x_1}(a, b)$

$\frac{\partial f}{\partial x_2}$ as $f_{x_2}(a, b)$

$$f(x_1, x_2) - f(a, b) \approx f_{x_1}(a, b) * (x_1 - a) + f_{x_2}(a, b) * (x_2 - b) \quad (8)$$

Or,

$$f(x_1, x_2) - f(a, b) = (x_1 - a, x_2 - b) \begin{pmatrix} f_{x_1}(a, b) \\ f_{x_2}(a, b) \end{pmatrix} \quad (9)$$

Hence,

$$f(x_1, x_2) - f(a, b) = (x_1 - a, x_2 - b) \nabla f \quad (10)$$

$$f(x_1, x_2) - f(a, b) = (\Delta x_1, \Delta x_2) \nabla f = -(f_{x_1}^2 + f_{x_2}^2) \quad (12)$$

The error becomes smaller

Each iteration as long as

$\Delta x_1, \Delta x_2$ defined by Eqn (5)

$$f(x_1, x_2) - f(a, b) = -(f_{x_1}^2 + f_{x_2}^2) < 0 \quad (13)$$

Now, Review the Python Sample Code

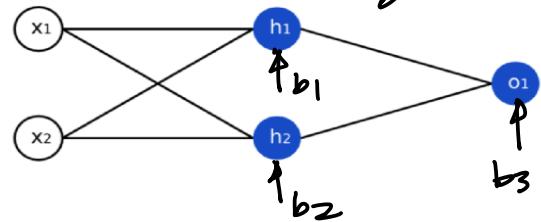
Ref



2022S-103c-#nn_sample_2022.py

Sample code for this NN.

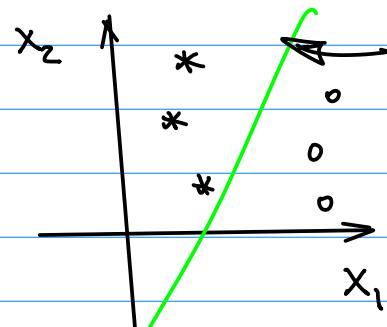
Input Layer Hidden Layer Output Layer



```

159 # -----
160 data = np.array([
161     [1, 2.5],      # person A
162     [1, 3],        # person A
163     [2.1, 3.4],   # Person A
164     [2.1, 1],     # person B
165     [3.3, 1],     # person B
166     [3, 2.3],     # person B
167     # HL 2020-9-7 Part E
168     # for the testing of adaptive learning
169     # (1) create a new program based on this o
170 ])

```



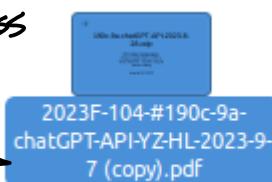
Sept. 28 (Th)

Note 1. LLM (Large Language Model) Based on ChatGPT.
Focus API Package,

Note: the 1st of the 3 PPT.

Deploy API with Python.

Ref: On the class
github.



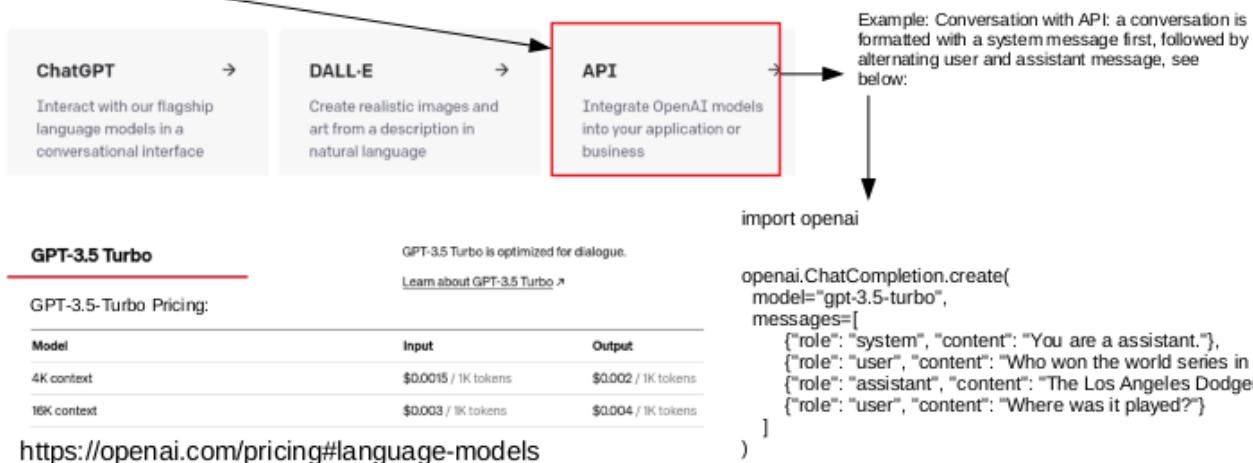
ChatGPT API from OpenAI

<https://community.openai.com/t/is-chat-gpt-provided-for-free/86249>

Step 1. Log in ChatGPT, then click on API



Step 2. Design your prompt to program your model



```
harry@harrys-gpu-laptop: ~/PycharmProjects/chatGPT
File Edit View Search Terminal Help
(base) harry@harrys-gpu-laptop:~/PycharmProjects/chatGPT$ tree -L 2
.
└── API_Key.json
    └── ChatHistory.json
        └── fine-tuning
            ├── fine-tuning.jsonl
            └── finetuning.py
            └── sample
                └── fine-tuning-10QA-2023-9-14
                    └── GPT_Fine_Tuning.zip
                    └── GPTWithHistorySaved.py

3 directories, 6 files
(base) harry@harrys-gpu-laptop:~/PycharmProjects/chatGPT$
```

Note 1. Be Sure to Save it on a Separate file from the ChatGPT(OpenAI) Page.

harry@harrys-gpu-laptop: ~/PycharmProjects/chatGPT

```

File Edit View Search Terminal Help
fine-tuning
  fine-tuning.jsonl
  finetuning.py
  sample
fine-tuning-10QA-2023-9-14
  GPT_Fine_Tuning.zip
GPTWithHistorySaved.py

3 directories, 6 files
(base) harry@harrys-gpu-laptop:~/PycharmProjects/chatGPT$ tree -L 3
.
└── API_Key.json
    ├── ChatHistory.json
    └── fine-tuning
        ├── fine-tuning.jsonl
        ├── finetuning.py
        └── sample
            └── train_data.jsonl
    └── fine-tuning-10QA-2023-9-14
        └── GPT_Fine_Tuning.zip
    GPTWithHistorySaved.py

3 directories, 7 files
(base) harry@harrys-gpu-laptop:~/PycharmProjects/chatGPT$ 
```

Notez: ".jsonl" file that user created by developer to provide training data for fine-tuning purpose

Note: Create ".jsonl" file to provide Training Data for fine-tuning.

harry@harrys-gpu-laptop: ~/PycharmProjects/chatGPT/fine-tuning

```

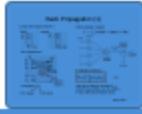
File Edit View Search Terminal Help
{
  "messages": [
    {
      "role": "system",
      "content": "CTI One Test: W100 is a smart walker to listen to your instruction and drive up to you when you call it"
    },
    {
      "role": "user",
      "content": "Tell me how can I use W100 walker."
    },
    {
      "role": "assistant",
      "content": "Make sure you have powered-up W100 and initialized it already. Then you can use your smartphone to call for its attention."
    }
  ]
} 
```

Homework, Due 1 week,
Oct. 5th.

- 1° Install openAI to
use ChatGPT API 3.5;
- 2° Create and Run "Hello,
theWorld" Python code;
- 3° Screen Captures. for 1°
and 2° with Personal ID.
- 4° Submission: On CANVAS.
as zip file. Use the
following Naming convention:
firstName_LastName_4SID_-
ChatGPT1.zip.

Consider Back propagation, or,
"Back Prop" Technique.

Ref: on the github



20225-106a-#20195-38-
lec13-BackProp-activation.
pdf

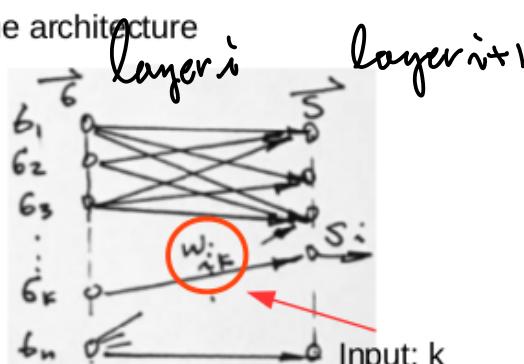
Note 1. Generalize the Architecture from a Single Neuron. Back Propagation (1)

1. input and output neurons

$$\vec{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix}_{n \times 1}$$

$$\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix}_{m \times 1}$$

The architecture



2. weights w_{ik}
i for output k for input

Note 2: for the Weights Notation

We adopted 2 subscripts

$$\vec{w} = (w_1, w_2, \dots, w_n)$$

$$(\dots, w_{ik}, \dots)$$

4. Transfer function h_i

let

$$h_i = \sum_k w_{ik} g_k \quad \dots (2)$$

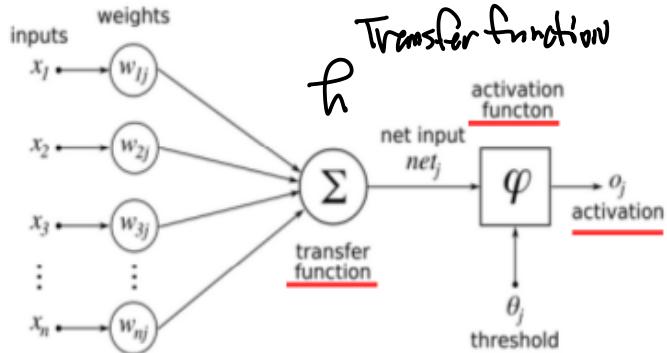
Index i for the output neuron

Neuron output function s_i

$$s_i = f(h_i) = f[h_i(w_{ik})] \quad \dots (2-1)$$

Other popular notation

$$Y = \sum (\text{weight} * \text{input}) + \text{bias}$$



3. Activation function

$$s_i = f\left(\sum_k w_{ik} g_k\right) \quad \dots (1)$$

Note the activation function $f(\cdot)$ and the bias (offset can be written in the unified summation term)

Note 3.

Harry Li, Ph.D.

$$s_i = f\left(\sum_k w_{ik} g_k\right)$$

"i" for the output layer.

Input feature vector, Excitation.

Note: The New Notation By Substitute Egn(2) into Egn(1).

5. Error at each neuron output (the difference between the true output Zeta (desired true output) and the current output S_i^M) at the experiment Mu

$$S_i^M - S_i^D$$

... (2-2)

Harry Li, Ph.D.

6. total error for all output neuron I and all experiments Mu

$$D = \frac{1}{2} \sum_m \sum_i (S_i^M - S_i^D)^2 \quad \dots (3)$$

MSE? (mean Square Error)

N : No. of Neurons

M : No. of Experiments / Epochs.

$\frac{1}{NM}$ Scaling factor

Note: Rewrite the output S_i^M by its functional form.

7. minimize the error wrt to w_{ik}

$$\begin{aligned} \frac{\partial D}{\partial w_{ik}} &= \frac{\partial}{\partial w_{ik}} \cdot \frac{1}{2} \sum_m \sum_i [S_i^M - f(h_i^M)]^2 \\ &= \sum_m [S_i^M - f(h_i^M)] f'(h_i^M) \frac{\partial h_i}{\partial w_{ik}} \end{aligned}$$

8. Training the NN by updating the weights

$$w_{ik}(t+1) = w_{ik}(t) + \delta w_{ik}(t)$$

... (4)

where

$$\delta w_{ik}(t) = -\epsilon \frac{\partial D}{\partial w_{ik}}$$

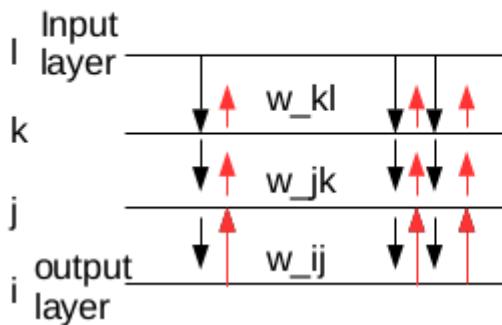
... (5)

Activation function (Transfer function)

Derivative Calculation is Based on "Chain Rule"

Note: Egn(5) on TP. 24.

10. feed forward NN (4 layers)



Blank: for Experiments.

Red: Training Update

Black for the input; red for the back prop training direction

Math. Description.

Back Propagation (4)

11. chain rule, updating the weights (training)

For layer i

$$\frac{\partial D}{\partial w_{ij}} = \sum_m [S_i^m - f(h_i^m)] f'(h_i^m) \frac{\partial h_i^m}{\partial w_{ij}}$$

Note: the training described here all related to the derivative to the activation function, $f'(\cdot)$. So selection of the activation function is important and will be discussed in details next.

For layer j

$$\frac{\partial D}{\partial w_{jk}} = \sum_n \sum_i [S_i^n - f(h_i^n)] f'(h_i^n) \frac{\partial h_i^n}{\partial S_j} \cdot \frac{\partial S_j}{\partial w_{jk}}$$

For layer k

$$\frac{\partial D}{\partial w_{kl}} = \sum_n \sum_i \sum_j [S_i^n - f(h_i^n)] f'(h_i^n) \frac{\partial h_i^n}{\partial S_j} \cdot \frac{\partial S_j}{\partial S_k} \cdot \frac{\partial S_k}{\partial w_{kl}}$$

Project 1 Due 2 weeks
from Next Monday.

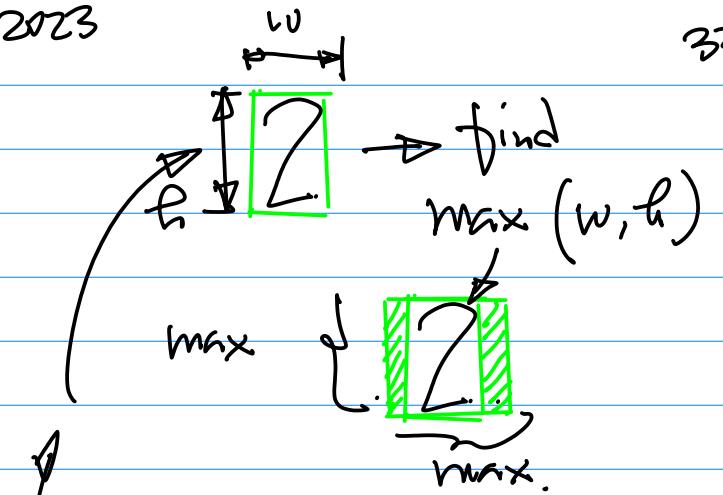
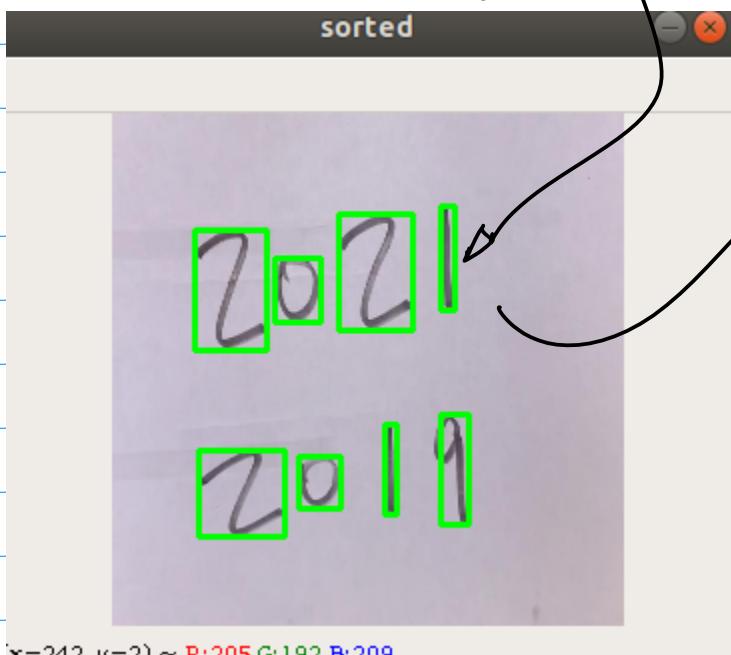
Oct. 3rd (Tue).

Note1 : Move the homework to the Sunday.

Note2: Project 1 Assignment for Handwritten SID Recognition System.

1. Objective To use either Live CAM Input or Video file input to Recognize Handwritten SID. Last 4 Digits Please keep Handwritten Digits as the inputs, No printed digits.

2. Preprocessing of the input image, to get Bounding Boxes



use OpenCV Bitwise Operations

3. Input Image Resolution (1pt).

Bounding Box $B_i(x, y; w, h)$

Width Height
Make sure in the preprocessing you will produce square image \rightarrow preserve the Aspect Ratio of the digits.

$$\lambda = \frac{h}{w} \dots w$$

Example: Preprocessing.

Ref: 2023S-101- ~ -D3-21.

Binarized Image,
Contour Analysis.
Sample Code

~~Ref: 2023S-101-Poutz-~
-D1-18.~~

Step 1. Capture Image
from a Video, Convert
it to a Gray Scale
image $I(x,y)$.

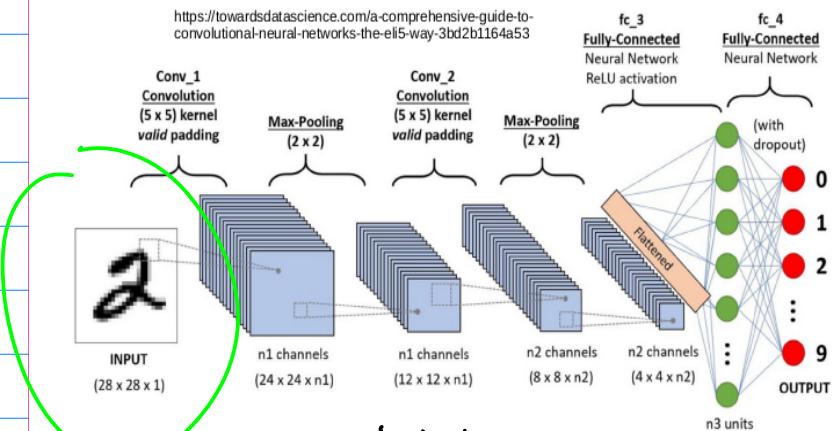
↓
Step 2. Canny Edge Detection
to produce Edge
map. $I_{\text{Canny}}(x,y)$

↓
Step 3. Contour Analysis.
Find 22 primitive
Contour features.

Draw the contours on the
input preprocessed image.

- a) W,h of a given object
- b) Bounding Box of an object
- c) Shape Descriptor
- d) Moments
- e) Others.

Illustration of A CNN for Digits Recognition



is to be
→ 28x28x1 Changed to Higher
Resolution. Such as 56x56x1
by selecting Proper Convolution
Layers / Maxpooling Layers.
please preserve the Rest of the
MNIST Architecture.

4. Comparison of your implementation
with the conventional MNIST.

5. Provide Readme file,
(good quality document),

Due Oct. 21st, Sat (11:59 PM).

Colab, and Online Tools are
Encouraged, However the submission
of the projects, Homework Require
Python code.

Step 4. Filtering Operation.

Based ON the Contour

$$\text{Mmp. } I_{\text{count}}(x, y)$$

Filter Out the Artifacts.

Look into the features

$$\text{of } I_{\text{count}}(x, y)$$

Step 5. Prepare the image ROI.

And Capture the Image(s)
from the ROI, then form
Square image, to feed
the image (digit) at
time to the CNN Engine.

Example: Continuation of Examples
on Binarized image

PP. 17 & 18

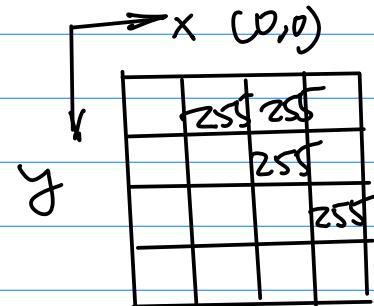
Motivation for Binarization: To
Extract Objects.

$$\bar{x} = \frac{\iint_S x \cdot B(x, y) dx dy}{\iint_S B(x, y) dx dy} \dots (1)$$

$$\downarrow \quad \iint_S B(x, y) dx dy$$

$$\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} B(x, y)$$

$$\left/ \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} B(x, y) \right. \quad \text{AREA}$$



Note: Redefine the top left
corner coordinate as (1,1)
for Hand Calculation. !

$$\begin{aligned} & \cancel{\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} B(x, y)} = \\ & \sum_{y=0}^4 \sum_{x=0}^4 B(x, y) = \\ & [B(0,0) + B(1,0) + B(2,0) + B(3,0) + B(4,0)] \\ & = B(1,1) + B(2,1) + B(3,1) + B(4,1) + \\ & B(1,2) + B(2,2) + B(3,2) + B(4,2) + \\ & B(1,3) + B(2,3) + B(3,3) + B(4,3) + \\ & B(1,4) + B(2,4) + B(3,4) + B(4,4) \end{aligned}$$

$$= 255 * 4$$

$$\begin{aligned} & \text{Similarly,} \\ & \text{Now, } \sum_{y=1}^4 \sum_{x=1}^4 x \cdot B(x, y) \\ & = \sum_{y=1}^4 [0 + B(1,1) + B(2,1) + \\ & 3 * B(3,1) + B(4,1)] \\ & = 255 * (2 * 1 + 3 * 2 + 4 * 1) \\ & = 255 * 12 \end{aligned}$$

Oct. 5 (Th)

Note 1: Check the CANVAS for the project 1, due Oct. 21st Sat;

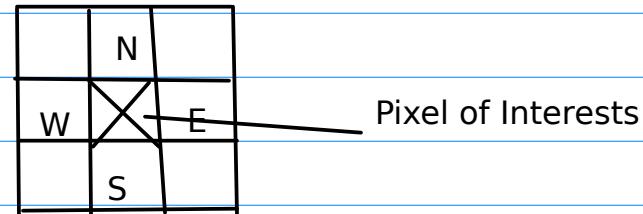
Consider edge detection, e.g., detection of sudden changes of pixel values

Example: X-bar calculation

Continuation from the previous example

hence,

$$X_{\text{bar}} = 12/4 = 3;$$

Similarly, for Y_{bar} , see the equation from the PPT of the class ref on github.

Define 4 connected neighbors N, S, E, W then the sudden change of the intensity value is defined as the difference between the Pol and any of one of its neighbors.

Denominator from the equation for Y_{bar} is the same as for X_{bar} (Area)

Similarly, we can also define 8 connected neighbors, by adding NE, NW, SE, and SW

| ... (4-a)

and $\bar{y} = \frac{\sum_{y=0}^{N-1} \sum_{x=0}^{M-1} y B(x,y)}{A}$

To pick up the sudden change, we use derivative, see pp. 22 of the lecture note

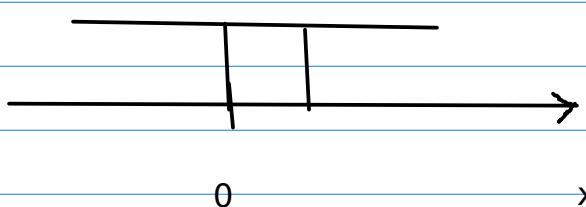
from pp. 18 of this note.

From eqn (2a) for x and (2b) for y, we have forward difference to compute derivatives

Now, let us derive convolutional kernel for this computation

Then the top part of this equation is evaluated:

$$\begin{aligned} & 255 * (1 * B(2,1) + 1 * B(3,1)) \\ & 2 * B(3,2) + 3 * B(4,3)) \\ & = 255 * 7 \end{aligned}$$



Therefore

$$Y_{\text{bar}} = 7/4$$

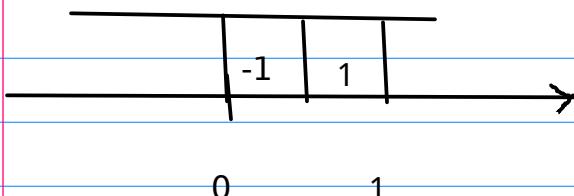
Note: keep the float for the coding implementation if needed.

Form a small grid with index 0 to map the equation (2a)

let $\Delta x = 1$ for digital image and we have $f(x+1) - f(x)$ for the right hand side of the equation,we can rewrite the above form into $1 * f(x+1, y) + (-1) * f(x, y)$ Consider edge detection technique.
ref. lecture note on the github

Map the coefficients to the template,
e.g., kernel

so we have



-1		1
1 / 2		
-1		1

Similarly, we can do the same for
y direction.

Note for $\text{LoG}(x,y)$

-1	1
----	---

0 1

Note this is a kernel but it is not
skewed, not balanced.
So with the modification of derivative
from the forward difference to
backward difference,

$$f(x,y) - f(x-1,y)$$

which give the kernel

-1	1
----	---

- 1 0

Add this kernel (BW) to the kernel (FW)
and divide the result by 2, we have

1 / 2	-1		1
	- 1	0	1

Oct. 10th (Tue).

Example: Consider Contour Analysis

Ref.

1° 2025-101-note ~ 03-24.pdf.

2° Pp. 23.

- 2025-101-example-binary.pdf
- 2025-109-contour-intro-2022-2-28.pdf
- 2025-109b-countours-roi.py
- 2025-109c-v2-localization-contours.py
- 2025-110--#lec5-1-mnist-hl-2021-3-1.pdf
- 2025-110a-MNIST-digits-saveTrained.py

PPT.

Python Sample
for
Contour
Analysis

Python Sample
Code which integrates
all the pre-processing task
Except the last one which is
the task to build a squared
Bounding Boxes

Note:

1° Notation for Contours

C — Contour

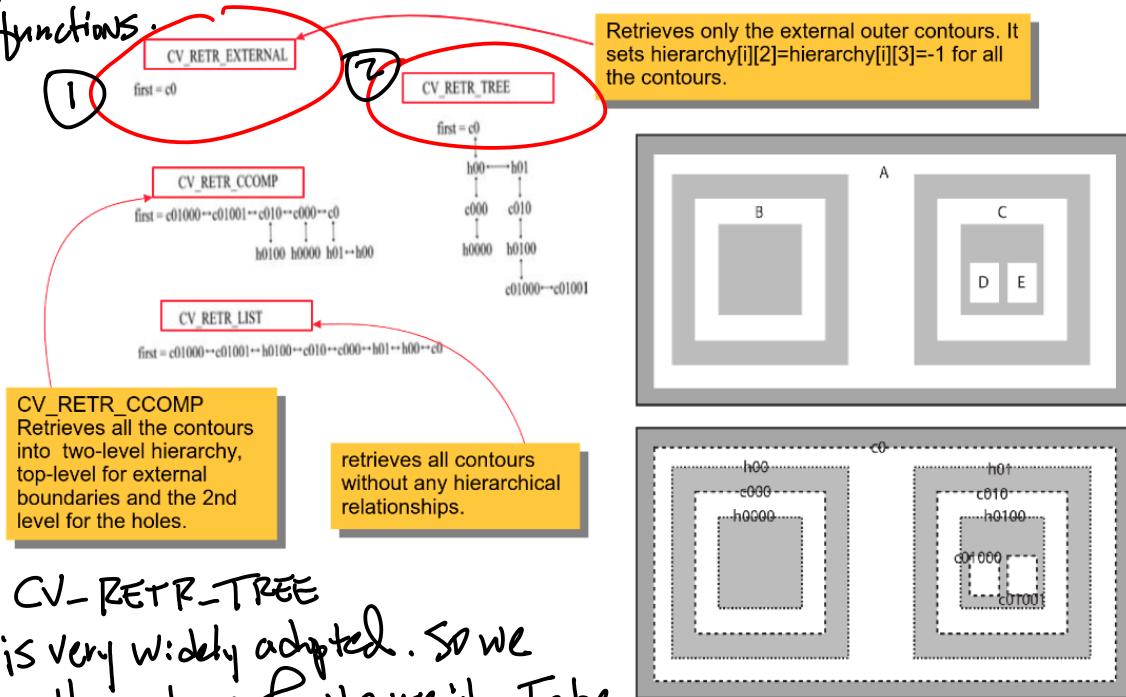
H — Hole

Note 1. Pay Attention to 2 of the

following 4 OpenCV

Contour functions:

4 Contours Models



Note 2: CV_RETR_TREE

is very widely adopted. So we will emphasize how to use it. To be able to carry out hand calculation

as well.

<http://opencvexamples.blogspot.com/2013/09/find-contour.html>

Note 3: CV_RETR_EXTERNAL can

be utilized for the SIR Handwritten

digit Recognition. Which provides

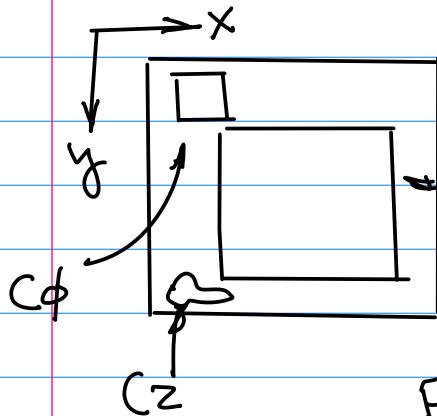
Simpler, But Adequate Contours

for Bounding Boxes finding.

Example:

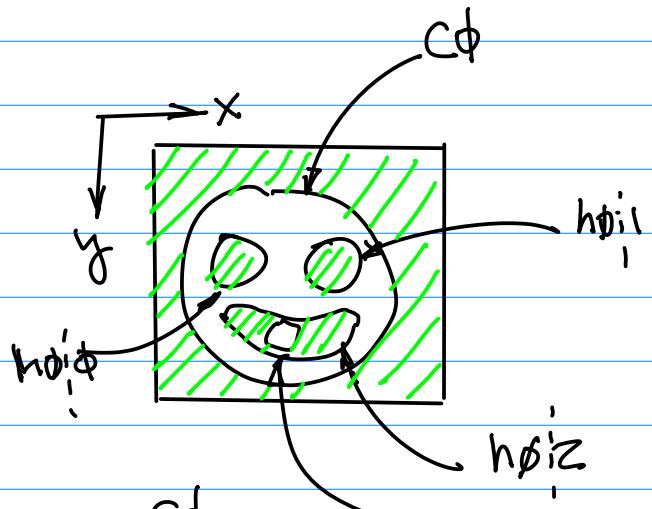
See the Example on pp. 24.

Lecture Notes 2022S-109 ~



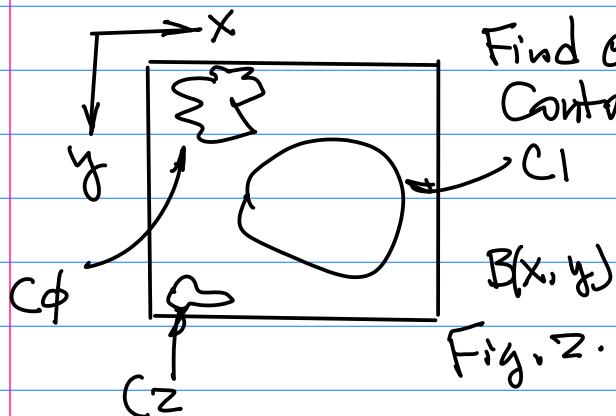
Find External
Contours.

Fig. 1.



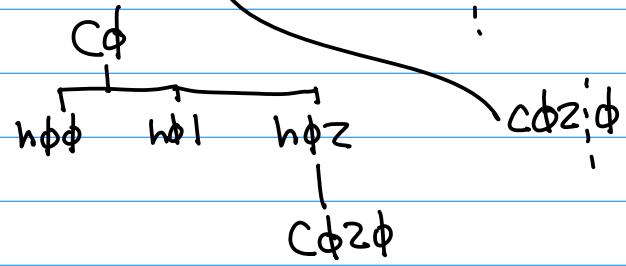
Now, find CV.RETR_CCOMP

Consider the following contour.



Find External
Contours.

Fig. 2.



Note: cv2.cvtColor();
cv2.Canny();

Sample code.

[opencv / deep-learning-2022s / 2022S-109c-v2-localization-contours.py](#)

hualili Add files via upload

Note: This Python code
Shows Contour Findings.

Code

Blame

53 lines (39 loc) · 1.5 KB



Code 55% faster with GitHub Copilot

```

1  # This is a sample Python code for preprocessing for contour analysis
2  #
3
4  def contour_example(name):
5      # Use a breakpoint in the code line below to debug your script.
6      print(f'Harry, {name}') # Press Ctrl+F8 to toggle the breakpoint.
7
8  def get_distance(x,y):
9      return math.sqrt(x*x+y*y)
10
11 # Press the green button in the gutter to run the script.
12 if __name__ == '__main__':
13
14     contour_example('Contour Example')

```

```

17 import cv2
18 import math
19
20 #img = cv2.imread('test.jpg')
21 image = input('Enter image file name:')
22
23 img = cv2.imread(image, cv2.IMREAD_COLOR)
24 img = cv2.resize(img, (256, 256))
25 cv2.imshow('original image', img)
26
27 imgray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
28 cv2.imshow('gray scale image', imgray)
29
30 thresh = cv2.Canny(imgray, 100, 200)
31 cv2.imshow('Canny', thresh)
32
33 contours, hierarchy = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
34 # Draw contour
35 thresh1 = cv2.cvtColor(thresh, cv2.COLOR_GRAY2BGR)
36 cv2.drawContours(thresh1, contours, -1, (0, 255, 0), 5) # all contours
37 cv2.imshow('All contours', thresh1)
38 #Bounding box

```

Note 3.

`cv.Canny()`
Edge Detection with
Upper and Lower Thresholds

Note 5: `CV2.drawContours()` Memorize this function!

Note 6. Filtering

```

40 for i in range(len(contours)):
41     [x, y, w, h] = cv2.boundingRect(contours[i])
42     # Filter out smaller noise regions
43     if (w > 5) & (h > 25):
44         print("i:", i)
45         cv2.rectangle(img, (x, y), (x + w, y + h), (0, 255, 0), 2)
46 cv2.imshow('sorted', img)
47

```

```
48 print(img.shape)
```

```

50 cropped2 = img[10:20, 10:30]
51 cv2.imshow('cropped2:', cropped2)
52 cv2.waitKey(0)
53 cv2.destroyAllWindows()

```

Note 1. Be Careful for `CV2.resize()`

: It forces the image to Resize to 256x256, which distorts the Original Image. Not Recommended, Should use $w/2, h/2$ as a better practice to preserve the Aspect Ratio

Note 2. OpenCV Color Image Format "BGR"

Not RGB.

Memorize this function!

Note 4. `contours, hierarchy = CV2.findContours()`

Keywords: `CV2.RETR_EXTERNAL`
`CV2.CHAIN_APPROX_SIMPLE`

Memorize this function!

Note 6. Bounding Boxes

Note 7. Crop Image To Prepare for Bitwise Operation in the Squeezed Image Construction.

Oct 12(Th).

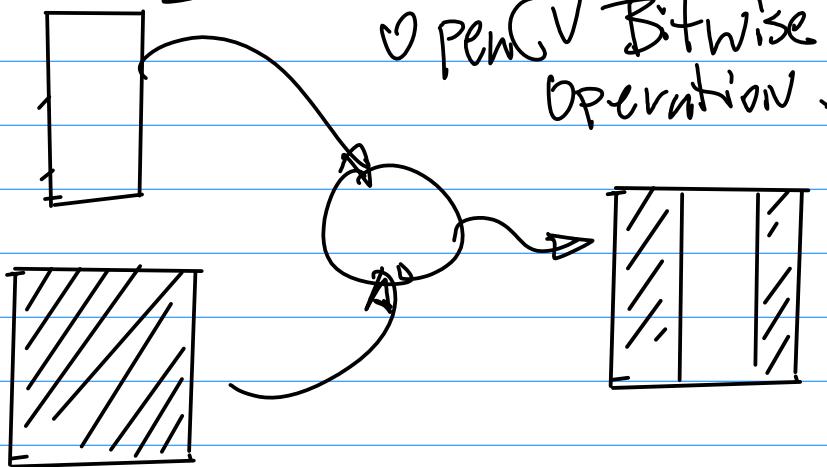
Today's Topics:

1° Sample code
for ROI localization.

2° Yolo

Example: Same code, Line 50.

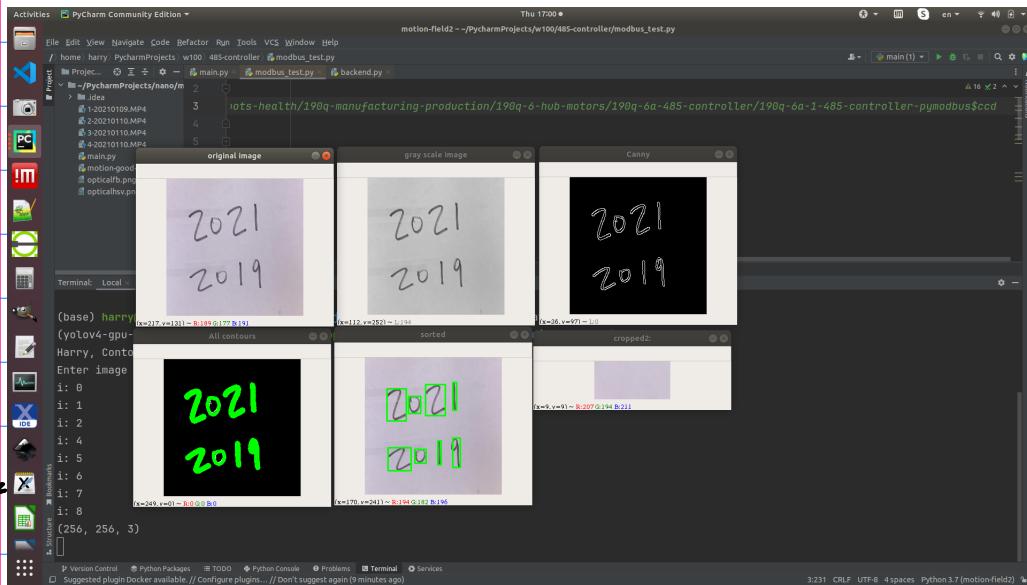
PP. 37. Cropped Image from the ROI.



$K \times K$
Square Image
(Background)

$$K = \max(w, h)$$

Example:



Note : The 2nd Sample code
Processed Results Displayed
With Caption .

Now, Consider YOLO Technique .
Note1: YOLO (You Only Look Once)

Note2: Motivation .

Among other Objectives
2 of them plays
Significant Roles here .

1. To Automate the
R.O.I / Bounding
Boxes Creation .

2. To Expand the

Object Recognition
Capability Beyond what
we have so far .



Base YOLO model runs at 45 FPS. A smaller version, Fast YOLO, runs astounding 155 FPS second, outperforms DPM (deformable parts models) and R-CNN.

Figure 1: The YOLO Detection System. (1) resizes image to 448×448 , (2) runs a convolutional network, and (3) thresholds the result by the model's confidence.

YOLO github Code and readme <https://github.com/huaweiopenlab/master/deep-learning-2022s/2022F-106-YOLO4-README-Tiny-YOLO-YY-2022-9-28.pdf>

Tutorial: <https://pjreddie.com/darknet/yolo/>

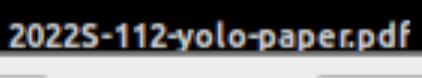
GitHub, Inc. (US) git clone <https://github.com/pjreddie/darknet>

The Approach :

1^o Research Paper.

2^o Github Code

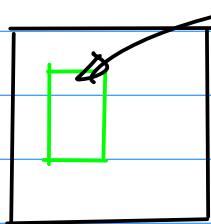
With Readme
provided from the
Class. On the Class
github.

2022S-112-yolo-paper.pdf

Oct. 17 (Tue).

Q & A. Bounding Boxes
Filtering.

Objective: To Detect
These Boxes which do
not contains a Digit



Any Digits.
↓
Single color

Hence, Any Pattern
inside the Box has more 2
Colors, than we want to Remove
it.

"Floodyfill" for Coloring operation

(1) "Seed" point is Needed.

(2) feed the Seed to the floodyfill.

Note 1. Homework on Yolo Installation.

Note 2. Oct. 24 (Tue). Quiz, 15 + 15

① Bring your Laptop. 15 minutes.
Run the Code including Prep of
OpenCV, NN Introduction, Submission
of your work.

② Screen Capture of the Code Execution.
With personal identifier.

③ Please Bring 1 Blank Paper to
Write your Answer.

④ Then take a photo. → pdf, then
integrates All pdf files into One pdf.
With following Naming Convention.

FirstName-LastName-SID-Quiz-

258.zip.

⑤ Submission to the
Canvas.

YOLO (You Only Look Once).

Ref: Yolo Paper, See the github.

Overview of the Paper.

Example: Notations & Formulation.

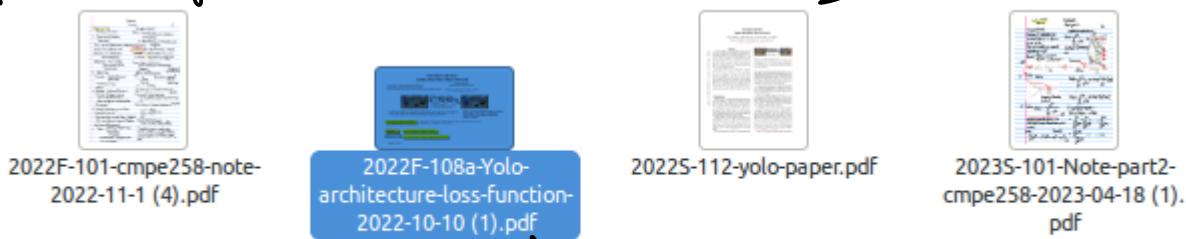
Cmpe258

F2023

Ref on the github.

Paper

40



↑ PPT.
↓ Overview.

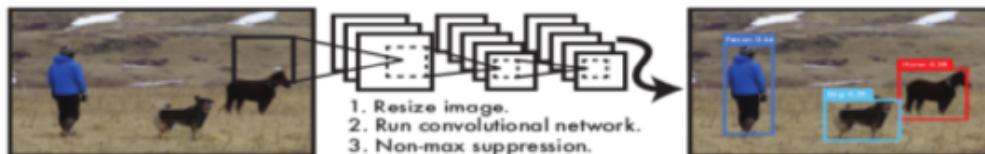
You Only Look Once: Unified, Real-Time Object Detection

Research paper: <https://arxiv.org/pdf/1506.02640v5.pdf>

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

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/>



Base YOLO model runs at 45 FPS. A smaller version, Fast YOLO, runs astounding 155 FPS second, outperforms DPM (deformable parts models) and R-CNN.

Figure 1: The YOLO Detection System. (1) resizes image to 448×448 , (2) runs a convolutional network, and (3) thresholds the result by the model's confidence.

YOLO github Code and readme

<https://github.com/huailili/opencv/blob/master/deep-learning-2022s/2022F-106-YOLO4-README-Tiny-Yolo4-GPU-YY-2022-9-28.pdf>

Tutorial:

<https://pjreddie.com/darknet/yolo/>

GitHub, Inc. (US)

`git clone https://github.com/pjreddie/darknet`

Harry Li, Ph.D.

Notations for the formulation.
See Ref Below.



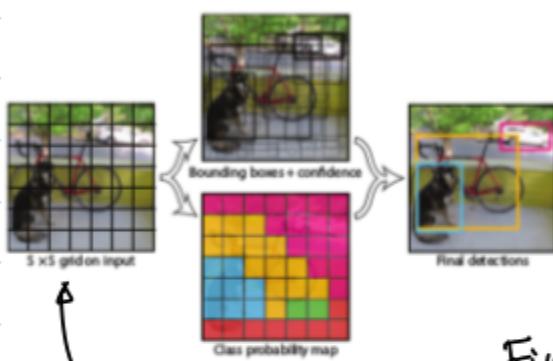
CMPE258
Oct.13,22

Fig.2c

1. Image $I(x,y)$ is divided into $S \times S$ Grids.

Denote it as $G_{p,q}(x,y), \dots (1)$
where $p,q = 0, 1, 2, \dots$ indicate the location of the grid. (See Ref. Lecture Note, pp40)

So, we have $\overset{x}{G}_{0,0}(x,y), G_{0,1}(x,y), \dots$ etc.
from Left to the Right.

$G_{p,q}(x,y)$ where $p,q = 0, 1, \dots, \dots (1)$

2. Bounding Box:

$B_{i,j}(x,y)$. $i, j = 0, 1, 2, \dots \dots (2)$

Note: A $B_{i,j}(x,y)$ may be inside $G_{p,q}(x,y)$, or Overlaps with $G_{p,q}(x,y)$, OR

Could be Across the Boundary of a $G_{p,q}(x,y)$, OR
occupy multiple $G_{p,q}(x,y)$

Note: Parameters to Characterize $B_{i,j}(x,y)$.

Location (x, y) . Top Left

Corner,

Shape, w, h width,
height;

$B_{i,j}(x,y, w, h)$

↓
Object classification for
the Bounding Box.

See Example Next Page.

Compare to the Objective function, loss function, $B_{ij}(x, y, w, h)$

$$\lambda_1 \sum_{i=0}^S \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$\lambda_2 + \lambda_{\text{coord}} \sum_{i=0}^S \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$

$$\lambda_1 + \lambda_2 = 1.0$$

Assume $\lambda_1 = \lambda_2 \rightarrow$ Equal weight,

Equally Important.

Note: Suppose make location of $B_{ij}(x, y)$ twice as important as the shape. So, we have

$$\lambda_2 = 2\lambda_1$$

Hence, $2\lambda_1 + \lambda_1 = 1.$

$$\lambda_1 = \frac{1}{3}.$$

$$\therefore \lambda_2 = \frac{2}{3}.$$

Oct 1(a)(Th).

Note: A week from today, presentation (Abstract) for the Semester Long Project Proposal. Please read the team Project Requirements on CANVAS, Prepare the presentation to address.

- 1° Title of the object ;
- 2° Technical Challenge(s).
- 3° Proposed Work/ Approach to address the challenges
- 4° Existing github code. and primary Research Paper.

- 5° Expected Results & Deliverables

Example: Continuation of the formulation for YOLO Technique. Ref: github ↗

2023S-101-Note-e-part 2-cmpe258-2023-04-18 (1).pdf

$$B_{ij}(x, y, w, h, f)$$

↑
Confidence Index to describe how like the object detected inside the Bounding Box Belongs to Object Class n.

From the Lecture Note, see Ref.



3. Five Parameters to define each Bounding Box.

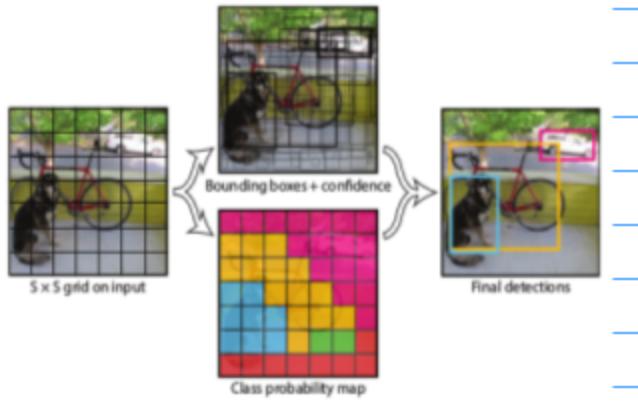
(x, y) : Location of the top left corner of $B_{ij}(x, y)$

w, h : Width and Height of $B_{ij}(x, y)$

f : Confidence level / Probability distribution to describe

The likelihood of the B.B (B^z) belongs to a certain Class of objects.

(x, y, w, h, f) ... (3)



Spiral

Note 1. α . \curvearrowleft Graph (x, y) Grid.

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.

$b. B_{ij}(x, y)$ Class probability.

$c. (x, y, w, h, f)$ Probability
Confidence.

Apriori (Th).

Example: Discussion on Notation/Formulation

From the lecture Note. Ref. on github.

Note 1. a. (Grid_(x,y)) Grid.

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.

b. B_j(x,y) Class probability.

c. (x,y,w,h,f)

Probability
Confidence.

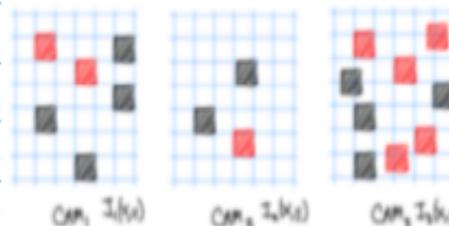
Appli(Th).

Example: Discussion on Notation/Formulation.

Ref.

2022F-101-cmpe258-note-2022-11-1.pdf

PP38



Camera 1: I₁(x,y) Camera 2: I₂(x,y) Camera 3: I₃(x,y)

$$R = RI_1 + RI_2 + RI_3 \quad \dots (1)$$

R: Red Squares, Persons.

B: (Blank) Vehicles . for "Union"
"U"

Intersection. "∩"

Consider the probability of the event

"R" (meaning Person(s) being Capture
on any one of these images).

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

36

$$I_1 \cap I_2 \cap I_3 = \emptyset \quad (\text{Empty set})$$

Consider Each Individual Camera:

$$\text{Prob}(RI_1) = \text{Prob}(R|I_1) \text{Prob}(I_1) \quad \dots (3a)$$

Similarly,

$$\text{Prob}(RI_2) = \text{Prob}(R|I_2) \text{Prob}(I_2) \quad \dots (3b)$$

$$\text{Prob}(RI_3) = \text{Prob}(R|I_3) \text{Prob}(I_3) \quad \dots (3c)$$

Rewrite Eqn (2):

$$\begin{aligned} \text{Prob}(R) &= \text{Prob}(R|I_1) \text{Prob}(I_1) \\ &+ \text{Prob}(R|I_2) \text{Prob}(I_2) \\ &+ \text{Prob}(R|I_3) \text{Prob}(I_3) \\ &= \sum_{i=1}^3 \text{Prob}(R|I_i) \text{Prob}(I_i) \quad \dots (4) \end{aligned}$$

Ref: 2022S-112-yolo-paper.pdf

37

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

Note 1. Condition Probability. $\Pr(\text{Class}_i | \text{Obj}) = \frac{\Pr(\text{Class}_i \cap \text{Obj})}{\Pr(\text{Obj})}$

1. IOU (Intersection of Union)

Index for the purpose of Comparing

2 Bounding Boxes at a time

pp.37

Example: pp.36

5. IOU (Intersection of Union). =

Example: Illustration of IOU

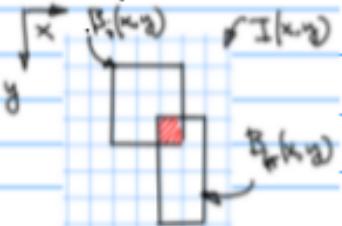


Fig.3

Fig.5

Note 1.0

$B_{ij}(x,y)$
for the i,j

The Union of $B_{ij}(x,y)$, $B_{gt}(x,y)$



Note: "Gr"

Ground Truth.

Sol. First, find the Number of pixels of the $B_{ij}(x,y) \cap B_{gt}(x,y)$,

$$B_{ij}(x,y) \cap B_{gt}(x,y) =$$

then, Find the Union

$$B_{ij}(x,y) \cup B_{gt}(x,y) =$$

$$N[B_{ij}(x,y)] + N[B_{gt}(x,y)]$$

$$- N[B_{ij}(x,y) \cap B_{gt}(x,y)]$$

$$= (3 \times 3) + (2 \times 4) - 1 = 9 + 8 - 1$$

$$= 17 - 1 = 16$$

$$\therefore \text{IOU} = \frac{1}{16}$$

$$\text{IOU} = \frac{\text{Intersection of } B_{ij}(x,y) \text{ and } B_{gt}(x,y)}{\text{The Union of } B_{ij}(x,y) \text{ and } B_{gt}(x,y)}$$

... (5)

pp.37

$$C = \frac{\text{IOU}}{\text{Confidence}}$$

(IOU)
Confidence,
denoted as C

... (4)

Now, from Eq.(1) of the Ref. (Research Paper)

Note,

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

$\text{IOU}^{\text{truth}} \leftarrow \text{G.T. B.B.}$

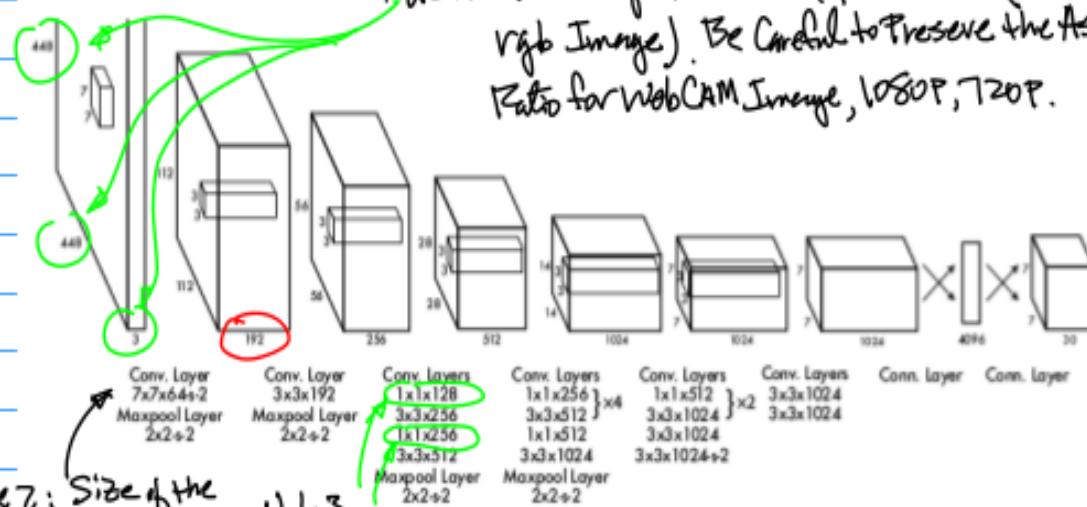
$\text{IOU}^{\text{pred}} \leftarrow \text{Predicted B.B.}$

Hand Calculation of IOU.

YODA
Example: Architecture:

Spring 2023

38



Note 2: Size of the Kernel: 7×7 , 64 of them.

Note 3: 1×1 Convolution is utilized here

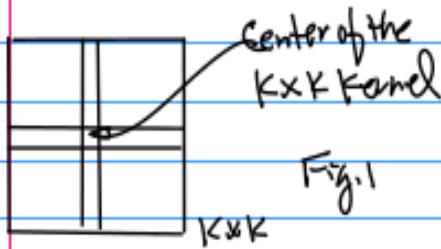


Fig.1

for $K \times K$ 2D Convolution,
the output of the convolution.

"Spatial Information", Neighbouring pixels under the $K \times K$ Kernel are counted for (for feature extraction @ the center of the Kernel)

Applies to:

Presentation (Brief) on Each Team Project.

Example: Continuation on

Math Formulation.

1×1 Convolution.

Note 1. Background on $K \times K$ Convolution.

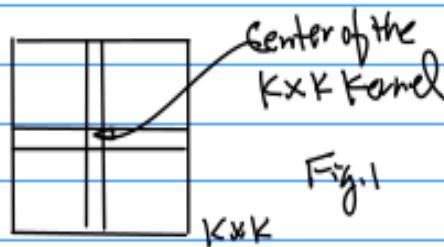


Fig.1

Output: 1 pixel
Input: $K \times K$ pixels.

Captures All Neighbouring pixels at a time, And Produce the pixel Output.

Note 2. for each convolution kernel, the convolution conducted will result in One Output Feature Layer

Oct. 24 (Tue).

Note 1: Semester Long Project

Abstract Presentation

On Oct. 26th.

1^o Check the CANVAS

Requirements. On Abstract.

2^o One page Abstract.

Title,

Technical Challenges

Methodology / Techniques

Hardware / Software Resources

Deliverables.

Table of Responsibility.

Consider the Architecture for Yolo.

Example: Continuation. Ref. PP41.

Combine

Note: Convolution + SubSampling operations into One.

a. $448 \times 448 \times 3$

b. Convolution with

Stride=2. "Scan" the input layer at every two pixels from left to right, then at every other row from top to bottom.

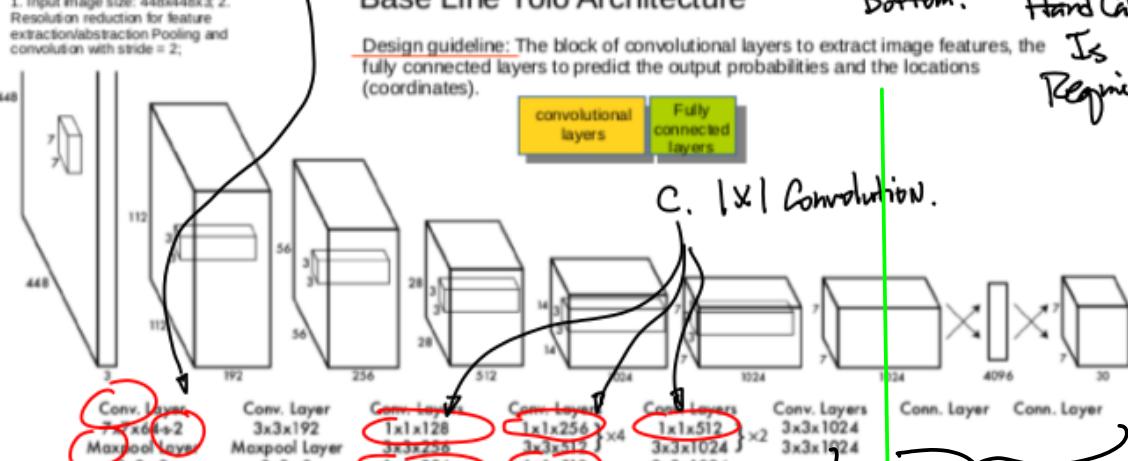
Base Line Yolo Architecture

Bottom. Hand Calculation

Is Required.

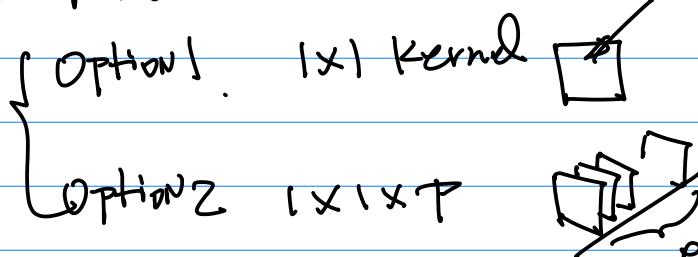
Design guideline: The block of convolutional layers to extract image features, the fully connected layers to predict the output probabilities and the locations (coordinates).

c. 1×1 Convolution.

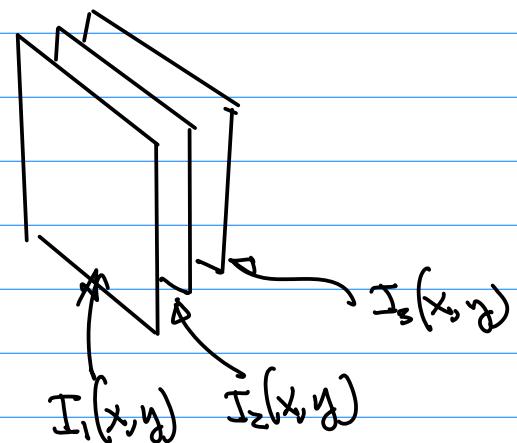


Decision Fig. 1 making.

For 1×1 Convolution.



Coefficient Suppose $P=3$.



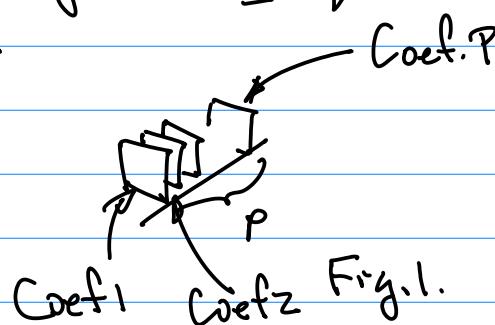
Convolution with the kernel in Option 1

produces duplication of the feature

layer but weighted by its
Coefficient d .

Convolution with Option 2 on

a set of feature layers will allow us to integrate multiple features layers into one feature layer.



$$\sum_i \text{Coef}_i = 1$$

In addition, we can assign different coefficient to emphasize the contributions from some feature layers than the others.

Oct. 26 (Th).

Project Presentation in Class.

Oct. 31 (Tue)

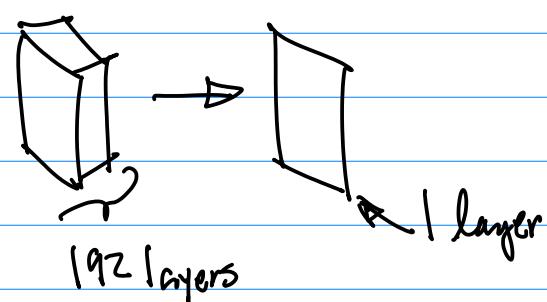
Subject today : $1 \times 1 \times K$ Convolution

Example: General Guideline :

To Reduce the Number of feature layers to A manageable level.

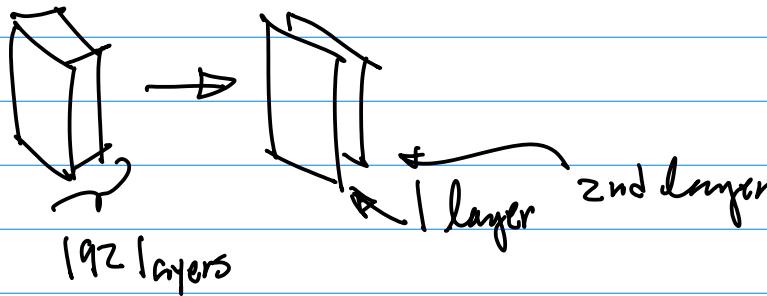
e.g. from K channels (Layers) to 1 from Fig. 1 where the layers = P .

$$1 \times 1 \times P \Big|_{P=192}$$



for the equal contribution from each layer, we can make

$$\text{Coefficient } \alpha_i = \frac{1}{192}, i=1, 2, \dots, 192.$$



For the above requirements,

Two 1×1 kernels

$$1 \times 1 \times P_1 \text{ and } 1 \times 1 \times P_2$$

$$P_1 + P_2 = 192 \text{ AND}$$

$$P_1 = P_2$$

Ref: On the class github.

2022F-108 a- ...

Note: To Combine $N=5$ Layers (Feature Layers)

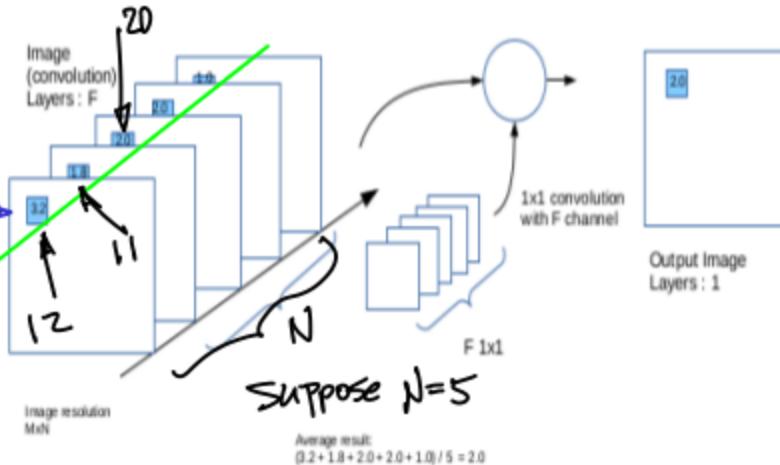
The 1×1 convolution enables dimension reduction by reducing the number of channels in convolution layers

$$1 \times 1 \times 5 \quad \begin{cases} \alpha_2 = 2\alpha_1, \dots (1) \\ \alpha_1 + \alpha_2 = 1 \dots (2) \end{cases}$$

- Suppose the input layers are $C \times H \times W$, where C is its channels. The 1×1 convolution generates one average result in shape $H \times W$. The filter is a vector of length C .
- Now if you have F 1×1 filters, you get F layers of output, the output shape is $F \times H \times W$. For input layer $C \times H \times W$ with F 1×1 convolution (with channel is C), you will get $F \times H \times W$ layers.

Note1.

One pixel across the entire stack of the feature layer.



Note2. Suppose N_b , No. of feature layers. We would like to have

2 output feature layers.

a) the 1st 3 features is defined as Group 1.

b) the 2nd 3 feature layers is defined as Group 2.

c) Output 1 feature Layer each from Group 1 and 2 respectively.

Design $1 \times 1 \times P_1$ and $1 \times 1 \times P_2$ convolutions.

d) Suppose Group 2 has twice as much contribution as Group 1.

$$\alpha_2 = 2\alpha_1 \dots (1)$$

$$\alpha_1 + \alpha_2 = 1 \dots (2)$$

Substitute (1) into (2) we have

$$\alpha_1 + 2\alpha_1 = 1$$

$$\therefore \alpha_1 = \frac{1}{3}$$

$$\alpha_2 = 2/3$$

Once the coefficients are set,
Convolution can be conducted
as follows. for Example for $\alpha_1 = \alpha_2 = \dots = \alpha_5 = \frac{1}{5}$
 $N=5$.

pp.40. from the Ref.

$$\frac{1}{N} [I_1(x_1, y_1) + I_2(x_1, y_1) + \dots + I_5(x_1, y_1)]$$

... (i)

$$\frac{1}{N} I_1(x_1, y_1) + \frac{1}{N} I_2(x_1, y_1) + \dots + \frac{1}{N} I_5(x_1, y_1)$$

Ref.

Example: pp.43 Equal contribution by

Given 3 Feature Layers

1 1 3	2 0 -1	2 3 2
2 0 -5	1 0 2	1 0 2
1 1 0	0 0 1	6 0 1

$$F_1(x_1) \quad F_2(x_1) \quad F_3(x_1)$$

3x3

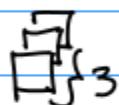
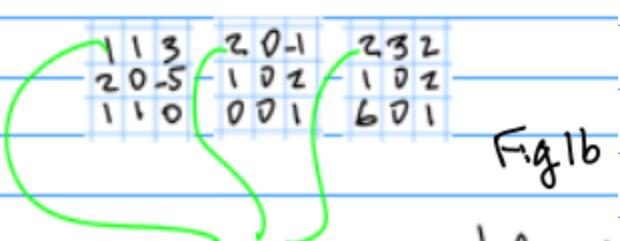


Fig.1a

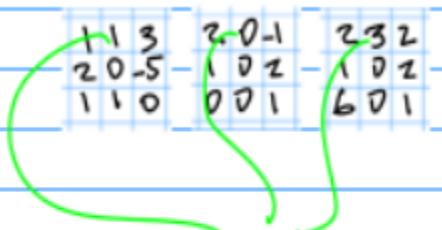
↓ Computation.

For Step 1.

pp.52



Step 2.



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

Fig.1c

5/3	14/3

Then. Carry out the Rest of the
Convolution.

Consider Probability Distribution
Map.

Nov. 2nd (Th).

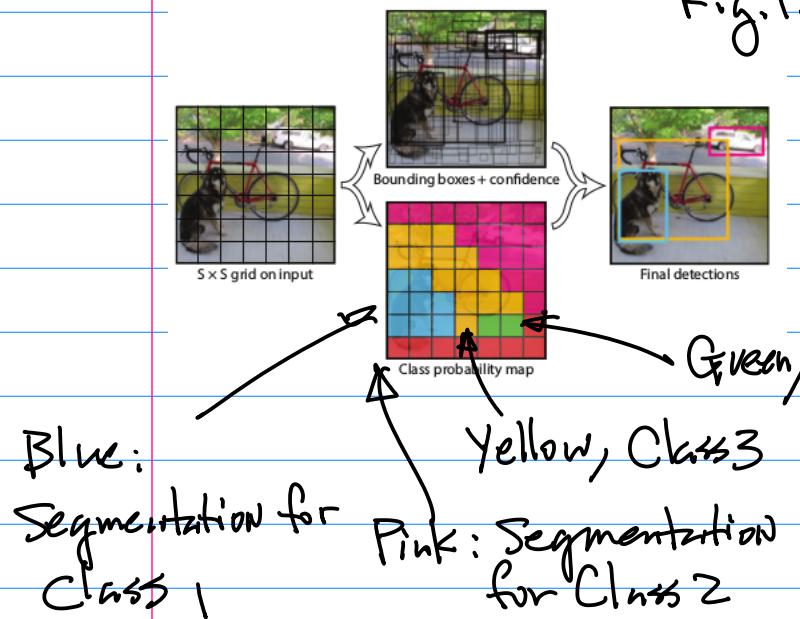
Consider the technique for finding
Probability Distribution map.

Example: From Fig.2 of the
Research paper. Conditions for
the discussion:

1. We have a segmented image.

$I(x, y)$. Example in the fig. below.

$$\begin{aligned} \text{Total No. of pixels of } I(x, y) &= 7 \times 7 \times 5 \times 5 \\ &= 49 \times 25 \end{aligned}$$



Hence,

$$\text{Prob(Class 1)}_{\text{Blue}} = \frac{8 \times 1}{49 \times 25} = \frac{8}{49}$$

Next,

$$\text{Prob(Class 2)}_{\text{pink}} = \frac{25}{49}$$

$$\text{Prob(Class 3)}_{\text{yellow}} = \frac{14}{49}$$

$$\text{Prob(Class 4)}_{\text{green}} = \frac{2}{49}$$

Note: To verify the calculation, we should have the following

Property:

$$\sum_{i=1}^4 \text{Prob}(C_i) = 1 \quad \dots (2).$$

from Our Calculation,

$$\sum_{i=1}^4 \text{Prob}(C_i) = \text{Prob}(C_1) +$$

$$\text{Prob}(C_2) + \dots + \text{Prob}(C_4)$$

$$= (8 + 25 + 14 + 2) / 49 = 1.$$

from the given Condition in Fig. 1

Since,

$$\left(\frac{\text{pixels belong to}}{\text{Class 1}} \right) = 8 \times (\underbrace{5 \times 5}_{\text{pixels ONE Grid}}) = 8 \times 25$$

The Next Step for the Probability Distribution Map Calculation: Segmentation, e.g. A Technique to allow us to classify

Objects into one of the class n.

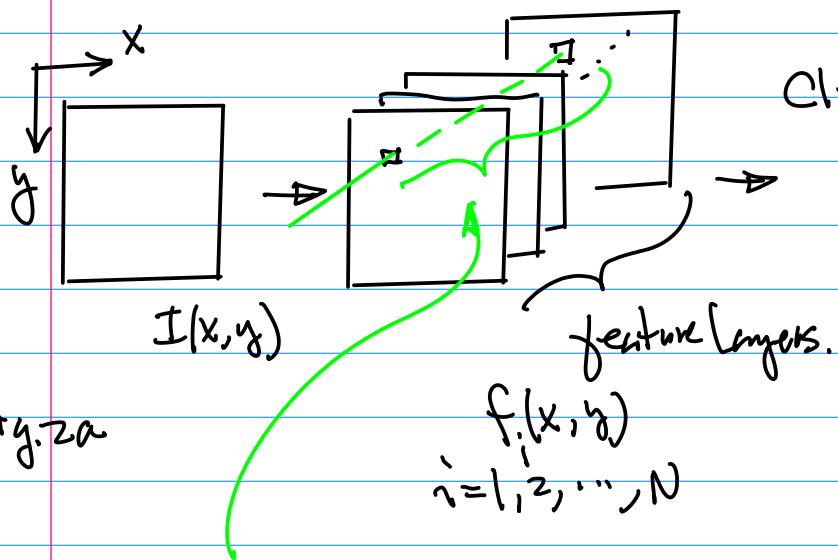


Fig.2a

Form Feature Vector(s)

2022F-101-cmpe258-note-2022-11-1 (4).pdf

$$\vec{X} = (x_1, x_2, \dots, x_N) \dots (3).$$

introduce Subscript.

$$\vec{x}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \dots (3-b)$$

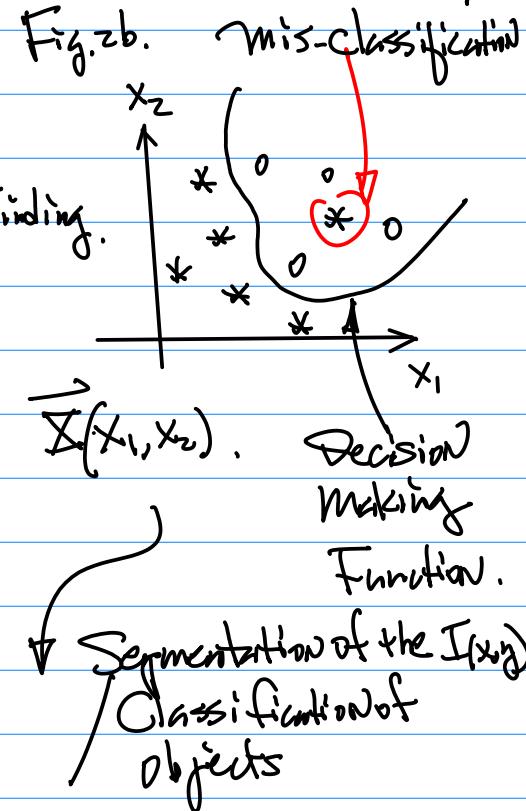
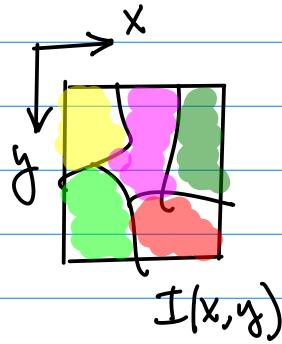


Fig.2c



Ref: k-mean Cluster Classification.

1. 2022F-101-cmpe258-note-2022-11-1 (4).pdf

2. github, 2022S, Search Keywords

"kmean", Two Relevant References

{PPT. — Formulation.

Hand Calculation.