

# 基于深度支撑值学习网络的遥感图像融合

李 红<sup>1,2)</sup> 刘 芳<sup>1,2)</sup> 杨淑媛<sup>2)</sup> 张 凯<sup>2)</sup>

<sup>1)</sup>(西安电子科技大学计算机学院 西安 710071)

<sup>2)</sup>(西安电子科技大学智能感知与图像理解教育部重点实验室、智能感知与计算国际联合研究中心、  
智能感知与计算国际合作联合实验室 西安 710071)

**摘 要** 该文将深度学习用于遥感图像融合,在训练深度网络时加入了结构风险最小化的损失函数,提出了一种基于深度支撑值学习网络的融合方法.为了避免图像融合过程中的信息损失,在传统卷积神经网络的基础上,取消了特征映射层的下采样过程,构建了深度支撑值学习网络(Deep Support Value Learning Networks, DSVL Nets),DSVL Nets 网络模型包含 5 个隐藏层,每一层的基本结构由卷积层和线性层构成,该基本单元提供了一种多尺度、多方向、各向异性、非下采样的冗余变换,该模型在网络训练完毕之后,取出各卷积层和第 5 个隐藏层的线性层作为网络模型的输出层.输出层的各卷积层图像融合采用绝对值取大法,得到融合后的各卷积层图像;另外,将线性层图像分别在过完备字典上进行稀疏表示,并对稀疏系数采用绝对值取大法进行融合,得到融合后的线性层图像;最后将融合后的各卷积层和线性层图像重构得到结果图像.文中使用 QuickBird 和 Geoeye 卫星数据验证本文所提方法的有效性,实验结果表明,与 PCA、AWLP、PN-TSSC 和 SVT 算法相比较,该文所提方法的融合结果无论在主观视觉还是客观评价指标上均优于对比算法,较好地保持了图像的光谱信息和空间信息.

**关键词** 深度学习;卷积神经网络;深度支撑值学习网络;过完备字典;遥感图像融合;机器学习  
**中图法分类号** TP391 **DOI 号** 10.11897/SP.J.1016.2016.01583

## Remote Sensing Image Fusion Based on Deep Support Value Learning Networks

LI Hong<sup>1,2)</sup> LIU Fang<sup>1,2)</sup> YANG Shu-Yuan<sup>2)</sup> ZHANG Kai<sup>2)</sup>

<sup>1)</sup>(School of Computer Science and Technology, Xidian University, Xi'an 710071)

<sup>2)</sup>(Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,  
International Research Center of Intelligent Perception and Computation,  
International Collaboration Joint Lab in Intelligent Perception and Computation, Xidian University, Xi'an 710071)

**Abstract** A novel method based on Deep Support Value Learning Networks (DSVL Nets) is proposed for fusion of remote sensing images. The loss function based on structural risk minimization is used in the training of deep learning network. In order to avoid the loss of information, we abandon the downsampling of feature mapping layer of traditional convolution neural network. The DSVL Nets contains five hidden layers, where each layer consists of convolution layer and linear layer. And each layer provides a redundant transform which is multi-scale, multi-direction, anisotropy and non-subsampled. All convolution layers and the fifth linear layer are regarded as the outputs of DSVL Nets. The convolution layers images are fused by abs-maximum model. The

收稿日期:2015-06-29;在线出版日期:2016-01-25. 本课题得到国家“九七三”重点基础研究发展计划项目基金(2013CB329402)、国家自然科学基金(61573267,61173090)、国家自然科学基金重大研究计划(91438201,91438103)、高等学校学科创新引智计划(111 计划)(B07048)、中央高校基本科研业务费专项资金(JB140317,BDY021429)、陕西省教育厅科学研究计划项目(16JK1823)资助. 李 红,女,1976 年生,博士研究生,副教授,中国计算机学会(CCF)会员,主要研究方向为图像处理、模式识别等. E-mail: honglishining@163.com. 刘 芳,女,1963 年生,教授,中国计算机学会(CCF)会员,主要研究领域为智能信息处理、模式识别等. 杨淑媛,女,1978 年生,教授,中国计算机学会(CCF)会员,主要研究领域为智能信息与图像处理、机器学习等. 张 凯,男,1992 年生,博士研究生,主要研究方向为图像融合、机器学习等.

linear layer images are sparsely represented on overcomplete dictionary, and then the coefficients are fused by abs-maximum model. The fused convolution layers images and the linear layer image are reconstructed, and one can obtain the fused result image. Some experiments are taken on several QuickBird and Geoeye satellite datasets. Compared with PCA, AWLP, PN-TSSC and SVT, the experimental results show that the proposed method outperforms some related pan-sharpening approaches in both visual results and numerical guidelines, and reduces the distortion in both the spectral and spatial domain.

**Keywords** deep learning; convolutional neural networks; deep support value learning networks; overcomplete dictionary; remote sensing image fusion; machine learning

## 1 引 言

由于传感器技术的限制,地球观测卫星很难获得同时具有高空间分辨率和高光谱分辨率的遥感图像.例如 QuickBird 卫星和 IKONOS 卫星获得的单通道的全色图像具有较高的空间分辨率,可详尽描述地物的细节特征;而多光谱图像虽具有较多的光谱信息,有利于对地物进行识别、分类和解译,但其空间分辨率较低.将具有高空间分辨率的全色图像与光谱信息丰富的多光谱图像融合,能够获得高分辨的多光谱图像,从而为后期的目标检测与识别等处理提供基础<sup>[1]</sup>.

传统的全色图像与多光谱图像融合技术可以分为:成分替换法,多分辨分析法和迭代复原法.成分替换法原理简单、易实现,常用的成分替换算法有主成分分析(Principal Component Analysis, PCA)、强度色彩饱和度(Intensity-Hue-Saturation, IHS)和施密特正交变化(Gram-Schmidt, GS)等<sup>[2-7]</sup>,该类融合方法能够获得较高的空间信息,但通常会严重的光谱扭曲.基于多分辨分析的融合方法利用 Mallat 提出的多尺度逼近思想,该方法首先对原始图像进行多尺度分解以得到图像在不同尺度分解层上的高频与低频系数,根据不同的融合策略分别对高频和低频系数进行融合,最后经过逆变换得到融合结果.多分辨分析方法在保持光谱信息的同时,较好地注入了全色图像的高频信息,它有效地解决了频谱失真的问题.常见的多分辨分析融合方法有小波变换<sup>[8-9]</sup>、曲线波变换<sup>[10]</sup>、轮廓波变换<sup>[11-12]</sup>和支撑值变换(Support Value Transform, SVT)<sup>[13-15]</sup>等.由于小波变换只能捕获点奇异性,且缺乏平移不变性,难以真实反映图像中的边缘和轮廓等空间特征,它对图像并不是“最稀疏”的表示方法.随之出现的

曲线波和轮廓波等多尺度分析工具较好地解决了二维或更高维奇异性,在图像处理方面得到了广泛且成功的应用.

迭代复原法通过建立全色图像与多光谱图像的观测模型,将融合问题转化为一个复原问题,近年来受到了众多学者的关注. Li 等人<sup>[16]</sup>首先提出了基于压缩感知的全色图像与多光谱图像融合方法,该方法对于不同卫星的多光谱图像和全色图像进行融合时,需要生成不同卫星的、规模较大的字典,实验结果表明,该方法取得了较为满意的融合结果,然而用来生成字典所用到的高分辨多光谱图像通常不易获得;文献<sup>[17]</sup>将全色图像和上采样后的低分辨多光谱图像作为训练图像集用以构造联合字典,这种方法训练字典时不需要高分辨多光谱图像,但却需要大量的训练图像;2013 年, Li 等人<sup>[18]</sup>又提出了在不需要训练图像集的基础上采用字典学习的方法构造所需字典;同年, Zhu 等人<sup>[19]</sup>提出了 SparseFI 融合方法,此方法在不需要大量训练图像集的情况下,仅由全色图像即可得到所需要的对偶 HR/LR 字典,其中 HR 字典是将全色图像进行分块而得到,对全色图像进行下采样,再取其相对应的块作为 LR 字典,这种构造字典的方法简单,但存在的缺点是如果图像块的结构信息较弱或没有结构信息,则从字典中选择适当的原子是一件比较困难的事情; Two-Step 方法的提出,有效利用了低分辨多光谱图像和高分辨全色图像的局部结构相似性<sup>[20]</sup>.

近年来,深度学习在图像处理领域显示出巨大的成功.加拿大多伦多大学的 Hinton 教授以及他的学生 Salakhutdinov<sup>[21]</sup>于 2006 年在《Science》杂志上发表了一篇关于深度学习的文章,文章指出深度学习网络结构具备以下两个特征:第一,包含多个隐藏层的人工神经网络具有优异的特征学习能力,经过网络学习得到的特征对数据本身有着本质的刻

画;第二,可以通过“逐层初始化”(layer-wise pre-training)来降低训练深度神经网络的难度,通过无监督学习实现“逐层初始化”。深度学习能够模拟视觉感知系统的层次结构,通过建立包含有多个隐藏层结构的机器学习模型,对大量数据的训练得到更为有用的特征;它起源于神经网络,是一种深层次的非线性网络结构,通过这一非线性结构更优地逼近复杂函数,深层的结构使深度学习具有极强的表达能力和学习能力,当它用于处理图像时,能够学习到图像的“部分-整体”的分解关系,它在人脸识别、字符识别、图像去噪<sup>[22-23]</sup>等方面取得了远优于传统浅层神经网络的结果。

卷积神经网络(Convolutional Neural Networks, CNNs)是一种典型的深度学习模型,它利用空间相对关系减少参数数目以提高训练性能,是第一个真正获得成功的深度架构<sup>[24]</sup>。CNNs 融合了局部感受野、共享权值和下采样来实现位移缩放和扭曲不变性,如 LeNet5 是一个典型的 CNNs,它有 2 个卷积层、2 个采样层和 2 个全连接层<sup>[25]</sup>。首先,局域感受野指的是每一个卷积层的神经元都只与它上的一个小邻域内的神经单元相连接,以提取初级的视觉特征;第二,权值共享指的是同一个特征图中的神经元可以共享网络权值,它的优点是可以减少网络结构中的参数;第三,下采样降低了特征图的分辨率,降低了位移缩放和扭曲的敏感度;由此可见, CNNs 的 3 个特征保证了卷积神经网络的平移不变性,减小了对位移缩放和扭曲的敏感度以及对提取出的特征出现位置的不敏感性。在卷积神经网络之后,也出现了许多其他的深度学习模型,如去噪自动编码<sup>[26-27]</sup>、DCN<sup>[28]</sup>和 Sumproduct<sup>[29]</sup>等。

尽管卷积神经网络可以识别有变换的模式,具有一定的鲁棒性,并已成功应用于人脸识别、车牌识别<sup>[30]</sup>、行为识别<sup>[24]</sup>、语音识别<sup>[31]</sup>和图像分类<sup>[32]</sup>等计算机视觉的研究领域,目前很少有将它用于图像融合的工作。本文将深度学习用于全色图像与多光谱图像融合,在传统卷积神经网络(Convolutional Neural Networks, CNNs)的基础上,构建了深度支撑值学习网络(Deep Support Value Learning Networks, DSVL Nets)模型,该模型每一层的基本结构包括卷积层和线性层,该基本单元提供了一种对输入图像的多尺度、多方向、各向异性、非下采样的冗余变换。同时该模型去掉了卷积神经网络中的下采样过程,使得输入图像在进行各级滤波时,自适应地学习出不同的支撑值滤波器,以获得对图像的

最优表示。

本文所提方法用于全色图像与多光谱图像融合时突显出了 3 个优点:(1) DSVL Nets 模型在学习过程中不仅仅依赖于经验风险,它能够自适应地学习各级分解时的最优支撑值滤波器;(2) 多尺度支撑值滤波器能够最优发现图像的本质特征,有效提取各级图像的高频信息;(3) 将全色图像的高频信息注入到光谱图像中,减少了图像的空间与光谱扭曲,与已有融合算法相比较,本文所提方法融合图像质量较高。

本文在第 2 节中详细介绍深度支撑值学习网络的构建模型;在第 3 节中给出深度支撑值学习网络的计算过程以及全色图像与多光谱图像的融合步骤;将深度支撑值学习网络用于全色图像与多光谱图像融合,并将融合结果与经典的融合方法作比较;在第 4 节中给出具体的实验结果,并对实验结果进行分析。

## 2 深度支撑值学习网络

### 2.1 深度卷积神经网络(CNNs)

深度学习网络模型可被视为包含了多个隐藏层的人工神经网络模型,为了得到有效的特征以提高预测与分类的准确性,需要利用海量的训练数据来训练网络结构。在深度神经网络中,通过对上一层输出的非线性变换来实现复杂函数的逼近性能。CNNs 模型是一个典型的深度学习模型,它的优势在于可以将图像直接作为网络的输入,它的权值共享网络结构降低了网络模型的复杂度,减少了权值的数量。卷积神经网络结构图如图 1 所示,从图 1 可以看出,多层的结构组成了 CNNs 模型,其中网络的每一层都由数个二维平面组成,每个二维平面又都由数个独立神经元构成,网络中的任一中间级都由 C 层和 S 层串接而成,其中 C 层为卷积层, S 层为特征映射层,通过映射面上神经元的权值共享,降低选择参数的复杂度和减少参数的个数。

### 2.2 深度支撑值学习网络

在深层神经网络的各模型包括上述 CNNs 中,各层中的滤波器大都通过最小化经验风险获得系数。然而,最小化经验风险的准则难以保证训练出的网络具有好的推广性能,类似于支撑向量机(Support Vector Machine, SVM)。在本节中我们使用如图 2 所示的网络基本单元。

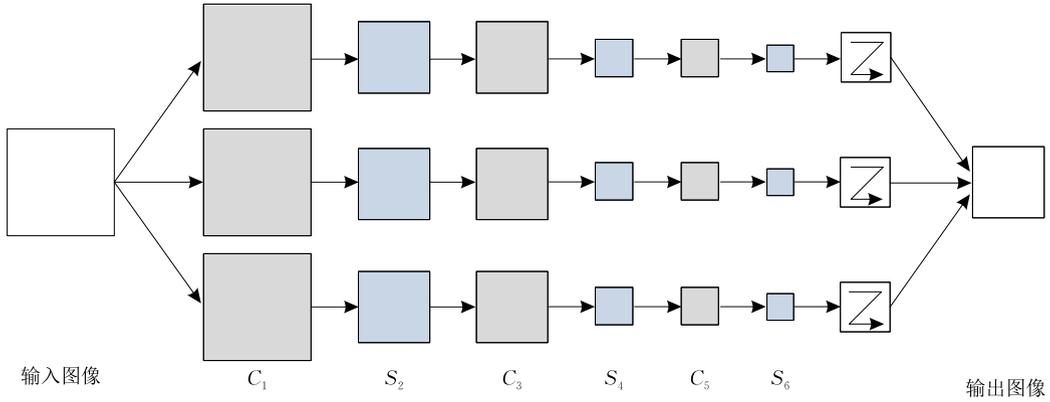


图 1 卷积神经网络结构示意图

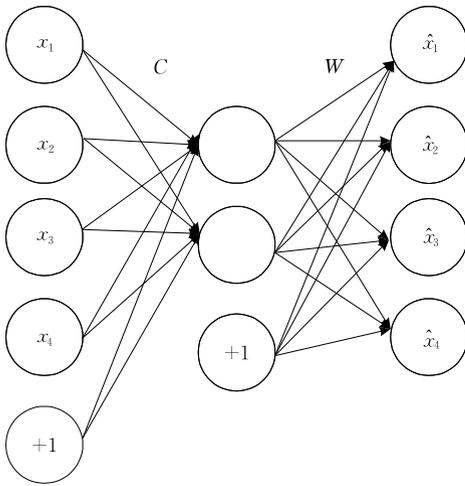


图 2 深度支撑值学习网络基本单元

该基本网络包括卷积层和线性层,令输入  $X \in R^d (y \in R), R^d$  为输入空间,  $d$  为维数,通过卷积操作  $C$  后,再经过隐藏层神经元的传递函数进行映射  $\phi(X); R^d \rightarrow R^q, q$  为特征空间的维数,其输出经过线性层,通过权值  $W$  和偏置  $b$  后得到输出结果.在基本单元的训练中,输入训练数据  $\{x_i, y_i\}_{i=1}^N$ ,输出的估计函数为  $f(C, W, x_i) = W^T \phi(CX) + b$ . 定义目标函数:

$$R(C, W) = \sum_{i=1}^N L[y_i, f(C, W, x_i)] + \gamma \frac{\|W\|^2}{2} \quad (1)$$

其中  $L[y_i, f(C, W, x_i)] = [y_i - W^T \phi(CX) - b]^2$ .

基于深度神经网络中表征学习的思路,令期望输出等于输入.类似于 SVM 中求解权值的思路,首先将基本网络的估计函数写为

$$f(X) = \sum_{i=1}^N \alpha_i K(CX, CX_i) + b \quad (2)$$

这里  $K(CX, CX_i) = \phi(CX)^T \phi(CX_i), i = 1, \dots, N$  是核函数,  $\alpha_i$  是支撑矢量  $X_i$  的支撑值.将估计函数写成矩阵形式:

$$\begin{bmatrix} 0, & \mathbf{1}^T \\ \mathbf{1}, & \Omega \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix} \quad (3)$$

其中:  $\Omega_{ij} = K_{ij} + I_{ij}/\gamma; K_{ij} = K(CX_i, CX_j); \mathbf{Y} = [y_1, \dots, y_N]^T, \alpha = [\alpha_1, \dots, \alpha_N]^T; \mathbf{1} = [1, \dots, 1]^T$ . 采用交替迭代算法优化网络参数,更新过程如下:首先固定  $C$ ,计算  $\alpha, b$  的解析解:

$$\begin{cases} b = \frac{\mathbf{1}^T \Omega^{-1} \mathbf{Y}}{\mathbf{1}^T \Omega^{-1} \mathbf{1}} \\ \alpha = \Omega^{-1} (\mathbf{Y} - b \mathbf{1}) \end{cases} \quad (4)$$

下一步,令  $A = \Omega^{-1}, \mathbf{B}^T = \frac{\mathbf{1}^T \Omega^{-1}}{\mathbf{1}^T \Omega^{-1} \mathbf{1}}$ ,则式(4)进一步表示为

$$\begin{cases} b = \mathbf{B}^T \mathbf{Y}, \\ \alpha = A(\mathbf{Y} - b \mathbf{1}) = A(I - \mathbf{1} \mathbf{B}^T) \mathbf{Y} \end{cases} \quad (5)$$

定义  $\mathbf{Q} = A(I - \mathbf{1} \mathbf{B}^T), \mathbf{Q}$  是一个  $(N \times N)$  的矩阵,假设各像素点的支撑值与以该点为中心的映射领域的中心点的支撑值相似,则取  $\mathbf{Q}$  的中央行向量组成方阵,获得卷积层中的权值  $C$ .在网络训练完毕之后,取出卷积层的输出,可以得到一系列的支撑值图像  $\{I_1, I_2, \dots, I_k, \dots, I_l\}$  和序列近似图像,序列近似图像则由原始图像  $P$  与它的支持值图像的差得到

$$\begin{cases} P_1 = P; \\ I_k = F_k \times P_k; \\ P_{k+1} = P_k - I_k \end{cases} \quad (6)$$

其中:  $k(k = 1, 2, \dots, l)$  为网络模型的隐含层层数;  $\{F_1, F_2, \dots, F_k, \dots, F_l\}$  为序列支撑值滤波器;“ $\times$ ”为卷积运算.

则图像的重构过程为

$$P = P_{l+1} + \sum_{k=1}^l I_k \quad (7)$$

### 3 基于 DSVL Nets 的融合算法

本文将深度学习理论用于遥感图像融合,基

于卷积神经网络结构构建了深度支撑值学习网络(Deep Support Value Learning Network, DSVL Nets),该网络模型在使用深度卷积神经网络时,去掉了隐藏层的下采样操作. DSVL Nets 网络模型可以有效选取输入图像在各级分解时使用的最优支撑值滤波器,使其能够获取图像的本质特征. 经 DSVL Nets 学习得到的序列卷积层图像和线性层图像,对序列卷积层图像采用绝对值取大法进行融合,得到融合后的序列卷积层图像;对线性层图像在过完备 DCT 字典下进行分块稀疏表示得到块稀疏系数,对稀疏系数使用绝对值取大法进行融合,进而得到融合后的线性层图像;最后,将融合后的序列卷积层图像与线性层图像按照式(7)进行重构,得到融合后的高分辨多光谱图像.

### 3.1 DSVL Nets 模型

全色图像与多光谱图像的融合要尽量保持光谱图像的光谱信息和全色图像的空间信息,因此本文对传统卷积神经网络结构(图 1)进行了修改,取消每一单元池化层  $S$  的下采样过程,将其修改为前层输出图像与卷积层图像  $C$  的差值层,修改后的网络结构保持了隐含层所有图像的大小与输入图像大小相同的特点,并将修改后的网络结构用于全色图像与多光谱图像的融合,网络计算过程如图 3 所示. 除去输入层和输出层,中间隐含层共 5 层,即由 5 对  $C$ - $S$  串接构成隐含层,其中  $C$  层为卷积层, $S$  层为线性层;首先对低分辨的多光谱图像进行上采样,使其大小与全色图像大小一致,将上采样后多光谱图像的 4 个波段与全色图像相拼接,构成网络模型的输入层;输出层输出的是网络结构的各卷积层图像以及最后一个隐藏层的线性层图像.

(1)  $C_1$  层为卷积层.  $C_1$  得到的图像为原始输入图像与学习得到的滤波器组卷积操作后的图像  $I_1$

( $I_1 = \{I_1^1, I_1^2, \dots, I_1^f\}$ ),  $f$  为滤波器的个数,  $I_1$  对应式(6)中的  $I_1$ .

(2)  $S_2$  为线性层. 获取对目标的局部平移不变性的描述,传统的卷积神经网络结构是对  $C_1$  层的图像进行下采样,进而得到  $S_2$  层的图像. 本文去掉了  $S_2$  层的下采样过程,这一层被视为线性层,它得到的是原始输入图像  $P$  与  $C_1$  层的卷积图像  $I_1$  的差值图像  $P_2$  ( $P_2 = \{P_2^1, P_2^2, \dots, P_2^f\}$ ),  $P_2$  对应式(6)中的  $P_2$ .

(3)  $C_3, C_5, C_7$  和  $C_9$  为卷积层. 实现原理与  $C_1$  层的实现原理相似,即将一组可训练滤波器组分别与  $S_2, S_4, S_6$  和  $S_8$  线性层图像进行卷积,分别得到卷积层图像  $C_3, C_5, C_7$  和  $C_9$ ,它们分别对应于式(6)中的  $I_k, k(k=2, 3, \dots, l)$  为网络模型的隐藏层的层数. 本文 DSVL Nets 网络结构包含 5 个隐藏层,故  $k$  的最大值为 5.

(4)  $S_4, S_6, S_8$  和  $S_{10}$  为线性层. 实现原理与  $S_2$  层的实现原理相似,将上一隐藏层的线性层图像视为该层的输入图像,将其与该层的卷积层图像相减,得到的差值即为该层的线性层图像,即将  $S_2, S_4, S_6$  和  $S_8$  层的图像分别与  $C_3, C_5, C_7$  和  $C_9$  层的图像相减,得到线性层的差值图像  $S_4, S_6, S_8$  和  $S_{10}$ ,它们分别对应于式(6)中的  $P_{k+1}$ .

(5) 输出层输出各卷积层图像以及  $S_{10}$  层的线性层图像.

文中采用自编码方式训练 DSVL Nets 网络模型. 自编码器可以通过无标签学习得到一个能更好地表示输入数据的特征模型. 它使用了反向传播算法,并让目标值等于输入值,自编码神经网络尝试学习一个  $h_{w,b}(x) \approx x$  的函数,它试图逼近一个恒等函数,从而使得输出  $\hat{x}$  接近于输入  $x$ . 网络训练过程为:依次训练网络的每一层,进而预训练整个深度神

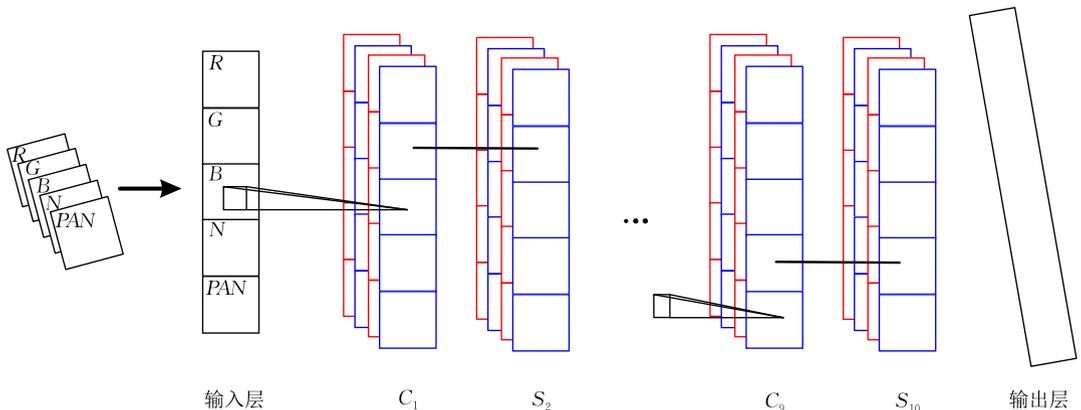


图 3 DSVL Nets 网络计算过程

经网络. 首先利用原始输入数据训练仅包含有一个隐藏层的网络结构, 训练结束之后扩展网络结构为包含有两个隐藏层的深度支撑值学习网络, 再次进行训练; 每一次训练过程都是采用自编码方式进行, 即固定前面  $t-1$  层, 增加第  $t$  层, 以此类推直到将整个网络训练完毕. 这些各层单独训练所得到的权重被用来初始化最终的深度网络的权重, 再对整个网络进行“微调”.

### 3.2 图像融合

由 3.1 节可知, 网络结构输出层输出的是多光谱图像 R、G、B、N 和全色图像 PAN 这 5 个波段图像的  $C_1, C_3, C_5, C_7, C_9$  和  $S_{10}$  层图像, 卷积层图像主要包含图像的边缘和轮廓等高频信息, 对其采用绝对值取大法进行融合, 得到融合后的各卷积层图像; 线性层主要包含图像的低频信息, 对其采用分块稀疏表示再融合的方法. 文中构造了过完备离散余弦变换 (Discrete Cosine Transform, DCT) 字典  $D_{DCT}$ , 将线性层图像分块后在  $D_{DCT}$  字典下使用正交匹配追踪 (Orthogonal Matching Pursuit, OMP) 算法进行稀疏表示, 对稀疏系数采用绝对值取大法进行融合, 得到融合后线性层图像; 最后, 将融合后的各卷积层图像和线性层图像按式 (7) 进行重构, 得到融合后的高分辨多光谱图像.

过完备 DCT 字典由 DCT 变换获得, 给定序号  $x(n) (n=0, 1, \dots, N-1)$ , 有

$$X_c(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \quad (8)$$

$$X_c(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N} \quad (9)$$

将其写成矩阵的形式:

$$X_c = C_N x \quad (10)$$

$C_N$  是  $N \times N$  变换矩阵, 其行向量为余弦基. 离散余弦变换后得到一个完备字典, 再将其扩展成为过完备离散余弦字典, 图 4 所示为已经构造好的过完备 DCT 字典.

全色图像与多光谱图像的融合过程具体描述如下:

(1) 由图 3 的 DSVL Nets 网络计算过程可以得到多光谱图像各波段的序列卷积层图像和线性层图像, 即  $C_1^d, C_3^d, C_5^d, C_7^d, C_9^d$  和  $S_{10}^d$ , 其中  $d(d=1, \dots, 4)$  表示为多光谱图像的波段数. 同时得到全色图像的序列卷积层图像和线性层图像, 即  $C_1^P, C_3^P, C_5^P, C_7^P, C_9^P$  和  $S_{10}^P$ , 其中  $P$  表示为全色图像;

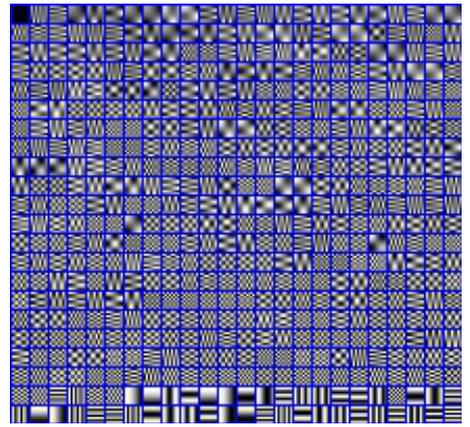


图 4 过完备 DCT 字典 ( $D_{DCT}$ )

(2) 将多光谱图像各波段的  $C_1^d, C_3^d, C_5^d, C_7^d$  和  $C_9^d$  图像与全色图像的  $C_1^P, C_3^P, C_5^P, C_7^P$  和  $C_9^P$  图像分别采用绝对值取大法进行融合, 得到融合后高分辨多光谱图像的  $C_1^{Fu}, C_3^{Fu}, C_5^{Fu}, C_7^{Fu}$  和  $C_9^{Fu}$  图像, 其中  $Fu$  表示为各波段融合后的图像;

(3) 将多光谱图像各波段的  $S_{10}^d$  图像与全色图像的  $S_{10}^P$  图像分别进行融合. 首先对各图像进行分块, 将图像块在  $D_{DCT}$  字典下稀疏表示, 对稀疏系数按绝对值取大法进行融合, 得到融合后高分辨多光谱图像的  $S_{10}^{Fu}$  图像;

(4) 将 (2) 和 (3) 得到的融合后的图像  $C_1^{Fu}, C_3^{Fu}, C_5^{Fu}, C_7^{Fu}, C_9^{Fu}$  和  $S_{10}^{Fu}$  按照式 (7) 进行重构, 最终得到融合后的高分辨多光谱图像.

其中 (1)~(4) 过程中和各卷积层图像  $C$  表示式 (7) 中图像  $I$ , 线性层图像  $S$  表示式 (7) 中的图像  $P$ .

## 4 实验结果与分析

文中使用 QuickBird 和 Geoeye 数据集, 与经典的主成分分析方法 (Principal Component Analysis, PCA)<sup>[2]</sup>、AWLP<sup>[9]</sup> 方法、PN-TSSC 方法<sup>[20]</sup> 和单尺度支撑值变换方法 (Support Value Transform, SVT)<sup>[14]</sup> 进行比较, 以验证本文所提方法的有效性, 各对比算法的参数设置与文章中的参数设置相同. 在对融合结果进行分析时采用了主观评价法与客观评价指标相比较的方法, 评价指标包括有相关系数 (CC), 通用图像质量指数 (UIQI), 均方根误差 (RMSE), 全局融合指标 ( $Q_4$ )<sup>[33]</sup>, 光谱角映射 (SAM)<sup>[34]</sup> 和全局相对光谱损失 (ERGAS)<sup>[35]</sup>.

(1) 相关系数 (CC)

相关系数 (Correlation Coefficient, CC) 反映的

是融合图像与参考图像对应谱带的相似程度,理想值为 1,值越接近于 1 表示融合后图像的光谱损失越少,与参考图像越相近.其表示形式为

$$CC = \frac{\sum_{i=1}^M \sum_{j=1}^N [F_{(i,j)} - \mu_F][X_{(i,j)} - \mu_X]}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N [F_{(i,j)} - \mu_F]^2 \sum_{i=1}^M \sum_{j=1}^N [X_{(i,j)} - \mu_X]^2}} \quad (11)$$

其中: $F_{(i,j)}$ 和 $X_{(i,j)}$ 分别表示图像大小均为 $M \times N$ 的融合后图像和参考图像; $\mu_F$ 和 $\mu_X$ 分别表示 $F_{(i,j)}$ 和 $X_{(i,j)}$ 的均值.

### (2) 通用图像质量指数(UIQI)

通用图像质量指数(UIQI)用于评价融合图像与参考图像的结构失真程度,其理想值为 1,越接近于 1 说明图像融合质量越好,UIQI 定义为

$$UIQI = \frac{\sigma_{FX}}{\sigma_F \sigma_X} \times \frac{2\mu_F \mu_X}{\mu_F^2 + \mu_X^2} \times \frac{2\sigma_F \sigma_X}{\sigma_F^2 + \sigma_X^2} \quad (12)$$

其中: $\sigma_{FX}$ 表示 $F_{(i,j)}$ 和 $X_{(i,j)}$ 的协方差; $\sigma_F$ 和 $\sigma_X$ 分别表示 $F_{(i,j)}$ 和 $X_{(i,j)}$ 的标准差.

### (3) 均方根误差(RMSE)

均方根误差(Root Mean Square Error, RMSE)表示融合图像与参考图像的差异程度,其理想值为 0,即 RMSE 值越小则表示融合结果越理想. RMSE 定义为

$$RMSE = \frac{1}{MN} \sqrt{\sum_{i=1}^M \sum_{j=1}^N [F_{(i,j)} - X_{(i,j)}]^2} \quad (13)$$

### (4) 全局融合指标(Q<sub>4</sub>)

全局融合指标(Q<sub>4</sub>)反映了融合图像与参考图像在空间和谱间的整体相似程度,融合结果越好, Q<sub>4</sub> 的值越接近于 1. 首先应对图像进行分块( $D \times D$ ), 对每一块按式(14)计算 Q<sub>4</sub>.

$$Q_4 = E(\|Q_{4(D \times D)}\|) \quad (14)$$

其中, $Q_{4(D \times D)}$ ,  $x$  和  $y$  由式(15)计算得到

$$\left\{ \begin{aligned} Q_{4(D \times D)} &= \frac{4[E(x \cdot y^*) - \bar{x} \cdot \bar{y}^*]}{E(\|x\|^2) - \bar{x} \cdot \bar{x} + E(\|y\|^2) - \bar{y} \cdot \bar{y}} \cdot \frac{\|\bar{x}\| \cdot \|\bar{y}\|}{\|\bar{x}\|^2 + \|\bar{y}\|^2} \\ x &= X_1(i, j) + \alpha X_2(i, j) + \beta X_3(i, j) + \gamma X_4(i, j) \\ y &= F_1(i, j) + \alpha F_2(i, j) + \beta F_3(i, j) + \gamma F_4(i, j) \end{aligned} \right. \quad (15)$$

### (5) 光谱角映射(SAM)

光谱角映射(Spectral Angle Mapper, SAM)表示融合图像与参考图像的光谱相近程度,其理想值为 0,即光谱越相近, SAM 越小. SAM 定义为

$$SAM = \arccos\left(\frac{\langle u_X, u_F \rangle}{\|u_X\|_2 \cdot \|u_F\|_2}\right) \quad (16)$$

其中, $u_F$ 和 $u_X$ 分别表示 $F_{(i,j)}$ 和 $X_{(i,j)}$ 的谱向量.

### (6) 全局相对光谱损失(ERGAS)

全局相对光谱损失(Erreur Relative Global Adimensionnelle de Synthèse, ERGAS)表示对融合图像各个谱带进行总体的性能评价,其理想值为 0, ERGAS 值越小,表示融合结果越好. ERGAS 定义为

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{k} \sum_{k=1}^K \left(\frac{RMSE}{\bar{x}_k}\right)^2} \quad (17)$$

其中: $\frac{h}{l}$ 表示全色图像与多光谱图像分辨率之比; $k$ 表示多光谱图像的谱带数, $\bar{x}_k$ 为各谱带的平均值.

## 4.1 QuickBird 图像集实验

QuickBird 卫星数据来源于印度的孙德尔本斯(拍摄于 2002 年 11 月 21 日)和中国的西安(拍摄于 2008 年 9 月 30 日),该卫星获取的全色图像的分辨率为 0.6-m,多光谱图像的分辨率为 2.4-m. 实验中将原始图像作为参考图像,将其进行采样因子为 4 的下采样,下采样后的图像作为实验图像,下采样后的全色图像的分辨率为 2.4-m,多光谱图像的分辨率为 9.6-m. 文中所用的多光谱图像和全色图像的大小均为  $64 \times 64$  和  $256 \times 256$ ,参考图像大小为  $256 \times 256$ ;图 5 所示为 10 组测试图像的多光谱图像,表 1 为这 10 组图像融合后的结果图像的平均数值指标;图 6 和图 7 所示为单幅图像的融合结果,其客观评价指标如表 2 和表 3 所示,表中最优指标均用加粗字体表示.

表 1 为来自 QuickBird 卫星的 10 组图像的平均数值指标,从该表中可以看出本文所提出的基于深度支撑值学习网络的图像融合方法得到的平均融合指标均大于对比算法中相应的数值指标. 图 6 所示为单幅图像的融合结果对比图,其中图 6(a)和(b)分别为多光谱图像和全色图像,(c)为参考图像,从图中可以看出 PCA 方法融合后的图像较暗,光谱扭曲较严重;AWLP 方法融合后的结果有较强的空间扭曲和光谱扭曲;PN-TSSC 方法的融合结果空间保持较好,但光谱信息有失真;单尺度 SVT 方法融合后的结果图像较模糊,且空间失真较严重;相较于对比算法,本文所提方法得到的融合图像较好地保持了图像光谱信息和空间信息. 表 2 为图 6 融合后的数值指标,从表 2 可以看出,本文方法的 CC, UIQI, RMSE, Q<sub>4</sub>, SAM 和 ERGAS 这 6 个指标均优于其它对比算法的相应数值指标.

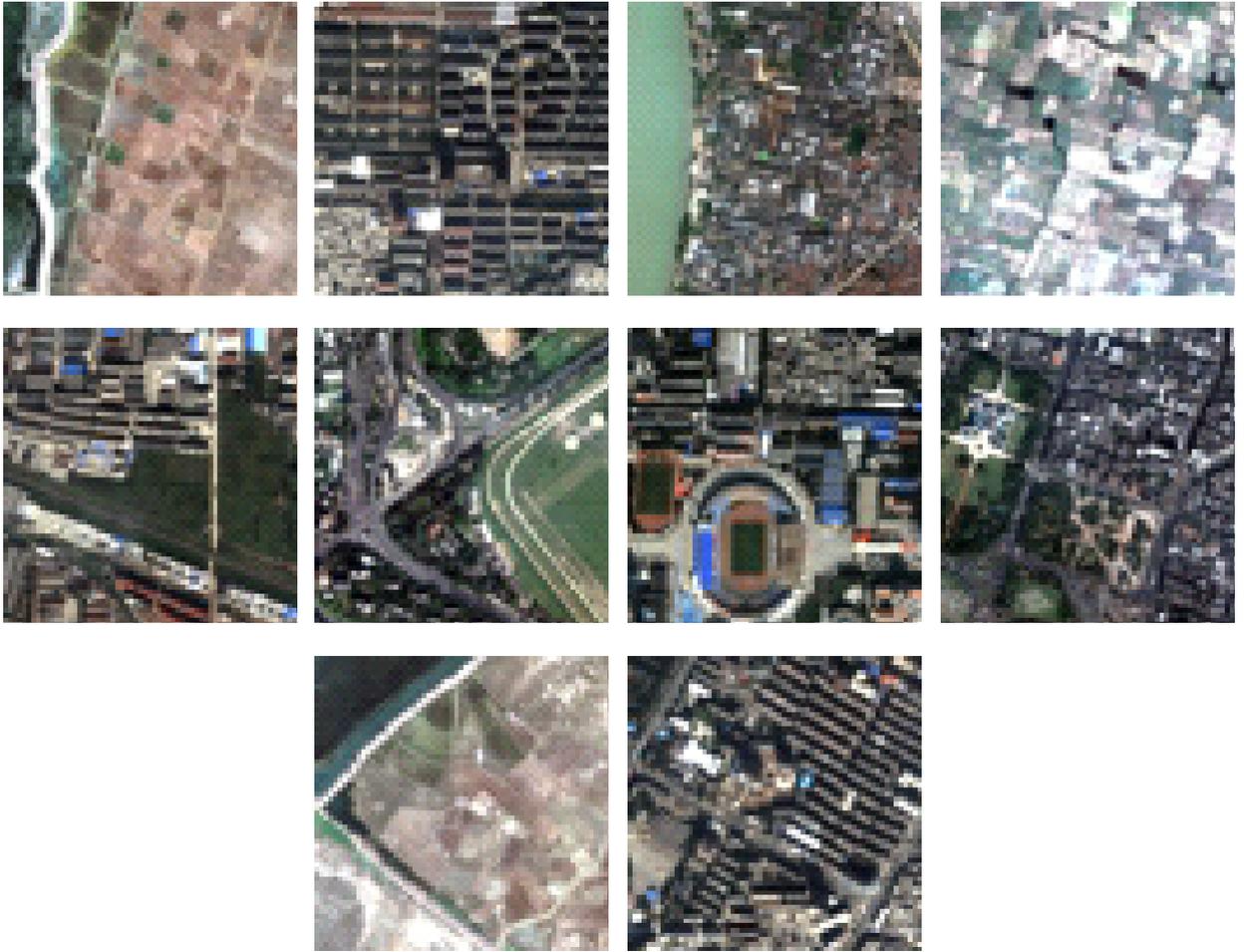


图 5 测试图像集

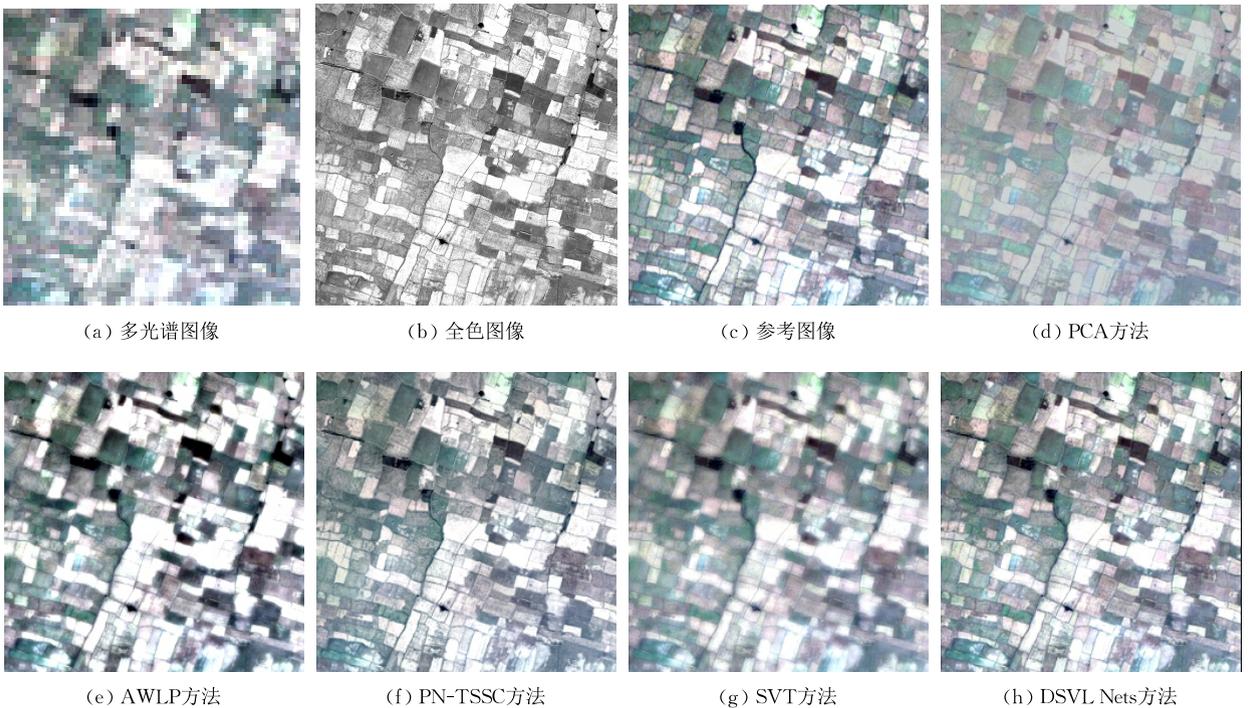


图 6 图像融合结果

表 1 图 5 的平均数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.8608	0.9188	0.9338	0.9389	<b>0.9482</b>
UIQI	0.6825	0.8766	0.9315	0.9317	<b>0.9475</b>
RMSE	0.1324	0.1260	0.0755	0.0736	<b>0.0680</b>
$Q_1$	0.6143	0.7906	0.8288	0.8063	<b>0.8487</b>
SAM	6.3610	14.5220	6.0721	4.6926	<b>4.6328</b>
ERGAS	4.2492	3.8451	2.4507	2.3590	<b>2.1893</b>

表 2 图 6 融合结果的数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.9116	0.9294	0.9488	0.9476	<b>0.9622</b>
UIQI	0.6833	0.8453	0.9480	0.9430	<b>0.9622</b>
RMSE	0.1184	0.1328	0.0604	0.0618	<b>0.0528</b>
$Q_1$	0.6487	0.8151	0.8944	0.8610	<b>0.9101</b>
SAM	3.5339	3.8653	2.8407	2.5603	<b>2.5341</b>
ERGAS	2.2542	2.4151	1.1475	1.1768	<b>1.0082</b>

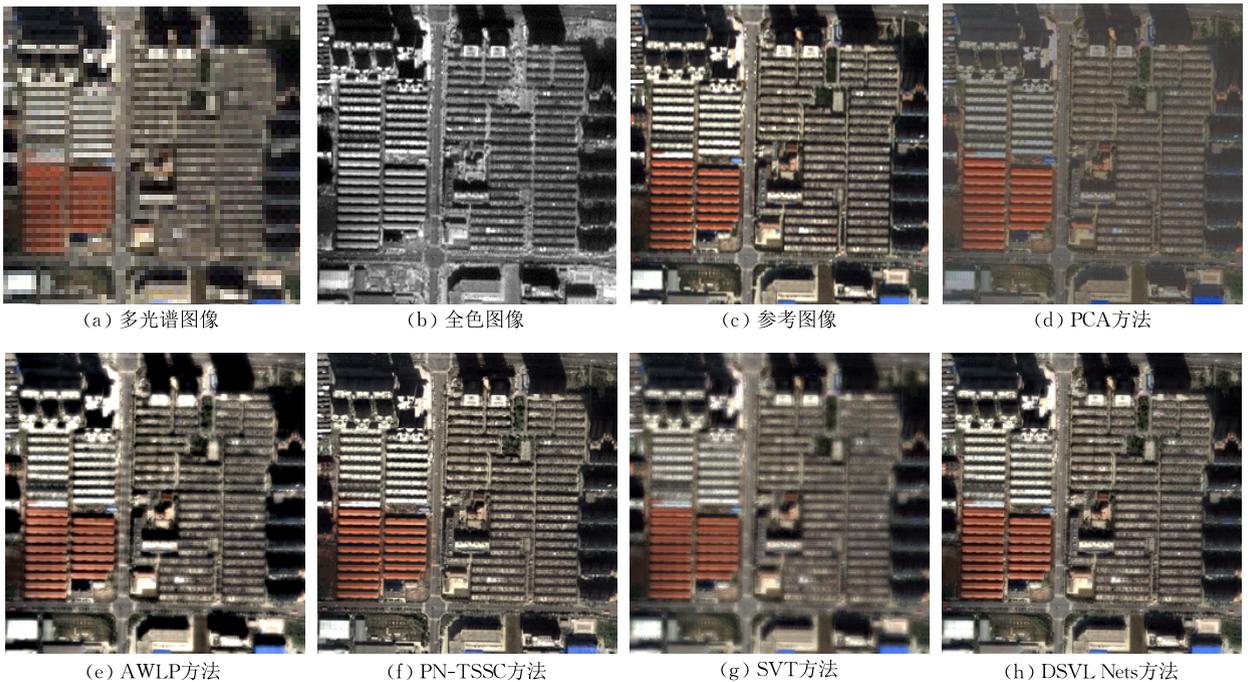


图 7 图像融合结果

表 3 图 7 融合结果的数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.9077	0.9268	0.9377	0.9083	<b>0.9495</b>
UIQI	0.6770	0.9006	0.9376	0.8910	<b>0.9473</b>
RMSE	0.1447	0.1183	0.0821	0.0993	<b>0.0729</b>
$Q_1$	0.6076	0.8301	0.8813	0.7546	<b>0.8816</b>
SAM	6.1020	22.5028	7.2061	5.6044	<b>5.5500</b>
ERGAS	5.2152	4.2548	2.9522	3.5671	<b>2.6242</b>

从图 7 可以看到,基于 PCA 方法融合后的图像较暗,光谱扭曲较严重;基于 AWLP 方法融合后的结果有边界模糊;单尺度 SVT 方法的方法融合图像模糊,空间扭曲较严重;而 PN-TSSC 方法和本文方法的结果图像光谱信息较接近于参考图像,融合效果较好.表 3 为图 7 融合后的数值指标,从表 3 可以看出,

本文方法的  $CC, UIQI, RMSE, Q_1, SAM$  和  $ERGAS$  这 6 个指标均优于其它对比算法的相应数值指标.

#### 4.2 Geosy 图像集实验

Geosy 卫星数据来源于澳大利亚的霍巴特(拍摄于 2009 年 2 月 24 日),此卫星获取的全色图像的分辨率为 0.5-m,多光谱图像的分辨率为 2-m. 实验中

将原始图像作为参考图像,将其进行采样因子为 4 的下采样,下采样后的图像作为实验图像,下采样后的全色图像的分辨率为 2-m,多光谱图像的分辨率为 8-m.文中所用的多光谱图像大小均为  $64 \times 64$ ,全色

图像大小均为  $256 \times 256$ ,参考图像大小为  $256 \times 256$ ;图 8 所示为 10 组测试图像的多光谱图像,表 4 为这 10 组图像融合后的结果图像的平均数值指标;图 9 和图 10 所示为单幅图像的融合结果,其客观评价指



图 8 测试图像集

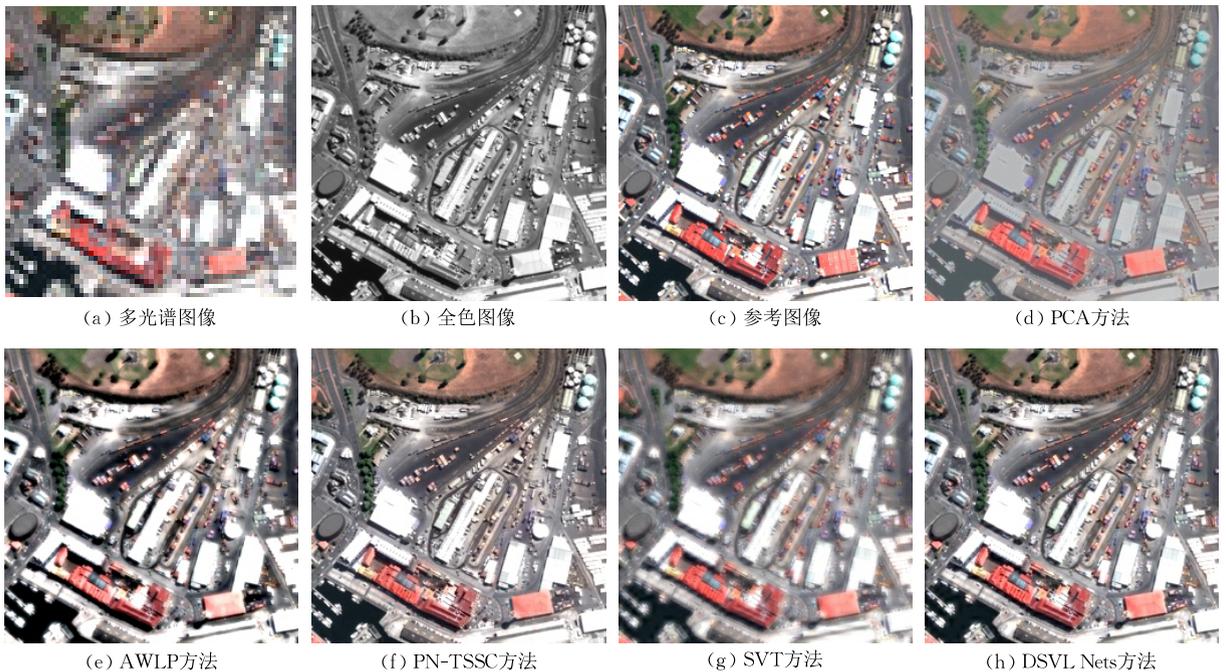


图 9 图像融合结果

表 4 图 8 的平均数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.9184	0.9229	0.9401	0.9302	<b>0.9529</b>
UIQI	0.7360	0.8861	0.9371	0.9186	<b>0.9522</b>
RMSE	0.1404	0.1415	0.0775	0.0901	<b>0.0709</b>
$Q_4$	0.6139	0.7619	<b>0.8493</b>	0.7677	0.8408
SAM	6.3091	16.1864	6.6955	5.9704	<b>5.6139</b>
ERGAS	3.8808	3.9480	2.1792	2.4908	<b>1.9712</b>

表 5 图 9 融合结果的数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.9404	0.9444	0.9707	0.9577	<b>0.9740</b>
UIQI	0.7704	0.8879	0.9704	0.9443	<b>0.9739</b>
RMSE	0.1366	0.1472	0.0597	0.0769	<b>0.0576</b>
$Q_4$	0.6728	0.8184	<b>0.9058</b>	0.8820	0.8955
SAM	3.7268	11.4881	4.2739	5.3056	<b>3.9421</b>
ERGAS	3.2862	3.5461	1.4816	1.9059	<b>1.4096</b>

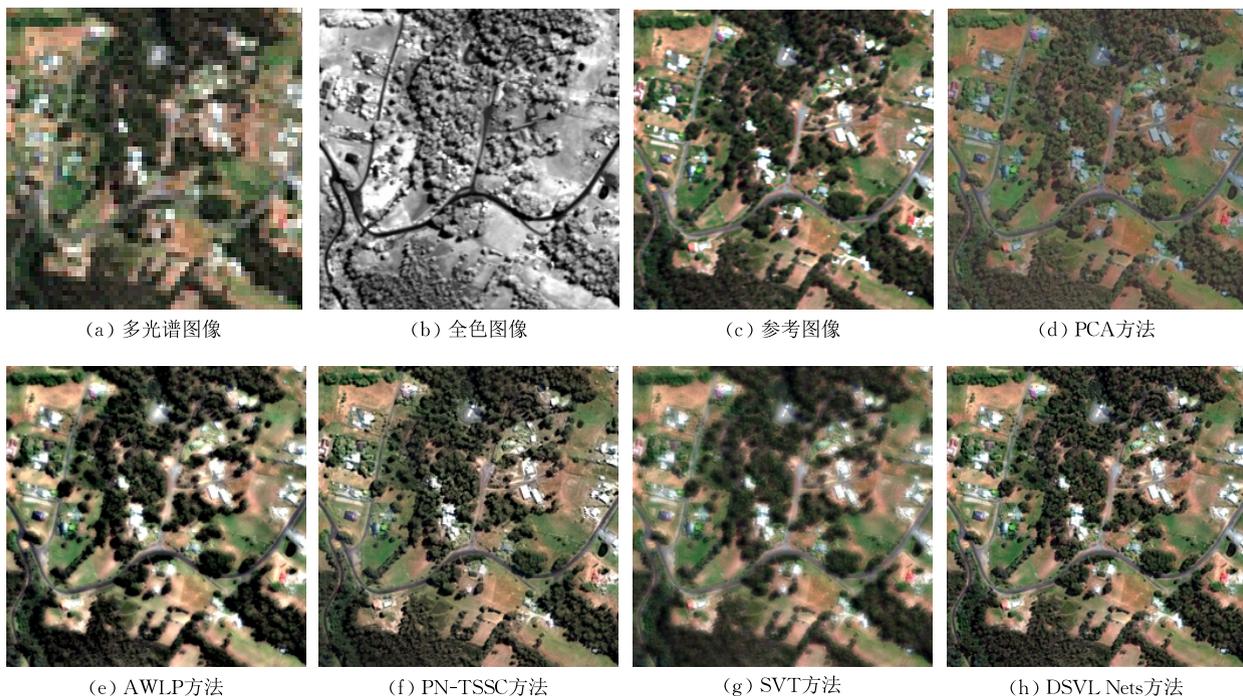


图 10 图像融合结果

表 6 图 10 融合结果的数值指标

指标	PCA 方法	AWLP 方法	PN-TSSC 方法	SVT 方法	DSVL Nets 方法
CC	0.9122	0.9242	0.9205	0.9488	<b>0.9554</b>
UIQI	0.7688	0.8903	0.9176	0.9431	<b>0.9550</b>
RMSE	0.1218	0.1225	0.0884	0.0711	<b>0.0664</b>
$Q_4$	0.6245	0.7407	0.7770	0.7328	<b>0.7785</b>
SAM	8.9343	17.5408	11.9260	<b>6.5865</b>	8.3798
ERGAS	4.0555	3.8763	2.8988	2.2499	<b>2.1647</b>

标如表 5 和表 6 所示, 表中最优指标均用加粗字体表示。

表 4 为来自 Geoeye 卫星的 10 组图像的平均数值指标, 从该表中可以看出本文所提出的基于深度支撑值学习网络的图像融合方法中的  $CC$ ,  $UIQI$ ,

$RMSE$ ,  $SAM$  和  $ERGAS$  这 5 个指标均优于对比方法的相应指标, 而  $PN-TSSC$  方法的  $Q_4$  指标最优, 本文方法的  $Q_4$  指标次之。图 9 所示为单幅图像的融合结果对比图, 其中图 9(a) 和 (b) 分别为多光谱图像和全色图像, (c) 为参考图像, 从图中可以看出

PCA 方法融合后的图像较暗,光谱扭曲较严重;AWLP 方法融合后的图像边界较模糊,有较少的空间扭曲和光谱扭曲;PN-TSSC 方法的结果空间保持较好,但图像较暗;单尺度 SVT 方法融合后的结果图像较模糊,空间失真较严重;而本文方法的结果比较接近于参考图像,它较好地保持了光谱信息和空间信息,表 5 为图 8 中各对比算法的数值指标,从表 5 中可以看出,本文方法的  $CC$ 、 $UIQI$ 、 $RMSE$ 、 $SAM$  和  $ERGAS$  这 5 个指标均优于对比方法的相应指标,而 PN-TSSC 方法的  $Q_4$  指标最优,本文方法的  $Q_4$  指标次优,但从视觉效果上看本文方法所得到的结果图像光谱信息和空间信息保持最优。

从图 10 可以看出,基于 PCA 方法融合后的图像较暗,光谱扭曲较严重;单尺度 SVT 方法融合后的结果图像较模糊,空间失真较严重;AWLP 方法的光谱信息较接近于参考图像,但图像边界较模糊;从视觉效果上看 PN-TSSC 方法和本文方法的结果空间保持较好,且光谱信息较接近于参考图像。但对图像从客观指标上整体评价,可以看出本文方法的  $CC$ 、 $UIQI$ 、 $RMSE$ 、 $Q_4$  和  $ERGAS$  这 5 个指标均优于对比方法的相应指标,而单尺度 SVT 方法的  $Q_4$  指标最优,本文方法的  $Q_4$  指标次优,但从视觉效果上看本文方法所得到的融合结果图像空间信息最优,光谱信息最接近于参考图像。

## 5 结 论

本文提出了基于深度支撑值学习网络(DSVL Nets)模型的全色图像与多光谱图像融合方法,网络模型各层中的滤波器通过最小化经验风险获得系数。DSVL Nets 网络模型取消了传统卷积神经网络结构中特征映射层的下采样过程,构建了由 5 个隐藏层构成的适合图像融合的新网络模型,每一层的基本结构包括卷积层和线性层,这样的基本单元提供了一种对输入图像的多尺度、多方向、各向异性、非下采样的冗余变换。为了验证本文方法的有效性,文中使用了 QuickBird 和 Geoeye 卫星数据集,实验结果表明,与 PCA、AWLP、PN-TSSC 和 SVT 算法相比较,本文所提方法的融合结果无论在主观视觉还是客观评价指标上均优于对比算法,较好地保持了图像的光谱信息和空间信息。

**致 谢** 感谢陕西省智能感知与大数据协同创新中心、教育部创新团队和首批陕西省科技创新团队计划以及国家“九七三”重点基础研究发展计划项目和国

家自然科学基金委重大研究计划的支持。计算机学报编辑部和评委老师给出了宝贵意见,在此一并感谢!

## 参 考 文 献

- [1] Thomas C, Ranchin T, Wald L, Chanussot J. Synthesis of multispectral Images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(5): 1301-1312
- [2] Chavez P S, Sides S C, Anderson J A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and Remote Sensing*, 1991, 57(3): 295-303
- [3] Jia Yong-Hong. Fusion of landsat TM and SAR images based on principal component analysis. *Remote Sensing Technology and Application*, 1998, 13(1): 46-49(in Chinese)  
(贾永红. TM 和 SAR 影像主分量变换融合法. *遥感技术与应用*, 1998, 13(1): 46-49)
- [4] Carper W J, Lillesand T M, Kiefer P W. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and Remote Sensing*, 1990, 56(4): 459-467
- [5] Tu T M, Su S C, Shyu H C, Huang P S. A new look at IHS-like image fusion methods. *Information Fusion*, 2001, 2(3): 177-186
- [6] Tu T M, Huang P S, Hung C L, Chang C P. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. *IEEE Geoscience and Remote Sensing Letters*, 2004, 1(4): 309-312
- [7] Craig A L, Bernard V B. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. *United States Patent*, 6 011 875, 2000
- [8] Li S T, Kwok J T, Wang Y N. Using the discrete wavelet frame transform to merge landsat TM and SPOT panchromatic images. *Information Fusion*, 2002, 3(1): 17-23
- [9] Otazu X, Gonzalez-Audicana M, Fors O, Nunez J. Introduction of sensor spectral response into image fusion methods. Application to Wavelet-Based Methods. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(10): 2376-2385
- [10] Nencini F, Garzelli A, Baronti S, Alparone L. Remote sensing image fusion using the curvelet transform. *Information Fusion*, 2007, 8(2): 143-156
- [11] Shah V P, Younan, N H, King R. Pan-sharpening via the contourlet transform//*Proceedings of IEEE International Geoscience and Remote Sensing Symposium*. Barcelona, Spain, 2007: 310-313
- [12] Amro I, Mateos J. Multispectral image pansharpening based on the contourlet transform. *Information Optics and Photonics*, 2010, 206(1): 247-261
- [13] Zheng S, Shi W Z, Liu J, et al. Multisource image fusion

- method using support value transform. *IEEE Transactions on Image Processing*, 2007, 16(7): 1831-1839
- [14] Zheng S, Shi W Z, Liu J, Tian J W. Remote sensing image fusion using multiscale mapped LS-SVM. *IEEE Transactions Geoscience and Remote Sensing*, 2008, 46(5): 1313-1322
- [15] Yang S Y, Wang M, Jiao L C. Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis. *Information Fusion*, 2012, 13(3): 177-184
- [16] Li S T, Yang B. A new pan-sharpening method using a compressed sensing technique. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(2): 738-746
- [17] Jiang C, Zhang H Y, Shen H F, Zhang L P. A practical compressed sensing-based pan-sharpening method. *IEEE Geoscience and Remote Sensing Letter*, 2012, 9(4): 629-633
- [18] Li S T, Yin H T, Fang L Y. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(9): 4779-4789
- [19] Zhu X X, Bamler R. A sparse image fusion algorithm with application to pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(5): 2827-2836
- [20] Jiang C, Zhang H Y, Shen H F, Zhang L P. Two-step sparse coding for the pan-sharpening of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(5): 1792-1805
- [21] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504-507
- [22] Xie J Y, Xu L L, Chen E H. Image denoising and inpainting with deep neural networks//*Proceedings of the Neural Information Processing Systems*. Lake Tahoe, USA, 2012: 350-358
- [23] Agostinelli F, Anderson M R, Lee H. Robust image denoising with multi-column deep neural networks//*Proceedings of the Neural Information Processing Systems*. Lake Tahoe, USA, 2013: 1493-1501
- [24] Ji S W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221-231
- [25] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324
- [26] Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting and composing robust features with denoising autoencoders//*Proceedings of the 25th International Conference on Machine Learning*. ACM New York, USA, 2008: 1096-1103
- [27] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 2010, 11(6): 3371-3408
- [28] Li D, Dong Y. Deep convex net: A scalable architecture for speech pattern classification//*Proceedings of the 12th Annual Conference of International Speech Communication Association*. Florence, Italy, 2011: 2285-2288
- [29] Poon H, Domingos P. Sum-product networks: A new deep architecture//*Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. Barcelona, Spain, 2011: 689-690
- [30] Zhao Zhi-Hong, Yang Shao-Pu, Ma Zeng-Qiang. License plate character recognition based on convolutional neural network LeNet-5. *Journal of System Simulation*, 2010, 22(3): 638-641(in Chinese)  
(赵志宏, 杨绍普, 马增强. 基于卷积神经网络 LeNet-5 的车牌字符识别研究. *系统仿真学报*, 2010, 22(3): 638-641)
- [31] Ossama A H, Mohamed A R, Jiang H, et al. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio speech and Language Processing*, 2014, 22(22): 1533-1545
- [32] Smirnov E A, Timoshenko D M, Andrianov S N. Comparison of regularization methods for ImageNet classification with deep convolutional neural network. *AASRI Procedia*, 2014, 6(1): 89-94
- [33] Alparone L, Baronti S, Garzelli A, Nencini F. A global quality measurement of pan-sharpened multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 2004, 1(4): 313-317
- [34] Alparone L, Wald L, Chanussot J, et al. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(10): 3012-3021
- [35] Wald L, Ranchin T, Mangolini M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 1997, 63(6): 691-699



**LI Hong**, born in 1976, Ph.D. candidate, associate professor. Her research interests include image processing, pattern recognition.

**LIU Fang**, born in 1963, professor. Her research interests include intelligent signal processing, pattern recognition.

**YANG Shu-Yuan**, born in 1978, Ph.D., professor. Her research interests include intelligent signal and image processing, machine learning.

**ZHANG Kai**, born in 1992, Ph.D. candidate. His research interests include image fusion, machine learning.

## Background

Many remote sensing applications require images with both high spatial and high spectral resolution. Due to technological limitations of remoter sensors, these images can not be obtained easily. And the earth observation satellites are composed of several multispectral (MS) channels and a panchromatic (PAN) channel. The MS images provide the more spectral information and less spatial information. The PAN image has the accurate geometric information. The remote sensing image fusion aims at creating a fused image of high spatial and spectral resolutions through merging a panchromatic image with a multispectral image. The common methods are component substitution, multiscale analysis and restoration based methods. Component substitution methods can recover high resolution MS images via the substitution. However, the spectral distortion is remarkable in the fused image for the complicated spectral response between low resolution MS and PAN images. Multiscale analysis methods have the characteristic of multiscale, multidirectional and anisotropy. However, the spectral distortion always arises for the mismatch of injected spatial details extracted from PAN image with that of low resolution MS images. In very recent years, restoration-based methods have been proved to achieve efficient pan-sharpening and become more and more popular. And several restoration-based pan-sharpening methods are developed by establishing representation models for source images.

Inspired by the deep learning and convolutional neural networks, a novel approach based on Deep Support Value

Learning Networks (DSVL Nets) is proposed. The loss function based on structural risk minimization is used in the training deep learning network. In order to avoid the loss of information, we abandon the downsampling of feature mapping layer of traditional convolution neural network. The DSVL Nets contains five hidden layers, which each layer consists of convolution layer and linear layer. And DSVL Nets provides a redundant transform which is multi-scale, multi-direction, anisotropy and non-sampled. Experimental results show the proposed method achieves better fusion efficiency, and reduces the distortion in both the spectral and spatial domain.

This work was supported by the National Key Basic Research and Department (973) Program of China (2013CB329402), the National Natural Science Foundation of China (61573267, 61173090), the Major Research Plan of the National Natural Science Foundation of China (91438201, 91438103), the Fund for Foreign Scholars in University Research and Teaching Programs (111 Project) (B07048), the Fundamental Research Funds for the Central Universities (JB140317, BDY021429), and the Science Basic Research Program in Shaanxi Province of China (16JK1823).

This group has been working on image processing, natural computing and machine learning for many years. So far, they have obtained some achievements. And parts of the research results have been published in the international proceeding and academic journals.