Generative Adversarial Networks

Yuanzhi LOU
Xiuqi CHEN
Kunning SHEN

INTRODUCTION

Generative Adversarial Networks (GANs) have dramatically transformed the field of artificial intelligence by enabling models to generate new, realistic samples that mimic the distribution of original training data. This innovative approach not only enhances the diversity of generated images but also deepens our understanding of the underlying data structure. This project delves into the intricacies of traditional GANs alongside two of its notable variants: the Auxiliary Classifier GAN (AC-GAN) and the Wasserstein GAN (WGAN), each tailored to leverage the CIFAR-10 dataset for image generation.

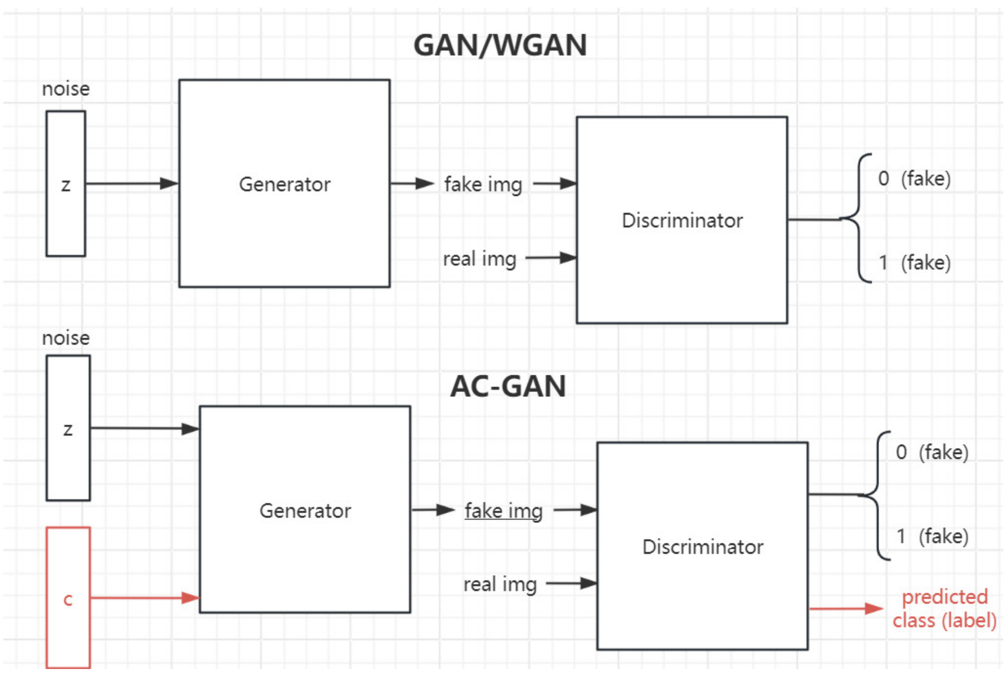
METHODOLOGY

Design of Traditional GAN, WGAN and AC-GAN

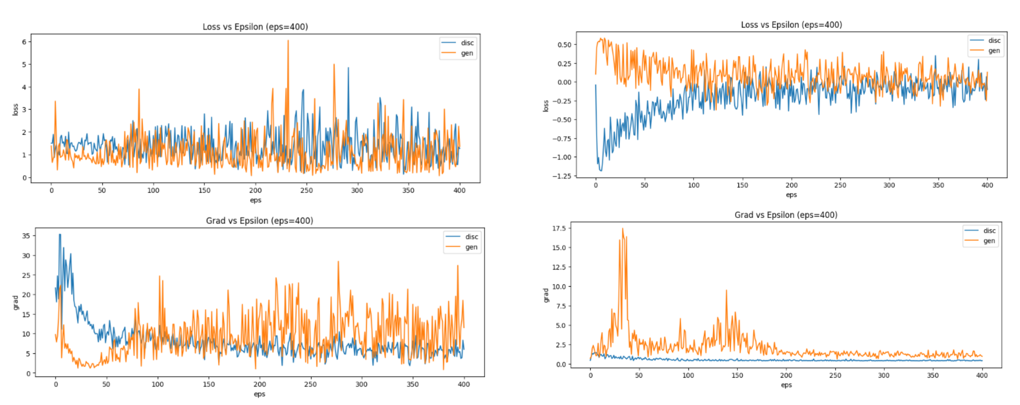
The **traditional GAN** architecture comprises two key components: a **Generator** and a **Discriminator**. The Generator creates fake images from a noise vector (z), while the Discriminator evaluates them against real images, learning to differentiate between genuine and artificial samples. The Discriminator outputs a binary classification, labeling images as real (1) or fake (0).

Building upon the traditional GAN, the **WGAN** modifies the training process to improve stability. WGAN alters the loss function (shown below), using a Wasserstein distance metric which provides more meaningful gradients and smoother training dynamics.

AC-GAN introduces an additional input called the class label (c) to the Generator. This enables the generation of class-specific images, and the Discriminator is extended to also predict the class of the input image. The AC-GAN not only discerns between real and fake images but also classifies the generated images into their respective categories, enhancing the control over the image generation process.



Traditional GAN vs WGAN



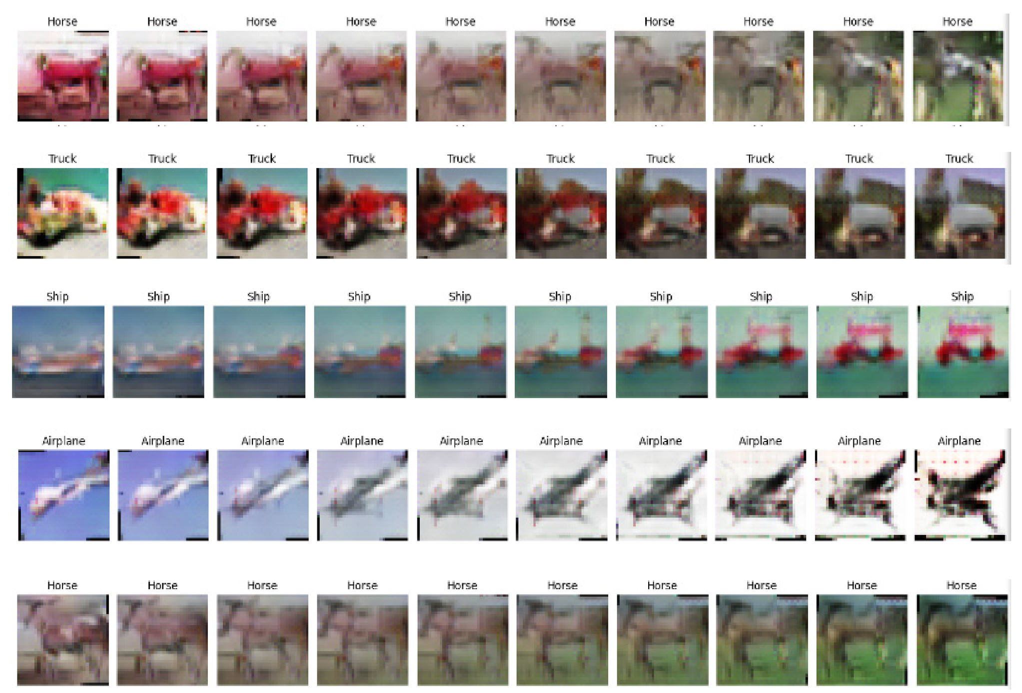
Traditional GAN

WGAN

As shown in the figure, WGAN is more stable. WGAN enhances training stability primarily through its use of the Wasserstein distance, providing steady gradients necessary for consistent learning, and weight clipping, which constrains the discriminator's output, preventing the gradients from exploding or vanishing.

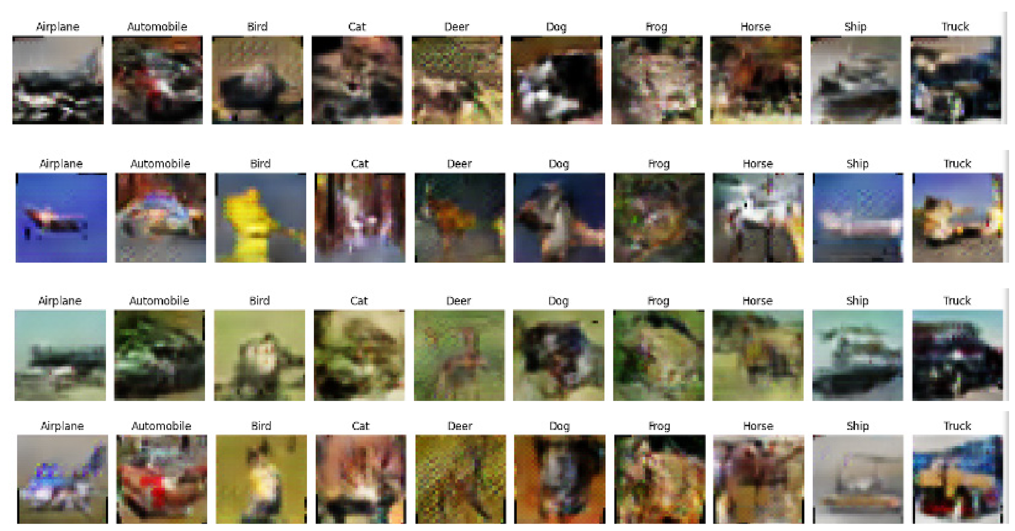
Latent space interpolations

The figure below highlights interpolations in the latent space between two different image samples within the same class. it can be seen that the transitions between the two image samples are relatively smooth.



Class-independent information

In this part, samples with the same "style" across several classes are produced when the AC-GAN is sampled with z fixed but the class label is changed, as shown below.



Performance evaluation

Generated Images

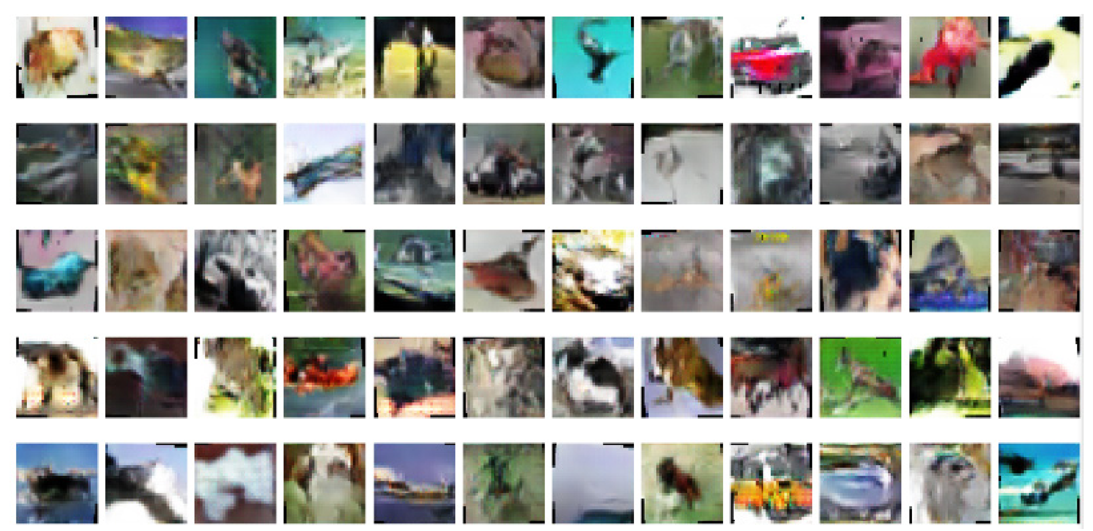
In the present analysis, we explore the efficacy of GANs in synthesizing photorealistic images, leveraging the CIFAR-10 dataset as a substrate for our generative tasks. Our exploration encompasses the conventional GAN framework as well as its two notable derivatives: the AC-GAN and the WGAN.

The foundational GAN model establishes a preliminary benchmark in our investigation. It exhibits a rudimentary capacity to capture and replicate the distribution inherent to the CIFAR-10 dataset. The images generated by this model, whilst exhibiting a degree of diversity, are characterized by a certain indistinctness, particularly in finer details. This is congruent with early GAN iterations, where the generator's capacity to synthesize detailed features is often limited by the discriminator's proficiency in differentiating between authentic and synthetic samples.

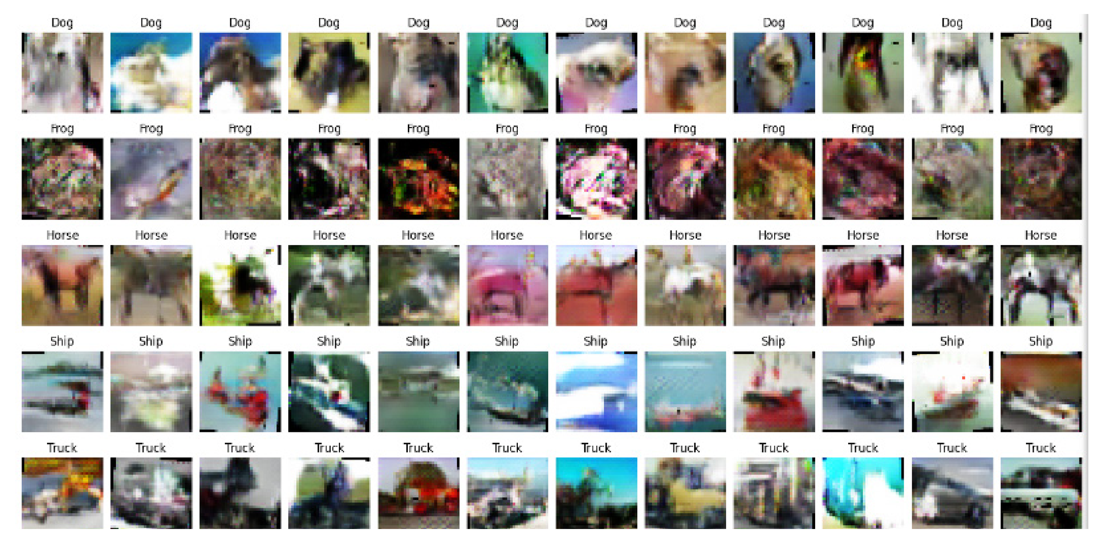
The implementation of the WGAN further refines the image synthesis process. The WGAN, with its innovative loss function grounded in the Wasserstein distance, aims to ameliorate the training stability and proffers a more interpretable gradient flow. The generated images exhibit an enhanced clarity with more structured and recognizable features, although some still display a residual blur—indicative of the challenges intrinsic to GAN-based image synthesis at higher resolutions.

Progressing to the AC-GAN architecture, there is a discernible enhancement in image clarity and specificity. This advancement is attributable to the AC-GAN's integration of class labels, which ostensibly guides the generative model towards producing images with improved definition that are class-coherent. The resultant images showcase better-defined contours and a more accurate color palette, indicating a sophisticated understanding of the dataset's categorical structure.

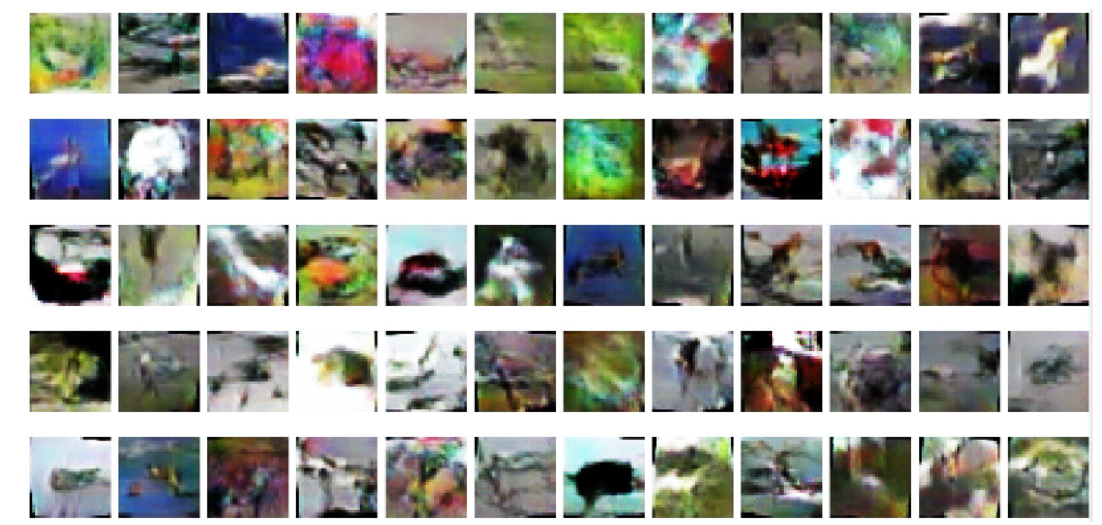
In summary, the comparative study of these GAN variants demonstrates a tangible progression in the model's ability to generate images that not› only depict diversity but also maintain fidelity to the original dataset's statistical properties. This progression is a testament to the incremental yet significant advancements in the field of deep generative models. Further research is warranted to optimize these models for even higher fidelity image generation, particularly in the context of fine details and textural accuracy.



Traditional GAN



AC-GAN



WGAN

Metrics

The table provided includes three metrics commonly employed to evaluate the performance of GANs:

Frechet Inception Distance (FID): This metric quantifies the difference between the distribution of generated images and real images. The FID calculates the distance between the feature vectors of the Inception network for both the real and generated images. The lower the FID, the closer the generated images are to the real image distribution in the feature space of the Inception model, implying higher fidelity and lower diversity within the generated images.

Inception Score (IS): The Inception Score uses the Inception network to classify generated images into predefined categories, thereby measuring two aspects: the clarity of images (as perceived by the pre-trained network) and the diversity of the generated dataset. A higher Inception Score indicates that the images are both distinct from each other (diversity) and recognizable as specific objects (clarity).

Kernel Inception Distance (KID): Similar to the FID, the KID measures the distance between the feature representations of real and generated images. However, KID uses a kernel-based method that is more robust to outliers and has a lower computational complexity. Like FID, a lower KID value is preferable, as it suggests the generated images are closer to the true data distribution.

Metric	Traditional GAN	WGAN	AC-GAN
Frechet Inception Distance	79.76	104.50	76.68
Inception Score	4.33	3.45	4.68
Kernel Inception Distance	0.06	0.08	0.06

Upon evaluating the provided metrics, it is discernible that the AC-GAN achieves the lowest Frechet Inception Distance and a high Inception Score relative to its counterparts, indicating a synthesis of images that are not only realistic but also diverse. Conversely, the WGAN manifests a higher FID, which suggests that the generated images deviate more significantly from the distribution of the real images. Nonetheless, it is crucial to note that while the WGAN underperforms in terms of FID, it does not necessarily imply inferiority in image quality, as this metric also encompasses diversity, which might be disproportionately represented.

In conclusion, based on the quantitative metrics presented, the AC-GAN appears to be the most effective model among the evaluated GAN variants, yielding the most favorable balance of image realism and diversity as quantified by these established benchmarks. It is recommended, however, to complement these quantitative assessments with qualitative human judgments to ensure a comprehensive evaluation of image quality.

CONCLUSION

The investigation into Generative Adversarial Networks showcases the advancement in stabilizing the training process and enhancing image generation quality. The Wasserstein GAN, with its innovative approach to measuring distances between distributions, has proved to be a pivotal development in stabilizing training. Simultaneously, the Auxiliary Classifier GAN's introduction of class labels has significantly enriched the diversity and fidelity of the generated images, offering a new dimension to the generative capabilities of these networks. This comparative analysis highlights the importance of architectural and loss function refinements in the progressive evolution of GANs.

REFERENCE

[1] Odena, Augustus, Christopher Olah, and Jonathon Shlens. "Conditional image synthesis with auxiliary classifier gans." International conference on machine learning. PMLR, 2017.

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

[3] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." International conference on machine learning. PMLR, 2017.