

文章编号: 2095-4980(2023)07-0939-13

基于多级特征补偿的遥感图像时空融合方法

刘文杰^{a,b}, 李雨珈^{a,b}, 白梦浩^{a,b}, 张莉萍^{a,b}, 雷大江^{*a,b}

(重庆邮电大学 a. 计算机科学与技术学院; b. 重庆市图像认知重点实验室, 重庆 400065)

摘要: 在许多地球科学应用中要用到大量的高时空分辨率的地球观测数据。时空图像融合方法为产生高时空分辨率的数据提供了一种可行且经济的解决方案。然而, 现有的一些基于学习的方法对于图像深层特征提取能力较弱, 对于高分辨力图像细节特征利用度不够。针对这些问题, 提出一种基于多级特征补偿的遥感图像时空融合方法。该方法使用 2 个分支进行多层级的特征补偿, 并提出了融合通道注意力机制的残差模块作为网络的基本组成单元, 可以将高分辨力输入图像的深层特征更为详尽地提取利用。提出一种基于拉普拉斯算子的边缘损失, 在节省预训练计算开销的同时取得了很好的融合效果。使用从山东和广东 2 个地区采集的 Landsat 和中分辨率成像光谱仪(MODIS)卫星图像对所提出的方法进行实验评估。实验结果表明, 提出的方法在视觉外观和客观指标方面都具有更高质量。

关键词: 时空融合; 注意力机制; 边缘损失; 特征补偿

中图分类号: TP751

文献标志码: A

doi: 10.11805/TKYDA2022191

Spatiotemporal fusion of remote sensing images based on multi-level feature compensation

LIU Wenjie^{a,b}, LI Yujia^{a,b}, BAI Menghao^{a,b}, ZHANG Liping^{a,b}, LEI Dajiang^{*a,b}

(a.School of Computer; b.Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: A large amount of earth observation data with the high spatial and temporal resolution is employed in many earth science applications. The spatiotemporal image fusion method provides a feasible and economical solution for generating high spatiotemporal resolution data. However, some of the existing learning-based methods are poor in extracting deep image features and utilizing the detail features of high-resolution image. A spatiotemporal fusion method is proposed for remote sensing images based on multi-level feature compensation. It uses two branches to perform multi-level feature compensation and proposes a residual module fused with a channel attention mechanism as the basic unit of the network, which can extract and utilize the deep features of high-resolution input images in more detail. An edge loss is proposed based on the Laplacian operator, which saves the computational cost of pre-training and achieves a good fusion effect. The proposed method is experimentally evaluated by using Landsat and Moderate-resolution Imaging Spectroradiometer(MODIS) satellite images collected from two regions in Shandong and Guangdong. Experimental results show that the proposed method bears higher quality in both visual appearance and objective metrics.

Keywords: spatiotemporal image fusion; attention mechanism; edge loss; feature compensation

由于传感器硬件条件等的限制, 在密集时间序列中连续获取高分辨力的卫星图像比较困难。对于许多的地

收稿日期: 2022-09-30; 修回日期: 2022-11-10

基金项目: 国家自然科学基金资助项目(61972060; 62027827; U1713213; 61902046); 国家重点研发计划资助项目(2019YFE0110800); 重庆市自然科学基金资助项目(cstc2020jcyj-zdxmX0025; cstc2019cxcyljrc-td0270); 重庆市留学人员回国创业创新支持计划资助项目(cx2018120); 重庆市高新技术研究计划资助项目(cstc2018jcyjAX0279)

*通信作者: 雷大江 email:leidj@cqupt.edu.cn

表动力学研究, 获取同时具有高空间和高时间分辨力的遥感卫星图像至关重要^[1], 如作物和森林监测^[2], 以及土地利用和土地覆盖变化检测^[3]。这些应用场景中都需要在密集的时间序列下捕捉地面的变化, 还需要通过高清晰度的图像获取地表细节和地貌特征。尽管近年来传感器技术的进步极大地提高了卫星观测的精确度, 但由于预算或其他方面条件的限制, 同时保证高时间分辨力和高空间分辨力比较困难^[4-5]。目前的研究重点放在数据后处理方面^[6-8]。遥感图像时空融合利用 2 个或多个不同的图像数据源合成具有高时空分辨力的图像^[9], 该文采用具有高时间分辨力但低空间分辨力的粗略空间分辨力图像(High Time Low Spatial, HTLS)和具有低时间分辨力但高空间分辨力的精细空间分辨力图像(Low Time High Spatial, LTHS)。其中粗略分辨力的图像采用 MODIS 卫星传感器每天采集的图像^[10]。而精细分辨力图像采用 Landsat 图像^[11], 它的大多数光谱波段的空间分辨力为 30 m, 但它的时间分辨力约为 16 天。

现有的时空融合算法主要可以分为 4 类: 基于变换的方法; 基于重建的方法; 基于贝叶斯的方法; 基于学习的方法^[12-13]。基于变换的方法采用的是数学变换方法, 如小波变换等, 通过在变换空间中整合多源信息, 获得融合后的高分辨力数据。这种方法不仅用于时空融合^[14-15], 还广泛用于全色和多光谱的融合^[16-19]。基于重建的方法是通过将已知的高空间低时间分辨力的图像和高时间低空间分辨力的图像进行权重赋值并组合, 预测某时刻的高空间分辨力图像。典型的模型有时空自适应融合模型(Spatiotemporal Adaptive Reflection Fusion Model, STARFM)^[6]等。基于贝叶斯的模型^[20-22]利用贝叶斯统计理论, 对多个参考图像进行预测, 利用先验理论的原理, 在已知图像和预测图像之间建立数学关系模型, 这种方法可以正向直观地阐释融合过程。目前这些传统的融合模型已得到了实际应用, 如植被季节监测、森林物候分析、农作物监测、日蒸散量评估^[23]等, 并取得了不错的效果。

近年来, 基于学习的模型方法成为更多研究者的选择。稀疏表示和深度学习方法是最常见的 2 种方法^[24-25]。基于稀疏表示方法的理论是假设在同一天获取的 HTLS 和 LTHS 图像对共享相同的稀疏码。通过预先联合学习对应于 HTLS 和 LTHS 图像块的 2 个字典, 可以用所学习的字典以及稀疏编码算法重建用于预测的 LTHS 图像^[26]。基于深度学习的技术则是通过模拟人类大脑中神经元的工作方式, 在输入和输出之间建立复杂的非线性映射关系。近些年的研究表明, 卷积神经网络(Convolutional Neural Networks, CNN)对于处理图像特征提取和识别问题很有成效^[27], 使其逐渐应用于图像融合领域, 如全色和多光谱图像的融合问题等。但对于时空融合, 目前在深度学习领域进行的探索比较有限。Tan 等^[12]提出了一种时空融合模型——深度卷积时空融合网络(Deep Convolutional Spatial-Temporal Fusion Network, DCSTFN), 对 MODIS 图像和 Landsat 图像进行融合。一对 LTHS 和 HTLS 图像作为参考进行输入, 另一个 HTLS 图像作为预测。这个网络的融合过程是在特征级进行的, 然后利用融合之后的特征图重建生成预测图像。Liu 等^[28]提出了一种残差融合网络(Spatio-temporal fusion Network, StfNet)。通过在预测数据和真实数据之间建立非线性映射学习像素差异。上述文献都是基于深度学习的方法进行预测, 这些现有的基于 CNN 的模型虽已经大幅提高了融合的精确度, 但同时也存在一些不足: 第一, 融合模式单一, 若只在特征级进行融合, 会损失更多的图像纹理细节; 若只在原始像素级进行融合, 则会出现预测图像的细节信息高度依赖于 LTHS 图像的问题。因为所有详细的高频信息都来自这些图像, 所以预测准确与否会很大程度上依赖于这些图像。如果在参考和预测期间地面信息发生重大变化, 预测情况可能会变得很差; 第二, 融合部分都基于一定的线性假设, 这会导致预测结果受参考值的影响。研究并解决这些问题, 可以进一步提高预测的图像质量。

1 相关工作

1.1 拉普拉斯算子

拉普拉斯算子是一种简单的各向同性的微分算子, 具有旋转不变性。在图像处理领域, 经常把拉普拉斯算子作为边缘检测算子之一。一个二维图像函数的拉普拉斯变换是各向同性的二阶导数。拉普拉斯算子是二阶微分线性算子, 在对图像边缘的处理上, 二阶微分的边缘定位能力更强, 锐化效果更好。它的基本思想是当邻域的中心像素灰度低于它所在邻域内的其他像素的平均灰度时, 此中心像素的灰度进一步降低; 当高于它所在邻域内的其他像素的平均灰度时, 进一步提高中心像素的灰度, 从而实现图像锐化处理。

当邻域内像素灰度相同时, 拉普拉斯模板的卷积运算结果为 0; 当中心像素灰度高于邻域内其他像素的平均灰度时, 模板的卷积运算结果为正数; 当中心像素的灰度低于邻域内其他像素的平均灰度时, 模板的卷积为负数。对卷积运算的结果用适当的衰弱因子处理并加在原中心像素上, 就可以实现图像的锐化处理。

在图像处理中, 可以利用它的特性来增强图像的边缘特征。如果图像上的某个位置特征性很强, 经过拉普

拉斯算子可以让这个特征更加突出。应用到遥感图像时空融合领域中，可以利用它进行对图像边缘特征的突出挖掘，使预测图像对于地表的细节特征更为敏感。因此，本文将它运用到了复合损失函数的特征感知损失部分。

1.2 分散注意力机制

自 AlexNet 提出以来，深度卷积神经网络一直主导着图像分类。随着这一趋势，已经从手工特征逐渐转移到网络架构的研究，一种模块化的网络设计策略(Visual Geometry Group, VGG)被提出。重复堆叠相同类型的网络块，简化了网络设计的工作流程。提出的 ResNet 残差神经网络引入了一种身份跳跃连接，它减轻了深度神经网络中梯度消失的困难，并允许网络学习更深层次的特征表示。ResNet 已成为视觉领域最成功的架构之一，并被用于各种计算机视觉应用中。作为 ResNet 的一种变体，分散注意力(Split Attention, SA)多路径表示的结构应运而生。SA 模块是一个计算单元，由特征映射组和分散注意力操作组成。特征通道可被分为几组，超参数 r 表示模块内的分裂数。每个单独的组经过模块的主干分支后最后进行相加融合，所以每个分组都相当于一个独立的注意力机制模块。与 ResNet 相比，它不需要增加额外高的计算量，却可以显著提高网络特征筛选的能力。

2 本文方法

2.1 时空融合模型简介

以 Landsat 和 MODIS 遥感影像融合为例，L 和 M 分别表示 Landsat 和 MODIS 图像。在同一地理区域内，有在 t_0 、 t_1 时刻采集的 MODIS 图像，以及在 t_0 时刻采集的 Landsat 图像。时空融合的目标是在 t_1 时刻预测具有精细空间分辨力的类似陆地卫星的图像。通过在预测日期上的 Landsat 图像与参考数据之间建立关系映射，预测出在 t_1 时刻具有精细空间分辨力的 Landsat 图像。映射函数可以表示为：

$$L_{t_1} = F(M_{t_1}, L_{t_0}, M_{t_0} | \theta) \tag{1}$$

式中参数 θ 表示深度学习通过建立复杂的映射函数，以逼近实际函数 F 而学习到的一系列参数。

2.2 总体架构

图 1 为本文所提出的网络结构，由 4 部分组成：用于提取图像浅层特征的双层卷积组合编码器；用于提取深层特征的残差注意力机制模块(Residual Attention, RA)组合；用于进行多通道特征过滤的 SA 模块；用于进行图像融合生成的双层卷积重建解码器。该模型的目的是借助于 t_0 时刻的高分辨率图像 L_0 和低分辨率图像 M_0 以及 t_1 时刻的低分辨率图像 M_1 来预测 t_1 时刻对应的高分辨率图像 L_1 。

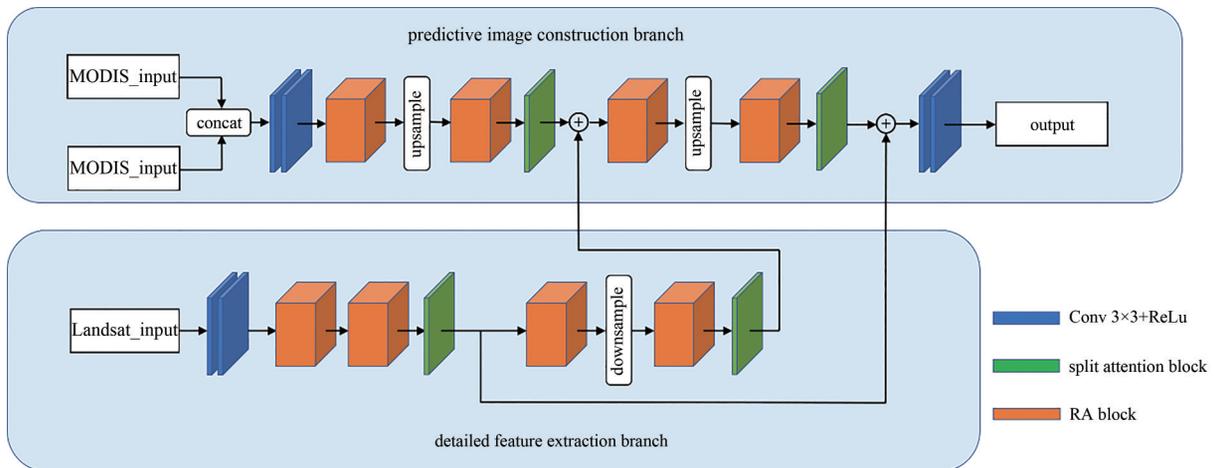


Fig.1 General network structure
图 1 网络总体结构

该网络结构由 2 个分支构成：一个是进行图像预测的主分支；一个是用于高分辨力图像细节特征提取的分支。本文在高分辨力图像 L_0 与低分辨率图像 M_0 、 M_1 之间建立层级映射关系，通过对低分辨率图像逐层上采样以及对高分辨力图像逐层下采样来生成相匹配的过渡图像，并将相对应的过渡图像特征元素进行逐像素相加，实现高分辨力图像对低分辨率图像的特征补偿。组合卷积用于图像浅层特征的提取，用来学习 Landsat 图像与 MODIS 图像间的差异。RA block 组合用于图像深层特征提取，高维卷积更有利于对图像细节信息的捕捉。该文

采用双三次插值法对初始 MODIS 图像进行上采样。使用插值法上采样不仅可以避免使用转置卷积上采样所带来的“棋盘伪陷”^[29]情况，也可以相应减少网络层的构建。SA 模块的输入为深层特征提取之后的特征组合，用来对特征通道进行筛选。最后经过重建卷积操作将这些高级特征恢复到原始像素空间衍生出预测。由式(2)~(4)表示：

$$P_1 = F_{sp} \left\{ F_{RA \times 2} \left[F_{en} \left(\text{concat}(M_0 + M_1) \right) \right] \right\} \quad (2)$$

$$P_2 = F_{sp} \left[F_{RA \times 2} (Q_2 + P_1) \right] \quad (3)$$

$$L_1 = F_{dc} (P_2 + Q_1) \quad (4)$$

式中： P_1 和 P_2 分别为图像预测分支中2次特征补偿的输入值； Q_1 和 Q_2 则为细节特征提取分支中补偿到预测分支中的2部分特征值； F_{en} 为双层卷积编码器； F_{dc} 为解码器； $F_{RA \times 2}$ 为2个连续的RA block，在2个RA block连续操作中间需要经过一次上采样操作； F_{sp} 为分散注意力机制模块； L_1 为模型的预测输出。特征提取分支为：

$$Q_1 = F_{sp} \left[F_{RA \times 2} (F_{en}(L_0)) \right] \quad (5)$$

$$Q_2 = F_{sp} \left[F_{RA \times 2} (Q_1) \right] \quad (6)$$

式中2个连续的RA block之间需要经过一次下采样。

在每次残差模块组进行特征提取之后，使用分散注意力机制来对融合特征进行校正。由于该文的网络结构中2个分支的上下采样操作都包含在注意力残差模块组中，而上下采样操作会对图像特征有很大的影响，所以需要在上下采样过程之后进行通道校正，且在特征补偿之前需要对关键特征通道进行捕获。与原本的分散注意力模块相比，SA模块将其中的批标准化(Batch Normal, BN)^[30]操作去掉，因为使用BN会破坏原有的数据分布，从而降低融合精确度。对于图像融合问题，本文希望模型能够学习到更多的细节，更好地适应图像融合等下游任务。经过实验验证，在去掉BN操作之后生成图像的清晰度及其指标方面均有一定提升。

引入并改进该模块的好处是，该模块相较于传统的注意力机制，只需要较少的参数量就可以达到相当的效果。按分组将权重赋值后的通道进行累加，可以将深层特征都详尽地进行表达，而不增加多余的计算量。在之后的网络训练中，可以关注到所有关键的通道特征信息。

Zhang等^[31]提出一种双向金字塔的Pan-Sharpening网络结构。类似地，该文中的特征补偿也是基于一种双向金字塔的结构。特征补偿发生在MODIS图像分别2次上采样之后。常见的模型都只进行对于高分辨力图像和低分辨力图像的一次特征融合。2种图像之间存在较大的放大系数，直接将MODIS图像放大到与Landsat图像同样的像素级别可能会导致不适定问题。通过进行多级的特征补偿，更好地解决了低分辨力图像中空间细节丢失的问题。

在实验中，使用过大的卷积核会增加计算中训练权重参数的数量，因此文中构建了多层卷积结构，并使用小卷积核代替大卷积核。对于大多数卷积层，将卷积核大小设置为 3×3 ，卷积步幅设置为1。对于提取浅层特征的卷积组合，将其3个卷积层的输出通道设置为32、64和128。重建操作采用滤波器大小为1的卷积层，减少计算量的同时，通过在前一层的学习表达上添加的非线性激励，更好地提升了网络的表达能力。

2.3 残差注意力机制模块

为了更好地进行特征提取和表达，设计了融合通道注意力机制残差模块(RA block)作为网络的基本组成结构。与原本的残差模块相比，通道注意力机制的引入可以更好地关注到图像的关键特征。将多维通道特征作为输入，设置残差网络的卷积通道数为64，所以每个RA模块的输入有64个通道的特征。模块分为2个分支：一个是用于深层特征提取的深度卷积分支。网络中通道数量固定为64，在每层卷积之后，由激活函数ReLU进行激活，连续的残差堆叠可以很大程度地对深层特征进行挖掘，但在图像上下采样以及初始化的随机值会导致一些噪声的引入，需要对上一层网络传来的特征进行校准取舍，引入通道注意力机制对特征通道进行加权筛选。另一个是通过通道注意力机制进行特征权重筛选的分支。上一层网络传递过来的多维特征通过压缩激励对特征通道进行筛选，在经过压缩之后将 $H \times W \times C$ 的输入转换成 $1 \times 1 \times C$ 的输出，经过全连接层之后计算为C组 1×1 特征权重，再用Sigmoid激励函数将权重值激活到0~1范围，起到对无用特征通道的抑制作用。最后通过 F_{scale} 将原始输入通道特征值与对应权重值进行相乘。RA block与传统残差模块的对比如图2所示。RA block内部操作可定义为：

$$z_c = F_{sq}(X) \quad (7)$$

$$s_c = F_{\text{ex}}(z_c, W) = \sigma[W_2(\text{ReLU}(W_1 z_c))] \quad (8)$$

$$X_1 = F_{\text{scale}}(X, s_c) = s_c X \quad (9)$$

式中： z_c 为对原始输入通道特征进行压缩之后的特征； s_c 为对 1×1 特征通道进行2个全连接层激励操作之后的特征权重值； X_1 为特征通道加权之后的输出。在全连接层中，为了降低模型复杂度且减少计算量，对第1层进行降维，然后在第2层中将特征权重数量还原到原来的维度 C 。

而单元模块的最终输出为特征经过2个分支之后进行相加的输出：

$$X_2 = \text{ReLU}[\text{ConV}(\text{ReLU}(\text{ConV}(X)))] \quad (10)$$

$$Y = X_1 + X_2 \quad (11)$$

式中 X_2 为对输入进行2组 3×3 卷积之后的结果，与 X_1 相加之后构成了输出 Y 。相比于在深度卷积之后进行通道筛选，这样设计的好处是，可以对特征通道进行小幅的调整。从数学角度来理解，加的加速度总是要小于乘的增加速度。在多次的迭代训练中，后者更容易陷入局部最小值，从而影响最终的效果。特征相加可表示为：

$$Y = \sum_{i=1}^n a_i + b_i \mu_i \quad (12)$$

式中： n 为原始特征通道数64； a_i 为深度卷积之后的特征； b_i 为原始输入特征； μ_i 为特征通道对应的权重值。

经过这样处理之后，特征通道可以赋予一个非线性调整，残差映射不再简单地快捷连接，经过通道注意力机制进行压缩激励之后的原输入数据，可以更好地保留浅层特征。在深度卷积过程中融合多元信息的能力也有一定增强。

2.4 损失函数

在传统的图像融合实验中都采用像素级损失作为损失函数，但是单一的像素级损失难以处理恢复丢失的高频细节纹理所固有的不确定性。为了预测图像的质量及其准确性，用自定义的复合损失函数来代替单一的像素级损失。这种新的复合损失由内容损失、特征感知损失、视觉损失组成。

内容损失是图像融合领域的基础，其计算方式为对预测图像与真实图像的逐像素损失进行均方误差(Mean Squared Error, MSE)计算，内容损失能够确保预测图像的基本质量，可表示为：

$$l_{\text{Content}} = \frac{1}{N} \sum_{i=1}^N (\hat{L}_i - L_i)^2 \quad (13)$$

式中： N 为图像的逐像素迭代编号； \hat{L}_i 为预测图像； L_i 为目标图像； $\hat{L}_i - L_i$ 为预测图像与目标图像进行逐像素作差。

Johnson等^[32]首次提出了特征感知损失。方式是将真实图像进行多层卷积处理之后得到的feature与预测生成的图像进行相同卷积之后得到的feature进行MSE计算，使预测图像的深层信息能够逼近真实图像。其中的卷积层使用预先训练好的卷积神经网络。这种方法的核心思想是利用预先训练好的网络来对深层细节信息进行感知，来比对预测数据与真实数据之间的差异。如果网络结构过于简单可能取得的效果不是很好，网络过于复杂又可能会出现过拟合现象，使得网络对于数据源的敏感性增强，对于某些差异性较大的数据源可能需要重新进行训练。这样不仅会耗费大量的计算资源，而且其取得的效果也不一定会达到预期。针对这个问题，Lei等^[33]提出了

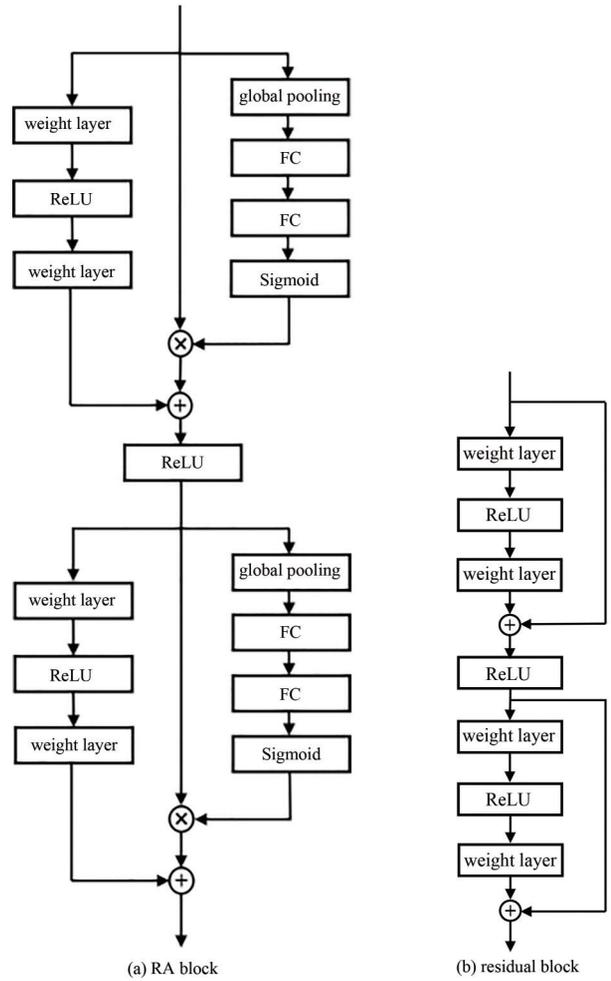


Fig.2 Comparison of RA block and residual block
图2 RA block与residual block对比

梯度损失, 利用一个自定义的卷积算子来代替预训练网络去获取真实图像和预测图像的 feature, 然后对 2 个 feature 进行 MSE 运算。提出一种基于拉普拉斯卷积算子的边缘损失。由于在经过网络迭代之后的特征图像, 其噪声对于特征内容来说是微乎其微的, 所以图像噪声并不会明显地影响到拉普拉斯算子对于图像特征感知的准确度。并且拉普拉斯算子是各向同性的, 能对任何走向的界线和线条进行锐化, 无方向性, 这种特性使其对图像边缘特征的感知能力比其他微分算子更强。

文中自定义卷积核设置为 3×3 , 卷积步长设置为 1。使用的非线性激活函数为修正线性单元(Rectified Linear Unit, ReLU)^[34]。边缘损失可以表示为:

$$l_{\text{Edge}} = \frac{1}{N} \sum_{i=1}^N (f_{\text{LP}}(\hat{L}_i) - f_{\text{LP}}(L_i))^2 \quad (14)$$

式中: N 为 2 个图像的逐像素迭代编号; f_{LP} 为拉普拉斯卷积算子。

视觉损失的加入可以提高预测图像的整体视觉感受和质量。关于视觉部分的损失该文采用在融合领域使用比较广泛的多尺度结构相似性(Multi-Scale Structural Similarity, MS-SSIM)指标计算方法来进行评估。SSIM 通过模拟人类的视觉感知过程, 从亮度、对比度和结构 3 个方面综合评价图像的相似性。在视觉损失方面, SSIM 评估出来的结果往往要比简单的 l_1 、 l_2 损失的结果更具有说服力, SSIM 表示为:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (15)$$

式中: x 、 y 分别为预测图像和真实图像; μ_x 和 μ_y 为预测图像和真实图像的平均值; σ_x^2 和 σ_y^2 为它们的方差; σ_{xy} 为图像 x 和 y 的协方差; c_1 和 c_2 是为了维持稳定而设置的 2 个变量参数。

MS-SSIM 是基于多层(图片按照一定规则, 由小到大缩放)的 SSIM 函数, 相当于考虑了分辨力。MS-SSIM 损失被定义为式(16), 以惩罚来自预测图像 y 得到的低 MS-SSIM。

$$l_{\text{Vision}} = 1 - [MS-SSIM(y)] \quad (16)$$

本文采用的复合损失函数整体可以用式(17)来表示, 其中 α 用来平衡视觉损失, 根据经验设置为 0.5。

$$l_{\text{Compose}} = l_{\text{Content}} + l_{\text{Edge}} + \alpha l_{\text{Vision}} \quad (17)$$

3 实验及结果分析

3.1 研究区域与数据集

时空融合实验中常用的一组数据源是 Landsat 和 MODIS 影像。实验中采用 2 级 Landsat-8 OLI 和为期 8 天的中分辨率成像光谱仪(Moderate-resolution Imaging Spectroradiometer, MODIS)合成数据 MOD09A1(地表反射率产品)来评估模型性能。选用为期 8 天的合成数据不仅大大减轻了大气和云层对实验精确度的影响, 并尽可能修复丢失的数据, 提供了更高的图像质量。为了验证本实验模型在各个数据源的通用性, 选择了中国的 2 个不同地区的影像作为数据源: 广东和山东, 时间跨越 2013 年 1 月 1 日至 2017 年 12 月 31 日。在广东省数据集中, 在 Landsat 全球参考系统(Worldwide Reference System, WRS)中选取坐标为 P122R043、P123R043 作为实验区域, 对应的 MODIS 正弦网格坐标为 h28v06。广东省地处沿海地区, 气候潮湿, 使得该区域地表大部分时间都被云层覆盖, 经过筛选之后可供参考使用的数据并不多。在山东省的数据集中, 选择 WRS 坐标为 P122R043 的 Landsat 图像, 相应的 MODIS 正弦网格坐标为 h27v05。山东省位于中国东部沿海, 境内山地突起, 平原盆地交错环列, 异质性相较广东省较低, 因此可以作为本文实验研究的重点数据集。本文主要针对山东省数据集来提升融合模型的性能, 为了之后实验中更好地证明各种方法的鲁棒性, 文中没有对实验数据集进行去云处理。

对于实验数据源的预处理, 由于满足实验条件的 MODIS 数据源相对较多, 可以对其进行较为严格的筛选。首先按照数据采集日期进行筛选, 使参考数据的日期与预测数据的日期尽可能相邻。对原始 MODIS 数据重采样到与 Landsat 相同的地理范围(空间分辨力为 480 m), 并提取了 4 个波段数据, 包括蓝色、绿色、红色和近红外(Near Infrared, NIR)波段。将原始的 Landsat 数据提取出的 4 个波段数据融合为新的 Landsat 图像, 然后把各个区域对应的 Landsat 图像都设置为固定大小 4800×4800 , 对应的 MODIS 图像裁剪为 300×300 。将数据预处理完毕后, 将对应的图像进行分组, 将 Landsat 图像和 MODIS 图像根据其参考日期和预测日期对应进行分组。将每个区域划分为 14 个数据组, 对于每个区域, 训练集和验证集都是随机划分的。

3.2 参数设置

文中网络框架是基于 Pytorch^[35]深度学习框架和 Python 语言进行编写搭建的。实验中，将 MODIS 图像的分块尺寸设置为 35×35 ，滑动步幅设置为 30×30 。与之相对应，Landsat 的设置是 MODIS 的 16 倍。将批处理的大小设置为 4，以适应图形处理器(Graphic Processing Unit, GPU)的内存，并使用 Adam 算法对提出的整个网络进行损失优化。学习率根据经验设置为 10^{-3} ，这是通过设置了多个参数 $\{10^{-4}$ 、 10^{-3} 、 $10^{-2}\}$ 进行比较确定出来的，在测试数据集获得了最低的均方根误差(Root Mean Square Error, RMSE)。整个训练过程迭代了 60 个 epoch 来确保收敛。实验在 Intel Xeon Gold 6148 CPU 2.40 GHz 和 NVIDIA Tesla V100-SXM2-16GB 的 PC 上训练整个网络。

3.3 比较和评估策略

与其他的图像融合方法进行比较时，采用 2 种传统的融合方法：时空自适应反射融合模型(STARFM)、灵活时空数据融合(Flexible Spatiotemporal Data Fusion, FSDAF)^[36-37]和 2 种基于深度学习的时空融合模型：深度卷积时空融合网络(Deep Convolutional Spatio-Temporal Fusion Network, DCSTFN)和增强型深度卷积时空融合网络(Enhanced Deep Convolutional Spatio-temporal Fusion Network, EDCSTFN)^[38]，验证本文所提出的新的时空融合网络结构的优越性。通过将预测图像与真实图像进行比较，对图像融合的结果进行定量和视觉评估。对于定量评估部分，采用 4 个指标对其进行评估：相关系数(Correlation Coefficient, CC)^[39]、结构相似性指数(SSIM)^[40]、光谱角度映射器(Spectral Angle Mapper, SAM)^[41]和综合的无量纲整体相对误差(Erreur Relative Global Adimensionnelle de Synthèse, ERGAS)^[42]。CC 通过将预测图像与真实图像进行线性联系来评价它们的相关性。SSIM 用来表示预测和真实图像之间总体结构的相似性。CC 和 SSIM 的理想值均为 1。SAM 通过光谱角度来评估光谱损失，SAM 较小，表示结果更好。ERGAS 通过评判预测误差对融合结果进行评价，它的值越小，表明融合结果越好。

3.4 实验结果

为验证本文所提出结构与所改进的模块的优越性与有效性，文中针对所提出结构进行了多组消融实验。

3.4.1 消融实验

在山东数据集上进行了消融实验。首先设计了基础网络结构 Baseline 作为对比，Baseline 中网络基础组成模块采用传统的 Residual Block。在 Baseline 中去掉复合损失中的边缘损失，其中损失函数只有内容损失和视觉损失两部分。Baseline 中去掉了 SA 模块，为了体现改进的合理性，文中将原分散注意力机制模块与改进之后的分散注意力机制模块进行对比。实验结果如表 1 所示。

表 1 不同模块结构在山东数据集上的性能测评
Table 1 Performance evaluation of different block structures on Shandong dataset

Baseline	RA	loss	SA	BN	SAM	ERGAS	CC	SSIM
√					5.058 4	6.500 8	0.740 0	0.777 7
√	√				4.375 1	4.778 3	0.745 0	0.788 5
√		√			4.403 1	5.856 4	0.742 6	0.784 7
√			√		4.430 2	5.126 6	0.751 4	0.790 6
√			√	√	4.588 7	5.782 8	0.745 1	0.789 7
√	√	√	√		4.313 2	4.302 6	0.752 1	0.795 4
		reference			0	0	1	1

表 1 中列出了不同模块对于结果所造成的影响，分析结果可得，梯度损失的加入可以使融合图像的质量有一定的提升。引入 RA 模块以及 SA 模块，对于结果也都有一定的提升。同时，通过改进后的分散注意力机制模块与原本的分散注意力机制模块的效果对比，证明了 BN 层操作对于此类任务的不适用性。消融实验证明了模型中各个部分的有效性。

使用决定系数 R^2 评估预测结果，通过它可以有效评估观察图像和预测图像之间的关联性。其目的是对预测图像和观测图像的像素值进行拟合，获得一个回归函数，并确定其统计量的接近程度，其定义如下：

$$R^2 = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x}_i)^2} \quad (18)$$

式中： \bar{x}_i 为观测图像的平均像素值； x_i 为预测图像的像素值； y_i 为观测图像的像素值； N 为图像的总像素数。当该值接近于 1 时，预测图像的性能会提高。

图 3 描述了每种消融方法的预测结果与真实图像近红外波段之间的相关性。MFCNet(Multi-stage Feature

Compensation Net)为本文所提出的完整方法, SA-old表示未进行改进的SA模块。这张图更直观地说明了在不同的实验下, 预测结果与观察结果的匹配情况。对于实验样品, 本文选择了近红外波段作为代表进行详细演示。可以看出, BatchNormal层对预测结果有很大的负面影响。MFCNet的点云更集中在拟合的直线上, 散落在外面的点更少, 轮廓线更平滑。图中的 R^2 值和拟合直线的斜率和截距也表明, MFCNet的预测结果更准确, 预测结果与观测图像的相关性更好。综上所述, 对统计指标和相关图的主观考察表明, 本文提出的完整网络的预测图像与观测图像更接近, 预测结果与观测结果更吻合。同时, 这也更加证明了文中提出的模块的有效性。

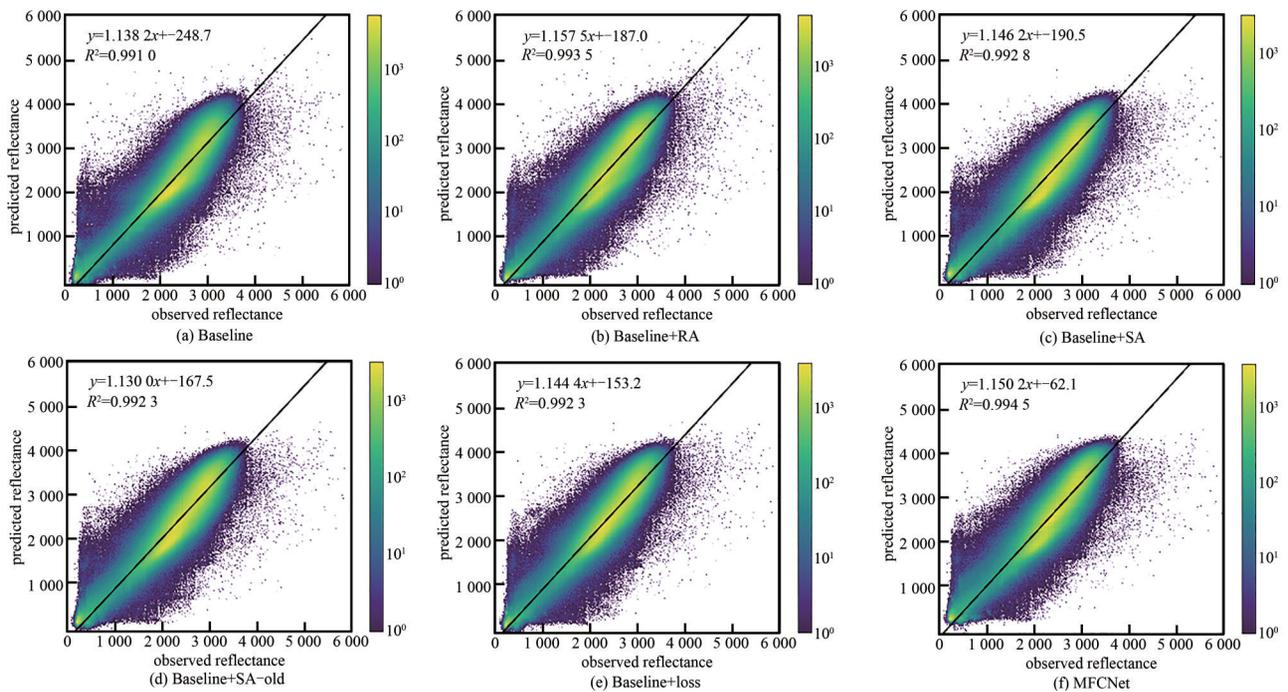


Fig.3 Correlation between the predicted and real image NIR bands for each ablation method
图3 各消融方法预测图像与真实图像近红外波段之间的相关性

3.4.2 对比实验

为了能够直观地看到各种方法所融合出来图像的对比差异, 将在同一时间内的融合图像在相同的位置截取相同大小的图像作对比。截取的图像块大小为 300×300 , 并把图像块中具有突出特征的 100×100 的部分进行放大, 以便能够更加清晰地看出各种方法的优缺点。由对比图可以看出, 无论是在哪个数据集上, 传统方法STARFM和FSDAF都存在严重的光谱失真问题, 且图像的纹理细节特征也没有很好地表达出来。这可能是由于数据集没有进行去云处理的原因, 且传统方法的鲁棒性较差, 导致预测结果的质量不高。跟传统方法相比, 深度学习可以比较好地解决这些问题, 这也体现出深度学习较于传统方法在图像融合领域的优势。

山东数据集上的融合结果如图4所示, 由于山东的气候比广东干燥, 可获得的无云或者少云的图像也比较多, 山东地区的数据集相对质量较高, 在此数据集上的融合图像更有参考价值。从图中可以看出, 使用深度学习的方法相比于传统方法, 预测的结果更为精确。EDCSTFN相比其他的深度学习模型, 其表现出的光谱信息更为丰富, 但本文方法可以更好地降低光谱的差异性, 更加准确地还原真实图像。山东数据集的指标测试结果如表2所示, 可以看到, 较于其他方法, 本文方法指标均有所提高。

表2 山东数据集上的平均量化指标

Table2 Average quantitative index on Shandong dataset

method	SAM	ERGAS	CC	SSIM
AMNet	6.562 2	7.398 8	0.698 9	0.680 2
DCSTFN	7.532 9	13.057 5	0.620 5	0.574 4
EDCSTFN	4.504 4	6.963 5	0.733 8	0.579 3
FSDAF	12.748 3	23.090 3	0.444 6	0.487 5
STARFM	14.035 3	26.749 2	0.462 4	0.419 0
MFCNet	4.313 2	4.302 6	0.752 1	0.795 4
reference	0	0	1	1

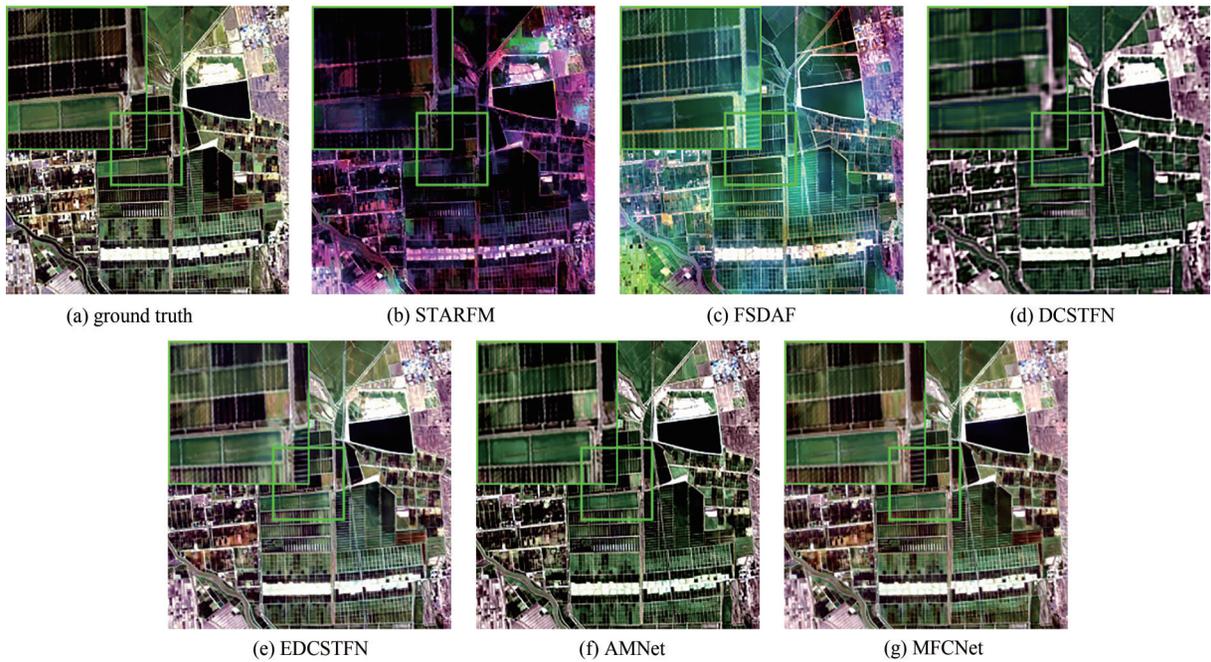


Fig.4 Comparison of predicted and actual images in Shandong province

图 4 山东省的预测图像和实际图像的比较

广东数据集上的融合结果如图 5 所示，广东地处沿海，气候较潮湿，地表大部分时间会被云层覆盖，因此相对于山东数据集，广东数据集的质量不高。观察图像可以发现，各个图像的表现结果与在山东数据集上大同小异。传统方法表现较差，在深度学习方法中，与其他方法相比，本文所提出的方法融合结果更加清晰，精确度更高。广东数据集的指标评估结果如表 3 所示。从指标数值上可以看出，本文方法在各个指标上都有明显提高。综上所述，本文方法所得到的预测图像在指标结果和视觉感受两方面都优于其他方法。

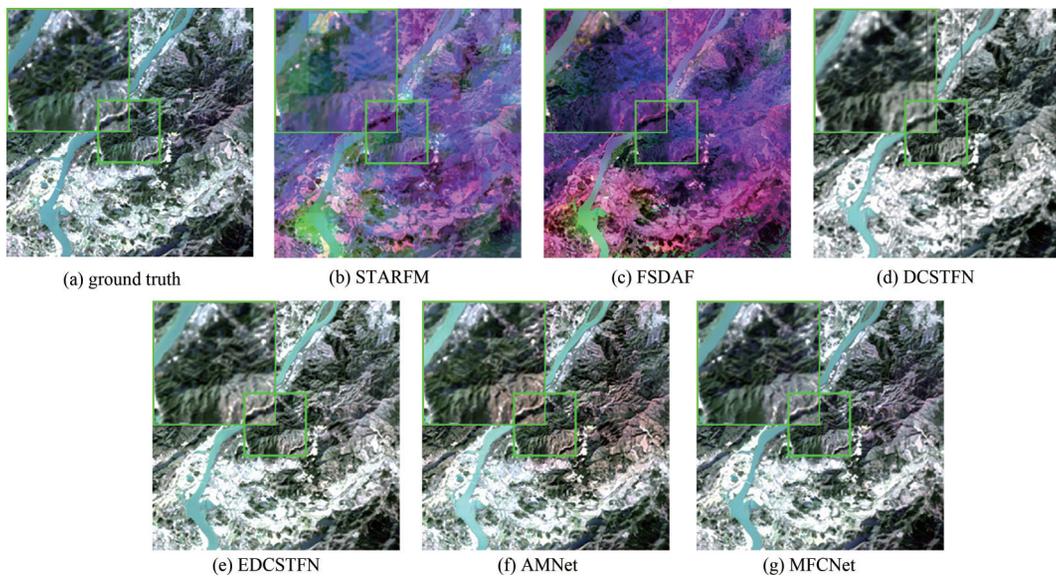


Fig.5 Comparison of predicted and actual images in Guangdong province

图 5 广东省的预测图像和实际图像的比较

此外，本文还从 2 个维度比较了基准实验中基于深度学习的模型，包括模型参数数量和浮点运算(FLOPs)。模型参数数量表示模型需要学习的参数数量；FLOPs 用于测量模型的计算复杂性，在 FLOPs(G)中，G 为 1×10^6 。结果示于表 4 中(Reference 为 0，表示越小越好；为 1，表示越大越好)。因为 EDCSTFN 要求预训练模型，具有更多的参数和更复杂的模型。MFCNet 在参数数量和 FLOPs 方面都远低于其他方法，可以认为，该文所提出的模型在参数数量和计算复杂度方面更具优势。

表3 广东数据集上的平均量化指标

Table3 Average quantitative index on Guangdong dataset

method	SAM	ERGAS	CC	SSIM
AMNet	3.473 2	4.657 3	0.922 9	0.912 2
DCSTFN	3.910 7	8.624 9	0.904 6	0.881 0
EDCSTFN	3.894 9	9.409 6	0.903 9	0.880 4
FSDAF	12.748 2	14.785 4	0.444 6	0.482 0
STARFM	14.035 3	11.883 2	0.462 4	0.476 0
MFCNet	3.394 8	3.495 5	0.942 2	0.940 0
reference	0	0	1	1

表4 模型参数量和模型复杂性的对比

Table4 Comparison of model parameter quantities and model complexity

method	parameters	FLOPs(G)
DCSTFN	408 961	150.481
EDCSTFN	762 856	111.994
AMNet	633 452	97.973
MFCNet	375 284	84.960
reference	0	0

3.4.3 残差实验

将预测图像与真实图像作残差，即逐像素相减，然后对图像的4个波段求残差的均值，观察融合图像的质量。图6和图7展示了山东和广东数据集上各种方法的预测图像与真实图像作差之后的结果。在残差实验中，残差图像显现出来的纹理细节越少，证明图像融合的效果越好。由结果可以看出，传统方法STARFM和FSDAF预测出来的图像作残差后，显现出的建筑细节比较明显，说明图像对于真实图像的光谱预测能力较差。这表明传统方法对于图像光谱细节的表达能力相对不足。而深度学习的几种方法虽较传统方法效果较好，但也表现出了较为清晰的地表纹理，表明它们的结果也有待提高。与它们相比，MFCNet方法的残差图显现出了很少的细节，说明该文方法的融合图像效果更好，证明了提出方法的有效性。

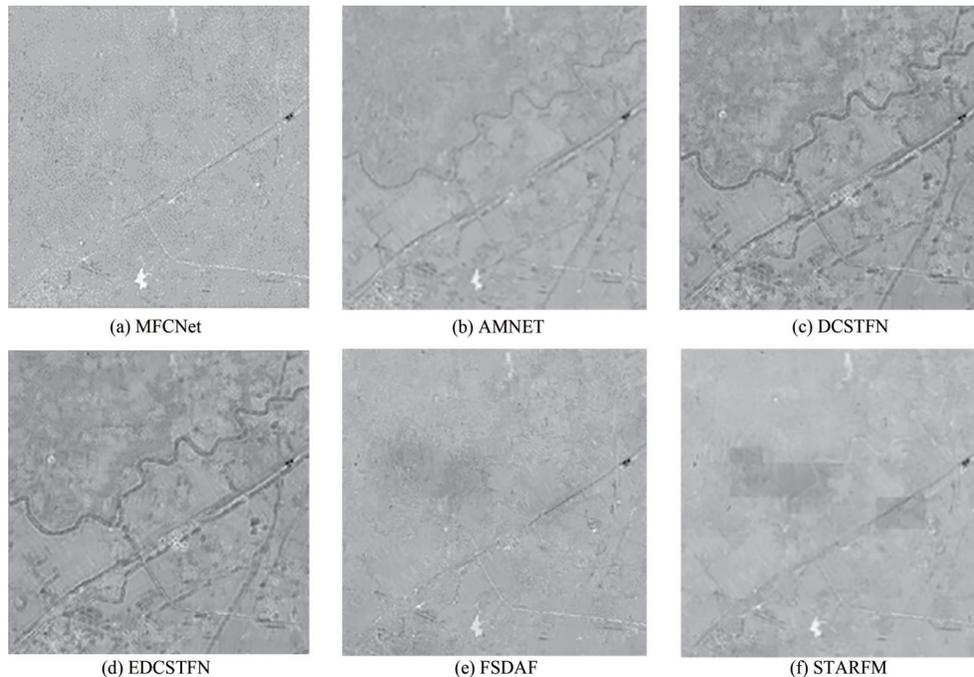


Fig.6 512×512 residual image on Shandong dataset

图6 山东数据集上的512×512残差图

4 结论

本文提出了一种新的基于深度学习的时空图像融合方法，利用多级特征补偿提高了模型还原细节纹理特征的能力。引入了通道注意力机制对基础残差模块进行改进，提高了它的特征提取能力。使用基于拉普拉斯算子

的边缘损失模块代替复合损失中传统的基于预训练模型的特征损失模块，使模型收敛的过程中对突出的特征信息更加敏感，预测图像细节纹理更接近于真实图像。通过数据集的对比实验可以看出，本文方法不仅明显优于传统的时空融合方法，且对比其他的一些深度学习方法，该方法也具有更为精确的预测能力。本文中模型还存在很多的不足，未来工作包括对于数据集的处理和改进，以及模型的继续优化。

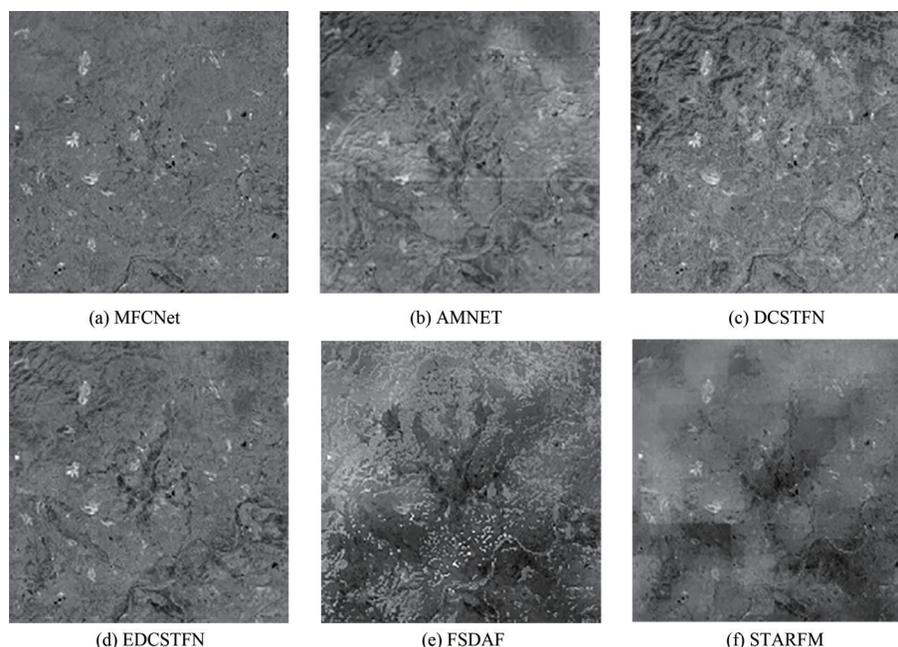


Fig.7 512×512 residual image on Guangdong dataset

图7 广东数据集上的512×512残差图

参考文献：

- [1] TAN Zhenyu,DI Liping,ZHANG Mingda,et al. An enhanced deep convolutional model for spatiotemporal image fusion[J]. Remote Sensing, 2019,11(24):2898. doi:10.3390/rs11242898.
- [2] PASTOR-GUZMAN J,BROWN L,MORRIS H,et al. The sentinel-3 OLCI Terrestrial Chlorophyll Index(OTCI):algorithm improvements, spatiotemporal consistency and continuity with the MERIS archive[J]. Remote Sensing, 2020,12(16):2652. doi: 10.3390/rs12162652.
- [3] GUHA S,GOVIL H. An assessment on the relationship between land surface temperature and normalized difference vegetation index[J]. Environment,Development and Sustainability, 2021,23(2):1944-1963. doi:10.1007/s10668-020-00657-6.
- [4] SHEN Huanfeng,MENG Xiangchao,ZHANG Liangpei. An integrated framework for the spatio-temporal-spectral fusion of remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016,54(12):7135-7148. doi:10.1109/TGRS.2016.2596290.
- [5] CHEN Bin,HUANG Bo,XU Bin. Comparison of spatiotemporal fusion models:a review[J]. Remote Sensing, 2015,7(2):1798-1835. doi:10.3390/rs70201798.
- [6] GAO F,MASEK J,SCHWALLER M,et al. On the blending of the Landsat and MODIS surface reflectance:predicting daily landsat surface reflectance[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006,44(8): 2207-2218. doi: 10.1109/TGRS.2006.872081.
- [7] HILKER T,WULDER M A,COOPS N C,et al. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model[J]. Remote Sensing of Environment, 2009,113(9):1988-1999. doi:10.1016/j.rse.2009.05.011.
- [8] KHALEGHI B,KHAMIS A,KARRAY F O,et al. Multisensor data fusion:a review of the state-of-the-art[J]. Information Fusion, 2013,14(1):28-44. doi:10.1016/j.inffus.2011.08.001.
- [9] BELGIU M,STEIN A. Spatiotemporal image fusion in remote sensing[J]. Remote Sensing, 2019, 11(7): 818. doi: 10.3390/rs11070818.
- [10] ZHANG Hongyan,SONG Yiyao,HAN Chang,et al. Remote sensing image spatiotemporal fusion using a generative adversarial network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020,59(5):4273-4286. doi:10.1109/TGRS.2020.3010530.

- [11] MASEK J G, WULDER M A, MARKHAM B, et al. Landsat 9: empowering open science and applications through continuity[J]. *Remote Sensing of Environment*, 2020(248):111968. doi:10.1016/j.rse.2020.111968.
- [12] TAN Zhenyu, YUE Peng, DI Liping, et al. Deriving high spatiotemporal remote sensing images using deep convolutional network[J]. *Remote Sensing*, 2018,10(7):1066. doi:10.3390/rs10071066.
- [13] 李昌洁, 宋慧慧, 张开华, 等. 条件生成对抗遥感图像时空融合[J]. *中国图象图形学报*, 2021,3(3):714–726. (LI Changjie, SONG Huihui, ZHANG Kaihua, et al. Spatiotemporal fusion of satellite images via conditional generative adversarial learning[J]. *Journal of Image and Graphics*, 2021,3(3):714–726.) doi:10.11834/jig.200219.
- [14] ACERBI–JUNIOR F W, CLEVERS J, SCHAEPMAN M E. The assessment of multi–sensor image fusion using wavelet transforms for mapping the Brazilian Savanna[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2006,8(4):278–288. doi:10.1016/j.jag.2006.01.001.
- [15] PENG Yidong, LI Weisheng, LUO Xiaobo, et al. Integrated fusion framework based on semicoupled sparse tensor factorization for spatio–temporal–spectral fusion of remote sensing images[J]. *Information Fusion*, 2021(65):21–36. doi:10.1016/j.inffus. 2020.08.013.
- [16] 石岩. 基于二维不可分形态小波变换的多光谱图像全色锐化方法[J]. *北京交通大学学报*, 2018,42(5):116–122. (SHI Yan. A multispectral pansharpening method based on 2D nonseparable morphological wavelet transform[J]. *Beijing Jiaotong University*, 2018,42(5):116–122.) doi:10.11860/j.issn.1673–0291.2018.05.016.
- [17] 陶旭婷, 和红杰, 陈帆, 等. 基于局部相关性的遥感图像全色锐化算法[J]. *光子学报*, 2014(3):139–144. (TAO Xuting, HE Hongjie, CHEN Fan, et al. Pan–sharpening algorithm for remote sensing images based on local correlation[J]. *Acta Photonica Sinica*, 2014(3):139–144.)
- [18] HE Lin, ZHU Jiawei, LI Jun, et al. Spectral–fidelity convolutional neural networks for hyperspectral pansharpening[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020(13):5898–5914. doi:10.1109/JSTARS.2020.3025040.
- [19] 王乐, 杨晓敏. 基于感知损失的遥感图像全色锐化反馈网络[J]. *计算机科学*, 2021,48(8):91–98. (WANG Le, YANG Xiaoming. Remote sensing image pansharpening feedback network based on perceptual loss[J]. *Computer Science*, 2021,48(8):91–98.) doi:10.11896/j.jsjcx.200700112.
- [20] HUANG Bo, ZHANG Hankui, SONG Huihui, et al. Unified fusion of remote–sensing imagery: generating simultaneously high–resolution synthetic spatial–temporal–spectral earth observations[J]. *Remote Sensing Letters*, 2013,4(6):561–569. doi:10.1080/2150704X.2013.769283.
- [21] XUE Jie, LEUNG Y, FUNG T. A Bayesian data fusion approach to spatio–temporal fusion of remotely sensed images[J]. *Remote Sensing*, 2017,9(12):1310. doi:10.3390/rs9121310.
- [22] CHEN Junfu, PAN Yue, CHEN Yang. Remote sensing image fusion based on Bayesian GAN[J/OL]. arXiv:2009.09465, 2020.
- [23] ZHUANG Qifeng, SHI Yintao, SHAO Hua, et al. Evaluating the SSEBop and RSPMPT models for irrigated fields daily evapotranspiration mapping with MODIS and CMADS data[J]. *Agriculture*, 2021,11(5):424. doi:10.3390/agriculture11050424.
- [24] HUANG Bo, SONG Huihui. Spatiotemporal reflectance fusion via sparse representation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012,50(10):3707–3716. doi:10.1109/TGRS.2012.2186638.
- [25] SONG Huihui, LIU Qingshan, WANG Guojie, et al. Spatiotemporal satellite image fusion using deep convolutional neural networks[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018,11(3):821–829. doi:10.1109/JSTARS.2018.2797894.
- [26] SONG Huihui, HUANG Bo. Spatiotemporal satellite image fusion through one–pair image learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012,51(4):1883–1896. doi:10.1109/TGRS.2012.2213095.
- [27] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. *计算机学报*, 2019,42(3):453–482. (ZHANG Shun, GONG Yihong, WANG Jinjun. The development of deep convolutional neural networks and its applications on computer vision[J]. *Journal of Computer Science and Technology*, 2019,42(3):453–482.) doi:10.11897/SP.J.1016.2019.00453.
- [28] LIU Xun, DENG Chenwei, CHANUSSOT J, et al. StfNet: a two–stream convolutional neural network for spatiotemporal image fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019,57(9):6552–6564. doi:10.1109/TGRS.2019.2907310.
- [29] ODENA A, DUMOULIN V, OLAH C. Deconvolution and checkerboard artifacts[J]. *Distill*, 2016, 1(10): e3. doi:10.23915/distill.00003.
- [30] LUO Chunjie, ZHAN Jianfeng, WANG Lei, et al. Extended batch normalization[J/OL]. arXiv:2003.05569, 2020.
- [31] ZHANG Yongjun, LIU Chi, SUN Mingwei, et al. Pan–sharpening using an efficient bidirectional pyramid network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019,57(8):5549–5563. doi:10.1109/TGRS.2019.2900419.

- [32] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]// European Conference on Computer Vision. [S.l.]:Springer, Cham, 2016:694-711. doi:10.1007/978-3-319-46475-6_43.
- [33] LEI Dajiang, BAI Menghao, ZHANG Liping, et al. Convolution neural network with edge structure loss for spatiotemporal remote sensing image fusion[J]. International Journal of Remote Sensing, 2022, 43(3):1015-1036. doi:10.1080/01431161.2022.2030070.
- [34] CHEN Yinpeng, DAI Xiyang, LIU Mengchen, et al. Dynamic ReLu[C]// European Conference on Computer Vision. [S.l.]:Springer, Cham, 2020:351-367. doi:10.1007/978-3-030-58529-7_21.
- [35] KOKHLIKYAN N, MIGLANI V, MARTIN M, et al. Captum: a unified and generic model interpretability library for pytorch[J/OL]. arXiv:2009.07896, 2020.
- [36] LIU Meng, YANG Wei, ZHU Xiaolin, et al. An Improved Flexible Spatiotemporal Data Fusion(IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series[J]. Remote Sensing of Environment, 2019 (227):74-89. doi:10.1016/j.rse.2019.03.012.
- [37] WANG Qunming, TANG Yijie, TONG Xiaohua, et al. Virtual image pair-based spatio-temporal fusion[J]. Remote Sensing of Environment, 2020(249):112009. doi:10.1016/j.rse.2020.112009.
- [38] AO Z, SUN Y, XIN Q C. Constructing 10 m NDVI time series from Landsat 8 and sentinel 2 images using convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(8):1461-1465. doi:10.1109/LGRS.2020.3003322.
- [39] ALCARAS E, PARENTE C, VALLARIO A. Automation of pansharpening methods for pleiades images using GIS basic functions[J]. Remote Sensing, 2021, 13(8):1550. doi:10.3390/rs13081550.
- [40] LI Weisheng, XIANG Minghao, LIANG Xuesong. MDCwFB: a multilevel dense connection network with feedback connections for pansharpening[J]. Remote Sensing, 2021, 13(11):2218. doi:10.3390/rs13112218.
- [41] XIE Weiyang, CUI Yuhang, LI Yunsong, et al. HPGAN: hyperspectral pansharpening using 3D generative adversarial networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(1):463-477. doi:10.1109/TGRS.2020.2994238.
- [42] SCARPA G, CIOTOLA M. Full-resolution quality assessment for pansharpening[J/OL]. arXiv:2108.06144, 2021. doi: 10.3390/rs14081808.

作者简介:

刘文杰(1998-), 男, 在读硕士研究生, 主要研究方向为遥感图像处理. email:s200201117@stu.cqupt.edu.cn.

李雨珈(2003-), 女, 学士, 主要研究方向为遥感图像处理.

白梦浩(1996-), 男, 在读硕士研究生, 主要研究方向为遥感图像处理.

张莉萍(1983-), 女, 博士, 讲师, 主要研究方向为遥感图像处理.

雷大江(1979-), 男, 博士, 教授, 主要研究方向为机器学习、遥感影像处理与解译.