

Paper Critique: Wasserstein Generative Adversarial Networks

Dan Peng
Department of Computer Science
May 14, 2024
danpeng@unc.edu

1. Research Problem

1.1. What research problem does the paper address?

This paper tackles the inherent instability of Generative Adversarial Networks (GANs) by reimagining the core mathematical foundation of how we measure the distance between distributions. Like an architect finding a fatal flaw in a building's foundation, the authors identify why standard GANs often collapse or produce chaotic results, and rebuild using a more robust mathematical framework.

1.2. What is the motivation of the research work?

Traditional GANs suffer from a Jekyll-and-Hyde syndrome—they can create breathtaking results when they work, but training them resembles taming wild animals. The motivation stems from the authors' insight that this instability isn't due to implementation quirks but fundamental theoretical limitations of the Jensen-Shannon divergence, which fails catastrophically when comparing distributions on low-dimensional manifolds—precisely what happens in image generation.

2. Technical Novelty

2.1. What are the key technical challenges identified by the authors?

The authors identify three major hurdles: vanishing gradients that make learning impossible, mode collapse where generators produce limited varieties of samples, and the lack of meaningful metrics to track progress. Traditional GANs are like navigating with a compass that only works when you're already close to your destination—useless for the actual journey.

2.2. How significant is the technical contribution of the paper?

This work represents a paradigm shift rather than an incremental improvement. The authors don't just patch the GAN framework; they reconceptualize it using Earth

Mover's distance (Wasserstein-1), creating a theoretically sound alternative that considers the underlying geometry of the data. This is like upgrading from a flat map to a 3D globe for navigation—a fundamentally better representation of reality.

2.3. Identify 1-5 main strengths of the proposed approach.

- The WGAN provides meaningful learning curves that correlate with sample quality, finally giving researchers a compass that works
- Eliminates mode collapse, allowing generators to capture the full diversity of the target distribution
- Significantly improves stability, making GANs less finicky about hyperparameters and network architectures

2.4. Identify 1-5 main weaknesses of the proposed approach.

- The weight-clipping approach to enforce Lipschitz constraints is rather crude and can potentially limit the critic's expressive power
- The theoretical analysis, while elegant, doesn't fully account for the limitations of finite-capacity networks approximating the true Wasserstein distance

3. Empirical Results

3.1. Identify 1-5 key experimental results, and explain what they signify.

- WGANs successfully train on architectures where standard GANs fail completely, demonstrating their robustness to model specification
- The WGAN loss consistently correlates with sample quality—a revolutionary feature for researchers who previously had to judge models by eye

- WGANs show no evidence of mode collapse across all experiments, suggesting they truly capture the full data distribution rather than cherry-picking easy modes

3.2. Are there any weaknesses in the experimental section?

Yes, while the experimental section convincingly demonstrates WGAN's superiority, it lacks quantitative comparisons using established metrics like Inception Score or Fréchet Inception Distance. Additionally, the paper doesn't thoroughly explore different approaches to enforcing the Lipschitz constraint beyond weight clipping, which was later shown to be suboptimal. I'm also curious how WGANs would perform on more diverse datasets beyond bedroom images.

4. Summary

I'm 90% impressed by this paper's contributions. The WGAN approach feels like discovering a new continent in the world of generative modeling. By replacing the fundamentally flawed Jensen-Shannon divergence with the Wasserstein distance, the authors provide not just a technical improvement but a conceptual breakthrough that transforms how we think about comparing distributions. The few weaknesses are minor compared to the major theoretical and practical contributions. This work launched a new era of stable, reliable generative models that researchers could actually work with systematically rather than through trial and error.

5. QA Prompt for a Paper Discussion

5.1. Discussion Question

Why does the Earth Mover's distance provide more useful gradients than Jensen-Shannon divergence for training generative models?

5.2. Your Answer

The Earth Mover's distance behaves like a patient teacher rather than a harsh critic. When two distributions have little overlap—as often happens during training—Jensen-Shannon basically says "completely wrong!" and provides no useful feedback (mathematically: vanishing gradients). In contrast, the Earth Mover's distance measures the actual "transportation cost" to move probability mass between distributions, providing meaningful guidance even when they're far apart. It's like the difference between someone telling you "you're on the wrong continent" versus giving you actual directions to your destination—one helps you improve, the other just leaves you stranded.