

文章编号 1004-924X(2024)02-0301-16

多感受野特征自适应融合及动态损失调整的 初烤烟叶等级检测

何自芬, 罗 洋, 张印辉*, 陈光晨, 陈东东, 徐 林
(昆明理工大学 机电工程学院, 云南 昆明 650500)

摘要:初烤烟叶等级的快速准确检测对开发烟叶智能分级设备以促进农产品精细化管理有着重要意义。针对相似度较高但等级不同的初烤烟叶难以区分的问题,本文提出多感受野特征自适应融合及动态损失调整的初烤烟叶等级检测网络(Flue-cured Tobacco Leaf Grade Detection Network, FTGDNet)。首先,FTGDNet采用CSPNet作为特征提取主干网络,采用GhostNet作为辅助特征提取网络以增强模型的特征提取能力;其次,在主干网络末端嵌入显式视觉中心瓶颈模块(Explicit Visual Center Bottleneck module, EVCB)以实现全局特征信息与局部细节特征信息融合;然后,构建多感受野特征自适应融合模块(Multi-Receptive Field Feature Adaptive Fusion module, MRFA_d),利用注意力特征融合机制(Attention Feature Fusion, AFF)将不同感受野特征图进行自适应加权融合,在增强模型局部感受野的同时突出有效通道信息;最后,设计了一种新的定位损失函数(More Complete IoU Loss, MCIOU_Loss),结合预测框与真实框面积损失以解决在回归定位过程中二者宽高比相等且中心点重合时CIOU_Loss性能退化导致定位精度下降问题,此外,引入矩形相似度衰减系数在训练过程中对真实框与预测框的相似度判别项进行动态调整,加快模型拟合。实验结果表明,FTGDNet对十个等级的初烤烟叶的验证精度达到90.0%,测试精度达到87.4%,且推理时间仅为12.6 ms。相较于多种先进目标检测算法,FTGDNet具有更高的检测精度和更快的检测速度,可为高精度初烤烟叶等级检测提供关键技术支撑。

关 键 词:初烤烟叶;目标检测;多感受野特征融合;动态损失调整

中图分类号:TP391.4;S24 **文献标识码:**A **doi:**10.37188/OPE.20243202.0301

Flue-cured tobacco leaf grade detection through multi-receptive field features fusing adaptively and dynamic loss adjustment

HE Zifen, LUO Yang, ZHANG Yin-hui*, CHEN Guang-chen, CHEN Dong-dong, XU Lin

(Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology,
Kunming 650500, China)

* Corresponding author, E-mail: zyhhzf1998@163.com

Abstract: Rapid and accurate detection of flue-cured tobacco leaf grade is integral to the advancement of tobacco intelligent equipment, promoting refined management of agricultural products. Aiming at the issue that it is difficult to distinguish flue-cured tobacco leaves with high similarity between different grades, a flue-cured tobacco leaf grade detection network (FTGDNet) through multi-receptive field feature fusing

收稿日期:2023-05-19;修订日期:2023-07-05.

基金项目:国家自然科学基金资助项目(No. 62171206, No. 62061022);中国烟草云南分公司烟叶智能分级项目资助(No. HZ2021K0462A)

adaptively and dynamic loss adjustment was proposed. Firstly, FTGDNet adopted CSPNet and GhostNet as feature extraction backbone network and auxiliary feature extraction network to enhance the model feature extraction ability, respectively; Secondly, to merge global feature information and local detail feature information, an explicit visual center bottleneck module (EVCB) was embedded at the end of backbone network; Moreover, a multi-receptive field feature adaptive fusion module (MRFA_d) was constructed, in which the attention feature fusion (AFF) mechanism adaptively fuses the weights of feature maps with different receptive fields to highlight the effective channel information while enhancing the local receptive fields of the model; In addressing the decrease of positioning accuracy due to CIOU_Loss performance degradation when the prediction box and real box shared the same aspect ratio and their centers align during the regression positioning process, a new positioning loss function MCIoU_Loss was designed, In addition, the rectangular similarity attenuation coefficient was introduced to dynamically adjust the similarity discriminant of prediction box and real box to accelerate the model fitting. The experimental results show that the verification accuracy and test accuracy of FTGDNet for 10 grades of flue-cured tobacco leaf reached 90.0% and 87.4%, respectively, with an inference time of 12.6 ms. Compared with various advanced object detection network, FTGDNet achieves higher detection accuracy and faster detection speed, which could provide technical support for high-precision flue-cured tobacco leaf grade detection.

Key words: flue-cured tobacco leaf; object detection; multi receptive field feature fusion; dynamic loss adjustment

1 引 言

2021 年中国烟草种植面积达 1 013 千公顷, 初烤烟叶产量达 202.01 万吨^[1], 贡献 1.2 万亿财政税收。初烤烟叶分级^[2]是烟叶实际收购过程中的重要环节, 分级结果直接决定了初烤烟叶的价格以及烟草种植农户的利益。目前, 由于分级精度不高, 烤烟分级自动化方法未得到广泛应用, 因此, 每年烤烟采收季节, 都需投入大量人力和财力对初烤烟叶进行人工分级。此外, 人工分级结果依赖于专家经验知识, 使得分级过程主观化、经验化, 从而导致分级结果缺乏客观依据。为解决以上初烤烟叶分级任务中所存在的问题, 研究一种精度高、速度快的初烤烟叶等级检测算法对提高自动化分级设备性能具有重要意义。

根据中国烤烟国家标准 GB2635-1998, 初烤烟叶颜色、成熟度、油分、长度、身份、残伤、叶片结构是决定其等级的重要因素。为了更好的对初烤烟叶进行分级, 研究人员提出了基于机器视觉的初烤烟叶分级方法, 主要分为传统视觉方法和深度学习方法。传统特征提取方法主要采用图像色彩阈值划分和形态学计算等方法来提取初烤烟叶图像的颜色、脉络等显式相关表层特

征, 再根据人工经验判断其等级。Zhang 等^[3]提出了一种表示初烤烟叶特征分布的二维特征空间, 利用传统数字图像处理技术提取初烤烟叶的颜色、纹理和形状参数等外观特征, 再与专家标准进行对比, 实现初烤烟叶分级。Tattersfield 和 Forbes^[4]提出一种颜色分析方法, 将 RGB 色彩空间转换为孟塞尔颜色体系来模拟人类视觉, 对初烤烟叶进行颜色分组。Han^[5]利用支持向量机对烟叶生长部位进行识别。Marcelo 等^[6]采用近红外高光谱成像技术结合支持向量机对初烤烟叶进行分级。尽管传统视觉分级方法相较于人工分级取得了很大进步, 但其在初烤烟叶图像预处理时需要人工调整大量超参数, 且无法提取叶片的深层语义特征信息, 从而限制了其分级精度和效率。

随着深度学习技术的发展, 可实现自动提取特征并进行分类的卷积神经网络被广泛应用于农业生产中。林相泽等^[7]结合字典学习与 SSD 目标检测网络, 对不完整稻飞虱图像进行识别, 为稻飞虱的预防和监督提供了技术支持。Zhang 等^[8]设计了一种基于区域分割和 AlexNet 的快速定位分类模型, 并将其部署于分拣机器人上, 用于茶叶分拣, 最高分选准确率达 92%。Shang

等^[9]利用轻量级深度学习算法对自然环境下的苹果花进行实时检测。在卷积神经网络的应用提高了农产品生产效率的同时,深度学习方法也被应用在初烤烟叶分级任务中,曾祥云^[10]改进了经典 AlexNet 网络,增强了模型的学习能力,提高了初烤烟叶分级的准确率;Dasari 和 Prasad^[11]使用卷积神经网络从三个等级的初烤烟叶图像中自动提取特征,分级准确率达到 85.10%,但作者仅应用了 120 张初烤烟叶图像,数据集图像数量较少,这难以使神经网络充分学习初烤烟叶特征;Chen 等^[12]使用 MobileNetV2 和 Swin Transformer 组成的并行双编码器结构对 8 个等级的初烤烟叶进行分级,分级精度达到 79.30%,但其模型图像处理时间达到每帧 95 ms,无法达到实时检测的需求。Lu 等^[13]搭建了一种细粒度视觉分类框架并将其应用于初烤烟叶分级任务中,六个等级的初烤烟叶分级准确率达到 80.65%,尽管单帧初烤烟叶图像处理时间缩短至 42.1 ms,但仍然达不到实时检测的需求。

YOLOv5 目标检测算法在检测精度和检测速度方面均具有优秀的性能,广泛应用于缺陷检测^[14]、行人检测^[15]、异物检测^[16]等方面。将 YOLOv5 用于初烤烟叶等级检测任务时,随着网络深度的增加,一些对初烤烟叶等级检测有益的深层特征信息随之丢失;缺乏对叶片的局部特征信息的关注,导致其无法有效区分相似度较高的不同等级的初烤烟叶;在模型定位过程中,真实框与预测框宽高比相同且中心点重合但真实框与预测框并未重合时,模型的定位损失函数性能退化。针对上述问题,本文提出一种多感受野特征自适应融合及动态损失调整的初烤烟叶等级检测网络(Flue-cured Tobacco Leaf Grade Detection Network, FTGDNet)。首先利用 CSPNet^[17]作为特征提取主干网络,利用 GhostNet^[18]作为辅助特征提取网络,在 CSPNet 和 GhostNet 末端将所提取的特征进行像素相加以增强模型特征提取能力,保留更多对初烤烟叶等级检测有益的深层特征信息;使用显式视觉中心瓶颈模块 EVCB 将全局信息与局部信息进行融合;嵌入多感受野特征自适应融合模块 MRFA 增强模型局部感受野的同时突出有效通道信息;使用本文提出的 MCIOU_Loss 代替 CIOU_Loss, MCIOU_Loss 结

合了真实框与预测框面积损失以及矩形相似度衰减系数,加快模型拟合的同时有效改善真实框与预测框宽高比相同且中心点重合时模型定位损失函数退化问题。

2 多感受野特征自适应融合及动态损失调整的初烤烟叶等级检测网络

2.1 FTGDNet 网络结构

FTGDNet 深度学习模型运用回归思想,以一阶段网络完成目标定位及分类。如图 1 所示,FTGDNet 沿用了 YOLOv5 的整体布局,其网络结构由 Input, Backbone, Neck 以及 Output 四部分组成。

FTGDNet 在 Input 部分将图像缩放为 640×640 。Backbone 的第一层是 Focus 模块,Focus 模块将输入图像进行 4 次切片操作,切片操作采用间隔像素点采样的方式,将原图像宽度 W 、高度 H 转换到通道空间,输入通道由原 RGB 3 通道变为 12 通道,再通过卷积操作,将输入图像由 $640 \times 640 \times 3$ 变为 $320 \times 320 \times 32$ 。受启发于 DenseNet^[19]的密集跨层连接思想,CSPNet 利用不同层的特征信息进行局部跨层融合以获得更为丰富的特征图。GhostNet 的核心思想是以更少的参数生成更多的特征信息,进而达到减少模型参数量同时加快模型推理速度的目的。为增强模型主干特征提取能力,FTGDNet 采用 CSPNet 作为特征提取主干,采用 GhostNet 进行辅助特征提取。空间金字塔池化模块 SPP 使用 $5 \times 5, 9 \times 9, 13 \times 13$ 池化核进行最大池化操作,再将不同尺度的特征图进行拼接,增加了网络提取全局信息的能力。在 Backbone 末端嵌入具有全局特征和局部特征相结合功能的 EVCB 模块作为瓶颈层,随后添加 MRFA_d 模块 EVCB 模块和 MRFA_d 模块将在 2.2 和 2.3 详细描述。Neck 部分由特征金字塔网络 (Feature Pyramid Networks, FPN) 和路径聚合网络 (Path Aggregation Network, PAN) 组成,FPN 自顶向下将高层特征信息与不同 C3 模块的输出进行聚合,PAN 自底向上聚合浅层特征,进而充分融合不同层的图像特征。Output 部分使用 MCIOU_Loss 作为

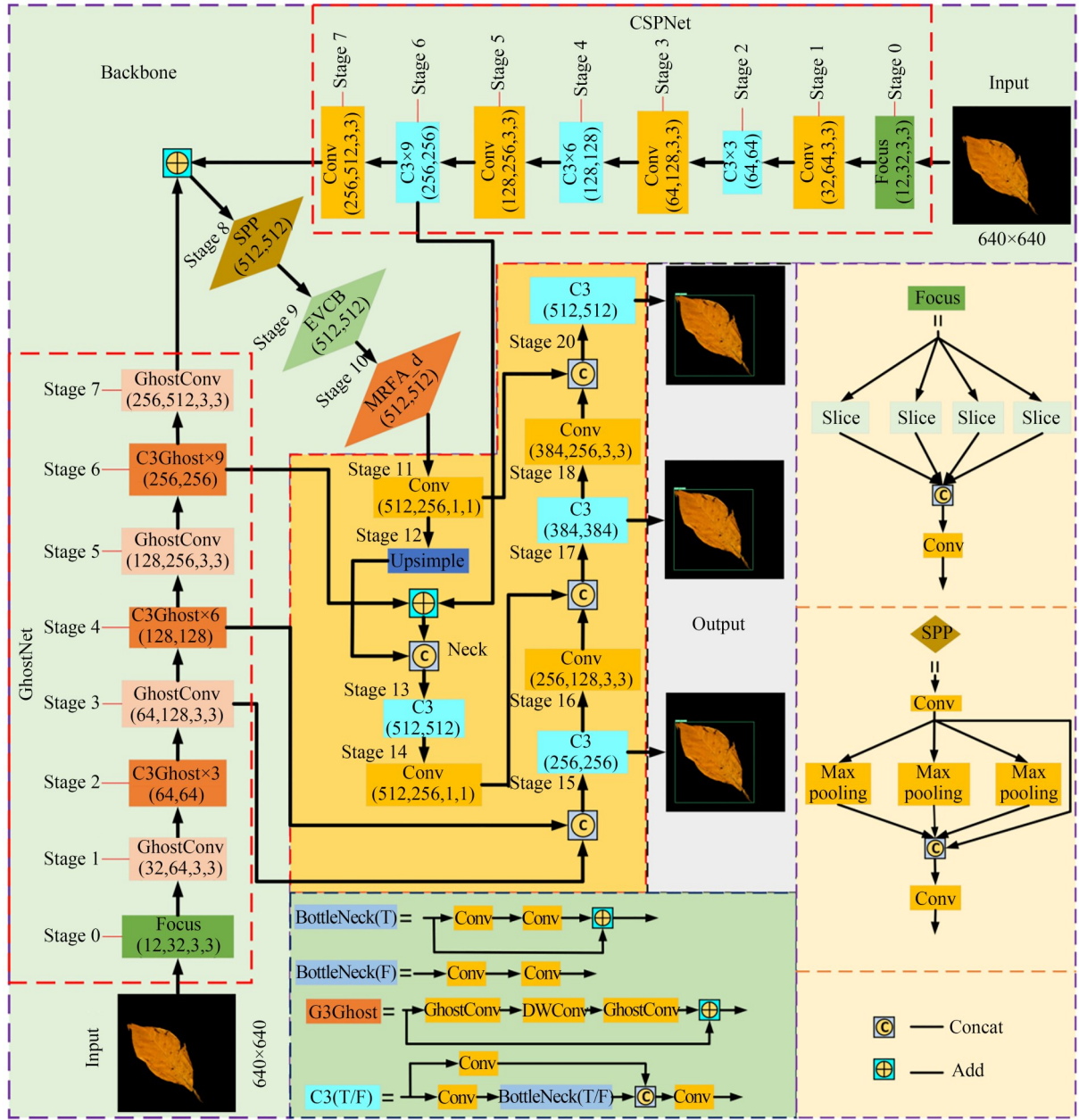


图1 FTGDNet网络结构

Fig. 1 FTGDNet network structure

定位损失函数,加快模型拟合的同时有效改善真实框与预测框宽高比相等且中心点重合时 CIoU_Loss 性能退化问题。

2.2 显式视觉中心瓶颈模块

显式视觉中心瓶颈模块(Explicit Visual Center Bottleneck module, EVCB)主要由可学习视觉中心模块(Learnable Visual Center, LVC)和轻量级多层感知机模块(Lightweight MLP, LMLP)两部分构成,其结构如图2所示。可学习视觉中

心模块中存在一个拥有固有码本和缩放因子的编码器,其中固有码本 $B = \{b_1, b_2, \dots, b_k\}$, 缩放因子 $S = \{s_1, s_2, \dots, s_k\}$ 为可学习的视觉中心。对于输入特征(在FTGDNet中,该输入来自SPP模块的输出),先使用 7×7 卷积平滑所输入的特征信息,再利用一组卷积(由 1×1 卷积、 3×3 卷积、 1×1 卷积、 3×3 卷积 + BN + ReLU 构成)对该特征进行编码,将编码后的特征 x_i 输入 Codebook 中,使用缩放因子 s_k 关联 x_i 与 b_k ,进而映射出 x_i

与 b_k 的位置信息,上述过程表达为如式(1)所示:

$$e_k = \sum_{i=1}^N \frac{e^{s_i(x_i - b_k)^2}}{\sum_{j=1}^K e^{s_j(x_i - b_j)^2}} (x_i - b_k), \quad (1)$$

其中: e_k 为输入特征图关于第 k 个码字的位置信息, N 为输入特征的像素点总数, K 为视觉中心的总数, x_i 为第 i 个像素点, b_k 为第 k 个码字, s_k 为第 k 个缩放因子。

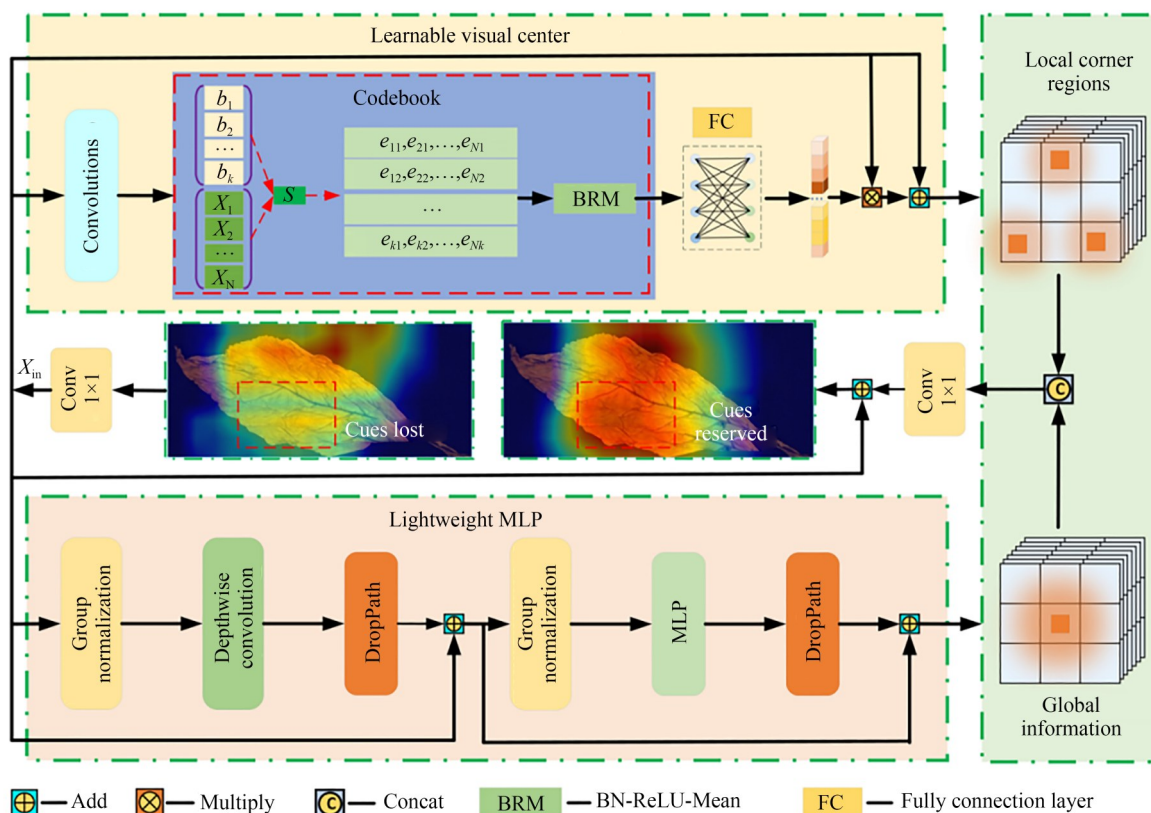


图2 EVCB模块结构

Fig. 2 EVCB module structure

使用BRM融合所有的 e_k 得到所有像素点关于全部码字的位置信息,将该信息输入全连接层中,突出具有类别区分性的局部信息,再使用view函数将FC输出的特征图大小变为 1×1 ,同时将 7×7 卷积的输出与view函数的输出 X_{in} 进行通道相乘,该过程可以表示为:

$$Y = X_{in} \otimes R(\text{view}(FC(BRM(e_k)))) \quad (2)$$

其中: \otimes 表示通道相乘, R 表示 ReLU 激活函数。

最后将 Y 与 X_{in} 进行通道拼接,得到包含局部特征信息的LVC的输出,上述过程表示为:

$$LVC(X_{in}) = X_{in} \oplus Y \quad (3)$$

其中, \oplus 表示像素相加。

轻量级多层感知机模块由深度卷积残差模块和通道多层感知机残差模块构成,其中,通道多层感知机残差模块的输入是深度卷积残差模

块的输出。在两个残差模块中均对输入特征沿着通道维度进行分组操作,同时使用DroupPath操作来提高模型的鲁棒性。具体来说,来自 7×7 卷积的特征信息 X_{in} 输入深度卷积残差模块,该过程表示为:

$$\tilde{X}_{in} = DConv(GN(X_{in})) \oplus X_{in} \quad (4)$$

其中: \tilde{X}_{in} 表示深度卷积残差模块的输出, DConv 表示深度卷积, GN 表示输入特征沿着通道维度进行分组的操作。

相较于空间多层感知机,通道多层感知机在保证模型检测性能的前提下,能够有效降低模型的计算复杂度。通道多层感知机残差模块信息处理过程可以表示为:

$$\hat{X}_{in} = CMLP(GN(\tilde{X}_{in})) \oplus \tilde{X}_{in} \quad (5)$$

其中: \hat{X}_{in} 表示通道多层感知机残差模块的输出, CMLP 表示通道多层感知机。 \hat{X}_{in} 也即 LMLP 模块的输出, 即 $LMLP(X_{in}) = \hat{X}_{in}$ 。

将 LVC 模块和 LMLP 模块的输出进行通道拼接, 随后使用 1×1 卷积将通道数减半, 最后与 X_{in} 进行像素相加得到 EVCB 模块的输出, 上述过程表示为:

$$EVCB(X_{in}) = \text{Conv}(\text{Concat}(LVC(X_{in}), LMLP(X_{in})) + X_{in}), \quad (6)$$

其中, Concat 表示通道拼接。

LVC 能够有效保留局部特征信息, 而 LMLP

则有效获取具有长距离依赖的全局特征信息, EVCB 模块将二者进行结合, 有效提高模型的特征能力进而提高模型检测精度。

2.3 多感受野特征自适应融合模块

多感受野特征自适应融合模块 (Multi-Receptive Field Feature Adaptive Fusion module, MRFA) 结构如图 3 所示。MRFA 包含三层标准卷积结构, 第一层和第三层均为 1×1 卷积, 主要用于减少通道数。第二层卷积结构由卷积核为 $3 \times 3, 5 \times 5, 7 \times 7$ 的卷积并行而成, 用于制造具有不同感受野的特征图。在卷积神经网络中, 感受野是指卷积神经网络每层输出的特征图上的像

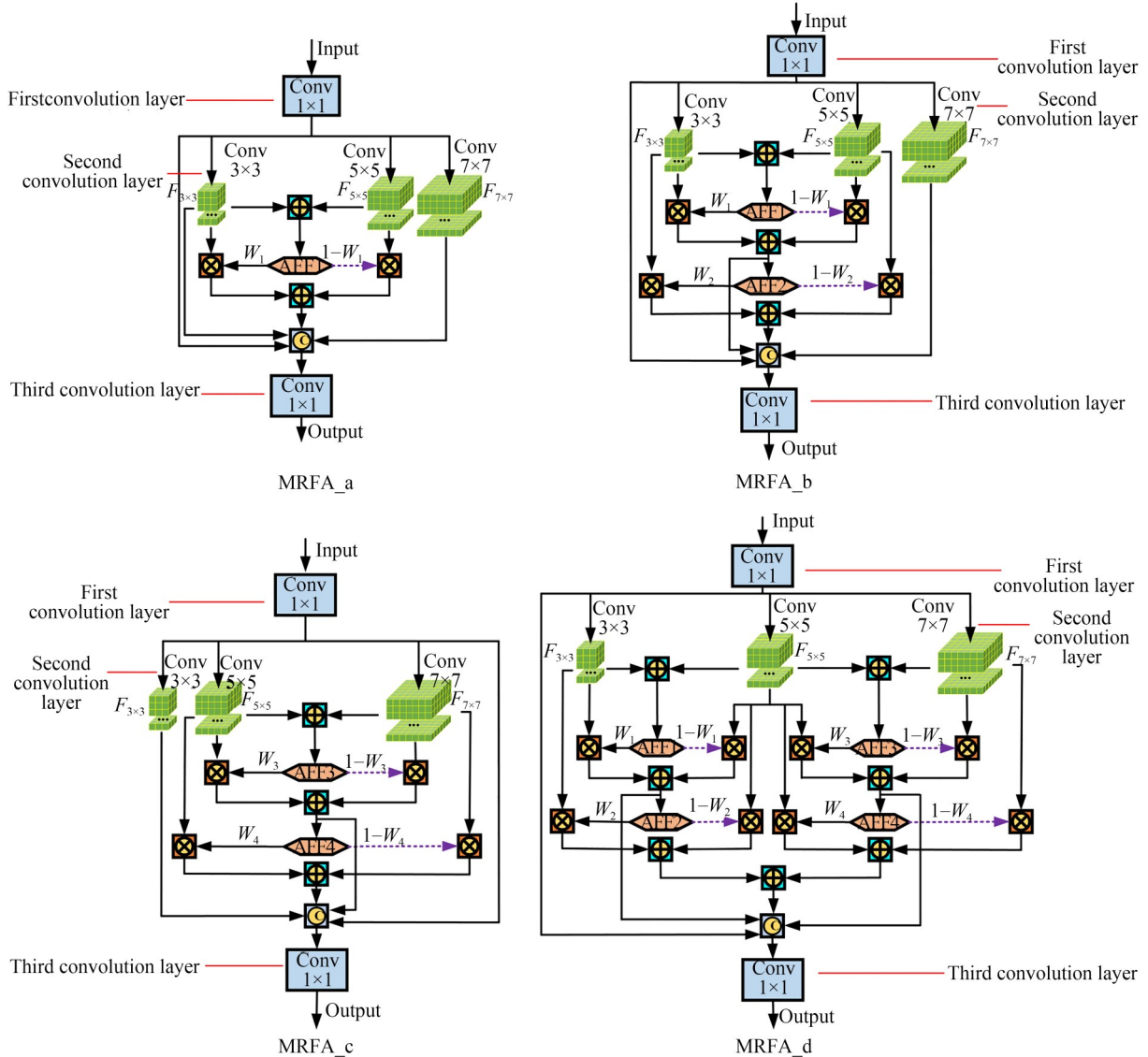


图 3 四种 MRFA 结构对比

Fig. 3 Comparison of four MRFA structures

素点在原始图像上映射区域的大小。感受野尺寸的求解相当于已知输出特征图大小反向求出输入特征图的过程。计算感受野大小时,常采用自顶向底的方式计算,即先计算最终特征图在前一层上的感受野,然后逐渐传递到第1层。MRFA中所使用卷积均为标准卷积且卷积步长 s 均为1,其感受野计算公式如式(7)所示:

$$l_m = (l_{m-1} - 1)s + n, \quad (7)$$

其中: l_m 为第 m 层感受野, l_{m-1} 为第 $m-1$ 层感受野, s 为第 m 层卷积步长, n 为第 m 层卷积核尺寸。

MRFA第三层卷积输出特征图感受野为 1×1 ,即 $l_3 = 1$,根据式(8)计算,第二层 $3 \times 3, 5 \times 5, 7 \times 7$ 卷积输出特征图感受野分别为 $3 \times 3, 5 \times 5, 7 \times 7$,即 $l_2^{3 \times 3} = 3, l_2^{5 \times 5} = 5, l_2^{7 \times 7} = 7$,卷积核尺寸为 7×7 的卷积输出特征图感受野最大,再次利用式(7)计算,得到MRFA模块输出的特征图的一个像素点在原始特征图的感受野为 7×7 。

MRFA利用不同卷积核尺寸卷积制造了感受野为 $3 \times 3, 5 \times 5, 7 \times 7$ 的特征图,定义相邻感受野特征图中较小感受野特征图为 $F_{a \times a}$,感受野较大特征图为 $F_{b \times b}$,将 $F_{a \times a}$ 与 $F_{b \times b}$ 进行像素相加,再将像素相加结果输入注意力特征融合AFF机制中,获得通道注意力权重 W 。上述过程表示为:

$$W = \text{AFF}(F_{a \times a} \oplus F_{b \times b}). \quad (8)$$

AFF是一种类似于压缩激励(Squeeze-and-Excitation, SE)^[20]注意力机制的结构,能够在通道方向上筛选出重要的特征信息,其原理如图4所示。在获得通道注意力权重 W 后,权重 W 与 $F_{a \times a}$ 进行通道相乘,同时使用 $1 - W$ 与 $F_{b \times b}$ 进行通道相乘,再将通道相乘结果进行像素相加,即对相邻感受野特征图进行加权融合,上述过程表示为:

$$Z = W \otimes F_{a \times a} \oplus (1 - W) \otimes F_{b \times b}, \quad (9)$$

其中: Z 表示相邻感受野特征图进行加权融合输出, W 表示通道注意力权重。

本文设计了四种MRFA结构,分别为图3所示的MRFA_a, MRFA_b, MRFA_c, MRFA_d。定义 $3 \times 3, 5 \times 5, 7 \times 7$ 感受野特征图为 $F_{3 \times 3}, F_{5 \times 5}, F_{7 \times 7}$ 。MRFA_a将 $F_{3 \times 3}, F_{5 \times 5}$ 进行自适应加权融合,随后将其与感受野为 3×3 和 7×7 以及第一层卷积输出的特征图进行通道拼接,最后

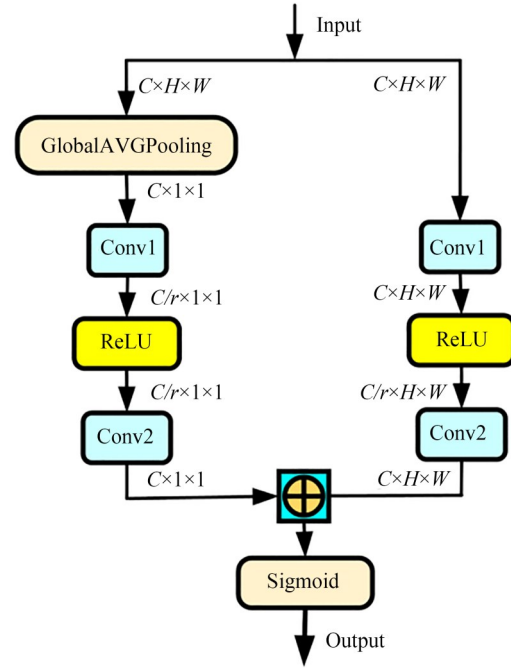


图4 AFF机制

Fig. 4 AFF mechanism

再输入 1×1 卷积中将通道数减半。MRFA_a特征处理过程可以表示为:

$$\text{Output}_{\text{MRFA}_a} =$$

$$\text{Conv}(\text{Concat}(C_{3 \times 3}, C_{\text{AFF1}}, C_{7 \times 7}, C_0)), \quad (10)$$

其中: Conv 表示卷积核尺寸为 1×1 的卷积, Concat 表示通道拼接, $C_{3 \times 3}, C_{5 \times 5}, C_{7 \times 7}$ 分别表示 $F_{3 \times 3}, F_{5 \times 5}, F_{7 \times 7}$ 对应通道, C_{AFF1} 表示 $F_{3 \times 3}$ 与 $F_{5 \times 5}$ 自适应加权融合特征对应通道, C_0 表示第一层卷积输出特征通道。

MRFA_b在MRFA_a的基础上进行二次加权特征融合,进一步增强特征图中的有用信息,MRFA_b特征处理过程表示为:

$$\text{Output}_{\text{MRFA}_b} =$$

$$\text{Conv}(\text{Concat}(C_{\text{AFF1}}, C_{\text{AFF2}}, C_{7 \times 7}, C_0)), \quad (11)$$

其中: C_{AFF2} 表示 $F_{3 \times 3}$ 与 $F_{5 \times 5}$ 二次加权融合特征对应通道。

MRFA_c与MRFA_b相似,所不同的是MRFA_c将 $F_{5 \times 5}$ 与 $F_{7 \times 7}$ 进行二次加权融合,其特征处理过程表示为:

$$\text{Output}_{\text{MRFA}_c} =$$

$$\text{Conv}(\text{Concat}(C_{\text{AFF3}}, C_{\text{AFF4}}, C_{3 \times 3}, C_0)), \quad (12)$$

其中: C_{AFF3} 表示 $F_{5 \times 5}$ 与 $F_{7 \times 7}$ 加权融合特征对应通道, C_{AFF4} 表示 $F_{5 \times 5}$ 与 $F_{7 \times 7}$ 二次加权融合特征对应通道。

MRFA_d 综合 MRFA_b 与 MRFA_c, 同时将 $F_{3 \times 3}$ 与 $F_{5 \times 5}$ 以及 $F_{5 \times 5}$ 与 $F_{7 \times 7}$ 进行二次加权特征融合, 并将二次加权融合结果进行像素相加, 再与 C_{AFF1}, C_{AFF3}, C_0 进行通道拼接, 其特征处理过程表示为:

$$\begin{aligned} & Output_{MRFA_d} = \\ & Conv(Concat(C_{AFF1}, C_{AFF3}, C_{(AFF2 \oplus AFF4)}, C_0)), \end{aligned} \quad (13)$$

其中: $C_{(AFF2 \oplus AFF4)}$ 表示将 AFF_2 与 AFF_4 像素相加后特征对应通道。

将多感受野特征自适应融合模块嵌入 FT-GDNet 中, 增加模型局部感受野的同时突出有效特征信息, 进一步提升模型表征能力。

2.4 目标定位损失

YOLOv5 网络中使用 CIoU_Loss 定位损失函数, 其定义如下:

$$\begin{cases} CIoU_Loss = 1 - IoU + \frac{\rho^2(c_p, c_g)}{d^2} + \alpha v \\ \alpha = \frac{v}{(1 - IoU) + v} \\ v = \frac{4}{\pi^2} \left(\arctan \frac{w_g}{h_g} - \arctan \frac{w_p}{h_p} \right)^2 \end{cases}, \quad (14)$$

其中: $\rho^2(c_p, c_g)$ 表示真实框与预测框的中心点距离, d 真实框与预测框最小外接矩形的对角线长度, w_g 和 h_g 是真实框的宽度和高度, w_p 和 h_p 是预测框的宽度和高度。

如图 5 所示, 网络在回归定位过程中真实框

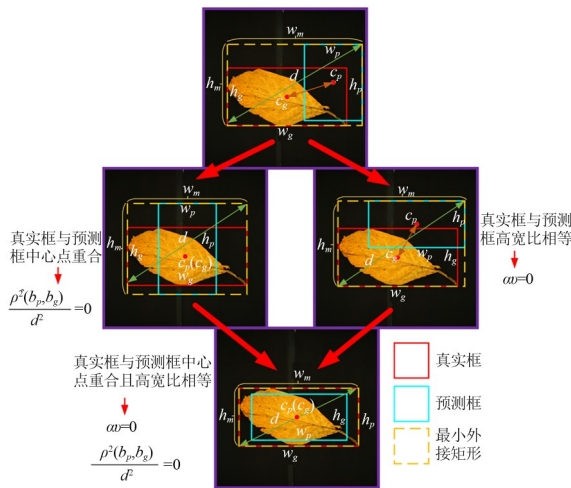


图 5 真实框与预测框拟合结果

Fig. 5 Fitting result of real box and prediction box

与预测框宽高比相等 (即 $\frac{w_g}{h_g} = \frac{w_p}{h_p}$) 且其中心点重合时, CIoU_Loss 中的 $\frac{\rho^2(c_p, c_g)}{d^2}$ 以及 αv 将失效, CIoU_Loss 退化为 IoU_Loss, 但此时真实框与预测框并未重合, 这降低了模型拟合速度且不利于模型检测精度的提升。针对上述问题, 本文提出 MCIOU_Loss, MCIOU_Loss 定义如下:

$$\begin{cases} MCIOU_Loss = 1 - IoU + \frac{\rho^2(c_p, c_g)}{d^2} + (\alpha v)^\lambda + \beta^\lambda \\ \alpha = \frac{v}{(1 - IoU) + v} \\ v = \frac{4}{\pi^2} \left(\arctan \frac{w_g}{h_g} - \arctan \frac{w_p}{h_p} \right)^2 \\ \beta = \left(\frac{1}{\exp\left(\frac{w_g h_g}{w_m h_m}\right)} - \frac{1}{\exp\left(\frac{w_p h_p}{w_m h_m}\right)} \right)^2 \\ \lambda = \frac{1}{1 - IoU} \end{cases}, \quad (15)$$

其中: β 为真实框与预测框面积损失, w_m 和 h_m 分别为真实框与预测框最小外接矩形面积的宽度和高度, 将 αv 和 β 定义为真实框与预测框的矩形相似度判别项, λ 为矩形相似度衰减系数。

本文在 MCIOU_Loss 中引入了真实框与预测框的面积损失, 有效解决真实框与预测框宽高比相等且其中心点重合时 CIoU_Loss 性能退化问题, 同时引入矩形相似度衰减系数 λ , 随着预测框与真实的框重合度增加, λ 增大, 真实框与预测框的矩形相似度判别项数值不断衰减, 在训练过程中对定位损失函数进行动态调整, 加快模型拟合速度同时进一步提高模型定位精度。

3 实验结果与分析

3.1 实验数据集、实验配置及评价指标

本文使用智能分级设备 5XYZ-9C 采集初烤烟叶图像, 如图 6 所示, 该设备使用黑色传送带将叶片运送至装有工业面阵相机和固定光源的暗室中进行图像采集, 工业面阵相机固定于暗室顶部, 其镜头距离传送带 845 mm。相机型号为 CA050-11U, 其分辨率为 $2\,384 \times 1\,528$, 镜头型号为 M0824-MPW2。光源设备型号 XC-BK-650-

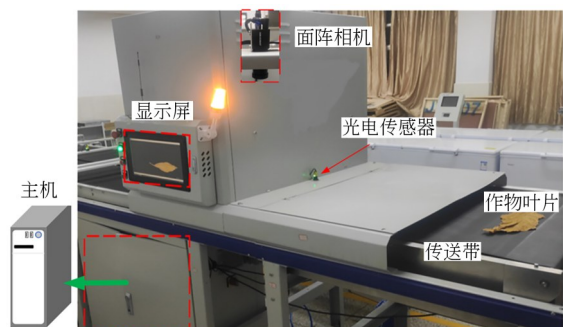


图6 智能分级设备5XYZ-9C

Fig. 6 Intelligent grading equipment 5XYZ-9C

1100,固定于暗室左右两侧顶部。利用5XYZ-9C获取代号为B1F, B2F, B3F, C2F, C3F, C4F, X2F, X3F, V, GY十个等级初烤烟叶图像数据,其中,等级代号相邻的叶片间叶形及颜色特征差异较小,相似度较高。通过数据筛选,得到高质量图像共计3 192幅,建立初烤烟叶分级数据集(Flue-cured Tobacco Leaf Grading Dataset, FTLGD),各等级初烤烟叶示例如图7所示。使用LabelImg对FTLGD中叶片进行锚框标注,得到包含叶片中心坐标、宽、高信息的xml文件,使用python编程将xml文件转换为txt标注文件,建立可以在YOLO模型中运行的数据集,并将其中2 700幅按照4:1比例随机划分训练集和验证集、492幅作为测试集用于网络训练和性能测试。

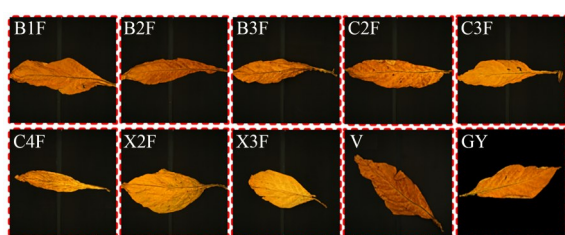


图7 FTLGD中初烤烟叶叶片

Fig. 7 Flue-cured tobacco leaves in FTLGD

实验配置:11th Gen Intel(R) Core(TM) i5-11400处理器,运行内存为16G,图形处理单元为NVIDIA GeForce RTX 3060(12G),深度学习框架为Pytorch1.8.0,使用CUDA11.1, cuDNN 8.0.4加快网络训练。Batch_sizes设置为16, Epoch设置为300。在训练过程中,通过Mosaic算法对输入数据进行在线增强,采用SGD对训练过程进行优化。

评价指标:采用模型参数量(Parameters)、浮点运算数(FLOPs)以及前向推理时间(Inference time)对模型的计算效率进行评价。采用均值平均精度(mean Average Precision, mAP)对模型检测精度进行评价,其计算公式如式(16)所示:

$$mAP = \frac{\sum_{i=1}^M AP_i}{M}, \quad (16)$$

其中: M 为类别数, AP_i 为第 i 类的识别精度。

3.2 实验结果与分析

3.2.1 FTGDNet消融实验

FTGDNet使用CSPNet作为特征提取主干网络,为增强主干的特征提取能力,使用GhostNet进行辅助特征提取。为验证上述改进的效果,本节对ShuffleNet^[21], GhostNet, CSPNet作为特征提取主干网络以及CSPNet作为特征提取主干的同时Shufflenet或GhostNet作为辅助特征提取网络进行实验对比,对比结果如表1所示。ShuffleNet和GhostNet具有模型轻量化的优点,但其特征提取能力有限。将ShuffleNet和GhostNet作为特征提取主干网络时,相较于CSPNet,浮点运算数分别下降6.9 G, 5.2 G, 参数量分别下降2.1 M, 1.4 M, 推理时间分别加快1.0 ms, 0.7 ms,但初烤烟叶等级检测精度大幅下降,在测试集上仅为56.0%和59.5%,在验证集上仅为51.8%和55.5%。将CSPNet作为特征提取主干网络,ShuffleNet和GhostNet分别作为辅助特征提取网络时,由于增加了一条辅助特征提取网络,模型的浮点运算数、参数量、推理时间均有不同程度的上升,但GhostNet作为辅助特征提取网络时,在验证集和测试集上,检测精度分别达到83.5%和81.2%,相较于单一CSPNet作为特征提取主干网络,上升4.7%和5.5%。

本文以CSPNet+GhostNet作为Baseline,以验证EVCB模块、MRFA_d模块、MCIoU_Loss定位损失函数的检测性能,各模块性能消融实验如表2所示。由表2可知,模型在采用EVCB模块后,在验证集上检测精度上升3.1%,在测试集上检测精度上升1.0%,验证了EVCB模块将全局与局部特征信息结合进而有效提高初烤烟叶等级检测精度的结论,但由于EVCB中存在全连接层,模型的浮点运算数、参数量分别上升2.1 G, 2.8 M,推理时间减慢1.6 ms。对MRFA_d模块进行

实验分析,在 Baseline+EVCB 的基础上加入 MRFA_d,在验证集和测试集上,模型检测精度分别达到 89.2% 和 85.3%,模型浮点运算数上升到 22.9 G,参数量上升到 12.4 M,模型推理时间减慢至 12.6 ms。对 MCIoU_Loss 定位损失函数进

行实验分析,在模型中使用 MCIoU_Loss 定位损失函数后,模型的验证精度上升 0.8%,达到 90%,测试精度上升 2.1%,达到 87.4%,证明 MCIoU_Loss 定位损失函数具有提高定位精度进而提高检测精度的能力。

表 1 主干网络对比实验

Tab. 1 Backbone network comparison study

Module	FLOPs/G	Parameters/M	mAP/%		Inference time/ms
			Val	Test	
ShuffleNet	9.5	5.0	56.0	51.8	6.0
GhostNet	11.2	5.7	59.5	55.5	6.3
CSPNet	16.4	7.1	78.8	75.7	7.0
CSPNet+ShuffleNet	18.1	7.7	77.3	77.0	9.0
CSPNet+GhostNet	19.4	8.0	83.5	81.2	9.3

表 2 模块消融实验

Tab. 2 Module ablation study

Module	EVCB	MRFA_d	MCIoU_Loss	FLOPs/G	Parameters/M	mAP/%		Inference time/ms
						Val	Test	
Baseline	×	×	×	19.4	8.0	83.5	81.2	9.3
Baseline+EVCB	✓	×	×	21.5	10.8	86.6	82.2	10.9
Baseline+EVCB+MRFA_d	✓	✓	×	22.9	12.4	89.2	85.3	12.6
Baseline+EVCB+MRFA_d+MCIoU_Loss	✓	✓	✓	22.9	12.4	90.0	87.4	12.6

3.2.2 显式视觉中心瓶颈层对比实验

本文将显式视觉中心瓶颈模块(EVCB)作为网络的瓶颈层(Stage 9)来提取初烤烟叶特征信息,为验证 EVCB 作为网络瓶颈层的性能,本节选用 YOLOv5 网络中作为瓶颈层结构的 Bottleneck, BottleneckCSP, C3 与 EVCB 进行对比,对比结果如表 3 所示。由表 3 可知,FTGDNNet 采用 Bottleneck 作为瓶颈层时,模型整体性能较差,对初烤烟叶等级的识别准确率最低,验证精度仅为 81.5%,测试精度仅为 81.4%。C3 模块作为

瓶颈层时,其浮点运算数、参数量均为最低,分别为 19.4 G, 8.0 M, 此外,其推理时间最快,仅为 9.3 ms。当 FTGDNNet 选择 EVCB 模块作为瓶颈层时,对初烤烟叶的识别精度最高,验证精度达到 86.6%,相较于 Bottleneck, BottleneckCSP, C3 模块分别提升 5.1%, 4.0%, 3.1%, 测试精度达到 82.2%,相较于 Bottleneck, BottleneckCSP, C3 模块分别提升 0.8%, 1.6%, 1.0%, 但由于 EVCB 中使用了全连接层结构,其浮点运算数和参数量均为最高,模型推理时间最慢,为 10.9 ms。

表 3 瓶颈模块效果对比

Tab. 3 Effect comparison of bottleneck module

Module	FLOPs/G	Parameters/M	mAP/%		Inference time/ms
			Val	Test	
Bottleneck	19.5	8.2	81.5	81.4	9.4
BottleneckCSP	19.4	8.1	82.6	80.6	9.4
C3	19.4	8.0	83.5	81.2	9.3
EVCB	21.5	10.8	86.6	82.2	10.9

3. 2. 3 多感受野特征自适应融合模块实验

本节在 Baseline+EVCB 框架下,验证多感受野特征自适应融合模块在初烤烟叶等级检测任务中的优越性能。本文设计了四种多感受野自适应特征融合结构 MRFA_a、MRFA_b、MRFA_c、MRFA_d,并将其嵌入网络的第 10 层(Stege 10),四种多感受野自适应特征融合结构性能对比如表 4 所示。根据表 4 可知,将 4 种多感受野特征自适应融合模块先后嵌入网络后,相较于 Baseline+EVCB,模型的验证精度分别上升 0.5%,2.0%,2.2%,2.6%,达到 87.1%,

88.6%,88.8%,89.2%,测试精度分别上升 2.2%,1.5%,2.8%,3.1%,达到 84.4%,83.7%,85.0%,85.3%。但多感受野特征自适应融合模块中使用了多个卷积层,模型的浮点运算数分别上升 1.3 G,1.3 G,1.3 G,1.4 G,参数量均上升 1.8 M,推理时间分别减慢 1.2 ms,1.4 ms,1.4 ms,1.7 ms。综合考虑下,MRFA_d 对初烤烟叶等级的检测精度最高,验证精度和测试精度分别达到 89.2%和 85.3%,更适合应用于初烤烟叶等级检测任务中。

表 4 四种 MRFA 结构性能对比
Tab. 4 Performance comparison of four MRFA structures

Module	FLOPs/G	Parameters/M	mAP/%		Inference time/ms
			Val	Test	
MRFA_a	22.8	12.4	87.1	84.4	12.1
MRFA_b	22.8	12.4	88.6	83.7	12.3
MRFA_c	22.8	12.4	88.8	85.0	12.3
MRFA_d	22.9	12.4	89.2	85.3	12.6

3. 2. 4 定位损失函数对比实验

为验证 MCIoU_Loss 定位损失函数的性能,本节在 FTGNet 最终模型的基础上,选用 GIoU_Loss,DIoU_Loss,SIoU_Loss,CIoU_Loss 四种定位损失函数与 MCIoU_Loss 进行实验对比,对比结果如表 5 所示。根据表 5,模型采用 GIoU_Loss 的检测精度最低,验证精度仅为 85.9%,测试精度仅为 82.4%;DIoU_Loss 在 GIoU_Loss 的基础上加入中心距损失,较 GIoU_Loss 其检测精度上升 1.8%和 2.4%;

SIoU_Loss 同时使用了角度损失、中心距损失、形状损失,模型采用 SIoU_Loss 定位损失函数后,测试精度和验证精度上升至 88.2%和 85.3%;CIoU_Loss 相较 DIoU_Loss,CIoU_Loss 引入了宽高比损失,进一步提高了目标定位精度,其检测精度达到了 89.2%和 85.3%;MC-IoU_Loss 中引入真实框与预测框面积损失,同时引入矩形相似度衰减系数对真实框与预测框相似度判别项进行动态调整,模型使用 MC-IoU_Loss 后,在验证集上,检测精度达到 90.0%,相较于 GIoU_Loss,DIoU_Loss,SIoU_Loss,CIoU_Loss 分别提升 4.1%,2.3%,1.8%,0.8%,在测试集上,检测精度达到 87.4%,相较于 GIoU_Loss,DIoU_Loss,SIoU_Loss,CIoU_Loss 分别提升 5.0%,2.6%,2.1%,2.1%。图 8 对比了训练过程中五种定位损失函数的数值变化,其中,本文提出的 MC-IoU_Loss 在 40 个 Epoch 后保持最低,图 9 对比了训练过程中 CIoU_Loss 与 MCIoU_Loss 的 mAP 值,进一步验证了 MCIoU_Loss 具有提高定位精度和加快模型拟合的能力。

表 5 不同损失函数性能对比
Tab. 5 Performance comparison of different Loss functions

Loss	mAP/%	
	Val	Test
GIoU_Loss	85.9	82.4
DIoU_Loss	87.7	84.8
SIoU_Loss	88.2	85.3
CIoU_Loss	89.2	85.3
MCIoU_Loss	90.0	87.4

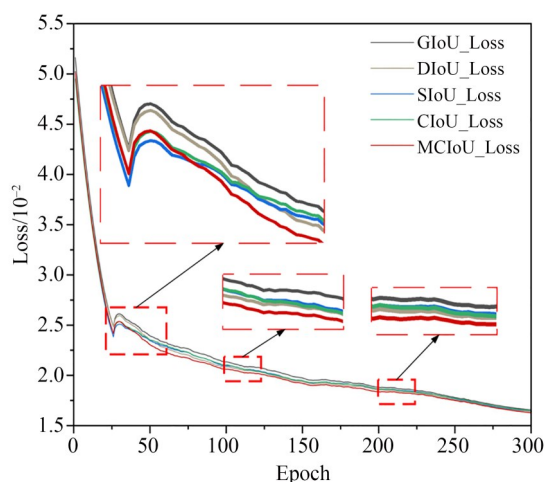


图 8 训练过程损失值变化曲线

Fig. 8 Change curves of Loss value in training process

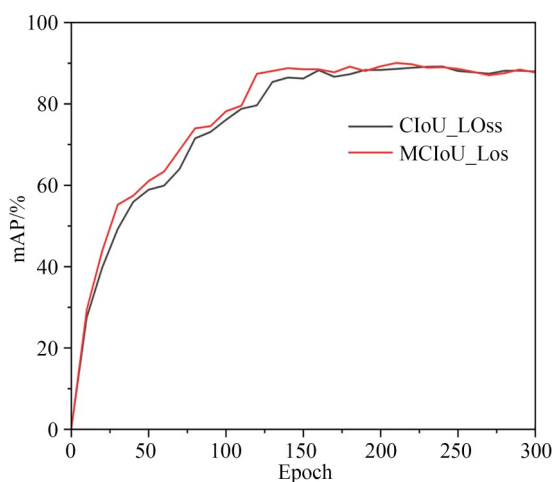


图 9 训练过程 mAP 值变化曲线

Fig. 9 Change curves of mAP value in training process

3.2.5 不同网络对比实验

为客观评价本文算法对初烤烟叶等级的检测性能,选用 Faster R-CNN^[22], Double_head R-CNN^[23], Dynamic R-CNN^[24], SOBL^[25], TOOD^[26], Sparse R-CNN^[27], YOLO 系列^[28-33] 共 20 种检测算法,在实验设备、数据集结构不变条件下与本文方法进行对比实验,实验结果如表 6 所示。

可得到如下结论:

(1)浮点运算数对比,本文模型的浮点运算数为 22.9G,与 TOOD, Faster R-CNN, Sparse R-CNN, Double_head R-CNN, Dynamic R-CNN,

SOBL, YOLOv3, YOLOr_CSP, YOLOr_CSPx, YOLOv5m, YOLOv6m, YOLOv7 相比, FTGDNet 存在较大优势。

(2)参数量对比, FTGDNet 参数量为 12.4M, 分别为 TOOD, Faster R-CNN, Sparse R-CNN, Double_head R-CNN, Dynamic R-CNN, SOBL, YOLOv3, YOLOr_CSP, YOLOr_CSPx, YOLOv5m, YOLOv6m, YOLOv7 的 39.0%, 30.1%, 11.7%, 26.5%, 30.1%, 17.4%, 20.2%, 13.4%, 12.9%, 59.3%, 43.8%, 34.0%。

(3)检测精度对比, FTGDNet 的验证精度达到 90.0%, 测试精度达到 87.4%, 均高于所对比的 20 种先进检测网络。

(4)推理时间对比, 本文模型的推理时间仅为 12.6 ms, 较 YOLOv3-Tiny, YOLOv5n, YOLOv5s, YOLOv6n, YOLOv7-Tiny 轻量化模型慢, 但仍具有较高的实时检测性。

综上分析, FTGDNet 在初烤烟叶等级检测任务有着更为优秀的性能。

3.3 检测结果与定性分析

由上述对比实验可知, 在初烤烟叶分级数据集 FTLGD 上, 本文所提方法有着更为优越的检测性能。图 10 对比了不同网络对初烤烟叶分级的测试结果, 可以看出, YOLOv3-Tiny, YOLOv3, YOLOv5s, YOLOv6s 均存在不同程度的检测缺失问题, 预测框无法覆盖叶片区域, YOLOv7-Tiny, YOLOv7 存在错检或漏检问题。与原始 YOLOv5s 网络相比, 本文模型在精度和泛化能力均有较大提升, 引入 MCIoU_Loss 定位损失函数后, 模型的定位精度有一定程度的提升。

为加深对 FTGDNet 理解, 本文对 FTGDNet 和 YOLOv5 特征提取过程进行可视化, 如图 11 所示, 可以看出, Stage 0 到 Stage 6 随着网络深度的加深, 网络提取的特征从叶片的轮廓结构信息逐渐过渡到深层语义信息, CSPNet 和 GhostNet 共同提取叶片特征信息, 在 Stage 9 到 Stage 20 由深层语义信息指导浅层轮廓信息, 使网络获取更完整的目标轮廓和位置信息, 在 Stage9-Stage17 阶段, 随着网络深度的加深, FTGDNet 明显较 YOLOv5 保留了更多的有效的深层特征信息。

表 6 模型效果对比
Tab. 6 Comparison of model effects

Module	FLOPs/G	Parameters/M	mAP/%		Inference time/ms
			Val	Test	
TOOD	181.0	31.8	69.2	68.5	63.3
Faster R-CNN	206.7	41.2	72.1	71.3	62.5
Sparse R-CNN	150.0	106.0	77.4	74.7	58.5
Double_head R-CNN	480.9	46.8	79.1	78.2	122.0
Dynamic R-CNN	206.7	41.2	82.6	80.9	64.9
SOBL	413.3	71.3	86.2	85.7	204.1
YOLOx-Tiny	19.0	5.4	62.3	61.2	13.1
YOLOx	33.3	8.9	81.7	78.2	24.6
YOLOv3-Tiny	12.9	8.6	50.8	49.1	3.4
YOLOv3	154.9	61.5	82.4	79.7	27.7
YOLOr_CSP	119.2	92.5	74.7	68.2	13.4
YOLOr_CSPx	225.3	96.4	74.5	67.1	25.4
YOLOv5n	4.2	1.8	77.7	73.1	5.4
YOLOv5s	16.4	7.1	78.8	75.7	7.0
YOLOv5m	48.1	20.9	84.1	83.0	13.0
YOLOv6n	11.4	4.7	77.9	77.0	3.6
YOLOv6s	30.5	13.1	88.0	85.7	12.7
YOLOv6m	72.2	28.3	89.5	85.8	15.7
YOLOv7-Tiny	13.2	6.0	86.9	83.9	4.7
YOLOv7	103.5	36.5	88.2	84.2	12.8
FTGDNet	22.9	12.4	90.0	87.4	12.6

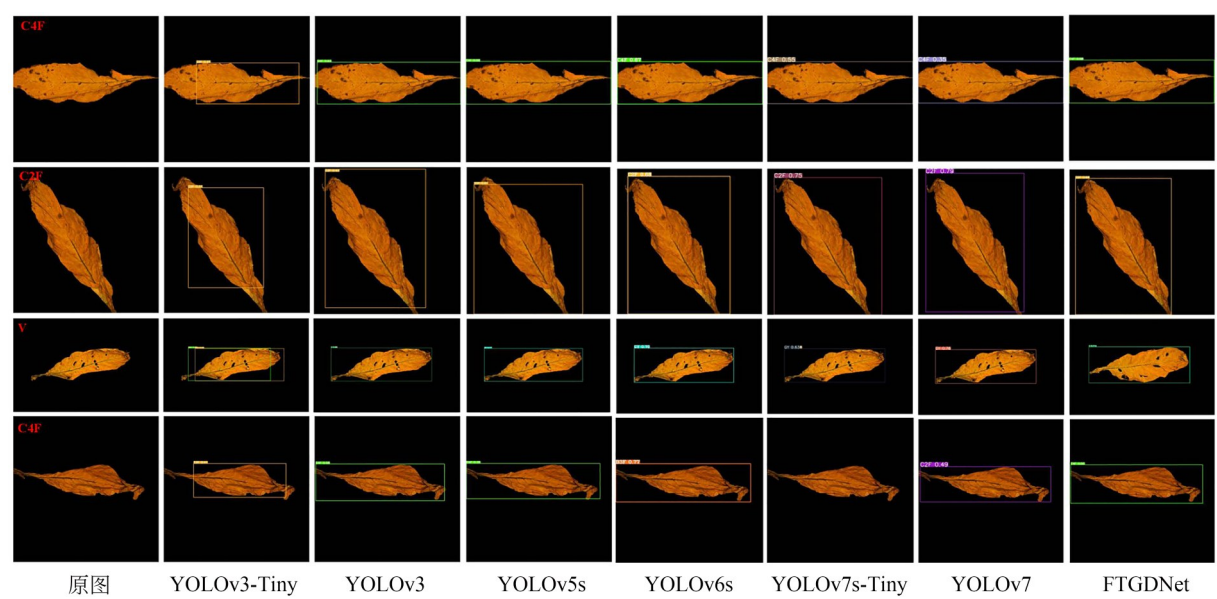


图 10 网络检测结果可视化
Fig. 10 Test result visualization

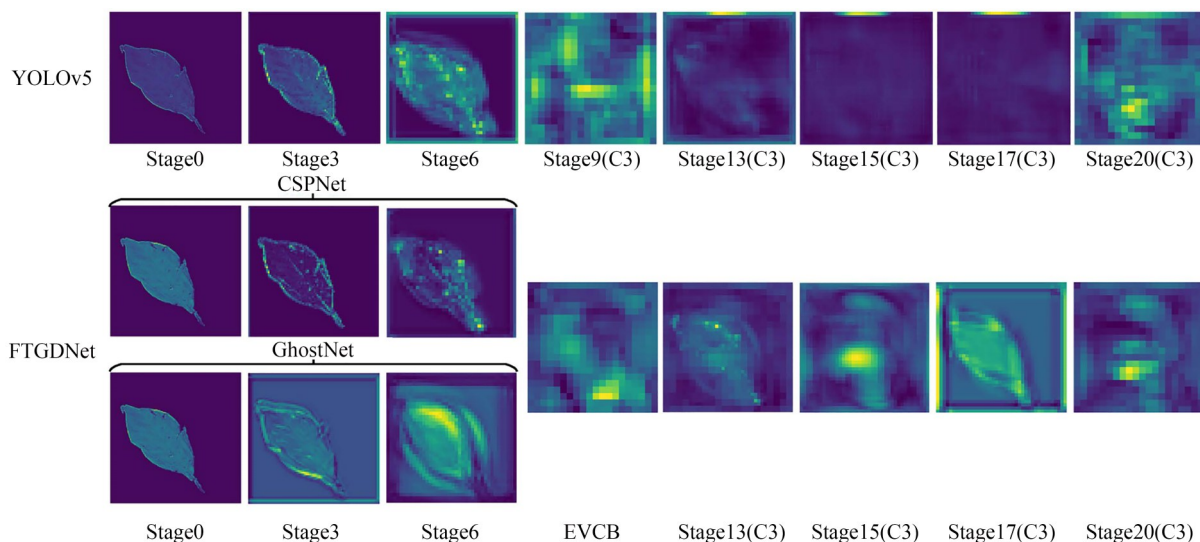


图 11 FTGDNet和YOLOv5特征提取过程可视化

Fig. 11 Visualization of FTGDNet and YOLOv5 feature extraction process

4 结 论

针对相似度较高但等级不同的初烤烟叶难以区分问题,本文提出了多感受野特征自适应融合及动态损失调整的初烤烟叶等级检测算法FTGDNet,算法采用CSPNet和GhostNet共同提取叶片特征信息以达到增强模型特征提取能力的目的;嵌入显式视觉中心瓶颈模块将全局特征与局部特征相融合;通过多感受野特征自适应融合模块将不同感受野的特征图进行自适应加权融合,增强模型的局部感受野的同时突出有效

通道信息;使用MCIoU_Loss定位损失函数解决模型在回归定位过程中真实框与预测框宽高比相等且中心点重合时CIoU_Loss定位性能退化问题。FTGDNet对FTGD中十个等级初烤烟叶的验证精度达到90%,测试精度达到87.4%,高于TOOD, Faster R-CNN, Sparse R-CNN, Double_head R-CNN, Dynamic R-CNN, SOBL, YOLOv3, YOLOx, YOLOr, YOLOv5, YOLOv6, YOLOv7等主流目标检测网络,同时,FTGDNet的推理时间分别仅为12.6 ms,具有较高的实时检测性能。

参考文献:

- [1] 国家统计局. 中国统计年鉴[M]. 北京:中国统计出版社, 2021.
National Bureau of Statistics. *China Statistical Yearbook* [M]. Beijing: China Statistics Press, 2021. (in Chinese)
- [2] 苏明秋. 基于烟叶图像的参数精确提取和分级识别系统研究与实现[D]. 成都:电子科技大学, 2020.
SU M Q. *Research and Implementation of Parameter Extraction and Classification Recognition System Based on Tobacco Image* [D]. Chengdu: University of Electronic Science and Technology of China, 2020. (in Chinese)
- [3] ZHANG J, SOKHANSANJ S, WU S, et al. A trainable grading system for tobacco leaves [J]. *Computers and Electronics in Agriculture*, 1997, 16 (3): 231-244.
- [4] TATTERSFIELD G, FORBES K. Classification of tobacco leaves by color and plant position [C]. *Proceedings of the Ninth Annual Symposium of the Pattern Recognition Association of South Africa, Cape Town*. 1998, 11.
- [5] HAN L Q. Recognition of the Part of Growth of flue-cured tobacco leaves based on support vector machine [C]. 2008 7th World Congress on Intelligent Control and Automation. Chongqing, China. IEEE, 2008: 3624-3627.
- [6] MARCELO M C A, SOARES F F, ARDILA J A, et al. Fast inline tobacco classification by near-in-

- frared hyperspectral imaging and support vector machine-discriminant analysis [J]. *Analytical Methods*, 2019, 11(14): 1966-1975.
- [7] 林相泽, 张俊媛, 徐啸, 等. 基于字典学习与SSD的不完整昆虫图像稻飞虱识别分类[J]. *农业机械学报*, 2021, 52(9): 165-171.
- LIN X Z, ZHANG J Y, XU X, *et al.* Recognition and classification of rice planthopper with incomplete image information based on dictionary learning and SSD[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(9): 165-171. (in Chinese)
- [8] ZHANG X C, WU Z M, CAO C M, *et al.* Design and operation of a deep-learning-based fresh tea-leaf sorting robot[J]. *Computers and Electronics in Agriculture*, 2023, 206: 107664.
- [9] SHANG Y Y, XU X S, JIAO Y T, *et al.* Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments[J]. *Computers and Electronics in Agriculture*, 2023, 207: 107765.
- [10] 曾祥云. 一个基于深度学习的烤烟分级系统的设计与实现[D]. 南京: 东南大学, 2017: 5-7.
- ZENG X Y. *Design and Implementation of a Flue-Cured Tobacco Classification System Based on Deep Learning* [D]. Nanjing: Southeast University, 2017: 5-7. (in Chinese)
- [11] DASARI S K, PRASAD V. A novel and proposed comprehensive methodology using deep convolutional neural networks for flue cured tobacco leaves classification[J]. *International Journal of Information Technology*, 2019, 11(1): 107-117.
- [12] CHEN D D, ZHANG Y H, HE Z F, *et al.* Feature-reinforced dual-encoder aggregation network for flue-cured tobacco grading[J]. *Computers and Electronics in Agriculture*, 2023, 210: 107887.
- [13] LU M Y, WANG C, WU W B, *et al.* Intelligent grading of tobacco leaves using an improved bilinear convolutional neural network [J]. *IEEE Access*, 2023, 11: 68153-68170.
- [14] 乔健, 陈能达, 伍雁雄, 等. 融合注意力机制的金属锅圆柱表面缺陷检测[J]. *光学精密工程*, 2023, 31(3): 404-416.
- QIAO J, CHEN N D, WU Y X, *et al.* Defect detection of cylindrical surface of metal pot combining attention mechanism [J]. *Opt. Precision Eng.*, 2023, 31(3): 404-416. (in Chinese)
- [15] 郝帅, 何田, 马旭, 等. 动态特征优化机制下的跨尺度红外行人检测[J]. *光学精密工程*, 2022, 30(19): 2390-2403.
- HAO S, HE T, MA X, *et al.* Cross-scale infrared pedestrian detection based on dynamic feature optimization mechanism[J]. *Opt. Precision Eng.*, 2022, 30(19): 2390-2403. (in Chinese)
- [16] 郝帅, 张旭, 马旭, 等. 基于CBAM-YOLOv5的煤矿输送带异物检测[J]. *煤炭学报*, 2022, 47(11): 4147-4156.
- HAO S, ZHANG X, MA X, *et al.* Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5[J]. *Journal of China Coal Society*, 2022, 47(11): 4147-4156. (in Chinese)
- [17] WANG C Y, MARK LIAO H Y, WU Y H, *et al.* CSPNet: a new backbone that can enhance learning capability of CNN[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA, USA. IEEE, 2020: 1571-1580.
- [18] HAN K, WANG Y H, TIAN Q, *et al.* GhostNet: more features from cheap operations [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 1577-1586.
- [19] IANDOLA F, MOSKEWICZ M, KARAYEV S, *et al.* DenseNet: Implementing Efficient ConvNet Descriptor Pyramids [EB/OL]. 2014: *arXiv*: 1404.1869. <http://arxiv.org/abs/1404.1869.pdf>
- [20] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-Excitation Networks [EB/OL]. 2017: *arXiv*: 1709.01507. <http://arxiv.org/abs/1709.01507.pdf>
- [21] ZHANG X Y, ZHOU X Y, LIN M X, *et al.* ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 6848-6856.
- [22] 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.
- [23] WU Y, CHEN Y P, YUAN L, *et al.* Rethinking classification and localization for object detection

- [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 10183-10192.
- [24] ZHANG H K, CHANG H, MA B P, *et al.* *Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training*[M]. Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 260-275.
- [25] WANG J Q, ZHANG W W, CAO Y H, *et al.* *Side-Aware Boundary Localization for More Precise Object Detection* [M]. Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 403-419.
- [26] FENG C J, ZHONG Y J, GAO Y, *et al.* TOOD: task-aligned one-stage object detection [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada. IEEE, 2021: 3490-3499.
- [27] SUN P Z, ZHANG R F, JIANG Y, *et al.* Sparse R-CNN: end-to-end object detection with learnable proposals [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 14449-14458.
- [28] REDMON J, FARHADI A. YOLOv3: an Incremental Improvement [EB/OL]. 2018: *arXiv*: 1804.02767. <http://arxiv.org/abs/1804.02767.pdf>
- [29] GE Z, LIU S T, WANG F, *et al.* YOLOX: Exceeding YOLO Series in 2021 [EB/OL]. 2021: *arXiv*: 2107.08430. <http://arxiv.org/abs/2107.08430.pdf>
- [30] KIZILAY E, AYDIN İ. A YOLOR based visual detection of amateur drones [C]. 2022 *International Conference on Decision Aid Sciences and Applications (DASA)*. Chiangrai, Thailand. IEEE, 2022: 1446-1449.
- [31] FANG Y M, GUO X X, CHEN K, *et al.* Accurate and automated detection of surface knots on sawn timbers using YOLO-V5 model [J]. *BioResources*, 2021, 16(3): 5390-5406.
- [32] LI C Y, LI L L, JIANG H L, *et al.* YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications [EB/OL]. 2022: *arXiv*: 2209.02976. <http://arxiv.org/abs/2209.02976.pdf>
- [33] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors [EB/OL]. 2022: *arXiv*: 2207.02696. <http://arxiv.org/abs/2207.02696.pdf>

作者简介:



何自芬(1976—),女,山西阳泉人,博士,教授,硕士生导师,2000年、2005年于西安理工大学分别获得学士和硕士学位,2013年于昆明理工大学获得博士学位,主要从事图像处理和机器视觉等方面的研究。E-mail: zyhzf1998@163.com