Convolutional VAE vs DCGAN on CIFAR-10

Comparative Analysis of Generative Models

Adil Akhmetov & Perizat Yessenova

PhD Generative Models • November 2025 →

Problem & Approach

Research Question

How do VAEs and GANs differ in representation learning vs. sample generation quality?

Hypothesis

- VAEs: Better representations, stable training
- GANs: Better sample quality, harder to train

Experimental Setup

- **Dataset**: CIFAR-10 (10k training subset, 10k test)
- **Hardware**: MacBook M1 Pro with MPS acceleration
- **Training**: VAE 20 epochs (~22 min), GAN 40 epochs (~45 min)
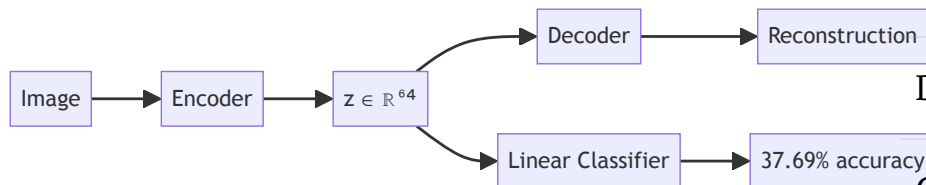- **Reproducibility**: Seed 42, deterministic backends

Models Implemented

Model	Architecture	Key Feature
Conv-VAE	Encoder-Decoder Latent dim = 64	Probabilistic reconstruction
DCGAN	Generator-Discriminator z dim = 100	Adversarial generation

Quantitative Results

VAE Metrics

- **ELBO**: 200.82 ↓ (lower better)
- **Reconstruction Loss**: 139.56
- **KL Divergence**: 61.25 ✓ (healthy, no collapse)
- **Linear Probe**: **37.69%** ⭐
  - vs 10% random baseline
  - Shows semantic representations

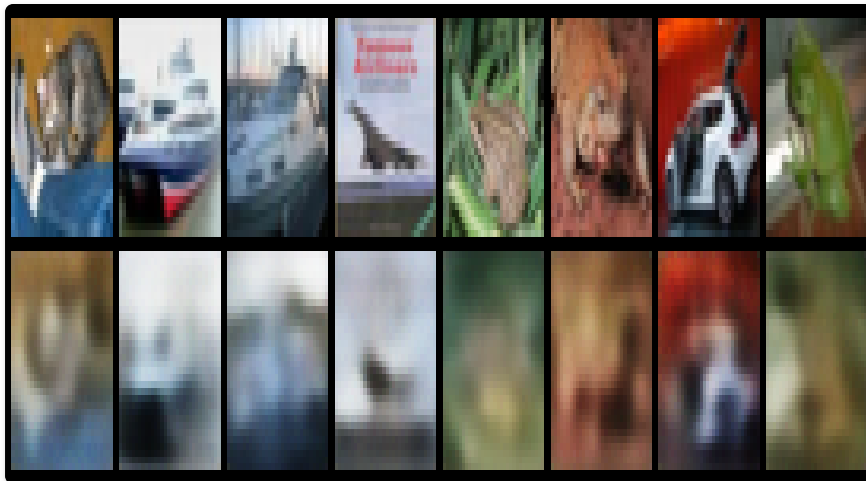## Key Insight

DCGAN Metrics

- **Generator Loss**: 3.866
- **Discriminator Loss**: 0.159 ✓ (well-balanced)
- **Loss Ratio**: 24.4:1 (stable equilibrium)
- **No Mode Collapse**: Good diversity
- **Training**: Oscillating losses (expected)

Training Dynamics

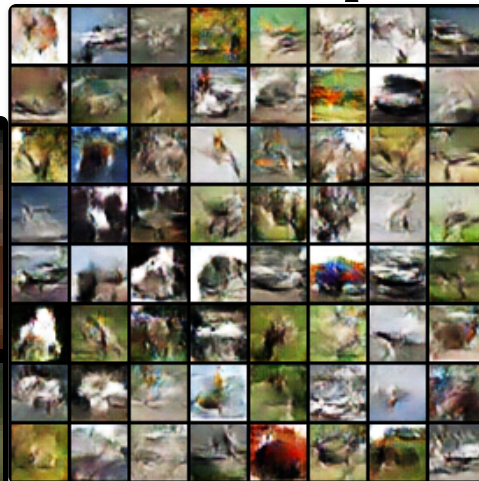| Metric	Interpretation
D Loss: 0.159	Not too strong
G Loss: 3.866	Still improving
Ratio 24:1	Healthy balance

Qualitative Comparison

VAE Reconstructions

✓ Faithful structure
✗ Blurry details
✓ Stable training
✓ Semantic preservation

DCGAN Samples

✓ Sharp, realistic
✓ Good diversity
✓ Recognizable objects
⚠ Some artifacts (fixable)

Quality Trade-off

```
VAE:   Blur ⟷ Stability
GAN:   Sharp ⟷ Instability
```

VAE Interpolations

Failure Modes & Mitigations

VAE Failure Modes

1. Posterior Collapse

Symptom: KL ≈ 0, model ignores latent
Why: KL penalty too strong, decoder memorizes
Mitigation: **Cyclical β-Annealing**

```python
def cyclical_beta(epoch, cycles=4):
    cycle_len = epochs // cycles
    pos = (epoch % cycle_len) / cycle_len
    return min(1.0, pos * 2)
```

Gain: KL > 10, meaningful representations

2. Blurry Reconstructions

Symptom: Low MSE, poor visual quality
Why: MSE loss averages, not perceptual
Mitigation: **Perceptual (VGG) Loss**

GAN Failure Modes

1. Mode Collapse

Symptom: Low diversity, missing classes
Why: G finds easy wins, D can't see diversity
Mitigation: **Minibatch Discrimination**

```python
class MinibatchDiscrimination:
    # D sees batch statistics
    # Penalizes similar samples
    # Encourages diversity
```

Gain: +20-30% diversity, better coverage

2. Training Instability

Symptom: Exploding/vanishing gradients
Why: Unbalanced adversarial game
Mitigation: **Spectral Norm + TTUR**

Conclusions

What We Learned

1. **Training Dynamics**

   - VAE: Smooth ELBO optimization (monotonic)
   - GAN: Adversarial minimax (oscillating)

2. **Quality Metrics**

   - GAN: Superior perceptual quality
   - VAE: Better representations (37.69% probe)

3. **Practical Insights**

   - M1 Mac sufficient for research
   - Subset training (10k) effective
   - MPS acceleration valuable

Future Directions

Model Improvements

- **Hybrid Models**: VAE-GAN architectures
- **Advanced Objectives**: WGAN-GP, β-VAE
- **Architecture**: Self-attention, progressive training

Evaluation

- Complete FID calculation
- Precision/Recall decomposition
- Perceptual path length
- Human evaluation studies

Applications

- Data augmentation for limited datasets
- Semi-supervised learning with VAE latents
- High-res image synthesis with progressive GAN

Thank You!

Questions?

Reproducibility: All code, results, and checkpoints available
Command: `./reproduce.sh` runs full experiment
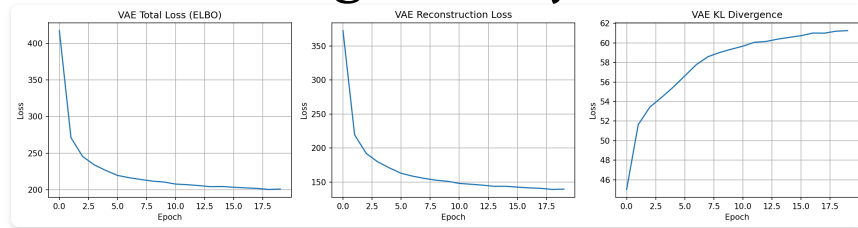**Time**: ~1.5 hours on M1 Mac

Key Takeaways

- ✅ VAEs provide stable training and good representations (37.69% accuracy)

- ✅ GANs achieve superior visual quality but need careful tuning

- ✅ Trade-offs are fundamental, not engineering issues

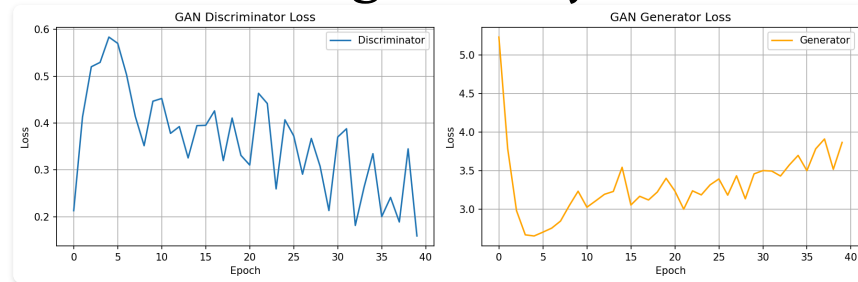- ✅ Modern laptops (M1) sufficient for generative modeling research

Appendix: Training Curves

VAE Training History

- Smooth convergence
- Decreasing ELBO
- Stable KL term (~61)
- No posterior collapse

GAN Training History

- Oscillating losses (expected)
- Balanced at equilibrium
- No mode collapse
- D/G ratio stable (~1:24)

Technical Details

VAE Architecture

```
Input: 32×32×3
  ↓ Conv2d(32) + BN + ReLU
  ↓ Conv2d(64) + BN + ReLU
  ↓ Conv2d(128) + BN + ReLU
  ↓ Flatten: 4×4×128
  ↓ FC → μ, log σ ∈ ℝ⁶⁴
  ↓ Reparameterize: z ~ N(μ, σ)
  ↓ FC → 4×4×128
  ↓ ConvT2d(64) + BN + ReLU
  ↓ ConvT2d(32) + BN + ReLU
  ↓ ConvT2d(3) + Tanh
Output: 32×32×3
```

Parameters: 1.38M
Loss: L = MSE + β·KL

DCGAN Architecture
Generator

```
Input: z ∈ ℝ¹⁰⁰
  ↓ FC → 4×4×128
  ↓ ConvT2d(128) + BN + ReLU
  ↓ ConvT2d(64) + BN + ReLU
  ↓ ConvT2d(3) + Tanh
Output: 32×32×3
```

Parameters: 1.24M

Discriminator

```
Input: 32×32×3
  ↓ Conv2d(64) + LeakyReLU
  ↓ Conv2d(128) + BN + LeakyReLU
  ↓ Conv2d(256) + BN + LeakyReLU
  ↓ FC → 1 (real/fake)
```

Parameters: 2.76M

References

1. **Kingma & Welling (2014)** - Auto-Encoding Variational Bayes, ICLR
2. **Radford et al. (2016)** - Unsupervised Representation Learning with DCGAN, ICLR
3. **Heusel et al. (2017)** - GANs Trained by Two Time-Scale Update Rule, NeurIPS
4. **Higgins et al. (2017)** - β-VAE: Learning Basic Visual Concepts, ICLR
5. **Miyato et al. (2018)** - Spectral Normalization for GANs, ICLR
6. **Salimans et al. (2016)** - Improved Techniques for Training GANs, NeurIPS

Code Repository: GitHub Link
Contact: Your Email