**National University of Computer and Emerging Sciences**

**FAST School of Computing** **Fall-2023** **Islamabad Campus**

# CS-4055: Digital Image Processing

Serial No:

## Final Exam

**Total Time: 3 Hours**

**Total Marks: 100**

Tuesday, 19th December, 2023

## Course Instructors

Akhtar Jamil

_____

**Signature of Invigilator**

| **Student Name** | **Roll No.** | **Course Section** | **Student Signature** |
|---|---|---|---|

**DO NOT OPEN THE QUESTION BOOK OR START UNTIL INSTRUCTED.**

**Instructions:**

1. Attempt on question paper. Read the question carefully, understand the question, and then attempt it.
2. No additional sheet will be provided for rough work.
3. Verify that you have **three? (3?)** different printed pages including this title page. There are **six (6)** questions.
4. Calculator sharing is strictly prohibited.
5. Use permanent ink pens only. Any part done using soft pencil will not be marked and cannot be claimed for rechecking.
6. Ensure that you do not have any electronic gadget (like mobile phone, smart watch, etc.) with you.

| | **Q-1** | **Q-2** | **Q-3** | **Q-4** | **Q-5** | **Q-6** | **Total** |
|---|---|---|---|---|---|---|---|
| **Marks Obtained** | | | | | | | |
| **Total Marks** | **50** | **10** | **10** | **10** | **10** | **10** | **100** |

## Question 1. MCQ Answer Sheet

**Cross (X)** the correct answer. Overwriting an MCQ will result in ZERO marks.

**[1 x 50 = 50]**

| S. No | A | B | C | D | S. No | A | B | C | D |
|---|---|---|---|---|---|---|---|---|---|
| 1. | | | | | 26. | | | | |
| 2. | | | | | 27. | | | | |
| 3. | | | | | 28. | | | | |
| 4. | | | | | 29. | | | | |
| 5. | | | | | 30. | | | | |
| 6. | | | | | 31. | | | | |
| 7. | | | | | 32. | | | | |
| 8. | | | | | 33. | | | | |
| 9. | | | | | 34. | | | | |
| 10. | | | | | 35. | | | | |
| 11. | | | | | 36. | | | | |
| 12. | | | | | 37. | | | | |
| 13. | | | | | 38. | | | | |
| 14. | | | | | 39. | | | | |
| 15. | | | | | 40. | | | | |
| 16. | | | | | 41. | | | | |
| 17. | | | | | 42. | | | | |
| 18. | | | | | 43. | | | | |
| 19. | | | | | 44. | | | | |
| 20. | | | | | 45. | | | | |
| 21. | | | | | 46. | | | | |
| 22. | | | | | 47. | | | | |
| 23. | | | | | 48. | | | | |
| 24. | | | | | 49. | | | | |
| 25. | | | | | 50. | | | | |

**Question 1. MCQs based questions have only one possible answer. Only answers marked on the answer sheet will be considered.**

1. Assume an 8-bit grayscale image of size 5 x 5 with all pixels having same value equal to 255. What will be the possible relative data redundancy in this image?
   a) 100%
   **b) 96%**
   c) 90%
   d) 92%
2. Discrete histogram equalization may not produce a perfectly flat histograms due to discrete nature of the images.
   **a) True**
   b) False

3. In multi-class classification with SVMs, what approach is commonly used to extend SVMs to handle multiple classes?
   a) One-vs-One (OvO)
   b) One-vs-All (OvA)
   c) Multi-class kernels
   **d) Both a) and b)**

4. If the compression ratio of an image is 4, the relative redundancy can be approximately calculated as:
   a) 0.70
   **b) 0.75**
   c) 0.80
   d) 0.85

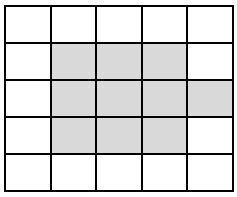
5. In HOG, if the gradient direction and magnitude at particular location in a cell are 165 and 136, respectively, then the correct value assigned to the two bins for oriented histogram with 9 bins will be:
   **a) 34 for $0^{th}$ bin and 102 for $160^{th}$ bin.**
   b) 102 for $0^{th}$ bin and 34 for $160^{th}$ bin.
   c) 136 for $165^{th}$ bin
   d) 136 for $180^{th}$ bin

6. Which operation is more suitable for smoothing the boundary of the following object?
   a) Hole filling
   b) Erosion
   **c) Dilation**
   d) None of the above

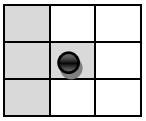
7. Which of the following statement is NOT correct about morphological processing?
   a) Dilation makes objects bigger
   b) Erosion makes objects smaller
   c) Dilation and Erosion are dual of each other
   **d) Dilation followed by Erosion and Erosion followed by Dilation will always produce same result.**

For the following two questions assume gray cells have value 1 and white have value 0. Moreover, the center pixel for kernel is shown with a filled circle.

8. Which of the following output will be obtained by applying erosion on the following input image using the structuring element SE shown?

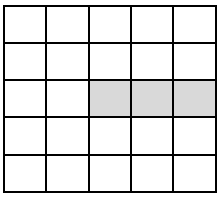


Input Image

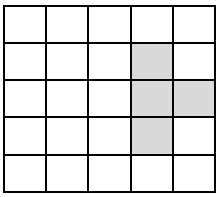


SE

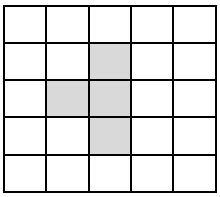
**(a)**



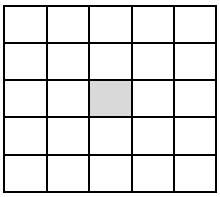
(b)



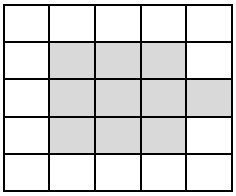
(c)



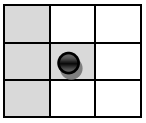
(d)



9. Which of the following output will be obtained by applying dilation on the following input image using the structuring element (SE)?

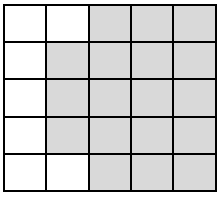


Input Image

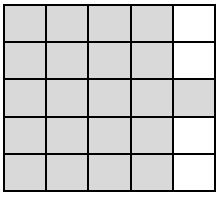


SE

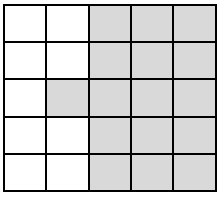
**(a)**



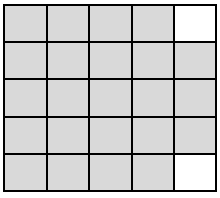
(b)



(c)



(d)



10. Which morphological operator is more suitable for removing gaps within characters shown below without enlarging them?

**a) Opening**
b) Closing
c) Erosion
d) Dilation

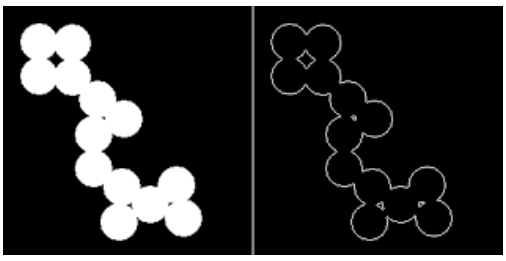
11. Information loss for image compression can be measure in terms of
a) Image derivatives
b) Difference of Gaussian
**c) Signal to noise ratio**
d) Gaussian Pyramids and their differences

12. What is the "margin" in an SVM classifier?
    a) The distance between the data points and the decision boundary
    **b) The width of the decision boundary**
    c) The number of support vectors
    d) The complexity of the kernel function

13. Batch Normalization in HOG will help make the histogram:
    a) Normalize values between 0 - 255
    b) Normalize values between 0.0 - 1.0
    **c) Illumination Independent**
    d) None of the above

14. The suitable morphological operation on the input image (A) that will produce the output shown in (b) with a structing element (B) is:

    a) correct $A - (A \oplus B)$
    b) $A + (A \oplus B)$
    c) $A \oplus B$
    d) $A \ominus B$



(a) Input (b) Output

15. The total number of comparision with neighborhood to find the key points in SIFT are:
    a) 4 comparisions
    b) 8 comparisions
    c) 24 comparisions
    **d) 26 comparisions**

16. Suppose that for an image of size 64 x 64, code-1 has average code length is 8-bits/pixel while for code-2 the average code length is 2.50 bits/pixel. The compression ratio obtained is:
    **a) 3.2**
    b) 4.2
    c) 5.2
    d) 6.2

17. In histogram equalization, strictly monotonically increasing function guarantees the inverse mapping will be:
    **a) Singled valued**
    b) Multi-valued
    c) Average valued
    d) Sum of all values.

18. Dimensionality reduction is an example of
    a) Supervised Machine Learning
    **b) Unsupervised Machine Learning**
    c) Reinforcement Machine Learning
    d) None of the Above

19. For HOG we calculate a histogram with 9 bins corresponding to following angles. Also, Magnitude and Direction for an image of size 2x2 are also given.

| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |
|---|---|---|---|---|---|---|---|---|

Histogram

| 50 | 40 |
|---|---|
| 40 | 80 |

Direction

| 10 | 6 |
|---|---|
| 4 | 4 |

Magnitude

The correct histogram of gradients corresponding to the above details is:

**a)**

| 0 | 0 | 15 | 5 | 4 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|
| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |

b)

| 4 | 0 | 10 | 5 | 4 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|
| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |

c)

| 0 | 0 | 6 | 10 | 4 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|
| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |

d)

| 0 | 10 | 6 | 4 | 4 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|
| 0 | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 |

20. What is the sum of all elements of a normalized histogram?
    a) -1
    b) 0
    **c) 1**
    d) None of the above

21. Which algorithm(s) will help count objects within an image?
    a) Connected component labeling
    b) Contours method
    c) Skeleton
    d) **Both a) and b)**

22. The average length for the code shown is:

| Letter | Probability | Codeword |
|---|---|---|
| $a_2$ | 0.4 | 1 |
| $a_1$ | 0.2 | 01 |
| $a_3$ | 0.2 | 00 |
| $a_4$ | 0.1 | 010 |
| $a_5$ | 0.1 | 011 |

    a) 2.8 bits/pixel
    b) 2.2 bits/pixel
    c) 3.0 bits/pixel
    **d) 1.8 bits/pixel**

23. Applying the following operation on image A with structuring element B (3 x 3 with all elements set to 1) will result in: **Output =** $A - (A \ominus B)$
    a) The edges of objects with holes only
    **b) Boundary of the objects**
    c) Zero padding of image
    d) Remove small holes

24. Given an input image of size 32 x 32 x 3. Assume two kernels of size 7 x 7 x 3 are applied with stride=1. What will be the size of the output feature map?
    a) 26×26×1
    **b) 26×26×2**
    c) 28×28×1
    d) 28×28×2

25. In deep learning, CNN is used for:
    a) Extracting low level features only
    b) For extracting high level features only.
    **c) Can extract both high level and low level features.**
    d) For classification only.

26. Which of the following is not a supervised algorithm?
    **a) PCA**
    b) Linear Regression
    c) Logistic Regression
    d) Support Vector Machines (SVMs)

27. Assume you have a highly imbalanced data. Which classifier you prefer to use:
    a) ANN
    **b) SVM**
    c) CNN
    d) None of the above.

28. How does the HOG descriptor handle changes in illumination?
    a) By converting the image to grayscale
    b) By ignoring the color information in the image
    **c) By normalizing the histograms in each cell**
    d) By applying a threshold to the image intensity

29. What is the significance of the 'orientation' in HOG features?
    a) It refers to the orientation of the camera with respect to the object.
    **b) It indicates the dominant direction of the gradient for a particular patch of the image.**
    c) It determines the shape of the object in the image.
    d) It is used to identify the color of the object in the image.

30. What makes SIFT descriptors scale-invariant?
    a) The use of a fixed-size window for all features
    **b) The detection of features at multiple scales using a Gaussian pyramid**
    c) The application of a threshold to the size of features
    d) The use of color information to maintain consistency

31. What characteristic of SIFT makes it particularly robust for object recognition in images?
    **a) Scale-invariance**
    b) Hue adjustment
    c) Edge sharpening
    d) Brightness normalization

32. How is the entropy of an image related to its compressibility?
    a) Images with high entropy are more compressible.
    **b) Images with low entropy are more compressible.**
    c) Entropy and compressibility are unrelated.
    d) The relationship varies depending on image size.

33. In subjective fidelity criteria, the image quality is measure by a human expert;
    **a) True**
    b) False

Consider the following input image and corresponding reconstructed image after compression.

| | | |
|---|---|---|
| **3** | **5** | **2** |
| **3** | **6** | **1** |
| **4** | **6** | **1** |

*Input Image* $(f)$

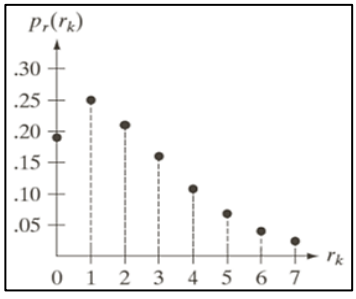
| | | |
|---|---|---|
| **3** | **4** | **1** |
| **2** | **4** | **1** |
| **4** | **6** | **1** |

*Reconstructed Image* $(\tilde{f})$

34. The root mean squared error between input image and reconstructed image after compression is (approximately):
    a) 3.0
    **b) 0.88**
    c) 2.65
    d) 0.25

35. If pixels in an image exhibit maximally correlation, then this shows the image pixels have high temporal redundancy.
    a) True
    **b) False**

36. Consider the following histogram of an image. What will happen if we apply the histogram equalization on this image directly?
    a) It will improve the contrast
    **b) It will improve the brightness too much and make image too bright**
    c) It will have not impact on it.
    d) It will make the histogram uniform



37. In an SVM classifier, what is a support vector?
    a) The data points that are farthest from the decision boundary
    **b) The data points that are closest to the decision boundary**
    c) The average of all data points
    d) Data points in the center of the feature space

38. What is the primary goal of a Support Vector Machine (SVM)?
    a) Regression
    b) Clustering
    **c) Classification**
    d) Dimensionality reduction

39. Rosenblatt's perceptron model was primarily designed for what type of machine learning task?
    a) Regression
    b) Clustering
    **c) Classification**
    d) Reinforcement learning

40. What happens when a Rosenblat's perceptron model encounters data that is not linearly separable?
    a) It can still perform classification accurately.
    **b) It fails to converge and cannot find a solution.**
    c) It requires a different activation function.
    d) It becomes computationally inefficient.

41. What is the most important feature of SIFT that is often used in applications such as image stitching and object recognition?.
    a) Its ability to work with low-resolution images
    **b) Its robustness to changes in viewpoint and lighting conditions**
    c) Its speed in real-time applications
    d) Its reliance on global image features

42. Which of the following is a potential drawback of SIFT in terms of computational efficiency?
    a) SIFT features are highly compressed, reducing computational load.
    b) SIFT requires a large number of key points to be effective.
    **c) SIFT involves extensive Gaussian blurring, which can be computationally expensive.**
    d) SIFT does not require any pre-processing steps.

43. SIFT descriptors are based on:
    a) Histograms of pixel intensities
    **b) Histograms of gradient orientations**
    c) Fourier transform coefficients
    d) Principal Component Analysis (PCA)

44. Which of the following is a key step in HOG feature extraction that helps improve invariance to changes in lighting?
    a) Gradient computation
    b) Non-maximum suppression
    c) Gamma correction
    **d) Histogram Block normalization**

45. What type of gradient information is typically used in HOG feature extraction?
    **a) Magnitude and orientation**
    b) Hue and saturation
    c) Edge detection responses
    d) Color histograms

46. Which of the following is a potential limitation of HOG features?
    a) Sensitivity to object rotation
    b) High computational complexity
    **c) Limited ability to handle occlusions**
    d) Strong reliance on color information

47. Lossless image compression techniques ensure that:
    **a) No information is lost during compression**
    b) Some information may be lost, but the image remains visually similar
    c) Image quality is significantly improved
    d) The image can be decompressed only with a password

48. In information theory, what does "entropy" represent?
    a) The amount of information in an image
    **b) The uncertainty or randomness in an image**
    c) The speed of data transmission
    d) The size of the image

49. What term is commonly used to describe the measure of how well an image compression method performs in preserving image quality?
    a) Entropy
    b) Compression ratio
    c) Bitrate
    **d) Signal-to-Noise Ratio (PSNR)**

50. Which of the following image regions is likely to have higher entropy?
    a) A uniform blue sky
    **b) A complex, detailed texture**
    c) A grayscale gradient
    d) A solid black background

**Question 2. Write short answers to the following questions. [2 x 10 = 20]**

**1) What is the difference between face detection and face landmark detection?**

Face detection is the process of identifying the presence and location of faces in a digital image or video stream, typically providing a bounding box around each face. On the other hand, face landmark detection is a more detailed analysis that identifies specific features or points on a face, such as the corners of the eyes, the tip of the nose, the edges of the mouth, and the contour of the jawline.

**2) What is the difference between stochastic and batch backpropagation approaches?**

Stochastic backpropagation updates the model's weights using the gradient of the error with respect to each training example individually. Batch backpropagation computes the gradient of the error using a batches of the dataset to make a single weight update. Both methods aim to minimize the loss function, but they differ in the frequency and granularity of the weight updates.

**3) Write two steps to stop model from further learning when the optimal parameters are obtained.**

1) Use early stopping
2) Reaching a required level of accuracy.

**4) SVM is a linear classifier, how can it be used for multi-class classification problem?**

One-vs-Rest (OvR) or One-vs-All (OvA): For K separate SVMs are trained. Each SVM distinguishes one class from all the remaining classes. For a given input, all SVMs predict a class membership, and the class associated with the SVM that gives the highest confidence (distance from the hyperplane) is chosen as the final class.

One-vs-One (OvO): In this approach, an SVM is trained for every pair of classes, resulting in K classes. For a new input, each SVM votes for one class, and the class with the most votes determines the input's label.

**5) Given the following weight matrix, calculate the margin:**

$$m = \frac{2}{||\mathbf{w}||}$$

$$\mathrm{w} = 1.3,\ 4.0,\ 2.5,\ 0.5$$

w is given by:

$$||w|| = \sqrt{w_1^2 + w_2^2 + w_3^2 + w_4^2}$$

$$margin = 0.4066.$$

**6) What is the difference between soft margin and hard margin classifiers?**

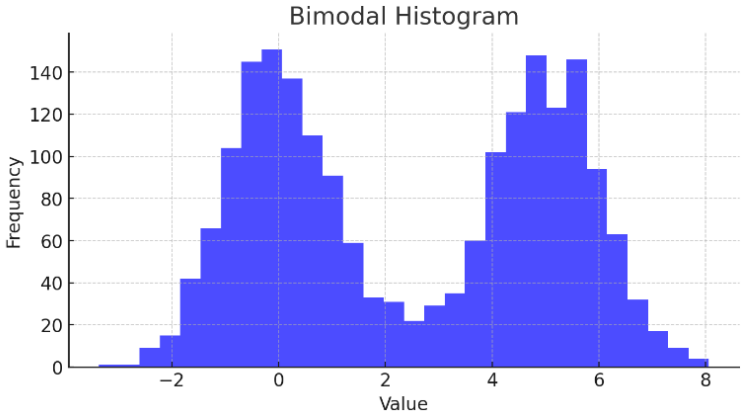
A hard margin classifier insists on perfectly separating the two classes using a hyperplane. This approach assumes that the data is linearly separable and does not allow any points to be within the margin boundaries.
A soft margin introduces a slack variable for each data point, permitting some points to fall within the margin and classifies if it is on right or wrong side of the hyperplane. The soft margin approach aims to find a balance between maximizing the margin and minimizing the classification error.

**7) What is the main difference between conventional machine learning and deep learning approaches?**

Conventional machine learning and deep learning differ primarily in their approach to feature extraction and model complexity. In conventional machine learning, feature extraction is performed manually while in deep learning, automates feature extraction through layers of neural networks.

**8) What is bimodal histogram? Draw a diagram to show a bimodal histogram.**

A bimodal histogram is a type of histogram that displays two distinct peaks or modes. These modes represent concentrations of values at different points in the data set. Such histograms are indicative of a distribution where data are concentrated around two central values, suggesting the presence of two different sub-populations within the overall dataset.




**9) Can histogram equalization be used on an image which has already a uniform histogram? Explain your reason.**

Histogram equalization would be ineffective and unnecessary to apply on an image with already uniform histogram. The purpose of histogram equalization is to redistribute the intensity values of an image so that they span the entire range more uniformly, enhancing the contrast of the image. When an image already has a uniform histogram, it implies that the intensities are already well distributed across the available range. Applying histogram equalization to such an image would not contribute any significant improvement in contrast or visibility.

**10) What is the difference between semantic segmentation and object detection?**

Semantic segmentation provides a detailed, pixel-wise classification of an image, object detection identifies and locates individual objects, typically with bounding boxes, without necessarily detailing the exact shape of the object at the pixel level.

**Question 3. [5+5]**

**a) Convolve the following image (left side) using the filter shown on right side. Assume stride of 2.**

| | | | | | |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 |

| | | |
|---|---|---|
| 1 | -1 | -1 |
| -1 | 1 | -1 |
| -1 | -1 | 1 |

| | |
|---|---|
| 3 | -3 |
| -3 | 0 |

**b) Consider the following histogram of an input image (left side) and the targe histogram (right side), perform histogram specification.**

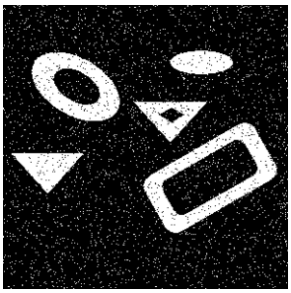
| Input Image | | | |
|---|---|---|---|
| $r_k$ | $n_k$ | $p(n_k)$ | CDF |
| 0 | 20 | 0.0667 | 0.0667 |
| 1 | 10 | 0.033 | 0.1 |
| 2 | 30 | 0.1 | 0.2 |
| 3 | 40 | 0.133 | 0.333 |
| 4 | 50 | 0.1667 | 0.5 |
| 5 | 60 | 0.2 | 0.7 |
| 6 | 70 | 0.233 | 0.933 |
| 7 | 20 | 0.0667 | 1.0 |

| Target image | | | |
|---|---|---|---|
| $z_k$ | $n_k$ | $p(n_k)$ | CDF |
| 0 | 50 | 0.227 | 0.227 |
| 1 | 60 | 0.2727 | 0.5 |
| 2 | 40 | 0.1818 | 0.682 |
| 3 | 30 | 0.1364 | 0.818 |
| 4 | 20 | 0.0909 | 0.909 |
| 5 | 10 | 0.0455 | 0.955 |
| 6 | 5 | 0.0227 | 0.977 |
| 7 | 5 | 0.0227 | 1.0 |

| Output Mapping | |
|---|---|
| $r_k$ | $s_k$ |
| 0 | 0 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 1 |
| 5 | 2 |
| 6 | 5 |
| 7 | 7 |

**Question 4. [5+5]**

**a) Write a program to count the number of objects with holes in the following image. Only consider objects which contains at least 100 pixels in it.**



```
import cv2

image_path = 'c:/aj/data/shapes.png'
image = cv2.imread(image_path, 0)

_, thresh = cv2.threshold(image, 128, 255, cv2.THRESH_BINARY)

# Find contours
contours, hierarchy = cv2.findContours(thresh, cv2.RETR_TREE, cv2.CHAIN_APPROX_SIMPLE)

def has_hole(hier, index):
    return hier[0, index, 2] != -1

# Initialize the number of objects with holes
objects_with_holes = 0

for i, contour in enumerate(contours):
    if has_hole(hierarchy, i) and cv2.contourArea(contour) > 100:
        objects_with_holes += 1
```

**b) Write code to determine the percentage of filled gray color in the following image.**

```
# Load the image
image_path = 'c:/aj/data/bar.png'
image = cv2.imread(image_path)
gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

non_white_pixels = np.sum(gray_image < 240)

# Calculate the total number of pixels in the image
```

```
total_pixels = image.shape[0] * image.shape[1]

# Calculate the percentage of filled gray color
percentage_filled = (non_white_pixels / total_pixels) * 100
```

**Question 5. [4+2+2+2]**
**Consider an image of size 10 x 10. The probability of its pixels are shown below:**

| Symbol (Pixel) | Probabilities |
|---|---|
| $a_1$ | 0.1 |
| $a_2$ | 0.4 |
| $a_3$ | 0.06 |
| $a_4$ | 0.1 |
| $a_5$ | 0.04 |
| $a_6$ | 0.3 |

a) Compress the image using Huffman coding. Show the final codes obtained properly for each symbol.

| Symbols (like intensity levels) | Probabilities (sorted) | Source Reduction (do till two values are left) (Maintain in sorted order here as well) | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| a2 | 0.4 1 | 0.4 | 0.4 | 0.4 | 0.6 0 |
| a6 | 0.3 00 | 0.3 | 0.3 | 0.3 00 | 0.4 1 |
| a1 | 0.1 011 | 0.1 | 0.2 010 | 0.3 01 | |
| a4 | 0.1 0100 | 0.1 0100 | 0.1 011 | | |
| a3 | 0.06 01010 | 0.1 0101 | | | |
| a5 | 0.04 01011 | | | | |

b) Calculate the average number of bits/pixel needed for Huffman coding.

$$L_{\text{avg}} = 0.4 \times 1 + 0.3 \times 2 + 0.1 \times 3 + 0.1 \times 4 + 0.06 \times 5 + 0.04 \times 5 = 2.2 \text{ bits/symbol}$$

c) Compute the entropy of the image.

$$H(X) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

$$Entropy = -(0.1 \times \log_2(0.1) + 0.4 \times \log_2(0.4) + 0.06 \times \log_2(0.06) + 0.1 \times \log_2(0.1) + 0.04 \times \log_2(0.04) + 0.3 \times \log_2(0.3))$$

Entropy = 2.144

d) Compute the redundancy and compression ratio for Huffman coding for above image.

$$C = \frac{10 \; x \; 10 \; x \; 8}{10 \; x \; 10 \; x \; 2.14} = 3.74$$

$$R = 1 - \frac{1}{C} = 0.7252 = 73.26\%$$

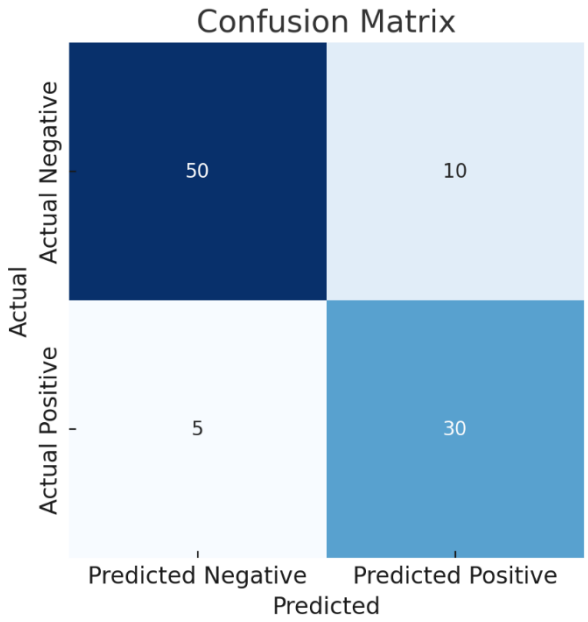
**Question 6. [3+2+5]**

**a) Assume a confusion matrix with two binary classes shown below. Calculate precision, recall and f-measure.**

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$




- $TP = 30$
- $TN = 50$
- $FP = 10$
- $FN = 5$

$$\text{Precision} = \frac{30}{30 + 10} = 0.75 \; or \; 75\%$$

$$\text{Recall} = \frac{30}{30 + 5} = 0.857 \; or \; 85.7\%$$

$$\text{F1 Score} = 2 \times \frac{0.75 \times 0.857}{0.75 + 0.857} = 0.799 \; or \; 79.9\% \approx 80\%$$

**b) How the selective search algorithm works? Why is it used in RCNN?**

Selective Search is a segmentation algorithm. It starts by segmenting the image into small regions, which are then hierarchically merged based on similarities like color, texture, and shape. This process results in a comprehensive set of region proposals of various sizes and shapes, effectively capturing objects at different scales and aspects.

In R-CNN, Selective Search is used for generating the regions proposal, making the object detection task more computationally efficient. R-CNN is then applied on these regions for object detection and localization. This approach not only improves efficiency of the model.

**c) With the help of a diagram, explain working of R-CNN.**

Region-based Convolutional Neural Networks (R-CNN) streamline object detection in images through a multi-step process. Initially, R-CNN utilizes an algorithm like Selective Search to scan the image and

generate a set of region proposals - these are areas within the image that potentially contain objects. Each of these proposed regions is then individually warped to a uniform size and passed through a Convolutional Neural Network (CNN) which acts as a feature extractor, converting the regions into a format suitable for classification.

Once the features are extracted, each region is classified into object categories using a set of Support Vector Machines (SVMs), one for each category. Alongside classification, a regression model adjusts the bounding boxes of each region to tightly fit the objects identified. This two-phase approach of first proposing regions and then classifying them allows R-CNN to accurately locate and identify objects within the image, marking them with labeled bounding boxes in the final output. Despite its effectiveness, this method is computationally demanding due to the independent processing of multiple regions.