

---

# Vector Quantized Variational Autoencoders on Novel Datasets

---

Shivanshu Gupta

Kolby Nottingham

Preethi Seshadri

## Abstract

We 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 experiment with VQ-VAEs on a novel dataset and design experiments to test the advantages and disadvantages of multiple VQ-VAE variations. Our results indicate that while the original VQ-VAE algorithm learns faster than some of its successors, it does not achieve the same level of performance.

## 1 Introduction

Generative machine learning models learn a distribution  $p(x, y)$  for data instances  $x$  and labels  $y$ . Compared to discriminative machine learning models that learn  $p(y|x)$ , generative models can be more difficult to learn but come with added benefits. One use case for generative machine learning models is mapping data instances to a latent space. Once mapped to a latent space, latent variables can be used as a compressed representation of a data instances or to sample and generate new data instances.

One generative machine learning model that learns latent variable representations of data is the variational autoencoder (VAE). VAEs learn a latent variable representation of each datapoint with an encoder module. Simultaneously, a decoder module learns to reconstruct the original image from the latent variables.

Traditional VAEs learn a latent space that is associated with a probability distribution of continuous variables. Vector quantized variational autoencoders (VQ-VAE) learn discrete latent variables instead. Discrete variables are sometimes advantageous because real world data can often be summarized by categorical features.

In this work, we compare two variations of VQ-VAEs. The second variation (VQ-VAE2) was released as a follow up to VQ-VAEs and introduces a hierarchical structure for the model. We go into further detail comparing the algorithms in section ???. We implement each of these methods and run experiments with each on the Kaokore dataset. To the best of our knowledge, this is the first time VAEs have been used to learn a latent space for this dataset.

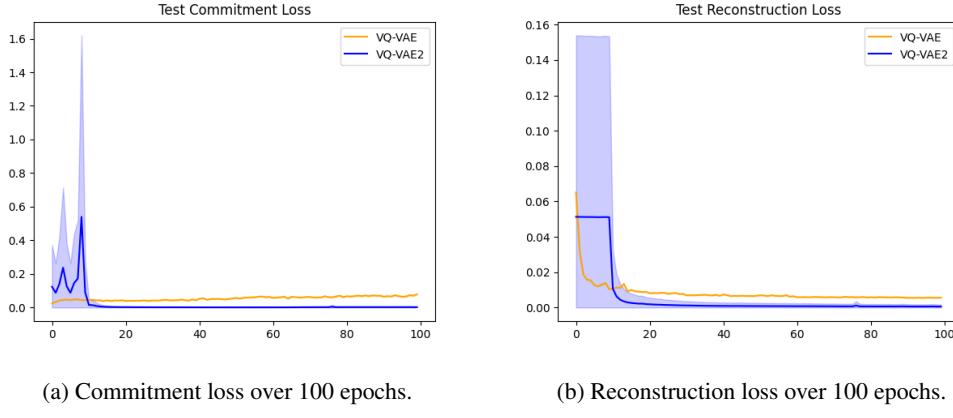
## 2 Background

Generative models...

Autoencoders...

	Commitment Loss	Reconstruction Loss
VQ-VAE	0.0793 (0.0027)	0.0056 (0.0001)
VQ-VAE2	0.0025 (0.0035)	0.0005 (0.0008)

Table 1: Commitment loss and reconstruction loss on the Kaokore testset for both algorithms.



(a) Commitment loss over 100 epochs.

(b) Reconstruction loss over 100 epochs.

### 3 Vector Quantized Variational Autoencoders

#### 3.1 VAE

#### 3.2 VQ-VAE

#### 3.3 VQ-VAE2

### 4 Experiments

In a series of experiments, we compare traditional VAEs to VQ-VAEs. We first...

#### 4.1 Kaokore Dataset

The Kaokore dataset consists of...

#### 4.2 Results

As indicated in table...

### 5 Conclusion

The authors of the VQ-VAE paper presented promising results. However, based on our experiments, we conclude that it is not strictly better than VAE on every dataset...

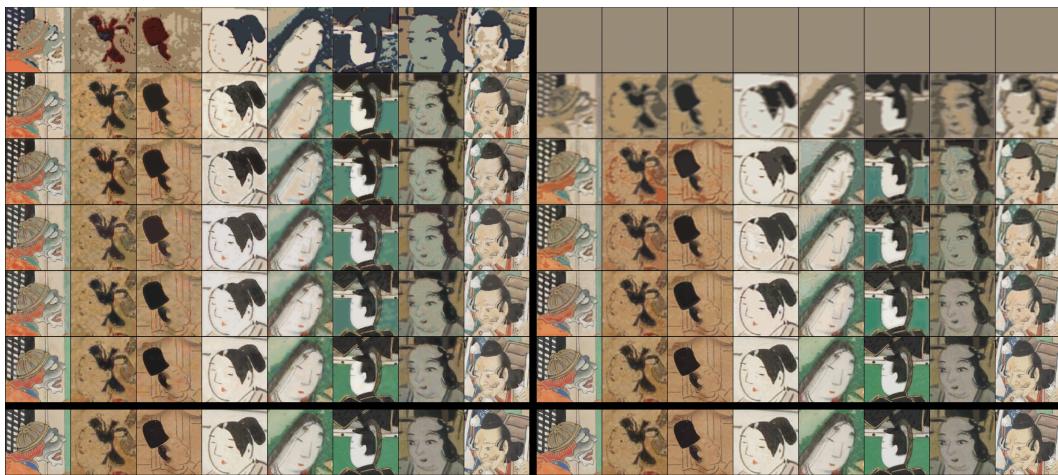


Figure 2: The first six rows of images show samples at 1, 12, 14, 16, 18, and 100 epochs for VQ-VAE (left) and VQ-VAE2 (right). The final row shows the corresponding images from the testset.