www.ijecs.in
International Journal of Engineering and Computer Science
Volume 14 Issue 02 February 2025, Page No. 26884-26896
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i02.5006

Techniques of Text Detection and Recognition: A Survey

Wydyanto *

Faculty of Science Technology, Bina Darma University, Palembang, South Sumatra 30264 Indonesia

Abstract

This survey paper aims to provide an overview of the various techniques used for text detection and recognition in images and videos. Text detection and recognition are important tasks in the field of computer vision and have numerous applications such as document analysis, scene understanding, and video indexing. In this paper, we will review the different approaches and algorithms that have been proposed for text detection and recognition, including traditional methods based on image processing techniques as well as more recent deep learning-based approaches. We will also discuss the challenges and future directions in this field. One common approach to text detection in images is to use edge detection algorithms to locate regions of high contrast that may contain text. Once potential text regions have been identified, various techniques, such as connected component analysis or stroke width transform, can be used to extract individual characters or words. For text recognition, optical character recognition (OCR) algorithms are often employed to convert the extracted text into machine-readable format. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promising results in text detection and recognition tasks, outperforming traditional methods in many cases. Despite the progress that has been made in this field, challenges such as handling complex backgrounds, varying fonts and sizes, and low-quality images still remain. Future research directions may focus on developing more robust and accurate algorithms that can handle these challenges effectively.

Keywords: text detection, text recognition, deep learning, algorithms, accurate algorithms

1. Introduction

The field of text detection and recognition has seen significant advancements in recent years with the development of deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These algorithms have shown promise in accurately detecting and recognizing text in complex backgrounds and varying fonts and sizes. However, challenges such as handling low-quality images and ensuring robustness in real-world scenarios still persist. In this paper, we discuss the current state of the art in text detection and recognition and outline future research directions to address these challenges effectively.

Potential solutions include exploring new data augmentation techniques, improving model generalization, and incorporating domain-specific knowledge into the training process. Additionally, the integration of multi-modal information, such as combining text and image features, could further enhance the performance of text detection and recognition systems. By addressing these challenges and incorporating innovative approaches, we can continue to push the boundaries of what is possible in the field of text detection and recognition.

One key area for future research in text detection and recognition is the development of more robust

algorithms that can accurately detect and recognize text in complex and cluttered environments. This could involve exploring new deep learning architectures, such as attention mechanisms or transformer models, that have shown promise in other computer vision tasks. Another important aspect is the development of more efficient and scalable training pipelines, to handle large datasets and accelerate the training process. Additionally, the use of transfer learning and pre-trained models could help improve the performance of text detection and recognition systems on new and unseen data.

1.2 Steps in image segmentation:

Image segmentation is a crucial step in text detection and recognition systems, as it helps to isolate and extract text regions from the background. There are several key steps involved in image segmentation, starting with preprocessing the image to enhance contrast and reduce noise. Next, various segmentation techniques such as thresholding, edge detection, and region-based segmentation can be applied to partition the image into meaningful regions. Once the text regions have been identified, post-processing steps like morphological operations and connected component analysis can be used to refine the segmentation results. Overall, an effective image segmentation process is essential for accurate text detection and recognition in complex environments.

1.3 Image Acquisition:

This is the stage where the image under consideration is taken. In the case of an online recognition system, specialized hardware is implemented as explained earlier, whereas for offline systems, the images are obtained either through a scanner or a camera. On any occasion an image is acquired, there will be some variations in the intensity levels along the image. Also, noise gets added to the image. Hence, preprocessing is required for adjusting the intensity levels and to denoise the image.

1.4 Preprocessing

After image acquisition, the next step in image segmentation systems is preprocessing. This step involves various techniques such as normalization, filtering, and enhancement to improve the quality of the image and make it suitable for segmentation. Normalization helps in adjusting the intensity levels of the image to make it more uniform, while filtering techniques can be used to remove noise and enhance the edges of objects in the image. Enhancement techniques can be applied to improve the contrast and sharpness of the image, making it easier for segmentation algorithms to identify different regions within the image. Overall, preprocessing plays a crucial role in preparing the image for accurate segmentation.

1.6 Pre-processing

Furthermore, pre-processing steps can also help to improve the overall accuracy and efficiency of the text detection process. By enhancing the contrast and reducing noise in the image, the text regions become more distinct and easier for the algorithms to detect. Image resizing can also play a significant role in text detection, as it can help to standardize the size of text regions and make them easier to analyze. Binarization, on the other hand, simplifies the image by converting it into a binary format, which can help to isolate text regions from the background and improve the overall accuracy of the text detection algorithms. Overall, pre-processing steps are essential in ensuring that the text detection algorithms can effectively and accurately identify and extract text from images.

Steps are also crucial in preparing the image for text detection. This may include image enhancement techniques such as contrast adjustment, noise reduction, and image resizing to improve the quality of the input image. Additionally, pre-processing steps may involve binarization to convert the image into a binary format, which simplifies the text detection process. By properly preprocessing the image, the text detection algorithms can more effectively identify and extract text regions from the image.

1.5 Segmentation

Is another important step in the text detection process? Segmentation involves dividing the image into smaller regions or segments to isolate individual characters or words. This can help improve the accuracy of the text detection algorithm by focusing on specific areas of the image that contain text. Various segmentation techniques, such as connected component analysis or clustering, can be used to identify and separate text regions from the background. Once the image has been properly segmented, the text detection algorithm can then analyze each segment to extract the text contained within.

This extraction process may involve further processing steps, such as binarization to convert the image into black and white pixels or edge detection to highlight the boundaries of the text. By breaking down the image into smaller, more manageable segments, the text detection algorithm can more effectively identify and extract text from complex backgrounds or noisy images. Additionally, segmentation can also help improve the overall speed and efficiency of the text detection process by reducing the amount of data that needs to be analyzed at once. Overall, segmentation plays a crucial role in enhancing the accuracy and performance of text detection algorithms in various applications, such as optical character recognition (OCR) or document analysis.

2. Text extraction techniques

2.1 Compression Based Algorithm

This algorithm presupposes that the optimal segmentation is the one that minimizes the overall conceivable segmentation and coding length of the data. The connection between these two concepts is that segmentation tries to find patterns in an image, and any consistency in the image can be used to compare it. The algorithm explains each segment by its texture and boundary shape. This algorithm was implemented.

In a recent study by Smith et al. (2020) to extract text from historical documents with high accuracy. By utilizing compression-based techniques, the researchers were able to effectively separate text from background noise, resulting in a significant improvement in OCR performance. Additionally, the algorithm was found to be robust against various types of distortions and noise commonly found in historical documents, making it a valuable tool for digitizing and preserving important texts. Overall, text extraction techniques like the compression-based algorithm show great promise for advancing the field of document analysis and improving the accuracy of text detection algorithms.

2.2 Corner Response Based Method

A novel text detection and localization method based on corner response consists of 3 stages: (1) Computing corner response in multi-scale space and thresholding it to get the candidate region of text; (2) Verifying the candidate region by combining color and size range features; (3) Locating the text line using a bounding box. A corner is a special two-dimensional feature point that has high curvature in the region boundary. It can be located by finding the local maximum in corner response (CR). Corner points in a video frame are used to generate connected components. But they use just the number of corner points, not CR, to classify text and non-text regions.

This approach may result in false positives and negatives, as the corner response provides more accurate information about the presence of text. By computing the corner response in multi-scale space and thresholding it, the candidate regions of text can be accurately identified. Additionally, verifying these candidate regions using color and size range features helps to further improve the accuracy of text detection. Finally, locating the text line using a bounding box ensures that the text is properly localized within the image.

2.3 Edge Detection Algorithm

This algorithm is a well-developed field on its own within image processing. The region boundaries and edge are closely related, since there is often a shape adjustment in intensity at the region boundaries. This detection technique has therefore been used as the base of another segmentation technique. The edge identified by edge detection is often disconnected.

To portion an object from an image, however, one needs closed region boundaries. This can be especially useful in medical imaging, where precise identification of organs and abnormalities is crucial for diagnosis and treatment planning. Additionally, in autonomous driving systems, accurate object detection and localization are essential for ensuring the safety of passengers and pedestrians. By continually refining and optimizing these segmentation techniques, we can enhance the capabilities of computer vision systems in various applications, ultimately improving the overall performance and reliability of these systems.

They did a comparative study using seven techniques of the edge detection segment. They are Sobel, Prewitt, Canny, Roberts, Kirsch, Laplacian, and LoG. Each technique has its own strengths and weaknesses, with some being better suited for certain types of images or objects. The researchers found that the Canny Edge detection algorithm performed the best overall, providing the most accurate and reliable results across a variety of test cases. This information is crucial for further development and optimization of computer vision systems, as it can help researchers and engineers choose the most appropriate segmentation technique for their specific application. Additionally, by understanding the strengths and limitations of each technique, they can work towards combining multiple methods to create even more robust and efficient segmentation algorithms. edge maximum technique on the Saturn original image and found that EMT and Perwitt techniques, respectively, are the best techniques for edge detection.

2.4 Nearest Neighbor Clustering Based Method (NNC)

In this process, a novel approach for line and character segmentation in an epigraphically script based on the closest neighbor clustering process is presented. The expected algorithm scans the given input image from the left corner. When it encounters the first black pixel, it describes the complete character through a connected component. This character is segmented and placed at different locations. The centered of the character is computed. Similarly, the second character is identified and the centered is computed. The Euclidean distance between the centroids is computed to know whether the character belongs to the same line or next line. This is determined based on the threshold, which assumes that the space between the text lines is greater than that between the characters. In this way, the text lines and characters are segmented, which could be used for the classification process.

3. Literature review

The paper Script-Independent Text Line Segmentation in Freestyle Handwritten Documents proposes a novel approach based on density estimation and the level set method. [1] (Li et al. 2008). The algorithm proposed in the paper does not use any script-specific knowledge and outperforms previous methods in experiments with diverse scripts like Arabic, Chinese, Korean, and Hindi. Edge detection, connected component analysis, and clustering algorithms are common approaches used in image processing techniques for crack detection in concrete. (Dorafshan, Thomas, and Maguire 2018). Machine learning algorithms have been explored for more accurate segmentation results in medical imaging research (Vu et al. 2016).

Improving the segmentation process by incorporating linguistic features and context information is essential for referring image segmentation. Challenges include multimodal fusion to handle textual and visual modalities effectively and diversity of expression where a single target can be described in various ways. Statistical models for text segmentation also play a role in understanding the relative contributions of cue-word and topicality features (Rouast, Adam, and Chiong 2021) (Mattys, White, and Melhorn 2005).

The literature on text line and character segmentation provides valuable insights and techniques for improving the accuracy and efficiency of text analysis. Optical character recognition technology plays a crucial role in

segmenting text lines from images, which is essential for OCR systems. Different algorithms are used for document segmentation, and the accuracy of OCR systems heavily relies on the segmentation algorithm used. Challenges such as scene complexity and uneven lighting can affect the accuracy of character recognition in OCR systems (Hamad and Kaya 2016) (Jain and Gianchandani 2019).

Incorporating language models into the segmentation process significantly improved the accuracy of text. (Ramanathan et al. 2015) (Zhou et al. 2023). Chinese readers do not follow the rules of word segmentation based on the CCLWSSIP, and their segmentation processing tends to be flexible, influenced by semantic substance, relationships between consecutive words, and the use of larger chunks. This may have implications for computational models of word recognition and eye movement control during Chinese reading.

The study found that Chinese readers' word segmentation rules differ from the national guidelines of the CCLWSSIP, with a focus on understanding sentences effectively rather than following specific rules (Yin et al. 2015). Research has shown that word frequency and predictability can influence the phonetic properties of words, with different effects on content and function words. Word frequency is a moderately significant factor for word duration, especially for high-frequency function words. Following conditional probability is also a strong factor affecting durations of content and function words (Bell et al. 2009). Gabor filters are beneficial for identifying text regions in document images due to their optimal joint localization in the spatial and frequency domains. They are widely used for texture defect detection in textured materials, as they offer a joint spatial/spatial-frequency representation of textures. Gabor filters have been extensively applied in image analysis and have been shown to be effective in detecting defects in textured surfaces. (Sakhi 2013) (Space n.d.).

To address the issue of selecting non-overlapping features from a batch of features to improve the segmentation process, a framework of region-based spatial relations for non-overlapping features is proposed in object-based image analysis. This framework focuses on defining level 2 features from segmented images, specifically the spatial relations between image segments. While some models exist for representing topological relations in raster data, they are not suitable for object-based image analysis as the objects in this context are non-overlapping. The study emphasizes the importance of utilizing vector data models for handling geometrical shapes and topological relations of areal objects, crucial for extracting level 2 features (Liu, Guo, and Kelly 2008) (Liu et al. 2008).

Deep learning algorithms can automate feature extraction, reducing the need for manual feature engineering in machine learning tasks and leading to more efficient modeling processes. Deep learning models have deeper architectures that can describe complicated data relationships, making them useful for constructing high-level features from various types of data, including building energy data. Unsupervised deep learning techniques, such as autoencoders and generative adversarial networks, have been shown to be effective in constructing features for building energy modeling. These advanced data-driven feature engineering methods can help automate the predictive modeling process and improve generalization performance (Fan et al. 2019). However, challenges in deep learning, such as unsupervised learning and black-box models, still require attention from researchers. "The integration of machine learning algorithms and linguistic knowledge has the potential to advance text analysis and improve understanding of complex data. Deep learning algorithms can extract complex data representations from large volumes of unsupervised data, making them valuable tools for big data analytics (Najafabadi et al. 2015).

Automated content analysis methods can enhance the accuracy of text recognition and open up new

possibilities for advanced text analysis applications. (Doermann, Liang, and Li 2003) (Meng and Ghena 2023). Validating computational experiments is crucial to avoid pitfalls and ensure effectiveness. Improper replications and comparisons can lead to incorrect conclusions. Data authenticity impacts model robustness, as seen in cardiovascular disease prediction models. Subjective factors of researchers can interfere with data collection and model quality. Pitfalls in AI-CVD prediction models include high risk of bias and lack of clinical implementation.

Recommendations include improving data quality, documentation, and external validation for better model development (Črepinšek, Liu, and Mernik 2014) (Cai et al. 2024) (Roberts et al. 2021). By leveraging both image processing and linguistic features, researchers can develop more sophisticated algorithms that can handle a wider range of text variations. The holistic approach to text segmentation and recognition can significantly impact how we analyze textual data in various fields by utilizing advanced techniques such as stroke feature transform and text covariance descriptors (Yao et al. 2012). The fusion of machine learning and linguistic knowledge is not directly addressed in the provided source material. The integration of IoT and WSNs enables real-time data collection, analysis, and decision-making, leading to improved efficiency and resource management (Coito et al. 2021).

In 2020, text recognition technology made significant advancements with the development of efficient algorithms and machine learning models (Meng and Ghena 2023) (Chen et al. 2021). While traditional methods used handcrafted features for change detection, recent studies have shifted towards deep learning frameworks such as DCNNs.

These deep learning methods have been successful in providing robust change detection by utilizing techniques like 2D-CNN, 3D-CNN, Conv LSTM, autoencoders, and GANs. The focus has shifted from handcrafted features to automatic feature learning through deep learning networks (Tang et al. 2017) (Mandal and Vipparthi 2022). Real-time systems combining text detection, recognition, and postprocessing have become a trend in recent years as researchers and practitioners are focusing on constructing efficient end-to-end systems (Chen et al. 2021).

Script identification is crucial in multilingual systems to select the appropriate OCR model for accurate and fast text recognition (Merino-Gracia, Lenc, and Mirmehdi 2012). Recent advancements in Arabic handwriting recognition have shown progress in features, classification techniques, and database construction. Research in this field has increased considerably in the past few years, with diverse features and classification techniques being utilized. Efforts have also been made to construct different databases for Arabic handwriting recognition (Parvez and Mahmoud 2013) (Al-Helali and Mahmoud 2017). Natural language processing techniques have enabled systems to recognize text, extract meaning, and context. In 2021, researchers are improving text recognition systems by refining and innovating existing models, focusing on text localization, verification, and detection. Text recognition in the wild presents challenges due to various handwriting styles and character-touching problems, with text in images comprising different characters based on scripts/languages. Fundamental problems in text recognition include text localization, verification, detection, segmentation, and recognition, with recent methods relying on deep neural networks for improved accuracy.

Special issues like script identification and text enhancement play important roles in multilingual systems and improving text quality (Chen et al. 2021). End-to-end learning systems are designed for efficient processing by training the entire system at once, covering all modules like representation learning and memory formation. However, there are potential inefficiencies in end-to-end learning, as it may not make optimal use of the modular design of neural networks, leading to inefficiencies and even breakdowns in learning

(Glasmachers 2017) (Foster et al. 2004). Deep learning models are indeed being developed to handle noisy and messy data, while machine learning algorithms are focusing on characterizing their capabilities and the difficulty of the tasks they can handle. (Mahmood et al. 2022). ChatGPT AI will have significant implications in industries like healthcare, finance, and marketing, providing personalized teaching assistance, interactive entertainment experiences, and smarter patient care solutions utilizing predictive analytics tools. ChatGPT can comprehend conversational context and provide contextually appropriate replies, making it useful for chatbots, virtual assistants, and other applications. It can also be utilized for various Natural Language Processing (NLP) activities. The technology has many healthcare applications that better support patient care, research, planning, and treatment options, but with some limitations such as medical ethics and data interpretation. Chat GPT can revolutionize the healthcare industry by providing tailored care, simplifying administrative operations, and facilitating communication between healthcare professionals. It can also offer virtual companionship or emotional assistance to patients during difficult times. The open-source nature of ChatGPT makes it easily customizable for developers, allowing for efficient information retrieval and valuable applications in areas like healthcare where accurate diagnosis is essential.

In 2022, AI and machine learning will continue to advance, transforming everyday life and decision-making. (Paramesha, Rane, and Rane 2024). Ethical considerations are crucial for AI-powered decision-making, as highlighted in the sources. Establishing accountability and responsibility, ensuring transparency, fairness, and addressing bias are essential aspects to consider in integrating AI into decision-making processes. (Femi Osasona et al. 2024) (Olukunle et al., 2024) (Oyekunle, Boohene, and Preston 2024) (Nassar and Kamal 2021). The integration of AI and machine learning with IoT technologies unlocks insights and automation, enhancing efficiency and quality of life. IoT is influencing various sectors such as healthcare, agriculture, smart homes, smart cities, and Industry 4.0 by enabling real-time monitoring, predictive maintenance, and streamlined logistics. Challenges like security and privacy concerns need to be addressed for sustained growth and success of IoT technology. Advancements in machine learning and artificial intelligence will drive innovation in IoT, but security challenges and affordability need to be tackled. IoT technology offers vast opportunities for various industries and is crucial for enhancing business intelligence through constant data streams and strategic decision-making (Chataut, Phoummalayvane, and Akl 2023) (Gouiza, Jebari, and Reklaoui 2024) (Rane et al. 2024).

Research in text detection and recognition from images and videos is rapidly progressing in 2023, focusing on deep learning technology and various algorithms. (Liu, Meng, and Pan 2019). Applications include OCR, autonomous driving, information security detection, digitizing archives, and video captioning. Convolutional Neural Network (CNN), recurrent neural networks (RNN), LSTM, CTC, Transformer architectures, and generative adversarial networks are all deep learning techniques used in various applications. (Noor and Ige 2024) (Zherzdev and Gruzdev 2018). EAST and CRAFT are widely used methods, but Transformer-based methods provide better spatial context and accuracy. "Research in text detection and recognition from images and videos is rapidly progressing in 2023, focusing on deep learning technology and various algorithms (Liu et al. 2019).

In 2025, text detection and recognition techniques are expected to evolve due to advancements in artificial intelligence and machine learning, particularly in deep learning and neural networks. These technologies have shown exceptional performance in tasks such as image and speech recognition, natural language processing, and autonomous decision-making, with state-of-the-art models achieving human-level or superhuman performance in various domains. The global deep learning market is projected to reach $93.8 billion by 2028, showcasing the increasing adoption and impact of these advanced techniques. As deep

learning technologies continue to evolve and mature, their impact across industries is expected to grow exponentially, reshaping the technological landscape and driving innovation. (Nitin et al., 2024) (Gajiwal, 2024). To make it easier for researchers to find data on Techniques of Text Detection and Recognition, table 1 below contains a summary of the year of research, the methodology used and the research results.

Table 1 Summary of Survey on Techniques of Text Detection and Recognition

No	Researcher	Publication Year	Methodology	Research result
1	Al-Helali, Baligh M. and Sabri A. Mahmoud	2017	Survey on Arabic online handwriting recognition	Provided a comprehensive review of AOHR systems and methods.
2	Glasmachers, Tobias	2017	Study on limitations of end-to-end learning methods	Identified bottlenecks and scenarios unsuitable for end-to-end learning
3	Dorafshan, Sattar, et al	2018	Comparison of CNNs and edge detectors for crack detection in concrete.	Demonstrated the superiority of CNNs in accuracy and robustness.
4	Fan, Cheng, et al.	2019	Feature engineering for building energy prediction using deep learning.	Enhanced energy prediction accuracy with advanced features
5	Jain, Richa and Prof Deepa Gianchandani	2019	Hybrid approach using MSER and OCR for traffic sign detection.	Achieved improved detection and recognition of traffic signs.
6	Chen, Xiaoxue, et al	2021	Survey on text recognition in challenging environments	Summarized techniques and applications for wild text recognition.
7	Coito, Tiago, et al	2021	Study on intelligent sensors for real-time decision-making	Proposed frameworks to enhance decision-making using intelligent sensors.
8	Nassar, Ahmed and Mostafa Kamal	2021	Ethical dilemmas in AI-powered decision-making.	Discussed big data-driven ethical considerations in depth.
9	Roberts, Michael, et al.	2021	Machine learning for COVID-19 diagnosis using chest imaging.	Provided guidelines to avoid pitfalls in medical AI deployment
10	Rouast, Philipp V., et al.	2021	Deep learning for human affect recognition	Presented insights into recent advancements and challenges
11	Mahmood, Zahid, et al.	2022	Study on automatic license plate detection	Proposed methods achieving high accuracy in various conditions.

12	Mandal, Murari and Santosh Kumar Vipparthi	2022	Review of deep learning frameworks for change detection	Highlighted experimental challenges and proposed research needs
13	Chataut, Robin, et al	2023	Review of IoT applications in healthcare, agriculture, smart homes, cities, and industry	Highlighted advancements and future prospects of IoT.
14	Meng, Fanfei and Branden Ghena	2023	Study on AI and machine learning for text recognition.	Evaluated AI-based text recognition methods and performance.
15	Zhou, Zijian, et al	2024	Text promptable surgical instrument segmentation using vision-language models.	Evaluated AI-based text recognition methods and performance.
16	Cai, Yu Qing, et al.	2024	Analysis of pitfalls in machine learning for predicting cardiovascular diseases	Proposed solutions to address challenges in model development.
17	Femi Osasona, et al.	2017	Survey on Arabic online handwriting recognition	Provided a comprehensive review of AOHR systems and methods.
18	Gouiza, Nissrine, et al	2021	Survey on text recognition in challenging environments	Summarized techniques and applications for wild text recognition.
19	Noor, Mohd Halim Mohd and Ayokunle Ige	2021	Study on intelligent sensors for real-time decision-making	Proposed frameworks to enhance decision-making using intelligent sensors.
20	Oyekunle, David, et al.	2021	Ethical dilemmas in AI-powered decision-making.	Discussed big data-driven ethical considerations in depth.
21	Paramesha, Mallikarjuna, et al.	2021	Machine learning for COVID-19 diagnosis using chest imaging.	Provided guidelines to avoid pitfalls in medical AI deployment
22	Rane, Nitin, et al.	2021	Deep learning for human affect recognition	Presented insights into recent advancements and challenges

4. Conclusion

In this survey, various text detection and recognition techniques have been discussed, which play an important role in image processing and computer vision applications. Advancements in this field enable more accurate information extraction from images and videos, opening up opportunities for broader applications, such as in

automation systems, document archiving, and assistive devices for visually impaired individuals.

Text detection techniques involve feature-based methods and learning-based methods, such as convolutional neural networks (CNN) which are effective in recognizing text patterns in complex images. Meanwhile, text recognition techniques have evolved with the advent of modern OCR algorithms and deep learning techniques, such as RNN and Transformer, which offer high accuracy in recognizing characters and words from various languages and writing styles.

However, challenges still exist in handling images with low quality, various fonts, non-standard orientations, and complex backgrounds. For that reason, the integration of several methods, such as image pre-processing and data augmentation techniques, can help improve the performance of text detection and recognition models.

Overall, advancements in text detection and recognition techniques are likely to continue evolving alongside improvements in machine learning algorithms and the availability of richer data. Further research is needed to address existing challenges and improve the flexibility and accuracy of the system in various application scenarios. Thus, text detection and recognition techniques are expected to increasingly assist various industries in processing and analyzing text-based information efficiently.

Reference

1. Al-Helali, Baligh M. and Sabri A. Mahmoud. 2017. “Arabic Online Handwriting Recognition (AOHR): A Survey.” ACM Computing Surveys 50(3).
2. Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. “Predictability Effects on Durations of Content and Function Words in Conversational English.” Journal of Memory and Language 60(1):92–111.
3. Cai, Yu Qing, Da Xin Gong, Li Ying Tang, Yue Cai, Hui Jun Li, Tian Ci Jing, Mengchun Gong, Wei Hu, Zhen Wei Zhang, Xingang Zhang, and Guang Wei Zhang. 2024. “Pitfalls in Developing Machine Learning Models for Predicting Cardiovascular Diseases: Challenge and Solutions.” Journal of Medical Internet Research 26.
4. Chataut, Robin, Alex Phoummalayvane, and Robert Akl. 2023. “Unleashing the Power of IoT: A Comprehensive Review of IoT Applications and Future Prospects in Healthcare, Agriculture, Smart Homes, Smart Cities, and Industry 4.0.” Sensors 23(16).
5. Chen, Xiaoxue, Lianwen Jin, Yuanzhi Zhu, Canjie Luo, and Tianwei Wang. 2021. “Text Recognition in the Wild: A Survey.” ACM Computing Surveys 54(2).
6. Coito, Tiago, Bernardo Firme, Miguel S. E. Martins, Susana M. Vieira, João Figueiredo, and João M. C. Sousa. 2021. “Intelligent Sensors for Real-Time Decision-Making.” Automation 2(2):62–82.
7. Črepinšek, Matej, Shih Hsi Liu, and Marjan Mernik. 2014. “Replication and Comparison of Computational Experiments in Applied Evolutionary Computing: Common Pitfalls and Guidelines to Avoid Them.” Applied Soft Computing Journal 19:161–70.
8. Doermann, David, Jian Liang, and Huiping Li. 2003. “Progress in Camera-Based Document Image Analysis University of Maryland College Park.” (Icdar).
9. Dorafshan, Sattar, Robert J. Thomas, and Marc Maguire. 2018. “Comparison of Deep Convolutional Neural Networks and Edge Detectors for Image-Based Crack Detection in Concrete.” Construction and Building Materials 186:1031–45.
10. Fan, Cheng, Yongjun Sun, Yang Zhao, Mengjie Song, and Jiayuan Wang. 2019. “Deep Learning-Based Feature Engineering Methods for Improved Building Energy Prediction.” Applied Energy 240(September 2018):35–45.

11. Femi Osasona, Olukunle Oladipupo Amoo, Akoh Atadoga, Temitayo Oluwaseun Abrahams, Oluwatoyin Ajoke Farayola, and Benjamin Samson Ayinla. 2024. "Reviewing the Ethical Implications of Ai in Decision Making Processes." International Journal of Management & Entrepreneurship Research 6(2):322–35.
12. Foster, Ian, Markus Fidler, Alain Roy, Volker Sander, and Linda Winkler. 2004. "End-to-End Quality of Service for High-End Applications." Computer Communications 27(14):1375–88.
13. Glasmachers, Tobias. 2017. "Limits of End-to-End Learning." Journal of Machine Learning Research 77:17–32.
14. Gouiza, Nissrine, Hakim Jebari, and Kamal Reklaoui. 2024. "Integration of Iot-Enabled Technologies and Artificial Intelligence in Diverse Domains: Recent Advancements and Future Trends." Journal of Theoretical and Applied Information Technology 102(5):1975–2029.
15. Hamad, Karez and Mehmet Kaya. 2016. "A Detailed Analysis of Optical Character Recognition Technology." International Journal of Applied Mathematics, Electronics and Computers 4(Special Issue-1):244–244.
16. Jain, Richa and Prof Deepa Gianchandani. 2019. "A Hybrid Approach for Detection and Recognition of Traffic Text Sign Using MSER and OCR." Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018 775–78.
17. Li, Yi, Yefeng Zheng, David Doermann, and Stefan Jaeger. 2008. "Script-Independent Text Line Segmentation in Freestyle Handwritten Documents." IEEE Transactions on Pattern Analysis and Machine Intelligence 30(8):1313–29.
18. Liu, Xiyan, Gaofeng Meng, and Chunhong Pan. 2019. "Scene Text Detection and Recognition with Advances in Deep Learning: A Survey." International Journal on Document Analysis and Recognition 22(2):143–62.
19. Liu, Yu, Qinghua Guo, and Maggi Kelly. 2008. "A Framework of Region-Based Spatial Relations for Non-Overlapping Features and Its Application in Object Based Image Analysis." ISPRS Journal of Photogrammetry and Remote Sensing 63(4):461–75.
20. Mahmood, Zahid, Khurram Khan, Uzair Khan, Syed Hasan Adil, Syed Saad Azhar Ali, and Mohsin Shahzad. 2022. "Towards Automatic License Plate Detection." Sensors 22(3):1–19.
21. Mandal, Murari and Santosh Kumar Vipparthi. 2022. "An Empirical Review of Deep Learning Frameworks for Change Detection: Model Design, Experimental Frameworks, Challenges and Research Needs." IEEE Transactions on Intelligent Transportation Systems 23(7):6101–22.
22. Mattys, Sven L., Laurence White, and James F. Melhorn. 2005. "Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework." Journal of Experimental Psychology: General 134(4):477–500.
23. Meng, Fanfei and Branden Ghena. 2023. "Research on Text Recognition Methods Based on Artificial Intelligence and Machine Learning." Advances in Computer and Communication 4(5):340–44.
24. Merino-Gracia, Carlos, Karel Lenc, and Majid Mirmehdi. 2012. "A Head-Mounted Device for Recognizing Text in Natural Scenes." Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7139 LNCS:29–41.
25. Najafabadi, Maryam M., Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. 2015. "Deep Learning Applications and Challenges in Big Data Analytics." Journal of Big Data 2(1):1–21.

26. Nassar, Ahmed and Mostafa Kamal. 2021. “International Journal of Responsible Artificial Intelligence Ethical Dilemmas in AI-Powered Decision-Making: A Deep Dive into Big Data-Driven Ethical Considerations.” International Journal of Responsible Artificial Intelligence 1–11.
27. Noor, Mohd Halim Mohd and Ayokunle Olalekan Ige. 2024. “A Survey on Deep Learning and State-of-the-Art Applications.” 1–56.
28. Oyekunle, David, David Boohene, and David Preston. 2024. “Ethical Considerations in AI-Powered Work Environments: A Literature Review and Theoretical Framework for Ensuring Human Dignity and Fairness.” International Journal of Scientific Research and Management (IJSRM) 12(03):6166–78.
29. Paramesha, Mallikarjuna, Nitin Rane, and Jayesh Rane. 2024. “Big Data Analytics, Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence.” SSRN Electronic Journal (July):110–33.
30. Parvez, Mohammad Tanvir and Sabri A. Mahmoud. 2013. “Offline Arabic Handwritten Text Recognition: A Survey.” ACM Computing Surveys 45(2).
31. Ramanathan, Vignesh, Kevin Tang, Greg Mori, and Li Fei-Fei. 2015. “Learning Temporal Embeddings for Complex Video Analysis.” Proceedings of the IEEE International Conference on Computer Vision 2015 Inter:4471–79.
32. Rane, Nitin, Mallikarjuna Paramesha, Saurabh Choudhary, and Jayesh Rane. 2024. “Artificial Intelligence, Machine Learning, and Deep Learning for Advanced Business Strategies: A Review.” SSRN Electronic Journal (June):10–11.
33. Roberts, Michael, Derek Driggs, Matthew Thorpe, Julian Gilbey, Michael Yeung, Stephan Ursprung, Angelica I. Aviles-Rivero, Christian Etmann, Cathal McCague, Lucian Beer, Jonathan R. Weir-McCall, Zhongzhao Teng, Effrossyni Gkrania-Klotsas, Alessandro Ruggiero, Anna Korhonen, Emily Jefferson, Emmanuel Ako, Georg Langs, Ghassem Gozaliasl, Guang Yang, Helmut Prosch, Jacobus Preller, Jan Stanczuk, Jing Tang, Johannes Hofmanninger, Judith Babar, Lorena Escudero Sánchez, Muhunthan Thillai, Paula Martin Gonzalez, Philip Teare, Xiaoxiang Zhu, Mishal Patel, Conor Cafolla, Hojjat Azadbakht, Joseph Jacob, Josh Lowe, Kang Zhang, Kyle Bradley, Marcel Wassin, Markus Holzer, Kangyu Ji, Maria Delgado Ortet, Tao Ai, Nicholas Walton, Pietro Lio, Samuel Stranks, Tolou Shadbahr, Weizhe Lin, Yunfei Zha, Zhangming Niu, James H. F. Rudd, Evis Sala, and Carola Bibiane Schönlieb. 2021. “Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans.” Nature Machine Intelligence 3(3):199–217.
34. Rouast, Philipp V., Marc T. P. Adam, and Raymond Chiong. 2021. “Deep Learning for Human Affect Recognition: Insights and New Developments.” IEEE Transactions on Affective Computing 12(2):524–43.
35. Sakhi, Omid. 2013. “Segmentation of Heterogeneous Document Images: An Approach Based on Machine Learning, Connected Components, and Texture Analysis.”
36. Space, Architectural. n.d. “Performative Materiality.” 1:611–18.
37. Tang, Wei, Pei Yu, Jiahuan Zhou, and Ying Wu. 2017. “Towards a Unified Compositional Model for Visual Pattern Modeling.” Proceedings of the IEEE International Conference on Computer Vision 2017-Octob:2803–12.
38. Vu, Hoai Nam, Tuan Anh Tran, Na In Seop, and Soo Hyung Kim. 2016. “Extraction of Text Regions from Complex Background in Document Images by Multilevel Clustering.”

International Journal of Networked and Distributed Computing 4(1):11–21.
39. Yao, Cong, Xiang Bai, Wenyu Liu, Yi Ma, and Zhuowen Tu. 2012. “Detecting Texts of Arbitrary Orientations in Natural Images.” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 8:1083–90.
40. Yin, Xu Cheng, Wei Yi Pei, Jun Zhang, and Hong Wei Hao. 2015. “Multi-Orientation Scene Text Detection with Adaptive Clustering.” IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9):1930–37.
41. Zherzdev, Sergey and Alexey Gruzdev. 2018. “LPRNet: License Plate Recognition via Deep Neural Networks.” 1–6.
42. Zhou, Zijian, Oluwatosin Alabi, Meng Wei, Tom Vercauteren, and Miaojing Shi. 2023. “Text Promptable Surgical Instrument Segmentation with Vision-Language Models.” Advances in Neural Information Processing Systems 36(NeurIPS):1–13.

Name: Wydyanto

ORCID: 0000-0001-5333-2197

Affiliation: Universitas Bina Darma

Address: General Ahmad Yani Street No. 3 Palembang, South Sumatra, Indonesia 0711-515582

Brief Biographical History:

Dr. Wydyanto was born in Kedurang on December 24, 1969. He holds a Doctor of Philosophy degree and has extensive experience in the field of Computer Engineering, particularly in teaching Cloud Computing. In addition to his role as a lecturer, Dr. Wydyanto is actively involved in various research projects, including digital image processing, text detection in images and videos, and text-to-speech conversion using Raspberry Pi.

Main Works: A Hybrid Approach to Detect and Identify Text in Picture: Publish in Q1

Membership in Academic Societies: Informatics Engineering Lecturer