Handwriting Generation and Animation with Deep Learning
-Midterm Report-

Anita Dash
MA20BTECH11001

Dhruv Srikanth
EE20BTECH11014

Shambhu Prasad Kavir
CS20BTECH11045

Taha Adeel Mohammed
CS20BTECH11052

Abstract

Handwriting style transfer for personalized text generation is a fascinating and evolving field that sits at the intersection of computer vision and natural language processing. It has several applications in various domains, ranging from enhancing digital communication and marketing to improving accessibility and education.

In this project, we aim to create a system that takes a written text prompt and a sample of the handwriting to mimic as the input, and generates realistic animated handwritten text that closely resembles the provided style while maintaining readability and coherence.

1. Introduction

Handwriting is a deeply personal and authentic form of expression. It holds a unique place in human communication and self-expression. It is also an interesting and challenging problem to work with for DL models. There has been extensive research in this field, both very recently, and in the past. There exist various state-of-the-art models solving the problems of handwriting recognition, handwrting classification, handwriting generation, handwriting style transfer, and handwriting trajectory recovery. In this project, we aim to combine these state-of-the-art techniques to create a model that can generate realistic animated handwritten text in a particular style.

2. Problem Statement

Implement a deep learning model that can take a text prompt and a sample image of a person's handwriting style as input, and generate the image of the text prompt, in the specified handwriting style as the output. We want the model to accurately mimic the unique characteristics, nuances, and idiosyncrasies of the provided handwriting style while maintaining the readability and coherence of the generated text. We could consider further diversification - calculating the temporal sequence of brush strokes: if our generation model is successful, we could try to calculate the pen trajectory, i.e. the successive coordinate points for every time step of the brush stroke, to be extracted. This trajectory can then be used to animate the brush stroke in real-time, creating a realistic animated handwritten text in the given style.

3. Literature Review

Preliminary Report Literature Review

3.1. Decoupled Style Descriptors [8]

3.2. GANwriting [4, 7]

3.3. Handwriting Transformers [3]

3.4. Dynamic CRNN (Recognizer) [1]

3.5. Handwriting Trajectory Recovery [2]

3.6. Domain-Adversarial Neural Network [6]

,

Midterm Report Literature Review

3.7. Stroke Separation & Writing Order Recovery

3.8. Writing Order Recovery [5]

Writing order recovery is a complex problem that relates to the intrinsic properties of human handwriting. However, integrating its solution with an observed handwritten text could tell us more about a subject's characteristic handwriting than ever before. This paper [5] proposes an innovative deterministic algorithm to recover the writing order of any thinned long static handwritten signature. Signatures have been chosen out of the belief that they are the most complex version of this problem.

The proposed method is completely intuitive and draws from the good continuity criteria of handwriting. The process of recovery has been split into 3 subprocesses - point classification, local examination, and global reconstruction. Strokes are believed to be entirely continuous so the authors have observed the 8-connected pixels, adjacent pixels surrounding a target pixel, to observe the target pixel's position within a stroke. A pixel with 2 connected pixels is believed to be a trace-point, one of the points along the trajectory of a stroke. A point with only 1 connected pixel is believed to be an end-point of a stroke. Any point with more than 2 connected pixels is believed to be a cluster point, a point found in the region of overlap between 2 distinct strokes/components. After this classification, the next step is described as local examination. Adjacent trace points are considered to be part of one stroke and hence form large groups reconstructing the strokes from an end-point, through trace-points, to another end-point. The only complexity remaining at this stage is that of overlapping strokes forming clusters. Branches exiting clusters are marked with anchor points and their exit angles are measured and characterized. The authors define a few commonly occurring cluster scenarios and match the present observed cluster to the predefined scenarios. In the global reconstruction stage, clusters are resolved by modelling the rapid change in direction (from branch angle and position) as energy alongside assigning priority to each branch and each scenario to perform energy minimization. After the internal cluster paths are separated, the authors use a Gaussian spread formulation to choose the leftmost starting point of a stroke and use a proximity criterion to connect a pen-up point to the next pen-down point thus deriving the order of strokes/components.

The main complexity of such a problem has been described as the interpolation of pen-up and pen-down amidst strokes causing distinct strokes and consequently their overlap. The solution proposed to this problem is intuitive and replicative of a human thought process which is why it captured our attention. Obtaining the temporal properties has been left as a future problem that could, in their opinion, aid more significantly in the field of handwriting recognition, analysis, and generation.

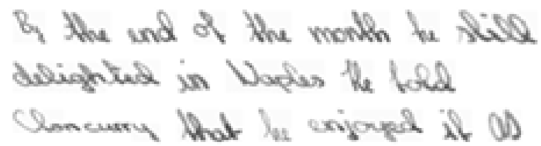
3.9. DL paper on trajectory recovery

4. Our approach

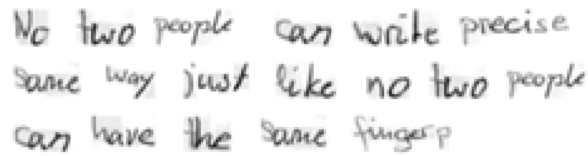
5. Architecture Overview

6. Replicated Results

GANwriting [7] is limited to generating singular words, and as a result we cannot have outputs of variable length. This is a major drawback as we want to generate entire lines of text. It doesn't encode style content entanglement at a character level and hence struggles to mimic character specific styles.



(a) **Desired Handwriting**

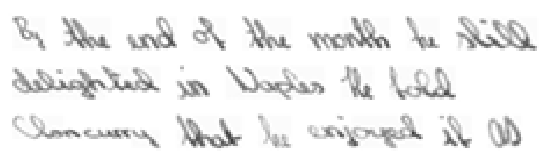


(b) **GANwriting**

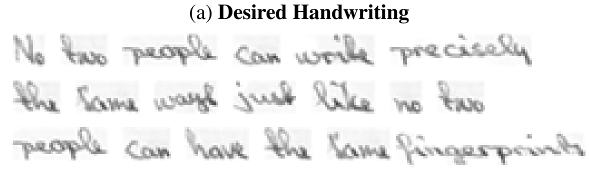
Figure 1. Since GANwriting is limited to generating fixed number of words, it fails to generate the entire line of text. It also fails to capture the style of the text.

While Decoupled Style Descriptors [8] is able to capture local and global style patterns, it fails to generate legible letters or connect cursive letters. This is because of the underlying inconsistencies in human writing, which is only partially addressed in this model. Additionally, the model works with online handwriting data and we want to capture the styles of offline handwriting.

Handwriting Transformers [3] is able to generate realistic styled handwritten text images and significantly outperforms other state-of-the-art models. It handles variable length input and captures local and global style patterns. It is also compatible with offline handwriting data. Hence we have decided to use this model as our base model.



(a) **Desired Handwriting**



(b) **HWT**

Figure 2. HWT is able to generate the entire line of text. It also captures the local and global style of the text.

6.1. FID Scores

The Fréchet Inception Distance (FID) is a metric used to evaluate the quality of generated images. It is based on the Inception Score (IS) and the Frechet Distance (FD).It is a metric that measures the similarity between two distributions. The FID score is the Frechet Distance between the distribution of the real images and the distribution of the generated images. The lower the FID score, the better the quality of the generated images. A score close to zero implies that the two groups of images are nearly identical.

Table 1: Comparison of HWT with GANwriting model with respect to their FID scores computed between the generated text images and the real text images from the IAM dataset. We generate datasets of 6 images each for both the models. The FID score for the HWT model is lower than that of the GANwriting model implying that the former generates better quality images than the latter.

Model	FID Score
HWT	70.67
GANwriting	94.77

While the size of the datasets used here was rather small, we can still observe that the HWT model generates better quality images than the GANwriting model. This is because the HWT model is able to capture the local and global style patterns of the text, while the GANwriting model is not. The purpose of this test was only to attest what was already proven in the papers. The code for the above test can be found here.

Table 2: Comparison of HWT model with Scrabble-GAN and Davis et al with respect to their FID scores. We generate a dataset of 10,000 images for the HWT model and then compute its FID score. For the other models, we take the FID scores as mentioned in the paper [3]. Once again, we observe that the HWT model has a lower FID score than the other models.

Model	FID Score
HWT	16.71
ScrabbleGAN	20.72
Davis et al	20.65

Since we have used a significantly smaller dataset for computing the FID score for the HWT model, the results may not be entirely accurate. However, we can still observe that the HWT model has a lower FID score than the other models. The code for the above test can be found here.

6.2. Testing

Add results

7. Preliminary Results

Testing with our own handwriting
Add changes made to make testing more accessibile

References

[1] Ahmed AL-Saffar, Suryanti Awang, Wafaa AL-Saiagh, Ahmed Salih AL-Khaleefa, and Saad Adnan Abed. A sequential handwriting recognition model based on a dynamically configurable crnn. *Sensors*, 21(21), 2021.

[2] Ayan Kumar Bhunia, Abir Bhowmick, Ankan Kumar Bhunia, Aishik Konwer, Prithaj Banerjee, Partha Pratim Roy, and Umapada Pal. Handwriting trajectory recovery using end-to-end deep encoder-decoder network. pages 3639–3644, 2018.

[3] Ankan Kumar Bhunia, Salman H. Khan, Hisham Cholakkal, Rao Muhammad Anwer, Fahad Shahbaz Khan, and Mubarak Shah. Handwriting transformers. *CoRR*, abs/2104.03964, 2021.

[4] Brian L. Davis, Chris Tensmeyer, Brian L. Price, Curtis Wigington, Bryan S. Morse, and Rajiv Jain. Text and style conditioned GAN for generation of offline handwriting lines. *CoRR*, abs/2009.00678, 2020.

[5] Moises Diaz, Gioele Crispo, Antonio Parziale, Angelo Marcelli, and Miguel A. Ferrer. Writing order recovery in complex and long static handwriting. *International Journal of Interactive Multimedia and Artificial Intelligence*, X(X):1–14, 2021.

[6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. 2016.

[7] Lei Kang, Pau Riba, Yaxing Wang, Marçal Rusiñol, Alicia Fornés, and Mauricio Villegas. Ganwriting: Content-conditioned generation of styled handwritten word images. 2020.

[8] Atsunobu Kotani, Stefanie Tellex, and James Tompkin. Generating handwriting via decouple style descriptors. *CoRR*, abs/2008.11354, 2020.

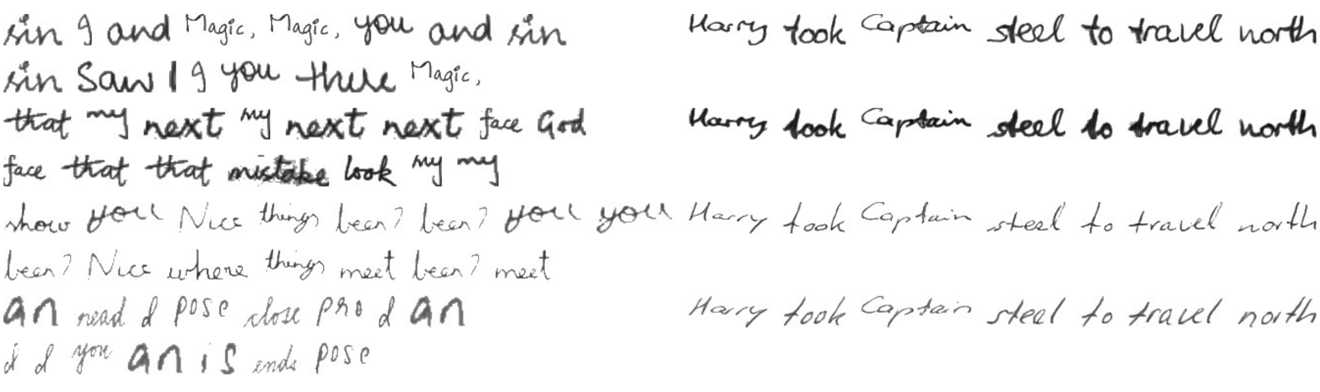


Figure 3. Your image caption

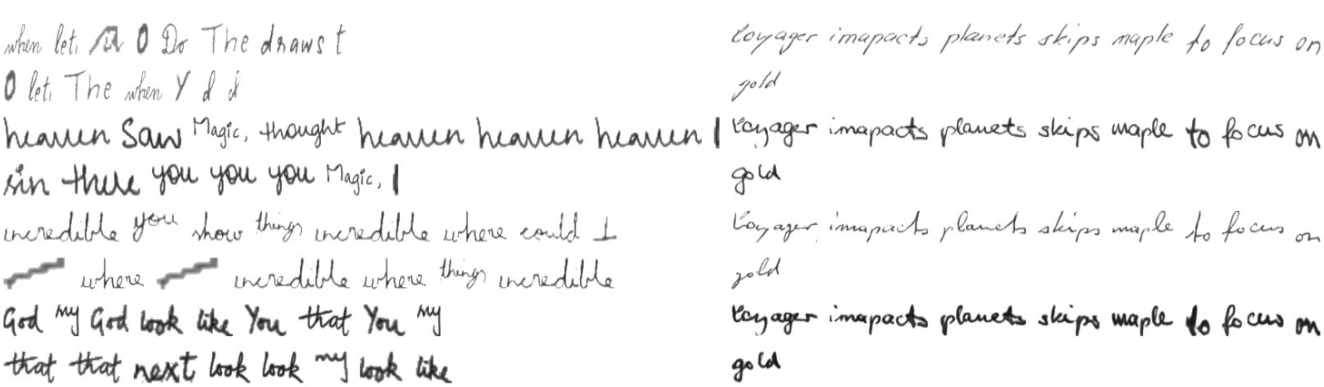


Figure 4. Your image caption