Q.6 Attempt any two:

i. What is Image Captioning, and how does it differ from traditional image recognition tasks? Explain the basic idea of generating a textual description of an image, and how this can be achieved using deep learning models. **5**

ii. Describe the process of training a Visual QA model, including data preparation, model architecture design, and optimization techniques. What are some common challenges and limitations in training Visual QA models, and how have researchers attempted to address them? **5**

iii. Describe the architecture of a Transformer Network, including the encoder and decoder components, and how they are trained using unsupervised or supervised learning objectives. What are some common optimization techniques used to improve the performance of Transformer Networks, such as adaptive learning rates or regularization? **5**

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No.....

Faculty of Engineering

End Sem Examination May-2023

RA3CO39 Computer Vision

Programme: B.Tech. Branch/Specialisation: RA

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Which corner detection algorithm is most suitable for detecting corners in images with varying lighting conditions? **1**

(a) Harris corner detector (b) Shi-Tomasi corner detector

(c) Moravec corner detector (d) FAST corner detector

ii. What is the difference between a Gaussian filter and a median filter? **1**

(a) A Gaussian filter is a low-pass filter, while a median filter is a high-pass filter

(b) A Gaussian filter preserves edges, while a median filter blurs an image

(c) A Gaussian filter blurs an image, while a median filter preserves edges

(d) A Gaussian filter and a median filter are the same thing

iii. What is the role of the threshold parameter in RANSAC? **1**

(a) To determine the number of iterations of the algorithm

(b) To determine the number of data points used to fit the model

(c) To determine the learning rate of the algorithm

(d) To determine the maximum distance a data point can be from the model and still be considered an inlier

iv. What is a pyramid in the context of Pyramid Matching? **1**

(a) A series of images at different resolutions

(b) A mathematical model used for feature matching

(c) A set of key points in an image

(d) A method for image segmentation

v. What is a Multi-layer Perceptron (MLP)? **1**
(a) A type of decision tree model
(b) A type of linear regression model
(c) A type of logistic regression model
(d) A type of deep learning model

vi. What is the main advantage of using an InceptionNet over a traditional convolutional neural network? **1**
(a) The ability to use different pooling methods in the same layer
(b) The ability to use different activation functions in the same layer
(c) The ability to use different filter sizes in the same layer
(d) The ability to use different optimization algorithms in the same layer

vii. What is the main advantage of using a CNN for object detection? **1**
(a) The ability to learn features automatically from data
(b) The ability to detect objects at different scales and orientations
(c) The ability to handle multiple object classes
(d) The ability to handle varying lighting conditions

viii. How does the RNN component of a CNN + RNN model extract temporal features from video frames? **1**
(a) By using pre-trained models to extract features automatically
(b) By concatenating the feature maps from the CNN component over time
(c) By applying a fully connected layer to the feature maps from the CNN component
(d) By processing the feature maps from the CNN component using a set of recurrent cells

ix. What are the two components of a Visual QA system? **1**
(a) An image encoder and a text decoder
(b) A question encoder and an image decoder
(c) A language model and a visual model
(d) A feature extractor and a classifier

x. What is the purpose of attention mechanisms in image captioning systems? **1**
(a) To reduce the dimensionality of the input data
(b) To highlight relevant parts of the image and the sentence
(c) To generate natural language descriptions of the image
(d) To handle complex sentences that require reasoning

Q.2 i. What is computer vision? Also state the applications of the CV. **2**
ii. What is linear filtering in image processing? Also state the applications of linear filtering in image processing. **3**
iii. Explain the concept of edge detection in image processing. How is it different from other feature detection techniques? **5**
OR iv. Explain the concept of Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Feature (SURF) extraction techniques. How do they work to detect and match features in images? **5**

Q.3 i. What is image processing? Also name the different types image processing techniques. **2**
ii. What is RANSAC? What problem does it solve? Explain the basic steps of the algorithm and how it is used to fit a model to noisy data. **8**
OR iii. Discuss some of the recent advances and extensions to the Bag-of-words and VLAD models, such as the Fisher Vector encoding method or deep learning-based feature extraction. What benefits do these approaches offer over traditional methods, and what are some of the challenges in implementing them? **8**

Q.4 i. What is deep learning? Also state the applications of deep learning. **3**
ii. What is a Multi-layer Perceptron (MLP), and how does it differ from a single-layer perceptron? Explain the basic architecture of an MLP, and how it can be used for classification or regression tasks. **7**
OR iii. What are DenseNets, and how do they differ from traditional CNNs? Describe the basic architecture of a DenseNet, including dense blocks, and how they can be used for image classification or semantic segmentation. **7**

Q.5 i. What Fast R-CNN? **2**
ii. What are CNNs, and how do they differ from traditional neural networks? Explain the basic architecture of a CNN, including convolutional layers, pooling layers, and fully connected layers, and how they can be used for image recognition tasks. **8**
OR iii. Explain CNN + RNN modes for video understanding in brief. **8**

Marking Scheme

RA3CO39 [T] Computer Vision

Q.1 i) (b) Shi-Tomasi corner detector **1**

ii) (c) A Gaussian filter blurs an image, while a median filter preserves edges **1**

iii) (d) To determine the maximum distance a data point can be from the model and still be considered an inlier **1**

iv) (a) A series of images at different resolutions **1**

v) (d) A type of deep learning model **1**

vi) (c) The ability to use different filter sizes in the same layer **1**

vii) (b) The ability to detect objects at different scales and orientations **1**

viii) (d) By processing the feature maps from the CNN component using a set of recurrent cells **1**

ix) (c) A language model and a visual model **1**

x) (b) To highlight relevant parts of the image and the sentence **1**

Q.2 i. What is computer vision? **1 Mark**
Also state the applications of the CV. **1 Mark**

ii. What is linear filtering in image processing? **2 Marks**
Also state the applications of linear filtering in image processing. **1 Mark**

iii. Explain the concept of edge detection in image processing. **3 Marks**
How is it different from other feature detection techniques? **2 Marks**

OR iv. Explain the concept of Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Feature (SURF) extraction techniques. **3 Marks**
How do they work to detect and match features in images? **2 Marks**

Q.3 i. What is image processing? **1 Mark**
Also name the different types image processing techniques. **1 Mark**

ii. What is RANSAC, and what problem does it solve? **4 Marks**

Explain the basic steps of the algorithm and how it is used to fit a model to noisy data. **4 Marks**

OR iii. Discuss some of the recent advances and extensions to the Bag-of-words and VLAD models, such as the Fisher Vector encoding method or deep learning-based feature extraction. **4 Marks**
What benefits do these approaches offer over traditional methods, and what are some of the challenges in implementing them? **4 Marks**

Q.4 i. What is Deep Learning? **2 Marks**
Also state the applications of Deep Learning. **1 Mark**

ii. What is a Multi-layer Perceptron (MLP) **1 Marks**
how does it differ from a single-layer perceptron? **2 Marks**
Explain the basic architecture of an MLP **2 Marks**
how it can be used for classification or regression tasks **2 Marks**

OR iii. What are DenseNets **1 Mark**
how do they differ from traditional CNNs? **2 Mark**
Describe the basic architecture of a DenseNet, including dense blocks **2 Marks**
how they can be used for image classification or semantic segmentation. **2 Marks**

Q.5 i. What Fast R-CNN? **2 Marks**

ii. What are CNNs **2 Marks**
how do they differ from traditional neural networks? **2 Marks**
Explain the basic architecture of a CNN, including convolutional layers, pooling layers, and fully connected layers, **3 Marks**
how they can be used for image recognition tasks. **1 Marks**

OR iii. Describe the process of training a CNN + RNN model for video understanding, including data preparation, model architecture design, and optimization techniques. **4 Marks**
What are some common challenges and limitations in training CNN + RNN models, and how have researchers attempted to address them? **4 Marks**

Q.6

i. What is Image Captioning **1 Mark**
how does it differ from traditional image recognition tasks? **2 Marks**
Explain the basic idea of generating a textual description of an image, and how this can be achieved using deep learning models. **2 Marks**

ii. Describe the process of training a Visual QA model, including data preparation, model architecture design, and optimization techniques. **2 Marks**
What are some common challenges and limitations in training Visual QA models **2 Marks**
how have researchers attempted to address them? **1 Marks**

iii. Describe the architecture of a Transformer Network, including the encoder and decoder components, and how they are trained using unsupervised or supervised learning objectives. **3 Marks**
What are some common optimization techniques used to improve the performance of Transformer Networks, such as adaptive learning rates or regularization? **2 Marks**

******