VQ-VEA FOR TIME SERIES PATTERN DISCOVERY

1 Introduction

The VQ-VAE Vector Quantized Variational Autoencoder (Van Den Oord et al., 2017) is a type of variational autoencoder that employs discrete latent representations instead of continuous ones.

The input, such as an image or time series, is passed through an encoder, which transforms it into a lower-dimensional latent representation, typically in the form of a set of latent vectors. This latent representation captures essential features of the input data while reducing dimensionality. In the quantization stage, each latent vector produced by the encoder is compared against all vectors in a learned codebook, or set of embeddings. The nearest vector from the codebook, determined by Euclidean distance, is selected as the quantized representation of the latent vector. This step forces the latent representation to align with discrete codebook entries, facilitating more structured representation learning. The quantized latent vectors (i.e., the selected codebook entries) are then passed through a decoder, which attempts to reconstruct the original input from this quantized latent space. The decoder is trained to minimize the reconstruction error, ensuring that the output resembles the input as closely as possible.

VQ-VAE employs two key loss functions:

Reconstruction Loss: This loss, typically computed using Mean Squared Error (MSE), measures how well the decoder can reconstruct the original input from the quantized latent vectors.

Commitment Loss: This term encourages the latent representations to "commit" to a single codebook vector, preventing the use of many different embeddings and ensuring that each vector in the latent space is clearly mapped to a specific codebook entry.

2 Experiments

We explored the use of VQ-VAE to find patterns in time series data. Here, we present results based on two datasets: synthetic time series data with injected noise and a plane dataset. Both datasets are described in the T2P paper.

To use VQ-VAE, we first modified the algorithm to make it compatible with time series. Because in practice the decoder activates multiple embeddings, we added a max function to select the highest-activation embedding. Using the Plane dataset, there are 7 predefined classes, thus we set the embedding space to 7 vectors.

We show the results below. In each case, we also report the silhouette score to quantify how well the samples are clustered. Higher scores indicate better separation between clusters and more cohesive clusters.

3 Discussion

Although VQ-VAE works effectively for tasks like image compression and generation, it presents significant shortcomings when applied to time series data, particularly in pattern recognition. Our experiments revealed several key issues. First, in the context of time series data, VQ-VAE struggles to identify and encode recurrent patterns. This occurred in the synthetic and plane data experiments, where VQ-VAE consistently failed to recognize recurring structures in the input data.

Second, VQ-VAE is sensitive to the number of embedding vectors. In our experiments, we tested the model with various embedding sizes, values that were smaller, equal to, and larger than the actual number of distinct patterns present in

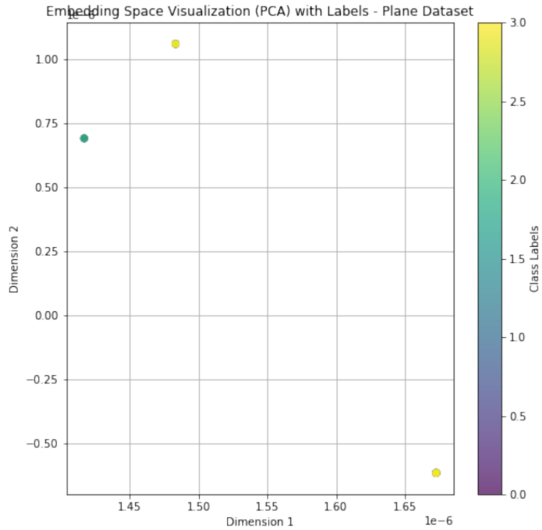

Figure 1: Visualization of VQ-VAE discoveries for synthetic data, no added noise, K=2 clusters. The silhouette score is -0.184.

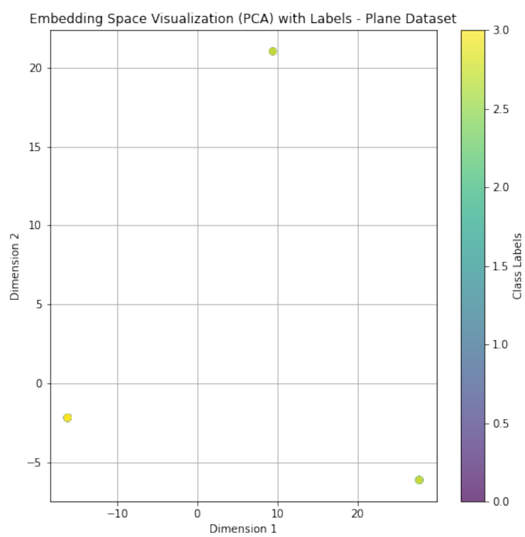

Figure 2: Visualization of VQ-VAE discoveries for synthetic data, no added noise, K=4 clusters. The silhouette score is -0.150.

the data. When the number of embedding vectors was less than the number of distinct patterns, the model failed to adequately capture the data structure, leading to poor reconstructions. This contrasts with T2P, which can still learn and represent the most frequent patterns even when the embedding size is not optimal.

Third, VQ-VAE fails in noisy conditions. As noise levels in the input data increased, VQ-VAE's ability to reconstruct the input degraded rapidly. In noisy settings, the model was unable to reconstruct any meaningful data. Conversely, T2P demonstrated a higher degree of resilience, managing to handle noise effectively and maintaining pattern recognition capabilities.

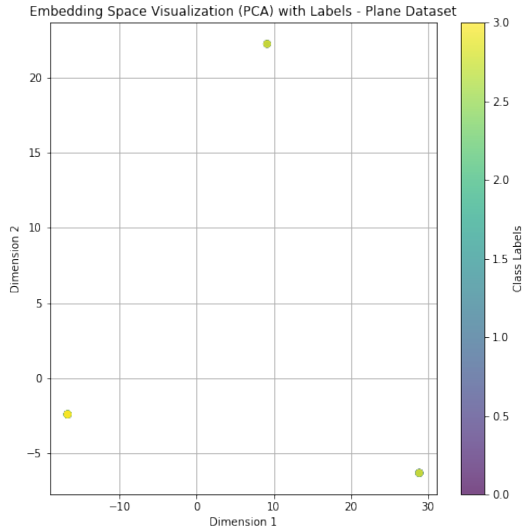

Figure 3: Visualization of VQ-VAE discoveries for synthetic data, no added noise, K=6 clusters. The silhouette score is -0.170.

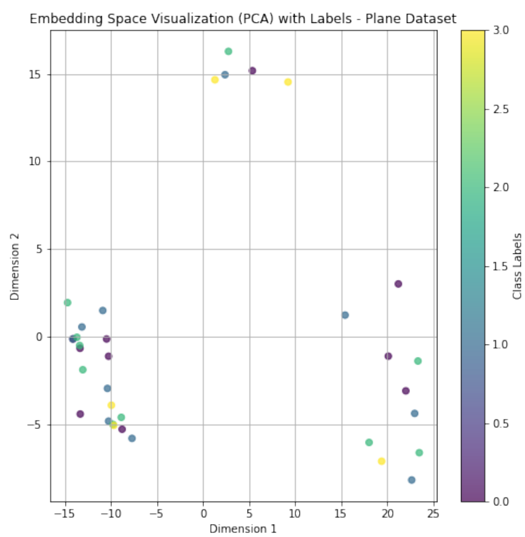

Figure 4: Visualization of VQ-VAE discoveries for synthetic data, 30% added noise, K=2 clusters. The silhouette score is -0.061.

Fourth, during reconstruction, we observed that more than one latent vector often became active for a given input, complicating the interpretability of the learned representations. To analyze whether the activated vectors aligned with meaningful clusters, we examined the use of the "max" vector (the most activated one) and compared it to the true labels of the data. However, we found that the model randomly assigned vectors during reconstruction, with no clear pattern emerging from the latent space representation.

Fifth, to further investigate the clustering ability of VQ-VAE, we visualized the latent space by coloring each point based on its true label, as shown in the figures. The silhouette score, a metric that measures the quality of clustering,

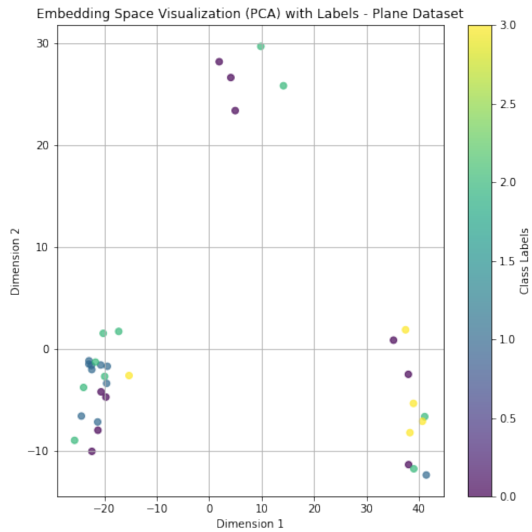

Figure 5: Visualization of VQ-VAE discoveries for synthetic data, 30% added noise, K=4 clusters. The silhouette score is -0.199.

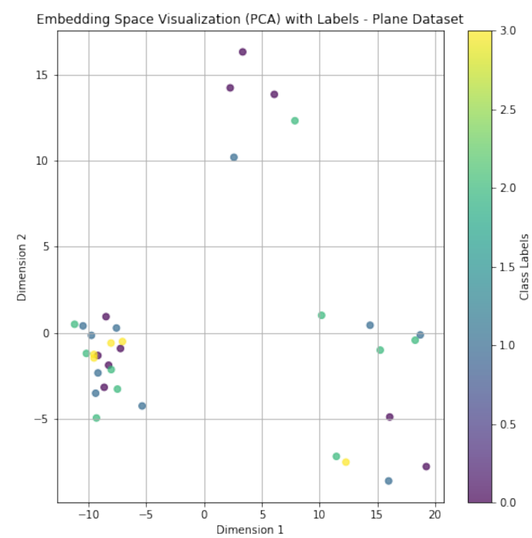

Figure 6: Visualization of VQ-VAE discoveries for synthetic data, 30% added noise, K=6 clusters. The silhouette score is -0.077.

was negative, confirming that the model failed to learn any meaningful patterns or clusters from the input data. This lack of structure in the latent space reinforces the claim that VQ-VAE is ill-suited for time series pattern recognition.

Finally, even in the original VQ-VAE papers, the authors did not showcase the model's ability to handle time series data for pattern recognition. This omission suggests that the method may not be intended or optimized for such tasks.

Based on these findings, we conclude that VQ-VAE is ineffective for pattern discovery in time series data and summarization of these data. The algorithm's inability to capture recurrent patterns, its sensitivity to embedding vector

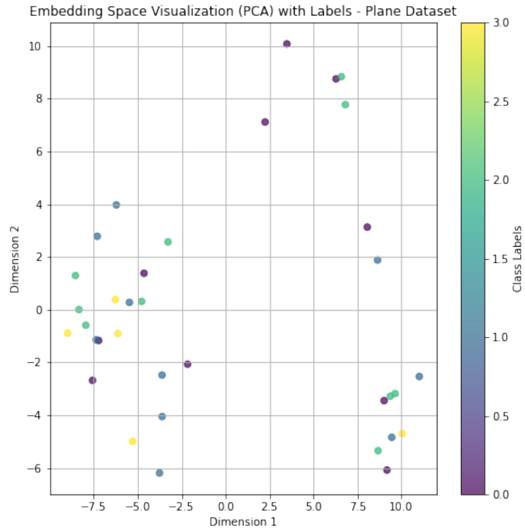

Figure 7: Visualization of VQ-VAE discoveries for synthetic data, 70% added noise, K=2 clusters. The silhouette score is -0.071.

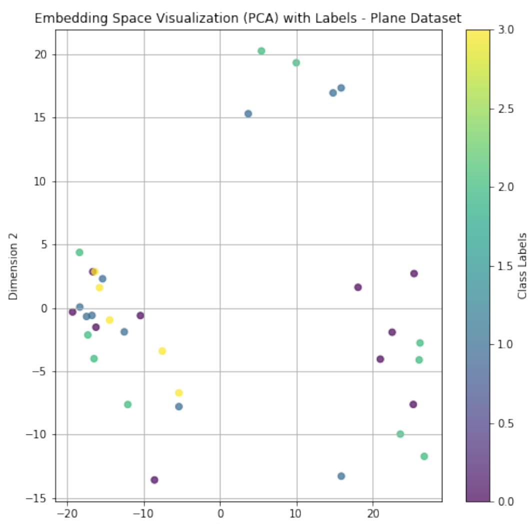

Figure 8: Visualization of VQ-VAE discoveries for synthetic data, 70% added noise, K=4 clusters. The silhouette score is -0.094.

sizes, and its poor performance in noisy conditions render it unsuitable for this task. Models like T2P, which are more robust in these scenarios, offer a better alternative for time series pattern discovery and data summarization.

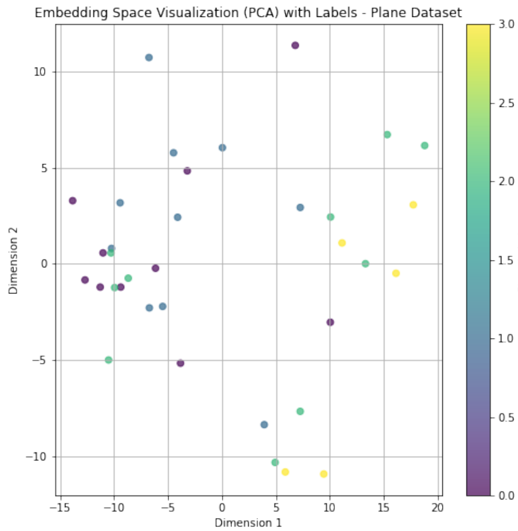

Figure 9: Visualization of VQ-VAE discoveries for synthetic data, 70% added noise, K=6 clusters. The silhouette score is -0.051.

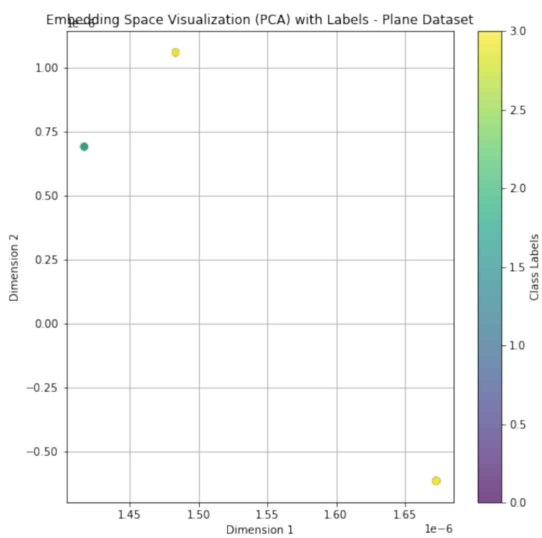

Figure 10: Visualization of VQ-VAE discoveries for synthetic data, 100% added noise, K=2 clusters. The silhouette score is -0.244.

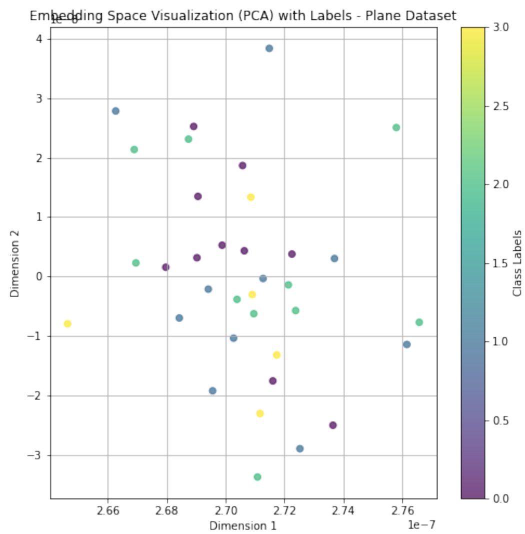

Figure 11: Visualization of VQ-VAE discoveries for synthetic data, 100% added noise, K=4 clusters. The silhouette score is -0.306.

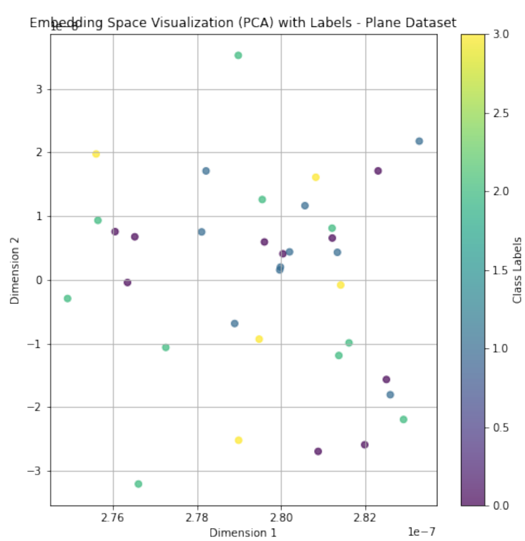

Figure 12: Visualization of VQ-VAE discoveries for synthetic data, 100% added noise, K=6 clusters. The silhouette score is -0.048.

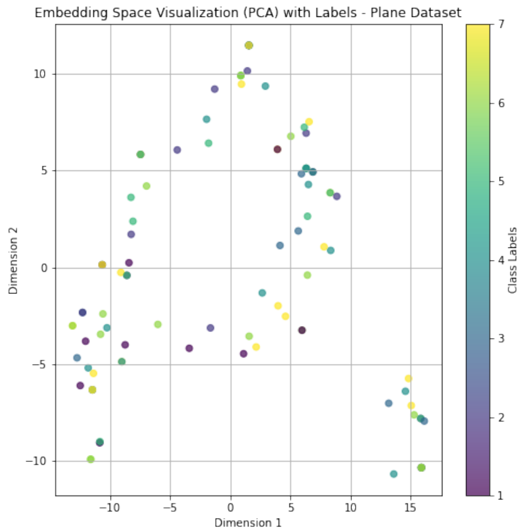

Figure 13: Visualization of VQ-VAE discoveries for plane data, K=4 clusters. The silhouette score is -0.089.

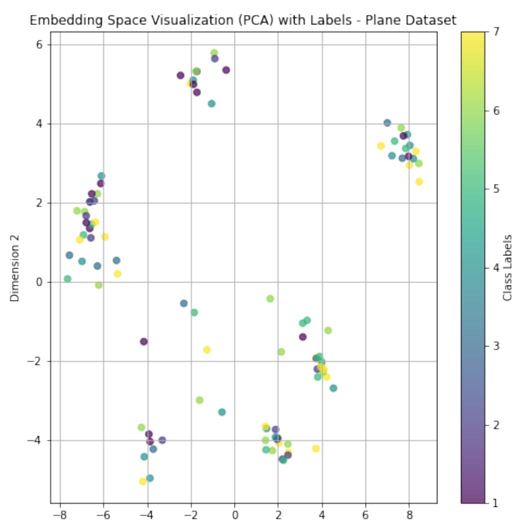

Figure 14: Visualization of VQ-VAE discoveries for plane data, K=7 clusters. The silhouette score is -0.104.

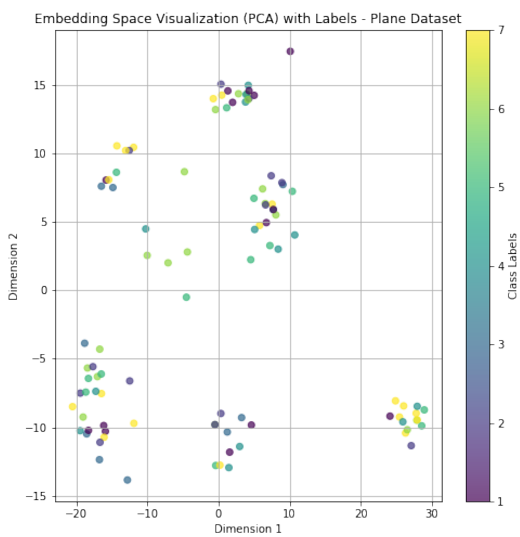

Figure 15: Visualization of VQ-VAE discoveries for plane data, K=14 clusters. The silhouette score is -0.102.