

基于GAN与NSST结合的遥感图像时空融合算法

王易飞¹, 保文星¹, 屈克文¹, 冯伟^{1,2}, 王文龙^{3,4}

1. 北方民族大学 计算机科学与工程学院, 银川 750021;

2. 西安电子科技大学 电子工程学院, 西安 710071;

3. 宁夏回族自治区遥感调查院, 银川 750021;

4. 高分辨率对地观测系统宁夏数据与应用中心, 银川 750021

摘要: 针对目前基于深度学习的时空融合方法在卷积操作时会平衡局部区域内像素误差, 致使融合图像纹理细节差, 区域边缘模糊的问题。本文提出了一种结合生成对抗网络 (GAN) 和非下采样剪切波变换 (NSST) 的时空融合模型。首先, 将膨胀卷积分支和差值一注意力模块 (SUB-AM) 引入生成器的编码器结构, 使精细特征提取时能兼顾全局性, 并利用 SUB-AM 模块捕获的粗略特征变化调整所提取的精细特征, 以提高预测精度。其次, 通过 NSST 分解与自适应重构块 (NSST-ARM) 对编码器—解码器生成的结果和参考时刻精细图像的高低频信息重构, 增强融合图像的高频细节; 最后, 在损失函数中引入清晰度损失项, 以减少融合图像的清晰度损失。实验结果表明, 与最先进方法相比, 本文方法结构相似性 (SSIM) 值平均提高了 1.5%, 相对无量纲全局误差 (ERGAS) 值平均减少了 14%, 融合图像在纹理细节表达、边缘结构清晰度方面均有提升。实验证明了本文方法的有效性。

关键词: 时空融合, 生成对抗网络, 膨胀卷积, 非下采样剪切波变换

中图分类号: TP701/P2

引用格式: 王易飞, 保文星, 屈克文, 冯伟, 王文龙. 2026. 基于 GAN 与 NSST 结合的遥感图像时空融合算法. 遥感学报, 30(6): 1775–1791

Wang Y F, Bao W X, Qu K W, Feng W and Wang W L. 2026. Spatiotemporal fusion algorithm for remote sensing images based on the combination of GAN and NSST. National Remote Sensing Bulletin, 30(6): 1775–1791 [DOI: 10.11834/jrs.20264591]

1 引言

随着遥感技术的发展, 遥感进入了一个更大范围、更高精度、更深层次的应用阶段 (Dhillon 等, 2022)。它成为了地球观测、环境监测、资源调查的重要手段 (Husman 等, 2023; Francini 等, 2022; Jafarbiglu 和 Pourreza, 2022)。但由于传感器硬件和技术等限制, 单个卫星获得的遥感图像分辨率特性往往是比较单一的 (Jing 等, 2023)。为满足各种应用需求, 需要将来自不同卫星的不同分辨率遥感图像融合。

遥感图像融合包括了高光谱图像 (HIS) 和多光谱图像 (MSI) 融合以及时空融合等。前者可以

获取到高空间和高光谱分辨率兼具的图像, 这类融合大都基于模型驱动、数据驱动或者模型数据驱动 (Yan 等, 2025)。其中模型数据驱动领域目前研究较多, 如基于无监督深度张量网络的高光谱—多光谱图像融合法 (UDTN) (Yang 等, 2024), 该类方法泛化性更高。后者时空融合可以对于单一卫星无法获得高时空分辨率的遥感图像提供帮助 (Li 等, 2020b)。

时空融合就是将高时间分辨率遥感图像的粗略特征变化和高空间分辨率图像的精细特征融合, 生成高时空分辨率的图像 (Li 等, 2021)。

时空融合方法分为 5 类: 基于解混的方法、基于权函数的方法、基于贝叶斯的方法、基于学习

收稿日期: 2024-12-30; 预印本: 2026-03-31

基金项目: 国家自然科学基金 (编号: 62461001); 宁夏自然科学基金项目重点项目 (编号: 2024AAC02035); 国家民委创新团队, 视觉感知与计算宁夏重点实验室

第一作者简介: 王易飞, 研究方向为遥感图像处理、机器学习。E-mail: 20227441@stu.nmu.edu.cn

通信作者简介: 保文星, 研究方向为数字图像处理、遥感图像分类和融合。E-mail: bwx71@163.com

的方法和基于混合的方法。基于解混的方法是根据线性光谱混合理论,将低空间分辨率像素分解来估计高空间分辨率像素。如时空数据融合法(STDFA)(Wu等,2012)和灵活时空数据融合法(FSDAF)(Zhu等,2016)等。该类方法不能适应地物覆盖突变情况。基于权函数的方法是通过权函数综合所有输入图像信息来估计精细像素。如时空自适应反射率融合模型(STARFM)(Gao等,2006)、增强自适应反射率融合模型(ESTARFM)(Zhu等,2010)等。这类方法提高了融合精度,但不适应异质区域。基于贝叶斯的方法融合了图像时间序列中的相关信息,将融合问题转为概率问题(Zhang等,2021)。如贝叶斯最大熵融合法(Li等,2013)、空间—光谱和时空融合的统一贝叶斯框架(Huang等,2013)等。它一定程度上解决了景观异质的问题,但融合效率低(Wu等,2020)。基于学习的方法大多是通过字典学习来实现时空融合的,可以很好地捕捉土地覆盖变化。如基于稀疏表示的时空反射率融合模型(SPSTFM)(Huang和Song,2012)。基于混合的方法则整合了解混策略、权重函数和贝叶斯理论来追求更好的性能,如空间和时间反射率解混模型(STRUM)(Gevaert和García-Haro,2015)、增强型线性时空融合方法(Ping等,2018)。

目前基于深度学习的时空融合模型,主要采用卷积神经网络(CNN)、残差网络(ResNet)、生成对抗网络(GAN)或Transformer机制,其泛化性和准确性更高。如Song等(2018)提出的深度卷积神经网络(STFDCNNs),通过超分辨CNN来解决粗细分辨率图像之间的巨大分辨率差距。该方法分为非线性映射卷积神经网络(NLMCNN)学习、超分辨卷积神经网络(SRCNN)学习和图像融合3个阶段。Tan等(2018)提出了一种深度卷积时空融合网络(DCSTFN)。该方法的输入仅需一对先验图像,单一参考依赖会影响预测精度。因此Tan等(2019)提出了增强的DCSTFN(EDCSTFN),以一种新的网络结构和混合损失函数进行时空融合,参考数据和预测数据之间的差异从真实数据中学习,大大提高了预测精度。Li等(2020b)通过考虑上下文信息,提出了一种基于膨胀卷积和多尺度机制的遥感影像时空融合网络(DMNet)。Song等(2022)提出了基于生成对抗网络的多级特征融合方法(MLFF-GAN),通过

多级特征提取和重构缩小粗—细图像间的巨大分辨率差距,并设计自适应实例归一化块(AdaIn)和注意力块(AM)学习特征的全局分布和局部变化。Tan等(2022)提出了一种基于条件式GAN的参考不敏感型融合网络(GAN-STFM),参考图像日期选择是随机的。随后,一种基于Swin Transformer和线性频谱混合理论的SwinSTFM模型出现(Chen等,2022),该模型充分利用Swin Transformer的特征提取优势并将混合分解理论融入到模型中,大大提高了融合精度。

虽然,基于深度学习的模型提升了预测精度,但仍存在一些问题。卷积操作过程中会平衡局部区域内每个像素间的误差,使特征损失达到最小(Ajit等,2020)。故预测图像较为平滑,且容易出现光谱失真和纹理细节丢失。同时,粗分辨率图像特征的时空变化不能完全反映精细图像变化。因此需要一种融合方法在提高预测精度的同时,增强图像清晰度,减少纹理细节丢失。

针对上述问题,本文提出了一种基于GAN和NSST结合的遥感图像时空融合模型(STF-GAN_NSST)。该模型生成器由编码器—解码器模块和NSST分解与自适应重构块(NSST-ARM)组成。编码器中,普通卷积和膨胀卷积两个分支用于提取不同感受野内的精细特征,差值—注意力模块(SUB-AM)用于捕获参考时刻与预测时刻之间粗分辨率特征变化,并结合上述提取的精细特征对待输出的精细特征进行调整。NSST-ARM负责将编码器—解码器生成的初步融合图像和参考时刻精细图像分解为多尺度的高低频系数,并按照自适应融合规则进行重构,加强初步图像的高频特征。同时在复合损失函数中引入了清晰度损失,提高模型训练精度。本文的主要贡献如下:

(1)提出了一种基于GAN和NSST结合的遥感图像时空融合模型。编码块采用 3×3 的普通卷积分支和感受野扩展到 5×5 大小的膨胀卷积分支,分别在不同感受野内提取精细特征,并利用了SUB-AM块输出的粗略特征变化信息对精细特征进行调整。可以兼顾局部细节和全局关系,提高预测精度。

(2)设计了一种NSST分解和自适应重构块。通过NSST将编码—解码器生成的初步融合图像与参考时刻精细图像的亮度分量分解为多尺度高低频系数。并对参考时刻精细图像的像素点进行云

区检测和匹配测度计算。根据云覆盖情况和像素点位置，对初步融合图像和参考时刻图像的高频信息和低频信息分别进行自适应融合，实现图像重构。在保持低频稳定的同时，增强初步融合图像的高频信息，使图像纹理细节更丰富，区域边缘更明显。

(3) 在损失函数中引入清晰度损失，由平均梯度损失和空间频率损失项共同构成，旨在减少训练过程中融合图像的清晰度损失。

(4) 在 CIA、LGC 以及 AHB 数据集上进行实验。与几种主流算法进行了比较，验证了本文所提方法的泛化性和有效性。

2 相关方法与基本原理

2.1 生成对抗网络(GAN)

GAN是一种无监督的深度学习模型 (Goodfellow 等, 2014)，它提供了一种合理的方案来隐式计算真实数据分布。GAN 包括生成器 G 和判别器 D ， G 生成样本来欺骗 D ，使其相信这些样本是真的， D 则区分样本的真伪。 G 和 D 在不断的零和博弈中达到纳什均衡，最终表现为极大极小优化问题。

$$\min_G \max_D \text{Loss} = E_{z \sim p_z} [\log(1 - D(G(z)))] + E_{x \sim p_{\text{data}}} [\log(D(x))] \quad (1)$$

式中， $z \sim p_z$ 和 $x \sim p_{\text{data}}$ 分别表示随机噪声向量和遵循真实的数据分布的输入图像。 $E(\cdot)$ 表示期望。 $D(x)$ 和 $G(z)$ 分别表示通过在真实输入向量上训练得到的 D 的 sigmoid 函数输出，以及通过随机噪声得到的 G 的生成数据。 $D(G(z))$ 给出了输入为 $G(z)$ 时 D 的实际期望。在 GAN 优化过程中， G 和 D 交替优化。 G 和 D 在持续不断地对抗训练中实现了纳什均衡，并产生了期望的结果。

2.2 非下采样剪切波变换(NSST)

NSST 是剪切波技术 (Easley 等, 2008) 的改进，属于多尺度多方向分解。图 1 为 NSST 三级分解过程。首先，通过非下采样金字塔滤波器组对源图像进行多级滤波，获得 1 个低频子带和 3 个高频子带。在每级分解前都会对图像上采样到上级图像的大小，使每级图像大小与源图像相同。然后，把标准的剪切波滤波器从伪极化网格系统映射到笛卡尔坐标系统，利用“Meyer”小波构造窗

函数，获得新剪切波滤波器，通过计算高频子带图像与剪切波滤波器的卷积来获取不同方向的子图像。最后，通过傅里叶逆变换得到平移不变的剪切波系数。

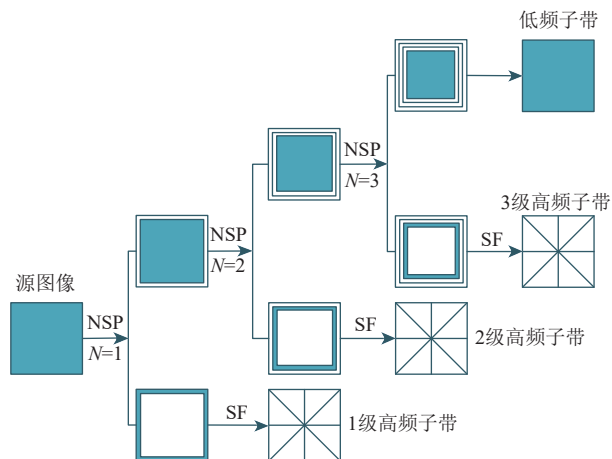


图 1 NSST 分解示意图

Fig. 1 Schematic diagram of NSST decomposition

3 本文所提模型

本文所提 STF-GAN_NSST 模型包含生成器和判别器两部分。以 GAN-STFM 为基线模型，在模型生成器的编码块中引入了膨胀卷积分支、SUB-AM 块和 NSST 分解与自适应重构块，生成器结构如图 2 所示。模型判别器沿用 GAN-STFM 的判别器结构。如图 3 所示。总体上表示为

$$\tilde{F}_t = f(C_{t^*}, F_{t^*}, C_t) \quad (2)$$

式中， C_{t^*} 和 F_{t^*} 分别为参考时刻粗分辨率图像和精分辨率图像。 C_t 和 $\tilde{F}_t$ 为预测时刻粗分辨率图像和预测的精分辨率图像。 t^* 与 t 应尽可能邻近。

3.1 生成器

生成器分两个阶段生成预测图像：首先利用编码器—解码器模块生成初步融合图像 $\hat{F}_t$ 。然后通过 NSST-ARM 模块对图像 $\hat{F}_t$ 和 F_{t^*} 的亮度分量进行多尺度多方向分解和高低频自适应融合，并完成重构，恢复多光谱形式，实现融合图像 $\hat{F}_t$ 高频信息的强化。

模型编码—解码器基于 ResNet (Babu Avvaru 等, 2023) 构建。其中，构成编码器的残差块命名为 GE-ResBlock，构成解码器的残差块命名为 GD-ResBlock。

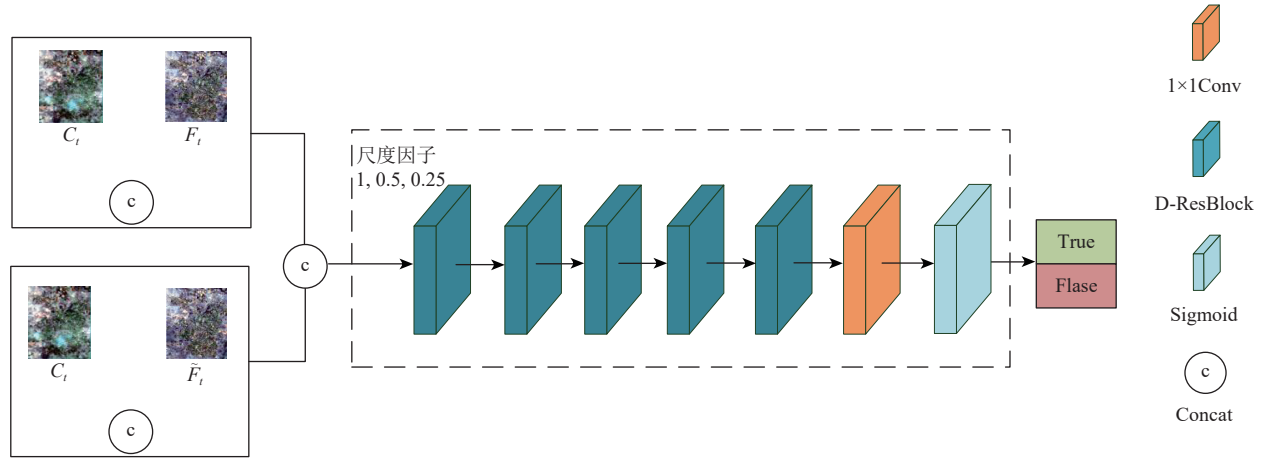
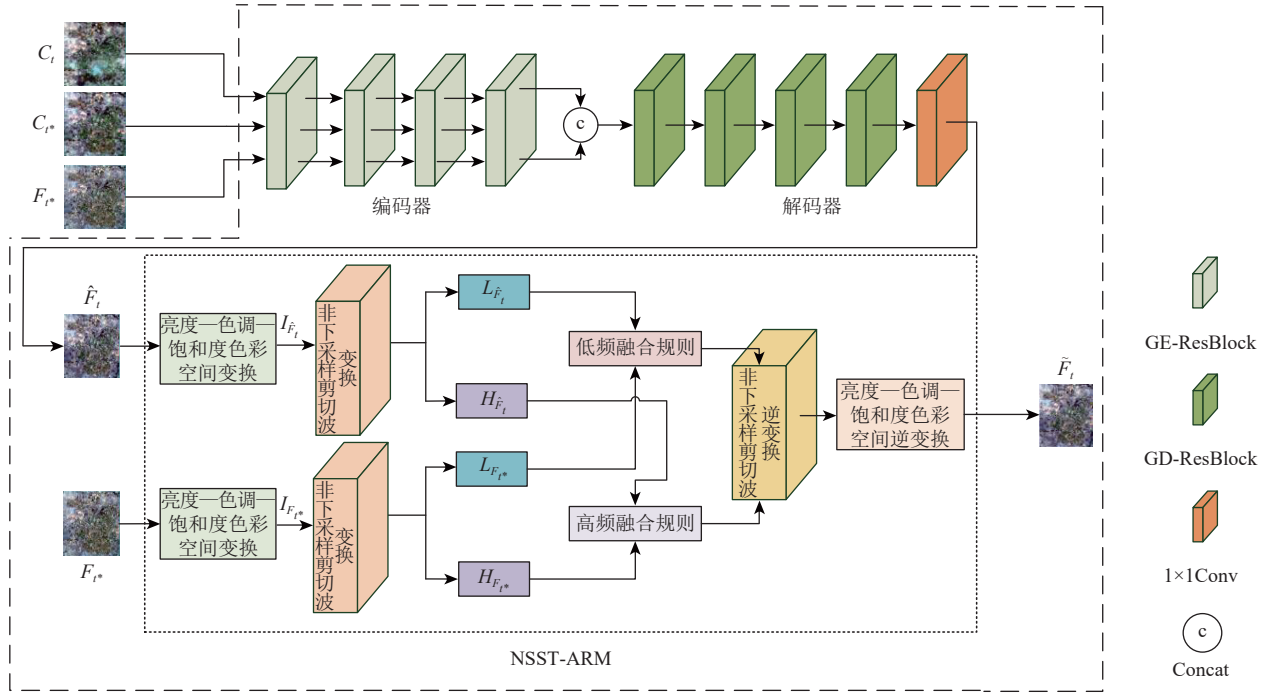


图2中编码器由4个GE-ResBlock串联而成。GE-ResBlock结构如图4(a)所示,分为主分支和侧分支两部分。主分支是在原先GAN-STFM的 3×3 普通卷积分支基础上,引入卷积核大小为 3×3 ,膨胀率为2的膨胀卷积分支,分别在 3×3 和 5×5 大小的感受野内提取参考时刻精细图像的高通特征并结合,通过结合不同感受野的精细特征来使所提取的精细特征兼具局部细节和全局关系。在侧分支引入SUB-AM块,捕获两个时刻粗略特征在时间维度上的变化,并加到主分支的提取结果中以生成中间精细特征。中间精细特征、预测和参考时刻的粗略特征送到下一块处理。GE-ResBlock的

计算见式(3)—(8)。

$$F_1^i = \text{ReLU} \left(\text{SN} \left(\text{Conv} \left(\text{ReLU} \left(\text{SN} \left(F^{i-1} \right) \right) \right) \right) \right) \quad (3)$$

$$F_2^i = \text{ReLU} \left(\text{SN} \left(\text{DConv} \left(\text{ReLU} \left(\text{SN} \left(F^{i-1} \right) \right) \right) \right) \right) \quad (4)$$

$$C_{SA}^i = \text{ReLU} \left(\text{SA} \left(\text{Conv1} \left(C_{t^*}^{i-1} \right), \text{Conv1} \left(C_t^{i-1} \right) \right) \right) \quad (5)$$

$$F^i = F_1^i + F_2^i + C_{SA}^i \quad (6)$$

$$C_{t^*}^i = \text{Conv1} \left(C_{t^*}^{i-1} \right) \quad (7)$$

$$C_t^i = \text{Conv1} \left(C_t^{i-1} \right) \quad (8)$$

式中, i 为GE-ResBlock编号。 F_1^i 和 F_2^i 分别表示普通卷积和膨胀卷积分支提取的高通特征。SA代表SUB-AM模块, C_{SA}^i 为该模块输出的粗特征变化信

息。 F^i 为调整后的精细特征。 F^{i-1} , C_t^{i-1} , C_{ts}^{i-1} 为上一块的输出。 SN 为可切换归一化 (Luo 等, 2018), 用于过滤特定样本信息仅提取常见高通细节。 $ReLU$ 表示LeakyRELU激活函数, 用于加快训练速度, 提高稳定性。 $Conv$ 表示卷积核大小为3,

步长为1的普通卷积, $Dconv$ 表示卷积核大小为3, 步长为1, 膨胀率为2的膨胀卷积。普通卷积采取1层边缘复制填充, 而膨胀卷积为两层边缘复制填充。 $Conv1$ 表示1×1的卷积, 用于调整侧分支特征图通道数, 使其与主分支的输出匹配。

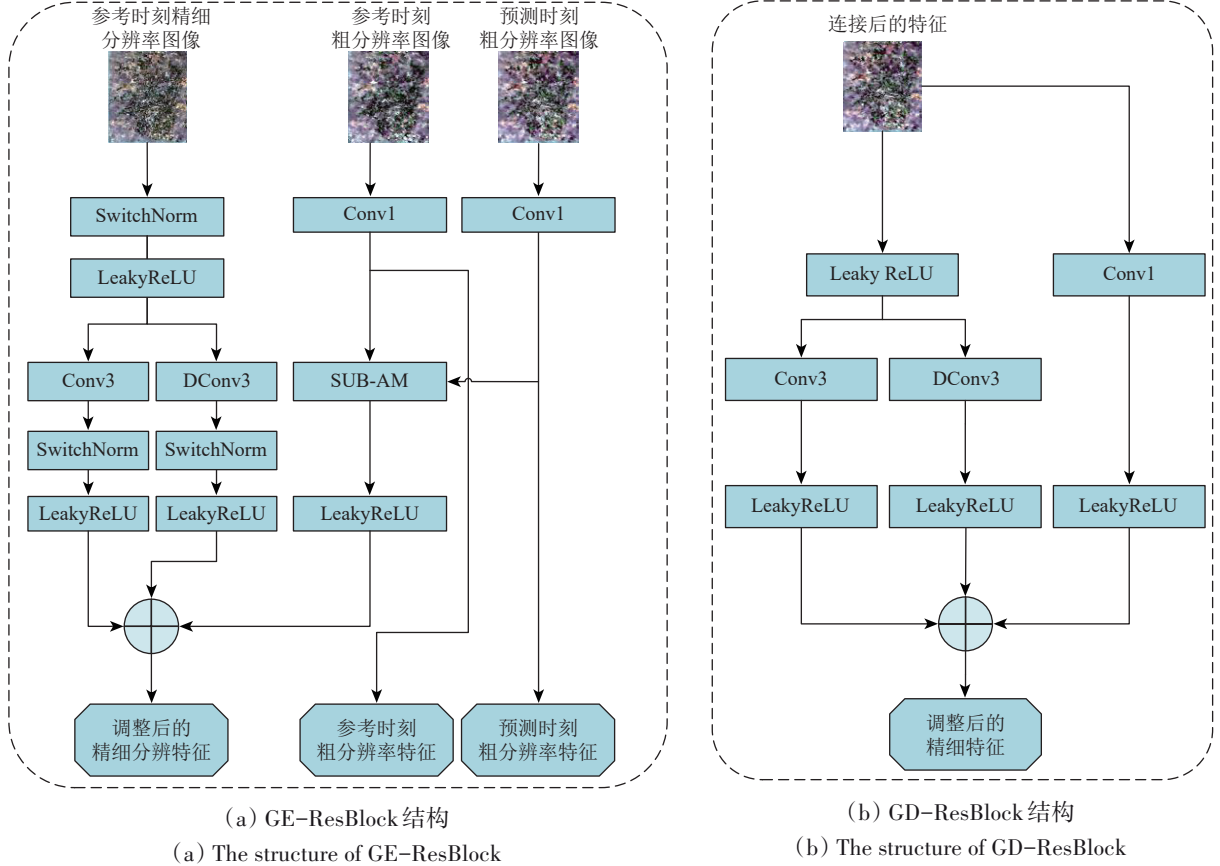


图4 编码器残差块和解码器残差块的内部结构($\oplus$ 表示像素级相加)

Fig. 4 The inner structure of the GE-ResBlock and GD-ResBlock ($\oplus$ represents pixel-level addition)

解码器部分由4个GD-ResBlock块和一个1×1Conv组成, 输入为编码结果中精细特征与预测时刻粗特征的连接。GD-ResBlock结构如图4(b), 除无归一化块和SUB-AM块外, 其余构件与GE-ResBlock类似。连接的特征被主分支和侧分支处理。侧分支输出的粗略特征添加到主分支提取的精细特征中形成中间特征送到下一解码块。GD-ResBlock表示为式(9)–(11)。

$$\hat{F}_1^k = ReLU\left(Conv\left(ReLU\left(\hat{F}^{k-1}\right)\right)\right) \quad (9)$$

$$\hat{F}_2^k = ReLU\left(DConv\left(ReLU\left(\hat{F}^{k-1}\right)\right)\right) \quad (10)$$

$$\hat{F}^k = \hat{F}_1^k + \hat{F}_2^k + LR\left(Conv1\left(\hat{F}^{k-1}\right)\right) \quad (11)$$

式中, k 为GD-ResBlock编号。 $\hat{F}_1^k$, $\hat{F}_2^k$ 分别为普通卷积分支和膨胀卷积分支提取的高通特征。 $\hat{F}^{k-1}$ 为上一块输出的精细特征。 $\hat{F}^k$ 为当前解码块的输

出。 $ReLU$ 、 $Conv$ 、 $DConv$ 、 $Conv1$ 表示的意义与GE-ResBlock相同。

3.2 判别器

判别器采用多尺度结构对预测结果的真伪进行判断。构成判别器的残差块命名为D-ResBlock。判别器由4个D-ResBlock块、1×1Conv和Sigmoid函数组成。输入有预测结果 $\tilde{F}_t$ 、预测时刻粗分辨率图 C_t 和真实精细分辨率图 F_t 。其中, C_t 作为条件来指导训练, 使网络朝正确方向进行。判别器在输入的下采样因子为1, 0.5, 0.25的3个尺度下进行训练。图5展示了D-ResBlock块的结构, 表示为式(12)。

$$F_M^i = ReLU\left(BN\left(SPN\left(StridedConv3\left(ReLU\left(BN\left(F_{stake}^{j-1}\right)\right)\right)\right)\right) + SPN\left(Conv1\left(F_{stake}^{j-1}\right)\right)\right) \quad (12)$$

式中, $ReLU$ 、 BN 、 SPN 分别表示LeakyRELU激活

函数、批量归一化和光谱归一化。 $ReLU$ 用于增强模型非线性。 BN 可以加快模型收敛,提高精度。 SPN 用于稳定GAN训练中的对抗过程,保证可学习隐函数的Lipschitz连续性。 $StridedConv3$ 表示步长为2的 3×3 卷积。 $Conv1$ 是大小为 1×1 的卷积,用于调整通道数。 F_M^j 为D-ResBlock的输出。 $j=1$ 时 F_{stake}^{j-1} 为判别器的原始输入,是 $\tilde{F}_t$, C_t , F_t 等3幅图像的维度连接。

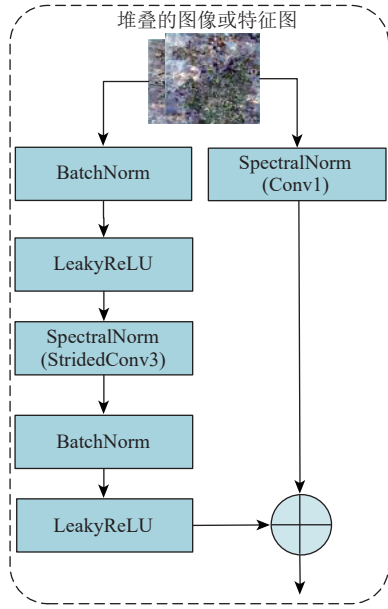


图5 D-ResBlock结构($\oplus$ 表示像素级相加)

Fig. 5 The structure of the D-ResBlock ($\oplus$ represents pixel-level addition)

3.3 SUB-AM 模块

SUB-AM是编码器的关键部分,用于捕获参考时刻和预测时刻粗分辨率特征在时间维度上的局部变化。它利用空间注意力机制从 C_t^i 和 $C_{t^*}^i$ 中学习权重,组成新的粗略特征 C_{SA}^i 。后续会被添加到主分支

提取的精细特征中,对精细特征进行调整。公式为

$$C_{SA}^i = M^i C_t^i + (1 - M^i) C_{t^*}^i \quad (13)$$

式中, i 为GE-ResBlock的编号。 M^i 表示SUB-AM块通过网络方向传播学习到的权重映射图,值范围为 $[0, 1]$,由式(18)求出。权重代表粗略特征的变化程度,值越大变化越大,分配给 C_t^i 的比例越大。 C_t^i 和 $C_{t^*}^i$ 分别表示参考时刻和预测时刻的粗略特征。

图6展示了SUB-AM的结构。 C_t^i 和 $C_{t^*}^i$ 作为该模块的输入。首先将它们进行通道维度上的连接,得到特征图 $F \in R^{H \times W \times C}$,其中 $H \times W$ 表示特征图大小, C 为通道数。 F 经过卷积层和 $Relu$ 层进行局部特征提取,以限制模型复杂度并提高泛化性。对提取的特征分别进行最大池化和平均池化,聚合特征图的通道信息。然后将最大池化和平均池化结果进行像素级相加,并经过单层卷积和 $Relu$ 激活。最后利用 $Sigmoid$ 函数输出空间注意力权重矩阵 M^i ,用于对粗图像 C_t^i 和 $C_{t^*}^i$ 进行组合。如式(14)–(18)。

$$F = Concat(C_t^i, C_{t^*}^i) \quad (14)$$

$$F_{CoRL} = ReLU(Conv(F)) \quad (15)$$

$$F_{avg} = AvgPool(F_{CoRL}) = \frac{\sum_{i=1}^H \sum_{j=1}^W F_{CoRL}(i, j)}{H \times W} \quad (16)$$

$$F_{max} = MaxPool(F_{CoRL}) = \max(F_{CoRL}) \quad (17)$$

$$M^i = Sigmoid(ReLU(Conv(F_{max} + F_{avg}))) \quad (18)$$

式中, $Concat(\cdot)$ 表示通道维度上的连接。 F_{CoRL} 为经过 3×3 的单层卷积 $Conv$ 和 $LeakReLU$ 激活后的特征。 F_{avg} 和 F_{max} 分别表示 F_{CoRL} 的平均池化和最大池化结果。 $Sigmoid$ 为激活函数, M^i 用于输出SUB-AM模块注意力权重图,以求解粗略特征 C_{SA}^i 。

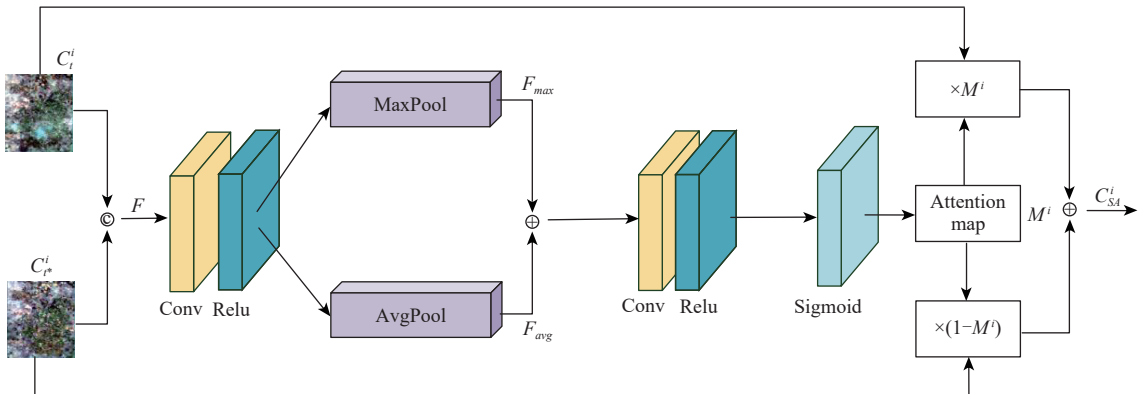


图6 SUB-AM模块结构($\odot$ 表示特征维度连接, $\oplus$ 表示像素级相加)

Fig. 6 The structure of SUB-AM($\odot$ represents a concat of the feature dimension, $\oplus$ represents pixel-level addition)

3.4 NSST-ARM 模块

本文设计了NSST分解与自适应融合模块，命名为NSST-ARM。在编码器—解码器模块生成图像 $\hat{F}_i$ 后，为进一步增强纹理细节，模型会通过NSST-ARM块对 $\hat{F}_i$ 和参考时刻精细图像 F_{i^*} 进行NSST多尺度分解，得到高低频系数，并在设计的自适应融合规则下对 $\hat{F}_i$ 实行重构，完成高频补充。

首先从图像 $\hat{F}_i$ 和 F_{i^*} 中分别取4, 3, 2波段组合为RGB图像，转换为IHS彩色空间 (Zhang 和 Hong, 2005)，并选择亮度分量 I_i 和 I_{i^*} ，单独进行融合。完成后再经过一系列反变换得到最终融合图像 $\tilde{F}_i$ 。

I_i 和 I_{i^*} 经NSST三级分解后，各自会得到一个低频子带和3个高频子带。低频子带记为 L_{F_i} 和 $L_{F_{i^*}}$ ，高频子带集记为 $H_{F_i}^{l,k}$ 和 $H_{F_{i^*}}^{l,k}$ 。

I_i 和 I_{i^*} 的融合流程为：首先采用改进的分形网络演化算法 (FNEA) (Gao 等, 2010) 对图像 F_{i^*} 进行聚类分割；接着综合光谱和形状特征对 F_{i^*} 进行云区检测；然后根据云覆盖情况和自适应融合规则对 I_i 和 I_{i^*} 的高低频系数进行融合。

其中，使用改进的FNEA目的是对图像 F_{i^*} 进行预处理，通过预先的区域分割来检测该精细图像是否有云覆盖以及云覆盖位置，以此来确定高频或低频特征融合时是否需要参考精细图像该位置的特征。由于数据集图片云覆盖要求低于10%，所以大部分图片没有或只有少量云，只需要进行简单分割和云检测辅助判断。云层有更显著的颜色特征，所以区域分割过程中更应倾向光谱权重。根据经验，对FNEA的光谱权重和形状权重分别设为0.7和0.3，尺度参数设置为1000。

融合规则的制定考虑引入匹配测度，通过计算每个像素点的匹配测度，确定像素点的位置类型，从而选择合适的高低频系数融合方式。匹配测度计算采用8邻域检测法，待处理的中心点设为 f_{x_0, y_0} ，它周围的相邻点设为 f_{x_i, y_j} ，其中 i 和 j 的取值为 $[1, 2, 3]$ 。依据图像区域特征，图像的局部特征可以简化为3种情况，如图7。

图7 (a) 表示中心点 f_{x_0, y_0} 为区域内部点时，周围像素点和该中心点相似，标记匹配测度 $w_{AB}=8$ 。图7 (b) 表示 f_{x_0, y_0} 为孤立点的时，该中心点与周围点完全不同，记 $w_{AB}=0$ 。图7 (c) 表示中心点

f_{x_0, y_0} 处于两区域交界处或边缘时，属于过渡带，周围点与 f_{x_0, y_0} 相似的个数为1—7，所以记 $1 \leq w_{AB} \leq 7$ 。

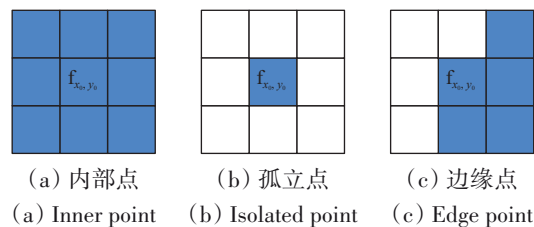


图7 像素点的3种位置类型

Fig. 7 The three types of positions for the pixel point

低频融合：低频分量的选择要考虑云区的影响和像素点是区域中心还是边缘。对于云遮挡区域，参考时刻亮度分量 I_{i^*} 的低频系数没有参考意义。对无云区域需计算 I_{i^*} 像素点的匹配测度，判断该点为区域中心、边缘还是孤立点，进而选择合适的低频融合方式。低频融合公式归纳为

$$L_{\tilde{F}_i, ij} = \begin{cases} L_{F_i, ij} & M = 1 \\ L_{F_i, ij} & M = 0, w_{AB} = 0 \\ L_{F_i, ij} & M = 0, w_{AB} = 8 \\ (1-w) \times L_{F_i, ij} + w \times L_{F_{i^*}, ij} & M = 0, 1 \leq w_{AB} \leq 7 \end{cases} \quad (19)$$

式中， $L_{\tilde{F}_i, ij}$ 为融合后的低频系数。 (i, j) 为像素位置， M 是局部区域云覆盖情况，值为1表示有云，0表示无云。 w_{AB} 为像素的匹配测度， w 为权值，根据 I_i 和 I_{i^*} 区域能量得出，公式为

$$E = \sum_i^m \sum_j^n f(x_i, y_j) \quad (20)$$

$$w = \frac{E_{I_{i^*}}}{E_{I_i} + E_{I_{i^*}}} \quad (21)$$

式中， E 表示区域能量。 m, n 表示区域大小。 E_{I_i} 和 $E_{I_{i^*}}$ 表示 I_i 和 I_{i^*} 的区域能量。

根据式 (19)，低频系数的融合规则可以表述为：

(1) 如果是云遮挡区 ($M=1$)，则亮度分量 I_{i^*} 的低频系数没有参考意义，保持原始图像 I_i 的低频系数 $L_{F_i, ij}$ 。

(2) 如果是无云区 ($M=0$)，且 $w_{AB}=0$ 或8，则该点为孤立点或区域内部点。为内部点时，要保持原始低频细节，故采用 $L_{F_i, ij}$ 。为孤立点时，被认为是噪声点，予以剔除，故也选择原始低频系数 $L_{F_i, ij}$ 。

(3) 如果是无云区 ($M=1$)，且 $1 \leq w_{AB} \leq 7$ ，

则该点为边缘点, 包含了两个区域丰富的低频信息, 有必要借鉴 I_{t^*} , 所以通过加权平均对 $L_{F_i, ij}$ 和 $L_{F_{t^*}, ij}$ 进行融合, 以区域能量 w 作为权值。

高频融合: 高频系数反映图像的纹理细节, 为了使初步融合图像 $\hat{F}_i$ 更清晰, 需要进行高频补充。高频选择同样要根据云区情况和像素点的匹配测度判断。对于有云区, 参考时刻分量 I_{t^*} 的高频无法借鉴, 强行融合会影响精度。对于无云区, 需基于匹配测度和改进的 PCNN 模型 (Blasch, 1999) 计算的模拟点火次数对高频进行选择。点火次数可反映图像高频特征的丰富度。点火次数越多, 越丰富。高频系数的绝对值作为 PCNN 网络的反馈输入 D_{ij} 。而迭代次数为 N 时, 高频系数 $H_{F_i}^{l, k}$ 和 $H_{F_{t^*}}^{l, k}$ 对应的模拟点火次数 $T_{F_i, ij}^{l, k}[N]$ 和 $T_{F_{t^*}, ij}^{l, k}[N]$ 可以根据 PCNN 模型求解 (Jiao 和 Wu, 2019), 每次迭代结束后模拟点火触发的次数总和 $T(n) = T(n-1) + Y(n)$ 得出。高频融合公式归纳为

$$H_{\hat{F}_i, ij}^{l, k} = \begin{cases} H_{F_i, ij}^{l, k} & M=1 \\ H_{F_i, ij}^{l, k} & M=0, w_{AB}=0 \\ w \times H_{F_i, ij}^{l, k} + (1-w) \times H_{F_{t^*}, ij}^{l, k} & M=0, w_{AB}=8, T_{F_i, ij}^{l, k}[N] \geq T_{F_{t^*}, ij}^{l, k}[N] \\ H_{F_{t^*}, ij}^{l, k} & M=0, 1 \leq w_{AB} \leq 7 \\ H_{F_i, ij}^{l, k} & M=0, w_{AB}=8, T_{F_i, ij}^{l, k}[N] < T_{F_{t^*}, ij}^{l, k}[N] \end{cases} \quad (22)$$

式中, $H_{F_i, ij}^{l, k}$ 为第 l 层 k 方向上融合后的高频系数。 (i, j) 为像素位置, M 为云量情况, 1 表示有云, 0 表示无云。 w 为 I_i 和 I_{t^*} 的结构相似值。根据公式 (22), 高频系数的具体融合规则表述为

(1) 如果是云遮挡区 ($M=1$), 则 I_{t^*} 的高频特征无法借鉴, 优先保持 I_i 的原始高频信息 $H_{F_i, ij}^{l, k}$ 。

(2) 如果是无云区 ($M=0$), 且匹配测度 $w_{AB}=0$, 说明所测像素点为孤立点, 异常点需剔除, 仍选择高频 $H_{F_i, ij}^{l, k}$ 。

(3) 如果是无云区 ($M=0$), 且匹配测度 $w_{AB}=8$, 说明该像素点为区域内点, I_{t^*} 的高频特征可以借鉴。然后比较 $H_{F_i, ij}^{l, k}$ 和 $H_{F_{t^*}, ij}^{l, k}$ 对应的模拟点火次数 $T_{F_i, ij}^{l, k}[N]$ 和 $T_{F_{t^*}, ij}^{l, k}[N]$, 判断高频信息丰富程度。若 $T_{F_i, ij}^{l, k}[N] \geq T_{F_{t^*}, ij}^{l, k}[N]$, 则选择 $T_{F_i, ij}^{l, k}[N]$ 。否则选择 $T_{F_{t^*}, ij}^{l, k}[N]$ 。

如果是无云区域 ($M=0$), 且 $1 \leq w_{AB} \leq 7$, 认为中心点为区域边缘, 边缘包含两个区域丰富的

高频细节, 需要进行加权组合强化 I_i 的原始高频, I_i 和 I_{t^*} 的结构相似性 SSIM 作为权值。

3.5 损失函数

本文模型生成器的损失函数主要包括 GAN 损失和图像内容损失。而判别器的损失函数与 LSGAN (Mao 等, 2017) 的相应部分一致。GAN 损失选择 LSGAN 的最小二乘损失项 $\mathcal{L}_{GAN}(G)$ 。图像内容损失由 L1 损失 $\mathcal{L}_I$ 、光谱损失 $\mathcal{L}_{\text{spectral}}$ 、结构损失 $\mathcal{L}_{\text{structure}}$ 、平均梯度损失 $\mathcal{L}_{\bar{G}}$ 和空间频率损失 $\mathcal{L}_{SF}$ 共同组成。表示为

$$\mathcal{L}_G = \alpha \mathcal{L}_{GAN}(G) + \beta \mathcal{L}_I + \gamma \mathcal{L}_{\text{spectral}} + \delta \mathcal{L}_{\text{structure}} + \varepsilon (\mathcal{L}_{\bar{G}} + \mathcal{L}_{SF}) \quad (23)$$

$$\mathcal{L}_D = \mathcal{L}_{GAN}(D) \quad (24)$$

式中, $\mathcal{L}_G$ 、 $\mathcal{L}_D$ 表示生成器损失函数和判别器损失函数。 $\mathcal{L}_{GAN}(G)$ 是 LSGAN 模型生成器的最小二乘损失项, 而 $\mathcal{L}_{GAN}(D)$ 为 LSGAN 判别器的损失函数, 如式 (25) 和 (26)。 α 、 β 、 γ 、 δ 、 ε 为权重系数。

$$\mathcal{L}_{GAN}(G) = \frac{1}{2} E_{F_i, C_i, C_t} [(D(G(F_i, C_i), C_t) - 1)^2] \quad (25)$$

$$\mathcal{L}_{GAN}(D) = \frac{1}{2} E_{F_i, C_i} [(D(F_i, C_i) - 1)^2] + \frac{1}{2} E_{F_i, C_i, C_t} [(D(G(F_i, C_i), C_t) - 1)^2] \quad (26)$$

$\mathcal{L}_I$ 用于捕捉预测图像和真实图像在同一位置的像素值差异, 以最小化像素值变化, 用均方误差 (MSE) 表示, 如式 (27)。 $\mathcal{L}_{\text{spectral}}$ 用于减少预测图像与真实图像之间的光谱失真, 通过光谱角 (SAM) 来判断光谱失真情况, 如式 (28)。 $\mathcal{L}_{\text{structure}}$ 用于保留图像结构细节, 通过多尺度结构相似度 (MS-SSIM) (Wang 等, 2003) 来计算, 如式 (29)。而 $\mathcal{L}_{\bar{G}}$ 和 $\mathcal{L}_{SF}$ 则是通过计算预测图像和真实图像之间的平均梯度和空间频率来减小图像的清晰度损失, 以取得更好的视觉效果, 如式 (30)–(31)。

$$\mathcal{L}_I = \frac{1}{N} \sum_{i=1}^N [X_i - Y_i]^2 \quad (27)$$

$$\mathcal{L}_{\text{spectral}} = I - \frac{X \times Y}{\|X\| \|Y\|} \quad (28)$$

$$\mathcal{L}_{\text{structure}} = I - [l_M(X, Y)]^{\alpha_u} \times \prod_{j=1}^M [c_j(X, Y)]^{\beta_j} [s_j(X, Y)]^{\gamma_j} \quad (29)$$

$$\mathcal{L}_{\bar{G}} = \frac{1}{N} \sum_{k=1}^N \|\bar{G}_{X_k} - \bar{G}_{Y_k}\|_1 \quad (30)$$

$$\mathcal{L}_{SF} = \frac{1}{N} \sum_{m=1}^N \|SF_{X_m} - SF_{Y_m}\|_1 \quad (31)$$

式中, N 为图像波段数。 X 和 Y 分别为真实图和预测图。 I 为1的张量。 M 表示最高尺度数。 l_j , c_j , s_j 分别第 j 个尺度下图像对的亮度、对比度和结构比较。 α_j , β_j , γ_j 为相关的权重。 $\bar{G}_{X_i}$, $\bar{G}_{Y_i}$ 表示 X 和 Y 的平均梯度。 SF_{X_i} , SF_{Y_i} 则分别为 X 和 Y 的空间频率。

4 实验

4.1 研究区域和数据集

本实验采用Coleambaly灌溉区(CIA)和Lower Gwydir Catchment (LGC) (Emelyanova等, 2013)以及阿科尔沁旗农牧区(AHB) (Li等, 2020a) 3个数据集来测试模型性能。

CIA位于澳大利亚新南威尔士州南部(34.0034°E, 145.0675°S), 属于现代化灌溉系统水稻种植区。其在作物生长季节周期内拥有较为丰富的物候变化。由17个2001年—2002年的Landsat-MODIS图像对组成, 光谱段数为6。图像的空间分辨率为25 m, 大小为2040×1720。

LGC区域位于澳大利亚新南威尔士州北部(149.2815°E, 29.0855°S), 由14对2004年4月至2005年4月的Landsat-MODIS图像对组成, 云覆盖均小于10%, 光谱带数为6。图像的空间分辨率为25 m, 大小为2720×3200。由于有洪水发生, 该区域具有显著的土地覆盖类型突变。

AHB区域位于中国内蒙古自治区阿科尔沁旗(106.3771°E, 38.9868°N), 属于中国东北部农牧区, 这里有大量的圆形牧场和农田。AHB数据集包含27对维度为6×2480×2800的无云Landsat-MODIS图像, 时间段为2013年5月30日到2018年12月6日。长达5年的时间跨度, 使AHB数据有着显著的物候变化。该区域光谱异质性较高。

4.2 实验设置

本文通过3部分实验验证STF-GAN_NSST的性能。首先, 将STF-GAN_NSST与现有的6种融合方法进行对比, 证明本文模型的有效性。对比算法包括STARFM、FSDAF、EDCSTFN、MLFF-GAN、GAN-STFM和SwinSTFM。其次, 对所提模型生成器的一些关键模块进行消融实验, 展示它们对模型的贡献。最后, 讨论了模型的运行效率。

为方便比较, 需要根据模型各自的输入要求

组织数据集。STARFM、FSDAF、EDCSTFN、MLFF-GAN、SwinSTFM和STF-GAN_NSST等需要至少一幅预测日期的粗分辨率图像和一对参考日期的粗—细图像对作为输入。而GAN-STFM模型由于参考不敏感性, 只需要预测时刻粗分辨率图像和任意时刻精细图像作为输入即可。除了GAN-STFM外, 其他模型都要求参考时刻与预测时刻应尽可能邻近。本实验3个数据集都分成训练集和测试集。对于CIA, 训练集日期为2001年10月至2002年2月, 共12个图像组。测试集为2002年3月至5月的3个图像组。对于LGC, 训练集为日期在2004年4月至12月之间的5个图像组, 测试集为2005年1月至4月的3个图像组。对于AHB, 除日期为2018年的6对数据用作测试集外, 其他数据全部划为训练集。测试集数据全程未参与训练。

模型训练时, 由于计算机内存大小有限, 输入图像会被裁剪成256×256大小的图像块后再送入模型。对于生成器的4个编码块和解码块输出通道数分别设置为16, 32, 64, 128, 128, 64, 32, 16。对于判别器, 输入是预测出的图像、预测时刻粗分辨率图像和真实的精细分辨率图像。输出为0—1的矩阵, 值越接近1, 图像越逼真。GAN-STFM、MLFF-GAN的初始学习率设置为 2×10^{-4} 。而EDCSTFN和SwinSTFM则设置为 1×10^{-3} 。每个epoch的批大小设置为16, 总训练200个epochs。损失函数式(23)中各损失项的权值设置为: $\alpha=0.01$, $\beta=\gamma=\delta=\varepsilon=1$ 。模型参数可根据情况进行调整。所有对比算法和本文方法都通过Python语言和PyTorch框架实现, 并使用GPU进行训练。具体见表1。

表1 实验软硬件环境

Table 1 Experiment with software and hardware environments

硬件		软件	
RAM	128G	Python	3.6
CPU	Intel Xeon Gold 6154 (2块)	CUDA	11.0
GPU	TITAN V 12G (2块)	PyTorch	1.7.1

4.3 实验结果与分析

4.3.1 评估指标

实验选择了6种典型的指标来评估融合性能, 分别为均方根误差(RMSE)、结构相似性指数

(SSIM) (Wang 等, 2004)、光谱角映射 (SAM)、平均绝对误差 (MAE)、相关系数 (CC) 和相对无量纲全局误差 (ERGAS) (Khan 等, 2009)。

其中, RMSE 和 MAE 分别用于衡量真实与预测图像之间的均方误差和平均绝对误差, 值越小, 图像质量越好。SSIM 和 CC 分别用于评估图像的结构相似性和整体相似度, 取值为 0 至 1, 值越大, 表明图像越相似。SAM 用于评估图像光谱失真度, 值越小, 光谱相似度越高。ERGAS 用来综合评估

融合图像质量, 值越小, 表明图像质量越好, 清晰度越高。

4.3.2 对比实验

将本文所提 STF-GAN_NSST 和 6 种对比算法在 CIA、LGC 和 AHB 数据集上展开比较, 并通过量化指标和主观视觉两方面对融合结果进行评估。为方便比较, 所有对比实验均采用单图像对作为参考, 见表 2、表 3 和表 4。GAN-STFM 对于参考日期的选取是灵活的, 不受规定的参考日期约束。

表 2 针对 CIA 数据集的量化评价结果
Table 2 Quantitative evaluation results for the CIA dataset

Prediction date (Reference date)	Model	MAE	RMSE	SAM	SSIM	CC	ERGAS
Apr 01, 2002 (Mar 16, 2002)	STARFM	0.0221	0.0347	0.2011	0.8332	0.8480	2.0130
	FSDAF	0.0198	0.0324	0.1874	0.8669	0.8657	1.8351
	EDCSTFN	0.0195	0.0293	0.1882	0.9023	0.9146	1.8862
	MLFF-GAN	0.0187	0.0288	<u>0.1654</u>	0.9445	<u>0.9555</u>	<u>1.6243</u>
	GAN-STFM	0.0250	0.0401	0.2044	0.9237	0.9289	1.7583
	SwinSTFM	<u>0.0176</u>	<u>0.0255</u>	0.1697	<u>0.9488</u>	0.9500	1.6600
	Ours	0.0168	0.0225	0.1556	0.9601	0.9699	1.4988
Apr 17, 2002 (Apr 10, 2002)	STARFM	0.0353	0.0422	0.1898	0.8504	0.8499	1.9560
	FSDAF	0.0296	0.0378	0.1977	0.9072	0.9176	1.6436
	EDCSTFN	0.0285	0.0366	0.2102	0.9415	0.9305	1.5869
	MLFF-GAN	0.0179	0.0245	<u>0.1659</u>	<u>0.9436</u>	<u>0.9499</u>	<u>1.2877</u>
	GAN-STFM	0.0324	0.0401	0.2117	0.9317	0.9387	1.4212
	SwinSTFM	0.0209	0.0269	0.1702	0.9381	0.9401	1.3546
	Ours	<u>0.0182</u>	<u>0.0257</u>	0.1546	0.9553	0.9613	1.2810
May 03, 2002 (Apr 26, 2002)	STARFM	0.0364	0.0524	0.2110	0.7861	0.8011	2.4211
	FSDAF	0.0321	0.0485	0.1844	0.8596	0.8657	2.2167
	EDCSTFN	0.0279	0.0427	0.1906	0.8625	0.8660	2.2020
	MLFF-GAN	0.0266	0.0396	<u>0.1523</u>	<u>0.8824</u>	0.9024	<u>1.8574</u>
	GAN-STFM	0.0275	0.0420	0.2239	0.8669	0.8713	1.9362
	SwinSTFM	<u>0.0264</u>	<u>0.0374</u>	0.1726	0.8433	0.8523	1.8664
	Ours	0.0258	0.0325	0.1510	0.9160	<u>0.9202</u>	1.8559

注: 粗体表示最优精度, 下划线表示次优精度。

表 3 针对 LGC 数据集的量化评估结果
Table 3 Quantitative evaluation results for the LGC dataset

Prediction date (Reference date)	Model	MAE	RMSE	SAM	SSIM	CC	ERGAS
Jan 29, 2005 (Jan 13, 2005)	STARFM	0.0376	0.0448	0.6377	0.8849	0.8911	1.3033
	FSDAF	0.0351	0.0403	0.4309	0.8954	0.9019	1.2987
	EDCSTFN	0.0204	0.0257	0.4296	0.9013	0.9135	1.2860
	MLFF-GAN	0.0168	0.0243	0.3976	0.9523	0.9441	1.0745
	GAN-STFM	0.0192	0.0265	0.3618	0.9344	0.9527	1.0874
	SwinSTFM	<u>0.0165</u>	<u>0.0176</u>	<u>0.3445</u>	<u>0.9619</u>	<u>0.9597</u>	<u>0.8429</u>
	Ours	0.0142	0.0168	0.2811	0.9679	0.9658	0.7318

							续表
Prediction date (Reference date)	Model	MAE	RMSE	SAM	SSIM	CC	ERGAS
Mar 02, 2005 (Feb 14, 2005)	STARFM	0.0223	0.0311	0.4028	0.9215	0.9189	3.7110
	FSDAF	0.0159	0.0225	0.3792	0.9387	0.9401	2.4381
	EDCSTFN	0.0160	0.0258	<u>0.2989</u>	0.9335	0.9469	1.3126
	MLFF-GAN	<u>0.0116</u>	0.0175	0.3212	<u>0.9579</u>	<u>0.9658</u>	0.8982
	GAN-STFM	0.0140	0.0210	0.3458	0.9216	0.9347	1.1456
	SwinSTFM	0.0100	0.0109	0.3125	0.9575	0.9622	<u>0.7351</u>
	Ours	0.0124	<u>0.0135</u>	0.2917	0.9622	0.9751	0.5116
Apr 03, 2005 (Mar 02, 2005)	STARFM	0.0184	0.0251	0.3967	0.9237	0.9288	2.3574
	FSDAF	0.0197	0.0304	0.4336	0.9197	0.9206	2.6861
	EDCSTFN	0.0179	0.0233	0.5671	0.8975	0.9144	1.2475
	MLFF-GAN	0.0159	0.0197	<u>0.2866</u>	0.9316	0.9471	1.0985
	GAN-STFM	0.0168	0.0228	0.3911	0.9298	0.9366	1.3143
	SwinSTFM	<u>0.0124</u>	<u>0.0195</u>	0.3725	<u>0.9513</u>	<u>0.9490</u>	<u>0.8732</u>
	Ours	0.0105	0.0116	0.1906	0.9677	0.9712	0.7434

注：粗体表示最优精度，下划线表示次优精度。

表 4 针对 AHB 数据集的量化评估结果

Table 4 Quantitative evaluation results for the AHB dataset

Prediction date (Reference date)	Model	MAE	RMSE	SAM	SSIM	CC	ERGAS
Feb 05, 2018 (Jan 04, 2018)	STARFM	0.3002	0.3521	0.4730	0.7266	0.6537	5.6971
	FSDAF	0.1958	0.2222	0.4818	0.7515	0.7756	5.3931
	EDCSTFN	0.1989	0.2547	0.6002	0.6834	0.5719	4.7692
	MLFF-GAN	<u>0.1297</u>	<u>0.1688</u>	<u>0.3971</u>	<u>0.8784</u>	<u>0.8616</u>	<u>3.6995</u>
	GAN-STFM	0.1534	0.1766	0.4679	0.8357	0.7548	3.9816
	SwinSTFM	0.1652	0.1985	0.4121	0.7966	0.8006	4.2767
	Ours	0.0975	0.1133	0.2501	0.8848	0.9145	2.7626
May 12, 2018 (Mar 25, 2018)	STARFM	0.2976	0.3257	0.5068	0.7641	0.8024	4.5460
	FSDAF	0.1667	0.1879	0.4871	0.7828	<u>0.8545</u>	4.6129
	EDCSTFN	0.1478	0.1546	0.5311	0.6467	0.7117	3.2276
	MLFF-GAN	0.0876	0.0921	<u>0.2810</u>	0.8727	0.6851	3.1604
	GAN-STFM	0.1034	0.1306	0.4025	0.8566	0.7954	2.6532
	SwinSTFM	<u>0.0745</u>	<u>0.0817</u>	0.3856	0.8053	<u>0.8257</u>	<u>1.9439</u>
	Ours	0.0418	0.0594	0.2433	<u>0.8579</u>	0.8653	1.5796
Dec 06, 2018 (Oct 19, 2018)	STARFM	0.2553	0.2865	0.5440	0.8170	0.7756	4.2014
	FSDAF	0.1846	0.2047	0.5090	0.8287	0.8340	4.1456
	EDCSTFN	0.1475	0.1537	0.6011	0.7829	0.6536	7.2095
	MLFF-GAN	0.1233	0.1469	<u>0.3940</u>	<u>0.8731</u>	0.7954	<u>2.9578</u>
	GAN-STFM	0.0877	0.1064	0.4762	0.8322	0.8437	4.7692
	SwinSTFM	<u>0.0854</u>	<u>0.0936</u>	0.4165	0.8459	<u>0.8566</u>	3.5324
	Ours	0.0679	0.0721	0.2678	0.8945	0.9025	2.0469

注：粗体表示最优精度，下划线表示次优精度。

表 2、表 3 和表 4 分别展示了各方法在 CIA、LGC 和 AHB 数据集上的量化评估。表中加粗部分为最优结果，下划线部分为次优结果。同时在 3 个数据集的融合结果中，选取 2002 年 4 月 17 日、2005 年 4 月 3 日和 2018 年 2 月 5 日的结果进行主观视觉展示，以 4，3，2 波段组成的彩色图像显示，

见图8、图9和图10。图中最顶部一行为融合图像的全局效果，中间行表示全局效果中黄框区域的

局部放大，而底部行则为融合图像与真实图像之间的误差情况，被拉伸在0.0—0.1之间显示。

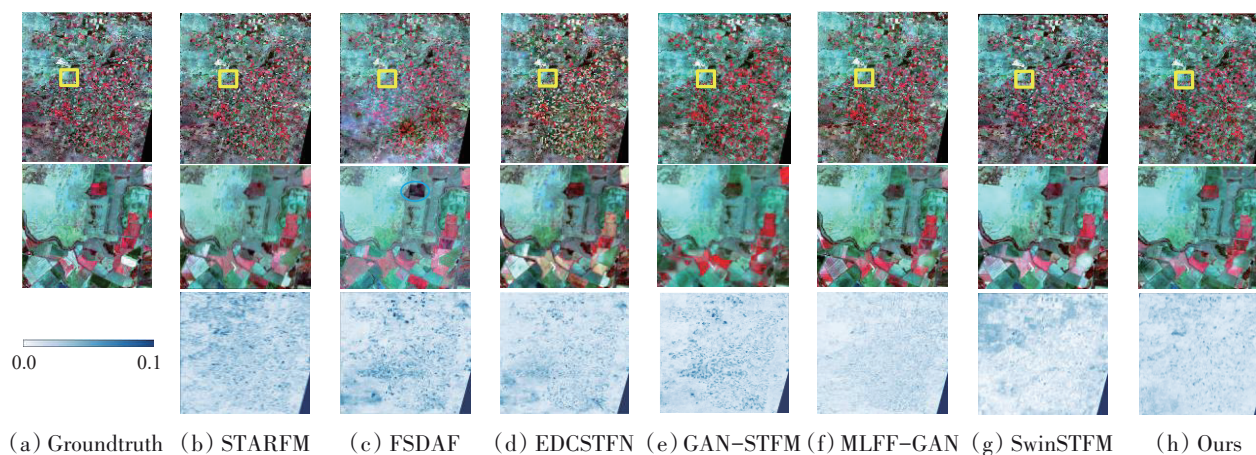


图8 2002年4月17日CIA数据的融合结果

Fig. 8 Fusion results of the CIA data on April 17, 2002

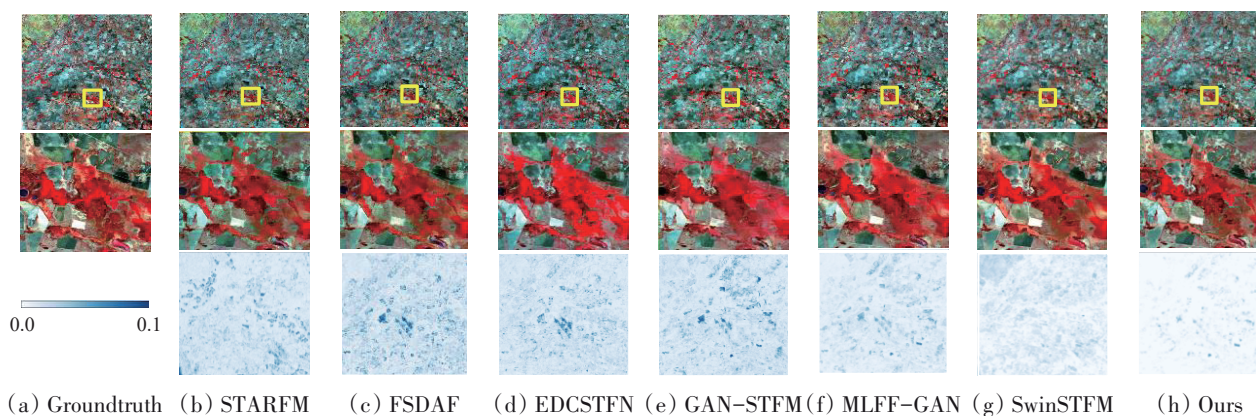


图9 2005年4月3日LGC数据的融合结果

Fig. 9 Fusion results of the LGC data on April 3, 2005

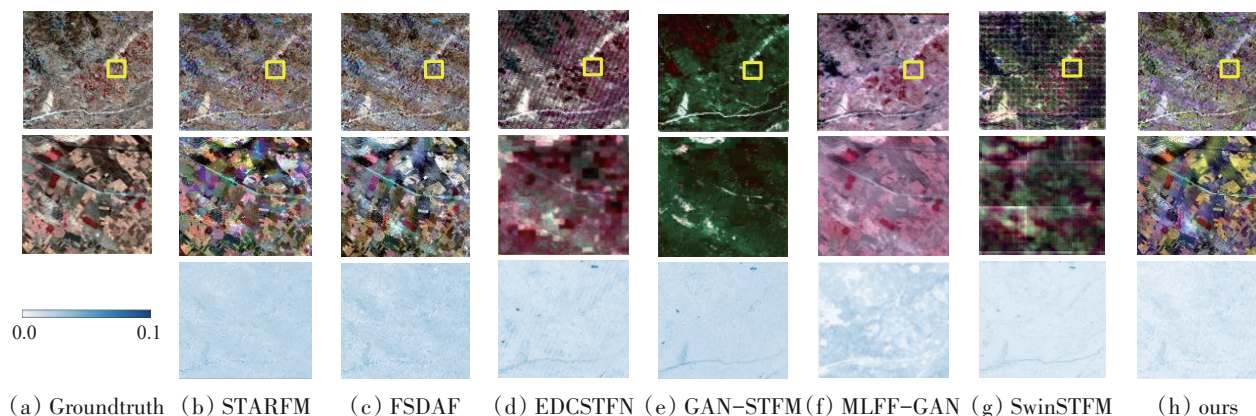


图10 2018年2月5日AHB数据的融合结果

Fig. 10 Fusion results of the AHB data on February 5, 2018

从表2可知，深度学习方法在CIA数据集上的表现总体上优于传统方法，本文所提方法在多数评价指标上优势显著。在2002年4月1日的预测

中，本文所提方法在所有评估指标上都最优。与次优的方法MLFF-GAN相比，CC值提高了1.4%，ERGAS值减少了0.13。同时MAE和RMSE等误差

指标领先于次优的SwinSTFM。本文所提方法的融合结果与真实图像的相关性更高,误差更小。在2002年4月17日的预测中,本文所提方法除MAE和RMSE等值稍大外,其他评估指标都比次优的MLFF-GAN更优。SSIM和CC值提高了1.2%,ERGAS值也略有减少,本文所提方法的融合结果纹理细节更丰富,清晰度更好。而误差指标稍大可能是偶然性导致。因为2002年5月3日的预测中,误差指标又比较低,比次优的SwinSTFM表现更好。图8展示了4月17日的融合结果。从全局看,所有融合结果都接近于真实图(Ground Truth),只有FSDAF受云层影响质量较差。但局部看有所不同。局部效果显示STARFM和FSDAF的结果中存在严重噪声,SwinSTFM和MLFF-GAN在光谱和空间细节上的表现比其他对比方法好,但仍面临纹理细节差,边缘模糊问题。而在本文所提方法的融合结果中该问题得到了有效解决。误差图显示本文所提方法是除MLFF-GAN外误差最小的。

从表3可见,对于LGC数据集,本文所提方法在大多数时候仍占优势。2005年1月29日的预测中,本文所提方法的评估指标最优。与次优的SwinSTFM相比,MAE和RMSE都有下降,误差更小。SAM值比SwinSTFM降低了0.0634,ERGAS值减少了0.11,在光谱一致性和图像清晰度方面有很大提升。2005年3月2日的预测中,除MAE和RMSE等指标外,本文所提方法在其他指标上有最好的性能。特别地,SSIM值比次优的MLFF-GAN提高了0.5%,ERGAS指标比次优的SwinSTFM减少了0.2,显示了本文所提方法在增强图像纹理细节和提高图像清晰度方面的优势。对于2005年4月3日,我们所提方法的SAM值比次优的MLFF-GAN低了0.0960,SSIM值和CC值分别提升了1.6%和2.2%,ERGAS减少了0.13。无论是光谱表现还是纹理细节方面都表现最好。图9展示了该日期的视觉效果。可见,噪声的存在使STARFM方法误差最高。从局部看,本文所提方法在空间细节表达和光谱方面与真实图最接近,细节最清晰。而其他深度学习方法中边缘细节都比较模糊。特别是EDCSTFN,过拟合严重。

从表4可见,对于AHB数据集,所有方法指标都比较低。因为AHB地物类型复杂,光谱变化大,所以预测难度增加。但本文所提方法仍有很大优势。在2018年2月5日的预测中,所有指标都超过了次优的MLFF-GAN,其中SAM值降低了

0.1470,CC值提升了5.29%,ERGAS值减少了0.9365。在2018年5月12日的预测中,除SSIM外,其他评估指标都最高。而2018年12月6日的预测中,所有指标都超过了次优方法,CC值和ERGAS值的表现更好。图10展示的2018年2月5日的视觉效果可见,各方法直观效果都比较差,光谱扭曲严重。EDCSTFN和SwinSTFM的整体效果甚至差于传统方法,显示出极差的泛化性。但GAN-STFM和MLFF-GAN优于传统方法,特别是在结构相似性方面。MLFF-GAN对于真实地物的逼近是准确和清晰的,但GAN-STFM的结果在光谱表现上比较差。本文所提方法在光谱表现上超过了GAN-STFM。本文所提方法融合结果的局部细节更清晰,噪声干扰更少。误差图也显示本文所提方法误差最小。无论是定量指标还是视觉效果,本文所提方法都呈现出了很好的表现。

4.3.3 消融实验

本文模型引入了膨胀卷积分支、SUB-AM块和NSST-ARM块,在兼顾特征分布的全局关系和局部变化的同时增强了融合图像的高频细节,并通过设计清晰度损失项来减少模型训练时的清晰度损失。因此,以含有普通卷积分支的GAN为基础,对以上组件展开消融,以了解它们对于模型的贡献。同时将损失函数中清晰度损失项去掉,并与之前进行对比,以验证清晰度损失项的有效性。表5展示了各模型在2005年4月3日LGC数据上的性能表现。可见,本文所提的STF-GAN_NSST几乎在所有指标上都表现最佳。而且随着组件的不断添加,各项指标表现越来越好,说明上述组件的添加对模型性能提高有所帮助。同时损失函数的清晰度损失项去掉以后,模型的性能有所下降,说明清晰度损失项在模型训练中有一定的正向作用。图11给出了各种消融结果的直观比较。由于清晰度损失项的优势只体现在了量化指标上,视觉效果与有清晰度损失项时的融合结果区别不大。所以图11仅展示了各模型在含清晰度损失项下训练时的结果。顶部一行为全局效果,中间行表示全局效果中黄框所示区域的放大,最后一行是融合结果与真实图像之间的像素差值图。可以看出,STF-GAN_NSST效果最好,局部细节和光谱表现最优。表5中的误差指标也说明了这一点。第三行的误差图显示,STF-GAN_NSST融合结果误差也是最小。这也证明,本文所提方法各组件在融合过程中发挥着重要作用。

表5 针对LGC数据集的消融实验量化评估结果
Table 5 Quantitative evaluation results of ablation experiments for the LGC dataset

Model	MAE	RMSE	SAM	SSIM	CC	ERGAS
GAN+Conv	0.0168	0.0228	0.3911	0.8310	0.9037	3.3345
GAN+Conv+DConv	0.0154	0.0200	0.4100	0.9023	0.9134	2.4365
GAN+Conv+DConv+SUB-AM	0.0129	0.0176	0.2340	0.9422	0.9510	1.6453
STF-GAN_NSST/no clarity loss	<u>0.0111</u>	<u>0.0124</u>	<u>0.1979</u>	<u>0.9654</u>	<u>0.9700</u>	<u>0.8049</u>
STF-GAN_NSST/clarity loss	0.0105	0.0116	0.1906	0.9677	0.9712	0.7434

注：粗体表示最优精度，下划线表示次优精度。

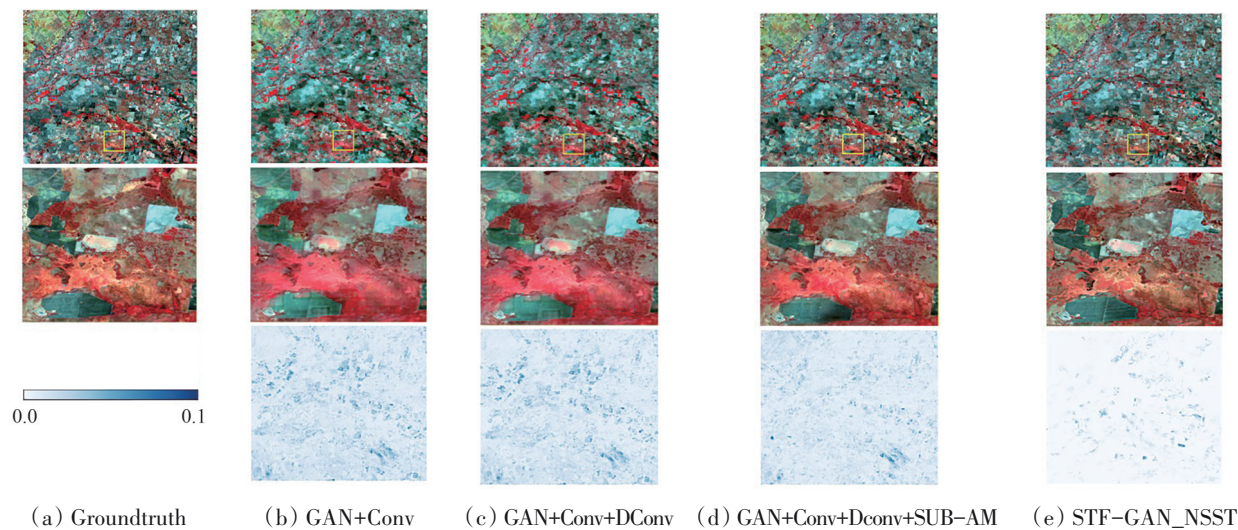


图11 2005年4月3日LGC数据的消融实验结果

Fig. 11 The ablation experiment results of LGC data on April 3, 2005

4.3.4 运算效率分析

为评估运行效率，本文分别计算了各深度学习方法的参数量、运算量以及时间成本，如表6所示。其中，计算量用MACs表示，MACs为计算一个256×256×6的图像所需的运算次数，时间成本则通过每个批次的运行时间来表现。表中，SwinSTFM参数最多，这主要是所采用的Transformer结构复杂所致。但SwinSTFM的MACs值比较低，故计算复杂度不是很高，可以看出每个批次运行时间仅次于MLFF-GAN。基于GAN的算法MACs值都比较大，因为模型对抗训练需消耗大量计算资源。MLFF-GAN受多级网络影响，比GAN-STFM和本文所提模型参数更多，但计算效率更高，所花时间成本是所有方法中最低的。STF-GAN_NSST中因为有膨胀卷积分支、SUB-AM块和NSST-AM块，导致MACs值很大，运算负载高，NSST分解和图像重构需消耗大量时间，所以本文方法的时间成本是最高的。但与SwinSTFM和MLFF-GAN相比，

参数更少，生成器的参数量只有2.6×10⁶。所以STF-GAN_NSST模型更小，有一定优势。

表6 各模型的参数量与运算复杂度分析
Table 6 Analysis of the number of parameters and computational complexity of each model

模型	参数量	运算复杂度(MACs)	时间/批次(s)
EDCSTFN	2.8×10 ⁵	1.8×10 ¹⁰	5.45
GAN-STFM	5.8×10 ⁵ (Generator)	3.8×10 ¹⁰	17.01
	3.7×10 ⁶ (Discriminator)	7.7×10 ⁶	
MLFF-GAN	5.9×10 ⁶ (Generator)	1.4×10 ¹⁰	1.08
	2.8×10 ⁶ (Discriminator)	3.8×10 ⁹	
SwinSTFM	4.0×10 ⁷	2.8×10 ¹⁰	2.51
Ours	2.6×10 ⁶ (Generator)	8.6×10 ¹⁰	20.33
	3.7×10 ⁶ (Discriminator)	7.7×10 ⁶	

5 结 论

本文所提的STF-GAN_NSST方法利用深度学习和NSST变换结合的思路来进行遥感图像时空融合。首先在编码器中通过普通卷积和膨胀卷积两个分支分别提取不同视野的精细特征并结合, 兼顾了局部细节和全局关系。SUB-AM块输出的粗特征变化对精细特征进行调整, 提高了初步融合结果的预测精度。同时NSST-ARM块将编码器—解码器生成的图像和参考时刻精细图像分解为多尺度高低频系数, 并按照自适应融合规则重构, 增强了融合图像的高频细节。该方法得到的融合图像空间特征丰富, 边缘结构清晰。在3个数据集的对比实验和LGC数据集的消融实验证明了它的优势和各模块对于模型的贡献。本文所提方法在遥感图像时空融合方面展现出了广阔的应用前景。

同时也认识到, 时空融合技术发展愈发成熟, 继续提高融合性能难度很大。对于地物覆盖突变情况, 现有模型的预测效果仍不理想。另外, 深度学习融合模型参数量大, 时间复杂度高, 实时性差也是亟需解决的问题。所以在接下来的工作中, 进一步优化模型结构提高模型鲁棒性和实时性, 减少参数量仍是我们需要继续努力的方向。

志 谢 此次实验所需的Landsat-Modis数据来源于NASA官方的空间数据网站, 在此表示衷心的感谢!

参考文献(References)

- Ajit A, Acharya K and Samanta A. 2020. A review of convolutional neural networks//2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). Vellore: IEEE: 1-5 [DOI: 10.1109/ic-ETITE47903.2020.049]
- Babu Avvaru H S, Dasarapalli S R and Kollu P K. 2023. Dehazing of multispectral remote sensing images using CNN with ResNet//2023 3rd Asian Conference on Innovation in Technology (ASIAN-CON). Ravet: IEEE: 1-6 [DOI: 10.1109/ASIANCON58793.2023.10269956]
- Blasch E P. 1999. Biological information fusion using a PCN-N and belief filtering//Proceedings of the International Joint Conference on Neural Networks. Washington: IEEE: 2792-2795 [DOI: 10.1109/IJCNN.1999.833523]
- Chen G Y, Jiao P, Hu Q, Xiao L J and Ye Z J. 2022. SwinSTFM: remote sensing spatiotemporal fusion using swin transformer. IEEE Transactions on Geoscience and Remote Sensing, 60: 5410618 [DOI: 10.1109/TGRS.2022.3182809]
- Dhillon M S, Dahms T, Kübert-Flock C, Steffan-Dewenter I, Zhang J and Ullmann T. 2022. Spatiotemporal fusion modelling using STARFM: examples of Landsat 8 and sentinel-2 NDVI in Bavaria. Remote Sensing, 14(3): 677 [DOI: 10.3390/rs14030677]
- Easley G, Labate D and Lim W Q. 2008. Sparse directional image representations using the discrete shearlet transform. Applied and Computational Harmonic Analysis, 25(1): 25-46 [DOI: 10.1016/j.acha.2007.09.003]
- Emelyanova I V, McVicar T R, Van Niel T G, Li L T and Van Dijk A I J M. 2013. Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: a framework for algorithm selection. Remote Sensing of Environment, 133: 193-209 [DOI: 10.1016/j.rse.2013.02.007]
- Francini S, D'Amico G, Vangi E, Borghi C and Chirici G. 2022. Integrating GEDI and Landsat: Spaceborne Lidar and four decades of optical imagery for the analysis of forest disturbances and biomass changes in Italy. Sensors, 22(5): 2015 [DOI: 10.3390/s22052015]
- Gao F, Masek J, Schwaller M and Hall F. 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE Transactions on Geoscience and Remote Sensing, 44(8): 2207-2218 [DOI: 10.1109/TGRS.2006.872081]
- Gao W, Liu X G, Peng P and Chen Q H. 2010. An improved method of high-resolution remote sense image segmentation. Earth Science, 35(3): 421-425 [DOI: 10.3799/dqkx.2010.050]
- Gevaert C M and García-Haro F J. 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sensing of Environment, 156: 34-44 [DOI: 10.1016/j.rse.2014.09.012]
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2020. Generative adversarial networks. Communications of the ACM, 63(11): 139-144 [DOI: https://doi.org/10.1145/3422622]
- Huang B and Song H H. 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 50(10): 3707-3716 [DOI: 10.1109/TGRS.2012.2186638]
- Huang B, Zhang H K, Song H H, Wang J and Song C Q. 2013. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial-temporal-spectral earth observations. Remote Sensing Letters, 4(6): 561-569 [DOI: 10.1080/2150704X.2013.769283]
- Husman S D R, Hu Z Y, Wouters B, Munneke P K, Veldhuijsen S and Lhermitte S. 2023. Remote sensing of surface melt on Antarctica: opportunities and challenges. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16: 2462-2480 [DOI: 10.1109/JSTARS.2022.3216953]
- Jafarbiglu H and Pourreza A. 2022. A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. Computers and Electronics in Agriculture, 197: 106844 [DOI: 10.1016/j.

- compag.2022.106844]
- Jiao J and Wu L D. 2019. Fusion of multispectral and panchromatic images via morphological filter and improved PCNN in NSST domain. *Journal of Image and Graphics*, 24(3): 435-446 (焦姣, 吴玲达. 2019. 形态学滤波和改进PCNN的NSST域多光谱与全色图像融合. *中国图象图形学报*, 24(3): 435-446) [DOI: 10.11834/jig.180399]
- Jing W P, Lou T T, Wang Z Y, Zou W T, Xu Z K, Mohaisen L, Li C and Wang J. 2023. A rigorously-incremental spatiotemporal data fusion method for fusing remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 6723-6738 [DOI: 10.1109/JSTARS.2023.3296468]
- Khan M M, Alparone L and Chanussot J. 2009. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11): 3880-3891 [DOI: 10.1109/TGRS.2009.2029094]
- Li A H, Bo Y C, Zhu Y X, Guo P, Bi J and He Y Q. 2013. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. *Remote Sensing of Environment*, 135: 52-63 [DOI: 10.1016/j.rse.2013.03.021]
- Li J, Li Y F, He L, Chen J and Plaza A. 2020a. Spatio-temporal fusion for remote sensing data: an overview and new benchmark. *SCIENCE CHINA Information Sciences*, 63(4): 140301 [DOI: 10.1007/s11432-019-2785-y]
- Li W S, Zhang X Y, Peng Y D and Dong M L. 2020b. DMNet: A network architecture using dilated convolution and Multiscale mechanisms for spatiotemporal fusion of remote sensing images. *IEEE Sensors Journal*, 20(20): 12190-12202 [DOI: 10.1109/JSEN.2020.3000249]
- Li Y F, Li J and Zhang S Q. 2021. A extremely fast Spatio-temporal fusion method for remotely sensed images//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels: IEEE: 4452-4455 [DOI: 10.1109/IGARSS47720.2021.9554794]
- Luo P, Ren J M, Peng Z L, Zhang R M and Li J Y. 2018. Differentiable learning-to-normalize via switchable normalization//7th International Conference on Learning Representations. New Orleans: OpenReview.net
- Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z and Smolley S P. 2017. Least squares generative adversarial networks//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE: 2813-2821 [DOI: 10.1109/ICCV.2017.304]
- Ping B, Meng Y S and Su F Z. 2018. An enhanced linear spatio-temporal fusion method for blending Landsat and MODIS data to synthesize Landsat-like imagery. *Remote Sensing*, 10(6): 881 [DOI: 10.3390/rs10060881]
- Song B Z, Liu P, Li J, Wang L Z, Zhang L, He G J, Chen L J and Liu J B. 2022. MLFF-GAN: A multilevel feature fusion with GAN for spatiotemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4410816 [DOI: 10.1109/TGRS.2022.3169916]
- Song H H, Liu Q S, Wang G J, Hang R L and Huang B. 2018. Spatio-temporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3): 821-829 [DOI: 10.1109/JSTARS.2018.2797894]
- Tan Z Y, Di L P, Zhang M D, Guo L Y and Gao M L. 2019. An enhanced deep convolutional model for spatiotemporal image fusion. *Remote Sensing*, 11(24): 2898 [DOI: 10.3390/rs11242898]
- Tan Z Y, Gao M L, Li X H and Jiang L C. 2022. A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5601413 [DOI: 10.1109/TGRS.2021.3050551]
- Tan Z Y, Yue P, Di L P and Tang J M. 2018. Deriving high spatiotemporal remote sensing images using deep convolutional network. *Remote Sensing*, 10(7): 1066 [DOI: 10.3390/rs10071066]
- Wang Z, Bovik A C, Sheikh H R and Simocelli 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]
- Wang Z, Simoncelli E P and Bovik A C. 2003. Multiscale structural similarity for image quality assessment//The 37th Asilomar Conference on Signals, Systems & Computers. Pacific Grove: IEEE: 1398-1402 [DOI: 10.1109/ACSSC.2003.1292216]
- Wu J G, Cheng Q, Li H F, Li S, Guan X B and Shen H F. 2020. Spatio-temporal fusion with only two remote sensing images as input. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 6206-6219 [DOI: 10.1109/JSTARS.2020.3028116]
- Wu M Q, Niu C, Wang C Y, Wu C Y and Wang L. 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing*, 6(1): 063507 [DOI: 10.1117/1.JRS.6.063507]
- Yan H F, Zhao Y Q, Chan J C W, Kong S G, Ei-Bendary N and Reda M. 2025. Hyperspectral and multispectral image fusion: When model-driven meet data-driven strategies. *Information Fusion*, 116: 102803 [DOI: 10.1016/j.inffus.2024.102803]
- Yang J X, Xiao L, Zhao Y Q and Chan J C W. 2024. Unsupervised deep tensor network for hyperspectral-multispectral image fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9): 13017-13031 [DOI: 10.1109/TNNLS.2023.3266038]
- Zhang H, Sun Y, Shi W Z, Guo D Z and Zheng N S. 2021. An object-based spatiotemporal fusion model for remote sensing images. *European Journal of Remote Sensing*, 54(1): 86-101 [DOI: 10.1080/22797254.2021.1879683]
- Zhang Y and Hong G. 2005. An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. *Information Fusion*, 6(3): 225-234 [DOI: 10.1016/j.inffus.2004.06.009]
- Zhu X L, Chen J, Gao F, Chen X H and Masek J G. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11): 2610-2623 [DOI: 10.1016/j.rse.2010.05.032]

Zhu X L, Helmer E H, Gao F, Liu D S, Chen J and Lefsky M A. 2016.
A flexible spatiotemporal method for fusing satellite images with

different resolutions. *Remote Sensing of Environment*, 172: 165-177 [DOI: 10.1016/j.rse.2015.11.016]

Spatiotemporal fusion algorithm for remote sensing images based on the combination of GAN and NSST

WANG Yifei¹, BAO Wenxing¹, QU Kewen¹, FENG Wei^{1,2}, WANG Wenlong^{3,4}

1. North Minzu University, School of Computer Science and Engineering, Yinchuan 750021, China;

2. Xidian University, School of Electronic Engineering, Xi'an 710071, China;

3. Ningxia Remote Sensing Survey Institute, Yinchuan 750021, China;

4. Ningxia Data and Application Center of High Resolution Earth Observation System, Yinchuan 750021, China

Abstract: Currently, most deep learning-based methods for the spatiotemporal fusion of remote sensing images employ convolutional operations to extract image features, which are then combined to produce higher-accuracy fused images. However, because of the convolution operation balancing pixel errors within local regions during the process of extracting image features, the high-precision fused image obtained after integrating various extracted features still suffers from issues such as weak texture details and blurred region boundaries.

To address such issues, in this paper, we propose a new spatiotemporal fusion method based on the combination of Generative Adversarial Network (GAN) and nonsubsampling Shearlet transform (NSST). This method is based on the GAN framework and consists of a generator and a discriminator. Specifically, a dilated convolution branch and a subtraction-attention module (SUB-AM) are first introduced into the encoder structure of the generator, so that the extracted fine features can take into account the global nature; the coarse feature changes captured by the SUB-AM block are used to adjust the extracted fine features to improve the prediction accuracy. Then, through NSST and an adaptive reconstruction module (NSST-ARM), the high- and low-frequency information of the results generated by the encoder-decoder and the fine image of the reference moment are reconstructed in accordance with adaptive fusion rules, enhancing the high-frequency details of the fused image. Finally, a clarity loss term is introduced into the loss function to reduce the clarity loss of the fused image. The clarity loss term consists of the average gradient loss and the spatial frequency loss.

Experimental results show that compared with a state-of-the-art method, the proposed method exhibits increased structural similarity value by 1.5% on average, reduced relative dimensionless global error in synthesis value by 14% on average, and improved texture detail expression and edge structure clarity of the fused image. Experiments prove the effectiveness of the proposed method.

In this paper, we propose a spatiotemporal fusion method for remote sensing images based on the combination of GAN and NSST. The method has the following advantages: (1) Dilated convolution and SUB-AM are introduced into the encoder to make the features of the fused image accurate and the image information comprehensive. (2) The introduction of NSST-ARM and the optimization of the loss function make the edges and contours of the fused image clear. Experimental results on three datasets show that the method proposed in this paper can achieve high fusion accuracy and obtain clear image details. Nonetheless, the applicability of the proposed method remains limited; that is, it cannot be applied to all types of land. Moreover, there is still considerable room for improvement in its efficiency. Given these limitations, how to make our method more generalizable and efficient is a topic worth exploring in future research.

Key words: spatiotemporal fusion, Generative Adversarial Network (GAN), dilated convolution, Nonsubsampling Shearlet Transform (NSST)

Supported by National Natural Science Foundation of China (No. 62461001); Natural Science Foundation of Ningxia Province of China (No. 2024AAC02035); Innovation Team of the National Ethnic Affairs Commission; Ningxia Key Laboratory of Visual Cognition and Computation