Generative Adversarial Networks

A Comprehensive Lecture

Amir

Computer Vision

Outline

What are Generative Models?

Definition

Models that learn the underlying probability distribution $p_{data}(x)$ from samples and can generate new samples from this distribution.

Discriminative Models:

- Learn $p(y|x)$
- Classify/regress
- Decision boundaries

Generative Models:

- Learn $p(x)$ or $p(x|y)$
- Generate new data
- Understand data structure

Applications

- Image synthesis (art, design)
- Data augmentation
- Anomaly detection
- Style transfer

Types of Generative Models

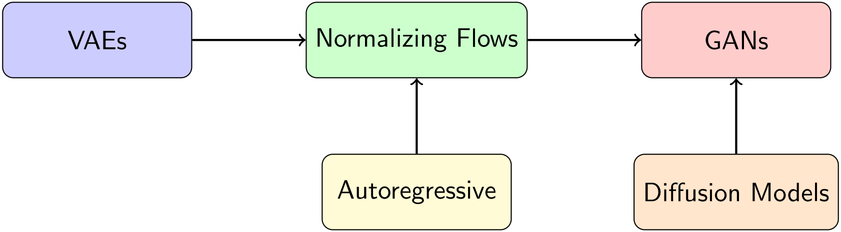


Figure: Evolution of generative models

GAN Characteristics

- **Adversarial training**: Two networks compete
- **No explicit likelihood**: Implicit distribution learning
- **State-of-the-art**: Best image quality (until diffusion models)

The GAN Framework

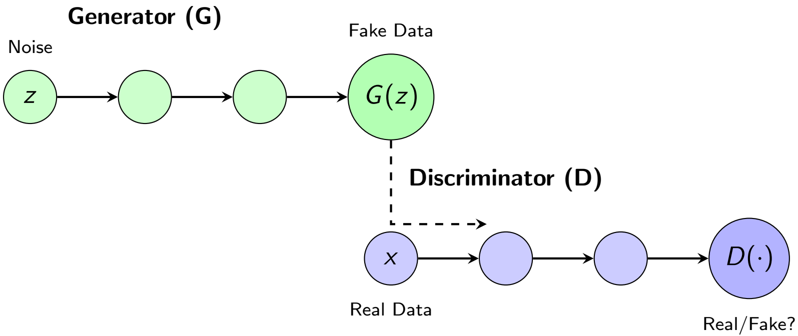
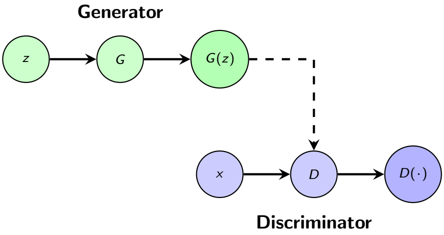


Figure: Adversarial training: G tries to fool D, D tries to detect fakes

The GAN MiniMax Game

A GAN is a **two-player minimax game** played between G and D:

- **Generator (G)**: Tries to minimize the probability that D identifies its samples as fake.
- **Discriminator (D)**: Tries to maximize the probability of correctly labeling real vs. fake data.



Original GAN Objective [1]

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

Both models are trained simultaneously in a competition to reach a Nash Equilibrium.

Nash Equilibrium in GANs

- In the GAN minimax game, the training process seeks a **Nash Equilibrium** where neither player can improve their outcome by changing their strategy.
- This point is reached when the generator produces the true data distribution: $p_g = p_{data}$.

Optimal Discriminator and Equilibrium

At the equilibrium point, the discriminator is unable to distinguish between real and fake data, resulting in:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} = \frac{1}{2}$$

- **Global Minimum:** Goodfellow et al. [1] proved that the global minimum of the training criterion is achieved if and only if $p_g = p_{data}$.
- **Challenge:** In practice, finding the Nash Equilibrium is difficult because gradient descent is designed to find local minima of a function, not a stationary point in a game.

Mathematical Formulation

Original GAN Objective [1]

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

Discriminator:

- Maximizes $V(D, G)$
- Distinguishes real from fake
- Outputs probability (0 to 1)

Generator:

- Minimizes $V(D, G)$
- Maximizes $\log D(G(z))$
- Learns to fool D

Theorem (Optimal Discriminator)

For fixed G, the optimal discriminator is:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

Training Algorithm

Algorithm GAN Training Algorithm

Require: Learning rates α_D, α_G, number of iterations N
Require: k: number of D steps per G step

for iteration $= 1$ to N **do**
  for $t = 1$ to k **do**
    Sample minibatch $\{x^{(i)}\}_{i=1}^{m} \sim p_{data}$
    Sample minibatch $\{z^{(i)}\}_{i=1}^{m} \sim p_z$
    Update D: $\theta_D \leftarrow \theta_D + \alpha_D \nabla_{\theta_D} \frac{1}{m} \sum_{i=1}^{m} [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$
  end for
  Sample minibatch $\{z^{(i)}\}_{i=1}^{m} \sim p_z$
  Update G: $\theta_G \leftarrow \theta_G - \alpha_G \nabla_{\theta_G} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)})))$
end for

Practical Note

In practice, update G to maximize $\log D(G(z))$ instead of minimizing $\log(1 - D(G(z)))$ to avoid vanishing gradients early in training.

Optimal Discriminator Derivation

For fixed G, we want to maximize:

$$V(D) = \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

Rewrite as integral:

$$V(D) = \int_x p_{data}(x) \log D(x) + p_g(x) \log(1 - D(x)) dx$$

Take functional derivative w.r.t $D(x)$ and set to zero:

$$\frac{\partial V}{\partial D(x)} = \frac{p_{data}(x)}{D(x)} - \frac{p_g(x)}{1 - D(x)} = 0$$

Solve for $D(x)$:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

Global Optimality and Convergence

Substitute $D^*(x)$ into value function:

$$
\begin{aligned}
C(G) &= \max_D V(G, D) \\
&= \mathbb{E}_{x \sim p_{data}}[\log D^*(x)] + \mathbb{E}_{x \sim p_g}[\log(1 - D^*(x))] \\
&= \mathbb{E}_{x \sim p_{data}}\left[\log \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}\right] + \mathbb{E}_{x \sim p_g}\left[\log \frac{p_g(x)}{p_{data}(x) + p_g(x)}\right]
\end{aligned}
$$

This equals:

$$C(G) = -\log 4 + 2 \cdot JSD(p_{data} \| p_g)$$

where JSD is Jensen-Shannon divergence.

Theorem

The global minimum of $C(G)$ is achieved if and only if $p_g = p_{data}$.

Training Instability: Common Issues

Non-Convergence

- Oscillations in loss
- No equilibrium reached
- Cyclic behavior

Vanishing Gradients

- D becomes too good too fast
- $\log(1 - D(G(z))) \to 0$
- G gets no useful gradient

Mode Collapse

- G produces limited variety
- Captures few modes of p_{data}
- Diversity loss

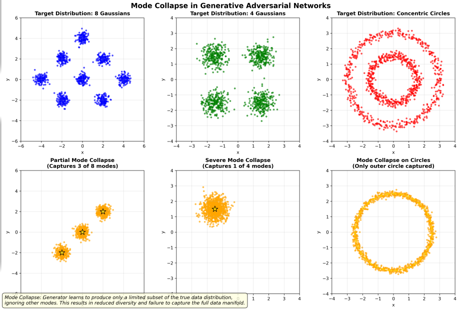


Figure: Mode collapse example

Wasserstein GAN (WGAN) [2]

Key Insight

Use Earth Mover (Wasserstein-1) distance instead of JS divergence:

$$W(p_{data}, p_g) = \inf_{\gamma \in \Pi(p_{data}, p_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]$$

Kantorovich-Rubinstein Duality

$$W(p_{data}, p_g) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x\sim p_{data}}[f(x)] - \mathbb{E}_{x\sim p_g}[f(x)]$$

WGAN Objective

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x\sim p_{data}}[D(x)] - \mathbb{E}_{z\sim p_z}[D(G(z))]$$

where $\mathcal{D}$ is set of 1-Lipschitz functions.

WGAN-GP: Gradient Penalty [3]

Problem with Weight Clipping

Enforcing Lipschitz via weight clipping leads to:

- Capacity underuse
- Exploding/vanishing gradients

Gradient Penalty Solution

Add penalty term:

$$L = \underbrace{\mathbb{E}_{\tilde{x}\sim p_g}[D(\tilde{x})] - \mathbb{E}_{x\sim p_{data}}[D(x)]}_{\text{Wasserstein loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x}\sim p_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}_{\text{Gradient penalty}}$$

Improved Stability

- Better gradient flow
- Higher quality samples
- More stable training

DCGAN: Deep Convolutional GAN [4]

Architecture Guidelines

1. Replace pooling with strided convolutions
2. Use BatchNorm in both G and D
3. Remove fully connected layers
4. Use ReLU (G) and LeakyReLU (D)

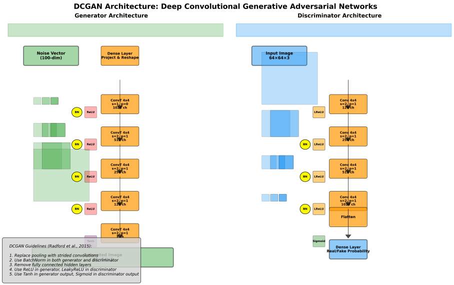


Figure: DCGAN generator architecture

Key Contributions

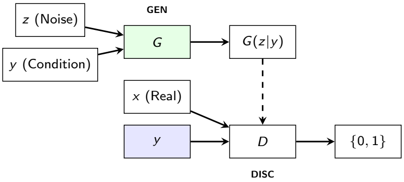
- First stable CNN-based GAN
- Learned meaningful latent representations
- Vector arithmetic in latent space

Conditional GANs [5]

Objective Function

$$\min_G \max_D V(D, G) =$$
$$\mathbb{E}_{x \sim p_{data}}[\log D(x|y)] +$$
$$\mathbb{E}_{z \sim p_z}[\log(1 - D(G(z|y)|y))]$$

Unlike standard GANs, both G and D receive auxiliary information y (e.g., class labels or images).



StyleGAN Series Evolution [6, 7, 8]

StyleGAN (2019)

- Style-based generator
- AdaIN (Adaptive Instance Norm)
- Stochastic variation via noise

StyleGAN2 (2020)

- Fixed artifacts (water droplets)
- Path length regularization
- No progressive growing

StyleGAN3 (2021)

- Aliasing-free architecture
- Equivariance to transformations
- Better interpolation

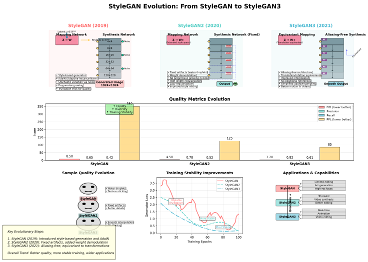


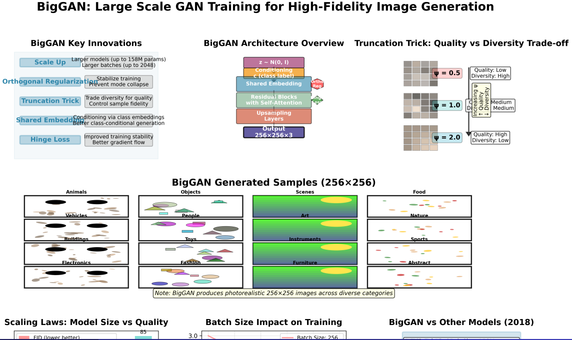
Figure: Improvements across StyleGAN versions

Image Generation: BigGAN [9]

Key Innovations

- **Scale**: Large batch sizes (2048) and models
- **Orthogonal Regularization**: Stabilize training
- **Truncation Trick**: Trade diversity for quality

$$R_{\beta}(W) = \beta \| W^{\top} W - I \|_F^2$$



Text-to-Image Synthesis: Proprietary Models

DALL-E 2 [10]

- Diffusion prior + decoder
- CLIP embeddings
- High semantic alignment

Imagen [11]

- T5-XXL text encoder
- Cascade diffusion
- State-of-the-art FID

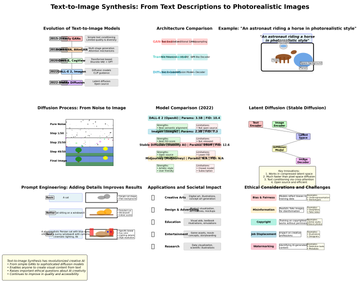


Figure: "An astronaut riding a horse"

Text-to-Image Synthesis: Stable Diffusion

Stable Diffusion [12]

- **Latent Diffusion Model (LDM)**: Operates in a compressed z-space.
- **Efficiency**: Drastically reduces VRAM usage vs. pixel-space models.
- **Conditioning**: Flexible integration of text via Cross-Attention.

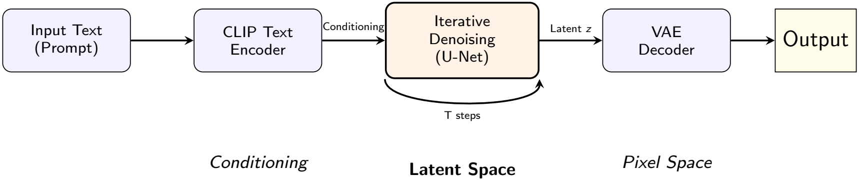


Figure: The Stable Diffusion Pipeline: Conditioning → Latent Denoising → Reconstruction

Medical Applications

Data Augmentation

- Generate rare disease cases
- Balance imbalanced datasets
- Improve classifier robustness

Anomaly Detection

- Learn normal distribution
- Flag deviations as anomalies
- Early disease detection

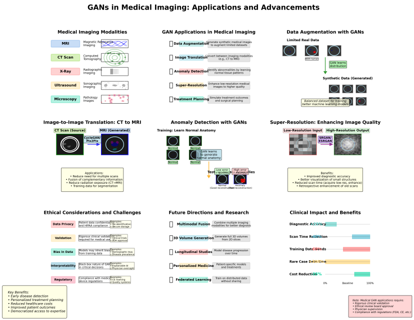


Figure: Generated MRI scans for data augmentation

Ethical Considerations

- Patient privacy
- Validation requirements
- Clinical approval needed

Diffusion Models vs GANs

Aspect	GANs	Diffusion Models
Training Stability	Low	High
Sample Quality	Excellent	Excellent
Sample Diversity	Good	Excellent
Training Speed	Fast	Slow
Sampling Speed	Fast	Slow
Mode Coverage	Partial	Full
Theoretical Understanding	Limited	Good

Table: Comparison of GANs and Diffusion Models (2023)

Hybrid Approaches

- GANs for fast sampling
- Diffusion for training stability
- Best of both worlds

Ethical Considerations

Deepfakes

- Face swapping
- Voice cloning
- Misinformation risks

Bias Amplification

- Dataset biases $\rightarrow$ model biases
- Underrepresentation issues
- Fairness concerns

Mitigation Strategies

- **Watermarking**: Embed invisible markers
- **Detection**: Train classifiers to detect fakes
- **Attribution**: Track model origins
- **Regulation**: Legal frameworks

Responsible AI Practices

- Transparency in generation
- Bias audits
- Ethical guidelines

Future Directions

3D Generation

- Neural radiance fields (NeRF)
- 3D-aware GANs
- Multi-view consistency

Multimodal Generation

- Text + Image + Audio
- Cross-modal retrieval
- Unified representations

Video Synthesis

- Temporal consistency
- Long-term dependencies
- Story generation

Theoretical Advances

- Better convergence guarantees
- Optimal architectures
- Generalization bounds

Grand Challenge

AGI-level creativity: Systems that can generate truly novel, valuable content across domains.

Conclusion

Key Takeaways

1. GANs revolutionized generative modeling through adversarial training
2. Theoretical foundation: Minimax game optimizing JS/Wasserstein distance
3. Practical challenges: Instability, mode collapse $\rightarrow$ many solutions (WGAN, StyleGAN, etc.)
4. State-of-the-art: Photorealistic image generation, many applications
5. Active research: 3D/video generation, ethical considerations, hybrid models

Resources for Further Study

- Awesome GAN Applications
- Original GAN paper
- DeepLearning.AI GAN Specialization

References I

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.
Generative adversarial nets.
Advances in Neural Information Processing Systems, 27:2672–2680, 2014.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou.
Wasserstein gan.
arXiv preprint arXiv:1701.07875, 2017.

[3] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville.
Improved training of wasserstein gans.
Advances in Neural Information Processing Systems, 30, 2017.

[4] Alec Radford, Luke Metz, and Soumith Chintala.
Unsupervised representation learning with deep convolutional generative adversarial networks.
arXiv preprint arXiv:1511.06434, 2015.

References II

[5] Mehdi Mirza and Simon Osindero.
Conditional generative adversarial nets.
arXiv preprint arXiv:1411.1784, 2014.

[6] Tero Karras, Samuli Laine, and Timo Aila.
A style-based generator architecture for generative adversarial networks.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.

[7] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila.
Analyzing and improving the image quality of stylegan.
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8110–8119, 2020.

[8] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila.
Alias-free generative adversarial networks.
Advances in Neural Information Processing Systems, 34:852–863, 2021.

References III

[9] Andrew Brock, Jeff Donahue, and Karen Simonyan.
Large scale gan training for high fidelity natural image synthesis.
arXiv preprint arXiv:1809.11096, 2018.

[10] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen.
Hierarchical text-conditional image generation with clip latents.
arXiv preprint arXiv:2204.06125, 2022.

[11] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al.
Imagen: Photorealistic text-to-image diffusion models with deep language understanding.
arXiv preprint arXiv:2205.11487, 2022.

[12] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer.
High-resolution image synthesis with latent diffusion models.
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10684–10695, 2022.