

DOI:10.11918/202005038

采用多层卷积稀疏表示的红外与可见光图像融合

张洲宇¹, 曹云峰¹, 丁萌², 陶江¹

(1. 南京航空航天大学 航天学院, 南京 210016; 2. 南京航空航天大学 民航学院, 南京 211106)

摘要: 为提升光学图像在低照度条件下的可用性,对红外图像与可见光图像进行融合从而结合两者的优势是一种有效的技术手段。稀疏表示理论在红外与可见光图像融合领域已经得到了广泛的应用,然而基于稀疏表示理论的图像融合方法所采用的局部建模方式易于导致语义信息损失和对误匹配的低容忍度两大缺陷。卷积稀疏表示的全局建模能力对克服上述不足具有巨大的优势,本研究借鉴卷积神经网络的结构设计了一种前馈式基于多层卷积稀疏表示的红外与可见光图像融合网络,该网络包含5层:第1、2层为卷积稀疏层,通过预训练的字典滤波器获取源图像的卷积稀疏响应;第3层为融合层,通过活性图评价以获取融合结果;第4、5层为重建层,基于融合结果结合字典滤波器重建融合图像。实验结果表明,所提出的图像融合方法有效抑制了稀疏表示理论应用于图像融合的两大不足,在客观评价指标方面明显优于基于稀疏表示、基于单层卷积稀疏表示和基于卷积神经网络的图像融合算法,在算法的计算复杂度和运行时间方面优于基于稀疏表示和基于卷积神经网络的图像融合算法。

关键词: 图像融合;卷积稀疏表示;稀疏表示;神经网络;红外图像

中图分类号: V249.3, TP391 文献标志码: A 文章编号: 0367-6234(2021)12-0051-09

Infrared and visible image fusion via multi-layer convolutional sparse representation

ZHANG Zhouyu¹, CAO Yunfeng¹, DING Meng², TAO Jiang¹

(1. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Integrating the advantages of infrared and visible images by image fusion is an effective means to enhance the applicability of optical images in low illumination conditions. Despite the wide application of sparse representation (SR) theory in the field of infrared and visible image fusion, the drawbacks including detail loss and low toleration with mis-registration caused by the local patch representation nature of SR have never been effectively solved. Different from SR, the global representation capability of the recently emerged convolutional sparse representation (CSR) model reveals huge potential to overcome the above mentioned deficiencies. Drawing on the convolutional neural network (CNN) architecture, a multi-layer CSR model was designed for pixel level image fusion. The image fusion model was constructed with five layers in a forward-feeding manner: the first two layers are CSR layers which acquire sparse coefficient maps with response to the pre-learned dictionary filter sets; the third layer is fusion layer which obtains fused results of the sparse coefficient maps; the last two are reconstruction layers which reconstruct the fused image step by step, and the fusion results are thus obtained. Experimental results indicate that the image fusion method proposed in this paper can effectively overcome the two drawbacks of SR. The method outperforms SR, CSR, and CNN in the aspect of objective assessment metrics, and outperforms SR and CNN in terms of computation complexity and computation time.

Keywords: image fusion; convolutional sparse representation; sparse representation; neural network; infrared image

由于传感器成像特性的差异,红外传感器与可

见光传感器拍摄的图像往往具有较强的互补性^[1]。

红外图像在低能见度条件下依然可以清晰的捕获目标,但是图像的边缘、纹理等细节不够丰富。可见光图像具有较强的细节保存能力,但是成像质量极易受到光照变化的影响。为了最大程度地结合两种图像类型的优点,将红外图像与可见光图像进行融合从而弥补传感器的固有缺陷成为了一条行之有效的思路。红外图像与可见光图像融合目前已广泛地应

收稿日期: 2020-05-11

基金项目: 国家自然科学基金(61673211);中央高校基本科研业务费(NP2019105);江苏省研究生科研与实践创新计划(KYCX18_0301);南京航空航天大学博士学位论文创新与创优基金(BCXJ18-11)

作者简介: 张洲宇(1992—),男,博士研究生;
曹云峰(1964—),男,教授,博士生导师

通信作者: 曹云峰, cyfac@nuaa.edu.cn

用于智能监控、目标监视、视频分析等领域^[2]。

现有的图像融合算法根据所选取的融合信息可分为像素级、特征级、决策级 3 个层级,其中像素级图像融合直接对源图像进行融合,所保存的细节信息最为完备,因此在过去数十年中得到了广泛的关注^[3]。像素级图像融合又可以进一步分为空间域融合与变换域融合两类,空间域融合无需对图像进行任何变换操作,因此实现较为简单,但是易于导致融合图像低对比度及块状分布等缺陷。为了克服上述不足,变换域图像融合采取的方案是对源图像进行适当的图像变换,并对转换后的图像信息进行融合。变换域图像融合的基本流程分为:1) 图像变换,将源图像由空间域映射到变换域;2) 活性度衡量,衡量变换后向量的重要程度;3) 融合规则设计,采取恰当的融合规则完成活性层的筛选。显然,变换域图像融合的核心在于由空间域到变换域的变换方式^[4]。

早期的变换域图像融合所采取的变换方法包括小波变换、金字塔变换等,通过人工构建的小波基与金字塔基实现源图像的变换。由于人工构建的图像变换的图像表示能力始终有限,该类方法始终难以获取图像中所有隐藏的信息。为克服早期图像变换的局限性,压缩感知 (compressive sensing, CS) 理论通过预训练的过完备字典实现源图像的变换,由于过完备字典是从大量的训练样本中学习得到的,CS 理论的图像表示能力具有显著的优势,在过去 10 年中受到了广泛的研究^[5]。基于 CS 理论的图像融合方法可分为 3 个步骤:首先将源图像分解为若干均等大小的图像块,对每一个图像块利用预训练的过完备字典计算与其对应的稀疏向量;其次,对同一位置对应的稀疏向量进行融合,获取融合后的向量;最后,利用融合后的向量结合过完备字典完成融合图像的重建。CS 理论应用于图像融合最大的不足在于:1) 基于图像块的建模方式破坏了源图像的语义结构,难以有效提取源图像中包含的空间上下文信息。2) 基于图像块的建模、融合与重建对于两幅图像的匹配关系提出了严苛的要求,因此对于误匹配的容忍度较低^[6]。为克服上述不足,近年来的相关研究工作开始尝试使用全局建模的方式对源图像进行变换,其中最为有效的变换方式为卷积神经网络 (convolutional neural network, CNN) 与卷积稀疏表示 (convolutional sparse representation, CSR)。

CNN 根据所需解决任务属性的不同可以分为分类式 CNN 与回归式 CNN 两类^[7]。分类式 CNN 目前已在视觉目标检测、识别与分类等任务中得到了广泛应用,传统的视觉识别流程总体上可分为 3 个步骤,即特征表示、特征选取与特征分类,分类

式 CNN 最大的优势在于通过多层卷积神经网络将上述 3 个步骤联合实现^[8-12]。Liu 等^[4]指出,早期的变换域图像融合方法所包含的 3 个步骤与传统视觉识别的 3 个步骤具有极大的相似性,因此将分类式 CNN 应用于图像融合成为了一条可行的思路。回归式 CNN 又称为全卷积神经网络 (fully convolutional neural network, FCN),一般采用端对端的方式实现视觉信息的分析与处理。回归式 CNN 目前已在低级视觉任务中得到了大量应用,例如图像分割、超分辨率重建等。因此,如果给定融合图像的真值图,即可训练回归式 CNN 通过端对端的方式获取融合结果。然而,无论是分类式 CNN 还是回归式 CNN,其共性不足在于难以获取大量带标签的训练样本用于网络训练^[13]。

CSR 的基本思路源于 Zeiler 等^[14]所设计的反卷积网络,其目的在于通过非监督的方式从自然图像中提取中级与高级特征。CSR 的基本原理是通过一组预先训练的卷积字典滤波器将源图像分解为一系列卷积稀疏响应图,每一张卷积稀疏响应图都包含了目标不同层级的信息。作为一种非监督机器学习方法,CSR 已被成功地用于解决许多视觉处理任务,例如目标跟踪、背景建模以及图像去噪等。在图像融合领域,CSR 可以视为一种有效的图像变换方法^[15]。由于 CSR 的图像建模方式无需对源图像进行分解,避免了基于稀疏表示 (sparse representation, SR) 的图像融合算法^[3]的局部建模所带来的语义信息缺失与对误匹配的低容忍度两大缺陷,因此 CSR 已在图像融合领域得到了成功的应用。Zeiler 等^[14]指出,当基于 CSR 的反卷积网络的层数加深时,网络学习得到的图像特征将由边缘向整个目标转移,考虑到红外与可见光图像融合的初衷在于凸显源图像中的目标,为此本文设计了一种面向红外与可见光图像融合的多层卷积稀疏表示网络。

1 面向红外与可见光图像融合的多层卷积稀疏网络结构

本文所设计的面向红外与可见光图像融合的多层卷积稀疏网络如图 1 所示,该网络共包含 5 层,采用前馈的方式实现红外与可见光源图像的融合。

网络的第 1、2 层为卷积稀疏层,通过预先训练的多层字典滤波器将源图像变换为一组卷积稀疏响应图。网络的第 3 层为融合层,通过对活性度衡量以获得卷积稀疏响应图的融合结果。网络的第 4、5 层为重建层,通过融合后的卷积稀疏响应图结合预先训练的多层字典滤波器实现融合图像的重建。

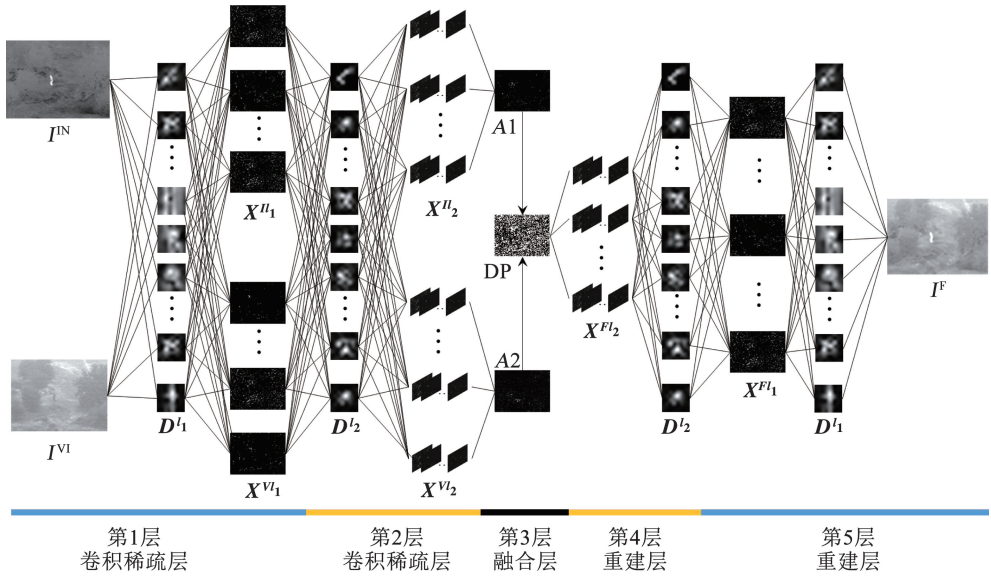


图 1 基于多层卷积稀疏表示的红外与可见光图像融合网络

Fig. 1 Multi-layer convolutional sparse representation-based network for infrared and visible image fusion

相比于 SR、CSR、CNN 等现有图像融合方法, 本文所设计的图像融合网络具有以下优势:

1) 与 SR 的局部变换方式不同, 本文所设计的图像融合网络采用全局变换的方式, 有效抑制了 SR 应用于图像融合所导致的语义信息损失以及对细节信息的低容忍度两大缺陷。

2) 相比于基于 CSR 的图像融合方法, 本文借鉴了卷积神经网络的设计思路。Papyan 等^[16]通过理论分析已经证明了多层卷积稀疏表示网络与 CNN 的结构存在着紧密的联系, 多层卷积稀疏表示可以实现更有效的图像变换。

3) 区别于 CNN 的监督学习特性, 本文所设计的多层卷积稀疏网络采取的学习方式是非监督的, 无需大量带有标签的训练样本完成网络的训练, 在实现上更为简单。

4) 本文算法的计算复杂度相比于 SR 具有明显的优势, 相比于 CSR 不会明显上升。假设 SR 与 CSR 所使用的字典维度均为 k , 输入图像的大小均为 $D \times D$, 基于 SR 的图像融合方法计算复杂度为 $O(D^2 \times k^2)$; 基于 CSR 的图像融合方法计算复杂度为 $O(k \times D \times \log(D))$, 明显低于 SR 的计算复杂度^[17]; 本文算法共包含两层卷稀疏层, 因此计算复杂度为 $O(2k \times D \times \log(D))$, 较 CSR 更高, 但依然低于 SR。

2 基于多层卷积稀疏表示的红外与可见光图像融合算法

2.1 多层卷积稀疏表示理论

给定输入图像 $I \in R^{A \times B}$, 以及一组相同大小的

图像滤波器 $f_i \in R^{a \times b}, i = 1, 2, \dots, m$, 这样一组滤波器被定义为卷积字典滤波器。CSR 的基本思想在于任意一幅输入图像 I 都可以表示为卷积字典滤波器以及与之对应的卷积稀疏响应图 $s_i \in R^{A \times B}, i = 1, 2, \dots, m$ 乘积的和。CSR 的目标函数如下式所示:

$$\arg \min_{\{s_i\}} \frac{1}{2} \left\| \sum_{i=1}^m f_i * s_i - I \right\|_2^2 + \lambda \sum_{i=1}^m \|x_i\|_1 \quad (1)$$

式中 λ 为稀疏正则项。

由于卷积稀疏响应 S 可被视为 m 张大小为 $A \times B$ 的图像集合, S 中的每一张卷积稀疏响应图可以进一步地被另一个卷积字典滤波器 $F' \in R^{a' \times b' \times m'}$ 和与之对应的卷积稀疏响应图表示, 以此类推, 可获得 K 层的卷积稀疏响应图。如定义 1 所示, 这种类型的 CSR 被称为多层 CSR。压缩感知理论的基本思想可视为通过对过完备字典 D 中原子的组合来表示输入信号, 多层 CSR 的基本思想与压缩感知理论相似, 但是区别于压缩感知理论所用到的字典原子, 多层 CSR 用于表征信号的是采用多层字典的复杂组合 $D^1 D^2 \dots D^K$, 与压缩感知理论中的字典原子对应, 这一类的字典组合被称为“分子”^[16]。

定义 1 给定输入图像 $I \in R^{A \times B}$, 对应与 K 层的卷积字典 $D^i = \{d_1^i, d_2^i, \dots, d_{m_i}^i\}, i = 1, 2, \dots, K$ 以及一组向量 $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ 。多层 CSR 被定义为寻找一组对应于 K 层的卷积稀疏响应图 $X^i = \{x_1^i, x_2^i, \dots, x_{m_i}^i\}, i = 1, 2, \dots, K$ 。

$$\text{layer 1: } \arg \min_{\{x_i^1\}} \frac{1}{2} \left\| \sum_{i=1}^{m_1} d_i^1 * x_i^1 - I \right\|_2^2 +$$

$$\lambda_1 \sum_{i=1}^{m_1} \|x_i^1\|_1$$

$$\text{layer 2: } \arg \min_{\{\mathbf{x}^{l_2}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_2} \mathbf{d}_i^{l_2} * \mathbf{x}_i^{l_2} - \mathbf{x}_i^{l_1} \right\|_2^2 + \lambda_2 \sum_{i=1}^{m_2} \|\mathbf{x}_i^{l_2}\|_1$$

.....

$$\text{layer } K: \arg \min_{\{\mathbf{x}^{l_K}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_K} \mathbf{d}_i^{l_K} * \mathbf{x}_i^{l_K} - \mathbf{x}_i^{l_{K-1}} \right\|_2^2 + \lambda_K \sum_{i=1}^{m_K} \|\mathbf{x}_i^{l_K}\|_1$$

给定一幅输入图像和预先学习的多层字典滤波器, 多层 CSR 的核心在于卷积稀疏响应图的估计。假设第 p 层的卷积字典滤波器为 $\mathbf{D}^{l_p} = \{\mathbf{d}_1^{l_p}, \mathbf{d}_2^{l_p}, \dots, \mathbf{d}_{m_p}^{l_p}\}$, 则计算卷积稀疏响应图的目标函数如下式所示:

$$\arg \min_{\{\mathbf{x}^{l_p}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_p} \mathbf{d}_i^{l_p} * \mathbf{x}_i^{l_p} - \mathbf{I} \right\|_2^2 + \lambda_p \sum_{i=1}^{m_p} \|\mathbf{x}_i^{l_p}\|_1 \quad (2)$$

式(2)可视为经典的基追踪问题的卷积形式, 可通过交替方向乘子算法 (Alternating direction method of multipliers, ADMM) 有效求解。

为了引入 ADMM 求解式(2)的优化问题, 将式(2)转换为如式(3)所示的交替形式, 对式(3)的优化过程通过迭代的方式进行, 从第 t 步至第 $t+1$ 步的迭代如式(4)~式(6)所示:

$$\arg \min_{\{\mathbf{x}^{l_p}, \mathbf{y}^{l_p}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_p} \mathbf{d}_i^{l_p} * \mathbf{x}_i^{l_p} - \mathbf{I} \right\|_2^2 + \lambda_p \sum_{i=1}^{m_p} \|\mathbf{y}_i^{l_p}\|_1$$

$$\text{s. t. } \mathbf{x}_i^{l_p} - \mathbf{y}_i^{l_p} = \mathbf{0} \quad (3)$$

$$\{\mathbf{x}_i^{l_p}\}^{(t+1)} = \arg \min_{\{\mathbf{x}^{l_p}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_p} \mathbf{d}_i^{l_p} * \mathbf{x}_i^{l_p} - \mathbf{I} \right\|_2^2 + \frac{\rho}{2} \sum_{i=1}^{m_p} \|\mathbf{x}_i^{l_p} - \mathbf{y}_i^{l_p(t)} + \mathbf{u}_i^{l_p(t)}\|_2^2 \quad (4)$$

$$\{\mathbf{y}_i^{l_p}\}^{(t+1)} = \arg \min_{\{\mathbf{y}^{l_p}\}} \lambda_p \sum_{i=1}^{m_p} \|\mathbf{y}_i\|_1 + \frac{\rho}{2} \sum_{i=1}^{m_p} \|\mathbf{x}_i^{l_p(t+1)} - \mathbf{y}_i^{l_p(t)} + \mathbf{u}_i^{l_p(t)}\|_2^2 \quad (5)$$

$$\mathbf{u}_i^{(t+1)} = \mathbf{u}_i^{(t)} + \mathbf{x}_i^{(t+1)} - \mathbf{y}_i^{(t+1)} \quad (6)$$

式中 \mathbf{u} 为引入的辅助变量。

对于式(5)的求解通过软阈值法即可完成, 而对于式(4)的求解相对复杂, 需进一步转换为如式(7)所示的形式。定义变量 $\mathbf{I}, \mathbf{d}_i^{l_p}, \mathbf{z}_i^{l_p}$ 和 $\mathbf{x}_i^{l_p}$ 在傅里叶频域对应的变量为 $\hat{\mathbf{I}}, \hat{\mathbf{d}}_i^{l_p}, \hat{\mathbf{z}}_i^{l_p}$ 和 $\hat{\mathbf{x}}_i^{l_p}$, 式(7)可转换为如式(8)所示的频域形式。

$$\arg \min_{\{\hat{\mathbf{x}}^{l_p}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_p} \hat{\mathbf{d}}_i^{l_p} * \hat{\mathbf{x}}_i^{l_p} - \hat{\mathbf{I}} \right\|_2^2 + \frac{\rho}{2} \sum_{i=1}^{m_p} \|\hat{\mathbf{x}}_i^{l_p} - \hat{\mathbf{z}}_i^{l_p}\|_2^2 \quad (7)$$

$$\arg \min_{\{\hat{\mathbf{x}}^{l_p}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_p} \hat{\mathbf{d}}_i^{l_p} * \hat{\mathbf{x}}_i^{l_p} - \hat{\mathbf{I}} \right\|_2^2 + \frac{\rho}{2} \sum_{i=1}^{m_p} \|\hat{\mathbf{x}}_i^{l_p} - \hat{\mathbf{z}}_i^{l_p}\|_2^2 \quad (8)$$

定义 $\hat{\mathbf{X}}^{l_p}, \hat{\mathbf{D}}^{l_p}$ 和 $\hat{\mathbf{z}}^{l_p}$ 如式(9)所示, 式(8)可进一步转换为式(10), 因此所求解的卷积稀疏响应的频域形式可通过式(11)获得。

$$\hat{\mathbf{X}}^{l_p} = (\hat{\mathbf{x}}_1^{l_p}, \hat{\mathbf{x}}_2^{l_p}, \dots, \hat{\mathbf{x}}_{m_p}^{l_p})$$

$$\hat{\mathbf{D}}^{l_p} = (\hat{\mathbf{d}}_1^{l_p}, \hat{\mathbf{d}}_2^{l_p}, \dots, \hat{\mathbf{d}}_{m_p}^{l_p}) \quad (9)$$

$$\hat{\mathbf{z}}^{l_p} = (\hat{\mathbf{z}}_1^{l_p}, \hat{\mathbf{z}}_2^{l_p}, \dots, \hat{\mathbf{z}}_{m_p}^{l_p})$$

$$\arg \min_{\hat{\mathbf{x}}^{l_p}} \frac{1}{2} \left\| \hat{\mathbf{D}}^{l_p} \hat{\mathbf{X}}^{l_p} - \hat{\mathbf{I}} \right\|_2^2 + \frac{\rho}{2} \left\| \hat{\mathbf{x}} - \hat{\mathbf{z}} \right\|_2^2 \quad (10)$$

$$(\hat{\mathbf{D}}^{l_p H} \hat{\mathbf{D}}^{l_p} + \rho \mathbf{E}) \hat{\mathbf{X}}^{l_p} = \hat{\mathbf{D}}^{l_p H} \hat{\mathbf{I}} + \rho \hat{\mathbf{z}}^{l_p} \quad (11)$$

2.2 多层卷积字典学习

定义 1 中卷积字典需要用训练样本中学习获得, 以第 p 层的卷积字典学习过程为例, 假设用于学习的训练样本为一组图像 $\mathbf{I}_i^{l_p} \in R^{A \times B}, i=1, 2, \dots, u_p$, 则第 p 层字典学习的优化目标函数如下式所示:

$$\arg \min_{\{\hat{\mathbf{d}}^{l_p}, \{\mathbf{x}^{l_p}\}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \lambda_p \sum_{i=1}^{u_p} \sum_{j=1}^{m_p} \|\mathbf{x}_{i,j}^{l_p}\|_1 \quad (12)$$

式中 $\|\mathbf{d}_j^{l_p}\|_1 = 1, j=1, 2, \dots, m_p$ 。

式(12)的优化可以拆分为对 $\{\mathbf{x}_{i,j}^{l_p}\}$ 的优化以及对 $\{\hat{\mathbf{d}}_j^{l_p}\}$ 的优化交替求解, 对 $\{\mathbf{x}_{i,j}^{l_p}\}$ 的优化及对 $\{\hat{\mathbf{d}}_j^{l_p}\}$ 的优化如式(13)、(14)所示。式(13)中的字典 $\{\hat{\mathbf{d}}_j^{l_p}\}$ 是固定的, 针对 $\{\mathbf{x}_{i,j}^{l_p}\}$ 的求解算法在多层卷积稀疏表示理论已经给出。因此, 重点介绍式(14)的优化算法。

$$\arg \min_{\{\mathbf{x}^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \lambda_p \sum_{i=1}^{u_p} \sum_{j=1}^{m_p} \|\mathbf{x}_{i,j}^{l_p}\|_1 \quad (13)$$

$$\arg \min_{\{\hat{\mathbf{d}}^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 \quad (14)$$

在式(14)中, 卷积稀疏响应 $\{\mathbf{x}_{i,j}^{l_p}\}$ 是固定的, 该问题可以视为多方向法 (Method of multiple directions, MOD) 的卷积形式。为了快速实现卷积操作, 采用傅里叶变换是一种有效的操作, 当在离散傅里叶频域计算卷积时, 需要对卷积字典 $\{\hat{\mathbf{d}}_j^{l_p}\}$ 进行零填充操作从而匹配 $\{\mathbf{x}_{i,j}^{l_p}\}$ 的维度。然而, 空间域的零填充操作无法在离散傅里叶频域得到清晰的表示, 为了解决该问题, 本文采用分离变量的方式将式(14)改写为式(15), 其中集合 C 和符号 τ 的定义如式(16)、(17)所示。通过求解式(15), 卷积字典

最终将以 $\{\hat{P}^T \hat{\mathbf{d}}_j^{l_p}\}$ 的形式获得, 其中 P 为零元素填充运算符。

$$\arg \min_{\{\hat{\mathbf{d}}_j^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \sum_{j=1}^{m_p} \tau C(\hat{\mathbf{d}}_j^{l_p}) \quad (15)$$

$$C = \{\mathbf{x} \in R^{A \times B} : (\mathbf{I} - PP^T)\mathbf{x} = 0, \|\mathbf{x}\|_2 = 1\} \quad (16)$$

$$\tau S(\mathbf{X}) = \begin{cases} 0 & \text{if } \mathbf{X} \in S \\ \infty & \text{if } \mathbf{X} \notin S \end{cases} \quad (17)$$

通过引入辅助变量 h 以及参数 ρ , 式 (15) 可以改写为式 (18) 的形式从而采用 ADMM 求解, 从第 t 步至第 $t+1$ 步的迭代过程如式 (19) ~ 式 (21) 所示。

$$\arg \min_{\{\hat{\mathbf{d}}_j^{l_p}, \hat{\mathbf{g}}_j^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \sum_{j=1}^{m_p} \tau C(\hat{\mathbf{g}}_j^{l_p}) \quad (18)$$

$$\text{s. t. } \forall j, \hat{\mathbf{d}}_j^{l_p} - \hat{\mathbf{g}}_j^{l_p} = 0 \quad (18)$$

$$\{\hat{\mathbf{d}}_j^{l_p}\}^{(t+1)} = \arg \min_{\{\hat{\mathbf{d}}_j^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \frac{\rho}{2} \sum_{j=1}^{m_p} \left\| \hat{\mathbf{d}}_j^{l_p} - \hat{\mathbf{g}}_j^{l_p(t)} + \mathbf{h}_j^{l_p(t)} \right\|_2^2 \quad (19)$$

$$\{\hat{\mathbf{g}}_j^{l_p}\}^{(t+1)} = \arg \min_{\{\hat{\mathbf{g}}_j^{l_p}\}} \sum_{j=1}^{m_p} \tau C(\hat{\mathbf{g}}_j^{l_p}) + \frac{\rho}{2} \sum_{j=1}^{m_p} \left\| \hat{\mathbf{d}}_j^{l_p(t+1)} - \hat{\mathbf{g}}_j^{l_p} + \mathbf{h}_j^{l_p(t)} \right\|_2^2 \quad (20)$$

$$\mathbf{h}_j^{l_p(t+1)} = \mathbf{h}_j^{l_p(t)} + \hat{\mathbf{d}}_j^{l_p(t+1)} - \hat{\mathbf{g}}_j^{l_p(t+1)} \quad (21)$$

对 $\{\hat{\mathbf{g}}_j^{l_p}\}$ 的优化通过软阈值法即可有效实现, 然而对 $\{\hat{\mathbf{d}}_j^{l_p}\}$ 的优化则相对复杂, 需进一步转换为式 (22) 的形式。定义 $\mathbf{I}_i^{l_p}, \mathbf{d}_j^{l_p}, \mathbf{x}_{i,j}^{l_p}$ 的频域形式为 $\hat{\mathbf{I}}_i^{l_p}, \hat{\mathbf{d}}_j^{l_p}, \hat{\mathbf{x}}_{i,j}^{l_p}$, 式 (22) 可以转换为式 (23) 的频域形式。

$$\arg \min_{\{\hat{\mathbf{d}}_j^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \mathbf{x}_{i,j}^{l_p} - \mathbf{I}_i^{l_p} \right\|_2^2 + \frac{\rho}{2} \sum_{j=1}^{m_p} \left\| \hat{\mathbf{d}}_j^{l_p} - \hat{\mathbf{z}}_j^{l_p} \right\|_2^2 \quad (22)$$

$$\arg \min_{\{\hat{\mathbf{d}}_j^{l_p}\}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \sum_{j=1}^{m_p} \hat{\mathbf{d}}_j^{l_p} * \hat{\mathbf{x}}_{i,j}^{l_p} - \hat{\mathbf{I}}_i^{l_p} \right\|_2^2 + \frac{\rho}{2} \sum_{j=1}^{m_p} \left\| \hat{\mathbf{d}}_j^{l_p} - \hat{\mathbf{z}}_j^{l_p} \right\|_2^2 \quad (23)$$

定义 $\hat{\mathbf{X}}_i^{l_p}, \hat{\mathbf{D}}^{l_p}, \hat{\mathbf{z}}^{l_p}$ 如式 (24) 所示, 式 (23) 可被改写为式 (25) 的形式, 卷积字典在傅里叶频域的解可以通过式 (26) 求得。

$$\begin{aligned} \hat{\mathbf{X}}_i^{l_p} &= (\hat{\mathbf{x}}_{i,1}^{l_p}, \hat{\mathbf{x}}_{i,2}^{l_p}, \dots, \hat{\mathbf{x}}_{i,m_p}^{l_p}) \\ \hat{\mathbf{D}}^{l_p} &= (\hat{\mathbf{d}}_1^{l_p}, \hat{\mathbf{d}}_2^{l_p}, \dots, \hat{\mathbf{d}}_{m_p}^{l_p}) \\ \hat{\mathbf{z}}^{l_p} &= (\hat{\mathbf{z}}_1^{l_p}, \hat{\mathbf{z}}_2^{l_p}, \dots, \hat{\mathbf{z}}_{m_p}^{l_p}) \end{aligned} \quad (24)$$

$$\arg \min_{\hat{\mathbf{D}}^{l_p}} \frac{1}{2} \sum_{i=1}^{u_p} \left\| \hat{\mathbf{X}}_i^{l_p} \hat{\mathbf{D}}^{l_p} - \hat{\mathbf{I}}_i^{l_p} \right\|_2^2 + \frac{\rho}{2} \left\| \hat{\mathbf{D}}^{l_p} - \hat{\mathbf{z}}^{l_p} \right\|_2^2 \quad (25)$$

$$\sum_{i=1}^{u_p} (\hat{\mathbf{X}}_i^{l_p H} \hat{\mathbf{X}}_i^{l_p} + \rho \mathbf{E}) \hat{\mathbf{D}}^{l_p} = \sum_{i=1}^{u_p} \hat{\mathbf{X}}_i^{l_p H} \hat{\mathbf{I}}_i^{l_p} + \rho \hat{\mathbf{z}}^{l_p} \quad (26)$$

至此, 多层卷积字典学习的定义归纳如下。

定义 2 给定对应于 K 层的训练样本 $\mathbf{I}_i^{l_p} \in R^{A_p \times B_p}, i=1, 2, \dots, u_p; p=1, 2, \dots, K$ 和向量 $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_k\}$, 多层卷积字典学习的定义为寻找对应于 K 层的卷积字典 $\mathbf{D}^{l_i} = \{\mathbf{d}_1^{l_i}, \mathbf{d}_2^{l_i}, \dots, \mathbf{d}_{m_i}^{l_i}\}, i=1, 2, \dots, K_0$ 。

$$\text{layer1: } \arg \min_{\{\hat{\mathbf{d}}_j^{l_1}, \hat{\mathbf{x}}_{i,j}^{l_1}\}} \frac{1}{2} \sum_{i=1}^{u_1} \left\| \sum_{j=1}^{m_1} \hat{\mathbf{d}}_j^{l_1} * \mathbf{x}_{i,j}^{l_1} - \mathbf{I}_i^{l_1} \right\|_2^2 + \lambda_1 \sum_{i=1}^{u_1} \sum_{j=1}^{m_1} \left\| \hat{\mathbf{x}}_{i,j}^{l_1} \right\|_1$$

$$\text{layer2: } \arg \min_{\{\hat{\mathbf{d}}_j^{l_2}, \hat{\mathbf{x}}_{i,j}^{l_2}\}} \frac{1}{2} \sum_{i=1}^{u_2} \left\| \sum_{j=1}^{m_2} \hat{\mathbf{d}}_j^{l_2} * \mathbf{x}_{i,j}^{l_2} - \mathbf{I}_i^{l_2} \right\|_2^2 + \lambda_2 \sum_{i=1}^{u_2} \sum_{j=1}^{m_2} \left\| \hat{\mathbf{x}}_{i,j}^{l_2} \right\|_1$$

.....

$$\text{layerK: } \arg \min_{\{\hat{\mathbf{d}}_j^{l_K}, \hat{\mathbf{x}}_{i,j}^{l_K}\}} \frac{1}{2} \sum_{i=1}^{u_K} \left\| \sum_{j=1}^{m_K} \hat{\mathbf{d}}_j^{l_K} * \mathbf{x}_{i,j}^{l_K} - \mathbf{I}_i^{l_K} \right\|_2^2 + \lambda_K \sum_{i=1}^{u_K} \sum_{j=1}^{m_K} \left\| \hat{\mathbf{x}}_{i,j}^{l_K} \right\|_1$$

2.3 红外与可见光图像融合算法

如图 1 所示, 给定一组配准后的红外图像 \mathbf{I}^N 与可见光图像 \mathbf{I}^V , 本文所设计的图像融合网络共包含 5 层, 以前馈的方式实现。

在第 1、2 层的卷积稀疏层, 给定对应于两层的预训练卷积字典 $\mathbf{D}^{l_1} = \{\mathbf{d}_1^{l_1}, \mathbf{d}_2^{l_1}, \dots, \mathbf{d}_{m_1}^{l_1}\}, \mathbf{D}^{l_2} = \{\mathbf{d}_1^{l_2}, \mathbf{d}_2^{l_2}, \dots, \mathbf{d}_{m_2}^{l_2}\}$ 和向量 $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$, 对应于两层的卷积稀疏响应可通过式 (27)、(28) 求解。

$$\begin{cases} \arg \min_{\{\hat{\mathbf{x}}_i^{l_1}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_1} \hat{\mathbf{d}}_i^{l_1} * \hat{\mathbf{x}}_i^{l_1} - \mathbf{I}^N \right\|_2^2 + \lambda_1 \sum_{i=1}^{m_1} \left\| \hat{\mathbf{x}}_i^{l_1} \right\|_1 \\ \arg \min_{\{\hat{\mathbf{x}}_i^{l_2}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_2} \hat{\mathbf{d}}_i^{l_2} * \hat{\mathbf{x}}_i^{l_2} - \mathbf{I}^V \right\|_2^2 + \lambda_2 \sum_{i=1}^{m_2} \left\| \hat{\mathbf{x}}_i^{l_2} \right\|_1 \end{cases} \quad (27)$$

$$\begin{cases} \arg \min_{\{\hat{\mathbf{x}}_i^{l_2}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_2} \hat{\mathbf{d}}_i^{l_2} * \hat{\mathbf{x}}_i^{l_2} - \hat{\mathbf{x}}_i^{l_1} \right\|_2^2 + \lambda_2 \sum_{i=1}^{m_2} \left\| \hat{\mathbf{x}}_i^{l_2} \right\|_1 \\ \arg \min_{\{\hat{\mathbf{x}}_i^{l_1}\}} \frac{1}{2} \left\| \sum_{i=1}^{m_1} \hat{\mathbf{d}}_i^{l_1} * \hat{\mathbf{x}}_i^{l_1} - \hat{\mathbf{x}}_i^{l_2} \right\|_2^2 + \lambda_2 \sum_{i=1}^{m_2} \left\| \hat{\mathbf{x}}_i^{l_1} \right\|_1 \end{cases} \quad (28)$$

在此之后, 如式 (27) 所示, 通过 l_1 最大绝对值融合规则来融合第 2 层卷积稀疏层输出的 $\{\hat{\mathbf{x}}_i^{l_2}\}$ 和 $\{\hat{\mathbf{x}}_i^{l_1}\}$ 。其中 $\mathbf{X}^{l_2} = (\mathbf{x}_1^{l_2}, \mathbf{x}_2^{l_2}, \dots, \mathbf{x}_{m_1}^{l_2})$ 和 $\mathbf{X}^{l_2}(a, b)$ 表示 \mathbf{X}^{l_2} 在位置 (a, b) 处的元素值。

$$\mathbf{X}^{Fl_2}(a, b) = \begin{cases} \mathbf{X}^{l_2}(a, b) & \left\| \mathbf{X}^{l_2}(a, b) \right\|_1 > \left\| \mathbf{X}^{l_1}(a, b) \right\|_1 \\ \mathbf{X}^{l_1}(a, b) & \left\| \mathbf{X}^{l_1}(a, b) \right\|_1 > \left\| \mathbf{X}^{l_2}(a, b) \right\|_1 \end{cases} \quad (29)$$

最终,第 4、5 层的重建层利用预先训练的卷积字典逐层地重建恢复融合图像,第 4 层的重建过程如式(30)所示,第 5 层的重建过程如式(31)所示。

$$\mathbf{x}_i^{Fl_1} = \sum_{i=1}^{m_2} \mathbf{d}_i^{l_2} * \mathbf{x}_i^{Fl_2} \quad (30)$$

$$\mathbf{I}^F = \sum_{i=1}^{m_1} \mathbf{d}_i^{l_1} * \mathbf{x}_i^{Fl_1} \quad (31)$$

至此,基于多层卷积稀疏网络的红外与可见光图像融合算法归纳如下。

算法:基于多层卷积稀疏的像素级红外与可见光图像融合。

输入:给定 $2K + 1$ 层的图像融合网络,输入信息包含一组源图像 \mathbf{I}^N 和 \mathbf{I}^V , 一组对应于 K 层的卷积字典 $\mathbf{D}^{l_i} = \{\mathbf{d}_1^{l_i}, \mathbf{d}_2^{l_i}, \dots, \mathbf{d}_{m_i}^{l_i}\}, i = 1, 2, \dots, K$ 以及向量 $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ 。

输出:融合图像 \mathbf{I}^F 。

步骤 1:利用 $\mathbf{D}^{l_i} = \{\mathbf{d}_1^{l_i}, \mathbf{d}_2^{l_i}, \dots, \mathbf{d}_{m_i}^{l_i}\}, i = 1, 2, \dots, K$ 从源图像 $\mathbf{I}^N, \mathbf{I}^V$ 中求解卷积稀疏响应 $\mathbf{X}^{NI_i} = \{\mathbf{x}_1^{NI_i}, \mathbf{x}_2^{NI_i}, \dots, \mathbf{x}_{m_i}^{NI_i}\}, i = 1, 2, \dots, K$ 和 $\mathbf{X}^{VI_i} = \{\mathbf{x}_1^{VI_i}, \mathbf{x}_2^{VI_i}, \dots, \mathbf{x}_{m_i}^{VI_i}\}, i = 1, 2, \dots, K$ 。

$$\begin{cases} \arg \min_{\{\mathbf{x}_j^{NI_{i+1}}\}} \frac{1}{2} \left\| \sum_{j=1}^{m_{i+1}} \mathbf{d}_j^{l_{i+1}} * \mathbf{x}_j^{NI_{i+1}} - \mathbf{X}^{NI_{i+1}} \right\|_2^2 + \\ \lambda_{i+1} \sum_{j=1}^{m_{i+1}} \left\| \mathbf{x}_j^{NI_{i+1}} \right\|_1 \\ \arg \min_{\{\mathbf{x}_j^{VI_{i+1}}\}} \frac{1}{2} \left\| \sum_{j=1}^{m_{i+1}} \mathbf{d}_j^{l_{i+1}} * \mathbf{x}_j^{VI_{i+1}} - \mathbf{X}^{VI_{i+1}} \right\|_2^2 + \\ \lambda_{i+1} \sum_{j=1}^{m_{i+1}} \left\| \mathbf{x}_j^{VI_{i+1}} \right\|_1 \end{cases}$$

$$i = 1, 2, \dots, K - 1, \mathbf{X}^{NI_K} = \mathbf{I}^N, \mathbf{X}^{VI_K} = \mathbf{I}^V$$

步骤 2:从卷积稀疏响应 \mathbf{X}^{NI_k} 和 \mathbf{X}^{VI_k} 中获取 \mathbf{X}^F 。

$$\mathbf{X}^F(a, b) = \begin{cases} \mathbf{X}^{NI_k}(a, b) \left\| \mathbf{X}^{NI_k}(a, b) \right\|_1 > \left\| \mathbf{X}^{VI_k}(a, b) \right\|_1 \\ \mathbf{X}^{VI_k}(a, b) \left\| \mathbf{X}^{VI_k}(a, b) \right\|_1 > \left\| \mathbf{X}^{NI_k}(a, b) \right\|_1 \end{cases}$$

步骤 3:从 \mathbf{X}^F 中利用卷积字典重建融合图像。

$$\mathbf{X}^{Fl_i} = \sum_{j=1}^{m_i} \mathbf{d}_j^{l_i} * \mathbf{x}_j^{Fl_i}, i = 1, 2, \dots, K, \mathbf{X}^{Fl_K} = \mathbf{I}^F$$

3 实验分析与验证

3.1 实验设置

如图 2 所示,3 组配准后的红外图像与可见光图像被选为实验用源图像。为了衡量图像融合的实验结果,本文采用主观评价与客观评价相结合的方式评判融合结果的优劣。主观评价单纯依靠人眼衡量融合结果,而客观评价通过一系列图像质量的评价指标来衡量融合结果。本文所选取的图像质量评价指标包括空间频率 (spatial frequency, SF), 熵 (Entropy, EN), 互信息 (mutual information, MI) 以及

梯度评价指标 $Q^{AB/F}$, 各项评价指标的含义如下:

1) SF 代表融合结果的清晰度与纹理丰富度, SF 越高, 融合效果越好。

2) EN 代表融合结果包含的信息量与纹理, EN 越高, 融合效果越好。

3) MI 代表融合结果的互信息, MI 越高, 融合效果越好。

4) $Q^{AB/F}$ 代表融合结果的梯度信息, $Q^{AB/F}$ 越高, 融合效果越好。

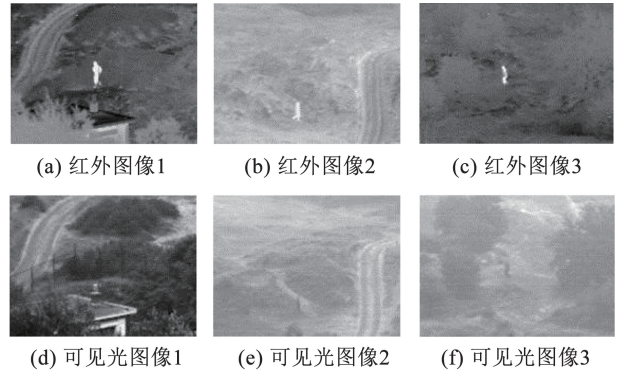


图 2 实验图像

Fig. 2 Images used for experiment

3.2 参数学习与分析

对于本文所设计的多层卷积稀疏表示网络而言,预训练的卷积稀疏字典对于网络的融合效果有着至关重要的影响,因此有必要针对字典的选取对网络融合效果的影响展开分析。由上一节的理论分析可知,给定训练样本,学习参数 λ , 迭代次数 t , 卷积字典的长度与卷积字典的大小对于卷积字典的学习有重要的影响。本文用于第 1 卷积稀疏层的训练样本为 200 幅自然图像,用于第 2 卷积稀疏层的训练样本为 200 幅卷积稀疏响应图,学习参数 λ , 迭代次数 t 设置为 500。

3.2.1 卷积字典的尺寸对融合效果的影响

如表 1 所示,当第 1、2 卷积稀疏层的卷积字典长度固定为 32 时,卷积字典的大小取值为 8×8 , 16×16 , 32×32 和 64×64 。由于第 2 卷积稀疏层的卷积字典训练样本为卷积稀疏响应图而非自然图像,因此可以直观地看出 \mathbf{D}^{l_2} 比 \mathbf{D}^{l_1} 更稀疏。

表 1 不同尺寸的卷积字典

Tab. 1 Convolutional dictionaries with different sizes

字典	长度固定为 32 时卷积字典尺寸的变化			
	8 × 8	16 × 16	32 × 32	64 × 64
\mathbf{D}^{l_1}				
\mathbf{D}^{l_2}				

源图像 1, 2, 3 在不同尺寸字典下的图像融合客观评价指标如图 3 ~ 5 所示。由图中结果可知, D^1 中字典尺寸变化带来的影响较 D^2 更大, 导致该现象的原因是: 1) 在字典训练的迭代次数一致的前提下, 当字典的尺寸增大时, 对于源图像变换的重建误差将随之上升, 从而影响最终图像融合的效果; 2) 与卷积神经网络类似, 浅层的网络对于提取图像的边缘信息更有效, 且底层网络所提取的边缘信息是后续网络提取语义信息的基础, 因此第 1 层的重建误差将传播至第 2 层, 影响最终的图像融合结果。

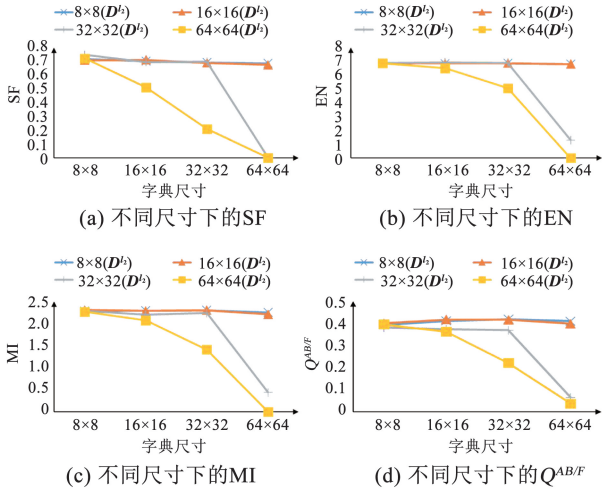


图 3 源图像 1 在不同字典尺寸下融合结果的客观评价指标
Fig. 3 Objective assessment metrics for fusion results of image 1 with different dictionary sizes

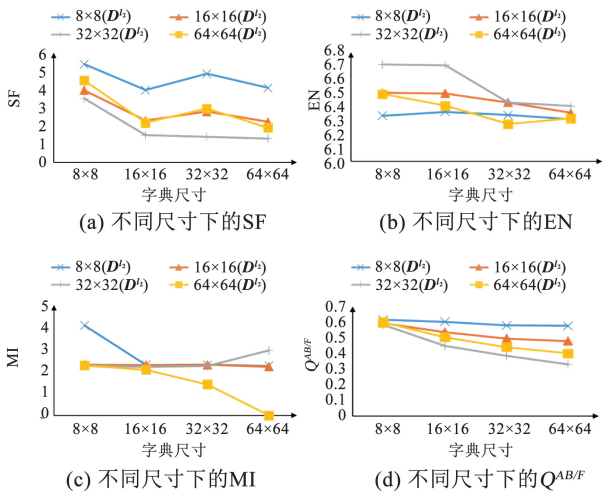


图 4 源图像 2 在不同字典尺寸下融合结果的客观评价指标
Fig. 4 Objective assessment metrics for fusion results of image 2 with different dictionary sizes

3.2.2 卷积字典的长度对融合效果的影响

如表 2 所示, 当第 1、2 卷积稀疏层的卷积字典尺寸固定为 16×16 , 卷积字典的长度为 16, 32, 64 和 128。源图像 1, 2, 3 在不同尺寸字典下的图像融合客观评价指标如图 6 ~ 8 所示。由图中结果可知, 相比于卷积字典的尺寸, 卷积字典的长度对融合结

果的影响较小。当卷积字典的长度增加时, 图像融合的效果会有微弱的改善, 且 D^2 对融合结果的影响较 D^1 更大。上述现象产生的原因为: 1) 图像变换对于信息的表示能力极大地取决于特征的维度, 而多层 CSR 的特征维度与字典的长度紧密相关, 因此当字典长度增加时, 融合效果会有相应的改善; 2) 与 CNN 的结构类似, 从网络深层提取的信息相较于浅层信息的表示能力更强, 因此 D^2 对融合结果有更大的影响。

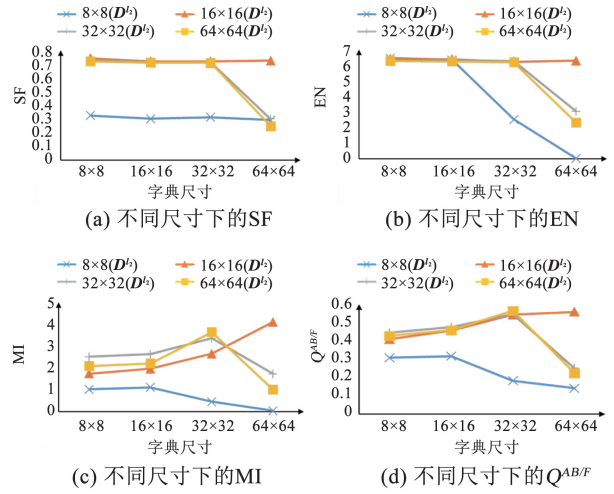


图 5 源图像 3 在不同字典尺寸下融合结果的客观评价指标
Fig. 5 Objective assessment metrics for fusion results of image 3 with different dictionary sizes

表 2 不同长度的卷积字典

Tab. 2 Convolutional dictionaries with different lengths

字典	尺寸固定为 16×16 时卷积字典的长度变化			
	8	16	32	64
D^1				
D^2				

3.3 对比实验分析

本文采用了 3 种经典的图像融合算法与本文设计的算法进行对比, 3 种对比算法分别是: 基于稀疏表示 (SR) 的图像融合算法^[3], 基于卷积稀疏表示 (CSR) 的图像融合算法^[15], 基于卷积神经网络 (CNN) 的图像融合算法^[9]。每一种融合算法所对应的融合结果刚见表 3, 根据融合结果采用主观评价可知, 本文所设计图像融合算法获得的结果在保留细节 (例如可见光图像中的植被、建筑等) 的同时显著增强了图像中的目标。

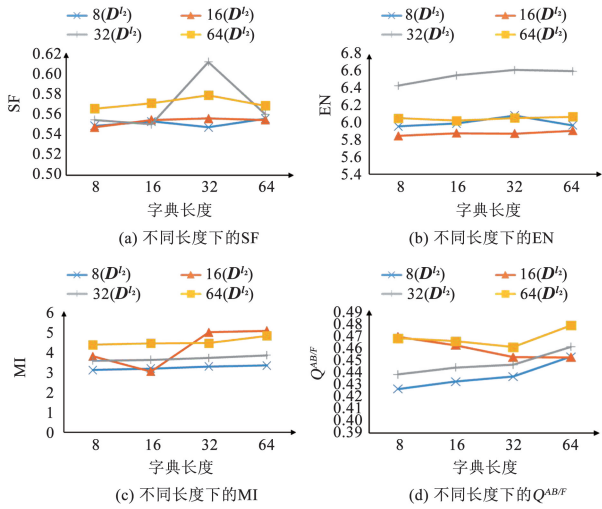


图 6 源图像 1 在不同字典长度下融合结果的客观评价指标
Fig. 6 Objective assessment metrics for fusion results of image 1 with different dictionary lengths

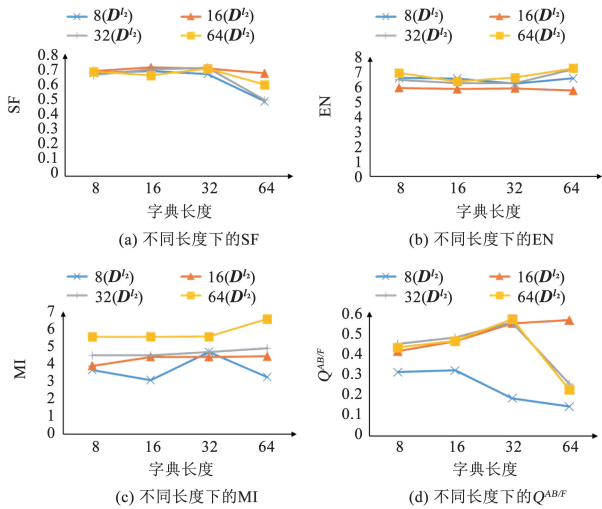


图 7 源图像 2 在不同字典长度下融合结果的客观评价指标
Fig. 7 Objective assessment metrics for fusion results of image 2 with different dictionary lengths

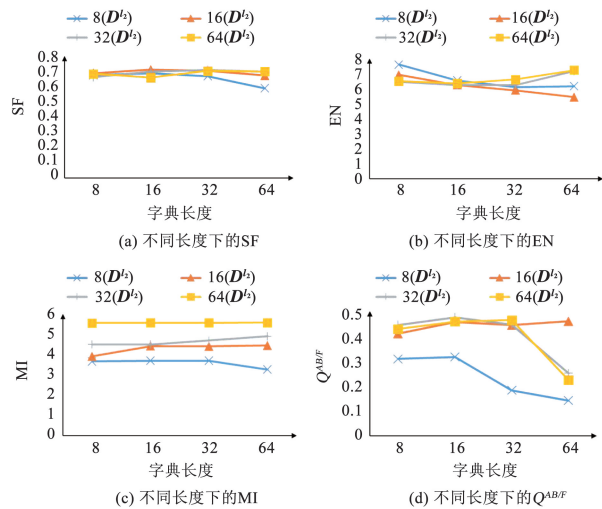


图 8 源图像 3 在不同字典长度下融合结果的客观评价指标
Fig. 8 Objective assessment metrics for fusion results of image 3 with different dictionary lengths

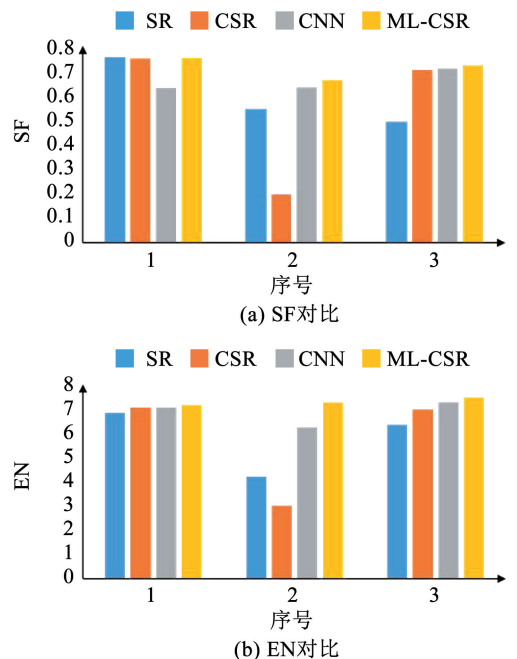
表 3 融合结果对比

Tab. 3 Comparison of fusion results

算法	1	2	3
输入图像	红外	可见光	
	可见光	红外	
SR			
CSR			
CNN			
本文算法			

进一步对本文所设计的算法进行客观评价,客观评价指标对比结果如图 9 所示。由客观评价指标的计算结果分析可知,本文所提出的图像融合算法相比于同类算法具有一定的优势。

针对 3 组实验图像,对 4 种算法的实时性进一步进行验证分析。本文的算法实现平台为 Matlab 2016b,计算机主频为 3.4GHz,内存为 8GB,采用



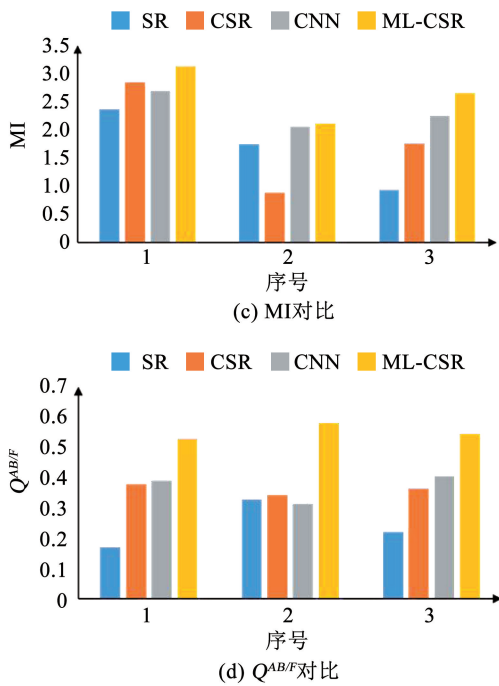


图 9 不同方法的融合结果客观评价指标对比

Fig. 9 Objective assessment metrics comparison of fusion results for different methods

Matlab 的 tic toc 命令, 对于 4 种算法的运行时间进行统计, 统计结果见表 4。由表 4 可知, 本文算法相比于 SR 和 CNN 在计算时间方面具有明显的优势; 相比于 CSR, 由于前馈式的网络结构导致需要进行两次卷积稀疏运算, 因此计算时间略有增长。

表 4 融合计算时间对比

Tab. 4 Comparison of computation time for image fusion

算法	第 1 组融合图像 运行时间/s	第 2 组融合图像 运行时间/s	第 3 组融合图像 运行时间/s
SR	859	633	757
CSR	33	29	31
CNN	132	96	117
本文算法	65	42	58

4 结 论

1) 本文设计了一种多层卷积稀疏表示网络, 且给出了针对该网络的卷积稀疏字典训练方法与卷积稀疏响应图求解方法, 作为一种有效的图像变换方法, 该网络不仅可用于红外与可见光图像融合, 同样可被扩展于目标检测、跟踪等领域。

2) 与基于稀疏表示的图像融合方法对比, 本文所设计的多层卷积稀疏表示网络所具备的全局建模能力在误匹配条件下具有明显的优势。

3) 作为一种基于非监督学习的融合网络, 本文所设计的图像融合方法无需大量带有标签的训练样

本即可完成参数的学习, 因此该网络同样可被用于解决其他类型的图像融合问题, 例如多焦点图像融合、医学图像融合等。

参考文献

- [1] MA Jiayi, MA Yong, LI Chang. Infrared and visible image fusion methods and applications; A survey[J]. Information Fusion, 2019, 45: 153. DOI: 10.1016/j.inffus.2018.02.004
- [2] LI Shutao, KANG Xudong, FANG Leyuan, et al. Pixel-level image fusion: A survey of the state of the art[J]. Information Fusion, 2017, 33: 100. DOI: 10.1016/j.inffus.2016.05.004
- [3] DING Meng, WEI Li, WANG Bangfeng. Research on fusion method for infrared and visible images via compressive sensing[J]. Infrared Physics & Technology, 2013, 57: 56. DOI: 10.1016/j.infrared.2012.12.014
- [4] LIU Yu, CHEN Xun, WANG Zengfu, et al. Deep learning for pixel-level image fusion: Recent advances and future prospects[J]. Information Fusion, 2018, 42: 158. DOI: 10.1016/j.inffus.2017.10.007
- [5] LIU Yu, LIU Shuping, WANG Zengfu. A general framework for image fusion based on multi-scale transform and sparse representation[J]. Information Fusion, 2015, 24: 147. DOI: 10.1016/j.inffus.2014.09.004
- [6] ZHANG Qiang, LIU Yi, BLUM R S, et al. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images; A review[J]. Information Fusion, 2018, 40: 57. DOI: 10.1016/j.inffus.2017.05.006
- [7] LIU Yu, CHEN Xun, PENG Hu, et al. Multi-focus image fusion with a deep convolutional neural network[J]. Information Fusion, 2017, 36: 191. DOI: 10.1016/j.inffus.2016.12.001
- [8] LIU Y, CHEN X, CHENG J, et al. A medical image fusion method based on convolutional neural networks[C]//Proceedings of the 20th International Conference on Information Fusion (Fusion). Xi'an: IEEE, 2017: 1. DOI: 10.23919/ICIF.2017.8009769
- [9] LIU Yu, CHEN Xun, CHENG Juan, et al. Infrared and visible image fusion with convolutional neural networks[J]. International Journal of Wavelets, Multiresolution and Information Processing, 2018, 16(3): 1850018. DOI: 10.1142/S0219691318500182
- [10] SUN Xudong, WU Pengcheng, HOI S C H. Face detection using deep learning: An improved faster RCNN approach[J]. Neurocomputing, 2018, 299: 42. DOI: 10.1016/j.neucom.2018.03.030
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV: IEEE, 2016: 779. DOI: 10.1109/CVPR.2016.91
- [12] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 21. DOI: 10.1007/978-3-319-46448-0_2
- [13] KALANTARI N K, RAMAMOORTHY R. Deep high dynamic range imaging of dynamic scenes[J]. ACM Transactions on Graphics, 2017, 36(4): 1. DOI: 10.1145/3072959.3073609
- [14] ZEILER M D, TAYLOR G W, FERGUS R. Adaptive deconvolutional networks for mid and high level feature learning[C]//Proceedings of the International Conference on Computer Vision. Barcelona: IEEE, 2011: 2018. DOI: 10.1109/ICCV.2011.6126474
- [15] LIU Yu, CHEN Xun, WARD R K, et al. Image fusion with convolutional sparse representation[J]. IEEE Signal Processing Letters, 2016, 23(12): 1882. DOI: 10.1109/LSP.2016.2618776
- [16] POPYAN V, ROMANO Y, ELAD M. Convolutional neural networks analyzed via convolutional sparse coding[J]. The Journal of Machine Learning Research, 2017, 18(1): 2887. DOI: arXiv:1607.08194.
- [17] HEIDE F, HEIDRICH W, WETZSTEIN G. Fast and flexible convolutional sparse coding[C]//Proceedings of the Computer Vision and Pattern Recognition. Boston, MA: IEEE, 2015. DOI: 10.1109/CVPR.2015.7299149