Opto-Electronic Engineering
光 电 工 程

Article
2022 年, 第 49 卷, 第 5 期

DOI: 10.12086/oee.2022.210382

基于自相似特征增强网络结构的图像超分辨率重建

汪荣贵，雷　辉，杨　娟*，薛丽霞

合肥工业大学计算机与信息学院，安徽 合肥 230601

摘要：深度卷积神经网络最近在图像超分辨率方面展示了高质量的恢复效果。然而，现有的图像超分辨率方法大多只考虑如何充分利用训练集中固有的静态特性，却忽视了低分辨率图像本身的自相似特征。为了解决这些问题，本文设计了一种自相似特征增强的网络结构 (SSEN)。具体来说，本文将可变形卷积嵌入到金字塔结构中并结合跨层次协同注意力，设计出了一个能够充分挖掘多层次自相似特征的模块，即跨层次特征增强模块。此外，本文还在堆叠的密集残差块中引入池化注意力机制，利用条状池化扩大卷积神经网络的感受野并在深层特征中建立远程依赖关系，从而深层特征中相似度较高的部分能够相互补充。在常用的五个基准测试集上进行了大量实验，结果表明，SSEN 比现有的方法在重建效果上具有明显提升。

关键词：超分辨率；自相似性；特征增强；可变形卷积；注意力；条状池化

中图分类号：TP391.4　　**文献标志码：**A

汪荣贵，雷辉，杨娟，等. 基于自相似特征增强网络结构的图像超分辨率重建 [J]. 光电工程，2022，**49**(5): 210382
Wang R G, Lei H, Yang J, et al. Self-similarity enhancement network for image super-resolution[J]. *Opto-Electron Eng*, 2022, **49**(5): 210382

Self-similarity enhancement network for image super-resolution

Wang Ronggui, Lei Hui, Yang Juan*, Xue Lixia

School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China

Abstract: Deep convolutional neural networks (DCNN) recently demonstrated high-quality restoration in the single image super-resolution (SISR). However, most of the existing image super-resolution methods only consider making full use of the inherent static characteristics of the training sets, ignoring the internal self-similarity of low-resolution images. In this paper, a self-similarity enhancement network (SSEN) is proposed to address above-mentioned problems. Specifically, we embedded the deformable convolution into the pyramid structure and combined it with the cross-level co-attention to design a module that can fully mine multi-level self-similarity, namely the cross-level feature enhancement module. In addition, we introduce a pooling attention mechanism into the stacked residual dense blocks, which uses a strip pooling to expand the receptive field of the convolutional neural network and establish remote dependencies within the deep features, so that the patches with high similarity in deep features can complement each other. Extensive experiments on five benchmark datasets have shown that the SSEN has a significant improvement in reconstruction effect compared with the existing methods.

收稿日期：2021−11−26；收到修改稿日期：2022−02−21
基金项目：国家重点研发计划资助项目 (2020YFC1512601)
*通信作者：杨娟，yangjuan6985@163.com。

Keywords: super-resolution; self-similarity; feature enhancement; deformable convolution; attention; strip pooling

1 引 言

单帧图像超分辨率旨在从观测的低分辨率图像重建出清晰的高分辨率图像，是计算机视觉领域中最经典的图像重建任务之一。清晰的高分辨率图像不仅可以直接用于实际生活中，还能给计算机视觉的其他任务提供帮助，例如目标检测、语义分割。

单帧图像超分辨率是一个病态的逆问题，即同一张低分辨率图像可由许多的高分辨率图像退化得到。目前，解决这一问题的方法主要有三类，基于插值的方法[1-2]、基于重构的方法[3]、以及最近基于实例学习的方法[4-6]。

Dong 等人[7] 在图像插值后使用三层卷积神经网络进行图像超分辨率，展示出比以往所有传统方法更优异的性能。于是在过去的几年里，一系列基于卷积神经网络的单帧图像超分辨率方法被提出来，学习从低分辨率图像输入到其相应高分辨率图像输出的非线性映射函数。通过充分利用训练数据集中固有的图像静态特性，神经网络在单帧图像超分辨率领域取得了显著的进步[8-9]。虽然图像超分辨率方法已经取得了很大的进展，但现有的基于卷积神经网络的超分辨率模型仍然存在一定的局限性：1) 大多数基于卷积神经网络的超分辨率方法主要关注设计更深或更广的网络来学习更有鉴别性的高级特征，而没有充分利用低分辨率图像内部的自相似特征；2) 许多模型没有合理的利用多层次的自相似特征，即使有些方法考虑到了多层次自相似特征的重要性，也没有一个很好的方法来融合它们；3) 大多数方法通过计算每个空间位置的大型关系矩阵来寻找自相似特征，性能往往较低。

本文提出了一种新的跨层次特征增强模块来解决上述的第一个问题和第二个问题。该模块在金字塔结构的每一层嵌入了可变形卷积，并配合跨层次协同注意力来加强跨层次特征传播的能力。由于可变形卷积有一个并行网络学习偏移量，使得卷积核在浅层特征的采样点发生偏移，从而大大提升了网络对浅层特征的建模能力，并且利用可变形卷积还可以积极地使用设计的偏移估计器搜索自相似特征。本文采用了感受野模块[10] 作为可变形卷积的偏移估计器，它以多尺度方式执行像素级别以及特征级别的相似性匹配。

对于第三个问题，许多网络模型引用了非局部网络模块以提高对卷积神经网络中对远程依赖关系建模的能力[11]。然而，单纯的非局部图像恢复方法只探索了相同尺度下的特征相似性，往往性能相对较低。随后，研究人员在此基础上改进成了跨尺度非局部图像恢复方法[12]，虽然性能上有很大的提升，但仍需消耗大量内存来计算每个空间位置的大型关系矩阵。在本文中，为了更有效地捕获这种远程依赖关系，本文提出了池化注意力机制。

实验结果表明，与以往算法的结果相比，本文的重建结果更加准确和真实。如图 1 所示，本文所提出的超分辨率重建网络的主要贡献如下：

1) 提出了一个跨层次特征增强模块 (cross-level feature enhancement module, CLFE)，该模块充分利用低分辨率图像的自相似特征来增强浅层特征。

2) 提出了跨层次协同注意力，在特征金字塔结构中加强了跨层次特征传播的能力。

3) 提出了池化注意力机制，以较低的计算量自适应捕获远程依赖关系，增强了自相似的深层特征，从而显著提高了重建效果。

2 相关工作

2.1 超分辨率中的自相似性

在自然图像中，相似的图案往往在同一图像中重复出现。关于如何利用自相似性进行图像重建，已有多种方法对此进行了研究[11-12]，这些方法试图利用内部信息作为参考来重建高质量的图像。STN[13] 提出了一种允许几何变换模型，该模型处理透视变形和仿射变换。然而，在基于深度学习的方法中利用自相似特征进行图像超分辨率重建的方法仍然是模糊的。为了解决这个问题，一些研究者提出了基于非局部先验的方法。例如 Dai 等人[11] 设计了一种基于 SENet 的二阶注意力机制，并引入了非局部神经网络来进一步提高图像重建的性能。Mei 等人[12] 引入了跨尺度非局部 (cross-scale non-local, CS-NL) 注意力模块，在低分辨率图像中挖掘更多的跨尺度特征相关性。非局部操作通过计算像素相关性，来捕捉全局相关性。相关性计算为输入要素图中所有位置的加权和。这些基于非局部网络的方法虽然一定程度上克服了传统卷积神经

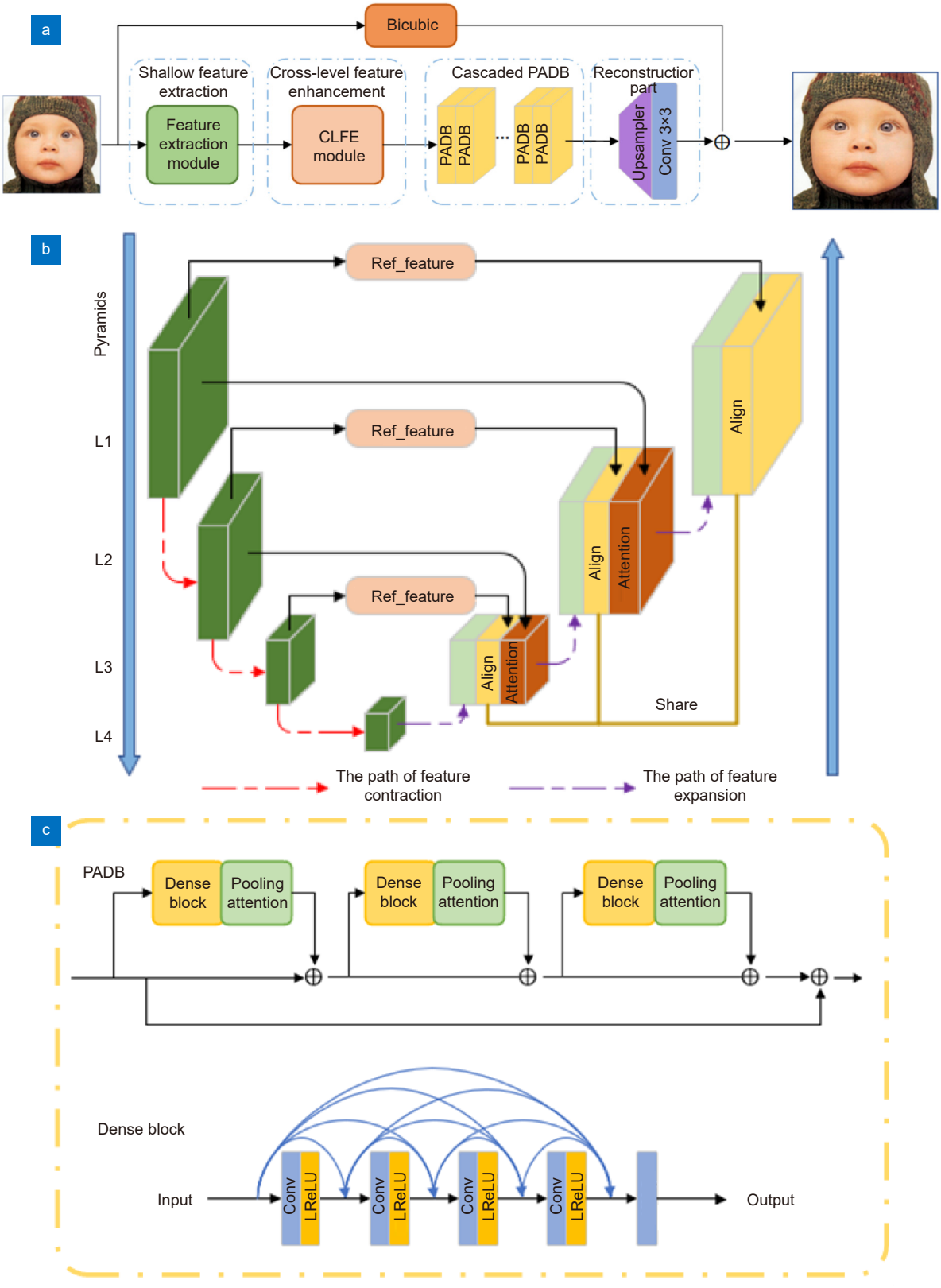


图 1　网络结构。

(a) 本文的基本网络结构; (b) 跨层次特征增强模块; (c) 池化注意力密集块

Fig. 1　Basic architectures.

(a) The architecture of our proposed self-similarity enhancement network;
(b) The cross-level feature enhancement module; (c) The pooling attention dense blocks

网络的局限性，但计算量大。因此，在计算能力有限的情况下，使用这些基于非局部网络的方法进行图像恢复并不是一个很好的选择。本文所提出的方法不仅多层次地搜索浅层特征的自相似性，还可以在池化注意力的帮助下搜索深层特征的自相似性。

2.2 注意力机制

神经网络中的注意机制的目的是为了对上一层输入中最有益和最重要的部分重新校准特征响应。近年来，注意力模块在图像分类[14]、图像生成、图像恢复[15]等一系列任务中取得的成功体现了其重要性。Hu 等人[14] 通过研究网络中卷积特征通道之间的相互依赖性，引入了一种称为挤压和激励 (squeeze-and-excitation, SE) 块的通道注意机制，自适应地重新校准用于图像分类的通道特征响应。受 SE 网络的启发，Zhang 等人[15] 提出了 RCAN，将通道注意力与残差块相结合，增强重要的通道特征，实现了 SISR 的卓越性能。此外，还有一些通过整合通道信息和空间信息来增加注意力的研究，例如，Sanghyun 等人[16] 引入了卷积块注意模块 (convolutional block attention module，CBAM)，该模块应用通道和空间注意来强调有意义的特征。然而，上述注意方法都是利用全局平均或最大池化来获取信道或空间上的统计信息。

与上述方法不同的是，本文提出了跨层次协同注意力来融合金字塔不同层次的特征，并且还提出了一种计算量较小的池化注意力来捕获深层特征的远程依赖关系，以便充分利用图像的自相似特征。

2.3 多尺度表示

多尺度的本质是对信号进行不同粒度的采样，即在不同的尺度下能够观测到不同的特征。源于多尺度这一特性，该结构已成为计算机视觉研究的热点之一。HR-Nets[17] 提出了精心设计的网络体系结构，其中包含多个分支，每个分支都有自己的空间分辨率。沙漏网络[18] 通过跳跃连接将分辨率从高到低过程中的所有低分辨率组合为相同分辨率的特征。多网格卷积神经网络[19] 提出了一种多网格金字塔特征表示方法，并定义了可以在整个网络中集成的 MG-Conv 算子。Oct-Conv[20] 与 MG-Conv 有相似的想法，但其动机是减少参数的冗余。

同时，一些学者也在探索多尺度在图像重建任务中的作用，Han 等人提出了双态递归网络 (dual-state recurrent networks，DSRN) [21]，通过联合低分辨率和高分辨率尺度上的信息来实现图像超分辨率。具体来说，DSRN 中的递归信号通过延迟反馈的方式来进行两个尺度间的信息交换。多尺度残差网络 (multi-scale residual network，MSRN)[22] 通过使用不同尺度的卷积核来提取图像在不同尺度下的特征。Yang 等人提出多级多尺度图像超分辨率网络 (M2SR)[23]，利用残差 U 型网络和注意力 U 型网络提取图像的多尺度特征，增强网络的表达能力。

在上述思想的基础上，本文设计了一个具有多尺度特征和不同层次特征之间信息交互的金字塔结构，进一步增强了提取多尺度特征的能力。

3 本文方法

3.1 方法概述

如图 1 所示，本文提出的网络结构 (self-similarity enhancement network，SSEN) 主要由四个部分组成：浅层特征提取模块、跨层次特征增强模块 (CLFE)、级联的池化注意力密集块以及重建模块。其中$I_{\rm LR}$和$I_{\rm SR}$表示为 SSEN 的输入和输出。如在文献中[9] 所研究的那样，本文仅使用一个卷积层从低分辨率的输入中提取浅层特征：

$$F_{\rm EF} = H_{\rm CLFE}\left(H_{\rm FE}\left(I_{\rm LR}\right)\right), \tag{1}$$

其中：$H_{\rm FE}(\cdot)$表示浅层特征提取模块，提取的浅层特征随后作为跨层次特征增强模块的输入。$H_{\rm CLFE}(\cdot)$表示本文提出的跨层次特征增强模块，它是一个嵌入了若干特征增强模块的金字塔结构，该模块可作为浅层特征提取的一种延伸。因此，本文将其视为一种增强的浅层特征。$F_{\rm EF}$从而替代浅层特征作为级联的池化注意力密集块的输入：

$$F_{\rm DF} = H_{\rm CPADB}\left(F_{\rm EF}\right), \tag{2}$$

其中：$H_{\rm CPADB}(\cdot)$表示本文提出的级联的池化注意力密集块，该模块包含G个池化注意力密集块。Hou 等人提出的条状池化在语义分割中能够有效的捕获远程依赖关系。所以，本文通过池化注意力密集块进行深度特征提取，提取的深度特征为$F_{\rm DF}$，深度特征随后被送入重建模块：

$$I_{\rm SR} = H_{\rm rec}\left(F_{\rm DF}\right) + H_{\rm bic}\left(I_{\rm LR}\right) = H_{\rm SSEN}\left(I_{\rm LR}\right), \tag{3}$$

其中：$H_{\rm rec}(\cdot)$和$H_{\rm bic}(\cdot)$分别表示重建模块和双立方插值函数。重建模块又包含上采样和重建两部分，先使用亚像素卷积进行上采样，然后用一个普通的3×3卷积重建放大的特征。

3.2 跨层次特征增强模块

董超在最近的工作$\rm MS^3-Conv$[24] 中强调了多尺度

特征对超分辨率重建的重要性，并根据多尺度的两个重要因素即特征传播和跨尺度通信，设计了一种通用高效的多尺度卷积单元。受其启发，本文提出了跨层次特征增强模块，其内部结构如图 1(b) 所示可分为三个部分，主体部分为提供多尺度特征的金字塔结构，以及嵌入的特征增强模块和跨层次协同注意力模块。

金字塔结构是一种多尺度特征提取的成熟方案，就是通过多次使用跨步卷积层对输入图像进行下采样，使得大多数计算都在低分辨率空间中完成，从而大大节省了计算成本，最后的上采样层会将特征大小调整为原始输入分辨率。如图 1(b) 中左下角的红色虚线所示，本文使用跨步卷积在第 (L-1) 金字塔层将特征下采样 2 倍，获得金字塔第 L 层的特征表示。本文将红色虚线所构成的路径称为特征收缩路径。同理，上采样过程如紫色的虚线所示，本文将紫色虚线所构成的路径为特征扩张路径。本文从收缩路径中所获得的参考特征一方面作为金字塔同一层次特征增强模块的输入，另一方面又可跨层次提供一些辅助信息。下面将详细阐述特征增强模块和跨层次协同注意力模块。

3.2.1 特征增强模块

首先简要回顾一下可变形卷积，文献 [25] 提出了可变形卷积，以提高卷积神经网络的几何变换的建模能力。它以可学习的偏移量进行训练，这有助于使用变形的采样网格对像素点进行采样。由于这个特性，它被广泛地用于特征配准或隐式运动估计。在这项工作中，本文利用收缩路径的参考特征对扩张路径的输入特征进行增强，采用调制可变形卷积[26]，该方法可另外学习带有调制标量的采样内核的动态权重。

对于输出特征图Y上的每个位置p，普通的卷积过程可以表示为

$$Y(p)=\sum_{k=1}^{K} w_k \cdot X(p+p_k)\,, \tag{4}$$

其中：X 是输入，p_k表示具有 K 个采样位置的采样网格，而w_k表示每个位置的权重。例如，$K=9$且 $p_k \in \{(-1,-1),(-1,0),\ldots,(1,1)\}$可定义一个$3\times3$的卷积核。而在调制的可变形卷积中，将预测的偏移量和调制标量添加到采样网格中，从而使可变形的内核在空间上变化。形式上，可变形卷积运算定义如下：

$$Y_L(p)=\sum_{k=1}^{K} w_k \cdot X_{L,S}\,(p+p_k+\Delta p_k)\cdot \Delta m_k\,, \tag{5}$$

其中：$X_{L,S}$是金字塔第 L 层的支撑特征作为输入，Y_L是金字塔第 L 层特征增强模块的输出，k 和 K 分别表示可变形卷积核的索引和数目。w_k, p, p_k和Δp_k分别是第k个核的权重，中心索引，固定偏移和第k个位置的可学习偏移。Δm_k为调制标量，这里它能够学习到下采样过程的参考特征与输入特征的对应关系。

这样可变形卷积将在具有动态权重的不规则位置上进行操作，以实现对输入特征的自适应采样。由于偏移量和调制标量都是可学习的，因此将每个收缩路径的参考特征与扩张路径的支撑特征连接起来从而生成相应的可变形采样参数：

$$(\Delta P_L, \Delta M_L)=f([R_L, Y_{L+1}\circ\uparrow])\,, \tag{6}$$

其中：[,]表示串联操作，下标 L 表示金字塔第 L 层。R_L表示金字塔第 L 层的参考特征。$Y_{L+1}\circ\uparrow$表示金字塔第 L+1 层的输出结果再上采样 2 倍。而$\Delta P=\{\Delta p_k\}$，$\Delta M=\{\Delta m_k\}$。由于Δp_k可能为分数，本文使用双线性插值，这与文献 [25] 中提出的相同。

特征增强模块由一个可变形卷积和一个给可变形卷积提供偏移量的并行网络组成，如图 2 所示。在特征增强模块中，一个参考特征和一个支撑特征被连接起来作为输入。然后，它们通过一个 3×3 的卷积层来减少通道，并通过一个感受野模块 (RFB) 来增加感受野的大小。接下来的 3×3 卷积层被用来获得可变形核的偏移ΔP_L和调制标量ΔM_L。

图 3 描述了 RFB 的结构。它引入一种类似 Inception 模块的多分支卷积模块，以相对低的计算成本有效地扩大感受野，这有助于处理高频信息较丰富的边缘和纹理。在 RFB 的膨胀卷积层中，每个分支都是一个普通卷积后面加上一个膨胀因子不同的膨胀卷积。因此在保持参数量和同样感受野的情况下，RFB 能够获取更精细的特征。关于 RFB 的更多细节可以在文献 [10] 中找到。RFB 的使用有利于获得有效的感受野，因此本文可以更有效地利用全局特征的自相似性来生成采样参数。

特征增强模块将可变形卷积和 RFB 感受野模块进行巧妙的结合，使得特征在传播过程中能够充分利用全局信息，从而提升特征的表达能力。

3.2.2 跨层次协同注意力

本文提出的跨层次协同注意力 (cross-level co-attention，CLCA) 的目的是自适应地调整来自金字塔不同层次 (图 1(a) 中的深橘色方块) 的重要特征，并为特征融合生成可训练的权重。CLCA 的结构如图 4 所示。

给定一个高层次特征X_L和一个低层次特征X_{L+1}，

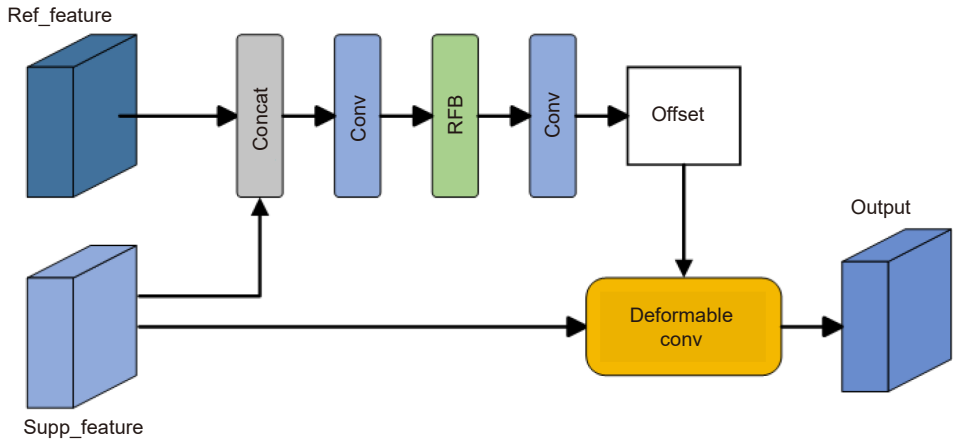


图 2 提出的特征增强模块

Fig. 2 The proposed feature enhancement module

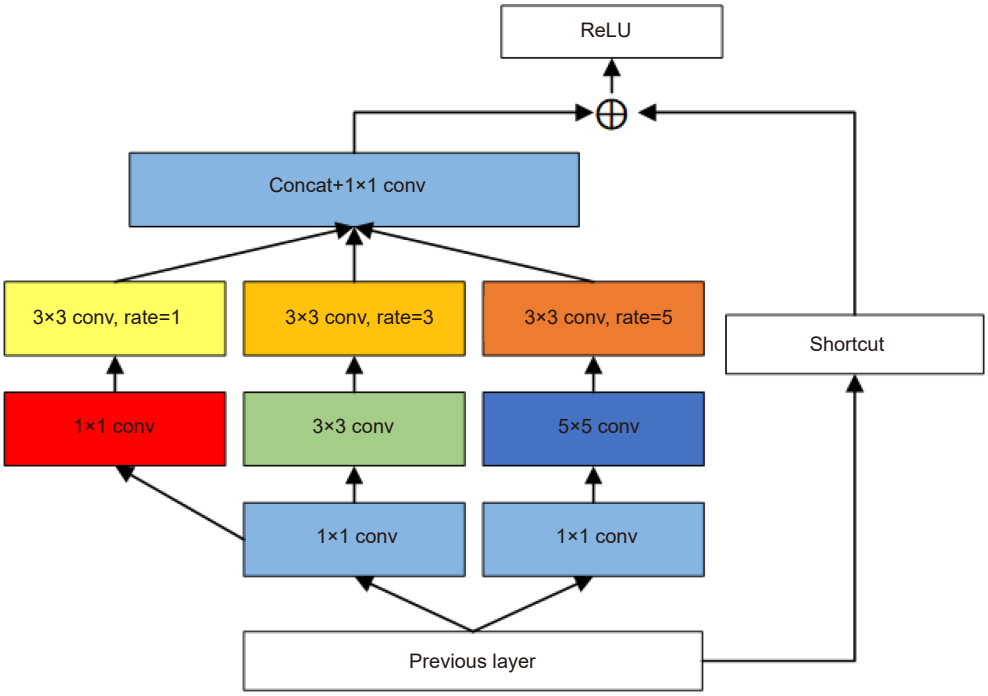


图 3 感受野模块

Fig. 3 Receptive field block

它们的大小分别为$C\times H\times W$和$C\times\frac{H}{2}\times\frac{W}{2}$。首先通过一个全局平均池化将特征$X_L$和$X_{L+1}$的全局空间信息分别压缩到两个信道描述符 z_1 和 z_2，它们第 c 个元素可分别由以下式子求出：

$$z_L^c = F_{\rm gp}(X_L^c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_L^c(i,j)\ , \tag{7}$$

$$z_{L+1}^c = F_{\rm gp}(X_{L+1}^c) = \left(\frac{H}{2}\times\frac{W}{2}\right)^{-1}\sum_{i=1}^{\frac{H}{2}}\sum_{j=1}^{\frac{W}{2}}X_{L+1}^c(i,j)\ , \tag{8}$$

其中：$F_{\rm gp}(\cdot)$表示全局平均池化操作，$X_L^c(i,j)$是X_L第c个通道且位置为(i,j)的值，$X_{L+1}^c(i,j)$是X_{L+1}第c个通道且位置为(i,j)的值。

然后将这两个信道描述符串联成一个信道汇总统计量$S\in\mathbb{R}^{2C\times1\times1}$，其中 $C_{\rm concat}(\cdot)$ 为串联函数。

$$S = C_{\rm concat}\ (z_L, z_{L+1})\ . \tag{9}$$

为了通过全局平均池从聚合信息中完全捕获通道依赖，本文引入了一种能够学习信道之间非线性交互的门控机制。在这里，本文选择利用 Sigmoid 函数 σ，信道统计量可以用以下公式计算：

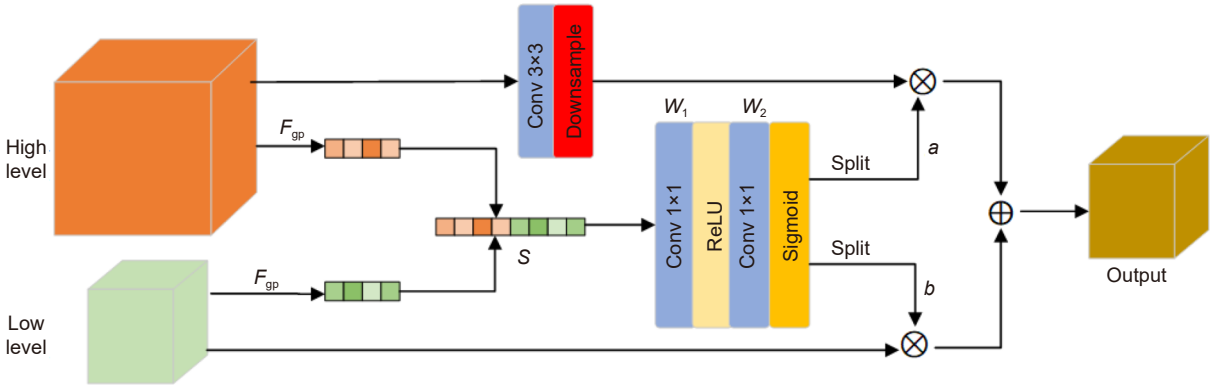


图 4 提出的跨层次协同注意力结构，其中 F_{gp} 表示全局平均池化

Fig. 4 The proposed Cross-Level Co-Attention architec-ture. "F_{gp}" denotes the global average pooling

$$(\boldsymbol{a},\boldsymbol{b}) = S_{\text{split}}(\sigma(W_2\delta(W_1 S))), \tag{10}$$

其中：$\delta(\cdot)$表示 ReLU 激活函数。W_1是第一个卷积层的权重，它作为降维层，具有压缩比 r。在被 ReLU 激活后，低维信号随后以比率 r 升维，其权重是W_2。最后将获得的信道统计量划分为$\boldsymbol{a},\boldsymbol{b}$两部分，用于重新标定不同层次特征的权重。然后将这些特征融合起来，过程如下：

$$F_{\text{output}} = a * \text{down}(\text{conv}(X_L)) + b * X_{L+1}, \tag{11}$$

其中：$S_{\text{down}}(\,)$ 表示下采样过程，$C_{\text{conv}}(\,)$ 表示普通的 3×3卷积，F_{output}表示跨层次协同注意力的输出。

3.3 池化注意力密集块

跨层次特征增强模块输出了增强的浅层特征并馈入后面级联的池化注意力密集块 (pooling attention dense blocks，PADB)。池化注意力密集块主要由具有池化注意机制的堆叠残差密集块组成，而堆叠残差密集块的更多细节可以在文献 [27] 中找到。

池化注意力密集块的结构如图 1(c) 所示。它结合了多级残差网络和密集连接。从而充分利用输入图像的层次特征，获得更好的恢复质量。

3.3.1 池化注意力

池化注意力机制利用空间池化来扩大卷积神经的感受野并收集提供有用信息的上下文，利用条状池化[28] 作为全局池化的替代方法，所谓条状池化就是使用条状池化窗口沿水平或垂直方向执行池化，如图 5 所示。数学上，给定二维张量$\boldsymbol{x}\in\mathbb{R}^{H\times W}$，在条状池化过程中，需要池化的空间范围为 $(H,1)$ 或 $(1, W)$。与二维平均池不同，条状池化对一行或一列中的所有特征值进行平均。因此，水平条状池化后的输出$y^{\text{h}}\in\mathbb{R}^{H}$可以写成：

$$y_i^{\text{h}} = \frac{1}{W}\sum_{0\leqslant j\leqslant W} x_{i,j}. \tag{12}$$

同理，垂直条状池化后的输出$y^{\text{v}}\in\mathbb{R}^{H}$可以写成：

$$y_j^{\text{v}} = \frac{1}{H}\sum_{0\leqslant i\leqslant H} x_{i,j}. \tag{13}$$

条状池化具有两个全局池化所没有的优点。一方面，它可以沿一个空间维度部署较长的内核空间，因

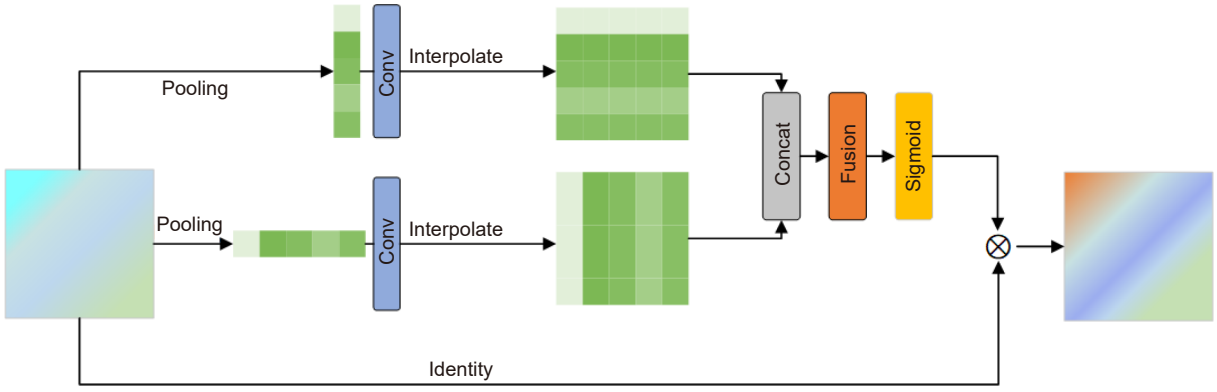


图 5 池化注意力示意图

Fig. 5 Schematic illustration of the pooling attention

此可以捕获离散区域的远程关系；另一方面，条状池化考虑的是长而窄的范围，而不是整个特征图，从而避免了在相距甚远的位置之间建立大多数不必要的连接。

图 5 描述了本文提出的池化注意力。设$x \in \mathbb{R}^{C\times H\times W}$为输入张量，其中 C 表示通道数。本文首先将x馈入两条并行路径，每条路径包含一个水平或垂直条状池化层，后面是一个内核大小为 3 的一维卷积层，用于调制当前位置及其相邻特征。从而给出了水平方向上的池化结果$y^{\mathrm{h}} \in \mathbb{R}^{H\times W}$和垂直方向上的池化结果$y^{\mathrm{v}} \in \mathbb{R}^{H\times W}$。为了获得包含更有用的全局信息输出$\boldsymbol{z} \in \mathbb{R}^{C\times H\times W}$，本文将$y^{\mathrm{h}}$和$y^{\mathrm{v}}$用双线性插值法膨胀为输入相同的大小，再将膨胀后的张量融合起来，得到$\boldsymbol{y} \in \mathbb{R}^{C\times H\times W}$，该过程可表示为

$$y_{c,i,j} = y^{\mathrm{h}}_{c,i} + y^{\mathrm{v}}_{c,j}\,, \tag{14}$$

于是，池化注意力的结果为

$$z = S_{\mathrm{Scale}}\left(x, \sigma(f(y))\right), \tag{15}$$

其中：$S_{\mathrm{Scale}}(,)$ 指的是逐元素乘法，σ是 Sigmoid 函数，f是1×1卷积。应当注意，有多种方式来组合由两个条状池化层提取的特征，例如计算两个提取的一维特征向量之间的内积。然而，考虑到效率并使池化注意力模块更加轻量，本文采用了上述操作，发现这些操作仍然具有不错的效果。

4 实验结果与分析

4.1 数据集与度量方法

根据文献 [9,15]，本文选用了 DIV2K[29] 作为网络的训练集，该数据集由 800 张训练集图片和 100 张验证集图片组成。为了测试模型的效果，本文选用 5 个标准的基准数据集，分别为：Set5[30]，Set14[31]，BSD100[32]，Urban100[5]，Manga109[33]。其中测试集 BSD100 包含有多种风格类型的图片，Urban100 为各种类型的建筑物图片，Manga109 为各种类型的卡通图片。这 5 个测试集具有丰富多样的信息，能够很好地验证超分辨率方法的有效性。为了评估超分辨率性能，本文采用两种常用的全参考图像质量评估标准来评估差异：峰值信噪比 (PSNR) 和结构相似性 (SSIM)。按照超分辨率的惯例，亮度通道被选择用于全参考图像质量评估，因为图像的强度比色度对人类视觉更敏感。

4.2 损失函数

本文采用L_1损失函数[9,15] 来优化 SSEN。对于给定的训练集$\{I^i_{\mathrm{LR}}, I^i_{\mathrm{HR}}\}_{i=1}^N$，包含了 N 个低分辨率和高分辨率图像对。本文的网络目标是训练图像对并利用L_1损失函数来进行优化，公式如下所示：

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\| H_{\mathrm{SSEN}}\left(I^i_{\mathrm{LR}}\right) - I^i_{\mathrm{HR}}\right\|_1, \tag{16}$$

其中：$H_{\mathrm{SSEN}}(\,)$表示网络重建结果。$\|\,\|_1$定义为L_1范数，而θ表示了网络中的参数。

4.3 实验细节

下面具体说明本文提出的 SSEN 的实验细节，在每一轮训练中，本文将低分辨率的 RGB 图像和对应高分辨率的 RGB 图像的切分为大小为 48×48 的块。通过随机旋转 90°、180°、270°和水平翻转来增加训练数据。本文在堆叠的池化注意力密集块中将密集块的个数设置为18，在每个池化注意力密集块中，本文有三个残差密集块和三个池化注意力块。其中残差密集块的增长率为 32，文中未说明的通道数均为 64，网络最后输出的通道数为 3。此外，本文的模型采用 ADAM 优化函数来优化网络，网络的初始学习率设置为2×10^{-4}，并且每迭代 2×10^5次学习率减半。本文所提出的方法实现测试的硬件环境搭配 Intel Core™ i9-9900K(3.6 GHz)、内存 8 GB、配置 NVIDIA GeForce GTX 2080 GPU 的计算机。软件环境为 64 位 Ubuntu 操作系统，PyTorch 框架和 Matlab R2019a。

4.4 实验结果与分析

实验中，本文将 SSEN 与现阶段一些具有代表性的方法作对比，其中包含 Bicubic、SRCNN[7]、VDSR[8]、LapSRN[34]、M2SR[23]、PMRN[35] 和 RDN[36]。为了比较的公平性，将所有的方法在 5 个基准数据集 Set5、Set14、BSD100、Urban100 和 Manga109 上进行实验测试，然后对于不同基准测试集上得到的 PSNR 和 SSIM 指标值分别取平均值。获得的结果列于表 1 中，表中红色字体表示最优结果，蓝色字体表示次优结果。从表中可以看出 SSEN 获得的 PSNR 和 SSIM 值都高于绝大部分其他的对比方法获得的结果值，比如在数据集 Set5 上放大 4 倍的情况下本文的模型重建图像的 PSNR 和 SSIM 值相比于 M2SR 方法分别提高了 0.19 dB 和 0.003，相比于 PMRN 方法分别提高了 0.08 dB 和 0.0011。在数据集 Set14 上放大2倍的情况下，本文的模型重建图像的 PSNR 和 SSIM值相比于 OISR-RK2 方法分别提高了 0.12 dB 和 0.0011，相比于 DBPN 方法分别提高了 0.07 dB 和

表 1 在数据集 Set5、Set14、BSD100、Urban100、Manga109 上放大倍数分别为 2、3、4 的平均 PSNR(dB) 和 SSIM 的结果比较

Table 1 The average results of PSNR/SSIM with scale factor 2×, 3× and 4× on datasets Set5, Set14, BSD100, Urban100 and Manga109

Scale	Method	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
2×	Bicubic	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8409	30.80/0.9339
	SRCNN[7]	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
	VDSR[8]	37.53/0.9590	33.05/0.9130	31.90/0.8960	30.77/0.9140	37.22/0.9750
	M2SR[23]	38.01/0.9607	33.72/0.9202	32.17/0.8997	32.20/0.9295	38.71/0.9772
	LapSRN[34]	37.52/0.9591	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
	PMRN[35]	38.13/0.9609	33.85/0.9204	32.28/0.9010	32.59/0.9328	38.91/0.9775
	OISR-RK2[37]	38.12/0.9609	33.80/0.9193	32.26/0.9006	32.48/0.9317	–
	DBPN[38]	38.09/0.9600	33.85/0.9190	32.27/0.9000	32.55/0.9324	38.89/0.9775
	RDN[36]	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353	39.18/0.9780
	SSEN(ours)	38.11/0.9609	33.92/0.9204	32.28/0.9011	32.87/0.9351	39.06/0.9778
3×	Bicubic	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.96/0.8546
	SRCNN[7]	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
	VDSR[8]	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
	M2SR[23]	34.43/0.9275	30.39/0.8440	29.11/0.8056	28.29/0.8551	33.59/0.9447
	LapSRN[34]	33.82/0.9227	29.79/0.8320	28.82/0.7973	27.07/0.8272	32.19/0.9334
	PMRN[35]	34.57/0.9280	30.43/0.8444	29.19/0.8075	28.51/0.8601	33.85/0.9465
	OISR-RK2[37]	34.55/0.9282	30.46/0.8443	29.18/0.8075	28.50/0.8597	–
	RDN[36]	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653	34.13/0.9484
	SSEN(ours)	34.64/0.9289	30.53/0.8462	29.20/0.8079	28.66/0.8635	34.01/0.9474
4×	Bicubic	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
	SRCNN[7]	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
	VDSR[8]	31.35/0.8838	28.02/0.7680	27.29/0.7260	25.18/0.7540	28.83/0.8870
	M2SR[23]	32.23/0.8952	28.67/0.7837	27.60/0.7373	26.19/0.7889	30.51/0.9093
	LapSRN[34]	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7551	29.09/0.8900
	PMRN[35]	32.34/0.8971	28.71/0.7850	27.66/0.7392	26.37/0.7950	30.71/0.9107
	OISR-RK2[37]	32.32/0.8965	28.72/0.7843	27.66/0.7390	26.37/0.7953	–
	DBPN[38]	32.47/0.8980	28.82/0.7860	27.72/0.7400	26.38/0.7946	30.91/0.9137
	RDN[36]	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
	SSEN(ours)	32.42/0.8982	28.79/0.7864	27.69/0.7400	26.49/0.7993	30.88/0.9132

0.0014。表 1 中的客观指标的实验对比结果证明了本文方法的有效性。

为了从视觉质量上对比不同超分辨率方法的重建性能，图 6 和图 7 分别展示了数据集 Urban100 中“Img048”和“Img092”图像在 4 倍放大时的超分辨率重建结果。图 8 和图 9 分别展示了数据集 B100 中“223061”和“253027”图像在 4 倍放大时的超分辨率重建结果。其中 GT (ground truth) 代表原始 HR 图像。为了突出对比效果，本文选取了图像的局部区域使用双三次插值的方法进行放大。通过观察图 7 和图 9 可以看出，虽然 RDN 方法[36] 能清晰地恢复图像中显著的纹理信息，但这些纹理信息存在明显的方向性问题，而 OISR-RK2 方法[37] 和 DBPN [38] 的方法虽在一定程度上恢复了正确的纹理信息，但难以抑制错误的纹理，并且这两种方法的纹理较为模糊。相比之下，本文方法在图中局部放大区域上能够产生方向正确的纹理和比较清晰的边缘，而且更加符合人眼视觉。这是由于跨层次特征增强模块中的可变形卷积有较强的特征对

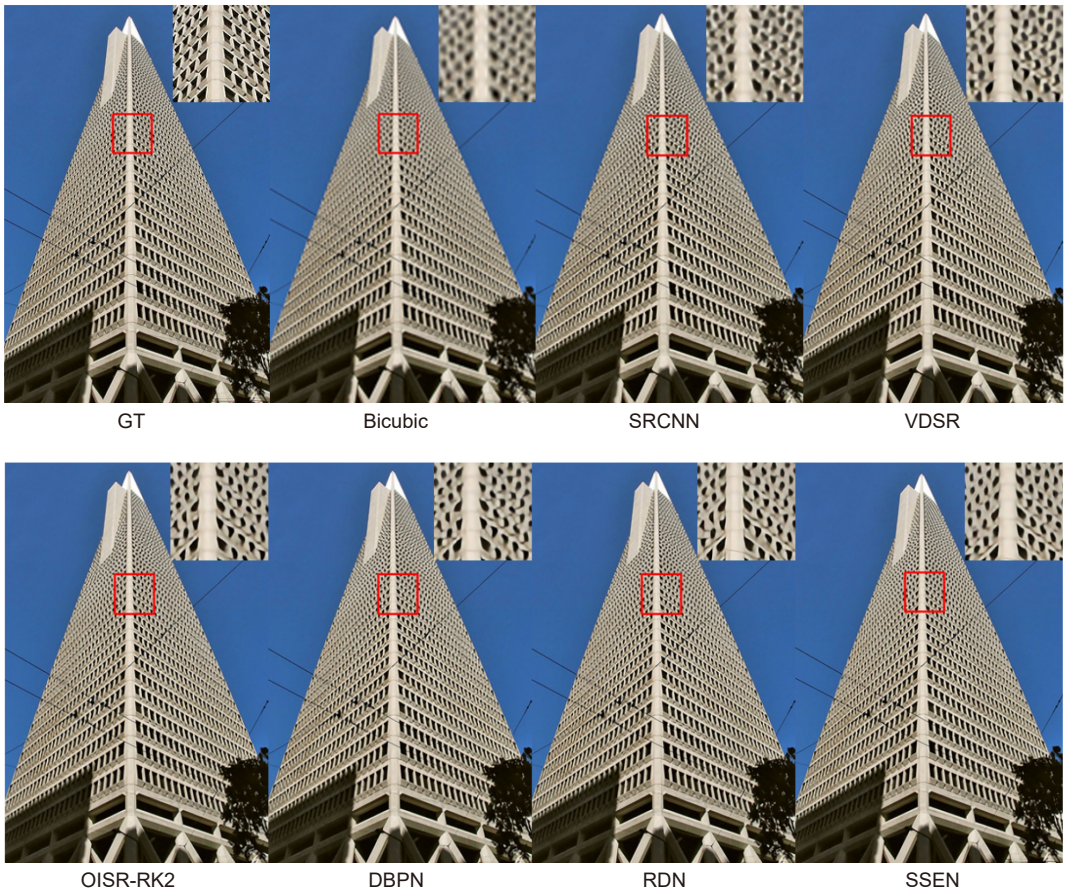


图 6 数据集 Urban100 中“Img048”放大 4 倍的超分辨率结果

Fig. 6 Super-resolution results of " Img048" in Urban100 dataset for 4× magnification

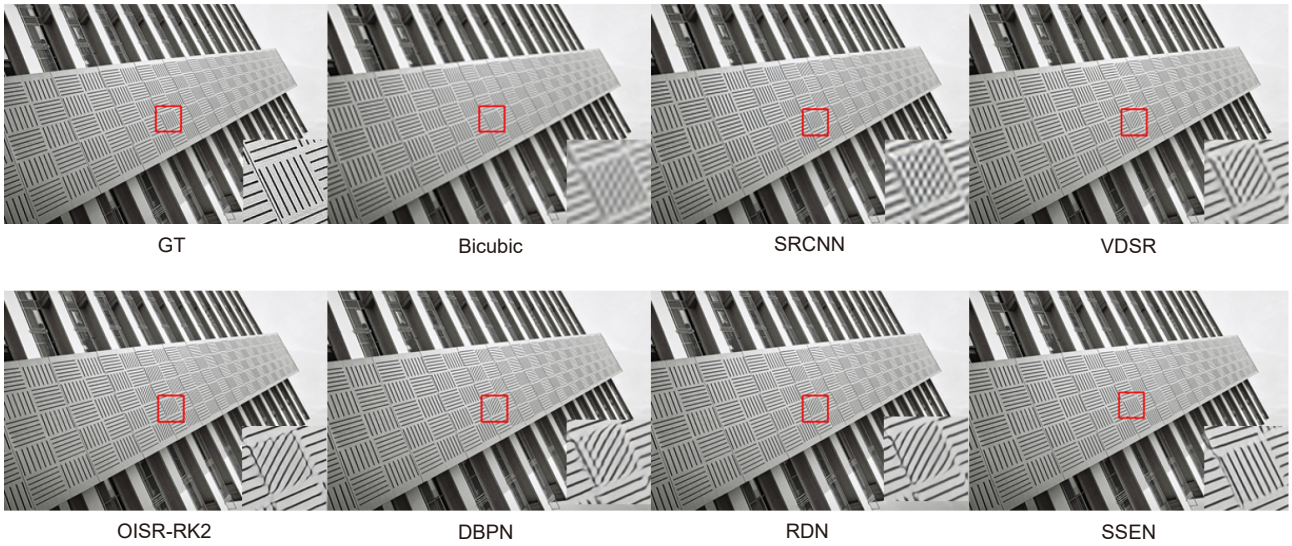


图 7 数据集 Urban100 中“Img092”放大 4 倍的超分辨率结果

Fig. 7 Super-resolution results of " Img092" in Urban100 dataset for 4× magnification

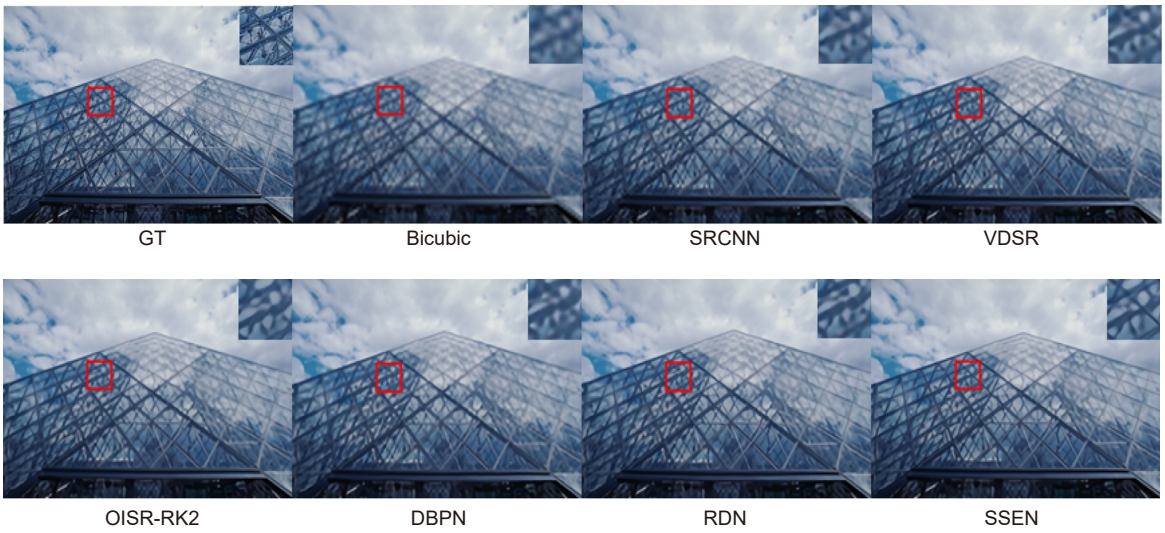


图 8 数据集 BSD100 中“223061”放大 4 倍的超分辨率结果

Fig. 8 Super-resolution results of " 223061" in BSD100 dataset for 4× magnification

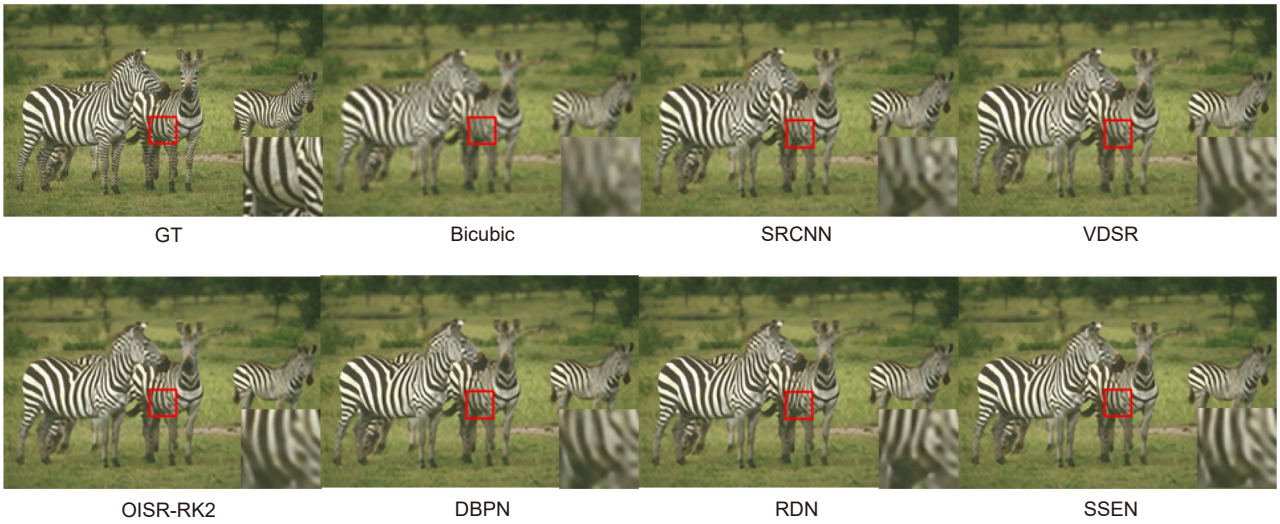


图 9 数据集 BSD100 中“253027”放大 4 倍的超分辨率结果

Fig. 9 Super-resolution results of " 253027" in BSD100 dataset for 4× magnification

齐能力，从而使得网络模型能够更正确的恢复不同图像中复杂的纹理结构。在图 8 的局部放大部分可以清晰地观察到，其他的方法重建出的图像细节很模糊，甚至无法重建出图像的边缘信息，而本文方法重建出的细节更加清晰，具有较好的识别度。这些结果也表明，本文方法在主观表现上取得了更优的效果。

4.5 消融实验

为了验证跨层次特征增强模块和池化注意力密集块的有效性，本文在测试集 Set5 中对图像放大4倍的情况下进行了消融实验来验证本文模型的优越性。

图 10 给出了这五种网络的收敛过程。本文选用 18 个 RRDB 块作为基线，这五种网络具有相同的 RRDB 数。当本文将跨层次特征增强模块和池化注意力密集块分别添加到基线中，得到了 Baseline + CLFE 和 Baseline + Cascaded PADB 这两条曲线。从而验证这两个模块均能有效地提高基线的性能。当本文在模块 CLFE 的基础上去掉跨层次注意力得到了曲线 Baseline + CLFE_no_attention，对比曲线 Baseline + CLFE 可以看出失去注意力的约束后，虽然网络收敛速度变快了，但最终的 PSNR 却下降了 0.03 dB，但仍比基线网络要高 0.04 dB，从而分别验证了特征增强模块和跨层次注意力模块的有效性。当本文同时

向基线网络添加了两个模块，得到曲线 Baseline + CLFE + Cascaded PADB。可以看出，两个模块的组合性能比只有一个模块性能更好。这些定量和可视化分析证明了本文提议的 CLFE 和 PADB 的有效性。

表 2 给出了网络包含跨层次特征增强模块和池化注意力密集块中一种或者两种的情况下的实验结果。从表中可以看出，当本文的网络同时包含跨层次特征增强模块和池化注意力密集块时PSNR值相比于只包含跨层次特征增强模块和只包含池化注意力密集块的情况下分别提高了 0.07 dB 和 0.05 dB，而在 SSIM 上也获得了最大值。

为了更好地展示网络中跨层次特征增强模块的效果，本文分别对只包含浅层特征提取的特征图和加入跨层次特征增强模块的特征图进行了可视化，其中图 11(a) 表示网络在第一层卷积输出的结果，图 11(b) 和图 11(c) 分别代表跨层次特征增强模块输出结果和堆叠的池化注意力密集块输出结果。从图 11(b) 和 11(c) 可以看出，跨层次特征增强模块学习到了图像大量的自相似特征，比如蝴蝶身上的圆形斑点得到了很好的恢复。而堆叠的池化注意力密集块则学习到了更多的图像纹理细节。实验结果表明，本文网络中的两个增强模块起到了很好的自相似特征增强的作用。

4.6 参数和计算量分析

为了进一步验证本文提出模型的有效性，本文在参数的数量方面和计算量方面将 SSEN 与当前公认取得效果比较好的一些深度学习的超分辨率方法进行了分析比较，这些方法包括 EDSR，RDN，OISR-RK3 和 DBPN，参数和计算量结果如表 3 所示。

从表中可以看出 SSEN 在取得了较好客观指标的同时，大幅缩减了网络的参数量和计算量。在数据集 Set14 上放大 2 倍的情况下 SSEN 模型参数量约等

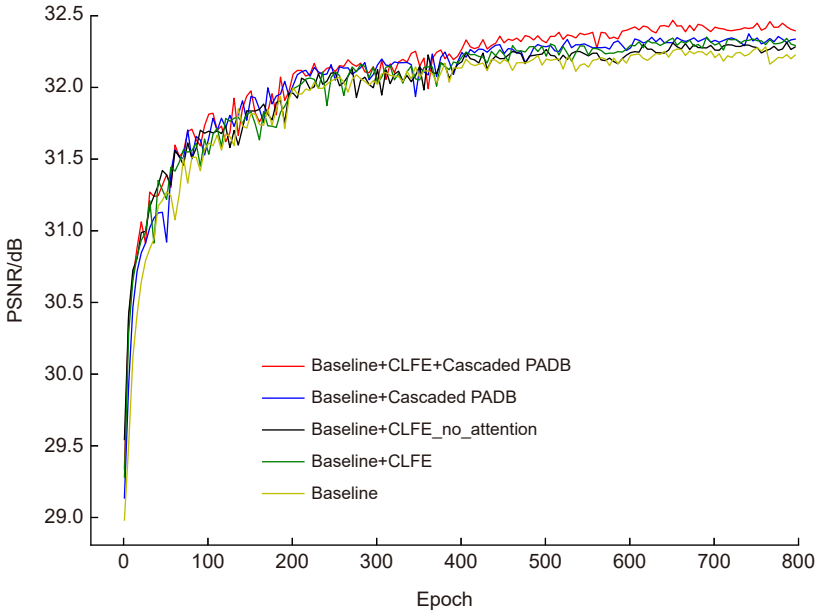


图 10 跨层次特征增强模块和池化注意力密集块聚合分析

每种组合的曲线均基于 Set5，放大因子为 4，共 800 epoch

Fig. 10 Convergence analysis on CLFE and PADB. The curves for each combination are based on the PSNR on Set5 with scaling factor 4× in 800 epochs.

表 2 跨层次特征增强模块和池化注意力密集块在数据集 Set5 放大 4 倍下结果比较

Table 2 The results of cross-level and feature enhancement module and pooling attention dense block with scale factor 4× on Set5

Baseline	√	√	√	√
CLFE	×	√	×	√
Cascaded PADB	×	×	√	√
PSNR/dB	32.28	32.35	32.37	32.42
SSIM	0.8962	0.8971	0.8972	0.8982

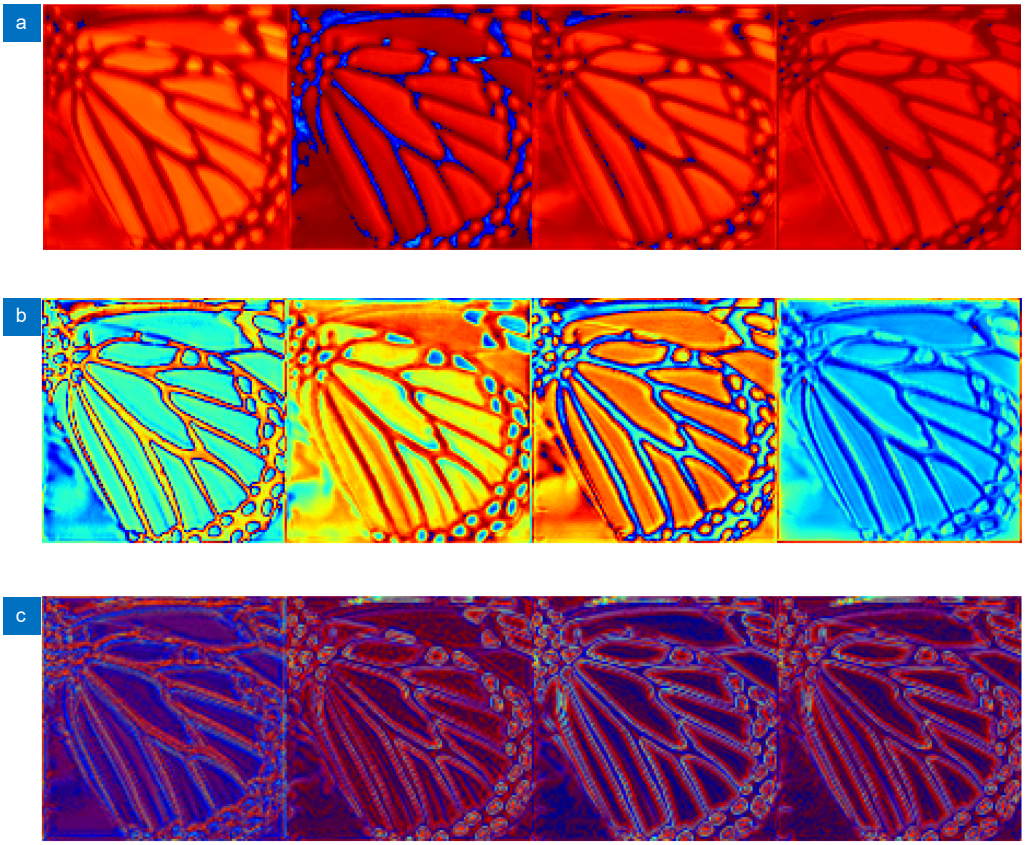


图 11 网络中各模块的输出结果。

(a) 第一层卷积输出结果; (b) 跨层次特征增强模块输出结果;

(c) 堆叠的池化注意力密集块输出结果

Fig. 11 Results of each module in the network.

(a) The result of first layer convolution; (b) The results of cross-level feature enhancement module;

(c) The results of Stacked pooling attention dense blocks

表 3 模型大小和计算量在数据集 Set14 放大 2 倍情况下的比较，计算量表示乘法操作和加法操作的数目之和

Table 3 Model size and MAC comparison on Set14 (2×), "MAC" denotes the number of multiply-accumulate operations

模型	参数	计算量	PSNR/dB	SSIM
RDN[36]	22M	5096G	34.01	0.9212
OISR-RK3[37]	42M	9657G	33.94	0.9206
DBPN[38]	10M	2189G	33.85	0.9190
EDSR[39]	41M	9385G	33.92	0.9195
SSEN	15M	3436G	33.92	0.9204

于 EDSR 和 OISR-RK3 参数量的 36%，计算量也只有它们的 37%，但获得的 PSNR 和 SSIM 结果却十分接近。虽然 SSEN 的参数量和计算量略高于 DBPN 方法，获得的 PSNR 和 SSIM 值相比于 DBPN 方法提高了 0.07 dB 和 0.0014。

由此可以证明，SSEN 在图像重建质量和模型压缩以及计算效率上取得了更好的平衡，即 SSEN 在参数较少时也能获得较好的 PSNR 和 SSIM 结果。在主观视觉效果上，如图 6–9 所示，SSEN 与目前客观指标上较优的 RDN 方法进行比较，取得了相近的重建质量，但 SSEN 参数却比它少了很多。

5 结 论

本文提出了一个基于自相似特征增强网络结构的单帧图像超分辨率重建网络。该方法着重对低分辨率图像内的自相似特征进行增强，本文将整个自相似特

征增强的过程设计成两个即插即用的模块，即跨层次特征增强模块和池化注意力密集块。其中跨层次特征增强模块可作为浅层特征增强模块，在CLFE中，金字塔结构的每一层都嵌入了可变形卷积，以便充分挖掘同一尺度下的自相似信息。金字塔的不同层次间也包含特征的传递，在一定程度上补充了跨尺度的自相似信息，为了防止不同层次的自相似信息相互之间产生干扰，本文提出了跨层次注意力来约束这种信息的传递。此外，本文还提出了池化注意力来挖掘中间特征的自相似特征。通过充分利用浅层特征和中间特征的自相似信息，本文提出的方法无论在客观指标还是在主观表现下都取得了较好的效果。

参考文献

[1] Zhang L, Wu X L. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. *IEEE Trans Image Process*, 2006, **15**(8): 2226−2238

[2] Li X Y, He H J, Wang R X, et al. Single image superresolution via directional group sparsity and directional features[J]. *IEEE Trans Image Process*, 2015, **24**(9): 2874−2888

[3] Zhang K B, Gao X B, Tao D C, et al. Single image super-resolution with non-local means and steering kernel regression[J]. *IEEE Trans Image Process*, 2012, **21**(11): 4544−4556

[4] Xu L, Fu R D, Jin W, et al. Image super-resolution reconstruction based on multi-scale feature loss function[J]. *Opto-Electron Eng*, 2019, **46**(11): 180419
徐亮, 符冉迪, 金炜, 等. 基于多尺度特征损失函数的图像超分辨率重建[J]. 光电工程, 2019, **46**(11): 180419

[5] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 5197–5206.

[6] Shen M Y, Yu P F, Wang R G, et al. Image super-resolution via multi-path recursive convolutional network[J]. *Opto-Electron Eng*, 2019, **46**(11): 180489
沈明玉, 俞鹏飞, 汪荣贵, 等. 多路径递归网络结构的单帧图像超分辨率重建[J]. 光电工程, 2019, **46**(11): 180489

[7] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[C]//*Proceedings of the 13th European Conference on Computer Vision*, 2014: 184–199.

[8] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1646–1654.

[9] 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*, 2019: 2024–2032.

[10] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 404–419.

[11] Dai T, Cai J R, Zhang Y B, et al. Second-order attention network for single image super-resolution[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 11057–11066.

[12] Mei Y Q, Fan Y C, Zhou Y Q, et al. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 5689–5698.

[13] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 2017–2025.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.

[15] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 294–310.

[16] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19.

[17] Sun K, Zhao Y, Jiang B R, et al. High-resolution representations for labeling pixels and regions[Z]. arXiv: 1904.04514, 2019. https://arxiv.org/abs/1904.04514.

[18] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 483–499.

[19] Ke T W, Maire M, Yu S X. Multigrid neural architectures[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4067–4075.

[20] Chen Y P, Fan H Q, Xu B, et al. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 3434–3443.

[21] Han W, Chang S Y, Liu D, et al. Image super-resolution via dual-state recurrent networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1654–1663.

[22] Li J C, Fang F M, Mei K F, et al. Multi-scale residual network for image super-resolution[C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*, 2018: 527–542.

[23] Yang Y, Zhang D Y, Huang S Y, et al. Multilevel and multiscale network for single-image super-resolution[J]. *IEEE Signal Process Lett*, 2019, **26**(12): 1877−1881

[24] Feng R C, Guan W P, Qiao Y, et al. Exploring multi-scale feature propagation and communication for image super resolution[Z]. arXiv: 2008.00239, 2020. https://arxiv.org/abs/2008.00239v2.

[25] Dai JF, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 764–773.

[26] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9300–9308.

[27] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]//*Proceedings of 2018 European Conference on Computer Vision*, 2018: 63–79.

[28] Hou Q B, Zhang L, Cheng M M, et al. Strip pooling: rethinking spatial pooling for scene parsing[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 4002–4011.

[29] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017: 1122–1131.

[30] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//*Proceedings of the British Machine Vision Conference*, 2012.

[31] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C]//*Proceedings of the 7th International Conference on Curves and Surfaces*, 2010: 711–730.

[32] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//*Proceedings Eighth IEEE International Conference on Computer Vision*, 2001: 416–423.

[33] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. *Multimed Tools Appl*, 2017, **76**(20): 21811−21838

[34] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 5835–5843.

[35] Liu Y Q, Zhang X F, Wang S S, et al. Progressive multi-scale residual network for single image super-resolution[Z]. arXiv: 2007.09552, 2020. https://arxiv.org/abs/2007.09552v3.

[36] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 2472–2481.

[37] He X Y, Mo Z T, Wang P S, et al. ODE-inspired network design for single image super-resolution[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 1732–1741.

[38] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1664–1673.

[39] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017: 1132–1140.

作者简介

汪荣贵 (1966-)，男，博士，教授，博士生导师，主要从事深度学习理论与应用、视频大数据与云计算、智能视频监控与公共安全、嵌入式多媒体技术等领域的研究，主持完成国家自然科学基金面上项目等多个纵向研究课题。

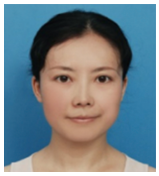
E-mail： wangrgui@hfut.edu.cn

【通信作者】杨娟 (1983-)，女，博士，讲师，硕士生导师，主要从事神经网络与深度学习技术、智能视频图像处理与分析、WEB 数据智能分析软件的研究与实现等的研究。

E-mail：yangjuan6985@163.com

雷辉 (1997-)，男，合肥工业大学在读研究生，本科毕业于武汉科技大学，主要研究方向是计算机视觉和机器学习。

E-mail： 2019170965@mail.hfut.edu.cn

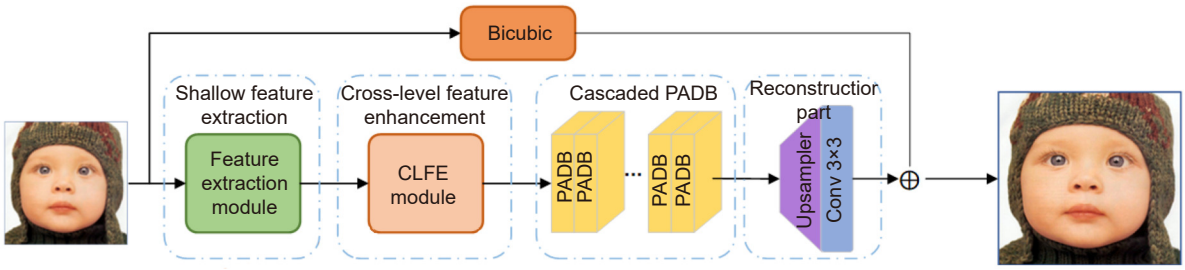


薛丽霞 (1976-)，女，博士，副教授，硕士生导师，主要从事神经网络与深度学习技术、智能视频图像处理与分析、WEB 数据智能分析软件的研究与实现、嵌入式多媒体技术等的研究。

E-mail：51003239@qq.com

Self-similarity enhancement network for image super-resolution

Wang Ronggui, Lei Hui, Yang Juan*, Xue Lixia



The architecture of our proposed self-similarity enhancement network

Overview: Single image super-resolution can not only be directly used in practical applications, but also benefits other tasks of computer vision, such as object detection and semantic segmentation. Single image super-resolution, with the goal of reconstructing an accurate high-resolution (HR) image from its observed low-resolution (LR) image counterpart, is a representative branch of image reconstruction tasks in the field of computer vision. Dong et al. firstly introduced a three-layer convolutional neural network to learn the mapping function between the bicubic-interpolated and HR image pairs, demonstrating the substantial performance improvements compared to those of conventional algorithms. Therefore, a series of single image super-resolution algorithms based on deep learning have been proposed. Although a great progress has been made in image super-resolution methods, existing convolutional neural network-based super-resolution models still have some limitations. First, most CNN-based super-resolution methods focus on designing deeper or wider networks to learn more advanced features of discriminability, but fail to make full use of the internal self-similarity information of the low-resolution images. In response to this problem, SAN introduced non-local networks and CS-NL proposed cross-scale non-local. Although these methods can take the advantage of self-similarity, they still need to consume a huge amount of memory to calculate the large relational matrix of each spatial location. Second, most methods do not make reasonable use of multi-level self-similarity. Even if some methods consider the importance of multi-level self-similarity, they do not have a good method to fuse them, so as to achieve a good image reconstruction effect.

To solve these problems, we propose a self-similarity enhancement network (SSEN). We embedded deformable convolution into the pyramid structure to mine multi-level self-similarity in the low-resolution images, and then introduced the cross-level co-attention at each level of the pyramid to fuse them. Finally, the pooling attention mechanism was utilized to further explore the self-similarity in deep features. Compared with other models, our network mainly has the following differences. First, our network searches self-similarity using an offset estimator of deformable convolution. At the same time, we use the cross-level co-attention to enhance the ability of cross-level feature transmission in the feature pyramid structure. Second, most models capture global correlation by calculating pixel correlation through non-local networks. However, the pooled attention mechanism is used in our network to adaptively capture remote dependencies with low computational cost, which enhances the deep features of self-similarity, thus significantly improving the reconstruction effect. Extensive experiments on five benchmark datasets have shown that the SSEN has a significant improvement in reconstruction effect compared with the existing methods.

Wang R G, Lei H, Yang J, et al. Self-similarity enhancement network for image super-resolution[J]. *Opto-Electron Eng*, 2022, **49**(5): 210382; DOI: 10.12086/oee.2022.210382

Foundation item: The National Key Research & Development Program of China (2020YFC1512601)

School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230601, China

* E-mail: yangjuan6985@163.com