A review on Automated Annotation System for Document Text Images

Ashlin Deepa R N
Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
rndeepa.pradeep@gmail.com

Boda Srija
Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
srijaboda5@gmail.com

Maria Jabeen
Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
jabeen.maria2003@gmail.com

Kushi Reddy Kankar
Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
kushireddykankar@gmail.com

Bollepalli Jhansi Lakshmi
Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
jhansibollepalli18@gmail.com

Atul Negi
School of Computer and Information Sciences
University of Hyderabad
Hyderabad, India
atul.negi@uohyd.ac.in

2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) | 979-8-3503-6178-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC-CGU58078.2024.10530793

***Abstract*—Automated Annotation Systems for text play a crucial role in automatically adding descriptive metadata or labels to textual data. Over recent years, there has been a notable surge in interest among researchers and practitioners in automated annotation systems, particularly those focused on text images, including handwritten text images. This comprehensive review delves into the latest advancements in Automated Annotation Systems tailored for text images. It meticulously examines various aspects such as data creation, pre-processing steps, word detection, word recognition, and annotation models as presented across diverse research papers.**

***Index Terms*—Automated Annotation, Optical Character Recognition, Handwritten Text Recognition, CNN, RNN, LSTM, CTC, Transformers**

I. Introduction

Optical Character Recognition (OCR) is a pivotal technology that enables the transformation of printed or handwritten text from physical documents or images into machine-readable and editable text. It plays a crucial role in converting books, articles, historical records, and various textual materials into digital formats. There have been numerous remarkable advancements in the field of OCR.

While OCR primarily focuses on printed text, HTR takes the recognition of text a step further by specialising in handwritten content. [3] introduces an optimised OCR system tailored for handwritten Marathi text document classification and recognition. Many historical documents, archives, and manuscripts contain handwritten text, which presents unique challenges for recognition due to variations in writing styles, ink fading, and paper degradation. HTR technology employs neural networks, machine learning algorithms, and pattern recognition techniques to decipher and transcribe handwritten content, preserving the authenticity and historical significance of such documents. Its primary aim is to convert handwritten text into digital format.

Handwritten documents have been a vital source of historical and cultural information, but the task of annotating and transcribing them manually is both time-consuming and error-prone. Automated Annotation prevails over manual annotation due to its superior efficiency and user-friendly nature.

Automated Annotation Systems[4] are used to streamline and improve various tasks related to text analysis, information retrieval, and data organisation. It enhances the organisation, searchability, and understanding of textual data. They typically employ techniques from Natural Language Processing (NLP), machine learning, and data mining to analyse and understand text, enabling the automated generation of tags, categories, or summaries for large volumes of textual content.

[5] proposes an ingenious automatic annotation-based approach for digitised text recognition. Annotating textual content within documents is a pivotal task in fields such as information management, content indexing, and text mining. Manual annotation of documents can be labour-intensive, time-consuming, and prone to errors. Automated annotation systems offer an efficient and user-friendly alternative. They not only expedite the annotation process but also improve its accuracy. It has wide-ranging applications, including in content recommendation, information retrieval, sentiment analysis, and content categorization.

II. Literature Survey

The evolution of OCR is marked by the adoption of models such as Convolutional Recurrent Neural Networks (CRNN), Bidirectional Long Short-Term Memory (BiLSTM), Connectionist Temporal Classification (CTC), and others, significantly enhancing OCR's accuracy and versatility.

A. CRNN

The Convolutional Recurrent Neural Network (CRNN) stands out for its ability to combine the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for sequential data analysis. Papers [6], and [7] showcase CRNNs effectiveness in recognizing Urdu text, handling sequential data like handwriting, and achieving precise text separation.

B. LSTM and Bi-LSTM

Long Short-Term Memory networks (LSTMs) and Bidirectional LSTMs prove powerful in sequence modelling, particularly for OCR. Addressing challenges in historical documents and diverse scripts, papers [9], and [10] demonstrate the efficacy of these architectures in word and character recognition, Sanskrit OCR, and transfer learning for various Indian languages.

C. End2End Network

End-to-end networks streamline OCR by directly processing input data, eliminating the need for manual feature extraction. Papers [11] and [12] illustrate their utility in word recognition, retrieval from challenging documents, and locating handwritten text within unstructured collections.

D. Transformers

Transformers have revolutionised OCR with their efficient handling of contextual information. Papers [13] and [14] discuss the challenges in adapting Transformers to handle sequential visual data and underrepresented Indian languages, respectively. Additionally, [15] introduces a model combining convolutional and Transformer networks, contributing to the automation of text recognition from images.

E. CTC

Connectionist Temporal Classification (CTC) introduces an innovative technique for recognizing uninterrupted sequences of data. Papers [16] and [17] highlight its superiority in tasks like Handwritten Text Recognition (HTR) and speech recognition, showcasing its promising approach for labelling unsegmented sequence data. Techniques like Word Beam Search (WBS) further enhance the applicability of CTC in decoding neural network outputs.

In summary, the diverse range of models and techniques presented in these papers reflects the continuous innovation in OCR, addressing challenges, and expanding its capabilities for various text recognition tasks in different languages and contexts.

III. METHODOLOGY

The methodology of a typical annotation system represents a systematic and strategic framework dedicated to the transformation of raw data into not just annotated but highly usable and insightful information. This comprehensive process involves several key steps: (i) Dataset creation (ii) Pre-processing (iii) Word Detection (iv) Word Recognition and Annotation

A. Dataset creation

Recent OCR research has led to advanced understanding beyond machines and computers. OCR types include those based on font restrictions, character recognition, and image extraction, catering to various customer needs. OCR can handle handwritten or machine-printed text. Earlier OCR was simpler due to consistent character styles and positioning, but it faces complexity with diverse writing patterns and languages. OCR can be categorised as online (process during writing) and offline (process on static data), with online systems being less complex and capturing writing speed. Offline systems involve complex pattern recognition. Table I highlights some of the publicly available benchmark datasets.

TABLE I
PUBLICLY AVAILABLE BENCHMARK DATASETS

Dataset [reference]	**Language**	**Lexicon sizes**	**Dataset type**
IAM[18]	English	13,353	Handwritten
IIIT5K[19]	English	50 and 1k	Printed
DIBCO[20]	Multilingual	-	Handwritten
IC03[21]	English	50, 1k and full	Printed
IIIT-HW-WORDS[22]	Multilingual (Indian Languages)	80,000	Handwritten
CEDAR[23]	English	2640	Handwritten
SVT-P[24]	English	50 and full	Printed
CTW1500[25]	English, Chinese	1500	Printed

B. Pre-processing

Pre-processing in text recognition refers to a set of techniques and operations applied to the raw input data, such as scanned documents or images, with the aim of improving the overall quality and readability of the text before it undergoes recognition. The primary goal of pre-processing is to enhance the clarity of text, remove noise, correct distortions, and prepare the data for accurate text recognition by OCR engines. Effective pre-processing is essential for achieving high OCR accuracy and is often tailored to the specific characteristics of the input data.

In Table II, various pre-processing techniques and their corresponding works are comprehensively presented, providing a detailed overview of different methodologies employed in the field.

In Fig. 1, we conducted a comparative assessment of three distinct binarization techniques: threshold-based noise reduction, Otsu, and DE-GAN. The results unequivocally demonstrate that DE-GAN distinguishes itself as the most effective

TABLE II
DIFFERENT PRE-PROCESSING TECHNIQUES AND THEIR RELATED WORKS

S. No.	Category[reference]	Methodology	Description	Dataset - Language
1	Binarization[26]	DE-GAN	Introduced a novel Document Enhancement Generative Adversarial Network, DE-GAN, designed to effectively restore severely degraded document images and compared it to existing methods proving its ability in image restoration.	Several DIBCO datasets - Urdu, English etc.
2	Deblurring[26]	DE-GAN	The comparative analysis concludes that DE-GAN surpasses the CNN and gives more accuracy than the pix2pix-HD model.	DIBCO - English
3	Binarization[27]	non-local p-Laplacian	Proposes a nonlinear p-Laplacian diffusion binarization method that enhances degraded images by estimating and then removing the undesirable background.	Several H-DIBCO datasets - Handwritten Multiple Languages Texts
4	Watermark Removal[28]	Advanced Unet	This work presents a deep learning-based technique AdvancedUnet neural network which extracts and removes watermarks.	Microsoft COCO val2014 dataset - Image Dataset
5	Skew-correction[29]	Spatial Transformer	Introduces Spatial Transformers, a novel module for CNN's, addressing the limitations of standard CNN's in handling spatial variations and transformations in input data.	MNIST - Handwritten Number English Dataset

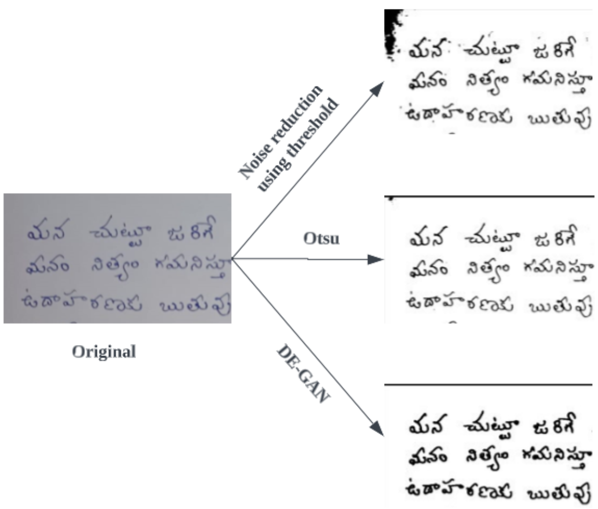

Fig. 1. Comparison of Binarization Techniques

method. Furthermore, it is worth noting the computational efficiency of these methods. Noise reduction using a threshold required 20 seconds for execution, Otsu only 5 seconds, whereas DE-GAN took 30 seconds to complete the process.

C. Word Detection

Word Detection or Localization is a fundamental step in the process of text recognition for documents. This crucial initial phase involves identifying and precisely locating individual words or text regions within a document image. The objective of word detection is to delineate the boundaries of each word, segmenting them from the surrounding content, such as background noise. Accurate word localization is instrumental in improving the overall accuracy and efficiency of automated systems. Table III presents various text detection models and their key attributes.

As mentioned in Table IV, a comprehensive evaluation of model performance across various datasets reveals that the CRAFT model clearly stands out, outperforming all other models across various performance metrics.

The CRAFT Model[38], short for Character Region Awareness for Text Detection, offers an innovative approach to text detection. CRAFT is designed to localise individual character regions within text and link these detected characters to form complete text instances. This approach capitalises on character-level region awareness, making it possible to represent text in various shapes with ease. Instead of treating text as a whole, CRAFT focuses on each character's location and its relationship with adjacent characters.

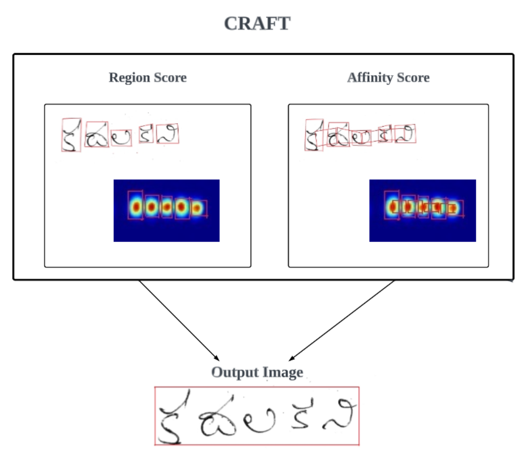

Fig. 2. Working of CRAFT

In the training process, for each input image, CRAFT generates ground truth labels for two critical aspects: the region score and the affinity score with character-level bounding

TABLE III
EXISTING TEXT DETECTION MODELS

Model [Reference]	Methodology	Dataset - Language	Text Type	Supervision
FCE: Fourier Contour Embedding [30]	FCE method for text contour representation and FCENet for text detection.	CTW1500, Total-Text - English and Chinese	Arbitrary-shaped texts	Yes
EAST: An Efficient and Accurate Scene Text Detector[31]	Fully Convolutional Network (FCN), Non-Maximum Suppression (NMS) merging stage.	ICDAR 2015, COCO-Text and MSRA-TD500 - English	Horizontal, Multi-Oriented, Curved	-
SegLink: Segment Linking [32]	Oriented text detection using an end-to-end trained, fully-convolutional neural network.	SynthText, ICDAR 2015, MSRA-TD500 - English, Chinese	Multi-oriented, Horizontal, Inclined, Arbitrary oriented strings	Yes
Shape Robust Text Detection with Progressive Scale Expansion Network[33]	Progressive Scale Expansion Network (PSENet).	CTW1500, Total-Text, ICDAR 2015, and ICDAR 2017 MLT	Curved, Multi-oriented and Horizontal texts	-
Real-time Scene Text Detection with Differentiable Binarization[34]	Differentiable Binarization (DB).	SynthText, MLT-2017, ICDAR 2015, MSRA-TD500, CTW1500, and Total-Text.	Curved, Multi-oriented, Multi-language	Yes
DeepText[35]	Ambiguous text category (ATC) information and multilevel region-of-interest pooling (MLRP) for text localization.	ICDAR 2011 and 2013 - English	Multi-oriented, Horizontal	-
ABCNet: Real-time Scene Text Spotting with Adaptive Bezier-Curve Network[36]	Bezier curve detection and BezierAlign with a recognition branch.	Total-Text and CTW1500	Curved	No
SCAST: Subcategory-aware self-training[37]	Adaptive scene text detection using SCAST.	SynthText, ICDAR13, ICDAR15, COCO-Text17, Total-Text - English	Multi-oriented	Yes
CRAFT: Character Region Awareness for Text Detection[38]	Convolutional neural network to generate character region score and affinity score.	ICDAR2013-English, ICDAR2015-English, ICDAR2017-9 languages, MSRA-TD500-English and Chinese, TotalText-English, CTW-1500 Chinese	Various text shapes	Weakly-supervised

TABLE IV
COMPARISON OF TEXT DETECTION MODELS

Model	Dataset	F-measure (%)	Precision (%)	Recall (%)
SCAST[37]	ICDAR13	77.36	76.48	78.26
EAST[31]	ICDAR 2015	78.2	83.6	73.5
FCENet[30]	CTW1500	85.5	87.6	83.4
PSENet-1s[33]	ICDAR 2015	85.7	86.9	84.5
DeepText[35]	ICDAR 2011	85	87	83
DB-ResNet-50[34]	ICDAR 2015	87.3	91.8	83.2
SegLink[32]	ICDAR 2013	85.3	87.7	83
ABCNet[36]	Inverse-Text	No Lexicon-45.2 Full Lexicon-34.3	-	-
CRAFT[38]	ICDAR 2013	95.2	97.4	93.1

boxes. The region score reflects the probability that a given pixel is the centre of a character, while the affinity score represents the likelihood that the pixel lies at the centre of the space between adjacent characters. This character-level detection approach enables convolutional filters to concentrate on the intra-character and inter-character aspects, rather than the entire text instance, leading to more precise results.

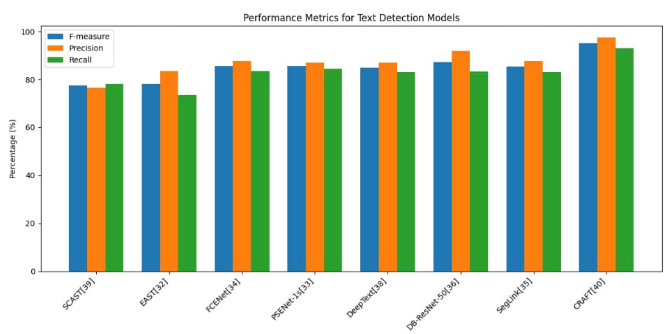

Fig. 3. Performance Metrics for Text Detection Models

D. Word Recognition and Annotation

Text recognition in document images is a challenging yet crucial task with numerous applications, ranging from digitising historical manuscripts to aiding in automatic form processing. This field involves developing algorithms and models that can accurately transcribe and understand the content of handwritten text within images. Table V gives an overview of some existing text recognition models.

Handwritten text recognition (HTR) models can be broadly categorised into two main types:

1) Statistical models: Models such as HMMs, SVMs, Random Forests etc. rely on statistical methods to extract features from handwritten text images and then use these features to classify the text.

TABLE V
EXISTING TEXT RECOGNITION MODELS

Author[paper]	Dataset - Language	Methodology	Accuracy
Vijaya Krishna[39]	UHTelPCC - Telugu Character Level[40]	MOMFO-DL	99.00%
Tejasree Ganji[41]	Own dataset - Telugu Character Level	CNN, VGGNET model	Model 1- 92%, Model 2- 80.5%
C. Suganthe[42]	HPLabs India - Tamil Character Level	CNN, Hyperparameter tuning	85.15%
Chauhan[43]	CMATERdb 3.4.1, Amrita MalCharDb, Kannada-mnist	HCR-Net: Deep Learning recognition network	Telugu - 99.13%, Malayalam - 94.87%, Kannada - 86.61%
V. Jayanthi[44]	DWP-H - Tamil Word Level	MMU-SNet	96.8%
DHANDRA B.V[45]	Own dataset - Telugu and Kannada numericals	KNN and SVM classifiers	KNN - 96.18%, SVM - 97.81%
Najwa[46]	Hijja, AHCD - Arabic Character Level	CNN	97% and 88%
Dezhi Peng[47]	ICDAR2013, SCUT-HCCDoc - Chinese Character Level	Segmentation-based, ConR	94.46% and 90.71%

2) Deep learning models: Models such as CNNs, RNNs, GANs etc. use deep neural networks to extract and learn features from handwritten text images.

Annotation:

After recognizing word images in a handwritten document, the task involves aligning the recognized text with a text sequence to automatically generate ground truth transcriptions corresponding to the word images.In [4], the recognized text is mapped to the corresponding word images based on a Levenshtein distance threshold.

TABLE VI
COMPARISION OF DIFFERENT ANNOTATION MODELS

Author[Paper]	Category	Precision	Recall (%)
Stork[8]	MLP-CNN-BLSTM - English	1.0	75
Gunna[9]	STAR-Net - Telugu	62.13	-
Dmytro[1]	LSTM - English	> 85	98.91

The paper[2] presents an interactive pipeline for digitizing handwritten English manuscripts using deep learning and user interaction. It combines a detection system with a custom recognition model, achieving high accuracy in annotating handwritten text. The system outperforms previous models on the IAM dataset and demonstrates robustness with challenging handwritten styles from the CVL dataset.

[48] investigates training models for handwritten text recognition with multiple noisy transcriptions. Using historical municipal registers from Belfort, France, it evaluates different training strategies and quality-based data selection methods. Results show that training with multiple transcriptions or computing a consensus is effective, but filtering based on agreement between annotators doesn't improve performance.

IV. CONCLUSION

In conclusion, this review provides a comprehensive exploration of automated annotation systems for document text images. The surveyed literature encompasses diverse methodologies, from dataset creation to sophisticated word detection and recognition techniques. It identifies diverse models across various languages and thoroughly analyses their performance. Notably, our findings indicate that binarization techniques such as DEGAN for preprocessing, CRAFT for word detection, deep learning models for recognition, and LSTM-based models for annotation demonstrate superior efficacy on handwritten images.

ACKNOWLEDGMENT

We thank the Department of Science and Technology(DST), Government of India for providing financial support for our project (Sanction Number: SP/YO/2021/2096 (C & G)).

REFERENCES

[1] Zhelezniakov, Dmytro et al. “A New Approach to Data Annotation Automation for Online Handwritten Mathematical Expression Recognition based on Recurrent Neural Networks.” 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (2021): 1125-1132.

[2] P. Guruprasad, S. S, V. C, and S. Chakravarthy, ”An end-to-end, interactive Deep Learning based Annotation system for cursive and print English handwritten text,” 2023.

[3] Ramteke, Surendra P. et al. “A Streamlined OCR System for Handwritten Marathi Text Document Classification and Recognition Using SVM-ACS Algorithm.” International Journal of Intelligent Engineering and Systems (2018): n. pag.

[4] Mondal, A., Tulsyan, K., & Jawahar, C.V. (2022). Automatic Annotation of Handwritten Document Images at Word Level. Proceedings of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing.

[5] Sankar, K. Pramod and C. V. Jawahar. ”Enabling Search over Large Collections of Telugu Document Images - An Automatic Annotation Based Approach.” Indian Conference on Computer Vision, Graphics & Image Processing (2006).

[6] Kong, L., Dyer, C., & Smith, N. A. ”Segmental Recurrent Neural Networks”. arXiv e-prints, 2015. doi: 10.48550/arXiv.1511.06018.

[7] Jubaer, Sheikh Mohammad, et al. ”BN-DRISHTI: Bangla Document Recognition through Instance-level Segmentation of Handwritten Text Images.” arXiv preprint arXiv:2306.09351 (2023).
[8] Stork, Lise, et al. ”Automated semantic annotation of species names in handwritten texts.” European Conference on Information Retrieval. Cham: Springer International Publishing, 2019.
[9] Gunna, Sanjana, Rohit Saluja, and C. V. Jawahar. ”Transfer learning for scene text recognition in Indian languages.” International Conference on Document Analysis and Recognition. Cham: Springer International Publishing, 2021.
[10] Kass, Dmitrijs, and Ekta Vats. ”AttentionHTR: handwritten text recognition based on attention encoder-decoder networks.” International Workshop on Document Analysis Systems. Cham: Springer International Publishing, 2022.
[11] Bansal, Siddhant, Praveen Krishnan, and C. V. Jawahar. ”Fused Text Recogniser and Deep Embeddings Improve Word Recognition and Retrieval.” International Workshop on Document Analysis Systems. Cham: Springer International Publishing, 2020.
[12] Gongidi, Santhoshini and C. V. Jawahar. “Handwritten Text Retrieval from Unlabeled Collections.” International Conference on Computer Vision and Image Processing (2021).
[13] Aberdam, Aviad, et al. ”Sequence-to-sequence contrastive learning for text recognition.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[14] Jain, Kushal, et al. ”Indic-transformers: An analysis of transformer language models for Indian languages.” arXiv preprint arXiv:2011.02323 (2020).
[15] Singh, S.S., Karayev, S. (2021). Full Page Handwriting Recognition via Image to Sequence Extraction. In: Lladós, J., Lopresti, D., Uchida, S. (eds) Document Analysis and Recognition – ICDAR 2021. ICDAR 2021. Lecture Notes in Computer Science(), vol 12823. Springer, Cham. https://doi.org/10.1007/978-3-030-86334-0_4
[16] Scheidl, Harald, Stefan Fiel, and Robert Sablatnig. ”Word Beam Search: A Connectionist Temporal Classification Decoding Algorithm.” 253-258. 10.1109/ICFHR-2018.2018.00052.
[17] Mathew, Minesh, and C. V. Jawahar. ”An empirical study of CTC based models for OCR of Indian languages.” arXiv preprint arXiv:2205.06740 (2022).
[18] Marti, U-V., and Horst Bunke. ”The IAM-database: an English sentence database for offline handwriting recognition.” International Journal on Document Analysis and Recognition 5 (2002): 39-46.
[19] Mishra, Anand, Karteek Alahari, and C. V. Jawahar. ”Scene text recognition using higher order language priors.” BMVC-British machine vision conference. BMVA, 2012.
[20] Mustafa, W. A., et al. ”A comprehensive review on document image (DIBCO) database.” IOP Conference Series: Materials Science and Engineering. Vol. 557. No. 1. IOP Publishing, 2019.
[21] Simon M Lucas, Alex Panaretos, Luis Sosa, Anthony Tang, Shirley Wong, and Robert Young. 2003. ”ICDAR 2003 robust reading competitions.” In Proceedings of ICDAR. 682–687.
[22] Gongidi, Santhoshini, and C. V. Jawahar. ”iiit-indic-hw-words: A Dataset for Indic Handwritten Text Recognition.” Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV 16. Springer International Publishing, 2021.
[23] Srinivasan, Harish et al. “Machine Learning for Signature Verification.” Indian Conference on Computer Vision, Graphics & Image Processing (2006).
[24] Trung Quy Phan, Palaiahnakote Shivakumara, Shangxuan Tian, and Chew Lim Tan. 2013. ”Recognizing text with perspective distortion in natural scenes.” In Proceedings of ICCV. 569–576.
[25] Yuliang, Liu & Lianwen, Jin & Zhang, Shuaitao & Sheng, Zhang. (2017). ”Detecting Curve Text in the Wild: New Dataset and New Solution.”
[26] Souibgui, M. A. and Kessentini, Y., ”DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement”, arXiv e-prints, 2020. doi:10.48550/arXiv.2010.08764.
[27] Bella, Fatim Zahra Ait et al. “An innovative document image binarization approach driven by the non-local p-Laplacian.” EURASIP Journal on Advances in Signal Processing 2022 (2022): n. pag.
[28] Wei, Rongfeng. ”Visual Watermark Removal Based on Deep Learning.” arXiv preprint arXiv:2302.11338 (2023).
[29] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. ”Spatial transformer networks.” Advances in neural information processing systems 28 (2015).
[30] Zhu, Yiqin et al. “Fourier Contour Embedding for Arbitrary-Shaped Text Detection.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 3122-3130.
[31] Zhou, X., “EAST: An Efficient and Accurate Scene Text Detector”, *arXiv e-prints*, 2017. doi:10.48550/arXiv.1704.03155.
[32] B. Shi, X. Bai and S. Belongie, ”Detecting Oriented Text in Natural Images by Linking Segments,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 3482-3490, doi: 10.1109/CVPR.2017.371.
[33] Wang, Wenhai, et al. ”Shape robust text detection with progressive scale expansion network.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
[34] Liao, Minghui, et al. ”Real-time scene text detection with differentiable binarization.” Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020.
[35] Z. Zhong, L. Jin and S. Huang, ”DeepText: A new approach for text proposal generation and text detection in natural images,” 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 1208-1212, doi: 10.1109/ICASSP.2017.7952348.
[36] Liu, Yuliang, et al. ”Abcnet: Real-time scene text spotting with adaptive bezier-curve network.” proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
[37] Tian, Zichen & Xue, Wenyuan & Zhang, Lei & Wang, Xinggang & Qiao, Yu. (2016). ”Detecting Text in Natural Image with Connectionist Text Proposal Network.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 4353-4360. 10.1109/CVPR.2016.488.
[38] Liu, W. et al. “SSD: Single Shot MultiBox Detector.” European Conference on Computer Vision (2015).
[39] Sonthi, Vijaya Krishna, and Krishnaraj N. ”An Intelligent Telugu Handwritten Character Recognition Using Multi-Objective Mayfly Optimization with Deep Learning–Based DenseNet Model.” ACM Transactions on Asian and Low-Resource Language Information Processing 22.3 (2023): 1-16.
[40] Kummari, Rakesh and Chakravarthy Bhagvati. “UHTelPCC: A Dataset for Telugu Printed Character Recognition.” International Conference on Recent Trends in Image Processing and Pattern Recognition (2018).
[41] Ganji, T., Sekhar Velpuru, M., and Dugyala, R., ”Multi Variant Handwritten Telugu Character Recognition Using Transfer Learning”, in ¡i¿Materials Science and Engineering Conference Series¡/i¿, 2021, vol. 1042, no. 1, p. 012026. doi:10.1088/1757-899X/1042/1/012026.
[42] R. C. Suganthe, K. Pavithra, N. Shanthi and R. S. Latha, ”A CNN model-based approach for offline handwritten Tamil text recognition system,” Natural Volatiles & Essential Oils Journal, vol. 8, no. 5, pp. 164–175, 2021.
[43] Chauhan, Vinod Kumar, Sukhdeep Singh, and Anuj Sharma. ”HCR-Net: A deep learning based script independent handwritten character recognition network.” arXiv preprint arXiv:2108.06663 (2021).
[44] V. Jayanthi and S. Thenmalar, ”Recognition of handwritten words from digital writing pad using mmu-snet,” Intelligent Automation & Soft Computing, vol. 36, no.3, pp. 3551–3564, 2023.
[45] Bv, Dhandra, Gururaj Mukarambi, and Mallikarjun Hangarge. ”A script independent approach for handwritten bilingual Kannada and Telugu digits recognition.” International Journal of Machine Intelligence 3 (2011): 155-159.
[46] Altwaijry, Najwa & Al-Turaiki, Isra. (2021). ”Arabic handwriting recognition system using convolutional neural network.” Neural Computing and Applications. 33. 10.1007/s00521-020-05070-8.
[47] P. Dezhi, J. Lianwen, M. Weihong, X. Canyu, Z. Hesuo et al., ”Recognition of handwritten Chinese text by segmentation: A segment-annotation-free approach,” IEEE Transactions on Multimedia, vol. 67, pp. 1520–9210, 2022.
[48] Tarride, Solène & Faine, Tristan & Boillet, Mélodie & Mouchère, Harold & Kermorvant, Christopher. (2023). Handwritten Text Recognition from Crowdsourced Annotations. 1-6. 10.1145/3604951.3605517.