中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-17

论文引用格式:Xing Yinan, Liu Bo, Zhang Yunfeng, Ren Yuehe. A spatial-frequency strongly-sparse guided diffusion model for S2O image translation[J/OL]. Journal of Image and Graphics, XXXX:1-17. DOI: 10.11834/jig.260188. (邢怡楠, 刘博, 张云峰, 任玥赫. 面向S2O图像翻译的空频强稀疏引导扩散模型[J/OL]. 中国图象图形学报,XXXX:1-17. DOI: 10.11834/jig.260188.)[DOI:10.11834/jig.260188]

面向S2O图像翻译的空频强稀疏引导扩散模型

邢怡楠,刘博,张云峰,任玥赫

山东财经大学计算机与人工智能学院,山东省济南市　250014

摘　要:**目的**　合成孔径雷达(Synthetic Aperture Radar,SAR)至光学图像(SAR-to-Optical,S2O)翻译是实现全天候对地观测的关键技术。现有方法面临两大瓶颈:一是未能有效解耦SAR乘性相干斑噪声与地物信号,导致跨模态特征融合与表达能力不足;二是主流扩散生成模型计算代价高昂,难以满足在资源受限遥感平台下的实时处理需求。针对上述问题,本文提出一种空频强稀疏引导的扩散模型(Spatial-Frequency Strongly-Sparse Guided Diffusion Model,SFSG-Diff),旨在实现更稳健高效的S2O转换。**方法**　设计多尺度空频去噪编码(Multi-scale Spatial-Frequency Denoising Encoder,MDE),利用空频域特征互补性显式分离噪声与有效信号,抑制噪声并增强地物结构表达;提出轻量化的强稀疏语义融合(Strong-Sparse Semantic Fusion,SIF),仅对部分特征流高效融合,以低计算代价实现多尺度特征精准引导;采用两阶段训练策略,融合感知、聚焦频率与对抗损失联合优化。**结果**　在SEN1-2,QXS-SAROPT和WHU-OPT-SAR三个数据集上的实验表明,本方法在峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)、结构相似度(Structural Similarity Index,SSIM)、学习感知图像块相似度(Learned Perceptual Image Patch Similarity,LPIPS)和弗雷歇初始距离(Fréchet Inception Distance,FID)上均取得最优的结果,其中在SEN1-2数据集上与次优结果相比PSNR与SSIM分别提升0.77dB与17.8%,LPIPS与FID分别降低8.0%和17.6%。模型参数量少、计算复杂度低,单次推理仅需0.21秒,较同类扩散模型最优结果提速约69.1%,效率接近传统生成对抗网络。**结论**　SFSG-Diff可有效抑制SAR斑点噪声,实现跨模态高质量图像生成,兼顾性能与计算效率,适用于计算资源受限的遥感平台,为实时SAR图像处理提供可行方案。

关键词:SAR至光学图像翻译;扩散模型;特征融合;遥感图像;多尺度编码;强稀疏引导

A spatial-frequency strongly-sparse guided diffusion model for S2O image translation

Xing Yinan, Liu Bo, Zhang Yunfeng, Ren Yuehe

School of Computer Science and Artificial Intelligence, Shandong University of Finance and Economics, Jinan City, Shandong Province 250014, China

Abstract: **Objective** Synthetic Aperture Radar (SAR) to optical image translation (SAR-to-Optical, S2O) is essential for full-time and all-weather earth observation. Optical images provide rich texture information and intuitive visual representations, which are important to a wide range of downstream tasks such as land-cover classification and urban monitoring. However, optical sensors are highly vulnerable to clouds, fog, and illumination changes, limiting continuous data acquisition. In contrast, SAR sensors operate in the microwave spectrum and can acquire images under all-weather and all-time

收稿日期:2026-04-10;**修回日期**:2026-06-03

基金项目:山东省自然科学基金(批准号:ZR2025QC1629)

Supported by: Project supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2025QC1629)

conditions. Consequently, S2O translation can significantly enhance the interpretability of SAR data and facilitate multi-modal remote sensing applications. Despite recent progress in cross-modal image translation, achieving high-quality S2O translation remains challenging. First, SAR images inevitably contain multiplicative speckle noise that is strongly coupled with structural information, and together with the substantial semantic gap caused by fundamentally distinct imaging mechanisms between SAR and optical sensors, this leads to insufficient cross-modal feature fusion and representation capability. Second, although diffusion models have recently demonstrated impressive performance in image generation tasks, their high computational overhead and inefficient conditional fusion mechanisms limit their applicability for real-time processing on resource-limited remote sensing platforms. To address these issues, we propose a Spatial-Frequency Strongly-Sparse Guided Diffusion Model (SFSG-Diff) for efficient and high-quality SAR-to-optical image translation. **Method** The proposed SFSG-Diff framework follows the standard diffusion generation paradigm, consisting of a forward diffusion process and a reverse denoising process. In the forward process, Gaussian noise is progressively added to a clean optical image until the data distribution approaches pure noise. During the reverse process, a conditional denoising network iteratively removes the noise to reconstruct the target optical image. To improve both generation quality and computational efficiency, SFSG-Diff introduces spatial-frequency guided feature extraction and a strongly sparse conditional fusion mechanism. To effectively suppress SAR speckle noise and enhance structural representation, a Multi-scale Spatial-Frequency Denoising Encoder (MDE) is designed to extract robust conditional features from the input SAR image. The encoder adopts a dual-branch architecture composed of a spatial feature extraction branch and a frequency feature extraction branch. The spatial branch focuses on capturing local textures and contextual information through multi-scale convolutional operations, while the frequency branch transforms SAR images into the frequency domain to capture global structural information that is less sensitive to speckle noise. Since speckle noise is mainly concentrated in high-frequency components, whereas meaningful structural information is typically distributed in low and mid-frequency regions, the spatial-frequency joint representation enables effective separation of noise and structural features. The outputs of both branches are fused across multiple scales to produce robust conditional representations that guide the diffusion model during image generation. To further improve conditional guidance efficiency, we introduce a lightweight Strong-Sparse Semantic Fusion (SIF) pattern. Instead of densely integrating conditional features into all feature channels, the SIF module performs channel-wise selection and adaptive attention-based modulation to identify and fuse the most informative feature components. This sparse guidance mechanism not only reduces computational overhead but also improves cross-modal semantic alignment by emphasizing the most relevant structural cues. Furthermore, to enhance training stability and generation quality, we adopt a two-stage training strategy, leveraging a joint loss, including a simplified mean squared error loss, perceptual loss, focus frequency loss, and adversarial loss. **Result** To evaluate the performance of the proposed framework, extensive experiments are conducted on three publicly available datasets, including SEN1-2, QXS-SAROPT, and WHU-OPT-SAR. The experimental results demonstrate that SFSG-Diff consistently outperforms several state-of-the-art GAN-based and diffusion-based image translation models. Quantitative evaluation results demonstrate that SFSG-Diff consistently outperforms existing state-of-the-art methods across multiple evaluation metrics. On the SEN1-2 dataset, the proposed SFSG-Diff achieves a PSNR improvement of 0.77 dB over the best competing method, while the Structural Similarity Index (SSIM) improves by 17.8%. In addition, the perceptual quality metrics show substantial improvements, with LPIPS and Fréchet Inception Distance (FID) reduced by 8.0% and 17.6%, respectively. These results indicate that the proposed model not only improves reconstruction fidelity but also produces images with higher perceptual realism. Similar performance gains are observed on the QXS-SAROPT and WHU-OPT-SAR datasets, which include more complex urban structures and higher spatial resolution imagery. The proposed SFSG-Diff demonstrates strong robustness across diverse scenes and maintains consistent advantages in both structural similarity and perceptual quality. Visual comparisons further confirm the superiority of SFSG-Diff. Compared with existing methods, the SFSG-Diff generates optical images with clearer structural boundaries, more realistic textures, and fewer noise artifacts. In particular, the model exhibits improved performance in challenging regions such as dense urban areas and vegetation-rich scenes. In addition to generation quality, the computational efficiency of the proposed framework is also evaluated. Owing to the lightweight SIF pattern and the two-stage training strategy, SFSG-Diff achieves significantly faster inference compared with conventional diffusion models. The average inference time per image is approximately 0.21

seconds，which represents a 69. 1% reduction in computation time compared with representative diffusion-based baselines. **Conclusion** We presents SFSG-Diff，a novel framework for one-step S2O translation，which addresses key challenges in cross-modal image generation by leveraging spatial-frequency joint feature modeling and lightweight Strong-Sparse conditional guidance. The MDE effectively suppresses speckle noise and enhances structural feature representation，while the SIF pattern improves cross-modal alignment and reduces computational complexity. Extensive experiments on multiple benchmark datasets demonstrate that the proposed method significantly outperforms existing approaches in both quantitative metrics and visual quality. Moreover，the framework achieves substantial improvements in inference efficiency，making it more suitable for practical remote sensing applications with limited computational resources. Overall，the proposed SFSG-Diff framework provides an effective solution for robust and efficient SAR-to-optical image translation and offers new insights into the integration of spatial-frequency modeling and sparse conditional guidance within diffusion-based generative models.

Key words: SAR-to-optical image translation; diffusion model; feature fusion; remote sensing imagery; multi-scale encoder; strongly sparse guidance

论文引用格式:[DOI:10. 11834/jig. 260188]

0 引　言

光学遥感影像凭借丰富的纹理信息与直观的视觉表达,在地物分类、变化检测等遥感解译任务中应用广泛。然而,光学传感器的成像过程高度依赖光照与大气条件,难以实现全天候连续观测(Zhao等,2025a)。相比之下,合成孔径雷达(Synthetic Aperture Radar,SAR)作为一种主动微波成像传感器,具备全天时、全天候成像能力,能够穿透云雾并在弱光条件下稳定获取地表信息(Huang等,2021)。因此,将SAR图像转换为具有直观视觉表达的光学图像(SAR-to-Optical,S2O)对融合多源遥感优势、构建全时空对地观测体系具有重要价值。

然而,现有S2O方法在实际应用中仍面临两大技术瓶颈:其一,SAR固有的乘性相干斑噪声严重制约了跨模态特征的对齐与融合。斑点噪声与地物信号在空间域高度耦合且非线性混合,直接导致网络特征提取的准确性下降。受此影响,现有方法在生成光学图像时,极易出现结构扭曲与纹理混淆等问题(Schmitt 等,2018)。其二,主流生成模型难以满足遥感平台的实时处理需求。为了弥补上述特征表达的缺陷,现有的改进方案往往依赖复杂的网络模块。加之扩散模型等主流架构本身受限于多步迭代采样的机制,导致模型整体计算代价高昂,难以适配计算资源受限的遥感平台。

早期研究主要聚焦于采用生成对抗网络架构(generative adversarial network,GAN)。例如,Pix2Pix(Isola等,2017)和CycleGAN(Zhu等,2017)通过构建图像映射关系,实现了S2O翻译的初步探索。后续研究主要从网络结构和损失函数两方面进行改进:引入多尺度残差连接抑制深层特征流失(Fu等,2021),使用扩张卷积扩大神经元感受野(Naderi Darbaghshahi等,2022),或在生成器中融入边缘保持卷积,强化轮廓细节特征(Guo等,2021;Yang等,2022b)。此外,引入细粒度非平衡生成器与结合感知、风格等多重监督损失也进一步优化了生成效果(Yang等,2022a)。然而,受限于卷积神经网络(convolutional neural networks,CNN)的局部感受野,这类方法在复杂场景中难以充分建模长距离依赖关系,缺乏对SAR图像中全局纹理一致性的充分感知能力,生成结果常出现纹理模糊或结构失真等问题(李美玲 等,2026)。此外,GAN架构存在训练稳定性差、模式崩溃等固有缺陷,限制了其在高精度S2O翻译中的应用(Bai等,2023)。

近年来,扩散模型(diffusion model,DM)凭借其稳健的渐进式加噪与去噪过程,在图像生成领域表现出比GAN更强大的拟合能力(Li等,2023;Su等,2024;黄颖 等,2025;王义杰 等,2025),并在S2O翻译任务中取得阶段性突破。其中,IDDPM(Nichol和Dhariwal,2021)通过优化反向过程的变分下界构建稳定的条件分布;Bai与Xu(2024)引入颜色监督损失纠正生成图像的色彩偏差;cBBDM(Kim和Chung,2025)利用布朗桥机制直接连接源域与目标域以强化重建细节;DGD-S2O(Du等,2025)通过噪声域特征对齐来强化边缘复原。尽管这些方法在一定程度提升了图像的生成质量,但它们多采用简单

特征拼接或浅层融合方式粗粒度地注入SAR条件信息，未能实现精细的跨模态语义对齐。同时，扩散模型多步迭代采样的特性导致推理效率较低，难以满足计算资源严格受限的遥感平台对实时图像处理的需求(Song等，2020)。

除空域优化外，引入频域信息为突破图像生成现存瓶颈提供新思路。频域分析擅长表征全局结构、区分噪声，在传统去噪与超分任务中已有验证：频域卷积可有效加速特征提取(Dong等，2016)，频率感知注意力机制则擅长复原细微纹理细节(Yu等，2018)。具体到S2O任务，HVT-cGAN(Zhao等，2025a)通过并行CNN与Transformer分支隐式捕获空频联合特征，SFDiff(Qin等，2024a)则通过在扩散模型去噪网络中增设空频互补学习模块以完善空域特征，相关研究验证了空频融合方案的可行性。但现有方法多依托纯数据驱动的隐式特征对齐，未有效利用SAR图像的频谱物理先验，即噪声主要分布在高频，地物结构信息则集中于中低频区域(Goodman，1976)，这也导致模型难以兼顾去噪效果与结构保留能力。

针对上述问题，本文提出一种空频强稀疏引导的扩散模型(Spatial-Frequency Strongly-Sparse Guided Diffusion Model，SFSG-Diff)。根据SAR图像的频谱分布特性，本文设计了多尺度空频去噪编码(Multi-scale Spatial-Frequency Denoising Encoder，MDE)，通过并行分支协同提取空域与频域特征，在最大限度保留图像空间结构完整性的基础上，对频域噪声进行显式判别与抑制，从根源上实现信号与噪声的高效解耦，为扩散生成过程提供更鲁棒的条件先验。为进一步提升计算效率、规避密集条件注入带来的噪声与冗余干扰，本文构建轻量级强稀疏语义融合模块(Strong-Sparse Semantic Fusion，SIF)。该模块突破传统全特征注入的固有局限，自适应将有效条件信息融入去噪网络的部分特征流，并结合双重注意力机制动态调控融合权重，在降低模型计算开销的同时，显著提升跨模态语义对齐精度。此外，针对扩散模型的计算瓶颈，设计两阶段训练策略，使得模型推理阶段仅需单步采样即可生成高质量图像，将推理速度提升至GAN架构水平，可适配算力受限的遥感平台。

本文主要贡献如下：1)提出空频强稀疏引导的扩散模型(SFSG-Diff)，利用扩散模型的稳定生成能力与空频协同部分引导的优势来实现稳健高效的S2O翻译；2)设计多尺度空频去噪编码(MDE)，通过多尺度空频并行特征提取与融合，协同建模SAR图像全局结构与局部细节，有效提升模型对相干斑噪声的鲁棒性；3)构建轻量级的强稀疏语义融合(SIF)，依托多尺度条件特征实现对部分特征流的精准引导，兼顾跨模态语义对齐精度与模型计算效率；4)在SEN1-2、QXS-SAROPT、WHU-OPT-SAR三个基准数据集上的实验表明该方法在视觉保真度与定量指标上均优于现有先进方法，且推理效率接近GAN模型，为实时SAR图像处理提供可行方案。

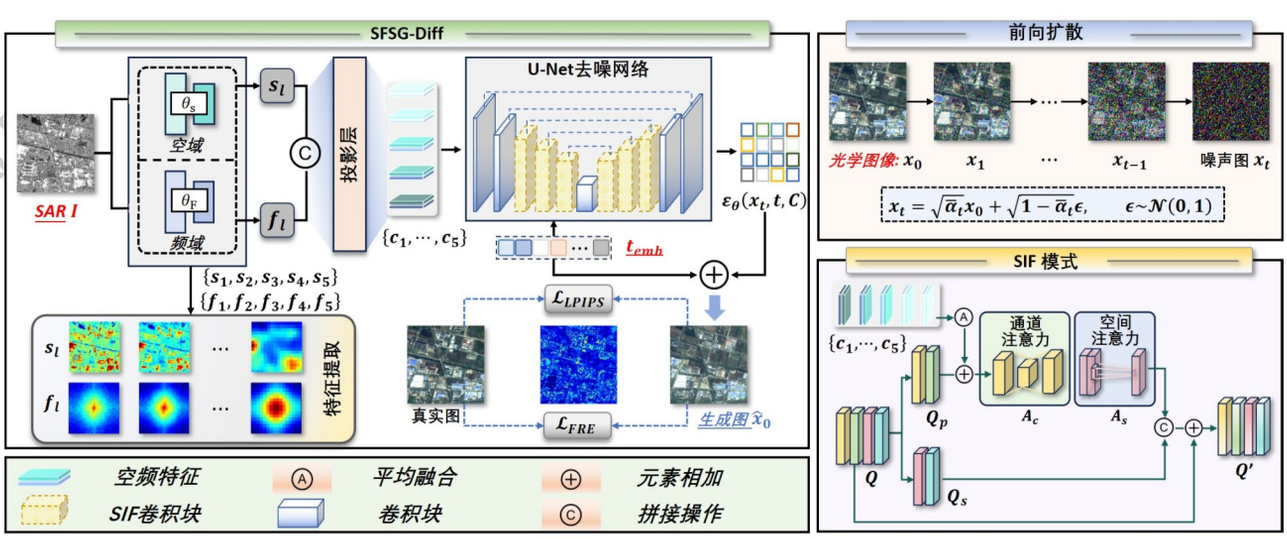


图1 SFSG-Diff模型整体框架图

Fig. 1 Overall framework diagram of the SFSG-Diff model

1 本文方法

1.1 整体框架

为应对现有方法在噪声鲁棒性与计算效率方面的局限，本文提出空频强稀疏引导扩散模型SFSG-Diff。SFSG-Diff的整体架构如图1所示，该框架通过前向扩散与反向去噪两阶段协同工作，并引入多尺度空频特征作为精细化的条件引导信号，从而在复杂场景下实现高质量、高效率的S2O转换。在前向扩散过程中，目标光学图像$\boldsymbol{x}_0 \in \mathbb{R}^{H\times W\times 3}$经过$T$步，通过遵循预定义好的噪声调度表$\beta_t \in (0,1)$，逐步注入高斯噪声$\boldsymbol{\epsilon}$，生成一系列噪声逐渐增强的中间图像$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T$。任意时刻$t$处的噪声图像$\boldsymbol{x}_t$可通过以下闭式解直接采样得到：

$$\boldsymbol{x}_t = \sqrt{\bar{a}_t}\,\boldsymbol{x}_0 + \sqrt{1-\bar{a}_t}\,\boldsymbol{\epsilon},\quad \boldsymbol{\epsilon}\sim\mathcal{N}(0,\boldsymbol{I}) \tag{1}$$

$$\bar{a}_t = \prod_{i=1}^{t}(1-\beta_i),\quad t\in\{1,2,\cdots,T\} \tag{2}$$

式中$\bar{a}_t$表示时刻t处的累计噪声方差，用于控制加噪进程。

反向去噪过程是模型实现图像重建的关键，由参数化的U-Net（U-shaped Convolutional Network）去噪网络$\boldsymbol{\theta}$以噪声图像$\boldsymbol{x}_t$、扩散时间步t以及由从SAR图像中提取的多尺度先验特征$\boldsymbol{C}$为输入，预测应被移除的噪声$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, \boldsymbol{C})$。

时间步调制机制被深度集成在整个网络中，以自适应地引导不同噪声水平的去噪过程。具体而言，模型首先将标量时间步t通过正弦位置编码被映射为高维向量$\boldsymbol{t}_{emb}$，并进一步生成与网络深度相适应的自适应缩放与偏置参数$\boldsymbol{\gamma}$和$\boldsymbol{\beta}$，以实现对不同噪声水平的自适应调制，增强生成过程的灵活性与可控性。这些调制参数与来自多尺度空频去噪编码的条件特征$\boldsymbol{C}$被协同注入到U-Net的各个层级。U-Net每一层的输入特征$\boldsymbol{h}$首先经历时间步调制，再与条件特征一同送入强稀疏语义融合进行深度融合：

$$\boldsymbol{\gamma},\boldsymbol{\beta} = \boldsymbol{F}_m(\boldsymbol{t}_{emb}) \tag{3}$$

$$\boldsymbol{h}_{\text{time}} = (1+\boldsymbol{\gamma})\cdot\boldsymbol{F}_c(\boldsymbol{h}) + \boldsymbol{\beta} \tag{4}$$

$$\boldsymbol{h}_{\text{out}} = \boldsymbol{F}_c(\boldsymbol{F}_{\text{SIF}}(\boldsymbol{h}_{\text{time}}, \boldsymbol{C})) + \text{Conv}_{1\times 1}(\boldsymbol{h}) \tag{5}$$

式中$\boldsymbol{h}_{\text{time}}$为时间步调制后的特征；$\boldsymbol{h}_{\text{out}}$为融合后的输出特征；$\boldsymbol{F}_c$代表标准卷积块（包含组归一化、Swish激活函数和3×3卷积）；$\boldsymbol{F}_m$是一个轻量的多层感知机；$\boldsymbol{F}_{\text{SIF}}$表示本文提出的轻量级SIF方法。公式中的残差连接确保了主要信息流的稳定性。通过这种将全局时间步调制与局部空频条件引导深度集成的设计，模型能够根据输入SAR图像的独特结构和内容，动态且精细地指导从纯噪声到逼真光学图像的重建全过程。

1.2 多尺度空频去噪编码MDE

SAR图像中的相干斑噪声本质上是由地形散射体回波的干涉引起的，属于地物信号严重耦合的乘性噪声，且在频谱上表现为弥散且广泛分布的宽带白噪声（主要集中于高频段）（Goodman，1976）。为减轻其干扰，MDE利用噪声和信号在频域可分离的物理特性，采用空间与频率双分支并行架构，旨在从不同域中提取互补的特征表示，通过显式的空频协同建模，为后续的扩散生成过程提供鲁棒且判别性强的条件引导。

由于相位信息对位置极其敏感且容易被高频噪声严重破坏，直接在复数域进行特征提取往往会因为相位误差而引发伪影。因此，MDE的频率分支舍弃相位，仅用对数幅度谱进行建模，使特征提取免受空间微小偏移与局部强噪声的干扰，从而在频域实现信号与斑点噪声的有效解耦。

MDE的输入为经过通道简单复制输入的SAR图像$\boldsymbol{I}\in\mathbb{R}^{H\times W\times 3}$。两个并行分支空间编码器$\boldsymbol{\theta}_{\text{S}}$与频率编码器$\boldsymbol{\theta}_{\text{F}}$均由卷积与堆叠残差块构成，以构建深层的多尺度特征提取能力。空间分支$\boldsymbol{\theta}_{\text{S}}$直接处理输入图像$\boldsymbol{I}$，通过连续下采样捕获对细节恢复至关重要的局部纹理与上下文信息。与此同时，频率分支$\boldsymbol{\theta}_{\text{F}}$则专注于提取对噪声不敏感的全局轮廓与结构，首先通过将SAR图像$\boldsymbol{I}$通过Gray(·)转换为灰度图像，再进行二维的傅里叶变换$\mathcal{F}$，并计算对数幅度谱$\boldsymbol{M}\in\mathbb{R}^{H\times W\times 1}$以增强数值稳定性与结构可见性：

$$\boldsymbol{M} = \log(1+|\mathcal{F}(\text{Gray}(\boldsymbol{I}))|) \tag{6}$$

通过复制得到三通道频率表示$\boldsymbol{M}'\in\mathbb{R}^{H\times W\times 3}$以匹配SAR图像$\boldsymbol{I}$的三通道，并送入频率编码器$\boldsymbol{\theta}_{\text{F}}$进行特征学习。

两个分支均输出五个尺度的特征，形成层次化的特征金字塔，其特征提取过程可形式化表示为：

$$\{\boldsymbol{s}_1,\boldsymbol{s}_2,\boldsymbol{s}_3,\boldsymbol{s}_4,\boldsymbol{s}_5\} = \boldsymbol{\theta}_{\text{S}}(\boldsymbol{I}) \tag{7}$$

$$\{\boldsymbol{f}_1,\boldsymbol{f}_2,\boldsymbol{f}_3,\boldsymbol{f}_4,\boldsymbol{f}_5\} = \boldsymbol{\theta}_{\text{F}}(\boldsymbol{M}') \tag{8}$$

式中$\boldsymbol{s}_l,\boldsymbol{f}_l\in\mathbb{R}^{\frac{H}{2^{l-1}}\times\frac{W}{2^{l-1}}\times C_l}$分别承载了U-Net第$l$层的空

间和频率特征，C_l表示通道维度。

为实现空频信息的深度融合与互补增强，MDE在每个尺度上引入了轻量级的特征融合模块。首先对同尺度的空间特征与频率特征进行通道拼接，随后通过一个1 × 1卷积投影层$\mathcal{P}_l$进行融合与降维：

$$\boldsymbol{c}_l = \mathrm{P}_l([\boldsymbol{s}_l, \boldsymbol{f}_l]),\ \ l = 1,2,\cdots,5 \tag{9}$$

式中[·, ·]表示通道维度拼接，$\boldsymbol{c}_l \in \mathbb{R}^{\frac{H}{2^{l-1}} \times \frac{W}{2^{l-1}} \times C_l}$是第$l$尺度的融合特征。最终输出是双域融合的多尺度条件特征集合$\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, \boldsymbol{c}_4, \boldsymbol{c}_5\}$，为后续去噪网络提供局部精准、全局一致的引导信息，从根源上提升强噪声场景特征提取与跨模态对齐的可靠性。

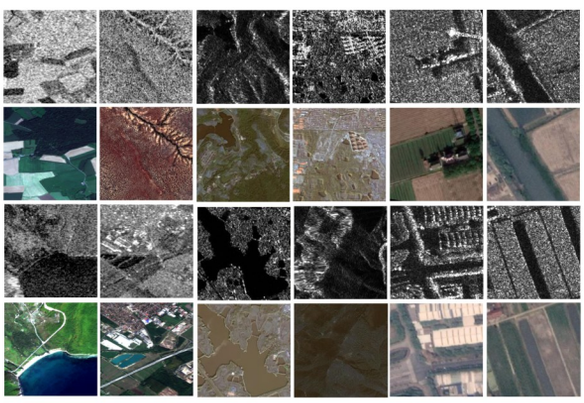


图2　各数据集示例图像

Fig. 2　Sample images from the datasets

1.3　强稀疏语义融合SIF

为解决传统条件注入中计算冗余和语义对齐不足问题，本文提出了轻量级的SIF模式。传统方法在进行S2O转换时，往往依赖密集特征注入，这容易引入SAR图像固有的斑点噪声和更多的无关信息，损害生成质量的同时带来较高的计算负担。实际上，S2O转换更依赖于稀疏而精准的跨模态语义引导，而非简单的特征堆叠。受跨阶段网络设计启发（Wang等，2020），SIF采用通道分割策略，仅用多尺度融合的条件信息引导部分特征流，并结合双重注意力机制完成精细化语义调制，实现精度与效率的协同优化，模块结构如图1所示。

首先，SIF采用通道分割策略将输入特征划分为处理路径与捷径路径，兼顾条件引导强度与模型训练稳定性。针对去噪网络U-Net$\boldsymbol{\theta}$主干特征$\boldsymbol{Q} \in \mathbb{R}^{B \times C_q \times H \times W}$，按分割比例$\rho$进行通道划分：

$$\boldsymbol{Q}_p, \boldsymbol{Q}_s = \mathrm{Split}(\boldsymbol{Q}, [\rho C_q, (1-\rho)C_q]) \tag{10}$$

式中$\boldsymbol{Q}_p \in \mathbb{R}^{B \times \rho C_q \times H \times W}$作为处理路径，接收MDE输出的多尺度条件特征进行调制，以增强特征多样性与语义判别力，而$\boldsymbol{Q}_s \in \mathbb{R}^{B \times (1-\rho) C_q \times H \times W}$作为捷径路径，保留原始输入信息以确保梯度传播的稳定性，避免因条件注入过强而导致的训练不稳定。实验默认分割比例$\rho = 0.5$，使模型能够以较低的计算代价引入条件信息，并保持主干信息流的通畅。

对于处理路径$\boldsymbol{Q}_p$，SIF对MDE输出的条件特征$\boldsymbol{c}_l \in \boldsymbol{C}$进行自适应融合。每个尺度的条件特征$\boldsymbol{c}_l$通过双线性插值统一空间尺寸得到$\tilde{\boldsymbol{c}}_l$，再经由轻量级线性层$\mathcal{W}_l$完成通道维度对齐，最终通过均值聚合得到多尺度融合特征$\boldsymbol{F} \in \mathbb{R}^{B \times \rho C_q \times H \times W}$：

$$\boldsymbol{F} = \frac{1}{\mathrm{N}} \sum_{l=1}^{\mathrm{N}} \mathcal{W}_l(\tilde{\boldsymbol{c}}_l),\ \ \mathrm{N} = 5 \tag{11}$$

式中N表示特征尺度的数量。多尺度融合机制可充分整合SAR图像局部细节与全局结构的跨模态先验信息，为后续注意力精细化调制提供丰富、分层的语义支撑。

接着，SIF通过双重注意力机制对融合后的特征进行精细化调制，以实现更精准的跨模态语义对齐。融合特征$\boldsymbol{F}$通过残差连接与处理路径特征$\boldsymbol{Q}_p$结合，并依次通过通道注意力$\mathcal{A}_c$与空间注意力$\mathcal{A}_s$：

$$\boldsymbol{Q}'_p = \mathcal{A}_s(\mathcal{A}_c(\boldsymbol{Q}_p + \boldsymbol{F})) \tag{12}$$

式中$\mathcal{A}_c$采用全局平均池化与全连接层生成通道维权重，强调与当前生成内容语义最相关的特征通道。$\mathcal{A}_s$则利用卷积操作生成空间权重图，使模型能够聚焦于由SAR条件特征所引导的关键图像区域。最后，调制后的处理路径特征$\boldsymbol{Q}'_p \in \mathbb{R}^{B \times \rho C_q \times H \times W}$与捷径路径特征$\boldsymbol{Q}_s$拼接，并通过一个融合卷积层$\mathcal{H}$结合残差连接，生成SIF的最终输出$\boldsymbol{Q}' \in \mathbb{R}^{B \times C_q \times H \times W}$：

$$\boldsymbol{Q}' = \mathcal{H}([\boldsymbol{Q}'_p, \boldsymbol{Q}_s]) + \boldsymbol{Q} \tag{13}$$

该设计实现了多尺度条件特征的动态、自适应集成，而不干扰主要特征流，以最小的计算成本实现了精确的跨模态对齐。具体步骤可见算法1。

1.4　两阶段训练与损失函数

为实现高质量与高效率的SAR到光学图像翻译，SFSG-Diff采用两阶段的训练策略，使用不同损失函数进行联合优化。

1.4.1　第一阶段

第一阶段侧重于让网络学习从噪声输入$\boldsymbol{x}_t$重建光学图像$\boldsymbol{x}_0$的条件分布。采用标准扩散模型训练范

算法 1：强稀疏语义融合算法

输入：查询特征 $\boldsymbol{Q}$，多尺度条件特征集合 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3,\boldsymbol{c}_4,\boldsymbol{c}_5\}$
输出：调制特征 $\boldsymbol{Q}'$

特征分割 $\boldsymbol{Q}_p,\boldsymbol{Q}_s=\mathrm{Split}(\boldsymbol{Q},[\rho C_q,(1-\rho)C_q])$
for $\boldsymbol{c}_l$ **in** $\{\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3,\boldsymbol{c}_4,\boldsymbol{c}_5\}$ **do**
    $\tilde{\boldsymbol{c}}_l\leftarrow$双线性插值调制 $\boldsymbol{c}_l$ 尺寸与 $\boldsymbol{Q}_p$ 相同
    $\boldsymbol{F}\leftarrow\boldsymbol{F}+\mathcal{W}_l(\tilde{\boldsymbol{c}}_l)$ #1×1投影
end for
$\boldsymbol{F}\leftarrow\boldsymbol{F}/5$ #平均融合
$\boldsymbol{Q}'_p\leftarrow\mathcal{A}_s(\mathcal{A}_c(\boldsymbol{Q}_p+\boldsymbol{F}))$ #双重注意力
特征重组 $\boldsymbol{Q}'=\mathcal{H}([\boldsymbol{Q}'_p,\boldsymbol{Q}_s])+\boldsymbol{Q}$
return $\boldsymbol{Q}'$

式，目标函数为去噪均方误差损失：

$$\mathcal{L}_{\mathrm{I}}=\mathbb{E}_{\boldsymbol{x}_0,t,}\ [\|\quad_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t,\boldsymbol{C})-\quad\|^2] \tag{14}$$

此阶段训练步数为1000步，使用50步DDIM（Denoising Diffusion Implicit Model）采样器（Song等，2020）进行验证。经过该阶段，去噪网络 $\boldsymbol{\theta}$ 已具备稳定的逐步去噪能力，且相邻时间步的去噪预测高度连续，为后续压缩采样步数提供了可靠的先验分布基础。

表1　实验参数配置

Table 1　Experimental parameters configuration

配置项	第一阶段	第二阶段
训练/推理步数	1000/50（DDIM）	1000/1（DDIM）
批量大小	11	10
迭代次数	140000	160000
损失权重	$\mathcal{L}_{\mathrm{I}}$	$\boldsymbol{\lambda}_1\mathcal{L}_{\mathrm{LPIPS}}+\boldsymbol{\lambda}_2\mathcal{L}_{\mathrm{FRE}}+\boldsymbol{\lambda}_3\mathcal{L}_{\mathrm{GAN}}$
学习率	5×10^{-5}	3×10^{-5}

注：DDIM步数指使用DDIM采样器的步数。

1.4.2　第二阶段

第二阶段旨在将第一阶段学到的高质量多步去噪轨迹，压缩为单步直接映射，从而实现高效推理。通过固定噪声端点 T，以单步DDIM采样公式直接计算最终生成结果，并与真实图像计算损失，使网络将完整的反向扩散链精简为一次前向传播。

具体的，给定纯噪声 $\boldsymbol{x}_T\sim\mathcal{N}(0,\boldsymbol{I})$ 和条件特征 $\boldsymbol{C}$，单步生成结果由DDIM一步更新公式得到：

$$\hat{\boldsymbol{x}}_0=\frac{\boldsymbol{x}_T-\sqrt{1-\overline{\alpha}_T}\,\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_T,\boldsymbol{C},T)}{\sqrt{\overline{\alpha}_T}} \tag{15}$$

式中 $\overline{\alpha}_T$ 表示时刻 T 处的累计噪声方差。该公式不含中间时间步迭代，模型仅需一次前向即可输出重建图像 $\hat{\boldsymbol{x}}_0$。由于第一阶段已确保去噪函数在相邻时间步间高度一致，仅在端点 T 引入施加多损失联合优化，即可利用轨迹连续性将完整去噪链隐式缩减为单步函数。这与一致性模型（Song等，2023）需在整个轨迹显示约束相邻输出相等的做法不同，无需专门设计离散策略或额外训练约束。

为在实现单步推理的同时提升视觉真实感、结构保真度与频谱一致性。模型引入的总体损失为：

$$\mathcal{L}_{\mathrm{II}}=\mathcal{L}_{\mathrm{I}}+\boldsymbol{\lambda}_1\mathcal{L}_{\mathrm{LPIPS}}+\boldsymbol{\lambda}_2\mathcal{L}_{\mathrm{FRE}}+\boldsymbol{\lambda}_3\mathcal{L}_{\mathrm{GAN}} \tag{16}$$

式中加权系数根据经验设置为 $\boldsymbol{\lambda}_1=10$，$\boldsymbol{\lambda}_2=5$ 和 $\boldsymbol{\lambda}_3=0.3$，以平衡各项损失的贡献。

感知损失 $\mathcal{L}_{\mathrm{LPIPS}}$ 使用预训练的VGG-16（Visual Geometry Group 16-layer network）网络 ϕ 提取多层特征图，计算逐像素加权L2距离，以增强生成图像的局部纹理一致性和视觉自然度：

$$\mathcal{L}_{\mathrm{LPIPS}}=\mathbb{E}[\|\phi(\hat{\boldsymbol{x}}_0)-\phi(\boldsymbol{x}_0)\|_2^2] \tag{17}$$

式中 $\boldsymbol{x}_0$ 是真实光学图像，$\hat{\boldsymbol{x}}_0$ 是合成输出。

聚焦频率损失 $\mathcal{L}_{\mathrm{FRE}}$ 则针对SAR图像频域特性设计，通过动态构建频谱权重矩阵（采用对数调整与逐样本最大值归一化）对重建困难的中高频分量施加更高权重，显式约束生成图像与真实图像在频谱空间的一致性，从而有效抑制SAR图像乘性噪声在频域残留的伪影：

$$\mathcal{L}_{\mathrm{FRE}}=\frac{1}{HW}\sum_{\substack{u\in H\\v\in W}}\tilde{w}(u,v)\left\|\mathcal{F}(u,v)-\hat{\mathcal{F}}(u,v)\right\|_2^2 \tag{18}$$

式中 H 和 W 分别表示图像高度和宽度；$\mathcal{F}(u,v)$ 表示在频率坐标 (u,v) 处计算的二维傅里叶变换；$\tilde{w}(u,v)$ 表示频谱坐标 (u,v) 处的重新加权系数，用于强调重建困难的频率分量。

对抗损失 $\mathcal{L}_{\mathrm{GAN}}$ 采用多级sigmoid损失，判别器以CLIP（Contrastive Language-Image Pre-training）为视觉主干，输出多个尺度的真实性概率图，通过生成器 D 与判别器的对抗训练进一步优化生成质量，促进单步推理的收敛：

$$\mathcal{L}_{\mathrm{GAN}}=\mathbb{E}_{\boldsymbol{x}_0}[\log D(\boldsymbol{x}_0)]+\mathbb{E}_{\hat{\boldsymbol{x}}_0}[\log(1-D(\hat{\boldsymbol{x}}_0))] \tag{19}$$

上述感知级损失均不关心中间去噪过程，只对最终输出施加压力，驱使网络将所有去噪能力压缩至一次前向。与需要数轮训练、保存教师模型的渐进蒸馏（Salimans 和 Ho，2022）相比，SFSG-Diff的两阶段训练无需额外引入学生网络或修改网络结构，

仅通过切换训练目标即可实现单步高效推理，更为轻量且天然适配强条件驱动的图像翻译任务。

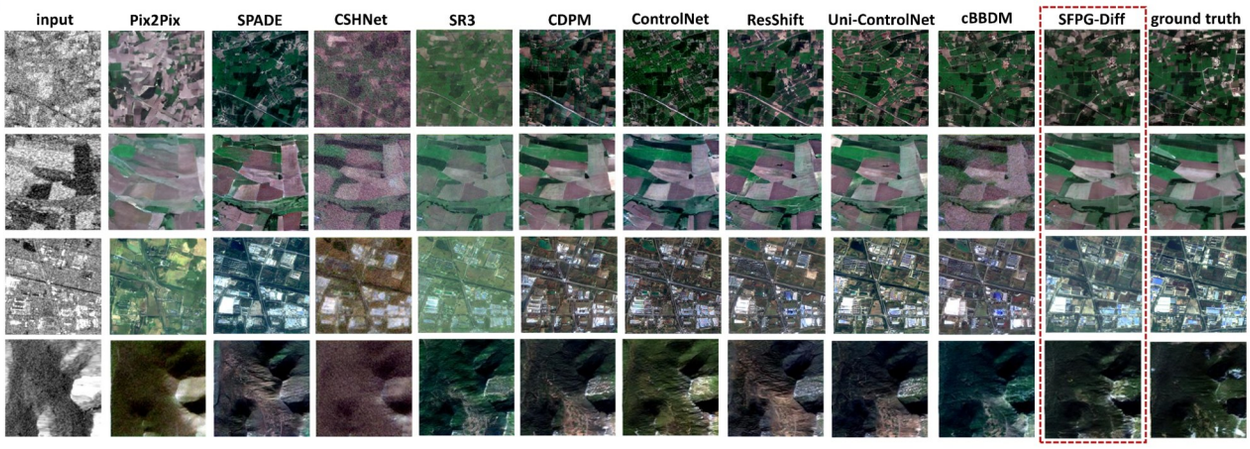


图3　在SEN1-2测试集上各方法的可视化对比结果

Fig. 3　Visual comparison results of different methods on the SEN1-2 test set

2　实　验

2.1　实验设置

2.1.1　实施细节

实验设置总扩散步数 $T = 1000$，采用线性噪声调度，其中 β_t 从 10^{-6} 增加到 10^{-2}。所有实验均使用单张NVIDIA V100 GPU(32GB)，并采用Adam优化器。训练采用两阶段策略：第一阶段使用均方误差损失(式(14))训练140K步，批大小为11，学习率按余弦衰减从 5×10^{-5} 降至 1×10^{-6}；第二阶段使用联合损失(式(15))进一步优化160K步，批大小调整为10，学习率从 3×10^{-5} 余弦衰减至 1×10^{-6}。详细的参数配置如表1所示。

2.1.2　数据集

为全面评估模型性能与泛化能力，本文在三类特性各异的公开遥感数据集上开展对比实验，数据样本示例如图2所示。QXS-SAROPT(Huang等，2021)包含20,000对配对的SAR-光学图像，以复杂港口与海岸线场景为主，存在显著的乘性相干斑噪声与海洋杂波干扰。本文从中选取4900对作为训练集、400对作为测试集。SEN1-2(Schmitt等，2018)为全球多时相遥感数据集，涵盖多类季节变化、地物类型与光照条件，对模型的泛化能力和噪声鲁棒性具有较高考验难度，本文选取5200对图像用于训练、400对用于测试。WHU-OPT-SAR(Li等，2022)包含100组大幅面高分辨率图像，覆盖城乡各类大尺度地形与复杂人工结构，适用于评估模型宏观几何结构与高频纹理细节的重建能力，本文筛选4900对用于训练、400对用于测试。实验图像均统一预处理为256×256分辨率并进行归一化处理，确保输入一致性。全部数据划分均采用随机采样，保证评估结果的客观性与公平性。

2.1.3　评估指标

实验采用四项通用的主流图像评价指标，多维度量化生成图像的综合质量：结构相似性指数(Structural Similarity Index，SSIM)衡量两幅图像在亮度、对比度和结构信息上的相似性，数值越接近1，代表图像结构保真效果越优；峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)量化生成图像与真实图像在像素强度上的误差，数值越高，像素级重建精度越高；学习感知图像块相似度(Learned Perceptual Image Patch Similarity，LPIPS)基于深度特征空间计算相似度，贴合人眼视觉评价机制，数值越低代表视觉感知效果越好；弗雷歇初始距离(Fréchet Inception Distance，FID)通过计算生成图像与真实图像的特征分布距离，反映生成样本的真实性与分布多样性，数值越低表明生成图像分布越贴合真实图像。

2.2　对比实验

为系统评估SFSG-Diff的综合性能，本文在上述三个数据集上与多种最具代表性的方法开展对比实验。对比方法包括GAN类方法：Pix2Pix(Isola等，

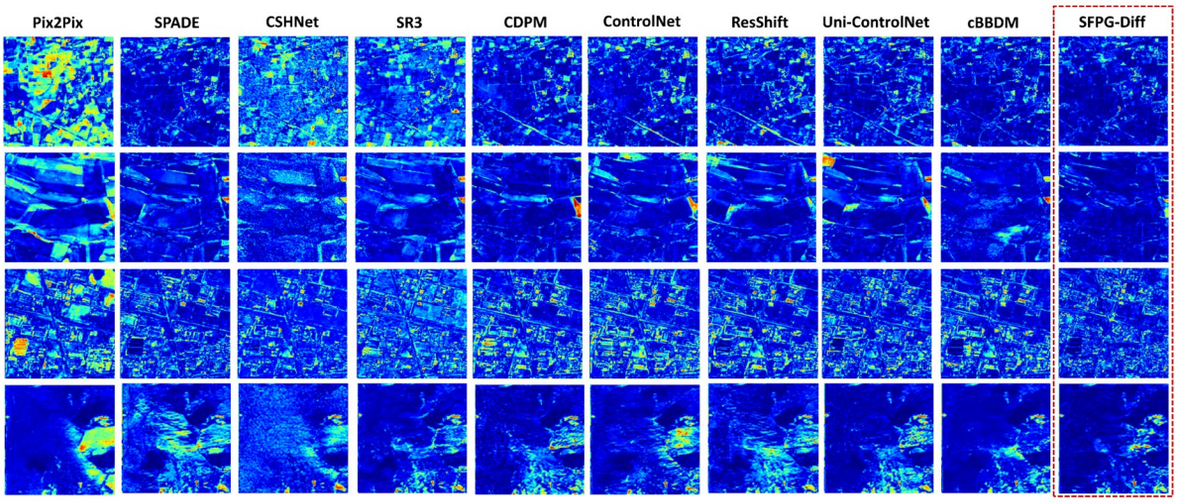


图4　在SEN1-2数据集上各方法的像素级误差热力图对比

Fig. 4　Comparison of pixel-level error heat maps for different methods on the SEN1-2 dataset

2017)、SPADE(Park 等，2019)和CSHNet(Yang 等，2026)；以及扩散模型类方法：SR3(Qin 等，2024b)、CDPM(Bai 等，2023)、ControlNet(Zhang 等，2023)、ResShift(Yue 等，2023)，Uni-ControlNet(Zhao 等，2023)和cBBDM(Kim和Chung，2025)。为确保对比的公平性，所有对比方法均根据官方开源代码在本文划分的统一数据集上重新训练测试。

表2　SEN1-2数据集上不同方法的定量比较

Table 2　Quantitative comparison of different methods on the SEN1-2 dataset

模型	PSNR↑	SSIM↑	LPIPS↓	FID↓
Pix2Pix	15.24	0.201	0.595	155.35
SPADE	16.35	0.272	0.498	119.44
CSHNet	16.37	0.289	0.635	90.51
SR3	15.85	0.365***	0.463	89.07
CDPM	16.65	0.352	0.455	82.76
ControlNet	16.15	0.285	0.485	105.41
ResShift	16.75**	0.345	0.467	97.85
Uni-ControlNet	16.05	0.278	0.493	103.73
cBBDM	16.66	0.351	0.424*	83.31
SFSG-Diff	**17.52**	**0.430**	**0.390**	**68.23**

注：***p<0.001，**p<0.01，*p<0.05(与次优方法进行逐样本的Wilcoxon符号秩检验)；加粗和下划线分别代表最优与次优结果，灰色背景突出本文模型结果，后续表格的标注含义与此相同。

2.2.1　在SEN1-2数据集上

SEN1-2数据场景丰富、地物多变，可有效检验模型噪声鲁棒性与跨模态对齐性能，可视化对比结果如图3所示。可以观察到，SFSG-Diff在相干斑抑制、地物轮廓保留与细节纹理复原上整体优于其他对比方法。例如，在植被和城市区域，本方法能更准确地重建光学影像的纹理特征，避免出现其他方法中常见的结构模糊或噪声残留现象。图4通过像素级误差热力图进一步量化重建偏差：其他方法误差在整图范围内弥散或集中于结构边缘，SFSG-Diff的误差热力图显示其误差更小且分布更为收敛，全局与局部重建结果均更贴近真值图像。

定量比较结果(表2)有力地支撑了上述视觉观察。SFSG-Diff在所有关键指标上均取得了最佳性能，PSNR提升约0.77dB，SSIM提升约17.8%，LPIPS与FID分别降低约8.0%和17.6%。为验证性能提升的统计可靠性，本文进一步对SFSG-Diff与次优方法在SEN1-2测试集上进行了Wilcoxon符号秩检验。结果表明，本文方法在PSNR(p=0.0012)、SSIM(p<0.001)、LPIPS(p<0.05)上均显著优于次优方法，证明提升并非由随机波动引起。

SR3、CDPM等扩散模型依托稳定的生成建模，FID指标整体优于GAN系列算法，规避了GAN普遍存在的模式崩溃、训练震荡缺陷，但在结构保真(SSIM)与人眼感知细节(LPIPS)上仍存在短板。SFSG-Diff则通过精细化条件引导策略补齐上述缺陷，实现全指标协同优化。具体而言，MDE模块通

表3　QXS-SAROPT与WHU-OPT-SAR数据集上不同方法的定量比较

Table 3　Quantitative comparison of different methods on the QXS-SAROPT and WHU-OPT-SAR datasets

模型	QXS-SAROPT				WHU-OPT-SAR			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
Pix2Pix	14.65	0.246	0.498	195.51	16.15	0.302	0.526	101.26
SPADE	15.23	0.336	0.551	186.77	17.26	0.356	0.510	98.15
CSHNet	15.35	0.354	0.580	120.32	17.76	0.365	0.531	106.42
SR3	15.92	0.358	0.463	103.95	19.85	0.368	0.489	85.49
CDPM	15.88	0.369	0.501	112.84	19.63	0.352	0.496	79.64
ControlNet	15.52	0.365	0.467	95.24	22.83	0.396	0.468	65.64
ResShift	15.66	0.355	0.450	108.42	22.95	0.401	0.492	64.29
Uni-ControlNet	15.45	0.332	0.483	101.55	21.16	0.372	0.484	55.12
cBBDM	16.15	0.305	0.461	79.47	19.98	0.415	0.472	68.22
SFSG-Diff	**17.78**	**0.404**	**0.438**	**62.30**	**25.44**	**0.436**	**0.444**	**36.21**

过融合空域局部细节与频域的全局抗噪结构，从源头抑制噪声干扰，提升PSNR与SSIM；与SR3中简单的特征拼接或ControlNet中的单尺度条件引导不同，SIF模块实现了精细化的多尺度条件引导，确保了跨模态语义的精准对齐，这直接反映在LPIPS指标的显著改善上；两阶段训练与联合损失进一步优化了生成图像的视觉自然度与频谱一致性，共同促成了FID分数的大幅领先。

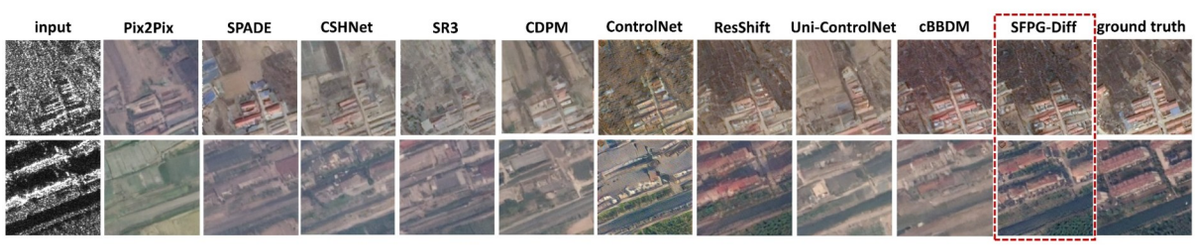


图5　在QXS-SAROPT数据集上各方法的可视化对比结果

Fig. 5　Visual comparison results of different methods on the QXS-SAROPT dataset

2.2.2　在QXS-SAROPT和WHU-OPT-SAR数据集上实验

QXS-SAROPT和WHU-OPT-SAR聚焦典型复杂场景，对模型的专项重建能力与场景适应性提出更高要求。

在QXS-SAROPT数据集上，SFSG-Diff取得了全面领先的定量结果（表3）：PSNR较次优方法cBBDM（16.15dB）提升达1.63dB；SSIM提升9.5%；LPIPS与FID分别降低2.7%和2.2%。这体现了MDE模块在处理复杂场景时的有效性。频率分支提供的全局结构先验有助于在强杂波背景下稳定重建大型结构的轮廓，而空间分支则能恢复码头、建筑物等目标的表面细节，两者的协同作用保障了生成图像在整体与局部的一致性。

在WHU-OPT-SAR数据集上，SFSG-Diff的性能优势最为显著（表3）：PSNR高达25.44dB，较次优方法ResShift（22.95dB）提升2.49dB；SSIM为0.436，提升5.1%；LPIPS降低5.1%；FID低至36.21，较次优方法Uni-ControlNet（55.12）降低约34.3%。这验证了SIF模块在大尺度城乡地形语义对齐方面的卓越能力。通过将多尺度条件特征动态、稀疏地注入到去噪网络，并结合注意力机制进行调制，模型能够

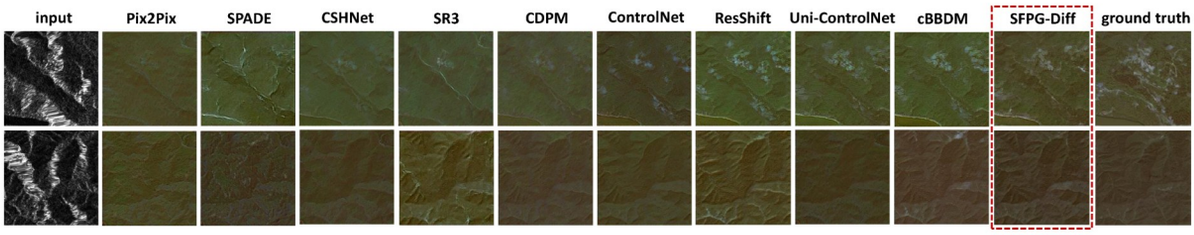


图6　在WHU-OPT-SAR数据集上各方法的可视化对比结果

Fig. 6　Visual comparison results of different methods on the WHU-OPT-SAR dataset

精准地依据SAR图像中的结构线索，在光学域中生成几何规整、丰富纹理的大尺度场景。

可视化结果直观地展示了上述优势。在QXS-SAROPT场景中（图5），本方法生成的整体轮廓锐利、结构清晰，且背景平滑，噪声抑制效果明显。在WHU-OPT-SAR场景中（图6），本方法重建的大尺度城乡区域排列整齐、纹理清晰，远优于其他方法可能出现的结构扭曲或纹理模糊现象。

表4　模型复杂度与推理效率分析

Table 4　Analysis of model complexity and inference efficiency

类型	模型	发表期刊	参数量/M	FLOPs/G	Memory/MB	Time/s
GANs	Pix2Pix	CVPR 2017	54.41	24.22	464.12	0.06
	SPADE	CVPR 2019	102.42	187.21	1024.12	0.09
	CSHNet	TCSVT 2026	30.60	55.56	1232.15	0.13
DMs	SR3	RADAR 2024	869.46	1964.86	4485.61	1.52
	CDPM	ICCV 2023	1312.72	2231.03	7567.83	4.50
	ControlNet	NeurIPS 2023	459.49	792.16	2963.64	0.68
	ResShift	NeurIPS 2023	1519.18	2295.20	8382.00	4.95
	Uni-ControlNet	NeurIPS 2023	756.23	1562.34	4215.28	1.81
	cBBDM	GRSL 2025	949.58	2122.49	6147.93	3.15
本文方法	SFSG-Diff	–	249.34	223.93	2211.84	0.21

2.3　效率分析

在实现卓越生成质量的同时，SFSG-Diff在计算效率方面也展现出显著优势。模型复杂度与推理效率的详细对比如表4所示，相较于现有先进扩散模型，SFSG-Diff的参数量（249.34M）和计算量（FLOPs 223.93G）均处于较低水平，显存占用（2211.84 MB）适中。更重要的是，得益于两阶段训练策略带来的单步采样推理特性，SFSG-Diff单图推理耗时仅0.21秒，可与传统GAN的方法相媲美，有效填补了扩散模型在图像翻译任务中固有的效率短板，具备在算力受限的遥感平台实时处理场景中部署的潜力。

2.4　消融实验

为深入探究SFSG-Diff中每个核心模块的设计合理性及其贡献，我们在SEN1-2数据集上进行了严谨的消融实验，以广泛使用的扩散模型SR3作为性能比较的基线。

2.4.1　多尺度空频去噪编码（MDE）效果分析

本文首先验证MDE中空频融合设计的必要性。在SR3基线基础上，模型分别引入仅含频率分支的MDE_f，仅含空间分支的MDE_s和完整的空频双分支MDE。消融实验结果如表5所示，与基线相比，单独引入频率分支MDE_f后，SSIM与FID分别改善约4.1%和5.9%，验证了频域全局结构信息对生成质量的积极作用，但其PSNR仅提升0.15dB，而LPIPS基本不变甚至有所回弹，表明仅靠幅度谱难以恢复精确的空间定位与局部细节。引入空间分支MDE_s

表5　基于SEN1-2数据集的MDE消融实验

Table 5　Ablation study of MDE on the SEN1-2 dataset

变量	PSNR↑	SSIM↑	LPIPS↓	FID↓
Baseline	15.85	0.365	0.463	89.07
+MDE$_s$	16.59	0.386	0.423	75.36
+MDE$_f$	16.01	0.380	0.465	83.78
+MDE	**17.20**	**0.412**	**0.413**	**70.44**

注：+MDE$_f$代表仅在基线上引入频率分支，+MDE$_s$代表仅在基线上引入空间分支，+MDE代表在基线上引入完整的空频双分支。

后，各项指标均有提升，证明了空间特征在像素级重建与细节恢复方面的有效性。当进一步引入完整的空频双分支MDE后，模型性能实现了全面且显著的飞跃：PSNR提升约0.61dB，SSIM提升约6.7%，LPIPS降低约2.4%，FID降低约6.5%。

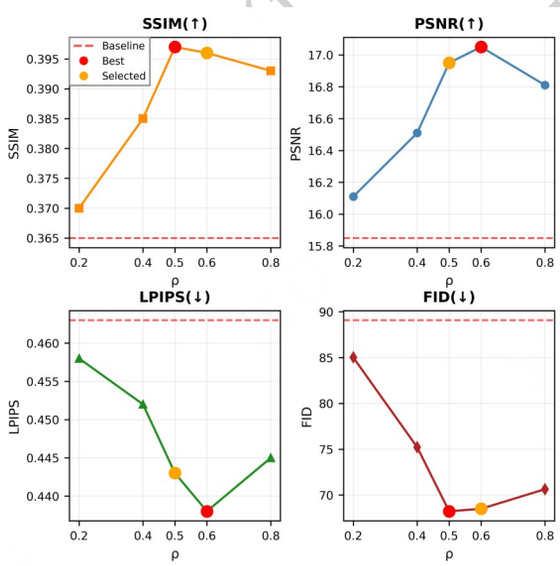


图7　分割比例ρ的影响

Fig. 7　Effect of Segmentation Ratio (ρ)

这一结果并非简单的特征叠加效应，而是源于空间域与频率域在信息表征上的本质互补性。空间卷积擅长捕捉局部纹理与上下文细节，但对全局结构建模能力有限且易受乘性噪声干扰；频率谱则能显式刻画图像的全局轮廓与结构分布，对高频噪声具有天然鲁棒性，但会损失精确的空间定位信息。单独使用任一分支均存在明显短板，双分支融合后更能发挥二者各自的优势。MDE通过并行提取并深度融合双域特征，实现了噪声与信号在表征层面的显式解耦与互补增强，从而为后续去噪过程提供了兼具局部精度与全局一致性的强判别性先验，这是模型性能全面提升的根本原因。

2.4.2　强稀疏语义融合(SIF)注意力机制分析

为探究SIF中双重注意力机制的具体作用，我们对比了四种不同的注意力配置变体：无注意力(SIF$_o$)、仅空间注意力(SIF$_s$)、仅通道注意力(SIF$_c$)以及完整双重注意力(SIF)。消融实验结果如表6所示。

表6　基于SEN1-2数据集的SIF注意力消融实验

Table 6　Ablation study of attention mechanisms in SIF on the SEN1-2 dataset

变量	PSNR↑	SSIM↑	LPIPS↓	FID↓
Baseline	15.85	0.365	0.463	89.07
+SIF$_o$	16.22	0.364	0.458	80.23
+SIF$_s$	**16.95**	0.397	0.443	77.81
+SIF$_c$	16.50	0.369	0.452	79.12
+SIF	16.85	**0.398**	**0.437**	**72.35**

注：+SIF$_o$代表不含任何注意力机制；+SIF$_s$代表只有空间注意力机制；+SIF$_c$代表只有通道注意力机制；+SIF代表完整的注意力机制

由实验结果可见：无注意力版本SIF$_o$性能虽优于基线，但显著落后于有注意力的变体，说明简单的特征融合无法实现有效的跨模态语义调制；仅使用空间注意力SIF$_s$在PSNR(16.95dB)上取得最高值，表明其通过聚焦关键空间区域，有利于像素级的精确重建；仅使用通道注意力SIF$_c$在各项指标上表现中等；完整的双重注意力机制SIF在SSIM、LPIPS和FID这三个关键指标上均取得了最佳或次优结果，展现出均衡优越的综合性能。

尽管空间注意力SIF$_s$在PSNR上略有优势，但其无法建模特征通道间的语义关联，易造成有效通道特征丢失。而SIF通过串联空间与通道注意力，既能保留对关键区域的精准空间定位，又能自适应地筛选与当前生成内容语义最相关的特征通道，实现更深层次的跨模态语义对齐，因此在SSIM、LPIPS和FID上表现更佳，这进一步验证了双重注意力机制对提升跨模态图像翻译性能的必要性。

2.4.3　强稀疏语义融合分割比例分析

SIF中的分割比例ρ控制着条件信息注入的强

度与范围，是平衡引导效果与计算效率的关键超参数。本文在0.2~0.8区间选取五组数值开展消融实验，通过图7可视化不同ρ值下的视觉对比结果，表7则展示了定量的对比结果。

实验表明，ρ过低（如0.2）时，处理路径通道数过少，条件信息注入不足，模型难以充分学习跨模态语义映射，导致生成图像结构模糊、细节缺失，各项指标均不理想；随着ρ增加，性能逐步提升。当ρ=0.5时，模型在SSIM与FID指标上达到最优；当ρ=0.6时，在PSNR（17.05dB）与LPIPS（0.438）上略优；然而，当ρ过高（如0.8）时，处理路径通道数过多，冗余通道会提升计算成本，也易引入噪声扰动主干特征传播，引发训练震荡，使模型性能回落（PSNR降至16.81dB）。虽然ρ=0.6在PSNR（17.05 dB）和LPIPS（0.438）上略有优势，但ρ=0.5在SSIM和FID上表现更佳，且ρ=0.5相比于ρ=0.6进一步降低了计算的复杂度。综合考虑定量精度与计算效率，本文将分割比例ρ默认设置为0.5。

表7　分割比例ρ对模型性能影响的消融实验

Table 7　Ablation study on the effect of segmentation ratio ρ on model performance

变量	PSNR↑	SSIM↑	LPIPS↓	FID↓
Baseline	15.85	0.365	0.463	89.07
ρ = 0.2	16.11	0.370	0.458	85.01
ρ = 0.4	16.51	0.385	0.452	75.22
ρ = 0.5	<u>16.95</u>	**0.397**	<u>0.443</u>	**68.23**
ρ = 0.6	**17.05**	<u>0.396</u>	**0.438**	<u>68.50</u>
ρ = 0.8	16.81	0.393	0.445	70.64

注：ρ代表分割比例，实验在SEN1-2数据集上进行

2.5　损失函数与推理步数消融分析

为验证单步推理的性能代价和两阶段训练策略中各损失函数的必要性与协同作用，本文在WHU-OPT-SAR数据集上以第一阶段预训练结果为基线（表8第一行实验结果），对不同损失组合与推理步数进行了联合消融实验。所有实验均以第一阶段训练完成的统一模型为起点，按表8所列设置分别进行第二阶段训练微调，实验结果如表8所示。

从第一阶段预训练结果可以看出，仅使用均方误差损失$\mathcal{L}_1$时，模型只能实现基础的图像生成，缺乏感知与频谱约束，PSNR仅为14.63dB，FID高达168.51，生成图像的视觉真实感与分布一致性表现较差。在预训练基础上引入联合损失且保持多步推理设置时，所有定量指标均显著改善，PSNR提升至25.82dB，FID降至35.67，该结果验证了预训练的必要性和多项损失联合优化的有效性。

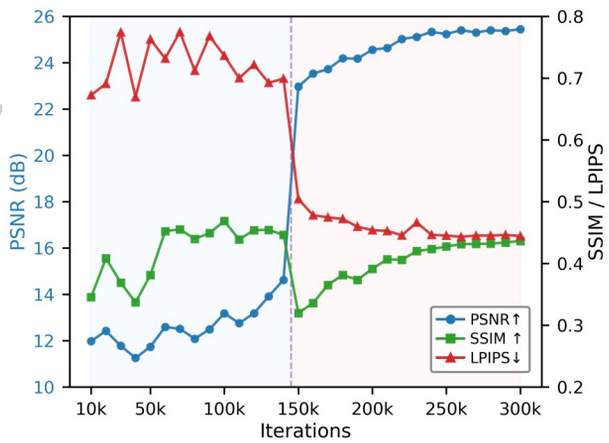


图8　两阶段训练指标演化曲线

Fig. 8　Two-Stage Training Metrics Evolution Curves

单步推理设置下的损失项消融实验进一步揭示了各损失的独特贡献。仅保留均方误差损失$\mathcal{L}_1$时，模型性能大幅回退，这表明单步采样的严格约束需要额外的感知和频谱监督来补偿，仅靠像素级均方误差损失无法将多步去噪知识压缩至一次前向。当去除对抗损失$\mathcal{L}_{GAN}$时，FID从36.21升至78.53，说明对抗损失对提升生成分布的真实性至关重要。若移除感知损失$\mathcal{L}_{LPIPS}$，LPIPS与FID分别升至0.584和75.28，验证了感知损失在提升人眼视觉自然度方面的关键作用。特别地，当去除聚焦频率损失$\mathcal{L}_{FRE}$时，LPIPS上升0.037，同时SSIM指标虚高，这是因为缺乏频谱约束造成的生成图像过度平滑，进一步验证了聚焦频率损失在保持细节与结构平衡方面的独特价值。而引入完整多项损失联合的第二阶段模型取得了较好的综合性能。虽然第二阶段单步采样与第一阶段直接联合优化相比在指标上略有逊色，但推理速度与GAN相当，有效弥合了扩散模型在S2O任务中的效率鸿沟。

2.6　模型训练过程分析与讨论

为深入剖析两阶段训练策略的动态影响，我们在WHU-OPT-SAR数据集上监测了关键指标随训练进程的演化趋势。如图8显示，第二阶段训练后，

PSNR与LPIPS得到了持续且显著的提升，而SSIM则基本保持稳定。

表8 损失函数与推理步数对模型性能影响的消融实验

Table 8 Ablation study on the effect of loss function and inference steps on model performance

训练设置	推理步数	$\mathcal{L}_{\text{LPIPS}}$	$\mathcal{L}_{\text{FRE}}$	$\mathcal{L}_{\text{GAN}}$	PSNR↑	SSIM↑	LPIPS↓	FID↓
预训练	50	▮	▮	▮	14.63	0.463	0.699	168.51
预训练+全损失优化	50	▮	▮	▮	**25.82**	0.473	**0.427**	**35.67**
预训练+单步推理	1	▮	▮	▮	11.55	0.257	0.824	216.63
预训练+单步推理+损失优化	1	▮	▮	▮	24.72	0.389	0.469	78.53
预训练+单步推理+损失优化	1	▮	▮	▮	24.96	0.425	0.584	75.28
预训练+单步推理+损失优化	1	▮	▮	▮	25.15	**0.517**	0.481	56.15
预训练+单步推理+全损失优化	1	▮	▮	▮	25.44	0.436	0.444	36.21

注：▮表示不使用该损失函数，▮表示使用该损失函数，预训练指文中的第一阶段，单步推理指文中第二阶段，损失优化与单步推理都是在预训练得到的模型权重基础上进行的后续训练，实验在WHU-OPT-SAR数据集上进行。

这一现象可通过指标本质与训练目标的差异得到解释。PSNR的提升主要得益于第二阶段引入的聚焦频率损失显式约束了生成图像与真实图像在频谱中高频分量上的一致性，从而进一步降低了像素级误差。LPIPS的显著改善则归功于针对感知质量设计的感知损失与对抗损失，它们共同推动了生成图像在视觉真实感与纹理细节上向真实图像靠近。SSIM的稳定则表明，模型在第一阶段已较好地掌握了图像的主体结构；第二阶段在优化高频细节与感知质量时引入的细微纹理变化，可能轻微改变了局部对比度，这恰好反映了SSIM与人类视觉感知在评价维度上的不同侧重。

综上所述，两阶段优化所引发的指标分化现象，本质上是模型在计算效率、感知真实性与像素/结构保真度上做出的最优权衡。模型以SSIM可接受的微小波动为代价，换取了LPIPS的显著改善与PSNR的同步提升，并最终实现了接近实时的推理速度，验证了该训练策略的有效性。

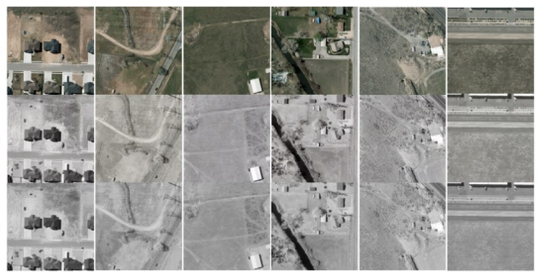
表9 基于VEDAI数据集的对比实验

Table 9 Quantitative comparison of different methods on the VEDAI dataset

模型	PSNR↑	SSIM↑	LPIPS↓	FID↓
CycleGAN	6.25	0.099	0.854	363.4
MUNIT	18.82	0.565	0.345	138.6
EnCo	13.02	0.516	0.488	209.0
SN-DCR	18.94	0.600	0.373	179.7
PatchGGL	16.20	0.664	0.442	240.3
SFSG-Diff	**25.73**	**0.767**	**0.234**	**75.1**

2.7 拓展到其他图像翻译任务

本文的SFSG-Diff模型可以应用于其他的图像翻译任务。为了验证该模型的通用性与可迁移性，在可见光至红外图像翻译任务上进行了实验验证。实验在VEDAI数据集（Razakarivony和Jurie，2016）上进行，仅需更换对应数据集并根据其规模合理调整训练迭代次数。



（(a)可见光图像；(b)生成图像；(c)红外图像）

((a)visible image; (b)synthetic image; (c)infrared image)

图9 在VEDAI上的可视化比较结果

Fig. 9 Comparison of visualization results on the VEDAI

表9展示了SFSG-Diff模型在VEDAI数据集上与其他先进方法的定量比较，可以发现SFSG-Diff模型在红外至可见光图像翻译任务上同样取得了良好的定量结果。对比方法包括CycleGAN（Zhu等，2017）、MUNIT（Huang等，2018）、EnCo（Wang等，

2022)，SN-DCR(Zhao 等，2025b)和 PatchGGL(Jung 等，2022)。可视化结果(图9)进一步显示，本方法生成的红外图像在突出热目标(如车辆、行人)的同时，背景细节自然，与真实红外图像在视觉上高度吻合。

该实验表明，SFSG-Diff所提出的空频特征提取与稀疏条件引导机制具有一定的普适性，能够有效迁移至其他具有不同成像物理机理的跨模态图像翻译任务中，为框架的进一步推广提供了实证支持。

3 结 论

本文针对SAR图像中乘性噪声干扰、跨模态语义对齐困难和计算效率低下等问题，提出了一种基于空频强稀疏引导的扩散模型SFSG-Diff以实现S2O跨模态翻译。该方法通过并行的多尺度空频去噪编码器(MDE)显式分离噪声与信号，并利用轻量的强稀疏语义融合模块(SIF)实现高效精准的条件引导。结合两阶段训练策略，在保证生成质量的前提下，将模型推理优化至单步采样，显著提升了模型的推理速度。实验结果表明，该方法在多项定量指标与视觉质量上均优于现有方法，同时在推理速度上达到了与GAN方法相当的水平，验证了其在精度与速度上的综合优势。

当前方法的局限主要在于对极端噪声场景的适应性和输入分辨率的灵活性有待提高。未来工作将集中于增强模型鲁棒性、探索自适应网络结构及进一步压缩推理成本，并将其拓展至更广泛的遥感跨模态任务中。

参考文献(References)

Bai X Y, Pu X Y and Xu F. 2023. Conditional diffusion for SAR to optical image translation. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2023.3343456]

Bai X Y and Xu F. 2024. SAR to optical image translation with color supervised diffusion model//IGARSS 2024 - 2024 IEEEInternational Geoscience and Remote Sensing Symposium, Athens, Greece, 2024, pp. 963-966 [DOI: 10.1109/IGARSS53475.2024.10640647]

Dong C, Loy C C and Tang X O. 2016. Accelerating the super-resolution convolutional neural network//European Conference on Computer Vision (ECCV). Amsterdam: Springer: 391-407 [DOI: 10.1007/978-3-319-46475-6_25]

Du D, Gu Y and Liu T. 2025. A diffusion-guided task decomposition framework for SAR-to-optical image translation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18: 27597-27614 [DOI: 10.1109/JSTARS.2025.3623105]

Fu S L, Xu F and Jin Y Q. 2021. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks. Science China Information Sciences, 64(2): 122301 [DOI: 10.1007/s11432-020-3077-5]

Goodman J W. 1976. Some fundamental properties of speckle. Journal of the Optical Society of America, 66(11): 1145-1150 [DOI: 10.1364/JOSA.66.001145]

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]

Huang X, Liu M Y, Belongie S and Kautz J. 2018. Multimodal unsupervised image-to-image translation//Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III. Munich: Springer: 179-196 [DOI: 10.1007/978-3-030-01219-9_11]

Huang M Y, Xu Y, Qian L X, Shi W L, Zhang Y Q, Bao W, Wang N, Liu X J and Xiang X S. 2021. The qxs-saropt dataset for deep learning in sar-optical data fusion[EB/OL]. [2025-08-05]. https://arxiv.org/pdf/2103.08259.pdf

Huang Y, Cheng B, Fang S J and Liu X. 2025. Shadow removal with wavelet-based non-uniform diffusion model. Journal of Image and Graphics, 30(01): 0066-0082 (黄颖，程彬，房少杰，刘歆. 2025. 用于阴影去除的小波非均匀扩散模型. 中国图象图形学报, 30(01): 0066-0082) [DOI: 10.11834/jig.230904]

Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 1125-1134 [DOI: 10.1109/CVPR.2017.632]

Jung C Y, Kwon G H and Ye J C. 2022. Exploring patch-wise semantic relation for contrastive learning in image-to-image translation tasks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 18239-18248 [DOI: 10.1109/CVPR52688.2022.01772]

Kim S H and Chung D. 2025. Conditional Brownian bridge diffusion model for VHR SAR to optical image translation. IEEE Geoscience and Remote Sensing Letters, 22: 1-5 DOI: 10.1109/LGRS.2025.3562401

Li B, Xue K T, Liu B and Lai Y K. 2023. BBDM: Image-to-image translation with Brownian bridge diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 1952-1961 [DOI: 10.1109/CVPR52729.2023.00194]

Li M L, Qian Z X and Zhang X P. 2026. Survey on artificial intelligence-

generated image detection. Journal of Image and Graphics, 31(1): 0013-0044 (李美玲, 钱振兴, 张新鹏. 2026. 人工智能生成图像检测技术综述. 中国图象图形学报, 31(1): 0013-0044) [DOI: 10.11834/jig.250053]

Li X, Zhang G, Cui H, Hou S S, Wang S Y, Li X, Chen Y J, Li Z J and Zhang L. 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]

Naderi Darbaghshahi F, Mohammadi M R and Soryani M. 2022. Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation. IEEE Transactions on Geoscience and Remote Sensing, 60: 4105309 [DOI: 10.1109/TGRS.2022.3146908]

Nichol A Q and Dhariwal P. 2021. Improved denoising diffusion probabilistic models[EB/OL]. [2025-06-10]. https://arxiv.org/pdf/2102.09672.pdf

Park T, Liu M Y, Wang T C and Zhu J Y. 2019. Semantic image synthesis with spatially-adaptive normalization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 2337-2346 [DOI: 10.1109/CVPR.2019.00244]

Qin J, Wang K, Zou B, Zhang L and van de Weijer J. 2024a. Conditional diffusion model with spatial-frequency refinement for SAR-to-optical image translation. IEEE Transactions on Geoscience and Remote Sensing, 62: 5226914 [DOI: 10.1109/TGRS.2024.3491826]

Qin J, Zou B, Zhang L M and Qiu Y. 2024b. SAR-to-optical image translation using conditional denoising diffusion probabilistic models//2024 International Radar Conference (RADAR). Marseille: IEEE: 1-5 [DOI: 10.1109/RADAR58632.2024.10837562]

Razakarivony S and Jurie F. 2016. Vehicle detection in aerial imagery: A small target detection benchmark. Journal of Visual Communication and Image Representation, 34: 187-203 [DOI: 10.1016/j.jvcir.2015.11.002]

Salimans T and Ho J. 2022. Progressive distillation for fast sampling of diffusion models//ICLR 2022 [DOI: 10.48550/arXiv.2202.00512]

Schmitt M, Hughes L H and Zhu X X. 2018. The sen1-2 dataset for deep learning in sar-optical data fusion[EB/OL]. [2025-08-15]. https://arxiv.org/pdf/1807.01569.pdf

Song J M, Meng C L and Ermon S. 2020. Denoising diffusion implicit models[EB/OL]. [2025-06-05]. https://arxiv.org/pdf/2010.02502.pdf

Song Y, Dhariwal P, Chen M and Sutskever I. 2023. Consistency models//Proceedings of the 40th International Conference on Machine Learning (ICML 2023. PMLR, 202: 32211-32252 [DOI: 10.48550/arXiv.2303.01469]

Su X Z, Jia D X, Wu F G, Zhao J S, Zheng C W and Qiang W W. 2024. Unbiased image synthesis via manifold guidance in diffusion models//2024 IEEE International Conference on Multimedia and Expo (ICME). Niagara Falls: IEEE: 1-6 [DOI: 10.1109/ICME57554.2024.10687809]

Wang C Y, Liao H Y M, Wu Y H, Chen P Y, Hsieh J W and Yeh I H. 2020. CSPNet: A new backbone that can enhance learning capability of CNN//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE: 390-391 [DOI: 10.1109/CVPRW50498.2020.00203]

Wang W, Zhou W, Bao J, Chen D and Li H. 2022. Contrastive learning for unpaired image-to-image translation//Advances in Neural Information Processing Systems 35. New Orleans: Curran Associates, Inc.: 23882-23895 [DOI: 10.48550/arXiv.2205.14267]

Wang Y J, Gong J X, Liang Z B, Chong Q P, Cheng X and Xu J D. 2025. Fuzzy diffusion model for seen-through document image restoration. Journal of Image and Graphics, 30(4): 1118-1129 (王义杰, 龚嘉鑫, 梁宗宝, 崇乾鹏, 程翔, 徐金东. 2025. 面向透射文档图像复原的模糊扩散模型. 中国图象图形学报, 30(4): 1118-1129) [DOI: 10.11834/jig.240350]

Yang X, Wang Z H, Zhao J Y and Yang D. 2022a. FG-GAN: A fine-grained generative adversarial network for unsupervised SAR-to-optical image translation. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-11 [DOI: 10.1109/TGRS.2022.3222867]

Yang X, Zhao J, Wei Z, Wang N and Gao X. 2022b. SAR-to-optical image translation based on improved CGAN. Pattern Recognition, 121: 108208 [DOI: 10.1016/j.patcog.2021.108208]

Yang X, Shi H, Wang Z, Wang N and Gao X. 2026. CSHNet: A novel information asymmetric image translation method. IEEE Transactions on Circuits and Systems for Video Technology, 36(2): 1862-1875 [DOI: 10.1109/TCSVT.2025.3612484]

Yue Z S, Wang J Y and Loy C C. 2023. ResShift: Efficient diffusion model for image super-resolution by residual shifting [EB/OL]. [2025-08-08]. https://proceedings.neurips.cc/paper_files/paper/2023/file/2ac2eac5098dba08208807b65c5851cc-Paper-Conference.pdf

Yu J, Lin Z, Yang J, Shen X, Lu X and Huang T S. 2018. Free-form image inpainting with gated convolution//IEEE International Conference on Computer Vision (ICCV). Seoul: IEEE: 4471-4480 [DOI: 10.1109/ICCV.2018.00468]

Zhang L M, Rao A Y and Agrawala M. 2023. Adding conditional control to text-to-image diffusion models//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 3836-3847 [DOI: 10.1109/ICCV51070.2023.00355]

Zhao S H, Chen D D, Chen Y C, Bao J M, Hao S Z, Yuan L and Wong K Y K. 2023. Uni-ControlNet: All-in-one control to text-to-image diffusion models[EB/OL]. [2025-09-13]. https://proceedings.neurips.cc/paper_files/paper/2023/file/2468f84a13ff8bb6767a67518fb596eb-Paper-Conference.pdf

Zhao W B, Jiang N N, Liao X X and Zhu J B. 2025a. HVT-cGAN: Hybrid Vision Transformer cGAN for SAR-to-Optical Image Trans-

lation. IEEE Transactions on Geoscience and Remote Sensing, 63: 1-17 [DOI: 10.1109/TGRS.2024.3523040]

Zhao C, Cai W L and Yuan Z. 2025b. Spectral normalization and dual contrastive regularization for image-to-image translation. The Visual Computer, 41(1): 129-140 [DOI: 10.1007/s00371-024-03314-5]

Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE: 2223-2232 [DOI: 10.1109/ICCV.2017.244]

作者简介

邢怡楠，2001年生，女，硕士研究生，研究方向为遥感图像处理。E-mail：xingyinan@mail.sdufe.edu.cn

刘博，通信作者，男，讲师，主要研究方向为多源遥感图像处理。E-mail：liubo24@sdufe.edu.cn

张云峰，男，教授，主要研究方向为计算机视觉与人工智能。E-mail：yfzhang@sdufe.edu.cn

任玥赫，男，硕士研究生，主要研究方向为计算机视觉。E-mail：242115017@mail.sdufe.edu.cn