universität freiburg

Reconstructing 3D Shapes

Muhammad Ali, Soham Basu, Yusuf Salcan

Deep Learning Lab

Motivation

3D representations are essential for applications in robotics, self-driving, and virtual/augmented reality. This has led to an increasing number of diverse tasks that rely on effective 3D representations such as single view 3D prediction, shape completion and conditional generation.

There have been recent attempts to unify the solving of these tasks using state-of-the-art architectures like transformers and diffusion models. However, training these models directly over the continuous and high-dimensional space of 3D shapes is computationally infeasible.

In this project, we **reduce high-dimensional 3D shapes to lower dimensional discrete representations** using a Vector-Quantized Variational Autoencoder (VQ-VAE). This reduced representation enables downstream tasks to be performed efficiently.

Technical Approach

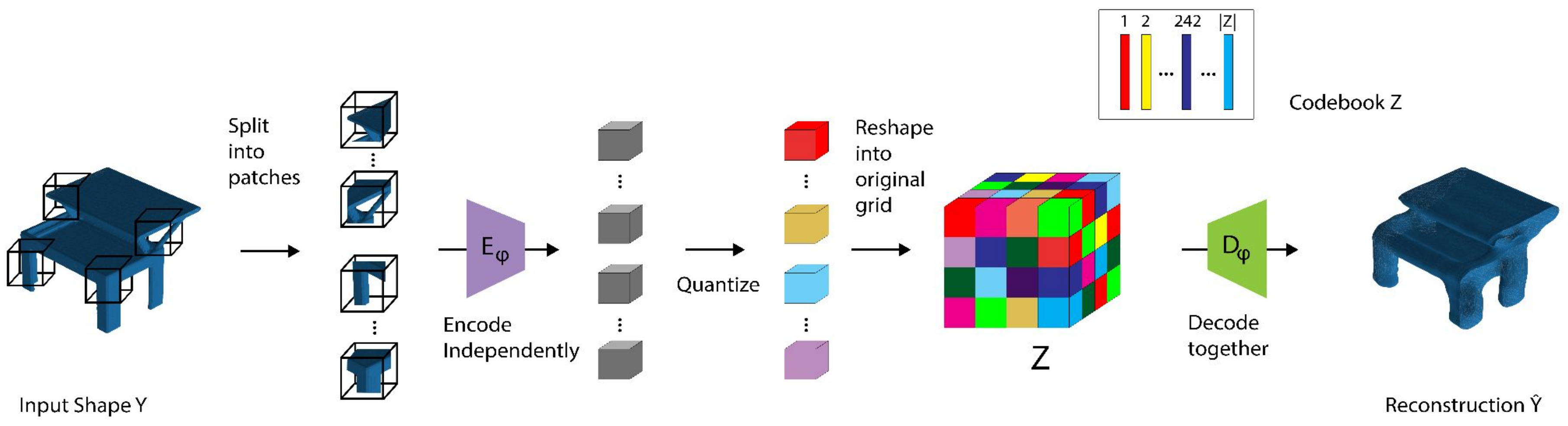

Figure 1: Overview of technical approach

We implemented the Patchwise VQ-VAE (P-VQ-VAE) architecture proposed by shown in Figure 1, based on Mittal et al.
3D meshes from the ShapeNet dataset are converted to a Truncated Signed-Distance-Function (T-SDF) representation and split into patches. These patches are independently encoded by the encoder network and subsequently quantized in the Vector Quantization step. In this step, the encoding of each patch is mapped to the nearest element in the codebook **Z**, which is jointly learned while training the VQ-VAE.

The codebook embeddings are folded into a cube and jointly decoded by the decoder network to obtain a reconstruction of the original mesh.

Novelty

A significant shortcoming in the P-VQ-VAE model proposed by Mittal et al. is **embedding collapse**, wherein only a few embeddings are repeatedly chosen from the codebook. This limits learning and results in sub-optimal performance.

To address this problem, we utilize selected approaches from Vali et al. and combine them with novel enhancements to obtain **better performance** and **faster training** than what we observed in exclusively using either approach.

Following Vali et al., we used a vector quantization technique called NSVQ (Noise Substitution in Vector Quantization). The NSVQ simulates vector quantization error with the product of the original quantization error magnitude with a normalized noise vector. Vali et al. show that this renders higher accuracy and faster convergence than straight through estimator (STE).

To mitigate the problem of embedding collapse, we used a novel approach called **codebook dropout**. We randomly mask a percentage of codebook embeddings while calculating similarity for each batch of patches, ensuring diversity in the selected codebook indices.

We trained the model to optimize **reconstruction loss,** with an L1 penalty for regularization.

Evaluation

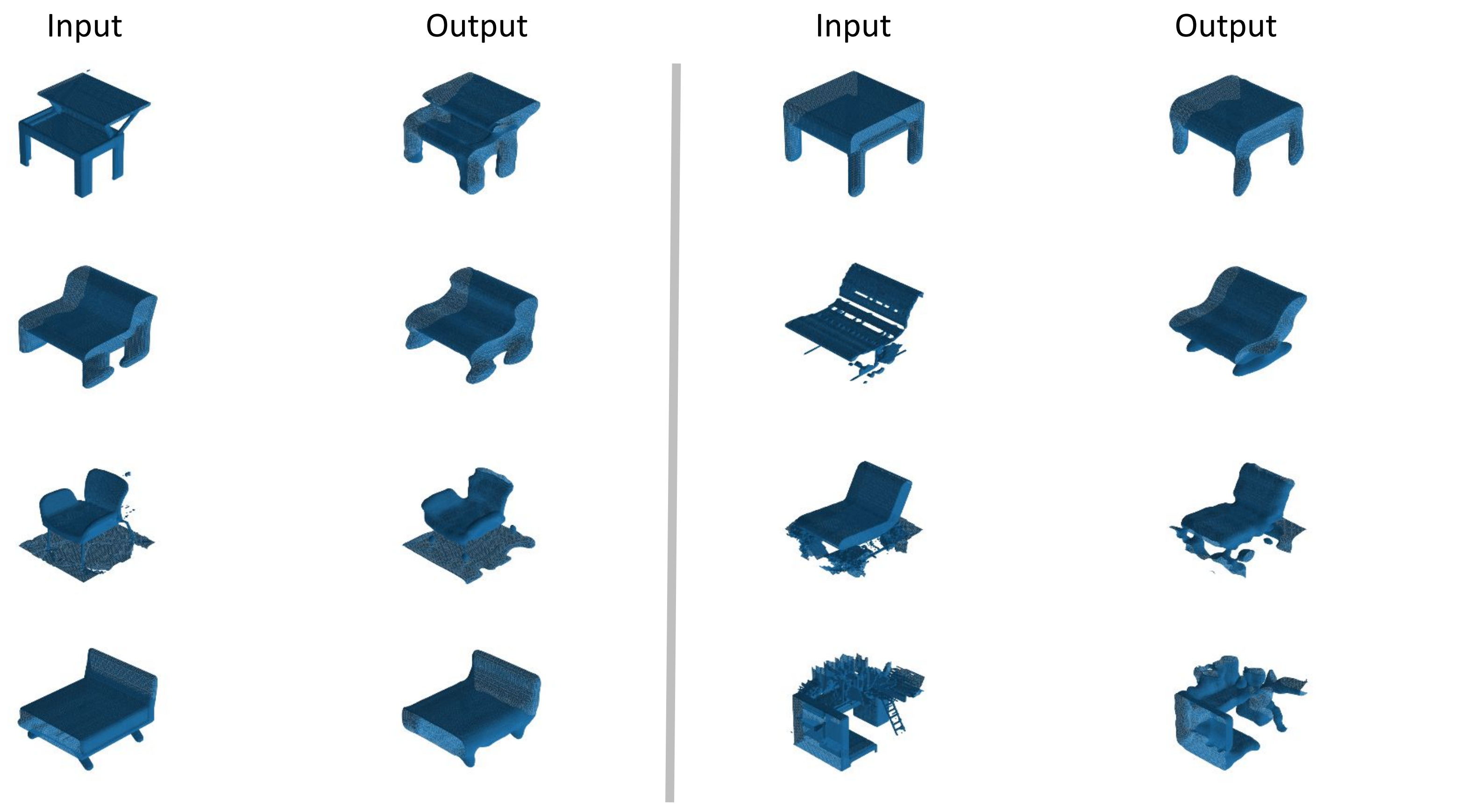

Figure 2: Reconstructions of the 3D objects using the VQ-VAE

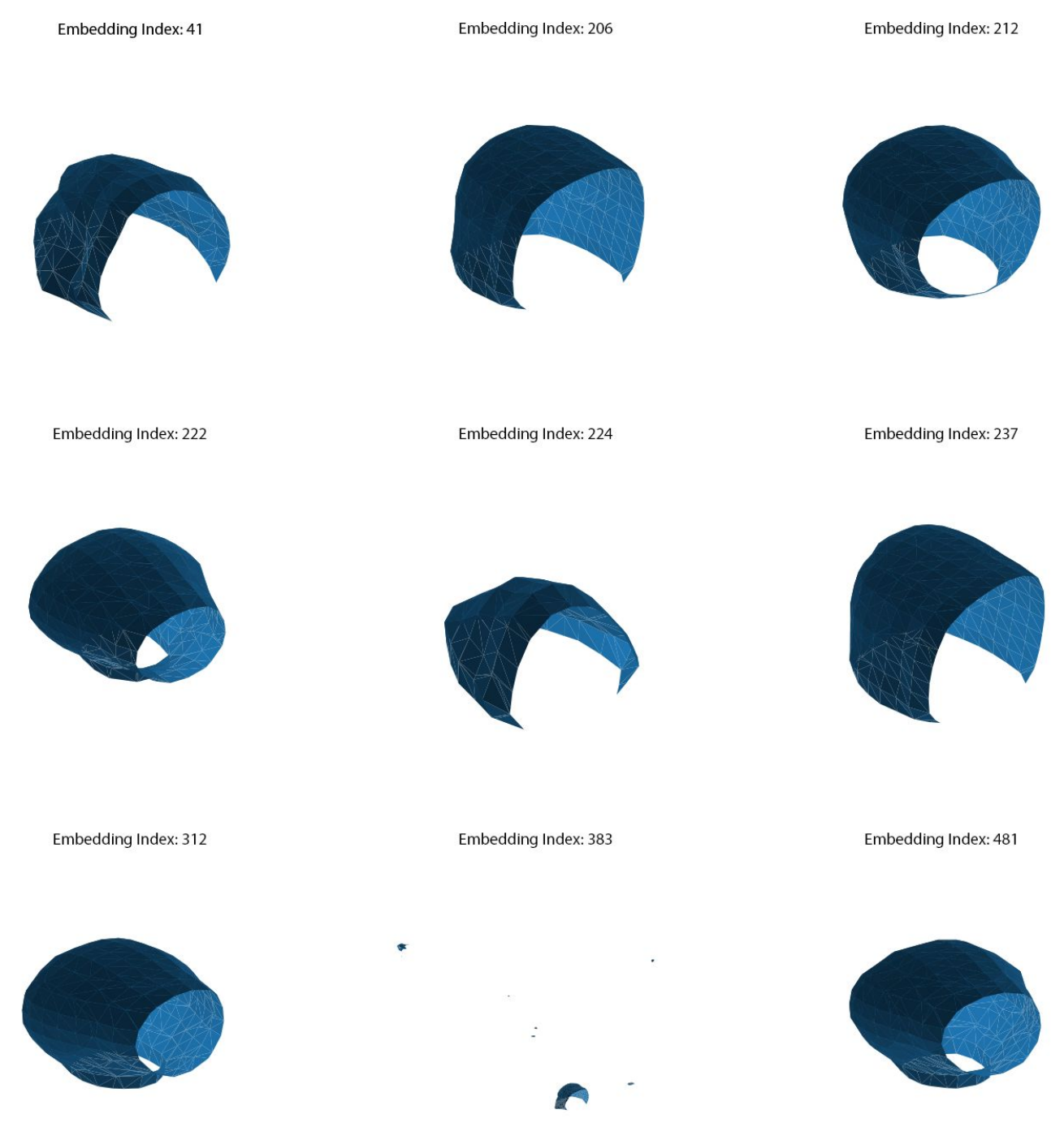

Figure 3: Visualization of codebook embeddings

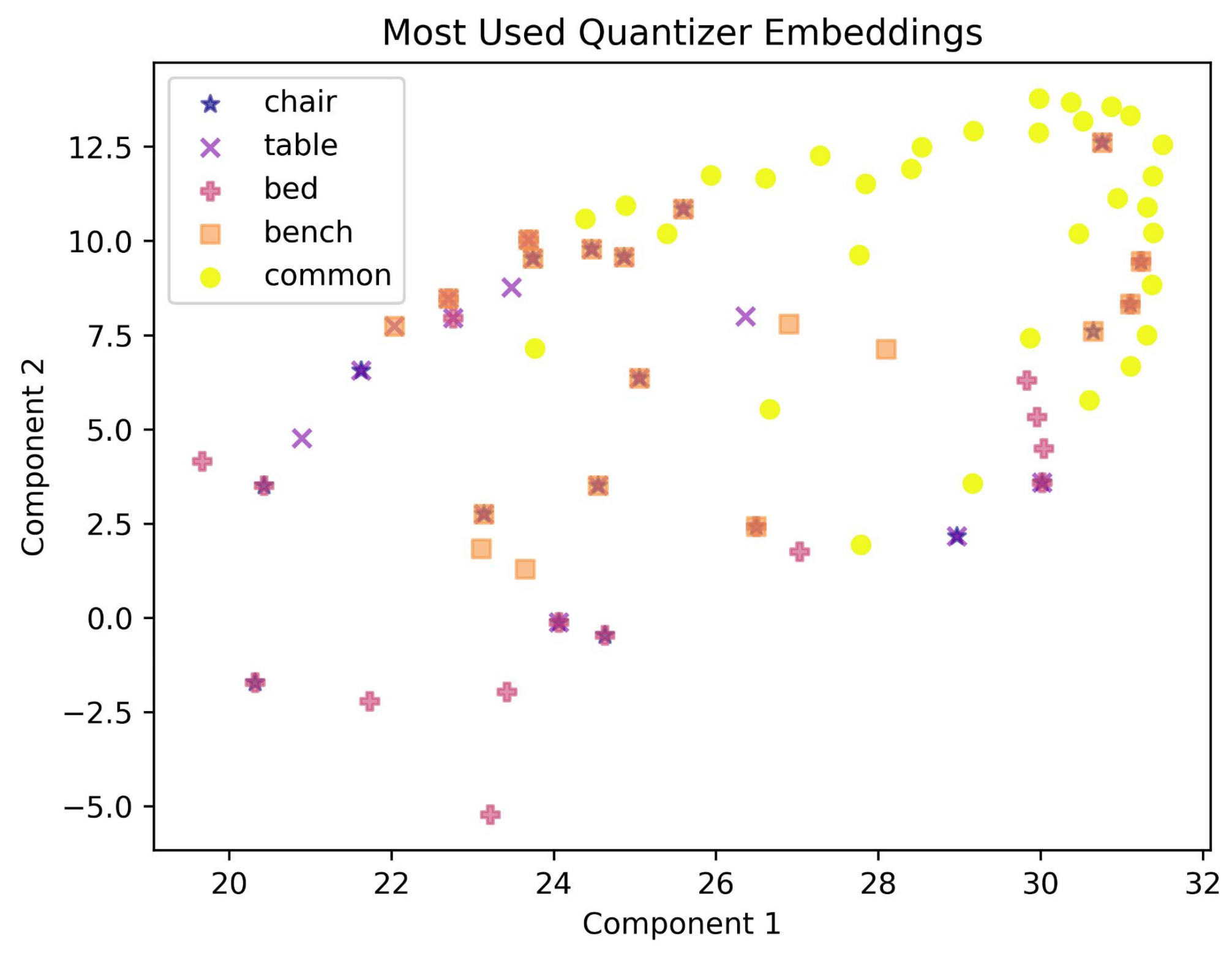

Figure 4: Distribution of the VQ embedding space by classes

Most frequently used codebook embeddings are plotted using t-SNE. This result shows that most of the frequently used embeddings are shared by all classes. Also, structurally similar classes, like table and bed, use mostly same embeddings.

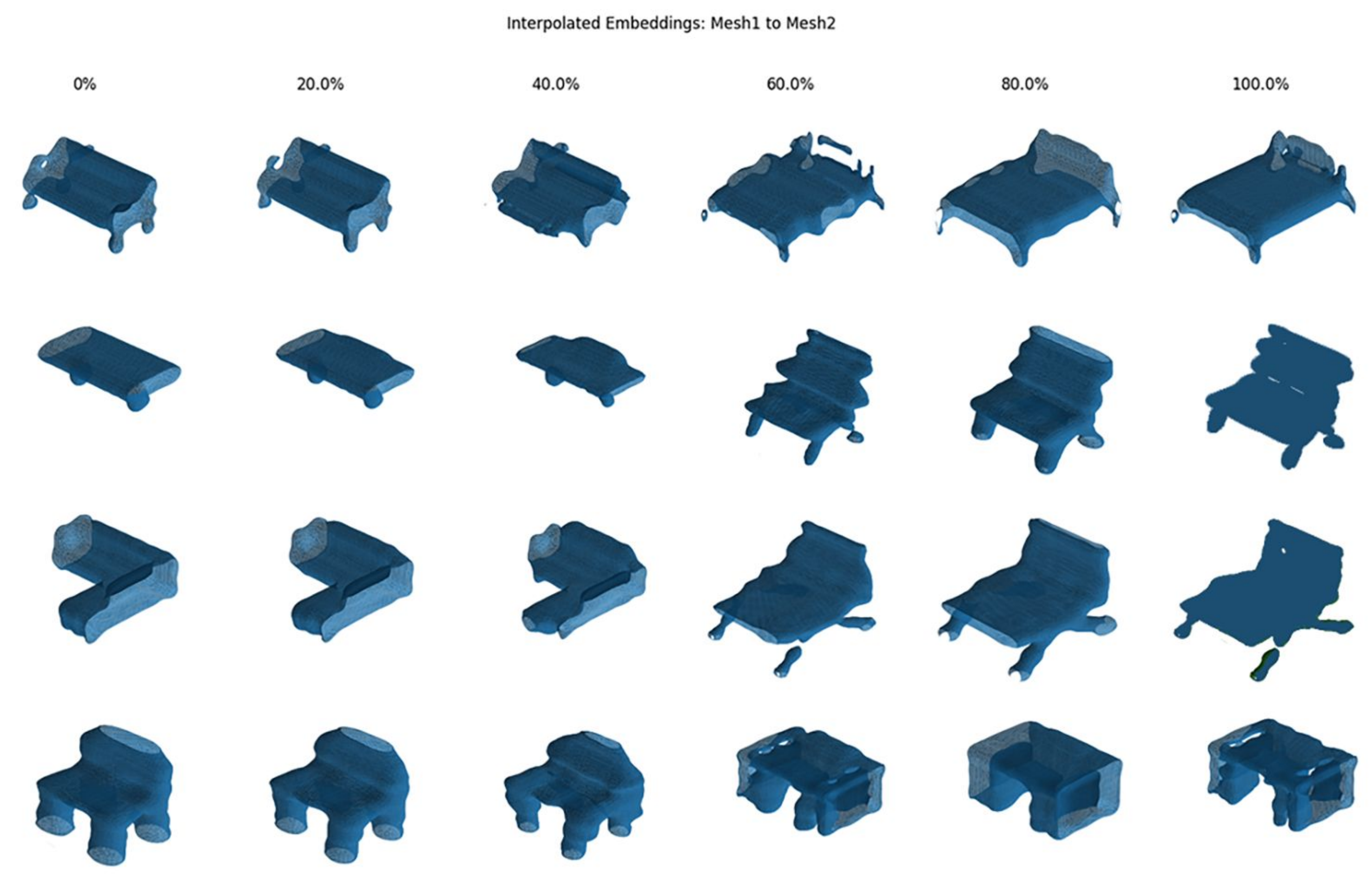

Figure 5: Interpolation of VQ embeddings of different classes

Embeddings of two different classes are linearly interpolated and resulting 3D reconstructions are displayed. This result shows that although the embedding space is discrete, continuous interpolations between discrete embeddings can be decoded as the 3D features for corresponding objects.

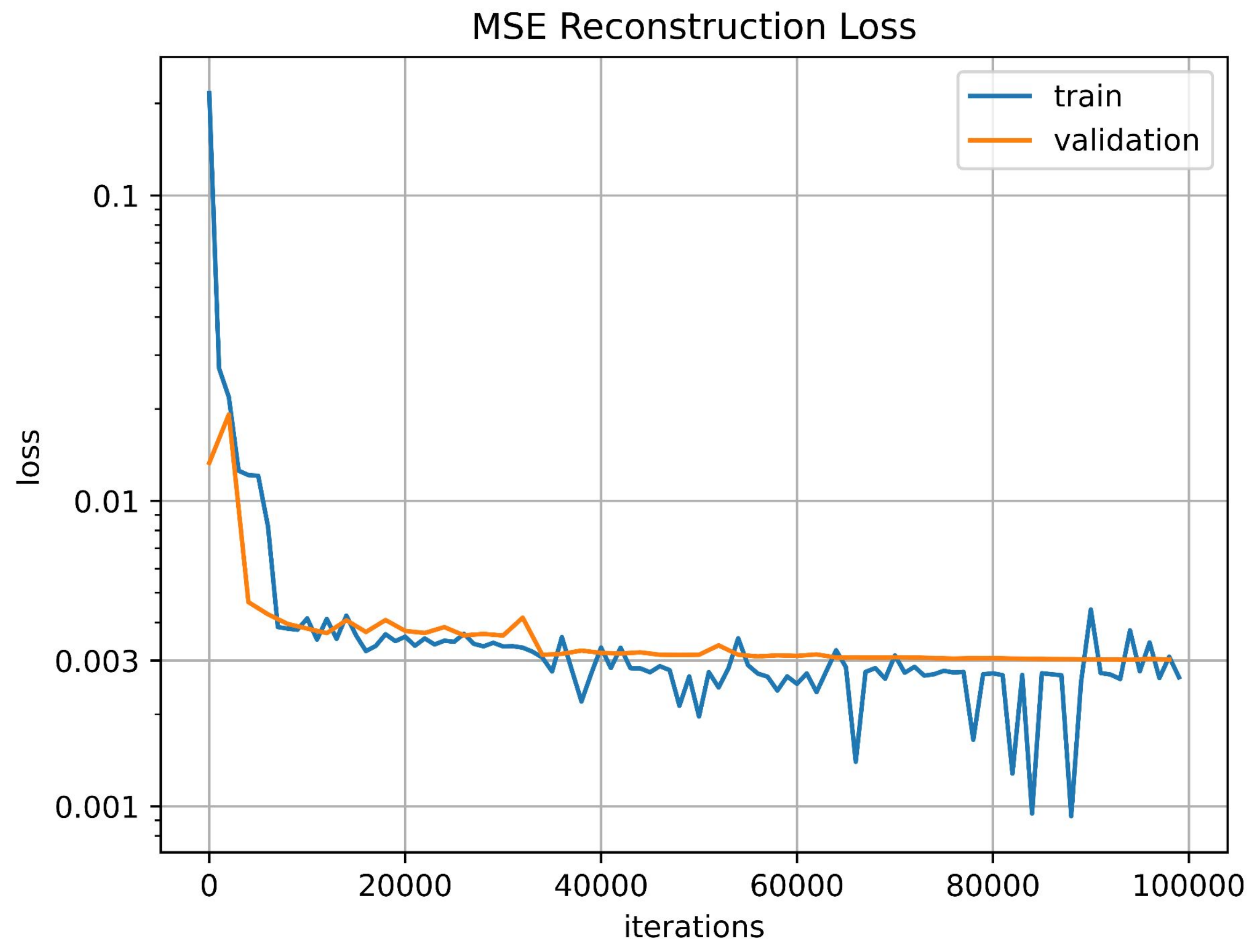

Figure 6: Training and validation reconstruction loss curves