Question 3

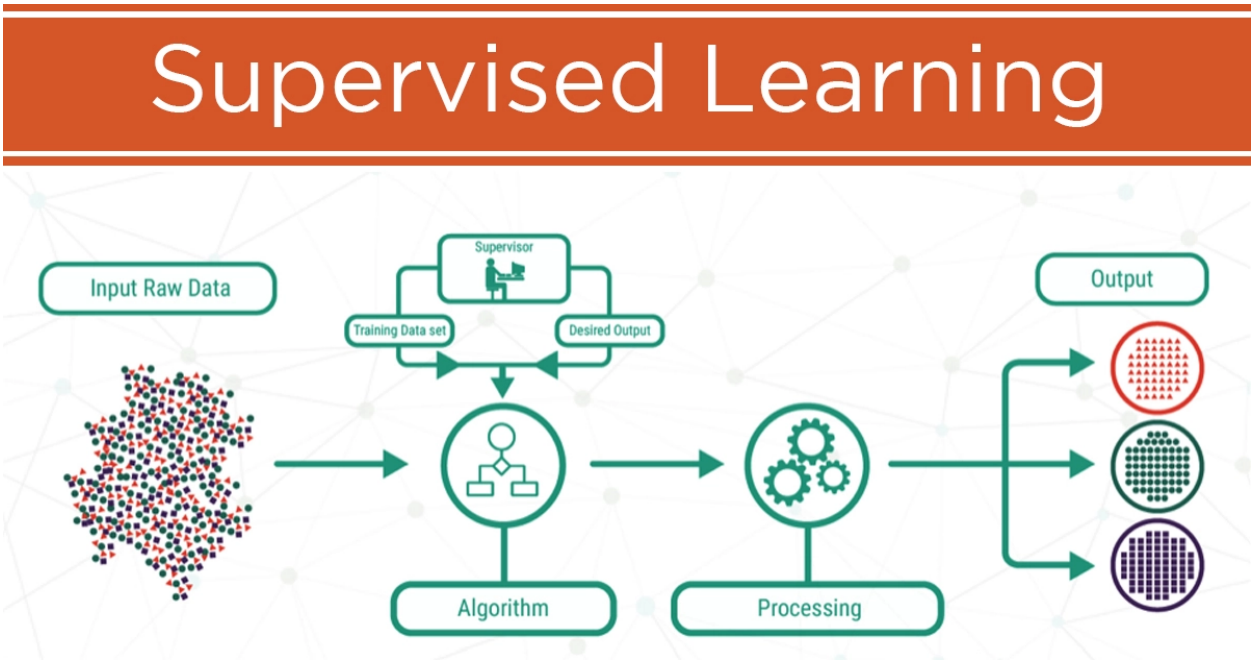
Candidate Number: 59069

Visual Intelligence - CS3VI18

3.

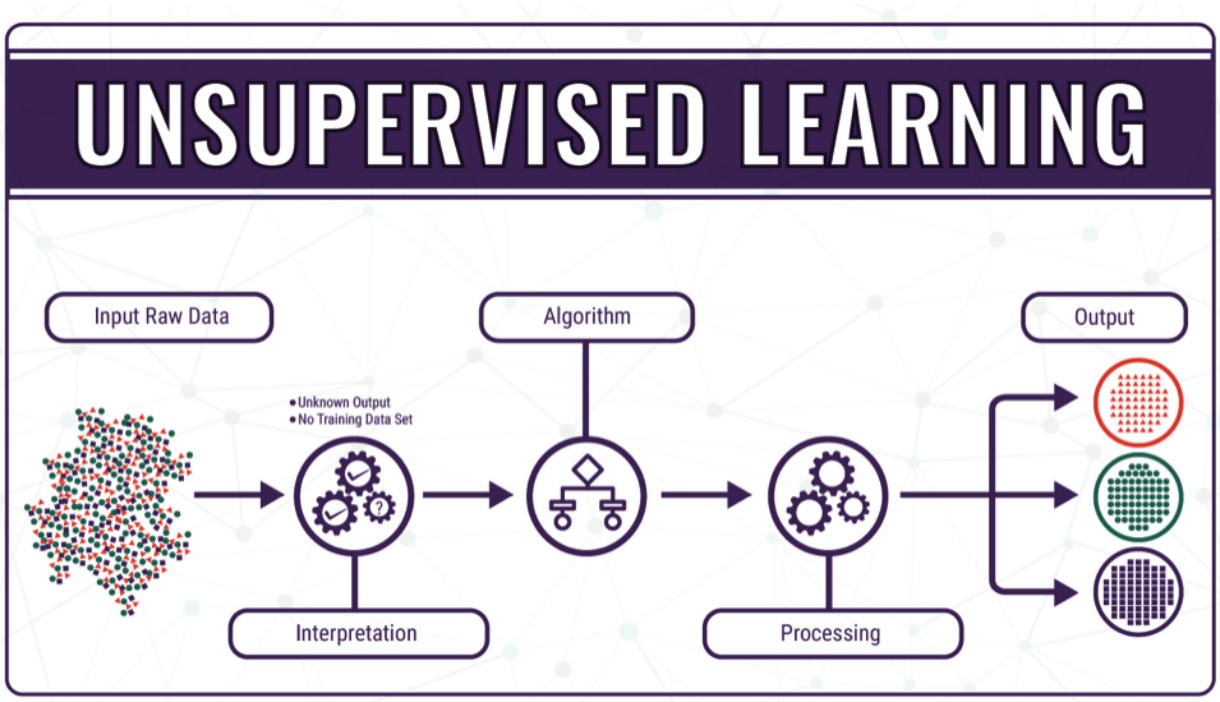
a) Describe the models, with diagram assistance, of “supervised learning” and “unsupervised learning” with the context of image based pattern classifications. Compare and contrast the two learning methods. (8 marks)

Supervised learning is a machine learning method that functions from a labelled data set, each item in the training data set consisting of an input value and a desired output value. This method can be assimilated to concept learning observed in human and animal psychology. Supervised learning is more than often chosen for classification problems, mapping an input to a continuous output. It follows a methodology of inputting the data, training the algorithm on a labelled data set, until it has learnt the problem, then presenting with an unseen dataset for validation purposes before being deployed for live use on more unseen datasets.



On the contrary unsupervised learning is often used associated clustering, representation learning and density estimation. The similarity between these methods is the desire to explore and learn the data without explicitly using labels. These generally include algorithms such as principal component analysis. Unsupervised learning is

favoured for exploratory data tasks due to the ability to identify structures throughout data - this makes it extremely effective for clustering multidimensional data and revealing trends that would be invisible to a humans methods.



Both supervised learning and unsupervised learning have their strengths and weaknesses and their own optimal use cases. Unsupervised learning models, cannot be trained using labelled dat aand therefore follow a lot more sequential order of events,

Input -> interpret -> algorithm, -> processing -> output.

	Supervised Learning	*Unsupervised Learning*
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

Supervised Learning models are those such as Logistic Regression models, Artificial Neural networks, Naive Bayes, and Random forest classifiers. Whereas Unsupervised make use of K-means clustering methods and autocoders.

b) In a video sequence, it contains a road acne with a moving car. A frame if the video sequence is shown in figure Q3-1 . You are asked to remove the moving car from the video sequence and recover the road scene in it. Design an integrated algorithm that can:

- Identify the moving car,
- Outline the shape of the car either automatically or semi-automatically,
- Remove the car from each frame of the video sequence, and finally fill the hole (where the car used to be) in each frame to recover the road scene.

Figure Q3-1. A video frame containing a road scene with a moving car.

State any assumptions you make in your answer, and identify possible flaws.

(12 marks)

- Assumptions made:
    - Camera is still
    - Wind is not sufficient to move the camera significantly.

Due to the assumptions made the camera will be sufficiently stable and that the object would be moving through the frame. Background subtraction was clearly one of the most reasonable choices of object detection.

Background subtraction is a method that only functions effectively with static cameras where the background is constant or at least relatively unchanged compared to the main article designed to be detected.

Background detection works by comparing the current frame to a pre-saved reference frame and highlighting the different zones. It is possible to use an aggregate of a selection of the previous frames in order to detect the change between them instead, this is known as “Temporal Average Filter” however it does not seem as effective for this situation as it a high computational cost, and is slower and less efficient than standard background subtraction.

Background subtraction can be used to create a foreground mask around the pixels which differ between the current frame, by using this, the car can be found and highlighted / outlined. This could also be extended on by altering the threshold values of the mask, or even by further implementing an edge detection algorithm as well, such as canny edge detection.

Once this mask has been created the object can be removed. The simplest solution herre, would be to remove the entirety of the object flagged in the foreground mask, and replace them directly with the pixels taken from the reference image.

Other options include, filling with a solid block colour, however this is very basic and seldom appropriate. Or the interpolation of the surrounding pixels to try and recreate a realistic and dynamic output. In this situation, aligning with the assumptions made

already, the most appropriate solution seems to be to fill with the pixels taken when the identical mask is placed over the reference image. If this method were to be improved slightly without causing too much computational stress on the system, multiple reference images could be stores accounting for change in sunlight and environment conditions and then the most appropriate fixed to fill the video, stopping the risk of an object being removed and replaced with a sunny day, when the image was taken on a dark winter's night.