LATENT SPACE EXPLORATION IN VARIATIONAL AUTOENCODERS USING FASHION MNIST

ADHIDEV MD
(RA2211056010039)

Abstract

Variational Autoencoders (VAEs) are powerful generative models capable of learning meaningful latent space representations. This paper explores the latent space of a trained VAE model on the Fashion MNIST dataset, demonstrating its ability to generate new samples and interpolate between data points. The study investigates how the learned latent space captures key features of the dataset and evaluates the model's performance based on reconstruction accuracy and smoothness of generated transitions.

Keywords: Variational Autoencoders, latent space, Fashion MNIST.

1. Introduction

1.1 Overview of Variational Autoencoders (VAEs)

Variational Autoencoders (VAEs) are a class of probabilistic generative models that extend traditional autoencoders by introducing a latent variable model. Unlike classical autoencoders, VAEs impose a probabilistic structure on the latent space, encouraging meaningful organization of data points. The encoder network maps input data to a probability distribution over the latent space, while the decoder network reconstructs the input from sampled latent vectors.

A key feature of VAEs is their ability to generate new data by sampling from the latent distribution, making them effective in image generation, data augmentation, and representation learning. The Fashion MNIST dataset, consisting of grayscale images of clothing items, serves as an ideal benchmark for evaluating the generative power of VAEs.

1.2 Latent Space Representation

The latent space in a VAE is a compressed, structured representation of input data. Each point in this space corresponds to a potential output image. By exploring this space, we can:

- □ Generate new samples by sampling random points from the latent space.
- □ Interpolate between two points to observe smooth transitions.
- □ Analyze how different regions correspond to distinct data features.

This study focuses on training a VAE on Fashion MNIST, visualizing the learned latent space, generating new samples, and performing interpolation between data points.

2. Implementation Details

2.1 Model Architecture

The VAE consists of:

- □ Encoder: Maps input images to a lower-dimensional latent representation (mean and variance).
- □ Latent Space Sampling: Uses the reparameterization trick to sample points from the learned latent distribution.
- □ Decoder: Reconstructs input images from sampled latent vectors.

2.1.1 Encoder Architecture

- □ Conv2D layers for feature extraction
- □ Dense layers to compute latent mean and variance
- □ Reparameterization trick for sampling

2.1.2 Decoder Architecture

- □ Dense layers to map latent vectors to an initial shape
- □ Conv2D Transpose layers for upsampling
- □ Sigmoid activation to reconstruct the grayscale image

2.2 Training & Loss Function

The loss function consists of two components:

- Reconstruction Loss: Measures how well the decoded images match the original inputs.
- KL Divergence Loss: Encourages the latent space to follow a standard normal distribution.

The total loss function is:

$L=\text{Reconstruction Loss}+\beta\cdot\text{KL Divergence}$ where

β is a weight factor balancing the two losses.

The model is trained using the Adam optimizer for 20 epochs with a batch size of 128.

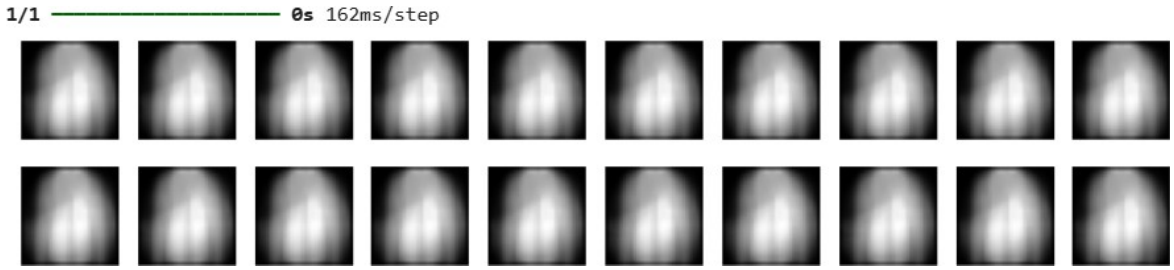
3. Results & Analysis

3.1 Latent Space Exploration

After training, we visualize the latent space by encoding test images and plotting their (z_mean, z_log_var) values. A well-structured latent space ensures smooth transitions between different clothing categories.

3.2 Sample Generation

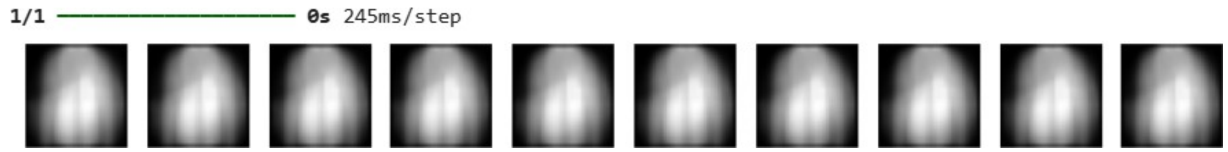
To generate new images, we sample random points from the latent space and pass them through the decoder. The generated images resemble real clothing items, demonstrating the VAE's ability to learn meaningful representations.



```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 162ms/step
```

3.3 Latent Space Interpolation

To examine the continuity of the learned representations, we interpolate between two latent points and generate intermediate images. The gradual transformation between images confirms that the VAE learns a smooth latent space.



```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 245ms/step
```

4. Conclusion

Our study demonstrates the ability of VAEs to capture meaningful latent representations in the Fashion MNIST dataset. Key findings include:

1. Latent Space Structure: The two-dimensional latent space shows a smooth, organized distribution of different fashion categories.
2. New Sample Generation: Sampling from the latent space produces realistic and diverse clothing items.
3. Interpolation Analysis: The smooth transitions between latent points confirm the model's effectiveness in learning structured feature representations.

These results highlight the potential of VAEs in generative tasks, especially in applications requiring data synthesis and manipulation. Future work can explore higher-dimensional latent spaces, improved architectures, and conditional VAEs for better control over generated samples.

References

1. Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.
2. Doersch, C. (2016). Tutorial on Variational Autoencoders. *arXiv preprint arXiv:1606.05908*.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.