Literature Survey

1. Handwritten Text Recognition (HTR)(Jan 2024)

In their comprehensive study, Wissam AlKendi, Franck Gechter, Laurent Heyberger, and Christophe Guyeux explore the crucial role of Handwritten Text Recognition (HTR) in digitizing various document types, from historical manuscripts to modern forms. Their work emphasizes HTR's importance in efficient data storage, retrieval, and analysis, highlighting its applications in preserving historical records, improving medical data accessibility, and enhancing document management.

The authors [1] trace HTR's evolution from traditional methods like Optical Character Recognition (OCR) and Hidden Markov Models (HMM) to advanced machine learning approaches. They note that while OCR and HMM often struggled with handwriting variability, modern techniques utilizing Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have significantly improved accuracy and robustness in handling diverse handwriting styles.

Despite technological advancements, the authors [1] point out persistent challenges in HTR. These include dealing with variability in handwriting styles, text quality issues, and alignment problems. Historical documents present unique difficulties such as overlapped characters and marginal annotations, necessitating continuous refinement of HTR techniques.

The researchers discuss how modern HTR systems, powered by sophisticated models like CNNs and RNNs, can recognize various handwriting styles across multiple languages. This capability has made HTR an indispensable tool in digitizing a wide range of handwritten documents.

Looking forward, Author [1] and his colleagues propose several research directions to enhance HTR systems further. They suggest improvements in preprocessing techniques to better handle text skew and document layout analysis. The authors also emphasize the need for refined segmentation processes, especially for preserving unique characteristics of historical texts.

Additionally, they advocate for advanced classification models that combine multiple classifiers and optimization techniques, particularly useful for hybrid documents containing both printed and handwritten text. The researchers stress the importance of enhanced post-processing methods to capture semantic relationships and reduce errors by filtering out less reliable predictions.

The authors [1] also highlight the scarcity of historical French datasets as a specific challenge, proposing data augmentation techniques to address this gap. They emphasize the importance of researchers possessing strong language skills, especially when working with historical documents, to ensure comprehensive understanding of grammar, vocabulary, and writing style variations across different periods.

In conclusion, the authors present a thorough overview of HTR's current state and future prospects, underscoring the need for continuous improvement to address the unique challenges posed by handwritten text recognition, particularly in historical contexts.

2. Scene Text Recognition (STR)(jul 2024)

Bangbang Zhou, Yadong Qu, Zixiao Wang, Zicheng Li, Boqiang Zhang, and Hongtao Xie have presented a significant advancement in Scene Text Recognition (STR) technology. Their research addresses crucial challenges in STR, particularly focusing on improving the recognition of challenging texts in natural environments.

The authors [2] identify two primary issues affecting STR accuracy: Large Intra-Class Variance (LICV) and Small Inter-Class Variance (SICV). LICV refers to significant visual differences within the same character category, while SICV involves high visual similarity between different character categories. These problems often lead to misrecognitions, especially in distorted or contextless text scenarios.

To tackle these challenges, the authors [2] propose a Character Features Enriched (CFE) model. This innovative approach incorporates two key components:

1. Character-Aware Constraint Encoder (CACE): This uses a decay matrix to guide the attention mechanism, focusing on morphological information at the character level. This enhancement improves feature discriminability.
2. Intra-Inter Consistency Loss (I2CL): This component addresses both LICV and SICV by learning long-term memory units for each character category, ensuring intra-class compactness and inter-class separability.

The researchers highlight that existing STR encoders, particularly those based on CNNs and Transformers, often struggle with limited receptive fields or fail to capture local character-level features effectively. Their CFE model aims to overcome these limitations.

In benchmarking their model, author [2] and colleagues compare the CFE against several state-of-the-art STR models, including RobustScanner, PARSeq, SVTR, CLIP-OCR, VisionLAN, ABINet, MATR, SRN, and MGP-Base. The CFE model demonstrates superior performance on common benchmarks and the challenging Union 14M-Benchmark, achieving impressive accuracy rates of 94.1% and 61.6%, respectively.

Looking towards future research, the authors suggest several directions:

1. Exploring new architectures to better capture character morphology and context nuances.
2. Improving data augmentation techniques to simulate more challenging real-world scenarios.
3. Integrating multimodal information by combining visual and linguistic cues more effectively.
4. Developing sophisticated loss functions that can dynamically adjust to different levels of character variance and similarity.

In conclusion, the work by Zhou, Qu, Wang, Li, Zhang, and Xie represents a significant step forward in STR technology. Their CFE model, with its innovative CACE and I2CL components, shows promise in enhancing the robustness and accuracy of STR systems, particularly in challenging real-world applications.

3. Comparison of OCR Algorithms(Jul 2024)

The comparative analysis of Google Vision OCR and Tesseract for newspaper text recognition was conducted by the authors [3] (2024). Their study, published in the Media of Computer Science journal, provides valuable insights into the current state and challenges of OCR technology.

The authors [3] developed a web application using PHP and Laravel to implement and compare these OCR systems. They captured high-quality newspaper images using a DSLR camera and processed them through both OCR algorithms. The researchers employed confusion matrix calculations to determine accuracy, precision, and sensitivity metrics for comparison.

The results obtained by the authors [3] demonstrated that Google Vision OCR outperformed Tesseract in terms of accuracy (97.89% vs. 90.16%), precision (98.62% vs. 96.62%), and sensitivity (99.23% vs. 92.82%). However, they found that Tesseract exhibited superior processing speed, averaging 12.96 seconds compared to Google Vision's 30.08 seconds, primarily due to its offline processing capability.

The authors [3] highlighted several ongoing challenges in OCR technology, particularly when dealing with complex newspaper layouts and varied text styles. They noted that achieving high accuracy, especially for low-quality or poorly lit images, remains a significant hurdle. The researchers also pointed out that processing speed and internet dependence for cloud-based OCR systems like Google Vision can be limiting factors in certain scenarios.

Looking towards future developments, the authors [3] suggested expanding testing to include a wider variety of document types, such as old newspapers, letters, and magazines. They proposed implementing OCR in educational settings, for instance, in library catalog search systems. The researchers recommended further work on optimizing image preprocessing techniques and exploring ways to improve Tesseract's accuracy while maintaining its speed advantage.

Overall, the study by the authors [3] (2024) provides a comprehensive benchmark for OCR performance on complex documents like newspapers, offering a solid foundation for future advancements in this crucial area of digital image processing. Their work contributes significantly to the understanding of current OCR capabilities and limitations, guiding future research and development in this field.

4. Text Extraction from Images(Jun 2023)

The authors [4] paper highlights the critical importance of text extraction from images in various applications such as document analysis, content-based image retrieval, and image captioning. The authors emphasize that this process is essential for converting visual information into machine-readable and searchable text, enabling automated data processing and analysis across multiple industries.

The researchers discuss several techniques employed in text extraction, including image preprocessing, text detection, localization, segmentation, and recognition. They particularly note the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in deep learning-based approaches, which have shown significant improvements over traditional OCR methods. The authors [4] and colleagues also mention the importance of post-processing techniques to enhance accuracy.

The paper identifies several challenges in text extraction, such as handling complex backgrounds, varying fonts, distorted or degraded text, and issues with multilingual text. They also highlight the difficulties in processing text in natural scene images, where lighting conditions, perspective, and occlusions can complicate extraction.

Regarding state-of-the-art systems, The authors [4] discuss the architecture of modern text extraction systems, which typically involve stages like object detection, text object detection, text classification, and text reconstruction. They note that current systems are increasingly leveraging machine learning and artificial intelligence to improve accuracy and efficiency.

Looking to the future, The authors [4] and colleagues suggest several avenues for advancement. They propose further improvements in accuracy, especially for complex image features and unconventional fonts. The authors also highlight the potential for real-time processing and multilingual capabilities, including on-the-fly translation. They envision text extraction technology revolutionizing industries like healthcare, finance, and retail by enabling quick data analysis and informed decision-making.

In conclusion, this comprehensive review by the authors [4] provides valuable insights into the current state and future directions of text extraction from images. Their work underscores the potential impact of this technology on various industries and everyday life, paving the way for more efficient data analysis and information utilization in our increasingly visual digital world.

5. Performance Evaluation of CNN Models for Text Detection(June 2024)

This paper, published by the authors [5], highlights the importance of text detection in digital advertisement images. The authors recognize that these images contain crucial information about products and services, and manual extraction of this information is time-consuming and error-prone. They emphasize the need for efficient and accurate text detection methods in the context of digital advertising.

The study employs Single-shot detection (SSD) architecture, which has been proven effective in balancing speed and accuracy for object detection tasks. The authors evaluate the performance of three Convolutional Neural Network (CNN) models—Resnet-50, Mobilenetv2, and Resnet-101—as feature extractors within the SSD framework. They created a dataset of 400 manually collected and annotated digital advertisement images for their experiments.

The researchers face several challenges in this domain, including the complex backgrounds of advertisement images and the varying sizes of text within them. They address these issues by using SSD's ability to detect objects of various sizes and by employing data augmentation techniques such as contrast enhancement, brightening, and darkening.

In terms of state-of-the-art systems, the paper compares the performance of the three CNN models. Their results show that Resnet-50 performs well in detecting small texts, achieving a mean Average Precision (mAP) of 0.736. Resnet-101 shows the highest overall mAP of 0.755, while Mobilenetv2 excels in detecting large texts and offers the fastest inference time.

For future work, the authors [5] suggest expanding the advertisement image dataset to improve the models' performance further. They also propose exploring other models to enhance the detection accuracy of small texts, which remains a challenging aspect of text detection in advertisement images.

This study contributes to the field by specifically focusing on text detection in advertisement images, an area that had not been extensively explored despite the vast literature on text detection in other contexts. The authors' work provides valuable insights into the performance of different CNN models for this specific application, paving the way for more efficient and accurate text extraction from digital advertisements

6. Text Extraction Using OCR Technology(May 2024)

The authors [6] paper highlights the critical importance of text extraction from images using Optical Character Recognition (OCR) technology in our increasingly digital world. The authors emphasize its wide-ranging applications in document digitization, image translation, and augmented reality. They present OCR as a crucial tool for converting scanned images and handwritten text into machine-readable formats, thereby enhancing accessibility and automating data processing across various industries.

The researchers discuss several techniques employed in OCR, including image preprocessing, pattern recognition, and the integration of machine learning algorithms. They particularly note the advancements brought about by deep learning, especially convolutional and recurrent neural networks, which have significantly improved OCR accuracy and efficiency.

The authors [6] and colleagues identify several challenges facing OCR technology, such as handling complex fonts and languages, addressing image noise, and managing computational complexity. They also point out the difficulties in extracting text from dynamic and non-standardized inputs like handwritten notes and video streams, highlighting this as a critical research gap.

The paper reviews state-of-the-art OCR systems, comparing the performance of different Convolutional Neural Network (CNN) models like Resnet-50, Mobilenetv2, and Resnet-101 within the Single-shot detection (SSD) framework. The authors also discuss the integration of OCR with other technologies such as computer vision and natural language processing, opening new frontiers for innovation.

Looking to the future, The authors [6] suggest several avenues for advancement. They propose expanding research into real-time text extraction from unconventional sources and improving OCR performance for underrepresented languages and dialects. The authors also recommend further integration of OCR with emerging technologies like augmented reality and privacy-preserving techniques.

In conclusion, this comprehensive review by The authors [6] and colleagues provides valuable insights into the current state and future directions of OCR technology. Their work underscores the potential impact of advanced text extraction techniques on various industries and everyday life, paving the way for more accessible and actionable information in our digital age.

7. Research Paper on Text Extraction using OCR(May 2023)

The authors [7] paper highlights the critical importance of text extraction from images using Optical Character Recognition (OCR) technology in various fields such as document digitization, information retrieval, and image understanding. The authors emphasize that accurate and efficient text extraction is essential for enabling automated processes and analyzing large volumes of visual data.

The researchers discuss several techniques employed in OCR, including image preprocessing (such as grayscale conversion, noise reduction, and binarization), text detection using convolutional neural networks (CNNs), character segmentation, and text recognition using recurrent neural networks (RNNs) or transformer-based models. They also mention post-processing techniques like language models and dictionary lookups to improve accuracy.

The author [7] and colleagues identify several challenges facing OCR technology, such as handling complex backgrounds, varying fonts, distorted or degraded text, and issues with small font sizes and low resolution, particularly in Chinese screen-rendered text. They also highlight the difficulties in processing tilted or skewed text in natural scenes.

The paper reviews some state-of-the-art OCR systems, mentioning advancements in deep learning techniques, particularly CNNs and RNNs, which have significantly improved text detection and recognition accuracy. The authors also discuss the integration of OCR with other technologies such as image enhancement and language processing.

Looking to the future, The authors [7] suggest several avenues for advancement. They propose developing OCR algorithms that can handle large databases of handwritten text, improving recognition of degraded or broken characters, and enhancing word recognition rates for compound characters. The authors also recommend exploring the conversion of extracted text to audio to aid visually impaired users. Furthermore, they suggest the need for creating standard databases for text recognition and improving OCR performance on various types of challenging text.

In conclusion, this comprehensive review by the author [7] and colleagues provides valuable insights into the current state and future directions of OCR technology. Their work underscores the potential impact of advanced text extraction techniques on various industries and everyday life, paving the way for more accessible and actionable information in our increasingly digital world.

8. Text Extraction and NLP in Medical Reports(Jun 2024)

This paper, published by the authors [8], highlights the importance of automating information extraction from medical report images using a combination of optical character recognition (OCR) and natural language processing (NLP) techniques. The authors emphasize the significance of this approach in streamlining healthcare decision-making processes by efficiently extracting valuable clinical data from scanned documents or photographs.

The researchers employ several key techniques, including genetic algorithms to fine-tune OCR hyperparameters, maximizing text extraction length. They also utilize NLP methods like named entity recognition (NER) to categorize extracted information into required entities. The adaptive model they developed adjusts document parameters to enhance entity recognition accuracy.

The study addresses many challenges, such as dealing with diverse formats of medical report images, handling unstructured text, and overcoming variations in document quality (e.g., darkness, blurriness, handwritten text). The authors also note difficulties in accurately identifying multiple instances of payment dates and amounts within a single document.

While the paper doesn't explicitly compare their system to others, it presents a novel approach that combines OCR parameter optimization with NLP-based entity extraction. This integrated method appears to be at the forefront of automated medical document processing.

Looking to the future, the authors suggest several areas for improvement and further research. They propose exploring the integration of state-of-the-art neural network architectures for image segmentation and text recognition tasks. Additionally, they recommend investigating the use of convolutional neural networks (CNNs) for image segmentation and recurrent neural networks (RNNs) for text recognition to potentially enhance document processing and information extraction.

The researchers also acknowledge the need for continued refinement and optimization of their adaptive model to address challenges in entity recognition and complex document format handling. They emphasize the potential of their approach to revolutionize document processing workflows across various sectors, including healthcare and finance.

In summary, the authors [8] present a promising approach to automated medical document processing, combining OCR optimization and NLP techniques. Their work contributes valuable insights to the field and opens up avenues for further research and development in document recognition and information extraction.

9. Plagiarism Checking with extracted text from image using NLP: A Comprehensive Overview

This paper by Daksha D, Bhuvana K, N N Siri, and Sanjana M explores an innovative approach to plagiarism detection by combining Natural Language Processing (NLP) techniques with text extraction from images. The authors [9] highlight the importance of this research in addressing the significant challenge of plagiarism in academic and professional settings. They emphasize that their proposed system can enhance plagiarism detection by analyzing textual content extracted from images, which is a crucial advancement in ensuring academic and professional integrity.

The techniques described in the paper involve using Optical Character Recognition (OCR) to extract text from images, followed by NLP algorithms to analyze and compare the extracted content. The authors [9] detail the process, including preprocessing steps like tokenization, cleaning, stop word removal, and stemming. They also discuss similarity analysis using measures such as cosine similarity to quantify the resemblance between document samples.

The researchers [9] acknowledge several challenges in this field, including the complexity of text recognition from diverse image types, the need for accurate OCR, and the intricacies of semantic analysis in plagiarism detection. They also note the potential for false positives and the importance of context consideration in plagiarism assessment.

Regarding state-of-the-art systems, the authors reference various approaches in text detection and recognition, including connected component-based and texture-based methods. They also mention the use of machine learning techniques in online plagiarism checking systems.

For future suggestions, the paper proposes developing a plagiarism system applicable to all kinds of images, not just specific types like flowcharts. The authors [9] suggest incorporating features such as color and gray scaling to improve accuracy. They also emphasize the need for further research in image-based plagiarism detection, as this area is currently underexplored compared to text-based methods.

Overall, this paper by Daksha D et al. presents a comprehensive overview of plagiarism checking using NLP and image text extraction. It contributes to the ongoing efforts in developing

more robust and inclusive strategies for ensuring academic and professional integrity in an increasingly digital world. The authors' work paves the way for future advancements in plagiarism detection technology, particularly in handling multimedia content.

References

[1] AlKendi, W.; Gechter, F.; Heyberger, L.; Guyeux, C.; Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey; J. Imaging 2024, 10, 18. https://doi.org/10.3390/jimaging10010018

[2] Bangbang Zhou , Yadong Qu , Zixiao Wang , Zicheng Li , Boqiang Zhang , Hongtao Xie; Focus on the Whole Character: Discriminative Character Modeling for Scene Text Recognition; arXiv:2407.05562v1 [cs.CV] 8 Jul 2024

[3]Bintang Timur Kusmanto, Nurcahya Pradana, Taufik Prakisya, Puspanda Hatta; Comparative Analysis of Google Vision OCR with Tesseract on Newspaper Text Recognition; 10.69616/mcs

[4] Dr. Bireshwar Ganguly, Yash Zade, Yash Bele , Siddhant Umare; Text Extraction from an Image; https://www.ijarsct.co.in/Paper11378.pdf

[5] Aanuoluwa O. Adio, Caleb O. Akanbi, Adepeju A Adigun, Abdulwakil Kasali, Solagbade Adisa; Assessing Selected CNN Models for Efficient Feature Extraction in SSD for Text Detection in Advertisement Images; https://doi.org/10.4314/dujopas.v10i2b.17

[6] Rutuja Krishna Lohar, Shreyas Sanjay Jadhav, Sammed Sunil Chougule, Ashitosh Shivaji Jagtap, Prof. P. D. Patil; Text Extraction From Image Using OCR Technology; https://www.jetir.org/papers/JETIR2405A50.pdf

[7] Prof. Anuradha Thorat, Mayur Zagade, Shivani More, Manish Pasalkar, Anand Narute; Research Paper on Text Extraction using OCR; https://www.ijarsct.co.in/Paper10793.pdf

[8] Ivan Malashin, Igor Masich, Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub and Aleksei Borodulin; Image Text Extraction and Natural Language Processing of Unstructured Data from Medical Reports; Mach. Learn. Knowl. Extr. 2024, 6, 1361–1377. https://doi.org/10.3390/make6020064

[9] Daksha D, Bhuvana K, N N Siri, Sanjana M; Plagiarism Checking with extracted text from image using NLP: A Comprehensive Overview; International Journal of Novel Research and Development (www.ijnrd.org), ISSN:2456-4184, Vol.9, Issue 4, page no.g238-g243, April-2024, Available :http://www.ijnrd.org/papers/IJNRD2404624.pdf

9. Text Detection and Recognition in X-ray Weld Seam Images(Mar 2024)

- **Objective**: Enhancing accuracy and efficiency in detecting and recognizing text in X-ray weld seam images using deep learning methods.
- **Methodology**: The DBNet algorithm for text detection and the MATRN model for text recognition were used, incorporating modules like CA, CBAM, and HFA.
- **Results**: Achieved high precision, recall, and accuracy, significantly improving automated weld seam information collection in industrial manufacturing.

10. Deep Learning in Text Recognition and Detection(Aug 2022)

- **Techniques**: Discusses the use of CNNs and other deep learning models for text detection and recognition, highlighting their effectiveness in handling complex tasks.
- **Processes**: Explains text detection (locating text in images) and text recognition (converting detected text into machine-readable format), emphasizing challenges and solutions provided by deep learning.
- **Datasets and Pre-processing**: Importance of datasets like MNIST, IAM, and others, along with pre-processing techniques to clean and normalize data for better model performance.
- **Challenges**: Outlines challenges such as background complexity, poor handwriting, and language identification, and how deep learning models address these issues.