DOI: 10.13382/j.jemi.B2306462

用于红外与可见光图像融合的注意力残差密集融合网络*

陈广秋 温奇璋 尹文卿 段 锦 黄丹丹
(长春理工大学电子信息工程学院 长春 130022)

摘 要:为了解决当前红外与可见光图像融合算法中易出现场景信息缺失、目标区域细节模糊、融合图像不自然等问题,提出一种用于红外与可见光图像融合的注意力残差密集融合网络(ARDFusion)。本文整体架构是一种自编码器网络,首先,利用存在最大池化层的编码器对源图像进行多尺度特征提取,然后,利用注意力残差密集融合网络分别对多个尺度的特征图进行融合,网络中的残差密集块可以连续存储特征并且最大程度地保留各层特征信息,注意力机制可以突出目标信息并获取更多与目标、场景有关的细节信息。最后,将融合后的特征输入到解码器中,通过上采样和卷积层对特征进行重构,得到融合图像。本文提出了一种用于红外与可见光图像融合的注意力残差密集融合网络,实验结果表明,较已有文献的其他典型融合算法,具有较好的融合效果,能够更好地保留可见光图像中的光谱特性且红外目标显著,并在主观评价和客观评价方面都取得了较好的融合性能。

关键词:红外与可见光图像融合;自编码器网络;残差密集连接;注意力机制;光谱特性
中图分类号:TP391.4 **文献标识码**:A **国家标准学科分类代码**:520.20

Attentional residual dense connection fusion network for infrared and visible image fusion

Chen Guangqiu Wen Qizhang Yin Wenqing Duan Jin Huang Dandan
(School of Electronics and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: In order to solve the problems in the current infrared and visible image fusion algorithm, such as missing scene detail information, unclear target region detail information and unnatural fusion image, an attentional residual dense fusion network (ARDFusion) for infrared and visible image fusion is proposed. The whole architecture of this paper is an auto-encoder network. First, the encoder with the largest pooling layer is used to extract multi-scale features of the source image, then the attention residual dense fusion network is used to fuse the feature maps of multiple scales. The residual dense blocks in the network can continuously store features and maximize the retention of feature information at each layer. The attention mechanism can highlight target information and obtain more detailed information related to the target and scene. Finally, the fused features are input into the decoder and reconstructed through upsampling and convolutional layers to obtain the fused image. This article proposes an attention residual dense fusion network for infrared and visible image fusion. The experimental results show that compared to other typical fusion algorithms in existing literature, it has better fusion performance, can better preserve the spectral characteristics in visible images, and has significant infrared targets. It has achieved good fusion performance in both subjective and objective evaluations.

Keywords: infrared and visible image fusion; auto-encoder network; residual dense connection; attention mechanism; spectral characteristic

收稿日期:2023-04-23 Received Date: 2023-04-23
*基金项目:国家自然科学基金重大仪器专项(62127813)、吉林省科技发展计划项目(20210203181SF)资助

0　引　言

红外图像具有丰富的热辐射目标信息,成像机制不会受到天气因素影响,而且目标相比于背景具有更高的对比度和辨识度,但其图像分辨率低,背景细节不突出,且目标成像极易受到温度变化影响。相比于红外图像,可见光图像具有丰富的背景和纹理信息,具有较高的图像分辨率,不易受到温度变化影响,但其成像易受到天气因素影响,在受到干扰的情况下,目标与背景的可辨识程度差,无法在场景内突出目标信息。因此具有一定空间和信息互补性的红外与可见光图像融合技术应用广泛[1],对于一些难以辨析的空间场景具有很重要的应用价值。

随着图像融合技术的发展,已经出现了许多不同的图像融合方法,根据其使用的理论知识主要分为,基于多尺度变换的方法[2]、基于稀疏表示的方法[3]、基于显著性的方法[4]、基于神经网络的方法[5]和其他融合方法[6],其中基于多尺度变换的融合方法在图像融合领域中使用较为活跃。这些融合方法经常使用金字塔变换[7]、小波变换[8]和非下采样轮廓波变换[9]等变换方法对图像进行分解,并按照特定的融合规则对各个层进行融合,最后重建目标图像。

近些年来,基于深度学习的图像融合方法发展迅速,因其具有很强的特征提取能力而得到了广泛的应用,并且各专家学者也加深了对深度学习图像融合的研究,其方法主要可分为 3 种:卷积神经网络[10]、自编码器网络[11]和生成对抗网络[12]。2017 年,Liu 等[13]将卷积神经网络引入到图像融合领域,利用模糊的背景图像与清晰的前景图像训练网络,通过二值化和两个一致性策略得到决策映射,并将其作为权重重建融合图像,虽然相比于传统算法有着明显的改进,但该方法中间丢失大量细节信息,导致网络无法进一步加深。为解决这些问题,2018 年,Li 等[14]提出了一种基于 VGG 网络融合方法,将传统方法与深度方法相结合,使用预训练网络提取图像特征,虽然强化了网络提取特征的能力,但其融合方法稍显简单,随着网络的逐步加深却无法充分利用深层信息,导致网络退化,效果变差。2019 年,Li 等[15]提出一种基于密集块的自编码器网络(DenseFuse),实现编码器提取特征、解码器重构图像的目的,并使用密集连接的方式充分利用中间层的特征信息,但由于其使用人工调制的融合规则进行融合,使得融合效果不佳。同年,Ma 等[16]将 GAN 网络与红外与可见光图像融合相结合,提出一种基于生成对抗网络的融合模型(FusionGAN),该网络利用了生成对抗网络的对抗性,使得生成器生成的融合图像尽可能多的包含真实图像,也就是源图像的分布特性,但由于超参数的调试过于复杂,使得源图像细节丢失严重,融合图像模糊。2020 年 Li 等[17]提出了一种基于巢连接的图像融合(NestFuse)方法,这是一种自编码器网络,使用编码器通过最大池化的方式将源图像信息下采样为多尺度深度特征,随后利用融合规则将特征融合,最后输入到解码器中重构图像,得到具有源图像多尺度特征信息的融合图像,但融合规则过于简单,无法为解码器提供充足的特征信息,导致重构后的图像部分背景纹理信息模糊,图像整体表现不自然。2021 年,Li 等[18]提出了一种端到端的残差融合网络(RFN-Nest),将简单的融合规则换为以残差方式连接的融合网络,虽然为解码器提供了比简单融合规则更多的特征信息,但在融合过程中损失了一定的精度,使得最后的融合图像失真严重。2022 年,Tang 等[19]基于照明感知来确定损失函数权重的方法(PIAFusion),网络分为特征提取,图像还原和光照感知网络 3 部分,主要利用光照感知网络判断可视图像是白天还是夜晚,然后得到白天概率和夜晚概率,并以这个概率作为损失函数的权重,不过,虽然考虑了整体的光照条件,但模型过于简单,无法在复杂环境中调整光照。

为解决红外与可见光图像融合算法中背景纹理信息缺失、红外目标细节不突出等问题,本文设计一种注意力残差密集融合网络的红外与可见光图像融合算法,利用残差密集连接保证了各层信息的充分利用及多尺度特征的传递性,使得每层的信息都可以被下一层所充分利用,从而达到更好的融合效果。此外,由于红外图像中具有大量的目标信息和一部分场景细节信息,可见光图像中具有大量的场景细节信息和一部分目标细节信息,所以设计了目标注意力模块和场景注意力模块对目标信息、场景细节信息和目标细节信息进行增强,在保证红外图像目标清晰的同时,完整地保留了可见光图像中关于目标区域的细节信息和来自背景的场景细节信息,使得融合性能进一步提升,融合图像更加自然,更符合人的视觉感知。

1　ARDFusion 算法模型

1.1　ARDFusion 网络架构

本文所提出的 ARDFusion 是一种自编码器网络,网络结构由两个编码器、注意力残差密集融合网络(ARDFNet)和解码器构成,整体网络架构如图 1 所示。

首先,红外与可见光图像分别输入到各自的编码器中,源图像被分解成具有不同尺度的特征图,随后注意力残差密集融合网络对输入的多尺度特征进行融合,解码器对融合后的各层特征进行重建,最后得到融合图像。整体训练过程分为自编码器网络训练和融合网络训练两部分,后文将更为详细地介绍这两部分的实验细节。

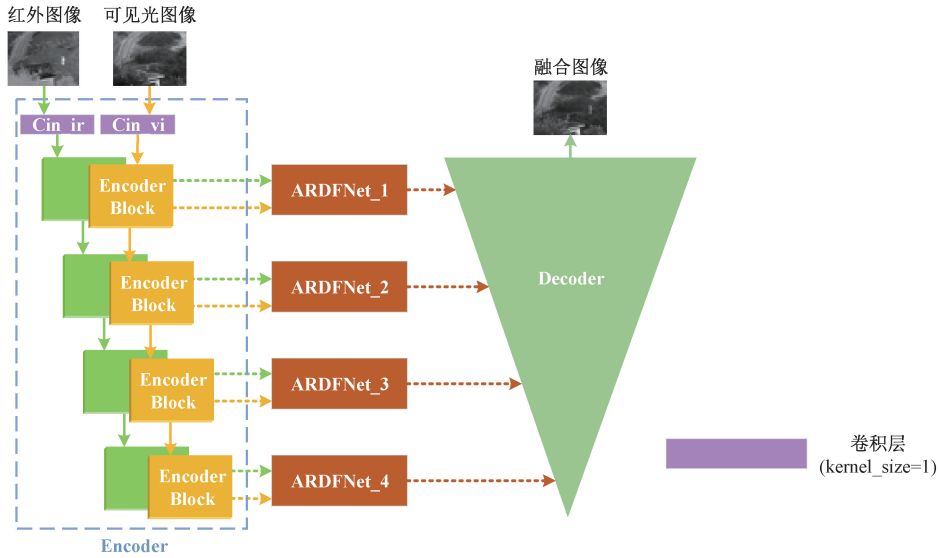


图 1　ARDFusion 网络架构

Fig. 1　ARDFusion network architecture

1.2　自编码器网络

1)编码器网络架构

编码器由 1 层 1×1 的卷积层和 4 个编码块(encoder block,EB)组成,如图 2 所示。其中前 3 个编码块中包含两层 3×3 的卷积层和 1 层最大池化层,最后一个编码块则不设最大池化层。首先经过具有池化层编码器的图像可以被分解为包含源图像信息的多尺度特征,浅层特征保留更多细节信息,深层特征则保留了更多的语义信息,随后原图像的多尺度特征将输入到融合网络 ARDFNet 中进行特征融合。

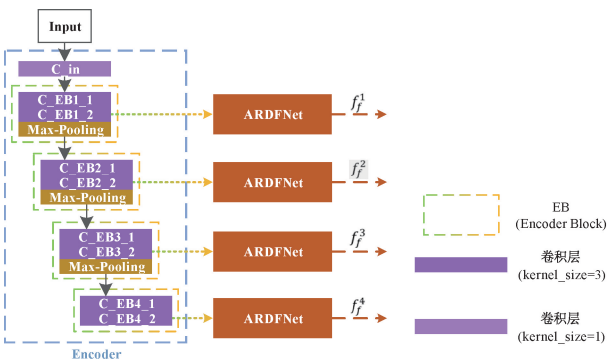


图 2　编码器网络架构

Fig. 2　Encoder network architecture

2)解码器网络架构

解码器由 6 个解码块(decoder block,DB)和 1 个步长为 1 的 1×1 卷积层构成,网络架构如图 3 所示。首先融合特征 f_f^k 输入到前 3 个解码块,而后为了可以使其方便快捷的进行图像重建,确保经融合网络输出的融合特征被充分利用,每一行 DB 都以密集连接形式连接,跨行 DB 则用上采样操作进行连接,最后经过 1 层步长为 1 的 1×1 卷积层得到最后的融合图像。

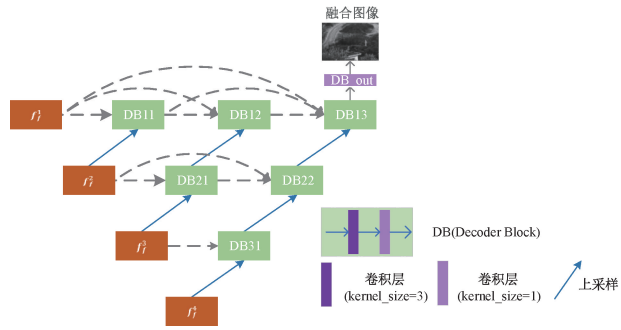


图 3　解码器网络架构

Fig. 3　Decoder network architecture

1.3　融合网络

ARDFNet(注意力残差密集融合网络)整体架构如图 4 所示,由目标注意力模块、场景特征注意力模块、残差密集块和 1 个步长为 1 的 1×1 降维卷积层组成。

残差密集块由 3 个步长为 1 的 1×1 卷积层和 4 个步长为 1 的 3×3 卷积层构成。并且为了保证各层信息的充分利用及多尺度特征的传递性,使用了残差密集连接方式,使得每层的信息都可以被下一层所充分利用。在目标注意力模块与场景注意力模块中,面对融合后的特征可能对于红外图像特征的目标信息和可见光图像特征的背景纹理信息不敏感,本文设计了一种增强红外特征和可见光特征的注意力模块,架构如图 5 和 6 所示。

这两种注意力模块都由通道注意力机制和像素注意

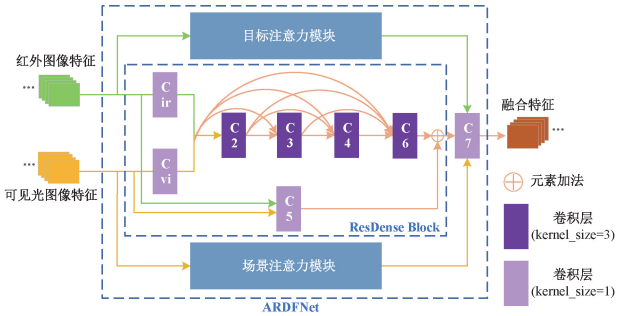


图 4　ARDFNet 架构

Fig. 4　ARDFNet architecture

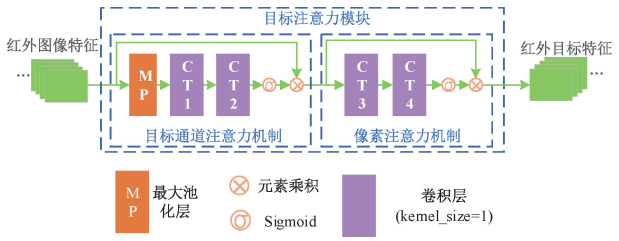


图 5　目标注意力模块

Fig. 5　Target attention module

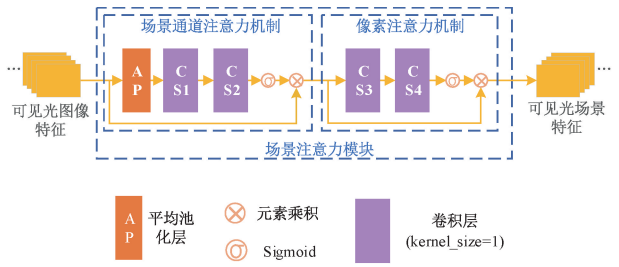


图 6　场景注意力模块

Fig. 6　Scene attention module

力机制组成,虽然两者的像素注意力机制则都为两个步长为 1 的 1×1 卷积层和 Sigmoid 组成,但有所不同的是,目标注意力模块中的目标通道注意力机制是由全局最大池化层、2 个步长为 1 的 1×1 卷积层和 Sigmoid 组成,而场景注意力模块中的通场景道注意力机制是由全局平均池化层、2 个步长为 1 的 1×1 卷积层和 Sigmoid 层组成。在目标注意力模块中,红外特征图输入到目标注意力模块的目标通道注意力机制中,经过最大池化操作将特征图形状从 $C \times H \times W$ 变为 $C \times 1 \times 1$,获得不同通道的权重,特征通过两个卷积层和 Sigmoid 进行计算,最后按元素将输入的特征图与通道权重相乘,得到各通道权重不同的特征图。由于红外图像侧重表达目标信息,所以本文使用像素注意力机制,使红外目标特征在像素层面可以得到更多的关注。

与通道注意力机制相似,目标通道注意力机制和场景通道注意力机制输出形状为 $C \times H \times W$ 的特征图,经过卷积层和 Sigmoid 的计算,改变形状为 $1 \times H \times W$ 的像素权重图,而后对两种通道权重不同的特征图和像素权重图使用元素乘法,得到最终的红外特征图。同理,在场景注意力模块中,可见光特征侧重表达场景信息,所以本文使用平均池化操作替换最大池化操作,输出的通道权重特征图通过相同结构像素注意力机制,使像素层面更注重场景信息的表达,得到最终的可见光特征图。

2　实验结果及分析

在本章中,对所提出的融合方法进行了验证。在详细介绍了训练阶段和测试阶段的实验设置之后,本文提出了一项消融研究,以研究不同元素对提出的融合网络的影响。最后,将本文的融合框架与其他现有算法进行定性比较。

在网络模型训练过程中,本文使用 PyTorch1. 10. 2 作为编程环境,在 NVIDIA RTX3090 GPU 和 Intel(R) Xeon(R) Platinum 8255C CPU @ 2. 50 GHz 上进行实验。

2. 1　自编码器网络训练

在自编码器网络训练过程中,只使用一个编码器参与整体训练首先将训练图像输入到编码器,经过编码器提取多尺度特征后,直接输入到解码器中进行特征重建,最后得到重建图像。其训练架构如图 7 所示。

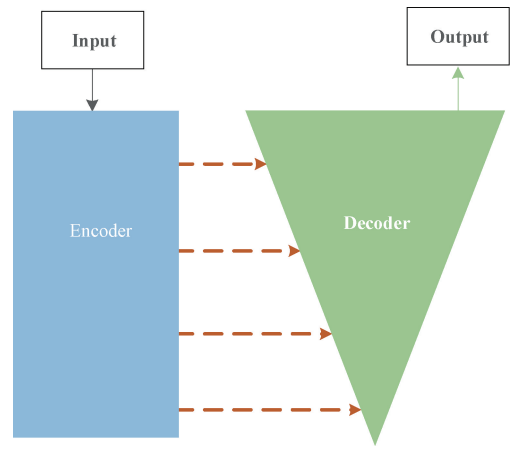


图 7　自编码器网络训练架构

Fig. 7　Auto-encoder network training architecture

本文使用数据集 MS-COCO[20] 作为训练集,训练图像数量为 80 000 张,先将训练图像转换为灰度图像,而后图像的尺寸重塑为 256×256,batch_size 设置为 4,epoch 设置为 2,学习率设置为 1×10^{-4},使用 Adam 作为优化器。

在自编码器网络训练过程中,本文设计了一种自编码器网络损失函数 L_{auto} 如式(1)所示:

$$L_{auto} = \alpha L_{structure} + L_{pixel} \tag{1}$$

其中，$L_{structure}$ 和 L_{pixel} 表示输入图像和输出图像之间的结构损失和像素损失，α 为两个损失函数之间的权重参数。

结构损失（$L_{structure}$）的计算公式如式(2)所示：

$$L_{structure} = 1 - SSIM(I_{output}, I_{input}) \tag{2}$$

其中，I_{output} 代表输出图像，I_{input} 代表输入图像。$SSIM(\cdot)$ 是结构相似性度量，通常用来图像质量评价，并且量化了输入图像和输出图像的结构相似性。$L_{structure}$ 在结构上约束输出图像和输入图像。

像素损失（L_{pixel}）的计算公式如式(3)所示：

$$L_{pixel} = \frac{1}{NM}\sum_{x\in N, y\in M}(I_{output}(x,y) - I_{input}(x,y))^2 \tag{3}$$

其中，I_{output} 代表输出图像，I_{input} 代表输入图像，M 和 N 为图像的高和宽，(x,y) 为像素位置。L_{pixel} 在像素级别上约束输出图像类似于输入图像。

2.2　融合网络训练

在融合模块训练过程中，固定自编码器网络及其参数，从而在自编码器架构中训练融合网络 ARDFNet。训练架构如图 8 所示。

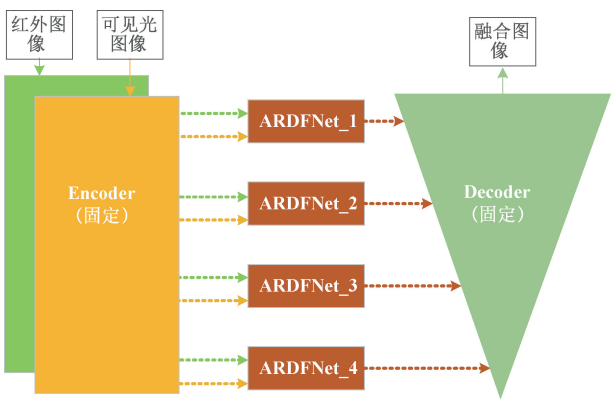


图 8　融合网络训练架构

Fig. 8　Fused network training architecture

本文使用数据集 KAIST[21] 作为训练集，训练图像数量为 80 000 张，同样先将训练图像转换为灰度图像，而后图像的尺寸重塑为 256×256，batch_size 设置为 4，epoch 设置为 2，学习率设置为 1×10^{-4}，使用 Adam 作为优化器。

在融合网络中，本文设计了一种 ARDFNet 的损失函数 L_{ARDF} 如式(4)所示：

$$L_{ARDF} = \alpha(L_{TD} + L_{SD}) + L_F \tag{4}$$

其中，L_{TD} 表示目标细节损失函数，L_{SD} 表示场景细节损失函数，L_F 表示特征损失函数，α 为目标细节损失函数和背景细节损失函数的增强权重参数，由于要在结构相似性上保留更多的信息，所以设置了增强权重。

在红外和可见光图像融合中，红外图像不仅包含精确的目标信息，而且还伴有一些场景纹理信息，同理可见光图像不仅包含大部分场景信息，而且还具有一定的目标细节信息，所以利用 L_{TD} 和 L_{SD} 可以在完整保留目标信息和场景信息的情况下，极大程度地保留红外图像中的一些场景纹理信息和可见光图像中的目标细节信息，L_{TD} 和 L_{SD} 分别如式(5)和(6)所示：

$$L_{TD} = 1 - SSIM(I_f, I_{ir}) \tag{5}$$

$$L_{SD} = 1 - SSIM(I_f, I_{vis}) \tag{6}$$

其中，I_f 代表融合图像，I_{ir} 代表红外图像，I_{vis} 代表可见光图像。其次，由于在多尺度特征提取过程中，红外特征包含着比可见光特征更多的目标信息，本文设计一种特征损失函数 L_F，L_F 如式(7)所示：

$$L_F = \frac{1}{HW}\sum_{k=1}^{K}(f_f^k - (\beta_{ir}f_{ir}^k + \beta_{vis}f_{vis}^k))^2 \tag{7}$$

其中，H 和 W 为特征图的高和宽，$K=4$ 为多尺度特征的层数，f_f 代表融合特征，f_{ir} 代表红外特征，f_{vis} 代表可见光特征，β_{ir} 和 β_{vis} 为红外特征和可见光特征的增强权重参数，由于很多算法的融合图像无法保留更多的可见光图像特征，导致图像细节辨识度差，所以设置红外特征和可见光特征的增强权重参数，旨在特征层次上保留更多的场景纹理信息和目标细节信息。

2.3　结果分析

1）测试数据集，本文在公开的红外与可见光数据集 TNO[22] 上进行测试，使用 3 对红外图像与可见光图像作为测试样本。

2）对比实验与评价指标，为了比较所提出方法与最先进算法的融合性能，选取了 7 种具有代表性算法，包括 GTF[23]、IMSVD[23]、DenseFuse[15]、FusionGAN[16]、PMGI[24]、RFN-Nest[18] 和 SDNet[25]，融合图像对比实验为图 9~13。其中 GTF 和 IMSVD 是典型的传统融合方法，DenseFuse、FusionGAN、PMGI、RFN-Nest 和 SDNet 都属于深度学习融合方法，在对比实验中，通过借鉴原始论文设置所有算法的参数。并且本文使用评价指标客观评估本文的融合算法，使用的评价指标包括：互信息（MI）[26]、差异相关性（SCD）[27]、视觉信息保真度（VIF）[28]、特征互信息（FMI）[29]、多层次结构相似度（MS-SSIM）[30] 和结构相似性度（SSIM）[31]，表 1~3 分别为本文方法与其他几种方法的客观评价指标对比结果。

本文的融合方法（ARDFusion）与其他融合方法的第 1 对融合图像对比实验如图 9 所示，GTF 与 IMSVD 生成的融合图像红外目标周围没有虚影，噪声较多，可见光图像的背景纹理信息不够突出。FusionGAN 与 RFN-Nest 的融合图像整体模糊不清，大大降低了红外图像与可见光图像特征的突出性，DenseFuse、PMGI 和 SDNet 虽然对于红外背景细节保留较好，但是可见光的背景纹理缺失

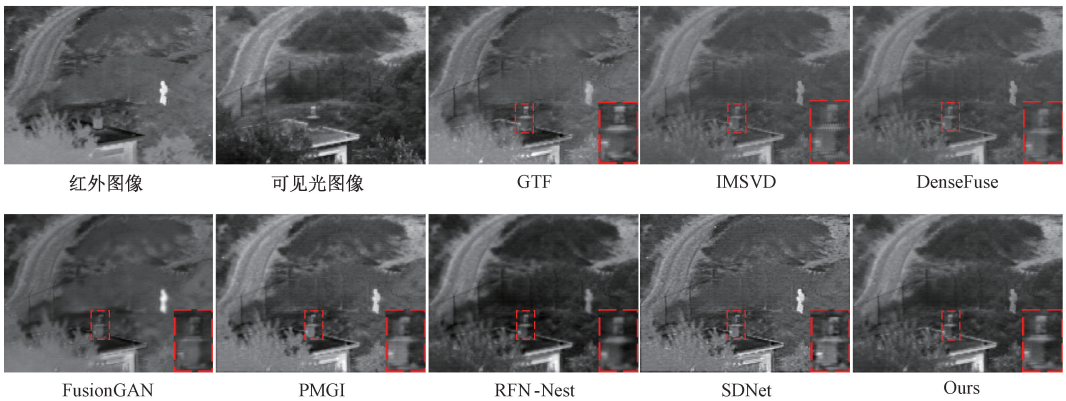


图 9　Camp 图像的融合结果

Fig. 9　Fusion results of camp images

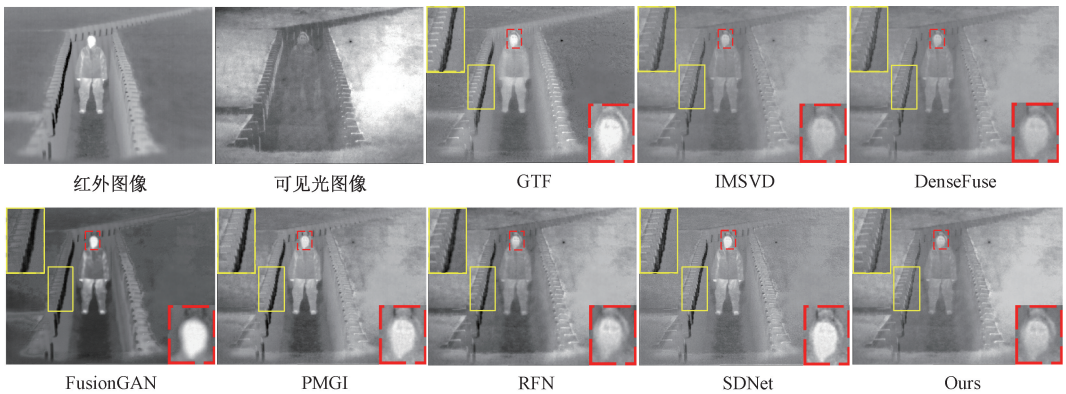


图 10　Bridge 图像融合结果

Fig. 10　Fusion results of bridge images

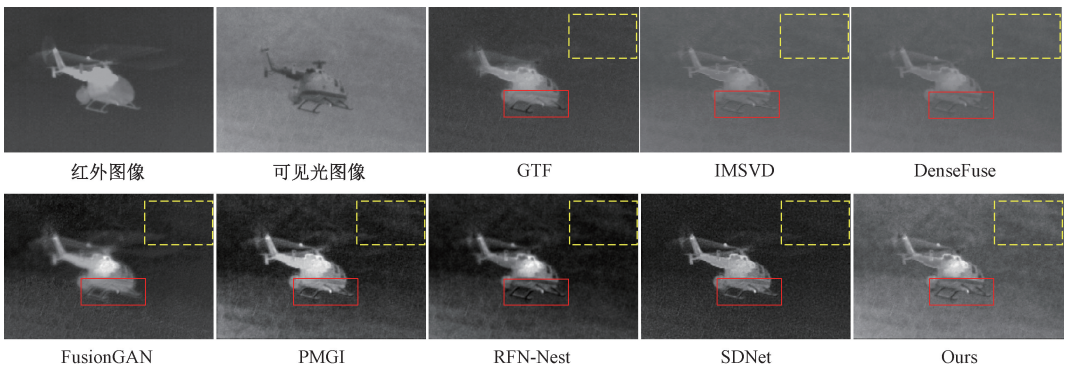


图 11　Helicopter 图像融合结果

Fig. 11　Fusion results of helicopter images

较为严重,而本文方法生成的融合图像在最大程度保留可见光的背景纹理信息的同时也保留了红外图像中的目标信息与红外背景信息,虽然红外目标并没有过多增强,但“树林”的纹理特征保留更加充分,“烟囱”更具有层次感,融合图像整体更符合人眼视觉感官。

客观评价结果表明,本文算法在 *VIF*、*MS-SSIM* 和 *SSIM* 中提升最多,对比其他 7 种算法分别提高 5.8%、1.0%和 2.5%,其他 3 项指标也为最优,客观评价指标如表 1 所示。其中斜粗体为最优值,正粗体为次优值。

表1　Camp图像融合的客观评价指标

Table 1　Objective evaluation index of Camp image fusion

方法	*MI*	*SCD*	*VIF*	*FMI*	*MS-SSIM*	*SSIM*
GTF	2.050 7	0.969 6	0.399 4	0.884 4	0.784 3	0.434 8
IMSVD	1.541 7	1.477 4	0.397 2	0.874 3	0.868 7	0.505 2
DenseFuse	1.612 4	1.484 0	0.449 0	0.882 4	0.870 0	**0.539 4**
FusionGAN	**2.131 7**	1.137 8	0.3532	0.878 0	0.668 7	0.363 4
PMGI	2.112 7	1.791 7	0.491 3	0.870 2	0.872 6	0.530 7
RFN-Nest	1.717 2	**1.888 9**	**0.493 9**	**0.888 6**	**0.924 2**	0.471 2
SDNet	1.950 4	1.565 3	0.443 4	0.870 2	0.856 5	0.519 2
本文	***2.140 1***	***1.899 7***	***0.524 4***	***0.888 7***	***0.933 7***	***0.553 5***

在图10中,GTF、FusionGAN和SDFusion这3种方法无法突出虚线框所框选包含在可见光图像中的人脸面部细节,在剩余的几种图像融合方法中,只有RNF-Nest和本文的融合图像能较为清晰地保留人脸面部细节。SDNet和PMGI对于目标特征保留充分,但缺失可见光图像特征,相比于上两者,IMSVD和DenseFuse红外目标特征平缓,同时也保留了一些可见光图像特征,但仍存在噪声。在实线框区域中,只有IMSVD、DenseFuse和本文的融合图像保留了更多的可见光背景纹理特征,其余方法的黄框区域更倾向于红外图像,而本文方法更均匀的融合了红外图像与可见光图像,使整体图像保证红外目标特征不丢失的前提下,更多的保留可见光图像中的目标细节信息和背景纹理信息。

客观评价结果表明,本文算法在6项评价指标中均取得了最高评分,客观评价指标如表2所示。其中斜粗体为最优值,正粗体为次优值。

表2　Bridge图像融合客观评价

Table 2　Objective evaluation of Camp image fusion

方法	*MI*	*SCD*	*VIF*	*FMI*	*MS-SSIM*	*SSIM*
GTF	**2.619 4**	1.099 8	0.497 7	**0.908 1**	0.784 3	0.404 1
IMSVD	1.231 3	1.935 9	0.348 6	0.883 0	0.868 7	0.479 7
DenseFuse	1.327 5	1.937 7	0.437 0	0.886 2	0.870 0	**0.554 7**
FusionGAN	2.587 9	1.889 2	0.370 7	0.888 4	0.668 7	0.342 3
PMGI	1.916 5	**1.982 3**	0.518 2	0.889 2	0.872 6	0.510 7
RFN-Nest	1.641 7	1.980 1	**0.582 1**	0.897 5	**0.924 2**	0.419 7
SDNet	2.476 8	1.726 0	0.452 1	0.891 4	0.856 5	0.478 1
本文	***2.628 1***	***1.983 1***	***0.591 4***	***0.915 9***	***0.933 7***	***0.556 9***

在图11中,由于可见光图像本身存在一定的噪声干扰,所以这几种融合方法的融合图像也会生成噪声,但是在背景区域,大量的噪声中具有一定的云雾信息,也就是可见光图像中存在一部分背景纹理信息,与本文对比的7种融合方法中只有PMGI和RFN-Nest生成的融合图像存在少量的可见光背景纹理特征,而本文方法生成的融合图像最大限度的保留了背景纹理特征,可以更清晰地看到背景云雾。对于实线框内的“起落架”,除了DenseFuse和IMSVD方法外,都可以看到较为清晰的轮廓,但本文方法生成的融合图像对于红框内的“起落架”,在突出目标的前提下,保留了更多可见光图像的目标纹理特征。

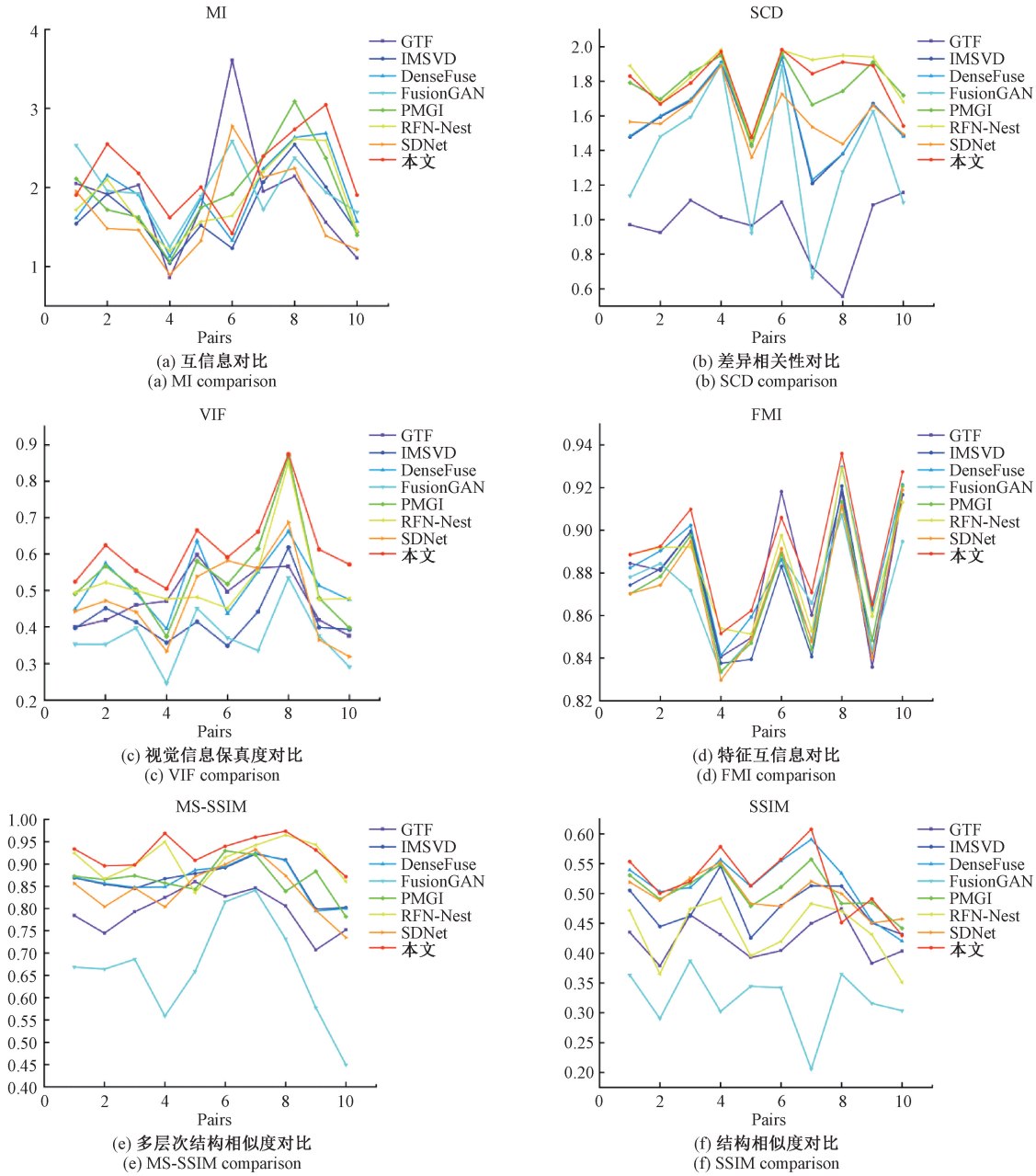
由本次对比实验可以看出,本文算法在6项评价指标中均为最优,在*VIF*中较GTF算法提升了5.5%,在*MI*中,本文算法的评分更是比排名第二位的FusionGAN算法提升了22.8%,客观评价指标如表3所示。其中斜体为最优值,正粗体为次优值。

表3　Helicopter图像融合客观评价

Table 3　Objective evaluation of Helicopter image fusion

方法	*MI*	*SCD*	*VIF*	*FMI*	*MS-SSIM*	*SSIM*
GTF	0.860 6	1.106 6	**0.477 3**	0.840 6	0.824 7	0.430 9
IMSVD	1.042 6	1.652 8	0.357 4	0.837 6	0.866 9	0.545 9
DenseFuse	1.125 4	1.647 4	0.395 6	0.841 4	0.848 5	**0.556 6**
FusionGAN	**1.249 4**	1.454 5	0.246 0	0.833 3	0.558 7	0.302 2
PMGI	1.062 9	1.863 3	0.374 7	0.833 8	0.857 4	0.548 5
RFN-Nest	1.195 6	**1.918 2**	0.476 7	**0.853 9**	**0.949 7**	0.491 5
SDNet	0.895 8	1.640 3	0.333 9	0.829 6	0.803 7	0.552 2
本文	***1.617 9***	***1.935 7***	***0.505 3***	***0.854 6***	***0.968 8***	***0.578 4***

在本节最后,使用现有的 7 种算法与本文算法在包含 10 对图像的测试集上进行客观评价指标对比实验,如图 12 所示。



(a) 互信息对比
(a) MI comparison

(b) 差异相关性对比
(b) SCD comparison

(c) 视觉信息保真度对比
(c) VIF comparison

(d) 特征互信息对比
(d) FMI comparison

(e) 多层次结构相似度对比
(e) MS-SSIM comparison

(f) 结构相似度对比
(f) SSIM comparison

图 12　各算法的十对图像融合客观评价

Fig. 12　Objective evaluation of ten pairs of image fusion algorithms

实验结果表明本文算法在 *VIF*、*MS-SSIM*、*SSIM* 中表现最为突出,而后是 *MI*、*SCD* 和 *FMI*。由于视觉信息保真度与主观视觉有更高的一致性,多层次结构相似度和结构相似度用于度量图像之间的相似度,3 种评价指标值越大,表明图像质量越好,所以本文算法在图像整体结构完整性和主观视觉感受性上表现更加优秀,其次互信息与像素特征互信息促使融合图像包含源图像的更多细节特征,差异相关性保证融合图像更加自然且最大程度减少人为噪声对融合图像的干扰,使得本文算法在图像噪声较大的情况下仍然可以保证整体结构完整、局部细节突出和主观视觉感受良好。

2.4　消融实验

本文为验证特征损失函数 L_F 中可红外特征权重 β_{ir} 和可见光特征权重 β_{vis} 对实验结果的影响,将两种权重值分别取值为 0.1、2、4、6,并对不同权值的训练模型进行

比较，主观对比结果如图 13 所示，横坐标与纵坐标的最后一幅图像分别为红外源图像与可见光源图像，客观评价指标对比结果如表 4 所示。从图 13 和表 4 中可以看到每个权重值的取值对于融合图像的影响，当 $\beta_{ir}=\beta_{vis}$ 时，融合图像包含源图像信息虽然较为均衡，但图像中“地砖”和“树丛”的纹理特征不明显，也就是说明部分可见光场景纹理细节表现较差。当 $\beta_{ir}=0.1$，$\beta_{vis}>0.1$ 时，融合图像更加倾向于突出场景信息的可见光图像，红外图像中的目标特征不明显。同理 $\beta_{vis}=0.1$，$\beta_{ir}>0.1$ 时，融合图像更加倾向于突出目标特征的红外图像，可见光图像在红外图像的高对比度下，几乎看不到任何场景纹理信息。当 $\beta_{ir}>\beta_{vis}>0.1$ 且 $\beta_{ir}\neq\beta_{vis}$ 时，融合图像包含红外图像的目标信息较多，过多的抑制了可见光图像特征，所以使得主观视觉上表现较差。相反，当 $\beta_{vis}>\beta_{ir}>0.1$ 且 $\beta_{ir}\neq\beta_{vis}$ 时，融合图像包含可见光图像的细节信息较多，虽然红外图像的目标特征不如 $\beta_{ir}>\beta_{vis}$ 的取值情况下突出，但符合主观视觉观看，对于可见光图像特征与红外图像特征的保留更加平衡。

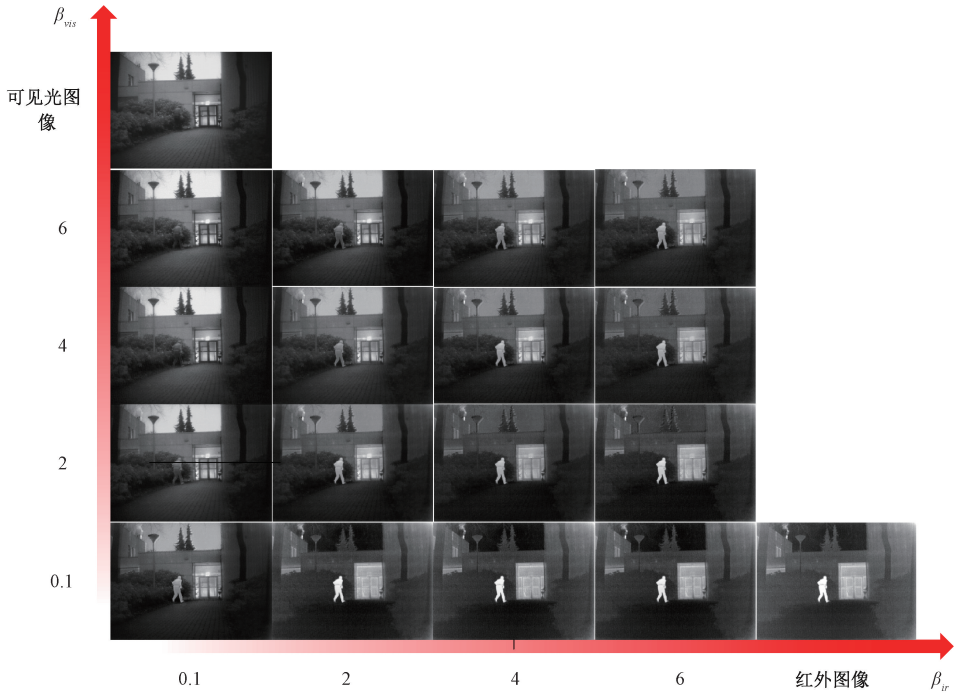


图 13　消融实验图像融合结果

Fig. 13　Ablation experimental image fusion results

表 4　消融实验图像融合客观评价

Table 4　Objective evaluation of ablation experimental image fusion

β_{ir}	β_{vis}	*MI*	*SCD*	*VIF*	*FMI*	*MS-SSIM*	*SSIM*
0.1	0.1	2.420 5	1.632 6	0.482 7	0.930 1	0.870 8	0.470 2
	2	2.221 1	1.399 2	0.554 0	0.938 8	0.787 0	0.544 9
	4	2.317 9	1.417 7	**0.653 8**	0.939 9	0.870 9	0.511 7
	6	2.597 2	1.370 1	0.618 1	**0.943 4**	0.870 5	0.510 5
2	0.1	2.346 5	1.433 7	0.528 7	0.933 1	0.823 2	0.545 2
	2	2.268 3	**1.835 0**	0.564 2	0.912 4	0.898 6	0.539 5
	4	**2.670 6**	1.820 8	0.613 2	0.921 5	0.916 9	0.526 8
	6	2.320 3	1.793 1	0.652 5	0.922 6	**0.927 8**	**0.554 0**
4	0.1	2.180 3	1.282 5	0.539 7	0.933 0	0.780 0	0.484 5
	2	2.258 7	1.721 9	0.573 2	0.923 3	0.921 6	0.537 4
	4	2.279 6	1.828 8	0.588 0	0.924 1	0.915 7	0.543 2
	6	***2.681 7***	***1.835 3***	***0.668 9***	***0.946 2***	***0.933 4***	***0.560 7***
6	0.1	2.194 2	1.798 7	0.550 2	0.934 4	0.829 5	0.535 2
	2	2.529 4	1.756 3	0.571 0	0.928 1	0.926 3	0.542 6
	4	2.149 6	1.719 3	0.537 8	0.931 3	0.913 4	0.523 7
	6	2.261 9	1.777 5	0.581 1	0.926 7	0.882 8	0.538 2

在消融实验中,仍然使用 *MI*、*SCD*、*VIF*、*FMI*、*MS-SSIM* 和 *SSIM* 6 种评价指标进行客观评价。从表 4 可以看出,在权重的客观评价对比过程中,当 $\beta_{ir}=4$、$\beta_{vis}=6$ 时,融合图像在这 6 项指标的综合评价中取得了最佳成绩。综合主观评价和客观评价,使用 $\beta_{ir}=4$、$\beta_{vis}=6$ 作为特征损失函数最终的权重取值。其中斜粗体为最优值,正粗体为次优值。

3　结　论

本文提出了一种自编码器网络和融合网络相结合的红外与可见光图像融合算法。其中注意力残差密集融合网络由密集残差模块、目标注意力模块以及场景注意力模块组成。首先经过编码器获得多尺度特征,而后注意力残差密集融合网络旨在充分融合红外与可见光特征中的目标信息、场景纹理信息和目标细节信息,从而进一步丰富融合图像的原图像信息。随后利用解码器重构图像,最大程度保留各个尺度的目标信息和细节信息。最后,实验结果表明,该算法在主观视觉评价和和客观指标评价中优于已有文献中具有代表性的算法,达到了较好的融合效果。

参考文献

[1] 罗 娟,王立平. 基于非下采样 Contourlet 变换耦合特征选择机制的可见光与红外图像融合算法[J]. 电子测量与仪器学报,2021,35(7):163-169.
LUO J, WANG L P. Infrared and visible image fusion algorithm based on nonsubsampled contourlet transform coupled with feature selection mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(7):163-169.

[2] 付 涵,严 华. 基于多尺度变换和 VGG 网络的红外与可见光图像融合[J]. 现代计算机,2021(16):112-117.
FU H, YAN H. Infrared and visible image fusion based on multi-scale transform and VGG network[J] Modern Computer, 2021(16):112-117.

[3] LIU C H, QI Y, DING W R. Infrared and visible image fusion method based on saliency detection in sparse domain[J]. Infrared Physics & Technology, 2017, 83: 94-102.

[4] 江兆银,王磊. 基于显著性检测与权重映射的可见光与红外图像融合算法[J]. 电子测量与仪器学报,2021,35(1):174-182.
JIANG ZH Y, WANG L. Visible and infrared image fusion algorithm based on significance detection and weight mapping[J]. Journal of Electronic Measurement and Instrumentation, 2021,35(1):174-182.

[5] 王娟,柯聪,刘敏,等. 神经网络框架下的红外与可见光图像融合算法综述[J]. 激光杂志,2020,41(7):7-12.
WANG J, KE C, LIU M, et al. Overview of infrared and visible image fusion algorithms based on neural network framework [J]. Laser Journal, 2020,41(7):7-12.

[6] 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.

[7] 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.

[8] 符贻,李俊霖,韦嵬. 基于小波变换的图像融合研究[J]. 电脑知识与技术,2022,18(34):32-34.
FU Y, LI J L, WEI W. Research on image fusion based on wavelet transform. [J]. Computer Knowledge and Technology, 2022,18(34):32-34.

[9] 余 腾,胡伍生,孙小荣,等. 基于非下采样 Contourlet 变换耦合能量相似制约的遥感图像融合算法[J]. 电子测量与仪器学报,2021,35(6):71-78.
YU T, HU W SH, SUN X R, et al. Remote sensing image fusion algorithm based on nonsubsampled contourlet transform combined with energy similarity restriction[J]. Journal of Electronic Measurement and Instrumentation, 2021,35(6):71-78.

[10] LI H, CEN Y, LIU Y, et al. Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion[J]. IEEE Transactions on Image Processing, 2021, 30: 4070-4083.

[11] XU H, ZHANG H, MA J. Classification saliency-based rule for visible and infrared image fusion [J]. IEEE Transactions on Computational Imaging, 2021, 7: 824-836.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] TANG L, YUAN J, ZHANG H, et al. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware[J]. Information Fusion, 2022, 83: 79-92.

[20] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]. Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.

[21] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: Benchmark dataset and baseline[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1037-1045.

[22] XU H, WANG X, MA J. DRF: Disentangled representation for visible and infrared image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.

[23] MA J, 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.

[24] 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.

[25] ZHANG H, MA J. SDNet: A versatile squeeze-and-decomposition network for real-time image fusion[J]. International Journal of Computer Vision, 2021, 129: 2761-2785.

[26] 陈昭宇,范洪博,马美燕,等. 基于结构重参数化的红外与可见光图像融合[J/OL]. 控制与决策:1-9[2023-04-17].
CHEN ZH Y, FAN H B, MA M Y, et al. Infrared and visible image fusion based on structural re-parameterization[J/OL]. Control and Decision: 1-9[2023-04-17].

[27] 李威,田时舜,刘广丽,等. M-SWF域红外与可见光图像结构相似性融合[J/OL]. 红外技术:1-8[2023-04-17].
LI W, TIAN SH SH, LIU G L, et al. Structural similarity fusion of infrared and visible image in M-SWF domain[J/OL]. Infrared Technology:1-8[2023-04-17].

[28] 陈彦林,王志社,邵文禹,等. 红外与可见光图像多尺度Transformer融合方法[J]. 红外技术,2023,45(3): 266-275.
CHEN Y L, WANG ZH SH, SHAO W Y, et al. Multi-scale Transformer fusion method for infrared and visible images[J]. Infrared Technology, 2023, 45(3): 266-275.

[29] LI K, QI M, ZHUANG S, et al. TIPFNet: A Transformer-based infrared polarization image fusion network[J]. Optics Letters, 2022, 47(16): 4255-4258.

[30] 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.

[31] REN L, PAN Z, CAO J, et al. Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation[J]. Infrared Physics & Technology, 2021, 117: 103839.

作者简介

陈广秋,1999年于吉林大学获得学士学位,2006年于吉林大学获得硕士学位,2015年于吉林大学获得博士学位,现为长春理工大学副教授,主要研究方向为图像处理与机器视觉。

E-mail: gaungqiu_chen@126.com

Chen Guangqiu received his B. Sc. degree from Jilin University in 1999, M. Sc. degree from Jilin University in 2006 and Ph. D. degree from Jilin University in 2015, respectively. Now he is an associate professor in Changchun University of Science and Technology. His main research interests include image processing and machine vision.

温奇璋,2016年于吉林工程技术师范学院获得学士学位,现为长春理工大学硕士研究生,主要研究方向为图像处理与机器视觉。

E-mail: wqz17843088635@163.com

Wen Qizhang received B. Sc. degree from Jilin Engineering Normal University in 2016. He is now a M. Sc. candidate at Changchun University of Science and

Technology. His main research interests include image processing and machine vision.

尹文卿,2016 年于吉林师范大学获得学士学位,现为长春理工大学硕士研究生,主要研究方向为图像处理与机器视觉。

E-mail: yinwenqing0037@126.com

Yin Wenqing received B. Sc. degree from Jilin Normal University in 2016. He is now a M. Sc. candidate at Changchun University of Science and Technology. His main research interests include image processing and machine vision.

段锦(通信作者),1993 年于北京理工大学获得学士学位,1998 年于沈阳工业学院获得硕士学位,2004 年于吉林大学获得博士学位,现为长春理工大学教授,主要研究方向为偏振成像探测、图像处理与模式识别、数字光学环境仿真。

E-mail: duanjin@vip.sina.com

Duan Jin (Corresponding author) received his B. Sc. degree from Beijing Institute of Technology in 1993, M. Sc. degree from Shenyang Institute of Technology in 1998 and Ph. D. degree from Jilin University in 2004, respectively. Now he is a professor in Changchun University of Science and Technology. His main research interests include polarization imaging detection, image processing and pattern recognition, digital optical environment simulation.

黄丹丹,2007 年于长春理工大学大学获得学士学位,2009 年于东北大学获得硕士学位,2014 年于大连理工大学获得博士学位,现为长春理工大学讲师,主要研究方向为计算机视觉和机器学习。

E-mail: hdd@cust.edu.cn

Huang Dandan received her B. Sc. degree from Changchun University of Science and Technology in 2007, M. Sc. degree from Northeastern University in 2009 and Ph. D. degree from Dalian University of Technology in 2014, respectively. Now she is a lecturer in Changchun University of Science and Technology. Her main research interests include computer vision and machine learning.