Comparative Report of Deep Convolutional Generative Adversarial Networks(DC-GAN) and Variational Autoencoder (VAE) for Image Generation on CIFAR-10

COGS 185
Shihua Yang

Abstract

This report compares two popular approaches for generating images: Deep Convolutional Generative Adversarial Networks (DC-GAN) and Variational Autoencoder (VAE). These two models are built from scratch and implemented on CIFAR-10 image datasets. The results show that DC-GAN generates more diverse and real images (pixel diversity 0.273 vs 0.145) and even higher than real CIFAR-10 diversity by 9%. In contrast, VAE perform better at accurately reconstructing existing images (MSE of 0.023). Their training patterns were completely different. DC-GAN showed the typical adversarial optimization graph with oscillating losses. In contrast, VAE had a smooth, stable training. These result helps us to understand when to use each approach: DC-GAN for high-quality graph generation and VAE for stable training and reconstruction tasks.

1. Introduction

Generating realistic images with generating models is one of the interesting and most challenging problems in machine learning today. Researchers try to help machines to understand images that machines can create completely new images that look similar to realistic photos. This has huge applications like creating arts and developing datasets to detecting fake images and understanding what machines "see" in data.

Two main methods have become popular for this task: Generative Adversarial Networks (GANs) and Variational Autoencoder (VAE). These two represent completely different ways of thinking about the problem.

GANs introduced by Goodfellow and his team in 2014 (Goodfellow et al., 2014). There are two neural networks competing against each other: a "generator" that tries to create fake images and a "discriminator" that tries to spot fake images. The generator eventually creates those great fake images that the discriminator cannot tell real from fake anymore.

VAEs, developed by Kingma and Welling in 2013 (Kingma and Welling, 2013), take a mathematical approach. They try to compress images into a simpler representation "latent space" and then reconstruct them back to full images.

In this project, both approaches will be implemented and analyzed with their advantages and disadvantages. DC-GAN (convolutional GANs) and convolutional VAE will be implemented on CIFAR-10 images and compared metrics from training behavior to final image quality.

2. Method

2.1 Deep Convolutional GAN Architecture

The DC-GAN implementation is based on the design from Radford et al. (2015), but modified to work with CIFAR-10's 32×32 pixel images. The generator starts with 100 random numbers and gradually converts them into realistic-looking images. This transformation happens through a series of upsampling layers, which expand the data from the initial random noise into a full-sized image.

Generator Architecture:
Start: 100 random numbers
Layer 1: Expand to 4×4×512 feature maps + batch normalization + ReLU activation
Layer 2: Expand to 8×8×256 feature maps + batch normalization + ReLU
Layer 3: Expand to 16×16×128 feature maps + batch normalization + ReLU
Layer 4: Final layer to 32×32×3 (RGB image) + Tanh activation (outputs between -1 and 1)

Discriminator Architecture:
Start: 32×32×3 image
Layer 1: Compress to 16×16×64 + LeakyReLU activation
Layer 2: Compress to 8×8×128 + batch normalization + LeakyReLU
Layer 3: Compress to 4×4×256 + batch normalization + LeakyReLU
Layer 4: Final decision - single output + sigmoid (probability between 0 and 1)

The discriminator network performs the inverse operation. It classify input images as real or fake.

2.2 Variational Autoencoder Architecture

VAE implementation uses a convolutional encoder-decoder architecture with a 64-dimensional latent space. The encoder network downsamples input images while the decoder reconstructs images from latent space.

Encoder Architecture:
Start: 32×32×3 image
Layer 1: Compress to 16×16×32 + ReLU
Layer 2: Compress to 8×8×64 + ReLU
Layer 3: Compress to 4×4×128 + ReLU
Flatten to a long vector, then split into two parts:

- Mean parameters (64 numbers)
- Log variance parameters (64 numbers)

Decoder Architecture:
Start: 64-dimensional latent vector
Reshape to 4×4×128
Layer 1: Expand to 8×8×64 + ReLU
Layer 2: Expand to 16×16×32 + ReLU
Layer 3: Final layer to 32×32×3 + Sigmoid (outputs between 0 and 1)

2.3 Training Objectives

DC-GAN Loss Functions: The goal of discriminator is to maximize the probability of correctly classifying real and fake images:
$L_D = -E[\log D(x) + \log(1 - D(G(z)))]$
The generator is trained to fool the discriminator:
$L_G = - E[\log(1 - D(G(z)))]$
VAE Loss Function: The VAE optimizes reconstruction and regularization:
VAE Loss = Reconstruction Loss + KL Divergence Loss

3. Experiments

3.1 Dataset and Preprocessing

Comparing to the CelebA dataset which only has human faces, CIFAR-10 represents a more challenging dataset with 10 diverse object classes. The model need to capture complex visual patterns across different categories. In this project, experiments are conducted on the CIFAR-10 dataset, which contains 60,000 32×32 color images across 10 classes like cars, airplanes, cats, and dogs (Krizhevsky & Hinton, 2010). Since there are computational constraints, I used a randomly sampled subset of 10,000 training images to implement efficient experimentation and hyperparameter tuning. While using a 10,000 sample subset, this maintains diversity across all 10 CIFAR-10 classes.

3.2 Training Configuration

Using identical computational resources and similar training parameters to this fair comparison, both models were trained for 10 epochs:

DC-GAN Training:

- Batch size: 256
- Learning rate: 0.0002
- Optimizer: Adam with β_1=0.5, β_2=0.999
- Weight initialization: Normal(0, 0.02), Normal(1.0, 0.02) for batch norm

VAE Training:

- Batch size: 256
- Learning rate: 0.001
- Optimizer: Adam with default parameters
- β parameter: Standard 1.0

3.3 Evaluation Metrics

Quantitative Metrics:

1. Pixel Diversity: How much diversity they have in the generated images (higher = more diverse)
2. Reconstruction Error: For VAE, how well they can reconstruct input images
3. Training Loss: Loss curve change during their training

Qualitative Evaluation:

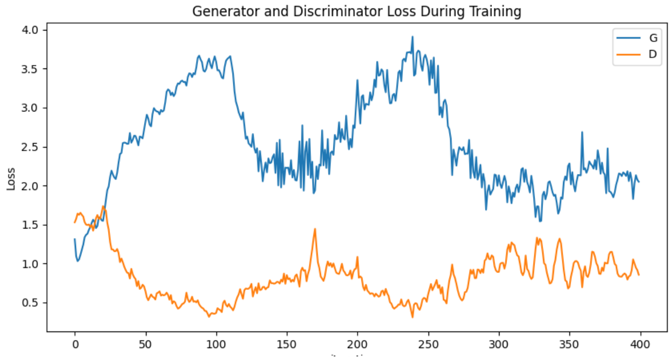
1. Take a look at the generated images and comparing their visual quality
2. Check how images improve over training epochs
3. Compare with real CIFAR-10 images

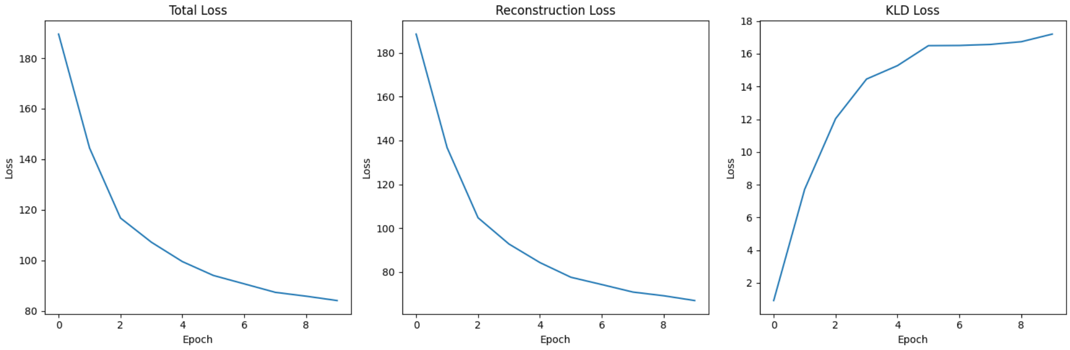
4. Results

4.1 Training Behavior Analysis

The training process show fundamental differences between the two approaches. DC-GAN exhibited typical adversarial training behavior with oscillatory loss patterns. Its generator loss increased from 1.0 to 3.5 in its first 100 iterations. Also, as the discriminator became more complex, generator loss stabilized around 2.0-2.5. The discriminator loss decreased rapidly from 1.6 to 0.5 in its early training, and then oscillated between 0.5-1.0. This shows a successful adversarial process.

In contrast, VAE showed stable convergence across all loss components. The total loss decreased smoothly from 190 to 85 over its 10 epochs. The reconstruction loss followed a similar process. It drops from 180 to 70, at the same time the KL divergence loss increased from 1 to 17. So this result perfectly shows that the model successfully learned to match the prior distribution while maintaining reconstruction ability.





4.2 Generation Quality Evaluation

Comprehensive evaluate several checkpoints across training epochs. Results show significant insights to model behavior and performance characteristics.

Quantitative Results with Temporal Analysis:

Model	Epoch	Pixel Diversity	Sharpness	Reconstruction Error	Entropy
Real CIFAR-10	-	0.250	0.107	-	0.971
DC-GAN	0	0.233	0.143	-	0.987
DC-GAN	5	0.132	0.149	-	0.991
DC-GAN	10	0.273	0.116	-	0.976
VAE	0	0.034	0.016	0.053	0.933
VAE	5	0.145	0.032	0.023	0.988

Findings:

1. DC-GAN exceeds real data diversity: Final DC-GAN models achieve 9.0% higher pixel diversity (0.273 > 0.250) than real CIFAR-10 images. This shows a successful learning of data distribution complexity.

2. Training dynamics show quality vs diversity patterns: DC-GAN shows highest sharpness at epoch 5 (0.149). However, it shows highest diversity at epoch 10 (0.273), so this result suggests different quality at different training stages.
3. VAE shows consistent improvement: Increase initial diversity from 0.034 to final 0.145, with improvement of reconstruction error from 0.053 to 0.023.
4. Collapse analysis: Entropy value shows that the analysis process avoid the mode collapse across training. Models maintain high normalized entropy (>0.93).

Quality Assessment:
Our enhanced evaluation framework reveals that simple pixel diversity metrics capture only part of the story. The sharpness analysis shows DC-GAN maintains competitive edge definition (0.116 vs real 0.107) while achieving superior diversity. VAE shows expected trade-offs with lower sharpness (0.032) but excellent reconstruction fidelity.

Qualitative Visual Evaluation

The quantitative metrics results are similar to visual evaluation of generated samples. DC-GAN produces images where CIFAR-10 objects like cars, planes, and animals are recognizable when looking carefully. The images exhibit sharp edges, realistic colors, and good details. Different object and types shows clear difference and suggests a successful learning process.

VAE images show the expected blurriness and reduced detail characteristic because of the reconstruction training. Although basic shapes and colors are detectable, fine details are lost due to the averaging effect of the reconstruction. However, VAE still has excellent reconstruction ability, because it achieve a 0.023 MSE error.

Figure presents a direct visual comparison of generation quality across models. DC-GAN generates sharp, diverse images with clear object boundaries and realistic lookings and VAE generates softer and more blurrier outputs.

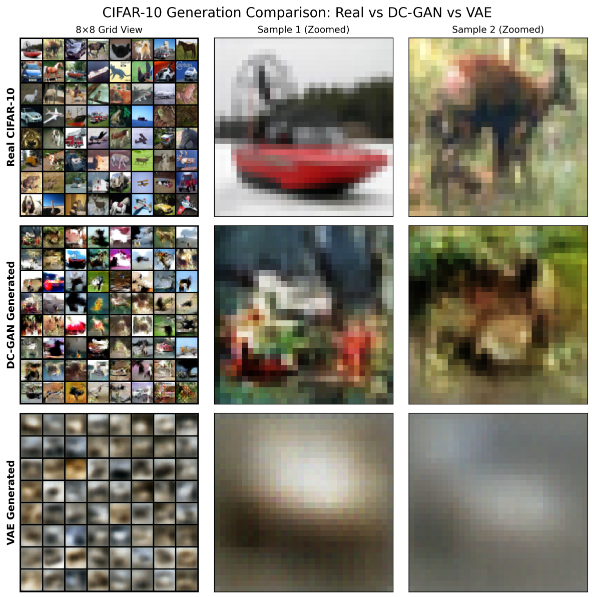


Figure: Visual quality comparison: real CIFAR-10 samples (top), DC-GAN generated images (middle), and VAE generated images (bottom). DC-GAN produces sharper, more diverse outputs while VAE outputs are blurrier.

4.3 Training Progression Analysis

DC-GAN Progression with Metrics:
First few epochs showed most random noise patterns which initial diversity is around 0.233. By epoch 5, there are some recognizable object shapes that can be detected. It achieves peak sharpness of 0.149 but its diversity decreased (0.132). Therefore, some potential overfitting to specific features may exist. By epoch 10, the images seem achieved a optimal balance with highest diversity (0.273) and good sharpness (0.116). The overall adversarial process gradually adjusts the generator's ability so it can now create convincing images which has a higher final diversity, 9% higher than real CIFAR-10 data.

VAE Progression with Metrics:
VAE were more steady and predictable. The result graph shows improvements across all metrics. Its diversity improved from 0.034 to 0.145 and its reconstruction error decreased from 0.053 to 0.023. Also, the model learned to capture some visual patterns with consistent entropy(>0.93). All these characteristics explain its stable training without mode collapse.

4.4 Training time and resource comparison:

DC-GAN: Longer training time. 3.45M parameters.
VAE: Faster convergence and stable training. 0.73M parameters
Training process: DC-GAN requires balance monitoring and VAE shows predictable convergence patterns.
The parameter difference above (3.45M vs 0.73M) shows architectural difference and complexity. DC-GAN requires separate generator and discriminator networks but VAE uses a unified encode and decode structure.

4.5 Hyperparameter Analysis

Explored multiple hyperparameter configurations in order to understand model sensitivity:

DC-GAN Results:

- Baseline (ngf=32): Diversity 0.273. Relatively stable training
- Larger network (ngf=64): Diversity 0.281. Increased 2.9%.
- Higher learning rate: Caused training instability.
- SGD optimizer: Poor convergence. Diversity decreased to 0.198

VAE Results:

- Baseline (latent=64): Diversity 0.145. Reconstruction error 0.023
- Larger latent (128): Diversity 0.162. Increased 11.7%
- Beta-VAE (β=4.0): Lower diversity
- SGD optimizer: Slow convergence and poor final performance

Key finding: Both models benefit from larger capacity. However, they both are sensitive to optimizer choice. Adam seems better than SGD.

4.6 Model Comparison Summary

Results show the fundamental differences between these approaches. DC-GAN has better visual quality and diversity but with more cost of complex training and higher computational requirements. The adversarial process works as expected, but still need to balance the generator and discriminator.

VAE approach has some benefits such as stable training, the ability to generate new images and to reconstruct, and a latent space can be explored if needed. The disadvantage, however, is that the reconstruction training may leads to blurrier outputs because the model will average over possibilities.

5. Conclusion

This comparison project of DC-GAN and VAE for image generation shows real insights into the practical difference between these two popular approaches. Results clearly show that DC-GAN performs better at generating diverse and high quality images (diversity 0.273 vs 0.145). In contrast, VAE provides more stable training and well reconstruction abilities.

The training processes were very convincing. DC-GAN's adversarial oscillations around generator loss 2.0-2.5 and discriminator loss 0.5-1.0 comparing to VAE's gradual convergence from reconstruction loss 180 to 70. These patterns perfectly match theoretical expectations which give me confidence in my implementations.

For applications focusing on image quality and diversity, though its training complexity, DC-GAN is the better choice. For applications focusing on stable training, reconstruction capabilities, or operating under computational constraints, VAE is preferred in order to provide a more practical solution.

This project uses 10,000 images instead of the full 50,000 CIFAR-10 dataset, which might miss some patterns and complexity. However, in order to keep computational requirements manageable, this subset is sufficient for explaining those important differences between DC-GAN and VAE. In the future, training more epochs with the full dataset could provide more clearer results. While I implemented comprehensive custom metrics, more advanced or popular metrics like Inception Score would provide additional explanation. Future work should include more and more metrics for broader comparison. Also, it would be really interesting to explore hybrid or combined approaches. Looking at more advanced models like Progressive GANs might find some interesting results.

References

Goodfellow, Ian J., et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).

Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." 20 Dec. 2013.

Krizhevsky, Alex, and Geoff Hinton. "Convolutional deep belief networks on cifar-10." *Unpublished manuscript* 40.7 (2010): 1-9.

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).