# Font Style Transfer with Conditional Variational Autoencoders

**2025 Spring ECE285 Project**
Kejia Ruan Hongjie Wang
Department of Electrical and Computer Engineering
University of California, San Diego
{keruan, jackwang}@ucsd.edu

## 1 Problem

We aim to re-implement a Conditional Variational Autoencoder (CVAE) architecture for font style transfer, focusing on generating diverse font styles for given characters. Existing font generation work often reuses standard datasets (e.g., EMNIST, VGG-Font) or relies on GAN-based models, while CVAE-based font synthesis remains relatively under-explored.

The innovation of our project lies in:

- Developing a CVAE model from scratch without using any pre-existing implementation.
- Applying the model to the Typography-MNIST (TMNIST) dataset, which has not been widely used in generative modeling.
- Exploring conditioning strategies for improved generation quality and conducting ablation studies on model variants.

## 2 Motivation

Font style transfer has significant applications in digital typography, personalized typeface design, and creative AI tools. By leveraging CVAEs, we aim to learn structured latent spaces that separately encode character identity and font style.

This project helps us solidify our understanding of variational inference and deep generative models taught in ECE285, and enables us to apply theoretical concepts such as KL divergence and the reparameterization trick to real-world visual data.

## 3 Approach

Our approach involves the following pipeline:

1. **Data Preparation**: Utilize the TMNIST dataset, which contains 565,292 grayscale images of 1,812 unique glyphs rendered in 1,355 Google Fonts.
2. **Model Architecture**: Implement a CVAE architecture where the encoder maps an image and a style label to a latent vector, and the decoder generates glyphs conditioned on the latent vector and target style.
3. **Training**: Train the model by minimizing the reconstruction loss and KL divergence between approximate posterior and prior.
4. **Evaluation**: Evaluate the generated glyphs using SSIM, pixel-wise loss, and visual quality checks. Latent space interpolation and style transfer tests will be conducted to validate generation controllability.

5. **Ablation and Optimization**: Explore different conditioning schemes (e.g., one-hot label, learned embeddings), and evaluate their impact through controlled experiments.

Our methodology is grounded in the foundational work by Doersch [1] and extends it to structured image synthesis with style control.

## 4 Dataset

We adopt the Typography-MNIST (TMNIST) dataset introduced by Magre et al. [2], which contains a large collection of character glyphs rendered using modern and historical fonts. TMNIST's diversity enables learning disentangled style-content representations, making it ideal for CVAE-based generation tasks.

## References


[1] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2021.

[2] Nimish Magre and Nicholas Brown. Typography-mnist (tmnist): an mnist-style image dataset to categorize glyphs and font-styles, 2022.