Domain Expansion

Generative Neural Networks for the Sciences

in M.Sc. Informatik

Apl. Prof. Ullrich Köthe

April 23, 2024

Participants	**Matriculation number**	**Course of study**
Yuchen Li	3759412	Data and Computer Science
Xichu Yu	4732331	Data and Computer Science

Contributions:

Xichu Yu: Abstract, Introduction, Background, Methods(first 4)

Yuchen Li: Methods(last 3), Experiments and Results, Conclusion and Outlook

Abstract

Generative Adversarial Networks (GANs) have revolutionized image generation, demonstrating remarkable capability in producing highly realistic images. Traditionally, these models excel within the confines of their training data, often struggling when tasked with generating images beyond their initial domain. This limitation underscores the necessity for models that can transcend these boundaries without the need for retraining, a concept known as domain expansion. Domain expansion builds upon the principles of domain adaptation, enabling GANs to adapt from generating images in their original domain to creating diverse outputs in previously unexplored domains. This process leverages the untapped potential within the latent space, particularly focusing on dormant directions that, when activated, facilitate the introduction of novel features into the generated images. Our research employs StyleGAN as a foundational model, not only to exploit its advanced image generation capabilities but also to explore domain expansion's potential in broadening the model's versatility across various domains. By factorizing the latent space and harnessing dormant directions, we aim to enhance the model's creative scope, enabling the generation of a wider array of images beyond its original training dataset. This exploration not only enriches the diversity of generated images but also offers insights into the structured yet expansive nature of the latent space, marking a significant advancement in the field of generative modeling.

Introduction

In current deep learning research, generative models, especially Generative Adversarial Networks (GANs), have made remarkable achievements in the field of image generation, capable of producing highly realistic images.

Domain adaptation in the context of GANs refers to the process of fine-tuning a model trained on one domain (source) to perform effectively on a different, yet related, domain (target) without necessitating retraining from scratch. An exemplary illustration of domain adaptation can be seen in the application of StyleGANs, where a model trained to generate human faces can be adapted to generate caricatures or stylized portraits. This adaptation leverages the inherent flexibility of the model's latent space, aligning it more closely with the characteristics of the target domain, thereby extending the model's applicability and versatility.

However, most of the existing models focus on specific data domains, which means that the models are only able to generate the image types present in the training data after training. The concept of domain expansion within this context signifies a strategic shift from merely adapting generative models within their trained confines to empowering them to generate outputs beyond their initial training data's scope. This expansion is pivotal for overcoming the limitations of traditional GAN models, which, despite their prowess, often falter outside their training domain, leading to a constrained variety of outputs. The exploration into domain expansion seeks to breach these confines, pushing generative models to innovate beyond the data they were trained on, thereby enhancing their versatility and application across a broader spectrum of domains without necessitating retraining from scratch.

One of the critical methodologies in achieving domain expansion is the factorization of the latent space to unearth dormant directions – latent vectors that, despite their potential, remain unexploited in influencing the generated outputs. By activating these dormant directions, one can introduce novel features into the generated outputs, enriching the diversity of the generative model's capabilities. This process not only augments the model's ability to generate a wider array of images but also deepens our understanding of the latent space's structure and the potential for precise control within these generative systems. Through domain expansion, we seek to empower these models to innovate beyond their initial training constraints, enhancing their utility across a wider range of applications without the need for retraining.

In our project, we use StyleGAN as the generated model, not just harnessing the power of StyleGAN (Karras, T, 2020) to replicate its already impressive capabilities, but attempting to push the boundaries further to reproduce the concept of domain expansion. The methods are clearly presented in the paper though, we still wonder if our computing resources could suffice the training effort required.

Background

Data

AFHQ

The Animal Faces-HQ (AFHQ) dataset is a high-quality image dataset comprising animal faces from three distinct categories: cats, dogs, and wild animals. The dataset is designed to support and challenge the development of generative adversarial networks (GANs) and other machine learning models in producing high-resolution, photorealistic images of animal faces. Each category within the AFHQ dataset includes approximately 5,000 images, totaling around 15,000 images. The images are curated to ensure diversity in terms of age, breed, and facial expressions, providing a robust foundation for training and evaluating machine learning models.

In our project, we mainly use the AFHQ dataset on cats. Within the AFHQ, the cat dataset stands out for its concentrated focus on feline faces, encompassing a diverse collection of breeds, ages, and facial expressions. This subset contains approximately 5,000 high-quality images, all standardized to a resolution of 512x512 pixels, providing an excellent basis for training and evaluating advanced machine learning models.

Given the dataset's composition, it offers a unique opportunity to address specific research questions related to the generation of animal faces, such as capturing and replicating the intricate details and variations found in the facial features of different species.

CIFAR-10

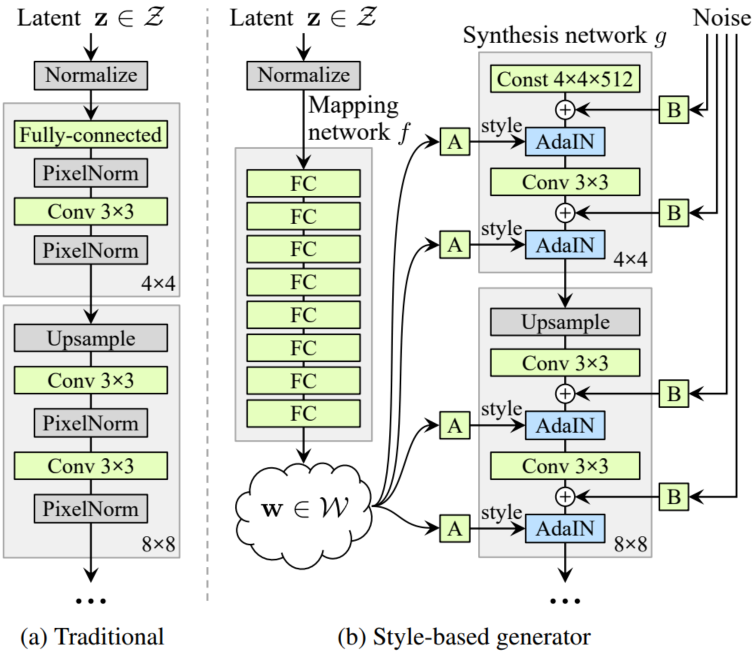
The CIFAR-10 dataset is a well-known collection of images commonly used for training machine learning and computer vision algorithms. It consists of 60,000 32x32 color images in 10 different classes, with 6,000 images per class. The dataset is divided into 50,000 training images and 10,000 testing images. The ten classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each class is mutually exclusive, meaning there is no overlap between them. CIFAR-10 is widely recognized for its utility in developing and benchmarking machine learning models, especially in the field of image recognition and classification. It provides a balanced dataset that challenges models to accurately distinguish between distinct object categories, making it a staple in the machine learning community for educational and research purposes.

StyleGAN

Style Generative Adversarial Network (Karras T, 2019) originates from the GAN framework, which introduces a novel architecture that allows for fine control over the generated images' style at different levels of detail, from coarse features like pose and shape to finer attributes like textures. The architecture of StyleGAN comprises two main components: a generator and a discriminator, following the standard GAN framework. However, the generator in StyleGAN is distinctively designed to incorporate a mapping network and a synthesis network.

The mapping network is the foundation of style. The mapping network $f : Z \to W$ takes a latent code z in a latent space Z, and transforms it into an intermediate latent space W, which encapsulates the "style" of the generated images. After this process we get a style code w. This transformation is pivotal as it distills random input noise into a structured style representation. The style code encapsulates the essence of the image's style, allowing for granular control over the generated image's appearance, from coarse attributes like overall structure and posture to finer details such as texture and color nuances.

The discriminator, designed to distinguish between real and generated images, aids in training by providing feedback to the generator. Additionally, StyleGAN introduces techniques like Adaptive Instance Normalization (AdaIN) and noise injection at different layers of the synthesis network to enhance detail and variability. The synthesis network is where the magic of image creation happens. Utilizing the style code, along with additional layers of noise for added detail, it intricately crafts the final image through a series of convolutional layers. Each layer is equipped with Adaptive Instance Normalization (AdaIN), a technique that adjusts layer activations based on the style code, thereby injecting the desired style into the image at various levels of detail. The synthesis network's ability to modulate features across different scales enables the creation of complex, textured, and visually appealing images that are rich in diversity.



Architecture of StyleGANs

Latent direction

In the realm of generative models, the concept of a latent space serves as a foundational pillar. It represents an abstract, high-dimensional mathematical space, wherein each point uniquely corresponds to a specific output, such as a detailed image. This latent space is derived through the model's learning process, which internalizes the distribution and intrinsic features of the input data, encapsulating the essence of the data's variability and complexity. Within this space, the proximity of points is indicative of the similarity in their corresponding outputs, allowing for nuanced interpolations and transitions between generated outcomes.

Critical to navigating this latent space are the latent directions—vectors signifying pathways along which traversal induces modifications in the output. These directions are akin to dials that adjust various attributes of the output images, encompassing broad aspects like posture or overall structure, down to minute details like textures or facial expressions. The existence of such controllable directions hints at an underlying structure within the latent space, offering the tantalizing prospect of meticulously steering the generative process.

However, not all latent directions wield this transformative power. Some, known as dormant directions, lie untapped, exerting negligible influence on the generated images. Their presence, though seemingly inert, harbors untold potential for innovation. The quest for domain expansion within generative models capitalizes on these dormant directions. By employing techniques to factorize the latent space, researchers aim to unearth and activate these dormant vectors, thereby extending the generative model's capability to forge new features and visual domains, all while preserving its core generative prowess.

This venture into domain expansion not only broadens the horizons of image generation but also deepens our comprehension of the latent space's architecture. It underscores the latent space's capacity not merely as a repository of data-derived features but as a rich, malleable canvas for creative exploration and expansion beyond the confines of the original training dataset.

SeFA

Closed-form factorization of Latent Semantics (Shen Y, 2021), so called SeFA, presents an innovative approach to uncover interpretable directions within the latent spaces of pre-trained Generative Adversarial Networks, notably focusing on the StyleGAN architecture. Unlike traditional methods that often rely on dataset annotations or optimization-based strategies, SeFA introduces a more straightforward, unsupervised method for latent space exploration through the application of closed-form factorization on the weight matrices of the generator's layers. This process aims to identify latent directions that correspond to meaningful semantic variations in the output images, thus enhancing the interpretability and utility of GAN models.

At the heart of SeFA is the analysis of the generator's fully connected layers, which transform initial latent codes into intermediate style vectors. The SeFA algorithm understands which directions are important in the potential space by analyzing the weights of the generator network, in particular the first layer weights that are directly connected to the potential vectors. Through singular value decomposition (SVD) applied to these weight matrices, SeFA identifies orthogonal directions that encapsulate the generator's most significant modes of variation. This process can be mathematically represented as:

$$W = U\Sigma V^T$$

where W is the weight matrix of a fully connected layer, U and V are matrices containing the left and right singular vectors, and Σ is a diagonal matrix with singular values. The columns of V (or U, depending on the layer structure) provide the directions in the latent space associated with the most variance in the data representation.
By leveraging these directions, SeFA facilitates a more nuanced control over the generated images, allowing users to manipulate specific image attributes such as facial expressions, age, and scenery elements without extensive manual intervention. This method not only propels the understanding of the GAN's latent space structure but also broadens the horizon for creative and practical applications of GANs in various domains.

Methods

Factorize latent space

In domain expansion, identifying dormant directions within a model's latent space is crucial. We need to find dormant directions in the latent space of the model. These are directions that the original model barely uses when generating an image and therefore have no significant effect on the resulting image. Utilizing such dormant directions offers a novel pathway for domain expansion, enabling the incorporation of new domain-specific knowledge without compromising the original domain's integrity.
Concretely, as we mention in the background, following SeFA, we obtain a semantic and orthogonal basis vector of the latent space, which contains a set of basis vectors. These basis vectors form a base framework for the latent space, and each basis vector represents a direction of change in the potential space. The directions unearthed by SeFA inherently align with semantically meaningful attributes in the generated images, such as changes in pose, facial expressions, lighting conditions, and textures. These directions, so-called dormant directions, which SeFA can efficiently uncover through its singular value decomposition-based approach, represent variations that are not explicitly modeled or utilized during the initial training process.

By identifying these dormant directions, domain expansion methods can activate previously unused aspects of the model's representational capacity, leading to the generation of new, diverse, and semantically rich outputs that extend beyond the original training dataset. Since these directions were hardly utilized during the original training process, new domain information can be encoded into these dormant directions without interfering with the model's ability to learn and generate the original domain. In this way, the generative model is able to cover both the original domain and the newly introduced domains and maintain good performance on their respective domains.

Since singular values are commonly sorted in decreasing orders, the last basis vectors are most dormant. So we dedicate a single dormant direction as v_i, which is used to repurpose. The rest of the directions in latent space form the base subspace Z_{base}, as

$$Z_{base} = span(v_{N+1}, ..., v_D) + \bar{z}$$

where the $\bar{z}$ is the mean of the latent distribution used to train the generator. And the repurposed subspace Z_i can be represented as

$$Z_i = Z_{base} + sv_i$$

The training process of domain expansion is based on the base subspace Z_{base} and repurposed subspace Z_i, and the repurposed subspace Z_i is the only part that affects the training object. Domain expansion tries to keep the variation factors inherited from Z_{base} and limit the new variation factors to a single repurposed direction as v_i.

This process hinges on a strategic balance between these subspaces. The goal is to retain the variation factors inherent to the base subspace while confining new variations to the repurposed direction. Through meticulous manipulation and optimization of these latent directions, generative models can transcend their original confines, offering richer and more diverse generative capabilities.

L2 pixel loss

L2 pixel loss is a common loss function used to calculate the Euclidean distance between the generated image and the target image at the pixel level. This loss helps to ensure that the generated image is close to the target image at the pixel level. In domain expansion, the L2 pixel loss can help keep the output of the expanded model on the original domain input visually consistent with the output of the original model. The formula can be expressed as:

$$L_{\mathrm{L2}} = \|\mathrm{G}_{\mathrm{exp}}(z) - \mathrm{G}_{\mathrm{orig}}(z)\|_2^2$$

where $\mathrm{G}_{\mathrm{exp}}(z)$ is the output of the expansion model, and $\mathrm{G}_{\mathrm{orig}}(z)$ is the output of the source generator.

LPIPS (Learned Perceptual Image Patch Similarity)

LPIPS (Zhang, R, 2018) is a more advanced loss function that measures the perceived similarity between two images. Unlike the L2 pixel loss, the LPIPS takes into account the structural and textural information of the image and better captures the human perception of image similarity. In domain extension scenarios, the use of LPIPS helps to ensure that the images generated by the extended model are perceptually consistent with the images generated by the original model. The formula for LPIPS can be expressed as:

$$L_{\mathrm{LPIPS}} = \mathrm{LPIPS}(\mathrm{G}_{\mathrm{exp}}(z), \mathrm{G}_{\mathrm{orig}}(z))$$

Regularization

In domain expansion, another crucial aspect is the maintenance of the original domain's performance while integrating new domain features. This balance is achieved through regularization techniques designed to prevent catastrophic forgetting (McCloskey, M, 1989), a phenomenon where a model loses its ability to perform tasks it was previously trained on. The regularization process ensures that while the model acquires capabilities in new domains, it retains its proficiency in the original domain. The goal of regularization is to add new domain features while ensuring that the model does not forget or compromise its performance in the original domain.

Domain adaptation loss

Firstly, after we define the repurposed subspace Z_i of the latent space Z, we should make sure that the generator can learn and generate data from the new domain D_i. Therefore, the latent code $z \in Z$, sampled from distribution $p(z)$ defined on the latent space Z, should be projected to the repurposed subspace Z_i. In this case, each new domain can define the new and independent training objectives, thus supporting the generation of multiple domains simultaneously in a single model. The projection formula is defined as

$$\mathrm{proj}_{Z_i}(z) = \sum_{j=N+1}^{D} \langle v_j, (z - \bar{z}) \rangle v_j + \bar{z} + s v_i$$

where v_j is orthogonal basis vectors, v_i is the single dormant direction, $\bar{z}$ is the mean of the latents distribution used to train the generator.

The generator should be continued to optimize to fit the original training data and maintain the quality of generation about the original domain Z, the training loss is defined as

$$L_{\text{expand}} = \sum_{i=1}^{N} \mathbb{E}_{z \sim p_{\text{i}}(z)} L_{\text{i}}(G(\mathrm{proj}_{Z_{\text{i}}}(z)))$$

Domain expansion loss

Optimizing L_{expand} could help the model generate data from a new domain D_i and constrain the base subspace Z_{base}. However, it cannot make sure the features the model learns from D_i would not "leak" to Z_{base}. "Leak" refers to the fact that the newly learned features or information not only affects the model's ability to process the new task, but also accidentally affects the model's ability to process the old task, which may lead to a decrease in the model's performance on the old task. Therefore, the regularization techniques are crucial to prevent this phenomenon. Here are two useful regularization techniques:

In order to ensure that the generated image not only closely matches the target domain at the pixel level, but also maintains perceptual integrity according to human judgement, domain expansion uses a weighted combination of an L2 pixel loss and LPIPS, as

$$L_{\text{recon}} = \lambda_{\text{L2}} L_{\text{L2}} + \lambda_{\text{LPIPS}} L_{\text{LPIPS}}$$

where $\lambda_{L2} = \lambda_{LPIPS} = 10$. In this way, models are encouraged to add new domain features while maintaining the quality of the generation of the original domain as constant as possible.

These regularization techniques should only be applied to the base subspace, as the aim is to allow other subspaces to vary to learn new behaviors. This is achieved by projecting latent code z onto the base subspace Z_{base} and then computing the regularization term. The overall regularization goal formula is:

$$L_{\text{reg}} = \mathbb{E}_{z \sim p_{\text{src}}(z)} \left[\lambda_{\text{src}} L_{\text{src}}(G(\text{proj}_{Z_{\text{base}}}(z))) + L_{\text{recon}}(G(\text{proj}_{Z_{\text{base}}}(z))) \right]$$

λ_{src} is a trade-off parameter that balances the importance between source and reconstruction loss. $\lambda_{src} = 1$

The final regularization goal of domain expansion is defined as

$$L_{full} = L_{expand} + L_{reg}$$

This formula combines an expansion loss (learning new domain features) and a regularization loss (retaining the original domain features) to ensure that the model does not lose mastery of the original domain as it extends to new domains.

The choice of these regularization strategies is rooted in a comprehensive understanding of how generative models interact with and learn from varied data domains. By preventing feature "leakage" — where enhancements for new domain tasks inadvertently degrade performance on original tasks — these techniques ensure a model's expansion does not dilute its established strengths. This careful preservation is critical for maintaining the generative model's diversity and utility across a wide spectrum of applications, without necessitating exhaustive retraining for each new domain encountered.

Regularization in domain expansion not only safeguards against the pitfalls of catastrophic forgetting but also paves the way for generative models to transcend their training confines. By leveraging these advanced regularization techniques, models can achieve a delicate equilibrium: evolving to capture the essence of new domains while upholding the integrity and performance standards set by their original training datasets.

Evaluation metrics

Fréchet Inception Distance(FID)

Fréchet Inception Distance (Yu Y, 2021), so called FID, is a metric used to evaluate the quality of images generated by generative models, like GANs (Generative Adversarial Networks). It measures the similarity between two datasets of images, typically between a set of generated images and a set of real images. The "50k" in "FID50k" refers to the evaluation being performed using 50,000 images from each dataset.

FID calculates the distance between the feature vectors of the real and generated images. These feature vectors are extracted using a specific layer of the Inception-v3 network, a pre-trained deep learning model. The distance metric used is the Fréchet distance (also known as the Wasserstein-2 distance) between the multivariate Gaussians of the feature vectors of the real and generated images.

A lower FID score indicates that the generated images are more similar to the real ones, suggesting better quality and realism of the generated images. FID50k is a common benchmark for assessing the performance of generative models, particularly in tasks like image generation and style transfer.

Kernel Inception Distance(KID)

Kernel Inception Distance (Bińkowski M, 2018), so-called KID, is an evaluation metric designed to measure the quality of images produced by generative models, such as GANs. Similar to FID, KID compares the distribution of generated images to that of real images. The "50k" in "KID50k" signifies that the metric is calculated using 50,000 images from both the generated and real datasets. The feature vectors for comparison are obtained from an intermediate layer of the Inception network, a deep learning model pre-trained on a large image dataset.

However, KID utilizes the Maximum Mean Discrepancy (MMD) with a polynomial kernel to compare the feature representations, offering a different approach from the Gaussian assumption used in FID.

A lower KID score indicates that the generated images are closer in distribution to the real images, suggesting higher quality and realism. Unlike FID, KID is unbiased and does not suffer from issues related to the size of the sample set, making it a robust alternative for evaluating the performance of image generative models, especially in scenarios where the assumption of Gaussian distribution may not hold.

Experiments and Results

We conducted several experiments to validate the implementation and explore the training process. The table below gives a short summary of our experimental results.

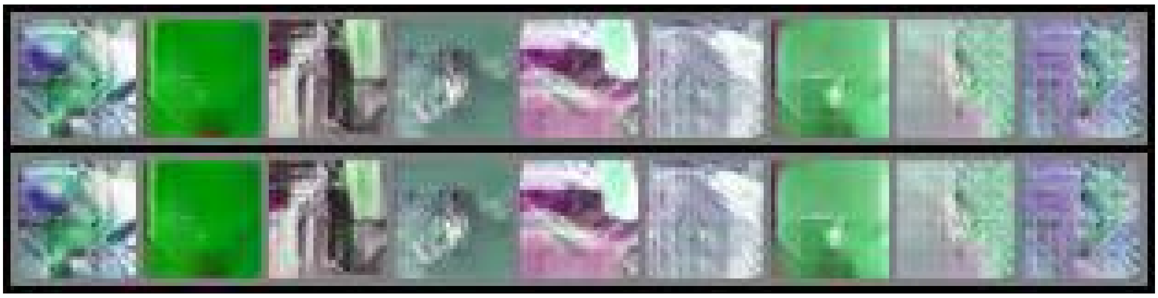
Dataset	Metric	Training Resolution	Training Set Size	Evaluation Result	Total Time
CIFAR-10	is50k	32	10k	3.64	1d 16h
AFHQ Cat	fid50k	256	5k	275.25	1d 17h
AFHQ Cat	fid50k	128	5k	287.87	19h 35m
AFHQ Cat	fid50k	512	5k	14.96	Lost due to disconnection from Colab

CIFAR-10 Dataset

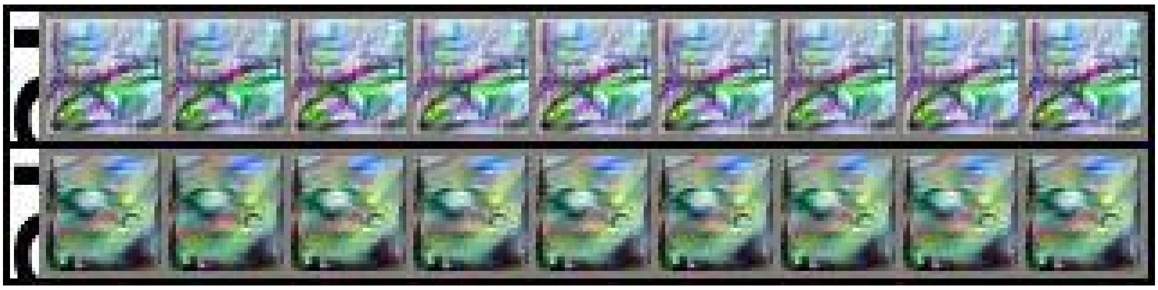
NADA Adaption task: “photo” -> “sketch”, “photo” -> “oil painting”

Training set augmentation: default augmentation pipeline

We first start with the CIFAR-10 dataset because we would like to start with a dataset with small-sized images to generate fast results. And we also halved the channels in both the generator network and the discriminator network to accelerate the training.



Base images



Expanded images

However, the results did not come off very well, even though the evaluation results appeared to be good. Apart from the quality, the diversity also disqualifies our model. The model produces almost the same in the repurposed dimensions.

There are many categories of objects present in CIFAR-10 and therefore much more training effort is probably required to generate images of satisfactory quality. This also answers why the expanded images are also unrecognizable.

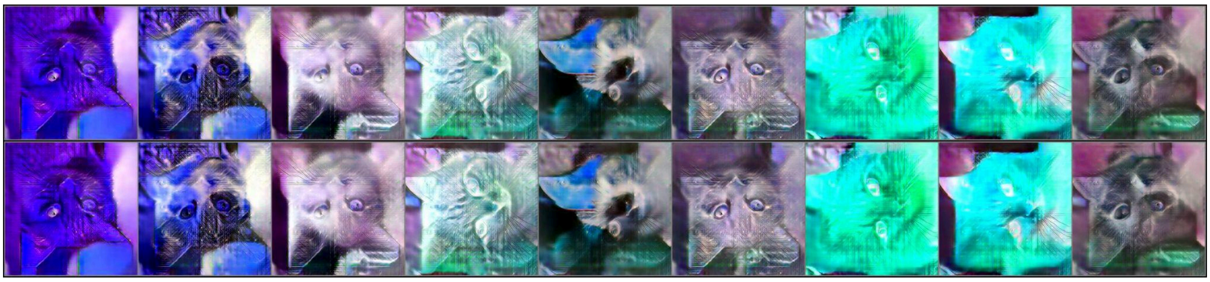
In response to these mixed results, we opted to shift our focus towards the AFHQ Cat dataset. This pivot was made in the hope that a more homogenous set of images would allow us to fine-tune our model's ability to generate high-quality and diverse images, thereby overcoming the obstacles we encountered with CIFAR-10. This step is anticipated to be a crucial move in refining our model's performance and achieving the high standards we set out to reach.

AFHQ Cat Dataset (compressed)

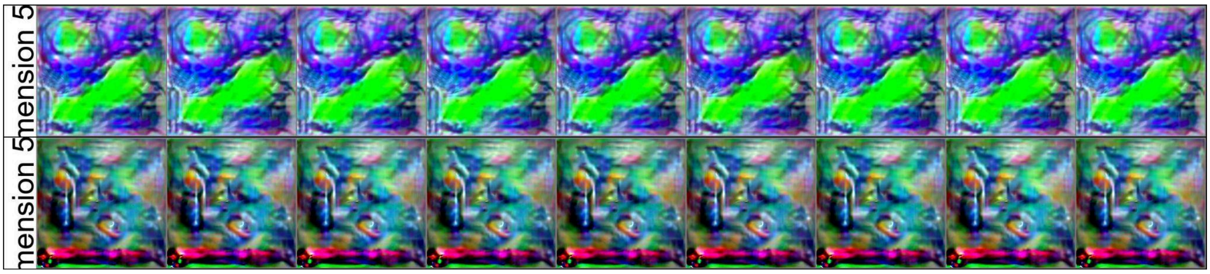
NADA Adaption task: “photo” -> “sketch”, “photo” -> “oil painting”

Training set augmentation: default augmentation pipeline

We first compressed the dataset from 512 pixels to 256 again with the aim of accelerating the training process. And we halved the channels again (¼ as original). We spent the same time on this training process and stopped manually. The result of this experiment is shown below.



Base images



Expanded images

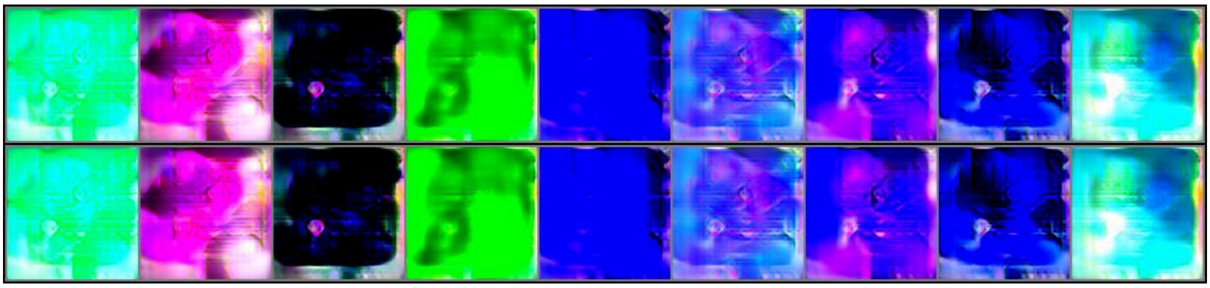
With limited training resources and time, the basic GAN generates images of cats that are recognizable but distorted in color. However, the adapted images still have the exact problem described in the previous experiment. The diversity is very low and the generated images are still of poor quality.

AFHQ Cat Dataset (further compressed)

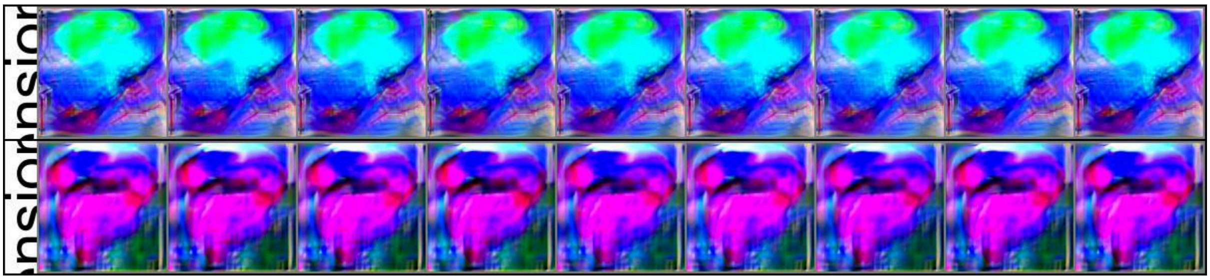
NADA Adaption task: “photo of a cat” -> “sketch”, “photo of a cat” -> “tongue out”

Training set augmentation: default augmentation pipeline

To solve these problems, we tried another different set of experiment settings. We use the AFHQ Cat dataset compressed to 128 pixels and use the same network architecture as in experiment 1, expecting the GAN model would produce better results.



Base images



Expanded images

However, the generated images turned out to be even worse. Moreover, we found the exploration of resolution-complexity is of very low efficacy which takes a lot of GPU runtime and produces little useful information.

AFHQ Cat Dataset (pre-trained)

NADA Adaption task: “photo of a cat” -> “sketch”, “photo of a cat” -> “tongue out”

Training set augmentation: default augmentation pipeline and mirroring

In the end, we chose to do our domain expansion on pre-trained StyleGAN models which are directly available on GitHub. But the full-sized GAN would take up to several hours for a 4080 GPU to finish only one epoch. To deal with this problem, we adapt our project to Google Colab Pro with more powerful computing resources.

Base images

Expanded images

We can observe from the results that with the pre-trained StyleGAN model, we achieved both low evaluation loss and rerlatively good generation quality. This validates our method to be correct.

Conclusion and Outlook

In our research project, we basically yield reasonable results within limited computing resources. However, the quality of the generated images is not very satisfactory. We have been stuck by the problem caused by CUDA compatibility for several days which restrained our time on more training experiments. Moreover, the training effort required is more than we expected. It was a huge work to train a good StyleGAN even with a simplified network architecture and compressed training set. However, expanding a pre-trained StyleGAN means no simplifications mentioned before are possible.

Generally speaking, our experimental results replicate the functionality of the original work. The trained latent dimensions can present certain features while keeping the original generated images not affected.

To keep our research in this direction, we could continue our attempt on the CIFAR-10 dataset to explore if this method could work on small images. More interestingly, we could try to generalize this method with more kinds of generative models such as diffusion models.

Reference

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8110-8119).

Shen, Y., & Zhou, B. (2021). Closed-form factorization of latent semantics in gans. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1532-1540).

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 586-595).

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Psychology of learning and motivation (Vol. 24, pp. 109-165). Academic Press.

Wu, C., Herranz, L., Liu, X., Van De Weijer, J., & Raducanu, B. (2018). Memory replay gans: Learning to generate new categories without forgetting. Advances in neural information processing systems, 31.

Yu, Y., Zhang, W., & Deng, Y. (2021). Frechet inception distance (fid) for evaluating gans. China University of Mining Technology Beijing Graduate School.

Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying mmd gans. arXiv preprint arXiv:1801.01401.