

# APPENDIX

---

As a supplementary asset to the primary survey, we offer extra visual representations and informative material related to the subjects being discussed. This enhances readers' understanding of the main concepts. To start, we display a more comprehensive visualization result about the non-blind/blind methods. Next, we showcase crucial data and visualizations of outcomes related to different methodologies used in various collaborative tasks. Furthermore, we include examples of the various technical frameworks described in the main text, including GAN architectures, Prior bootstrapping structures, and more. Lastly, we provide comprehensive information about commonly used datasets for facial restoration, with the goal of improving readers' understanding of the dataset tables presented in the main text.

## VISUALIZATION OF NON-BLIND/BLEND TASKS.

Owing to spatial constraints, only a subset of the visualization results is presented in the main text. Thus, we provide additional visualization outputs of the non-blind methods using the CelebA and Helen test sets in Fig. 1, the blind methods using the CelebA-HQ test sets in Fig. 2, along with visualization outputs of the blind methods applied to a broader range of real-world datasets in Fig. 3.

## JOINT TASKS.

In this section we give examples of some visual comparisons of joint face restoration tasks summarized in the main text. Fig. 5 provides a visual representation of the outcomes yielded by certain approaches concerning the Joint Face Completion task. Moving to Fig. 6, it showcases the outcomes achieved by various methods in the context of the Joint Face Frontalization task. Fig. 7 offers visual insights into the results obtained by some techniques for the Joint Face Alignment task. Similarly, Fig. 8 illustrates the results obtained by some methodologies for the Joint Face Recognition task. Furthermore, we introduce TABLE 1 and TABLE 2 as supplementary resources, aiding readers in better comprehending the assistance rendered by these methods

in enhancing face recognition rates. Shifting to Fig. 9, it presents visual depictions of the outputs generated by several techniques targeting the Joint Face Illumination Compensation. Lastly, Fig. 11 offers visual results of some methods concerning the Joint Face Fairness.

## TECHNICAL FRAMEWORKS

### Prior Guide Approach.

In this section, we first delve into the examination of bootstrapping methods aimed at enhancing the validity and robustness of prior knowledge. Inspired by [1], we categorize priors bootstrapping can be categorized into the following five categories: pre-prior, in-prior, parallel-prior, input-prior, and post-prior. Each of these categories represents different strategies for incorporating face prior into the restoration, potentially leading to varying outcomes.

- **Pre-prior.** As depicted in Fig. 4 (a), pre-prior methods [2]–[6] typically involve estimating the prior of the LQ face image before the restoration process. They guide the restoration by utilizing extracted prior information as inputs to the restoration network. It enables the network to produce more accurate and contextually relevant results based on the prior knowledge. However, this type of method extracts the prior directly from the LQ face image, which can limit the accuracy of the prior extraction and subsequently restrict the overall performance of the restoration.

- **In-prior.** In-prior approach [5], [7]–[9] is developed to address the disadvantage of inaccurate prior estimation. As depicted in Fig. 4 (b), it involves adding a restoration network before the prior estimation network. This initial restoration network is responsible for roughly recovering the LQ face image. Then, the prior information is extracted from this intermediate feature. Finally, an additional restoration network is used to complete the fine face restoration, utilizing the extracted prior information. While the In-prior approach brings performance gains in FR tasks, it comes with a significant increase in computational consumption.

- **Parallel-prior.** The above structure overlooks the correlation between prior estimation and face recovery.

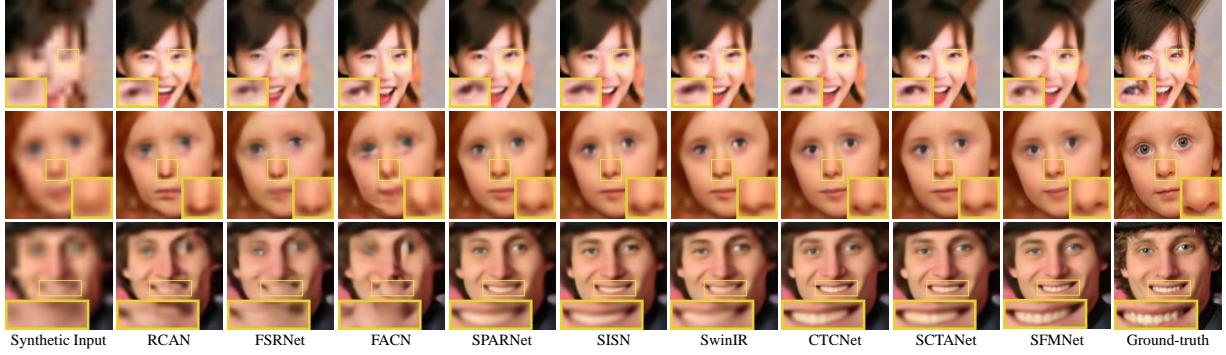


Figure 1: Visual comparison of different non-blind methods on the CelebA test set and Helen test set.

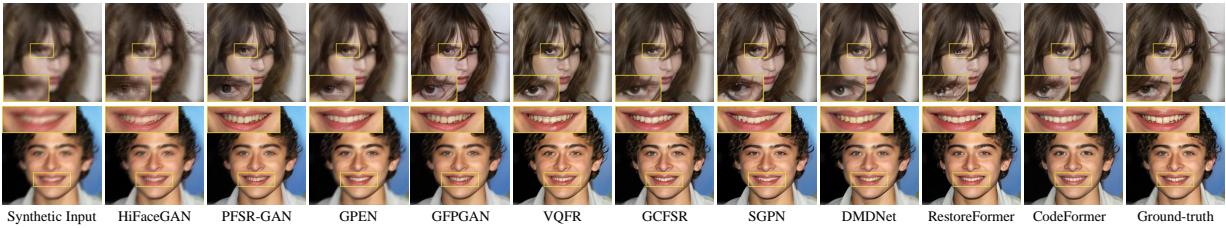


Figure 2: Visual comparison of different blind methods on the CelebA-HQ test set for blind face restoration.

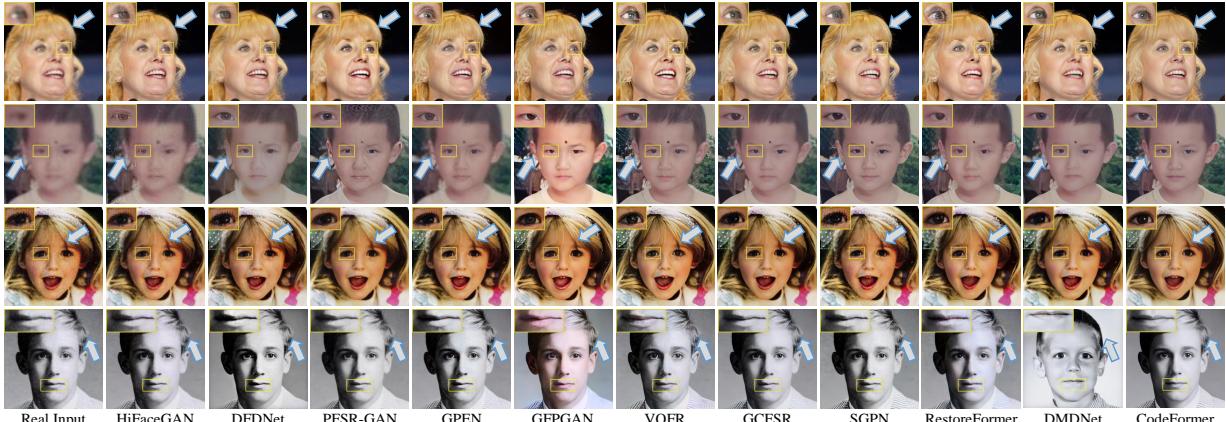


Figure 3: Qualitative comparison of restoration for the real test sets, including LFW-Test (first row), WebPhoto (second row), Celeb-Child (third row), and Celeb-Adult (fourth row).

In the Parallel-prior approach [10]–[13], as shown in Fig. 4 (c), the method first extracts the main features of the face using a shared feature extraction network. Then, it jointly trains both the recovery network and the prior estimation network for the two tasks. Some methods [12] in this category also feed features from the prior estimation branch into the recovery network to leverage the correlation and achieve better results. But this type of methods may fail to fully utilize the power of prior guidance.

- **Input-prior.** To address the challenge of estimating

prior directly or indirectly from LQ faces, the input-prior approach [14]–[17], as shown in Fig. 4 (d), takes both a LQ face and a other HQ face image of the same person as inputs, utilizes a reference HQ face as prior to aid the restoration process. This approach leverages more bootstrapping features, which can help restoration and alleviate computational consumption. However, a drawback is that it requires different face images of the same person, which increases the difficulty of training and inference.

- **Post-prior.** As shown in Fig. 4 (e), post-prior [18]–[20]

approach estimates the prior from the recovered face images and utilizes the feedback of the estimation accuracy to regulate the recovery process. This approach mainly focuses on designing the loss function for the prior estimation. This approach has the advantage of eliminating the prior estimation process during inference, making it faster in inference than other methods. However, the indirect feedback process may lead to untimely updates and potentially result in suboptimal restoration quality.

### Frameworks.

Then we will give some examples of the methodological structure of the framework summarized in the main text. Regarding the GAN structure, Fig. 12, Fig. 13, and Fig. 14 show examples of using the general GAN structure, the Pre-trained GAN structure, and using the unpaired images GAN structure, respectively. In addition, regarding the prior guided structure, Fig. 15, Fig. 16, Fig. 17 Fig. 18 and Fig. 19 show examples of the use of the Pre-prior framework, the In-prior framework, the Parallel-prior framework, the Input-prior framework, and the Post-prior framework, respectively.

## DATASETS

In this section, we provide a detailed overview of the dataset utilized for face restoration.

- **LFW** [21] contains a large collection of labeled face images collected from the internet, with a focus on unconstrained, real-world conditions. And it consists of more than 13,000 images of over 5,700 individuals, capturing a wide range of variations in pose, lighting, expression, and occlusion.
- **Multi-PIE** [22] features a large collection of face images captured under controlled lighting conditions and from multiple viewing angles. The dataset contains images of over 337 subjects, with each subject captured under 15 different viewpoints and 20 different lighting conditions.
- **AFLW** [23] contains a large collection of facial images that are taken "in the wild", and includes images of various ethnicities, ages, and gender, making it suitable for training and evaluating algorithms in real-world scenarios.
- **SCFace** [24] is a set of 4,160 still face images containing 130 subjects, and the face images were captured in an uncontrolled indoor environment using five video surveillance cameras of different quality.
- **Helen** [25] contains 2,330 face images and is specifically designed for the task of intensive landmark annotation with landmarks covering a variety of facial features, including eyes, eyebrows, nose, mouth, and jawline.
- **300W** [26] is a widely used benchmark dataset in the field of facial landmark detection and face alignment.
- **CASIA-WebFace** [27] consists 494,414 high-quality aligned face images of over 10,000 unique subjects, each subject has multiple images taken under different conditions.
- **CelebA** [28] consists of more than 200,000 celebrity images, known for their diversity in terms of gender, age and ethnicity. And each image is labeled with 40 different attribute annotations and bounding box annotations around the face.
- **Widerface** [29] contains 32,203 images containing bounding boxes of face locations and encompasses different scenes with faces of different scales and orientations.
- **IMDB-WIKI** [30] consists of over 500,000 face images collected from IMDb and Wikipedia databases, with name, age and gender. However, it may contain some bias, as these images are mainly of individuals from certain regions.
- **VGGFace** [31] contains 2.6 million images of over 2,600 people, including celebrities, public figures and ordinary people. And these images vary in terms of pose, lighting conditions and facial expressions.
- **LS3D-W** [32] consists of a large number of 2D face images and corresponding 3D facial landmark annotations, with the main purpose of facilitating the development and evaluation of face alignment and 3D face reconstruction algorithms.
- **VGGFace2** [33] consists of more than 3.3 million images from over 9,000 people and is one of the largest publicly available face recognition datasets.
- **IJB-C** [34] consists of over 138,000 still images and 1,400 video clips covering a wide range of variations in pose, expression, lighting conditions, occlusion, and resolution.
- **FFHQ** [35] contains over 70,000 face images with a resolution of  $1024 \times 1024$  pixels, covering a diverse range of individuals, including people of different ages, genders, races and facial expressions.
- **CelebAMask-HQ** [36] is an extension of the CelebA [28] and contains over 30,000 celebrity face images with a resolution of  $512 \times 512$  pixels. Each of its images is annotated with pixel-level semantic segmentation masks and provides detailed information about different facial regions and attributes.
- **EDFace-Celeb-1M** [37] is a benchmark dataset dedicated to face restoration, containing 1.7 million pairs of low-quality and high-quality face images, covering faces of different ethnicities from different regions.
- **CelebRef-HQ** [17] contains 10,555 images of 1,005 identities covering a wide range of ages, genders, ethnicities, backgrounds, poses and expressions.

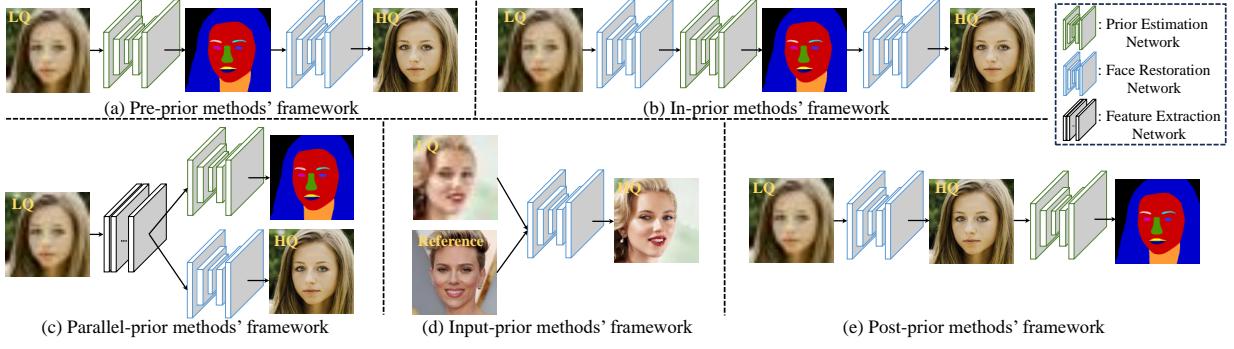


Figure 4: Summary of prior-guided methods’ architecture used for face restoration.

Table 1: Comparison of face recognition rates evaluated by OpenFaces [53] for HR faces reconstructed by various **joint face recognition methods** on [27] by upscaling: (a) from  $8 \times 8$  to  $32 \times 32$ ; (b) from  $16 \times 16$  to  $64 \times 64$

Method	Top-1	Top-5	Top-10
HR ( $32 \times 32$ )	30.4%	51.2%	59.6%
LR ( $8 \times 8$ )	10.7%	19.5%	33.1%
Bicubic	10.8%	20.1%	34.4%
DFCG [2]	9.3%	17.7%	21.4%
UR-DGN [54]	9.9%	18.6%	22.7%
DCGAN [55]	4.6%	10.9%	16.8%
PRSR [56]	10.8%	18.8%	24.4%
SR-GAN [44]	8.8%	11.1%	19.4%
Wavelet-SRNet [57]	12.8%	20.2%	30.3%
SiGAN	15.8%	27.5%	40.4%

Method	Top-1	Top-5	Top-10
HR ( $64 \times 64$ )	36.8%	55.9%	63.8%
LR ( $16 \times 16$ )	12.4%	27.4%	37.1%
Bicubic	11.6%	27.5%	37.6%
DFCG [2]	9.6%	23.7%	34.8%
UR-DGN [54]	12.2%	29.0%	38.7%
DCGAN [55]	9.3%	24.9%	33.9%
PRSR [56]	13.3%	29.7%	40.1%
SR-GAN [44]	11.6%	23.2%	36.3%
Wavelet-SRNet [57]	12.0%	25.5%	38.8%
SiGAN	17.9%	32.9%	48.1%

## REFERENCES

- | Method                | Top-1 | Top-5 | Top-10 |
|-----------------------|-------|-------|--------|
| HR ( $32 \times 32$ ) | 32.2% | 50.8% | 56.7%  |
| LR ( $8 \times 8$ )   | 9.3%  | 17.4% | 30.9%  |
| Bicubic               | 9.6%  | 17.7% | 30.4%  |
| DFCG [2]              | 9.3%  | 16.9% | 27.5%  |
| UR-DGN [54]           | 7.9%  | 16.8% | 20.1%  |
| DCGAN [55]            | 4.7%  | 9.9%  | 14.6%  |
| PRSR [56]             | 10.3% | 19.8% | 26.1%  |
| SR-GAN [44]           | 9.1%  | 13.3% | 22.6%  |
| Wavelet-SRNet [57]    | 13.1% | 22.7% | 32.0%  |
| SiGAN (proposed)      | 14.5% | 26.7% | 39.2%  |
- 
- | Method                | Top-1 | Top-5 | Top-10 |
|-----------------------|-------|-------|--------|
| HR ( $64 \times 64$ ) | 35.4% | 51.4% | 60.1%  |
| LR ( $16 \times 16$ ) | 14.8% | 26.6% | 35.3%  |
| Bicubic               | 15.0% | 26.4% | 35.6%  |
| DFCG [2]              | 13.2% | 25.4% | 34.7%  |
| UR-DGN [54]           | 15.9% | 30.2% | 39.4%  |
| DCGAN [55]            | 11.6% | 24.3% | 32.6%  |
| PRSR [56]             | 18.3% | 32.6% | 45.5%  |
| SR-GAN [44]           | 12.6% | 26.5% | 38.8%  |
| Wavelet-SRNet [57]    | 15.1% | 27.1% | 40.2%  |
| SiGAN (proposed)      | 21.5% | 40.5% | 50.2%  |

- [1] J. Jiang, C. Wang, X. Liu, and J. Ma, “Deep learning-based face super-resolution: A survey,” *ACM Computing Surveys*, vol. 55, no. 1, pp. 1–36, 2021. [1](#)
- [2] 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. [1, 4, 9](#)
- [3] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, “Deep semantic face deblurring,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269. [1, 9](#)
- [4] R. Kalarot, T. Li, and F. Porikli, “Component attention guided face super-resolution network: Cagface,” in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 370–380. [1](#)

Table 2: Comparison of face recognition rates evaluated by OpenFaces [53] for HR faces reconstructed by various **joint face recognition methods** on LFW [21] by upscaling: (a) from  $8 \times 8$  to  $32 \times 32$ ; (b) from  $16 \times 16$  to  $64 \times 64$

- [5] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, “Exploiting semantics for face image deblurring,” *International Journal of Computer Vision*, vol. 128, pp. 1829–1846, 2020. [1](#)
- [6] C. Chen, X. Li, L. Yang, X. Lin, L. Zhang, and K.-Y. K. Wong, “Progressive semantic-aware style transformation for blind face restoration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 896–11 905. [1](#)
- [7] 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. [1, 8, 13](#)
- [8] Y. Zhang, Y. Wu, and L. Chen, “Msfsr: A multi-stage face super-resolution with accurate facial representation via enhanced facial boundaries,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 504–505. [1](#)
- [9] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou, “Deep face super-resolution with iterative collaboration between attentive



Figure 5: Qualitative comparison of some state-of-the-art **joint face completion methods** on the CelebA-HQ dataset. From left to right, we demonstrate the ground-truth image, the masked image, and the inpainting results from EC [38], RFR [39], Lafin [40], and FT-TDR [41] with the predicted mask, respectively.

- recovery and landmark estimation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5569–5578. [1, 13](#)
- [10] S. Zhu, S. Liu, C. C. Loy, and X. Tang, “Deep cascaded bi-network for face hallucination,” in *Proceedings of the European Conference on Computer Vision*. Springer, 2016, pp. 614–630. [2, 6, 8](#)
- [11] Y. Yin, J. Robinson, Y. Zhang, and Y. Fu, “Joint super-resolution and alignment of tiny faces,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 693–12 700. [2, 14](#)
- [12] M. Li, Z. Zhang, J. Yu, and C. W. Chen, “Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement,” *IEEE Transactions on Multimedia*, vol. 23, pp. 468–483, 2020. [2](#)
- [13] J. Li, B. Bare, S. Zhou, B. Yan, and K. Li, “Organ-branched cnn for robust face super-resolution,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6. [2](#)
- [14] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang, “Learning warped guidance for blind face restoration,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 272–289. [2, 14](#)
- [15] B. Dogan, S. Gu, and R. Timofte, “Exemplar guided face image super-resolution without facial landmarks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0. [2](#)
- [16] X. Li, W. Li, D. Ren, H. Zhang, M. Wang, and W. Zuo, “Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2706–2715. [2](#)
- [17] X. Li, S. Zhang, S. Zhou, L. Zhang, and W. Zuo, “Learning dual memory dictionaries for blind face restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. [2, 3](#)
- [18] A. Bulat and G. Tzimiropoulos, “Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 109–117. [2, 9, 14](#)
- [19] D. Kim, M. Kim, G. Kwon, and D.-S. Kim, “Progressive

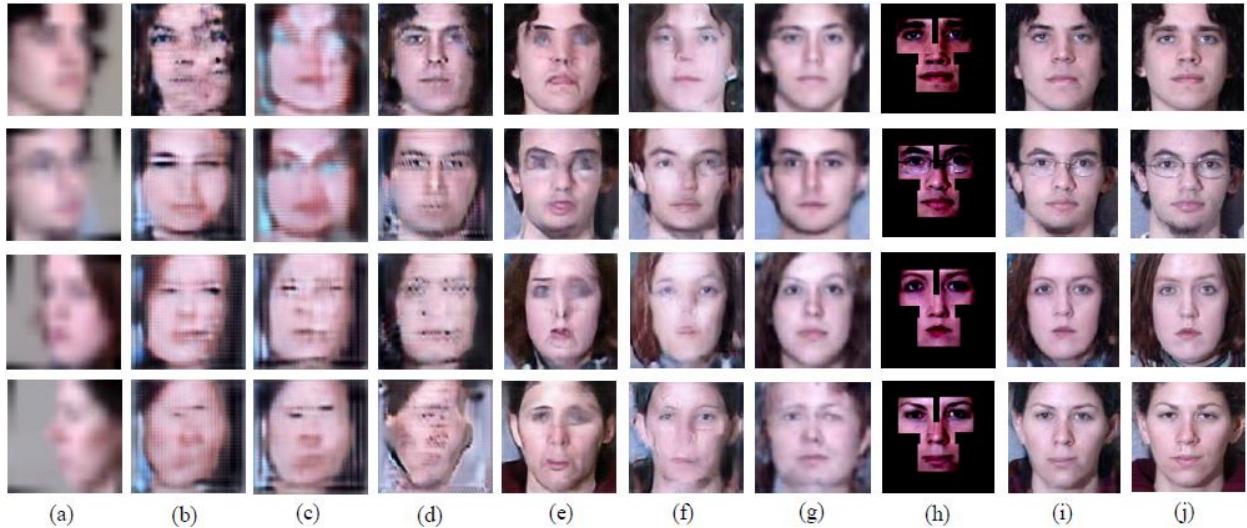
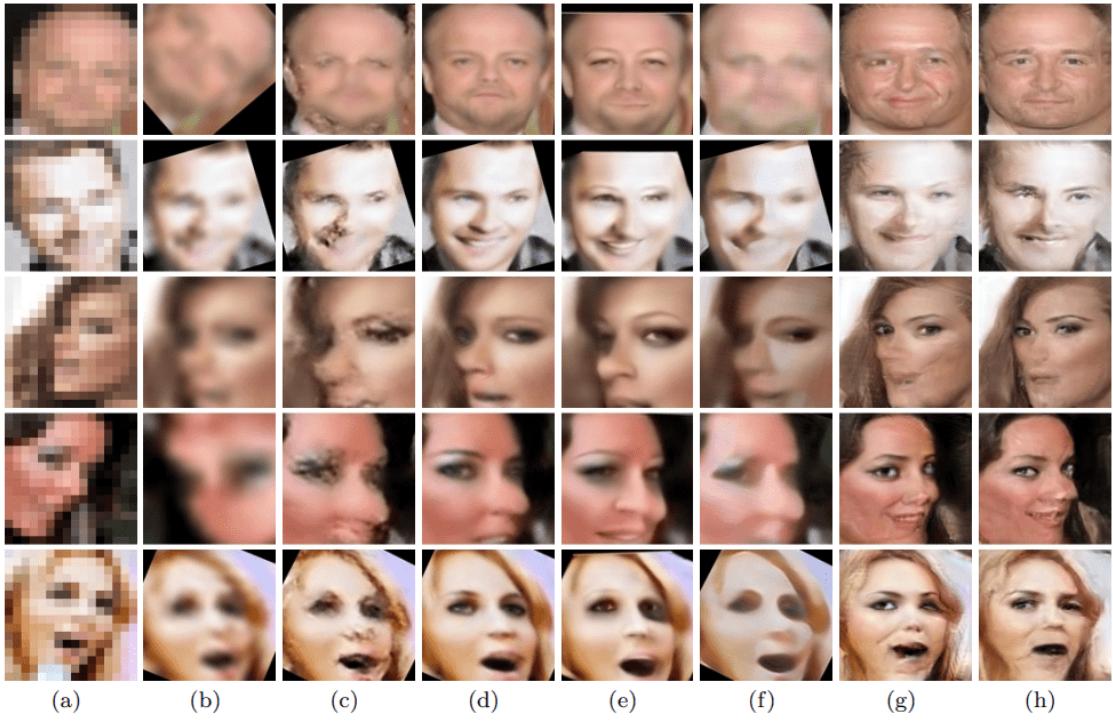


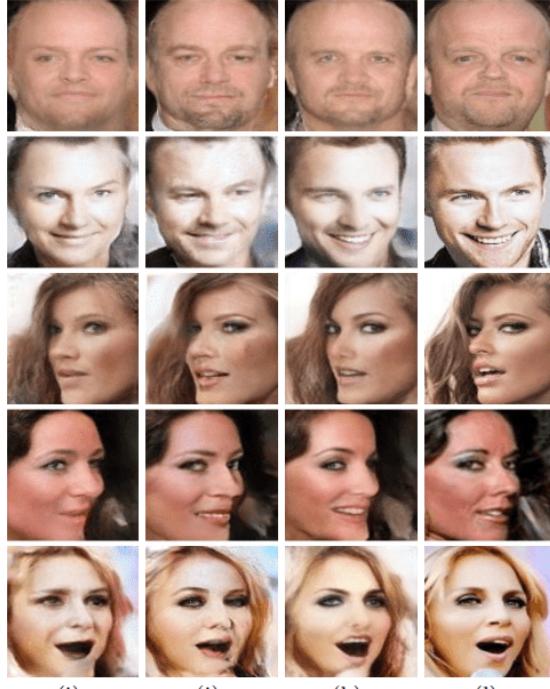
Figure 6: Qualitative comparisons of some state-of-the-art **joint face frontalization methods** on the Multi-PIE database. Columns: (a) Unaligned LR inputs under various poses (Rows:  $+60^\circ$ ,  $+45^\circ$ ,  $-75^\circ$  and  $-90^\circ$ ). (b) Bicubic + [42] (c) [42] + [10]. (d) [42] + [43]. (e) [44] + [42]. (f) [45] + [42]. (g) [46]. (h) Fine-grained facial components. (i) VividGAN [47]. (j) Ground-truths.

- face super-resolution via attention to facial landmark," *arXiv preprint arXiv:1908.08239*, 2019. 2
- [20] L. Li, J. Tang, Z. Ye, B. Sheng, L. Mao, and L. Ma, "Unsupervised face super-resolution via gradient enhancement and semantic guidance," *The Visual Computer*, vol. 37, pp. 2855–2867, 2021. 2
- [21] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proceedings of Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008. 3, 4, 9
- [22] 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. 3
- [23] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2011, pp. 2144–2151. 3
- [24] M. Grgic, K. Delac, and S. Grgic, "Scface—surveillance cameras face database," *Multimedia Tools and Applications*, vol. 51, pp. 863–879, 2011. 3
- [25] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proceedings of the European Conference on Computer Vision*, 2012, pp. 679–692. 3
- [26] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403. 3
- [27] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014. 3, 4
- [28] 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. 3, 9
- [29] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5525–5533. 3
- [30] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 10–15. 3
- [31] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference*. British Machine Vision Association, 2015. 3
- [32] 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. 3
- [33] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2018, pp. 67–74. 3
- [34] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney *et al.*, "Iarpa janus benchmark-c: Face dataset and protocol," in *Proceedings of the International Conference on Biometrics*. IEEE, 2018, pp. 158–165. 3
- [35] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410. 3
- [36] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5549–5558. 3
- [37] K. Zhang, D. Li, W. Luo, J. Liu, J. Deng, W. Liu, and S. Zafeiriou, "Edface-celeb-1 m: Benchmarking face hallucination with a million-scale dataset," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 3
- [38] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edgeconnect: Generative image inpainting with adversarial edge learning," *arXiv preprint arXiv:1901.00212*, 2019. 5
- [39] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, "Recurrent feature reasoning for image inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7760–7768. 5

- [40] Y. Yang and X. Guo, "Generative landmark guided face inpainting," in *Proceedings of the Pattern Recognition and Computer Vision*. Springer, 2020, pp. 14–26. 5
- [41] J. Wang, S. Chen, Z. Wu, and Y.-G. Jiang, "Ft-tdr: Frequency-guided transformer and top-down refinement network for blind face inpainting," *IEEE Transactions on Multimedia*, 2022. 5
- [42] L. Tran, X. Yin, and X. Liu, "Representation learning by rotating your faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 3007–3021, 2018. 6
- [43] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet domain generative adversarial network for multi-scale face hallucination," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 763–784, 2019. 6
- [44] 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. 4, 6, 8, 9
- [45] X. Yu and F. Porikli, "Face hallucination with tiny unaligned images by transformative discriminative neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017. 6, 8
- [46] X. Yu, F. Shiri, B. Ghanem, and F. Porikli, "Can we see more? joint frontalization and hallucination of unaligned tiny faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2148–2164, 2019. 6
- [47] Y. Zhang, I. W. Tsang, J. Li, P. Liu, X. Lu, and X. Yu, "Face hallucination with finishing touches," *IEEE Transactions on Image Processing*, vol. 30, pp. 1728–1743, 2021. 6
- [48] J. Kim, J. K. Lee, and K. M. 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. 8
- [49] X. Yu and F. Porikli, "Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3760–3768. 8
- [50] 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. 8
- [51] 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*, 2018, pp. 217–233. 8
- [52] X. Yu, F. Porikli, B. Fernando, and R. Hartley, "Hallucinating unaligned face images by multiscale transformative discriminative networks," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 500–526, 2020. 8
- [53] B. Amos, B. Ludwigczuk, M. Satyanarayanan *et al.*, "Openface: A general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, vol. 6, no. 2, p. 20, 2016. 4
- [54] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *Proceedings of the European Conference on Computer Vision*. Springer, 2016, pp. 318–333. 4, 9, 12
- [55] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015. 4, 9
- [56] R. Dahl, M. Norouzi, and J. Shlens, "Pixel recursive super resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5439–5448. 4, 9
- [57] 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. 4, 9
- [58] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung, "Sigan: Siamese generative adversarial network for identity-preserving face hallucination," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6225–6236, 2019. 9
- [59] R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE Transactions on Image Processing*, vol. 29, pp. 6251–6263, 2020. 9
- [60] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 8878–8887. 9
- [61] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 399–415. 9
- [62] 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 ACM International Conference on Multimedia*, 2020, pp. 1551–1560. 9
- [63] R. Yasarla, H. R. V. Jozé, and V. M. Patel, "Network architecture search for face enhancement," *arXiv preprint arXiv:2105.06528*, 2021. 9
- [64] Z. Zhang, Y. Ge, Y. Tai, X. Huang, C. Wang, H. Tang, D. Huang, and Z. Xie, "Learning to restore 3d face from in-the-wild degraded images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4237–4247. 10
- [65] S. Wu, C. Rupprecht, and A. Vedaldi, "Unsupervised learning of probably symmetric deformable 3d objects from images in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1–10. 10
- [66] Z. Zhang, Y. Ge, R. Chen, Y. Tai, Y. Yan, J. Yang, C. Wang, J. Li, and F. Huang, "Learning to aggregate and personalize 3d face from in-the-wild photo collection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 214–14 224. 10
- [67] X. Zeng, X. Peng, and Y. Qiao, "Df2net: A dense-fine-finer network for detailed 3d face reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2315–2324. 10
- [68] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, "Learning an animatable detailed 3d face model from in-the-wild images," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–13, 2021. 10
- [69] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2437–2445. 11
- [70] A. Jalal, S. Karmalkar, J. Hoffmann, A. Dimakis, and E. Price, "Fairness for image generation with uncertain sensitive attributes," in *Proceedings of the International Conference on Machine Learning*. PMLR, 2021, pp. 4721–4732. 11
- [71] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9168–9178. 12
- [72] H. Hou, J. Xu, Y. Hou, X. Hu, B. Wei, and D. Shen, "Semi-cycled generative adversarial networks for real-world face super-resolution," *IEEE Transactions on Image Processing*, vol. 32, pp. 1184–1199, 2023. 13



(a) (b) (c) (d) (e) (f) (g) (h)



(i) (j) (k) (l)

Figure 7: Qualitative comparisons of some state-of-the-art **joint face alignment methods** on the input images of size  $16 \times 16$  pixels. The results are obtained in the scenario of first upsampling LR faces and then aligning the super-resolved faces by Bulat et al.'s method (Bulat and Tzimiropoulos, 2017). (a) Unaligned LR inputs. (b) Bicubic interpolation. (c) VDSR [48]. (d) SRGAN [44]. (e) CBN [10]. (f) FSRNet [7]. (g) TDN [45]. (h) TDAE [49]. (i) Yu et al. [50]. (j) Yu et al. [51]. (k) MTDN [52]. (l) Original HR.



Figure 8: Subjective visual quality comparison of some state-of-the-art **joint face recognition methods** for five faces with unknown identities selected from LFW [21] and CelebA [28]: (a) The LR face images ( $16 \times 16$ ). (b)–(i) are the reconstructed  $64 \times 64$  HR faces using (b) bicubic interpolation, (c) DFCG [2], (d) DCGAN [55], (e) UR-DGN [54], (f) PRSR [56], (g) SR-GAN [44], (h) Wavelet-SRNet [57], (i) SiGAN [58], and (j) the ground-truths ( $64 \times 64$ .)

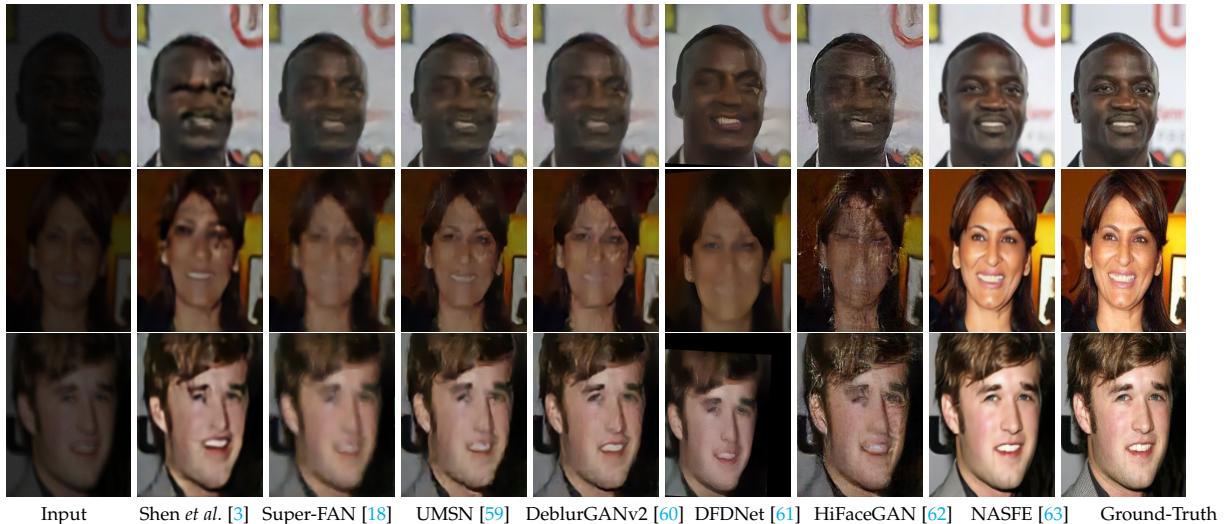


Figure 9: Qualitative comparisons of some state-of-the-art **joint face illumination compensation methods** on synthetic test set *Test-BNL*.



Figure 10: Qualitative comparisons of some state-of-the-art **joint face 3D reconstruction methods**. From left to right, we demonstrate the ground-truth image, the LQ input, and the results from L2R [64], Unsup3D [65], LAP [66], DF2Net [67] and DECA [68], respectively.

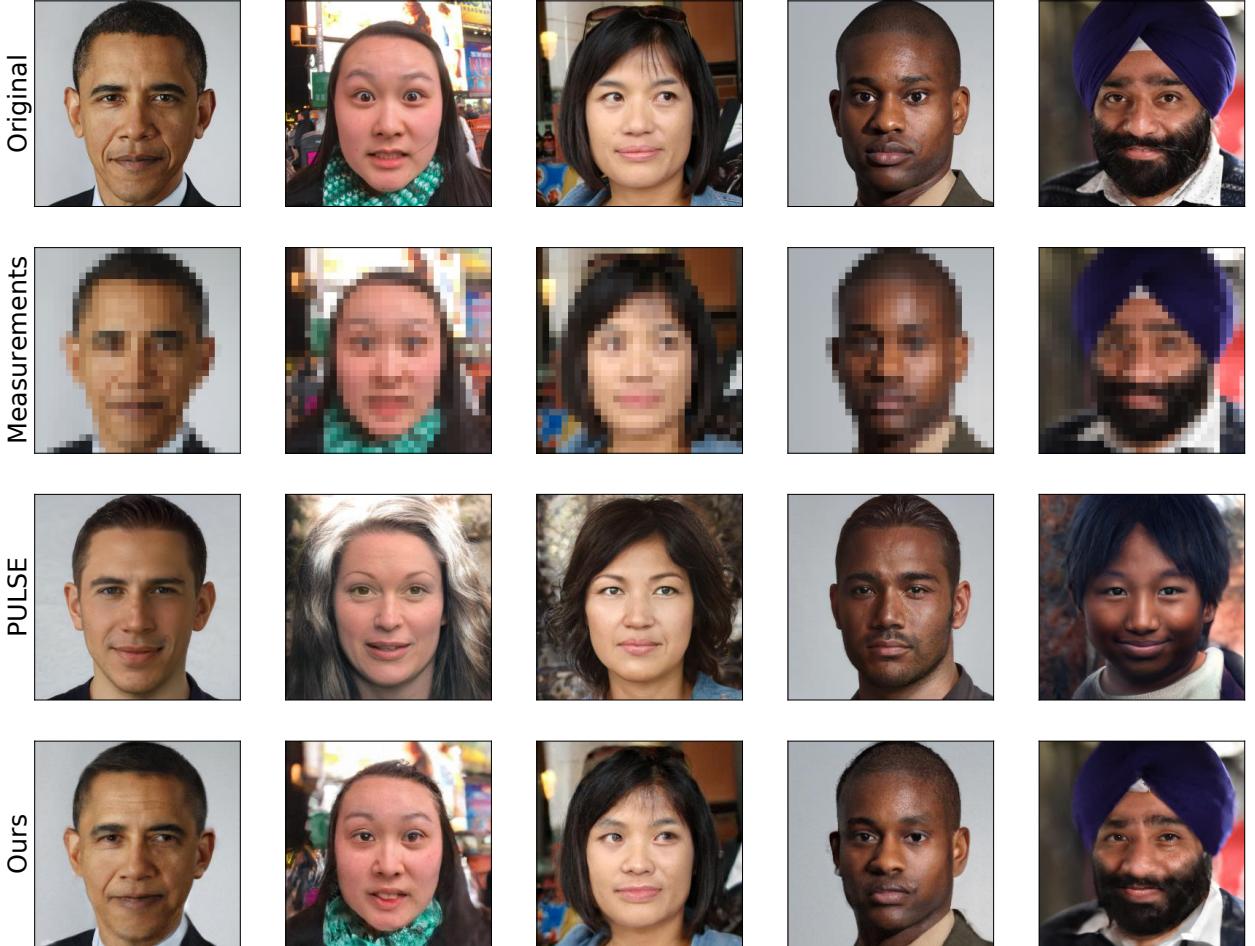


Figure 11: **Joint face fairness** comparisons on Barack Obama and four faces from the FFHQ dataset. The top row shows original images, the second row shows what the algorithms observe: blurry measurements after downsampling by 32x in each dimension. The third row shows reconstructions by PULSE [69], and the last row shows reconstructions by Posterior Sampling via Langevin dynamics, the algorithm [70] are advocating for. These faces were chosen to compare performance on various ethnicities.

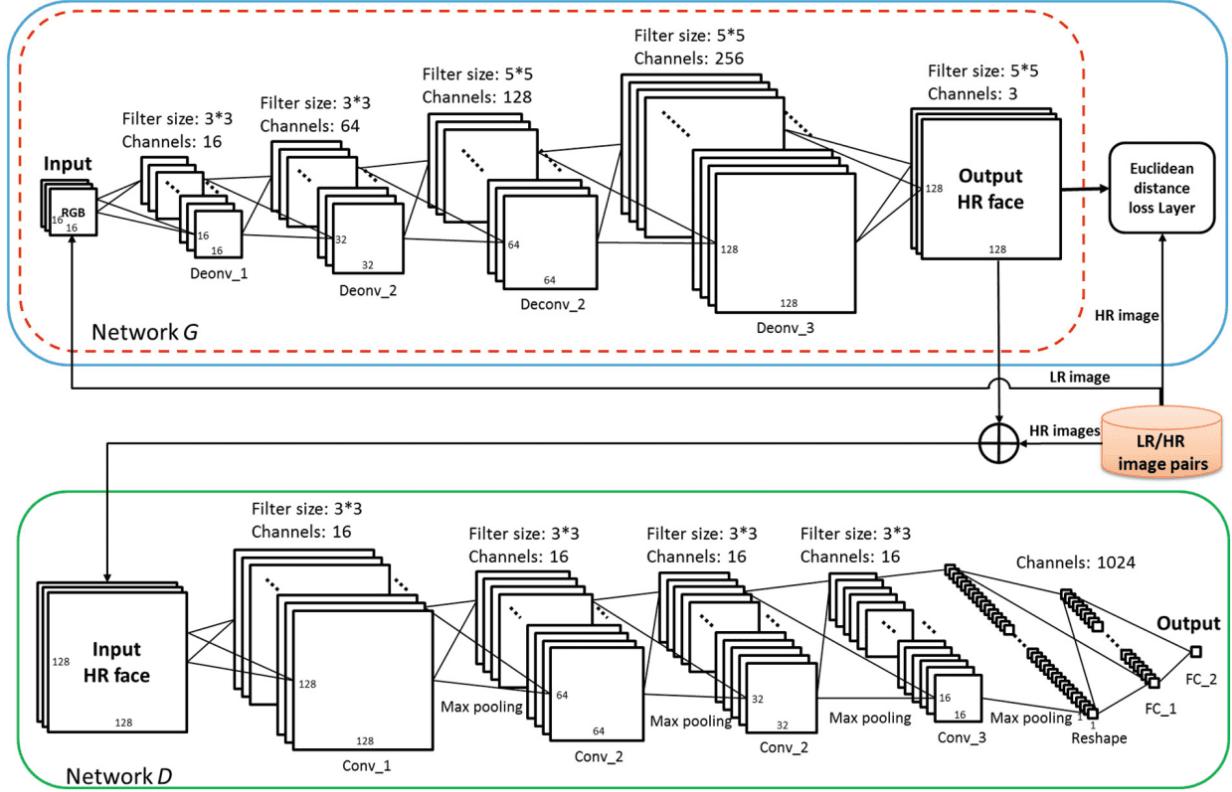


Figure 12: The method URDGN [54] using the **general GAN structure**. In the testing phase, only the generative network in the red dashed block is employed. (Color figure online)

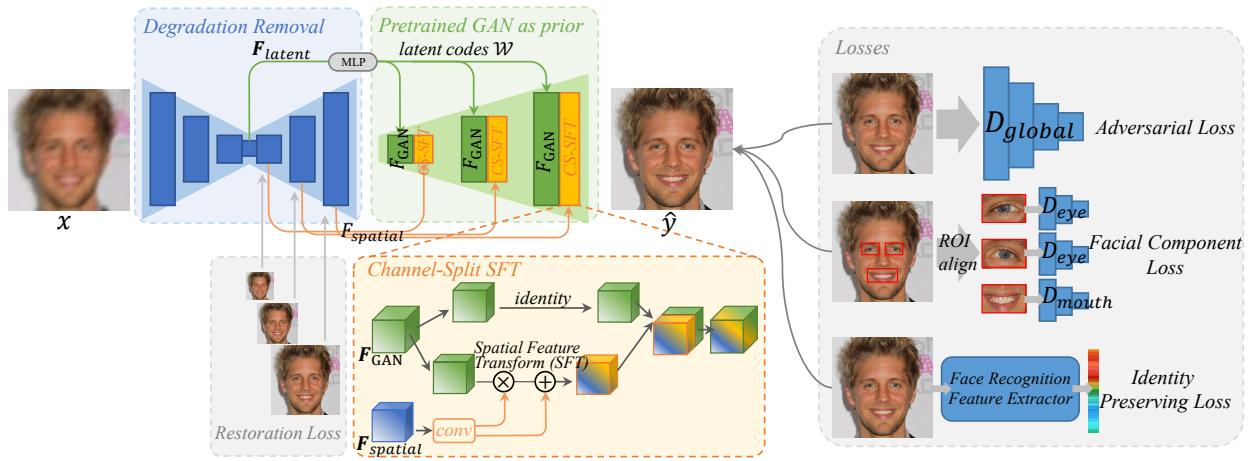


Figure 13: The method GFPGAN [71] using the **Pre-trained GAN structure**. It consists of a degradation removal module (U-Net) and a pretrained face GAN as facial prior.

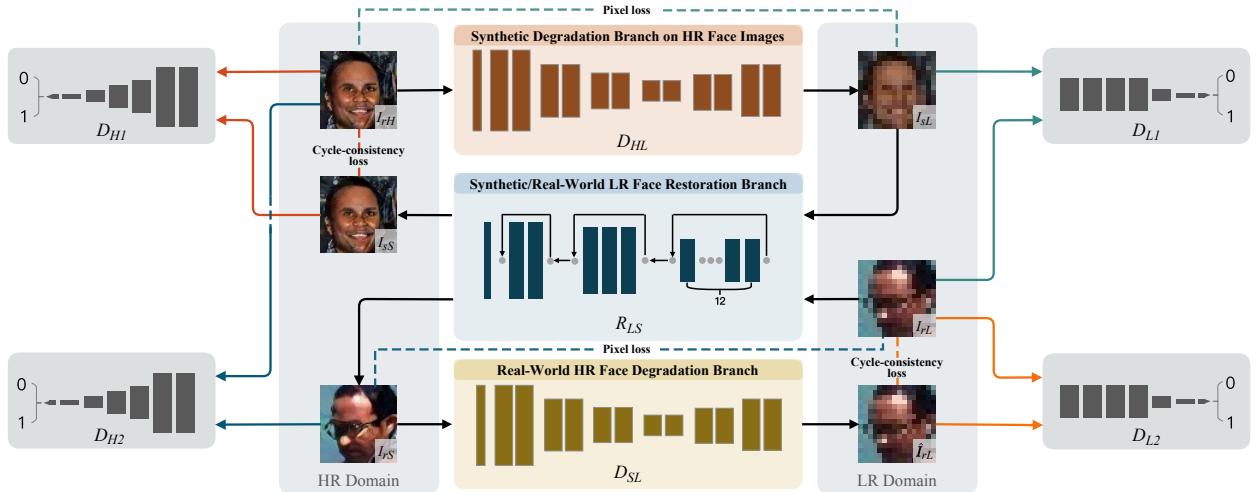


Figure 14: The method SCGAN [72] using the **unpaired images GAN structure**.

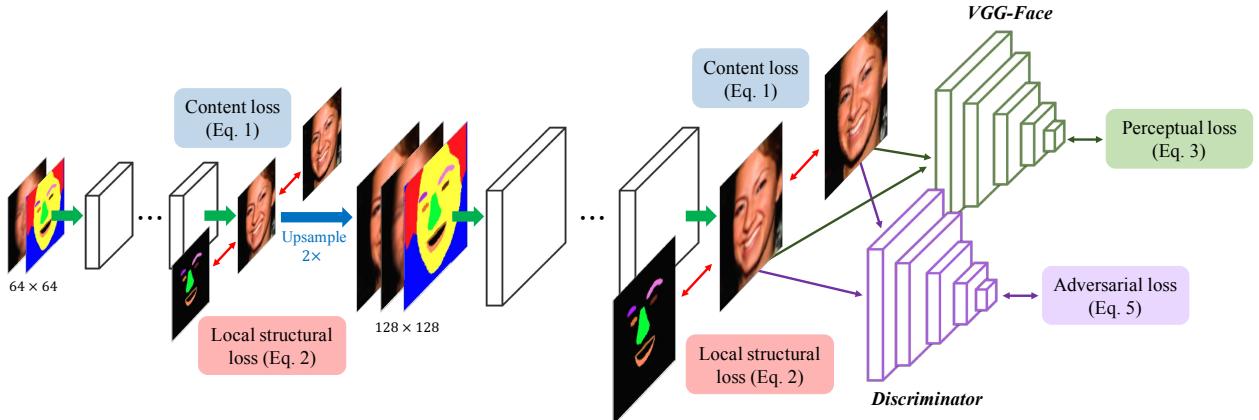


Figure 15: The method DIC [9] using the **Pre-prior framework**.

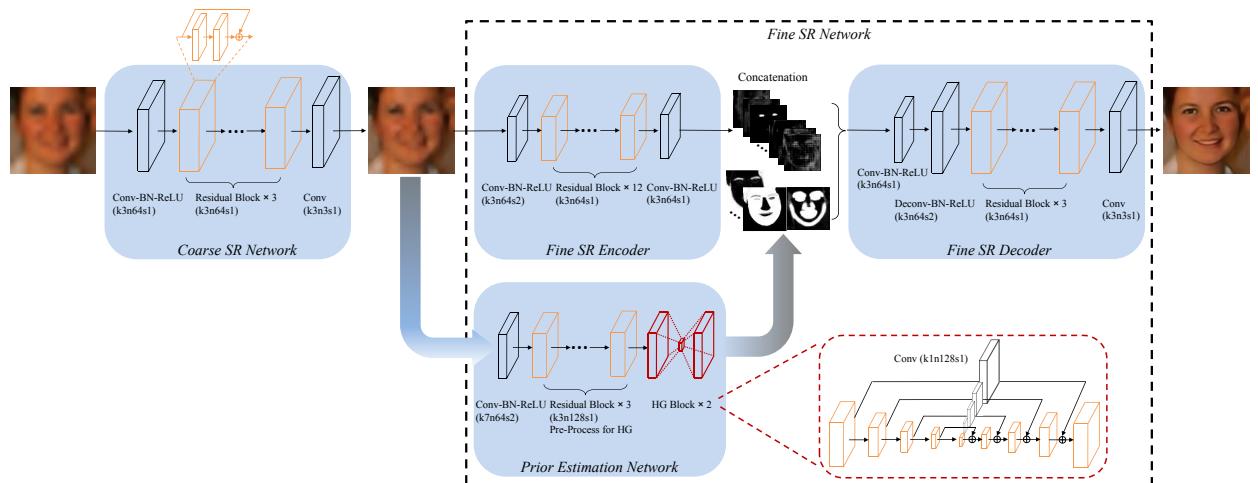


Figure 16: The method FSRNet [7] using the **In-prior framework**.

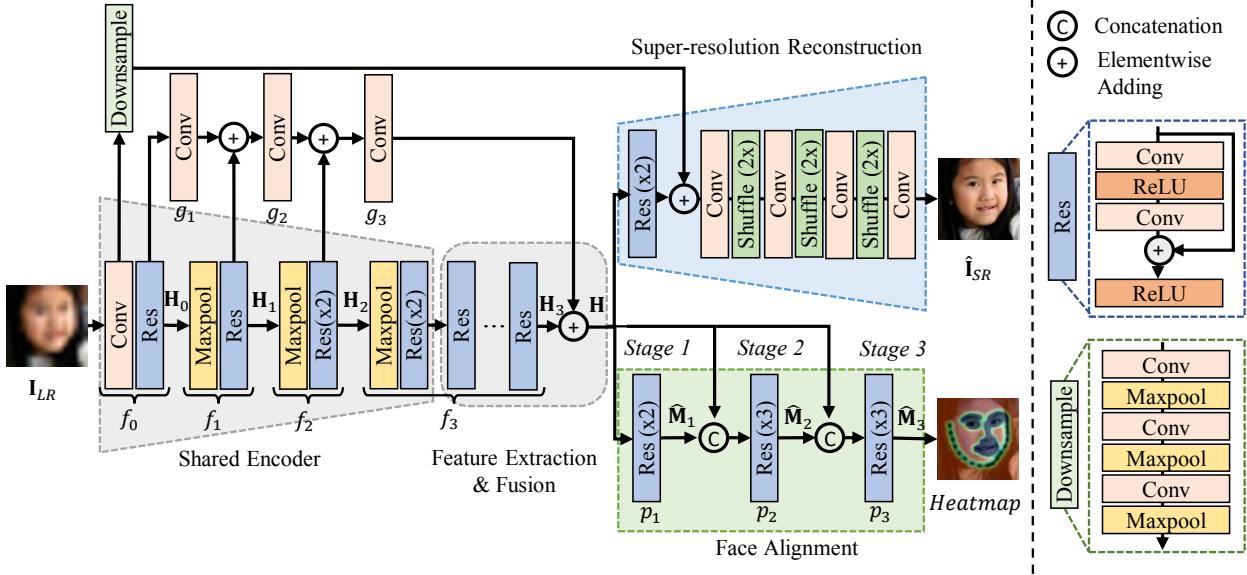


Figure 17: The method JASRNet [11] using the **Parallel-prior framework**.

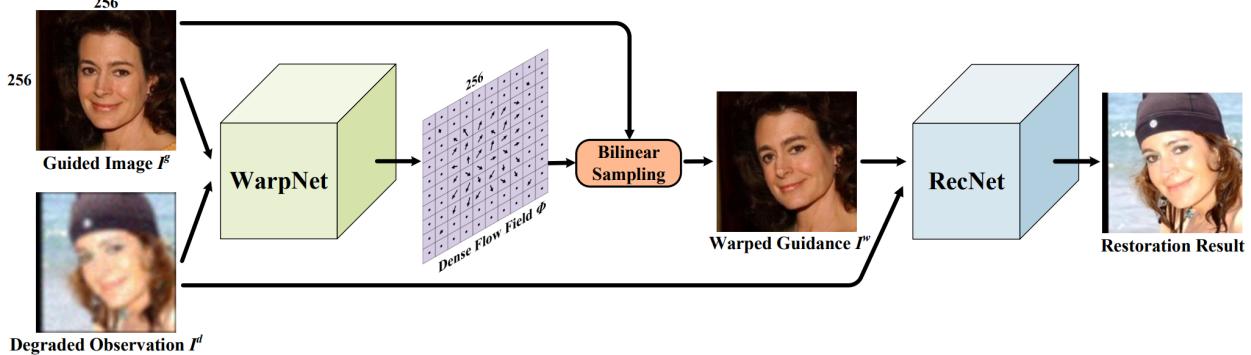


Figure 18: The method GFRNet [14] using the **Input-prior framework**.

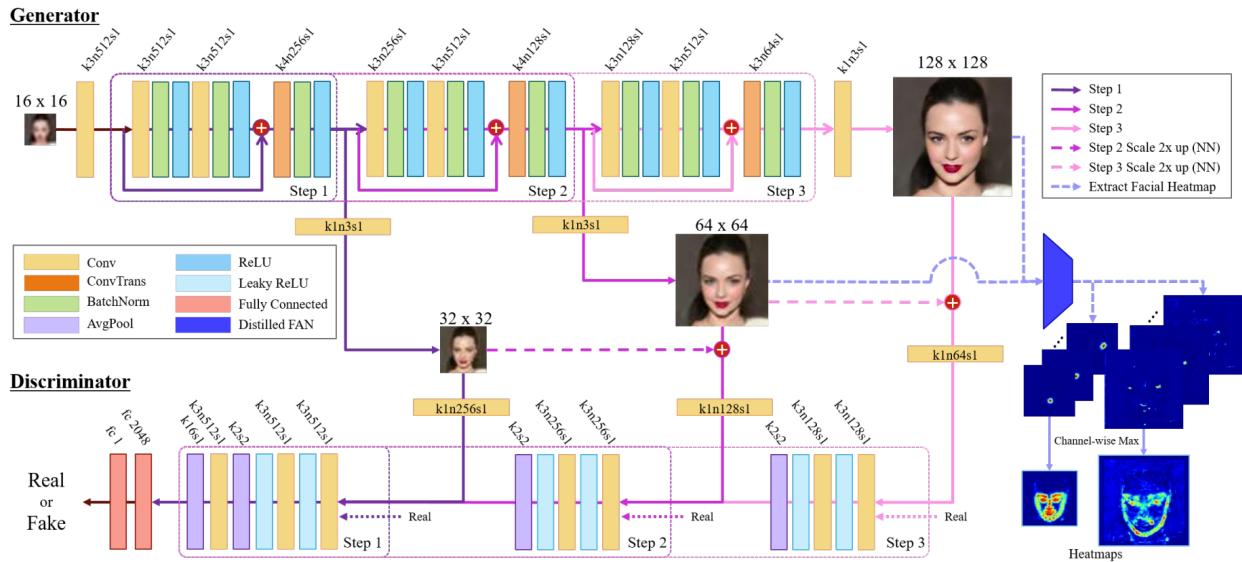


Figure 19: The method FAN [18] using the **Post-prior framework**.