

Dated: - Dec 2025

Registration no: _____

1. Please **return** the question paper
2. Attempt all questions.
3. Don't try to write unnecessary details, it will be considered as a negative point.
4. Mobile phones or digital devices are not allowed.

(10 marks)

10 marks)

- In a digital image, the intensity of a pixel is represented using 8 bits. i) Calculate the total number of possible intensity levels in the image. ii) If an image has a resolution of 640×480 pixels, compute the total number of pixels. iii) Explain briefly how increasing the number of intensity levels improves visual quality in computer vision applications such as medical imaging or object recognition.
- Explain how image filtering and image warping differ in computer vision. Use the expressions $g(x)=h(f(x))$ and $g(x)=f(h(x))$ to support your explanation. where, $f(x)$ is original image intensity at pixel location x , $g(x)$: output image, $h(\cdot)$: transformation function, x : spatial coordinates (pixel location)
- Explain the differences between Gaussian and average filters in terms of: kernel weighting, edge preservation and noise reduction
- Why is rectification important in stereo vision?
- If a point shifts from (10, 20) to (13, 22) in two frames over 0.1 seconds, compute the 2D velocity (along x axis and y axis).

Q# 02: a) In an encoder-decoder-based approach for semantic segmentation, describe the process of applying downsampling and upsampling for a given scenario.

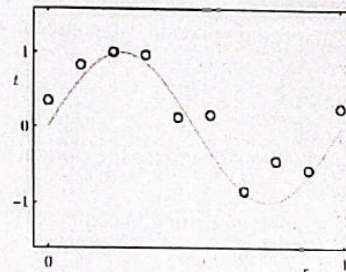
2	3	10	8					(4)
6	7	0	3					
8	5	0	1					
3	1	9	4					

b) In stereo matching algorithm (for the situation in right $\rightarrow$). Clearly elaborate how we can compute depth from disparity (disparity=baseline*f/z). Given two cameras with a baseline of Baseline 20 cm, Focal length $f=600$ pixels, and Disparity 30 pixels. Compute the depth in meters. (3)

(3)

c) In the given scenario (for the situation in right $\rightarrow$), how you can interpolate a 1D data using polynomial curve fitting (3)

(3)



Green: the ground truth curve

$$\sin(2\pi x)$$

blue points: observation values
with noise
number of observations: 10

CLO 3 Apply modern neural networks and

probabilistic graphical models to represent and model visual patterns. (15 marks).

Q# 03: In the given scenario, provide a step-by-step explanation of your calculations, for a simple CNN model with the following architecture. (5 marks)

- Input layer: $128 \times 128 \times 3$ (image size and channels)
- Convolutional layer: 8 filters, 3×3 kernel, stride 1, padding 1
- Max-pooling layer: 2×2 pooling window, stride 2
- Fully connected layer: 256 neurons

(5 marks)

- a) Calculate total number of trainable parameters in the model.
b) Calculate the output volume size after each layer.

(2,5)
(2,5)

Q#04 Compute for the following 6x6 input image.
(5 marks)

- a) Step 1: Compute the convolution of a 6x6 image with 3x3 filter,
(1,5)
b) Step 2: Apply the max pooling with 2x2 filters with stride of 2.
(1,5)
c) Step 3: Apply 2x2 transposed convolution (using conv kernel), with stride of 1 and a pad of 0 for upsampling.

3	1
1	5

Conv
Kernel

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

6 X 6 image

(2)

*

1	0	-1
1	0	-1
1	0	-1

3 X 3 filter

Q#05 You are designing a vision-based deep learning system using Transformers to model complex visual patterns. Apply your understanding of attention mechanisms and explain where and why you would use the following in computer vision tasks: a) Self-Attention, b) Cross-Attention and c) Masked Attention.
(5 marks)

CLO 4 Analyze generative models and motion estimation techniques used in computer vision to interpret and model complex visual data. (15 marks).

Q# 06: In the given scenario, two images $I_1(x,y)$ and $I_2(x,y)$, are given. Use sum of squared difference (SSD) formula to analyse the motion based on given intensity values.

$$SSD(dx, dy) = \sum_{x,y} (I_1(x, y) - I_2(x + dx, y + dy))^2 \quad (5 \text{ marks})$$

$$I_1 \begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}$$

$$I_2 \begin{bmatrix} 4 & 6 \\ 3 & 8 \end{bmatrix}$$

- a) Calculate SSD between I_1 and I_2 . For the translational alignment at $(dx, dy) = (0, 0)$ (1,5)
b) Compute SSD for the shift of $(dx, dy) = (1, 0)$ (1 pixel right shift in x-direction & zero padding) i.e. $I_2[6 \ 0 \ 8 \ 0]$ (1,5)
c) Which alignment gives the lowest SSD value? (2)

Q# 07: Given the following values at a certain pixel in two consecutive video frames. Use the optical flow constraint equation, with $I_x=4$, $I_y=3$, $I_t=-10$. (5 marks)

$$I_x u + I_y v + I_t = 0 \quad (1,5)$$

(x-flow) (y-flow)

- a) If $u=1$, compute the value of v .
b) If $v=2$, compute the value of u .
c) Check if motion vector $u=1$ and $v=2$ satisfy the optical flow constraint or not. (2)

Q# 08: a) Explain the concept of generative modelling in computer vision and describe how it differs from discriminative modelling. (1,5)

b) Discuss at least three major generative model families, such as: (1,5)

- I. Variational Autoencoders (VAEs)
- II. Generative Adversarial Networks (GANs)
- III. Diffusion Models
- IV. Normalizing Flows

c) For each model family, analyze: (2)

- I. Core architecture and mathematical foundations
- II. Strengths and limitations
- III. Example applications in computer vision (e.g., image synthesis, denoising, segmentation, anomaly detection, image-to-image translation)

Good Luck