第 42 卷 第 9 期/2022 年 5 月/光学学报 研究论文

光学学报

基于双流融合网络的单兵伪装偏振成像检测

王荣昌[1,2]，王峰[1,2*]，任帅军[1,2]，王勇[1,2]

[1]中国人民解放军陆军炮兵防空兵学院信息工程系，安徽 合肥 230031；

[2]偏振光成像探测技术安徽省重点实验室，安徽 合肥 230031

摘要 单兵伪装目标与背景之间在颜色上有高度的相似性，目标具有高度复杂的姿态，而且存在遮挡问题，这些问题使得单兵伪装目标检测较传统目标检测有很大的挑战性。针对上述问题，提出基于偏振信息和 RGB(Red，Green，Blue)信息的深度学习算法，同时构建单兵伪装目标偏振图像数据集 CIP3K(Multicam 型迷彩伪装数据集和 Woodland 型迷彩伪装数据集)。基于 Faster R-CNN(Faster Region-Convolutional Neural Network)提出一种双流特征融合网络 TSF-Net，其能够融合目标偏振特征信息和 RGB 特征信息。在 CIP3K 数据集上进行大量实验，用来测试 TSF-Net 模型与其他检测模型的性能。实验结果表明，相较于 Faster R-CNN，TSF-Net 模型在两个数据集上的平均检测精度分别提高了 8.2 个百分点和 8.8 个百分点，且优于一些主流目标检测模型。

关键词 机器视觉；单兵伪装检测；偏振成像；卷积神经网络；数据集

中图分类号 TP751.1 **文献标志码** A **doi**：10.3788/AOS202242.0915001

Polarization Imaging Detection of Individual Camouflage Based on Two-Stream Fusion Network

Wang Rongchang[1,2], Wang Feng[1,2*], Ren Shuaijun[1,2], Wang Yong[1,2]

[1]*Department of Information Engineering, PLA Army Artillery Air Defense Force College, Hefei* 230031, *Anhui, China*;

[2]*Key Laboratory of Polarized Light Imaging Detection Technology of Anhui Province, Hefei* 230031, *Anhui, China*

Abstract There is a high degree of color similarity between the individual camouflage target and the background, the target has a highly complex posture, and there are occlusion problems, which make individual camouflage target detection more challenging than traditional target detection. In order to solve the above problems, a depth learning algorithm based on polarization information and RGB (Red, Green, Blue) information is proposed, and the polarization image dataset CIP3K (Multicam type camouflage dataset and Woodland type camouflage dataset) is constructed. Based on Faster R-CNN (Faster Region-Convolutional Neural Network), a dual-stream feature fusion network TSF-Net is proposed, which can integrate target polarization feature information and RGB feature information. A large number of experiments are carried out on the CIP3K dataset to test the performance of the TSF-Net model and other detection models. The experimental results show that, compared with Faster R-CNN, the average detection accuracy of the TSF-Net model on the two datasets is increased by 8.2 percentages and 8.8 percentages, respectively, and is better than some mainstream object detection models.

Key words machine vision; individual camouflage detection; polarization imaging; convolutional neural network; dataset

收稿日期：2021-09-03；**修回日期**：2021-10-08；**录用日期**：2021-11-17

通信作者：*wfissky7202@sina.com

1 引 言

伪装目标检测是用来识别、检测图像中融于背景环境的伪装物体。从目标伪装方式来看，伪装目标检测通常可分为自然伪装目标检测和人造伪装目标检测，而本文所研究的单兵伪装目标检测属于人造伪装目标检测的范畴。同时，单兵伪装目标检测是特种作战中最常见的反侦察技术之一，通过单兵伪装目标的颜色和纹理信息与背景的高度相似性来达到反侦察的目的。相较于计算机视觉中的通用目标检测，单兵伪装目标检测中的难点与挑战如下：1)单兵伪装目标的颜色与背景相似，甚至相同；2)大部分单兵伪装目标存在遮挡问题；3)单兵伪装目标的姿态具有高度复杂性和不确定性。上述问题与挑战增加了获得具有可区分性的特征的难度，从而降低单兵伪装目标检测的正确率。

目前，对于基于深度学习的单兵伪装目标检测的研究较少，现有研究中常用的方法包括引入注意力机制[1-2]、增强语义信息[3-4]和扩大感受野[5-6]等。Fang 等[7]构建了一个新的完整的伪装人员检测数据集，并提出了一种基于端到端结构的强语义扩展网络，通过特征图的求和来增强语义信息，并添加扩张卷积层以扩大模型的感受野。Zheng 等[8]通过提取深度卷积神经网络中的高级语义特征，并在反卷积阶段引入短连接，从而构建密集的反卷积网络，进而实现语义信息的使用和融合。邓小桐等[9]在 RetinaNet 检测框架[10]的基础上，针对迷彩目标所具有的特性嵌入了空间注意力模块和通道注意力模块，并基于定位置信得分提出了新的预测框过滤算法。王杨等[11]在 YOLOv5(You Only Look Once v5)算法的基础上，为特征提取网络添加注意力模块并将网络中不同层的特征图进行融合。

现有基于深度学习的单兵伪装目标检测的研究都是基于 RGB(Red，Green，Blue)图像开展的，单兵伪装目标的颜色与背景的相似程度高，甚至相同，导致以 RGB 图像作为数据源进行检测的效果不佳。而单兵伪装目标表面的材质与背景有着较大的差异，使得它们的偏振特性也有较大的差异，因此本文提出了基于偏振图像与 RGB 图像的双流特征融合检测算法，该算法可以解决单兵伪装目标中颜色相似、遮挡和姿态复杂的问题，进而提升单兵伪装目标检测的正确率。本文的主要贡献有如下两个方面。1)本文构建了单兵伪装偏振图像数据集(CIP3K)，其包含 Multicam 型迷彩伪装数据集(1500 张)和 Woodland 型迷彩伪装数据集(1500 张)两个子数据集，子数据集中均包含 20～25 个自然背景，其中迷彩伪装人员被设计了多种姿态及遮挡情况。通过解析该数据集可以得到 RGB 图像数据集、偏振图像数据集(偏振方向图数据集和各偏振参量图数据集)。2)本文提出了一种双流特征融合网络(TSF-Net)，其可以提取图像偏振特征信息和 RGB 特征信息，并能融合两种特征。在 CIP3K 数据集上进行性能测试，相较于当今主流检测网络，所设计的网络的检测精度有所提升。

2 相关工作

2.1 偏振成像

2.1.1 光的偏振

光波本质上是一种电磁波，其在传输过程中，电矢量 $\boldsymbol{E}$ 的振动方向垂直于光的传播方向，如图 1 所示。自然光通过媒介发生反射、折射、吸收和散射后会出现某一个方向的振动比其他方向有优势，导致光的振动分布不再具有对称性，这就是偏振光。

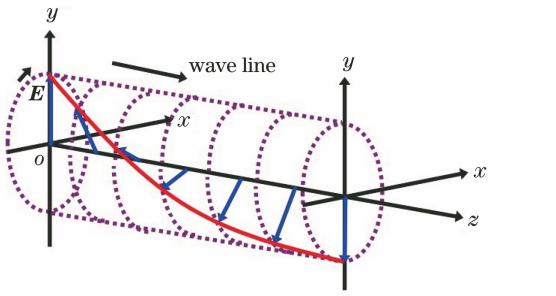

图 1 光传播示意图

Fig. 1 Schematic of light propagation

由图 1 可以看到，光沿着 z 轴传播，电矢量 $\boldsymbol{E}$ 在不同方向上可表示为

$$\begin{cases} E_x = E_{0x}\cos(\tau+\delta_x) \\ E_y = E_{0y}\cos(\tau+\delta_y) \end{cases}, \tag{1}$$

式中：$\boldsymbol{E}_0$ 表示初始时刻的电矢量；$\tau+\delta$ 表示相位，其中 $\tau=\omega t-kz$，ω 表示角频率，k 表示传播矢量大小，t 表示传播时间，δ 表示相位偏置量，$\delta=\delta_x-\delta_y$。为了方便描述光的电矢量运动轨迹，对式(1)进行整理，可以得到

$$\left(\frac{E_x}{E_{0x}}\right)^2+\left(\frac{E_y}{E_{0y}}\right)^2-2\left(\frac{E_x}{E_{0x}}\right)\left(\frac{E_y}{E_{0y}}\right)\cos\delta=\sin^2\delta。 \tag{2}$$

当 $\delta=n\pi(n=0,\pm1,\pm2,\cdots)$ 时，电矢量的振动方向保持固定的一个方向，且在同一个平面内为线偏振光，其包括 0°、45°、90°和 135°线偏振光。

2.1.2 彩色分焦平面偏振成像原理

图 2 为彩色分焦平面偏振像元阵列的排布示意图，排布规则：像元阵列由彩色偏振像元组成，每个彩色偏振像元由 4×4 个基础像素点组成，其中由 0°、45°、90°和 135°4 个偏振方向的像元组成一个偏振像元，在此基础上将 4 个偏振像元按照拜尔排布规则组成一个彩色偏振像元。由该偏振成像系统获得的原始图通过解析处理可以得到 4 个偏振方向的彩色三通道强度图，进而计算得到彩色三通道偏振参量图。

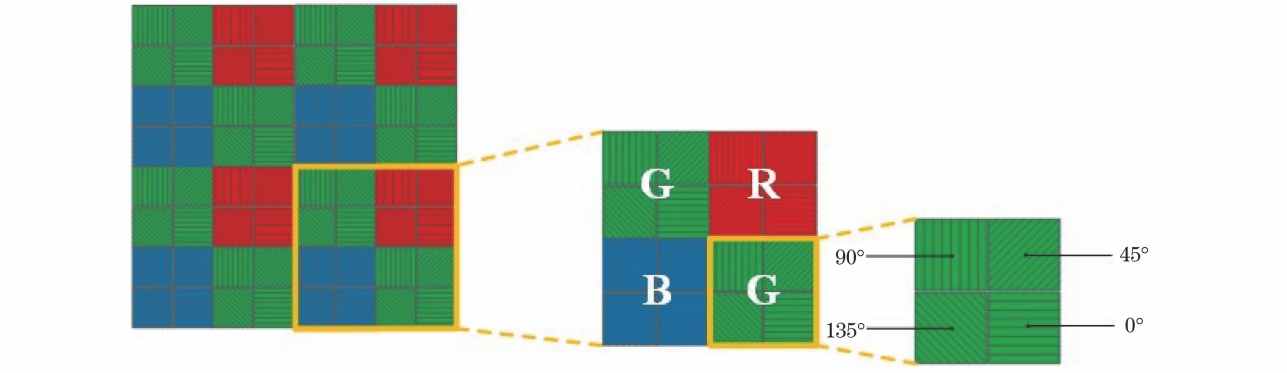

图 2 彩色分焦平面偏振像元阵列的排布示意图

Fig. 2 Layout diagram of color focal plane polarized pixel array

彩色分焦平面偏振相机所获取的原始图通过解析可以得到彩色 RGB 偏振方向强度图，计算偏振参量的过程中 R、G 和 B 3 个通道互不影响，计算偏振参量的公式为

$$\begin{cases} I=\dfrac{1}{2}(I_0+I_{45}+I_{90}+I_{135}) \\ Q=I_0-I_{90} \\ U=I_{45}-I_{135} \\ P=\dfrac{\sqrt{Q^2+U^2}}{I} \\ A=\dfrac{1}{2}\arctan\left(\dfrac{U}{Q}\right) \end{cases}, \tag{3}$$

式中：I、Q、U、P 和 A 分别表示辐射强度、水平方向偏振分量强度、45°方向偏振分量强度、偏振度和偏振角。

2.2 单兵伪装检测

2.2.1 数据集

在单兵伪装目标检测领域中，现有的公开数据集有如下几个：1）CAMO[12]是一个迷彩伪装数据集，它有 2500 张图像，图像中涵盖 8 个类别，而数据集中含有 CAMO 和 MS-COCO 两个子数据集，每个子数据集均包含 1250 张图像；2）Fang 等[7]构建了另一种迷彩伪装人员数据集，包含 26 种迷彩样式，有 2600 张图像（每种样式有 100 张）；3）邓小桐等[9]在此基础上增加了 7 种迷彩样式，有 700 张图像，总计构建 33 种迷彩样式，有 3300 张图像。

本文构建了 CIP3K 数据集，该数据集通过解析可以得到偏振方向图像数据集、偏振参量图像数据集和 RGB 图像数据集，标注信息可以互相通用。CIP3K 数据集有 3000 张原始偏振图，包含多样的自然背景和复杂的伪装人员姿态。

2.2.2 单兵伪装类型

单兵伪装类型主要包括迷彩伪装、伪装网和伪装涂料等，最常用的是迷彩伪装。迷彩伪装的类型较多：根据季节分为春夏季型、秋季型和冬季型三种类型；根据应用场景分为丛林型、荒漠型和雪地型。本文的研究对象为春夏季丛林型迷彩伪装。

2.2.3 单兵伪装目标检测规定

单兵伪装目标检测是与类无关的任务，因此单兵伪装目标检测的公式简单且易于定义。给定一张图像，该任务需要一种单兵伪装对象检测方法为每个像素 i 分配一个置信度 $p_i\in[0,1]$，其中 p_i 表示像素 i 的概率分数。不属于单兵伪装对象的像素得分为 0，而得分为 1 表示像素完全分配给单兵伪装对象。本文重点介绍对象级单兵伪装检测任务。

3 基于双流特征融合网络的单兵伪装偏振成像检测算法

3.1 算法总述

所提算法是在 Faster R-CNN（Faster Region-Convolutional Neural Network）检测网络[13]的基础上改进的，本文充分利用偏振信息和 RGB 信息并进行联合互补，从而提高单兵伪装目标检测的正确率。本文算法的总体框架如图 3 所示，其中 APP-Net 表示自适应偏振参量网络，FPN 表示特征金字塔网

络，RPN 表示区域生成网络，ROI 为感兴趣区域，FC 为全连接层，Ⓒ表示特征融合操作。TSF-Net 负责融合偏振信息和 RGB 信息，后续网络负责进一步的检测任务。

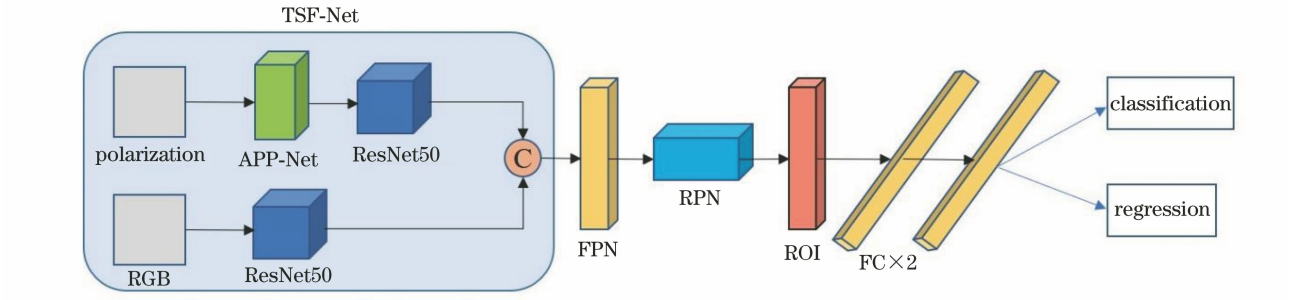

图 3 TSF-Net 的结构图

Fig. 3 Structure diagram of TSF-Net

3.2 TSF-Net

如图 3 所示，TSF-Net 包括偏振流网络(Pol-Net)和 RGB 流网络(RGB-Net)两个分支，其中 Pol-Net 负责从偏振图像中提取特征信息，RGB-Net 负责从 RGB 图像中提取特征信息，最后将两个分支所提取的特征进行融合操作。

3.2.1 Pol-Net

Pol-Net 包括 APP-Net 和 ResNet-50，其中 APP-Net 负责生成自适应偏振参量，这是 Pol-Net 的主体部分，ResNet-50 负责对 APP-Net 生成的自适应偏振参量进行特征提取。

1) APP-Net 的设计

由彩色分焦平面偏振成像原理可知，分焦平面获取的是原始偏振图在 0°、45°、90°和 135°4 个偏振方向的强度信息，这 4 个偏振方向的强度信息可以通过式(3)计算得到目标的偏振参量信息。此外，理论上还可以将 4 张偏振方向图通过某种运算操作来得到代表目标偏振特性的参量，该过程可以描述为

$$S_f^{(x,y)}=f(d_0^{(x,y)},d_{45}^{(x,y)},d_{90}^{(x,y)},d_{135}^{(x,y)}),\quad(4)$$

式中：$S_f^{(x,y)}$ 表示计算得到的偏振参量；$f(\cdot)$表示某种运算的过程；$(d_0^{(x,y)},d_{45}^{(x,y)},d_{90}^{(x,y)},d_{135}^{(x,y)})$表示 4 个方向偏振图像中坐标为$(x,y)$的像素点。那么，通过不同的 $f(\cdot)$，即可得到具有不同目标偏振特性的偏振参量。

上述思路可以用人工神经网络(ANN)来表达。如图 4 所示，包含偏振参量的 ANN 包括输入层、若干数量的隐藏层和输出层。输入层有 4 个节点，代表 4 张偏振方向图上的像素点；输出层有 n 个节点，代表不同偏振参量图上的像素点；隐藏层代表 $f(\cdot)$的运算过程，其中每个节点都有权重和偏振项。假设隐藏层中的节点数量足够多，层数够深，ANN 就可以拟合出任意复杂的运算过程，通过输入 4 个不同偏振方向图中的像素值可以计算得到大量的偏振参量。

根据 ANN 原理，本文设计了适用于处理偏振图像的 APP-Net。APP-Net 的结构如图 5 所示，其中 $\boldsymbol{F}_v^{(k)}$ 表示 v 通道特征图输入网络后输出 k 通道特征图，BN 表示批量归一化。

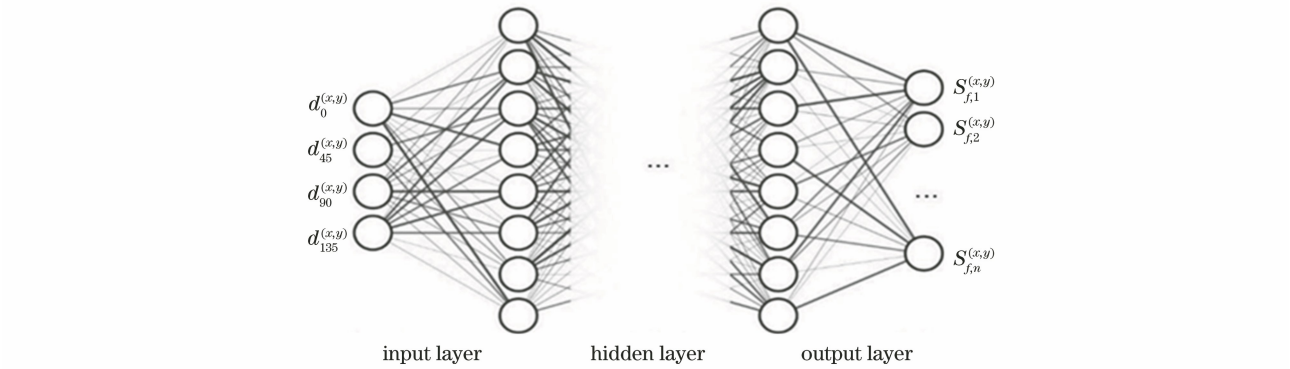

图 4 ANN 的结构图

Fig. 4 Structure diagram of ANN

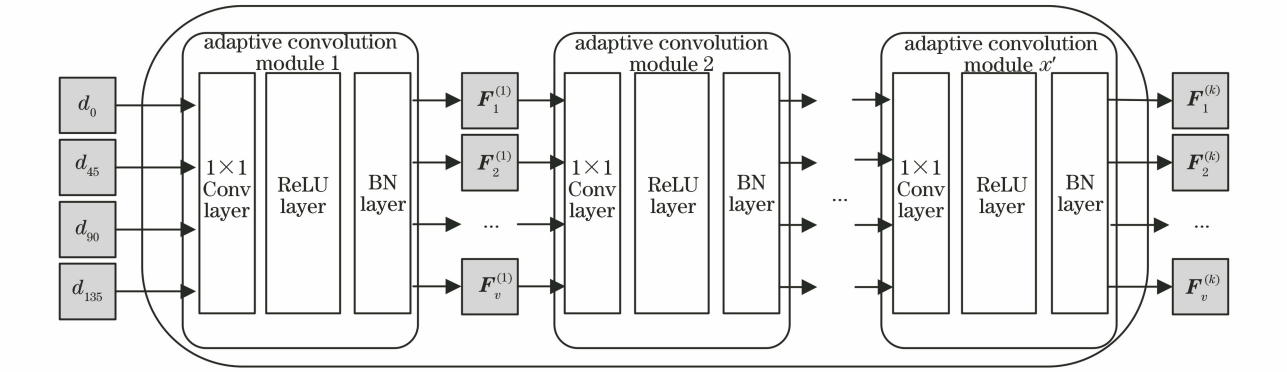

图 5 APP-Net 的结构图

Fig. 5 Structure diagram of APP-Net

APP-Net 的输入是 4 通道偏振方向图，图像经过 x' 个自适应卷积模块处理后，得到 v 幅自适应偏振参量图。其中每个自适应卷积模块均包括 k' 个 1×1 卷积层、激活层(ReLU)[14]和 BN 层[15]，其中每一个卷积模块均通过 k' 个 1×1 大小的卷积核对输入图像进行线性运算，可以输出一组自适应偏振参量图，计算过程可表示为

$$\boldsymbol{F}_v^{(k)}=\mathrm{ReLU}\left(\sum_{i=1}^{k'}\boldsymbol{w}_i\boldsymbol{f}_n\right), \tag{5}$$

式中：$\boldsymbol{f}_n$ 表示输入的 n 维特征图，初始的 $\boldsymbol{f}_0=[d_0\ d_{45}\ d_{90}\ d_{135}]$，代表 4 个偏振方向的图像；$\boldsymbol{w}_i(i=1,2,\cdots,k')$ 表示每一个卷积核的权重矩阵，每一个权重矩阵的维度均为 $n\times n$。最后一个自适应卷积模块中的 $k'=v$，v 的取值决定了输出图像的维度，也就是输出自适应偏振参量图的数量。输入图像为 4 通道偏振方向图，输出为 v 幅自适应偏振参量图。

2) APP-Net 结构的设置

APP-Net 中的自适应卷积模块决定了网络的结构和容量，而自适应卷积模块是可以人为调节的。为了能够较好地拟合出自适应偏振参量的网络结构，本文设计实验，通过调节相关参数来验证网络的性能。首先构建实验所需的数据集，其包括 200 组不同的偏振图像，每组图像包含 0°、45°、90°和 135°4 幅偏振方向强度图。然后分别解析出 200 组偏振图像的 I 图、P 图和 A 图(分别进行线性运算、非线性运算和较复杂的反三角函数运算，有助于测试网络的拟合性能)，将计算后的 200 组图像作为标签。最后设置网络层级和网络容量，即验证自适应卷积模块的数量以及卷积核的数量对网络性能的影响。分别设置三组网络结构：第一组为(8,16,8,3)和(16,8,8,3)；第二组为(96,48,32,3)和(48,96,32,3)；第三组为(128,96,64,32,3)和(96,128,64,32,3)。其中，括号中的每个数字代表一个自适应卷积模块；数字的值代表自适应卷积模块中卷积核的数量；包含 4 个自适应卷积模块的网络称为四级网络，包含 5 个自适应卷积模块的网络称为五级网络。共计 6 个网络结构。

训练过程中，定义的损失函数为

$$x_{\mathrm{Loss}}=\frac{1}{N}\sum_{n'=1}^{N}\left[\frac{1}{W\cdot H}\left(\|\hat{\boldsymbol{Y}}_I^{(n')}-\boldsymbol{Y}_I^{(n')}\|_2+\|\hat{\boldsymbol{Y}}_P^{(n')}-\boldsymbol{Y}_P^{(n')}\|_2+\|\hat{\boldsymbol{Y}}_A^{(n')}-\boldsymbol{Y}_A^{(n')}\|_2\right)\right], \tag{6}$$

式中：$\|\cdot\|_2$ 表示 l_2 范数；N 表示样本总数；n' 表示样本索引；W 和 H 分别表示图像的宽和高；$\boldsymbol{Y}_I$、$\boldsymbol{Y}_P$ 和 $\boldsymbol{Y}_P$ 分别表示 I 图、P 图和 A 图的真实值；$\hat{\boldsymbol{Y}}_I$、$\hat{\boldsymbol{Y}}_P$ 和 $\hat{\boldsymbol{Y}}_A$ 表示对应真实值的网络拟合结果。相应地，网络的拟合能力越强，损失越小。在批大小为 2、学习率为 0.003 和输入图像的分辨率为 1024 pixel×1024 pixel 的情况下，训练 200 轮后的结果如表 1 所示。

表 1 不同结构的训练结果

Table 1 Training results with different structures

Structure	GPU memory usage /MB	Time/min	Loss
(8,16,8,3)	1453	225	1.23×10^{-2}
(16,8,8,3)	1453	216	9.51×10^{-3}
(96,48,32,3)	3817	207	7.48×10^{-3}
(48,96,32,3)	3817	229	5.79×10^{-3}
(128,96,64,32,3)	6109	255	5.55×10^{-3}
(96,128,64,32,3)	6109	282	4.27×10^{-3}

表 1 的结果表明，当网络层级较高且网络容量较大时，网络性能更优，拟合能力更强；但随着网络层级的增加，网络的复杂程度也会增加，相应的计算量也会增长。表 1 中五级网络的内存占用达到了 6109 MB，相较于四级网络的内存占用增加了 1/2

左右。当自适应卷积模块中的卷积核数量较多时，也会增加内存占用，消耗较多的训练时间。因此，不能一味地增加网络层级和容量，需要考虑到计算机硬件条件限制和效率问题。

综上所述，综合考虑 GPU 内存大小和网络性能等问题，本文将 APP-Net 的结构设置为(48，96，32，16，v)，如图 6 所示。APP-Net 包含 5 个自适应卷积模块，其中前 4 个模块的卷积核数量分别为 48、96、32 和 16。考虑到实际应用场景的复杂多变性，将最后一层设置为可人为调节的参数 v，通过设置参数 v 的大小来选择输出 v 个自适应偏振参量。整体网络输入 $W\times H\times 4$ 大小的偏振方向图，经卷积后输出 $W\times H\times v$ 大小的自适应偏振参量图。

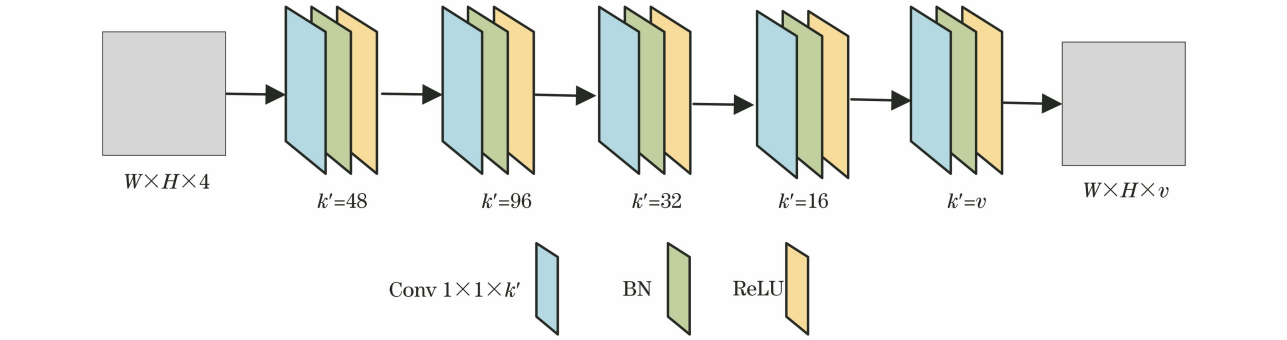

图 6 APP-Net 的结构图

Fig. 6 Structure diagram of APP-Net

APP-Net 是输入 Pol-Net 中的 4 通道偏振方向图，输出 v 幅偏振参量图，然后进一步输入 ResNet-50 中进行特征提取，最后输出 256、512、1024 和 2048 4 个尺度的特征图。

3.2.2 RGB-Net

RGB-Net 是将 RGB 图像输入到 ResNet-50 中[16]进行特征提取。如图 7 所示，输入图像首先经过 7×7 大小的卷积层和 3×3 大小的最大池化层，然后进入 4 个不同的卷积组中，卷积组包含若干个瓶颈式残差模块，最后分别输出 256、512、1024 和 2048 4 个尺度的特征图，其中 s 为步长。

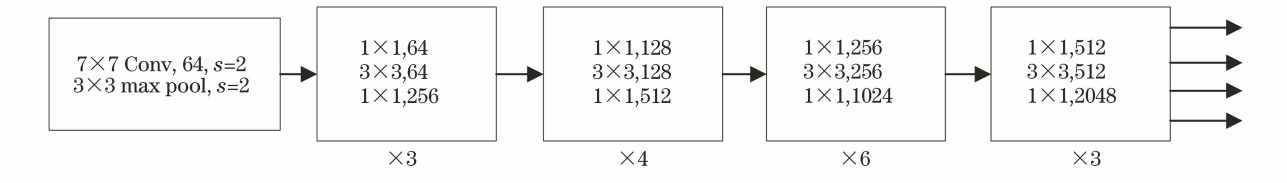

图 7 RGB 流网络的结构图

Fig. 7 Structure diagram of RGB-Net

3.2.3 特征融合

Pol-Net 和 RGB-Net 分别从偏振图像和 RGB 图像中提取特征信息，然后采用通道融合的方式将两个分支输出的特征图进行融合，得到新的特征图。特征提取和特征融合的过程涉及网络中大量单元节点的复杂运算，此处仅使用抽象的方式来描述这种过程。特征提取和特征融合的过程如图 8 所示，其中 $\delta_{\text{Roughness}}$、$\delta_{\text{Texture}}$ 和 δ_{Margion} 分别表示粗糙度特征、纹理特征和边缘特征，δ_{Color} 表示目标色敏特征，$\beta_{\text{Fusion}}=f(\delta_{\text{Roughness}},\delta_{\text{Texture}},\delta_{\text{Margion}},\delta_{\text{Color}})$ 表示最终得到的特征。

如图 8 所示，Pol-Net 分支是从偏振图像中提取目标特征信息，单兵伪装目标与背景有极高的相似度，但其表面材质是由人工制成的，故与自然背景材质有着较大的差异。材质的差异具体表现为粗糙度、纹理和边缘，这些差异是导致目标与背景产生偏振的主要因素。因此从偏振图像中提取到的特征信息包括粗糙度特征、纹理特征和边缘特征，即 $\delta_{\text{Roughness}}$、$\delta_{\text{Texture}}$ 和 δ_{Margion}。

RGB-Net 分支是从 RGB 图像中提取目标特征信息，即单兵伪装目标的色敏特征。迷彩伪装单兵目标的颜色与背景十分相近，现有的迷彩伪装技术难以实现伪装目标的颜色信息与所有背景完全相同，因此可以通过特征提取网络来提取目标色敏特征，即 δ_{Color}。

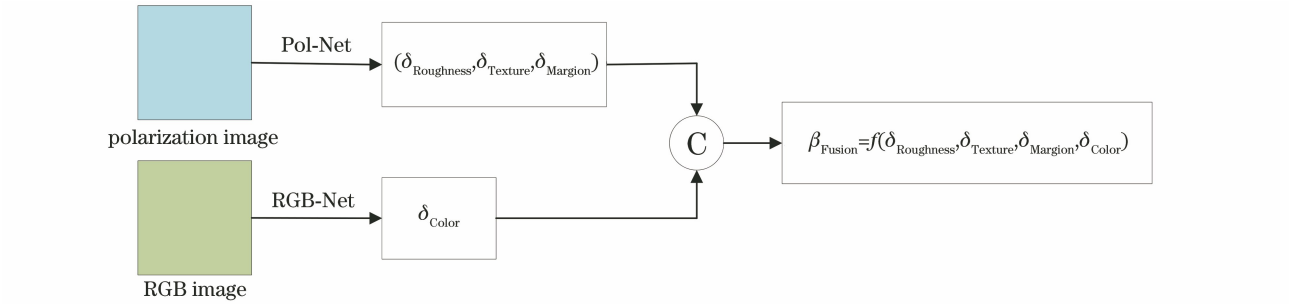

图 8 特征提取和特征融合的过程

Fig. 8 Process of feature extraction and feature fusion

单兵伪装目标检测的难点是在图像中能够区分目标与背景的有效信息十分有限，因此检测算法需要从有限的图像信息中尽可能提取更多的目标特征信息，这也是 TSF-Net 采用双分支结构的根本原因。利用通道融合的方式将两个分支的特征进行融合，最后得到的特征可表示为 $\beta_{Fusion}=f(\delta_{Roughness},\delta_{Texture},\delta_{Margion},\delta_{Color})$。

3.3 损失函数

当网络训练时，区域生成网络的损失函数由分类和回归两部分组成，表示为

$$L(\{p_{i'}\},\{t_{i'}\})=\frac{1}{N_{cls}}\sum_{i'}L_{cls}(p_{i'},p_{i'}^{*})+\lambda\frac{1}{N_{reg}}\sum_{i'}p_{i'}^{*}L_{reg}(t_{i'},t_{i'}^{*}),\tag{7}$$

式中：$\frac{1}{N_{cls}}\sum_{i'}L_{cls}(p_{i'},p_{i'}^{*})$ 表示分类损失；$p_{i'}$ 表示第 i' 个锚框的类别真值；$p_{i'}^{*}$ 表示预测概率；$\frac{1}{N_{reg}}\sum_{i'}p_{i'}^{*}L_{reg}(t_{i'},t_{i'}^{*})$ 表示锚框几何特征的回归损失；$\boldsymbol{t}$ 表示一个向量，包括锚框的 4 个坐标参数；λ 表示加权系数，用于平衡两类损失的权重。

在网络的训练与测试阶段，本文算法主要选择随机梯度下降优化算法来训练网络，网络训练与测试流程如图 9 所示，将训练数据输入网络中进行训练可以得到参数权重和相应损失，将测试数据输入网络中可以得到代表类别和预测框位置的五维数组。

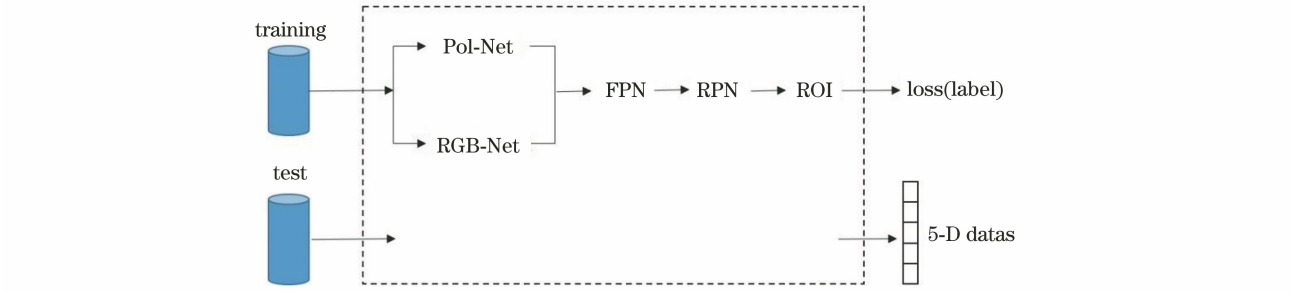

图 9 训练与测试流程示意图

Fig. 9 Schematic of training and test process

4 实验分析

4.1 数据集

4.1.1 数据集的建立

为了采集数据集，搭建了便携式采集设备，如图 10 所示。采集设备主要由搭载采集程序的树莓派板和彩色分焦平面偏振相机组成。相机的主要参数如表 2 所示。

为了尽可能地拍摄包含更多、更接近应用环境背景的图片，采集地点选为某野外训练场。

表 2 彩色分焦平面相机的参数

Table 2 Parameters of color focal plane camera

Category	Parameter
Camera model	FLIR BFS-U3-51S5PC-C
Resolution /(pixel×pixel)	2448×2048
Frame rate /(frame·s^{-1})	75
Chip model	Sony IMX250MYR, Polar-RGB
Data interface	USB3.1 Gen1
Size and weight /(mm×mm×mm)	29×29×30
Mass /g	36
Lens interface	C-Mount

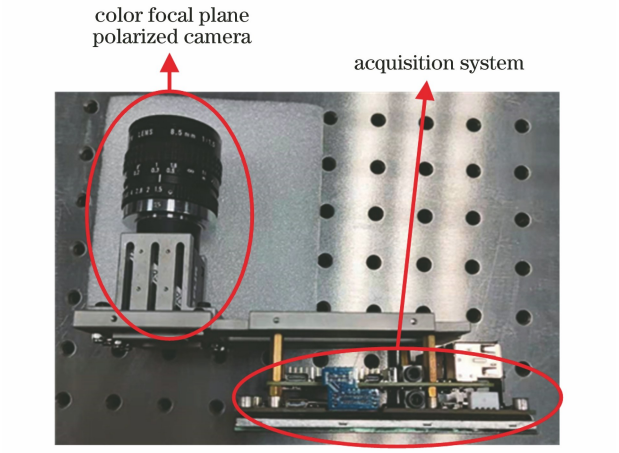

图 10　便携式采集设备的实物图

Fig. 10　Physical drawing of portable acquisition equipment

4.1.2　数据集的标注

由原始数据集解析得到的 RGB 图的尺寸与 4 通道偏振方向灰度图相等，为了更精准地标注伪装人员位置，选用 RGB 图进行标注，标注工具为精灵标图助手，由数据拍摄者和伪装者共同标注（部分图片难以较快找到伪装人员，拍摄者和伪装者对场景和位置熟悉，共同进行标注可提高效率）。

4.1.3　数据集的介绍

现有研究中主要有迷彩伪装人员的普通光学图像数据集，但公共数据集中还没有偏振图像数据集，因此本文构建了 CIP3K 数据集，其有 3000 张原始图，包含 20～25 个不同的自然背景。在构建数据集的过程中设计了多种伪装人员的活动姿态，具体分类如图 11 所示。

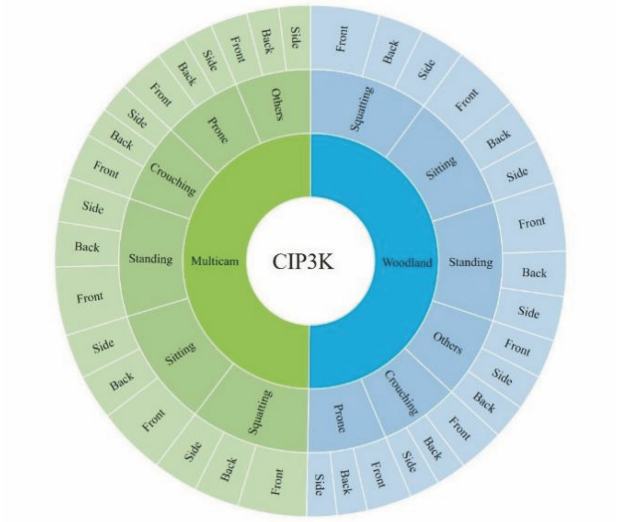

图 11　单兵伪装偏振图像数据集的分类示意图

Fig. 11　Schematic of classification of individual camouflage polarization image dataset

CIP3K 数据集包含 Woodland 型迷彩伪装和 Multicam 型迷彩伪装（每张图像包含一个伪装人员）两大类。根据伪装人员的姿态分为 6 个类别，即站姿、蹲姿、坐姿、伏姿、卧姿和其他自然活动姿态，其中每个姿态又可细分为正面、背面和侧面，划分比例为 4∶3∶3。

4.2　实验实施

实验中 Woodland 数据集和 Multicam 数据集各包含 1500 张图像，分别取 1400 张作为训练集，100 张作为测试集。实验平台为 Ubuntu16.04，8 个 NVIDIA GeForce GTX 1080TiGPU，8 个 Intel Xeon E5-1660 v4CPU。

4.2.1　参数设定

初始学习率设置为 0.02，学习率下降策略采用均匀下降，步长为{5，50，100}，优化算法采用随机梯度下降（SGD）法，动量设置为 0.9；权重衰减率取 0.0001；批训练图像数为 16，最大训练迭代次数为 14100，参数迭代次数设置为 150。

4.2.2　评价指标

本文采用平均检测精度均值（mAP）作为模型性能的评价指标，其反映了所有类标签的平均精确率。mAP 是所有类别平均检测精度（AP）的均值，AP 由精确率和召回率计算得到，精确率和召回率由目标检测中预测结果的正负例情况来决定，如表 3 所示。

表 3　正负例情况

Table 3　Positive and negative cases

Case	Prediction (positive)	Prediction (negative)
Ture(true)	TP	TN
Ture(false)	FP	FN

表 1 中 TP 表示被正确地预测为正例的个数，TN 表示被错误地预测为正例的个数，FP 表示被正确地预测为负例的个数，FN 表示被错误地预测为负例的个数。

根据表 3，精确率 P 的计算公式为

$$P=\frac{x_{\mathrm{TP}}}{x_{\mathrm{TP}}+x_{\mathrm{FP}}}。\tag{8}$$

召回率 R 的计算公式为

$$R=\frac{x_{\mathrm{TP}}}{x_{\mathrm{TP}}+x_{\mathrm{FN}}}。\tag{9}$$

AP 和 mAP 的计算公式分别为

$$\begin{cases}x_{\mathrm{AP}}=\dfrac{1}{\mu}\displaystyle\sum_{r\in\mu}p_{\mathrm{interp}}(r)\\ x_{\mathrm{mAP}}=\dfrac{1}{K}\displaystyle\sum_{k^{*}\in K}x_{\mathrm{AP}}(k^{*})\end{cases},\tag{10}$$

式中：μ 表示给定召回率 R 的个数；p_{interp} 表示大于给定召回率在所有召回率中对应的最大精确率；K 表示类别数。

另外还用交并比(IOU)来评价模型对目标定位的准确率，IOU 表示预测框与真实标注框之间的重叠度。将真实标注框区域面积记为 A，将预测框区域面积记为 B，则 IOU 可表示为

$$x_{\text{IOU}}=\frac{S(A\cap B)}{S(A\cup B)}。\tag{11}$$

IOU 值越接近 1，表明位置定位越准确。当判定预测框中的目标类别时，需要设置 IOU 的阈值。如果预测框与标注框之间的 IOU 值大于该阈值，那么判定该预测框是正例，反之为负例。实验中，IOU 值通常取 0.5。

4.3 实验结果

4.3.1 不同模型对比

将所提模型与 5 种主流检测模型在 Multicam 和 Woodland 两种类型的迷彩伪装人员数据集上的检测效果进行实验对比，包括 SSD(Single Shot MultiBox Detector)[17]、YOLO[18] 系列中的 YOLOv4[19] 和 YOLOv5、RetinaNet[10] 和 Faster R-CNN 5 种检测模型，结果如表 4 所示。经实验验证，当 TSF-Net 模型中的参数 v 取 109 时，所提算法的检测精度最优。

由表 4 可以看出，所提模型的检测精度明显高于另外 5 种检测模型，在两种数据集上的检测结果分别有 8.8～15.0 个百分点和 8.2～13.7 个百分点的提升。这是由于 TSF-Net 能够提取目标的偏振特征信息与 RGB 特征信息并能将它们融合，迷彩伪装目标与自然背景在颜色上的相似度很高，但它们的材质有较大差异，材质上的差异具体表现为粗糙度、纹理和边缘差异，这种差异可以通过偏振特性反映出来。因此，TSF-Net 的检测精度优于以 RGB 图像作为输入源的检测模型。另外，6 种模型在两种类型迷彩伪装数据集上的检测精度有所不同，在 Woodland 类型迷彩伪装数据集上的检测精度均略高于在 Multicam 类型迷彩伪装数据集上的检测精度。这是由于两种类型迷彩伪装目标的颜色和材质存在差异，且 Woodland 类型迷彩伪装与背景的颜色差异略大于 Multicam 类型迷彩伪装与背景的颜色差异。

表 4 不同模型的检测精度对比

Table 4 Comparison of detection accuracy of different models

Model	mAP /%	
	Multicam dateset	Woodland dataset
SSD	70.9	73.5
YOLOv4	71.5	73.4
YOLOv5	73.1	74.6
RetinaNet	75.2	77.5
Faster R-CNN	77.1	78.9
TSF-Net	85.9	87.1

接下来对比 6 种模型的实际检测效果，图 12 为两种迷彩伪装目标测试图，图 13 和图 14 为两种迷彩伪装目标的检测结果。

图 12 两种类型的迷彩伪装目标测试图。(a) Multicam 型迷彩；(b) Woodland 型迷彩

Fig. 12 Two types of camouflage target test diagram. (a) Multicam type camouflage; (b) Woodland type camouflage

从图 13 和图 14 可以看出，虽然 6 种模型都能大致检测出伪装人员的位置，但除了 TSF-Net 模型以外，对比的 5 种检测模型均存在误检的情况，而且检测到真实迷彩伪装人员的预测框置信得分均低于 TSF-Net 的预测框置信得分。由此表明，TSF-Net 的检测精度更高，其提取到的目标偏振特征信息在检测任务中起到了重要作用。

4.3.2 模型稳定性验证

为了验证单兵伪装人员的不同活动姿态对模型检测精度的影响，从 Multicam 型迷彩伪装数据集中挑出“坐姿正面”图像作为训练集，测试集分为三种情况：第一种是将“站姿正面”图像作为测试集，记为 SF(Sanding Front)；第二种是将“站姿侧面”图像作为测试集，记为SS(Sanding Side)；第三种是将

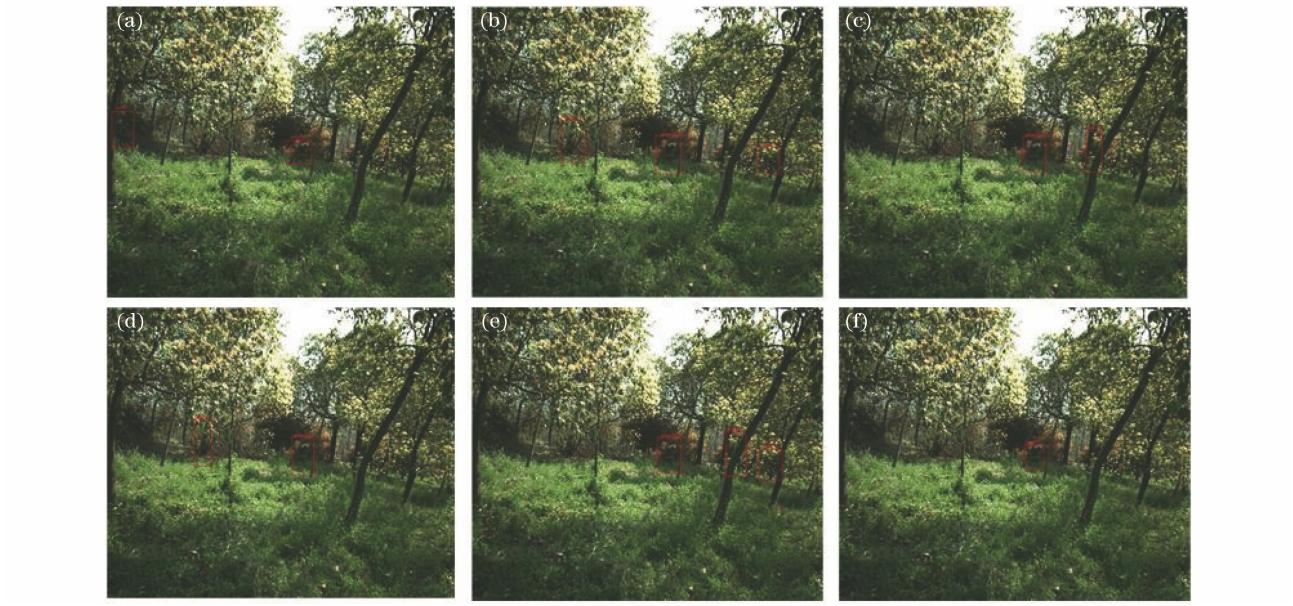

图 13 不同模型在 Multicam 数据集中的检测效果。(a) SSD 模型;(b) YOLOv4 模型;(c) YOLOv5 模型;(d) RetinaNet 模型;(e) Faster R-CNN 模型;(f) TSF-Net 模型

Fig. 13 Detection effects of different models in Multicam dataset. (a) SSD model; (b) YOLOv4 model; (c) YOLOv5 model; (d) RetinaNet model; (e) Faster R-CNN model; (f) TSF-Net model

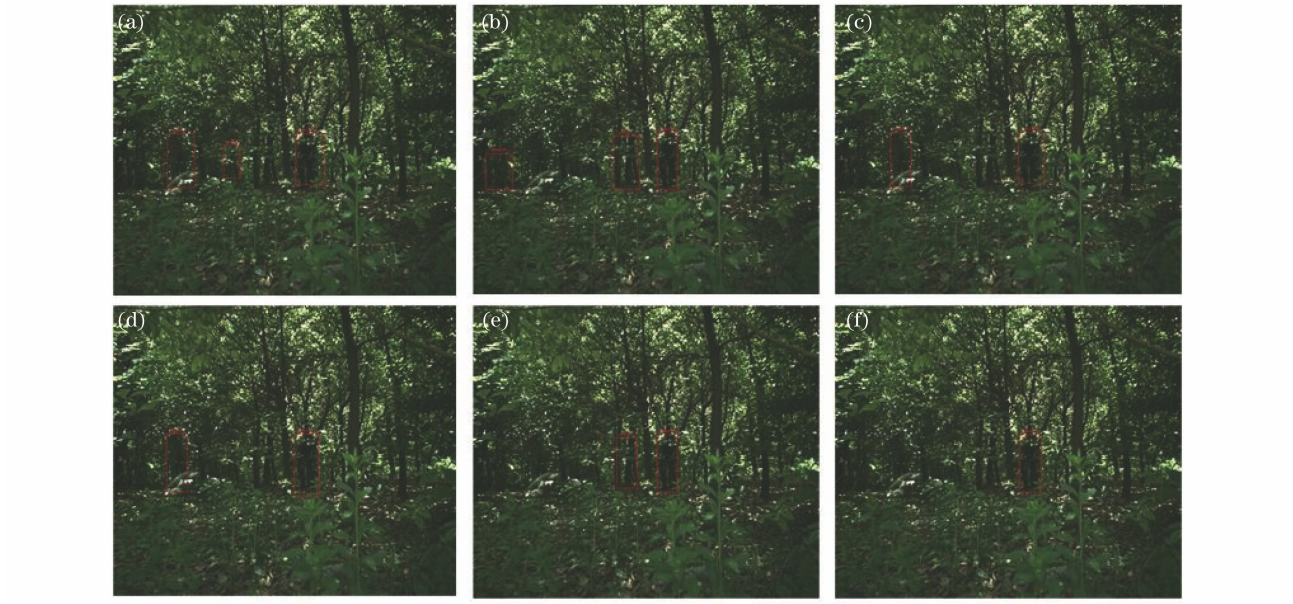

图 14 不同模型在 Woodland 数据集中的检测效果。(a) SSD 模型;(b) YOLOv4 模型;(c) YOLOv5 模型;(d) RetinaNet 模型;(e) Faster R-CNN 模型;(f) TSF-Net 模型

Fig. 14 Detection effects of different models in Woodland dataset. (a) SSD model; (b) YOLOv4 model; (c) YOLOv5 model; (d) RetinaNet model; (e) Faster R-CNN model; (f) TSF-Net model

“站姿背面”图像作为测试集,记为 SB(Sanding Back),结果如表 5 所示。实验过程中,模型参数的设置不变。

由表 5 可以看出,当训练集为“坐姿正面”图像,测试集为“站姿正面”图像、“站姿侧面”图像和“站姿背面”图像时,TSF-Net模型的检测精度均略低于

表 5 不同伪装人员的姿态测试结果对比

Table 5 Comparison of posture test results of different camouflage personnel

Model	mAP /%		
	SF	SS	SB
TSF-Net	84.9	84.3	84.6

在 Multicam 整体数据集上的测试结果。当测试集为“站姿正面”图像时，TSF-Net 模型的检测精度受到的影响最小，检测精度降低了 1 个百分点；当测试集为“站姿侧面”图像时，TSF-Net 模型的检测精度受到的影响最大，检测精度降低了 1.6 个百分点。由此表明，当单兵伪装人员的活动姿态比较复杂时，TSF-Net 模型能够保持较稳定的检测效果。这是由于 TSF-Net 模型是通过检测迷彩服来检测伪装人员，迷彩服与背景的材质差异较大，单兵伪装人员姿态的复杂性不会从根本上影响其表面材质与背景的偏振特征差异，因此 TSF-Net 模型能够保持较好的稳定性。

4.3.3　独立验证分析

1）参数验证

TSF-Net 模型的偏振流分支网络中的参数 v 决定了网络自适应得到的偏振参量数量，改变 v 值后测试模型性能，选取 Multicam 型迷彩伪装数据集进行实验，结果如图 15 所示。

如图 15 所示，v 共取 11 个值，其中随着 v 值的增加，TSF-Net模型的检测精度逐渐增加；当v值为 109 时，TSF-Net 模型的检测精度达到峰值 85.9%，据此可确立当 TSF-Net 模型中的偏振流分支网络结构为(48,96,32,16,109)时检测精度最优。

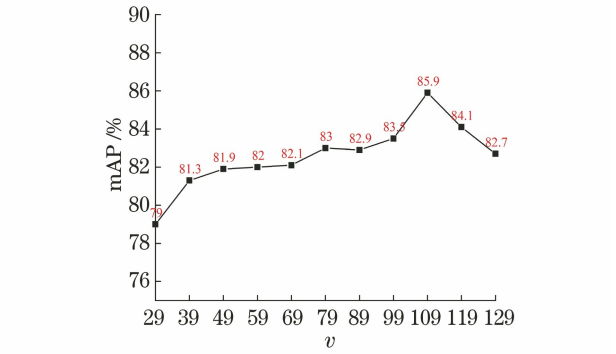

图 15　参数验证结果

Fig. 15　Parameter verification result

2）不同分支验证

将 Pol-Net 和 RGB-Net 两个分支分开进行性能测试，Pol-Net 同样改变参数 v 的取值后测试，选取 Multicam 型迷彩伪装数据集进行实验，得到的结果如图 16(a)所示，图 16(b)为 Pol-Net 与 RGB-Net 的 IOU-mAP 曲线图。

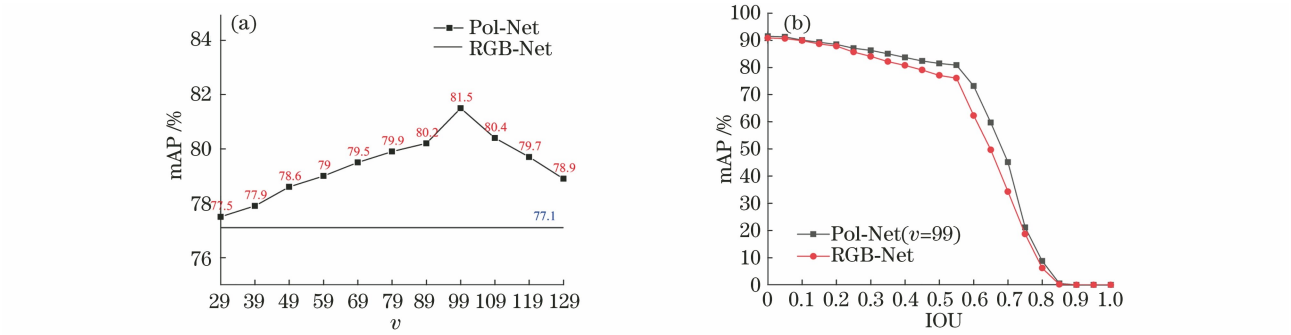

图 16　不同分支的验证结果。(a)不同 v 值的检测精度；(b) IOU-mAP 曲线

Fig. 16　Verification results for different branches. (a) Detection accuracy of different v values; (b) IOU-mAP curves

图 16(a)中的直线代表 RGB-Net 模型的检测精度，曲线代表 Pol-Net 结构中 v 取不同数值的检测精度。从图 16 可以看到：无论 v 取任何值，Pol-Net 模型的检测精度始终大于 RGB-Net 模型，单兵伪装目标与背景的材质差异大于颜色差异，目标与背景的偏振特征信息差异大于 RGB 特征信息；当 v 值为 29 时，RGB-Net 模型的检测精度接近 Pol-Net 模型，表明当自适应偏振参量较少时，Pol-Net 模型不能较好地反映目标偏振特性。从图 16(b)可以看出，当 IOU 在 0.3～0.7 之间时，Pol-Net 模型的检测精度比 RGB-Net 模型大。

4.3.4　交叉验证分析

对两种类型的数据集进行交叉验证实验，分为两种情况。第一种情况(M/W)：采用 Multicam 型迷彩伪装数据集中的 1400 张图像作为训练集，Woodland 型迷彩伪装数据集中的 100 张图像作为测试集；第二种情况(W/M)：采用 Woodland 型迷彩伪装数据集中的 1400 张图像作为训练集，Multicam 型迷彩伪装数据集中的 100 张图像作为测试集。实验过程中，模型的参数设置不变。交叉验证对比结果如表 6 所示。

表 6　交叉验证对比

Table 6　Cross-validation comparison

Model	mAP /%	
	M/W	W/M
Faster R-CNN	23.5	15.7
TSF-Net	48.8	35.5

从表 6 可以看出，交叉验证实验中两种模型的检测精度都有明显下降，但在 M/W 和 W/M 两种情况下，TSF-Net 模型的检测精度分别是 Faster R-CNN 模型的 2.08 倍和 2.26 倍，TSF-Net 模型的泛化能力明显高于 Faster R-CNN 模型。原因在于 TSF-Net 模型不仅提取了单兵伪装目标的 RGB 特征信息，还提取了单兵伪装目标的偏振特征信息，单兵伪装目标与背景的颜色差异小于其材质与背景的差异，即 RGB 特征差异小于偏振特征差异。

在交叉验证中，影响 TSF-Net 检测精度的原因有如下两部分。

1) 迷彩服材质差异。TSF-Net 主要通过提取单兵伪装人员表面迷彩服的偏振特征信息来达到检测伪装人员的目的，不同种类的迷彩服，材质不同，材质的差异会导致其偏振特征也具有差异性。材质差异越大，偏振特性差异越大，在交叉验证实验中模型的检测精度也会越差。

2) 数据集采集环境差异。目标的偏振特征不仅会受到材质的影响，也会受到光照的影响，其中材质为主要影响因素。两种迷彩类型数据集的采集环境并不是完全一致的，当采集环境不同时，光照情况有所差异，从而导致迷彩伪装偏振特性产生差异，因此在交叉验证实验中模型的检测精度也会受到影响。

5 结　论

本文建立了基于偏振图像与 RGB 图像的单兵伪装目标数据集，即 CIP3K 数据集，其包含 20～25 种自然背景并包含丰富的伪装人员姿态。在 Faster R-CNN 的基础上提出了一种基于双流特征融合的单兵伪装偏振成像检测网络 TSF-Net，其根据单兵伪装目标表面材质与自然背景材质差异较大的特点，通过双流特征融合网络来学习丰富稳定的伪装目标特征信息，其中偏振流网络负责提取目标偏振特征信息，这有利于区分目标与背景，RGB 网络负责提取目标的 RGB 特征信息，该特征信息对于边缘细节的定位更准确，最后将偏振特征与 RGB 特征进行融合，实现特征信息的互补。通过在 CIP3K 数据集上的大量实验验证，TSF-Net 的检测精度比 Faster R-CNN 高了 8.2 个百分点和 8.8 个百分点，优于一些主流目标检测模型，且模型的泛化能力优于 Faster R-CNN。在后续工作中，扩展数据集以及优化偏振特征信息与 RGB 特征的融合是本课题组研究的重点。

参　考　文　献

[1] 关世豪, 杨桄, 卢珊, 等. 基于注意力机制的多目标优化高光谱波段选择[J]. 光学学报, 2020, 40(21): 2128002.
Guan S H, Yang G, Lu S, et al. Multi-objective optimization of hyperspectral band selection based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(21): 2128002.

[2] 赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测[J]. 光学学报, 2020, 40(5): 0504001.
Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica, 2020, 40(5): 0504001.

[3] 闫芬婷, 王鹏, 吕志刚, 等. 基于视频的实时多人姿态估计方法[J]. 激光与光电子学进展, 2020, 57(2): 021006.
Yan F T, Wang P, Lü Z G, et al. Real-time multi-person video-based pose estimation[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021006.

[4] Zhang L, Xie W, Zhao F, et al. Deep learning based classification using semantic information for POLSAT image[C]//IEEE International Geoscience and Remote Sensing Symposium, September 26-October 2, 2020, Waikoloa, HI, USA. New York: IEEE Press, 2020: 196-199.

[5] 张祥东, 王腾军, 朱劭俊, 等. 基于扩张卷积注意力神经网络的高光谱图像分类[J]. 光学学报, 2021, 41(3): 0310001.
Zhang X D, Wang T J, Zhu S J, et al. Hyperspectral image classification based on dilated convolutional attention neural network[J]. Acta Optica Sinica, 2021, 41(3): 0310001.

[6] Zhao Y M, Zhao J Z, Zhao C Y, et al. Robust real-time object detection based on deep learning for very high resolution remote sensing images[C]//IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 1314-1317.

[7] Fang Z, Zhang X, Deng X, et al. Camouflage people detection via strong semantic dilation network[C]// Proceedings of the ACM Turing Celebration Conference, May 17-19, 2019, Chengdu, China. New York: ACM, 2019.

[8] Zheng Y F, Zhang X W, Wang F, et al. Detection of people with camouflage pattern via dense deconvolution network[J]. IEEE Signal Processing Letters, 2018, 26(1): 29-33.

[9] 邓小桐, 曹铁勇, 方正, 等. 改进 RetinaNet 的伪装

人员检测方法研究[J]. 计算机工程与应用, 2021, 57(5): 190-196.
Deng X T, Cao T Y, Fang Z, et al. Research on detection of people with camouflage pattern via improving RetinaNet[J]. Computer Engineering and Applications, 2021, 57(5): 190-196.

[10] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 42(2): 318-327.

[11] 王杨, 曹铁勇, 杨吉斌, 等. 基于 YOLO v5 算法的迷彩伪装目标检测技术研究[J]. 计算机科学, 2021, 48(10): 226-232.
Wang Y, Cao T Y, Yang J B, et al. Camouflaged object detection based on improved YOLO v5 algorithm[J]. Computer Science, 2021, 48(10): 226-232.

[12] Le T N, Nguyen T V, Nie Z L, et al. Anabranch network for camouflaged object segmentation[J]. Computer Vision and Image Understanding, 2019, 184: 45-56.

[13] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[14] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on Machine Learning, July 6-11, 2015, Lille, France. Cambridge: JMLR, 2015: 448-456.

[15] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//14th International Conference on Artificial Intelligence and Statistics, April 11-13, 2011, Fort Lauderdale, USA. Cambridge: JMLR, 2011: 315-323.

[16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[17] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[18] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[19] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-04]. https://arxiv.org/abs/2004.10934.