Vector Quantized Variational Autoencoders on Novel Datasets

Shivanshu Gupta   **Kolby Nottingham**   **Preethi Seshadri**

Abstract

We will explore using the Vector Quantized Variational Autoencoder (VQ-VAE) to generate discrete representations for the Kaokore dataset, which contains images of facial expressions from traditional Japanese illustrations (`https://github.com/rois-codh/kaokore`). The framework VQ-VAE is built on, Variational Autoencoders (VAE), learn continuous latent representations. While continuous representations are flexible, many real world attributes are better defined discretely, and some current state-of-the art model architectures, like transformers, only work with discrete data. Additionally, VAEs have been shown to exhibit posterior collapse, which means that latent codes are ignored. In this project, we will experiment with VQ-VAEs on a novel dataset and design experiments to test the advantages and disadvantages of this approach in terms of generation quality and learned latent structure. Since the VQ-VAE paper uses CIFAR10 and 128x128 ImageNet images, there might be additional experimentation and training required to achieve good performance on the Kaokore dataset. A successful project will demonstrate the strengths and weaknesses of VQ-VAEs over traditional VAEs.