

FVC-Net: 高空间分辨率植被覆盖度的融合网络

张智昊, 王群明, 丁欣宇

同济大学 测绘与地理信息学院, 上海 200092

摘要: 植被覆盖度 (FVC) 是描述地表植被分布的定量指标之一。通过遥感卫星 (如 Landsat 和 Sentinel-2) 获取大尺度下的高空间分辨 (如 10 m 级) FVC, 能为全球生态系统研究提供重要基础数据。然而, 由于云雾干扰以及卫星重返时间分辨率有限等问题, 高空间分辨 FVC 在时域上存在大量缺失。本文考虑协同 30 m Landsat 8 和 10 m Sentinel-2 数据, 实现二者在时域上的互补。为解决二者空间分辨率不一致的问题, 本文提出了一种基于 FVC-Net 的深度学习方法, 通过融合 10 m Sentinel-2 归一化植被指数 (NDVI) 数据, 将 30 m Landsat FVC 降尺度至 10 m。FVC-Net 方法构建双分支结构下的通道注意力模块用于 FVC 和 NDVI 的多尺度特征采集与融合, 随后利用空间注意力模块将选择的特征进行细节增强, 以有效描述不同获取时间下的 10 m NDVI 与 30 m FVC 之间的非线性映射关系。实验中, 与 4 种典型非深度学习方法及 4 种深度学习方法相比, FVC-Net 获得了更高精度的融合结果。FVC-Net 有望应用于全球尺度下的 30 m Landsat FVC 产品的降尺度, 为相关领域研究提供更为精细的数据支撑。

关键词: 遥感, 植被覆盖度 (FVC), 归一化植被指数 (NDVI), 深度学习, 降尺度, 数据融合

中图分类号: P2

引用格式: 张智昊, 王群明, 丁欣宇. 2024. FVC-Net: 高空间分辨率植被覆盖度的融合网络. 遥感学报, 28(12): 3184–3196

Zhang Z H, Wang Q M and Ding X Y. 2024. FVC-Net: A fusion network for producing fine spatial resolution fractional vegetation cover. National Remote Sensing Bulletin, 28(12): 3184–3196 [DOI: 10.11834/jrs.20243112]

1 引言

中国拥有储量丰富的林业资源以及城市绿地生态系统。科学准确地掌握植被资源信息, 统筹资源、环境与经济发展相协调, 将会为中国绿色经济带来巨大的价值, 对“双碳”目标的实现具有十分重要的意义 (匡文慧, 2019)。植被覆盖度 FVC (Fractional Vegetation Cover) 定义为植被 (包括叶面、枝干等) 垂直投影面积占投影下垫面总面积的百分比, 是反映地表植被覆盖情况与生态环境的重要指标。基于地面验证样本数据的传统 FVC 获取方法需要大量的人力资源和抽样统计工作, 能够获取较为精确的地面实测数据, 但是巨大的时间和人工成本限制了传统地面调查方法的大范围应用。通过遥感手段, 能快速获取大范围地表覆盖信息。因而, 利用遥感数据反演地表生态参数, 如森林清查、储蓄量调查以及 FVC 估算等已受到越来越多的关注。

目前基于遥感数据的 FVC 反演算法主要有模型法、相对植被丰度法以及机器学习方法等。基于模型的方法利用遥感数据建立与 FVC 的模型来估算区域内的植被覆盖信息 (Boyd 等, 2002)。回归模型法利用植被与非植被之间在反射率上存在的差异, 将单波段或波段组合的遥感数据和 FVC 建立线性或非线性的回归关系, 进而能估算区域生长情况 (Xiao 和 Moody, 2005)。基于像元分解模型法利用植被与非植被的光谱响应特性, 将地物进行二分或多分类, 统计像元组成和比例, 通过像元分解模型分析得到研究区域的 FVC (陈云浩 等, 2002; Guerschman 等, 2009; Degerickx 等, 2019)。像元分解法的关键是对植被和非植被像元的划分, 但是实际复杂环境中很难获取纯净像元。对于相对植被丰度法, 其假设一个像素中只包含绿色植被和土壤, 通过植被指数 (通常是 NDVI) 和 FVC 构建二者的线性或非线性函数关系,

收稿日期: 2023-04-12; 预印本: 2023-10-30

基金项目: 国家自然科学基金 (编号: 42222108, 42171345)

第一作者简介: 张智昊, 研究方向为遥感数据融合。E-mail: zhangzh3520@ecepd.com

通信作者简介: 王群明, 研究方向为遥感信息智能处理与应用。E-mail: wqml1111@126.com

核心是确定合适的全FVC和裸土归一化值 (Gutman 和 Ignatov, 1998)。由于相对植被指数法物理意义明确、模型简单, 因而得到较多应用 (Zeng 等, 2003; 崔天翔 等, 2013)。此外, 伴随计算机硬件和技术的快速发展, 机器学习方法被广泛的应用到FVC的估算 (Su, 2009; Huang等, 2008; Hansen 等, 2002; Chen 和 Ge, 2022)。

在全球尺度上, 中粗和中等空间分辨率遥感数据是现今较为常用的遥感数据源, 其中应用最为频繁的传感器是 Landsat 系列 (包括 TM、ETM 和 OLI) 数据, 以及 MODIS 数据, 分别能提供空间分辨率为 30 m 和 250 m 的植被覆盖度估算数据。相较于 MODIS 数据, Landsat 数据因其较高的空间分辨率, 对 FVC 细节的表达更为丰富, 因而广泛应用于区域尺度下的 FVC 估算。基于此, 现已有研究生产 30 m FVC 数据集 (Song 等, 2022)。然而, Landsat 数据重访周期为 16 d, 且易受到云层和阴影的污染, 其时间序列数据无法监测 FVC 的快速变化。另一方面, 欧洲航天局发射的 Sentinel-2 卫星能周期性地全球尺度覆盖, 同时 Sentinel-2A 和 2B 卫星的联合观测能将重返周期提高至 5 d。相应地, Sentinel-2 MSI 数据已逐渐成为 FVC 估算主要遥感数据源之一 (Wang 等, 2020)。然而, Sentinel-2 MSI 数据同样易受到云层干扰而存在大量数据缺失, 使得有效观测数据的时间间隔通常远多于 5 d。为获取更高频次的高分辨率观测数据, 可考虑协同利用 Landsat 与 Sentinel-2 数据。为解决二者分辨率不一致的问题, 可通过融合方法, 将 30 m FVC 数据降尺度至 10 m。具体地, 为降低 10 m Sentinel-2 FVC 反演过程不确定性带来的影响, 本文将通过直接融合 30 m Landsat FVC 与 10 m Sentinel-2 NDVI 数据, 实现 30 m FVC 的降尺度。

现有通过融合方式获取高空间分辨率图像的方法主要基于全色锐化 (Wang 等, 2016; Wang 等, 2023)。全色锐化将高空间分辨率全色 (PAN) 图像的空间特征与低分辨率多光谱 (MS) 图像的光谱特征进行融合, 以提高 MS 图像的空间分辨率 (Ranchin 等, 2003)。全色锐化方法一般假设 PAN 图像和 MS 图像的多个通道存在线性关系。然而, FVC 与 NDVI 之间通常存在非线性关系, 尤其是在 FVC 极高和极低区域。另一方面, Landsat 与 Sentinel-2 数据获取时间不一致, 植被变化 (如农作物的生长) 会进一步降低二者之间相关性。

现有研究中, 已有学者利用深度学习网络处理图像融合问题 (Chen 和 Ge, 2022; Jia 等, 2021)。深度学习方法采用卷积等操作提取图像特征, 通过网络训练参数模型, 利用损失函数衡量融合结果与真实图像之间的差异, 通过迭代的形式进行优化, 得准确结果。针对上述传统方法存在的难点, 本文采用深度学习理论, 提出一种双分支融合网络 FVC-Net, 融合 10 m Sentinel-2 NDVI 与 30 m Landsat 8 FVC 得到 10 m Landsat 8 FVC 数据。

2 方 法

在本节中将详细介绍提出的 FVC-Net 方法。首先, 将对网络结构进行概述, 包括网络框架、损失函数定义等。然后, 重点介绍其中的通道注意力模块 CA-block (Channel Attention Block) 以及空间注意力模块 SA-block (Spatial Attention Block)。

2.1 FVC-Net概述

本文目标是从 30 m Landsat FVC 数据中生成具有 10 m 空间分辨率的 FVC 数据。具体地, 通过融合其他时间点的 10 m Sentinel-2 NDVI 数据, 构建 NDVI 与 FVC 之间的关系实现 FVC 空间分辨率的提升。提出的 FVC-Net 在原始数据的约束下, 在退化和降尺度过程中, 构建同一区域不同尺度之间的关系实现降尺度 (Wang 等, 2020)。在 FVC-Net 网络中, 10 m 空间分辨率的 NDVI 图像为降尺度过程提够了纹理信息, 使得重建后的 FVC 能够获得更加清晰的空间细节。如图 1 所示, FVC-Net 的两个分支分别是 FVC 分支和 NDVI 分支。FVC 分支主要实现原始低空间分辨率 FVC (LRFVC) 到高空间分辨率 FVC (HRFVC) 的光谱 (强度) 信息重建。NDVI 分支的作用是提取和选择空间纹理特征, 然后单向的注入到对应的 FVC 分支的层结构中, 保证网络融合具有与 LRFVC 相同的强度分布的同时, 获得与高分辨率 NDVI 相同的空间细节。网络各层的具体细节如表 1 所示, 其中 P 为像素大小。

具体地, 我们通过反卷积过程将 LRFVC 上采样到与 NDVI 相同的大小。随后在两个分支中分别使用了不同大小的通道注意力模块和空间注意力模块提取和保存图像特征。利用通道注意力保证 LRFVC 强度分布不发生改变, 利用空间注意力模块提取 NDVI 的空间特征重建 HRFVC 的细节纹理。不同大小的通道注意力模块和空间注意力模块的感受区域大小不同, 可以接受不同尺度下的特征

信息。由于网络卷积层的加深，NDVI分支中获得的空
间特征在不断缩小，高特征层存在一定的模
糊和信息丢失，因此我们将各个尺度下提取的空
间特征信息全部保存，减少空间细节的损失。

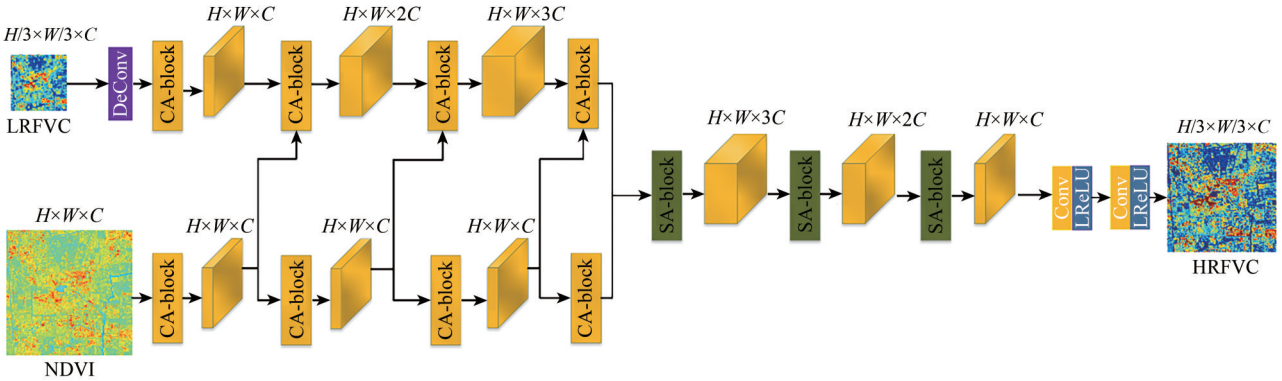


图1 FVC-Net网络结构
Fig. 1 The structure of the proposed FVC-Net

表1 FVC-Net网络各层具体实现
Table 1 Details of FVC-Net in each layer

模块	层数	输入大小	卷积核大小	输出大小
反卷积	Deconv	$\frac{1}{3}P \times \frac{1}{3}P \times 1$	$5 \times 5 \times 1$	$P \times P \times 1$
CA-block (for Layer1_NDVI)	Layer1_NDVI_3x3	$P \times P \times 1$	$3 \times 3 \times 1 \times 16$	$P \times P \times 16$
	Layer1_NDVI_5x5	$P \times P \times 1$	$5 \times 5 \times 1 \times 16$	$P \times P \times 16$
	Layer1_NDVI_7x7	$P \times P \times 1$	$7 \times 7 \times 1 \times 16$	$P \times P \times 16$
	Layer1_NDVI_cat	$P \times P \times 16$	—	$P \times P \times 48$
	Layer1_NDVI_max1	$P \times P \times 48$	$1 \times 1 \times 48 \times 12$	$P \times P \times 12$
	Layer1_NDVI_mean1	$P \times P \times 48$	$1 \times 1 \times 48 \times 12$	$P \times P \times 12$
	Layer1_NDVI_max2	$P \times P \times 12$	$1 \times 1 \times 12 \times 48$	$P \times P \times 48$
	Layer1_NDVI_mean2	$P \times P \times 12$	$1 \times 1 \times 12 \times 48$	$P \times P \times 48$
CA-block (for Layer1_FVC)	Layer1_FVC_3x3	$P \times P \times 1$	$3 \times 3 \times 1 \times 16$	$P \times P \times 16$
	Layer1_FVC_5x5	$P \times P \times 1$	$5 \times 5 \times 1 \times 16$	$P \times P \times 16$
	Layer1_FVC_7x7	$P \times P \times 1$	$7 \times 7 \times 1 \times 16$	$P \times P \times 16$
	Layer1_FVC_cat	$P \times P \times 16$	—	$P \times P \times 48$
	Layer1_FVC_max1	$P \times P \times 48$	$1 \times 1 \times 48 \times 12$	$P \times P \times 12$
	Layer1_FVC_mean1	$P \times P \times 48$	$1 \times 1 \times 48 \times 12$	$P \times P \times 12$
	Layer1_FVC_max2	$P \times P \times 12$	$1 \times 1 \times 12 \times 48$	$P \times P \times 48$
	Layer1_FVC_mean2	$P \times P \times 12$	$1 \times 1 \times 12 \times 48$	$P \times P \times 48$
CA-block (for Layer2_NDVI)	—	$P \times P \times 96$	—	$P \times P \times 96$
CA-block (for Layer2_FVC)	—	$P \times P \times 48$	—	$P \times P \times 48$
CA-block (for Layer3_NDVI)	—	$P \times P \times 144$	—	$P \times P \times 144$
SA-block1	Layer1_3x3	$P \times P \times 144$	$3 \times 3 \times 144 \times 96$	$P \times P \times 96$
	Layer1_5x5	$P \times P \times 96$	$5 \times 5 \times 96 \times 16$	$P \times P \times 16$
	Layer1_7x7	$P \times P \times 112$	$7 \times 7 \times 16 \times 16$	$P \times P \times 16$
	Layer1_max	$P \times P \times 16$	—	$P \times P \times 1$
	Layer1_mean	$P \times P \times 16$	—	$P \times P \times 1$
	Layer1_out	$P \times P \times 16$	—	$P \times P \times 16$
SA-block2	—	$P \times P \times 16$	—	$P \times P \times 16$
SA-block3	—	$P \times P \times 16$	—	$P \times P \times 16$
图像重建	Re-conv1	$P \times P \times 16$	$5 \times 5 \times 16 \times 4$	$P \times P \times 4$
	Re-conv2	$P \times P \times 4$	$5 \times 5 \times 4 \times 1$	$P \times P \times 1$

注: P 为像素大小。

2.2 通道注意力模块

如图2所示,本文使用的通道注意力模块包括两部分:多尺度特征提取和注意力感知。为了提高模块的表达能力,我们将原始数据通过不同大小的卷积运算层获取强度特征。不同维度的卷积可以获取不同尺度下的特征信息,多尺度下的通道注意力模块可以在不同接受域中提取特征,有利于局部结构的维护。已有研究表明增加网络的宽度可以提高模型对特征提取的性能(Szegedy等, 2015)。因此我们设计了具有3个通道的并行结构,

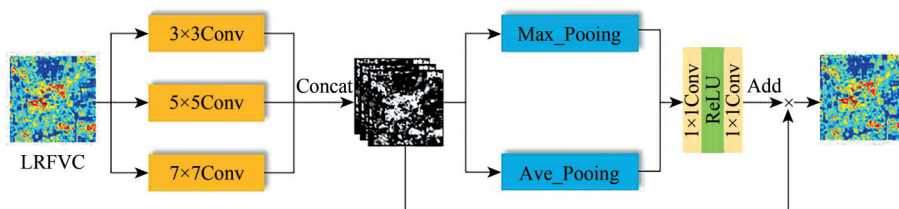


图2 通道注意力模块

Fig. 2 The channel attention block (CA-block)

2.3 空间注意力模块

为获取高质量的高空间分辨率的FVC数据,空间纹理的细节对重建FVC数据至关重要。LRFVC和NDVI经过通道注意力模块处理后,可以获得各自不同尺度下的通道增强特征。与LRFVC相比,NDVI的通道增强特征具有更多的空间细节,将其单向注入各个尺度下的FVC分支中,可以提高FVC通道增强特征的空间质量。为进一步提高FVC的

如表1所示,大小分别为3×3、5×5以及7×7。并行结构不同大小的卷积模块可以在不同的感受野中提取图像特征信息,在较小训练数据的前提下保证网络对数据的有效学习。注意力感知模块是将输入的特征中的信息进行空间转换,保留其中的关键信息,抑制其他无用信息。具体地,注意力感知通过全局最大池化层和全局平均池化层的并行结构,对特征信息进行聚合,将并行结果相加生成通道注意图,最后将多尺度特征提取结果与通道注意图相乘,获得通道增强特征。

空间细节,加强对微小纹理的保存,我们采用了包含空间强化层和空间感知层的空间注意力模块,如图3所示。空间强化层由三层卷积组成,大小为3×3、5×5和7×7。三层卷积线性串联,通过加深结构层次,提取局部空间细节特征。随后,将提取到的空间特征注入由全局最大池化和平均池化组成的并行结构中生成空间注意图,最后将空间特征与空间注意图相乘得到空间增强特征。

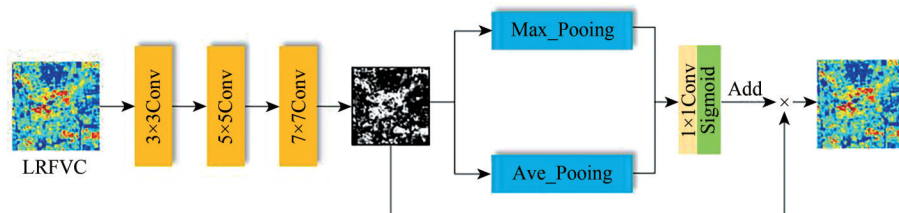


图3 空间注意力模块

Fig. 3 The spatial attention block (SA-block)

2.4 损失函数

为保证融合结果的准确性,强调空间纹理细节,设计了由强度损失描述 L_I 和空间损失描述 L_S 组成的损失函数:

$$L = L_I + \alpha L_S \quad (1)$$

式中, α 为经验参数,用于平衡损失描述(为确定 α 对结果的影响,我们将在实验中比较不同 α 下融合

结果的精度,如表3所示)。 L_I 用于约束重建的FVC图像,使其近似于参考图像的强度分布。 L_I 表示为

$$L_I = \frac{1}{n} |I_x - I_y| \quad (2)$$

式中, n 为图像像素个数, I_x 表示融合结果, I_y 表示标签。 L_I 为第一范式损失,与第二范式相比,第一范式损失可以减少回归带来的平滑效应。

通常,如果仅应用强度损失描述仍会造成局

部空间细节的模糊。为评价融合过程对空间细节的保存,引入了空间损失描述,利用通用图像质量指标 UIQI (Universal Image Quality Index) (Wang 和 Bovik, 2002) 对融合过程加以约束,表示为

$$L_s = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x} \cdot \bar{y}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (3)$$

式中, \bar{x} 、 \bar{y} 表示融合结果和标签的均值; σ_x^2 和 σ_y^2 表示结果和标签的标准差; $\sigma_x \sigma_y$ 表示两者的协方差。利用图像质量指数作为损失描述,可以较好地约束融合结果的结构相似性。

3 实验结果

3.1 实验设计

为评估降尺度的 FVC 数据的精度,基于 Wald 协议的评估方法 (Wald, 2000),我们设计了 30 m 模拟实验 (3.2 节),并通过目视和定量的方式对融合的结果做出评价。在 3.3 节的模拟实验中,分别将 30 m Landsat FVC 和 10 m Sentinel-2 NDVI 数据分别退化至 90 m 和 30 m。通过模拟的 90 m Landsat FVC 与 30 m Sentinel-2 NDVI 融合,得到 30 m Landsat FVC 结果。该结果可参照原始已有 30 m FVC 数据,进行客观评价。为验证网络的有效性以及确定损失函数的经验参数,我们设计了消融实验 (3.3 节) 对网络并行结构、特征层大小排列以及参数设置进行分析。最后,在 3.4 节给出了 10 m 真实实验的结果。

本文选择以上海市为中心的长三角城市群作为研究区。如图 4 所示,中心区域建筑密度较高,主要地物包括居民住房、不透水面 (城市街道)、农田以及河流等。采用的数据包括 Landsat 8 以及 Sentinel-2 数据。其中, Landsat 8 OLI 数据获取时间为 2021 年 2 月 8 日, Sentinel-2 MSI 数据获取时间为 2021 年 3 月 16 日。利用 ENVI 5.3,以 Landsat 8 数据为基准,选取同名像点对两种数据进行几何校正,基于 Landsat 8 数据采用相对植被丰度法获得 30 m FVC 数据。为获得质量较好的数据作为训练数据集,我们选取了覆盖主要地物的 7 个子区域 (如图 4 中红框所示,依次标记为区域 1—7) 作为训练数据,并挑选其中 1 个区域作为测试。需要注意的是,对于任一测试区域,对应区域并不参与训练,即选择剩余 6 个区域用于网络训练。

在定量评价方面,我们采用相关系数 CC (Correlation Coefficient)、通用图像质量指标 (UIQI)、

均方根误差 RMSE (Root Mean Square Error) 以及对全局综合误差 ERGAS (Relative Global-dimensional Synthesis Error) 评价指标。具体地,在模拟实验中,以原始 FVC 数据作为参考图像,对融合生成的 30 m FVC 融合结果进行定量评价。通过与 HPF (Chavez 等, 1991)、Indusion (Khan 等, 2008)、SFIM (Liu, 2000)、ATPRK (Wang 等, 2016)、PanNet (Yang 等, 2017)、PNN (Masi 等, 2016)、HPGAN (Xie 等, 2021) 以及 NDVI-Net (Zhang 等, 2020) 等 8 种方法进行对比,验证提出的 FVC-Net 方法的优势。最后,通过真实 10 m 数据融合实验,比较原始 30 m FVC 数据和降尺度后的 10 m FVC 结果的空间细节差异,展示 FVC-Net 方法的应用潜力。

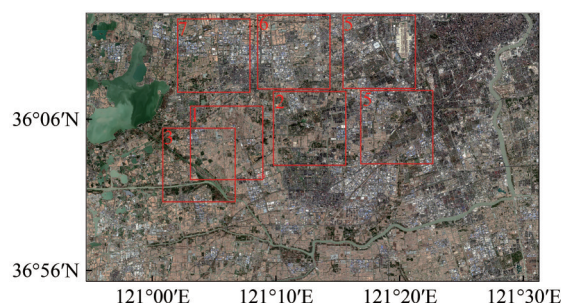


图 4 研究区范围(真彩色合成;共 7 个区域用于训练或测试,即对于任一测试区域,其并不参与训练,剩余 6 个区域用于训练)

Fig. 4 The study area (True color; seven regions were selected for training or test, that is, for each tested region, it was not involved in training, and the other six regions were used for training)

3.2 30 m 模拟数据融合结果

图 5 给出了各种方法的整体结果。可以看到,采用深度学习方法融合的结果,目视上更加接近真实图像。基于传统非学习方法 (如 Indusion 和 HPF 等) 的结果存在明显的平滑问题。为更加直观地比较各种方法结果的差异,局部放大结果如图 6 所示。从局部显示结果可以看到,传统非学习方法存在较为明显的色差,整体上数据与参考数据存在明显误差。这是由于传统方法无法有效顾及不同时相数据之间的地物变化。进一步地,在这些深度学习方法中,提出的 FVC-Net 的效果优于其他方法,其结果最接近于参考数据。从误差图 (以绝对值表示) (图 7) 也可以显示出各种融合方法的性能差异。其中, FVC-Net 结果中细节纹理重建的误差较少,整体误差小于其他方法。

从图8中的散点图可以看出, Indusion 和 HPF 等传统方法存在明显偏差, 且深度学习方法的整体结

果更加接近真实数据, 这与融合结果及误差图反应的结论一致。

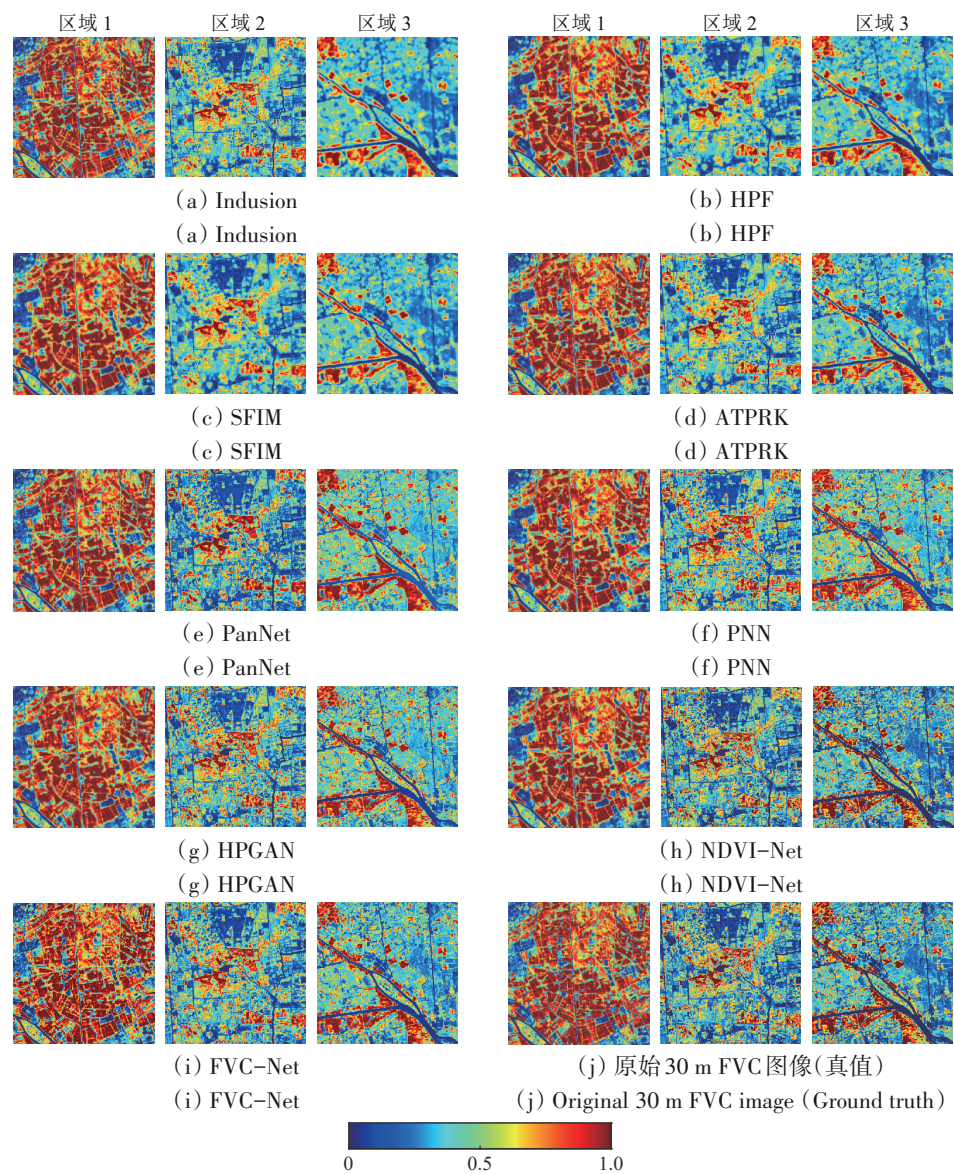


图5 3个区域各种方法的30 m融合结果
Fig. 5 The 30 m fusion results of different methods for the three regions

表2展示了各种融合方法在3个区域上的4项精度指标。从表2中可以得到以下3条主要结论。(1) 在4种传统非机器学习方法中, ATPRK整体精度相对最高。以区域1为例, ATPRK的UIQI高出另外3种非机器学习方法0.05以上, 且RMSE低出0.02以上。在区域2中, ATPRK相较于其他3种非机器学习方法, CC高出约0.02。(2) ATPRK的精度仍低于5种深度学习方法。以区域3为例, 4种深度学习方法相较于ATPRK, CC高出0.04以上。(3) 在4种深度学习方法中, FVC-Net在各项指标下均呈现出了最高精度, 即在所有9种方法中,

FVC-Net精度最高。

3.3 消融实验结果

本节设计了一些消融实验来证明网络的有效性。所有网络均在3.1节介绍的数据中训练得到。除特别说明的参数不同外, 其余训练结构和超参数均相同。表3对3个方面的网络设计进行了对比, 并通过CC和RMSE验证其融合精度。第1列比较了并行网络和串行网络对该融合结果的影响。第2列对比了通道注意力模块以及空间注意力模块特征层的设置, 包括卷积核大小以及排列方式。第3列对比了损失函数设置对精度的影响。

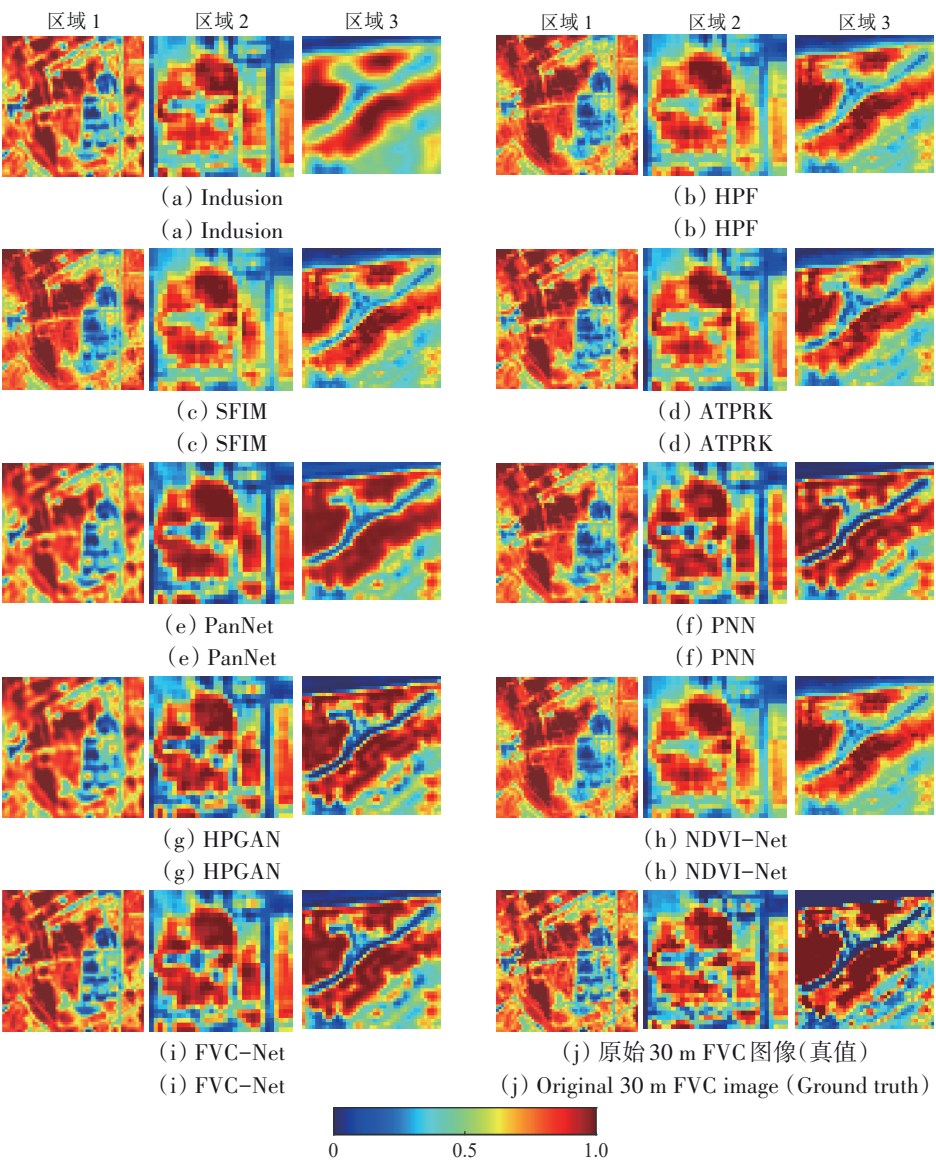
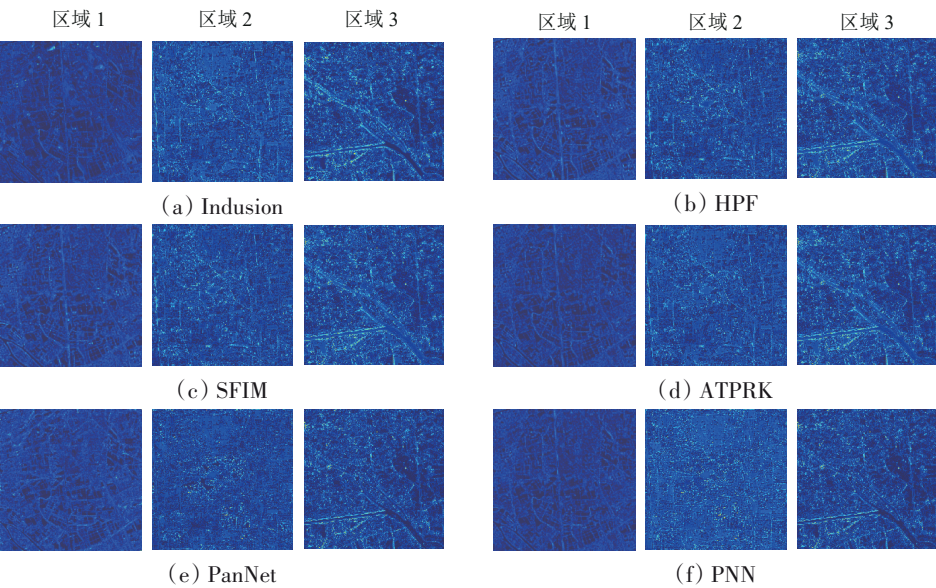


图6 图5中子区域结果

Fig. 6 The results of the sub-area in Fig. 5



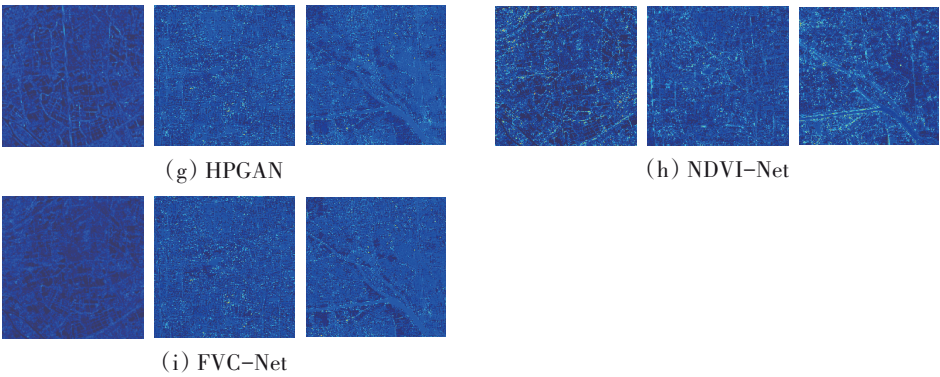
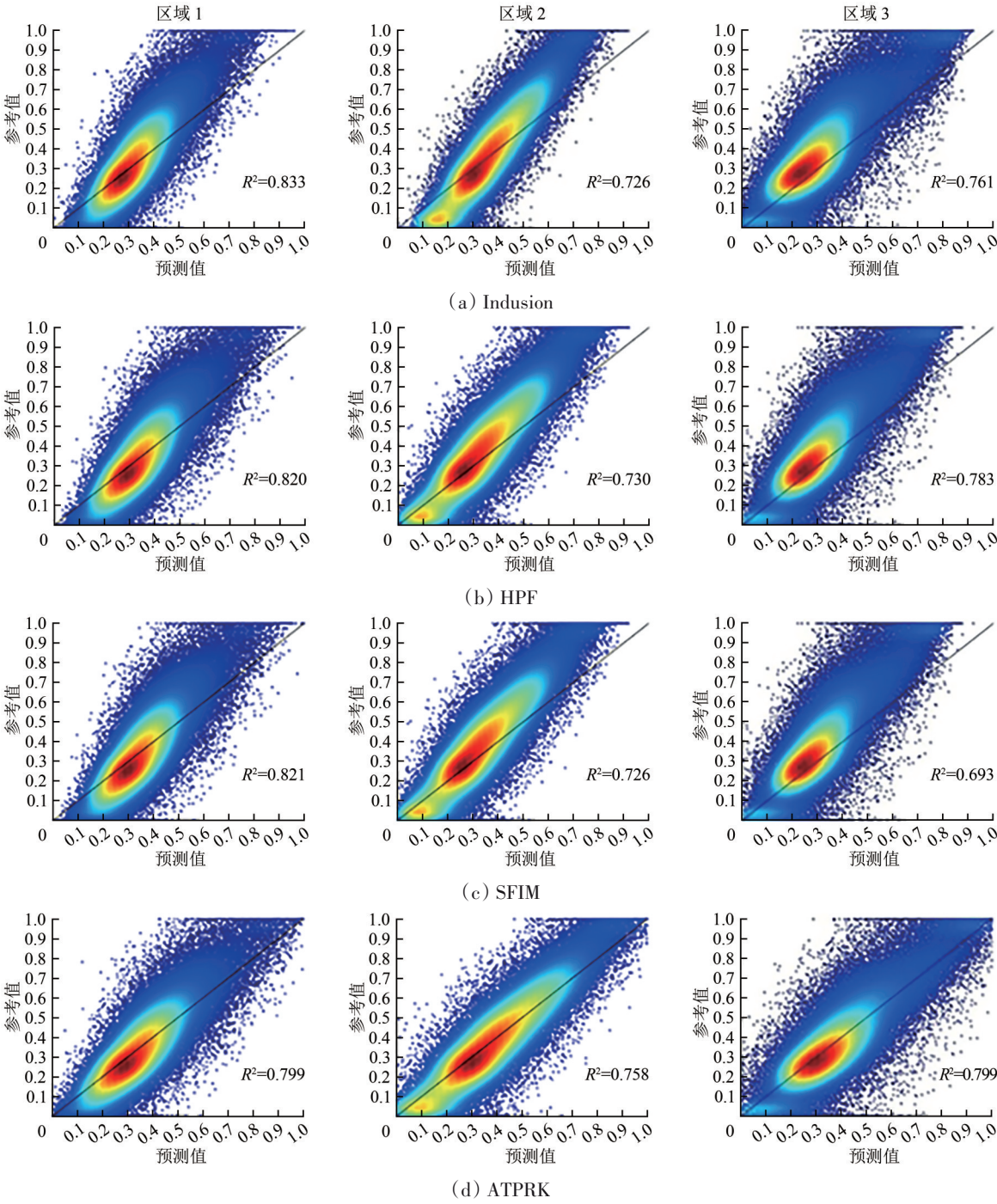


图7 3个区域各种融合方法 30 m 误差结果

Fig. 7 The 30 m error images of different fusion methods for the three regions



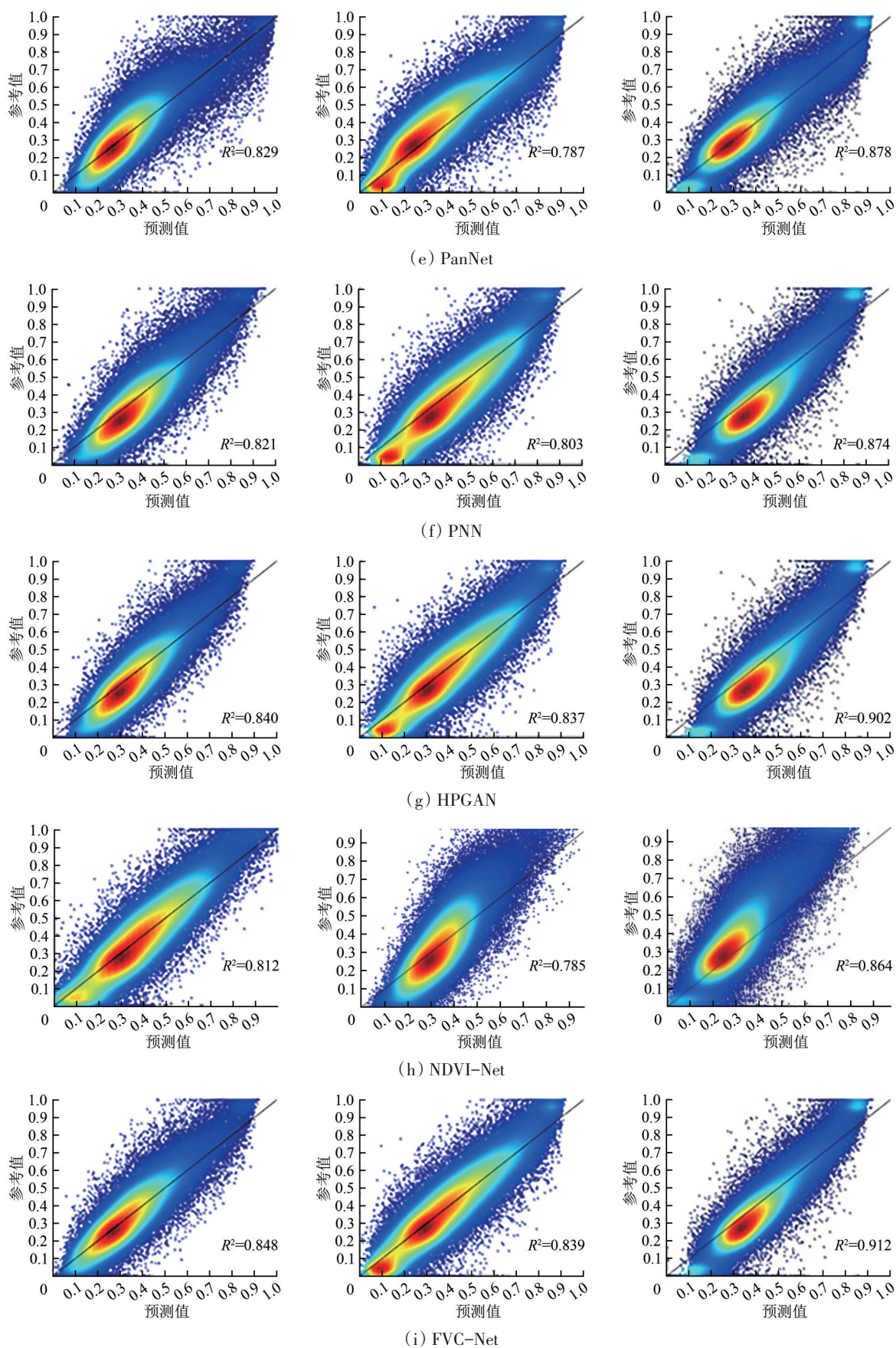


图8 3个区域各种融合方法的散点图

Fig. 8 Scatter plot of different fusion methods for the three regions

表 2 各区域融合精度

Table 2 Fusion accuracy of the three regions

方法		精度指标			
		CC	UIQI	RMSE	ERGAS
区域 1	Indusion	0.9125	0.8110	0.1412	12.0498
	HPF	0.9056	0.8419	0.1255	10.7153
	SFIM	0.9059	0.8434	0.1246	10.6364
	ATPRK	0.8940	0.8936	0.1029	10.9889
	PanNet	0.9107	0.8969	0.1188	10.1376
	PNN	0.9062	0.8572	0.1540	13.1459
	HPGAN	0.9164	0.8793	0.1382	11.7973
	NDVI-Net	0.9011	0.8432	0.1617	14.5876
	FVC-Net	0.9211	0.9044	0.1287	8.78290
区域 2	Indusion	0.8522	0.7642	0.1326	10.8794
	HPF	0.8539	0.7749	0.1322	11.9875
	SFIM	0.8522	0.8126	0.1204	10.8794
	ATPRK	0.8705	0.8377	0.1192	11.2418
	PanNet	0.8874	0.8779	0.1196	11.3580
	PNN	0.8960	0.8551	0.1259	11.3816
	HPGAN	0.9151	0.8469	0.1515	13.6983
	NDVI-Net	0.8859	0.8322	0.1644	14.0039
	FVC-Net	0.9157	0.8831	0.1177	10.6403
区域 3	Indusion	0.8721	0.8422	0.1361	11.2075
	HPF	0.8851	0.8153	0.1402	11.5463
	SFIM	0.8326	0.8025	0.1692	15.0241
	ATPRK	0.8939	0.8804	0.1227	11.1248
	PanNet	0.9369	0.8924	0.1248	10.2810
	PNN	0.9348	0.8642	0.0970	13.5635
	HPGAN	0.9495	0.8935	0.0883	11.9798
	NDVI-Net	0.9296	0.9076	0.0914	14.8064
	FVC-Net	0.9552	0.9085	0.0952	10.1089

注：黑体为各项精度指标中的最优值。

表 3 消融实验定量评价结果

Table 3 Quantitative evaluation results for the ablation experiments

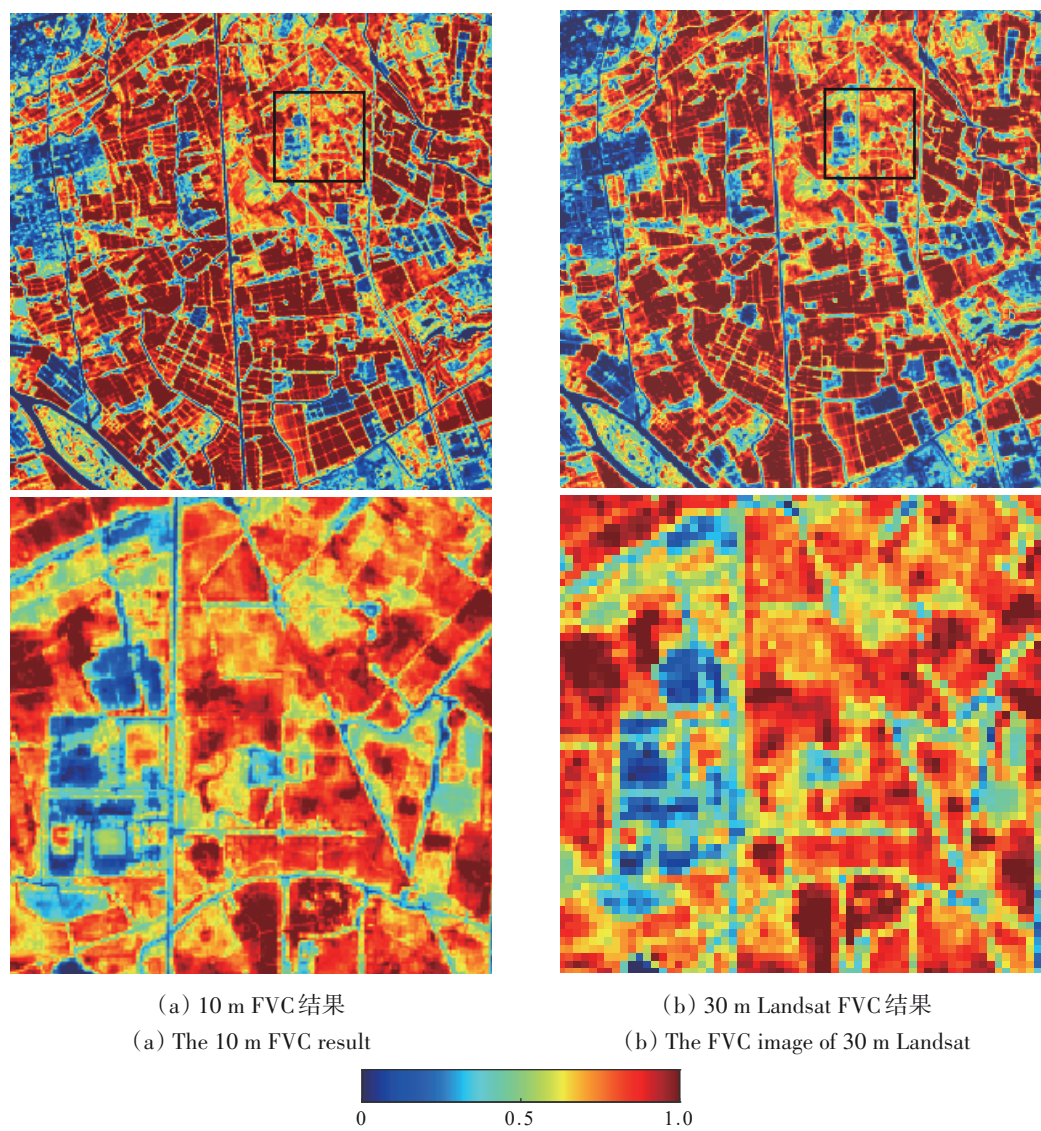
并行结构	特征大小设置	损失函数设置	CC	RMSE
√	3×3、5×5、7×7	$\alpha = 0.2$	0.9211	0.1287
√	7×7、7×7、7×7	$\alpha = 0.2$	0.9055	0.1415
√	5×5、5×5、5×5	$\alpha = 0.2$	0.9047	0.1531
√	3×3、3×3、3×3	$\alpha = 0.2$	0.9024	0.1577
√	5×5、7×7、3×3	$\alpha = 0.2$	0.8676	0.1978
√	3×3、7×7、5×5	$\alpha = 0.2$	0.8411	0.2105
×	3×3、5×5、7×7	$\alpha = 0.2$	0.8045	0.1988
√	3×3、5×5、7×7	$\alpha = 0.6$	0.9100	0.1349
√	3×3、5×5、7×7	$\alpha = 0.8$	0.8966	0.1541

注：黑体为消融实验在不同参数设置下 CC 和 RMSE 的最优值。

为验证本文提出的并行结构的作用，我们与串行结构进行对比，即利用反卷积的形式将FVC和NDVI图像大小统一，随后将两种数据在维度（通道）上进行堆叠，从而实现后续操作的一并处理。实验结果表明，并行结构的设计相比于堆叠处理的形式，具有一定的精度优势。同时，在通道特征提取模块卷积核大小设计和排列中，3×3、5×5和7×7的多尺度特征提取模块的形式可以获得更好的融合结果。

3.4 10 m真实数据融合结果

本文通过训练30 m NDVI与90 m FVC之间的拟合关系，将其迁移至10 m NDVI与30 m FVC的融合过程中，得到10 m FVC融合结果。图9给出了区域1的10 m FVC融合结果。从局部放大图中可以明显看出，融合后的10 m FVC较原始30 m FVC图像更加清晰，呈现了更多的空间细节信息，便于更准确了解各空间位置下的植被覆盖信息。



(a) 10 m FVC结果
(a) The 10 m FVC result

(b) 30 m Landsat FVC结果
(b) The FVC image of 30 m Landsat

图9 10 m融合结果(区域1)
Fig. 9 The 10 m fusion result by FVC-Net (Region 1)

4 结 论

本文提出一种FVC-Net方法，直接融合10 m Sentinel-2 NDVI与30 m Landsat FVC时序数据，将后者降尺度至10 m空间分辨率，解决二者数据空间

分辨率不一致的问题，实现二者在时域上的互补，提高植被生长的精细动态监测能力。FVC-Net网络采用双分支多尺度结果对低空间分辨率FVC以及高空间分辨率NDVI进行特征提取，并通过注意力模块提取NDVI的通道特征单向注入FVC分支以

提高细节表达能力。FVC-Net能有效描述 10 m NDVI 与 30 m FVC 之间的非线性映射关系, 克服 NDVI 与 FVC 本身物理层面的固有差异, 以及不同时间点的 Sentinel-2 与 Landsat 数据中植被覆盖的变化。实验结果表明, FVC-Net 有效提升了 Landsat FVC 数据的空间细节表达能力, 且与 8 种经典方法相比, FVC-Net 从视觉和定量评估角度均呈现出更高精度。在本文研究中, 损失函数较为依赖经验参数的设置, 未来可研究更为可靠的损失函数范式。

参考文献(References)

- Boyd D S, Foody G M and Ripple W J. 2002. Evaluation of approaches for forest cover estimation in the Pacific Northwest, USA, using Remote Sensing. *Applied Geography*, 22(4): 375-392 [DOI: 10.1016/S0143-6228(02)00048-6]
- Chavez Jr P S, Sides S C and Anderson J A. 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and Remote Sensing*, 57(3): 295-303
- Chen Y H and Ge Y. 2022. Spatiotemporal image fusion using multiscale attention-aware two-stream convolutional neural networks. *Science of Remote Sensing*, 6: 100062 [DOI: 10.1016/j.srs.2022.100062]
- Chen Y H, Li X B and Shi P J. 2002. Landscape spatial-temporal pattern analysis on change in the fraction of green vegetation based on remotely sensed data: a case study in Haidian District, Beijing. *Acta Ecologica Sinica*, 22(10): 1581-1586 (陈云浩, 李晓兵, 史培军. 2002. 基于遥感的植被覆盖变化景观分析——以北京海淀区为例. *生态学报*, 22(10): 1581-1586) [DOI: 10.3321/j.issn: 1000-0933.2002.10.002]
- Cui T X, Gong Z N, Zhao W J, Zhao Y L and Lin C. 2013. Research on estimating wetland vegetation abundance based on spectral mixture analysis with different endmember model: a case study in Wild Duck Lake wetland, Beijing. *Acta Ecologica Sinica*, 33(4): 1160-1171 (崔天翔, 宫兆宁, 赵文吉, 赵雅莉, 林川. 2013. 不同端元模型下湿地植被覆盖度的提取方法——以北京市野鸭湖湿地自然保护区为例. *生态学报*, 33(4): 1160-1171) [DOI: 10.5846/stxb201204270604]
- Degerickx J, Roberts D A and Somers B. 2019. Enhancing the performance of Multiple Endmember Spectral Mixture Analysis (MES-MA) for urban land cover mapping using airborne lidar data and band selection. *Remote Sensing of Environment*, 221: 260-273 [DOI: 10.1016/j.rse.2018.11.026]
- Guerschman J P, Hill M J, Renzullo L J, Barrett D J, Marks A S and Botha E J. 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment*, 113(5): 928-945 [DOI: 10.1016/j.rse.2009.01.006]
- Gutman G and Ignatov A. 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, 19(8): 1533-1543 [DOI: 10.1080/014311698215333]
- Hansen M C, Defries R S, Townshend J R G, Sohlberg R, Dimiceli C and Carroll M. 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83(1/2): 303-319 [DOI: 10.1016/S0034-4257(02)00079-2]
- Huang C Q, Song K, Kim S, Townshend J R G, Davis P, Masek J G and Goward S N. 2008. Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112(3): 970-985 [DOI: 10.1016/j.rse.2007.07.023]
- Jia D, Cheng C X, Song C Q, Shen S, Ning L X and Zhang T Y. 2021. A hybrid deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions. *Remote Sensing*, 13(4): 645 [DOI: 10.3390/rs13040645]
- Khan M M, Chanussot J, Condat L and Montanvert A. 2008. Indusion: fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience and Remote Sensing Letters*, 5(1): 98-102 [DOI: 10.1109/LGRS.2007.909934]
- Kuang W H. 2019. Mapping global impervious surface area and green space within urban environments. *Science China Earth Sciences*, 49(7): 1151-1168 (匡文慧. 2019. 全球城市人居环境不透水面与绿地空间特征制图. *中国科学: 地球科学*, 49(7): 1151-1168) [DOI: 10.1007/s11430-018-9342-3]
- Liu J G. 2000. Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18): 3461-3472 [DOI: 10.1080/014311600750037499]
- Masi G, Cozzolino D, Verdoliva L and Scarpa G. 2016. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7): 594 [DOI: 10.3390/rs8070594]
- Ranchin T, Aiazzi B, Alparone L, Baronti S and Wald L. 2003. Image fusion—the ARSIS concept and some successful implementation schemes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1/2): 4-18 [DOI: 10.1016/S0924-2716(03)00013-3]
- Song D X, Wang Z H, He T, Wang H and Liang S L. 2022. Estimation and validation of 30 m fractional vegetation cover over China through integrated use of Landsat 8 and Gaofen 2 data. *Science of Remote Sensing*, 6: 100058 [DOI: 10.1016/J.SRS.2022.100058]
- Su L H. 2009. Optimizing support vector machine learning for semi-arid vegetation mapping by using clustering analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4): 407-413 [DOI: 10.1016/j.isprsjprs.2009.02.002]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]
- Wald L. 2000. Quality of high resolution synthesised images: is there a simple criterion?//*Proceedings of the Third Conference*. Nice, France:

- SEE/URISCA, 99-103
- Wang Q M, Shi W Z and Atkinson P M. 2016. Area-to-point regression kriging for pan-sharpening. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114: 151-165 [DOI: 10.1016/j.isprsjprs.2016.02.006]
- Wang Q M, Tang Y J and Atkinson P M. 2020. The effect of the point spread function on downscaling continua. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168: 251-267 [DOI: 10.1016/j.isprsjprs.2020.08.016]
- Wang Q M, Zhang Z H and Zhang C Y. 2023. Downscaling GF-5 hyperspectral images by fusing with Sentinel-2 images. *National Remote Sensing Bulletin*, 27(8): 1936-1950 (王群明, 张智昊, 张成媛. 2023. 融合 Sentinel-2 数据的高分五号高光谱数据降尺度. *遥感学报*, 27(8): 1936-1950) [DOI: 10.11834/jrs.20211420]
- Xiao J F and Moody A. 2005. A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA. *Remote Sensing of Environment*, 98(2/3): 237-250 [DOI: 10.1016/j.rse.2005.07.011]
- Xie W Y, Cui Y H, Li Y S, Lei J, Du Q and Li J J. 2021. HPGAN: hyperspectral pansharpening using 3-D generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 463-477 [DOI: 10.1109/TGRS.2020.2994238]
- Yang J F, Fu X Y, Hu Y W, Huang Y, Ding X H and Paisley J. 2017. PanNet: a deep network architecture for pan-sharpening//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE: 1753-1761 [DOI: 10.1109/ICCV.2017.193]
- Zeng X B, Rao P, Defries R S and Hansen M C. 2003. Interannual variability and decadal trend of global fractional vegetation cover from 1982 to 2000. *Journal of Applied Meteorology and Climatology*, 42(10): 1525-1530 [DOI: 10.1175/1520-0450(2003)042<1525:IVADTO>2.0.CO;2]
- Zhang H, Ma J Y, Chen C and Tian X. 2020. NDVI-Net: a fusion network for generating high-resolution normalized difference vegetation index in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168: 182-196 [DOI: 10.1016/j.isprsjprs.2020.08.010]
- Wang Z and Bovik A C. 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9(3): 81-4 [DOI: 10.1109/97.995823]

FVC-Net: A fusion network for producing fine spatial resolution fractional vegetation cover

ZHANG Zhihao, WANG Qunming, DING Xinyu

College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

Abstract: Fractional Vegetation Cover (FVC) is defined as the percentage of the vertical projected area of vegetation to the total area of the projected subsurface. It is an important indicator to characterize the spatial distribution of vegetation on the land surface, which plays an essential role in quantifying the capacity of terrestrial ecosystems for carbon sequestration. Remote sensing satellites (such as Landsat and Sentinel-2) can acquire fine spatial resolution FVC data at the 10 m level, which are crucial sources for studies on the global ecosystem. However, a large amount of fine spatial resolution FVC data are unavailable in the temporal domain due to the relatively coarse temporal resolution of satellites, coupled with cloud contamination. This study considers the combination of 30 m Landsat 8 and 10 m Sentinel-2 to increase the temporal frequency of 10 m FVC data to obtain vegetation information timely.

A deep learning-based method called FVC-Net was proposed to address the difference in spatial resolution. FVC-Net fuses 30 m Landsat FVC with the 10 m Sentinel-2 Normalized Difference Vegetation Index (NDVI) directly to produce 10 m Landsat FVC. Specifically, a two-branch network based on the multiscale attention mechanism is designed. In this network, channel enhancement blocks are used in FVC and NDVI branches for feature extraction and fusion. Then, the spatial attention blocks are employed to increase the spatial details of the fused FVC features. The scheme designed with FVC-Net can help to characterize the nonlinear relationship between 10 m NDVI and 30 m FVC.

In the experiments, the proposed FVC-Net method was validated based on three regions selected in the urban area in Shanghai, China. FVC-Net was compared with four typical non-deep learning-based (i.e., HPF, Indusion, SFIM and ATPRK) and four deep learning-based (i.e., PanNet, PNN, HPGAN and NDVI-Net) fusion methods. Both visual and quantitative evaluation reveals that: 1) in non-deep learning-based methods, ATPRK is more accurate than the other three methods; 2) the results of the deep learning-based methods are closer to the reference FVC data; 3) the proposed FVC-Net method outperforms the eight benchmark methods in terms of the fusion accuracy, with results presenting the smallest errors. Finally, the experiments on fusion of real 30 m Landsat FVC and Sentinel-2 10 m NDVI show that the 10 m FVC produced by FVC-Net presents more spatial detailed texture than original 30 m input Landsat FVC.

The proposed FVC-Net is an effective solution to downscale 30 m Landsat FVC to 10 m by fusion with 10 m Sentinel-2 NDVI, thus effectively overcoming the differences between Sentinel-2 and Landsat data at different time points. The generated 10 m FVC data by FVC-Net not only enhance the ability of Landsat data to express spatial details, but also serve as an additional supplement resource to increase the temporal resolution of the existing 10 m Sentinel-2 observations. In future, FVC-Net can be potentially applied to downscale existing 30 m Landsat FVC products at a larger scale (e.g., the global scale), and the predictions may support studies in various related fields.

Key words: remote sensing, Fractional Vegetation Cover (FVC), Normalized Difference Vegetation Index (NDVI), deep learning, downscaling, data fusion

Supported by National Natural Science Foundation of China (No. 42222108, 42171345)