文章编号:1001-9014(2021)05-0696-13 DOI:10. 11972/j. issn. 1001-9014. 2021. 05. 017

基于边缘保持和注意力生成对抗网络的红外与可见光图像融合

朱雯青[1,2,3], 汤心溢[1,3*], 张 瑞[1,2,3], 陈 潇[1,2,3], 苗 壮[1,2,3]
(1. 中国科学院上海技术物理研究所,上海 200083;
2. 中国科学院大学,北京 100049;
3. 中国科学院红外探测与成像技术重点实验室,上海 200083)

摘要:由于红外与可见光图像特征差异大,并且不存在理想的融合图像监督网络学习源图像与融合图像之间的映射关系,深度学习在图像融合领域的应用受到了限制。针对此问题,提出了一个基于注意力机制和边缘损失函数的生成对抗网络框架,应用于红外与可见光图像融合。通过引入对抗训练和注意力机制的思想,将融合问题视为源图像和融合图像对抗的关系,并结合了通道注意力和空间注意力机制学习特征通道域和空间域的非线性关系,增强了显著性目标特征表达。同时提出了一种边缘损失函数,将源图像与融合图像像素之间的映射关系转化为边缘之间的映射关系。多个数据集的测试结果表明,该方法能有效融合红外目标和可见光纹理信息,锐化图像边缘,显著提高图像清晰度和对比度。

关 键 词:图像融合;生成对抗网络;边缘损失;注意力机制
中图分类号:TP391.41 **文献标识码**:A

Infrared and visible image fusion based on edge-preserving and attention generative adversarial network

ZHU Wen-Qing[1,2,3], TANG Xin-Yi[1,3*], ZHANG Rui[1,2,3], CHEN Xiao[1,2,3], MIAO Zhuang[1,2,3]
(1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. Key Laboratory of Infrared System Detection and Imaging Technology, Chinese Academy of Sciences, Shanghai 200083, China)

Abstract: Infrared and visible image features are quite different, and there are no ideal fused images supervise neural networks to learn the mapping relationship between the source images and the fused images. Thus, the application of deep learning is limited to the field of image fusion. To solve this problem, a generative adversarial network framework based on attention mechanism and edge loss is proposed, which is applied to the infrared and visible image fusion. Derived from the thoughts of attention mechanism and adversarial training, the fusion problem is regarded as an adversarial game between the source images and the fused images, and combining channel attention and spatial attention mechanism can learn nonlinear relationship between channel domain features and spatial domain features, which enhances the expression of salient target features. At the same time, an edge-based loss function is proposed, which converts the mapping relationship between the source image pixels and the fused image pixels into the mapping relationship between the edges. Experimental results on multiple datasets demonstrate that the proposed method can effectively fuse infrared target and visible texture information, sharpen image edges, and significantly improve image clarity and contrast.

Key words: image fusion, generative adversarial network, edge-based loss function, attention mechanism

收稿日期:2020- 10- 29,**修回日期**:2020- 12- 05 **Received date**:2020- 10- 29,**Revised date**:2020- 12- 05

作者简介(Biography):朱雯青(1994—),女,浙江杭州人,博士研究生,主要研究领域为计算机视觉与深度学习。E-mail: fyzhuwenqing@163.com

[*] **通讯作者**(Corresponding author): E-mail: gq227@mail. sitp. ac. cn

引言

近年来,红外与可见光图像融合技术一直是图像融合领域研究的热点。红外成像对拍摄的景物热辐射成像,能在恶劣天气、光照不足、强闪光等条件下发现目标,但红外图像具有边缘模糊、细节较少和不符合人眼视觉感受的缺点;可见光图像对地物反射成像,图像分辨率高、细节丰富,但容易受天气、光照的干扰。因此,该技术能够有效融合不同波段图像的信息,弥补单一波段成像的不足,提高图像生成质量,对目标检测、识别、跟踪和分割等任务有着重要的意义[1].

目前研究者已经提出多种融合方法,如小波变换[2]、非下采样 Contourlet 变换(NSCT)[3]、脉冲神经网络(PCNN)法[4]、引导滤波法[5]、稀疏表示法[6]等。近几年,深度学习在计算机视觉领域大放异彩,也被应用到图像融合领域中,算法根据融合过程的不同可分成分类式和端到端式。分类式融合方法是将图像融合问题视为分类问题[7],例如 Liu 等人提出的基于卷积网络的融合算法[8]先提取源图像对的特征,然后利用神经网络对特征进行分类得到信息权重图从而决定融合策略。Li 等人提出的算法先使用基于引导滤波的融合算法(GFF)[5]将源图像分解成基础部分和细节部分,基础部分用权重取平均的策略,细节部分用神经网络提取深层特征生成权重图,最后进行融合图像重建[9]。这些方法的缺点在于神经网络只进行融合策略的选择,需要用传统算法先对图像进行特征分解,算法复杂度较高。端到端式融合方法是将源图像对输入网络进行端到端训练后输出融合图像。Li 等人提出的算法(DenseFuse)将编解码器和残差块相结合实现红外与可见光图像融合[10]。Ma 等人首次提出基于生成对抗网络的红外与可见光图像融合算法(FusionGAN)[11],接着研究者提出了多种基于生成对抗网络的图像融合算法。Xu 等人提出的基于生成对抗网络的融合算法(LBP-BEGAN)设计了基于局部二值模式(LBP)的损失函数[12]。Li 等人提出的耦合生成对抗网络与相对鉴别器相结合的融合算法(RCGAN),利用了预生成的融合图像作为标签进行学习[13]。图像融合是一个无监督问题,没有理想的融合图像能作为真值(ground truth)实现监督学习,因此这些方法生成的图像效果主要依赖于单一源图像或者预生成的融合图像,不能有效提取多源图像信息并实现融合,导致图像细节缺失较为严重。

针对以上问题,本文提出了一种基于边缘损失和注意力机制的生成对抗网络框架(EAGAN),用于红外与可见光图像融合。首先算法将注意力机制和生成对抗网络相结合,实现了网络在提取特征的同时能学习通道域和空间域的上下文信息,提高了融合图像特征表达能力;其次提出了一种边缘损失函数,将融合问题转化为源图像对与融合图像边缘之间的映射关系,并结合了感知损失和对抗损失,为解决像素损失无法衡量源图像与融合图像映射关系的问题提供了一种新思路,同时解决了端到端式融合算法重建的图像细节缺失严重的问题,提高了融合图像生成质量。

1　基于边缘保持和注意力生成对抗网络的融合算法

1.1　生成对抗网络原理

生成对抗网络(Generative Adversarial Nets, GAN)是一种概率生成模型,它通过对抗训练的方式促使生成的样本分布服从真实数据分布。传统 GAN 模型由一个生成模型 G 和一个判别模型 D 组成,为了生成模型 G 能够学习真实数据 x 的分布 p_g,先定义一个噪声变量 $p_z(z)$ 通过生成模型 G 映射到数据空间 $G\left(z,\theta_g\right)$;判别模型 $D\left(z,\theta_d\right)$ 是一个二值分类网络,用于辨别生成模型 G 产生的数据是否来自真实数据 P_g。在模型训练过程中,生成模型尽可能生成真实的样本试图“欺骗”判别模型,判别模型尽可能辨别出假样本,两者相互博弈对抗。换言之,生成对抗网络是一个极小极大优化问题,最优值是一个鞍点,生成模型在该鞍点上达到最小值,判别模型则达到最大值。传统生成对抗网络的目标函数 $V(D,G)$[14]如式(1)所示:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right], \tag{1}$$

其中,p_{data} 表示真实训练数据的分布,$p_z(z)$ 表示输入的噪声变量,$D(x)$ 表示判别模型辨别样本来自真实样本的概率,$\log D(x)$ 表示 $[\,1\quad 0\,]^{\mathrm{T}}$ 和 $[\,D(x)\quad 1-D(x)\,]^{\mathrm{T}}$ 的交叉熵,$\log(1-D(G(z)))$ 表示 $[\,0\quad 1\,]^{\mathrm{T}}$ 和 $[\,D\left(G(z)\right)\quad 1-D(G(z))\,]^{\mathrm{T}}$ 的交叉熵。

然而在传统 GAN 模型训练初期,生成模型学习性能很差,生成的数据分布与真实的数据分布差异较大,使得判别模型辨别出生成数据的置信度很高,导致式(1)的目标函数反向传播给生成模型的

梯度很小，产生梯度消失的问题。本文引入的最小二乘生成对抗网络（Least Squares Generative Adversarial Nets，LSGAN）[15]在传统GAN模型的基础上提出了最小二乘损失函数，判别模型D产生的新决策边界会惩罚远离决策边界的生成样本从而在训练过程中为生成模型G提供更大梯度，克服了原始GAN模型梯度消失的问题，LSGAN目标函数如式（2）所示：

$$\min_D V_{LSGAN}(D) = E_{x\sim p_{data}(x)}\left[\left(D(x)-b\right)^2\right] + E_{z\sim p_z(z)}\left[\left(D\left(G(z)\right)-a\right)^2\right]$$

$$\min_G V_{LSGAN}(G) = E_{z\sim p_z(z)}\left[\left(D\left(G(z)\right)-c\right)^2\right], \quad (2)$$

其中a、b分别表示生成样本和真实样本的标签，c表示判别模型D辨别为生成样本的边界值。

1.2 算法框架

基于深度学习的红外与可见光融合算法的关键在于能够有效提取源图像对的特征，并通过合理的融合策略完成融合图像的重建，提升图像表达能力。本文提出的基于边缘保持和注意力生成对抗网络的红外与可见光融合算法（EAGAN）将融合问题转化为源图像对和融合图像对抗的过程，算法网络结构如图1所示。生成模型G是一个基于注意力机制的卷积神经网络，由特征提取和特征重建两个部分组成。特征提取部分网络由N个注意力模块组成，注意力模块包括了卷积层、Leaky ReLU激活层、通道注意力模块和空间注意力模块，能够有效学习深层特征分布。随着网络深度加深，特征越来越抽象，细节高频信息也越来越少。为了防止梯度消失和细节模糊的问题，在特征重建阶段先利用长连接将浅层特征和深层特征在通道维度级联，同时使用通道注意力模块学习浅层和深层特征之间的非线性关系，再用卷积层和Tanh激活函数将特征映射到像素灰度值空间。为了更好地保留图像细节，生成模型中没有使用批归一化层加速参数的迭代。判别模型是基于VGGNet-16[16]改进的二值化分类网络，本文将可见光图像和生成的样本图像作为判别模型的输入，使得重建的图像具有更多的可见光细节纹理，更符合人眼的视觉感受。

在模型训练阶段，先将可见光图像I_v和红外图像I_r在通道维度进行合并后输入生成模型G中输出融合图像I_f，在设计的注意力模块（见1.3节）和损失函数（见1.4节）作用下，生成模型G提取源图像对的结构信息和边缘信息。然后将融合图像I_f和可见光图像I_v输入判别模型D，让判别模型从样本中辨别出融合图像。生成模型和判别模型交替训练，从而驱使生成模型生成的图像具有红外图像的显著性目标和可见光图像的纹理细节。在测试推理阶段，只需要将可见光图像I_v和红外图像I_r在通道维度合并后输入生成模型G得到融合图像I_f即可。

1.3 注意力模块

融合算法的关键在于如何提取红外与可见光图像中的互补信息，目前已经提出的基于深度学习的红外与可见光图像融合算法都将通道域、空间域的特征重要性同等对待，导致网络不能聚焦于显著

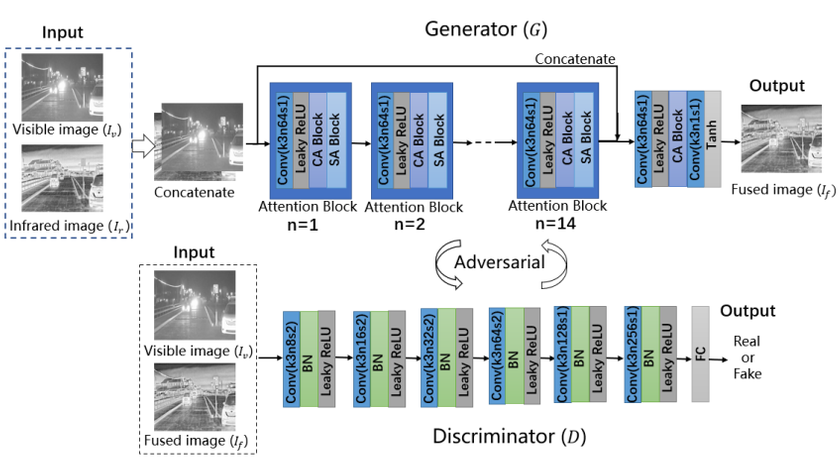

图1 EAGAN网络结构示意图，CA Block为通道注意力模块，SA Block为空间注意力模块，BN为批归一化，FC为全连接层，k为卷积核尺寸，n为卷积核数量，s为步长

Fig. 1 Architecture of the proposed EAGAN. CA Block: channel attention block, SA Block: spatial attention block, BN: batch normalization, FC: fully connected layer, Conv: corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer

性区域的特征提取与融合。红外图像通常有明显的目标边缘信息,可见光图像具有丰富的细节纹理等高频信息,两者的特征差异较大却互为补充;其次,卷积层的每个卷积只能捕捉到局部信息,感受野的大小依赖于卷积的大小,单用卷积难以捕捉到感受野外的上下文信息。而注意力机制不仅可以用来选择聚焦位置,还能增强该位置上目标的不同表示,提升该区域的表达能力。本文将通道注意力[17]和空间注意力[18]组成注意力模块引入生成模型中,提高了网络显著性特征的传递。生成模型网络包含了N个注意力模块,每个注意力模块由卷积层、激活层、通道注意力模块和空间注意力模块级联而成,能够学习到通道域、空间域之间的上下文信息,模块结构如图2所示。令F_{n-1}和F_n分别表示第n个注意力模块的输入和输出,H_n表示卷积和Leaky ReLU激活过程,H_c表示通道注意力模块,H_s表示空间注意力模块,则注意力模块F_n计算过程如式(3)所示:

$$F_n = H_s\left(H_c\left(H_n\left(F_{n-1}\right)\right)\right). \tag{3}$$

通道注意力能对不同通道特征的重要性进行校正,增强重要特征并抑制无关特征,同时实现了可见光和红外图像特征的选择。将输入的特征$X \in R^{H\times W\times C}$用卷积层映射到空间$U \in R^{H\times W\times C}$,接着用全局平均池化(Global Average Pooling)得到通道维度的全局统计信息$Z \in R^{1\times 1\times C}$,为了更好地表达通道域的非线性关系,然后用$1\times 1$卷积和ReLU[19]激活函数对$Z$进行尺度因子为$r$的卷积映射得到大小为$1\times 1\times \frac{C}{r}$的通道特征$Z_1$,再用$1\times 1$卷积和Sigmoid激活函数将通道特征$Z_1$重采样后映射到$[0,1]$区间得到大小为$1\times 1\times C$的通道描述子$S_c$,利用门控机制学习通道域的上下文信息。令输入的特征图为$X=[x_1,x_2,\cdots,x_C]$,高度为H,宽度为W,通道数为C。$V=[v_1,v_2,\cdots,v_C]$表示数量为C的卷积核,$B=[b_1,b_2,\cdots,b_C]$表示偏置,则$U=\left[u_1,u_2,\cdots,u_C\right]$可表示如下:

$$u_c = v_c * X + b_c\ , \tag{4}$$

其中,u_c、v_c、b_c分别表示第c个通道的特征响应值、二维卷积核和偏置。每个通道的特征统计值是通过计算每个通道空间特征的平均值得到的,第c层通道特征图的统计值z_c表示如式(5)所示:

$$z_c = H_{GAP}\left(u_c\right) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}u_c\left(i,j\right), \tag{5}$$

其中,$x_c\left(i,j\right)$表示第c个通道特征图x_c中位置坐标为$\left(i,j\right)$的特征元素,$H_{GAP}(\cdot)$表示全局平均池化。接着,用两次卷积和激活函数学习通道维度的全局统计特征的非线性相互作用,通道描述子S_c计算公式如式(6)所示:

$$S_c = F\left(z,W,B\right) = \sigma\left(W_2 * \left(\delta\left(W_1 * z + B_1\right)\right) + B_2\right), \tag{6}$$

其中,δ表示ReLU激活函数,σ表示sigmoid激活函数,$W_1 \in R^{\frac{C}{r}\times C}$、$W_2 \in R^{C\times\frac{C}{r}}$表示卷积核,$B_1$、$B_2$表示偏置向量,$*$表示卷积。为了防止网络深度增加出现梯度消失现象,使用了残差结构将输入的特征传递到深层[20],注意力模块的输出X_{CA}如式(7)所示:

$$X_{CA} = X + F_{CA}\left(u_c,S_c\right) = X + S_c u_c\ . \tag{7}$$

与通道注意力机制类似,空间注意力是通过学习空间域的上下文信息从而找出空间显著性目标,增强网络对空间显著性区域的表达能力。为了计算空间注意力,本文将输入的特征X_{CA}沿着通道维度用全局平均池化(Global average pooling)和全局最大池化(Global max pooling)分别得到空间特征图的平均值和最大值,并将两个特征统计图在通道维度级联,然后用卷积计算空间注意力特征图,用Sig-

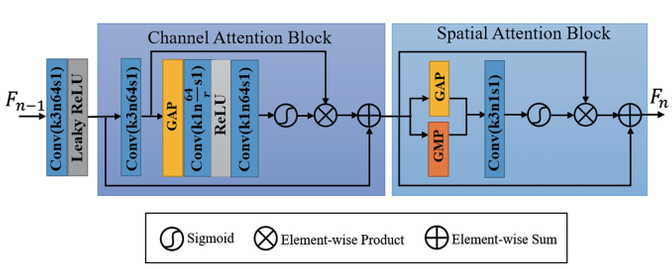

图2　注意力模块结构示意图,GAP为全局平均池化,GMP为全局最大池化,r为尺度因子,k为卷积核尺寸,n为卷积核数量,s为步长

Fig. 2　Architecture of Attention Block. GAP: Global Average Pooling, GMP: Global Max Pooling, r: scaling factor, Conv: corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer

moid激活函数将空间注意力特征映射到[0,1]区间得到空间注意力描述子[18]，同时利用残差结构将输入的特征馈送到输出。空间注意力模块X_{SA}的计算过程如式(8)所示：

$$X_{SA}=X_{CA}+\sigma(W_S*\left[H_{GAP}\left(X_{CA}\right);H_{GMP}\left(X_{CA}\right)\right])\times X_{CA}\ , \tag{8}$$

其中，输入的特征为X_{CA}，H_{GAP}表示全局平均池化，H_{GMP}表示全局最大池化，W_S表示卷积核，*表示卷积，σ表示Sigmoid激活函数。

1.4 损失函数设计

融合问题本质上是一个无监督问题，不存在理想的融合图像作为标签监督模型训练。因此本文提出了一种边缘损失函数，将源图像对与融合图像像素灰度值之间的映射关系转化成图像边缘特征之间的关系，并结合了感知损失与对抗损失函数，摆脱了对像素损失的依赖。损失函数分为生成模型G损失函数$\mathcal{L}_G$和判别模型D损失函数$\mathcal{L}_{EAGAN}(D)$两个部分，生成模型的损失函数$\mathcal{L}_G$由对抗损失$\mathcal{L}_{EAGAN}(G)$、边缘损失$\mathcal{L}_{edge}$、感知损失$\mathcal{L}_{perceptual}$组成，λ_1、λ_2为权重系数，如式(9)所示。

$$\mathcal{L}_G=\mathcal{L}_{EAGAN}(G)+\lambda_1\mathcal{L}_{perceptual}+\lambda_2\mathcal{L}_{edge}\ , \tag{9}$$

$\mathcal{L}_{EAGAN}(G)$是生成模型和判别模型对抗的损失函数，如式(10)所示：

$$\mathcal{L}_{EAGAN}(G)=\frac{1}{N}\sum_{n=1}^{N}(D\left(I_f^n\right)-c)^2\ , \tag{10}$$

其中，I_f^n表示融合图像，c为判别模型辨别输入图像是假样本的阈值，N表示输入判别模型的图像数量。

因不存在理想的融合图像作为标签监督网络的学习，本文提出的边缘损失函数使用特征描述算子将红外图像、可见光图像和融合图像从像素空间映射到浅层梯度空间，通过计算融合图像与源图像间浅层边缘特征距离，学习红外与可见光图像的边缘特征分布，实现融合图像同时保留红外目标边缘信息和可见光细节纹理信息。本文用拉普拉斯算子提取源图像的边缘信息，并计算融合图像梯度与红外与可见光图像梯度之间的Manhattan距离，边缘损失函数$\mathcal{L}_{edge}$表示如式(11)所示：

$$\mathcal{L}_{edge}=\frac{1}{HW}\left\|\nabla I_f-\nabla I_v-\nabla I_r\right\|_1\ , \tag{11}$$

其中，H、W分别表示输入图像的高度、宽度，I_f表示融合图像，I_v表示可见光图像，I_r表示红外图像，∇表示拉普拉斯算子，$\|\cdot\|_1$表示L_1范数。

像素损失只能表示图像浅层的像素信息差异，而感知损失[21]能反映图像间深层的特征信息差异。本文中的感知损失是将VGG-16预训练模型作为特征提取器，在模型训练阶段将红外图像、可见光图像和生成的融合图像输入VGG-16网络中，分别计算网络前4层的特征图，接着得到特征图的Gram矩阵，再分别计算红外图像、可见光图像与融合图像对应层之间的欧氏距离。Gram矩阵计算如式(12)所示，其中$\phi_n(x)_{h,w,c}$为计算得到的特征图，计算时需要将$C_n\times H_n\times W_n$的特征图尺寸转换为$C_n\times\left(H_nW_n\right)$。

$$G_n^{\phi}(x)=\phi_n(x)_{h,w,c}\phi_n(x)_{h,w,c}\ . \tag{12}$$

感知损失$\mathcal{L}_{perceptual}$如式(13)所示：

$$\mathcal{L}_{perceptual}=\frac{1}{C_nH_nW_n}(\left\|G_n^{\phi}\left(vgg_n\left(I_f\right)\right)-G_n^{\phi}\left(vgg_n\left(I_v\right)\right)\right\|_{\mathrm{F}}^2+\left\|G_n^{\phi}\left(vgg_n\left(I_f\right)\right)-G_n^{\phi}\left(vgg_n\left(I_r\right)\right)\right\|_{\mathrm{F}}^2)\ , \tag{13}$$

式(13)中，I_f表示融合图像，I_v表示可见光图像，I_r表示红外图像。vgg_n表示第n层特征图，G_n^{ϕ}表示Gram矩阵计算，C_n、H_n、W_n分别表示第n层特征图的通道数、高度和宽度，$\|\cdot\|_F^2$表示平方Frobenius范数。

判别模型负责辨别输入的图像是融合图像还是可见光图像，判别模型损失函数$\mathcal{L}_{EAGAN}(D)$如式(14)所示：

$$\mathcal{L}_{EAGAN}(D)=\frac{1}{N}\sum_{n=1}^{N}((D\left(I_f^n\right)-a^n)^2+(D\left(I_v^n\right)-b^n)^2)\ , \tag{14}$$

其中，I_f^n、I_v^n分别表示融合图像和可见光图像，a^n、b^n分别表示融合图像和可见光图像的标签。

2 实验结果分析

2.1 实验数据及细节

目前，自动驾驶场景下红外与可见光图像数据集较少，本文选择了RoadScene[22]数据集作为训练集，并从RoadScene数据集、TNO数据集和INO数据集中选取部分图像进行了模型泛化性能测试。RoadScene数据集是从FLIR数据集(FREE FLIR Thermal Dataset for Algorithm Training)中选择了221对红外与可见光图像，不仅包含了街景、行人、车辆、建筑、交通标志等，还包括了日间过曝光、夜间会车眩光等场景，而且对红外和可见光图像进行了配准。其中，本文选择了200对图像作为训练集(包含94对日间和106对夜间图像)进行训练，另外21对图像(包含10对日间和11对夜间图像)作为测试

集验证算法的性能。此外,从TNO和INO数据集中分别选择了25和10对不同场景的图像进行泛化性能测试。

200对图像不足以进行神经网络的训练,因此本文通过裁剪图像的方法增加数据的多样性。本文将每对红外与可见光图像间隔14个像素裁剪成64 × 64的图像块,总图像块数量为113833,大大扩充了训练集数据,防止网络过拟合。本文先将可见光图像转换成灰度图,再将可见光图像和红外图像的像素值归一化至[- 1,1]输入生成模型,经过运算输出64 × 64的融合图像,然后将融合图像和可见光图像输入判别模型判断样本来源。算法训练环境为NVIDIA 1080 Ti GPU,除了在通道注意力模块中通道放大和缩小时使用了1 × 1卷积,其余卷积尺寸为3 × 3,优化过程选取参数为$\beta_1 = 0.9$,$\beta_2 = 0.999$的Adam[23]算法,epoch为20,迭代总次数为71140,批数量为32,学习率设为$1e^{-4}$,损失函数的权重系数λ_1、λ_2分别设置为1、10,空间注意力模块中的采样因子r为4,注意力模块数N为14。融合图像边界值a、可见光图像边界值b和阈值c是软标签,不是确定的数值,a为[0,0.3]区间内的随机数,b为[0.7,1.2]区间内的随机数,c为[0.7,1,2]区间内的随机数。

本文算法与引导滤波融合算法(GFF)[5]、自适应稀疏表示融合算法(ASR)[6]、梯度转换融合算法(GTF)[24]、基于卷积网络的融合算法(DenseFuse)[10]、基于对抗生成网络的融合算法(FusionGAN)[11]和鉴别器的耦合生成对抗网络融合算法(RCGAN)[13]这6种经典算法进行了比较,实验过程中使用了该6种算法论文中的参数。

图像质量评价一般可分为主观评价和客观评价,主观评价虽符合人眼视觉感受,但难以用指标量化,因此需要用客观评价指标对评价效果进行补充。本文选择了信息熵(entropy,EN)[25]、差异相关性总量(The sum of the correlations of differences,SCD)[26]、空间频率(spatial frequency,SF)[27]和边缘强度(edge intensity,EI)[28]4个评价指标对融合图像进行图像质量评价。

2.2　实验结果

2.2.1　RoadScene测试集结果

本文从RoadScene测试集中选取了4对典型的红外与可见光图像(2对夜间图像和2对日间图像)进行主观评价结果展示,如图3所示。从细节上观察行人、车辆、路灯、建筑等,GFF、GTF、FusionGAN和RCGAN算法不能正确反映天空背景等真实场景信息,GFF算法在日间过度曝光场景下不能突出红外目标特性,GTF 和FusionGAN算法不能有效表达红外目标并且提取的可见光细节较少,ASR、Densefuse和RCGAN算法生成的图像边缘较为模糊并且图像对比度较低。本文算法可以突出红外图像的显著性目标,如有效融合了会车眩光和过度曝光区域的车辆、行人等目标;其次,本文算法可以保留可见光图像中丰富的纹理细节,如人行横道、路灯和树木等,并且可以反映真实场景信息,利于实现自动驾驶场景下的目标识别、图像分割等任务。

此外,对RoadScene测试集21对图像进行测试,用图像质量评价指标计算平均值评价融合算法性能,结果如表1所示。该算法在EN、SF和EI三项指标上远超其它算法,说明本文算法生成的融合图像信息量大、清晰度高、细节丰富。在SCD指标上表现与DenseFuse接近,说明该算法对融合源图像中的信息转化率较高。该算法生成的融合图像边缘清晰可辨、对比度高,视觉效果较好。

2.2.2　TNO数据集测试结果

为了验证该算法模型的泛化性能,在TNO数据集中选择Duine sequence、Nato_camp_sequence、Kaptein_1123、men in front of house和soldier_behind_smoke_3场景下的5对红外与可见光图像进行测试,结果如图4所示。综合观察,ASR、GTF算法生成的图像较为平滑,GFF算法在烟雾场景下不能有效融合红外目标,FusionGAN、RCGAN算法融合的高频信息较少,DenseFuse算法生成的图像红外目标边缘较模糊。而该算法模型生成的图像目标轮廓明显,保留了可见光图像的细节纹理,同时增强了图像的对比度,更能适应多种复杂场景。此外,选取了25对不同场景下的红外与可见光图像进行图像质量评价,评价指标结果对比如表2所示。从表2可得,本文算法模型在EN、SF和EI指标上表现最优,在SCD指标上与DenseFuse较为接近。测试结果表明,该算法能有效提取源图像中的显著目标,同时能保留较多的纹理信息,在TNO数据集上迁移效果较好。

2.2.3　INO数据集测试结果

本文从INO数据集ParkingSnow、GroupFight、MultipleDeposit和ClosePerson场景下选取了4对典型图像进行主观成像质量对比,如图5所示。同时

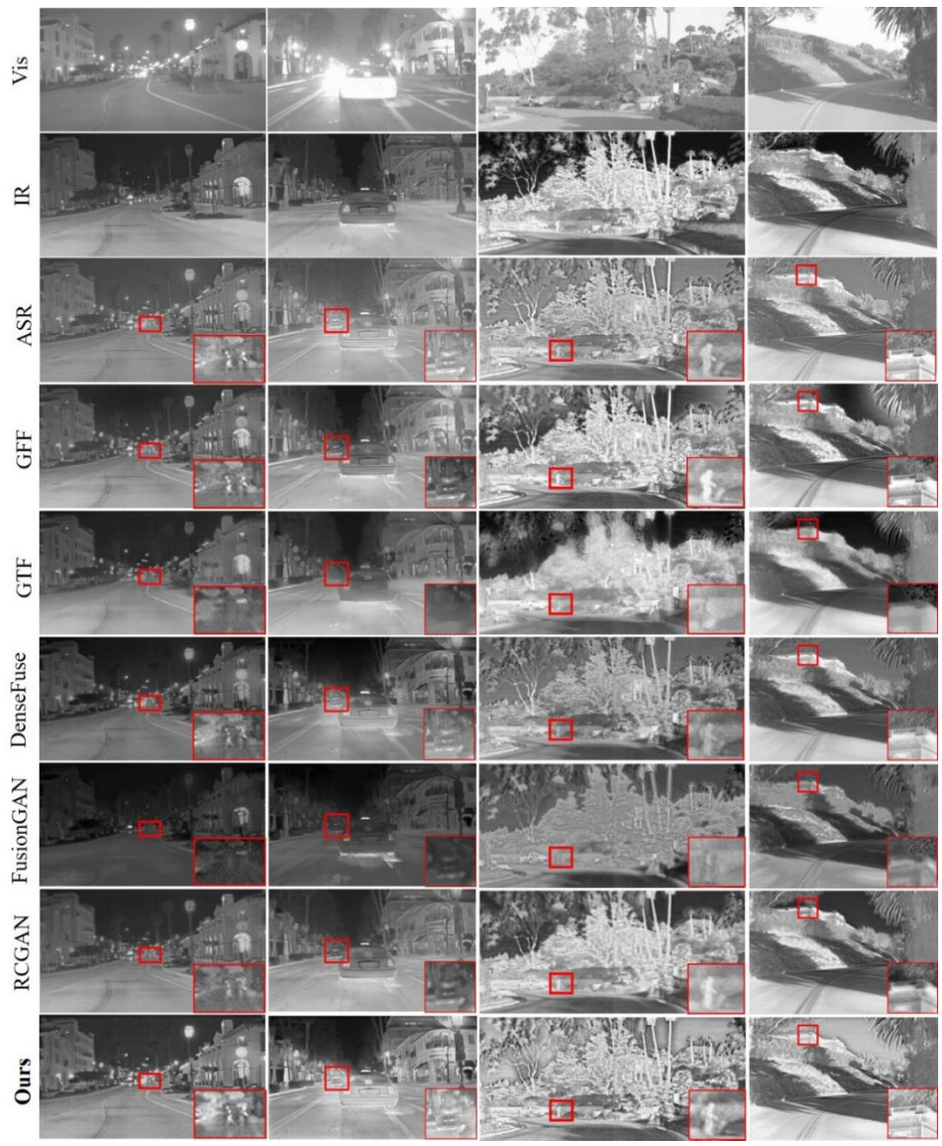

图3 RoadScene测试集4对典型红外与可见光图像不同算法融合结果对比

Fig. 3 Qualitative comparison of different algorithms on 4 typical infrared and visible image pairs. From top to bottom: visible image, infrared image, fusion results of ASR, GFF, GTF, DenseFuse, FusionGAN, RCGAN and our algorithm.

表1 RoadScene测试集上不同算法融合图像的评价指标对比

Table 1 Quantitative comparison of different algorithms on RoadScene dataset

	ASR	GFF	GTF	DenseFuse	FusionGAN	RCGAN	**Ours**
EN	6.92	7.26	7.26	7.24	6.84	7.14	**7.30**
SCD	1.27	1.23	0.98	**1.71**	0.86	1.19	1.54
SF	13.11	13.60	9.12	11.97	9.35	9.50	**15.62**
EI	0.22	0.22	0.16	0.20	0.16	0.18	**0.27**

从10个不同场景中选择了10对典型的红外与可见光图像对本文模型和其他六个算法进行客观质量评价，结果如表3所示。图5中红外图像中行人、车辆等目标轮廓突出，但是树木、草坪、建筑等纹理较少；与之相反，可见光图像包含了丰富的纹理信息，但行人目标受光照影响不容易被人眼发现。其他6种算法在图像融合过程中，容易因为红外与可见光特征差异较大导致目标边缘模糊，而该算法生成的融合图像边缘最为清晰，能保持红外图像目标的热辐射分布和源图像的对比度。该算法依旧在EN、SF和EI指标上表现最佳，SCD指标与DenseFuse算法接近，说明该算法模型鲁棒性好，能适应多种复杂场景。

2.2.4 注意力机制分析

为了直观地说明注意力模块的作用效果，本文

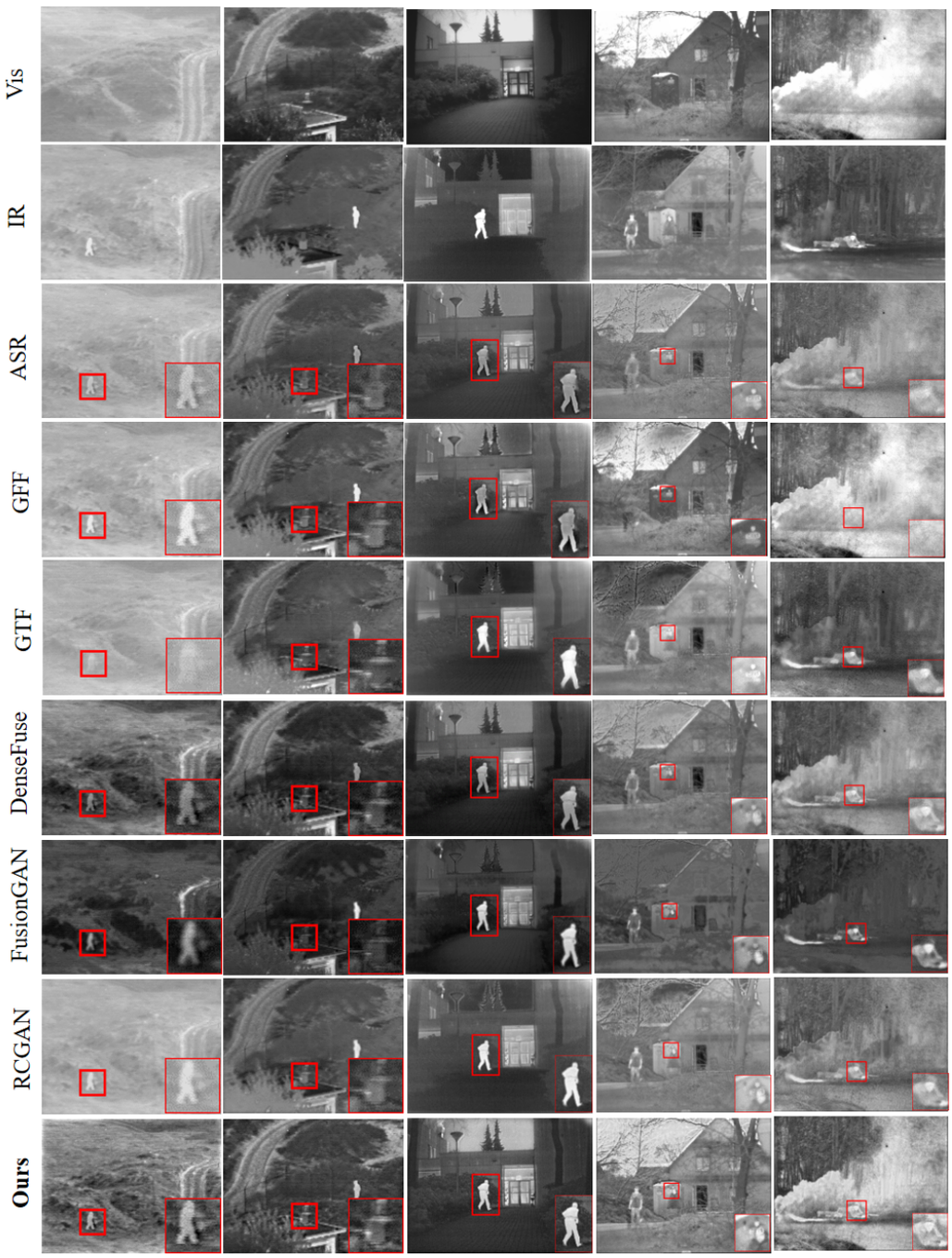

图 4　TNO 数据集 5 对典型红外与可见光图像融合结果对比(从左向右图像来源为 Duine sequence、Nato_camp_sequence、Kaptein_1123、men in front of house 和 soldier_behind_smoke_3)

Fig. 4　Qualitative comparison of different algorithms on 5 typical infrared and visible image pairs from TNO dataset. From left to right: Duine sequence, Nato_camp_sequence, Kaptein_1123, men in front of house and soldier_behind_smoke_3. From top to bottom: visible image, infrared image, fusion results of ASR, GFF, GTF, DenseFuse, FusionGAN, RCGAN and our algorithm.

对网络的通道注意力权重和空间注意力权重可视化,如图 6 所示。图 6(a)、(b)为模型输入的红外与可见光图像,图 6(c)为本文模型生成的融合图像;图 6(d)为第 3 个注意力模块输出的结果图,分别对应 64 个通道;图 6(e)显示了大小为 14 × 64 的通道注意力权重图,其中 14 行分别对应为第 1 至 14 个注意力模块中通道方向上的特征图权重,表明不同通道的特征自适应地学习到了不同的权重;图 6(f)为第 4 个注意力模块的空间注意力权重图,在车辆、行人等目标具有较大的权重,道路和天空等背景区域权重较小,说明网络更加关注显著性目标区域。

为了验证注意力机制对该方法的作用,本文在

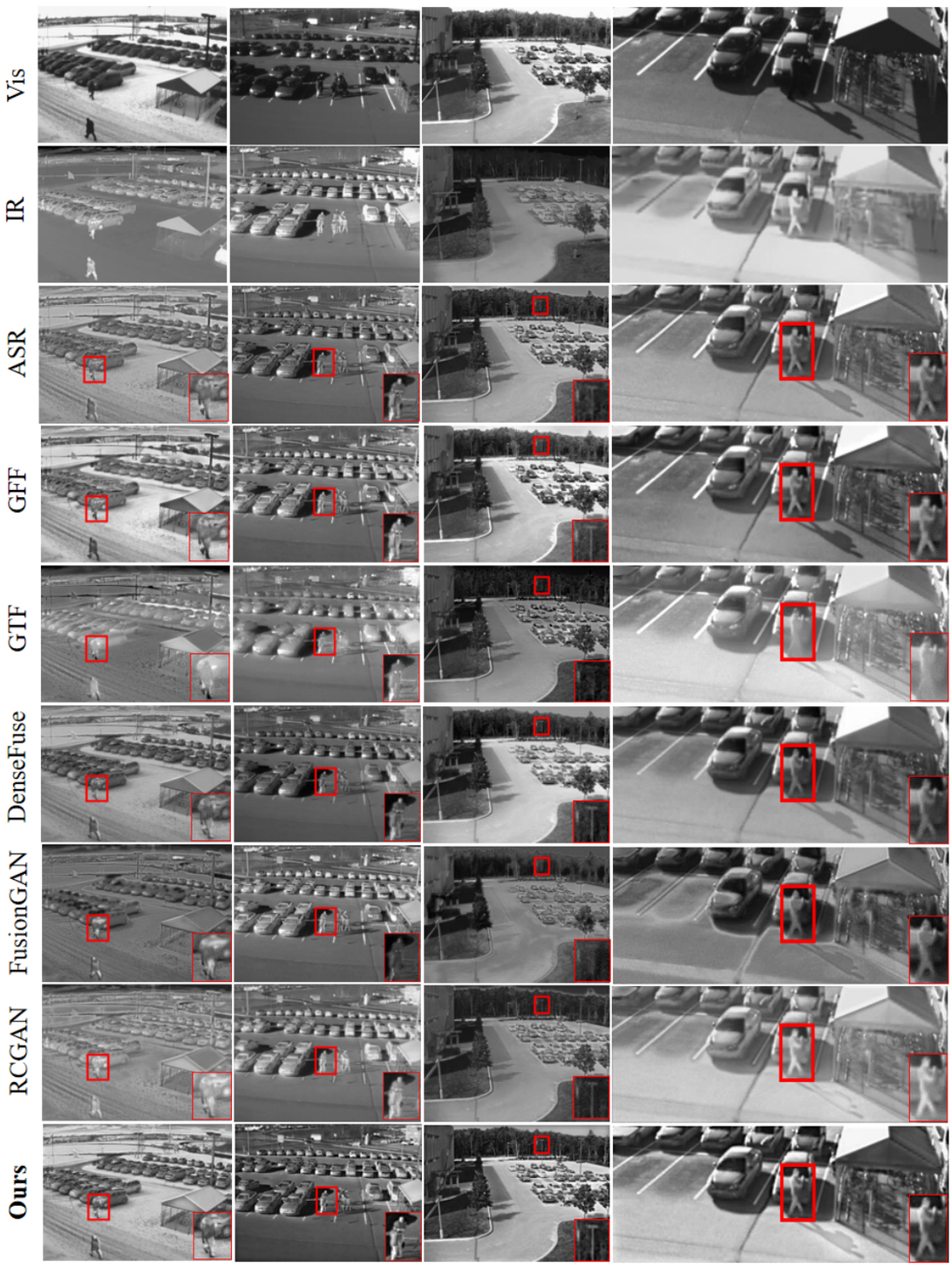

图5 INO数据集4对典型红外与可见光图像融合结果对比(从左向右图像来源为ParkingSnow、GroupFight、MultipleDeposit、ClosePerson)

Fig. 5 Qualitative comparison of different algorithms on 4 typical infrared and visible image pairs from INO dataset. From left to right: ParkingSnow, GroupFight, MultipleDeposit, ClosePerson. From top to bottom: visible image, infrared image, fusion results of ASR, GFF, GTF, DenseFuse, FusionGAN, RCGAN and our algorithm.

损失函数和训练参数不变情况下只去除了注意力模块中的通道注意力模块和空间注意力模块,在RoadScene训练集上进行训练,在TNO数据bench场景下的测试结果如图7所示。图7(a)、(b)对比可得,有注意力机制的模型生成的融合图像中的红外目标显著性高于无注意力机制的模型。此外,将训

表 2　TNO 数据集上不同算法融合结果指标对比

Table 2　Quantitative comparison of different algorithms on TNO dataset

	ASR	GFF	GTF	DenseFuse	FusionGAN	RCGAN	**Ours**
EN	6.44	6.84	6.93	6.87	6.35	6.77	**7.08**
SCD	1.61	1.36	0.97	**1.79**	1.30	1.41	1.67
SF	8.93	9.55	8.31	8.54	6.59	7.41	**11.59**
EI	0.13	0.14	0.13	0.14	0.11	0.13	**0.19**

表 3　INO 数据集上不同算法融合结果指标对比

Table 3　Quantitative comparison of different algorithms on INO dataset

	ASR	GFF	GTF	DenseFuse	FusionGAN	RCGAN	**Ours**
EN	6.94	7.14	7.02	7.09	6.62	6.97	**7.23**
SCD	1.40	1.29	1.03	**1.69**	1.02	1.18	1.53
SF	16.80	17.33	14.72	14.34	12.71	13.12	**19.40**
EI	0.25	0.26	0.21	0.22	0.19	0.21	**0.30**

练好的模型在 RoadScene、TNO 测试集上和该方法进行对比,客观评价指标结果如表 4 所示。有注意力机制的模型在 EN 和 SCD 指标上表现好于无注意力机制的模型,SF 和 EI 指标上两者接近,综合来看,注意力机制能增加融合图像的信息量,提高源图像的信息转化量。

2.2.5　生成模型损失函数分析

为了验证生成模型损失函数中对抗损失 $\mathcal{L}_{EAGAN}(G)$、边缘损失 $\mathcal{L}_{edge}$ 和感知损失 $\mathcal{L}_{perceptual}$ 对本文方法的重要性,本文在网络结构和其余参数不变的情况下通过改变 $\mathcal{L}_G$ 损失函数进行了对比实验:(a) $\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual}$,只有感知损失;(b) $\mathcal{L}_G = \lambda_2\mathcal{L}_{edge}$,只有边缘损失;(c) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G)$,只有对抗损失,因为缺少其他条件限制,为防止出现模式崩坏,在生成模型每个注意力模块的第一层卷积层后增加了批归一化层;(d) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_1\mathcal{L}_{perceptual}$,不含边缘损失;(e) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_2\mathcal{L}_{edge}$,不含感知损失;(f) $\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual} + \lambda_2\mathcal{L}_{edge}$,不含对抗损失。该 6 个模型在相同的 RoadScene 训练集上训练,并与本文方法进行了对比,在 TNO 数据集 Kaptein_1123 图像对上的测试结果如图 8 所示,图 8 红框内图像是对比区域的放大显示。从图中可得,只有对抗损失的模型能学习到低频信息,缺点在于伪影较多、细节较少,同时损失函数中含有对抗损失能使生成的图像更符合人眼的视觉感受,没有对抗损失的融合图像会在红外目标周围出现不必要的梯度变化;感知损失能学习到源图像对中的基本信息,更符合可见光图像的分布,没有感知损失的模型降低了融合图像的对比度;边缘损失能学习到红外显著性目标和可见光纹理细节,训练过程中不使用边缘损失的模型生成的图像非常平滑,红外目标边缘非常模糊,高频纹理信息较少。结合了对抗损失、边缘损失和感知损失,该方法能保持图像边缘并融合丰富的细节纹理信息。

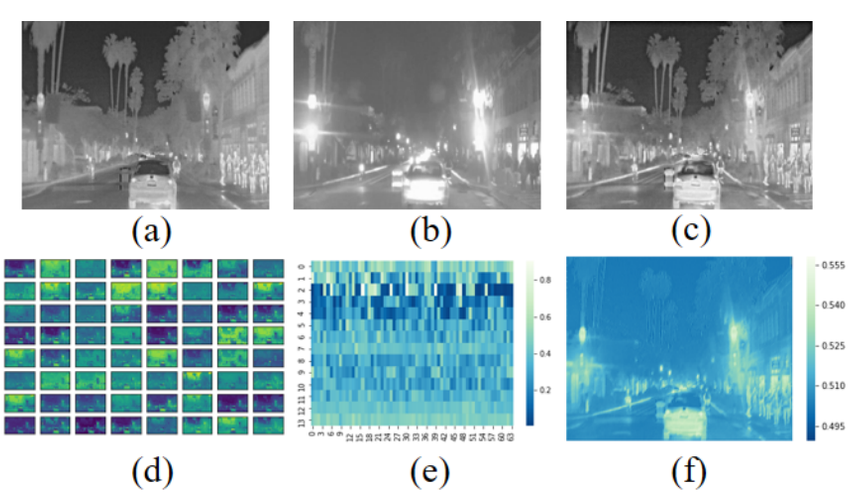

图 6　注意力权重图:(a)红外图像;(b)可见光图像;(c)本文方法的融合结果;(d)第 3 个注意力模块输出结果图;(e)通道注意力权重图;(f)空间注意力权重图

Fig. 6　Attention weight maps: (a) the infrared image; (b) the visible image; (c) the fused result of our proposed EAGAN; (d) Output result of the third attention block; (e) Channel Attention weight map; (f) Spatial Attention weight map

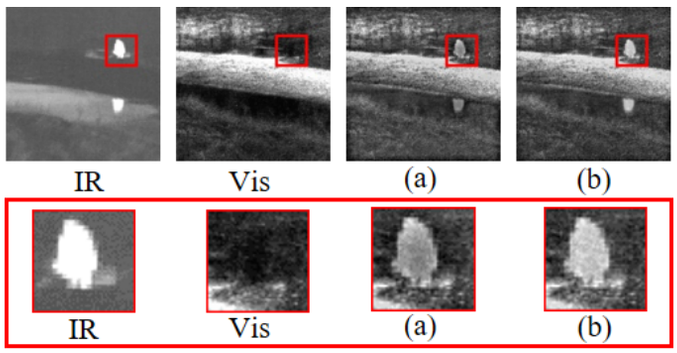

图7 注意力机制对融合效果的影响:(a) 无注意力机制模型测试结果 (b)本文方法测试结果

Fig. 7 The effect of attention mechanism on fusion results: (a) fusion result of the network without attention mechanism; (b) fusion result of our algorithm.

表4 注意力机制对融合效果影响对比

Table 4 Comparison of effects of attention mechanism on fusion results

		EN	SCD	SF	EI
RoadScene	无注意力机制方法	7.26	1.52	**16.02**	0.27
	本文方法	**7.30**	**1.54**	15.62	0.27
TNO	无注意力机制方法	6.93	1.63	**11.62**	0.19
	本文方法	**7.08**	**1.67**	11.59	0.19

3 结论

针对融合算法缺少真值难以用深度学习方法进行学习的问题,本文提出了一种基于边缘保持和注意力机制的生成对抗网络,应用于红外与可见光融合算法。本文将注意力机制应用到最小二乘生成对抗网络中,提高了网络显著性目标的特征提取能力。同时提出了一种边缘损失函数,并结合了感知损失和对抗损失,将融合问题从像素空间分别映射到边缘和深层特征空间。该算法生成的图像良好融合了源图像对信息,红外目标突出、可见光纹理丰富,同时算法提高了融合图像清晰度和对比度,增强了图像边缘,有利于后续检测、识别等任务。然而在实际应用中,红外与可见光图像无法实现绝对配准,配准算法精度对成像效果有一定影响,因此在非严格配准条件下的图像融合问题亟待解决。此外,源图像对和融合图像之间的映射关系还可以进一步探索,并且通过优化网络结构和损失函数等方法,本文算法仍有提升空间。

References

[1] Ma J Y, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey [J]. *Information Fusion*, 2019, **45**: 153-178.

[2] Liu G X, Yang W H. Image fusion scheme of pixel-level and multi-operator for infrared and visible light images [J]. *J. Infrared Millim. Waves*(刘贵喜,杨万海.一种像素级多算子红外与可见光图像融合方法.**红外与毫米波学报**), 2001, **20**(3): 207-210.

[3] Upla K P, Joshi M V, Gajjar P P. An Edge Preserving Multiresolution Fusion: Use of Contourlet Transform and MRF Prior [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, **53**(6): 3210-3220.

[4] Liu S P, Fang Y. Infrared image fusion algorithm based on contourlet transform and improved pulse coupled neural network [J]. J. *Infrared Millim. Waves*(刘盛鹏,方勇.基于Contourlet变换和IPCNN的融合算法及其在可见光与红外线图像融合中的应用.**红外与毫米波学报**), 2007, **26**(3): 217-221.

[5] Li S, Kang X, Hu J. Image fusion with guided filtering [J]. *IEEE Transactions on Image Processing*, 2013, **22**(7): 2864-2875.

[6] Liu Y, Wang Z F. Simultaneous image fusion and denoising with adaptive sparse representation [J]. *Image Processing Iet*, 2015, **9**(5): 347-357.

[7] Li S T, Kwok J T, Wang Y N. Multifocus image fusion using artificial neural networks [J]. *Pattern Recognition Letters*, 2002, **23**(8): 985-997.

[8] 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).

[9] Li H, Wu X J, Kittler J. Infrared and Visible Image Fusion using a Deep Learning Framework [J]. *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018: 2705-2710.

[10] Li H, Wu X J. DenseFuse: A Fusion Approach to Infrared and Visible Images [J]. *IEEE Transactions on Image Processing*, 2018:2614-2623.

[11] Ma J Y, Yu W, Liang P W, *et al*. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. *Information Fusion*, 2019, **48**: 11-26.

[12] Xu J T, Shi X P, Qin S Z, *et al*. LBP-BEGAN: A generative adversarial network architecture for infrared and visible image fusion [J]. *Infrared Physics & Technology*, 2020, 104.

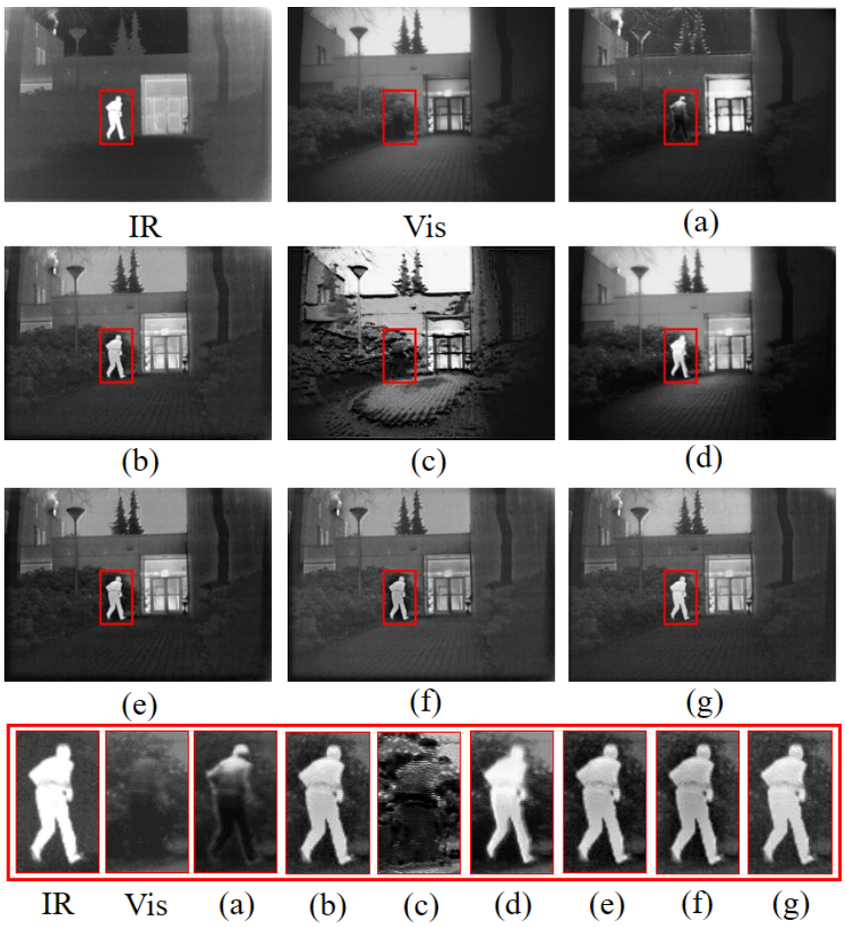

图 8　损失函数对比实验结果:(a)$\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual}$;(b)$\mathcal{L}_G = \lambda_2\mathcal{L}_{edge}$;(c)$\mathcal{L}_G = \mathcal{L}_{EAGAN}(G)$;(d)$\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_1\mathcal{L}_{perceptual}$;(e)$\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_2\mathcal{L}_{edge}$;(f)$\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual} + \lambda_2\mathcal{L}_{edge}$;(g)本文方法实验结果

Fig. 8　Fusion results when the loss function of the generator changes: (a) $\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual}$; (b) $\mathcal{L}_G = \lambda_2\mathcal{L}_{edge}$; (c) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G)$; (d) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_1\mathcal{L}_{perceptual}$; (e) $\mathcal{L}_G = \mathcal{L}_{EAGAN}(G) + \lambda_2\mathcal{L}_{edge}$; (f) $\mathcal{L}_G = \lambda_1\mathcal{L}_{perceptual} + \lambda_2\mathcal{L}_{edge}$; (g) result of EAGAN.

[13] Li Q, Lu L, Li Z, *et al*. Coupled GAN with Relativistic Discriminators for Infrared and Visible Images Fusion [J]. *IEEE Sensors Journal*, 2019: 1-1.

[14] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al*. Generative Adversarial Nets [J]. *Advances in Neural Information Processing Systems 27* (*NIPS* 2014), 2014, 27.

[15] Mao X D, Li Q, Xie H R, *et al*. Least Squares Generative Adversarial Networks [J]. *2017 IEE International Conference on Computer Vision* (*ICCV*), 2017: 2813-2821.

[16] Simonyan K, Zisserman A J C e. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. *2014 International Conference on Learning Representations*, 2014.

[17] Zhang Y, Li K, Li K, *et al*. Image Super-Resolution Using Very Deep Residual Channel Attention Networks [J]. *2018 European Conference on Computer Vision*, 2018: 294-310.

[18] Zagoruyko S, Komodakis. N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer [J]. *International Conference on Learning Representations*, 2017.

[19] Nair V, Geoffrey E H. Rectified linear units improve restricted boltzmann machines [J]. *Proceedings of the 27th International Conference on Machine Learning*, 2010: 807-814.

[20] Wang F, Jiang M Q, Qian C, *et al*. Residual Attention Network for Image Classification [J]. *30th IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR* 2017), 2017: 6450-6458.

[21] Johnson J, Alahi A, Li F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution [J]. *European Conference on Computer Vision*, 2016: 694-711.

[22] Xu H, Ma J, Jiang J, *et al*. U2Fusion: A Unified Unsupervised Image Fusion Network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[23] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization [J]. *International Conference on Learning Representations*, 2014.

[24] Ma J Y, Chen C, Li C, *et al*. Infrared and visible image fusion via gradient transfer and total variation minimiza-

tion [J]. *Information Fusion*, 2016, **31**: 100-109.
[25] Roberts J W, van Aardt J, Ahmed F. Assessment of image fusion procedures using entropy, image quality, and multispectral classification [J]. *Journal of Applied Remote Sensing*, 2008, 2.
[26] 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): 160-166.
[27] Eskicioglu A M, Fisher P S. Image quality measures and their performance [J]. *IEEE Transactions on Communications*, 1995, **43**(12): 2959-2965.
[28] Rajalingam B, Priya R. Hybrid Multimodality Medical Image Fusion Technique for Feature Enhancement in Medical Diagnosis [J]. *International Journal of Engineering Science Invention*, 2018.