

面向暗光场景的目标偏振/可见光融合检测方法

马如钺¹, 王晨光¹, 曹慧亮², 申冲², 唐军², 刘俊²

(1. 中北大学 信息与通信工程学院, 太原 030051; 2. 中北大学 仪器与电子学院, 太原 030051)

摘要: 为解决偏振暗光场景下常见目标识别结果准确性不高的问题, 提出了基于卷积神经网络的偏振度图像与可见光图像融合算法, 设计了新的损失函数以形成无监督学习过程, 引入拉普拉斯算子提高融合图像的质量, 最终将被测目标的偏振信息与可见光信息有效结合; 提出了基于改进的 YOLOv5 算法对融合后的目标进行目标检测, 在网络框架中加入 CA 注意力机制, 将通道注意力机制与空间注意力机制相结合; 在自制的数据集上对提出的网络进行训练测试, 结果表明, 融合图像在主客观上都达到了较好的视觉效果, 将改进的 YOLOv5 算法相比最优的 YOLOv5s 模型, 精确率和召回率分别达到了 89.3% 和 82.5%, 均值平均精度分别提高了 2.6% 和 1.8%。

关键词: 偏振成像; 目标检测; 图像融合; YOLOv5; 图像增强; 注意力机制

Target Polarization/Visible Light Fusion Detection Method for Dark Light Scenes

MA Ruyue¹, WANG Chengaung¹, CAO Huiliang², SHEN Chong², TANG Jun², LIU Jun²

(1. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China;

2. School of Instrument and Electronics, North University of China, Taiyuan 030051, China;)

Abstract: In order to solve the low accuracy of common target recognition results in polarized dark light scenes, this paper proposes a fusion algorithm based on convolutional neural network of polarization degree images and visible light images, designs a new loss function to form an unsupervised learning process, introduces Laplace operator to improve the quality of fused image, and finally the polarization information of the target to be measured is effectively combined with the visible light information; A fused target detection algorithm is proposed based on the improved YOLOv5 algorithm for the target detection of the fused target, the coordinate attention (CA) attention mechanism of the network framework is added to combine the channel attention mechanism with the spatial attention mechanism. The proposed network is trained and tested on a homemade dataset, the results show that the fused image achieves better visual effects subjectively and objectively, compared with the optimal YOLOv5s model, the precision and recall rate of the improved YOLOv5 algorithm reach 89.3% and 82.5%, respectively, and the mean average precisions increase by 2.6% and 1.8%, respectively.

Keywords: polarization imaging; target detection; image fusion; YOLOv5; image enhancement; attention mechanism

0 引言

偏振是光的重要物理特性, 它可以表征场景中任何物体反射的光波, 即使在弱光照或强反射的情况下, 偏振也能够描述物体的重要物理特性, 包括其表面几何结构、材料性质及其粗糙度。反射光的偏振态与物体的强度、形状和反射特性等物理特性高度相关^[1]。偏振成像技术是通过获取线偏振度 (DoLP, degree of linear polarization) 图像和偏振角 (AoP, angle of linear polarization) 图像来获得目标的内在属性, 作为一种先进的技术, 它不仅能获得目标

的偏振信息, 还能提供二维空间的光强分布特征, 在许多领域都有广泛的潜在应用, 包括目标探测^[2-3]、通信^[4]、水下探测^[5]和医疗成像^[6]等领域。

然而光的偏振特性容易受到环境 (如雾霾、阴雨、弱光) 的影响, 使得偏振图像质量难以达到适用水平。由于可见光图像和偏振图像往往具有互补特性, 研究人员通常采用将 DoLP 图像与 AOP 图像融合或将 DoLP 图像与可见光图像融合等方法来增加图像信息, 从而提高单幅偏振图像的多维细节信息。偏振图像融合方法主要分为传统融合

收稿日期: 2023-12-07; 修回日期: 2023-12-26。

基金项目: 国家自然科学基金 (61973281, 51821003, 51922009); 山西省重点研发计划项目 (202003D111003); 山西省优秀青年培育项目 (202103021222011); 光电信息控制和安全技术重点实验室基金 (2021JJCJQLB055010); 山西省量子传感与精密测量重点实验室 (201905D121001); 山西省 1331 工程项目。

作者简介: 马如钺 (1999-), 女, 硕士研究生。

通讯作者: 王晨光 (1981-), 男, 博士, 副教授。

引用格式: 马如钺, 王晨光, 曹慧亮, 等. 面向暗光场景的目标偏振/可见光融合检测方法[J]. 计算机测量与控制, 2024, 32(4): 46-53.

方法和神经网络融合方法两类, 目前较为常见且应用广泛的为神经网络融合方法。2021 年, 文献 [7] 提出了一种新型的自学习策略深度神经网络, 实验验证该方法在视觉质量和定量测量方面均优于几种最先进的方法; 2022 年, 文献 [8] 利用改进的小波模式最大值算法提取光强图像和偏振图像的边缘, 然后对两个边缘进行细化和融合得到最终的边缘信息, 最终可获得边缘更清晰的融合图像; 文献 [9] 提出了一种基于 Transformer 的深度神经网络, 对红外偏振图像的长程特征进行编码, 利用自注意力机制获取全局上下文信息, 以提高红外偏振图像融合的性能; 文献 [10] 通过像素信息引导和注意力机制提出了一种新型无监督偏振和可见光图像融合网络, 设计了损失函数来执行融合图像与源图像之间的像素分布约束, 展现出更丰富的偏振信息和更优良的亮度。这些融合方法都为后期偏振成像在目标检测等领域打下坚实的基础。

目标检测是计算机视觉领域内的热门研究课题, 其目的是在图像中定位出目标的位置并识别出目标物的类别, 近年来深度学习算法在目标检测中的广泛应用足以证明其优越性, 不仅可以高精度检测物体, 还在处理速率上达到质的飞跃。目前基于深度学习网络的目标检测算法主要分为两类: 第一类是两阶段目标检测算法, 例如 Fast R-CNN (Region-convolutional Neural Network)^[11]、Faster R-CNN^[12]、Mask R-CNN^[13]等, 这类算法将检测任务分为两步, 相比传统目标检测算法能够提取到更加丰富、深层的图像特征信息, 同时检测精度和检测速度都得到了大幅提升, 但两阶段算法仍然难以满足实时目标检测的要求; 因此第二类单阶段目标检测算法应运而生, 例如 SSD (Single Shot MultiboxDetector)^[14]、YOLO (You Only Look Once) 系列算法^[15-18]等, 此类算法直接将图像输入检测网络, 能够达到实时检测的要求, 因此单阶段算法成为目前热门的研究方向, 应用领域也更加广泛。文献 [19] 针对小目标物体提出了一种特征金字塔结构 RetinaNet, 有效提升了小目标的检测精度; 文献 [20] 提出了基于中心点的方法 CenterNet, 使用关键点估算来寻找中心点, 并对所有其他物体属性进行回归, 如大小、三维位置、方向甚至姿态等; 文献 [21] 使用单个卷积神经网络将物体边界框检测为一对关键点, 将目标检测分为图像分类和目标定位两个子任务, 并且引用了新型池化层, 解决遮挡和多目标问题。文献 [22] 将卷积模块替换为基于自注意力机制的 swinblock 结构, 将注意力机制计算限制在非重叠的局部窗口, 同时允许跨窗口连接, 获得较好的检测效果。

尽管目前的图像融合算法和目标检测算法已经取得了较好的性能, 但针对偏振目标的检测, 尤其是极端天气下的目标检测研究还不够充分, 同时目前也没有公开使用的偏振数据集, 这对偏振成像下的目标检测的研究有一定阻碍。为研究暗光条件下的目标检测, 提高基于偏振成像的目标检测精度, 本文首先构建了暗光场景下的偏振数据集, 提出了一种基于卷积神经网络的可见光图像与偏振图像融

合网络, 对损失函数进行了优化, 以更大程度地还原偏振图像细节信息。还引入了一种基于拉普拉斯算子的图像增强方法, 利用待测目标灰度与偏振图像中背景噪声灰度之间的差值来提高偏振图像的质量, 最终得到同时具有偏振信息和可见光信息的待测目标。在偏振图像增强的基础上, 基于 YOLOv5s 提出了一种轻量化自注意力机制的目标检测模型。通过引入 CA (Coordinate Attention) 注意力机制, 不仅同时引入空间注意力和通道注意力, 还考虑到长程依赖的问题, 有效实现了检测精度的提升。最后, 我们使用主观和客观的评估标准对融合算法以及目标检测算法进行评估, 并将提出的目标检测算法与几种经典目标检测算法进行比较, 使用多个评价指标来评估所提出的网络的可靠性和可用性。

1 图像融合网络

本文所提出的图像融合网络总体架构如图 1 所示, 旨在实现更好的图像融合效果。该模型基于卷积神经网络 (CNN, convolutional neural network), 主要由 3 部分组成: 编码器、融合模块和解码器。首先, 将可见光图像 I_S 和偏振图像 I_{DoLP} 同时输入编码器, 分别提取其图像特征, 然后由图像融合模块进行融合, 最后, 融合后的特征图进入解码器, 以便解码器重建最终的融合图像。在该网络结构中, 滤波器和步长分别为 3×3 和 1。在整个过程中, 图像大小不会改变, 并使用零填充操作来确保图像大小的一致性。表 1 为具体的网络配置参数表。接下来将详细介绍这 3 部分的结构与作用。

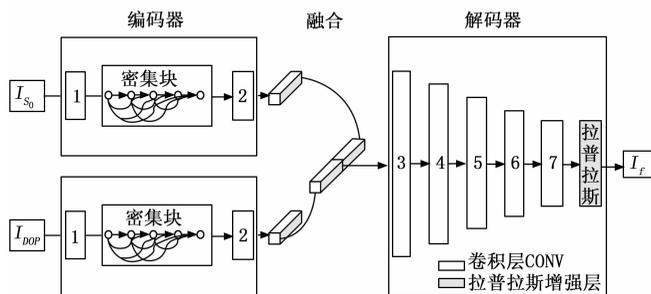


图 1 本文提出的图像融合网络架构

1.1 网络结构介绍

编码器部分: 首先, 源图像 I_S 和 I_{DoLP} 被送入一个 16 通道的卷积层, 紧接着通过 ReLU 激活函数^[23], 用于提取初始图像特征; 然后将输出输入 DenseBlock 模块^[24], 该模块包含 3 个卷积层, 每个卷积层后都有一个 ReLU 激活函数, 其结构如图 2 所示。使用 DenseBlock 结构可以提取更高维、更深层次的图像特征, 并减少图像的过度拟合, 这种结构允许输入任意大小的图像, 确保所有有效特征都能应用到融合策略中; 最后, 输出经过一个具有 64 个输出通道的卷积层, 以尽可能保留所有深层和浅层特征, 最终输出特征图 I_S^F 和 I_{DoLP}^F 并作为输入发送到融合模块。

表 1 网络配置参数

	层级	尺寸	输入	输出	激活函数
编码器	Conv	3	1	16	ReLU
	DenseBlock	3	16	16	ReLU
		3	32	16	ReLU
		3	48	16	ReLU
	Conv	3	64	16	ReLU
融合网络	Conv	3	64	128	—
解码器	Conv	3	64	128	ReLU
	Conv	3	128	64	ReLU
	Conv	3	64	32	ReLU
	Conv	3	42	16	ReLU
	Conv	3	16	1	—
	拉普拉斯	—	1	1	—

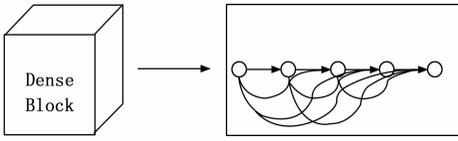


图 2 Dense Block 网络架构

融合网络部分：将两个特征图进行物理拼接，得到一个包含 128 个通道的融合特征图，然后将其作为解码器的输入。其中活动水平测量和融合规则在随后的卷积层中自主学习，无需人工设计。

解码器部分：该模块有 5 个主要卷积层，除最后一层外，所有卷积层都采用 ReLU 激活函数。输出通道数分别为 128、64、32、16 和 1，每层的学习能力 c 值^[25]大于 1/6， I_f^r 最上层卷积层的感受野大小为 19。由于使用了零填充操作，本文的网络没有使用上层采样层。此外，还引入了拉普拉斯增强层来增强融合图像，并输出最终的融合图像 I_f ，得到偏振信息与可见光信息相融合的待测目标信息。

在数字图像处理领域，拉普拉斯算子^[26]常用于图像增强、边缘检测和模糊判定等任务。它也是工程数学中常见的积分变换，具有旋转不变性，即无论图像如何旋转，得到的响应都是不变的。当图像灰度发生剧烈变化时，对其进行一阶微分会形成局部极值，而对其进行二阶微分会形成过零点，这个零点的一边会产生波峰，另一边会产生波谷，因此可以设置一个阈值来检测这个过零点。函数的一阶微分和二阶微分如图 3 所示。

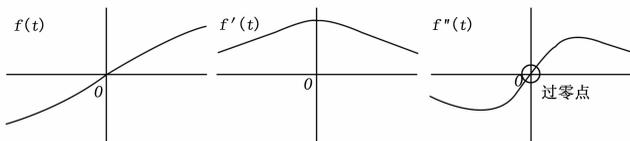


图 3 函数的一阶微分和二阶微分表示

二维图像的拉普拉斯变换是每个同相项的二阶导数，其定义如下：

$$\nabla^2(f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (1)$$

在数字图像处理中，平面图像可以看作是像素点在 x 和 y 两个方向上的集合排列，在分布上是离散的，因此拉普拉斯算子一般使用微分近似，如下式所示：

$$\begin{aligned} \nabla^2(f) &= f(x+1,y) + f(x-1,y) + \\ &f(x,y+1) + f(x,y-1) - 4f(x,y) \end{aligned} \quad (2)$$

拉普拉斯算子对原始图像进行变换时，保留了图像的高频成分，抑制了图像的低频成分，从而增强了图像灰度跳变处的对比度，增强了图像的细小细节部分，保留了图像的背景色调，使图像的细节比原始图像更加清晰。

1.2 损失函数

损失函数用于评估模型预测值与真实值之间的差异程度^[27]，差异值越小则表明模型的效果越好，因此损失函数的选择在卷积神经网络训练中非常重要。结构相似度 (SSIM, structure similarity) 的概念于 2004 年提出^[28]，它从图像中提取 3 个主要特征：结构、亮度和对比度，然后根据这 3 个特征比较两幅图像的相似度，其定义为：

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \quad (4)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (5)$$

其中： μ_i 为给定图像的均值， σ_x^2 为给定图像的方差， σ_{xy} 为两幅图像的协方差， C_1 和 C_2 为常数，分别设为 1×10^{-4} 和 9×10^{-4} ，以确保分母为零时的稳定性。而在本文提出的方法中，为了处理不同尺度图像的特征信息，提高图像融合质量，采用了多尺度加权结构相似度 (MWSSIM, multi-scale weighted structure similarity) 来计算原始图像和融合图像中的数据损失，其定义如下：

$$Loss_{MWSSIM} = 1 - \frac{1}{|\omega|}$$

$$\sum_{\omega} (\beta_{\omega} \cdot SSIM(I_s, I_f; \omega) + (1 - \beta_{\omega}) \cdot SSIM(I_{DOP}, I_f; \omega)) \quad (6)$$

其中： $\omega \in \{3, 5, 7, 9, 11\}$ 表示不同的窗口，同时用不同的窗口提取不同尺度的特征信息， $SSIM(x,y;\omega)$ 表示窗口 ω 下两幅图像的结构相似度， β_{ω} 为权重系数，计算公式如下：

$$\beta_{\omega} = \frac{g(\sigma_s^2)}{g(\sigma_s^2) + g(\sigma_{DOP}^2)} \quad (7)$$

当 $\sigma_s^2 > \sigma_{DOP}^2$ 时，可见光图像在局部区域有更多的细节信息，因此相应的可见光图像的权重系数应该更大。此外，权重系数 $g(x) = \max(x, 0.0001)$ 是一个校正函数，用于提高解的鲁棒性。

$Loss_{MWSSIM}$ 主要从 3 个方面比较源图像和融合图像的结构相似性，但忽略了一些细节信息，如像素信息、边缘信息等。然而，细节信息在图像应用中至关重要。为了更好地训练模型，减少源图像和融合图像之间的细节信息差异，我们在图像边缘损失函数中引入了多尺度加权融合质量指

标 Q_w [29], 定义如下:

$$Loss_{Q_w} = 1 - \frac{1}{|\omega|} \sum_w c(\omega) (\lambda(\omega) \cdot Q_o(x, f; \omega) + [(1 - \lambda(\omega)) \cdot Q_o(y, f; \omega)]) \quad (8)$$

$$c(\omega) = \frac{\max[s(x; \omega), s(y; \omega)]}{\sum_w \max[s(x; \omega'), s(y; \omega')]} \quad (9)$$

$$\lambda(\omega) = \frac{s(x; \omega)}{s(x; \omega) + s(y; \omega)} \quad (10)$$

$$Q_o = \frac{4\sigma_{xy}\mu_x\mu_y}{(\mu_x^2 + \mu_y^2) \cdot (\sigma_x^2 + \sigma_y^2)} \quad (11)$$

其中: $s(x; \omega)$ 反映了窗口 ω 中图像 x 的局部相关性, 它一般取决于对比度、清晰度或熵等因素。由于本算法中的图像增强更侧重于图像边缘, 因此我们选择对比度来表示图像特征。 $\lambda(\omega)$ 表示图像 x 相对于图像 y 的相对重要性, 其取值介于 0 和 1 之间。给相对于输入图像具有较高显著性的窗口分配更多权重, 如公式 (9) 所示, $c(\omega)$ 为窗口的总体显著性权重。此外, $Q_o(x, y; \omega)$ 是窗口 ω 中 x 和 y 的相似度度量, 取值在 -1 和 1 之间, 当图像 x 和 y 相等时, 最大值为 1。

最终的损失函数表示如下:

$$Loss = Loss_{MWSSIM} + \alpha Loss_{Q_w} \quad (12)$$

其中: α 是平衡参数, 用于平衡损失函数的数量级, 在实验中设定为 $\alpha = 0.1$ 。

2 目标检测网络

2.1 YOLOv5 网络框架

YOLO 算法最初是由 Redmon 等人 [15] 提出的一种目标检测算法, 区别于传统目标检测算法, 该框架直接根据检测性能进行端到端的优化, 并且达到了实时处理图像的要求, 优于领域内其他检测方法。随着研究人员的深入研究拓展, YOLO 系列算法已经发展到 v8 模型。YOLOv5 在 YOLO 系列中是当前使用最广泛也是效果最好的目标检测模型之一, 本文通过对其改进以实现暗光条件下偏振成像的目标检测。

如图 4 所示, YOLOv5 主要由 4 个部分构成: Input (输入端)、Backbone (主干网)、Neck (颈部) 和 Head (头部)。输入端采用了 Mosaic 数据增强, 一般使用 4 张图片按照一定比例进行拼接, 从而缩小目标识别范围; 采用自适应锚框计算, 不同于前 YOLO 系列的模型使用单独脚本进行锚框计算, YOLOv5 在训练前会根据不同数据集来自适应地计算锚框; 采用自适应图片缩放, 自动计算图片缩放比例并缩短短边以提高检测速度。Backbone 模块采用 Focus 结构和 CSP 结构提高了计算力, 且不会丢失有用信息。Neck 模块中采用了 CSPNet 设计的 CSP2 结构, 加强网络的图像特征融合能力。最后在 Head 结构中采用了 Boundingbox 损失函数和 NMS 非极大值抑制, NMS 主要用来消除检测时冗余的框, 提高检测效率。YOLOv5 算法通过灵活的参数配置和超参优化策略, 达到优异性能的同时体量相比后 YOLO 系列算法更小, 适合应用于实时偏振目标检测场景。

2.2 引入 CA 注意力机制

一般的注意力机制在求取通道注意力的时候, 通道的处理一般是采用全局最大池化或者平均池化, 而这样会损失掉物体的空间信息, 因此引用 CA (Coordinate Attention) 注意力机制 [30], CA 通过获取精确的位置信息进行编码, 将位置信息嵌入到了通道注意力中, 也就是引入了空间注意力机制。CA 注意力机制的算法流程图 5 所示, 具体操作主要为 Coordinate 信息嵌入和 Coordinate Attention 生成。

1) 首先对全局平均池化进行分解, 获取具有精准位置信息的远程空间交互信息。Input 为输入的特征图像, 尺寸为 $C \times H \times W$, 然后对 Input 分别进行 X 方向与 Y 方向的池化操作, 从而生成尺寸为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图, 通过这种方法产生的特征图可以使 CA 注意力在一个通道内捕获长距离的依赖关系, 并且有助于保留精确的位置信息, 从而使网络能够更加准确地定位对象, 图 6 展示了特征图

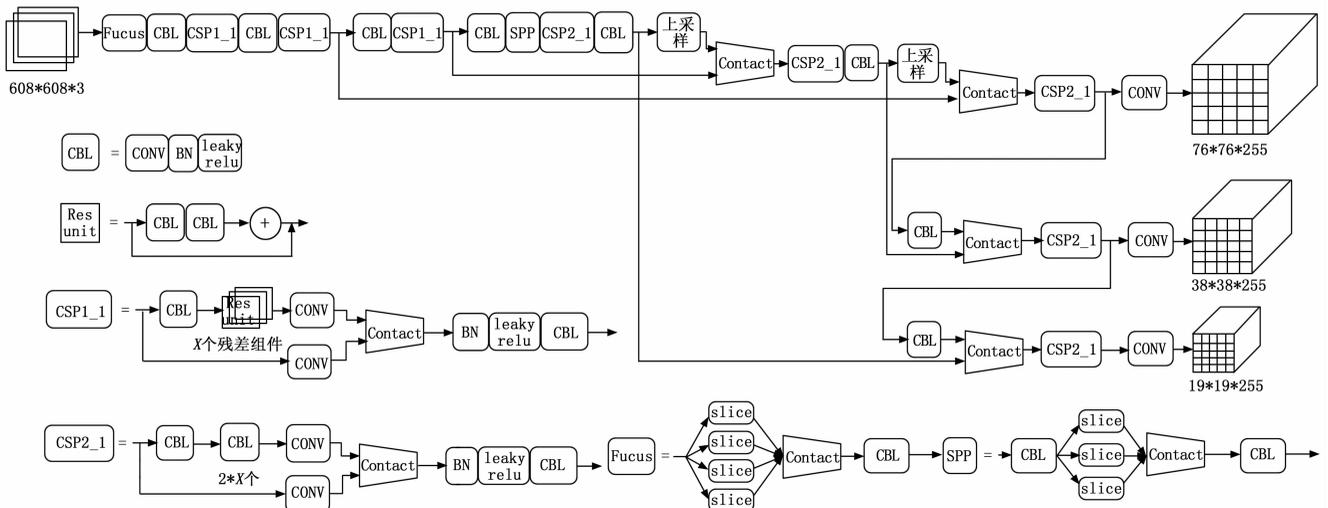


图 4 YOLOv5 网络结构

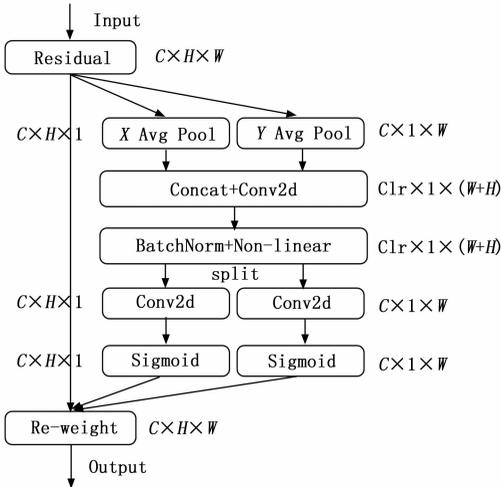


图 5 CA 注意力机制流程图

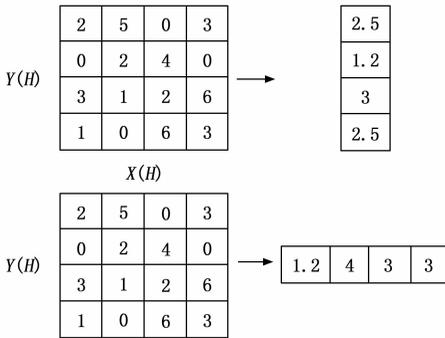


图 6 特征图的平均池化过程

的平均池化过程。

2) 将生成的尺寸为 $C \times 1 \times W$ 的特征图 z^h 和 z^w 进行 Concat 操作，即按空间维度进行拼接：

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (13)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (14)$$

$$f = \delta\{F_1[(z^h, z^w)]\} \quad (15)$$

拼接后得到特征图 $f \in R^{C/r \times (H+W) \times 1}$ ，其中 r 用于控制块的减小率。再将特征图经过 F_1 卷积变换函数 (1×1 卷积) 和非线性激活函数产生中间特征图 f 。

3) 基于空间维度，将中间特征图 f 进行 split 操作拆分成两个张量，分别表示为 $f^h \in R^{C/r \times H \times 1}$ 和 $f^w \in R^{C/r \times 1 \times W}$ ，然后利用 F_h 和 F_w 卷积变换函数 (1×1 卷积) 升维度，再结合 sigmoid 激活函数得到最终的坐标注意力向量 $g^h \in R^{C \times H \times 1}$ 和 $g^w \in R^{C \times 1 \times W}$ ：

$$g^h = \sigma[F_h(f^h)] \quad (16)$$

$$g^w = \sigma[F_w(f^w)] \quad (17)$$

4) 将 $g^h \in R^{C \times H \times 1}$ 和 $g^w \in R^{C \times 1 \times W}$ 与源输入相乘，即 CA 注意力机制的输出公式为：

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (18)$$

CA 不仅同时引入空间注意力和通道注意力，解算二者之间的关系，还考虑到长程依赖的问题，有效实现了检测精度的提升；同时其参数量、计算量相比其他注意力机制也较少，这种足够轻量与灵活的特点使其能够更便捷地插入到轻量级网络的模块中。

3 实验结果与分析

3.1 网络训练与参数设置

由于没有足够成熟且庞大的公开偏振图像数据集，本文自制了偏振数据集用于训练与测试网络。数据集来自 Lucid 的 Phoenix PHX050S-PC 偏振相机，Phoenix 相机传感器芯片集成了 4 个方向的纳米级阵列偏振片，可以同时获取 4 个方向的偏振图像。共获取了 500 幅大小为 $1\,224 \times 1\,024$ 的偏振图像，主要场景包括阴天暗光环境下建筑物、汽车、行人与单车等，图 7 展示了数据集中的一些代表性图像。偏振图像按照相邻的 4 个像素进行分解，得到 0° 、 45° 、 90° 和 135° 共 4 个不同偏振方向的图像，分解后的图像大小为 612×512 ，同时进行归一化处理，以便统一像素范围。用于训练和测试的计算机配置为 Intel (R) Core (TM) i5-7300HQ CPU @ 2.50 GHz 和 (NVIDIA) GeForce GTX 1050。



图 7 部分数据集中的图像

对于图像融合网络：在网络模型的训练阶段，选取了 200 对图像作为融合网络数据集，其中 150 对图像用于训练图像融合网络，40 对图像用于检验网络，剩余的图像则用于测试网络性能，在训练之前先对可将光图像与偏振度图像的像素范围规范在 $[1, 0]$ 内以便模型处理。在训练、验证和测试过程中均不对源图像进行任何变动，直接输入模型。在参数设置方面，学习率设为 0.000 1，训练轮数设为 30，批次大小设为 128，并在 TensorFlow 平台上实现网络模型，使用 Adam 优化器训练模型。

对于目标检测网络：由于数据集的数量相对较少，不足以充分训练改进的 YOLOv5 模型，本文对数据集进行增强处理，通过水平翻转、垂直翻转、裁剪拼接、改变亮度等方式将原有的 500 幅图像增强到 3 273 张，达到了充足训练网络的水平。在训练前先将数据集以 $7:2:1$ 的比例分成 3 部分，分别作为训练集、测试集和验证集用于模型中，选取汽车、行人与单车 3 个标签作为目标检测类别。权重衰减率为 0.000 5，初始学习率设为 0.01，训练轮数设为 100，批次大小为 16，动量因子设置为 0.937，在 Pytorch 平台上实现网络模型。

3.2 结果与分析

首先使用主观方法对图像融合网络的效果进行评估, 图 8 展示了两个场景下的图像融合结果, 其中 S_0 为可见光图像, $DoLP$ 为偏振度图像, AoP 为偏振角图像, Ave 为 S_0 和 $DoLP$ 的平均图像, Our 则为 S_0 和 $DoLP$ 的平均图像。从主观上来看, 本文所提出的图像融合方法可以将可见光图像的目标信息与偏振度图像的偏振细节信息有效融合, 并且融合后的图像呈现出较好的视觉效果, 对比度增强且边缘清晰, 融合后的图像可以应用到目标检测场景中。

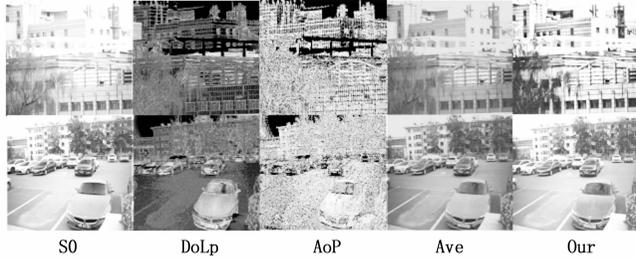


图 8 图像融合网络效果图

使用主观方法对图像融合网络的效果进行评估, 本文建立了不同的验证集: 可见光图像验证集、偏振度图像验证集和融合图像验证集, 使用改进的 YOLOv5 目标检测算法在不同验证集下的检测结果对比如图 9 所示。第一组图像为室内光照不足条件下的人像图, 由于图像融合后获得清晰的边缘信息, 因此对人像的检测有一定提升效果; 第二组图像中, 由于左下角车辆玻璃的材质以及反光影响, 产生一定的偏振噪声, 导致融合后的精度降低, 但其他目标的精度均有提升, 尤其是对于小弱目标的检测精度; 第三组图像中最左侧的车辆在可见光图像下未识别出, 待测目标有一定遮挡且尺度较小, 肉眼也很难观测到, 而融合后的图像具有更多特征信息, 能够被有效检测; 在第四组图像中, 拍摄场景在多云阴天, 有较多环境干扰信息, 经过图像融合后有效抑制了环境造成的冗余特征, 将关注点更多地放在有用区域。

接下来对改进的 YOLOv5 模型进行可行性分析, 改进模型的损失函数曲线如图 10 所示, 其中 box_loss 为目标锚框的定位面向暗光场景的目标偏振可见光融合检测方法损失, obj_loss 是置信度损失, $loss$ 为总损失。观察图中曲线可看出 3 种损失均在训练过程中逐渐收敛, 最终达到一个稳定值, 总损失 $loss$ 在前 15 轮迅速下降到 0.03 以下, 在之后的轮次也逐渐趋于稳定, 最终达到 0.000 5 左右, 说明了模型的稳定性与有效性。

为更好地验证本文改进的 YOLOv5 模型的优势, 选取几种目前主流的目标检测算法: Faster-RCNN^[11]、YOLOv4^[18]、YOLOv5s、YOLOv5m 和 YOLOv5l, 在相同数据集下进行对比实验, 并且使用针对目标检测的评价标准进行客观性评价, 评价标准主要有以下几个: 召回率 μ_R , 表示模型预测的正样本数占实际正样本数的比例, 也称为查全率; 精确率 μ_P , 表示预测正确样本数占有识别为正样本的样本



图 9 不同验证集下的检测结果对比图

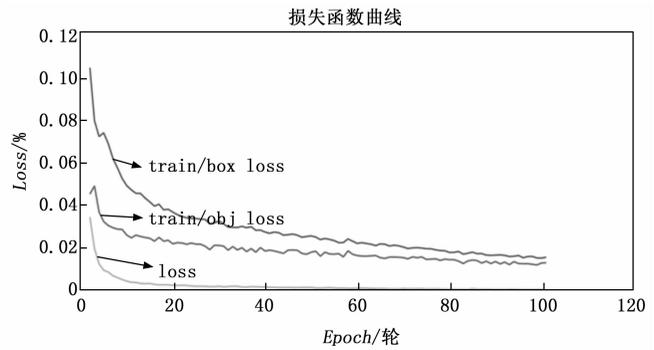


图 10 损失函数曲线

比例; 均值平均精度 $mAP@0.5$, 代表模型在 IOU 阈值为 0.5 时的平均精度, 其中 IOU 是交并比, 表示模型预测的目标框和人工标定目标框的交集之间的比例; 均值平均精度 $mAP@0.5 : 0.95$, 代表 IOU 阈值从 0.5~0.95, 步长为 0.05 时的平均精度; 推理时间 t , 表示模型检测一张图像所需要的时间; 平均精度 AP , 指 $P-R$ 曲线的面积, 代表平均精度, 计算公式如下:

$$AP = \int_0^1 P(R) dR \quad (19)$$

$$\mu_P = \frac{TP}{TP + FP} \quad (20)$$

$$\mu_R = \frac{TP}{TP + FN} \quad (21)$$

其中: $P(R)$ 表示 $P-R$ 曲线; FN 表示样本的真实类别是正样本, 但是模型将其预测为负样本; TP 表示样本的真实类别是正样本, 模型将其预测的结果也是正样本; FP 表示样本的真实类别是负样本, 但是模型将其预测为正样本。

不同算法的性能指标对比结果如表 2 所示, 可以看出本文算法与其他经典算法相比, 精确率和召回率都更高, 分别为 89.3% 和 82.5%; $mAP@0.5$ 只有本文算法和 YOLOv5s 达到了 90% 以上, 但本文算法效果更好, 相比 YOLOv5s 提高了 2.6%; $mAP@0.5 : 0.95$ 达到 70% 的有

4 种算法, 本文算法略高与其他 3 种, 相比于第二优秀的 YOLOv5s 提高了 1.8%; 在检测速率方面 YOLOv5s 算法的速度最快, 由于本文算法引入了 CA 注意力机制, 使检测速率平均慢了 2.9 ms, 但在几种算法中仍然具有一定优势, 能够满足实时监测的要求。整体来看, 本文改进的 YOLOv5 算法兼顾了实时性与检测精度, 通过引入注意力机制, 有效抑制了阴天暗光等环境因素对目标检测的影响, 将更多的注意力集中在待测目标上, 从而提升微小目标的检测精度与准确度。

表 2 不同算法的性能指标对比结果

模型名称	$\mu_P / \%$	$\mu_R / \%$	$mAP@0.5 / \%$	$mAP@0.5 : 0.95 / \%$	t / ms
Faster-R-CNN	66.1	72.3	88.5	68.4	95.3
YOLOv4	85.2	77.9	85.9	69.3	102
YOLOv5s	87.9	80.7	91.2	77.8	84.5
YOLOv5m	85.3	80.2	88.7	75.5	125
YOLOv5l	86.6	79.7	87.1	75.4	176
本文算法	89.3	82.5	93.6	79.2	88.6

4 结束语

为了提高阴天暗光场景下的偏振目标检测效果, 首先从改善可见光和偏振图像的融合效果入手, 提出了一种基于 CNN 的有效融合网络, 该网络以无监督的方式进行训练, 无需手动设计复杂的融合规则和活动水平指标, 而是可以在卷积层中自主学习。还引入了 DenseBlock 来充分提取图像特征, 设计了一个新的损失函数来确保网络的融合性能, 在更大程度上缩小了融合图像与源图像之间的差距, 引入拉普拉斯算子来增强融合图像的对比度和边缘强度, 最终得到同时具有偏振信息与可见光信息的待测目标。然后提出了一种基于改进的 YOLOv5s 模型, 通过添加 CA 注意力机制将空间注意力机制与通道注意力机制相结合, 使网络更容易捕捉到待检测目标物, 提升模型检测精度, 并且该模型足够轻量, 实时性高, 全局感知能力强。

为验证所提算法的有效性, 从主观和客观两个方面评估了融合算法的有效性, 通过与 5 种经典目标检测算法进行对比实验, 使用 5 种评估指标对目标检测网络进行评估。实验结果表明, 本文所提出的偏振目标检测网络相比最优的 YOLOv5s 模型, 精确率和召回率分别达到了 89.3% 和 82.5%, $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提高了 2.6% 和 1.8%, 因此本文提出的图像融合算法与目标检测算法解决了暗光条件下目标信息缺失难以被检测到的问题, 先通过融合网络将目标的偏振信息与可见光信息相融合, 构成信息丰富的融合图像, 然后通过改进的目标检测网络将注意力集中在目标物区域, 从而提升目标检测的准确度与速度。相比于 YOLOv5 和其他主流目标检测模型在检测精度、检测时间等方面有较大提升, 适合一些极端条件下的目标检测场景。本文在检测速率以及模型轻量化方面仍有继续提升的空间, 也缺少对更多极端场景下的目标检测研究, 这

也是后续算法的进一步优化方向。

参考文献:

- [1] 罗海波, 张俊超, 盖兴琴, 等. 偏振成像技术的发展现状与展望 (特邀) [J]. 红外与激光工程, 2022, 51 (1): 20210987.
- [2] SHORT N J. Effects of surface materials on polarimetric-thermal measurements: applications to face recognition [J]. Applied Optics, 2016, 55 (19): 5226-5233.
- [3] ZHONG B, WANG X, WANG D, et al. Target-background contrast enhancement based on a multi-channel polarization distance model [J]. Bioinspir Biomim, 2021, 16 (4): 046009.
- [4] GAIARIN S, PEREGO A M, DA SILVA E P, et al. Dual-polarization nonlinear Fourier transform-based optical communication system [J]. Optica, 2018, 5 (3): 263-270.
- [5] LIU F, WEI Y, HAN P, et al. Polarization-based exploration for clear underwater vision in natural illumination [J]. Optics Express, 2019, 27 (3): 3629-4361.
- [6] LIU Y R, HE H H, WU J. Differentiation of human GBM from non-GBM brain tissue with polarization imaging technique [J]. Front Oncol, 2022, 12: 863682.
- [7] ZHANG J, SHAO J, CHEN J, et al. Polarization image fusion with self-learned fusion strategy [J]. Pattern Recognition, 2021, 118: 108045.
- [8] GU Y, LÜ J, BO J, et al. An improved wavelet modulus algorithm based on fusion of light intensity and degree of polarization [J]. Applied Sciences, 2022, 12 (7): 3558.
- [9] LI K, QI M, ZHUANG S, et al. TIPFNet: a transformer-based infrared polarization image fusion network [J]. Optics Letters, 2022, 47 (16): 4255-4258.
- [10] XU H, SUN Y, MEI X, et al. Attention-guided polarization image fusion using salient information distribution [J]. IEEE Transactions on Computational Imaging, 2022, 8: 1117-1130.
- [11] GIRSHICK R. Fast r-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [12] REN S, HE K, GIRSHICK R, et al. Faster r-CNN: towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 111-119.
- [13] DOLLÁR K, GIRSHICK R. Mask r-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Computer Vision-ECCV 2016; 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part I 14. 2016: 21-37.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

- [16] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 - 7271.
- [17] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. Arxiv Preprint Arxiv: 180402767, 2018: 231 - 236.
- [18] BOCHKOVSKIY A, WANG C, LIAO H. Yolov4: optimal speed and accuracy of object detection [J]. Arxiv Preprint Arxiv: 2004.10934, 2020: 29 - 45.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980 - 2988.
- [20] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [J]. Arxiv Preprint Arxiv: 1904.07850, 2019: 582 - 593.
- [21] LAW H, DENG J. CORNERNET: Detecting objects as paired keypoints [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734 - 750.
- [22] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 - 10022.
- [23] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C] // Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010: 807 - 814.
- [24] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images [J]. IEEE Trans Image Process, 2018, 28 (5): 2614 - 2623.
- [25] CAO X. A practical theory for designing very deep convolutional neural networks [J]. Unpublished Technical Report, 2015: 847 - 853.
- [26] SONKA M, HLAVAC V, BOYLE R. Image processing, analysis, and machine vision [M]. Cengage Learning, 2014.
- [27] ZHOU Q, GUO K, SUN J, et al. Research on tool wear detection method using deep residual network [C] // Intelligent Robotics and Applications; 14th International Conference, ICIRA 2021, Yantai, China, Proceedings, Part IV 14. Springer International Publishing, 2021: 565 - 575.
- [28] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans Image Process, 2004, 13 (4): 600 - 612.
- [29] PIELLA G, HEIJMANS H. A new quality metric for image fusion [C] // Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). IEEE, 2003, 3: 164 - 173.
- [30] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713 - 13722.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.
- [15] 端木春江, 石亮. 基于特征融合的图像超分辨率 [J]. 计算机时代, 2023 (4): 120 - 122.
- [16] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks [C] // Advances in Neural Information Processing Systems, 2016: 4905 - 4913.
- [17] LIU S, HUANG D. Receptive field block net for accurate and fast object detection [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385 - 400.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 - 9.
- [19] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 - 11542.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 - 7141.
- [21] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146 - 157.
- [22] 曾文鑫. 基于机器学习的车标图像识别研究 [D]. 西安: 西安石油大学, 2021.

(上接第 28 页)

- [4] 徐国新, 李雷孝, 何嘉彬, 等. 驾驶员安全带检测方法研究综述 [J]. 计算机工程与应用, 2023, 59 (17): 48 - 66.
- [5] 宁健, 马森, 柴立臣, 等. 深度学习的目标检测算法综述 [J]. 信息记录材料, 2022, 23 (10): 1 - 4.
- [6] 殷远齐, 徐源, 邢远新. 基于 YOLO v4 的车辆目标检测算法 [J]. 计算机与现代化, 2022 (7): 8 - 14.
- [7] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 - 19.
- [8] 牛洪超, 胡晓兵, 罗耀俊. 基于 M-YOLO 的自动驾驶下目标识别算法 [J]. 计算机工程与设计, 2022, 43 (8): 2213 - 2220.
- [9] 田坤, 李冠, 赵卫东. 基于 YOLO 和极限学习机的驾驶员安全带检测模型研究 [J]. 计算机应用与软件, 2019, 36 (11): 196 - 201.
- [10] HOSSEINI S, FATHI A. Automatic detection of vehicle occupancy and driver's seat belt status using deep learning [J]. Signal, Image and Video Processing, 2022: 1 - 9.
- [11] 刘展威, 陈慈发, 董方敏. 基于 YOLOv5s 的航拍小目标检测改进算法研究 [J]. 无线电工程, 2023, 53 (10): 2286 - 2294.
- [12] 李世伟. 基于 YOLOv5 算法的目标检测与车牌识别系统 [J]. 电子技术与软件工程, 2022 (1): 138 - 141.
- [13] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759 - 8768.