# NANYANG TECHNOLOGICAL UNIVERSITY

## SEMESTER 1 EXAMINATION 2023-2024

## EE6427 – VIDEO SIGNAL PROCESSING

November / December 2023 Time Allowed: 3 hours

**INSTRUCTIONS**

1. This paper contains 4 questions and comprises 5 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) The two-dimensional Discrete Cosine Transform (2-D DCT) matrix of an $N \times N$ pixel block is given by:

$$\mathrm{T}(i,j) = \begin{cases} \dfrac{1}{\sqrt{N}}, & \text{if } i = 0 \\ \sqrt{\dfrac{2}{N}} \cos \dfrac{(2j+1)i\pi}{2N}, & \text{if } i > 0 \end{cases},$$

where $i$ and $j$ are the row and column indices, respectively.

(i) Determine the 2-D DCT matrix **T** for a 4×4 pixel block. Round your answer to 4 decimal places.

(4 Marks)

(ii) Based on your result in part (a)(i), calculate the 2-D DCT of the following pixel block **A**. Round your answer to 3 decimal places.

$$\mathbf{A} = \begin{bmatrix} 20 & 20 & 0 & 0 \\ 20 & 20 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(6 Marks)

Note: Question No. 1 continues on page 2.

(b) In a new image compression scheme, a student would like to follow similar steps in the baseline JPEG to perform grayscale image compression. However, the student proposes to partition image into multiple 4×4 pixel blocks, perform 4×4 DCT for each pixel block, followed by corresponding quantization and entropy encoding.

(i) Briefly discuss a main similarity and a main difference between the basis functions of the DCT used in this new compression scheme with the basis functions used in the baseline JPEG compression.

(5 Marks)

(ii) Write down a suitable quantization table for this new compression scheme. Briefly justify your answer.

(5 Marks)

(iii) Two pixel blocks **B1** and **B2** undergo DCT to obtain the corresponding DCT coefficient blocks **C1** and **C2**, respectively as follows.

$$\mathbf{C1} = \begin{bmatrix} 100 & 52 & 43 & 22 \\ 54 & 44 & 23 & 20 \\ 41 & 24 & 22 & 18 \\ 25 & 18 & 16 & 15 \end{bmatrix}, \quad \mathbf{C2} = \begin{bmatrix} 100 & 40 & 20 & 12 \\ 42 & 22 & 13 & 8 \\ 21 & 14 & 4 & 0 \\ 15 & 0 & 0 & 0 \end{bmatrix}$$

The DCT coefficient blocks **C1** and **C2** subsequently go through the quantization and entropy encoding. State clearly which pixel block **B1** or **B2** will likely experience more reconstruction error during image decompression. Briefly justify your answer.

(5 Marks)

2. (a) A Long Short-Term Memory (LSTM) network has the following settings.

Initial hidden state, $\mathbf{h}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, Initial cell state, $\mathbf{c}_0 = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$,

Forget gate weight matrix, $\mathbf{W}_f = \begin{bmatrix} \mathbf{W}_{hf} & \mathbf{W}_{xf} \end{bmatrix} = \begin{bmatrix} 0.1 & 0.2 & 0.5 & 0.6 \\ 0.3 & 0.4 & 0.7 & 0.8 \end{bmatrix}$,

Input gate at timestep $t = 1$, $\mathbf{i}_1 = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$,

Gate gate at timestep $t = 1$, $\mathbf{g}_1 = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}$,

Output gate at timestep $t = 1$, $\mathbf{o}_1 = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$,

Note: Question No. 2 continues on page 3.

Input at timestep $t=1$, $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Assume no bias is used in the computation of the LSTM. The sigmoid and tanh functions are given as follows.

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

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

(i) Find the forget gate $\mathbf{f}_1$ at timestep $t=1$. Comment on your obtained result.

(ii) Find the cell state $\mathbf{c}_1$ at timestep $t=1$.

(iii) Find the hidden state $\mathbf{h}_1$ at timestep $t=1$.

(13 Marks)

(b) Briefly describe the function(s) of (i) transformer encoder and (ii) position embedding in a Vision Transformer (ViT).

(6 Marks)

(c) Briefly discuss the key advantage(s) of using Swin Transformer when compared with Vision Transformer (ViT) to perform image classification.

(6 Marks)

3. (a) Briefly discuss the reasons why different encoder products (hardware or software) following the same video coding standard may have different coding performance.

(5 Marks)

(b) In the MPEG video coding, a frame is often coded as either an I-frame, a P-frame, or a B-frame. Rank these frames in terms of coding efficiency and briefly discuss why it is not suitable to use too many B-frames in the MPEG video coding.

(7 Marks)

(c) Figure 1 on the next page shows a group of pictures (GOP) structure that uses four B frames and two P frames in the GOP, i.e., IBBPBBP. Briefly explain the decoder order of these frames and the reasons behind it.

Note: Question No. 3 continues on page 4.

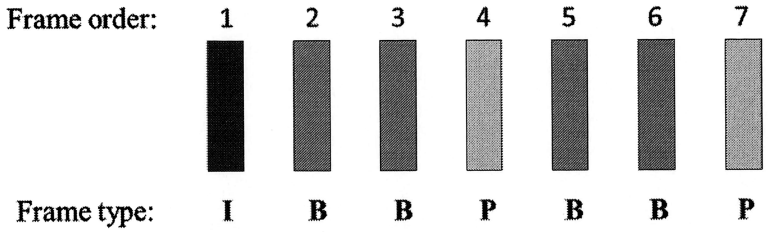


**Figure 1**

(7 Marks)

(d) Multiple object tracking (MOT) aims at predicting trajectories of multiple targets in video sequences. Briefly explain two main types of methods for multiple objects tracking and their differences.

(6 Marks)

4. (a) A vector $\boldsymbol{v}$ is given by

$$\boldsymbol{v} = [10,\ \ 13,\ \ 25,\ \ 26,\ \ 29,\ \ 21,\ \ 7,\ \ 15].$$

Find the outputs of one-level Haar wavelet transform, two-level Haar wavelet transform and three-level Haar wavelet transform of $\boldsymbol{v}$.

(8 Marks)

(b) Given the brightness constancy equation:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0,$$

explain what each of the five derivatives in the brightness constancy equation measures.

(5 Marks)

(c) Given a file containing the following characters with the frequencies as shown in Table 1, if Huffman Coding is used for data compression, determine the Huffman code for each character.

Note: Question No. 4 continues on page 5.

**Table 1**

| Characters | a | b | c | d |
|---|---|---|---|---|
| Frequencies | 40 | 20 | 10 | 10 |

(4 Marks)

(d) Consider a generative adversarial network (GAN) with a generator $G(z)$ and a discriminator $D(G(z))$, which is trained to produce images of apples. The perfect discriminator outputs 1 for a real apple instance, and 0 for a fake apple instance.

(i) In the early stage of training, is the value of $D(G(z))$ closer to 0 or 1? Explain why.

(ii) When the GAN is successfully trained, can the discriminator classify the generated images as apple or non-apple correctly? Briefly justify your answer.

(8 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.