



JK Lakshmipat University Jaipur
INSTITUTE OF ENGINEERING AND TECHNOLOGY

End Term Examination, May 2026
B. Tech. (CSE/CCE), Semester VII, 2025-26

Roll No.....

CS1228: Computer Vision

Time: 2 hours

Max. Marks: 25

Instructions to students:

1. Do not write anything other than your roll number on the question paper.
2. Mention all the assumptions for your answers clearly.
3. Use of simple non-programmable calculators is allowed.

Q1.	Explain how the Difference of Gaussian (DoG) is used in the Scale-Invariant Feature Transform (SIFT) operator for keypoint detection. How does DoG differ from the Laplacian of Gaussian (LoG) in terms of computation and application? $\begin{matrix} 10 & 10 & 10 \\ 10 & 50 & 10 \\ 10 & 50 & 10 \end{matrix}$ Given an image $I = \begin{matrix} 10 & 50 & 10 \\ 10 & 50 & 10 \\ 10 & 50 & 10 \end{matrix}$ a) Apply Gaussian smoothing using kernels with $\sigma=1.0$ and $\sigma=1.5$. b) Compute the corresponding blurred images. c) Calculate the Difference of Gaussian (DoG) by subtracting the two results. d) Comment on how the DoG highlights intensity changes in the image.	5.0 (CS1228.2)
Q2.	Design a Convolutional Neural Network (CNN) architecture that takes an input image of size 448×448 and produces an output feature map of size 7×7 with 1024 feature channels. You must use 3×3 convolution kernels. a) Propose a suitable layer-by-layer architecture, clearly indicating: • Number of filters • Stride and/or pooling operations • Output size at each stage b) Justify your design choices for spatial down-sampling and feature extraction. c) Would you recommend any normalization techniques in this architecture? If yes, explain which technique(s) and why. d) Analyze how the architecture and performance would change if 5×5 kernels were used instead of 3×3 kernels.	5.0 (CS1228.3)

Q3.	Lucas Kanade Optical Tracking is used for an Image whose Image Gradients are represented by a matrix $I_{x,y} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ and $I_t = \begin{bmatrix} -1 \\ -3 \end{bmatrix}$ Compute the velocity vectors in the x and y directions. What is your interpretation of the result?	5.0 (CS1228.4)																		
Q. 4	Explain why the YOLO model is preferred for video-based object detection/classification tasks. In a YOLO model, the predicted bounding box parameters and the corresponding ground truth values are provided in the table below, where x_c, y_c is the centre, w and h are the width and height of the bounding boxes, and c is the confidence score. <table><tr><td></td><td>x_c</td><td>y_c</td><td>w</td><td>h</td><td>c</td></tr><tr><td>Ground Truth</td><td>2</td><td>2</td><td>4</td><td>5</td><td>1.0</td></tr><tr><td>Prediction</td><td>3</td><td>3</td><td>5</td><td>6</td><td>0.7</td></tr></table> a) Show the visualization of the bounding boxes. b) Compute the intersection area between the predicted and ground truth bounding boxes. c) Calculate the Intersection over Union (IoU). d) If the prediction confidence is 0.7, compute the IoU-based loss (assume a suitable IoU loss formulation).		x_c	y_c	w	h	c	Ground Truth	2	2	4	5	1.0	Prediction	3	3	5	6	0.7	5.0 (CS1228.5)
	x_c	y_c	w	h	c															
Ground Truth	2	2	4	5	1.0															
Prediction	3	3	5	6	0.7															
Q. 5	Compare convolutional kernels and attention mechanisms. Derive the equations and provide pseudocode for backpropagation through the attention module to update the Query, Key, and Value weight matrices. Additionally, analyze the computational complexity of training attention weights in a Vision Transformer for an input image of size $n \times n$.	5.0 (CS1228.6)																		

All the best!