SHAPE GENERATION AND LATENT INFORMATION IN AUTOENCODERS

Swati Swati

Victoria University of Wellington, NZ

1. INTRODUCTION

Variational Autoencoders (VAEs) are a class of generative models that are particularly well-suited for tasks involving the synthesis of data, such as image generation. Unlike traditional autoencoders, VAEs introduce a probabilistic approach to the latent space, allowing for the generation of new data points by sampling from a continuous latent distribution. This makes VAEs powerful for applications requiring diversity in generated outputs, such as generating images of different objects or shapes [4].

In this study, we explore the application of a VAE and a basic autoencoder for generating images containing geometric shapes, such as circles, triangles, and rectangles, randomly placed in 28x28 pixel images. The objective is to evaluate the performance of these autoencoders in generating realistic and diverse shapes, while also estimating the information passing through the latent layer.

2. THEORY

In this section, we focus on the theoretical background relevant to the implemented system.

2.1. Basic Autoencoders

A basic autoencoder is a neural network for unsupervised learning, designed to encode input data into a latent space and reconstruct it accurately. It consists of an encoder that compresses input $\mathbf{x}$ into a lower-dimensional latent representation $\mathbf{z}$, and a decoder that reconstructs the input from $\mathbf{z}$. The objective is to minimize the reconstruction error, typically measured using Mean Squared Error (MSE):

$$\mathcal{L}_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 \tag{1}$$

where $\mathbf{x}_i$ is the input, $\hat{\mathbf{x}}_i$ is the reconstructed output, and N is the number of samples. This loss helps the network learn efficient latent representations for accurate reconstruction [3].

2.2. Variational Autoencoders

Variational Autoencoders (VAEs) extend the basic autoencoder by introducing a probabilistic approach to the latent space. Instead of directly encoding an input $\mathbf{x}$ into a fixed latent representation $\mathbf{z}$, the VAE encodes it as a distribution, typically Gaussian, with a mean μ and variance σ^2. The loss function combines the reconstruction error and a regularization term (KL divergence) to encourage the latent variables to follow a normal distribution:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + D_{\text{KL}}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})) \tag{2}$$

where $\mathcal{L}_{\text{recon}}$ is the reconstruction loss, and D_{KL} is the Kullback-Leibler divergence between the learned latent distribution $q(\mathbf{z}|\mathbf{x})$ and the prior $p(\mathbf{z})$ [1].

2.3. t-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a dimensionality reduction technique particularly effective for visualizing high-dimensional data. It maps data points to a lower-dimensional space while preserving local structures, making it ideal for exploring latent representations [2].

2.4. Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is a perceptual metric used to evaluate the quality of reconstructed images by comparing their structural information, luminance, and contrast. It provides a more accurate reflection of perceived visual quality compared to traditional metrics like Mean Squared Error (MSE), as it considers human visual perception.

2.5. Information Passing

The information passing through the latent layer in autoencoders can be quantified using the mutual information formula:

$$I(Y;Z) = \frac{1}{2} \log \frac{\sigma_Z^2}{\sigma_\epsilon^2}$$

where $I(Y;Z)$ represents the mutual information between the input Y and the latent representation Z. This metric reflects the amount of relevant information retained during the encoding process, providing insights into the efficiency of the learned representation [5].

3. EXPERIMENTS

In this section, we investigate the performance of a VAE and basic autoencoder designed to generate images containing geometric shapes, such as circles, triangles, and rectangles, randomly placed in 28x28 pixel images. We set up an experimental framework to evaluate the effectiveness of the VAE and modified autoencoder in generating realistic and diverse shapes.

3.1. Experimental Setup

A dataset of 50,000 images, each measuring 28x28 pixels, was created, featuring a single geometric shape randomly positioned. The Variational Autoencoder (VAE) architecture consisted of an encoder with three convolutional layers and a decoder with three transposed convolutional layers, which compressed input images into a latent representation and reconstructed them from this latent space. Key hyperparameters, including latent dimension, learning rate, and batch size, were optimized for performance.

A modified autoencoder was implemented with the same architecture as the VAE, adding Gaussian noise (with a standard deviation of 0.1) to the latent code before decoding. This setup evaluated the impact of noise on reconstruction quality.

To assess performance quantitatively, suitable measures included reconstruction loss for accuracy and the Structural Similarity Index (SSIM) for perceptual quality. These metrics helped evaluate how well the VAE generated new data and preserved the integrity of the original shapes [5].

3.2. Training Process

The VAE was trained on the generated image dataset over a specified number of epochs, utilizing the Adam optimizer and the Evidence Lower Bound (ELBO) as the loss function. After initial training, hyperparameters were fine-tuned to enhance model performance, as detailed in Table 1, and training versus validation loss curves were plotted, as shown in Fig. 1. New images were subsequently generated by sampling from the latent space.

The VAE was then modified into a basic autoencoder that employed Mean Squared Error (MSE) as the loss function (Fig. 2), while ensuring that the latent layer's distribution remained independent and identically distributed (iid) Gaussian. Gaussian noise, with a fixed variance of 0.1, was introduced into the latent layer, allowing for an exploration of its impact on reconstruction quality. Additionally, the information passing through the latent layers was estimated in bits [4].

Finally, the losses, Structural Similarity Index (SSIM), interpolation analyses, and t-SNE visualizations were calculated for both the VAE and the modified autoencoder, enabling a clear distinction between their respective performances.

3.3. Results

After training for 50 epochs, the standard VAE achieved a validation loss of 12.2581, while the modified autoencoder demonstrated a significantly lower loss of 0.5018. The average information passing through the latent layer of the modified VAE was 11.8005 bits, indicating that it retained a substantial amount of relevant information, which facilitated the effective reconstruction of the input shapes.

The Mean Squared Error (MSE) was 0.00065 for the modified VAE, whereas the standard VAE achieved a slightly lower MSE of 0.00051. In terms of the Structural Similarity Index (SSIM), the modified VAE recorded a value of 0.9934, in comparison to 0.9931 for the standard VAE.

Additionally, the VAE exhibited smooth transitions between generated images during interpolation in the latent space, resulting in more natural image variations, Fig. 7. In contrast, the basic autoencoder showed less smooth interpolation, leading to less natural transitions between images, Fig. 8.

Furthermore, t-SNE visualizations revealed that the standard VAE closely resembled the distribution of real images, while the modified VAE exhibited a less distributed representation that did not align with the distribution of the real images, as per Fig. 9 and Fig. 10.

4. CONCLUSION

From the conducted experiments, several key conclusions were drawn:

In terms of reconstruction, basic autoencoders demonstrated a slight advantage in reproducing images that closely resembled the input, maintaining reasonable reconstruction accuracy, as illustrated in Fig. 3 and Fig. 4. However, Variational Autoencoders (VAEs) excelled in generating new images due to their probabilistic latent space, which allowed for sampling from the learned distribution to produce diverse outputs, as illustrated in Fig. 5 and Fig. 6.

t-SNE visualizations revealed that the standard VAE exhibited a more distributed latent space, closely aligning with the distribution of real images. In contrast, the modified VAE's latent space was less distributed, lacking a similar representation.

VAEs offered smooth interpolations in their continuous latent space, resulting in coherent transitions between images. In contrast, basic autoencoders' discrete latent space led to more abrupt transitions, underscoring VAEs' suitability for tasks requiring nuanced outputs.

In summary, VAEs proved better suited for interpolation tasks and generating diverse outputs, while basic autoencoders were suitable for reconstructing images from original inputs. The choice between models ultimately depended on the specific requirements for shape generation quality and computational efficiency.

5. STATEMENT OF TOOLS USED

The tools used in this study include PyTorch for autoencoders implementation, PIL for generating shapes, matplotlib for visualization, and Google Colab for executing the code. The code was written by me, and I verify that it is my original work. The code can be accessed at the following link: *Final-Code*.

6. REFERENCES

[1] Max Welling. Diederik P Kingma. Auto-encoding variational bayes. International Conference on Learning Representations, 2014.

[2] G. E. Hinton* and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. www.science.org, 2006.

[3] Prof Bastiaan Kleijn. Basic autoencoder. Victoria University Lecture in L9_autoEnc.pdf, 2024.

[4] Prof Bastiaan Kleijn. Variational autoencoder (vae): generative structure. Victoria University Lecture in L9_autoEnc.pdf, 2024.

[5] Chris Varano. Disentangling variational autoencoders for image classification. Stanford, 2017.

7. APPENDIX

Table 1: VAE Parameters Manual Optimization.

Parameter	Values	Best Value
epochs	25, 50, 100	50
latent_dim	3, 4, 8, 16, 32	3
learning_rate	0.01, 0.001, 0.0001	0.001
batch_size	32, 64, 128	64

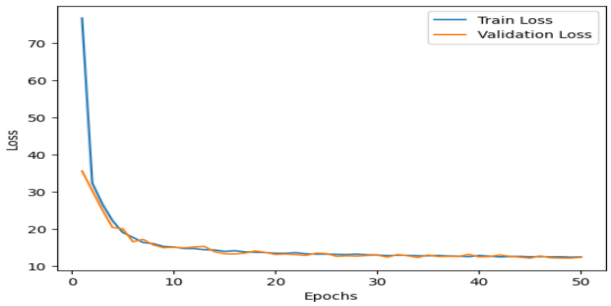

Fig. 1: Training vs Validation Loss Curve for VAE.

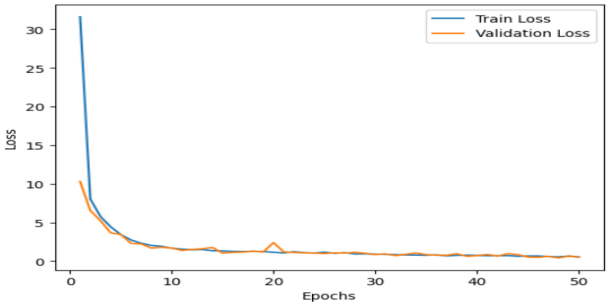

Fig. 2: Training vs Validation Loss Curve for Modified Autoencoder.

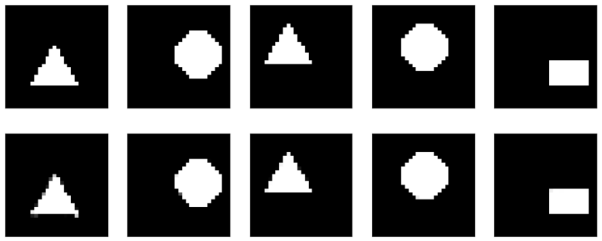

Fig. 3: Original vs Reconstructed Shapes for VAE.

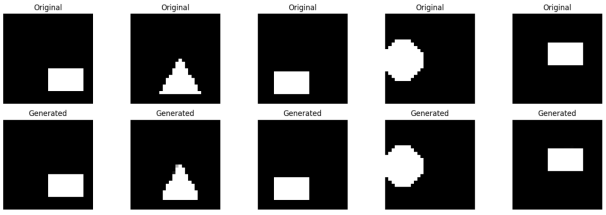

Fig. 4: Original vs Reconstructed Shapes for Modified Autoencoder.

Fig. 5: Newly Generated Images for VAE.

Fig. 6: Newly Generated Images for Modified Autoencoder.

Fig. 7: Interpolation between 2 latent vectors for VAE.

Fig. 8: Interpolation between 2 latent vectors for Modified Autoencoder.

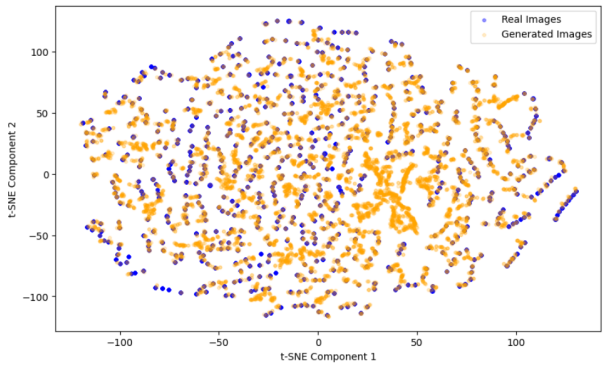

Fig. 9: t-SNE Visualization for VAE.

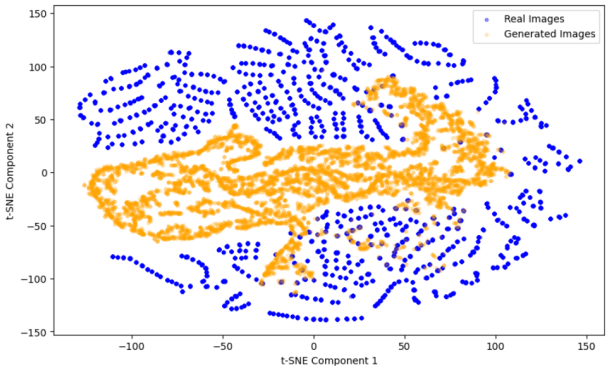

Fig. 10: t-SNE Visualization for Modified Autoencoder.