# PERFACE: METRIC LEARNING IN PERCEPTUAL FACIAL SIMILARITY FOR ENHANCED FACE ANONYMIZATION

*Haruka Kumagai[1], Leslie Wöhler[1], Satoshi Ikehata[2,1], Kiyoharu Aizawa[1]*

The University of Tokyo[1] NII[2]

## ABSTRACT

In response to rising societal awareness of privacy concerns, face anonymization techniques have advanced, including the emergence of face-swapping methods that replace one identity with another. Achieving a balance between anonymity and naturalness in face swapping requires careful selection of identities: overly similar faces compromise anonymity, while dissimilar ones reduce naturalness. Existing models, however, focus on binary identity classification "the same person or not", making it difficult to measure nuanced similarities such as "completely different" versus "highly similar but different." This paper proposes a human-perception-based face similarity metric, creating a dataset of 6,400 triplet annotations and metric learning to predict the similarity. Experimental results demonstrate significant improvements in both face similarity prediction and attribute-based face classification tasks over existing methods. Our dataset is available at `https://github.com/kumanotanin/PerFace`.

***Index Terms***— face anonymization, face similarity, face swapping, human perception

## 1. INTRODUCTION

Face anonymization is a technique used to protect individual privacy in facial images while maintaining the usability of the underlying information. Among the various approaches such as occlusion, blurring and face swapping—replacing the target's facial identity with the source's while preserving target's identity-irrelevant attributes (e.g., pose, expression, or background) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]—is considered an effective anonymization method. Unlike traditional methods such as occlusion or blurring, which compromise image quality and utility while effectively concealing the identity of the original person, face swapping ensures realism and coherence by preserving the facial structure in images.

When employing face swapping for anonymization, it is crucial that the source and the target share key attributes such as gender and age [4] to avoid unnatural outcomes (See Fig. 2). To address this, a recent study [12] leveraged distances between embedded features from pretrained face recognition models (e.g., ArcFace [13]). Their algorithm identifies the most distant face in latent space that nonetheless retains similar attributes (e.g., age, gender), thereby ensuring both anonymity and a natural appearance.

However, leveraging pretrained face recognition models [14, 15, 16, 17, 18, 19, 13, 20, 21, 22, 11, 23] for facial anonymization poses two major challenges. First, these models were optimized via metric learning to cluster images of the same identity and separate those of different identities—thereby considering all different identities
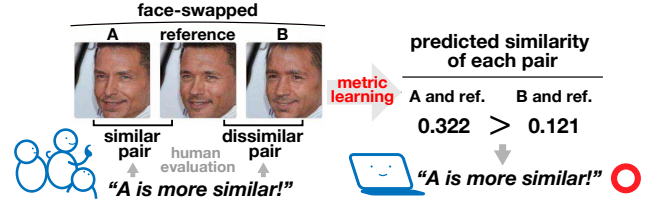
**Fig. 1**: We developed a model that predicts perceptual facial similarity using metric learning, based on our SimCelebA dataset.

as dissimilar, even when they appear perceptually similar. In contrast, face swapping for anonymization necessitates an accurate assessment of similarity between different identities to guarantee that the swapped face is perceptually distinct from the original. Second, existing models were trained exclusively on genuine face images. In face anonymization, it is imperative to evaluate the perceptual distance between the original image and its face-swapped counterpart, rather than merely measuring the distance between the target and source images. This is essential because a face's overall impression is influenced not only by its intrinsic parts (*e.g.*, eyes, noses) but also by factors such as facial contours and hairstyle. Moreover, face-swapped images often contain artifacts and attribute inconsistencies, resulting in a significant domain gap from natural images.

In this work, we present *PerFace*, a feature extractor tailored for evaluating facial similarity. Unlike traditional face recognition models that emphasize identity matching, PerFace is trained through metric learning on human-annotated similarity assessments specifically derived from *face-swapped images*. To address the challenge of directly quantifying similarity, we introduced a pairwise comparison task in which annotators select the face-swapped image that most closely resembles a reference face-swapped image (Fig. 1). Building on PerFace, we propose a comprehensive face anonymization framework that leverages these refined features. Our framework outperforms conventional pretrained models (e.g., [13]) on facial similarity assessment and more effectively selects images that are perceptually dissimilar to the original while preserving key attributes such as age and gender. Extensive validation using our human-annotated SimCelebA dataset underscores the effectiveness and specificity of our contributions to face anonymization.

## 2. RELATED WORKS

### 2.1. Face Recognition Models

In recent years, machine learning models have come to dominate face recognition, a technology used in applications from smartphone authentication to law enforcement. Typically, these systems identify the face in a database that matches the identity of a given test image.

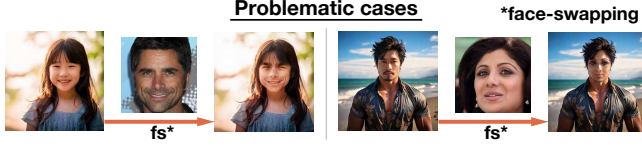**Problematic cases**          ***face-swapping**

**Fig. 2**: Example of unnatural face swapping. Swapping a young child's face with an adult male's can produce unrealistic images with deep wrinkles and age-inappropriate contours. Similarly, gender-based facial differences can also lead to unnatural results.

To achieve this, feature extractors are trained to compute distances between feature vectors—proxies for facial similarity—using losses such as contrastive loss [14, 24, 25, 26], triplet loss [15, 16, 27], and angular margin-based losses [13, 19, 18]. However, because these similarity measures are optimized for verification and identification—tasks focused on confirming whether two faces share the same identity—they fall short when it comes to effectively quantifying the degree of difference between faces of different individuals.

### 2.2. Face Anonymization via Face Swapping

Face swapping, in the context of facial anonymization, replaces only identity-specific facial features while preserving non-identifying attributes such as expression, pose, and lighting [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. In contrast to conventional anonymization methods—such as occlusion or blurring—that obscure the entire face and often degrade image quality, face swapping modifies only the features critical for identity recognition. In practical applications, this process typically involves replacing a target face with one drawn from a set of source faces that have sufficiently different identities. Although state-of-the-art face recognition systems can detect and quantify differences between target and source faces through deep feature embeddings [28], human observers may still perceive a high degree of similarity or even believe the target and face-swapped images belong to the same individual, even when the model has assigned them different identities.

## 3. METHOD

The primary goal of this work is to develop an effective facial feature extractor that accurately measures facial similarity, facilitating the selection of appropriate source and target faces for face-swapping. Building on this, we propose a facial anonymization framework that replaces a target face with a source face that shares key attributes (*e.g.*, gender, age) while ensuring that the original and face-swapped images remain as perceptually distinct as possible. Traditional facial anonymization methods (*e.g.*, [12]) typically rely on pretrained face recognition models (*e.g.*, ArcFace [13]). However, these models are not optimized for evaluating similarity between images of entirely different identities or for assessing similarity distances in face-swapped images. To address these limitations, we conduct human assessments of similarity in face-swapped images and construct a dataset, *SimCelebA*, to train a facial feature extractor, *PerFace*, using metric learning.

### 3.1. SimCelebA Dataset

Firstly, we conducted a human assessment to create a training dataset containing *face-swapped* images annotated with perceptual facial similarity scores. Since assessing the absolute similarity between

two face-swapped images is inherently challenging, we adopted a triplet-based approach. Each triplet comprised three images: a reference image (C) and two comparison images (A and B). Participants were tasked with determining which of the two, A or B, bore a closer resemblance to C.

To prevent similarity judgments from being overly influenced by key facial attributes of the target (*e.g.*, hair style, contour) or artifacts caused by face-swapping, we kept the target face constant and selected three distinct source images for swapping. This ensured that all images presented to the participants were face-swapped, eliminating potential biases arising from the presence of natural images.

We utilized the CelebAMask-HQ dataset [29], a large-scale collection of 30,000 high-resolution face images. From this dataset, we manually selected 80 target images (40 male and 40 female) and selected 240 source images (120 male and 120 female), forming 80 triplets. To minimize bias, images with obstructions such as glasses were excluded from the dataset. Using these chosen target and source images, we applied the SimSwap method [2] to generate face-swapped images. This process resulted in a dataset of 6,400 samples.

We recruited 18 participants, ensuring that each triplet was annotated by at least three individuals to capture human-perceived facial similarity. To enhance the reliability of the annotations, we embedded multiple dummy samples within the triplets. In these dummy samples, two images depicted the same individual, ensuring that a careful examination would unambiguously yield the correct answer. Consequently, only annotations from participants who answered all of these dummy samples correctly were considered valid. As a result, each sample ultimately received three high-quality annotations. This dataset was divided into training, validation, and test sets for use.

### 3.2. PerFace: facial similarity extractor trained on SimCelebA

Our goal is for the model to acquire the ability to evaluate facial similarity in a human-like manner, going beyond mere identity matching. To achieve this, we fine-tuned ArcFace [13] on our SimCelebA dataset, leveraging its strong discriminative power in face recognition tasks to better capture human-perceived facial similarity in face-swapped images. Given an annotated triplet (*i.e.*, A, B, and C) of face-swapped images, we propose using the following triplet loss to train the model:

$$\mathcal{L} = \max\left(0, \left(\cos(x_i, x_i^-) - \cos(x_i, x_i^+)\right) + m\right). \quad (1)$$

Here, $x_i \in \mathbb{R}^d$ represents the embedded feature of the reference image (*i.e.*, C) in the $i$-th sample. $x_i^+ \in \mathbb{R}^d$ is the embedded feature of the image chosen as more similar to the reference image $x_i$ by the majority of annotators (*i.e.*, A or B). Similarly, $x_i^- \in \mathbb{R}^d$ is the feature vector of the image chosen by the minority (*i.e.*, A or B). The feature vector dimension $d$ is set to 512, following prior studies [30, 13, 18, 19]. $\cos(\cdot, \cdot)$ denotes cosine similarity, and $m$ represents the margin. This triplet loss is minimized to ensure that the distance between similar pairs chosen by annotators decreases, while the distance between non-selected pairs increases. As a result, the model is fine-tuned such that the distances between embedded features of similar faces, as perceived by humans, become smaller. To facilitate face similarity comparisons, our dataset is designed to select relatively similar faces, as described in Sec.3.1. Accordingly, the loss function is also designed to focus on relative distances.

**1. Select a set with closer attributes. (gender/age)**

Query — Male / Female — Candidates

Query — Young / Older — Candidates
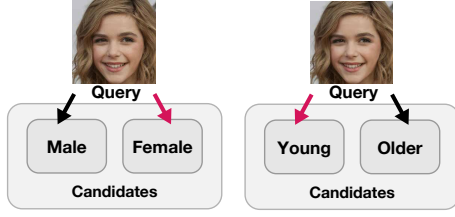
**2. Select the items with low similarity within the chosen set.**

Female / Young — Candidates
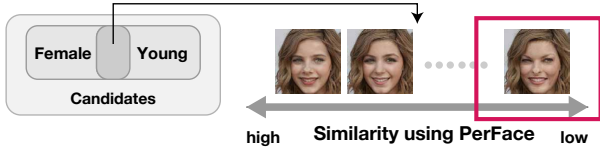
high — **Similarity using PerFace** — low

**Fig. 3**: Overview of the method for selecting a suitable source for face swapping with a query image.

### 3.3. Face Anonymization with PerFace

In most existing work (e.g., ArcFace [13]), when employing face swapping for anonymization, only the similarity between the source and target faces is evaluated. In contrast, we aim to assess the similarity between the target face and the *face-swapped* face using our PerFace feature extractor. However, simply comparing all pairs of source and target faces is prohibitively resource demanding, so we first align facial attributes before and after swapping and then compare faces only within each group.

The overview of the selection method, which consists of two steps, is shown in Fig. 3. Considering practical applications, a pre-defined set of face-swap candidates may sometimes be prepared. Therefore, this study assumes that the face-swap candidates have been annotated with attributes in advance. Conversely, it is assumed that the target attributes are unknown. Since it is impossible to predict which face the user wishes to anonymize, pre-annotating the target is impractical.

**STEP 1: Group Selection.** To ensure that the face-swapping process remains natural, suitable face-swapping candidates are identified through facial similarity evaluation. Assume there is a face image that the user wants to anonymize, referred to as the query image. The attributes of the query image are determined. Within the set of face-swapping candidates, images are grouped based on attributes. Following previous studies [31], age and gender are considered as attributes. We created 4 attribute groups (male, female, young, and older) and their intersection sets considering both gender and age (e.g.,young∩male), resulting in a total of 8 groups. For each attribute group, the similarity with the query image is calculated, and the group with the highest similarity is selected as the face-swapping candidates.

**STEP 2: Anonymization.** Next, the anonymization process is performed. As long as the selected group from STEP1 is used, the face-swapped image is expected to maintain a certain degree of naturalness. The goal is to achieve anonymization while maintaining this naturalness. Using the proposed method, our model trained via metric learning is employed to estimate the similarity between the images of the query and the face-swapping candidates within the group. At this stage, an image with low similarity in the group is chosen as the source for face swapping. By choosing less similar face from the same attribute group, face anonymization can be achieved while
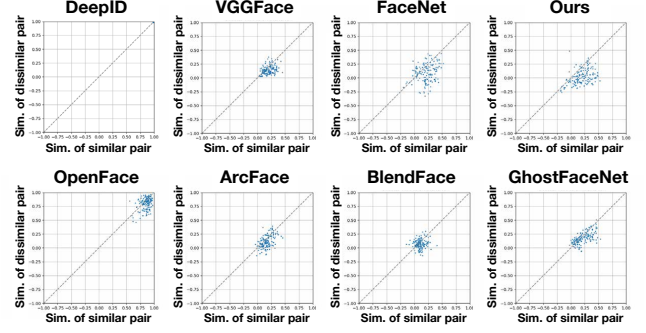


**Fig. 4**: Scatter plot of predicted similarity scores for similar and dissimilar pairs by major face recognition models. The X-axis represents the predicted similarity between similar pairs, while the Y-axis represents the predicted similarity between dissimilar pairs. The gray dashed line represents the graph of $Y = X$. Points below $Y < X$ indicate that the pair judged as more similar has been assigned a higher similarity score by the model.

ensuring naturalness.

## 4. EXPERIMENTS AND DISCUSSION

### 4.1. Training details

We adopted ArcFace [13] as the pretrained model, which we trained using the MS1MV3 [32] dataset. For the feature extractor, we employed a ResNet50[33]. We considered two cases for training: (1) using all triplet data in the training set ($D1$), and (2) using only triplet data with consistent annotations ($D2$).
**Common Settings for Training Data D1 and D2.** The batch size was 32, momentum was set to 0.9, weight decay to 5e-4, and the learning rate to 0.01. SGD was used as the optimizer, and the loss margin in Equation (1) was set to 0.1.

### 4.2. Evaluation method

Using the test triplet data and the annotation results, similar pairs and dissimilar pairs were created. Similar pairs consisted of the reference image and the image selected as more similar. Dissimilar pairs consisted of the reference image and the image not selected as more similar. If the similarity score for the similar pair was higher than that for the dissimilar pair, the model prediction was considered correct for that sample. Only samples with consistent annotations across all three responses were adopted as evaluation data.

### 4.3. Comparison with other methods

Here, we compare our method with existing approaches capable of measuring the distance between faces. We used $D2$ to ensure the highest annotation quality. As for BlendFace [11], we used the official code and weights. Other similarity evaluations were conducted through the DeepFace framework [34], with RetinaFace [35] as the detector and cosine similarity as the metric.

Fig. 4 presents scatter plots of similarity scores for similar and dissimilar pairs. The proportion of points below $Y < X$ corresponds to the accuracy values shown in Table 1. Our proposed method significantly outperformed existing methods in terms of accuracy. Models such as ArcFace [13], VGG Face [16], GhostFaceNets [23],

**Table 1**: Similarity evaluation results in comparison with previous methods.

| Method | Acc |
|---|---|
| DeepID[14] | 0.604 |
| VGG Face[16] | 0.750 |
| FaceNet[15] | 0.715 |
| OpenFace[17] | 0.660 |
| ArcFace[13] | 0.701 |
| BlendFace[11] | 0.576 |
| GhostFaceNets[23] | 0.604 |
| **Ours** | **0.917** |

**Table 2**: Examples of evaluating similar and dissimilar pairs in triplet samples using our method compared to others. For each triplet, the central image and the left-side image form a similar pair, while the central image and the right-side image form a dissimilar pair.

| Method | similar | dissimilar | similar | dissimilar |
|---|---|---|---|---|
| DeepID[14] | 0.993 | 0.997 | 0.998 | 0.998 |
| VGG Face[16] | 0.252 | 0.349 | 0.198 | 0.376 |
| FaceNet[15] | 0.207 | 0.261 | 0.405 | 0.339 |
| OpenFace[17] | 0.905 | 0.948 | 0.864 | 0.823 |
| ArcFace[13] | 0.287 | 0.273 | 0.288 | 0.314 |
| BlendFace[11] | 0.105 | 0.143 | 0.144 | 0.051 |
| GhostFaceNets[23] | 0.333 | 0.364 | 0.171 | 0.257 |
| **Ours** | **0.339** | **-0.006** | **0.183** | **0.055** |

and BlendFace[11] predicted similarity scores for both similar and dissimilar pairs around 0.00 to 0.50. In contrast, OpenFace [17] and DeepID [14] predicted high similarity scores for both similar and dissimilar pairs. For the triplet samples used in this evaluation, all three face images inherited the hairstyle and skin tone of target, making it difficult for these models to understand differences in facial parts. Specific examples are presented in Table2. Comparison methods make incorrect predictions, often assigning higher similarity scores to the dissimilar pair or consistently predicting nearly identical scores.

### 4.4. Analyzing the effects on source/target knowledge

Considering the task of face anonymization, there may be cases where the source used for anonymization is known in advance. Therefore, we investigated how prior knowledge of identity affects face similarity evaluation. Additionally, we also examined the impact of training data quality on model performance.

**Evaluation Data: [i], [ii], and [iii].** We evaluated with three types of evaluation datasets: **[i]** Neither the source nor the target were included in the training data. **[ii]** The target was not included in the training data, but the source was included. **[iii]** The source was not included in the training data, but the target was included. As summarized in Fig. 5, set [ii] achieved the highest accuracy, suggesting that familiarity with source faces improves performance. In contrast, [i] and [iii], which involved unknown source, resulted in lower accuracy. This highlights the advantage of pretraining on known face-swapping candidates.

**Table 3**: Mean±SD of facial similarity prediction accuracy.

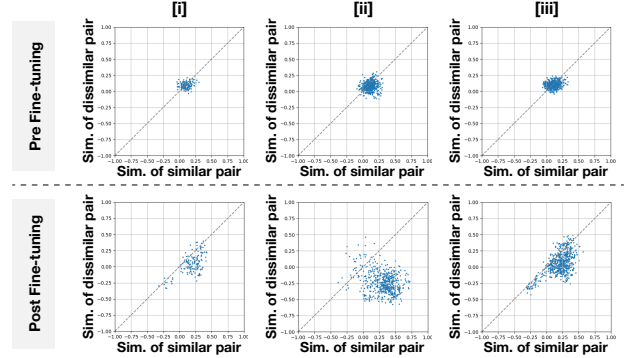| Dataset | Acc in [i] | Acc in [ii] | Acc in [iii] |
|---|---|---|---|
| - (Original) | 0.690 | 0.600 | 0.644 |
| D1 | 0.753 ±0.025 | 0.795 ±0.006 | 0.770 ±0.014 |
| D2 | 0.862 ±0.031 | 0.922 ±0.007 | 0.864 ±0.018 |



**Fig. 5**: Scatter plot of the similarity scores output by the model for similar and dissimilar pairs in triplet data.

**Training Data:** $D1$ **and** $D2$**.** As shown in Table 3, we found that while $D1$ benefits from a larger dataset, its inconsistent annotations hinder performance. In contrast, $D2$ achieves higher accuracy due to its superior annotation quality. For the following experiments, $D2$ was employed.

**Key Insight.** High-quality training data ($D2$) and familiarity with source triplets ([ii]) are critical for optimal model performance, emphasizing the importance of dataset curation and pretraining strategies.

### 4.5. Selection of attribute groups

This section presents the results of the performance on attribute group classification task. We determined the most similar attribute group for a given face query image, based on our perceptual similarity metric.

**Details of the attribute classification experiment.** In this experiment, we defined the following eight attribute groups: male, female, young, older, and their combinations: young∩male, older∩male, young∩female, and older∩female. These attribute groups were defined based on annotations from CelebAMask-HQ [29]. Additionally, we rechecked the annotations and corrected several annotation errors. For the four subdivided groups (young∩male, older∩male, young∩female, older∩female), 100 face images were randomly selected for each group. For the broader male, female, young, and older groups, each group was constructed as the union of the corresponding subdivided groups, containing 200 face images per group. By dividing the dataset into these groups, we prepared to analyze how similarity evaluation changes across attribute groups.

For an image $I_{G_i,j}$ in an attribute group $G_i$ and a query image $I_q$, the distance $d_{I_q,I_{G_i,j}}$ is defined as: $d_{I_q,I_{G_i,j}} = 1 - \cos(f(I_q), f(I_{G_i,j}))$ where $f(\cdot)$ is the feature extractor in our proposed model. The distance between an attribute group $G_i$ and the query image $I_q$ is denoted as $D_{I_q,G_i}$. The attribute group $G^*$ most similar to the query image $I_q$ is defined as: $G^* = \underset{G_i}{\arg\min}\, D_{I_q,G_i}$
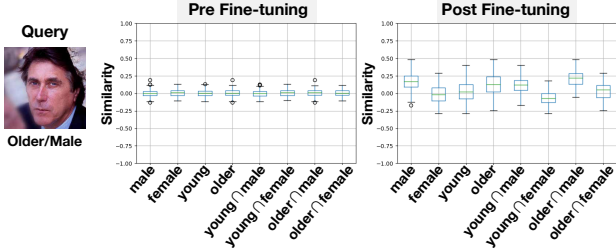
**Fig. 6**: Distributions of similarity between query images and face swap candidates within each attribute group.

**Table 4**: Classification accuracy when the distance $D_{I_q, G_i}$ is defined as the upper limit of the 95% confidence interval of distances $d_{I_q, I_{G_i, j}}$.

| | Category | Precision | Recall | Accuracy | AUC |
|---|---|---|---|---|---|
| | Male | 0.930 | 0.936 | 0.933 | 0.933 |
| | Female | 0.936 | 0.930 | | |
| | Young | 0.829 | 0.812 | 0.822 | 0.822 |
| Pre | Older | 0.816 | 0.832 | | |
| Fine-tuning | Young∩Male | 0.698 | 0.812 | 0.865 | 0.847 |
| | Young∩Female | 0.807 | 0.652 | 0.874 | 0.800 |
| | Older∩Male | 0.801 | 0.788 | 0.898 | 0.861 |
| | Older∩Female | 0.770 | 0.804 | 0.891 | 0.862 |
| | Male | 0.959 | 0.972 | 0.965 | 0.965 |
| | Female | 0.972 | 0.958 | | |
| | Young | 0.871 | 0.754 | 0.821 | 0.821 |
| Post | Older | 0.783 | 0.888 | | |
| Fine-tuning | Young∩Male | 0.716 | 0.856 | 0.879 | 0.871 |
| | Young∩Female | 0.822 | 0.664 | 0.880 | 0.808 |
| | Older∩Male | 0.872 | 0.788 | 0.918 | 0.875 |
| | Older∩Female | 0.791 | 0.864 | 0.909 | 0.894 |

We used a total of 1000 query images, selecting 250 images from each of the attribute groups: young∩male, young∩female, older∩male, and older∩female. As the evaluation, for each query image $I_q$, the group $G^*$ with the minimum distance was identified, and the accuracy was evaluated by calculating the proportion of queries where the predicted attribute label matched the actual attribute label.

**Results.** Fig. 6 shows the similarity distribution between the query image and each attribute group. Here, we use an image with the attributes "male ∩ older" as an example to illustrate the analysis. As a result, in the fine-tuned model, the variance within attribute groups and the variance between groups both increased compared to before fine-tuning. It was observed that "male," "older," and "male ∩ older" had higher similarities than other groups.

Based on the similarity distribution, we classified the closest attribute group to the query image using the 95% upper confidence interval as the distance. We then verified whether the attribute of the selected group matched the actual attribute of the query image. The classification method was evaluated under two conditions: considering single attribute and considering multiple attributes simultaneously. For single attribute, we perform binary classification of gender (male or female) and age group (young or older). When considering multiple attributes simultaneously, we classify into four groups: "young ∩ male," "young ∩ female," "older ∩ male," and "older ∩ female."

As shown in Table 4, after fine-tuning, the AUC increased in most cases compared to before fine-tuning, indicating improved classification accuracy. Focusing on the differences in classification accuracy by attribute, the accuracy for male/female classification is higher than that for young/older classification. This suggests that the model prioritizes gender (male/female) over age group (young/older) when evaluating similarity. This can likely be traced back to the annotators' judgments in the training dataset, which may have placed greater emphasis on gender distinctions.

### 4.6. Selection of face swap candidates

Based on the attributes of the query image determined in Sec. 4.5, suitable face swap candidates for anonymization were selected from those sharing the same attributes. To select candidates with lower similarity to the query, the face swap candidates in the selected group were sorted by similarity to the query, as examples shown in Fig. 7.

Given the diversity of human faces, even among faces that are not very similar to the query image, various types of "dissimilarity" may exist. To provide users with more flexibility in selecting a "dissimilar" face, this selection algorithm can effectively recommend multiple face swap candidates with relatively low similarity.
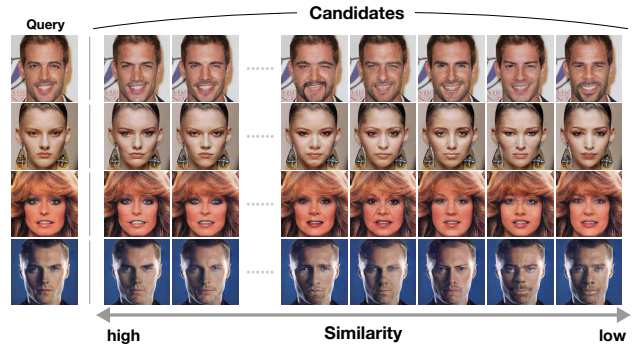


**Fig. 7**: Sorted face swap candidates in the selected attribute group, showing the top two most similar and the bottom five least similar candidates.

## 5. CONCLUSION

We propose a novel method to address limitations in face similarity prediction for face anonymization via natural face swapping. Conventional methods struggle to evaluate nuanced similarities, such as distinguishing "completely different" from "highly similar but different individuals."

To overcome this, we introduce a new task to assess "how similar a different identity is," using our perceptual similarity model PerFace. We constructed an evaluation dataset via user studies and developed a transfer-learning-based model optimized to capture subtle inter-individual similarities.

Our PerFace model significantly outperformed baseline models in face similarity judgment tasks. Additionally, it achieved superior accuracy in attribute classification, highlighting the influence of facial attributes on perceptual similarity judgments and offering insights into perception-based evaluations.

A limitation of this work is its inclusion of subjective factors, especially since facial similarity inherently involves personal perception. Consequently, broader and more diverse perspectives are crucial to capture the sense of similarity that people share. We hope this paper will inspire the creation of larger and more inclusive datasets.

# 6. REFERENCES

[1] "deepfakes," https://github.com/deepfakes/faceswap, Accessed: Dec. 26, 2024.

[2] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge, "Simswap: An efficient framework for high fidelity face swapping," in *MM*, 2020, pp. 2003–2011.

[3] Kihong Kim, Yunho Kim, Seokju Cho, Junyoung Seo, Jisu Nam, Kychul Lee, Seungryong Kim, and KwangHee Lee, "Diffface: Diffusion-based face swapping with facial guidance," *arXiv preprint arXiv:2212.13344*, 2022.

[4] Leslie Wöhler, Susana Castillo, and Marcus Magnor, "Personality analysis of face swaps: can they be used as avatars?," in *IVA*. 2022, IVA '22, Association for Computing Machinery.

[5] Yuval Nirkin, Yosi Keller, and Tal Hassner, "Fsgan: Subject agnostic face swapping and reenactment," in *ICCV*, 2019, pp. 7184–7193.

[6] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen, "Advancing high fidelity identity swapping for forgery detection," in *CVPR*, 2020, pp. 5073–5082.

[7] Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu, and Zhenan Sun, "One shot face swapping on megapixels," in *CVPR*, 2021, pp. 4834–4844.

[8] Gege Gao, Huaibo Huang, Chaoyou Fu, Zhaoyang Li, and Ran He, "Information bottleneck disentanglement for identity swapping," in *CVPR*, 2021, pp. 3404–3413.

[9] Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji, "Hififace: 3d shape and semantic prior guided high fidelity face swapping.," in *IJCAI*, 2021, pp. 1136–1142.

[10] Wenliang Zhao, Yongming Rao, Weikang Shi, Zuyan Liu, Jie Zhou, and Jiwen Lu, "Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion," in *CVPR*, 2023, pp. 8568–8577.

[11] Kaede Shiohara, Xingchao Yang, and Takafumi Taketomi, "Blendface: Re-designing identity encoders for face-swapping," in *ICCV*, 2023, pp. 7634–7644.

[12] Umur A Ciftci, Gokturk Yuksek, and Ilke Demir, "My face my choice: Privacy enhancing deepfakes for social media anonymization," in *WACV*, 2023, pp. 1369–1379.

[13] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699.

[14] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang, "Deep learning face representation by joint identification-verification," *NeurIPS*, vol. 27, 2014.

[15] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015, pp. 815–823.

[16] Omkar Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep face recognition," in *BMVC*. British Machine Vision Association, 2015.

[17] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency, "Openface: an open source facial behavior analysis toolkit," in *WACV*. IEEE, 2016, pp. 1–10.

[18] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017, pp. 212–220.

[19] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu, "Cosface: Large margin cosine loss for deep face recognition," in *CVPR*, 2018, pp. 5265–5274.

[20] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou, "Magface: A universal representation for face recognition and quality assessment," in *CVPR*, 2021, pp. 14225–14234.

[21] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao, "Synface: Face recognition with synthetic data," in *ICCV*, 2021, pp. 10880–10890.

[22] Minchul Kim, Anil K. Jain, and Xiaoming Liu, "Adaface: Quality adaptive margin for face recognition," in *CVPR*, 2022, pp. 18750–18759.

[23] Mohamad Alansari, Oussama Abdul Hay, Sajid Javed, Abdulhadi Shoufan, Yahya Zweiri, and Naoufel Werghi, "Ghostfacenets: Lightweight face recognition model from cheap operations," *IEEE Access*, vol. 11, pp. 35429–35446, 2023.

[24] Yi Sun, Xiaogang Wang, and Xiaoou Tang, "Deeply learned face representations are sparse, selective, and robust," in *CVPR*, 2015, pp. 2892–2900.

[25] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[26] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[27] Changxing Ding and Dacheng Tao, "Robust face recognition via multimodal deep face representation," *IEEE transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015.

[28] Jingyi Cao, Xiangyi Chen, Bo Liu, Ming Ding, Rong Xie, Li Song, Zhu Li, and Wenjun Zhang, "Face de-identification: State-of-the-art methods and comparative studies," *arXiv preprint arXiv:2411.09863*, 2024.

[29] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *CVPR*, 2020.

[30] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*. Springer, 2016, pp. 499–515.

[31] Jiseob Kim, Jihoon Lee, and Byoung-Tak Zhang, "Smoothswap: A simple enhancement for face-swapping with smoothness," in *CVPR*, 2022, pp. 10779–10788.

[32] Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi, "Lightweight face recognition challenge," in *ICCVW*, 2019, pp. 2638–2646.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[34] Sefik Ilkin Serengil and Alper Ozpinar, "Lightface: A hybrid deep face recognition framework," in *INISTA*. IEEE, 2020, pp. 1–5.

[35] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *CVPR*, 2020, pp. 5203–5212.