

A Survey of Modern Handwriting Generation Models

Thrishaa J , Agamy David , Neha Venkatesh , Dr. Kiran Y C

¹ Student, Dept. of Information Science and Engineering, Global Academy of Technology, Bangalore, India

² Student, Dept. of Information Science and Engineering, Global Academy of Technology, Bangalore, India

³ Student, Dept. of Information Science and Engineering, Global Academy of Technology, Bangalore, India

⁴ Head of Department , Dept Of Information Science and Engineering , Global Academy of Technology, Bangalore , India

Abstract

Advances in handwriting generation techniques through deep learning have allowed for automatic reproduction of an individual's unique handwriting style with little input data. This paper examines the most current models used in handwriting generation including: GAN, Transformer, VAE, Diffusion and others together with their relative strengths, weaknesses and how they are developing toward the goals of one-shot and zero-shot personalization. The Emuru architecture, which utilizes a VAE style encoder to provide a user-specific writing experience by combining it with an autoregressive Transformer model, serves as a practical option for creating a customized handwriting output based upon one example. The InkPersona system is based on the Emuru model and provides a comprehensive overview of the challenges and future opportunities for real-time generation of highly personalized handwriting.

Index Terms—Personalized handwriting, handwriting synthesis, few-shot learning, zero-shot learning, Emuru, variational autoencoder, transformer, diffusion models

1. INTRODUCTION

In our current digital age, the majority of both professional and personal communication will involve typed text. However, writing remains an intrinsically personal and emotional medium to convey individuality, emotion, and authenticity that typed text simply cannot. Whether it be in a greeting card, a signature, or personal notes, writing style carries an emotional element to communicate deeper, more personal messages and enhance human interaction.

Handwritten communication, though expressive, has suffered due to faster, more efficient digital alternatives, such as emails, forms, and online communication platforms. Today's digital handwriting emulation tools commonly offer limited handwriting fonts or templates that attempt to recreate customary segmentation, slant, and other stylistic variations found among individual writers' handwriting but miss these more nuanced variations. Therefore, a primary challenge of research is to develop methodologies

to recreate a person's handwriting from minimal data while retaining the unique stylistic attributes of that individual.

New advancements in deep generative modelling, including Generative Adversarial Networks (GANs), Transformers, Variational Autoencoders (VAEs), and Diffusion Models, have led to significant improvements in the ability to learn visual and stylistic patterns with very little data. These advancements result in systems that can create personalized handwriting in a zero- or one-shot manner with only a single image input.

Research Objective: This work presents a survey of the current landscape of personalized handwriting synthesis, as well as the introduction of InkPersona, a web-based implementation of the Emuru model. InkPersona leverages a VAE encoder to code a user's handwriting sample and, using a Transformer decoder, generates text authored by the user's handwriting style. InkPersona works with both typed and spoken text input and generates handwriting images that are independent of background content.

Contributions:

- A thorough review of handwriting synthesis techniques that span GAN, Transformer, VAE, and Diffusion methods.
- Technical characterization of the Emuru architecture for handwriting style transfer in zero-shot settings.
- Setup of a clear and interactive architecture (InkPersona) for real-time letterform generation.
- Identification of open research questions, ethical concerns, and implications for future work and research.

2. RELATED WORK

Research in handwriting synthesis has progressed through three principal generations of techniques: 1. GAN-based multi-sample synthesis, 2. Transformer and VAE-based few-shot approaches, and 3. Diffusion and zero-shot synthesis methods.

A. Early GAN-based Handwriting Generation

The initial phase of handwritten text generation (HTG) research was largely driven by the use of Generative Adversarial Networks (GANs). One of the pioneering works, GANwriting [1], explored content-conditioned handwriting generation by jointly encoding textual information and writer specific style representations obtained from several handwriting samples. Later, ScrabbleGAN [2] expanded this approach to handle words of varying lengths, though it offered only limited control over stylistic diversity. While these GAN-based methods produced handwriting that appeared realistic, they depended heavily on large multi-sample datasets and faced challenges with unstable adversarial training, which made them less practical for real world use.

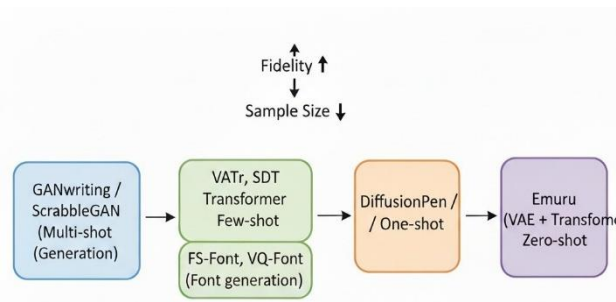


Fig. 1. Evolution of handwriting synthesis techniques: GAN-based multi-shot → Transformer/VAE few-shot → Diffusion and Emuru-based zero-shot personalization.

B. Transformer-Era Advances

The introduction of the Transformer architecture [3] was a pivotal moment in handwriting synthesis research, enabling more robust representation of long-range dependencies between the text content and style features. The Handwriting Transformer (HWT) and VTr [4] both created style embeddings from a convolutional neural network (CNN) encoder that tailored the Transformer decoder at generation time. More recently, VTr++ [5] improved the stability of training, improved dataset balance, and allowed for greater zero-shot generalization with fewer references.

C. Style Disentanglement and Local Aggregation

Subsequent research began to emphasize the difference between representations of handwriting style in global and local contexts. The Style Disentangled Transformer (SDT) [6] distinguished global, writer-level features (e.g. stroke slant, and line thickness) from local, glyph-level features like curvature and spacing. Building on this, FS-Font [7] introduced a cross attention-based Style Aggregation Module to reach a better representation of style by considering the context of the glyphs above. The authors in [13] leveraged stylistic markers, whereas DS-Font [8] utilized contrastive learning to refine the full style shape manifold. Recently, VQ-Font [9] incorporated vector quantization into the style encoding process, removing all human annotation and seamlessly representing the compositional elements of characters.

D. VAEs as Robust Style Encoders

Variational Autoencoders (VAEs) [10] have shown great ability to encode handwriting into a structured latent space, denoising input data while preserving the geometric properties of individual strokes. Advancing this idea, the Emuru model [11] grows the VAE framework to learn writer-specific latent representations from a single sample of handwriting, which makes it a very promising model for zero-shot handwriting generation.

E. Autoregressive Generation and Emuru

The Emuru model includes a style encoder that utilizes VAE as well as an autoregressive Transformer decoder, which supports generating handwriting latent sequences conditioned on style and text. This model supports:

- Zero-shot handwriting personalization from a single example.

- Generation of handwriting with variable lengths without retraining.
- Faster inference times compared to diffusion-based models.

F. Diffusion Models: Fidelity and One-Shot Advances

Recently, diffusion-based methods have set a new benchmark for image quality. Earlier models like CTIG-DM and WordStylist [12], [13] relied on conditional diffusion but needed several style samples to work effectively. To address this, DiffusionPen [14] introduced a metric-learning style encoder, which made the model better at handling different styles. Later, One-DM [15] pushed things further by using Laplacian contrastive losses, allowing it to capture even the smallest stroke details and produce high-quality images from just a single example.

G. Positioning InkPersona

InkPersona is a one-shot handwriting synthesis system based on hybrid VAE-Transformer architecture. Built independently of large training data, it can provide synthetic handwriting based on as little as one handwriting exemplar.

Although generative diffusion models might produce slightly higher quality in terms of visible marks, InkPersona stands out in practice through its accessibility and higher speed, making it practical and efficient for real-world personalization scenarios.

3. PROPOSED SYSTEM

InkPersona is a handwriting generation technique that takes advantage of recent advancements to provide a solution that is both powerful and easy to use. Its framework is built on three core components that capture and reproduce a person’s handwriting style accurately and flexibly.

TABLE I

COMPARISON OF HANDWRITING SYNTHESIS MODELS

Model	Samples		Fidelity		
	Style Control	Latency	Multi-shot	Medium	Low
GANwriting	Moderate	Fast	Medium	Low	Fast
ScrabbleGAN	Multi-shot	Medium	High	High	Moderate
VATr	Few-shot	High	High	High	Moderate
VATr++	shot	Very	High	High	Moderate
DiffusionPen	Few-shot	High	Very	High	Slow
One-DM	shot	High	Very	High	Slow
Emuru	One-shot	High	High	High	Fast
	One-shot	High	High	High	
	Zero-shot	High	High	High	

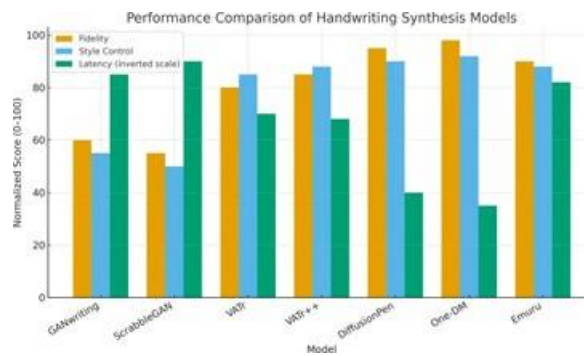


Fig. 2. Comparison of performance among handwriting generation models based on three performance criteria: fidelity, style control, and latency. Emuru is a strong middle ground between visual fidelity and slower generation times than the diffusion-based models.

A. Style Acquisition

The first step consists of the user uploading a handwritten image sample. The handwritten image will go through pre- processing steps to render a noise-free background, normalize the height, and then slice into slices based on the characters. The sample will be sent to an encoder based on Variational Autoencoder (VAE) that produces a latent representation of the original handwriting that captures the writer’s style information in compressed form.

B. Content Encoding

The system will now process the text - either written or verbal - that the user will provide. The user text will be tokenized and converted into embeddings with which the system will connect with the necessary visual archetypes of characters. This process allows the system to accept user type inputs in various modalities with semantic and stylistic consistency.

C. Handwriting Generation

To conclude the Transformer decoder generates latent handwriting sequences in an autoregressive fashion, conditioned on the textual input and the extracted handwriting style. The VAE plug-in decoder then reconstructs latent sequences back into image space, resulting in synthetic handwriting that represents the individual’s penmanship style.

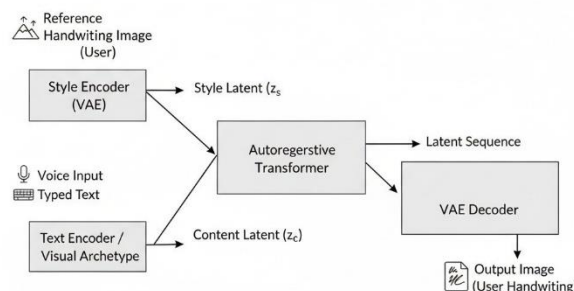


Fig. 3. Overview of the InkPersona architecture based on the Emuru model. The VAE encoder captures handwriting style from a single sample image, while the Transformer decoder generates handwriting conditioned on textual or spoken input.

IV. DISCUSSION AND FUTURE WORK

InkPersona achieves a balance between realism and computational viability in handwriting synthesis. While diffusion models can provide high-quality results, their high computational complexity and slow inference time make them impractical for real-time or interactive applications. The Emuru based VAE-Transformer design, on the other hand, represents a more pragmatic compromise, providing high-quality handwriting generation while also being amenable to fast, user-accessible execution.

Future research directions include:

- Increasing versatility for cursive styles and multiple writing language styles.
- Reducing model latency to enable smoother real-time rendering.
- Increasing robustness of model to handwriting samples of different qualities and orientations.
- Embedding digital forensic watermarks to delineate between generated handwriting and handwritten text.

V. CONCLUSION

This article provided a robust overview of developments in handwriting technology in the context of handwritten words produced by machine-generated methods like GAN, Transformer, VAE, and Diffusion. We introduced **InkPersona**, an efficient, easy-to-use, zero-shot handwriting generation system, utilizing the Emuru architecture, as a feasible solution for handwriting generation for personalization. In combining the benefits of VAE encoder with the Transformer decoder, InkPersona is capable of generating convincingly written output in the style of a user, based solely on one example.

Our design is intended to help the industry move towards a better relationship between emotional expressiveness and digital communication in an age of automation.

REFERENCES

1. F. Alonso et al., “GANwriting: Content-Conditioned Generation of Styled Handwritten Text,” ECCV, 2020.
2. R. Fogel et al., “ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation,” CVPR, 2020.
3. D. Kang et al., “Handwriting Transformer: Transformer-based Handwritten Text Generation,” arXiv preprint arXiv:2201.XXXX, 2022.
4. B. Vanherle et al., “VATr: Visual Archetypes Transformer for Handwritten Text Generation,” WACV, 2023.
5. B. Vanherle et al., “VATr++: Choose Your Words Wisely for Handwritten Text Generation,” arXiv preprint arXiv:2406.08130, 2024.
6. Y. Li et al., “Style-Disentangled Transformer for Handwriting,” ICCV, 2023.
7. C. Zhu et al., “FS-Font: Few-Shot Font Generation,” CVPR, 2022.



8. H. Jiang et al., “DS-Font: Disentangled Style Representation for Few- Shot Font Generation,” CVPR, 2022.
9. W. Park et al., “VQ-Font: Few-Shot Font Generation with Vector Quantization,” NeurIPS, 2022.
10. D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” ICLR, 2014.
11. V. Pippi et al., “Zero-shot styled text image generation, but make it autoregressive,” arXiv preprint arXiv:2503.17074, 2025.
12. A. Author et al., “CTIG-DM: Conditional Text Image Generation with Diffusion Models,” 2023.
13. K. Author et al., “WordStylist: Latent Diffusion for Handwriting Syn- thesis,” 2023.
14. K. Nikolaidou et al., “DiffusionPen: Controlling the Style of Handwrit- ten Text Generation,” ECCV, 2024.
15. X. Author et al., “One-DM: One-shot Latent Diffusion for Styled Text Generation,” 2024.