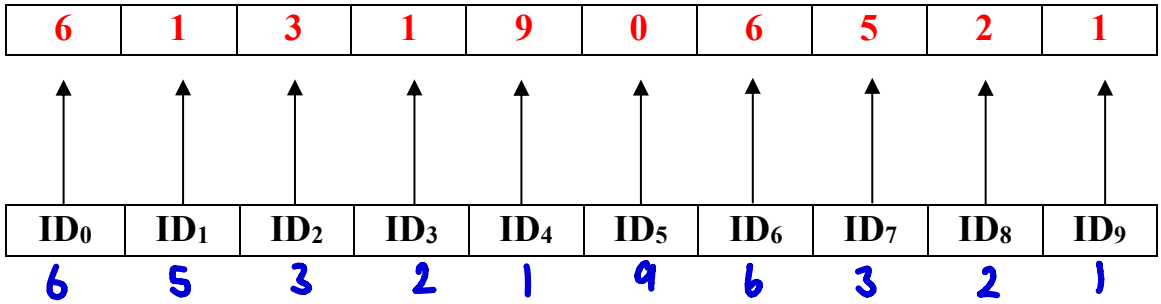
Final Exam Sample

1.

Preliminaries:

Your ID sequence will be used to create an image for Problem 1. ID_0 – ID_9 correspond each of your id as shown below, for example,



Then, use the quantization table below to quantize your ID into an intensity value in a range [0, 3]. The quantized IDs denote **qID.**

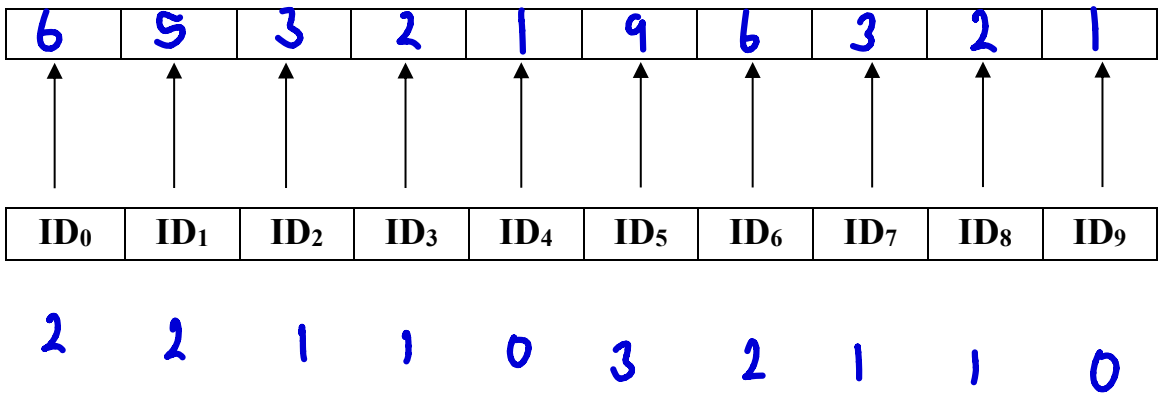
Original ID (ID_n)	**Quantized ID (qID_n)**
0-1	0
2-3	1
4-6	2
7-9	3

A 4x10 image created from your quantized ID has 4 different intensity levels in a range of [0,3] below.

qID_0	qID_1	qID_2	qID_3	qID_4	qID_5	qID_6	qID_7	qID_8	qID_9
qID_0	qID_1	qID_2	qID_3	qID_4	qID_5	qID_6	qID_7	qID_8	qID_9
qID_9	qID_8	qID_7	qID_6	qID_5	qID_4	qID_3	qID_2	qID_1	qID_0
qID_9	qID_8	qID_7	qID_6	qID_5	qID_4	qID_3	qID_2	qID_1	qID_0

(Write in the answer sheet provided)

Fill in your ID here



Fill in your 4x10 quantized image here

2	2	1	1	0	3	2	1	1	0
2	2	1	1	0	3	2	1	1	0
0	1	1	2	3	0	1	1	2	2
0	1	1	2	3	0	1	1	2	2

Otsu's threshold is a non-parametric and unsupervised method for automated threshold selection. An image has L total levels in a range $[0, \dots\ L-1]$. To derive Otsu's threshold, the gray-level histogram is normalized and regarded as a probability distribution:

L = intensity level
here 4 so [0,3]

$$p_i = n_i/N$$

Where n_i is the number of pixels at level i and N is the total number of pixels. A pixel in the image can be classified into two classes, C_0 and C_1 (background and objects) by a threshold level k; C_0 denotes pixels with level $[0, \dots\ k]$ and C_1 denotes pixels with levels $[k+1, \dots, L-1]$. The probability of class occurrence of C_0 and C_1 denotes $\omega_0\ (k)$ and $\omega_1\ (k)$, respectively, and are given by

$$\omega_0\ (k) = \sum_{i=0}^{k} p_i$$

$$\omega_1\ (k) = \sum_{i=k+1}^{L-1} p_i = 1 - \omega_0\ (k)$$

Total mean μ_T is given by

$$\mu_T\ = \sum_{i=0}^{L-1} ip_i$$

The first-order cumulative moment $\mu(k)$ of the histogram up to level k is given by

$$\mu(k) = \sum_{i=0}^{k} ip_i$$

The optimal (Otsu's) threshold is selected from the value that can maximize the separability of the resultant classes or the between-class variance in the formula below

$$\text{Between-class variance } \sigma_B^2(k) = \frac{(\mu_T\omega_0\ (k) - \mu(k))^2}{\omega_0\ (k)\omega_1\ (k)}$$

Answer questions 1.1.1-1.1.4 in the answer sheet provided.

1.1.1 Determine total mean (μ_T)

Total mean (μ_T) = (0x0.2) + (1x0.4) + (2x0.3) + (3x0.1) = 1.3

0 + 0.4 + 0.6 + 0.3

1.1.2 Determine the values in the table

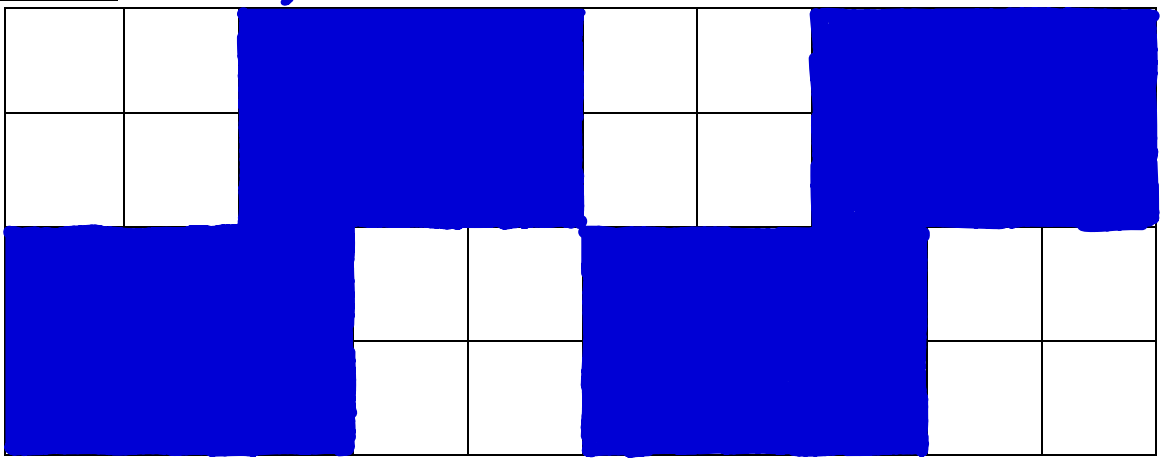
Intensity Level i	**0**	**1**	**2**	**3**
n_i	8	16	12	4
p_i	0.2	0.4	0.3	0.1

For each selected threshold level k, calculate the between-class variance $\sigma_B^2(k)$ and other parameters below.

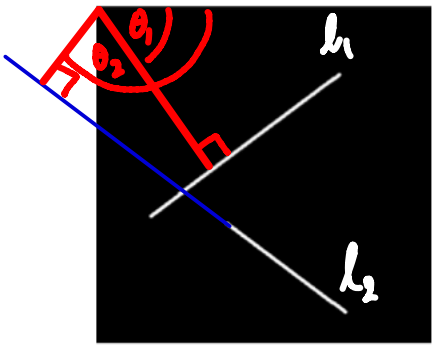
Threshold level k	**0**	**1**	**2**	**3**
$\omega_0(k)=\sum_{i=0}^{k} p_i$	0.2	0.6	0.9	1
$\mu(k) = \sum_{i=0}^{k} ip_i$	0	0.4	1	1.3 = μ_T
$\omega_1(k) = 1 - \omega_0(k)$	0.8	0.4	0.1	0
Between-Class Variance $\sigma_B^2(k)$	0.4225	0.6017	0.3211	Undefined

1.1.3 The selected threshold k or Otsu's threshold = 1

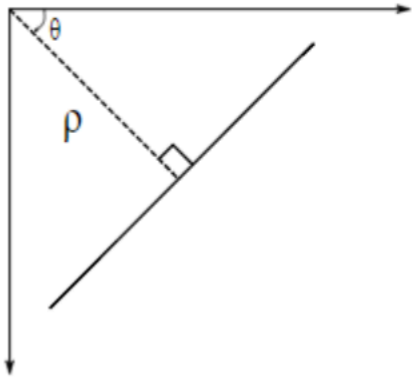
1.1.4 Fill in values of the thresholded image after apply inverse thresholding using this Otsu's threshold by **highlighting black color if the value = 0, and remain white box for value = 1.**



2. Considering the 100 x 100 image (left) with two main lines, draw an output of applying Hough Line transform to the image. The y-axis of the output represents the perpendicular distance (r) from the line to the origin (0,0), and the x-axis represents the angle (q) that this perpendicular makes with the horizontal axis as seen q-r relationship in the right image.

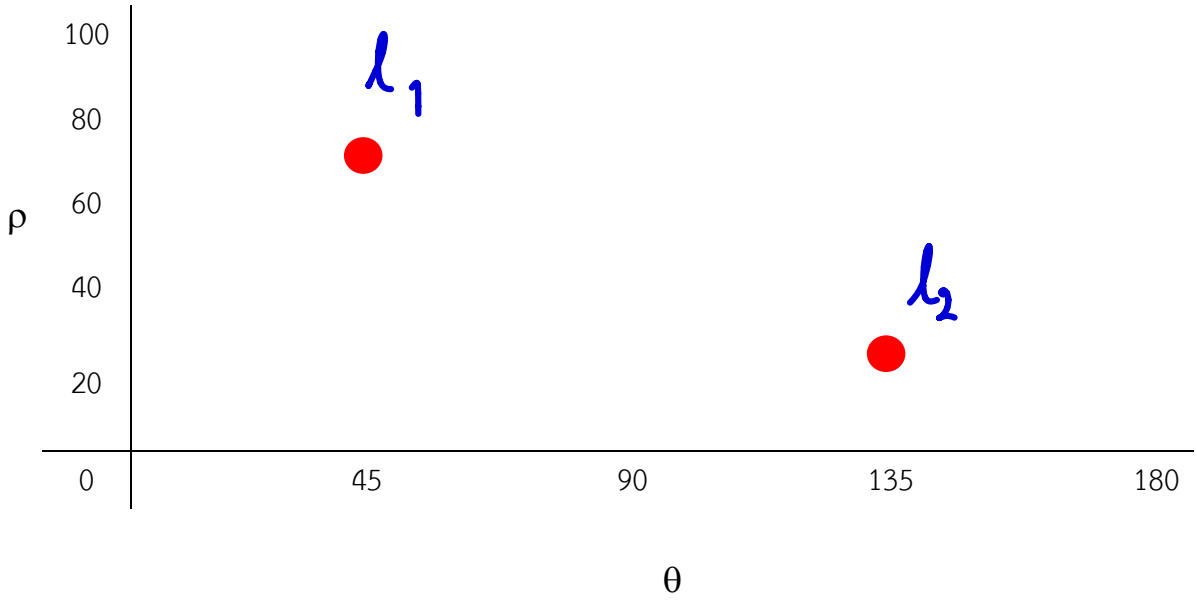


Original Image



θ-ρ relationship

Answer here (in red):



Matching and Multiple Choice
Select the correct answer for each question and provide it on the **answer sheet ONLY**.

Use these options and select only one best answer A.-K. for questions 3–5.

A. Erosion
B. Dilation
C. Opening
D. Closing
E. Hit-or-Miss transformation
F. Boundary extraction
G. Region filling
H. Extraction of connected components
I. Convex Hull
J. Thinning
K. Pruning

3. Find the locations (those pixels) whose neighbourhood matches the shape of a targeted structuring element
Ans: E Hit-or-Miss Transformation

Hit or Miss

4. To clean up some noise in the image below, but maintaining the same size of the image.

Erosion

Dilation

White = 1
Black = 0

Erosion → Dilation : Opening

Ans: C Opening (Erode -> dilate)

5. Which morphological technique should be used to merge between two separated objects below?
Ans: B. Dilation

fill small hole : Dilation

6. A method to find object structure in only single-pixel wide lines.
Ans. J.Thinning

Thinning

7. Highlighting the image output resulted from A ⊛ B where ⊛ is Hit-or-Miss transformation. You can ignore the boundary (partial matching).

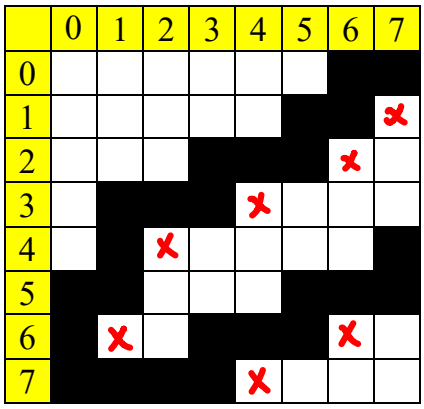
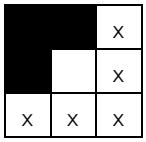


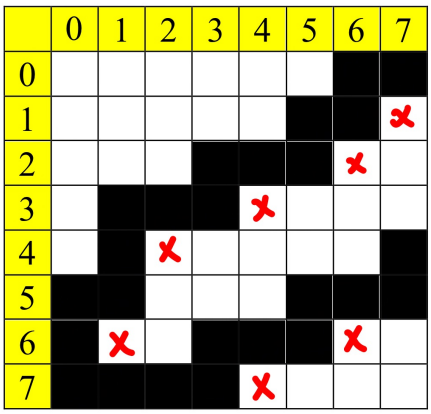
Image A



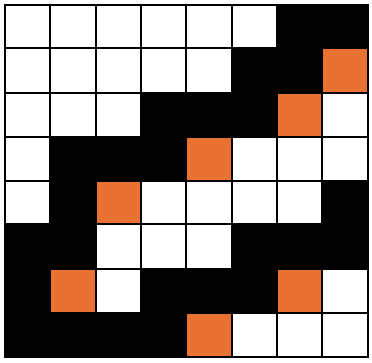
Structural Element B

■ =1 □ =0
x = don't care

Highlight your answer here:



Ans. Only orange color



8. Convolution Neural Networks

8.1) What is the purpose of Convolution? (2 points)

use learnable filter (mask) to extract propriate features from image.

Ans. Extract features from the image using learnable filters.

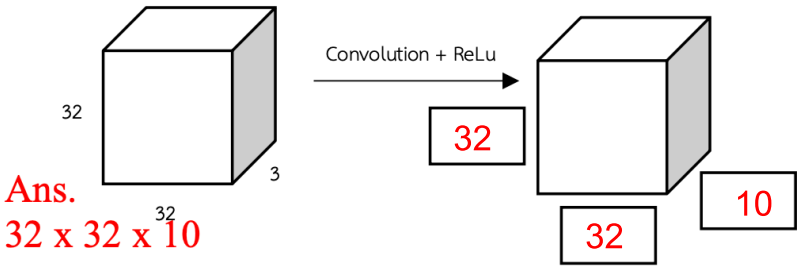
reduce spatial size by choosing max in mask

8.2) What is max-pooling? (2 points)

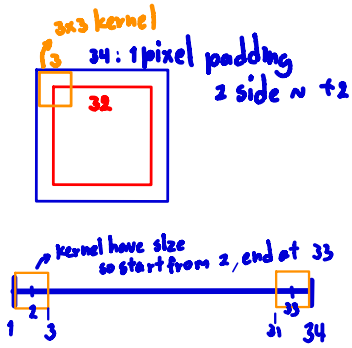
Ans. The process to reduce dimension of the feature size by selecting the max value from each kernel.

8.3) Fill the size of the output in the blank box below when convolve the image (32 x 32 x 3) by 10 filters (3 x 3 x 3) with stride = 1, with 1 pixel padding (3 points)

mean 10 output channels (depth)



Ans.
32 x 32 x 10



8.4 From 8.3) How many parameters (weights and bias) for convolutional layer?

ANS : 32x32x10

(3 x 3 x 3 + 1) x 10 = 280 params

Every filter have sub filter

InCh = 3, mask = 3x3, OutCh = 10

for every input channel : OutCh x InCh x mask

((3x3x3)+1)x10 = 280 params

Every filter have 1 bias : + OutCh

Formula : ((InCh x mask)+1) OutCh

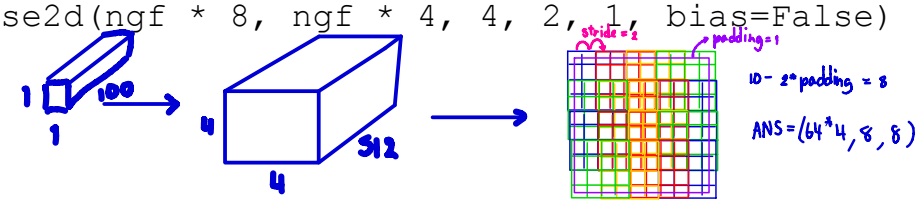
9. Generative Image

9.1 Describe the diffusion process , Pure Noise → denoise → Image

9.2 Consider the following layer from the DCGAN generator:

in out kernel stride padding

```
nn.ConvTranspose2d(nz, ngf * 8, 4, 1, 0, bias=False)
nn.ConvTranspose2d(ngf * 8, ngf * 4, 4, 2, 1, bias=False)
```



Given that
-nz = 100
-ngf = 64
-input tensor shape = (batch_size, 100, 1, 1)

What is the output size (excluding batch dimension) after the second layer?
Ans : (256, 8, 8)

10. Design a classification model for recognizing 20 kinds of Thai fruits, for example, banana, grape, watermelon, coconut, etc. Describe the process and write a flow diagram from collecting datasets, pre-processing, modelling, parameters, and evaluation method used in the application (10 points).

11. Classification

11.1 Histogram of Oriented Gradients is extracted from a 100 x 100 RGB image for a classification problem with a cell size of 5 x 5 pixels, each block has 2 x 2 cells, each cell has 5 orientation bins. Calculate the size of this HoG feature descriptor and also demonstrate how do you get this size?

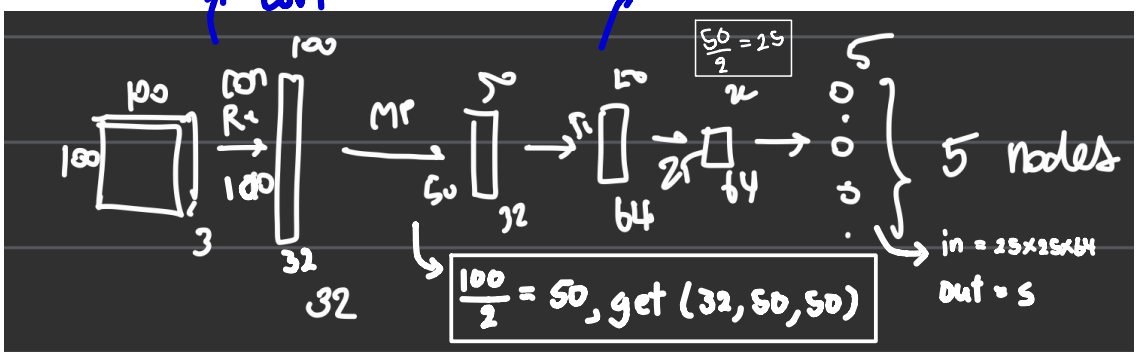
Ans 19 x 19 x 2 x 2 x 5 = 7220

11.2) Design a multilayer perceptron that accepts an HoG feature descriptor as input and outputs the classification for a problem with 5 classes.

Ans. 7220 input nodes and 5 output nodes, hidden layer is free style

11.3) A basic convolutional network takes an input image from 11.1), comprising two convolutional layers, each convolved with 3 x 3 filters and 1-pixel padding to maintain the input size, followed by ReLU activation and 2 x 2 max-pooling. The number of filters in the two layers is 32 and 64, respectively. Include one final fully connected layer. Draw this CNN architecture, showing the input image size (W x H x C) and the output.

Ans.



11.4) How many learnable parameters required in the final layer of 5.3).

Ans. (25 x 25 x 64 + 1) x 5 = 200,005 parameters

Use these options and select only one best answer A.-F. for question 12-14

A. ResNet-34
B. ResNet-50
C. EfficientNet

D. GoogleNet
E. Vision Transformer
F. AlexNet

12. Which model uses compound scaling to achieve state-of-the-art performance by balancing resolution, depth, and width, but may require careful tuning for new datasets?
Ans. C. EfficientNet

13. Which model uses an Inception module for computational efficiency, includes auxiliary outputs to improve training convergence?
Ans. GoogleNet

14. Which model is based on self-attention mechanisms, offering high accuracy on large-scale datasets but often requires extensive pretraining and large datasets to perform optimally?
Ans E. Vision transformer

Use these options and select only one best answer A.-D. for question 15-16

A. Binary Thresholding
B. Inverse Binary Thresholding
C. Otsu's Threshold
D. Adaptive Threshold

15. Given a grayscale image of scanned text documents with dark text on a light background, describe the segmentation technique used to create a binary image where the background is black and the text remains white, using a manually specified threshold.

Ans B. Inverse Binary Thresholding

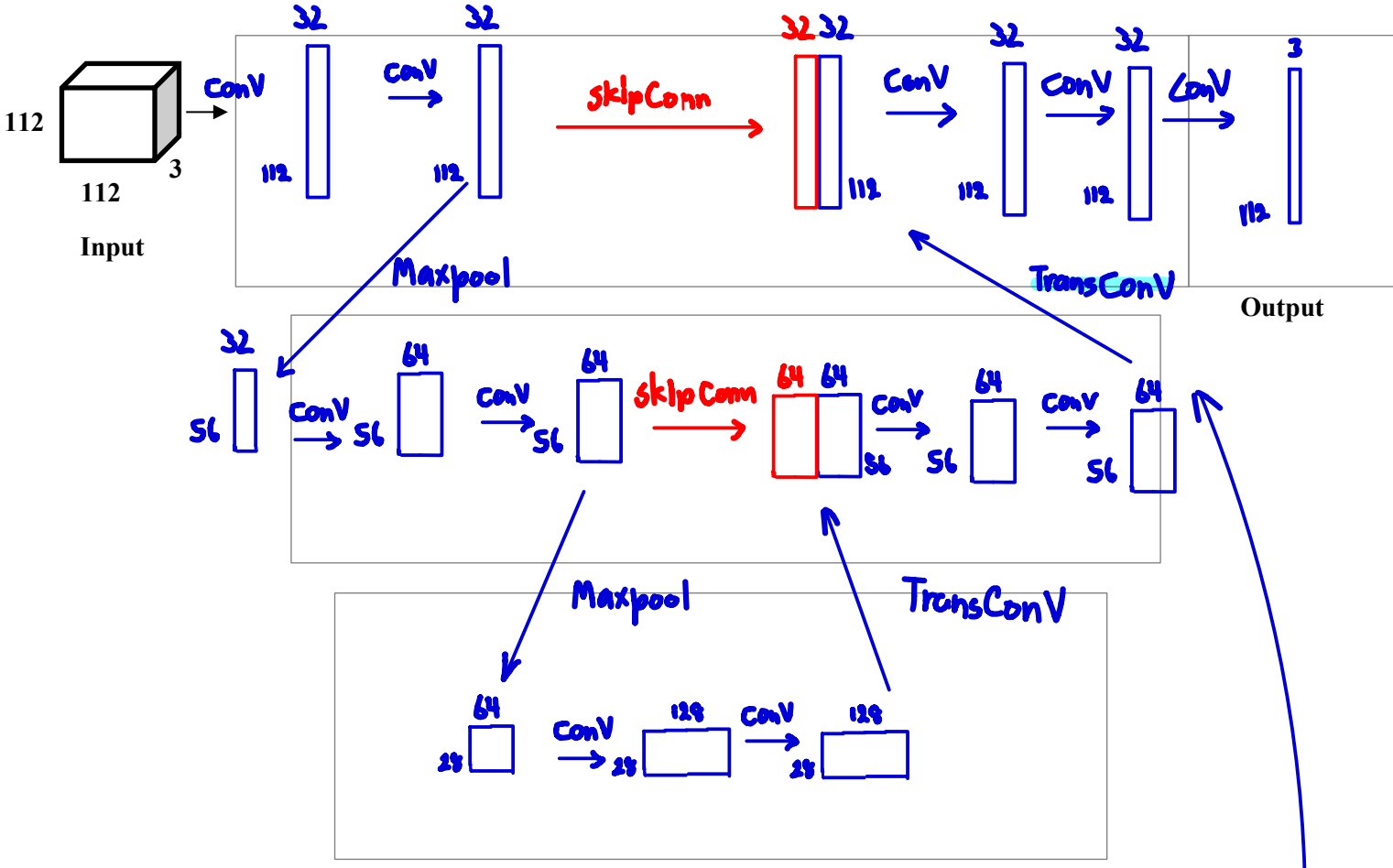
16. In a grayscale image with uneven lighting, explain the segmentation technique that dynamically determines the threshold value for each region to separate foreground from the background effectively.

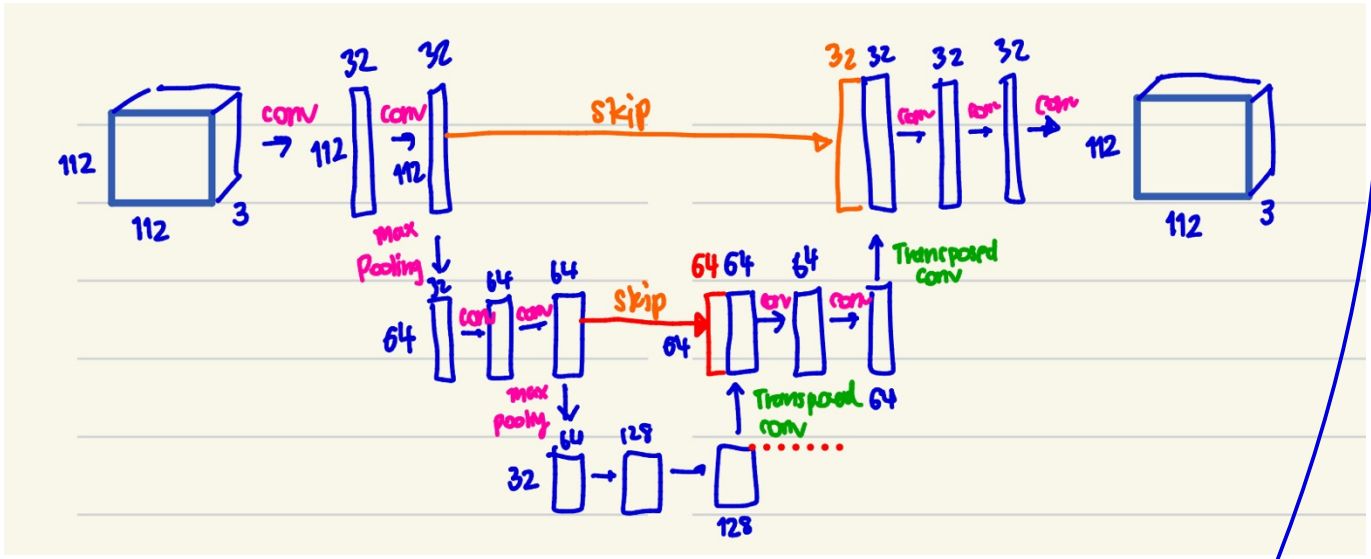
Ans C. ~~Otsu's Threshold~~

17. U-Net

17.1) Draw a U-Net architecture for three-class image segmentation that takes a 112 x 112 input image. The network should have an encoder-decoder structure with three levels as seen in a rough structure below. Each level in the encoder and decoder contains two convolutional layers with 3 x 3 kernels and 1-pixel padding to maintain the input size, followed by 2 x 2 max-pooling in the

encoder. The number of filters for each level is 32, 64, and 128, respectively for both encoder and decoder. For the decoder, use a 2 x 2 kernel for the up-sampling process. Clearly label the feature input size and the feature output at each stage (W x H x C) of the architecture including the final output. (7 points)





17.2) Calculate the total number of parameters (including bias) in the last deconvolution layer of the decoder (only the deconvolution process), given its kernel size, input channels, and output channels in 5.1). (2 points) kernel = 2x2, input_ch = 64, output_ch = 32

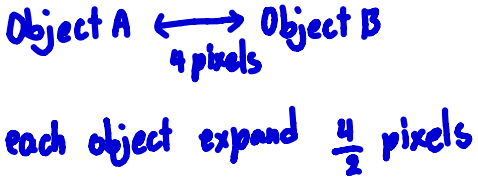
params = (input x kernel) x output + bias ← same as out

= ((input x kernel) + 1) output = ((64x4)+1)32 = 8,224 params

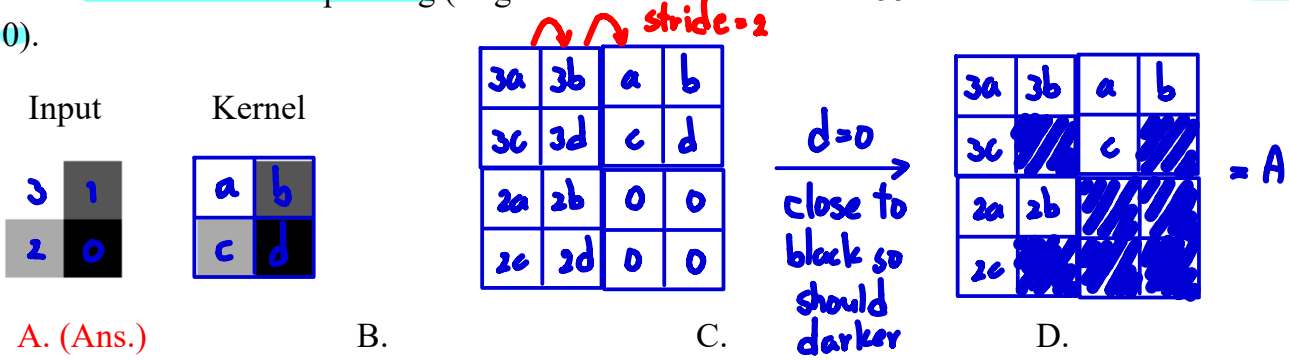
(64 x (2 x 2) + 1) x 32 = 8,224

18. Design a structural element that can merge between two binary objects that are at most separated by 4 pixels using dilation process.

Ans.
[[1, 1, 1, 1, 1],
[1, 1, 1, 1, 1],
[1, 1, 1, 1, 1],
[1, 1, 1, 1, 1],
[1, 1, 1, 1, 1]]



19. What is the output of transposed convolution between a 2x2 input and a 2x2 kernel below with stride of 2 and no padding (bright color value is close to 255 and dark color value is close to 0).



A. (Ans.) B. C. D.

20. In Canny edge detection, we will get more discontinuous edges if we make the following change to the hysteresis thresholding:

A. Increase the high threshold
B. Decrease the high threshold
C. Increase the low threshold
D. Decrease the low threshold
Ans. C

21. In region growing, what criteria are commonly used to determine whether a neighboring pixel should be added to a growing region?

A. The histogram of the entire image.
B. A similarity measure such as intensity difference or texture.
C. The absolute position of the pixel.
D. The dimensionality of the image.
Ans. B

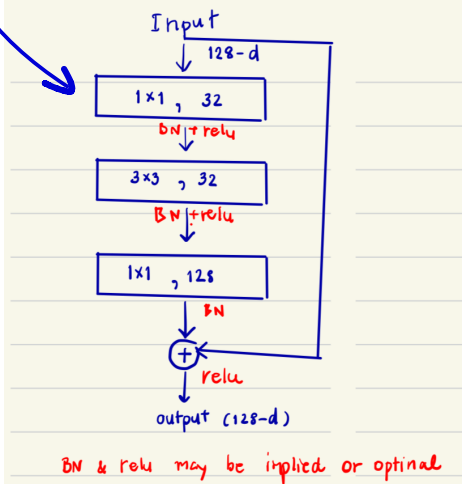
22. Design a **ResNet-style bottleneck block** with the following specifications:

- Input depth: 128
- Output depth: 128
- Inner (bottleneck) depth: 32
- Convolutions: 1×1 → 3×3 → 1×1
- Include a skip (residual) connection

Tasks:

Draw the block diagram showing the depth at each stage and the skip connection.

(Optional) You may include Batch Normalization (BN) and ReLU layers in the diagram, or leave them implied.



23. Select one application discussed by the invited speaker and describe it in at least 40 words. (Bonus 5 points) Nope, our guest talk about Advice In AI era.