Project Proposal: Handwritten Text Generation using Transformers

Ashvin Patidar
Hritvija Singh
Khush Gupta
Kshitij Bagga
Shashwat Agarwal

October 6, 2025

1 Introduction

Handwritten text generation has gained significant attention due to its applications in document digitization, personalized content creation, and digital signature synthesis. The primary goal of this project is to generate handwritten text from digital text input while preserving the style of a given handwriting sample. The span of the project extends from english to Indian languages like Hindi. The proposed project builds upon the **Handwriting Transformers** framework.

2 Related Work

2.1 Handwriting Transformers (ICCV 2021, IEEE)

This paper presents a transformer-based architecture designed for generating handwritten text conditioned on style examples. The model uses a sequence-to-sequence approach and achieves style consistency while producing realistic handwriting. It has been trained on the IAM dataset and demonstrates strong layout and word placement.

Link to the original paper : `https://openaccess.thecvf.com/content/ICCV2021/papers/Bhunia_Handwriting_Transformers_ICCV_2021_paper.pdf`

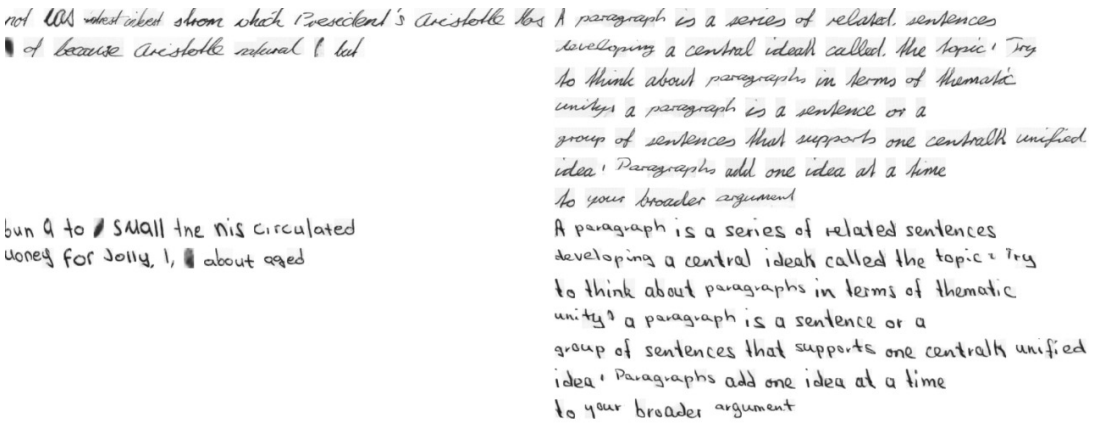

Figure 1: Sample results from Handwriting Transformers.

2.2 WriteViT: Handwritten Text Generation with Vision Transformer (Preprint)

WriteViT proposes a Vision Transformer-based GAN framework to improve handwritten text generation, especially for languages with diacritical marks like Vietnamese. It introduces attention-based mechanisms to better capture style and structure in handwriting. Although currently a preprint, its ideas could inspire extensions to our project. Link to the preprint paper : `https://arxiv.org/abs/2505.13235`

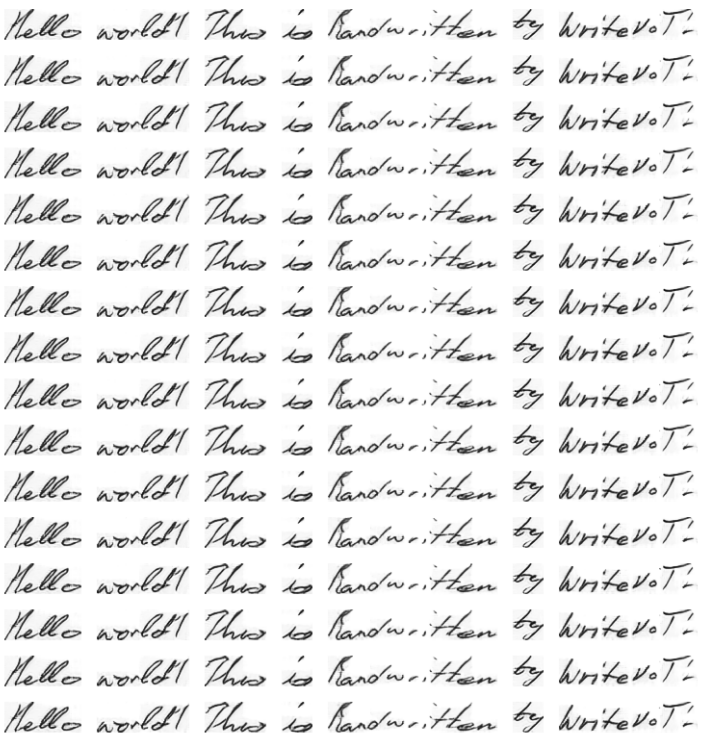

Figure 2: Sample results from WriteViT.

3 Repository and Dataset

The source code and dataset for the Handwriting Transformers paper are publicly available. We successfully cloned the repository, examined the structure, and built the environment. All code modules, dataset scripts, and model checkpoints are accessible and were used for our replication study.

4 Reproduction of Results

We reproduced the experiments of the original paper on a small subset of the IAM dataset. While the original model was trained for **100,000 epochs**, our replication was limited to **500 epochs** considering the computational resource limitations and model training time. Consequently, the generated images do not produce fully readable words, but the *word size and placement* closely match the outputs from the original model.

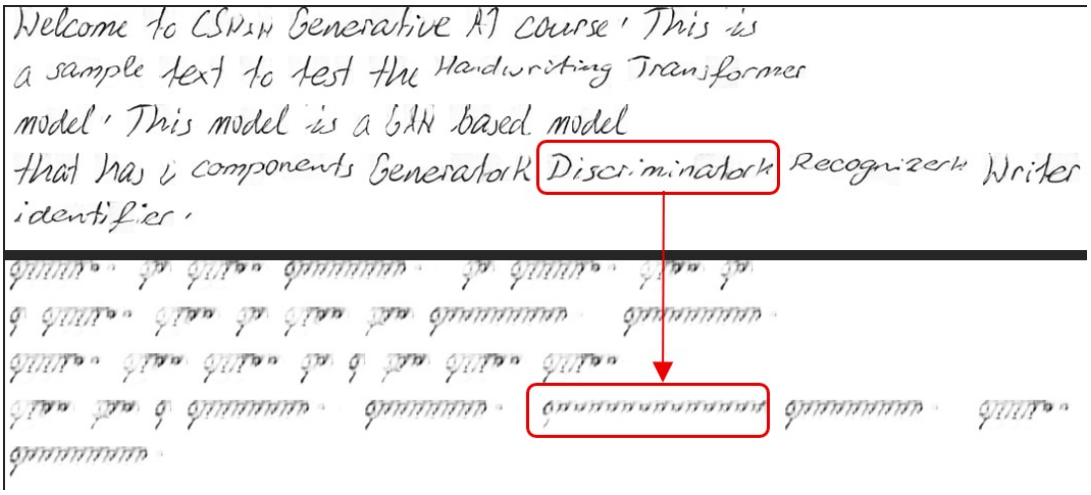

Figure 3: Generated image comparisons between 500 epochs model and provided model.

4.1 Evaluation Metrics

We evaluated both the original model and our reproduced model using **SSIM** (Structural Similarity Index Measure) and **MSE** (Mean Squared Error):

Model	SSIM	MSE
Original Paper	0.8195	0.017447
Our Reproduction	0.8446	0.011253

Table 1: Comparison of evaluation metrics between the original model and our reproduction.

Interpretation:

- SSIM: Measures perceptual similarity between generated and reference images. Our model achieved slightly higher SSIM, indicating that the word layout, spacing, and overall style closely match the reference handwriting.
- MSE: Measures pixel-level differences between images. Our lower MSE value indicates that our generated images are numerically closer to the style reference, though readability is not fully achieved due to fewer epochs.

4.2 Generated Results

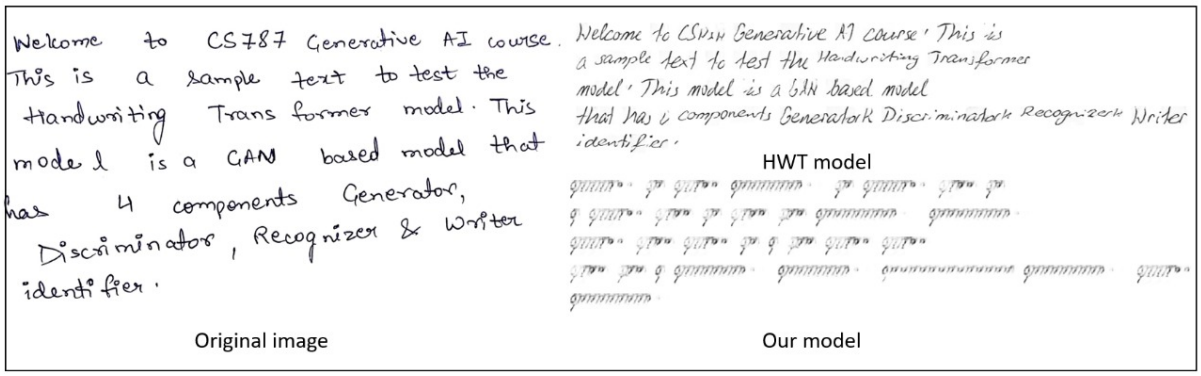

Figure 4: Generated handwriting image from our model.

5 Proposed Project Plan

Our project aims to develop a system where a user can input:

- Digital text that they want to generate.
- A handwriting sample to serve as the style reference.

The model will then generate the input text in the style of the provided handwriting sample at runtime. Current models extract text from the input handwritten text image using OCR and then generate the output based on the learned handwritings and doesn't copy the handwritings from the input image.

5.1 Handwriting Features to Consider for Training

To ensure realistic and personalized handwriting generation, we plan to focus on the following features:

- **Letter Form:** Curves, slants, proportional sizes of letters, slope, and linking lines between letters. The model should consider how letters appear in different positions within a word.

- **Line Form:** Smoothness, darkness, and pressure of lines, which indicate writing speed and pressure applied.
- **Formatting:** Spacing between letters, words, line placement, and margins. Also, how strokes from one line interact with strokes from lines above and below.

6 Conclusion

The preliminary results demonstrate that the Handwriting Transformers framework can be replicated successfully. The proposed project will enable user-driven handwritten text generation while incorporating detailed handwriting features for improved style consistency and realism.