Task 1

Identify which frames in the clip have the staff present?

Base Assumptions

1. We are only given the videos of the man walking around (no sensors to detect the badge)
2. There are videos of the man from different angles (so the badge and the face of the man is always visible, as well as not having the pose estimation problem)
3. The cameras used to record the man are identical.

Solution

Implement a multiple object tracking algorithm that uses a neural network to identify if someone is wearing a badge.

Multiple Object Tracking (MOT) Algorithm

The MOT algorithm can be defined as the process of locking on to multiple moving objects (in the case of more than 1 person wearing the badge) within a frame and being able to determine if the object is the same as the one from a previous frame. Below are some examples of algorithms which could be used.

1. **You Only Look Once (YOLO)**: Detects moving objects by predicting bounding boxes, class probabilities and confidence scores for multiple objects simultaneously in an input image.
2. **Region-based Convolutional Neural Network (R-CNN)**: Performs object localization that groups regions based on their pixel intensities.
3. **Single Shot Multibox Detector (SSD)**: Uses a neural network to predict bounding boxes and class labels for objects in an image.

In the case of the objects being covered by another object or briefly disappear, MOT algorithms will be able to predict its position based on its previous trajectory, even when the object is temporary obscured.

Comparison

Evaluating the performance of multiple object tracking (MOT) algorithms involves a systematic assessment of their ability to track objects accurately and robustly in various scenarios. The criterion for a practical MOT algorithm should involve:

1. Multiple Object Tracking Precision (MOTP): Measures the accuracy of localization of detection boxes
2. Multiple Object Tracking Accuracy (MOTA): Measures the accuracy of both the tracker and detection
3. Inference Speed: The ability of the algorithm to process input data (in frames) and generate tracking results within a certain amount of time.

Training and Testing the Model

After selecting a suitable MOT algorithm. The next task is to train the neural network, the following steps should be:

1. Data Preparation: Prepare a large dataset of employees wearing the accessory of different angles, different environments, etc.
2. Data Preprocessing: Crop, scale, and normalize the images if required.
3. Training the Model: Split the dataset into training, testing and validation dataset. The training dataset will be used to train the model, the validation dataset will be used to tune the hyperparameters and the testing dataset will be used to test the accuracy of the model.
4. Once the model has been tested to fit certain criteria, continue to monitor the performance of the model when necessary.

Task 2

Find the man's xy coordinates.

Centroid Tracking (CT) Algorithm

In the initial question, an object detection algorithm (YOLO, R-CNN, etc.) was used to detect the objects of interest (the people wearing the badge). The detection algorithms would then place bounding boxes around the detected objects.

1. Compute the centroid of the bounding box in the initial frame. The centroid represents the center point of each object with coordinates (x0, y0).
2. In the first frame, the CT algorithm initializes tracking for each detected object by associating a unique ID with the centroid. This ID will be used to track the corresponding object for each passing frame.
3. As the MOT algorithm continues to detect the object and update the bounding box, the CT algorithm would then attempt to compute the centroid for the updated bounding boxes.
4. This matching process then uses a distance measuring metric such measuring the Euclidean distance between centroids. By doing this, it allows the algorithm to identify which object in the current frame corresponds to the object in the previous frame and track their 2D coordinates (x,y).