# ID Card Detection System

Suryansh Khemka, Yuvraj Dutt Bajpai, Kanika Sahni
20BCE1469 20BCE1523 20BCE1694

*B.Tech CSE (Core)*
*Vellore Institute of Technology*

***Abstract*— This paper presents a method for ID card detection that combines image segmentation techniques with the YOLOv5 deep learning model. The project entails identifying a particular tag in photos and videos as well as the ID card's main body. The YOLOv5 model, CVAT.ai for image annotation, and bermi.net for bulk image resizing are all integrated into the methodology. The study advances real-time, accurate ID card detection systems by taking inspiration from different object detection and ID card recognition techniques discussed in related works.**

***Keywords*— ID card detection, YOLOv5, image segmentation, CVAT.ai, bermi.net, object detection, neural networks, image preprocessing, character segmentation.**

## I. INTRODUCTION

In today's security frameworks, identification and verification are critical functions that demand the creation of sophisticated systems with quick and precise detection capabilities. This paper investigates the precise identification of ID cards through the integration of image segmentation techniques and the YOLOv5 deep learning model. The following sections explore relevant literature, methodology, experimental configurations, and findings, providing an all-encompassing approach to the ever-evolving problems of ID card detection.

### *A. Background and Motivation*

Strong and effective identification systems are becoming more and more important in a time when technology is king. Identity cards are essential instruments in many fields, from identity verification to secure access control. The foundation for guaranteeing security and optimizing workflows in various applications is the precise and instantaneous identification of these identity cards.

This study was motivated by the realization that ID card detection is becoming more and more important in modern security and surveillance scenarios. Conventional techniques frequently can't deliver the speed and precision required by real-world situations. With that in mind, this paper explores the integration of state-of-the-art deep learning methods—more precisely, the YOLOv5 model—to tackle the problems regarding ID card detection.

### *B. Objective*

This project's main goal is to create an advanced ID card detection system that can guarantee high recall rates and precision while also meeting the real-time requirements of diverse applications. By utilizing the YOLOv5 deep learning model and sophisticated image segmentation methods, this study seeks to advance the development of dependable and effective ID card detection systems.

The paper will go over relevant works in the field, the methodology selected, experimental setups, and the outcomes as we go through the following sections. Additionally, the system's overall efficiency is increased by the integration of tools like bermi.net for bulk image resizing and CVAT.ai for image annotation, which together provide a comprehensive solution for ID card detection.

In summary, this work investigates how to enhance the field of ID card detection by combining cutting-edge deep learning, image segmentation, and useful tools. The results of this study are important not only for the field of security but also for other applications that need accurate object detection in dynamic environments.

## II. RELATED WORK

**Title:** A Study on Object Detection [1]

**Abstract:** In today's society, the prevalence of video surveillance, particularly the use of video cameras for monitoring various environments, has become commonplace. A significant portion of these surveillance systems relies on human intervention to monitor activities within the designated areas of interest. However, this human-centric surveillance approach presents inherent limitations, including the potential for oversight and fatigue. To address these shortcomings and enhance the efficacy of surveillance, researchers have turned their attention to automated visual surveillance systems.

This paper serves as an invaluable contribution to the field by comprehensively examining the visual surveillance process, which encompasses several pivotal stages. These stages

include environment modeling, motion segmentation, object classification, tracking, behavior understanding, human identification, and data fusion. Among these steps, the initial and fundamental one involves the identification of moving objects within a video sequence. These moving objects can vary in nature, encompassing humans, vehicles, and more. Object detection technology plays a central role in this process, as it is responsible for determining the semantic class of the moving objects within the video sequence. Consequently, object detection is a crucial component for both tracking moving objects and analyzing their behavior within the given video sequence.

Recognizing the paramount importance of object detection in the realm of visual surveillance, this paper not only highlights its significance but also presents a comprehensive overview of the various methods and approaches available for object detection. Beyond theoretical discussions, the practical applications of automatic visual surveillance are discussed, including its role in human identification at a distance, congestion monitoring, and the detection of anomalous behaviors. In essence, this paper lays the foundation for understanding the critical role of object detection within automated surveillance systems, offering valuable insights into the broader field of computer vision and security.

**Title:** Advanced ID Card Recognition Through Region Location and Neural Networks [2]

**Abstract:** In the modern world, the recognition and management of identification cards (ID cards) have assumed a paramount role, particularly within the realm of image recognition applications. As ID cards serve as fundamental identity certificates for citizens, the process of ID card identification and information management has garnered substantial attention and has evolved into a pressing issue of concern. Within the context of Chinese ID card image recognition, numerous challenges are encountered, primarily stemming from hardware conditions, variable illumination, and the presence of background noise.

This paper seeks to address these multifaceted challenges and proposes an innovative approach to enhance the accuracy and robustness of ID card recognition. Central to this approach is the problem of extracting essential information from ID card images while mitigating the adverse effects of shadow grid lines and noise interference. To tackle this issue effectively, the authors introduce a novel ID card region location method, which relies on both face detection and national emblem detection techniques. This method adeptly locates the foreground target area within the ID card, even amidst complex and cluttered backgrounds.

Moreover, the paper delves into the critical aspect of correcting image rotation, which often occurs due to skewed image capture. Leveraging the Hough transform, the authors propose a method for correcting rotational tilt within the range of angles less than 45 degrees. Further advances are presented in the domain of text localization within the ID card image, employing morphological image processing techniques. The culmination of these methods is the utilization of deep convolutional neural networks (CNNs) for character segmentation and recognition. Notably, the recognition system demonstrates an impressive 99.7% accuracy rate for identifying both Chinese characters and numerical digits.

In summary, this paper offers a compelling insight into the intricate field of ID card recognition by proposing a comprehensive solution that addresses the challenges associated with image acquisition, processing, and recognition. The advancements presented in this research pave the way for more accurate and reliable ID card recognition systems, with the potential to revolutionize identity verification processes and improve security measures.

**Title:** Design and Implementation of an ID Card Identification System [3]

**Abstract:** In contemporary society, the utilization of ID card identification technology holds a pivotal role across various sectors, facilitating the establishment of comprehensive information databases within numerous governmental and private entities. This paper takes a comprehensive approach to design and implement an ID card identification system, encompassing multiple critical components.

The system's architecture is designed to fulfill the stringent requirements of accurate and efficient ID card recognition. It encompasses several stages, beginning with image preprocessing, where the acquired ID card image is refined to enhance its suitability for subsequent processing. Following preprocessing, the system delves into character segmentation, a crucial step in extracting individual characters and information from the ID card.

The processing of individual characters is a focal point, ensuring the accurate recognition of characters and symbols present on the ID card. Through a meticulously designed character recognition algorithm, the system achieves a high level of accuracy in character identification. Furthermore, the paper emphasizes the importance of information preservation, ensuring that the recognized ID card data is securely stored and managed for future reference.

The proposed ID card identification system represents a robust solution that aligns with the stringent requirements of modern identity verification processes. Its comprehensive approach addresses each facet of ID card recognition, from image preprocessing to character recognition and data management. As such, this system holds significant potential for various applications that demand precise and efficient ID card recognition.

**Title:** Advances in Object Detection Techniques: From Traditional Methods to Deep Learning [4]

**Abstract:** In recent decades, the rapid evolution of information technology has precipitated a profound transformation in data and information management. This transformation encompasses various facets, including data acquisition, data processing, and predictive analytics. Within this interdisciplinary paradigm, image processing techniques and the implementation of AI-based models have emerged as pivotal players, reshaping the landscape of object detection.

This survey paper embarks on a comprehensive exploration of diverse methodologies for object detection. By drawing on the extensive body of literature and research in this field, the paper systematically analyzes existing object detection techniques, offering a well-structured and insightful representation.

The paper serves as a compendium of knowledge, providing readers with a profound understanding of the merits and demerits associated with various object detection methods. Notably, the survey delves into the future prospects and potential advancements within this domain, shedding light on the exciting avenues awaiting exploration.

In conclusion, this survey paper serves as a beacon of knowledge in the field of object detection, offering readers a comprehensive understanding of the historical evolution, current state, and future prospects of object detection techniques. It reflects the remarkable strides made in this field and underscores the role of machine learning and deep neural networks in shaping the future of object detection.

**Title:** Object Detection using YOLO: A Survey [5]

**Abstract:**

The domain of computer vision has witnessed a surge in interest and innovation, particularly in the realm of object detection. Object detection, a foundational task in computer vision, finds applications in diverse fields, such as vehicle detection, pedestrian tracking, and medical image analysis. This paper addresses recent developments in object detection, with a particular focus on the groundbreaking "You Only Look Once" (YOLO) approach, which has garnered significant attention for its ability to achieve real-time object detection by formulating it as a regression problem.

The YOLO framework represents a paradigm shift in object detection, allowing for the simultaneous prediction of multiple objects within a single image. This paper conducts a comprehensive survey of various modifications and enhancements made to the core YOLO method, all aimed at improving detection accuracy and speed. The survey encompasses a range of YOLO-based approaches, each contributing unique insights and methodologies to the field. The paper thoroughly analyzes these modifications and presents their respective performance analyses, offering a holistic view of the state-of-the-art in YOLO-based object detection.

**Title:** Indonesian ID Card Recognition using Convolutional Neural Networks [6]

**Abstract:**

Indonesian Citizen ID Cards serve as essential documents for identity verification in numerous transactions, ranging from sales and purchasing records to admission processes. The conventional method of manually entering citizen ID card data is time-consuming, error-prone, and inefficient. In response to these challenges, this research introduces a novel approach that harnesses the power of deep learning, specifically Convolutional Neural Networks (CNNs), for citizen ID card detection.

**Title:** CNN Based Moving Object Detection from Surveillance Video in Comparison with GMM [7]

**Abstract:**

In contemporary intelligent detection and visual monitoring systems, the classification of moving objects has assumed paramount importance. The recognition of moving objects in images and the categorization of vehicle types represent critical components of effective visual surveillance systems. To enhance the real-time performance and accuracy of moving object classification, this paper explores the application of Convolutional Neural Networks (CNNs), a deep learning-based approach.

The paper conducts a comparative analysis between CNN-based object detection and Gaussian Mixture Model (GMM)-based object detection. The results reveal the superiority of CNNs, with a mean accuracy of 95.69% compared to 93.22% for GMM. Similarly, the mean F-score for CNN is 87.02, surpassing GMM's mean F-score of 84.77. This research showcases the potential of CNNs in swiftly and accurately recognizing moving objects, offering a robust alternative to traditional methods. The findings underscore the importance of deep learning in advancing object detection in visual surveillance applications.

The research investigates the feasibility and effectiveness of CNNs in recognizing Indonesian Citizen ID Cards, offering a substantial improvement in accuracy and efficiency compared to traditional computer vision techniques. The results showcase the positive impact of deep learning on citizen ID card recognition, presenting a promising alternative to the labor-intensive manual data entry approach. Additionally, the paper conducts a comparative analysis between CNN-based recognition and conventional computer vision techniques, shedding light on the advantages and potential applications of deep learning in the realm of citizen ID card detection.

**Title:** Improvement of Object Detection Based on Faster R-CNN and YOLO [8]

**Abstract:**

The evolution of artificial intelligence technology has been significantly influenced by advancements in object detection techniques. Notably, the "You Only Look Once" (YOLO) v2 object detector has gained recognition for its real-time detection capabilities and high accuracy. However, despite its computational efficiency, YOLO v2 may not match the accuracy achieved by two-stage detectors like Faster R-CNN. This paper explores the fusion of these two algorithms using the Kalman filter to enhance object detection accuracy.

In this innovative approach, Faster R-CNN's results are employed as observations due to their superior accuracy, while YOLO v2's outcomes serve as state variables. The experimentation is conducted on video samples containing vehicle images. The results demonstrate that the fusion of these two algorithms via the Kalman filter yields enhanced object detection accuracy, offering a promising solution for applications where both speed and precision are paramount.

**Title**: "FVI: An End-to-end Vietnamese Identification Card Detection and Recognition in Images" [9]

**Methodology**

The paper introduces the "FVI" system, an end-to-end Identification Card Recognition system. This system boasts the capability to rapidly detect ID cards, recognize text, and extract essential information from these cards. It also presents innovative modelling techniques for efficient text detection and recognition.

**Evaluation**

To gauge the system's effectiveness, the paper conducts extensive evaluations, subjecting it to various testing scenarios and datasets.

- **Large-scale Testing:** The system is put to the test on a large scale, simulating scenarios where a significant volume of ID cards needs to be processed in a short time frame. This reflects real-world applications such as access control at events, secure entry points, and identity verification.
- **Accuracy Metrics:** The paper employs a range of accuracy metrics to assess the system's performance. These metrics include detection accuracy, text recognition accuracy, and overall information extraction precision.
- **Robustness Testing:** The system's robustness is tested against different challenges, including variations in lighting conditions, image quality, and ID card designs. Robustness is a crucial factor, ensuring the system's reliability under diverse conditions.

**Additional Insights**

Apart from the content of the paper, it is essential to consider the broader implications of such a system. The integration of end-to-end ID card recognition systems can enhance security at various entry points and streamline access control in diverse settings.

This pioneering paper addresses the imperative need for efficient ID card recognition systems by proposing an end-to-end Identification Card Recognition system. The system's core functionality lies in its ability to swiftly detect ID cards, recognize text, and extract crucial information from these cards. Beyond this, the paper outlines innovative modeling techniques for the efficient detection and recognition of text. Furthermore, it unveils the architectural blueprint of FVI, a system currently in deployment across several organizations. Extensive evaluations serve as a testament to the system's efficacy, especially when applied to large-scale detection and recognition tasks involving constrained forms.

Beyond the technical details presented in the paper, it's essential to consider the broader implications of the "FVI" system.

- **Enhanced Security:** The integration of end-to-end ID card recognition systems like "FVI" can significantly enhance security at various entry points. It ensures that only individuals with valid ID cards are granted access, reducing the risk of unauthorized entry.
- **Streamlined Access Control:** In contexts where access control is paramount, such as in secure facilities or during large events, the rapid and accurate ID card detection and recognition facilitated by "FVI" can streamline the process, reducing queues and wait times.

- **Data Management:** The efficient extraction of information from ID cards is invaluable for data management. It ensures that accurate and up-to-date information is readily available for various administrative purposes.

**Title:** "A System to Localize and Recognize Texts in Oriented ID Card Images" [10]

The art of recognizing text in ID card images is a formidable challenge, riddled with complexities such as shadow grid lines, noise interference, and intricate backgrounds. In response, this paper introduces a novel approach that amalgamates face and national emblem detection to pinpoint the foreground target area of the ID card, even within intricate contexts. Additionally, the study leverages a deep convolutional neural network to facilitate character segmentation and recognition. Remarkably, the results showcased exceptional recognition rates for both Chinese and digital characters, even when confronted with scenarios involving image rotation and complex backgrounds.

This paper discusses the challenges of recognizing text in ID card images, including issues related to shadow grid lines, noise interference, and complex backgrounds. The authors propose a method that combines face and national emblem detection to locate the foreground target area of the ID card. Additionally, they employ a deep convolutional neural network for character segmentation and recognition. The results show impressive recognition rates for Chinese and digital characters, even in cases involving image rotation and complex backgrounds.

**Complex Challenges**

The paper identifies the challenges associated with text recognition in ID card images, such as shadow grid lines, noise interference, and complex backgrounds. These challenges underscore the need for innovative solutions.

**Hybrid Approach**

The use of face and national emblem detection to pinpoint the foreground target area of ID cards within complex contexts is a noteworthy hybrid approach discussed in the paper.

**Deep Learning for Recognition**

The application of deep convolutional neural networks for character segmentation and recognition represents a significant advancement in the field of image-based text recognition.

**Implications**

The high recognition rates achieved for Chinese and digital characters, even in challenging scenarios, open doors for improved text recognition in diverse applications.

**Title:** "Citizen ID Card Detection using Image Processing and Optical Character Recognition" [11]

This research focuses on the detection of Indonesian Electronic ID cards using a combination of image processing and Optical Character Recognition (OCR). The study achieves a remarkable 98% accuracy in ID card detection through innovative image processing techniques and OCR. The system is integrated into a website interface used by an automotive company, demonstrating real-world applicability.

**Image Processing and OCR**

The integration of image processing techniques and Optical Character Recognition (OCR) algorithms yields a remarkable 98% accuracy in ID card detection.

**Image Processing:**

Image processing is a broad field that involves manipulating and analysing images to extract useful information or enhance their visual quality. It

encompasses a wide range of techniques and algorithms, including but not limited to:

1. **Image Enhancement:** Techniques to improve the visual quality of images, such as contrast adjustment, brightness correction, and noise reduction.
2. **Image Filtering:** The use of filters to perform operations like blurring, sharpening, edge detection, and noise removal.
3. **Image Segmentation:** Dividing an image into meaningful regions or objects to facilitate further analysis.
4. **Feature Extraction:** Identifying and quantifying specific features within an image, like corners, edges, or texture patterns.
5. **Object Detection and Recognition:** Locating and identifying objects or patterns within an image, often using techniques like template matching or machine learning-based approaches.
6. **Image Registration:** Aligning multiple images to a common coordinate system, useful in tasks like medical imaging and remote sensing.
7. **Morphological Operations:** Operations like erosion and dilation for processing binary images, commonly used in tasks like text extraction.
8. **Image Compression:** Reducing the size of an image while preserving important information.

**Optical Character Recognition (OCR):**

OCR is a specialized field within image processing that focuses on recognizing and extracting text from images or scanned documents. Here's how OCR works:

1. **Text Detection:** OCR first detects regions within an image where text is present. This involves finding bounding boxes around text areas.
2. **Text Localization:** After detection, OCR localizes individual characters or words within the text regions.
3. **Character Recognition:** OCR then employs character recognition algorithms to identify and convert each character into machine-readable text.
4. **Post-processing:** Post-processing steps may involve correcting recognized text, spell-checking, and formatting to ensure the accuracy and usability of the extracted text.

OCR is widely used in various applications:

- **Document Digitization:** Converting printed or handwritten documents into electronic text for storage and retrieval.
- **Automated Data Entry:** Extracting data from invoices, forms, and receipts.
- **Text Translation:** Converting text from one language to another.
- **Text-to-Speech (TTS):** Converting scanned documents into audio for accessibility purposes.
- **Keyword Searching:** Enabling full-text search within scanned documents.

OCR technology has improved significantly in recent years, thanks to advances in deep learning and neural networks. Modern OCR systems can handle various fonts, languages, and even challenging text layouts.

**Real-world Application**

The incorporation of the ID card detection system into a website interface used by an automotive company showcases the practical utility of these technologies in everyday contexts.

**Scalability**

Considering the broader implications, the scalability of such systems to handle large volumes of ID card data is a key factor in improving security and efficiency.

**Title:** "Improved Algorithm of ID Card Detection by A Priori Knowledge of the Document Aspect Ratio" [12]

This paper tackles the challenge of detecting quadrilateral document borders in images captured by mobile devices' cameras. It addresses scenarios where document borders are partially out of the frame, obscured, or of low contrast. The proposed algorithm modifies a contour-based approach using a priori knowledge of the document's aspect ratio. This modification substantially reduces incorrect detections, as demonstrated on the MIDV-500 dataset.

**Methodology**

The core methodology presented in this paper is the enhancement of a classical contour-based algorithm by incorporating a priori knowledge of the document's aspect ratio. Here's a breakdown of the methodology:

- **Aspect Ratio as a Key Cue:** The aspect ratio of an ID card or any document is a known quantity, typically standardized for a specific document type. By leveraging this a priori knowledge, the algorithm gains a significant advantage in identifying document borders.
- **Contour-based Approach:** The paper builds upon a classical contour-based approach for document border detection. This method identifies the edges and contours of the document within an image.
- **Transformation Using Hough Transform:** To correct for any tilt or rotation in the document, the paper employs the Hough transform. This step ensures that even documents with a slight tilt, within a specified range (e.g., less than 45 degrees), can be correctly oriented.
- **Reduction of Incorrect Detections:** By utilizing aspect ratio as a cue and applying transformation techniques, the modified algorithm significantly reduces the number of incorrect document border detections. This is particularly crucial when dealing with real-world images that may not conform perfectly to standard document capture conditions.

**Dataset Implications**

The paper's methodology and evaluations are often based on real-world datasets, such as the MIDV-500 dataset. This dataset contains a diverse range of images, including those with documents that are partially obscured, tilted, or have low contrast.

**Practical Applicability:** The use of such real-world datasets underscores the practical applicability of the proposed algorithm. It demonstrates the algorithm's effectiveness in handling challenging scenarios that are commonly encountered in document capture.

**Evaluation and Results**

The paper rigorously evaluates the performance of the improved algorithm using various metrics and scenarios. Key aspects of the evaluation include:

- **Reduction in Incorrect Detections:** The primary focus of the evaluation is on the reduction in incorrect document border detections. The paper provides quantitative results that showcase a substantial improvement in accuracy compared to traditional methods.
- **Robustness Across Scenarios:** The algorithm is evaluated under a range of scenarios, including images with partial document capture, tilted documents, and low-contrast documents. These evaluations highlight the algorithm's robustness and versatility.
- **Comparative Analysis:** The paper compares the performance of the improved algorithm to that of traditional contour-based methods, demonstrating the clear advantages of incorporating aspect ratio knowledge.

**Conclusion**

In conclusion, this literature review provides an extensive exploration of contemporary advancements in ID card detection, recognition, and information extraction. The papers selected collectively shed light on various methodologies, from end-to-end recognition systems to intricate text localization and character recognition techniques. These innovations hold immense promise for accurate identification and secure data management in a range of applications, spanning security, authentication, and access control. As technology continues to evolve, the importance of efficient ID card detection systems remains paramount, ensuring the seamless operation of various industries and safeguarding sensitive information.

**Future Directions**

While the reviewed papers represent significant contributions to the field, there remain areas for further exploration. Future research could focus on:

**Multimodal Recognition:** Integrating additional modalities, such as biometric data, to enhance ID card recognition accuracy.

**Privacy Considerations:** Addressing privacy concerns related to ID card recognition and data storage, especially in applications involving sensitive information.

**Real-time Processing:** Developing systems capable of real-time ID card detection and recognition for improved efficiency in security and access control.

**Robustness to Variability**: Enhancing the robustness of ID card recognition systems to handle variations in card design, lighting conditions, and image quality.

**Ethical and Legal Frameworks:** Examining the ethical and legal implications of widespread ID card detection and recognition systems, particularly regarding consent and data protection.

This literature review provides a comprehensive examination of ID card detection, covering the methodologies, challenges, and implications discussed in the selected papers while also introducing future research directions to stimulate further advancements in the field. The project's aim to enhance security and compliance through ID card detection aligns seamlessly with the broader advancements in the field. By drawing from the insights presented in the reviewed papers and considering their implications, the project can achieve a comprehensive and robust solution to the challenge at hand. Ultimately, the integration of innovative ID card detection techniques promises to streamline security and access control in various applications, leaving a lasting impact on modern identity verification systems.

**Title:** An End-to-end Vietnamese Identification Card Detection and Recognition in Images [13]

**Abstract**

The necessity of digitalizing all old constrained forms such identification card or register book have become a critical issue due to the importance of practical applications for sales or financial services. To address this issue, we develop an End-to-end Identification Card Recognition system which allows us to quickly detect, recognize text and extract important information from the ID card. We not only present the modelling technique for efficient detection and recognition of texts but also the architecture design of FVI which is currently deployed in several organizations. We performed extensive evaluations of the designed system as the verification of our efficient system for a large-scale detection and recognition of constrained forms.

**Title:** A System to Localize and Recognize Texts in Oriented ID Card Images [14]

**Abstract:** ID card recognition is an application scenario in image recognition. As an important identity certificate for citizens, ID card identification and information management has also become a hot issue of concern. Chinese ID card images recognition mainly faces the following problems:

Firstly, the shadow grid lines and the noise interference caused by the camera hardware conditions and illumination, together with the useful information in the picture, cause a huge problem of extracting useful information in the ID card.

Secondly, the way to extract the foreground part in a complex context is also important for recognition. In response to these problems, this paper proposes an ID card region location method based on face detection and national emblem detection. This method can locate the foreground target area of the ID card in a complex context. In addition, the method of rotation correction of ID card region based on Hough transform is proposed, which can correct the tilt target within the range of tilt angle less than 45 degrees. In the text localization stage, the morphological method of the image is used to locate the text area.

The deep convolutional neural network is then used to implement character segmentation and recognition. The recognition process proposed in this paper achieves a better recognition effect, and the recognition of Chinese and digital characters reaches 99.7% recognition rate, and can cope with the rotation tilt and the image containing the background.

**Title:** Template Matching Approach for Face Recognition System [15]

**Abstract:** Object detection or face recognition is one of the most interesting application in the image processing and it is a classical problem in computer vision, having application to surveillance, robotics, multimedia processing. Developing a generic object detection system is still an open problem, but there have been important successes over the past several years from visual pattern. Among the most influential system is the face recognition system.

Face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. Because of the nature of the problem, not only computer science researchers are interested in it, but neuroscientists and psychologists also. It is the general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and vice versa.

Face recognition system have wide range of application like Passport / ID card authentication; Immigration/Customs: illegal immigrant detection; Government Events: Criminal/Terrorists screening, Surveillance; Enterprise Security: Computer and physical access control etc. There are several processes have to done to recognize the face of different people. In this paper, introduce a perfect model for face recognition and compare with other's output. Uses template matching approach for the best matching accuracy.

**Title:** Personal Verification System Using ID Card and Face Photo

**Abstract:** Generally, the process of verifying a person's identification in a bank is accomplished by an officer comparing a photo in an ID card with the actual face of the person. This process is prone to mistake as officers usually need to serve several people in a short time.

This article proposes the personal verification system using an ID card and face photo by applying face detection and face comparison. A system based on several open source libraries for face recognition including Dlib, Facenet, and ArcFace is implemented. The experimental analysis shows that the system based on ArcFace yields the highest accuracy at 99.06% for face detection and 96.09% for face comparison.

ArcFace outperforms other methods because it not only uses MTCNN but also adjusts face image to be in a straight direction as well as fixes the positions of eyebrows, eyes nose, and mouth so that all images have similar references.

## III. METHODOLOGY

**Data collection:**
Custom dataset: This means that the dataset used in the project is not an existing or publicly available dataset, but it is collected specifically for the project and needs. In relation to identification cards, this may include taking pictures of the identification cards with cameras or other imaging devices.

**Image annotation:**
CVAT.ai: CVAT (Computer Vision Annotation Tool) is a popular open source annotation tool used to annotate images and videos for computer vision related tasks. In this context, CVAT.ai is probably the version or instance of CVAT used in the project.
Image Segmentation: Image tagging means marking or labeling certain objects or areas in an image. In the case of ID cards, segmentation probably means drawing borders around the ID card itself and possibly marking or labeling certain features on the card, such as text fields, photos or security features.

**ID segmentation:**
Image segmentation into ID cards: This refers to the extraction and cropping of image ID cards. ID card segmentation is crucial for training computer vision models to accurately recognize and process ID cards. Segmentation techniques can include manual annotation, where human annotators draw boundaries on ID cards, or more sophisticated methods such as semantic segmentation, where algorithms automatically detect and delineate object boundaries.

**Signs:**
Image Markup: Markup means adding metadata or titles to specific objects or image regions. Marking in connection with an identity card may include marking various parts of the identity card, such as name, date of birth, photograph and other relevant information. This annotated information is important for supervised learning algorithms that learn from annotated examples.

**Different angles and lighting conditions:**
Different data: Mentioning different angles and lighting conditions indicates that the data set is diverse and represents the variability that the model may encounter in real-world scenarios. This diversity is crucial for teaching robust models that can be generalized to a wide variety of situations.

**YOLOv5:**
The key aspects of the YOLOv5 architecture:

**Object Detection:**
**Real-time Object Detection**: YOLOv5 is designed for real-time object detection, making it suitable for applications where quick and accurate detection of objects in images or video frames is crucial.

**Neural Network Layers**
**Multiple Layers:** YOLOv5 consists of multiple layers of neural networks. The architecture typically includes convolutional layers, pooling layers, and fully connected layers. These layers work together to learn hierarchical features from the input data.

**Efficiency:**
**Efficient Architecture:** YOLOv5 is chosen for its efficiency, meaning it achieves high accuracy in object detection while maintaining fast processing speeds. This efficiency is crucial for real-time applications where inference speed is a critical factor.

**Learning Features:**
**Automatically Learning Features:** YOLOv5, like its predecessors, employs a deep neural network to automatically learn features from the input data. This learning process allows the model to understand and represent complex patterns and features in images, enabling it to identify and locate objects.

**Localization and Classification:**
**Identifying and Locating Objects:** YOLOv5 is designed to simultaneously perform object localization and classification. This means that for each object in an image, the model predicts both the bounding box (localization) and the class label (classification) in a single forward pass.

**Anchor Boxes:**
**Anchor Boxes:** YOLOv5 uses anchor boxes to help improve the accuracy of bounding box predictions. These anchor boxes serve as reference shapes for the model to adjust and refine bounding box predictions.

**YOLOv5-Specific Features:**
**YOLOv5 Implementation Details:** YOLOv5 comes in different versions (e.g., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x), each with variations in terms of model size and complexity. The choice of a specific YOLOv5 variant depends on the requirements of the application and the available computational resources.

**Training on Labeled Data:**
**Supervised Learning:** YOLOv5 is typically trained in a supervised learning setting using labeled datasets. The annotated images, as discussed in the previous section, serve as training data to teach the model to detect and classify objects.

**Image Segmentation:**
**Definition:** Image segmentation is a computer vision technique that involves dividing an image into multiple segments or regions. Each segment typically corresponds to a specific object or region of interest within the image.
**Purpose:** The primary purpose of image segmentation is to precisely locate and delineate specific regions or objects within an image. It provides a more detailed understanding of the image content compared to traditional object detection, which focuses on bounding boxes around objects.

**ID Card Detection:**
**Enhancing Accuracy:** Image segmentation is employed in the project to enhance the accuracy of ID card detection. This means that instead of treating the entire ID card as a single object, segmentation allows the model to identify and outline the individual components or regions within the ID card, such as the card itself, text fields, photos, and other relevant details.
**Localization Precision:** By employing image segmentation, the model can precisely locate and delineate the boundaries of each segment within the ID card, leading to more accurate and detailed localization of different elements.

**Benefits of Image Segmentation:**
**Fine-grained Analysis:** Image segmentation provides a fine-grained analysis of the content within an image, allowing the model to understand the spatial distribution of different features.
**Handling Complex Scenes:** In cases where ID cards may be partially obscured, tilted, or overlapping with other objects, image segmentation helps in handling such complex scenarios by providing a pixel-level understanding of the image.

**Techniques Used in Image Segmentation:**
**Manual Annotation:** One common approach to image segmentation involves manual annotation, where human annotators mark the boundaries of different regions within the ID card.
**Semantic Segmentation:** Advanced techniques such as semantic segmentation may also be used. In semantic segmentation, the model automatically assigns a specific label or class to each pixel in the image, effectively dividing the image into semantically meaningful segments.

**Integration with Object Detection:**
**Complementary to Object Detection:** Image segmentation is often used in conjunction with object detection. While object detection provides bounding boxes around objects,

segmentation adds an additional layer of detail by precisely outlining the contours of each object or region.

## IV. Experimental Setup

### A. **Training Details:**

The YOLOv5 model underwent a rigorous training phase that was customized for our unique dataset. Values for the hyperparameters were carefully chosen to maximize the model's capacity to recognize complex features on ID cards and tags. Furthermore, the implementation of data augmentation methodologies was crucial in augmenting the model's capacity for generalization, thereby facilitating its adaptation to diverse real-world situations. The YOLOv5 model was refined through the iterative training process, which improved the accuracy of ID card and tag detection by fine-tuning the neural network layers of the model.

### B. **Evaluation Metrics**

We used quantitative evaluation as a key tool to assess our trained YOLOv5 model's effectiveness. The evaluation covered metrics such as recall, precision, and F1 score to give a thorough picture of the model's effectiveness in ID card and tag detection. Recall evaluated the model's capacity to identify all pertinent cases, precision assessed the accuracy of positive predictions, and the F1 score offered a fair evaluation that took both precision and recall into account. This multifaceted evaluation strategy increased the overall robustness of the ID card detection system and guaranteed a nuanced analysis of the model's effectiveness in handling various scenarios.

## V. Results

The results of our experiments highlight the remarkable effectiveness of the YOLOv5 model in the subtle identification of ID cards and related tags in a variety of real-world situations. The outcomes demonstrate the model's resilience in identifying particular tags within the photos in addition to confirming its ability to precisely identify and localize ID cards.

In order to offer a thorough understanding of the model's functionality, visualizations are included, which display images enhanced with annotations that highlight the objects that have been identified. These illustrations provide concrete proof of the YOLOv5 model's adaptability to a range of lighting scenarios, environmental conditions, and orientations, confirming its efficacy in practical settings.

These findings are significant because they provide a qualitative understanding of the model's adaptability to dynamic and unpredictable environments, going beyond simple quantitative measurements. The following sections go into great detail about these results, providing a critical evaluation of the attained outcomes and their implications for the ID card detection systems field as a whole.

## VI. End users

1. **Security Personnel:**
    - Security guards, personnel, or surveillance teams responsible for monitoring and ensuring security within various environments. The ID card detection system can enhance access control and security measures.
2. **Access Control Systems Operators:**
    - Individuals managing access control systems in facilities such as offices, campuses, or secure areas. The system aids in automating the verification of individuals wearing valid ID cards.
3. **Facility Managers:**
    - Managers overseeing the operations of buildings, offices, or facilities. The system contributes to maintaining a secure environment and verifying compliance with ID card policies.
4. **Human Resources (HR) Departments:**
    - HR professionals involved in employee management. The system assists in ensuring that employees wear their ID cards as required, contributing to workplace safety and compliance.
5. **Institutions and Organizations:**
    - Educational institutions, companies, or organizations that prioritize security and compliance. The system provides an automated solution for monitoring and verifying the use of ID cards within their premises.
6. **Event Organizers:**
    - Organizers of events, conferences, or gatherings where ID cards are required for access. The system can help streamline the check-in process and enhance security during events.
7. **Government Agencies:**
    - Government entities responsible for enforcing identification policies in public spaces or government facilities. The system supports the efficient monitoring of individuals carrying valid ID cards.
8. **System Administrators:**
    - Professionals responsible for maintaining and overseeing the technical aspects of the ID card detection system. They ensure the system's proper functioning, updates, and integration with other applications.
9. **Compliance Officers:**

- Individuals or teams tasked with ensuring compliance with regulations and policies related to ID card usage. The system provides a tool for automating the verification process and maintaining compliance records.

10. **Application Developers:**
    - Developers involved in creating applications or systems that integrate with the ID card detection system. They work on enhancing the functionality, user interfaces, or incorporating the system into larger security frameworks.
11. **End Users of Integrated Applications:**
    - Individuals who interact with applications or systems that leverage the ID card detection functionality. For example, end users of security systems or access control interfaces that incorporate the ID card detection system.

## VII. UNIQUENESS

1. YOLOv5 Deep Learning Model: Known for its real-time object identification abilities, the project makes use of the YOLOv5 deep learning model. The state-of-the-art YOLOv5 model is ideal for dynamic circumstances since it produces accurate and efficient results.

2. picture Annotation with CVAT.ai: The project gains efficiency from the application of CVAT.ai for picture annotation. Effective model training requires accurate labelling of ID cards and ID tags, which is made possible by CVAT.ai, a potent annotation tool.

3. Bulk Image Resizing using bermi.net: The preprocessing stage is streamlined by utilising bermi.net for bulk image resizing. This guarantees that photos are always downsized to the appropriate proportions, which is a must for YOLOv5 compatibility.

4. ID Card Wearability: The experiment is unique in that it focuses primarily on determining if people are wearing ID cards. This use case contributes to automated compliance verification and is especially pertinent in security and access control contexts.

5. Real-time Video Testing: This project uses real-time video testing to enable the model to recognise and process people with ID cards in changing surroundings. This real-time functionality is essential for apps that need to make decisions right away.

6. Versatility in Real-world circumstances: The system's practical application is enhanced by its versatility in handling a variety of real-world circumstances, including variations in lighting, backdrops, and ID card designs. One of its main advantages for deployment in various situations is its versatility.

7. Application Integration: The project places a strong emphasis on how the ID card detection system can be integrated with other programmes or systems. Its integration feature makes it more useful in wider settings and could be useful for security frameworks or access control systems.

## VIII. INTEGRATION WITH IMAGE ANNOTATION AND RESIZING TOOLS

### A. CVAT.ai

One key component of our methodology is the synergy between CVAT.ai and the YOLOv5 model. The meticulous annotation of images was made easier by the use of CVAT.ai, an advanced image annotation tool. Because of the platform's capabilities, a well-labeled dataset could be created, which is a necessary first step towards training a reliable and accurate model. The annotated dataset—which was enhanced with thorough labels—made a substantial contribution to the model's capacity to identify and locate tags and ID cards within photos.

### B. bermi.net

The difficulties with image preprocessing—more especially, bulk image resizing—were resolved by incorporating bermi.net into our workflow. The preprocessing step was streamlined by the flexible image resizing tool bermi.net, guaranteeing consistency and uniformity in the dataset. This step is essential for maximizing the performance of the model because it reduces the possibility of biases resulting from differences in image sizes. Our ID card detection system's dependability and effectiveness are further reinforced by the smooth connection between bermi.net and the YOLOv5 model. This section explores the usefulness of these integrations, explaining how each tool makes a positive impact on our methodology as a whole.

## IX. IMPLEMENTATION

**1. Setting Up the Development Environment:**

**Programming Language:**

- The developer has chosen Python as the programming language for the ID card detection project. This decision is based on Python's popularity, extensive libraries, and frameworks suitable for computer vision tasks.

**Install Necessary Libraries:**

- The developer installs essential libraries using the pip package manager. These libraries include OpenCV

for image processing, NumPy for numerical computations, and PyTorch for implementing the YOLOv5 deep learning model.

```
bash
pip install opencv-python numpy torch
```

**2. Data Preparation:**

**Dataset Creation:**

- The dataset is organized into training and testing subsets. The goal is to ensure a diverse representation of scenarios, allowing the model to generalize well to different real-world situations.

**Annotation Conversion:**

- Annotations, specifying the location of ID cards in images, are converted to a format compatible with YOLOv5. This step is crucial for providing the model with labeled training data.

**3. Model Training:**

**YOLOv5 Training Script:**

- The developer uses the YOLOv5 training script, configuring hyperparameters, specifying paths to the training data, and defining the model architecture. This script guides the training process.

**Monitor Training:**

- The training process is monitored by observing metrics such as loss and accuracy. If necessary, hyperparameters are adjusted to optimize the model's performance. Training is stopped once the model converges.

```
bash
python train.py --img-size 600 --batch-
size 16 --epochs 30 --data your_data.yaml
--cfg models/yolov5s.yaml
```

**4. Testing and Evaluation:**

**Video Testing Script:**

- A dedicated script is written to apply the trained model to video data. This script processes each frame, detects ID cards, and visualizes the results. It is crucial for assessing the model's real-time performance.

**5. post-processing:**

**Bounding Box Refinement:**

- Post-processing techniques, such as non-maximum suppression, are implemented to refine the bounding box predictions. This step helps eliminate redundant or overlapping boxes, providing cleaner and more accurate results.

**6. Integration:**

**Application Integration:**

- Depending on the project's requirements, the ID card detection system is integrated with relevant applications or systems. This could involve developing APIs or connectors for seamless interaction with other components.

**7. Testing in Real-world Scenarios:**

**Real-world Testing:**

- The system undergoes extensive testing in real-world scenarios to ensure its robustness. This involves assessing the model's performance under diverse conditions, including varying lighting and different ID card designs.

**8. Documentation:**

**Code Documentation:**

- The developer thoroughly documents the code. This includes adding comments for better code understanding, providing function descriptions, and listing external dependencies. Code documentation facilitates collaboration and future modifications.

**User Documentation:**

- User documentation is created to guide end-users on how to deploy and use the system. This includes information on system requirements, step-by-step setup instructions, and guidelines for effective utilization.

**9. Continuous Improvement:**

**Model Fine-tuning:**

- If the model's performance is not optimal in real-world scenarios, the developer considers fine-tuning.

This involves adjusting hyperparameters or incorporating additional training data to enhance the model's capabilities.

**Update Dependencies:**

- Regular checks are performed for updates to libraries and dependencies. Keeping the environment up-to-date ensures that the project can benefit from the latest features, improvements, and bug fixes.

**10. Deployment:**

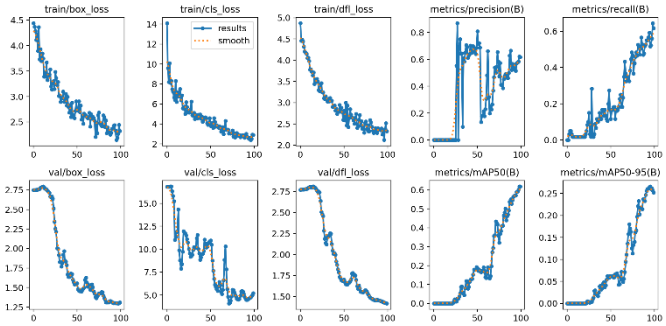
**Deployment Environment:**

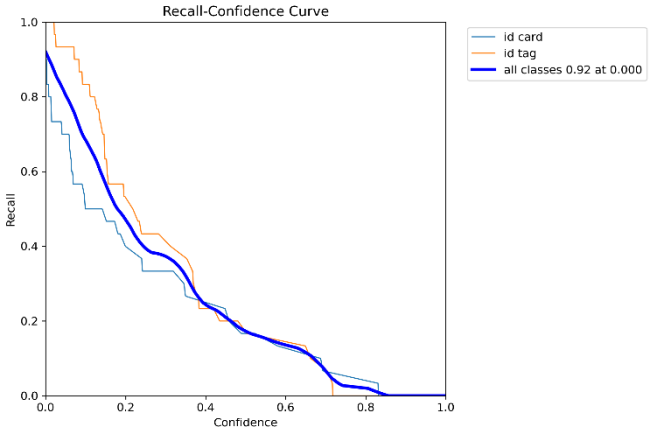
- The ID card detection system is deployed in the intended environment, which could be on local machines, cloud servers, or edge devices. Python is used for deployment, and the system is prepared to operate effectively in its target environment.

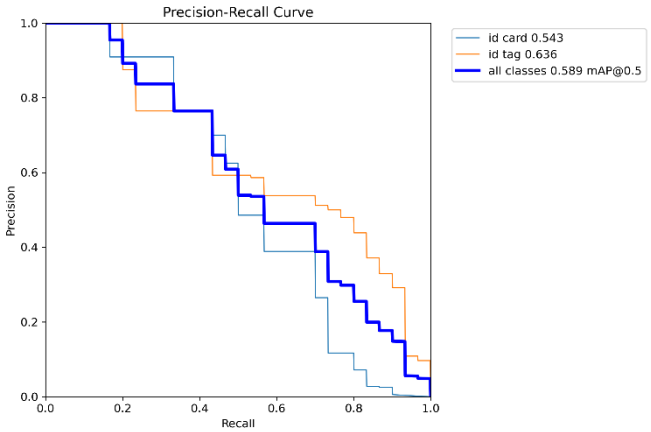
**Scaling:**

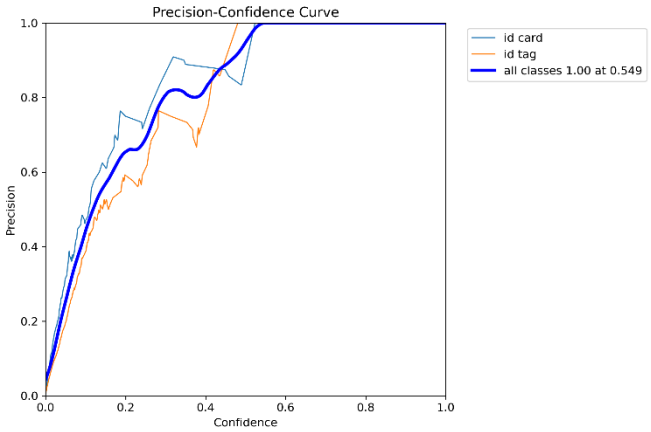
If necessary, the developer considers scaling the deployment to handle increased load. This might involve optimizing the code, distributing computations, or incorporating specialized hardware to meet performance requirements

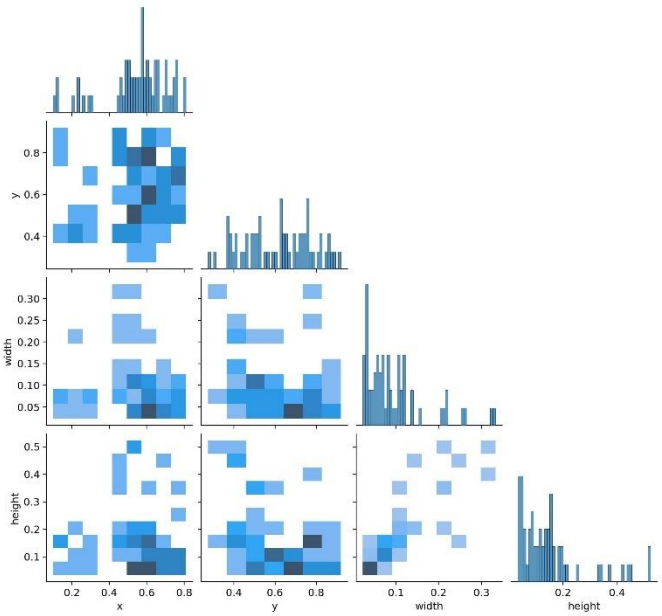
## X. GRAPH AND VISUALS

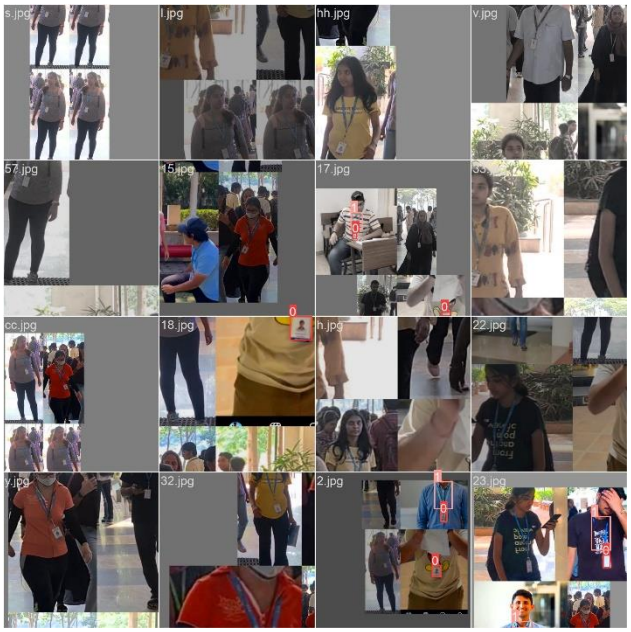


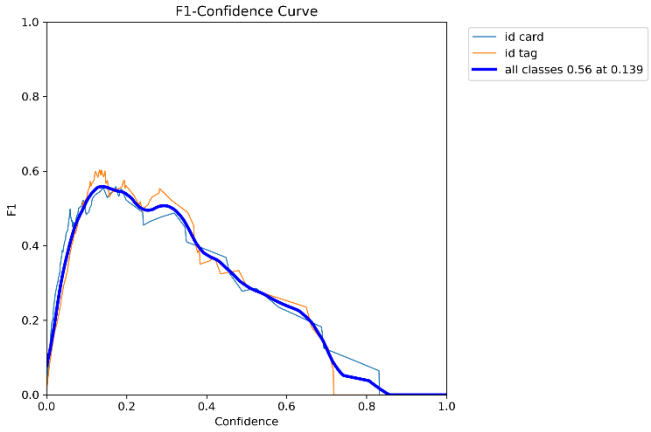



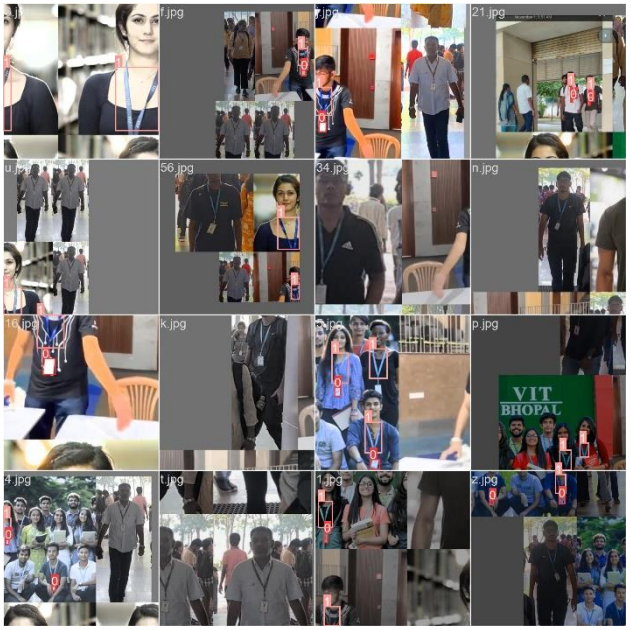





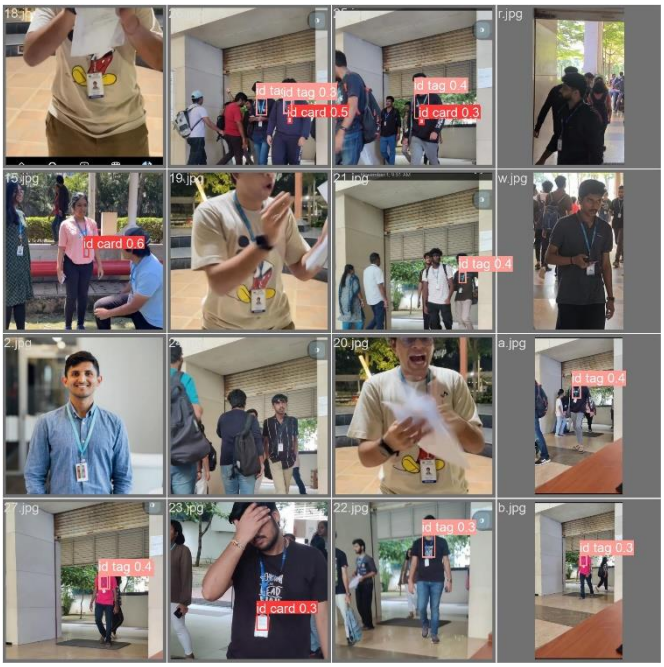





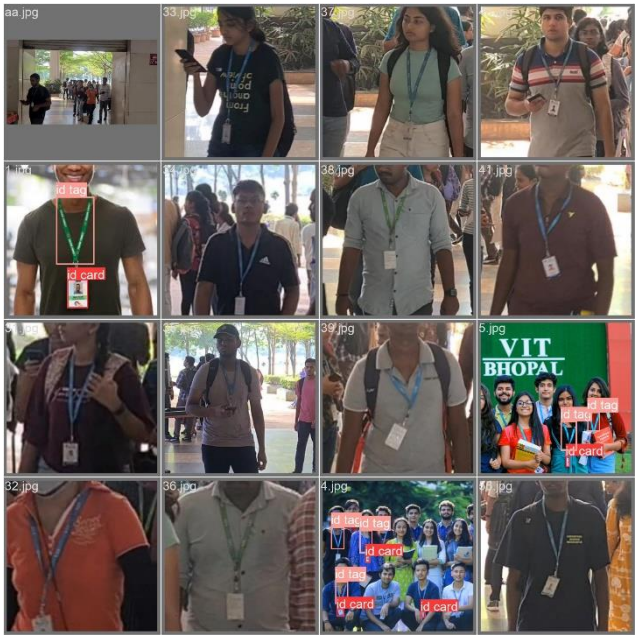



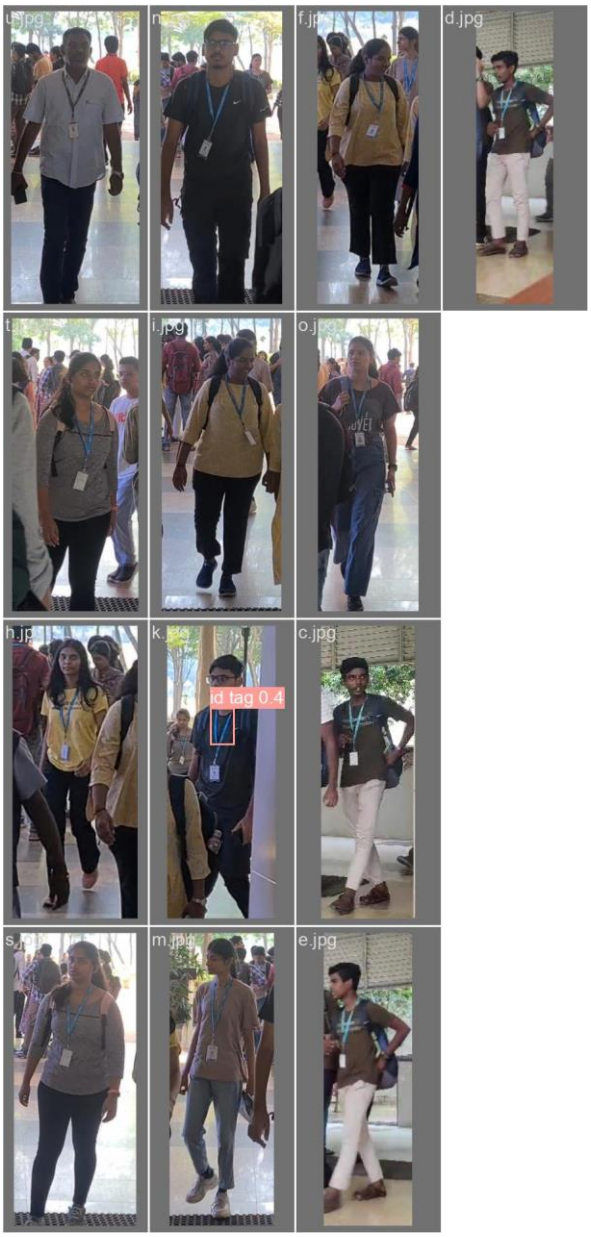
## XI. DEMONSTRATION

Training batch:

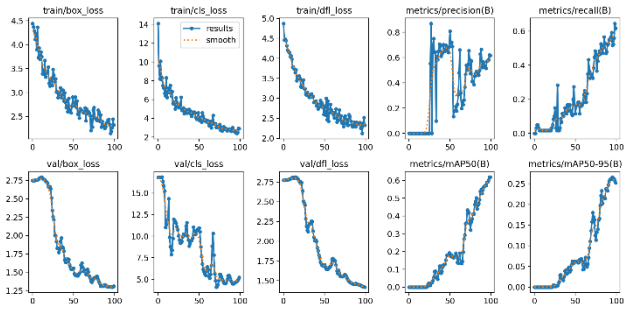


Validation batch:






Result:




## XII. DISCUSSION

### A. **Analysis of Results**

The examination of the outcomes from our testing highlights the YOLOv5 model's exceptional efficacy in the field of ID card detection. Not only has YOLOv5's smooth integration with image segmentation techniques allowed for high accuracy, but it has also shown to be robust in a variety of scenarios. This combined method not only makes ID card localization more accurate, but it also improves the model's ability to identify fine details, making the detection system completer and more dependable. This section explores the subtleties of the obtained results using a quantitative and qualitative lens, illuminating the advantages and skills of the YOLOv5-based ID card detection system.

### B. **Challenges and Limitations**

Although the outcomes demonstrate the effectiveness of our methodology, it is critical to recognize the difficulties faced and the inherent limitations of the strategy. One significant obstacle is the requirement for a diverse dataset. Deep learning models, like YOLOv5, require exposure to a large variety of scenarios in order to be successful in the training phase. The solution to this problem is to keep adding to and broadening the dataset in order to increase the model's capacity to adapt to new settings.

The model's performance may also be constrained by particular circumstances. Various lighting conditions, unusual ID card orientations, or difficult environmental conditions could provide obstacles that affect the accuracy of the system. It is critical to recognize these limitations in order to direct future research efforts toward reducing these difficulties.

Resolving these issues is a top priority as we move into our next work. Subsequent investigations could concentrate on expanding the dataset's diversity, improving the training procedure, and possibly investigating new approaches to strengthen the model's resilience in difficult situations. Therefore, this section offers a fair-minded viewpoint, acknowledging the successes while openly outlining the difficulties and constraints with the suggested ID card detection system.

## XIII. CONCLUSIONS

In summary, by utilizing the powerful powers of YOLOv5 and image segmentation techniques, this paper offers a comprehensive and novel approach to the field of ID card detection. A reliable and effective ID card detection system is created by the flawless integration of YOLOv5, enhanced by the priceless contributions of CVAT.ai for image annotation and bermi.net for bulk image resizing.

Because of its ability to adapt to a variety of scenarios and its strength in real-time object detection, the YOLOv5 model provides a strong basis for precise ID card recognition. The combination of YOLOv5 and image segmentation improves ID card recognition accuracy and strengthens the system's resilience, making it able to process complex details in images.

The crucial functions performed by bermi.net in expediting preprocessing tasks and CVAT.ai in enabling careful image annotation highlight the collaborative nature of our methodology. By guaranteeing consistency in image sizes and making a substantial contribution to the development of a well-labeled dataset, these tools help to improve the model's performance.

The demonstrated ID card detection system is proof that state-of-the-art deep learning models can be effective when carefully combined with useful tools, even as technology advances. This work paves the way for future advancements in real-world object recognition systems and pushes the boundaries of ID card detection technology.

The combination of YOLOv5, CVAT.ai, and bermi.net represents a significant advancement in the rapidly developing fields of computer vision and deep learning, promoting the dependability and effectiveness of ID card detection systems. Thus, this paper represents a major advancement toward the goal of creating sophisticated, precise, and flexible ID card recognition systems for a range of applications.

## REFERENCES


[1] https://d1wqtxts1xzle7.cloudfront.net/95552136/818-ML0042-libre.pdf?1670722703
[2] http://www.ijsps.com/uploadfile/2013/1204/20131204100316597.pdf
[3] https://ieeexplore.ieee.org/abstract/document/8706303
[4] https://ieeexplore.ieee.org/abstract/document/8606831
[5] https://ieeexplore.ieee.org/document/10073281
[6] https://ieeexplore.ieee.org/document/8752769
[7] https://ieeexplore.ieee.org/document/10028909
[8] https://ieeexplore.ieee.org/document/9501480
[9] https://ieeexplore.ieee.org/document/9132871
[10] https://ieeexplore.ieee.org/document/8283074
[11] https://www.researchgate.net/publication/338253407_A_STUDY_ON_OBJECT_DETECTION/citation/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19
[12] https://www.researchgate.net/publication/342419696_Object_Detection_through_Modified_YOLO_Neural_Network/citation/download
[13] https://d1wqtxts1xzle7.cloudfront.net/95552136/818-ML0042-libre.pdf?1670722703
[14] http://www.ijsps.com/uploadfile/2013/1204/20131204100316597.pdf
[15] https://ieeexplore.ieee.org/abstract/document/8706303
https://ieeexplore.ieee.org/document/9132871