

De Novo Reconstruction of 3D Human Facial Images from DNA Sequence

Mingqi Jiao, Jiarui Li, Bingxu Zhong, Siyuan Du, Shuning Li, Manfei Zhang, Qibin Zhang, Zhongming Liang, Fan Liu,* Chunman Zuo,* Sijia Wang,* and Luonan Chen*

Facial morphology is a distinctive biometric marker, offering invaluable insights into personal identity, especially in forensic science. In the context of high-throughput sequencing, the reconstruction of 3D human facial images from DNA is becoming a revolutionary approach for identifying individuals based on unknown biological specimens. Inspired by artificial intelligence techniques in text-to-image synthesis, it proposes Difface, a multi-modality model designed to reconstruct 3D facial images only from DNA. Specifically, Difface first utilizes a transformer and a spiral convolution network to map high-dimensional Single Nucleotide Polymorphisms and 3D facial images to the same low-dimensional features, respectively, while establishing the association between both modalities in the latent features in a contrastive manner; and then incorporates a diffusion model to reconstruct facial structures from the characteristics of SNPs. Applying Difface to the Han Chinese database with 9,674 paired SNP phenotypes and 3D facial images demonstrates excellent performance in DNA-to-3D image alignment and reconstruction and characterizes the individual genomics. Also, including phenotype information in Difface further improves the quality of 3D reconstruction, i.e. Difface can generate 3D facial images of individuals solely from their DNA data, projecting their appearance at various future ages. This work represents pioneer research in de novo generating human facial images from individual genomics information.

personal identification in forensic applications. DNA profiling, a fundamental strategy in forensic science, which identifies individuals by comparing unknown biological samples (profiled DNA) with known biological specimens,^[1–3] playing a critical role in identification. It is the gold standard in forensic investigations, providing definitive results where traditional methods may falter. As a complementary method, DNA phenotyping analyzes physical traits such as hair and eye color, aiding targeted forensic investigations and streamlining the process. A frontier lies in extracting facial imagery from genetic sequences, poised to transform identification strategies. However, the human face, composed of eyes, nose, chin, and mouth, is shaped by both genetic and environmental factors. Especially in genetically homogeneous Chinese populations, the shape and pigmentation of the face are less variable than in immigrant populations such as those in the United States. Therefore, reconstructing facial images only from genetic data are challenging^[4]: 1) limited knowledge of facial genetics; 2) technological difficulties due to high-dimensional data and small

1. Introduction

Facial morphology represents a unique and genetically inherited biometric characteristic, providing critical insights for

sample size. Currently, many studies have been developed to study facial genetics.^[5–16] Huang et al. introduce a genome-wide association study of facial morphology that identifies novel genetic loci in Han Chinese.^[17] Xiong et al. introduce C-GWAS,

M. Jiao, B. Zhong, S. Li, Q. Zhang, Z. Liang, L. Chen
Key Laboratory of Systems Health Science of Zhejiang Province
Hangzhou Institute for Advanced Study
University of Chinese Academy of Sciences
Chinese Academy of Sciences
Hangzhou 310024, China
E-mail: lnchen@sjtu.edu.cn

J. Li, S. Du, S. Wang
CAS Key Laboratory of Computational Biology
Shanghai Institute of Nutrition and Health
University of Chinese Academy of Sciences
Chinese Academy of Sciences
Shanghai 200031, China
E-mail: wangsijia@picb.ac.cn

S. Du, M. Zhang
Bio-X Institutes
Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders
Ministry of Education
Shanghai Jiao Tong University
Shanghai 200030, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202414507>

© 2025 The Author(s). Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202414507

refining genome-wide association studies to identify new genetic loci for facial traits.^[18] Zhang et al.'s research^[19] conducts genome-wide analysis to explore facial morphology variations between East Asian and European populations, enhancing our understanding of genetic influences across ancestries. Overall, our understanding of facial genetics remains incomplete. On the other hand, some studies have tried to construct the association between genetics data and facial phenotypes. Specifically, Gurovich et al.^[20] present a framework for facial image analysis by identifying genetic disorders. Claes et al.^[21] uncover relations between facial variation and the effects of sex, genomic ancestry using bootstrapped response-based imputation modeling (BRIM). Sero et al.^[22] and Mahdi et al.^[23] proposed models to match facial images with unidentified demographic and genomic attributes of identified individuals. Facial image classification is also an active area of research in machine learning (i.e., 3D face based prediction of sex,^[24] age,^[25] ancestry,^[26] and sexual orientation^[27]). Although these methods have revealed interesting findings, they cannot generate facial images from genetic data.

Inspired by the advanced techniques in text-to-image alignment and synthesis,^[28–30] we proposed Difface which explores both generative diffusion process and contrastive multi-modal alignment, to enable direct to de novo 3D facial image reconstruction from SNPs. Applying Difface to our Han Chinese database with the paired SNP phenotypes and 3D facial images of 9674 individuals demonstrates excellent performance in DNA-to-3D image alignment and reconstruction and characterizes the individual of genomics. Further augmenting the model's performance, integrating auxiliary phenotype data—encompassing age, gender, and Body Mass Index (BMI) into Difface significantly enhanced the precision of the 3D facial reconstructions. In addition, we also identified new SNPs features significantly associated with the related facial regions. This multifaceted work not

only validates the robustness of Difface but also underscores its potential as a versatile tool in personalized medicine and forensic identification.

However, this groundbreaking technology challenges current conceptions of genomic privacy, raising significant ethical and legal implications. By associating de-identified genomic data with phenotypic measurements, our research aims to facilitate a deeper discussion on the impacts of DNA phenotyping. We invite commentary and deliberation on the implications of these findings for genomics research, investigatory practices, and broader societal ethical considerations. Although some scholars have addressed the implications of DNA phenotyping, this work emphasizes the necessity of thorough analysis to navigate the complex interplay of innovation and privacy, ultimately contributing to more informed and ethical genomic research.

2. Results

2.1. Overview of Difface

We proposed Difface, a multi-modality model designed to reconstruct 3D facial images from SNP phenotypes. Specifically, Difface employs a transformer^[31] and spiral convolution network^[32] to map high-dimensional Single Nucleotide Polymorphisms (SNPs) and 3D point clouds onto a unified low-dimensional feature space. At the same time, a spiral convolutional network was trained to generate facial images from the low-dimensional feature space. During the training process a contrastive learning was used to construct the association between genetic and facial images in the low-dimensional features, while during constructing process, Difface uses a diffusion network to generate 3D human facial images only from genetic features on each individual (Figure 1).

2.2. Difface Enables the Alignment of DNA and 3D Facial Images

To evaluate whether Difface can accurately construct the association between DNA and 3D facial images, we employed Difface to analyze large-scale 3D facial surface scans of 9674 samples collected from three independent cohorts in China (Methods). Using the conventional genome-wide significance threshold ($P = 5 \times 10^{-8}$), we identified 7842 genome-wide significant variants associated with 3D facial images. These variants were treated as input SNPs, constituting the input SNP vector of Difface.

Difface differs from Mahdi's^[23] and Sero's^[22] approaches by employing a unique implementation of contrastive learning. Unlike Mahdi's reliance on traditional regression models and Sero's use of basic machine learning without a focus on feature space alignment, Difface focuses on aligning genetic and phenotypic data in a low-dimensional space, efficiently bridging complex genetic information with visible facial features. Difface enhances the specificity and sensitivity of feature extraction, providing a deeper understanding of the relationship between SNPs and facial features, and broadening the model's applicability to different datasets.

To comprehensively compare Difface with recently published methods (Mahdi's and Sero's), we trained Difface using the train-

F. Liu
Naif Arab University for Security Sciences
Riyadh 11433, Saudi Arabia
E-mail: flu@nauss.edu.sa
C. Zuo
School of Life Sciences
Sun Yat-sen University
Guangzhou 510275, China
E-mail: zuochm@mail.sysu.edu.cn
S. Wang
Taizhou Institute of Health Sciences
Fudan University
Taizhou, Jiangsu 200433, China
S. Wang
Center for Excellence in Animal Evolution and Genetics
Chinese Academy of Sciences
Kunming 650223, China
L. Chen
Key Laboratory of Systems Biology
Shanghai Institute of Biochemistry and Cell Biology
Center for Excellence in Molecular Cell Science
Chinese Academy of Sciences
Shanghai 200031, China
L. Chen
School of Mathematical Sciences and School of AI
Shanghai Jiao Tong University
Shanghai 200240, China

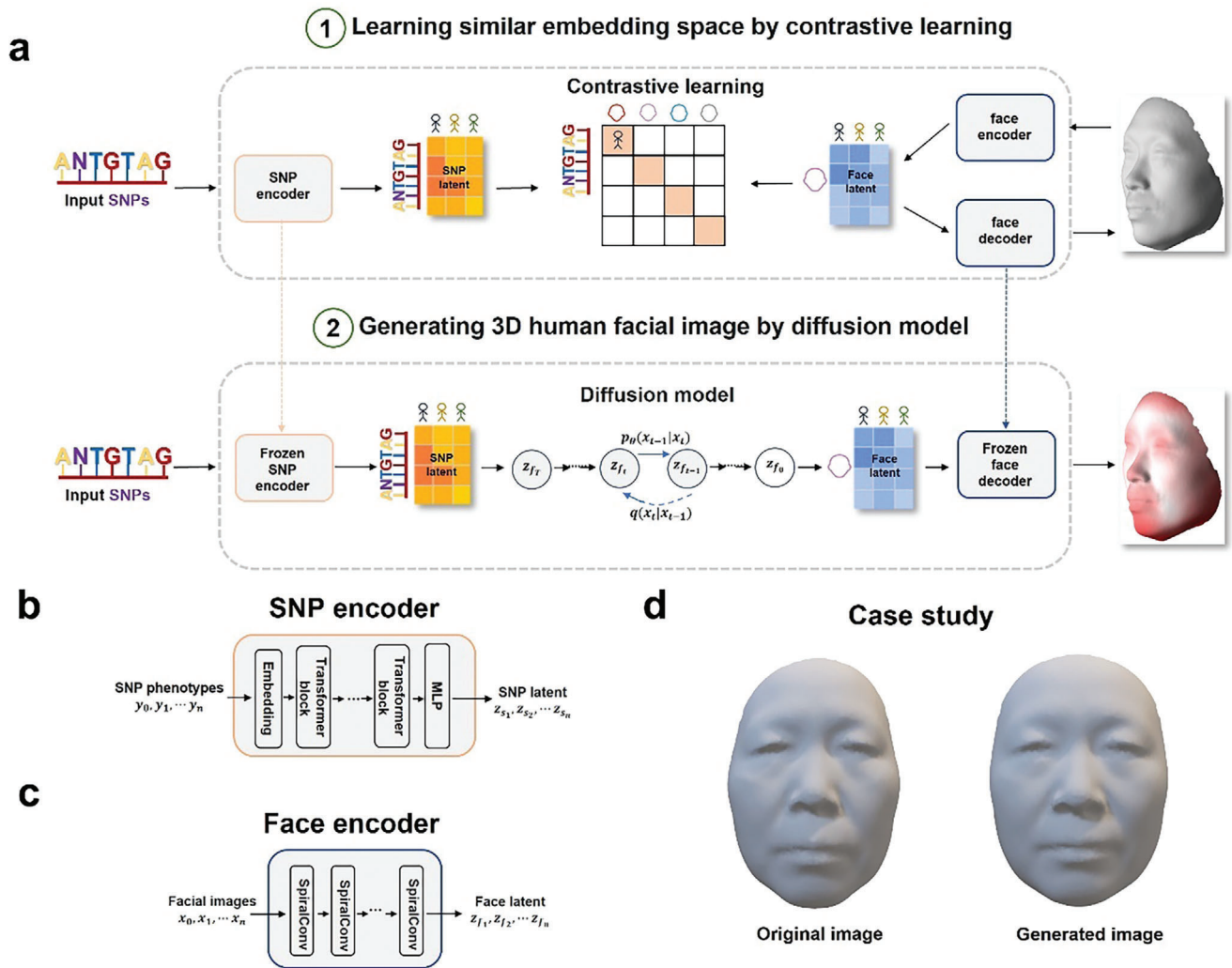


Figure 1. The overall framework of Difface. a) depicts the operation of Difface on a dataset comprising N SNP sets and their corresponding facial data. The initial phase of the model's training involves the merging of SNP-face pairs into a unified representation space, laying the foundation for the generation phase. This entails the construction of a multimodal encoding space through the training of SNP and face encoders. The encoders enhance the cosine similarity between the correct SNP and face pair embeddings within each batch, while simultaneously reducing the similarity between the $N \times N$ incorrect pairs. Concurrently, a decoder is trained to generate 3D faces from face embeddings. Subsequently, an SNP embedding is introduced into a diffusion model, resulting in the generation of a face embedding. This face embedding is then utilized by the decoder to create the final image. It is noteworthy that the encoders and decoder remain fixed throughout the training of the diffusion model. b) Architecture of the SNP encoder. c) Architecture of the face encoder. d) A case study demonstrating the original and generated images.

ing dataset and then used two trained encoders to learn the low-dimensional features of SNPs and facial images. Cosine similarity was employed to assess the alignment between facial images and SNP data. To validate the efficacy of our contrastive learning model, we utilized PCA^[33] and CCA^[34] to downscale SNP phenotypes and facial images. The comparison results demonstrated that Difface exhibited the highest alignment score among competing methods, indicating that contrastive learning is an efficient approach.

Specifically, in terms of identification, Difface achieved a rank-1 identification rate (R1) of 3.33%, a rank-10 identification rate (R10) of 23.33%, and a rank-20 identification rate (R20) of 42.53%, significantly outperforming other methods such as Mahdi's (R1: 2.48%, R10: 21.90%, R20: 38.49%) and

Sero's (R1: 3.00%, R10: 23.03%, R20: 39.00%). In verification tasks, Difface also demonstrated superior performance with an equal error rate (EER) of 27.6% and an area under the curve (AUC) of 80.7%, compared to Mahdi's (EER: 35.7%, AUC: 72.1%) and Sero's (EER: 36.8%, AUC: 70.6%) (Figure 2a,b) Table 1.

These results indicate that contrastive learning enables Difface to not only capture complex genetic features but also differentiate between correct SNP-face pairs and incorrect ones by pulling the correct pairs closer together in the feature space. This allows for more precise facial reconstructions compared to retrieval-based or traditional methods. Difface, by utilizing advanced contrastive learning techniques, significantly improves the alignment and reconstruction of 3D facial images from DNA data, providing a

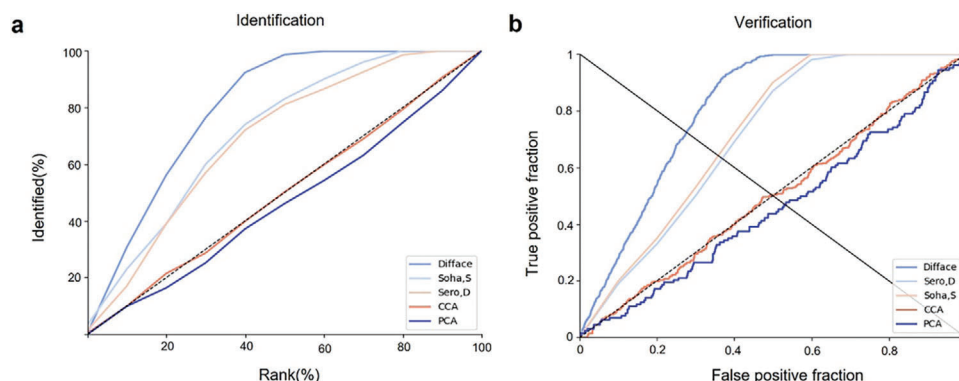


Figure 2. Performance comparison of Difface with other methods for identification and verification tasks. a) Identification performance: The y-axis represents the percentage of correctly identified samples, while the x-axis represents the rank percentage. Difface shows superior performance compared to Soha.S, Sero.D, CCA, and PCA. b) Verification performance: The y-axis represents the true positive fraction, and the x-axis represents the false positive fraction. The ROC curves indicate that Difface outperforms other methods in terms of verification accuracy.

Table 1. Average identification and verification results for different methods.

Method	EER	AUC	R1	R10	R20
Difface	27.6	80.7	3.33	23.33	42.53
Sero,D	35.7	72.1	2.48	21.90	38.49
Soha, S*	36.8	70.6	3.00	23.03	39.00
PCA	49.4	49.9	0	6.00	11.00
CCA	53.2	45.7	1.50	10.00	18.50

EER verification equal error rate, AUC verification area under the curve, R1 rank 1% identification rate, R10 rank 10% identification rate, R20 rank 20% identification rate.

more accurate and robust tool for forensic identification and personalized medicine.

2.3. Difface Enables the Generation of DNA and 3D Facial Images

To further assess the ability of Difface for reconstructing 3D facial images from single nucleotide polymorphisms (SNPs), we leveraged the Euclidean metric to evaluate performance by calculating the distance between the real and generated 3D point clouds. Our findings indicate that Difface, when leveraging only SNPs, achieved a mean error of 3.52 mm, which establishes a solid baseline for facial reconstruction fidelity. This result demonstrates that Difface can effectively reconstruct facial images from SNPs (Figure 3a). We similarly evaluated the results using chamfer distance and RMSE (Figure S1, Supporting Information).

A detailed analysis of error metrics at critical facial landmarks showed that Difface consistently achieved significantly lower error distances, confirming its exceptional precision in reconstructing key facial features. The heatmap in Figure 3d illustrates that with the diffusion model, errors at specific facial points such as the glabella, nasion, and left/right alare were notably reduced, with errors mostly staying below 4.0 mm, whereas without the diffusion model, some points showed errors exceeding 4.5 mm.

In our evaluation of Difface, we assessed the contribution of the diffusion network to the model's performance. The diffusion technique has gained recognition for its effectiveness in gener-

ative tasks across various fields. The results, depicted in Figure 3c, show a significant enhancement in feature generation within the low-dimensional space attributed to the diffusion network. Specifically, the mean Euclidean distance error with the diffusion network was 3.517 mm, compared to 4.125 mm without it, highlighting the crucial role of the diffusion network in achieving accurate feature generation.

Inspired by the flexible framework of Difface, we additionally incorporated SNP embeddings with more biological and demographic variables such as sex, age, and BMI. The comparison results showed that the addition of these variables systematically improved the quality of the reconstructed facial images. For instance, when combining SNPs with sex data, the mean error decreased to 3.06 mm, and further inclusion of age and BMI data reduced the error to 2.93 mm. Strikingly, when Difface operated without SNP data, relying solely on age, BMI, and sex, the mean error surged to 7.03 mm, reinforcing the critical role of genetic information in precise facial reconstruction. We also explored the impact of age as an additional variable. The quantitative analysis of these transitions revealed a change magnitude of 1.430 from youth to middle age, 1.689 from middle age to old age, and a total transformation of 2.415 from youth to old age. These values reflect the extent of facial structural change, such as the drooping of facial features and other age-related changes, and are visually demonstrated in Figure S6 (Supporting Information). The progression from youth to old age is shown through the corresponding 3D reconstructions, illustrating how Difface captures age-related morphological transformations.

These quantitative results not only validate the robustness of Difface in aligning and reconstructing 3D facial images from DNA data but also underscore its potential as a versatile tool in personalized medicine and forensic identification.

2.4. Difface Facilitates Individual Reconstruction of 3D Facial Images by Diffusion Process

We further demonstrated that Difface is able to individually reconstruct facial images with the diffusion process. In particular, we leveraged the latent features of the real and generated facial

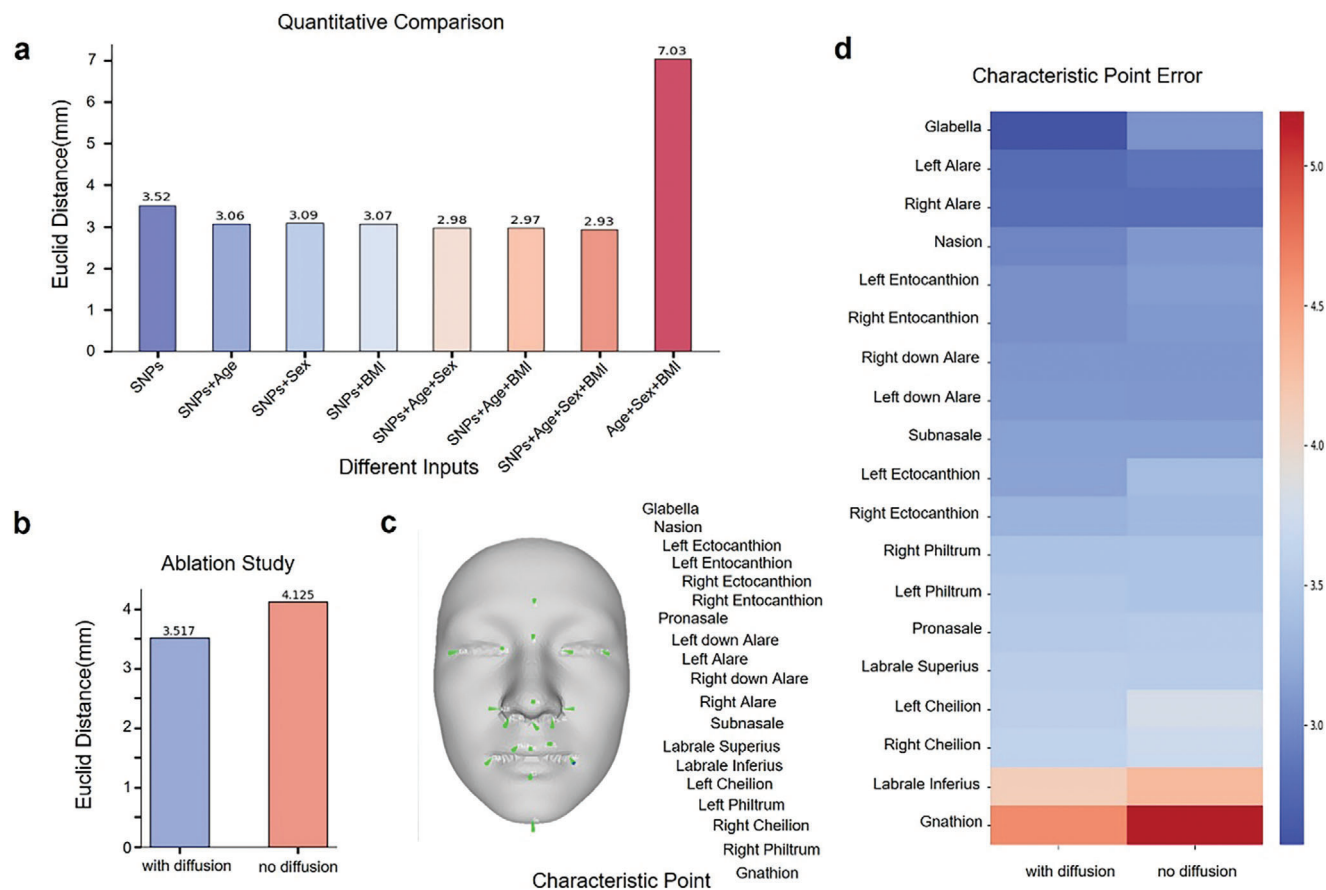


Figure 3. Euclidean Distance Comparison and Error Analysis for Facial Reconstruction Models. a) Quantitative comparison of different inputs on Euclidean distance (mm) between predicted and true facial landmarks. Inputs include SNPs, age, sex, BMI, and their combinations. b) Ablation study showing the impact of the diffusion model on Euclidean distance (mm). c) Visualization of characteristic points on the face used for error measurement. d) Heatmap showing the characteristic point error for models with and without diffusion. The color intensity represents the magnitude of the error, with red indicating higher error values.

images by the trained encoder to predict the facial phenotypes. For each phenotype characteristics, we first trained each classifier by the training dataset, and then calculated the classification accuracy for the test dataset. Considering the pronounced hereditary influence on nasal morphology, our study stratified facial features into two principal categories: those associated with the nose and those about other facial regions. We conducted distinct classification tests for various features within these two categories.

The classification results, showcased in **Figure 4a**, reveal that the correct classification rate for reconstructed faces is impressively close to that of real faces across multiple feature classifications. This result highlights the model's capability to precisely reproduce a broad range of facial features, demonstrating the strength and accuracy of Difface in capturing complex facial characteristics. We also present a direct comparison between original and generated 3D facial reconstructions, focusing on features such as nasion depression and nose wing protrusion. The aim is to demonstrate the model's capacity to mimic critical facial characteristics effectively. (Fig 4b). Similarly, We also compared the generated 3D faces with the original 3D faces and 2D facial photographs (Figure S2, Supporting Information).

In an innovative experimental setup designed to evaluate the efficacy of the Difface model in generating lifelike facial images, we conducted a blinded test involving 10 volunteers. Each participant was tasked with identifying the real face corresponding to a synthetically generated face from a lineup of real faces. To increase the challenge, we provided lineups of 5, 10, and 20 faces, which included the correct match alongside random real faces (**Table 2**).

The experiment was structured to ascertain the likelihood of the participants correctly identifying the real face corresponding to the synthetic version at different lineup sizes. The results revealed a clear trend: as the number of choices increased, the accuracy of identification decreased, underscoring the challenges of facial recognition in larger lineups. The specific findings are summarized in the table below:

Furthermore, we leveraged the Determinantal Point Process (DPP) (a sophisticated method traditionally applied in subset selection problems with diversity constraints^[35]) to quantify the diversity in the facial features generated by Difface. We calculated the DPP score for actual facial datasets and compared it with the DPP score of the average diversity derived from multiple random samplings of facial features. This allowed us to establish

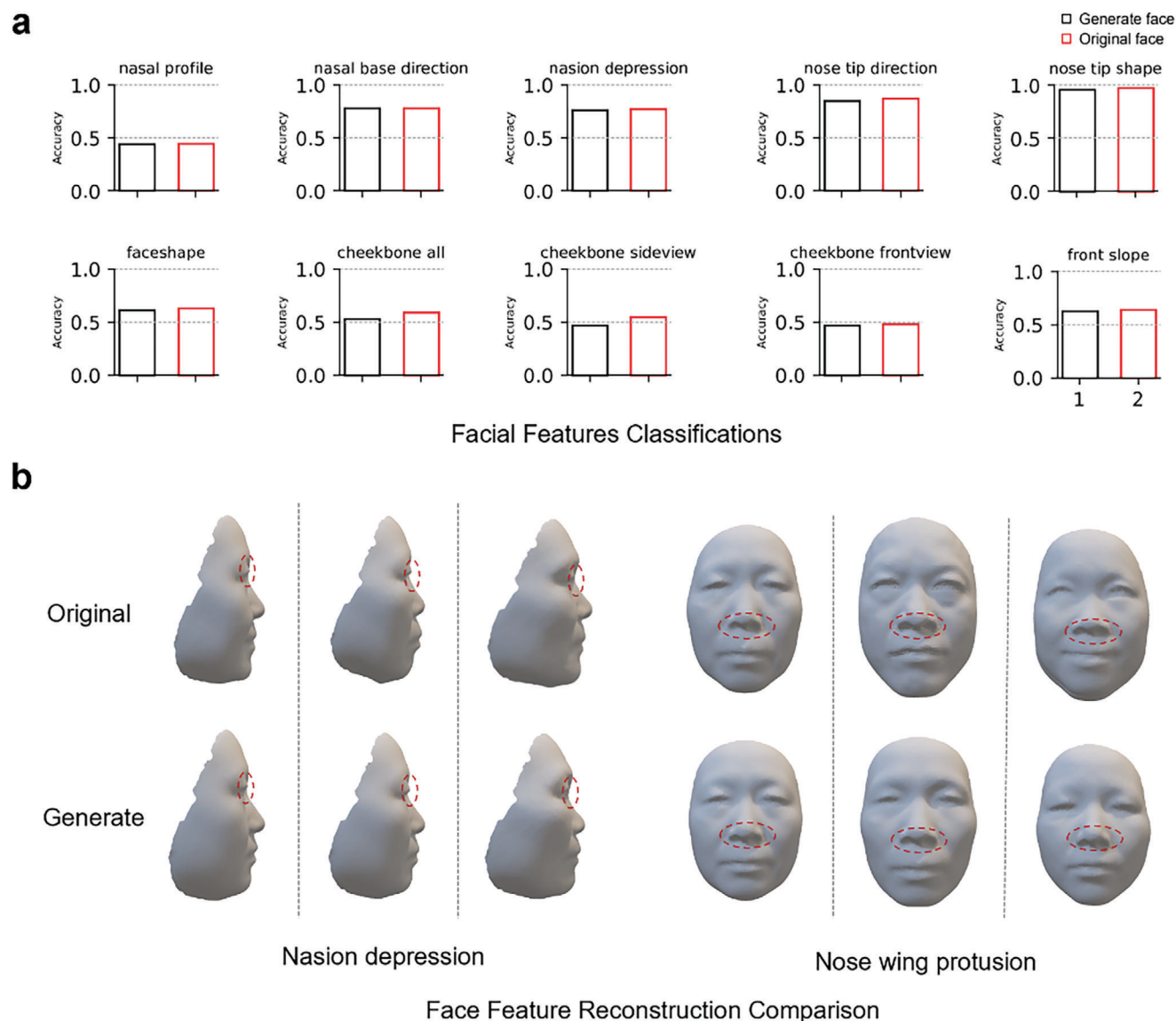


Figure 4. Comparison of Facial Feature Classification and Reconstruction Accuracy Using Difface and Ground Truth Data. a) Face feature classification results, Difface (SNPs), is the outcome of the classification task after utilizing Difface to produce faces from SNPs. Ground truth is the result of the original face for the classification task. b) Comparing the Difface restored face image with the real face image, the top is the real face image, and the bottom is the Difface restored face image.

two robust baselines for the analysis. Additionally, we computed the DPP diversity for the facial images generated by Difface.

The results notably indicate that the DPP diversity score of the facial features generated by Difface is within a 5% margin of the diversity score calculated from the real facial data set. This proximity in diversity scores suggests a high level of variance in the features generated by Difface, closely mirroring the natural variability observed in human facial features. Furthermore, a threshold for desirable diversity at a DPP score that exceeds 95% of the real data's diversity score, which Difface consistently meets or surpasses. Consequently, it can be concluded that Difface successfully achieves a desirable level of diversity in its generated outcomes, effectively capturing the complex diversity in human facial features (Table 3).

Table 2. Accuracy of Facial Recognition Across Different Lineup Sizes.

Lineup Size	Correct Identification Rate [%]
5 Choices	75.6
10 Choices	53.3
20 Choices	51.1

2.5. Generative Effect Test for SNP Deletions

In practical applications, particularly in the field of forensics, it is often challenging to obtain a comprehensive set of single nucleotide polymorphisms (SNPs) from DNA samples. To as-

Table 3. Analysis Results of Data Diversity.

Data	DPP-Diversity
Difface (SNPs)	0.9960
Real images	0.9999
Mean value	0.6012

sess the generative efficacy of Difface under conditions of SNP scarcity, evaluations were conducted with inputs ranging from 90% to 10% of the total SNP phenotypes. This approach was designed to simulate the constraints commonly encountered in forensic analyses.

The efficacy of facial reconstruction was evaluated across a spectrum of SNP completeness levels, ranging from 10% to 90%. Recent studies^[19] have demonstrated that the nose exhibits the strongest genetic associations among facial features, with 107 genetic variants identified surpassing other regions of the face. The distinctive nose shape commonly observed in East Asians is largely attributed to local adaptations to environmental factors and genetic drift. While natural selection did play a role in shaping this trait, the selective pressures were less intense compared to those experienced by European populations. Specific genetic variants in East Asians contributed to the evolution of their characteristic flatter and broader nose shape, despite the underlying biological processes being similar across different populations. Given these findings, we focused our analysis on how the nose region changes as the SNP input decreases. As illustrated in **Figure 5**, the nose, particularly the nasal profile and tip shape, remains relatively resilient at higher SNP levels but starts to show increased error as the SNP input drops below 70%. We also analyzed the variation in error across the face and eye region, giving us a broader understanding of the effect of SNPs on different features. The Euclidean distance—which is used to quantify the discrepancy between the reconstructed and actual facial structures—increases as the level of input SNPs decreases. With complete SNP input (100%), the error is minimized at 3.517 mm. Conversely, as the proportion of SNP inputs is reduced to 10%, the error increases significantly to 3.957 mm, thereby demonstrating a marked decline in reconstruction accuracy. The heatmap provides a detailed analysis of the errors at various characteristic facial points at different SNP levels. As the SNP completeness diminishes, the error amplification at points such as the nasion, subnasale, and cheilion increases. Furthermore, the capacity to accurately categorize specific facial features, such as the nasal profile, nose tip shape, and cheekbone appearance, is diminished as the SNP input is reduced. Notably, the classification accuracy for certain features, such as the nasal profile, remains relatively high even at lower SNP levels. This suggests that certain facial aspects are more resilient to SNP scarcity than others. Extended data 3 shows more details of the change in error for generating surfaces as the number of SNPs decreases.

Despite the retention of some facial features with limited SNP data, a critical threshold was observed. When the completeness of the SNP data fell below 70% (equating to fewer than 5489 SNPs from the full set of 7842), the reconstructions began to lose individual specificity and converged toward a more generic facial structure. This observation highlights the necessity of a suffi-

ciently comprehensive SNP dataset to preserve the distinctive, individualised characteristics essential for forensic applications. It is of paramount importance to strike a balance between the robustness of the model and the availability of genetic data in order to ensure that facial reconstructions remain both accurate and person-specific. This balance is of vital importance for the enhancement of the utility and reliability of SNP-based facial reconstruction in real-world forensic settings.

2.6. Enhancing Model Interpretability in Facial Reconstruction with SHAP and GWAS

In order to further refine and elucidate the mechanism behind the Difface model, we employed SHapley Additive exPlanations^[36] (SHAP), a method based on cooperative game theory. SHAP assigns quantitative values to individual SNP features, allowing for the precise determination of how each feature influences the model's predictions. By decomposing a prediction into the contributions of specific SNPs, SHAP enables us to better understand the role of each genetic marker in shaping facial morphology.

This approach was complemented by an enrichment analysis using genome-wide association studies (GWAS), focusing on SNPs with high SHAP values and significant p-values ($p < 0.05$). Through this dual analysis, we identified key SNPs that not only play a prominent role in the model's output but also have strong genetic associations with facial structure. For instance, SNPs like *nasal_base_direction* and *nasal_root_height* consistently exhibited low hypergeometric p-values across multiple sample sizes, underscoring their relevance in both SHAP and GWAS analyses.

As shown in **Figure 6a**, the workflow begins with input SNPs processed through the frozen SNP encoder, followed by a multi-layer perceptron (MLP) and softmax classifier to predict facial feature classifications. SHAP values are then calculated via backward loss propagation to assess the impact of individual SNPs on the model's predictions. The SNPs identified through SHAP are further analyzed for biological significance through enrichment analysis.

The results, presented in **Figure 6b**, demonstrate a high degree of overlap between SHAP-determined SNP importance and SNPs identified in GWAS, confirming the robustness and interpretability of the model. SNPs relevant to features like nasal profile and cheekbone sideview were highlighted in both SHAP and GWAS results, providing strong evidence that the genetic associations are consistent with known biological influences.

Moreover, the analysis extended beyond statistical significance to explore the biological implications of these findings. Gene Ontology (GO) enrichment analysis revealed that many of these SNPs are involved in processes critical to facial development, such as “regulation of epithelial to mesenchymal transition” and “morphogenesis of embryonic epithelium.” These processes are vital to facial structure formation, influencing both bone and muscle development, as well as vascularization and symmetry. Notably, pathways like “skeletal system development” and “muscle organ development” further underscore the role of these SNPs in defining the face's structural framework. (Figure S4, Supporting Information)

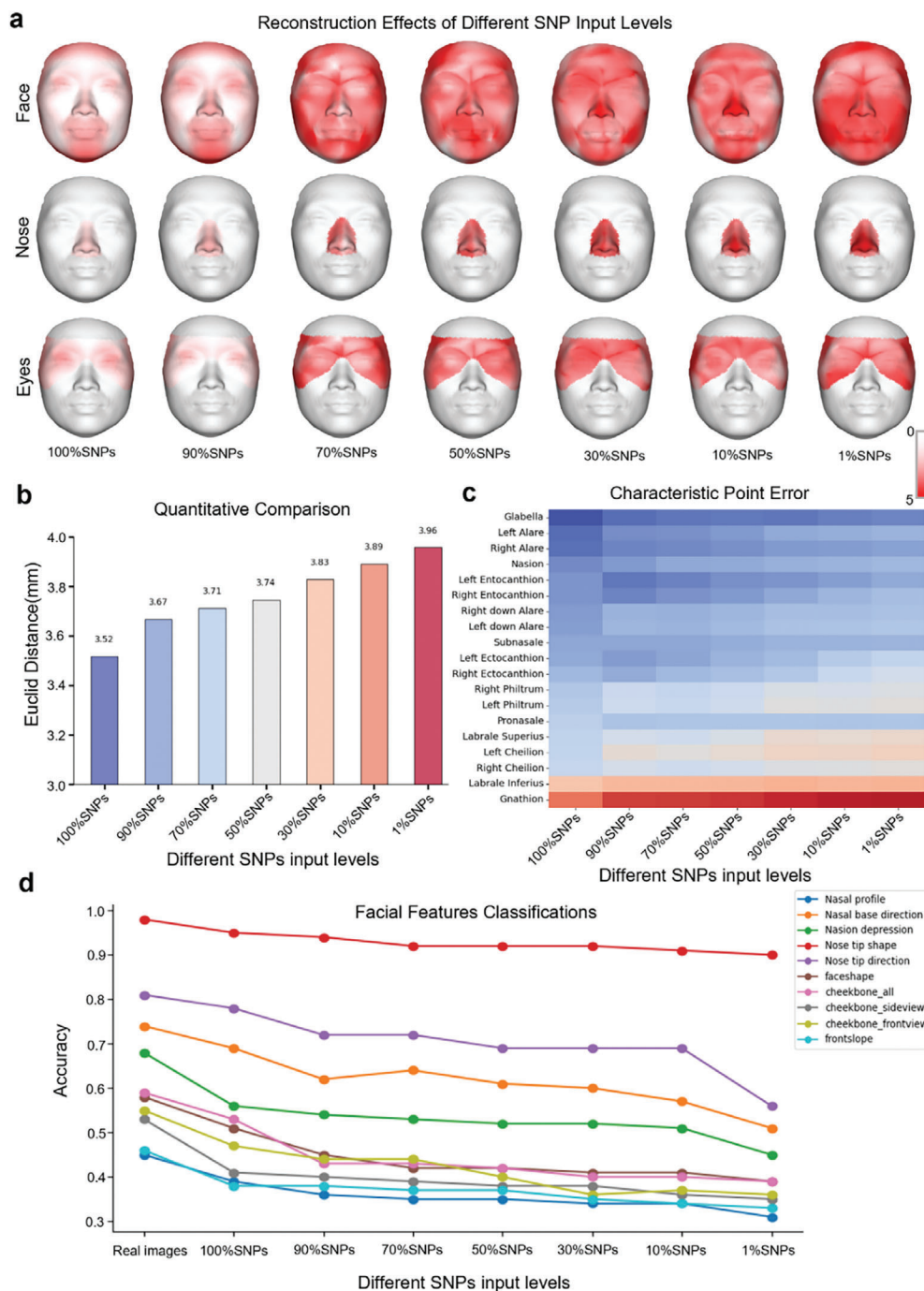


Figure 5. Impact of SNP Input Density on Facial Reconstruction Accuracy and Feature. a) Visual representation of facial reconstructions with different levels of SNP inputs, from 100% to 1%, illustrating the impact of SNP completeness on the visual accuracy of generated facial models. The first row is the whole face; the second row is the nose region; and the third row is the eye region. b) A bar graph shows the mean accuracy of facial reconstructions at different SNP densities, indicating a trend of decreasing accuracy with lower SNP percentages. c) A heat map displays the mean error rates for different facial regions across varying SNP densities, with darker shades indicating higher error values. d) Line graphs represent the stability of facial feature predictions (nasal profile, nasal base direction, nasal depression, nose shape, and nose tip direction) across different SNP input densities. Also, line graphs depict the accuracy of specific facial regions (face shape, cheekbone, cheekbone_altview, cheekbone_frontview, and frentislope) at different SNP input percentages.

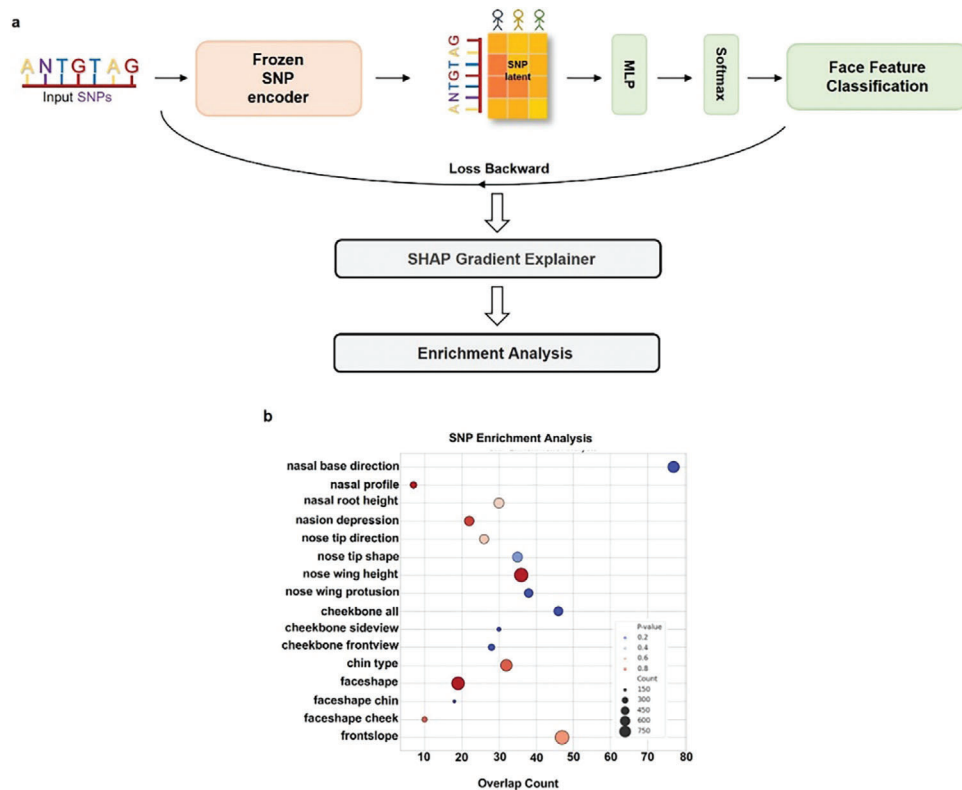


Figure 6. a) The diagram illustrates the Difface model's workflow, starting from the input of SNPs through to facial feature classification. The process includes the transformation of SNPs by a Frozen SNP encoder, SNP latent feature extraction, feature classification, and the back-propagation of losses for SHAP value calculation. The workflow concludes with an enrichment analysis to correlate SNP effects with biological processes. b) SNP Enrichment Analysis: This graph shows the overlap count between SNPs identified by SHAP values and those found significant in GWAS studies for various facial features. Each point represents a facial feature with its corresponding overlap count, indicating the statistical significance (p -value) through color coding.

This combination of SHAP and GWAS highlights the interpretability of the Difface model by showing how the SNPs identified as crucial by SHAP analysis are also statistically significant in genetic studies. The congruence between these two methods provides compelling evidence that Difface not only reconstructs facial features with high accuracy but does so in a manner consistent with established genetic influences on facial morphology.

By integrating SHAP and GWAS into our analysis, we have demonstrated that the Difface model is both interpretable and biologically grounded. These findings pave the way for future research into the genetic underpinnings of facial architecture, offering potential targets for exploring the effects of genetic variations on facial structure.

3. Discussion

This study has introduced Difface, a de novo multi-modality model to reconstruct 3D facial images from DNA with remarkable precision, by a generative diffusion process and a contrastive learning scheme. Through comprehensive analysis and SNP-FACE matching tasks, Difface demonstrated superior performance in generating accurate facial reconstructions from genetic data. In particular, Difface could generate/predict 3D facial images of individuals solely from their DNA data at various future

ages. Notably, the model's integration of transformer networks with spiral convolution and diffusion networks has set a new benchmark in the fidelity of generated images to their real images, as evidenced by its outstanding accuracy in critical facial landmarks and diverse facial feature reproduction.

Difface's novel approach, combining advanced neural network architectures, significantly outperforms existing models in genetic-to-phenotypic facial reconstruction. This superiority is attributed to its unique contrastive learning method of aligning high-dimensional SNP data with 3D facial point clouds in a unified low-dimensional feature space, a process further enhanced by adopting diffusion networks for phenotypic characteristic generation. Such advancements contribute to the model's exceptional precision and ability to capture the subtle genetic variations influencing facial morphology, a feat less pronounced in previous methodologies.

Despite Difface's demonstrated strengths, there remain directions for improvement. Addressing these limitations will require a focused effort to increase the model's robustness and adaptability to diverse datasets. Future research should aim to incorporating variables like age and BMI would allow Difface to simulate age-related changes, enabling the generation of facial images at different life stages an application that holds significant potential in both forensic science and medical diagnostics. Similarly, BMI could help the model account for variations in body composition,

improving its ability to generate accurate facial reconstructions across a range of body types.

While our study focused on a genetically homogeneous East Asian cohort, the underlying structure of the Difface model is flexible enough to be applied to different ethnic groups. By expanding our dataset to include individuals from a wide range of ethnic backgrounds, Difface has the potential to learn the specific relationships between SNPs and facial features in various populations. This would enable the model to accurately generate corresponding facial images based on the SNP input from individuals of different ethnicities. Future studies should focus on this crucial extension by testing and adapting the model to encompass more genetically diverse populations. Validating Difface with datasets from multiple ethnic groups and exploring whether additional genetic loci are necessary for certain facial features will be key steps to ensuring that the model generalizes effectively across diverse populations. This broader application will not only enhance the model's robustness but also significantly increase its utility in real-world forensic investigations and personalized medicine on a global scale.

However, this innovative technology brings forth critical ethical and legal challenges, particularly concerning genomic privacy. The ability to associate de-identified genomic data with phenotypic characteristics raises substantial risks, including potential re-identification, unauthorized data usage, and complex legal issues surrounding data ownership. By enabling the prediction of physical traits from genetic information, DNA phenotyping intensifies concerns about the misuse of sensitive personal data, which could lead to genetic-based discrimination and privacy infringements. It is therefore essential for researchers, policymakers, and industry leaders to engage in proactive dialogue to address these ethical and legal complexities.

Beyond academic exploration, the societal impacts of DNA phenotyping are extensive, touching on sensitive fields such as forensic science, healthcare, and insurance. In forensic applications, while the technology has potential benefits in identifying individuals from genetic material, improper use could result in significant harm, including wrongful convictions and racial or ethnic profiling. In healthcare and insurance, there is the risk of genetic information being used to discriminate against individuals based on perceived health risks.

To mitigate these risks, we emphasize the urgent need for comprehensive interdisciplinary analysis. Establishing clear, enforceable ethical frameworks and guidelines is paramount to ensuring that genomic research advances responsibly, safeguarding individual privacy and rights while harnessing the potential benefits of this technology. Such frameworks will help balance scientific progress with societal responsibility, creating a foundation for ethical and transparent application of DNA phenotyping across various fields.

4. Experimental Section

Difface: We proposed Difface, a multi-modality model designed to reconstruct 3D facial images from SNPs with a generative diffusion process and a contrastive learning scheme. Specifically, Difface separately employs a transformer^[31] and spiral convolution network^[32] to map high-dimensional SNPs (Y) and 3D point clouds (X) into the same low-

dimensional feature spaces (z_f and z_s) while ensuring the paired SNPs and facial images are embedded nearly compared to unpaired by contrastive learning. Subsequently, Difface uses a diffusion model to transform SNP embeddings z_s into facial image embeddings z_f conditioned on SNPs y . Finally, 3D facial images are generatively reconstructed from facial image embeddings z_f by a decoder $P(x|z_f, y)$, as shown in Figure 1.

Learning Related Features Between SNPs and 3D Point Cloud by Contrastive Learning: Our contrastive learning approach is designed to map high-dimensional SNP data and 3D facial point clouds into a unified low-dimensional feature space, leveraging a cosine similarity metric to ensure accurate alignment of true SNP-face pairs while reducing the similarity of incorrect pairings. To achieve this goal, we have designed a face encoder and a SNP encoder, as well as a decoder to generate faces from embeddings, which further constrain the potential space. The spiral convolution network was chosen to capture localized spatial features in 3D point clouds effectively. Compared to traditional convolutions, spiral convolution allows for better handling of the non-grid structure of 3D data, such as facial point clouds, by maintaining the spatial relationships between points. The specific model is described as follows:

The face encoder design was inspired by SpiralConv++^[32] which designs a spiral neighbor as:

$$0 - \text{ring}(\nu) = \nu \quad (1)$$

$$k - \text{disk}(\nu) = \cup_{i=0, \dots, k} i - \text{ring}(\nu) \quad (2)$$

$$(k+1) - \text{ring}(\nu) = N(k - \text{ring}(\nu)) \setminus k - \text{disk}(\nu) \quad (3)$$

where $n(\nu)$ is the set of all vertices adjacent to any vertex in set V , V corresponds to the vertices a point cloud network, and the length of the Spiral is represented as l . Hence, $S(\nu, l)$ is an ordered set of k -rings connected by l vertices.

The spiral convolution operator for node i is defined as follows:

$$f_{(k)}^\alpha = \gamma^{(k)} \left(\sum_{j \in S(i, l)} f_{(k-1)}^\alpha \right) \quad (4)$$

In this context, γ denotes MLPs. The encoder's structure is: $3 \times \{\text{Conv}(32) \rightarrow \text{Pool}(4)\} \rightarrow \{\text{Conv}(64) \rightarrow \text{Pool}(4)\} \rightarrow \text{FC}(128)$, with ReLU activation function after each Conv layer.

The structure of the decoder is the reversed order of the encoder with the replacement of pooling layers to unpooling layers. Note that one more convolutional layer with the output dimensional of 3 should be added to the end of the decoder to reconstruct 3D shape coordinates.

A transformer block is utilized to construct the SNP encoder. The model architecture was determined based on the original Transformer^[31] with a model dimension of 128, a feed-forward layer dimension of 512, and 2 layers in the encoder. ReLU activation was used, and the encoder had a dropout ratio of 0.1. The 7842 SNP phenotypes were embedded as a sequence and then input into the transformer block after dimensionality reduction by a Multilayer Perceptron (MLP).

We use L_1 loss for the generation loss $\text{Loss}_{\text{mesh}}$ and cross-entropy is used as the loss Loss_{con} for contrastive learning. Formally, we have

$$\text{Loss}_{\text{mesh}} = \frac{1}{N} \sum_{i=1}^N |\hat{x}_i - x_i| \quad (5)$$

$$\text{Logit}_1 = \alpha(z_f, x_{i_T}^z) \quad (6)$$

$$\text{Logit}_2 = \alpha(z_{s_i}, f_{i_T}^z) \quad (7)$$

$$\text{Loss}_{\text{con}} = \frac{l_{\text{CE}}(\text{Logit}_1, \text{labels}) + l_{\text{CE}}(\text{Logit}_2, \text{labels})}{2} \quad (8)$$

where α is a hyperparameter, and z_f, z_s are face embeddings and SNP embeddings, respectively. l_{CE} is the cross-entropy loss, and $labels$ is a vector just like $[0, 1, 2, \dots, N - 1]$, where N is the batch size.

Thus, our overall loss function is

$$\text{Loss}_{\text{ALL}} = \beta_1 \text{Loss}_{\text{con}} + \beta_2 \text{Loss}_{\text{mesh}} \quad (9)$$

where β_1, β_2 are hyperparameters.

Reconstructing 3D Facial Images from SNPs by Diffusion Model: We trained an encoder to generate SNP embeddings and facial embeddings, along with a decoder for generating faces from facial embeddings. However, we require a prior model to generate facial embeddings from SNP embeddings. To achieve this, we employ a diffusion model^[37–39] in this context.

The diffusion model operates through a two-step process: the forward diffusion and the reverse denoising. During the forward diffusion process, we introduce Gaussian noise into the facial embeddings $z_{f_0} \sim q(z_f)$ over T steps. To accomplish this, a series of hyperparameters representing Gaussian distribution variances $t = \frac{1}{T} \{\beta_t \in (0, 1)\}$ are required. The forward process can be viewed as a Markov process since each time step t only depends on the previous time step $t - 1$. So, it can also be regarded as a Markov process.

$$\begin{aligned} q(z_{f_t} | z_{f_{t-1}}) &= N(z_{f_t}; \sqrt{1 - \beta_t} z_{f_{t-1}}, \beta_t I), q(q(z_{f_{1:T}} | z_{f_{t-1}}) | x_0) \\ &= \prod_{t=1}^T q(z_{f_t} | z_{f_{t-1}}) \end{aligned} \quad (10)$$

In this process, as t increases, z_{f_t} becomes progressively closer to pure noise. When $T \rightarrow \infty$, z_{f_T} becomes entirely Gaussian noise. If we consider the forward process as the noise introduction process, then the reverse process is the denoising diffusion process of the diffusion model. If we could obtain the reverse distribution step by step $q(z_{f_{t-1}} | z_{f_t})$, we could recover the original distribution x_0 from the fully standard Gaussian distribution $z_{f_T} \sim N(0, I)$. However, inferring $q(z_{f_{t-1}} | z_{f_t})$ is not straightforward. Therefore, we use a deep learning model to predict a reverse distribution p_θ .

$$p_\theta(X_{0:T}) = p(z_{f_T}) \prod_{t=1}^T p_\theta(z_{f_{t-1}} | z_{f_t}) \quad (11)$$

$$p_\theta(z_{f_{t-1}} | z_{f_t}) = N\left(z_{f_{t-1}}; \mu_\theta(z_{f_t}, t), \sum_{\theta} (z_{f_t}, t)\right) \quad (12)$$

While we may not have access to the reverse distribution $q(z_{f_{t-1}} | z_{f_t})$, if we know x_0 , we can indeed use Bayes' theorem to obtain $q(z_{f_{t-1}} | z_{f_t}, z_{f_0})$.

$$q(z_{f_{t-1}} | z_{f_t}, z_{f_0}) = N(z_{f_{t-1}}; \tilde{\mu}(z_{f_t}, z_{f_0}), \tilde{\beta} I) \quad (13)$$

To achieve the generation of facial images from SNPs, SNP embeddings were utilized as a conditional bootstrap diffusion model to generate facial embeddings. The forward process of the conditional diffusion model is identical to that of the unconditional diffusion model. Hence, the joint probability of the backward process is

$$p_\theta(X_{0:T} | \gamma) = p(z_{f_T}) \prod_{t=1}^T p_\theta(z_{f_{t-1}} | z_{f_t}, \gamma) \quad (14)$$

We find it better to train our model to predict the unnoised z_{f_T} directly, and thus use a mean-squared error loss on this prediction:

$$L_{\text{diffusion}} = E_{t \sim [1, T], z_{f_t}^f \sim q_t} \left[\|f_\theta(z_{f_t}^f, t, \gamma) - z_{f_t}\|^2 \right] \quad (15)$$

Model Architecture and Components: The architecture of the Difface model is composed of several key components that work together to achieve high performance in reconstructing facial features from genetic data. At the core of the model is the integration of convolutional and transformer-based networks, enabling effective feature extraction and the modeling of complex spatial relationships.

The face encoder employs a spiral convolutional network with a sequence length of $[9, 9, 9, 9]$ and a dilation factor of $[1, 1, 1, 1]$, allowing it to capture localized spatial features while expanding the receptive field without loss of resolution. Additionally, the SNP encoder consists of two transformer layers, each with a model dimension of 128, an 8-headed self-attention mechanism, and a feed-forward network with a hidden dimension of 512. These layers incorporate a dropout rate of 0.1 and apply layer normalization with an epsilon value of $1e-5$, ensuring stable training and robust generalization.

The diffusion prior model is designed to predict the starting state without conditional dropout, operating with an hidden embedding dimension of 128 and 1 hidden channel. It features a 4-layer depth and an attention mechanism with 4 heads, each with a head dimension of 64. The feed-forward network has an expansion multiplier of 2, with a 0.1 dropout rate applied to both the attention and feed-forward layers. The diffusion model operates over 1000 timesteps, allowing the model to handle input data directly and reconstruct facial features with high precision.

Optimization and Parameters: To ensure robust performance and improve learning efficiency, we employed several optimization strategies in training the Difface model, including gradient descent algorithms and dynamic learning rate scheduling.

The Adam optimizer was selected as the primary optimization method. Different components of the model required tailored configurations. For the core model parameters, we utilized a learning rate of 7.3×10^{-6} and a weight decay of 0.30, preventing overfitting while enhancing generalization capabilities. For the decoder, which plays a key role in reconstructing 3D facial images from learned embeddings, a higher learning rate of 3.2×10^{-5} and a lower weight decay of 0.001 were applied. To further refine the training process, we incorporated a StepLR scheduler that dynamically adjusted the learning rate during training. The learning rate was reduced by a factor of 0.99 after each epoch, enabling a gradual decrease that fine-tuned the model weights.

The diffusion prior model was trained with a learning rate of 0.31×10^{-4} and a weight decay of 0.752. A maximum gradient norm of 0.5 was enforced to prevent gradient explosion, ensuring stable convergence throughout the process. Additionally, an exponential moving average (EMA) was implemented with a beta value of 0.98, and updates were applied every 1000 steps, with EMA updates occurring every 100 steps to smooth the learning process and stabilize model weights.

We utilized a dataset of 9674 samples, each containing 3D facial images, SNP data, age, BMI, and other relevant information. The dataset was randomly split into training and testing sets, with 80% allocated for training and 20% for testing. The model was trained using four NVIDIA 3080Ti GPUs, with a total training time of ≈ 48 h.

Quantification of Diversity of the Learned SNP Features: We measure this diversity using a metric based on the determinant of a kernel matrix:

$$\text{dpp diversity} = \det(K) \quad (16)$$

where $K_{ij} = \frac{1}{1 + \text{dist}(c_i, c_j)}$ and $\text{dist}(c_i, c_j)$ represents the distance metric between real faces and generated faces. In practice, to avoid ill-conditioned determinants, we introduce small random perturbations to the diagonal elements when computing the determinant.

Evaluating Feature Importance: The Difface model employs a component known as the GradientExplainer to leverage SNP phenotypes and their corresponding facial images within the analytical framework. This tool is designed for use with models that incorporate gradient data in order to elucidate the relationship between input variables and output responses. By capturing and analyzing the gradient signals, the GradientExplainer identifies the extent to which the model's predictions are sensitive to variations in SNP inputs. The resulting SHAP values for each SNP pro-

vide a direct metric of their influence, offering a detailed and quantitative understanding of the genetic determinants affecting facial structure.

SNPs Enrichment Analysis: We conducted two distinct enrichment analyses to better understand the genetic associations of SNPs identified through the Difface model:

Analysis of Top 200 SNPs: For each phenotype related to somatic anthropology, we selected the top 200 SNPs that showed the strongest correlation based on SHAP values. These SNPs were filtered to retain independent loci, which were then subjected to genetic annotation and subsequent enrichment analysis. This process aimed to identify biological pathways and processes that are significantly influenced by these SNPs.

Analysis of Top 500 SNPs: In this analysis, we selected the top 500 SNPs with the highest correlation, as determined by SHAP values. These SNPs were analyzed for enrichment against genome-wide association studies (GWAS), focusing on those with a significance threshold of p -value < 0.05 . This allowed us to identify SNPs that are not only strongly correlated with the phenotypes but also statistically significant in broader genetic studies.

Datasets and Processing Ethics Statement Sample and Recruitment Details: Ethics statement. All participants provided written informed consent, and all study protocols were approved by the institutional review boards of the pertinent research institutions. The National Survey of Physical Traits (NSPT) is a subproject of The National Science & Technology Basic Research Project approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (14117). The Northern Han Chinese (NHC) cohort was approved by the Ethics Committee of Human Genetic Resources at the Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410-A1801). The Taizhou Longitudinal Study (TZL) was approved by the Ethics Committee of Human Genetic Resources at the Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410). Written informed consent was granted to each participant before enrollment in the study. We confirm that our research is compliant with the Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources.

The samples in this study were collected from three independent cohorts, the NSPT cohort ($n = 3322$), the NHC cohort ($n = 4767$) and the TZL cohort ($n = 2881$). For the NSPT sample, individuals were recruited in three Chinese cities: Nanning, Guangxi province ($n = 1326$); Taizhou, Jiangsu Province ($n = 986$); Zhengzhou, Henan Province ($n = 1010$). In the NHC cohort, participants were recruited in Tangshan, Hebei province.

3D Image Acquisition, Registration, and Quality Control: The 3D images of all individuals in the three cohorts were captured and acquired using the 3dMDface (3dMD) camera system. Participants were asked to close their mouth, open their eyes and hold faces with a neutral expression when capturing. The 3D surface images were registered using MeshMonk (v.0.0.6)^[40] in MATLAB 2018a. This performed a homologous configuration of 7906 spatially dense landmarks, allowing the 3D image data to be standardized. We performed generalized procrustes analysis (GPA) and symmetrization, then investigated every mapped image manually and identified outlier images. Any 3D facial images with poor quality were removed or reprocessed, with details available in the Supplementary Note. As a result, 6968 ($n = 4089$ in the NHC cohort, $n = 2879$ in the NSPT cohort) and 2706 unrelated individuals with good quality 3D images in the discovery and replication dataset were used for further analysis.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

M.J. and J.L. contributed equally to this work. This work was supported by the National Natural Science Foundation of China (Nos. T2341007, T2350003, 12131020, 42450084, 42450135, 12326614, 12426310 to L.C.; 32300523, 62132015 to C.Z.; and 32325013, 92249302 to S.W.), National Key R&D Program of China (2022YFA1004800 to C.L. and

2024YFC3405800 to S.W.), Zhejiang Province Vanguard Goose-Leading Initiative (2025C01114) to C.L., Hangzhou Institute for advanced study of UCAS (2024HIAS-P004) to C.L., the Science and Technology Commission of Shanghai Municipality (No. 23JS1401300) to C.L., and the JST Moonshot R&D (No. JPMJMS2021) to C.L., CAS Young Team Program for Stable Support of Basic Research (YSBR-077) to S.W., CAS Interdisciplinary Innovation Team to S.W., the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB38020400) to S.W., Shanghai Municipal Science and Technology Major Project (2017SHZDZX01) to S.W., Shanghai Science and Technology Innovation Action Project (24JS2810300) to S.W., the National Key Research and Development Project (2018YFC0910403) to S.W., Shanghai Science and Technology Commission Excellent Academic Leaders Program (22XD1424700) to S.W.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Keywords

artificial Intelligence in genetics, deep multimodal integration, DNA-guided 3D face generation, genetic phenotyping, representational learning

Received: November 7, 2024
Published online: May 7, 2025

- [1] M. Kayser, *Forensic Sci. Int. Genet.* **2015**, *18*, 33.
- [2] P. Gill, A. J. Jeffreys, D. J. Werrett, *Nature* **1985**, *318*, 577.
- [3] L. Roewer, *Invest. Genet.* **2013**, *4*, 1.
- [4] J. Roosenboom, G. Hens, B. C. Mattern, M. D. Shriver, P. Claes, *Biomed Res. Int.* **2016**, *2016*, 3054578.
- [5] F. Liu, A. J. S. W. Rayner, C. P. J. Wright, M. T. Smith, L. R. L. Dunham, A. M. K. Browning, D. P. J. N. G. Houghton, L. M. Bowles, C. R. Richmond, S. D. B. Smith, D. M. Evans, *PLOS Genet.* **2012**, *8*, e1002932.
- [6] L. Paternoster, A. I. Zhurov, A. M. Toma, J. P. Kemp, B. St. Pourcain, N. J. Timpson, G. McMahon, W. McArdle, S. M. Ring, G. D. Smith, S. Richmond, D. M. Evans, *Ame. J. Human Genet.* **2012**, *90*, 478.
- [7] K. Adhikari, G. Reales, A. J. P. Smith, E. Konka, J. Palmen, M. Quinto-Sanchez, V. Acuña-Alonso, C. Jaramillo, W. Arias, M. Fuentes, M. Pizarro, R. Barquera Lozano, G. Macín Pérez, J. Gómez-Valdés, H. Villamil-Ramírez, T. Hunemeier, V. Ramallo, C. C. Silva de Cerqueira, M. Hurtado, V. Villegas, V. Granja, C. Gallo, G. Poletti, L. Schuler-Faccini, F. M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, F. Rothhammer, G. Bedoya, R. Calderón, et al., *Nat. Commun.* **2015**, *6*, 7500.
- [8] J. B. Cole, M. Manyama, E. Kimwaga, J. Mathayo, J. R. Larson, D. K. Liberton, K. Lukowiak, T. M. Ferrara, S. L. Riccardi, M. Li, W. Mio, M. Prochazkova, T. Williams, H. Li, K. L. Jones, O. D. Klein, S. A. Santorico, B. Hallgrímsson, R. A. Spritz, *PLoS Genet.* **2016**, *12*, e1006174.
- [9] M. K. Lee, J. R. Shaffer, E. J. Leslie, E. Orlova, J. C. Carlson, E. Feingold, M. L. Marazita, S. M. Weinberg, *PLoS One* **2017**, *12*, e0176566.

- [10] S. Cha, J. E. Lim, A. Y. Park, J.-H. Do, S. W. Lee, C. Shin, N. H. Cho, J.-O. Kang, J. M. Nam, J.-S. Kim, K.-M. Woo, S.-H. Lee, J. Y. Kim, B. Oh, *BMC Genomics* **2018**, *19*, 481.
- [11] P. Claes, J. Roosenboom, J. D. White, T. Swigut, D. Sero, J. Li, M. K. Lee, A. Zaidi, B. C. Mattern, C. Liebowitz, L. Pearson, T. González, E. J. Leslie, J. C. Carlson, E. Orlova, P. Suetens, D. Vandermeulen, E. Feingold, M. L. Marazita, J. R. Shaffer, J. Wysocka, M. D. Shriver, S. M. Weinberg, *Nat. Genet.* **2018**, *50*, 414.
- [12] D. J. M. Crouch, B. Winney, W. P. Koppen, W. J. Christmas, K. Hutnik, T. Day, D. Meena, A. Boumertit, P. Hysi, A. Nessa, T. D. Spector, J. Kittler, W. F. Bodmer, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E676.
- [13] Lu Qiao, Y. Yang, P. Fu, S. Hu, H. Zhou, S. Peng, J. Tan, Y. Lu, H. Lou, D. Lu, S. Wu, J. Guo, L. Jin, Y. Guan, S. Wang, S. Xu, K. Tang, *J. Genet. Genom.* **2018**, *45*, 419.
- [14] Yi Li, W. Zhao, D. Li, X. Tao, Z. Xiong, J. Liu, W. Zhang, A. Ji, K. Tang, F. Liu, C. Li, *Human Genet.* **2019**, *138*, 681.
- [15] W. Wu, J. Wang, X. Li, Z. Yang, M. Zhou, H. Yu, C. Zhang, Z. Xu, Y. Zhang, Y. Li, C. J. Qiu, P. Li, D. Li, Y. Wu, L. Yang, Y. Li, L. Chen, L. Wu, Y. Qian, Y. Zhang, Y. Wang, F. Li, F. Ma, D. H. Chen, Y. Xu, L. Li, X. Zhang, Y. Wu, C. Zhang, Q. Li, et al., *Human Genet.* **2019**, *138*, 601.
- [16] Z. Xiong, J. Li, L. Zhang, Y. He, X. Huang, Q. Yu, J. Jiang, X. Chen, W. Zhang, Y. Liu, J. Wei, S. X. Qian, Y. Wu, J. Zhang, X. Yang, H. Tang, J. Dong, Y. Chen, L. Xie, H. Liu, P. Ma, Z. Wu, D. Zhou, X. Gu, Z. Zhang, X. He, Y. Shen, H. Song, X. Li, *Elife* **2019**, *8*, e49898.
- [17] Y. Huang, D. Li, L. Qiao, Y. Liu, Q. Peng, S. Wu, M. Zhang, Y. Yang, J. Tan, S. Xu, L. Jin, S. Wang, K. Tang, S. Grünwald, *J. Genet. Genom.* **2021**, *48*, 198.
- [18] Z. Xiong, X. Gao, Y. Chen, Z. Feng, S. Pan, H. Lu, A. G. Uitterlinden, T. Nijsten, A. Ikram, F. Rivadeneira, M. Ghanbari, Y. Wang, M. Kayser, F. Liu, *Nat. Commun.* **2022**, *13*, 7832.
- [19] M. Zhang, S. Wu, S. Du, W. Qian, J. Chen, L. Qiao, Y. Yang, J. Tan, Z. Yuan, Q. Peng, Y. Liu, N. Navarro, K. Tang, A. Ruiz-Linares, J. Wang, P. Claes, L. Jin, J. Li, S. Wang, *Nat. Genet.* **2022**, *54*, 403.
- [20] Y. Gurovich, Y. Hanani, O. Bar, G. Nadav, N. Fleischer, D. Gelbman, L. Basel-Salmon, P. M. Krawitz, S. B. Kamphausen, M. Zenker, L. M. Bird, K. W. Gripp, *Nat. Med.* **2019**, *25*, 60.
- [21] P. Claes, D. K. Liberton, K. Daniels, K. M. Rosana, E. E. Quillen, L. N. Pearson, B. McEvoy, M. Bauchet, A. A. Zaidi, W. Yao, H. Tang, G. S. Barsh, D. M. Absher, D. A. Puts, J. Rocha, S. Beleza, R. W. Pereira, G. Baynam, P. Suetens, D. Vandermeulen, J. K. Wagner, J. S. Boster, M. D. Shriver, *PLoS Genet.* **2014**, *10*, e1004224.
- [22] D. Sero, A. Zaidi, J. Li, J. D. White, T. B. G. Zarzar, M. L. Marazita, S. M. Weinberg, P. Suetens, D. Vandermeulen, J. K. Wagner, M. D. Shriver, P. Claes, *Nat. Commun.* **2019**, *10*, 2557.
- [23] S. S. Mahdi, N. Nauwelaers, P. Joris, G. Bouritsas, S. Gong, S. Walsh, M. D. Shriver, M. Bronstein, P. Claes, *IEEE Trans. Biom. Behav. Identity Sci.* **2022**, *4*, 163.
- [24] A. J. O'toole, T. Vetter, N. F. Troje, H. H. Bülthoff, *Perception* **1997**, *26*, 75.
- [25] B. Xia, B. B. Amor, M. Daoudi, *Image Vision Comput.* **2017**, *64*, 90.
- [26] D. Huang, H. Ding, C. Wang, Y. Wang, G. Zhang, L. Chen, *Image Vision Comput.* **2014**, *32*, 1181.
- [27] M. N. Skorska, S. N. Geniole, B. M. Vrysen, C. M. McCormick, A. F. Bogaert, *Arch. Sexual Behav.* **2015**, *44*, 1377.
- [28] A. Radford, L. Metz, D. P. Kingma, P. R. Day, J. M. Shinn, I. Sutskever, J. M. Susskind, P. K. G. Pritchard, R. Salakhutdinov, A. S. P. Wibisono, S. K. Lee, Z. A. Miller, *Proc. 38th Int. Conf. on Machine Learning*, PMLR **2021**, 8748–8763.
- [29] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, *arXiv* **2022**, 2204.06125.
- [30] C. Saharia, A. Gupta, S. A. Mohamed, B. A. Larson, A. M. Khan, A. P. Goyal, R. S. Rajendran, A. M. Shankar, R. S. Rosenthal, P. P. Khatri, P. M. Zappella, D. J. Lee, B. D. Chinn, A. A. Mehrotra, A. G. Singh, N. R. Sharma, M. R. Johnson, *Adv. Neural Informat. Proc. Syst.* **2022**, *35*, 36479.
- [31] A. Vaswani, N. Shazeer, N. Parmar, L. Uszkoreit, J. Kaiser, M. Polosukhin, A. Jones, A. Gomez, L. Kaiser, M. K. K. K. Lee, S. Tsai, J. H. K. Chuang, W. Wu, Z. Yu, K. Kaiser, A. Dehghani, A. V. Schuster, P. Vaswani, *Adv. Neural Informat. Proc. Syst.* **2017**, *30*, 1.
- [32] S. Gong, L. Chen, M. Bronstein, S. Zafeiriou, in *Proc. IEEE/CVF Int. Conf. on Computer Vision Workshops*, IEEE, Piscataway, NJ **2019**.
- [33] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.
- [34] H. Hotelling, *Relations Between Two Sets of Variates*, Springer, New York **1992**, pp. 162–190.
- [35] A. Kulesza, B. Taskar, *Foundat. Trends® Mach. Learn.* **2012**, *5*, 123.
- [36] S. M. Lundberg, S.-I. Lee, *Adv. Neural Informat. Proc. Syst.* **2017**, *30*, 1.
- [37] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, presented at *Int. Conf. Machine Learning*, Lille, France, **2015**.
- [38] J. Ho, A. Jain, P. Abbeel, *Adv. Neural Informat. Proc. Syst.* **2020**, *33*, 6840.
- [39] L. Zhang, A. Rao, M. Agrawala, in *Proc. IEEE/CVF Int. Conf. on Computer Vision*, IEEE, Piscataway, NJ **2023**, pp. 3836–3847.
- [40] J. D. White, M. S. Finkelstein, L. H. Fields, C. E. Klein, N. A. Anderson, J. L. Smith, D. G. Lee, P. J. Kim, L. A. Sullivan, T. G. Barnes, J. H. Zhang, A. T. McDonald, M. R. Gonzalez, B. W. Rogers, T. A. Thomas, E. M. Walters, *Sci. Rep.* **2019**, *9*, 6085.