technologies

MDPI

Article

Comparing Classical and Quantum Generative Learning Models for High-Fidelity Image Synthesis

Siddhant Jain [1], Joseph Geraci [2,3,4,5]* and Harry E. Ruda [6]

[1] Division of Engineering Science, University of Toronto, Toronto, ON M5S 1A1, Canada
[2] Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON K7L 3N6, Canada; geracij@queensu.ca
[3] Visiting Scientist for Quantum Computation and Neuroscience, University of California San Diego, La Jolla, CA 92093, USA
[4] Center for Biotechnology and Genomics Medicine, Medical College of Georgia, Augusta, GA 30912, USA
[5] Chief Technology Officer, NetraMark Holdings, Toronto, ON M6P 3T1, Canada
[6] Stanley Meek Chair in Nanotechnology, Centre for Nanotechnology, Center for Quantum Information and Quantum Control, Department of Electrical Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada; harry.ruda@utoronto.ca
* Correspondence: siddhant.jain@utoronto.ca

Abstract: The field of computer vision has long grappled with the challenging task of image synthesis, which entails the creation of novel high-fidelity images. This task is underscored by the Generative Learning Trilemma, which posits that it is not possible for any image synthesis model to simultaneously excel at high-quality sampling, achieve mode convergence with diverse sample representation, and perform rapid sampling. In this paper, we explore the potential of Quantum Boltzmann Machines (QBMs) for image synthesis, leveraging the D-Wave 2000Q quantum annealer. We undertake a comprehensive performance assessment of QBMs in comparison to established generative models in the field: Restricted Boltzmann Machines (RBMs), Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Denoising Diffusion Probabilistic Models (DDPMs). Our evaluation is grounded in widely recognized scoring metrics, including the Fréchet Inception Distance (FID), Kernel Inception Distance (KID), and Inception Scores. The results of our study indicate that QBMs do not significantly outperform the conventional models in terms of the three evaluative criteria. Moreover, QBMs have not demonstrated the capability to overcome the challenges outlined in the Trilemma of Generative Learning. Through our investigation, we contribute to the understanding of quantum computing's role in generative learning and identify critical areas for future research to enhance the capabilities of image synthesis models.

Keywords:

Citation: Jain, S. ; Geraci, J.; Ruda, H.E. Comparing Classical and Quantum Generative Learning Models for High-Fidelity Image Synthesis. *Technologies* **2023**, *1*, 0. https://doi.org/

Academic Editor: Pedro Antonio Gutiérrez

Received: 27 September 2023
Revised: 11 November 2023
Accepted: 21 November 2023
Published:

1. Introduction

1.1. Generative Modeling

Generative modeling is a class of machine learning that aims to generate novel samples from an existing dataset. Image synthesis is a subset of generative modeling applications relating to the generation of novel high-fidelity images that mimic an underlying distribution of images, known as the training set. The main types of generative models are Generative Adversarial Networks (GANs), probabilistic models, and Variational Autoencoders (VAE), all of which are capable of high-fidelity image synthesis. In 2020, a new methodology for producing image synthesis using diffusion models was shown to produce high-quality images [1]. In 2021, OpenAI demonstrated Denoising Diffusion Probabilistic Models' (DDPM) superiority in generating higher image sample quality than the previous state-of-the-art GANs [2].

Quantum annealers, namely the D-Wave 2000Q, have also been shown to perform generative modeling with varied success [3,4]. By taking advantage of quantum sampling

and parallelization, D-Wave 2000Q can hold an embedding of the latent space relating to a set of training data in an architecture of coupled qubits [5]. There are still significant research gaps relating to utilizing generative modeling on the quantum processing unit for image synthesis, especially as it relates to measuring their performance against other generative models on standard scoring methods, namely the Inception score, FID, and KID. This research aims to close this gap by investigating the efficacy of the D-Wave 2000Q quantum annealer on the problem of image synthesis.

1.2. Trilemma of Generative Learning

Xiao et al. describe the Trilemma of Generative Learning as the inability of any single deep generative modeling framework to solve the following requirements for wide adoption and application of image synthesis: (i) high-quality sampling, (ii) mode coverage and sample diversity, and (iii) fast and computationally inexpensive sampling [6]. Current research primarily focuses on high-quality image generation and ignores the real-world sampling constraints and the need for high diversity and mode coverage. Fast sampling allows for the generative models to be utilized in greater fast-learning applications, which require quick image synthesis, e.g., interactive image editing [6]. Diversity and mode coverage ensure generated images are not direct copies of, but are also not significantly skewed from, the training data.

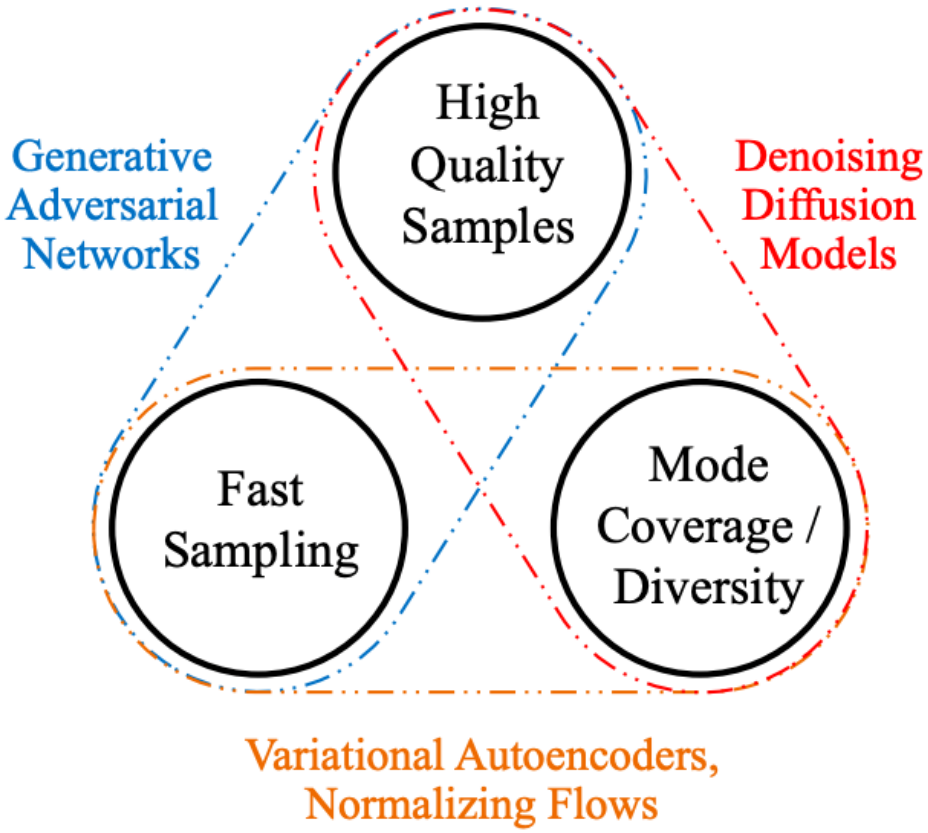


Figure 1. Generative Learning Trilemma [6]. Labels show frameworks that tackle two of the three requirements well.

This paper reviews research that aims to tackle this trilemma with the D-Wave quantum annealer and attempts to determine the efficacy of modeling on the three axes of the trilemma. In doing so, the success of the quantum annealer will be tested against other classical generative modeling methodologies. Success in showing the quantum annealer's ability to produce (i) high-quality images, (ii) mode coverage and diversity, and (iii) fast

sampling will demonstrate the supremacy of quantum annealers over classical methods for the balanced task of image synthesis.

2. Background

The trajectory of artificial intelligence in the domain of image synthesis, evolving from Restricted Boltzmann Machines (RBMs) to Denoising Diffusion Probabilistic Models (DDPMs), marks a significant technical progression. This advancement, intermediated by Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), has driven improvements in the fidelity, diversity, and realism of generated images, while also introducing a host of model-specific challenges and computational complexities.

Before exploring generative modeling within quantum computing environments, let us provide background into classical image synthesis models, namely RBMs, VAEs, GANs, and DDPMs.

Following this, we will delve into the research of quantum annealing and its application in machine learning. The ultimate goal is to create a blueprint for image synthesis on a quantum annealer.

2.1. Classical Image Synthesis

2.1.1. Restricted Boltzmann Machine

Boltzmann Machines are a class of energy-based generative learning models. A Restricted Boltzmann Machine, a subset of Boltzmann Machines, is a fully connected bipartite graph that is segmented into visible and hidden neurons.

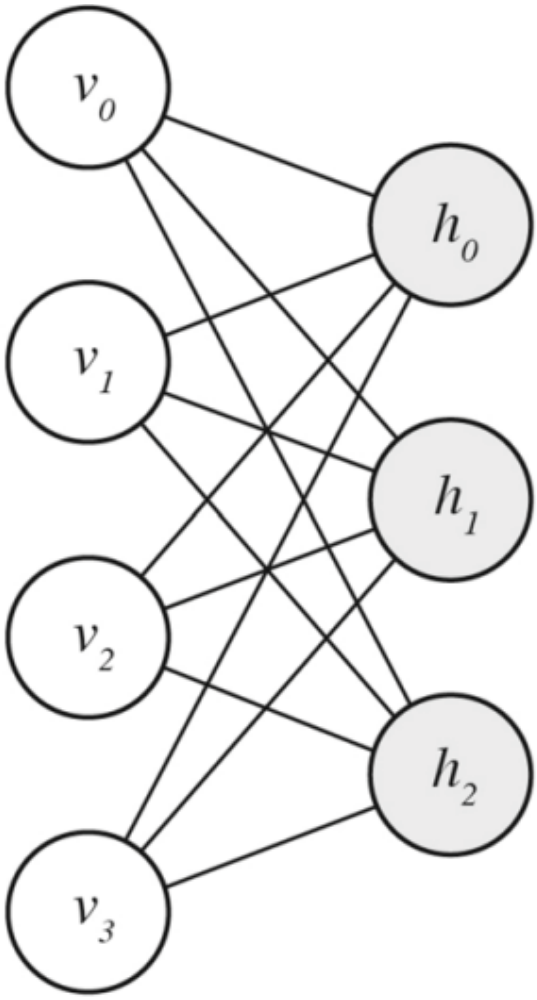


Figure 2. Restricted Boltzmann Machine architecture [3].

RBMs are generative models that embed the latent feature space in the weights between the visible and hidden layers. RBMs were first introduced in 1986 by Smolensky and were further developed by Freund and D. Haussler in 1991 [7,8]. The energy function to minimize when training an RBM is the following [9]:

$$E(v,h) = -a^T v - b^T h - v^T W h \tag{1}$$

Training is the process of tuning the weights matrix W and bias vectors a and b on the visible v and hidden h layers, respectively. v represents the visible units, i.e., the observed values or a training sample. The network assigns a probability to every possible pair of a visible and a hidden vector via this energy function [10]:

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{2}$$

Z is the partition function given by summing over all possible pairs of v and h [10]. Thus, the probability of a given v is:

$$p(v) = \frac{1}{Z} \sum_h e^{-E(v,h)} \tag{3}$$

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{4}$$

The difficulty in evaluating the partition function Z introduces the need to use Gibbs sampling with Contrastive Divergence Learning, introduced by Hinton et al. in 2005 [11]. By utilizing such methods, one can train the RBM quickly via gradient descent, similar to other neural networks. By adding more hidden layers, a deeper embedding can be captured by the model; such a system is called a Deep Belief Network (DBN).

RBMs, while of little note in the modern landscape of machine learning research due to their limited performance and relatively slow training times, are of particular note to this research, as they have direct parallels with both the architecture of the D-Wave 2000Q quantum processor and the method by which they reduce the total energy of their respective systems. RBMs also have limited applications in computer vision but were an important advancement in the field of generative modeling as a whole.

2.1.2. Variational Autoencoder

A Variational Autoencoder (VAE) is a generative machine learning model developed in 2013 composed of a neural network that is able to generate novel high-fidelity images, texts, sounds, etc. [12].

Autoencoders seek to compress an input space into a compressed latent representation from which the original input space can be recovered [12]. Variational Autoencoders improve upon traditional Autoencoders by recognizing the input space has an underlying distribution and seeks to learn the parameters of that distribution [12]. Once trained, VAEs can be used to generate novel data, similar to the input space, by removing the encoding layers and exploring the latent space [12]. Exploring the latent space is simply treating the latent compression layer as an input layer and observing the output of the VAE for various inputs. VAEs marked the first reliable way to generate somewhat high-fidelity images using machine learning [13].

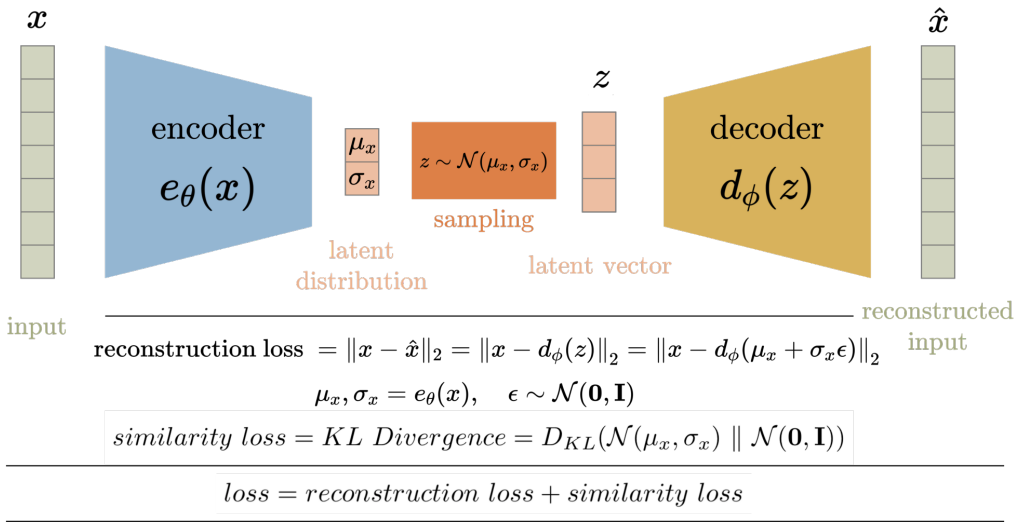


Figure 3. Variational Autoencoder architecture [13].

2.1.3. Generative Adversarial Networks

The most significant development in high-fidelity generative image synthesis was in 2014 with the introduction of GANs by Ian Goodfellow et al. [14]. Goodfellow et al. propose a two-player minimax game composed of a generator model (G) and a discriminator model (D). As the game progresses, both the generator and discriminator models improve.

GANs are trained via an adversarial contest between the generator model (G) and discriminator model (D) [14]. x contains samples from both the training set and p_g, the images generated by G. $D(x;\theta_d)$ outputs the probability that x originates from the training dataset as opposed to p_g. Meanwhile, $G(z,\theta_g)$ outputs p_g given noise z. G's goal is to fool D while D aims to reliably differentiate real training data from data generated by G. The loss function for G is $log(1 - D(G(z)))$. Thus, the value/loss function, error, of a GAN is represented as:

$$\min_{G} \max_{D} V(D,G) = E_{x \sim p_{data}(x)}[logD(x)] + E_{z \sim p_z(z)}[log(1 - D(G(z)))] \quad (5)$$

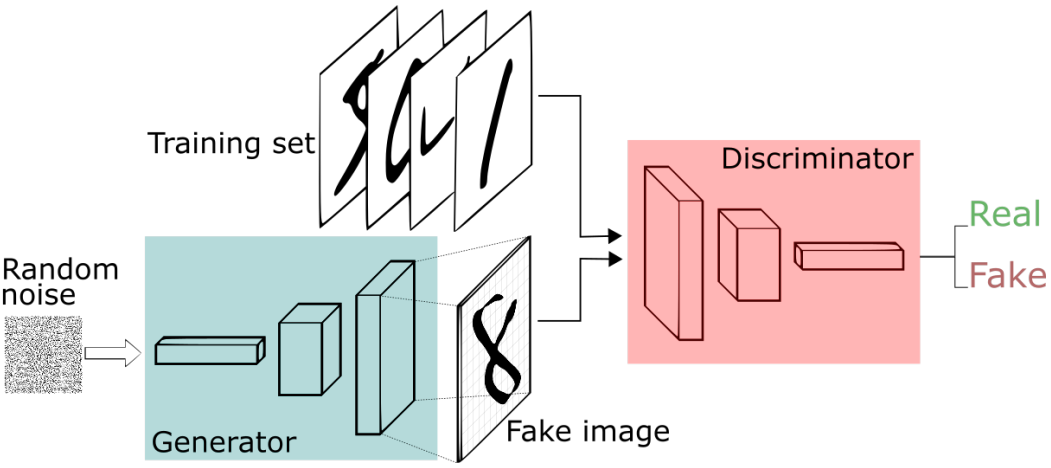


Figure 4. GANs architecture [15].

Both G and D are trained simultaneously. This algorithm allows for lock-step improvements to both G and D. Towards the conclusion of training, G becomes a powerful image generator, which closely replicates the input space, i.e., training data.

GANs have several shortcomings that make them difficult to train. Due to the adversarial nature of GANs, training the model can face the issue of Vanishing Gradients, when the discriminator develops more quickly than the generator, consequently correctly

predicting every x and leaving no error to train on for the generator [16]. Another common issue is Mode Collapse, when the generator learns to generate a particularly successful x such that the discriminator is consistently fooled and the generator continues to only produce that singular x and has no variability in image generation [16]. Both Vanishing Gradients and Mode Collapse are consequences of one of the adversarial models improving faster than the other.

2.1.4. Denoising Diffusion Probabilistic Model

DDPMs are a recent development proposed by Jonathan Ho et al. (2020) inspired by nonequilibrium thermodynamics that produces high-fidelity image synthesis using a parameterized Markov chain [1]. Beginning with the training sample, each step of the Markov chain adds a single layer of Gaussian noise. A neural network is trained on parameterizing these additional Gaussian noise layers to reverse the process from random noise to a high-fidelity image.

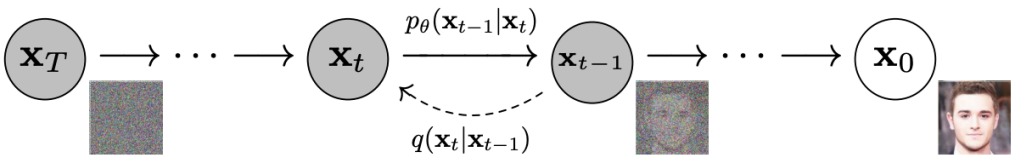


Figure 5. DDPM Markov chain [1].

$q_\theta(x_t|x_{t-1})$ represents the forward process, adding Gaussian noise, and $p_\theta(x_{t-1}|x_t)$ represents the reverse process, denoising. The reverse process is captured by training.

$$p_\theta(x_0) := \int p_\theta(x_{0:T})dx_{1:T} \tag{6}$$

where

$$p_\theta(x_{0:T}) := p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x) \tag{7}$$

and

$$p_\theta(x_{t-1}|x) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t); \Sigma_\theta(x_t, t)) \tag{8}$$

For clarity, we remind the reader that $\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t); \Sigma_\theta(x_t, t))$ is the normal distribution with mean $\mu_\theta(x_t, t)$ and covariance matrix $\Sigma_\theta(x_t, t)$. The loss function for a DDPM is as follows:

$$L := \mathbf{E}_q[-\log p(x_T) - \sum_{t\geq 1} \log \frac{p_\theta(x_{t-1}|x)}{q_\theta(x|x_{t-1})}] \tag{9}$$

Using a U-Net and a CNN with upsampling, with stochastic gradient descent and $T = 1000$, Ho et al. were able to generate samples with an impressive, but not state-of-the-art, FID score of 0.317 on the CIFAR10 dataset. On CelebA-HQ 256×256, the team generated the novel images in Figure 6.

Figure 6. Generated samples on CelebA-HQ 256 × 256 by DDPM [1].

In 2021, Dhariwal et al. at OpenAI made improvements upon the original DDPM parameters, and it achieved state-of-the-art FID scores of 2.97 on ImageNet 128 × 128, 4.59 on ImageNet 256 × 256, and 7.72 on ImageNet 512 × 512 [2].

The first improvement is not to set $\Sigma_\theta(x_t, t)$ as a constant but rather as the following:

$$\Sigma_\theta(x_t, t) = \exp(v\log\beta_t + (1-v)\log\tilde{\beta}_t) \tag{10}$$

where β_t and $\tilde{\beta}_t$ correspond to the upper and lower bounds of the Gaussian variance.

Dhariwal et al. also explore the following architectural changes; note: attention heads refer to embedding blocks in the U-Net [2]:

- Increasing depth versus width, holding model size relatively constant.
- Increasing the number of attention heads.
- Using attention at 32 × 32, 16 × 16, and 8 × 8 resolutions rather than only at 16 × 16.
- Using the BigGAN residual block for upsampling and downsampling the activations.
- Rescaling residual connections with $\frac{1}{\sqrt{2}}$.

With these changes, Dhariwal et al. were able to demonstrate their DDPM beating GANs in every single class by FID score and establishing DDPMs as the new state-of-the-art for image synthesis [2].

2.2. Quantum Machine Learning

2.2.1. Quantum Boltzmann Machine

Energy-based machine learning models, such as the RBM, seek to minimize an energy function. Recall:

$$p(v) = \frac{\sum_h e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}} \tag{11}$$

is maximized when $E(v, h)$ is minimized.

$$E(v, h) = -a^T v - b^T h - v^T W h \tag{12}$$

or, in its expanded form

$$E(v, h) = -\sum_i v_i \cdot a_i - \sum_j h_j \cdot b_j - \sum_i \sum_j v_i \cdot W_{ij} \cdot h_j \tag{13}$$

Recall also that this energy function is intractable for all v and h, thus RBMs are trained via Contrastive Divergence [17].

The D-Wave 2000Q via the Ising model is able to minimize an energy function via coupled qubits, taking advantage of entanglement. The energy function for the Ising model is the following Hamiltonian:

$$E_{\text{ising}}(\mathbf{s}) = \sum_{i=1}^{N} h_i s_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} J_{i,j} s_i s_j \tag{14}$$

The $s_i \in \{-1, +1\}$ represents the qubit spin state, with spin up and spin down effectively. h_i is the bias term provided by the external magnetic field, and $J_{i,j}$ captures the coefficients for the coupling between qubits [18]. Solving for the ground state of an Ising model is NP-hard, but by taking advantage of the QPU's ability to better simulate quantum systems, we can solve this problem more efficiently [19].

Clamping neurons is the process of affixing certain qubits to specific values, namely the data being trained on. By clamping the neurons v and h onto the qubits, applying an external magnetic field equivalent to the biasing parameters a and b, and setting the coupling parameters to match those of W (and to 0 for absent or intralayer edges), the RBM can be effectively translated into a format suitable for a quantum annealer. The resulting model is known as a Quantum Boltzmann Machine (QBM) and is similarly trained using QPU-specific Gibbs sampling methods [18,20].

Increased sampling from the quantum annealer leads to a more comprehensive representation of the Hamiltonian's energy landscape. The process of training a QBM involves adjusting the couplings based on this acquired information. The D-Wave 2000Q has the qubit coupling architecture in Figure 7.

2.2.2. Image Classification

The field of Quantum Machine Learning (QML) applied to computer vision is still quite nascent. Most QML research focuses on classification tasks, particularly using quantum support vector machines, decision trees, nearest neighbors, annealing-based classifiers, and variation classifiers [21]. Wei et al. propose a Quantum Convolutional Neural Network with capabilities for spatial filtering, image smoothing, sharpening, and edge detection, along with MNIST digit recognition, with a lower computation complexity than classical counterparts [22]. Such research provides a valuable precursor to the exploration of QML for image synthesis.

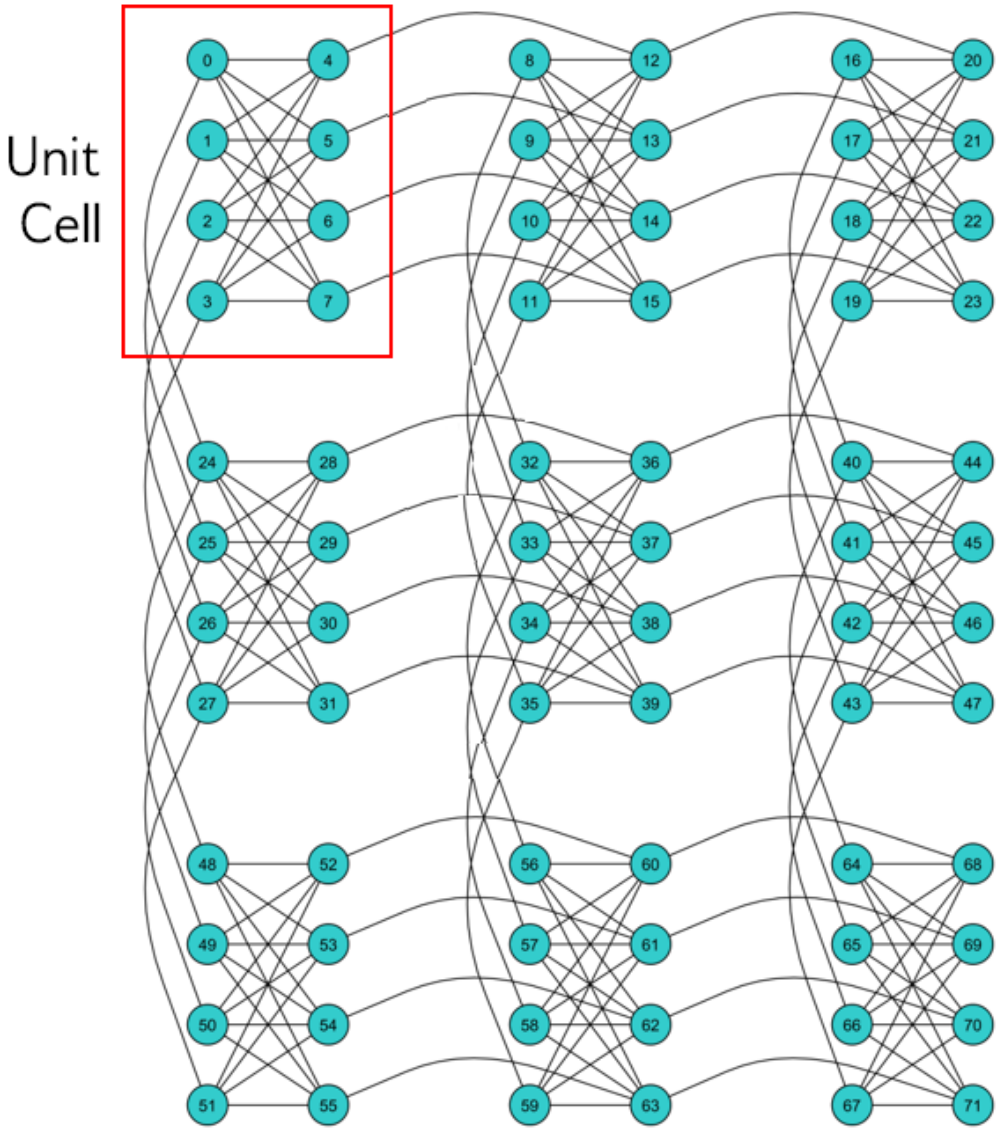


Figure 7. D-Wave Quantum Processing Unit (QPU) topology Chimera graph [18].

2.2.3. Image Synthesis

In 2020, Sleeman et al. demonstrated the D-Wave QUBO's ability to generate images mimicking the MNIST hand-drawn digits and Fashion MNIST datasets [23]. Due to the limited number of qubits available, Sleeman et al. create an encoding of the images via a convolutional autoencoder, feed the encoding to a QBM, and finally reverse the process to perform image synthesis. The model architecture is provided in Figure 8.

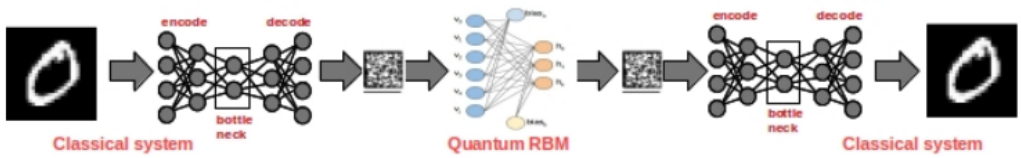


Figure 8. Hybrid Approach that used a Classical Autoencoder to map the image space to a compressed space [23].

In their research, Sleeman et al. contrast the performance of their QBM with that of a traditional RBM, in addition to assessing the efficacy of the autoencoder's encoding

capabilities. Despite showcasing the potential of the D-Wave 2000Q in aiding image synthesis, the authors do not juxtapose their findings with those of other classical generative modeling methods. Furthermore, the omission of FID, KID, and Inception scores for their proposed models restricts the breadth of comparison between the QBM and its classical counterparts.

3. Methods

3.1. Goal

To reiterate, the goal of this research is to train the D-Wave 2000Q quantum annealer on image synthesis (generative image creation) and compare the results both quantitatively and qualitatively against existing classical models. Secondly, the goal is to determine the quantum annealer's efficacy at cracking the challenges outlined in Section 1.2, specifically the Trilemma of Generative Learning.

Additionally, our research aims to close many of the gaps in Sleeman et al.'s study. Namely:

- perform the image synthesis directly on the QBM,
- evaluate the performance of the QBM against a(n) RBM, VAE, GAN, and DDPM,
- evaluate various generative modeling methods on FID, KID, and Inception scores,
- model a richer image dataset, CIFAR-10.

3.2. Data

We utilize a standardized dataset, CIFAR-10, for all of our experiments. The CIFAR-10 dataset consists of sixty thousand 32 by 32 three-channel (color) images in ten uniform classes [24]. The data were initially collected in 2009 by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton and have become the standard for machine learning research relating to computer vision [25]. One of the primary reasons CIFAR-10 is so popular is because the small image sizes allow for quick training and testing of new models [26]. In addition, the ubiquity of testing models on CIFAR-10 allows researchers to quickly benchmark their model performance against prior research [26].

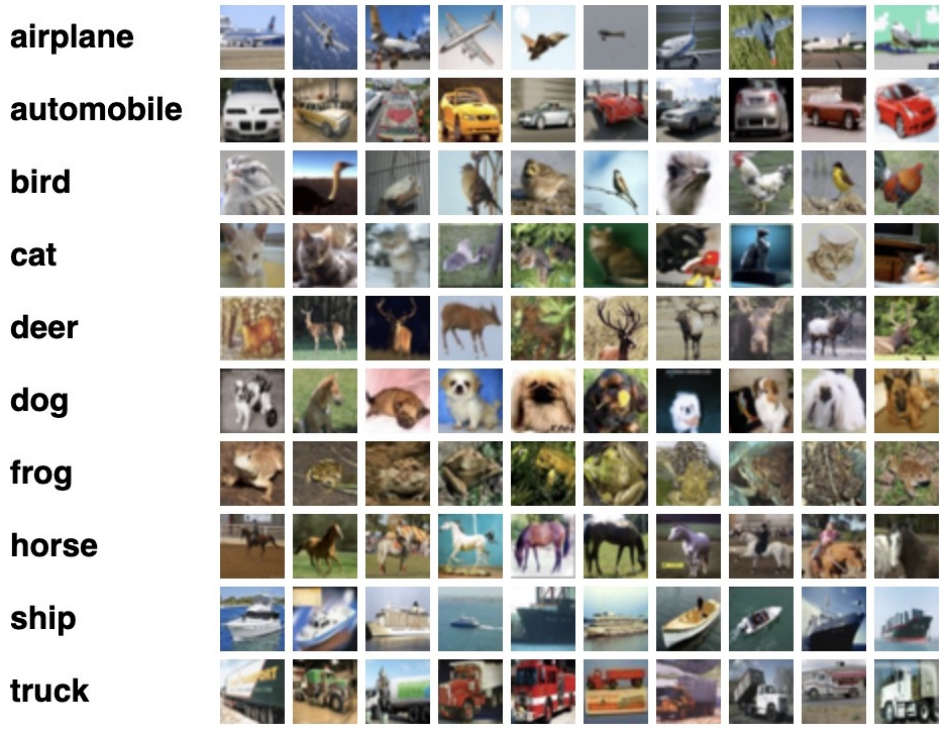


Figure 9. Ten random images from each class of CIFAR-10 with respective class labels [24].

The images in CIFAR-10 are exclusive to photographs of discrete distinct objects on a generally neutral background. The dataset contains photographs, which are two-dimensional projections of three-dimensional objects, from various angles.

3.3. Classical Models

To establish a benchmark and facilitate the comparison of results between novel quantum machine learning methods and existing generative image synthesis techniques, we initially trained and tested a series of classical models on the CIFAR-10 dataset. The classical models we trained on were the following: (i) RBM, (ii) VAE, (iii) GANs, and (iv) DDPM. Initially, we adopted a uniform approach, training each model with the same learning rate, batch size, and number of epochs to standardize results. However, this method led to significant challenges due to the varying rates of convergence among the models, causing an imbalance in result quality and impeding our analysis. Consequently, we adjusted our approach to individually optimize the hyperparameters for each model within the bounds of available time and resources. This adjustment yielded higher-quality results, offering a more equitable comparison across models. We concluded the training of each model when additional epochs resulted in insignificant improvements in model loss, a term left intentionally vague to accommodate training variability across models. An exception to this approach was made for the DDPM, which demanded considerable computational power, prompting us to conclude the experiment after 30,000 iterations.

3.4. Quantum Model

For the quantum model, the training images were also normalized by mean and variance, identically to the preprocessing for the classical models. Since quantum bits can only be clamped to binary values and not floating point numbers, the data also had to be binarized. This process involves converting each input vector into 100 vectors where the representation of 1s in each row reflects the floating point number between 0–1, as pictured in Figure 10.

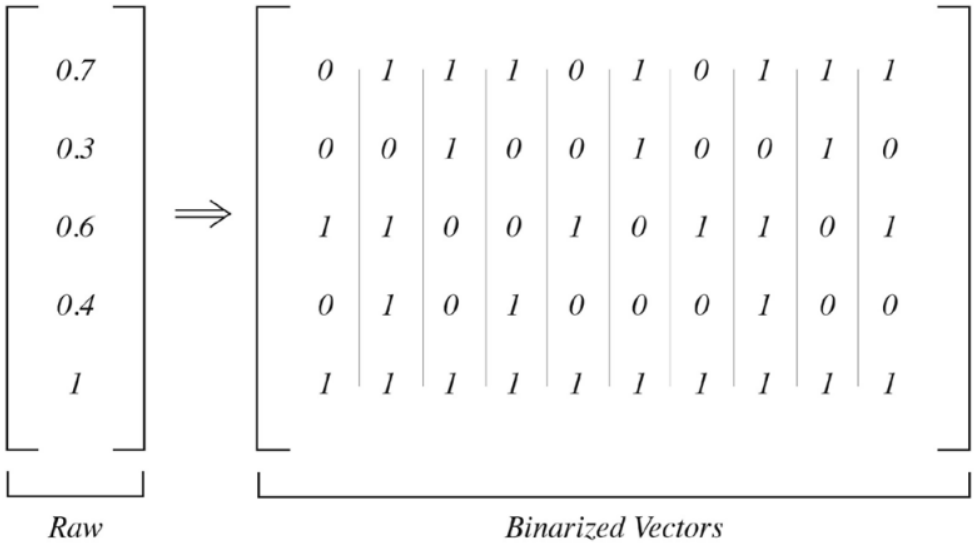


Figure 10. Binarization of a normalized vector to a set of binary vectors [3].

The D-Wave 2000Q quantum annealer is trained by mapping the architecture of an RBM onto the QPU Chimera graph, thus creating a QBM [18]. The visible, i.e., input, nodes are clamped with the training data, and the hidden layer is sampled from. As we increase sampling, we gain a better understanding of the energy landscape and can better update the weights (i.e., inter-qubit coupling coefficients) [3].

Due to limitations with the number of available qubits on the D-Wave 2000Q being 2048, and user resource allocation challenges, our experiments are limited. To resolve this constraint, each image was split into four distinct squares along the x and y axes. Thus, each training image was $16 \times 16 \times 3$ for an input vector size of 768.

3.5. Hyper-Parameters

The hyper-parameters were determined by conducting grid search hyper-tuning. Since DDPMs are trained via an iterative process, unbatched, they require significantly more epochs, as reflected in Table 1.

Table 1. Final hyper-parameters used for respective model training.

	QBM	RBM	VAE	GAN	DDPM
Epochs	10	10	50	50	30,000
Batch Size	256	256	512	128	-
# of Hidden Nodes	128	2500	32	64	32
Learning Rate (10^{-3})	0.0035	0.0035	0.2	0.2	0.2

3.6. Metrics

3.6.1. Inception Score

Inception score measures two primary attributes of the generated images: (i) the fidelity of the images, i.e., the image distinctly belongs to a particular class and (ii) the diversity of the generated images [27]. The Inception classifier is a Convolutional Neural Network (CNN) built by Google and trained on the ImageNet dataset consisting of 14 million images and 1000 classes [28].

(i) Fidelity is captured by the probability distribution produced as classification output by the Inception classifier on a generated image [27]. Note that a highly skewed distribution with a single peak indicates that the Inception classifier is able to identify the image as belonging to a specific class with high confidence. Therefore, the image is considered high fidelity.

(ii) Diversity is captured by summing all the probability distributions produced for individually generated classes. The uniform nature of the resultant sum of distributions is indicative of the diversity of the generated images. E.g., a model trained on CIFAR-10 that only manages to produce high-fidelity images of dogs would severely fail to be diverse.

The average of the K-L Divergences between the produced probability distribution and the summed distribution is the final Inception score, capturing both diversity and fidelity. Rigorously, each generated image x_i is classified using the Inception classifier to obtain the probability distribution $p(y|x_i)$ over classes y [29]. The marginal distributions are provided by:

$$p(y) = \frac{1}{N} \sum_{i=1}^{N} p(y|x_i) \tag{15}$$

From which the K-L Divergence may be computed by the following [29]:

$$D_{\mathrm{KL}}(p(y|x_i)||p(y)) = \sum_{y} p(y|x_i) \log\left(\frac{p(y|x_i)}{p(y)}\right) \tag{16}$$

Take the expected value of these K-L Divergences over all N generated images [29]:

$$\mathbb{E}_x[D_{\mathrm{KL}}(p(y|x)||p(y))] = \frac{1}{N} \sum_{i=1}^{N} D_{\mathrm{KL}}(p(y|x_i)||p(y)) \tag{17}$$

Finally, we exponentiate the value above to evaluate an Inception score [29]:

$$\mathrm{IS}(G) = \exp(E_{x \sim pg} \mathcal{D}_{KL}(p(y|x)||p(y)) \tag{18}$$

3.6.2. Fréchet Inception Distance (FID)

Fréchet Inception Distance improves upon the Inception score by capturing the relationship between the generated images against the training images, whereas the Inception score only captures the characteristics of the generated images against each other and their classifications. The Inception classifier, used to determine the Inception score, also embeds a feature vector. I.e., the architecture of the Inception classifier captures the salient features of the images it is trained on.

The FID score is determined by taking the Wasserstein metric between the two multivariate Gaussian distributions of the feature vectors for the training and generated images on the Inception model [30]. Simply, the dissimilarity between the features found in the training and generated data. This is an improvement upon the Inception score since it captures the higher-level features that would be more human-identifiable when comparing model performance. The Gaussian distributions of the feature vector for the generated images and the training images are $N(\mu, \Sigma)$ and $N(\mu_w, \Sigma_w)$, respectively [31]. The Wasserstein metric, resulting in the FID score, is as follows [31]:

$$\mathrm{FID} = ||\mu - \mu_w||_2^2 + \mathrm{tr}(\Sigma + \Sigma_w - 2(\Sigma^{1/2}\Sigma_w\Sigma^{1/2})^{1/2}) \tag{19}$$

3.6.3. Kernel Inception Distance (KID)

KID measures the maximum mean discrepancy of the distributions of training and generated images by randomly sampling from them both [32]. KID does not specifically account for differences in high-level features and rather compares the raw distributions more directly.

Specifically, for generator X with probability measure $\mathbb{P}$ and random variable Y with probability measure $\mathbb{Q}$, we have [32]:

$$\mathcal{D}_F(\mathbb{P}, \mathbb{Q}) = \sup_{f \in F} \mathbb{E}_{\mathbb{P}} f(X) - \mathbb{E}_{\mathbb{Q}} f(Y) \tag{20}$$

3.6.4. Quantitative Metrics

The following table summarizes the three quantitative metrics used to evaluate model performance:

Table 2. Summary of quantitative metrics for generative image synthesis evaluation.

Metric	Description	Performance
Inception	K-L Divergence between conditional and marginal label distributions over generated data	Higher is better
FID	Wasserstein distance between multivariate Gaussians fitted to data embedded into a feature space	Lower is better
KID	Measures the dissimilarity between two probability distributions P_r and P_g using samples drawn independently from each distribution	Lower is better

3.6.5. Qualitative Metrics

Our qualitative evaluation was performed by analyzing the visual discernment of generated images in relation to their respective classes less stringently. This approach aims to foster a broader discussion about the applicability of such models and their effectiveness.

4. Results

4.1. Restricted Boltzmann Machine (RBM)

The generated images by the RBM include a high degree of brightly-colored noise. Interestingly this noise is concentrated in sections of the image with high texture, i.e., high variance of pixel values. Notice the image of the cat in the bottom-center on Figure 11b has a great deal of noise at the edges of and inside the boundaries of the cat itself, but not in the blank white space surrounding it. This demonstrates a high degree of internode interference in the hidden layer. That is, areas with large pixel variance influence the surrounding pixels greatly and often cause bright spots to appear as a result.

4.2. Variational Autoencoder (VAE)

The generated images from the VAE are incredibly high fidelity. Notably, the VAE results liken superresolution. Notice the decrease in image blur/noise from the input images. Since the VAE encodes an embedding of the training data, some features, such as the exact color of the vehicle in the top left corner in Figure 12b, are lost, but the outline of the vehicle and the background are sharpened. This demonstrates the VAE is capturing features exceptionally well.

4.3. Generative Adversarial Networks (GANs)

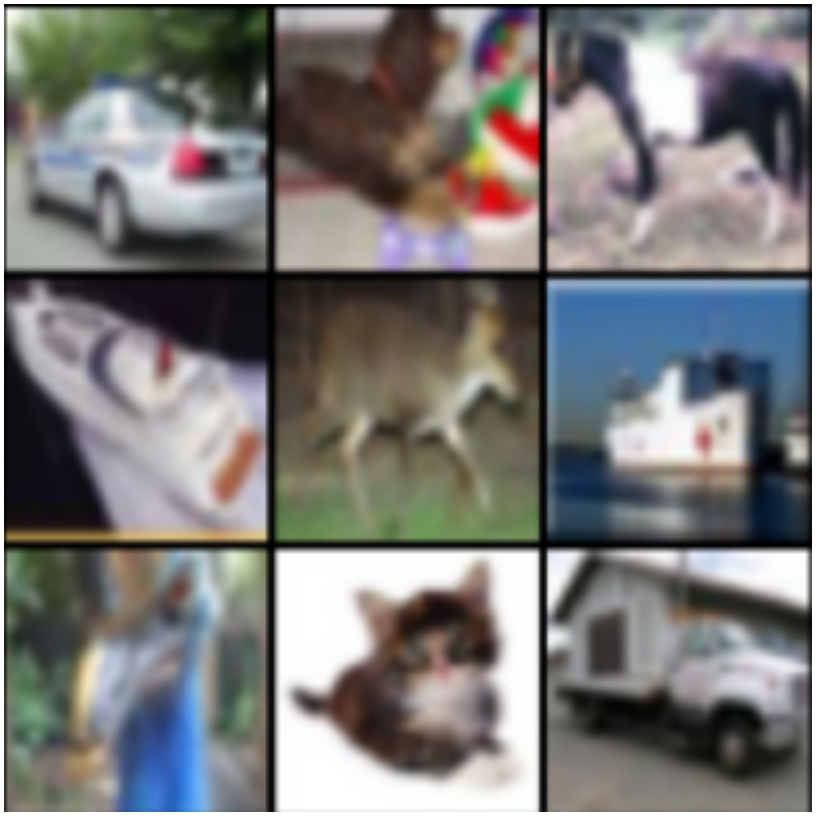
The GAN is able to produce some images with high fidelity, namely the cat in the top left corner and the dog in the bottom right corner of Figure 13b, but struggles with the sharpness of the images. Humans looking at the majority of the images produced could easily determine they are computer generated. In addition, the GAN was uniquely difficult to train, requiring retraining dozens of times in order to avoid Vanishing Gradients and Mode Collapse. Recall from Section 2.1.3, Vanishing Gradients and Mode Collapse are issues that arise from the discriminator or generator improving significantly faster than the counterpart and dominating future training, thus failing to improve both models adequately and defeating the adversarial training nature of the network.

4.4. Denoising Diffusion Probabilistic Model (DDPM)

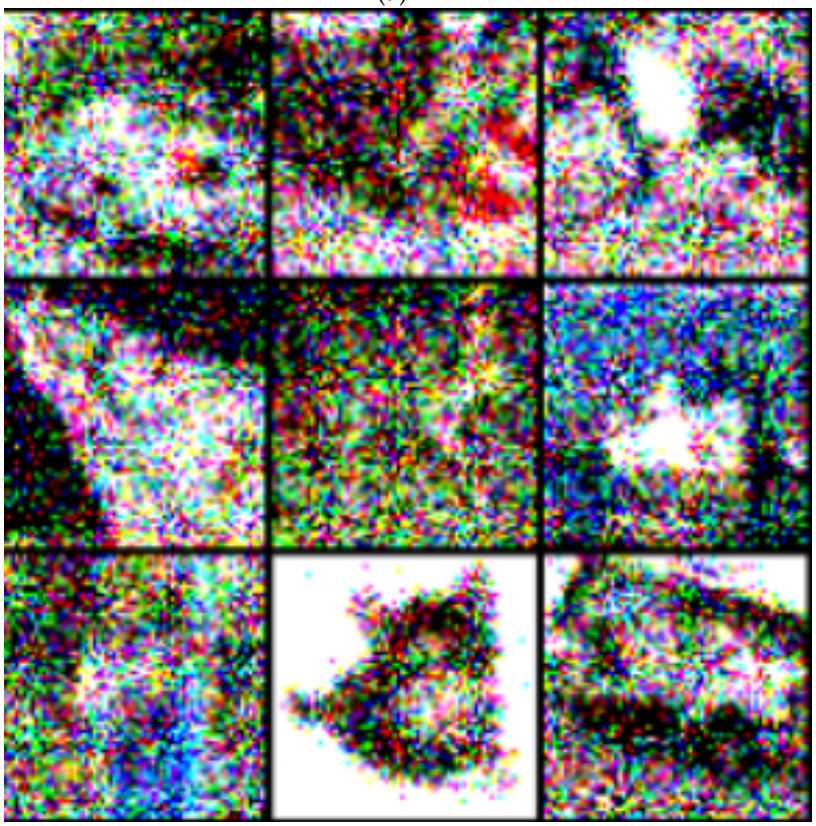
The quality of the results for the DDPM are limited by the computational power available to run the experiment. DDPMs are state-of-the-art for image generation when scored on fidelity but require several hours of training on a Tensor Processing Unit (TPU). A TPU can perform one to two orders of magnitude more operations than an equivalent GPU per second [33]. Without access to these Google-exclusive TPUs, we were unable to replicate state-of-the-art generation results.

4.5. Quantum Boltzmann Machine (QBM)

Recall the QBM required training images to be split and restitched into four independent squares for training due to qubit limitations. This splitting and restitching has a distinct influence on the resultant generated images. Notice the generated images have distinct features in each quadrant of the image. These features are often from various classes and appear stitched together because they are. Notice how the image in the bottom row, second from the rightmost column has features of a car, a house, and a concrete background.



(**a**)



(**b**)

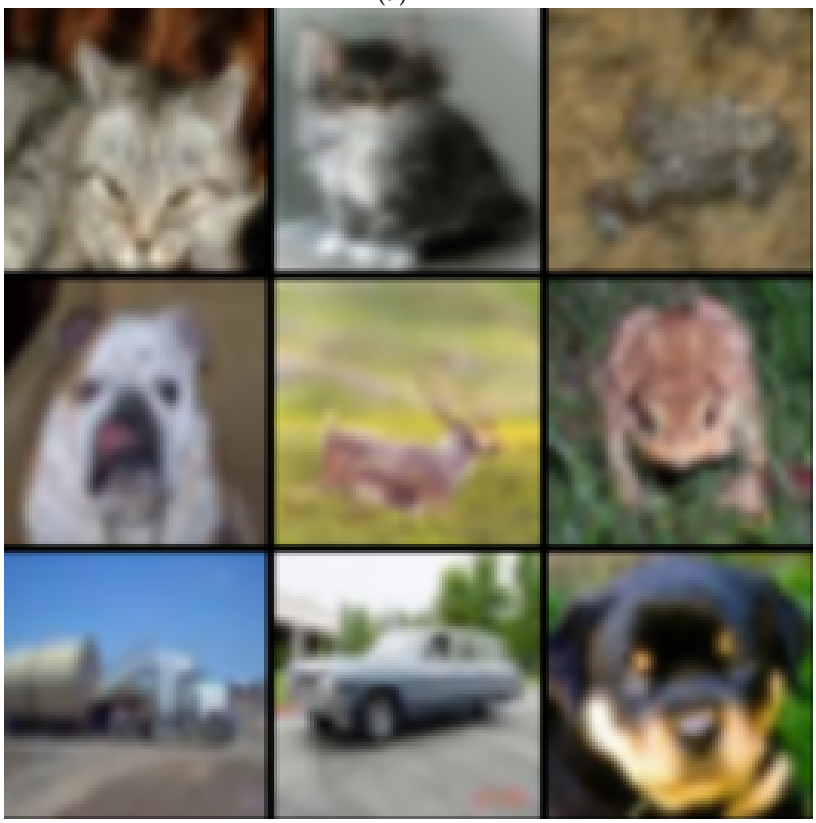
Figure 11. RBM-generated image synthesis output from respective input. (**a**) RBM input images; (**b**) RBM output images.

(**a**)

(**b**)

Figure 12. VAE-generated image synthesis output from respective input. (**a**) VAE input images; (**b**) VAE output images.

(**a**)



(**b**)

Figure 13. GAN-generated image synthesis output from respective input. (**a**) GAN input images; (**b**) GAN output images.

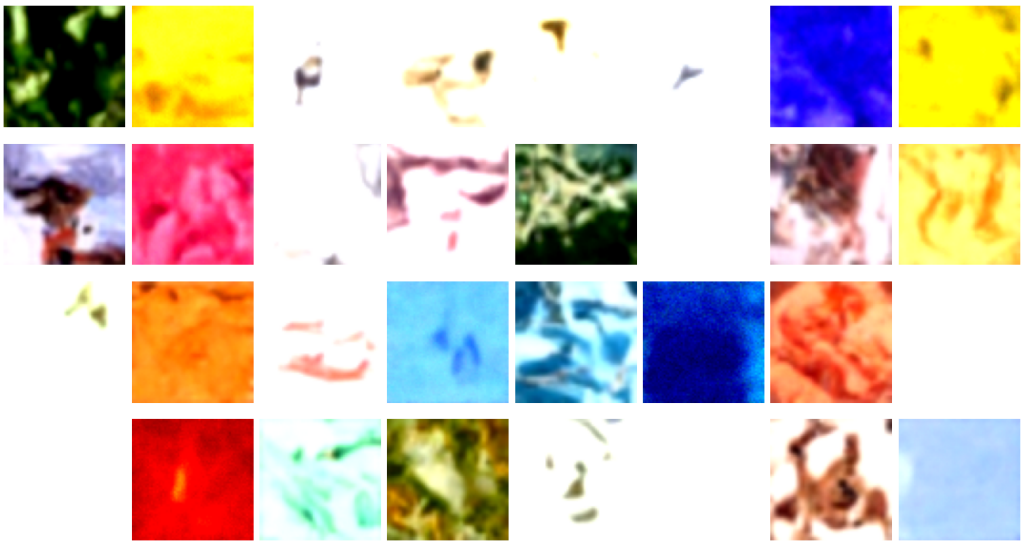


Figure 14. DDPM-generated image synthesis output from random noise inputs.

5. Analysis

5.1. Scores

The following analyses reference the results captured in Table 3.

Table 3. Quantitative results of generative modeling on Inception score, FID, and KID metrics.

	QBM	RBM	VAE	GAN	DDPM
Inception	1.77	3.84	7.87	2.72	3.319
FID	210.83	379.65	93.48	122.49	307.51
KID	0.068	0.191	0.024	0.033	0.586

5.1.1. Inception Score

On Inception scores, the QBM performed significantly worse than the classical models. This means that the diversity and fidelity of the QBM-generated images were significantly worse than those produced via existing classical methods. The VAE produced an exceptionally high Inception score, suggesting the images were both distinctly single-class labelled and the results produced an equal variety of results across classes. Qualitative observation of the produced samples is consistent with this score, as the produced images are of high fidelity and varied classes. Note that Figure 12b has distinct images of vehicles, animals, planes, etc.

Interestingly the DDPM produced a middling Inception score despite producing images that were of exceptionally low fidelity. This is because the Inception score measures the K-L Divergence between the single sample classification probability distribution and the summed distribution. While the image fidelity may be low, the overall summed distribution is fairly uniform due to the high variance of results, resulting in a higher K-L Divergence than naively expected.

5.1.2. Fréchet Inception Distance

The QBM produced the median FID score on the generated images, performing better than the RBM and DDPM but worse than the GAN and VAE. Recall the primary difference between the FID score and other metrics is the model's ability to extract and replicate salient features of the training data. The VAE and GAN do this exceptionally well, producing images that have distinct features that are easily observable. Notice Figures 12b and 13b both contain images that have easily identifiable features, namely the animals and vehicles in each set of generated images. Despite these images mimicking the input image very

closely, especially for Figure 12, the FID score only captures the distance between the features present in produced vs. training images, not the diversity of the images themselves.

Alternatively, the images produced by the DDPM and RBM have a distinct lack of identifiable features. To the human eye, Figure 11b does reflect the general lines and edges of the input found in Figure 11a, yet the Inception classifier fails to capture these features in its embedding, likely due to the high levels of surrounding noise with bright values. Note that brightly colored pixels are caused by large RGB (red-green-blue) values, which will have a larger effect upon the convolutional filters, which rely on matrix multiplication. This can have an undue negative effect on feature extraction and thus lead to lower FID scores. DDPMs face issues relating to a general lack of distinctive features produced. As discussed in Section 4.4, the computational limitations did not allow for adequate training and can thus account for the lack of effective feature generation.

As discussed in Section 4.5, the stitching and restitching of images cause features from multiple classes to be present in a single image, despite each feature being of moderately high fidelity. This restitching has negative consequences on the FID score and, given more qubits, could be improved upon but clamping entire images to the QPU directly.

5.1.3. Kernel Inception Distance

As with the FID score, the QBM produced the median score on the generated images yet skewed lower and thus achieved better results than the DDPM and GAN. The DDPM once again suffers from a lack of computing power and thus performs significantly worse than other models. The VAE and RBM performed exceptionally well, indicating the models' superior ability to generate samples that are distributed similarly to the training set.

KID is the metric on which the QBM performed comparatively best. This means that while the QBM lacks the ability to represent features in its generated images well and struggles to produce diverse, high-fidelity images, it can capture the underlying distribution of training images with its generated images moderately well. This result is significant because the fidelity of generated images should improve with increases in the number of qubits and better error correction, but a promising KID score is indicative that the QBM is adequately capturing the essence of image generation. Qualitatively, from Figure 15, it is clear the QBM can capture some meaningful image features from the training set but struggles with fidelity, i.e., distinct objects, clear boundaries, textured backgrounds, etc.

Figure 15. QBM-generated image synthesis output from random noise inputs.

5.2. Feature Extraction

Since QBMs and RBMs both lack convolutional layers, which are especially effective at capturing image features via convolution and image filters, it is expected that they would in turn score poorly for FID scores. This limitation of RBMs and QBMs can be solved by

transfer learning. Transfer learning allows a pre-trained model to be detached between two layers and then reconnected to an untrained model. That way, the embeddings, i.e., learned weights of the pre-trained model, can improve the performance of the untrained model [34]. Transfer learning with the convolutional layers from a CNN can be detached and reattached to the visible nodes of the RBM and QBM. However, for this strategy to work as intended with the QBM, a binarization layer, discussed in Section 3.4, would need to interface between the output of the CNN layers and the visible nodes.

5.3. Trilemma of Generative Learning

Recall the trilemma consists of the following: "(i) high-quality sampling, (ii) mode coverage and sample diversity, and (iii) fast and computationally inexpensive sampling" [6].

5.3.1. High-Quality Sampling

High-quality sampling is captured by FID and Inception scores. The QBM performed terribly on the Inception score and only moderately well on FID scores. Thus, it would be inaccurate to say the quantum annealer is uniquely producing high-quality samples. We hypothesize the main contributor to this result is the lack of convolutional layers and the image stitching required for training. This will be further discussed in 6.

5.3.2. Mode Coverage and Diversity

Mode coverage and diversity are captured by Inception and KID scores. While the QBM performed poorly on the Inception score, the KID score was promising. From qualitative observations of the generated images, it seems the QBM is managing to produce a diversity of images representative of the training data. The Inception score is certainly poorer than expected due to image stitching causing the Inception classifier to fail at classifying the images into one class.

5.3.3. Fast Sampling

The QBM thoroughly and unequivocally fails at fast sampling. The quantum annealer is extremely slow at sampling. This is partially due to hardware constraints, partially due to the high demand for quantum resources, and partially due to computational expensiveness. Regardless, the process of quantum sampling from an annealer is prohibitively slow and expensive. We hope to see this improve over time.

5.3.4. Conclusions

The QBM currently fails to improve on the Trilemma of Generative Learning (Section 1.2) in any of the three axes in any meaningful way. Despite this lack of improvement, it is important to note that quantum annealers are still in their infancy and have a limited number of qubits, require significant error correcting, are a shared resource, and are not the same as (or have the universality of) a general quantum computer. With hardware improvements, we expect to see further improvements and can revisit the trilemma once significant progress has been made.

6. Conclusions and Future Work

In conclusion, our team attempted to determine the efficacy of the D-Wave 2000Q quantum annealer on image synthesis, evaluated by industry-standard metrics compared to classical model counterparts, and determined if QBMs can crack the Trilemma of Generative Learning (Section 1.2). The quantum annealer, operating under a Quantum Boltzmann Machine (QBM) architecture, was assessed based on several performance metrics, including the Inception score, Fréchet Inception Distance (FID), and Kernel Inception Distance (KID). Its performance was compared against a suite of classical models comprising:

- Restricted Boltzmann Machine
- Variational Autoencoder
- Generative Adversarial Network

- Denoising Diffusion Probabilistic Model

The quantitative results of these experiments can be found in Table 3. The results showed that the QBM struggled to generate images with a high Inception score but managed to show promise in FID and KID scores, indicating an ability to generate images with salient features and a similar distribution to that of the training set.

The QBM implemented on the D-Wave 2000Q quantum annealer is not significantly better than the state-of-the-art classical models in the field. While the QBM outperformed a few classical models on FID and KID scores, it is important to note the difficulty of comparing models with different architectures trained on different hyper-parameters. The QBM did show great promise in its ability to represent the underlying distribution of the training data in its generated samples, and we hope to see this improve with more hardware improvements.

6.1. Image Preprocessing

A significant challenge in developing the QBM was the lack of qubits. This limitation forced us to split each image into a set of four squares, as described in Section 3.4, leading to the issue of stitching generated images in post. This issue can be somewhat resolved in the future in a few different ways.

Firstly, one could wait until hardware improvements are made to the quantum annealer in the form of an increase in the number of qubits and error-correcting abilities. With these improvements, one should see an increase in image synthesis quality. As more pixels can be embedded directly onto the QPU, the need for stitching will diminish and the QBM will be able to encode a richer embedding with features from the entire image in the correct locations.

Secondly, a CNN could be introduced and pre-trained via transfer learning. This would limit the input vector size required for the visible nodes on the QBM, thus allowing the CNN to pick up the bulk of the feature extraction. While this would not be a purely "quantum" solution, it would allow for the quantum annealer to specialize in embedding and sampling from a distribution of features as opposed to pixel values. This ought to improve performance, as CNNs are the gold standard in image processing for machine learning applications.

6.2. Quantum Computing

As quantum annealers improve, our team expects the ability to sample more often and in greater numbers will improve. With a greater number of samples, the QBM can evaluate a richer energy landscape and capture a more sophisticated objective function topology. With faster sampling, additional hyper-tuning could also be performed in a more timely manner, allowing for greater convergence upon a more ideal architecture.

Author Contributions:

Funding:

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Acknowledgments: We would like to thank D-Wave for providing access to their quantum computing resources as well as their continued support for Quantum Machine Learning research. This research would not be possible without the deep collaboration with Nurosene Health and their Chief Scientific Officer Dr. Joseph Geraci, Assistant Professor at Queen's Molecular Medicine. Lastly, a special thank you to Prof. Harry Ruda, Stanley Meek Chair in Advanced Nanotechnology, and Professor in the Department of Material Science and Engineering at the University of Toronto, for supervising this research.

Conflicts of Interest:

References

1. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 6840–6851.
2. Dhariwal, P.; Nichol, A. Diffusion Models Beat GANs on Image Synthesis. *arXiv* **2021**, arXiv:2105.05233.
3. Jain, S.; Ziauddin, J.; Leonchyk, P.; Yenkanchi, S.; Geraci, J. Quantum and classical machine learning for the classification of non-small-cell lung cancer patients. *SN Appl. Sci.* **2020**, *2*, 1088. https://doi.org/10.1007/s42452-020-2847-4.
4. Thulasidasan, S. *Generative Modeling for Machine Learning on the D-Wave*; Technical Report; Los Alamos National Lab. (LANL): Los Alamos, NM, USA, 2016. https://doi.org/10.2172/1332219.
5. Amin, M.H.; Andriyash, E.; Rolfe, J.; Kulchytskyy, B.; Melko, R. Quantum Boltzmann Machine. *Phys. Rev. X* **2018**, *8*, 021050. https://doi.org/10.1103/PhysRevX.8.021050.
6. Xiao, Z.; Kreis, K.; Vahdat, A. Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. *arXiv* **2021**, arXiv:2112.07804.
7. Smolensky, P. Information processing in dynamical systems: Foundations of harmony theory. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*; MIT Press: Cambridge, MA, USA, 1986; Volume 1.
8. Freund, Y.; Haussler, D. Unsupervised learning of distributions on binary vectors using two layer networks. In *Advances in Neural Information Processing Systems*; Moody, J., Hanson, S., Lippmann, R., Eds.; Morgan-Kaufmann: Burlington, MA, USA, 1991; Volume 4.
9. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. https://doi.org/10.1073/pnas.79.8.2554.
10. Hinton, G.E. A Practical Guide to Training Restricted Boltzmann Machines. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 599–619. https://doi.org/10.1007/978-3-642-35289-8_32.
11. Carreira-Perpiñán, M.Á.; Hinton, G.E. On Contrastive Divergence Learning. In Proceedings of the AISTATS, Bridgetown, Barbados, 6–8 January 2005.
12. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2014**, arXiv:1312.6114.
13. Rocca, J. Understanding Variational Autoencoders (VAES). 2021. Available online: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73 (accessed on).
14. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661.
15. A Beginner's Guide to Generative Adversarial Networks (Gans). Available online: https://wiki.pathmind.com/generative-adversarial-network-gan (accessed on).
16. Arjovsky, M.; Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv* **2017**, arXiv:1701.04862.
17. Hinton, G.E. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Comput.* **2002**, *14*, 1771–1800. https://doi.org/10.1162/089976602760128018.
18. What is Quantum Annealing? D-Wave System Documentation. Available online: https://docs.dwavesys.com/docs/latest/c_gs_2.html (accessed on).
19. Lu, B.; Liu, L.; Song, J.Y.; Wen, K.; Wang, C. Recent progress on coherent computation based on quantum squeezing. *AAPPS Bull.* **2023**, *33*, 7. https://doi.org/10.1007/s43673-023-00077-4.
20. Wittek, P.; Gogolin, C. Quantum Enhanced Inference in Markov Logic Networks. *Sci. Rep.* **2017**, *7*, 45672. https://doi.org/10.1038/srep45672.
21. Li, W.; Deng, D.L. Recent advances for quantum classifiers. *Sci. China Phys. Mech. Astron.* **2022**, *65*, 220301. https://doi.org/10.1007/s11433-021-1793-6.
22. Wei, S.; Chen, Y.; Zhou, Z.; Long, G. A quantum convolutional neural network on NISQ devices. *AAPPS Bull.* **2022**, *32*, 2. https://doi.org/10.1007/s43673-021-00030-3.
23. Sleeman, J.; Dorband, J.E.; Halem, M. A hybrid quantum enabled RBM advantage: Convolutional autoencoders for quantum image compression and generative learning. *arXiv* **2020**, arXiv:2001.11946.
24. Krizhevsky, A.; Nair, V.; Hinton, G. CIFAR-10 (Canadian Institute for Advanced Research).
25. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. 2009. Available online: https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf (accessed on).
26. Peter Eckersley, Y.N.e.a. EFF AI Progress Measurement Project. 2017.
27. Mack, D. A Simple Explanation of the Inception Score. 2019. Available online: https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a (accessed on).
28. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.
29. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved Techniques for Training GANs. *arXiv* **2016**, arXiv:1606.03498.
30. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.

31. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *arXiv* **2017**, arXiv:1706.08500. https://doi.org/10.48550/ARXIV.1706.08500.
32. Bińkowski, M.; Sutherland, D.J.; Arbel, M.; Gretton, A. Demystifying MMD GANs. *arXiv* **2018**, arXiv:1801.01401. https://doi.org/10.48550/ARXIV.1801.01401.
33. Cloud Tensor Processing Units (TPUS) | Google Cloud. Available online: https://cloud.google.com/tpu/docs/tpus (accessed on).
34. Dhillon, P.S.; Foster, D.; Ungar, L. Transfer Learning Using Feature Selection. *arXiv* **2009**, arXiv:0905.4022. https://doi.org/10.48550/ARXIV.0905.4022.