

# 基于整体注意力的图像超分辨率反投影网络

戚唐尧, 张东晓

(集美大学理学院, 福建 厦门 361021)

**[摘要]** 针对图像超分辨率重建问题, 提出了一种基于整体注意力的反投影网络。该网络中的上下迭代投影模块与整体注意力模块相互协同, 共同捕获更多特征信息。通过上下迭代投影模块的相互转化, 获取不同分辨率特征之间的关联性, 帮助学习到高频纹理细节。使用层注意力模块, 将上下投影输出的特征层分别进行层之间的注意力加权, 用以寻找不同层、通道和位置之间相互依存的关系。此外, 针对最后的上投影模块使用了通道空间注意力模块, 可以引导网络学习到通道内部和通道间的相关信息。消融实验表明, 两种注意力模块对特征信息的增强有助于提升重建效果。对比实验表明, 本文提出的基于整体注意力的反投影网络模型具有较好的重建效果。

**[关键词]** 超分辨率; 反投影; 整体注意力; 层注意力; 通道空间注意力

**[中图分类号]** TP 391.41; TP 18

## Image Super-Resolution Back-Projection Network Based on Holistic Attention

QI Tangyao, ZHANG Dongxiao

(School of Science, Jimei University, Xiamen 361021, China)

**Abstract:** Aiming at the problem of image super-resolution reconstruction, this paper proposes a back-projection network based on the overall attention. The up-and-down iterative projection module and the overall attention module in the network work together to capture more feature information. Through the mutual conversion of the up and down iterative projection modules, the correlation between different resolution features is obtained, which helps to learn high-frequency texture details. In this paper, the layer attention module is used to weight the attention between the layers of the feature layers output by the upper and lower projections to find the interdependent relationship between different layers, channels and positions. In addition, the spatial channel attention module is used for the final up-projection module, which can guide the network to learn relevant information within and between channels. Ablation experiments show that the enhancement of feature information by the two attention modules helps to improve the reconstruction effect. Comparative experiments show that the back-projection network model based on holistic attention proposed in this paper has a good reconstruction effect.

**Keywords:** super-resolution; back-projection; holistic attention; layer attention; channel spatial attention

**[收稿日期]** 2022-07-21

**[基金项目]** 福建省自然科学基金项目“基于深度过完备字典的大倍数超分辨率重建研究”(2020J01710); 集美大学国家基金培育计划项目“基于深度过完备字典和快速稀疏分解的单帧图像超分辨率重建研究”(ZP2020063)

**[作者简介]** 通信作者: 张东晓(1982—), 副教授, 从事视频与图像处理、机器学习方向研究。E-mail: 200661000115@jmu.edu.cn

<http://xuebaobangong.jmu.edu.cn/zkb>

## 0 引言

高分辨率 (high resolution, HR) 图像通常包含更清晰的边缘结构以及更丰富的纹理细节, 在娱乐、安全、医疗等领域有广泛的需求。但是在采集、传输和储存等过程中, 往往会受设备、环境等因素影响, 图像分辨率会降低。图像超分辨率 (super resolution, SR) 技术正是瞄着这一问题, 通过特定算法由低分辨率 (low resolution, LR) 图像重建出 HR 图像, 进而提升分辨率。目前, SR 已成为计算机视觉和图像处理中的研究热点, 同时该技术有广阔的应用前景, 已经应用在医学<sup>[1]</sup>、遥感<sup>[2]</sup>和公共安全<sup>[3]</sup>等多个领域。

由于卷积神经网络 (convolutional neural network, CNN)<sup>[4]</sup>具有强大的特征自动表征能力, 目前, 基于 CNN 的重建方法<sup>[5]</sup>往往能够学习到 LR 图像到 HR 图像之间复杂的映射关系, 重建效果已经全面超越经典算法。例如, Shi 等<sup>[6]</sup>提出亚像素卷积神经网络 (efficient Sub-Pixel convolutional neural network, ESPCNN), 通过使用亚像素卷积替换反卷积的方式对图像进行上采样操作, 降低了计算成本。Kim 等<sup>[7]</sup>将残差结构引入模型, 提出了 VDSR (very deep super-resolution) 网络, 且将网络加深到 20 层, 同时加入残差学习和梯度裁剪, 解决了由网络加深带来的收敛速度慢和梯度弥散等问题。但目前基于 CNN 的多数重建方法往往不考虑中间特征层的相关性, 而且在图像上采样阶段通常只使用单一的前馈方式。这会使纹理细节在重建过程中趋于平滑, 同时也会削弱 LR 和 HR 之间的联系, 致使模型难以学到从 LR 到 HR 的精准映射。此外, 以往的网络只是将不同卷积层提取出的特征进行简单的串联合并, 没有强调不同层之间的相关性。而一些浅层的特征层信息在深度网络结构的后期影响通常会被削弱, 虽然加入一个长的跳跃连接可以将浅层信息直接传递到深层, 但长跳跃连接平等对待浅层特征和深层特征, 这并不太合理。

为了解决上述问题, 本文引入通道注意力机制<sup>[8-9]</sup>, 同时借鉴反投影网络<sup>[10]</sup>中交替投影的思想, 提出了整体注意力反投影 (holistic attention back projection, HABP) 网络。该网络包含三个方面的创新。首先, 在重建过程中, 交替使用上采样算子和下采样算子, 调整特征的空间维度, 以寻找 HR 图像和 LR 图像在相互转化时的内在联系, 同时将高频信息向前反馈以细化低频信息, 从而增强 LR 图像和 HR 图像之间的联系; 使用这种误差反馈机制指导重建, 可以重建出更好的结果。其次, 从整体注意力网络<sup>[11]</sup>中引入了层注意力模块 (layer attention module, LAM), 以加权的方式合并多尺度特征, 实现不同特征层之间的信息融合。最后, 为了关注每个特征层中不同位置的重要性, 引入通道空间注意力模块 (channel-spatial attention module, CSAM), 用来协同增强每个特征层的高频信息。

## 1 整体注意力反投影网络设计思路

在众多重建算法中, 迭代投影方法是由 Haris 等<sup>[12]</sup>提出, 其主要结构是深度反向投影网络 (deep back-projection networks, DBPN)。虽然该网络在大倍数因子的超分辨率重建任务中获得良好的效果, 但没有考虑到不同特征层对结果的影响。Hu 等<sup>[13]</sup>在 2018 年首次提出通道注意力机制的概念, 对特征通道之间相互依赖的联系进行显式建模, 从而学习到每个特征通道的权重, 以此来提升网络的特征表达能力。然而, 基于注意力机制的方法通常只考虑通道和空间的相关性, 而忽略了多尺度层之间的相互依赖性。本文设计了一种通道注意力与上下迭代投影<sup>[10]</sup>相互融合的结构, 使得网络能够充分利用层次结构之间非线性特征交互特点, 同时细化低频信息, 从而在大倍数因子的超分辨率恢复任务中取得提升效果。

## 2 研究方法

针对超分辨率重建问题, 本文提出的 HABP 网络结构如图 1 所示。网络的核心组件包括特征提取模块、上下投影块模块、整体注意力模块和重建模块。图像首先经过特征提取模块提取特征, 然后通

过上下投影模块在学习高低分辨率图像相互转化间残差的同时提取出不同分辨率的次级特征,接着分别对次级特征进行注意力操作,最后将输出图像合并重建。

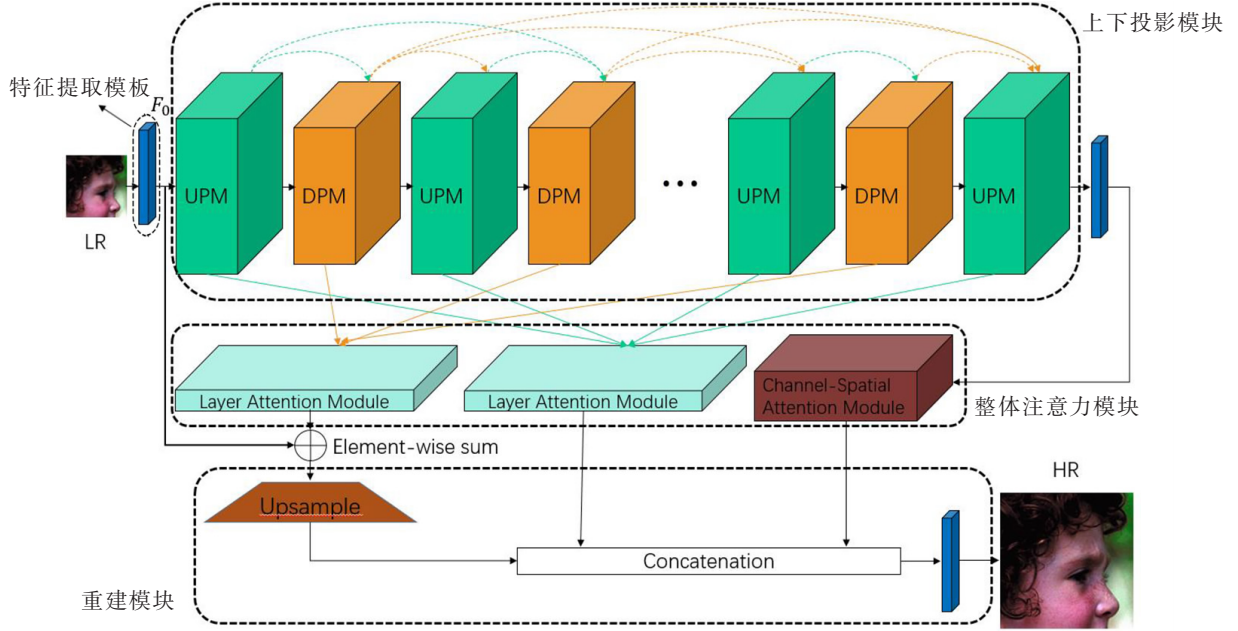


图 1 HABP 网络框架示意图

Fig.1 Schematic diagram of HABP network framework

## 2.1 特征提取模块

针对输入的 LR 图像  $I_{LR}$ , 使用一层卷积层提取  $I_{LR}$  的浅层特征  $F_0$ , 公式为  $F_0 = \text{conv}(I_{LR})$ 。

## 2.2 上下投影模块

在经过特征提取模块得到浅层特征  $F_0$  后, 将  $F_0$  输入到交替出现的上投影模块 (up projection module, UPM) 和下投影模块 (down projection module, DPM), 以此来进一步提取它的次级特征  $F_i$ , 公式为  $\begin{cases} F_i = \text{UPM}_i(F_{i-1}), i = 1, 3, \dots, 2N + 1, \\ F_i = \text{DPM}_i(F_{i-1}), i = 2, 4, \dots, 2N. \end{cases}$  其中:  $\text{UPM}_i$  和  $\text{DPM}_i$  分别代表第  $i$  层上投影模块和下投影

模块;  $2N + 1$  表示上、下投影模块的总量。这里  $F_{2N+1}$  是上下迭代投影模块得到的最终特征, 而  $F_1, F_2, \dots, F_{2N}$  都是该网络的中间特征;  $F_1, F_3, \dots, F_{2N-1}$  是由上投影模块得到的特征, 而  $F_2, F_4, \dots, F_{2N}$  是下投影模块得到的特征。通过  $\text{UPM}_i$  和  $\text{DPM}_i$  的依次堆叠, 使得模型在迭代过程中学习图像的残差并反馈给网络, 不断优化网络。

### 2.2.1 UPM 结构

由于卷积可以降低特征的空间维度, 而反卷积可以提升特征的空间维度。所以在上下投影模块中, 借鉴 DBPN<sup>[12]</sup> 的思想, 本文交替使用卷积和反卷积来实现特征空间维度的提升 (降低)。但是与 DBPN 的结构有所不同, 本文在每一次降低或提升特征空间维度 (卷积或反卷积) 时, 均引入注意力机制, 目的是强化重要特征通道, 同时抑制非重要特征通道。为了叙述方便, 这里将它们分别称为卷积通道注意块 (convolution channel attention block, CCAB) 和反卷积通道注意块 (deconvolution channel attention block, DCAB)。这两个块的结构如图 2 所示。由于针对卷积和反卷积的注意力结构是相同的, 所以这里只给出统一的结构图。

如图 2 所示, 在降低 (或提升) 特征空间维度时, 先对输入的特征进行卷积 (或反卷积), 得到通道数为  $C$  的特征, 然后使用全局最大池化, 得到  $1 \times 1 \times C$  的特征。随后将其输入到一个三层

神经网络中, 该网络的输入层有  $C$  个神经元, 隐藏层有  $C/r$  个神经元, 输出层有  $C$  个神经元, 其中  $r$  是一个超参数, 用来调节隐藏层的神经元个数。隐藏层的激活函数为 Relu, 输出层的激活函数为 Sigmoid。之所以使用 Sigmoid, 是为了对三层神经网络的输出做归一化处理, 从而得到取值范围在  $0 \sim 1$  之间的权重系数。最后将这个权重系数与经过卷积层 (反卷积层) 的特征相乘, 实现对降维 (或升维) 特征的通道加权, 最后将加权特征叠加到卷积 (或反卷积) 特征上, 得到信息增强后的新特征。

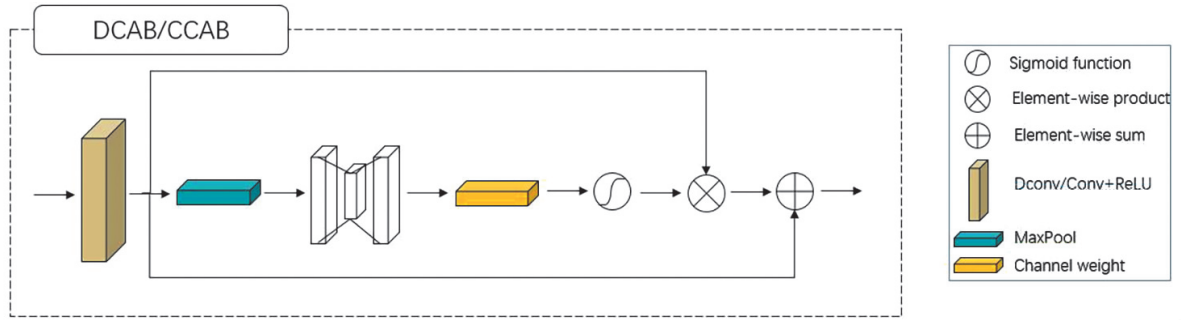


图 2 卷积 (反卷积) 通道注意块示意图

Fig.2 Convolution (deconvolution) channel attention block schematic

整个重建的目的是提升图像分辨率, 为此需要重建出更多的高分辨率特征。如果只使用 CCAB 提升特征的空间维度, 无法确保高分辨率特征与低分辨率特征之间很好的对应关系。为此, 本文借鉴 DBPN 的思想, 交替使用 DCAB 和 CCAB, 以使特征在高、低分辨率之间相互转换, 进而学习到高、低分辨率特征之间更好的对应关系。这种提升特征空间分辨率的 UPM 的结构如图 3 所示。与 DBPN 的结构有所不同, 本文引入了一个  $1 \times 1$  卷积来自由训练参数, 对全局和局部进行加权, 这样可以缓解深度网络训练的难度, 使高频信息从输入到输出自由流动, 有助于达到更好的效果。

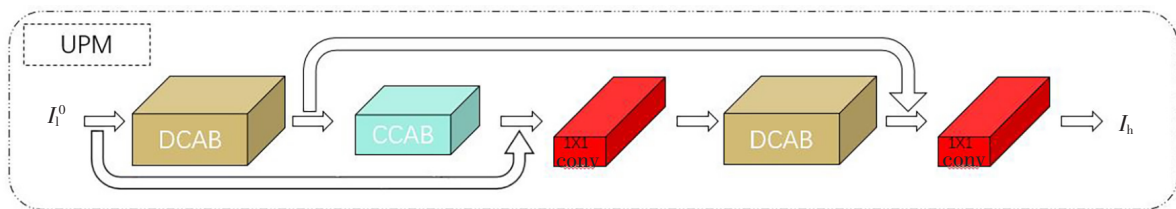


图 3 UPM 的结构示意图

Fig.3 Schematic diagram of the structure of UPM

UPM 可表示为  $I_h = C_{1 \times 1} \{ \mathcal{D}[C_{1 \times 1}(\mathcal{C}[\mathcal{D}(I_1^0)], I_1^0)], \mathcal{D}(I_1^0) \}$ 。其中:  $\mathcal{D}$  表示反卷积通道注意块运算;  $\mathcal{C}$  表示卷积通道注意块运算;  $C_{1 \times 1}(\cdot, \cdot)$  表示对输入的两组特征先在通道维度串联, 然后做  $1 \times 1$  卷积。

UPM 通过上下交替的方式实现特征上采样。将低分辨率特征  $I_1^0$  输入该模块后,  $I_1^0$  通过 DCAB 提升了维度, 得到的高分辨率特征图记为  $I_h^1$ 。然后,  $I_h^1$  通过 CCAB 又得到了新的低分辨率特征, 将其记为  $I_1^1$ 。接着, 将两个低分辨率特征  $I_1^0$  和  $I_1^1$  沿通道串联, 并通过  $1 \times 1$  卷积得到重新整合的特征, 将其记为  $I_1^{\text{con}}$ 。随后, 将  $I_1^{\text{con}}$  输入 DCAB 提升特征维度, 在这里得到的高分辨率特征记作  $I_h^2$ 。最后, 将两次反卷积得到的特征  $I_h^1$  和  $I_h^2$  串联再次经过一个  $1 \times 1$  卷积, 得到最后的高分辨率特征  $I_h$ 。

### 2.2.2 DPM 结构

与上投影模块中上下交替的思想一样, 本文进一步使用交替投影的方式来重建高分辨率特征。因



此, DPM 结构与 UPM 结构类似, 如图 4 所示。DPM 公式为  $I_l = C_{1 \times 1} \{ \mathcal{C}[C_{1 \times 1}(\mathcal{D}[\mathcal{C}(I_h^0)], I_h^0)], \mathcal{C}(I_h^0) \}$ 。

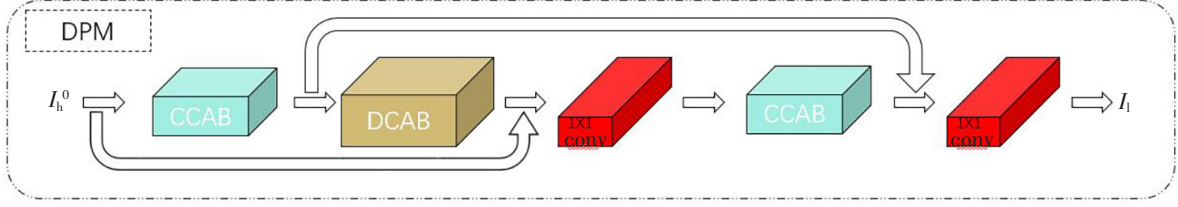


图 4 DPM 的结构示意图

Fig.4 Schematic diagram of the structure of DPM

### 2.3 整体注意力模块

经过上下投影模块后输出的中间特征  $F_1, F_2, \dots, F_{2N}$  以及最终特征  $F_{2N+1}$  分类输入整体注意力模块, 这个模块包含 LAM 和 CSAM。LAM 公式为  $\begin{cases} F_{LA}^H = \text{LAM}(\text{concatenate}(F_1, F_3, \dots, F_{2N+1})) \\ F_{LA}^L = \text{LAM}(\text{concatenate}(F_2, F_4, \dots, F_{2N})) \end{cases}$ , 其中: LAM 表示层注意力模块; concatenate ( $\cdot$ ) 表示将特征按通道拼接。

整个 LAM 的过程为: 首先将 UPM 得到的特征  $F_1, F_3, \dots, F_{2N-1}, F_{2N+1}$  沿着通道维度串联起来, 然后将串联的结果输入到 LAM, 得到的新特征记为  $F_{LA}^H$ 。类似地将 DPM 得到的特征  $F_2, F_4, \dots, F_{2N}$  串联后输入 LAM, 得到的结果记为  $F_{LA}^L$ 。这里, LAM 的作用是增强贡献大的特征层, 同时抑制贡献小的特征层。

另一方面, 本文使用 CSAM 自适应地捕捉通道间和通道内部更多重要的信息来调整特征, 其定义为  $F_{CS} = \text{CSM}(F_{2N+1})$ 。其中: CSM 表示通道空间注意模块;  $F_{CS}$  表示通道空间注意模块输出的特征。为了减少计算量, 本文只对最后一个特征层做通道空间注意。

#### 2.3.1 LAM 结构

如图 1 所示, 当特征交替通过上下投影模块后, 为了充分利用中间模块输出的特征, 这里分别合并 UPM 和 DPM 输出的中间特征。中间特征对重建结果起到关键作用, 但是它们的影响应该有所不同。为了强化重要特征而抑制非重要特征, 这里使用图 5 所示的 LAM 学习所有特征层的权重, 以期增大重要特征层的权重, 减小非重要特征层的权重。

如图 5 所示, 该模块的输入可以通过 UPM 得到的高分辨率特征, 也可以是 DPM 得到的低分辨率特征。为了叙述方便, 用  $F_{G_l}$  表示所有 DPM 输出特征串联合并后的特征组, 用  $F_{G_h}$  表示所有 UPM 输出特征串联合并后得到的特征组。

这里以下投影特征组  $F_{G_l}$  为例来介绍 LAM。将  $F_{G_l}$  的维度记为  $N \times H \times W \times C$ , 这里  $N$  表示特征组的数量,  $H$  表示特征的高度,  $W$  表示特征的宽度,  $C$  表示特征的通道数。

为了得到不同层的权重系数, 这里需要计算相关性矩阵。首先, 将特征组  $F_{G_l}$  重塑成一个维数为  $N \times HWC$  的二维矩阵, 将矩阵与该矩阵的转置相乘, 得到层注意力相关矩阵  $w_{i,j} = \sigma(\mu(F_{G_l})_i \cdot [\mu(F_{G_l})]_j^T), i, j = 1, 2, \dots, N$ 。其中:  $\mu(\cdot)$  表示调整特征维度的操作;  $\sigma(\cdot)$  表示 softmax 函数;  $\cdot$  是向量内积;  $\mu(F_{G_l})_i$  表示特征组  $\mu(F_{G_l})$  的第  $i$  行;  $[\mu(F_{G_l})]_j^T$  表示  $\mu(F_{G_l})$  转置的第  $j$  列。

最后, 把得到的相关性矩阵与特征组  $\mu(F_{G_l})$  相乘, 并且乘以一个比例因子  $\alpha$ , 重塑为原尺寸后再与特征组  $F_{G_l}$  相加, 得到 LAM 最终的输出结果。该过程为  $F_{LA}^L = \alpha \mu^{-1}[w_{i,j} \times \mu(F_{G_l})] + F_{G_l}$ 。其中:  $\mu(\cdot)$  和  $\mu^{-1}(\cdot)$  分别表示改变特征尺寸的运算和它的逆运算;  $\times$  表示矩阵的乘法;  $\alpha$  是待学习参数, 其初始值为 0。

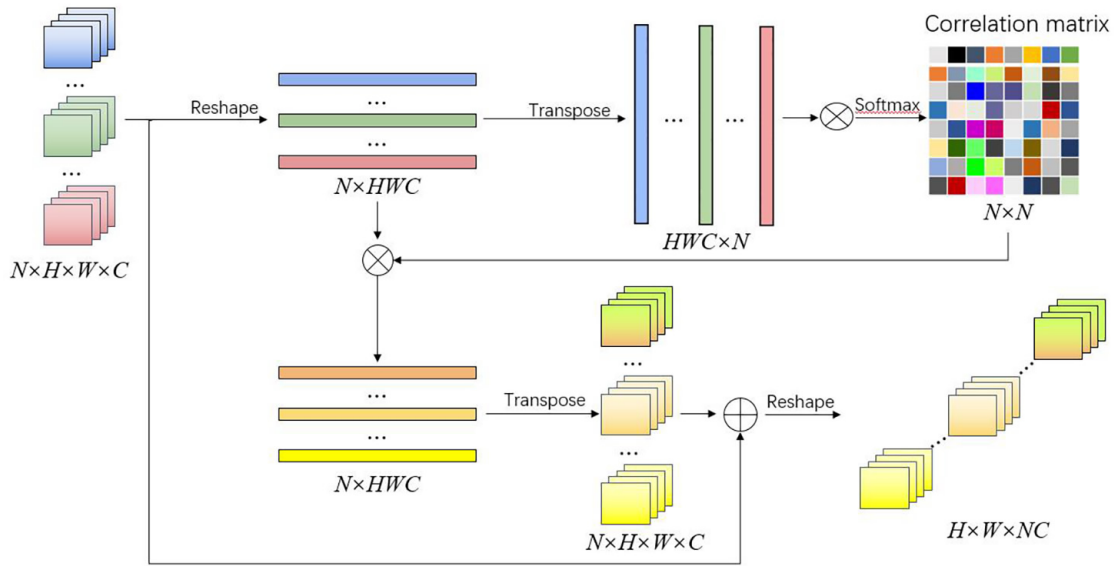


图 5 LAM 的结构示意图

Fig.5 Schematic diagram of the structure of LAM

由上叙述可看出, LAM 通过特征间的加权, 加大了特征之间的关联性, 同时也增强了贡献较高的特征层。

### 2.3.2 CSAM 结构

目前, 现有的空间注意力机制主要关注特征的空间维数 (宽高), 却忽视了通道信息。而最新的通道注意力机制<sup>[8]</sup>关注特征的通道维数, 却忽略了特征的空间信息。通道注意力机制重点关注哪一层的特征是重要的, 而空间注意力机制则强调特征的哪一块区域是重要的。因此本文应用了一种最新的 CSAM, 该模块能够自适应地学习通道间和空间上各个维度的特征信息, 结构如图 6 所示。

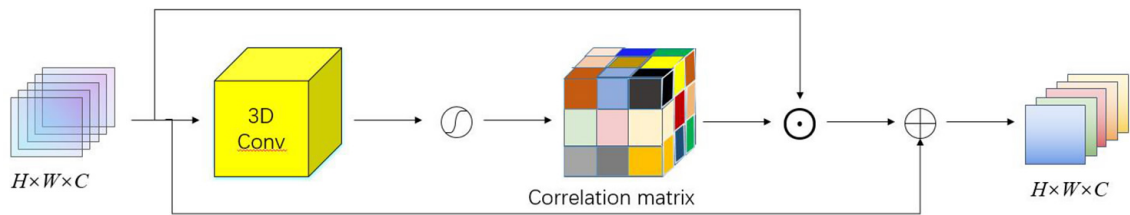


图 6 CSAM 的结构示意图

Fig.6 Schematic diagram of the structure of CSAM

由图 1 可知, 上下投影模型紧临重建模块, 所以上下投影模块最后一层的输出对重建结果有直接影响, 考虑到计算成本, 本文只对上下投影模块的最后一层做通道空间注意操作。如 2.2 节所述, 由上下投影模块最后一层得到的特征为  $F_{2N+1} \in \mathbb{R}^{H \times W \times C}$ , 所以这里只对  $F_{2N+1}$  操作。

如图 6 所示, 首先, 将  $F_{2N+1}$  的三组连续通道分别与一个核为  $3 \times 3 \times 3$ 、步长为 1 的三维卷积层<sup>[14]</sup>做 3D 卷积, 得到三组通道空间注意映射图  $W_{csa}$ 。然后, 将注意映射图  $W_{csa}$  经过 sigmoid 函数得到通道空间注意相关性矩阵 (correlation matrix)。最后, 将相关性矩阵和输入特征  $F_{2N+1}$  进行逐元素相乘, 并加权乘以比例因子  $\beta$ , 与输入特征  $F_{2N+1}$  求和, 得到  $F_{2N+1}$  的通道空间注意结果, 该过程表示为  $F_{cs} = \beta \varphi(W_{csa}) \odot F_{2N+1} + F_{2N+1}$ 。其中:  $\varphi(\cdot)$  代表 sigmoid 函数;  $\odot$  代表逐元素乘法;  $\beta$  是待学习参数, 其初始值为 0。与原始特征相比,  $F_{cs}$  额外包含了所有通道空间位置的加权信息。与传统的空

间注意和通道注意相比,CSAM同时考虑了通道维数和空间维数上特征信息之间的相互依赖性,使得网络能够自适应地学习通道内部和通道间的相关信息。

## 2.4 重建模块

整个网络的最后是图像的重建模块,该模块使用亚像素卷积<sup>[8]</sup>对所得特征上采样。亚像素卷积通过卷积和多通道间的重组得到高分辨率的特征图,它将多个特征通道上的单个像素组合在一个特征上,以使每个特征上的像素作为新特征的亚像素。具体表示为

$$F_U^H = U_{\uparrow}(\text{concatenate}(F_0, F_{LA}^L))。 \quad (1)$$

其中: $U_{\uparrow}$ 表示亚像素卷积。特别需要说明的是,由图1可知,在式(1)中引入了一个长跳跃连接来确保该网络在训练过程中的稳定性。这里为了强调浅层特征和深层特征的不同作用,分别在两个特征前乘以可训练系数权重,来调节两个特征对结果的作用。

最后,将 $F_U^H$ 、 $F_{LA}^H$ 、 $F_{CS}$ 进行串联,然后使用一层卷积层得到超分辨率重建结果,公式为 $I_{SR} = \text{conv}(\text{concatenate}(F_U^H, F_{LA}^H, F_{CS}))$ 。

## 2.5 损失函数的设计

除了上述提到网络模型的结构外,如何设计损失函数也是机器学习的重要问题,在基于深度学习的超分辨率重建任务中,损失函数的选择也至关重要。

现有的大多数方法选用 $L_1$ 或者 $L_2$ 损失函数,由于 $L_2$ 损失函数无法捕捉到HR图像的潜在多模态分布,这往往会导致重建后的HR图像过于平滑。相较于 $L_2$ 损失函数, $L_1$ 损失函数具有更强的鲁棒性,在一定程度上可以去除斑点和伪影,使得图像更加接近人类的视觉感受,但其重建时间往往相对较长,且会出现最优解不稳定的情况。因此,本文选用Charbonnier损失函数<sup>[15]</sup>,该损失函数同时解决了 $L_1$ 损失函数在零点不可导和 $L_2$ 损失函数对于重建图像过于平滑的问题,具体公式为 $L_{SR} = \frac{1}{HWC} \sum_{i,j,k}^{H,W,C} \sqrt{(\text{HABP}(\text{LR})_{i,j,k} - \text{HR}_{i,j,k})^2 + \varepsilon^2}$ 。其中: $\text{HABP}(\cdot)$ 表示本文所提重建网络;LR表示输入到重建网络的低分辨率图像;HR表示与低分辨率图像对应的高分辨率参考图像; $H$ 、 $W$ 、 $C$ 分别代表重建图像的高度、宽度和通道数; $\text{HR}_{i,j,k}$ 表示高分辨率参考图像HR在 $(i,j,k)$ 处的颜色值;常量 $\varepsilon$ 是超参数,一般取 $10^{-3}$ 。

# 3 实验结果及分析

## 3.1 实验数据集

本文使用DIV2K数据集<sup>[11]</sup>来训练网络,选取常用的5个标准参考数据集作为测试集,即Set5、Set14、B100、Urban100和Manga109。本文通过数据增强增加训练集样本的数量,增强方法包括旋转( $90^\circ$ 、 $180^\circ$ 和 $270^\circ$ )、比例缩放(一般是0.6~0.9)、垂直和水平翻转。

## 3.2 模型超参数设置

卷积参数设置:在上下投影模块中,本文在2倍数下使用的卷积核大小为6,步长设置为2,padding设置为2;在3倍数下使用的卷积核大小为7,步长设置为3,padding设置为2;在4倍数下使用的卷积核大小为8,步长设置为4,padding设置为2;在其他模块中的卷积核大小均为3,步长设置均为1。并且除特征提取模块和最后的重建模块以外,所有的卷积和反卷积层后面都伴随一个ReLU激活函数。

上下投影模块参数设置:采用9个DPM和10个UPM。

优化器参数设置:采用Adam优化器,其中初始学习率为 $10^{-4}$ 且学习率为固定步长衰减;动量 $\beta_1$ 和 $\beta_2$ 经过测试,分别设置为0.9和0.999,且 $\varepsilon = 10^{-3}$ 。

## 3.3 实验结果

本节挑选10种典型重建方法在5个测试集上与本文的HABP模型比较,这些方法是Bicubic、

SRCNN<sup>[16]</sup>、ESPCN<sup>[6]</sup>、FSRCNN<sup>[17]</sup>、VDSR<sup>[9]</sup>、LapSRN<sup>[14]</sup>、MemNet<sup>[18]</sup>、SRMDNF<sup>[19]</sup>、IMDN<sup>[20]</sup> 和 MAFFSRN<sup>[21]</sup>。其中: Bicubic 是双三次线性插值, IMDM 和 MAFFSRN 是当前最新的模型, 其余的是比较经典的超分模型。表 1 给出了上述重建模型与本文提出的 HABP 模型在 2 倍、3 倍和 4 倍放大因子上的峰值信噪比 (peak signal-to-noise ratio, PSNR) 的值和结构相似性 (structural similarity, SSIM) 的值, 加粗的值表示最高值, 带有下划线的值表示第二高值。其中, 表格中的数据是在对应测试集上计算的平均指标。

表 1 双三次降采样下测试集上的定量结果

Tab. 1 Quantitative results on the test set under bicubic downsampling

尺寸	方法	Set5		Set14		Urban100		Manga109		B100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
2	Bicubic	33.66	0.930	30.24	0.869	26.88	0.840	30.80	0.934	29.56	0.843
	SRCNN	36.66	0.954	32.45	0.907	29.50	0.895	35.60	0.966	31.36	0.888
	ESPCN	37.00	0.956	32.75	0.910	29.87	0.907	36.21	0.969	31.51	0.894
	FSRCNN	37.05	0.956	32.66	0.909	29.88	0.902	36.67	0.971	31.53	0.892
	VDSR	37.53	0.959	33.05	0.913	30.77	0.914	37.22	0.975	31.90	0.896
	LapSRN	37.52	0.959	33.08	0.913	30.41	0.910	37.27	0.974	31.08	0.895
	MemNet	37.78	0.960	33.28	0.914	31.31	0.920	37.72	0.974	32.08	0.898
	SRMDNF	37.79	0.960	33.32	0.916	31.33	0.920	38.07	0.976	32.05	0.896
	IMDN	<u>38.00</u>	<u>0.961</u>	<b>33.63</b>	<b>0.918</b>	<u>32.17</u>	<u>0.928</u>	<b>38.88</b>	<b>0.977</b>	<u>32.19</u>	<u>0.900</u>
	MAFFSRN	<b>38.07</b>	<b>0.961</b>	<u>33.59</u>	<u>0.918</u>	<b>32.38</b>	<b>0.931</b>			<b>32.23</b>	<b>0.901</b>
3	HABP	37.82	0.960	33.56	0.917	31.92	0.926	<u>38.23</u>	<u>0.976</u>	32.12	0.899
	Bicubic	30.39	0.868	27.55	0.774	24.46	0.735	26.95	0.856	27.21	0.739
	SRCNN	32.75	0.909	29.30	0.822	26.24	0.799	30.48	0.912	28.41	0.786
	ESPCN	33.02	0.914	29.49	0.827	26.41	0.816	30.79	0.918	28.50	0.794
	FSRCNN	33.18	0.914	29.37	0.824	26.43	0.808	31.10	0.921	28.53	0.791
	VDSR	33.67	0.921	29.78	0.832	27.14	0.829	32.01	0.934	28.83	0.799
	LapSRN	33.82	0.923	29.87	0.832	27.07	0.828	32.21	0.935	28.82	0.798
	MemNet	34.09	0.925	30.00	0.835	27.56	0.838	32.51	0.937	28.96	0.800
	SRMDNF	34.12	0.925	30.04	0.838	27.57	0.840	33.00	0.940	28.97	0.803
	IMDN	34.36	0.927	30.32	0.842	28.17	0.852	<b>33.61</b>	<b>0.945</b>	29.09	0.805
4	MAFFSRN	<u>34.45</u>	<u>0.928</u>	<u>30.40</u>	<u>0.843</u>	<b>28.26</b>	<b>0.855</b>			<u>29.13</u>	<u>0.806</u>
	HABP	<b>34.46</b>	<b>0.930</b>	<b>30.42</b>	<b>0.845</b>	<u>28.24</u>	<u>0.854</u>	<u>33.49</u>	<u>0.944</u>	<b>29.14</b>	<b>0.807</b>
	Bicubic	28.42	0.810	26.10	0.704	23.15	0.659	24.92	0.789	25.96	0.669
	SRCNN	30.49	0.862	27.61	0.754	24.53	0.724	27.66	0.858	26.91	0.712
	ESPCN	30.66	0.865	27.71	0.756	24.60	0.736	27.70	0.856	26.98	0.712
	FSRCNN	30.72	0.866	27.61	0.755	24.62	0.728	27.90	0.861	26.98	0.715
	VDSR	31.35	0.882	28.03	0.770	25.18	0.753	28.82	0.886	27.29	0.726
	LapSRN	31.54	0.885	28.19	0.772	25.21	0.756	29.09	0.890	27.32	0.728
	MemNet	31.74	0.889	28.26	0.772	25.50	0.763	29.42	0.894	27.40	0.728
	SRMDNF	31.96	0.893	28.35	0.779	25.68	0.773	30.09	0.902	27.49	0.734
4	IMDN	<u>32.21</u>	<u>0.895</u>	28.58	0.781	26.04	0.784	<u>30.45</u>	<u>0.908</u>	27.56	0.735
	MAFFSRN	32.20	0.895	<u>28.62</u>	<u>0.782</u>	<b>26.16</b>	<b>0.789</b>			<u>27.59</u>	<u>0.737</u>
	HABP	<b>32.22</b>	<b>0.896</b>	<b>28.63</b>	<b>0.784</b>	<u>26.13</u>	<u>0.787</u>	<b>30.48</b>	<b>0.909</b>	<b>27.63</b>	<b>0.740</b>



表1显示了11种重建模型在2倍、3倍和4倍尺度下的重建结果。在4倍重建的情况下,本文的HABP的重建指标总体上领先其他模型。从PSNR指标来看,在测试集B100上比第二高出0.04 dB,在测试集Manga109上高出0.03 dB,在Set5和Set14数据集上均高出第二名0.01 dB;不过在Urban100测试集上的效果没有达到最优,排名第二。在3倍重建的情形中,本文方法的指标在Set5、Set14和B100这三个数据集上达到最优,PSNR指标比第二分别高出0.01和0.02 dB;不过在Urban100和Manga109这两个数据集上的效果略差,排在第二,但PSNR仍超过第三名0.07 dB。分析2倍重建的指标,HABP的效果略低于最新的模型,整体排名第三。

对比2倍、3倍和4倍的重建效果,可以发现本文方法在较大倍数的重建任务中,比现有算法有较为明显的优势,说明本文使用的交替上下投影在较大倍数的重建中确实能够学习到更多的信息。

对比各数据集上的效果,可以发现本文方法在Urban100上的重建效果不及现有方法。详细考察Urban100,可以发现这个数据集中建筑物类型的图像居多,这也间接反映了在4倍重建时本文模型可能不适合处理建筑物外观类型的数据,而在重建人物、动物和风景上较有优势。

图7~9为2倍重建任务的视觉效果对比结果,图中只展示了三种典型模型的视觉结果。从图中鸟喙周围的构造,动画人物中人眼和帽子上银网链的款式以及考拉眼睛周围的纹理的重建效果,可以看出本文提出的HABP模型能够获得更清晰的细节构造和低频纹理。



图7 Set5中的bird图像2倍重建结果及PSNR、SSIM值

Fig.7 The results of the 2 times reconstruction of the bird image in Set5, with PSNR and SSIM values

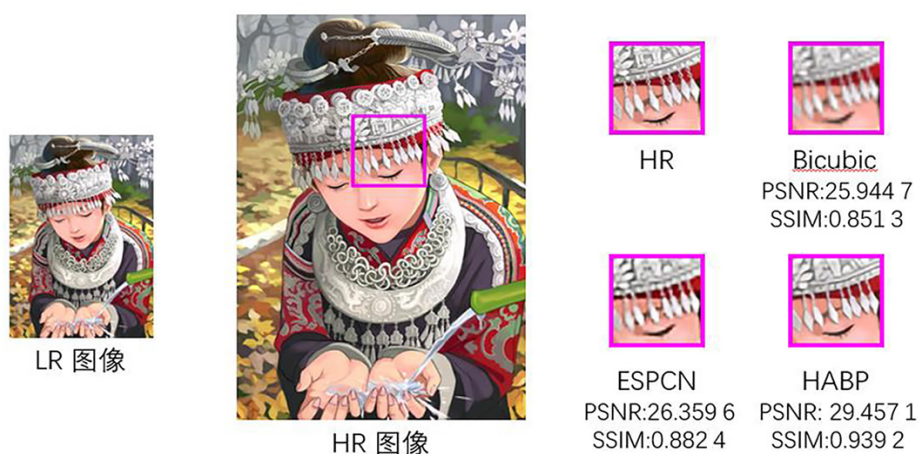


图8 Set14中的comic图像2倍重建结果及PSNR、SSIM值

Fig.8 The results of the 2 times reconstruction of the comic image in Set14, with the PSNR and SSIM values



图 9 B100 中的 69015 图像 2 倍重建结果及 PSNR、SSIM 值

Fig.9 The results of the 2 times reconstruction of 69015 image in B100, with PSNR and SSIM values

图 10~12 展示了 3 倍放大因子的重建效果，通过对图像裁剪区域的比较能够发现，本文的模型通过适当利用特征空间中的通道和空间注意力来恢复低分辨率图像中缺失的结构化细节，有效地抑制了伪影。



图 10 Set5 中的 woman 图像 3 倍重建结果及 PSNR、SSIM 值

Fig.10 The result of the 3 times reconstruction of the woman image in Set5, with PSNR and SSIM values



图 11 Set14 中的 baboon 图像 3 倍重建结果及 PSNR、SSIM 值

Fig.11 The result of the 3 times reconstruction of the baboon image in Set14, with PSNR and SSIM values



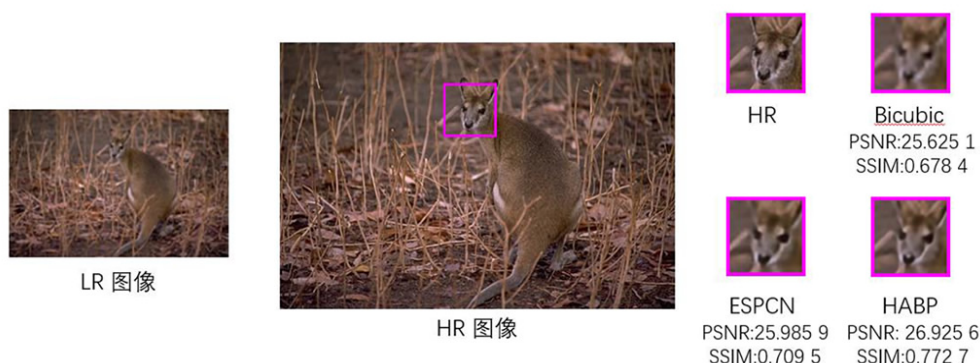


图 12 B100 中的 69040 图像 3 倍重建结果及 PSNR、SSIM 值

Fig.12 The result of the 3 times reconstruction of 69040 image in B100, with PSNR and SSIM values

图 13~15 展示了 4 倍因子的放大效果, 通过观察重建图像的细节部分, 可以发现本文在 4 倍的模型重建出了更高质量的图像。如图 13 的 head 图像中眼皮和睫毛的纹理结构更清晰, 图 14 的 monarch 图像中蝴蝶翅膀处的花纹更光滑, 图 15 的 013 图像中建筑外观的网格构造也更清楚。

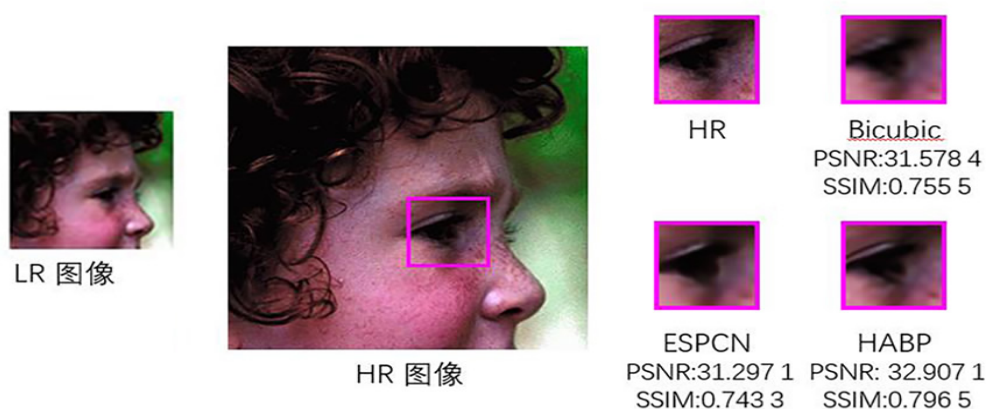


图 13 Set5 中的 head 图像 4 倍重建结果及 PSNR、SSIM 值

Fig.13 The result of the 4 times reconstruction of the head image in Set5, with PSNR and SSIM values



图 14 Set14 中的 monarch 图像 4 倍重建结果及 PSNR、SSIM 值

Fig.14 The result of the 4 times reconstruction of the monarch image in Set14, with PSNR and SSIM values



图 15 Urban100 中的 013 图像 4 倍重建结果及 PSNR、SSIM 值

Fig.15 The result of the 4 times reconstruction results of 013 image in Urban100, with PSNR and SSIM values

### 3.4 LAM 和 CSAM 的消融实验

为了验证 LAM 和 CSAM 这两种注意机制对特征的作用, 本节在数据集 B100 上进行消融实验, 其中放大因子为 4。在 HABP 的网络中, 将去掉 LAM 和 CSAM 的模型作为基线方法, 得到 PSNR 为 27.39 dB。然后比较去掉上投影输出后的 LAM (PSNR 为 27.48 dB)、去掉下投影输出后的 LAM (PSNR 为 27.50 dB)、去掉所有 LAM (PSNR 为 27.45 dB), 以及去掉 CSAM (PSNR 为 27.53 dB) 时的效果。经过删减不同模块的实验, HABP 网络的 PSNR 能达到 27.57 dB。

由以上结果可知, LAM 对网络的影响是较大的, 本文网络的 PSNR 与基线方法相比提升能达到 0.18 dB。其中, LAM 在测试数据集上提高了 0.14 dB, 而 CSAM 在测试数据集上提高了 0.06 dB。消融实验表明所提出的整体注意力提升了 HABP 网络的效果。

### 3.5 上下投影模块数量的消融实验

为了验证投影块的有效性这里考虑投影模块数量的影响。为此考虑 3 种模型的效果: 6 个 UPM + 5 个 DPM (U6D5), 8 个 UPM + 7 个 DPM (U8D7) 和 10 个 UPM + 9 个 DPM (U10D9)。这里只考虑了放大因子为 4 倍的情况, 3 种模型在 5 个测试集上的 PSNR 如表 2 所示。分析表 2, 可以发现随着投影块数量的增加, 结果也有小幅度的提升, 考虑到计算量与运行速度对实验带来的影响和实验结果带来的提升效果, 最终网络选择了 10 个 UPM 和 9 个 DPM 的网络结构。

表 2 不同上下投影模块数量消融实验的 PSNR

	dB				
模型	Set5	Set14	Urban100	Manga109	B100
U6D5	31.88	28.44	25.75	29.94	27.44
U8D7	32.04	28.48	25.86	30.14	27.47
U10D9	32.15	28.57	26.06	30.39	27.57

## 4 结论

本文提出的 HABP 网络改进了上下迭代投影模块, 以此更好地学习高低分辨率特征在上下迭代投影过程中所生成的高频纹理细节。通过密集连接的方式, 使信息在网络中高效流动。与此同时, 利用整体注意力自适应地学习不同深度、通道和位置之间的全局依赖关系。其中, CSAM 从特征的各个维度生成注意力映射图, 从而得到通道间和通道内部的相关性, 并指导特征中不同维度信息的表达。LAM 通过寻找不同层之间和层内部的相关性, 指导不同层的权重分配, 强化重要的特征通道, 抑制非重要特征通道, 使得网络在不断加深时还能够尽可能地捕捉更多特征信息。从实验数据来看, HABP 网络在 PSNR、SSIM 和视觉上确实有较好的效果。



## [ 参考文献 ]

- [1] 游森榕. 基于生成对抗网络的医学影像重建技术[D]. 深圳:中国科学院大学(中国科学院深圳先进技术研究院),2022.
- [2] 王毓武. 基于超分辨率重建的遥感图像多目标识别算法研究[D]. 哈尔滨:黑龙江大学,2022.
- [3] RASTI P, UIBOUPIN T, ESCALERA S, et al. Convolutional neural network super resolution for face recognition in surveillance monitoring[C]//International Conference on Articulated Motion and Deformable Objects. Palma de Mallorca, Spain: Springer International Publishing,2016:175-184.
- [4] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [J]. Advances in Neural Information Processing Systems,2014,27:3320-3328.
- [5] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA:IEEE Press,2016:1637-1645.
- [6] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA:IEEE Press,2016:1874-1883.
- [7] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//IEEE Conference on Computer Vision & Pattern Recognition(CVPR). Las Vegas, NV, USA:IEEE Press,2016:1646-1654.
- [8] ZHANG Y L, LI K P, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany:Springer,2018:286-301.
- [9] WOO S, PARK J, LEE J Y, et al. CBAM:convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich, Germany:Springer,2018:3-19.
- [10] BAREJA M N, MODI C K. An effective iterative back projection based single image super resolution approach[C]//International Conference on Communication Systems & Network Technologies. Rajkot, Gujarat, India:IEEE Press,2012:95-99.
- [11] NIU B, WEN W L, REN W Q, et al. Single image super-resolution via a holistic attention network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Glasgow, UK:Springer,2020:191-207.
- [12] HARIS M, SHAKHNAROVICH G, UKITA N. Deep back-projection networks for super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, UT, USA:IEEE Press,2018:1664-1673.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE Eonference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA:IEEE Press,2018:7132-7141.
- [14] JI S W, XU W, YANG M, et al. 3d convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,35(1):221-231.
- [15] LAI W S, HUANG J B, AHUJA N, et al. Fast and accurate image super-resolution with deep laplacian pyramid networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,8(4):2599-2613.
- [16] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//Proceedings of the European Conference on Computer Vision (ECCV). Zurich, Switzerland:Springer,2014:184-199.
- [17] DONG C, LOY C C, TANG X O. Accelerating the super-resolution convolutional neural network[C]//14th European Conference. Amsterdam, The Netherlands :Springer,2016:391-407.
- [18] TAI Y, YANG J, LIU X M, et al. Memnet:a persistent memory network for image restoration[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy:IEEE Press,2017:4539-4547.
- [19] ZHANG K, ZUO W M, ZHANG L. Learning a single convolutional super-resolution network for multiple degradations[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA:IEEE Press,2018:3262-3271.
- [20] HUI Z, GAO X B, YANG Y C, et al. Lightweight image super-resolution with information multi-distillation network[C]//Proceedings of the 27th ACM International Conference on Multimedia(ACMMM). New York, NY, United States:Association for Computing Machinery,2019:2024-2032.
- [21] MUQEET A, HWANG J, YANG S, et al. Multi-attention based ultra lightweight image super-resolution[C]//Proceedings of the European Conference on Computer Vision(ECCV). Glasgow, UK:Springer,2020:103-118.

(责任编辑 彭海滨 英文审校 黄振坤)