Invertible Conditional GAN Revisited: Photo-to-Manga Face Translation with Modern Architectures (Student Abstract)

Taro Hatakeyama[1], Ryusuke Saito[1], Komei Hiruta[1], Atsushi Hashimoto[1,2], Satoshi Kurihara[1]

[1]Keio University
[2]OMRON SINIC X Corp.
{taro74, ryusuke.saitocp3, komei, atsushi.hashimoto, satoshi}@keio.jp

Abstract

Recent style translation methods have extended their transferability from texture to geometry. However, performing translation while preserving image content when there is a significant style difference is still an open problem. To overcome this problem, we propose Invertible Conditional Fast GAN (IcFGAN) based on GAN inversion and cFGAN. It allows for unpaired photo-to-manga face translation. Experimental results show that our method could translate styles under significant style gaps, while the state-of-the-art methods could hardly preserve image content.

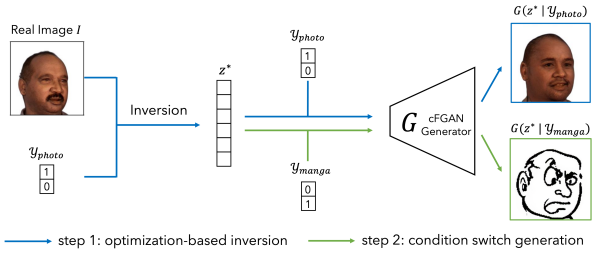

Figure 1: Architecture of IcFGAN.

Introduction

It is challenging to translate styles over significant differences in texture and geometry. Photo-to-manga face translation is such a task. Face photos are colored and include fine elements, while manga faces are generally depicted with black lines on a white background, and their parts are deformed with geometrical exaggerations. Therefore, photo-to-manga face translation requires significant changes to the manga style while preserving face identities. Existing translation methods are limited to photo-realistic manga faces or colored anime faces.

In this paper, we propose a novel style translation approach called Invertible Conditional Fast GAN (IcFGAN) that revisits IcGAN (Perarnau et al. 2016). IcGAN is a very different approach from other recent style translation methods, although it is underexplored until today. We enlighten the method again, and reaffirm that applying modern architectures, IcGAN is effective for translations with significant style changes.

Related Work

MangaGAN(Su et al. 2021) is the first method for unpaired photo-to-manga face translation. The target manga style for MangaGAN is relatively similar in geometry to face photos. Thus, its effectiveness for datasets that require more geometric deformation is an open problem. In addition, it is not possible to train MangaGAN with the landmark-less manga face dataset used in this study.

Layer swapping-based methods, such as UI2I-Style (Kwong, Huang, and Liao 2021), involve a trade-off between geometric deformation and identity preservation. Models that disentangle images into style and content vectors, such as GNR (Chong and Forsyth 2021), aim to perform geometric deformation and identity preservation simultaneously. However, it is difficult to disentangle them completely under significant style gaps.

Our method achieves translations with geometric deformation and identity preservation because it does not require challenging feature disentanglement.

Approach

Invertible Conditional Fast GAN

Fig. 1 illustrates the architecture of our method. IcFGAN learns each image style with shared parameters and explicitly conditions by using condition code represented as one-hot vectors. IcFGAN uses a pre-trained cGAN for translation. We used cFGAN, an extended model of FastGAN(Liu et al. 2020) to cGAN, and the extension to cGAN is straightforward. Condition labels are concatenated with the inputs of FastGAN's generator and discriminator. cFGAN can generate high-quality paired images with a small number of parameters. FastGAN has less than half the parameters of StyleGAN2 (Karras et al. 2020). The use of fewer parameters improves the adaptation of features across conditions and preserves the identities of the images generated under different conditions derived from the same latent vector.

The translation in IcFGAN follows two steps, optimization-based inversion and condition switch generation, as illustrated in Fig. 1. First, an input real image is

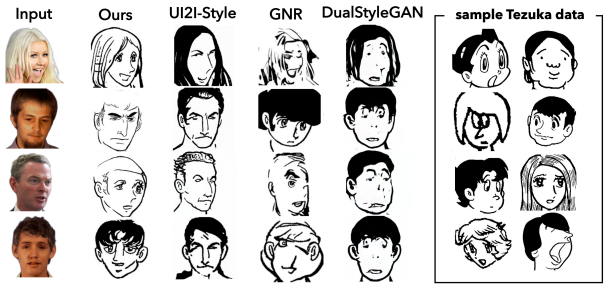

Figure 2: *Left*: results of comparison with state-of-the-art translation methods. *Right*: samples of target manga works.

inverted to the latent vector z^*. During the inversion, the condition code y_{photo} is not included in the optimization and does not change. Next, the inverted latent vector is fed to cFGAN (with the condition code y_{manga}), which results in a manga-style face image.

For GAN inversion, the condition code is represented as one-hot vectors and set to the same conditions as the input images when inverting. Our loss L can be described as

$$L = L_{LPIPS}(G(z^* \mid y_{photo}), I) + \lambda ||G(z^* \mid y_{photo}) - I||_2^2, \quad (1)$$

where G, y and I are the generator of cFGAN, the condition code, and an input image, respectively.

In summary, IcFGAN updated IcGAN in its encoder-based inversion based on the knowledge of recent GAN inversion techniques, and the GAN part to cFGAN, which improves the adaptation of features across conditions.

Dataset and Training

We used the MUCT Face Database (Milborrow, Morkel, and Nicolls 2010) for photos and the Tezuka Dataset (Saito et al. 2021) for manga faces as training datasets. We filled the background of the photos with white and then cropped the face region. We used Adam optimizer to inverse input images and set the learning rate to $\eta = 0.1$, $\beta = (0.5, 0.999)$ and the number of training iterations to 100. We set λ in Eq. 1 to 0.1.

Evaluation

Fig. 2 presents a qualitative comparison with three state-of-the-art methods: UI2I-Style, GNR and DualStyleGAN (Yang et al. 2022). Our method allows for geometric deformation while preserving the face identity. The baselines' hyperparameters were determined with our best effort. We excluded DualStyleGAN from the quantitative evaluation due to the lowest diversity. As a result of the quantitative evaluation, the Fréchet Inception Distance (FID) between the translated images and the Tezuka Dataset is shown in Table 1. The table shows that our method achieved the best FID score.

We further conducted two types of user studies, and 26 students participated in the evaluation. The results are also shown in Table 1. In user study 1, we presented manga faces translated by three different methods from a photo and asked users to select which one was most like the target manga style. In user study 2, we showed a face photo with five manga faces and asked them to find the one translated from the photo. The results show that our method also performed best in these user studies.

Method	FID↓	User Study 1↑	User Study 2↑
GNR	43.04	20.72%	49.05%
UI2I_Style	114.10	28.12%	62.22%
Ours	**33.15**	**51.16%**	**64.60%**

Table 1: Fréchet Inception Distance (FID) between the translated images and Tezuka Dataset and results of user studies.

Conclusion

We proposed an IcGAN-based unpaired photo-to-manga face translation method using the modern architectures of optimization-based GAN inversion and cFGAN. Learning each image style with shared parameters enables translation to a target manga style under significant geometric style gaps without feature disentanglement. The use of cFGAN, which has only a few parameters, also allowed for the preservation of face identities.

Acknowledgments

This research was supported by the NEDO project "Development of Interactive Story-Type Contents Creation Framework."

References

Chong, M. J.; and Forsyth, D. 2021. GANs N' Roses: Stable, Controllable, Diverse Image to Image Translation (Works for Videos Too!). *arXiv:2106.06561 [cs]*.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *CVPR*.

Kwong, S.; Huang, J.; and Liao, J. 2021. Unsupervised Image-to-Image Translation via Pre-trained StyleGAN2 Network. *IEEE Transactions on Multimedia*.

Liu, B.; Zhu, Y.; Song, K.; and Elgammal, A. 2020. Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis. In *ICLR*.

Milborrow, S.; Morkel, J.; and Nicolls, F. 2010. The MUCT Landmarked Face Database. In *PRASA*.

Perarnau, G.; van de Weijer, J.; Raducanu, B.; and Álvarez, J. M. 2016. Invertible Conditional GANs for Image Editing. *NIPS Workshop*.

Saito, R.; Hiruta, K.; Hashimoto, A.; and Kurihara, S. 2021. Conditional GAN for learning small datasets [Translated from Japanese.]. IBIS.

Su, H.; Niu, J.; Liu, X.; Li, Q.; Cui, J.; and Wan, J. 2021. MangaGAN: Unpaired Photo-to-Manga Translation Based on The Methodology of Manga Drawing. *AAAI*.

Yang, S.; Jiang, L.; Liu, Z.; and Loy, C. C. 2022. Pastiche Master: Exemplar-Based High-Resolution Portrait Style Transfer. In *CVPR*.