Pegon AI: Toward A Smart Transliteration System For The Preservation Of Pegon Manuscripts

Muhammad Fikri Hidayattullah[1*], M. Ibrahim Hanif[2], Ananda Adhitama Afriansyah[3],
Sharfina Febbi Handayani[4], Yustia Hapsari[5]

[1,2,3,4]Bachelor of Applied Informatics Engineering Study Program, University of Harkat Negeri, Tegal, Indonesia

[5]Digital Business Study Program, Faculty of Economics and Business, Universitas Pancasakti Tegal, Indonesia

*Corresponding author email: fikri@harkatnegeri.ac.id

Abstract

The Pegon script is a valuable cultural and religious literacy heritage within Islamic communities in Indonesia. However, its usage has significantly declined, primarily due to the dominance of the Latin alphabet, which has been extensively digitized. Various preservation efforts such as the development of transliteration tools and virtual keyboards have been introduced, yet most still face limitations in terms of linguistic accuracy, contextual understanding, and user accessibility. To date, no system has been available that can transliterate Pegon text from images either through takeing a photo or selecting from gallery and there is no integration within word processing software to facilitate efficient and accurate digital writing in Pegon script. This study proposes the development of Pegon AI, an intelligent transliteration system powered by artificial intelligence, consisting of two main components: (1) a mobile application that supports image based transliteration of Pegon script through photo capture and gallery selection, and (2) a Microsoft Word add ins capable of accurately converting Latin script to Pegon, thereby supporting culturally contextual digital writing practices. The system utilizes Computer Vision to identify the positions of text lines in images and employs a Vision Language Model (VLM) to perform accurate and context aware transliteration. This research adopts a qualitative methodology, employing case studies and participatory observation to understand user needs and evaluate the potential of the system within relevant social and cultural settings. Such an approach provides a strong foundation for the development of a transliteration system that is adaptive, user oriented, and capable of supporting the preservation of the Pegon script in the digital era.

Key words: Pegon Script, Cultural Heritage Preservation, Artificial Intelligence, Computer Vision, Vision Language Model.

1. Introduction

The Pegon script represents a culturally rich form of Islamic literacy in the Indonesian archipelago, adapted from Arabic script to suit Javanese and Sundanese phonology. Since the 15th century, it has been widely used in pesantren for writing classical Islamic texts and religious commentaries, as exemplified by scholars such as KH Ahmad Rifa'i and KH Sholeh Darat. Its legacy persists in select pesantren, including Babakan Ciwaringin, where Pegon remains central to Islamic education. However, in the digital era its use has declined sharply. Surveys show that only a small fraction of younger generations are literate in Pegon, a decline exacerbated by the absence of Pegon in social media, word processing tools, and online learning platforms [1].

Previous studies highlight similar concerns in other contexts. Research on Gorontalo manuscripts shows that Pegon literacy has become inaccessible to most young readers, leading to proposals for digital transliteration models tailored to Pegon Arabic orthography [2]. More broadly, traditional Javanese manuscripts, many of which employ Pegon, are often regarded as obscure relics with little relevance for modern communication [3]. In Islamic education, the integration of digital technology has been proposed as a means of sustaining culturally embedded literacy practices [4], while community engagement programs in Nganjuk have sought to revitalize Pegon literacy as part of the Salafiyah pesantren tradition [5].

Several preservation efforts such as virtual keyboards and manual transliteration systems have been introduced, yet they remain limited in linguistic accuracy, contextual awareness, and accessibility. Notably, no system has been developed to transliterate Pegon script directly from images, nor has Pegon been integrated into widely used word processors. This gap underscores the urgent need for digital solutions capable of both recognition and transliteration of Pegon script.

Advances in computer vision and natural language processing offer promising pathways. Optical Character Recognition (OCR) enables the conversion of text in images into digital characters [6], while NLP facilitates interpretation of textual data in complex layouts [7]. Richo demonstrated that YOLOv11m combined with PaddleOCR could accurately detect textual information in noisy environments with high precision, suggesting potential for Pegon segmentation [8]. More recently, multimodal models such as Vision Language Models (VLMs) have shown strong performance in diacritic rich Arabic recognition [9]. Lightweight VLMs like

SmolVLM further enable efficient deployment in mobile and embedded platforms, aligning with the needs of real world Pegon applications [10].

Building on these advances, this research introduces Pegon AI, a dual component system consisting of (1) a mobile application for image based transliteration and (2) a Microsoft Word add in for Latin to Pegon conversion. Beyond technical innovation, this work seeks to bridge traditional Islamic literacy with digital ecosystems. Employing qualitative methods including case studies and participatory observation, the study also assesses the sociocultural impact of Pegon digitization. In doing so, it aims to contribute to both the revitalization of Pegon literacy and the broader discourse on inclusive, culturally grounded digital transformation.

2. Method

The proposed methodology integrates a qualitative case study approach with system design and quantitative evaluation. It covers the research approach, system workflow, dataset preparation, implementation environment, and evaluation metrics. The methodological steps are presented in a logical order consistent with the reporting of results.

2.1. Research Approach

This study adopts a qualitative case study approach to capture the cultural and linguistic context of Pegon script usage in Islamic educational settings. Insights were primarily drawn from literature review, secondary sources, and informal discussions with individuals familiar with Pegon. This approach informed the system design, particularly in prioritizing features and linguistic rules for transliteration modules, ensuring cultural relevance and practical applicability. At the same time, a quantitative evaluation was incorporated to assess the system performance. By combining both approaches, the study ensures that the solution is technically robust while maintaining contextual alignment with real world Pegon script practices.

2.2. Implementation Environment

The implementation of Pegon AI was conducted using Python as the primary programming language. Python was selected due to its versatility, extensive library support, and its dominant role in scientific computing and machine learning research [11]. Its simplicity and integration with

deep learning frameworks make it highly suitable for rapid prototyping and deployment in natural language and computer vision systems. For object detection and training, Ultralytics YOLOv11 was adopted as the core framework. YOLOv11 provides a robust and efficient architecture for real-time detection, integrating advanced modules such as spatial pyramid pooling and transformer-based components to enhance both accuracy and computational efficiency [12]. This makes it particularly effective for text line localization in Pegon manuscripts. For image preprocessing and segmentation, Pillow was utilized as it offers efficient and lightweight image manipulation functionalities [13]. Pillow enables cropping, thresholding, and segmentation of Pegon script lines prior to recognition, supporting the modular pipeline of the system. The multimodal transliteration tasks employed SmolVLM, a compact vision language model, which was fine-tuned using the Unsloth framework [14]. Unsloth enables memory efficient fine-tuning strategies, such as parameter-efficient adaptation and quantization, making it feasible to train large multimodal models even under constrained GPU environments. This optimization ensures scalability and adaptability for the Pegon AI system. The backend was implemented using Python and MongoEngine due to their simplicity, flexibility, and seamless integration with AI models, enabling efficient data management and scalability. The mobile application was developed with Flutter, which according to recent studies, offers near-native performance and reduced development complexity, making it well-suited for the Pegon AI system [15].

2.3. Pipeline Architecture And User Interface Design

The proposed Pegon AI system is designed as a dual pipeline transliteration framework, supporting both image based Pegon to Latin transliteration and text based Latin to Pegon transliteration. Each pipeline leverages modern deep learning components tailored for efficiency, cultural relevance, and linguistic accuracy.

a. Image based Pegon to Latin transliteration pipeline

The image based Pegon to Latin transliteration pipeline of Pegon AI operates through sequential computer vision and multimodal inference stages to convert Pegon script images into machine readable Latin text. The process begins with image acquisition, either from direct capture or static upload, which is then analyzed by a YOLOv11n detector to identify

horizontal text lines. These lines are segmented using Pillow to ensure precise extraction, and each segment is passed to SmolVLM, a lightweight vision language model designed to interpret complex Pegon glyphs in context. This modular workflow enables accurate and efficient end to end transliteration, ensuring both recognition reliability and scalability. The complete architecture of the system is depicted in Figure 1.
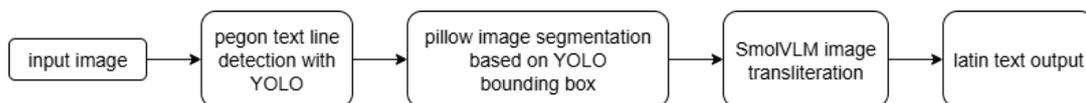

Figure 1. Image based Pegon to Latin transliteration pipeline

b. Text based Latin to Pegon transliteration pipeline

The Latin to Pegon transliteration module in Pegon AI is implemented as a purely text driven pipeline, independent of visual input. Users can provide Latin text directly through the application interface or via Microsoft Word add ins. The input is processed by SmolVLM in text to text generation mode, where its transformer based architecture interprets the linguistic structure and produces Pegon script with semantic, syntactic, and phonological fidelity. This design ensures accurate and culturally aware conversion for educational, religious, and general use cases, supporting the system's mission of digitally preserving Pegon based orthographies. The complete architecture of this module is illustrated in Figure 2.
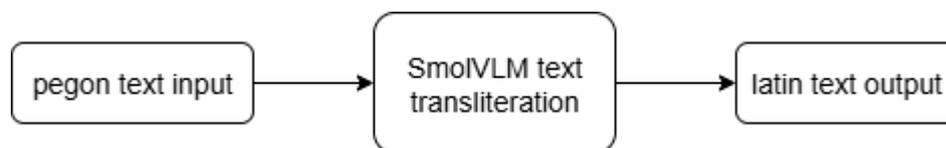

Figure 2. Text based Latin to Pegon transliteration pipeline

c. Mobile application user interface

The mobile application of Pegon AI is designed with a simple and user–friendly interface to support both transliteration pipelines. The design emphasizes clarity, accessibility, and seamless interaction for end–users. Three key user interface modules are provided:

1) Photo Transliteration Module

This interface allows users to capture or upload images containing Pegon script. As shown in Figure 3, users can directly take a photo which is then processed by the Pegon to Latin transliteration pipeline. The processed output, including the recognized text, transliteration result, and historical records, is displayed in a structured format with date and result details for future reference.
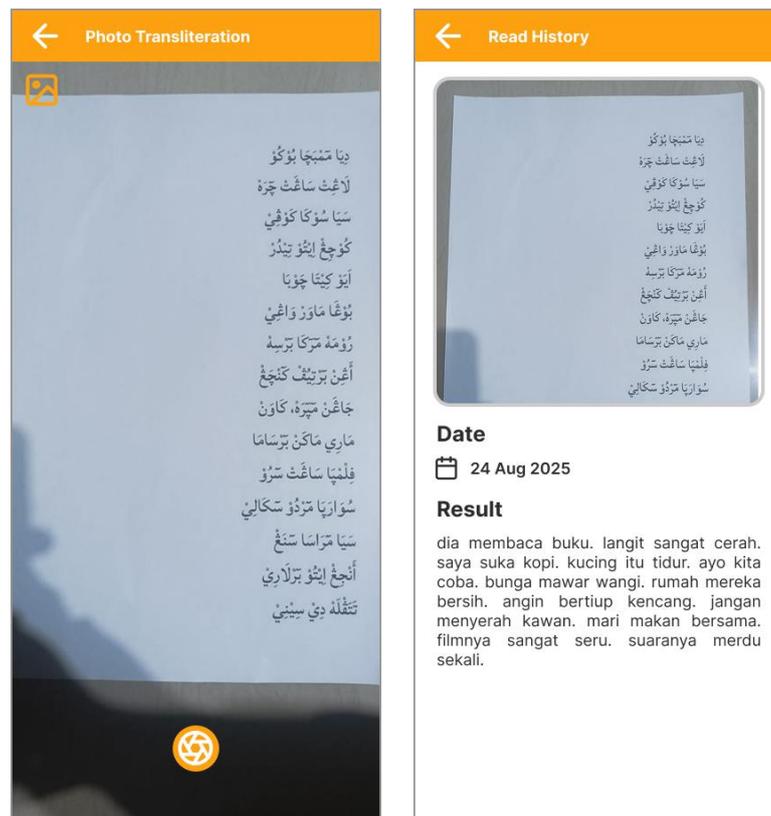

Figure 3. Design of an Image-Based Interface for Pegon-to-Latin Transliteration

2) Text Transliteration Module

The text–based module, illustrated in Figure 4, enables users to input Latin text manually. After submission, the system generates the Pegon transliteration in real–time. The interface is divided into two panels: the Latin input area and the Pegon output area. This dual–pane layout provides immediate feedback and supports copy functionality for easy reuse in other applications.

Figure 4. Design of a Text-Based Interface for Latin-to-Pegon Text Transliteration

3) Microsoft Word Add–in Integration

To extend usability beyond mobile devices, Pegon AI integrates with Microsoft Word through an add–in interface, as presented in Figure 5. The add–in provides a side panel where users can input Latin text, execute transliteration, and insert the Pegon output directly into the document. This integration ensures that Pegon AI is not only practical for mobile use but also effective for academic and professional text editing workflows.
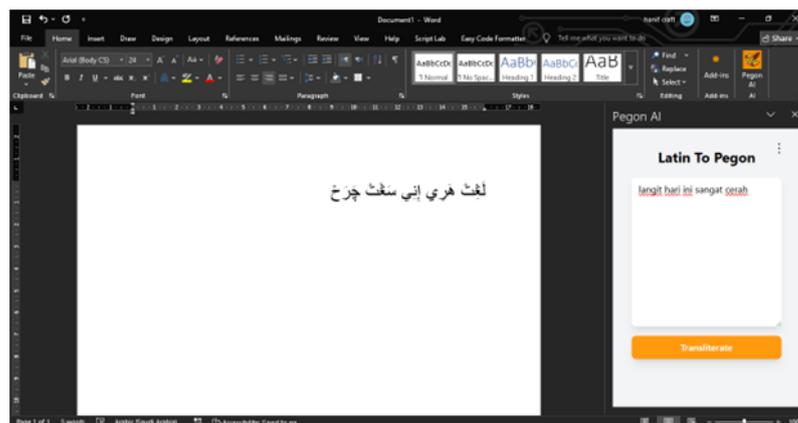

Figure 5. Design of the Pegon AI Add-in Interface for Microsoft Word

2.4. Dataset Preparation

The dataset for YOLOv11n training was derived from printed Pegon manuscripts, with manual annotation performed in Roboflow to generate bounding boxes, ensuring precise ground truth for line level detection. To improve model robustness and generalization, several image augmentations were applied: grayscale conversion on 15% of images, brightness adjustments between -15% and +15%, exposure changes between -10% and +10%, slight blur up to 0.5 px, and noise affecting up to 0.5% of pixels. These augmentations simulate real world variations in manuscript quality and lighting conditions, enhancing the model's ability to detect text under diverse scenarios. Figure 6 presents an example of the annotated dataset used for YOLOv11n training.
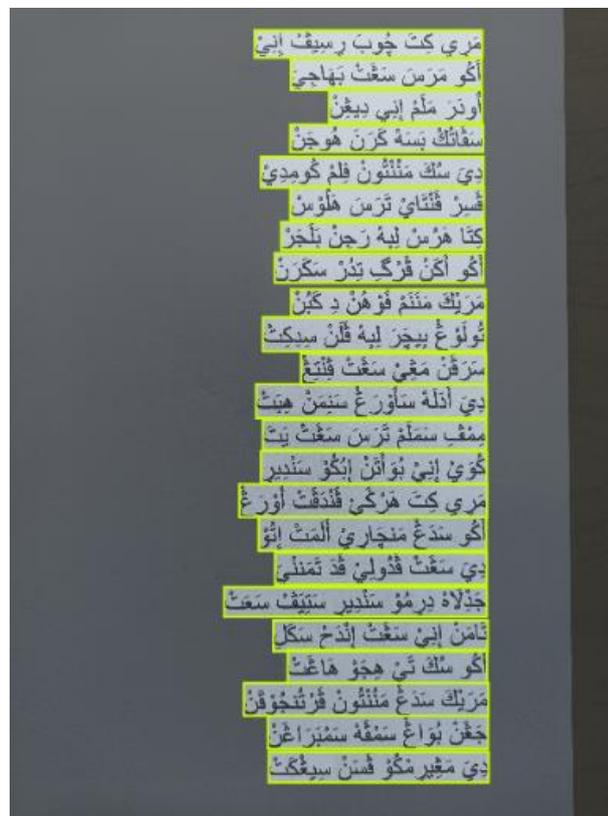

Figure 6. Example anotated dataset for train YOLO

For SmolVLM fine tuning, a multimodal dataset was constructed, consisting of Pegon script images paired with Latin transliterations (Pegon → Latin) and Latin text paired with Pegon

outputs (Latin → Pegon). The dataset underwent targeted augmentations, including brightness modification, blur, binarization, and noise addition, to improve the model's resilience to visual and textual variability. Stored in Parquet format, this dataset supports efficient processing and integration with chat based applications, enabling reliable two way transliteration and comprehensive coverage of both text and image modalities. Figure 7 presents an example of the SmolVLM training dataset, showing paired Pegon and Latin text samples along with their corresponding images and outputs, highlighting the model's two-way transliteration capability.
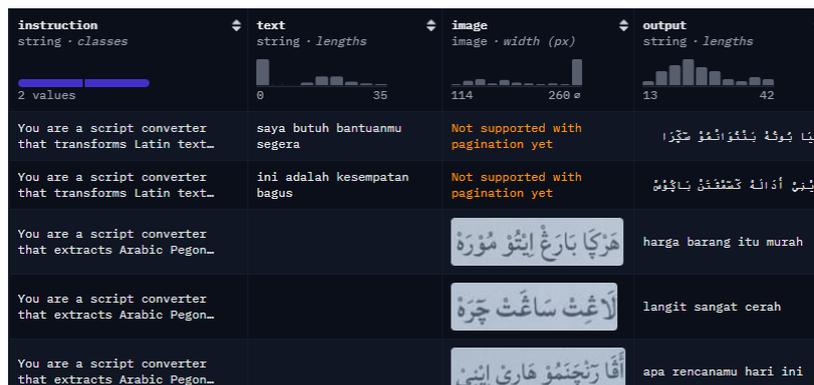

Figure 7. Example dataset for train SmolVLM

2.5. Evaluation Metrics

System performance was evaluated using Character Error Rate (CER), which measures insertions, deletions, and substitutions at the character level [16]. This metric is widely applied in OCR and transliteration systems due to its sensitivity to character level precision. The evaluation was first conducted on printed Pegon text to establish baseline performance. This controlled dataset allows measurable outcomes before expansion into handwritten or regional script variations.

3. Result and Discussion

The evaluation of Pegon AI demonstrates clear strengths in text line detection but also reveals significant limitations in the transliteration stage. The detection module consistently achieved near perfect results, while the transliteration modules exhibited mixed performance depending on input characteristics.

3.1. Quantitative Evaluation

The YOLOv11n detection achieved outstanding performance in identifying Pegon text lines, with precision of 0.9990, recall of 1.0000, mAP@0.5 of 0.9950, and mAP@0.5:0.95 of 0.8293, indicating highly reliable localization across the dataset. In image based transliteration, the SmolVLM model obtained a Character Error Rate (CER) of 0.09 on standard length inputs, demonstrating strong potential for accurate conversion. However, this performance was not consistent: when processing extended lines of text, the model frequently failed, producing outputs that deviated entirely from the source content. In text based transliteration, results were considerably weaker, with CER values exceeding 0.8 due to frequent insertion errors and the generation of incorrect word forms.

3.2. Qualitative Observations

Qualitative outcomes align with the quantitative findings. In the image based task, short and clearly segmented lines could be transliterated with high accuracy, but longer inputs often resulted in complete errors, highlighting the model's sensitivity to sequence length. In the text based task, the outputs frequently contained additional words or altered lexical structures not present in the original input, reducing fidelity and readability. These observations confirm the robustness of the detection stage while emphasizing the need for more effective handling of long sequences and tighter control over transliteration outputs.

4. Conclusion

This study introduced Pegon AI, an intelligent transliteration system designed to support the preservation and revitalization of Pegon script in the digital era. The system integrates two main components: (1) an image-based Pegon-to-Latin transliteration pipeline and (2) a text-based Latin-to-Pegon transliteration module, including a Microsoft Word add-in. Experimental results showed that the detection stage using YOLOv11n achieved near-perfect accuracy, confirming its robustness in localizing Pegon text lines. However, the transliteration stage demonstrated mixed performance. The image-based module produced acceptable results for short and clear text segments but struggled with longer inputs, while the text-based module frequently generated

inaccurate or inconsistent outputs. Despite these limitations, Pegon AI represents a significant step toward bridging traditional Islamic literacy with modern digital ecosystems. By combining computer vision, vision-language models, and user-oriented design, this work provides a foundation for scalable and context-aware Pegon digitization tools. Future research should focus on improving sequence handling in long text transliteration, refining language modeling to reduce insertion and substitution errors, and expanding the dataset to cover handwritten and regional script variations. With continued development, Pegon AI has the potential to become a sustainable solution for preserving cultural heritage while enabling inclusive digital transformation.

Furthermore, the authors recommend that future research explores the integration of Qwen2-VL-2B, a vision-language model equipped with native image size features and a strong foundational knowledge of the Arabic language. Leveraging this model has the potential to enhance transliteration accuracy, particularly in handling long text sequences and improving semantic alignment in Pegon–Arabic contexts.

5. Acknowledgement

6. References

[1] Z. H. Sulistiani, D. N. Rosidin, A. Saefullah, and M. Mujizatullah, "Aksara Pegon dan Transmisi Keilmuan Islam: Potret dari Pesantren Babakan Ciwaringin Cirebon," *EDUKASI J. Penelit. Pendidik. Agama dan Keagamaan*, vol. 21, no. 2, pp. 117–137, 2023, doi: 10.32729/edukasi.v21i2.1415.

[2] I. R. N. Hula, M. Arif, S. Sarif, N. T. Amin, B. Arsyad, and S. R. Saleh, *Digital Transliteration of Pegon Arabic Orthography for Language Preservation in Gorontalo Manuscripts*, vol. 1, no. ICoeSSE. Atlantis Press SARL, 2023. doi: 10.2991/978-2-

38476-142-5_22.

[3]   M. D. Cookson and P. M. R. Stirk, "Pelestarian Bahasa Lokal Studi Kasus Pada Naskah-Naskah Jawa," *Repository.Usd.Ac.Id*, pp. 1–17, 2019, [Online]. Available: https://repository.usd.ac.id/40900/1/7365_Makalah%2BPELESTARIAN%2BBBAHASA%2BLOKAL.pdf

[4]   N. U. Muna, S. Mawaddah, U. I. N. Ar-raniry, P. A. Islam, U. I. N. Ar-raniry, and P. A. Islam, "Integrasi Teknologi Digital dalam Pembelajaran Baca Tulis Arab Melayu: Studi Kasus di Aceh," pp. 1–14, 2023.

[5]   S. Arofik, J. Iswanto, and I. Zumrotus, "Pelestarian Tradisi Salafiyah Di Era 4 . 0 melalui Pengembangan Baca Tulis Pegon Pada Santri Madrasah Diniyah Ulul Hakim Desa Mlorah Kecamatan Rejoso Kabupaten Nganjuk," *J. Pengabdi. Masy.*, vol. 06, no. 01, pp. 13–22, 2023.

[6]   H. Garmouch and O. Abdoun, *Enhancing crane and gate OCR efficiency at container terminal using a hybrid genetic algorithm and neural network model: case study of tangier med port*, vol. 7, no. 7. Springer International Publishing, 2025. doi: 10.1007/s42452-025-07289-3.

[7]   M. Nadas, L. Diosan, and A. Tomescu, "Synthetic Data Generation Using Large Language Models: Advances in Text and Code," *IEEE Access*, vol. PP, p. 1, 2025, doi: 10.1109/ACCESS.2025.3589503.

[8]   R. Richo, "Sistem Identifikasi Informasi Expired Date Produk Kemasan Menggunakan Kolaborasi Metode Yolo-V11M Dan Paddleocr," *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 2, no. 3, pp. 886–900, 2025, doi: 10.70248/jrsit.v2i3.1719.

[9]   A. Wasfy *et al.*, "QARI-OCR: High-Fidelity Arabic Text Recognition through Multimodal Large Language Model Adaptation," 2025, [Online]. Available: http://arxiv.org/abs/2506.02295

[10]   A. Nassar *et al.*, "SmolDocling: An ultra-compact vision-language model for end-to-end multi-modal document conversion," pp. 1–24, 2025, [Online]. Available: http://arxiv.org/abs/2503.11576

[11]   M. Reza *et al.*, "Artifical Intelligence : Image Processing & Application with Python," *Semin. Nas. Pengabdi. Masy. LPPM UMJ*, vol. 1, no. 1, pp. 1–8, 2022, [Online]. Available: http://jurnal.umj.ac.id/index.php/semnaskat

[12]   J. Zophie and H. Himawan Triharminto, "9. Implemetasi Algoritma You Only Look

Once (YOLO) menggunakan Web Camera untuk Mendeteksi Objek Statis dan Dinamis," *TNI Angkatan Udar.*, vol. 1, no. 1, pp. 98–109, 2023, doi: 10.62828/jpb.v1i1.50.

[13] R. Rajab Asaad, R. Ismael Ali, Z. Arif Ali, and A. Ahmad Shaaban, "Image Processing with Python Libraries," *Acad. J. Nawroz Univ.*, vol. 12, no. 2, pp. 410–416, 2023, doi: 10.25007/ajnu.v12n2a1754.

[14] A. Marafioti *et al.*, "SmolVLM: Redefining small and efficient multimodal models," pp. 1–20, 2025, [Online]. Available: http://arxiv.org/abs/2504.05299

[15] P. Srinivasa Rao, B. Pavan, A. Srivastava, K. Venkata Amani, and A. Sharma, "Distinction of Mobile Frameworks: Flutter Vs Native Apps," *Int. Res. J. Mod. Eng. Technol. Sci.*, no. 06, pp. 2582–5208, 2022, [Online]. Available: www.irjmets.com

[16] C. Park, H. Kang, and T. Hain, "Character Error Rate Estimation for Automatic Speech Recognition of Short Utterances," *Eur. Signal Process. Conf.*, pp. 131–135, 2024, doi: 10.23919/eusipco63174.2024.10715433.