Generative Modelling of Shapes using Variational Autoencoders

Noppachon Chaisongkhram
School of Engineering and Computer Science, Victoria University of Wellington
chaisonopp@myvuw.ac.nz

***Abstract*—This report implements and analyses a Variational Autoencoder (VAE) for generative modelling of synthetic shape data. A convolutional VAE was trained to learn a compressed, probabilistic latent representation of images containing circles, triangles, and rectangles. The model successfully generated coherent shapes from its structured latent space. A modified VAE, enforcing a strict Gaussian latent distribution with added noise, was compared to the standard model. An information-theoretic analysis estimated ∼10 bits of information traversed the modified model's latent layer. The standard VAE achieved superior reconstruction fidelity (BCE: 0.0346) compared to the modified model (BCE: 0.0968), demonstrating the trade-off between reconstruction quality and latent space organisation. Code is publicly accessible via GitHub and Google Colab.**

***Index Terms*—Deep Learning, Generative Modelling, Variational Autoencoders, Latent Space, Information Theory, Kullback-Leibler Divergence**

I. Introduction

Generative modelling aims to learn the underlying probability distribution of data, enabling the synthesis of novel, realistic samples. Autoencoders provide a framework for learning compressed data representations but often lack a structured latent space suitable for generation. Variational Autoencoders (VAEs) address this limitation by introducing a probabilistic formulation, enforcing a known prior distribution (e.g., Gaussian) on the latent space [1]. This facilitates smooth interpolation and sampling. This report details the implementation of a VAE for a synthetic dataset of 2D shapes, investigating the VAE's core mechanics, its performance as a generative model, and analyses the information flow through its bottleneck. A modified VAE is constructed to dissect the components of the VAE objective function, and a comparative analysis is performed.

II. Theory

A. Problem Formulation

Let $\mathbf{x} \in \mathbb{R}^{784}$ be a flattened 28×28 greyscale image. The goal is to learn the true data distribution $p(\mathbf{x})$. A VAE achieves this by introducing a latent variable $\mathbf{z} \in \mathbb{R}^d$ where $d \ll 784$ and defining the generative process as:

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \int p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \tag{1}$$

where $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the prior and $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ is the probabilistic decoder governed by parameters $\boldsymbol{\theta}$ [1].

B. Variational Inference and the ELBO

The posterior $p(\mathbf{z}|\mathbf{x})$ is intractable. VAEs use variational inference, introducing an approximate posterior $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ (the encoder) parameterised by $\boldsymbol{\phi}$. To train the model, we maximise the Evidence Lower BOund (ELBO) [1, 2]:

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) \geq \text{ELBO} = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})] - D_{\text{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})) \tag{2}$$

The first term is the reconstruction loss, encouraging decoded samples to match the input $\mathbf{x}$. For binary image data, this is implemented as binary cross-entropy. The second term is the KL divergence, a regulariser that pushes the encoder's distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ towards the prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, ensuring a structured and continuous latent space.

C. Reparameterisation

To allow gradient-based optimisation through the stochastic sampling operation $\mathbf{z} \sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$, reparameterisation is used:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{3}$$

This separates the stochastic node from the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ (outputs of the encoder), enabling backpropagation [1].

D. Information Theory

The mutual information $I(X; Z)$ between the input X and latent representation Z measures the information transmitted through the bottleneck. For a Gaussian channel where $Z = Y + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma_n^2$, the information is:

$$I(Y; Z) = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_y^2}{\sigma_n^2}\right) \tag{4}$$

bits [3]. In the modified VAE, adding noise of variance σ_n^2 to a latent variable with signal variance σ_y^2 allows for an estimate of the information rate, providing insight into what the model has learned to preserve.

III. Experiments

A. Experimental Setup

A dataset of 15,000 28×28 pixel images was generated programmatically. Each image contained a single white shape (circle, triangle, or rectangle) on a black background. Shapes were assigned random sizes and positions. The dataset was

split into training (70%), validation (15%), and test (15%) sets, ensuring balanced class representation.

The model architecture consists of the encoder and decoder. The encoder contains two convolutional layers (32 and 64 filters, 3x3 kernels, stride 2) with ReLU activation, followed by a dense layer (16 units, ReLU). The network outputs the mean μ and log-variance log $\boldsymbol{\sigma}^2$ for an 8-dimensional latent space. The decoder contains a dense layer projects the latent vector $\mathbf{z}$ to a 7x7x64 feature map. Two transposed convolutional layers (64 and 32 filters, 3x3 kernels, stride 2) with ReLU activation upsample the features. A final transposed convolution with a sigmoid activation outputs the reconstructed image.

The standard VAE was trained for 50 epochs using the Adam optimiser (learning rate is $1e-3$) with a batch size of 128. The loss was the negative ELBO. The modified VAE used a higher KL loss weight ($\beta = 10$) and added Gaussian noise ($\sigma = 0.5$) to the latent vector before decoding. It was trained for 30 epochs.

B. Results and Analysis

The standard VAE training converged stably. The final test binary cross-entropy (BCE) was 0.0346. Sampling from the 2D latent subspace revealed a smooth transition between different shapes, confirming the latent space was well-organised and continuous, enabling meaningful generation. The modified VAE, with its strong KL constraint and added noise, achieved a higher test BCE of 0.0968, indicating worse reconstruction fidelity. This is the expected trade-off: enforcing a "clean" latent distribution comes at the cost of detail preservation. The information passing through its noisy latent channel was estimated to be 10.93 bits, quantifying the information content of the compressed shape representation.

IV. Conclusion

The project successfully demonstrated the principles of variational autoencoders. The standard VAE learned a generative model capable of producing novel shapes from a structured latent space. The analysis of the modified VAE confirmed the theoretical tension between reconstruction accuracy and latent space regularisation. The information-theoretic estimate provided a quantitative measure of the information content within the latent representations. The results underscore the VAE's effectiveness as a generative model and the critical balance governed by its objective function.

V. Statement

This work represents my original implementation. The solution was developed using TensorFlow, TensorFlow Probability, NumPy, Matplotlib, and Scikit-Learn within PyCharm Jupyter Notebook. The complete code, including data generation and visualisation routines, is accessible via GitHub and Google Colab at:

- https://github.com/jescsk/AIML425/tree/main/Assignment3
- https://colab.research.google.com/drive/1K2tiWDSMvrMjSDTjveXFMThikQiaIOgp?usp=sharing

References

[1] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv.org, Dec. 20, 2013. https://arxiv.org/abs/1312.6114

[2] C. Bishop, Pattern recognition and machine learning. Springer Verlag, 2006.

[3] T. M. Cover and J. A. Thomas, Elements of information theory. Hoboken, N.J.: Wiley-Interscience, 2006.