

GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

The journal is indexed in

EBSCO

ISSN 2545-4803 on line

DOI: 10.46763/BJAMI

**BALKAN JOURNAL
OF APPLIED MATHEMATICS
AND INFORMATICS
(BJAMI)**



YEAR 2025

VOLUME VIII, Number 1

AIMS AND SCOPE:

BJAMI publishes original research articles in the areas of applied mathematics and informatics.

Topics:

1. Computer science;
2. Computer and software engineering;
3. Information technology;
4. Computer security;
5. Electrical engineering;
6. Telecommunication;
7. Mathematics and its applications;
8. Articles of interdisciplinary of computer and information sciences with education, economics, environmental, health, and engineering.

Managing editor

Mirjana Kocaleva Vitanova Ph.D.

Zoran Zlatev Ph.D.

Editor in chief

Biljana Zlatanovska Ph.D.

Lectoure

Snezana Kirova

Technical editor

Biljana Zlatanovska Ph.D.

Mirjana Kocaleva Vitanova Ph.D.

**BALKAN JOURNAL
OF APPLIED MATHEMATICS AND INFORMATICS
(BJAMI), Vol 8**

**ISSN 2545-4803 online
Vol. 8, No. 1, Year 2025**

EDITORIAL BOARD

- Adelina Plamenova Aleksieva-Petrova**, Technical University – Sofia,
Faculty of Computer Systems and Control, Sofia, Bulgaria
- Lyudmila Stoyanova**, Technical University - Sofia , Faculty of computer systems and control,
Department – Programming and computer technologies, Bulgaria
- Zlatko Georgiev Varbanov**, Department of Mathematics and Informatics,
Veliko Tarnovo University, Bulgaria
- Snezana Scepanovic**, Faculty for Information Technology,
University “Mediterranean”, Podgorica, Montenegro
- Daniela Veleva Minkovska**, Faculty of Computer Systems and Technologies,
Technical University, Sofia, Bulgaria
- Stefka Hristova Bouyuklieva**, Department of Algebra and Geometry,
Faculty of Mathematics and Informatics, Veliko Tarnovo University, Bulgaria
- Vesselin Velichkov**, University of Luxembourg, Faculty of Sciences,
Technology and Communication (FSTC), Luxembourg
- Isabel Maria Baltazar Simões de Carvalho**, Instituto Superior Técnico,
Technical University of Lisbon, Portugal
- Predrag S. Stanimirović**, University of Niš, Faculty of Sciences and Mathematics,
Department of Mathematics and Informatics, Niš, Serbia
- Shcherbacov Victor**, Institute of Mathematics and Computer Science,
Academy of Sciences of Moldova, Moldova
- Pedro Ricardo Morais Inácio**, Department of Computer Science,
Universidade da Beira Interior, Portugal
- Georgi Tuparov**, Technical University of Sofia Bulgaria
- Martin Lukarevski**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Ivanka Georgieva**, South-West University, Blagoevgrad, Bulgaria
- Georgi Stojanov**, Computer Science, Mathematics, and Environmental Science Department
The American University of Paris, France
- Iliya Guerguiev Bouyukliev**, Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences, Bulgaria
- Riste Škrekovski**, FAMNIT, University of Primorska, Koper, Slovenia
- Stela Zhelezova**, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria
- Katerina Taskova**, Computational Biology and Data Mining Group,
Faculty of Biology, Johannes Gutenberg-Universität Mainz (JGU), Mainz, Germany.
- Dragana Glušac**, Tehnical Faculty “Mihajlo Pupin”, Zrenjanin, Serbia
- Cveta Martinovska-Bande**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Blagoj Delipetrov**, European Commission Joint Research Centre, Italy
- Zoran Zdravev**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Aleksandra Mileva**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Igor Stojanovik**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Saso Koceski**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Natasa Koceska**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Aleksandar Krstev**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Biljana Zlatanovska**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Natasa Stojkovik**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Done Stojanov**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Limonka Koceva Lazarova**, Faculty of Computer Science, UGD, Republic of North Macedonia
- Tatjana Atanasova Pacemska**, Faculty of Computer Science, UGD, Republic of North Macedonia

CONTENT

Sara Kostevska, Biljana Chitkuseva Dimitrovska, Todor Chekerovski, Maria Chekerovska and Sara Srebrenkoska

SMART CITY: A REVIEW OF CURRENT DEVELOPMENTS AND IMPLEMENTATION OF SMART GRID TECHNOLOGY.....7

Aleksandra Risteska-Kamcheski

GENERALIZATION OF APPLICATION OF FUNDAMENTAL LEMMA OF VARIATIONAL CALCULUS.....19

Goce Stefanov, Vasilija Sarac

MONITORING OF AC MOTOR SPEED CONTROLLER PARAMETERS IN AN IoT NETWORK.....27

Elena Jovanovska, Marjan Kotevski, Blagoj Kotevski, Saso Koceski

AUTOMATED DOOR STATE DETECTION USING DEEP LEARNING: A COMPUTER VISION APPROACH WITH ROBOFLOW PLATFORM.....41

José Alejandro Ramón Rocha, Elena Jovanovska, Marjan Kotevski, Blagoj Kotevski and Saso Koceski

DEEP LEARNING-BASED DETECTION AND CLASSIFICATION OF DOCUMENT ELEMENTS USING ROBOFLOW.....51

Anastasija Antova, Elena Karamazova Gelova, Dushko Josheski, Mirjana Kocaleva Vitanova

ANALYSIS OF THE MOVEMENT OF FLUCTUATIONS AND TRENDS OF THE GROSS DOMESTIC PRODUCT IN THE REPUBLIC OF NORTH MACEDONIA AND FORECASTS.....61

Aleksandar Kotevski

INTEGRATING AI AND CLOUD COMPUTING FOR EFFICIENT AUDIO ANALYSIS73

Rexhep Mustafovski

STATE-OF-THE-ART COMPARISON OF MOBILESECURECOMM WITH MODERN SECURE COMMUNICATION PLATFORMS FOR TACTICAL OPERATIONS.....87

DEEP LEARNING-BASED DETECTION AND CLASSIFICATION OF DOCUMENT ELEMENTS USING ROBOFLOW

JOSÉ ALEJANDRO RAMÓN ROCHA, ELENA JOVANOVSKA, MARJAN KOTEVSKI, BLAGOJ KOTEVSKI AND SASO
KOCESKI

Abstract. Recent advances in deep learning have enabled significant improvements in document understanding and processing. This paper presents a workflow for detecting and classifying key document elements—paragraphs of text, tables, signatures, and stamps—using deep learning models implemented through the Roboflow platform. We address the unique challenges of document element detection, including complex spatial relationships, overlapping boundaries, and varying formats across different document types. Our methodology leverages the Roboflow's comprehensive toolkit for dataset preparation, model training, and deployment, with particular attention to data augmentation techniques specific to document processing. The proposed system achieves robust detection and classification of document elements. Experimental results demonstrate the system's effectiveness in real-world applications. Our findings indicate that the proposed approach offers significant accuracy and efficiency of automated document processing, particularly in scenarios requiring the precise identification and extraction of structured and unstructured document components.

1. Introduction

The automated analysis of document elements—including paragraphs of text, tables, signatures, and stamps—represents a fundamental challenge in intelligent document processing. While traditional optical character recognition (OCR) systems excel at converting text into machine-readable format, the detection and classification of distinct document elements require a more sophisticated approach. The recent advancements in artificial intelligence have made a huge impact and transformed almost all sectors in society such as: tourism [1], medicine [2-4], biology [5], education [6-7], robotics [8-12], and economy [13]. Moreover, the development of deep learning and computer vision technologies have revolutionized document analysis by enabling automated systems to detect, classify, and understand complex document elements with remarkable accuracy, transforming the traditionally manual and time-consuming processes across industries into efficient, scalable operations. This paper presents a deep learning-based solution, leveraging the Roboflow platform [14], for accurate identification and classification of key document components.

Document element detection presents unique challenges that distinguish it from conventional computer vision tasks. Unlike natural images where objects are typically well-separated and distinct, document elements often exhibit complex spatial relationships, overlapping boundaries, and varying formats. Tables may contain embedded text, signatures might overlap with stamps, and paragraphs can be formatted in multiple columns or interrupted by other elements. These characteristics necessitate specialized approaches to achieve reliable detection and classification.

The advent of deep learning architecture has revolutionized computer vision tasks, offering new possibilities for document analysis. Recent advances in object detection networks, particularly those optimized for document processing, have demonstrated remarkable capability in identifying and classifying document elements with high precision. Our approach utilizes these advances, implementing a custom-trained model through the Roboflow platform to achieve robust detection of four critical document elements: paragraphs of text, tables, signatures, and stamps.

The Roboflow platform provides crucial advantages in developing and deploying document analysis solutions. Its comprehensive toolkit facilitates efficient dataset preparation, model training and deployment, while offering capabilities for data augmentation and preprocessing that are particularly valuable for document processing. These features enable the creation of more robust models capable of handling variations in document quality, orientation, and formatting that are common in real-world applications.

Tables present a particular challenge in document processing due to their variable structure and the need to preserve their logical organization. Our approach not only detects the presence and boundaries of tables but also maintains the structural relationships necessary for downstream data extraction. Similarly, the detection of signatures and stamps—crucial for document verification and workflow routing—requires careful consideration of their unique visual characteristics and potential overlap with other elements.

The applications of this technology span across numerous domains where document processing plays a critical role. In financial services, it enables automated processing of forms and contracts by accurately identifying and extracting information from specific document sections. Healthcare organizations can benefit from improved processing of medical records, where the accurate detection of tables containing test results and official stamps for validation is crucial. Legal firms can streamline document review processes by automatically identifying key paragraphs and verification elements like signatures and stamps.

This paper presents a workflow for developing and implementing an efficient detection and classification system, including preparation of training data, model architecture selection, and optimization strategies. We evaluate our approach on a diverse dataset of real-world documents, demonstrating its effectiveness in accurately detecting and classifying document elements across various formats and quality levels. Through detailed analysis, we examine the system's performance and discuss its practical applications in automated document processing workflows.

2. Related work

The automated detection and classification of document elements has evolved significantly with the advancement of deep learning techniques. This section reviews relevant work in document layout analysis, element detection, and the application of deep learning in document processing.

Early approaches to document layout analysis relied primarily on rule-based systems and traditional computer vision techniques. Chen et al. [15] proposed X-Y cut algorithms for page segmentation, while O'Gorman [16] introduced the docstrum algorithm for document layout analysis. These methods, while foundational, struggled with complex layouts and varying document formats.

The emergence of deep learning has transformed the document layout analysis. LayoutLMv2 [17] introduced a multimodal pre-training approach that jointly models text and layout information, achieving state-of-the-art performance on several document understanding tasks. DocFormer [18] extended this work by incorporating visual features and spatial relationships between document elements, demonstrating improved performance in document structure understanding.

Recent years have seen significant advances in applying object detection architectures to document element detection. YOLO-based approaches have been particularly successful in this domain. Huang et al. [19] adapted YOLOv5 for document layout analysis, achieving high accuracy in detecting text blocks, tables, and figures. Similarly, FastDoc [20] proposed a lightweight detection framework specifically optimized for document elements.

Table detection has received special attention due to its complexity. CascadeTabNet proposed dual neural architecture for table detection and structure recognition and presented an improved approach upon this with a cascade mask R-CNN, demonstrating superior performance in detecting complex table structures [21].

The detection of signatures and stamps presents unique challenges due to their variable appearance and potential overlap with other elements. Signature and logo detection using deep CNN and YOLO v2 for document image retrieval was proposed in [22]. It was demonstrated that the proposed solution could be very useful for document retrieval using signature or logo information. For seal text detection, Sun et al. [23] developed an improved YOLO v8 detection method based on a receptive-field attention and efficient multi-scale attention (RFEMA) module.

The development of platforms like Roboflow has significantly streamlined the implementation of deep learning solutions for document processing. Similar platforms such as Fedvision [24] and MMDetection [25] have been widely used in academic research for object detection tasks. Fedvision is based on federated learning approach and is suitable for edge-devices in IoT networks but compared to Roboflow, it lacks the support for multiple architectures (YOLO, SSD, etc.) as well as auto-labeling tools. MMDetection is a powerful, highly configurable open-source framework suited for advanced users who need full control over training and customization of state-of-the-art object detection models but it still lacks the auto-labeling features compared to Roboflow. Moreover, Roboflow's specialized features for document processing, including automated data augmentation and preprocessing pipelines, have made it particularly suitable for document element detection tasks.

Recent work has increasingly focused on multi-modal approaches that combine visual and textual information. DocEnTr [26] introduced a transformer-based architecture that jointly processes visual and textual features for improved document understanding.

Similarly, DocFormer [27] demonstrated the benefits of incorporating spatial information alongside visual and textual features. However, both of them are lacking the support of various object detection models as well as auto-labeling features that are significantly improving the training of models.

Despite significant progress, several challenges remain in document element detection. Binmakhashen et al. [28] highlighted issues with handling degraded documents and complex layouts. The integration of deep learning models with downstream processing tasks, such as information extraction and document classification, remains an active area of research.

3. Document element detection workflow

The workflow for creating and training a document element detection model is presented in Figure 1. It begins with comprehensive data collection. This initial phase involves gathering a diverse set of document images that represent various layouts, quality levels, and orientations. For robust model training, we collected 390 images with a minimum of 80 images per class to ensure sufficient representation of different document elements, despite the fact that the text paragraphs are dominant in the text documents and their number is significantly higher.

Once the dataset is assembled, the next step involves uploading the collected images to the Roboflow platform. The platform accepts common image formats such as JPG and PNG, making it flexible for various data sources.

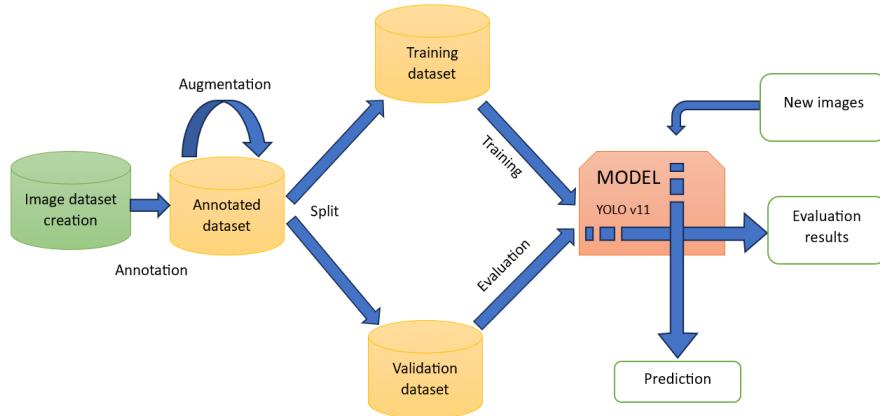


Figure 1. Document element detection workflow based on YOLO v11 and Roboflow

Image preprocessing follows the upload phase, where several crucial transformations are applied. All images are resized to 640x640 pixels to meet YOLOv11's input requirements. The platform automatically handles image orientation correction and performs pixel normalization to ensure consistent input quality. During this stage, any corrupt images are identified and handled appropriately to maintain dataset integrity.

The annotation process represents a critical phase in the workflow. Using Roboflow's annotation tool, each image is carefully labeled with bounding boxes for four distinct classes: paragraphs of text, tables, signatures, and stamps. This process requires careful attention to maintain consistency in annotation style and accuracy across the entire dataset.

Following annotation, the dataset is strategically split into three subsets. The training receives 70% of the data, while the validation receives 20% and test sets receive 10%. This division ensures sufficient data for training while maintaining independent sets for validation and final performance evaluation.

Data augmentation techniques are then applied to enhance the model's robustness. The process includes rotating images within a $\pm 15^\circ$ range, adjusting brightness by $\pm 20\%$, adding random noise, and applying slight blur effects of 0.5 pixels. These augmentations effectively triple the dataset size and help the model generalize better to various document conditions. After the augmentation phase, the dataset size increased to 936 images.

The training phase utilizes the YOLOv11 architecture with carefully tuned hyperparameters. The model trains use a batch size of 64 across 300 epochs. Early stopping with a patience of 10 epochs is implemented to prevent overfitting. Model checkpoints are saved throughout training to preserve the best-performing iterations.

Model evaluation employs three key metrics: mean Average Precision (mAP), precision, and recall. These metrics provide comprehensive insight into the model's detection capabilities across all classes. A confusion matrix analysis helps identify specific areas where the model might be struggling with document elements.

If the evaluation reveals suboptimal performance, the workflow enters an iteration phase. This might involve adjusting hyperparameters, expanding the training dataset, modifying augmentation strategies, or reviewing annotation quality. Once the model achieves satisfactory performance across all metrics, it moves to the deployment phase.

The final deployment stage involves exporting the model in the required format, setting up API endpoints for inference, and implementing the necessary pipeline for production use.

4. Experimental evaluation

The training of the model was conducted using the Roboflow platform, leveraging the You Only Look Once (YOLO) v11 architecture [29]. This model addresses one of the most complex challenges in computer vision: the precise identification and localization of objects in real-time applications. Since its introduction, YOLO has gained recognition for its straightforward yet effective design. Unlike conventional multi-stage methods, YOLO processes the entire image in a single pass through a neural network, simultaneously predicting bounding boxes and class probabilities without requiring prior priming. This approach not only enhances efficiency but also ensures rapid image processing without sacrificing accuracy.

A notable limitation in earlier YOLO versions was the Cross Stage Partial (CSP) mechanism. To overcome this, YOLOv11 incorporates the C3K2 block, which improves the information flow within the network. By dividing the feature map and applying a sequence of 3x3 convolutions, the C3K2 block achieves faster and more computationally

efficient operations compared to larger kernels. This design allows for superior feature representation while reducing the number of parameters relative to YOLOv8.

YOLOv11 retains the Spatial Pyramid Pooling Fast (SPFF) module, which aggregates features from different regions and scales within an image. This capability strengthens the network's ability to detect objects of diverse sizes. Another significant innovation in YOLOv11 is the Cross Stage Partial with Spatial Attention (C2PSA) block. This component employs attention mechanisms to direct the model's focus toward critical areas in the image, particularly useful for detecting small or partially obscured objects.

The model was trained with a batch size of 64 over 300 epochs. After training the model on the curated dataset, its performance was assessed using the Roboflow platform. The real-time detection engine was tested on custom images, with some results illustrated in Fig. 2.

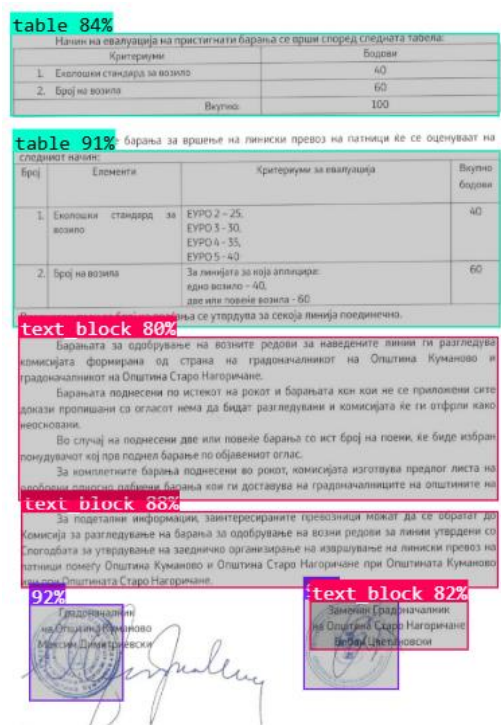


Figure 2. Real-time detection of document elements in a sample image.

5. Results and discussion

The training and validation metrics are depicted in Fig. 3. The training graphs demonstrate steady progress in key metrics. Both the box loss and classification loss curves exhibit a consistent decline for training and validation, indicating improved accuracy in bounding box localization and object classification. Similarly, the focal loss distribution shows ongoing refinement in predicting box boundaries.

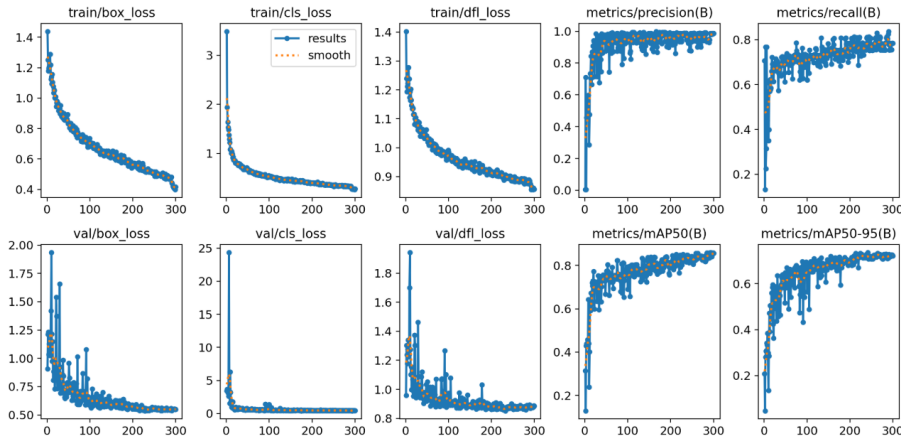


Figure 3. Training and validation performance graphs.

Post-training evaluation yielded the following metrics: mAP: 98.5%, Precision, 98.7% and Recall 95.6%. The steady rise in precision and recall suggests a reduction in false positives and an increase in correctly identified objects. Additionally, the mAP50 and mAP50-95 values show continuous improvement, underscoring the model's robust performance even under stringent evaluation criteria. These trends indicate effective learning without overfitting, achieving a harmonious balance between training and validation (Fig. 4).

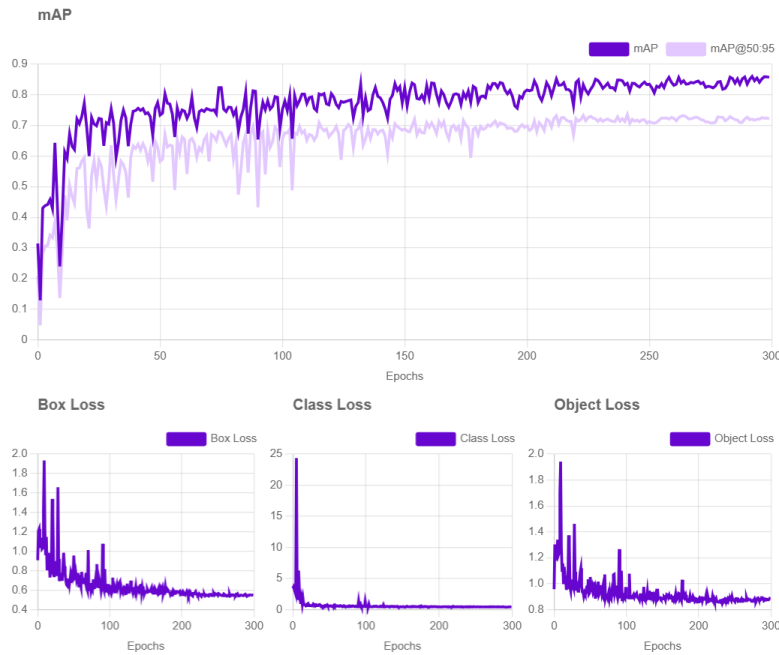


Figure 4. mAP50 and mAP50-95 values show continuous improvement.

6. Conclusion

In this paper, we have demonstrated a robust and efficient workflow for detecting and classifying key document elements—such as paragraphs, tables, signatures, and stamps—using deep learning models implemented via the Roboflow platform. By addressing challenges such as complex spatial relationships, overlapping boundaries, and diverse document formats, our approach leverages advanced data augmentation techniques and a comprehensive toolkit to achieve high accuracy and adaptability across various document types and quality levels. The experimental results highlight the system's effectiveness in real-world applications, showcasing its ability to precisely identify an extract of both structured and unstructured document components. This work underscores the potential of deep learning in advancing document understanding and processing, offering a scalable solution for automating tasks that require high precision and efficiency. Future research could explore extending this framework to additional document types and further optimizing model performance for edge deployment scenarios as well as its direct integration with Large Language Models (LLMs).

References

- [1] Koceski, S. and Petrevska, B., 2012. Empirical evidence of contribution to e-tourism by application of personalized tourism recommendation system. *Scientific Annals of the “Alexandru Ioan Cuza” University of Iasi—Economic Sciences Section*, (1), pp.371-383.
- [2] Trajkovic, V., Vlahu-Gjorgievska, E., Koceski, S. and Kulev, I., 2015. General assisted living system architecture model. In *Mobile Networks and Management: 6th International Conference, MONAMI 2014, Würzburg, Germany, September 22-26, 2014, Revised Selected Papers 6* (pp. 329-343). Springer International Publishing.
- [3] Stojanov, D. and Koceski, S., 2014, September. Topological MRI prostate segmentation method. In *2014 Federated Conference on Computer Science and Information Systems* (pp. 219-225). IEEE.
- [4] Koceska, N., Komadina, R., Simjanoska, M., Koteska, B., Strahovnik, A., Jošt, A., ... & Trontelj, J. (2020). Mobile wireless monitoring system for prehospital emergency care. *European Journal of Trauma and Emergency Surgery*, 46, 1301-1308.
- [5] Stojanov, D., Mileva, A. and Koceski, S., 2012. A new, space-efficient local pairwise alignment methodology. *Advanced Studies in Biology*, 4(2), pp.85-93.
- [6] Koceski, S., Koceska, N., Lazarova, L. K., Miteva, M., & Zlatanovska, B. (2025). Exploring the performance of ChatGPT for numerical solution of ordinary differential equations. *Journal of Technology and Science Education*, 15(1), 18-34.
- [7] Velinov, A., Koceska, N., & Koceski, S. (2024). Virtual Tour Using Telepresence Robot and MQTT Protocol. *TEM Journal*, 13(1), 750-756.
- [8] Koceski, S., Koceska, N. and Koccev, I., 2012. Design and evaluation of cell phone pointing interface for robot control. *International Journal of Advanced Robotic Systems*, 9(4), p.135.
- [9] Koceski, S., Panov, S., Koceska, N., Zobel, P.B. and Durante, F., 2014. A novel quad harmony search algorithm for grid-based path finding. *International Journal of Advanced Robotic Systems*, 11(9), p.144.
- [10] Koceska, N., Koceski, S., Durante, F., Zobel, P.B. and Raparelli, T., 2013. Control architecture of a 10 DOF lower limbs exoskeleton for gait rehabilitation. *International Journal of Advanced Robotic Systems*, 10(1), p.68.
- [11] Koceska, N., & Koceski, S. (2022). Supporting elderly's independent living with a mobile robot platform. *The Journal of Universal Computer Science-JUCS*, 28(5), 475-498.
- [12] Koceska, N., Koceski, S., Beomonte Zobel, P., Trajkovic, V., & Garcia, N. (2019). A telemedicine robot system for assisted and independent living. *Sensors*, 19(4), 834.

- [13] Koceska, N. and Koceski, S., 2014. Financial-Economic Time Series Modeling and Prediction Techniques—Review. *Journal of Applied Economics and Business*, 2(4), pp.28-33.
- [14] Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from <https://roboflow.com>. computer vision.
- [15] Chen, K., Yin, F., & Liu, C. L. (2013, August). Hybrid page segmentation with efficient whitespace rectangles extraction and grouping. In 2013 12th International Conference on Document Analysis and Recognition (pp. 958-962). IEEE.
- [16] O’Gorman, L. (1993). The document spectrum for page layout analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 15(11), 1162-1173.
- [17] Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., ... & Zhou, L. (2020). Layoutlmv2: Multi-modal pre-training for visually-rich document understanding. *arXiv preprint arXiv:2012.14740*.
- [18] Appalaraju, S., Jasani, B., Kota, B. U., Xie, Y., & Manmatha, R. (2021). Docformer: End-to-end transformer for document understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 993-1003).
- [19] An, N., Huang, L., Hu, M., Zhu, J., & Wang, C. (2024). Improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5. *AI EDAM*, 38, e22.
- [20] Nandy, A., Kapadnis, M. N., Patnaik, S., Butala, Y. P., Goyal, P., & Ganguly, N. (2023). FastDoc: Domain-Specific Fast Continual Pre-training Technique using Document-Level Metadata and Taxonomy. *arXiv preprint arXiv:2306.06190*.
- [21] Prasad, D., Gadpal, A., Kapadni, K., Visave, M., & Sultanpure, K. (2020). CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 572-573).
- [22] Sharma, N., Mandal, R., Sharma, R., Pal, U., & Blumenstein, M. (2018, August). Signature and logo detection using deep CNN for document image retrieval. In 2018 16th international conference on frontiers in handwriting recognition (ICFHR) (pp. 416-422). IEEE.
- [23] Sun, H., Tan, C., Pang, S., Wang, H., & Huang, B. (2024). RA-YOLOv8: An Improved YOLOv8 Seal Text Detection Method. *Electronics*, 13(15), 3001.
- [24] Liu, Y., Huang, A., Luo, Y., Huang, H., Liu, Y., Chen, Y., ... & Yang, Q. (2020, April). Fedvision: An online visual object detection platform powered by federated learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 08, pp. 13172-13179).
- [25] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., ... & Lin, D. (2019). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.
- [26] DocEnTr [13] introduced a transformer-based architecture that jointly processes visual and textual features for improved document
- [27] Appalaraju, S., Jasani, B., Kota, B. U., Xie, Y., & Manmatha, R. (2021). Docformer: End-to-end transformer for document understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 993-1003).
- [28] Binmakhashen, G. M., & Mahmoud, S. A. (2019). Document layout analysis: a comprehensive survey. *ACM Computing Surveys (CSUR)*, 52(6), 1-36.
- [29] Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* 2024, *arXiv:2410.17725*.

José Alejandro Ramón Rocha
Universidad de Granada,
E.T.S. de Ingenierías
Informática y de
Telecomunicación, Granada,
Spain

E-mail address: alejandroramon@correo.ugr.es

Elena Jovanovska
Goce Delcev University,
Faculty of computer science
Stip, North Macedonia
E-mail address: elena.210224@student.ugd.edu.mk

Marjan Kotevski
Goce Delcev University,
Faculty of computer science
Stip, North Macedonia
E-mail address: marjan.210223@student.ugd.edu.mk

Blagoj Kotevski
Goce Delcev University,
Faculty of computer science
Stip, North Macedonia
E-mail address: blagoj.210209@student.ugd.edu.mk

Saso Koceski
Goce Delcev University,
Faculty of computer science
Stip, North Macedonia
E-mail address: saso.koceski@ugd.edu.mk