

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-18

论文引用格式: Zhang Zhihao, Fu Zhitao, Ji Yashuai, Zhang Xinshan, Tang Bohui. UKT-GAN: unidirectional knowledge transfer GAN for dual-fidelity SAR-to-optical image translation[J/OL]. Journal of Image and Graphics, XXXX: 1-18. DOI: 10.11834/jig.250573. (张智豪, 付志涛, 嵇娅帅, 张歆山, 唐伯惠. UKT-GAN: 面向SAR到光学图像双保真转换的单向知识迁移生成对抗网络[J/OL]. 中国图象图形学报, XXXX: 1-18. DOI: 10.11834/jig.250573.) [DOI: 10.11834/jig.250573]

UKT-GAN: 面向SAR到光学图像双保真转换的单向知识迁移生成对抗网络

张智豪¹, 付志涛^{1,2,3*}, 嵇娅帅¹, 张歆山¹, 唐伯惠^{1,2,3}

1. 昆明理工大学国土资源工程学院; 昆明 650093; 2. 云南省定量遥感重点实验室; 昆明 650093; 3. 云南省山地灾害天空地一体化智慧监测国际联合实验室; 昆明 650093

摘要: 目的 针对现有SAR到光学图像转换方法采用单一生成器难以兼顾全局语义一致性与局部纹理真实性的问题, 本文提出一种单向知识迁移生成对抗网络(unidirectional knowledge transfer generative adversarial network, UKT-GAN), 通过双分支网络间的单向知识迁移, 实现生成图像在全局与局部层面的双重保真。方法 UKT-GAN由细节重建子网络与语义保持子网络构成。其中, 细节重建子网络专注于图像局部纹理结构细节的生成, 语义保持子网络则负责保障全局语义信息的一致性。通过在两个子网络的输出中添加单向一致性约束, 将细节重建子网络的细节生成能力迁移至语义保持子网络, 来对语义保持子网络的生成图像进行局部纹理结构细节的优化调整。从而确保语义保持子网络生成的图像在全局语义信息与局部纹理结构细节上均具备高度的保真度。结果 在本文的5个数据集上, 相比于当前主流的6种转换方法, UKT-GAN取得了更优的定量结果。在SEN1-2的农田、山地子数据集和WHU-OPT-SAR数据集上, 其PSNR, SSIM, LPIPS和RMSE四项指标全面领先; 在SEN1-2的建筑物和森林子数据集上, 其SSIM和LPIPS两项指标取得了最优值。结论 实验结果表明, 在SAR到光学图像转换方面, UKT-GAN通过单向知识迁移能够有效整合细节重建子网络和语义保持子网络间的互补性优势, 生成结构更清晰、细节更丰富的光学图像。本文代码开源地址: <https://www.scidb.cn/s/YNjqIf>。

关键词: 图像转换; 生成对抗网络; SAR图像; 光学图像; 知识迁移; 双分支网络

UKT-GAN: unidirectional knowledge transfer GAN for dual-fidelity SAR-to-optical image translation

Zhang Zhihao¹, Fu Zhitao^{1,2,3*}, Ji Yashuai¹, Zhang Xinshan¹, Tang Bohui^{1,2,3}

1. Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China; 2. Yunnan Key Laboratory of Quantitative Remote Sensing, Kunming 650093, China; 3. Yunnan International Joint Laboratory for Integrated Sky-Ground Intelligent Monitoring of Mountain Hazards, Kunming 650093, China

收稿日期: 2025-11-13; 修回日期: 2026-04-16

* 通信作者: 付志涛 zhitao@kust.edu.cn

基金项目: 国家自然科学基金项目(62561034)、云南省“兴滇英才支持计划”项目(KKRD202221041)、云南省科技厅基础研究计划面上项目(202301AT070463)、森林生态大数据国家林业和草原局重点实验室开放基金重点项目(2022-BDK-01)、云南省农业遥感与数字技术国际联合实验室(202503AP140020)

Supported by: the National Natural Science Foundation of China under Grant (62561034); “Xingdian” Talent Support Program Project (grant no. KKRD202221041); Yunnan Fundamental Research Projects under Grant (202301AT070463) and the Key Laboratory of State Forestry and Grassland Administration on Forestry and Ecological Big Data Open Fund Priority Project (2022-BDK-01); Yunnan Provincial International Joint Laboratory of Agricultural Remote Sensing and Digital Technology (202503AP140020)

Abstract: Objective Synthetic Aperture Radar (SAR), as an active microwave remote sensing technology, is capable of all-weather and day-and-night data acquisition of the Earth's surface. However, due to its coherent imaging principle, the SAR receiver inevitably introduces significant speckle noise when processing the backscattered signals. Furthermore, SAR imagery primarily reflects the dielectric properties and geometrical structures of targets, rather than the spectral characteristics familiar to the human visual system. This leads to fundamental differences between SAR and optical images in terms of imaging mechanisms, physical properties, and image characteristics. Such differences severely limit their joint application and analysis in subsequent tasks. SAR-to-optical image translation, as a data processing technique capable of converting heterogeneous remote sensing images into data with homogeneous image characteristics, can effectively address the modality gap between SAR and optical imagery. It provides crucial technical support for subsequent applications and analyses such as SAR and optical image matching and SAR and optical image change detection. Therefore, research on SAR-to-optical image translation is of significant importance. However, existing SAR-to-optical image translation methods typically employ a single generator structure, which struggles to simultaneously maintain global semantic consistency and local textural-structural details. This often leads to generated images with semantic distortions and blurred details, consequently limiting their reliability and practical utility in real-world applications. To address this issue, this paper proposes a unidirectional knowledge transfer generative adversarial network (UKT-GAN) for dual-fidelity SAR-to-optical image translation. The model aims to achieve dual fidelity in both global and local dimensions of the generated imagery through unidirectional knowledge transfer between its dual-branch network architecture. **Method** The proposed method employs a dual-branch generation framework that allocates two fundamentally distinct tasks of local texture-structure detail reconstruction and global semantic information preservation to a Detail Reconstruction Subnetwork and a Semantic Preservation Subnetwork, respectively. This architectural design structurally overcomes the inherent limitation of single-branch generative adversarial networks in simultaneously maintaining global semantic consistency and local textural-structural fidelity. The detail reconstruction branch consists of a generator and a discriminator. The generator employs a Unet architecture embedded with CBAM modules, which utilizes skip connections to achieve multi-scale fusion of feature maps from the encoder and decoder, thereby preserving richer spatial details. The discriminator adopts a shallow two-layer convolutional structure to focus on assessing the authenticity of local textural and structural details, while pixel-level loss constraints are applied to ensure the fidelity of local detail reconstruction. Similarly, the semantic preservation branch also comprises a generator and a discriminator. Its generator utilizes a ResNet framework integrated with PE-Transformer modules, leveraging the multi-head attention mechanisms and residual blocks within the PE-Transformer to perform multi-dimensional and deep extraction and integration of semantic features from the imagery. This ensures the complete and accurate retention of semantic information throughout the feature extraction and transmission process. The discriminator employs a deeper four-layer convolutional network to evaluate the overall structural rationality and global semantic consistency of the generated images, supplemented by a combined pixel-level and feature-level loss constraint to maintain global semantic information consistency. Finally, a unidirectional consistency loss transfers the detail capture capability from the detail reconstruction branch to the semantic preservation branch. This process optimizes and refines the local textural and structural details of the optical images generated by the Semantic Preservation Subnetwork, ensuring that these generated images can simultaneously maintain both global semantic consistency and local textural-structural fidelity, thereby enhancing the overall quality of the imagery produced by the Semantic Preservation Subnetwork. **Result** To validate the effectiveness of UKT-GAN in SAR-to-optical image translation, this paper conducts comparative analysis with six mainstream image translation methods, including Pix2pix, DCLGAN, Parallel-GAN, Conditional Diffusion, StegoGAN and HVT-cGAN, on two public datasets: SEN1-2 and WHU-OPT-SAR. The evaluation employs four metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Root Mean Square Error (RMSE), assessing the quality of generated images from four aspects: global quality, structural similarity, deep feature fidelity, and pixel-level accuracy. On the farmland and mountain subsets of SEN1-2 and the WHU-OPT-SAR dataset, the optical images generated by UKT-GAN achieved optimal results across all four metrics: PSNR, SSIM, LPIPS, and RMSE. Particularly notable improvements were observed in the LPIPS metric, with an accuracy improvement of at least 2.3% compared to the second-best results (where lower LPIPS values indicate better performance). On the building and forest subsets of SEN1-2, the

optical images generated by UKT-GAN achieved optimal results on SSIM and LPIPS. These findings demonstrate that the proposed UKT-GAN can generate higher-quality optical images compared to six mainstream image translation methods.

Conclusion This paper proposes a Unidirectional Knowledge Transfer Generative Adversarial Network (UKT-GAN) for dual-fidelity SAR-to-optical image translation. Through unidirectional consistency loss, the model effectively transfers the local textural-structural reconstruction capability from the Detail Reconstruction Subnetwork to the Semantic Preservation Subnetwork. This enables the generation of optical images that simultaneously maintain both global semantic information and local textural-structural details during testing, while requiring only the single Semantic Preservation Subnetwork to be deployed. Furthermore, experimental results demonstrate that compared to other methods, the proposed UKT-GAN can generate optical images with clearer structures and superior overall quality when processing different land cover types. Moreover, when confronted with data distribution shifts caused by varying sensor characteristics and imaging configurations, UKT-GAN maintains stable translation performance, exhibiting strong generalization capability. The open source code for this paper is available at: <https://www.scidb.cn/s/YNjqIf>.

Key words: image translation; generative adversarial network; SAR image; optical image; knowledge transfer; dual-branch network

论文引用格式: Zhang Zhihao, Fu Zhitao, Ji Yas-huai, Zhang Xinshan, Tang Bohui. UKT-GAN: Unidirectional Knowledge Transfer GAN for Dual-Fidelity SAR-to-Optical Image Translation [J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.250573. (张智豪, 付志涛, 嵇娅帅, 张歆山, 唐伯惠. UKT-GAN: 面向SAR到光学图像双保真转换的单向知识迁移生成对抗网络[J/OL]. 中国图象图形学报. DOI: 10.11834/jig.250573.)

0 引言

SAR图像与光学图像在成像机制、物理特性和图像形态方面存在显著差异,这种差异严重制约了二者在后续任务中的应用与分析。SAR到光学图像转换作为一种能够将异源遥感图像转换为具有同源图像特征的数据处理技术,能够有效地解决SAR图像与光学图像之间的模态差异问题,为后续地SAR与光学图像匹配、SAR与光学图像变化检测、SAR图像目标识别、SAR与光学图像野火评估及作物监测、SAR图像语义分割等遥感应用提供重要的技术支撑(Toriya等,2019;Li等,2021;Sun等,2022;Hu等,2023;Kwak等,2024;杜文亮等,2026)。因此,进行SAR图像到光学图像的转换研究具有重要的意义。但现有的图像转换方法在进行SAR到光学图像转换时,仍存在重大挑战。首先,由于遥感图像中地物类型复杂多样,不同地物在特征表达上存在显著差异,要求模型必须具备强大的全局语义理解与转换

能力;其次,SAR成像特有的相干斑噪声不仅会干扰图像质量,还会导致地物边缘结构信息的模糊,使得模型在转换过程中需要平衡去噪效果与边缘保持的精度。此外,不同成像条件、传感器参数和区域特性带来的域间差异,也对模型的泛化能力提出了更高要求。

为了克服这些挑战,研究者们陆续提出了一系列面向跨模态图像转换任务的生成对抗网络,以提升生成图像的质量与保真度(余佩伦等,2021)。研究者们首先聚焦于生成器的优化上(如图1(a)),通过在生成器中引入特征增强模块(Guo等,2021)、注意力机制(Weng等,2023)、多尺度特征融合(Turnes等,2022;Chen等,2024)等方法来增强生成器对SAR图像特征的捕获与利用能力,从而提高生成图像的质量。然而,这类优化方法存在显著局限性,模型难以同时兼顾全局语义信息与局部纹理结构细节重建这两种截然不同的任务,导致网络在优化过程中面临难以调和的内部冲突,制约了生成质量的进一步提升。为了克服这一局限,研究者们又提出了两类新的优化方法:一是在优化生成器的基础上引入多尺度判别器(Yang等,2022;Wei等,2023),提供从全局到局部的分层监督(如图1(b));二是在生成器的瓶颈层中分别提取图像的全局语义特征与局部纹理特征,再进行特征级的互补融合(Zhao等,2025),以同时提升全局语义信息与局部纹理细节的重建质量(如图1(c))。但这两类优化方法仍存在一定的局限性,首先,引入多尺度判别器的方法虽然从不同空间尺度上对生成的光学图像进行判别,提

升了生成图像的视觉保真度与结构完整性,但这种多尺度判别方法与生成器内部功能模块间相脱节,难以对生成器中承担不同任务的特定模块进行定向优化。其次,双分支特征融合方法虽然构建了并行的特征提取通道,分别提取图像的全局语义特征与局部纹理特征,提升了生成图像的特征丰富度,但由于损失函数仅作用于融合后的最终输出,无法对双

分支所提取的全局语义和局部纹理特征分别施加定向约束,导致全局分支可能因网络过度关注局部细节而弱化了全局语义信息的一致性,局部分支也可能因缺乏明确的细节重建约束而丢失精细的纹理结构细节,从而限制了生成图像在全局语义准确性与局部纹理结构真实性两方面的同步提升,制约了生成图像的整体视觉保真度。

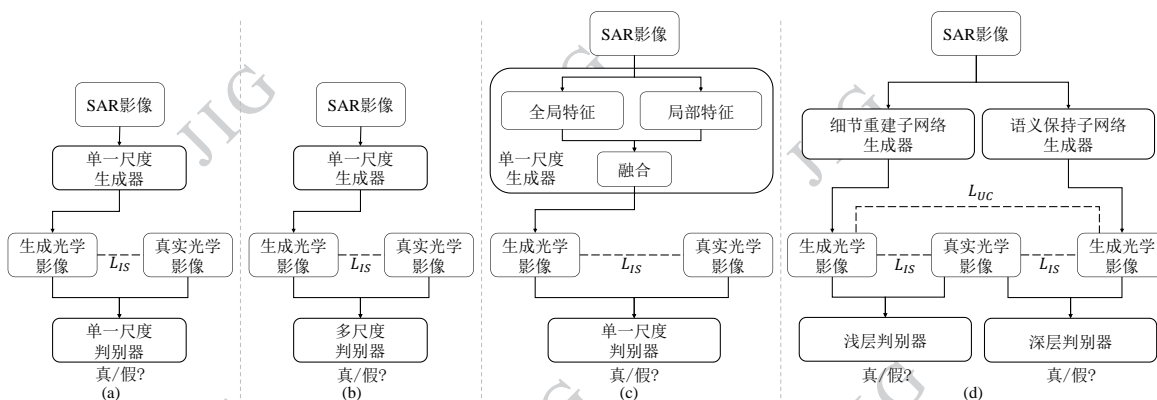


图1 SAR到光学图像转换的条件生成对抗网络框架示意图。(a)基于生成器优化的网络框架;(b)引入多尺度判别器的网络框架;(c)双分支特征融合方法的网络框架;(d)本文方法所采用的网络框架。所有方法均在有监督学习范式下进行,其中 L_{IS} 表示图像相似性损失, L_{UC} 表示单向一致性损失

Fig. 1 Schematic diagram of the conditioned generative adversarial network framework for SAR-to-optical image translation. (a) The network framework based on generator optimization; (b) The network framework incorporating a multi-scale discriminator; (c) The network framework of the dual-branch feature fusion method. (d) The network framework of the proposed method. All methods operate under a supervised learning paradigm, where L_{IS} denotes the image similarity loss and L_{UC} denotes the image consistency loss

因此,为了克服现有单一分支生成对抗网络在全局语义保持与局部纹理结构细节重建之间的内在矛盾,多尺度判别器与生成器内部功能模块间相脱节的难题以及双分支特征融合方法中因缺乏定向约束导致的优化目标冲突。本文提出了一种用于SAR到光学图像双保真转换的单向知识迁移生成对抗网络(UKT-GAN),该模型通过单向一致性损失来构建单向知识迁移通道,将细节重建子网络所承载的结构纹理信息定向迁移至语义保持子网络中,来保证生成图像在全局与局部层面的双重保真。如图1(d)所示,UKT-GAN首先采用双分支生成框架,将局部纹理结构细节重建与全局语义信息保持两个本质迥异的任务分配到细节重建子网络和语义保持子网络中,从结构上克服了单一分支生成对抗网络难以同时兼顾全局语义信息一致性和局部纹理结构细节保真度的固有局限性。并在此基础上,通过采用专有的双分支判别器与任务导向的损失函数,为细节重建子网络配备专注于局部纹理结构判别的浅层

判别器并施加像素级损失约束保证局部纹理结构细节重建的真实性,同时为语义保持子网络配备负责全局语义信息一致性评估的深层判别器并辅以像素级与特征级联合损失约束保证全局语义信息的一致性,从而实现了两个子网络的定向优化,又从根本上解决了多尺度判别器与生成器内部功能模块间相脱节的难题,避免了双分支特征融合方法中因缺乏定向约束导致的优化目标冲突。最后,利用单向一致性损失将细节重建子网络的纹理结构细节重建能力迁移至语义保持子网络,使其生成的光学图像能够兼顾全局语义与局部细节的双重保真性,从而提升语义保持子网络的整体生成质量。

1 研究方法

1.1 网络框架

用于SAR到光学图像转换的UKT-GAN网络框架,如图2所示。该模型采用有监督的学习方法,借

助成对的SAR与光学图像开展训练,主要由并行训练的
细节重建子网络和语义保持子网络组成。这两个子网络均基于GAN(Goodfellow等,2014)架构进

行设计,其中生成器部分采用了差异化的框架设计,而判别器部分虽框架相同,但在卷积层数的具体配置上有所区别。

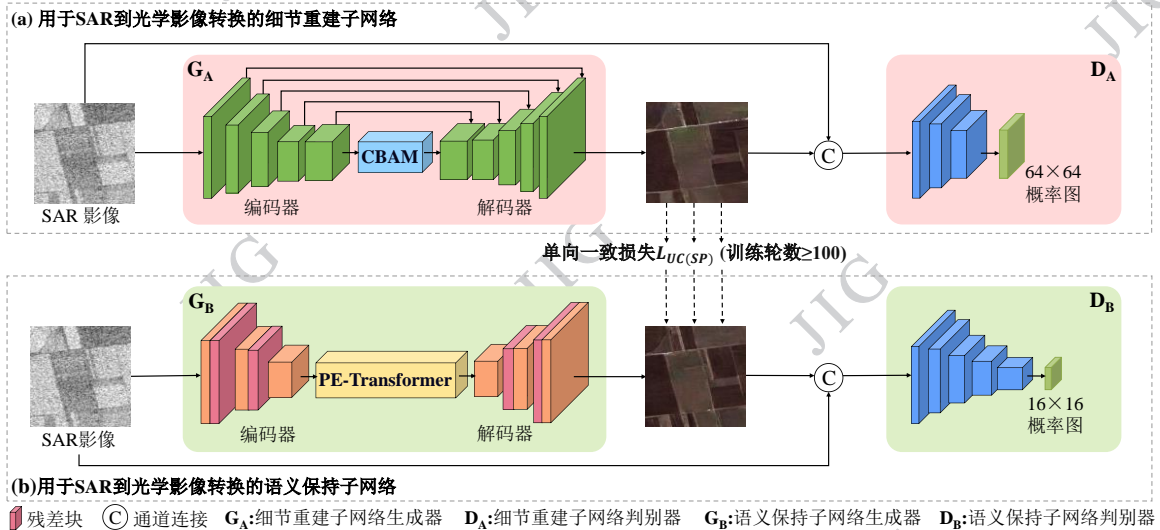


图2 UKT-GAN的总体框架

Fig. 2 Overall framework of the UKT-GAN

在训练过程中,SAR图像数据首先被同步输入到细节重建子网络和语义保持子网络,以生成两组分别关注于图像局部纹理结构细节的光学图像和关注于图像全局语义信息的光学图像。然后,利用具有不同感受视野的判别器分别对生成的两组光学图像进行对抗性判别,以分别引导细节重建子网络与语义保持子网络专注于局部纹理结构细节的重建与全局信息的保持。最后,在细节重建子网络与语义保持子网络分别完成基础能力训练的基础上(训练轮数超过100轮),利用设计的单向一致性损失函数构建基于细节特征约束的单向知识迁移通道,确保语义保持子网络在对抗训练过程中能够持续学习细节重建子网络所提取的精细纹理结构细节,从而促使语义保持子网络获得兼顾全局语义信息与局部纹理结构细节的图像转换能力。

在测试过程中,如图3所示,网络仅使用训练完成的语义保持子网络,来将输入的SAR图像转换为相应的光学图像。关于细节重建子网络与语义保持子网络的详细设计,本文将在第1.2节与第1.3节分别进行系统性阐述。

1.2 用于SAR到光学图像转换的细节重建子网络

细节重建子网络主要关注于捕捉并重建图像的局部纹理细节,为语义保持子网络提供局部纹理结

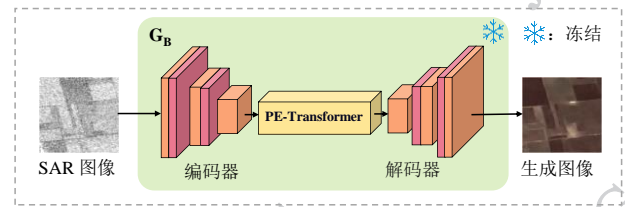


图3 UKT-GAN测试阶段网络流程图

Fig. 3 Network flow chart of UKT-GAN in the testing phase

构优化参考。如图2(a)所示,其主要由一个负责图像转换的生成器和一个用于判别生成图像真伪的判别器构成。

1.2.1 细节重建子网络的生成器

细节重建子网络的生成器采用U型网络(u-shaped network, U-Net)(Ronneberger等,2015)网络,通过跳跃连接将编码器中的特征图与解码器中的特征图进行多尺度融合,从而保留更丰富的空间细节信息。并且为了避免UNet网络瓶颈层中下采样对图像信息的过度压缩,采用了一种渐进式特征扩展策略,在保持特征图空间尺度不变的前提下,仅逐步增加特征的通道维度,来避免信息损失,同时增强特征的表达能力。此外,为进一步提升浅层编解码网络的特征表达能力,我们还利用卷积块注意力(convolutional block attention module, CBAM)(Woo等,2018)模块(如图4)将通道注意力机制和空间注意

力机制嵌入到浅层编解码网络的最内层中,通过通道注意力机制自动学习各特征通道的权重,增强关键特征通道的表达,同时结合空间注意力机制聚焦于特征图中的重要区域,从而进一步提升生成图像的局部纹理结构质量。

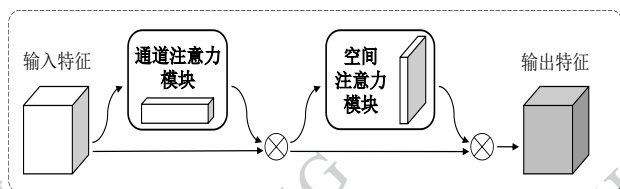


图4 CBAM模块示意图

Fig. 4 Schematic diagram of CBAM module

1.2.2 细节重建子网络的判别器——浅层判别器

细节重建子网络的判别器采用PatchGAN(Isola等,2017)架构,其卷积层数设置为2层,使其具备受限的局部感受视野,专注于判别图像中局部纹理结构细节的真实性。并通过生成器与判别器之间的对抗训练,引导生成器学习生成具有丰富纹理结构细节的光学图像。

1.3 用于SAR到光学图像转换的语义保持子网络

语义保持子网络则主要关注于保持图像的全局语义信息一致性,并利用单向一致性约束,将细节重建子网络对图像局部纹理结构细节的捕获和重建能力迁移到语义保持子网络中,以提升语义保持子网络生成图像的总体质量。如图2(b)所示,其同样由一个执行图像转换的生成器与一个判别图像真伪的判别器组成。

1.3.1 语义保持子网络的生成器

语义保持子网络的生成器采用残差网络(residual network, ResNet)(He等,2016)作为主干网络,旨在利用其残差连接确保语义信息在编码解码过程中稳定传递,保障全局语义信息的完整性。但不同于传统ResNet框架在瓶颈层中简单堆叠残差模块的设计(He等,2016),本方法在瓶颈层中嵌入了带位置编码的Transformer(Vaswani等,2017)编码模块(transformer encoder module with positional encoding, PE-Transformer),来增强网络对全局上下文信息的感知能力,从而保障生成图像在全局尺度上的语义一致性。如图5(a)所示,该模块首先通过可学习的位置编码机制保留特征图的关键空间结构信息,避免因纯注意力操作造成的空间位置信息丢

失,进而确保地物的几何轮廓与相对位置得以保持。随后,利用多头注意力机制在多个独立的表征子空间中并行计算注意力权重,同时捕获不同尺度的长程依赖关系,进一步弥补传统卷积神经网络在全局上下文建模方面的固有局限性。此外,在编码器与解码器的下采样和上采样过程中,我们嵌入了残差模块(如图5(b)所示),确保信息能够在多层网络中稳定传递,从而保障语义特征在跨层传递过程中的完整性并缓解深层网络中梯度消失的问题。

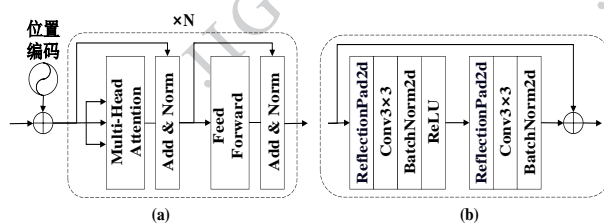


图5 PE-Transformer模块与残差模块示意图。(a)表示PE-Transformer模块;(b)表示残差模块

Fig. 5 Schematic diagram of the PE-Transformer module and the residual module. (a) represents the PE-Transformer module; (b) represents the residual module

1.3.2 语义保持子网络的判别器——深层判别器

语义保持子网络的判别器同样基于PatchGAN架构,但其卷积层数设置为4层,相比细节重建子网络的判别器具有更大的感受野,从而能够有效评估生成图像的整体结构合理性与全局语义一致性。然后通过生成器与判别器之间的对抗训练,引导生成器学习生成在全局尺度上具备更高视觉真实性的光学图像。

1.4 损失函数

所提出的UKT-GAN主要由细节重建子网络和语义保持子网络两部分组成。其总损失函数主要分为细节重建子网络的损失函数、语义保持子网络的损失函数和单向一致性损失函数三类。

在细节重建子网络进行SAR到光学图像转换的训练过程中,其损失函数主要包含图像相似性损失、生成对抗损失两类。其中图像相似性损失为:

$$L_{c_i} = \|y - G_A(x)\|_1 \quad (1)$$

式中, $\|\cdot\|_1$ 表示为1范数, x 表示真实的SAR图像, y 表示真实的光学图像, $G_A(x)$ 表示细节重建子网络生成的光学图像。

细节重建子网络的生成对抗损失设计如下:

$$L_{adv_1} = E_{x \sim p_{optical(x)}} [\log D_A(x, y)] + E_{x \sim p_{optical(x)}} [\log(1 - D_A(x, G_A(x)))] \quad (2)$$

式中, D_A 表示细节重建子网络的判别器(浅层判别器)。

在语义保持网络中,总的损失函数也分图像相似性损失和生成对抗损失两类。在图像的相似性损失方面,为了进一步提高语义保持网络的全局语义保持能力,我们不仅利用 L_l 损失构建内容一致性损失 $L_{content}$,从像素层面保障生成图像的准确性,还引入了特征一致性损失 $L_{feature}$,在深层特征空间保障生成图像的真实性。其图像相似性损失设计如下:

$$L_{content} = \|y - G_B(x)\|_1 \quad (3)$$

$$L_{feature} = \sum_{l=1}^L \frac{1}{H_l W_l} \sum_{h,w} \|\omega_l \odot (y_{hw}^l - G_B(x)_{hw}^l)\|_2^2 \quad (4)$$

$$L_{G_B} = \lambda_1 L_{content} + \lambda_2 L_{feature} \quad (5)$$

式中, $G_B(x)$ 表示语义保持子网络生成的光学图像, $\|\cdot\|_2$ 表示 2 范数, y^l 表示采用视觉几何组(visual geometry group, VGG)(Simonyan 等, 2014)提取的真实光学图像的第 l 层特征图, $G_B(x)^l$ 表示采用 VGG 提取的语义保持子网络生成的光学图像图像的第 l 层特征图, ω_l 表示特征图之间的加权参数, $\lambda_1=100$ 和 $\lambda_2=30$ 分别表示内容一致性损失 $L_{content}$ 和特征一致性损失 $L_{feature}$ 的占比权重。

语义保持子网络的生成对抗损失设计如下:

$$L_{adv_2} = E_{x \sim p_{optical(x)}} [\log D_B(x, y)] + E_{x \sim p_{optical(x)}} [\log(1 - D_B(x, G_B(x)))] \quad (6)$$

式中, D_B 表示语义保持子网络的判别器(深层判别器)。

为了将细节重建子网络对图像局部纹理结构细节的捕获和重建能力迁移到语义保持子网络中,我们设计了细节重建子网络到语义保持子网络的单向一致性损失 $L_{UC(SP)}$,该损失在细节重建子网络与语义保持子网络分别完成基础能力训练的基础上(训练轮数超过 100 轮),通过定向约束语义保持子网络与细节重建子网络输出间的差异,促使语义保持子网络能够持续学习细节重建子网络所提取的精细纹理结构细节。并且为了在消融实验中验证单向知识迁移网络的有效性,我们还设计了将语义保持子网络的全局语义信息保持能力迁移到细节重建子网络中的单向一致性损失 $L_{UC(DR)}$,其同样是在训练轮数超过 100 轮后,进行语义保持子网络到细节重建子网络

的单向知识迁移。这两个单向一致性的损失函数设计如下:

$$L_{UC(SP)} = \|sg[G_A(x)] - G_B(x)\|_1 \quad (7)$$

$$L_{UC(DR)} = \lambda_3 \|G_A(x) - sg[G_B(x)]\|_1 + \lambda_4 \sum_{l=1}^L \frac{1}{H_l W_l} \sum_{h,w} \|\omega_l \odot \{G_A(x)_{hw}^l - sg[G_B(x)]_{hw}^l\}\|_2^2 \quad (8)$$

式中, $sg[\cdot]$ 表示停止梯度操作,模型参数不回传。 $\lambda_3=10$ 和 $\lambda_4=5$ 表示占比权重。

由于 UKT-GAN 是利用单向一致性约束,将细节重建子网络对图像局部纹理结构细节的捕获和重建能力迁移到语义保持子网络中,因此 UKT-GAN 的总损失函数设计如下:

$$L_{total} = \lambda_5 L_{G_A} + \lambda_6 L_{adv_1} + \lambda_7 L_{G_B} + \lambda_8 L_{adv_2} + \lambda_9 L_{UC(SP)} \quad (9)$$

式中, $\lambda_5=100$, $\lambda_6=1$, $\lambda_7=1$, $\lambda_8=1$, $\lambda_9=10$ 依次表示各部分损失函数的占比权重。

2 实验设计

2.1 实验数据

为了验证所提出 UKT-GAN 的有效性,我们在 SEN1-2(Schmitt 等, 2018)和 WHU-OPT-SAR(Li 等, 2022)两个公开数据集上来进行对比实验。对于 SEN1-2 数据集,我们选取了农田、建筑物、山地和森林四种典型地物类型的子数据集(如图 6(a)所示),来验证所提出 UKT-GAN 在不同地物类型下的转换效果。对于 WHU-OPT-SAR 数据集(其地物类型分布见图 6(b)),我们将其作为跨传感器验证的补充数据(与 SEN1-2 数据集的差异详见表 1),并将其原始图像统一裁剪为 256×256 标准尺寸,构建适用于模型训练与测试的配准图像对,以评估 UKT-GAN 在面对不同卫星传感器成像特性差异时,是否仍能保持优异的泛化能力与鲁棒性。

此外,为了评估所提出 UKT-GAN 的有效性,我们基于上述两个数据集采用了 6 种图像转换方法进行对比分析,依次是 Pix2pix(Isola 等, 2017)、DCLGAN(Han 等, 2021)、Parallel-GAN(Wang 等, 2022)、Conditional Diffusion(Bai 等, 2024)、StegoGAN(Wu 等, 2024)和 HVT-cGAN(Zhao 等, 2025)。

2.2 评价指标

为了全面评估 UKT-GAN 和其他 6 种转换方法在 SAR 到光学图像转换中的效果,我们将基于峰值

表1 SEN1-2数据集和WHU-OPT-SAR数据集的参数对照表

Table 1 Comparison of parameters between the SEN1-2 and WHU-OPT-SAR datasets

	光学图像源	SAR图像源	真实图像尺寸	实验图像尺寸	空间分辨率
SEN1-2	Sentinel-2	Sentinel-1	256×256	256×256	10m
WHU-OPT-SAR	GF-1	GF-3	5556×3704	256×256	5m

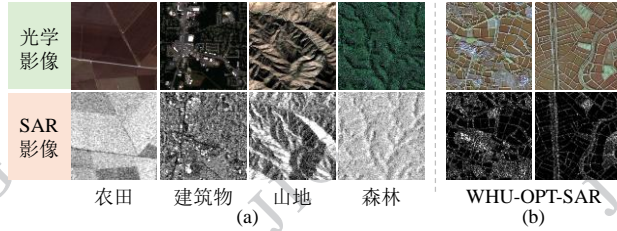


图6 SEN1-2数据集和WHU-OPT-SAR数据集中不同地物类型的图像示例。(a)表示SEN1-2数据集中农田、建筑物、山地和森林四种典型地物类型的图像示例；(b)表示WHU-OPT-SAR数据集中地物类型的图像示例

Fig. 6 Image examples of different land cover types from the SEN1-2 and WHU-OPT-SAR datasets. (a) shows examples of four typical land cover types (farmland, mountain, building, and forest) in the SEN1-2 dataset; (b) presents examples of land cover types in the WHU-OPT-SAR dataset

信噪比 PSNR (Chan 等, 2006)、结构相似性指数 SSIM (Wang 等, 2004)、学习感知图像块相似性 LPIPS (Zhang 等, 2018) 和均方根误差 RMSE (Tanchenko 等, 2014) 四种评价指标, 从全局质量、结构相似性、深层特征保真度和像素级精度四个方面对生成图像的质量进行评估, 以确保对图像转换效果的全面分析。

PSNR: 使用均方误差来判断两幅图像之间的差异, PSNR 的值越高, 表示重建图像的质量越好, 与原始图像越接近, 其计算公式如下:

$$PSNR(y, y') = 10 \log \frac{MAX^2}{MSE} \quad (10)$$

$$= 10 \log \left[\frac{MAX^2 WH}{\sum_{m=0}^{W-1} \sum_{n=0}^{H-1} [y(m, n) - y'(m, n)]^2} \right]$$

式中, y 和 y' 分别表示真实的光学图像和生成的光学图像, MAX 表示合成图像中所能取得的最大灰度值, MSE 表示图像之间的均方误差, W 和 H 表示图像的行数和列数, (m, n) 表示像素点的位置。

$$SSIM(y, y') = [l(y, y')^\alpha \cdot c(y, y')^\beta \cdot s(y, y')^\gamma] \quad (11)$$

$$= \frac{(2\mu_y \mu_{y'} + c_1)(2\sigma_{yy'} + c_2)}{(\mu_y^2 + \mu_{y'}^2 + c_1)(\sigma_y^2 + \sigma_{y'}^2 + c_2)}$$

$$l(y, y') = \frac{2\mu_y \mu_{y'} + c_1}{\mu_y^2 + \mu_{y'}^2 + c_1} \quad (12)$$

$$c(y, y') = \frac{2\sigma_y \sigma_{y'} + c_2}{\sigma_y^2 + \sigma_{y'}^2 + c_2} \quad (13)$$

$$s(y, y') = \frac{\sigma_{yy'} + c_3}{\sigma_y \sigma_{y'} + c_3} \quad (14)$$

式中, μ_y 和 $\mu_{y'}$ 表示 y 和 y' 的均值, σ_y 和 $\sigma_{y'}$ 表示 y 和 y' 的方差, $l(y, y')$ 、 $c(y, y')$ 和 $s(y, y')$ 分别表示图像之间的亮度、对比度和结构, α 和 β 设置为 1。

LPIPS: 通过深度卷积神经网络来提取图像的特征, 并利用加权的 L2 距离来评估两幅图像之间相似度, 其值越低表明两幅图像越相似。计算公式如下:

$$LPIPS(y, y') = \sum_{l=1}^L \frac{1}{H_l W_l} \sum_{h,w} \|\omega_l \odot (y_{hw}^l - y'_{hw}^l)\|_2^2 \quad (15)$$

式中, y^l 和 y'^l 分别表示评估网络中 y 和 y' 图像的第 l 层特征图, ω_l 表示特征图之间的加权参数。

RMSE: 用于评价两幅图像之间的像素值的偏差, 对图像中细微的变化比较敏感, RMSE 的值越低, 表明两幅图像越相似性。其计算公式如下:

$$RMSE(y, y') = \sqrt{\frac{1}{WH} \sum_{m=0}^{W-1} \sum_{n=0}^{H-1} [y(m, n) - y'(m, n)]^2} \quad (16)$$

2.3 实验设置

本方法基于 PyTorch 框架, 在 NVIDIA RTX 3090 单卡环境下实现。生成器和判别器均使用 $\beta_1=0.5$ 和 $\beta_2=0.9$ 的 Adam 来进行优化, 迭代次数为 200, 批量大小设为 1。在网络前 100 个迭代周期内生成器和判别器均以 $lr=0.0002$ 的固定学习率进行训练, 并在后续的 100 个迭代周期内依次递减为 0。

3 实验结果与分析

3.1 定量分析

在 SEN1-2 数据集集中的农田、建筑物、山地和森林四种典型地物类型的子数据集上, 不同方法的定量结果如表 2 和表 3 所示。通过对比不同图像转换方法在 PSNR、SSIM、LPIPS 和 RMSE 四组评价指标上

的表现可以看出,UKT-GAN在农田和山地子数据集上均取得最优结果;在建筑物和森林子数据集上,UKT-GAN在SSIM和LPIPS两项指标上取得最优值,整体展现出卓越的性能优势。这种优势主要可归因于以下两个方面:首先,UKT-GAN采用了基于单向知识迁移的双分支训练框架,将细节重建子网络对图像局部纹理结构细节的捕获和重建能力整合至语义保持子网络,确保了语义保持子网络生成的图像在全局语义信息与局部纹理结构细节上均具备高度的保真度。其次,UKT-GAN采用了多层次优化机制对细节重建子网络和语义保持子网络进行了针对性优化,提高了细节重建子网络和语义保持子网络生成图像的质量。此外,通过对比SEN1-2的4个子数据集的定量结果也可以发现, Pix2pix、Parallel-GAN

和HVT-cGAN同样也具备较优的结果,这主要得益于这些模型采用了一种有监督的图像转换方法,在模型的训练过程中通过真实的光学图像对生成的光学图像添加了图像相似性约束,提高了网络生成图像的质量。相比之下,无监督的图像转换方法(DCLGAN和StegoGAN)通常只能依赖循环一致性损失进行间接性约束,在复杂场景下容易出现细节模糊、语义失真或模态混淆等问题,导致其在定性和定量指标上的表现普遍逊色于有监督的图像转换方法。同时,基于扩散模型的有监督方法Conditional Diffusion,由于存在颜色偏移等问题,在SAR到光学图像转换中的图像生成质量显著低于其他有监督方法,甚至不及部分无监督方法。

表2 SEN1-2中农田与建筑物子数据集上不同方法的定量结果

Table 2 Quantitative results of different methods on the farmland and building subdatasets of the SEN1-2

方法	农田				建筑物			
	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
Pix2pix (2017,CVPR)	17.1571	0.4197	0.4882	35.0384	17.9074	0.3487	0.4653	32.8678
DCLGAN (2021,CVPR)	15.2554	0.3400	0.4750	45.7139	16.2268	0.1970	0.5061	39.2062
Parallel-GAN (2022,TGRS)	18.0499	0.4428	0.4565	31.9261	<u>18.4061</u>	0.3615	0.4897	<u>30.3080</u>
Conditional Diffusion (2024,GRSL)	12.0829	0.1969	0.6102	66.6022	11.3224	0.1559	0.5312	69.4884
StegoGAN (2024,CVPR)	13.9877	0.2759	0.4977	51.8477	15.8222	0.1504	0.5464	41.5423
HVT-cGAN (2025,TGRS)	<u>18.1783</u>	<u>0.4486</u>	<u>0.4484</u>	<u>31.8978</u>	18.5018	<u>0.3645</u>	<u>0.4378</u>	30.1665
UKT-GAN	18.2613	0.4592	0.4345	31.8672	18.3970	0.3654	0.4359	30.3278

注:加粗字体为每行最优值,下划线表示取得次优值。

3.2 定性分析

图7展示了在SEN1-2数据集中的农田、建筑物、山地和森林四种典型地物类型通过不同方法生成的图像可视化结果,通过对比可以发现,UKT-GAN相较于其他6种方法,能够生成与地面真实光学图像纹理结构更加相似的光学图像。并且如图8所示,在图7的红色标记区域内,UKT-GAN相比于其它方法生成的光学图像能够更清晰地呈现不同地物之间的边界细节。

此外,为进一步分析生成图像与真实图像之间的差异性,本文还在SEN1-2和WHU-SAR-OPT两个数据集上对多种方法生成的光学图像和真实光学图像的残差分布进行了可视化。如图9与图10所示,

深蓝色表示生成的光学图像与真实图像的残差接近于零,颜色轴由蓝到黄表示残差结果逐步增大,残差值越小表示生成的光学图像与真实光学图像越相似。通过SEN1-2数据集上生成的光学图像和真实光学图像残差图(图9)之间的对比,可以发现UKT-GAN生成的光学图像和真实光学图像之间的残差图整体色调以蓝色为主,并且在地形复杂区域(图9的第三行和第四行红色方框区域),UKT-GAN对应的残差依然以蓝色为主色调,仅零星分布着少量黄色斑块,表明UKT-GAN生成的光学图像在细节和整体结构上都更贴近于真实光学图像,进一步验证了UKT-GAN在进行SAR到光学图像转换任务时的有效性。

表3 SEN1-2中山地与森林子数据集上不同方法的定量结果

Table 3 Quantitative results of different methods on the mountain and forest subdatasets of the SEN1-2

方法	山地				森林			
	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
Pix2pix (2017,CVPR)	15.5418	0.2659	0.4928	43.5106	16.0789	0.1776	0.4865	40.2478
DCLGAN (2021,CVPR)	12.1411	0.1369	0.5500	58.2652	13.4435	0.1020	0.5701	55.4445
Parallel-GAN (2022,TGRS)	16.0706	0.2733	0.4889	39.5872	16.5887	0.1843	0.4818	37.6306
Conditional Diffusion (2024,GRSL)	13.7014	0.2646	0.4812	52.9071	9.2248	0.1318	0.6886	91.8559
StegoGAN (2024,CVPR)	11.6078	0.0920	0.5340	67.9221	13.0321	0.1187	0.5658	59.1558
HVT-cGAN (2025,TGRS)	<u>16.1592</u>	<u>0.2858</u>	<u>0.4645</u>	<u>39.5563</u>	<u>16.3655</u>	<u>0.1854</u>	<u>0.4678</u>	<u>38.0885</u>
UKT-GAN	16.1979	0.2889	0.4498	39.5325	<u>16.3714</u>	0.1860	0.4565	<u>38.0755</u>

注:加粗字体为每行最优值,下划线表示取得次优值。

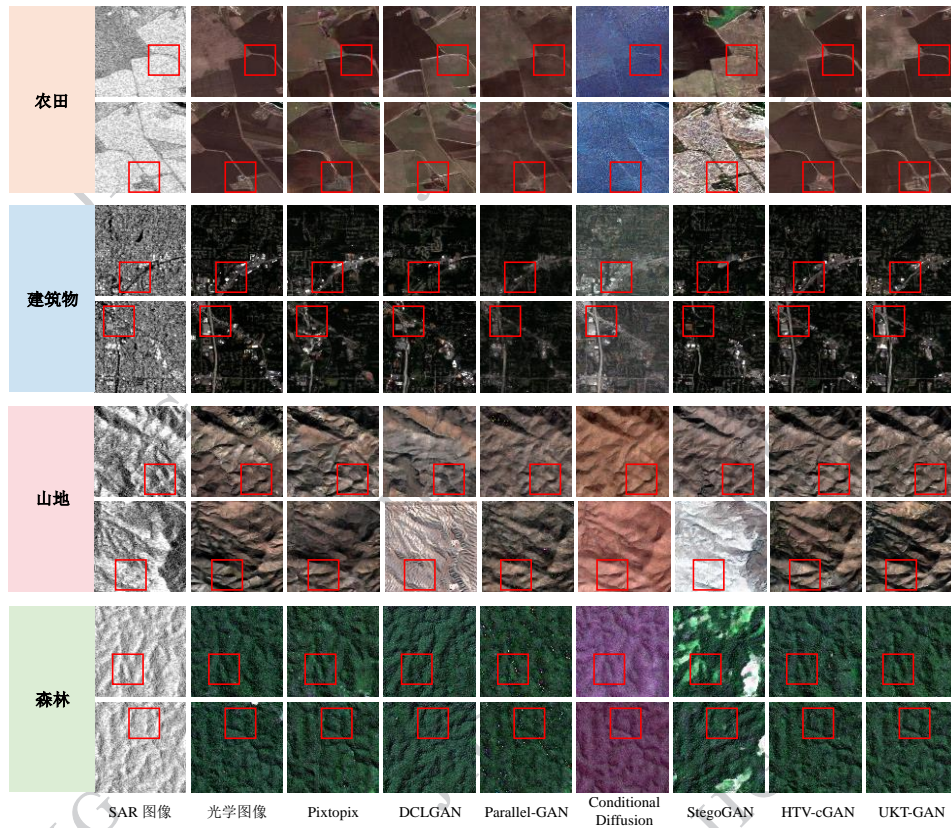


图7 SEN1-2中农田、建筑物、山地和森林四种典型地物测试数据集上不同方法获得的转换图像样本

Fig. 7 Transformed image samples obtained by different methods on the test datasets of four typical land cover types (Farmland, Mountain, Building, and Forest) in SEN1-2

3.3 泛化性分析

为了验证 UKT-GAN 在不同传感器特性与成像配置下的泛化能力,我们基于 WHU-OPT-SAR 数据集开展了进一步的对比实验与分析。不同方法在该数据集上的定量结果、生成图像可视化结果和残差

可视化结果分别如表4和图10所示。由表4可知, UKT-GAN 在 PSNR、SSIM、LPIPS 和 RMAE 四项指标上均优于其他对比方法,表明其在 WHU-OPT-SAR 数据集上仍能生成结构更清晰、整体质量更高的光学图像。此外,从可视化结果(图10)来看,UKT-

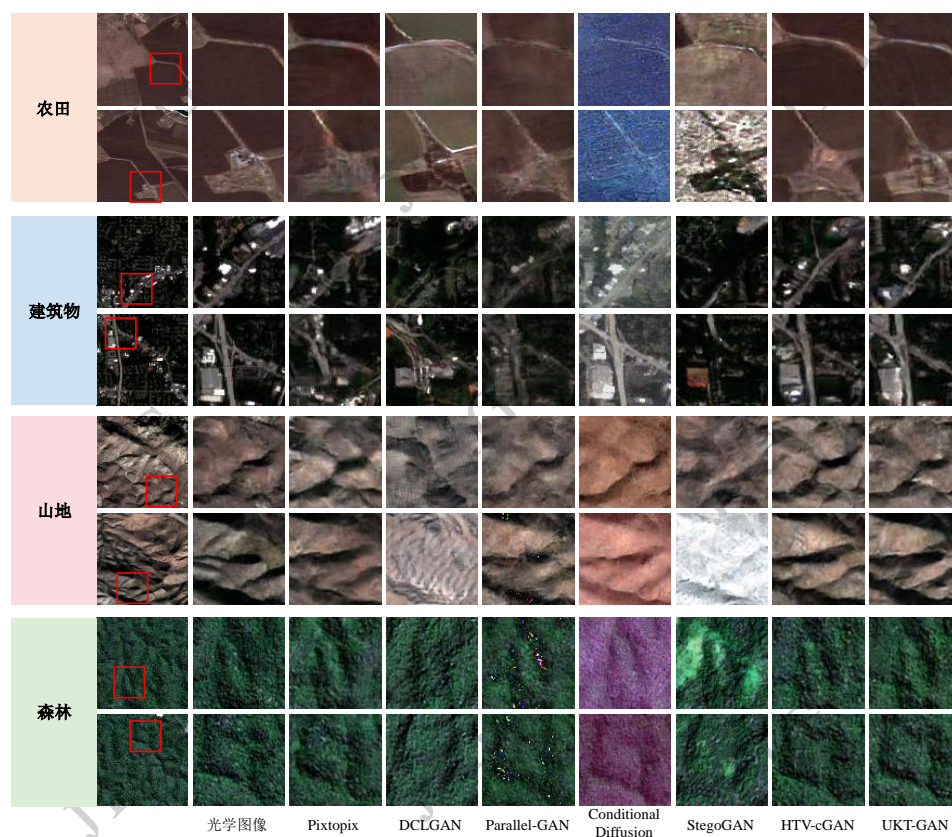


图8 不同方法的转换图像样本在图7红框区域的局部放大对比图

Fig. 8 Local magnified comparison of image translation results by different methods in the red box area of Fig. 7

表4 WHU-SAR-OPT数据集上不同方法的定量结果

Table 4 Quantitative results of different methods on the WHU-SAR-OPT dataset

方法	评价指标			
	PSNR(dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
Pix2pix (2017,CVPR)	15.2067	0.2682	0.4811	44.0448
DCLGAN (2021,CVPR)	14.6328	0.2286	0.4678	46.4971
Parallel-GAN (2022,TGRS)	15.3623	0.2878	0.4651	42.4781
Conditional Diffusion (2024,GRSL)	13.5861	0.2699	0.4556	53.7643
StegoGAN (2024,CVPR)	14.4419	0.2294	0.4755	47.6507
HVT-cGAN (2025,TGRS)	<u>15.3876</u>	<u>0.2894</u>	<u>0.4501</u>	<u>42.2106</u>
UKT-GAN	15.4560	0.2904	0.4393	41.8099

注:加粗字体为每行最优值,下划线表示取得次优值。

GAN生成的光学图像能够更好的保留与真实图像相近的纹理结构细节,并且其生成的光学图像与真实光学图像的残差图呈现出更显著的蓝色分布趋势,表明其生成误差低于用于对比的6种图像转换方法,生成的光学图像在像素级别上与真实图像更为接近。这些结果综合表明,UKT-GAN在面对由不同传感器特性与成像配置所引起的数据分布差异

时,仍能保持优异的性能,具备良好的泛化能力。

3.4 模型复杂度与效率分析

为了全面评估各方法的实用性,本节从模型复杂度与实际推理速度两个维度对所有对比方法进行定量分析。所有方法的参数量(Params)、单次前向传播计算量(FLOPs)以及在相同硬件环境下(RTX 3090,批次大小为1)的单样本平均推理时间(Times)

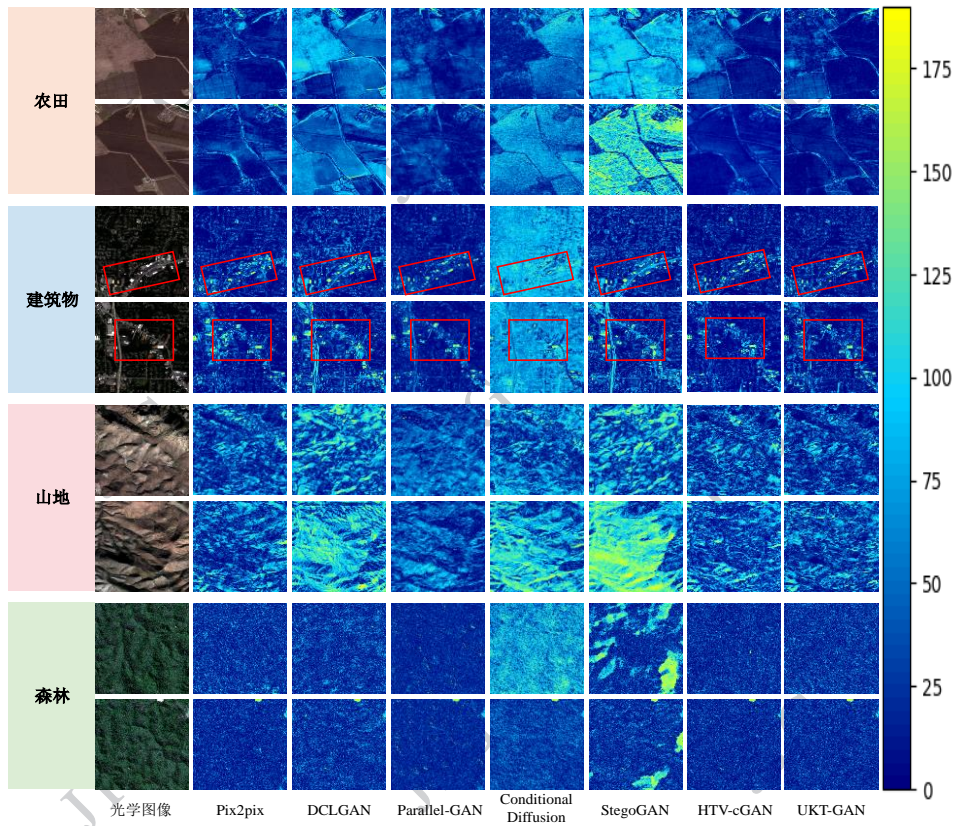


图9 SEN1-2测试数据集上不同方法生成光学图像与真实光学图像残差可视化结果

Fig. 9 Visualization of the residuals results between generated optical images and real optical images by different methods on the SEN1-2 test dataset

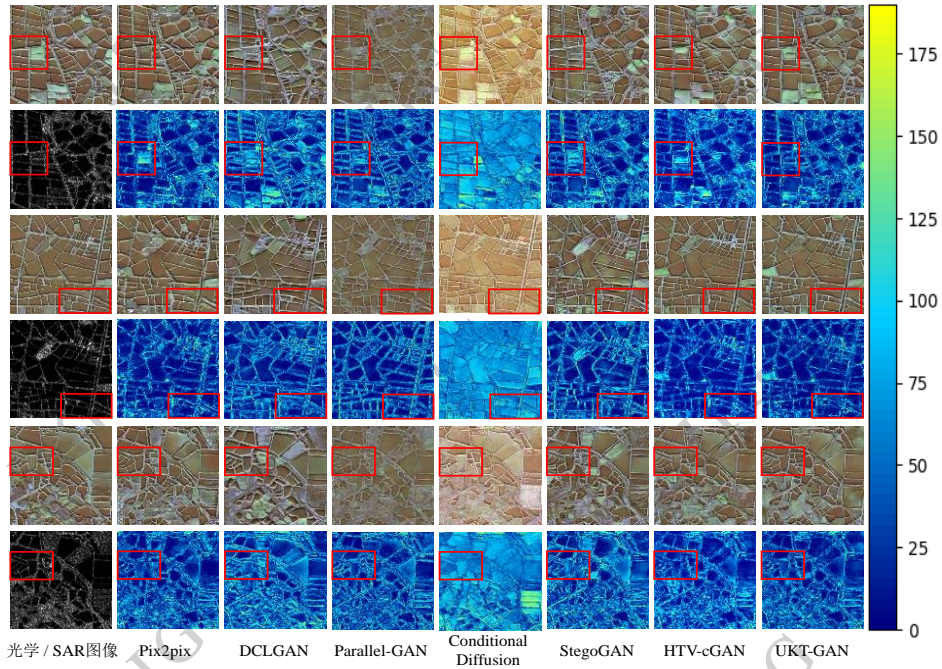


图10 WHU-OPT-SAR测试数据集上不同方法获得的转换图像样本和残差可视化结果

Fig. 10 Transformed image samples and visualization of the residuals results obtained by different methods on the WHU-OPT-SAR test dataset

如表5所示。通过对比可以发现,我们的方法UKT-GAN在参数量、计算量与推理速度之间取得了较好的平衡。相比于图像转换效果相近的Parallel-GAN和HVT-cGAN,UKT-GAN的单次前向传播计算量和

单样本平均推理时间均远低于二者,且参数量显著低于Parallel-GAN。这些结果综合表明,UKT-GAN在保持较低模型复杂度的同时,具备高效的推理性能。

表5 不同图像转换方法的模型复杂度结果

Table 5 Model complexity results of different image translation methods

方法	评价指标		
	Params(M) ↓	FLOPs(G) ↓	Times(MS) ↓
Pix2pix (2017,CVPR)	54.4140	18.1522	1.7902
DCLGAN (2021,CVPR)	11.3782	64.1309	32.0899
Parallel-GAN (2022,TGRS)	532.2004	96.1538	12.7068
Conditional Diffusion (2024,GRSL)	164.2259	332.2115	11278.2193
StegoGAN (2024,CVPR)	11.3782	56.8643	47.0452
HVT-cGAN (2025,TGRS)	11.1685	55.8875	23.6306
UKT-GAN	21.0930	28.1828	5.3421

3.5 消融实验

为了验证UKT-GAN中基于单向知识迁移的双分支生成框架,细节重建子网络和语义保持子网络中CBAM模块和PE-Transformer模块以及特征一致性损失和单向一致性损失对生成图像质量的影响,我们在SEN1-2数据集中的农田子数据集上进行了消融实验。

3.5.1 知识迁移策略的有效性验证

为了验证基于单向知识迁移的双分支生成框架的有效性,我们将单独训练的细节重建子网络、单独训练的语义保持子网络、基于双向知识迁移(损失函数未设置停止梯度操作)的细节重建子网络、基于双向知识迁移的语义保持子网络、基于单向知识迁移的细节重建子网络以及基于单向知识迁移的语义保持子网络(UKT-GAN)在SEN1-2数据集中的农田子数据集上的定量结果记录于表6。通过对比可以发现,在双向知识迁移机制下,细节重建子网络与语义保持子网络会趋于一致,从而导致细节重建子网络转换效果的提升,语义保持子网络转换效果的下降。而在单向知识迁移机制下,基于单向知识迁移的细节重建子网络能够成功嵌入语义保持子网络对全局语义信息的保持能力,基于单向知识迁移的语义保持子网络也能够成功嵌入细节重建子网络对局部纹理结构细节的生成能力,从而使得生成图像的质量均得到了有效提升。但由于基于单向知识迁移的语

义保持子网络拥有更好的图像转换效果,因此本文采用基于单向知识迁移的语义保持子网络(UKT-GAN)作为核心主干网络,以确保获得兼具全局语义信息一致性与丰富局部纹理结构细节的光学图像输出。

3.5.2 细节重建子网络和语义保持子网络中关键模块的有效性验证

为了验证细节重建子网络和语义保持子网络中CBAM模块和PE-Transformer模块对生成图像质量的影响,我们在单独训练的细节重建子网络和单独训练的仅采用内容损失和生成对抗损失的语义保持子网络(基础语义保持子网络)上进行了对比实验。通过表7可以发现,在细节重建子网络中引入CBAM模块后,PSNR、SSIM、LPIPS和RMSE四项指标均获得全面提升,表明该注意力机制通过通道与空间维度的自适应加权,能够有效增强细节重建子网络对局部纹理与细节的特征表达能力。而在语义保持子网络中集成PE-Transformer模块后,生成的光学图像在PSNR、SSIM和RMSE三项指标上取得了提升,而在LPIPS指标上表现欠佳。表明通过在语义保持网络中嵌入PE-Transformer模块能够有效建立长程依赖关系,能够显著改善生成图像的全局结构合理性与像素级精度,但牺牲了部分深层特征保真度。

3.5.3 细节重建子网络中不同注意力机制的对比

表6 不同知识迁移损失函数配置下细节重建子网络与语义保持子网络在农田子数据集上的定量对比

Table 6 Quantitative comparison of detail reconstruction and semantic preservation subnetworks with different knowledge transfer loss configurations on the farmland subdataset

方法	评价指标			
	PSNR(dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
细节重建子网络	17.8269	0.4430	0.4614	33.2151
语义保持子网络	<u>18.1140</u>	<u>0.4537</u>	<u>0.4353</u>	<u>31.9020</u>
基于双向知识迁移 $L_{UC(SP)}$ 的细节重建子网络	17.8428	0.4440	0.4600	33.1388
基于双向知识迁移 $L_{UC(SP)}$ 的语义保持子网络	18.0721	0.4514	0.4432	32.2927
基于双向知识迁移 $L_{UC(DR)}$ 的细节重建子网络	17.9044	0.4427	0.4555	33.0831
基于双向知识迁移 $L_{UC(DR)}$ 的语义保持子网络	18.0533	0.4504	0.4453	32.3706
基于单向知识迁移 $L_{UC(DR)}$ 的细节重建子网络	17.8700	0.4446	0.4531	33.1129
UKT-GAN (基于单向知识迁移 $L_{UC(SP)}$ 的语义保持子网络)	18.2613	0.4592	0.4345	31.8672

注:加粗字体为每行最优值,下划线表示取得次优值。

表7 细节重建子网络和语义保持子网络有无加入CBAM模块和PE-Transformer模块对生成图像质量的影响

Table 7 Impact of integrating CBAM and PE-Transformer modules in the detail reconstruction and semantic preservation subnetworks on generated image quality

方法	模块配置	评价指标			
		PSNR(dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
细节重建子网络	(w/o CBAM)	17.7135	0.4421	0.4641	33.3293
	(w/ CBAM)	17.8269	0.4430	0.4614	33.2151
基础语义保持子网络	(w/o PE-Transformer)	17.6899	0.4306	0.4551	34.0015
	(w/ PE-Transformer)	17.8838	0.4442	0.4606	32.8948

分析

为了进一步验证细节重建子网络中选择CBAM的合理性,我们在保持细节重建子网络其他配置不变的情况下,分别将CBAM替换为压缩激励网络(squeeze-and-excitation network, SE)(Hu等, 2018)、高效通道注意力(efficient channel attention, ECA)(Wang等, 2020)和坐标注意力(coordinate attention, CA)(Hou等, 2021)进行对比实验,结果如表8所示。实验结果显示,CBAM在PSNR、RMSE和SSIM指标上优于SE、ECA和CA,在LPIPS指标上取得次优值。这些结果综合表明,CBAM通过串联通道注意力与空间注意力,实现了特征图在通道和空间维度的双重增强,相比于SE、ECA和CA,能够更加充分的保留生成图像的纹理与结构信息,对于细节子网络具备更优的适配性,进一步验证了其在细节重建中的有效性。

表8 细节重建子网络中不同注意力机制的对比结果

Table 8 Comparison results of different attention mechanisms in the detail reconstruction subnetwork

注意力模块	评价指标			
	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
Base(无注意力)	17.7135	0.4421	0.4641	33.3293
SE	17.7010	0.4426	0.4615	33.4040
ECA	17.8061	0.4421	0.4621	33.2191
CA	<u>17.8200</u>	<u>0.4429</u>	0.4597	<u>33.2168</u>
CBAM	17.8269	0.4430	<u>0.4614</u>	33.2151

注:加粗字体为每行最优值,下划线表示取得次优值。

3.5.4 特征一致性损失函数和单向一致性损失函数设计的有效性验证

为了提升语义保持网络对深层特征的保真度和整合细节重建子网络对纹理结构细节的捕获能力,

我们引入了基于特征感知相似性的特征一致性损失 $L_{feature}$ 与面向知识迁移的单向一致性损失 $L_{UC(SP)}$, 并其定量结果记录在表9中。实验结果表明, 在语义保持子网络中加入特征一致性损失 $L_{feature}$ 后, 语义保持子网络在PSNR、SSIM、LPIPS与RMSE四项指标上均获得了提升, 并在LPIPS指标上提升显著(从0.4606降至0.4353), 表明 $L_{feature}$ 通过约束语义保持子网络输出与真实光学图像在深度特征空间上的一

致性, 能有效的增强生成图像的深层特征保真度。当进一步引入单向一致性损失 $L_{UC(SP)}$ 后, 语义保持子网络在PSNR、SSIM、LPIPS和RMSE指标上得到了提升, 表明通过加入单向一致性损失 $L_{UC(SP)}$ 将细节重建子网络对局部纹理结构细节的捕获与重建能力迁移至语义保持子网络, 也有助于进一步提高生成图像的总质量, 证明了单向一致性损失 $L_{UC(SP)}$ 设计的有效性。

表9 特征一致性损失与单向一致性损失对生成图像质量的影响

Table 9 Impact of feature consistency loss and unidirectional consistency loss on generated image quality

方法	评价指标			
	PSNR(dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
基础语义保持子网络	17.8838	0.4442	0.4606	32.8948
基础语义保持子网络+ $L_{feature}$	<u>18.1140</u>	<u>0.4537</u>	<u>0.4353</u>	<u>31.9020</u>
UKT-GAN (基础语义保持子网络+ $L_{feature}$ + $L_{UC(SP)}$)	18.2613	0.4592	0.4345	31.8672

注: 加粗字体为每行最优值, 下划线表示取得次优值。

表10 特征一致性损失权重 λ_2 不同取值的对比结果Table 10 Comparison results of feature consistency loss weight λ_2 with different values

λ_2 取值	评价指标			
	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
$\lambda_2=0$	17.8838	0.4442	0.4606	32.8948
$\lambda_2=10$	18.0918	0.4489	0.4421	32.1768
$\lambda_2=20$	18.0364	<u>0.4531</u>	0.4362	32.0767
$\lambda_2=30$	18.1140	0.4537	0.4353	31.9020
$\lambda_2=40$	<u>18.0365</u>	0.4521	<u>0.4359</u>	<u>31.9489</u>

注: 加粗字体为每行最优值, 下划线表示取得次优值。

表11 单向一致性损失权重 λ_0 不同取值的对比结果Table 11 Comparison results of unidirectional consistency loss weight λ_0 with different values

λ_0 取值	评价指标			
	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	RMSE ↓
$\lambda_0=0$	18.1140	0.4537	0.4353	31.9020
$\lambda_0=5$	<u>18.2546</u>	<u>0.4566</u>	0.4349	<u>31.8817</u>
$\lambda_0=10$	18.2613	0.4592	0.4345	31.8672
$\lambda_0=15$	18.1529	0.4528	<u>0.4348</u>	31.9092
$\lambda_0=20$	18.1181	0.4514	0.4393	31.9586

注: 加粗字体为每行最优值, 下划线表示取得次优值。

3.5.5 特征一致性损失函数和单向一致性损失函数不同权重对生成图像质量的影响

1) 不同特征一致性损失函数权重对语义子网络生成图像质量的影响。

为验证不同特征一致性损失函数权重对语义子网络生成图像质量的影响, 我们在保证语义子网络其他配置不变的情况下, 对特征一致性损失权重 λ_2 的取值分别设置为0、10、20、30、40进行对比实验, 其定量结果如表10所示。实验结果表明, $\lambda_2=30$ 时模型在PSNR、SSIM、LPIPS和RMSE指标上取得最佳平衡。当 $\lambda_2=0$ 时, 语义子网络缺乏特征一致性约束, LPIPS指标较高, 感知质量有限; 当 $\lambda_2=40$ 时, 感知损失过度主导优化, 导致像素级精度下降。这些结果表明, 当特征一致性损失函数权重 $\lambda_2=30$ 时, 模型能够有效平衡像素级精度与感知质量, 因此本文采用 $\lambda_2=30$ 作为特征一致性损失函数的权重。

2) 不同单向一致性损失函数权重对总网络UKT-GAN生成图像质量的影响。

为验证不同单向一致性损失函数权重对总网络UKT-GAN生成图像质量的影响, 我们在保证总网络其他配置不变的情况下, 对单向一致性损失权重 λ_0 的取值分别设置为0、5、10、15、20进行对比实验, 其定量结果如表11所示。实验结果表明, $\lambda_0=10$ 时模型在各项指标上均取得最优结果。当 $\lambda_0=0$ (即无单

向一致性迁移)时,模型性能明显下降,验证了单向一致性损失的有效性;当 λ_0 取值过大(15或20)时,性能出现衰减,表明过高的迁移权重会破坏语义保持子网络的原有语义一致性。这些结果表明, $\lambda_0=10$ 作为适中权重,能够有效平衡迁移强度与语义保持稳定性,因此本文采用 $\lambda_0=10$ 作为单向一致性损失函数的权重。

4 结论

针对现有基于生成对抗网络的SAR到光学图像转换方法采用单一的生成器,难以同时兼顾全局语义信息一致性与局部纹理结构细节真实性的问题,本文提出了一种用于SAR到光学图像双保真转换的单向知识迁移生成对抗网络UKT-GAN。该网络由分工明确的细节重建子网络和语义保持子网络组成,其中,细节重建子网络采用嵌入CBAM模块的U-Net生成器、浅层判别器以及像素级的定向约束,专注于局部纹理结构细节的重建。语义保持子网络则采用嵌入PE-Transformer模块的ResNet生成器,深层判别器以及像素级和特征级的定向约束保障生成图像的全局语义一致性,并通过单向一致性损失将细节重建子网络对局部纹理结构细节的重建能力迁移到语义保持子网络中,从而实现在测试中仅部署单一语义保持子网络的情况下,生成同时兼具全局语义信息与局部纹理结构细节的光学图像。为了全面评估UKT-GAN的性能,本文还选取6种主流的图像转换方法作为对比基准。在SEN1-2数据集中的农田、建筑物、山地和森林四种典型地物类型的子数据集以及WHU-OPT-SAR数据集上,进行了定量评估与定性分析。实验结果表明,我们所提出的UKT-GAN在面对不同地物类型时,相比于其他方法,能够生成结构更清晰、总体质量更好的光学图像;在面对由不同传感器特性与成像配置所引起的数据分布差异时,UKT-GAN仍能保持稳定的转换性能,具备良好的泛化能力。

参考文献(References)

- Toriya H, Dewan A and Kitahara I. 2019. SAR2OPT: Image alignment between multi-modal images using generative adversarial networks// Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan: IEEE: 923-926 [DOI: 10.1109/IGARSS.2019.8898605]
- Li X H, Du Z S, Huang Y Y and Tan Z Y. 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. ISPRS Journal of photogrammetry and remote sensing, 179: 14 - 34 [DOI: 10.1016/j.isprsjprs.2021.07.007]
- Sun Y C, Jiang W, Yang J Y and Li W Z. 2022. SAR target recognition using cGAN-based SAR-to-optical image translation. Remote Sensing, 14(8): 1793 [DOI: 10.3390/rs14081793]
- Hu X K, Zhang P Z, Ban Y F and Rahmehoonfar M. 2023. GAN-based SAR and optical image translation for wildfire impact assessment using multi-source remote sensing data. Remote Sensing of Environment, 289: 113522 [DOI: 10.1016/j.rse.2023.113522]
- Kwak G H and Park N W. 2024. Assessing the potential of multi-temporal conditional generative adversarial networks in SAR-to-optical image translation for early-stage crop monitoring. Remote Sensing, 16(7): 1199 [DOI: 10.3390/rs16071199]
- Du W L, Guo B, Zhao J Q, Yao R and Zhou Y. 2026. Semantic-guided contrastive learning for SAR and optical image translation. Journal of Image and Graphics, 1-16 (杜文亮, 郭波, 赵佳琦, 姚睿, 周勇. 2026. 语义引导对比学习的SAR与光学图像转换. 中国图象图形学报, 1-16) [DOI: 10.11834/jig.250526]
- Yu P L, Shi Q and Wang H. 2021. Infrared-to-visible image translation based on parallel generator network. Journal of Image and Graphics, 26(10): 2346-2356 (余佩伦, 施佳, 王晗. 2021. 并行生成网络的红外-可见光图像转换. 中国图象图形学报, 26(10): 2346-2356) [DOI: 10.11834/jig.200113]
- Guo J, He C Y, Zhang M J, Li Y S, Gao X B and Song B Y. 2021. Edge-preserving convolutional generative adversarial networks for SAR-to-optical image translation. Remote Sensing, 13(18): 3575 [DOI: 10.3390/rs13183575]
- Weng Y C, Ma Y, Chen F, Shang E P, Yao W T, Zhang S Yet al. 2023. Temporal co-attention guided conditional generative adversarial network for optical image synthesis. Remote Sensing, 15(7): 1863 [DOI: 10.3390/rs15071863]
- Turnes J N, Castro J D B, Torres D L, Vega P J, Feitosa R Q and Happ P N. 2022. Atrous cGAN for SAR to optical image translation. IEEE Geoscience and Remote Sensing Letters, 19: 1-5 [DOI: 10.1109/LGRS.2020.3031199]
- Chen Y K, Zhu Z G, Huang Y, Wang P, Huang B and Dalla Mura M. 2024. MSF: A multi-scale fusion generative adversarial network for SAR-to-optical image translation//Proceedings of 2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Athens, Greece: IEEE: 9058-9061 [DOI: 10.1109/IGARSS.53475.2024.10640486]
- Yang X, Zhao J Y, Wei Z Y, Wang N N and Gao X B. 2022. SAR-to-optical image translation based on improved CGAN. Pattern Recognition, 121: 108208 [DOI: 10.1016/j.patcog.2021.108208]
- Wei J, Zou H X, Sun L, Cao X, He S T, Liu S and Zhang Y Q. 2023.

- CFRWD-GAN for SAR-to-optical image translation. *Remote Sensing*, 15(10): 2547 [DOI: 10.3390/rs15102547]
- Zhao W B, Jiang N N, Liao X X and Zhu J B. 2025. Hvt-cgan: Hybrid vision transformer cgan for SAR-to-optical image translation. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 5202017 [DOI: 10.1109/TGRS.2024.3523040]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//*Proceedings of the 2014 Advances in Neural Information Processing Systems (NIPS)*. Montreal, Canada: ACM: 2672-2680
- Ronneberger O, Fischer P and Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//*Proceedings of 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Berlin, Germany: Cham:9351: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. Cbam: Convolutional block attention module//*Proceedings of 2018 European Conference on Computer Vision (ECCV)*. Munich, Germany: Cham:11211: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//*Proceedings of 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Identity mappings in deep residual networks//*Proceedings of 2016 European Conference on Computer Vision (ECCV)*. Amsterdam, Netherlands: Cham: 9908: 630-645 [DOI: 10.1007/978-3-319-46493-0_38]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//*Proceedings of the 2017 Annual Conference on Neural Information Processing Systems (NIPS)*. Long Beach, CA, USA: ACM: 630-645
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2014-09-04]. <https://arxiv.org/pdf/1409.1556.pdf>
- Schmitt M, Hughes L H and Zhu X X. 2018. The SEN1-2 dataset for deep learning in SAR-optical data fusion//*Proceedings of 2018 ISPRS TC I Mid-term Symposium on Innovative Sensing - From Sensors to Methods and Applications*. Karlsruhe, Germany: Cham: 4-1: 141-146 [DOI: 10.5194/isprs-annals-IV-1-141-2018]
- Li X, Zhang G, Cui H, Hou S S, Wang S Y, Li X et al. 2022. MCANet: A joint semantic segmentation framework of optical and SAR images for land use classification. *International Journal of Applied Earth Observation and Geoinformation*, 106: 102638 [DOI: 10.1016/j.jag.2021.102638]
- Han J L, Shoeiby M, Petersson L and Armin M A. 2021. Dual contrastive learning for unsupervised image-to-image translation//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE: 746-755 [DOI: 10.1109/CVPRW53098.2021.00084]
- Wang H X, Zhang Z G, Hu Z Y and Dong Q L. 2022. SAR-to-optical image translation with hierarchical latent features. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5233812 [DOI: 10.1109/TGRS.2022.3200996]
- Bai X Y, Pu X Y and Xu F. 2024. Conditional diffusion for SAR to optical image translation. *IEEE Geoscience and Remote Sensing Letters*, 21: 4000605 [DOI: 10.1109/LGRS.2023.3337143]
- Wu S D, Chen Y, Mermel S, Hurni L, Schindler K, Gonthier N et al. 2024. Stegogan: Leveraging steganography for non-bijective image-to-image translation//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE: 7922-7931 [DOI: 10.1109/CVPR52733.2024.00757]
- Chan T F, Shen J and Zhou H M. 2006. Total variation wavelet inpainting. *Journal of Mathematical Imaging and Vision*, 25(1): 107-125 [DOI: 10.1007/s10851-006-5257-3]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]
- Tanhenko A. 2014. Visual-PSNR measure of image quality. *Journal of Visual Communication and Image Representation*, 25(5): 874-878 [DOI: 10.1016/j.jvcir.2014.01.008]
- Hu J, Shen L and Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W H and Hu Q H. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]
- Hou Q B, Zhou D Q and Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

作者简介

张智豪,男,硕士研究生,主要研究方向为异源遥感图像转换。E-mail:zhang_zhi_hao@stu.kust.edu.cn

付志涛,通信作者,男,博士,副教授,硕士生导师,主要研究方向为深度学习方法及其应用、异构遥感图像匹配、多源遥感图像融合及应用、三维重建及其可视化。E-mail:zhi-taofu@kust.edu.cn

嵇娅帅,男,硕士研究生,主要研究方向为图像与点云配准。

E-mail:jiyashuai@stu.kust.edu.cn

张歆山,男,硕士研究生,主要研究方向为异源图像匹配。E-mail:xinshanz@163.com

唐伯惠,男,博士,教授,中科院地理资源所研究员,博士生导师,主要研究方向为地表参数的定量反演、地表净辐射遥感估算、高原遥感。E-mail:tangbh@kust.edu.cn