HonestFace: Towards Honest Face Restoration with One-Step Diffusion Model
Supplementary Materials

Jingkai Wang[1], **Wu Miao**[1], **Jue Gong**[1], **Zheng Chen**[1],
Xing Liu[2], **Hong Gu**[2], **Yutong Liu**[1]*, **Yulun Zhang**[1]*
[1]Shanghai Jiao Tong University, [2]vivo Mobile Communication Co., Ltd

This supplementary material provides additional results to support the main manuscript. Section A first explores the application scenarios of HonestFace. Subsequently, Section B examines potential broader impacts. It considers both positive and negative aspects and discusses strategies to mitigate potential negative effects. Experiments detailed in Sec. C demonstrate the effectiveness of using multiple reference images; these images can offer more comprehensive identity information and improve ID consistency. The mathematical properties of the proposed affined landmark distance are then investigated in Sec. D. Further, Section E outlines the limitations of the current work and directions for future research. Finally, Section F offers additional visual comparisons with state-of-the-art methods. Notably, this section also includes results from tests of real-world data.

A Application Scenarios

Reference-guided restoration fits neatly into today's mobile photography. Most phones already group portraits by identity in the gallery. Thus, when a new photo is taken, the device can fetch a matching high-quality image, align it, and blend its fine textures into the live frame. The process runs on the device, so privacy is preserved and latency stays low. It is especially helpful in dim light, strong back-light, or front-camera selfies where detail is often lost. Manufacturers can also ship an opt-in library of public celebrity portraits. Then the fans could capture clear, faithful shots of singers or athletes even from noisy zoom images. In a studio workflow, editors can apply the same tool after filming; each frame is matched to an actor's portfolio to keep the appearance stable across takes. Across all these cases, the system raises visual quality without heavy computation, bringing identity-aware enhancement to everyday hardware.

B Broader Impacts

Our reference-guided face restoration model brings clear societal benefits by enhancing image quality in both everyday photography and professional workflows. It improves visual clarity under challenging conditions, producing more faithful and pleasing images with relatively low computational cost. While current deployments may rely on cloud processing, the model's lightweight design makes future on-device deployment on mobile phones feasible, potentially reducing latency and improving user experience.

At the same time, the technology poses risks, including misuse for creating realistic manipulated images or deepfakes, which could contribute to misinformation and privacy concerns. There is also a risk of biased performance if the model is trained on unbalanced datasets, potentially affecting fairness across different demographic groups.

To mitigate these risks, we emphasize responsible deployment strategies, including opt-in use, limiting model access, and ongoing monitoring for misuse. We encourage transparency about the technology's capabilities and limitations to prevent unintended harms. Overall, we believe the positive applications outweigh potential negatives when appropriate safeguards are in place.

*Corresponding authors: Yutong Liu <isabelleliu@sjtu.edu.cn> and Yulun Zhang <yulun100@gmail.com>.

C Comparisons of HonestFace with Varying Numbers of Exemplars

As shown in Tab. 1, the number of provided reference images positively correlates with the improvement in identity consistency. This observation indicates that providing more high-quality reference exemplars enhances the model's ability to preserve identity, highlighting the practical benefits of using multiple references in real-world applications. More reference information generally leads to a more honest restoration of identity. It demonstrates that HonestFace can effectively capture salient identity cues from the reference input and accurately apply this information to the face being restored.

# Max Exemplars	Deg.↓	L2-LD↓	ALD-e↓	ALD-m↓
1	27.6316	1.7955	13.6345	8.9821
3	26.6206	1.7929	13.6932	8.9996
5	26.2673	1.7978	13.6701	9.0007
7	26.1139	1.7823	13.5469	8.9880
N/A	**26.1060**	**1.7798**	**13.4444**	**8.9793**

Table 1: Comparisons of HonestFace on different exemplar numbers. ALD-e and ALD-m are the affine landmark distances for the eyes and mouth. N/A indicates that max exemplars is not limited.

D Detailed A-LMD Properties Analysis

The affine landmark distance ($d_{\text{A-LD}}$) serves as a robust measure for quantifying the geometric dissimilarity between a set of reconstructed facial landmark points and their ground truth counterparts. Defined for a set of restored landmarks $L = \{l_k\}_{k=1}^N$, a ground truth set $H = \{h_k\}_{k=1}^N$ (with $l_k, h_k \in \mathbb{R}^2$), it is formulated as the minimum sum of weighted squared Euclidean distances. This minimization is performed over all possible 2D affine transformation matrices $A \in \mathbb{R}^{2\times 3}$ that map the homogeneous coordinates $l'_k = [l_{k,x}, l_{k,y}, 1]^\intercal$ of the restored landmarks to the corresponding ground truth landmarks h_k:

$$d_{\text{A-LD}}(L, H; W) = \min_A \sum_{k=1}^{N} w_k \|Al'_k - h_k\|^2. \tag{1}$$

Analyzing its fundamental properties, $d_{\text{A-LD}}$ satisfies non-negativity as it is defined as the minimum of a sum of non-negative weighted squared Euclidean distances, ensuring $d_{\text{A-LD}}(L, H; W) \geq 0$. Furthermore, it exhibits a weak form of identity of indiscernibles: $d_{\text{A-LD}}(L, H; W) = 0$ if and only if the ground truth landmark set H can be perfectly obtained by an affine transformation of the restored landmark set L. This property is particularly advantageous in image restoration. It allows the metric to robustly assess the structural fidelity of the restored face, by effectively factoring out global affine differences (*e.g.*, translation, rotation, and scaling) that might arise from reconstruction processes or initial image alignment. A zero distance thus signifies perfect structural recovery up to such transformations, which is often the desired outcome for facial feature preservation.

We further provide the proof of non-negativity and the weak form of identity of indiscernibles.

Non-negativity: $d(L, H) \geq 0$.

Proof. The term $w_k\|Al'_k - h_k\|^2$ represents the weighted squared Euclidean distance between two 2D points. Since $w_k > 0$ and the Euclidean norm $\|\cdot\|$ always yields a non-negative value, its square is also non-negative. Therefore, the sum of non-negative terms, $\sum_{k=1}^N w_k\|Al'_k - h_k\|^2$, is always non-negative. Consequently, the minimum of this sum over A must also be non-negative. □

Identity of Indiscernibles (weak): $d(L, H) = 0 \iff L = H$.

Proof. (a) If L=H: Meaning l_k=h_k for all k=$1, \ldots, N$. We need to show that $d_{\text{A-LD}}(L, L; W) = 0$. Consider the identity affine transformation

$$A_{\text{id}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{2}$$

For this A_{id}, $A_{\text{id}}l'_k = l_k$. Substituting this into the sum: $\sum_{k=1}^N w_k\|A_{\text{id}}l'_k - l_k\|^2 = \sum_{k=1}^N w_k\|l_k - l_k\|^2 = \sum_{k=1}^N w_k\|0\|^2 = 0$. Since the minimum value cannot be less than 0, and we found an A that yields 0, it must be that $d_{\text{A-LD}}(L, L; W) = 0$.

(b) If $d_{\text{A-LD}}(L, H; W) = 0$: This implies that $\min_A \sum_{k=1}^N w_k\|Al'_k - h_k\|^2 = 0$. Since $w_k > 0$ and $\|Al'_k - h_k\|^2 \geq 0$, this sum can only be zero if each term is zero. Thus, for the optimal A^*, we must have $\|A^*l'_k - h_k\|^2 = 0$ for all $k = 1, \ldots, N$. This means $A^*l'_k = h_k$ for all k. In other words, there exists an affine transformation A^* that perfectly maps each landmark in L to its corresponding landmark in H. This indicates that L and H are affinely equivalent.

Therefore, this distance metric satisfies a weaker form of identity: $d_{\text{A-LD}}(L, H; W) = 0$ if and only if H can be perfectly obtained by an affine transformation of L. It does not require strict pointwise equality ($L = H$). □

However, $d_{\text{A-LD}}$ is generally not symmetric, meaning $d_{\text{A-LD}}(L, H; W) \neq d_{\text{A-LD}}(H, L; W)$, as the least-squares optimization is asymmetrical with respect to the roles of the source and target point sets. Despite these deviations from the axioms of a strict mathematical metric space, $d_{\text{A-LD}}$ remains a highly effective dissimilarity measure for its intended application. Its primary advantage lies in its inherent robustness to global geometric variations, enabling it to specifically quantify the non-affine distortions present in the restored facial structure. This makes it a more perceptually relevant and robust indicator of restoration quality compared to raw L2 distances, which are overly sensitive to mere positional or scale shifts.

E Existing Limitations and Future Work

Model size and inference speed. The identity encoder proposed integrates information from reference images. However, its structure can be simplified, which would cut memory use and shorten runtime. A leaner encoder is expected to run faster than OSDFace [9], pushing HonestFace closer to real-time use on mobile and edge devices.

Trade-offs between foundation models. HonestFace is built on Stable Diffusion 2.1. Larger backbones, such as SDXL (2.6 B parameters) [6] and Flux (12 B parameters) [1], can lift restoration quality but make on-device deployment harder. In future work, we will test the performance of our modules and loss in other foundation models, study the capacity-size balance, and design compact variants of large backbones.

More effective reference. We will refine the way reference images guide restoration. The goal is to keep helpful cues and discard outdated or irrelevant details. For example, a hairstyle in a month-old photo may be obsolete, or a slight camera angle shift may call for a different reference. If the reference image shows glasses but the target photo does not, a better substitute should be chosen. We plan to combine automatic filtering with optional user hints. A scoring module will rank candidate references and keep only those that match the current target.

Tone enhancement. Current methods correct color shift but still leave room for finer detail. We will adjust facial tone to improve naturalness, focusing on (a) higher contrast and haze removal; (b) balanced global and local contrast that fits the background; (c) lower lighting ratio so highlight-shadow gaps shrink and the face looks soft, even, and three-dimensional. All tone edits will respect the scene lighting to keep the whole image consistent.

Texture-aware IQA. Existing no-reference IQA metrics are not portrait-aware, which explains why our method does not rank as well as it looks to the eye. These metrics often score images such as DiffBIR [4] or FaceMe [5] highly, even when they appear over-synthetic or lack texture. We will develop new portrait-specific IQA and aesthetic metrics that consider texture, tone, and overall photographic appeal, giving a human-aligned view of visual quality.

F Further Visual Comparisons

Figures 1 and 2 present visual comparisons of HonestFace on real-world images. As for these real-world faces, we sourced low-quality facial images and corresponding high-quality reference images from online. Notably, in specific examples such as image "01" in Fig. 1, HonestFace effectively restores challenging features, including dark spots on the subject's skin. Similarly, for image "02" in Fig. 1, the face restored by HonestFace exhibits more natural teeth color and skin texture when compared to results from other methods.

Additionally, we provide further visual results on the CelebHQRef-Test and Reface-Test datasets, as illustrated in Figs. 3, 4, and 5. These examples highlight our method's capability to accurately restore subtle yet important details, such as eye bags (*e.g.*, image "101" in Fig. 4). Moreover, HonestFace generally achieves a more faithful restoration of diverse facial expressions (*e.g.*, image "374" in Fig. 5). Overall, faces restored by HonestFace possess richer and more natural textures. Our approach successfully avoids common issues like over-smoothing and does not introduce the artificial or repetitive patterns frequently observed in outputs from other methods.

References

[1] Black Forest Labs. Flux. https://github.com/black-forest-labs/flux, 2024. 3

[2] Xiaoming Li, Wenyu Li, Dongwei Ren, Hongzhi Zhang, Meng Wang, and Wangmeng Zuo. Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In *CVPR*, 2020. 5, 6, 7, 8, 9

[3] Xiaoming Li, Shiguang Zhang, Shangchen Zhou, Lei Zhang, and Wangmeng Zuo. Learning dual memory dictionaries for blind face restoration. *IEEE TPAMI*, 2022. 5, 6, 7, 8, 9

[4] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Towards blind image restoration with generative diffusion prior. In *ECCV*, 2024. 3, 5, 6, 7, 8, 9

[5] Siyu Liu, Zheng-Peng Duan, Jia OuYang, Jiayi Fu, Hyunhee Park, Zikun Liu, Chun-Le Guo, and Chongyi Li. FaceMe: Robust blind face restoration with personal identification. In *AAAI*, 2025. 3, 5, 6, 7, 8, 9

[6] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *ICLR*, 2024. 3

[7] Keda Tao, Jinjin Gu, Yulun Zhang, Xiucheng Wang, and Nan Cheng. Overcoming false illusions in real-world face restoration with multi-modal guided diffusion model. In *ICLR*, 2025. 7, 8, 9

[8] Yu-Ju Tsai, Yu-Lun Liu, Lu Qi, Kelvin CK Chan, and Ming-Hsuan Yang. Dual associated encoder for face restoration. In *ICLR*, 2024. 5, 6, 7, 8, 9

[9] Jingkai Wang, Jue Gong, Lin Zhang, Zheng Chen, Xing Liu, Hong Gu, Yutong Liu, Yulun Zhang, and Xiaokang Yang. OSDFace: One-step diffusion model for face restoration. In *CVPR*, 2025. 3, 5, 6, 7, 8, 9

[10] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *NeurIPS*, 2024. 5, 6

[11] Peiqing Yang, Shangchen Zhou, Qingyi Tao, and Chen Change Loy. PGDiff: Guiding diffusion models for versatile face restoration via partial guidance. In *NeurIPS*, 2023. 5, 6, 7, 8, 9

[12] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *NeurIPS*, 2022. 5, 6

Figure 1: Visual comparison of real-world faces. Please zoom in for a better view.

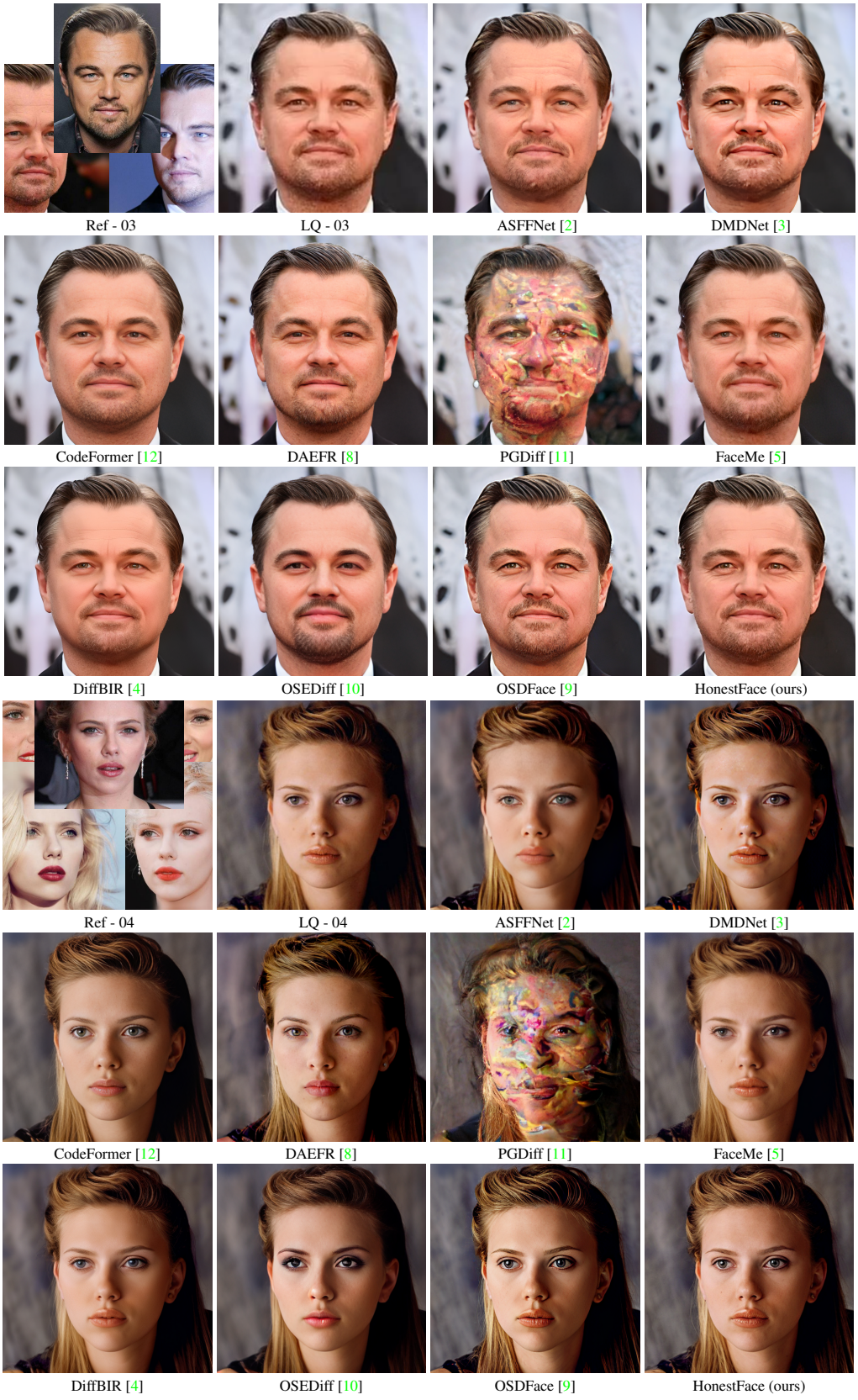

Figure 2: Visual comparison of real-world faces. Please zoom in for a better view.

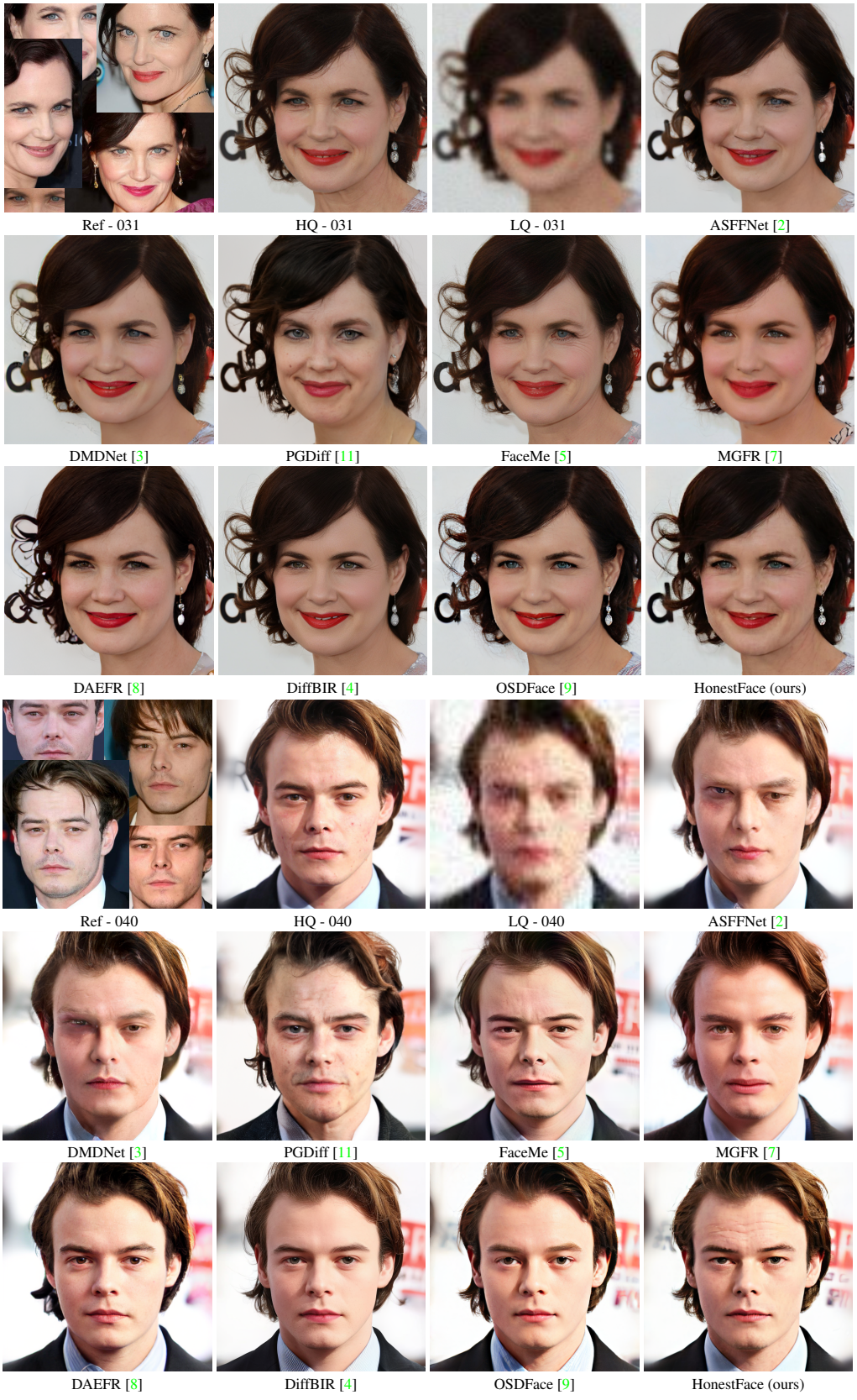

Figure 3: More visual comparison of CelebHQRef-Test. Please zoom in for a better view.

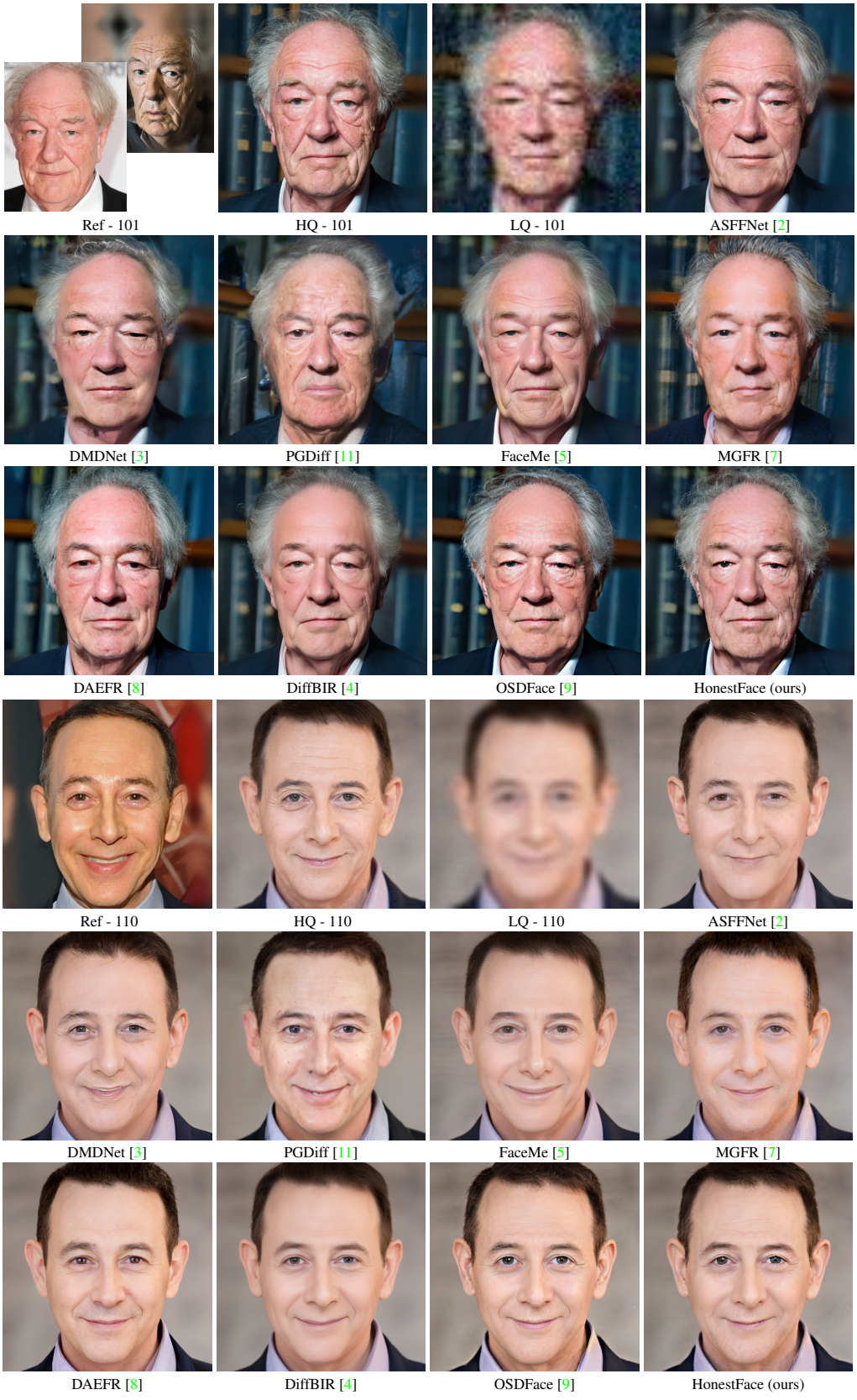

Figure 4: More visual comparison of Reface-Test. Please zoom in for a better view.

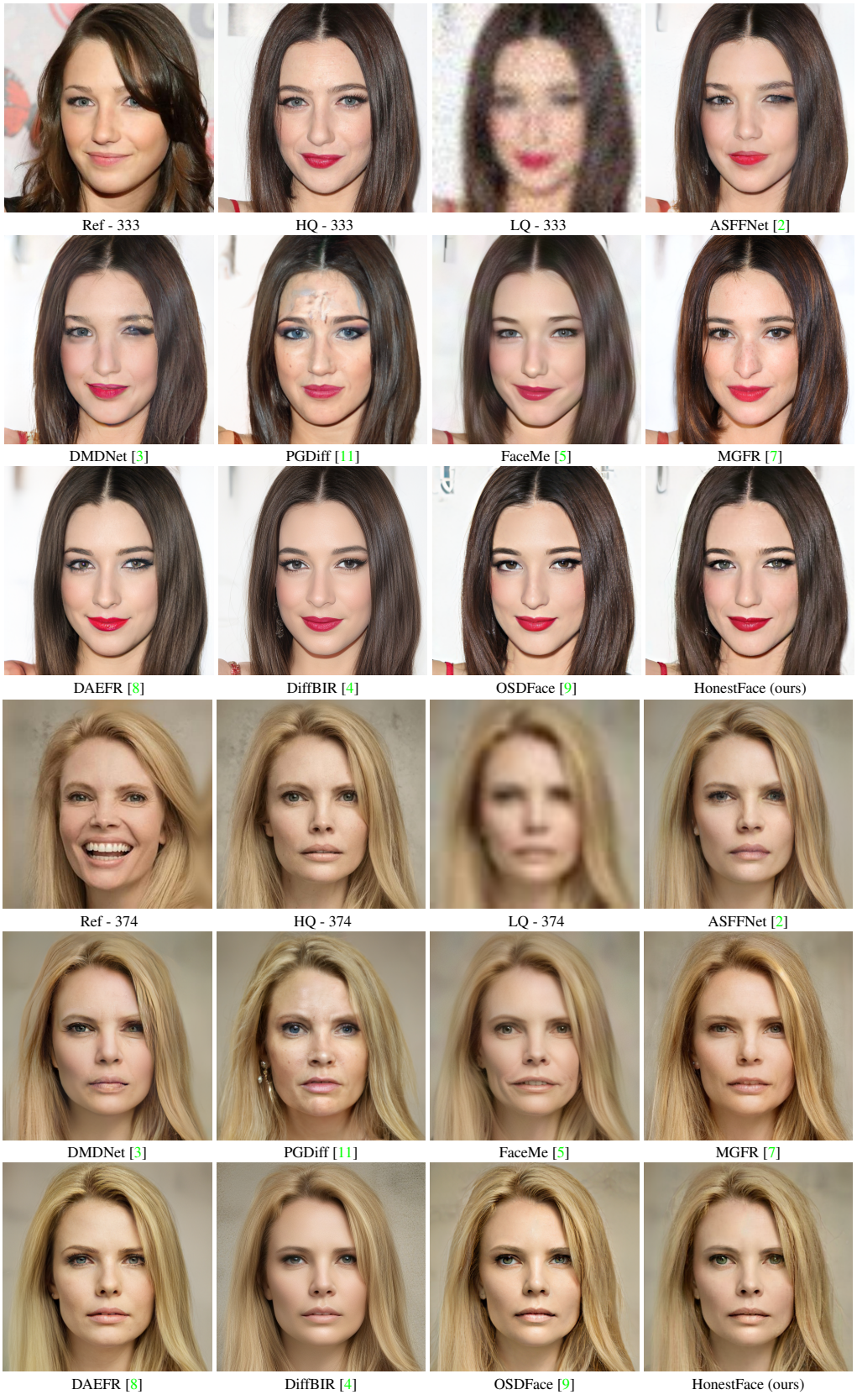

Figure 5: More visual comparison of Reface-Test. Please zoom in for a better view.