

doi:10.16112/j.cnki.53-1223/n.2026.01.202411040007

引用格式:蔡梦浩,唐鑫,李云平,等.面向真实场景的图像超分辨率网络双三次特征调制和多分类鉴别器的设计[J].昆明理工大学学报(自然科学版),2026,51(1):126-136.

Citation:CAI Menghao,TANG Xin,LI Yunping,et al.Design of Bicubic Feature Modulation and Multi-Classification Discriminator for Real-World Image Super-Resolution Network[J].Journal of Kunming University of Science and Technology(Natural Science),2026,51(1):126-136.

# 面向真实场景的图像超分辨率网络双三次特征调制和多分类鉴别器的设计

蔡梦浩<sup>1</sup>,唐鑫<sup>2</sup>,李云平<sup>2\*</sup>,戴银波<sup>2</sup>,张亚飞<sup>1</sup>

(1.昆明理工大学信息工程与自动化学院,云南昆明650500;2.红云红河烟草集团有限责任公司昆明卷烟厂,云南昆明650032)

**摘要:**大多数真实场景的图像超分辨率方法通常需要预估图像的退化参数。一旦退化参数估计不准确,网络重建的图像视觉效果往往欠佳。现有的基于生成对抗网络(Generative Adversarial Network, GAN)的方法大多采用全局鉴别器,但由于真实场景的自然图像包含多种复杂且未知的分布,全局鉴别器容易使超分辨率网络仅学到图像的粗粒度信息,进而影响重建图像的局部细节质量。为解决上述问题,提出了双三次特征调制和多分类鉴别器图像超分辨率网络。该网络由双三次特征调制模块和多分类鉴别器组成。双三次特征调制模块能够将真实场景中未知退化的图像特征调制为超分辨率网络更易于处理的双三次特征,以此来提高未知退化图像的重建质量。另外,本研究还设计了多分类鉴别器,通过设置多个鉴别器,使各鉴别器分别专注于图像的一种分布特征进行鉴别,多个鉴别器协同合作,增强对图像不同分布特征的辨识能力,从而提升重建图像的细节表现。实验结果显示,与现有先进的面向真实场景的图像超分辨率方法相比,本文的方法在真实场景测试集上具有显著优势。

**关键词:**图像超分辨率;深度学习;特征调制;鉴别器

**中图分类号:**TP391.41 **文献标识码:**A **文章编号:**1007-855X(2026)01-0126-11

## Design of Bicubic Feature Modulation and Multi-Classification Discriminator for Real-World Image Super-Resolution Network

CAI Menghao<sup>1</sup>,TANG Xin<sup>2</sup>,LI Yunping<sup>2\*</sup>,DAI Yinbo<sup>2</sup>,ZHANG Yafei<sup>1</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;  
2. Kunming Cigarette Factory of Hongyun Honghe Tobacco (Group) Co., Ltd., Kunming 650032, China)

**Abstract:** Most real-world image super-resolution methods usually need to estimate the degradation parameters of the image. Once the degradation parameter estimation is inaccurate, the image visual effect of network reconstruction is often suboptimal. In addition, most of the existing methods based on Generative Adversarial Network (GAN) use global discriminators. Due to the fact that natural images in real-world scenarios contain a variety of complex and unknown distributions, the global discriminator can easily cause the super-resolution network to only learn the coarse-grained information of the image, which in turn affects the quality of the local details of the reconstructed image. In order to solve the above problems, this paper proposes a bicubic feature modulation and multi-classification discriminator image super-resolution network. The network consists of a bicubic feature modulation module and a multi-classification discriminator. The bicubic feature modulation module can modulate the image features in real-world scenarios into bicubic features that are easier for the super-resolution network to process, thereby improving the reconstruction quality of unknown degraded images. In addition, this study also designed a multi-classification discriminator, by setting multiple discriminators, each discriminator focuses on identifying a specific distribution feature of the image, and multiple discriminators cooperate to enhance the ability to identify different distribution features of the image, thereby improving the detail performance of the reconstructed image. Experimental results show that compared with existing advanced real-world image super-resolution methods, the proposed method has a significant advantage on the real-world scene test set.

**收稿日期:**2024-11-04. **基金项目:**国家自然科学基金项目(62161015),云南省基础研究专项(202301AV070004).

**作者简介:**蔡梦浩(1999-),男,硕士研究生.主要研究方向:计算机视觉. **E-mail:**15139504182@163.com

**通信作者:**李云平(1973-),男,工程师.主要研究方向:计算机视觉. **E-mail:**hyhhlyp@163.com

lation and multi-classification discriminator image super-resolution network. The network consists of a bicubic feature modulation module and a multi-classification discriminator. The bicubic feature modulation module can modulate the unknown degraded image features in real-world scenarios into bicubic features that are easier to process for super-resolution networks. This improves the reconstruction quality of unknown degraded images. In addition, this paper designs a multi-classification discriminator. By setting up multiple discriminators, each discriminator focuses on one image distribution feature for identification, and the multiple discriminators work together to enhance the ability to identify different distribution features of the image, so as to improve the detail performance of the reconstructed image. Experimental results show that the method proposed in this paper has significant advantages over the existing advanced image super-resolution methods for real-world scenarios.

**Keywords:** image super-resolution; deep learning; feature modulation; discriminator

## 0 引言

不同于传统的基于插值的图像超分辨率方法<sup>[1]</sup>,近些年来基于深度学习的图像超分辨率方法<sup>[2-7]</sup>在计算机视觉和图像处理领域成为热门话题,引发了广泛的关注与讨论.在众多实际应用场景中,如图像压缩、医学成像、遥感成像、公共安防等领域,理想的高分辨率图像起着至关重要的作用.然而在实际情况下,受采集设备与环境的影响,往往只能获取低分辨率图像.于是,图像超分辨率技术应运而生,旨在从退化的低分辨率图像中重建出具有丰富纹理细节的高分辨率图像.

大多数早期的图像超分辨率方法<sup>[8-15]</sup>通常仅仅考虑最为简单的图像退化方式,即双三次退化.在此前提下,模型采用 L1 损失函数或者 MSE 损失函数时,能够表现出良好的性能,然而生成的图像往往过于平滑.为了提升图像的视觉效果,引入了生成对抗网络<sup>[16]</sup>以缓解图像过度平滑的问题.生成对抗网络包含生成器和鉴别器两部分,生成器的目的是生成足够逼真的图像,从而骗过鉴别器;鉴别器的目的则是区分图像的真实性.通过对抗训练的策略,生成器会逐渐生成包含更多细节的图像,进而提升图像的视觉质量.

尽管图像超分辨率方法在双三次退化上取得了成功,但由于真实场景中的图像退化类型繁多且难以确定,利用合成的双三次退化训练集训练出的模型在对真实场景的低分辨率图像进行测试时,通常效果欠佳.这种不尽人意的视觉效果主要是因为合成训练集和真实测试集之间存在域差异.为了解决这种域差异问题,Cai 等<sup>[17]</sup>和 Wei 等<sup>[18]</sup>提出运用特定焦距的相机镜头来采集真实场景的高分辨率低分辨率图像对;Wang 等<sup>[3]</sup>明确构建了一个包含模糊、下采样、噪音以及 JPEG 压缩的高阶退化模型,生成更接近真实场景的图像对.BSRGAN 模型则通过随机组合不同的模糊、下采样策略、噪音和 JPEG 压缩方法,进一步拓展了退化模型的建模空间<sup>[19]</sup>.这些方法合成的图像被用于生成对抗网络的训练,有效提高了生成图像的视觉质量.此外,DASR 引入了一个小型递归网络来预测输入图像的退化参数,结合多个卷积专家的联合优化,通过专家的非线性组合确定网络参数,使得模型具有处理多种退化的能力<sup>[20]</sup>.MM-RealSR 模型提出无监督退化估计策略,通过度量学习将真实场景中无法量化的退化映射到度量空间,进行无监督训练<sup>[21]</sup>.SeeSR 模型采用扩散模型方法,先训练语义提取器从图像中提取语义信息,以引导扩散模型提升处理图像局部细节的能力<sup>[22]</sup>.尽管扩散模型生成的图像质量较高,但是由于其需要多次采样,所以推理时间较长.

虽然 Wang 等<sup>[3]</sup>提出了一种高阶图像退化模型以模拟真实场景中的复杂退化情况,然而,由于真实场景的退化过程极为复杂,难以进行准确建模,因此通过这种模拟真实场景的合成训练对训练得到的模型在真实场景的测试集上的表现依旧不尽人意.此外,大多数基于生成对抗网络的图像超分辨率方法<sup>[3,5,6,23-25]</sup>所采用的是全局鉴别器.全局鉴别器仅仅将一张富含丰富纹理细节的图像归一化为一个概率值,试图学习整张图像的复杂分布,却往往忽略了图像中的局部分布特征.这种全局鉴别器会使生成图像的局部细节丢失或者模糊,进而影响视觉效果.

为解决上述问题,本文提出了双三次特征调制和多分类鉴别器图像超分辨率网络,以应用于真实场景的图像超分辨率.鉴于真实场景的高分辨率低分辨率图像对难以获取,而通过双三次退化核合成的训练对相对容易,因此本文的目标是充分利用合成的双三次退化的训练对,使模型能够对真实场景的低分辨率图

像进行恢复. 为此, 本文设计了双三次特征调制模块, 该模块将未知退化的图像特征调制为主干网络更易于恢复的双三次特征. 由于图像通常涵盖多种不同的分布, 传统的全局鉴别器难以对整张图像的复杂分布进行精确建模. 为化解这一难题, 本文引入了多分类鉴别器, 设置多个鉴别器, 使每个鉴别器专注于鉴别图像中的一种分布特征, 通过将各鉴别器的输出映射至不同类别, 以实现图像分布的细粒度建模.

## 1 方法

### 1.1 网络结构

由于人工合成的数据集与真实场景中未知复杂退化的数据集存在巨大的域差异, 因此直接将人工合成的数据集训练好的模型用于重建真实场景的低分辨率图像时, 就会由于这种差异而表现不佳. 为了解决这个问题, 本文创新性地设计了双三次特征调制模块, 同时, 为了尽可能地重建出图像细粒度的纹理细节, 进一步提出了多分类鉴别器以对抗训练的方式联合优化双三次特征调制模块. 网络结构如图 1 所示, 所提出的网络主要由三部分组成: SwinIR 主干网络、双三次特征调制模块以及多分类鉴别器. 其中, SwinIR 主干网络的作用是将低分辨率图像重建为高分辨率图像; 双三次特征调制模块负责把真实场景中未知退化的低分辨率图像特征调制为主干网络易于处理的双三次特征, 以此来克服未知退化给图像恢复带来的难题; 多分类鉴别器用于对整张图像的不同分布区域进行鉴别, 从而助力恢复图像的局部细节. 模型训练分为两个阶段: 第一阶段, 先用双三次退化的低分辨率图像训练 SwinIR 主干网络, 随后固定其参数, 以保留其恢复双三次退化图像的能力; 第二阶段, 将双三次退化的低分辨率图像输入到主干网络中, 逐层提取中间层的特征作为标签. 与此同时, 将模拟真实场景的未知退化的图像输入主干网络与双三次特征调制模块中, 提取经过调制后的特征, 并与对应的双三次特征标签进行损失计算. 最后, 通过多分类鉴别器以对抗训练的方式联合优化双三次特征调制模块, 使其能够更好地适应真实场景中的退化特征.

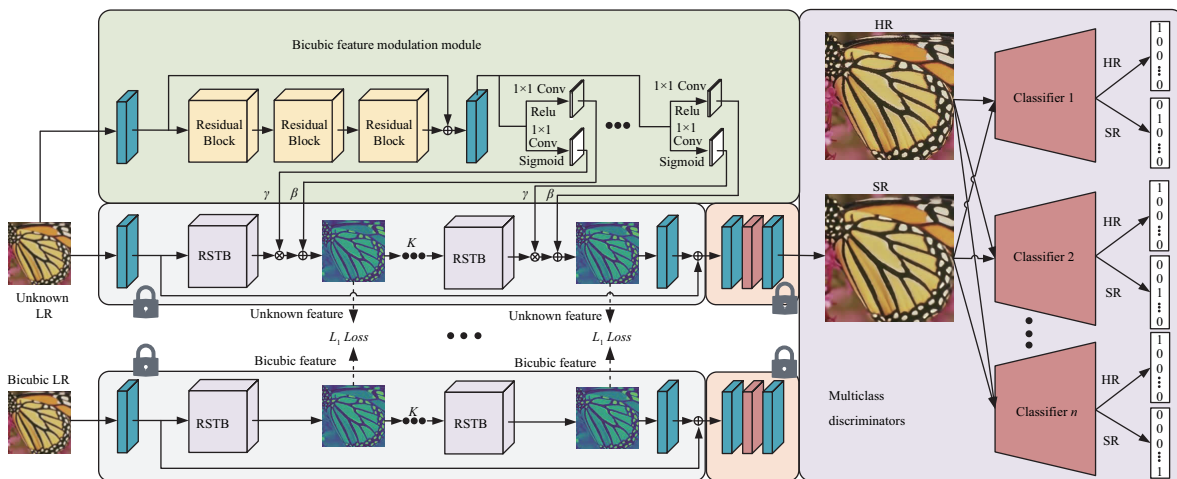


图 1 网络的总体框架

Fig. 1 Overview framework of the network

#### 1.1.1 主干网络

SwinIR 是一种基于 Transformer 结构的主干网络架构, 专为通用图像恢复任务设计, 包含浅层特征提取、深层特征提取和高质量图像重建三个模块. 深层特征部分提取由多个残差连接的 Swin Transformer 块 (Residual Swin Transformer Block, RSTB) 组成, 每个 RSTB 内部含有六个 Swin Transformer 模块. 鉴于 SwinIR 网络在双三次退化的低分辨率图像上展现了优异的超分辨率重建效果, 本文选用 SwinIR 作为本模型的主干网络.

首先, 先利用合成的双三次退化训练集来训练 SwinIR 主干网络. 给定一个双三次退化的低分辨率图像  $I_{LR}^b \in \mathbb{R}^{H \times W \times 3}$  (其中,  $H$  和  $W$  分别表示图像的高度和宽度, 3 代表输入的通道数), 通过  $3 \times 3$  卷积层  $H_{SF}(\cdot)$  提取浅层特征  $F_0 \in \mathbb{R}^{H \times W \times C}$ :

$$F_0 = H_{SF}(I_{LR}^b) \quad (1)$$

其中: $I_{LR}^b$ 表示输入网络的双三次退化的低分辨率图像; $H_{SF}(\cdot)$ 表示卷积层; $F_0$ 表示卷积层提取的浅层特征; $C$ 表示特征通道数.卷积层的作用是将输入的图像空间映射到高维空间.接下来,从 $F_0$ 中提取深层特征 $F_{DF} \in \mathbb{R}^{H \times W \times C}$ :

$$F_{DF} = H_{DF}(F_0) \quad (2)$$

式中: $F_{DF}$ 表示深层特征提取模块提取的深层特征; $H_{DF}(\cdot)$ 是SwinIR主干网络中的深层特征提取模块,包含 $K(K=6)$ 个残差Swin Transformer块(RSTB)和一个 $3 \times 3$ 卷积层.具体而言,从RSTB块中提取中间层特征 $F_1, F_2, \dots, F_K$ 以及最终的深层特征 $F_{DF}$ :

$$F_i = H_{RSTB_i}(F_{i-1}), i = 1, 2, \dots, K \quad (3)$$

$$F_{DF} = H_{CONV}(F_K) \quad (4)$$

式中: $H_{RSTB_i}(\cdot)$ 表示第 $i$ 个RSTB块; $F_{i-1}$ 表示第 $i$ 个RSTB块的输入特征; $F_i$ 表示第 $i$ 个RSTB块的输出特征; $F_K$ 表示第 $K$ 个RSTB块的输出特征, $H_{CONV}$ 表示最后的卷积层.最终,通过图像重建模块聚合浅层特征 $F_0$ 和深层特征 $F_{DF}$ 得到高分辨率图像 $I_{SR}$ :

$$I_{SR} = H_{REC}(F_0 + F_{DF}) \quad (5)$$

式中: $I_{SR}$ 表示重建出来的高分辨率图像; $H_{REC}(\cdot)$ 表示图像重建模块.当完成主干网络的训练后,固定其参数,使其保留对双三次退化的低分辨率图像的重建能力.

### 1.1.2 双三次特征调制模块

由于合成的双三次退化数据集和真实场景中的退化图像在域上存在差异,所以直接将预训练好的主干网络应用于真实场景的低分辨率图像时,效果欠佳.为了解决这个问题,本文提出了双三次特征调制模块.当输入同一场景的未知退化的低分辨率图像时,该模块的主要作用是将未知退化的网络中间层特征调制为主干网络易于处理的双三次特征.

给定一张同一场景的未知退化的低分辨率图像 $I_{LR}^u \in \mathbb{R}^{H \times W \times 3}$ (其中 $H, W$ 表示图像的高度和宽度,3代表通道数),将图像输入到固定参数的预训练主干网络中,同时也输入到双三次特征调制模块中,该模块用学习到的调制参数将主干网络中间层的特征调制为更易于恢复的双三次特征.具体而言,首先通过一个 $3 \times 3$ 卷积层 $H'_{SF}(\cdot)$ 来提取浅层特征 $F'_0 \in \mathbb{R}^{H \times W \times C}$ :

$$F'_0 = H'_{SF}(I_{LR}^u) \quad (6)$$

式中: $I_{LR}^u$ 表示输入的未知退化的低分辨率图像; $H'_{SF}$ 表示卷积层; $F'_0$ 表示经卷积层提取的浅层特征; $C$ 表示通道数.然后,从 $F'_0$ 中提取深层特征 $F'_{DF} \in \mathbb{R}^{H \times W \times C}$ :

$$F'_{DF} = H'_{DF}(F'_0) \quad (7)$$

式中: $F'_{DF}$ 表示双三次特征调制模块提取的深层特征; $H'_{DF}$ 表示双三次特征调制网络的深层特征提取模块,由 $N(N=15)$ 个残差块(Residual Block, RB)组成,每个残差块包含三层卷积层.具体而言,从RB块中提取中间层的特征 $F'_1, F'_2, \dots, F'_N$ 以及输出的深层特征 $F'_{DF}$ 如下所示:

$$F'_j = H_{RB_j}(F'_{j-1}), j = 1, 2, \dots, N \quad (8)$$

$$F'_{DF} = H_{CONV}(F'_0 + F'_N) \quad (9)$$

式中: $H_{RB_j}(\cdot)$ 表示第 $j$ 个残差块; $F'_{j-1}$ 表示第 $j$ 个残差块的输入特征; $F'_j$ 表示第 $j$ 个残差块的输出特征; $F'_N$ 表示第 $N$ 个残差块的输出特征; $H_{CONV}$ 表示最后一层的卷积层.接下来,通过空间特征转换层<sup>[26]</sup>学习一个映射函数:

$$M: F'_{DF} \rightarrow (\gamma_i, \beta_i), i = 1, 2, \dots, K \quad (10)$$

式中: $(\gamma_i, \beta_i)$ 表示从提取出来的深层特征中预测出的第 $i$ 对调制参数,在得到 $(\gamma_i, \beta_i)$ 后,通过对主干网络中间层的特征进行缩放和平移来实现仿射变换:

$$SFT(F_i | \gamma_i, \beta_i) = \gamma_i \odot F_i + \beta_i, i = 1, 2, \dots, K \quad (11)$$

式中: $SFT$ 表示仿射变换操作; $F_i$ 表示主干网络中第 $i$ 个RSTB块输出的中间层特征,其维度与 $\gamma_i$ 和 $\beta_i$ 一样;符号 $\odot$ 表示逐元素相乘操作.通过对未知退化图像的每个中间层特征进行仿射变换,调制参数可以自

适应地调制这些特征.

为了使主干网络中调制后的中间层特征接近双三次特征,该方法首先将双三次退化的图像输入到主干网络中,将其中间层的特征  $F_1^b, F_2^b, \dots, F_K^b$  提取出来,同时提取调制后的特征  $F_1^u, F_2^u, \dots, F_K^u$ ,然后计算相应层特征的损失:

$$L_{\text{feature}} = L_1 \text{Loss}(F_1^u, F_1^b) + L_1 \text{Loss}(F_2^u, F_2^b) + \dots + L_1 \text{Loss}(F_K^u, F_K^b) \quad (12)$$

式中:  $L_1 \text{Loss}$  表示  $L_1$  损失函数,  $L_{\text{feature}}$  表示中间层特征损失函数.

由于主干网络的参数已固定,所以其重建双三次退化特征的能力得到保留,因此该方法只训练双三次特征调制模块,使其具备将未知退化的特征调制为双三次特征的能力.

### 1.1.3 多分类鉴别器

为了解决自然图像中多种复杂分布的问题,本文进一步提出了多分类鉴别器.通过设置多个鉴别器,以期不同的鉴别器能够对不同的图像分布特征进行分类,以此更精细地建模图像的细粒度分布特征.具体而言,以  $n$  个鉴别器实现  $n+1$  分类为例:当输入高分辨率图像时,所有  $n$  个鉴别器均输出类别  $y_{\text{HR}} = [1, 0, 0, \dots, 0]$ .当重建图像输入时,第一个鉴别器的输出类别为  $y_{\text{SR}_1} = [0, 1, 0, \dots, 0]$ ,第二个鉴别器的输出类别为  $y_{\text{SR}_2} = [0, 0, 1, \dots, 0]$ ,以此类推,第  $n$  个鉴别器的输出类别为  $y_{\text{SR}_n} = [0, 0, 0, \dots, 1]$ .这种设计使每个鉴别器能够专注于图像的特定分布,从而增强重建图像局部细节的可视化效果.损失函数如下所示:

$$\begin{aligned} \{ & L_{D_1} = \text{CELoss}(D_1(I_{\text{HR}}), y_{\text{HR}}) + \text{CELoss}(D_1(I_{\text{SR}}), y_{\text{SR}_1}) \\ & L_{D_2} = \text{CELoss}(D_2(I_{\text{HR}}), y_{\text{HR}}) + \text{CELoss}(D_2(I_{\text{SR}}), y_{\text{SR}_2}) \\ & \dots\dots \\ & L_{D_n} = \text{CELoss}(D_n(I_{\text{HR}}), y_{\text{HR}}) + \text{CELoss}(D_n(I_{\text{SR}}), y_{\text{SR}_n}) \} \end{aligned} \quad (13)$$

$$L_D = L_{D_1} + L_{D_2} + \dots + L_{D_n} \quad (14)$$

式中:  $\text{CELoss}$  表示交叉熵损失函数;  $D_1$  表示第一个分类器;  $D_2$  表示第二个分类器;  $D_n$  表示第  $n$  个分类器;  $L_D$  表示多分类鉴别器总的损失函数;  $I_{\text{HR}}$  表示高分辨率图像标签,  $I_{\text{SR}}$  表示重建的高分辨率图像.

## 1.2 损失函数

该方法首先对 SwinIR 主干网络进行训练,使其具备恢复双三次退化低分辨率图像的能力,训练主干网络的损失函数如下:

$$L_{\text{SR}} = L_{\text{recon}} + \lambda_p L_{\text{per}} \quad (15)$$

式中:  $L_{\text{recon}}$  表示像素级的  $L_1$  重建损失函数<sup>[4]</sup>;  $L_{\text{per}}$  表示感知损失函数<sup>[27]</sup>; 超参数  $\lambda_p$  控制感知损失函数的权重.在主干网络训练完成后,固定其参数,然后利用多分类鉴别器以生成对抗的策略训练双三次特征调制网络,此阶段的损失函数如下所示:

$$L_{\text{SR}} = L_{\text{recon}} + \lambda_p L_{\text{per}} + \lambda_f L_{\text{feature}} + \lambda_d L_D \quad (16)$$

式中:  $L_{\text{feature}}$  表示约束网络中间层特征的损失函数;  $L_D$  表示多分类鉴别器的损失函数;  $\lambda_f$  和  $\lambda_d$  分别表示这两项损失函数的权重系数.

## 2 实验

遵循先前的研究<sup>[3,19]</sup>,本文聚焦实现 4 倍的图像超分辨率工作.

### 2.1 实验设置

训练集:本文采用 DIV2K<sup>[28]</sup>, Flickr2K<sup>[28]</sup> 和 OST<sup>[26]</sup> 训练集,先将训练集中的低分辨率图像裁剪为  $120 \times 120$  的图像块,将高分辨率图像裁剪为  $480 \times 480$  的图像块;然后使用 BSRGAN 合成未知退化的低分辨率图像.训练期间,该方法随机地对图像进行  $90^\circ, 180^\circ, 270^\circ$  的旋转以及水平翻转来增强数据.

测试集:该方法采用以下测试集评估模型性能.(1)针对真实场景,本文选用 RealSR<sup>[17]</sup> 和 DRealSR<sup>[18]</sup> 测试集,并参照 SeeSR<sup>[29]</sup> 的操作,将数据集中的低分辨率图像中心裁剪为  $128 \times 128$  的子图,将高分辨率图像相应地裁剪为  $512 \times 512$  的子图.(2)为进行可视化对比,使用了论文[30]中收集的无标签真实场景的测试集 DPED.

实施细节:本文采用 SwinIR 作为主干网络,双三次特征调制模块中残差块的数量设置为 15. 首先,使用 Adam<sup>[31]</sup> 优化器对主干网络进行 100K 次迭代训练, batch\_size 为 8, 学习率为 0.000 2; 接着, 固定主干网络的参数, 训练双三次特征调制模块 100K 次, batch\_size 和学习率设置为 8 和 0.000 2, 设置感知损失函数的权重  $\lambda_p$  为 1, 中间层特征损失函数权重  $\lambda_f$  为 1, 多分类鉴别器损失函数权重  $\lambda_d$  为 0.1. 训练均在 4 张 NVIDIA 24G 4090 GPU<sub>s</sub> 上完成.

评价指标:为全面评估不同方法,本文采用一系列参考和非参考指标. 由于 PSNR 和 SSIM<sup>[32]</sup> 指标没有充分考虑到人眼视觉系统对图像结构、纹理和内容的感知方式,因此本文只对比 LPIPS<sup>[33]</sup>、DISTS<sup>[34]</sup> 和 NIQE<sup>[35]</sup> 三个感知指标. LPIPS 和 DISTS 是基于参考的感知质量评价指标, NIQE 是基于非参考的图像感知质量评价标准,三个感知指标的值越小越好.

对比方法:本文将提出的方法和最先进的真实场景的图像超分辨率方法进行对比,包括:ESRGAN<sup>[6]</sup>、RealSR<sup>[36]</sup>、DAN<sup>[37]</sup>、LDL<sup>[24]</sup>、DASR<sup>[20]</sup>、FeMaSR<sup>[38]</sup>、ReDSR<sup>[39]</sup>.

## 2.2 定量和定性对比实验

### 2.2.1 定量对比

表 1 显示了本文提出的方法在三个真实场景的数据集上的定量对比结果. 在对比实验中, 本文将多分类鉴别器的数目设置为 2. 由表可知, 在 RealSR 数据集上, 本文的方法在三个最主要的感知指标 LPIPS、DISTS 和 NIQE 上均实现了最佳结果, 分别比第二名提升了 0.43%、4.28% 和 0.23%, 这得益于本文提出的双三次特征调制模块和多分类鉴别器. 而在 DRealSR 数据集上, 本文的方法在 LPIPS 和 DISTS 指标上也取得了最佳的结果, 分别比第二名提升了 0.94% 和 3.48%, 在 NIQE 指标上则位列第二. 综合来看, 本文的方法在真实场景的测试集上表现出色.

### 2.2.2 定性对比

图 2 和图 3 展现了不同真实场景的图像超分辨率方法在 RealSR 和 DRealSR 数据集上的可视化结果. 首先, 正如图 2 第一行和第二行所示, 对比的几种真实场景超分辨率方法所恢复出来的图像的整体建筑细节依然很模糊, 而本文的方法在建筑物纹理的恢复上表现出色. 在第三行中, ESRGAN、RealSR、DAN、LDL 所生成的图像在人脸皱纹上的细节仍很模糊, 未能恢复出清晰的细节. 而 FeMaSR 恢复出来的图像过于锐化, 以至于效果显得不自然. 而本文的方法恢复出来的皱纹细节和高分辨率原图最为接近. 和图二的结果相似, 图三第一行中的几种对比方法并不能清晰地恢复出草的细节, 而本文的方法恢复出来的图像在细节方面具有明显优势. 在第二行中, 采用 FeMaSR 方法恢复的花瓣纹路颜色失真, 本文的方法恢复出来的最为清晰. 在第三行的对比中, FeMaSR 和 ReDSR 甚至产生了严重的失真变形, 而本文的方法对阳台条纹细节的生成和高分辨率标签最为接近.

为了评估该模型在真实场景中的恢复能力, 本文在 DPED<sup>[30]</sup> 测试集上进行可视化效果对比. 由于 DPED 测试集只有低分辨率图像, 没有相对应的高分辨率图像标签, 因此本文只进行了可视化对比实验. 如图 4 所示, 在第一行的可视化结果中, 只有本文的方法能将窗户的细节恢复出来, 其他方法产生的图像细节依旧很模糊. 在第二行的可视化对比图像中, 其他方法引入了一些伪影, 尽管 FeMaSR 方法恢复的图像相对清晰, 但棚子的条理却显得扭曲, 相比之下, 本文的方法恢复出来的棚子的细节更加真实. 第三行的可视化结果中, 所对比的几种方法在很暗的环境下无法有效地恢复出栏杆的细节, 而本文的方法则成功地重建了栏杆的形状及细节, 具有很好的视觉效果. 综合来看, 本文提出的方法在真实场景的低分辨率图像上表现出色.

表 1 在真实场景的数据集上和最先进的方法的定量对比

Tab. 1 Quantitative comparison of state-of-the-art methods on real-world datasets

数据集	方法	LPIPS ↓	DISTS ↓	NIQE ↓
RealSR	ESRGAN	0.187 3	0.238 5	5.916 6
	RealSR	0.186 1	0.238 8	6.370 6
	DAN	0.178 1	0.233 5	8.382 1
	LDL	0.184 3	0.236 6	<u>5.649 6</u>
	DASR	0.165 1	0.206 9	6.687 7
	FeMaSR	<u>0.162 4</u>	<u>0.193 9</u>	5.768 8
	ReDSR	0.165 6	0.206 7	7.594 9
	<b>Ours</b>	<b>0.161 7</b>	<b>0.185 6</b>	<b>5.636 6</b>
DRealSR	ESRGAN	0.189 2	0.232 4	9.585 8
	RealSR	0.192 1	0.235 1	9.764 6
	DAN	0.185 5	0.224 2	10.001 3
	LDL	0.186 7	0.230 2	9.336 3
	DASR	<u>0.159 5</u>	<u>0.192 6</u>	7.848 5
	FeMaSR	0.168 9	0.202 2	6.041 6
	ReDSR	0.180 8	0.212 5	9.107 6
	<b>Ours</b>	<b>0.158 0</b>	<b>0.185 9</b>	<b>6.864 7</b>

注:最好的和次好的分别用粗体和下划线标出.

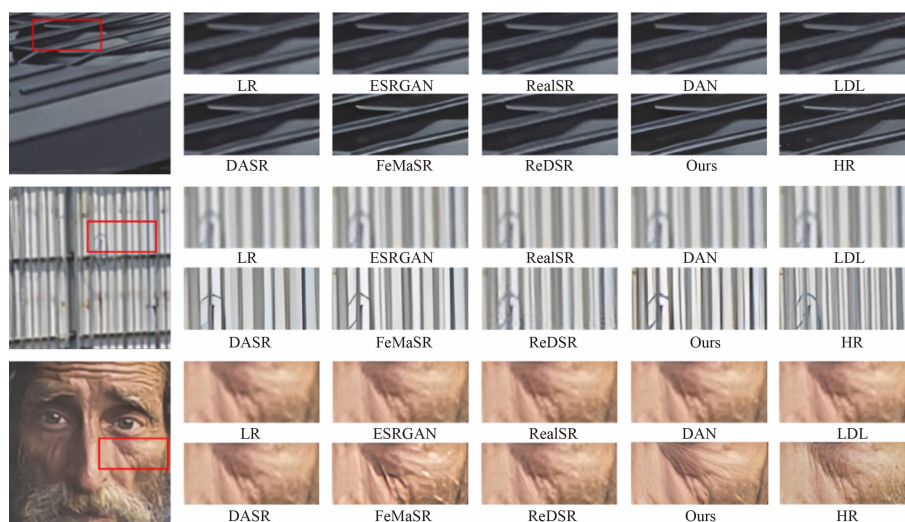


图 2 不同方法在 RealSR 数据集上的可视化结果

Fig. 2 Visualization results of different methods on the RealSR dataset

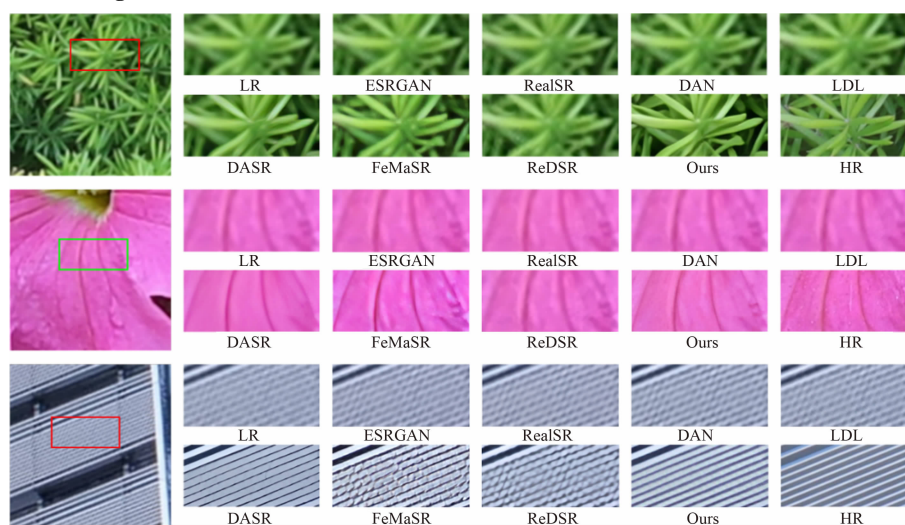


图 3 不同方法在 DRealSR 数据集上的可视化结果

Fig. 3 Visualization results of different methods on the DRealSR dataset

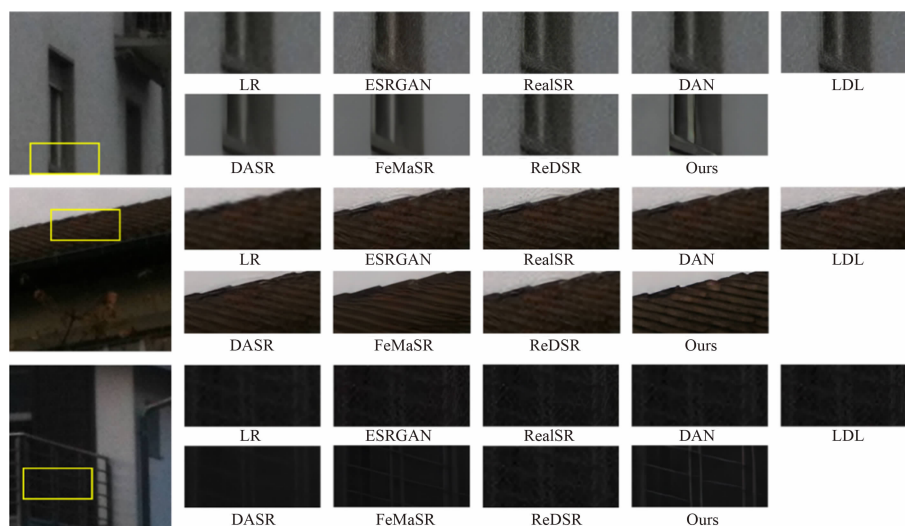


图 4 不同方法在真实场景图像的可视化结果

Fig. 4 Visualization results of different methods on real - world images

### 2.3 消融实验

为了探究本文提出的双三次特征调制模块和多分类鉴别器对模型性能的影响,以及多分类鉴别器从几个角度进行鉴别效果最佳,本文在 RealSR 和 DRealSR 测试集上进行消融实验,该实验只对比 LPIPS、DISTS 和 NIQE 这三个感知指标.在实验中,本文将 SwinIR 主干网络表示为 Baseline,将双三次特征调制模块表示为 FMN,将多分类鉴别器表示为 MCD,将全局鉴别器表示为 GD.为了验证每个模块的作用,逐步将它们添加到 Baseline 中.

#### 2.3.1 双三次特征调制模块的有效性

为了将真实场景的未知退化的低分辨率图像的特征调制为主干网络易于处理的双三次特征,本文引入了双三次特征调制模块 FMN.由表 2 可知,在 Baseline 的基础上添加 FMN,同 Baseline 进行对比,在 RealSR 测试集上, LPIPS 和 DISTS 指标分别提升了 1.5% 和 12.41%;在 DrealSR 数据集上 LPIPS、DISTS、NIQE 三个指标分别提升了 7.48%、11.41% 和 1.4%.由图 5 前两张图像可视化结果对比可知,加上 FMN,图像的层次感也更丰富.由此可证明,本文提出的双三次特征调制模块的有效性.

表 2 特征调制网络和多分类鉴别器对模型性能的影响

**Tab. 2 Influence of feature modulation network and multi-classification discriminator on model performance**

数据集	方法	LPIPS ↓	DISTS ↓	NIQE ↓
RealSR	Baseline	0.179 9	0.236 1	8.543 7
	Baseline + FMN	0.177 2	0.206 8	9.316 8
	Baseline + FMN + GD	0.169 4	0.198 1	6.075 4
	Baseline + FMN + MCD	<b>0.161 7</b>	<b>0.185 6</b>	<b>5.636 6</b>
DRealSR	Baseline	0.188 5	0.231 4	10.988 2
	Baseline + FMN	0.174 4	0.205 0	10.834 3
	Baseline + FMN + GD	0.166 1	0.204 9	7.264 3
	Baseline + FMN + MCD	<b>0.158 0</b>	<b>0.185 9</b>	<b>6.864 7</b>

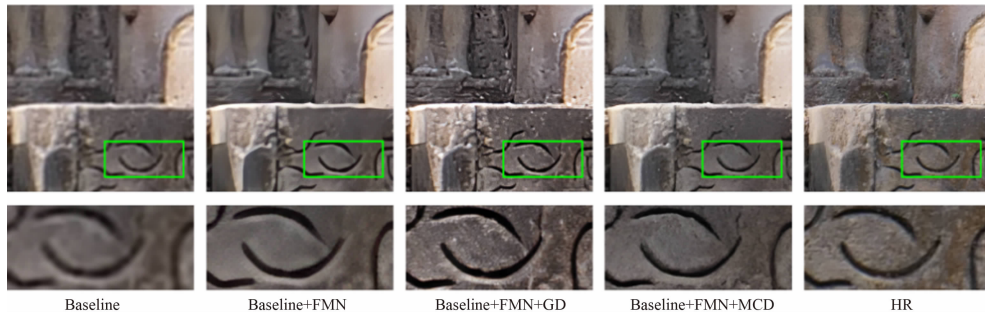


图 5 双三次特征调制模块和多分类鉴别器消融实验的可视化结果

Fig. 5 Visualization results of the ablation experiment of the bicubic feature modulation module and the multi-classification discriminator

#### 2.3.2 多分类鉴别器的有效性

为了对整张图像的不同分布进行鉴别,本文引入了多分类鉴别器.为了证明多分类鉴别器 MCD 的有效性,在进行消融实验时,本文将多分类鉴别器 MCD 替换为全局鉴别器 GD.由表 2 可知,与全局鉴别器 GD 相比,多分类鉴别器 MCD 在 RealSR 测试集上三个感知指标分别提升了 4.55%、6.31% 和 7.22%;在 DrealSR 测试集上分别提升了 4.88%、9.27% 和 5.5%.由图 5 的可视化结果可知,采用全局鉴别器 GD,图像有明显的过于锐化的痕迹,而采用本文提出的 MCD,恢复的图像更加真实,由此可证明多分类鉴别器的有效性.

#### 2.3.3 多分类鉴别器分类数目的影响

为了研究从图像的几个角度对图像进行鉴别效果最佳,本文将多分类鉴别器的分类数分别设置为 2、

3、4、5、6 进行消融实验. 由表 3 可知, 在 RealSR 和 DRealSR 测试集上, 从图像的两个角度对图像进行鉴别时, 三个定量指标最佳, 鉴别效果最优; 随着判别角度的增加, 性能逐渐下降, 本文认为是随着判别器数目增加的过多, 生成器和多个判别器的对抗训练变得越来越不稳定, 从而导致感知指标也越来越差. 图 6 也验证了该实验的结果. 和从两个角度可视化出来的结果相比, 从三个角度可视化的结果在绿色衣服布料的缝隙上弱了一些, 随着判别器数目的增多, 对抗训练越来越不稳定, 以至于超分辨率生成的图像越来越模糊, 从六个角度进行鉴别图像时, 甚至出现了伪影.

表 3 多分类器分类数对模型性能的影响

Tab. 3 Effect of the number of multi-classifier categories on model performance

数据集	分类器分类数量	<i>LPIPS</i> ↓	<i>DISTS</i> ↓	<i>NIQE</i> ↓
RealSR	2	0.161 7	0.185 6	5.636 6
	3	0.164 6	0.188 7	5.844 3
	4	0.172 3	0.197 0	5.929 7
	5	0.180 6	0.207 5	5.683 6
	6	0.187 4	0.210 7	5.735 6
DRealSR	2	0.158 0	0.185 9	6.864 7
	3	0.158 1	0.183 3	7.090 9
	4	0.167 9	0.194 4	7.847 4
	5	0.176 2	0.203 2	7.530 5
	6	0.178 5	0.206 0	7.463 3

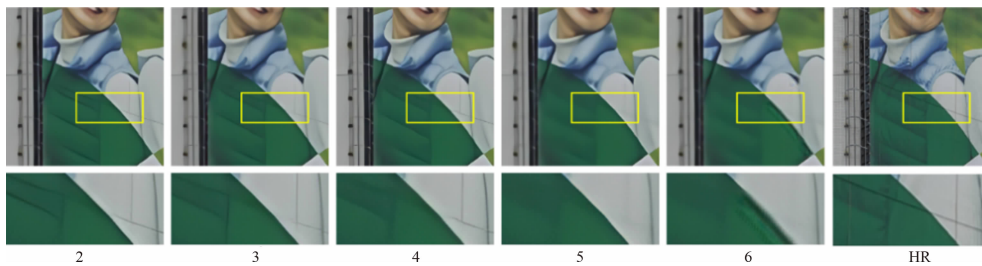


图 6 多分类鉴别器分类数目消融实验的可视化结果

Fig. 6 Visualization results of the ablation experiment on the number of categories of the multi-class discriminator

### 3 结论

本文提出了双三次特征调制和多分类鉴别器图像超分辨率网络, 双三次特征调制模块通过将真实场景复杂退化的图像特征调制为主干网络易于恢复的双三次特征, 有效解决了真实场景未知退化图像的恢复难题. 而多分类鉴别器通过设置多个鉴别器, 使每个鉴别器专注于鉴别图像中的一种分布特征, 通过将各鉴别器的输出映射至不同类别, 实现对图像分布的细粒度建模. 本文的实验证明了所提出的方法超出了以往大多数方法.

#### 参考文献:

- [1] 蔡念, 张海员, 张楠. 基于 Contourlet 的改进加权抛物线插值图像超分辨率算法[J]. 上海交通大学学报, 2011, 45(10): 1509 - 1515.  
CAI N, ZHANG H Y, ZHANG N. Using improved weighted parabolic interpolation and contourlet transformation to zoom images for super-resolution[J]. Journal of Shanghai Jiao Tong University, 2011, 45(10): 1509 - 1515.
- [2] PARK K, SOH J W, CHO N I. Efficient attention-sharing information distillation transformer for lightweight single image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(6): 6416 - 6424.
- [3] WANG X T, XIE L B, DONG C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 11 - 17 October 2021, Montreal,

- BC, Canada. IEEE, 2021; 1905 – 1914.
- [4] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super – resolution [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops ( CVPRW ), 21 – 26 July 2017, Honolulu, HI, USA. IEEE, 2017; 1132 – 1140.
- [5] HUANG J Q, LI K, JIA J F, et al. Single image super – resolution through image pixel information clustering and generative adversarial network [J]. *Big Data Mining and Analytics*, 2025, 8(5) : 1044 – 1059.
- [6] WANG X T, YU K, WU S X, et al. ESRGAN: Enhanced super – resolution generative adversarial networks [C]//Computer Vision – ECCV 2018 Workshops. Cham: Springer, 2019; 63 – 79.
- [7] 刘志坚, 陶韵旭, 刘航, 等. 融合残差密集与生成对抗网络的红外巡检图像超分辨率重建 [J]. *昆明理工大学学报(自然科学版)*, 2023, 48(5) : 120 – 129.
- LIU Z J, TAO Y X, LIU H, et al. Super – resolution reconstruction of infrared inspection images by integrating residual dense and generative adversarial networks [J]. *Journal of Kunming University of Science and Technology ( Natural Science )*, 2023, 48(5) : 120 – 129.
- [8] CAI Q, LI J X, LI H F, et al. TDPN: Texture and detail – preserving network for single image super – resolution [J]. *IEEE Transactions on Image Processing*, 2022, 31 : 2375 – 2389.
- [9] KIM J, LEE J K, LEE K M. Accurate image super – resolution using very deep convolutional networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition ( CVPR ), 27 – 30 June 2016, Las Vegas, NV, USA. IEEE, 2016; 1646 – 1654.
- [10] LAI W S, HUANG J B, AHUJA N, et al. Deep Laplacian pyramid networks for fast and accurate super – resolution [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition ( CVPR ), 21 – 26 July 2017, Honolulu, HI, USA. IEEE, 2017; 5835 – 5843.
- [11] LI H F, YUAN M, LI J X, et al. Focus affinity perception and super – resolution embedding for multifocus image fusion [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(3) : 4311 – 4325.
- [12] WANG Q Q, GAO Q X, WU L L, et al. Adversarial multi – path residual network for image super – resolution [J]. *IEEE Transactions on Image Processing*, 2021, 30 : 6648 – 6658.
- [13] YAN Y Y, REN W Q, HU X B, et al. SRGAT: Single image super – resolution with graph attention network [J]. *IEEE Transactions on Image Processing*, 2021, 30 : 4905 – 4918.
- [14] ZHANG Y F, HUANG Y Q, WANG K Z, et al. Single image super – resolution reconstruction with preservation of structure and texture details [J]. *Mathematics*, 2023, 11(1) : 1 – 17.
- [15] 黄裕青, 李华锋, 原铭, 等. 基于卷积神经网络梯度和纹理补偿的单幅图像超分辨率重建 [J]. *数据采集与处理*, 2023, 38(5) : 1112 – 1124.
- HUANG Y Q, LI H F, YUAN M, et al. Super – resolution reconstruction of single image based on convolutional neural network gradient and texture compensation [J]. *Journal of Data Acquisition and Processing*, 2023, 38(5) : 1112 – 1124.
- [16] GOODFELLOW I, POUGET – ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. *Communications of the ACM*, 2020, 63(11) : 139 – 144.
- [17] CAI J R, ZENG H, YONG H W, et al. Toward real – world single image super – resolution: A new benchmark and a new model [C]//2019 IEEE/CVF International Conference on Computer Vision ( ICCV ), 27 October 2019 – 2 November 2019, Seoul, Korea. IEEE, 2020; 3086 – 3095.
- [18] WEI P X, XIE Z W, LU H N, et al. Component divide – and – conquer for real – world image super – resolution [C]//Computer Vision – ECCV 2020. Cham: Springer, 2020; 101 – 117.
- [19] ZHANG K, LIANG J Y, VAN GOOL L, et al. Designing a practical degradation model for deep blind image super – resolution [C]//2021 IEEE/CVF International Conference on Computer Vision ( ICCV ), 10 – 17 October 2021, Montreal, QC, Canada. IEEE, 2022; 4771 – 4780.
- [20] LIANG J, ZENG H, ZHANG L. Efficient and Degradation – adaptive network for Real – world image super – resolution [C]//Computer Vision – ECCV 2022. Cham: Springer, 2022; 574 – 591.
- [21] MOU C, WU Y Z, WANG X T, et al. Metric learning based interactive modulation for Real – world super – resolution [C]//Computer Vision – ECCV 2022. Cham: Springer, 2022; 723 – 740.

- [22] WU R Y, YANG T, SUN L C, et al. SeeSR: Towards semantics – aware real – world image super – resolution [ C ] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition ( CVPR ), 16 – 22 June 2024, Seattle, WA, USA. IEEE, 2024; 25456 – 25467.
- [23] LEDIG C, THEIS L, HUSZÁR F, et al. Photo – realistic single image super – resolution using a generative adversarial network [ C ] // 2017 IEEE Conference on Computer Vision and Pattern Recognition ( CVPR ), 21 – 26 July 2017, Honolulu, HI, USA. IEEE, 2017; 105 – 114.
- [24] LIANG J, ZENG H, ZHANG L. Details or artifacts: A locally discriminative learning approach to realistic image super – resolution [ C ] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition ( CVPR ), 18 – 24 June 2022, New Orleans, LA, USA. IEEE, 2022; 5647 – 5656.
- [25] ZHANG K, VAN GOOL L, TIMOFTE R. Deep unfolding network for image super – resolution [ C ] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition ( CVPR ), 13 – 19 June 2020, Seattle, WA, USA. IEEE, 2020; 3214 – 3223.
- [26] WANG X T, YU K, DONG C, et al. Recovering realistic texture in image super – resolution by deep spatial feature transform [ C ] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18 – 23 June 2018, Salt Lake City, UT, USA. IEEE, 2018; 606 – 615.
- [27] BELL S, UPCHURCH P, SNAVELY N, et al. Material recognition in the wild with the materials in context database [ C ] // 2015 IEEE Conference on Computer Vision and Pattern Recognition ( CVPR ), 07 – 12 June 2015, Boston, MA, USA. IEEE, 2015; 3479 – 3487.
- [28] AGUSTSSON E, TIMOFTE R. NTIRE 2017 challenge on single image super – resolution: Dataset and study [ C ] // 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops ( CVPRW ), 21 – 26 July 2017, Honolulu, HI, USA. IEEE, 2017; 1122 – 1131.
- [29] WANG J Y, YUE Z S, ZHOU S C, et al. Exploiting diffusion prior for real – world image super – resolution [ J ]. International Journal of Computer Vision, 2024, 132( 12 ): 5929 – 5949.
- [30] IGNATOV A, KOBYSHEV N, TIMOFTE R, et al. DSLR – quality photos on mobile devices with deep convolutional networks [ C ] // 2017 IEEE International Conference on Computer Vision ( ICCV ), 22 – 29 October 2017, Venice, Italy. IEEE, 2017; 3297 – 3305.
- [31] KINGMA D P, BA J. Adam: A method for stochastic optimization [ C ] // Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA: ICLR, 2015. <https://arxiv.org/abs/1412.6980>.
- [32] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity [ J ]. IEEE Transactions on Image Processing, 2004, 13( 4 ): 600 – 612.
- [33] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [ C ] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18 – 23 June 2018, Salt Lake City, UT, USA. IEEE, 2018; 586 – 595.
- [34] DING K Y, MA K D, WANG S Q, et al. Image quality assessment: Unifying structure and texture similarity [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44( 5 ): 2567 – 2581.
- [35] ZHANG L, ZHANG L, BOVIK A C. A feature – enriched completely blind image quality evaluator [ J ]. IEEE Transactions on Image Processing, 2015, 24( 8 ): 2579 – 2591.
- [36] JI X Z, CAO Y, TAI Y, et al. Real – world super – resolution via kernel estimation and noise injection [ C ] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops ( CVPRW ), 14 – 19 June 2020, Seattle, WA, USA. IEEE, 2020; 1914 – 1923.
- [37] LUO Z X, HUANG Y, LI S, et al. Unfolding the alternating optimization for blind super resolution [ EB/OL ]. Arxiv, 2020: 2010. 02631. <https://arxiv.org/abs/2010.02631>.
- [38] CHEN C F, SHI X Y, QIN Y P, et al. Real – world blind super – resolution via feature matching with implicit high – resolution priors [ C ] // Proceedings of the 30th ACM International Conference on Multimedia, Lisboa Portugal. ACM, 2022; 1329 – 1338.
- [39] LIU H D, WANG L G, ZHANG Y, et al. Preserving full degradation details for blind image super – resolution [ EB/OL ]. Arxiv, 2024; 2407. 01299. <https://arxiv.org/abs/2407.01299>.

( 编 辑: 朱银周 )