DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING
Indian Institute of Technology Jodhpur

March 29, 2022

Image and Vision Computing
Ph.D. Comprehensive Examination

Time: 4 hours

Maximum Marks: 100

1.   Total: 25

   (a) What are three technical differences between Gaussian Pyramid, DCT and DWT? Be technical - no stories. [3]

   (b) Define corners, lines, and edges in a given image and technically describe the algorithms to detect each one of them. [6]

   (c) Given an image F = [130 11 67; 10 10 50; 80 90 100]. What is the output if you apply (a) mean filter, and (b) median filter? [2]

   (d) Consider an image [2 1; 3 1]. Resize the image to three times by using (i) bilinear interpolation and (ii) bicubic interpolation. [2]

   (e) Consider a one dimensional image f(x) = [60 60 60 100 100 100]. What are the first and second order derivatives? [2]

   (f) Detecting photo of a photo: If a photo is given to you, you need to tell if the photo is real or "photo-of-a-photo". An example of photo-of-a-photo is shown in Figure 1. You are asked to develop two novel algorithms for detecting if a given photo contains original content or it is a photo-of-a-photo: one algorithm using handcrafted features and one algorithm using deep learning pipeline. For each of those, design a conceptual diagram, provide justifications for each step and the specific order and techniques you adopt in combining different components. Try to provide as much information as needed. [10]

2. Answer the following questions: Total: 25

   (a) What is the vanishing gradient problem in object detection? Can you propose a strategy to deal with this problem? If your proposed solution is used in any of the existing well-known neural network models, write the name of the model. [6]

Figure 1: Photo of a photo.

(b) What is Densenet? What are its major advantages over its predecessors? Explain the [6]
main role of the transition layer in Densenet.

(c) What is mean average precision (mAP) for an object detection algorithm? Is mAP [4]
dependent on the IoU threshold? Explain.

(d) Why is faster RCNN faster than RCNN? Explain the advantage of YOLO over faster [4]
RCNN.

(e) Consider an automated video surveillance system installed in a toll plaza. The system [5]
is intended to analyze the videos/ images in each toll booth and count the total number
of persons crossing the toll plaza per day. Assume that only cars pass through that toll
plaza. Explain how you would design a system for this task mentioning the potential
challenges.

3. Answer the following questions with reference to the paper 'P-CNN: Part-Based Convolutional
Neural Networks for Fine-Grained Visual Categorization ':   Total: 25

(a) Write a technical summary of the proposed algorithm. It should also include technical [5]
limitations of the proposed algorithm.

(b) Present a technical solution to address the limitation of P-CNN that you have just [4]
identified (it should be your solution). Argue why your algorithm should handle the
limitations of P-CNN.

(c) Extend the proposed algorithm if the input is part-wise randomly masked? For instance, [6]
the input image is 50% masked in patches using a 3×3 mask with all values 0. Mathematically
explain your extended algorithm along with a block diagram.

(d) The paper defines a duplex focal loss in Equations 7, 8 and 9. In place of the current [6]
Equation 8, replace it with SVM style margin based loss. Extend the formulation of
duplex focal loss with the margin based loss. Technically argue why your proposed loss
is better than the existing one.

(e) Present a technical description of a handcraft approach of Fine-Grained Visual Categorization. [4]
Compare and contrast it with P-CNN.

4. Answer the following questions with reference to the paper 'Color Balance and Fusion for
Underwater Image Enhancement':   Total: 25

(a) What is gray world assumption? How does it help in the design of the proposed method? [3]

(b) How can the problem of saturation of color channel arise? [2]

(c) Explain the following statement using equation 4: 'In the proposed method, the red [5]
channel compensation depends on the global characteristics of the image'. Explain a
potential problem of making the red channel compensation globally dependent considering
that gray world assumption is not valid.

(d) Propose a strategy (and write a pseudocode) to make the red channel compensation locally [4]
dependent.

(e) Consider a situation where the underwater environment is such that it causes unequal [2]
channel attenuation at different regions of the image. You may either apply a locally
dependent or a globally dependent red channel compensation strategy. Which of these
two would you apply? Justify your answer.

(f) What is Gamma correction? How does it help the proposed method? What is the problem [5] associated with Gamma correction in the context of the proposed method? How do the authors take care of that problem?

(g) 'As we move upward in a Gaussian Pyramid, we get more and more high frequency [4] components'. Do you agree with this statement? Justify your answer. In this context, also justify that a Laplacian pyramid is a bandpass pyramid

_____