doi：10. 12068/j. issn. 1005－3026. 2026. 20249051

基于空间自适应特征调制的遥感图像超分辨率重建算法

赵小虎[1]，张 悦[2]，慕灯聪[2]
（1. 中国矿业大学 矿山互联网应用技术国家地方联合工程实验室，江苏 徐州 221008；
2. 中国矿业大学 信息与控制工程学院，江苏 徐州 221008）

摘 要：针对遥感图像分辨率较低、细节模糊且重建难度大的问题，提出了一种基于空间自适应特征调制的超分辨率重建网络（SARGAN）. 首先，在生成器方面，提出了空间自适应多尺度特征提取模块，并设计多级残差嵌套结构，对提取到的全局和局部特征逐层调制；提出了特征优化模块，有效利用并整合多尺度信息. 其次，在判别器方面，SARGAN优化了传统视觉几何组（VGG）网络的判别器结构，进一步提高了网络训练的稳定性和生成图像的质量. 最后，实验结果表明，相较于其他算法，SARGAN在峰值信噪比和结构相似性上均有提高，且展现了良好的泛化能力，重建图像的清晰度和结构完整性均有提升.

关 键 词：图像超分辨率；空间自适应特征调制；遥感图像；生成对抗网络；注意力机制
中图分类号：TP 391 **文献标志码**：A **文章编号**：1005-3026(2026)03-0019-11

Super-Resolution Reconstruction Algorithm of Remote Sensing Image Based on Spatial Adaptive Feature Modulation

ZHAO Xiao-hu[1], *ZHANG Yue*[2], *MU Deng-cong*[2]
（1. The National Joint Engineering Laboratory of Internet Applied Technology of Mines, China University of Mining and Technology, Xuzhou 221008, China; 2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221008, China. Corresponding author: ZHAO Xiao-hu, E-mail: xiaohuzhaozxh@163.com）

Abstract: To address the problems of low resolution, blurred details, and high reconstruction difficulty in remote sensing images, a super-resolution reconstruction algorithm based on spatial adaptive feature modulation, named SARGAN, was proposed. Firstly, regarding the generator, a spatial adaptive multi-scale feature extraction module was proposed, and a multi-level residual nested structure was designed to modulate the extracted global and local features layer by layer. A feature optimization module was introduced to effectively utilize and integrate multi-scale information. Secondly, regarding the discriminator, SARGAN optimized the discriminator structure of the traditional visual geometry group (VGG) network, further improving the stability of network training and the quality of generated images. Finally, experimental results show that compared with other algorithms, SARGAN not only has improvements in peak signal-to-noise ratio and structural similarity but also exhibits good generalization ability. The clarity and structural integrity of the reconstructed images are improved.

Key words: image super-resolution; spatial adaptive feature modulation; remote sensing image; generative adversarial network; attention mechanism

遥感图像是利用无人机、航空器或者卫星等遥感平台获取的地球表面图像，其能够捕捉到地

收稿日期：2024-10-21
作者简介：赵小虎（1976—），男，江苏徐州人，中国矿业大学教授，博士生导师.
通信作者：赵小虎，E-mail：xiaohuzhaozxh@163.com.

表的多种信息,包括地形、植被、水体、土地利用等.与自然图像不同,自然图像通常具有丰富的颜色、纹理,分辨率较高且噪声较少.相较之下,遥感图像以俯视视角拍摄地表信息,通常涵盖大范围的地理区域,不同场景中差异较大且分辨率较低.遥感图像的采集过程容易受到天气和大气条件的影响,从而导致图像质量下降或细节纹理缺失.虽然遥感图像在不断提高分辨率,但受限于传感器的轨道高度和成像条件,以及采集过程中地形起伏等因素的影响,部分细小地物无法在遥感图像中清晰呈现.如何提高图像分辨率是遥感图像处理领域的难点之一.最直接的方法是升级用于图像采集的物理设备,但这往往需要高昂的成本.通过超分辨率方法提升遥感图像分辨率能够在避免增加设备成本的前提下,为后续的分析任务提供更详细的地表信息.遥感图像重建的难点主要体现在以下几个方面:首先,遥感图像的获取过程容易受到大气干扰、成像设备噪声等因素的影响,导致图像质量较差;其次,遥感图像往往具有复杂的地表特征,而这些特征在低分辨率图像中容易丢失,使得精确重建变得困难;此外,遥感图像的分辨率差异较大且具有复杂的尺度变化,需要重建算法以便准确捕捉并恢复跨尺度的细节信息.

图像的超分辨率重建是指将给定的低分辨率图像通过特定的算法恢复成相应的高分辨率图像,在医学诊断、卫星遥感、显微镜图像处理等领域应用广泛.其旨在克服或补偿由于图像采集系统或采集环境本身的限制所导致的成像图像模糊、质量低下、感兴趣区域不显著等问题.传统的超分辨率重建方法(如插值方法、基于学习的方法等)无法有效捕获图像中的复杂结构,生成的图像质量往往较低.随着深度学习的发展,其能够利用大规模数据来增强图像建模能力的特点,使得更多的学者聚焦于使用深度学习方法完成图像的超分辨率重建.Dong等[1]提出的超分辨率卷积神经网络(SRCNN)首次将卷积神经网络应用在图像超分辨率领域;之后提出的增强深度残差网络(EDSR)[2]、超深层卷积网络(VDSR)[3]和残差通道注意力网络(RCAN)[4]都采用了深度卷积神经网络架构,通过引入残差学习策略来有效提高超分辨率模型的重建能力和性能.虽然卷积神经网络在超分辨率领域中表现良好,但它们局限于局部感受野,难以有效捕捉全局特征.超分辨率生成对抗网络(SRGAN)[5]的提出将生成对抗网络引入超分辨率领域,通过生成器和判别器的对抗训练,生成更真实、细节更丰富的高分辨率图像.此后基于生成对抗网络的算法相继提出,增强型超分辨率生成对抗网络(ESRGAN)[6]、基于排序学习的超分辨率生成对抗网络(Rank-SRGAN)[7]等模型都取得了不错的重建效果,但这些方法生成的图像往往存在伪影.Transformer最初在自然语言处理领域取得成功后,被引入视觉任务,应用在超分辨率领域时,通过其自注意力机制来捕捉图像的长距离依赖关系和全局特征,有效提升了对复杂细节的重建能力.基于滑动窗口变换器的图像复原网络(SwinIR)[8]、纹理-结构联合建模的双分支网络(TTST)[9]、基于置换自注意力的超分辨率网络(SRFormer)[10]等基于Transformer的模型利用自注意力机制来捕捉图像的全局特征,但在处理局部特征时存在不足,难以捕捉精细的纹理信息.

综上所述,自然图像的超分辨率方法在处理遥感图像时无法充分利用其特有的空间和光谱信息,导致重建效果不理想.设计专门用于遥感图像的超分辨率方法显得尤为重要.基于变换器特征增强的图像超分辨率网络(TransENet)[11]引入基于Transformer的多阶段增强结构,结合传统超分辨率框架,通过在上采样层后增强高维特征表示来提升图像超分辨率效果;分层注意力引导网络(HAUNet)[12]采用混合注意力机制,通过单尺度特征提取模块和跨尺度交互模块以桥接不同尺度特征间的差异;混合尺度自相似性利用网络(HSENet)[13]利用遥感图像中重复出现的目标特征来增强特征表示能力,该网络引入单尺度自相似利用模块和跨尺度连接结构,并结合两者构建混合尺度自相似利用模块,能够同时挖掘图像的单尺度和多尺度信息,从而提升超分辨率性能.虽然这些方法相较于自然图像的重建算法能够提取更多的地面细节,但仍然不能有效应对遥感图像复杂的地形变化,难以充分恢复细节纹理,并容易产生伪影.为解决上述问题,本文提出了一种基于空间自适应特征调制的遥感图像超分辨率重建网络(spatial adaptive residual generative adversarial network, SARGAN),其改进于SRGAN,能够有效地提取和融合多尺度特征信息,提升模型对复杂场景的理解能力.本文的主要贡献:

1)提出了空间自适应多尺度特征提取模块,其包含双分支注意力机制和并行卷积通道混合

器的设计,可以灵活调整特征提取策略,高效捕捉图像中的多层次特征.

2) 设计了一种多级残差嵌套的深层特征融合结构,有效融合多尺度信息,加速网络收敛,解决了SRGAN的残差卷积层捕获全局特征能力较弱、难以融合多尺度特征的问题.

3) 提出了特征优化模块,确保高频信息得到充分捕捉,进一步提升模型在多尺度信息处理上的灵活性,减少冗余特征,提高重建效果.

4) 针对生成对抗网络重建图像易产生伪影的问题,优化了SRGAN中判别器的结构,在提高训练稳定性的同时进一步减少了伪影的产生.

1　相关理论

1.1　SRGAN

Ledig 等[5]在 2017 年提出 SRGAN. SRGAN的生成器将低分辨率(low resolution,LR)图像重建成超分辨率(super resolution,SR)图像后,同真实的高分辨率(high resolution,HR)图像一起送入判别器进行真假判别,并将结果反馈给生成器,通过对抗训练输出细节纹理最真实的重建图像.

生成器网络如图1所示.其主要结构是由残差块组成的深度网络,每个残差块包括卷积层、批归一化(batch normalization,BN)层和PReLU激活函数.低分辨率图像传入生成器后先经过一层卷积进行浅层特征提取,然后经过N个残差块提取深层特征.每个残差块用残差连接加强信息流动,防止因网络深度增加导致的梯度消失问题.最后用亚像素卷积(sub-pixel convolution,SPC)重建得到高分辨率图像.

判别器网络如图2所示.判别器采用VGG网络,中间部分使用了多个卷积模块进行特征提取,激活函数使用LeakyReLU以允许模型在负输入时也有输出,避免神经元的失活.网络的末端使用密集连接层融合各层特征,并使用Sigmoid函数得到图像的分类概率.

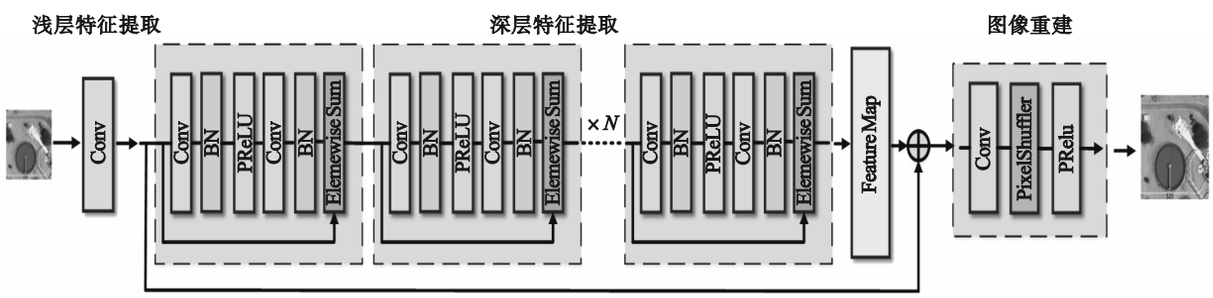


图1　生成器网络

Fig. 1　Generator network

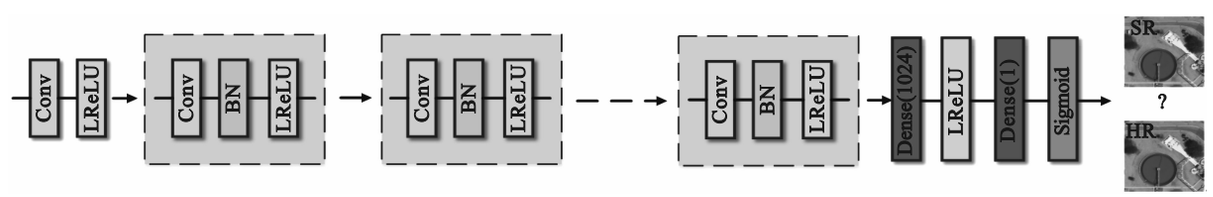


图2　判别器网络

Fig. 2　Discriminator network

1.2　空间自适应特征调制

SAFM[14](spatial adaptive feature modulation)是一种基于视觉Transformer架构的创新设计,能够从多尺度特征中学习长距离的依赖关系,从而更好地挖掘有用特征.如图3所示,SAFM应用一个特征金字塔来生成空间自适应特征调制的注意力图.

首先将特征X在通道维度上进行拆分,得到多个在空间维度上与特征X相同但通道维度缩减的子特征图:

$$[X_0, X_1, X_2, X_3] = \mathrm{Split}(X). \tag{1}$$

式中:X_i为特征X拆分得到的子特征图;Split(·)表示在通道维度上进行拆分.对子特征图进行不同倍率的空间下采样和上采样处理:其中X_0特征图保持不变,其余3个特征图分别进行×2, ×4, ×8的下采样操作.随后,对这些多尺度特征图均应用深度可分离卷积,并对经过下采样的特征图通过最近邻插值法进行上采样,以恢复到原始的空间尺寸:

$$\hat{X}_0 = \mathrm{DW_Conv}_{3\times3}(X_0), \tag{2}$$

$$\hat{X}_i = \uparrow_p (\mathrm{DW_Conv}_{3\times3}(\downarrow_{\frac{p}{2^i}}(X_i))), 1 \leqslant i \leqslant 3. \tag{3}$$

其中：$\mathrm{DW_Conv}_{3\times3}$ 表示 3×3 深度可分离卷积；$\hat{X}_0$ 表示对 X_0 特征图应用深度可分离卷积；$\uparrow_p(\cdot)$ 表示通过最近邻插值将特征上采样到原始分辨率 p；$\downarrow_{\frac{p}{2^i}}(\cdot)$ 表示将输入特征池化到 $\frac{p}{2^i}$；$\hat{X}_i$ 表示对 X_1, X_2, X_3 特征图先池化、再应用深度可分离卷积、最后上采样后的输出. 通道拼接完成后，将多尺度特征连接并通过 1×1 卷积层聚合局部和全局关系，最后与原始输入特征相乘，完成整个空间自适应特征调制流程：

$$\hat{X} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}([\hat{X}_0, \hat{X}_1, \hat{X}_2, \hat{X}_3])), \tag{4}$$

$$\bar{X} = \phi(\hat{X}) \odot X. \tag{5}$$

其中：$\mathrm{Concat}(\cdot)$ 为通道拼接；$\mathrm{Conv}_{1\times1}(\cdot)$ 为 1×1 卷积层；$\hat{X}$ 为特征聚合后的结果；$\phi(\cdot)$ 为 GELU 激活函数；$\odot$ 为特征相乘；$\bar{X}$ 为特征 X 经过空间自适应特征调制后的结果.

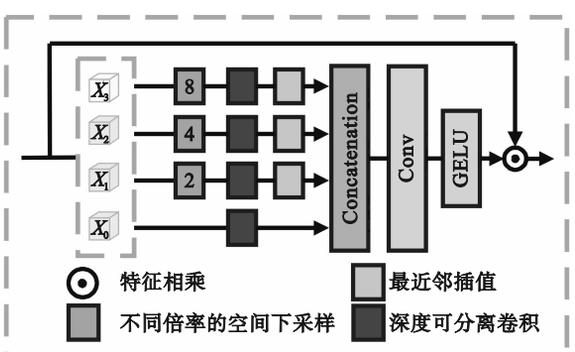


图3 空间自适应特征调制结构

Fig. 3 Spatial adaptive feature modulation structure

2 本文方法

SRGAN能够生成更符合视觉感知的高分辨率图像，显著保留图像中的细腻纹理，但仍存在因细节捕捉不足而导致的部分纹理模糊和重建图像产生伪影的问题. 为了解决这些问题，本文提出了一个基于空间自适应特征调制的遥感图像超分辨率重建网络(SARGAN)，该网络包含生成器和判别器两部分. 生成器接收LR，通过特征提取和重建模块生成SR. 判别器对生成的SR和真实的HR进行区分判断. SARGAN网络结构如图4所示.

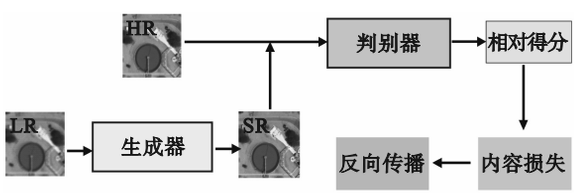


图4 SARGAN网络结构

Fig. 4 SARGAN network structure

2.1 生成器结构

考虑到SRGAN生成器的基本残差块难以应对复杂遥感图像的细节，因此设计了如下模块：并行卷积通道混合器(parallel convolutional channel mixer, PCM)、空间自适应多尺度特征提取模块(spatially adaptive multi-scale feature extraction module, SMFE)、残差融合模块(residual fusion module, RFM)和特征优化模块(feature enhancement module, FEM). 总体结构如图5所示.

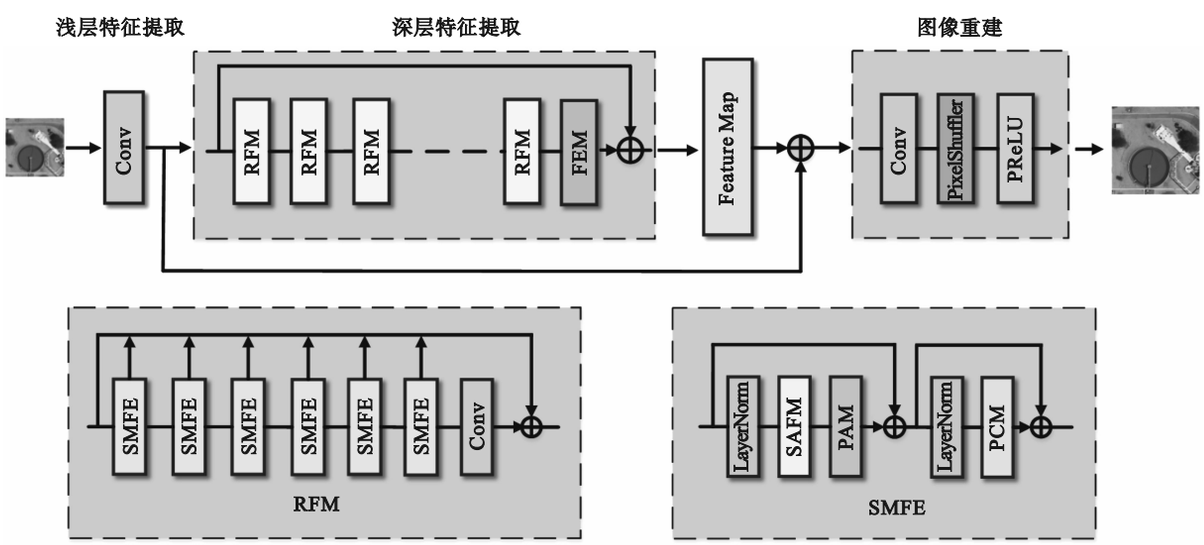


图5 生成器结构

Fig. 5 Generator structure

浅层特征传入RFM后,SMFE模块对其进行特征调制以确保多尺度下特征的充分表达,RFM将每个SMFE的输出特征用残差网络连接在一起,经过多次残差融合后,传入FEM模块进行特征优化,以达到更好的重建效果.

2.1.1　并行卷积通道混合器

虽然SAFM在捕捉全局特征方面表现出色,但对局部信息缺乏有效的处理.因此,本文设计了并行卷积通道混合器(PCM)以解决SAFM对局部信息处理不足的问题.PCM的设计基于FMBConv[15],由于SAFM已经具备了在通道维度上进行动态调节的能力,单独的压缩与激励模块(squeeze-and-excitation networks,SE)[16]在功能上存在冗余,所以将FMBConv中的SE模块替换为GELU激活函数以增强非线性特征表达能力,并引入并行分支以便在不同路径上处理特征.

如图6所示,特征经过这两条分支的并联融合后,有效弥补了SAFM机制处理局部信息时的不足,增强了特征的区分性和表达能力,既保留了细节信息又提升了局部感知能力,方便后续全面提取和整合多尺度信息.

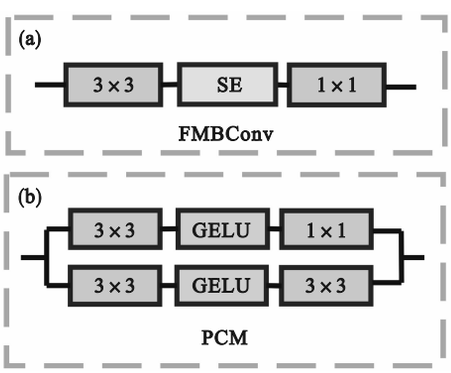


图6　FMBConv和并行卷积通道混合器结构

Fig. 6　FMBConv and parallel convolutional channel mixer structure

(a)—FMBConv;(b)—并行卷积通道混合器.

2.1.2　空间自适应多尺度特征提取模块

在遥感图像处理任务中,图像通常包含丰富且复杂的地物信息,不同尺度的特征对于准确解析图像具有重要意义,但是在重建时会出现由于多尺度特征提取不充分导致无法准确还原图像细节特征的问题.由上文可知,SAFM专注于捕捉全局特征,PCM专注于局部信息的处理,为了最大限度发挥两者的独特优势,本文提出了空间自适应多尺度特征提取模块(SMFE),以提升对全局和局部信息的表达能力.为提高模块对重要特征的关注程度,在SMFE捕获全局信息的分支后面串联双分支注意力机制(parallel attention mechanism,PAM),PAM由无参数注意力机制(SimAM)[17]和SE并联组成.虽然SE在PCM中功能冗余,但是为了增强PCM与SAFM结构之间的适配性,引入SE模块学习通道间的相互依赖关系;同时考虑到自适应特征选择的重要性,将SimAM和SE并联.SimAM模拟神经元激活的原理,专注于特征的有效性评估,该并联结构可以加强通道间的相互依赖关系,增强模型对关键区域和细节的感知能力.整体结构如图7所示.

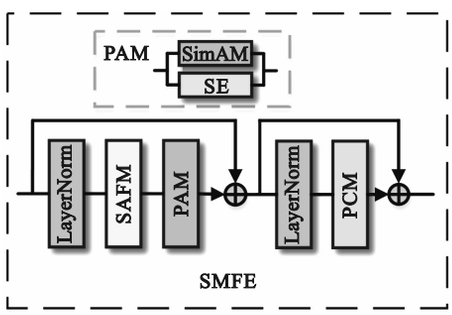


图7　空间自适应多尺度特征提取模块结构

Fig. 7　Spatial adaptive multi-scale feature extraction module structure

浅层特征首先通过层归一化(layer normalization,LN)以标准化特征分布,然后进入SAFM层,动态选择具有代表性的特征.接着,特征通过双分支注意力机制PAM,进一步增强有用特征.之后,原始特征与经过SAFM和PAM调制后的特征相加,形成残差连接,这一过程可以用式(6)表示:

$$F_Y=\mathrm{PAM}(\mathrm{SAFM}(\mathrm{LN}(F_X)))+F_X. \tag{6}$$

随后再经过LN,确保特征的稳定性.这些特征进入PCM,编码局部上下文信息并进行通道混合,最后处理后的特征与前一阶段的特征进行残差连接,该过程如式(7)所示.这里F_X,F_Y,F_Z均为中间特征.

$$F_Z=\mathrm{PCM}(\mathrm{LN}(F_Y))+F_Y. \tag{7}$$

通过将局部和全局信息相融合,SMFE模块实现了多尺度特征的充分表达,为后续多层次残差融合和细节增强奠定了基础.

2.1.3　残差融合模块

尽管SMFE模块能够在一定程度上有效提取遥感图像中的多尺度特征,但遥感图像的复杂性导致单一模块的特征融合能力受限.仅靠单一模块的处理可能导致细节特征被全局特征淹没,或者全局特征未能充分表征,无法避免信息丢失和特征过度调制的问题.为了解决这个问题,本文

设计了残差融合模块(RFM).如图8所示,该模块包含6个SMFE模块和一个卷积层,进一步整合和精炼来自各个SMFE模块提取的特征.考虑到特征间的相关性与信息流动存在不足,RFM引入嵌套式残差连接来综合浅层和深层特征,充分复用不同层次的特征信息,增强了特征间的流动性.

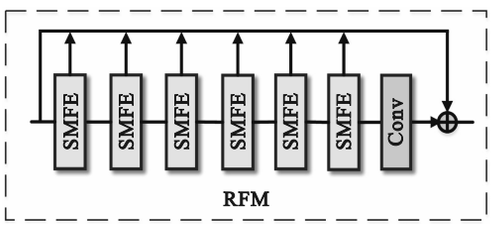


图8 残差融合模块结构

Fig. 8 Residual fusion module structure

该模块内部经过每个SMFE模块得到的输出和经过卷积层的输出为

$$\left.\begin{aligned} F_{\mathrm{SMFE1}} &= R_{\mathrm{SMFE}}(F_{n-1}),\\ F_{\mathrm{SMFE2}} &= R_{\mathrm{SMFE}}(F_{\mathrm{SMFE1}}),\\ &\vdots\\ F_{\mathrm{SMFE5}} &= R_{\mathrm{SMFE}}(F_{\mathrm{SMFE4}}),\\ F_{\mathrm{SMFE6}} &= R_{\mathrm{SMFE}}(F_{\mathrm{SMFE5}}),\\ F_{\mathrm{Conv}} &= \mathrm{Conv}(F_{\mathrm{SMFE6}}). \end{aligned}\right\} \tag{8}$$

式中:$R_{\mathrm{SMFE}}(\cdot)$代表多尺度特征提取模块;F_{n-1}代表上一个残差融合模块的输出特征;$F_{\mathrm{SMFE1}},F_{\mathrm{SMFE2}},\cdots,F_{\mathrm{SMFE6}}$表示经过各个多尺度特征提取模块后的特征.最后,将提取出的不同层次的特征在通道维度上拼接在一起:

$$\boldsymbol{F}_{\mathrm{RFM}} = \left[F_{n-1},F_{\mathrm{SMFE1}},\cdots,F_{\mathrm{SMFE6}},F_{\mathrm{Conv}}\right]. \tag{9}$$

其中,$\boldsymbol{F}_{\mathrm{RFM}}$表示残差融合模块的输出.

2.1.4 特征优化模块

文献[4-5]在深层特征提取的末端放置单个3×3卷积层以优化提取到的特征,然而感受野较小的单一卷积层难以对深度特征进行有效整合,同时RFM模块可能会因为过多的信息处理而导致某些细节的丢失,因此本文提出了特征优化模块(FEM),旨在最大化信息捕捉与整合的效果.FEM主要包括两层卷积、GELU激活函数以及SAFM模块,如图9所示.3×3卷积和GELU激活函数用来强化局部特征,确保关键细节在后续处理中得以保留,避免因特征压缩而丢失重要信息;SAFM模块关注关键信息,提高特征的有效性;1×1卷积用来整合不同路径的特征.为确保结果不受冗余特征的影响,在FEM中加入残差连接来平衡各类信息,从而使得重建图像在细节与边缘处理上表现更加优秀.特征优化模块计算式为

$$\left.\begin{aligned} F_1 &= \sigma(\mathrm{Conv}_{3\times3}(\boldsymbol{F}_{\mathrm{RFM}})),\\ F_2 &= \mathrm{SAFM}(F_1),\\ \boldsymbol{H}_{\mathrm{FEM}} &= \mathrm{Conv}_{1\times1}(F_1+F_2+\boldsymbol{F}_{\mathrm{RFM}}). \end{aligned}\right\} \tag{10}$$

式中:F_1,F_2代表网络不同分支的输出;$\sigma(\cdot)$表示激活函数;$\boldsymbol{H}_{\mathrm{FEM}}$表示特征优化模块的输出.FEM有效解决了特征提取过程中的细节丢失和信息冗余问题.最后,经特征优化后的输出与浅层特征共同送入重建模块,通过亚像素卷积将低分辨率图像的像素拆分成更小的亚像素,并对其进行精细排列和重新组合,获得更高分辨率的重建图像.

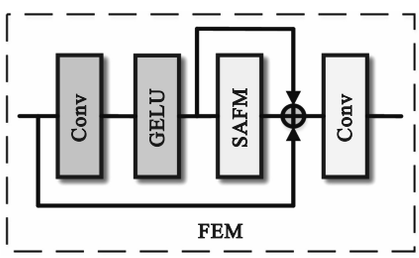


图9 特征优化模块结构

Fig. 9 Feature optimization module structure

2.2 判别器结构

本文判别器的设计基于SRGAN的判别器.SRGAN采用传统的BN层来加速训练收敛并稳定模型输出.BN层通过归一化输入数据的均值和方差,有效地减少了内部协变量偏移的问题.然而,BN层会引入不稳定性,使生成的图像产生不自然的伪影.为了克服BN的这些问题,判别器改用了谱归一化(spectral normalization,SN)层,如图10所示.SN层通过对权重矩阵的奇异值进行约束,直接控制权重的最大奇异值,从而保证判别器的利普希茨连续性(Lipschitz continuity).这种方式可以有效地稳定生成对抗网络的训练过程,不依赖于批次数据的统计特性,同时也能减少伪影的产生.

SRGAN采用的绝对判别器结构直接对生成样本和真实样本进行二分类,不考虑样本之间的相对关系.而相对平均判别器结构则引入了相对性的比较,不仅判断输入样本的真实性,还比较生成样本与真实样本之间的相对差异,因此SARGAN采用相对平均判别器结构.相对平均判别器,它的优势在于使得判别器学习的不只是单一样本的真假,而是生成样本相对于真实样本的相对可信度.这种相对性的信息可以更好地引导生成器学习更加接近真实分布的特征,从而生成纹理细腻的图像.

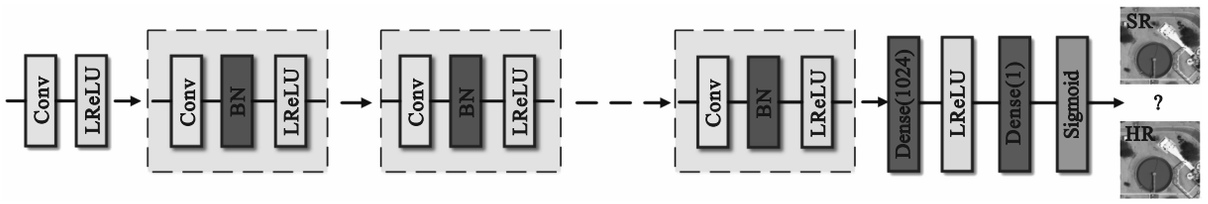


图 10　判别器结构

Fig. 10　Discriminator structure

3　实验分析

3.1　数据集

本文训练集采用DIV2K,Flickr2K和遥感数据集NWPU-RESISC45(NWPU)进行联合训练.使用双三次下采样方法由HR采样得到LR,从而得到HR-LR图像对进行训练和测试.受文献[18]采用DIV2K训练的启发,本文首先选取DIV2K的800张训练图像和Flickr2K的2 550张图像为训练集,这两种数据集图像清晰,有助于模型学习细节信息和纹理特征.选取DIV2K的100张验证图像和Flickr2K的100张图像作为验证集,共计训练集3 350张,验证集200张.训练400个epoch后,从NWPU选取500张图像加入训练集,50张图像加入验证集,继续训练100个epoch得到最终结果.

本文测试集采用遥感数据集和自然图像数据集进行测试.其中,遥感数据集包含UC Merced Land-Use Dataset(UC Merced),RSSCN7 Dataset(RSSCN7)和NWPU,如表1所示.UC Merced用于研究土地利用,具有较高的空间分辨率,类别较为丰富;RSSCN7的图像富有多样性,这些图像来源于不同的季节和天气变化,并以不同的比例进行采样,场景类别较少;NWPU有着极为丰富的场景类型,蕴含了大量的环境信息和特征多样性.

表1　遥感测试集配置

Table 1　Remote sensing test set configuration

数据集	类别数	图像总数/张	尺寸/像素	空间分辨率/m	测试集数量/张
UC Merced	21	2 100	256×256	0.3	150
RSSCN7	7	2 800	400×400	—	150
NWPU	45	31 500	256×256	0.2~30	200

3.2　实验配置

本文实验所用深度学习框架为PyTorch 2.0.0,CUDA 11.8,训练环境为Python 3.8,实验平台的CPU为Intel(R) Xeon(R) Gold 6130,GPU为V100-32 GB(32 GB).采用Adam优化器训练网络,$\beta_1=0.9$,学习率为0.000 1.设定batch size为32,图像裁剪成88像素×88像素大小送入网络,训练400个epoch后加入遥感图像继续训练100个epoch.

3.3　不同方法对比

本文采用峰值信噪比(peak signal-to-noise ratio,PSNR)和结构相似性(structural similarity index,SSIM)来评价图像的重建效果.为了验证SARGAN在遥感图像上超分辨率重建的有效性,将本文方法与SRGAN[5],SwinIR[8],DBPN[19],TransENet[11],AMFFN[20],HAUNet[12],HSENet[13]7种方法进行比较(其中前3种为SR经典算法,后4种为遥感专用算法),在NWPU,RSSCN7和UC Merced 3种数据集上分别进行放大因子为2倍、4倍的重建对比,结果如表2所示.

由表2可知,SARGAN在2倍、4倍放大下的PSNR和SSIM的均值分别为31.69 dB/0.865,27.29 dB/0.688.相比于其他方法,本文提出的方法在3个遥感公开数据集上都取得了最优或次优的PSNR和SSIM.具体来说,从2倍重建结果来看,本文方法相较于原始基线模型SRGAN,PSNR最多提高了0.90 dB,SSIM最多提高了0.010;与4种遥感专用算法相比最多提高了0.47 dB/0.02,在场景类型多变、细节信息丰富的RSSCN7数据集上重建效果显著.这表明本文提出的空间自适应多尺度特征提取机制能够有效地处理复杂场景和细节丰富的图像.SARGAN在UC Merced数据集上的重建性能略低于HAUNet,这是由于HAUNet的跨尺度交互模块更擅长捕捉复杂的上下文信息,在处理蕴含大量细节纹理的UC Merced数据集时的效果更为明显.4倍重建结果延续了2倍重建的优异表现,进一步证明了本文

模型在不同尺度下的稳定性.相比SRGAN,PSNR最多提高了0.66 dB,SSIM最多提高了0.025;与4种遥感专用算法相比,最高提高了0.51 dB/0.022.由于HSENet的混合尺度自相似度开发模块在处理场景类别较少的RSSCN7数据集上更有优势,SARGAN的重建效果略低于HSENet.整体来看,本文算法在不同放大倍数下的重建效果优于其他算法,进一步验证了该算法的有效性,强调了图像多尺度信息的细粒化挖掘和空间自适应特征调制对图像重建的重要性.

表2 不同SR算法在遥感图像上的重建性能指标

Table 2 Reconstruction performance indices of different SR algorithms on remote sensing images

模型	放大因子	NWPU		RSSCN7		UC Merced	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
SRGAN	×2	30.802 0	0.834 3	31.354 2	0.856 2	31.230 6	0.867 8
SwinIR		31.270 1	0.835 3	32.071 5	0.865 2	31.932 1	0.874 6
DBPN		31.243 7	0.834 3	31.792 4	0.861 5	31.344 1	0.864 1
TransENet		31.255 1	0.834 6	31.802 1	0.861 7	31.381 1	0.863 1
AMFFN		31.261 3	0.834 9	31.806 3	0.861 6	31.442 9	0.862 6
HAUNet		31.236 2	0.833 9	31.779 3	0.860 7	**31.557 2**	0.864 0
HSENet		31.281 3	0.840 2	32.061 2	0.861 6	31.509 1	0.871 2
SARGAN		**31.295 0**	**0.843 1**	**32.250 8**	**0.867 2**	31.515 3	**0.883 4**
SRGAN	×4	26.803 2	0.674 1	27.144 0	0.682 1	26.372 9	0.664 2
SwinIR		27.151 2	0.676 6	27.505 1	0.688 8	26.564 9	0.669 8
DBPN		27.122 5	0.666 9	27.156 1	0.676 6	26.564 9	0.669 8
TransENet		27.085 1	0.664 9	27.491 3	0.673 6	26.594 4	0.667 9
AMFFN		27.108 0	0.666 0	27.528 3	0.675 2	26.523 4	0.667 8
HAUNet		27.100 0	0.665 7	27.628 4	0.674 9	27.033 7	0.684 2
HSENet		27.064 6	0.664 3	**27.653 3**	0.673 7	26.641 3	0.667 2
SARGAN		**27.194 1**	**0.682 9**	27.646 3	**0.692 9**	**27.036 7**	**0.689 2**

注:黑体表示最优结果,下同.

3.4 主观视觉分析

为进一步验证本文方法对遥感图像重建性能的提升效果,分别从3个测试集中各挑选一张4倍放大的图片进行可视化分析,方框标出重点区域并放大.从图11可看到,本文方法在细节轮廓以及纹理结构上都表现出了更出色的还原能力,相比其他方法,本文方法能够更准确地重建遥感图像的关键特征.DBPN和SwinIR重建的图像较为平滑,沙漠纹理以及草场边缘轮廓不够明显;SRGAN重建的图像相较之下更符合人眼视觉,但是存在伪影;用遥感专用模型重建的图像相比经典SR方法效果更好,轮廓更清晰,但是在某些关键区域比如沙滩的边缘较为模糊,存在细节丢失的问题.相比之下,本文模型重建效果最优,最为接近真实的高分辨率图像,对复杂结构仍然能够实现较精准的重构,这验证了空间自适应多尺度特征提取机制的有效性.

综上所述,不同模型在特征提取和信息融合能力上的差异使得重建结果存在明显差异,各模型都表现出不同程度的细节丢失和边缘模糊现象.本文提出的基于空间自适应调制的多尺度特征提取模型能够捕捉更多的细节,得益于动态调整特征响应和多尺度信息抓取,实现了全局和局部信息的平衡捕获,从而增强了模型在复杂场景下的表现能力,证明了其在遥感图像重建任务中的优势.

3.5 消融实验

为了验证不同模块对于SARGAN重建效果的影响,对其进行消融实验.选择4倍重建,使用DIV2K和Flickr2K数据集训练400个epoch后,加入NWPU数据集继续训练100个epoch,选择NWPU的200张图像作为测试集,结果如表3所示.

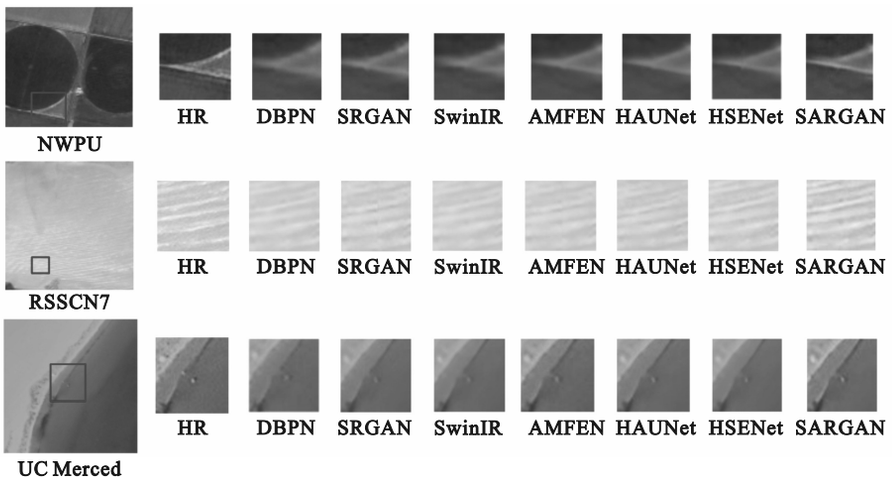


图 11　在遥感数据集上的可视化结果

Fig. 11　Visualization results on remote sensing datasets

表 3　各模块的消融实验

Table 3　Ablation experiments on each module

SENet	SimAM	PCM	多级残差	相对判别器+SN	参数量×10⁶(生成器/判别器)	PSNR/SSIM
	√	√	√	√	8. 675/5. 211	27. 151/0. 679
√		√	√	√	8. 749/5. 211	27. 154/0. 679
		√	√	√	8. 675/5. 211	27. 145/0. 675
√	√		√	√	3. 441/5. 211	27. 098/0. 673
√	√	√		√	8. 749/5. 211	27. 146/0. 673
√	√	√	√		8. 749/5. 215	27. 178/0. 679
√	√	√	√	√	**8. 749/5. 211**	**27. 194/0. 683**

3. 6　泛化性验证

为了验证本文模型在非遥感领域的表现,本节将在遥感数据集之外选择自然基准图像数据集进行额外测试分析.通过对比其他算法在自然图像上的表现,全面验证其在不同场景和图像分布下的鲁棒性和泛化性,实验结果如表 4 所示.

表 4　不同 SR 算法在自然图像上的重建性能指标

Table 4　Reconstruction performance indices of different SR algorithms on natural images

模型	放大因子	Set5		Set14		BSD100		Urban100	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
TransENet	×2	34. 50	0. 92	31. 12	0. 87	30. 41	0. 88	28. 57	0. 90
AMFFN		34. 91	0. 93	31. 10	0. 89	30. 44	0. 88	28. 56	0. 91
HAUNet		**34. 97**	**0. 94**	31. 13	0. 89	**30. 47**	0. 88	28. 60	0. 91
HSENet		34. 29	0. 93	31. 25	0. 88	30. 26	0. 88	28. 40	0. 90
SARGAN		34. 93	**0. 94**	**31. 32**	**0. 89**	30. 36	**0. 88**	**28. 65**	**0. 91**
TransENet	×4	28. 89	0. 82	25. 28	0. 72	25. 14	0. 64	23. 15	0. 69
AMFFN		29. 21	0. 84	25. 75	0. 71	25. 50	0. 69	23. 08	0. 73
HAUNet		29. 23	0. 84	25. 64	0. 75	25. 42	0. 69	23. 51	0. 72
HSENet		28. 25	0. 80	25. 14	0. 72	25. 34	0. 68	**23. 58**	0. 68
SARGAN		**29. 26**	**0. 84**	**25. 85**	**0. 76**	**25. 51**	**0. 70**	23. 53	**0. 73**

从表 4 可以看出,与遥感算法相比,本文模型在大多数情况下取得了不同倍数重建的最优结果.为了直观展示与其他模型的重建差异,从 4 种自然测试集中各选取一张 4 倍重建图像进行可视化分析,如图 12 所示.从图 12 可以看出,本文模型可以很好地重建细节信息,比如蝴蝶的花纹和企鹅的喙.相比其他模型出现的伪影和模糊,本文模型重建的图像更符合人眼视觉,最接近高分辨率图像.这也验证了本文模型具有良好的泛化性,不仅能够在不同数据集上还原整体结构,还能有效重建图像细节.

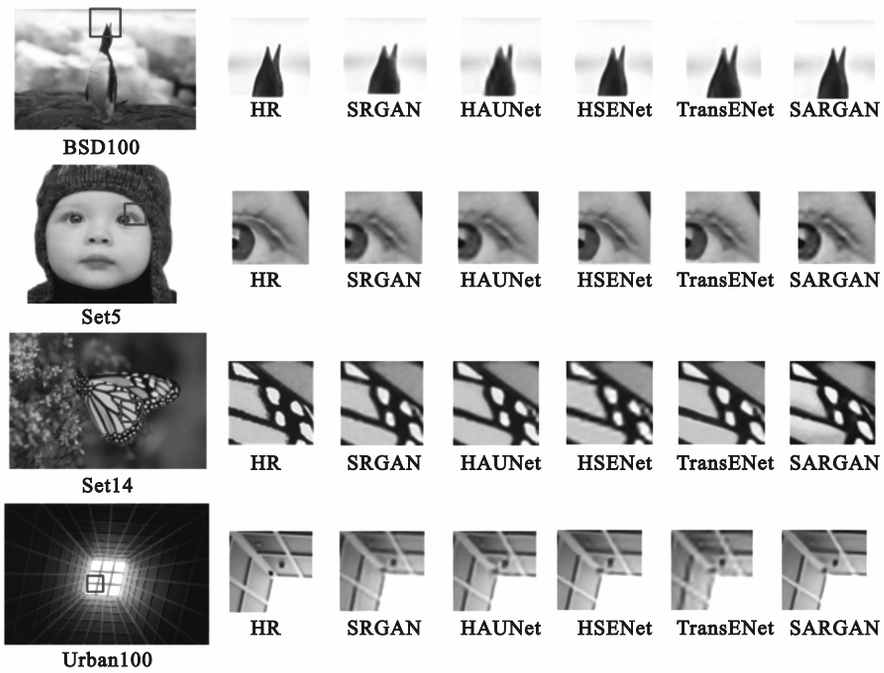


图 12 在自然数据集上的可视化结果

Fig. 12 Visualization results on natural datasets

4 结 语

遥感图像通常具有场景多变、细节缺失以及成像模糊的特征，使得遥感图像在重建时存在多尺度特征提取不充分、细节恢复困难的问题.因此，本文提出了一种基于空间自适应特征调制的遥感图像超分辨率重建算法.首先，针对算法的生成器，提出了空间自适应多尺度特征提取模块以捕捉全局和局部信息；为了避免信息丢失和特征过度调制，设计了残差融合模块用于整合不同层次的特征信息；并提出了特征优化模块，实现了最大化保留关键特征，减少冗余特征.其次，优化了判别器以减少伪影的产生，从而得到更真实的图像.最后，通过仿真实验证明了本文所提方法在遥感图像重建任务中表现出色，PSNR 和 SSIM 都有所提升.相较于其他算法，本文算法重建出的图像纹理细节更为丰富，且具有良好的泛化性.后续工作将研究如何加强判别器的判别能力以及进一步提高图像中复杂密集区域的细节重建效果.

参考文献：

[1] Dong C, Loy C C, He K M, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.

[2] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, 2017: 1132-1140.

[3] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 1646-1654.

[4] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]// Computer Vision-ECCV 2018. Cham: Springer, 2018: 294-310.

[5] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 105-114.

[6] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]// Computer Vision-ECCV 2018 Workshops. Cham: Springer, 2019: 63-79.

[7] Zhang W L, Liu Y H, Dong C, et al. RankSRGAN: super resolution generative adversarial networks with learning to rank[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(10): 7149-7166.

[8] Liang J Y, Cao J Z, Sun G L, et al. SwinIR: image restoration using swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, 2021: 1833-1844.

[9] Xiao Y, Yuan Q Q, Jiang K, et al. TTST: a top-k token selective transformer for remote sensing image super-resolution[J]. *IEEE Transactions on Image Processing*, 2024, 33: 738-752.

[10] Mehri A, Behjati P, Carpio D, et al. SRFormer: efficient yet powerful transformer network for single image super resolution[J]. *IEEE Access*, 2023, 11: 121457-121469.

[11] Lei S, Shi Z W, Mo W J. Transformer-based multistage enhancement for remote sensing image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*,

2022，60：5615611.
［12］ Zhang H K，Lian J，Yi Z T，et al. HAU-Net：Hybrid CNN-Transformer for breast ultrasound image segmentation［J］. *Biomedical Signal Processing and Control*，2024，87：105427.
［13］ Lei S，Shi Z W. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution［J］. *IEEE Transactions on Geoscience and Remote Sensing*，2022，60：5401410.
［14］ Sun L，Dong J X，Tang J H，et al. Spatially-adaptive feature modulation for efficient image super-resolution［C］// 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Paris，2024：13144-13153.
［15］ Sun L，Pan J S，Tang J H. ShuffleMixer：an efficient ConvNet for image super-resolution［EB/OL］.（2022-05-30）［2024-10-01］2022：arXiv：2205.15175. https://arxiv.org/abs/2205.15175.
［16］ Jin X，Xie Y P，Wei X S，et al. Delving deep into spatial pooling for squeeze-and-excitation networks［J］. *Pattern Recognition*，2022，121：108159.
［17］ Liang L H，Zhang Y，Zhang S Q，et al. Fast hyperspectral image classification combining transformers and SimAM-based CNNs［J］. *IEEE Transactions on Geoscience and Remote Sensing*，2023，61：5522219.
［18］ Wang Z Y，Li L L，Xue Y，et al. FeNet：feature enhancement network for lightweight remote-sensing image super-resolution［J］. *IEEE Transactions on Geoscience and Remote Sensing*，2022，60：5622112.
［19］ Haris M，Shakhnarovich G，Ukita N. Deep back-projecti networks for single image super-resolution［J］. *IEEE Transactions on Pattern Analysis and Machine Intelligence*，2021，43(12)：4323-4337.
［20］ Wang X Y，Wu Y D，Ming Y，et al. Remote sensing imagery super resolution based on adaptive multi-scale feature fusion network［J］. *Sensors*，2020，20(4)：1142.

（上接第 18 页）

［21］ 温萍梅，叶志炜，丁文健，等. 命名实体消歧研究进展综述［J］. 数据分析与知识发现，2020，4(9)：15-25.
（Wen Ping-mei，Ye Zhi-wei，Ding Wen-jian，et al. A review of research progress on named entity disambiguation［J］. *Data Analysis and Knowledge Discovery*，2020，4(9)：15-25.）
［22］ Leng J，Tang X J. Graph attention networks for multiple pairs of entities and aspects sentiment analysis in long texts［J］. *Journal of Systems Science and Information*，2022，10(3)：203-215.
［23］ Zhu L，Li N，Bai L Y. Embedding-based entity alignment between multi-source temporal knowledge graphs［J］. *Engineering Applications of Artificial Intelligence*，2024，133：108451.
［24］ Hu P H，Ye Q，Zhang W Y，et al. Integration of multiple terminology bases：a multi-view alignment method using the hierarchical structure［J］. *Bioinformatics*，2023，39(11)：689.
［25］ Ma X J，Li Z Y，Song G J，et al. Learning discrete adaptive receptive fields for graph convolutional networks［J］. *Science China Information Sciences*，2023，66（12）：222101.
［26］ Lin L，Zu L Z，Guo F，et al. Using combinatorial optimization to solve entity alignment：an efficient unsupervised model［J］. *Neurocomputing*，2023，558：126802.
［27］ 段宗涛，李菲，陈柘. 实体消歧综述［J］. 控制与决策，2021，36(5)：1025-1039.
（Duan Zong-tao，Li Fei，Chen Zhe. A survey of entity disambiguation［J］. *Control and Decision*，2021，36(5)：1025-1039.）
［28］ Zhang Y X，Li Y X，Wei X，et al. Graph matching for knowledge graph alignment using edge-coloring propagation［J］. *Pattern Recognition*，2023，144：109851.
［29］ Xie C X，Zhang L M，Zhong Z G. Entity alignment method based on joint learning of entity and attribute representations［J］. *Applied Sciences*，2023，13(9)：5748.
［30］ Wang M，Shi Y H，Yang H，et al. Probing the impacts of visual context in multimodal entity alignment［J］. *Data Science and Engineering*，2023，8(2)：124-134.