

MobileMRZNet: Efficient and Lightweight MRZ Detection for Mobile Devices

Necmettin Bayar, Kubra Guzel and Deniz Kumlu


Abstract—The Machine Readable Zone (MRZ) is a standardized section found on identification documents (IDs) that adhere to the International Civil Aviation Association (ICAO) Document 9303. The MRZ region contains sensitive personal information about the document holder, and a portion of this information is utilized to establish communication between the passive chip within the ID and a mobile device via Near Field Communication (NFC) protocol. This communication is crucial as the data retrieved from the ID's chip is subsequently used in authentication steps such as face or fingerprint recognition. Thus, accurate detection of the personal information within the MRZ region is vital. In this research study, we propose a fast and lightweight approach for MRZ region detection called MobileMRZNet, which is based on the BlazeFace model. The MobileMRZNet architecture is specifically designed for mobile Graphical Processing Units (GPUs) and enables rapid and precise detection of MRZ regions. To train and evaluate the model, a dataset consisting of both simulated and real data was created using Turkish national IDs. The BlazeFace model was reconfigured and trained specifically for MRZ region detection. The detector, based on BlazeFace and trained on augmented real and simulated data, demonstrates excellent generalization capabilities for deployment with real IDs. Both qualitative and quantitative results confirm the superiority of our proposed method. The mean Intersection over Union (IoU) for the first frame, without utilizing any layout alignment for IDs, achieves an accuracy of approximately 81%. For character recognition, the method achieves 100% accuracy after three consecutive frames. The model operates in less than 10 milliseconds on a mobile device, and its size is around 400 KB, making it significantly fast, lightweight, and robust compared to any existing MRZ detection methods.

Index Terms—Biometric identification, Blazeface model, ID Cards, MRZ Detection, Travel Documents.


I. INTRODUCTION

IDENTITY DOCUMENTS (IDs) play a vital role in ensuring national security and enabling citizens to engage in


Necmettin Bayar, is with Artificial Intelligence Department, Kobil Technology, Istanbul, Turkey, (e-mail: necmettin.bayar@kobil.com).

 <https://orcid.org/0000-0003-2367-828X>

Kubra Guzel, is with Artificial Intelligence Department, Kobil Technology, Istanbul, Turkey, (e-mail: kubra.guzel@kobil.com).

 <https://orcid.org/0009-0002-8401-6983>

Deniz Kumlu, is with Artificial Intelligence Department, Kobil Technology, Istanbul, Turkey, (e-mail: deniz.kumlu@kobil.com).

 <https://orcid.org/0000-0002-7192-7466>

Manuscript received Jan 08, 2024; accepted Apr 15, 2024.

DOI: [10.17694/bajece.1416404](https://doi.org/10.17694/bajece.1416404)

various activities such as voting, traveling, and accessing government and private sector benefits. The accurate verification of IDs allows governments to effectively monitor their citizens, identify illegal immigrants, prevent identity theft, track criminals, monitor terrorism activities, and identify individuals who may pose a risk to national security. Moreover, governments provide services specifically tailored to their citizens, and IDs serve as a means to streamline service delivery by ensuring that only eligible citizens receive these services, minimizing any potential errors or misuse by non-citizens.

When the brief history of IDs are checked, it is evident that they first appeared around 1876. However, they did not become widely accessible until the early 20th century [1]. The introduction of photographic IDs occurred in 1915, following the well-known Lody-Spy scandal [2]. Prior to 1985, there was no global standardization for IDs. It was in 1985 that ISO/IEC 7810 established standardized guidelines for the shape, size, and content of IDs, which were further refined in 1988 through ISO/IEC 7816. As technology advanced, radio frequency identification (RFID) chips were integrated into IDs, enabling the storage of sensitive personal information alongside biometric data like photographs and fingerprints. The most recent standards for IDs are defined by the International Civil Aviation Association (ICAO), and the majority of countries worldwide have aligned their ID systems with these standards [3].

The International Civil Aviation Organization (ICAO) document 9303 establishes the standardized guidelines for machine-readable travel documents, including national IDs, passports, and more. Presently, over 190 countries have adopted these standards and incorporate machine-readable zones (MRZs) within their respective IDs.

The MRZ is a specific section on IDs that contains sensitive personal information about the document holder. It is designed to streamline and expedite the scanning process for government-issued documents such as IDs and passports. The information within the MRZ can be read using optical character recognition (OCR) methods, such as the Tesseract OCR engine, as the characters in the MRZ region have a unique font called OCR-B. To enhance the accuracy and performance of the OCR engine, it is crucial to accurately detect the exact MRZ region on IDs beforehand [4].

The MRZ region is commonly used as part of the authentication process, with some of its information utilized to access the chip within IDs via the Near Field Communication (NFC) module of mobile phones. NFC comprises a collection of short-range wireless technologies that establish a reliable

connection between the passive chip on IDs and mobile devices, typically requiring a distance of 4cm or less [5]. To initiate communication via NFC and retrieve sensitive personal or biometric information from the chip, control over the data written on the ID's MRZ region, such as the date of birth, document number, and expiration date, is necessary. Once this information is provided to the chip, a communication process known as the hand-shake protocol begins, allowing the extraction of highly secure personal information, facial biometric images, and fingerprints from the ID's chip. This information is highly reliable for identity verification, as it is difficult to manipulate, tamper with, or falsify the data stored within the chip. Additionally, passive and active authentication processes implemented on the chip help ensure the integrity of the information, guarding against cloning, tampering, and manipulation [6]. As a result, the use of NFC-based IDs authentication has gained popularity, particularly in financial and banking mobile applications.

Several methods have been proposed for MRZ detection and character reading [7-17]. The initial method lacked a dedicated detection module and required users to manually align their IDs with a specific template on the device screen, followed by manual cropping to obtain the MRZ region. However, this heuristic approach heavily relied on the user's ability to align the ID properly, leading to inaccuracies if alignment was not precise [7]. Subsequently, methods were introduced that employed binarization of the MRZ region using adaptive thresholding. Horizontal and vertical histogram projections were utilized to decompose the document image into character parts, and character recognition algorithms were applied to extract the desired information [8, 9]. However, these methods were found to be less robust in scenarios with nonuniform illumination and situations involving occlusion, making them unsuitable for sensitive mobile applications. In a different approach, a combination of Fuzzy Adaptive Resonance Theory (ART)-based Radial Basis Function (RBF) network and Principal Component Analysis (PCA) algorithm was proposed for passport recognition [10]. This method involved isolation and connected component analysis, filtering and selecting clustered lines of text, and converting the input image to a binary image prior to component search. However, this method relied on video frames and fusion of results from multiple frames, which increased processing time. Furthermore, the complexity of the algorithm made it unsuitable for mobile implementation, and it was specifically designed for passport documents.

Neural network (NN) based models have been proposed for MRZ detection as well. In a study by Khan et al. [11], a combination of Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) was employed for accurate recognition of the passport MRZ region, even when dealing with passport images of varying sizes and skewed orientations. To achieve effective character segmentation, the method included MRZ line detection using a connected component analysis algorithm and skew correction using a perspective transform algorithm. The results of this approach showed

promise, but it should be noted that the method is complex and specifically applied to passport images, limiting its applicability to other types of IDs.

A recent MRZ region detection model named MRZNet, proposed by Li et al. [12], has demonstrated remarkable performance compared to other existing MRZ detection models. MRZNet is based on MobileNetV2 architecture [13]. The authors conducted a comprehensive comparison with Tesseract-based methods [14], a deep learning-based commercial solution [15, 16], and end-to-end NN based text spotting approaches [17, 18]. They reported that their method outperforms existing solutions by a significant margin and can effectively extract MRZ information from passports of various sizes and content. However, similar to other existing solutions, the running time of the method and the size of the model may not be optimal for mobile device implementation. Additionally, it should be noted that the evaluation of MRZNet's results was specifically focused on passports and not on other types of travel documents, limiting its generalizability to those document types.

Our proposed model has been rigorously compared with two well-known models, PassportEye [14] and UltimateMRZ [15] utilizing two publicly available datasets (MIDV-500 [19] and MIDV-2019 [20]), in addition to a dataset specifically generated for this study. These comprehensive comparisons offer insights into the performance of our model in terms of character error rate (CER), model size, and inference time, highlighting its capabilities and advantages for the mobile platforms.

A. Types of Graphics

The MRZ is a designated section within IDs that holds sensitive personal information of the document holder. Currently, the majority of countries incorporate MRZ regions in their IDs. The specifications and formats of the MRZ region are defined by the ICAO document 9303 [3]. The specific layout and content of the MRZ region may vary depending on the type of document, such as passports, national IDs, or other travel documents.

- "Type 1" refers to a common format for national IDs, which are typically similar in size to credit cards. In this format, the MRZ region of the ID contains three lines, each consisting of 30 characters.
- "Type 2" refers to a less common format for IDs, which deviates from the typical credit card-size. In this format, the MRZ region of the ID contains two lines, and each line consists of 36 characters.
- "Type 3" refers to the standard format used for passports. In this format, the MRZ region of the passport contains two lines, with each line consisting of 44 characters.

The commonly used ID formats are Type 1 and Type 3 IDs. Type 1 IDs is 85.6×54.0 mm (3.37×2.13 in) in size such as credit cards. The MRZ data in Type 1 IDs consists of three rows, with each row containing 30 characters. The characters used in the MRZ region are limited to uppercase letters "A-Z", digits "0-9", and the filler character "<". Type 3 is a passport

detector (SSD) models [23]. The first model is exploited for feature extraction and the second model is exploited for GPU-friendly anchor scheme. In addition, alternative approach is applied to the non-maximum suppression (NMS) algorithm for additional refinement at the final stage. BlazeFace is primarily used for face detection purposes and it is also combined with mobilefacenets [24] in the recent study [25].

Many models are using traditional convolution networks (generally, it is 3×3 kernel size) on the input image and the computational cost can be significant. From the computational view, input image has a size $s \times s \times c$, where s is the image dimensions and c is the number of channel and the convolution kernel is $k \times k \times d$ where k is the kernel dimension and d is the number of kernels. Then, total number of operations will be ds^2ck^2 for the normal convolution case. In MobileNetV1/V2 depth-wise separable convolution computation is used which consists of depth-wise and point-wise convolutions. For the same size input image and kernel like in the normal convolution case, depth-wise and point-wise convolutions' computations will become s^2ck^2 and s^2cd , respectively. Thus, total computation for the depth-wise separable convolution will be $s^2ck^2 + s^2cd$ and their ratio becomes factor of $\frac{d}{s^2}$. There is factor of $\frac{k^2d}{k^2+d}$ decrease between the normal convolution and depth-wise separable convolution [22]. This reduction in computation makes the MobileNetV1/V2 models more computationally efficient while still maintaining good performance. It allows for faster inference and makes these models suitable for resource-constrained environments, such as mobile devices.

In addition, between the depth-wise and point-wise convolution operations, the latter one takes more time not only due to the arithmetic operations but also due to fixed costs and memory access factors. Thus, increasing the kernel size of the depth-wise part is cheap and does not explode running time of the model. In [22], the use of a 5×5 kernel size in the depth-wise convolution operation has been shown to be effective in increasing the receptive field size and improving the performance of the model. This approach provides a balance between computational efficiency and capturing more comprehensive information from the input data. By optimizing the kernel sizes and balancing computational costs, the BlazeFace model achieves a good trade-off between model efficiency and receptive field enlargement, making it suitable for real-time applications on mobile devices.

To further accelerate the progression of the receptive field size, a double BlazeBlock is introduced in addition to the single BlazeBlock. The double BlazeBlock is designed to effectively capture larger contextual information and enhance the understanding of complex visual patterns. The double BlazeBlock consists of two consecutive sets of convolutional layers, depth-wise separable convolutions, and non-linear activation functions. Each set of convolutions is followed by a down-sampling operation, typically achieved through max pooling or strided convolutions, which reduces the spatial dimensions of the feature maps. By stacking two BlazeBlocks together, the receptive field of the model is increased, allowing it to capture more global information and contextual dependencies. The double BlazeBlock architecture enables the

network to learn more robust and discriminative features, enhancing its ability to accurately detect and classify objects. This progression in receptive field size is particularly beneficial for tasks that involve objects with varying scales or complex spatial relationships. Both single and double Blazeblock architectures are shown in Fig. 2

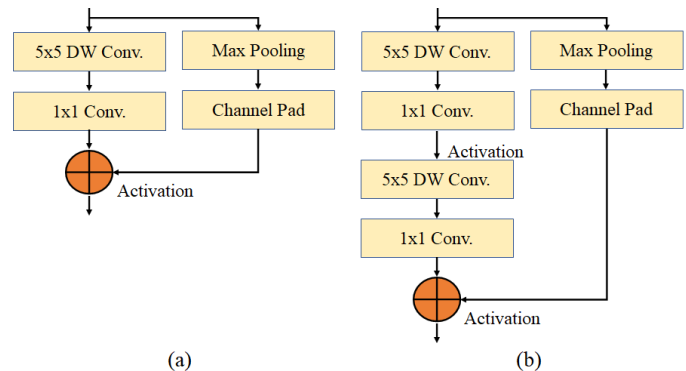


Fig.2. (a) Single Blazeblock, (b) double Blazeblock.

By incorporating the double BlazeBlock into the network architecture, the model can effectively handle larger and more diverse visual inputs, making it suitable for applications such as object detection, scene understanding, and image classification.

For the feature extraction part, the input RGB image (face image) is 128×128 , BlazeFace model architecture contains 5 single BlazeBlocks and 6 Double BlazeBlocks. The highest channel resolution is 96 and the lowest spatial resolution is 8×8 .

In addition, the classical non-maximum suppression algorithm (the final bounding box depends only one of the candidate anchor) is replaced with blending strategy that predicts the regression parameters of bounding boxes as weighted mean between overlapping estimations. In [21], it is mentioned that this replacement brought 10% increase in the accuracy for the detection results.

III. PROPOSED METHOD

The BlazeFace model [21], originally developed for face detection and facial landmark detection on mobile devices, has proven to be effective with its lightweight design and high-speed architecture. It performs comparably well to other popular face detection models while offering the advantages of its lightweight nature. Leveraging the performance of the BlazeFace model in face detection, we have modified it to detect the MRZ region in IDs.

In our modified model which is shown in Fig. 3, the layers of the BlazeFace model remain the same, but the input sizes have been increased from 128 to 320. This allows for the collection of features from the 9th layer output and the last layer output, increasing the total anchor count. While a lower anchor count may reduce inference time, it can also compromise model performance. Thus, anchor counts and schemes have been updated to achieve better detection performance.

Typically, face detection models assume that the aspect ratio of the detected face's width and height is equal to 1, and prior boxes are set accordingly. However, in the case of the MRZ region, it has been determined that the aspect ratio for width to height is approximately 4. Therefore, the aspect ratio value for

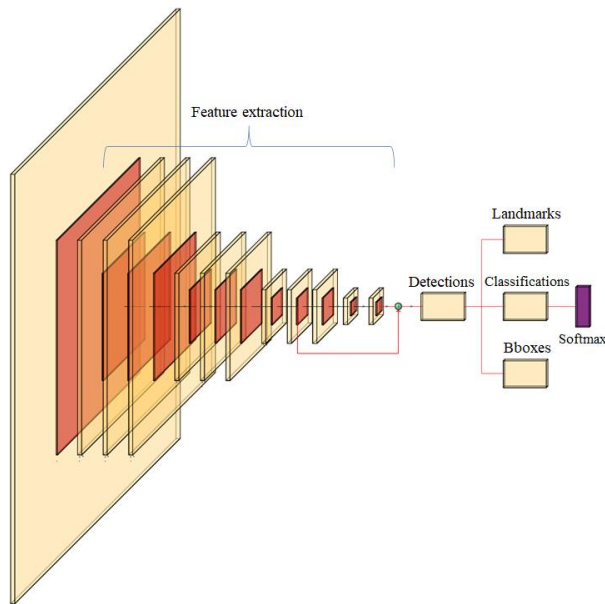


Fig.3. Block diagram of our proposed model.

the prior boxes has been changed to 4, and specialized anchor calculations have been performed to improve the results. In the experimental results, it has been observed that setting the aspect ratio as 3 performs better than 4 due to an additional non-maximum suppression step in the final stage. A higher aspect ratio may result in wider bounding boxes, which can negatively impact subsequent processes on the detected region.

Another difference between the face detection model and the MRZ region detection model is that the MRZ region model does not require the detection of landmarks. Therefore, the landmark layer can be dropped to reduce model weight and increase inference speed.

Overall, the modified MRZ region detection model based on BlazeFace offers accurate and efficient detection of the MRZ region in IDs, leveraging the strengths of the original BlazeFace model while making necessary adjustments for the specific requirements of MRZ region detection.

IV. EXPERIMENTAL RESULTS

In the experimental part of this study, BlazeFace model [21] was trained on a dataset comprising real and simulated IDs. The model was then evaluated on a range of test datasets, including simulated IDs, real IDs, and publicly available ID samples. We conducted both quantitative and visual analyses to assess the performance of our proposed model in detecting the MRZ region.

Quantitative results were obtained by measuring various performance metrics such as accuracy, precision, recall, and F1 score. These metrics provide objective measures of the model's ability to accurately detect the MRZ region in different types of IDs.

Additionally, we presented visual results to demonstrate the effectiveness of our proposed model. This involved showcasing the model's detection results on sample ID images, highlighting the accurately detected MRZ regions. Visual representations help provide a clear understanding of how well the model

performs in identifying the MRZ region within various ID samples.

By combining quantitative analysis with visual evidence, it is aimed to provide a comprehensive evaluation of our proposed model's MRZ region detection performance. These results contribute to validating the accuracy and effectiveness of our model in reliably detecting the MRZ region in IDs, regardless of whether they are simulated, real, or sourced from publicly available samples.



Fig.4. (a) Blank Turkish national ID, (b) randomly filled Turkish national ID.

A. Simulated Dataset

To address the security concerns surrounding sensitive personal information on national IDs, the availability of public datasets or real IDs for research purposes is limited. In order to overcome this limitation, we have created a new simulated dataset specifically tailored for national Turkish IDs, while ensuring compatibility with other national IDs adhering to ICAO standards [3].

During the construction of the simulated dataset, the static text was kept consistent across all IDs, only modifying the dynamic text components. Certain information on the IDs, such as the date of birth, expiration date, and national identification number, maintain a fixed length of characters. However, other text fields, such as the name and surname, vary in length. To account for this variability, we randomly generate the length of the name and surname using a Gaussian distribution to simulate realistic variations.

By utilizing this simulated dataset, we are able to train and evaluate our MRZ detection model effectively while respecting privacy and security concerns associated with real IDs. The simulated dataset allows us to capture the essential characteristics and variability of national IDs, enabling robust and accurate performance assessment of our proposed model.

Firstly, the dynamic text part of the ID, which contains sensitive information, is removed using an inpainting algorithm. This algorithm fills in the areas corresponding to the dynamic text with appropriate background patterns, resulting in an empty national Turkish ID as depicted in Fig. 4(a). Next, the empty ID is populated with randomly selected digits, letters, and the "<" symbol to replicate the actual content and order of the dynamic text. This process ensures that the simulated data closely resembles the characteristics and structure of real IDs. The resulting simulated data for the ID is illustrated in Fig. (b).

By employing this method, diverse range of simulated IDs were generated while preserving the privacy and security of individuals. The inpainting algorithm enables the removal of sensitive information, and the subsequent random filling process ensures the variability and realism of the simulated data.



Fig.5. Sample Turkish national IDs with different backgrounds in the simulated dataset.



Fig.6. MRZ detection results for the simulated Turkish national IDs.

To create a diverse and comprehensive simulated dataset, 40 different fake national Turkish IDs are randomly generated based on the template shown in Fig. (b). These simulated IDs are then printed with high quality to resemble actual physical IDs. In order to capture variations in image quality, background, and hand grip positions, multiple photos of each

printed ID are taken using different scenarios. This is done using three different mobile devices: Samsung, iPhone, and Xiaomi. As a result, a total of 1,300 photos are obtained, covering approximately 50 different backgrounds. To further enhance the variability and robustness of the dataset, data augmentation techniques are applied. This includes

manipulating factors such as color, illumination, sharpness, blurriness, and other relevant parameters. Through data augmentation, the size of the simulated dataset is increased to 130,000 IDs. Fig. 5 illustrates different sample IDs from our simulated dataset, showcasing the diversity and realism achieved through the combination of random generation, high-quality printing, diverse photo scenarios, and data augmentation techniques.

B. Real Dataset

To address the limitations of using only simulated data, we recognize the importance of incorporating real data into the training dataset. However, due to security concerns and the sensitive nature of personal information on real IDs, it is challenging to find publicly available real ID datasets.

In this study, we managed to obtain a small set of 4 different genuine national Turkish IDs. These IDs were recorded in video format, capturing them in various scenarios with 100 different backgrounds. From each recorded video, 7 frames were extracted, resulting in a total of 700 real ID images. To increase the diversity and quantity of the real dataset, data augmentation techniques were applied to the extracted real ID images. This augmented the total number of real IDs to 70,000, providing a more comprehensive training dataset.

While increasing the number of real data can be beneficial for model performance, it was necessary to keep the quantity limited due to security concerns and the sensitive nature of the information contained in real IDs. By carefully balancing the inclusion of real and simulated data, it is aimed to enhance the quality, quantity, and diversity of the training dataset while ensuring the privacy and security of individuals' personal information.

C. Training Configuration

To enhance the quality, quantity, and diversity of the training data, we combined the simulated dataset with the limited real dataset. This approach is commonly used in the literature to improve model performance. The real dataset consisted of a small number of unique identity cards, which were augmented to increase the data diversity.

The initial training dataset comprised 2000 images before augmentation. Data augmentation techniques were applied to both the simulated and real datasets, resulting in a final dataset of approximately 200,000 images. The augmentation numbers were optimized by assessing the model's performance.

During training, an initial learning rate of 0.001 was set, and four learning rate decay steps were implemented over the course of 200 epochs. The input images had three channels and a size of 320×320 pixels. A batch size of 512 was used in training. Three different optimizers, namely Adam, stochastic gradient descent (SGD), and root mean squared propagation (RMSProp), were tested. The best results were achieved with RMSProp.

Model performance evaluation was conducted at different intervals during the training process. Up to the 100th epoch, evaluation was performed every 10 epochs, and after that, it was done every 5 epochs.

The training process was carried out on a workstation equipped with an Ubuntu 20.04 operating system, 128 GB of RAM, and 2 Nvidia Rtx A5000 GPUs. The total training time

for the approximately 200,000 images was approximately 4 days. It was observed that the model's performance reached a convergence point after the 130th epoch, and there was no significant improvement beyond that point.

D. Detection Results

In this study, BlazeFace model [21] was trained for the MRZ detection for the first time. For the training part, simulated and real ID card dataset is constructed where the following scenario is tested;

- Scenario: Training on the simulated and real Turkish national IDs dataset and testing on the simulated, real and publicly available IDs.

The performance of the proposed model was evaluated using both visual and quantitative measures. The visual results are presented in Fig. 6, and Fig. 7 (a)-(c), showcasing the MRZ region detection for different types of IDs, including simulated IDs, real Turkish national IDs, and IDs, passports, and visas from various countries.

Fig. 6 demonstrates the successful MRZ region detection for simulated IDs with different backgrounds. The bounding boxes accurately enclose the MRZ region, ensuring that all relevant text information is contained within them. This allows for subsequent processing with OCR engines.

Fig. 7(a) shows the visual results for real Turkish national IDs, with sensitive personal information hidden for security reasons. The bounding boxes are well-fitted to the MRZ region, even in cases with different orientations, distances, and backgrounds. The model is capable of handling challenging scenarios such as IDs held by hand, where the shape of the ID card may be distorted or occluded. Unlike methods that rely on the shape of the ID card, our model focuses solely on the MRZ region, enabling successful detection even in distorted or occluded cases.

Figs. 7(b) and (c) present the MRZ region detection results for IDs, passports, and visas from different countries worldwide. Despite the training dataset being solely based on Turkish IDs, the proposed model exhibits excellent generalization ability. It can successfully detect the MRZ region in various travel documents and IDs from different countries, demonstrating its superior performance and generalization capability.

Overall, the visual results demonstrate the effectiveness and robustness of the proposed model in detecting the MRZ region in a wide range of ID types and scenarios.

Quantitative results were also obtained to evaluate the performance of the model. Fig. 8 illustrates the average precision, which exceeds 96%. This high precision score is achieved by setting a high confidence threshold during the testing stage. Despite the high threshold, there are very few cases of missed detection. The precision-recall curve shows that the model maintains a consistently high precision even as the recall increases, indicating its robustness in detecting the MRZ region.

Accurate localization of the MRZ region is crucial for the OCR process. To evaluate this aspect, the Accuracy vs. IoU threshold is presented in Fig. 9.



(a)



(b)



(c)

Fig.7. MRZ detection results (a) the real Turkish national IDs, (b) publicly available ID cards for different nations, and (c) publicly available other travel documents such as passports and visas.

The red point on the graph shows that our model achieves an accuracy above 80% when the IoU threshold is set at 0.75. Our experimental results indicate that the Tesseract OCR engine performs well with MRZ regions that have an IoU threshold of 0.75 or higher, based on the average results of our tests.

80%. This indicates that using templates for IDs allows for more accurate MRZ region detection, especially in the first few frames and often in the first frame.

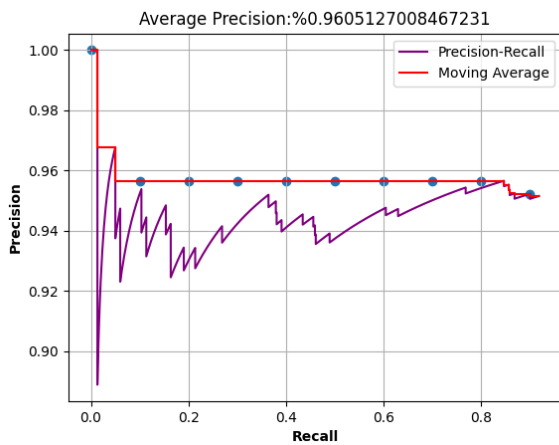


Fig.8. Precision vs Recall curve of trained model with average precision.

Furthermore, the model's performance was assessed in capturing MRZ regions in the wild using a mobile device camera. Fig. 10 demonstrates that even without aligning to a specific template, our model achieves an accuracy rate of over

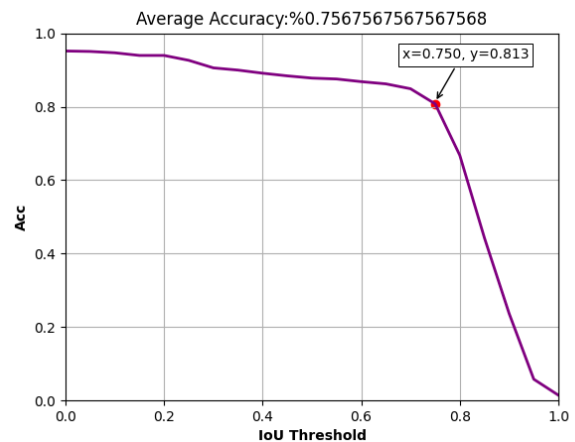


Fig.9. Accuracy vs. IoU threshold curve of trained model with average accuracy.

In summary, the quantitative results confirm the high performance and accuracy of the proposed model in detecting the MRZ region, and its compatibility with OCR engines for extracting text information from the MRZ regions of various IDs captured by mobile devices.

E. Quantitative Analysis and Comparison

For this experiment, we employed MIDV-500 [19] and MIDV-2019 [20] datasets, and our dataset, comparing the results with existing models from the literature. While MIDV-500 and MIDV-2019 include passport MRZ regions, our model is solely trained on Turkish ID cards. Consequently, the results obtained on these datasets demonstrate the generalizability and cross-dataset performance of our proposed model. It's worth noting that PassportEye is designed for use with passports exclusively, and therefore, it did not yield results in our ID card dataset. As for PassportEye [14] and UltimateMRZ [15], we utilized the pre-trained models as provided in their original publications. The performance comparisons are tested over videos since for our use case, video clip of the ID card is taken until it recognizes the characters in the MRZ region. For each individual video, if character error rate (CER) is lower than 50% for one of the frame, it is considered as a successful detection.

The results are outlined in Table I. In publicly available datasets like MIDV-500 and MIDV-2019, UltimateMRZ obtains the top position, with PassportEye following closely. In these datasets, our proposed model exhibits lower performance compared to these two models. Several factors contribute to this difference. 1) Our model is primarily trained on ID card MRZ regions, and while it generalizes well to passports, it has not been exposed to passport MRZ regions during training. 2) We deliberately excluded extreme cases in MRZ videos from our dataset. Since our mobile application guides users to capture videos of their ID cards with a good quality. This guided approach streamlines our model's architecture but may result in reduced performance when dealing with more diverse angles and scenarios. In our dataset, our model outperforms UltimateMRZ. PassportEye, designed primarily for passports, doesn't yield valid results with ID cards, as expected. Since our dataset only contains ID card videos, it's reasonable to conclude that incorporating passport images into our training data could potentially lead to competitive results for datasets like MIDV-500 and MIDV-2019.

TABLE I
MRZ DETECTION RESULTS (%) BASED ON CER

Dataset	UltimateMRZ	PassportEye	MobileMRZNet
MIDV-500	93.33	90.66	90
MIDV-2019	90	88.33	75
Ours	88.46	-	90.80

In Table II, we provide a comparison of model sizes and inference times. All methods were tested on the following hardware and environment: Intel Core i7-7820HQ @ 2.9GHz, 16GB DDR4-2400 MHz, NVIDIA Quadro M620, on a Windows 10 64-bit platform. Our proposed model is significantly more lightweight than UltimateMRZ, being nearly 2.5 times smaller, and it surpasses PassportEye by a factor of 50 in terms of model size. Furthermore, for inference time comparison, our proposed model outperforms both UltimateMRZ and PassportEye, with the latter exhibiting a particularly substantial difference in running-time performance.

The results presented in both Table I and Table II provide strong evidence supporting the suitability of our proposed

methods for mobile platforms, aligning with the claims made in this study.

TABLE II
MODEL SIZE AND INFERENCE TIME COMPARISON

Model	Size (MB)	Time (Sec.)
UltimateMRZ	2.6	0.16
PassportEye	23.5	1.11
MobileMRZNet	1.06	0.12

V. CONCLUSION

The proposed MRZ region detection model offers an efficient and lightweight solution specifically tailored for mobile devices. The construction of a realistic simulated dataset addresses the challenge of limited publicly available MRZ datasets, considering the need to protect sensitive information. By training the BlazeFace model on the combined real and simulated IDs dataset, the proposed model outperforms existing MRZ region detection models in terms of detection performance, robustness, and computational efficiency. The model is well-suited for mobile devices and demonstrates generalization capabilities to other travel documents such as passports and visas from different countries. Overall, the proposed model represents a significant advancement in MRZ region detection, offering improved accuracy and efficiency for various applications.

VI. ACKNOWLEDGEMENT

This work was supported by TUBITAK-TEYDEB under Project No. 3201086, and actively use in Know Your Customer (KYC) product by KOBIL Technology LTD. STI.

REFERENCES

- [1] R. Hall, G. Dodds, S. Triggs, "The World of William Notman," David R. Godine, pp. 46-47, 1993.
- [2] J. Douman, D. Lee, "Every Assistance & Protection: A History of the Australian Passport," Federation Press, p. 56, 2008.
- [3] International Civil Aviation Organization (ICAO), "Machine Readable Travel," 7th ed., Parts 2-7, 2015.
- [4] J. Monnerat, S. Vaudenay, M. Vuagnoux, "About machine-readable travel documents," Springer, 2007.
- [5] R. Want, "Near field communication," IEEE Pervasive Computing, vol. 10, no. 3, pp. 4-7, 2011.
- [6] P. K. Chan, C. S. Choy, C. F. Chan, K. P. Pun, "Preparing smartcard for the future: from passive to active," IEEE Transactions on Consumer Electronics, vol. 50, no. 1, pp. 245-250, 2004.
- [7] A. Hartl, C. Arth, D. Schmalstieg, "Real-time detection and recognition of machine-readable zones with mobile devices," 10th International Conference on Computer Vision Theory and Applications, vol. 3, pp. 79-87, 2015.
- [8] Y. V. Visilter, S. Y. Zheltov, A. A. Lukin, "Development of OCR system for portable passport and visa reader," In Document Recognition and Retrieval VI, Vol. 3651, pp. 194-199, 1999.
- [9] Y.-B. Kwon, J.-H. Kim, "Recognition based verification for the machine-readable travel documents," In 7th International Workshop on Graphics Recognition, pp. 1-10, 2007.
- [10] K.-B. Kim, S. Kim, "A passport recognition and face verification using enhanced fuzzy ART based RBF network and PCA algorithm," Neurocomputing, vol. 71, pp. 3202-3210, 2008.
- [11] J. Kim, "Recognition of Passport MRZ Information Using Combined Neural Networks," Journal of Korea Society of Digital Industry and Information Management, vol. 15, no. 4, pp. 149-157, 2019.
- [12] Y. Liu, H. James, O. Gupta, D. Raviv, "MRZ code extraction from visa and passport documents using convolutional neural networks," International Journal on Document Analysis and Recognition (IJ DAR), vol. 25, no.1, pp. 29-39, 2022.

- [13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520, 2018.
- [14] K. Tretyakov, "PassportEye: Extraction of machine-readable zone information from passports, visas, and ID-cards via OCR," GitHub, 2016. [Online]. Available: <https://github.com/konstantint/PassportEye>.
- [15] UltimateMRZ, doubango.org, 2020. [Online]. Available: <https://github.com/DoubangoTelecom/ultimateMRZ-SDK>. Accessed: Oct. 01, 2023.
- [16] D. Kostro, M. Zasso, "MRZ detection via image-js," GitHub, 2020. [Online]. Available: <https://github.com/image-js/mrz-detection>.
- [17] P. Lyu, M. Liao, C. Yao, W. Wu, X. Bai, "Masktextspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes," In Proceedings of the European Conference on Computer Vision (ECCV), pp. 67-83, 2018.
- [18] L. Xing, Z. Tian, W. Huang, M. R. Scott, "Convolutional Character Networks," arXiv:1910.07954, 2019.
- [19] V. V. Arlazarov, K. Bulatov, T. Chernov, V. L. Arlazarov, "MIDV-500: A dataset for identity document analysis and recognition on mobile devices in video stream," Computer Optics, vol. 43, pp. 818-824, 2019.
- [20] K. Bulatov, D. Matalov, V. V. Arlazarov, "MIDV-2019: Challenges of the Modern Mobile-Based Document OCR," ICMV 2019, pp. 114332N1-114332N6, 2020.
- [21] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, M. Grundmann, "BlazeFace: Sub-millisecond neural face detection on mobile GPUs," arXiv preprint arXiv:1907.05047, 2019.
- [22] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
- [23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, "SSD: Single shot multibox detector," In European Conference on Computer Vision, pp. 21-37, Springer, Cham, 2016.
- [24] S. Chen, Y. Liu, X. Gao, Z. Han, "Mobilefacenet: Efficient CNNs for accurate real-time face verification on mobile devices," In Chinese Conference on Biometric Recognition, pp. 428-438, 2018.
- [25] N. Bayar, K. Güzel, D. Kumlu, "A Novel BlazeFace Based Pre-processing for MobileFaceNet in Face Verification," 45th International Conference on Telecommunications and Signal Processing (TSP), pp. 179-182, 2022. M. Yilmaz, "The Prediction of Electrical Vehicles' Growth Rate and Management of Electrical Energy Demand in Turkey," Green Technologies Conference (GreenTech), 2017 Ninth Annual IEEE. Denver, US, 2017.



Deniz Kumlu received B.S. degree from the Department of Electrical and Electronics Engineering, Turkish Naval Academy, Istanbul, Turkey, in 2007 with honor degree, M.S. degree from the Ming Hsieh Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA, in 2012, where he studied with the special fellowship granted from the Turkish Naval Forces. He obtained Ph.D. degree from the Department of Electronic and Communication Engineering, Istanbul Technical University, Istanbul, Turkey, in 2018 with best thesis award of the year. He is currently serving as a senior researcher in private sector. He is Senior IEEE member and his research interests include image processing, signal processing, machine learning, deep learning, and radar signal processing applications.

BIOGRAPHIES



Necmettin Bayar received B.S. degree from the Department of Electronic Engineering, Gebze Technical University, Kocaeli, Turkey, in 2018 and completed his master degree from Istanbul Technical University in 2023. He is working as AI Developer in private company. His research interests cover image processing, signal processing, machine learning, deep learning, ai based radar imaging, SAR/ISAR imaging.



Kübra Güzel received B.S. degree from the Department of Electronic and Communication Engineering, Kocaeli University, Kocaeli, Turkey, in 2022. She works as AI Developer in private company. Her interests cover image processing, signal processing, machine learning, deep learning.