NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER 2 EXAMINATION 2023-2024

EE6427 – VIDEO SIGNAL PROCESSING

April / May 2024 Time Allowed: 3 hours

INSTRUCTIONS

1. This paper contains 4 questions and comprises 4 pages.
2. Answer all 4 questions.
3. All questions carry equal marks.
4. This is a closed book examination.
5. Unless specifically stated, all symbols have their usual meanings.

---

1. (a) In a compression scheme, a data source consists of eight symbols, with the probability distribution given in Table 1.

Table 1

Symbol	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Probability of occurrence	0.02	0.05	0.08	0.10	0.14	0.16	0.19	0.26

(i) Design a suitable set of Huffman codewords for the eight symbols. Clearly show all the key steps and calculations.

(8 Marks)

(ii) A student originally uses 8 bits to represent each symbol in an uncompressed scheme. Find the compression ratio of the Huffman coding scheme developed in part (i) when compared with the original uncompressed scheme.

(6 Marks)

Note: Question No. 1 continues on page 2.

(iii) Find the entropy of the data source. Briefly discuss whether it is possible to design a codeword set which can achieve a target of less than 2.5 bits/symbol.

(5 Marks)

(b) Draw a simple block diagram of the baseline JPEG decoder. Clearly label all the key components in the diagram.

(6 Marks)

2. (a) In a simple Convolutional Neural Network (CNN), an input image **A** passes through a convolutional layer, followed by an activation layer and a max pooling layer. The output from the max pooling layer is then used for further processing. The grayscale image **A** is given by:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 1 \\ 4 & 0 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

The convolutional layer has the following settings: the current filter is given by **F** below, the amount of zero padding at each side of the image is 1, and the stride both horizontally and vertically is 2.

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

The activation function used in the activation layer is given by a sigmoid function:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad .$$

The max pooling layer uses 2×2 max pooling with a stride of 2.

(i) Find the output after the convolution layer.

(ii) Briefly discuss the effect of filter **F** when applied to the input image.

(iii) Find the output after the activation layer.

(iv) Find the output after the max pooling layer.

Note: Question No. 2 continues on page 3.

(v) A student would like to make the following changes to the input image and the convolutional layer:

- Change the grayscale image **A** to an RGB image **B** with a spatial dimension of 100×100.
- Change the channel number of the output feature maps to 6 for the new convolutional layer. Assume the spatial dimension of the filter remains the same.

Find the number of trainable parameters of the new convolutional layer after the changes. Assume no bias is used in the calculation.

(18 Marks)

(b) A student would like to develop an Artificial Intelligence (AI) model to perform short video clip genre classification with high accuracy performance. The type of genre may include comedy, action, romance, etc. Assume the visual feature from each video frame has been embedded into a feature vector. State clearly which of the following models is most likely to satisfy the student's need: (i) CNN, (ii) Vanilla Recurrent Neural Network (RNN), or (iii) Transformer. Briefly justify your answer.

(7 Marks)

3. (a) State clearly whether the following object detectors are one-stage detector or two-stage detector: (i) R-CNN and (ii) YOLOv7. Which of the two object detectors above is a more suitable choice if speed is the key consideration in an object detection application? Briefly justify your answer.

(7 Marks)

(b) Sketch a diagram of window-based self-attention and shifted window-based self-attention in the Swin Transformer. Briefly describe the objectives of these windows in the Swin Transformer.

(6 Marks)

(c) List the key steps in the tracking-by-detection multiple-object tracking in video.

(6 Marks)

(d) Briefly describe how Temporal Shift Module (TSM) can achieve good computational efficiency in video action recognition.

(6 Marks)

4. (a) Briefly describe the purpose of chroma subsampling in the MPEG-1 video compression.

(4 Marks)

(b) Draw a flowchart of the P-frame encoding in the MPEG-1 standard. Clearly label all the key steps in the flowchart. Briefly explain the role of entropy encoding in the process.

(9 Marks)

(c) Rank from the best to the worst the following 3 motion estimation methods in terms of: (i) computational speed and (ii) accuracy:

- Three-step search
- Full search
- 2D logarithm search.

(6 Marks)

(d) Briefly discuss how MPEG-2 scalability can be used in networks with variable bitrate channels.

(6 Marks)

END OF PAPER

EE6427 VIDEO SIGNAL PROCESSING

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.