

红外与可见光图像深度学习融合方法综述

李霖 王红梅 李辰凯

A review of deep learning fusion methods for infrared and visible images

Li Lin, Wang Hongmei, Li Chenkai

在线阅读 View online: <https://doi.org/10.3788/IRLA20220125>

您可能感兴趣的其他文章

Articles you may be interested in

红外与可见光图像融合技术的研究进展

Research progress of infrared and visible image fusion technology

红外与激光工程. 2021, 50(9): 20200467 <https://doi.org/10.3788/IRLA20200467>

采用卷积神经网络的红外和可见光图像块匹配

Infrared-visible image patches matching via convolutional neural networks

红外与激光工程. 2021, 50(5): 20200364 <https://doi.org/10.3788/IRLA20200364>

基于FDST和双通道PCNN的红外与可见光图像融合

Infrared and visible image fusion based on FDST and dual-channel PCNN

红外与激光工程. 2019, 48(2): 204001 <https://doi.org/10.3788/IRLA201948.0204001>

红外与可见光图像融合的汽车抗晕光系统

Vehicles anti-halation system based on infrared and visible images fusion

红外与激光工程. 2017, 46(8): 818005 <https://doi.org/10.3788/IRLA201746.0818005>

多输入融合对抗网络的水下图像增强

Multi-input fusion adversarial network for underwater image enhancement

红外与激光工程. 2020, 49(5): 20200015 <https://doi.org/10.3788/IRLA20200015>

改进IHS-Curvelet变换融合可见光与红外图像抗晕光方法

Anti-halation method of visible and infrared image fusion based on improved IHS-Curvelet transform

红外与激光工程. 2018, 47(11): 1126002 <https://doi.org/10.3788/IRLA201847.1126002>

红外与可见光图像深度学习融合方法综述

李霖, 王红梅, 李辰凯

(西北工业大学 航天学院, 陕西 西安 710072)

摘要: 红外与可见光图像融合技术充分利用不同传感器的优势, 在融合图像中保留了原图像的互补信息以及冗余信息, 提高了图像质量。近些年, 随着深度学习方法的发展, 许多研究者开始将该方法引入图像融合领域, 并取得了丰硕的成果。根据不同的融合框架对基于深度学习的红外与可见光图像融合方法进行归类、分析、总结, 并综述常用的评价指标以及数据集。另外, 选择了一些不同类别且具有代表性的算法模型对不同场景图像进行融合, 利用评价指标对比分析各算法的优缺点。最后, 对基于深度学习的红外与可见光图像融合技术研究方向进行展望, 总结红外与可见光融合技术, 为未来研究工作奠定基础。

关键词: 图像融合; 红外图像; 可见光图像; 卷积神经网络; 自编码器网络; 生成对抗网络
中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20220125

A review of deep learning fusion methods for infrared and visible images

Li Lin, Wang Hongmei, Li Chenkai

(School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Infrared and visible image fusion technology makes full use of the advantages of different sensors, retains the complementary information and redundant information of the original image in the fused image, and improves the image quality. In recent years, with the development of deep learning methods, many researchers have begun to introduce this method into the field of image fusion, and have achieved fruitful results. According to different fusion frameworks, the infrared and visible image fusion methods based on deep learning are classified, analyzed and summarized, the commonly used evaluation indicators and data sets are reviewed. In addition, some representative algorithm models of different categories are selected to fuse different scene images, the advantages and disadvantages of each algorithm are compared and analyzed by evaluation indicators. Finally, the research direction of infrared and visible image fusion technology based on deep learning is prospected, infrared and visible fusion technology is summarized, which is the basis for future research work.

Key words: image fusion; infrared image; visible image; convolutional neural network; autoencoder network; generative adversarial network

收稿日期: 2022-02-23; 修订日期: 2022-04-25

基金项目: 国家自然科学基金 (61771400); 陕西省重点研发计划 (2020GY-014)

作者简介: 李霖, 男, 硕士生, 主要从事深度学习、图像融合方面的研究。

导师简介: 王红梅, 女, 教授, 硕士生导师, 主要从事深度学习、目标识别以及图像融合方面的研究。

0 引言

图像融合技术是将来自不同场景的图像合并成一幅融合图像,该融合图像具有多幅原图像特征。可以参与融合的图像有很多,如红外和可见光图像^[1-3]、红外和 SAR 图像^[4-5]、可见光和 SAR 图像^[6-7]、医学中的 CT 和 MRI 图像^[8-9]等。其中,红外图像具有良好的目标检测、识别能力,可以避免外界环境的影响,比如:烟雾、光照、雨天等^[10]。但是它也存在一些不足,比如:像素分辨率低、对比度较差、背景纹理模糊等。可见光图像与人眼视觉特征一致,具有高的分辨率,可以反映丰富的场景信息,比如:纹理信息和细节信息。但是可见光图像容易受到天气、烟雾、遮挡等^[10]环境因素的影响,在受到干扰情况下,可见光图像不能突出场景目标。为了同时得到红外图像目标以及可见光图像纹理信息,融合技术成为必要选择,它可以结合二者的互补性优势,从而得到一幅目标明亮、背景丰富的融合图像。当前,红外和可见光融合技术被广泛地应用于图像增强^[11]、目标识别^[12]、目标检测^[13]、目标跟踪^[14]、农业自动化^[15]、遥感检测^[16]等领域。

在过去的几十年中,已经提出了许多不同的图像融合方法。依据融合方法的主干思路,这些融合方法可以分为不同的类别,包括基于多尺度变换的方法^[17-21]、基于稀疏表示的方法^[22-24]、基于神经网络的方法^[25-26]、基于子空间的方法^[27-29]、基于显著性的方法^[30-31]、混合模型方法^[32-33]和其他融合方法^[2, 34-35]。基于多尺度变换融合方法包括多尺度分解、多尺度融合和多尺度重构 3 部分,其分解重组形式包括小波^[17,19]、金字塔^[18,20]、曲波^[21]等。基于稀疏表示方法是从一组训练图像中学习一个过完备字典进行图像融合,构造字典是该类方法的关键。基于神经网络方法模仿人脑的感知行为来处理数据信息,结构简单,且具有一定的容错率。基于子空间方法使用类似想法,但是在完整基础的空间中实现信息处理,主要包括主成分分析^[27]、独立成分分析^[28]、非负矩阵分解^[29]等。显著性方法利用人类视觉系统对像素弱弱的敏感性来捕获感兴趣区域。基于上述方法,混合模型结合其优点来提高融合性能。其他方法主要基于总变差^[2]、模糊理论^[34]、熵^[34-35]等。整体而言,上述传统方

法为红外和可见光图像融合技术奠定了基础,为新方法的研究提供了帮助。

虽然传统融合方法取得了一定的进展,但是依然存在一些问题,总的来说,这些方法依赖手动设计特征提取和融合规则,而仅仅依靠人工设计的特征提取和融合规则使融合过程越来越复杂^[10,36]。另外,这种方式提取的特征缺乏多样性,往往导致融合图像对比度低^[23]、纹理模糊^[26,37]、目标存在伪影^[37]等。基于深度学习的图像融合方法借助网络的学习能力,自适应地训练更新模型参数,形成端对端的输入-输出模式。与传统方法相比,该方法避免了活动水平测量和融合规则设计,大大降低了人为因素对融合结果的影响^[36]。另外,深度学习方法发挥了网络特征提取的能力,在融合图像中充分保存原图像的互补信息,提高了融合图像质量。

近年来,基于深度学习的融合算法、理论以及应用得到快速发展,并取得一系列研究成果。文中将对近几年的深度学习融合算法展开研究,并进行归类、对比等。内容安排如下:引言部分介绍红外和可见光图像融合的实际意义、应用以及研究背景。第 1 节归类、总结近几年的深度学习融合方法。第 2 节概述常用于衡量融合图像质量的指标,并对各指标进行详细阐述。第 3 节展示了该领域一些典型数据集,并对比分析不同类别的典型算法在这些数据集上的融合结果。第 4 节分析当前基于深度学习的红外与可见光图像融合方法的不足以及发展方向。在第 5 节中,笔者总结概括了整篇文章。

1 基于深度学习的红外和可见光图像融合方法

与传统方法依靠人工活动水平测量和融合规则设计相比,基于深度学习的融合方法通过网络自主学习训练,避免了人为操作所带来的影响,实现了端对端的融合方式。在近些年,越来越多的深度神经网络融合方法被应用于图像融合领域,其方法大致分为 3 类:卷积神经网络方法、自编码解码网络方法以及生成对抗网络方法^[36]。为了进一步总结和探索基于深度学习的红外和可见光图像融合方法,下面笔者将对近些年来提出的深度学习融合方法进行详细描述。

1.1 基于卷积神经网络的融合方法

在 1998 年, LeCun 等^[38] 提出了卷积神经网络, 并在文中提出了 Le-Net5 模型, 此模型可以很好地识别 Mnist 手写字, 之后被应用于银行钞票识别。2012 年, Krizhevsky 等^[39] 利用卷积神经网络模块搭建了一种更深层次的 AlexNet 网络, 它被应用于图像分类, 并成功赢得当年 ILSVRC 比赛冠军, 成为了深度学习方法的开山之作。2017 年, Liu 等^[40] 将深度卷积网络引入到图像融合领域, 他们利用模糊的背景和前景图像来训练网络, 并得到一张二值化的权重图谱。在测试阶段, 原图像结合权重图谱得到一张融合的多焦点图像。但该网络是一个分类网络, 并不适用于红外和可见光图像融合, 因此, 研究者们开始尝试把深度学习方法引入红外和可见光图像融合领域。

在初始阶段, 大多数研究者尝试在传统方法中引入深度学习模块, 这种方式给融合图像注入了丰富的语义信息。比如: Li 等^[41] 利用 VGG19 网络^[42] 进一步处理多尺度分解后的细节部分, 从而在融合图像中保

留丰富的纹理信息; Liu 等^[43] 发现卷积神经网络提取到的特征在一定程度上可以反映原图像在融合过程中的占比, 因此, 他们把卷积后权重图的下采样序列作为两个支路下采样序列的融合比例图, 避免了人为设计融合策略, 其网络结构图如图 1 所示; 类似地, Li 等^[44] 采用零相分量分析及 L1 范数等操作得到一张反映原图像占比的权重图, 克服了 Liu 等下采样原图像过程中信息丢失的问题。这些方法借助网络强大的特征提取能力, 在融合图像中保留了丰富的细节信息。但是, 上述深度学习模块主要用于特征提取阶段, 原图像多尺度分解或者融合策略依然采用传统算法。该类方法的主要不足包括: (1) 多数算法依然需要活动水平测量或者融合规则设计; (2) 卷积神经网络在整个图像融合过程参与度不够, 仅用于提取融合需要的权重图, 融合规则采用简单的加权融合。造成上述问题的原因在于深度学习融合方法没有摆脱传统方法的限制, 从而不能最大程度上发挥网络自身优势。

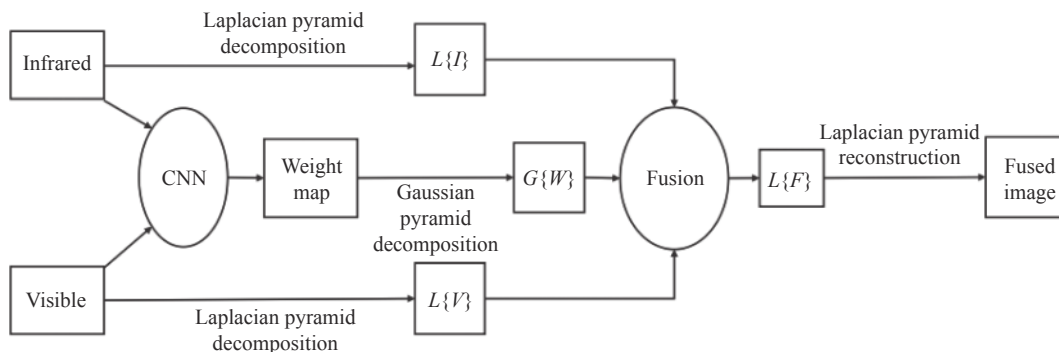


图 1 基于卷积神经网络的红外与可见光图像融合框架^[43]

Fig.1 Infrared and visible image fusion framework based on convolutional neural network^[43]

在此之后, 红外和可见光图像的端对端卷积神经网络融合模型被逐步建立起来, 它的输入大致分为两种: 单通道级联输入^[45] 以及多通道输入^[46-47], 网络模型主要利用简单卷积神经网络^[38]、残差网络^[48]、密集残差网络^[49] 等, 结合合适的损失函数, 实现了端对端的融合模式。比如: Li 等^[50] 采用两个通道分别输入原图像到两个的卷积神经网络, 并将得到的特征图级联后输入到一个卷积层得到融合图像。虽然该方法可以完成图像融合工作, 但融合结构简单, 训练容易产生过拟合, 因此, 融合图像质量易受影响。参考文

献 [47,51] 在两通道输入的基础上, 将密集连接网络作为特征提取模块, 同时设计更深层的重组模块以增加信息的保留量。除此之外, 基于密集残差网络, Long 等^[52] 提出一种深度密集残差网络, 增加了融合图像信息量; 参考文献 [51,53] 在网络中嵌入注意力机制模块, 选择性地为融合图像提供了更多的亮度和梯度信息。针对上述深度学习融合模型泛化性弱的问题, Xu 等^[54] 提出自适应信息保留度, 即自动评估并得到原图像的重要性参数, 训练得到的网络具有良好的泛化能力。进一步地, 参考文献 [55] 直接对原图像

进行评价,同样得到自适应权重系数,相对前者,融合方法被进一步简化。

经历了萌芽、兴起和发展阶段,基于卷积神经网络的红外和可见光图像融合方法从初始仅借助简单卷积神经网络提取图像特征向网络结构的多元化发

展,比如:网络自身的融合功能被挖掘,与其他有效模型结合等。整体而言,它的结构简单,模型参数少,更容易被训练和优化。另外,文中对典型卷积神经网络红外和可见光图像融合方法的局限性进行分析,如表 1 所示。

表 1 典型卷积神经网络融合方法局限性

Tab.1 Limitations of typical CNN-based fusion methods

References	Limitation
[40]	Being suitable for mutiil-focus image fusion, only the last convolutional layer features are used to calculate the fusion result
[46]	The information in the middle layer is lost, and the fusion strategy has no theoretical support
[50]	The structure is simple and prone to overfitting
[54]	The model mainly saves detailed texture information and cannot highlight infrared targets

1.2 基于自编码器的融合方法

2017 年, Prabhakar 等^[56]提出了一种新颖的深度学习网络 (DeepFuse), 并实现了对多曝光图像的融合, 该算法首先利用卷积神经网络提取原图像的特征图谱, 之后对各维度上的特征图谱进行融合, 最后通过解码层还原得到一张融合图像。这种编码解码结构在一定程度上增强了融合图像的信息丰富度, 但融合图像的质量主要依赖编码层最后一层网络提取的特征。

2018 年, 在 DeepFuse 的基础上, Li 等^[57]提出了一种更加复杂的深度学习网络, 网络结构主要由编码层、融合层以及解码层组成, 其网络结构如图 2 所示。在编码层, 该方法将每层提取到的特征图谱级联到下一层的输入, 其增加了信息流动, 网络更容易被训练, 最终的融合图像保留了大量原图像信息。但是, 由于数据集的缺乏, 该网络采用 MS-COCO 数据集^[58]的灰度图像训练自编码网络, Zhao 等^[59]利用红外和可见光图像作为训练数据, 结合跳级连接操作, 预训练网络更能突出红外和可见光图像特有信息。即便上述方法提高了融合图像质量, 但采用相同的卷积操作分别采样原图像会导致融合结果细节信息丢失, 参考文献 [60-61] 分别设计用于红外和可见光图像训练的编码-解码网络, 因此, 网络更具针对性。另外, Liu 等^[62]针对残差网络, 提出自适应权重分配策略, 该方法给每张传递的特征图谱赋予权重, 网络模型得到优化, 同时融合结果可以保留更多细节信息。为

防止中间层信息在传递过程中丢失, 参考文献 [63-65] 采用嵌套式网络充分提取每层卷积得到的特征序列, 因此融合图像保留了不同尺度的特征; Fu 等^[66]设计颜色损失和感知损失增加融合图像的细节和语义信息; Zhang 等^[67]则利用路径转移模块融合各支路卷积层之间的信息。和卷积神经网络融合方法区别在于, 自编码器融合方法需要设计融合策略, 而当前多数网络的融合策略依然是简单相加、平均或 L1 范数^[57,59-61], Jian 等^[68]用 Softmax 操作得到不同通道特征图对应的权重, 融合图像很好地保留了红外图像的热辐射信息。Li 等^[64]设计了一个新颖的残差融合模块, 通过分别训练编码-解码器和残差融合模块, 从而避免了人工设计融合策略。类似地, Zhao 等^[69]也设计了一种两个阶段训练的融合模型, 第一阶段利用自监督策略在编码区的不同通道保留原图像信息, 第二阶段把解码器替换为融合增强模块, 最终得到一幅细节增强的融合图像。另外, 为了让网络对目标以及背景更敏感, Ma 等^[70]结合语义分割方法, 得到目标和背景分离的二值掩膜, 用该掩膜结合原图像得到训练数据。Raza 等^[71]结合传统方法, 训练过程给网络融合特征模块提供边缘特征, 测试阶段叠加编码器生成的融合图像和红外特征, 组成新的融合图像。面向应用, Tang 等^[72]加入语义损失, 融合图像可有效地促进高级视觉任务。

上述基于自编码器的融合方法, 网络结构主要包括 3 部分: 编码器、解码器以及融合模块。其中, 编码

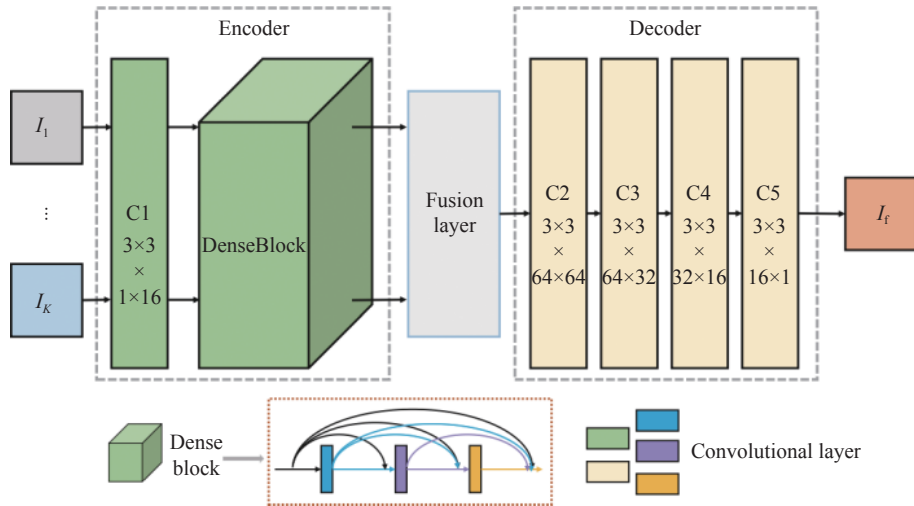


图 2 基于自编码器网络的红外与可见光图像融合框架^[54]

Fig.2 Infrared and visible image fusion framework based on autoencoder network^[54]

器可以对图像有用特征进行提取;融合层既可以利用网络融合特征图,也可以人为设计融合策略;解码器完成图像的重构。整体而言,该网络模块可拆分和重

组,迁移能力强,更容易被修改以及监测。另外,文中对典型自编码器红外和可见光图像融合方法的局限性进行分析,详见表 2。

表 2 典型自编码网络融合方法局限性

Tab.2 Limitations of typical autoencoder-based fusion methods

References	Limitation
[57]	The model is not targeted enough to highlight the infrared target, and the fusion strategy is simple
[64]	Insufficient attention to texture information, large amount of network parameters are not conducive to application
[68]	Network channels share weights, pre-training models focus on common information, and unique information may be lost
[67]	Abundant texture details cannot be obtained

1.3 基于生成对抗网络的融合方法

2014 年, Goodfellow 等^[73]提出了生成对抗网络(GAN),网络模型分为两个部分:生成器和鉴别器。生成器可以利用随机噪声产生一个新的数据样本;鉴别器是一个二分类器,它的输入是真实数据以及生成器产生的样本数据。训练过程中,生成器和鉴别器会形成一个对抗关系,训练过程中鉴别器给生成器生成的样本打一个低的分数,但是,随着训练过程参数的更新,生成器生成的新数据样本和真实样本越来越相近,鉴别器最终无法判别虚假样本。基于该网络结构的特点,它被广泛应用于风格转换^[74]、目标检测^[75]、图像增强^[76]、图像融合^[10,77-78]等领域。

在图像融合领域,由于融合图像的真值图像不能被得到,更多的融合过程只能依靠无监督或者半监督

来实现,生成对抗网络的出现给无监督条件下的图像融合工作带来便利。2019 年, Ma 等^[10]提出一个新颖的基于生成对抗网络的融合模型(FusionGAN),实现了红外和可见光图像的融合,其网络框架如图 3 所示,主要分为两个部分:生成器和鉴别器,生成器的输入为级联后的红外和可见光图像,鉴别器的输入为生成器生成的图像和可见光图像,损失函数分为两部分:生成器损失和鉴别器损失,生成器的损失分为对抗损失和背景损失,鉴别器损失迫使融合图像和可见光图像更相似。该网络利用生成对抗网络的特点,为融合图像保留了一定的纹理和目标信息,但其对比度较低、融合图像暗淡,主要原因为:(1)生成器忽视了亮度信息的保留;(2)“标签”设置为可见光图像。因此,参考文献[79]设计两个鉴别器,同时保留纹理、亮度

信息; Li 等^[80]把引导滤波融合算法的融合图像^[81]作为标签,迫使融合图像同时保留可见光图像和红外图像互补信息。除此之外,研究者们还为融合图像边缘、细节、优化模型等方面设计针对性的损失函数,比如:设计目标边缘增强损失^[82]、局部二进制损失^[83-84]保留原图像边缘细节; Li 等^[85]惩罚浅层判别器的原图像和融合图像注意力图谱,旨在保存更多原图像注意区域信息;参考文献^[84,86]把带有梯度惩罚的 Wasserstein 距离作为鉴别器的损失函数,模型更容易被训练。在网络结构上,参考文献^[87]采用多尺度注意力机制强制生成器专注原图像最具辨别力的区域;为了

使融合结果更加均衡, Ma 等^[88]把鉴别器设置为一个分类网络,最终鉴别器不再能分辨出融合图像和真实数据。另外, Yang 等^[86]结合传统方法,首先通过引导滤波分解原图像到多个尺度,之后对细节模块和基础模块分别进行特征提取、融合及重组,融合图像在保留强度纹理信息的同时,也保留了背景和细节信息。考虑到红外和可见光图像的融合最重要的目的是突出红外图像目标并保留可见光图像背景,参考文献^[89-90]用语义分割方法^[91]得到原图像的掩膜,结合该掩膜图像进一步得到鉴别器的“标签”,因此预训练网络对红外目标和可见光纹理信息更加敏感。

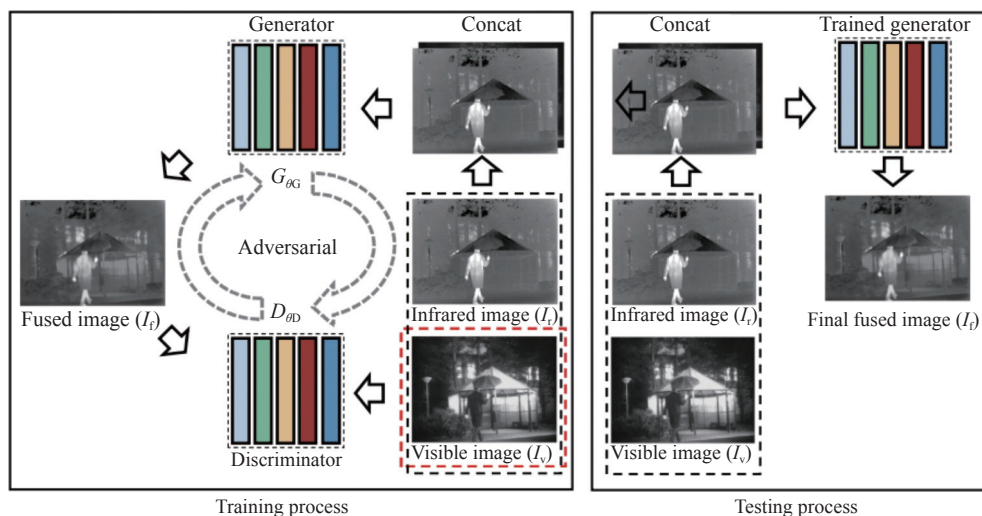


图 3 基于生成对抗网络的红外与可见光图像融合框架^[10]

Fig.3 Infrared and visible image fusion framework based on generative countermeasure network^[10]

上述基于生成对抗网络融合方法,主要依靠生成器和鉴别器的对抗关系生成融合图像。训练过程中,“对抗游戏”可以迫使融合图像的某些特征和原图像保持一致,比如:融合图像的纹理、亮度信息等,因此,网络可以实现无监督的融合任务。但是,该方

法由于模型的优化不稳定,训练相对较困难,下面对典型生成对抗网络红外和可见光图像融合方法的局限性进行汇总,详见表 3。另外,文中对近年来基于深度学习的红外和可见光融合方法进行分析,详见表 4。

表 3 典型生成对抗网络融合方法局限性

Tab.3 Limitations of typical GAN-based fusion methods

References	Limitation
[10]	Insufficient consideration of infrared brightness information.
[79]	The two adversarial losses are difficult to balance, and fusion image target is distorted
[86]	Under the two-discriminator condition, the Wasserstein distance adversarial loss does not enhance the target brightness
[89]	The lack of well-segmented datasets, the quality of the pre-training model depends on the accuracy of semantic segmentation

表 4 基于深度学习的红外和可见光图像融合方法汇总

Tab.4 Summary of infrared and visible image fusion methods based on deep learning

Type	Typical methods	Characteristic
Input method	Single channel [10], [45], [52], [54-55], [79], [82], [83]	Cascading source images, mining the fusion ability of the network
	Multi-channel [46], [47], [50-51], [53], [56-72]	Distinguishing the source images, but need to design a fusion strategy
	Multi-image multi-channel [67], [88-89]	Inputting the source images in proportion, keeping the same category information of the source images
	Preprocess image [70-71], [86], [89-90]	Providing more useful information for fused images
Common block	Attention network [45], [51], [53], [63], [65], [85], [87]	Enhancing feature maps from channels and spaces, it can be embedded in any network
	Nest network [63-65]	The network structure is complex, and focusing on the shallow and middle layers of the network
	Skip connection [59], [68], [77], [87]	Based on residual and dense networks, it prevents loss of useful shallow information
	Perceptual loss [55], [66], [82], [87]	Balancing feature error between reconstructed image and input
LossFunction	TV loss [47], [79]	constraining the fused image to exhibit similar gradient variation with the visible image
	Edge detail loss [69], [82], [83-84]	Enhancing fusion image edge detail
	Sematic loss [72]	More targeted to different information of the scene

2 融合图像评价

红外和可见光图像的融合技术已经被广泛应用到多个领域,随着进一步发展,融合图像评价准则成为这个领域的焦点。对不同的融合方法,保留信息的方式可能不同,在实际应用过程中可能存在显著差异,为了对融合图像做出评价,许多研究人员开始研究融合图像的评估指标,并设计出一些评价标准,其可以对融合图像进行定性和定量分析。现有的融合准则大致分为主观评价和客观评价,通过这两方面的分析可以充分评估红外与可见光融合图像的质量。

2.1 主观评价方法

主观评价方法在融合图像质量评价过程中发挥着重要作用,其主要依靠人类肉眼的观察,在评价过程中,人类视觉系统可以观察到图像的细节、对比度、图像完整性和失真程度等,进而可以做出初步的评价。这种做法比较直观、简洁。尽管如此,主观评价方法由于没有定性地给出评价参数,很难具有绝对的说服力。同时,主观评价者对于同一张图像可能给出的评价不同,原因是个人的喜好、侧重面不同。为了克服人为因素给融合图像评价结果带来干扰,在主观评价的基础之上,需设计出定性测量的指标对融合图像进行进一步的评估。

2.2 客观评价方法

客观评价可以克服主观评价说服力低的问题。对于红外和可见光融合图像,研究人员针对不同方面,设计出相应的融合评价指标,从而对融合图像做出准确的评价。客观融合指标有很多种,一般通过一些数学公式来定义,此节将对相关评价指标进行详细介绍,常用的客观评价指标包括:信息熵^[92]、结构相似性测量^[93]、标准差^[94]、互信息^[95]、均方误差^[1,79]、空间频率^[96]、峰值信噪比^[97]、平均梯度^[98]、视觉信息保留度^[99]、基于梯度的融合性能^[100]、相关系数^[101]以及其他准则^[1]。计出相应的融合评价指标,从而对融合图像做出准确的评价。客观融合指标有很多种,一般通过一些数学公式来定义,此节将对相关评价指标进行详细介绍,常用的客观评价指标包括:信息熵^[92]、结构相似性测量^[93]、标准差^[94]、互信息^[95]、均方误差^[1,79]、空间频率^[96]、峰值信噪比^[97]、平均梯度^[98]、视觉信息保留度^[99]、基于梯度的融合性能^[100]、相关系数^[101]以及其他准则^[1]。

(1) 信息熵

信息熵反映图像中信息的丰富程度。一般情况下,图像信息量越大,信息熵越大,其数学表达式如下:

$$EN = - \sum_{x=0}^L p(x) \log_2 p(x) \quad (1)$$

式中: x 表示图像灰度级数; $p(x)$ 表示该灰度值的概率分布。

(2) 结构相似度测量

结构相似度测量是从图像亮度、对比度和结构 3 个方面对比原图像和融合图像的相似度。一般情况下, 融合图像和原图像越相似, 结构相似度测量值越大, 其数学表达式如下:

$$SSIM(x, y) = \left(\frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right)^\alpha \cdot \left(\frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right)^\beta \cdot \left(\frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \right)^\gamma \quad (2)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (3)$$

式中: x 、 y 分别代表原图像和融合图像; μ_x 、 μ_y 、 σ_x^2 、 σ_y^2 、 σ_{xy} 分别表示原图像和融合图像的均值、方差和协方差; $c_i(i=1,2,3)$ 是一个较小的正数, 避免分母为零; α 、 β 、 γ 用于调节亮度、对比度和结构 3 部分的比例。

(3) 标准差

标准差反映融合图像信息丰富程度, 用于测量融合图像像素强度的变化。当融合图像的标准差越大, 图像对比度就越高, 语义信息也更加丰富。其数学表达式如下:

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [H(i, j) - \bar{H}]^2} \quad (4)$$

$$\bar{H} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N H(i, j) \quad (5)$$

式中: M 、 N 分别代表融合图像的宽和高; \bar{H} 表示图像灰度值的平均。

(4) 互信息

互信息衡量原图像向融合图像转移的信息量。一般情况下, 该值越大, 说明融合图像和原图像的共同信息越多, 信息保留量越大。其数学表达式如下:

$$MI = MI(x_{ir}, x_f) + MI(x_{vis}, x_f) \quad (6)$$

$$MI(x_i, x_f) = \sum_{x,f} p_{x_i, x_f}(x, f) \log \frac{p_{x_i, x_f}(x, f)}{p_{x_i}(x)p_{x_f}(f)} \quad (7)$$

式中: x_{ir} 、 x_{vis} 、 x_f 分别表示红外图像、可见光图像和融合图像; x_i 表示原图像; $p_{x_i}(x)$ 、 $p_{x_f}(f)$ 分别表示原图像和融合图像的边缘直方图; $p_{x_i, x_f}(x, f)$ 表示原图像和融合图像的联合直方图。

(5) 均方误差

均方误差反映融合图像和原图像之间的像素差。一般情况下, 均方误差值越小, 则融合图像和原图像在像素强度方面越接近。其数学表达式如下:

$$MSE_{XF} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i, j) - F(i, j))^2 \quad (8)$$

式中: X 、 F 分别表示原图像和融合图像。

(6) 空间频率

空间频率反映图像细节和纹理信息, 分为空间行频率和列频率。空间频率值越大, 融合图像的纹理细节和边缘信息丰富, 容易被人类视觉感知。其数学表达式如下:

$$SF = \sqrt{RF^2 + CF^2} \quad (9)$$

$$RF = \sqrt{\sum_{i=1}^M \sum_{j=2}^N (x_{i,j} - x_{i,j-1})^2} \quad (10)$$

$$CF = \sqrt{\sum_{i=2}^M \sum_{j=1}^N (x_{i,j} - x_{i-1,j})^2} \quad (11)$$

式中: SF 、 CF 分别表示行、列的空间频率。

(7) 峰值信噪比

峰值信噪比通过峰值功率和噪声功率之比反映融合图像失真的度量。一般情况下, 峰值信噪比的值越大, 表明融合过程产生的失真越小, 融合后的图像和原图像更相似。其数学表达式如下:

$$PSNR = 10 \log_{10} \frac{r^2}{MSE} \quad (12)$$

式中: r 表示融合图; MSE 表示融合图像和原图像的均方误差。

(8) 平均梯度

平均梯度用于衡量融合图像的清晰度。该值越大, 图像中包含的纹理细节越多, 融合图像质量越好。其数学表达式如下:

$$AG = \frac{1}{(M-1)(N-1)} \cdot \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\frac{[I_f(i+1, j) - I_f(i, j)]^2 + [I_f(i, j+1) - I_f(i, j)]^2}{2}} \quad (13)$$

(9) 视觉信息保留度

视觉信息保留度反映融合图像包含原图像的有

效视觉信息。一般情况下,视觉信息保留度值越大,表明融合图像的视觉效果更好。其数学表达式如下:

$$VIFF(I_1, I_2, I_f) = \sum_k p_k VIFF_k(I_1, I_2, I_f) \quad (14)$$

$$VIFF_k(I_1, I_2, I_f) = \frac{\sum_b FVID_{k,b}(I_1, I_2, I_f)}{\sum_b FVIND_{k,b}(I_1, I_2, I_f)} \quad (15)$$

式中: p_k 是第 k 个子带的加权系数; $FVID_{k,b}$ 、 $FVIND_{k,b}$ 分别表示在第 k 个子带第 b 个模块失真和没有失真的融合视觉信息。

(10) 基于梯度的融合性能

基于梯度的融合性能通过局部度量估计融合图像中原图像边缘信息的保留程度。该值越小,边缘信息丢失越多;相反,边缘信息越完整。其数学表达式如下:

$$Q^{AB/F} = \frac{\sum_{n=1}^N \sum_{m=1}^M Q^{AF}(n,m)w_A(n,m) + Q^{BF}(n,m)w_B(n,m)}{\sum_{i=1}^N \sum_{j=1}^M (w_A(i,j) + w_B(i,j))} \quad (16)$$

$$Q^{XF}(n,m) = Q_g^{XF}(n,m)Q_a^{XF}(n,m) \quad (17)$$

式中: $Q_g^{XF}(n,m)$ 和 $Q_a^{XF}(n,m)$ 分别表示位置 (n,m) 处的边缘强度和方向值; w_A 和 w_B 表示权重。

(11) 相关系数

相关系数反映融合图像和原图像的线性关系,该值越大,说明融合图像和原图像的关系越密切,融合性能更好。其表达式如下:

$$CC = \frac{(r_{I,F} + r_{v,F})}{2} \quad (18)$$

$$r_{X,F} = \frac{\sum_{i=1}^H \sum_{j=1}^W (X(i,j) - \bar{X})(F(i,j) - \bar{F})}{\sqrt{\sum_{i=1}^H \sum_{j=1}^W (x(i,j) - \bar{X})^2 \left(\sum_{i=1}^H \sum_{j=1}^W F(i,j) - \bar{F} \right)^2}} \quad (19)$$

式中: X 表示原图像; \bar{X} 、 \bar{F} 分别表示原图像和融合图像的像素均值; H 、 W 分别表示图像的宽和高。

(12) 其他准则

除上述评价准则外,还存在一些其他的评价方法。比如:交叉熵值^[60]反映了融合图像和原图像之间

的相异性;差异相关性和^[102]用于评估融合图像存在的伪信息;特征互信息^[103]从特征空间角度评价原图像和融合结果的互信息量;运行时间反映了算法的实时性。

综上所述,像素级图像融合评价准则主要从融合图像的信息量、亮度、纹理丰富度、视觉效果等方面评价融合图像质量,下面进一步对比分析不同评价指标特点、区别以及组合策略:(1)信息量:信息熵反应融合图像的信息量,互信息反映原图像向融合图像转移的信息量,在实际评价过程中,当互信息值较高时,信息熵相对较高,反之则不然。(2)亮度:标准差关注融合图像的对比度,区分红外目标和背景,均方误差反映融合图像和原图像的像素信息相似程度,一般用于损失函数。(3)纹理丰富度:空间分辨率和平均梯度计算融合图像整体的梯度,二者在评价结果上表现相似,基于梯度的融合性能则更关注局部边缘梯度。(4)视觉效果:视觉信息保留度反映融合图像视觉信息的保留量。(5)其他方面:峰值信噪比反映融合图像失真情况;结构相似性度量从多方面对比融合图像和原图像的相似性;综合考虑上述评价准则的特点,提出以下组合策略:互信息、标准差、空间分辨率、视觉信息保留度、基于梯度的融合性能、峰值信噪比。另外,大多数情况下,互信息可以代替信息熵,空间分辨率和平均梯度可相互替换,结构相似性度量及相关系数可适度使用。在实验对比阶段,文中将结合评价结果进一步验证各评价指标特点及相似性。

3 实验结果及分析

3.1 数据集

3.1.1 TNO 数据集

TNO 数据集^[104]来自不同波段的相机系统,其包含不同场景下的多光谱图像,比如:视觉增强图像、近红外图像、可见光图像、长波红外图像等。该数据集的相机系统分别是 Athena、DHV、FEL 和 TRICLOBS, Athena 系统提供了一些军事方面的目标和背景图像,比如:飞机、士兵等;DHV、FEL 系统提供了一些场景图像,比如:日出、湖畔等;TRICLOBS 提供了一些日常生活图像,比如:住房、汽车等。文中展示了来自 TNO 数据集不同系统的几幅图像,如图 4 所示。

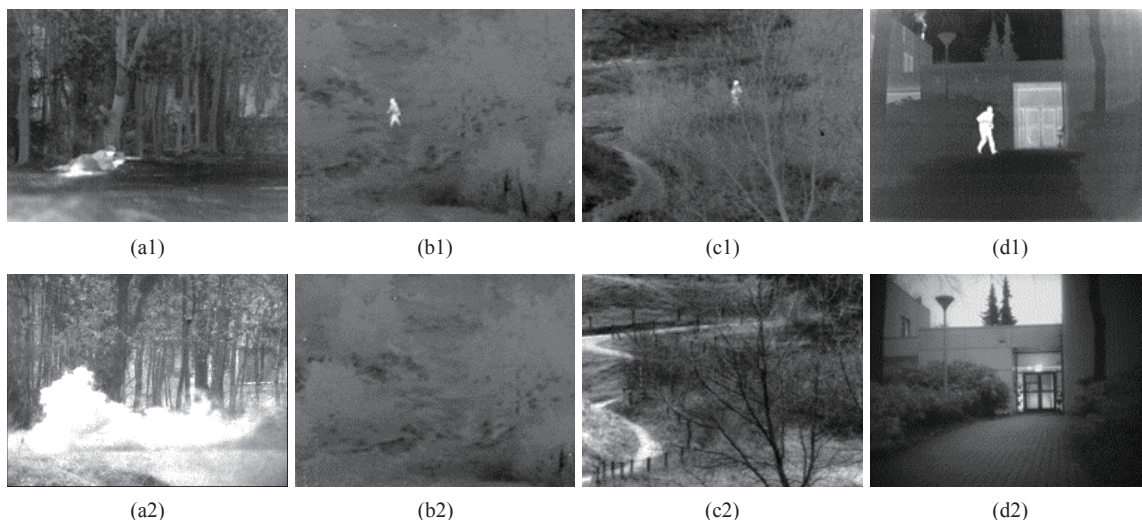


图 4 Athena、DHV、FEL 以及 TRICLOBS 系统采集到的红外和可见光图像。(a1)~(d1) 红外图像;(a2)~(d2) 可见光图像

Fig.4 Infrared and visible images collected by Athena, DHV, FEL and TRICLOBS systems respectively. (a1)-(d1) Infrared images; (a2)-(d2) Visible images

3.1.2 RoadScene 数据集

RoadScene 数据集^[54]包含 221 对红外和可见光图像,它们选自于 FLIR 视频序列,包含了丰富的生活场景,比如:马路、交通工具、行人等。该数据集对原始

的红外图像的背景热噪声进行了预处理,并准确对齐红外和可见光图像对,最终裁剪出精确的配准区域以形成该数据集。文中展示了 RoadScene 数据集不同场景下的图像,如图 5 所示。

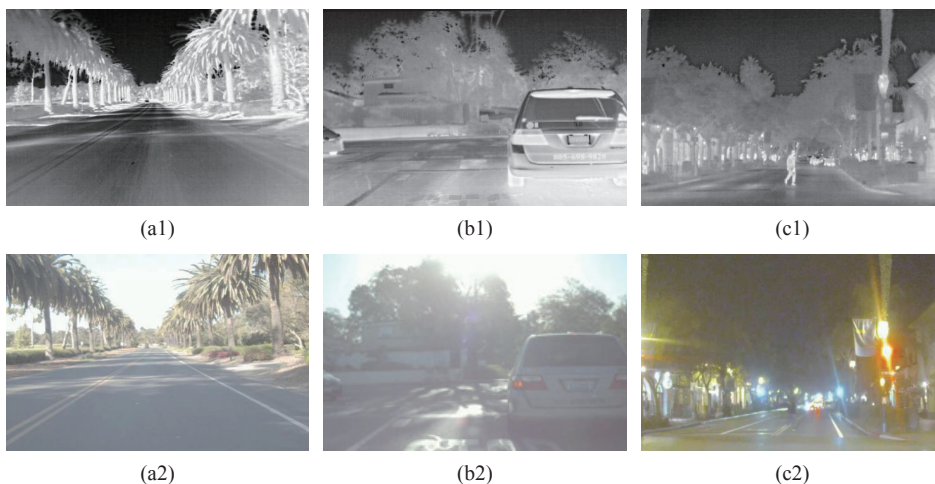


图 5 马路、交通工具、行人图像场景下的图像。(a1)~(c1) 红外图像;(a2)~(c2) 可见光图像

Fig.5 The Images in roads, vehicles, and pedestrians scenes respectively. (a1)-(c1) Infrared images; (a2)-(c2) Visible images

3.1.3 INO 数据集

INO 数据集是由加拿大光学研究所发布的,它包含了几对在不同天气和环境下的可见光和红外视频。比如: BackyardRunner、CoatDeposit、GroupFight、

MultipleDeposit 等。在对预训练模型测试过程中,一般从几个视频序列中随机挑选一些帧来验证模型的有效性。文中展示了该数据集不同视频序列中的一些帧率图像,如图 6 所示。

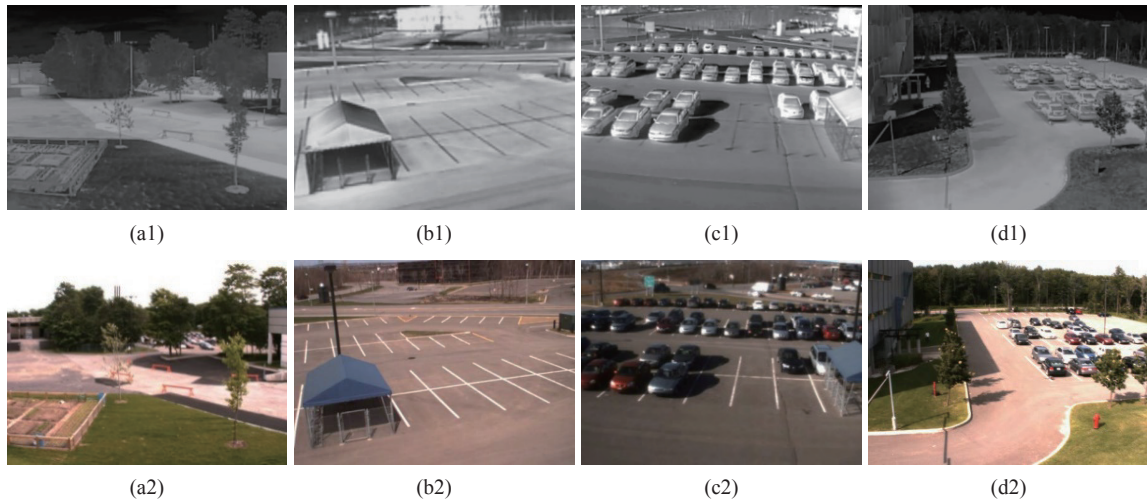


图 6 BackyardRunner、CoatDeposit、GroupFight、MulitpleDeposit 场景下的图像。(a1)-(d1) 红外图像; (a2)-(d2) 可见光图像

Fig.6 The images in BackyardRunner, CoatDeposit, GroupFight, and MulitpleDeposit scenes respectively. (a1)-(d1) Infrared images; (a2)-(d2) Visible images

3.1.4 OTCBVS 数据集

OTCBVS 数据集^[105]用于测试和评估一些新颖和先进的计算机视觉算法,它包括了多个子数据集,比如:热目标行人数据集、红外与可见光人脸数据集、自动驾驶数据集、红外与可见光行人数据集等。其中红外与可见光行人数据集拍摄于俄亥俄州立大学校园内繁忙的道路交叉口,包含了 17089 对红外与可见光图像对,图像大小为 320×240,文中展示了这个子数据集中的个别场景图像,如图 7 所示。

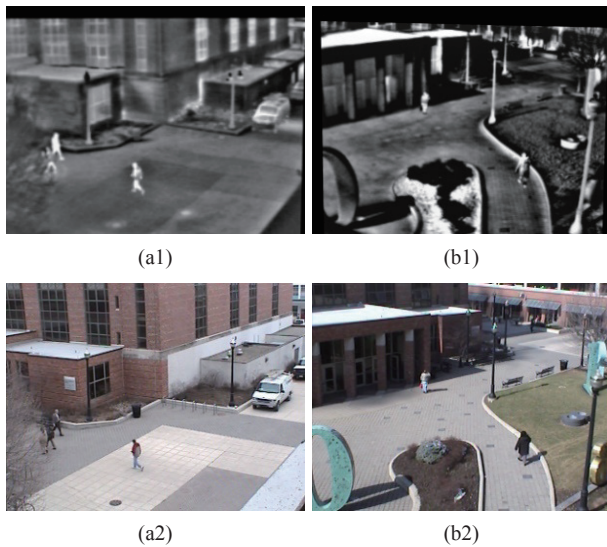


图 7 两个视频序列中的第一帧场景图像; (a1)、(b1) 红外图像; (a2)、(b2) 可见光图像

Fig.7 The first frames in two video sequences respectively. (a1), (b1) Infrared images; (a2), (b2) Visible image

3.1.5 其他数据集

除以上数据集外,还存在一些公开的数据集,比如:MS-COCO 数据集^[58],由于红外图像和可见光图像数据缺乏,研究者用 MS-COCO 数据集的灰度图像训练模型;为了获得红外图像目标,有研究者提出基于语义分割标签的红外和可见光图像数据集^[89],使得网络模型对红外目标和可见光图像背景更敏感。

3.2 实验对比

文中对比分析了近几年一些典型的深度学习融合方法,其中包括:DenseFuse^[57]、FusionDN^[55]、U2 Fusion^[54]、FusionGAN^[10]、DDcGAN^[79]、GANMcC^[88]、RFN-Nest^[64]、STDFusionNet^[70]、SDDGAN^[90]及 SeA Fusion^[72]方法。另外分别从 TNO 数据集、OTCBVS 数据集、INO 数据集各随机选择两张含有背景以及目标的图像作为测试集,对比各类算法的优缺点。实验部分使用的设备为 4.0 GHz AMD Ryzen Threadripper PRO 3945 WX, GPU RTX3080 10 G。

3.2.1 定性比较

文中展示 6 个场景下 10 个深度学习算法的融合结果,如图 8 所示,从左至右分别是 Kaptein_1654、Sandpath、campus_1、campus_2、MulitpleDeposit 以及 VisitorParking 场景。根据融合结果,可以发现每一种方法都可以保留红外图像和可见光图像的各自特点,融合图像既有纹理特征又有红外目标。然而,各类方法之间也存在着各自的优势和不足。

(1) 对比 TNO 数据集 Kaptein_1654、Sandpath 两个场景, DenseFuse、FusionDN、U2 Fusion、FusionGAN

以及 RFN-Nest 方法没有突出目标像素信息, FusionGAN、DDcGAN、GANMcC、SDDGAN 方法融合图像



图 8 定性融合结果。(a)、(b) 红外、可见光图像；(c)~(l) DenseFuse、FusionDN、U2 Fusion、FusionGAN、DDcGAN、GANMcC、RFN-Nest、STDFusionNet、SDDGAN 以及 SeAFusion 方法融合结果

Fig.8 Qualitative fusion results. (a), (b) Infrared and visible images; (c)-(l) Fusion methods of DenseFuse, FusionDN, U2 Fusion, FusionGAN, DDcGAN, GANMcC, RFN-Nest, STDFusionNet, SDDGAN and SeAFusion

纹理信息不够清晰,虽然 DDcGAN 方法目标和背景对比度明显,但和红外原图像相比,融合图像目标边缘存在伪影。

(2) 对比 OTCBVS 数据集的 campus_1、campus_2 场景, DenseFuse、FusionDN、U2 Fusion、RFN-Nest 方法不能突显目标,原因在于这些方法为了适应多源图像融合任务,训练集并没有单独使用红外和可见光图像,导致网络在红外和可见光图像融合过程中针对性不足。FusionGAN、GANMcC、DDcGAN 方法生成的融合图像模糊,目标存在伪影,这些问题来源于不确定的对抗关系。STDFusionNet、SeAFusion 方法既保留了红外显著目标,也保留了可见光图像清晰的背景信息,前者对比度高于后者。

(3) 对比 INO 数据集的 MultipleDeposit 场景,由于目标较小,除 SeAFusion 方法可以看到清晰的目标像素和结构信息外,其他算法均不能突出目标像素和结构信息。相对而言, DenseFuse、FusionDN 更能突出目标结构信息, U2 Fusion、STDFusionNet 更能突出目标像素信息。在 INO 数据集的 VisitorParking 场景下, FusionDN、STDFusionNet、SeAFusion 目标明亮且纹理清晰。

综上所述,4 种生成对抗网络融合方法都可以突出目标,但 FusionGAN、DDcGAN、GANMcC 方法融合结果不清晰, SSDGAN 方法融合图像除目标区域外,整幅图像亮度低。DenseFuse、FusionDN、U2 Fusion、STDFusionNet 以及 SeAFusion 方法可以保留清晰的纹理信息,在视觉上,STDFusionNet、SeAFusion

方法的对比度高、目标明亮, SeAFusion 方法拥有锐化的目标边缘。

3.2.2 定量比较

为了进一步对比不同融合方法的优缺点,文中选择 10 个评价指标对融合图像进行评价,其包括: EN、MI、SSIM、SD、AG、SF、PSNR、VIFF、CC 以及 $Q^{AB/F}$ 。

文中分别展示了 Kaptein_1654、Sandpath、campus_1、campus_2、MultipleDeposit、VisitorParking 6 个场景下融合图像的客观评价数据,如表 5~表 10 所示。在 6 个场景中, FusionDN、SeAFusion 方法对原图像信息的保存能力相对其他算法好,而且可以发现,互信息值比较大时,融合图像在信息熵上的表现也相对较好; DenseFuse 方法在 SSIM 指标上的表现最好,说明该方法在亮度、对比度以及结构 3 个方面的综合评价相对更好,但融合结果反映该算法对目标亮度关注度不够; FusionDN、SeAFusion 方法在 SD 指标上值相对其他算法较高,可见这两种方法得到的融合图像对比度更高,在视觉上算法保留的像素信息更好; FusionDN 方法在 AG、SF、VIFF 指标上的值相对于其他算法较高, SeAFusion、U2 Fusion、STDFusionNet 方法次之,说明上述方法纹理清晰,更容易被人类视觉系统接受,同时可以发现 AG 和 SF 指标在融合结果中的表现的确相似,说明多数情况下二者可以相互替换; FusionGAN 方法在 CC 指标上表现相对较好,但该指标在所有算法融合结果上表现的并不稳定,而且 FusionGAN 方法融合结果在其他指标上表现一般,因此该评判标准的适用性还需要进一步研究; STD-

表 5 不同方法在 Kaptein_1654 场景下的客观评价指标

Tab.5 Objective evaluation indicators of different methods in the Kaptein_1654 scene

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	6.42	12.83	0.72	29.74	3.62	6.93	16.39	0.34	0.53	0.36
FusionDN	7.19	14.37	0.64	46.48	6.67	12.97	14.56	0.55	0.52	0.42
U2 Fusion	6.58	13.16	0.70	28.68	4.61	8.62	16.17	0.35	0.53	0.40
FusionGAN	5.74	11.47	0.67	17.10	3.29	6.28	17.05	0.08	0.64	0.17
DDcGAN	6.96	13.93	0.59	37.17	6.29	11.63	15.15	0.32	0.52	0.38
GANMcC	6.06	12.11	0.69	25.36	2.13	4.44	15.38	0.21	0.56	0.14
RFN-Nest	6.54	13.09	0.68	31.47	2.39	4.99	15.69	0.32	0.52	0.28
STDFusionNet	6.70	13.41	0.65	52.90	5.29	11.22	15.17	0.40	0.51	0.54
SSDGAN	5.85	11.70	0.64	23.06	1.49	3.66	13.15	0.19	0.56	0.08
SeAFusion	6.71	13.43	0.67	41.07	5.86	11.25	13.92	0.41	0.56	0.49

表 6 不同方法在 Sandpath 场景下的客观评价指标

Tab.6 Objective evaluation indicators of different methods in the Sandpath scenario

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	6.68	13.35	0.69	29.06	6.61	10.89	19.86	0.54	0.68	0.36
FusionDN	7.42	14.84	0.55	45.13	10.95	18.35	15.75	0.82	0.68	0.29
U2 Fusion	6.34	12.67	0.69	20.46	6.42	10.49	19.19	0.36	0.69	0.33
FusionGAN	6.43	12.85	0.60	21.12	6.22	10.30	17.05	0.13	0.66	0.29
DDcGAN	7.25	14.50	0.49	37.85	10.23	16.91	15.03	0.44	0.69	0.37
GANMcC	6.33	12.65	0.69	21.12	3.40	5.66	19.44	0.22	0.71	0.19
RFN-Nest	6.89	13.79	0.64	32.01	4.88	8.13	19.24	0.50	0.68	0.42
STDFusionNet	6.82	13.64	0.59	35.09	6.83	11.64	18.59	0.22	0.60	0.56
SSDGAN	5.95	11.91	0.62	16.26	2.08	3.56	16.32	0.18	0.71	0.09
SeAFusion	6.82	13.64	0.65	33.02	7.43	12.26	17.81	0.35	0.66	0.42

表 7 不同方法在 campus_1 场景下的客观评价指标

Tab.7 Objective evaluation indicators of different methods in the campus_1 scenario

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	7.13	14.26	0.64	37.50	7.32	17.07	14.99	0.29	0.87	0.44
FusionDN	7.56	15.12	0.60	50.79	11.32	25.36	14.55	0.33	0.86	0.42
U2 Fusion	7.16	14.32	0.62	37.79	9.04	19.94	14.95	0.30	0.88	0.40
FusionGAN	6.15	12.29	0.55	18.65	5.56	13.03	12.98	0.08	1.08	0.14
DDcGAN	7.38	14.76	0.53	44.16	11.48	24.23	14.19	0.23	0.83	0.38
GANMcC	7.16	14.33	0.59	36.89	6.21	11.02	15.49	0.21	0.88	0.21
RFN-Nest	7.19	14.39	0.58	39.47	5.04	11.19	14.83	0.27	0.87	0.22
STDFusionNet	7.39	14.79	0.68	49.13	11.30	28.52	15.67	0.17	0.85	0.50
SSDGAN	6.70	13.40	0.52	30.07	3.46	8.11	13.56	0.18	0.95	0.12
SeAFusion	7.56	15.11	0.59	53.36	11.82	27.87	14.21	0.29	0.86	0.47

表 8 不同方法在 campus_2 场景下的客观评价指标

Tab.8 Objective evaluation indicators of different methods in the campus_2 scenario

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	7.41	14.81	0.62	50.76	8.76	20.71	15.13	0.50	0.93	0.44
FusionDN	7.44	14.88	0.60	51.99	10.33	23.62	14.85	0.47	0.91	0.46
U2 Fusion	7.32	14.64	0.60	52.49	10.57	23.99	15.05	0.52	0.92	0.50
FusionGAN	6.70	13.40	0.49	27.51	6.02	14.02	12.23	0.23	0.97	0.16
DDcGAN	7.36	14.72	0.52	46.74	10.12	23.09	14.16	0.28	0.90	0.37
GANMcC	7.42	14.84	0.55	48.77	6.21	13.88	15.53	0.45	0.93	0.26
RFN-Nest	7.42	14.84	0.55	49.85	6.07	13.71	14.80	0.45	0.94	0.26
STDFusionNet	7.34	14.68	0.58	54.27	11.33	28.35	13.99	0.32	0.88	0.49
SSDGAN	7.02	14.03	0.46	42.71	4.89	11.78	13.70	0.35	0.93	0.17
SeAFusion	7.71	15.42	0.60	66.00	13.79	32.00	13.50	0.48	0.56	0.92

表 9 不同方法在 MultipleDeposit 场景下的客观评价指标

Tab.9 Objective evaluation indicators of different methods in the MultipleDeposit scenario

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	7.66	15.31	0.78	71.02	6.65	15.86	16.64	0.72	0.96	0.55
FusionDN	7.46	14.91	0.75	53.44	7.22	16.87	16.61	0.59	0.95	0.52
U2 Fusion	7.23	14.64	0.77	52.49	10.57	23.99	15.05	0.52	0.97	0.50
FusionGAN	7.13	14.27	0.70	43.06	5.98	14.21	15.88	0.28	0.97	0.38
DDcGAN	7.29	14.58	0.67	47.60	6.85	15.47	15.74	0.36	0.95	0.43
GANMcC	7.71	15.42	0.75	73.35	4.44	10.38	15.27	0.57	0.96	0.36
RFN-Nest	7.70	15.39	0.75	72.45	4.96	11.86	16.14	0.67	0.99	0.47
STDFusionNet	7.50	15.01	0.72	72.25	8.86	23.66	19.84	0.69	0.92	0.62
SSDGAN	7.67	15.35	0.71	67.96	3.77	8.58	15.90	0.51	0.96	0.25
SeAFusion	7.79	15.59	0.74	76.36	8.80	21.46	17.97	0.80	0.95	0.62

表 10 不同方法在 VisitorParking 场景下的客观评价指标

Tab.10 Objective evaluation indicators of different methods in the VisitorParking scenario

Methods	EN	MI	SSIM	SD	AG	SF	PSNR	VIFF	CC	$Q^{AB/F}$
DenseFuse	6.77	13.54	0.74	34.64	4.41	10.97	18.18	0.51	0.67	0.42
FusionDN	7.50	15.00	0.63	52.26	7.92	19.62	14.32	0.70	0.63	0.40
U2 Fusion	6.53	13.05	0.74	31.37	4.55	11.30	19.06	0.41	0.66	0.40
FusionGAN	6.19	12.39	0.65	32.06	3.90	10.00	14.59	0.21	0.63	0.27
DDcGAN	7.26	14.52	0.61	42.80	7.02	16.99	14.84	0.55	0.66	0.39
GANMcC	6.76	13.53	0.72	39.09	2.76	6.60	18.88	0.37	0.65	0.22
RFN-Nest	7.18	14.36	0.68	44.54	3.57	9.41	15.22	0.61	0.66	0.40
STDFusionNet	6.28	12.55	0.68	26.09	4.83	14.43	20.30	0.19	0.61	0.49
SSDGAN	6.49	12.97	0.72	29.54	2.09	5.01	19.27	0.34	0.67	0.13
SeAFusion	6.79	13.58	0.70	35.86	5.96	14.84	16.46	0.47	0.66	0.47

FusionNet 方法在 PSNR 指标上相对于其他算法较好, 在 $Q^{AB/F}$ 指标上表现最好, 因此融合图像清晰、纹理边缘丰富。分析这 10 个评价指标, 可以从客观上分析图像融合的质量, 比如: SeAFusion、STDFusionNet 方法更能突出红外目标像素信息, FusionDN 方法更能突出整体信息。因此, 这 3 种方法在客观评价指标上优于其他算法。

另外, 为了对比算法实时性, 文中测试不同方法在 GPU 环境下融合 Sandpath 场景的红外和可见光图像运行时间, 如表 11 所示, 其中 STDFusionNet 方法运行时间最短, 说明该模型更加轻量化。

综上, 根据主观评价和客观评价结果可以判别,

表 11 不同方法的运行时间

Tab.11 Running time of different methods

Methods	Run time/s
DenseFuse	3.7234
FusionDN	2.5158
U2 Fusion	1.0212
FusionGAN	0.5221
DDcGAN	2.4545
GANMcC	1.0142
RFN-Nest	1.1682
STDFusionNet	0.0480
SDDGAN	0.1970
SeAFusion	0.1605

基于深度学习的生成对抗网络融合方法结果还不够理想,原因在于理论上的对抗关系在实际训练过程中并没有使模型绝对收敛,因此,该方法还需要进一步完善。基于卷积神经网络的融合方法在逐步向深层网络发展,并为自编码器及生成对抗融合网络奠定了基础。

4 展 望

伴随着信息数字化技术的发展,图像融合方法已经取得了一定的成绩,尤其在红外和可见光图像融合领域,从传统的空域法和频域法进阶到深度学习方法,取得显著成果。但是,基于深度学习的红外和可见光图像融合方法也存在很多难点问题,例如,可见光图像细节信息丢失严重、红外目标不突出、评价体系不完善、算法实时性等问题。因此,在未来红外与可见光图像融合技术需要进一步研究改进,下面笔者提出一些需要解决的问题及改进的方向:

(1) 当前深度网络模型的良好性能已在融合任务中得到发挥,但是整体上网络模型依然沿用一些常用模块,比如:简单级联、注意力机制模块等。因此,需要结合前沿技术更新网络模型,比如:语义分割、Transform 模型等;

(2) 基于深度学习的红外和可见光图像网络融合模型通常并不对红外通道和可见光通道的结构进行区分,仅依靠损失函数来约束各通道权重参数,这种方式可能导致网络提取到的信息不具针对性。因此,需要提出一种针对红外与可见光图像融合的网络模型,以分别保留原图像特有信息;

(3) 在协调网络损失函数比例系数时,现存方法大多通过人为设定,而设定值的大小将直接影响融合结果。因此,研究者需要考虑各损失函数在融合过程中的重要性,更合理地得到适应的比例系数;

(4) 大多数融合网络采用一次性训练方式,而损失函数下降到一定阶段会趋于稳定,模型达到拟合状态,理论上的训练完成并不能有好的测试结果。因此,研究者需考虑除损失函数外,是否可以在模型、训练方式上进行创新,增加损失函数训练过程中的下降幅度,从而得到高质量的测试图像;

(5) 图像融合领域没有标准的融合评价指标。比如:融合结果在几个评价指标上的结果显眼,但在个

别指标上比较差。因此,需要提出一个相对完备的评价标准模式;

(6) 基于深度学习的红外和可见光图像融合方法主要依赖像素级融合,该种方式可以提高图像质量且适合人类视觉系统。但是该融合方式忽视了融合后的应用问题,特征级融合以及决策级融合方式可能更有利于落地应用。

(7) 融合模型需要尽可能地轻量化,在保存信息的同时需权衡实时性、工程应用便捷性等。

5 结 论

近几年,随着深度学习技术热度一直攀升,相应领域也得到了飞速发展。在红外和可见光图像融合领域,伴随着科研人员的不懈努力,深度学习融合技术达到一个新的高度。该篇文章回顾了基于深度学习的红外与可见光图像融合技术的发展历程以及研究成果,并对涉及到的融合方法进行凝练总结。这些方法大致分为 3 类:基于卷积神经网络融合方法、基于自编码器融合方法以及基于生成对抗网络融合方法。文中简述了融合图像的主观评价方法,并搜集了该领域的典型客观评价指标,对近几年一些经典融合算法的融合结果进行对比分析。之后,文中基于现存方法存在的普遍问题,提供了一些改进的方向。基于上述总结和分析,笔者更加肯定深度学习方法在该领域的应用,其方法也逐渐趋于成熟。最后,红外和可见光图像融合技术将被应用于其他已知和未知领域,促进人类科技进步和发展。

参考文献:

- [1] Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey [J]. *Information Fusion*, 2019, 45: 153-178.
- [2] Ma J, Chen C, Li C. Infrared and visible image fusion via gradient transfer and total variation minimization [J]. *Information Fusion*, 2016, 31: 100-109.
- [3] Shen Ying, Huang Chunhong, Huang Feng, et al. Research progress of infrared and visible image fusion technology [J]. *Journal of Infrared and Millimeter Waves*, 2021, 50(9): 20200467. (in Chinese)
- [4] Ji X, Zhang G. Image fusion method of SAR and infrared image based on curvelet transform with adaptive weighting [J].

- Multimedia Tools and Applications*, 2017, 76(17): 17633-17649.
- [5] Li H, Zhou Y T, Chellappa R. SAR/IR sensor image fusion and real-time implementation[C]//Conference Record of The Twenty-Ninth Asilomar Conference on Signals, Systems and Computers. IEEE, 1995, 2: 1121-1125.
- [6] Ye Y, Zhao B, Tang L. SAR and visible image fusion based on local non-negative matrix factorization[C]//2009 9th International Conference on Electronic Measurement & Instruments. IEEE, 2009: 4263-4266.
- [7] Ali M A, Clausi D A. Automatic registration of SAR and visible band remote sensing images[C]//IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2002, 3: 1331-1333.
- [8] Parmar K, Kher R K, Thakkar F N. Analysis of CT and MRI image fusion using wavelet transform[C]//2012 International Conference on Communication Systems and Network Technologies. IEEE, 2012: 124-127.
- [9] Liu X, Mei W, Du H. Structure tensor and nonsubsampling shearlet transform based algorithm for CT and MRI image fusion [J]. *Neurocomputing*, 2017, 235: 131-139.
- [10] Ma J, Yu W, Liang P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. *Information Fusion*, 2019, 48: 11-26.
- [11] Bai L, Zhang W, Pan X, et al. Underwater image enhancement based on global and local equalization of histogram and dual-image multi-scale fusion [J]. *IEEE Access*, 2020, 8: 128973-128990.
- [12] Rashid M, Khan M A, Alhaisoni M, et al. A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection [J]. *Sustainability*, 2020, 12(12): 5037.
- [13] Tang Cong, Ling Yongshun, Yang Hua, et al. Decision-level fusion detection for infrared and visible spectra based on deep learning [J]. *Journal of Infrared and Millimeter Waves*, 2019, 48(6): 0626001. (in Chinese)
- [14] Shen Y. RGBT bimodal twin tracking network based on feature fusion [J]. *Journal of Infrared and Millimeter Waves*, 2021, 50(3): 20200459. (in Chinese)
- [15] Adamchuk V I, Rossel R V, Sudduth K A, et al. Sensor fusion for precision agriculture [M]//Thomas C. Sensor Fusion-Foundation and Applications. Rijeka, Croatia: InTech, 2011: 27-40.
- [16] Wang Z, Li G, Jiang X. Flood disaster area detection method based on optical and SAR remote sensing image fusion [J]. *Journal of Radar*, 2020, 9(3): 539-553. (in Chinese)
- [17] Yang Xie, Tong Tao, Lu Songyan, et al. Fusion of infrared and visible images based on multi-features [J]. *Optical Precision Engineering*, 2014, 22(2): 489-496. (in Chinese)
- [18] Chen J, Li X, Luo L, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition [J]. *Information Sciences*, 2020, 508: 64-78.
- [19] Liu Y, Jin J, Wang Y, et al. Region level based multi-focus image fusion using quaternion wavelet and normalized cut [J]. *Signal Process*, 2014, 97: 9-30.
- [20] Chen Hao, Wang Yanjie. Research on image fusion algorithm based on laplace pyramid transform [J]. *Laser & Infrared*, 2009, 39(4): 439-442. (in Chinese)
- [21] Choi M, Kim R Y, Nam M R, et al. Fusion of multispectral and panchromatic satellite images using the curvelet transform [J]. *IEEE Geoscience and Remote Sensing Letters*, 2005, 2(2): 136-140.
- [22] Yang B, Li S. Multifocus image fusion and restoration with sparse representation [J]. *IEEE Transactions on Instrumentation and Measurement*, 2009, 59(4): 884-892.
- [23] Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation [J]. *Information Fusion*, 2015, 24: 147-164.
- [24] Liu Xianhong, Chen Zhibin, Qin Mengze. Fusion of infrared and visible light images combined with guided filtering and convolutional sparse representation [J]. *Optical Precision Engineering*, 2018, 26(5): 1242-1253. (in Chinese)
- [25] Du X, El-Khamy M, Lee J, et al. Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection[C]//2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017: 953-961.
- [26] Dai Jindun, Liu Yadong, Mao Xianyin, et al. Infrared and visible image fusion based on FDST and dual-channel PCNN [J]. *Infrared and Laser Engineering*, 2019, 48(2): 0204001.
- [27] Fu Z, Wang X, Xu J, et al. Infrared and visible images fusion based on RPCA and NSCT [J]. *Infrared Physics & Technology*, 2016, 77: 114-123.
- [28] Mitianoudis N, Stathaki T. Pixel-based and region-based image fusion schemes using ICA bases [J]. *Information Fusion*, 2007, 8(2): 131-142.
- [29] Kong W, Lei Y, Zhao H. Adaptive fusion method of visible light and infrared images based on non-subsampling shearlet transform and fast non-negative matrix factorization [J]. *Infrared Physics & Technology*, 2014, 67: 161-172.
- [30] Wang A, Wang M. RGB-D salient object detection via

- minimum barrier distance transform and saliency fusion [J]. *IEEE Signal Processing Letters*, 2017, 24(5): 663-667.
- [31] Wang Xin, Ji Tongbo, Liu Fu. Fusion of infrared and visible light images combined with object extraction and compressed Sensing [J]. *Optical Precision Engineering*, 2016, 24(7): 1743-1753. (in Chinese)
- [32] Cui Xiaorong, Shen Tao, Huang Jianlu, et al. Infrared and visible image fusion based on BEMD and improved visual saliency [J]. *Infrared Technology*, 2020, 42(11): 1061. (in Chinese)
- [33] Lewis J J, O'Callghan R J, Nikolov S G, et al. Pixel-and region-based image fusion with complex wavelets [J]. *Information Fusion*, 2007, 8(2): 119-130.
- [34] Rajkumar S, Mouli P C. Infrared and visible image fusion using entropy and neuro-fuzzy concepts[C]//ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I. Cham: Springer, 2014: 93-100.
- [35] Zhao J, Cui G, Gong X, et al. Fusion of visible and infrared images using global entropy and gradient constrained regularization [J]. *Infrared Physics & Technology*, 2017, 81: 201-209.
- [36] Sun C, Zhang C, Xiong N. Infrared and visible image fusion techniques based on deep learning: A review [J]. *Electronics*, 2020, 9(12): 2162.
- [37] Ma J, Chen C, Li C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization [J]. *Inf Fusion*, 2016, 31: 100-109.
- [38] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [39] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA. 2012: 1097-1105.
- [40] 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.
- [41] 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). IEEE, 2018: 2705-2710.
- [42] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2033-02-23]. <https://arxiv.org/abs/1409.1556>.
- [43] Liu Y, Chen X, Cheng J, et al. Infrared and visible image fusion with convolutional neural networks [J]. *International Journal of Wavelets, Multiresolution and Information Processing*, 2018, 16(3): 1850018.
- [44] Li H, Wu X J, Durrani T S. Infrared and visible image fusion with ResNet and zero-phase component analysis [J]. *Infrared Physics & Technology*, 2019, 102: 103039.
- [45] Cui Y, Du H, Mei W. Infrared and visible image fusion using detail enhanced channel attention network [J]. *IEEE Access*, 2019, 7: 182185-182197.
- [46] 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.
- [47] Hou R, Zhou D, Nie R, et al. VIF-Net: An unsupervised framework for infrared and visible image fusion [J]. *IEEE Transactions on Computational Imaging*, 2020, 6: 640-651.
- [48] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [49] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
- [50] Li L, Xia Z, Han H, et al. Infrared and visible image fusion using a shallow CNN and structural similarity constraint [J]. *IET Image Processing*, 2020, 14(14): 3562-3571.
- [51] Li Y, Wang J, Miao Z, et al. Unsupervised densely attention network for infrared and visible image fusion [J]. *Multimedia Tools and Applications*, 2020, 79(45): 34685-34696.
- [52] Long Y, Jia H, Zhong Y, et al. RXDNFuse: A aggregated residual dense network for infrared and visible image fusion [J]. *Information Fusion*, 2021, 69: 128-141.
- [53] Zhu J, Dou Q, Jian L, et al. Multiscale channel attention network for infrared and visible image fusion [J]. *Concurrency and Computation: Practice and Experience*, 2021, 33(22): e6155.
- [54] Xu H, Ma J, Jiang J, et al. U2 fusion: A unified unsupervised image fusion network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(1): 502-518.
- [55] Xu H, Ma J, Le Z, et al. FusionDn: A unified densely connected network for image fusion[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(7): 12484-12491.
- [56] Prabhakar K R, Srikar V S, Babu R V. Deepfuse: A deep unsupervised approach for exposure fusion with extreme

- exposure image pairs[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 4714-4722.
- [57] Li H, Wu X J. DenseFuse: A fusion approach to infrared and visible images [J]. *IEEE Transactions on Image Processing*, 2018, 28(5): 2614-2623.
- [58] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755.
- [59] Zhao Z, Xu S, Zhang C, et al. DIDFuse: Deep image decomposition for infrared and visible image fusion [EB/OL]. (2020-03-20)[2022-02-23]. <https://arxiv.org/abs/200309210>.
- [60] Pan Y, Pi D, Khan I A, et al. DenseNetFuse: A study of deep unsupervised DenseNet to infrared and visual image fusion [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2021, 12(11): 10339-10351.
- [61] Wang H, An W, Li L, et al. Infrared and visible image fusion based on multi - channel convolutional neural network [J]. *IET Image Processing*, 2022, 16(6): 1575-1584.
- [62] Liu L, Chen M, Xu M, et al. Two-stream network for infrared and visible images fusion [J]. *Neurocomputing*, 2021, 460: 50-58.
- [63] Li H, Wu X J, Durrani T. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(12): 9645-9656.
- [64] Li H, Wu X J, Kittler J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images [J]. *Information Fusion*, 2021, 73: 72-86.
- [65] Wang Z, Wang J, Wu Y, et al. UNFusion: A unified multi-scale densely connected network for infrared and visible image fusion [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(6): 3360-3374.
- [66] Fu Y, Wu X J. A dual-branch network for infrared and visible image fusion[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 10675-10680.
- [67] Zhang H, Xu H, Xiao Y, et al. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(7): 12797-12804.
- [68] Jian L, Yang X, Liu Z, et al. SEDRFuse: A symmetric encoder-decoder with residual block network for infrared and visible image fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 1-15.
- [69] Zhao F, Zhao W, Yao L, et al. Self-supervised feature adaption for infrared and visible image fusion [J]. *Information Fusion*, 2021, 76: 189-203.
- [70] Ma J, Tang L, Xu M, et al. STDFusionNet: An infrared and visible image fusion network based on salient target detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-13.
- [71] Raza A, Liu J, Liu Y, et al. IR-MSDNet: Infrared and visible image fusion based on infrared features and multiscale dense network [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 3426-3437.
- [72] Tang L, Yuan J, Ma J. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network [J]. *Information Fusion*, 2022, 82: 28-42.
- [73] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014: 2672-2680.
- [74] Wei L, Zhang S, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.
- [75] Li J, Liang X, Wei Y, et al. Perceptual generative adversarial networks for small object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1222-1230.
- [76] Rabbi J, Ray N, Schubert M, et al. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network [J]. *Remote Sensing*, 2020, 12(9): 1432.
- [77] Fu Y, Wu X J, Durrani T. Image fusion based on generative adversarial network consistent with perception [J]. *Information Fusion*, 2021, 72: 110-125.
- [78] Yang Y, Liu J, Huang S, et al. Infrared and visible image fusion via texture conditional generative adversarial network [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(12): 4771-4783.
- [79] Ma J, Xu H, Jiang J, et al. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4980-4995.
- [80] Li Q, Lu L, Li Z, et al. Coupled GAN with relativistic discriminators for infrared and visible images fusion [J]. *IEEE Sensors Journal*, 2019, 21(6): 7458-7467.

- [81] Li S, Kang X, Hu J. Image fusion with guided filtering [J]. *IEEE Transactions on Image Processing*, 2013, 22(7): 2864-2875.
- [82] Ma J, Liang P, Yu W, et al. Infrared and visible image fusion via detail preserving adversarial learning [J]. *Information Fusion*, 2020, 54: 85-98.
- [83] Xu J, Shi X, Qin S, et al. LBP-BEGAN: A generative adversarial network architecture for infrared and visible image fusion [J]. *Infrared Physics & Technology*, 2020, 104: 103144.
- [84] Li J, Huo H, Liu K, et al. Infrared and visible image fusion using dual discriminators generative adversarial networks with Wasserstein distance [J]. *Information Sciences*, 2020, 529: 28-41.
- [85] Li J, Huo H, Li C, et al. AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks [J]. *IEEE Transactions on Multimedia*, 2020, 23: 1383-1396.
- [86] Yang X, Huo H, Li J, et al. DSG-fusion: Infrared and visible image fusion via generative adversarial networks and guided filter [J]. *Expert Systems with Applications*, 2022, 200: 116905.
- [87] Li J, Huo H, Li C, et al. Multigrained attention network for infrared and visible image fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 1-12.
- [88] Ma J, Zhang H, Shao Z, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 1-14.
- [89] Hou J, Zhang D, Wu W, et al. A generative adversarial network for infrared and visible image fusion based on semantic segmentation [J]. *Entropy*, 2021, 23(3): 376.
- [90] Zhou H, Wu W, Zhang Y, et al. Semantic-supervised infrared and visible image fusion via a dual-discriminator generative adversarial network[J/OL]. *IEEE Transactions on Multimedia* (Early Access), (2021-11-22)[2022-02-23]. <https://ieeexplore.ieee.org/document/9623476>.
- [91] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [92] Roberts J W, Van Aardt J A, Ahmed F B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification [J]. *Journal of Applied Remote Sensing*, 2008, 2(1): 023522.
- [93] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2: 1398-1402.
- [94] Rao Y J. In-fibre Bragg grating sensors [J]. *Measurement Science and Technology*, 1997, 8(4): 355.
- [95] Qu G, Zhang D, Yan P. Information measure for performance of image fusion [J]. *Electronics Letters*, 2002, 38(7): 313-315.
- [96] Eskicioglu A M, Fisher P S. Image quality measures and their performance [J]. *IEEE Transactions on Communications*, 1995, 43(12): 2959-2965.
- [97] Guo W, Xiong N, Chao H C, et al. Design and analysis of self-adapted task scheduling strategies in wireless sensor networks [J]. *Sensors*, 2011, 11(7): 6533-6554.
- [98] Cui G, Feng H, Xu Z, et al. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition [J]. *Optics Communications*, 2015, 341: 199-209.
- [99] Han Y, Cai Y, Cao Y, et al. A new image fusion performance metric based on visual information fidelity [J]. *Information Fusion*, 2013, 14(2): 127-135.
- [100] Xydeas C S, Petrovic V. Objective image fusion performance measure [J]. *Electronics Letters*, 2000, 36(4): 308-309.
- [101] Deshmukh M, Bhosale U. Image fusion and image quality assessment of fused images [J]. *International Journal of Image Processing (IJIP)*, 2010, 4(5): 484.
- [102] 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.
- [103] Haghighat M B A, Aghagolzadeh A, Seyedarabi H. A non-reference image fusion metric based on mutual information of image features [J]. *Computers & Electrical Engineering*, 2011, 37(5): 744-756.
- [104] Toet A. The TNO multiband image data collection [J]. *Data in Brief*, 2017, 15: 249-251.
- [105] Davis J W, Sharma V. OTCBVS benchmark dataset collection [EB/OL]. (2007)[2022-02-23]. <http://www.cse.ohio-state.edu/otcbvs-bench>.