

论文引用格式:

# 遥感图像全色锐化的卷积神经网络方法研究进展

邓良剑, 冉燃<sup>1</sup>, 吴潇, 张添敬

电子科技大学, 四川省成都市 611731

**摘要:** 近年来, 随着计算机科学、遥感科学、大数据科学等领域的迅速发展, 基于卷积神经网络的方法在图像处理、计算机视觉等任务上发挥着越来越重要的作用。而在遥感图像全色锐化领域, 卷积神经网络由于其优秀的融合效果, 已得到研究学者的广泛关注并有大量的研究成果产生。尽管如此, 依然有一些亟待解决的问题, 例如缺乏全色锐化数据集的仿真细节描述、公平公开的训练-测试数据集、简单易懂的统一代码编写框架等。对此, 本文主要从以下几方面回顾当前遥感图像全色锐化问题在卷积神经网络方面的一些进展, 并针对前述问题发布相关数据集和代码编写框架。1) 详细介绍七种典型的基于卷积神经网络的全色锐化方法, 并在统一数据集下进行公平比较(包括与典型传统方法的比较); 2) 详细介绍训练-测试数据集的仿真细节, 并发布相关卫星(如WorldView-3, QuickBird, GaoFen2, WorldView-2)的全色锐化训练-测试数据集; 3) 针对本文介绍的七种基于卷积神经网络的方法, 发布基于Pytorch深度学习库的Python代码统一编写框架, 便于后来初学者的入门、开展研究以及公平比较; 4) 发布统一的全色锐化传统-深度学习方法MATLAB测试软件包, 便于后来学者进行公平的实验测试对比; 5) 对本领域的未来研究方向进行讨论和展望。本文的相关数据集和代码详见课题主页:  
<https://liangjiandeng.github.io/PanCollection.html>

**关键词:** 全色锐化; 卷积神经网络; 典型卷积神经网络方法比较; 数据集发布; 代码框架发布; 进展综述

## A review on CNN-based remote sensing pansharpening

Liang-Jian Deng, Ran Ran, Xiao Wu, Tian-Jing Zhang

University of Electronic Science and Technology of China, Chengdu Sichuan 611731, China

**Abstract:** The goal of pansharpening is to record and process data obtained from various remote sensing satellites using machine learning and signal processing techniques. Its major task is to merge a multispectral (MS) image with low spatial resolution and a panchromatic (PAN) image with high spatial resolution to produce a high-quality multispectral image. In various academic fields, machine learning (ML) is impacting the literature and frequently suggesting state-of-the-art methods. Pansharpening, as an image fusion technique based on the merging of a multispectral image, which is distinguished by its medium/low spatial resolution, has recently been investigated using machine learning. Therefore, machine learning for pansharpening represents a developing research area that merits more study. In recent years, with the rapid development of computer science, remote sensing science, big data science and other fields, convolutional neural networks (CNNs) based techniques, belonging to a category of ML, have become increasingly important role in image processing, computer vision and other tasks. In the field of pansharpening of remote sensing images, due to its excellent fusion effect, CNN has received extensive attention from experts in the field and a large number of research results have been

收稿日期: ; 修回日期:

基金项目:国家自然科学基金项目(项目编号: 12271083, 12171072), 四川省自然科学基金面上项目(项目编号: 2022NSFSC0501); 国家重点研发计划项目(项目编号: 2020YFA0714001), 四川省应用基础研究重点项目(项目编号: 2020YJ0216)

Supported by: NSFC (Grant No. 12271083, 12171072), Natural Science Foundation of Sichuan Province (2022NSFSC0501), National Key Research and Development Program of China (Grant No. 2020YFA0714001), and Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216).

produced, but there are still some problems to be solved, such as the detailed introduction to typical pansharpening CNNs, data simulation, fair and open training-test datasets, simple and easy-to-understand unified code writing framework, etc. In this regard, this paper mainly reviews the current progress in CNNs for remote sensing pansharpening from the following aspects. 1) Detailed introduction of seven typical CNN-based pansharpening methods and fair comparisons (including comparisons with representative traditional methods) under the unified datasets will be given; 2) The simulation of training-test datasets will be introduced in details and we intend to release the pansharpening datasets of related satellites (such as WorldView-3, QuickBird, GaoFen2, World View-2 satellites); 3) For the seven CNN-based methods introduced in this article, Python codes based on Pytorch library will be released in a unified framework to facilitate the complete and fair comparisons for later scholars; 4) We will publish a unified pansharpening MATLAB software package for the test of deep-learning and traditional approaches, which is convenient for later researchers to conduct fair tests; 5) Also, future directions in this field will be discussed. Please find the homepage of this project from the website: <https://liangjiandeng.github.io/PanCollection.html>.

As a conclusion of this work, in situations that are similar to those that were shown during the training phase, CNN-based techniques have proven to have exceptional performance. Instead, when a completely different situation is employed in the testing phase, it has been shown that performance has decreased (particularly in contrast with contemporary state-of-the-art traditional methods), demonstrating a limited ability of these approaches to generalize. The fine-tuning technique, on the other hand, has demonstrated its effectiveness in resolving the aforementioned problem, ensuring great performance even in these difficult test scenarios. Even when compared to the quickest conventional methods, the computing load of the studied CNN-based algorithms, as determined during the testing phase, can be deemed appropriate. Finally, we would like to offer some suggestions for the design of fresh CNN-based pansharpening methods. It is true that the skip connection operation can aid ML-based methods in achieving a faster convergence when focused on the analyzed CNN-based pansharpening approaches. Instead, a stronger feature extraction and learning can be supported by the design of multiscaled architectures (even including the bidirectional structure). Furthermore, the capacity of the networks to generalize can be improved by using a fine-tuning technique and learning in a particular domain (instead of the original image domain). However, there are still certain difficulties that can be resolved, giving future research initiatives room to grow. In particular, as was already mentioned above, the computational weight is still a problem that drives researchers to create networks with fewer parameters (even achieving a quick convergence) while maintaining the network's performance. Furthermore, the majority of the recent advancements in CNN for pansharpening have a restricted capacity for generalization. To deploy machine learning products for remote sensing picture fusion in a commercial setting, it is imperative that this issue be resolved. Finally, the initial vision of using a lower resolution to provide labels for network training is beneficial. It is based on the theory that scales are invariant, hence that cannot be true. As a result, new (unsupervised) techniques based on loss functions evaluating similarities at full resolution have been developed. Even with the requirement for fresh study on more precise quality criteria at full resolution, this research line still needs to be developed further.

**Key words:** pansharpening; convolutional neural networks; comparison of typical CNN methods; datasets releasing; coding-framework releasing; pansharpening survey

# 0 引言

## 0.1 遥感图像全色锐化的概念

遥感卫星搭载的成像传感器旨在采集能够精确记录和反应地物特征的多/高光谱图像数据。提升传感器在光谱域与空间域的响应范围可以获得高分辨率的图像数据，然而受到搭载卫星信噪比、衍射极限等因素的限制，很难获得高分辨率的遥感图像（Zhang 等, 2011）。解决这一问题的一个有效方案就是利用全色锐化技术将特征互补的图像数据进行融合，进而得到高分辨率的遥感图像（Xiao 等, 2020）。

全色锐化（即遥感图像全色锐化，Pansharpening）旨在应用信号处理、机器学习等方法将遥感卫星搭载的不同传感器所记录的同一目标区域的单波段高空间分辨率全色图像（Panchromatic, PAN）与低空间分辨率的多光谱图像（Low-Spatial Resolution Multispectral, LRMS）进行信息融合、优势互补，从而得到优质的高空间分辨率的多光谱图像（High-Spatial Resolution Multispectral, HRMS）（Wang 等, 2021），其具体过程如图 1 所示。



图 1. 全色锐化问题示例（来源：WorldView-3 卫星）。将同一地表场景获取的高空间分辨率全色图像（左）和低空间分辨率的多光谱图像（中）进行融合，最终得到高空间分辨率的融合图像（右）。

Figure 1. Example of pansharpening problem (Source: WorldView-3 satellite). The high spatial resolution panchromatic image (left) and the low spatial resolution multispectral image (middle) from the same scene are fused to finally obtain the high spatial resolution fused image (right).

## 0.2 遥感图像全色锐化的研究意义

空间和光谱分辨率是衡量遥感图像应用价值的两个重要指标。空间分辨率指的是遥感图像中能够识别的两个相邻地物之间的最小距离，而光谱分辨率则是指传感器可以分辨的最短波长间隔。因此 HRMS 图像拥有丰富的空间信息和光谱信息，能够准确地观察和识别到目标地物的大小、形状等几何特征，同时精细地反映地物的内在物理特性。全色锐化技术使得 LRMS 和 PAN 图像的有效信息被充

分利用，突破了单一成像传感器的局限性，极大提高了遥感图像的应用价值。目前世界各国发射的很多卫星都搭载了各种不同的成像传感器，能够同时获得 LRMS 与 PAN 图像，例如美国“WorldView”系列卫星，以及中国“高分”系列卫星。将这些图像进行融合得到 HRMS 图像，能够为后续的分析和应用提供更好的数据支撑，为遥感监测技术提供有利保障。因此，全色锐化技术具有广阔的应用前景和重要的社会经济效益。例如，多光谱遥感技术与军事发展息息相关，这带来了巨大的商业价值。美国军事影像情报部门使用的世界上首颗空间分辨率达到 1 米的遥感卫星是由美国 DigitalGlobe 公司于 1999 年发射的 IKONOS 卫星，成为太空时代的一个重大里程碑。2006 年，我国启动了“高分辨率对地观测系统重大专项”，构建了我国稳定运行的高分辨率对地观测体系，推动了空间信息产业的发展。此外，随着遥感技术的迅猛发展，人类不断地获取到丰富的遥感图像数据，通过全色锐化技术获得的 HRMS 图像在地质勘探、环境监测、精准农业分析和国防安全等领域都发挥着重要作用，是相关产业技术领域所倚重的重要信息技术手段。目前针对全色锐化问题已有大量的研究工作，特别是近年来受到广泛关注的基于卷积神经网络（CNN）的方法，展示了其出色的融合能力（Yang 等, 2021）。然而，这个领域依然存在一些亟待解决的问题，如缺少公开公平的数据集、简单高效的代码编写框架等，从而制约了基于 CNN 的全色锐化方法的发展。综上，本文关注的基于 CNN 的全色锐化问题具有重要的科学的研究和实际应用价值。

## 0.3 当前主流方法

全色锐化研究发展至今，大量关于全色锐化的方法被提出。通常全色锐化方法被大致分为成分替代法（Component Substitution, CS），多分辨率分析法（Multiresolution Analysis, MRA），变分优化方法（Variational Optimization, VO）和机器学习方法（Machine Learning, ML）四大类。本文在此分类的基础上，按照提出时间将前三类归为传统方法，一些优秀综述文献如（Vivone 等, 2015）对传统方法进行了深入的分析。本文将进一步讨论当前一个重要的主流方法—深度学习方法。

### 0.3.1 传统方法

本文将 CS 方法、MRA 方法和基于 VO 的方法视作传统方法。其中 CS 方法主要思想是将 LRMS

图像投影在变换域,用PAN图像的空间细节替换LRMS的空间信息分量,同时尽量保持其原有的光谱信息。代表性方法包括强度-色调-饱和度方法(Carper等,1990;Laben等,2000),主成分分析方法(Chavez等,1989;Shah等,2008;Shettigara,1992),格拉姆-施密特方法(Tu等,2001),自适应成分替换方法(Choi等,2010)等。MRA方法主要思想是通过对LRMS图像进行多分辨率分解,提取其空间成分,用高频细节信息丰富的全色图像进行替换。代表性方法包括拉普拉斯金字塔分解(Burt等,1987)、小波变换方法(Mallat,1989;Nasaton,1995)和轮廓波方法(Do等,2005)等。此外,基于VO的方法利用已知的先验信息构建正则项对模型进行合理约束,通过高效的求解算法得到最终的全色锐化结果。代表性方法包括基于稀疏先验构造正则项的全色锐化模型(Fang等,2013),基于图像的非局部相似性(Buades等,2014)和分片光滑性(Deng等,2018)的正则化模型,这些方法可明显提高模型的光谱与空间维度的保真能力。上述三类方法的优缺点详见表1。

### 3.3.2 基于ML的方法

得益于计算机软硬件技术的发展,基于ML的方法在全色锐化问题上展示出了巨大潜力。主流的ML方法主要包括字典学习(Dictionary Learning)方法(Cheng等,2014;Fang等,2013;Li等,2013;Zhu等,2016)和基于深度学习(Deep Learning,DL)的方法。

字典学习是一种特殊的数据表示方法,主要基于稀疏编码从输入数据中找到稀疏线性表示,形成所谓的字典矩阵和相应系数。但这类方法通常也是基于一些人工先验假设(比如稀疏性)下建立模型,如果假设并不是非常精确,那此类方法就会像传统VO方法一样效果受到限制。

深度学习(Deep Learning,DL)的概念源于人工神经网络的研究,随着DL技术在众多领域取得突破,将深度学习技术应用到全色锐化问题成为一个重要的探索领域。2015年,Huang等人(Huang等,2017)受到稀疏去噪任务的自动编码器方案启发,开创性地将DL技术用于全色锐化问题。2016年,Masi等人(Masi等,2016)搭建并训练了第一个基于CNN的全色锐化网络(称为PNN),该架构主要由三个卷积层组成,其灵感来自于单图超分辨率问题的卷积神经网络SRCNN(Dong等,2015)。PNN方法作为第一个基于CNN的全色锐

化方法获得了当时最优秀的融合结果,同时也启发后续学者沿着此工作不断开展研究,不过由于PNN方法只有三个卷积层,并且没有使用能加速收敛的跳跃连接技术,因此这个方法在效果和收敛速度上相较于后来的方法都是较弱的。同时,Zhong等人(Zhong等,2016)在2016年也提出了一种基于GS变换的全色锐化方法,其利用高效的超分辨率CNN对LRMS图像进行超分辨率融合。得益于CNN强大的非线性拟合和特征表示能力,在这些开创性的方法取得非常优秀的效果之后,这一方向受到了更多研究人员的关注,大量基于DL的全色锐化工作被提出。与PNN简单的网络结构的不同,后续的全色锐化网络结构得到了加深和拓宽,在训练阶段拥有越来越复杂的模型结构。例如,基于残差学习的网络结构能够有效缓解梯度消失和爆炸现象,从而加快网络的收敛,被广泛应用于全色锐化问题(Wei等,2017;Yang等,2017)。这些算法虽然在网络的深度和宽度方面不断加强并且取得较好的效果,但是其还未充分挖掘图像的另一个重要性质—即多尺度性质。随后,多尺度金字塔结构的网络能够获取不同尺度下图像所包含的信息,取得了优越的效果(Yuan等,2018;Zhang等,2022;Zhang等,2019)。

在最近的研究中,无监督学习策略也被引入到全色锐化领域(Liu等,2020;Luo等,2020;Ma等,2020;Qu等,2020)。无监督学习旨在探索没有任何标记数据下的网络潜在特征,这意味着不需要标签数据集进行训练,其效果主要依赖于损失函数的构建。无监督学习是未来研究的重要方向,其优点是可以根据实际数据进行学习,进而得到效果优良、泛化性好的结果;而其弱点在于缺乏足够好的损失函数造成不充分的训练学习,通常效果并不比基于监督学习的方法效果优秀。此外,Liu等人(Liu等,2020)首次利用生成对抗(Generative Adversarial Network,GAN)网络来解决全色锐化问题(称为PSGAN);该方法由一个用于输出HRMS图像的生成器和一个判断图像真实性的鉴别器组成,实验证明了其在真实数据上的有效性。此外,Ma等人(Ma等,2020)提出了一种新的无监督全色锐化方法,可以避免高分辨率图像退化仿真时带来的分辨率损失。而Qu等人(Qu等,2020)提出了一种基于自注意力机制的无监督学习全色锐化技术,实验证明了这类方法具有良好的泛化能力。基于GAN的方法具有很好的全色锐化效果,尤其是对真实数

据效果通常较好；但这类方法有时候会造成“无中生有”的效果，产生一些不真实存在的图像细节。

### 0.3.5 其他方法介绍

除了上述四种主流全色锐化方法外，近期研究

表1：主要方法的优缺点粗略比较（优: ★★★；良: ★★；差: ★）。

Table 1: Comparison of the advantages and drawbacks of the main methods (excellent: ★★★; good: ★★; poor: ★).

	CS 方法	MRA 方法	VO 方法	DL 方法
空间保真	★★	★	★★	★★★
光谱保真	★	★★	★★	★★★
泛化能力	★★	★★	★★★	★
运算速度	★★★	★★★	★	★★
模型可解释性	★★★	★★	★★★	★

人员还从不同角度展开了相关研究。例如基于非负矩阵分解的全色锐化方法（Berne 等, 2010; Kawakami 等, 2011; Yokoya 等, 2011），基于贝叶斯的全色锐化方法（Hardie 等, 2004; Molina 等, 2008; Zhang 等, 2009）等。

此外，一些将传统 VO 方法和 DL 方法结合的混合方法最近成为全色锐化等图像融合领域的一个前景方向（Dian 等, 2020; Feng 等, 2021; Shen 等, 2019; Wu 等, 2020; Wu 等, 2022; Xie 等, 2019），这种结合可以共享不同方法的优点，如深度学习的高精度图像融合效果和传统 VO 方法优秀的数据泛化性。将 DL 方法和 VO 方法组合的技术路线目前主要分为三类：变分优化模型展开策略（Unfolding VO Model）（Feng 等, 2021; Xu 等, 2021; Xie 等, 2019）；即插即用策略（Plug-and-Play，PnP）（Teodoro 等, 2017）和深度嵌入式变分优化模型的策略（Shen 等, 2019; Wu 等, 2020; Wu 等, 2022）。

针对模型展开的方法，Xie 等人（Xie 等, 2019）通过将构建的传统变分优化模型展开为若干个子问题，对每个子问题的求解构建合适的网络近似对应，最后通过端到端的训练获得优秀的高光谱和多光谱图像融合效果。此外，Feng 等人（Feng 等, 2021）首先提出一个基于空间细节分解的两步优化模型，然后在梯度下降的框架下将给定的模型进行展开，进一步构建出相应的端到端 CNN 网络结构。变分优化模型展开成网络的方法综合了优化模型良好泛化性和深度学习方法大数据训练的优点，获得非常优秀的融合结果；不过据作者所了解，因为模型展开的方法涉及到求解优化模型的大量外迭代，从而使这

类模型的算法参数和计算量相较于其他深度学习方法更多。而针对 PnP 的方法，Teodoro 等人（Teodoro 等, 2017）首先假设出隐式的先验正则项，然后通过变量替换等技术将原始带隐式先验的优化问题转化为若干个子问题，其中一个为带隐式先验的去噪子问题，通过其他已有的高效去噪器（如 BM3D）可快速求解此子问题，最后在外迭代框架下不断更新问题的解，最终获得融合图像。这类方法设计巧妙，并且能有效结合传统优化模型和深度学习方法的优点，但是其依然和模型展开方法一样可能会遭遇模型参数和计算量大的缺点。针对深度嵌入式变分优化模型的方法，Shen 等人（Shen 等, 2019）将从 DL 模型学习到的全色锐化结果嵌入到梯度域下的变分优化模型框架中，该策略虽然简单，但在实际应用中效果非常优秀。Wu 等人（Wu 等, 2022）进一步考虑构建刻画深度学习全色锐化图像和潜在高分辨率图像之间距离的深度嵌入项，并设计一个自适应权重算子用于约束这两个图像之间像素级的距离关系；最终的变分优化模型由两个传统的保真项和所提的深度嵌入项构成；这类方法将传统方法和深度学习方法融合在统一的优化模型框架下，进而针对优化模型设计相关算法；这种方法取得了优异的融合效果，并且深度先验可以直接来自于预训练模型，可以不消耗巨大的网络模型参数和计算量。

## 0.4 本文的贡献

本文的主要贡献如下：

1) 从全色锐化发展的角度，对其概念、意义以及国内外研究趋势进行回顾。详细介绍七种典型的基于 CNN 的全色锐化方法，并在统一的数据集、代码框架下进行公平比较。

2) 发布全色锐化训练和测试数据集“PanCollection”，并对数据集的产生细节进行介绍。据作者了解，发布的数据集是国内外第一个可以直接用于深度学习训练和测试的全色锐化数据集。读者可以非常容易地将其应用于深度学习方法。

3) 针对以往全色锐化深度学习代码编写方式不统一，进而可能造成的不公平比较问题，本文将发布一个基于 Pytorch 深度学习库的统一 Python 代码编写框架（上述比较的七种深度学习方法均在此框架下进行）。此代码框架具有统一的数据输入/输出结构，以及简单、高效的代码编写框架。读者只需要将框架中的核心部分更改自己的方法模块，即可迅速获得模型的输出。

4) 发布统一的全色锐化传统-深度学习方法

MATLAB 测试软件包，便于后来学者进行公平的测试。此外，本文将对本领域未来的研究方向进行讨论和展望。

## 0.5 本文的组织架构

本文的组织架构如下。章节 1 将详解介绍七种近年来基于 CNN 的全色锐化方法，并对这些方法进行讨论。章节 2 将详细介绍本文涉及的数据集情况，包括详细的图像拍摄地点、时间，数据集的仿真细节、数量、尺寸等等。在章节 3，本文将针对几种代表性的传统方法和七种深度学习方法进行公平比较，并对这两类方法进行深入讨论。进一步，本文将在章节 4 对全色锐化未来的方向和发展趋势进行探讨。最后，章节 5 将对发布的数据集和代码框架进行详细介绍，并开源在公开网址。

# 1 七种基于 CNN 的全色锐化方法

本文将详细介绍七种具有代表性的基于 CNN 的全色锐化方法，包括网络结构、方法细节和一些相关的讨论。这七种方法分别为 PNN (Masi 等, 2016)，PanNet (Yang 等, 2017)，DiCNN (He 等, 2019)，MSDCNN (Yuan 等, 2018)，BDPN (Zhang 等, 2019)，FusionNet (Deng 等, 2020) 和 LAGConv (Jin 等, 2022)。这些方法都属于监督学习的范畴，因此他们的表现主要依赖于在仿真数据集上的训练。本文将在相同的训练和测试数据集上对它们进行比较。同时值得强调的是本文并未选择无监督的方法或者基于 GAN 的方法来进行比较，因为他们需要和监督学习不一样的数据集，进而无法进行公平的比较。

## 1.1 PNN 方法

作为基于 PNN 的全色锐化开山之作，Masi 等人 (Masi 等, 2016) 在 2016 年首次将一个简单的三层 CNN 结构应用于全色锐化任务（简称 PNN），并获得优秀的融合效果。这个三层结构来自于自然图像单图超分辨率方法 SRCNN (Dong 等, 2015)。针对全色锐化这个新任务，PNN 首先对低分辨率多光谱图像 MS 进行插值上采样到全色图的尺寸。然后将上采样后的图像与 PAN 图沿谱维度进行叠加，形成网络的输入。这个输入已经是高分辨率目标图像的尺寸，因此在网络的中间层不需要使用任何上采样操作。当网络的输入通过 PNN 的三层卷积后，网络获得和多光谱图像 MS 具有相同光谱波段数的

输出。最终网络的输出和高分辨率多光谱图像在  $L_2$  损失函数下进行网络训练，得到网络的训练参数。具体网络结构详见图 2(a)。

尽管 PNN 只是简单利用 SRCNN 中的三层 CNN 网络架构并将其扩展到全色锐化任务，但这个方法的提出在基于深度学习的全色锐化领域扮演着相当重要的角色，已经成为基于深度学习全色锐化方法的一个基准方法。由于 PNN 的主要结构只涉及三个简单的没有任何跳跃连接的卷积层，它的收敛速度相对较慢，不过其网络参数较少。

## 1.2 PanNet 方法

Yang 等人 (Yang 等, 2017) 在 2017 年提出 PanNet 全色锐化网络，其主要思想是将全色锐化任务分为两个目标，即光谱保真和空间细节注入。具体来说，为了光谱保真，PanNet 将具有较好光谱信息的上采样 MS 图直接和空间细节学习网络的输出相加，这样可以有效地将光谱信息直接传播到输出图像。空间细节学习网络主要包含一个预处理卷积层，其可以增加特征通道；然后采用了 4 个 ResNet 块，其可以有效加深网络深度以获得更好的特征提取。PanNet 网络同样采用  $L_2$  损失函数进行网络参数的训练。具体网络结构详见图 2(b)。

通常大多数基于深度学习的全色锐化技术都是在图像域上设计网络，而 PanNet 的设计思想是在高频域上设计网络架构。这种在高频域上的网络设计策略可以获得优秀的数据泛化性，因为从不同传感器获得的图像具有类似的高频信息分布。此外，由于高频细节特征图大部分像素值接近于零，使得映射空间减少，进而网络更加容易训练。

## 1.3 DiCNN 方法

He 等人 (He 等, 2019) 在 2019 年提出一种基于图像细节注入的全色锐化深度学习网络架构，其可有效学习出多光谱图像的细节信息并将其注入到低分辨率多光谱图像，这一过程与传统的滤波方法物理过程相似。作者在文中提出两种基于细节注入的全色锐化模型，它们和 PNN 相似，都是受到 SRCNN 的启发，构建两个三层的卷积网络。但是不同于 PNN，DiCNN 的两个模型都加上了跳跃连接。这种策略有助于缓解梯度爆炸并加速网络的收敛。本文所采用的模型即为 DiCNN 的第一个模型（为了方便，这里称其为 DiCNN 方法）。DiCNN 方法首先将 PAN 图和上采样的低分辨率 MS 图串联在一起，然后进入

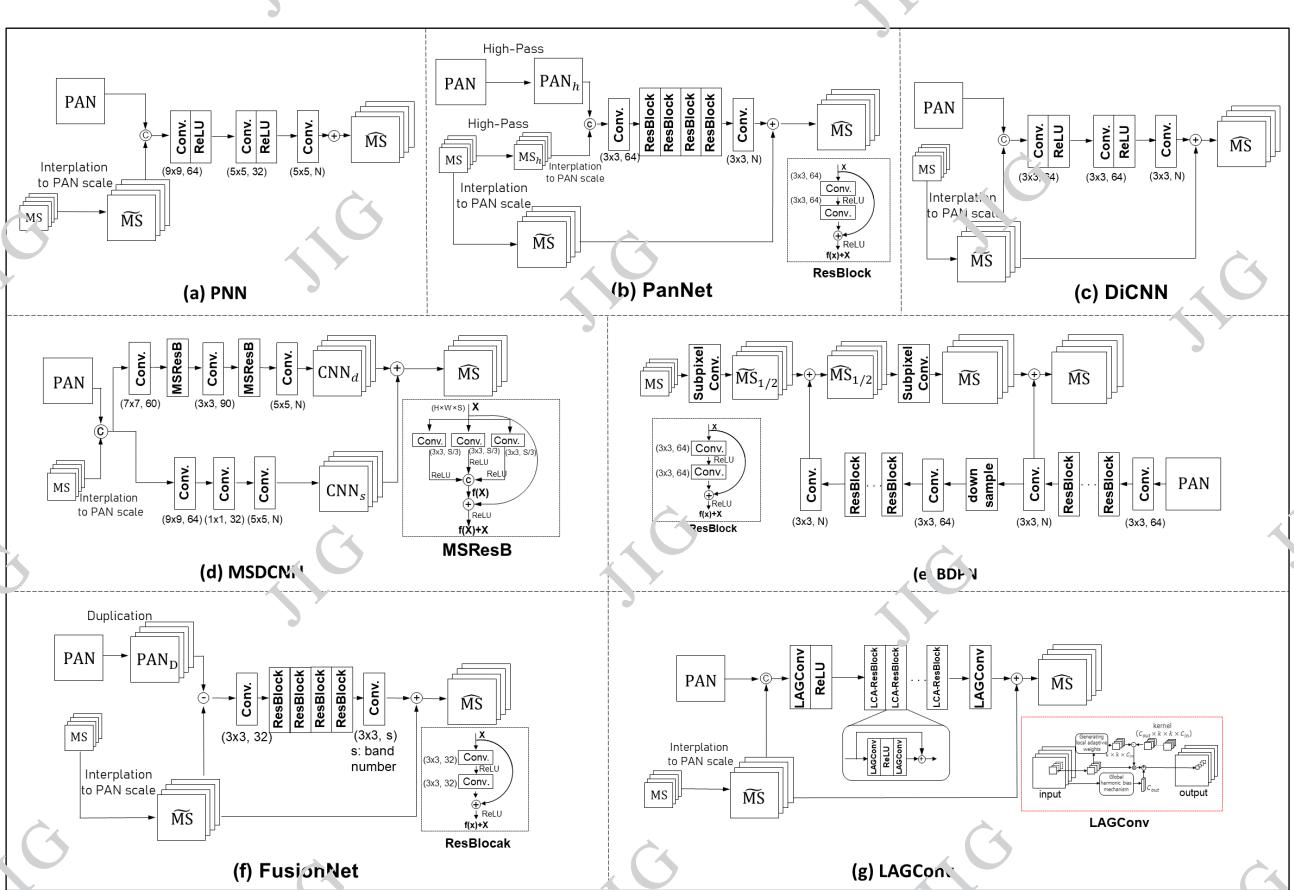


图 2: 本文涉及的七种代表性的基于 CNN 的全色锐化方法, 即为(a) PNN, (b) PanNet, (c) DiCNN, (d) MSDCNN, (e) BDPN, (f) FusionNet 和 (g) LAGConv。

Figure 2: The seven CNN-based pansharpening methods mentioned in this paper, namely (a) PNN, (b) PanNet, (c) DiCNN, (d) MSDCNN, (e) BDPN, (f) FusionNet and (g) LAGConv.

三个简单的卷积层用于学习图像的残差细节, 最后网络的输出直接与上采样的低分辨率 MS 图像相加得到最终的融合图像输出, 并在  $L_2$  损失函数下进行网络的训练。这个方法的结构非常简单、高效, 并且由于其简单的网络结构所涉及的参数量很小。具体网络结构详见图 2(c)。

总的来说, 和其他图像处理任务相似, 跳跃连接的策略对于全色锐化任务非常有效, 其可以有效提高网络训练的收敛速度甚至精度, 也是后来全色锐化深度学习网络最常使用的策略。DiCNN 方法在全色锐化表现优秀, 并且参数量少, 不过也是由于其网络结构简单参数量少, 不能更加有效地提取特征, 因此在融合效果上稍微逊于其他具有更深层网络的方法。

#### 1.4 MSDCNN 方法

Yuan 等人 (Yuan 等, 2018) 在 2018 年提出了 MSDCNN 全色锐化网络结构, 其核心思想就是通过不同大小的卷积核去卷积特征图, 进而提取能表征不同尺度并具有不同感受野的特征, 从而增强网络

的表示能力。在全色锐化问题中, 图像的结构和纹理细节对修复的结果非常重要。而由于遥感图像获取的时间不同, 地物目标的大小变化很大、种类众多等因素, 因此如何有效刻画表示图像的多尺度特征显得尤为重要。针对这个问题, 作者提出了一种包含三个并行卷积层的多尺度模块, 每个并行卷积层的卷积核大小分别为 3、5 和 7。此外, 作者还针对每个多尺度块采用跳跃连接进而形成多尺度残差块(分支一)。另一方面, MSDCNN 还采用具有三个卷积层的浅层网络去提取浅层特征, 这三个卷积层的卷积核大小分别为 9、1 和 5(分支二)。最终将两个分支的输出结果相加得到融合图像。具体网络结构见图 2(d)。

总的来说, MSDCNN 受益于通过不同大小的卷积核获得不同尺度的特征(称为多尺度操作)。通过这种策略, 具有不同感受野的特征图被整合在一起去改进特征提取的精度。此外, 第二个分支结构通过简单的卷积层去灵活学习所谓的浅层特征, 但实际上这个分支效果不明显, 反而会让两个分支过于

灵活，导致深浅层特征学习的不确定性。

### 1.5 BDPN 方法

传统基于 MRA 的方法主要是通过手工提取 PAN 图的多尺度细节信息来改进 MS 图的空间分辨率。而多尺度细节信息提取对全色锐化的质量尤为重要，如果细节提取不足会造成图像模糊，但如果细节提取过多则会导致图像伪影和光谱失真现象。为了更精确地提取多光谱图像中的多尺度细节信息，Zhang 等人在 (Zhang 等, 2019) 中提出一种双方向多尺度学习的全色锐化方法（称为 BDPN）。这种方法根据传统的 MRA 方法思想，设计了一种基于金字塔结构的双方向网络架构去分别处理低分辨 MS 图和高分辨率 PAN 图。通过这个网络，PAN 图的多尺度细节可以有效被提取出来并注入到 MS 图得到高分辨率输出。具体地，整个网络结构在不同尺度间转换，其中提取细节的网络使用了数个经典的残差网络块（ResNet Block），而重构图像的网络采用子像素卷积层（Subpixel Convolutional Layer）去上采样 MS 像。因为网络中间会产生多个不同尺度的输出，所以这种方法采用了多个损失函数监督训练的策略。具体网络结构见图 2(e)。

虽然双方向网络结构的想法之前已在其他多分辨率融合任务上提出（如深度图像超分辨率（Depth Image SR）(Hui 等, 2016)），但是 BDPN 首次将这种思想引入到全色锐化领域，并证明了这种结构的有效性。但是由于其使用的多尺度卷积层太多，导致 BDPN 方法具有大量参数。未来可以研究如何采用更有效的方法来减轻这类方法的参数量。

### 1.6 FusionNet 方法

传统基于 CS 和 MRA 的全色方法取得了非常优秀的效果，然而这些传统方法是在假设线性注入模型的基础上建立的，实际上并不适用于传感器的光谱响应特性。而 CNN 方法具有优秀的非线性表征能力，这激发了研究人员利用非线性技术（如 CNN）来替代传统的线性全色锐化模型的想法。在文献 (Deng 等, 2020) 中 Deng 等人联合 CNN 方法和传统方法（如 CS 和 MRA）的物理模型思想，提出一种新型全色锐化融合网络架构（称为 FusionNet）。其主要策略就是通过深度网络结构去估计关于上采样 MS 图和 PAN 图细节的非线性注入模型。不同于 PanNet 采用高通滤波器提取图像细节，FusionNet 直接用原始的上采样 MS 图与 PAN 图作差提取图像细节，这种方式可以有效保持图像的空间信息和潜在

的光谱信息。所提取的细节被输入到数个残差网络块进行特征提取和细节学习，最后的输出与上采样 MS 图进行相加获得融合图像。FusionNet 依然采用常用的  $L_2$  损失函数进行网络的训练。具体网络结构见图 2(f)。

得益于传统方法启发的 FusionNet 可以在输入和输入之间进行更好更快的数据回归，在训练和测试时获得有竞争力的结果。

### 1.7 LAGConv 方法

前述基于 CNN 的全色锐化方法都是基于传统卷积方式的方法，而基于传统卷积方式的深度模型在空间全局区域都采用一组相同的卷积核来提取特征，这样限制网络特征的表达能力并使得网络难以优化。例如在空间的不同局部区域可能存在不同的目标特征，而整个空间上只采用一组相同的卷积核提取特征显然不是最优的。为了解决这个问题，Jin 和 Zhang 等人在文献 (Jin 等, 2022) 中提出一种能够对局部内容自适应的卷积模块 LAGConv，其主要包括局部自适应卷积核生成和全局偏置机制。局部自适应卷积核的生成方式采用在传统卷积核上乘上可学习的自适应权重矩阵来实现，其主要涉及两个维度规模较小的全连接网络。而全局偏置机制主要是为了补充前面的局部自适应卷积造成的全局信息丢失问题，其实现也是通过两个规模较小的全连接网络来实现。所提的整体网络框架详见图 2(g)，而 LAGConv 模块的细节见图 2(g) 的右下角。

LAGConv 方法能非常高效的提取并表示特征，并且网络的参数也不高，其可以非常快速和有效地拟合训练数据，因此在相应的测试数据上表现优秀。

上述七种方法的优缺点可以在表 2 中找到。除了上述七种代表性的基于 CNN 的全色锐化方法，近期还有很多优秀方法从不同角度被提出，这里不作详细介绍（更多方法详见引言“国内外研究进展”部分）。本文的目的在于针对这些 CNN 方法在统一公开的数据集、代码框架上进行比较，以给读者对于全色锐化问题一个全局的视角，推动本领域的的发展。

表 2. 七种深度学习方法的优缺点总结。

Table 2. Summary of the advantages and drawbacks of the seven deep learning methods.

方法	优点	缺点	其它
PNN	结构简单、参数量少	效果一般、收敛慢、泛化性差	第一个基于 CNN 的全色锐化网络
PanNet	效果好、泛化性好	参数量较多	

化性好、收敛快 结构简单、参数量少、收敛快 充分利用多尺度信息、效果好 充分利用多尺度信息 结构简单、效果好 自适应卷积、效果好	效果一般、泛化性差 参数多、收敛慢、泛化性差 参数多、收敛慢、泛化性差 泛化性差 泛化性差	— — — — 自适应卷积
---	---	---------------------------

## 2 数据集详情

数据集是基于深度学习的方法最重要的因素之一，以往的工作有时数据集各异、细节不明，造成读者对本领域的困惑，也抑制了本领域的健康发展。本文一个重要的目的就是努力讲清楚基于深度学习的全色锐化数据集的来龙去脉，为后续初学者作一个参考。本文将具体从如下方面进行说明。

### 2.1 数据具体信息

本文将在三种卫星传感器获取的数据集上进行 CNN 的训练和测试，这四种数据集的具体信息如下。

1) **WorldView-3 (WV3) 数据集:** 此数据集主要由 WV3 卫星在可见光和近红外光谱范围获取的全色图像和多光谱图像（波段数为 8）数据组成，全色图像和多光谱图像的空间采样间隔（Spatial Sampling Interval, SSI）分别为 0.3 米和 1.2 米，它们之间的空间分辨率比例为 4，辐射测量分辨率（Radiometric Resolution）为 11 位。本数据集主要包含由 WV3 卫星拍摄的两个地方的图像数据，即：WV3 Rio 和 WV3 Tripoli。为了表征不同季相同一地点的地表变化情况，进一步考虑这两个地方不同季节的图像数据。最终获得如图 3 右边的四幅图像用于网络的训练和测试，具体的数据详情（如地点、月份、多光谱图像 MS 的尺寸）可在图中找到。其中，前两幅 WV3 图像数据（即 WV3 Rio（5 月）和 WV3 Tripoli（8 月）右边的 1/4 部分被切割出来制造仿真测试数据集，制造过程采用 Wald 准则（具体细节详见图 4）。此外，真实测试数据集也在与仿真数据集相同的图像上进行，但其制造过程不需要进行下采样，而是直接对图像进行切割获得。

2) **QuickBird (QB) 数据集:** 此数据集主要由

QB 卫星在可见光和近红外光谱范围获取的全色图像和多光谱图像（波段数为 4）数据组成，全色图像和多光谱图像的空间采样间隔 SSI 分别为 0.61 米和 2.44 米，空间分辨率比例为 4，辐射测量分辨率为 11 位。本数据集主要包含由 QB 卫星在 Indianapolis 拍摄的一幅图像数据，即 QB Indianapolis。最终获得如图 3 左边的一幅图像用于网络的训练和测试，具体的数据详情（如地点、月份、多光谱图像的尺寸）可在图中找到。其中，这幅图（即 QB Indianapolis）右边的 1000 个像素被切割出来制造仿真和真实测试数据集，其制造过程与 WV3 数据集相似。

3) **GaoFen-2 (GF2) 数据集:** 此数据集由 GF2（高分 2 号）卫星在可见光和近红外光谱范围获取的全色图像和多光谱图像（波段数为 4）数据组成，全色图像和多光谱图像的空间采样间隔 SSI 分别为 1 米和 4 米，空间分辨率比例为 4，辐射测量分辨率为 10 位。本数据集包含由 GF2 卫星在 Guangzhou 拍摄的一幅图像数据，即 GF2 Guangzhou。最终获得如图 3 中间的一幅图像用于网络的训练和测试，具体的数据详情（如地点、月份、多光谱图像的尺寸）可在图中找到。其中，这幅图（即 GF2 Guangzhou）右边的 1000 个像素被切割出来制造仿真和真实测试数据集，其制造过程依然与 WV3 数据集相似。

### 2.2 数据集仿真

由于上述获得的几种遥感图像数据集缺乏真实的高分辨率多光谱图像，因此需要仿真出训练数据（包括输入数据和标签数据）和测试数据。具体的仿真过程如图 3 所示（以 WV3 数据为例），主要包括数据切割和数据仿真两个步骤。**(a) 数据切割:** 原始获取的全色图（PAN）和低分辨率多光谱图（MS）首先被切割为 N 个尺寸为  $256 \times 256$  和  $64 \times 64 \times 8$  的 PAN 和 MS 图像块。**(b) 数据仿真 (利用 Wald 协议):** 将 (a) 中图像块利用每个卫星特有的调制传递函数（MTF）首先进行滤波操作，然后采用“Nearest”插值下采样获得低分辨率上的全色图像块（pan）和多光谱图像块（ms），进一步利用“23-tap polynomial”插值（Aiazzi 等, 2002）获得上采样的多光谱图像块（lms），下采样前的 MS 图像块作为标签数据(gt)。最终获得如图 4(c) 用于网络训练/验证的数据集。特别地，测试数据的仿真也与此过程相同。

### 2.3 最终的数据集

通过章节 2.2 的数据仿真过程（即图 4），章节 2.1 中的数据（即图 3）被仿真为大量用于网络训练和测试

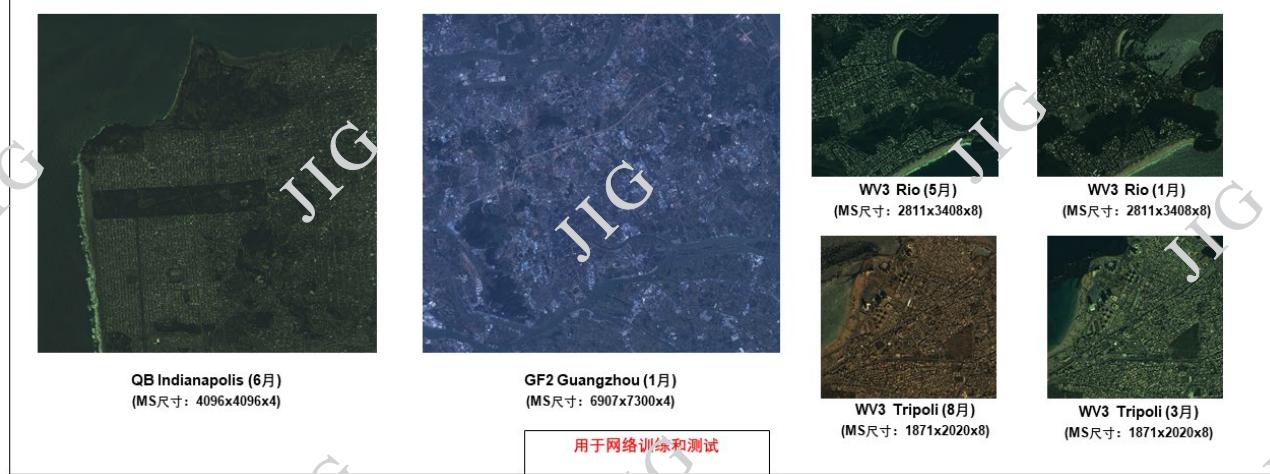


图 3. 本文涉及的网络训练和测试数据集，包括：1) 4 幅位于 Rio 和 Tripoli 并处于不同季节的 WV3 图像数据；2) 1 幅位于 Indianapolis 的 QB 图像数据；3) 1 幅位于 Guangzhou 的 GF2 图像数据；三个图像数据用于网络的训练和测试。所涉及的图像数据包含多种地表特征，如建筑、植被、海洋或河流等，能覆盖大部分卫星观测目标的类型。

Figure 3. The training and testing datasets involved in this paper, include: 1) 4 WV3 images in Rio and Tripoli in different seasons; 2) A QB image in Indianapolis; 3) A GF2 image in Guangzhou; The three images are used for training and testing of the network. The involved image data contain a variety of surface features, such as buildings, vegetation, oceans or rivers, which can cover most of the types of satellite observation targets.

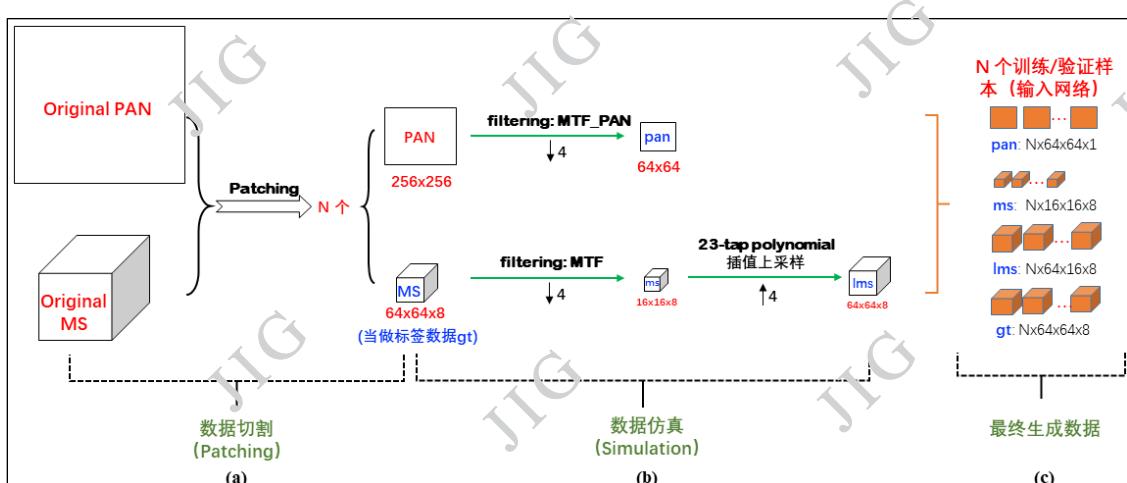


图 4. 数据仿真详细流程（以 WV3 数据为例），主要包括数据切割和数据仿真两个步骤。

Figure 4. Detailed flow of data simulation (with WV3 data as an example), which mainly includes two steps of data cutting and data simulation.

的数据集，最终生成的数据集详情见图 5。通过图 5 可以清晰地了解整个数据仿真的过程。总的来说，本数据集的优点是卫星数据种类多，包含不同波段、不同分辨率、不同地点和地表内容的卫星数据，并详细展示数据获取的地点和时间、哪部分数据用于训练、哪部分用于测试、数据的总样本个数（包括训

练和验证样本）、测试数据（包括仿真数据和真实数据）的选取原则和样本个数。这些细节有益于读者从整体角度了解数据的构造流程。而数据集的难点在于如何通过 MTF（调制传递函数）来仿真数据集，如何将数据集的仿真细节描述清楚，便于读者理解。此外测试数据集选取了不同的场景内容来进行测试，

	数据源	训练数据	测试数据
(A) WorldView-3 (8 bands)	<p><b>WV3 Rio (5月)</b> (MS尺寸: 2811x3408x8)    <b>WV3 Tripoli (3月)</b> (MS尺寸: 1871x2020x8)    <b>WV3 Rio (1月)</b> (MS尺寸: 2811x3408x8)    <b>WV3 Tripoli (3月)</b> (MS尺寸: 1871x2020x8)</p>	<p>总样本数: 10,794 训练集样本数: 9,714 验证集样本数: 1,080</p> <p>(1) 仿真数据集: 选取PAN尺寸为256x256的20个样例进行测试 (2) 真实数据集: 选取PAN尺寸为512x512的20个样例进行测试</p>	
(B) QuickBird (4 bands)	<p><b>QB Indianapolis (6月)</b> (MS尺寸: 4096x4096x4)</p>	<p>总样本数: 19,044 训练集样本数: 17,139 验证集样本数: 1,905</p> <p>(1) 仿真数据集: 选取PAN尺寸为256x256的20个样例进行测试 (2) 真实数据集: 选取PAN尺寸为512x512的20个样例进行测试</p>	
(C) Gaofen-2 (4 bands)	<p><b>GF2 Guangzhou (1月)</b> (MS尺寸: 6907x7300x4)</p>	<p>总样本数: 22,010 训练集样本数: 2,201 验证集样本数: 19,809</p> <p>(1) 仿真数据集: 选取PAN尺寸为256x256的20个样例进行测试 (2) 真实数据集: 选取PAN尺寸为512x512的20个样例进行测试</p>	

图 5. 最终生成的数据详情，包括训练数据集和测试数据集的类型、数量等。

Figure 5. Details of the final generated data, including the type, number, etc. of the training and test datasets.

对深度学习方法来说更具挑战性。特别地，本文将在最后章节 5 给读者提供相关的数据集下载和使用细节。

### 3 实验比较

本节将在上述数据集的基础上对七种典型的基于 CNN 的全色锐化方法进行比较，同时也针对代表性的传统方法进行比较，具体情况如下所述。特别地，本文所涉及的方法比较均在相同的设备和环境下进行，具体为在 CPU 为 i9-10900X, GPU 为 GeForce RTX 3090，操作系统为 Ubuntu 20.04 的设备上进行实验，深度学习方法的训练和测试在函数库 PyTorch 1.9.0 平台下进行，最终所有的方法测试比较在 MATLAB (2022a) 上进行。

#### 3.1 比较方法

本文将在相同的测试数据集上对 13 种分属不同类别的全色锐化方法进行比较，主要包括：

- **EXP 方法：**作为基准的上采样多光谱图，即图 4 中经过“23-tap polynomial”插值上采样的多光谱图。
- **CS 方法**

**BT-H：**带雾霾校正的优化 Brovey 变换方法

(Lolli 等, 2017)。

**BDSD-PC：**带物理约束的波段相关空间细节方法 (Vivone, 2019)。

- **MRA 方法**

**MTF-GLP-FS：**基于 MTF 匹配滤波器的广义拉普拉斯金字塔方法 (Vivone 等, 2018b)。

**MTF-GLP-HPM-R：**基于 MTF 匹配滤波器的广义拉普拉斯金字塔高通调制注入模型方法 (Vivone 等, 2018)。

- **VO 方法**

**TV：**基于全变分模型的全色锐化方法 (Palsson 等, 2014)。

- **DL 方法**

**PNN：**基于三层卷积的 CNN 全色锐化方法 (Masi 等, 2016)。

**PanNet：**基于高频细节学习的 CNN 全色锐化方法 (Yang 等, 2017)。

**DiCNN：**基于细节注入的 CNN 全色锐化方法 (He 等, 2019)。

**MSDCNN：**基于多尺度多深度的 CNN 全色锐化方法 (Yuan 等, 2018)。

**BDPN：**基于双方向网络的全色锐化方法。

**FusionNet：**基于机器学习和传统方法的 CNN 全色锐化方法 (Leng 等, 2020)。

**LAGConv:** 基于局部内容自适应的 CNN 全色锐化方法 (Jin 等, 2022)。

特别地, 传统方法的源代码已经在网站开源:  
<https://openremotesensing.net/knowledgebase/a-critical-comparison-among-pansharpening-algorithms/>, 而 CNN 方法的源代码网址请见本文摘要的课题主页。

### 3.2 全色锐化的量化指标

对全色锐化图像进行量化评估是一个备受争议的问题。通常做法是在 Wald 协议下对锐化结果的一致性和合成性进行评估来衡量结果的优劣。而针对仿真数据和真实数据各有不同的量化指标。

1) 仿真数据的量化指标。由于仿真数据集中包括高分辨的多光谱 (即标签数据 gt), 在这种情况下可以通过三个常用的全色锐化指标对融合图像质量进行评估 (Vivone 等, 2021)。这三个指标分别为 SAM (Spectral Angle Mapper), ERGAS (the Relative Dimensionless Global Error in Synthesis), Q2n (the Multi band Extension of the Universal Image Quality Index) (注: 当为 4 波段数据时使用 Q4, 当为 8 波段数据时使用 Q8)。其中 SAM 和 ERGAS 越小越好 (0 为理想值), 而 Q2n 越大越好 (1 为理想值)。

2) 真实数据的量化指标。由于真实数据集中没有高分辨的多光谱图像 (即标签数据 gt), 在这种情况下可以通过另外三个全色锐化指标对融合图像质量进行评估。这三个指标是用于衡量空间失真的指标  $D_s$ , 用于衡量光谱失真的指标  $D_\lambda$ , 以及无参考质量指标 HQNR。其中  $D_s$  和  $D_\lambda$  越小越好 (0 为理想值), 而 HQNR 越大越好 (1 为理想值)。更多关于指标的详情参见文献 (Arienzo 等, 2022)。

### 3.3 所比较 CNN 方法的参数详情

此小节主要说明 CNN 方法的超参数和其他设置详情, 包括批数量、学习率、优化算法、损失函数等。这些细节可以从表 3 中查询。特别地, 由于本文将所有比较的 CNN 方法统一在 Pytorch 深度框架下, 一些原来在 TensorFlow 框架下的方法由于编程环境的变化 (比如两种框架下的参数初始化不同), 需要对部分超参数进行细调以达到合理效果。注意, 由于本文对不同数据集采用了一组相同的超参数, 因为可能一些深度方法并没有达到最优 (调参并不是本文的主要目的), 后来读者可在本文代码框架下对每类数据集分别调参以实现最优效果。

### 3.4 WV3 数据集实验结果

为验证各类方法在全色锐化中的有效性, 本节将分析各类方法在 WV3 数据集上的表现。针对这七种基于 CNN 的全色锐化方法, 首先将其在 WV3 数据集上训练, 然后在 20 个 WV3 仿真数据集和 20 个真实数据上进行测试, 并与其他当前最优的传统方法进行比较。数据集详情请见第 2 节, 选取的测试数据来自不同的地表特征 (如海岸、植被、建筑等), 具有良好的代表性。

#### 3.4.1 仿真实验结果

本节将评估所有方法在 20 个 WV3 卫星 (拍摄于 Rio 和 Tripoli 地区) 仿真测试数据集上的表现, 这些数据集与用于训练的数据集具有相似的地理区域 (参见图 5 中的数据介绍)。表 4 展示了各类比较方法在 WV3 测试数据集上的定量评价结果。很容易能够观察到, 深度学习方法比传统方法获得了更好的平均指标, 也拥有更小的标准差 (std), 表明深度学习方法相较于传统方法拥有更好的鲁棒性。具体来说, LAGConv 在这些测试数据上明显优于其他比较方法。此外, FusionNet、MSDCNN 和 DiCNN 也拥有很强的竞争力。总的来说, 由于训练数据集与测试数据集具有相似的数据结构和图像特征, 基于深度学习方法的结果比传统方法具有明显的优势。这证实了 CNN 在训练阶段强大的非线性拟合能力和特征提取能力, 因此很容易针对全色锐化任务学习端到端的关系。

为了更直观的进行方法间的视觉比较, 图 6 展示了不同全色锐化方法在 WV3 测试数据上的视觉效果 (从 20 个测试数据中选取一个样例进行显示, 后同)。由图可知, 视觉效果与表 4 中的量化指标表现一致, 基于深度学习的方法对比传统方法拥有更强的光谱保真和空间细节刻画能力。而传统方法 (如 TV) 存在明显的伪影和光谱失真现象。特别地, 深度学习方法中 BDPN 的谱保真能力较差, 对高频边缘区域的重建能力较弱。相反, LAGConv 和 FusionNet 等方法获得的图像很好地保留了光谱信息, 并且能够有效保持图像的纹理等信息。

#### 3.4.2 真实实验结果

除了对仿真数据集进行测试外, 本节将在真实数据集 WV3 Rio 和 WV3 Tripoli 上进行各种方法的性能比较, 数据详情请参见图 5。表 5 和图 7 展示了不同方法取得的定量和视觉结果。很容易观察到, 大多数先进的传统方法都能获得优秀的表现, 甚至

可以将它们与基于深度学习的方法进行比较，证明了其空间和光谱的保真能力。具体来看，传统方法

中 MTF-GLP-HPM-R 方法得到的 HQNR 最高，仅次于

表 3 基于 CNN 方法的最优参数设置。缩略词含义如下：**Epo. #** (epoch number), **Bs** (mini-batch size), **Algo.** (optimization algorithm), **Initi. Lr** (initial learning rate), **Lr Tun.** (tuning strategy of learning rate), **Fs** (filter size for each layer), **Filt. #** (filter number for each layer), **Loss Ty.** (type of loss function), **Ly. #** (number of layers). 这里使用 WV3 数据集作为示例。

Table 3 Optimal parameter settings for CNN-based methods. Abbreviations mean the following: Epo. # (epoch number), Bs (mini-batch size), Algo. (optimization algorithm), Initi. Lr (initial learning rate), Lr Tun. (tuning strategy of learning rate), Fs (filter size for each layer), Filt. # (filter number for each layer), Loss Ty. (type of loss function), Ly. # (number of layers). The WV3 data set is used here as an example.

Para.	PNN	PanNet	DiCNN	MSDCNN	BDPN	FusionNet	LAGConv
<b>Epo. #</b>	12,000	450	5000	500	1.00o	400	1000
<b>Bs</b>	64	32	64	64	8	32	32
<b>Algo</b>	SGD	Adam	Adam	Adam	Adam	Adam	Adam
<b>Initi. Lr</b>	0.0289*bands	0.001	$2 \times 10^{-4}$	$1 \times 10^{-4}$	0.0001	0.0003	0.001
<b>Lr Tun.</b>	fixed initi. Lr(Fl)	FIL	$\times 0.5$ per 200 Epos.	$\times 0.5$ per 60 Epos.	$\times 0.8$ per 100 Epos.	FIL	$\times 0.1$ per 500 Epos.
<b>Fs</b>	$9 \times 9, 5 \times 5$	$3 \times 3$	$3 \times 3$	$3 \times 3$	$3 \times 3$	$3 \times 3$	$3 \times 3, 1 \times 1$
<b>Filt. #</b>	64, 32	64	32	32	64	32	32, 9
<b>Loss Ty.</b>	$\ell_2$	$\ell_2$	$\ell_2$	$\ell_2$	Charbonnier	$\ell_2$	$\ell_2$
<b>Ly. #</b>	3	10	3	12	43	10	7

于深度学习方法中最优越的 PanNet。而对于  $D_\lambda$  指标，MTF-GLP-FS 和 MTF-GLP-HPM-R 都取得了优秀的表现，证明了其空间保真能力强。另一方面，基于深度学习的方法在  $D_s$  指标上的表现普遍优于传统方法，证明了深度学习方法杰出的光谱保真能力。总的来说，基于深度学习的方法在真实数据集上的性能相比仿真数据集显著降低，造成这种现象的可能原因是真实数据与仿真数据的分辨率不一致，训练好的网络参数不能很好地适应测试数据。因此，多数先进的传统方法都能获得优秀的表现，甚至可以将它们与基于深度学习的方法进行比较，证明了其

表 4 WV3 数据集仿真实验量化指标结果。

Table 4 Results of quantitative metrics for simulation experiments on WV3 dataset.

	Q8 ( $\pm$ std)	SAM ( $\pm$ std)	ERGAS ( $\pm$ std)
<b>BT-H</b>	0.8150 $\pm$ 0.1090	4.7744 $\pm$ 1.5429	4.7656 $\pm$ 1.5164
<b>BDSD-PC</b>	0.8115 $\pm$ 0.1106	5.3252 $\pm$ 1.9009	4.9155 $\pm$ 1.6381
<b>MTF-GLP-FS</b>	0.8157 $\pm$ 0.1067	5.1961 $\pm$ 1.8537	4.9014 $\pm$ 1.5878
<b>MTF-GLP-HPM-R</b>	0.8174 $\pm$ 0.1065	5.1993 $\pm$ 1.8732	5.3312 $\pm$ 2.7113
<b>TV</b>	0.7780 $\pm$ 0.1263	5.5780 $\pm$ 1.8898	5.0630 $\pm$ 1.4475

PNN	0.8775 $\pm$ 0.048	3.5816 $\pm$ 0.8754	2.7731 $\pm$ 0.6924
PanNet	0.8733 $\pm$ 0.1096	3.5287 $\pm$ 0.8534	2.7747 $\pm$ 0.7509
DiCNN	0.8846 $\pm$ 0.1033	3.4936 $\pm$ 0.8751	2.7748 $\pm$ 0.7192
MSDCNN	0.8743 $\pm$ 0.1046	3.6785 $\pm$ 0.9010	2.8647 $\pm$ 0.7463
BDPN	0.8528 $\pm$ 0.1120	4.0448 $\pm$ 0.9067	3.1283 $\pm$ 0.7581
FusionNet	0.8884 $\pm$ 0.1061	3.2338 $\pm$ 0.7998	2.5683 $\pm$ 0.7048
LAGConv	0.8942 $\pm$ 0.1049	3.0266 $\pm$ 0.6609	2.3732 $\pm$ 0.6332

空间和光谱的保真能力。具体来看，传统方法中 MTF-GLP-HPM-R 方法得到的 HQNR 最高，仅次于深度学习方法中最优越的 PanNet。而对于 指标，MTF-GLP-FS 和 MTF-GLP-HPM-R 都取得了优秀的表现，证明了其空间保真能力强。另一方面，基于深度学习的方法在 指标上的表现普遍优于传统方法，证明了深度学习方法杰出的光谱保真能力。总的来说，基于深度学习的方法在真实数据集上的性能相比仿真数据集显著降低，造成这种现象的可能原因是真实数据与仿真数据的分辨率不一致，训练好的网络参数不能很好地适应测试数据。因此，多数先进的传统方法都能获得优秀的表现，甚至可以将它们与基于深度学习的方法进行比较，证明了其空间和光谱的保真能力。具体来看，传统方法中 MTF-

GLP-HPM-R 方法得到的 HQNR 最高, 仅次于深度学习方法中最优越的 PanNet。而对于  $D_\lambda$  指标, MTF-GLP-FS 和 MTF-GLP-HPM-R 都取得了优秀的表现, 证明了其空间保真能力强。另一方面, 基于深度学习的方法在  $D_s$  指标上的表现普遍优于传统方法, 证明了深度学习方法杰出的光谱保真能力。总的来说, 基于深度学习的方法在真实数据集上的性能相比仿真数据集显著降低, 造成这种现象的可能原因是真实数据与仿真数据的分辨率不一致, 训练好的网络参数不能很好地适应测试数据。因此, 如何提高深度学习方法的泛化能力是未来亟待解决的问题。

### 3.5 QB 数据集实验结果

与前面的实验设计相似, 本节主要测试不同方法在 QB 数据集上的全色锐化表现。测试数据集包括含有代表性地物目标的 20 个仿真数据和 20 个真实数据。

表 5 WV3 数据集真实实验量化指标结果。

Table 5 Results of real experimental quantitative metrics for the WV3 dataset.

	$D_\lambda$ ( $\pm$ std)	$D_s$ ( $\pm$ std)	HQNR ( $\pm$ std)
<b>BT-H</b>	0.1966 $\pm$ 0.1369	0.1855 $\pm$ 0.1551	0.6722 $\pm$ 0.2481
<b>BDSD-PC</b>	0.1556 $\pm$ 0.1219	0.1647 $\pm$ 0.1317	0.7186 $\pm$ 0.1988
<b>MTF-GLP-FS</b>	0.0741 $\pm$ 0.0640	0.1699 $\pm$ 0.1528	0.7770 $\pm$ 0.1865
<b>MTF-GLP-HPM-R</b>	0.0743 $\pm$ 0.0637	0.1689 $\pm$ 0.1535	0.7777 $\pm$ 0.1865
<b>TV</b>	0.1414 $\pm$ 0.1399	0.1701 $\pm$ 0.1717	0.7349 $\pm$ 0.2499
<b>PNN</b>	0.1593 $\pm$ 0.1446	0.0791 $\pm$ 0.0603	0.7811 $\pm$ 0.1726
<b>PanNet</b>	0.0851 $\pm$ 0.0899	0.1155 $\pm$ 0.1149	0.8188 $\pm$ 0.1737
<b>DiCNN</b>	0.1066 $\pm$ 0.1011	0.0798 $\pm$ 0.0758	0.8279 $\pm$ 0.1457
<b>MSDCNN</b>	0.1281 $\pm$ 0.1245	0.0755 $\pm$ 0.0457	0.8112 $\pm$ 0.1515
<b>BDPN</b>	0.2233 $\pm$ 0.2233	0.1400 $\pm$ 0.1343	0.6950 $\pm$ 0.2761
<b>FusionNet</b>	0.1169 $\pm$ 0.1095	0.1146 $\pm$ 0.1280	0.7925 $\pm$ 0.1899
<b>LAGConv</b>	0.1815 $\pm$ 0.1710	0.0763 $\pm$ 0.0569	0.7637 $\pm$ 0.1939

#### 3.5.1 仿真实验结果

本节主要展示所有方法在 QB 仿真数据集上的表现。由表 6 可知所有基于 CNN 的深度学习方法均表现出优秀的定量效果, 其效果大幅度领先于传统方法。这主要是由于深度学习方法针对大量的 QB 数据进行训练, 而测试的数据与训练数据类型相似, 则学习出的模型能有效融合测试数据。其中, 基于自适应卷积的 LAGConv 方法表现最为亮眼, 这也

是由于其优秀的局部特征提取能力以及数据拟合能力。而经典的 PNN 方法的量化指标甚至超过 PanNet 方法, 这可能是因为刚好在这个数据集上, PNN 相较于 PanNet 能更好地拟合训练数据。传统方法中 MTF-GLP-HPM-R 在 Q4 和 SAM 指标上表现最好, 但是 ERGAS 弱于其他方法, 显示其相较于其他传统方法虽然光谱保真能力较强, 但是空间保真能力较弱。图 8 展示不同方法的仿真测试视觉效果, 其结果基本是定量指标一致。

表 6 QB 数据集仿真实验量化指标结果。  
Table 6 Results of quantitative metrics for the QB dataset simulation experiments.

	Q4 ( $\pm$ std)	SAM ( $\pm$ std)	ERGAS ( $\pm$ std)
<b>BT-H</b>	0.8308 $\pm$ 0.0926	7.2112 $\pm$ 1.5117	7.4567 $\pm$ 0.7380
<b>BDSD-PC</b>	0.8284 $\pm$ 0.0936	8.1025 $\pm$ 1.9410	7.5668 $\pm$ 0.7071
<b>SR-D</b>	0.8276 $\pm$ 0.0945	7.4305 $\pm$ 1.6952	7.5415 $\pm$ 0.8135
<b>MTF-GLP-FS</b>	0.8335 $\pm$ 0.0914	7.8010 $\pm$ 1.7755	7.4135 $\pm$ 0.6601
<b>MTF-GLP-HPM-R</b>	0.8404 $\pm$ 0.0980	7.7808 $\pm$ 1.7449	9.7803 $\pm$ 8.1258
<b>TV</b>	0.8197 $\pm$ 0.0873	7.5266 $\pm$ 1.6195	7.7492 $\pm$ 0.8110
<b>PNN</b>	0.9158 $\pm$ 0.0975	5.2135 $\pm$ 0.9343	4.5004 $\pm$ 0.3602
<b>PanNet</b>	0.8830 $\pm$ 0.0952	5.8013 $\pm$ 1.1600	5.9076 $\pm$ 0.8512
<b>DiCNN</b>	0.9011 $\pm$ 0.0991	5.3862 $\pm$ 0.9999	5.1764 $\pm$ 0.4438
<b>MSDCNN</b>	0.9188 $\pm$ 0.0966	5.1471 $\pm$ 0.9342	4.3828 $\pm$ 0.3400
<b>BDPN</b>	0.9005 $\pm$ 0.0923	6.1225 $\pm$ 1.2106	5.2756 $\pm$ 0.6870
<b>FusionNet</b>	0.9231 $\pm$ 0.0943	4.9279 $\pm$ 0.8878	4.1754 $\pm$ 0.3017
<b>LAGConv</b>	0.9314 $\pm$ 0.0917	4.5548 $\pm$ 0.8155	3.8436 $\pm$ 0.4032

#### 3.5.2 真实实验结果

本小节将所有方法在 QB 真实数据集上的实验结果进行了量化指标对比, 参见表 7。从表中可以清晰地看出, 一些基于 CNN 的深度学习方法(如 PanNet、DiCNN 和 FusionNet)都取得了很高的 HQNR 量化表现。此外, 大多数基于深度学习的方法比其他传统技术(如 BT-H 和 BDSD-PC)获得更好的指标。在所有基于深度学习的方法中, PanNet、DiCNN 和 FusionNet 都属于最佳性能一类。特别地, PanNet 在所有指标上都取得最优表现, 原因与 WV3 真实数据集的一致。而相比于在仿真实验中的表现, LAGConv 在真实数据上的表现并不出色, 这是由于对训练数据集的强大拟合性能导致模型在性质不同的数据集上出现不匹配的现象, 这也说明了基于 CNN 的模型泛化性与复杂度之间难以平衡。

表 7 QB 数据集真实实验量化指标结果。

Table 7 Results of real experimental quantitative metrics for the QB dataset.

	$D_\lambda$ ( $\pm$ std)	$D_s$ ( $\pm$ std)	HQNR ( $\pm$ std)
<b>BT-H</b>	0.2788 $\pm$ 0.1323	0.1835 $\pm$ 0.0861	0.5965 $\pm$ 0.1535
<b>BDSD-PC</b>	0.2245 $\pm$ 0.0588	0.1789 $\pm$ 0.1051	0.6420 $\pm$ 0.1222
<b>MTF-GLP-FS</b>	0.0719 $\pm$ 0.0353	0.1558 $\pm$ 0.0800	0.7851 $\pm$ 0.0922
<b>MTF-GLP-HPM-R</b>	0.0674 $\pm$ 0.0295	0.1708 $\pm$ 0.0783	0.7751 $\pm$ 0.0903
<b>TV</b>	0.0797 $\pm$ 0.0344	0.1337 $\pm$ 0.0835	0.7997 $\pm$ 0.1026
<b>PNN</b>	0.1085 $\pm$ 0.0373	0.1100 $\pm$ 0.0791	0.7960 $\pm$ 0.1003
<b>PanNet</b>	0.0612 $\pm$ 0.0329	0.1031 $\pm$ 0.0805	0.8443 $\pm$ 0.1004
<b>DiCNN</b>	0.0717 $\pm$ 0.0238	0.1015 $\pm$ 0.0783	0.8350 $\pm$ 0.0847
<b>MSDCNN</b>	0.0832 $\pm$ 0.0340	0.094 $\pm$ 0.0902	0.8189 $\pm$ 0.1063
<b>BDPN</b>	0.1415 $\pm$ 0.0911	0.0761 $\pm$ 0.0343	0.7952 $\pm$ 0.1032
<b>FusionNet</b>	0.0917 $\pm$ 0.0498	0.0868 $\pm$ 0.0482	0.8314 $\pm$ 0.0820
<b>LAGConv</b>	0.1214 $\pm$ 0.0548	0.0763 $\pm$ 0.0216	0.8118 $\pm$ 0.0564

### 3.6 GF2 数据集实验结果

为了更加全面的评估所有方法的性能，本节在 4 波段的 GF2 数据集上进行实验，与之前的实验类似，分别研究了在仿真数据集和真实数据集上的性能表现。

#### 3.6.1 仿真实验结果

本节主要展示所有方法在 GF2 仿真数据集上的表现。由表 8 可知，与上一节的结果相似，所有基于 CNN 的深度学习方法的量化效果都能够超越传统方法。值得注意的是，传统方法中的 BT-H 有着极具竞争力的表现，与基于大数据训练的 BDPN 方法十分相近。同时，基于 CNN 的深度学习方法中除 LAGConv 拥有明显的优势之外，基于高频细节特征提取的 PanNet 和结合传统方法的 FusionNet 在所有指标上都有优越的表现。为了进行视觉比较，图 9 中展示了所有方法的融合结果。从融合后的高分辨率图像可以看出，所有基于 CNN 的深度学习方法都能较好地保留图像内容的纹理和其他空间细节，这表明了这些方法的有效性。

表 8 GF2 数据集仿真实验量化指标结果。

Table 8 Results of quantitative metrics for the GF2 data set simulation experiments.

	Q4 ( $\pm$ std)	SAM ( $\pm$ std)	ERGAS ( $\pm$ std)
<b>BT-H</b>	0.9156 $\pm$ 0.0250	1.6531 $\pm$ 0.3479	1.5300 $\pm$ 0.3957
<b>BDSD-PC</b>	0.8890 $\pm$ 0.0307	1.6886 $\pm$ 0.3464	1.6693 $\pm$ 0.4312

<b>MTF-GLP-FS</b>	0.8938 $\pm$ 0.0305	1.6616 $\pm$ 0.3730	1.5964 $\pm$ 0.3848
<b>MTF-GLP-HPM-R</b>	0.8969 $\pm$ 0.0298	1.6559 $\pm$ 0.3801	1.5941 $\pm$ 0.3941
<b>TV</b>	0.9050 $\pm$ 0.0291	1.9127 $\pm$ 0.4305	1.7384 $\pm$ 0.4327
<b>PNN</b>	0.9592 $\pm$ 0.0098	1.091 $\pm$ 0.2196	1.0592 $\pm$ 0.2270
<b>PanNet</b>	0.9660 $\pm$ 0.0101	0.9984 $\pm$ 0.2064	0.9223 $\pm$ 0.1854
<b>DiCNN</b>	0.9582 $\pm$ 0.0099	1.0536 $\pm$ 0.2237	1.0821 $\pm$ 0.2442
<b>MSDCNN</b>	0.9601 $\pm$ 0.0106	1.0477 $\pm$ 0.2139	1.0423 $\pm$ 0.2217
<b>BDPN</b>	0.9242 $\pm$ 0.0217	1.4642 $\pm$ 0.3133	1.5224 $\pm$ 0.4118
<b>FusionNet</b>	0.9628 $\pm$ 0.0095	0.9742 $\pm$ 0.2051	0.9892 $\pm$ 0.2138
<b>LAGConv</b>	0.9796 $\pm$ 0.0093	0.7888 $\pm$ 0.1436	0.6916 $\pm$ 0.1102

#### 3.6.2 真实实验结果

与在 QB 数据集上的实验流程一致，本节同样在 GF2 的真实数据集上进行了方法测试和对比。表 9 展示了所有方法的定量指标。对表 9 中的结果进行分析，PanNet 远优于其他对比方法。总的来说，传统的方法（除了 MTF-GLP-HPM-R 和 MTF-GLP-FS）获得的性能远远低于大多数基于 CNN 的深度学习方法。所有深度学习的方法中，通过 FusionNet 方法得到的 HQNR 是最低，表明 FusionNet 学到的模型不能适应真实数据。

表 9 GF2 数据集真实实验量化指标结果。

Table 9 Results of quantitative indicators of real experiments for the GF2 dataset.

	$D_\lambda$ ( $\pm$ std)	$D_s$ ( $\pm$ std)	HQNR ( $\pm$ std)
<b>BT-H</b>	0.0891 $\pm$ 0.0335	0.1712 $\pm$ 0.0388	0.7558 $\pm$ 0.0517
<b>BDSD-PC</b>	0.092 $\pm$ 0.0292	0.1652 $\pm$ 0.0362	0.7582 $\pm$ 0.0509
<b>MTF-GLP-FS</b>	0.0370 $\pm$ 0.0138	0.1539 $\pm$ 0.0351	0.8150 $\pm$ 0.0404
<b>MTF-GLP-HPM-R</b>	0.0364 $\pm$ 0.0131	0.1531 $\pm$ 0.0353	0.8163 $\pm$ 0.0400
<b>TV</b>	0.0702 $\pm$ 0.0299	0.1448 $\pm$ 0.0378	0.7952 $\pm$ 0.0444
<b>PNN</b>	0.0496 $\pm$ 0.0171	0.1329 $\pm$ 0.0340	0.8245 $\pm$ 0.0430
<b>PanNet</b>	0.0152 $\pm$ 0.0053	0.0746 $\pm$ 0.0221	0.9113 $\pm$ 0.0223
<b>DiCNN</b>	0.0212 $\pm$ 0.0070	0.0884 $\pm$ 0.0235	0.8924 $\pm$ 0.0233
<b>MSDCNN</b>	0.0289 $\pm$ 0.0093	0.0890 $\pm$ 0.0195	0.8846 $\pm$ 0.0190
<b>BDPN</b>	0.0372 $\pm$ 0.0133	0.0987 $\pm$ 0.0289	0.8678 $\pm$ 0.0326
<b>FusionNet</b>	0.0519 $\pm$ 0.0170	0.1306 $\pm$ 0.0315	0.8247 $\pm$ 0.0405
<b>LAGConv</b>	0.0363 $\pm$ 0.0121	0.1057 $\pm$ 0.0319	0.8621 $\pm$ 0.0372

### 3.7 实验讨论

本节将进一步对前述的比较方法进行实验讨论，

主要包括测试时间比较，以及不同深度学习方法的训练时间、参数量以及计算量 GFlops 的比较。

### 3.8.1 测试时间比较

表 10 展示了所有比较方法的测试时间对比，由表可知传统方法（除了 TV）表现出非常少的测试时间，证明其方法的有效性。基于深度学习的方法也

具有较少的测试时间。所有方法中只有 TV 方法测试时间最多，因为其需要迭代求解一个最优化模型，在迭代的过程中耗费了大量的计算时间。

### 3.8.2 训练时间-参数量-计算量比较

表 11 展示了七种 CNN 方法在 WV3 训练数据

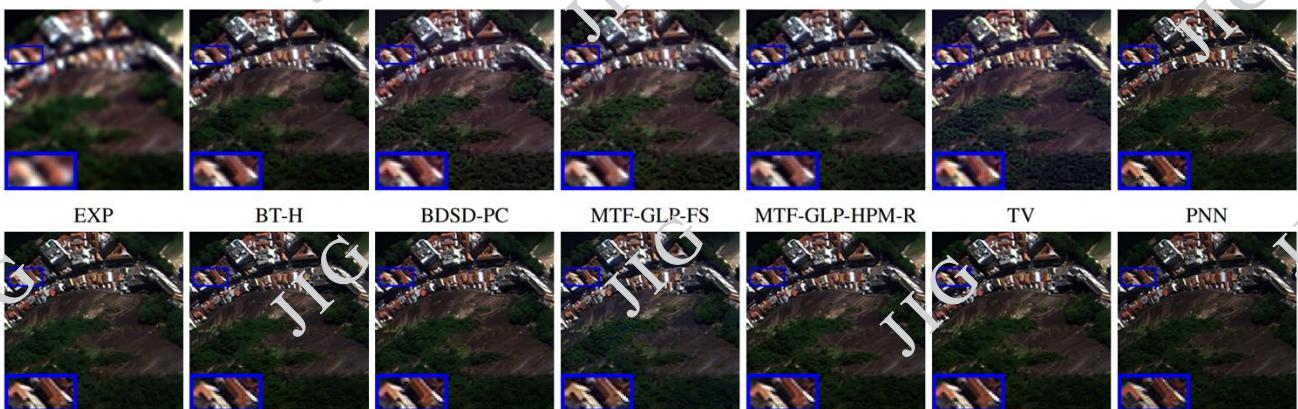


图 6. 不同方法在 WV3 仿真数据的视觉效果展示。

Figure 6. Visual presentation of different methods on WV3 simulation data.

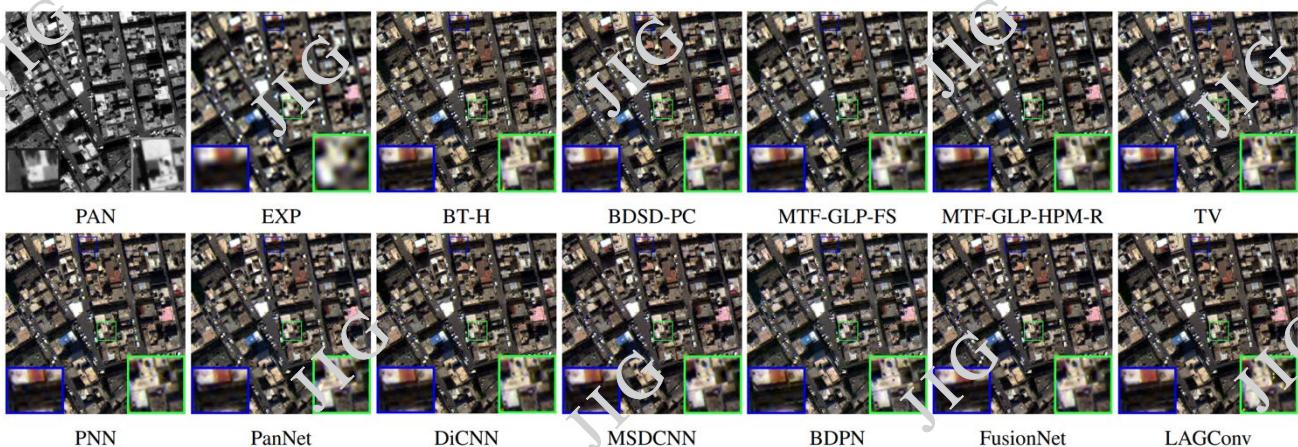


图 7. 不同方法在 WV3 真实数据的视觉效果展示。

Figure 7. Visual presentation of the different methods on WV3 real data.

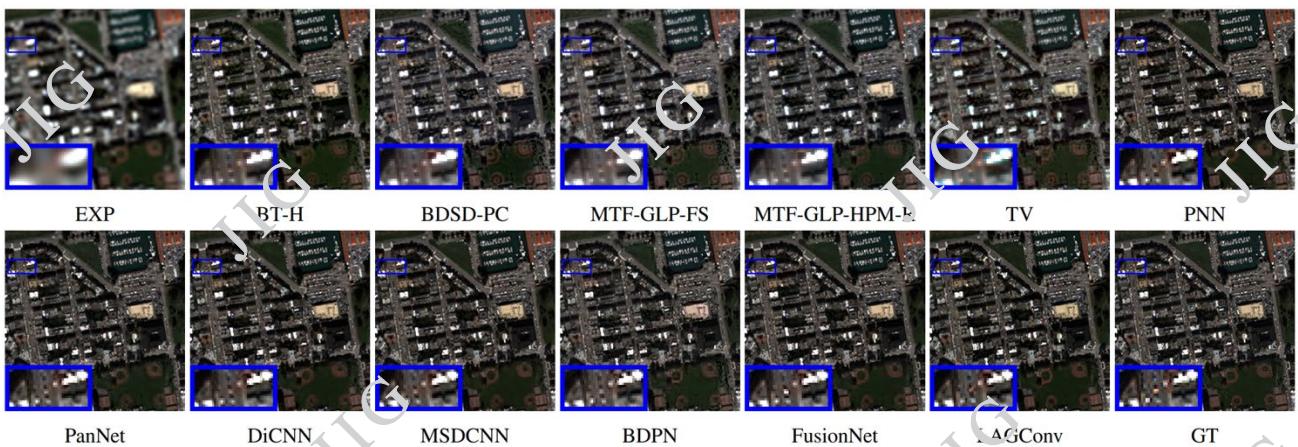


图 8. 不同方法在 QB 仿真数据的视觉效果展示。

Figure 8. Visual presentation of different methods on QB simulation data.

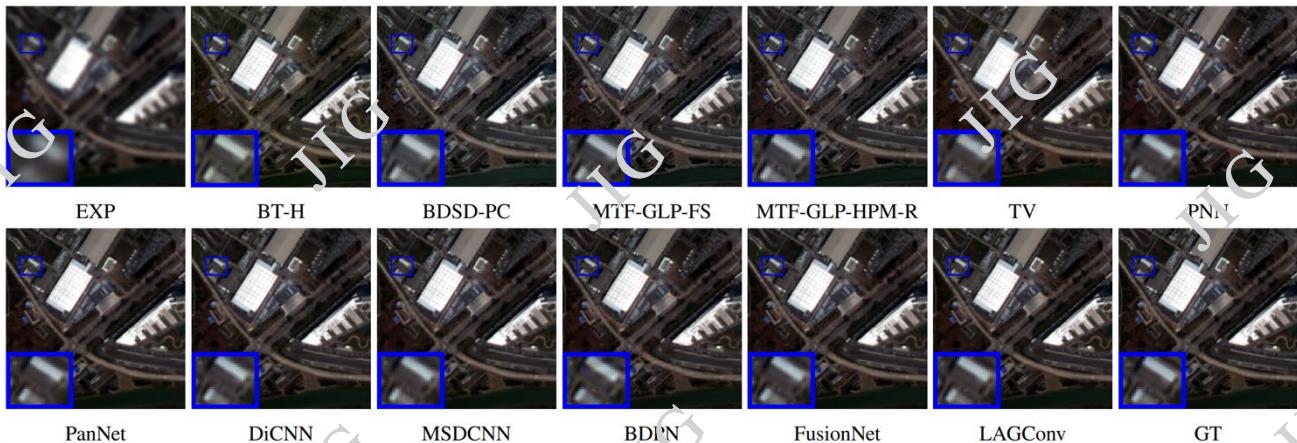


图 9. 不同方法在 GF2 仿真数据的视觉效果展示。

Figure 9. Visual presentation of the different methods in GF2 simulation data.

表 10 所有方法在 WV3 仿真测试数据集上的测试时间比较。传统的方法（第一行）是在 CPU 上测试，而基 CNN 的方法（第二行）是在 GPU 上测试。（单位：秒）

Table 10 Comparison of test times for all methods on the WV3 simulation test dataset. The conventional method (first row) is tested on the CPU, while the base CNN method (second row) is tested on the GPU. (Unit: seconds)

	EXP	BT-H	BDSD-PC	MTF-GLP-FS	MTF-GLP-HPM-R	TV	
Testing Time	0.07	0.092	0.234	0.314	0.246	31.232	
	PNN	PanNet	DiCNN	MSDCNN	BDPN	FusionNet	LAGConv
Testing Time	0.47	0.599	0.581	0.417	0.463	0.588	1.366

表 11 基于 CNN 的方法在 WV3 训练数据集上的训练时间（小时:分钟）、参数量和 GFlops 比较。

Table 11 Comparison of training time (hours:minutes), number of parameters and GFlops for CNN-based methods on the WV3 training dataset.

	PNN	PanNet	DICNN	MSDCNN	BDPN	FusionNet	LAGConv
Train. Time	26:41	2:38	11:23	1:29	24:51	1:11	19:18
Para.#	104,360	83,024	46,792	228,556	1,484,412	78,632	54,341
GFlops	0.29	0.23	0.19	0.91	3.8	0.32	0.51

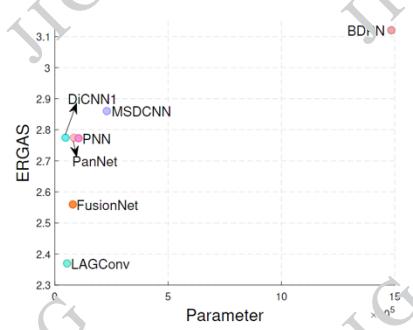
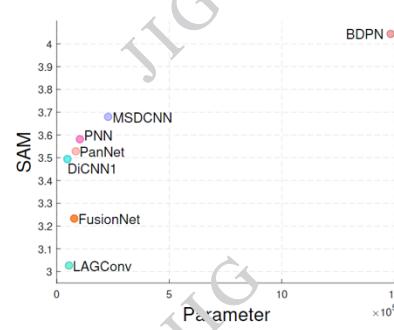
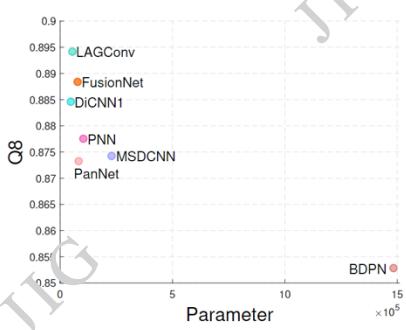


图 10. 基于 CNN 的方法在 WV3 仿真数据集上的参数量和 Q8、SAM、ERGAS 之间的关系。

Figure 10. Relationship between the number of parameters and Q8, SAM, and ERGAS for the CNN-based approach on the WV3 simulation dataset.

集上的训练时间、参数量和计算量 GFlops 的比较。由表可知 PNN 和 BDPN 耗费的训练时间最多，而 FusionNet 则最少；而 DiCNN 的参数量最少，对比之下 BDPN 的参数量最多，超过了 100 万；DiCNN

的 GFlops 最少，而 BDPN 的 GFlops 最多。为了更形象表示参数量和性能指标之间的关系，图 10 分别展示了这七种 CNN 方法在 WV3 仿真数据集上的参数量和 Q8、SAM、ERGAS 之间的关系。

## 4 讨论与展望

### 4.1 未来研究挑战

从实验结果可知，基于深度学习方法（本文主要是 CNN 的方法）当训练数据和测试数据性质接近时，其可以获得出色的图像融合效果。而当训练数据与测试数据性质差异较大时，CNN 获得的结果往往不尽如人意，即数据泛化性差。这实际上也是在意料之中的，因为 CNN 是针对当前训练数据去“学习”或“归纳”（或数学上称之为拟合）出其潜在“规律”（或称之为先验），因此若在新的具有不同特征的数据下测试，通常学习到的“规律”并不适用此类数据，进而不能获得优秀效果（有时甚至差于传统方法）。而传统方法由于没有基于大量数据的训练，可能对数据潜在的先验挖掘并不充分，因此通常效果会弱于深度学习的方法。此外，基于深度学习方法的资源耗费问题（如大量数据集需求、训练时间长等）也是其一大弱点之一。

### 4.2 未来研究方向

传统方法的优势在于其良好的数据泛化性，只需一对 PAN 和 MS 图像输入即可获得融合图像（即零样本），并且不需要通过下采样获得的标签数据（注：下采样过程会降低图像分辨率，因此训练得到的模型并不是在原始分辨率上获得的，即无监督）。因此，基于深度学习的全色锐化方法在未来可以朝着这两个具有潜力和挑战性的方向前进（即无监督、零样本方法），进而逐渐克服当前深度学习方法的弱点。

同样针对上述传统方法和深度学习方法的优缺点，另一个有潜力的方向是通过合理的策略将两类方法嵌入到一个框架下，共享两类方法的优点。这个思路已有研究人员进行过尝试，大体包括两种思路。一是通过将变分优化模型进行模型展开（Unfolding），进而针对展开的子问题设计相似的网络结构（主要是通过网络去近似模拟子问题的求解），最后在大量训练数据集上进行训练获得优异的测试结果（Xie 等，2019）。另一种方法是针对变分优化模型的未知正则项进行合适的变量替换，进而构建带隐式正则项的子问题（通常可视作一个去噪问题），子问题的求解过程还是按照传统子问题交替更新的方式进行，只不过在求解带隐式正则项子问题的时候通常使用一个预训练好的深度学习去噪器（也可以是传统效果优良的去噪器，如 BM3D）（Teodoro 等，2017），这类方法称之为即插即用（Plug-and-Play），

PnP）方法。近来，一种更为简单直接的结合传统方法和深度学习方法的思路是直接将深度学习预训练模型的结果嵌入到传统变分优化模型中（Shen 等，2019；Wu 等，2020；Wu 等，2022）。其主要通过构造一个潜在融合图像和深度学习预训练图像之间的近似项将传统变分模型和深度学习结果连接在一起，后续的模型求解过程和传统变分模型求解方式一致。其主要研究的科学问题集中在如何去构造连接潜在融合图像和深度学习预训练图像的权重距离函数。上述所提到的方法都是近年全色锐化方向传统变分方法和深度学习方法结合的主要思路，这些方法有效提高了深度学习方法的效果，尤其是数据泛化性。但依然面临如参数量过大、训练/测试时间过长等问题，这些问题后续可以进一步研究的方向。

## 5 公开资源分享

本节分享的相关数据代码资源请见本课题主页：  
<https://liangjiandeng.github.io/PanCollection.html>

### 5.1 训练-测试数据集下载

正如前文所提到的那样，基于深度学习的全色锐化领域当前并没有统一公平的训练数据集和测试数据集，严重制约本领域的发展。本文另一个重要的目的就是正式发布全色锐化相关的训练和测试数据集，并且详细给出数据集的生成细节（包括数据集获取地点和时间、训练和测试数据的截取情况、仿真细节等，详见章节 2）。

### 5.2 七种 CNN 方法的 Pytorch 统一框架源代码下载

亦如前文提到那样，当前基于深度学习的各种方法代码编写方式、平台框架等不尽相同，造成后续研究学者入门困难，更重要可能会造成方法比较的不公平性。为了有效处理这个问题，本文将正式发布一个基于 Pytorch 的统一代码编写框架。此框架具有统一的数据输入和输出模块、简单易懂的编写模式、清晰的使用指南，后续研究者只需要将方法模块替换为自己的工作即可有效的和已有方法进行比较。特别地，本文比较的七种 CNN 方法将在此框架下运行，读者可通过运行相关方法的代码轻松入门本领域。

### 5.3 全色锐化传统-深度学习方法 MATLAB 测试软件包下载

通常深度学习方法得到的结果需要和传统方法进行公平比较，但是传统方法并没有 Python 的软件

包。常用的方法是将深度学习的结果（Python 代码得到）放入已有的全色锐化 MATLAB 软件包中进行比较。本文亦会发布一个更简单的全色锐化 MATLAB 测试软件包，在此软件包中将预留存放深度学习结果的文件路径，读者只需要将深度学习结果存入相应路径，然后运行测试函数，即可一键得到所有传统方法和深度学习方法的比较结果（包括量化结果和视觉结果），并将结果存储在相应的文件夹，便于读者使用。

## 6 结 论

本文针对深度学习全色锐化问题中几个亟待解决的问题（如数据集的细节呈现、公平的训练-测试数据集、统一的编程代码框架、方便的全色锐化测试 MATLAB 软件包）展开讨论，并开源相关的数据和代码。正如引言部分所提到的贡献，本文：1) 详细介绍七种基于 CNN 的全色锐化方法的基础上，在统一数据集下进行实验比较；2) 对涉及的训练-测试数据集的仿真细节进行讨论，发布相关全色锐化数据集；3) 发布基于 Pytorch 深度学习库的统一编程代码框架，便于后来初学者的入门以及研究开展；4) 发布统一的全色锐化传统-深度学习方法 MATLAB 测试软件包，方便研究人员的测试；5) 对本领域的未来研究方向进行讨论和展望。作者相信本文可以推动全色锐化问题的发展，尤其是在实验比较的前提下，能尽量厘清当前传统方法和深度学习方法各自的优缺点，进一步探索当前全色锐化面临的挑战，希望成为其他同行优秀研究工作的一个参考。

## 参考文献 (References)

- Arienzo A, Vivone G, Garzelli A, Alparone L and Chanussot J. 2022. Full-Resolution Quality Assessment of Pansharpening: Theoretical and Hands-On Approaches. *IEEE Geoscience and Remote Sensing Magazine*: 2–35[DOI:10.1109/MGRS.2022.3170092]
- Aiazzi B, Alparone L, Baronti S and Garzelli A. 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10): 2300–2312[DOI:10.1109/TGRS.2002.803623]
- Buades A, Coll B, Duran J and Sbert C. 2014. Implementation of nonlocal pansharpening image fusion. *Image Processing On Line*, 4: 1–15[DOI:10.5201/ipol.2014.98]
- Berne O, Helens A, Pilleri P and Jobin C. 2010. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy//2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. Reykjavik, Iceland: IEEE: 1–4[DOI:10.1109/WHISPERS.2010.5594900]
- Burt P J and Adelson E H. 1987. The laplacian pyramid as a compact image code. *Readings in Computer Vision*, 31(4): 671–679[DOI:10.1016/B978-0-08-051581-6.50065-9]
- Chavez P S, Kwarteng A Y. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogrammetric Engineering & Remote Sensing*, 55(3): 339–348
- Carper W J, Lillesand T M and Kiefer R W. 1990. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing*, 56(4): 459–467
- Choi J, Yu K Y and Kim Y. 2010. A new adaptive component-substitution based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1): 295–309[DOI:10.1109/TGRS.2010.2051674]
- Cheng M, Wang C and Li J. 2014. Sparse representation based pansharpening using trained dictionary. *IEEE Geoscience and Remote Sensing Letters*, 11: 293–297[DOI:10.1109/LGRS.2013.2256875]
- Dong C, Loy C C, He K M and Tang X O. 2015. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295–307[DOI:10.1109/TPAMI.2015.2439281]
- Deng L J, Feng M and Tai X C. 2019. The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior. *Information Fusion*, 52: 76–89[DOI:10.1016/j.inffus.2018.11.014]
- Deng L J, Vivone G, Jin C and Chanussot J. 2020. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8): 6955–7010[DOI:10.1109/TGRS.2020.3031366]
- Deng L J, Vivone G, Paoletti M E, Scarpa G, He J, Zhang Y J, Chanussot J, and Plaza A J. 2022. Machine Learning in Pansharpening: A Benchmark, from Shallow to Deep Networks. *IEEE Geoscience and Remote Sensing Magazine*: 2–38[DOI:10.1109/MGRS.2022.3187652]
- Do M N and Vetterli M. 2005. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12): 2091–2106[DOI:10.1109/TIP.2005.859376]
- Dian R W, Li S T and Kang X D. 2020. Regularizing hyperspectral and multispectral image fusion by CNN denoiser. *IEEE Transactions on Neural Networks and Learning Systems*, 32(3): 1124–1135[DOI:10.1109/TNNLS.2020.2980398]
- Fang F M, Li F, Shen C M and Zhang G X. 2013. A variational approach for pan-sharpening. *IEEE Transactions on Image Processing*, 22(7): 2822–2834[DOI:10.1109/TIP.2013.2258355]
- Feng Y Q, Liu J M, Chen K, Wang B and Zhao Z X. 2021. Optimization algorithm unfolding deep networks of detail injection model for

- pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5[DOI:10.1109/LGRS.2021.3077183]
- He L, Rao Y Z, Li J, Chanussot J, Plaza A J, Zhu J W and Li B. 2019. Pansharpening via Detail Injection Based Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4): 1188–1204[DOI:10.1109/JSTARS.2019.2898574]
- Hardie R C, Eismann M T and Wilson G L. 2004. MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. *IEEE Transactions on Image Processing*, 13(9): 1174–1184[DOI:10.1109/TIP.2004.829779]
- Hui T W, Loy C C and Tang X. 2016. Depth Map Super-Resolution by Deep Multi-Scale Guidance//Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands: Springer: 353–369
- Huang W, Xiao L, Wei Z H, Liu H Y and Tang S Z. 2017. A new pansharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5): 1037–1041[DOI:10.1109/LGRS.2014.2376034]
- Jin Z R, Zhang T J, Jiang T X, Vivone G and Deng L J. 2022. LAGConv: Local-context Adaptive Convolution Kernels with Global Harmonic Bias for Pansharpening//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual: AAAI: 1-9[DOI: 10.1609/aaai.v36i1.19996]
- Kuwakami R, Matsushita Y, Wright J, Ben E M, Tai Y W and Ikeuchi K. 2011. High-resolution hyperspectral imaging via matrix factorization//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE: 2329–2336[DOI:10.1109/CVPR.2011.5995457]
- Laben C A and Brower B V. 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. U.S., No. 6,011,875.
- Liu Q J, Zhou H Y, Xu Q Z, Liu X Y and Wang Y H. 2020. PSGAN: A generative adversarial network for remote sensing image pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 10227–10242[DOI:10.1109/TGRS.2020.3042974]
- Loqli S, Alparone L, Garzelli A and Vivone G. 2017. Haze correction for contrast-based multispectral pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 14(12): 2255–2259[DOI:10.1109/LGRS.2017.2761021]
- Li S T, Yin H T and Fang L Y. 2013. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9): 4779–4789[DOI:10.1109/TGRS.2012.22230332]
- Luo S Y, Zhou S B, Feng Y and Xie J G. 2020. Pansharpening via unsupervised convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 4295–4310[DOI:10.1109/JSTARS.2020.3008047]
- Masi G, Cozzolino D, Verdoliva L and Scarpa G. 2016. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7): 594[DOI:10.3390/rs8070594]
- Ma J Y, Yu W, Chen C, Liang P W, Guo X J and Jiang J J. 2020. PanGAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62: 110–120[DOI: 10.1016/j.inffus.2020.04.006]
- Molina R, Vega M, Mateos J and Katsaggelos A K. 2008. Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images. *Applied and Computational Harmonic Analysis*, 24(2): 251–267[DOI: 10.1016/j.acha.2007.03.006]
- Mallat S G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(4): 674–693[DOI:10.1109/34.192463]
- Nason G. 1995. The stationary wavelet transform and some statistical applications. *Lecture Notes in Statistics* 103, Wavelet and Statistics.
- Palsson F, Sveinsson J R and Ulfarsson M O. 2014. A New Pansharpening Algorithm Based on Total Variation. *IEEE Geoscience and Remote Sensing Letters*, 11(1): 318–322[DOI:10.1109/LGRS.2013.2257669]
- Qu Y, Baghbaderani R K, Qi H P and Kwan C M. 2020. Unsupervised pansharpening based on self-attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 59(4): 3192–3208[DOI:10.1109/TGRS.2020.3009207]
- Shen H F, Jiang M H, Li J, Yuan Q Q, Wei Y C and Zhang L P. 2019. Spatial-spectral fusion by combining deep learning and variational model. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 6169–6181[DOI:10.1109/TGRS.2019.2904659]
- Shettigar V K. 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogrammetric engineering and remote sensing*, 58(5): 561–567
- Shah V P, Younan N H and King R L. 2008. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5): 1323–1335[DOI:10.1109/TGRS.2008.916211]
- Teodoro A, Bioucas-Dias J M and Figueiredo M. 2017. Sharpening Hyperspectral Images Using Plug-and-Play Priors//International Conference on Latent Variable Analysis and Signal Separation. Cham, Switzerland: Springer: 392–402[DOI:10.1007/978-3-319-53547-0\_37]
- Tu T M, Su S C, Shyu H C and Huang P S. 2001. A new look at IHS-like image fusion methods. *Information Fusion*, 2(3): 177–186[DOI: 10.1016/S1566-2535(01)00036-7]
- Vivone G. 2019. Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9): 6421–6433[DOI:10.1109/TGRS.2019.2906073]
- Vivone G, Dalla-Mura M, Garzelli A, Restaino R, Scarpa G, Ulfarsson M O, Alparone L, and Chanussot J. 2021. A New Benchmark Based on Recent Advances in Multispectral Pansharpening: Revisiting Pansharpening With Classical and Emerging Pansharpening Methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1): 53–81.[DOI:10.1109/MGRS.2020.3019315]
- Vivone G, Restaino R and Chanussot J. 2018. A Regression-Based High-Pass Modulation Pansharpening Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 984–996[DOI:10.1109/TGRS.2017.2757508]

- Vivone G, Restaino R and Chanussot J. 2018. Full Scale Regression-Based Injection Coefficients for Panchromatic Sharpening. *IEEE Transactions on Image Processing*, 27(7): 3418–3431[DOI:10.1109/TIP.2018.2819501]
- Vivone G, Simoes M, Dalla-Mura M, Restaino R, Bioucas-Dias J M, Licciardi G A and Chanussot J. 2015. Pan sharpening Based on Semiblind Deconvolution. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4): 1997–2010[DOI:10.1109/TGRS.2014.2351754]
- Wang L and Yang X M. 2021. Remote Sensing Image Pan sharpening Feedback Network Based on Perceptual Loss. *Compute Science*, 48(8): 91-98 (王乐, 杨晓敏. 2021. 基于感知损失的遥感图像全色锐化反馈网络. *计算机科学*, 48(8): 91–98)[DOI:10.11896/jsjkx.200700112]
- Wei Y C, Yuan Q Q, Shen H F and Zhang L P. 2017. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geoscience and Remote Sensing Letters*, 14(10): 1795–1799[DOI:10.1109/LGRS.2017.2736020]
- Wu Z C, Huang T Z, Deng L J, Hu J F and Vivone G. 2022. VO+ net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–16[DOI:10.1109/TGRS.2021.3066425]
- Wu Z C, Huang T Z, Deng L J, Vivone G, Miao J Q, Hu J F and Zhao X L. 2020. A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 6277–6290[DOI:10.1109/JSTARS.2020.3030129]
- Xiao L, Liu P F and Li H. 2020. Progress and challenges in the fusion of multisource spatial-spectral remote sensing images. *Journal of Image and Graphics*, 25(5): 851-863 (肖亮, 刘鹏飞, 李恒. 2020. 多源空—谱遥感图像融合方法进展与挑战. *中国图象图形学报*, 25(5): 851-863)[DOI:10.11834/jig.190520]
- Xu S, Zhang J, Zhao Z, Sun K, Liu J and Zhang C. 2021. Deep Gradient Projection Networks for Pan-sharpening//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE: 1366–1375
- Xie W Y, Lei J, Cui Y H, Li Y S and Du Q. 2019. Hyperspectral pan sharpening with deep priors. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5): 1529–1543[DOI:10.1109/TNNLS.2019.2920857]
- 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 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 5449-5457
- Yokoya N, Yairi T and Iwasaki A. 2011. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2): 528–537[DOI:10.1109/TGRS.2011.2161320]
- Yuan Q Q, Wei Y C, Meng X C, Shen H F and Zhang L P. 2018. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3): 978–989[DOI:10.1109/JSTARS.2018.2794888]
- Yang Y, Xu Z, Huang S Y, Wan W G, Tu W and Lu H Y. 2022. Survey of Deep-Learning Approaches for Pixel-level Pansharpening. *National Remote Sensing Bulletin*, 1-22 (杨勇, 苏昭, 黄淑英, 万伟国, 涂伟, 卢航远. 2022. 基于深度学习的像素级全色图像锐化研究综述. *遥感学报*, 1-22)[DOI:10.11834/jrs.20211325]
- Zhong J Y, Yang B, Huang G Y, Zhong Fei and Chen Z Z. 2016. Remote sensing image fusion with convolutional neural network. *Sensing and Imaging*, 17(1): 1-16[DOI:10.1007/s11220-016-0135-v]
- Zhang L P and Zhang L F. 2011. Hyperspectral Remote Sensing. Beijing. Surveying and Mapping Press (张良培, 张立福. 2011. 高光谱遥感. 北京. 测绘出版社)
- Zhang T J, Deng L J, Huang T Z, Chanussot J and Vivone G. 2022. A triple-double convolutional neural network for panchromatic sharpening. *IEEE Transactions on Neural Networks and Learning Systems*: 1-14[DOI:10.1109/TNNLS.2022.3155655]
- Zhu X X, Grohnfeldt C and Bamler R. 2016. Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5): 2664–2681[DOI:10.1109/TGRS.2015.2504261]
- Zhang Y F, De B S and Scheunders P. 2009. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11): 3834–3843[DOI:10.1109/TGRS.2009.2017737]
- Zhang Y J, Liu C, Sun M W and Ou Y J. 2019. Pan-sharpening using an efficient bidirectional pyramid network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 5549–5563[DOI:10.1109/TGRS.2019.2900419]
- Zhuo Y W, Zhang T J, Hu J F, Dou H X, Huang T Z, Deng L J. 2022. A Deep-Shallow Fusion Network With Multidetail Extractor and Spectral Attention for Hyperspectral Pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:7539-7555[DOI:10.1109/JSTARS.2022.3202866]

## 作者简介

第一作者(通讯作者)照片

邓良剑, 1987 年生, 电子科技大学特聘研究员。主要研究方向为基于深度学习、最优化建模、PDE 建模的图像处理和计算机视觉问题研究。E-mail: [liangjian.deng@uestc.edu.cn](mailto:liangjian.deng@uestc.edu.cn)。更多信息详见主页: <https://liangjiandeng.github.io/>

其他作者:

冉燃, 男, 电子科技大学硕士研究生, 基于深度学习的图像融合。E-mail: [ranran@std.uestc.edu.cn](mailto:ranran@std.uestc.edu.cn)

吴潇, 男, 电子科技大学硕士研究生, 基于深度学习的图像复原。E-mail: [wxwsx1997@gmail.com](mailto:wxwsx1997@gmail.com)

张添敬, 女, 电子科技大学本科生, 基于深度学习的图像融合。E-mail: [zhangtianjinguestc@163.com](mailto:zhangtianjinguestc@163.com)