

# 基于深度学习的无监督红外图像与可见光图像融合算法

陈国洋<sup>1,2</sup>, 吴小俊<sup>1,2\*</sup>, 徐天阳<sup>2</sup>

<sup>1</sup>江南大学人工智能与计算机学院, 江苏 无锡 214122;

<sup>2</sup>江南大学江苏省模式识别与计算智能工程实验室, 江苏 无锡 214122

**摘要** 卷积神经网络(CNN)由于出色的性能,被逐渐应用于图像融合领域。对于红外图像和可见光图像的融合任务而言,由于没有标签数据,对其进行无监督的学习建模具有重要意义。针对这个问题,提出了一种无监督的端到端的深度融合算法,该算法可以由输入的红外源图像和可见光源图像,直接预测出包含源图像显著信息的融合图像。所提算法构建了一个自编码网络,并使用真实的数据集进行训练。网络所用损失函数为广泛应用于图像融合任务的图像结构相似度量(SSIM)。具体地说,设计了一种改进的无参考图像评价指标来计算损失函数,达到对网络进行无监督训练的目的。此外,将注意力机制引入到模型中进一步提升了融合结果。将所提算法与多种融合算法进行了对比实验,实验结果表明,所提算法融合结果无论在主观评价还是客观指标评价方面都十分具有竞争力。

**关键词** 图像处理; 图像融合; 红外图像; 可见光图像; 无监督学习; 卷积神经网络; 注意力机制

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0410010

## Unsupervised Infrared Image and Visible Image Fusion Algorithm Based on Deep Learning

Chen Guoyang<sup>1,2</sup>, Wu Xiaojun<sup>1,2\*</sup>, Xu Tianyang<sup>2</sup>

<sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China;

<sup>2</sup>Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, Jiangsu 214122, China

**Abstract** Convolutional neural network (CNN) is gradually applied to the field of image fusion because of its excellent performance. For the fusion task of infrared image and visible image, because there is no label data, unsupervised learning modeling is of great significance. To solve this problem, an unsupervised end-to-end depth fusion algorithm is proposed. The algorithm can directly predict the fused image containing significant information of the source image from the input infrared source image and visible light source image. The proposed algorithm constructs an auto-encoder network and uses real datasets for training. The loss function used in the network is the image structure similarity index measure (SSIM), which is widely used in image fusion tasks. Specifically, an improved non reference image evaluation index is designed to calculate the loss function, so as to achieve the purpose of unsupervised training of the network. In addition, the attention mechanism is introduced into the model to further improve the fusion results. The proposed algorithm is compared with many fusion algorithms, experimental results

收稿日期: 2021-02-26; 修回日期: 2021-03-18; 录用日期: 2021-03-31

基金项目: 国家自然科学基金(62020106012, U1836218, 61672265)、中国教育部 111 项目(B12018)、国家重点研发计划(2017YFC1601800)

通信作者: \*wu\_xiaojun@jiangnan.edu.cn

show that the fusion results of the proposed algorithm are very competitive in both subjective evaluation and objective index evaluation.

**Key words** image processing; image fusion; infrared image; visible image; unsupervised learning; convolutional neural network; attention mechanism

## 1 引言

图像融合是将两张或多张源图像的显著特征集成到一张图像以此增强感知性的过程<sup>[1]</sup>。红外图像和可见光图像由于获取容易且有很好的信息互补性,是图像融合领域中的一个热门研究方向。红外图像是热辐射图像,可以通过亮度的形式反映不同物体所包含的热辐射信息差异,但是图像分辨率往往较低,视觉效果差。可见光图像是光反射图像,包含丰富的纹理信息和背景信息,并且符合人类的视觉系统,但是可见光传感器容易受到天气等因素的影响。对这两种图像的优点进行融合可以得到场景信息更加丰富的融合图像。红外图像与可见光图像融合技术已经广泛应用到军事侦察、社会安全、目标跟踪等众多领域<sup>[2-3]</sup>。

图像融合的过程如下:首先对源图像特征进行抽取;然后使用合理的策略对抽取到的特征进行融合;最后对融合后的特征进行重建。传统的融合方法主要为基于多尺度分解的方法<sup>[4-5]</sup>和基于稀疏表示的方法<sup>[6-7]</sup>,这些方法主要存在以下问题:1)需要大量的手工设计;2)在数据量比较多的情况下,效率较低。近年来许多基于深度学习的融合算法被提出,在一定程度上解决了上述问题。相比于传统方法,卷积神经网络(CNN)可以通过学习的方式抽取到更加鲁棒的图像特征,成为图像处理过程中的有力工具<sup>[8-11]</sup>。

2017年,Liu等<sup>[12]</sup>使用清晰图像和模糊图像作为标签数据训练网络,将CNN应用到多聚焦图像融合任务中。但是红外图像与可见光图像没有可参考的融合图像,无法制作标签数据集。2018年,Li等<sup>[13]</sup>用预训练的VGG网络作为特征提取器,实现了红外图像与可见光源图像多层深度特征的提取和融合。2019年,Li等<sup>[14]</sup>又提出了DenseFuse,DenseFuse用COCO数据集<sup>[15]</sup>训练一个图像重建网络模型,然后用该模型实现了源图像的融合。2020年,Zhang等<sup>[16]</sup>提出了一个通用的图像融合框架IFCNN,该模型实现了多任务融合。这些算法虽然都有良好的融合效果,但是需要在其他有标签的数据集进行学习。而红外图像与可见光源图像并

未参与训练,因此网络模型不能自适应地对源图像特征进行提取和融合。

2017年,Prabhakar等<sup>[17]</sup>提出针对多曝光图像的深度融合算法Deepfuse,该算法通过一个无参考图像评价指标实现无监督学习。受DeepFuse启发,Mustafa等<sup>[18]</sup>在2019年通过构建一个无监督深度模型实现了多聚焦图像的融合。同年,Ma等<sup>[19]</sup>提出FusionGAN,该模型通过无监督的生成对抗网络实现了红外图像与可见光图像的融合。这些算法通过样本数据自身学习构建融合图像,使得网络中的参数与融合任务高度相关,模型具有更强的特征自适应提取能力。此外,这些算法都是端到端的融合算法,融合效率进一步提高。

综上所述,将无监督深度模型应用到红外图像与可见光图像融合中是十分有意义的一项工作。基于此,本文提出了一种无监督的端到端的深度融合算法,该算法可以由输入的红外源图像和可见光源图像直接预测出包含源图像显著信息的融合图像。所提算法使用基于方差加权的方式计算损失函数,并引入注意力模块进一步优化实验结果。实验结果表明,所提算法融合结果无论在主观评价还是客观指标评价方面都十分具有竞争力。

## 2 所提算法模型

### 2.1 网络模型

所提网络模型结构如图1所示,网络由编码器、融合层和解码器3部分组成。 $I_1$ 和 $I_2$ 分别代表一对配准好的红外图像和可见光图像, $I_f$ 代表融合图像。

编码器的作用是对输入的源图像进行特征提取。编码器由4个卷积层组成,卷积核大小均为 $3 \times 3$ ,卷积步长为1,激活函数为ReLU。将 $I_1$ 和 $I_2$ 输入到编码器,对于 $I_1$ 而言,首先用两个卷积层C1和C2对低层特征进行提取,然后用两个带有跳跃连接的卷积层C3、C4进行高层特征提取,跳跃连接可以增加网络低层和高层之间的信息流动。对 $I_2$ 用相同参数的编码器进行特征提取,这样操作的优点在于不但可以减小编码器的参数量,还可以在对应位置产生相同属性的特征,便于后续的融合操作。

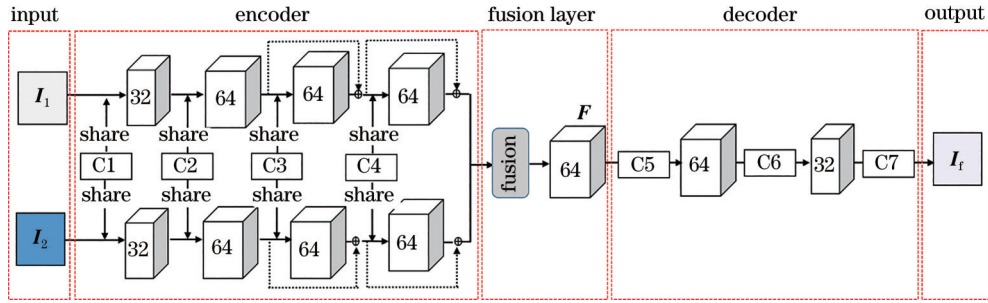


图 1 所提网络结构

Fig. 1 Structure of proposed network

融合层的作用是对编码器提取到的图像特征进行融合。 $\Phi_1$ 和 $\Phi_2$ 分别表示编码器提取到的红外图像与可见光图像的特征,因为 $\Phi_1$ 和 $\Phi_2$ 是用相同编码器提取的,所以在对应位置的特征类型是相同的,可以直接进行相加融合。但是通过这种方式得到的融合特征比较粗糙,所以所提算法在对特征进行融合后引入了注意力机制,如图 2 所示。具体来说,先将 $\Phi_1$ 和 $\Phi_2$ 相加得到初步融合特征 $F'$ ,然后将 $F'$ 输入 squeeze-and-excitation(SE)模块进行精细化

的调整。SE模块能够通过训练自动学习到各个特征通道的重要程度,然后给不同通道特征分配不同的权重,重新调整每个通道的重要性。具体过程如下:首先对 $F'$ 进行全局平均池化得到一个通道维度的全局特征,接着用两个全连接层(FC)学习每个通道之间的相关性,然后用 Sigmoid 函数进行激活得到一个 0~1 之间的权重系数,最后将得到的权重系数用相乘的方式对 $F'$ 进行调整得到最终的融合特征 $F$ 。

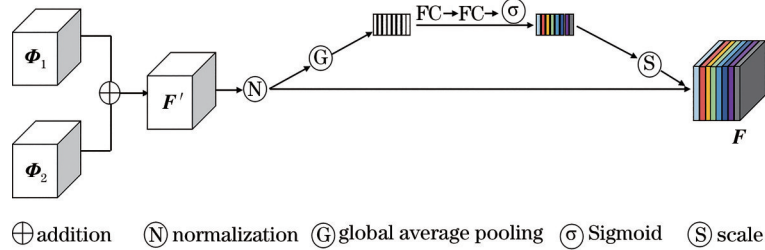


图 2 融合层示意图

Fig. 2 Diagram of fusion layer

解码器的任务是对融合后的特征进行重建得到期望的融合图像。解码器包含 3 个卷积层(C5, C6, C7),卷积核大小为  $3 \times 3$ ,步长为 1,激活函数为 ReLU,融合图像  $I_f$  由最后的卷积层 C7 产生。

### 2.2 损失函数

损失函数对网络训练是至关重要的,也是所提算法的核心。所提算法使用无监督的方式对网络进行训练,因此选择结构相似度量(SSIM)<sup>[20]</sup>为损失函数。SSIM是一个衡量两张不同图像相似性的函数,由于对图像评价更符合人眼的视觉系统,被广泛应用到图像重建和融合任务中。用 $x$ 和 $y$ 代表两张不同的图像,则 SSIM 函数的表达式为

$$\text{SSIM}(x, y|\omega) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (1)$$

式中: $\mu$ 表示像素平均值; $\sigma^2$ 表示方差; $\sigma_{xy}$ 为协方差;

$C_1$ 和 $C_2$ 是用来维持稳定的常数, $C_1 = (k_1L)^2$ , $C_2 = (k_2L)^2$ ,取 $k_1 = 0.01$ , $k_2 = 0.03$ , $L = 255$ 。 $\omega$ 是一个滑动窗口,SSIM是通过 $\omega$ 逐像素计算每个像素所在区域的 SSIM 值得到的,最终计算结果是一个矩阵。SSIM( $x, y|\omega$ )  $\in [-1, 1]$ ,越靠近 1 时,说明两张图像在 $\omega$ 中的部分越相似,令 $\omega$ 为  $11 \times 11$ 。

因为 SSIM 是用一个滑动窗口 $\omega$ 去逐像素滑动计算的,所以可以用无参考图像评价指标去判断哪张源图像在这一窗口内的图像质量更优,然后选优与融合图像计算 SSIM。考虑纹理特征和亮度特征的特点,将图像方差和均值有机组合作为一个新的无参考图像评价指标计算损失函数(SSIM),从而实现对源图像显著特征的融合。

图像的方差和均值都是典型的无参考图像质量评价指标。图像的方差可以表示图像的纹理信



息,方差越大,表明图像灰度级越分散,图像的纹理就越多。图像的均值可以表示图像的亮度信息,均值越大,表明图像亮度越高,图像包含的信息越多。

所提算法计算 SSIM 的方式为

$$S(I_1, I_2, I_f | \omega) = \begin{cases} \text{SSIM}(I_1, I_f | \omega), \alpha_1 \cdot \sigma^2(I_1 | \omega) \geq \alpha_2 \cdot \sigma^2(I_2 | \omega) \\ \text{SSIM}(I_2, I_f | \omega), \alpha_1 \cdot \sigma^2(I_1 | \omega) < \alpha_2 \cdot \sigma^2(I_2 | \omega) \end{cases}, \quad (2)$$

式中:  $\sigma^2(I_1 | \omega)$  代表红外图像在某一滑动窗口  $\omega$  内所有像素方差;  $\sigma^2(I_2 | \omega)$  代表可见光图像在对应位置滑动窗口  $\omega$  中的方差;  $\alpha_1$  和  $\alpha_2$  是权重系数。

对于  $I_1$  和  $I_2$  所对应的某一  $\omega$  区域,首先分别计算这两个区域的方差并比较大小。若在  $I_1$  中这一  $\omega$  区域方差大于  $I_2$  对应  $\omega$  区域的方差,说明  $I_1$  中这一区域包含的纹理信息更多,则该区域的 SSIM 值由融合图像中的对应区域与  $I_1$  中的这一区域计算得到,融合图像中的该区域与  $I_1$  中该区域相近,反之则与  $I_2$  中对应区域相近。

但是只用方差判断是不准确的,因为对于像素值变化不是很明显的连续区域,两个对应窗口的方差可能相差很小。因此所提算法用一个权重系数对方差进行加权。

$$\alpha_i = \frac{\mu(I_i | \omega)}{\sum_{i=1}^2 \mu(I_i | \omega)}, \quad (3)$$

式中:  $\mu(I_i | \omega)$  代表在图像滑动窗口  $\omega$  中的像素均值,  $i = \{1, 2\}$ 。对于方差相差不明显的区域用均值对方差进行加权,使得网络可以选择亮度较明显的区域进行优化。这样的计算方法使得融合图像既可以保留足够的纹理特征,又可以捕捉到亮度区域,实现对源图像显著特征的融合。最终的损失函数可以表示为

$$L_{\text{loss}}(I_1, I_2, I_f) = 1 - \frac{1}{N} S(I_1, I_2, I_f | \omega), \quad (4)$$

式中:  $N$  表示一张图像中滑动窗口  $\omega$  的数量。

### 2.3 实验环境及参数设置

TNO 数据集<sup>[21]</sup>是一个配准好的真实的红外图像和可见光图像数据集。从 TNO 数据集中收集了 15 对图像作为训练数据。由于数据量过少,这些图像不能直接用来训练网络,用一个  $64 \times 64$  的滑动窗口对原始图像通过裁剪的方式进行数据增强处理,滑动步长为 20,得到 10000 多张子图像,然后随机选取其中的 10000 张图像作为训练数据集。

实验在操作系统 Ubuntu 16.04 上进行,编程语言为 Python 3.7,深度学习框架为 Pytorch, CPU 为 Intel i7-6850K, GPU 为 NVIDIA TITAN XP。网络优化器为 Adam,初始学习率为  $1 \times 10^{-4}$ ,权重衰减率为  $1 \times 10^{-5}$ 。网络共训练 32 个周期,批量池大小为 8。

## 3 实验及结果分析

从文献[21]和文献[22]中随机选出 21 对红外图像和可见光图像作为测试集。该测试集中的图像均为真实的红外图像和可见光图像,其中选取的许多图像也被现有的红外图像和可见光图像的融合算法用作测试数据,具有代表性。

对融合结果从主观评价和客观指标两个方面分别进行评价。首先进行消融实验来探究所提算法的合理性,然后将所提算法融合结果与 9 种典型的算法融合结果进行对比,包括基于离散余弦波的小波变换(DCHWT)算法<sup>[23]</sup>、基于联合稀疏表示(JSR)算法<sup>[7]</sup>、基于梯度转移和总变差最小化(GTF)的算法<sup>[24]</sup>、基于视觉显著性图和加权最小二乘优化(WLS)算法<sup>[22]</sup>、基于卷积稀疏表示的算法<sup>[25]</sup>(CSR)、基于 VGG 网络多层深度特征融合(VGGML)的算法<sup>[13]</sup>、基于密集连接的自编码网络(DenseFuse)融合算法<sup>[14]</sup>、基于生成对抗网络(FusionGAN)的算法<sup>[15]</sup>、基于通用融合(IFCNN)网络融合算法<sup>[16]</sup>。

### 3.1 评价指标

采用 6 种图像客观评价指标来衡量融合图像的质量,这些指标为融合图像质量评价的常用指标,能够客观说明图像的融合质量。 $X_1$ 、 $X_2$  和  $X_f$  分别代表红外图像、可见光图像和融合图像,  $X$  表示某一张源图像。

1) 信息熵(En)<sup>[26]</sup>。信息熵可以衡量图像中包含信息量。熵的值越大,表明一张图像包含的信息越多。图像信息熵的表达式为

$$E_{\text{En}} = - \sum_{l=0}^{L-1} p_l \cdot \log_2 p_l, \quad (5)$$

式中:  $L$  表示灰度级;  $p_l$  表示每一级灰度出现的概率。

2) 标准差(SD)<sup>[27]</sup>。标准差反映图像灰度值相对于均值的离散程度。标准差越大表明图像像素分布差异越大,图像的质量越好。

$$D_{SD} = \sqrt{\sum_{m=1}^M \sum_{n=1}^N (P_{m,n} - \mu)^2}, \quad (6)$$

式中:  $P_{m,n}$  为图像某一点像素值。

3) 视觉信息保真度(VIF)<sup>[28]</sup>。视觉信息保真度可以衡量融合图像包含源图像相关信息的量,还可以反映融合图像的失真情况。VIF 值越高表明融合图像质量越好,具体可参考文献[31]。

4) 互信息量(MI)<sup>[29]</sup>。互信息量是用来衡量融合图像包含源图像信息量的指标,值越大表明融合图像包含源图像的特征越多,融合结果也更优。互信息的表达式为

$$I_{MI}(\mathbf{X}, \mathbf{F}) = \sum \sum p(\mathbf{X}, \mathbf{F}) \cdot \log_2 \left[ \frac{p(\mathbf{X}, \mathbf{F})}{p(\mathbf{X})p(\mathbf{F})} \right], \quad (7)$$

式中:  $p(\mathbf{X})$  表示图像  $\mathbf{X}$  的灰度分布直方图;  $p(\mathbf{F})$  表示  $\mathbf{F}$  的灰度分布直方图;  $p(\mathbf{X}, \mathbf{F})$  表示  $\mathbf{X}$  和  $\mathbf{F}$  的联合分布直方图。所提算法的互信息表达式为

$$I_{MI}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{F}) = I_{MI}(\mathbf{X}_1, \mathbf{F}) + I_{MI}(\mathbf{X}_2, \mathbf{F}). \quad (8)$$

5) 边缘信息度量( $Q_{abf}$ )<sup>[30]</sup>。边缘信息度量是用来衡量从源图像转移到融合图像的边缘特征的量,此指标越大,表明融合图像保存的源图像边缘信息越多,具体计算可参考文献[30]。

6) 差异相关总和(SCD)<sup>[31]</sup>。差异相关总和是用来衡量融合图像包含源图像中互补信息的多少

的量,值越大表明融合图像包含源图像的互补信息越多。差异相关总和的表达式为

$$D_{SCD}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{F}) = r(\mathbf{X}_1, \mathbf{F}) + r(\mathbf{X}_2, \mathbf{F}), \quad (9)$$

$$r(\mathbf{X}, \mathbf{F}) = \frac{\sum_m \sum_n [D(m, n) - D_\mu][X(m, n) - X_\mu]}{\sqrt{\sum_m \sum_n \{ [D(m, n) - D_\mu]^2 \cdot [X(m, n) - X_\mu]^2 \}}}, \quad (10)$$

$$D(m, n) = F(m, n) - X(m, n). \quad (11)$$

### 3.2 消融实验

设置了3个实验来验证所提算法的合理性。实验1只用方差作为无参考指标计算损失函数,特征融合方式为相加。实验2采用所提基于方差加权的方式计算损失函数,特征融合方式为相加。实验3在实验2基础上将注意力机制引入特征融合过程中,即所提完整算法。

在主观评价方面,从测试集中选出两对具有代表性的图像来展示3个实验的融合效果,两对图像名称分别为“街道”和“人”,如图3、4所示。图中用实线框来标记亮度信息,虚线框来标记纹理信息。左下角和右上角框中内容为放大两倍后的对应颜色矩形框中的标记信息,便于视觉方面的分析。

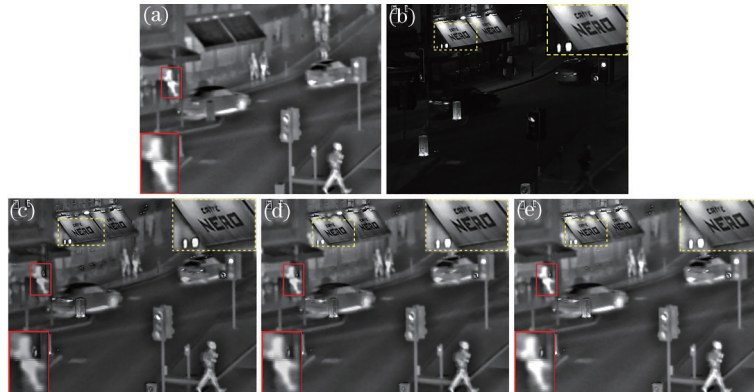


图3 “街道”图像实验。(a)红外图像;(b)可见光图像;(c)实验1的融合结果;(d)实验2的融合结果;(e)实验3的融合结果  
Fig. 3 Experiments on “street” images. (a) Infrared image; (b) visible light image; (c) fusion result of experiment 1; (d) fusion result of experiment 2; (e) fusion result of experiment 3

图3是一张夜晚的街道图。图3(a)和图3(b)分别代表待融合的红外图像和可见光图像。从图中可以看到,可见光图像光线较暗但是纹理信息比较丰富,红外图像中包含丰富的亮度信息,车辆行人及信号灯的轮廓分明,但是纹理信息较为模糊。图3(c)是实验1的融合结果,实验1实现了对红外图像和可见光图像的融合,纹理信息融合得较好,

但是亮度信息不够明显。图3(d)是实验2的融合结果,相比于实验1,实验2的融合图像中红外图像的红外特征有所增强,纹理信息也融合较好。从图3(e)可以看出,亮度信息显著增强,同时纹理信息足够清晰,达到最优的融合结果。

图4为3个实验在“人”图像的融合结果。图4(a)是红外图像,图像清晰地显示了一个人的红外信息,

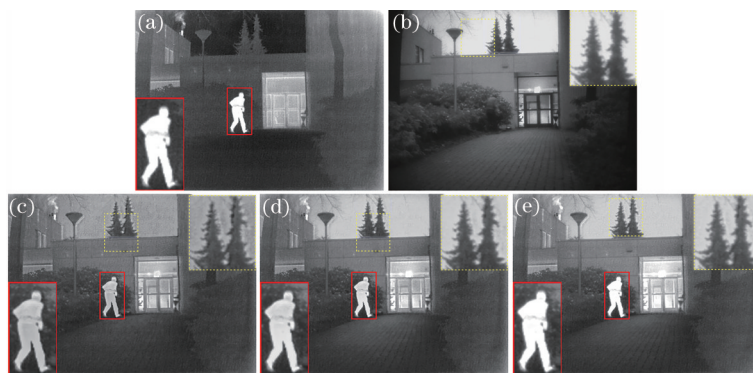


图 4 “人”图像实验。(a)红外图像;(b)可见光图像;(c)实验 1 的融合结果;(d)实验 2 的融合结果;(e)实验 3 的融合结果  
Fig. 4 Experiments on “man” images. (a) Infrared image; (b) visible light image; (c) fusion result of experiment 1; (d) fusion result of experiment 2; (e) fusion result of experiment 3

但是图像纹理信息较少。图 4(b)是可见光图像,图像有丰富的纹理信息,但是图中的人和环境颜色较为接近,容易发生混淆。从融合结果可以看出,所提算法能够最好地突出红外目标的亮度信息,同时充分地保留了可见光图像的纹理特征,融合结果是最优的。

为了进一步说明实验的有效性,从测试集中随机挑选了 3 对图像,图 5 为不同实验的可视化结果,图中用框标注了差异区域。从图中可以看出,相比于实验 1,所提算法亮度部分有明显的增强,实验 2

融合结果和所提算法差别不是很显著。

表 1 是不同实验中 21 张融合图像的 6 个客观指标的均值,表中每个指标的最优值用粗体标注,次优值用斜体标注。从表 1 可以看到,实验 2 融合结果相比于实验 1,在图像熵、标准差、视觉信息保真度和互信息 4 个指标都有显著的提升。值得注意的是,所提算法在所有指标都取得最优结果,说明其融合图像质量最优。

主观感受和客观评价指标说明,所提损失函数

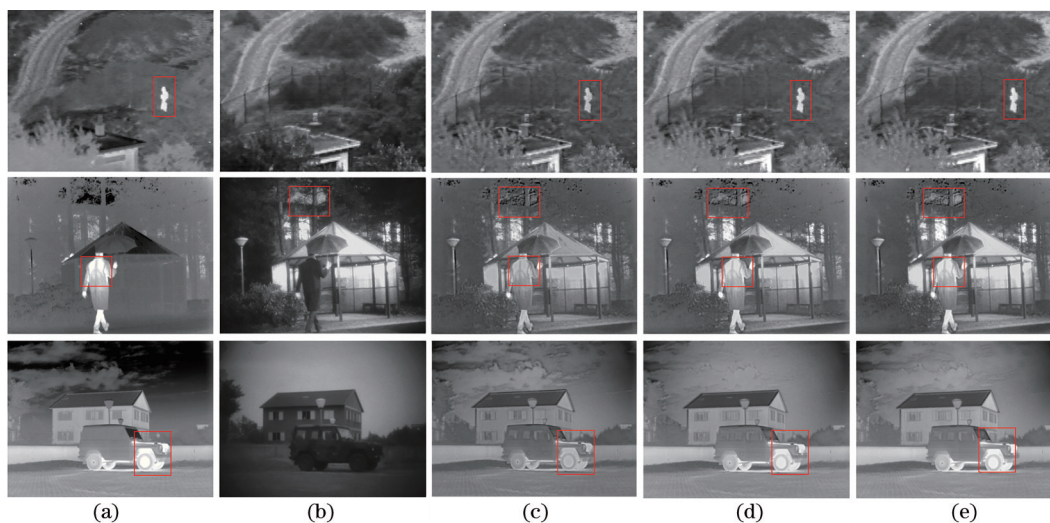


图 5 在 3 对图像上的实验。(a)红外图像;(b)可见光图像;(c)实验 1 的融合结果;(d)实验 2 的融合结果;(e)实验 3 的融合结果  
Fig. 5 Experiments on 3 pairs of images. (a) Infrared images; (b) visible light images; (c) fusion results of experiment 1; (d) fusion results of experiment 2; (e) fusion results of experiment 3

表 1 不同实验中 21 张融合图像评价指标均值  
Table 1 Average quality metrics of 21 fused images in different experiments

Experiment	EN	SD	VIF	MI	$Q_{abf}$	SCD
Experiment 1	6.5949	63.7026	0.6107	13.1898	0.5764	1.7738
Experiment 2	6.7190	71.7095	0.6483	13.4379	0.5683	1.7631
Experiment 3	<b>6.7969</b>	<b>79.6500</b>	<b>0.6739</b>	<b>13.5938</b>	<b>0.6023</b>	<b>1.8067</b>



计算方式和在特征融合过程中引入的注意力机制都是有效的。

### 3.3 对比实验

将所提算法与现有融合算法进行了对比。首先分别在“街道”“人”图像中进行可视化比较,结果如图 6、7 所示。图中用实线框标记亮度特征,虚线框标记纹理特征,左下角和右上角框中内容为放大两倍后的标记信息。

图 6 是不同算法在“街道”图像上的融合结果,

从图中可以看到,DCHWT 融合图像包含较多的噪声信息,视觉感受较差。JSR 融合图像包含的热辐射信息过多,图像亮度值偏大。GTF、FusionGAN 融合图像的热辐射信息比较显著,但是纹理细节信息严重丢失,如虚线标记框中的广告牌字母及实线标记框中目标人物轮廓都很模糊。CSR、VGGML、DenseFuse 的融合图像比较平滑,纹理信息不够清晰。WLS、IFCNN 和所提算法融合结果相对较好,与 IFCNN 相比,所提算法的热辐射信息更加显著。

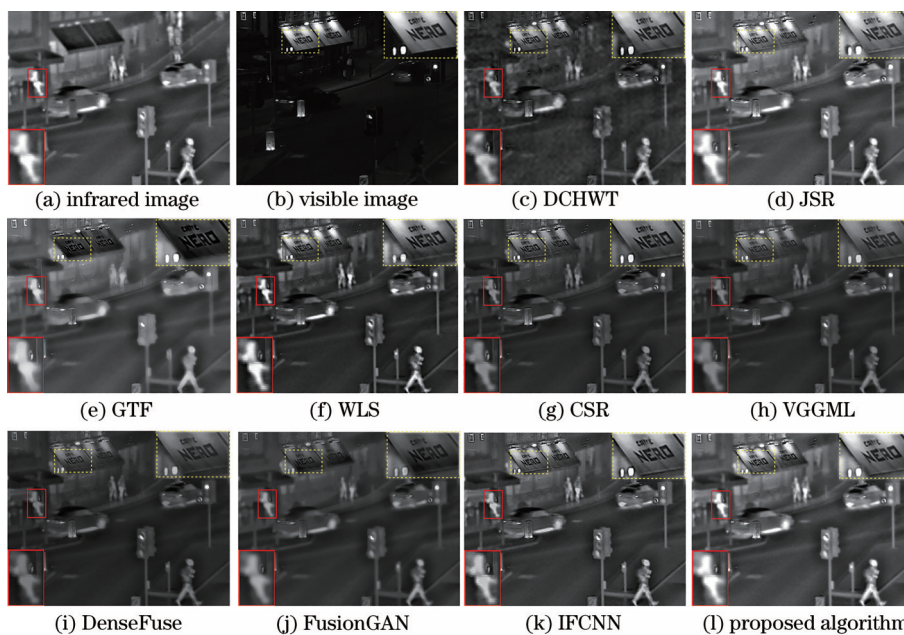


图 6 不同算法在“街道”图像实验

Fig. 6 Experiments with different algorithms on “street” images

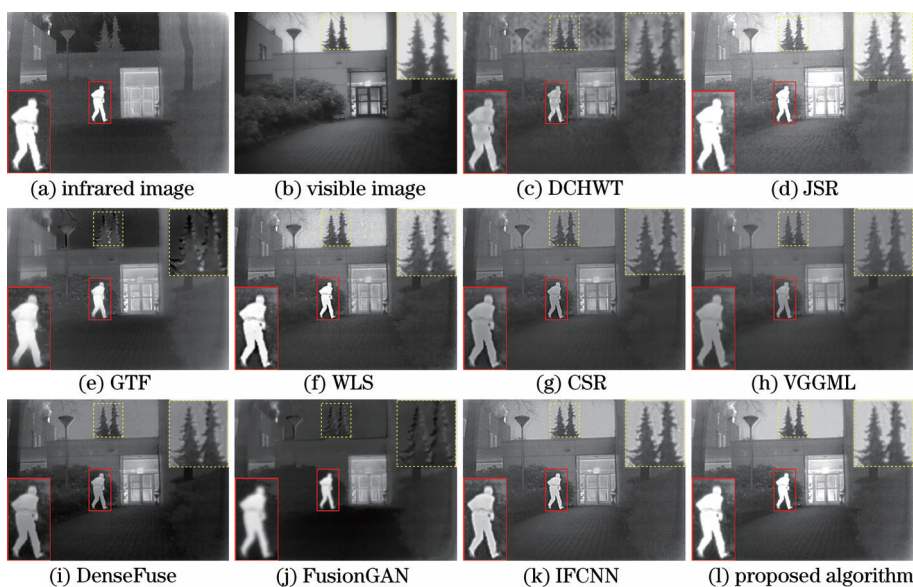


图 7 不同算法在“人”图像实验

Fig. 7 Experiments with different algorithms on “man” images



图 7 是不同算法在“人”图像融合结果。DCHWT 融合图像仍然有大量的噪声信息,JSR 融合图像的亮度失真。GTF、FusionGAN、CSR、VGGML 和 DenseFuse 对纹理的刻画不足,树的轮廓不够清晰(虚线矩形框)。IFCNN 的热辐射信息不够显著。WLS 及所提算法的融合图像都有良好的视觉效果。

通过对图 6 和图 7 分析可以得出,所提融合算法对源图像中的纹理细节信息和亮度信息能够有较好的保留,融合结果超过大多数算法。为了进一步探究所提算法的有效性,从测试集中随机选取 6 对图像,然后用不同算法进行融合,融合结果如图 8 所示。从图中可以看出,所提算法对红外图像和可

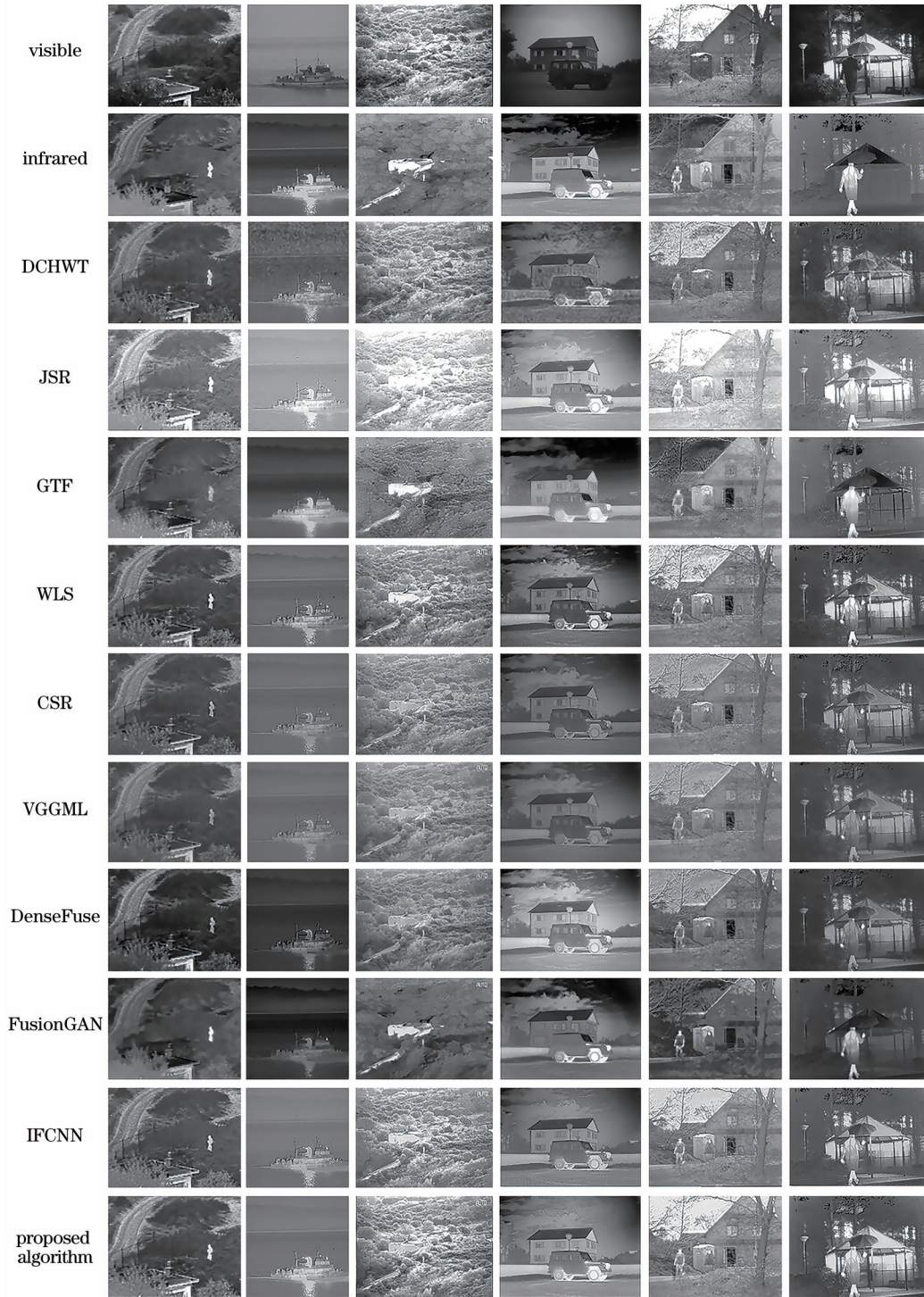


图 8 不同算法在 6 对图像上的实验

Fig. 8 Experiments with different algorithms on 6 pairs of images



见光图像进行了有效的融合,表明所提算法对红外图像与可见光图像的融合任务是普适的。

最后从客观角度分析不同算法融合图像的质量。表 2 是不同算法融合的 21 张融合图像的平均指标。表中用粗体标注每个指标的最优值,斜体标注每个指

标的次优值。从表中可以看出,所提算法在 6 个指标中,4 个指标达到最优,2 个达到次优,表明在所比较的 10 种不同的融合算法中,所提算法的融合图像信息最丰富,融合图像包含源图像中的互补信息最多,且对源图像的边缘纹理信息也能较好保留。

表 2 21 张融合图像的平均指标  
Table 2 Average index of 21 fused images

Algorithm	EN	SD	VIF	MI	$Q_{abf}$	SCD
DCHWT	6.5678	64.9789	0.5056	13.1355	0.4659	1.6099
JSR	6.3644	<i>74.1159</i>	0.6407	12.7287	0.3593	1.7517
GTF	6.6353	67.6260	0.4136	13.2707	0.4247	1.0134
WLS	6.6407	70.5889	<b>0.7287</b>	13.2814	0.5008	1.7961
CSR	6.2587	50.7437	0.3922	12.5174	<i>0.5348</i>	1.6482
VGGML	6.1819	48.1385	0.2949	12.3639	0.3677	1.6348
DenseFuse	<i>6.7247</i>	66.0013	0.6601	<i>13.4493</i>	0.4009	<b>1.8083</b>
FusionGAN	6.3629	54.3575	0.4535	12.7257	0.2183	1.4569
IFCNN	6.5955	66.8758	0.5903	13.1909	0.5033	1.7137
Proposed algorithm	<b>6.7969</b>	<b>79.6500</b>	<i>0.6739</i>	<b>13.5938</b>	<b>0.6023</b>	<i>1.8067</i>

## 4 结 论

对红外图像和可见光图像的融合任务,提出了一个无监督的融合算法。所提算法使用基于方差加权的方式计算损失函数,并引入注意力模块进一步优化实验结果。实验结果表明,所提损失函数计算方式能够融合可见光图像和红外图像的显著特征,融合结果与现有的算法相比无论在可视方面还是在客观指标方面都有较出色的表现。所提算法仍有不足之处,例如对编码器提取到的特征只对最后 1 层网络提取到的特征进行融合而忽略了浅层特征的融合,这在一定程度上会丢失一些信息。此外,注意力机制模块只考虑了通道维度,没有考虑空间维度。这些问题都是今后研究中所要改进的方向。

## 参 考 文 献

- [1] Piella G. A general framework for multiresolution image fusion: From pixels to regions[J]. Information Fusion, 2003, 4(4): 259-280.
- [2] Tang C, Ling Y S, Yang H, et al. Decision-level fusion tracking for infrared and visible spectra based on deep learning[J]. Laser & Optoelectronics Progress, 2019, 56(7): 071502.  
唐聪, 凌永顺, 杨华, 等. 基于深度学习的红外与可见光决策级融合跟踪[J]. 激光与光电子学进展, 2019, 56(7): 071502.
- [3] Ma J Y, Ma Y, Li C. Infrared and visible image fusion methods and applications: a survey[J]. Information Fusion, 2019, 45: 153-178.
- [4] Ben H A, He Y, Krim H, et al. A multiscale approach to pixel-level image fusion[J]. Integrated Computer-Aided Engineering, 2005, 12(2): 135-146.
- [5] Xu L, Cui G M, Zheng C P, et al. Fusion method of visible and infrared images based on multi-scale decomposition and saliency region extraction[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111003.  
许磊, 崔光芒, 郑晨浦, 等. 基于多尺度分解和显著性区域提取的可见光红外图像融合方法[J]. 激光与光电子学进展, 2017, 54(11): 111003.
- [6] Yang B, Li S T. Multifocus image fusion and restoration with sparse representation[J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(4): 884-892.
- [7] Zhang Q H, Fu Y L, Li H F, et al. Dictionary learning method for joint sparse representation-based image fusion[J]. Optical Engineering, 2013, 52(5): 057006.
- [8] Zhang T F, Zhong S C, Lian C M, et al. Deep learning feature fusion-based retina image classification [J]. Laser & Optoelectronics Progress, 2020, 57(24): 241025.  
张添福, 钟舜聪, 连超铭, 等. 基于深度学习特征融合的视网膜图像分类[J]. 激光与光电子学进展,

- 2020, 57(24): 241025.
- [9] Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 120005.  
段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法及其主流框架综述[J]. *激光与光电子学进展*, 2020, 57(12): 120005.
- [10] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [12] Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network[J]. *Information Fusion*, 2017, 36: 191-207.
- [13] Li H, Wu X J, Kittler J. Infrared and visible image fusion using a deep learning framework[C] // 2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 2705-2710.
- [14] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2614-2623.
- [15] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M] // Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8693: 740-755.
- [16] Zhang Y, Liu Y, Sun P, et al. IFCNN: a general image fusion framework based on convolutional neural network [J]. *Information Fusion*, 2020, 54: 99-118.
- [17] Prabhakar K R, Srikanth V S, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4724-4732.
- [18] Mustafa H T, Yang J, Zareapoor M. Multi-scale convolutional neural network for multi-focus image fusion [J]. *Image and Vision Computing*, 2019, 85: 26-35.
- [19] Ma J Y, Yu W, Liang P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, 48: 11-26.
- [20] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: From error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [21] TOET. A. TNO image fusion dataset, 2014[EB/OL]. [2021-02-20]. [https://figshare.com/articles/TN\\_Image\\_Fusion\\_Dataset/1008029](https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029).
- [22] Ma J L, Zhou Z Q, Wang B, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. *Infrared Physics & Technology*, 2017, 82: 8-17.
- [23] Kumar B K S. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform[J]. *Signal, Image and Video Processing*, 2013, 7(6): 1125-1143.
- [24] Ma J Y, Chen C, Li C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information Fusion*, 2016, 31: 100-109.
- [25] Liu Y, Chen X, Ward R K, et al. Image fusion with convolutional sparse representation[J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1882-1886.
- [26] van Aardt J. Assessment of image fusion procedures using entropy, image quality, and multispectral classification[J]. *Journal of Applied Remote Sensing*, 2008, 2(1): 023522.
- [27] Rao Y J. In-fibre Bragg grating sensors[J]. *Measurement Science and Technology*, 1997, 8(4): 355-375.
- [28] Han Y, Cai Y Z, Cao Y, et al. A new image fusion performance metric based on visual information fidelity[J]. *Information Fusion*, 2013, 14(2): 127-135.
- [29] Sulaiman M A, Labadin J. Feature selection based on mutual information[C] // 2015 9th International Conference on IT in Asia (CITA), August 4-5, 2015, Sarawak, Malaysia. New York: IEEE Press, 2015: 1-6.
- [30] Piella G, Heijmans H. A new quality metric for image fusion[C] // Proceedings 2003 International Conference on Image Processing, September 14-17, 2003, Barcelona, Spain. New York: IEEE Press, 2003: 111-117.
- [31] Aslantas V, Bendes E. A new image quality metric for image fusion: the sum of the correlations of differences[J]. *AEU-International Journal of Electronics and Communications*, 2015, 69(12): 1890-1896.