DOI: 10.13382/j.jemi.B2307071

基于双重注意力机制生成对抗网络的偏振图像融合*

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

摘 要:针对单一强度图像缺少偏振信息,在恶劣天气条件下无法提供充足场景信息的问题,本文提出了一种基于双重注意力机制生成对抗网络用于强度图像和偏振度图像进行融合。算法网络由一个包含编码器、融合模块和解码器的生成器和一个鉴别器组成。首先源图像输入到生成器的编码器中,经过一个卷积层和密集块进行特征提取,然后通过含有注意力机制的纹理增强融合模块中进行特征融合,最后通过解码器得到融合图像。鉴别器主要由两个卷积模块和两个注意力模块组成,在网络训练过程中,通过不断博弈,迭代优化生成器网络参数,使生成器输出既保留偏振度图像的稀疏特征又不损失强度图像信息的高质量融合图像。实验表明,该方法得到的融合图像在主观上纹理信息更丰富,更符合人眼的视觉感受,并且在客观评价指标中 SD 提升约 18.5%,VIF 提升约 22.4%。

关键词: 图像融合;偏振图像;生成对抗网络;注意力机制
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

Polarization image fusion based on dual attention mechanism for generating adversarial networks

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

Abstract: Aiming at the problem that a single intensity image lacks polarization information and cannot provide sufficient scene information under bad weather conditions, this paper proposes a dual-attention mechanism to generate an adversarial network for fusion of intensity and polarization images. The algorithmic network consists of a generator containing an encoder, a fusion module and a decoder and a discriminator. First, the source image is fed into the encoder of the generator, after a convolutional layer and dense block for feature extraction, then feature fusion is performed in the texture enhancement fusion module containing the attention mechanism and finally the fused image is obtained by the decoder. The discriminator is mainly composed of two convolutional modules and two attention modules, and the generator network parameters are iteratively optimised by constant gaming during the network training process, so that the generator outputs a high-quality fused image that retains the sparse features of the polarimetric image without losing the intensity image information. Experiments show that the fused images obtained by this method are subjectively richer in texture information and more in line with the visual perception of the human eye, and that the SD is improved by about 18.5% and the VIF by about 22.4% in the objective evaluation index.

Keywords: image fusion; polarization images; generative adversarial networks; attention mechanisms

0 引 言

强度图像包含了图像的主要能量,背景信息丰富,但在复杂光场条件下,强度成像难以区分目标和背景,同时雾霾、烟尘、雨雪、海雾等传输条件导致低能见度,强度成像可视距离变短,图像清晰度和作用距离矛盾突出。偏振成像技术是在传统成像的基础上增加了偏振维度信息,不仅能提供二维空间的光强分布,还能获得目标和背景的偏振信息,当目标与背景辐射强度相近或被遮挡时,

收稿日期:2023-11-23 Received Date: 2023-11-23
*基金项目:国家自然科学基金重大仪器专项(62127813)资助

偏振成像可以清晰地将目标从背景中突显出来。为补充和丰富目标检测及识别的光电探测手段,提高雨、雪、雾霾和逆光等复杂条件下的探测识别能力,增加目标的探测距离,提升目标识别的准确率,可以利用偏振图像融合技术,融合强度图像和偏振度图像,增强边缘纹理清晰,提高图像的对比度,解决“看不远”、“认不清”和“辨不出”的问题。

在过去的几十年中,许多图像融合方法被提出,主要分为传统融合方法和基于深度学习融合方法两类,传统融合算法主要包括基于空间域融合方法和基于变换域融合方法,基于空间域融合方法主要有主成分分析法[1]、引导滤波[2]等,基于变换域融合方法主要有拉普拉斯金字塔(Laplacian pyramid, LP)[3]、梯度金字塔(gradient pyramid, GP)[4]、离散小波(discrete wavelet transform, DWT)[5]、双树复小波(dual-tree complex wavelet transforms, DTCWT)[6]、非下采样轮廓波(non-subsampled contourlet transform, NSCT)[7]、非下采样剪切波(nonsubsampled shearlet transform, NSST)[8]等。空间域融合方法的结果取决于图像块的选择,在不同亮度的图像上具有较大的差别;变换域融合方法需要选择多尺度分解工具并手工设计融合规则,无法保证在不同场景下都能自适应地取得好的融合效果。

近几年,随着深度学习技术的不断发展,学者们将其引入图像融合领域,并取得了许多优于传统方法的融合成果,根据网络结构,目前基于深度学习的融合方法主要包括基于卷积神经网络[9](convolutional neural network, CNN)、基于自编码器(auto-encoder, AE)[10-12]和基于生成对抗网络[13-14](generative adversarial network, GAN)的方法。其中基于 GAN 的融合方法采用无监督学习,通过对抗训练生成模型可以有效的平衡源图像的特征信息分配,取得了较好的融合效果。

偏振图像融合作为融合领域的一个分支,基于深度学习的方法相对研究较少。2021 年 Zhang 等[15]首次将 CNN 用于可见光单波段偏振融合中,提出了基于无监督深度融合框架 PFnet,取得了较好的效果。随后,作者又借鉴残差结构设计了带有自学习融合策略的融合网络,进一步提升了融合效果。Wang 等[16]提出了一种基于密集连接的生成对抗融合网络,在生成器中加入密集连接块,同时把梯度损失函数作为结构相似度损失函数的权重进行训练,使生成的图像具有更多梯度信息。2023 年,Liu 等[17]通过加入信息量判别块,提出了基于双鉴别器 GAN 的语义引导偏振图像融合方法,增强了场景表示的充分性。上述基于深度学习的方法解决了传统方法需要手工设计融合规则和适应性差的不足,提升了融合质量,但仍然存在融合结果图像纹理不清晰,亮度和对比度较低等问题。为此本文结合残差块结构、注意力机制和 GAN 网络的优势,提出了一种基于双重注意力机制生成对抗网络的偏振图像融合方法,在生成器的编码器中加入密集块用于提取源图像特征,融合模块中构建了加入注意力机制的梯度残差结构,增加融合图像的梯度信息和网络稳定性,在解码器中,引入编码器中各卷积层的特征图,增强特征图的利用率,丰富融合图像的细节信息,在鉴别器中,加入空间注意力和通道注意力机制以聚焦图像场景信息以及偏振特征,促使融合图像既能保留偏振度图像的稀疏特征又不损失强度图像的背景信息。

1　本文方法

1.1　数据预处理

分焦平面偏振相机将光电探测器和微偏振片器件集成在同一个焦平面上,探测器上的每个像元对应着一个微偏振片,在相邻 2×2 像元组合中,微偏振片角度分别为 0°、45°、90° 和 135°,构成一个超像元,所以分焦平面偏振相机获得的图像是类似于马赛克的图像,采用文献[18]提出的牛顿多项式算法对图像进行去马赛克和降噪处理,利用 Stokes 矢量来定量描述强度和偏振度信息,公式如下:

$$\begin{cases} S_0 = (I_{0^\circ} + I_{45^\circ} + I_{90^\circ} + I_{135^\circ})/2 \\ S_1 = I_{0^\circ} - I_{90^\circ} \\ S_2 = I_{45^\circ} - I_{90^\circ} \end{cases} \tag{1}$$

式中:S_0 是强度图像,S_1 表示 0°和 90°方向线性偏振的强度差;S_2 表示 45°和 135°方向线性偏振的强度差。利用上述 3 种图像可以获得描述偏振特性的偏振度图像,公式如下:

$$DoLP = \frac{\sqrt{S_1^{\,2} + S_2^{\,2}}}{S_0} \tag{2}$$

式中:$DoLP$ 为偏振度图像,是一个介于 0~1 之间的标量,从上述定义中可以看出,强度图像 S_0 捕捉反射光和透射光,表达了物体的光谱信息,而偏振度图像 $DoLP$ 与强度无关,表述的是人造目标表面形状、阴影和粗糙度等详细特征,所以为了获得含有强度和偏振信息的图像,有必要将强度图像和偏振度图像进行融合,提高复杂背景下人造目标的检测能力。

1.2　网络结构

融合网络包括生成器和鉴别器,生成器主要由编码器(Encoder)、融合模块(Fusion Block)、解码器(Decoder)3 部分组成,如图 1 所示。编码器对输入图像中进行特征提取,融合模块对特征图进行融合处理,编码器对融合后的特征图进行重构,得到最后的融合图像 I_f。鉴别器的作用是在网络训练阶段区分融合图像 I_f 和强度图像 S_0,把鉴别结果反馈给生成器进行网络参数优化,形

成一种对抗训练，当鉴别器区分不出融合图像时，训练结束。在测试阶段，将强度图像 S_0 和偏振度图像 *DoLP* 输入到优化参数后的生成器中，得到融合图像。

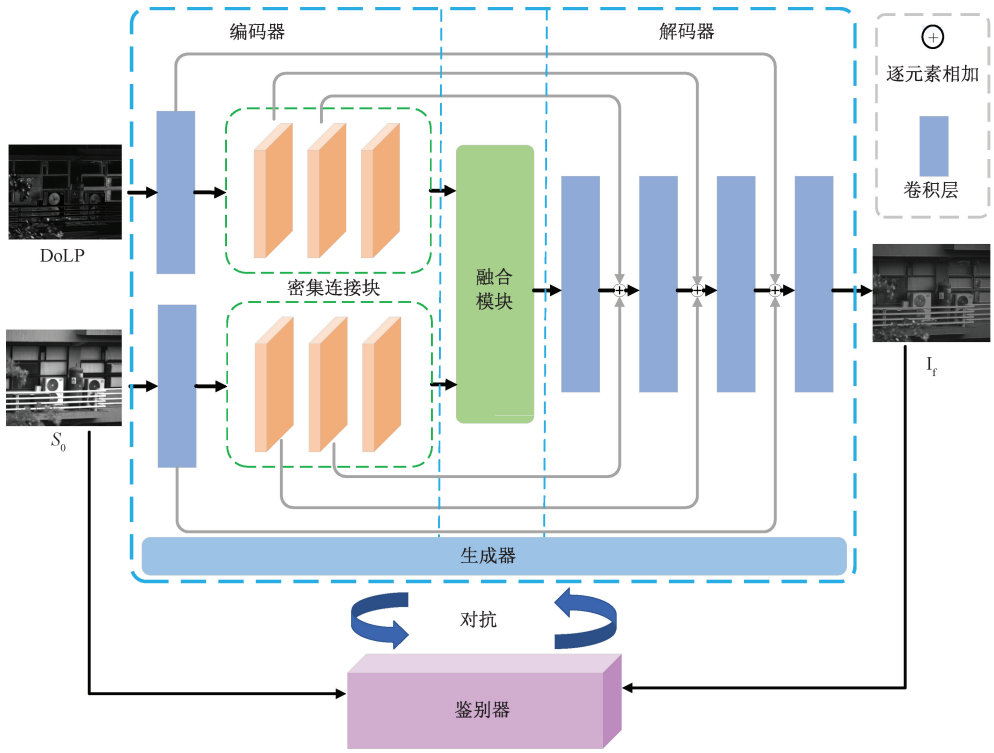


图 1　整体网络结构

Fig. 1　Overall network structure

1）生成器结构

生成器由编码器、融合模块、解码器组成，如图 1 所示，编码器如图 1 中 Encoder 模块所示，由 1 个卷积层和 1 个密集块组成，密集块结构如图 2 所示，由 3 个 3×3 的卷积层组成，每个卷积层把前面卷积层的输出进行级联作为本层的输入，从而实现浅层特征的再次利用和深度特征的提取，最后输入到融合模块中。强度图像 S_0 和偏振度图像 *DoLP* 输入到编码器中，先经过一个 3×3 卷积层的浅层特征提取，得到 16 通道的特征图，记为 $F_X^{1:16}$，$X \in \{S_0, DoLP\}$，然后输入到密集块中进行深层特征提取，得到 64 通道的特征图，记为 $\hat{F}_X^{1:64}$。

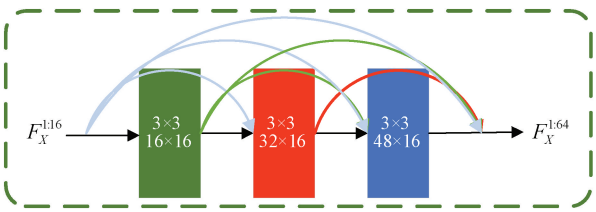


图 2　密集块结构

Fig. 2　Dense block construction

融合模块结构如图 3 所示，特征图 $F_X^{1:64}$ 的走向分成两个支路，在主支路的强度特征通道中加入通道注意力机制，自适应地预测和增加通道重要特征的比例，在偏振度特征通道中加入空间注意力机制，自动捕获特征图的重要区域，然后再分别通过 3×3 卷积层进行特征提取，并与强度通道的特征图进行级联，使用 1 个 1×1 的卷积层和 2 个 3×3 的卷积层进行降维和特征提取，得到 64 通道特征图，记为 $\hat{F}_A^{1:64}$；在分支路中，利用 Sobel 算子对特征图 $\hat{F}_{S_0}^{1:64}$ 和 $\hat{F}_{DoLP}^{1:64}$ 进行梯度提取，得到梯度特征图，利用像素取大策略合并特征图，得到 64 通道梯度特征图，记为 $\hat{F}_D^{1:64}$，然后经过两个卷积层和 1 个 Sigmoid 归一化函数得到含有梯度信息的权重图 $\hat{F}_g$，$\hat{F}_g$ 与 $\hat{F}_A^{1:64}$ 相乘得到 $\hat{F}_G^{1:64}$，最后，为了减少信息丢失，将 $\hat{F}_A^{1:64}$ 与 $\hat{F}_G^{1:64}$ 进行逐像素相加，得到含有丰富纹理信息的 64 通道特征图，记为 $\hat{F}_P^{1:64}$。

（1）通道注意力机制

通道注意力模块如图 4 所示，分别利用全局最大池化和全局平均池化对输入的 64 通道特征图 $\hat{F}_{S_0}^{1:64}$ 进行空间维度压缩，获得两个描述通道特征的 64 维向量。通过多层感知器（multilayer perceptron，MLP）学习，得到注意

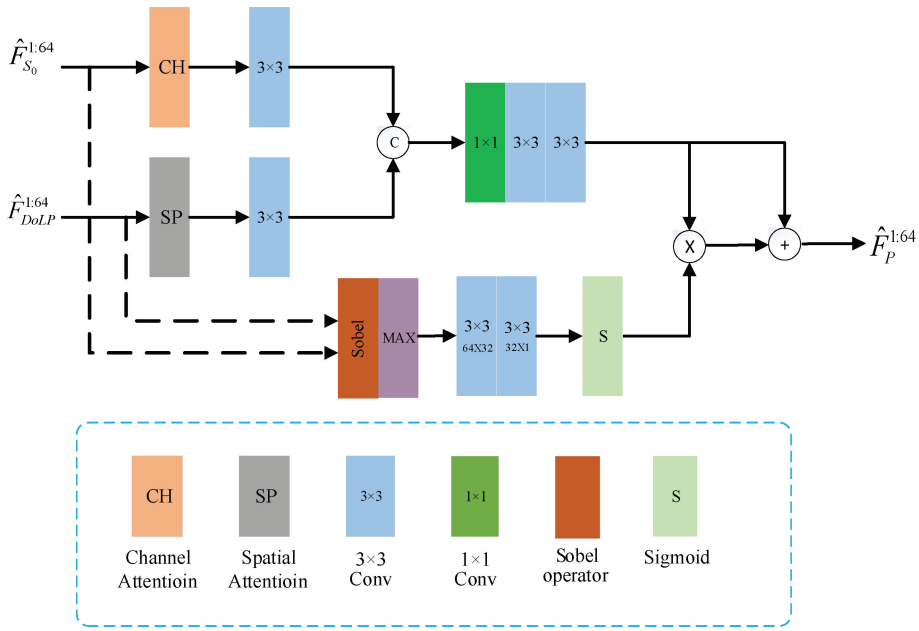


图 3　融合模块

Fig. 3　Fusion module

力特征图,然后逐像素相加,再经过 Sigmoid 激活函数将特征图的每个通道的权重值归一化到 0~1 之间,与通道特征图 $\hat{F}_{S_0}^{1:64}$ 相乘,得到 64 通道显著特征图 $\hat{F}_{CH}^{1:64}$。

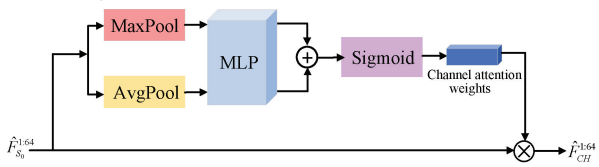


图 4　通道注意力机制

Fig. 4　Channeling attention mechanisms

(2)空间注意力机制

空间注意力机制在空间维度上细化特征,关注图像中的重要区域,增强特征图中的显著信息,结构如图 5 所示,首先在通道维度下,对输入的 64 通道特征图 $\hat{F}_{DoLP}^{1:64}$ 做最大池化和平均池化,得到两组 1 维特征向量并进行级联,然后,使用 5×5 大小的卷积核融合通道信息和降维处理,结果经过 Sigmoid 激活函数进行归一化处理,得到空间注意力权重矩阵,最后,将权重矩阵与通道特征图 $\hat{F}_{DoLP}^{1:64}$ 相乘,得到空间显著特征图 $\hat{F}_{SP}^{1:64}$。

解码器结构如图 1 中 Decoder 模块所示,由 4 个 3×3 卷积层组成,输入通道数分别为 64、48、32、16,输出通道数分别为 48、32、16、1,最后输出的融合图像为 I_f。了使图像的纹理细节信息不丢失,将编码器密集块中特征图通过逐像素取大策略获得的特征图,利用跳跃连接引入解码器的对应卷积层中,通过逐像素相加的方式,将更多的

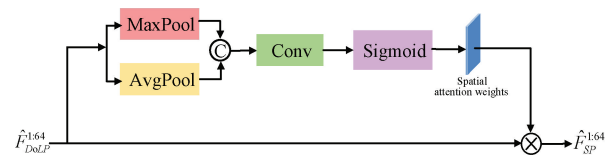


图 5　空间注意力机制

Fig. 5　Spatial attention mechanisms

结构和纹理特征信息从编码器转移到解码器中,同时这种跳跃连接使算法网络更容易训练,提升收敛速度。

2)鉴别器结构

鉴别器本质上是一种分类器(Classifier),当输入一幅图像时,网络输出当前图像属于某一类的条件概率,这个结果作为反馈输入到生成器中,以更新生成器的参数。

为了更好的区分融合图像和 S_0 本文引入通道注意力模块(channel attention block,CA Block)和空间注意力模块(spatial attention block,SA Block)对输入图像进行更细致的通道维度和空间维度的特征提取,提高鉴别器的分辨能力。鉴别器结构如图 6 所示,融合图像 I_f 或强度图像 S_0 输入到鉴别器中,鉴别器依次由卷积块(Conv Block)、通道注意力模块、卷积块、空间注意力模块和线性层(FC)组成。输出通道分别为 32、64、128、256。卷积块包括 3×3 卷积层、批量归一化层(BN)和 Leaky ReLU 激活函数。通道注意力模块和空间注意力模块结构相似都是由 3×3 卷积层、批量归一化层(batch normalization,BN)、通道注意力机制或者空间注意力机制组成。最后将特征图输入到线性层进行分类,输出一个标签判定输

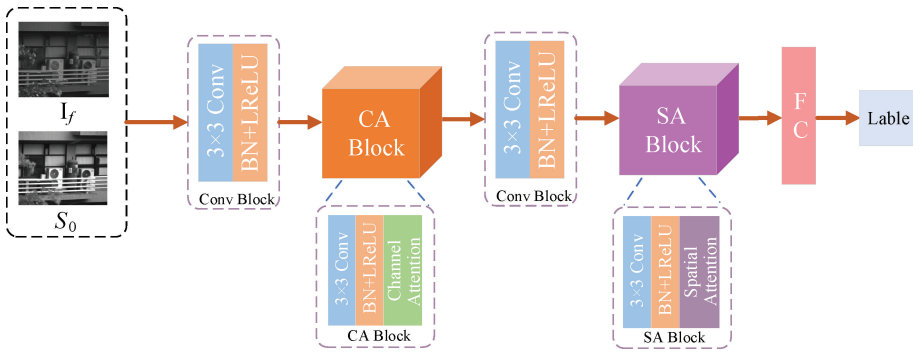


图6　鉴别器网络结构

Fig. 6　Discriminator network architecture

入图像是 I_f 还是 S_0。

1.3　损失函数

本文的损失函数主要分为两个部分:生成器损失 L_G、鉴别器损失 L_D。

1)生成器损失函数

生成器损失函数 L_G 由4部分组成,分别为生成器与鉴别器之间的对抗损失函数 L_{adv}、梯度损失函数 L_{Grad}、强度损失函数 L_{int} 和多尺度加权结构相似度损失函数 $L_{MSWSSIM}$ 组成,如式(3)所示:

$$L_G = L_{adv} + \alpha L_{MSWSSIM} + \beta L_{Grad} + \gamma L_{int} \tag{3}$$

式中:α、β、γ 是平衡参数,本文中 $\alpha=100$,$\beta=10$,$\gamma=10$。对抗损失函数 L_{adv} 如式(4)所示:

$$L_{adv} = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_f^n) - h\right)^2 \tag{4}$$

式中 $D_{\theta_D}(\cdot)$ 为鉴别器输出结果,本文表示输入图像是 S_0 图像的概率,I_f^n 表示融合图像,n 表示融合图像的个数,h 是生成器希望鉴别器相信的假数据,本文将其设为0.7~1.2之间的随机值。多尺度加权结构相似度损失函数 $L_{MSWSSIM}$ 如式(5)所示:

$$L_{MSWSSIM} = 1 - \frac{1}{5}\sum_{w\in\{3,5,7,9,11\}}\left(\eta_w \cdot L_{SSIM}(S_0, I_f; w) + (1-\eta_w)\cdot L_{SSIM}(DoLP, I_f; w)\right) \tag{5}$$

式中:η_w 为权重系数,其计算公式如式(6)所示:

$$\eta_w = \frac{g(\sigma^2_{w_{S_0}})}{g(\sigma^2_{w_{S_0}}) + g(\sigma^2_{w_{DOLP}})} \tag{6}$$

式中:$g(x)=\max(x, 0.0001)$,表示修正函数,$\sigma^2_{w_{S_0}}$ 表示 S_0 在窗口 w 中的方差。

$L_{SSIM}(\cdot)$ 为结构相似度函数,其基于像素点的计算如式(7)所示:

$$L_{SSIM}(x,y;w) = \frac{(2\bar{w}_x\bar{w}_y + C_1)(2\sigma_{w_xw_y} + C_2)}{(\bar{w}_x^2 + \bar{w}_y^2 + C_1)(\sigma^2_{w_x} + \sigma^2_{w_y} + C_2)} \tag{7}$$

其中,C_1、C_2 为常数,本文中 $C_1=1\times10^{-4}$,$C_2=9\times10^{-4}$,w_x、w_y 表示图像 x、y 在窗口 w 中的部分,$\bar{w}_x$、$\bar{w}_y$ 表示所有 w_x、w_y 的均值的平均,$\sigma^2_{w_x}$、$\sigma^2_{w_y}$ 表示 w_x 与 w_y 的方差,$\sigma_{w_xw_y}$ 表示 w_x 与 w_y 的协方差。当滑动窗口遍历整个图像后,得到 L_{SSIM} 的值。

为了增强融合图像的纹理信息,引入梯度损失函数 L_{Grad},如式(8)所示:

$$L_{Grad} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left(D_{S_0(x,y)} - D_{f(x,y)}\right)^2 \tag{8}$$

式中:W 为图像的宽度,H 为图像的高度,$D_{S_0(x,y)}$ 表示强度图像 S_0 的梯度,$D_{f(x,y)}$ 表示融合图像的梯度。强度损失函数 L_{int} 如式(9)所示:

$$L_{int} = \frac{1}{HW}\| w_P(DoLP - I_f)\|_F^2 + \frac{1}{HW}\| w_I(S_0 - I_f)\|_F^2 \tag{9}$$

其中,$\|\cdot\|_F$ 表示F范数,用于度量两个矩阵之间的距离。w_P 和 w_I 为权重系数,用来约束融合图像和两个源图像之间像素分布的相似性。当偏振图像的像素值比强度图像的像素值更大时,这意味着像素的偏振信息更显著,即权重系数 w_P 应该更大,反之应该更小。w_P 和 w_I 定义为:

$$w_P = \frac{\exp(DoLP/\tau)}{\exp(DoLO/\tau) + \exp(S_0/\tau)},\quad w_I = 1 - w_P \tag{10}$$

其中,τ 表示常系数。本文中 $\tau=0.1$,以放大像素值的差异。

2)鉴别器损失函数

虽然鉴别器仅用于图像分类,但是鉴别器中强度图像和融合图像之间的损失会迫使生成器生成与强度图像强度分布相似的融合图像,使图像更符合人的视觉感受。因此,为了更好的区分融合图像和强度图像,本文使用的鉴别器损失函数如式(11)所示:

$$L_D = \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(S_0^n) - a\right)^2 + \frac{1}{N}\sum_{n=1}^{N}\left(D_{\theta_D}(I_f^n) - b\right)^2 \tag{11}$$

式中：$D_{\theta_D}(S_0)$，$D_{\theta_D}(I_f)$ 分别为图像 S_0 和融合图像 I_f 经过鉴别器之后的分类结果，N 为融合图像数量，a 和 b 分别为融合图像 I_f 和强度图像 S_0 的标签，分别设定为 1 和 0。

2　实验与分析

2.1　实验设置

本文网络训练过程采用与文献[19]相同的方法，首先采用 MS-COCO[20] 数据集进行编码器和解码器之间的训练，再分别使用 ZJU-RGB-P[21] 数据集和文献[22]数据集进行融合模块和整体网络的训练。从经过预处理的 ZJU-RGB-P 数据集和文献[22]数据集中选取 70% 的 S_0 和 *DoLP* 图像作为网络的训练集，为了加快训练速度，把每幅图像都进行裁剪，得到 120×120 像素大小的图像，分别得到大约 11 000 张 S_0 和 *DoLP* 训练图像；其余 30% 的图像作为测试集图像使用。训练阶段首先用 Adam 训练鉴别器，然后再训练生成器，直到训练次数达到 32 次，取最佳模型进行测试。学习率设置为 1×10^{-4}，衰减率设置为 0.9，计算机硬件配置为 2.0 GHz Intel(R) Xeon(R) CPU 和 NVIDIA RTX 3090 GPU。

2.2　对比实验与分析

为验证本文方法的有效性，选取基于曲波变换(CVT)[10]、梯度转移融合(GTF)[23]、低通比率金字塔(RP)[24]、IFCNN[9]、FusionGAN[25]、DenseFuse[19]、PFnet[15] 7 种方法与本文方法进行对比实验。其中 CVT、GTF 为传统融合方法，IFCNN 是基于卷积神经网络的融合方法，DenseFuse 是基于自编编码器的融合方法，FusionGAN 是基于 GAN 网络的融合方法，PFnet 是基于无监督深度学习框架的融合方法，以上 4 种基于深度学习的方法都是使用数据集 ZJU-RGB-P 和文献[22]数据集进行训练后得到输出结果。

本文使用信息熵(information entropy，EN)[26]、标准差(standard deviation，SD)[27]、均方误差(mean-square error，MSE)[28]、峰值信噪比(peak signal-to-noise ratio，PSNR)[29]、互信息(mutual Information，MI)[30] 和视觉信息保真度(visual information fidelity，VIF)[31] 6 种客观评价指标对融合性能进行评价。EN 值越大，表明融合图像包含更多的信息；SD 值越大，表明融合图像具有更高的对比度；MSE 值越小，表明融合图像与源图像间相似度越高；PSNR 值越大，表明融合图像失真成度越小；MI 值越大，表明融合图像保留了更多的源图像信息；VIF 是结合自然图像统计模型、图像失真模型和人眼视觉系统模型的图像质量评价指标，VIF 值越大，表明融合图像符合人的视觉感受。实验所用的源图像和融合结果如图 7 和 8 所示，客观评价指标数据如表 1 所示。

从图 7 中可以看出，CVT、IFCNN、DenseFusion 融合结果中出现了明显的目标边缘模糊现象，RP 融合结果中丢失了 S_0 图像信息，保留了过多的 *DoLP* 图像信息，GTF 和 FusionGAN 融合结果过于平滑，丢失了许多能量，PFnet 融合结果中在较暗区域出现细节丢失、对比度低的情况，而本文融合结果在保留足够多源图像纹理细节信息的基础上具有更自然的视觉感受。

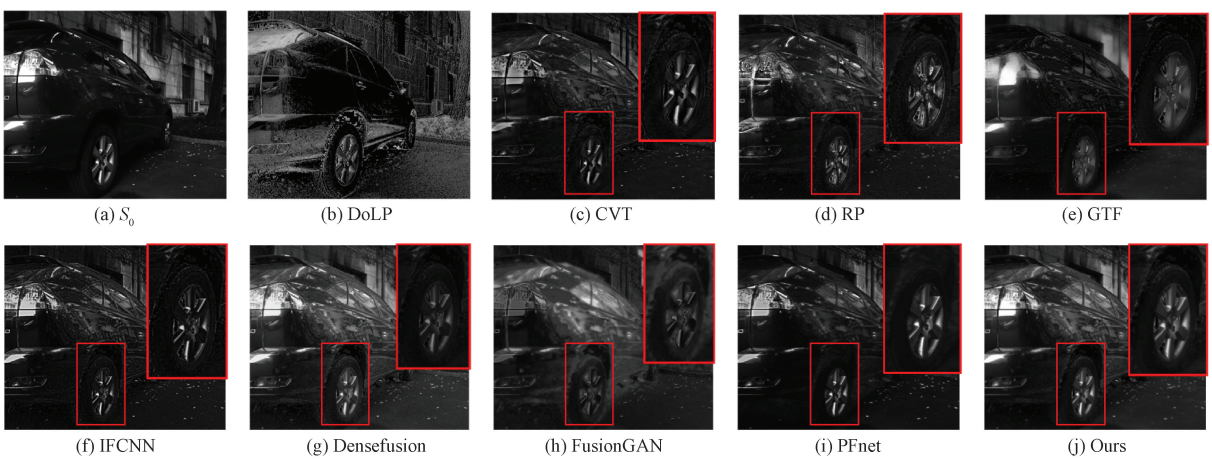


图 7　ZJU-RGB-P 数据集场景 1 对比实验

Fig. 7　ZJU-RGB-P dataset scene 1 comparison experiment

在图 8 所示文献[22]数据集场景 1 的对比实验中，CVT 融合结果中背景信息出现一定程度的丢失，RP 融合结果出现了严重的伪影现象，GTF 和 FusionGAN 融合结果过于平滑，IFCNN 出现了 *DoLP* 信息的过度保留，

Densefusion 融合结果亮度较低，丢失了强度图像 S_0 的语义信息，不便于理解场景信息，PFnet 融合结果对比度较低，边缘纹理出现丢失，本文融合结果对比度适中，纹理边缘信息保留完整，符合人眼的视觉感受。

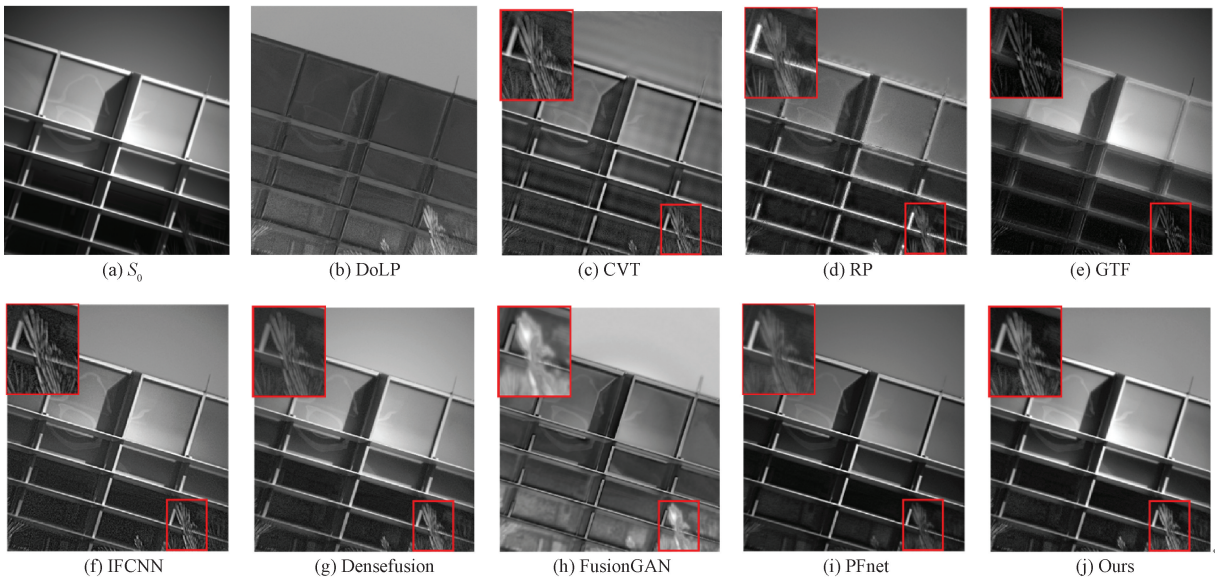


图 8 文献[22]数据集场景 1 对比实验

Fig. 8 Literature [22] dataset scenario 1 comparison experiment

表 1 不同融合方法融合结果的客观评价

Table 1 Objective evaluation of fusion results of different fusion methods

	EN	SD	MSE	PSNR	MI	VIF
CVT	6. 798	34. 063	0. 015	66. 130	1. 270	0. 502
	7. 081	40. 403	0. 028	63. 665	2. 245	0. 549
GTF	6. 976	36. 263	0. 016	64. 161	1. 208	0. 435
	7. 080	43. 565	0. 044	63. 446	2. 829	0. 372
RP	6. 624	34. 919	0. 024	66. 209	1. 553	0. 633
	6. 053	42. 265	0. 028	60. 377	2. 602	0. 561
IFCNN	6. 731	33. 684	0. 025	63. 186	1. 458	0. 482
	7. 2854	46. 479	0. 028	63. 602	3. 174	0. 606
Fusion-GAN	6. 755	33. 145	0. 016	66. 924	1. 397	0. 235
	6. 993	54. 305	0. 059	60. 377	2. 937	0. 312
Dense-fuse	6. 647	27. 056	0. 031	63. 187	2. 547	0. 505
	6. 681	32. 392	0. 024	64. 303	3. 496	0. 557
PFnet	6. 891	30. 070	0. 013	65. 591	1. 637	0. 321
	7. 022	40. 335	0. 027	63. 783	1. 647	0. 495
Ours	7. 234	39. 873	0. 010	73. 385	3. 358	0. 750
	7. 909	60. 306	0. 019	67. 611	3. 704	0. 645

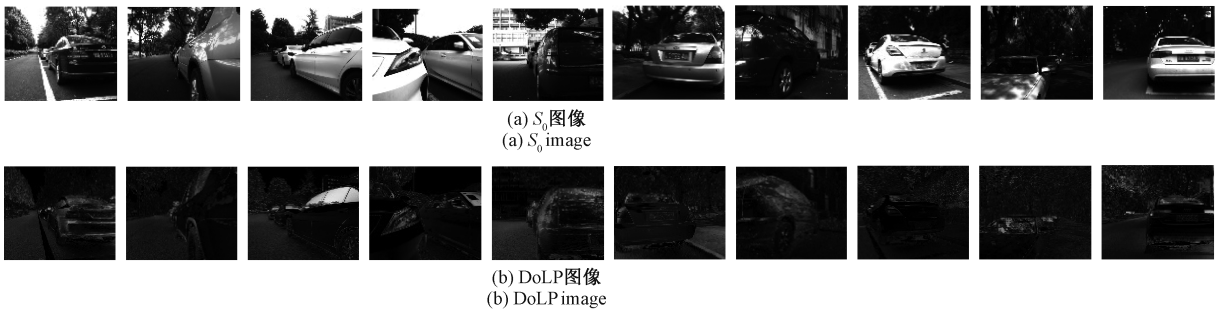
为了验证本文融合方法的稳定性，从 ZJU-RGB-P 数据集选取 10 幅偏振图像进行融合实验，源图像如图 9 所示。

融合结果的平均定量客观评价如表 2 所示，表中数据为各评价指标的平均值，图 10 所示为不同融合方法的 6 种客观评价指标折线图。表 2 中加粗的字体表示此项指标中的最佳值。从表 2 和图 10 中可得出，本文融合结果在客观评价数据方面均优于其他 7 种方法，其中 MI 增幅最明显达到 36. 70%，这说明本文融合方法能够稳定地得到高质量的融合结果。

2. 3 消融实验与分析

为验证本文网络结构的有效性，设计了 7 种网络结构进行消融实验，1）去除生成器中通道注意力机制，记为 W/O G(Ch)；2）去除生成器中空间注意力机制，记为 W/O G(SP)；3）去除生成器中融合模块，记为 W/O G(AGR)；4）去除生成器中通道和空间注意力机制，记为

W/O G(Att);5)去除鉴别器中空间注意力机制,记为 W/O D(SP);6)去除鉴别器中通道注意力机制,记为 W/O D(Ch);7)去除鉴别器中通道和空间注意力机制,记为 W/O D(Att)。源图像和融合结果如图 11 所示,客观指标如表 3 所示。



(a) S_0图像
(a) S_0 image

(b) DoLP图像
(b) DoLP image

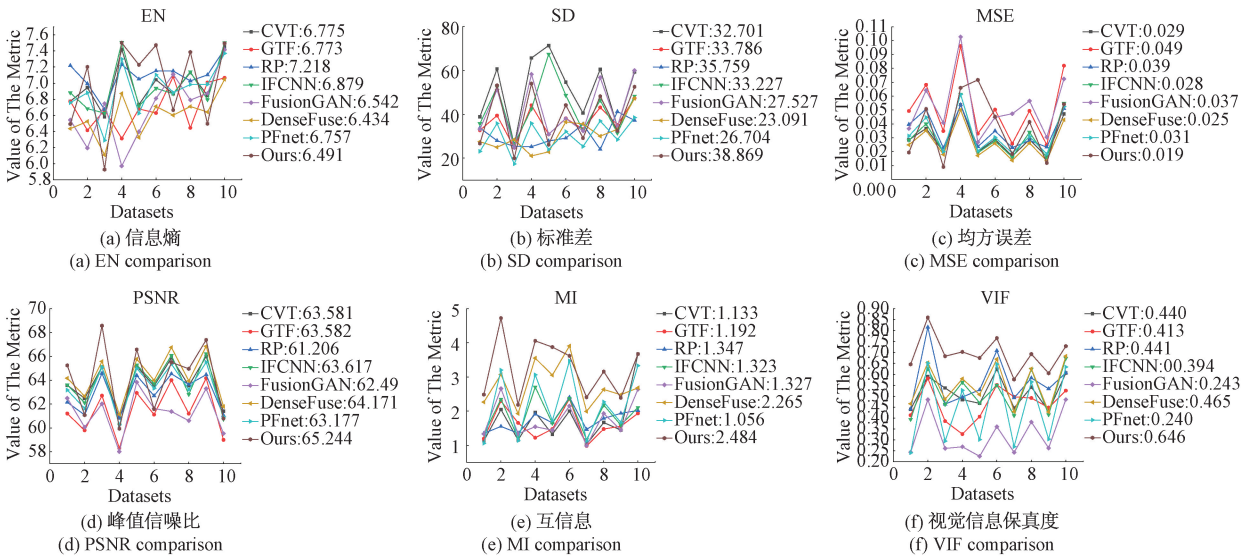
图 9　10 对偏振源图像

Fig. 9　10 pairs of polarization source images

表 2　不同融合方法融合结果的平均定量客观评价

Table 2　Average quantitative objective evaluation of fusion results of different fusion methods

	EN	SD	MSE	PSNR	MI	VIF
CVT	6.944	36.891	0.030	64.011	1.529	0.517
RP	7.101	43.195	0.351	62.943	1.712	0.567
GTF	6.707	30.723	0.051	61.716	1.578	0.463
IFCNN	6.966	42.013	0.032	63.916	1.838	0.519
FusionGAN	6.985	30.748	0.072	61.483	1.647	0.322
DenseFusion	6.597	29.743	0.043	64.540	2.065	0.552
PFnet	6.917	37.12	0.031	63.619	2.079	0.443
Ours	**7.886**	**51.186**	**0.027**	**68.162**	**2.842**	**0.694**



(a) 信息熵　(a) EN comparison
(b) 标准差　(b) SD comparison
(c) 均方误差　(c) MSE comparison
(d) 峰值信噪比　(d) PSNR comparison
(e) 互信息　(e) MI comparison
(f) 视觉信息保真度　(f) VIF comparison

图 10　10 对偏振图像评价指标

Fig. 10　10 pairs of polarization image evaluation metrics

从图 11 的融合结果中可以看出,W/O G(Ch)融合结果中保留了更多的 *DoLP* 图像的信息,而 S_0 中的场景信息保留较少,W/O G(SP) 融合结果中则保留了 S_0 中的语义信息,忽略了 *DoLP* 中的纹理信息,W/O G(AGR)和 W/O G(Att) 融合结果中出现了背景较暗,纹理信息丢失等问题。在鉴别器的消融实验中如图中红框部分所

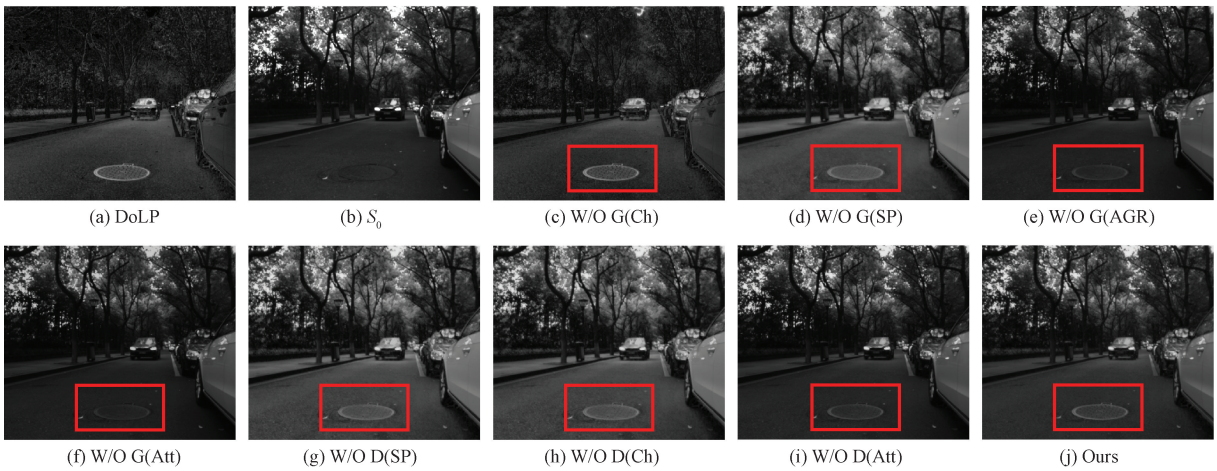


图 11　消融实验结果

Fig. 11　Results of ablation experiments

示没有注意力机制的引导或者缺少一种注意力机制的引导都会导致场景信息与纹理信息之间不平衡,本文网络结构能够以最优的比例将 S_0 和 *DoLP* 图像信息转移到融合图像中,融合结果既符合人眼的视觉感受,又能最大限度的保证纹理信息的不丢失。从表 3 可以看出,客观评价指标数据值的大小和分布与主观评价一致,本文融合结果的各项评价指标均取得最佳值,可以验证本文提出的融合网络结构能够提升整体融合效果。

表 3　消融实验评价指标

Table 3　Evaluation indicators for ablation experiments

	EN	SD	MSE	PSNR	MI	VIF
w/o G(Ch)	5.144	29.790	0.041	60.628	2.429	0.416
w/o G(SP)	6.101	30.142	0.032	62.245	2.597	0.607
w/o G(AGR)	4.764	25.323	0.056	51.946	1.678	0.475
w/o D(Att)	5.986	22.958	0.054	50.887	1.869	0.419
w/o D(SP)	6.054	32.246	0.022	61.532	2.667	0.612
w/o D(Ch)	6.582	32.843	0.019	64.540	2.756	0.632
w/o D(Att)	5.427	27.126	0.031	58.879	2.289	0.549
Ours	**6.610**	**35.618**	**0.015**	**65.714**	**2.996**	**0.667**

3　结　论

本文提出了一种基于双重注意力机制生成对抗网络,融合强度图像和偏振度图像融合。为了更好的保留偏振度图像的纹理信息和强度图像的背景信息,在生成器的编码器中采用了密集块结构深度提取特征,在融合模块中使用了注意力机制(通道和空间)和梯度残差结构,聚焦图像的强度分布和纹理信息。在解码器中通过跳跃连接方式引入编码器中的特征图,增加特征图的利用率,在鉴别器中引入了注意力机制(通道和空间),生成器不断博弈,迭代优化网络参数,激励生成器的输出既能保留偏振度图像的稀疏特征又不损失强度图像信息,实验结果表明,本文提出融合方法在整体亮度、局部对比度、信息丰富度和视觉感知等方面均能得到稳定的最优结果,在主观视觉评价和客观指标评价方面均优于已有文献中具有代表性的融合方法。

参考文献

[1] 杨彬,黄润才,王从澳. 基于鲁棒主成分分析法的红外可见光图像融合[J]. 制造业自动化,2022,44(6):4-7,16.
YANG B, HUANG R C, WANG C AO. Infrared and visible light image fusion based on robust principal component analysis [J]. Manufacturing Automation, 2022,44(6):4-7,16.

[2] GAO C, SONG C, ZHANG Y, et al. Improving the performance of infrared and visible image fusion based on latent low-rank representation nested with rolling guided image filtering [J]. IEEE Access, 2021, 9: 91462-91475.

[3] WANG Z, CUI Z, ZHU Y. Multi-modal medical image

fusion by Laplacian pyramid and adaptive sparse representation[J]. Computers in Biology and Medicine, 2020, 123: 103823.

[4] LIU Y, WANG L, CHENG J, et al. Multi-focus image fusion: A survey of the state of the art[J]. Information Fusion, 2020, 64: 71-91.

[5] 姜文斌,孙学宏,刘丽萍. 基于离散小波变换和梯度锐化的遥感图像融合算法[J]. 电光与控制, 2020, 27(5):47-51.
JIANG W B, SUN X B, LIU L P. A remote sensing image fusion algorithm based on DWT and gradient sharpening[J]. Electronics Optics & Control, 2020, 27(5):47-51.

[6] 王少杰,潘晋孝,陈平. 基于双树复小波变换的图像融合[J]. 核电子学与探测技术, 2015, 35(7): 726-728,737.
WANG SH J, PAN J X, CHEN P. Image Fusion Based on Dual-Tree Complex Wavelet Transform [J]. Nuclear Electronics & Detection Technology. 2015, 35(7): 726-728, 737.

[7] PENG H, LI B, YANG Q, et al. Multi-focus image fusion approach based on CNP systems in NSCT domain [J]. Computer Vision and Image Understanding, 2021, 210: 103228.

[8] JOSE J, GAUTAM N, TIWARI M, et al. An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion[J]. Biomedical Signal Processing and Control, 2021, 66: 102480.

[9] ZHANG Y, LIU Y, SUN P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.

[10] HOU R, ZHOU D, NIE R, et al. VIF-Net: An unsupervised framework for infrared and visible image fusion [J]. IEEE Transactions on Computational Imaging, 2020, 6: 640-651.

[11] ZHANG J, SHAO J, CHEN J, et al. Polarization image fusion with self-learned fusion strategy [J]. Pattern Recognition, 2021, 118: 108045.

[12] 陈广秋,温奇璋,尹文卿,等. 用于红外与可见光图像融合的注意力残差密集融合网络[J]. 电子测量与仪器学报, 2023, 37(8):182-193.
CHEN G Q, WEN Q ZH, YIN W Q, et al. Attentional residual dense connection fusion network for infrared and visible image fusion [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37 (8): 182-193.

[13] LI J, HUO H, LI C, et al. AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks [J]. IEEE Transactions on Multimedia, 2020, 23: 1383-1396.

[14] MA J, LIANG P, YU W, et al. Infrared and visible image fusion via detail preserving adversarial learning [J]. Information Fusion, 2020, 54: 85-98.

[15] ZHANG J, SHAO J, CHEN J, et al. PFNet: an unsupervised deep network for polarization image fusion[J]. Optics letters, 2020, 45(6): 1507-1510.

[16] WANG T, CHENG J, LIU T, et al. Visible Light Polarization Image Fusion based on Dense Connection Generative Adversarial Network[C]. Journal of Physics: Conference Series. IOP Publishing, 2021, 2035(1): 012028.

[17] LIU J, DUAN J, HAO Y, et al. Semantic-guided polarization image fusion method based on a dual-discriminator GAN[J]. Optics Express, 2022, 30(24): 43601-43621.

[18] LI N, ZHAO Y, PAN Q, et al. Demosaicking DoFP images using Newton's polynomial interpolation and polarization difference model[J]. Optics express, 2019, 27(2): 1376-1391.

[19] LI H, WU X. DenseFuse: A fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614-2623.

[20] LIN T, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]. Computer Vision-ECCV 2014: 13th European Conference,. Springer International Publishing, 2014: 740-755.

[21] XIANG K, YANG K, WANG K. Polarization-driven semantic segmentation via efficient attention-bridged fusion[J]. Optics Express, 2021, 29: 4802-4820.

[22] QIU S, FU Q, WANG C, et al. Linear polarization demosaicking for monochrome and colour polarization focal plane arrays[C]. Computer Graphics Forum. 2021, 40(6): 77-89.

[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, 31100-109.

[24] Karim S, Tong G, Li J, et al. Current advances and future perspectives of image fusion: A comprehensive review[J]. Information Fusion, 2023, 90: 185-217.

[25] 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.

[26] RAO D, XU T, Wu X. Tgfuse: An infrared and visible image fusion approach based on transformer and generative adversarial network[J]. IEEE Transactions on Image Processing, 2023.

[27] LI H, ZHAO J, LI J, et al. Feature dynamic alignment and refinement for infrared-visible image fusion: Translation robust fusion[J]. Information Fusion, 2023, 95: 26-41.
[28] ZHANG X, DEMIRIS Y. Visible and infrared image fusion using deep learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
[29] LIAO H, JIANG Q, JIN X, et al. MUGAN: thermal infrared image colorization using mixed-skipping UNet and generative adversarial network [J]. IEEE Transactions on Intelligent Vehicles, 2022.
[30] YU R, CHEN W, ZHU B. Infrared and visible image fusion algorithm based on a cross-layer densely connected convolutional network [J]. Applied Optics, 2022, 61(11): 3107-3114.
[31] WANG J, XI X, LI D, et al. FusionGRAM: An Infrared and Visible Image Fusion Framework Based on Gradient Residual and Attention Mechanism [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12.

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. He is now an associate professor in 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.

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.

张晨洁,2004 年于长春理工大学获得学士学位,2008 年于长春理工大学获得硕士学位,现为长春理工大学副教授,主要研究方向为模式识别与智能信息处理。

Zhang Chenjie received her B. Sc. degree from Changchun University of Technology in 2004, M. Sc. degree from Changchun University of Technology in 2008. She is now an associate professor in Changchun University of Technology. Her main research interests include pattern recognition and intelligence information processing.

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. He is now 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.