Paper Proposal: Context aware GAN with Data Augmentation

Jiugeng Sun[1,*], Jiaqing Xie[2,*], Sofie Daniëls[3,*]

[*]All authors contributed equally to this research.

Email: [1]jiusun@student.ethz.ch, [2]jiaxie@student.ethz.ch, [3]sdaniels@student.ethz.ch

Keywords: GAN, Kinship Generation, Image Augmentation, Image Segmentation

1 Introduction

Generative adversarial networks (GANs) have become increasingly popular in recent years, in particular, due to their ability to solve image-based problems, such as image generation [1], image translation [2], and explainable representation learning [3]. For example, GANs can be used to generate realistic faces in high resolution [4–6]. In this project, we focus on kinship face generation, and more specifically, the generation of children faces from their parents.

So far, only a few papers have been written on this topic. Nevertheless, promising results were achieved using GANKin, DNA-net, and ChildGAN [7–9]. After latent space projection of the image, GANKin uses a four-layer fully connected network to determine the child features [7], while DNA-net maps features to genes so as to utilise the genetic knowledge [8]. Similarly, ChildGAN [9] uses genetic knowledge combined with a semantic learning framework.

Most papers perform feature selection in the latent space to account for the imbalance between the limited dataset size and the number of learnable parameters. The most common choices for such image-to-latent and latent-to-image space projection are Image2StyleGAN [10] and StyleGAN [11].

2 Methods

Despite the breakthroughs in the field of image synthesis using deep neural networks such as GANs, the training typically requires a large amount of data [12]. Likewise, dataset size is also a problem for GAN-based kinship generation. Therefore, before we forward the images of the parents to the GAN, we propose to separately apply three image augmentation techniques to our kinship dataset.

The first image augmentation technique, Mixup [13], generates a weighted combination of image pairs from the training data. AugMix, on the other hand, [14] mixes augmented images through linear interpolations. Lastly, SmartAugment [15] trains two networks in tandem such that the first network, Network-A, translates two input images into new images and forward them to train the second network, Network-B, and conversely improves itself by learning the loss of Network-B.

Furthermore, we use a random function for MixUp and AugMix to determine whether to use any of them or maintain the original image. For SmartAugment, We will begin with the most fundamental MLP networks for Network-A. Regarding the performance of the prior model and our time allotment, we reserve the right to decide not to deploy the SmartAugment.

In the second preprocessing step, we use segmentation to facilitate the downstream extraction of the main facial features of the parents. Earlier face segmentation works mainly rely on error-prone preprocessing steps [16–19], but Lin et al. [20] instead propose a scale invariant, rotation and transformation equivariant model that can provide competitive accuracy. We therefore opt to use their pretrained network in our segmentation step.

After preprocessing, we convert the segmented faces of the parents into latent space vectors using [10]. Similar to GANkin [7], we then use four fully connected layers to predict the latent vector of the child, and convert the result back into the image space using StyleGAN [11].

To reverse the predicted segmentation of the child face into a real image at the end of the pipeline, we rely on a pix2pix GAN as proposed in [21]. Early experiments on a CPU with a small-scale image dataset showed promising results and supported the feasibility of training the final 'desegmentation' step with a pix2pix GAN.

The full pipeline can be found in Figure 1.

3 Evaluation

We train and test our model with the TSKinFace Dataset [22]. The TSKinFace dataset is grouped into three main family compositions, Father-Mother-Daughter (FM-D), Father-Mother-Son (FM-S), and Father-Mother-Son-Daughter (FM-SD), which allows straightforward extraction of the parent and corresponding child images. The pix2pix GAN will be trained separately with the iBugMask dataset consisting of over 22 000 image-segmentation pairs [20].

For evaluation, we compare four different scenarios. Our baseline consists of the original GAN implementation with a four-layer network used for child feature prediction. In the first experiment, we apply image augmentation to the parent images, in order to overcome the limited dataset size. In the second experiment, we test the effects of segmentation and parse the parent faces with 11 labels. The resulting child segmentation is converted back into a final realistic image of the child with the pix2pix GAN. Finally, for our last experiment we combine both augmentation and segmentation.

The cosine similarity between the prediction and the real child image will be used as evaluation metric.

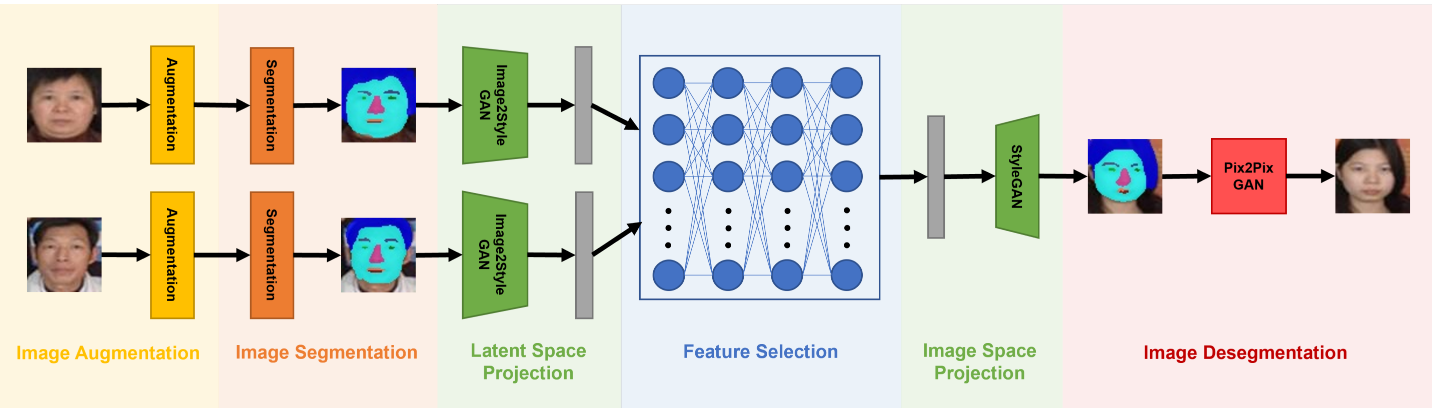

Figure 1: Pipeline of context-aware GAN with data augmentation.

References

[1] Bao, J., Chen, D., Wen, F., Li, H., and Hua, G. "CVAE-GAN: fine-grained image generation through asymmetric training". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2745–2754.

[2] Liu, M.-Y., Breuel, T., and Kautz, J. "Unsupervised image-to-image translation networks". In: *Advances in neural information processing systems* 30 (2017).

[3] Tran, L., Yin, X., and Liu, X. "Disentangled representation learning gan for pose-invariant face recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1415–1424.

[4] Wang, X., Guo, H., Hu, S., Chang, M.-C., and Lyu, S. "Gan-generated faces detection: A survey and new perspectives". In: *arXiv preprint arXiv:2202.07145* (2022).

[5] Kammoun, A., Slama, R., Tabia, H., Ouni, T., and Abid, M. "Generative Adversarial Networks for face generation: A survey". In: *ACM Computing Surveys (CSUR)* (2022).

[6] Wu, X., Xu, K., and Hall, P. "A survey of image synthesis and editing with generative adversarial networks". In: *Tsinghua Science and Technology* 22.6 (2017), pp. 660–674.

[7] Ghatas, F. and Hemayed, E. "GANKIN: generating Kin faces using disentangled GAN". In: *SN Applied Sciences* 2 (Feb. 2020). DOI: `10.1007/s42452-020-1949-3`.

[8] Gao, P., Robinson, J., Zhu, J., Xia, C., Shao, M., and Xia, S. "Dna-net: Age and gender aware kin face synthesizer". In: *2021 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2021, pp. 1–6.

[9] Cui, X., Zhou, W., Hu, Y., Wang, W., and Li, H. "Heredity-aware Child Face Image Generation with Latent Space Disentanglement". In: *arXiv preprint arXiv:2108.11080* (2021).

[10] Abdal, R., Qin, Y., and Wonka, P. "Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?" In: *CoRR* abs/1904.03189 (2019). arXiv: `1904.03189`. URL: `http://arxiv.org/abs/1904.03189`.

[11] Karras, T., Laine, S., and Aila, T. "A style-based generator architecture for generative adversarial networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4401–4410.

[12] Shorten, C. and Khoshgoftaar, T. M. "A survey on Image Data Augmentation for Deep Learning". In: *Journal of Big Data* 6.1 (July 2019), p. 60. ISSN: 2196-1115. DOI: `10.1186/s40537-019-0197-0`. URL: `https://doi.org/10.1186/s40537-019-0197-0` (visited on 11/22/2022).

[13] Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. "mixup: BEYOND EMPIRICAL RISK MINIMIZATION". en. In: (2018), p. 13.

[14] Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., and Lakshminarayanan, B. *AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty*. arXiv:1912.02781 [cs, stat]. Feb. 2020. DOI: `10.48550/arXiv.1912.02781`. URL: `http://arxiv.org/abs/1912.02781` (visited on 11/22/2022).

[15] Lemley, J., Bazrafkan, S., and Corcoran, P. "Smart Augmentation - Learning an Optimal Data Augmentation Strategy". In: *IEEE Access* 5 (2017). arXiv:1703.08383 [cs, stat], pp. 5858–5869. ISSN: 2169-3536. DOI: `10.1109/ACCESS.2017.2696121`. URL: `http://arxiv.org/abs/1703.08383` (visited on 11/22/2022).

[16] Khan, K., Mauro, M., and Leonardi, R. "Multi-class semantic segmentation of faces". In: Aug. 2015. DOI: `10.1109/ICIP.2015.7350915`.

[17] Khan, K., Ahmad, N., Ullah, K., and Din, I. "Multiclass semantic segmentation of faces using CRFs". In: *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCI-*

ENCES 25 (Jan. 2017), pp. 3164–3174. DOI: 10.3906/elk-1607-332.

[18] Zhou, L., Liu, Z., and He, X. "Face Parsing via a Fully-Convolutional Continuous CRF Neural Network". In: (Aug. 2017).

[19] Zhou, Y. "Top-down sampling convolution network for face segmentation". In: Dec. 2017, pp. 1893–1897. DOI: 10.1109/CompComm.2017.8322867.

[20] Lin, Y., Shen, J., Wang, Y., and Pantic, M. "RoI Tanh-polar transformer network for face parsing in the wild". In: *Image and Vision Computing* 112 (2021), p. 104190. ISSN: 0262-8856. DOI: https://doi.org/10.1016/j.imavis.2021.104190. URL: https://www.sciencedirect.com/science/article/pii/S0262885621000950.

[21] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. "Image-to-Image Translation with Conditional Adversarial Networks". In: *CVPR* (2017).

[22] *The Closed Eyes in the Wild (CEW) dataset*. URL: http://parnec.nuaa.edu.cn/_upload/tpl/02/db/731/template731/pages/xtan/TSKinFace.html (visited on 11/30/2022).