Electronic Letters on Computer Vision and Image Analysis 25(1):83-97, 2025

Optimized Detection of Urdu Signatures in Real-World Images Using YOLO v7

Muzammal Hussain* and Muhammad Ahsan Rafiq+

* *Department of Computer Science, Government College University Faisalabad, Sahiwal Campus, Sahiwal, Pakistan*
+ *Department of Physics and Robotics, Punjab Group of Colleges, Faisalabad Campus, Faisalabad, Pakistan*

Received 8 July 2025; accepted 6th of March 2026

Abstract

Authentication plays an important role in managing security. Signature is one of the first broadly practiced method to authenticate an individual. However, existing research is solely based on the English signature detection and recognition with limited work on low-resource languages. Although being desirable for the document forensics and security purposes, it remains a challenging task to detect signatures in Urdu script in realistic settings due to different styles of the signatures and presence of noise, background, and other nuisance factors. Moreover, the lack of annotated datasets hindered the signature detection in natural environments. To address these challenges, this paper proposes an UrduSig dataset consisting of more than 5000 official and real-world scanned documents of different genres, i.e., publicly available official government letters, feedback given by the public in relation to various departments, and civil court orders of the Sahiwal region. The You Only Look Once (YOLO v7) model was employed for the signature detection task. The findings show that the proposed YOLO v7 is effective and accurate in detection even under different lighting conditions, background complexities, and distortions. The YOLOv7 model achieved the highest (mean Average Precision) mAP@0.5:0.95 rate of 0.975.

Key Words: Urdu signatures (UrduSig), Handwritten signature detection, Signature verification, YOLOv7, Object detection, Deep learning, Computer vision

1 Introduction

Signatures play a crucial role in legal documents for identifying and verifying individuals [1]. Recognized as a widely used biometric identity verification method in banking, administration, and financial institutions, signatures are valued for their simplicity and unique characteristics. However, the surge in signed documents has led to an increase in crimes related to forgery, presenting a significant security challenge for organizations like insurance companies, banks, and government entities [2]. Urdu, often known as Standard Urdu (Mayari Urdu), is an Indo-Aryan language that is spoken in South Asia. It has been recognized as the second official language along with English in Pakistan, and it is the medium of communication in that country [3]. Urdu is one of the eight scheduled languages of India.

Correspondence to: muzammal_hussain@outlook.com

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2267
ELCVIA ISSN:15108-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

Nonetheless, because the Urdu language has unique features, there are some problems with language processing that can be encountered. In response to this challenge, there is the need for an automated and reliable system for detection of signatures in Urdu script. While digital identity verification has emerged as a new option, many official documents are still manually signed on paper, necessitating the crucial task of digitization [4][5]. This shift from historical documents to digital formats poses a major challenge in document management and knowledge extraction [6]. With the expedited progress of deep learning methods, there's an opportunity to enhance signature detection and recognition through data-driven approaches. Notably, signature detection is considered a vital prerequisite to ensure the accuracy of subsequent stages in this evolving landscape.

Depending on the type of signature acquisition, there are two systems developed for the purpose of automatic signature recognition: online and offline [7]. The online system, currently known as the dynamic method, assumes the use of electronic devices, such as tablets or pressure-sensitive pens, to obtain signatures. The dynamic mode of operation gives comprehensive information pertaining to time in pressure and position of the pen in the creation of the signature. However, the data collection aspect of online signatures is expensive in terms of resources. On the other hand, offline signatures are captured via scanning that transforms the signatures into a bitmap [8]. Acquiring offline signature data is simple, but its detection is difficult because of the lack of time information, fewer features, and a noisy environment. The offline signature detection in Urdu text is also affected by the absence of prior information on the position of signatures and variations in the scanners used with different resolutions of scan. This difference poses some uniqueness to the offline signature system, where it becomes a challenge to extract signatures from the scanned images. Therefore, this paper is particularly dedicated to the offline signature detection challenge in Urdu script.

Extensive work has been done in the general area of object detection [9][10][11], but the field of signature detection, sealed even more within the context of signatures in Urdu script, is rather ambiguous and has been less explored. However, the pre-existing work largely concerns signatures in Latin datasets [9][10][11][12][13] while leaving out the non-Latin scripts such as Arabic, Persian, and Chinese domains. This research gap means that the identification of non-Latin signatures remains a problem, thus requiring more research to be undertaken. Furthermore, offline signature detection, which aims at detecting signature positions in the images of the scanned documents, has different challenges distinct from those encountered while performing object detection [14]. Firstly, there are limited training datasets available freely to the public and privacy issues, which makes it a really big challenge, let alone the scenario of Urdu document datasets. Secondly, unlike English, signatures in Urdu script are often complex combinations of shapes and forms, which have no correlating names, making detection complex. Moreover, incorporated handwritten signatures present certain issues, as the signatures can overlap with logos, seals, dates, or other handwritten messages within the document. In cases like banking documents, there may be multiple signatures placed in close vicinity of each other, and hence it will not be easy for models to detect such eminent features.

To overcome these challenges of detection of signatures in Urdu script, this study presents the major contribution to the field:

- Developed a real-world dataset of scanned documents containing signatures in Urdu script, addressing privacy considerations and capturing variability in official Urdu documents.

- Investigated the effectiveness of YOLO v7 for detecting signatures in Urdu script, considering the unique challenges posed by the complex Urdu script.

The structure of this paper is as follows: Section 2 presents a comprehensive review of signature detection in Urdu script, with particular emphasis on existing datasets relevant to this domain. Section 3 describes the proposed methodology for detection of signatures in Urdu script, including the technical approach and model architecture. Section 4 provides an in-depth evaluation of the experimental results. Section 5 concludes the paper with future implications.

2 Related Work

The detection of signatures is frequently approached as an object detection task, leveraging the swift progress in deep learning and the increasing enthusiasm for object detection. This allows the utilization of sophisticated techniques developed in the realm of object detection to enhance the detection of signatures in Urdu script. The authors in a study [15], tested the Faster Region-based Convolutional Neural Network (Faster R-CNN), Detection Transformer (DETR), and YOLOv3 [16] models for detecting offline signature forgery in Chinese documents using the ChiSig [17] benchmark. Their findings revealed that Faster R-CNN surpassed alternative detection models regarding effectiveness. In another study, the authors of [18] applied Faster R-CNN and YOLOv2 models to the Tobacco 800 database, comprising scanned documents with signatures and annotated logos. Additionally, the authors of another study [19] explored the effectiveness of Faster R-CNN, YOLOv5, and RetinaNet models in urban construction applications in Vienna, revealing that YOLOv5 [20] exhibited better performance in comparison to alternative models. The authors of [21] explored three models for object detection: DETR, YOLOv5, and Faster R-CNN, each representing different approaches. Their research showed that YOLOv5, especially when enhanced with Copy–Paste augmentation, demonstrated superior performance compared to the other models.

The Tobacco800 dataset [22] is widely used for signature detection, but it is limited by its English language exclusivity and a small size of 1,290 images. Detection of signatures in Urdu script faces challenges as a result of the limited availability of publicly accessible datasets, primarily in English, hindering progress in this field. Currently, three datasets for signatures are available: UTSig (University of Tehran Signature) dataset [23], PHBC (Persian Handwritten Bank Checks) dataset [24] and FUM (Ferdowsi University of Mashhad) signature dataset [25]. Table 1 provides more information about offline signature datasets. Challenges persist, including privacy concerns limiting access to datasets and a severe shortage of datasets for official Urdu documents. To help address these constraints, a special collection of Urdu documents was gathered in response. To the best of knowledge, there is no work done on detection of signatures in Urdu script, and there is no dataset available publicly.

Database	**Number of Signatures**	**Genuine Signers**
UTSig [17]	8280	115
PHBC [18]	1200	100
FUM [19]	600	20

Table 1: Available offline datasets in Persian and Arabic signatures

3 Material and Methods

In this research, a real-world dataset named UrduSig was developed, comprising scanned images of signatures collected from diverse official documents. The dataset captures the variability and complexity inherent in Urdu handwriting, including differences in writing styles, stroke thickness, and document quality. To address the challenges of accurately detecting signatures, the effectiveness of the YOLOv7 object detection framework was investigated. The proposed methodology leverages YOLO v7's advanced feature extraction and detection capabilities to identify signatures in Urdu script with high precision, overcoming the unique challenges posed by the intricate structure of Urdu script. Overall architecture of proposed methodology is shown in Figure 1.

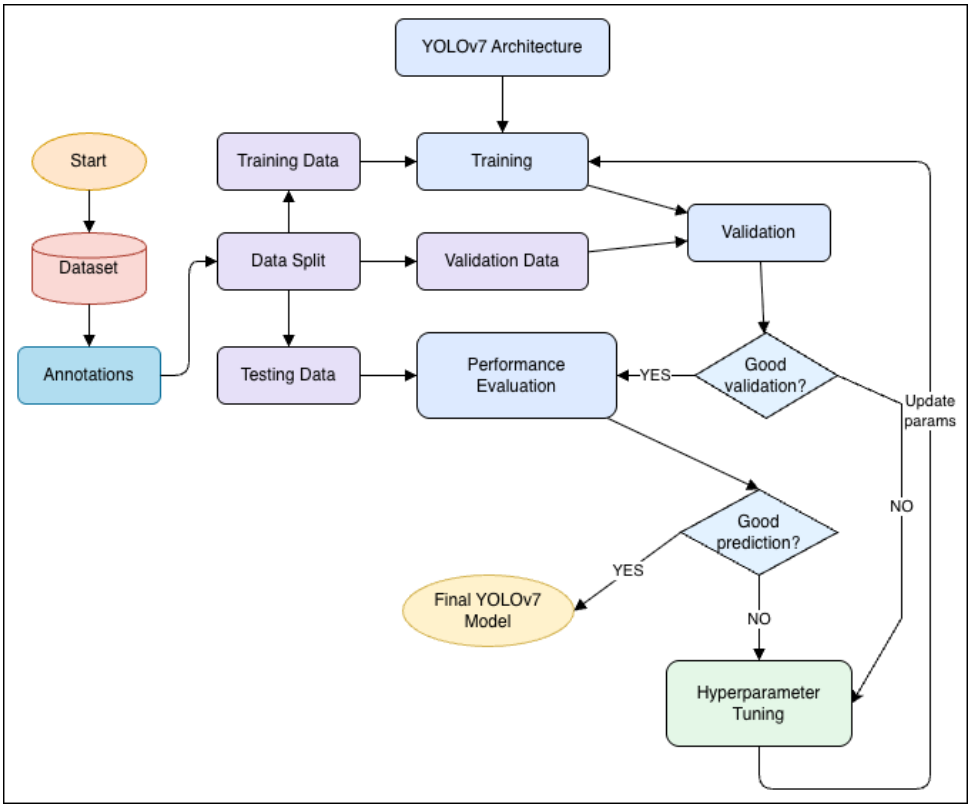

Figure 1: Architecture of proposed methodology for Urdu signature detection

3.1 UrduSig Dataset Creation

In the context of real life, Pakistan's official documents being digital, getting clean signatures in Urdu script from images proves to be a very difficult task [26][27]. In these documents, signatures do not lie within limited geographic area and can cover the elements like logos, stamps, date, and more texts [28]. These interactions create genuine signatures that are intrinsically stochastic, and those signatures exist against a backdrop of rich graphical patterns. As it is well established that only a limited number of datasets of official Urdu documents are available in the research domain, a dataset of Pakistani documents containing signatures was incorporated into this study. A Latin and signatures in Urdu script samples are presented in Figure 2.



Figure 2: Samples of Urdu (b & d) and latin nignatures (a & c)

To achieve this goal, the proposed approach aims to bridge the gap between highly controlled dataset environments and real-world conditions. The dataset comprises various official documents commonly found in Pakistan, including national identity cards, passports, bank documents, and legal papers, as illustrated in Figure 3. By incorporating diverse document types, the dataset captures a broad range of signature styles, document formats, and background complexities associated with official signatures. This diversity contributes to developing a more robust and realistic signature detection system capable of addressing the complexities inherent in real-world signatures in Urdu script.

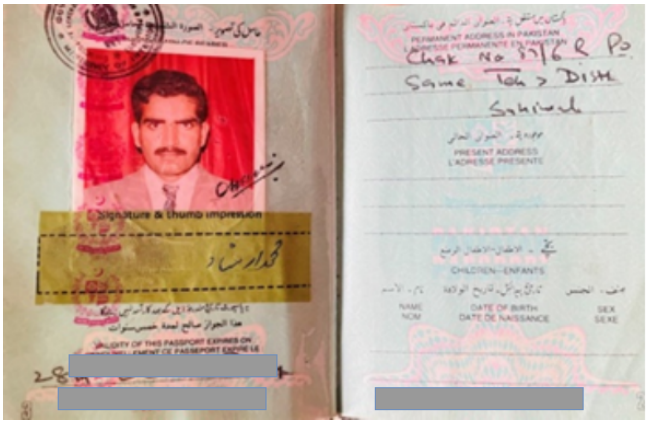

Figure 3: A sample from the UrduSig dataset of a passport with Urdu nastaleeq signatures

The forms were completed by 110 volunteers (40 male and 70 female), aged between 18 and 65 years, representing diverse geographical backgrounds across four regions of Pakistan. Participants were instructed to complete the forms in real-world scenarios, without including any personal information. Various pen colors were used to simulate natural variation found in real documents. The signature areas were intentionally left blank to enable the dataset's applicability in additional research domains such as Optical Character Recognition (OCR) for handwritten Urdu or integration of signatures into different document contexts. After completion, the forms were scanned at high resolution to generate digital versions. A total of 450 sheets were completed, supplemented by 80 public service feedback forms and 60 government correspondence letters, resulting in a total of 590 document samples. These documents include real-world artifacts such as dates, official logos, ink blots, and varied handwriting, capturing the complexity and variability typical of practical administrative documentation in Pakistan. A separate dataset of Urdu handwritten signatures was developed specifically for this study. A total of 10,000 genuine signatures in Urdu script were collected from 500 individuals across different regions of Pakistan. This dataset represents a wide spectrum of signature styles, capturing the diversity inherent in Urdu calligraphy and individual signing habits. Signatures were collected using standard signature cards and subsequently digitized to ensure uniformity and high-quality representation. Statistics of dataset is presented in Table. 2, providing a solid foundation for training YOLO v7 model.

UTSig (Persian), PHBC (Arabic), and FUM (Persian) were primarily designed for signature verification and forgery detection under controlled acquisition settings, UrduSig is the first dataset dedicated to Urdu in real-world documents. Unlike the aforementioned datasets, which consist of signatures collected on constrained forms with predefined boxes, UrduSig contains more than 5,000 scanned official documents (e.g., passports, bank forms, court orders, government correspondence, and feedback forms) featuring over 9,000 signatures embedded in heterogeneous backgrounds. This ensures the dataset captures authentic complexities such as logos, stamps, seals, handwritten notes, ink blots, and overlapping content that are absent in previous collections. Furthermore, UrduSig represents the signing habits of 500 individuals from diverse regions of Pakistan, incorporating variations in pen colour, stroke thickness, signature density (1–5 signatures per page), and multiple scanning resolutions. These characteristics make UrduSig uniquely suited for the upstream task of signature detection in cluttered, naturalistic environments, thereby complementing existing datasets focused on verification.

Description	Value
Dataset Information	
Total Number of Images	5,000
Total Number of Signatures	9,000
Image Resolution (Width $\times$ Height)	
1920×1080	1,200 images
1280×720	1,800 images
640×480	1,500 images
Other	500 images
Signature Characteristics	
Min Number of Signatures per Image	1
Max Number of Signatures per Image	5
Mean Number of Signatures per Image	2.4
Median Number of Signatures per Image	2

Table 2: Statistics of the proposed dataset

3.2 Experimental Settings

In this study, the YOLOv7 experiments were conducted on the Google Cloud Platform (GCP) [29] using the Compute Engine service. The computational environment comprised an a3-highgpu-8g virtual machine configuration equipped with 8 high-performance GPUs, 208 virtual CPUs (vCPUs), 1,872 GB of RAM, and a 6 TiB local SSD for fast data access and storage. This infrastructure was selected to ensure efficient training and inference for large-scale signature detection tasks.

3.3 Model

YOLOv7 [30][**?**] is a state-of-the-art model for object detection which uses a single-stage neural network that predicts objects' positions and classes in one pass [**?**]. The model is learned from a training dataset containing lots of images with labels and employs enhanced features such as a new loss function which incorporates localization and classification losses among others. The losses are calculated by following Eq. 1 to Eq. 3. The following loss functions are employed to optimize the object detection model:

$$L_{\text{loc}} = \alpha \sum_{p=0}^{S^2} \sum_{q=0}^{B} \delta_{\text{pos}}^{pq} \left[(\tilde{x}_p - x_p)^2 + (\tilde{y}_p - y_p)^2 + \left(\sqrt{\tilde{w}_p} - \sqrt{w_p} \right)^2 + \left(\sqrt{\tilde{h}_p} - \sqrt{h_p} \right)^2 \right] \tag{1}$$

$$L_{\text{conf}} = \sum_{p=0}^{S^2} \sum_{q=0}^{B} \left(\tilde{C}_p - C_p \right)^2 \tag{2}$$

$$L_{\text{bg}} = \beta \sum_{p=0}^{S^2} \sum_{q=0}^{B} \delta_{\text{neg}}^{pq} \left(\tilde{C}_p - C_p \right)^2 \tag{3}$$

In these Eqs., the terms L_{loc}, L_{conf}, and L_{bg} represent the localization loss, object confidence loss, and background (no-object) loss, respectively. The indicator functions δ_{pos}^{pq} and δ_{neg}^{pq} determine whether an object is present or absent in a specific grid cell and bounding box prediction. The hyperparameters α and β control the relative importance of the different loss components during training.

The dataset was split into a training set comprising 70% of the images and a validation set containing the remaining 30%. All the described labelling processes represent a novel contribution to the field. Subsequently, the proposed YOLOv7 deep learning model was employed to train on the labelled images.

3.3.1 Hyperparameter tuning

In the training process of a deep learning model, the selection of hyperparameters has a significant impact on the model's performances. For the YOLOv7 model for example, the initial learning rate (LR) of 0.005 is choosed to achieve the stable and sharp convergence through using the lesser update rates for the weights of the network during the back propagation. A lower LR helps to stop the model from overshooting the minima particularly in datasets which are more difficult to train, this is because it will ensure faster convergence of the minimum in the long run though the process might take longer time. The momentum value of 0.95 helps the optimizer maintain velocity in the right direction, speeding up the convergence process and reducing oscillations around local minima. A weight decay of 0.0001 is set to control overfitting more gently, allowing the model's weights to adjust slightly more freely, which can improve generalization on the training data. The box loss gain increased to 0.1 and class loss gain of 0.4 are adapted to shift the focus to both the localization and classification aspects of learning, to make sure that the model accurately identifies the boundaries of objects of interest, as well as assigns them to the correct category. An Intersection over Union (IoU) training threshold of 0.3 is used to apply a higher IoU between the predicted and the ground truth to make a valid prediction. This adjustment makes the model less likely to generate large numbers of false positives, which can improve precision, making it more conservative in its detection. List of hyper parameters are shown in Table 3. The YOLOv7 model

Parameter	**Set Value**
LR (start)	0.006
Momentum Coefficient	0.92
Weight Regularization	0.0002
Localization Loss Weight	0.12
Classification Loss Weight	0.35
IoU Threshold for Training	0.28
Random Image Blending	0.25
Scaling Factor Range	0.75
Horizontal Flip Probability	0.45

Table 3: Hyperparameter settings for the YOLOv7 model in signature detection experiments

was configured and trained using a carefully designed set of parameters optimized for the UrduSig dataset. The model architecture utilized the standard YOLOv7 backbone, CSPDarknet53, along with a PANet-based feature aggregation neck, to enhance multi-scale feature representation. All input images, originally of varying resolutions ranging from 640×480 to 1920×1080, were uniformly resized to 640×480 for training to ensure consistency. A batch size of 16 was used, and the model was trained for 300 epochs. The initial LR was set to 0.005 and decayed using a cosine annealing scheduler. Stochastic Gradient Descent (SGD) served as the optimizer with a momentum of 0.95 and a weight decay of 0.0001 to prevent overfitting. The loss function incorporated a box loss gain of 0.1 and a classification loss gain of 0.4, while the objectness loss gain followed the default YOLOv7 setting. The IoU threshold for positive sample matching during training was set to 0.3.

To increase the robustness and generalization of the model, several data augmentation techniques were applied. These included random horizontal flipping with a probability of 0.4, random scaling in the range of [0.8, 1.2], and random translation with a probability of 0.2. In addition, the training pipeline incorporated standard YOLOv7 augmentations such as color jittering, HSV shifts, and mosaic augmentation. Default anchor boxes were used initially and were subsequently fine-tuned based on the statistical distribution of signature bounding box sizes within the dataset.

To address the potential bias of over-detection, negative samples were explicitly incorporated into the training pipeline. While YOLOv7 inherently treats all non-annotated regions within an image as background, this principle was further extended by incorporating a dedicated subset of documents that contained no signatures. These negative samples comprised blank administrative forms, official letters without signatures, and document regions containing only printed text, stamps, or logos. Such images were annotated with zero bounding boxes

and added to the training set to reinforce the distinction between genuine signature regions and non-signature content. During training, these samples were interleaved with positive examples to maintain a balanced distribution, ensuring the detector was equally exposed to both scenarios.

4 Performance Evaluation and Results

On a global scale, the mean Average Precision (mAP) serves as a widely recognized metric in the field of object detection and is employed as the primary tool for evaluating the effectiveness of the proposed methodology. The evaluation process commences with an assessment of the accuracy of bounding box detections, which involves calculating the IoU between predicted and ground truth bounding boxes, as defined in Eq. 4. The IoU is calculated to measure the overlap between predicted and ground truth bounding boxes and is defined as:

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{4}$$

Precision and recall metrics are computed using Eqs. 5 and 6, where false negatives (FN), true positives (TP), false positives (FP) and true negatives (TN) are employed.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

Following these calculations, the Average Precision (AP) is determined, representing the area under the precision-recall curve. The mAP is computed as the mean of AP values across all categories. Given that the present experiments are focused on a single class, mAP and AP are treated as equivalent metrics. In accordance with the Common Objects in Context (COCO) [**?**] evaluation protocol, an IoU threshold of 0.5 is adopted, as it has demonstrated effectiveness in mAP calculations within the object detection domain. To provide a more comprehensive evaluation, mAP values are also computed across IoU thresholds ranging from 0.5 to 0.95, in increments of 0.05, commonly referred to as mAP@0.5:0.95. As illustrated in Figure 4, the proposed model achieves an mAP@0.5 of 0.975 and an mAP@0.5:0.95 of 0.885 after 350 training epochs.

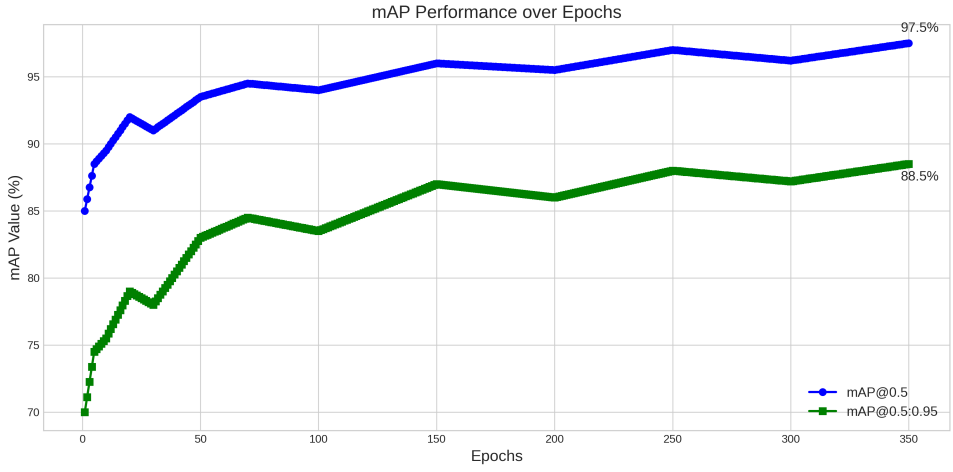

Figure 4: Evolution of mAP@0.5 and mAP@0.5:0.95 over 350 training epochs

In Figure 5(a), it is show that the YOLOv3 is capable of identifying most signature regions but often produces coarse bounding boxes and occasionally misclassifies graphical elements or handwritten strokes as signatures. The presence of overlapping seals and cluttered text leads to false positives, demonstrating the model's limited

ability to differentiate signatures from similar background patterns. Figure 5(b) illustrates the outputs from YOLOv5, which shows clear improvement over YOLOv3. The bounding boxes are tighter and more consistent, and the number of false detections is reduced. YOLOv5 demonstrates a stronger ability to generalize to noisy and heterogeneous backgrounds, though in complex cases it still misses some subtle or faint signatures and struggles with closely spaced instances. Figure 5(c) illustrates the training and validation progress of the YOLOv7 model for signature detection. The plots show consistent declines in the box loss, classification loss, and distribution focal loss (DFL) across training epochs, indicating effective learning and convergence of the model. Precision and recall curves demonstrate upward trends, suggesting improved detection reliability as training advances. Furthermore, metrics for mAP@0.5 and mAP@0.5:0.95 increase steadily, reflecting the model's growing ability to localize and classify signatures accurately. Quantitatively, the trained YOLOv7 model achieved mAP of 0.975 at an IoU threshold of 0.5 and an mAP@0.5:0.95 of 0.885, confirming robust performance under varying detection criteria.

Table 4 presents a comparative evaluation of YOLOv3, YOLOv5, and YOLOv7 models for signature detection. The YOLOv7 model achieved superior performance, with a mAP@0.5 of 0.975 and mAP@0.5:0.95 of 0.885, outperforming both YOLOv3 and YOLOv5. These results highlight YOLOv7's improved capability in accurate detection and localization of signatures under varying IoU thresholds.

Model	**mAP@0.5**	**mAP@0.5:0.95**
YOLOv3	0.891	0.742
YOLOv5	0.936	0.803
YOLOv7	0.975	0.885

Table 4: Performance comparison of YOLO models for signature detection

ChatGPT said:

In addition to mAP, precision, recall, and F1-score at IoU = 0.5 are reported to provide a more comprehensive evaluation of the model performance as presented in 5. The results show that YOLOv7 achieves the highest scores across all three metrics, with precision of 0.975, recall of 0.96, and F1-score of 0.97, indicating a strong balance between accurate detections and coverage of true signatures. YOLOv5 follows with precision of 0.945, recall of 0.92, and F1-score of 0.93, while YOLOv3 attains precision of 0.90, recall of 0.88, and F1-score of 0.89. The transformer-based DETR and the two-stage Faster R-CNN perform reasonably well, but both fall behind the YOLO series in terms of precision–recall balance. These results reinforce the superiority of YOLOv7, which consistently provides the most reliable and accurate detection of signatures in real-world documents. Figure 6 presents a radar chart comparing Faster R-CNN, DETR, YOLOv3, YOLOv5, and YOLOv7 across

Model	**Precision (P)**	**Recall (R)**	**F1-score**
Faster R-CNN	0.88	0.84	0.86
DETR	0.92	0.85	0.88
YOLOv3	0.90	0.88	0.89
YOLOv5	0.945	0.92	0.93
YOLOv7	0.975	0.96	0.97

Table 5: Precision, Recall, and F1-scores of different detection models for signature detection in Urdu script at IoU = 0.5

seven evaluation metrics: model size, parameters, training Video Random Access Memory (VRAM), Floating Point Operations (FLOPs), inference time, inference Frames Per Second (FPS), and mAP. Faster R-CNN shows the heaviest complexity with a model size of approximately 240 MB, 63M parameters, and high computational cost (> 200 GFLOPs), requiring more than 10 GB of VRAM and resulting in slow inference (120 ms, ~ 8 FPS) despite achieving a reasonable accuracy of mAP@0.5 = 0.891. DETR performs slightly better with a smaller model ($\sim$180 MB, 45M parameters), moderate VRAM usage ($\sim$8 GB), and inference time of $\sim$90 ms ($\sim$11 FPS), yielding mAP@0.5 $\approx$ 0.91. In contrast, YOLOv3 reduces the complexity to $\sim$120 MB and 40M

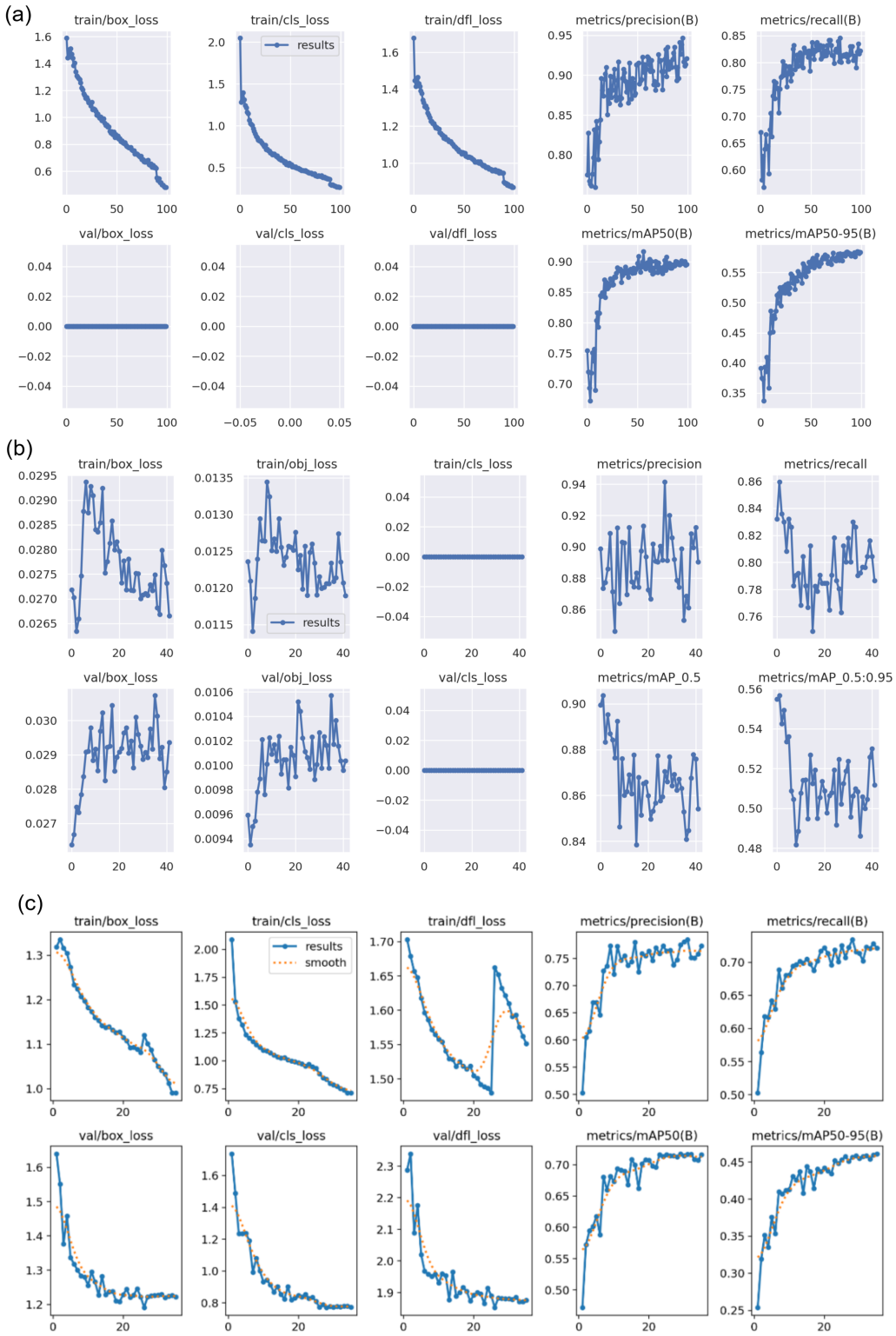

Figure 5: Comparison of detection results for YOLOv3 (a), YOLOv5 (b), and YOLOv7 (c), showing YOLOv7 achieving superior accuracy and localization

parameters, running at ∼40 ms per image (∼25 FPS) with mAP@0.5 = 0.891. YOLOv5 further improves efficiency with a 90 MB model, 27M parameters, and ∼6 GB VRAM, reaching ∼25 ms inference time (∼40 FPS) and mAP@0.5 = 0.936. Finally, YOLOv7 demonstrates the most balanced performance, with a compact 75 MB model and 25M parameters requiring only ∼5 GB of VRAM, fast inference (∼15 ms, ∼65 FPS), and the highest detection accuracy (mAP@0.5 = 0.975, mAP@0.5:0.95 = 0.885). These results clearly indicate that YOLOv7 offers the best trade-off between accuracy, speed, and computational efficiency, making it the most suitable choice for real-world signature detection.

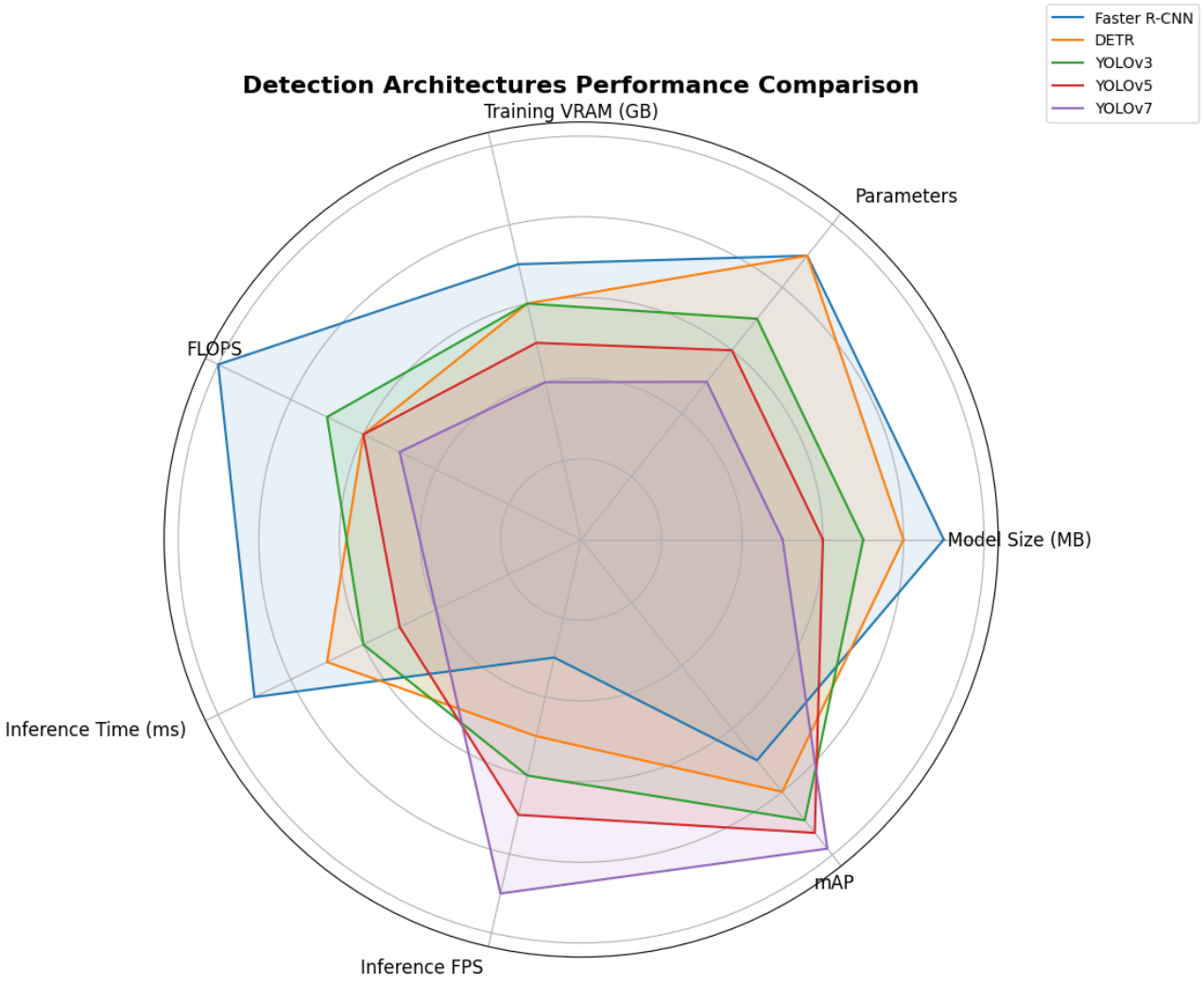

Figure 6: Performance comparison of detection models, with YOLOv7 achieving the best balance of accuracy and efficiency

Figure 7 visually demonstrates the result of the signature detection process using the YOLOv7 model. As shown, the model successfully localizes the signature within the Urdu document by drawing a precise bounding box around the region of interest. The detected box corresponds accurately to the handwritten signature area, indicating a high-confidence prediction. This visual result affirms the model's ability to distinguish signature patterns from other textual and graphical components present in real-world documents. The bounding box output not only reflects the spatial accuracy of detection but also serves as a foundation for downstream tasks such as signature verification, extraction, or segmentation.



Figure 7: Results of bounding box detection of signatures using the YOLOv7 model

The UrduSig dataset contains documents with varying numbers of signatures, ranging from single-signature pages to multi-signature documents containing several instances. This variation reflects real-world scenarios, where official documents often require endorsements from multiple individuals. During training, such diversity provides the model with exposure to different spatial distributions, scales, and densities of signatures, thereby improving its generalization capability. However, it also introduces challenges: documents with a higher number of signatures increase the likelihood of overlapping bounding boxes and clutter, which can lead to false positives or missed detections. The experiments show that while YOLOv3 and YOLOv5 occasionally struggle with dense multi-signature pages, YOLOv7 is more robust, accurately localizing each instance even when signatures overlap or are closely positioned. Thus, the inclusion of documents with different signature counts not only enhances the realism of the dataset but also serves as a critical factor in evaluating the scalability and robustness of detection models.

5 Conclusion and Future Work

5.1 Conclusion

This study proposed an effective approach for detecting signatures in real-world scanned document images by leveraging the YOLOv7 deep learning model. A new dataset, UrduSig, was introduced, comprising diverse official documents from Pakistan, along with genuine and forged signatures in Urdu script were collected from

individuals across various regions. The dataset captures significant variations in document layouts, background complexities, and handwriting styles, contributing to the realism and practical relevance of the experiments. Experimental results demonstrated that YOLOv7 achieved a high detection accuracy, yielding a mAP@0.5 of 0.975 and a mAP@0.5:0.95 of 0.885. These results notably outperformed earlier models such as YOLOv3 and YOLOv5, underscoring YOLOv7's superior capability in handling the unique challenges posed by the complex Urdu script and diverse document formats. The high precision and recall scores reflect the robustness of the proposed methodology in accurately localizing signatures, even in cluttered and heterogeneous backgrounds. The findings indicate that advanced object detection architectures like YOLOv7 offer a promising solution for automated signature detection in practical applications, including document verification, fraud prevention, and archival processes

5.2 Future Work

While the proposed approach demonstrates strong performance in detecting signatures, several directions remain open for further research. Future work may focus on integrating a robust signature verification module to distinguish between genuine and forged signatures, enabling end-to-end authentication systems. Expanding the dataset to include a broader range of document types, regional handwriting variations, and multilingual content could further enhance the model's generalizability. Optimizing the model for lightweight deployment on mobile or embedded devices would also facilitate real-time signature detection in field environments, such as banking, legal, or administrative settings.

References

[1] O.-A. Ticleanu, I. D. Hunyadi, and N. Constantinescu, "Temperate blind signature scheme for particular subspaces," *Applied Sciences*, vol. 15, no. 13, p. 7180, 2025.

[2] K. Yan, Y. Zhang, H. Tang, C. Ren, J. Zhang, G. Wang, and H. Wang, "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 (CVPR) Workshops*, pp. 5163–5172, June 2022.

[3] M. R. Ashraf, M. Hussain, M. A. Jaffar, W. Y. Ramay, and M. Faheem, "Revolutionizing urdu sentiment analysis: Harnessing the power of xlm-r and gpt-2," *IEEE Access*, vol. 12, pp. 99779–99793, 2024.

[4] H. Wani, K. P. Rane, V. M. Deshmukh, K. D. Badgujar, J. Jeyvel, and S. Wategaonkar, "Ai-synergistic effect in c-rnn based online signature profiling and hash-qr encoding for forge signature detection, documents protection and applications authentication," *Multimedia Tools and Applications*, pp. 1–42, 2025.

[5] L. Cui, Q. Shen, Y. Xiao, D. Liu, and H. Wang, "Sparse graph structure fusion convolutional network for machinery remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 254, p. 110592, 2025.

[6] M. Ehrmann, A. Hamdi, E. L. Pontes, M. Romanello, and A. Doucet, "Named entity recognition and classification in historical documents: A survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–47, 2023.

[7] W. Xiao and H. Wu, "Learning features for offline handwritten signature verification using spatial transformer network," *Scientific Reports*, vol. 15, no. 1, p. 9453, 2025.

[8] N. Arab, H. Nemmour, M. L. Bouibed, and Y. Chibani, "1d-gan for improving offline handwritten signature verification based on small sets of real samples," *Multimedia Tools and Applications*, vol. 84, no. 21, pp. 24541–24561, 2025.

[9] J. Pan, S. Xu, Z. Cheng, and S. Lian, "C2F-YOLO: A coarse-to-fine object detection framework based on YOLO," in *Proceedings of the 3rd Asia Conference on Algorithms, Computing and Machine Learning (CACML 2024)*, pp. 150–157, ACM, 2024.

[10] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: A review and its applications," *Multimedia Tools and Applications*, vol. 83, pp. 83535–83574, 2024.

[11] A. Anitha, M. Priya, B. Kamaraj, N. K. Yadav, and S. Srivastav, "Handwritten signature forgery identification and prediction using harris corner detection and siamese network," in *Proceedings of the 2024 2nd International Conference on Emerging Trends in Information Technology and Engineering (IC-ETITE)*, 2024.

[12] S. Singh, S. Chandra, and A. R. Verma, "Enhancing offline signature verification via transfer learning and deep neural networks," *Augmented Human Research*, vol. 9, no. 1, p. 4, 2024.

[13] H. Li, P. Wei, Z. Ma, C. Li, and N. Zheng, "TransOSV: Offline signature verification with transformers," *Pattern Recognition*, vol. 145, p. 109882, 2024.

[14] K. Vijiyakumar, V. Govindasamy, and V. Akila, "An effective object detection and tracking using automated image annotation with inception based faster R-CNN model," *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 343–356, 2024.

[15] K. Binagekar and A. Pai, "Real-time structural crack detection in buildings using YOLOv3 and autonomous unmanned aerial systems," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 5, pp. 1874–1887, 2024.

[16] A. Giazitzis and E. N. Zois, "SigmML: Metric meta-learning for writer independent offline signature verification in the space of SPD matrices," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6300–6310, 2024.

[17] V. K. Sharma and R. N. Mir, "Saliency guided faster-RCNN (SGFr-RCNN) model for object detection and recognition," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 5, pp. 1687–1699, 2022.

[18] Y. Zhang, J. Zhang, K. Yan, H. Wang, and G. Wang, "Handwritten Chinese signature detection with simple copy–paste augmentation on power plants technical documents," *Service Oriented Computing and Applications*, vol. 17, no. 4, pp. 293–302, 2023.

[19] W. Wu, H. Liu, L. Li, Y. Long, X. Wang, Z. Wang, J. Li, and Y. Chang, "Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image," *PLOS One*, vol. 16, no. 10, p. e0259283, 2021.

[20] M. Moustapha, M. Taşyürek, and C. Öztürk, "A novel YOLOv5 deep learning model for handwriting detection and recognition," *International Journal on Artificial Intelligence Tools*, vol. 32, no. 4, p. 2350016, 2023.

[21] A. Soleimani, K. Fouladi, and B. N. Araabi, "UTSig: A Persian offline signature dataset," *IET Biometrics*, vol. 6, no. 1, pp. 1–8, 2017.

[22] Y. Akbari, M. J. Jalili, J. Sadri, K. Nouri, I. Siddiqi, and C. Djeddi, "A novel database for automatic processing of Persian handwritten bank checks," *Pattern Recognition*, vol. 74, pp. 253–265, 2018.

[23] M. Schlichtig, S. Sassalla, K. Narasimhan, and E. Bodden, "FUM – a framework for API usage constraint and misuse classification," in *Proceedings of the 2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pp. 673–684, 2022.

[24] S. A. A. Hashmi, S. Upadhyay, and R. Kumar, "Comparative study of forged Urdu signatures done by persons not familiar to language (region of Sikkim and Kashmir)," *Journal of Punjab Academy of Forensic Medicine & Toxicology*, 2021.

[25] U. Tariq, Z. Hu, R. Tariq, M. S. Iqbal, and M. Sadiq, "High-performance embedded system for offline signature verification problem using machine learning," *Electronics*, vol. 12, no. 5, p. 1243, 2023.

[26] M. Sethi, M. Kumar, and M. K. Jindal, "Forensic handwriting analysis: a hybrid classification framework for writer identification in devanagari script," *Multimedia Tools and Applications*, vol. 84, no. 27, pp. 33183–33194, 2025.

[27] Google Cloud, "Google cloud platform." `https://cloud.google.com/`, 2018.

[28] S. Mammeri, M. Amroune, M.-Y. Haouam, I. Bendib, and A. C. Silva, "Early detection and diagnosis of lung cancer using YOLO v7, and transfer learning," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30965–30980, 2024.

[29] Y. Chen, H. Liu, and F. Fang, "A novel pansharpening method based on cross stage partial network and transformer," *Scientific Reports*, vol. 14, no. 1, p. 12631, 2024.

[30] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, 2020.