

Efficient Signature Localization and Detection from Authenticated ID Cards

Thejashwini B L¹, H S Nagendraswamy¹, Lakshmiprasad K and Shivani P R

¹Department of Studies in Computer Science, University of Mysore, Mysuru, India, 0009-0005-7273-8754

¹Department of Studies in Computer Science, University of Mysore, Mysuru, India, 0009-0001-2614-6169

Abstract. In this digital era, the verification of a person's identity based on authenticated ID cards plays an important role. The verification can be done in many ways; out of that, the signature present on the documents is used for personal authenticity. The main aim of this research article is to localize and detect signatures from various types of authenticated ID cards, such as PAN cards, driving licenses, college IDs, and so on. The deep learning models were used, such as YOLO variants (YOLOv5, YOLOv8, YOLOv9, YOLOv10, YOLOv11), DETR (Detection Transformer), RT-DETR (Real-time Detection Transformer), and Detectron2. The proposed model detects the signature from newly collected datasets, some images sourced from the MASK-RCNN-Dataset and the Roboflow platform. The model was tested with our newly collected datasets without any preprocessing. To measure the effectiveness of the models, we have used the most common metric in detection phase as, mean average precision (mAP50 and mAP50-95). All the models detect signatures with above 96% accuracy. This article also compares all the models and observes that the RT-DETR model gives efficient results. Specifically, using the RT-DETR_r101vd_coco_0365 backbone model, RT-DETR achieved a detection accuracy of 100% with a pre-processed dataset and also observed that all models obtained an IoU ranging from 92% to 97%.

Keywords: Object Detection, Computer Vision, YOLO, Detection Transformer.

1 Introduction

In recent times, modern technology has disrupted various aspects of daily life, including education, the economy, and transportation, owing to its unique intricacies and advancements. Among the many methods of verifying personal identity, a person's handwritten signature remains one of the most widely used and trusted forms. Traditionally, signature verification involves human adjudicators who manually compare a given signature against the original one on file. However, this process is both time-consuming and expensive. As a result, there is a growing need for more cost-effective and scalable methods that can automate the detection and verification of signatures on identification documents with high precision.

To address these issues, this work proposes a comprehensive approach for signature detection on identity documents using a variety of state-of-the-art deep learning models. Specifically, the proposed framework employs advanced single-stage, two-stage, and transformer-based detection architectures. For single-stage detection, we utilize the real-time object detection algorithm You Only Look Once (YOLO), including its recent versions 8, 9, 10, and 11. For two-stage detection, we integrate the Detectron2 open-source library developed by Facebook AI Research (FAIR), exploring various backbone networks. Furthermore, to incorporate the benefits of attention mechanisms and global feature learning, we adopt transformer-based architectures such as the DETECTION TRANSFORMER (DETR) and its real-time variant, Real-Time DETECTION TRANSFORMER (RT-DETR), both implemented with different backbone configurations. We evaluated with a paired t-test between YOLOv8 and RT-DETR models on detection accuracy across test folds, resulting in a p-value < 0.05, confirming significant performance improvement of

the model. These models are rigorously evaluated for their ability to accurately detect handwritten signatures on diverse and challenging ID document images.

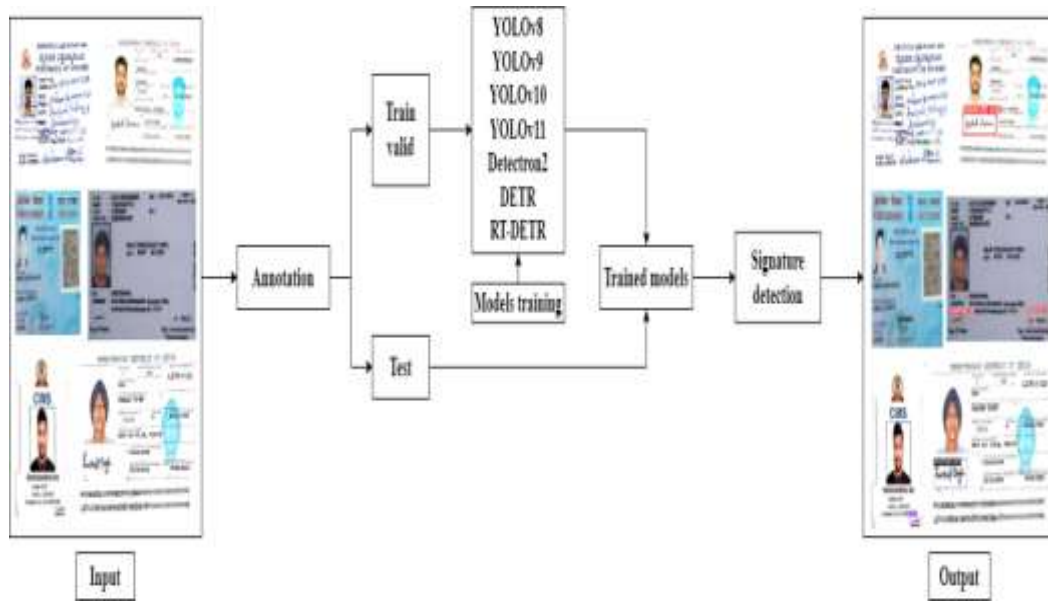
The remaining paper is organized as follows: Section II presents the Related Works; in Section III, we describe the Proposed Methodology; Section IV presents the Experimental Results and finally, in Section V we present the conclusions obtained from this work.

2 Related works

Recent research shows various methods for ID card recognition, signature detection, and object localization using deep learning. BJ et al. [1] used YOLOv5s to identify multiple elements, such as names, photos, and unique numbers from 1,000 ID proofs, achieving 94.6% accuracy. Li et al. [2] created a CNN-based OCR with GoogLeNet and YOLOv3 for 13,070 Chinese characters, reaching 99.39% accuracy through transfer learning. For handwritten signatures, Silva et al. [3] modified U-Net for segmentation. Junior et al. [4] enhanced it with an encoder-decoder and refinement layers for better precision on DSSigDataset-2. Zuo et al. [5] applied SVM-based frameworks, achieving 94% accuracy. Satyawani et al. [6] reached 98% accuracy in ID card detection using OCR. Rusli et al. [7] combined OCR with NLP for Indonesian IDs, obtaining an F-score of 0.78. Zhu et al. [8] used CNN on 55,000 images, achieving 99% accuracy. Van Hoai et al. [9] proposed CRNN for Vietnamese ID cards, achieving 96.7% character-level accuracy. Zhang et al. [10] used YOLOv5 with Copy-Paste augmentation to detect handwritten signatures. It did better than Faster R-CNN and DETR. Das Neves et al. [11] improved document segmentation with an extended U-Net. Rezkiani et al. [12] applied YOLOv4 for diploma logo detection and achieved 93.73% accuracy. Saha et al. [13] improved graphical element localization using transfer learning across benchmark datasets. Yan et al. [14] tested forgery detection methods with YOLOv2/v5 and Faster R-CNN, finding YOLOv5 at 84%. Sohan et al. [15] reviewed advancements in YOLOv8. Carion et al. [16] introduced DETR for transformer-based detection. Zhao et al. [17] enhanced real-time detection using hybrid encoders. Thejashwini et al. [18] combined YOLO with GANs for reliable signature detection. Tall et al. [19] used YOLOv11 to detect fish species in real time. This supports sustainability.

3 Proposed methodology

The proposed work describes the signature detection task performed by using single-stage deep models, such as four different versions of the YOLO model, namely YOLOv8 [16], YOLOv9 [17], YOLOv10 [18], and YOLOv11 [19], along with Detectron2, a two-stage deep model, and transformer-based architecture DETR (DEtection TRansformer) and its real-time variant RT-DETR (Real-Time DEtection TRansformer) [11]. This proposed work guarantees a thorough assessment of each model's accuracy and robustness in signature detection. Fig. 1 provides an overview of the proposed methodology.

Fig. 1. Architecture of the proposed work

3.1 Dataset

The dataset is crucial for addressing the challenges of signature detection and localization. The dataset used for our proposed work contains 300 images, among them 125 real-time ID card images that were manually collected from various individuals that include handwritten and digital college ID cards, as well as multiple government ID proofs such as PAN cards, driving licenses, vehicle RCs, and voter IDs collected from friends, relatives, and 25 images of MASK-RCNN-Dataset and the remaining 150 images sourced from the Roboflow platform. The dataset was annotated by labeling tool with rectangular bounding box and considered its coordinate values for training. Table I shows dataset information and Fig. 2 shows different dataset samples.

Table I. Dataset description

ID type	No. Of images
MASK-RCNN-Dataset[25]	25
Images sourced from the Roboflow platform[24]	150
Handwritten ID(college ID cards)	30
Digital ID (employee ID cards)	33
PAN card	29
Driving license	22
Vehicle RC	3
Voter ID	8
Total images	300

Fig.2. Samples from Custom Dataset

3.2 YOLOv8

YOLOv8 [16] is designed for computer vision tasks involving real-time object detection. It is an improved version of the previous YOLO models. It is based on a convolutional neural network, and is divided into two main components: the backbone and the head. It uses a variant of the CSPDarknet53 architecture as the backbone, which includes 53 convolutional layers with cross-stage partial connections. The head consists of multiple convolutional and fully connected layers that predict bounding boxes, confidence scores, and class probabilities. It incorporates a self-attention mechanism in its network head and uses a feature pyramid network to identify objects across various scales. This allows the model to concentrate on various parts of an image and effectively identify objects of different sizes. The speed of the model is crucial for applications requiring rapid responses, such as driverless cars and monitoring systems.

3.3 YOLOv9

YOLOv9's [17] real-time object detection capabilities continue the YOLO series tradition by quickly processing images or video streams to accurately identify objects without compromising speed. It incorporates the concept of Programmable Gradient Information (PGI) to ensure that deep features retain the necessary information for target tasks. This approach prevents information loss during input processing by using a reversible auxiliary branch that helps maintain a stable gradient flow. It is optimized for parameter efficiency, computational complexity, accuracy, and inference speed through the use of the Generalized ELAN (GELAN) architecture. The architecture allows users to choose the best computing components for different types of devices, providing more flexibility and efficiency. ResNet-101 was considered because of its capacity to extract detailed spatial information from document images, which is necessary for signature region recognition. GELAN, a lightweight backbone, was used to assess performance tradeoffs in computationally limited for less complex situations. Results show that YOLOv9 outperforms other real-time object detection algorithms in accuracy, speed, and overall performance on benchmark dataset, such as MS COCO. It is designed for adaptability, so it can be easily integrated into many systems, making it suitable for applications like surveillance, self-driving cars, and robotics.

3.4 YOLOv10

YOLOv10 is the newest version of the YOLO (You Only Look Once) series. It is faster and more accurate than previous versions and introduces several new features. One of the best changes is eliminating the requirement for Non-Maximum Suppression (NMS) during inference, which simplifies and speeds up the detection process. It reduces inference latency and enhances the alignment between training and inference by using a dual assignment approach during training. It reduces inference latency and enhances the alignment between training and inference by using a dual assignment approach during training. It provides a systematic model design technique that balances accuracy with efficiency. To reduce computing costs, it uses depth-wise separable convolutions for its classification heads. The model also utilizes a technique known as spatial channel decoupled down-sampling.

3.5 YOLOv11

Ultralytics latest model in the YOLO series is YOLOv11, which provides advancements in real-time object detection. With an improved backbone and neck architecture, it improves feature extraction and leads to more accurate object detection. It provides faster training pipelines for faster processing without losing accuracy, and it is optimized for speed and efficiency. YOLOv11 is more efficient and obtaining a higher mAP on the COCO benchmark while utilizing 22% fewer parameters than YOLOv8m. Its flexibility makes it suited for use on cloud servers, edge devices, and systems with NVIDIA GPUs. It also supports a wide range of tasks, such as oriented object detection (OBB), image classification, pose estimation, object detection, and instance segmentation.

3.6 Detectron2

Detectron2 is an open-source library created by Facebook AI Research for advanced object detection and segmentation tasks. It is an extensible and efficient framework based on Pytorch. The SDK includes models such as Faster R-CNN, Mask R-CNN, RetinaNet, and various ResNet architectures, enabling capabilities for object detection, instance segmentation, keypoint detection, and panoptic segmentation. There are great performance enhancing features in Detectron2 to use in research as well as for industry. So, numerous computer vision tasks use and incorporate this toolkit. Detectron2 has extensive applications in industries, ranging from autonomous driving to medical image analysis, and provides powerful visualization tools to examine model predictions. It integrates well with PyTorch to provide efficient training and inference of models, harnessing GPU acceleration to deliver high performance.

3.7 DETR (DEtection TRansformer)

Facebook AI Research's new object detection method [11], DETR, removes the need for complex post-processing steps like region proposals, anchor generation, and non-maximum suppression (NMS) by utilizing transformers for end-to-end object detection. It directly predicts bounding boxes and class labels using a set-based global loss that enforces one-to-one predictions through bipartite matching. Its architecture consists of three components: a feed-forward neural network (FFN) head, a transformer encoder-decoder network, and a CNN backbone. The transformer encoder processes the input image after it passes through the backbone—typically ResNet-50 or ResNet-101—to extract features and capture global relationships among all pixels. The transformer decoder uses learnable positional embeddings (object queries) and applies multi-head self-attention and cross-attention mechanisms to ensure the model focuses on the

relevant parts of the image. This enables DETR to predict varying numbers of objects without relying on fixed anchor boxes or predefined sizes.

3.8 RT-DETR (Real-Time Detection Transformer)

RT-DETR is an improved version of DETR, optimized for real-time object detection. It preserves the transformer design while including improvements for real-time applications. Leaning on a CNN backbone such as ResNet, it uses feature maps that are fed through a transformer's self-attention encoder to develop global context. the decoder, through cross-attention, predicts object locations and class labels using learnable object queries. RT-DETR has maintained DETR's cut-free detection through the addition of sparse attention, which improves efficiency by reducing computational complexity. Furthermore, it has innovative feature fusion and advanced loss functions that ensure superior classification and localization precision. The use of additional techniques enhances inference speed, enabling real-time processing capabilities.

4 Results

Results of YOLO models: The proposed versions of YOLO models, including YOLOv8, YOLOv9, YOLOv10, and YOLOv11, were trained on 100 epochs. YOLOv8 detects signatures accurately in IDs, but in some negligible cases, it shows multiple bounding boxes for a single signature and detects other objects as signatures. Table 2 depicts the performance evaluation metrics of the proposed YOLO models. Fig. 3, 4, 5, and 6 show the experimental results of detecting signatures from various IDs from versions v8, v9, v10, and v11.

Table 2 .PERFORMANCE EVALUATION RESULTS OF PROPOSED YOLO MODELS.

Model	mAP50	mAP50-95	Accuracy
YOLOv8	0.994	0.778	93.33
YOLOv9	0.995	0.827	98.33
YOLOv10	0.995	0.768	93.33
YOLOv11	0.994	0.772	96.66

Fig. 3. Sample results from signature detection using YOLOv8



Fig. 4. Sample results from signature detection using YOLOv9



Fig.5. Sample results from signature detection using YOLOv10**Fig. 6. Sample results from signature detection using YOLOv11**

Results of Detectron2: Detectron2, known for its robust object detection capabilities, was trained for 1500 iterations using the same dataset proportion as the YOLO models. To ensure a thorough evaluation of the model, we experimented with different Faster R-CNN backbones of the model. The model detects signatures using all the different backbones, but similar to the YOLOv8 model, the Detectron2 model with different backbones shows multiple bounding boxes for a single signature. Table 3 shows the performance evaluation metrics for the Detectron2 model. Fig. 7 describe the experiment results of detecting signatures from various IDs from the Detectron2 model with different backbones.

Table3. PERFORMANCE EVALUATION RESULTS OF DETECTRON2 MODEL

Backbone	AP 50	mAP50-95	Accuracy
Faster_RCNN_R_50_C4_1x	0.896	0.591	98.33
Faster_RCNN_R_50_C4_3x	0.889	0.615	93.33
Faster_RCNN_R_50_FPN_1x	0.905	0.647	96.66
Faster_RCNN_R_50_FPN_3x	0.917	0.658	95.00
Faster_RCNN_R_50_DC5_1x	0.938	0.688	98.33
Faster_RCNN_R_50_DC5_3x	0.944	0.692	95.00
Faster_RCNN_R_101_C4_3x	0.925	0.599	98.33
Faster_RCNN_R_101_DC5_3x	0.919	0.634	98.33
Faster_RCNN_R_101_FPN_3x	0.916	0.642	95.00
Faster_RCNN_X_101_32x8d_FPN_3x	0.929	0.652	98.33

Fig. 7. Sample results of signature detection by Detectron2

Results of DETR: DETR, known for its transformer-based architecture in object detection, was trained for 50 epochs using the same dataset proportion as the YOLO models and Detectron2 model. We thoroughly evaluated the effectiveness of the DETR model by testing its two backbones. The model successfully detects signatures with various backbone configurations. Among the two backbones, the DETRmodel detects signatures more accurately with a detr-resnet-101 backbone. In some IDs, the model with both backbones detects other objects as signatures similar to the YOLO models. Below Table 4 provides the performance evaluation metrics of the DETR model. Fig. 8 describe the experiment results of detecting signatures from various IDs from the DETR model with different backbones.

Table4. PERFORMANCE EVALUATION RESULTS OF DETR MODEL

Backbone	mAP50	mAP50-95	Accuracy
detr-resnet-50	0.859	0.602	75.00
detr-resnet-101	0.938	0.609	86.66

Fig.8. Sample results of signature detection by DETR

Results of RT-DETR: The RT-DETR model was trained for 50 epochs using the same dataset split proportions as the DETR model. It is known for its efficient transformer architecture designed for real-time object detection and its ability to recognize signatures. This ability was verified using both backbone setups as seen in Fig. 9 and summarized in Table 5. The model performed notoriously well and was able to label signature detections, especially with the rt- detr_r101vd_coco_o365 backbone. Nonetheless, both backbones of the model exhibit a characteristic structurally like DETR where non-signature objects are misclassified in certain IDs as signature objects. To fully assess the model, we implemented two different configurations for the rest of the backbone detection at level 2 (8 class detection) riddled throughout selected configurations for the RT-DETR model.

Table 5. PERFORMANCE EVALUATION RESULTS OF RT-DETR MODEL

Backbone	mAP50	mAP50-95	Accuracy
rt-detr_r50vd_coco_o365	0.98	0.73	70.00
rt-detr_r101vd_coco_o365	0.84	0.64	98.33

Fig .9. Sample results of signature detection by RT-DETR

RT-DETR shows higher mAP (0.98) but lower accuracy (70.00%) due to better localization yet more false positives, while DETR achieves balanced mAP and higher accuracy by reducing false detections. This trade-off highlights RT-DETR's strength in accurate localization and DETR's advantage highlights in overall detection reliability.

5 Conclusion

The assessment evaluated several models of object detection such as YOLO, Detectron2, DETR and RT-DETR in the ability to detect and locate signatures on various types of ID cards. The set comprised 300 images with various flavored ID card samples to help evaluate the models performance in real life situations. The best performer was YOLOv9 which had the highest mean Average Precision (mAP) and accuracy in signature detection. Detectron2 performed well with many different backbones but, unlike other cases with many bounding boxes on single signatures, failed. The other models such as DETR, and RT- DETR, while not as good as YOLOv9, the results were achieved excellently. This framework could support automated KYC and document verification systems in government and finance. The work notes that greater diversity within the models coupled with dataset refinement are certainly required towards enhancing how robust such models are to false positives and general ID card formats and conditions.

References

1. BJ, B. N., & Jose, M. (2023, February). Multiple Object Recognition from Smart Document Images using YOLOv5s. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 824-828). IEEE.
2. Li, Y. Q., Chang, H. S., & Lin, D. T. (2022). Large-scale printed Chinese character recognition for ID cards using deep learning and few samples transfer learning. Applied Sciences, 12(2), 907.
3. Silva, P. G., Junior, C. A., Lima, E. B., Bezerra, B. L., & Zanchettin, C. (2019, September). Speeding-up the handwritten signature segmentation process through an optimized fully convolutional neural network. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 1417-1423). IEEE.

4. Junior, C. A., da Silva, M. H. M., Bezerra, B. L. D., Fernandes, B. J. T., & Impedovo, D. (2020, July). Fcn+ rl: A fully convolutional network followed by refinement layers to offline handwritten signature segmentation. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.
5. Zuo, L., Chen, W., Qu, H., Huang, L., Wang, Z., & Chen, Y. (2019). An intelligent knowledge extraction framework for recognizing identification information from real-world ID card images. *IEEE Access*, 7, 165448-165457.
6. Satyawan, W., Pratama, M. O., Jannati, R., Muhammad, G., Fajar, B., Hamzah, H., ... & Kristian, K. (2019, June). Citizen ID card detection using image processing and optical character recognition. In *Journal of Physics: Conference Series* (Vol. 1235, No. 1, p. 012049). IOP Publishing.
7. Rusli, F. M., Adhiguna, K. A., & Irawan, H. (2021, August). Indonesian id card extractor using optical character recognition and natural language post-processing. In 2021 9th International Conference on Information and Communication Technology (ICoICT) (pp. 621-626). IEEE.
8. Zhu, J., Ma, H., Feng, J., & Dai, L. (2018, April). ID card number detection algorithm based on convolutional neural network. In *AIP Conference Proceedings* (Vol. 1955, No. 1). AIP Publishing.
9. Van Hoai, D. P., Duong, H. T., & Hoang, V. T. (2021). Text recognition for Vietnamese identity card based on deep features network. *International Journal on Document Analysis and Recognition (IJDAR)*, 24(1), 123-131.
10. Zhang, Y., Zhang, J., Yan, K., Wang, H., & Wang, G. (2023). Handwritten Chinese signature detection with simple Copy-Paste augmentation on power plants technical documents. *Service Oriented Computing and Applications*, 17(4), 293-302.
11. das Neves, R. B., Verçosa, L. F., Macêdo, D., Bezerra, B. L. D., & Zanchettin, C. (2020, July). A fast fully octave convolutional neural network for document image segmentation. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). IEEE.
12. Rezkiani, K., & Nurtanio, I. (2022, November). Logo Detection Using You Only Look Once (YOLO) Method. In 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS) (pp. 29-33). IEEE.
13. Saha, R., Mondal, A., & Jawahar, C. V. (2019, September). Graphical object detection in document images. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 51-58). IEEE.
14. Yan, K., Zhang, Y., Tang, H., Ren, C., Zhang, J., Wang, G., & Wang, H. (2022). Signature detection, restoration, and verification: A novel chinese document signature forgery detection benchmark. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5163-5172).
15. Sohan, M., Sai Ram, T., Reddy, R., & Venkata, C. (2024). A review on yolov8 and its advancements. In *International Conference on Data Intelligence and Cognitive Informatics* (pp. 529-545). Springer, Singapore.
16. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213-229). Cham: Springer International Publishing.
17. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., & Chen, J. (2024). Detsr beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16965-16974).
18. Thejashwini, B. L., Nagendraswamy, H. S., Rakshitha, P., & Manjuchethan, N. (2024). Integrating YOLO and GAN for robust signature detection and similarity assessment. *International Journal of Research and Analysis in Science and Engineering*, 4(5), Sept 2024. ISSN: 2582-8118.

19. M. M. Tall, I. Ngom, O. Sadio, I. Diagne, A. Coulibaly and M. Ndiaye, "Detection of Fish Species Captured by Artisanal Canoes Using YOLOv11," 2024 International Conference on Intelligent Computing and Next Generation Networks (ICNGN), Bangkok, Thailand, 2024, pp. 1-5, doi: 10.1109/ICNGN63705.2024.10871297.