

基于 GAN 的多光谱遥感图像超分辨率重建算法

王 宏¹ 刘车琛^{1,2*}
WANG Hong LIU Chechen

摘 要

基于 Real-ESRGAN 超分辨率算法, 对遥感影像进行针对性优化: 根据遥感影像具有原生高-低分辨率图像的特性, 针对性优化退化模型, 并增加了一层微调训练。在数据集方面, 对卫星 0 级 RAW 数据进行切分、对齐、通道转换后进行训练, 最终得到了改进后的算法模型。根据实验结果, 在遥感影像上, 文章所提模型在客观评价指标与主观视觉上均优于其他比较成熟的算法, 并且仍具有一定的优化空间, 对遥感影像后期生产流程的超分辨率流程具有一定的参考价值。

关键词

超分辨率重建; 遥感图像; 多光谱; 生成对抗网络

doi: 10.3969/j.issn.1672-9528.2025.09.034

0 引言

遥感能够大面积、快速、动态地采集地表信息, 它与多个行业存在着紧密联系^[1], 在地质勘测、水域探测、灾害监测以及军事应用等多领域存在广阔的发展前景^[2]。在使用遥感图像进行目标检测^[3]和分类^[4]、场景变化检测^[5]等应用时, 为了获得更加准确的结果, 往往需要高分辨率的图像。为了获取高分辨率遥感图像, 目前最普遍、经济代价最小的即在后期生产对低质量图像进行优化即遥感图像超分辨率技术, 以低分辨率遥感图像数据为基础, 运用算法重建高分辨率遥感图像数据。

当下, 深度学习技术发展已经较为成熟^[6]。在图像超分辨率重建任务领域, 由于深度学习模型能够处理大量的数据、具有更强的映射学习能力及特征表达能力, 相比于传统方法可以恢复更多的细节信息, 重建效果显著提高。

基于插值实现的超分辨率算法^[7,8]作为最早被提出的一种能够实现图像超分辨率重建任务的手段。其中应用较为广泛的方法有: 最近邻插值、双线性插值、双三次插值等^[9]。三种方法实现难度较低, 图像重建速度较快。但重建得出的图像在灰度变换明显的区域会出现锯齿状, 图像整体过于平滑, 图像的纹理细节模糊等问题。

在基于学习重建方法中, 基于 GAN 的方法重建图像的纹理细节更丰富, 更加符合人眼视觉效果。2017 年, Ledig 等人^[10]从“无监督”角度出发, 将生成对抗网络首次应用于图像超分辨率, 提出 SRGAN 模型 (super-resolution

generative adversarial network), 生成了纹理细节更丰富、更符合人类视觉感知的图像。2018 年, Wang 等人^[11]提出了 ESRGAN (enhanced super-resolution generative adversarial networks) 模型, ESRGAN 在 SRGAN 的基础上进行改进, 采用残差密集连接模块作为网络的基本块, 同时去除网络结构中的 BN 层, 减少了重建图像中的伪影问题。2020 年, Ji 等人^[12]提出了 RealSR (real-world super-resolution) 模型, 探索了一种图像真实退化过程, 通过该框架可生成与 HR 图像共享域的 LR 图像。2021 年, Wang 等人^[13]提出 Real-ESRGAN (real-world enhanced super resolution GAN) 模型, 在考虑过程中常见振铃和超调伪影的基础上, 以高阶退化模型模拟复杂的图像退化过程。

1 方法

本文通过对 Real-ESRGAN^[14] 算法进行精简和特异性优化, 使得原算法针对遥感影像进行针对性处理。算法流程如图 1 所示。

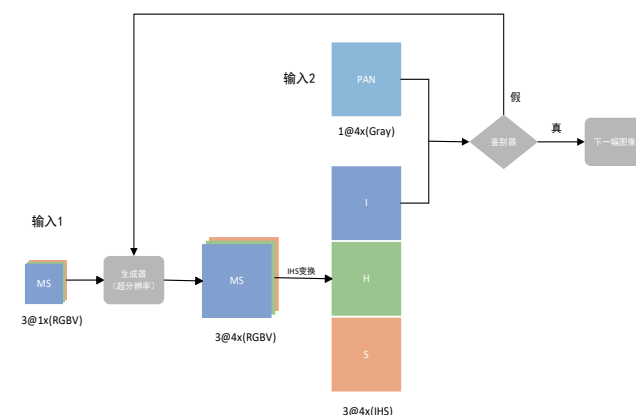


图 1 算法处理流程示意图

1. 西安石油大学计算机学院 陕西西安 710065
2. 陕西航天技术应用研究院有限公司 陕西西安 710061

算法通过学习多光谱影像的 I 通道与全色影像，输出高分辨率的多光谱影像；以便后期可以对超分辨率后的多光谱影像进行 IHS 变换，与全色影像进行融合，可输出质量更好的全色多光谱融合影像。

1.1 退化

原算法设计了一种高阶退化模型，以解决现实世界中的一些复杂退化问题，尤其是未知的噪声和复杂的伪影。但对于本文的处理目标 0 级数据而言，在星上原始数据经过解调解密解压缩后，可能造成图像成像质量损失仅两点：误码和压缩。

误码会带来少则几帧，多则成百上千帧的丢失或损坏，小面积的误码可能会带来一条或小面积花块，大面积的误码可能在处理压缩码流就会将其丢弃，造成图像的不连贯或大面积花块，这种对图像的损伤修复起来比较复杂；而部分卫星为节省带宽，可能在星上采用较高的压缩比对图像进行压缩，通常采用 JPEG2000 或 JPEG-LS 压缩算法，而这两种压缩算法在 8:1 或更高的压缩比下，对图像质量的损失往往比低压缩比带来的损失更加明显。故本文使用 8:1 的压缩比对图像进行压缩解压，以模拟这部分带来的图像质量的降低。

而使用全色影像作为原始高分辨率图像，通道转换后的多光谱影像作为原模型退化后的低分辨率影像，算法使用通道转换对多光谱图像进行处理，从而得到与全色图像近似的低分辨率图像。

1.2 通道转换

遥感卫星 0 级数据为未经处理的 RAW 文件，图像以 16 bits 存储，不仅查看需要知道其数据帧长，在进行处理时，由于 RAW 图像特性，读取器颜色信息时还需要进行大小端转换。

除此之外，多光谱影像与全色影像在色彩上每一通道均有差异，这是因为每个通道拍摄的波段信息不一致，这时候需要对多光谱影像将 RGB 通道转换为 IHS 通道，转换后的 I 通道与全色影像近似，通常对 0 级数据的生产流程更为复杂，但此法可以简单模拟基本处理方法。

RGB 是较为常见的一种色彩空间，其是 Red-Green-Blue 的缩写，意即 RGB 有红、绿、蓝 3 个通道。IHS 也是一种色彩空间，即 IHS 有强度、色调、饱和度 3 个通道。IHS 色彩空间转换为 RGB 色彩空间称为 IHS 正变换，反之则称为 IHS 逆变换^[14]。

IHS 正变换：

$$\begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{\sqrt{2}}{6} & -\frac{\sqrt{2}}{6} & \frac{2\sqrt{2}}{6} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = T \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

IHS 逆变换：

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = T^{-1} \begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} \quad (2)$$

式中：I 为强度分量； v_1 、 v_2 为中间变量。

色调分量 H 和饱和度分量 S 的计算公式：

$$H = \arctan\left(\frac{v_2}{v_1}\right) \quad (3)$$

$$S = \sqrt{v_1^2 + v_2^2} \quad (4)$$

1.3 生成器

ESRGAN 提出了一种密集残差块模型（residual in residual dense block, RRDB），去除了传统 SRGAN 网络中残差块（residual blocks）所有的批归一化（batch normalization）层，因为实验发现 BN 层可能会导致重建图像的伪影问题。RRDB 的引入有效增加了网络容量，如图 2 所示。

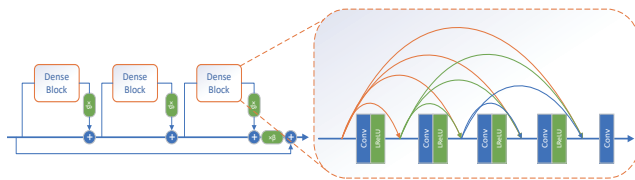


图 2 RRDB 网络结构

RRDB 采用了多级残差密集连接，首先由 5 个卷积激活块通过密集连接构成密集块（dense block），再通过残差连接将 3 个密集块组成 RRDB，为了防止模型在训练过程中出现不稳定现象，在残差连接时加入了残差缩放因子 β ，取值在 0~1 之间。

将 23 个 RRDB 密集连接，便构成了 ESRGAN 的生成器网络模型。

ESRGAN 是一个重型网络，大部分计算是在较小的分辨率空间中进行的，这可以减少显存和计算资源的消耗。如图 3 所示。

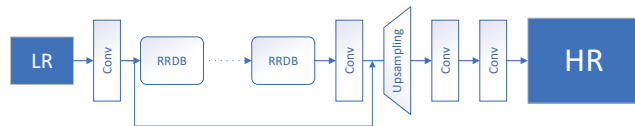


图 3 ESRGAN 生成器网络模型

1.4 判别器

ESRGAN 中采用了 VGG 的判别器，而现在数据量更大，空间细节更丰富，所以引入了 U-Net 判别器。但是 U-Net 和更复杂的退化都加大了训练的不稳定性。为缓解该问题，采用带有频谱归一化^[15]（spectral normalization, SN）的 U-Net

判别器，让其网络参数变动在一个范围内从而稳定判别器的训练。如图 4 所示。

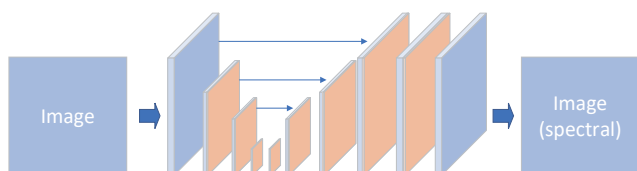


图 4 带有频谱归一化的 U-Net 判别器

判别器分三部分：

(1) 下采样：三层带有谱归一化的卷积层，每层通道翻倍，宽高减半。

(2) 上采样：使用 nearest 插值做上采样，三层带有谱归一化的卷积层，每层通道减半，宽高翻倍，同时与下采样残差边相连。

(3) 输出层：两层有谱归一化的卷积、一层卷积输出层。

2 实验结果及分析

2.1 数据来源

数据集采用资源卫星应用中心高分系列 GF1B 卫星 0 级数据，数据有全色图像和多光谱图像，一轨数据全色和多光谱各 6 片，每片 CCD 两种图像内容相对应。全色图像为单通道，分辨率 6 144，多光谱图像为横向拼接的四通道，每个通道分辨率为 1 536。图像内容包括道路、农田、森林、海岸、云层等常见地貌和现象，如图 5 所示。

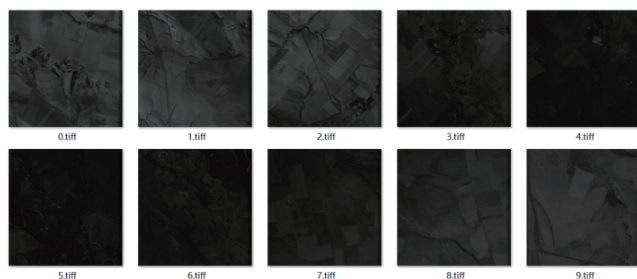


图 5 预处理后的数据集（部分）

数据集大小约为 45 GB，图像格式为 RAW 格式，图像数据每像元为 12 bit，用 16 bit 存储，其中前 12 bit 均有效，后 4 bit 填 0。

使用数据进行训练前，需要对数据进行预处理，即对齐、切分、通道转换。

由于数据多光谱四个通道间拍摄时间或相机略有差异，通常，并排存储的四通道多光谱图像在重合时会有几十个像素的误差，为保证通道转换后图像的正确性，需要对图像进行对齐。对齐后，再根据全色图像和多光谱图像的特征，将多光谱图像切分成 4 个 1 536×1 536 的小块，将全色切分成 6 144×6 144 的大块，同时确保全色和多光谱图像的图像特征一致。

2.2 评价指标

图像质量评价标准（image quality assessment, IQA）是评估图像处理技术性能优劣的重要指标，主要从主观和客观两个层面出发对图像进行科学、系统地研究，以评估图像的质量优劣。通过 IQA 指标的评估，可以帮助人们更好地了解图像处理的效果，并且有效地指导相关技术的研究和发展。

客观图像质量评价，即通过图像分析中能够进行量化的指标如：图像完整性、图像亮度、图像对比度等参数对图像进行定量、定性分析评价图像质量的优劣。一般图像质量评价指标如下：

(1) 均方误差（MSE）：均方误差类似平方距离，均方根误差类似欧氏距离，与人们直观的距离感知较为接近，因此常用于度量人类感知的差异。

$$MSE = \frac{\sum_{x,y,c} e(x,y,c)^2}{W \times H \times C} \quad (5)$$

其中：

$$e(x,y,c) = \hat{l}(x,y,c) - l(x,y,c) \quad (6)$$

式中： l 为图像像素取值； W 、 H 、 C 分别为宽度、高度、通道数。

(2) 峰值信噪比（PSNR）：PSNR 主要用于衡量原始图像与重建图像之间的差异，其数值越高表示重建图像与原始图像越接近。为了评价处理过后图像质量的优劣，通常使用 PSNR 值来代表处理后图像的质量，其通过计算原始图像和重建后的图像之间的均方误差来衡量两者之间的相似程度。信噪比和峰值信噪比与人类对信息的心理感知具有较好的线性关系，常用于模拟人类的信息感知情况。

$$PSNR = 10 \cdot \lg \left(\frac{(2^b - 1)^2}{MSE} \right) \quad (7)$$

式中： b 为图像位数。

(3) 结构相似度（SSIM）：SSIM 是一种基于结构相似性的图像质量评价方法。SSIM 的原理是通过图像的结构、亮度和对比度等多个方面的影响衡量图像之间的相似性。SSIM 不仅能够评价图像的失真程度，还能够反映图像的感知质量，能够更准确地反映图像的质量，其由图像亮度、对比度、结构对比 3 部分组成。

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (8)$$

式中： L 为图像像素最大值，如 8 bits 取 255； μ 为均值； σ^2 为方差； σ 为协方差； c_1 、 c_2 为常数， $c_1=(0.01 \times L)^2$ ， $c_2=(0.03 \times L)^2$ 。

2.3 实验设置

本文实验硬件环境为：Intel Xeon 6226R，NVIDIA RTX

A6000, 192 GB 内存; 软件环境为: Windows 10, Python 3.8, PyTorch 2.1.0, CUDA 12.0。

训练分为三个阶段, 首先训练 Real-ESRNet 生成器模型, 使用 L1 损失训练面向 PSNR 的模型, 基于预训练的 ESRGAN 模型进行训练, batch size 为 12, 学习率 0.000 2, 迭代 4 000 次。然后将上述模型作为生成器初始化, 并结合 L1 损失, 感知损失^[16], GAN 损失^[17] 再次进行训练, batch size 为 12, 学习率 0.000 1, 迭代 4 000 次, 得到 Real-ESRGAN 生成器和鉴别器模型。最后使用配对的高-低分辨率图像进行微调训练, batch size 为 12, 学习率 0.000 1, 迭代 4 000 次, 得到终版改进后的模型。

相比原 Real-ESRGAN, 本文实验增加了一次微调训练, 使超分辨率后的图像尽可能地与全色高分辨率图像保持一致。

2.4 结果对比

本次实验对比的算法选取 Real-ESRGAN, Real-SR, SRMD 和 Waifu2x, 这几种算法都对真实图像具有较好的支持性, 图像及细节、质量评价指标如图 6~8 和表 1~3 所示。

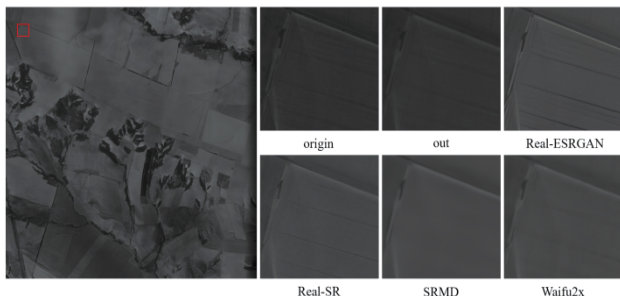


图 6 图像序号 0 对比图

表 1 图像序号 0 质量评价对比表

	PSNR	SSIM
Ours	36.940 660	0.948 380
Real-ESRGAN	36.130 723	0.851 637
Real-SR	36.130 674	0.827 598
SRMD	36.128 013	0.890 736
Waifu2x	36.128 041	0.884 037

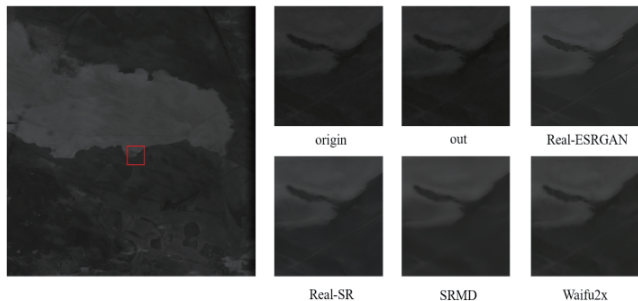


图 7 图像序号 48 对比图

表 2 图像序号 48 质量评价对比表

	PSNR	SSIM
Ours	37.270 260	0.983 283
Real-ESRGAN	36.129 098	0.929 247
Real-SR	36.136 372	0.911 827
SRMD	36.125 737	0.946 147
Waifu2x	36.124 961	0.943 568

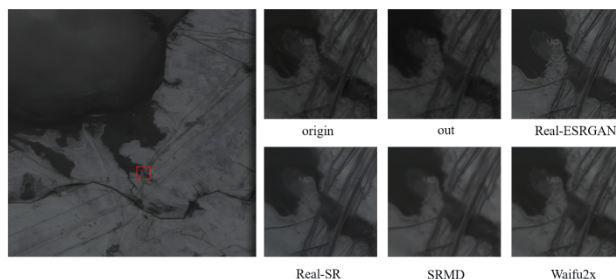


图 8 图像序号 295 对比图

表 3 图像序号 295 质量评价对比表

	PSNR	SSIM
Ours	36.807 900	0.957 448
Real-ESRGAN	36.128 953	0.832 891
Real-SR	36.129 108	0.794 246
SRMD	36.123 413	0.885 621
Waifu2x	36.123 161	0.886 093

由对比可见, 本优化算法对纹理有较好的还原, 如图 6 的农田具有较多纹理的陆面, 除此之外这几乎可以体现在每一组对比样张中。但对于边缘, 尤其是密集的林, 容易出现模糊的情况, 该问题同样发生在边界比较多的图中。尽管边缘处理不尽如人意, 但由于整体效果优于其他算法, PSNR 仍具有优势, SSIM 指标体现的是图像整体的观感, 这一指标优于其他算法也能体现这一特性。

3 未来与展望

首先是图像边界处理问题, 后期可能会加上锐化处理以改善这一现象, 但同时这一问题也可能源于相机传感器的偏差, 使得高-低分辨率图像不能 100% 对齐所导致, 这一现象无法避免, 只能由后期处理来改善。其次, Real-ESRGAN 算法用到了预训练的 ESRGAN 模型, 这一模型是根据真实世界图像训练而来, 对本文针对的使用场景不具有普适性。此外, 该卫星图像分辨率为 2 m 和 8 m, 而且发射时间比较远, 传感器无法避免出现老化, 如果采用比较新的型号的更高分辨率的图像, 效果可能还会进一步提升。

4 总结

本文提出了一种基于 Real-ESRGAN 的改进的针对遥感影像特化的超分辨率算法,使用遥感影像本身具有高-低分辨率图像的特性,以此优化退化流程。同时在残差网络中去除了 BN 层,使用了带有频谱归一化的判别器。最后增加了一层微调训练。在遥感影像数据上实验,结果表明客观指标和主观效果均优于其他主流算法,且仍具有优化空间。

参考文献:

- [1] 熊鹰飞. 基于生成对抗网络的多源跨区域遥感图像超分辨率[D]. 深圳: 中国科学院大学, 2021.
- [2] 成科扬, 荣兰, 蒋森林, 等. 面向目标检测双驱自适应遥感图像超分重建[J]. 北京航空航天大学学报, 2022, 48(8): 1343-1352.
- [3] WANG H Z, WANG Y, ZHANG Q, et al. Gated convolutional neural network for semantic segmentation in high-resolution images[J]. Remote sensing, 2017, 9(5): 446.
- [4] SHI W Z, CABALLERO J, LEDIG C, et al. Cardiac image super-resolution with global correspondence using multi-atlas patch match[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2013: 9-16.
- [5] 王程, 陈峰, 汶德胜, 等. 视觉传感成像技术与数据处理进展[J]. 中国图像图形学, 2021, 26(6): 1450-1469.
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521: 436-444.
- [7] ZHOU F, YANG W M, LIAO Q M. Interpolation-based image super-resolution using multisurface fitting[J]. IEEE transactions on image processing, 2012, 21(7): 3312-3318.
- [8] RUKUNDO O, CAO H Q. Nearest neighbor value interpolation[EB/OL]. (2019-03-04)[2025-04-15]. <https://doi.org/10.48550/arXiv.1211.1768>.
- [9] WEI Z H, HUANG Y D, CHEN Y, et al. A-ESRGAN: training real-world blind super-resolution with attention U-Net discriminators[EB/OL]. (2021-12-19)[2025-03-25]. <https://doi.org/10.48550/arXiv.2112.10046>.
- [10] LEDIG C, THEIS L, HUSZAR 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). Piscataway: IEEE, 2017: 4681-4690.
- [11] WANG X T, YU K, WU S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. Berlin: Springer, 2018: 8-14.
- [12] JI X Z, CAO Y, TAI Y, et al. Real-world super-resolution via kernel estimation and noise injection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2020: 466-467.
- [13] WANG X T, XIE L B, DONG C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data[C]// 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway: IEEE, 2021: 1905-1914.
- [14] 王万同, 李锐. 基于 IHS 的高分辨率遥感影像自适应融合算法[J]. 计算机工程与应用, 2010, 46(31): 22-24.
- [15] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[EB/OL]. (2018-02-16)[2025-06-23]. <https://inspirehep.net/literature/2729520>.
- [16] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision-ECCV 2016. Berlin: Springer, 2016: 694-711.
- [17] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM, 2014: 2672 - 2680.

【作者简介】

王宏(1968—), 男, 陕西西安人, 硕士, 副教授、硕士生导师, 研究方向: 人工智能、大数据应用、计算机视觉、物联网技术。

刘车琛(1999—), 通信作者(email: 15129233899@163.com), 男, 陕西西安人, 硕士研究生, 研究方向: 人工智能、计算机视觉、遥感影像处理。

(收稿日期: 2025-03-31 修回日期: 2025-09-08)