National University of Computer and Emerging Sciences

Islamabad Campus

Computer Vision (AI4002)

Sessional-II Exam

Course Instructor(s):
Ms. Khadija Mahmood
Section(s): (AI-K and AI-J)

Total Time (Hrs): 1
Total Marks: 40
Total Questions: 7

Date: Nov 4, 2024

_____	_____	_____
Roll No	**Course Section**	**Student Signature**

Do not write below this line.

Attempt all the questions.

[CLO:3 & 5. Apply appropriate image processing methods for image filtering, image restoration, image reconstruction, segmentation, classification and representation]

Q1: Attempt the MCQ's on provided bubble sheet **[5 Marks]**

1. **In the context of neural networks, what is the primary benefit of using skip connections (or residual connections)?**

   a) They decrease the model's complexity.

   b) They allow gradients to flow more easily through the network during backpropagation.

   c) They increase the overall depth of the network.

   d) They eliminate the need for activation functions.

2. **You are tasked with classifying images of rare animal species. Which of the following approaches is most suitable if you have limited labeled data?**

   a) Training a model from scratch

   b) Using transfer learning with a pre-trained model

   c) Using a simpler model with fewer parameters

   d) Increasing the number of layers in your current model

3. **In a self-supervised learning task, you apply a jigsaw puzzle approach to train your model. What is the primary objective of this technique?**

   a) To predict the next frame in a video sequence

   b) To learn spatial relationships and context from image patches

c) To improve classification accuracy on labeled data

d) To reduce the number of parameters in the model

4. **You're tasked with training a deep model for recognizing various wildlife species, but the dataset is small. A large, pre-trained model trained on ImageNet is available, which includes many classes but not specifically for wildlife. Given these constraints, you decide to use transfer learning.**

What would be the most effective approach to use the pre-trained model for this task?

a) Freeze all layers in the pre-trained model and add a new classifier layer for the wildlife species.

b) Fine-tune only the last few layers of the pre-trained model after adding a new classifier layer.

c) Randomly initialize the weights and train from scratch to adapt specifically to wildlife species.

d) Use the pre-trained model directly without modification since it already has feature representations.

5. **An AI research team is building a model that learns representations of videos by training on unlabeled video data. They decide to use a self-supervised learning approach where the model predicts the order of shuffled frames as a pretext task. After training, they observe that the learned representations generalize well to action recognition tasks with minimal fine-tuning.**

What is the likely reason for this pretext task leading to effective representations for action recognition?

a) Predicting frame order forces the model to capture temporal relationships, which is crucial for understanding actions.

b) Shuffling frames ensures that the model learns each frame in isolation, improving its spatial feature extraction.

c) This approach avoids learning temporal patterns, focusing on object recognition within individual frames.

d) The task directly labels actions, which simplifies the transfer to action recognition tasks.

[CLO:3 & 5. Apply appropriate image processing methods for image filtering, image restoration, image reconstruction, segmentation, classification and representation]

Short Questions: Write brief short answers. Long answers will be negatively marked. [20 Marks]

Q2. Explain the term bottleneck under the concept of U-Net? **[3 Marks]**

In U-Net, the bottleneck is the central, lowest-resolution part of the network that connects the encoder and decoder. It captures the most compressed representation of the input, retaining essential global features while discarding irrelevant details for effective reconstruction in the decoder.

Q3. Given the diagram illustrating the effects of stacking deeper layers in a "Plain" convolutional network, provide a one-line statement describing the hypothesis or motivation behind the ResNet model, without explaining how ResNet addresses this issue. **[3 Marks]**

A deeper model should be able to perform at least as well as the shallower model.

Q4. A traditional CNN architecture is used to convolve the input image, but how does it map back to the original image pixels for object detection? For example, if the input image is 80×80 and the object center point is at (30, 40), and the feature map size after convolution is 20×20, how can the same center point be accurately detected, given that no up-sampling technique is used? Write your answer in precise words. **[3 Marks]**

In a traditional CNN without up-sampling, the mapping back to the original image pixels is achieved by calculating the **receptive field** of each feature map cell.

Q5. Apply the transposed convolution on given image (2 by 2) using kernel (4 by 4). Resultant output must be of size 7 by 7. **[5 Marks]**

Image:

1	4
3	1

Kernel:

1	4	3	2
4	1	2	4
3	2	1	3
2	3	4	1

Date

1	4
3	1

image

1	4	3	2
4	1	2	4
3	2	1	3
2	3	4	1

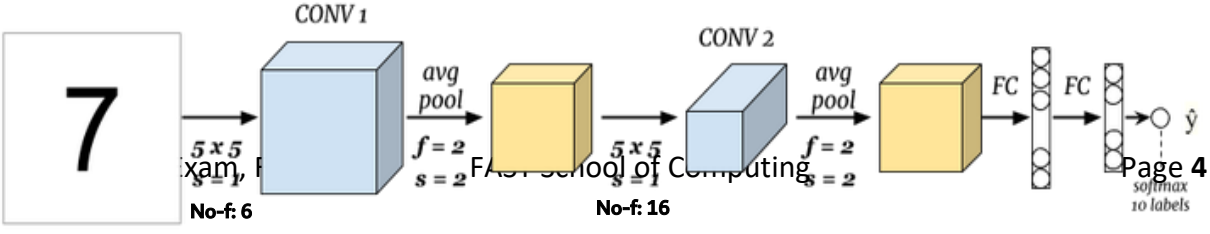
Kernel

1 4 3 2 4 16 12 8
4 1 2 4 16 4 8 16
3 2 1 3 12 8 4 12
2 3 4 1 8 12 16 4
3 12 9 6 1 4 3 2
12 3 6 12 4 1 2 4
9 6 3 9 3 2 1 3
6 9 12 3 2 3 4 1

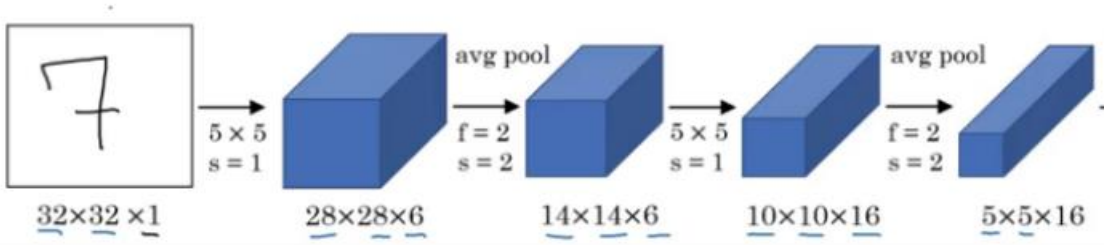
1 4 3 6 16 12 8
4 1 2 20 4 8 16
3 2 1 15 8 4 12
5 15 13 16 16 19 6
12 3 6 16 1 2 4
9 6 3 12 2 1 3
6 9 12 5 3 4 1

Q6. Consider the following CNN architecture taking input of handwritten digit image and classify the digit from 0 to 9 numbers, write the output dimension at every step. What are the total parameters of this network? **[6 Marks]**

7



Dimensions if we take input size = 32*32*1

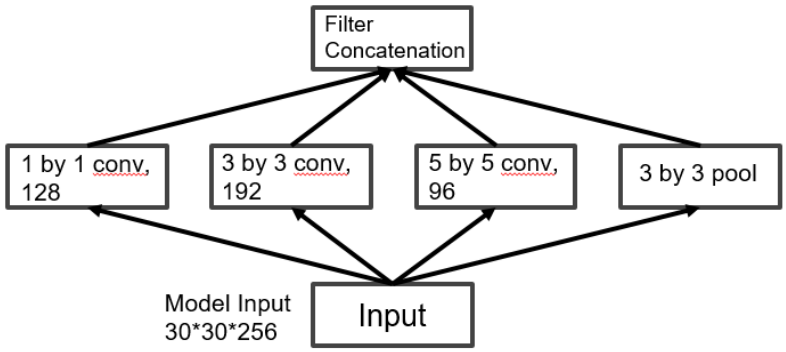


Total parameter = 572

Long Question . **[15 Marks]**

Q7. Consider the following Google Net model taking input of size 30*30*256

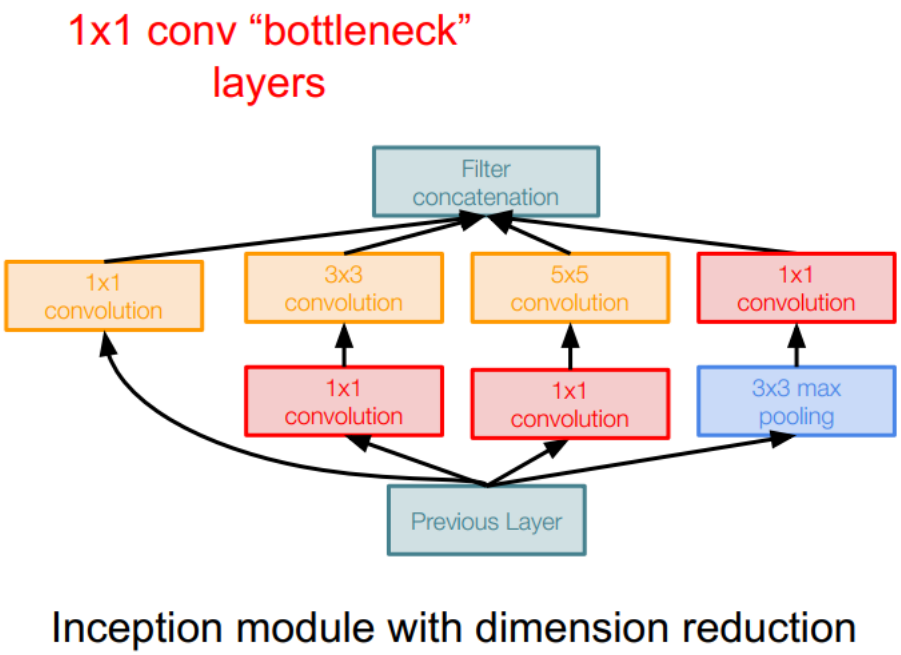
a) what is the output size after filter concatenation?



First Conv = 30*30*128
Second Conv = 30*30*192
Third Conv = 30*30*96
Pool = 30*30*256

Filter Concatenation: 30*30(128+192+96+256) = 30*30*672

b) Reconstruct the above architecture incorporating dimensionality reduction concept for making it less computation expensive.



Inception module with dimension reduction

c) Can we use Google Net model for segmentation task? If Yes, then what changes you'll make? If No write the reason.

Yes, the GoogleNet model (Inception architecture) can be adapted for segmentation tasks. Replace the fully connected layers with a **segmentation head** that outputs a pixel-wise classification.