计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202110639
60(3): 690−704, 2023

基于特征空间多分类对抗机制的红外与可见光图像融合

张 浩 马佳义 樊 凡 黄 珺 马 泳
(武汉大学电子信息学院 武汉 430072)
(zhpersonalbox@gmail.com)

Infrared and Visible Image Fusion Based on Multiclassification Adversarial Mechanism in Feature Space

Zhang Hao, Ma Jiayi, Fan Fan, Huang Jun, and Ma Yong
(*Electronic Information School*, *Wuhan University*, *Wuhan* 430072)

Abstract To break the performance bottleneck caused by traditional fusion rules, an infrared and visible image fusion network based on multiclassification adversarial mechanism in the feature space is proposed. Compared with existing methods, the proposed method has more reasonable fusion rule and better performance. First, an autoencoder introducing attention mechanism is trained to perform the feature extraction and image reconstruction. Then, the generative adversarial network (GAN) is adopted to learn the fusion rule in the feature space extracted by the trained encoder. Specifically, we design a fusion network as the generator to fuse the features extracted from source images, and then design a multi-classifier as the discriminator. The multiclassification adversarial learning can make the fused features approximate both infrared and visible probability distribution at the same time, so as to preserve the most salient characteristics in source images. Finally, the fused image is reconstructed from the fused features by the trained decoder. Qualitative experiments show that the proposed method is in subjective evaluation better than all state-of-the-art infrared and visible image fusion methods, such as GTF, MDLatLRR, DenseFuse, FusionGAN and U2Fusion. In addition, the objective evaluation shows that the number of best quantitative metrics of our method is 2 times that of U2Fusion, and the fusion speed is more than 5 times that of other comparative methods.

Key words image fusion; fusion rule; deep learning; autoencoder; generative adversarial network

摘 要 为突破传统融合规则带来的性能瓶颈,提出一个基于特征空间多类别对抗机制的红外与可见光图像融合网络.相较于现存方法,其融合规则更合理且性能更好.首先,训练一个引入注意力机制的自编码器网络实现特征提取和图像重建.然后,采用生成式对抗网络(generative adversarial network, GAN)在训练好的编码器提取的特征空间上进行融合规则的学习.具体来说,设计一个特征融合网络作为生成器融合从源图像中提取的特征,然后将一个多分类器作为鉴别器.这种多分类对抗学习可使得融合特征同时逼近红外和可见光2种模态的概率分布,从而保留源图像中最显著的特征.最后,使用训练好的译码器从特征融合网络输出的融合特征重建出融合图像.实验结果表明:与最新的所有主流红外与可见光图像融合方法包括GTF, MDLatLRR, DenseFuse, FusionGAN, U2Fusion相比,所提方法主观效果更好,客观指标最优个数为U2Fusion的2倍,融合速度是其他方法的5倍以上.

关键词 图像融合;融合规则;深度学习;自编码器;生成式对抗网络

中图法分类号 TP391

收稿日期:2021-06-11;**修回日期**:2022-06-07

基金项目:国家自然科学基金项目(62075169, 62003247, 62061160370);湖北省自然科学基金项目(2019CFA037, 2020BAB113)
This work was supported by the National Natural Science Foundation of China (62075169, 62003247, 62061160370) and the Natural Science Foundation of Hubei Province (2019CFA037, 2020BAB113).

通信作者:马佳义(jyma2010@gmail.com)

图像融合旨在从不同传感器或不同拍摄设置捕获的图像中提取最有意义的信息，并将这些信息融合生成单幅信息更完备、对后续应用更有利的图像[1-3]. 红外与可见光图像融合是应用最为广泛的图像融合任务之一. 具体来说，红外传感器对成像环境较鲁棒，所捕获的红外图像具有显著的对比度，能有效地将热目标与背景区分开. 然而，红外图像往往缺乏纹理细节，不符合人类的视觉感知习惯. 相反，可见光图像往往包含丰富的纹理细节，但容易受天气、光照等因素影响，且无法有效突出目标. 红外与可见光图像融合致力于同时保留这 2 种模态的优异特性，以生成既具有显著对比度又包含丰富纹理细节的图像. 由于融合图像的优良特性，红外与可见光图像融合已被广泛应用于军事探测、目标监控以及车辆夜间辅助驾驶等领域[4-5].

现存的红外与可见光图像融合方法根据其原理可分为传统方法和基于深度学习的方法. 传统方法通常利用相关的数学变换在空间域或变换域进行活动水平测量，并设计相应的融合规则来实现图像融合[6]. 代表性方法有：基于多尺度变换的方法[7-8]、基于稀疏表示的方法[9]、基于子空间的方法[10]、基于显著性的方法[11]以及混合方法[12]. 一般来说，这些传统方法手工设计的活动水平测量及融合规则具有较大的局限性：一方面，源图像的多样性势必会使这些手工设计越来越复杂；另一方面，这也限制了融合性能的进一步提升，因为不可能以手工设计的方式考虑所有因素.

近年来，深度学习的快速发展推动了图像融合领域的巨大进步. 基于深度学习的融合方法凭借神经网络强大的特征提取和图像重建能力，不断提升融合性能[13]. 根据图像融合的实现过程，现存的基于深度学习的图像融合方法可以分为端到端融合方法和非端到端融合方法. 端到端融合方法[14-17]通常在损失函数的引导下隐式地实现特征提取、特征融合及图像重建，其损失函数被定义为图像空间中融合图像与源图像绝对分布（如像素强度、梯度等原始图像属性）之间的距离，如图 1 所示. 在这一类方法中，图像融合网络的优化实际上是寻求红外与可见光图像绝对分布的中和比例，这势必会造成有益信息被削弱，如纹理结构和热目标被中和.

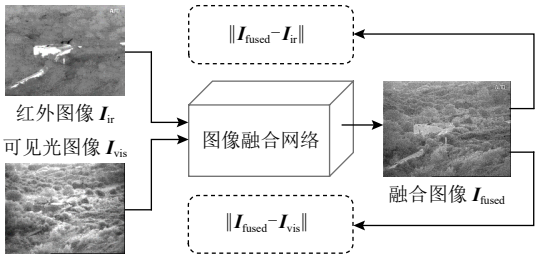


Fig. 1 Schematic of the end-to-end fusion method

图 1 端到端融合方法示意图

非端到端融合方法一般基于自编码网络，其先用编码器实现特征提取，然后使用融合策略聚合提取到的特征，最后使用译码器对融合特征进行译码实现图像重建. 然而，在现存非端到端图像融合方法中，所采用的中间特征融合策略仍然是传统的[18]，如 Mean 策略、Max 策略以及 Addition 策略等，如图 2 所示. 这些融合策略是全局的，不能根据输入图像来自适应地调整，融合性能十分有限. 比如，Mean 策略对输入特征直接取平均，会造成显著目标的亮度被中和；Addition 策略直接将输入特征相加，会造成部分区域亮度中和或饱和.

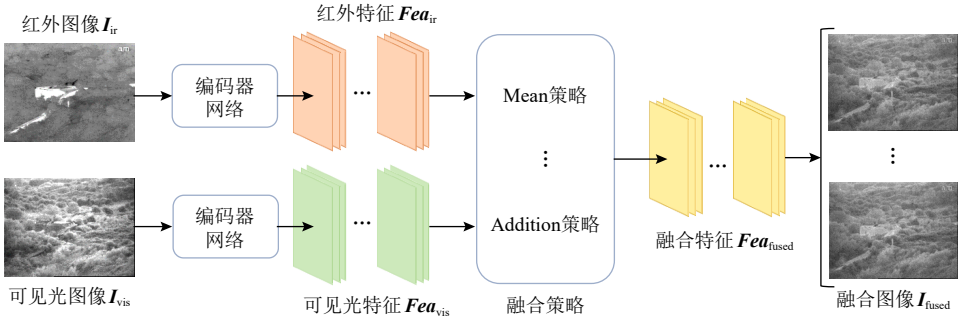


Fig. 2 Schematic of the non-end-to-end fusion method

图 2 非端到端融合方法示意图

为了解决上述挑战，本文提出一个基于特征空间多类别对抗机制的红外与可见光图像融合网络，显著提升了融合性能. 首先，该方法基于自编码网络，利用编码器网络和译码器网络显式地实现特征提取和图像重建. 其中，编码器网络引入了空间注意力机制来关注更重要的区域，如显著目标区和丰富纹理区；译码器网络引入通道注意力机制来筛选对重建图像本身更有利的通道特征，如高频特征通道和包

含了显著性目标的低频特征通道. 此外, 译码器网络还采用了多尺度卷积, 其可以从不同尺度处理特征, 从而在重建过程中更好地保留细微纹理. 然后, 采用生成式对抗网络 (generative adversarial network, GAN) 实现中间特征融合策略的可学习化. 具体来说, 本文设计了一个特征融合网络作为生成器来融合由训练好的编码器提取的特征, 其致力于生成同时符合红外和可见光 2 种模态概率分布的融合特征. 提出一个多分类器鉴别器, 其致力于区分红外特征、可见光特征以及融合特征. 特征融合网络和多分类器鉴别器持续地进行多分类对抗学习, 直到多分类器鉴别器认为融合特征既是可见光特征, 又是红外特征. 此时, 特征融合网络便能保留红外图像和可见光图像中最显著的特性, 从而生成高质量的融合特征. 最终的融合图像由训练好的译码器网络对融合特征译码得到. 值得注意的是, 所提方法采用的多分类对抗机制区别于传统 GAN[19] 的二分类对抗, 其更符合图像融合任务的多源信息拟合需求. 与当前基于 GAN 的图像域对抗融合方法[16] 也不同, 所提方法首次将生成对抗机制引入特征空间, 对技术路线中的“特征融合”环节更具针对性. 更重要的是所提方法摆脱了当前几乎所有的基于 GAN 的融合方法都需要的距离(内容)损失, 仅在 GAN 分类决策所捕获的模态概率分布(如对比度、纹理等模态属性)之间构建损失, 有效地避免了有益信息的削弱, 从而实现显著热目标和丰富纹理结构的自适应保留.

所提方法有两大优势: 1)相较于现存端到端的融合方法, 本文方法没有使用融合图像与源图像绝对分布之间的距离作为损失函数, 而是在分类决策捕获的模态概率分布之间建立对抗损失, 从而避免有益信息被削弱. 2)相较于现存非端到端的融合方法, 所提方法将中间特征融合策略可学习化, 能够根据输入图像自适应地调整融合规则, 较好地保留了源图像中的显著对比度和丰富纹理细节. 这种智能融合策略可以避免传统融合策略造成的亮度中和或饱和以及信息丢失等问题. 为了直观展示所提方法的优势, 选取了代表性的端到端融合方法 U2Fusion[15] 和非端到端融合方法 DenseFuse[18] 来对比显示, 其中 DenseFuse 按照原始论文建议选取了性能相对较好的 Addition 策略, 融合结果的差异如图 3 所示. 可以看出, U2Fusion 的融合结果中出现了典型的亮度中和现象, 目标建筑物的亮度没有被保持, 纹理结构也很不自然. DenseFuse 使用 Addition 融合策略, 虽然能较好地维持纹理结构的显著性, 但目标建筑物的亮度依旧被削弱. 相比之下, 本文方法能显著地改善这些问题, 融合结果不但准确地保持了目标建筑物的亮度, 而且包含丰富的纹理细节. 这得益于所提方法中特征融合网络的优异性能, 其能自适应地保留红外与可见光的模态特性.

本文的主要贡献有 3 个方面: 1)提出了一个新的红外与可见光图像融合网络, 其利用多分类对抗机制将传统融合策略扩展为可学习, 具有更好的融合性能. 2)所提模型将现存方法中融合图像与源图像绝对分布之间的距离损失扩展为模态概率分布之间的对抗损失, 有效避免了现存融合方法中有益信息被削弱的问题. 3)本文方法具有良好的泛化性, 可以推广到任意红外与可见光图像融合数据集.

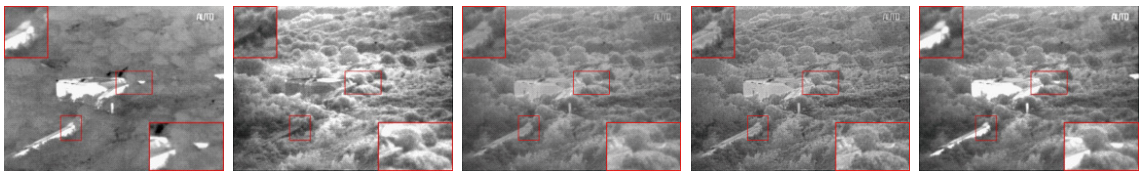


Fig. 3 Comparison of fusion performance

图 3 融合性能对比

1 相关工作

本节回顾和所提方法最相关的技术和工作, 包括基于深度学习的融合方法及 GAN.

1.1 基于深度学习的图像融合方法

近几十年, 基于深度学习的融合方法凭借神经网络强大的特征提取和图像重建能力, 获得了远超传统方法的性能[20]. 现存的基于深度学习的图像融合方法可以分为端到端融合方法以及非端到端融合方法.

端到端融合方法通常直接使用一个整体网络将输入的红外和可见光图像进行融合. 换句话说, 融合的各个阶段如特征提取、特征融合以及图像重建都

是隐式的. 端到端融合方法可根据所采取的架构分为基于卷积神经网络的融合方法[21-22]和基于GAN的融合方法[23-25]. 这些方法的共性在于依赖融合图像与源图像绝对分布之间的距离损失. 例如, PMGI[14]在融合图像和2幅源图像间建立强度和梯度距离损失, 并通过调节损失项的权重系数来调整信息融合过程中的保留比例, 从而控制融合结果绝对分布的倾向性. U2Fusion[15]则在融合图像和2幅源图像间建立强度和结构相似度损失[26], 并通过度量特征图的信息质量来自适应地调整损失项系数, 从而引导融合图像保留有效信息. 不幸的是, 这种融合图像与2幅源图像绝对分布之间的距离损失会建立一个博弈, 导致最终融合图像是2幅源图像原始属性(如像素强度、梯度)的折中, 不可避免地造成有益信息被削弱. 除此以外, 武汉大学的Ma等人[16]将GAN架构引入到图像融合领域并提出了引起广泛关注的FusionGAN, 其中网络的优化不仅依赖图像绝对分布之间的距离损失, 还依赖模态概率分布之间的对抗损失. 随后, 文献[16]的作者引入双鉴别器来平衡红外与可见光信息以进一步提升融合性能[17], 但是网络优化仍离不开图像绝对分布之间的内容损失, 这意味有益信息的丢失问题仍然存在.

非端到端融合方法主要是基于自编码架构[27], 其特征提取、特征融合以及图像重建3个阶段都是非常明确的, 由不同的网络或模块来实现. 现存非端到端图像融合方法的融合质量一直受融合策略的性能制约. 具体来说, 现存的基于自编码结构的融合方法采用的融合规则都是手工制作的, 且不可学习. 例如, DenseFuse[18]采用Addition策略和l_1-norm策略; SEDRFuse[28]采用最大值策略. 这些策略不能根据输入图像自适应地调整, 可能会造成亮度中和或过饱和、信息丢失等问题, 因此, 研究可学习的融合规则非常有意义.

1.2 GAN

原始GAN由Goodfellow等人[19]于2014年提出, 其由一个生成器和一个鉴别器组成. 生成器是目标网络, 致力于生成符合目标分布的伪数据; 鉴别器是一个分类器, 其负责准确分辨出真实数据和生成器伪造的假数据. 因此, 生成器和鉴别器之间是敌对关系. 也就是说, 生成器希望生成鉴别器无法区分的伪数据, 而鉴别器则希望能准确鉴别出伪数据. 生成器和鉴别器不断迭代地优化, 直到鉴别器无法区分是真实数据还是由生成器产生的伪数据. 此时, 生成器便具备生成符合目标分布数据的能力. 下面, 我们形式化上述对抗学习过程.

假设生成器被表示为G, 鉴别器被表示为D, 输入到生成器的随机数据为$Z=\{z_1, z_2, \cdots, z_n\}\sim P_z$, 目标数据为$X=\{x_1, x_2, \cdots, x_n\}\sim P_X$. 那么, 生成器致力于估计目标数据$X$的分布$P_X$, 并尽可能生成符合该分布的数据$G(Z)$, 而鉴别器$D$需要对真实数据$X$和生成的伪数据$G(Z)$进行准确区分. 总而言之, GAN的目的就是在不断地对抗训练中使得伪数据的分布P_G不断逼近目标数据分布P_X. 因此, GAN的目标函数被定义为

$$\min_G \max_D E_{x_i\sim P_X}[\log(D(x_i))]+ E_{z_i\sim P_Z}[\log(1-D(G(z_i)))]. \tag{1}$$

随着研究的深入, 研究者发现使用交叉熵损失的原始GAN在训练过程中非常不稳定, 且生成结果质量不高. 最小二乘GAN[29]的提出改善了这一现象, 其使用最小二乘损失作为损失函数, 引入标签来引导生成器和鉴别器的优化. 最小二乘GAN的目标函数被定义为

$$\min_D V_{\mathrm{LSGAN}}(D)=\frac{1}{2}E_{x_i\sim P_X}\left[D(x_i)-r^2\right]+ \frac{1}{2}E_{z_i\sim P_Z}\left[D(G(z_i))-s^2\right], \tag{2}$$

$$\min_G V_{\mathrm{LSGAN}}(G)=\frac{1}{2}E_{z_i\sim P_Z}\left[D(G(z_i))-t^2\right], \tag{3}$$

其中r, s, t是对应的概率标签. 具体来说, r是鉴别器判定目标数据集合X中数据x_i对应的标签, 设定$r=1$; s是鉴别器判定由生成器构造的伪数据$G(z_i)$对应的标签, 设定$s=0$; t是生成器希望鉴别器判定伪数据$G(z_i)$对应的标签, 设定$t=1$.

2 本文方法

本节详细描述提出的基于特征空间多分类对抗机制的红外与可见光图像融合网络. 首先, 我们给出问题建模, 然后介绍网络详细结构, 最后提供损失函数的具体设计.

2.1 问题建模

从定义上来说, 图像融合是从源图像中提取最有意义的特征, 将它们融合并重建包含更丰富信息的单幅图像. 因此, 图像融合的整个过程可以分为3个阶段: 特征提取、特征融合以及图像重建. 基于上述思想, 本文提出一个基于特征空间多分类对抗机制的红外与可见光图像融合网络, 其总体框架如图4所示.

首先, 鉴于自编码器网络的“低维—高维—低维”

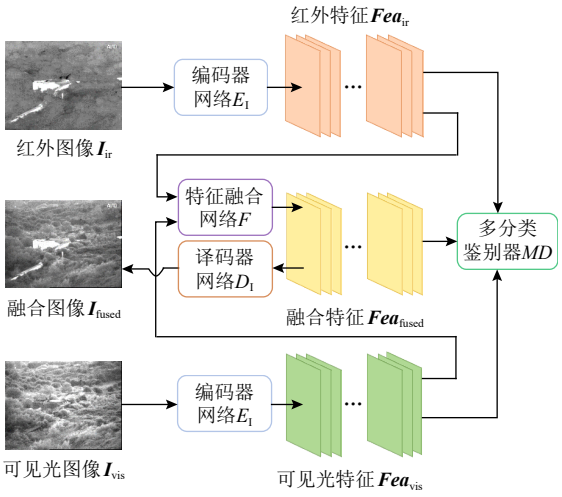


Fig. 4 Overall framework of the proposed method

图 4 本文方法的总体框架图

映射理念非常契合特征提取和图像重建这 2 个环节，所提方法提出一个引入注意力机制的自编码器网络来实现特征提取和图像重建. 其中，编码器网络E_I中的空间注意力机制能使得低维向高维映射时更关注那些包含重要信息的区域，如包含丰富纹理或显著热目标的区域；而译码器网络D_I中的通道注意力机制则使得高维向低维映射时更关注对重建图像更有利的通道特征，如高频特征通道和包含显著目标的低频特征通道. 除此以外，译码器网络D_I还引入了多尺度卷积来加强对细微空间纹理的保留.

其次，使用训练好的编码器网络E_I从红外和可见光图像中提取特征，并设计一个特征融合网络 F 来融合这些特征，这种可学习的特征融合策略比现存方法所使用的传统融合策略具有更强的性能. 具体来说，所提的特征融合网络 F 被当作生成器，然后结合使用 1 个多分类鉴别器 MD，二者构成特征空间上的生成式对抗网络. 特征融合网络 F 致力于同时估计红外与可见光 2 种模态特征概率分布，以生成同时符合这 2 种模态概率分布的融合特征；而多分类鉴别器 MD 则致力于准确区分可见光特征、红外特征以及特征融合网络生成的融合特征. 经过持续的对抗学习，直到多分类鉴别器认为融合特征既是红外特征又是可见光特征，此时该融合特征便具备了红外和可见光 2 种模态中最显著的特性. 值得注意的是，所提模型中生成式对抗网络的优化仅依赖于模态概率分布之间的对抗损失，不依赖绝对分布之间的距离损失，这极大地避免了现存方法中存在的有益信息被削弱问题. 最终，将特征融合网络 F 生成的融合特征经训练好的译码器网络D_I译码得到高质量的融合图像$\boldsymbol{I}_{fused}$. 整个融合过程可以被形式化为

$$\boldsymbol{I}_{fused}=D_I(F(E_I(\boldsymbol{I}_{ir}),E_I(\boldsymbol{I}_{vis}))), \tag{4}$$

其中$\boldsymbol{I}_{ir}$和$\boldsymbol{I}_{vis}$分别表示红外图像和可见光图像；$E_I(\cdot)$表示编码器网络对应的功能函数，$F(\cdot)$表示特征融合网络对应的功能函数，$D_I(\cdot)$表示译码器网络对应的功能函数.

2.2 网络结构

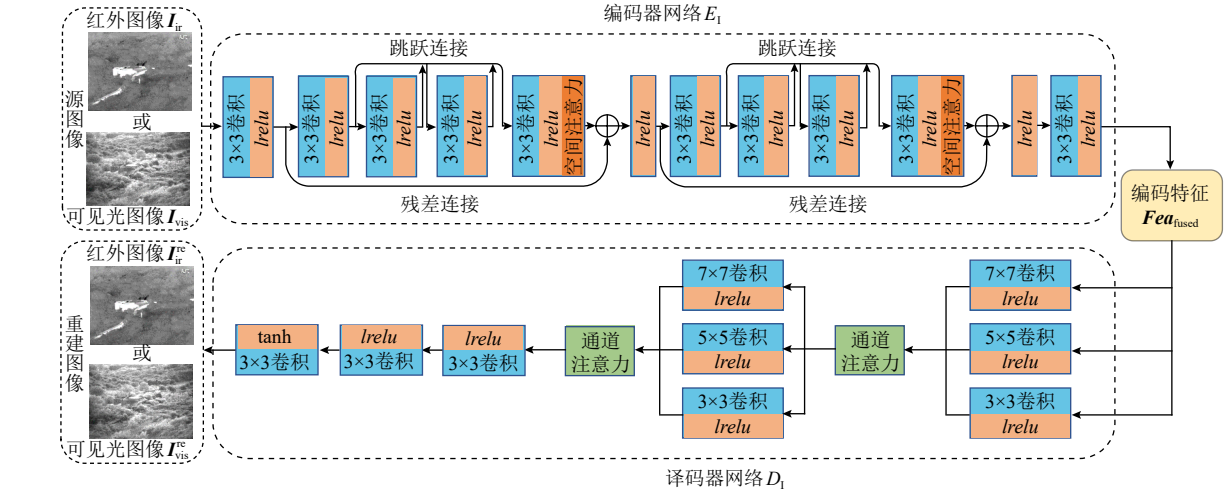
本文所提红外与可见光图像融合网络包括 2 部分：负责特征提取和图像重建的自编码器网络；负责融合规则学习的 GAN.

2.2.1 自编码器网络

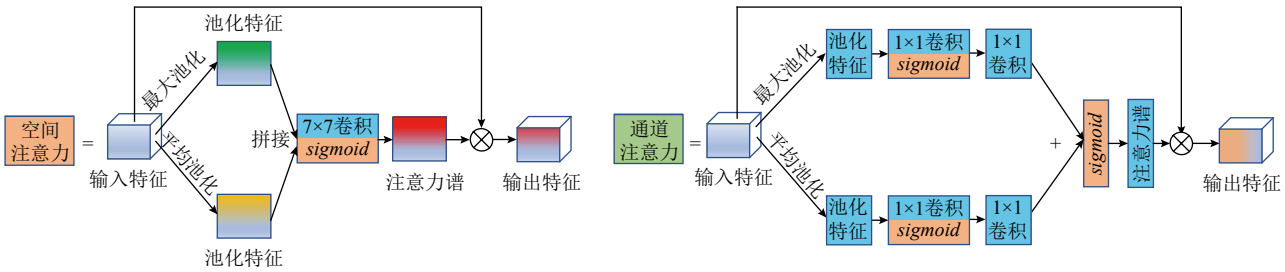
自编码器网络是一种经典的自监督网络，其以重建输入数据为导向，先利用编码器网络将图像映射到高维特征空间，再利用译码器网络将高维特征重新映射为图像. 因为译码器网络重建图像的质量依赖于中间高维特征的质量，所以编码器网络必须能提取具有高表达能力的特征，而译码器网络必须具备从中间特征准确重建出源图像的能力. 本文提出了一种新的自编码器网络来实现融合过程中的特征提取和图像重建，如图 5(a)所示.

编码器网络E_I使用 10 个卷积层从源图像中提取特征，其中卷积核尺寸均为3×3，激活函数均为 *lrelu*(leaky relu). 在第 5 和第 9 层后，使用空间注意力模块对所提特征沿空间位置加权，以增强特征中重要的空间区域(如显著目标、结构纹理). 空间注意力模块[30]的网络结构如图 5(b)所示，可以看到，空间注意力模块先使用最大池化和平均池化对固定空间位置不同通道的信息进行聚合，然后使用 1 个卷积层处理串接的聚合特征，以生成与原始特征空间尺寸相同的注意力谱. 该注意力谱本质上是一系列学习到的权重，对输入特征沿着空间维度进行选择性加权，从而实现感兴趣区域特征的增强. 在编码器中使用空间注意力模块可以有效满足对感兴趣特征的提取偏好，提升编码特征的表达能力. 此外，编码器还将密集连接[31]和残差连接[32]相结合，其一方面把浅层特征不断跳跃连接到深层网络以增强后续特征表达能力和增加特征利用率，另一方面残差连接也避免了特征提取过程出现的梯度消失和爆炸问题.

在译码器网络D_I中，先使用 2 个结合通道注意力模块的多尺度卷积层处理由编码器网络E_I提取的中间特征. 在每个多尺度卷积层，3 个具有不同尺寸卷积核的卷积层并行处理输入特征，其卷积核尺寸分别为7×7, 5×5, 3×3, 激活函数均为 *lrelu*. 通道注意力模



(a) 编码器和译码器网络结构图



(b) 空间注意力模块网络结构图　　(c) 通道注意力模块网络结构图

Fig. 5　Structures of the autoencoder network for feature extraction and image reconstruction

图 5　用于特征提取和图像重建的自编码器网络结构图

块[30]的网络结构如图 5(c)所示，其先使用最大池化和平均池化对固定通道不同空间位置的特征进行聚合，然后在 2 个分支中分别使用 2 个卷积层对聚合特征进行处理，最后将 2 个分支的处理结果求和得到最终的注意力谱. 该注意力谱是一个长度与输入特征通道数相同的向量，表示将为输入特征每个通道分发的权重. 在自监督重建的优化导向下，译码器将自适应地关注对重建更重要的特征通道，从而提升重建精度. 最后，使用 3 个卷积核尺寸为3×3的卷积层来重建源图像. 其中，除了最后一层，其他卷积层均使用 *lrelu* 作为激活函数，最后一层使用 tanh 作为激活函数. 在上述特定设计下，所提自编码网络具有强大的特征提取和图像重建能力.

2.2.2　GAN

本文设计了一种新颖的特征融合规则构建方式，其利用 GAN 将融合策略可学习化，从而获得更好的融合性能，如图 6 所示.

首先，特征融合网络 F 在对抗架构中扮演生成器的角色，其将训练好的编码器网络 E_1 提取的红外特征 $\boldsymbol{Fea}_{ir}$ 和可见光特征 $\boldsymbol{Fea}_{vis}$ 进行融合，生成融合特征 $\boldsymbol{Fea}_{fused}$. 在特征融合网络 F 中，先使用 3 个卷积核尺寸为 3×3、激活函数为 *lrelu* 的卷积层来处理输入的红外特征与可见光特征. 然后，采用 3 个分支来分别预测融合权重 ω_{ir}, ω_{vis} 以及偏差项 ε. 每个分支包含 2 个卷积层，其卷积尺寸均为 3×3. 在融合权重预测分支，2 个卷积层分别使用 *lrelu* 和 *sigmoid* 作为激活函数；在偏差预测分支，2 个卷积层的激活函数均为 *lrelu*. 融合特征可以被表示为

$$\begin{aligned}\boldsymbol{Fea}_{fused} &= F(\boldsymbol{Fea}_{ir}, \boldsymbol{Fea}_{vis}) = \\ &\omega_{ir} \cdot \boldsymbol{Fea}_{ir} + \omega_{vis} \cdot \boldsymbol{Fea}_{vis} + \varepsilon.\end{aligned} \tag{5}$$

其次，使用 1 个多分类鉴别器 *MD* 来区分红外特征 $\boldsymbol{Fea}_{ir}$、可见光特征 $\boldsymbol{Fea}_{vis}$ 以及特征融合网络 F 合成的融合特征 $\boldsymbol{Fea}_{fused}$. 在多分类鉴别器 *MD* 中，先使用 4 个卷积层来处理输入特征，它们的卷积核尺寸均为 3×3，激活函数均为 *lrelu*. 然后，处理后的特征被重塑为 1 个 1 维向量，并使用 1 个线性层来输出 1 个 1×2 的预测向量，分别表示输入特征为红外特征的概率 P_{ir}，以及输入特征为可见光特征的概率 P_{vis}. 特征融合网络 F 和多分类鉴别器 *MD* 连续地对抗学习，直到多分类鉴别器 *MD* 认为生成器产生的融合特征既

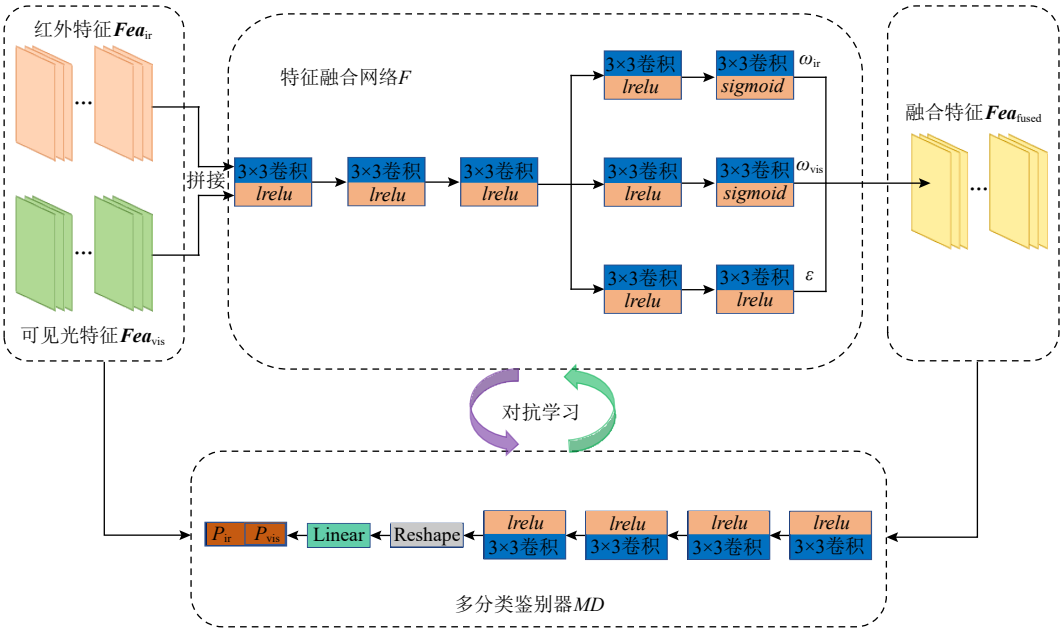


Fig. 6 Structures of generative adversarial network for fusion rule learning

图 6 用于融合规则学习的生成式对抗网络结构图

是红外特征又是可见光特征，此时特征融合网络F便学会合理的融合规则.

2.3 损失函数

本文的损失函数包括 2 部分：自编码器网络损失函数和生成式对抗网络损失函数.

2.3.1 自编码器网络损失函数

自编码器网络先利用编码器网络$E_{\rm I}$将低维图像映射为高维特征，再利用译码器网络$D_{\rm I}$重新将高维特征映射为低维图像. 也就是说，自编码器网络致力于重建输入图像. 所提方法在强度域和梯度域构建重建图像与输入图像的一致性损失：

$$\mathcal{L}_{\rm AE}=\mathcal{L}_{\rm int}+\beta\mathcal{L}_{\rm grad}, \tag{6}$$

其中$\mathcal{L}_{\rm AE}$是自编码器网络的总损失，$\mathcal{L}_{\rm int}$是强度损失，$\mathcal{L}_{\rm grad}$是梯度损失，β是平衡强度损失项和梯度损失项的参数. 值得注意的是，自编码器网络的训练同时在红外图像与可见光图像上进行，即红外图像和可见光图像共享编码器网络$E_{\rm I}$和译码器网络$D_{\rm I}$的权重，因此对应的强度损失和梯度损失被定义为：

$$\mathcal{L}_{\rm int}=\left|\boldsymbol{I}_{\rm ir}^{\rm re}-\boldsymbol{I}_{\rm ir}\right|+\left|\boldsymbol{I}_{\rm vis}^{\rm re}-\boldsymbol{I}_{\rm vis}\right|, \tag{7}$$

$$\mathcal{L}_{\rm grad}=\left|\nabla\boldsymbol{I}_{\rm ir}^{\rm re}-\nabla\boldsymbol{I}_{\rm ir}\right|+\left|\nabla\boldsymbol{I}_{\rm vis}^{\rm re}-\nabla\boldsymbol{I}_{\rm vis}\right|, \tag{8}$$

其中$\boldsymbol{I}_{\rm ir}$和$\boldsymbol{I}_{\rm vis}$是输入源红外和可见光图像，$\boldsymbol{I}_{\rm ir}^{\rm re}$和$\boldsymbol{I}_{\rm vis}^{\rm re}$是自编码网络重建的红外和可见光图像，其可以表示为$\boldsymbol{I}_{(\cdot)}^{\rm re}=D_{\rm I}(E_{\rm I}(\boldsymbol{I}_{(\cdot)}))$. 此外，$|\cdot|$是$\ell_1$范数，$\nabla$是 Sobel 梯度算子，其从水平和竖直 2 个方向来计算图像的梯度. 在上述损失的约束下，编码器网络$E_{\rm I}$能较好地从源图像中提取特征，译码器网络$D_{\rm I}$则能从编码特征中准确地重建源图像.

2.3.2 GAN 损失函数

生成式对抗网络通过连续地对抗学习构建高性能融合规则，其网络优化仅依赖于模态概率分布之间的对抗损失，不依赖融合图像与源图像绝对分布之间的距离损失，极大地避免了有益信息被削弱.

对于特征融合网络 F，其目的是产生可以骗过多分类鉴别器 MD 的融合特征$\boldsymbol{Fea}_{\rm fused}$，即让 MD 认为所生成的融合特征$\boldsymbol{Fea}_{\rm fused}$既是红外特征$\boldsymbol{Fea}_{\rm ir}$又是可见光特征$\boldsymbol{Fea}_{\rm vis}$. 因此，特征融合网络 F 的损失 $\mathcal{L}_F$ 为

$$\mathcal{L}_F=(MD(\boldsymbol{Fea}_{\rm fused})[1]-a)^2+(MD(\boldsymbol{Fea}_{\rm fused})[2]-a)^2, \tag{9}$$

其中$MD(\cdot)$表示多分类鉴别器的函数，其输出是 1 个 1×2 的概率向量. $MD(\boldsymbol{Fea}_{\rm fused})[1]$ 指的是该向量的第 1 项，表示多分类鉴别器判定输入特征是红外特征的概率 $P_{\rm ir}$；$MD(\boldsymbol{Fea}_{\rm fused})[2]$ 指的是该向量的第 2 项，表示多分类鉴别器判定输入特征是可见光特征的概率 $P_{\rm vis}$. a是概率标签，设定 $a=0.5$，即特征融合网络希望通过自身的优化使得多分类鉴别器无法区分融合特征是红外特征还是可见光特征.

与特征融合网络 F 成敌对关系，多分类鉴别器 MD 希望能准确判断输入特征是红外特征、可见光特征还是由特征融合网络 F 产生的融合特征. 因此，多分类鉴别器损失 $\mathcal{L}_{MD}$ 包括 3 部分：判定红外特征的损失 $\mathcal{L}_{MD_{\rm ir}}$、判定可见光特征的损失 $\mathcal{L}_{MD_{\rm vis}}$ 以及判定融

合特征的损失 $\mathcal{L}_{MD_{\text{fused}}}$，即

$$\mathcal{L}_{MD}=\alpha_1\mathcal{L}_{MD_{\text{ir}}}+\alpha_2\mathcal{L}_{MD_{\text{vis}}}+\alpha_3\mathcal{L}_{MD_{\text{fused}}},\tag{10}$$

其中，$\alpha_1,\alpha_2,\alpha_3$是平衡这些损失项的参数.

当输入特征为红外特征$\boldsymbol{Fea}_{\text{ir}}$，多分类鉴别器判定的 P_{ir} 应趋于 1，P_{vis} 应趋于 0. 对应的损失函数 $\mathcal{L}_{MD_{\text{ir}}}$ 被定义为

$$\begin{aligned}\mathcal{L}_{MD_{\text{ir}}}=&(MD(\boldsymbol{Fea}_{\text{ir}})[1]-b_1)^2+\\&(MD(\boldsymbol{Fea}_{\text{ir}})[2]-b_2)^2,\end{aligned}\tag{11}$$

其中b_1和b_2是红外特征对应的概率标签，设定$b_1=1$，$b_2=0$，即多分类鉴别器应该准确识别出输入特征是红外特征而不是可见光特征.

类似地，当输入特征为可见光特征$\boldsymbol{Fea}_{\text{vis}}$，对应的损失函数$\mathcal{L}_{MD_{\text{vis}}}$被定义为

$$\begin{aligned}\mathcal{L}_{MD_{\text{vis}}}=&(MD(\boldsymbol{Fea}_{\text{vis}})[1]-c_1)^2+\\&(MD(\boldsymbol{Fea}_{\text{vis}})[2]-c_2)^2,\end{aligned}\tag{12}$$

其中c_1和c_2是可见光特征对应的概率标签，设定$c_1=0$，$c_2=1$，即多分类鉴别器应该准确识别出输入特征是可见光特征而不是红外特征.

当输入特征为融合特征$\boldsymbol{Fea}_{\text{fused}}$，多分类鉴别器输出的 P_{ir} 和 P_{vis} 都应趋于 0. 对应的损失函数 $\mathcal{L}_{MD_{\text{fused}}}$ 被定义为

$$\begin{aligned}\mathcal{L}_{MD_{\text{fused}}}=&(MD(\boldsymbol{Fea}_{\text{fused}})[1]-d_1)^2+\\&(MD(\boldsymbol{Fea}_{\text{fused}})[2]-d_2)^2,\end{aligned}\tag{13}$$

其中d_1和d_2是融合特征对应的概率标签，d_1和d_2都被设为 0，即 *MD* 应能准确识别出输入特征既不是红外特征也不是可见光特征.

3 实　　验

本节将在公开数据集上评估所提方法. 5 个最先进的红外与可见光图像融合方法被挑选作为对比，包括 GTF[12], MDLatLRR[33], DenseFuse[18], FusionGAN[16], U2Fusion[15]. 值得注意的是，在后续实验中，DenseFuse 使用推荐的性能更好的 Addition 策略. 首先，提供实验配置，如实验数据、训练细节以及评估指标. 其次，从定性和定量 2 方面实施对比实验. 本节还提供泛化性实验、效率对比及消融实验来验证所提方法的有效性.

3.1 实验设置

3.1.1 实验数据

本文选用 TNO 数据集[34] 和 MFNet 数据集[35] 作为对比实验的数据，TNO 数据集和 MFNet 数据集用于测试的图像对数量分别为 20 和 200，用于训练的数据分别为裁剪得到的 45 910 对和 96 200 对 80×80 的图像块. 此外，选用 RoadScene[36] 数据集作为泛化性实验的数据，用于测试的图像对数量为 20. 以上 3 个数据集中的图像对都已被严格配准[37].

3.1.2 训练细节

首先训练自编码器网络. 在自编码器网络的训练过程中，批大小被设置为 s_1，训练 1 期需要 m_1 步，一共训练 M_1 期. 在实验中，设置为 $s_1=48$，$M_1=100$，m_1 是训练图像块总数量和批大小 s_1 的比率. 自编码器网络训练好后冻结其参数，然后在训练好的编码器网络提取的特征空间中训练 GAN. 在 GAN 的训练过程中，批大小被设置为 s_2，训练 1 期需要 m_2 步，一共训练 M_2 期. 在实验中，设置 $s_2=48$，$M_1=20$，m_2 是训练图像块总数量和批大小 s_2 的比率. 无论是自编码器网络还是 GAN，都采用 Adam 优化器来更新参数. 在整个训练结束后，将编码器网络、特征融合网络以及译码器网络级联组成完整的图像融合网络. 值得注意的是，因为该图像融合网络是一个全卷积神经网络，输入可以是任意尺寸源图像对，即测试时不需要像训练那样对源图像进行裁剪. 此外，根据经验，设定式(6)中的参数 $\beta=10$，式(10)中的参数 $\alpha_1=0.25$，$\alpha_2=0.25$，$\alpha_3=0.5$. 所有的实验均在 GPU NVIDIA RTX 2080Ti 及 CPU Intel i7-8750H 上实施.

3.1.3 评估指标

本文从定性和定量 2 个方面评估各方法的性能. 定性评估是一种主观评估方式，其依赖于人的视觉感受，好的融合结果应同时包含红外图像的显著对比度和可见光图像的丰富纹理. 定量评估则通过一些统计指标来客观评估融合性能，本文选用了 7 个在图像融合领域被广泛使用的定量指标，如视觉信息保真度[38](visual information fidelity, *VIF*)、信息熵[39](entropy, *EN*)、差异相关和[40](the sum of the correlations of differences, *SCD*)、互信息[41](mutual information, *MI*)、质量指标[42](quality index, $Q^{\boldsymbol{AB/F}}$)、标准差[43](standard deviation, *SD*)及空间频率[44](spatial frequency, *SF*). *VIF* 测量融合图像保真度，大的 *VIF* 值表示融合图像保真度高；*EN* 测量融合图像的信息量，*EN* 值越大，融合图像包含的信息越多；*SCD* 测量融合图像包含的信息与源图像的相关性，*SCD* 越大意味着融合过程引入的伪信息越少；*MI* 衡量融合图像中包含来自源图像的信息量，*MI* 越大意味着融合图像包含来自源图像的信息越多；$Q^{\boldsymbol{AB/F}}$衡量融合过程中边缘信息的保持情况，$Q^{\boldsymbol{AB/F}}$越大，边缘被保持得越好；*SD* 是对融合图像对比度的反映，大的 *SD* 值表示良好的对比度；*SF* 测量融合图像整体细节丰富度，*SF* 越大，融合图像包含的纹理越丰富.

3.2 TNO 数据集上的对比实验

3.2.1 定性对比

首先，在 TNO 数据集上进行定性对比. 5 组典型的结果被挑选来定性地展示各方法的性能，如图 7 所示. 可以看出，本文所提方法有 2 方面的优势：一方面，本文方法能非常精确地保留红外图像中的显著目标，它们的热辐射强度几乎没有损失，且边缘锐利；另一方面，所提方法也能很好地保留可见光图像中的纹理细节.

从融合结果的倾向性可以把对比方法分为 2 类：第 1 类是融合结果倾向于可见光图像的方法，如 MDLatLRR, DenseFuse, U2Fusion. 从图 7 中可以看到，这一类方法的融合结果虽然包含丰富的纹理细节，但其对比度较差，热辐射目标被削弱. 例如，在第 1 组结果中，MDLatLRR, DenseFuse, U2Fusion 对树木纹理保留得较好，但却削弱了目标建筑物的亮度. 类似的还有第 2 组中的水面、第 3 组和第 4 组中的人以及第 4 组中的坦克. 第 2 类是融合结果倾向于红外图像的方法，如 GTF 和 FusionGAN. 这一类方法能较好地保留热目标，但纹理细节不够丰富，它们的结果看起来很像是锐化的红外图像. 如在图 7 中的第 1 组结果中，GTF 和 FusionGAN 较好地保留了目标建筑物的显著性，但周边树木的纹理结构却不够丰富. 类似地还有第 2 组中的灌木、第 3 组中的路灯以及第 4 组中的树叶. 本文所提方法综合了这 2 类方法的优势. 具体来说，所提方法既能像第 1 类方法那样保持场景中的纹理细节，又能像第 2 类方法那样准确保持热辐射目标. 值得注意的是所提方法对热目标边缘保持得比第 2 类方法更锐利. 总的来说，本文方法在定性对比上优于这些最新方法.

3.2.2 定量对比

进一步，在 20 幅测试图像上的定量对比结果如表 1 所示. 可以看出，本文所提方法在 *EN*, *SCD*, *MI*, $Q^{AB/F}$, *SD*, *SF* 这 6 个指标上都取得最高平均值；在 *VIF* 上，本文方法排行第 2，仅次于方法 U2Fusion. 这些结果说明：本文方法在融合过程中从源图像传输到融合图像的信息最多、引入的伪信息最少、能最好地保持边缘. 生成的融合结果包含的信息量最大、有最好的对比度、具有最丰富的整体纹理结构. 总的来说，本文方法相较于这些对比方法在定量指标上也是有优势的.

3.3 MFNet 数据集上的对比实验

3.3.1 定性对比

在 MFNet 数据集上实施定性对比实验，同样提供 5 组代表性的结果来展示各种方法的性能，如图 8

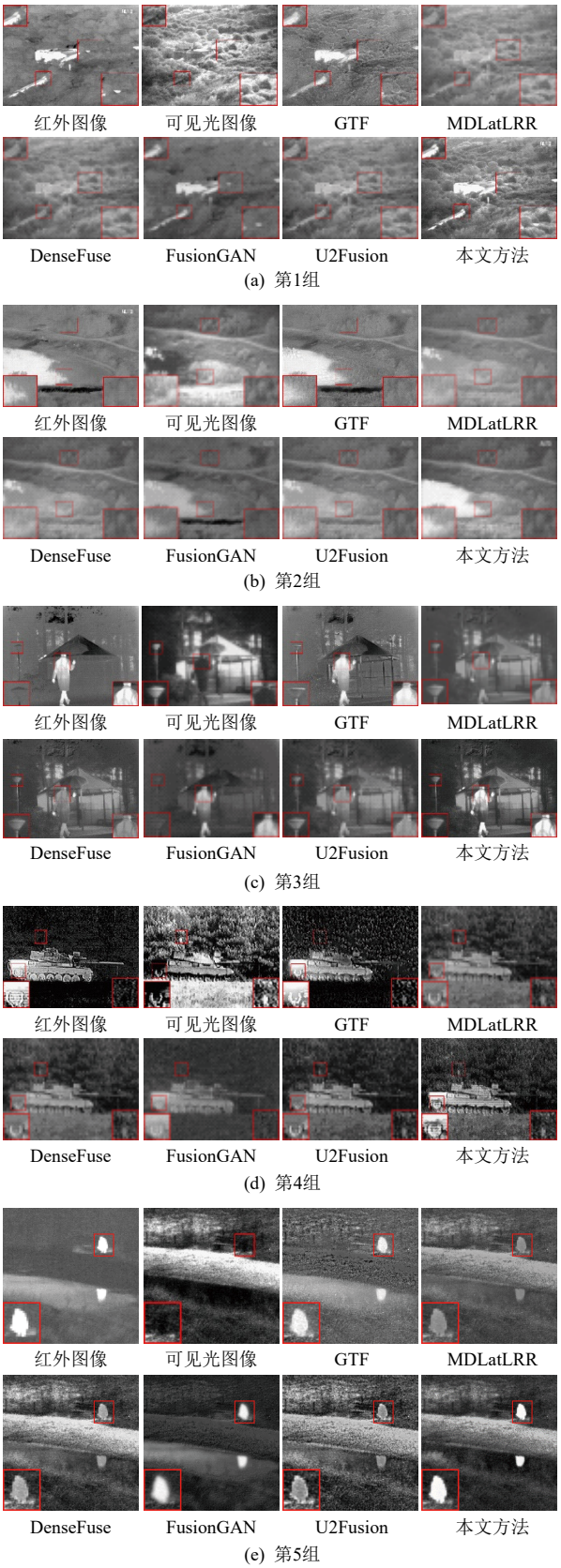


Fig. 7 Qualitative results of the comparative experiment on TNO dataset

图 7 TNO 数据集上对比实验的定性结果

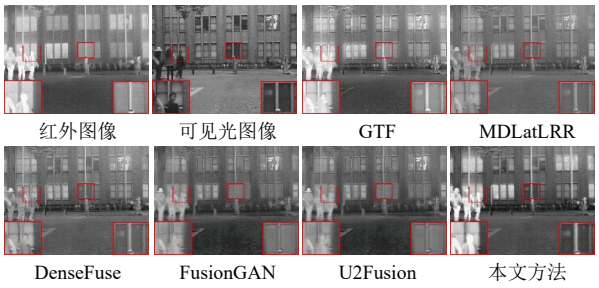
Table 1 Quantitative Results of the Comparative Experiment on TNO Dataset

表 1 TNO 数据集上对比实验的定量结果

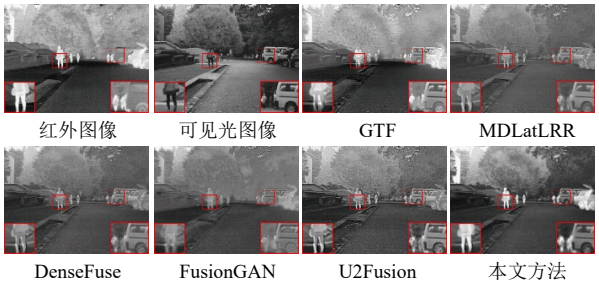
融合方法	*VIF*↑	*EN*↑	*SCD*↑	*MI*↑	$Q^{AB/F}$↑	*SD*↑	*SF*↑
GTF	0.350±0.052	6.753±0.396	0.985±0.165	1.200±0.440	0.423±0.100	35.157±11.405	10.315±5.268
MDLatLRR	0.346±0.051	6.438±0.408	1.663±0.135	1.037±0.225	0.435±0.077	26.148±6.242	7.930±3.587
DenseFuse	0.386±0.091	6.836±0.273	1.835±0.128	1.114±0.269	0.440±0.103	35.144±8.891	9.296±3.806
FusionGAN	0.231±0.046	6.450±0.323	1.512±0.228	1.099±0.207	0.210±0.055	27.683±6.052	6.075±2.051
U2Fusion	**0.423±0.106**	6.923±0.251	1.808±0.094	0.906±0.197	0.430±0.068	34.446±7.659	11.928±4.681
本文方法	0.414±0.103	**7.183±0.283**	**1.936±0.060**	**1.240±0.275**	**0.446±0.110**	**48.605±8.671**	**13.203±4.792**

注：↑表示值越高越好，加粗表示最优结果，加下划线表示次优结果.

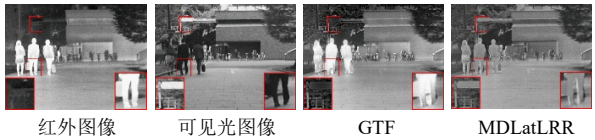
所示. 可以看到，只有 GTF, FusionGAN 以及本文方法能较好地维护红外图像中热辐射目标的显著度，但相较于这 2 种方法，本文方法能更好地保持热目标边缘的锐利性，呈现良好的视觉效果. 例如，在第 3, 4, 5 组结果中，本文方法能较好地保持热目标行人的姿态，而 GTF, FusionGAN 均由于边缘扩散导致轮廓模糊. 相反，MDLatLRR, DenseFuse, U2Fusion 太过于偏重于保留结构纹理，而忽视了热辐射目标保留，这导致一些场景中目标削弱或丢失. 例如，在第 2 组结果中，汽车旁边的微小行人在这些方法的结果中被丢失. 相较而言，本文方法能在热目标和结构纹理的保留上取得较好的平衡. 例如，第 1 组结果中，所提方法既维持了窗户的显著性，又保留了墙壁的纹理细节. 总体而言，本文方法在 MFNet 数据集的定性对比上比这些最新方法有优势.

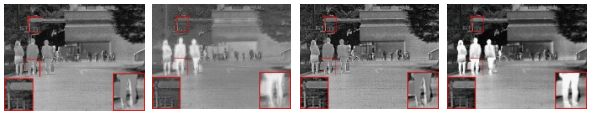


(a) 第1组

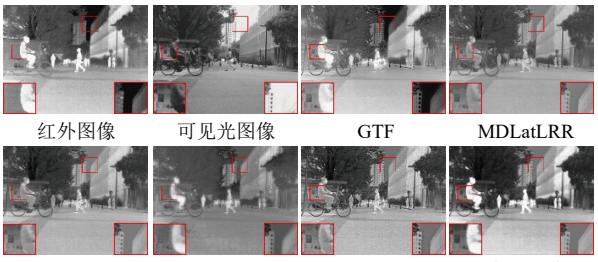


(b) 第2组

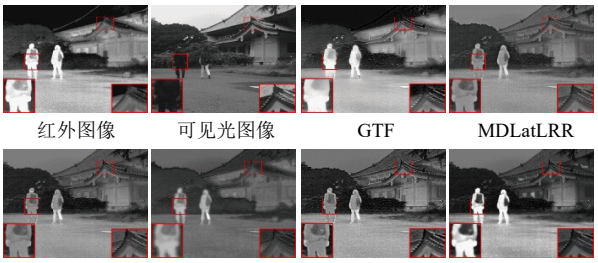




(c) 第3组



(d) 第4组



(e) 第5组

Fig. 8 Qualitative results of the comparative experiment on MFNet dataset

图 8 MFNet 数据集上对比实验的定性结果

3.3.2 定量对比

在 MFNet 数据集中的 200 幅测试图像上定量地对比这些最新方法以及本文所提方法，结果如表 2 所示. 本文方法在 *EN, SCD, MI, SD* 这 4 个指标上排行第 1，在指标 *SF* 上排行第 2，仅次于 U2Fusion. 这些客观结果表明本文方法所得结果包含的信息量最丰富、引入的伪信息最少，与源图像的相关性最大，以及具有最好的对比度，这些定量结果和图 8 展示的视觉结果相符合. 总的来说，本文方法在 MFNet 数据集上的定量对比上比其他方法性能更好.

3.4 泛化性实验

本文所提方法能较好地迁移到其他数据集，也

Table 2 Quantitative Results of the Comparative Experiment on MFNet Dataset

表 2 MFNet 数据集上对比实验的定量结果

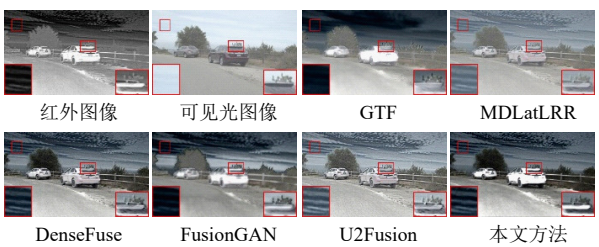
融合方法	*VIF*↑	*EN*↑	*SCD*↑	*MI*↑	$Q^{AB/F}$↑	*SD*↑	*SF*↑
GTF	0.311±0.022	7.458±0.197	1.027±0.186	1.575±0.230	0.399±0.059	55.343±8.671	10.501±1.866
MDLatLRR	0.327±0.025	6.896±0.225	1.306±0.239	1.325±0.233	0.461±0.034	39.477±7.181	9.016±0.986
DenseFuse	0.326±0.030	7.131±0.229	1.653±0.149	1.398±0.241	0.475±0.038	48.696±8.163	10.200±1.265
FusionGAN	0.178±0.022	6.882±0.300	0.609±0.594	1.424±0.186	0.234±0.044	35.397±5.765	7.299±1.288
U2Fusion	**0.350±0.035**	7.253±0.198	1.657±0.115	1.266±0.232	**0.496±0.028**	50.794±8.582	**14.072±1.546**
本文方法	0.319±0.027	**7.562±0.205**	**1.731±0.085**	**1.609±0.246**	0.422±0.036	**65.392±8.494**	10.749±1.242

注：↑表示值越高越好，加粗表示最优结果，加下划线表示次优结果.

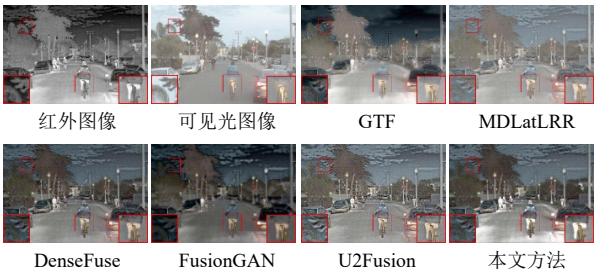
可以处理彩色可见光和红外图像融合. 为了评估本文方法的泛化性，实施了泛化性实验. 具体来说，使用 RoadScene 数据集中的图像测试在 TNO 数据集上训练得到的模型. 由于 RoadScene 数据集中的可见光图像是彩色图像，先将可见光图像从 RGB 转换到 YCbCr 色彩空间，然后融合 Y 通道与红外图像. 最后，将融合结果与 Cb 和 Cr 通道拼接在一起，并重新转换到 RGB 色彩空间得到最终的融合结果. 上述 5 种对比方法在泛化性实验中仍然被采用，且评估仍然从定性和定量 2 个方面来进行.

3.4.1 定性对比

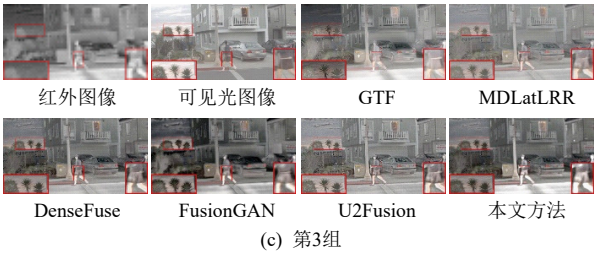
5 组代表性的定性结果被挑选来展示各方法的泛化能力，如图 9 所示. 可以看出，本文所提方法在 RoadScene 数据集上仍具有良好性能，且相较于对比方法在纹理保持和显著目标保留 2 个方面的优势仍十分明显. 首先，在显著目标保持上，本文所提方法表现最好，如第 1 组图像中的车辆、第 2 组和第 4 组中的骑行者，以及第 3 组和第 5 组中的行人. 相反，在 MDLatLRR, DenseFuse, U2Fusion 的融合结果中，这些显著目标被削弱. 虽然 GTF 和 FusionGAN 相对这些方法能更好地保留显著目标，但其在目标边缘保护上却不如所提方法. 其次，本文方法也能保证可见光图像中的纹理细节被很好地传输到融合图像中，如第 1 组和第 4 组结果中的云朵、第 2 组和第 3 组结果中的树木，以及第 5 组结果中的广告牌，而 GTF 和 FusionGAN 做不到这些. 因此，这些定性结果可以说明本文方法具有良好的泛化性，其能被迁移到 RoadScene 数据集，并得到高质量的融合图像.



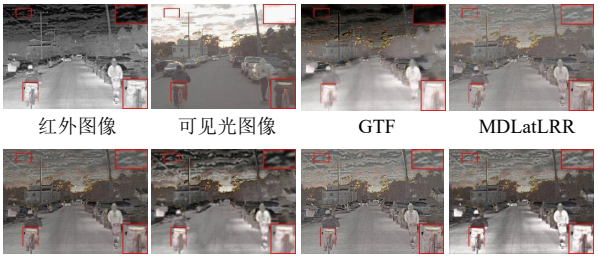
(a) 第1组



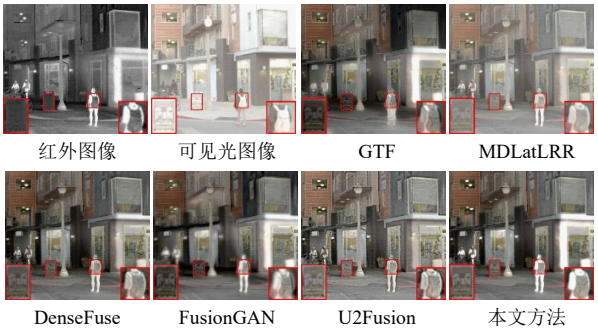
(b) 第2组



(c) 第3组



(d) 第4组



(e) 第5组

Fig. 9 Qualitative results of the generalization experiment

图 9 泛化实验的定性结果

3.4.2 定量对比

定量实验被实施来进一步验证所提方法的泛化

性能，结果如表 3 所示. 本文方法在 *EN, SCD, MI, SD* 这 4 个指标上取得了最好的结果，在 *SF* 上取得了第 2 好的结果. 对于 *VIF* 和 $Q^{AB/F}$，所提方法分别排行第 4 和第 3. 总的来说，本文所提方法在 RoadScene 数据集上的定量结果最好，这进一步说明了所提方法优良的泛化性.

Table 3　Quantitative Results of the Generalization Experiment

表 3　泛化实验的定量结果

融合方法	*VIF*↑	*EN*↑	*SCD*↑	*MI*↑	$Q^{AB/F}$↑	*SD*↑	*SF*↑
GTF	0.303±0.031	7.486±0.190	1.047±0.101	1.563±0.241	0.340±0.045	48.911±6.487	8.247±1.342
MDLatLRR	0.320±0.036	6.933±0.298	1.257±0.324	1.445±0.286	0.506±0.055	32.647±6.336	9.287±2.158
DenseFuse	0.329±0.048	7.283±0.245	1.669±0.218	1.503±0.280	0.534±0.042	43.337±6.869	11.228±2.197
FusionGAN	0.204±0.022	7.111±0.158	1.057±0.393	1.377±0.172	0.280±0.038	39.024±4.354	8.203±1.024
U2Fusion	**0.344±0.052**	7.249±0.263	1.546±0.236	1.293±0.259	**0.535±0.037**	40.279±7.032	**14.406±2.668**
本文方法	0.316±0.039	**7.575±0.185**	**1.726±0.135**	**1.641±0.303**	0.506±0.036	**54.533±6.577**	11.774±2.274

注：↑表示值越高越好，加粗表示最优结果，加下划线表示次优结果.

3.5　效率对比

运行效率是评估方法性能的重要依据之一，为此，统计各方法在 TNO, MFNet, RoadScene 数据集上的平均运行时间来比较运行效率，结果如表 4 所示. 本文所提方法在 3 个数据集上都取得了最快的平均运行速度，比对比方法快 5 倍以上.

Table 4　Mean of Running Time of Each Method on Three Datasets

表 4　各方法在 3 个数据集上的平均运行时间　s

融合方法	TNO	MFNet	RoadScene
GTF	5.302	3.259	1.644
MDLatLRR	35.569	28.052	15.188
DenseFuse	0.358	0.299	0.562
FusionGAN	0.360	0.196	0.403
U2Fusion	0.613	0.264	0.643
本文方法	**0.066**	**0.038**	**0.029**

注：加粗表示最优结果.

3.6　消融实验

在所提方法中，最终实现红外与可见光图像融合的框架包括编码器网络、特征融合网络以及译码器网络. 为了验证它们的有效性，相应的消融实验被实施.

3.6.1　特征融合网络分析

特征融合网络的作用是将中间特征的融合策略可学习化，从而使得融合特征同时符合红外与可见光 2 种模态特征的概率分布. 相较于现存方法使用的传统特征融合策略，所提的特征融合网络具有更强的性能. 为了验证这一点，将本文提出的用于特征提取和特征重建的编码器网络和译码器网络固定，中间特征融合规则分别用 Mean 策略、Max 策略、Addition 策略、l_1-norm 策略及所提特征融合网络，实验结果如图 10 所示.

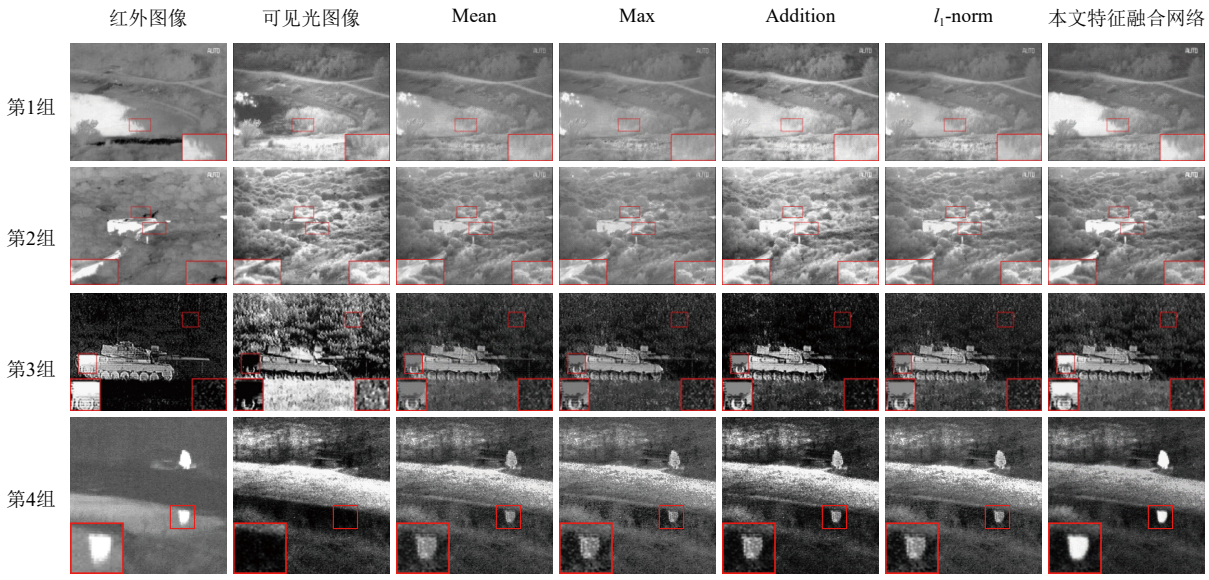


Fig. 10　Ablation experiment results of feature fusion network

图 10　特征融合网络的消融实验结果

首先, 由 Mean 策略、Max 策略以及 l_1-norm 策略得到的结果的对比度都非常差, 不仅显著目标被削弱, 场景纹理结构也不够自然. Addition 策略虽然比这些策略的性能更好, 但其无法自适应地选择性保留有益信息, 以至于某些区域出现过度饱和或亮度中和的现象. 比如, 在第 1 组结果中, 由 Addition 策略得到的结果没有充分保留水面的亮度; 在第 2 组结果中, 建筑物的亮度又是过度饱和的. 相反, 采用本文所提的特征融合网络可以自适应地保留红外与可见光源图像中的显著特征, 其不仅具有显著的对比度, 而且包含丰富的纹理细节. 这些结果可以说明所提特征融合网络的有效性.

3.6.2 自编码器网络分析

自编码器网络的作用是实现特征提取和图像重建, 其性能对融合结果的质量影响很大. 为了验证本文设计的自编码器网络的优势, 将其与经典的 DenseFuse[13] 进行对比. 具体地, 我们控制融合规则保持相同(Addition 策略以及 l_1-norm 策略), 使用本文提出的自编码器网络或 DenseFuse 中的自编码器网络来提取特征和重建图像, 实验结果如图 11 所示.

对于 Addition 策略, DenseFuse 结果中的显著目标被严重削弱, 而本文方法却可以很好地保持它们. 对于 l_1-norm 策略, DenseFuse 丢失了一些纹理结构, 如云朵和椅子, 而所提方法很好地保持了这些细节. 总的来说, 本文设计的自编码器网络的性能比 DenseFuse 中的更强, 其可以较好地实现特征提取和图像重建.

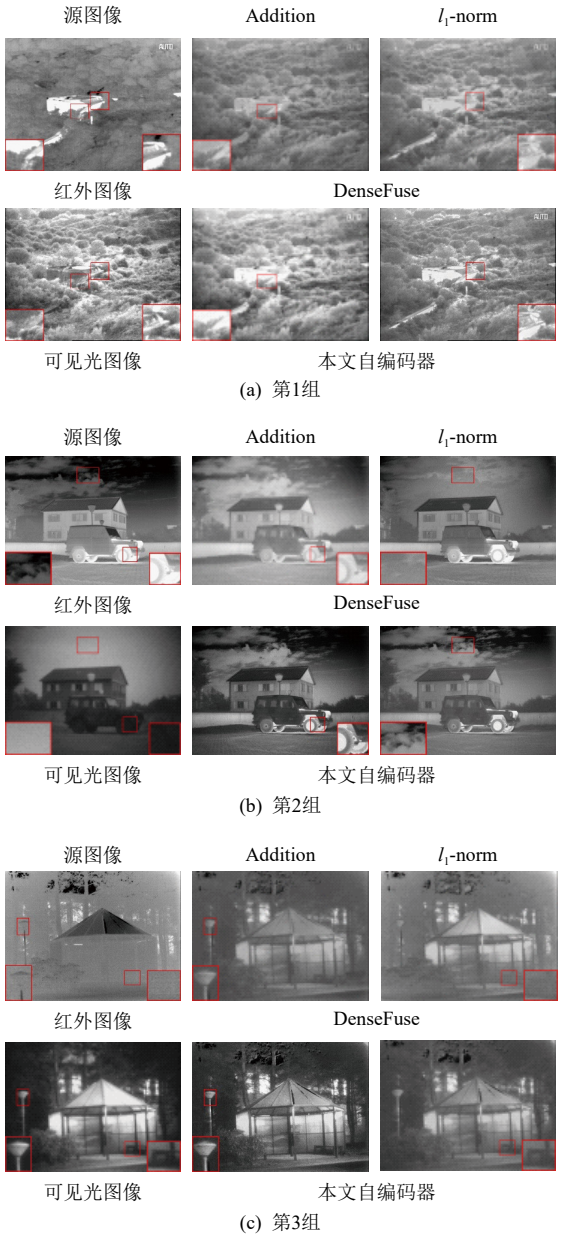


Fig. 11 Ablation experiment results of autoencoder

图 11 自编码器的消融实验结果

4 结 论

综上所述, 本文提出的基于特征空间多分类对抗机制的红外与可见光图像融合网络主要包括 2 个部分: 1) 引入空间注意力机制, 同时使用密集连接和残差连接来构建高性能编码器网络实现特征提取; 引入通道注意力机制, 同时使用多尺度卷积来构建高性能译码器网络实现图像重建. 2) 将融合策略网络化, 引入多分类生成对抗机制使融合特征同时符合红外与可见光 2 种模态的概率分布, 从而构建高性能的深度融合策略. 最终的红外与可见光图像融合网络由训练好的编码器、特征融合网络以及译码器级联而成, 生成高质量的融合结果. 相较于现存方法, 所提方法可解释性好, 能根据输入图像自适应生成更合理的融合策略, 从而在源图像中同时保留丰富的纹理细节和显著的对比度, 有效避免融合过程中有益信息的削弱问题. 广泛的实验表明, 所提方法与最新的融合方法 GTF, MDLatLRR, DenseFuse, FusionGAN, U2Fusion 相比, 融合性能更好. 同时, 本文方法的融合速度很快, 较对比方法快 5 倍以上.

作者贡献声明: 张浩提出方法思路, 设计及实施实验, 分析结果并撰写论文; 马佳义构思和设计研究方向, 对论文提出建设性的意见, 并对手稿进行修改; 樊凡协助实施实验, 并分析实验结果; 黄珺对论文提出建设性的意见, 并对稿件进行了修改; 马泳完善方法理论, 并对稿件进行了修改.

参 考 文 献

[1] Ma Jiayi, Ma Yong, Li Chang. Infrared and visible image fusion methods and applications: A survey[J]. Information Fusion, 2019, 45: 153−178

[2] Wang Xianghai, Wei Tingting, Zhou Zhiguang, et al. Research of remote sensing image fusion method based on the Contourlet coefficients' correlativity[J]. Journal of Computer Research and Development, 2013, 50(8): 1778−1786 (in Chinese)
(王相海, 魏婷婷, 周志光, 等. Contourlet四叉树系数方向相关性的遥感图像融合算法[J]. 计算机研究与发展, 2013, 50(8): 1778−1786)

[3] Chen Tao, Yi Mo, Liu Zhongxuan, et al. Image fusion at similar scale[J]. Journal of Computer Research and Development, 2005, 42(2): 2126−2131 (in Chinese)
(陈涛, 易沫, 刘忠轩, 等. 相似尺度图像融合算法[J]. 计算机研究与发展, 2005, 42(2): 2126−2131)

[4] Muller A C, Narayanan S. Cognitively-engineered multisensor image fusion for military applications[J]. Information Fusion, 2009, 10(2): 137−149

[5] Ma Jiayi, Zhang Hao, Shao Zhenfeng, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation, 2021, 70: 5005014

[6] Gao Xueqin, Liu Gang, Xiao Gang, et al. Fusion algorithm of infrared and visible images based on FPDE[J]. Acta Automatica Sinica, 2020, 46(4): 796−804 (in Chinese)
(高雪琴, 刘刚, 肖刚, 等. 基于FPDE的红外与可见光图像融合算法[J]. 自动化学报, 2020, 46(4): 796−804)

[7] Li Shutao, Yang Bin, Hu Jianwen. Performance comparison of different multi-resolution transforms for image fusion[J]. Information Fusion, 2011, 12(2): 74−84

[8] Lin Suzhen, Zhu Xiaohong, Wang Dongjuan, et al. Multi-band image fusion based on embedded multi-scale transform[J]. Journal of Computer Research and Development, 2015, 52(4): 952−959 (in Chinese)
(蔺素珍, 朱小红, 王栋娟, 等. 基于嵌入式多尺度变换的多波段图像融合[J]. 计算机研究与发展, 2015, 52(4): 952−959)

[9] Li Shutao, Yin Haitao, Fang Leyuan. Group-sparse representation with dictionary learning for medical image denoising and fusion[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(12): 3450−3459

[10] Kong Weiwei, Lei Yang, Zhao Huaixun. Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization[J]. Infrared Physics Technology, 2014, 67: 161−172

[11] Zhang Xiaoye, Ma Yong, Fan Fan, et al. Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition[J]. Journal of the Optical Society of America, 2017, 34(8): 1400−1410

[12] Ma Jiayi, Chen Chen, Li Chang, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. Information Fusion, 2016, 31: 100−109

[13] Zhou Huabing, Hou Jilei, Wu Wei, et al. Infrared and visible image fusion based on semantic segmentation[J]. Journal of Computer Research and Development, 2021, 58(2): 436−443 (in Chinese)
(周华兵, 侯积磊, 吴伟, 等. 基于语义分割的红外和可见光图像融合[J]. 计算机研究与发展, 2021, 58(2): 436−443)

[14] Zhang Hao, Xu Han, Xiao Yang, et al. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity[C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 12797 − 12804

[15] Xu Han, Ma Jiayi, Jiang Junjun, et al. U2Fusion: A unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2020, 44(1): 502−518

[16] Ma Jiayi, Yu Wei, Liang Pengwei, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11−26

[17] Ma Jiayi, Xu Han, Jiang Junjun, 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

[18] Li Hui, Wu Xiaojun. DenseFuse: A fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614−2623

[19] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C] //Proc of the 28th Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2672−2680

[20] Bao Fumin, Li Aiguo, Qin Zheng. Image fusion using SGNN[J]. Journal of Computer Research and Development, 2005, 42(3): 417−423 (in Chinese)
(鲍复民, 李爱国, 覃征. 基于SGNN的图像融合[J]. 计算机研究与发展, 2005, 42(3): 417−423)

[21] Long Yongzhi, Jia Haitao, Zhong Yida, et al. RXDNFuse: A aggregated residual dense network for infrared and visible image fusion[J]. Information Fusion, 2021, 69: 128−141

[22] Hou Ruichao, Zhou Dongming, Nie Rencan, et al. VIF-Net: An unsupervised framework for infrared and visible image fusion[J]. IEEE Transactions on Computational Imaging, 2020, 6: 640−651

[23] Li Jing, Huo Hongtao, Li Chang, et al. Multigrained attention network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation Measurement, 2021, 70: 5002412

[24] Li Jing, Huo Hongtao, Li Chang, et al. AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks[J]. IEEE Transactions on Multimedia, 2020, 23: 1383−1396

[25] Ma Jiayi, Liang Pengwei, Yu Wei, et al. Infrared and visible image fusion via detail preserving adversarial learning[J]. Information Fusion, 2020, 54: 85−98

[26] Wang Zhou, Bovik Alan C A. Universal image quality index[J]. IEEE Signal Processing Letters, 2002, 9(3): 81−84

[27] Li Hui, Wu Xiaojun, Durrani T. Nestfuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models[J]. IEEE Transactions on Instrumentation Measurement, 2020, 69(12): 9645−9656

[28] Jian Lihua, Yang Xiaomin, Liu Zheng, et al. SEDRFuse: A symmetric

encoder-decoder with residual block network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation Measurement, 2021, 70: 5002215

[29] Mao Xudong, Li Qing, Xie Haoran, et al. Least squares generative adversarial networks[C] //Proc of the 16th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2794-2802

[30] Woo S H, Park J C, Lee J Y, et al. CBAM: Convolutional block attention module[C] //Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 3-19

[31] Huang Gao, Liu Zhuang, Laurens V D M, et al. Densely connected convolutional networks[C] //Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4700-4708

[32] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C] //Proc of the 34th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778

[33] Li Hui, Wu Xiaojun, Josef K. MDLatLRR: A novel decomposition method for infrared and visible image fusion[J]. IEEE Transactions on Image Processing, 2020, 29: 4733-4746

[34] Alexander T, Hogervorst M A. Progress in color night vision[J]. Optical Engineering, 2012, 51(1): 010901

[35] Ha Q, Kohei W, Karasawa T, et al. MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes [C] //Proc of the 30th Int Conf on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2017: 5108-5115

[36] Xu Han, Ma Jiayi, Le Zhuliang, et al. FusionDN: A unified densely connected network for image fusion[C] //Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 12484-12491

[37] Jiang Xingyu, Ma Jiayi, Xiao Guobao, et al. A review of multimodal image matching: Methods and applications[J]. Information Fusion, 2021, 73: 22-71

[38] Hamid R S, Bovik A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444

[39] 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

[40] 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

[41] Qu Guihong, Zhang Dali, Yan Pingfan. Information measure for performance of image fusion[J]. Electronics Letters, 2002, 38(7): 313-315

[42] Piella G, Heijmans H. A new quality metric for image fusion[C] //Proc of the 10th IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2003: 173-176

[43] Rao Yunjiang. In-fibre Bragg grating sensors[J]. Measurement Science Technology, 1997, 8(4): 355-375

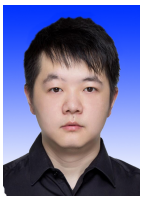
[44] Eskicioglu A M, Fisher P S. Image quality measures and their performance[J]. IEEE Transactions on Communications, 1995, 43(12): 2959-2965

Zhang Hao, born in 1996. PhD candidate. His main research interests include computer vision and machine learning.

张　浩, 1996 年生. 博士研究生. 主要研究方向为计算机视觉、机器学习.

Ma Jiayi, born in 1986. PhD, professor. His main research interests include computer vision, pattern recognition, and machine learning.

马佳义, 1986 年生. 博士, 教授. 主要研究方向为计算机视觉、模式识别、机器学习.



Fan Fan, born in 1989. PhD, associate professor. His main research interests include infrared thermal imaging and machine learning. (fanfan@whu.edu.cn)

樊　凡, 1989 年生. 博士, 副教授. 主要研究方向为红外热成像、机器学习.

Huang Jun, born in 1985. PhD, associate professor. His main research interests include infrared image processing and infrared spectrum processing. (junhwong@whu.edu.cn)

黄　珺, 1985 年生. 博士, 副教授. 主要研究方向为红外图像处理、红外光谱处理.

Ma Yong, born in 1971. PhD, professor. His main research interests include image processing and infrared remote sensing. (mayong@whu.edu.cn)

马　泳, 1971 年生. 博士, 教授. 主要研究方向为图像处理、红外遥感探测.