

基于信息蒸馏机制的水下图像超分辨率重建算法

袁红春, 孔令栋*, 张闪闪, 高凯, 杨钰锐

上海海洋大学信息学院, 上海 201306

摘要 快速而准确的水下图像超分辨率重建技术可以帮助水下航行器更好地感知水下场景, 从而作出导航决策。基于此, 提出一种轻量级的基于信息蒸馏机制的水下图像超分辨率重建算法 (SRIDM)。该算法在普通残差网络的基础上, 引入全局特征融合结构、信息蒸馏机制和空间注意力模块, 进一步提高了模型的特征表达能力。通过模型消融实验, 验证了每个模块的有效性并找到了最佳的模块组合和蒸馏速率。在 USR-248 测试集上的对比实验结果表明, 通过所提算法恢复的图像不管是在主观视觉效果还是在客观评价质量上均优于其他对比算法, 在放大因子为 4 时, 其峰值信噪比和结构相似度分别达到了 27.7640 dB 和 0.7640。另外, 所提算法也是一种轻量化算法, 在保持性能的同时大大降低了模型参数数量和计算复杂度。

关键词 图像处理; 超分辨率重建; 轻量级; 特征融合; 信息蒸馏机制; 空间注意力

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP221324

Super-Resolution Reconstruction Algorithm of Underwater Image Based on Information Distillation Mechanism

Yuan Hongchun, Kong Lingdong*, Zhang Shanshan, Gao Kai, Yang Yurui

School of Information, Shanghai Ocean University, Shanghai 201306, China

Abstract Fast and accurate underwater image super-resolution reconstruction technology can help underwater vehicles better perceive underwater scenes and make navigation decisions. Based on this, a lightweight underwater image super-resolution reconstruction algorithm (SRIDM) based on an information distillation mechanism is proposed. Based on an ordinary residual network, the algorithm presents a global feature fusion structure, information distillation mechanism, and spatial attention module, which further enhances the feature expression ability of the model. The effectiveness of each module was validated through model ablation experiments, and the best module combination and distillation rate were discovered. The experimental results on the USR-248 test set show that the proposed algorithm restores images better than other contrast algorithms in terms of subjective visual effect and objective evaluation quality. When the magnification factor is 4, its peak signal-to-noise ratio and structural similarity reach 27.7640 dB and 0.7640 respectively. Furthermore, the proposed algorithm is also a lightweight algorithm, which significantly reduces the number of model parameters and computational complexity while maintaining performance.

Key words image processing; super-resolution reconstruction; lightweight; feature fusion; information distillation mechanism; spatial attention

1 引言

视觉引导的自主水下航行器在许多重要应用中需要图像合成和场景理解, 如海洋物种和珊瑚礁监测^[1]、海底电缆和残骸检查^[2]、人机协作^[3]等, 其利用合成的水下图像进行视觉注意力建模, 从而作出导航决策。然而, 由于水下能见度低以及水的吸收和散射作用, 再

加上物理成像设备的限制, 水下图像的成像质量普遍较差, 缺乏重要细节, 感兴趣的对象可能会显得异常模糊。当设备无法接近物体以获得更近的视野时 (例如在跟踪快速移动的目标或者探查距离较远的珊瑚礁或海床时), 这一现象就变得尤为突出。快速而准确的水下图像超分辨率重建技术则可以放大感兴趣区域, 实现水下场景的详细感知。

收稿日期: 2022-04-18; 修回日期: 2022-05-21; 录用日期: 2022-06-28; 网络首发日期: 2022-07-08

基金项目: 国家自然科学基金(41776142)

通信作者: *2067718653@qq.com

单幅图像超分辨率(SISR)重建是指将一个低分辨率(LR)图像恢复成一个具有良好视觉效果的高分辨率(HR)图像的过程,然而SISR是一个高度不逆问题,解空间并不唯一,这是因为在原始图像降采样获得低分辨率图像的过程中,大量高频信息丢失,导致恢复过程缺乏足够的可用信息。为了解决这一逆问题,众多超分辨率重建方法已经被提出。目前,这些方法大致可以分为3类:基于插值的方法^[4-8]、基于重建的方法^[9-12]、基于学习的方法^[13-26]。基于插值的方法是在图像连续性假设的前提下,利用周围像素信息去预测未知像素点,简单快速,但其恢复出的图像常常会出现边缘与轮廓比较模糊的状况,纹理也无法得到较好恢复。基于重构的方法主要通过图像先验信息对重构结果进行约束,使其尽可能接近原始图像,但它对于复杂结构图像的处理效果仍旧欠佳。基于学习的方法从样本中学习高/低分辨率图像之间的映射关系,尤其是近些年来基于深度学习的方法在图像超分领域取得了显著成果。

Dong等^[13]首次将包含三层的卷积网络(SRCNN)引入图像超分领域,获得了巨大成功。SRCNN直接学习LR图像到HR图像的非线性映射关系,但由于该网络先利用双三次插值(Bicubic)^[7]将LR图像上采样至目标尺寸后再进行训练,所以整个重建过程是在HR空间上进行的,导致网络计算量大、训练速度慢。为了解决这一问题,Dong等^[14]又提出一种改进的SRCNN算法(FSRCNN),该算法直接将LR图像作为重建网络的输入,并采用反卷积的方式进行上采样,从而获得最终的SR图像。该算法大大加快了网络训练速度,其重建时间仅为SRCNN的1/38^[15]。Shi等^[16]提出的efficient sub-pixel convolutional network(ESPCN)采用一种独特的上采样方式——亚像素卷积,首先对输入特征进行卷积,扩充特征通道,得到 r^2 个特征图,然后将这些特征图沿通道轴进行像素重排,最终得到放大 r 倍的特征图。相比于反卷积,亚像素卷积能够减少人工痕迹,获得更大的感受野信息。因此,现在的许多先进算法都使用亚像素卷积操作进行上采样。受ResNet的启发,Kim等^[17]提出一种非常深的超分辨重建算法(VDSR),将网络层数提升至20层,进一步提高了网络的特征表达能力。随后,一些更深的模型不断涌现,如DRRN^[18]、EDSR^[19]、RCAN^[20]等。EDSR模型在SRResNet^[21]的基础上去除残差块中的批归一化(BN)层,提升了训练速度;考虑到不同通道之间特征表示的相互依赖性和相互作用,RCAN引入通道注意力机制,自适应地学习更多有用的通道特征,但却并未考虑到不同空间位置上的重要性同样存在差异。不同层次网络提取的特征具有不同尺寸的感受野信息,为了充分利用这些分层特征,Zhang等^[22]提出密集残差网络(RDN),在每个残差密集块内,通过融合每层的输入特征和输出特征,加强层与层之间的信息传递,但

是这种叠加式的局部融合方式会极大地增加计算量。而Li等^[23]设计了一种多尺度特征融合网络(MSRN),它仅从全局角度对不同深度的残差块输出特征进行融合,这样一来就大大降低了网络计算量。近些年来,越来越多的研究者开始致力于设计更加高效的轻量级模型。Kong等^[24]在原有SR模型的基础上加入分类器,根据携带高频信息量的高低将输入图像块的恢复难度划分为简单、中等、困难等3个等级,不同等级对应不同复杂度的SR网络。Song等^[25]开创性使用加法网络来实现图像超分学习,避免了卷积过程中的大量乘法操作,大幅降低了浮点运算量。Hui等^[26]提出一种信息蒸馏网络(IDN),对信息蒸馏模块(DBlock)中的中间特征进行截取,将其中的部分特征与输入特征合并,然后通过跳跃连接将其传递至模块尾部。这样虽然可以减少后续的特征通道,降低模型的参数数量和计算量,但模块只进行了单次信息蒸馏,并不能够细致、准确地区分待提炼的特征和需要跨层传递的特征。

为了以尽可能低的网络参数数量和计算负担获取更优的SR重建效果,本文提出一种轻量级的基于信息蒸馏机制的水下图像超分辨率重建网络。该网络主体由多个级联的残差特征蒸馏块(RFDB)组成。在RFDB内部,设计一种新的信息蒸馏方式并加入自定义的空间注意力模块,为图像边缘、纹理等高频细节的恢复提供更多可用信息。除此之外,还引入全局特征融合结构,对每个残差块的输出特征进行全局融合,加强层间信息流动和特征复用。

2 所提方法

2.1 降质退化模型

水下图像降质退化模型描述了水下高分辨率图像转换成人眼观测到的低分辨率图像的全过程,超分辨率重建即为该过程的逆过程,为图像超分技术提供了坚实的理论基础。与陆地成像不同,水下拍摄环境更加复杂多变(水的吸收、散射以及湍流影响等),成像质量容易受到多种因素的综合影响,实际退化过程无法准确定义。但运动模糊、降采样和噪声是其中重要的3个因素,所以一般可以将退化模型表示为

$$L = (k \otimes H) \downarrow_s + n, \quad (1)$$

式中: L 表示观测到的低分辨率图; H 表示高分辨率图像; k 表示模糊核,一般由环境或成像系统本身引起; \downarrow_s 表示缩放因子为 s 的降采样操作,通常由成像系统的分辨率决定; n 表示加性的高斯白噪声,通常来自成像环境或成像过程。由于本实验仅针对双三次下采样图像进行超分辨率重建,所以所提水下图像降质退化模型表示为

$$L = H \downarrow_s^{\text{Bicubic}} + n. \quad (2)$$

2.2 网络概述

针对上述退化模型,提出基于信息蒸馏机制的超

分辨率重建算法,采用端到端的方式自适应地学习低分辨率图像到高分辨率图像的映射关系。该网络主要包含浅层特征提取、特征蒸馏与融合、上采样等 3 个模块。浅层特征提取模块用于提取输入的低分辨率图像的浅层特征。特征蒸馏与融合模块主要由 3 个 RFDB 组成,在每个 RFDB 内部,采用多阶段信息蒸馏机制进

一步提取深度特征,并对 RFDB 输出的分层特征进行聚合。此外,在获得这些分层特征之前,使用空间注意力模块将网络注意力更多地放在携带高频信息的区域。上采样模块利用亚像素卷积将聚合特征放大至目标尺寸,获得最终的 RGB 重建图像。具体网络结构如图 1 所示。

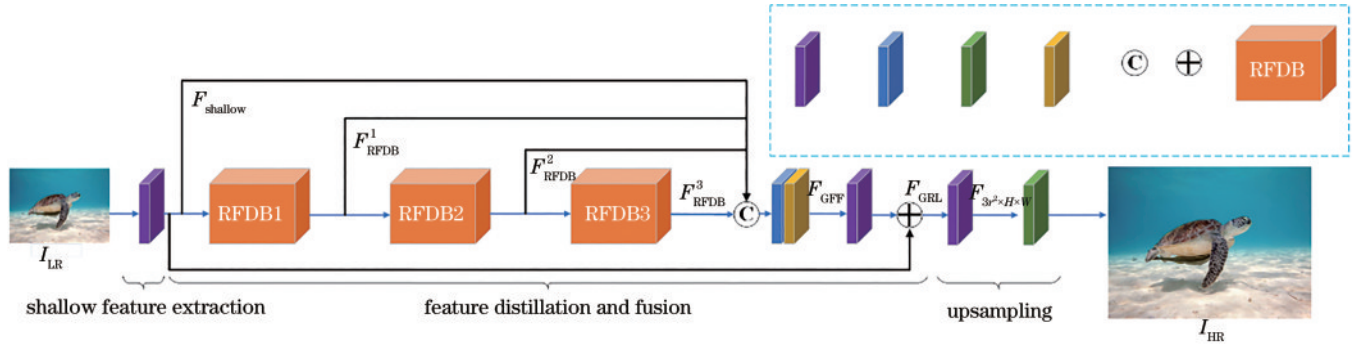


图 1 基于信息蒸馏机制的残差网络结构示意图

Fig. 1 Schematic diagram of residual network structure based on information distillation mechanism

2.3 浅层特征提取

输入图像 I_{LR} 经过 1 层 3×3 的卷积后可提取到浅层特征,该过程可以表示为

$$F_{\text{shallow}} = W_{\text{shallow}}^{3 \times 3} \times I_{LR} + B_{\text{shallow}}, \quad (3)$$

式中: F_{shallow} 表示提取到的浅层特征; $W_{\text{shallow}}^{3 \times 3}$ 表示卷积核的权重参数; B_{shallow} 表示偏置。

2.4 特征蒸馏与融合

首先对 3 个残差特征蒸馏块提取到的全局特征 F_{RFDB}^1 、 F_{RFDB}^2 、 F_{RFDB}^3 及浅层特征 F_{shallow} 进行通道拼接,然后融合得到 F_{GFF} , 并使用 Leaky ReLU 激活。该过程可以描述为

$$F_{\text{GFF}} = \text{Leaky ReLU} \left[W_{\text{GFF}}^{1 \times 1} \times C(F_{\text{RFDB}}^1, F_{\text{RFDB}}^2, F_{\text{RFDB}}^3, F_{\text{shallow}}) + B_{\text{GFF}} \right], \quad (4)$$

$$\text{Leaky ReLU}(x) = \begin{cases} x & x \geq 0 \\ \frac{x}{\alpha} & x < 0 \end{cases}, \quad (5)$$

式中: $C(\cdot)$ 表示通道拼接; $W_{\text{GFF}}^{1 \times 1}$ 表示卷积核的权重参数; B_{GFF} 表示偏置; α 表示负斜率,取 0.05。接着,使用 3×3 的卷积进一步提取特征,最后使用一个全局残差

学习(GRL)得到输出特征 F_{GRL} , 该过程可以表示为

$$F_{\text{GRL}} = (W_{\text{GRL}}^{3 \times 3} \times F_{\text{GFF}} + B_{\text{GRL}}) + F_{\text{shallow}}, \quad (6)$$

式中: $W_{\text{GRL}}^{3 \times 3}$ 表示卷积核的权重参数; B_{GRL} 表示偏置。

2.5 上采样

为了减少网络参数,上采样层仅包含一个可学习层和无参数操作(亚像素卷积)。首先,使用 3×3 的卷积对全局残差学习得到的 F_{GRL} 进行通道压缩或扩张,使其输出通道数变为 $3r^2$, 然后使用亚像素卷积进行像素重排,得到重建图像 \hat{I}_{HR} , 该过程可以表示为

$$F_{3r^2 \times H \times W} = W_{\text{up}}^{3 \times 3} \times F_{\text{GFF}} + B_{\text{up}}, \quad (7)$$

$$\hat{I}_{HR} = M_{\text{PixelShuffle}}(F_{3r^2 \times H \times W}), \quad (8)$$

式中: $W_{\text{up}}^{3 \times 3}$ 表示上采样使用的卷积核的权重参数; B_{up} 表示偏置; $M_{\text{PixelShuffle}}(\cdot)$ 表示像素重排操作。

2.6 RFDB

所设计的 RFDB 如图 2 所示。为了提高图像重建质量,在普通深度卷积的基础上又引入信息蒸馏机制以及空间注意力模块。首先将输入 F_{in} 经过所有通道分离后的蒸馏特征进行融合得到 F_{LFF} , 然后通过空间注意力(SA)模块,获得空间注意力值 A_{SA} , 并使用该

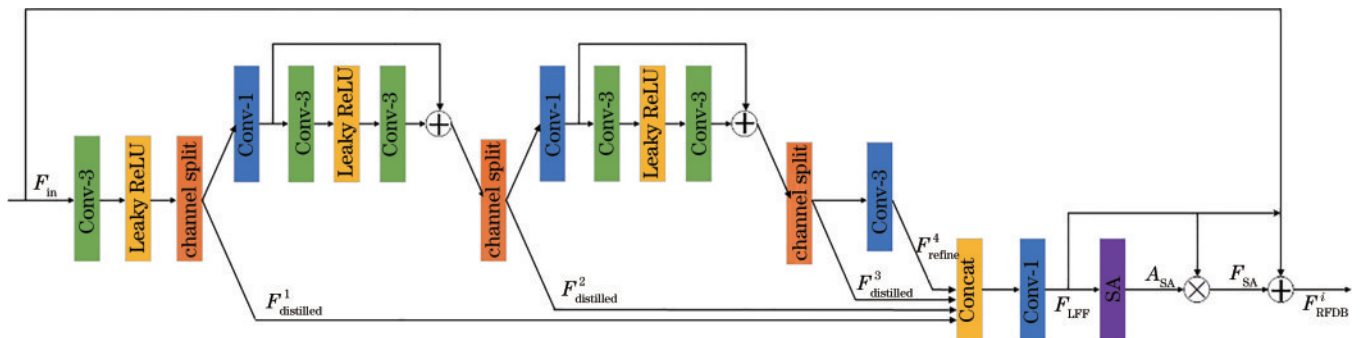


图 2 残差特征蒸馏块

Fig. 2 Residual feature distillation block

值对 F_{LFF} 不同空间位置进行加权,得到 F_{SA} ,为了更加顺利地将前一层特征传递到下一层,又加入一个短跳连接,最后使用残差学习获得第 i 个 RFDB 的输出 F_{RFDB}^i ,该过程可以表示为

$$A_{SA} = S(F_{LFF}), \quad (9)$$

$$F_{SA} = F_{LFF} \otimes A_{SA}, \quad (10)$$

$$F_{RFDB}^i = F_{SA} + F_{LFF} + F_{in}, i = 1, 2, 3. \quad (11)$$

2.6.1 信息蒸馏机制

为此,设计了一种新的多级信息蒸馏机制:输入特征 F_{in} 经过一层卷积后按照给定的蒸馏速率进行通道分离,获得一级蒸馏特征 $F_{distilled}^1$ 和待提炼特征 F_{refine}^1 。

在 F_{refine}^1 进入下一个蒸馏单元之前,首先使用 1×1 卷积对待提炼特征进行通道升维(使其与输入特征的通道数相同),然后使用包含两层卷积的残差块进一步提取深度特征,最后采用通道分离的方式获得二次蒸馏特征 $F_{distilled}^2$ 和待提炼特征 F_{refine}^2 。同样地,可以获得三级蒸馏特征 $F_{distilled}^3$ 和待提炼特征 F_{refine}^3 ,值得注意的是,最后一次通道分离后的待提炼特征直接进行 3×3 卷积得到精炼特征 F_{refine}^4 ,不再进行通道分离。该过程可以表示为

$$F_{distilled}^1, F_{refine}^1 = \text{split}\left\{\text{Leaky ReLU}\left[\text{Conv}^{3 \times 3}(F_{in})\right]\right\}, \quad (12)$$

$$F_{distilled}^2, F_{refine}^2 = \text{split}\left\{\text{Conv}^{3 \times 3}\left\{\text{Leaky ReLU}\left\{\text{Conv}^{3 \times 3}\left[\text{Conv}^{1 \times 1}(F_{refine}^1)\right]\right\}\right\} + \text{Conv}^{1 \times 1}(F_{refine}^1)\right\}, \quad (13)$$

$$F_{distilled}^3, F_{refine}^3 = \text{split}\left\{\text{Conv}^{3 \times 3}\left\{\text{Leaky ReLU}\left\{\text{Conv}^{3 \times 3}\left[\text{Conv}^{1 \times 1}(F_{refine}^2)\right]\right\}\right\} + \text{Conv}^{1 \times 1}(F_{refine}^2)\right\}, \quad (14)$$

$$F_{refine}^4 = \text{Conv}^{3 \times 3}(F_{refine}^3), \quad (15)$$

式中: $\text{split}(\cdot)$ 表示通道分离; $\text{Conv}^{3 \times 3}(\cdot)$ 表示卷积核为 3×3 的卷积操作; $\text{Conv}^{1 \times 1}(\cdot)$ 表示卷积核为 1×1 的卷积操作。最后利用 1×1 卷积将所有蒸馏特征 $F_{distilled}^1$ 、 $F_{distilled}^2$ 、 $F_{distilled}^3$ 和最终的精炼特征 F_{refine}^4 沿着通道维度进行融合,该过程可以描述为

$$F_{LFF} = W_{LFF}^{1 \times 1} \times C(F_{distilled}^1, F_{distilled}^2, F_{distilled}^3, F_{refine}^4) + B_{LFF}. \quad (16)$$

2.6.2 空间注意力模块

在 LR 图像的不同空间位置,低频信息和高频信息的分布并不一致。一些区域比较平滑,恢复起来也比较容易;而一些区域则包含大量的边界、纹理等高频细节,恢复起来相对困难。因此,有必要对这些区域加以区别对待,将更多的注意力放在携带高频信息的区域。为此,设计了一种关注空间区域的注意力模块,其结构如图 3 所示。

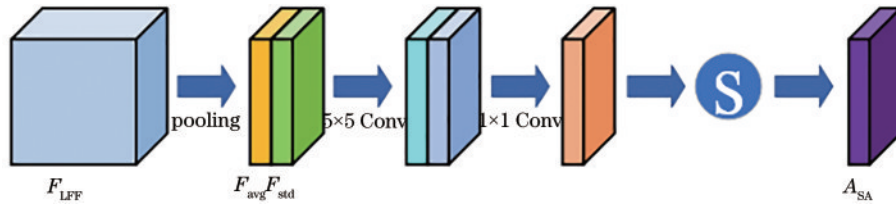


图 3 空间注意力模块

Fig. 3 Spatial attention module

空间注意力模块首先对特征 F_{LFF} 沿着通道轴进行平均池化和标准差池化:

$$\bar{X}(i, j) = \frac{1}{C} \sum_{c=1}^C X_c(i, j), \quad (17)$$

$$\sigma(i, j) = \sqrt{\frac{1}{C} \sum_{c=1}^C [X_c(i, j) - \bar{X}(i, j)]^2}, \quad (18)$$

式中: $\bar{X}(i, j)$ 表示在空间位置 (i, j) 处通道平均池化的结果; $\sigma(i, j)$ 表示在空间位置 (i, j) 处通道标准差池化的结果; $X_c(i, j)$ 表示第 c 个通道 (i, j) 位置上的特征值; C 表示通道数。然后对获取的两个池化结果进行通道拼接,得到两组空间特征描述符 F_{avg} 和 F_{std} ,接着使用一层卷积(卷积核 5×5 ,步长为 1)对特征描述符中不同位置的特征值进行融合,并将其压缩为单通道,最后经 Sigmoid 激活函数将其归一化到 $0 \sim 1$ 之间,得到空间注意力值 A_{SA} 。该过程可以描述为

$$A_{SA} = \text{Sigmoid}\left\{\text{Conv}^{1 \times 1}\left\{\text{ReLU}\left\{\text{Conv}^{5 \times 5}\left[C(F_{avg}, F_{std})\right]\right\}\right\}\right\}, \quad (19)$$

式中, $\text{Sigmoid}(\cdot)$ 、 $\text{ReLU}(\cdot)$ 分别表示 Sigmoid 激活函数、ReLU 激活函数,其定义分别为

$$\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)}, \quad (20)$$

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (21)$$

2.7 损失函数

在图像超分辨领域,损失函数的定义多种多样。常见的有平均绝对误差(L1 损失),均方误差(L2 损失)等。L2 损失会惩罚较大的误差,但更能容忍较小的误差,因此,通常会导致恢复出来的图像过于平滑,缺乏高频细节。相比于 L2 损失,L1 损失表现出更高

的重建性能和收敛性^[27],所以采用 L1 损失函数来优化模型参数:

$$L_1(\hat{I}_{HR}, I_{HR}) = \frac{1}{N} \sum_{i=1}^N |\hat{I}_{HR}^{(i)} - I_{HR}^{(i)}|, \quad (22)$$

$$\hat{\theta} = \operatorname{argmin}_{\theta} L_1(\hat{I}_{HR}, I_{HR}), \quad (23)$$

式中: $\hat{I}_{HR}^{(i)}$ 、 $I_{HR}^{(i)}$ 分别表示第 i 个样本对应的重建图像和其真实高分辨率图像; N 表示数据集样本量; θ 表示网络需要学习的参数; $\hat{\theta}$ 表示网络迭代更新后的参数。

3 实验结果与分析

3.1 实验数据

使用由 Minnesota University 交互式机器人与视觉实验室最新公布的水下图像数据集 USR-248^[28],该数据集包含大量的水下实体,包括珊瑚、鱼、残骸、水下航行器以及潜水员等。训练集包含 1060 张 HR 图像 (480×640) 以及对应的 LR 图像(使用双三次插值降采样获得);测试集包含 248 张 HR 图像及对应的 LR 图像。因为只讨论针对简单退化模型的超分辨率重建,所以去除了原 LR 图像 20% 的高斯噪声,只保留下采样操作(双三次插值)。从 USR-248 训练集中选择编号 1~1040 的图像作为训练集,从编号 1041~1060 中选取 5 张内容丰富、纹理清晰的图像作为验证集,使用 USR-248 全部的测试集对所提模型进行测量评估。

3.2 训练细节

将 LR 图像随机裁剪成 40×40 的图像块作为输入,并根据放大因子 2、4、8 分别将对应 HR 图像裁剪成 80×80 、 160×160 、 320×320 的图像块。为了避免出现欠拟合现象,采用数据增强的方法将输入图像块随机旋转 90° 、 180° 、 270° 及水平翻转,使得数据扩充为原来的 8 倍。将模型宽度设定为 64,批处理大小为 8,模型参数使用 Adam 算法更新迭代,算法中参数设置如下: $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ 、 $\epsilon = 10^{-8}$,初始学习率设定为 0.0001,模型训练 1000 个 epoch,每 250 个 epoch 后学习率衰减为原来的一半。实验条件如下:CPU 为 AMD Ryzen 9 5950X 16-Core Processor, GPU 为 NVIDIA GeForce RTX 3090,操作系统为 Ubuntu 18.04.5,深度学习框架为 PyTorch。

3.3 结果与分析

3.3.1 评价指标

使用峰值信噪比 (R_{PSNR})^[29] 以及结构相似度 (M_{SSIM})^[30] 这两个参数作为评价指标。 R_{PSNR} 是基于误差敏感的图像评价指标,主要由均方误差 (E_{MSE}) 决定:

$$E_{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2, \quad (24)$$

$$R_{PSNR} = 10 \log_{10} \left(\frac{x_{\max}^2}{E_{MSE}} \right), \quad (25)$$

式中: x_i 表示 HR 图像第 i 个位置上的像素值; \hat{x}_i 表示 SR 图像第 i 个位置上的像素值; N 表示图像总的像素

个数; x_{\max} 表示图像像素值最大可能取值。因为 R_{PSNR} 度量的是相同位置上的逐像素差异,并未考虑到人的视觉感知,所以又引入了另一个评价指标 M_{SSIM} 作为其补充。 M_{SSIM} 度量图像之间的结构相似性,由图像间的亮度、对比度和结构所共同决定:

$$M_{SSIM} = \frac{(2\mu_x \mu_x + c_1)(\sigma_{xx} + c_2)}{(\mu_x^2 + \mu_x^2 + c_1)(\sigma_x^2 + \sigma_x^2 + c_2)}, \quad (26)$$

式中: x 表示参考图像; \hat{x} 表示重构图像; μ_x 、 μ_x 分别为 x 和 \hat{x} 的均值; σ_x 、 σ_x 为方差; σ_{xx} 为 x 和 \hat{x} 的协方差;维持稳定的常数 $c_1 = (k_1 L)^2$ 、 $c_2 = (k_2 L)^2$, L 是像素值的动态范围,取 255, k_1 和 k_2 取 0.01 和 0.03。所有关于 R_{PSNR} 、 M_{SSIM} 的实验结果均是先将 RGB 图像转换成 YCbCr 空间,然后在 Y 通道上测得的。

3.3.2 模型分析

图 4 展示了在放大因子为 4 时,模型训练过程中的损失值的变化情况。从图中可以看出,在大约 100 个 epoch 后,损失值基本稳定在一定范围,不再明显下降,说明此时模型在训练集上已经开始收敛。但为了达到更高的精度,继续进行迭代训练直至完成 1000 个 epoch。

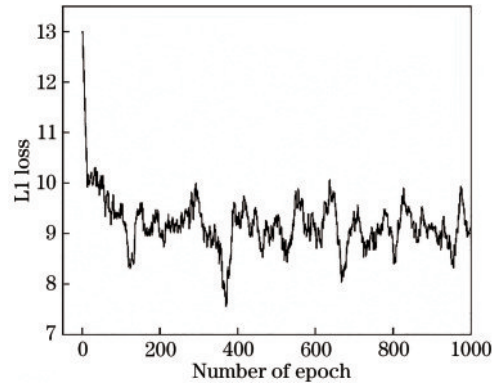


图 4 训练损失

Fig. 4 Training loss

为了探究所使用的全局特征融合结构和空间注意力模块对最终重建结果的影响,进行了消融实验。图 5 表示在训练过程中 4 种模型在验证集上的效果,

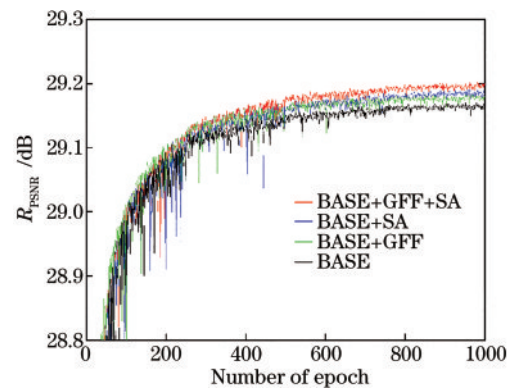


图 5 关于 GFF 和 SA 的消融实验结果(验证集)

Fig. 5 Results of ablation experiments on GFF and SA (validation set)

Base 表示基础模块, GFF 表示全局特征融合模块, SA 表示空间注意力模块。从图中可以看出, 随着训练迭代次数的增加, 4 种模型对应的 R_{PSNR} 均在稳步提高。最终的实验结果表明: GFF 和 SA 能进一步提高模型性能; 将 SA 和 GFF 耦合使用才能使模型作用发挥最大化。值得一提的是, 在训练初始阶段可以明显看到, Base+GFF 模型取得了领先的优势, 主要是因为它采用了全局特征融合结构, 能够快速地将图像低频特征传递至网络末端, 加快重建过程。

将 4 种模型分别在 USR-248 测试集上进行定量分析, 结果如表 1 所示, 可以看出, 不管是在 Base 上应用 GFF 还是 SA, R_{PSNR} 值均能有所提升, 尤其是在两者组合使用的情况下, R_{PSNR} 值直接提升 0.02 dB 左右。这与验证集上的实验结果是相一致的。

表 1 关于 GFF 和 SA 的消融实验结果(测试集)
Table 1 Results of ablation experiments on GFF and SA (test set)

Scale	Base	GFF	SA	R_{PSNR} /dB	M_{SSIM}
×4	✓	×	×	27.5629	0.7142
	✓	✓	×	27.5660	0.7143
	✓	×	✓	27.5690	0.7146
	✓	✓	✓	27.5851	0.7150

除此之外, 进一步分析了 SA 模块, 把经过 SA 模块的某一个前后特征通道以及空间注意力值进行可视化(为了更好的直观效果, 这里进行了伪彩色映射), 结果如图 6 所示。可以看到: 获得的空间注意力值 A_{SA} 是合理且有效的: 在越复杂的区域, 注意力值越大; 在越平滑的区域, 注意力值越小。经过 SA 模块后的特征图与之前相比, 边缘变得更加锐利、清晰。

为了进一步分析多级信息蒸馏机制对重建效果的影响, 将 RFDB 模块 3 个阶段的蒸馏特征进行可视化,

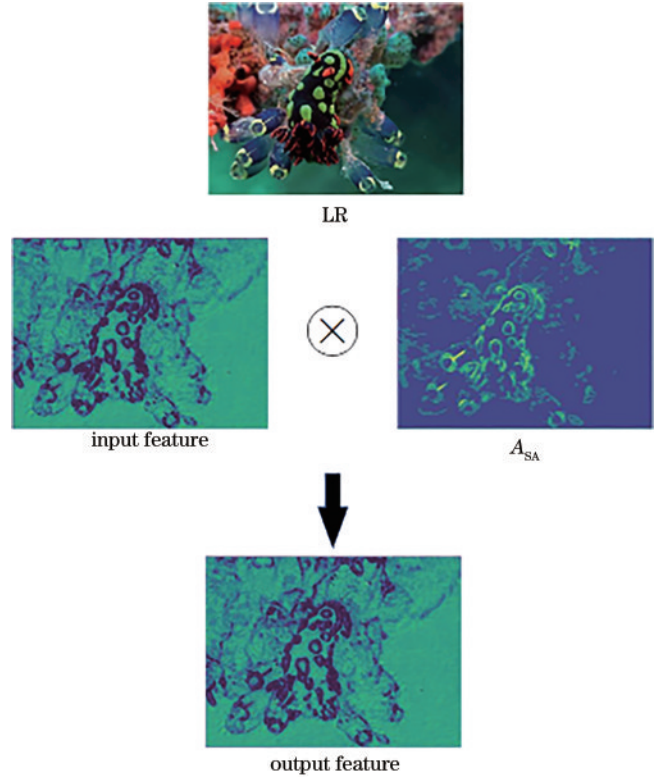


图 6 SA 可视化结果
Fig. 6 Visualization of SA

结果如图 7 所示。从图中可以看出, 随着蒸馏阶段的递进, 该网络也越来越关注于图像的高层语义特征: 在 stage 1 和 stage 2 阶段, 提取到的特征多是关于图像的整体轮廓; 在 stage 3 阶段, 提取到的特征则包含了更多的边界和纹理信息。蒸馏速率代表着在通道切片操作中蒸馏特征占总特征的比例, 表 2 展示了模型在 3 种蒸馏速率下在测试集上的定量分析结果。结果表明, 在蒸馏速率 d 为 0.25 时, 其获得的 R_{PSNR} 和 M_{SSIM} 值最高。最终模型也采用了这一蒸馏速率。

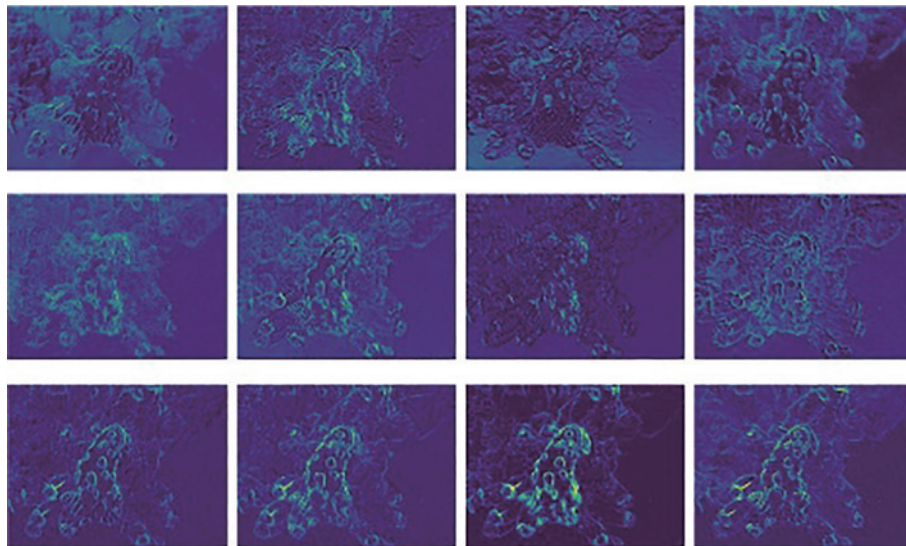


图 7 不同阶段的蒸馏特征图(部分)
Fig. 7 Distilled feature maps from different stages (part)

表 2 参数 d 对 R_{PSNR} 和 M_{SSIM} 的影响

Table 2 Influence of parameter d on R_{PSNR} and M_{SSIM}

d	R_{PSNR} /dB	M_{SSIM}
0.25	27.7640	0.7640
0.5	27.7566	0.7635
0.75	27.7480	0.7631

3.3.3 基准测试

为了验证所提算法的性能,将其与 SRCNN、DSRCNN^[31]、EDSR、RCAN、CARN^[32]、IMDN^[26]等 6 种算法进行对比实验。由于所提算法是轻量化算法,所以对公平的公平性,对 EDSR 和 RCAN 模型进行简化,将其网络主体部分的残差块数目全部设置为 12。所有算法均使用与本研究相同的训练数据和训练技巧进行训练。这些算法在图像超分领域具有广泛代表性,在自然图像数据集上也取得了不错的效果,但很少有人把这些模型应用于水下图像超分,因此其泛化性有待检验。表 3 展示了在放大因子分别为 2、4 和 8 时,不同算法在 USR-248 测试集上的定量分析结果。通过对比可以发现:所提算法在 $\times 2$ 和 $\times 4$ 测试集上能够以相对较低参数量(Params)、浮点运算量(FLOPs)取得最高的 R_{PSNR} 和 M_{SSIM} 指标;在 $\times 8$ 测试集上得到的重建效果依旧良好,但领先其他算法的优势已不够明显。这是因为所提空间注意力模块是基于低分辨率特征图设计的,所以当放大因子较大时,SRIDM 对重建结果的提升相对较小。

图 8、图 9 和图 10 展示了在 USR-248 测试集上由不同算法(SRCNN、DSRCNN、EDSR、CARN^[32]、IMDN、SRIDM)获得的超分结果,并给出了真实的 HR 图像以供参考。为了更清晰地进行比较,对左侧矩形框中的内容进行了局部放大。观察图 8 可以看出,SRIDM 对潜水员的重建轮廓更加清晰且具有更好的视觉效果。观察图 9 和图 10 可以看出,当放大因子较大时,所有算法均出现了不同程度的模糊、失真(尤

表 3 超分辨算法在测试集上的 R_{PSNR} 和 M_{SSIM} 平均值

Table 3 Comparison of average R_{PSNR} and M_{SSIM} by different super-resolution algorithms on test datasets

Algorithm	Scale	Params / 10^3	FLOPs / 10^9	R_{PSNR} /dB	M_{SSIM}
Bicubic	$\times 2$			32.5489	0.9083
SRCNN	$\times 2$	8	3.1	32.7391	0.9097
DSRCNN	$\times 2$	297	2896.0	33.5831	0.9186
EDSR	$\times 2$	1074	92.0	33.6245	0.9189
RCAN	$\times 2$	1192	101.6	33.6207	0.9191
CARN	$\times 2$	964	92.7	33.6721	0.9193
IMDN	$\times 2$	694	62.2	33.6810	0.9194
SRIDM	$\times 2$	668	60.1	33.6983	0.9198
Bicubic	$\times 4$			26.7702	0.7432
SRCNN	$\times 4$	8	3.1	27.1537	0.7402
DSRCNN	$\times 4$	297	2896.0	27.6512	0.7572
EDSR	$\times 4$	1222	36.0	27.7067	0.7628
RCAN	$\times 4$	1340	38.4	27.7003	0.7626
CARN	$\times 4$	1112	36.2	27.7426	0.7634
IMDN	$\times 4$	715	16.0	27.7336	0.7637
SRIDM	$\times 4$	689	15.5	27.7640	0.7640
Bicubic	$\times 8$			23.5757	0.6080
SRCNN	$\times 8$	8	3.1	24.1015	0.6043
DSRCNN	$\times 8$	297	2896.0	24.4117	0.6211
EDSR	$\times 8$	1370	22.1	24.4836	0.6304
RCAN	$\times 8$	1487	22.7	24.4814	0.6295
CARN	$\times 8$	1260	22.1	24.5004	0.6310
IMDN	$\times 8$	798	4.4	24.4907	0.6303
SRIDM	$\times 8$	772	4.3	24.5018	0.6304

其是 SRCNN),在 8 倍放大情况下,甚至出现了局部噪点(DSRCNN 最为明显)。但通过对比可以发现,SRIDM 不论是在局部纹理恢复还是在重建图像的颜色饱和度、锐度以及对比度方面,都比其他算法表现出了更优的处理效果。这主要得益于 SRIDM 具有更强大的表征能力,可以从 LR 空间提取更加复杂的特征。

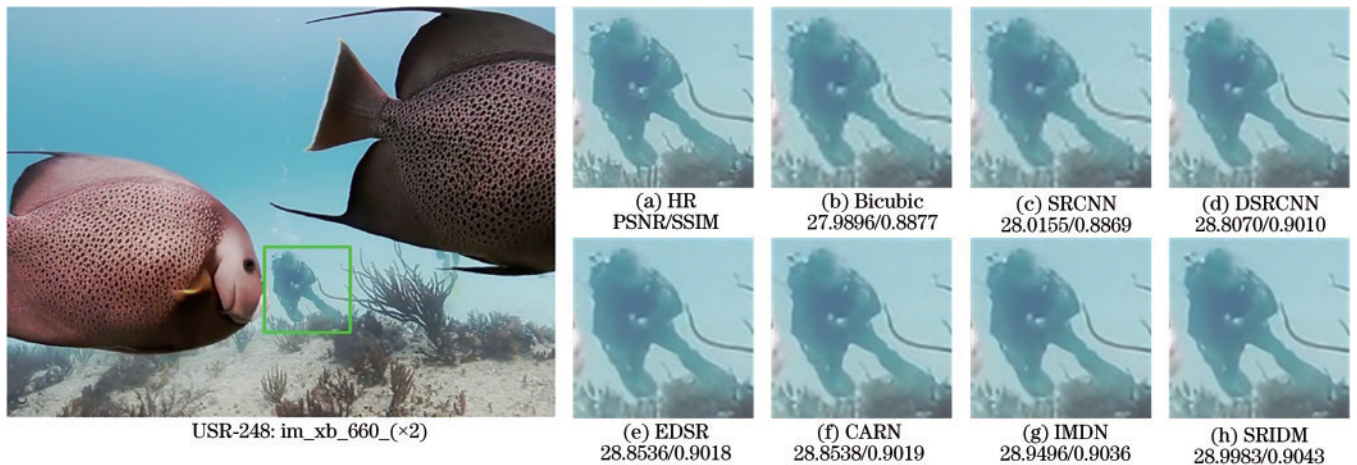


图 8 不同算法对 USR-248 数据集中图像 im_xb_660_ 在 2 倍上采样情况下的效果(R_{PSNR} 和 M_{SSIM})比较

Fig. 8 Comparison of visual results (R_{PSNR} and M_{SSIM}) of im_xb_660_ from USR-248 dataset among different algorithms ($\times 2$)

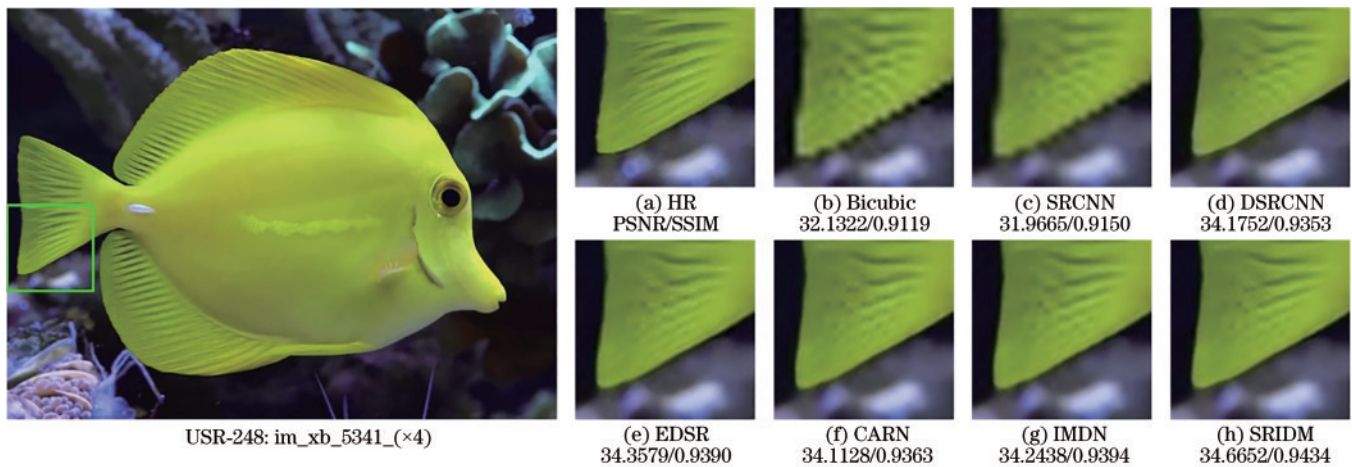


图 9 不同算法对 USR-248 数据集中图像 $im_xb_5341_$ 在 4 倍上采样情况下的效果 (R_{PSNR} 和 M_{SSIM}) 比较

Fig. 9 Comparison of visual results (R_{PSNR} and M_{SSIM}) of $im_xb_5341_$ from USR-248 dataset among different algorithms ($\times 4$)

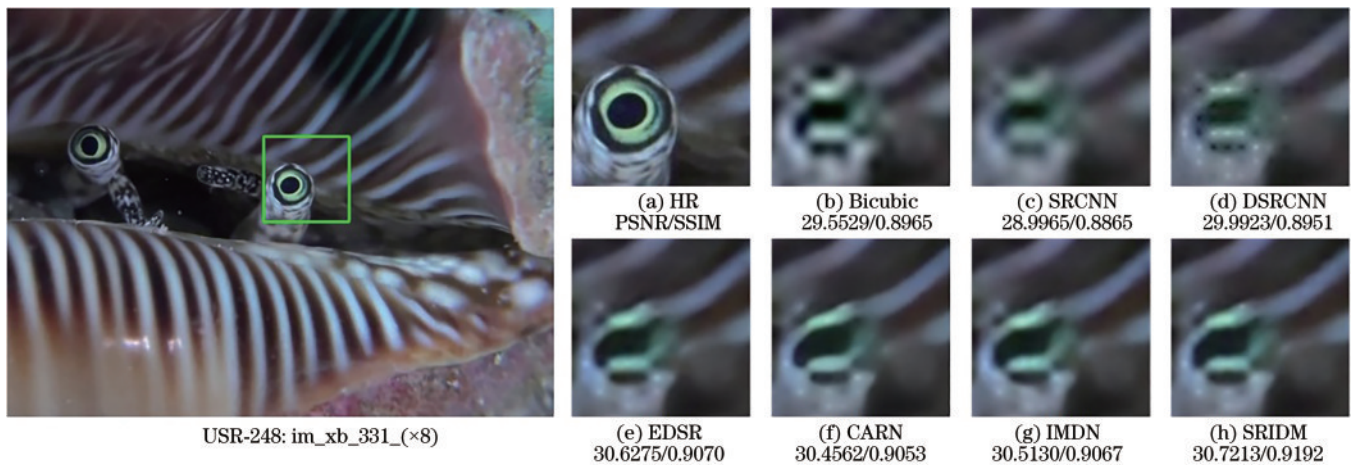


图 10 不同算法对 USR-248 数据集中图像 $im_xb_331_$ 在 8 倍上采样情况下的效果 (R_{PSNR} 和 M_{SSIM}) 比较

Fig. 10 Comparison of visual results (R_{PSNR} and M_{SSIM}) of $im_xb_331_$ from USR-248 dataset among different algorithms ($\times 8$)

3.3.4 真实水下场景测试

由于在真实的水下环境中,图像的降质模型并不能简单地被人为设定,这也使得面向真实应用场景的水下图像 SR 任务更具有挑战性。为了进一步评价 SRIDM 在现实场景中的表现,收集了部分真实的水下低分辨率图像,并分别通过 Bicubic 方法和 SRIDM 方法对其进行 4 倍超分,结果如图 11 所示。每个子图的左边表示使用 Bicubic 方法得到的放大结果;右边表示使用 SRIDM 方法得到的放大结果。视觉对比结果表明,所提 SRIDM 方法可以在真实的水下场景中重建出具有良好视觉质量的 HR 图像。在未来的工作中,将进一步探索面向真实数据集的超分辨率模型研究。

3.4 模型复杂度分析

水下图像超分辨率重建可帮助水下自动驾驶车辆或远程操作车辆实现对特定感兴趣区域的放大,以促进水下场景的详细理解。若想在这些嵌入式机器人平台上使用深度学习模型并满足实时性的要求,就必然

要考虑到算法的复杂度。从模型参数量、浮点运算量以及运行时间等 3 个方面来分析模型的复杂度(以 HR 图像大小进行衡量),结果如表 4 和图 12 所示。SRIDM 模型大小只有 2.6~3.0 MB,在 4 倍采样率下,在 RTX 3090 上处理速度可达到 260 frame/s 并且每帧的浮点运算量仅为 15.48×10^9 。图 12 展示了 SRIDM 与其他算法在参数量、浮点运算量以及运行时间上的对比结果($\times 4$)。可以看出:SRIDM 在参数量上处于中间水平,与 IMDN 接近而远低于 CARN、EDSR 和 RCAN;在浮点运算量上保持了与 IMDN、EDSR、RCAN 和 CARN 相近的量级,远低于 DSRCNN;在运行时间上仅次于 SRCNN,但其获得的 R_{PSNR} 却远远高于 SRCNN。与轻量化模型 CARN 对比发现,SRIDM 仅花费了 CARN 约 53% 的参数量和浮点运算量,却可在 CARN 的基础上进一步获得 0.02 dB 的 R_{PSNR} 增益。所提算法在实现高质量重建效果的同时,又大大减小了网络规模和计算量。

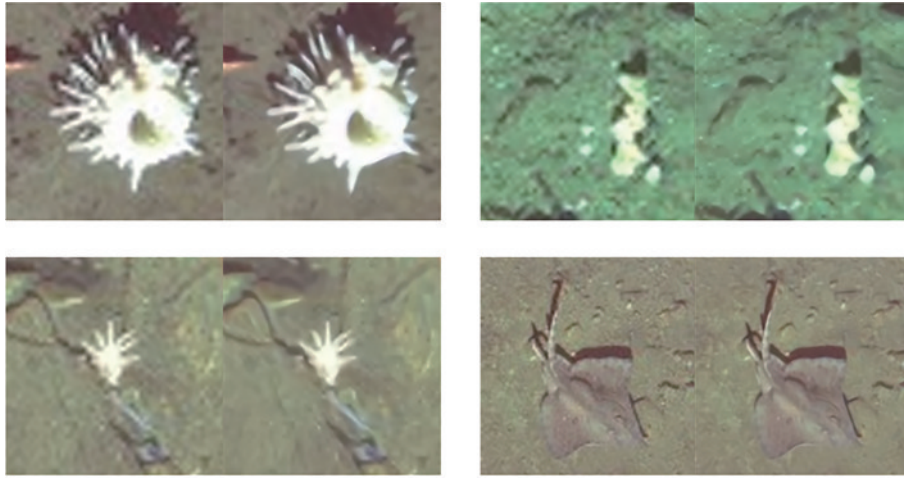


图 11 SRIDM 在真实水下场景中的超分辨率重建结果

Fig. 11 SR result generated with SRIDM in real underwater scenes

表 4 SRIDM 在 RTX 3090 上的参数量、FLOPs、运行速度和存储大小

Table 4 Number of params, FLOPs, run-time, and memory requirement of SRIDM on RTX 3090

Model	×2	×4	×8
Params / 10 ³	668	689	772
FLOPs / 10 ⁹	60.15	15.48	4.31
Run-time / ms	3.85	3.73	3.59
Frames per second	259.74	268.10	278.55
Model_size / MB	2.6	2.7	3.0

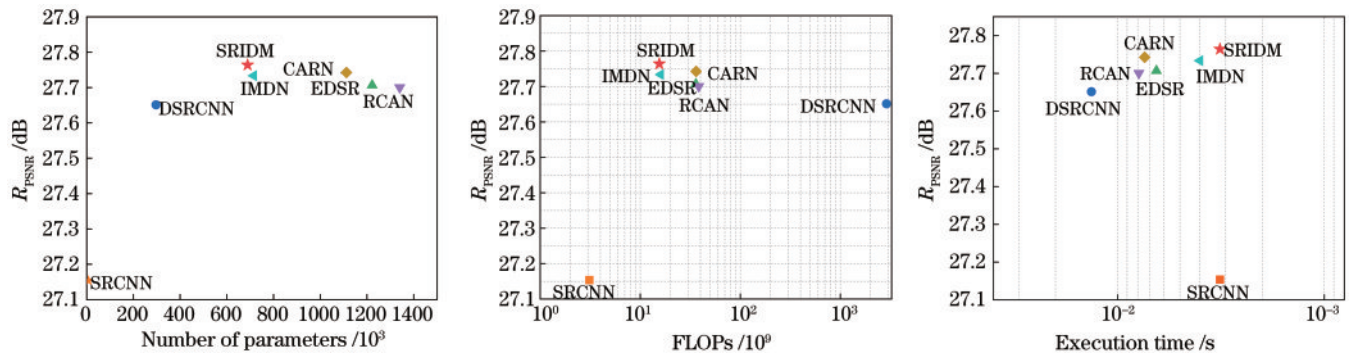


图 12 不同算法在参数量、浮点运算量以及运行时间上的对比

Fig. 12 Comparison of parameters, floating-operations, and running time of different algorithms

4 结 论

为了在性能和复杂度之间取得更好的平衡,提出一种轻量级的基于信息蒸馏机制的水下图像超分辨率重建算法。首先在传统的残差网络上加入全局特征融合结构,加强层间信息流动和特征重用;其次通过设计的信息蒸馏结构,逐阶段地提取特征,将模型的学习任务更多地放在图像纹理细节的重建;最后在 RFDB 中加入自定义的空间注意力模块,使得网络能够自适应地对图像空间分配权重,将更多的注意力放在携带高频信息的区域,以帮助网络快速地恢复高频细节。

通过模型消融实验可知,所采用的全局特征融合

结构和空间注意力模块都可以有效地提升重建图像的质量,尤其将两者耦合使用时提升的效果更好。模型蒸馏速率设置为 0.25 时可以获得最佳性能。通过基准测试和算法复杂度分析发现,与其他先进算法相比,所提算法不管是在主观视觉效果还是在客观量化评价上均能以相对较低的计算量获得更优的结果。所提算法只是针对简单退化图像的超分辨率重建,并未考虑噪声和模糊的影响。未来的工作将研究如何在包含未知噪声及未知模糊的复杂应用场景下进一步提升网络模型的鲁棒性。

参 考 文 献

[1] Hoegh-Guldberg O, Mumby P J, Hooten A J, et al.

- Coral reefs under rapid climate change and ocean acidification[J]. *Science*, 2007, 318(5857): 1737-1742.
- [2] Bingham B, Foley B, Singh H, et al. Robotic tools for deep water archaeology: surveying an ancient shipwreck with an autonomous underwater vehicle[J]. *Journal of Field Robotics*, 2010, 27(6): 702-717.
- [3] Islam M J, Ho M, Sattar J. Understanding human motion and gestures for underwater human-robot collaboration[J]. *Journal of Field Robotics*, 2019, 36(5): 851-873.
- [4] Singh A, Singh J. Content adaptive single image interpolation based super resolution of compressed images [J]. *International Journal of Electrical and Computer Engineering (IJECE)*, 2020, 10(3): 3014-3021.
- [5] Ni K S, Nguyen T Q. An adaptable k -nearest neighbors algorithm for MMSE image interpolation[J]. *IEEE Transactions on Image Processing*, 2009, 18(9): 1976-1987.
- [6] Hwang J W, Lee H S. Adaptive image interpolation based on local gradient features[J]. *IEEE Signal Processing Letters*, 2004, 11(3): 359-362.
- [7] Keys R. Cubic convolution interpolation for digital image processing[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, 29(6): 1153-1160.
- [8] Hou H, Andrews H. Cubic splines for image interpolation and digital filtering[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1978, 26(6): 508-517.
- [9] Wei S F, Zhou X Z, Wu W, et al. Medical image super-resolution by using multi-dictionary and random forest[J]. *Sustainable Cities and Society*, 2018, 37: 358-370.
- [10] Ma X Y, Zhang J S, Li T, et al. Super-resolution geomagnetic reference map reconstruction based on dictionary learning and sparse representation[J]. *IEEE Access*, 2020, 8: 84316-84325.
- [11] Xu Y, Wu Z B, Chanussot J, et al. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution[J]. *IEEE Transactions on Image Processing*, 2019, 28(6): 3034-3047.
- [12] Mishra D, Majhi B, Sa P K, et al. Development of robust neighbor embedding based super-resolution scheme [J]. *Neurocomputing*, 2016, 202: 49-66.
- [13] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[M]// Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8692: 184-199.
- [14] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9906: 391-407.
- [15] 唐艳秋, 潘泓, 朱亚平, 等. 图像超分辨率重建研究综述[J]. *电子学报*, 2020, 48(7): 1407-1420.
Tang Y Q, Pan H, Zhu Y P, et al. A survey of image super-resolution reconstruction[J]. *Acta Electronica Sinica*, 2020, 48(7): 1407-1420.
- [16] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [17] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [18] Tai Y, Yang J, Liu X M. Image super-resolution via deep recursive residual network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2790-2798.
- [19] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.
- [20] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[M]// Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 294-310.
- [21] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 105-114.
- [22] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2472-2481.
- [23] Li J C, Fang F M, Mei K F, et al. Multi-scale residual network for image super-resolution[M]// Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11212: 527-542.
- [24] Kong X T, Zhao H Y, Qiao Y, et al. ClassSR: a general framework to accelerate super-resolution networks by data characteristic[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 12011-12020.
- [25] Song D H, Wang Y H, Chen H T, et al. AdderSR: towards energy efficient image super-resolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 15643-15652.
- [26] Hui Z, Gao X B, Yang Y C, et al. Lightweight image super-resolution with information multi-distillation network [C]//MM '19: Proceedings of the 27th ACM International Conference on Multimedia, October 21-25, 2019, Nice, France. New York: ACM Press, 2019: 2024-2032.
- [27] Tai Y, Yang J, Liu X M, et al. MemNet: a persistent

- memory network for image restoration[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4549-4557.
- [28] Islam M J, Enan S S, Luo P G, et al. Underwater image super-resolution using deep residual multipliers[C]//2020 IEEE International Conference on Robotics and Automation, May 31-August 31, 2020, Paris, France. New York: IEEE Press, 2020: 900-906.
- [29] Horé A, Ziou D. Image quality metrics: PSNR *vs.* SSIM [C]//2010 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, Turkey. New York: IEEE Press, 2010: 2366-2369.
- [30] 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.
- [31] Mao X J, Shen C H, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections[EB/OL]. (2016-06-29)[2021-02-04]. <https://arxiv.org/abs/1606.08921>.
- [32] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11214: 256-272.