engineering proceedings

MDPI

Proceeding Paper

A Hybrid Synthetic Dataset Generation for Robust Document Recognition Using Image Rendering and Domain Randomization †

Plamen Nakov [1,*], Petar Petrov [2], Georgi Kotov [1], Milena Lazarova [1] and Ognyan Nakov [1]

[1] Department of Computer Systems, Technical University of Sofia, 1000 Sofia, Bulgaria; gkotov@tu-sofia.bg (G.K.); milaz@tu-sofia.bg (M.L.); nakov@tu-sofia.bg (O.N.)
[2] Digital and Software Solutions (DSS), 1000 Sofia, Bulgaria; petar.petrov@dss.bg
* Correspondence: p.nakov@tu-sofia.bg
† Presented at the 15th International Scientific Conference TechSys 2026—Engineering, Technologies and Systems, Plovdiv, Bulgaria, 14–16 May 2026.

Abstract

Automated recognition of identity documents is a critical component in digital identity verification systems. The development of robust recognition models is often constrained by the limited availability of large, diverse, high-quality, and publicly accessible ID card datasets. Collecting and annotating real-world ID card images is time-consuming, and often restricted due to privacy, legal, and security concerns. The paper proposes a novel approach for generating a large-scale synthetic dataset for ID card recognition by merging real ID card images with a texture dataset through a structured data fusion pipeline that introduces realistic visual variations as illumination effects, geometric distortions, and noise patterns while preserving the semantic integrity of the original ID card content.

Keywords: computer vision; document analysis; ID card recognition; dataset generation; synthetic data; data augmentation; texture fusion; deep learning

1. Introduction

Automated document recognition is an essential component of modern digital identity verification systems. Organizations across multiple sectors as governments, financial institutions, and private organizations increasingly rely on automated solutions to verify quickly and securely the authenticity and content of identity documents such as passports, national identity cards, and driver's licenses. These systems are utilized in applications as online banking account creation, electronic Know Your Customer (eKYC) procedures, border control automation, digital onboarding platforms, access control systems, and secure identity authentication services. Traditionally, the identity verification during customer onboarding processes required physical presence and manual verification by trained personnel. Nowadays digital platforms and automated solutions capable of extracting information from identity documents captured under uncontrolled environmental conditions have been widely adopted. The captured images often contain distortions caused by perspective shifts, varying illumination, shadows, reflections, occlusions, motion blur, and background clutter.

Recent advances in computer vision and deep learning have significantly improved the performance of document recognition systems. Despite the rapid advances in deep learning-based recognition models, the development of reliable document recognition systems remains challenging as the effectiveness of these models strongly depends on the availability of large, diverse, high-quality and well-annotated training datasets of real

Check for updates

Academic Editors: Nikola G. Shakev, Valyo N. Nikolov, Nikolay R. Kakanakov and Sevil A. Ahmed-Shieva

Published: 16 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

identity documents. Due to privacy regulations, legal restrictions, and security concerns, sharing and public release of real ID card datasets is severely limited. Consequently, researchers and developers often have access only to small proprietary datasets that may not sufficiently represent the variability encountered in real-world scenarios, having limited diversity and imbalance.

Identity documents differ across countries, issuing authorities and document versions in terms of layout, design elements, fonts, security features, holograms and background patterns. Furthermore, in the identity verification systems document images are captured using mobile devices by non-expert users. Thus, the acquired images are affected by significant variability in lighting conditions, camera angles, image quality, and background environments. In addition, shadows, motion blur, and partial occlusions further complicate the recognition process. Models trained using limited or homogeneous datasets that do not adequately capture the image variations tend to overfit to the training data and exhibit poor generalization for real inputs.

To address these challenges, synthetic data generation allows training datasets augmentation that can improve the robustness and generalization capabilities of machine learning models while maintaining control over visual properties of the generated images such as illumination, geometric distortions, noise patterns, and background environments. Unlike real-world datasets, synthetic data can be created without exposing sensitive personal information.

Recent advances in computer graphics and rendering tools have enabled the generation of highly realistic synthetic image datasets by simulating complex visual conditions through programmable pipelines and domain randomization techniques.

The primary objective of this research is to demonstrate that texture-enhanced synthetic data can effectively improve the generalization capabilities of document recognition models. The paper proposes a hybrid synthetic dataset generation pipeline for document recognition that combines real ID card images with texture datasets through a structured rendering process. The pipeline enables the generation of large-scale image datasets that simulate diverse environmental conditions and the visual appearance commonly encountered in real-world document capture scenarios. The suggested approach increases the image diversity by utilizing controlled variations in background textures, illumination conditions, geometric transformations, and noise patterns while preserving the semantic integrity of the original document content. The presented image generation pipeline provides a scalable framework for generating labeled synthetic datasets suitable for training document recognition models as the generation process is fully automated and large volumes of training images can be produced efficiently maintaining consistent annotation of document regions and textual elements.

The main contributions of this work can be summarized as follows:

- A hybrid synthetic dataset generation framework is proposed that merges real ID card images with diverse texture datasets using an image rendering pipeline;
- A domain randomization strategy is described that introduces realistic variations in lighting, geometry, background environments, and image noise;
- A scalable approach for producing large volumes of labeled synthetic datasets without relying on sensitive personal information is presented;
- An experimental evaluation is made to demonstrate improved robustness and generalization of deep learning-based document recognition models trained using the generated dataset.

The remainder of the paper is organized as follows. Section 2 reviews related work in document recognition and synthetic dataset generation. Section 3 describes the proposed synthetic data generation pipeline and rendering framework. Section 4 presents the ex-

perimental setup, and discusses the obtained results and their implications for document recognition systems. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Related Works

2.1. Synthetic Data in Computer Vision

Synthetic data generation has become an important approach for addressing the lack of large annotated datasets in computer vision in tasks such as object detection, facial recognition, and document analysis. Training modern deep learning models often requires millions of labeled samples, which are expensive or impractical to collect in many domains. Synthetic datasets provide a scalable alternative by enabling the automatic generation of labeled data while allowing precise control over scene parameters such as lighting, geometry, background, and noise. Techniques such as Generative Adversarial Networks (GANs), domain randomization, and style transfer have shown promising results in reducing dataset dependency [1–4].

One widely adopted technique for improving the robustness of models trained on synthetic data is domain randomization used as a strategy for bridging the reality gap by exposing models to enough synthetic variability that real images appear as just another variation at test time. Tobin et al. [5] demonstrate that models trained on highly randomized synthetic environments can generalize effectively to real-world images, even when the synthetic data does not perfectly match real-world distributions. The work shows that introducing large variations in textures, lighting, and camera parameters during training allows neural networks to learn features that remain robust when deployed in real environments.

Rendering engines and simulation frameworks have further expanded the possibilities of synthetic dataset generation. Blender-based programmatic rendering and ray tracing tools allow researchers to create realistic scenes and simulate complex imaging conditions. In the document analysis domain, a framework Sim2Real Docs [6] uses ray-traced rendering to generate synthetic documents placed in realistic environments. The suggested approach leverages physically based rendering and domain randomization to simulate variations in illumination, camera perspective, and background clutter and demonstrates that synthetic data can significantly improve model robustness when real-world datasets are limited.

A method to generate synthetic document images with automatic layout annotations is proposed in [7], allowing document layout recognition models to be trained without manually labeled data. Realistic documents and their labels are generated automatically through stochastic templates and structured dependencies between layout components and models trained purely on synthetic documents can approach the performance of models trained on real data for layout tasks.

An approach for richer visual document understanding by generating bilingual synthetic documents containing text, tables, charts, and images is presented in [8], positioning synthetic generation as a scalable response to the annotation bottleneck.

Although these methods demonstrate the value of synthetic data, most of them target general document understanding, such as reading order, layout parsing, or multimodal document parsing, rather than the narrower and more deployment-critical problem of ID card recognition from realistic captured images. Identity cards present additional constraints: their semantic structure must remain fixed and legible, document-specific fields must be preserved, and the generated images must reflect real capture artifacts such as background clutter, pose changes, illumination variation, and surface appearance changes without destroying the validity of the underlying card content. This makes ID-card-focused synthesis more restrictive than generic document generation, because realism

must be increased while preserving the template semantics and field consistency needed for supervised recognition.

2.2. Document Recognition Systems

Research on automated identity-document analysis has advanced steadily in recent years, driven by the expansion of digital onboarding, remote verification, and document-centric fraud prevention. Modern systems typically combine document localization, optical character recognition (OCR), and structured information extraction to process identity documents and other structured forms [9,10]. Previous studies have focused on deep learning-based OCR models for ID card recognition using convolutional neural networks (CNNs), YOLO-based object detection, and transformer-based architectures [4,11]. However, most approaches rely heavily on real-world datasets and a central obstacle is the scarcity of public datasets containing realistic identity documents and field-level annotations.

The recent development of automated identity-document analysis systems is supported by several publicly available datasets. One of the most widely used benchmarks is MIDV-500 which contains 500 video sequences of identity documents captured using mobile devices [12]. The dataset includes various document types and capture conditions and has been used extensively for evaluation of document localization, OCR, and field extraction, but it also highlights how difficult it is to build representative datasets for this domain because identity documents contain sensitive personal information. To further expand the scope of identity-document benchmarks an extended dataset MIDV-2020 is introduced in [13] containing additional document templates and image acquisition scenarios. MIDV-2020 includes mock identity documents, scans, photos, and videos designed to simulate real-world identity cards and to support more comprehensive identity-document analysis tasks under variable capture conditions while avoiding privacy concerns associated with real personal data.

Despite these contributions, document recognition research still faces significant limitations related to dataset availability. The benchmark datasets confirm a broader limitation of the field: publicly accessible identity-document datasets remain relatively small, constrained, and privacy-sensitive compared with datasets in more general computer vision domains. Identity documents contain sensitive personal information and, consequently, strict privacy regulations often prevent the publication of large-scale datasets containing real documents. This limitation reduces the diversity of available training data and can negatively affect the generalization capabilities of document recognition models.

More recent work has further emphasized that privacy restrictions and the limited diversity of public corpora continue to hinder progress in robust identity verification tasks and fraud analysis. The IDNet dataset is proposed explicitly in response to the limitations of prior identity-document benchmarks arguing that existing datasets provide too few samples, insufficient fraud diversity, and limited alteration of critical personal fields [14]. The SIDTD dataset [15] focuses on forged identity and travel document detection and is motivated by the same core issue: real identity documents cannot generally be published at scale because of personal data protection requirements, while forged examples are rare and highly variable. However, even these datasets remain limited in size and diversity compared to typical computer vision benchmarks. The field increasingly relies on synthetic or mock-document generation not merely as augmentation, but as a practical necessity for privacy-preserving research and model development.

2.3. Texture-Based Data Augmentation

Data augmentation techniques are commonly used to improve the generalization performance of deep learning models by artificially increasing the diversity of training data.

Traditional augmentation methods include geometric transformations, color adjustments, and noise injections. While these techniques can improve robustness to small variations, they may not sufficiently simulate the wide range of environmental conditions encountered in real-world document capture.

Texture-based augmentation has emerged as a more advanced strategy for generating realistic variations in image appearance. Instead of simply modifying existing images, texture-based approaches combine object templates with diverse texture datasets to simulate complex visual environments. This technique can introduce variations in background surfaces, lighting reflections, and material properties while preserving the semantic content of the original object. Texture-based approaches are particularly relevant for identity document recognition, where images are often captured in uncontrolled environments. Documents may be placed on different surfaces, photographed under varying illumination conditions, or partially occluded by shadows and reflections. Integrating diverse textures such as paper grain, plastic reflections, and background noise during synthetic data generation can therefore help simulate real-world capture conditions more effectively than traditional augmentation methods.

In the context of document understanding, synthetic document generators such as SynthDoG [16] use texture sampling and layout generation to create large-scale synthetic document generators used for model pre-training, specifically to reduce dependence on large-scale real document images. The approach combines sampled backgrounds, document textures, text corpora, and rendering effects to create diverse synthetic documents at scale, showing that synthetic data can be effective not only for augmentation but also for representation learning.

The previous works on identity document research suggests three consistent conclusions: (i) ID recognition is bottlenecked by the limited availability of public, privacy-safe datasets; (ii) synthetic generation has already proven useful in broader document AI and in privacy-preserving identity-document tasks; (iii) rendering-based domain randomization offers a way to incorporate realistic variability into synthetic samples. However, many existing synthetic document generation methods focus primarily on document layout and textual content rather than modeling realistic capture environments. This creates an opportunity for approaches focused specifically on scalable generation of labeled synthetic ID card images that preserve document semantics while enriching appearance through realistic texture and scene variation suitable for training robust recognition models.

3. Proposed Methodology

The goal of the proposed methodology is to create a scalable process for producing large volumes of labeled training data while preserving the semantic structure of identity documents.

3.1. Pipeline Overview

The proposed hybrid synthetic dataset pipeline generates synthetic images by combining real ID card templates with randomly sampled textures and applies a series of rendering transformations to simulate realistic capture conditions applying. The pipeline preserves the structural integrity of the ID card while introducing variations that mimic real-world capture conditions. It consists of four main stages as shown in Figure 1:

1. Scene Construction

A 3D ID card model with physically based materials is placed within procedurally generated environments. Background geometry, lighting sources, and camera configuration are defined within the scene.

2. Domain Randomization

To ensure variability, the fusion process introduces variability in several aspects: card pose and orientation, camera position and focal parameters, illumination intensity and color, background composition, surface material properties.

3. Physically Based Rendering

Images are rendered using a rendering engine to simulate realistic light transport, shadows, reflections, and perspective effects.

4. Image post-processing

Rendered images undergo minor image-space augmentations including low-level noise injection and compression artifacts to approximate real capture imperfections.

The pipeline is designed to be fully automated, enabling large-scale synthetic dataset generation while maintaining consistent labeling.

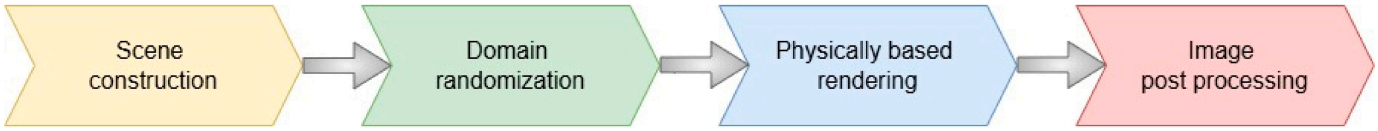


Figure 1. Proposed pipeline for synthetic ID card dataset generation using texture–ID card data fusion.

3.2. Datasets Used

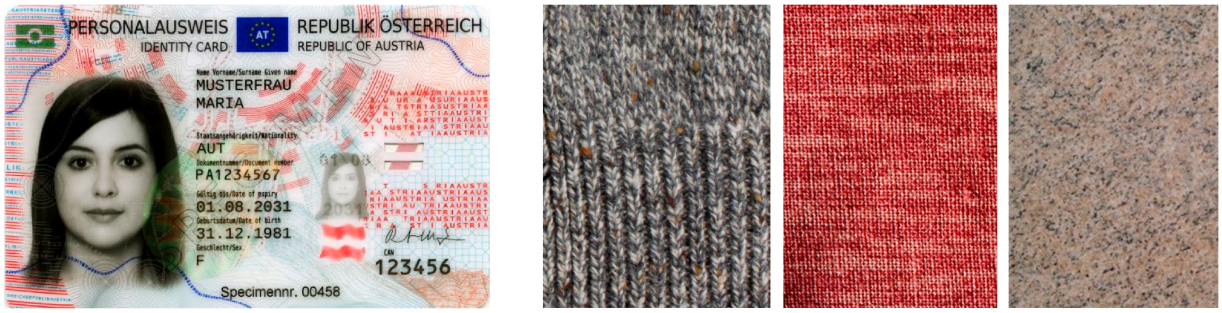
The proposed method requires two primary input data sources: ID card templates and texture images. The study employs two data sources: (i) a real ID card image dataset derived from scanned and approved document materials, and (ii) an open-source texture dataset used for environmental simulation and augmentation [17,18].

The real ID card dataset represents the base documents used in the generation process and consists of front-view scans of identification cards collected under controlled conditions that comprise a variety of documents [19]. The images capture realistic document layouts, lighting variations, and scanning artifacts. The images contain the layout structure and visual elements of identity cards, including text fields, photographs, and graphical elements. Each image is accompanied by metadata describing the spatial coordinates of relevant document regions such as text fields, portrait photograph areas, machine-readable zones, security features. These structural annotations enable automatic generation of labels during the synthesis process.

The texture dataset provides a diverse set of background surfaces, material and environmental patterns used to simulate real-world environments and surfaces on which identity documents may be placed during image capture, including paper textures, plastic surfaces, fabric and wood patterns, scratches, stains, and noise artifacts, lighting gradients and background patterns. The textures are used for compositing and data augmentation that simulate real-world environmental variations such as surface wear, background interference, and illumination inconsistencies. Sampling textures from a diverse dataset helps introduce realistic background variability and prevents models from overfitting to simple backgrounds. As the texture image is independent of the ID card content, the pipeline can generate a large number of visually distinct samples from a limited set of templates.

Examples from the real ID card dataset and the texture dataset used in the study are given in Figure 2.

https://doi.org/10.3390/engproc2026150010



(**a**) (**b**)

Figure 2. Examples from image datasets used: (**a**) real ID card dataset and (**b**) texture dataset.

3.3. Dataset Fusion Process

The core component of the proposed methodology is the texture fusion mechanism, which combines ID card templates with background textures to produce realistic document scenes. The suggested methodology for a synthetic dataset generation consists of the following steps (Figure 3):

Step 1. A real ID card template is selected from the input dataset and a texture image is randomly sampled from a texture dataset;

Step 2. A 3D scene is defined using the ID card and the texture image. Background geometry, lighting sources, and camera configuration are defined within the scene model to model variability of the card position and orientation, the camera position and focal parameters, the illumination intensity and color, and the background composition.

Step 3. The 3D scene is rendered using Blender's rendering engine [20] to produce a rendered image of the raw ID card placed over the raw texture surface.

Step 4. The rendered image is post-processed using Python (3.14.4)-implemented low-level image processing to add adjustments and noise effects that are introduced to emulate real capture conditions

Step 5. The final rendered image is exported together with automatically generated annotations.

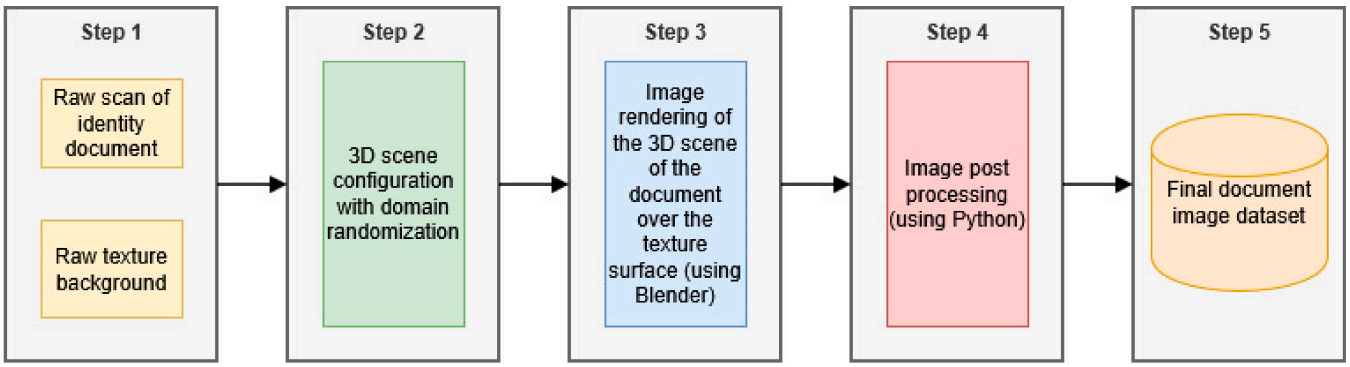


Figure 3. Processing steps of the data fusion process for synthetic ID card dataset generation.

Examples of the texture fusion process (before vs after) are given in Figure 4. Figure 5 shows samples of the generated synthetic ID card dataset.

Using the proposed pipeline, a synthetic ID card image dataset is generated with the following variations: different lighting conditions, different background textures, different card orientations, and simulated wear and tear. The proposed pipeline is designed to support large-scale dataset generation. By combining multiple ID card templates with thousands of background textures and randomized transformations, the number of possible generated images grows combinatorically. This scalability allows the pipeline to produce

datasets containing tens or hundreds of thousands of images from a relatively small number of base templates.



Figure 4. Illustration of the texture fusion process: (**a**) original ID card; (**b**) selected texture; (**c**) fused synthetic ID card.



Figure 5. Samples of the generated synthetic ID card dataset.

4. Results and Discussion

Starting from a limited set of sensitive real ID samples, the proposed Blender-Python framework is used to generate a large-scale synthetic dataset through procedural layout rendering, domain randomization, and automated mask extraction. Python is used to implement a procedural data generation pipeline within Blender via the bpy API. Custom scripts are developed to randomly sample scene parameters (lighting, camera pose, object orientation), assign textures, and render both RGB images and corresponding segmentation masks. Rendered outputs are stored in WebP format. Source and background images are sampled from PNG and JPEG files. Compression quality is set to Blender's default WebP encoder settings. All rendered images are generated at a fixed resolution of 640 × 640 pixels in RGB color space.

A dataset of 10,000 synthetic images is produced using the suggested hybrid synthetic dataset generation pipeline derived from 645 base templates corresponding to an expansion factor of approximately 15× relative to the original dataset.

The synthetic dataset expanded coverage utilizes illumination intensity and direction, camera tilt and perspective distortion, motion and Gaussian blur levels, background variability and sensor noise. Table 1 summarizes the parameter space used to introduce controlled variability into the synthetic dataset. Each parameter is sampled from a predefined numerical range. The synthetic dataset is generated by randomizing scene parameters within predefined ranges: object rotation (±15° around the z-axis), light orientation (±45° around x and z axes), light intensity (2.5–7.5 units), camera vertical position (7–13 units) and

lateral offset (−3 to +3 units). Background and foreground textures are sampled uniformly from predefined image sets. This increase in variability extends beyond what is observable in the limited real dataset, suggesting improved coverage of the document appearance while maintaining structural consistency.

Table 1. Parameter configuration for synthetic dataset variability.

Parameter	Range/Strategy	Purpose
Document rotation	±15 degrees	Simulates natural misalignment and orientation variance during capture
Lighting rotation	±45 degrees	Introduces directional lighting changes and shadow variation
Lighting intensity	[2.5, 7.5] units	Models illumination strength
Camera height	[7, 13] units	Controls scale and perspective distortion relative to the document
Camera offset	[−3, 3] units	Simulates lateral camera displacement and viewpoint shifts
Background & foreground selection	Uniform random sampling	Ensures diverse environmental contexts

A key advantage of the proposed pipeline is deterministic ground-truth generation. Rendering and annotation are fully automated without manual labeling effort. Segmentation masks are extracted directly from object index passes within Blender, thus ensuring pixel-level correspondence between object geometry and labels. This eliminates human annotation noise and guarantees boundary precision which is particularly valuable for training segmentation models on document layouts.

The average rendering time per image is 100 s using Blender Cycles (CPU-only) on a virtual machine with 4 vCPUs (Intel Xeon Gold 5315Y @ 3.19 GHz) and 12 GB RAM, enabling scalable dataset production with minimal human intervention. This demonstrates the feasibility of synthetic amplification when real data collection is constrained by privacy or regulatory limitations.

To assess whether the synthetic data approximates real-world document characteristics, statistical comparisons are performed between the real and synthetic subsets. Pixel intensity distributions, bounding box area ratios, and aspect ratio statistics show similar ranges with synthetic samples exhibiting broader variance due to controlled domain randomization.

To validate practical utilization of the suggested pipeline for synthetic document dataset generation, a YOLOv11 segmentation model is trained following the latest advancements of the model architecture and training for the identity document domain [21–24]. The model is trained on 10 000 synthetic images (9000/1000 train/validation split) for 73 epochs on a high-performance cluster and is evaluated on a held-out real test set of 100 images. The performance of the model trained with synthetic augmentation is assessed on the test dataset and is compared against a baseline model trained on manually labeled data evaluated under identical conditions. The model achieves mAP@0.5 = 99.50% on the test dataset and demonstrates improved robustness to non-frontal perspectives, blur, and illumination variation compared to a model trained solely on the limited real data.

The suggested hybrid synthetic dataset generation is additionally validated on passport documents producing comparable segmentation performance. The experimental results indicates that the increased variability introduced during synthetic generation contributes to improved generalization.

The training and the validation curves for the YOLO segmentation model when trained exclusively on the generated synthetic dataset are given in Figure 6 and the evaluation

metrics for precision, recall, mAP50 and mAP50-95 for the training and the validation splits are given in Table 2.

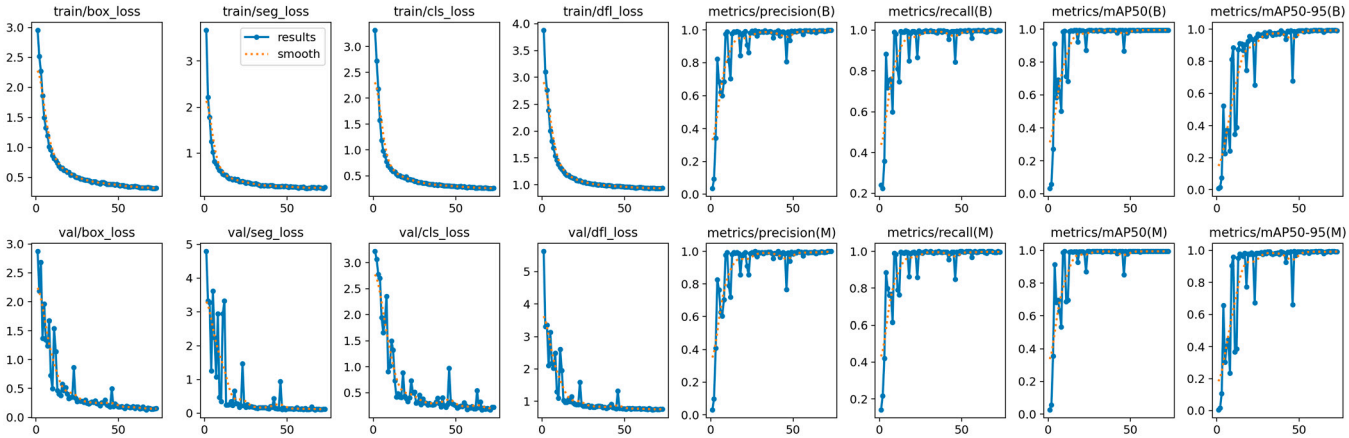


Figure 6. Training and validation curves for YOLOv11 segmentation model trained exclusively on the generated synthetic dataset.

Table 2. Training and validation metrics for YOLO segmentation trained exclusively on the generated synthetic dataset.

Split	Precision	Recall	mAP50	mAP50-95
Training	99.972%	99.671%	99.67%	99.50%
Validation	99.972%	99.671%	99.50%	99.22%

The plots show steady convergence of both bounding box and segmentation losses, alongside consistent increases in precision, recall, and mAP metrics. The training and validation curves of the YOLO segmentation model show stable and efficient learning behavior on the generated synthetic dataset.

The training losses (box loss, segmentation loss, classification loss, and distribution focal loss) decrease rapidly during the first few epochs and then gradually converge, indicating that the model quickly learns the patterns present in the dataset and the optimization process is stable. Similarly, the validation losses follow a comparable downward trend and remain close to the training losses after the initial epochs, which suggests that the model generalizes well to the validation data and there are no clear signs of overfitting within the synthetic dataset. The evaluation metrics further support this observation: precision and recall increase rapidly and reach values close to 1.0, showing that the model produces very few false positives and successfully detects most objects. The mAP metrics (mAP50 and mAP50-95) for both bounding boxes and segmentation masks also rise quickly and stabilize at very high values, indicating excellent detection accuracy and high-quality segmentation performance. The limited divergence between training and validation performance indicates that the synthetic samples provide sufficient diversity and coherent annotations to enable stable learning. These dynamics support the observation that the dataset's procedural variability contributes to improved model robustness on real-world test images.

Qualitative evaluation on real-world document images confirms accurate segmentation masks. Example outputs are omitted due to privacy restrictions on personal identity documents. Furthermore, the trained model has been successfully deployed in a production identity verification system aIDentix [25], further validating the practical applicability of the proposed synthetic training approach.

5. Conclusions and Future Work

The presented study introduces an approach for generating synthetic ID card datasets by merging real ID card images with texture datasets using domain randomization and physically based rendering. The results prove the effectiveness of suggested approach and confirm that texture-enhanced synthetic data can effectively mitigate data scarcity issues while reducing dependency on sensitive personal information.

While the proposed approach shows promising results, certain limitations remain as synthetic images may still fail to capture the full complexity of real-world conditions, and certain texture patterns can appear artificial. An ablation study isolating the impact of individual augmentation factors such as lighting, pose, texture, blur, and noise is planned for future work.

Although the experiments focused on ID cards and passports, the proposed pipeline is document-agnostic and can be extended to other document types, including driver's licenses, credit cards, and various travel documents.

The integration of more advanced generative techniques, such as Generative Adversarial Networks (GANs) or diffusion models, could further enhance the realism of the generated data. Future research will therefore focus on incorporating GANs or diffusion-based generation methods to improve visual realism and system robustness under challenging conditions such as lighting variations, motion blur, and partial occlusion. In addition, future work will explore deeper integration of face recognition and document recognition modules to strengthen end-to-end identity verification performance, improve security, and enhance fraud detection capabilities.

Author Contributions: Conceptualization, P.N., P.P., G.K., M.L. and O.N.; methodology, P.N., P.P. and G.K.; software, P.P. and G.K.; validation, P.P., P.N. and G.K.; resources, P.P. and P.N.; data curation, P.P. and G.K.; writing—original draft preparation, P.P., P.N. and G.K.; writing—review and editing, M.L. and O.N.; visualization, P.P. and M.L.; supervision, P.N., M.L. and O.N.; project administration, M.L. and O.N.; funding acquisition, M.L. and O.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research work presented in the paper is funded by the European Union-NextGenerationEU via the National Recovery and Resilience Plan of the Republic of Bulgaria under project BG-RRP-2.004-0005 "Improving the research capacity and quality to achieve international recognition and reSilience of TU-Sofia (IDEAS)". The research that led to these results was carried out using the infrastructure purchased under the National Roadmap for RI, financially coordinated by the MES of the Republic of Bulgaria (Grant No. DO1-98/26.06.2025).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The raw data is not publicly available due to privacy restrictions.

Conflicts of Interest: Author Petar Petrov was employed by the company Digital and Software Solutions. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Goyal, M.; Mahmoud, Q.H. A systematic review of synthetic data generation techniques using generative AI. *Electronics* **2024**, *13*, 3509. [CrossRef]
2. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.-H. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.* **2023**, *56*, 1–39. [CrossRef]

3. Tanveer, N.; Ul-Hasan, A.; Shafait, F. Diffusion models for document image generation. In *Document Analysis and Recognition—ICDAR 2023*; Fink, G.A., Jain, R., Kise, K., Zanibbi, R., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2023; Volume 14189. [CrossRef]
4. Veerasekar, P.A.; Sindha, M.M.R.; Pandiyan, U.M.; Vijayarajan, V. A transfer learning approach for college ID card detection. *AIP Conf. Proc.* **2025**, *3258*, 020030. [CrossRef]
5. Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; Abbeel, P. Domain randomization for transferring deep neural networks from simulation to the real world. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 23–30. [CrossRef]
6. Raman, N.; Shah, S.; Veloso, M. Synthetic document generator for annotation-free layout recognition. *Pattern Recognit.* **2022**, *128*, 108660. [CrossRef]
7. Maddikunta, N.; Zhao, H.; Keswani, S.; Samuel, A.; Guo, F.-M.; Srishankar, N.; Pardeshi, V.; Huang, A. Sim2Real Docs: Domain randomization for documents in natural scenes using ray-traced rendering. *arXiv* **2021**, arXiv:2112.09220. [CrossRef]
8. Ding, C.; Liu, X.; Tang, W.; Li, J.; Wang, X.; Zhao, R.; Nguyen, C.-T.; Tan, F. Synthdoc: Bilingual documents synthesis for visual document understanding. In *Proceedings of the 2nd Workshop on Large Generative Models Meet Multimodal Applications, (LGM3A '24)*; Association for Computing Machinery: New York, NY, USA; pp. 16–25. [CrossRef]
9. Suddul, G.; Seguin, J. A custom-built deep learning approach for text extraction from identity card images. *Int. J. Inform. Commun. Technol.* **2026**, *13*, 34–41.
10. Gulnara, B.; Yerassyl, A. Using image processing and optical character recognition to recognise ID cards in the online process of onboarding. In Proceedings of the 2022 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 28–30 April 2022; pp. 1–6. [CrossRef]
11. Ramachandrula, P.; Kunisetty, J.; Subramanian, S.; Palaniswamy, S.; Pati, P.B. A novel system for authentication of identity cards using YOLOv5 and YOLOv8. In *Advances in Data Science and Artificial Intelligence, ERCICA 2025*; Lecture Notes in Electrical Engineering; Springer: Singapore, 2025; Volume 1399. [CrossRef]
12. Arlazarov, V.V.; Bulatov, K.B.; Chernov, T.S.; Arlazarov, V.L. MIDV-500: A dataset for identity document analysis and recognition on mobile devices in video stream. *Comput. Opt.* **2019**, *43*, 818–824. [CrossRef]
13. Bulatov, K.; Emelianova, E.V.; Tropin, D.V.; Skoryukina, N.S.; Chernyshova, Y.S.; Sheshkus, A.V.; Usilin, S.A.; Ming, Z.; Burie, J.-C.; Luqman, M.M.; et al. MIDV-2020: A comprehensive benchmark dataset for identity document analysis. *Comput. Opt.* **2022**, *46*, 252–270. [CrossRef]
14. Guan, H.; Wang, Y.; Xie, L.; Nag, S.; Goel, R.; Swamy, N.E.N.; Yang, Y.; Xiao, C.; Prisby, J.; Maciejewski, R.; et al. IDNet: A novel dataset for identity document analysis and fraud detection. *arXiv* **2024**, arXiv:2112.09220. [CrossRef]
15. Boned, C.; Talarmain, M.; Ghanmi, N.; Chiron, G.; Biswas, S.; Awal, A.M. Synthetic dataset of ID and travel documents. *Sci. Data* **2024**, *11*, 1356. [CrossRef] [PubMed]
16. Kim, G.; Hong, T.; Yim, M.; Nam, J.; Park, J.; Yim, J.; Hwang, W.; Yun, S.; Han, D. OCR-Free document understanding transformer. In *Computer Vision—ECCV*; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13688, pp. 498–517. [CrossRef]
17. Sharan, L.; Rosenholtz, R.; Adelson, E.H. Material perception: What can you see in a brief glance? *J. Vis.* **2009**, *14*, 784. [CrossRef]
18. Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; Vedaldi, A. Describing textures in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3606–3613.
19. PRADO—Public Register of Authentic Identity and Travel Documents Online. Available online: https://www.consilium.europa.eu/prado/en/prado-start-page.html (accessed on 10 March 2026).
20. Blender. Available online: https://www.blender.org (accessed on 10 March 2026).
21. Liu, C.; Tao, Y.; Liang, J.; Li, K.; Chen, Y. Object detection based on YOLO network. In Proceedings of the 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference, Chongqing, China, 14–16 December 2018; pp. 799–803. [CrossRef]
22. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
23. Subramanian, S.; Kekatpure, V.; Raymond, G.; Parab, K.; Dugad, S.; Shirke, A. TEYSuR—Text extraction with YOLO and super resolution. In Proceedings of the 2022 International Conference for Advancement in Technology, Goa, India, 21–22 January 2022. [CrossRef]
24. GokulaKrishnan, E.; Malathi, G. Exploring hand wrist feature as biometric identifier: A deep learning perspective using YOLOv11. *Results Eng.* **2025**, *28*, 107107. [CrossRef]
25. aIDentix. Available online: https://aidentix.com (accessed on 10 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.