CSE 4120
Technical Writing & Seminar

Handwritten Document Recognition

Presented By:

Sagar Dutta

Roll : 1907085

Department of CSE, KUET

Supervised by:

Dr. K. M. Azharul Hasan

Professor

Department of CSE, KUET

Sunanda Das

Assistant Professor

Department of CSE, KUET

Department of Computer Science and Engineering
Khulna University of Engineering & Technology

Source Details

Table 1: Source Details

Serial No	Title	Author	Source	Published Year
1	Improving Handwritten Mathematical Expression Recognition via Similar Symbol Distinguishing	Zhe Li , Xinyu Wang , Yuliang Liu , Lianwen Jin , Yichao Huang, and Kai Ding	IEEE TRANSACTIONS ON MULTIMEDIA Vol-26	2024
2	DAN: A Segmentation-Free Document Attention Network for Handwritten Document Recognition	Denis Coquenet , Clément Chatelain , and Thierry Paquet	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE VOL. 45, NO. 7	2023
3	End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network	Denis Coquenet , Clement Chatelain , and Thierry Paquet	IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE VOL. 45, NO. 1	2023

Outline

- ✓ Introduction
- ✓ Publication Details
- ✓ Literature Review
- ✓ Methodology
- ✓ Result Analysis
- ✓ Limitations
- ✓ Comparative Discussion
- ✓ Recommendation & Findings
- ✓ Conclusion
- ✓ References

Introduction

The Problem:

Handwritten documents in many Sector:-

- Form Processing
- Banking and Finance
- Postal Service
- Medical Records

Converting them manually in Digital is a big hazard.

The Solution:

Handwritten Document Recognition.

Introduction (Cont.)

What it is actually:

❑ The process of converting handwritten text or characters into digital form using various techniques

The Techniques:

- Image processing

- Pattern recognition

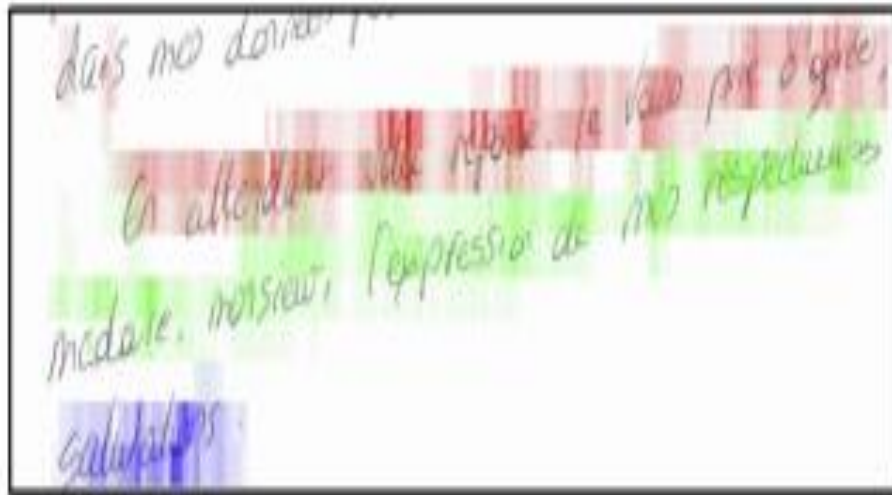- Machine learning algorithms

Introduction (Cont.)
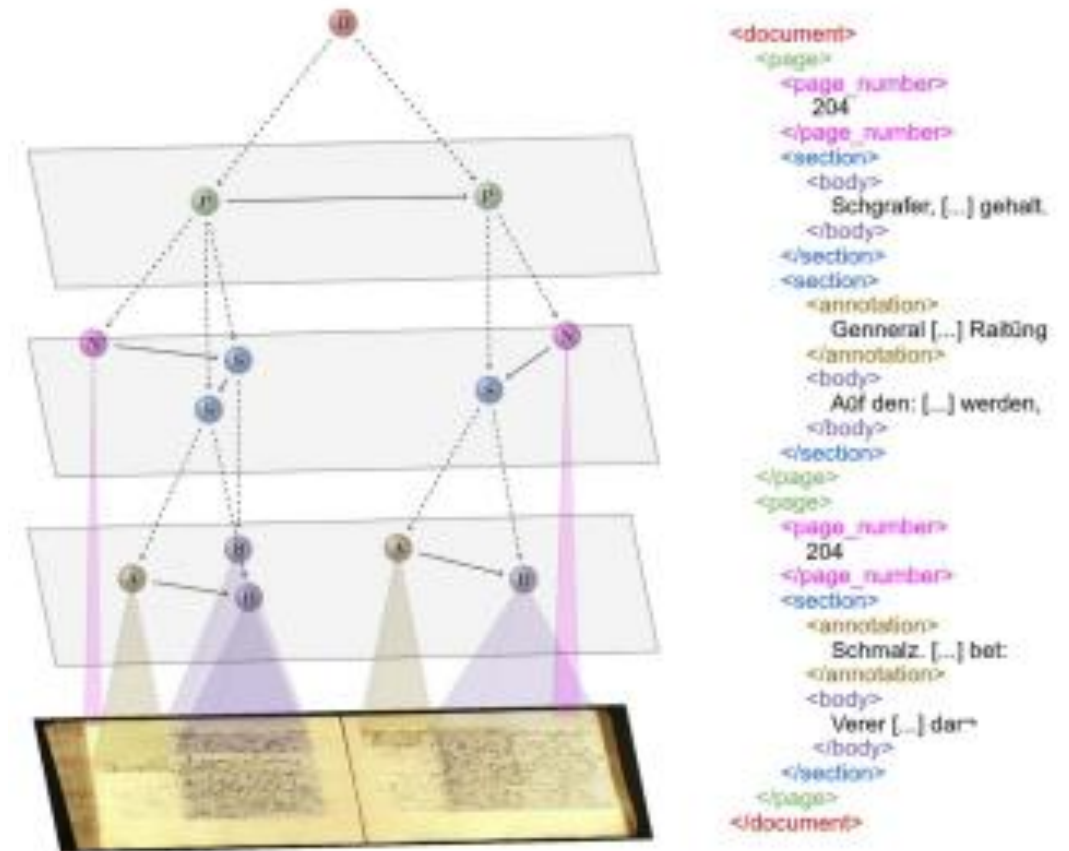

Figure 1: Visualization of input

Figure 2: Document image with associated layout Graph

Literature Review

➢ Handwritten document recognition research has evolved significantly, encompassing text and layout analysis. Traditional methods relied on grammar-based approaches, while recent advancements have seen a shift towards deep learning-based methodologies.

➢ These modern techniques leverage sequence and graph modeling, including encoder-decoder frameworks and attention mechanisms, to achieve superior performance. Additionally, ensemble methods have emerged as a strategy to enhance model robustness.

➢ The focus lies on developing end-to-end solutions that can accurately extract and understand textual and structural information from handwritten documents, marking a significant stride towards more efficient and effective document processing systems.

Literature Review

➢ Traditional methods, as highlighted across the surveyed papers, often relied on grammar-based frameworks, such as graph and context-free grammars, to dissect handwritten documents.

➢ Recent years have witnessed a surge in deep learning techniques, particularly sequence and graph modeling, which have revolutionized the field. These modern approaches, exemplified by encoder-decoder architectures, attention mechanisms, and graph neural networks.

➢ The focus lies on developing end-to-end solutions that can accurately extract and understand textual and structural information from handwritten documents, marking a significant stride towards more efficient and effective document processing systems.
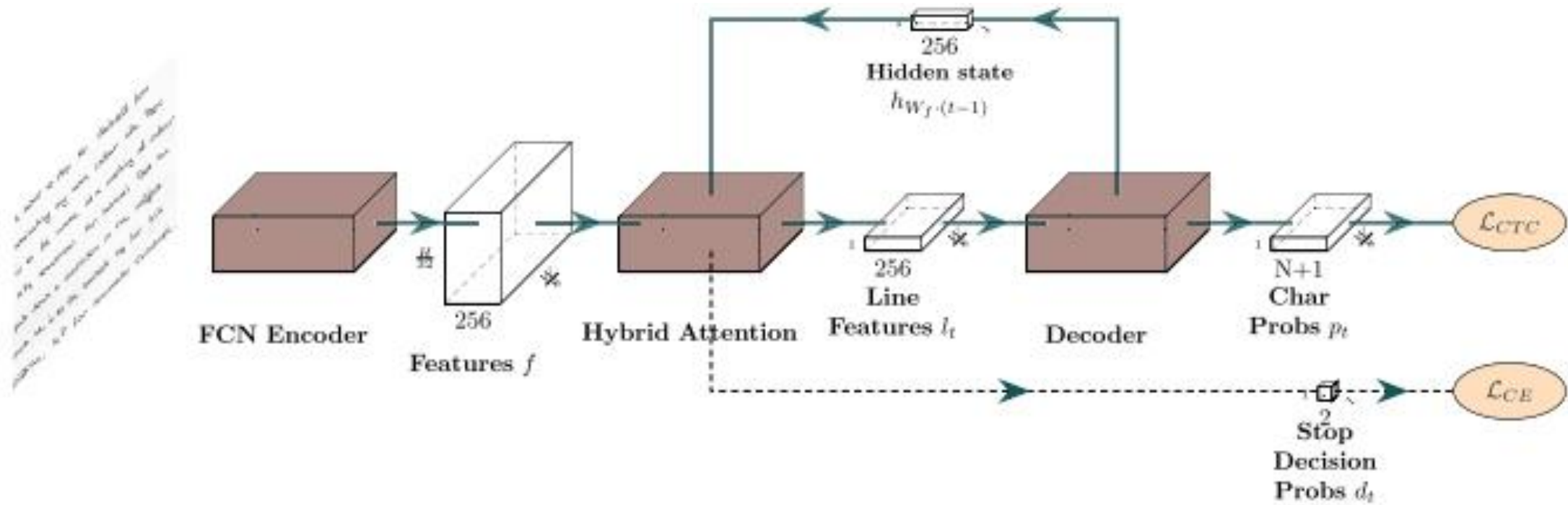
Methodology(Datasets)

Figure 3: Datasets

Methodology(Architecture)

Figure 4:Architecture Overview

Methodology Comparison

Table 3: Methodology

SL	Dataset	Preprocessing	Method
1.	CROHME 2014 Dataset	No preprocessing were done	• Method was implemented in PyTorch and optimized on 6 Nvidia GTX 1080Ti GPUs distributional • A 6-layer transformer encoder with eight heads was adopted • The learning rate was initialized to 0.1 and decreased by 10 times every 20,000 iterations
2.	RIMES, and READ2016 Handwritten Dataset	Contrast enhancement and Noise reduction were used	• Trained with 90% of synthetic documents. • Increased both the length of the target sequence, through the number of text lines, and the input image size
3.	The IAM-database	The encoder and the last convolutional layer of the decoder of the VAN are pretrained on line-level images	• Isolated text lines were avoided by using a pretrained VAN on another dataset • Different stopping strategies were implemented on the IAM dataset

Methodology Cont.

Table 4: Model Comparison

SL	Paper	Model
1.	Improving Handwritten Mathematical Expression Recognition via Similar Symbol Distinguishing	CNN based encoder and DTW-based algorithm ensembled
2.	DAN: A Segmentation-Free Document Attention Network for Handwritten Document Recognition	Document Attention Network (DAN), an endto-end encoder-decoder architecture that jointly recognizes both text and layout, from whole documents
3.	End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network	CNN+MDLSTMa , RPN+CNN+BLSTM+LM

Result Analysis

Table 5: Result Analysis

SL	Paper	Result
1.	Improving Handwritten Mathematical Expression Recognition via Similar Symbol Distinguishing	Achieves a superior ExpRate of 60.34%, 59.98%, and 64.22% with a single model on all three CROHME test splits
2.	DAN: A Segmentation-Free Document Attention Network for Handwritten Document Recognition	Reached very satisfying results at page level for both text and layout recognition with a CER of 4.54%, a WER of 11.85%, a LOER of 3.82% and a mAPCER of 93.74%
3.	End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network	Line level model achieve higher accuracy of 90.5% than all other model

Limitations

Table 6: Limitations

SL	Paper	Limits
1.	Improving Handwritten Mathematical Expression Recognition via Similar Symbol Distinguishing	• Lackings in Language Model Rectification • Can't predict implicit symbols
2.	DAN: A Segmentation-Free Document Attention Network for Handwritten Document Recognition	• Lack of large-scale public datasets of documents including both layout annotation and text annotation
3.	End-to-End Handwritten Paragraph Text Recognition Using a Vertical Attention Network	• Obtained results are very dependent on the quality of the synthetic data, as they must be close to the target dataset, notably in terms of layout

Comparative Discussion

Table 7: Discussion

Aspect	Paper 1	Paper 2	Paper 3
Key Contribution	End-To-End architecture for document-level recognition and layout labeling	Recognition of paragraph text without explicit segmentation	Implicit line segmentation for paragraph recognition
Approach	Unified statistical model, end-to-end architecture for data generation	Vertical attention mechanism, segmentation free neural network	Hybrid attention mechanism, implicit line segmentation, dropout strategies
Final Standing	Achieved state-of-the art levels at page and paragraph level	Achieved state-of-the art levels at paragraph level only	Achieved state-of-the art levels at page and paragraph level only

Recommendation and Findings

➢ Among the three papers, "DAN: A Segmentation-Free Document Attention. Network for Handwritten Document Recognition" this paper is found as the most promising

Reason behind:

- Better Accuracy-> 93.74%(highest)

- Potential for Document Understanding in a smooth way

- Innovative Training Strategy

- State-of-the-Art recognizing both text and layout

Conclusion

➢ These three papers collectively represent significant advancements in handwritten document recognition

➢ The Document Attention Network (DAN) introduced a groundbreaking segmentation-free architecture capable of recognizing text and layout information

➢ The Vertical Attention Network (VAN) proposed an end-to-end approach for paragraph text recognition

➢ These papers Offer efficient solutions that bridge the gap between text recognition and layout understanding.

References

➤ Z. Li, X. Wang, Y. Liu, L. Jin, Y. Huang, and K. Ding, "Improving handwritten mathematical expression recognition via similar symbol distinguishing," IEEE Transactions on Multimedia, 2023

➤ J. Chung and T. Delteil, "A computationally efficient pipeline approach to full page offline handwritten text recognition," in 2019 International conference on document analysis and recognition workshops (ICDARW), vol. 5. IEEE, 2019, pp. 35–40.

➤ P. Voigtlaender, P. Doetsch, and H. Ney, "Handwriting recognition with large multidimensional long short-term memory recurrent neural networks," in 2016 15th international conference on frontiers in handwriting recognition (ICFHR). IEEE, 2016, pp. 228–233.

References (Cont.)

> D. Coquenet, C. Chatelain, and T. Paquet, "Span: a simple predict & align network for handwritten paragraph recognition," in International Conference on Document Analysis and Recognition. Springer, 2021, pp. 70–84.

> D. Coquenet, C. Chatelain, and T. Paquet, "Dan: a segmentation-free document attention network for handwritten document recognition," IEEE transactions on pattern analysis and machine intelligence, 2023.

> End-to-end handwritten paragraph text recognition using a vertical attention network," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 508–524, 2022.

Thank You