

Indonesian handwritten word segmentation using a neural word detector

Zulkarnaen Hatala

Informatics Department, Politeknik Negeri Ambon, Indonesia

ARTICLE INFO

Article history:

Received Oct 1, 2026
Revised Oct 15, 2026
Accepted Oct 21, 2026

Keywords:

Deep Learning;
Document Image;
Indonesian Handwritten
Recognition;
Resnet;
Word Segmentation.

ABSTRACT

Word segmentation is an essential preprocessing step in offline handwritten text recognition because accurate word localization directly affects subsequent recognition performance. This study presents an Indonesian handwritten word segmentation approach using a neural word detector designed to localize individual words in handwritten student assignment documents. The dataset consists of real handwritten student assignments, with word-level ground truth annotations manually created using LabelMe in the form of rectangular bounding boxes. A pretrained neural word detector was trained for five epochs and evaluated using an Intersection over Union (IoU) threshold of 0.50. Predicted word bounding boxes were matched with ground-truth boxes using the Hungarian assignment algorithm, with the matching cost defined from the IoU between corresponding boxes. Evaluation was conducted on 120 paired files containing 9,267 ground-truth word boxes and 9,308 evaluated predicted boxes. The detector achieved a precision of 92.35%, recall of 92.76%, and F1-score of 92.55%, while the mean IoU of matched bounding boxes reached 86.27%. These results indicate that the neural word detector can effectively localize handwritten Indonesian words with high detection consistency and spatial overlap. The study demonstrates the feasibility of neural word detection as a practical approach for word-level segmentation of real Indonesian handwritten documents. The resulting segmentation can provide localized word regions for subsequent handwritten word recognition and document analysis applications.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Zulkarnaen Hatala,
Informatics Department,
Politeknik Negeri Ambon,
Jl. Ir. M. Putuhena, Rumah Tiga, Kec. Tlk. Ambon, Kota Ambon, Maluku, 97234, Indonesia
Email: dzulkarnaenhatala@gmail.com

1 INTRODUCTION

Offline handwritten document processing requires reliable localization of textual units before recognition. Word segmentation is particularly important because errors at this stage can propagate to subsequent handwriting recognition systems. Recent research has increasingly used deep learning to detect and segment handwritten text under variations in writing style, spacing, stroke structure, and document layout (de Sousa Neto et al., 2024; Shivakumara et al., 2022).

Indonesian handwritten documents present variability because they are produced by writers with different writing habits and are often captured in unconstrained student-document environments. Existing handwriting research has addressed word segmentation, recognition, writer characteristics, and writer-related representation learning using convolutional and transformer-based models (Berriche et al., 2024; Elkhayati et al., 2022; Wolf & Fink, 2024).

Document image analysis may also involve preprocessing and structural segmentation before textual regions are analyzed. Document image binarization has been investigated using a selectional auto-encoder to separate document foreground content from the background (Calvo-

Zaragoza & Gallego, 2019). Contour detection and hierarchical image segmentation have also been studied as general approaches for decomposing images into meaningful regions (Arbeláez et al., 2011). Residual convolutional architectures such as ResNet provide hierarchical feature extraction for computer vision applications (K. He et al., 2016). EAST provides an efficient approach for directly predicting text regions in images (Zhou et al., 2017). Handwritten document analysis has progressed through methods addressing segmentation, recognition, and writer-related representation learning. Recent studies emphasize robust representations for variations in handwriting and document appearance (Dhiyf et al., 2023; Jindal & Ghosh, 2023; Kang et al., 2022; Rakshit et al., 2023).

Recent studies have also addressed broader aspects of handwritten document understanding beyond direct word segmentation. Neural models have been investigated for semantic analysis of handwritten document images, emphasizing learned representations for interpreting handwritten document content (Tüselmann & Fink, 2024). Historical document image datasets have also been systematically reviewed, highlighting the importance of dataset characteristics when evaluating document analysis methods (Nikolaidou et al., 2022). Reading-order detection has further been studied in handwritten documents, demonstrating the importance of spatial organization and document structure in handwritten document analysis (Quirós & Vidal, 2022). Axler and Wolf investigated dataset-agnostic word segmentation, treating segmentation as a general localization problem (Axler & Wolf, 2018). The present study applies this perspective to handwritten Indonesian student assignments with manually annotated word bounding boxes.

At the word level, deep neural architectures have been applied to handwritten document word spotting, demonstrating the potential of learned visual representations for word-level localization (Boudraa et al., 2022). More recent research has investigated lightweight U-Net architectures for handwritten word segmentation, including approaches designed for Devanagari handwriting (Kaur et al., 2025). Deep-learning segmentation has also been investigated for handwritten diagrams, showing that segmentation requirements can vary according to the structure and visual characteristics of handwritten content (Pravalpruk & Dailey, 2024). These studies indicate the importance of developing segmentation methods that are appropriate for the specific textual structure and characteristics of the target handwritten documents.

Research on handwritten document processing has also addressed different scripts and recognition conditions. Feature-based methods have been investigated for processing handwritten Arabic text (Saloum & García-Magariño, 2021). Convolutional neural networks have also been applied to Arabic handwriting recognition, demonstrating the use of deep visual representations for handwritten text analysis (Altwaijry & Al-Turaiki, 2021). Retrieval-based language-model adaptation has been investigated for handwritten Chinese text recognition, highlighting the influence of language characteristics on handwritten recognition (Hu et al., 2022). Segmentation methods for unconstrained Malayalam handwritten documents have also been studied to support optical character recognition (Anju et al., 2024). These studies demonstrate that handwriting analysis can be influenced by script, writing characteristics, and document conditions.

Despite progress in handwritten document analysis, evidence concerning word-level segmentation of contemporary Indonesian student handwriting remains limited. Existing studies have investigated document segmentation using deep-learning architectures (Oliveira et al., 2018; Ronneberger et al., 2015), baseline detection in historical document datasets such as cBAD and READ-BAD (Diem et al., 2017; Grüning et al., 2018), and document enhancement for degraded archival material (S. He & Schomaker, 2019). Indonesian handwriting research has also addressed regional-script recognition, including handwritten Javanese character recognition (Wibowo et al., 2017). However, these studies do not directly establish the performance of a neural word detector on real Indonesian student-assignment documents. This leaves a practical evaluation gap concerning word-level localization accuracy under the writing styles, spacing variations, and page layouts present in this document setting.

To address this gap, this study evaluates a pretrained ResNet18-based neural word detector using real Indonesian student-assignment documents with manually annotated word-level bounding boxes. Detection performance is assessed using precision, recall, F1-score, and mean Intersection over Union (IoU), with predicted and ground-truth boxes matched using the Hungarian assignment algorithm and an IoU acceptance threshold of 0.50. The contribution is an empirical evaluation of word-level localization in this specific Indonesian handwriting setting, together with an annotated collection and a defined evaluation protocol.

2 RESEARCH METHOD

This study employs a quantitative experimental approach using deep learning for handwritten document image segmentation.

Dataset

The dataset consists of real handwritten student assignment documents. The images represent naturally produced handwriting rather than synthetic handwriting. Each document contains multiple handwritten words whose locations vary according to writing style, line spacing, word spacing, and page layout.

Word-level ground truth was manually created using LabelMe. For each handwritten word, an annotator drew a rectangular bounding box covering the corresponding word region. LabelMe is a widely used image-annotation environment (Russell et al., 2008). The manually annotated collection provides word-level localization targets for evaluating neural detectors on real Indonesian student handwriting. It contains 120 paired evaluation files and 9,267 ground-truth word boxes, providing a basis for investigating detection performance under variations in writing style, word spacing, and page layout. The annotations may support future comparative evaluations of word detectors, matching procedures, and localization metrics, provided that the corresponding data and annotation protocols can be made available under appropriate privacy and consent conditions. The present study does not claim that the dataset has been publicly released or independently validated as a general-purpose benchmark.

Method

The proposed system uses a neural word detector with a ResNet18-based feature extraction backbone. The backbone provides hierarchical visual representations while the detection component predicts word regions. Pretrained weights were used to initialize the model before adaptation to handwritten student-assignment images (K. He et al., 2016). The detector was fine-tuned for five epochs using pretrained weights as initialization. This setting represents the training configuration evaluated in the present experiment; the study did not establish five epochs as an optimal training duration because no systematic comparison of epoch counts or convergence analysis was conducted. The implemented training pipeline uses an internal training/validation split. Optimization uses Adam (Kingma & Ba, 2015). Evaluation is reported on available paired evaluation images rather than being described as an independent held-out test set.

The principal matching parameter is the IoU threshold, set to 0.50. Predicted and ground-truth word boxes were assigned using Hungarian matching, with assignment cost defined as one minus IoU. Equations (1) and (2) define the IoU measure and Hungarian assignment cost (López et al., 2026; Wu et al., 2021).

$$\text{IoU}(B_{GT}, B_{Pred}) = \frac{|B_{GT} \cap B_{Pred}|}{|B_{GT} \cup B_{Pred}|} \quad (1)$$

$$C_{ij} = 1 - \text{IoU}_{ij} \quad (2)$$

A matched pair is accepted as a valid detection when its IoU is at least 0.50. Hungarian assignment provides one-to-one correspondence between predicted and ground-truth boxes. The experimental parameters used for training and evaluation are summarized in Table 1.

Table 1. Experimental parameters

Parameter	Configuration
IoU acceptance threshold	0.5
Assignment method	Hungarian assignment
Training epochs	5
Initialization	Pretrained weights
Ground truth	Manual LabelMe rectangular word boxes

The evaluation compares predicted word boxes with manually annotated ground-truth boxes. The evaluation contains 120 paired files, 9,267 ground-truth boxes, and 9,308 evaluated prediction boxes. Three prediction-only files were ignored and no ground-truth-only files were recorded. Hungarian matching was used for one-to-one correspondence, and matches below the 0.50 IoU threshold were treated as unmatched.

3 RESULTS AND DISCUSSION

Results

The quantitative performance of the neural word detector on the evaluation dataset is presented in Table 2.

Table 2. Quantitative evaluation results of the neural word detector

Metric	Result
Evaluation files	120 paired files
Ground-truth boxes	9,267
Evaluated prediction boxes	9,308
True positives (TP)	8,596
False positives (FP)	712
False negatives (FN)	671
Precision	92.35%
Recall	92.76%
F1-score	92.55%
Mean IoU	86.27%

The detector achieved a precision of 92.35%, indicating that most evaluated predicted boxes were accepted as valid detections under the specified matching protocol. The recall of 92.76% indicates that most ground-truth word boxes were matched successfully, while the F1-score of 92.55% summarizes the balance between precision and recall. The mean IoU of 86.27% indicates substantial spatial overlap between accepted predicted boxes and their corresponding ground-truth boxes. These metrics measure detection and localization performance; they do not measure the accuracy of recognizing the textual content within each segmented word. Figure 1 provides a visual comparison between the predicted word bounding boxes (b) and the corresponding manually annotated ground-truth regions (a).

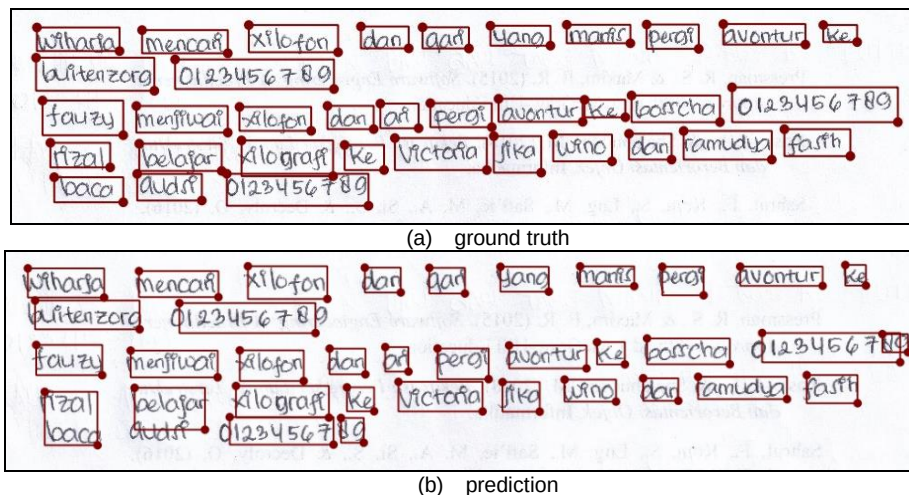


Figure 1. Comparison of ground truth versus predicted boxes

Precision (Hand et al., 2021) measures the proportion of accepted detections among all evaluated predicted boxes:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{8596}{8596 + 712} = \frac{8596}{9308} = 0.9235 = 92.35\%$$

Recall measures the proportion of ground-truth word boxes successfully detected:

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{8596}{8596 + 671} = \frac{8596}{9267} = 0.9276 = 92.76\%$$

The F1-score combines precision and recall through their harmonic mean:

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ = 2 \frac{(0.9235)(0.9276)}{0.9235 + 0.9276} = 92.55\%$$

For each accepted matched pair, IoU quantifies spatial overlap:

$$IoU_i = \frac{|B_{GT,i} \cap B_{Pred,i}|}{|B_{GT,i} \cup B_{Pred,i}|}$$

Mean IoU is calculated over all accepted matched pairs:

$$\text{Mean IoU} = \frac{1}{N_m} \sum_{i=1}^{N_m} IoU_i = 86.27\%$$

Discussion

IoU and Hungarian Assignment evaluation separates detection quality from localization quality. The IoU threshold of 0.50 is an evaluation parameter that determines whether a predicted box is sufficiently aligned with a ground-truth box. Mean IoU is a reported result metric that summarizes spatial overlap of accepted matched pairs. Hungarian assignment provides one-to-one matching so multiple predictions do not independently claim the same ground-truth word.

The evaluation produced 712 false positives and 671 false negatives under the one-to-one Hungarian matching protocol with an IoU acceptance threshold of 0.50. A false positive represents a predicted box that was not accepted as a valid match, whereas a false negative represents a ground-truth word box that remained unmatched. These counts identify the extent of detection errors but do not, by themselves, establish their underlying visual causes. To determine whether the errors were associated with closely spaced words, inaccurate box boundaries, missed words, or detections of non-word regions, representative false-positive and false-negative examples must be inspected. Such explanations should be reported as confirmed findings only when supported by the error analysis.

A mean IoU of 86.27% indicates that the predicted boxes generally overlapped substantially with the corresponding manually annotated regions among accepted matched pairs. The IoU acceptance threshold of 0.50 determines whether a pair qualifies as a valid detection, whereas mean IoU summarizes the average overlap of the accepted pairs. Consequently, the two values serve different purposes: the threshold governs matching acceptance, and mean IoU measures localization quality among accepted matches. The reported mean IoU should not be interpreted as a word-recognition accuracy of 86.27% or as evidence that every predicted box was accurately localized.

The experiment uses real student handwriting and manually annotated word regions. Therefore, the reported performance reflects the evaluated assignment images and should not automatically be generalized to all Indonesian handwriting datasets or document types.

4 CONCLUSION

This study evaluated a pretrained ResNet18-based neural word detector for segmenting handwritten Indonesian words in real student-assignment documents using manually annotated word-level bounding boxes. On the reported evaluation set of 120 paired files, the detector achieved a precision of 92.35%, recall of 92.76%, F1-score of 92.55%, and mean IoU of 86.27%, using Hungarian matching with an IoU acceptance threshold of 0.50. These results provide an empirical baseline for word-level detection and localization in the evaluated document setting. The study's contribution lies in evaluating the detector on this handwriting data and reporting detection and localization metrics under a defined matching protocol, rather than proposing a new detector architecture.

Future work should investigate the causes of false positives and false negatives through systematic error categorization, compare alternative detector architectures and training configurations, and evaluate the robustness of localization under different writing styles and document layouts. Further experiments should also document independent data-splitting

procedures and leakage controls to strengthen the reliability of performance estimates. Finally, the segmented word images should be evaluated in a downstream handwritten text-recognition pipeline to determine whether improvements in word localization translate into better recognition performance. Any wider use of the annotated data should be subject to appropriate privacy, consent, and data-access arrangements.

REFERENCES

- Altwaijry, N., & Al-Turaiki, I. (2021). Arabic Handwriting Recognition System Using Convolutional Neural Network. *Neural Computing and Applications*, 33, 2249–2261. <https://doi.org/10.1007/s00521-020-05070-8>
- Anju, A. T., Chacko, B. P., & Mohamed Basheer, K. P. (2024). Advancements in Segmentation of Unconstrained Malayalam Handwritten Documents for Enhanced OCR. *Indian Journal of Science and Technology*, 17(36), 3763–3775. <https://doi.org/10.17485/IJST/v17i36.2220>
- Arbeláez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 898–916. <https://doi.org/10.1109/TPAMI.2010.161>
- Axler, G., & Wolf, L. (2018). Toward a Dataset-Agnostic Word Segmentation Method. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2635–2639. <https://doi.org/10.1109/ICIP.2018.8451124>
- Berriche, L., Alqahtani, A., & Rekik, S. (2024). Hybrid Arabic Handwritten Character Segmentation Using CNN and Graph Theory Algorithm. *Journal of King Saud University – Computer and Information Sciences*, 36, 101872. <https://doi.org/10.1016/j.jksuci.2023.101872>
- Boudraa, O., Michelucci, D., & Hidouci, W. K. (2022). PUNet: Novel and Efficient Deep Neural Network Architecture for Handwritten Documents Word Spotting. *Pattern Recognition Letters*, 155, 19–26. <https://doi.org/10.1016/j.patrec.2022.01.019>
- Calvo-Zaragoza, J., & Gallego, A.-J. (2019). A selectional auto-encoder approach for document image binarization. *Pattern Recognition*, 86, 37–47. <https://doi.org/10.1016/j.patcog.2018.08.011>
- de Sousa Neto, A. F., Bezerra, B. L. D., Duarte de Moura, G. C., & Toselli, A. H. (2024). Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review. *SN Computer Science*, 5, 258. <https://doi.org/10.1007/s42979-023-02583-6>
- Dhiaf, M., Cheikh Rouhou, A., Kessentini, Y., & Ben Salem, S. (2023). MSdocTr-Lite: A Lite Transformer for Full Page Multi-Script Handwriting Recognition. *Pattern Recognition Letters*, 169, 28–34. <https://doi.org/10.1016/j.patrec.2023.03.020>
- Diem, M., Kleber, F., Fiel, S., Grüning, T., & Gatos, B. (2017). cBAD: ICDAR2017 competition on baseline detection. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 1, 1355–1360. <https://doi.org/10.1109/ICDAR.2017.222>
- Elkhayati, M., Elkettani, Y., & Mourchid, M. (2022). Segmentation of Handwritten Arabic Graphemes Using a Directed Convolutional Neural Network and Mathematical Morphology Operations. *Pattern Recognition*, 122, 108288. <https://doi.org/10.1016/j.patcog.2021.108288>
- Grüning, T., Labahn, R., Diem, M., Kleber, F., & Fiel, S. (2018). READ-BAD: A Dataset and Evaluation Scheme for Baseline Detection in Archival Documents. *International Conference on Document Analysis and Recognition (ICDAR)*. <https://doi.org/10.1109/ICDAR.2017.307>
- Hand, D. J., Christen, P., & Kirielle, N. (2021). F*: An Interpretable Transformation of the F-Measure. *Machine Learning*, 110(3), 451–456. <https://doi.org/10.1007/s10994-021-05964-1>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- He, S., & Schomaker, L. (2019). DeepOtsu: Document enhancement and binarization using iterative deep learning. *Pattern Recognition*, 91, 379–390. <https://doi.org/10.1016/j.patcog.2019.01.025>
- Hu, S.-Y., Wang, Q.-F., Huang, K., Wen, M., & Coenen, F. (2022). Retrieval-Based Language Model Adaptation for Handwritten Chinese Text Recognition. *International Journal on Document Analysis and Recognition*, 26, 109–119. <https://doi.org/10.1007/s10032-022-00419-2>
- Jindal, A., & Ghosh, R. (2023). Word and Character Segmentation in Ancient Handwritten Documents in Devanagari and Maithili Scripts Using Horizontal Zoning. *Expert Systems with Applications*, 225, 120127. <https://doi.org/10.1016/j.eswa.2023.120127>
- Kang, L., Riba, P., Rusiñol, M., Fornés, A., & Villegas, M. (2022). Pay Attention to What You Read: Non-Recurrent Handwritten Text-Line Recognition. *Pattern Recognition*, 129, 108766. <https://doi.org/10.1016/j.patcog.2022.108766>
- Kaur, R., Rani, R., & Pahuja, R. (2025). LUNet: Lightweight U-Net with adaptive scaling for Devanagari handwritten word segmentation. *Engineering Applications of Artificial Intelligence*, 159, 111618. <https://doi.org/10.1016/j.engappai.2025.111618>

- Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1412.6980>
- López, P., Zubasti, P., García, J., & Molina, J. M. (2026). A UAV-Based Framework for Visual Detection and Geospatial Mapping of Real Road Surface Defects. *Drones*, 10(2), 119. <https://doi.org/10.3390/drones10020119>
- Nikolaidou, K., Seuret, M., Mokayed, H., & Liwicki, M. (2022). A Survey of Historical Document Image Datasets. *International Journal on Document Analysis and Recognition*, 25, 305–338. <https://doi.org/10.1007/s10032-022-00405-8>
- Oliveira, S. A., Seguin, B., & Kaplan, F. (2018). dhSegment: A generic deep-learning approach for document segmentation. *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 7–12. <https://doi.org/10.1109/ICFHR-2018.2018.00011>
- Pravalpruk, B., & Dailey, M. N. (2024). Deep Handwritten Diagram Segmentation. *IET Image Processing*, 18(13), 4207–4220. <https://doi.org/10.1049/ipr2.13243>
- Quirós, L., & Vidal, E. (2022). Reading Order Detection on Handwritten Documents. *Neural Computing and Applications*, 34, 9593–9611. <https://doi.org/10.1007/s00521-022-06948-5>
- Rakshit, P., Halder, C., Sk, M. O., & Roy, K. (2023). A Generalized Line Segmentation Method for Multi-Script Handwritten Text Documents. *Expert Systems with Applications*, 212, 118498. <https://doi.org/10.1016/j.eswa.2022.118498>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision*, 77, 157–173. <https://doi.org/10.1007/s11263-007-0090-8>
- Saloum, S. S., & García-Magariño, I. (2021). Algorithm Based on Normal Coordinate Vectors with 16 Segments for the Data Fusion from Hand-Written Arabic Text Implemented with MATLAB. *PeerJ Computer Science*, 7, e705. <https://doi.org/10.7717/peerj-cs.705>
- Shivakumara, P., Jain, T., Pal, U., Surana, N., Antonacopoulos, A., & Lu, T. (2022). Text Line Segmentation from Struck-Out Handwritten Document Images. *Expert Systems with Applications*, 210, 118266. <https://doi.org/10.1016/j.eswa.2022.118266>
- Tüselmann, O., & Fink, G. A. (2024). Neural Models for Semantic Analysis of Handwritten Document Images. *International Journal on Document Analysis and Recognition*, 27, 245–263. <https://doi.org/10.1007/s10032-024-00477-8>
- Wibowo, M. A., Soleh, M., Pradani, W., Hidayanto, A. N., & Arymurthy, A. M. (2017). Handwritten Javanese Character Recognition Using Discriminative Deep Learning Technique. *2017 2nd International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 325–330. <https://doi.org/10.1109/ICITISEE.2017.8285521>
- Wolf, F., & Fink, G. A. (2024). Self-Training for Handwritten Word Recognition and Retrieval. *International Journal on Document Analysis and Recognition*, 27, 225–244. <https://doi.org/10.1007/s10032-024-00484-9>
- Wu, H., Du, C., Ji, Z., Gao, M., & He, Z. (2021). SORT-YM: An Algorithm of Multi-Object Tracking with YOLOv4-Tiny and Motion Prediction. *Electronics*, 10(18), 2319. <https://doi.org/10.3390/electronics10182319>
- Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: An Efficient and Accurate Scene Text Detector. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5551–5560. <https://doi.org/10.1109/CVPR.2017.283>