Carnegie Mellon University

Digital Transformation: Exploring DBNet and CRNN for Handwritten Records Detection, Recognition and Rendering

Applied Computer Vision Final Project Presentation Spring 2024

Team

Group Members:

- Damilare OLATUNJI
- Atabonfack Lorier Bernes
- Dirac Murairi
- Tanghang Elvis Tata

Faculty Advisor: Prof Moise Busogi & Assane Gueye

Introduction:

- How is data record and stored?
- Could this records be transformed and stored in better format?
- How do we used computer vision to relate this problem?

Preamble: The reality in Africa

- **Undigitalized records:**
  - Health sector,
  - Education sector,
  - Government parastatals and
  - Cultural sector

15-35%

WHO: Adoption of Electronic Medical Records in Africans [1]

Carnegie Mellon University

Problem Statement / Motivation

- Handwritten records remain prevalent across various sectors in Africa [2][3].
- Therefore, digitizing these records is essential for enhancing:
  - Data organization,
  - Improving record-keeping, and
  - Enabling data-driven solutions to tackle developmental challenges.

Previous Work (Literature Review)

1. Text Detection

2. Text Recognition

Previous Work (Text Detection)

- [4] introduced a differentiable binarization (DB) within segmentation network for scene text detection. By integrating DB, the need for post-processing binarization was eliminated.

  Experimental results presented in terms of detection accuracy and speed, particularly notable with a lightweight ResNet-18 backbone

  F-measure of 82.8 at 62 FPS on the MSRA-TD500 dataset.

Previous Work (Text Detection)

- [5] presented fast (faster arbitrarily-shape text detector) to scene text detection, emphasizing both accuracy and efficient.

  Result: FAST-T achieves 81.6% F-measure at 152 FPS on Total-Text, outperforming previous methods in both accuracy and speed. With TensorRT optimization, inference speed can exceed 600 FPS.

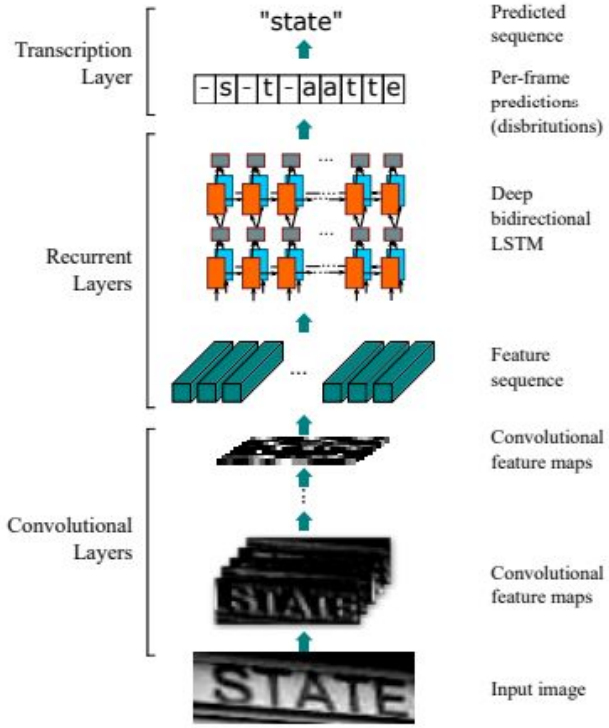
  Dataset: Total Text, CTW1500, ICDAR 2015, and MSRA-TD500.

  Reference: https://arxiv.org/pdf/2111.02394

Previous Work (Text Recognition)

- [3] used CRNN for scene text recognition, a challenging task in image-based sequence recognition. This architecture combines feature extraction, sequence modeling, and transcription into a unified framework.
- **Dataset:** IIIT-5K, Street View Text, and ICDAR datasets
- **Experimental result:**

	Clean	Synthesized	Real-World
Capella Scan [3]	51.9%/1.75	20.0%/2.31	43.5%/3.05
PhotoScore [4]	55.0%/2.34	28.0%/1.85	20.4%/3.00
CRNN	**74.6%/0.37**	**81.5%/0.30**	**84.0%/0.30**



Our Work ?

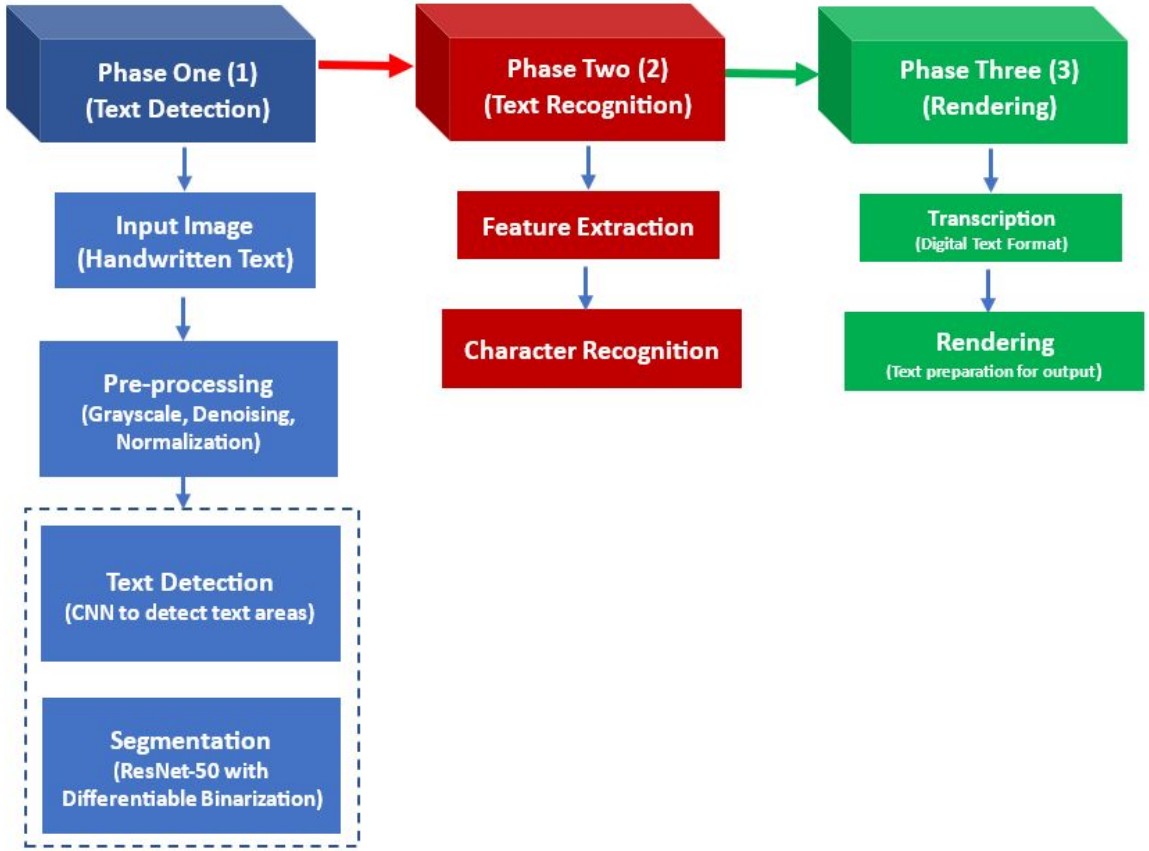
Data Composition

Dataset:
IAM Dataset for
Training Recognition

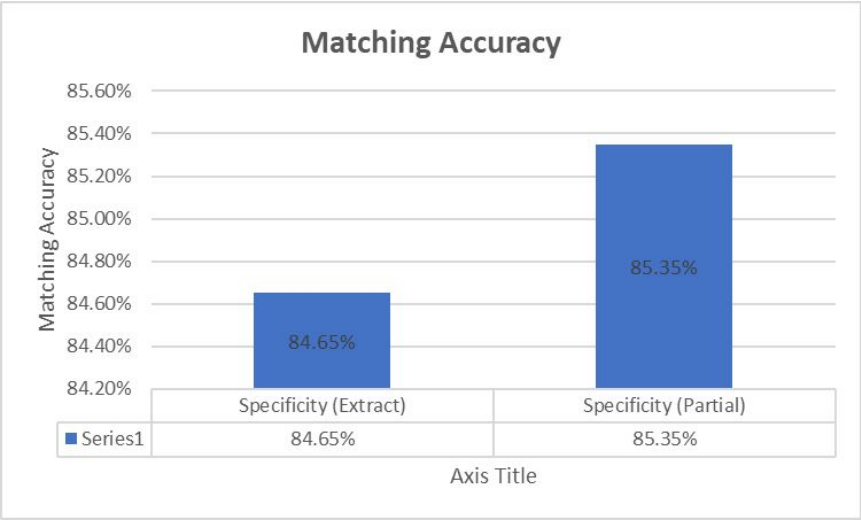
62,504
Training images

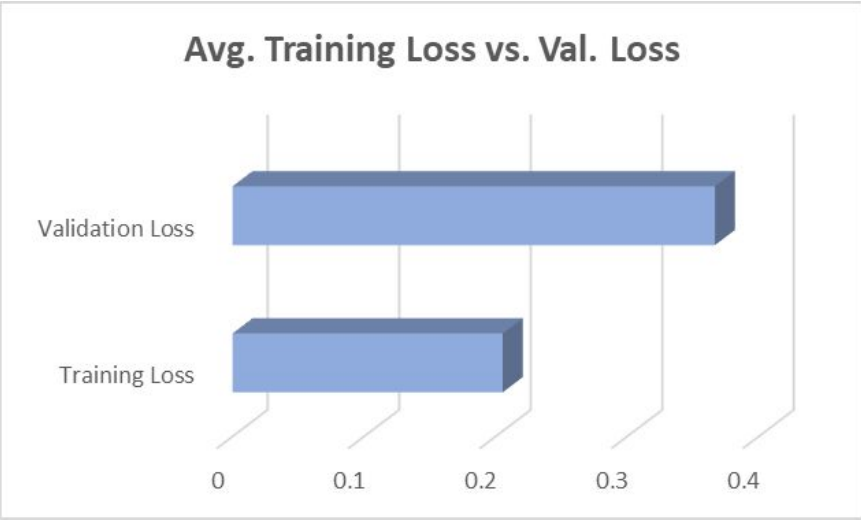
976
Validation images

Methodology



Evaluation Metrics



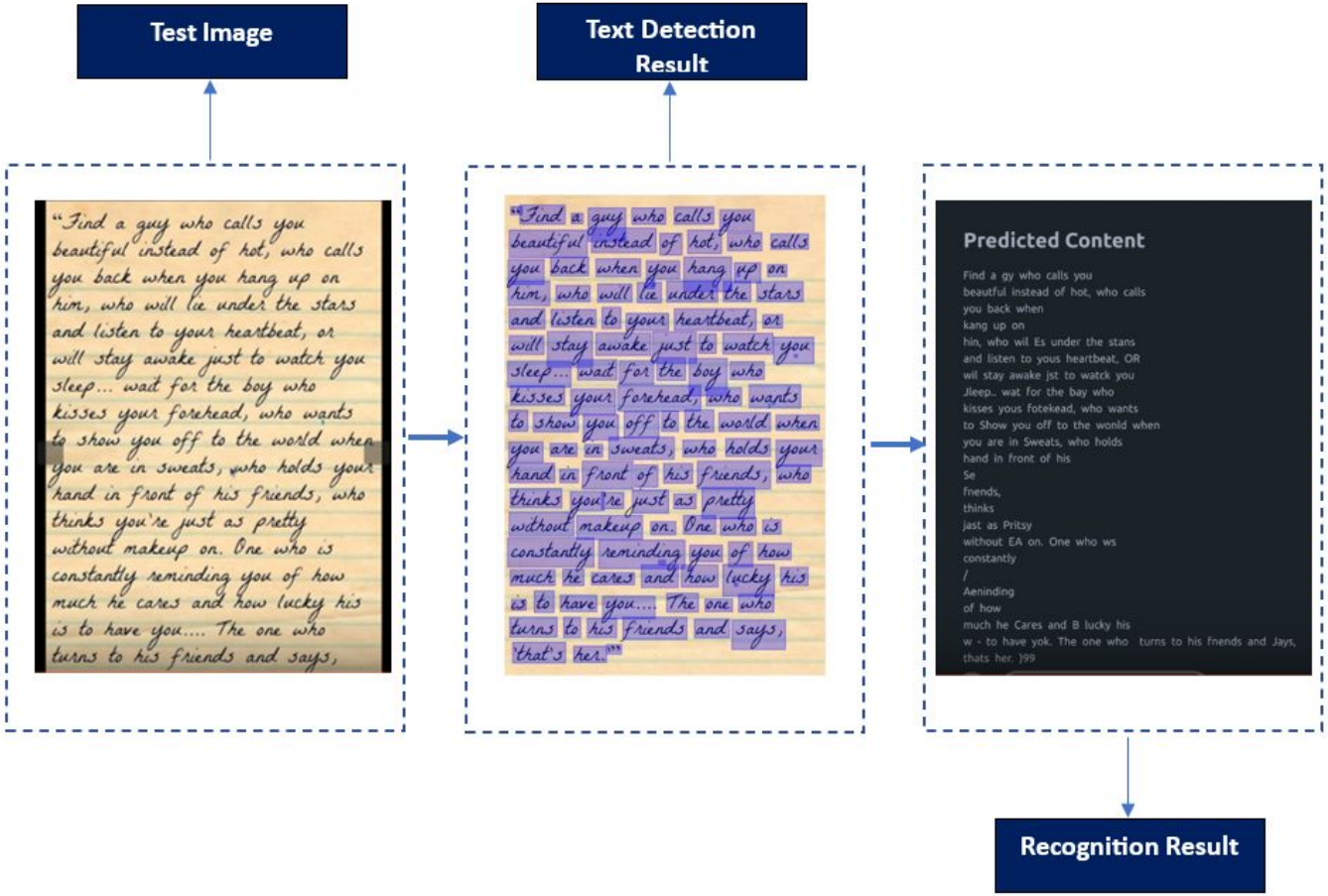


Dataset: IAM words dataset

Metrics: Text-Matching (Words matching)

Loss: Connectionist Temporal Classification Loss (CTC Loss)

Results (Visual Inspection)



Project Demo:

https://www.loom.com/share/e184c3b82b5f438692b6abf2d5a27d32?sid=847280e0-3f05-4e36-95ba-103b12d765ff

Achievements

- Text detection
- Text recognition
- Conversion into digital format

Future Work

- Multilingual Support (Process text in multiple languages.).
- Robustness to varied conditions (Different font styles, handwriting, backgrounds, contents - equations or table…)

Conclusion

- In Africa, the urgent demand for digitized records is crucial.
- Hence, handwriting detection and recognition technology can help narrow the digital divide and expedite digitization initiatives in the healthcare sector.

References

1. C. El Jabari, M. Macedo, and M. O. Al-jabari, “Towards a New Paradigm of Federated Electronic Health Records in Palestine,” Informatics, vol. 7, no. 4, p. 41, Oct. 2020, doi: https://doi.org/10.3390/informatics7040041.
2. Wikipedia Contributors, “Africa,” Wikipedia, Feb. 25, 2019. https://en.wikipedia.org/wiki/Africa
3. UNESCO, “Cultural heritage: 7 successes of UNESCO’s preservation work | UNESCO,” www.unesco.org, Jun. 16, 2022. https://www.unesco.org/en/cultural-heritage-7-successes-unescos-preservation-work
4. Journal Of L A T E X Class and Files, “FAST: Faster Arbitrarily-Shaped Text Detector with Minimalist Kernel Representation,” vol. 14, no. 8, 2021, Accessed: Mar. 18, 2024. [Online]. Available: https://arxiv.org/pdf/2111.02394.pdf
5. B. Shi, X. Bai, and C. Yao, “An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298–2304, Nov. 2017, doi: https://doi.org/10.1109/tpami.2016.2646371.
6. “doctr.models - docTR documentation,” *mindee.github.io*. https://mindee.github.io/doctr/latest/modules/models.html#doctr-models-detection (accessed Apr. 26, 2024).