# StyleGene: Crossover and Mutation of Region-level Facial Genes for Kinship Face Synthesis

Hao Li[1], Xianxu Hou[2,5], Zepeng Huang[1], Linlin Shen[1,2,3,4*]
[1]Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University
[2]National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University
[3]Shenzhen Institute of Artificial Intelligence and Robotics for Society
[4]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University
[5]School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University
{haoli2021, huangzepeng2021}@email.szu.edu.cn, hxianxu@gmail.com, llshen@szu.edu.cn

## A. StyleGene Implementation Details

### A.1. Fine-grained Face Segmentation

In this work, we follow the pipeline proposed by DatasetGAN [9] to obtain the fine-grained facial region annotations. Since original implementation was based on the StyleGAN [3], we first re-implement it based on the StyleGAN2 [4]. Fig. s1 shows an example of a segmentation mask used to train DatasetGAN, where 34 annotated regions are used. To build our training set, we first use an image encoder [7] to embed the real image into the $\mathcal{W}^+$ space of StyleGAN2. Then we use the StyleGAN2 generator to reconstruct the input image and adopt our DatasetGAN to obtain the corresponding facial masks.

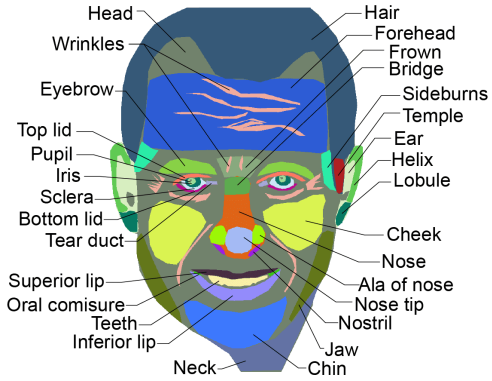


Figure s1. An example of facial region annotations [9].

### A.2. Building a Gene Pool

In this work, we introduce the gene pool into the mutation process to increase the diversity of generated descendants. Fig. s2 shows a schematic diagram of our gene pool.

*Corresponding Author

We employ FairFace [2] to obtain the age, gender and race of faces in FFHQ dataset, and then group the Region-level Facial Genes (RFGs) with the obtained labels. The gene pool divides age into 9 groups (0-2, 3-4, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, and over 70 years old) and contains 7 races (White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino) and 2 genders (Male and Female). During the mutation process, we sample RFGs from the gene pool for each facial region according to the target age, gender and race.

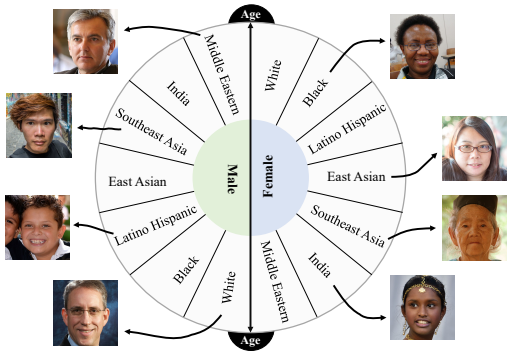


Figure s2. The schematic diagram of our proposed Gene Pool. We group face images by age, gender, and race, and extract their region-level facial genes to build the gene pool.

## B. Sensitivity Analysis

In this section, we analyze three hyper-parameters used in our model. The first two are $\eta$ and $\gamma$ in Eq.10 in the main text to control the degree of gene mutation. The other is $l$ in Eq.11 in the main text to decide how many layers are used to generate the latent code.

### B.1. Effect of $\eta$ and $\gamma$

Recall that we design a gene mutation process to increase the gene diversity. $\eta$ and $\gamma$ control the degree of genetic variation of the descendants. Thus, when fewer gene mutations are performed, the resulting descendants will be more similar to their parents. We expect to find the most appropriate $\eta$ and $\gamma$ to balance the kinship verification accuracy (ACC) and LPIPS. In particular, we calculate the ACC and LPIPS metrics using different $\eta$ $(0, 10\%, 20\%, 40\%, 50\%, 60\%)$ and $\gamma$ $(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$, respectively. We further normalize the resultant values to obtain Relative ACC (RACC) and Relative LPIPS (RLPIPS).

Inspired by Average precision (AP) [1], we propose Average Relative LPIPS (ARL) to first find the optimal $\eta$. ARL is defined as the mean RLPIPS at a set of 11 equally spaced RACC levels $[0, 0.1, \ldots, 1]$:

$$ARL = \frac{1}{11} \sum_{a \in \{0, 0.1, \ldots, 1\}} l_{\text{interp}}(a) \tag{1}$$

The RLPIPS at each RACC level $a$ is interpolated by taking the maximum RLPIPS measured for a method for which the corresponding RACC exceeds $a$:

$$l_{\text{interp}}(a) = \max_{\tilde{a}: \tilde{a} \geq a} l(\tilde{a}) \tag{2}$$

where $l(\tilde{a})$ is the measured RLPIPS at RACC $\tilde{a}$.

Tab. s1 shows the corresponding ARL for different values of $\eta$. The maximal value is obtained when $\eta = 40\%$.

Table s1. ARL for different values of $\eta$.

| $\eta\%$ | 0 | 10 | 20 | 30 | 40 | 50 | 60 |
|---|---|---|---|---|---|---|---|
| ARL | 0.39 | 0.45 | 0.45 | 0.45 | 0.47 | 0.37 | 0.39 |

Fig. s3 shows the RLPIPS and RACC for different $\gamma$ values when $\eta$ is equal to $40\%$. We can see that similar high accuracy can be achieved when using small $\gamma$ ($< 0.47$), and the accuracy starts to drop dramatically when $\gamma$ is bigger than $0.47$. In our experiments, we choose $\gamma = 0.47$ to increase the gene diversity of the descendants while maintaining high kinship similarity with their parents.

### B.2. Effect of $l$

As shown in previous work [6], high-level semantics such as face identity are controlled by lower layers (1-8) while color scheme and microstructure are decided by higher layers. Thus, the choice of $l$, which denotes the number of layers when generating the latent code of descendants, will have an effect on the diversity of generated descendants. To this end, we report the LPIPS values between synthetic descendants by choosing different $l$ on

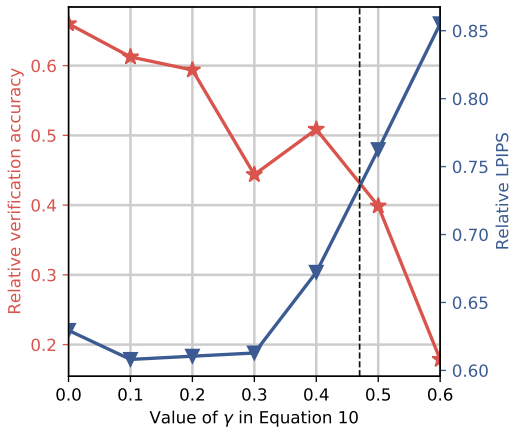


Figure s3. Sensitivity analysis of genetic variation intensity $\gamma$.

FF-Database [10]. As shown in Fig. s4, the LPIPS value increases with larger $l$ values, then the progress stops and LPIPS remains similar values after more than 8 layers are used. Therefore, we focus on controlling the first 8 layers for Kinship face synthesis.

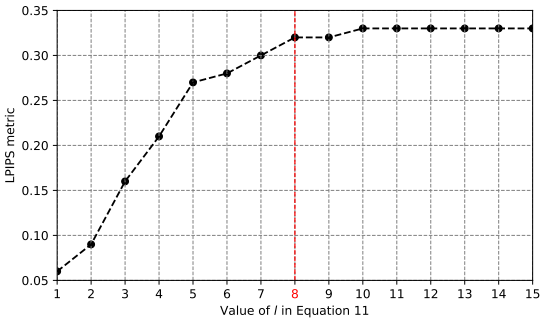


Figure s4. The effect of $l$ for mixing of parents' styles.

## C. Qualitative results

### C.1. Comparison with the state-of-the-art methods

Fig. s5 shows qualitative results on TSKinFace dataset. We compare our method with StyleDNA [5], ChildPredictor [10], and CDFS [8]. We use different methods for each pair of parents to produce a son and a daughter. As can be seen, the descendants generated by our method are more realistic than other methods.

### C.2. Age and gender control

In order to control the age and gender of the descendant, we use a gene pool constructed by grouping faces based on age, gender and race. As described in Section A.2, the target age and gender of descendants are used to identify the faces with similar attributes in gene pool, which are then selected and involved in the crossover and mutation process to gener-

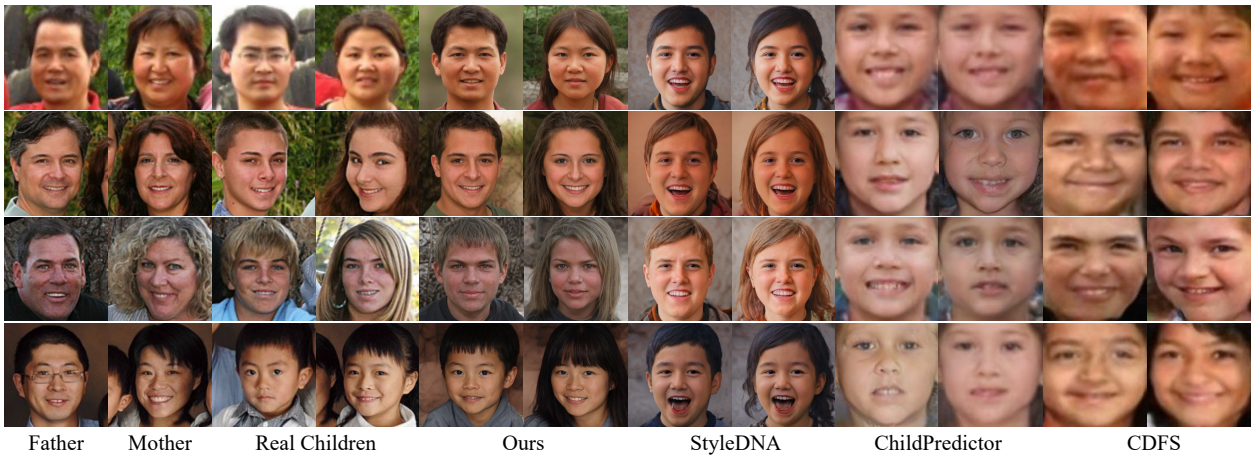


Figure s5. Comparison of children faces synthesized by our StyleGene and other methods. The two leftmost columns show the images of father and mother. We also show the real children and the faces synthesized by ours, StyleDNA, ChildPredictor and CDFS, respectively

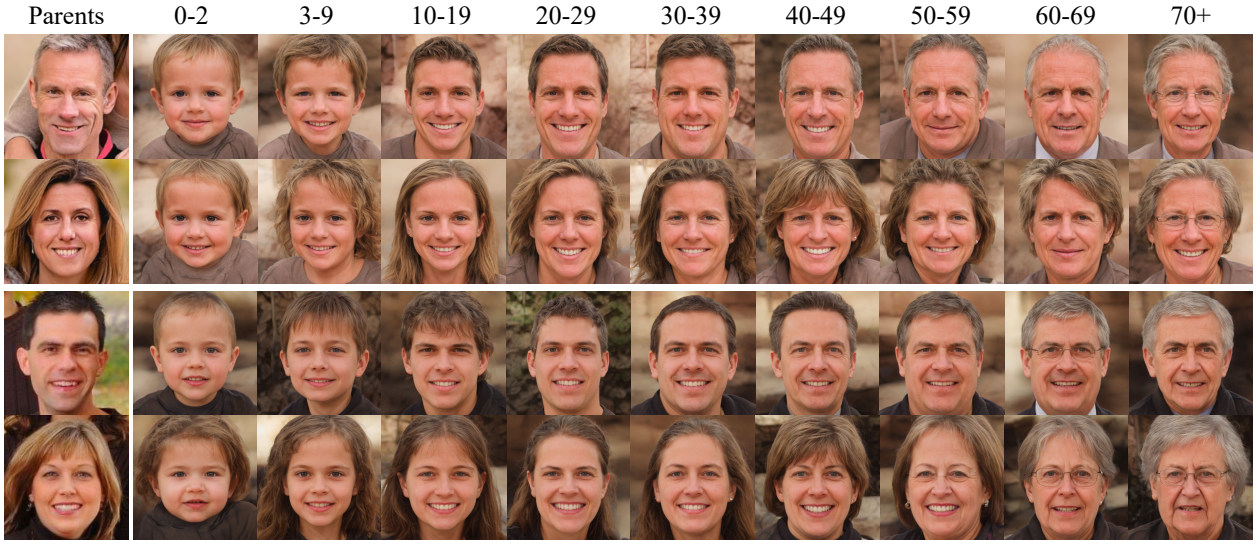


Figure s6. Results of age and gender control. The leftmost column represents the parents and the rest columns show the results with different age groups. The first and third rows are the generated sons, and the second and fourth rows are the generated daughters.

ate the RFGs of the descendants and faces with desired age and gender. Fig. s6 shows the generated descendants with different ages and genders. It can be seen that our method can accurately control the age and gender of the generated descendants and maintain a high degree of diversity.

## D. Limitations

We identify two limitations of our technique. First, our method needs to inversely project real images into the StyleGAN latent space to extract region-level facial genes (RFGs). Thus, this could be challenging for our method when the GAN inversion method cannot faithfully perform the projection. Second, we recognize that our current approach cannot perform well on face images with occlusion, as it is difficult to extract reliable RFGs from occluded regions.

## E. Societal Impact

This paper focuses on kinship face synthesis using Generative Adversarial Networks. Despite utilizing exclusively public datasets for research and adhering to their licenses, the potential for misuse of our method, particularly for deep fake generation, warrants concern. From the perspective of academia, these risks may be mitigated by advancing research on deep fake detection.

## References


[1] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 2

[2] Kimmo Karkkainen and Jungseock Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1548–1558, 2021. 1

[3] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 1

[4] 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. 1

[5] Che-Hsien Lin, Hung-Chun Chen, Li-Chen Cheng, Shu-Chuan Hsu, Jun-Cheng Chen, and Chih-Yu Wang. Styledna: A high-fidelity age and gender aware kinship face synthesizer. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, 2021. 2

[6] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021. 2

[7] Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. Designing an encoder for stylegan image manipulation. *ACM Transactions on Graphics (TOG)*, 40(4):1–14, 2021. 1

[8] Yong Zhang, Le Li, Zhilei Liu, Baoyuan Wu, Yanbo Fan, and Zhifeng Li. Controllable descendant face synthesis. *arXiv preprint arXiv:2002.11376*, 2020. 2

[9] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. Datasetgan: Efficient labeled data factory with minimal human effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10145–10155, 2021. 1

[10] Yuzhi Zhao, Lai-Man Po, Xuehui Wang, Qiong Yan, Wei Shen, Yujia Zhang, Wei Liu, Chun-Kit Wong, Chiu-Sing Pang, Weifeng Ou, et al. Childpredictor: A child face prediction framework with disentangled learning. *IEEE Transactions on Multimedia*, 2022. 2