|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Question | **1** | **2** | **3** | **4** |
| Marks | **5** | **5** | **5** | **10** |
| Total | **25** | | | |

**Name (***as in registration***): Kar Chaudhuri Anirban**

**Course:** Vision Systems

**Date:** 23 January 2021

**Written Test**

**Q1.** The images we get in general are in RGB format. This is the type of colour image we see in our screens. In image processing, often we need to convert these colour images into gray-scale images, which have less information. Why do we need to perform this conversion?

Sometimes, coloured information doesn't help us identify important edges or other features and are deemed to be noise. It is easier to convert to grayscale with binarized threshold operation such that edge detection and feature identification is smooth and easy with sufficient information.

Colored images deceptive to human eye, more difficult to visualize and analyze too, especially with problems involving illumination and occlusion. Algorithm processing and decision-making time is faster with grayscale images for various preprocessing and feature extraction steps like segmentation, noise reduction and thresholding involving grayscale images that have fewer dimensions than RGB format counterparts. Difference in speed becomes obvious with large number of images.

**Q2.** YOLO v3 is one of the solutions we use to perform object detection. The output of the model consists of the positions and the sizes of bounding boxes, the class scores and the objectness scores. Generally, we need to apply non-maximum suppression on the boxes detect by YOLO v3? Why is this step required?

Non-max suppression selects the best bounding box for an object and rejects or “suppress” all other bounding boxes. The non-max suppression identifies the bounding box with max objectiveness score and then remove other boxes with high overlapping regions. This process is iterative until there is no more reduction of boxes.

**Q3.** Suppose after a series of processing steps, you get a binary mask. But you realize there are little dots scattered around the mask, what should you do? Why is it important to produce an accurate mask?

The little dots are noise and should be filtered out/removed. Median filtering is a popular method. The median filter considers neighbourhood around a pixel, sort values in increasing order, determines and output the median values on a noisy pixel. An accurate mask is a noise free one where only Region of Interest of original image (indicated as 1) is portrayed in white while background is black. Noisy patches do not add value in identifying shape of object under consideration for further detection, recognition and analysis.

**Q4.** Both Viola-Jones algorithm and Inception v1 can detect the presence of human in an image. Compare the strength and weakness of each algorithm. Specify a suitable use case for each algorithm.

Viola-Jones algorithms is easy to understand and implement. It takes little time to detect objects for a well-trained YOLO model accurately. Lesser data needed than complex CNN based algorithms also used for image and video analytics. No rescaling of images required which saves computation memory and time.

However, there are downsides of Viola-Jones algorithm as well. Training time is slow and its suitability is limited to binary classification. It becomes ineffective with objects rotated or tilted with sideways facing the front. It is sensitive to noise and high or low brightness.

Inception V1 is a neural-network based model. It needs lots of data and time to be well-trained unlike Viola-Jones algorithm. Sometimes, data quantity is limited. However, it is highly accurate and can be used for multiclass detection and recognition as well. It is capable of auto feature engineering and selecting important features with the help of convolutional and pooling layers. However, inception v1 like other neural networks can be prone to overfitting too and harder to finetune than Viola-Jones.

Viola-Jones can be used in simple anxiety detection cases (indicating if a patient is depressed) or not. Inception V1 suitable for road object sign detection where there are multiple class labels and need higher accuracy as well to avoid life-endangering calamity.