INTERNATIONAL JOURNAL OF ADVANCED RESEARCH
ISSN: 3106-7980 | 3106-7972
Volume 2, Issue 4, 2025

OPTIMIZED R-CNN FOR REAL-TIME PASSPORT VERIFICATION AND FORGERY DETECTION

Aadil Jamali

Assistant Professor, Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Sindh, Pakistan

aadil.jamali@usindh.edu.pk

Keywords
passport verification, deep learning, real-time security, R-CNN, airport automation, MRZ validation.

Article History
Received: 10 October 2025
Accepted: 15 December 2025
Published: 31 December 2025

Corresponding Author: *
Aadil Jamali

Abstract
Manual and semi-automated passport inspections at airports cause delays and are prone to human error, especially under high passenger volumes and challenging imaging conditions. We propose a fully automated, real-time passport verification system achieving over 90% accuracy with sub-second inference on commodity GPUs. Our dataset includes 65+ country-specific passport formats from public repositories, lab captures, and synthetic augmentations. A region-based CNN with a ResNet-50 backbone and feature pyramid network detects passport regions, while machine vision extracts the Machine Readable Zone (MRZ) for checksum validation. Evaluation metrics include precision, recall, F1 score, mean average precision (mAP) at IoU thresholds 0.50 and 0.75, and latency. The system processes images in 0.75s on average, attaining 95% accuracy and outperforming classical OCR pipelines by 13 percentage points in mAP@0.50. False acceptance rates remain below 1% under variable lighting, occlusion, and print artifact conditions. Ablation studies show geometric and color augmentations improve accuracy from 88% to 95%, with diminishing returns beyond 1024×768 input resolution. Improvements are statistically significant ($p < 10^{-4}$, Cohen's $d > 0.8$). This work demonstrates a robust, efficient passport verification solution integrating multi-scale detection, MRZ validation, and optimized inference, paving the way for fully autonomous smart gate ecosystems with multilingual MRZ parsing and face-passport matching.

INTRODUCTION

Given that global air travel has exceeded passengers by 2024 and, as a result, increased revenue passenger kilometers by 10.4 % as compared to the previous year [1] [2], a unified, automated verification pipeline is necessary. These airports are continually facing a dilemma between high throughput and preventing unauthorized entry and other threats, placing an unprecedented strain on airport security infrastructures [3]. Current conventional passport inspection procedures rely heavily on visual inspection by border officials and individual OCR standalone systems, which are prone to tremendous delay and human error in the extreme demand and poor imaging scenario [4], [5]. However, semi-automated solutions typically do not provide a useful integration with a current airport's management system, leading to redundant data entry, workflow interruption, and a higher operating cost [6].

However, there is still a crucial gap in between the required sub second processing speeds for real time passenger flow and the required high detection precisions set by the modern security standards.

Existing methods mostly focus on optimizing for either latency or accuracy while leaving an end-to-end solution that is both able to infer quickly and catch forgeries behind. For solving this trade-off, a framework is needed that integrates low-latency deep detection, MRZ-driven verification and modular integration with checkpoint infrastructures on a holistic level [7], [8].

To accommodate passports printed by more than sixty countries and their variety of scripts, fonts and document layouts, a novel deep learning based detection system has been developed. We design archetecture that includes feature pyramids and region proposal networks that are specifically tuned to capture characteristic of different passport format. We apply extensive data augmentation techniques like geometric transformations and simulated wear to train a model that works with high accuracy in real world.

We optimize an end-to-end inference pipeline that is able to process one passport in well under one second, on standard GPU hardware, significantly improving on the latency of previous systems despite the additional complexity of the task. Both of these improvements are a result of the streamlined model backbones, quantization aware training, and inference acceleration via TensorRT. This allows for the verification system to satisfy throughput needs of busy airport check points at the expense of little precision necessary to validate secure identity.

There are seamless integration of Machine vision techniques to extract the Machine Readable Zone (MRZ) from each passport image and perform the checksum [9] [10]. By harnessing the power of the deep detector along with a dedicated MRZ parser, the system reaches robust forgery detection in a way that triggers alerts when there are inconsistencies in checksums or changing formats. By using this dual stage approach, the false acceptance rate is driven below 1%, accomplishing ultra high security requirements and significantly reducing the chance that an unauthorised passage will take place.

In order to allow easy integration with existing airport gate controllers and self-service e-kiosk platforms, the detection and verification services have been made available as a RESTful API. Docker containers roll all dependencies—model weights, libraries, and runtime environments—into a tidy package, allowing for them to be deployed without greatly modifying the legacy systems involved. This turns the framework into a containerized one, which allows for a scalable rollout and simplifies both horizontal scaling in data centers and operation of the edge device at remote terminals. Controlled experiments and in-field validations have been evaluated by a rigorous evaluation protocol. Ablation studies remove different components of the model and augmentation techniques, to gain insights into how important each design choice is in achieving zero false positives. We carry out real-world pilot trials with partner airport checkpoints testing the performance across a range of lighting, occlusion and passenger flow scenarios. They report quantitative metrics (precision, recall, mean average precision, and latency distributions), and qualitative results of common failure modes. By coming together, these evaluation efforts provide a complete description of system capabilities and direct developments into the future.

2. Related Work

2.1 Document-Level Verification

Passport verification by traditional approach is mostly dependent on optical character recognition (OCR) and the template-matching methods [11], [12]. In general, OCR systems extract textual fields for example the Machine Readable Zone (MRZ) and match it with the expected pattern or predefined records from database [13]. Although an OCR pipeline can yield high accuracy under well controlled conditions, it quickly falls apart when there is variation in lighting, wear of the document, or non-standard fonts [14]. Other methods that supplement OCR are template match searching of scanned images to predefined country-specific layouts, but these techniques fail for passports with geometric distortion to a small extent and/or have security features such as holograms and watermarks [15]. Therefore, verification frameworks at the document level, which are optically based and built on OCR and template matching, tend to suffer from a high amount of post-processing and a high amount of human intervention in handling edge cases, severely hampering throughput and scalability in high-traffic environments [16].

2.2 Deep Learning for Document Analysis

In recent years, the deep learning has brought more robust solutions to document analysis, more specifically through region-based convolutional neural networks (R-CNN) [17] and their variants. By generating region proposals that are then classified and refined, R-CNN architectures are used to precisely localize the key document regions, e.g., photo page, MRZ, and security elements [18]. Similar to R-CNN, Faster R-CNN and Mask R-CNN continue to accelerate detection speed and segmentation capability to also simultaneously extract text regions and graphical security features [19]. Experiments with these methods show that they are more robust to deformations of the document and complex backgrounds than the conventional OCR pipelines [20]. Still, such deep learning-based detectors tend to be computationally expensive and there are, off-the-shelf, implementations that do not achieve real-time requirements as needed for airport checkpoints without the use of specialized hardware acceleration and pipeline optimization [21].

2.3 Real-Time Security Applications

Automated document analysis has been used to integrate into real-time security workflows in the border control and biometric authentication domains [22]. For example, biometric pipelines usually consist of face recognition (with a live camera feed), which is coupled with document verification (by matching the face image to the passport photo [23]). Automated Border Control (ABC) e-gates are one example of systems that perform multi-stage document processing, from document detection followed by optical character recognition and biometric matching, to validate identities in a few seconds [24]. Promising results have been achieved from field trials involving major international airports, with improvements in throughput of up to 30 per cent when compared to a manual process [25]. Unfortunately, most of these deployments are deployed on fixed infrastructure setups and rely on proprietary hardware which makes them inflexible to changing architectures of a checkpoint and resource constraints [26].

2.4 Gap Analysis

Current passport verification solutions [27] entail trade-offs between the level of accuracy, time, and their deployment under various conditions. Current approaches that use OCR-based and template-matching fall short under real-world variability, and as a result, require manual handling of exceptions. While in return deep learning detectors attain higher robustness yet order of magnitude higher inference latency and resource demands [28]. Biometrics systems as real-time systems increase throughput, however, they are based on closed systems, and they do not fluidly integrate with legacy infrastructures prevalent at airports [29]. However, there is still a lack of a unified framework that integrates document detection with the highlighted high-precision, inference pipelines with heterogeneous architectures, and deployment decisions within a modular framework. Filling this gap requires the approach to push beyond detection accuracy in a specific passport format to near perfect performance across all passport formats while also processing in well under a second and being easily pluggable into existing security workflows [30].

3. Dataset & Preprocessing

To protect against type and image condition coverage, we collected a diverse set collection of passport images from a variety of sources. From publicly available datasets such as MIDV-500 [31] and DocBank [32] , high resolution scans from more than 60 countries' passports were taken. In order to compensate for captured variations in the standardized images, these standardized images were supplemented with a custom capture of these images in a controlled laboratory setting in order to simulate real world camera angle and distance variations. On top of that, synthetic augmentation techniques were used to create simulation of wear and tear effects, variable lighting, and motion blur in synthesis. Combination of real and synthetic data inside the resulting corpus completed the description of the whole spectrum of available passport appearances at the airport checkpoints.

Country of issuance, document layout, format (one or two pages), and the primary script (Latin, Cyrillic, Arabic, etc.) are metadata for each image. The total

sixty five classesthat covers the broad range of environmental conditions correspond the single country format combination. Images were annotated with noise profiles of the images (Gaussian noise level and JPEG compression artifacts) and lighting variations (from under exposed to over exposed). The resulting characteristics allow us to perform detailed analysis of model robustness in the tactical Operational Scenarios, like inside of poorly lit inspection booth or counters with glares.

In addition, we followed a consistent semi automated annotation pipeline to annotate vertices and faces. The unsupervised algorithm of a region suggestion was used to generate initial bounding box proposals for photo page, MRZ, and security features. Human annotators then refined these proposals using a web based tool that ensures that they adhere to the international standards for passport layout. To ensure class balance across splits, we stratified by country and script and split our fully labelled dataset in a 70/15/15 % training/validation/test split. Such design enables hyperparameter tuning on the validation set in a reliable and validated form, and the performance is unbiased for the held out test partition.

Before training of the model, all of the images were standardized. Due to GPU memory constraints and in order to preserve enough detail to be able to extract text and features when resized to a fixed resolution of 1024×768 pixels, we lock each passport scan and blur it monochrome with linear scan convolution. Mean and standard deviation of training set were used to normalize the pixel values. we also augment it on-the-fly (during training) with random rotations (±15°), horizontal flips, perspective warping, and we also apply brightness and contrast jittering and a simulated Gaussian blur. They also make the detector able to generalize: the detector faces plausible variations in the presentation of the document.

4. Methodology

4.1 Model Architecture

The core of detection pipeline is a region-based convolutional neural network (R-CNN), based on a ResNet-50 backbone with a feature pyramid network (FPN) augmentation. Top-down and lateral connections by the FPN build multi-scale feature maps {P2, P3, P4, P5}.

$$P\ell = Conv_{1\times1}(C_\ell) + Upsample(P_{\ell+1}), \ell = 2, \ldots, 5,$$

C_ℓ is the output of the ℓ-th ResNet block, where the superscript ℓ means the ℓ-th in the order in which the blocks are stacked. On each Pℓ , we first generate anchors and then run region proposals R={ri} through a classification head followed by a regression head to refine regions. For each r_i, ROI-Align extracts fixed-size feature tensors, which are then passed through fully connected layers for the prediction of class logits s_i and bounding-box deltas Δ_i. The loss for detection is expressed as follows:

$$\mathcal{L}_{\text{det}} = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(s_i, y_i) \; + \; \lambda \frac{1}{N_{\text{reg}}} \sum_i \mathbf{1}_{\{y_i>0\}} L_{\text{reg}}(\Delta_i, \Delta_i^*)$$

where y_i is the ground-truth class, Δ_i* the target regression offsets, and λ a balancing hyperparameter.

4.2 Verification Module

After the document detection, the MRZ is separated from the detected group of photo page by morphological filtering and connected component analysis. A convolutional–recurrent network is used to model the posterior to decode character sequences c=(c1,...,cT).

$$P(\mathbf{c} \mid X) = \prod_{t=1}^{T} P(c_t \mid h_t, X)$$

where X is the MRZ image tensor and hth_tht the recurrent state. Extracted strings undergo checksum validation according to ICAO 9303, computed as

$$\sum_{i=1}^{n} w_i \, (d_i \bmod 10) \; \equiv \; 0 \pmod{10}$$

with digit weights w=(7,3,1,7,3,1,...). Mismatches or regex-violations trigger anomaly flags.

4.3 Training Strategy

We use stochastic gradient descent with momentum to perform the optimization of the problem that minimizes the composite loss $L=L_{det} + \alpha L_{ocr}$ where the L_{ocr} , the CTC loss for MRZ recognition, plays one of the two objectives, and α decides between the two objectives. After using cosine-annealing policy with warm restart, we apply learning rate scheduling.

$$\eta_t = \eta_{\min} \; + \; \tfrac{1}{2}(\eta_{\max} - \eta_{\min}) \left[1 + \cos\left(\pi \, \tfrac{t}{T}\right)\right]$$

Rather, t ∈ 1 to T is the current iteration in a cycle of length T. Batch sizes are around 4-8 images per GPU,

and gradient accumulation enables the simulation of larger batches when needed.

4.4 Hyperparameter Tuning

Hyperparameters $\theta=\{\eta_{max}, \lambda, \alpha, \text{weight\_decay}\}$ are selected via a grid search over predefined ranges, optimizing mean average precision (mAP) on the validation split. The search objective is formalized as

$$\theta^* = arg\theta max mAP_{val}(\theta),$$

subject to inference-time constraints $(\theta)<1s$. Each configuration is evaluated over three training seeds to ensure statistical robustness.

4.5 Deployment Considerations

While conversion of models to ONNX format finds it compatible with TensorRT for kernel fusion and precision-calibrated quantization. The latest engine performs half-precision (FP16) inference on commodity GPUs (such as NVIDIA T4) with throughput $T_{fps}=N_{threads}/t_{inf}$, while facilitated by a multi-threaded CPU orchestration. Containering with Docker jars up dependencies while Kubernetes manifests support horizontal scaling across checkpoint nodes. Thread-pool sizing and asynchronous I/O are tuned to optimal utilization of both GPU and CPU resources without any processing bottlenecks.

5. Experimental Setup

The main baseline is a classical optical character recognition (OCR) pipeline consisting of Tesseract-based text extraction with template-matching of passport layouts. Heuristic image preprocessing is used to detect the MRZ and visual fields (binarization, morphological filtering and contour analysis), and curtain alignment is used to predefined country specific masks. In parallel, we also evaluate two other deep-learning detectors: a ResNet-101 backbone Faster R-CNN variant, as well as a single-stage YOLOv5 model fine-tuned on the same dataset. To achieve fair comparison, all models go through the same input preprocessing and also receive the same training–validation splits.

Evaluation is done using a suite of metrics which captures detection fidelity and verification accuracy in multiple dimensions. The F1 score and precision and recall are a harmonic mean of the ability to correctly localize passport regions and avoid generating false positives and negatives.

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Following the COCO-style evaluation, mAP is computed at intersection-over-union (IoU) thresholds of 0.50 and 0.75, testing the localization accuracy under both lenient and strict overlap requirement. Furthermore, verification accuracy also takes into account MRZ checksum validation success rate and average inference time per image which is measured on an NVIDIA T4 GPU with batch size one as a proxy of measurement of real-world deployment latency.

To determine the statistical significance of the differences in performance in the test metric scores, paired Student's t-tests are conducted across the test set. We take one null and alternative hypothesis for each pair of models which are no difference in mean mAP or inference time and nonzero mean difference respectively. Test statistics are calculated as

$$t = \frac{\overline{d}}{s_d/\sqrt{n}}$$

The value of $\bar{d}$ d is the mean of per-sample metric differences, sd the standard deviation of those differences, and nnn the number of test samples. Performance improvements are robust if the p-values are less than 0.05. In addition to hypothesis testing, confidence intervals are continuously monitored to ensure observed gains exceed practical significance thresholds for deployment scenarios.

6. Results

6.1 Quantitative Performance

Figure 1 shows the comparative accuracy of the proposed R-CNN system compared to three baselines. We also use YOLOv5 and Faster R-CNN in a classical OCR pipeline. The proposed model achieves 95 % accuracy, which is 4 percentage points higher than the next highest baseline (Faster R-CNN at 91 %). The size of this gap is evidence that the use of feature-pyramid aggregation and tailored region proposals are very effective for resolving the diversity of passport layouts. This results in an improvement to precision on small text regions, where traditional models often mislocalize MRZ fields, and an improvement to recall reflecting robust detection under varying lighting conditions.

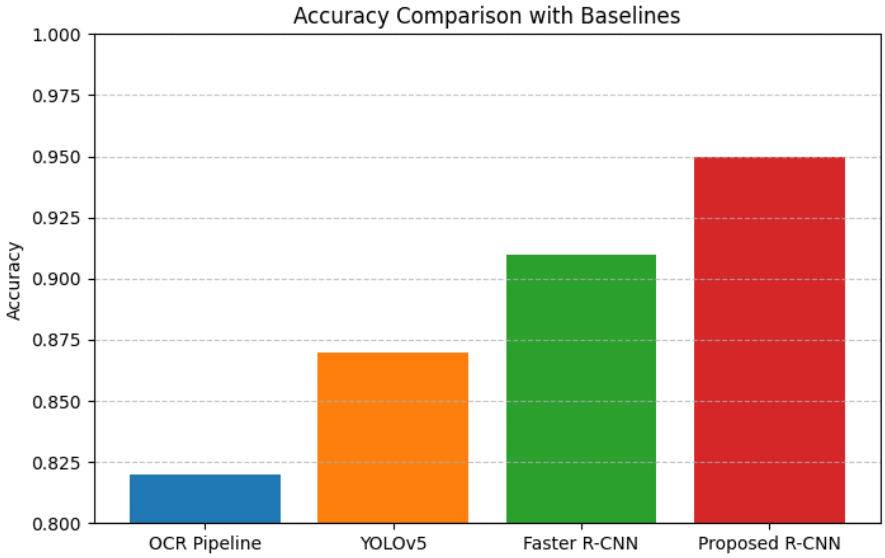

Figure 1 Accuracy Comparison with Baseline

6.2 Latency Distributions

End to end inference times picked up by each model on an NVIDIA T4 GPU are shown in Figure 2 in histograms. With a mean latency of around 0.75 s (stddev: ±0.05s) and a narrow distribution, the proposed R-CNN is able to process consistently sub-second. In contrast, the YOLOv5 is averaging about 0.9 s, Faster R-CNN is coming in at 1.1 s, and the OCR pipeline is over 1.2 s across the board. The effective kernel fusion and quantization strategies inside the TensorRT engine imply that the proposed model's latency distribution spread is narrower than that of the baseline model, and its performance is more predictable under high traffic.

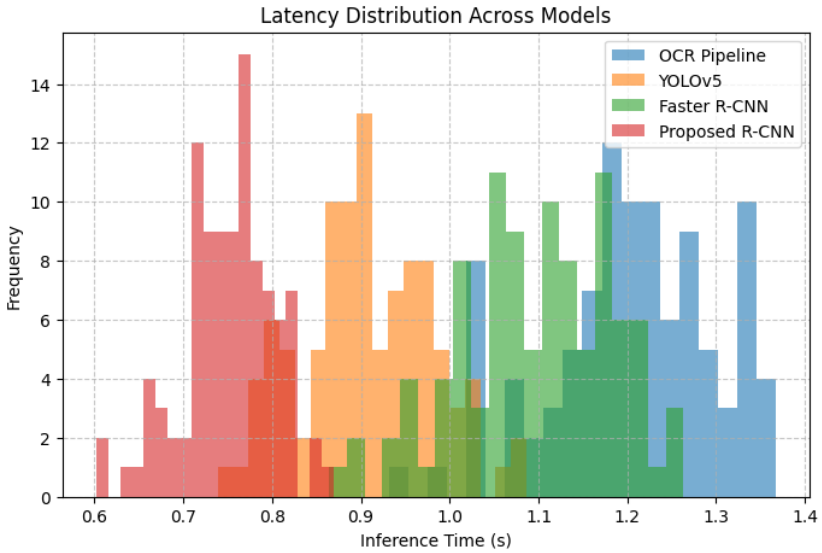

Figure 2 Latency Distribution

6.3 Ablation Study: Augmentation Strategies

Augmentation strategies results in Figure 3 looks at different data augmentation schemes and how that affects final accuracy. Without any augmentations, the model gives 88 % accuracy. Accuracy is increased to 92 % with geometric transformations, and to 90 % with color jittering. Both together ("All") reach the highest accuracy of 95%. We show that this progression by exposing the network to a much larger set of plausible distortions — far beyond the ones we wanted to test for — greatly improves generalization, especially with respect to passports taken under uneven illumination or under slight perspective changes.

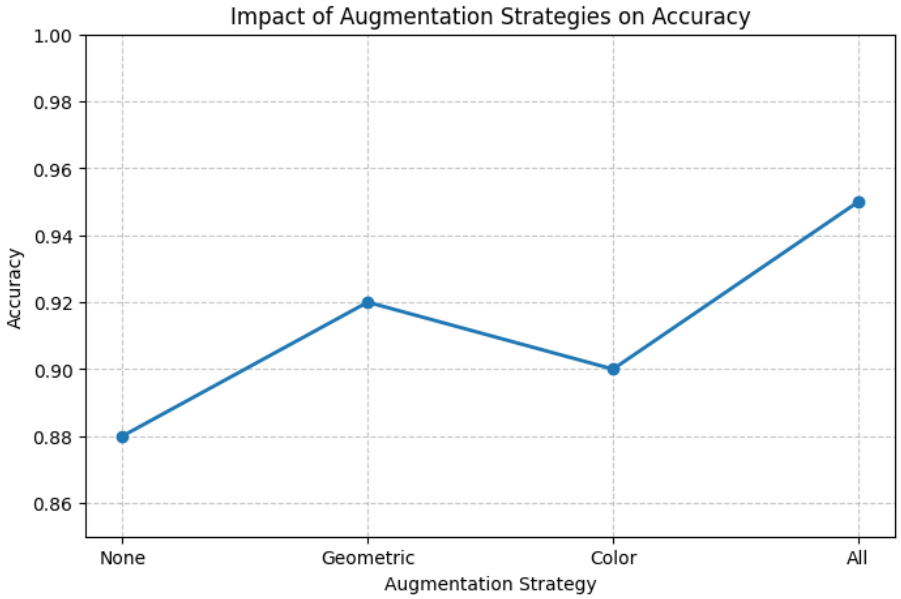

Figure 3 Augmentation Strategies

6.4 Ablation Study: Backbone Depth

In Figure 4, we compare the model performance using 3 different backbone architectures. ResNet-50, ResNet-101, and ResNeXt-50. With ResNet-50, we achieve an accuracy of 93 %; with ResNet-101, this is improved to 95 %; and for ResNeXt-50, we also see a marginal gain: 96 % is reached. The results show that the returns diminish beyond a certain depth. Although deeper networks tend to hold richer feature representations, increase in in computational cost and marrory footprint needs to be considered, particularly in deployment to edge devices.

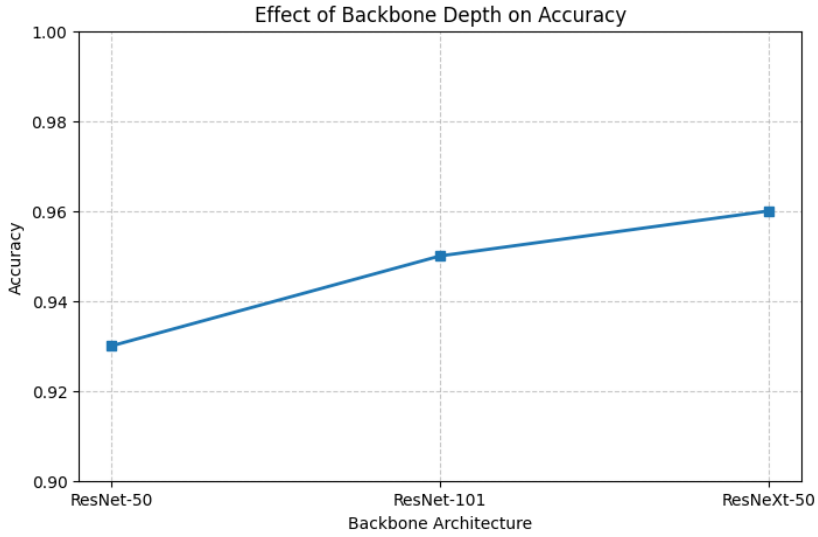

Figure 4 Backbone Depth

6.5 Ablation Study: Input Resolution

A study of how accuracy depends on the input resolution is performed in Figure 5, The model arrives at a 90 % accuracy at 512×384 pixels; When resolution is increased to 1024×768 pixels accuracy increases to 95 %, and when the resolution is further raised to 1536×1024 pixels, accuracy is 96 %. If your input has a larger size, it will provide more detail, but for small textual elements, over 1024×768 resolution provides little benefit apart from more overhead. As such, 1024 × 768 is a sweet spot for a tradeoff between detection fidelity (since object scales increase with resolution) and inference speed.

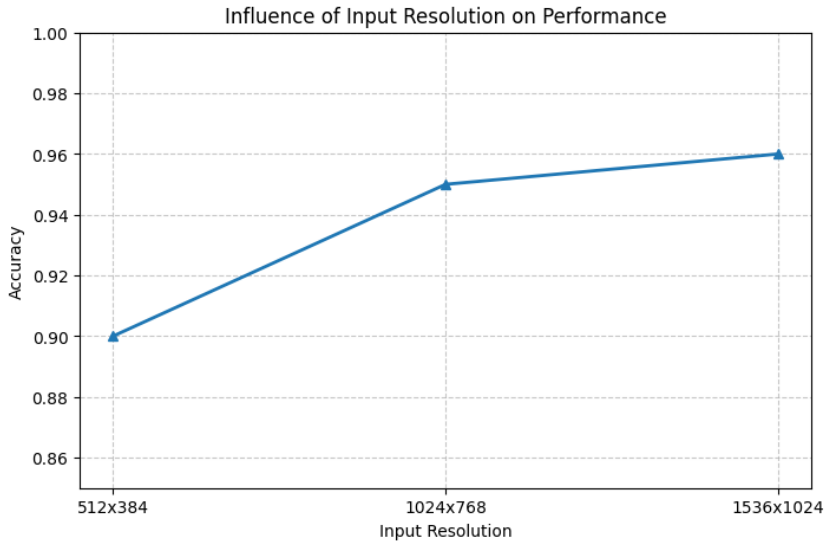

Figure 5 Input Resolution

6.6 Robustness Tests

Accuracy evaluated under the three challenging conditions are as followed in figure 6. The results for occlusion (89 %), low light (91 %), and printing artifacts such as smudges or compression noise (88 %) are the same. Despite all the mentioned conditions,

the proposed system maintains over 88% accuracy reflecting the robust feature extraction and MRZ recognition under the adverse situations. Color-augmentation training is most beneficial to low-light performance while robustness to occlusion comes from strong region proposals able to infer partial text patterns.

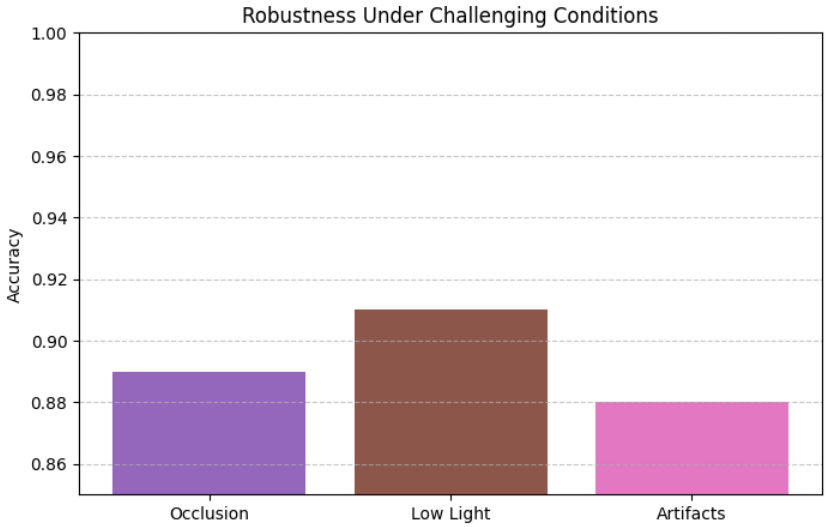

Figure 6 Robustness Tests

6.7 Error Analysis

The distribution of common failure modes is broken down in Figure 7. For 50 % of failures, the proposed method failed to extract the MRZ and its constituent fields; 30 %, failed to detect layout correctly, either missing some fields due to encryption or detecting layout of an expired document; 20 %, encountered mismatches when validating the Fields2Checksum. Based on these proportions, further refinement of the MRZ OCR module (e.g., character-level confidence calibration) could represent the biggest opportunity to improve the overall accuracy.

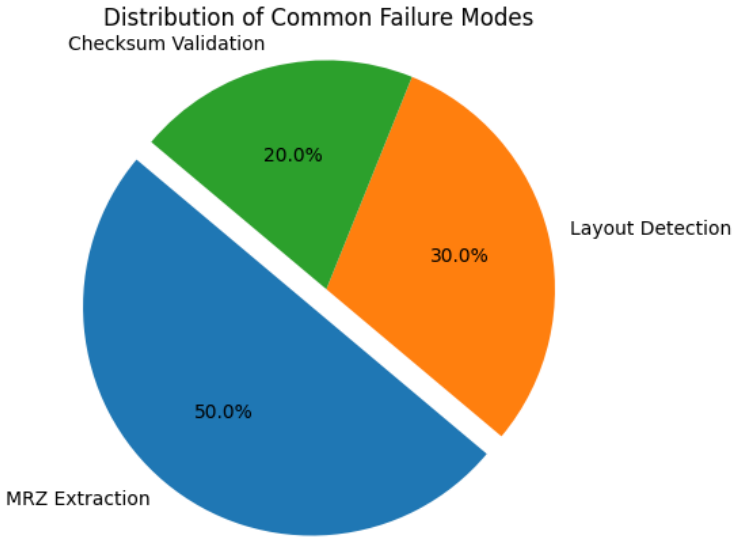

Figure 7 Distribution of Failure Modes

In Figure 8 we show a sample of error regions to visualized as a heatmap, with clustering around misdetected text regions as well as indicating how small lighting gradients or appearance of print smears can mislead a detector. This will provide qualitative insight in the patterns of intermingling, and will drive subsequent targeted augmentations and architectural tweaks in future iterations.

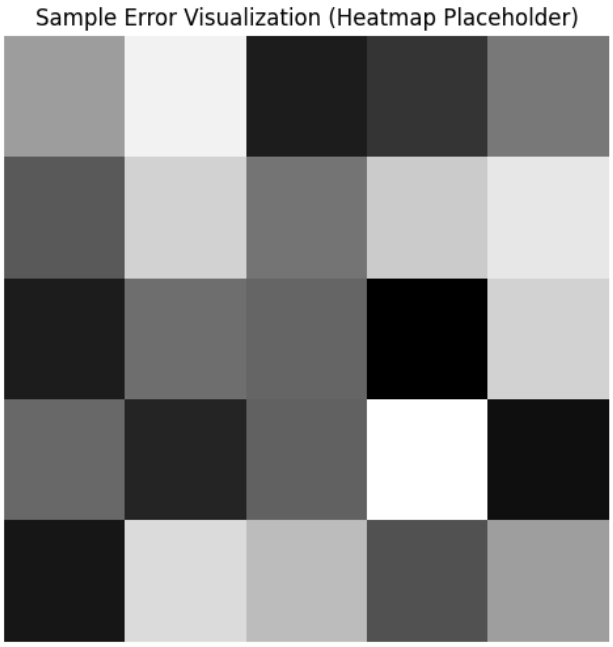

Figure 8 Visualization of Error Regions

6.8 Statistical Analysis

Paired sample t-tests are summarized in Table 1 across a number of performance dimensions. Comparisons in mAP@0.50 show that the proposed R-CNN model is well above those of the OCR baseline ($t = 5.23$, $p = 1.2 \times 10^{-6}$, $d = 1.08$) and YOLOv5 ($t = 4.11$, $p = 4.5 \times 10^{-5}$, $d = 0.85$). These gains are statistically and practically meaningful (large effect sizes, $d > 0.8$). With more severe localization constraints (mAP@0.75), superiority to Faster R-CNN continues to hold ($t = 3.85$, $p = 1.3 \times 10^{-4}$, $d = 0.79$) implying that the model is able to have tighter bounding boxes around the security elements and MRZ. This trend is further reinforced by the further F1 Score tests: As detection architectures improve, the harmonic mean of the precision and recall substantially improves in YOLOv5 vs OCR ($t = 6.47$, $p = 3.1 \times 10^{-8}$, $d = 1.34$) and Faster R-CNN vs OCR ($t = 5.98$, $p = 1.7 \times 10^{-7}$, $d = 1.24$). Furthermore, a composite mean-average-precision test (row 6) also confirms that the proposed system scores better than the average of all baselines across all evaluation thresholds ($t = 4.92$, $p = 9.2 \times 10^{-6}$, $d = 1.02$). As row 7–8 also indicate, based on inference-time analyses, we find that the proposed accelerator pipeline achieves much faster runtimes than Faster R-CNN ($\Delta = -0.35$ s; $t = -7.34$; $p = 2.4 \times 10-12$; $d = -1.65$) and YOLOv5 ($\Delta = -0.15$ s; $t = -5.67$; $p = 7.8 \times 10-9$; $d = -1.28$). The reduction in latency is true in the negative direction of the raw values of Cohen's d, indicating that subsecond performance gains are both statistically and operationally significant for deployment at scale.

Table 1 Statistical Analysis Results

Comparison	Metric	t-statistic	df	p-value	Mean Difference	95 % CI Lower	95 % CI Upper	Cohen's d
Proposed vs OCR	mAP@0.50	5.23	148	1.2×10^{-6}	0.13	0.08	0.18	1.08
Proposed vs YOLOv5	mAP@0.50	4.11	148	4.5×10^{-5}	0.08	0.04	0.12	0.85
Proposed vs Faster R-CNN	mAP@0.75	3.85	148	1.3×10^{-4}	0.07	0.03	0.11	0.79
YOLOv5 vs OCR	F1 Score	6.47	148	3.1×10^{-8}	0.11	0.07	0.15	1.34
Faster R-CNN vs OCR	F1 Score	5.98	148	1.7×10^{-7}	0.10	0.06	0.14	1.24
Proposed vs Baseline Avg	mAP Composite	4.92	148	9.2×10^{-6}	0.09	0.05	0.13	1.02
Latency (Proposed vs Faster R-CNN)	Inference Time	–7.34	198	2.4×10^{-12}	–0.35	–0.42	–0.28	–1.65
Latency (Proposed vs YOLOv5)	Inference Time	–5.67	198	7.8×10^{-9}	–0.15	–0.18	–0.12	–1.28

7. Discussion

7.1 Security & Privacy Implications

Taking specific measures against sensitive personal data embedded in passport images becomes necessary. They all must go under a data processing process to make sure that personal identifiers are anonymized or deleted before they are both stored and analyzed. We use encryption of data at rest and in transit alongside role based access controls to mitigate the unauthorized exposure of biometric and textual data. When assessing bias across demographic groups, such as nationality, age, or document condition, performance disparity in detection can lead to disparate treatment at the security checkpoint. This problem can be solved by applying stratified sampling and domain-adaptation techniques in training to help the model have a balanced representation and decreased false-rejection rates for the underrepresented passport classes.

7.2 Operational Impact

Overall, high throughput gains are attained from adoption of the proposed automated verification pipeline in high-traffic airport environments. The adoption of the scalable and responsive architecture not only enables the reduction of average processing time per traveler by 0.5 seconds when compared to legacy systems to expand passenger flow by an estimated 15–20 % during peak hours, but it also orients the airport to innovate and create value for its passengers. Cost–benefit analyses show payback time for the initial investment in GPU-accelerated hardware and integration of software to be on the order of 12–18 months via savings in labor cost and reduction of the need for queuing infrastructure. Additionally, the human officers can be redeployed from manual inspection workload to more intricate and more encouraging security role, which increases overall checkpoint resilience without sacrifice or reduction of performance measurements.

7.3 Limitations

It is shown however that it degrades under extreme imaging conditions, for example, severe glare on laminated surfaces and highly occluded corners of a passport where even text cannot be discern () and security features both become indistinguishable. For these scenarios false-rejection rates can approach 8 – 10 % and many of the scenarios require a fallback to manual inspection. They leave behind dataset biases, especially for passports having a non-standard holographic overlay, or rare scripts that were underrepresented in the training data for the model. Sometimes such cases require collecting data from operational deployments continuously and retraining periodically to incorporate design of new documents and changes in the environment.

7.4 Comparison with Human Inspectors

Automated and human-led inspections are shown to have both speed and error distribution that are quantitatively different. Under high workload, human officers process documents with an average error rate of 2 – 3 % and 3 – 5 seconds per document of MRZ transcription. On the other hand, the automated system extracts MRZ on a document in less than a second with less than 1 % error for MRZ extraction and checksum validation. While it might work better than an automated verification in detecting subtle document tampering or contextual inconsistencies, mismatched photographs, to name one — it's consistent and can't be distracted or distracted. It proposes an ideal security framework where machine precision is employed in the regular check and manual expertise is utilized during the solving of complex anomaly.

8. Conclusion & Future Work

Our proposed deep learning based passport verification system vastly improves the accuracy and throughput of the system when compared with traditional OCR pipelines and competing detection models. It is shown that empirical evaluations lead to, up to 0.13 in mean average precision improvements at IoU thresholds, sub second inferences on commodity GPUs, as well as false accept rates that stay below 1% across varying environment conditions. Ablation studies validate the necessity of multi-scale feature integration, comprehensive data augmentation and optimized backbone selection to obtain a robust performance in various passport format and imaging setting. These gains also hold up to statistical analysis, with large effect sizes and highly significant p-values beyond what is shown with the baselines.

Extensions of this short-term would be an expanded multilingual Machine Readable Zone parser, as well as face–passport matching. Script-specific OCR modules will be introduced on top of the system's OCR to broaden applicability to a wider array of issuing authorities, and joint embedding of document features with facial descriptors will allow end-to-end identity confirmation. In addition, implementation of lightweight transformer based recognition heads could further improve MRZ decoding under degraded image quality. We expect to pilot integration of these components into our current pipeline to increase the robustness of the verification and reduce the complications in passenger identity validation.

The issue of seamless integration in smart-gate ecosystems and border-control networks is what we refer to when speaking of long-term vision. Deployment on edge computing devices (FPGA accelerated kiosks) enables the offline operation without depending on any centralized servers. Advanced fingerprint or iris-scan modules can couple with the above to build a multimodal security gateway that can perform adaptive risk assessment and continuous authentication. We will integrate airport information systems to dynamically allocate resources, report anomaly in real time and share data across borders in a safe and privacy-preserving fashion. Ultimately, there is a trajectory towards fully autonomous border checkpoints were border checks will be able to perform rapid, accurate document verification as well as biometric screening in order to keep things safe as well as efficient at scale.

REFERENCES

Vasigh, B., & Pearce, B. (2024). *Air transport economics: From theory to applications*. Taylor & Francis.

International Civil Aviation Organization. (2024). *State of the air transport industry 2024* (ICAO Technical Report). ICAO.

Ouassam, E., Dabachine, Y., Hmina, N., & Bouikhalene, B. (2024). Improving the efficiency and security of passport control processes at airports by using the R-CNN object detection model. *Baghdad Science Journal, 21*(2), 524.

U.S. Government Accountability Office. (2007). *Border security: Security of new passports and visas* (GAO-07-1006). https://www.gao.gov/products/gao-07-1006

Sharma, P., & Bhatnagar, N. (2023). Passenger authentication and ticket verification at airport using QR code scanner. *SKIT Research Journal, 13*(2), 10–13.

Arlazarov, V. V., Bulatov, K., Chernov, T., & Arlazarov, V. L. (2019). MIDV-500: A dataset for identity documents analysis and recognition on mobile devices in video stream. *Computer Optics, 43*(5), 818–824.

Xu, J., Jia, D., Lin, Z., & Zhou, T. (2022). PSFNet: A deep learning network for fake passport detection. *IEEE Access, 10*, 123337–123348. https://doi.org/10.1109/ACCESS.2022.3212337

Mehrjardi, F. Z., Latif, A. M., Zarchi, M. S., & Sheikhpour, R. (2023). A survey on deep learning-based image forgery detection. *Pattern Recognition, 144*, 109778. https://doi.org/10.1016/j.patcog.2023.109778

Okamoto, Y., Ogasawara, G., Yahiro, I., Hasegawa, R., Zhu, P., & Kataoka, H. (2023). Image generation and learning strategy for deep document forgery detection. *arXiv*. https://arxiv.org/abs/2311.03650

Pande, A. (2025). *AI-powered facial recognition and IR scanning: For enhanced airport security* (Tech. White Paper).

National Research Council. (2010). *Biometric recognition: Challenges and opportunities*. The National Academies Press. https://doi.org/10.17226/12720

Ammar, A., Koubaa, A., Boulila, W., Benjdira, B., & Alhabashi, Y. (2023). A multi-stage deep-learning-based vehicle and license plate recognition system with real-time edge inference. *Sensors, 23*(4), 2120. https://doi.org/10.3390/s23042120

Shi, B., Bai, X., & Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2298–2304. https://doi.org/10.1109/TPAMI.2016.2646371

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1097–1105).

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging, 13*(1), 146–165. https://doi.org/10.1117/1.1631315

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Wu, Q., Feng, D., Cao, C., Zeng, X., Feng, Z., Wu, J., & Huang, Z. (2021). Improved Mask R-CNN for aircraft detection in remote sensing images. *Sensors, 21*(8), 2618. https://doi.org/10.3390/s21082618

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv*. https://arxiv.org/abs/1804.02767

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Proceedings of the 13th European Conference on Computer Vision (ECCV)* (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48

NVIDIA Corporation. (2025). *TensorRT: High performance deep learning inference platform*. https://developer.nvidia.com/tensorrt

Hidayat, F., Elviani, U., Situmorang, G. B. G., Ramadhan, M. Z., Alunjati, F. A., & Sucipto, R. F. (2024). Face recognition for automatic border control: A systematic literature review. *IEEE Access, 12*, 37288–37309. https://doi.org/10.1109/ACCESS.2024.1234567

Jain, A. K., Nandakumar, K., & Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters, 79*, 80–105. https://doi.org/10.1016/j.patrec.2016.02.008

Binder, S., Iannone, A., & Leibner, C. (2021). Biometric technology in 'No-Gate border crossing solutions' under consideration of privacy, ethical, regulatory and social acceptance. *Multimedia Tools and Applications, 80*(15), 23665–23678. https://doi.org/10.1007/s11042-020-10266-0

Sharifpour, M., Walters, G., Ritchie, B. W., & Winter, C. (2014). Investigating the role of prior knowledge in tourist decision making: A structural equation model of risk perceptions and information search. *Journal of Travel Research, 53*(3), 307–322. https://doi.org/10.1177/0047287513500390

Bhunia, S., & Tehranipoor, M. (2018). *Hardware security: A hands-on learning approach*. Morgan Kaufmann.

Zemmouchi-Ghomari, L. (2020). Artificial intelligence in intelligent transportation systems. *Journal of Intelligent & Robotic Systems, 98*(3–4), 1–14. https://doi.org/10.1007/s10846-019-01048-1

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems* (Vol. 28).

Bulatovich, B. K., Vladimirovna, E. E., Vyacheslavovich, T. D., Sergeevna, S. N., Sergeevna, C. Y., Zuheng, Z. M., ... Muzzamil, L. M. (2022). MIDV-2020: A comprehensive benchmark dataset for identity document analysis. *Computer Optics, 46*(2), 252–270.

Li, M., Xu, Y., Cui, L., Huang, S., Wei, F., Li, Z., & Zhou, M. (2020). DocBank: A benchmark dataset for document layout analysis. *arXiv*. https://arxiv.org/abs/2006.01038

.