

# 基于真实退化估计与高频引导的内窥镜图像超分辨率重建

李 嫣<sup>1,2</sup> 任文琦<sup>3,4</sup> 张长青<sup>5</sup> 张金刚<sup>3</sup> 聂云峰<sup>6</sup>

**摘 要** 内窥镜是诊断人体器官疾病的重要医疗设备, 然而受人体内部腔环境影响, 内窥镜图像分辨率一般较低, 需对其进行超分辨率处理. 目前多数基于深度学习的超分辨率算法直接使用双三次插值下采样从高质量图像中获取低分辨率 (Low-resolution, LR) 图像以进行配对训练, 此种方式会导致纹理细节丢失, 不适用于医学图像. 为解决该问题, 针对医学内窥镜图像开发了一种新颖的退化框架, 首先从真实低质量内窥镜图像中提取丰富多样的真实模糊核与噪声模式, 之后提出一种退化注入算法, 利用提取的真实模糊核与噪声将高分辨率 (High-resolution, HR) 内窥镜图像退化为符合真实域的低分辨率图像. 同时, 提出一种高频引导的残差密集超分辨率网络, 采用基于双频率信息交互的频率分离策略, 并设计多层级融合机制, 将提取的多级高频信息逐层嵌入残差密集模块的多层特征, 以充分恢复内窥镜图像的高频细节和低频内容. 在合成与真实数据集上的大量实验表明, 我们的方法优于对比方法, 具有更好的主客观质量评价.

**关键词** 内窥镜图像超分辨率, 退化估计, 高频引导, 卷积神经网络

**引用格式** 李嫣, 任文琦, 张长青, 张金刚, 聂云峰. 基于真实退化估计与高频引导的内窥镜图像超分辨率重建. 自动化学报, 2024, 50(2): 334-347

**DOI** 10.16383/j.aas.c230070

## Super-resolution of Endoscopic Images Based on Real Degradation Estimation and High-frequency Guidance

LI Yan<sup>1,2</sup> REN Wen-Qi<sup>3,4</sup> ZHANG Chang-Qing<sup>5</sup> ZHANG Jin-Gang<sup>3</sup> NIE Yun-Feng<sup>6</sup>

**Abstract** Endoscopes are effective medical devices for diagnosing diseases of human organs. However, due to the influence of the internal cavity environment of the human body, the resolution of endoscope images is generally low. Most existing deep learning-based super-resolution algorithms directly use bicubic interpolation downsampling to obtain low-resolution (LR) images from high-quality images for paired training. However, these methods will lead to texture details loss and are not suitable for medical images. To solve this problem, this paper proposes a novel degradation framework for medical endoscopic images. First, diverse realistic blur kernels and noise patterns are extracted from real-world low-quality endoscopic images, and then a degradation injection algorithm is proposed. The extracted real blur kernels and noise degrade the high-resolution (HR) endoscopic image into a low-resolution image. In addition, this paper proposes a high-frequency guided residual dense super-resolution network, which adopts a frequency separation strategy based on dual-frequency information interaction. And a multi-level fusion mechanism is designed to embed the extracted multi-level high-frequency information into the multi-layer features of the residual dense module layer by layer. This helps recover the high-frequency details and low-frequency content of the endoscopic image. Extensive experiments on synthetic and real-world datasets show that our method outperforms the contrastive methods with better subjective and objective quality evaluations.

**Key words** Endoscopic image super-resolution, degradation estimation, high-frequency guidance, convolutional neural network

**Citation** Li Yan, Ren Wen-Qi, Zhang Chang-Qing, Zhang Jin-Gang, Nie Yun-Feng. Super-resolution of endoscopic images based on real degradation estimation and high-frequency guidance. *Acta Automatica Sinica*, 2024, 50(2): 334-347

收稿日期 2023-02-23 录用日期 2023-10-21

Manuscript received February 23, 2023; accepted October 21, 2023

中国科学院网络安全和信息化专项 (CAS-WX2022SF-0102), 深圳市科技计划 (JCYJ20220530145209022) 资助

Supported by Chinese Academy of Sciences Cyber Security and Informatization Project (CAS-WX2022SF-0102) and the Shenzhen Science and Technology Program (JCYJ20220530145209022)

本文责任编辑 张道强

Recommended by Associate Editor ZHANG Dao-Qiang

1. 中国科学院信息工程研究所 北京 100093 中国 2. 中国科学院大学网络空间安全学院 北京 100190 中国 3. 中国科学院大学

未来技术学院 北京 100039 中国 4. 中山大学网络空间安全学院 深圳 518107 中国 5. 天津大学智能与计算学部 天津 300350 中国 6. 布鲁塞尔自由大学应用物理与光子学系 布鲁塞尔 1050 比利时

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China 2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100190, China 3. School of Future Technology, University of Chinese Academy of Sciences, Beijing 100039, China 4. School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen 518107, China 5. College of Intelligence and Computing, Tianjin University, Tianjin 300350, China 6. Department of Applied Physics and Photonics, Vrije Universiteit Brussel, Brussels 1050, Belgium

医用内窥镜是一种基于图像的医学检查仪器,能够准确观察人体内部器官形态和病理变化,辅助医师进行疾病的诊断及治疗<sup>[1-2]</sup>。然而对于耳鼻喉等内腔直径较小的部位,医用内窥镜的直径受到限制,其输出图像的分辨率普遍较低。此外,人体内部环境昏暗,仅凭内窥镜自带光源照亮,会导致采集的图像存在噪声、纹理特征不清晰、模糊等退化问题<sup>[3-4]</sup>。低分辨率的内窥镜图像不仅影响医生对病灶的判断,还阻碍后续研究,如异常区域分割、深度估计等。因此,如何提升低分辨率内窥镜图像的质量已经成为一个重要的医学图像问题。

图像超分辨率重建是提高图像质量的一种有效手段,其能将低分辨率图像 (Low-resolution, LR) 转换为高分辨率图像 (High-resolution, HR)。近年来,基于深度学习的图像超分辨率算法展现出了良好的性能,其利用非线性变换自动学习多层次特征,更深入地挖掘了 LR 与 HR 之间的内在联系<sup>[5]</sup>。Dong 等<sup>[6]</sup>提出了第一个使用卷积神经网络的端到端超分辨率算法 SRCNN,相较于传统方法,所获得的图像边缘更清晰,包含的伪影更少。Ledig 等<sup>[7]</sup>将生成对抗网络引入超分辨率算法,并提出了新的感知损失函数,使重建图像更具有真实感。Lim 等<sup>[8]</sup>通过去除批归一化层对残差块进行优化,并扩大了网络深度,使重构图像的质量有效提升。Ahn 等<sup>[9]</sup>提出了一种轻量级的级联残差网络 CARN,提升了超分辨率模型的效率和性能。以上述方法为代表的深度超分辨率算法大多针对自然图像开发和训练,而内窥镜图像与自然图像存在较大差异。首先是噪声干扰不同,自然图像通常由相机拍摄,噪声多为高斯噪声;内窥镜图像由内窥镜设备拍摄,噪声多为条纹噪声<sup>[10]</sup>。其次是图像场景内容不同,内窥镜图像的场景多为人体器官,色彩信息单一,并且包含重要的、细小的、不明显的医学纹理细节;而自然图像场景丰富,纹理细节较为明显。此外,公开的医学图像数据集种类和包含的图像数量有限,一般不存在大量成对的 LR 和 HR 图像,故而不适用于常见的超分辨率算法。

目前已有一些针对医学图像的超分辨率算法,如 Chen 等<sup>[11]</sup>提出一种生成对抗网络指导训练的多级密集连接超分辨率网络 (mDCSRN),实现了磁共振 MRI 图像的超分辨。Park 等<sup>[12]</sup>使用改进的 U-Net 网络实现了二维人脑 CT 图像的超分辨。Mahapatra 等<sup>[13]</sup>提出一种基于生成对抗网络的眼底图像超分辨率算法。现有医学研究的影像模态主要为 CT 图像和 MRI 图像<sup>[14]</sup>,不同医疗领域的医学图像模态差异大,图像特点各不相同,不能直接互用。对

于内窥镜图像,基于深度学习的超分辨率算法<sup>[15-18]</sup>,虽然在一定程度上实现了内窥镜图像的超分辨,但是这些算法都是通过插值下采样的方法获得用于训练的 LR-HR 图像对。使用这样一种单一的、与现实不符的退化方式生成 LR 图像,一方面导致模型的泛化性差,另一方面与真实图像分辨分布差异大,无法重建出高频纹理细节,而这些丢失的细节内容对医学图像至关重要,可能影响医生对病情的判断。

为解决上述问题,本文提出一种针对真实内窥镜图像的超分辨率算法,该算法分为两个模块:退化模块和超分辨率模块。退化模块提出一种新颖的内窥镜模糊核和噪声的估计及注入算法,可直接从真实 LR 内窥镜图像提取丰富多样的模糊核和噪声,然后通过随机注入和多噪声拼贴的方式作用于真实 HR 内窥镜图像,以生成与真实 LR 内窥镜图像分布相近的训练数据。超分辨率模块则提出一种高频引导的残差密集网络,使用基于高低频信息交互的频率分离策略,提取多级高频特征;再采用多层级融合策略,将多级高频特征逐层融入残差密集模块的分层特征,以引导其恢复出良好的纹理细节。此外网络还采用了残差嵌套结构,进一步加深了网络深度,提升了网络性能。本文的贡献总结如下:

1) 本文提出一种针对真实内窥镜图像的退化估计及注入算法,基于真实低分辨率图像提取内窥镜相关的模糊核与噪声模式,用于合成与真实域更接近的低分辨率内窥镜图像。

2) 本文提出一种新颖的高频信息引导的超分辨率网络结构,采用基于双频率交互的频率分离策略,以及多层级融合机制,提取的多级高频信息引导构建残差密集块的多层特征,重建出细节保留较好的高分辨率内窥镜图像。

3) 一系列的对比和消融实验证明了本文提出的退化框架以及超分网络的有效性,相比其他方法具有更好的视觉效果和更高的客观质量评价。

## 1 相关工作

### 1.1 超分辨率算法的退化估计

许多基于卷积神经网络的超分辨率算法取得了优于传统方法的性能和效果。这些算法<sup>[8, 19-25]</sup>通常使用双三次插值下采样的退化方法构建训练所需的 {LR, HR} 图像对。经过双三次下采样得到的图像会丢失高频细节,同时低频内容更清晰<sup>[26]</sup>,而真实图像存在多种模糊和噪声<sup>[27]</sup>,故这类合成图像与真实图像存在域差异,导致超分辨率算法在合成图像上取得很好的结果,但在真实测试图像上效果很差。

为解决该问题,一些研究者通过向 HR 图像注入预定义的模糊核与噪声,以生成具有复杂退化的 LR 图像,再用此图像对训练网络.如 Zhang 等<sup>[28]</sup>采用各向同性和各向异性的高斯模糊核,噪声为水平在  $[0, 75]$  之间的高斯白噪声合成 LR 图像. Zhang 等<sup>[29]</sup>采用高斯模糊核、运动模糊核、散焦模糊核、高斯白噪声合成 LR 图像. BSRGAN<sup>[30]</sup>将模糊、下采样、噪声这三种退化赋予多种类型,如模糊分为各向同性、各向异性高斯模糊核,下采样分为最近邻、双线性、双三次插值,噪声分为不同水平的高斯噪声、不同质量因子的 JPEG 压缩、相机传感器噪声,再将以上各种类型的退化随机组合作为 HR 图像的退化.

另外一些研究者则提出了自适应学习退化的方法,旨在生成与真实图像同域的 LR 图像. Fritsche 等<sup>[31]</sup>提出一个生成对抗网络 DSGAN 模拟退化过程,使生成的 LR 图像保留双三次下采样 LR 图像的低频信息,并拥有真实图像的高频信息. Lugmayr 等<sup>[32]</sup>通过生成对抗网络和循环一致性损失学习双三次下采样 LR 图像与真实 LR 图像之间的分布映射,以克服它们之间的域差异. Chen 等<sup>[33]</sup>将退化过程建模为两个步骤,HR 图像到合成 LR 图像的退化,以及合成 LR 图像到真实 LR 图像的域映射,之后使用生成对抗网络学习此退化过程.

本文研究的是内窥镜图像的超分辨,内窥镜采用小型图像传感器实现成像,图像通常采用高速串行信号或模拟信号传输,在传输过程中容易受到干扰,常导致图像出现条纹噪声<sup>[10]</sup>,故而使用人工构造的不同模糊程度的模糊核以及不同级别的加性随机噪声,不符合真实内窥镜图像的退化过程.而自适应学习退化的方法模型复杂,且大多基于稳定性低的生成对抗网络,难以应用到实际医学场景.本文设计了一种新颖的退化估计算法,从真实 LR 内窥镜图像估计其专有的模糊核与噪声模式.之后在数据准备阶段又设计了一种退化注入算法,将估计的退化注入高质量的内窥镜图像,以生成与真实内窥镜图像同域的 LR 图像.

## 1.2 基于频率的深度超分辨方法

近年来,已经提出了一些基于频率和深度学习的图像超分辨率算法,大致可以分为两类.一类将图像由空域转换至频域,进而获取不同频率信息<sup>[34-38]</sup>.如 Xie 等<sup>[34]</sup>通过离散余弦变换 (Discrete cosine transform, DCT) 将图像转至频域,并将特征分解为高、中、低频信息,分别应用不同复杂度的卷积层,再融合结果,获得最终的超分辨率结果. Yun 等<sup>[35]</sup>通过 DCT 变换将图像转换至频域,使用超参数提

取高频分量,再转换至空间域并应用通道注意力机制获得重建结果.在超分辨率领域,因傅立叶变换涉及复数,故大多数算法使用 DCT 进行频域转换. DCT 变换会获得多个频率子带,中间部分的子带隶属于低频抑或是高频,这一点难以界定,此外,图像在空域和频域之间的频繁转换,增加了网络的复杂性.另一类是在空域里直接分离频率信息<sup>[31, 39-41]</sup>. Fritsche 等<sup>[31]</sup>使用线性滤波器分离低频和高频信息,并制作了相应的低频、高频损失用于超分辨网络. Pang 等<sup>[39]</sup>把 LR 图像逐步输入到三个卷积层,认为每一层的特征图代表不同的频率信息,使用注意力机制处理并融合这些信息,恢复出高分辨率图像. Liu 等<sup>[40]</sup>使用卷积均值滤波器进行频率分离,用不同的模块对分离的频率进行处理,再通过加法运算组合,最终使用上采样块重构整幅图像.然而,这些方法完全孤立了低频与高频信息,没有考虑低频信息与高频信息的相互作用,而频率间交互是十分重要的<sup>[41]</sup>.本文在空间域分离高低频信息,但并非一次性分离低频与高频信息并分而治之.而是低频与高频信息多次交互通信,分离出多级高频信息.再使用多层级融合策略,使多级高频信息逐层融入残差密集块,充分引导其构建包含高频信息的特征.

## 2 基于真实退化估计与高频引导的内窥镜图像超分辨网络模型

本文模型分为两个部分:退化框架和超分模型.退化框架从真实 LR 内窥镜图像估计退化,再将退化注入真实 HR 内窥镜图像,以生成逼真的、与之配对的低分辨率内窥镜图像,退化框架示意图如图 1 所示.超分模型则利用退化模型构造的成对数据,通过本文提出的高频引导的残差密集超分辨网络,获得细节保留完好、内容清晰的超分辨率内窥镜图像,超分模型网络结构示意图如图 2 所示.

### 2.1 退化框架

#### 2.1.1 退化估计

本文从真实 LR 内窥镜图像直接估计模糊核和噪声,以确保医学图像的真实性和准确性.

**模糊核估计.**本文使用 KernelGAN<sup>[42]</sup>作为模糊核估计算法,KernelGAN 的生成器是深度线性网络,不包含任何非线性的激活函数,所以训练后 KernelGAN 的网络权重参数能够以矩阵的形式输出,作为一个固定的模糊核.本文利用 KernelGAN 这一原理,以单张真实低分辨率内窥镜图像作为输入,训练本张图像对应的 KernelGAN,训练结束后生成器的网络模型参数即为该张图像的模糊核.我



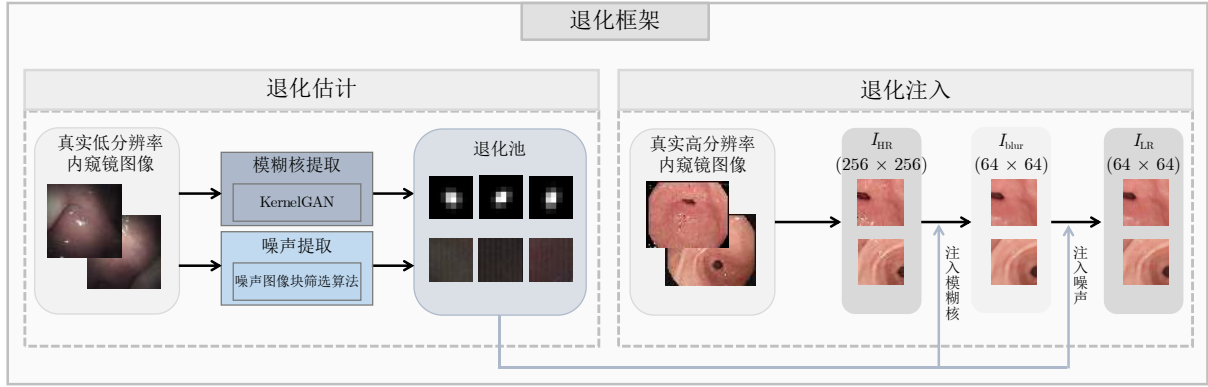


图 1 退化框架示意图

Fig.1 Overview of the degradation framework

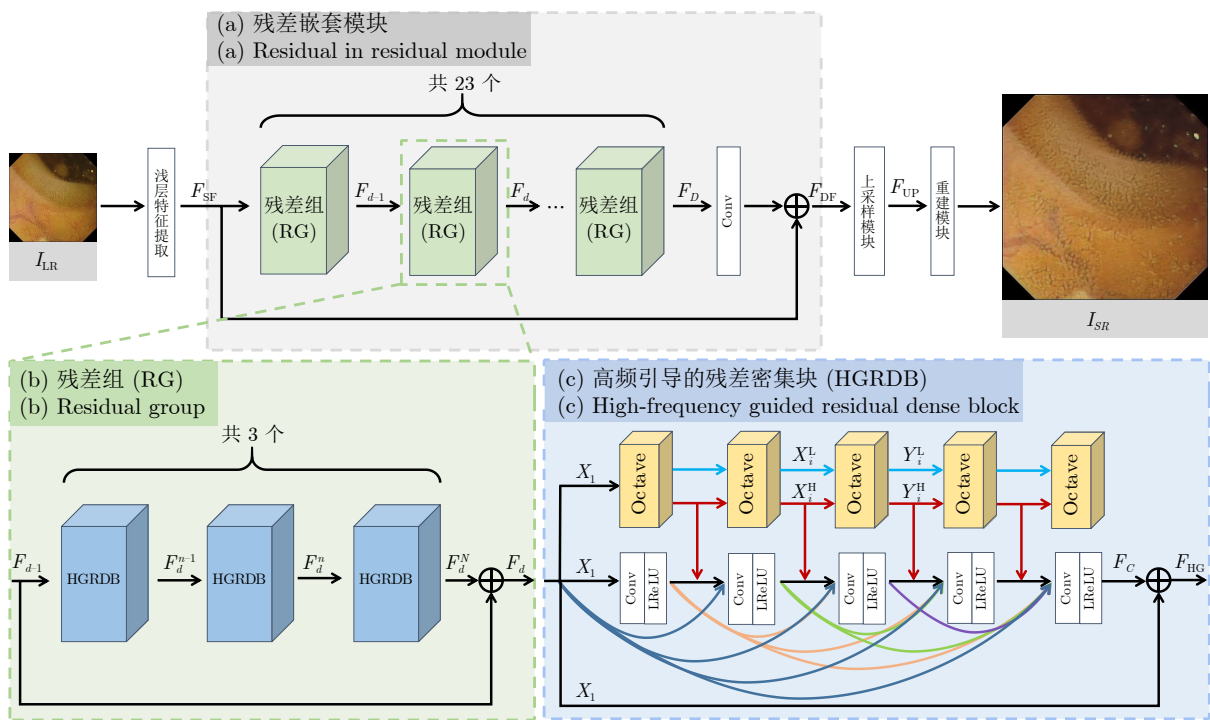


图 2 HGRDN 示意图

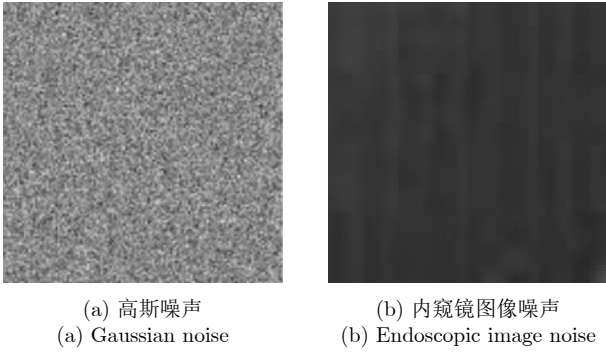
Fig.2 Overview of the HGRDN

们将所有真实低分辨率内窥镜图像单独对应的模糊核放入一个称作模糊核退化池的容器, 以供后续步骤使用。

**噪声估计.** 文献 [26, 43] 表明从原始噪声图像中筛选背景内容少、内部像素相似的区域可作为噪声图像块. 受其启发, 我们设计了一种专门针对内窥镜图像的噪声筛选算法, 用以直接从真实 LR 内窥镜图像提取医学内窥镜图像的噪声模式. 根据研究, 图像的方差可代表图像的对比度, 均值可代表图像的亮度[44]. 由于噪声图像块内部像素相似, 说明其内部像素对比度不高, 所以本文设定最大对

比度, 也即最大方差 ( $maxVar$ ), 来筛选对比度在一定范围的图像块. 由于噪声主要存在于图像低亮度区域[45], 所以本文设定最大亮度, 也即最大均值 ( $maxMean$ ), 来筛选亮度在一定范围的图像块. 因真实 LR 内窥镜图像的分辨率一般较小, 故而提取的图像块尺寸不能过大, 否则容易引入图像结构信息. 具体的筛选算法如算法 1 所示, 其中  $maxVar$  和  $maxMean$  均为经验设定. 最终提取的噪声图像块如图 3(b) 所示, 可直观看出提取的噪声不同于一般的加性高斯噪声, 并且与内窥镜常见的条纹噪声[10]具有一致的分布特性.





(a) 高斯噪声  
(a) Gaussian noise  
(b) 内窥镜图像噪声  
(b) Endoscopic image noise

图 3 常见加性高斯噪声与内窥镜噪声对比

Fig.3 Comparison of common additive Gaussian noise with endoscopic noise

### 算法 1. 真实 LR 内窥镜图像提取噪声图像块

**输入.** 真实低分辨率内窥镜图像  $I_{src}$ , 图像块  $patch$ , 最大方差  $maxVar$ , 最大均值  $maxMean$

**输出.** 所有满足要求的噪声图像块  $noisePatch$

**初始化:**  $patch = 32 \times 32$ ,  $maxVar = 20$ ,  $maxMean = 60$

for all  $patch$  in  $I_{src}$  do

    计算  $patch$  的方差  $patchVar$

    计算  $patch$  的均值  $patchMean$

    if  $patchVar < maxVar$  and  $patchMean < maxMean$

        该  $patch$  视作  $noisePatch$

$noisePatch$  添加进噪声退化池

    end

end

return 噪声退化池

#### 2.1.2 退化注入

退化注入前, HR 内窥镜图像会进行数据处理, 具体操作过程见第 3.1 节训练集部分, 最终获得尺寸为  $256 \times 256$  的图像块. 设置  $I_{HR}$  代表 HR 内窥镜图像的训练图像块,  $k$  代表模糊核,  $n$  代表噪声,  $\downarrow_s$  代表尺度系数为  $s$  的下采样,  $\otimes$  代表二维卷积, 其退化过程可用如下公式表示<sup>[26, 46]</sup>,  $I_{LR}$  为经退化所获得的 LR 内窥镜图像的图像块.  $\{I_{HR}, I_{LR}\}$  为配对图像.

$$I_{LR} = (I_{HR} \otimes k) \downarrow_s + n \quad (1)$$

**注入模糊核.** 从模糊核退化池随机选取一个模糊核, 由于该模糊核为一个给定数值的矩阵, 可看作是一种卷积核, 依照式 (1) 将其与  $I_{HR}$  进行卷积, 之后进行下采样操作, 取尺度系数为 4, 则获得模糊且下采样四倍的模糊图像  $I_{blur}$ , 其尺寸为  $64 \times 64$ .

**注入噪声.** 模糊图像  $I_{blur}$  拆分为多个  $16 \times 16$  的图像块, 在退化池中随机选取多个噪声图像块, 裁剪尺寸至  $16 \times 16$  并进行归一化操作, 之后与模糊图

像  $I_{blur}$  的多个图像块分别相加, 获得最终的  $I_{LR}$ . 将退化池的噪声裁剪为  $16 \times 16$  后再进行注入, 旨在向模糊图像  $I_{blur}$  注入数量更多、种类更繁的噪声, 以充分还原真实图像的情况.

通过从退化池中随机选取真实图像的模糊核和噪声, 极大地提高了合成图像  $I_{LR}$  的多样性和真实性. 需要说明的是, Ji 等<sup>[26]</sup> 也提出了一种退化估计框架, 但本文与其存在较大不同: 1) 退化估计的来源不同. Ji 等<sup>[26]</sup> 从 HR 图像估计模糊核和噪声, 再将其注入 HR 下采样后的图像. 本文从真实 LR 图像估计模糊核与噪声, 再将其注入 HR 图像, 真实性和准确性更高. 2) 提取噪声的理论依据、具体算法, 以及噪声注入算法均不一致. Ji 等<sup>[26]</sup> 主要针对自然图像设计, 本文则专门针对内窥镜图像设计.

## 2.2 内窥镜图像超分辨率网络模型

### 2.2.1 网络结构

如图 2 所示, 本文对 ESRGAN<sup>[47]</sup> 进行改进, 针对医学图像需要保留高频细节纹理这一特点, 以及频率信息对空域信息有补充作用这一优势, 提出使用频率分离策略和多层级融合机制的高频引导残差密集网络 (High-frequency guided residual dense network, HGRDN), 该网络由四部分组成: 浅层特征提取、残差嵌套模块、上采样模块、重建模块. 假设  $I_{LR}$  和  $I_{SR}$  为 HGRDN 的输入和输出, 首先使用一个  $3 \times 3$  的卷积对  $I_{LR}$  进行浅层特征提取, 使用零填充策略保持特征图尺寸不变:

$$F_{SF} = H_{SF}(I_{LR}) \quad (2)$$

其中,  $H_{SF}(\cdot)$  表示卷积操作,  $F_{SF}$  表示浅层特征. 之后  $F_{SF}$  输入残差嵌套模块, 该模块由 23 个残差组 (Residual group, RG) 和长跳跃连接组成, 每个残差组包含 3 个高频引导的残差密集块 (High-frequency guided residual dense block, HGRDB) 和短跳跃连接:

$$F_{DF} = H_{RIR}(F_{SF}) \quad (3)$$

其中,  $H_{RIR}(\cdot)$  表示残差嵌套模块的函数,  $F_{DF}$  为经过该模块后得到的深层特征. 有关 HGRDB 的具体结构在第 2.2.3 节详细阐述. 之后  $F_{DF}$  输入上采样模块, 该模块由两个尺度因子为  $\times 2$  的最近邻插值及  $3 \times 3$  卷积构成:

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

其中,  $H_{UP}(\cdot)$  表示上采样模块的复合函数,  $F_{UP}$  为上采样后的特征. 重建模块由一个卷积核为  $3 \times 3$  的卷积组成:

$$I_{SR} = H_{REC}(F_{UP}) \quad (5)$$

其中,  $H_{\text{REC}}(\cdot)$  表示重建模块的复合函数,  $I_{\text{SR}}$  为重建后的超分辨率图像.

### 2.2.2 残差嵌套模块

文献 [8, 48–49] 观察到更多的层和连接可以提高网络的性能, 而残差嵌套结构则使得训练层次更深的卷积神经网络成为可能. 文献 [49] 也提到残差组和长跳跃连接使得网络能够关注到 LR 图像的更多信息.

基于上述残差嵌套的优点, 本文采用 RCAN<sup>[49]</sup> 中关于残差嵌套部分的设计, 图 2(a) 展示了本文的残差嵌套模块, 其采用残差组 RG 作为基本模块, 第  $d$  个 RG 可公式化为:

$$F_d = H_{\text{RG},d}(F_{d-1}) = H_{\text{RG},d}(H_{\text{RG},d-1}(\cdots(H_{\text{RG},1}(F_{\text{SF}}))\cdots)) \quad (6)$$

其中,  $H_{\text{RG},d}$  代表第  $d$  个 RG 的函数,  $F_{d-1}$  和  $F_d$  代表第  $d$  个 RG 的输入和输出,  $F_{\text{SF}}$  表示浅层特征. 与文献 [49] 一样, 本文在多个残差组之后使用长跳跃连接和残差学习, 以稳定网络的训练:

$$F_{\text{DF}} = F_{\text{SF}} + WF_D = F_{\text{SF}} + WH_{\text{RG},d}(H_{\text{RG},d-1}(\cdots(H_{\text{RG},1}(F_{\text{SF}}))\cdots)) \quad (7)$$

其中,  $F_{\text{SF}}$  代表浅层特征,  $W$  代表残差嵌套模块末尾卷积层的权重, 此处省略了偏置项,  $F_D$  代表最后一个 RG 的输出特征,  $F_{\text{DF}}$  为经过残差嵌套模块后得到的深层特征.

如第 2.2.1 节所述, 每个残差组 RG 内部包含 3 个高频引导的残差密集块 (HGRDB) 和短跳跃连接, 如图 2(b) 所示, 第  $d$  个 RG 的第  $n$  个 HGRDB 可以公式化为:

$$F_d^n = H_{\text{RG},d}^n(F_d^{n-1}) = H_{\text{RG},d}^n(H_{\text{RG},d}^{n-1}(\cdots(H_{\text{RG},d}^1(F_{d-1}))\cdots)) \quad (8)$$

其中,  $F_d^{n-1}$  和  $F_d^n$  是第  $d$  个 RG 中第  $n$  个 HGRDB 的输入和输出,  $H_{\text{RG},d}^n$  是第  $d$  个 RG 中第  $n$  个 HGRDB 的函数,  $F_{d-1}$  是第  $d$  个 RG 输入的特征图. 加入短跳跃连接后, 最终残差组 RG 的输出可公式化为:

$$F_d = F_{d-1} + F_d^N = F_{d-1} + H_{\text{RG},d}^N(H_{\text{RG},d}^{N-1}(\cdots(H_{\text{RG},d}^1(F_{d-1}))\cdots)) \quad (9)$$

其中,  $F_{d-1}$  和  $F_d$  代表第  $d$  个 RG 的输入和输出特征,  $F_d^N$  代表第  $d$  个 RG 的最后一个 HGRDB 的输出特征.

### 2.2.3 高频引导的残差密集块 HGRDB

图 2(c) 展示了 HGRDB 的具体结构, 其由高频引

导层、残差密集块组成. 值得一提的是, 文献 [47–49] 也提出了残差密集块, 然而这些方法没有考虑图像频率信息的差异性, 对图像的低频与高频信息平等地对待, 没有充分利用 LR 图像的高频信息重建图像. 为克服以上缺点, 我们采用基于双频率信息交互的频率分离策略, 提取多级高频信息, 再使用多层级融合机制, 使高频信息与残差密集块提取的分层特征相结合, 以增强重建图像的细节恢复能力. 本文基本块能关注到更多的高频信息, 如细节、边缘、纹理等, 而这恰恰是医学图像不可或缺的重要信息.

**高频引导层.** 一幅图像由高频和低频信息组成, 其中高频部分代表局部的细节, 低频部分则代表全局的内容且一般存在大量冗余. 类似的, 卷积层输出的特征图也可以分为低频和高频. 根据这一现象, Chen 等提出一种新颖的 Octave 卷积<sup>[41]</sup>, 其能够分离频率, 获取不同频率的信息, 并且在 Octave 卷积过程中, 低频与高频按一定规则进行信息交流与更新. 我们采用这一思想, 设计了一种基于双频率信息交互的频率分离策略. 该策略级联多个 Octave 卷积, 以此进行多轮高低频信息交互和更新. 相较于一次性彻底分离高低频的方法<sup>[31, 39–40]</sup>, 这种双频率因素交互式作用更新的方法, 可以实现较全面的信息补偿, 以便充分挖掘频率信息. 每轮 Octave 卷积输出的高频信息被嵌入下方密集块的对应层级, 以引导其构建保留高频信息的特征. 我们将上述频率分离策略的具体实现结构定义为高频引导层, 如图 2(c) 所示.

在本文中, 高频引导层共级联 5 个 Octave 卷积. 假设第  $i$  个 Octave 卷积的输入特征图为  $X_i \in \mathbf{R}^{c_{\text{in}} \times h \times w}$ , 输出特征图为  $Y_i \in \mathbf{R}^{c_{\text{out}} \times h \times w}$ .  $c_{\text{in}}$  和  $c_{\text{out}}$  表示输入特征图和输出特征图的通道数,  $h$  表示特征图的高,  $w$  表示特征图的宽. 依前文所述, 特征图可看作由低频分量和高频分量组成, 则  $X_i$  可表示为  $\{X_i^{\text{H}}, X_i^{\text{L}}\}$ ,  $Y_i$  可表示为  $\{Y_i^{\text{H}}, Y_i^{\text{L}}\}$ , 其中 H 代表高频分量, L 代表低频分量. 按照 Octave 卷积<sup>[41]</sup> 的定义, 输入特征图的低频分量会进行 2 倍空间下采样以压缩信息, 故而输入输出特征图的低频和低频分量可以表示如下:

$$\begin{cases} X_i^{\text{H}} \in \mathbf{R}^{(1-\alpha_{\text{in}})c_{\text{in}} \times h \times w}, & \text{输入特征图的高频分量} \\ X_i^{\text{L}} \in \mathbf{R}^{\alpha_{\text{in}}c_{\text{in}} \times \frac{h}{2} \times \frac{w}{2}}, & \text{输入特征图的低频分量} \\ Y_i^{\text{H}} \in \mathbf{R}^{(1-\alpha_{\text{out}})c_{\text{out}} \times h \times w}, & \text{输出特征图的高频分量} \\ Y_i^{\text{L}} \in \mathbf{R}^{\alpha_{\text{out}}c_{\text{out}} \times \frac{h}{2} \times \frac{w}{2}}, & \text{输出特征图的低频分量} \end{cases}$$

其中,  $\alpha_{\text{in}}$  和  $\alpha_{\text{out}}$  表示输入和输出特征图中低频部分所占的比例, 是范围为  $[0, 1]$  的超参数.

单个 Octave 卷积的过程如图 4 所示. 其输入是

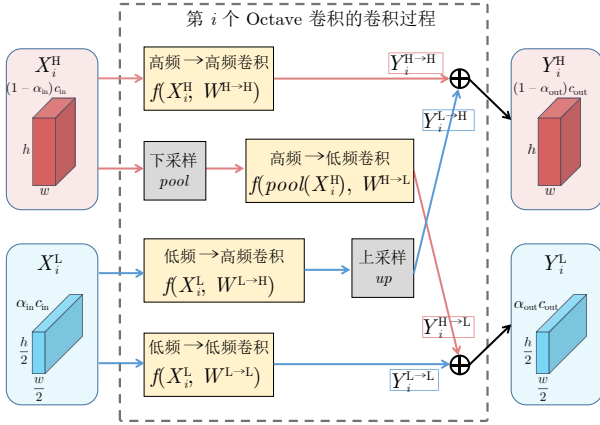


图 4 Octave 卷积之高低频率信息的交互

Fig. 4 The high-low frequency information interaction of Octave convolution

前段所述的  $X_i^H$  和  $X_i^L$ . 之后同频率的特征进行信息更新, 不同频率的特征进行信息交流. 最终输出更新后的高频信息  $Y_i^H$  和低频信息  $Y_i^L$ , 其中  $Y_i^H$  被嵌入进下方残差密集块的对应层级. 整个过程可用公式表达如下:

$$Y_i^H = Y_i^{H \to H} + Y_i^{L \to H} = f(X_i^H, W^{H \to H}) + up(f(X_i^L, W^{L \to H})) \quad (10)$$

$$Y_i^L = Y_i^{L \to L} + Y_i^{H \to L} = f(X_i^L, W^{L \to L}) + f(pool(X_i^H), W^{H \to L}) \quad (11)$$

其中,  $Y_i^{H \to H}$  代表高频之间的信息更新,  $W^{H \to H}$  为该操作使用的卷积核.  $Y_i^{L \to L}$  代表低频之间的信息更新,  $W^{L \to L}$  为该操作使用的卷积核.  $Y_i^{H \to L}$  和  $Y_i^{L \to H}$  代表不同频率间的信息交流,  $W^{H \to L}$  和  $W^{L \to H}$  为该操作使用的卷积核.  $f(X, W)$  表示输入为  $X$ , 卷积核为  $W$  的卷积.  $pool(\cdot)$  表示下采样操作, 此处采用全局平均池化方法.  $up(\cdot)$  表示上采样操作, 此处采用最近邻上采样方法. 上述提及的卷积核具体参数如下所示:

$$\begin{cases} W^{H \to H} \in \mathbf{R}^{(1-\alpha_{in})c_{in} \times (1-\alpha_{out})c_{out} \times k \times k} \\ W^{L \to H} \in \mathbf{R}^{\alpha_{in}c_{in} \times (1-\alpha_{out})c_{out} \times k \times k} \\ W^{H \to L} \in \mathbf{R}^{(1-\alpha_{in})c_{in} \times \alpha_{out}c_{out} \times k \times k} \\ W^{L \to L} \in \mathbf{R}^{\alpha_{in}c_{in} \times \alpha_{out}c_{out} \times k \times k} \end{cases}$$

值得注意的是, 按照文献 [41] 的定义, 第一个 Octave 卷积的输入一般为普通卷积的特征图, 为将其转换为多频特征表示, 设置该特征图只包含高频分量, 不包含低频分量, 即第一个 Octave 卷积层的  $\alpha_{in}$  设置为 0, 则第一个 Octave 卷积层的输入特征图  $X_1$  表示为  $\{X_1^H\}$ ,  $X_1^H \in \mathbf{R}^{c_{in} \times h \times w}$ .

**残差密集块.** 受文献 [47-48] 启发, 残差密集块

能够充分利用所有卷积层的分层特征, 并且提高网络的容量, 故本文采用残差密集块来提取 LR 图像的丰富局部特征, 其具体结构如图 2(c) 下半部分所示, 总共包含五个卷积层和一个短跳跃连接. 在本模块, 提出了一种多层级融合机制, 即从第二个卷积层开始, 每个卷积层的输入特征不仅包含密集块的输入特征、前序卷积层的输出特征, 还包含高频引导层的高频信息, 这样每层卷积的特征会被融入对应层级的高频信息, 由此起到高频引导的作用. 每个卷积层的输出特征图可以用以下公式表示:

$$F_c = \begin{cases} \sigma(W_1(X_1)) = F_1, & c = 1 \\ \sigma(W_c[X_1, F_1, \dots, F_{c-1}, Y_i^H]), & c = 2, 3, 4, 5 \end{cases} \quad (12)$$

其中,  $F_c$  表示第  $c$  个卷积层的输出特征图,  $\sigma$  表示 LReLU 激活函数,  $W_c$  表示第  $c$  个卷积的权重, 此处省略了偏置项.  $X_1$  表示残差密集块的输入特征图,  $[\cdot]$  表示特征图的连接,  $Y_i^H$  表示高频引导层的第  $i$  个 Octave 卷积提取的高频分量,  $i = 1, 2, 3, 4$ . 加入残差后, 最终 HGRDB 的输出为:

$$F_{HG} = X_1 + F_C \quad (13)$$

其中,  $F_{HG}$  表示此 HGRDB 模块的输出特征,  $X_1$  表示此 HGRDB 的输入特征,  $F_C$  表示经过残差密集块最后一个卷积层的输出特征图.

### 2.3 训练细节

为充分验证本文退化框架及超分辨率网络的有效性, 我们借鉴 ESRGAN<sup>[47]</sup> 和 BSRGAN<sup>[30]</sup> 的设置, 首先训练面向峰值信噪比 (Peak signal noise ratio, PSNR) 的网络模型 HGRDNet, 然后训练面向视觉质量的模型 HGRDGAN, 这两种模型均使用 HGRDN 的网络结构, 不同点仅在于损失函数. HGRDNet 使用  $L_1$  损失, HGRDGAN 使用  $L_1$  损失、感知损失<sup>[47]</sup>、对抗损失<sup>[47]</sup>, 其权重分别为 0.01, 1, 0.005. 面向 PSNR 的 HGRDNet 模型能在数值上客观体现出模型的优劣, 但是其往往由于像素平均问题而产生过度平滑的视觉结果<sup>[7]</sup>, 面向视觉质量的模型 HGRDGAN 使用了感知损失和对抗损失, 这种感知驱动的方式会在视觉上达到最优效果, 但无法体现在 PSNR 这些传统指标上. 因此在后续实验中, 我们主要使用 HGRDNet 模型进行定量实验以验证本文网络模型在 PSNR 等传统指标上的数值结果, 使用 HGRDGAN 模型进行定性实验以验证本文模型的视觉效果.

训练设置 *batchsize* 为 16 张图像. 初始的学习率为  $1 \times 10^{-4}$ , 且在第  $[50k, 100k, 200k, 300k]$  迭代周期时学习率减半,  $k = 1000$ . 对于优化, 使用 Adam



优化器<sup>[50]</sup>, 参数设为  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , 交替更新生成器和鉴别器网络, 直到模型收敛. 本文实验采用 PyTorch 框架, 使用单张 NVIDIA TITAN RTX 的 GPU 卡训练网络.

### 3 实验

#### 3.1 训练集

本文退化模型的输入是非配对的真实 HR 内窥镜图像和真实 LR 内窥镜图像. 其中真实 HR 内窥镜图像采用胃肠道内窥镜图像数据集 Kvasir v2<sup>[51]</sup> 食管炎类别和正常幽门类别的图像, 总共 2000 张, 分辨率为  $1280 \times 1024$ . 真实 LR 内窥镜图像由合作医院提供, 总共 296 张, 分辨率为  $400 \times 400$ .

真实 HR 内窥镜图像进行数据处理, 首先中心裁剪尺寸至  $1000 \times 1000$ , 之后进行 clean-up 操作<sup>[26]</sup>, 以产生近似无噪声图像, 然后裁剪至  $256 \times 256$ , 作为真实 HR 内窥镜图像的训练图像块  $I_{HR}$ . 真实 LR 内窥镜图像按照退化模型的退化估计模块, 进行模糊核和噪声估计, 退化估计完成后,  $I_{HR}$  按照退化模型的退化注入模块获得  $I_{LR}$ , 最终  $\{I_{LR}, I_{HR}\}$  配对图像为超分模型的训练集, 共 2000 对.

#### 3.2 测试集

因为在内窥镜图像领域, 目前尚不存在现成的真实低分辨率内窥镜图像及其对应的 GT (Ground truth, GT) 图像这种配对数据集, 为客观公正, 本文构造了两个定量数据集, 构造过程遵循本文第 2 节的退化估计及注入步骤. 首先从合作医院获取 15 张未在训练集里出现的、由同一个内窥镜设备采集的真实 LR 内窥镜图像, 从中提取多个模糊核和噪声图像块.

**定量测试集 A.** 从数据集 Kvasir v2<sup>[51]</sup> 正常 Z 线类别获取图像 100 张, 中心裁剪尺寸至  $800 \times 800$ , 作为 GT. 上述提取的多个模糊核和噪声按照第 2 节的退化注入模块与 GT 图像作用, 生成对应的 LR 图像, 尺寸为  $200 \times 200$ .

**定量测试集 B.** 为进一步验证本文方法的效果, 从合作医院获取真实 HR 内窥镜图像 100 张, 中心裁剪尺寸至  $800 \times 800$ , 作为 GT. 上述提取的多个模糊核和噪声按照第 2 节的退化注入模块与 GT 图像作用, 生成对应的 LR 图像, 尺寸为  $200 \times 200$ .

为验证本文方法的鲁棒性, 构建了两个只进行简单下采样而非用本文退化框架的测试集, 分别为定量测试集 C 和定量测试集 D.

**定量测试集 C.** 从数据集 Kvasir v2<sup>[51]</sup> 正常 Z 线类别获取图像 100 张, 中心裁剪尺寸至  $800 \times$

$800$ , 作为 GT. 对 GT 做双三次下采样, 生成对应的 LR 图像, 尺寸为  $200 \times 200$ .

**定量测试集 D.** 从合作医院获取真实 HR 内窥镜图像 100 张, 中心裁剪尺寸至  $800 \times 800$ , 作为 GT. 对 GT 做双三次下采样, 生成对应的 LR 图像, 尺寸为  $200 \times 200$ .

**定性测试集.** 从数据集 Kvasir-Capsule<sup>[52]</sup> 选取 100 张图像作为定性数据集, 图像尺寸为  $336 \times 336$ . 该数据集通过胶囊内窥镜获取, 采集的部位集中于胃肠道区域, 采集的图像为真实内窥镜 LR 图像, 无对应的 GT 图像.

#### 3.3 对比实验

本文提出的方法与多个先进的图像超分辨算法进行比较, 其中包含六种针对自然图像的超分辨方法: 基于实用退化模型的盲图像超分辨率网络 (Designing a practical degradation model for deep blind image super-resolution, BSRGAN)<sup>[30]</sup>, 基于真实退化框架的超分辨率网络 (Realistic degradation framework for super-resolution, RealSR)<sup>[26]</sup>, 基于纯合成数据的真实盲超分辨率网络 (Real-world blind super-resolution with pure synthetic data, RealESRGAN)<sup>[53]</sup>, 频率分离的真实超分辨率网络 (Frequency separation for real-world super-resolution, FSSR)<sup>[31]</sup>, 深度展开的图像超分辨率网络 (Deep unfolding network for image super-resolution, USRNet)<sup>[54]</sup>, 学习退化分布的盲图像超分辨率网络 (Learning the degradation distribution for blind image super-resolution, PDMSR)<sup>[55]</sup> 以及一种针对医学图像的超分辨方法: 反馈自适应加权密集超分辨率网络 (Feedback adaptive weighted dense network, FAWDN)<sup>[56]</sup>. 其中, 对于退化模型, BSRGAN<sup>[30]</sup>, RealSR<sup>[26]</sup>, RealESRGAN<sup>[53]</sup>, USRNet<sup>[54]</sup>, FAWDN<sup>[56]</sup> 采用的是人为设定的特定退化模型, FSSR<sup>[31]</sup>, PDMSR<sup>[55]</sup> 采用的是不配对的真实 HR 和 LR 图像生成的退化模型. 对于超分辨模型, RealSR<sup>[26]</sup> 和 FSSR<sup>[31]</sup> 是面向视觉质量的模型, USRNet<sup>[54]</sup>, FAWDN<sup>[56]</sup>, PDMSR<sup>[55]</sup> 是面向 PSNR 的模型, 而 BSRGAN<sup>[30]</sup>, RealESRGAN<sup>[53]</sup> 与本文一样, 具有两种模型, 即面向 PSNR 的模型 BSRNet<sup>[30]</sup>, RealESRNet<sup>[53]</sup> 和面向视觉质量的模型 BSRGAN<sup>[30]</sup>, RealESRGAN<sup>[53]</sup>. 为保证公平, 我们使用其面向 PSNR 的模型进行定量实验, 使用其面向视觉质量的模型进行定性实验. 上述对比方法均使用本文训练集重新训练或对其预训练模型微调, 超分辨的比例因子为 4, 重建图像的客观质量对比结果如表 1 所示, 最优值用粗体表示, 次优值用下划线表示. 在

表 1 各方法在定量测试集的客观结果  
Table 1 The objective results of different methods in quantitative testsets

方法	定量测试集 A		定量测试集 B		定量测试集 C		定量测试集 D	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
PDMSR <sup>[55]</sup>	29.21	0.723	28.60	0.773	27.78	0.761	24.40	0.776
RealSR <sup>[26]</sup>	28.08	0.652	28.09	0.621	25.41	0.581	25.16	0.561
RealESRGAN <sup>[53]</sup>	31.08	0.790	30.01	0.863	<u>32.60</u>	0.801	32.17	0.879
USRNet <sup>[54]</sup>	30.17	0.787	28.50	0.864	31.32	0.801	29.91	0.882
FSSR <sup>[31]</sup>	26.46	0.670	28.31	0.663	25.93	0.612	24.32	0.574
FAWDN <sup>[56]</sup>	<u>31.62</u>	<u>0.792</u>	<u>32.96</u>	<u>0.894</u>	32.33	<u>0.802</u>	<u>33.58</u>	<u>0.905</u>
BSRGAN <sup>[30]</sup>	30.73	0.777	29.86	0.848	31.30	0.792	29.89	0.864
<b>HGRDN (Ours)</b>	<b>31.78</b>	<b>0.797</b>	<b>33.22</b>	<b>0.902</b>	<b>32.61</b>	<b>0.808</b>	<b>33.90</b>	<b>0.913</b>

表 1 中为方便起见, 统一使用 BSRGAN, RealESRGAN, HGRDN 的算法名称来代表各自对应的两种模型.

### 3.3.1 定量实验

本文使用 PSNR 和结构相似度 (Structural similarity, SSIM) 指标来客观衡量各方法在定量测试集 A、B、C、D 上的超分辨率结果, 结果如表 1 所示, 我们的模型 HGRDN 取得了最优的结果, 在四个数据集上 PSNR 和 SSIM 均取得了最高值. 由于本文算法着重恢复内窥镜图像的细节, 为验证高频细节的恢复效果, 本文引入了高频指标: 高频 PSNR 和高频 SSIM. 首先我们使用高通滤波器<sup>[57]</sup> 处理各方法的输出图像, 获得各方法的高频图. 再使用高通滤波器处理定量数据集 A、B 的 GT 图像, 获得对应的高频图. 最后计算各 GT 高频图与各方法的高频图之间的 PSNR 值、SSIM 值, 获得各方法的高频 PSNR 值和高频 SSIM 值, 结果如表 2 所示. 我们的模型 HGRDN 取得了最优的结果.

表 2 各方法在定量测试集的高频结果  
Table 2 The high-frequency results of different methods in quantitative testsets

方法	定量测试集 A		定量测试集 B	
	高频 PSNR ↑	高频 SSIM ↑	高频 PSNR ↑	高频 SSIM ↑
PDMSR <sup>[55]</sup>	26.95	0.573	26.07	0.573
RealSR <sup>[26]</sup>	27.52	0.523	27.13	0.513
RealESRGAN <sup>[53]</sup>	28.21	0.600	28.53	0.630
USRNet <sup>[54]</sup>	27.67	0.590	27.76	0.625
FSSR <sup>[31]</sup>	28.34	0.574	27.21	0.539
FAWDN <sup>[56]</sup>	<u>29.51</u>	<u>0.601</u>	<u>29.80</u>	<u>0.649</u>
BSRGAN <sup>[30]</sup>	27.13	0.543	27.65	0.580
<b>HGRDN (Ours)</b>	<b>29.79</b>	<b>0.603</b>	<b>30.26</b>	<b>0.664</b>

除数值结果外, 我们还展示了定量测试集上的视觉结果, 如图 5 所示, HGRDNet, BSRNet<sup>[30]</sup>, RealESRNet<sup>[53]</sup>, USRNet<sup>[54]</sup>, PDMSR<sup>[55]</sup>, FAWDN<sup>[56]</sup> 这些面向 PSNR 模型, 虽然客观评价指标较高, 但往往存在过度平滑现象, 这是面向 PSNR 模型本身

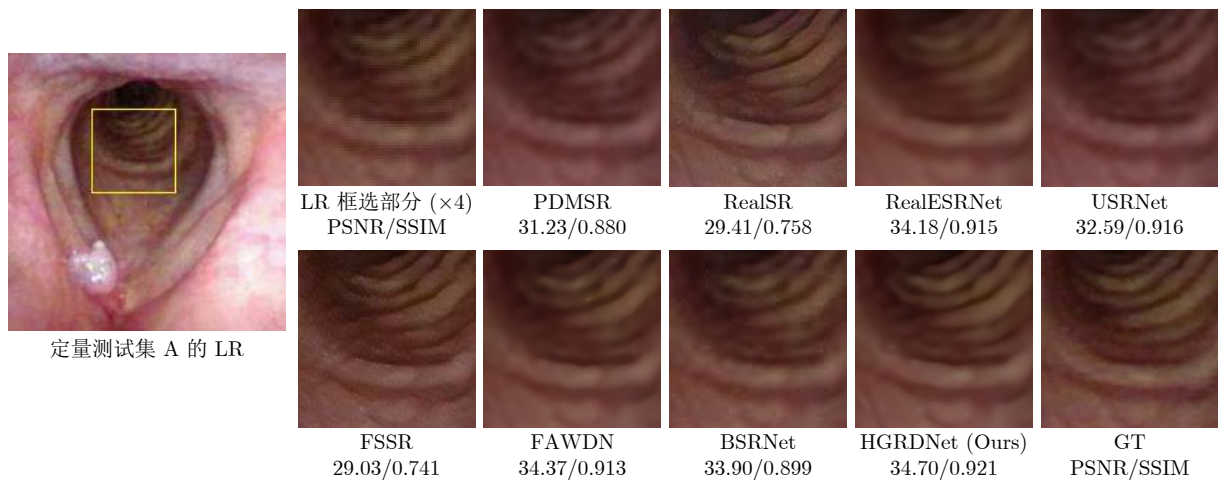


图 5 各方法在定量测试集上的视觉结果

Fig.5 The visual results of different methods in quantitative testsets

特性导致的. 而 RealSR<sup>[26]</sup> 和 FSSR<sup>[31]</sup> 这两个面向视觉质量的模型, 虽然具有较低的客观评价指标, 但是视觉效果却相对优秀. 这说明 PSNR 和 SSIM 指标虽可在一定程度上反映模型重建图像内容的能

力, 但却不符合人眼的视觉感知, 因此我们更关注符合实际应用的面向视觉质量模型 HGRDGAN.

### 3.3.2 定性实验

图 6 展示了本文模型 HGRDGAN 在定性数据

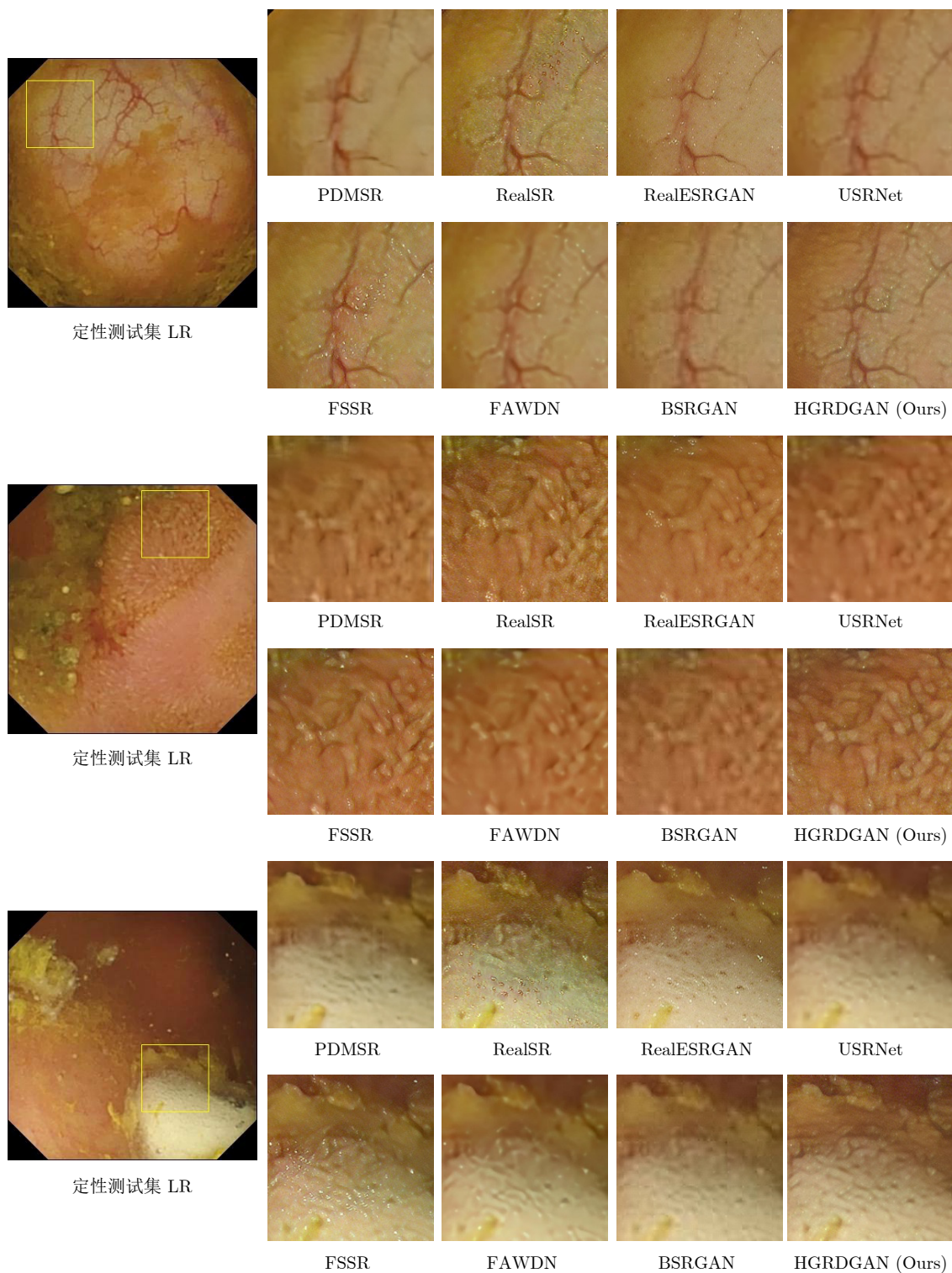


图 6 各方法在定性测试集上的视觉结果

Fig.6 The visual results of different methods in qualitative testsets



集上与其他方法的视觉对比结果,我们着重放大了重建图像的局部内容,从中可直观地看出本文模型在细节纹理的重建上优于其他方法,结果更清晰更富有层次性.从图中可以看出,所有对比方法都呈现出模糊不清晰的现象,其中 FSSR 方法因对噪声估计不准确出现了许多白点,RealSR 方法因对纹理细节关注不够出现了一些伪影.而本文模型纹理边缘清晰,无伪影和噪声,同时还充分恢复出了很多微小之处,如毛细血管以及各个组织表面的纹理,这是其他对比方法不曾关注的地方.这说明这些对比方法的退化模型和超分模型不适用于医学内窥镜图像.我们的退化模型是直接从真实内窥镜图像提取退化,所以真实性更高,更符合实际情况.我们的超分模型设计了高频引导层,所以增强了高频信息,充分挖掘了医学图像的纹理细节.与图 5 定量数据集的视觉效果对比,也可得出结论,使用面向视觉质量的感知损失和对抗损失令模型更加关注图像的视觉特征,得到的结果也比使用像素损失的结果更清晰.

由于定性数据集没有 GT,故我们采用无参考图像的质量评估指标 NIQE<sup>[58]</sup> 和 PI<sup>[59]</sup> 进行定量评估. NIQE 和 PI 的值均是越低代表感知质量越好,由表 3 所示本文模型在 NIQE 和 PI 的数值均为最低,即人眼主观感受最好.

表 3 不同方法在定性测试集的客观结果  
Table 3 The objective results of different methods in the qualitative testsets

方法	定性测试集	
	NIQE ↓	PI ↓
PDMSR <sup>[55]</sup>	7.96	6.73
RealSR <sup>[20]</sup>	<u>4.53</u>	<u>3.58</u>
RealESRGAN <sup>[53]</sup>	5.76	4.52
USRNet <sup>[54]</sup>	9.78	8.87
FSSR <sup>[81]</sup>	5.31	4.02
FAWDN <sup>[56]</sup>	9.54	8.29
BSRGAN <sup>[30]</sup>	6.74	5.57
<b>HGRDN (Ours)</b>	<b>4.40</b>	<b>3.20</b>

### 3.4 消融实验

为进一步验证模糊核与噪声的估计和注入在退化框架中的必要性,验证高频引导层在超分辨率模型中的有效性,我们进行了以下消融实验:

**去除噪声模块.**在退化框架中删除真实噪声的估计与注入模块,仅保留模糊核的估计与注入模块,以验证噪声模块的有效性.

**去除模糊模块.**在退化框架中删除真实模糊的估计与注入模块,以双三次插值下采样代替之,并保留真实噪声的估计与注入模块,以验证模糊核模块的有效性.

**去除高频模块.**设置超分模型高频引导的残差密集块不包含高频引导层,仅保留残差密集块.此时网络模型与基线网络 ESRGAN<sup>[47]</sup> 一致.

以上实验的训练设置与第 2.3 节一致.由表 4 可知完整模型在定量数据集 A 和 B 上的 PSNR 和 SSIM 数值仍为最优,这说明真实噪声与模糊核的退化估计与注入,以及高频引导模块均提升了模型的性能,缺一不可.图 7 展示了本文完整模型与各消融实验在定性数据集上的重建结果.如图 7(a) 所示,去除噪声模块的图像出现了许多伪影,而去除模糊模块和完整模型未曾出现这类现象,这表示缺乏真实噪声的注入,模型无法识别真实测试图像的噪声,进而重建结果出现偏差,这说明了真实噪声估计与注入的有效性.如图 7(b) 所示,去除模糊模块的图像非常模糊,这说明常见的双三次下采样方法不符合内窥镜图像的真实退化过程,故而在对真实内窥镜图像重建时无法恢复具体细节,说明了真实模糊估计的重要性.如图 7(c) 去除高频模块的图像纹理与细节边缘不如完整模型清晰明显,这说明高频引导模块的确加强了边缘、纹理等高频信息,使模型拥有更细腻的细节重建能力.

表 4 消融实验定量结果  
Table 4 The quantitative results of the ablation experiments

方法	定量测试集 A		定量测试集 B	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
去除噪声模块	30.50	0.761	32.03	0.857
去除模糊模块	30.97	0.789	32.54	0.892
去除高频模块	30.61	0.792	32.29	0.898
<b>完整模型</b>	<b>31.78</b>	<b>0.797</b>	<b>33.22</b>	<b>0.902</b>

## 4 结论

本文提出一种基于真实内窥镜图像退化估计的高频引导超分辨率网络.针对内窥镜图像的模糊核和噪声,设计了一种新颖的退化框架,旨在解决常规下采样方式生成的配对图像存在域差异的问题,获得符合真实域的退化图像.同时,在超分辨率阶段,本文开发了基于双频率信息交互的频率分离策略以及多层级融合机制,以重建完整清晰的重要纹理与细节.大量实验表明,与其他算法相比,本文提出的方法可以产生更高质量的重建效果.

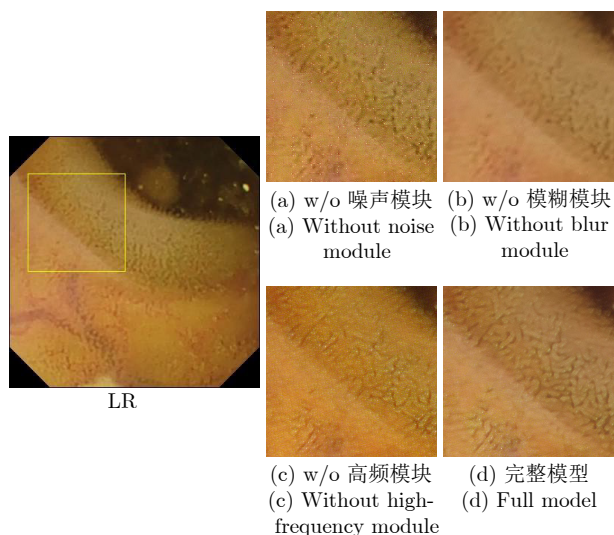


图 7 消融实验的视觉结果

Fig. 7 The visual results of the ablation experiments

## References

- Li B Z, Liao C R, Cai Z H, Zhou J, Zhao C, Jing L Q, et al. Femtosecond laser 3D printed micro objective lens for ultrathin fiber endoscope. *Fundamental Research*, DOI: [10.1016/j.fmre.2022.05.026](https://doi.org/10.1016/j.fmre.2022.05.026)
- Yang X R, Chen Y, Tao R, Zhang Y, Liu Z W, Shi Y G. Endoscopic image deblurring and super-resolution reconstruction based on deep learning. In: Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Engineering. Beijing, China: IEEE, 2020. 168–172
- Wang Zhi-Yin, Zhang Er-Hu, Shi Zheng-Hao, Duan Jing-Hong. Deep learning based backlight image enhancement method derived of zero-reference samples. *Journal of Image and Graphics*, 2022, **27**(5): 1589–1603  
(王知音, 张二虎, 石争浩, 段敬红. 零参考样本下的逆光图像深度学习增强方法. *中国图象图形学报*, 2022, **27**(5): 1589–1603)
- Yang Zhen, Di Shuan-Hu, Zhao Yu-Qian, Liao Miao, Zeng Ye-Zhan. Automatic liver tumor segmentation based on cascaded Dense-UNet and graph cuts. *Journal of Electronics & Information Technology*, 2022, **44**(5): 1683–1693  
(杨振, 邸拴虎, 赵于前, 廖苗, 曾业战. 基于级联 Dense-UNet 和图割的肝脏肿瘤自动分割. *电子与信息学报*, 2022, **44**(5): 1683–1693)
- Zhang Fang, Zhao Dong-Xu, Xiao Zhi-Tao, Geng Lei, Wu Jun, Liu Yan-Bei. Research progress of single image super-resolution reconstruction technology. *Acta Automatica Sinica*, 2022, **48**(11): 2634–2654  
(张芳, 赵东旭, 肖志涛, 耿磊, 吴骏, 刘彦北. 单幅图像超分辨率重建技术研究进展. *自动化学报*, 2022, **48**(11): 2634–2654)
- Dong C, Loy C C, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 295–307
- Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, 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. Honolulu, USA: IEEE, 2017. 105–114
- Lim B, Son S, Kim H, Nah S, Lee K M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE, 2017. 1132–1140
- Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 256–272
- Qu J L, Jin K, Wang M, Huang G M. Real-time stripe noise removal method for endoscope image. In: Proceedings of the IEEE 4th International Conference on Electronics Technology. Chengdu, China: IEEE, 2021. 865–870
- Chen Y H, Shi F, Christodoulou A G, Xie Y B, Zhou Z W, Li D B. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In: Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer, 2018. 91–99
- Park J, Hwang D, Kim K Y, Kang S K, Kim Y K, Lee J S. Computed tomography super-resolution using deep convolutional neural network. *Physics in Medicine & Biology*, 2018, **63**(14): Article No. 145011
- Mahapatra D, Bozorgtabar B. Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution. arXiv preprint arXiv: 1710.04783, 2017.
- Jiang Xi, Yuan Yi-Xuan, Wang Ya-Ping, Xiao Zhen-Xiang, Zhu Mei-Lu, Chen Ze-Hua, et al. A 20-year retrospect and prospect of medical imaging artificial intelligence in China. *Journal of Image and Graphics*, 2022, **27**(3): 655–671  
(蒋希, 袁奕萱, 王雅萍, 肖振祥, 朱美芦, 陈泽华, 等. 中国医学影像人工智能 20 年回顾和展望. *中国图象图形学报*, 2022, **27**(3): 655–671)
- Gu X G, Zhou F X, Chen R F, Ren X Z, Zhou W J. Endoscopic single image super-resolution based on transformer and convolutional neural network. In: Proceedings of the 7th International Conference on Life System Modeling and Simulation. Hangzhou, China: Springer, 2021. 24–32
- Song X W, Tang H, Yang C F, Zhou G Q, Wang Y G, Huang X J, et al. Deformable transformer for endoscopic video super-resolution. *Biomedical Signal Processing and Control*, 2022, **77**: Article No. 103827
- Turan M. A generative adversarial network based super-resolution approach for capsule endoscopy images. *Medicine Science*, 2021, **10**(3): 1002–1007
- Almalioglu Y, Ozyoruk K B, Gokce A, Incetan K, Gokceler G I, Simsek M A, et al. EndoL2H: Deep super-resolution for capsule endoscopy. *IEEE Transactions on Medical Imaging*, 2020, **39**(12): 4297–4309
- Dai T, Cai J R, Zhang Y B, Xia S T, Zhang L. Second-order attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 11057–11066
- Pan J S, Liu S F, Sun D Q, Zhang J W, Liu Y, Ren J, et al. Learning dual convolutional neural networks for low-level vision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3070–3079
- He X Y, Mo Z T, Wang P S, Liu Y, Yang M Y, Cheng J. ODE-inspired network design for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 1732–1741
- Hu X C, Mu H Y, Zhang X Y, Wang Z L, Tan T N, Sun J. Meta-SR: A magnification-arbitrary network for super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 1575–1584
- Li Z, Yang J L, Liu Z, Yang X M, Jeon G, Wu W. Feedback network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 3862–3871
- Qiu Y J, Wang R X, Tao D P, Cheng J. Embedded block residual network: A recursive restoration model for single-image su-

- per-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 4179–4188
- 25 Yin X, Tai Y, Huang Y G, Liu X M. FAN: Feature adaptation network for surveillance face recognition and normalization. In: Proceedings of the 15th Asian Conference on Computer Vision. Kyoto, Japan: Springer, 2020. 301–319
- 26 Ji X Z, Cao Y, Tai Y, Wang C J, Li J L, Huang F Y. Real-world super-resolution via kernel estimation and noise injection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE, 2020. 1914–1923
- 27 Wang W, Zhang H C, Yuan Z H, Wang C H. Unsupervised real-world super-resolution: A domain adaptation perspective. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 4298–4307
- 28 Zhang K, Zuo W M, Zhang L. Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3262–3271
- 29 Zhang K, Zuo W M, Zhang L. Deep plug-and-play super-resolution for arbitrary blur kernels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 1671–1681
- 30 Zhang K, Liang J Y, Van Gool L, Timofte R. Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 4771–4780
- 31 Fritsche M, Gu S H, Timofte R. Frequency separation for real-world super-resolution. In: Proceedings of IEEE/CVF International Conference on Computer Vision Workshop. Seoul, South Korea: IEEE, 2019. 3599–3608
- 32 Lugmayr A, Danelljan M, Timofte R. Unsupervised learning for real-world super-resolution. In: Proceedings of IEEE/CVF International Conference on Computer Vision Workshop. Seoul, South Korea: IEEE, 2019. 3408–3416
- 33 Chen S J, Han Z, Dai E Y, Jia X, Liu Z L, Liu X, et al. Unsupervised image super-resolution with an indirect supervised path. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE, 2020. 1924–1933
- 34 Xie W B, Song D H, Xu C, Xu C J, Zhang H, Wang Y H. Learning frequency-aware dynamic network for efficient super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 4288–4297
- 35 Yun J S, Yoo S B. Single image super-resolution with arbitrary magnification based on high-frequency attention network. *Mathematics*, 2022, **10**(2): Article No. 275
- 36 Xu R Y, Kang X J, Li C X, Chen H, Ming A L. DCT-FANet: DCT based frequency attention network for single image super-resolution. *Displays*, 2022, **74**: Article No. 102220
- 37 Guo T T, Mousavi H S, Monga V. Adaptive transform domain image super-resolution via orthogonally regularized deep networks. *IEEE Transactions on Image Processing*, 2019, **28**(9): 4685–4700
- 38 Hung K W, Wang K, Jiang J M. Image up-sampling using deep cascaded neural networks in dual domains for images down-sampled in DCT domain. *Journal of Visual Communication and Image Representation*, 2018, **56**: 144–149
- 39 Pang Y X, Li X, Jin X, Wu Y J, Liu J Z, Liu S, et al. FAN: Frequency aggregation network for real image super-resolution. In: Proceedings of the European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 468–483
- 40 Liu Z B, Yuan L, Sun L. Frequency separation-based multi-scale cascading residual block network for image super resolution. *Multimedia Tools and Applications*, 2022, **81**(5): 6827–6848
- 41 Chen Y P, Fan H Q, Xu B, Yan Z C, Kalantidis Y, Rohrbach M, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 3434–3443
- 42 Bell-Kligler S, Shocher A, Irani M. Blind super-resolution kernel estimation using an internal-GAN. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019. Article No. 26
- 43 Chen J W, Chen J W, Chao H Y, Yang M. Image blind denoising with generative adversarial network based noise modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3155–3164
- 44 Sudiro S A, Kardian A R, Madenda S, Hermanto L. Mean and variance statistic for image processing on FPGA. *International Journal of Applied Science and Engineering*, 2021, **18**(1): Article No. 2020115
- 45 Chen Ye-Yao, Jiang Gang-Yi, Shao Hua, Jiang Hao, Yu Mei. Noise suppression algorithm in the process of high dynamic range image fusion. *Opto-Electronic Engineering*, 2018, **45**(7): Article No. 180083 (陈晔曜, 蒋刚毅, 邵华, 姜浩, 郁梅. 高动态范围图像融合过程中的噪声抑制算法. *光电工程*, 2018, **45**(7): Article No. 180083)
- 46 Luo Z X, Huang Y, Li S, Wang L, Tan T N. Unfolding the alternating optimization for blind super resolution. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020. Article No. 473
- 47 Wang X T, Yu K, Wu S X, Gu J J, Liu Y H, Dong C, et al. ES-RGAN: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018. 63–79
- 48 Zhang Y L, Tian Y P, Kong Y, Zhong B N, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2472–2481
- 49 Zhang Y L, Li K P, Li K, Wang L C, Zhong B N, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 294–310
- 50 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015.
- 51 Pogorelov K, Randel K R, Griwodz C, Eskeland S L, de Lange T, Johansen D, et al. KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference. Taipei, China: Association for Computing Machinery, 2017. 164–169
- 52 Smedsrud P H, Thambawita V, Hicks S A, Gjostang H, Nedrejord O O, Naess E, et al. Kvasir-Capsule, a video capsule endoscopy dataset. *Scientific Data*, 2021, **8**(1): Article No. 142
- 53 Wang X T, Xie L B, Dong C, Shan Y. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE, 2021. 1905–1914
- 54 Zhang K, Van Gool L, Timofte R. Deep unfolding network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 3214–3223
- 55 Luo Z X, Huang Y, Li S, Wang L, Tan T N. Learning the degradation distribution for blind image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 6063–6072
- 56 Chen L H, Yang X M, Jeon G, Anisetti M, Liu K. A trusted medical image super-resolution method based on feedback ad-



aptive weighted dense network. *Artificial Intelligence in Medicine*, 2020, **106**: Article No. 101857

- 57 Gonzalez R C, Woods R E [Author], Ruan Qiu-Qi, Ruan Yu-Zhi [Translator]. *Digital Image Processing*. Beijing: Publishing House of Electronics Industry, 2011. 169–182 (Gonzalez R C, Woods R E [著], 阮秋琦, 阮宇智 [译]. 数字图像处理. 第3版. 北京: 电子工业出版社, 2011. 169–182)
- 58 Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 2013, **20**(3): 209–212
- 59 Blau Y, Mechrez R, Timofte R, Michaeli T, Zelnik-Manor L. The 2018 PIRM challenge on perceptual image super-resolution. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018. 334–355

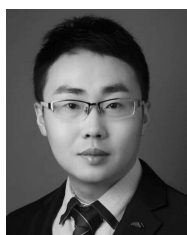


**李 嫣** 中国科学院信息工程研究所硕士研究生. 主要研究方向为医学图像处理, 计算机视觉, 智慧医疗.

E-mail: liyan1999@iie.ac.cn

**(LI Yan** Master student at Institute of Information Engineering, Chinese Academy of Sciences. Her

research interest covers medical image processing, computer vision, and intelligent healthcare.)



**任文琦** 中山大学网络空间安全学院副教授. 主要研究方向为人工智能, 计算机视觉, 图像处理, 网络空间内容安全. 本文通信作者.

E-mail: renwq3@mail.sysu.edu.cn

**(REN Wen-Qi** Associate professor at the School of Cyber Science and

Technology, Sun Yat-sen University. His research interest covers artificial intelligence, computer vision, image processing, and content security in cyberspace.)

Corresponding author of this paper.)

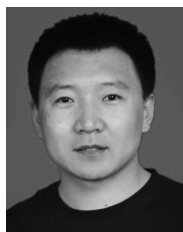


**张长青** 天津大学智能与计算学部副教授. 主要研究方向为机器学习, 计算机视觉, 智能医疗.

E-mail: zhangchangqing@tju.edu.cn

**(ZHANG Chang-Qing** Associate professor at the College of Intelligence and Computing, Tianjin Uni-

versity. His research interest covers machine learning, computer vision, and intelligent healthcare.)



**张金刚** 中国科学院大学未来技术学院副教授. 主要研究方向为智能成像技术, 医学内窥成像, 智能医学健康.

E-mail: zhangjg@ucas.ac.cn

**(ZHANG Jin-Gang** Associate professor at the School of Future Technology, University of Chinese

Academy of Sciences. His research interest covers intelligent imaging technology, medical endoscopy imaging, and intelligence medical health.)



**聂云峰** 布鲁塞尔自由大学应用物理与光子学系教授. 主要研究方向为自由曲面光学设计算法, 成像光谱仪, 生物医学成像.

E-mail: yunfeng.nie@vub.be

**(NIE Yun-Feng** Professor in the Department of Applied Physics and

Photonics, Vrije Universiteit Brussel. Her research interest covers freeform optical design algorithms, imaging spectrometers, and biomedical imaging.)