EDFace-Celeb-1M: Benchmarking Face Hallucination with a Million-scale Dataset

Kaihao Zhang, Dongxu Li, Wenhan Luo, Jingyu Liu, Jiankang Deng, Wei Liu, Stefanos Zafeiriou

Abstract—Recent deep face hallucination methods show stunning performance in super-resolving severely degraded facial images, even surpassing human ability. However, these algorithms are mainly evaluated on non-public synthetic datasets. It is thus unclear how these algorithms perform on public face hallucination datasets. Meanwhile, most of the existing datasets do not well consider the distribution of races, which makes face hallucination methods trained on these datasets biased toward some specific races. To address the above two problems, in this paper, we build a public Ethnically Diverse Face dataset, EDFace-Celeb-1M, and design a benchmark task for face hallucination. Our dataset includes 1.7 million photos that cover different countries, with relatively balanced race composition. To the best of our knowledge, it is the largest-scale and publicly available face hallucination dataset in the wild. Associated with this dataset, this paper also contributes various evaluation protocols and provides comprehensive analysis to benchmark the existing state-of-the-art methods. The benchmark evaluations demonstrate the performance and limitations of state-of-the-art algorithms. https://github.com/HDCVLab/EDFace-Celeb-1M

Index Terms—Face Hallucination; Face Super-resolution; Benchmarking; Million-scale Dataset; EDFace-Celeb-1M.

1 Introduction

Human faces contain important identity information and are central to various vision applications, such as face alignment [1], [2], [3], face parsing [4], [5] and face identification [6], [7]. However, most of these applications require high-quality images as input and the approaches perform less favorably in low-resolution conditions. To alleviate the issue, the task of face hallucination, or face super-resolution, aims to super-resolve low-resolution face images to their high-resolution counterparts, thus facilitating effective face analysis.

As a special case of Single Image Super-Resolution (SISR) [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], face hallucination is a fundamental and challenging problem in face analysis. Different from general SISR, which deals with super-resolving pixels in arbitrary scenes, the face hallucination task tackles only facial images. Therefore, the facial special prior knowledge in face images could help recover accurate face shape and rich facial details. As a result, the face hallucination methods often achieve better performance than general single image super-resolution ones in terms of higher up-scaling factors. Previous face hallucination methods [18], [19], [20], [21], [22], [23], [24], [25], [26] utilize facial priors to restore high-resolution images by approaches which are typically not end-to-end ones. Currently, a number of deep learning based methods [27], [28], [29], [30], [31], [32], [33], [34] are proposed to greatly boost the performance of the task of face hallucination, even surpassing human ability.

Is face hallucination solved? Firstly, many current deep hallucination methods evaluate their methods on *non-public* synthetic datasets. Especially, they typically download public datasets and synthesize pairs of low-resolution and high-resolution images, and then *randomly* select some faces as training and testing samples. After that, they evaluate their methods and re-train previous methods on their synthesized datasets. However, using this way for evaluation presents two obvious problems. (1) Given that the division of training/testing groups is random, the following researchers cannot strictly follow and reproduce the previous experiments like [35], [36]. (2) To verify that the new face hallucination methods outperform the previous methods, researchers have to synthesize new datasets by themselves and re-train previous methods again, which greatly increases the unnecessary workload and reduces the credibility of their results like [29], [33]. Therefore, it is not clear how these algorithms would perform on public face hallucination datasets.

On the other hand, though it is popular that the current deep face hallucination methods are evaluated on synthesized face datasets, these datasets face the problem of being biased toward specific races, and other races are significantly ignored. This scheme along with these datasets not only fails to accurately evaluate the performance of face hallucination methods, but also will raise the ethical problem. Therefore, a large-scale face hallucination dataset with relatively balanced race composition is necessary for face analysis in the community.

To address the above problems, in this paper, we introduce an Ethnically Diverse Face dataset called EDFace-Celeb-1M, and design benchmark protocols along with analysis to evaluate and encourage the development of face hallucination algorithms. There exist three key objectives of creating the EDFace-Celeb-1M dataset for face hallucination.

- *K. Zhang and D. Li are with the College of Engineering and Computer Science, Australian National University, Canberra, Australia. E-mail: super.khzhang@gmail.com, dongxu.li@anu.edu.au*
- *W. Luo is with Sun Yat-sen University, Guangzhou, China. E-mail: whluo.china@gmail.com.*
- *J. Liu and W. Liu are with the Tencent, Shenzhen, China. E-mail: jyliu0329@gmail.com; wl2223@columbia.edu.*
- *J. Deng and S. Zafeiriou are with Department of Computing, Imperial College London, London, UK. E-mail: {j.deng16, s.zafeiriou}@imperial.ac.uk.*

(1) It should contain a large-scale set of face images in the wild with unconstrained pose, emotion and exposure. Specially, the proposed EDFace-Celeb-1M dataset includes 1.7 million photos of more than $40,000$ unique celebrity subjects. (2) The dataset should include faces from as many different countries and races as possible to mitigate the race bias in the current face hallucination datasets. More specifically, the EDFace-Celeb-1M dataset contains different race groups including White, Black, Latino and Asian with a relatively balanced composition. (3) It should be publicly available, enabling benchmarking current and future face hallucination methods with a unified dataset protocol to ensure fair and effortless evaluation. In this paper, the proposed dataset is public available with fixed training and testing sets for fair comparison.

To investigate the performance of current deep face hallucination algorithms on the constructed dataset with relatively balanced race composition, we design and provide concrete evaluation protocols, and evaluate four publicly available face hallucination methods and four SISR methods. We also introduce in detail our experiment setup and report baseline models to benefit and drive future research for the face hallucination task and inspire other related tasks in the computer vision community.

Our main contributions are summarized as follows.

- First, to the best of our knowledge, we build the *first large-scale publicly available* face hallucination dataset with relatively balanced race composition. The dataset includes 1.7 million face images collected from different race groups, providing *fixed* training and testing groups, pairs of low-resolution and high-resolution images with different scale factors (*e.g.*, $2\times$, $4\times$, $8\times$), and aligned and non-aligned face images, which makes the future comparison more convenient, repeatable and credible.
- Second, we design a *benchmark* task to evaluate the performance of some current deep face hallucination and SISR methods to super-resolve low-resolution images. By doing so, it is clear how these algorithms perform on public face hallucination datasets (See Section 4).
- Third, we address *fundamental questions* of face hallucination and obtain several key findings.
  - How well do the current face hallucination and SISR methods perform in the case of different upsampling factors? (See Table 2 and Figure 8,9,10,11)
  - How do the noise and blur kernels affect the performance of face hallucination methods? (See Table 2 and Figure 8,9,10,11)
  - Is the size of training data important? (See Table 4 and Figure 7)
  - Using the super-resolved images for the facial analysis tasks like landmark detection and identity preservation, is the gap significant, compared with using the ground truth high-resolution face images? (See Table 3)

2 Related Work

2.1 Face Hallucination

Many approaches have been proposed for face hallucination, which can be classified into two categories: non-deep learning methods and deep learning methods. For the non-deep learning methods, holistic-based [19], [37], [38] and part-based [22], [39], [40] techniques are two popular models, which upsample face images via representing faces by parameters and extracting facial regions, respectively.

Recently, deep neural networks have been successfully applied to various computer vision tasks including face hallucination. Yu *et al.* [35] investigate the Generative Adversarial Network (GAN) to super-resolve face images of very low resolution and create perceptually realistic high-resolution face images. Huang *et al.* [29] introduce wavelet coefficients prediction into deep networks to generate super-resolution face images with different upscaling factors. To train deep face hallucination networks, Zhang *et al.* [41] propose a super-identify loss function to measure the difference of identity information. Cao *et al.* [30] design a novel attention-aware face hallucination framework and use deep reinforcement learning to optimize its parameters.

As a domain-specific super-resolution problem, there are also many face hallucination methods that use facial prior knowledge to help super-resolve low-resolution face images. Song *et al.* [42] propose a two-stage framework, which firstly generates facial components to represent the basic facial structures and then synthesizes fine-grained facial structures through a component enhancement method. Yu *et al.* [32] present a multi-task upsampling network to employ the image appearance similarity and exploit the face structure information with the help of the proposed facial component heat maps. Chen *et al.* [31] introduce a FSRNet model to make use of facial landmark heat maps and parsing maps. In addition, the attention mechanism [43] and bi-network [28] are also applied to make use of facial prior knowledge to train a high-resolution face generator.

2.2 Single Image Super-Resolution

The advancement of deep neural networks has achieved great success on image super-resolution, and most of the state-of-the-art SISR methods are based on deep learning [44]. As a pioneer work of deep SISR methods, Dong *et al.* [8], [45] propose a Super-Resolution Convolutional Neural Network (SRCNN) to super-resolve low-resolution images by firstly adopting deep learning for SISR. After that, many improvements have been explored. For example, Kim et al. [46] propose a deeply-recursive CNN to make use of skip connections to train their proposed a Deeply-Recursive Convolutional Network (DRCN). Lim *et al.* [11] design an Enhanced Deep Super-Resolution Network (EDSR) to remove redundant modules and combine with multi-scale processing. To reduce the computational cost, many efficient SISR methods are proposed [9], [10], [12]. To make the generated images more realistic, GAN based SISR methods are introduced to improve the perceptual quality of HR images [47], [48], [49]. Recently, Hairs *et al.* [15] develop a Deep Back-Project Network (DBPN) to exploit the mutual dependencies with a feedback mechanism. Zhang *et al.* [13] introduce dense connections to make use of cues. Residual Channel Attention Networks (RCAN) is proposed in [14] to introduce a residual-in-residual structure and a channel attention module. More recently, there are many SISR methods utilizing novel attention modules [17], [50]

TABLE 1
Representative face datasets. Most of the current public face datasets do not consider the race problem. Meanwhile, none of them provides large-scale pairs of LR and HR samples for evaluating deep hallucination methods.

Dataset	Size	Public	Race	HR-LR
LFW	13K	✓	-	×
CelebA	200K	✓	-	×
CASIA-WebFace	500K	✓	-	×
FB-DeepFace	4.4M	×	-	×
VGGFace	2.6M	✓	-	×
VGGFace2	3.3M	✓	-	×
UMDFaces	367K	✓	-	×
FaceScrub	100K	✓	-	×
MegaFace	1M	✓	-	×
FairFace	10K	✓	✓	×
MS-Celeb-1M	10M	✓	-	×
Ours	**1.7M**	✓	✓	✓

or feedback mechanisms [51] to further improve the performance of SISR. Even without considering facial structures, these above methods can also super-resolve low-resolution face images to their corresponding high-resolution versions.

2.3 Face Datasets

Currently, there does not exist publicly available face hallucination datasets. In this section, we briefly review some popular face datasets which have been constructed recently. The Labeled Faces in the Wild (LFW) dataset [52] is created in 2007, and it contains $13,000$ images. In 2014, the CelebA [53] and CASIA-WebFace [54] datasets are released, including about 20K and 500K images, respectively. The VGGface [55] dataset released in 2015 includes 2.6 million images.

More recently, Kemelmacher-Shlizerman *et al.* [56] assemble a dataset of 4.7 million images to evaluate how face recognition algorithms perform with a very large number of images. Cao *et al.* [57] release the VGGface2 dataset. Compared to the VGGface dataset, VGGface2 has 3.3 million images to cover a larger number of identities. The largest-scale face dataset is MS-Celeb-1M, which contains 10 million images for training and testing.

Even though there exist many face datasets, none of them can be directly utilized to evaluate the current face hallucination approaches, due to the following reasons. Firstly, none of these datasets provides large-scale pairs of low-resolution and high-resolution face images. However, the current deep face hallucination methods mainly rely on supervised learning and thus pairs of training face images are necessarily required. Secondly, researchers in the computer vision community have paid increasing attention to the race bias problem. However, these existing datasets are strongly biased toward specific races. Face hallucination models trained on these datasets will generate high-resolution face images with inappropriate race information.

The FairFace [58] contains face images from different races. However, it includes only 10K images without pairs of face images for the evaluation of face hallucination methods. A summary of current face datasets is listed in Table 1 to give a clear view.

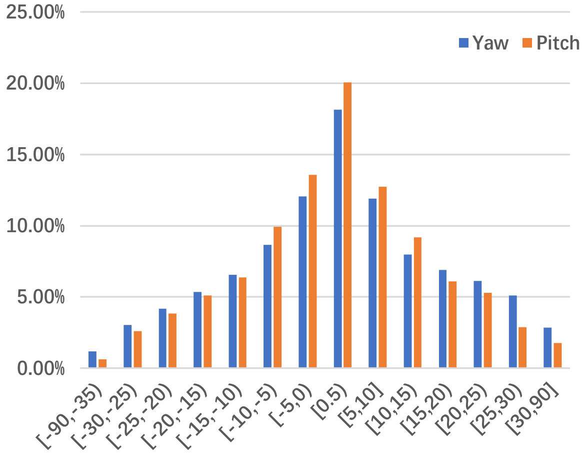


Fig. 1. **The facial pose statistics of the proposed EDFace-Celeb-1M dataset**.

To overcome the problem of lacking face hallucination datasets, current face hallucination researchers synthesize pairs of training and testing samples to evaluate the previous methods and their proposed ones. For example, Yu *et al.* [32], [36] and Kim *et al.* [43] train and test their proposed methods on pairs of face images synthesized from CelebA [59]. Zhang *et al.* [60] synthesize pairs of face images from Multi-PIE [61] and CelebA [59]. WaveletSR [29] is evaluated on synthesized face images from VGGface2 [57], and Li *et al.* [62] train and test their methods on the FFHQ [63], VGGface2 [57] and CelebA [59] datasets. However, most of these synthesized face hallucination datasets are not public. Meanwhile, given that some of them are randomly split training and testing samples, the face hallucination community faces the following challenge: all the following methods cannot directly use the pre-trained models and their results like (Peak Signal-to-Noise Ratio (PSNR)) and (Structural Similarity Index (SSIM)) values, and thus have to synthesize datasets by themselves again and re-train previous face hallucination methods. Therefore, it increases a lot of unnecessary duplication of works. In order to benchmark current deep face hallucination methods and provide convenience to the future face hallucination research, we provide a large-scale publicly available face hallucination dataset with relatively balanced race composition and contribute several baseline models in this paper.

3 EDFace-Celeb-1M Collection

In this section, we provide an overview of the EDFace-Celeb-1M dataset and introduce how it is collected in detail. We build the dataset to benchmark the current deep face hallucination methods and drive the development of the face hallucination task in the future. As mentioned above, we aim to build a publicly available large-scale face hallucination dataset, which provides pairs of low-resolution and high-resolution face images with a fixed setting of training and testing samples.

The dataset collection processing includes: how a list of candidate identities is obtained, how the candidate images are collected, how to detect the face in images, and how

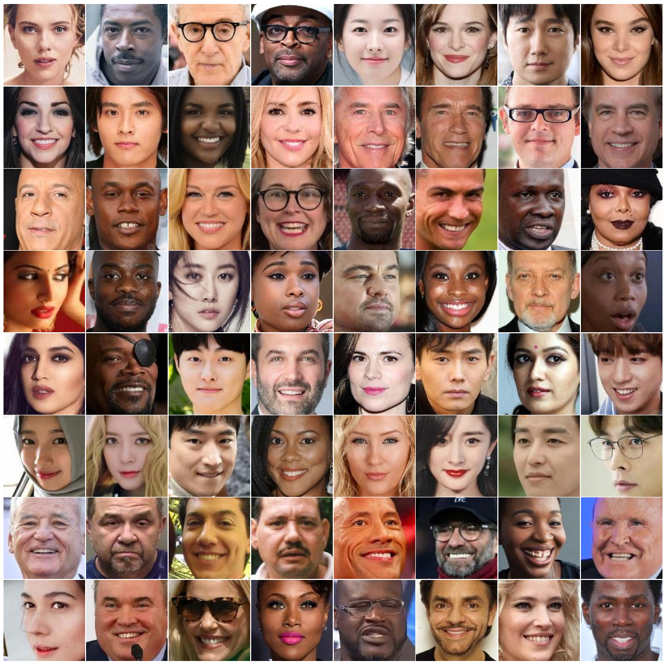


Fig. 2. **Representative face images from the proposed dataset.** These face images exhibit relatively balanced race distribution , evident pose and appearance variations.

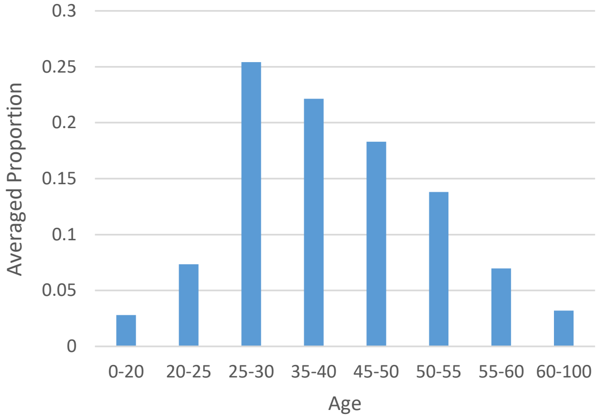


Fig. 3. **The age statistics of the proposed EDFace-Celeb-1M dataset**.

to synthesize pairs of low-resolution and high-resolution images. In addition, we provide attribute statistics of the proposed dataset such as race and gender.

3.1 Stage I: Obtain a Name List

The first step from scratch is to have a list of subject names whose faces we aim to collect. As mentioned before, race balance is the top priority when we build this dataset. To obtain such a name list with race balance, we split the races into four groups: White, Black, Asian and Latino. Based on this, we collect names for different groups. More specifically, for each group, we collect as many names of celebrities as possible from different countries. The names in our list are from diverse countries. For example, the Asian group includes names of people from East Asian, Southeast Asian, Middle East and India from Asian countries. In addition, it also includes some Asian-Americans. In summary, our EDFace-Celeb-1M dataset contains more than $20,000$ names and the ratio (name lists) of the above four groups is about $31.1\% : 19.2\% : 19.6\% : 18.3\%$.

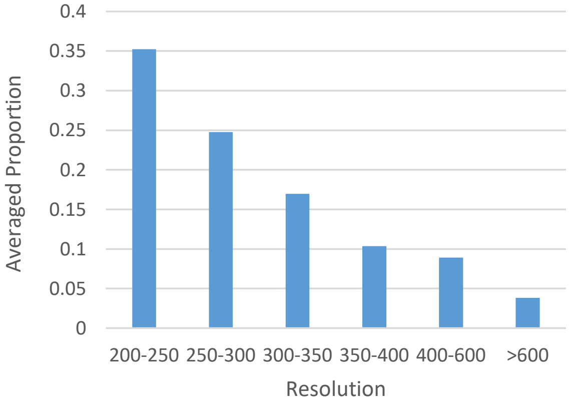


Fig. 4. **The resolution of face images before resizing**.

3.2 Stage II: Select Images for Each Identity

After obtaining a name list, we use the Google Image Search engine to download $100 \sim 1000$ images for each identity. Moreover, to obtain diverse images with age variations, we further add the keyword young to each subject and further download the corresponding images.

3.3 Stage III: Face Detection

We then detect faces in images via the Dlib detector [64]. In this way, we can obtain a facial image dataset, containing faces of different poses/angles, variations of appearance (like glasses, hat). In addition, we manually remove some non-face images.

3.4 Stage IV: Synthesize Pairs of Images

With the above steps, we have obtained about 10M facial images in total. Given the obtained facial images, we construct two subsets. The first one is composed of real-world low resolution facial images, which are dedicated to the qualitative study of existing face hallucination methods, as there is no ground truth available. Specifically, we choose images whose resolution is smaller than 50×50 to compose this subset.

The second set is dedicated to quantitatively evaluating the existing face hallucination methods, consisting of pairs of high-resolution and low-resolution facial images. To this end, we have to choose high-resolution images and synthesize the corresponding low-resolution images. To be specific, we choose 1.5M face images whose resolution is larger than 128×128 and resize them to 128×128. These images serve as high-resolution images. To synthesize the corresponding low-resolution images, we employ the strategies employed in most of the existing face hallucination methods. More specifically, we simulate the degradation process via specific operations like downsampling. The developed subset includes five different degradation settings, named as $2\times$, $4\times$, $4\times\_BD$, $4\times\_DN$ and $8\times$. The numbers indicate the downsampling factor, "D" stands for downsampling, "B" indicates blur operation, and "N " stands for Gaussian noise that is added to the LR images. The order of letters indicates the order of operations. For example, "BD" means that the blur artifact is applied prior to the downsampling operation. We use bicubic interpolation for downsampling. 2×, 4× and 8× mean that only bicubic downsampling operation is applied. When creating blurry and noisy face images, Gaussian blur and Gaussian noise are added to images. Among the 1.5M image pairs, we choose 1.36M pairs of images for training and 0.14M pairs of images for quantitative testing.

In summary, by conducting the above four steps, we derive a dataset including two subsets of non-aligned facial images. The first one contains 200K real-world low-resolution images for qualitative testing, and the second one includes 1.36M and 140K pairs of images for training and quantitative testing.

3.5 The Statistics of EDFace-Celeb-1M

Lastly, we present the statistics of specific attributes of the developed EDFace-Celeb-1M dataset as follows.

- First, we show some representative high-resolution face images from the EDFace-Celeb-1M dataset in Figure 2. It is obvious that our dataset includes different races including White, Black, Asian and Latino, from different countries. In addition, these faces exhibit evident pose and appearance variations (e.g. glasses). Figure 1 presents the statistics of the face pose.
- Second, we demonstrate the ratios of races and gender. 64% of the images of our dataset are from males and the rest 36% images are from females. The ratios of White, Black, Asian and Latino are about 31.1%, 19.2%, 19.6% and 18.3%, respectively, which are relatively balanced compared with existing datasets of face images.
- Third, we also analyze the distribution of age of the celebrities we include in our dataset, and show the results in Figure 3. The age of celebrities is estimated by the model from Insightface. Roughly, the majority of age is between 25 and 55, which aligns well with the property of demography. Notably, our dataset includes also celebrities younger than 20 and older than 60.
- Fourth, Figure 4 presents the statistics of the resolution of face images before the resizing operation. In general, all the ground-truth high-resolution images are resized from images with resolution greater than 128 × 128.
- Fifth, the proposed dataset provides fixed training and testing subsets. Each subsets includes high-quality images and corresponding low-quality images. All images are labeled as HR and LR with factors(*e.g.*, 2×, 4×, $4 \times \_BD$, $4 \times \_DN$ and 8×). Based on the labels and fixed subsets, the dataset can make the reproducibility and the fair comparison easier.

4 Experiments

In this section, we introduce the evaluation protocols and benchmark the existing face hallucination methods and representative SISR methods on the proposed EDFace-Celeb-1M dataset.

4.1 Evaluated Methods

In this benchmark study, we investigate four state-of-the-art face hallucination methods (*i.e.*, Deep Iterative Collaboration Network (DICNet) [33], Deep Iterative Collaboration GAN (DICGAN) [33], Wavelet-based Super-Resolution Network (WaveletSRNet) [29] and HiFaceGAN [65]), and four SISR methods (*i.e.*, EDSR [11], Holistic Attention Network (HAN) [17], Residual Dense Network (RDN) [13], and Residual Channel Attention Network (RCAN) [14]). All methods are based on deep learning. Specifically, DICGAN [33] is an iterative framework of recurrently estimating landmarks and recovering high-resolution images. Wavelet coefficients are introduced in WaveletSRNet [29] to deal with very-low-resolution facial images. HiFaceGAN [65] is proposed to formulate the face restoration task as a generation problem guided by semantics, and this problem is addressed by a multi-stage framework containing several units of collaborative suppression and replenishment. For the SISR task, EDSR is an enhanced deep super-resolution network containing several Resblocks. HAN is proposed in [17] to model the correlation among different convolution layers with a layer attention module and channel-spatial attention module. RDN [13] is a residual dense network to exploit the hierarchical features from both the local and global perspectives. RCAN [14] is a deep residual channel attention network with both short and long skip connections. Channel attention is also adopted in this network for better performance. In summary, we select the above eight methods for three criteria. First, these methods achieve high values of the commonly used metrics like PSNR and SSIM. Second, the codes of these methods are publicly available. Third, these methods are proposed for face hallucination or image super-resolution.

4.2 Implementation Details

Our dataset has four different degradation settings. Each setting corresponds to pairs of low-resolution and high-resolution face images, which are used to train different models. We use the code released from the original publications. For fair comparisons, the learning rate and epoch number for all methods are set as 0.0001 and 20, respectively. All models are trained using V100 GPUs. We conduct the calculation of different metrics in the RGB space to access the results. During the training stage, all models are trained on the training subset and the testing subset is not used. After training, we evaluate the models on the testing subset.

4.3 Face Super-Resolution

To evaluate the four methods on face hallucination, we provide the quantitative results of PSNR and SSIM on the EDFace-Celeb-1M dataset in Table 2. Based on the PSNR and SSIM values, DICNet achieves the best performance on 4×, $4 \times \_BD$, $4 \times \_DN$ and 8× degradation settings. RCAN achieves the best performance on 2×. In terms of SSIM, the best performance values on 4×, $4 \times \_BD$, $4 \times \_DN$, and 8× are also obtained by WaveletSR. The Table 2 does not provide the results of DICNet and DICGAN on $X2$ setting because the two methods do not work on this setting.

TABLE 2

Performance comparison of representative methods for face hallucination on the EDFace-Celeb-1M dataset. Results are reported in terms of both PSNR and SSIM. The highest, second highest and lowest results are highlighted in bolded, blue and red, respectively.

Scale	Metrics	DICNet [33]	DICGAN [33]	WaveletSR [29]	HiFaceGAN [65]	EDSR [11]	RDN [13]	RCAN [14]	HAN [17]
$2\times$	PSNR	-	-	30.60	29.68	31.23	31.39	**31.42**	31.41
	SSIM	-	-	**0.9119**	0.8836	0.8869	0.8889	0.8892	0.8888
$4\times$	PSNR	**29.06**	28.41	26.35	25.37	26.99	27.56	27.64	27.62
	SSIM	**0.8453**	0.8261	0.8211	0.7727	0.8035	0.8153	0.8161	0.8168
$4\times\_BD$	PSNR	**29.68**	28.58	26.52	24.23	27.22	27.83	27.89	27.88
	SSIM	**0.8213**	0.7988	0.7940	0.7009	0.7825	0.7955	0.7969	0.7964
$4\times\_DN$	PSNR	**27.96**	27.38	25.72	22.94	26.08	26.59	26.66	26.62
	SSIM	**0.8117**	0.7922	0.7815	0.6856	0.7673	0.7817	0.7835	0.7851
$8\times$	PSNR	**25.29**	24.64	22.33	21.88	23.24	23.74	23.77	23.73
	SSIM	**0.7453**	0.7134	0.6758	0.6408	0.6890	0.7117	0.7114	0.7114

TABLE 3

Performance comparison of representative methods for face hallucination on the EDFace-Celeb-1M dataset. Results are reported in terms of the errors of face alignment, face parsing and identity information. The highest, second highest and lowest results are highlighted in bolded, blue and red, respectively.

Scale	Metrics	DICNet [33]	DICGAN [33]	WaveletSR [29]	HiFaceGAN [65]	EDSR [11]	RDN [13]	RCAN [14]	HAN [17]
$2\times$	Alignment	-	-	0.0218	0.0223	0.0220	0.0219	0.0220	**0.0210**
	Identity	-	-	**0.9813**	0.9773	0.9800	0.9811	0.9812	**0.9813**
$4\times$	Alignment	0.0259	0.0272	0.0262	0.0295	0.0264	0.0254	0.0253	**0.0250**
	Identity	0.8360	0.8184	0.8338	0.7875	0.8262	0.8490	**0.8543**	0.8527
$4\times\_BD$	Alignment	0.0257	0.0277	0.0258	0.0312	0.0261	0.0252	**0.0250**	0.0251
	Identity	0.7846	0.7726	0.7055	0.6758	0.7786	0.8045	**0.8109**	0.8087
$4\times\_DN$	Alignment	0.0289	0.0312	0.0292	0.0362	0.0298	0.0281	0.0280	**0.0279**
	Identity	0.7035	0.6797	0.7857	0.5344	0.6969	0.7291	**0.7332**	0.7302
$8\times$	Alignment	0.0366	0.040	0.0407	0.0440	0.0399	**0.0353**	0.0354	0.0354
	Identity	0.4830	0.448	0.4305	0.3851	0.4600	0.5094	**0.5116**	0.5081

Figures 8, 9, 10 and 11 show the visual comparison of different methods on the EDFace-Celeb-1M dataset.

4.4 Face Alignment

Face alignment [66] is an important task in the field of computer vision. Apart from the PSNR and SSIM, we also evaluate the performance of face hallucination methods via face alignment. Specially, we first use a popularly used alignment model from Dlib to estimate the 68 landmarks of high-resolution face images, and the super-resolved face images generated via different methods. Then, we use a metric commonly employed in face alignment, Normalized Root Mean Squared Error (NRMSE), to evaluate the error between the landmarks from the HR (High-Resolution) and SR (Super-Resolution) face images. A smaller NRMSE value indicates better face alignment performance, which corresponds to a better face hallucination method. We can see from Table 3 that the best performance on $2\times$, $4\times$, $4\times\_BD$, $4\times\_DN$ and $8\times$ are obtained by HAN, HAN, RCAN, HAN and RDN, respectively.

We have an interesting observation that image SR methods achieve better performance than face hallucination methods regarding the metric of face alignment. This is because image SR methods mainly consider to super-resolve the LR images and ignore the adjustment of high-level information (like landmarks). Therefore, these methods can maintain the original landmarks better. For face hallucination methods, they usually consider the high-level information to help update the super-resolving process, which may cause the change of landmark.

4.5 Face Identity Information

Facial identity information is important during face hallucination. An ideal face hallucination approach should ensure that the ID information of HR face images and SR face images are the same. However, in the real world, it is difficult for the face hallucination methods to generate SR results that are the same as HR images. In this section, we use a popular face ID information extractor [7] to extract ID information from HR face images and SR face images generated by different face hallucination methods. We then calculate the cosine distance between them. The larger value indicates that the ID information loss is less, which corresponds to a better face hallucination method. Table 3 shows the results of different models. We can see that the best performance on $2\times$, $4\times$, $4\times\_BD$, $4\times\_DN$ and $8\times$ are obtained by HAN & WaveletSR, RCAN, RCAN, RCAM and RCAN, respectively.

Based on results from Table 3, we can find that face hallucination methods do not show better performance than image SR methods regarding the metric of face identification. We suspect that both of the two kinds methods do not consider the loss functions about identification information during the training stage. In the future, it may be a meaningful direction for researchers to study face identification information for the task of face hallucination. Finally, we extract similarity scores across different people's images before hallucination, and compare that with extracted similarity scores across different people's images after hallucination (by RCAN method) and their high-resolution versions. The Figure 6 shows the similarity.

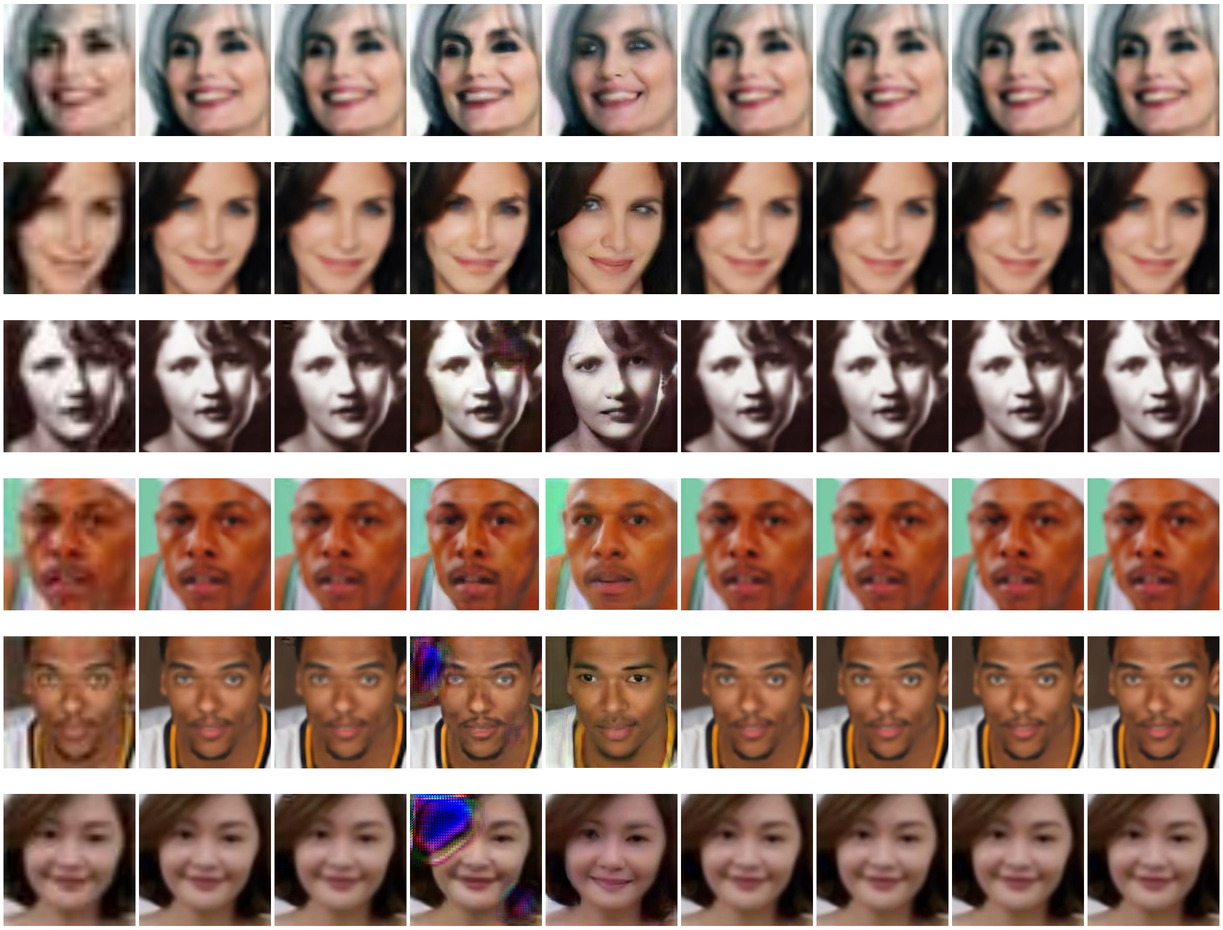


Fig. 5. **Face hallucination results on the real-world images.**. From left to right: results of bicubic, DICNet, DICGAN, WaveletSR, HiFaceGAN, EDSR, RDN, RCAN and HAN.

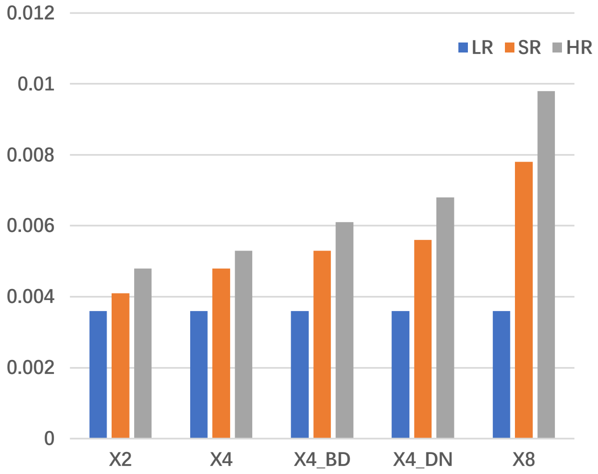


Fig. 6. **An increase in similarity across different people after the hallucination.**

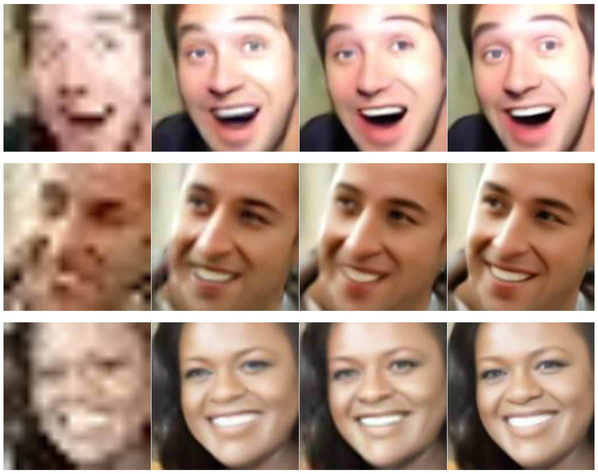


Fig. 7. **Face hallucination results trained on different numbers of training samples.** From left to right, the columns show results of the DICNet trained using $0.3M$, $0.7M$ and $1.35M$ training samples from the EDFace-Celeb-1M dataset.

4.6 Comparison on the Real LR Face Images

In addition, we also show the performance of the evaluated methods in the case of real-world scenarios based on our EDFace-Celeb-1M dataset. Taking a real-world low-resolution face image from our proposed dataset, we process it by different methods to generate SR face images, and the results are shown in Figure 5. We can find that most of the current face hallucination methods can improve the quality of low-resolution images.

4.7 Impacts of Training Samples

Given that the proposed EDFace-Celeb-1M is a large-scale public face hallucination dataset, we conduct an experimen-

TABLE 4
Performance comparison of DICNet [33] trained using different numbers of training samples from the EDFace-Celeb-1M dataset. Results are reported in terms of PSNR and SSIM. $0.3M$, $0.7M$ and $1.35M$ represent the size of training samples. The highest, second highest and lowest results are highlighted in bolded, blue and red, respectively.

Scale	Metrics	$0.3M$	$0.7M$	$1.35M$
$4\times$	PSNR	28.42	28.78	**29.06**
	SSIM	0.8265	0.8362	**0.8453**
$4\times\_BD$	PSNR	28.93	29.36	**29.68**
	SSIM	0.8053	0.8162	**0.8213**
$4\times\_DN$	PSNR	27.23	27.68	**27.96**
	SSIM	0.7986	0.8073	**0.8117**
$8\times$	PSNR	24.63	24.95	**25.29**
	SSIM	0.7216	0.7376	**0.7453**

tal study to explore the effect of the size of training samples for face hallucination. Figure 7 and Table 4 show that models can achieve better performance with an increasing number of training samples.

5 Discussions

In this section, we discuss the quality of the dataset and the effects we have made to fairly evaluate the current methods.

5.1 The Quality of the Dataset

In summary, we provide an important dataset for the face hallucination community. The literature currently lacks a dataset specific to the face hallucination task and the proposed dataset makes the reproducibility and the fair comparison much easier for further research. To ensure the quality of this dataset, we make several efforts. First, to improve the quality of facial images in our dataset, we use a face detector to obtain facial images. In addition, we also manually remove some low-quality facial images. Second, to improve models' reproducibility on the dataset, we fixed the two training and testing subsets. As a comparison, previous methods randomly choose the two subsets. Third, different from previous methods which only provide one or two degraded settings, the provided dataset provides five degraded settings. In this way, the proposed dataset can better evaluate the performance of different face hallucination methods. Fourth, different from previous methods which ignore the problem of ethnicity, the proposed dataset provides a relatively balanced race composition. Fifth, different from some previous methods which evaluate their methods on relatively small-scale datasets, the provided dataset is the largest publicly available face hallucination dataset. In future, we will consider building more datasets to benefit the development of face hallucination, including using stronger face detectors to better detect faces and providing video sequences to help learn spatio-temporal information.

5.2 Fairness

In summary, to fairly evaluate the current methods, this paper makes some significant efforts. First, we build a large-scale publicly available dataset and fix the training and testing sets to conduct experiments, rather than randomly choose samples. Second, we use the popular metrics including PSNR and SSIM to compare different methods. Third, to further evaluate the current methods, we design two task-driven ablation studies including landmark detection and identify preservation.

6 Conclusions

In this paper, we first propose the largest publicly available face hallucination dataset with relatively balanced race composition. It contains 1.5 million pairs of LR and HR face images for training and testing, and 140K real-world tiny face images for quantitative comparisons. Thanks to the proposed EDFace-Celeb-1M dataset, the following face hallucination can evaluate their methods on a public and fixed division of training and testing samples, which significantly makes the comparison convenient and improves the reliability.

In addition, given that the current face hallucination methods are evaluated on privately synthesized datasets, we **benchmark** four public available face hallucination methods, and four SISR methods on the proposed EDFace-Celeb-1M dataset. The proposed dataset will be made publicly available to encourage the development of face hallucination algorithms.

References

[1] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476–3483.

[2] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 146–155.

[3] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1021–1030.

[4] S. Liu, J. Yang, C. Huang, and M.-H. Yang, "Multi-objective convolutional learning for face labeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3451–3459.

[5] J. Lin, H. Yang, D. Chen, M. Zeng, F. Wen, and L. Yuan, "Face parsing with roi tanh-warping," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5654–5663.

[6] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5265–5274.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*. Springer, 2014, pp. 184–199.

[9] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 391–407.

[10] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[11] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

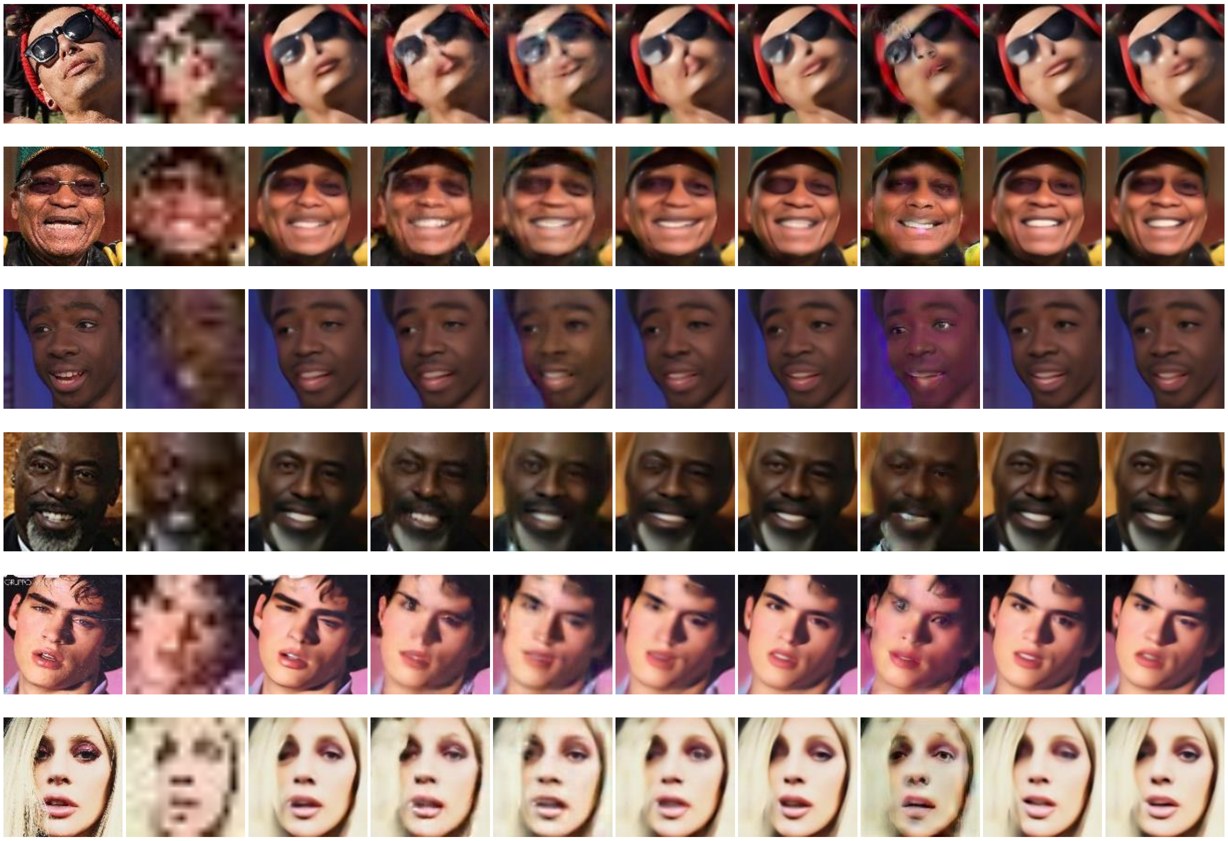


Fig. 8. **Visual results of BI models ($\times 8$) on the EDFace-Celeb-1M dataset**. From left to right: HR, results of bicubic, DICNet, DICGAN, WaveletSR, HiFaceGAN, EDSR, RDN, RCAN and HAN. “BI” means the bicubic interpolation.

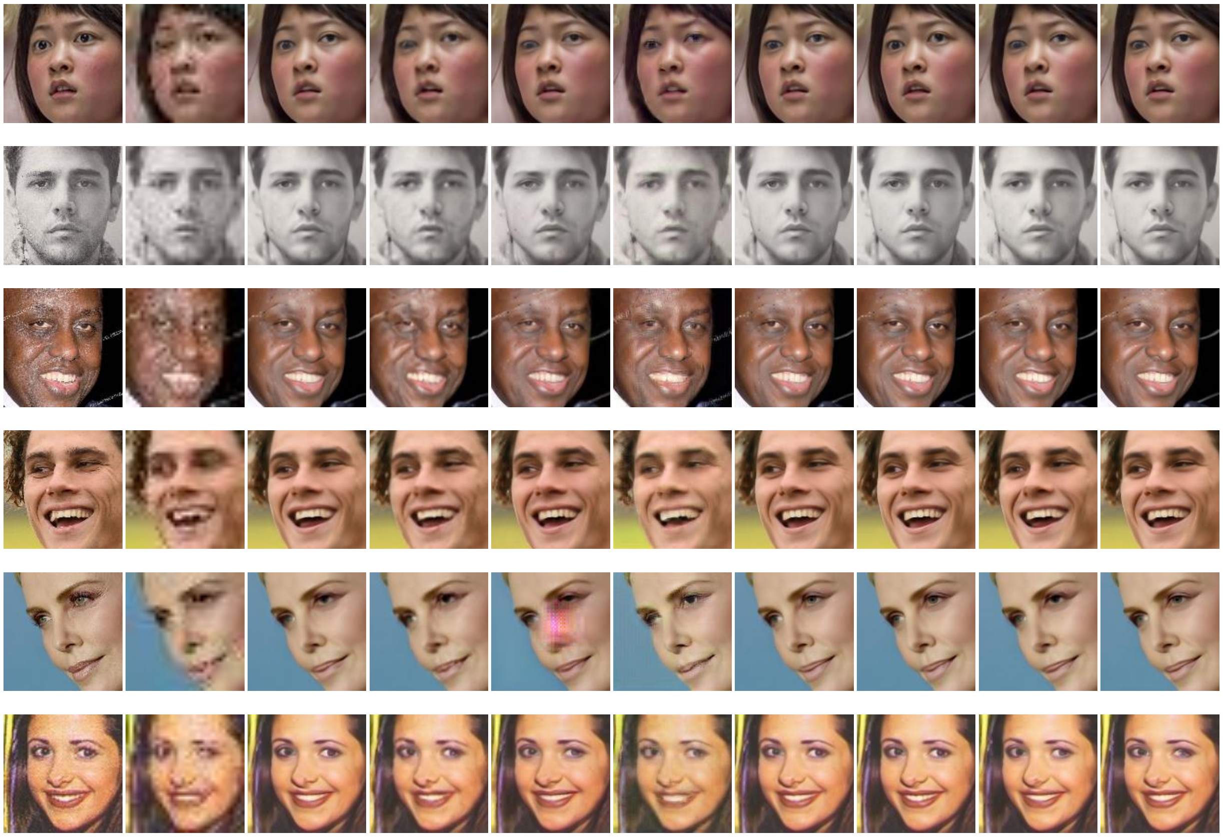


Fig. 9. **Visual results of BI models ($\times 4$) on the EDFace-Celeb-1M dataset**. From left to right: HR, results of bicubic, DICNet, DICGAN, WaveletSR, HiFaceGAN, EDSR, RDN, RCAN and HAN. “BI” means the bicubic interpolation.

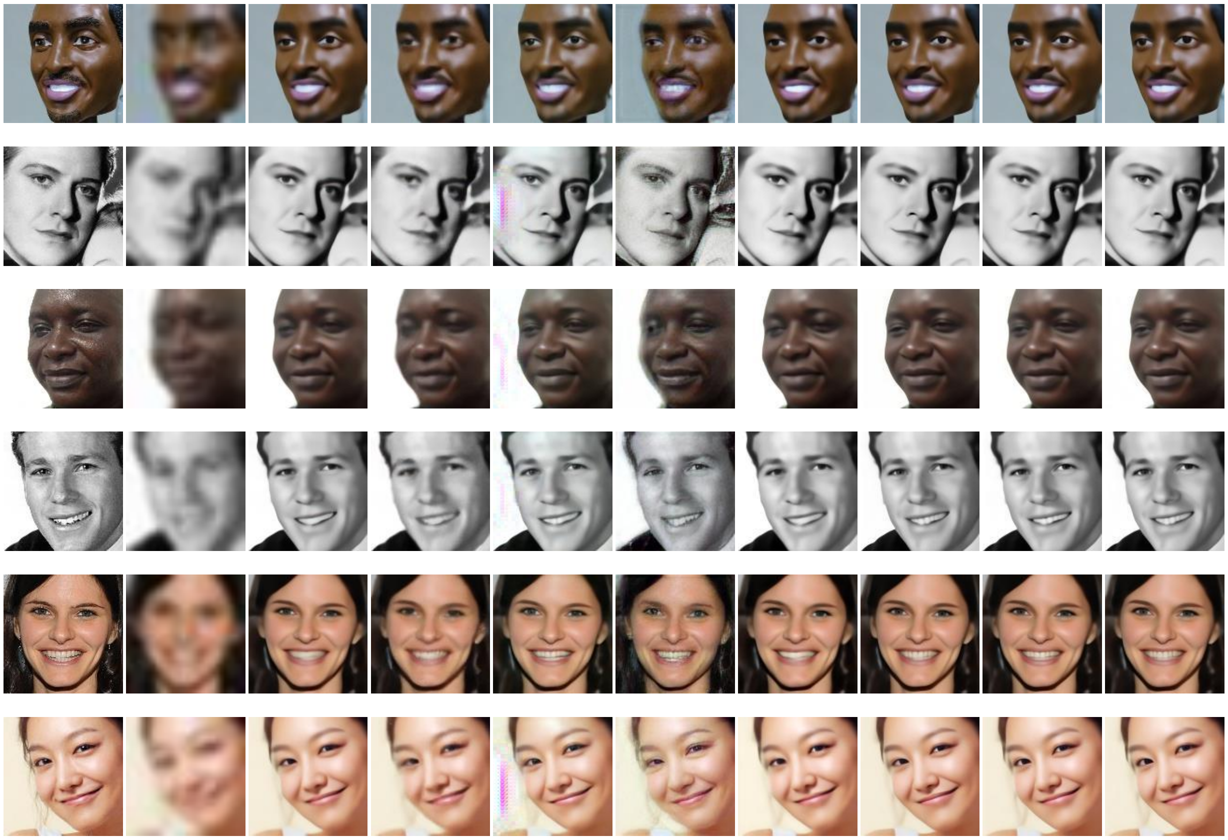


Fig. 10. **Visual results of BD models ($\times 4$) on the EDFace-Celeb-1M dataset**. From left to right: HR, results of bicubic, DICNet, DICGAN, WaveletSR, HiFaceGAN, EDSR, RDN, RCAN and HAN. “BD” means that the blur artifact is applied prior to the downsampling operation.

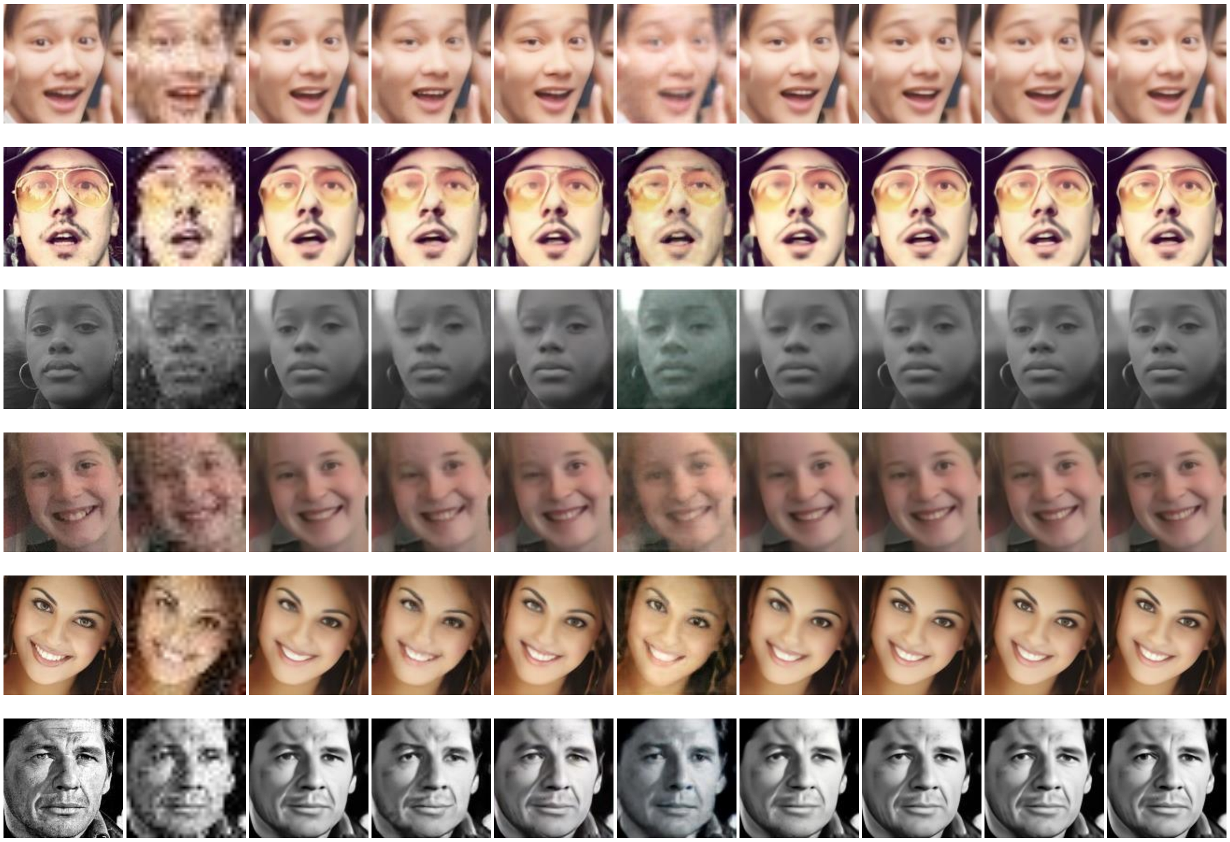


Fig. 11. **Visual results of DN models ($\times 4$) on the EDFace-Celeb-1M dataset**. From left to right: HR, results of bicubic, DICNet, DICGAN, WaveletSR, HiFaceGAN, EDSR, RDN, RCAN and HAN. “DN” means that Gaussian noise is added to the LR images after the downsampling operation.

[12] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 624–632.

[13] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[14] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 286–301.

[15] M. Haris, G. Shakhnarovich, and N. Ukita, “Deep back-projection networks for super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1664–1673.

[16] S. Anwar and N. Barnes, “Densely residual laplacian super-resolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[17] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen, “Single image super-resolution via a holistic attention network,” in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 191–207.

[18] A. Chakrabarti, A. Rajagopalan, and R. Chellappa, “Super-resolution of face images using kernel pca-based prior,” *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007.

[19] C. Liu, H.-Y. Shum, and W. T. Freeman, “Face hallucination: Theory and practice,” *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[20] K. Jia and S. Gong, “Generalized face super-resolution,” *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 873–886, 2008.

[21] H. Huang, H. He, X. Fan, and J. Zhang, “Super-resolution of human face image using canonical correlation analysis,” *Pattern Recognition*, vol. 43, no. 7, pp. 2532–2543, 2010.

[22] X. Ma, J. Zhang, and C. Qi, “Hallucinating face by position-patch,” *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.

[23] C. Jung, L. Jiao, B. Liu, and M. Gong, “Position-patch based face hallucination using convex optimization,” *IEEE Signal Processing Letters*, vol. 18, no. 6, pp. 367–370, 2011.

[24] C.-Y. Yang, S. Liu, and M.-H. Yang, “Structured face hallucination,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1099–1106.

[25] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, “A comprehensive survey to face hallucination,” *International journal of computer vision*, vol. 106, no. 1, pp. 9–30, 2014.

[26] J. Jiang, R. Hu, Z. Wang, and Z. Han, “Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning,” *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4220–4231, 2014.

[27] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, “Learning face hallucination in the wild,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.

[28] S. Zhu, S. Liu, C. C. Loy, and X. Tang, “Deep cascaded bi-network for face hallucination,” in *European conference on computer vision*. Springer, 2016, pp. 614–630.

[29] H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1689–1697.

[30] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li, “Attention-aware face hallucination via deep reinforcement learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 690–698.

[31] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, “Fsrnet: End-to-end learning face super-resolution with facial priors,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2492–2501.

[32] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, “Face super-resolution guided by facial component heatmaps,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 217–233.

[33] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou, “Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5569–5578.

[34] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, “Dual-path deep fusion network for face image hallucination,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[35] X. Yu and F. Porikli, “Ultra-resolving face images by discriminative generative networks,” in *European conference on computer vision*. Springer, 2016, pp. 318–333.

[36] X. Yu, B. Fernando, R. Hartley, and F. Porikli, “Super-resolving very low-resolution face images with supplementary attributes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 908–917.

[37] X. Wang and X. Tang, “Hallucinating face by eigentransformation,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 425–434, 2005.

[38] S. Kolouri and G. K. Rohde, “Transport-based single frame super resolution of very low resolution face images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4876–4884.

[39] M. F. Tappen and C. Liu, “A bayesian approach to alignment-based image hallucination,” in *European conference on computer vision*. Springer, 2012, pp. 236–249.

[40] C.-Y. Yang, S. Liu, and M.-H. Yang, “Hallucinating compressed face images,” *International Journal of Computer Vision*, vol. 126, no. 6, pp. 597–614, 2018.

[41] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang, “Super-identity convolutional neural network for face hallucination,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 183–198.

[42] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang, “Learning to hallucinate face images via component generation and enhancement,” *arXiv preprint arXiv:1708.00223*, 2017.

[43] D. Kim, M. Kim, G. Kwon, and D.-S. Kim, “Progressive face super-resolution via attention to facial landmark,” *arXiv preprint arXiv:1908.08239*, 2019.

[44] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[45] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[46] J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1637–1645.

[47] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.

[48] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “Esrgan: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 0–0.

[49] M. S. Sajjadi, B. Scholkopf, and M. Hirsch, “Enhancenet: Single image super-resolution through automated texture synthesis,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4491–4500.

[50] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, “Second-order attention network for single image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 065–11 074.

[51] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, “Feedback network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.

[52] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,” in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[53] Y. Sun, X. Wang, and X. Tang, “Deep learning face representation from predicting 10,000 classes,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1891–1898.

[54] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Learning face representation from scratch,” *arXiv preprint arXiv:1411.7923*, 2014.

[55] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.

[56] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, “The megaface benchmark: 1 million faces for recognition at

scale," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4873–4882.

[57] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[58] K. Kärkkäinen and J. Joo, "Fairface: Face attribute dataset for balanced race, gender, and age," *arXiv preprint arXiv:1908.04913*, 2019.

[59] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.

[60] Y. Zhang, I. W. Tsang, Y. Luo, C.-H. Hu, X. Lu, and X. Yu, "Copy and paste gan: Face hallucination from shaded thumbnails," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7355–7364.

[61] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and vision computing*, vol. 28, no. 5, pp. 807–813, 2010.

[62] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *European Conference on Computer Vision*. Springer, 2020, pp. 399–415.

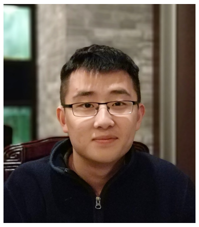
[63] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[64] D. E. King, "Dlib-ml: A machine learning toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[65] L. Yang, S. Wang, S. Ma, W. Gao, C. Liu, P. Wang, and P. Ren, "Hifacegan: Face renovation via collaborative suppression and replenishment," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1551–1560.

[66] V. Albiero, X. Chen, X. Yin, G. Pang, and T. Hassner, "img2pose: Face alignment and detection via 6dof, face pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7617–7627.

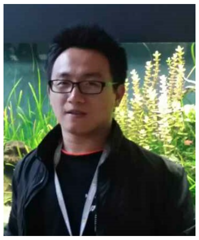
Kaihao Zhang is currently pursuing the Ph.D. degree with the College of Engineering and Computer Science, The Australian National University, Canberra, ACT, Australia. His research interests focus on computer vision and deep learning. He has more than 30 referred publications in international conferences and journals, including CVPR, ICCV, ECCV, NeurIPS, AAAI, ACMMM, TPAMI, IJCV, TIP, TMM, etc.



Dongxu Li is a Ph.D candidate at The Australian National University. His research interests are mainly computer vision and deep learning, including visual sequence representation learning, vision-language learning and multi-modal learning. Before starting PhD, Dongxu obtained his Bachelor degree from The Australian National University with first-class honours in Computing.

Wenhan Luo is currently an Associate Professor in Sun Yat-sen University. Prior to that, he worked as a research scientist for Tencent and Amazon. He has published over 40 papers in top conferences and leading journals, including ICML, CVPR, ICCV, ECCV, ACL, AAAI, ICLR, TPAMI, IJCV, TIP, etc. He also has been reviewer, senior PC member and Guest Editor for several prestigious journals and conferences. His research interests include several topics in computer vision and machine learning, such as image/video synthesis, image/video quality restoration, reinforcement learning. He received the Ph.D. degree from Imperial College London, UK, 2016, M.E. degree from Institute of Automation, Chinese Academy of Sciences, China, 2012 and B.E. degree from Huazhong University of Science and Technology, China, 2009.

Jingyu Liu is currently a research engineer in Tencent. His research interests include object recognition and multi-modal learning. He obtained his PhD degree from Institute of Automation, Chinese Academy of Science in 2018.



Jiankang Deng obtained his PhD degree from Imperial College London (ICL), supervised by Prof. Stefanos Zafeiriou and funded by the Imperial President's PhD Scholarships. His research topic is deep learning-based face analysis, including detection, alignment, reconstruction, recognition and generation etc. He is a reviewer in prestigious computer vision journals and conferences including T-PAMI, IJCV, CVPR, ICCV and ECCV. He is one of the main contributors to the widely used open-source platform Insightface. He is a student member of the IEEE.

Wei Liu is currently a Distinguished Scientist of Tencent AI Lab and a director of Computer Vision Center. Prior to that, he received the Ph.D. degree in EECS from Columbia University, New York, NY, USA, and was a research scientist of IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. Dr. Liu has long been devoted to research and development in the fields of machine learning, computer vision, information retrieval, big data, etc. Till now, he has published more than 150 peer-reviewed journal and conference papers, including Proceedings of the IEEE, TPAMI, TKDE, IJCV, NIPS, ICML, CVPR, ICCV, ECCV, KDD, SIGIR, SIGCHI, WWW, IJCAI, AAAI, etc. His research works win a number of awards and honors, such as the 2011 Facebook Fellowship, the 2013 Jury Award for best thesis of Columbia University, the 2016 and 2017 SIGIR Best Paper Award Honorable Mentions, and the 2018 "AI's 10 To Watch" honor. Dr. Liu currently serves as an Associate Editor to several international leading AI journals and an Area Chair to several international top-tier AI conferences, respectively.

Stefanos Zafeiriou is currently a Professor in Machine Learning and Computer Vision with the Department of Computing, Imperial College London, London, U.K, and an EPSRC Early Career Research Fellow. He served Associate Editor and Guest Editor in various journals including TPAMI, IJCV, TAC, CVIU, and IVC. He has been a Guest Editor of 8+ journal special issues and co-organised over 16 workshops/special sessions on specialised computer vision topics in top venues. He has co-authored 70+ journal papers mainly on novel statistical machine learning methodologies applied to computer vision problems, such as 2-D/3-D face analysis, deformable object fitting and tracking, shape from shading, and human behaviour analysis, published in the most prestigious journals in his field of research, such as TPAMI, IJCV, and many papers in top conferences, such as CVPR, ICCV, ECCV, ICML. His students are frequent recipients of very prestigious and highly competitive fellowships, such as the Google Fellowship x2, the Intel Fellowship, and the Qualcomm Fellowship x4. He has more than 20K+ citations to his work, h-index 64. He was the General Chair of BMVC 2017. He is a member of the IEEE.