

自适应滤波器的神经网络生成及 遥感图像处理新应用

唐娉¹, 刘璇^{1,2}, 金兴^{1,2}, 张正¹

1. 中国科学院空天信息创新研究院, 北京 100094;
2. 中国科学院大学 电子电气与通信工程学院, 北京 100049

摘要: 图像自适应滤波是非线性的图像变换, 有广泛的应用场景。传统的图像自适应滤波器均是专家设计的, 如双边滤波器和形状自适应滤波等。CNN 作为特征提取和非线性能力表达的有效工具, 可用于学习构造图像自适应滤波器。本文首先介绍了图像自适应滤波器的生成网络, 继而介绍了图像自适应滤波的两个图像处理的新应用: 不同时相间的图像变换用于图像插值、不同波段间的图像变换用于图像融合。从这两类应用中, 可以窥见图像自适应滤波在构建图像非线性变换方面的应用能力。

关键词: 遥感, 图像自适应滤波, 滤波器生成网络, 动态滤波器网络, 图像插值, 图像融合

中图分类号: P2

引用格式: 唐娉, 刘璇, 金兴, 张正. 2023. 自适应滤波器的神经网络生成及遥感图像处理新应用. 遥感学报, 27(7): 1523–1533

Tang P, Liu X, Jin X and Zhang Z. 2023. Neural network generation of adaptive filter and new applications in remote sensing image processing. National Remote Sensing Bulletin, 27(7): 1523–1533 [DOI: 10.11834/jrs.20232174]

1 引言

图像滤波是常用的图像处理方法, 也是一种图像增强方法, 目的是突出感兴趣的图像信息, 抑制或者去除不感兴趣的图像信息。图像去噪、图像平滑、图像锐化等都是通过图像滤波运算实现的。图像滤波可以通过卷积实现, 不同的卷积核代表不同的滤波方法, 比如 Laplace 卷积核表示一种图像边缘提取方法。因此卷积核的设计对滤波很重要。卷积核分两种, 一种与内容相关, 一种与内容无关。与内容有关的滤波也被认为是自适应滤波 (Su 等, 2019)。

图像自适应滤波指可以根据图像的局部结构和内容决定滤波器的形状、滤波器大小和权值。如双边滤波器 (Bilateral filter) (Tomasi 和 Manduchi, 1998)、形状自适应滤波 (Foi 和 Katkovnik, 2006) 等都是图像自适应滤波的例子。以双边滤波器为例, 双边滤波是一种非线性的滤波方法, 是一种可以保边去噪的滤波器, 能够在平滑去噪的同时

还能够很好地保存边缘。图像 f 双边滤波器的滤波公式如下:

$$g(i, j) = \frac{\sum_{k, l} f(k, l) \omega(i, j, k, l)}{\sum_{k, l} \omega(i, j, k, l)} \quad (1)$$

式中, $g(i, j)$ 是滤波结果, $\omega(i, j, k, l)$ 是权值或者说是滤波器的核, 它表示为空域核 $d(i, j, k, l)$ 和值域核 $r(i, j, k, l)$ 的乘积:

$$\omega(i, j, k, l) = d(i, j, k, l) \cdot r(i, j, k, l) = \exp \left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_d^2} - \frac{\|f(i, j) - f(k, l)\|^2}{2\sigma_r^2} \right) \quad (2)$$

式中, 参数 σ_d 、 σ_r 分别表示高斯函数钟状的宽度, 分别表示空间域和值域像元临近的程度; (i, j) 、 (k, l) 表示像元位置。在式 (1) 表达的双边滤波器中, 权值与像素的空间距离相关, 距离越近权重越大, 也与像素值的距离相关, 它根据像素值的差异分配权值, 像素值越接近, 权重越大。如果两个像

收稿日期: 2022-03-20; 预印本: 2022-06-20

基金项目: 国家自然科学基金重大项目 (编号: 42192584)

第一作者简介: 唐娉, 研究方向为遥感图像处理、遥感图像智能处理和大数据技术、非线性关系建模和表达。E-mail: tangping@aircas.ac.cn

通信作者简介: 张正, 研究方向为遥感数据挖掘、高性能遥感数据处理系统的理论和应用。E-mail: zhangzheng@aircas.ac.cn

素值接近,即使相距较远,也比差异大而距离近的像素点权重重大。正是该部分权值使得边缘特性得以保留。这就是与像素内容相关的部分。

自适应滤波器共有的缺点是运算时间长,不适合对大量的滤波器使用。因此,深度学习的卷积神经网络(CNN)并没有采取自适应滤波器,而是采用了权值共享的滤波器组(Goodfellow等,2016),已在图像处理、视频分析、自然语言处理等方面获得成功应用。空间权值共享和内容无关性是CNN标准空间卷积块的两个重要特征。

空间权值共享的优点是为CNN提供了平移不变性,同时显著减少了参数量。缺点是训练滤波器时图像所有像元位置的梯度损失被全局汇总了,迫使CNN学习的滤波器组需要一次性将所有像素位置的平均误差降至最低,这导致在特定位置的结果可能都不理想。

内容无关性的优点是可将相同的卷积滤波器组应用于所有图像和所有像素,而不管其内容如何,这使得它实现简单,CPU/GPU可高度优化实现,因而使CNN成为图像特征提取和特征集成的重要方式。缺点是由于图像内容在图像和像素之间存在很大差异,单个训练的CNN对具有不同复杂度内容(如水面、城区)的图像可能都不是最佳的。理想情况下,希望CNN滤波器组能适应各种类型的图像内容,而标准CNN并非如此。尽管这些缺点可以通过学习试图捕获图像和像素变化的大量滤波器来解决。但这又会增加参数的数量,需要占用更大的内存空间和大量的标记数据(Su等,2019)。

图像自适应滤波是非线性的图像变换。非线性变换具有强大的非线性数据的表示能力。其中,具有空间扭曲的能力的核方法,如支持向量机(SVM)的非线性核函数定义了从低维空间到高维空间的映射,把原始低维空间中线性不可分的数据变成高维空间中线性可分的数据,从而解决了非线性数据的表示问题。CNN的非线性能力是在一系列滤波器组的线性变换后使用非线性激活函数,如ReLU、Sigmoid函数、Tanh函数等,导致在值域的一些范围被激活,其余范围被抑制而实现的。图像自适应滤波的非线性则是通过明确指定应用滤波器的位置,在一些空间区域被激活,而在其他区域被抑制来实现的。显然,图像自适应滤波模型同时具有更高的可解释性的优点。

传统的图像自适应滤波器均是专家设计的,如双边滤波器和形状自适应滤波等。专家设计的自适应滤波器具有直观、易解释的特点,但设计出这样的滤波器并非易事,不是所有非线性变换类的图像处理任务都能设计出好的图像自适应滤波器,如遥感图像不同时相间的变换,因植被随时间的生长变化是非线性的,导致不同时相图像变换间的关系也是非线性的;再如多光谱图像的全色图像和多光谱波段,尽管全色波段的波长区间可能涵盖各多光谱波段的波长区间,但因波段响应函数的缘故,全色波段并非是多光谱波段的线性组合,而是非线性的组合关系,而这类非线性关系很难人为设计出来。

CNN作为特征提取和非线性能力表达的有效工具,可用于学习构造图像自适应滤波器。滤波器生成网络FGN(Filter Generation Networks)就是利用CNN学习构造滤波器的一类方法,基于它可生成动态滤波器网络DFN(Dynamic Filter Networks)(Wu等,2019;Xue等,2016)、图像几何变换网络(Dai等,2017)等,也可用于生成内容自适应的滤波器,学习各种滤波操作,包括局部空间变换,去模糊或自适应去噪等。

FGN是通用的模式,但它的计算策略很难扩展到具有大量滤波器组的整个网络,一般并不会用以替代CNN的标准卷积层,但它学习的图像自适应滤波在图像处理方面有广泛的应用前景。

本文首先介绍了图像自适应滤波器的生成网络,继而介绍了图像自适应滤波的两个图像处理的新应用:不同时相间的图像变换和时相间的插值、图像融合,从这两类应用中,可以窥见图像自适应滤波在构建图像非线性变换方面的应用能力。

2 图像自适应滤波器生成网络

图像自适应滤波有广泛的应用场景。卷积运算作为滤波的空间域运算,可表达多种物理意义。卷积运算可代表一个系统,也可代表某个过程,如成像系统的点扩展函数可用卷积核表示,图像变换(空间变换、辐射变换)过程都可以用卷积核表示。本节主要考虑深度神经网络如何构建图像自适应滤波的非线性变换。

2.1 自适应滤波器生成网络的模型结构

自适应滤波器生成网络的一般应用场景如图1

所示, 其中假定滤波器 ω 是使 $f * \omega = g$, 则滤波器 ω 是 f 和 g 的函数。滤波器生成网络就是通过网络模型实现从已知的 f 和 g 生成 ω 的模型。对于自适应的滤波器而言, ω 的每一个元素都可以不同, 其值只与 f 和 g 的对应邻域相关。

图2是一个自适应滤波器生成网络结构示例。该示例网络采用编码器—解码器结构学习位置依赖的内核, 使用单独的子网预测每个像素处的卷积滤波器权重。网络主要由3部分组成, 包括特征提取(主要提取训练样本的特征)、特征恢复(主要恢复所提取样本的特征)及滤波器(卷积核)估计。

示例中的特征提取部分包含5个卷积层(Conv-1、Conv-2、Conv-3、Conv-4、Conv-5)和池化层(Pool-1、Pool-2、Pool-3、Pool-4、Pool-5)。每个

卷积层中的滤波器数量分别为32、64、128、256和512。

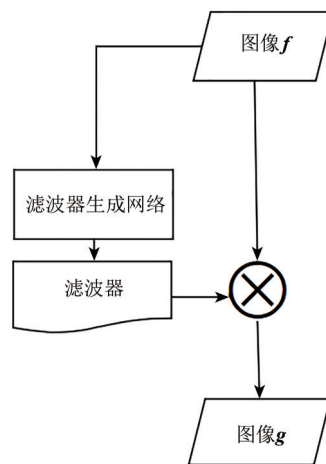


图1 滤波器生成网络架构示意

Fig. 1 Architecture diagram of filter generation network

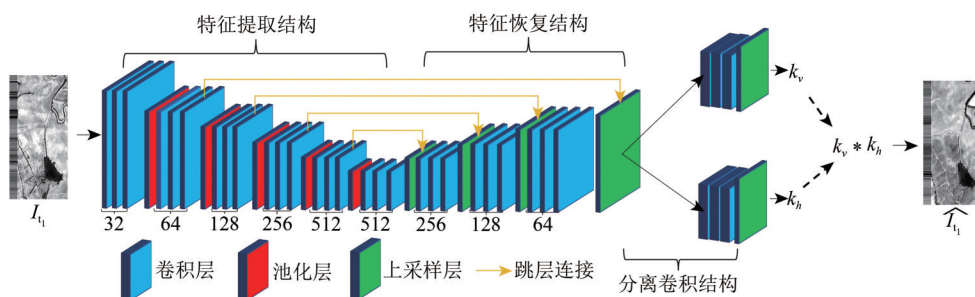


图2 一个自适应滤波器的生成网络结构示例

Fig. 2 One example of architecture of adaptive filter generation network

示例中的特征恢复部分包含4个反卷积层(Deconv-1、Deconv-2、Deconv-3、Deconv-4)和上采样层(Upsample-1、Upsample-2、Upsample-3、Upsample-4)。每个反卷积层中的滤波器数量分别为512、256、128和64。

示例中的滤波器估计部分原本使用一个子网络估计一个像元依赖的2D滤波器。但由于1D滤波器计算简单, 故转而估计两个1D滤波器, 利用两个1D滤波器的乘积来逼近一个2D滤波器, 这么做将每个滤波器的参数量(假定滤波器大小是 $n \times n$) 从 n^2 减少到 $2n$, 极大地减少了运算量, 从而加快速度。因此, 滤波器估计部分主要估计成对的1D卷积核, 同时将最后特征恢复层中的特征信息输入到2个子网络(Subnet-1、Subnet-2)中, 每个子网络包含一个上采样层并估计一个1D卷积核。网络中每层特征图的细节信息如表1所示(以影像块尺寸 128×128 , 卷积核为 3×3 , 影像波段数

量 b 为例)。

需要说明, 可以使用多种方式执行上采样层, 例如最近邻像元、双线性内插和三次卷积内插(Dong等, 2016), 同时利用跳过连接(Long等, 2015)让上采样层结合自适应滤波器的生成网络提取部分的特征。

2.2 自适应滤波器生成网络的目标函数

网络中的损失函数依据目标设立, 主要考虑滤波结果 \hat{g} 与对应参考图像的差异。差异的度量常用 L_2 范数和 L_1 范数。如果 \hat{g} 是滤波后的图像, g 是参考图像, 则 L_2 和 L_1 范数的损失函数可分别定义为

$$\begin{aligned} L_2 &= \|\hat{g} - g\|_2^2 \\ L_1 &= \|\hat{g} - g\|_1 \end{aligned} \quad (3)$$

另一个常用的损失函数被成为是视觉损失函数, 因为该损失函数被发现能够产生视觉上更好的图

像, 该函数通常基于图像的高阶特征定义如下

$$L_p = \|\varphi(\hat{g}) - \varphi(g)\|_2^2 \quad (4)$$

式中, φ 是图像的特征提取算子, φ 可以是如神经网络 VGG-19 的某个 relu 层。

表 1 网络中特征图的尺寸

Table 1 Size of feature maps in the network

类型	卷积核尺寸	步长	特征图尺寸
Input	—	—	$b \times 128 \times 128$
Conv-1	3×3	1×1	$32 \times 128 \times 128$
Pool-1	2×2	2×2	$32 \times 64 \times 64$
Conv-2	3×3	1×1	$64 \times 64 \times 64$
Pool-2	2×2	2×2	$64 \times 32 \times 32$
Conv-3	3×3	1×1	$128 \times 32 \times 32$
Pool-3	2×2	2×2	$128 \times 16 \times 16$
Conv-4	3×3	1×1	$256 \times 16 \times 16$
Pool-4	2×2	2×2	$256 \times 8 \times 8$
Conv-5	3×3	1×1	$512 \times 8 \times 8$
Pool-5	2×2	2×2	$512 \times 4 \times 4$
Deconv-1			$512 \times 4 \times 4$
Upsample-1			$512 \times 8 \times 8$
Deconv-2			$256 \times 8 \times 8$
Upsample-2			$256 \times 16 \times 16$
Deconv-3	3×3	1×1	$128 \times 16 \times 16$
Upsample-3			$128 \times 32 \times 32$
Deconv-4			$64 \times 32 \times 32$
Upsample-4			$64 \times 64 \times 64$
Subnet-1			$3 \times 128 \times 128$
Subnet-2			$3 \times 128 \times 128$

3 自适应滤波器的图像处理新应用

滤波是一种图像变换方法, 也是信息提取方式。自适应滤波器的生成网络在图像处理中的应用本质上取决于自适应滤波器的图像变换能提取怎样的信息。本节介绍的新应用是图像自适应滤波器的两个变换: 不同时相间的图像变换、不同波段间的图像变换, 分别应用于图像插值与图像融合。这区别于之前已知的滤波器的应用。在不同时相间的图像变换中, 自适应滤波器被训练用于表达不同时刻的多光谱图像之间的变换关系, 从而用于图像插值; 在不同波段间的图像变换中, 自适应滤波器被训练用于表达不同波段间的图像变换, 从而用于全色和多光谱间的图像拟合和融合。

3.1 应用实例 1—时相间的图像变换和图像插值

3.1.1 图像插值的新模型

空间分辨率较高的图像插值问题涉及图像在不同时刻变化的描述, 图像在不同时刻的变化不只有空间变化还有光谱变化, 这里我们暂时忽略空间变化, 主要考虑光谱变化, 从光谱变换的角度考虑光谱变化。假定时刻 $t-1$ 的图像 g_{t-1} 经过变换 f_1 , 变成了时刻 t 的图像 g_t , 即 $g_t(i, j) = f_1(g_{t-1}(i, j))$; 时刻 $t+1$ 的图像 g_{t+1} 经过变换 f_2 , 变成了时刻 t 的图像 g_t , 即 $g_t(i, j) = f_2(g_{t+1}(i, j))$, 则图像 g_t 可以通过简单平均插值得到 ($\alpha_1 + \alpha_2 = 1$):

$$g_t(i, j) = \alpha_1 f_1(g_{t-1}(i, j)) + \alpha_2 f_2(g_{t+1}(i, j)) \quad (5)$$

式中, g_t 可以通过前后景的图像 g_{t-1} 和 g_{t+1} 插值得到。

由于不同时间图像之间的变换是非线性的, 因此可假定该非线性变换是自适应滤波器表征的。将图像变换 $\alpha_1 f_1$ 和 $\alpha_2 f_2$ 采用图像自适应滤波的表示, 式 (6) 提供了一个新的图像插值方法 (Jin 等, 2021):

$$g_t(x, y) = g_{t-1} * k_1(x, y) + g_{t+1} * k_2(x, y) \quad (6)$$

式中, k_1 和 k_2 表示二维卷积核, $*$ 表示局部卷积操作。式 (6) 表达的涵义是插值图像在像素位置 (x, y) 的值是 g_{t-1} 和 g_{t+1} 在 (x, y) 位置处的图像块与核函数中心在 (x, y) 处的对应元素相乘的和。如前所述 2D 卷积核 k_1 和 k_2 , $i = 1, 2$ 能够用成对的 1 维卷积核的乘积近似, 成对的 1 维卷积核表示两个不同的方向 (水平、垂直), k_i 可以表示为 $k_{i,v} * k_{i,h}$, $i = 1, 2$ 。在这种假设的条件下, 自适应滤波器的生成网络主要任务是估计每对 1 维卷积核 $k_{i,v} * k_{i,h}$, $i = 1, 2$ 。卷积核的对数与图像的大小一致。不失一般性, 可假设这些 1 维核函数是具有紧支集的, 支集的大小可根据具体情况设定。

3.1.2 图像插值实验与结果

这里图像插值考虑图像 3 个时相之间的关系, 因此网络训练在三元组图像上展开 (Jin 等, 2021)。

实验主要分为 3 组, 每组实验包含 3 景训练影像和 2 景测试影像, 实验的目的是验证不同月份 (6 月、7 月、8 月) 下自适应滤波器的生成网络的有效性以及在相邻月份影像上的泛化能力, 每个训练样本都是一个三元组影像。

这里用 $f_{[I_1, I_2, I_3]}$ 表示由训练影像 I_1, I_2, I_3 产生的非线性网络映射模型, $f_{[I_1, I_2, I_3]}(I_1, I_2)$ 表示该映射对影像 I_1, I_2 的插值结果。

实验数据是一组无人机 (UAV) 影像, 图3展示了无人机影像位置。无人机影像位于圣米歇尔山上

游的库斯农河平原 (法国西部, 48.52°N , 1.53°W)。这个地点是一片面积约 174 ha 的水淹大草原 (Alvarez-Vanhard 等, 2020)。影像投影类型为兰伯特投影、空间分辨率为 0.02 m、波段数为 4 个 (绿色、红色、红边及近红外), 尺寸为 3072×5632 。

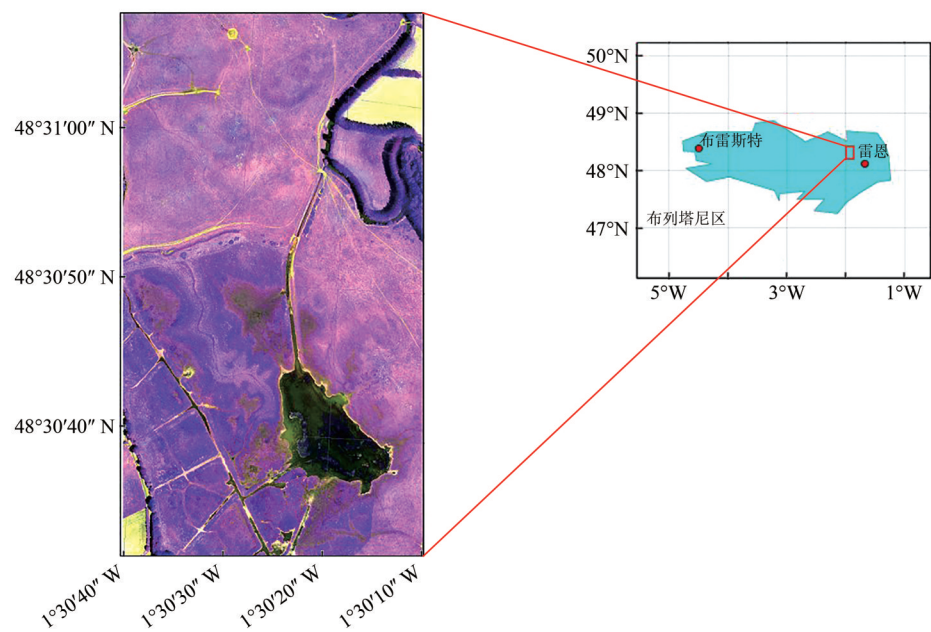


图3 无人机影像的位置(124波段合成)

Fig. 3 Location of UAV image (1, 2 and 4 band composite)

实验数据的名称及日期如表2所示, 表2中 I_4, I_5 是和 I_4, I_5 同月不同年拍摄的图像。实验数据的视觉效果如图4 (a) 所示。

训练样本由训练数据分块制作而成。实验中训练数据被分为 264 块, 94% (250 块) 用于网络的训练, 6% (14 块) 用于网络的验证, 相邻序列中与训练样本同月影像用于网络的测试。实验采用深度学习框架 (Pytorch) 运行自适应滤波器的生成网络。

滤波器估计子网络的卷积核尺寸设置为 3×3 , 网络损失函数使用均方误差, 优化器使用 Adamax, 学习率设置为 0.001 (Kingma 和 Ba, 2017)。

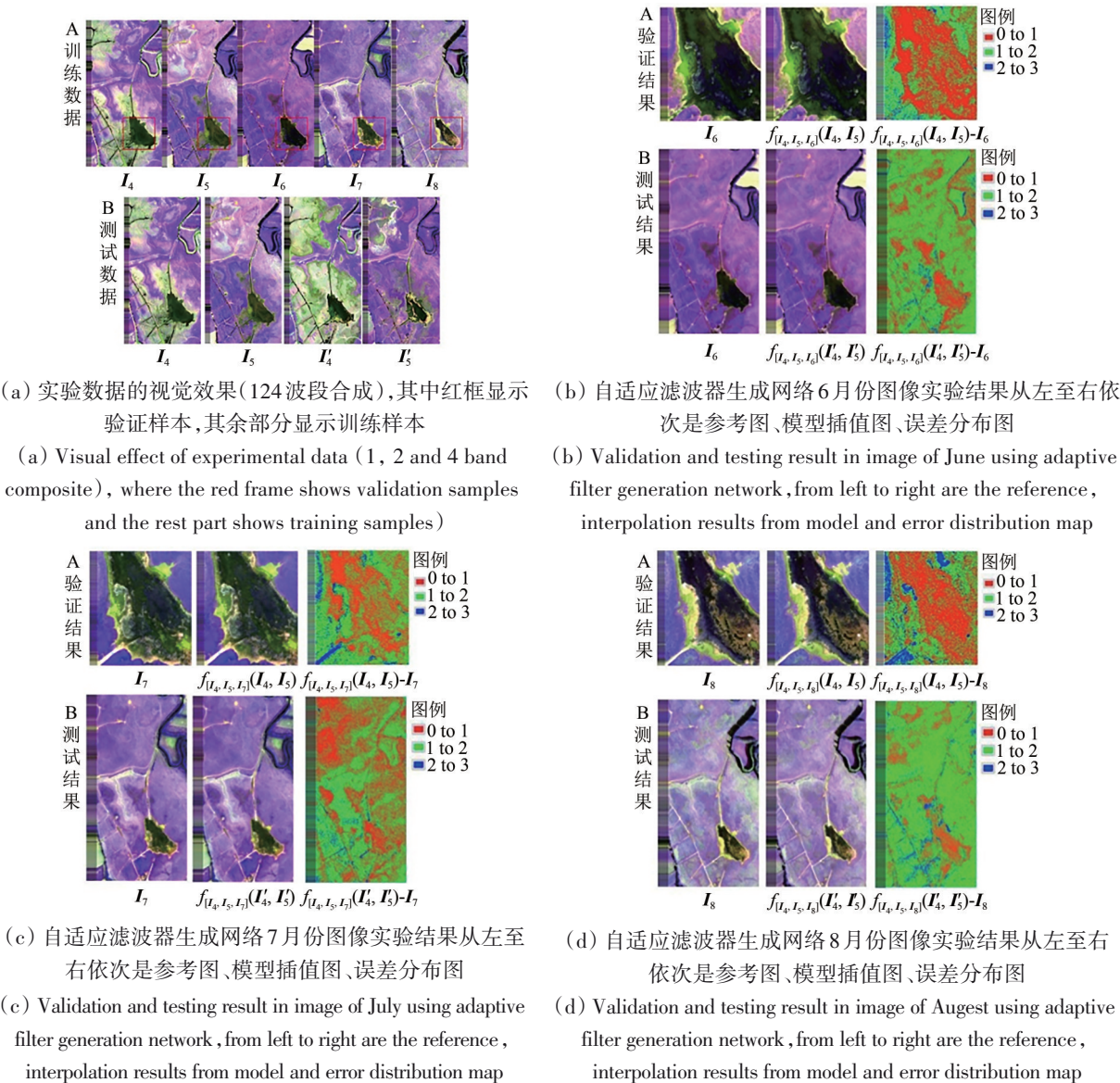
实验图像插值结果的评价采用视觉评价和定量指标评价相结合。

(1) 视觉评价呈现模型插值图像和参考图像的视觉对比, 并给出参考图像和模型插值图像对应像素的差值的像素误差图。不同月份下自适应滤波器的生成网络验证及测试结果如图4 (c) —

(d) 所示。

表2 实验数据的相关信息 (名称和成像日期)
Table 2 Information related to experimental data (name and imaging date)

实验组数	影像用途	名称	成像日期
第1组	训练	I_4	2019-04
		I_5	2019-05
		I_6	2019-06
	测试	I_4'	2018-04
		I_5'	2018-05
第2组	训练	I_4	2019-04
		I_5	2019-05
		I_7	2019-07
	测试	I_4'	2018-04
		I_5'	2018-05
第3组	训练	I_4	2019-04
		I_5	2019-05
		I_8	2019-08
	测试	I_4'	2018-04
		I_5'	2018-05



3.2 应用实例2—波段间的图像变换和图像融合

3.2.1 图像自适应滤波的分量替换模型

图像自适应滤波在图像融合上的应用主要考虑像素级的图像融合, 其中以低空间分辨率的多光谱图像和全色高空间分辨率图像融合形成高空间分辨率多光谱图像的问题为代表。该问题也可以认为是多光谱图像的锐化问题。

分量替换法是一种传统的多光谱图像锐化方法, 如HIS (Chavez等, 1991; Tu等, 2001)、PCA (Chavez等, 1991; Chavez和Kwarteng, 1989)、GS (Li等, 2004; Wang等, 2005; Maurer, 2013)、BDS (Garzelli等, 2008)等, 这一类方法从回归的角度看, 核心是全色图像与多光谱图像多个波段的回归问题, 其融合过程主要包括3个步骤:

利用超像素多光谱波段线性回归拟合或仿真全色波段:

$$P_L(x, y) = \sum_{i=1}^B \alpha_i \cdot \tilde{M}S_i(x, y) \quad (8)$$

全色波段和仿真的全色波段的差值图像作为全色波段的细节图像; 将细节图像依照一定比例添加到各多光谱波段:

$$MS(x, y) = \tilde{M}S_i(x, y) + \beta_i \cdot (P - P_L)(x, y) \quad (9)$$

从上述过程看, 分量替换法的融合方法致力于解决两个子问题, 即细节图像如何提取出来和细节图像如何添加到多光谱波段中去。假定细节添加的比例都取为1, 则上述方法的区别就是细节图像

提取方式不同, 但共同点是都假设了低分辨率的全色图像是超像素多光谱波段的线性组合, 区别只在于组合系数的不同。从滤波的角度看, 低分辨率的全色图像是每个超像素多光谱波段都经历了一个 1×1 核的滤波(或卷积)后叠加在一起的。

基于图像自适应滤波的分量替换方法是一个新的获得低分辨率的全色图像的模型(PDSDNet) (Liu等, 2022), 也即是一个新的获得全色图像细节的模型。该模型用以下公式表示:

$$P_L(x, y) = \sum_{i=1}^B \tilde{M}S_i(x, y) * k_i(x, y) \quad (10)$$

式中, $k_i(x, y)$ 表示波段 i 的像素中心位置在 (x, y) 的卷积核, 该卷积核的大小为 $n \times n$, 本文后续的实验中所用的卷积核大小是 5×5 。取 $\beta_i = 1.0$, 融合后的多光谱图像仍然为

$$MS(x, y) = \tilde{M}S_i(x, y) + (P - P_L)(x, y) \quad (11)$$

式中, 每个波段每个像素的卷积核 $k_i(x, y)$ 通过训练数据集训练获得, 训练使仿真的低分辨率的全色波段 P_L 充分逼近全色波段 P , 训练时的损失函数为:

$$Loss = \frac{1}{N} \sum_{j=1}^N \left\| P_{\{j\}} - \left\{ \sum_{i=1}^B \tilde{M}S_i(x, y) * k_i(x, y) \right\}_{\{j\}} \right\|_F^2 \quad (12)$$

式中, N 表示训练样本数, $\|\cdot\|_F$ 表示Frobenius范数, $P_{\{j\}}$ 表示参考图像中的第 j 个样本。图5是PDSDNet图像融合的流程 (Liu等, 2022)。

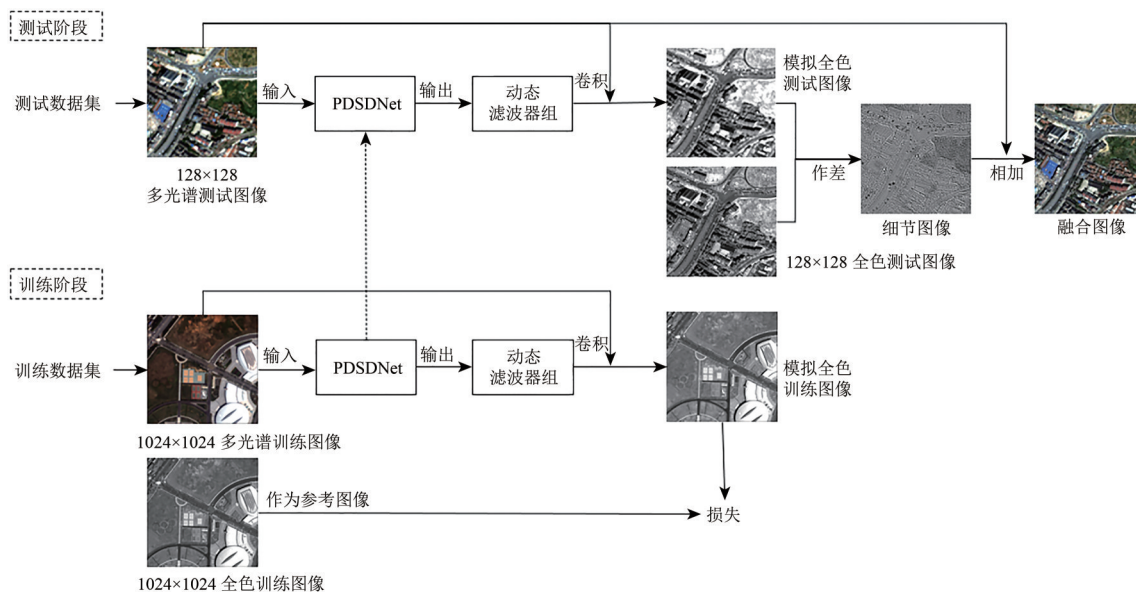


图5 PDSDNet 图像融合流程图

Fig. 5 Flowchart of PDSDNet image fusion

3.2.2 图像融合的实验结果与评价

通常的多光谱锐化方法是传感器相关的，每个传感器可以共享一套参数。

本实验采用的数据来自一个公开的用于多光谱算法锐化效果评价的基准数据集（Meng 等，2021）。数据集包括200个IKONOS、500个QuickBird、160个WorldView-3图像块及其他传感器的图像块数据。其中的高分辨率全色图像为大小1024×1024，低分辨率多光谱图像大小为256×256，二者之间的成像区域相同，已精确配准。基准数据集包含城区、水域、植被区等多种场景。

训练样本的制作时首先将多光谱图像上采样4倍到1024×1024，形成和全色图像相同大小的超像

素图像，之后对上采样后的多光谱图像以及全色图像进行Z-Score标准化处理，实现归一化。随机取出10%图像作为测试集，剩下图像裁剪出若干128×128大小的图像块用于网络训练和验证，其中训练集和验证集比例为5:1，多光谱图像为网络输入，全色图像为真值。

图像自适应滤波器网络中采用的损失函数为MSE函数，初始学习率为0.001，优化器为Adam。网络输出的结果为自适应滤波器，与输入的多光谱图像分波段进行卷积后加和，得到模拟全色图像，再将模拟图像与全色真值得到细节图像后加到多光谱图像中，得到最终融合图像。实验结果如图6所示。

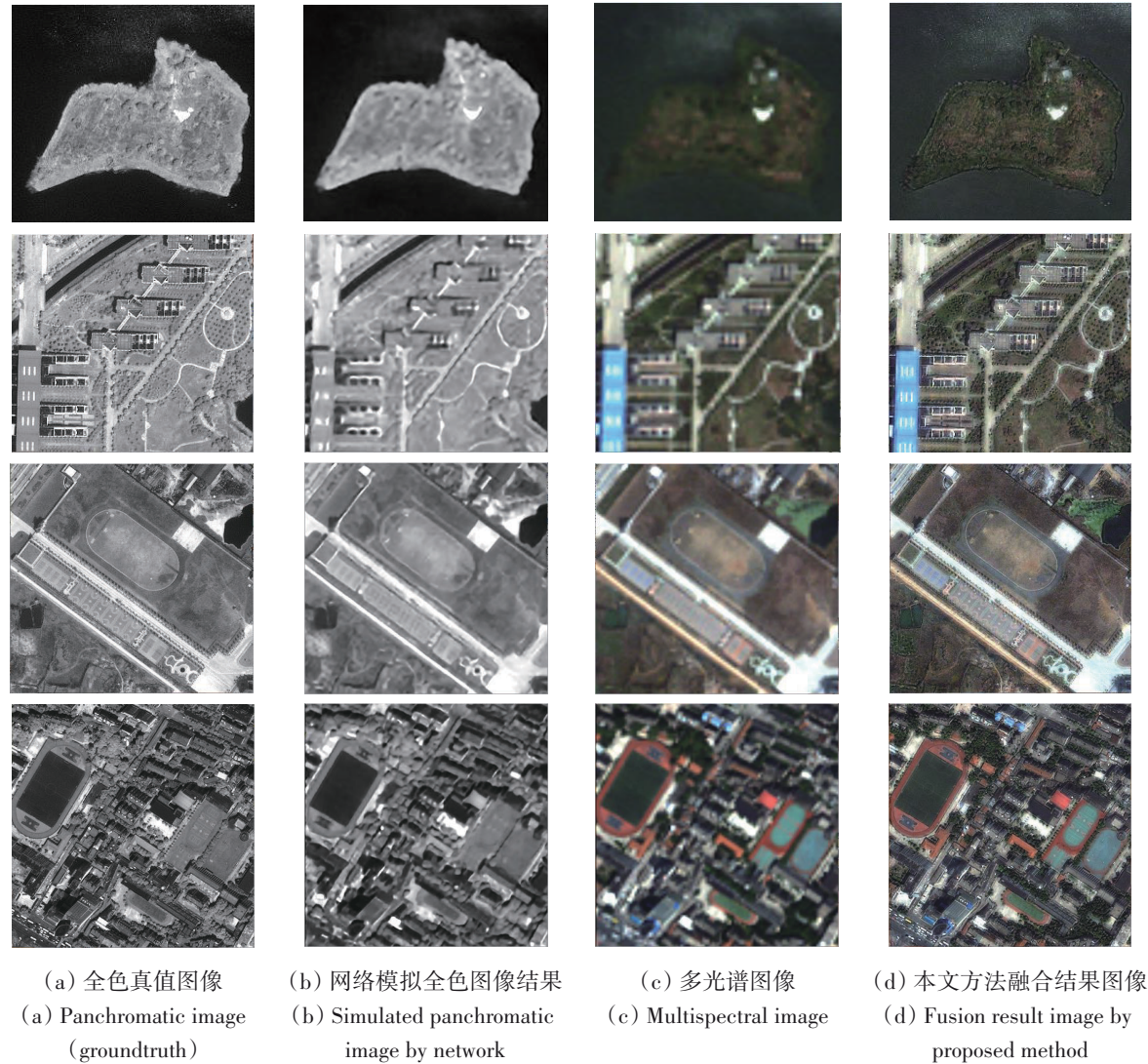


图6 图像自适应滤波器网络全色锐化结果

Fig. 6 Pan-sharpening results of image adaptive filter network

图6共有4组测试图像结果，每组从上至下、从左至右分别为全色真值图像、网络输出的模拟

全色图像、原始多光谱图像、融合结果图像。

从目视效果来看，融合图像能很好地保持高

光谱图像的细节和多光谱图像的光谱特性, 图像不存在明显的光谱扭曲和细节模糊问题。

选取了4个指标SSIM、ERGAS、SAM、MI分别对实验的20张测试图像的融合结果进行定量评价, 定量结果如表4。

表4 图像自适应滤波方法的定量评价指标
Table 4 Quantitative evaluation index of image adaptive filtering

测试图像	SSIM	ERGAS	SAM	MI
1	0.9263	3.9619	0.0043	9.2594
2	0.9269	7.0458	0.0057	10.1087
3	0.9730	9.7233	0.0049	7.2995
4	0.8910	8.8491	0.0076	9.9983
5	0.9469	6.3206	0.0038	9.2730
6	0.9230	6.3202	0.0088	9.0208
7	0.9493	7.4922	0.0072	9.6516
8	0.9124	8.3886	0.0078	9.9859
9	0.9781	7.6031	0.0066	9.1745
10	0.9368	9.5489	0.0088	8.6962
11	0.9514	7.0637	0.0052	9.5263
12	0.9127	5.3262	0.0027	9.5071
13	0.9435	5.0393	0.0042	9.2946
14	0.9087	7.7040	0.0053	10.1009
15	0.0385	9.8473	0.1814	7.6403
16	0.9656	7.3884	0.0041	9.4429
17	0.9571	5.9761	0.0041	9.3463
18	0.2307	4.9624	0.0559	9.3108
19	0.9266	4.4735	0.0058	9.5462
20	0.0606	4.8276	0.1142	9.2927

从定量评价指标来看, 图像自适应滤波方法的融合图像的质量较高, 光谱保真度和空间细节丰富程度都很好。

Liu等(2022)将PDSNet与另外几种融合效果好的方法BDS(BGarzelli等, 2008), MTF-GLP-HPM-PP(Lee和Lee, 2010)和PanNet(Yang等, 2017)的融合结果做了视觉上和定量的评价比较。这3种方法中, BDS是一种分量替换类方法, 它提取的空间细节是波段依赖的, 它通过比较全色波段和波段依赖的广义强度的差值提取空间细节, 并在最小二乘意义下精确线性注入多光谱波段。BDS方法的融合结果通常存在一些光谱畸变, 但是没有显著的空间畸变。MTF-GLP-HPM-PP则是一类基于多分辨率分析MRA(Multi-

Resolution Analysis)的有效的融合方法, 它的基础是广义Laplacian金字塔GLP(Generalized Laplacian Pyramid)模型, 带着调制传递函数MTF(Modulation Transfer Function)匹配的滤波器, 乘性注入模型(Aiazzi等, 2003)和后处理。基于MRA的方法容易有空间畸变很少光谱畸变。PanNet是一个基于深度学习的融合方法, 训练网络参数在高通滤波域而不是图像域, 从多光谱和全色波段的高频信息学习空间细节, 然后添加注入到超像素的多光谱图像。这些方法有个共同的预处理步骤, 即使用MTF先处理图像。PDSNet则没有进行这一步处理。

与上述3种方法相比得到如下结论: 从视觉效果看, PDSNet很有优势, 没有明显的光谱畸变, 也没有明显的空间畸变, 不同于BDS存在的空间畸变和和PanNet在QuickBird和IKONOS数据集上的模糊; 量化指标表明PDSNet只获得了和MTF-GLP-HPM-PP类似的效果, 不是定量指标最好的融合方法。比较分析同样也说明了图像融合视觉效果和量化指标并不匹配, 视觉效果好的PDSNet方法并没有得到最好的量化指标。寻找和视觉效果一致的量化指标仍然是一个开放的问题。

PDSNet之所以能取得比较好的融合质量, 与采用图像自适应滤波器拟合多光谱波段获得了对全色波段更好的低分辨率近似有关, 这种非线性拟合必然优于线性拟合。图像自适应滤波器是内容相关的, 能根据不同的内容自适应生成不同的滤波器, 能更好的学习和模拟全色图像的灰度分布情况, 使得提取出来的细节既充分包含了不同分辨率间的空间细节, 又不会引起光谱的畸变。

4 结 论

传统的图像自适应滤波器需要专家的精心设计, 是非常有效的滤波方式, 它可以根据图像的局部结构和内容决定滤波器的形状、滤波器大小和权值, 常用于抑制噪声同时保持图像的结构特征。作为非线性的图像变换, 图像自适应滤波器有更广泛的应用场景。本文给出了两种不同于传统的应用场景, 一种是将图像自适应滤波器作为多光谱波段回归全色波段的非线性拟合方式用于图像融合, 另一种是用于不同时相间进行图像插值的非线性变换。在这两类场景中, 图像自适应滤波器是基于具体的应用场景利用滤波器生成网

络学习而来的, 然后进行应用。这是图像自适应滤波器在图像处理方面的两个新应用。从这两类应用中, 可以窥见图像自适应滤波在构建图像非线性变换方面的应用能力, 也表明了它的应用潜力有待进一步挖掘。

参考文献(References)

- Aiazzi B, Alparone L, Baronti S, Garzelli A and Selva M. 2003. An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas// Proceedings of the 2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas. Berlin: IEEE: 90-94 [DOI: 10.1109/DFUA.2003.1219964]
- Alvarez-Vanhard E, Houet T, Mony C, Lecoq L and Corpetti T. 2020. Can UAVs fill the gap between in situ surveys and satellites for habitat mapping?. *Remote Sensing of Environment*, 243: 111780 [DOI: 10.1016/j.rse.2020.111780]
- Chavez P S Jr and Kwarteng A Y. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogrammetric Engineering and Remote Sensing*, 55(3): 339-348
- Chavez 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): 265-303
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017. Deformable convolutional networks. *arXiv*: 1703.06211 [DOI: 10.48550/arXiv.1703.06211]
- Dong C, Loy C C, He K M and Tang X O. 2016. 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]
- Foi A and Katkovnik V. 2006. From local polynomial approximation to pointwise shape-adaptive transforms: an evolutionary nonparametric regression perspective// Proceedings of the 2006 International TICSP Workshop on Spectral Methods and Multirate Signal Processing, SMMSP2006, Florence, Italy. TICSP Series: 51-58
- Garzelli A, Nencini F and Capobianco L. 2008. Optimal MMSE pan sharpening of very high resolution multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1): 228-236 [DOI: 10.1109/TGRS.2007.907604]
- Goodfellow I, Bengio Y and Courville A. 2016. *Deep Learning*. Cambridge: MIT Press
- Jin X, Tang P, Houet T, Corpetti T, Alvarez-Vanhard E G and Zhang Z. 2021. Sequence image interpolation via separable convolution network. *Remote Sensing*, 13(2): 296 [DOI: 10.3390/rs13020296]
- Kingma D P and Ba J. 2017. Adam: a method for stochastic optimization. *arXiv*: 1412.6980 [DOI: 10.48550/arXiv.1412.6980]
- Lee J and Lee C. 2010. Fast and efficient panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1): 155-163 [DOI: 10.1109/TGRS.2009.2028613]
- Li C J, Liu L Y, Wang J H, Zhao C J and Wang R C. 2004. Comparison of two methods of the fusion of remote sensing images with fidelity of spectral information//2004 IEEE International Geoscience and Remote Sensing Symposium. Anchorage: IEEE: 2561-2564 [DOI: 10.1109/IGARSS.2004.1369819]
- Liu X, Tang P, Jin X and Zhang Z. 2022. From regression based on dynamic filter network to pansharpening by pixel-dependent spatial-detail injection. *Remote Sensing*, 14(5): 1242 [DOI: 10.3390/rs14051242]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Maurer T. 2013. How to pan-sharpen images using the Gram-Schmidt pan-sharpen method—A recipe. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W1: 239-244 [DOI: 10.5194/isprsarchives-XL-1-W1-239-2013]
- Meng X C, Xiong Y M, Shao F, Shen H F, Sun W W, Yang G, Yuan Q Q, Fu R D and Zhang H Y. 2021. A large-scale benchmark data set for evaluating pansharpening performance: overview and implementation. *IEEE Geoscience and Remote Sensing Magazine*, 9(1): 18-52 [DOI: 10.1109/MGRS.2020.2976696]
- Su H, Jampani V, Sun D Q, Gallo O, Learned-Miller E and Kautz J. 2019. Pixel-adaptive convolutional neural networks//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 11158-11167 [DOI: 10.1109/CVPR.2019.01142]
- Tomasi C and Manduchi R. 1998. Bilateral filtering for gray and color images//Sixth International Conference on Computer Vision. Bombay: IEEE: 839-846 [DOI: 10.1109/ICCV.1998.710815]
- 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]
- Wang Z J, Ziou D, Armenakis C, Li D and Li Q Q. 2005. A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6): 1391-1402 [DOI: 10.1109/TGRS.2005.846874]
- Wu J L, Li D, Yang Y, Bajaj C and Ji X Y. 2019. Dynamic sampling convolutional neural networks. *arXiv*: 1803.07624 [DOI: 10.48550/arXiv.1803.07624]
- Xue T F, Wu J J, Bouman K L and Freeman B T. 2016. Visual dynamics: probabilistic future frame synthesis via cross convolutional networks//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc.: 91-99
- Yang J, Fu X, Hu Y, Huang Y, Ding X, Paisley J. 2017. PanNet: a deep network architecture for pan-sharpening//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE: 1753-1761 [DOI: 10.1109/ICCV.2017.193]

Neural network generation of adaptive filter and new applications in remote sensing image processing

TANG Ping¹, LIU Xuan^{1,2}, JIN Xing^{1,2}, ZHANG Zheng¹

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Image adaptive filtering is a nonlinear image transformation, which has a wide range of applications. Traditional image adaptive filters are designed by experts, such as bilateral filter and shape adaptive filter. They can determine the shape, size, and weight of the filter based on the local structure and content of the image. They are commonly used to suppress noise while preserving the structural characteristics of the image. Convolutional Neural Networks (CNNs) are an effective tool for feature extraction and nonlinear expression. They can be used to learn and construct image adaptive filters. And this paper explores the application of nonlinear image adaptive filters generated by convolutional neural networks in image interpolation and image fusion.

This paper introduces the generation network of an image adaptive filter, including its model structure and objective function. The common network structure usually employs an encoder-decoder architecture, which is mainly composed of three parts: feature extraction, feature recovery, and filter (convolution kernel) estimation. Then, the paper presents two different application scenarios of image adaptive filters: image interpolation and image fusion. The adaptive filter for images enables transformation between different phases during image interpolation and transformation between different bands during image fusion. In these two scenarios, the image adaptive filters are learned by the filter generation network based on the specific application scenario and then applied. In image interpolation applications, the image adaptive filter is used as a nonlinear transformation between two temporal images. The interpolated image is looked at as the mean of adaptive filtering of the previous temporal image and adaptive filtering of the latter temporal image. In image fusion applications, the image adaptive filter is used as a nonlinear fitting method to regress multispectral bands to the panchromatic band. It then extracts spatial details from the difference of the panchromatic band and the simulated panchromatic band, and finally adds spatial details to all the multispectral bands.

We conducted experiments in two application scenarios. The first involved nonlinear transformation for image interpolation with different phases simultaneously. The second utilized an image adaptive filter as a nonlinear fitting method for multi-spectral band regression panchromatic band in image fusion. In image interpolation applications, the experimental results show that the interpolated results are consistent with the reference image in spatial and spectral characteristics, and the RMSE of the interpolated image with the reference image is relatively small. The experimental results for image fusion applications indicate that the low-resolution panchromatic band obtained through adaptive filter fitting of the multi-spectral band is more accurate than the traditional component replacement method. The fusion result obtained by nonlinear image adaptive filters has neither obvious spectral distortion nor obvious spatial distortion.

From the application of nonlinear image adaptive filters generated by convolutional neural networks in image interpolation and image fusion, we have a glimpse of its application potential of image adaptive filter in constructing image nonlinear transformation. The filter generation network can generate adaptive filters for particular application scenarios, resulting in more accurate and visually pleasing images.

Key words: remote sensing, image adaptive filtering, filter generation network, dynamic filter network, image interpolation, image fusion

Supported by National Natural Science Foundation of China(No. 42192584)