

基于动态卷积的高光谱图像融合网络

冯佳琪

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2024年4月28日; 录用日期: 2024年5月22日; 发布日期: 2024年5月31日

摘要

基于深度学习的高光谱(Hyperspectral Image, HSI)和多光谱图像(Multispectral Image, MSI)融合技术已经被广泛研究, 以提高高光谱图像的分辨率。但大多数方法在融合时不能充分提取高光谱和多光谱图像的有效信息。针对这个问题, 本文提出了一种基于动态卷积的高光谱图像融合网络DHIF, 主要通过三个分支进行图像的特征提取和融合重构。首先, 在特征提取阶段, 引入动态卷积来提取高光谱和多光谱图像的串联信息, 然后针对其他两个图像分支, 设计高光谱动态卷积子网络HDCN和多光谱动态卷积子网络MDCN, 分别提取HSI和MSI的光谱与空间信息。此外, 提取到的信息一方面用于HR-HSI的重构, 另一方面作为损失函数的一部分来约束网络的训练。本文在三个高光谱数据集Pavia University (PU)、Pavia Centre (PC)和Botswana上实现了DHIF, 并和其他九种目前较好的融合算法进行比较, 证明本文提出的模型不论是在数量上还是质量上, 都实现了最好的融合效果。

关键词

高光谱图像融合, 深度学习, 动态卷积

Hyperspectral Image Fusion Network Based on Dynamic Convolution

Jiaqi Feng

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 28th, 2024; accepted: May. 22nd, 2024; published: May. 31st, 2024

Abstract

In order to improve the resolution of hyperspectral image (HSI), many hyperspectral and multispectral image (MSI) fusion methods based on deep learning have been researched widely. Howev-

er, when fusing, most of them don't make full use of the important information of HSI and MSI. To tackle this issue, we propose a hyperspectral and multispectral image fusion network based on dynamic convolution DHIF, where the features of HSI and MSI are extracted from three branches and then used to reconstruct HR-HSI. First, in the stage of feature extraction, we introduce dynamic convolution to capture the series information of HSI and MSI, and then design hyperspectral dynamic convolutional subnetwork HDCN and multispectral dynamic convolutional subnetwork MDCN to utilize the spectral and spatial features of two both, respectively. Third, the extracted features from three branches are not only inputted into the stage of image reconstruction but also added to the loss function to constrain the network training. Extensive experiments on three datasets of Pavia University (PU), Pavia Centre (PC), and Botswana demonstrate that the proposed DHIF surpasses other nine state-of-the-art methods and performs best both in quantitative and qualitative terms.

Keywords

Hyperspectral Image Fusion, Deep Learning, Dynamic Convolution

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

高光谱图像(Hyperspectral Image, HSI)是通过成像技术对目标区域的几十甚至几百个波段同时成像得到的图像, 不仅覆盖了丰富的地表空间信息也包含广泛的光谱信息, 因此被广泛应用于图像分类[1] [2]、目标检测[3] [4]和变化侦测[5] [6]等多种图像后处理领域。但由于成像系统的限制, 为了保证较高的信噪比, 捕获图像在覆盖到大量且连续光谱带的同时, 会以一定的空间信息为代价, 所以高光谱图像往往会有较低的空间分辨率, 从而限制了它的应用和发展。为了更好地进行高光谱图像研究, 我们需要获取同时具有高空间分辨率和高光谱分辨率的高光谱图像(HR-HSI)。

随着高光谱领域的不断研究, 大多数学者采用高空间分辨率的多光谱图像(HR-MSI)与高光谱图像(LR-HSI)的融合技术来提高高光谱图像的分辨率。这种融合方式通常可以分为两类: 基于模型的融合方法和基于深度学习的融合方法。基于模型的融合方法[7] [8] [9]通常需要手工设置先验信息再进行融合, 但这些先验信息不一定能很好地代表高光谱图像的内部结构; 而基于深度学习的融合方法通过借助深度学习强大的学习能力, 目前已有很大的发展[10] [11] [12]。虽然上述深度学习的方法已经实现了一定的融合效果, 但在融合过程中, 大多数方法仅使用静态卷积来提取图像的特征, 没有很好地关注不同图像中的重要信息, 从而限制了融合效果。为解决这个问题, 本文引入了动态卷积[13] [14], 提出了一种基于动态卷积的高光谱图像融合网络(DHIF)。该模型主要从三个分支来用于 HSI、MSI 和串联图像的特征提取与图像重构。对于特征提取阶段的串联分支, 本文使用动态卷积代替大部分网络中的静态卷积来提取串联特征; 而对于其他两个单图像的输入分支, 分别设计高光谱动态卷积子网络 HDCN 和多光谱动态卷积子网络 MDCN 来提取它们的光谱与空间信息。在图像重构阶段, 三个分支提取的特征信息一方面被用于图像融合, 另一方面还作为损失函数的一部分来约束网络模型的训练。综上所述, 本文的主要贡献为:

(1) 本文提出一种基于动态卷积的高光谱图像融合算法(DHIF), 用于高光谱和多光谱图像的融合, 并引入动态卷积对图像中有效信息进行提取。

(2) 本文分别设计了高光谱动态卷积子网络 HDCN 和多光谱动态卷积子网络 MDCN 提取原始 HSI

和 MSI 的空间光谱信息，并将提取的先验信息用于损失函数中，更好地保留两类图像中的原始信息。

(3) 本文在三个高光谱数据集 PU、PC 和 Botswana 上实现了所提出的算法 DHIF，并与其他九种目前较好的融合算法进行比较，从主观视觉和客观数据两方面都证明，本文模型实现了最好的融合效果。

2. 本文模型

本文将 HR-HSI 表示为 $Z \in R^{W \times H \times C}$ ，其中 W 、 H 和 C 分别表示 HR-HSI 的空间宽度、高度和通道数；LR-HSI 和 HR-MSI 分别表示为 $X \in R^{w \times h \times C}$ 和 $Y \in R^{W \times H \times c}$ ， w 、 h 和 C 为 LR-HSI 的宽度、高度和光谱通道数，而 W 、 H 和 c 为 MSI 的空间宽度、高度和光谱通道数。因为 HSI 包含大量的光谱信息，MSI 包含丰富的空间信息，因此 HSI 的 w 和 h 远小于 MSI 的 W 、 H ，而通道数 C 远大于 c ，而 HR-HSI 则同时具有较高的空间分辨率 W 、 H 和光谱分辨率 C 。 W/w 表示 HSI 与 MSI 的缩放因子。通常情况下，在进行融合时，将观测到的 HSI 作为 HR-HSI，即参考图像 GT (GroundTruth)，而用于融合的 LR-HSI 和 HR-MSI 分别由其退化而成，所以三者的关系可以表示为：

$$\begin{cases} X = fold(unfold(Z) \times BD), \\ Y = fold(R \times unfold(Z)), \end{cases} \quad (1)$$

其中， $unfold$ 表示将三维张量展开成二维矩阵，而 $fold$ 为它的逆操作，即将二维矩阵还原成三维张量。 $B \in R^{WH \times WH}$ 、 $D \in R^{WH \times wh}$ 分别表示模糊矩阵和空间下采样矩阵，是将 Z 在空间维度进行模糊和下采样生成 X ， $R \in R^{C \times c}$ 为光谱下采样矩阵，是将 Z 在光谱维度进行下采样生成 Y 。

本文主要是通过 X 和 Y 经融合网络 DHIF 生成 \hat{Z} ，来近似高分辨率的高光谱图像 Z ，因此整个融合过程可表示为：

$$\hat{Z} = Net_{DHIF}(X, Y) \quad (2)$$

其中， Net_{DHIF} 表示本文所使用的融合网络 DHIF， \hat{Z} 则表示融合生成的 HR-HSI。

图 1 展示了本文 DHIF 的整体网络模型结构，主要分为两个阶段进行高光谱图像的融合：特征提取阶段与图像重构阶段。在特征提取阶段，主要从三个分支分别提取 HSI、MSI 和它们的串联特征的信息，然后各分支提取的特征被串联输入图像重构阶段，用于高分辨率高光谱图像的融合和重构。

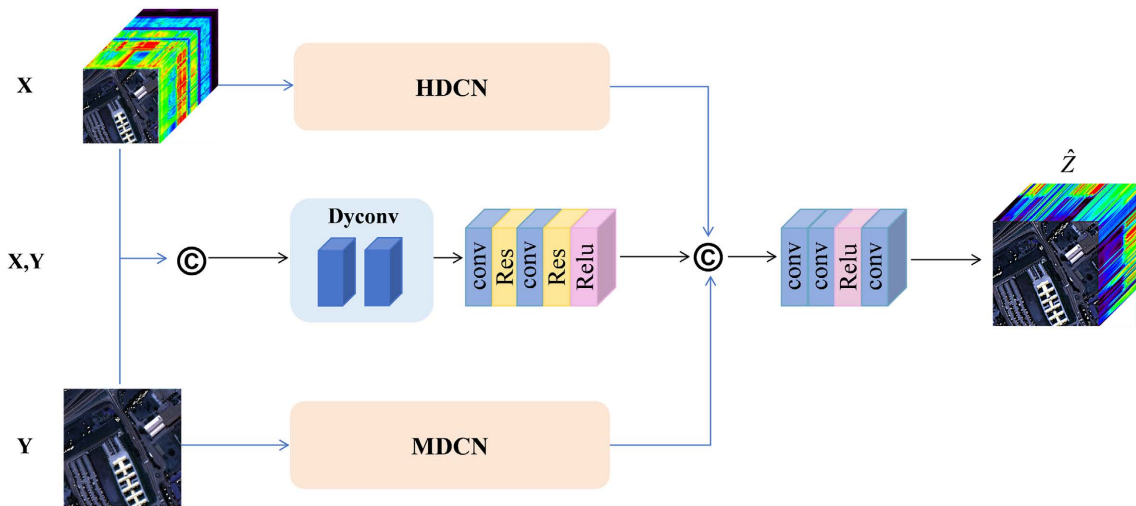


Figure 1. The overall structure of DHIF

图 1. DHIF 的整体结构

2.1. 特征提取阶段

受文献[13]中动态卷积的影响, 本文在特征提取阶段引入动态卷积, 即在传统静态卷积的基础上增加注意力机制, 通过输入特征的加权来更好提取图像中的重要信息。

如图 1 所示, 对于中间分支即高光谱和多光谱的联合特征, 本文将 HSI 通过双三次上采样后与 MSI 串联然后输入进网络, 再设计动态卷积块和静态卷积块, 共同实现串联特征和光谱空间相关性特征的提取。串联之后的特征可表示为:

$$XY = \text{Concatenate}(\text{Bicubic} \uparrow (X), Y) \quad (3)$$

其中 *Concatenate* 表示串联操作, *Bicubic* \uparrow 表示双三次上采样操作, *XY* 表示 HSI 和 MSI 串联之后的图像特征。

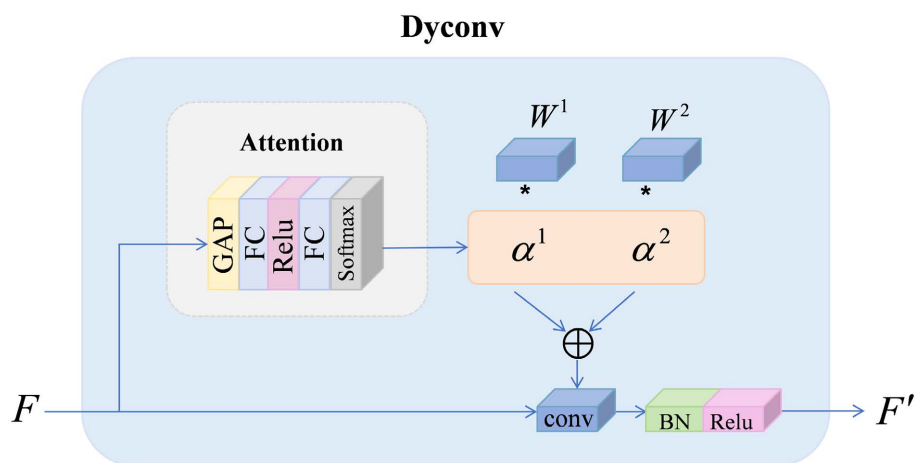


Figure 2. The structure of Dynamic convolution Block (Dyconv)
图 2. 动态卷积块(Dyconv)的结构

图 2 展示动态卷积块的结构, F 表示动态卷积块的输入特征, F' 表示动态卷积块的输出特征。从图中可以看出, 本文采用的动态卷积块包括 2 个 3×3 的卷积核 W^1 和 W^2 , 而且注意力机制 Attention 采用 [13] 中的经典模式, 即包括全局平均池化 GAP、全连接层 FC、激活函数 Relu、全连接层 FC 和激活函数 Softmax 五个部分。全局平均池化首先将输入的信息进行压缩, 然后通过全连接层和激活函数生成正则化的权重 α^1 和 α^2 。计算到的权重分别与对应卷积核进行相乘并逐元素相加后, 形成一个新的卷积操作, 再通过正则化 BN 和激活函数 Relu, 得到经动态卷积后提取的图像的浅层特征。可表示为:

$$\begin{cases} \alpha_0 = \text{Softmax}\left(\text{FC}\left(\text{Relu}\left(\text{FC}\left(\text{GAP}(XY)\right)\right)\right)\right), \alpha_0 = [\alpha_0^1, \alpha_0^2] \\ F'_{XY} = \text{Relu}\left(\text{BN}\left(\text{conv}\left(\alpha_0 \otimes W_0, XY\right)\right)\right), W_0 = [W_0^1, W_0^2] \end{cases} \quad (4)$$

其中 α_0 表示串联分支 XY 经动态卷积块后得到的权重, 包括 2 个权重参数 α^1 和 α^2 , W_0 表示卷积核矩阵, 即 W_0^1 和 W_0^2 。 \otimes 表示向量之间的乘积操作, $\alpha_0 \otimes W_0$ 表示加权后的卷积核, F'_{XY} 表示经动态卷积后所提取的串联分支的浅层特征。

提取完串联分支的浅层特征后, 本文再使用静态卷积块进行深层特征的提取。如图 1 中所示, 静态卷积块包含两个 3×3 的卷积核和一个 Relu 激活函数, 且每个卷积后跟着一个残差块 Res Block [15], 通过信息的传递来减少因网络加深带来的模型退化问题。静态卷积块的操作可表示为:

$$F_{XY} = Relu\left(Res\left(conv\left(Res\left(conv\left(F'_{XY}\right)\right)\right)\right)\right) \quad (5)$$

其中 Res 表示残差块, F_{XY} 表示该分支提取的深层 HSI-MSI 特征。

对于其他两个分支, 本文分别设计了高光谱动态卷积子网络 HDCN 和多光谱动态卷积子网络 MDCN, 同样引入了动态卷积, 来提取 HSI 和 MSI 独有的光谱和空间信息。图 3 分别展示了 HDCN 和 MDCN 的具体结构。

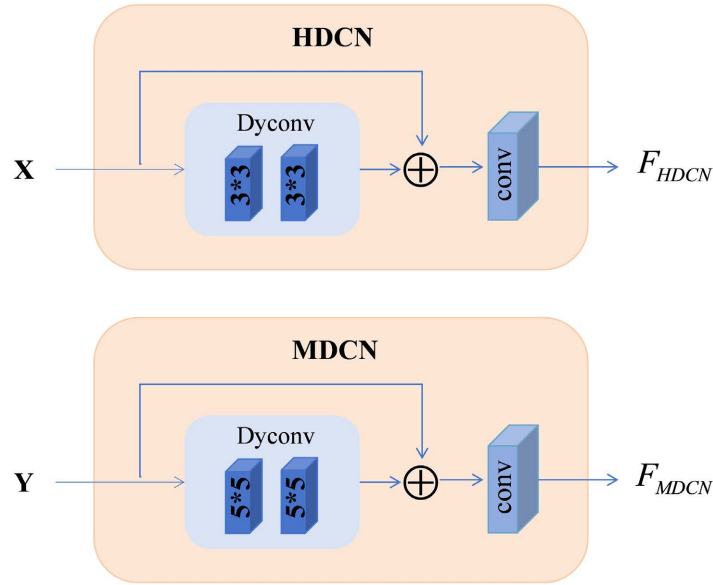


Figure 3. The structure of HDCN and MDCN
图 3. HDCN 和 MDCN 的结构

从图中可以看出, 高光谱动态卷积子网络 HDCN 和多光谱动态卷积子网络 MDCN 都包含一个动态卷积块和一个静态卷积。此处的动态卷积块也采用了图 2 中动态卷积块的结构, 只是在 HDCN 中采用了 2 个 3×3 的卷积核, 而在 MDCN 中采用了 2 个 5×5 的卷积核。这是考虑了 HSI 和 MSI 尺度不同的结果。而且在动态卷积块之后, HSI 和 MSI 也通过跳连接被加到网络中, 以增加原始特征的传递。最后一个卷积被用来进行光谱维度的转换, 将提取特征的光谱维度调整到 64, 便于图像重构阶段的串联与融合。所以 HDCN 的操作可表示为:

$$\begin{cases} \alpha_1 = \text{Softmax}\left(\text{FC}\left(\text{Relu}\left(\text{FC}\left(\text{GAP}(X)\right)\right)\right)\right), \alpha_1 = [\alpha_1^1, \alpha_1^2] \\ F'_{HDCN} = \text{Relu}\left(\text{BN}\left(\text{conv}\left(\alpha_1 \otimes W_1, X\right)\right)\right), W_1 = [W_1^1, W_1^2] \\ F_{HDCN} = \text{conv}\left(F'_{HDCN} \oplus X\right) \end{cases} \quad (6)$$

MDCN 的操作可以表示为:

$$\begin{cases} \alpha_2 = \text{Softmax}\left(\text{FC}\left(\text{Relu}\left(\text{FC}\left(\text{GAP}(Y)\right)\right)\right)\right), \alpha_2 = [\alpha_2^1, \alpha_2^2] \\ F'_{MDCN} = \text{Relu}\left(\text{BN}\left(\text{conv}\left(\alpha_2 \otimes W_2, Y\right)\right)\right), W_2 = [W_2^1, W_2^2] \\ F_{MDCN} = \text{conv}\left(F'_{MDCN} \oplus Y\right) \end{cases} \quad (7)$$

其中, α_1 和 α_2 分别表示 X 和 Y 在输入动态卷积块后经注意力机制得到的权重, W_1 和 W_2 分别表示两个子网络中的卷积核矩阵。 $\alpha_1 \otimes W_1$ 和 $\alpha_2 \otimes W_2$ 表示在 HDCN 和 MDCN 中两个动态卷积块形成的新卷积核, \oplus 表示向量之间的求和操作。 F'_{HDCN} 和 F'_{MDCN} 表示 X 和 Y 经动态卷积块后提取的光谱空间特征, 而 F_{HDCN} 和 F_{MDCN} 表示经过 HDCN 和 MDCN 两个子网络后提取的 HSI 和 MSI 的所有特征。

2.2. 图像重构阶段

在进行图像重构时, 通过特征提取阶段的三个分支分别得到的 HSI、MSI 和串联图像的特征, 被一起作为图像重构阶段的输入, 然后经过一系列操作, 实现最终 HR-HSI 的融合。这些操作主要包括两个 3×3 的静态卷积核, 一个 Relu 激活函数和一个 3×3 的静态卷积核。Relu 激活函数在经过两个卷积核的特征提取后将图像信息映射到非线性空间, 然后再通过一个卷积来匹配原来的 HSI 维度。上述操作可表示为:

$$Z_0 = \text{Concatenate}(F_{HDCN}, F_{XY}, F_{MDCN}) \quad (8)$$

$$\hat{Z} = \text{conv}\left(\text{Relu}\left(\text{conv}\left(\text{conv}(Z_0)\right)\right)\right) \quad (9)$$

其中, Z_0 表示图像重构阶段的输入特征, \hat{Z} 表示最后融合的 HR-HSI。

2.3. 损失函数

在进行模型训练时, 本文不仅使用融合图像与真实图像的差值来约束网络, 还使用提取的光谱空间先验信息与原始 HSI、MSI 的差值共同作为 DHIF 的损失函数, 以进一步提高它的融合能力。本文的损失函数可表示为:

$$L = L_1 + L_{HDCN} + L_{MDCN} \quad (10)$$

$$\begin{cases} L_1 = \frac{1}{N} \sum_{i=1}^N \|\hat{Z} - Z\|_1, \\ L_{HDCN} = \frac{1}{N} \sum_{i=1}^N \|F_{HDCN} - X\|_1, \\ L_{MDCN} = \frac{1}{N} \sum_{i=1}^N \|F_{MDCN} - Y\|_1, \end{cases} \quad (11)$$

其中, L_1 表示融合图像与真实图像的差值, L_{HDCN} 和 L_{MDCN} 分别表示 HDCN 和 MDCN 子网络中提取的空间光谱信息与原始 HSI 和 MSI 的差异, L 则表示模型 DHIF 整体的损失函数, N 为像素的总个数。

3. 实验

为更好地评估本文所提出网络模型 DHIF 的融合性能, 本文分别在三个高光谱数据集上进行了消融实验和对比实验, 并从数量和质量两方面分析模型的实验结果。下面先介绍实验所用的数据集、相关评价指标和实现细节, 再分析具体的实验结果。

3.1. 实验设置

3.1.1. 数据集

- Pavia University (PU)

PU 数据集是通过 ROSIS 传感器在意大利北部的 Pavia 地区获取的。它包含 103 个光谱带和 610×610 的空间像素, 且空间分辨率为 1.3 m。

- Pavia Centre (PC)

和 PU 数据集类似, PC 数据集也是由 ROSIS 传感器在 Pavia 地区捕获到的。但这个图像包含 102 个光谱带和 1096×1096 的空间像素。

- Botswana

Botswana 数据集是由 NASA EO-1 卫星传感器在 Botswana 上空所捕获的。它包含 242 个光谱带和 30 m 的像素分辨率, 且覆盖的光谱范围从 400 nm 到 2500 nm 不等。去掉未校准和噪声光谱带后, 剩余 145 个光谱带。

在去除图像中大量空白的无效信息后, 本文分别选择三个数据集每个光谱带左上部分的 610×340 、 1096×715 和 1476×256 的空间像素, 用于模型的训练和测试。

3.1.2. 评价指标

本文采用了 5 个广泛使用的数量评价指标来对所提出的模型 DHIF 进行定量评估, 它们分别为均方根误差(RMSE)、峰值信噪比(PSNR)、光谱角匹配(SAM) [16]、结构相似性(SSIM) [17]和相对无量纲全局误差(ERGAS) [18]。其中 RMSE 用来估计融合图像和真实图像的强度差, PSNR 评估图像整体的融合质量, SAM 和 SSIM 分别反映融合图像的光谱保真度和空间结构相似性, ERGAS 用来反映全局质量的好坏。对于指标 RMSE、SAM 和 ERGAS, 数据值越小, 融合效果越好; 相反, PSNR 和 ERGAS 则是值越大, 结果越好。

3.1.3. 实现细节

通常情况下, 将所获取的高光谱图像作为参考图像即 HR-HSI, 然后根据 Wald 准则[19]合成相应的 LR-HSI 和 HR-MSI。在具体实验中, HR-HSI 被放进网络后首先按照 32 的步长被切割成 64×64 的小块, 然后 LR-HSI 由 HR-HSI 经过大小为 8×8 , 方差为 2 的高斯模糊核下采样得到。根据下采样因子的不同, 得到的 HSI 大小也不同。本文的实验采用了 4、8 和 16 的缩放因子, 所以获得的 HSI 的大小分别 16×16 、 8×8 和 4×4 。而 HR-MSI 是通过 HR-HSI 乘以相应的光谱转换矩阵得到的。此外, 在模型训练过程中, 设置 batchsize 为 16, 即每次以 16 个块同时进行训练, 并且在 300 epochs 后结束训练。同时使用 ADAM 优化器最小化损失函数, 设置初值 β_1 为 0.9, 学习率 ε 为 $1e-4$ 。在测试阶段, 采用不重叠且大小为 64×64 的图像块来进行模型结果的评估。本文在 PyTorch 实现了所提出的网络 DHIF, 并且使用的软件系统平台为 Ubuntu 18.04.3 LTS, 硬件设备 CPU 为 32 GB RAM 的 i9-9900 KF, GPU 为 11 GB 的 NVIDIA GeForce RTX 2080Ti。

3.2. 量化分析

3.2.1. 消融实验

因为 DHIF 相较于以往的融合网络引入了动态卷积, 所以本文在 PU 数据集上分别对特征提取阶段三个分支中的动态卷积块进行消融实验。“DHIF_S0”和“DHIF_S1”是将串联分支和其他两个子网络 HDCN 和 MDCN 中的动态卷积块分别替换成了 2 个静态卷积核, 且卷积核的大小与原来相同; 而“DHIF_S2”仅使用融合图像和真实图像的差值即 L_f 作为模型的损失函数, 不包括子网络中提取的空间光谱先验。

表 1 显示了在 HSI 和 MSI 的缩放因子为 4 时 PU 数据集的消融实验结果, 最好的结果被加粗显示。从表中可以看出, 不使用动态卷积的情况下, “DHIF_S0”和“DHIF_S1”的 5 个指标结果都低于 DHIF, 说明在本文引入的动态卷积中, 通过对重要信息权重的增加, 可以更好提取图像中的有效特征, 抑制无效特征; 而且“DHIF_S2”的所有结果也都低于本文提出的模型, 说明从 HSI 和 MSI 中提取的先验信息在也有助于图像的融合重构。综上所述, 上述三个消融实验的结果都证明了本文所设计的动态卷积网络

和损失函数的有效性。

Table 1. The quantitative results of ablation study on PU data set ($\times 4$)

表 1. DHIF 在 PU 数据集上的消融实验结果($\times 4$)

模型	factor	RMSE↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓
DHIF_S0	4	0.0192	32.5140	4.4776	0.9577	3.0786
DHIF_S1	4	0.0191	32.5228	4.4708	0.9589	3.0874
DHIF_S2	4	0.0181	32.5603	4.4635	0.9691	2.9018
DHIF	4	0.0171	33.2078	4.1013	0.9782	2.8059

3.2.2. 对比实验

在对比实验部分, 本文使用了一个基于模型的融合方法 IR-TenSR [20]和 8 个目前为止最优的基于深度学习的融合方法 TFNet [21]、HyperPNN [22]、SSRNET [23]、MSDCNN [24]、DARN [25]、DBDNet [26]、SSFCNN [27]以及 MSSJFL [28], 与所提出的模型 DHIF 进行比较。上述深度学习方法都在 Pytorch 上被重新实现, 并且在训练时使用了和 DHIF 相同的参数设置。下面具体分析各对比实验在 3 个数据集上的结果。

- Pavia University (PU)

首先, 在 PU 数据集上实现了 DHIF 和其他对比方法。表 2 分别展示了当缩放因子为 4、8 和 16 时所有方法在 PU 数据集上的对比结果, 最好的结果被加粗显示。

Table 2. The quantitative results obtained by using different methods on PU data set

表 2. 各对比方法在 PU 数据集上的实验结果

模型	factor	RMSE↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓
IR-TenSR	4	0.0268	31.0804	4.6833	0.8766	3.3305
MSSJFL	4	0.2530	9.9021	37.7176	0.4594	48.6553
SSFCNN	4	0.1263	15.6854	39.6333	0.3581	21.9003
DBDNet	4	0.0836	19.4232	17.9799	0.6578	16.1640
SSRNET	4	0.0300	28.1822	7.0414	0.9006	4.7713
MSDCNN	4	0.0242	30.0647	5.7692	0.9303	3.8622
DARN	4	0.0247	29.8835	5.9761	0.9257	3.9435
HyperPNN	4	0.0233	30.4197	5.5660	0.9337	3.6935
TFNet	4	0.0202	31.6072	4.9212	0.9492	3.3355
DHIF	4	0.0179	32.5340	4.4637	0.9582	2.9816
IR-TenSR	8	0.0337	26.9917	7.2149	0.7768	5.0610
MSSJFL	8	0.2152	11.2428	36.9905	0.4781	40.6851
SSFCNN	8	0.1439	14.5430	46.6571	0.2266	24.1806
DBDNet	8	0.1353	15.0095	26.0991	0.4105	25.7282
SSRNET	8	0.0356	26.6933	8.3942	0.8615	5.7293
MSDCNN	8	0.0297	28.3166	6.9335	0.8962	4.6353

续表

DARN	8	0.0304	28.1271	7.3338	0.8917	4.7760
HyperPNN	8	0.0289	28.5710	6.8560	0.9004	4.4894
TFNet	8	0.0282	28.7404	6.4862	0.9075	4.4202
DHIF	8	0.0237	29.7739	5.6709	0.9362	4.0130
IR-TenSR	16	0.0383	24.7683	9.7128	0.7301	6.6964
MSSJFL	16	0.2401	10.4834	37.4915	0.3854	46.9414
SSFCNN	16	0.1333	15.2178	41.8337	0.2812	22.7150
DBDEnet	16	0.1560	13.8719	32.2484	0.3073	27.4385
SSRNET	16	0.0412	25.4413	9.8149	0.8121	6.2779
MSDCNN	16	0.0395	25.8563	9.3293	0.8256	6.0205
DARN	16	0.0383	26.1271	9.0652	0.8339	5.9025
HyperPNN	16	0.0369	26.4113	8.7333	0.8413	5.6301
TFNet	16	0.0384	26.1502	8.9651	0.8325	5.8234
DHIF	16	0.0289	27.6208	8.0122	0.8735	5.2630

从表中的结果可以看出，不论缩放因子是 4、8 还是 16，DHIF 在所有的数据指标上都表现出最好的结果，尽管其他对比方法采用了编码解码、多尺度和残差块等结构。当缩放因子为 4 时，本文提出的 DHIF 的 RMSE、PSNR、SAM、SSIM 和 ERGAS 的结果分别为 0.0179、32.5340、4.4637、0.9582 和 2.9816，相较于其他方法中次好的结果 0.0202、31.6072、4.6833、0.9492 和 3.3305，分别相差 0.0023、0.9268、0.2196、0.0090 和 0.3489。并且随着缩放因子的增大，这些差距也会越来越大，说明本文提出的模型在高光谱图像的缩放因子较大时仍具有很好的适应性。

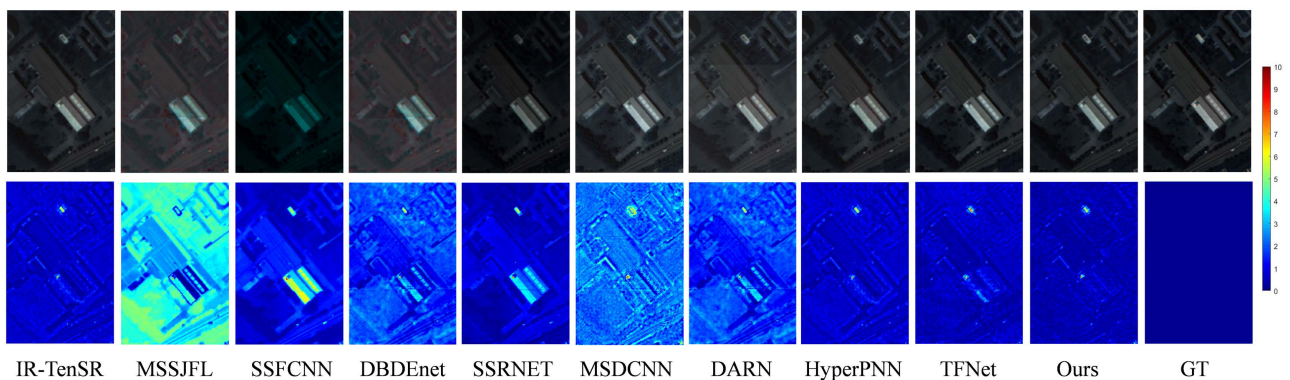


Figure 4. Visual fusion results and error maps on PU data set with the scaling factor $\times 4$

图 4. 各对比方法在缩放因子为 4 时 PU 数据集的融合图像和差值图像

除了数量上的评估结果，本文随机选取了 PU 数据集中的一部分测试结果来进行视觉上的评估。图 4 中第一行分别展示了缩放因子为 4 时，各对比方法和 DHIF 的融合结果，第二行表示的是融合结果与 GT 参考图像的差值图像。从图中可以看出，MSSJFL 产生了最模糊的融合结果，与 GT 的差别最大。SSFCNN、DBDEnet、SSRNET、MSDCNN 和 DARN 产生了一定程度的色彩偏差和伪影，而 HyperPNN、TFNet 和 IR-TenSR 的融合结果相对较好，但在局部细节的恢复上仍有不同程度的差异。相比之下，DHIF

不管是融合结果还是差值图像都和 GT 最接近，说明本文所设计结构的有效性。

- Pavia Centre (PC)

Table 3. The quantitative results obtained by using different methods on PC data set
表 3. 各对比方法在 PC 数据集上的实验结果

模型	factor	RMSE↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓
IR-TenSR	4	0.0258	36.5985	7.3441	0.8561	4.2324
MSSJFL	4	0.0492	23.3330	15.3268	0.8667	16.4339
SSFCNN	4	0.0853	17.2016	41.9369	0.5235	22.6243
DBDNet	4	0.1022	16.3639	32.7977	0.5827	18.0733
SSRNET	4	0.0186	30.2156	10.0815	0.9413	7.0849
MSDCNN	4	0.0129	33.7001	6.7907	0.9689	4.6218
DARN	4	0.0129	33.8513	6.5151	0.9701	4.3708
HyperPNN	4	0.0135	33.5500	6.5497	0.9686	4.4918
TFNet	4	0.0131	33.8374	6.5574	0.9697	4.4302
DHIF	4	0.0118	35.9862	6.2018	0.9786	4.1300
IR-TenSR	8	0.0332	33.5185	10.0813	0.7578	6.7536
MSSJFL	8	0.0621	21.4680	17.4525	0.8490	20.8893
SSFCNN	8	0.0774	18.0555	37.5294	0.5927	20.6615
DBDNet	8	0.1325	13.3770	38.8602	0.4431	41.0272
SSRNET	8	0.0200	29.8615	9.5743	0.9340	7.6581
MSDCNN	8	0.0172	31.5757	7.6856	0.9544	5.6150
DARN	8	0.0173	31.6330	7.6248	0.9544	5.5722
HyperPNN	8	0.0175	31.5968	7.4643	0.9540	5.5135
TFNet	8	0.0176	31.6751	7.2792	0.9540	5.3983
DHIF	8	0.0156	34.5439	7.1082	0.9603	4.9803
IR-TenSR	16	0.0391	30.0060	13.4068	0.6153	9.4932
MSSJFL	16	0.0540	21.5190	18.2255	0.8036	21.7359
SSFCNN	16	0.0786	17.8297	39.9639	0.5228	22.0164
DBDNet	16	0.1355	12.7891	46.0543	0.4229	33.8694
SSRNET	16	0.0248	28.3757	10.6454	0.9101	8.6331
MSDCNN	16	0.0228	29.3348	9.0760	0.9294	7.3709
DARN	16	0.0227	29.6359	8.7905	0.9322	6.9686
HyperPNN	16	0.0223	29.7222	8.7463	0.9328	6.8227
TFNet	16	0.0241	29.2407	8.7071	0.9249	7.1436
DHIF	16	0.0203	30.8902	8.3547	0.9372	6.3705

表 3 展示了 DHIF 和其他对比方法在不同缩放因子的情况下，PC 数据集的各项指标结果。从表中可

可以看出,当缩放因子为4时,IR-TenSR的PSNR优于本文的方法DHIF,但差距不是很大,仅为0.6123。而在其他情况下,所提出方法DHIF的各项指标仍然具有最优的结果。整体而言,DHIF的各项指标结果相较于其他方法,仍保持有最低的RMSE、SAM和ERGAS,与最高的PSNR和SSIM,再次证明本文引入的动态卷积和设计的卷积网络的特征提取能力。

本文随机截取了PC数据集中的一部分测试图像,用于评估各方法的融合效果。图5分别展示了在缩放因子为4时,各融合算法在PC数据集上的融合图像和差值图像。可以直观地看出,IR-TenSR、MSSJFL、SSFCNN、DBDEnet、SSRNET和DARN与GT都存在一定的差异,其中MSSJFL与GT的差值最大,融合结果也最模糊。而MSDCNN、HyperPNN和TFNet的融合结果相对较好,这与数量上的评估结果一致。但总体上来讲,DHIF在中间细节与纹理的恢复上都实现了更好的效果,且与GT最接近,证明DHIF具有更好的特征提取与融合能力。

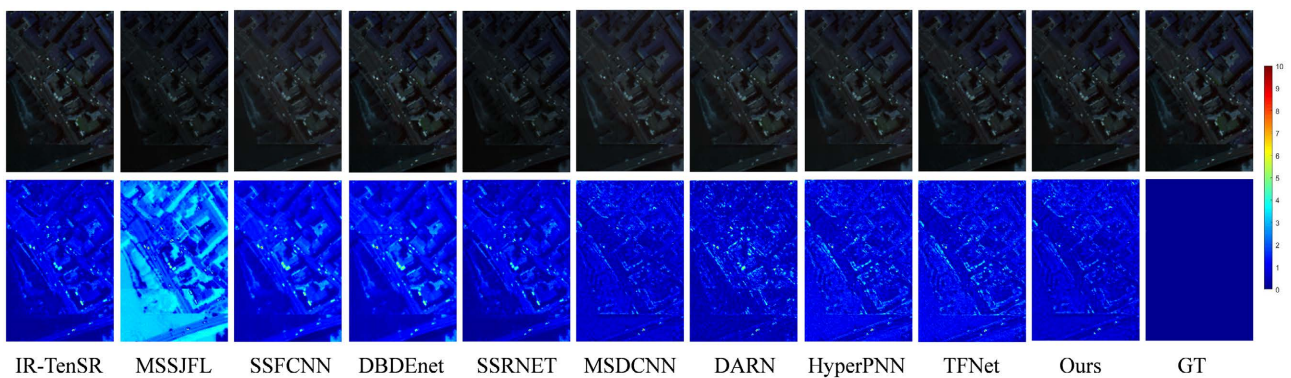


Figure 5. Visual fusion results and error maps on PC data set with the scaling factor $\times 4$

图5. 各对比方法在缩放因子为4时PC数据集的融合图像和差值图像

- Botswana

最后,本文分析了各方法在Botswana数据集的数量结果和视觉质量结果。表4展示了各方法在不同的缩放因子的条件下Botswana数据集的5个评价指标结果。可以看出,Botswana数据集也是在缩放因子为4时,IR-TenSR的PSNR相比于本文的模型略好。但整体而言,在其他情况下,DHIF的各项数量指标的结果上依旧优于其他的方法,表现出最好的结果。

Table 4. The quantitative results obtained by using different methods on Botswana data set

表4. 各对比方法在Botswana数据集上的实验结果

模型	factor	RMSE↓	PSNR↑	SAM↓	SSIM↑	ERGAS↓
IR-TenSR	4	0.0153	40.7072	2.0602	0.9672	2.5040
MSSJFL	4	0.0304	13.4987	26.1786	0.7782	37.6503
SSFCNN	4	0.0173	18.4304	21.0168	0.7089	19.7739
DBDEnet	4	0.0679	6.5030	35.8856	0.3481	108.5578
SSRNET	4	0.0032	32.7875	7.7808	0.9734	12.3203
MSDCNN	4	0.0023	35.7673	2.5616	0.9947	3.2274
DARN	4	0.0019	37.2391	2.1991	0.9960	3.6962
HyperPNN	4	0.0021	36.7331	2.2842	0.9955	2.4910
TFNet	4	0.0020	37.0839	2.2801	0.9961	2.9389

续表

DHIF	4	0.0016	39.3580	2.0335	0.9979	2.4520
IR-TenSR	8	0.0172	36.6985	2.3702	0.9509	2.8958
MSSJFL	8	0.0264	14.7285	23.9103	0.7578	43.8371
SSFCNN	8	0.0194	17.4237	23.7034	0.6861	20.1288
DBDNet	8	0.0584	7.7985	24.7270	0.4089	83.7772
SSRNET	8	0.0040	30.9428	4.8078	0.9618	12.0851
MSDCNN	8	0.0025	35.2620	2.6946	0.9940	3.3423
DARN	8	0.0021	36.4742	2.4078	0.9952	3.8328
HyperPNN	8	0.0023	36.2020	2.4806	0.9950	2.6112
TFNet	8	0.0022	36.2547	2.4646	0.9952	3.2167
DHIF	8	0.0020	36.9822	2.3024	0.9967	2.4830
IR-TenSR	16	0.0192	36.6600	3.0100	0.9445	3.6272
MSSJFL	16	0.0717	6.1445	43.6564	0.6413	96.0242
SSFCNN	16	0.0212	16.6694	26.1238	0.6668	20.4255
DBDNet	16	0.0392	11.3976	29.3273	0.5675	90.6760
SSRNET	16	0.0035	31.9685	4.0954	0.9744	13.4709
MSDCNN	16	0.0026	34.8916	2.8440	0.9937	3.2859
DARN	16	0.0023	35.8528	2.5308	0.9946	3.9563
HyperPNN	16	0.0023	35.8525	2.5839	0.9946	2.6044
TFNet	16	0.0022	36.0472	2.5528	0.9951	2.9570
DHIF	16	0.0019	36.7390	2.2093	0.9958	2.3025

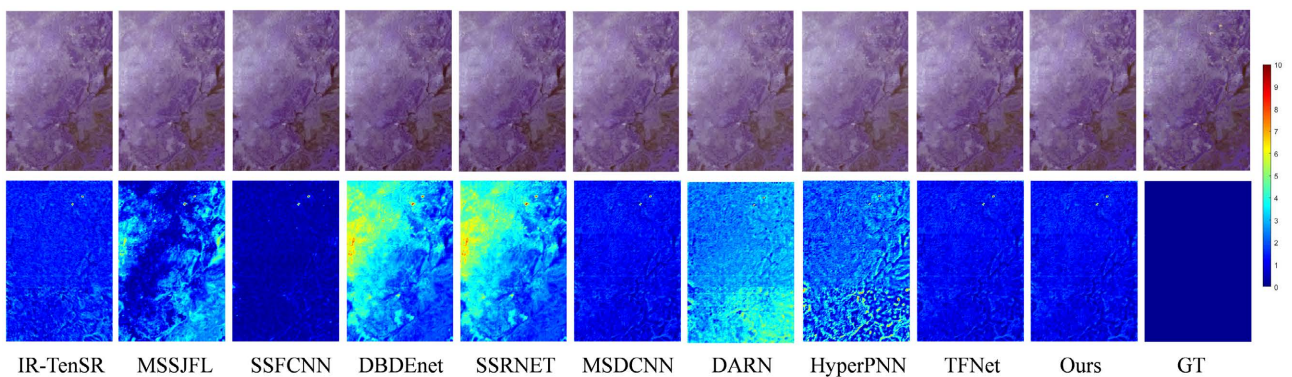


Figure 6. Visual fusion results and error maps on Botswana data set with the scaling factor $\times 4$
图 6. 各对比方法在缩放因子为 4 时 Botswana 数据集的融合图像和差值图像

图 6 展示了在缩放因子为 4 时, 各方法在 Botswana 数据集上的融合效果和差值图像, 本文随机选取了其中一个测试结果进行说明。从结果来看, MSSJFL、DBDNet、SSRNET、DARN 和 HyperPNN 几乎很难得到可视化的融合结果, 而且 SSFCNN 也产生了一定的色彩偏差。MSDCNN、TFNet 和 IR-TenSR 相较于 DHIF 在局部边缘信息的保留上, 都与 GT 存在或多或少的差别。而本文的方法 DHIF 相较于 GT,

仅存在一些细微的差距，并且其他部分都和 GT 最接近，再次证明了本文提出模型的优越性和强大的融合性能。

4. 结论

为更好地提取 HSI 和 MSI 中的有效信息，并用于高光谱和多光谱图像的融合，本文提出了一种基于动态卷积的高光谱图像融合网络 DHIF，其主要包含两个阶段：特征提取与图像重构。在特征提取阶段，本文引入动态卷积，并通过三个分支分别设计了高光谱动态卷积子网络 HDCN、多光谱动态卷积子网络 MDCN 和动态卷积块，以提取 HSI、MSI 和串联特征的光谱空间信息。在图像重构阶段，两个子网络中提取的图像信息一方面被作为该部分的输入，另一方面还被加到损失函数中，进一步约束模型的训练。通过在三个高光谱数据集 PU、PC 和 Botswana 上所做的相关消融实验和对比实验，证明不论是在数量指标还是主观视觉上，本文的模型 DHIF 都具有最好的融合效果。

参考文献

- [1] Mou, L., Ghamisi, P. and Zhu, X.X. (2017) Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 3639-3655. <https://doi.org/10.1109/TGRS.2016.2636241>
- [2] Wang, Q., He, X. and Li, X. (2018) Locality and Structure Regularized Low Rank Representation for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **57**, 911-923. <https://doi.org/10.1109/TGRS.2018.2862899>
- [3] Xie, Z., Li, Y., Niu, J., Yu, X. and Shi, L. (2019) Hyperspectral Face Recognition Using Block Based Convolution Neural Network and AdaBoost Band Selection. 2019 *6th International Conference on Systems and Informatics (ICSAI)*, Shanghai, 2-4 November 2019, 1270-1274.
- [4] Yan, H., Zhang, Y., Wei, W., et al. (2016) Salient Object Detection in Hyperspectral Imagery Using Spectral Gradient Contrast. 2016 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, 10-15 July 2016, 1560-1563. <https://doi.org/10.1109/IGARSS.2016.7729398>
- [5] Zhou, J., Kwan, C., Ayhan, B. and Eismann, M.T. (2016) A Novel Cluster Kernel RX Algorithm for Anomaly and Change Detection Using Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 6497-6504. <https://doi.org/10.1109/TGRS.2016.2585495>
- [6] Li, S., Zhang, K., Hao, Q., Duan, P. and Kang, X. (2018) Hyperspectral Anomaly Detection with Multiscale Attribute and Edge-Preserving Filters. *IEEE Geoscience and Remote Sensing Letters*, **15**, 1605-1609. <https://doi.org/10.1109/LGRS.2018.2853705>
- [7] Hu, J.-F., Huang, T.-Z., Deng, L.-J., Jiang, T.-X., Vivone, G. and Chanussot, J. (2021) Hyperspectral Image Super-Resolution via Deep Spatiospectral Attention Convolutional Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 7251-7265. <https://doi.org/10.1109/TNNLS.2021.3084682>
- [8] Dian, R., Li, S. and Kang, X. (2020) Regularizing Hyperspectral and Multispectral Image Fusion by CNN Denoiser. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 1124-1135. <https://doi.org/10.1109/TNNLS.2020.2980398>
- [9] Dong, W., Fu, F., Shi, G., Cao, X., Wu, J., Li, G. and Li, X. (2016) Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation. *IEEE Transactions on Image Processing*, **25**, 2337-2352. <https://doi.org/10.1109/TIP.2016.2542360>
- [10] Lin, C.-H., Ma, F., Chi, C.-Y. and Hsieh, C.-H. (2017) A Convex Optimization-Based Coupled Nonnegative Matrix Factorization Algorithm for Hyperspectral and Multispectral Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **56**, 1652-1667. <https://doi.org/10.1109/TGRS.2017.2766080>
- [11] Han, X.-H., Shi, B. and Zheng, Y. (2018) Self-Similarity Constrained Sparse Representation for Hyperspectral Image Super-Resolution. *IEEE Transactions on Image Processing*, **27**, 5625-5637. <https://doi.org/10.1109/TIP.2018.2855418>
- [12] Wang, X., Cheng, C., Liu, S., Song, R., Wang, X. and Feng, L. (2023) SS-INR: Spatial-Spectral Implicit Neural Representation Network for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-14. <https://doi.org/10.1109/TGRS.2023.3317413>
- [13] Chen, Y., Dai, X., Liu, M., et al. (2020) Dynamic Convolution: Attention over Convolution Kernels. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11027-11036. <https://doi.org/10.1109/CVPR42600.2020.01104>

- [14] Zhao, Z. and Dong, M. (2023) Channel-Spatial Dynamic Convolution: An Exquisite Omni-Dimensional Dynamic Convolution. 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, 21-23 April 2023, 1707-1711. <https://doi.org/10.1109/ICSP58490.2023.10248781>
- [15] Gross, S. and Wilber, M. (2016) Training and Investigating Residual Nets. Facebook AI Research.
- [16] Yuhas, R.H., Goetz, A.F.H. and Boardman, J.W. (1992) Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, Vol. 1, Pasadena, 1-5 June 1992, 147-149.
- [17] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/TIP.2003.819861>
- [18] Wald, L. (2000) Quality of High Resolution Synthesised Images: Is There a Simple Criterion? *Proceedings of the Third Conference "Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images"*, Sophia Antipolis, 26-28 January 2000, 99-103.
- [19] Han, X.-H., Zheng, Y. and Chen, Y.-W. (2019) Multi-Level and Multi-Scale Spatial and Spectral Fusion CNN for Hyperspectral Image Super-Resolution. 2019 *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, 27-28 October 2019, 4330-4339.
- [20] Xu, T., Huang, T.-Z., Deng, L.-J. and Yokoya, N. (2022) An Iterative Regularization Method Based on Tensor Subspace Representation for Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-16. <https://doi.org/10.1109/TGRS.2022.3176266>
- [21] Liu, X., Liu, Q. and Wang, Y. (2020) Remote Sensing Image Fusion Based on Two-Stream Fusion Network. *Information Fusion*, **55**, 1-15. <https://doi.org/10.1016/j.inffus.2019.07.010>
- [22] He, L., Zhu, J., Li, J., Plaza, A., Chanussot, J. and Li, B. (2019) HyperPNN: Hyperspectral Pansharpening via Spectrally Predictive Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **12**, 3092-3100. <https://doi.org/10.1109/JSTARS.2019.2917584>
- [23] Zhang, X., Huang, W., Wang, Q. and Li, X. (2020) SSR-NET: Spatial-Spectral Reconstruction Network for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **59**, 5953-5965. <https://doi.org/10.1109/TGRS.2020.3018732>
- [24] Yuan, Q., Wei, Y., Meng, X., Shen, H. and Zhang, L. (2018) A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-Sharpener. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **11**, 978-989. <https://doi.org/10.1109/JSTARS.2018.2794888>
- [25] Zheng, Y., Li, J., Li, Y., Guo, J., Wu, X. and Chanussot, J. (2020) Hyperspectral Pansharpening Using Deep Prior and Dual Attention Residual Network. *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 8059-8076. <https://doi.org/10.1109/TGRS.2020.2986313>
- [26] Qu, J., Hou, S., Dong, W., Xiao, S., Du, Q. and Li, Y. (2021) A Dual-Branch Detail Extraction Network for Hyperspectral Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-13. <https://doi.org/10.1109/TGRS.2021.3130420>
- [27] Han, X.H., Shi, B. and Zheng, Y. ((2018) SSF-CNN: Spatial and Spectral Fusion with CNN for Hyperspectral Image Super-Resolution. 2018 25th *IEEE International Conference on Image Processing (ICIP)*, Athens, 7-10 October 2018, 2506-2510. <https://doi.org/10.1109/ICIP.2018.8451142>
- [28] Min, Z., Wang, Y. and Jia, S. (2021) Multiscale Spatial-Spectral Joint Feature Learning for Multispectral and Hyperspectral Image Fusion. 2021 *IEEE 23rd International Conference on High Performance Computing & Communications; 7th International Conference on Data Science & Systems; 19th International Conference on Smart City; 7th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, Haikou, 20-22 December 2021, 1265-1270. <https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00193>