Final Report - Generalized one-shot domain adaptation via GANs

Abstract

This project presents a novel approach for the one-shot adaptation of Generative Adversarial Networks (GANs) from a source domain to a target domain. Unlike previous works that primarily focus on style transfer, the proposed framework tackles both style and entity transfer. The adaptation process is decoupled into two parts: transferring the global style, such as texture and color, and generating new entities that do not exist in the source domain. To achieve this, we introduce a reference image along with its binary entity mask. The main idea is to minimize the gap between the distributions of the reference image and the generated images using the sliced Wasserstein distance. To facilitate this process, an auxiliary network separates the entity and style transfer in the generator. The results demonstrate the effectiveness of the proposed approach in achieving generalized one-shot adaptation for both style and entity transfer tasks.

Introduction

GAN adaptation has emerged as a promising research area that aims to improve the knowledge gained from pre-trained GANs trained [1,16]. This adaptation process involves addressing data scarcity and accelerating training in new domains. It encompasses different scenarios, such as few-shot [18, 17, 3, 2], one-shot, and zero-shot, each presenting its own challenges and opportunities for transferring knowledge.

Recent efforts have focused on the more complex one-shot adaptation scenario. These endeavors concentrate on adapting GANs to target domains that share similar artistic styles with a given image. By utilizing GAN inversion techniques [4], these adapted models have found applications in creative tasks like image style transfer and manipulation [7].

However, existing one-shot adaptation approaches have limitations within their task settings [5, 7]. They often fail to fully consider the content of the source exemplar, which includes elements beyond artistic style, such as color, texture, and objects. Previous works primarily focused on style transfer and overlooked the transfer of entities. This oversight leads to two issues: firstly, in cases where entities are essential domain features, they should be transferred alongside the style; secondly, large entities, like masks, can bias the style extraction process for facial regions, resulting in unwanted artifacts in the adapted output.

To overcome these limitations and enable a more comprehensive and flexible adaptation process, we propose a novel task called generalized one-shot GAN adaptation. This task builds upon existing approaches by incorporating an additional binary mask that precisely defines the target domain and labels the entities of interest. By incorporating this binary mask, our approach facilitates the transfer of both artistic style and entities, making it suitable for more complex and diverse scenarios. Notably, previous works can be seen as special cases where the binary mask is entirely filled with zeros.

We incorporate an auxiliary network dedicated to entity generation, while the original generator focuses on generating stylized images. To learn the domain knowledge, our framework directly minimizes the divergence between the distributions of exemplars and generated outputs using sliced Wasserstein distance. We also introduce style fixation, anchoring all generated outputs with the exemplary style through transformations in the latent space. Furthermore, inspired by cross-domain consistency [10] and classic manifold learning techniques [12 ,2], we propose a variant Laplacian regularization to preserve the geometric structure of the source generative manifold and prevent content distortion during training.

Our experiments, with and without entities, validate the effectiveness of our proposed framework. The results demonstrate that our approach successfully leverages cross-domain correspondence and achieves high-quality transfer. Furthermore, the adapted models showcase their versatility in performing diverse image manipulation tasks.

Methodology and Simulation

We have developed a framework for one-shot domain adaptation, allowing adaptation of pre-trained StyleGAN models on FFHQ Dataset [16] to diverse domains, as showcased in Figure 1, which demonstrates the impressive results. It includes a modified network architecture with a target generator (G_t) and an auxiliary network (aux) for handling entities. The training process involves style fixation, exemplar reconstruction, distribution learning using sliced Wasserstein distance (SWD) [9] for style and entity transfer, and manifold regularization to preserve the content structure.

Style fixation and exemplar reconstruction: StyleGAN [16] is known for its disentangled latent spaces, where one of them controls the style and the other controls the content of the synthesis. In this framework, the exemplary style is roughly transferred to other syntheses by fixing the style code. The content part of the latent code is preserved, and the style part is replaced with the exemplary style code. This style fixation process ensures that the content of the original synthesis remains unchanged. The framework utilizes a reconstruction loss to narrow the visual gap between the reconstructed image and the exemplar, considering structural similarity, perceptual loss, and mask similarity. The reconstruction loss acts on the target generator G_t. It is important to note that the framework incorporates style fixation as a pre-processing step to enhance subsequent learning and improve style quality.

Fig. 1. The generalized one-shot GAN adaptation task examples.

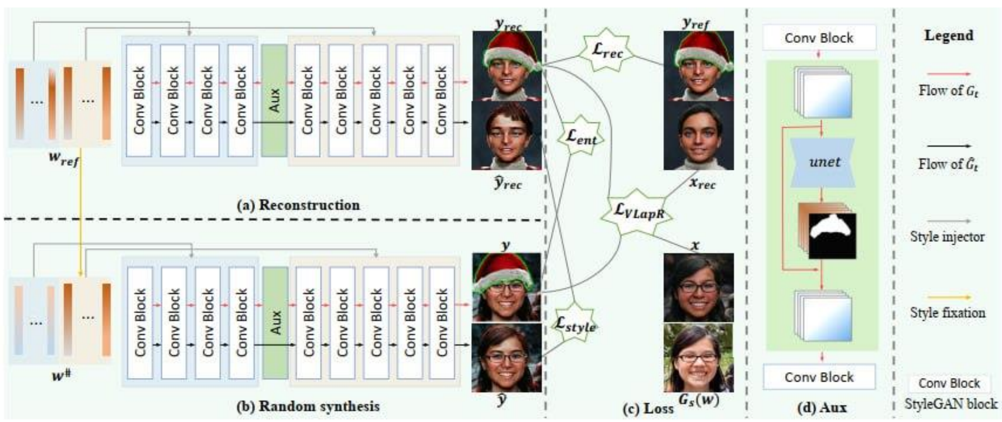

Fig. 2. The generalized one-shot GAN adaption framework. [17].

Internal distribution: The internal patch distribution of an image contains valuable information and minimizing the divergence of these distributions is crucial for tasks like image generation and style transfer. An efficient approach called sliced Wasserstein distance (SWD) [9] is introduced which measures the divergence between empirical internal distributions in the framework which SWD is employed to minimize the divergence between the internal distributions of synthesized images and the reference. The style loss is computed using SWD applied to feature maps extracted by pre-trained convolutional layers, ensuring accurate style transfer. The entity loss measures the divergence between the entity's internal distribution and the reconstructed entity from the target, preventing overfitting and enabling precise entity transfer. By incorporating SWD-based style and entity losses, the framework achieves effective style and entity transfer while overcoming the limitations of traditional GAN training methods.

To mitigate content distortion, a variant Laplacian regularization (LV LapR) is introduced in the framework. LV LapR preserves cross-domain correspondence and geometric structure by ensuring smooth semantic differences and isometric relationships between syntheses. Compared to the cross-domain distance consistency loss (LCDC), LV LapR offers advantages such as efficient computation, better optimization, and prevention of mode collapse. Replacing LCDC with LV LapR effectively addresses mode collapse issues.

Experiments and Results

The implementation of the project involves using the Artstation-Artistic-face-HQ Dataset (AAHQ) dataset for target domain to showcase the accuracy of artistic style and mask entity transfer. Figure 3 illustrates the influence of various loss functions utilized in one-shot domain adaptation, presenting each step of the process. Moreover, Figure 4 exhibits a range of domain adaptation results, all of which demonstrate the successful transfer of artistic style while preserving masked entities effectively throughout the procedure.

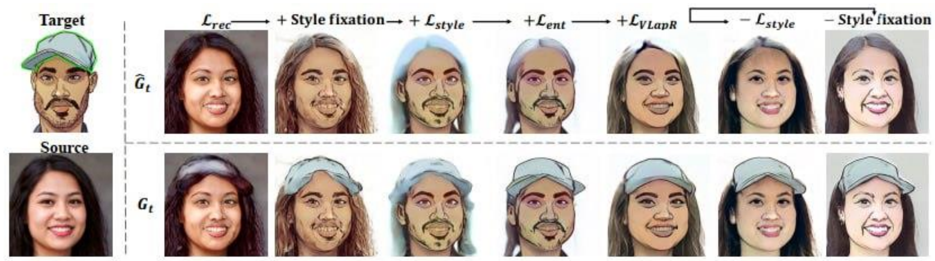

Fig. 3. The arrows show the differences between the current model and the prior model.

Fig. 4 displays the qualitative results of our approach. The top row showcases the original natural face images, which we subsequently invert using the e4e technique [11] to obtain their corresponding latent codes. Through careful examination of the outcomes, the following conclusions can be drawn: Firstly, our synthesis exhibits a style that competes favorably with the provided references, our results prominent high-frequency details, resulting in highly realistic appearances. Secondly, our method achieves remarkable cross-domain correspondence, effectively preserving the content and shape of the source images. This represents a substantial advancement compared to other techniques that lack this capacity, ensuring that the diversity of the source model remains intact without any content collapse.

Conclusion

In this project, a novel one-shot adaptation approach for Generative Adversarial Networks (GANs) is introduced, addressing both style and entity transfer challenges. The adaptation process is divided into global style transfer and generating new entities by utilizing a reference image and binary entity mask. Sliced Wasserstein distance (SWD) is employed to minimize the gap between the internal distributions of the reference image and synthesized outputs. Style fixation and an auxiliary network facilitate style and entity transfer in the generator. A variational Laplacian regularization technique ensures cross-domain correspondence and prevents content distortion during training. Extensive experiments validate the effectiveness of the proposed approach in achieving generalized one-shot adaptation for both style and entity transfer tasks, showcasing its versatility in diverse image manipulation applications. Future research could explore complex entity representations, conditional adaptation, and extension to other GAN architectures, as well as investigating adaptation across significantly different domains and conducting human perception studies to assess visual quality and artistic coherence.

Fig. 4. Qualitative results generated with an entity

Opinion section

<u>Strength</u>: The proposed approach addresses both style and entity transfer, making it more versatile and suitable for diverse scenarios. It ensures that entities present in the source domain are effectively transferred to the target domain alongside the artistic style. Moreover, by employing sliced Wasserstein distance (SWD), the proposed method efficiently measures the divergence between internal distributions for accurate style and entity transfer. Besides, they have proposed a methodology for one shot domain adaptation by providing a mask vector which can accurately change also the location of entity for the future works. Additionally, the introduction of style fixation enhances control over style transfer and improves the overall quality of adapted images.

<u>Weakness</u>: However, the paper lacks detailed discussion of evaluation metrics and may be relatively complex for some readers. Future research could explore more complex and multi entity representations, conditional adaptation, and real-world applications, as well as investigating adaptation across significantly different domains and conducting human perception studies to assess the visual quality and artistic coherence of the adapted images. Exploring the application of the proposed approach to other modalities, such as text-to-image synthesis or image-to-text generation, could further demonstrate its versatility and potential in different domains and could be a possible future work.

Final Report - Generalized one-shot domain adaptation via GANs

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." In International Conference on Learning Representations, 2019.

[2] Yijun Li, Richard Zhang, Jingwan Lu, and E. Shechtman. "Few-shot image generation with elastic weight consolidation." ArXiv, abs/2012.02780, 2020.

[3] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. "Freeze discriminator: A simple baseline for fine-tuning GANs." ArXiv, abs/2002.10964, 2020.

[4] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. "GAN inversion: A survey." arXiv preprint arXiv:2101.05278, 2021.

[5] Jiayu Xiao, Liang Li, Chaofei Wang, Zhengjun Zha, and Qingming Huang. "Few-shot generative model adaptation via relaxed spatial structural alignment." ArXiv, abs/2203.04121, 2022.

[6] Ceyuan Yang, Yujun Shen, Zhiyi Zhang, Yinghao Xu, Jiapeng Zhu, Zhirong Wu, and Bolei Zhou. "One-shot generative domain adaptation." ArXiv, abs/2111.09876, 2021.

[7] Peihao Zhu, Rameen Abdal, John C. Femiani, and Peter Wonka. Mind the gap: Domain gap control for single shot domain adaptation for generative adversarial networks. ArXiv, abs/2110.08398, 2021.

[8] Mikhail Belkin and Partha Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation." Neural Computation, 15:1373–1396, 2003. [12] Geoffrey E. Hinton and Sam T. Roweis. "Stochastic neighbor embedding." In NIPS, 2002.

[9] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. Journal of Mathematical Imaging and Vision, 51:22–45, 2014.

[10] Amit R.Sharma and Prakash.R.Devale Prakash.R.Devale. Face photo-sketch synthesis and recognition. International Journal of Applied Information Systems, 1:46–52, 2012.

[11] Mingcong Liu, Qiang Li, Zekui Qin, Guoxin Zhang, Pengfei Wan, and Wen Zheng. Blendgan: Implicitly gan blending for arbitrary stylized face generation. In Advances in Neural Information Processing Systems, 2021.

[12] Min Jin Chong and David Forsyth. Jojogan: One shot face stylization. ArXiv, abs/2112.11641, 2021.

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

[14] Yinglu Liu, Hailin Shi, Hao Shen, Yue Si, Xiaobo Wang, and Tao Mei. A new dataset and boundary-attention semantic segmentation for face parsing. In AAAI, 2020.

[15] Gihyun Kwon and Jong-Chul Ye. One-shot adaptation of gan in just one clip. ArXiv, abs/2203.09301, 2022.

[16] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4401–4410, 2019.

[17] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: Effective knowledge transfer from gans to target domains with few images. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 9329–9338, 2020.

[18] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data. In European Conference on Computer Vision, 2018.