DEPARTMENT OF COMPUTER SCIENCE

ROLLWALA COMPUTER CENTER, GUJARAT UNIVERSITY

DEEP LEARNING

ASSIGNMENT-II - 2024

INSTRUCTIONS:

• You are expected to provide **detailed** responses to each question.

• Remember to make **honest** attempt.

• There are **4 sections**:

1. Long answer type
2. Logical thinking type
3. Real life example or Application based type
4. Complex and difficult type

• There are total **40 questions**, each section contains **10 questions**. Out of these 4 Sections, only **Long Answers & Logic Based** question will be covered in exams.

Section I - Long-Answer Questions

1. Explain the concept of convolution in the context of Convolutional Neural Networks (CNNs). How does the convolution operation work on image data, and what is the significance of using kernels or filters in this process? Provide mathematical explanations and examples.

2. Describe the role of pooling in CNNs. Compare and contrast different pooling techniques such as max pooling, average pooling, and global pooling. How do these techniques affect the performance and computational efficiency of CNNs?

3. Discuss the impact of stride on the convolution operation in CNNs. How does changing the stride length influence the output feature map size and the computational load? Illustrate your answer with examples.

4. Analyze the architecture and function of convolutional modules in modern CNNs. How do these modules contribute to the hierarchical feature extraction process in deep learning? Discuss the design choices involved in creating efficient convolutional modules.

5. Evaluate the significance of efficient convolution algorithms, such as depthwise separable convolutions and group convolutions, in modern neural network architectures. How do these algorithms help in reducing computational complexity while maintaining model accuracy?

6. Discuss the concept of random or unsupervised feature learning in CNNs. How can CNNs be used to extract meaningful features from unlabeled data? Provide examples of techniques and applications where unsupervised feature learning has been successfully applied.

7. Explore the applications of convolutional networks in computer vision tasks. Choose two specific tasks (e.g., object detection, image segmentation, or facial recognition) and explain how CNNs are designed and optimized to perform these tasks effectively.

8. Critically analyze the trade-offs between model accuracy and computational efficiency in CNNs. How do design choices such as kernel size, depth, and the use of efficient convolutional techniques impact these trade-offs? Provide examples from recent research or practical implementations.

9. Examine the challenges associated with training deep convolutional networks. Discuss issues such as vanishing gradients, overfitting, and the need for large labeled datasets. How have modern techniques like batch normalization, dropout, and data augmentation helped address these challenges?

10. Discuss the future directions of research in convolutional networks. What are the emerging trends in the development of CNN architectures, and how are they likely to impact fields such as computer vision, natural language processing, and beyond? Provide examples from recent advances in the field.

Section II - Logical Thinking Questions

1. If a convolutional layer has a kernel size of 3x3, a stride of 2, and no padding, how will the output feature map size change compared to the input feature map? Explain the logical steps used to determine the output dimensions.

2. Consider a CNN with multiple layers of convolutions and pooling. If the input image size is 224x224, and the first convolutional layer uses a 7x7 kernel with a stride of 2 and padding of 3, followed by max pooling with a 2x2 kernel and stride of 2, calculate the size of the feature map after these two layers.

3. Given a scenario where a CNN is used for image classification, how would increasing the stride in the initial convolutional layers affect the model's ability to capture fine-grained details in the images? Provide a logical argument based on the concepts of receptive fields and feature extraction.

4. If a CNN's performance on a test dataset is significantly lower than on the training dataset, what logical steps would you take to diagnose and address the issue? Consider factors such as overfitting, data distribution, and model architecture in your response.

5. In a CNN architecture, if you replace a standard convolution operation with a depthwise separable convolution, how will the computational cost and the number of parameters change? Use logical reasoning to support your answer.

6. Assume you are tasked with designing a CNN for a specific computer vision task with limited computational resources. What logical approach would you take to balance model complexity with performance? Consider factors like kernel size, number of layers, and efficient convolution techniques in your answer.

7. If you apply a pooling operation with a kernel size of 2x2 and stride of 2 on a feature map, how will it affect the translation invariance of the features? Discuss the logical implications of this operation on the robustness of the model to small input translations.

8. You are using a CNN for unsupervised feature learning on a dataset of unlabeled images. How would you logically evaluate the quality of the features learned by the CNN? Discuss potential techniques or methods for this evaluation.

9. Imagine you have a convolutional module consisting of multiple convolutional layers followed by a ReLU activation function. What logical reasoning would you use to determine the optimal number of layers and the kernel size for extracting hierarchical features?

10. If you observe that your CNN is struggling with vanishing gradients during training, what logical steps would you take to mitigate this issue? Discuss potential modifications to the network architecture, initialization, or training process.

Section III - Application based questions

1. In the medical field, CNNs are increasingly used for diagnosing diseases from medical images such as X-rays, MRIs, and CT scans. How would you design a CNN-based system to detect early signs of cancer from a dataset of medical images? Discuss the challenges involved, such as data variability and the need for high accuracy, and propose solutions.

2. Autonomous vehicles rely heavily on CNNs for real-time object detection and classification to navigate safely. How would you design a CNN to detect and classify objects such as pedestrians, vehicles, and traffic signs in various weather conditions (e.g., rain, fog, snow)? Discuss the considerations for ensuring the model's robustness and reliability in diverse environments.

3. In smart city applications, CNNs can be used to monitor urban areas through CCTV footage for purposes such as traffic management or crime detection. How would you develop a CNN-based system to detect and analyze traffic congestion in real-time from street camera feeds? Consider the challenges of handling large-scale data, real-time processing, and varying lighting conditions.

4. CNNs are widely used in agriculture for tasks such as crop disease detection and yield estimation from drone or satellite imagery. Propose a CNN-based system that can monitor large agricultural fields for signs of crop stress or disease. Discuss how you would deal with challenges like different crop types, varying growth stages, and environmental factors.

5. In environmental conservation, CNNs can be used to identify and monitor endangered species through camera traps or aerial imagery. How would you design a CNN model to automatically detect and classify different animal species from camera trap images in the wild? Discuss the ethical considerations and the need for minimizing false positives and negatives.

6. In the retail industry, CNNs can be used for customer behavior analysis by analyzing video feeds from store cameras. Propose a CNN-based solution to track customer movement patterns in a retail store and identify areas of high engagement or interest. Discuss how this information can be used to optimize store layout and improve sales.

7. CNNs are applied in disaster management to assess damage from satellite images after events such as earthquakes, floods, or wildfires. How would you develop a CNN to automatically assess the extent of damage in an urban area following a natural disaster? Discuss the challenges of working with post-disaster imagery and the importance of rapid, accurate assessments.

8. In the fashion industry, CNNs can be used for virtual try-on applications where customers can see how clothes would look on them. How would you design a CNN that accurately maps clothing items onto a person's image, considering variations in body shape, pose, and lighting? Discuss the potential impact of such technology on the retail sector.

9. CNNs are utilized in healthcare to assist in the analysis of retinal images for diagnosing eye diseases like diabetic retinopathy. Design a CNN-based system that can be deployed in rural clinics with limited access to specialists to automatically screen for such conditions. Consider the challenges of creating a model that is both accurate and interpretable by non-expert healthcare workers.

10. In the field of archaeology, CNNs can help identify historical artifacts or structures from aerial or satellite imagery. How would you design a CNN to detect and classify archaeological sites from satellite images, especially in regions where these sites are obscured by vegetation or urban development? Discuss the potential impact of such a system on cultural heritage preservation.

Section IV - Complex and Somewhat Challenging Questions

1. Consider a CNN designed for real-time object detection on a mobile device. The network uses depth-wise separable convolutions to reduce computation. However, you notice a significant drop in accuracy compared to a standard convolutional network. Propose a detailed plan to improve the accuracy while keeping computational efficiency in mind, considering techniques such as hybrid architectures or multi-scale feature extraction.

2. You are given a pre-trained CNN model that performs well on a specific computer vision task. However, the model performs poorly when deployed in a different environment with a slightly different data distribution. Discuss how you would logically approach the problem of domain adaptation or transfer learning to improve the model's performance in the new environment. Consider data augmentation, fine-tuning, and adversarial training techniques.

3. In a research project, you are required to design a novel CNN architecture for 3D image segmentation. Discuss the challenges involved in extending 2D convolutions to 3D, and propose a detailed architecture that handles these challenges. Explain how your design would efficiently handle the increased computational complexity and memory usage associated with 3D data.

4. A CNN-based facial recognition system is facing challenges with generalization due to biases in the training data (e.g., demographic imbalances). Propose a strategy to reduce bias in the model and ensure fair representation across different demographic groups. Discuss the implications of your strategy on model training, evaluation, and deployment.

5. Given a large-scale dataset with high-resolution images, you are tasked with designing a CNN that efficiently handles the input size while maintaining high accuracy. How would you leverage concepts such as dilated convolutions, multi-scale feature aggregation, and attention mechanisms to build a model that balances performance and efficiency? Provide a detailed architectural proposal.

6. In a scenario where labeled data is scarce, discuss the potential of combining CNNs with unsupervised or self-supervised learning techniques to improve feature extraction. Propose an end-to-end pipeline that incorporates these techniques and justifies the choice of specific methods (e.g., contrastive learning, clustering, or generative modeling) for the task at hand.

7. You are exploring the use of CNNs for anomaly detection in video surveillance. Discuss the challenges associated with detecting anomalies in spatiotemporal data and propose a model architecture that combines CNNs with other techniques, such as recurrent neural networks (RNNs) or transformers, to effectively detect and localize anomalies in video streams.

8. In a complex computer vision task, such as image captioning or visual question answering (VQA), CNNs are used in conjunction with other models like LSTMs or transformers. Discuss the role of CNNs in these tasks and propose an architecture that optimally integrates CNNs with sequence models. Explain how attention mechanisms could be employed to enhance the model's performance.

9. Consider a scenario where you need to deploy a CNN on edge devices with extremely limited computational power and memory. Discuss the trade-offs involved in model compression techniques such as quantization, pruning, and knowledge distillation. Propose a strategy to compress the model while maintaining acceptable accuracy for a specific computer vision task.

10. You are tasked with designing a CNN for a multi-modal fusion task where the input consists of both images and textual descriptions. Discuss how you would architecturally integrate these different modalities within a CNN framework. Explain the role of feature fusion, cross-modal

attention, and joint embedding spaces in your design, and how these components contribute to improving task performance.