

基于结构先验的人脸图像超分辨率技术概述

王晨阳,江俊君,刘贤明

(哈尔滨工业大学, 计算学部, 哈尔滨 150000)

摘要: 人脸图像超分辨率技术, 又名人脸幻觉, 可根据给定的低分辨率人脸图像中恢复出对应的高分辨率人脸图像。该技术无论是在学术界还是在工业界都具有非常广泛的应用前景。人脸, 作为一种具有高度结构先验的对象, 其结构先验可以为网络提供结构信息, 从而辅助人脸图像超分辨率任务的执行, 改善人脸图像超分辨率性能。因而许多基于结构先验的人脸图像超分辨率方法被提出。为了了解和掌握基于结构先验的人脸图像超分辨率技术的发展状况, 本文对其进行了系统的总结与归类, 主要从先先验、并行先验、中间先验和后先验, 四个方面对基于结构先验的人脸图像超分辨率技术进行概述。最后分析基于结构先验的人脸图像超分辨率技术存在的问题与挑战。

关键词: 人脸图像超分辨率; 人脸幻觉; 结构先验

中图分类号: O422 **文献标识码:** A

An Overview of Face Image Super-Resolution Based on Structural Priors

Wang Chenyang, Jiang Junjun, Liu Xianming

(Faculty of Computing, Harbin Institute Technology, Harbin 150000, China)

Abstract: Face super-resolution technology, also known as face hallucination, can recover the corresponding high-resolution face images from the given low-resolution face images. It has a very broad application prospect both in academia and industry. Human face is a highly structured object, and its structural prior can provide the network structural information and then facilitates the execution of face super-resolution task and improve face super-resolution performance. In order to understand and master the development status of structural prior-based face super-resolution technology, this paper systematically summarizes and categorizes it. From pre-prior, parallel-prior, in-prior, post-prior, four aspects of the structural prior-based face image super-resolution technology are summarized. Finally, the paper analyzes the problems and challenges deep learning face super-resolution technology facing.

Key words: face super-resolution; face hallucination; structural prior

1 引言

人脸图像超分辨率技术，又名人脸幻觉，是一项致力于根据给定的低分辨率人脸图像恢复出具有高频细节的高质量人脸图像的技术。人脸图像超分辨率技术不仅可以提升人脸图像的分辨率，还可以增加人脸图像的辨识度。在工业界和学术界人脸图像超分辨率技术都发挥着举足轻重的作用。在智能安保和监控系统中，由于人物与监控的远距离，或者成像环境（环境光照条件、噪声、压缩等因素）的影响，捕获到的人脸图像往往是低质量且缺少高频细节的，无法提供较高的辨识度，从而对后续的人脸检测与人脸识别任务带来了巨大挑战。人脸图像超分辨率技术则可以有效改善这一问题^[1]。近年来，科学技术飞速发展，手机相机功能日益强大。由于旧手机在像素和硬件条件上远远差于新手机，旧手机拍摄的人脸图像在新手机上显示时，往往模糊不强，无法提供良好的视觉体验。人脸超分辨率则可以对这些低质量人脸图像进行修复得到视觉上令人愉悦的人脸图像^[2,3]。在学术界，很多人脸相关的任务，如人脸属性编辑、人脸识别、人脸表情分析等任务均高度依赖高质量的人脸图像数据集^[4]。因而人脸图像超分辨率技术具有工业界和学术界都有广泛应用前景。因而人脸图像超分辨率技术也始终是图像处理和计算机视觉领域的研究热点。

最早，人脸超分辨率是由Baker和Kanade二人在2000年首次提出^[5]。随后，越来越多学者投身于人脸图像超分辨率技术的研究中，很多里程碑式的人脸超分辨率方法脱颖而出，包括基于全局脸的人脸图像超分辨率方法^[6,7]、基于局部脸的人脸图像超分辨率方法^[8,9]。近几年，具有强大表达能力的深度学习技术在计算机视觉领域大显身手^[10-12]，基于深度学习人脸图像超分辨率方法更是层出不穷并取得了重大突破。而人脸是一种具有高度结构先验的对象，其结构先验的利用可以帮助超分辨率恢复出具有更真实结构信息的人脸图像。因而学者们往往将目光放在如何估计和利用人脸结构先验来促进人脸图像修复，基于结构先验的人脸图像超分辨率方法应运而生。而本文则主要对基于结构先验的人脸图像超分辨率方法进行综述。

本文的主要内容安排如下：第2节首先对深度学习人脸图像超分辨率背景进行粗略地介绍，然后第3节根据现有基于结构先验的人脸超分辨率方法的特征将其归类为，先先验、并行先验、中间先验、后先验四大类，如图1所示，并分别介绍其方法框架。接着第4节对基于结构先验的人脸图像超分辨率技术面临的关键挑战和核心问题进行分析、总结与归纳。最后第5节对全文进行了总结与展望。



图1 基于结构先验的人脸图像超分辨率方法分类

2 人脸图像超分辨率背景

2.1 问题的定义

我们用 I_{LR} 表示低分辨率人脸图像，用 I_{HR} 表示高分辨率人脸图像，图像降质过程可以定义为：

$$I_{LR} = \phi(I_{HR}, \theta)$$

其中 ϕ 是退化模型，而 θ 代表退化模型的参数，包括模糊核、下采样、噪声等。而人脸图像超分辨率技术旨在从给定的低分辨率人脸图像恢复出其对应的高分辨率人脸图像，得到超分辨率人脸图像 I_{SR} ，即图像降质过程的逆过程，

$$I_{SR} = F(I_{LR}, \delta)$$

其中 F 是超分辨率网络模型， δ 则是超分辨率模型的网络参数。为了获取足够大的高低分辨率人脸图像对来满足深度学习技术的需求，学者们通常用数学模型来模拟真实环境的图像降质过程，从而生成大量的高低分辨率图像对训练深度学习模型。

2.2 人脸图像的先验知识

不同于自然图像，人脸图像的主体就是人脸，而人脸本身是具有一些独特的结构先验知识的。结构先验知识可以给网络提供结构信息，从而辅助人脸图像超分辨率。大体上看，人脸图像的结构先验知

识可以分为三种：人脸特征点，人脸热力图和人脸解析图。如图 2 所示，从左到右依次展示了 人脸图像，该人脸图像的人脸特征点、热力图以及解析图先验。除了图示的几种人脸结构先验以外，人脸先验还包括 3D 结构先验^[13]，该先验可以提供比二维结构先验更丰富的信息。

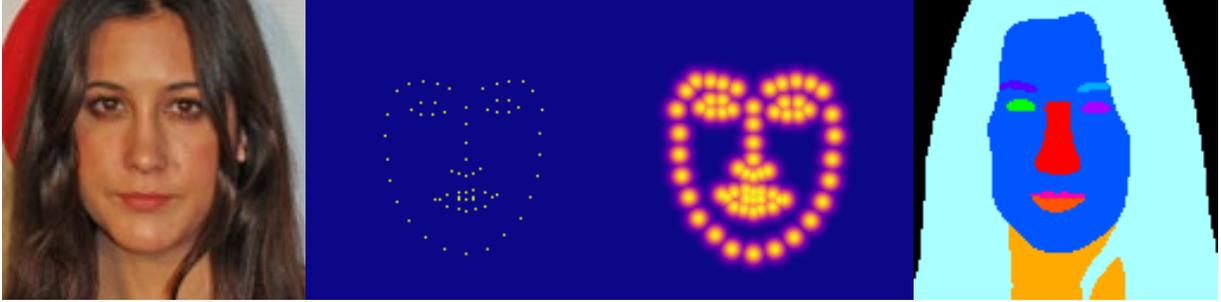


图 2 从左到右：人脸图像、人脸特征点、人脸热力图、人脸解析图

2.3 基于结构先验的损失函数

在深度学习人脸图像超分辨率中，像素级别的 L_1 、 L_2 、Huber^[14]和 Carbonnier 损失^[15]是最常用的损失函数。然而由于像素级别的损失函数通常导致过于光滑而缺少高频细节的超分辨率人脸图像，陆续有学者提出，对抗损失^[16]、感知损失^[17]、风格损失^[18]等损失函数来增加超分辨率人脸图像的高频细节。针对人脸图像超分辨率问题，为了利用人脸图像独有的结构信息，很多基于结构先验的损失函数被提出，基于结构先验的损失函数用于约束估计的先验应该跟真实的先验保持一致。基于结构先验的损失函数可以定义为，

$$L_p = \|P - P_{HR}\|_F$$

其中 L_p 代表基于结构先验的损失函数， p 表示不同的先验包括热力图、特征点和解析图， P 表示估计的先验，而 P_{HR} 则为真实的先验， F 是 1 或者 2。

2.4 常用包含结构先验的数据集

基于结构先验的人脸图像超分辨率方法，因为需要利用人脸结构先验，所以学者们往往需要选择包含结构先验的数据集，因此本文将常用的数据集以及改数据集提供的结构先验信息列在表 1 中。注意人脸热力图是由人脸特征点生成的，因而表 1 中省略了人脸热力图这一项。

表 1 基于结构先验方法常用数据集

数据集	特征点	解析图
CelebA ^[19]	✓	×
Helen ^[20]	✓	✓
CelebAMaskHQ ^[21]	×	✓
FFHQ ^[22]	✓	×

3 基于结构先验的深度学习人脸图像超分辨率技术

具有超强表达能力的深度学习技术的出现，为 人脸图像超分辨率领域带来了巨大进步，也解决了传统人脸图像超分辨率方法的许多问题。最近几年，深度学习技术在人脸图像超分辨率问题上更是大显神通。不同于自然图像，人脸图像的主体是人脸，而人脸具有其独特的先验知识。人脸图像超分辨率技术不仅探索网络模型的设计，还注重对人脸图像先验知识的利用。本文则对基于结构先验的深度学习人脸图像超分辨率方法进行综述。

在第 2 节中，本文已经介绍了人脸图像具有的结构先验，包括人脸特征点、人脸热力图和人脸解析图。这些信息可以帮助网络恢复出具有清晰人脸结构的超分辨率结果。为了利用人脸结构先验，基于结构先验的人脸图像超分辨率方法应运而生。基于结构先验的方法通常设法提取人脸结构先验，然

后利用结构先验来帮助改善人脸图像超分辨率性能。本文根据先验估计和超分辨率任务的执行顺序，将基于结构先验的人脸超分辨率方法归为四类：1) 先先验：先估计先验知识再执行超分辨率；2) 并行先验：并行地执行超分辨率和先验估计任务；3) 中间先验：在超分辨率任务的中间阶段估计先验知识；4) 后先验，在超分辨率完成后估计先验知识。下面本文将对这四类方法进行一一介绍。

3.1 先先验人脸图像超分辨率方法

先先验方法直接从低分辨率人脸图像中估计先验，然后利用该先验知识辅助人脸图像超分辨率任务的完成。早期，Jiang^[23]和 Song^[24]先用预训练好的模型从低分辨率人脸图像中估计人脸特征点，然后根据人脸特征点将低分辨率人脸图像划分为不同的部分，然后为其恢复出高频细节。然而从低分辨率人脸图像检测人脸特征点是极其困难的，且准确度也很低，导致模型性能不够理想，于是学者们将目光转向解析图。如图3所示，PSFRGAN^[25]首先预训练好了一个人脸解析网络，该网络可以直接生成

低分辨率人脸图像的解析图，然后将解析图和低分辨率人脸图像拼接在一起作为超分辨率模型的输入，从而利用人脸解析图促进人脸图像修复过程。此外，他们还利用解析图将超分辨率人脸图像分割成不同的部分，并计算每一部分的结构损失，促使网络可以根据人脸组件的差异恢复出不同的高频细节。而FSR3D则先从低分辨率人脸图像中估计3D结构先验，然后再利用3D结构先验提供的结构信息和低分辨率人脸图像恢复出高质量的超分辨率人脸图像。不同于之前的方法，在FSR3D^[13]中，3D先验的估计和人脸图像的超分辨率是联合训练的。

3.2 并行先验人脸图像超分辨率方法

为了让先验知识估计和人脸图像超分辨率任务相互利用相互促进，并行先验方法提出同时执行超分辨率和先验估计任务。如图4所示，JASRNet^[26]设计了一个编码器去提取人脸图像特征，然后将提取的特征同时送入超分辨率分支和先验估计分支，得到先验和超分辨率结果，以此使得超分辨率和先验估计两个任务相互促进，共同进步。

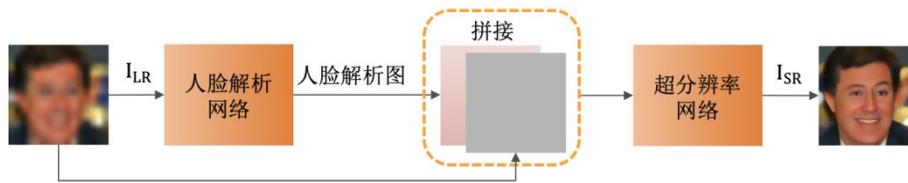


图3 PSFRGAN的框架图

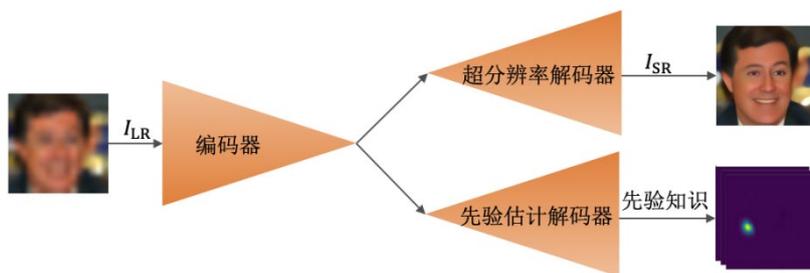


图4 JASRNet的框架图

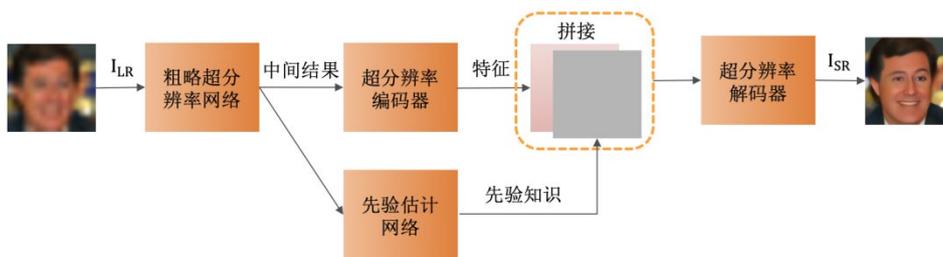


图5 FSRNet的框架图

3.3 中间先验人脸图像超分辨率方法

然而低分辨率人脸图像所包含的信息是非常有限的，从低分辨率人脸图像中估计先验知识非常困难。于是中间先验方法被提出。为了降低先验估计的难度，中间先验方法先对低分辨率人脸图像进行一步上采样，得到中间结果，然后从增强的中间结果估计先验知识，最后再利用先验知识和中间结果恢复出高分辨率人脸图像。CBN^[27]渐进地完成高分辨率人脸图像的恢复和人脸结构先验的估计。于是FSRNet^[28]先设计一个粗略超分辨率网络对低分辨率人脸图像进行一步粗略的超分，生成一个中间结果，接着再从粗略的中间结果中估计人脸先验知识（包括人脸特征点、人脸热力图、人脸解析图），同时利用超分辨率编码器提取中间结果的特征，最后将中间结果和先验知识拼接起来送入编码器中恢复出最终的超分辨率结果，如图5所示。相似地，FSRFCH^[29]也先对低分辨率人脸图像进行上采样，然后从上采样后的特征估计人脸热力图，再将估计的热力图与中间特征相结合重建出最终的超分辨率人脸图像。这两个方法在网络优化过程中均采用基于结构先验的损失来指导先验估计网络可以估计准确的人脸结构先验。

人脸图像质量的改善可以提高人脸先验知识估计的准确性，反过来，准确的人脸先验知识估计也可以提升人脸图像的质量。考虑到这一点，DIC^[30]提出交替执行人脸超分辨率任务和热力图估计任务。如图5所示DIC先将低分辨率人脸图像送入一个超分辨率网络，得到其超分辨率结果 I_{SR1} ，然后将该结果送入这热力图估计网络估计热力图 P_1 。这个过程即为超分辨率促进热力图估计。接着将估计的热力图 P_1 与低分辨率人脸图像一起送入到超分辨率网络中恢复下一次超分辨率结果 I_{SR2} 。这个过程即为热力图辅助超分辨率网络恢复更高质量人脸图像的过程。以此类推，经过多次迭代后，超分辨率任务和先验估计任务相互促进，共同进步。此外，因为直接拼接中间结果和结构先验（如FSRNet何FSRFCH）并不能有效地利用人脸结构先验，DIC在超分辨率网络中专门为热力图的利用设计了一个注意力融合模块。具体来看，该模块首先将热力图按照人脸图像进行分组，然后计算每个组的热力图的加和，在利用softmax对其进行激活后生成一组人脸组件注意力模。为了生成表示不同人脸组件的特征同时不带来过多的参数计算量，DIC选择用组卷积来自适应地生成人脸组件特征。在得到人脸组件特征和人脸组件注意力模后，将对应的组件特征和组件注意力模点乘在求和，即得到人脸组件注意力后的特征。此时人脸特征和人脸热力图有效地结合，超分辨率性能显著提升。

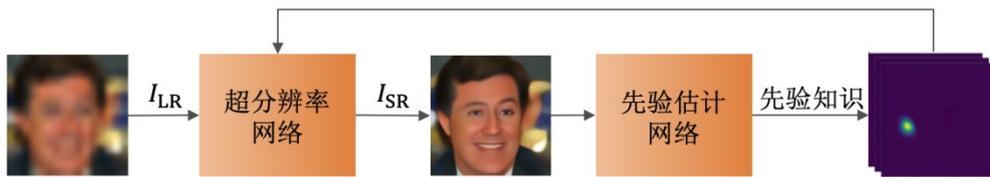


图6 DIC的框架图

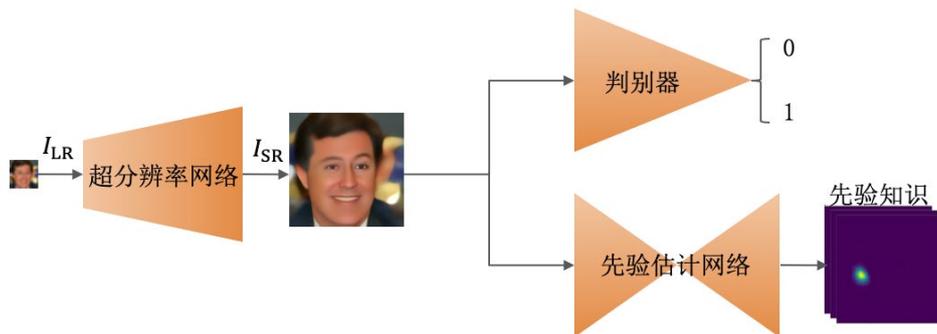


图7 Super-FAN的框架图

3.4 后先验人脸图像超分辨率方法

后先验方法则从最终的超分辨率结果中估计人脸结构先验。最具有代表性的后先验方法是 Super-

FAN^[31]，其网络框架如图6所示。它首先恢复一个超分辨率结果，然后设计先验估计网络了来估计超分辨率结果的人脸热力图信息，并利用基于结构先验的损失函数约束超分辨率人脸图像和高分辨率人

脸图像热力图信息要保持一致，从而促进超分辨率人脸图像结构的恢复。同时 Super-FAN 利用判别器计算对抗损失，来恢复出更真实的具有丰富高频细节人脸图像。本质上，Super-FAN 是利用基于结构先验的损失函数来辅助超分辨率模型的优化，引导超分辨率模型生成结构意识的人脸图像。在推理阶段，超分辨率人脸图像的恢复不再需要结构先验的估计。

4 问题与挑战

在上一节，本文已经介绍了基于结构先验的人脸图像超分辨率方法。本文则在这一节对基于结构先验的人脸图像超分辨率方法进行分析。首先先验直接从低分辨率人脸图像中估计先验信息，有低分辨率人脸图像包含的信息量极少，先验估计的准确性很难保证，从而限制了模型的性能。而并行先验的方式虽然可以使先验估计和超分辨率相互作用相互影响，但是本质上先验估计依然是从低分辨率人脸图像获得的，因此性能改善非常有限。中间先验的方式先恢复一个中间结果，再估计先验知识，再完成超分辨率任务，这一过程尽管可以显著提升模型性能，但也因为多次超分辨率加大了模型对空间和计算资源的需求。除此以外，上述方法均对结构先验具有较强的依赖性和较高的敏感度，一旦估计的结构先验存在较大的错误，模型的性能也会直接受到负面影响。相比于其他方法，后先验概率只需要在训练时估计先验，而在推理阶段只需完成超分辨率任务即可。这同样是一把双刃剑。推理阶段不涉及先验估计可以减少模型对空间和计算资源的需求，但也导致模型无法充分挖掘人脸结构先验的信息。因此设计一个既可以充分利用人脸结构先验知识，又不需要额外的计算和空间资源的基于结构先验的人脸图像超分辨率方法仍是当前人脸图像超分辨率领域的难题。此外人脸先验知识和低分辨率人脸图像的有效融合也是基于结构先验的人脸图像超分辨率方法的核心问题。

5 结论

本文对现有基于结构先验的人脸图像超分辨率技术进行了简单地回顾与总结。首先对人脸图像超分辨率技术背景进行简要介绍，主要包括人脸图像超分辨率问题的定义、人脸结构先验知识以及基于

结构先验的损失函数。接着本文对基于结构先验的人脸图像超分辨率技术进行了划分和归类，分别介绍了先验、并行先验、中间先验和后先验四个类别中具有代表性的方法。最后本文分析了现有方法中急需解决的问题。

参考文献

- [1] Aakerberg A, Nasrollahi K, Moeslund T B. Real-World Super-Resolution of Face-Images from Surveillance Cameras[J]. arXiv preprint arXiv:2102.03113, 2021.
- [2] Wan Z, Zhang B, Chen D, et al. Bringing old photos back to life[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2747-2757.
- [3] Wang X, Li Y, Zhang H, et al. Towards Real-World Blind Face Restoration with Generative Facial Prior[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 9168-9178.
- [4] Lai S C, He C H, Lam K M. Low-resolution face recognition based on identity-preserved face hallucination[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 1173-1177.
- [5] Baker S, Kanade T. Hallucinating faces[C]// Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2000:83-88.
- [6] Gunturk B K, Batur A U, Altunbasak Y, et al. Eigenface-domain super-resolution for face recognition[J]. IEEE transactions on image processing, 2003, 12(5):597-606.
- [7] Yang C Y, Liu S, Yang M H. Structured face hallucination[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1099-1106.
- [8] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, 1: I-I.
- [9] Chen L, Pan J, Jiang J, et al. Multi-stage Degradation Homogenization via Image Re-expression for Noisy Face Super-Resolution[J]. IEEE Transactions on Image Processing, 2021

- [10] Wang Z, Chen J, Hoi S C H. Deep learning for image super-resolution: A survey[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [11] Anwar S, Khan S, Barnes N. A deep journey into super-resolution: A survey[J]. *ACM Computing Surveys (CSUR)*, 2020, 53(3): 1-34.
- [12] Liu H, Ruan Z, Zhao P, et al. Video super resolution based on deep learning: A comprehensive survey[J]. *arXiv preprint arXiv:2007.12928*, 2020.
- [13] Hu X, Ren W, LaMaster J, et al. Face super-resolution guided by 3d facial priors[C]//*European Conference on Computer Vision*. Springer, Cham, 2020: 763-780.
- [14] Kalarot R, Li T, Porikli F. Component attention guided face super-resolution network: Cagface[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020: 370-380.
- [15] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 624-632.
- [16] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//*Advances in neural information processing systems (NIPS)*, 2014(2):2672-2680.
- [17] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 4681-4690.
- [18] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 2414-2423.
- [19] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild[C]//*Proceedings of the IEEE international conference on computer vision*. 2015: 3730-3738.
- [20] Le V, Brandt J, Lin Z, et al. Interactive facial feature localization[C]//*European conference on computer vision*. Springer, Berlin, Heidelberg, 2012: 679-692.
- [21] Lee C H, Liu Z, Wu L, et al. Maskgan: Towards diverse and interactive facial image manipulation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 5549-5558.
- [22] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 4401-4410.
- [23] Jiang J, Yu Y, Hu J, et al. Deep CNN Denoiser and Multi-layer Neighbor Component Embedding for Face Hallucination[C]// *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 2018(8):771-778.
- [24] Song Y, Zhang J, He S, et al. Learning to Hallucinate Face Images via Component Generation and Enhancement[C]// *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017: 4537-4543.
- [25] Chen C, Li X, Yang L, et al. Progressive Semantic-Aware Style Transformation for Blind Face Restoration[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 11896-11905.
- [26] Yin Y, Robinson J P, Zhang P, et al. Joint Super-Resolution and Alignment of Tiny Faces[C]//*Proceeding of the Association for the Advancement of Artificial Intelligence*, 2019.
- [27] Zhu S, Liu S, Loy C C, et al. Deep cascaded bi-network for face hallucination[C]//*European conference on computer vision*. Springer, Cham, 2016: 614-630.
- [28] Chen Y, Tai Y, Liu X, et al. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors[C]// *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018:2492-2501
- [29] Yu X, Fernando B, Ghanem B, et al. Face Super-resolution Guided by Facial Component Heatmaps[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018:217-233.
- [30] Ma C, Jiang Z, Rao Y, et al. Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 5569-5578.
- [31] Bulat A, Tzimiropoulos G. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 109-117.