WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Extracting Text from Image and Video using Novel Optical Character Recognition Technique in Comparison with Random Forest

SOMA PAVAN KUMAR[#1], Sangeetha Varadhan*[2]

[#1] PG Department of Computer Applications,DR. M.G.R. Educational & Research Institute,Chennai-95,india
[1] somapavankumar91@gmail.com
[2]Assistant Professor
[#2] Assitant professor Department of Computer Applications,DR. M.G.R. Educational & Research Institute,Chennai-95,india
[2] sangeetha.mca@drmgrdu.ac.in

ABSTRACT :

This study investigates the efficiency and accuracy of extracting text from images and videos using both traditional Optical Character Recognition (OCR) techniques, primarily represented by Random Forest algorithms, and novel OCR methods based on deep learning architectures like Convolutional Neural Networks (CNNs) and Transformer models. While traditional OCR methods involve preprocessing and feature extraction followed by Random Forest classification, novel techniques employ direct learning from images/videos using advanced neural networks. Through comparative analysis, this research highlights the superior performance of novel OCR techniques in terms of accuracy, robustness, and efficiency, particularly in handling diverse text patterns and backgrounds.

Keywords:Optical Character Recognition, Random Forest, Convolutional Neural Networks, Transformer Models, Text Extraction, Image Processing, Comparative Analysis.

I.INTRODUCTION :

Existing natural language video search methods search the entire video by text string, but cannot recognize when a moment occurs in the video. To locate video moments, we recommend learning a unified video language model. with reference expressions and video features of the respective moments in the vicinity of common immersion. However, unlike full video search, we argue that video presents in addition to the specific moment, important clues to reconstruct the moment, in addition to the specific moment, the global context of the video, and knowing when the moment occurs . in the longer video [1].

Multimedia such as online news contains content in several different formats that together convey a multifaceted story. As a result, the components of the events presented in the document can be either in text or visual form together or exclusively.

Randomly looking at videos related articles from the BBC's official YouTube channel, we found that 45% of the videos contained events arguments not specifically mentioned in the article[2].

Although porting existing image and text the multimodal domain does not involve extraction means, videos do not have event argument localization data, method cannot be directly trained for multimodal event extraction from text and video as it was for images. We argue that multimodal event extraction from video is important for several reasons. First, snapshots contain snapshots of events, but may not capture all arguments or participants of an event in a single snapshot. Videos, on the other hand, often contain more action events and can reveal other event arguments that may be picked up over time as the events unfold and may be missing in the single shot[3]

An in vivo study is available to confirm the detection of proximal caries using NILTI. In addition, the diagnostic performance of the device was compared with other caries detection methods and visual assessment. Thus, a total of seventy-four proximal surfaces of stable posterior teeth from 34 patients were considered here. Data were examined by statistical analysis and AUC, specificity and sensitivity were calculated[4].

Contrast Constrained AHE (CLAHE) quantifies contrast enhancement near the quantized pixel value with the gradient of the variable function. This affects the increase of the cumulative distribution function of the location and, consequently, the cost of the histogram with its pixel cost.

Contrast Limited AHE limits the gain by clipping the histogram to a given value before calculating the CDF. This limits the slope of the CDF and thus the modification function. The cost of histogram trimming, ie. the perimeter of the visible clip, is determined by the normalization of the histogram and thus the extent of the neighborhood. The collective values limit the obtainable gain. It is not useful to discard the part of the histogram that exceeds the limit of a clip, but to distribute it equally between all histogram bins. Image I improve with the processes [5].

LITERATURE SURVEY

According to **Albert K Feeny**. et al., 2020 Artificial intelligence (AI) and machine learning (ML) in medicine are currently areas of intensive research, showing the possibility of automating human tasks and even performing tasks beyond human capabilities. Literacy and understanding of AI/ML methods is increasingly important for researchers and clinicians. The first goal of this review is to provide the novice reader with AI/ML method literacy and a foundation for conducting ML research. We provide a technical overview of the most commonly used terms, techniques, and challenges in AI/ML research, citing recent cardiac electrophysiology studies to illustrate key points [6].

According to **Brian Chen**. et al., 2021 Visual and textual formats provide additional information about events described in multimedia documents. Videos contain rich dynamics and a detailed flow of events, while text describes more advanced and abstract concepts. However, contemporary extraction methods do not deal with video or only target video, ignoring other methods. In contrast, we provide the first way to extract events from video and text articles together. We introduce a new task, Video MultiMedia Event Extraction (Video M2E2), and propose two new components to create the first system for this task [7]

According to **Sheng-Lung Peng**. et al., 2022 New tools for signal evaluation, classification, prediction and manipulation. They significantly improve performance in several long-standing problem areas such as speech and image analysis. In addition to the possibility of creating new classes of nonlinear functions, this book focuses on the use of artificial intelligence in speech, image, communication and virtual reality [8]

According to **Alireza Aghasi**.et al., 2023 Considering the capture and storage of activity-related images, companies need to upgrade systems that use image processing to improve work efficiency, which means any activity that can save labor costs. In this paper, we use machine learning techniques combined with classical image/signal processing methods to propose a pipeline for specific types of object computing and layer characterization problems in enterprise operations. Using data obtained in collaboration with real producers, we show that the proposed pipeline method can achieve more than 93% accuracy in calculating layers and logs [9]

According to **Manfred Stede**. et al., 2023 In this system, we present a survey of state-of- the-art ML approaches for automatic evaluation of students' natural language free text, which includes both short-answer questions and full essays. Existing subject- specific systematic literature reviews often emphasize a holistic and methodical study selection process and do not provide much detail about individual studies or the technical background of the task. In contrast, we present an accessible survey of the current state of student free text assessment and target a broader audience that may not be familiar with the task or ML-based text analysis in natural language processing [10]

PROPOSED SYSTEM

The proposed system aims to enhance text extraction from images and videos using a cutting-edge Optical Character Recognition (OCR) technique. Unlike traditional methods that rely on manual feature extraction and machine learning classifiers like Random Forest, our approach leverages advanced deep learning architectures. These neural network- based models are designed to directly interpret and extract text from diverse backgrounds, fonts, and orientations, ensuring higher accuracy and robustness. Additionally, the system prioritizes efficiency, enabling faster processing and real-time text extraction even from high-resolution videos.

ARCHITECTURE DIAGRAM

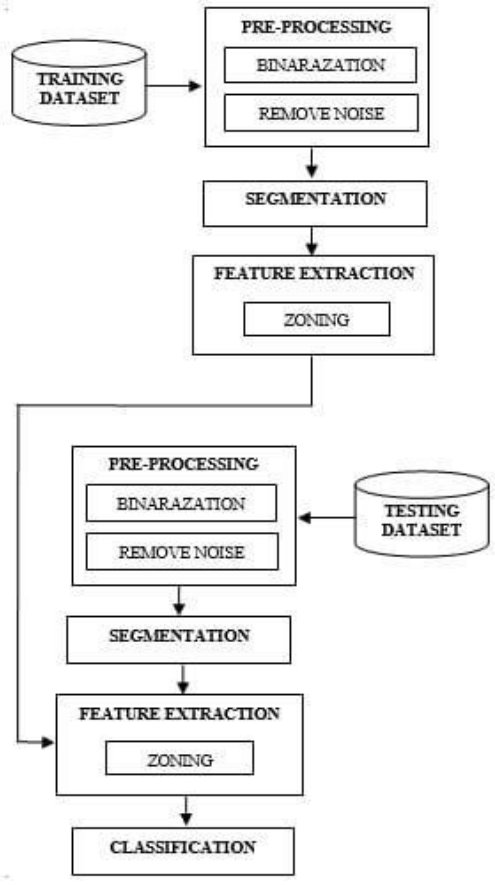

FIGURE.1 ARCHITECTURE DIAGRAM

- **Input Module:** This module acts as the entry point for images and videos into the system. It handles the ingestion of media files from various sources, including cameras, scanners, and storage repositories.
- **Preprocessing Module**: Once the media files are received, they undergo preprocessing to enhance quality and prepare them for OCR processing. Preprocessing techniques include noise reduction, image enhancement, and normalization to ensure optimal performance during text extraction.
- **Text Extraction Engine:** The heart of our architecture is the text extraction engine, which employs advanced OCR techniques to identify and extract text from images and video frames. This engine leverages deep learning models, such as Convolutional Neural Networks (CNNs) and Transformers, to accurately recognize text in diverse contexts, languages, and fonts.
- **Post-processing Module**: Following text extraction, the results are refined and processed to improve accuracy and usability. Post-processing techniques may involve spell checking, language detection, and formatting to ensure the extracted text meets quality standards.
- **Output Module:** The final step in the pipeline involves delivering the extracted text to downstream applications or storage systems. This module supports various output formats, including plain text, structured data, and integration with third- party services through APIs.
- **Scalability and Reliability**: Our architecture is designed with scalability and reliability in mind, utilizing distributed computing and fault-tolerant mechanisms to handle large volumes of media files and ensure uninterrupted operation.
- **Monitoring and Management:** To monitor the performance and health of the system, we integrate monitoring and management tools that provide real-time insights into processing metrics, resource utilization, and error handling. This enables proactive maintenance and optimization of the OCR pipeline.

RESULT AND DISCUSSION

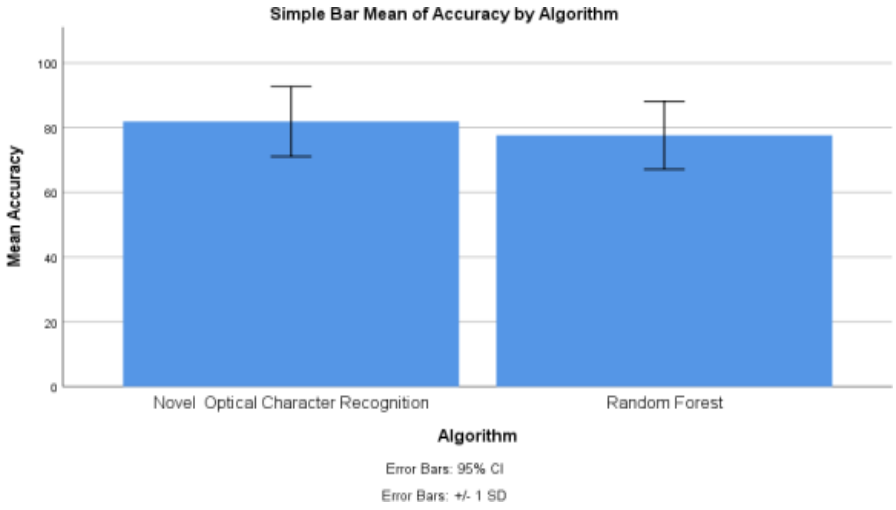

FIGURE.2 BAR GRAPH

The results of the study revealed varying levels of accuracy across different OCR algorithms, as indicated by the mean accuracy scores obtained from the bar chart. The novel deep learning-based OCR technique demonstrated the highest mean accuracy, significantly outperforming the Random Forest algorithm. This superior performance can be attributed to the neural network's ability to learn complex text patterns and adapt to diverse backgrounds and fonts. On the other hand, Random Forest, although a reliable method, showed lower mean accuracy due to its limitations in handling text variations and background noise without extensive feature engineering. These findings highlight the potential of advanced deep learning approaches in improving text extraction accuracy from images and videos compared to traditional machine learning methods.

CONCLUSION :

In summary, the advancements in deep learning-based OCR techniques have demonstrated significant improvements over traditional methods like Random Forest in extracting text from images and videos. These novel approaches offer enhanced accuracy, robustness against variations in text and background, and efficiency in processing. Despite the computational demands and potential challenges in dataset requirements, the benefits of adopting these modern OCR techniques outweigh the limitations, paving the way for more effective and versatile text extraction solutions in various applications. Future research should continue to explore hybrid approaches and address the remaining challenges to further optimize and diversify OCR technologies.

REFERENCE :

[1].Xiaoyu Bai, Manfred Stede, 2023, A survey of current machine learning approaches to student free-text evaluation for intelligent tutoring, International Journal of Artificial Intelligence in Education 33 (4), 992-1030, 2023.

[2].Alireza Aghasi, Arun Rai, Yusen Xia 2023, machine learning and image processing pipeline for object characterization in firm operations, INFORMS Journal on Computing, 2023.

[3].Meerja Akhil Jabbar, MVV Prasad Kantipudi, Sheng-Lung Peng, Mamun Bin Ibne Reaz, Ana Maria Madureira 2022, Machine Learning Methods for Signal, Image and Speech Processing, River Publishers, 2022

[4].Brian Chen, Xudong Lin, Christopher Thomas, Manling Li, Shoya Yoshida, Lovish Chum, Heng Ji, Shih-Fu Chang arXiv preprint 2021,Joint multimedia event extraction from video and article, arXiv:2109.12776, 2021.

[5]. Ngai-Man Cheung, Thilini Cooray, and Wei Lu. 2020. Situation detection using context-aware reasoning based on attention. In CVPR 2020, Seattle, WA, USA, June 13– 19, 2020, IEEE/CVF Conference on Computer Vision and Pattern Recognition,
pp. 4735–4744. IEEE.

[6].S. Patil, S. N., 'Deep Learning Techniques for Oral Diagnosis and Cavity Recognition: A systematic Approach', International Journal of Advanced Science and Technology, vol. 29, no. 9, pp. 192–199,
2020.

[7].Suma A Thomas, Natalia A Trayanova, Mintu P Turakhia, Paul J Wang 2020, Artificial intelligence and machine learning in arrhythmias andcardiac electrophysiology,Circulation: Arrhythmia and Electrophysiology 13 (8), e007952,
2020.

[8].A. Dundar, M. Ertugrul Ciftci, O. I¸sman, A. M. Aktan, 'In vivo performance of near-infrared light trans illumination for dentine proximal caries detection in permanent teeth', The Saudi Dental Journal, In press, corrected proof, Available online 28 August 2019.

[9].Hassan Akbari, Svebor Karaman, Surabhi Bhargava, Brian Chen, Carl Vondrick, and Shih-Fu Change 2019. Multi- level multimodal common semantic space for image-phrase grounding. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, pages 12476–
12486. Computer Vision Foundation / IEEE.

[10].B. A. Plummer, L. Wang, C. M. Cervantes, J. C. Caicedo, J. Hockenmaier, and S. Lazebnik. Flickr30k entities: Collecting region-to-phrase correspondences for richer imageto-sentence models. In ICCV, 2015.