SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-023-2518-8

RESEARCH PAPER

Coordinate-wise monotonic transformations enable privacy-preserving age estimation with 3D face point cloud

Xinyu Yang[1,2], Runhan Li[2], Xindi Yang[3], Yong Zhou[4], Yi Liu[3] & Jing-Dong J. Han[2*]

[1]School of Life Sciences, Peking University, Beijing 100871, China;
[2]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing 100871, China;
[3]Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;
[4]Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

*Corresponding author (email: jackie.han@pku.edu.cn)

Received 26 September 2023; Accepted 25 December 2023; Published online 2 April 2024

The human face is a valuable biomarker of aging, but the collection and use of its image raise significant privacy concerns. Here we present an approach for facial data masking that preserves age-related features using coordinate-wise monotonic transformations. We first develop a deep learning model that estimates age directly from non-registered face point clouds with high accuracy and generalizability. We show that the model learns a highly indistinguishable mapping using faces treated with coordinate-wise monotonic transformations, indicating that the relative positioning of facial information is a low-level biomarker of facial aging. Through visual perception tests and computational 3D face verification experiments, we demonstrate that transformed faces are significantly more difficult to perceive for human but not for machines, except when only the face shape information is accessible. Our study leads to a facial data protection guideline that has the potential to broaden public access to face datasets with minimized privacy risks.

face point cloud | age estimation | face verification | privacy | coordinate-wise monotonic transformation

INTRODUCTION

Facial characteristics are valuable predictors of various health statuses, such as coronary artery disease (Lin et al., 2020), obesity (Chanda and Chatterjee, 2021), and syndromes with distinctive facial phenotypes (Hsieh et al., 2022). With aging now formally classified as a disease that contributes to multiple non-communicable diseases (The Lancet Healthy Longevity, 2022; World Health Organization, 2019), efficient monitoring of age-related changes has become increasingly important. The human face is a convenient, non-invasive, and effective biomarker of aging, compared with fundus imaging (Zhu et al., 2023) or biopsy-based measures such as telomere length (Vaiserman and Krasnienkov, 2020), methylome (Bell et al., 2019), proteome (Lehallier et al., 2020), transcriptome (Meyer and Schumacher, 2021) and microbiome (Galkin et al., 2020). Over decades, computational models designed to estimate age using 2D facial images have evolved from machine learning algorithms based on explicit facial feature extraction (Lanitis et al., 2004) or implicit aging patterns (Geng et al., 2007) to deep learning techniques (Rothe et al., 2018; Tan et al., 2019), enabling automatic and high throughput age estimation. Advances in 3D body scanning technology (Bartol et al., 2021) now allow for accurate collection of 3D facial data (Lübbers et al., 2010), which provides shape boundary independent of pose and illumination (Mu et al., 2019; Xia et al., 2013) and has been shown to be more informative than 2D images (Eng et al., 2017; Ko et al., 2021; Matthews et al., 2018). Most 3D scanners generate point cloud output through system software (Kersten et al., 2018); however, the lack of large datasets or suitable deep learning architectures for face point cloud data has prevented direct deep learning on 3D face. Current approaches to age estimation with 3D facial data involve machine learning on knowledge-based facial features (Imai and Okami, 2019; Windhager et al., 2019; Xia et al., 2017) or template-registered vectors (Chen et al., 2015; Matthews et al., 2018), or deep learning on RGB-D data (Xia et al., 2020). However, these methods are either limited in predictive power or subject to error-prone preprocessing procedures, which constrains their application for industry-scale use. In 2017, deep learning architectures suitable for point cloud data format (Guo et al., 2021b; Qi et al., 2017a; Qi et al., 2017b) began to emerge, and later the attention mechanism (Vaswani et al., 2017) was introduced to improve performance in 3D tasks (Engel et al., 2021; Guo et al., 2021a; Zhao et al., 2021). To date, there have been limited attempts to apply deep learning directly to 3D face point cloud data in the field of face recognition (Zhang et al., 2019) and facial expression detection (Nguyen et al., 2021), while the applicability of this methodology for age estimation remains unclear.

Despite the immense value, the human face, as a hard biometric, is easily accessible and encounters the hurdle of data security once collected or shared (Nandakumar and Jain, 2009; Niinuma et al., 2010; Sundararajan et al., 2019). The occurrence of facial data leakage (Ilia et al., 2015; Liu et al., 2021) and spoofing (Boulkenafet et al., 2016) events has prompted intense ethical, security and privacy concerns (Ilia et al., 2015; Liu et al., 2021; Roussi, 2020; Smith and Miller, 2022). Recent research has also found that individuals with similar appearances share common genotypes, identifying single nucleotide polymorphisms associated with facial structures and

gene-expression changes. This suggests the potential for an individual's multiomics to be revealed through their facial features, linking facial data privacy directly to genetic data privacy (Joshi et al., 2022), adding more importance to guarding the privacy of the facial, in particular, 3D facial identity information. Responding to these concerns, public face datasets have been decommissioned (e.g., VGGFace (Parkhi et al., 2015)) or fallen under strict restrictions (e.g., datasets maintained by the Notre Dame Computer Vision Research Lab). These actions alleviate the danger of facial data misuse while inevitably raising the threshold for data driven facial research. In addition to policy constraints, facial data masking techniques have been proposed from both visual confusion (Kaur and Khanna, 2016; McPherson et al., 2016; Yang et al., 2022a) and data encryption (Patel et al., 2015; Shan et al., 2020) perspective to enable data protection. However, most of these techniques are designed for 2D images and indiscriminately eliminate all facial information, including both ID-dependent and ID-independent features (Patel et al., 2015). The possibility of decomposing identity-independent age features was posited but not validated (Xu et al., 2017). One study developed an algorithm to overlay digital masks on data subjects' faces in images, preserving certain facial features crucial for syndrome diagnosis (Yang et al., 2022b). This hints at the potential of separating specific facial phenotypes from identity. Overall, there remains a significant gap between the widespread accessibility of 3D facial data and the scarcity of 3D facial data masking techniques, and it is of great interest to develop facial data masking approaches that not only ensure privacy but also retain information of interest, such as age-related features.

In this study, we establish a FPCT-age (Face Point Cloud Transformer for age estimation) model to learn age estimation on a dataset of 3D face point clouds, which exhibits high accuracy and generalizability over a broad age range. We develop an algorithm called coordinate-wise rank transformation (CRT), which eliminates the Euclidean magnitudes of vertices and substantially distorts face shapes. We show that FPCT-age learns an indistinguishable mapping to age using faces treated with CRT or other coordinate-wise monotonic transformations (CMTs), indicating that the relative positioning of facial information (e.g., the normal vector and color) encodes age-related facial phenotypes. Through visual perception tests, we confirm that perceiving the identity, sex and age of CRT-treated faces is significantly more difficult. We also establish a minimal 3D face verification model named FPCT-ID, which, when trained with face shape data, achieves accurate verification, and can be invalidated by CRT. Finally, we propose a facial data protection guideline that has the potential to broaden data access for both face recognition applications and other face-based applications with minimized privacy leakage risks.

RESULTS

Direct age estimation using 3D face point clouds and a measure of model indistinguishability

We collected a dataset of 16,580 three-dimensional human face point clouds of Han Chinese, using the 3dMDface system (Lübbers et al., 2010). The dataset spanned 68 years in age (see Figure S1A in Supporting Information for demographics). We applied a 10-fold cross-validation approach for model training and evaluation (see Figure S1B in Supporting Information for age group divisions and Figure S1C in Supporting Information for sex distribution in the training sets). We applied simple preprocessing to the raw data (see Materials and Methods) to generate standard face point clouds containing vertex coordinates, normal vector components, and RGB color components, denoted as the V, N, and C fields, respectively. We customized the deep learning architecture of point cloud transformer (PCT) (Guo et al., 2021a) to adapt to the input dimensionality and the regression task. A range of hyperparameters was fine-tuned by grid search (see Materials and Methods) to fit the subtle characteristics of the human face. Figures S2 and S3 in Supporting Information illustrate the impact of training data quantity and point cloud sampling rate on model accuracy, respectively. We used distributed data parallel across two GPUs to expedite the training process, which also helped facilitate convergence (see Figure S4 in Supporting Information). The resulting model was referred to as the FPCT-age. Figure 1A provides a visual summary of the workflow. Figure 1B presents the optimal validation set mean absolute error (MAE) of individual checkpoints (with each checkpoint defined as a dictionary storing model parameters at a given epoch), the test set MAE obtained from the average ensemble specific to each training set, and the test set MAE from the average ensemble of all checkpoints (see Materials and Methods). The test set MAE was consistently lower than the validation set MAE, indicating the robust generalizability of FPCT-age. The ensemble of all checkpoints registered a MAE of <2.5 and a Pearson correlation coefficient (PCC) of 0.970 against chronological age, showcasing FPCT-age's superior performance on par with our earlier convolutional neural network models (Xia et al., 2020). Data augmentation facilitated FPCT-age in capturing the rotation-invariance intrinsic to 3D face point clouds, as manifested by the tight standard deviations in 100 iterations of age estimation, each applying varied random rotations of test faces (Figure 1C).

In deep learning, different model checkpoints or different model ensembles are unlikely to generate identical output for the same input. The reasons include the randomness in model initialization and optimization, and embedding strategies, e.g., the farthest point sampling (FPS) steps in FPCT-age's hidden layers. Additionally, it has been proposed that the perceived age of a face with a certain actual age possesses a wide distribution (Chen et al., 2015; Geng et al., 2014; Xia et al., 2020), which makes age estimation with facial data inherently non-deterministic. Therefore, to evaluate the closeness of two age estimation models, a proper metric was needed. That knowledge distillation (or model compression) (Gou et al., 2021; Hinton et al., 2015; Lopez-Paz et al., 2015) requires a student model to mimic the teacher model's activation pattern and/or outputs on the same inputs, and that machine unlearning (Bourtoule et al., 2021; Guo et al., 2020; Neel et al., 2021) requires a statistical indistinguishability to minimize performance shrinkage after data deletion, both implicitly point to the concept that describes the extent to which two models are indistinguishable in terms of making inferences. However, in both fields, the primary focus has been on designing the loss function to guide the training process, rather than a direct and conclusive measurement based on inference results. Therefore, here we propose the model indistinguishability (MI) score, to evaluate the similarity in pattern and magnitude of two sets of real-valued output. Specifically, PCC and the mean absolute difference (MAD) were calculated using

https://doi.org/10.1007/s11427-023-2518-8

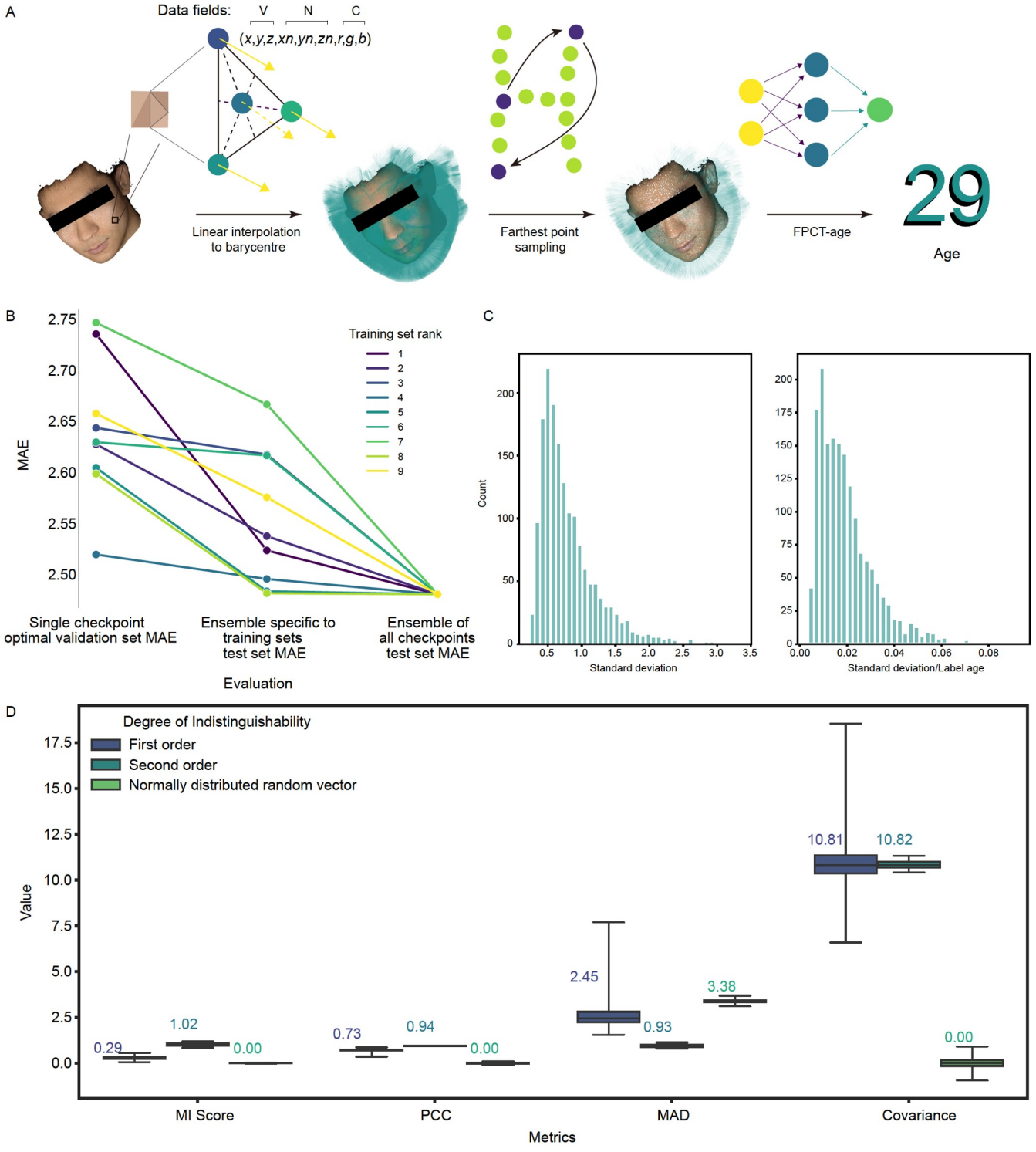

Figure 1. Data preprocessing and model performance. A, Overview of the data preprocessing procedures. The raw facial point cloud data, captured by the 3dMDface system, was initially organized in triangulation and contained over 2 million floating point numbers. To compress the data quantity required to represent a 3D face with sufficient age-related information, triangle information was linearly interpolated to mass centers and down-sampled by FPS. This reduced the number of floating-point numbers to approximately 0.1 million. The reduced point cloud data then underwent common data augmentation steps (see Materials and Methods) before being used for training the FPCT-age model. B, Comparison of age estimation accuracy as indicated by MAE. The lowest MAE on the validation set throughout the training process for each training set is shown in the first column. The second column displays the MAE values on the test set for training set-specific ensembles, which are consistently better than the counterparts in the first column, indicating favorable generalizability. The third column shows averaging over the nine training set-specific ensembles converged and further decreased test set MAE. C, Rotation-invariant age estimation. Histogram showcasing the frequency distribution of standard deviations for data subjects' age estimations (left) and the frequency distribution of the ratio between data subjects' age estimation standard deviations and their labeled ages (right). Age estimates were derived using the average ensemble from the first training set. D, Comparison of indistinguishability metrics. First-order metrics were evaluated using 404,550 pairs of model checkpoints. Second-order metrics were evaluated using 36 pairs of training set-specific model ensembles. For random array simulations, the random seed was set to 0.

age estimation residuals (the difference between the estimated and actual age for each data subject) generated by two inferences, and the ratio between them was defined as the MI score. Eq. (1) provides a full formulation of the MI score,

$$\text{MI score}=\frac{\text{PCC}}{\text{MAD}}=\frac{n\sum_{k=1}^{n}(x_k-\mu_x)(y_k-\mu_y)}{\sqrt{\sum_{k=1}^{n}(x_k-\mu_x)^2\sum_{k=1}^{n}(y_k-\mu_y)^2}\sum_{k=1}^{n}|x_k-y_k|}, \quad (1)$$

where k counts from 1 to the total number of sample subjects (n), x_k stands for the estimate of a subject k by model X, and y_k stands for the estimate by model Y. μ_x stands for the sample mean of model X's estimates, and μ_y for model Y.

We calculated the average MI score across all pairwise combinations of checkpoints to represent a threshold level of indistinguishability, termed as the first-order MI. All MI scores were statistically significant with P-values<1×10^{-5} (see Materials and Methods). The pairwise PCC, MAD and covariance values for these checkpoint combinations were also determined for comparison. Based on Eq. (2), the MI score for the first-order MI was 0.29 (the denominator 404,550 corresponds to the binominal coefficient C_{900}^{2}, which sums up the different ways to select 2 items out of a total of 900).

$$\text{First order MI}=\frac{\sum_{i=0}^{900}\sum_{j=i+1}^{900}\text{MIscore}\left(\text{Checkpoint}_i,\text{Checkpoint}_j\right)}{404550}. \quad (2)$$

Subsequently, the second-order MI was defined as the average MI score among pairs of training set-specific ensembles, representing a peak level of indistinguishability. The average MI score was 1.02 (all MI scores had P<1×10^{-5}), as derived from Eq. (3). As a control experiment, the MI score, PCC, MAD and covariance of two normally distributed random arrays (with μ=0.0, σ=3.0, to emulate the scale of age estimation residuals) of equivalent test set size were simulated 100,000 times. These results are presented in Figure 1D.

$$\text{Second order MI}=\frac{\sum_{i=0}^{9}\sum_{j=i+1}^{9}\text{MIscore}\left(\text{Ensemble}_i,\text{Ensemble}_j\right)}{36}. \quad (3)$$

The discrepancy between the first and second orders of MI highlights a reduction in estimate fluctuation beyond accuracy enhancement by model ensembling. The combined strengths of PCC (low variance) and MAD (high contrast) validate our MI score formulation. The MI score provides a more comprehensive perspective than either PCC or MAD alone, capturing both trend and magnitude differences.

Coordinate-wise monotonic transformations preserve age-related 3D facial features

Given the convenience of using non-registered point cloud data, we explored the sufficiency of different data fields for age estimation with 3D face point cloud. We trained FPCT-age with data fields or their combinations with the same model architecture, adjusting only the input channel. In this approach, the model was determined only by the input data fields. Therefore, we used the capital letters of the data fields or their combinations to denote the type of input and the specific FPCT-age models trained on that input.

The MAE values and pairwise MI scores of these data field-specific models are presented in Figure 2A. As expected, removing vertices from the facial data greatly undermined the effectiveness of the N and C models. The MI scores between models with and without the N field were relatively large, and the VNC and the VC models had the largest MI score (0.616), falling between the first and second order MI. The inclusion of the C field significantly improved model performance, and the MI scores were relatively low between models with and without the C field, suggesting a strong complementary role of color information against normal vector. By contrast, including the N field led to a smaller improvement, indicating that a portion of the normal vector information has been implicitly contained in the V field. This is not unexpected, as on a differentiable surface, normal vectors can be deduced from coordinates mathematically.

We sought to examine the information complementarity between the N and C fields using multiple linear regression. Specifically, we regressed the estimation residuals of the VN and VC models against that of the VNC model. Table 1 summarizes the regression statistics, which demonstrates a highly linear regression without collinearity. The MI score indicated indistinguishability between the VNC model and the linear combination of the VN and VC models, suggesting little nonlinear interaction existed between colors and normal vectors.

We then explored the possibility of eliminating geometric measures and extracting the framework information from the V field. We hypothesized that as a framework, it would be necessary to maintain the relative positioning of points but not their absolute locations. Since the term "relativeness" was ambiguous, we developed an idea of using coordinate-wise ranks to replace the original Euclidean distance, such that the order of vertices in each dimension remained unchanged. Specifically, we ranked the original vertex coordinates along each axis (X, Y, and Z) and replaced the coordinate components with their corresponding ranks in each axis. We then normalized the rank values and used them as the new magnitude-free, relativeness-preserved coordinates. Our literature search revealed a similar existing operation in mathematics called "rank transformation" for the 1D case (Conover and Iman, 1981; Headrick and Sawilowsky, 2000; Saltelli and Sobol, 1995). Therefore, we adopted this convention and named our transformation algorithm CRT. Figure 2B illustrates a 2D version of CRT. It is important to note that mathematically, CRT is monotonic and irreversible, although vertex correspondence could be saved and tracked for individual faces.

To examine the importance of relative positioning against Euclidean magnitudes for age estimation with 3D face point cloud, we incorporated CRT into the data loading module immediately following the rotation augmentation for training, and labeled the transformed vertices as the rV (relative vertex) field. The operational sequence increased the variability of transformation results, as the ranking of coordinates often changed significantly after large-angle rotations. Figure 2C provides visual examples of the transformation effect. We evaluated models trained on the rV field alone, or in conjunction with the N and/or C fields. These models' MAE values and MI scores to Euclidean counterparts are summarized in Figure 2D. Remarkably, despite the pronounced visual changes due to CRT, the age estimates from these models were comparably accurate and indistinguishable from the outputs of their corresponding

https://doi.org/10.1007/s11427-023-2518-8

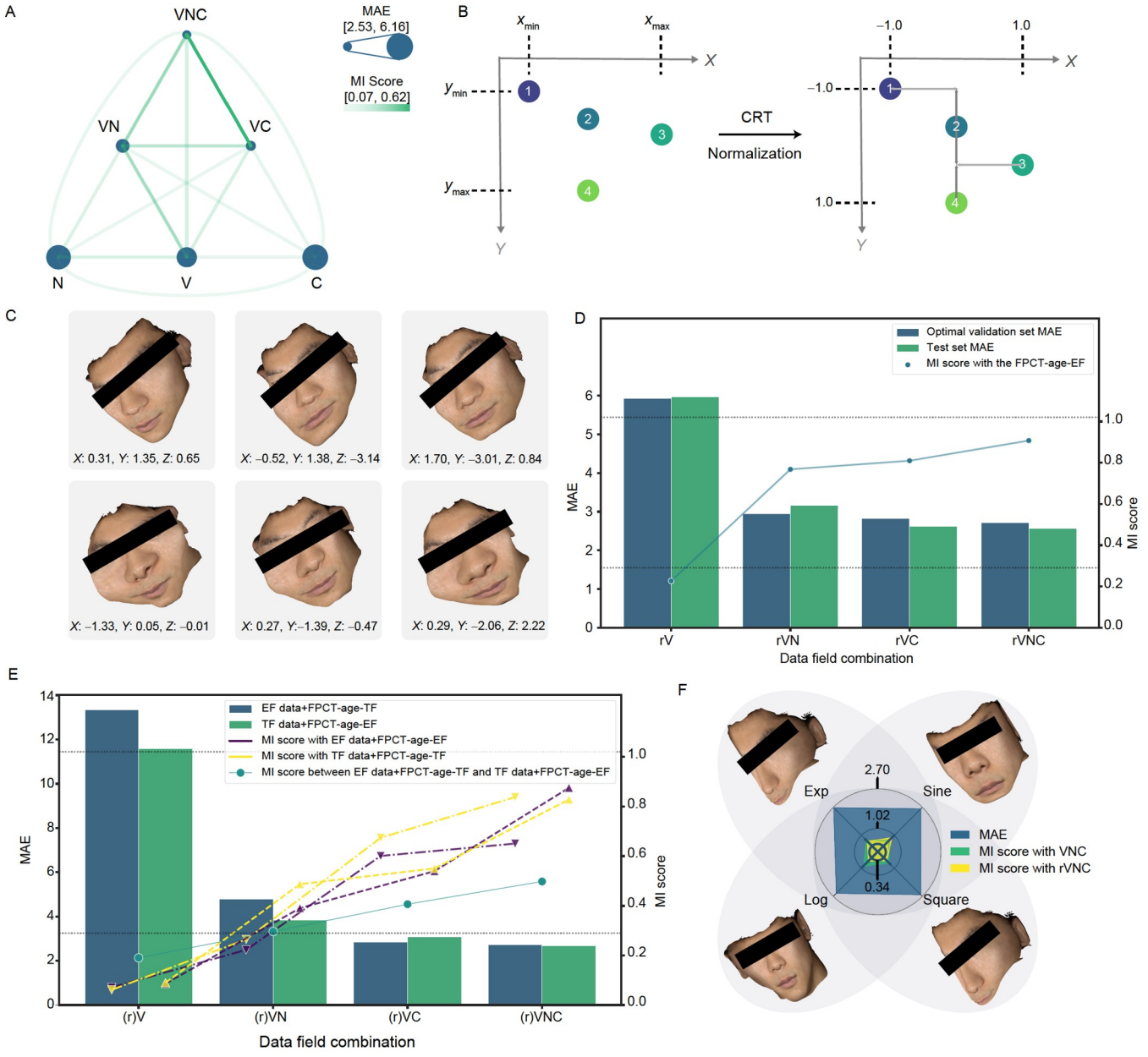

Figure 2. Modeling on data fields and transformed data. A, Hierarchical view of data field-specific model evaluation results. The dot size of a vertex is proportional to the MAE value and rescaled to a parameter value range of [50, 500]. The transparency (alpha parameter) of an edge is proportional to the squared MI score and rescaled to a parameter value range of [0.05, 1.0]. B, Illustration of a 2D CRT. Points 1, 2, 3, 4 originally had irregular ascending y-coordinates, which were transformed and rescaled to an arithmetic progression from −1 to 1 by CRT and normalization. Points 2 and 4 originally shared the same x-coordinate, and they continued to do so after CRT and normalization. C, Examples of face shape distortion. The same face underwent random spatial rotations and then CRT. Transformed faces are shown with a reverse rotation for the convenience of comparison. The corresponding arc values for each random rotation are shown below the face. D, The impact of CRT on age estimation. The bars represent MAE values based on the left-hand y-axis, and the dots connected by lines represent MI scores based on the right-hand y-axis. All MI scores have P-values$<1\times10^{-5}$. E, Cross-model inference on Euclidean and Transformed Data. The FPCT-age model trained on CRT-treated faces (FPCT-age-TF) was used to estimate ages of Euclidean faces (EF data), while the FPCT-age model trained on Euclidean faces (FPCT-age-EF) estimated ages for CRT-treated faces (TF data). The down triangle-connected line represents MI scores for the EF data+FPCT-age-TF group, the up triangle-connected line for the TF data+FPCT-age-EF group, and the circle-connected line compares MI scores between the two groups. In D and E, two horizontal dotted lines mark the first- and second-order MI. F, The impact of CMTs on age estimation. Four elementary monotonic functions (exponential, logarithmic, square, and sine) were applied immediately after CRT and normalized to [−1, 1] in a coordinate-wise manner. The three circles on the radar plot correspond to the first-order MI, the second-order MI and the MAE of exp(CRT) model (the largest among the four), from the inside out. All MI scores have P-values$<1\times10^{-5}$ and fall between the first- and second-order MI, and all MAEs are no greater than 2.70.

Euclidean models.

Additionally, we designed an experiment to assess the interchangeability of Euclidean and transformed data/models, to further elucidate the commonality in age-related features across both face forms. Specifically, we swapped the face form for FPCT-age training and inference and determined MAE and MI scores in this setting. The (r)VNC and (r)VC models exhibited consistent mappings across face forms with minimal differences. In contrast, (r)V models required strict face form matching, while (r)VN models presented an intermediate scenario between the

Table 1. Statistical conclusions: regression of age estimate residuals (VN and VC models on VNC model)

Main statisitcs	Value and interpretation
R^2	0.80. The VNC model estimate residuals can be well explained by the linear combination of the VN and VC model estimate residuals.
MAD	1.21.
MI score	0.74. Calculated using R over MAD, with $q<1\times10^{-5}$.
F-statistics	3,232.0. The regression is highly linear.
Coefficient of VN	0.22. Contribution of VN model estimate residuals to the linear combination.
Coefficient of VC	0.66. Contribution of the VC model estimate residuals to the linear combination.
Durbin-Waston statistics	1.977. No autocorrelation.
Condition number	5.03. No multicollinearity.

two. These observations are encapsulated in Figure 2E. Synthesizing these findings, we propose that the relative positioning of facial information, with emphasis on color information, is the underlying low-level biomarker of age.

To further confirm the dispensability of exact face shape for age estimation, we conducted additional experiments by applying multiple subsequent elementary monotonic transformations following CRT on the input to generate further distorted face shapes. These composite transformations were collectively referred to as CMTs, with CRT being a special case of CMT. Figure 2F presents visual examples for each of these composite CMTs and summarizes the model MAE values and MI scores. As expected, the resulting models achieved comparable accuracy and were highly indistinguishable from both the rVNC and VNC models, which led us to a hypothesis that any CMT applied to 3D faces should preserve age-related features. We emphasize that the composite transformation log(CRT) made FPCT-age more responsive to the transformed facial data, as evidenced by a lower MAE compared with that of the VNC model, suggesting that rational transformation design may improve estimation accuracy even for deep learning architectures. A software implementing CMTs with interactive visualization, referred to as VisualCMT, is provided (see Figure S5 in Supporting Information).

Coordinate-wise rank transformation inspires a facial data masking strategy

The resulting visual changes suggested that CMTs might enable facial data masking while preserving information of interest. We used CRT as an example to test our hypothesis that face shape distortions resulting from CMTs might challenge both human visual perception and computational face verification systems.

To measure human visual perception of the CRT-treated images, we conducted a test in which participants completed 3 types of tasks for 46 3D facial images covering an age range from 25 to 70: sex estimation, age estimation, and face matching. Participants' answers and response time were recorded. A random half of the subjects' 3D faces underwent CRT (see Materials and Methods). To improve convenience and accessibility, we programmed a graphical user interface (GUI) as illustrated in Figure 3A, for participants to complete the test on their mobile phones. We use Euclidean face (EF) to denote untransformed faces and transformed face (TF) to denote CRT-treated faces for brevity.

Results from 21 participants showed that CRT caused significant estimation compromise. Specifically, sex estimation accuracy dropped by 25.8% (t-test P-value$<1\times10^{-16}$), identification accuracy dropped by 58.1% ($P<1\times10^{-24}$), and age estimation error rose by 15.0% (P=0.0012). Figure 3B and C show the accuracy results calculated for individual 3D faces. It is worth noting that perception difficulty varied widely among both EFs and TFs. We saw a 22.0% increase in sex estimation time ($P<1\times10^{-12}$), a 2.3% increase in age estimation time (P=0.38), and a 53.8% increase in identification time ($P<1\times10^{-23}$), as shown in Figure 3D. We conclude that TFs are considerably more difficult for human to perceive, especially in identification tasks.

To investigate the impact of CRT on computational face verification systems, we incorporated the ArcFace loss module (Deng et al., 2019) into the PCT architecture to establish a minimal 3D face point cloud verification model, which we referred to as FPCT-ID. For this task, we assembled training, gallery and probe sets using data subjects collected over multiple years (see Materials and Methods). Transfer learning from trained FPCT-age models was used to offset the limited amount of available data for this task. For evaluation, we derived the embedded facial features (256-dimensional vectors) from the gallery and probe sets, and constructed receiver operating characteristic (ROC) curves. To do this, we scanned the separation threshold based on cosine distances between the feature vectors of the gallery and probed facial data. The procedures are demonstrated in Figure 4A. We observed diminishing returns in the area under curve (AUC) over 2,000 epochs of training, as shown in the subpanel of Figure 4B, and decided to evaluate AUC using the model checkpoint at the 2,000th epoch. Interestingly, in contrast to the age estimation task, we found that removing the N and/or C field had only limited consequence on face verification performance, as shown in Figure 4B, suggestive of the differential facial characteristics related to age and identity.

With a working face verification model, we investigated the impact of face form on models' response. Initially, when we modeled rVNC data for both training and inference, the accuracy was surprisingly better, as can be seen by comparing the leftmost and rightmost bars in the 4th group of bars in Figure 4C. We then decided to conduct exhaustive experiments with all possible face form combinations. Specifically, we trained FPCT-ID models on either EFs or TFs (the resulting models were denoted as FPCT-ID-EF and FPCT-ID-TF, respectively) and probed the test set facial features (inferred from either EFs or TFs) against the gallery set facial features (inferred from either EFs or TFs). Given the two training options and the two inference options for each set, this resulted in a total of eight possible ROC curves for each data field-specific model (see Figure 4A for data flow). We found that alternating between VNC and rVNC input only caused a negligible compromise in accuracy, contrary to what was observed in the visual perception test. When we experimented with (r)VN and (r)VC data, we observed larger but not highly discriminative gaps in verification accuracy for mixed face forms. Only the V and rV models were substantially confused by input faces in the opposite form, as evidenced by a marked drop in accuracy when the face forms were not the same throughout. The AUC values of all combinatorial inferences are summarized in Figure 4C. Under conventional practices, data is typically split

https://doi.org/10.1007/s11427-023-2518-8

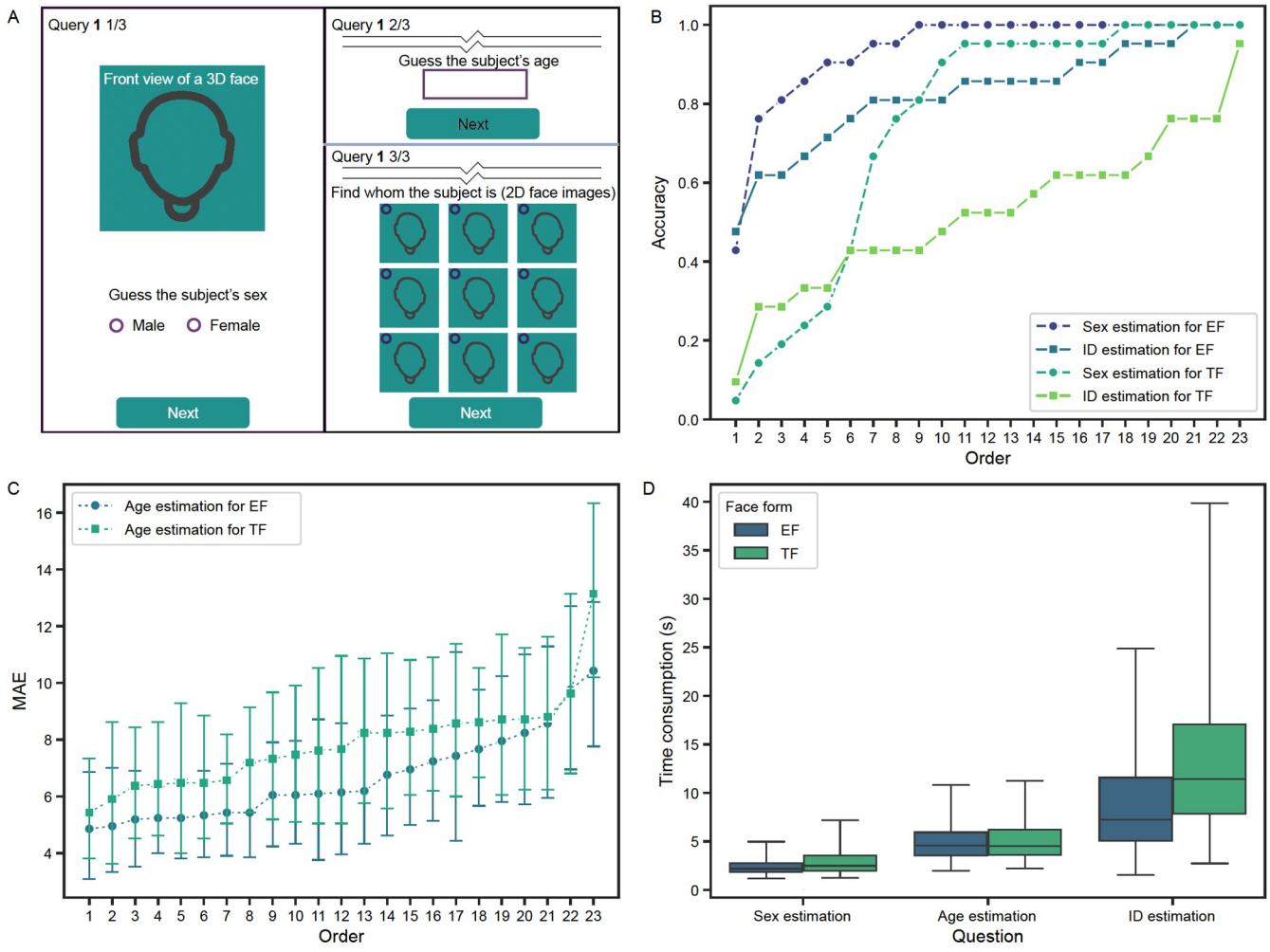

Figure 3. Visual perception test. A, Illustration of the test GUI. Each subject spans three pages to allow for sequential answering on sex estimation (binary-choice), age estimation (fill-in-the-blank) and identification (multiple-choice), while the subject's face remains in the same position throughout. B, Sex estimation and identification accuracy comparison, and C, Age estimation error comparison. Results were calculated for individual 3D faces, categorized by face form, and sorted within the same face form. D, Response time comparison. Data points more than 1.5 times the interquartile range above the third quartile or below the first quartile per type of task have been excluded for statistics and visualization.

into separate training, gallery and probe sets, for training face verification models and evaluating their performance. However, given the limited sample size available for this task, we opted to use the training data directly as the gallery to maximize the available training data. To validate our approach, we provided evidence that using the training set as the gallery did not introduce an overfitting risk. While there was a noticeable decrease in FPCT-ID's performance with reduced training data, the overall pattern of performance variation based on face form combinations remained consistent (Figure S6 in Supporting Information).

We posited that when color or normal vector information was available, the relative positioning of the information was sufficient for identification, similar to age estimation. To examine this hypothesis, we used the FPCT-ID-EF model to make inference on test data in both face forms, and collected facial feature vectors for dimension reduction using non-linear (UMAP and t-SNE, see Figure 4D and Figure S7B in SupportingInformation, respectively) or linear (PCA, Figure S7A in Supporting Information) algorithms. When the N and/or C data fields were available alongside the (r)V field, the extracted facial features from EFs and TFs were mostly inseparable. However, when learning was based solely on V/rV data, the inference was sensitive to face form.

Based on these observations CRT and other CMTs have the potential to be effective facial data masking techniques that obscure identity for both human visual perception and face verification systems, while preserving ID-independent information, such as age. To enhance 3D facial data protection, we propose a general guideline that features CMTs-based data transformation, as illustrated in Figure 5. Specifically, we suggest that for face recognition applications, only the V field of face point cloud data should be distributed and used to train models. For other face-based applications that have been confirmed to rely minimally on Euclidean face shape, facial data with transformed shape should be distributed. The entire guideline requires regular audits to ensure that the data repository and provisioning adheres to regulations. By streamlining the management of face datasets, this guideline is potential to accelerate advancements in both research and industry.

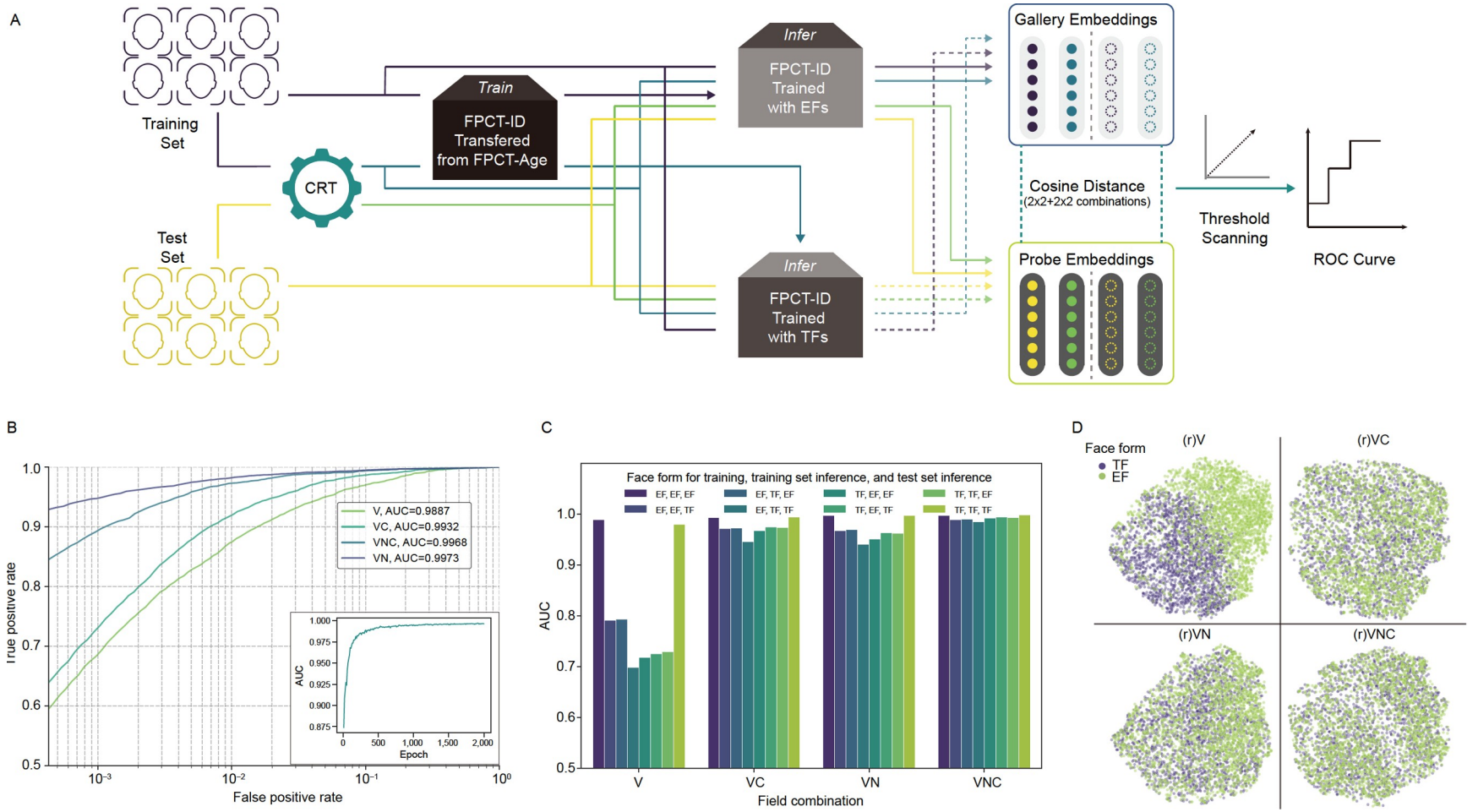

Figure 4. Computational experiments for face verification. A, Illustration of the model evaluation scheme. FPCT-ID was trained with either EFs or TFs, and inferences were made on training and test data in the form of either EF or TF. Dark blue and cyan lines indicate the flow of training data in the form of EF and TF, while yellow and green lines represent the flow of test data in the form of EF and TF. Translucent lines indicate inference by FPCT-ID-EF, and solid circles indicate the inferred facial features. Dashed lines indicate inference by FPCT-ID-TF, and hollow dashed circles indicate the inferred facial features. Arrows indicate the completion of training or inference. B, ROC curves for data field-specific FPCT-ID-EF models. Each curve corresponds to a model checkpoint at the 2,000th epoch. The *x*-axis is in logarithmic scale, and deciles between every two adjacent major *x*-ticks are shown in dashed vertical lines, following the convention of the ArcFace paper (Deng et al., 2019). The curves are truncated at a false positive rate equal to 1/2,333, where the mathematical expectation of a false positive is one person. The subpanel displays the AUC escalation during FPCT-ID model training on VNC data. C, Verification accuracies based on face form combinations. Model checkpoints at the 2,000th epoch are used. D, Facial feature embedding by UMAP. The input facial feature vectors are 256-dimensional. Each dot represents one 3D face. Transparency (alpha parameter) was set at 0.4 for better visualization.

https://doi.org/10.1007/s11427-023-2518-8

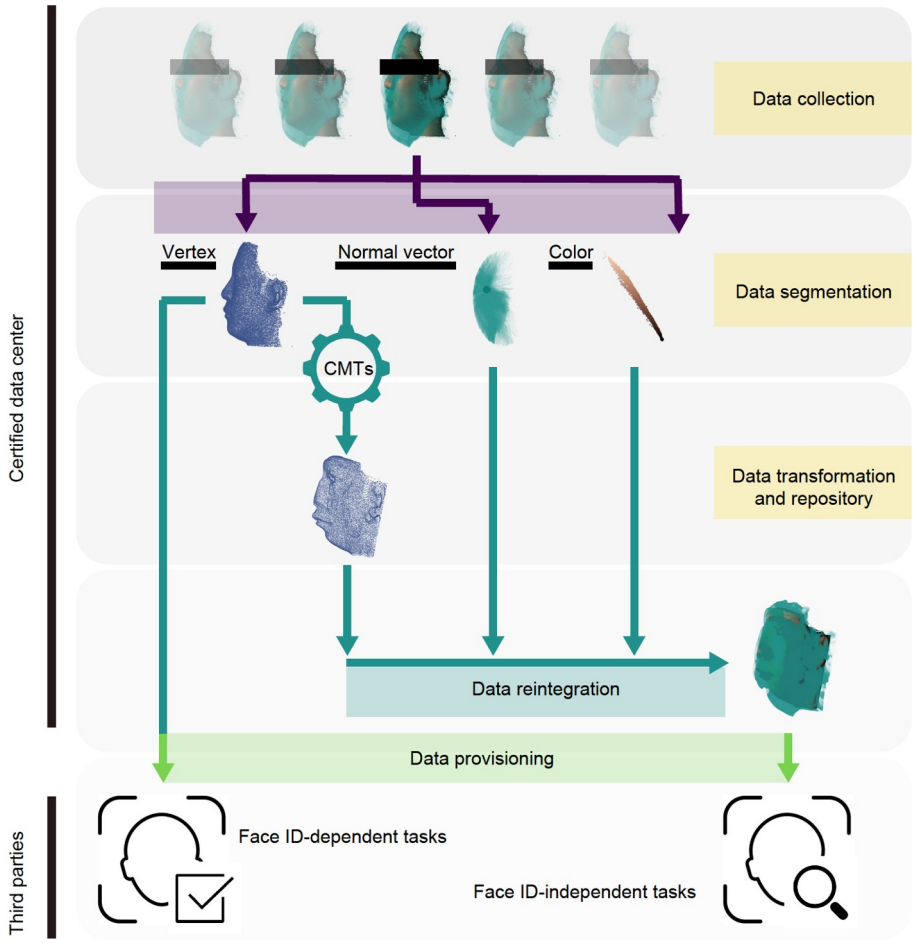

Figure 5. 3D facial data protection guideline. Data collection: ethical approval should be obtained, and informed written consent should be acquired. Data segmentation: this can be automatically executed through embedded camera software or by segmenting fields from the raw data at the certified data center. Data transformation and repository: the Euclidean vertices, transformed vertices, normal vector coordinates, and color values should be stored separately, each secured with its own encryption key. Data provisioning: for requests dependent on facial ID, the Euclidean vertices should be distributed. For requests independent of facial ID, the transformed vertices and other necessary data should be distributed.

DISCUSSION

In this study, our initial objective was to estimate age using 3D face point cloud data and explore the impact of data fields on modeling. However, we made a surprising observation that a rigid Euclidean description of the human face was not necessary for accurate age and identity estimation. Based on these discoveries, we proposed a 3D facial data protection guideline. Modeling directly on 3D face point cloud data was not a common practice to date. However, we showed that this data structure did not require extensive preprocessing to be compatible with deep regression models. This allows for easy deployment of such models to digital applications and hardware. We developed CRT and CMT algorithms to distort face shape. These algorithms are easy to implement and could hinder both human visual perception and computational face verification systems. Our findings fill a gap in the urgently needed 3D facial data masking techniques. We have verified that it is possible to preserve age-related features while removing identity features. The development of similar techniques has the potential to facilitate the sharing of facial data for research. Additionally, we highlight the theoretical importance of our discoveries that the relative positioning of facial features can be informative enough for multiple tasks.

In our prior study (Xia et al., 2020), we were constrained to reshape the original facial structures to a square ratio due to the limitations of classical convolutional neural networks. Notably, even with this reshaping, the accuracy of age estimation remained high. In hindsight, this can be attributed to the fact that the squaring process is a special instance of CMTs, which does not compromise age-related features. Furthermore, CMTs can be readily adapted to 2D imagery. To achieve this, one can map a 2D image to 3D space using reconstruction techniques or by simply assigning a constant Z coordinate to all facial pixels. After random rotation and CMT treatment, the data can be reoriented and projected back to 2D. It might be worth exploring whether 2D facial content can potentially retain age-related features when subjected to CMTs.

Currently, CMT methodologies presented in this study for age estimation and data protection with 3D face point cloud is tested only in Asians of ages from 18 to 85 years, further tests will be needed to see whether it is similarly applicable to other ethnic and age groups.

MATERIALS AND METHODS

Ethical approval

The collection, use and ethical review of the facial data used in this study were approved by the Ethical Committees of the Staff Hospital of Jidong oil-field of China National Petroleum Corporation in 2015, and renewed in 2020. Hard copies of signed consents from all subjects were archived and stored in a secure location.

Data preparation

Data collection and data format. This study used anonymized facial data from a Han Chinese cohort collected between 2016 and 2020 (excluding 2017) in Tangshan City, Hebei Province, China. The subjects were registered residents of Tangshan City with low mobility, and data collection was concentrated between June and November of each year. After data cleaning, a total of 16,580 samples were eligible for downstream use, with some subjects enrolled in multiple years.

The 3D facial data were captured using a 3dMDface system (3dMD Inc., USA). Raw data were exported as OBJ-formatted 3D facial information, which included vertex coordinates, normal vectors, texture assignments, and triangular element correspondences, along with paired texture images. The camera system was set up and calibrated in a room with stable lighting. During image capture, subjects were directed to sit at an appropriate height and distance, facing the camera with a neutral expression. They were required to remove any accessories such as glasses or hats, and hair on the forehead was secured with a headband to avoid obstruction. If a subject's eyes were closed or the image capture program failed, the subject was re-shot to ensure data integrity.

Data splits. In the initial subsection of the Results, we employed a 10-fold cross validation approach to train and evaluate FPCT-age. This involved performing a data split of the entire face dataset into ten subsets of nearly equal size, each with a similar age distribution. One of these subsets was reserved as the untouched test set during training. Of the remaining nine subsets, each was used in turn as the validation set, with the other eight combined to form the training set, yielding nine paired training and validation sets. Given the comparable age distributions between males and females, the data were split only based on age, without considering sex. To address imbalances in eleven age labels where data counts ranged between 10 and 20, we categorized data into 24 age groups based on 25 percentiles, ensuring a minimum age span of one year for each group. Data in each age group were then randomly divided into ten subsets. Due to the consistent model performance across different training sets, only the first training, validation, and the test sets were used in subsequent subsections of the Results.

For the FPCT-ID model, we excluded data subjects with fewer than three samples. For each of the remaining data subjects, we randomly chose one sample to be included in the probe, which is a query dataset containing faces to be verified. The other samples were used for training and establishing the gallery, a reference dataset containing faces considered known. As a result, both the training set and the gallery contained 5,164 samples, while the probe comprised 2,333 samples.

Data preprocessing. The 3dMDface system generates point cloud data that originally represents the face as a mesh of triangle elements, with each triangular element naturally a 27-dimensional vector consisting of vertex coordinates, normal vectors, and color assignments of its three constituent vertices. However, in our preliminary experiments, we found that the training and convergence of deep learning architectures were significantly slowed by this high-dimensional input. To address this issue, we applied error-free linear interpolation to compress the channel-wise values of each triangular element onto its mass center, resulting in a 9-dimensional vector. The interpolated point cloud, with mass centers as elements, has a visualization indistinguishable from the raw data and allows for more effective model training.

The working principle of the 3dMDface system introduces randomness in the quantity of captured data points for each face. The resulting data clouds typically contain more than 30,000 triangular elements, which can cause out-of-memory issues for large batch size. To address this issue, we adopted the FPS method, which had been widely used in the field of point cloud modeling since PointNet++ (Qi et al., 2017b). After mass center interpolation, the FPS method uniformly sampled the data in the spatial domain, maximally preserving information from both dense and sparse regions of the original data. We evaluated the visualization effect, model performance and training efficiency with different sampling rates in a preliminary experiment, as shown in Figure S4A–C in Supporting Information, respectively. In this article, we used a sampling rate of 12,288 for all other experiments. Some samples contained neck and collar, which were not filtered out, as these would be treated as random noise by the model and did not impair model performance.

Data augmentation. During training, we applied common data augmentation techniques used in point cloud modeling. First, we applied shuffling by randomly reordering the records of points. Next, we applied rotation by independently performing random rotations along each axis, ranging in $[-\pi, \pi]$, for both vertex coordinates and normal vectors. Additionally, we applied shift by randomly shifting the vertex coordinates along each axis with a random value in the range $[-1, 1]$. Finally, we applied scaling to all dimensions of the vertex coordinates and normal vectors within the range of $[0.8, 1.25]$. No data augmentation was conducted during the inference phase.

Deep learning

Geometric models. We conducted preliminary age estimation experiments using several base models, including PointNet (Qi et al., 2017a), two versions of point transformer (Engel et al., 2021; Zhao et al., 2021), and PCT (Guo et al., 2021a). Through a preliminary grid search on optimizer and learning rate and limiting training to 500 epochs, we selected PCT as the most promising base model for further optimization, modeling and analysis due to its superior performance. PCT is a novel architecture designed for point cloud data and achieves state-of-the-art performance on 3D object databases. It captures global context using symmetric function and attention modules, and captures local context using neighbor embedding modules (Guo et al., 2021a). To balance model performance and time consumption, we set the number of training epochs for FPCT-age to be 1,000.

Average ensembling. We saved model checkpoints (dictionaries storing model parameters values) at every fifth epoch

https://doi.org/10.1007/s11427-023-2518-8

starting from the 500th epoch. The age estimation for each subject was derived by averaging over a total of 100 checkpoints (for each training set) as the ensemble result.

ArcFace loss function. Face recognition is a challenging classification task with a large number of classes, which makes it difficult for traditional classification loss functions to perform well. Improving the loss function for such large-scale classification tasks is a critical research challenge. The Additive Angular Margin Loss (ArcFace) is a novel loss function proposed to enhance the discriminative power of feature embeddings. It uses geodesic distance margins in the softmax loss function to maximize class separability while promoting intra-class compactness and inter-class discrepancy. This approach has achieved state-of-the-art results in comprehensive face recognition experiments (Deng et al., 2019). To balance model performance and time consumption, we set the number of training epochs for FPCT-ID to be 2,000.

Hyperparameter grid search. The geometric models were initially optimized for object point clouds. However, due to the subtler spatial distribution of points in face point clouds compared with object point clouds, we conducted a comprehensive grid search on important hyperparameters for our face dataset. For efficiency, we limited the number of training epochs to be 500, and trained and evaluated the models using the first pair of training and validation sets.

(1) Optimizer and initial learning rate. We evaluated learning rates of 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, and 1×10^{-2}, and optimizers including stochastic gradient descent (SGD, with momentum 0.9) and Adam (with beta1=0.9 and beta2=0.99). Adam optimizer with an initial learning rate of 1×10^{-4} achieved the best performance for both FPCT-age and FPCT-ID. In contrast, SGD and initial rate of 1×10^{-2} were optimal for ModelNet40 classification task.

(2) Scheduler. We evaluated learning rate schedulers including Step, ReduceOnPlateau, OneCycle (Smith and Topin, 2017), CosineAnnealing (Smith, 2015), and no scheduler. Adam without a scheduler outperformed the complicated schedulers after approximately 300 epochs, and we used it for both FPCT-age and FPCT-ID.

(3) Feature embedding. Two implementations of FPS were incorporated in the shallow layers of PCT, each followed by a K-nearest neighbors (KNN) embedding. We evaluated various combinations of sampling rates and K values. To balance the trade-off between resource consumption and accuracy gain, we selected 1,024 and 512 as the two sampling rates and 32 and 32 as the two K values for FPCT-age, and 512 and 256 as the two sampling rates and 32 and 32 as the two K values for FPCT-ID.

(4) Weight decay. We evaluated weight decay rates of 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 1×10^{-2}, 1×10^{-1} and 0. A weight decay rate of 1×10^{-2} was optimal for both FPCT-age and FPCT-ID.

(5) Batch size and data parallelism. We evaluated batch sizes of 8 and 16, and GPU numbers of 1, 2 and 4. The GPU was Nvidia Quadro GV100 with CUDA 11.2. Using two GPUs, each loaded with a batch size of 16, was optimal for both FPCT-age and FPCT-ID.

Statistics

We used a permutation-based approach to calculate the P-value to evaluate the statistical significance of MI scores. Given numeric arrays a and b, we first calculate the MI score s_0 between them according to Eq. (1). We then apply a number of random permutations on b, each permutation using a different random seed, and calculate the MI score s_k between a and each permuted b. The number of cases where $s_k \geqslant s_0$ was counted and divided by the total number of permutations. We set the number of permutations to be 100,000, as a larger number requires too much computation time. The P-value is denoted as $<1\times10^{-5}$ if there was no observation for $s_k \geqslant s_0$.

Visual perception test

We randomly selected one subject per age from 25 to 70 years old, and implemented CRT on a random half of their 3D facial data. All 3D facial data were displayed from the front view. For the identification task, raw 2D facial images (captured by the 3dMDface system alongside 3D data) of the subject in question and eight other subjects of the same age and sex were displayed. The entire test took approximately 20 min to complete. All test 3D and 2D candidate images were randomly shuffled for each participant to eliminate the influence of changes in participants' attention during the test. Response time was defined as the time interval between the entrance and the exit of each page and was recorded invisibly.

Data and materials availability

The data that support the findings of this study are divided into two groups: shared data and restricted data. Shared data are provided by one of the authors and available from the manuscript and supplementary data. Restricted data relating to subjects in this study are subject to an agreement that requires full confidentiality. Therefore, such data cannot be shared. VNC model checkpoints for age estimation and face verification at the end of training are provided with code. All code on facial data preprocessing and modeling are made publicly available at https://github.com/JackieHanLab/CMTs-3DFace, together with an executable software named VisualCMT for implementing CMTs on face point cloud data. The code and software are provided for non-commercial use.

Compliance and ethics

The author(s) declare that they have no conflict of interest.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (92049302, 92374207, 32088101, 32330017) and the National Key Research and Development Program of China (2020YFA0804000). We thank Dr. Shiqiang Zhang and Haonan Zheng for visualization suggestions, and Dr. Hao Fu for discussion on mathematical modeling.

Supporting information

The supporting information is available online at https://doi.org/10.1007/s11427-023-2518-8. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

Bartol, K., Bojanic, D., Petkovic, T., and Pribanic, T. (2021). A review of body measurement using 3D scanning. IEEE Access 9, 67281–67301.

Bell, C.G., Lowe, R., Adams, P.D., Baccarelli, A.A., Beck, S., Bell, J.T., Christensen, B.C., Gladyshev, V.N., Heijmans, B.T., Horvath, S., et al. (2019). DNA methylation aging clocks: challenges and recommendations. Genome Biol 20, 249.

Boulkenafet, Z., Komulainen, J., and Hadid, A. (2016). Face spoofing detection using colour texture analysis. IEEE Trans Inform Forensic Secur 11, 1818–1830.

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C.A., Jia, H., Travers, A., Zhang, B., Lie, D., and Papernot, N. (2021). Machine unlearning. In: 2021 IEEE

Symposium on Security and Privacy (SP). San Francisco. 141–159.
Chanda, A., and Chatterjee, S. (2021). Predicting Obesity Using Facial Pictures during COVID-19 Pandemic. Biomed Res Int 2021, 6696357.
Chen, W., Qian, W., Wu, G., Chen, W., Xian, B., Chen, X., Cao, Y., Green, C.D., Zhao, F., Tang, K., et al. (2015). Three-dimensional human facial morphologies as robust aging markers. Cell Res 25, 574–587.
Conover, W.J., and Iman, R.L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat 35, 124–129.
Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). ArcFace: additive angular margin loss for deep face recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach. 4685–4694.
Eng, Z.H.D., Yick, Y.Y., Guo, Y., Xu, H., Reiner, M., Cham, T.J., and Chen, S.H.A. (2017). 3D faces are recognized more accurately and faster than 2D faces, but with similar inversion effects. Vision Res 138, 78–85.
Engel, N., Belagiannis, V., and Dietmayer, K. (2021). Point transformer. IEEE Access 9, 134826–134840.
Galkin, F., Mamoshina, P., Aliper, A., Putin, E., Moskalev, V., Gladyshev, V.N., and Zhavoronkov, A. (2020). Human gut microbiome aging clock based on taxonomic profiling and deep learning. iScience 23, 101199.
Geng, X., Wang, Q., and Xia, Y. (2014). Facial age estimation by adaptive label distribution learning. In: 2014 22nd International Conference on Pattern Recognition. Stockholm. 4465–4470.
Geng, X., Zhou, Z.H., and Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. IEEE Trans Pattern Anal Mach Intell 29, 2234–2240.
Gou, J., Yu, B., Maybank, S.J., and Tao, D. (2021). Knowledge distillation: a survey. Int J Comput Vis 129, 1789–1819.
Guo, C., Goldstein, T., Hannun, A., and Maaten, L.V.D. (2020). Certified data removal from machine learning models. In: Proceedings of the 37th International Conference on Machine Learning. JMLR.org.
Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., and Hu, S.M. (2021a). PCT: point cloud transformer. Comp Visual Media 7, 187–199.
Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. (2021b). Deep learning for 3D point clouds: a survey. IEEE Trans Pattern Anal Mach Intell 43, 4338–4364.
Headrick, T.C., and Sawilowsky, S.S. (2000). Properties of the rank transformation in factorial analysis of covariance. Commun Stat Simul Comput 29, 1059–1087.
Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. arXiv, 1503.02531.
Hsieh, T.C., Bar-Haim, A., Moosa, S., Ehmke, N., Gripp, K.W., Pantel, J.T., Danyel, M., Mensah, M.A., Horn, D., Rosnev, S., et al. (2022). GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet 54, 349–357.
Ilia, P., Polakis, I., Athanasopoulos, E., Maggi, F., and Ioannidis, S. (2015). Face/off: preventing privacy leakage from photos in social networks. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: Association for Computing Machinery.
Imai, T., and Okami, K. (2019). Facial cues to age perception using three-dimensional analysis. PLoS ONE 14, e0209639.
Joshi, R.S., Rigau, M., García-Prieto, C.A., Castro de Moura, M., Piñeyro, D., Moran, S., Davalos, V., Carrión, P., Ferrando-Bernal, M., Olalde, I., et al. (2022). Look-alike humans identified by facial recognition algorithms show genetic similarities. Cell Rep 40, 111257.
Kaur, H., and Khanna, P. (2016). Biometric template protection using cancelable biometrics and visual cryptography techniques. Multimed Tools Appl 75, 16333–16361.
Kersten, T.P., Lindstaedt, M., and Starosta, D. (2018). Comparative geometrical accuracy investigations of hand-held 3D scanning systems—an update. Int Arch Photogramm Remote Sens Spat Inf Sci XLII-2, 487–494.
Ko, K., Gwak, H., Thoummala, N., Kwon, H., and Kim, S. (2021). SqueezeFace: integrative face recognition methods with LiDAR sensors. J Sensors 2021, 1–8.
Lanitis, A., Draganova, C., and Christodoulou, C. (2004). Comparing different classifiers for automatic age estimation. IEEE Trans Syst Man Cybern B 34, 621–628.
Lehallier, B., Shokhirev, M.N., Wyss-Coray, T., and Johnson, A.A. (2020). Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. Aging Cell 19, e13256.
Lin, S., Li, Z., Fu, B., Chen, S., Li, X., Wang, Y., Wang, X., Lv, B., Xu, B., Song, X., et al. (2020). Feasibility of using deep learning to detect coronary artery disease based on facial photo. Eur Heart J 41, 4400–4411.
Liu, Y., Yan, W., and Hu, B. (2021). Resistance to facial recognition payment in China: The influence of privacy-related factors. TeleCommun Policy 45, 102155.
Lopez-Paz, D., Bottou, L., Schölkopf, B., and Vapnik, V. (2015). Unifying distillation and privileged information, arXiv, 1511.03643.
Lübbers, H.T., Medinger, L., Kruse, A., Grätz, K.W., and Matthews, F. (2010). Precision and accuracy of the 3dMD photogrammetric system in craniomaxillofacial application. J Craniofac Surg 21, 763–767.
Matthews, H., Penington, A., Clement, J., Kilpatrick, N., Fan, Y., and Claes, P. (2018). Estimating age and synthesising growth in children and adolescents using 3D facial prototypes. Forensic Sci Int 286, 61–69.
McPherson, R., Shokri, R., and Shmatikov, V. (2016). Defeating image obfuscation with deep learning. arXiv, 1609.00408.
Meyer, D.H., and Schumacher, B. (2021). BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. Aging Cell 20, e13320.
Mu, G., Huang, D., Hu, G., Sun, J., and Wang, Y. (2019). Led3D: a lightweight and efficient deep approach to recognizing low-quality 3D faces. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach. 5766–5775.
Nandakumar, K., and Jain, A.K. (2009). Soft biometrics. In: Li, S.Z., and Jain, A., eds. Encyclopedia of Biometrics. Boston: Springer US. 1235–1239.
Neel, S., Roth, A., and Sharifi-Malvajerdi, S. (2021). Descent-to-delete: Gradient-based methods for machine unlearning. In: Proceedings of the 32nd International Conference on Algorithmic Learning Theory. PMLR. 931–962.
Nguyen, D.P., Ho Ba Tho, M.C., and Dao, T.T. (2021). Enhanced facial expression recognition using 3D point sets and geometric deep learning. Med Biol Eng Comput 59, 1235–1244.
Niinuma, K., Park, U., and Jain, A.K. (2010). Soft biometric traits for continuous user authentication. IEEE Trans Inform Forensic Secur 5, 771–780.
Parkhi, O., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In: Proceedings of the British Machine Vision Conference (BMVC). Durham: British Machine Vision Association.
Patel, V.M., Ratha, N.K., and Chellappa, R. (2015). Cancelable biometrics: a review. IEEE Signal Process Mag 32, 54–65.
Qi, C.R., Su, H., Mo, K.C., and Guibas, L.J. (2017a). PointNet: deep learning on point sets for 3D classification and segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu. 77–85.
Qi, C.R., Yi, L., Su, H., and Guibas, L.J. (2017b). PointNet plus plus: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Red Hook: Curran Associates Inc. 5105–5114.
Rothe, R., Timofte, R., and Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. Int J Comput Vis 126, 144–157.
Roussi, A. (2020). Resisting the rise of facial recognition. Nature 587, 350–353.
Saltelli, A., and Sobol, I.M. (1995). About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 50, 225–239.
Shan, S., Wenger, E., Zhang, J., Li, H., Zheng, H., and Zhao, B.Y. (2020). Fawkes: protecting privacy against unauthorized deep learning models. arXiv, 2002.08327.
Smith, L.N. (2015). Cyclical learning rates for training neural networks. arXiv, 1506.01186.
Smith, L.N., and Topin, N. (2017). Super-convergence: very fast training of neural networks using large learning rates. arXiv, 1708.07120.
Smith, M., and Miller, S. (2022). The ethical application of biometric facial recognition technology. AI Soc 37, 167–175.
Sundararajan, A., Sarwat, A.I., and Pons, A. (2019). A survey on modality characteristics, performance evaluation metrics, and security for traditional and wearable biometric systems. ACM Comput Surv 52, 1–36.
Tan, Z., Yang, Y., Wan, J., Guo, G., and Li, S.Z. (2019). Deeply-learned hybrid representations for facial age estimation. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press. 3548–3554.
The Lancet Healthy Longevity (2022). Is ageing a disease? Lancet Healthy Longev 3, e448.
Vaiserman, A., and Krasnienkov, D. (2020). Telomere length as a marker of biological age: state-of-the-art, open issues, and future perspectives. Front Genet 11, 630186.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc. 6000–6010.
Windhager, S., Mitteroecker, P., Rupić, I., Lauc, T., Polašek, O., and Schaefer, K. (2019). Facial aging trajectories: a common shape pattern in male and female faces is disrupted after menopause. Am J Phys Anthropol 169, 678–688.
World Health Organization. (2019). MG2A ageing associated decline in intrinsic capacity. International Statistical Classification of Diseases and Related Health Problems. 11th ed. Geneva: World Health Organization.
Xia, B., Amor, B.B., Huang, D., Daoudi, M., Wang, Y., and Drira, H. (2013). Enhancing gender classification by combining 3D and 2D face modalities. In: 21st European Signal Processing Conference (EUSIPCO 2013). Marrakech. 1–5.
Xia, B., Ben Amor, B., and Daoudi, M. (2017). Joint gender, ethnicity and age estimation from 3D faces. Image Vision Comput 64, 90–102.
Xia, X., Chen, X., Wu, G., Li, F., Wang, Y., Chen, Y., Chen, M., Wang, X., Chen, W., Xian, B., et al. (2020). Three-dimensional facial-image analysis to predict

heterogeneity of the human ageing rate and the impact of lifestyle. Nat Metab 2, 946–957.
Xu, C., Liu, Q., and Ye, M. (2017). Age invariant face recognition and retrieval by coupled auto-encoder networks. Neurocomputing 222, 62–71.
Yang, K., Yau, J.H., Li, F., Deng, J., and Russakovsky, O. (2022a). A study of face obfuscation in ImageNet. In: Proceedings of the 39th International Conference on Machine Learning. Baltimore. 25313–25330.
Yang, Y., Lyu, J., Wang, R., Wen, Q., Zhao, L., Chen, W., Bi, S., Meng, J., Mao, K., Xiao, Y., et al. (2022b). A digital mask to safeguard patient privacy. Nat Med 28, 1883–1892.
Zhang, Z., Da, F., and Yu, Y. (2019). Data-free point cloud network for 3D face recognition. arXiv, 1911.04731.
Zhao, H., Jiang, L., Jia, J., Torr, P.H., and Koltun, V. (2021). Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE. 16259–16268.
Zhu, Z., Shi, D., Guankai, P., Tan, Z., Shang, X., Hu, W., Liao, H., Zhang, X., Huang, Y., Yu, H., et al. (2023). Retinal age gap as a predictive biomarker for mortality risk. Br J Ophthalmol 107, 547–554.