Abstract:

**Introduction**:

Identification cards are an easy but indispensable aspect of security in most settings like offices, schools, and restricted zones. Conventionally, verification of whether individuals have their ID cards has been done manually, which is time-consuming, erratic, and impractical for high-density or high-speed conditions. Given the increasing demand for automation and real-time surveillance, computer vision presents a robust alternative to human checks.

Current object detection techniques such as the YOLO family allow for real-time accurate and fast detections. The majority of current work, however, centers on detecting the ID card itself from images or merging ID card detection with other biometric techniques such as face recognition. Though effective in document scanning or multi-modal authentication, these approaches tend to necessitate additional processing steps like cropping or rectification, which complicate real-time deployment.Here, we approach the problem differently by posing the task as a dual-class detection problem. Rather than only detecting the card, our model detects individuals and classifies them into two groups: person with ID card and person without ID card. Employing a tailored YOLO model that has been trained on a specialized dataset, we propose to attain reliable and efficient classification under diverse real-world scenarios of illumination, pose, and crowd density. This methodology streamlines the detection pipeline, lessens processing time, and offers a viable basis for real-time surveillance systems.

**Reference**:

Recent object detection progress has dramatically enhanced computerized ID verification and surveillance. Much current research is aimed at identifying the ID card itself on an image. Kusuma et al. [1] designed an Identity Card Detection System based on YOLOv3 and Image Rectification, where the YOLOv3 model is utilized to detect the ID card and perspective transformation is used for rectifying the taken picture. This method has high detection precision but is mainly used for static, near-range ID card images, rather than detecting if someone is wearing one in real-world environments.

Several other works have tried fusing ID card detection with other biometric modalities. The YOLO-Powered Smart Entry System [2] integrates YOLO-based ID card detection with face recognition to construct secure smart entry systems. Although this enhances the accuracy of identity verification, it has greater computational complexity and relies on several subsystems to function in real time. Comparative research [3] have investigated dual-detection structures for security automation in which more than one YOLO model or conjoint detection tasks are employed to boost reliability. These contributions emphasize the benefit of multi-class or multi-modal detection but are mostly concerned with detecting various objects (e.g., face + ID) instead of classifying a single subject into dual categories.

Some studies target hardware-limited environments. For instance, a YOLOv2-based campus student identity identification system [4] shows that YOLOv2 can be designed for embedded systems to support real-time and lightweight student ID verification. Although these techniques are efficient, they are still based on the explicit detection of ID cards instead of classifying individuals by the presence of ID cards.

Lastly, initial research such as [5] investigated real-time human detection with YOLO for security purposes. These studies form the basis for identifying people correctly in diverse scenarios but do not cover ID card presence or compliance classification.

Different from these methods, our method is centered around dual-class detection, detecting people directly as person with ID card or person without ID card. This makes the detection pipeline simpler, eliminates rectification or secondary modalities, and allows for real-time surveillance using a single model trained to perform dual-class classification.

**Methodology:**

The suggested system is based on a YOLO-based object detection model trained specifically for identifying two classes: (i) person holding ID card and (ii) person not holding ID card. For robust performance in practical situations, a custom dataset of about 10,000 labeled images was created. The images were captured in different institutional and office settings using CCTV cameras and smartphone cameras, promoting view

point, background, and illumination variability.

The data was annotate carefully with bounding boxes for both target classes. The annotation strictly adhered to YOLO format standards to ensure consistency and match with the training pipeline. To enhance the model's ability to generalize, data augmentation processes were extensively applied, including horizontal and vertical flipping, random rotation, scaling, brightness, and contrast adjustments, as well as slight blurring. This was achieved to mimic various environmental conditions like change in camera angle, distance, and light.

Training was performed using transfer learning through fine-tuning a pre-trained YOLO backbone. The input size for the model was fixed at 640 × 640 pixels for trading off accuracy and real-time inference speed. The dataset was divided into 70% training, 20% validation, and 10% testing to systematically analyze performance during training.

The model was trained to 100 epochs with a batch size of 16, the Stochastic Gradient Descent (SGD) optimizer with the initial learning rate set to 0.01, momentum set to 0.937, and weight decay set to 0.0005. The loss function was a combination of three components:

Localization loss for bounding box regression,

Objectness loss for object presence detection, and

Classification loss for separation between the two classes.

Training was conducted on a workstation with GPU to speed up computation. During training, the primary metrics like precision, recall, and mean Average Precision (mAP) were tracked to assess the convergence and accuracy of the model. Early stopping and learning rate scheduling methods were used to avoid overfitting and to stabilize training.

The system to be proposed is set to identify automatically in real time persons with ID card and persons without ID card employing the dual-class YOLO-based detection model. There are three main components in the architecture: image acquisition, detection and classification, and output visualization.

Image Acquisition:

The input frames are obtained from surveillance cameras or webcams located in the monitoring location. The camera keeps streaming frames to the backend, where each frame is preprocessed and forwarded to the trained YOLO model for inference. This provides real-time monitoring without human intervention.

Detection and Classification:

The center of the system is a specially designed YOLO model that is trained using two classes: (1) person with ID card and (2) person without ID card. Each incoming frame is processed by the model and it does bounding box detection as well as class prediction. By casting the task as a dual-class detection task, the model concentrates on detecting the person and determining their ID card status in a single pass without requiring extra card detection or rectification steps. The model is real-time, thus appropriate for continuous surveillance environments.

Output and Visualization:

After detection, the processed frames are marked with bounding boxes and class labels, and finally presented on a web interface for live monitoring. Person without ID card detections can be automatically logged or flagged for follow-up action. Easy integration with existing camera networks and frontend dashboards is provided by the modular system design.