PROGRESSIVE GROWING CYCLEGAN FOR HIGH-RESOLUTION IMAGE TRANSLATION

Mu Niu
Duke University
Durham, NC 27708, USA
mu.niu@duke.edu

Hongyi Duan
Duke University
Durham, NC 27708, USA
hd162@duke.edu

ABSTRACT

This study investigates the integration of CycleGAN with Progressive Growing GAN to achieve high-resolution image-to-image translation using unpaired datasets. The proposed model employs progressive resolution scaling during training, enhancing both stability and image quality. The approach begins with low-resolution inputs and incrementally incorporates higher-resolution layers. The implementation focuses on replicating the brushstroke techniques and color palette characteristic of Claude Monet, transforming ordinary images into Monet-style paintings. Experimental evaluation highlights the model's ability to translate CIFAR-10 images into Monet-style artworks, demonstrating superior translation quality and stability. Performance metrics, including Fréchet Inception Distance and Inception Score, show notable improvements compared to the standard CycleGAN approach.

1 INTRODUCTION

Unpaired image-to-image translation aims to map images from a source domain to a target domain without paired samples. While CycleGAN has been effective in various tasks, its scalability to high-resolution images often results in unstable training and suboptimal outputs due to the complexity of high-resolution feature distributions and computational challenges associated with large-scale image processing.

This work introduces Progressive Growing CycleGAN, which integrates the resolution-scaling strategy from Progressive Growing GANs. By incrementally increasing image resolution and model complexity during training, the proposed method enables the model to first learn coarse features at lower resolutions and subsequently refine details at higher resolutions. This progressive training strategy enhances stability, mitigates mode collapse, and improves the fidelity of generated images.

Key contributions of this study include:

- Integrating progressive growing with CycleGAN for high resolution unpaired image-to-image translation.
- Demonstrating superior style transfer performance by translating CIFAR-10 images into Monet-inspired artworks.
- Providing a comprehensive evaluation using quantitative metrics such as Fréchet Inception Distance and Inception Score, as well as qualitative visual comparisons.

2 RELATED WORK

Unpaired image-to-image translation methods have gained significant attention for their ability to learn mappings between two image domains without paired training examples. CycleGAN Zhu et al. (2017b) introduced cycle consistency to ensure that translating from one domain to another and back again recovers the original input image. This key insight enabled stable training without paired data, allowing for a wide range of applications including style transfer, domain adaptation, and data augmentation.

However, directly applying CycleGAN at high resolutions poses training instability and image fidelity challenges. Progressive Growing GANs Karras et al. (2017) addressed similar issues in unconditional image synthesis by incrementally increasing the resolution of both generator and discriminator networks during training. This progressive scaling strategy results in more stable training dynamics and higher-quality outputs.

Our work extends these foundational ideas by integrating the progressive growing approach into the CycleGAN framework. By doing so, we aim to leverage the cycle consistency mechanism of CycleGAN while mitigating the instability issues that arise when attempting high-resolution translations. This integration allows for more stable, high-fidelity image-to-image translation, particularly evident in our artistic style transfer experiments.

3 METHODOLOGY

3.1 MODEL ARCHITECTURE

The Progressive Growing CycleGAN follows the standard CycleGAN formulation with two generators ($G_{A\to B}$ and $G_{B\to A}$) and two discriminators (D_A and D_B). However, unlike a fixed-resolution CycleGAN, our approach incrementally grows the network architecture to handle increasing image resolutions. We define a set of target resolutions (e.g., 16×16, 32×32, 64×64, 128×128, 256×256, 512×512) and progressively move through these stages during training.

Progressive Growing Generator:

As shown in the provided code, the generator starts with a low-resolution input block followed by:

- Downsampling Layers: These gradually increase the number of feature maps while halving the spatial resolution until reaching a bottleneck representation.
- Residual Blocks: The generator uses multiple residual blocks at each stage to refine features and learn complex mappings. Each residual block consists of convolutional layers with instance normalization and LeakyReLU activation, facilitating stable training and preserving fine details.
- Upsampling Layers: Symmetric to the downsampling path, these layers use transposed convolutions to increase the resolution back to the current target stage, reconstructing a refined, high-resolution image in the target domain.
- Final Convolution Layer and Tanh Activation: The final output of the generator is passed through a convolutional layer and a Tanh activation to produce images in the range [-1, 1].

At each progressive stage, additional residual and up/downsampling layers are added or adjusted based on the current target resolution. This modular design is reflected in the code, where the resolution parameter controls how many layers of each type are instantiated.

Progressive Discriminator:

The discriminator also adapts to increasing resolutions. At lower resolutions, it uses fewer layers, while higher resolutions introduce additional convolutional blocks:

- Convolutional and instance normalization layers are incrementally added.
- The number of layers scales with resolution, ensuring that the discriminator learns to distinguish fine details as the image size grows.

This progressive approach allows both the generator and discriminator to stably transition from learning coarse structures to capturing intricate, high-resolution details.

3.2 LOSS FUNCTIONS

Our loss functions remain consistent with standard CycleGAN:

- Adversarial Loss: Encourages the generated images to be indistinguishable from real target domain images.

- Cycle Consistency Loss: Ensures that translating from one domain to another and then back again recovers the original image. This maintains structural coherence.
- Identity Loss: Preserves color and overall tone, useful for style transfer tasks where stylistic changes are desired while retaining fundamental image composition.

These losses are computed at each resolution stage, guiding the model from coarse structure learning at low resolution to fine detail rendering at high resolution.

3.3 Training Pipeline and Implementation Details

We first train the model at a low resolution (e.g., 16x16), enabling it to grasp the fundamental domain mappings without being overwhelmed by high-frequency details. After 15 epochs, we increase the resolution (e.g., to 32x32) and continue training from the previously learned weights. This stepwise process continues until the desired high resolution (e.g., 512x512) is reached.

Data and Setup:

- Datasets: CIFAR-10 images serve as the source domain, and Claude Monet's artworks Zhu et al. (2017a) form the target domain.
- Framework: Implemented in PyTorch, utilizing nn.InstanceNorm2d and nn.LeakyReLU for stable training.
- Hyperparameters: Adam optimizer with a learning rate of 0.0002, batch size decreasing as resolution increases to manage computational load.
- Metrics: Fréchet Inception Distance and Inception Score are computed at each resolution stage.

The training loop, as illustrated in the provided code, trains each stage for 15 epochs. After completing a stage, the model and optimizers are updated for the next resolution.

4 Experiments

4.1 Dataset

- Source Domain (CIFAR-10): : A dataset comprising images from diverse classes, used as the input domain for image-to-image translation.
- Target Domain (Monet Artwork): High-resolution Monet paintings used as the target domain.
- All images are resized to match the current resolution stage during training.

4.2 Baseline Comparison: Standard CycleGAN

As a baseline, we employ the original CycleGAN framework without progressive growing. The standard CycleGAN model comprises two generator networks and two discriminator networks, each operating at the resolution of 512x512 from the start of training.

Generators

The generator adopts an encoder-bottleneck-decoder architecture. The encoder comprises five downsampling layers with instance normalization and ReLU activation, reducing the input image to a lower-resolution feature space. Six residual blocks refine these features, preserving spatial resolution while applying non-linear transformations to capture domain-specific patterns. The decoder progressively upsamples the features through five layers, reconstructing the transformed image at the target resolution.

Discriminators

The discriminator evaluates whether an image is real or generated, promoting local realism and sharp details. It employs stacked convolutional layers with instance normalization and LeakyReLU activation, progressively reducing spatial dimensions until producing a single output.

Both the baseline and the proposed Progressive Growing CycleGAN utilize identical hyperparameters, including learning rate, batch size, optimizer settings, and training iterations, ensuring a fair comparison. The baseline processes high-resolution images from the outset, while the Progressive Growing CycleGAN incrementally introduces higher resolutions, enabling coarse-to-fine learning and improved training stability.

By comparing the standard CycleGAN with the Progressive Growing CycleGAN, the study isolates the impact of progressive resolution scaling on training stability, convergence behavior, and image quality.

5 RESULTS

5.1 QUANTITATIVE METRICS 1

The reduced Fréchet Inception Distance of the Progressive Growing CycleGAN demonstrates improved alignment between generated images and real Monet artworks. While the Inception Score shows a minor decrease, qualitative analysis confirms that the proposed method produces images with greater stylistic coherence and visual appeal.

Table 1: Quantitative metrics at 512x512 resolution

Model	**FID**	**Inception Score**
Progressive Growing CycleGAN	270.4202	4.4124
Standard CycleGAN	273.0924	4.4329

5.2 QUALITATIVE ANALYSIS

The Progressive Growing CycleGAN 1 demonstrates superior performance over the Standard CycleGAN 2 in terms of visual quality and accurately replicating Monet's artistic style. The images generated by Progressive Growing CycleGAN exhibit smoother transitions, natural brushstroke-like textures, and a closer resemblance to Monet's artwork.

In contrast, the Standard CycleGAN faces challenges with high-resolution training, resulting in noisier outputs with visible artifacts and inconsistencies. For instance, the textures in scenes like the horse and truck appear chaotic and noisy, whereas the Progressive Growing CycleGAN maintains detail and balance in color throughout the image.

Additionally, the Progressive Growing CycleGAN effectively captures Monet's use of color and light, producing harmonious palettes and refined gradients. In comparison, the Standard CycleGAN often distorts colors or generates overly saturated images, failing to replicate Monet's style. Overall, the progressive approach ensures smoother, more detailed, and stylistically accurate outputs, establishing it as a superior method for high-quality image generation.

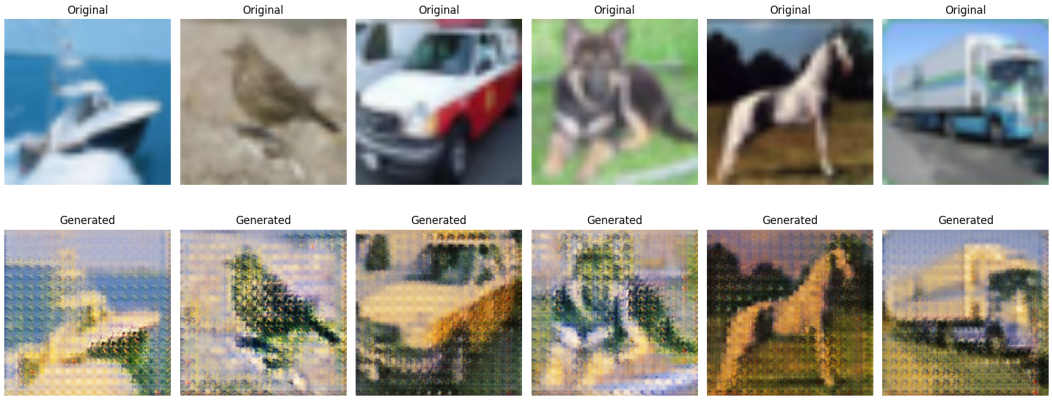

Figure 1: Progressive Growing CycleGAN Generated Samples

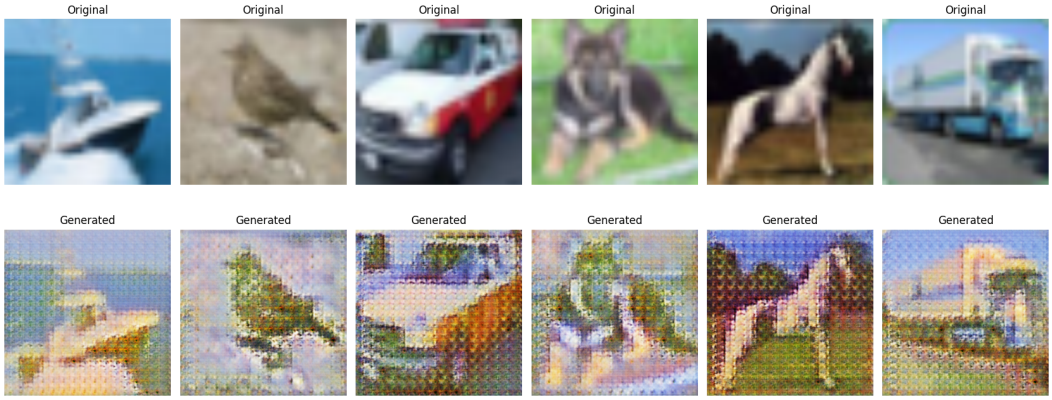

Figure 2: Standard CycleGAN Generated Samples

6 Limitations and Future Work

Despite the demonstrated improvements in training stability and image fidelity, several limitations remain. One notable constraint is the computational complexity associated with progressive resolution scaling. As the model progresses to higher resolutions, it necessitates additional training epochs per stage to ensure that each resolution level is sufficiently learned. This requirement can lead to longer training times and more substantial computational resource demands, limiting the feasibility of the approach for researchers and practitioners with less powerful hardware. Additionally, while our final resolution of 512x512 marks a significant step forward, pushing beyond this threshold would yield even more detailed and accurate images. However, scaling to such higher resolutions would further exacerbate computational demands, making training prohibitively expensive.

Another limitation lies in the choice of the initial image domain. We utilized the CIFAR-10 dataset, a collection of relatively small and diverse images that may not optimally align with Monet's distinct painterly style. Employing source domains with more visually compatible imagery—such as natural landscapes or photographs of gardens and water scenes, which more closely match Monet's thematic subjects—could improve the quality and coherence of the translated images. By starting from source images that share more stylistic attributes with the target domain, future research might achieve superior performance with less effort.

Looking ahead, future work can address these challenges on multiple fronts. First, exploring more efficient training schemes, such as adaptive resolution scheduling or leveraging distributed training techniques, may reduce computational overhead and streamline the learning process. Experiment-

ing with different normalization strategies, architectures, or regularization methods may further enhance performance and expedite convergence. Ultimately, these efforts will help push progressive image-to-image translation models toward producing even clearer, more detailed, and contextually appropriate high-resolution outputs.

7 ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Professor Vahid Tarokh, Yuwen Chen, and Andy Wang for their invaluable guidance, encouragement, and support throughout this research project. Their teaching and insights helped shape the direction of the work, and their patience and willingness to engage with questions and ideas provided a strong foundation for continuous learning and improvement. We are truly grateful for their expertise, mentorship, and the time they dedicated to helping us achieve our goals.

REFERENCES

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Cyclegan datasets. `https://efrosgans.eecs.berkeley.edu/cyclegan/datasets/`, 2017a.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017b.