1. Features of CNN/Yolo useful for video proctoring

1. Object Detection: Both CNN and YOLO can perform object detection tasks. They can accurately identify and locate various objects, such as faces, hands, and objects on a desk, allowing for real-time monitoring of the test-taker's actions.

2. Real-time Processing: Monitoring and analyzing the video feed in real-time is necessary for video proctoring. Because CNN and YOLO are both built to process frames rapidly and effectively, any questionable activity during the exam may be swiftly identified and addressed.

3. High Accuracy: CNN and YOLO models can recognise objects with a high degree of accuracy because they have been trained on big datasets. This is crucial for recognising particular behaviors or items that can contravene exam policies, such glancing at forbidden materials or utilizing forbidden equipment.

4. Flexibility: In particular, YOLO is renowned for its adaptability and capacity to detect numerous objects in a single pass. It is helpful for spotting multiple instances of cheating behavior or keeping tabs on multiple test-takers in a single video feed because it can detect and track multiple items concurrently.

5. Flexibility: Both CNN and YOLO are flexible to varied video proctoring settings and rule sets since they can be trained and improved on particular datasets. They can be programmed to recognise particular movements, gestures, or anything that are pertinent to the exam being proctored.

6. Integration with Existing Surveillance Systems: By integrating CNN and YOLO with current surveillance systems, it is possible to easily include video proctoring capabilities into an existing infrastructure. This integration enables easy deployment and scalability for large-scale video proctoring setups.

2. How are frames extracted in video processing?

In video processing, frames are extracted from a video to perform various operations, such as analysis, object detection, or image processing. The process of extracting frames involves the following steps:

1. Video Loading: A video processing library or framework is used to load the video file into memory. OpenCV (Open Source Computer Vision Library) and FFmpeg are two well-known libraries.

2. Frame Sampling: Frames can be extracted at predetermined intervals or in response to certain criteria, depending on the needs. The most popular method is to take screenshots at

regular intervals, like every second or a few frames. A representative sampling of the video frames is guaranteed by this technique.

3. Frame Decoding: The video decoder in the processing library decodes each frame into its raw pixel data after the frames have been chosen for extraction. This phase is required to extract the visual data from each frame.

4. Frame Processing: The extracted frames may be subjected to a variety of actions following decoding, including scaling, cropping, filtering, or feature extraction. Depending on the precise needs of the video processing activity, these activities are carried out utilizing image processing techniques or specialized algorithms.

5. Analysis or Output: Using computer vision algorithms, machine learning models, or other methods, the processed frames can be further analyzed. Any operation pertinent to the video processing task, including object detection, motion tracking, facial recognition, etc., may be included in this analysis. The transformed frames can also be delivered as output for visualization or stored to disk.