

Core methodological classes of text extraction and localization-a snapshot of approaches

Dayananda Kodala Jayaram, Puttegowda Devegowda

Department of Computer Science and Engineering, ATME College of Engineering, Visvesvaraya Technological University,
Mysuru, India

Article Info

Article history:

Received Sep 6, 2024
Revised Mar 18, 2025
Accepted Mar 26, 2025

Keywords:

Data analytics
Methodologies
Text detection
Text extraction
Text localization

ABSTRACT

The motivation to work on text extraction and localization is quite a substantial that potentially influences a larger area of application right from business intelligence to advanced data analytics. At present, there are massive archives of literatures addressing varying ranges of problems associated with text extraction and localization with an effective realization of respective contribution as well as on-going issues. However, problem statement is that all these massive implementation studies are further required to converge down in order to realize the core classes of methodologies involved in text extraction. Hence, this manuscript uses desk research methodology to address this issue by presenting a compact insight of core methodological classes where all the recent implementation work are converged down to understand its strength and weakness. The research outcome contributes towards facilitating information of current research trend and identified research gap. The proposed review study assists in undertaking decision of suitable approach of text extraction, localization, detection, recognition, and classification.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Dayananda Kodala Jayaram
Department of Computer Science and Engineering, ATME College of Engineering
Visvesvaraya Technological University
Bannur Rd, Mysuru, Karnataka, 570028, India
Email: dayananda.kem@gmail.com

1. INTRODUCTION

The usage of text extraction and localization are widely noticed in the application area of translation services, content management, and software development. In text extraction, the particular pieces of information are identified followed by retrieval from large corpus of text that can be subjected to both unstructured and structured data using optical character recognition (OCR), natural language processing (NLP), and regular expression [1]. The process of text extraction is widely adopted in search engines, web scraping, document parsing, and data mining. In text localization, the position of the text is identified in order to meet the contextual demands and linguistic requirements where the text contents are translated and customized to meet the demands of local users. The process of text localization is carried out using machine translation, translation memory, and localization management system [2]. There are various challenges associated with text extraction and location where the primary one is related to sub-optimal text quality and varied formats. Variability of structure also poses a significant challenge in extraction of suitable text where it is quite challenging to understand the context. Further, more complexities are added during extraction of text with multiple languages as well as domain-specific terminologies. While performing extraction from scene text or from complex background, there are also higher possibilities of missed text elements or inappropriately identified text eventually leading to errors. There is an immense future scope of text

extraction that are more likely to be formed by the pace of evolving technologies. Text detection when combined with artificial intelligence (AI) can offer more capability towards categorizing complex classes of information [3]. Another potential future scope of text extraction is related to proliferated demands of real-time data analysis from various consistent updating sources, feeds of social media, and live streams. essential for analyzing and monitoring trends and events of real-time [4]. Content generation and knowledge graphs are another evolving future scope of text extraction where various forms of reports and summaries can be automatically be generated, thereby, saving manual effort, time, enhances productivity. Enhanced handling of dialects and varied languages are another essential scope of text extraction to offer efficient multilingual capabilities and language-particular frameworks. A distinct form of linguistic feature and cultural context can be effectively handled by language-particular frameworks while multilingual capabilities can assist in handling multiple languages. There is also an immense need of text extraction tool specifically for various domains of services e.g., financial reports, medical records, and legal documents. In such aspect, the text extraction system can furnish more customized information on the basis of utility, enhanced relevance, behavior, and user preference. Finally, the more frequent usage of text extraction is witnessed from various scenes where multi-modalities are adopted. Such perspective of text extraction calls for integrating text extraction schemes with varying datatypes e.g., video, audio, and images. will offer extensive comprehensive insights of knowledge from the acquired text. Further, the textual contents can be correlated to AI with various forms of data in order to provide more extraction of data along with holistic understanding. Although, the conventional meaning of text localization is more focused towards translation-based processes, but the importance of identifying and converging the textual area from the scene text can be highly complementary towards text extraction process. However, the core challenges that almost all the research communities are encountering currently is to introduce scalable performance of text extraction. This is really a bigger issue as acquiring textual contents from massive volumes of data demands a precise tool with higher robustness as well as efficient infrastructures. The biggest challenge still resides for providing accurate and faster extraction of text from larger data volumes in current era.

There are various existing review works towards text extraction process which are required to be discussed in study background to have more fair idea of existing contribution. A unique work is designed by Ibrahim [5] which can perform detection of plagiarism text using AI. Cao *et al.* [6] have presented vivid discussion of text detection approaches from natural scene along with comprehensive discussion of protocols used for evaluation and dataset trends. Discussion of different variants of algorithms towards text classification have been presented by Gasparetto *et al.* [7] with a special emphasis on deep learning-based schemes. Adoption of deep learning approaches in research work towards similar direction of extraction of information is presented by Yang *et al.* [8] considering extraction of multi-modal information, event-related information, and relationship among entities based information. The paper has also presented discussion of multilevel dataset suitable for such algorithms. The work carried out by Alkendi *et al.* [9] have presented review of multiple handwritten text identification approaches in context of both research-based works as well as certain commercial system. A review on unique topic associated with matching of textual entities has been presented by Jiang and Cai [10] where the authors highlights ongoing challenges associated with newly evolving approaches of AI. Malashin *et al.* [11] have presented discussion of an adaptive model where search-based optimization is used for text extraction.

The identified problems arrived after reviewing the above-mentioned studies are as follows: i) there are highly scattered form of the review work at present which offers descriptive information about all individual approaches; however, fails to state a proper taxonomies of the frequently evolved approaches; ii) there are few research work which reports of limitations of core classes of methodologies involved in text extraction and localization; iii) there are various research papers with similar core technologies but using different research methodologies that has not been yet identified in the area of text extraction; and iv) simplification of research gap and current trends of methodological classes are few to be identified in existing review work.

Hence, the aim of proposed study is to carry out a insight description of the core methodological classes for text extraction and localization approaches from recent studies. The value-added information stated in this manuscript as a part of contribution are as follows: i) to highlights the core classes of frequently adopted methods towards text extraction and localization that has not yet been reported in prior studies; ii) to highlight the strength and weakness of associated methodological classes of text extraction in order to infer its degree of effectiveness; iii) to highlight the current research trends for identifying the frequently adopted methods as well as their individual methods along with identification of evolving approaches; and iv) to highlight the research gap in form of unsolved problems that are yet awaiting for an effective solution. The next section presents discussion of research methods adopted to construct this review study.

2. METHODS

The proposed method targets to construct a proper taxonomy of methodologies by reviewing the existing approaches of text extraction and localization. Figure 1 showcases the research method adopted for this purpose. The first step is towards performing database exploration associated with text extraction, text detection, localization, and classification. The initial screening is performed by reviewing the abstract and title followed by eliminating the duplicates. The study considers two different papers with exactly similar methods to be duplicates apart from two similar manuscripts. The inclusion criteria are: i) only journal papers published between 2019-2024 are included, and ii) journals from reputed publishes e.g., MDPI, Springer, ArXiv, IJCE, EURASIP are used. The exclusion criteria are: any theoretical or discussion or review papers and conference or proceedings papers.

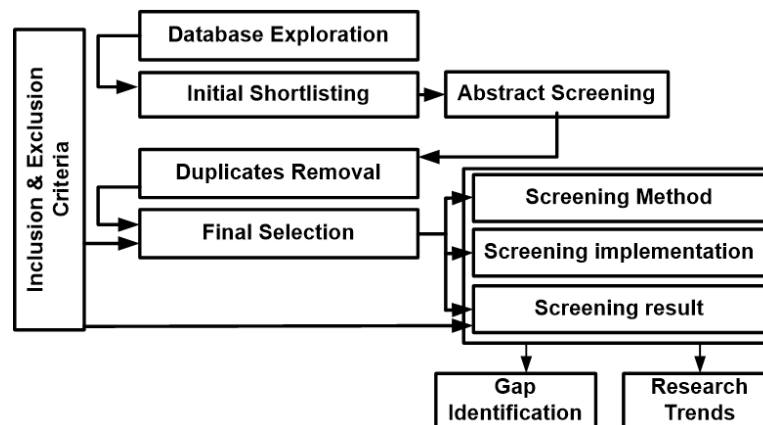


Figure 1. Adopted methodology

After duplicates removal, the final screening is carried out by screening the complete research method reported within the manuscript as well as implementation and result sections too. Further, a research trend analysis is performed to understand the frequencies of publications associated with each individual approaches under core methodological classes of text extraction and localization. The final step is to review the problems that have not reportedly being mitigates over a consecutive period of time in order to acquire research gap. The next section discusses about the outcome of proposed review work.

3. RESULTS AND DISCUSSION

This section presents various types of text detection as well as localization schemes from the video identified from the existing literatures. It has been noted that usage of such scheme has been witnessed towards developing and investigating automatic subtitles, content analysis, video indexing, and many more applications. Various distinct methodologies have been adopted towards detection of textual contents from multimedia addressing different variants of research problem. The highlights of the identified methods of text detection and localization are categorized and discussed as follows:

3.1. Reviewed approaches

3.1.1. Conventional computer vision methods

Table 1 highlights the summary of effectiveness of conventional computer vision methods. There are three essential approaches towards text detection and localization using computer vision methods viz. edge detection, connected component analysis, and region-based methods. Edge detection assists in identifying the structure and boundaries of text regions using multiple approaches viz. noise reduction, gradient calculation non-maximum suppression, double thresholding, and edge tracking by hysteresis. Conventional methods e.g., canny edge, Sobel operator, Prewitt operator, Laplacian of Gaussian, and Hough transform are used for edge detection. Connected component analysis is capable of determining and analyzing regions with connected pixels either in grayscale or binary and can significantly assists in extracting and isolating text area. Region-based method performs segmentation of an image to more logical region followed by text extraction and applying OCR. It also performs merging regions and bounding box refinement as post processing prior to text extraction.

Table 1. Summary of effectiveness of computer vision methods

Approaches	Papers	Advantage	Limitations
Edge detection	[12]	Ability to detect edges with high accuracy under different text orientation	Computationally more intensive
Connected component analysis	[13]	Highly flexible and effective for simple text.	Involves computational complexities, sensitive to noise
Region-based methods	[14], [15]	Better scalability and can handle diverse fonts and layouts of text	Highly sensitive to noise, involves complexity due to multiple steps

3.1.2. Learning-based approaches

In contrast to conventional approaches of text extraction and localization, machine learning and deep approaches are slowly evolving to prove its capability towards improving accuracy and adaptability. Table 2 summarizes the effectiveness of reviewed learning-based approaches. These learning approaches involves using a standard dataset followed by performing training operation to extract and recognize text region as well as textual characters. Following is more information about approaches:

- From perspective machine learning approach, sliding window and feature-based methods are frequently adopted towards text detection. The sliding window assists in detecting and isolating the text by moving a window of fixed size over an image for determining the region of interest. Prior to text extraction, it also performs filtering and bounding box refinement. This approaches also calls for applying various classifiers e.g., support vector machine (SVM) or random forest. Feature-based methods are another frequently adopted approaches in machine learning where detection of text is carried out using varied visual characteristics e.g., color and intensity, texture, shape, and morphology.
- From the perspective of deep learning approach, there are various frequently used approaches based on potential neural network architecture. This approach offers a capability to learning sophisticated features and complex patterns. The preprocessing operation in deep learning offers an enhanced model performance using noise reduction, normalization, and data augmentation. Deep learning methods also involves efficient pre-trained models that are highly specific to domain in order to enhance the robustness and accuracy performance. PyTorch and TensorFlow are more frequently adopted framework reported in existing literatures. Convolutional neural network (CNN) is one of the prominent deep learning methods known for its superior capability of acquiring hierarchical features for assisting in efficient text detection as well as recognition. CNN is mainly used for character recognition and text-region detection. Region-based CNN (R-CNN) performs classification of identified region of interest using pre-trained models mainly. You only look once (YOLO) can perform real-time detection of text by predicting class labels and bounding box using neural network. Single shot multibox detector (SSD) using similar approach of YOLO method towards text extraction.

Table 2. Summary of effectiveness of learning methods

Approaches	Papers	Advantage	Limitations
Sliding Window	[16]	Adaptable to varied form of text, simplicity	Fixed window size, overlapping issue, computationally intensive
Feature-based methods	[17], [18]	Interpretable, adaptable	Needs manual tuning of threshold and parameters, sensitive to noise and fluctuating text styles
CNN	[19]-[21]	Highest accuracy, autonomous feature learning	Demands massive trained and annotated data, highly computationally intensive
R-CNN	[22], [23]	Flexible, and highly accurate detection	Computationally intensive, induces higher complexities towards implementation
YOLO	[24], [25]	Performs end-to-end detection, real-time performance	Complex text localization, higher training complexity
Single shot multibox detector	[26], [27]	Can extract text at multiple scales, real-time extraction	Demands maximum fine tuning, demands higher number of annotated data, higher resource consumption
RNN and LSTM	[28], [29]	Effective in understanding context, capable of handling varying text length	RNN suffers from vanishing gradient problem
Transformer	[30], [31]	State-of-the Art performance	Demands large trained data, demands substantial memory and resource

Apart from above-mentioned learning approaches, attention mechanism is also reportedly used for managing sequences of input with varying length. It is noted that when transformer or RNN is integrated with attention layer, it improvises the performance of text recognition. However, such models is witnessed with

loopholes of higher computational resource dependencies for inference and training with increased complexities although, these models are reportedly characterized with better context handling and enhanced focus on features. There are also end-to-end models that combines both extraction and identification in one model used specifically for recognition of scene text. Such models can be developed by integrating convolution RNN and text bounding boxes where text are obtained from images. These approaches offer reduced computational effort towards extraction and recognition with higher streamlined workflow; however, they also suffer from extensive training data demands and model complexity. Object detection methods are also reportedly adopted in deep learning approached which is meant for localizing and classifying objects within images. Various frameworks like YOLO, SSD, and faster R-CNN are used for this purpose. These models are known for offering real-time detection using YOLO along with detection capability of multiple objects at same time; however, they suffer from overlapping region of text problem and demands higher number of labelled data for training.

3.1.3. Text recognition and localization approaches

Table 3 showcases essential properties of this method from existing reviewed studies. This method is known for its robust extraction process of text as well as interpreting them obtained from image source. This model performs joint operation of extraction of text, localization of textual content, and recognition of text. Various approaches used under this method are efficient and accurate scene text detector (EAST), character region awareness for text detection (CRAFT), TextBoxes++, tesseract OCR, convolution RNN, and attention-based sequence-to-sequence text recognition (ASTER).

Table 3. Summary of effectiveness of text recognition and localization model

Approaches	Papers	Advantage	Limitations
EAST	[32]	Robust, better accuracy, speed and accuracy	Do not perform text recognition, post processing is quite complex
CRAFT	[33]	Versatility, precise character-level detection	-do-
TextBoxes++	[34]	Enhanced localization, multi-scale detection	Implementation complexity, training complexity
Tesseract OCR	[35]	Simpler usage, supports multiple language, open source	Limited to text recognition, inconsistent performance
CRNN	[36]	Contextual information extraction, can handle varying text shape and length	Complex training and implementation, demands extensive annotated training data
ASTER	[37]	Attention methods increase accuracy, can handle varying text shape and length	Demands apriori knowledge for tuning, resource intensive

3.1.4. End-to-end approaches

Table 4 showcases properties studied after reviewing existing implementation schemes towards these end-to-end approaches. This approach consists of mainly two core method TextNet and connectionist text proposal network (CTPN) which is meant for perform extraction, detection, and recognition of the text. TextNet emphasizes towards both identification of text followed by recognition of text with lesser training operations involved in it. CTPN is responsible for yielding text proposals and they are highly ideal for complex backgrounds with textual contents. TextNet, using CNN approach, is specially used for detection of real-time text and document analysis while CTPN, using RNN, is used for digitization of document with complex form of layouts and detection of scene text.

Table 4. Summary of effectiveness of end-to-end approaches

Approaches	Papers	Advantage	Limitations
TextNet	[38]	Adaptable, higher accuracy, unified approach	Demands higher computational resources, demands extensived labelled data
CTPN	[39]	End-to-end solution, effective for multiple text lines	Tedius post processing is required to eliminate non-text region and extract text, computationally intensive for complex scenes

3.1.5. Temporal consistency approaches

Table 5 highlights the characteristic of temporal consistency approaches noted from existing studies. This approach is specifically meant for extracting text from videos as well as from image sequences targeting to retain efficient localization and recognition of text. This approach is mainly used for detecting live text from video streams and scanning document with consistent input. Typical methods used for this purpose are moving average filtering, Kalman filtering, frame-by-frame matching, simple online and realtime tracking (SORT), text-reidentification, and multi-frame fusion. The practical consideration towards this approach is

framerate, text movement, and computational resources. Out of various variants of this approaches, they have been categorized mainly based on tracking-based method and optical flow methods for simplification in taxonomies of literatures. The tracking-based method mainly uses SORT methodology for emphasizing on movement of extracted region of text while optical flow method is meant for evaluating motion vector of textual object between sequences of frames. Farneback algorithm and Lucas-Kanade algorithm are frequently adopted approaches for optical flow-based text extraction.

Table 5. Summary of effectiveness of temporal consistency approaches

Approaches	Papers	Advantage	Limitations
Tracking-based method	[40]	Can handle occlusion, consistency across frames, enhanced accuracy	Gradual deviation of tracked position, demands on previous detection
Optical Flow	[41]	Supports motion estimation of text, extracts low-level information, doesn't demand previous detection, simplified deployment	Highly sensitive to noise, limited to smooth motion, high resolution videos give increased computational complexity

3.2. Essential findings of study

From the prior section, it has been noted that there are 5 different classes of methods used for text detection and localization. It has been also noted that each class of methods have reported of its beneficial features and limitations too. However, it is required to acquire an overall picture of current trends of existing methodologies that can offer a potential insight towards frequently adopted methods. It is also necessary to converge to a specific point of on-going research issue extracted after review of existing classes of text detection and localization methods. This section offers highlight of essential findings of study with respect to research trend visualization and identification of research gap.

3.2.1. Research trend

It has been noted that there are approximately 162,301 publications towards the above-mentioned five discrete classes of methodologies of text extraction and localization. Only the papers published between 2019-2024 available in IEEE Xplore, Springer, and MDPI has been observed and recorded. Various other journals also have been referred to witnessed nearly the similar trends of publications. Figures 2 to 6 showcases the graphical outcomes of these classes of methodologies. It is noted that maximum publication is noted for feature-based methods ($n=63,862$) followed by CNN approach ($n=20,567$), and CTPN method ($n=20,571$). The lowest count is observed for Sliding window methods ($n=1$) while other methods are quite scattered from cardinality of publication viewpoint. There are no publications towards connected component analysis, TextNet, and ASTER methods with respect to implementation papers.

Some of the essential findings from the research trend exhibits witnessed from Figures 2 to 6 are as follows: i) adoption of deep learning approaches [CNN ($n=20567$), R-CNN ($n=16544$), YOLO ($n=2513$), SSD ($n=396$), RNN ($n=6023$), LSTM ($n=9974$), and transformer ($n=7173$)] has been on consistent rise, ii) Although, there are less number of recorded works for temporal consistency-based approaches, but they are next on rise of adoption after deep learning-based approaches, iii) The overall publications trends towards text recognition and localization methods are quite less in contrast to other methods even bearing some potential advantageous features towards text extraction.

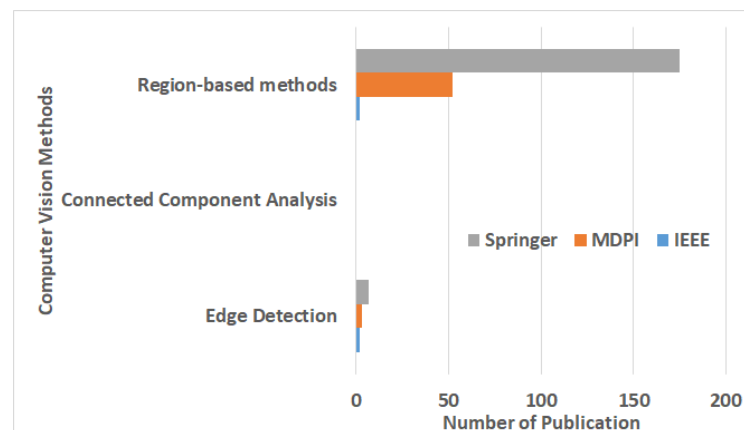


Figure 2. Trends of computer vision methods

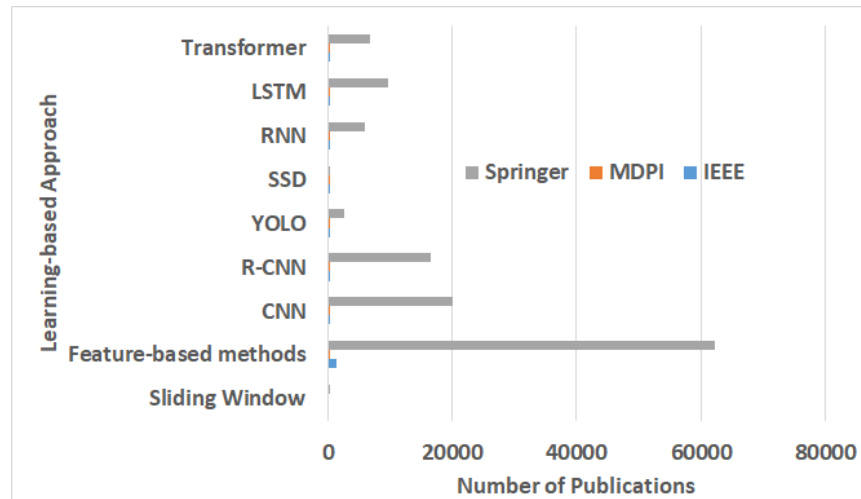


Figure 3. Trends of learning-based methods

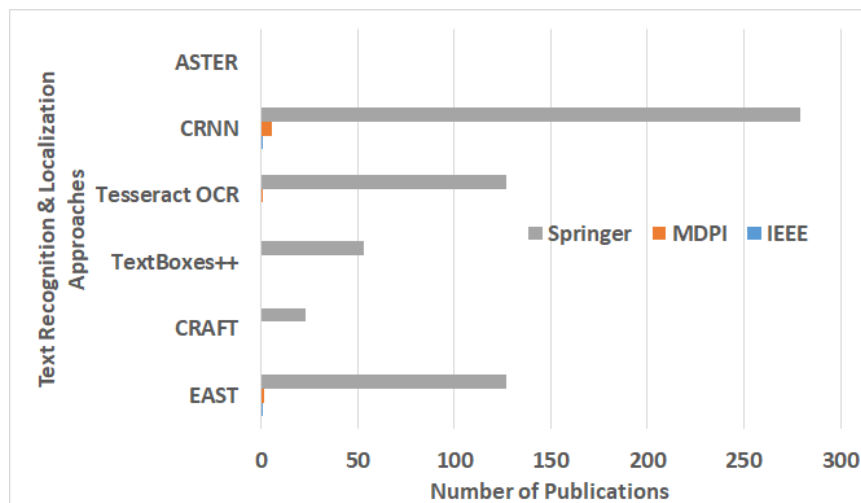


Figure 4. Trends of text recognition and localization methods

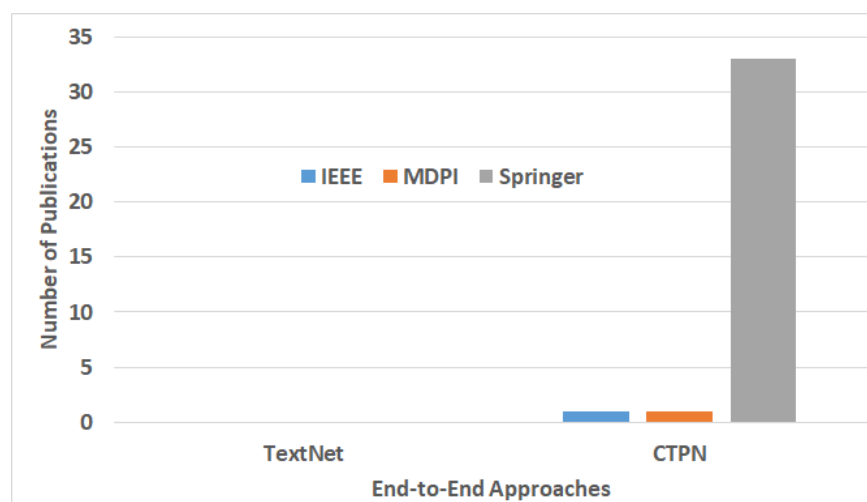


Figure 5. Trends of end-to-end methods

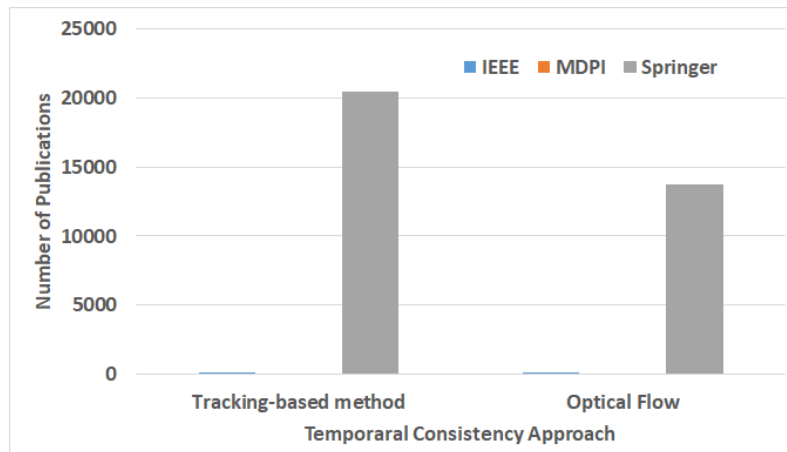


Figure 6. Trends of temporal consistency methods

3.2.2. Research gap

The identified research gap obtained after reviewing the existing classes of methodologies towards text extraction and localization are as follows: i) existing research models encounters significant challenges while extracting textual contents from documents characterized by complex background and layouts. ii) There are lesser reporting of study model considering complex form of document (in form of distortion or low-resolution) for text extraction leading to sub-optimal accuracy performance. iii) There are few algorithm which claims of faster and lightweight operation considering practical environment either towards localization or towards extraction of text. iv) Existing system has not witnessed with any solution where text extraction process is integrated with semantics for better localization accuracy. v) Although there are algorithm reported to overcome text extraction issues from natural scenes, but they are associated with increased computational burden. vi) Existing system have less number of benchmarked models considering varying test environment to prove its applicability on near real world applications.

3.3. Discussion

The outcomes of this study provide critical insights into the various approaches used for text identification and localization across multimedia sources. The data shows a clear trend of rising usage of deep learning-based systems, such as CNNs and region-based CNNs (R-CNNs), due to their improved text recognition accuracy. A crucial piece of supporting evidence is the observed increase in publications about CNN-based models, with over 20,000 studies published between 2019 and 2024. These deep learning models excel at handling complicated text recognition problems, overcoming many of the limits faced by traditional methods like edge detection and connected component analysis, which are computationally more costly and less resistant in noisy situations.

When compared to past studies, the findings of this study are consistent with the patterns highlighted in other works that emphasize the dominance of machine learning and deep learning models in text extraction tasks. However, it also notes some gaps in the current literature, particularly in methods such as temporal consistency approaches, which have received little research attention despite their expanding importance in video and sequence-based text detection. The study's merits include its extensive analysis of various approaches and detailed categorization of procedures. However, one restriction is the little emphasis on the integration of semantic understanding into text localization, which might potentially improve accuracy even more. Furthermore, while the study highlights significant advancements in text detection, it also uncovers the surprising finding that the number of publications in text recognition and localization is lower compared to other methods, suggesting a potential underrepresentation of this area in ongoing research.

In conclusion, the purpose of this study was to provide an in-depth assessment of current trends in text identification and localization technologies, as well as important insights into their effectiveness and limitations. The study's significance stems from its capacity to combine diverse methodologies and identify topics for further research, particularly in resolving the constraints of complex backgrounds, low-resolution text, and real-time application contexts. Unanswered questions include how to further minimize processing overhead while boosting accuracy, as well as how to include contextual and semantic understanding into text identification systems. Future study should look into lightweight models for practical applications, more robust methods for a variety of real-world settings, and the use of multimodal data to improve text localization across several media types.

4. CONCLUSION

This study looked at the various approaches for text detection and localization in multimedia, emphasizing their importance in applications like automatic subtitles, video indexing, and content analysis. The study stresses the growing importance of deep learning-based methods such as CNNs and R-CNNs, which provide more accuracy and adaptability in a variety of text recognition tasks. While some may argue that classical methods remain relevant because of their lower computational requirements, deep learning methodologies' improved performance and scalability make a compelling case for wider adoption. The paper contributes towards the following novelties viz. The manuscript presents 5 categories of methodological classes of text extraction (computer vision methods, learning-based approaches, text recognition and localization approaches, end-to-end approaches, temporal consistency approaches) that has not been reported before in any prior review works; thereby offering a compact snapshot of methods. The review has studied 22 individual methodologies belonging to each of the above-mentioned five methodological classes in order to understand its potential strength and weakness. The review work contributes towards a simplified and novel trend analysis of existing system to find increasing number of sophisticated approaches (e.g., learning based approaches and temporal consistency approaches) while less innovation has been yet witnessed for more simplified schemes towards text extraction. The paper finally contributes towards explicitly identified updated research gap based on last 5 years study models. The future work will be oriented towards presenting a simplified computational model towards addressing the identified research gap for evolving more cost-effective text extraction and localization approaches.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dayanand Kodala	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Jayaram														
Puttegowda		✓				✓		✓	✓	✓	✓	✓		
Devegowda														

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY




Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES




- [1] B. Hashemzadeh and M. Abdolrazzagah-Nezhad, "Improving keyword extraction in multilingual texts," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 5909–5916, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5909-5916.
- [2] R. M. Jayanth and M. Kapanaiiah, "Dominating set based arbitrary oriented bilingual scene text localization," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3730–3738, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3730-3738.
- [3] S. V. Mahadevkar, S. Patil, K. Kotecha, L. W. Soong, and T. Choudhury, "Exploring AI-driven approaches for unstructured document analysis and future horizons," *Journal of Big Data*, vol. 11, no. 1, p. 92, Jul. 2024, doi: 10.1186/s40537-024-00948-z.
- [4] G. Liao, Z. Zhu, Y. Bai, T. Liu, and Z. Xie, "PSENet-based efficient scene text detection," *Eurasip Journal on Advances in Signal Processing*, vol. 2021, no. 1, p. 97, Dec. 2021, doi: 10.1186/s13634-021-00808-5.
- [5] K. Ibrahim, "Using AI-based detectors to control AI-assisted plagiarism in ESL writing: 'The Terminator Versus the Machines,'" *Language Testing in Asia*, vol. 13, no. 1, p. 46, Oct. 2023, doi: 10.1186/s40468-023-00260-2.
- [6] D. Cao, Y. Zhong, L. Wang, Y. He, and J. Dang, "Scene text detection in natural images: a review," *Symmetry*, vol. 12, no. 12, pp. 1–26, Nov. 2020, doi: 10.3390/sym12121956.
- [7] A. Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, "Survey on text classification algorithms: from text to predictions," *Information (Switzerland)*, vol. 13, no. 2, p. 83, Feb. 2022, doi: 10.3390/info13020083.
- [8] Y. Yang, Z. Wu, Y. Yang, S. Lian, F. Guo, and Z. Wang, "A survey of information extraction based on deep learning," *Applied Sciences (Switzerland)*, vol. 12, no. 19, 2022, doi: 10.3390/app12199691.
- [9] W. Alkendi, F. Gechter, L. Heyberger, and C. Guyeux, "Advancements and challenges in handwritten text recognition: a comprehensive survey," *Journal of Imaging*, vol. 10, no. 1, p. 18, Jan. 2024, doi: 10.3390/jimaging10010018.

- [10] P. Jiang and X. Cai, "A survey of text-matching techniques," *Information (Switzerland)*, vol. 15, no. 6, p. 332, Jun. 2024, doi: 10.3390/info15060332.
- [11] I. Malashin, I. Masich, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Image text extraction and natural language processing of unstructured data from medical reports," *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 1361–1377, Jun. 2024, doi: 10.3390/make6020064.
- [12] N. A. Rehman and F. Haroon, "Adaptive gaussian and double thresholding for contour detection and character recognition of two-dimensional area using computer vision †," *Engineering Proceedings*, vol. 32, no. 1, 2023, doi: 10.3390/engproc2023032023.
- [13] M. Umair *et al.*, "A multi-layer holistic approach for cursive text recognition," *Applied Sciences (Switzerland)*, vol. 12, no. 24, p. 12652, Dec. 2022, doi: 10.3390/app122412652.
- [14] A. Mirza, O. Zeshan, M. Atif, and I. Siddiqi, "Detection and recognition of cursive text from video frames," *Eurasip Journal on Image and Video Processing*, vol. 2020, no. 1, p. 34, Dec. 2020, doi: 10.1186/s13640-020-00523-5.
- [15] J. Diaz-Escobar and V. Kober, "Natural scene text detection and segmentation using phase-based regions and character retrieval," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–17, Jun. 2020, doi: 10.1155/2020/7067251.
- [16] A. Drobny, B. K. Barakat, R. Alaasam, B. Madi, I. Rabaev, and J. El-Sana, "Text line extraction in historical documents using mask R-CNN," *Signals*, vol. 3, no. 3, pp. 535–549, Aug. 2022, doi: 10.3390/signals3030032.
- [17] M. Ibrayim, Y. Li, and A. Hamdulla, "Scene text detection based on two-branch feature extraction," *Sensors*, vol. 22, no. 16, p. 6262, Aug. 2022, doi: 10.3390/s22166262.
- [18] T. C. Phan, A. C. Phan, H. P. Cao, and T. N. Trieu, "Content-based video big data retrieval with extensive features and deep learning," *Applied Sciences (Switzerland)*, vol. 12, no. 13, p. 6753, Jul. 2022, doi: 10.3390/app12136753.
- [19] B. Kim, Y. Yang, J. S. Park, and H. J. Jang, "A convolution neural network-based representative spatio-temporal documents classification for big text data," *Applied Sciences (Switzerland)*, vol. 12, no. 8, p. 3843, Apr. 2022, doi: 10.3390/app12083843.
- [20] A. Rawat, M. A. Wani, M. ElAffendi, A. S. Imran, Z. Kastrati, and S. M. Daudpota, "Drug adverse event detection using text-based convolutional neural networks (TextCNN) technique," *Electronics (Switzerland)*, vol. 11, no. 20, p. 3336, Oct. 2022, doi: 10.3390/electronics11203336.
- [21] A. Sayeed, J. Shin, M. A. M. Hasan, A. Y. Srizon, and M. M. Hasan, "BengaliNet: a low-cost novel convolutional neural network for bengali handwritten characters recognition," *Applied Sciences (Switzerland)*, vol. 11, no. 15, p. 6845, Jul. 2021, doi: 10.3390/app11156845.
- [22] P. Preethi and H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," *Artificial Intelligence and Applications*, vol. 1, no. 2, pp. 119–127, Sep. 2022, doi: 10.47852/bonviewaia2202293.
- [23] Y. Wu, Y. Hu, and S. Miao, "Object detection based handwriting localization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12917 LNCS, 2021, pp. 225–239.
- [24] X. Wang, S. Zheng, C. Zhang, R. Li, and L. Gui, "R-yolo: a real-time text detector for natural scenes with arbitrary rotation," *Sensors (Switzerland)*, vol. 21, no. 3, pp. 1–21, Jan. 2021, doi: 10.3390/s21030888.
- [25] H. Sun, C. Tan, S. Pang, H. Wang, and B. Huang, "RA-YOLOv8: an improved YOLOv8 seal text detection method," *Electronics (Switzerland)*, vol. 13, no. 15, p. 3001, Jul. 2024, doi: 10.3390/electronics13153001.
- [26] J. Ryu and S. Kim, "Chinese character boxes: single shot detector network for Chinese character detection," *Applied Sciences (Switzerland)*, vol. 9, no. 2, p. 315, Jan. 2019, doi: 10.3390/app9020315.
- [27] S. Qu, K. Huang, A. Hussain, and Y. Goulermas, "A multipath fusion strategy based single shot detector," *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–16, Feb. 2021, doi: 10.3390/s21041360.
- [28] D. Olaniyan, R. O. Ogundokun, O. P. Bernard, J. Olaniyan, R. Maskeliūnas, and H. B. Akande, "Utilizing an attention-based LSTM model for detecting sarcasm and irony in social media," *Computers*, vol. 12, no. 11, p. 231, Nov. 2023, doi: 10.3390/computers12110231.
- [29] A. Amanat *et al.*, "Deep learning for depression detection from textual data," *Electronics (Switzerland)*, vol. 11, no. 5, p. 676, Feb. 2022, doi: 10.3390/electronics11050676.
- [30] J. Lim, I. Sa, H. S. Ahn, N. Gasteiger, S. J. Lee, and B. Macdonald, "Subsentence extraction from text using coverage-based deep learning language models," *Sensors*, vol. 21, no. 8, p. 2712, Apr. 2021, doi: 10.3390/s21082712.
- [31] Y. Ma *et al.*, "STEF: a swin transformer-based enhanced feature pyramid fusion model for dongba character detection," *Heritage Science*, vol. 12, no. 1, p. 206, Jun. 2024, doi: 10.1186/s40494-024-01321-2.
- [32] M. Lu, Y. Mou, C. L. Chen, and Q. Tang, "An efficient text detection model for street signs," *Applied Sciences (Switzerland)*, vol. 11, no. 13, p. 5962, Jun. 2021, doi: 10.3390/app11135962.
- [33] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, vol. 2019-June, pp. 9357–9366, doi: 10.1109/CVPR.2019.00959.
- [34] M. Liao, B. Shi, and X. Bai, "TextBoxes++: a single-shot oriented scene text detector," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3676–3690, Aug. 2018, doi: 10.1109/TIP.2018.2825107.
- [35] D. Sporici, E. Cuşnir, and C.-A. Boiangiu, "Improving the accuracy of tesseract 4.0 OCR engine using convolution-based preprocessing," *Symmetry*, vol. 12, no. 5, p. 715, May 2020, doi: 10.3390/sym12050715.
- [36] Y. Liu, Y. Wang, and H. Shi, "A convolutional recurrent neural-network-based machine learning for scene text recognition application," *Symmetry*, vol. 15, no. 4, p. 849, Apr. 2023, doi: 10.3390/sym15040849.
- [37] Y. Li, M. Du, and S. He, "Attention-based sequence-to-sequence model for time series imputation," *Entropy*, vol. 24, no. 12, p. 1798, Dec. 2022, doi: 10.3390/e24121798.
- [38] Y. Sun, C. Zhang, Z. Huang, J. Liu, J. Han, and E. Ding, "TextNet: irregular text reading from images with an end-to-end trainable network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11363 LNCS, 2019, pp. 83–99.
- [39] H. Xu, Y. He, X. Li, X. Hu, C. Hao, and B. Jiang, "Joint subtitle extraction and frame inpainting for videos with burned-in subtitles," *Information*, vol. 12, no. 6, p. 233, May 2021, doi: 10.3390/info12060233.
- [40] H. Liu, "Video text tracking for dense and small text based on pp-yoloe-r and sort algorithm," *arXiv preprint arXiv:2304.00018*, 2023, [Online]. Available: <https://arxiv.org/abs/2304.00018>0Ahttps://arxiv.org/pdf/2304.00018.
- [41] Y. Zhao, W. Wu, Z. Li, J. Li, and W. Wang, "FlowText: synthesizing realistic scene text video with optical flow estimation," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2023, vol. 2023-July, pp. 1517–1522, doi: 10.1109/ICME55011.2023.00262.

BIOGRAPHIES OF AUTHORS

Dayananda Kodala Jayaram    is working as associate professor and Head in the Department of Computer Application, GSSS SSFGC, Mysuru, Karnataka, India. He has received his M.Tech. and B.E. from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. He is pursuing his Ph.D. from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. His teaching and research interests are in the field of data mining, machine learning, and image processing. He has around total teaching experience of 10 years. He can be contacted at email: dayananda.kem@gmail.com.



Dr. Puttegowda Devegowda    is working as professor and Head in the Department of Computer Science and Engineering, ATME College of Engineering, Mysuru, Karnataka, India. He has received his Ph.D. from Mysuru University, Mysuru, Karnataka, India. His teaching and research interests are in the field of data mining, video processing, machine learning, and image processing. He has around total teaching experience of 20 years. He can be contacted at email: puttegowda.77@gmail.com.