

基于双支路拮抗网络的偏振方向图像融合方法

风 瑞¹, 袁宏武^{2,3}, 周玉叶¹, 王 峰³

(1. 安徽建筑大学 电子与信息工程学院, 安徽 合肥 230601; 2. 安徽新华学院 大数据与人工智能学院, 安徽 合肥 230088;
3. 偏振光成像技术安徽省重点实验室, 安徽 合肥 230031)

摘要: 为了提升偏振方向图像融合效果, 构建了一种偏振方向图像的双支路拮抗融合网络 (Double-branch Antagonism Network, DANet), 该网络主要包括特征提取、特征融合和特征转化3个模块。首先, 特征提取模块由低频支路和高频支路组成, 将 0° 、 45° 、 90° 和 135° 偏振方向图像连接输入到低频支路, 提取图像能量特征, 将2组拮抗图像差分输入到高频支路, 提取图像细节特征; 其次, 将得到的能量特征和细节特征进行特征融合; 最后, 将融合后的特征转化整合为融合图像。实验表明, 通过此网络得到的融合图像, 其视觉效果和评价指标均有较为显著的提升, 与合成强度图像 I 、偏振拮抗图像 S_d 、 S_{dd} 、 S_h 、 S_v 相比, 在平均梯度、信息熵、空间频率和图像灰度均值上, 分别至少提升了22.16%、9.23%、23.44%和38.71%。

关键词: 图像融合; 深度学习; 偏振图像

中图分类号: TP391

文献标识码: A

文章编号: 1001-8891(2024)03-0288-07

Fusion Method for Polarization Direction Image Based on Double-branch Antagonism Network

FENG Rui¹, YUAN Hongwu^{2,3}, ZHOU Yuye¹, WANG Feng³

(1. School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China;

2. School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei 230088, China;

3. Key Laboratory of Polarization Imaging Detection Technology in Anhui Province, Hefei 230031, China)

Abstract: To improve the quality of the fused image, the study presents a double-branch antagonism network (DANet) for the polarization direction images. The network includes three main modules: feature extraction, fusion, and transformation. First, the feature extraction module incorporates low and high-frequency branches, and the polarization direction images of 0° , 45° , 90° , and 135° are concatenated and imported to the low-frequency branch to extract energy features. Two sets of polarization antagonism images (0° , 90° , 45° , and 135°) are subtracted and entered into the high-frequency branch to extract detailed features and energy. Detailed features are fused to feature maps. Finally, the feature maps were transformed into fused images. Experiment results show that the fusion images obtained by DANet make obvious progress in visual effects and evaluation metrics, compared with the composite intensity image I , polarization antagonistic image S_d , S_{dd} , S_h , and S_v , the average gradient, information entropy, spatial frequency, and mean gray value of the image are increased by at least 22.16%, 9.23%, 23.44% and 38.71%, respectively.

Keywords: image fusion, deep learning, polarization image

0 引言

偏振成像^[1]是一种新型的光学成像手段, 能够获

取多个偏振方向图像, 通过偏振信息解析, 能够实现同场景多个不同偏振方向图像来表征目标的偏振信息, 从而实现对目标的检测。偏振信息解析是偏振成

收稿日期: 2022-12-01; 修订日期: 2023-03-17.

作者简介: 风瑞 (1998-), 女, 安徽六安人, 硕士研究生, 主要研究方向为图像处理、深度学习。E-mail: rui_feng98@163.com.

通信作者: 袁宏武 (1979-), 男, 安徽六安人, 副教授, 博士, 中国计算机学会专业会员, 主要研究方向为偏振图像处理、数据分析与挖掘和基于深度学习理论的应用。E-mail: yuanhongwu@axhu.edu.cn.

基金项目: 国家自然科学基金资助项目 (61906118); 安徽省自然科学基金资助项目 (2108085MF230); 偏振光成像技术安徽省重点实验室开放基金 (KFJJ-2020-2)

像中的重要环节，目前有很多学者针对不同的应用场景提出了多种偏振信息解析的方法^[2-5]，其中，沈洁^[6]等人根据螳螂虾的复眼能充分利用偏振信息实现水下复杂环境的猎物捕获，提出了基于拮抗机制的偏振仿生信息解析方法，将 0° 、 45° 、 90° 、 135° 四个偏振方向图像分成 2 组拮抗图像，通过使 2 组图像的拮抗信息熵最大来获得偏振方向图像的加权系数，得到偏振拮抗图像，从而实现水下目标探测。但这种方法本质是偏振方向图像间的线性操作，存在计算效率低、解析结果不确定、目标不够突出等问题。

深度学习具有强大的表征能力，可以从高维、复杂以及非线性数据中提取有用的特征，目前在很多领域都有着广泛的应用，如自然语言处理^[7]、语音识别^[8]、图像处理^[9]等。Li^[10]等人提出一种基于卷积神经网络 (Convolutional Neural Networks, CNN) 和残差神经网络 (Residual Networks, ResNets) 的深度学习架构，用于红外和可见光图像的融合，较好地突显了目标信息，并解决了传统的基于 CNN 中随着网络深度增加特征信息退化的问题。

偏振图像拮抗过程属于偏振方向图像融合范畴，本文充分利用深度学习在图像处理上的优势，提出了一种偏振方向图像的双支路拮抗融合网络，主要包括特征提取、特征融合和特征转化 3 个模块，输入 4 个不同方向的偏振图像，分成两个支路，低频支路通过合成图像来减少能量的损失，高频支路通过差分图像来突显图像的细节信息。将两个支路处理的结果分别通过深度融合网络进行处理，获取效果更好的融合图像，提高后续目标检测与识别效果。

1 基本原理

1.1 分焦平面型偏振成像

偏振成像需要获取多个偏振方向图像，常用的获取方式可以分为：分时型、分振幅型、分孔径型和分焦平面型。相对于其他的成像方式，分焦平面型偏振成像具有体积小、重量轻、成本低、能同时获取多个偏振方向图像等优点，成为目前偏振成像获取方法的主流。其基本原理是：在探测器的芯片上集成微型偏振分析器 (如图 1 所示)，在探测器的每 4 个像元上，耦合 0° 、 45° 、 90° 、 135° 等 4 个线偏振方向的偏振分析器，利用 4 个像元实现 1 个像素的偏振信息采集。按照微型偏振分析器排列顺序，对探测器输出图像进行重新整合，即可得到 0° 、 45° 、 90° 、 135° 四个偏振方向图像，如图 2(a)~(d) 所示。

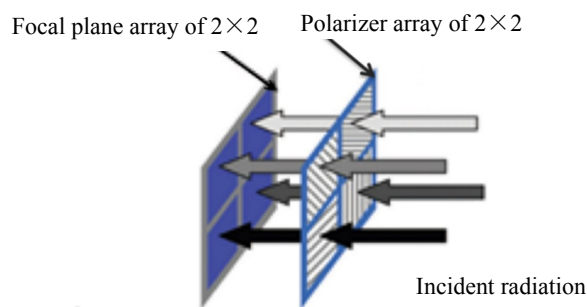


图1 分焦平面型偏振成像原理

Fig.1 Polarization image split-focus planar principle

利用偏振成像理论，可以得到合成强度图像 I ^[6]，如图 2(e) 所示：

$$I = I(0^\circ) + I(90^\circ) \quad (1)$$

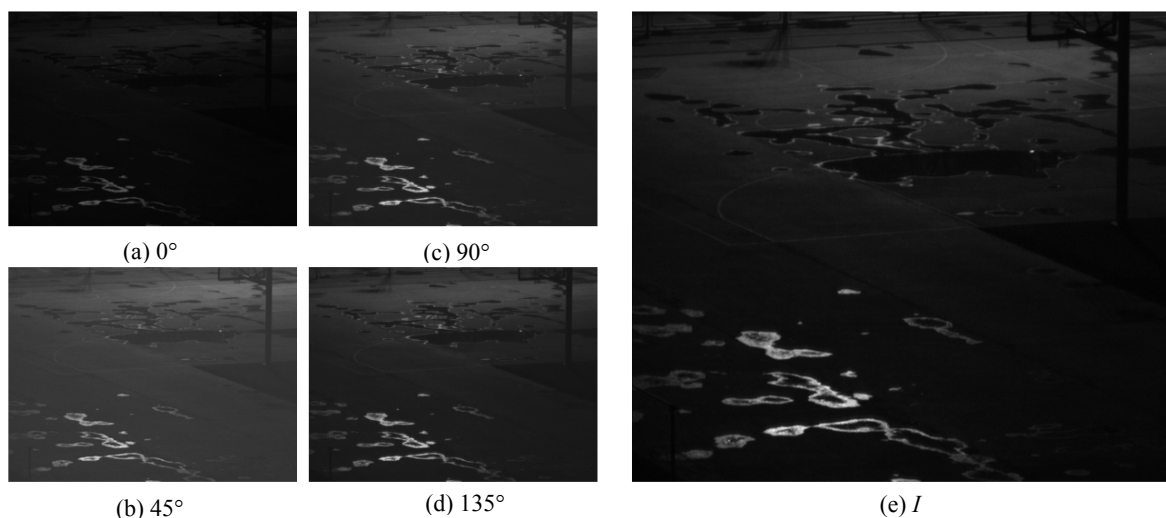


图2 0° 、 45° 、 90° 和 135° 的偏振方向图和合成强度图

Fig.2 Polarization and intensity pictures of 0° , 45° , 90° and 135°

1.2 仿生偏振拮抗图像

根据螳螂复眼的偏振拮抗机制^[6], 一对正交偏振图像的输入可以形成一个拮抗, 如 0° 和 90° 偏振方向图像、 45° 和 135° 偏振方向图像, 将采集到的 4 组正交偏振图像形成 4 个偏振拮抗通道, 每个通道由相互正交的一对偏振信号组成, 各组拮抗信号通过拮抗运算方式可以得到偏振拮抗图像, 如下所示:

$$S_d = k_1 \times I(45^\circ) - k_2 \times I(135^\circ) \quad (2)$$

$$S_{dd} = k_3 \times I(135^\circ) - k_4 \times I(45^\circ) \quad (3)$$

$$S_h = k_5 \times I(0^\circ) - k_6 \times I(90^\circ) \quad (4)$$

$$S_v = k_7 \times I(90^\circ) - k_8 \times I(0^\circ) \quad (5)$$

式中: $I(0^\circ)$ 、 $I(90^\circ)$ 、 $I(45^\circ)$ 和 $I(135^\circ)$ 分别表示 0° 、 90° 、 45° 和 135° 的偏振方向图像, $k_i (i=1, \dots, 8)$ 为拮抗系数, 起到对图像的增强和抑制作用, k 的取值范围由人为设定, $k_m \geq 1 (m=1, 3, 5, 7)$, $0 < k_n \leq 1 (n=2, 4, 6, 8)$, 文献^[6]中 k 是通过遍历所有范围内可能的值, 求取偏振拮抗图像信息熵最大来确定, 存在计算效率低、结果不确定的问题, 使得求取的偏振拮抗图像目标可能不够突出。

1.3 基于深度学习的图像融合

图像融合就是通过处理不同传感器所拍摄的源图像, 提取有用的信息或特征, 将其整合来改善图像的品质和清晰度^[11]。传统图像融合需要人工提取特征, 指定融合规则, 而基于深度学习的图像融合, 利用深度网络对输入图像进行卷积, 提取出目标高层特征, 再利用卷积将融合后特征转换为融合图像, 如图 3 所示。相对于可监督图像融合的方法, 无监督的方法通过约束融合图像和原图像之间的相似性, 克服了

大多数图像融合中无参考度量的普遍问题。

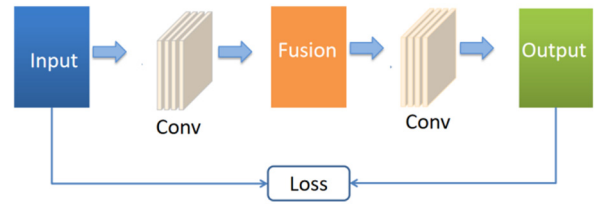


图3 深度学习图像融合网络一般过程

Fig.3 General process of deep learning image fusion network

2 DANet 网络设计

2.1 网络结构

由偏振成像机理可知, 每个偏振方向图像能量损失一半。为了提高融合后的图像的亮度, 如图 4 所示, 我们设计了一个低频支路, 将 4 个偏振方向图像通过 Concat 操作进行连接输入, 用于提取每个偏振方向图像的低频特征; 根据 Tyo^[12]的研究结论, 偏振差分成像可以突显目标细节信息, 因此本文设计了另一个高频支路, 将 2 组拮抗图像进行差分输入, 用于提取差分图像的目标高频特征。Huang^[13]等人提出了一种密集块结构, 其中使用了从任何层到所有后续层的直接连接。这种体系结构可以保存尽可能多的信息, 该模型可以改善网络中的信息流和梯度, 使网络易于训练, 同时, 密集连接具有正则化效果, 减少了任务的过拟合。受此启发, 本文将密集连接加入到差分图像细节特征提取中, 用于降低细节信息的损失。设计的网络结构主要包括特征提取、特征融合和特征转化 3 个模块。

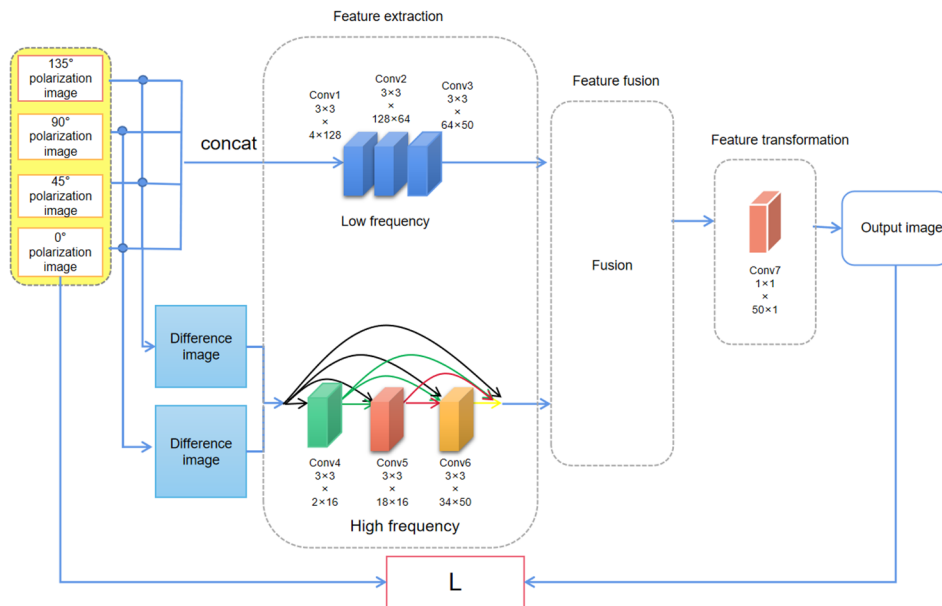


图4 偏振拮抗融合网络

Fig.4 Polarization antagonistic fusion network

图 4 中，特征融合模块将两个支路提取的特征图进行对应像素融合，得到融合特征，特征转化模块利用 1×1 卷积将融合的特征整合得到输出图像。低频和高频支路均有 3 个 3×3 的卷积层，网络参数如表 1 所示。

表 1 网络参数
Table 1 Network parameters

	Layer		Input channel	Output channel
Feature extraction	Low frequency	Conv1	4	128
		Conv2	128	64
		Conv3	64	50
	High frequency	Conv4	2	16
		Conv5	18	16
		Conv6	34	50
Feature fusion	Fusion		50	50
Feature transformation	Conv7		50	1

2.2 损失函数

在偏振拮抗图像获取中，利用信息熵最大作为评价标准^[6]，因此本文将信息熵损失 L_{entropy} 加入到损失函数中，用于增大融合图像的信息量；结构相似性度量方法作为图像领域使用最广泛的指标之一，该方法基于图像的亮度、对比度和结构 3 部分来衡量图像之间的相似性，本文将结构相似性损失 L_{ssim} 加入到损失函数中，用于保持融合图像结构特征；感知损失通常用于图像重建中，恢复出来的图像视觉效果较好，本文将感知损失 $L_{\text{perceptual}}$ 加入到损失函数中，用于提高融合图像的人眼视觉效果。因此，本文的损失函数 L 计算公式如下：

$$L = L_{\text{entropy}} \times \sigma + L_{\text{ssim}} \times \beta + L_{\text{perceptual}} \times \gamma \quad (6)$$

式中： σ 、 β 、 γ 为各损失的权重，本文分别取 0.1、10、0.1。

1) 信息熵损失 L_{entropy}

信息熵 (Information Entropy, IE) 越大，图像包

含的信息越多，为使融合图像信息熵最大，信息熵损失 L_{entropy} 定义为：

$$L_{\text{entropy}} = \frac{1}{\text{IE} + \varepsilon} \quad (7)$$

$$\text{IE} = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (8)$$

式中： ε 为极小量； x_i 为随机变量； $p(x_i)$ 为输出概率函数； n 为灰度等级。

2) 结构相似性损失 L_{ssim}

结构相似性 (structural similarity index, SSIM) 用于度量两幅图像的相似度，结构相似性损失 L_{ssim} 定义为：

$$L_{\text{ssim}} = 1 - \text{SSIM}(\text{output}, I) \quad (9)$$

式中： $\text{SSIM}(\cdot)$ 表示结构相似性运算；output 为输出图像； I 为强度图像。

3) 感知损失 $L_{\text{perceptual}}$

$$L_{\text{perceptual}} = \|\Phi_i(\text{output}), \Phi_i(I)\|_2^2 \quad (10)$$

式中： $\Phi_i(\cdot)$ 为 VGG16 网络第 i 层特征图。

3 实验与分析

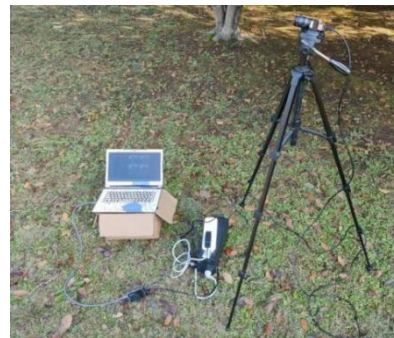
3.1 实验环境与数据

实验环境如下：训练与测试图像集采用处理器为 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30 GHz，系统运行内存为 32 GB，GPU 显卡为 8 GB 显存容量的 NVIDIA GeForce RTX3080 的图形工作站。训练软件运行环境为 Python3.9，编程框架为 Torch 1.11.0，搭建 Cuda10.0 用于实验加速。

本文采用 LUCID 公司的分焦平面型偏振相机 (型号: PHX050S-P)，如图 5 所示。该相机能够同时获取 0° 、 45° 、 90° 、 135° 四个偏振方向图像，图像分辨率为 2448×2048 。我们拍摄采集了多种场景目标的 9320 组偏振方向图像，构建出本文的数据集。



(a) 偏振相机实物图
(a) Polarization camera image



(b) 实验场景图
(b) Experimental scene image

图 5 LUCID 的分焦平面型偏振相机

Fig.5 LUCID's split-focus planar polarization camera

3.2 评价指标及训练参数

本文采用主观定性和客观定量相结合的方法对融合图像进行综合评价,主观评价主要通过人眼观察图像亮度和细节信息,客观评价采用平均梯度^[14]、信息熵^[15]、空间频率^[16]、均值^[17]等4个评价指标,定量评估融合效果。

1) 平均梯度 (Average Gradient, AG)

平均梯度能有效反映出图像层次信息,其值越大,图像层次越丰富,其计算公式为:

$$AG = \frac{1}{(M-1)(N-1)} \times \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{\frac{[F(i,j) - F(i+1,j)]^2 + [F(i,j) - F(i,j+1)]^2}{2}} \quad (11)$$

式中: $F(i, j)$ 为图像的第 i 行、第 j 列的灰度值; M 、 N 分别为图像的总行数和总列数。

2) 空间频率 (Spatial Frequency, SF)

空间频率是图像质量经典的标准之一,其值越大,代表图像质量越高,越清晰,其计算公式为:

$$SF = \sqrt{RF^2 + CF^2} \quad (12)$$

$$RF = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=2}^N [I_p(i, j) - I_p(i, j-1)]^2} \quad (13)$$

$$CF = \sqrt{\frac{1}{MN} \sum_{i=2}^M \sum_{j=1}^N [I_p(i, j) - I_p(i-1, j)]^2} \quad (14)$$

式中: RF 是行频率; CF 是列频率; M 、 N 为图片的宽高; $I_p(i, j)$ 为图像在 (i, j) 处的像素值。

3) 图像均值 (Image Mean, IM)

均值即图像像素的平均值,反应图像的平均亮度,平均亮度越大,能量越高,其计算公式为:

$$IM = \sum_{k=0}^{L-1} z_k p\left(\frac{n_k}{MN}\right) \quad (15)$$

式中: z_k 为图像的第 k 个灰度级; L 表示图像的灰度等级数目; n_k 是 z_k 在图像中出现的次数。

本文数据集共计 9320 组,其中训练集和测试集按照 9:1 划分,算法的模型由 Adam optimizer 训练,训练轮次为 20,初始学习率为 $1e^{-4}$,每训练 4 轮学习率衰减一半,详细参数如表 2 所示。

表 2 训练参数

Table 2 Training parameters

Parameters	Values
Training set	8388
Testing set	932
Training round	20
Epoch	4
Optimizer	Adam
Activation function	ReLU
Initial learning rate	$1e^{-4}$
Learning rate decay rate	$0.5 * lr / 4 \text{ round}$

3.3 实验结果分析

为了验证本文算法的有效性,从测试集中随机选取了 4 组数据,每组数据包含 0° 、 45° 、 90° 和 135° 偏振方向图像,第 1 组为室内沙地伪装板目标,第 2 组为室内标定装置目标,第 3 组为室外草地伪装板,第 4 组为水下珊瑚目标,如图 6 所示。

将上述数据输入到本文训练好的模型中,得到对应的融合图像,根据公式(1)~(5)分别得到合成强度图像 I 、偏振拮抗图像 S_d 、 S_{dd} 、 S_h 、 S_v ,如图 7 所示。

从图 7 可以看出,本文的融合图像亮度最高,能量最大,说明网络中的低频支路对图像能量的提升效果明显,有效解决了偏振成像中能量降低的问题。从图像细节来说,融合图像的细节效果有较为明显的提升,如第 1 组本文融合图像中的沙粒更加有颗粒感,伪装板的边缘更加突出,第 2 组本文融合图像中标定装置的线缆显现出来,背景板的线条更加清晰,而其他图像不太明显,第 3 组本文融合图像的 3 块伪装板全部从背景中区分出来,而其他图像只有部分能够区分开,第 4 组本文融合图像中,珊瑚整体目更亮,珊瑚边缘也更明晰。由于图像能量提升较为明显,可能会造成目标的对比度有所下降,如第一组本文融合图像的伪装板的对比度相较于 S_d 图和 S_h 图有所下降,但不影响目标整体的检测效果。

本文将测试集中 932 组图像输入到模型中,得到对应的 932 幅融合图像,并根据公式(1)~(5)分别得到 932 幅合成强度图像 I 和相应的偏振拮抗图像 S_d 、 S_{dd} 、 S_h 、 S_v 。利用平均梯度 (AG)、信息熵 (IE)、空间频率 (SF) 和图像灰度均值 (IM) 指标对其计算均值并进行评价,如表 3 所示。

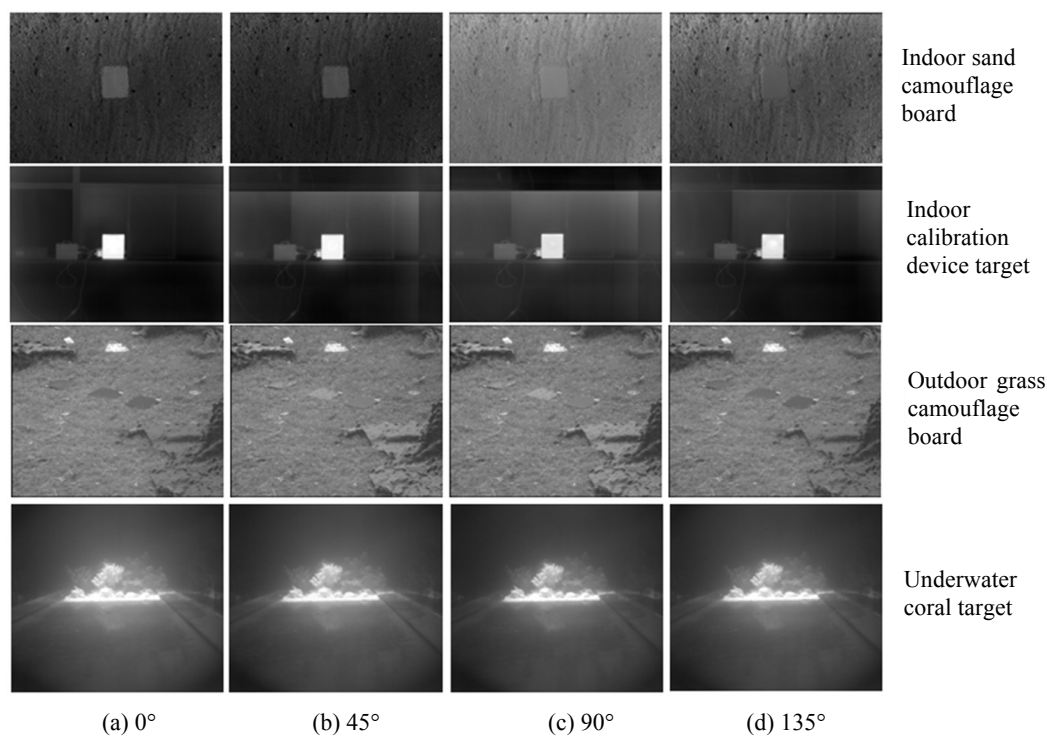


图 6 输入的 0°、45°、90°和 135°偏振方向图像

Fig.6 0°, 45°, 90° and 135° polarization direction images of input

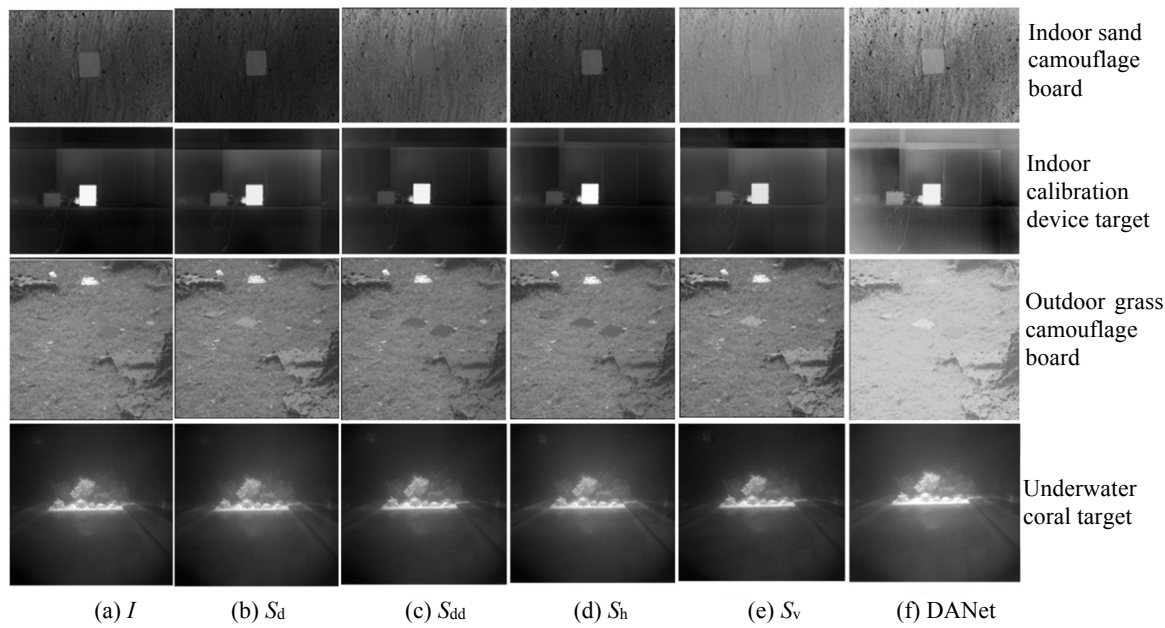


图 7 结果对比图

Fig.7 Results comparison diagram

表 3 输出结果的各项评价指标

Table 3 Evaluation indexes of the output results						
	I	S_d	S_{dd}	S_h	S_v	DANet
\overline{AG}	0.0099	0.0128	0.0119	0.0144	0.0126	0.0185
\overline{IE}	6.06	6.18	6.08	6.15	6.39	7.04
\overline{SF}	0.35	0.49	0.40	0.46	0.45	0.64
\overline{IM}	41	49	47	46	57	93

从表3中可以看出,在4个评价指标上,本文的方法都是最高的,在平均梯度上最少提高了22.16%,最多提高了46.49%;在信息熵上最少提高了9.23%,最多提高了13.92%;在空间频率上最少提高了23.44%,最多提高了45.31%;在图像灰度均值上最少提高了38.71%,最多提高了55.91%。实验结果表明,本文方法得到的融合图像亮度更高,包含的信息量更丰富,可以显现出更多的细节信息。

4 结论

针对偏振方向图像融合效果不明显的问题,本文提出了一种基于双支路拮抗融合网络的偏振信息解析方法,可以解决现有基于拮抗机制的仿生偏振信息解析方法中存在计算效率低、解析结果不确定、目标不够突出等问题,为偏振信息解析提供了一个新的技术途径。本文设计的DANet主要包括特征提取、特征融合和特征转化3个模块。首先,特征提取模块由低频支路和高频支路组成,将 0° 、 45° 、 90° 和 135° 偏振方向图像连接输入到低频支路,提取能量特征,将2组拮抗图像差分输入到高频支路,提取图像细节特征;其次,将得到的能量特征和细节特征进行特征融合;最后,将融合后的特征转化整合为融合图像。实验表明,通过DANet得到的融合图像在视觉效果和评价指标上均有较为显著提升,在平均梯度、信息熵、空间频率和图像灰度均值上分别至少提升了22.16%、9.23%、23.44%、38.71%。下一步,我们将进一步优化网络结构,以平衡能量支路和细节支路,改善融合图像的对比度;加大水下偏振方向图像在数据集中的比重,优化数据集。

参考文献:

- [1] 周强国,黄志明,周炜,等.偏振成像技术的研究进展及应用[J].红外技术,2021,43(9):817-828.
ZHOU Qiangguo, HUANG Zhiming, ZHOU Wei, et al. Research progress and application of polarization imaging technology[J]. *Infrared Technology*, 2021, 43(9): 817-828.
- [2] HU H, ZHANG Y, LI X, et al. Polarimetric underwater image recovery via deep learning[J]. *Optics and Lasers in Engineering*, 2020, 133: 106152.
- [3] WEI Y, HAN P, LIU F, et al. Enhancement of underwater vision by fully exploiting the polarization information from the Stokes vector[J]. *Optics Express*, 2021, 29(14): 22275-22287.
- [4] DING X, WANG Y, FU X. Multi-polarization fusion generative adversarial networks for clear underwater imaging [J]. *Optics and Lasers in Engineering*, 2022, 152: 106971.
- [5] 寻华生,张晶晶,刘晓,等.基于偏振图像的低照度场景多目标检测算法[J].红外技术,2022,44(5):483-491.
XUN Huasheng, ZHANG Jingjing, LIU Xiao, et al. Multi-object detection algorithm for low-illumination scene based on polarized image[J]. *Infrared Technology*, 2022, 44(5): 483-491.
- [6] SHEN J, WANG H, CHEN Z, et al. Polarization calculation and underwater target detection inspired by biological visual imaging[J]. *Sensors & Transducers*, 2014, 169(4): 33.
- [7] Otter D W, Medina J R, Kalita J K. A survey of the usages of deep learning for natural language processing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(2): 604-624.
- [8] Nassif A B, Shahin I, Attili I, et al. Speech recognition using deep neural networks: A systematic review[J]. *IEEE Access*, 2019, 7: 19143-19165.
- [9] JIAO L, ZHAO J. A survey on the new generation of deep learning in image processing[J]. *IEEE Access*, 2019, 7: 172231-172263.
- [10] LI H, WUX J. DenseFuse: A fusion approach to infrared and visible images [J]. *IEEE Transactions on Image Processing*, 2018, 28(5): 2614-2623.
- [11] 王霞,赵家碧,孙晶,等.偏振图像融合技术综述[J].航天返回与遥感,2021,42(6):9-21.
WANG Xia, ZHAO Jiabi, SUN Jing, et al. A review of polarized image fusion techniques[J]. *Space Return and Remote Sensing*, 2021, 42(6): 9-21.
- [12] Tyo J S, Rowe M P, Pugh E N, et al. Target detection in optically scattering media by polarization-difference imaging[J]. *Applied Optics*, 1996, 35(11): 1855-1870.
- [13] HUANG G, LIU Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700-4708.
- [14] SHEN Yu, WU Zhongdong, WANG Xiaopeng, et al. Tetrolet transform images fusion algorithm based on fuzzy operator[J]. *Journal of Frontiers of Computer Science and Technology*, 2015, 9(9): 1132-1138.
- [15] Eskicioglu A M, Fisher P S. Image quality measures and their performance[J]. *IEEE Transactions on communications*, 1995, 43(12): 2959 - 2965.
- [16] Roberts J W, Aardt J A van, 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.
- [17] JIN Xin, NIE Rencan, ZHOU Dongming, et al. Multifocus color image fusion based on NSST and PCNN[J/OL]. *Journal of Sensors*, 2016, <https://doi.org/10.1155/2016/8359602>.