文章编号:1671-4598(2024)08-0040-07 DOI:10.16526/j.cnki.11-4762/tp.2024.08.007 中图分类号:TP183 文献标识码:A

融合转置卷积的 YOLOv3 吸烟检测算法

龚英杰，沈希忠
（上海应用技术大学 电气与电子工程学院，上海 201418）

摘要：为预防公共场所因吸烟而引发的安全事故，在 YOLOv3 框架的基础上提出了改进的吸烟检测算法；首先针对传统上采样操作丢失像素信息等问题，设计出一种卷积－转置卷积模块进行替换；在特征融合部分加入坐标注意力机制，使网络更好关注小目标；使用改进的 k-means＋＋优化先验框；最后将 GIoU 替换 IoU 作为算法的损失函数，进一步提高检测精度；此外，构建了一个多场景的抽烟数据集，并对数据集进行数据增强与扩充；实验结果表明，改进后的算法较原算法在 AP@0.5 和 AP@0.5∶0.95 上分别提高 5.58%和 3.34%，FPS 降低 3 左右。

关键词：深度学习；目标检测；小目标；吸烟；转置卷积；注意力机制

YOLOv3 Smoking Detection Algorithm Fused with Transposed Convolution

GONG Yingjie, SHEN Xizhong
(School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract: To prevent safety incidents caused by smoking in public places, an improved smoking detection algorithm based on YOLOv3 framework is proposed. Firstly, to address the missing pixels in traditional upsampling, a convolution module with convolutional transpose for the replacement is designed, In the feature fusion part, the coordinate attention mechanism is added to make the network better focus on small targets. The improved k-means＋＋ is used to optimize the prior box. Finally, the generalized intersection over union (GIoU) is replaced with the intersection over union (IoU), it is taken as the loss function of the algorithm to further improve the detection accuracy. In addition, the multi-scene smoking dataset is constructed to achieve the data augmentation and expansion on the dataset. Experimental results show that compared to the original algorithm, the improved algorithm increases the AP@0.5 and AP@0.5∶0.95 by 5.58% and 3.34%, respectively, and the frames per second (FPS) decreases by about 3 points.

Keywords: deep learning; target detection; small target; smoking; transposed convolution; attention mechanism

0 引言

近年来，深度学习在目标检测中得到了广泛的应用和发展。目前主流的目标检测框架主要分为两类[1]：以 R-CNN、Fast R-CNN、Faster R-CNN 等为代表的双阶段目标检测框架[2]，以 SSD、YOLO 为代表的单阶段目标检测框架[3]。文献［4］首次提出 Faster R-CNN 对吸烟目标进行检测，相对于传统的 Faster R-CNN 算法所提出算法具有较低的误检率、检测时间和 CPU 占用率。近几年对于吸烟行为的检测采用较多的是 YOLO 系列的算法，YOLO 系列算法相比双阶段的检测算法具有实时性强，精度高等优点。文献［5］提出了结合行为先验的 YOLOv3 算法，该算法与单纯的深度学习端到端行为检测方法相比具有识别精度高、检测速度快的特点，能有效解决误检问题。文献［6］通过借鉴 YOLO 算法的单阶段检测思想，设计出一种轻量化的吸烟检测模型，可以较好地检测吸烟行为并在公共数据集上取得不错效果。

在吸烟行为检测中，香烟的外观多呈现为细长型，在图像中所占比例很小且很容易被手部面部遮挡。而传统的 YOLOv3 算法对于小目标检测效果并不是很理想，针对存在的问题本文提出改进的 YOLOv3 算法。设计一种卷积－反卷积模块替代上采样操作，融合于网络动态的学习中不断调整使得得到高分辨率特征图信息更加完全。对输入进特征融合层中的小目标信息可能存在丢失等问题加入（CA，coordinate attention）坐标注意力机制，使得网络更加聚焦物体本身。根据数据集中标注的物体分布特点重新进行 k-means＋＋聚类生成新的 anchors。采用 GIoU 损失函数更好地描述预测框与真实框的位置进一步提高检测精度。

1 YOLOv3 算法介绍

1.1 网络结构

YOLOv3 是由 Redmon 等人[7]于 2018 年提出目标检测算法，因其兼备精度与检测速度而得到广泛应用。YOLOv3 主要由 3 部分组成：主干网络（Backbone）、特征融合层（Neck）、检测头（Head）。具体结构如图 1 所示。主干网络采用了 Darknet-53 结构用来进行特征提取，Darknet-53 主要由 5 个残差模块 Residual Block 构成，每个残差模块都由对应数量的残差单元 res unit 和一个 CBL 模块组成。将输

收稿日期:2023-08-03； 修回日期:2023-09-08。
作者简介:龚英杰(1999-),男,硕士研究生。
通讯作者:沈希忠(1968-),男,博士,教授。
引用格式:龚英杰,沈希忠.融合转置卷积的 YOLOv3 吸烟检测算法[J].计算机测量与控制,2024,32(8):40-46,54.

入分辨率为 416×416 的 RGB 图像通过尺寸为 3×3，步长为 2 的卷积层进行下采样，在第三、第四和第五次下采样输出 3 个大小分别为 52×52×256、26×26×512、13×13×1 024 的特征图。将得到的 3 种特征图送入特征融合层中，YOLOv3 使用了特征金字塔网络（FPN，feature pyramid network）进行特征融合，将各个尺寸特征拼接融合再经过 head 部分进行卷积操作，最后得到 3 种尺寸分别为 13×13×［3×（*number _ classes*＋5）］，26×26×［3×（*number _ classes*＋5）］，52×52×［3×（*number _ classes*＋5）］的预测图。其中 *number _ classes* 表示为分类数目，3 表示每个预测点对应 3 种尺寸的锚框，5 表示为预测置信度和预测框中心点位置 x，y 以及预测框宽高 w，h。输出的 3 种预测图分别负责检测大、中、小 3 种尺度的物体。

1.2　损失函数

YOLOv3 的损失函数 Loss 由定位损失函数 L_{obj}，置信度损失函数 L_{conf}，分类损失函数 L_{cls}构成，如式（1）所示：

$$Loss = L_{\text{obj}} + L_{\text{conf}} + L_{\text{cls}} \tag{1}$$

式中，定位损失函数由中心坐标误差和宽高误差组成，可表示为：

$$L_{\text{obj}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} [(x_i^j - \hat{x}_i^j)^2 + (y_i^j - \hat{y}_i^j)^2] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left[\left(\sqrt{w_i^j} - \sqrt{\hat{w}_i^j} \right)^2 + \left(\sqrt{h_i^j} - \sqrt{\hat{h}_i^j} \right)^2 \right] \tag{2}$$

置信度损失使用交叉熵函数表示为：

$$L_{\text{conf}} = -\sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{noobj}} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{noobj}} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \tag{3}$$

分类损失同样使用交叉熵函数表示为：

$$L_{\text{cls}} = -\sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - P_i^j)] \tag{4}$$

式中，S 为网格尺寸，即 13、26、52。B 为 anchor 个数，即一种网格尺寸对应 3 个 anchor。I_{ij}^{obj}用来表示矩形框是否预测一个目标，负责预测则为 1；反之为 0。I_{ij}^{noobj}用来表示不负责预测则为 1；反之为 0。C_i^j 和 P_i^j 分别表示预测物体概率和类别概率。$\hat{C}_i^j$ 用来表示矩形框是否预测一个真实框，负责预测则为 1；反之为 0。$\hat{P}_i^j$ 用来表示预测目标类别是否为 c 类，是则为 1；反之为 0。λ 为各项损失函数的贡献权重值。本文研究的是单类目标的检测故分类损失函数 L_{cls}为 0，因此损失函数为定位损失加上置信度损失得到 $Loss = L_{\text{obj}} + L_{\text{conf}}$。

2　改进的 YOLOv3 算法

2.1　卷积－转置卷积模块

因小目标在图像中所占比例小、具备的信息有限等特点，所以在较深层的网络中进行多次卷积池化等操作会对小目标的信息造成严重丢失，这使得后续的特征提取难以得到有效信息[8]。而在特征融合部分需要对输入进来的特征图进行上采样操作，由于在此之前的特征图已经在多次步长为 2 的卷积下采样操作中损失信息，因此再将低分辨率图像通过上采样得到高分辨率的图像无疑会放大目标的细节损失。在 YOLO 检测算法中通常采用插值的方式完成上采样操作。例如：最邻插值（Nearest）和双线性插值法（Bilinear）。其中最为常用的为最邻插值法，即将最邻近的像素值赋给新插入像素点的像素值。线性插值的方法优点在于计算量小、计算速度快，缺点在于对目标细节特征无法保留，只是进行简单的数值处理。

减少信息丢失和重构细节特征是增加小目标检测精度的关键。转置卷积是基于深度学习的一种上采样方式，转置卷积可以认为是一种特殊的卷积操作，其计算公式如式（5）所示：

$$Input * C = Output \tag{5}$$

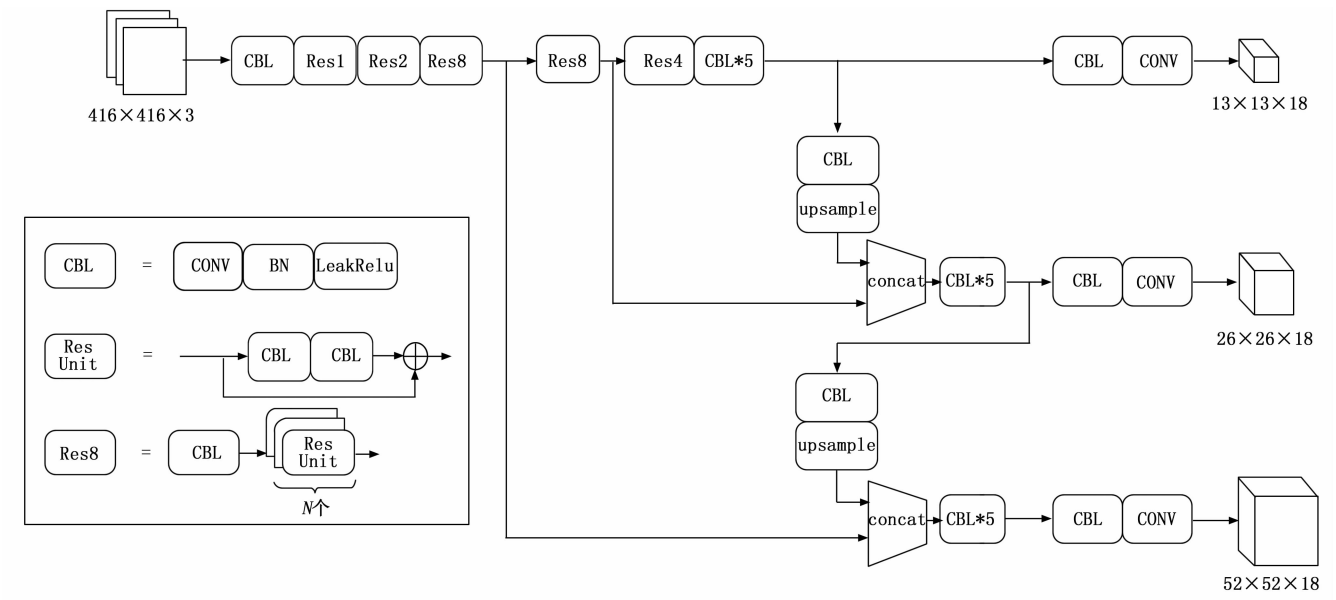

图 1　YOLOv3 结构

$$Output * C^T = Input \tag{5}$$

转置卷积优点在于融合于网络之中这样可以在训练过程中经过反向传播学习来调整参数[9]。相比于传统的插值算法这是一种动态可学习的上采样方式，因此得到的像素值更加合理，从而获得细节更加丰富的高分辨率特征图。本文设计了一种卷积－转置卷积模块（CTC，convolution transposed convolution），该模块由一个 CBL 模块和转置卷积层加 BN 层经 SiLu 激活函数输出组成。其中 CBL 中的卷积大小为 1×1，步长为 1，通过该模块使得输出特征图宽高不变的同时使通道维度降为原来的 1/2，减少了参数量且增加了网络的非线性，提升了网络特征表达能力[10]。转置卷积用于将通道数恢复的同时使宽高变为原来二倍完成上采样操作。考虑到转置卷积在运算过程中会导致一些部分重复堆叠而造成局部出现颜色过重的棋盘效应[11]，当卷积核尺寸可以整除步长时便可以缓解这种效应，减少对特征图信息的破坏。因此 CTC 模块中的转置卷积参数设置为步长为 2，卷积核大小为 4，边缘补充尺寸为 1。转置卷积的输出尺寸计算公式如下：

$$Output = stride * (input - 1) + kernel\_size - 2 * padding \tag{6}$$

添加 BN 层可以减少训练过拟合的风险，加快模型收敛速度。Silu 激活函数具备无上界有下界、平滑、非单调的特性，且在深层的网络中 Silu 的表现要比 ReLu 激活函数更好[11]。使用 Silu 激活函数增强了网络的非线性能力，重分组合了不同通道之间的特征。

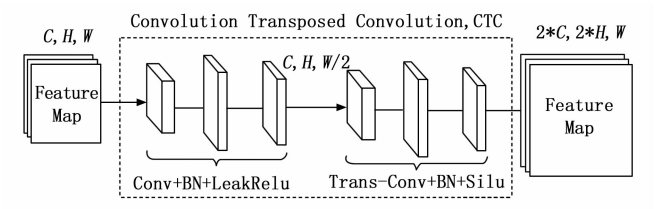

图 2 卷积－转置卷积模块

2.2 坐标注意力机制

人类可以在复杂环境下可以轻松地聚焦到显著区域，受此启发注意力机制被提出。在深度学习中常见的注意力机制有 CBAM（Convolutional Block Attention Module）[12]、ECA（Efficient Channel Attention）[13]、SE（Squeeze-and-Excitation）[14]、CA（Coordinate Attention）[15] 等。其中坐标注意力机制 CA 是一种新颖高效且计算量小的注意力机制，其思想是将位置信息嵌入到通道注意力中，这不仅捕获跨通道的关键信息，还捕获方向感知和位置信息，使得模型能够更为精确地定位和识别目标对象。因此，本文引入 CA 注意力机制，其结构如图 3 所示。

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{7}$$

其缺点在于无法保留确切的位置信息。因此为了促使注意力模块能够捕捉具有精确位置信息的远程空间交互，CA 注意力机制分解了全局池化，转化为一对一维特征编码操作，具体操作为：给定 x 的输入信息，首先使用尺寸为（H，1）或（1，W）的卷积核分别沿着水平坐标和垂直坐标对每个通道进行编码。因此，得到高度为 h 的第 c 通道和宽度为 w 的第 c 通道的输出分别可以表示为：

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leqslant i < W} x_c(h,c) \tag{8}$$

$$Z_c^w(h) = \frac{1}{H} \sum_{0 \leqslant i < H} x_c(j,w) \tag{9}$$

在生成高宽方向的聚合特征映射后，通过拼接操作和大小为 1×1 的卷积变换函数 F 最后通过非线性激活函数。为实现比例为 r 的降维，减少计算开支。具体实现如式（10）：

$$F = \delta[F^{1\times1}(z^h, z^w)] \tag{10}$$

随后将得到的输出 f 沿高宽两位维度拆分为两个独立张量，再分别经过大小为 1×1 的卷积变换和 sigmoid 函数输出。

$$G^h = \sigma[F_h^{1\times1}(f^h)] \tag{11}$$

$$G^w = \sigma[F_w^{1\times1}(f^h)] \tag{12}$$

最后将输入的信息 x 与沿高宽方向编码信息进行加权得到最后输出。

$$Y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{13}$$

FPN 的自顶向下分别对应大、中、小目标的预测，其底层具有图像大量的浅层语义包括小目标的特征和位置信息。采用 CTC 模块替换原网络中的上采样操作，使得小目标减少信息丢失帮助重构细节特征，对自顶向下输入的特征中加入 CA 注意力机制可以有效地使得网络更加关注小目标。得到改进后的特征融合层如图 4 所示。

2.3 锚框的改进

原 YOLOv3 中锚框的尺寸是基于 COCO 数据集通过 K-Means 聚类获得的，虽说具备较好的鲁棒性，但对于小目

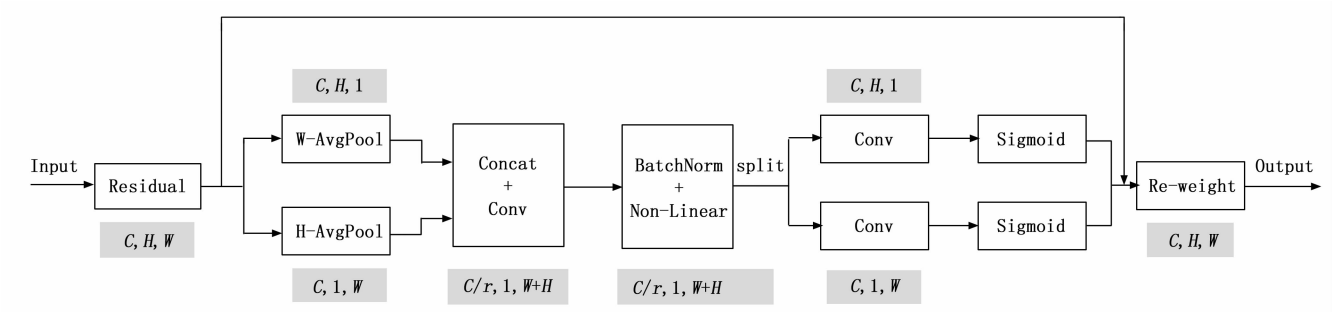

图 3 坐标注意力模块

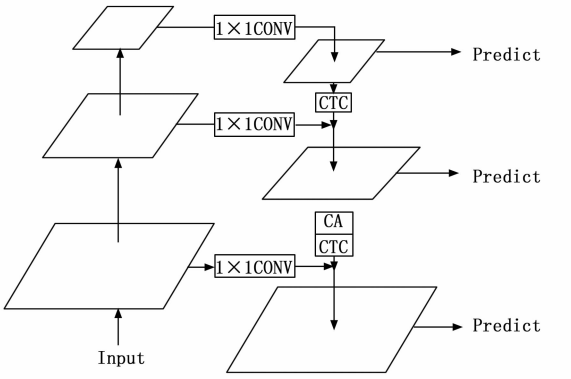

图 4　改进的特征融合层

标居多的吸烟数据集而言并不适合[16]。为了更加符合数据集中吸烟目标的锚框匹配，对聚类算法进行改进。针对 K-Means 聚类会随机选取不同的初始点，可能会造成不同的聚类结果这一问题本文提出使用 K-Means＋＋对锚框进行聚类分析，在衡量距离的指标上采用平均交并比替代标准 K-means＋＋中使用的欧氏距离，如式（14）所示：

$$D(i)=\frac{1}{n}\sum_{i=1}^{n}\frac{Box_i \cap Center}{Box_i \cup Center} \tag{14}$$

改进后的聚类算法步骤如下：

1）从输入的数据点集中随机选择一个真实框作为第一聚类中心。

2）计算数据集中所有真实框与聚类中心的 $D(i)$ 值。

3）遍历每个数据点作为新的聚类中心，选取概率最大的点成为聚类中心。

$$P(i)=\frac{D(i)^2}{\sum_{i\in I}D(i)^2} \tag{15}$$

4）重复进行 2）和 3），直到选出 k 个中心为止。

YOLOv3 共有 3 个检测层，每层需有 3 种锚框尺寸，因此聚类中心个数设为 9。最终重新得到的锚框尺寸如表 1 所示，计算样本中先验框与 k 个聚类中心的平均交并比，得到的结果如图 5 所示。

表 1　改进的锚框尺寸分配

特征图	感受野	锚框尺寸
52×52	小	(18,15)
		(25,46)
		(44,20)
26×26	中	(51,72)
		(86,135)
		(88,37)
13×13	大	(156,76)
		(184,190)
		(320,243)

2.4　损失函数的改进

在 YOLOv3 目标检测算法中，使用 IoU 作为边界损失函数来判断正负样本，并根据此依据来计算置信度损失。

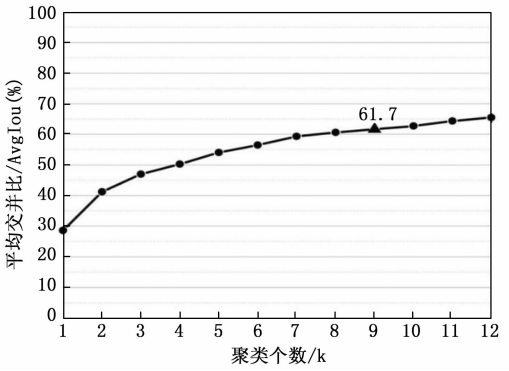

图 5　聚类个数与平均交并比关系图

IoU 是两个区域重叠部分除以两个区域集合部分得出的结果，但是以 IoU 作为损失函数有以下缺点：首先，IoU 并不能准确表达真实框与预测框的位置关系，不同方式的重叠可能会得到相同的 IoU 值。此外，当预测框与真实框没有重叠部分时 IoU 为 0，这将无法反映二者之间的距离，就会导致回传的损失值为 0 无法进行调整。因此引入了 $GIoU$[17] 对损失函数进行优化，$GIoU$ 考虑到了目标的非重叠区域，能够充分反应目标重叠的方式，弥补了 IoU 边界损失函数无法量化真实框与预测框不相交时的不足。

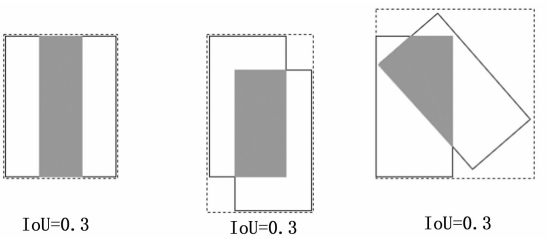

图 6　相同 IoU 下不同位置关系

$GIoU$ 损失函数的可描述为：对于真实框 A 和预测框 B，找到一个可以将它们包围住的最小外接框 C（图 6 中虚线框）。用 C 面积与 $A\cap B$ 面积的比值再比上 C 面积，再用 IoU 减去得到的值即为 $GIoU$ 公式：

$$GIoU=IoU-\frac{\left|\frac{C}{A\cap B}\right|}{|C|} \tag{16}$$

3　实验结果与分析

3.1　数据集的制作

由于目前关于吸烟的公开数据集较少且场景单一，本文所使用的数据集部分来源于 Kaggle 官网的 cigarette smoker 数据集以及通过互联网搜集和公共场合实拍获得。通过筛选最终得到 3 224 张图片，其中包含负样本图片（如进食棍状类物体、模拟抽烟姿势等），图片尺寸在训练时统一缩放为 416×416 大小。将数据集按 7∶2∶1 划分为训练集、验证集、测试集。通过使用开源软件 labelimg 对图像进行标注，如图 7 所示。将图片中标注物体的种类信息和位置信息保存至 XML 文件中，然后通过程序将 XML 格式转换为 YOLO 所需的 TXT 格式进行训练[18]，其中对负样本图片操作为：不标注任何物体，生成不含种类和位置信息的 XML 文件。为了

增强数据的鲁棒性以及减少过拟合，本文采用了在线数据增强的方式对数据集进行扩充。采用的方式为[19]：旋转、镜像、色调饱和度变换以及 mosaic 增强。

图 7 labelimg 标注数据集

3.2 实验环境与参数设置

实验环境与参数设置分别如表 2 和表 3 所示。

表 2 实验环境配置

操作系统	Windows10
CUDA	10.1
Python	3.9.13
pytorch	1.8.1
GPU	2060super
CPU	i-7-12700F

表 3 参数设置

优化器	SGD
Momentum	0.937
Init_lr	0.005
Mini_lr	0.000 5
Weight_decay	0.000 5
Batch_size	16
epoch	200

3.3 评价指标

深度学习中常用 TP、TN、FP、FN 对模型预测结果进行分类。具体定义如下，TP 为被模型预测为正类的正样本，TN 为被模型预测为负类的负样本，FP 为被模型预测为正类的负样本，FN 为被模型预测为负类的正样本。

模型的准确率（P，precision）计算公式为：

$$P = \frac{TP}{TP + FP} \tag{17}$$

模型的召回率（R，recall）计算公式为：

$$R = \frac{TP}{TP + FN} \tag{18}$$

当取不同的置信度时可以得到不同的 P 和 R，由此，得到平均准确率 AP 的计算公式为：

$$AP = \int_0^1 P(R)dR \tag{19}$$

FPS 用来衡量模型检测的速度，其中 t 为检测单张图片所用时长：

$$FPS = \frac{1}{t} \tag{20}$$

$AP@0.5$ 表示 IoU 阈值取 0.5 的情况下 AP 值。$AP@0.5:0.95$ 表示在 IoU 阈值分别取 0.5～0.95，步长为 0.05 情况下的平均 AP 值。

3.4 实验结果

本文进行了锚框尺寸改进前后对比试验、CTC 模块替换上采样实验、在特征融合部分加入注意力机制实验、损失函数对比试验以及不同检测算法的性能对比试验。根据已搭建的实验环境和参数设置，使用自建吸烟数据集中的测试集进行性能测试，且实验中各个模型均采用同样的数据增强策略。

采用 K-means＋＋聚类得到新的锚框尺寸与原锚框尺寸的对比试验性能表现如表 4 所示。

表 4 锚框尺寸替换的对比实验

模型	$AP@0.5$/%	$AP@0.5:0.95$/%	FPS/帧
YOLOv3	82.45	39.21	62
YOLOv3-N	85.79	40.42	62

注：YOLOv3-N 为替换新锚框后的模型。

在替换原 YOLOv3 中锚框尺寸后，$AP@0.5$ 和 $AP@0.5:0.95$ 较使用默认锚框尺寸的 YOLOv3 模型分别提升了 3.34%和 1.21%，在 FPS 表现上基本持平。

在采用新的锚框尺寸基础上，进行最邻上采样（Nearest）与用 CTC 模块替换的上采样对比试验，其性能表现如表 5 所示。

表 5 加入 CTC 模块的对比实验

模型	$AP@0.5$/%	$AP@0.5:0.95$/%	FPS/帧
YOLOv3-N	85.79	40.42	62
YOLOv3-N＋CTC	86.86	42.46	59

使用 CTC 模块替换传统上采样后，模型性能得到明显提高。相比与采用 YOLOv3 网络 $AP@0.5$ 比提高了 1.08%，$AP@0.5:0.95$ 提高了 2.04%。由于 CTC 模块的引入导致计算负担增大使得 FPS 下降 3 左右。

在加入 CTC 模块基础上，选择在 FPN 中最低层特征融合的部分加入不同的注意力机制，对比实验性能表现如表 6 所示。利用 Grad-CAM（Gradient-weighted Class Activation Mapping）[20]将注意力机制效果以热力图形式更好地呈现，如图 8 所示。

表 6 加入不同注意力机制的对比实验

模型	$AP@0.5$/%	$AP@0.5:0.95$/%	FPS/帧
YOLOv3-N＋CTC＋CBAM	85.35	39.08	56
YOLOv3-N＋CTC＋ECA	87.12	40.37	57
YOLOv3-N＋CTC＋CA	87.49	41.48	57

加入 CBAM 注意力机制使得模型的性能有所下降，$AP@0.5$ 和 $AP@0.5:0.95$ 分别下降了 1.51%和 3.38%，

图 8　不同注意力机制的热力图

FPS 下降了 3 左右。加入 ECA 注意力机制使得模型 $AP@0.5$ 提升了 0.26%，$AP@0.5:0.95$ 下降了 2.09%，FPS 下降了 2 左右。加入 *CA* 注意力机制使得模型 $AP@0.5$ 提升了 0.63%，$AP@0.5:0.95$ 下降了 0.98%，FPS 下降了 2 左右。其中 ECA 和 CA 注意力机制参量较 CBAM 更少因此推理速度相对更快。

图 8 中的 4 名吸烟者手上均持有香烟，从热力图中可以看出添加 CBAM 注意力机制加大了关注手部香烟的范围同时也过度关注到许多与香烟无关的地方。添加 ECA 注意力机制使得网络仅集中关注右侧 3 名吸烟者，而忽视了第一名吸烟者。添加 CA 注意力机制使得网络关注到了所有吸烟目标并忽略了无关目标但是关注的范围与真实范围仍然有所偏差。

在添加 CTC 模块和 CA 注意力机制的基础上对替换的损失函数进行对比试验，分别尝试将 SIoU[19]、CIoU、DIoU 和 GIoU 对 IoU 进行替换。对比实验性能表现如表 7 所示。

表 7　加入不同损失函数的对比实验

模型	$AP@0.5$/%	$AP@0.5:0.95$/%	*FPS*/帧
YOLOv3－N＋CTC＋CA＋SIoU	85.98	40.30	56
YOLOv3－N＋CTC＋CA＋CIoU	85.87	42.04	59
YOLOv3－N＋CTC＋CA＋DIoU	86.93	40.38	59
YOLOv3－N＋CTC＋CA＋GIoU	88.03	42.55	59

使用 SIoU 损失函数替换后 $AP@0.5$ 和 $AP@0.5:0.95$ 分别下降了 1.51%和 1.18%，*FPS* 下降了 1 左右。使用 CIoU 损失函数替换后 $AP@0.5$ 下降了 1.62%，$AP@0.5:0.95$ 上升了 0.56%，*FPS* 上升了 2 左右。使用 *DIoU* 损失函数替换后 $AP@0.5$ 提升了 0.56%，$AP@0.5:0.95$ 下降了 1.1%，*FPS* 上升了 2 左右。使用 *GIoU* 损失函数替换后 $AP@0.5$ 提升了 0.54%，$AP@0.5:0.95$ 上升了 1.07%，*FPS* 上升了 2 左右。图 9 为整合所有改进后的模型训练损失曲线，在 *epoch* 迭代到 150 次后损失曲线逐渐平稳，说明实验所取的超参数设置合理模型训练完成。表 8 为各个模块消融实验对比结果。

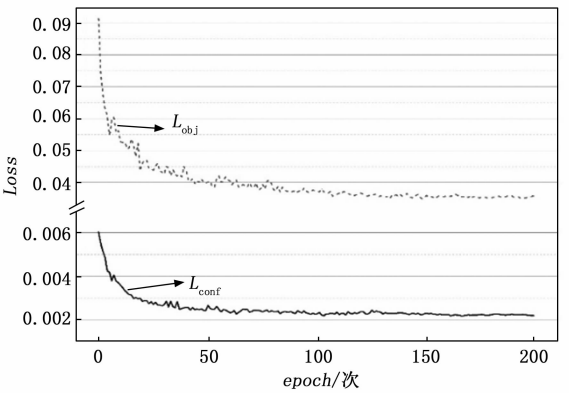

图 9　定位损失与置信度损失变化曲线

表 8　消融实验结果对比

改进锚框	CTC 模块	CA 模块	*GIoU* 损失函数	$AP@0.5$/%	$AP@0.5:0.95$/%	*FPS*/帧
—	—	—	—	82.45	39.21	62
√	—	—	—	85.79	40.42	62
√	√	—	—	86.86	42.46	59
√	√	√	—	87.49	41.48	57
√	√	√	√	88.03	42.55	59

为了更直观地对比 YOLOv3 模型改进前后的检测效果，选取多目标、中小目标以及远距离目标 3 种情况下的吸烟图片进行测试得到结果如图 10～12 所示。从图 10 可以看出，改进后算法较 YOLOv3 算法在第 3 个吸烟目标检测置信度上有略微降低，在其余目标上均有不同程度的提高，且在物体较多背景复杂的情况下未发生误检的情况。从图 11 可以看出，原 YOLOv3 算法忽略了较小的检测目标，只关注到了较大的目标进行检测，改进后的算法可以将两个吸烟目标更好地识别，但二者检测精度均较低其原因可能是所用图片像素低。从图 12 可以看出，原 YOLOv3 算法将

(a) YOLOv3检测效果　　(b) 本文算法检测效果

图 10　多目标识别效果

(a) YOLOv3检测效果　　(b) 本文算法检测效果

图 11　中小目标识别效果

(a) YOLOv3检测效果 (b) 本文算法检测效果

图 12 远距离目标识别效果

人物手指误识别为香烟，而改进后算法较好地关注到了手部香烟，做出了正确的检测且置信度，较原 YOLOv3 提高了 12%。

为验证改进后模型的先进性，与 YOLOv3、SDD、Faster R-CNN 和 YOLOv5 以及近年来一些吸烟检测模型进行对比实验。得到的实验结果如表 9 所示。

表 9 不同算法的对比实验

模型	主干网络	*AP*@0.5/%	*AP*@0.5∶0.95/%	*FPS*/帧
YOLOv3	Darknet53	82.45	39.21	62
Faster R-CNN	Resnet50	74.67	30.28	12
SSD	VGG16	72.34	28.35	79
YOLOv5	CSPDarknet53	83.52	38.02	143
文献[5]	Darknet53	86.46	/	/
文献[6]	文献[6]	86.30	/	103
本文模型	Darknet53	88.03	42.55	59

注：本文模型即为 YOLOv3－N＋CTC＋CA＋GIoU。

通过对不同目标检测算法进行比较发现：Faster R-CNN 和 SDD 在吸烟数据集上表现较差，本文算法相较于 Faster R-CNN 在 *AP*@0.5 和 *AP*@0.5∶0.95 分别提高了 13.36%和 12.27%，*FPS* 值提高了 47 左右。相较于 SSD 在 *AP*@0.5 和 *AP*@0.5∶0.95 分别提高了 15.69%和 14.20%，*FPS* 值下降了 20 左右。YOLO 系列算法在吸烟数据集上表现较好，本文算法相较于原 YOLOv3 算法提高明显，在综合所有改进之后 *AP*@0.5 和 *AP*@0.5∶0.95 分别提高了 5.58%和 3.34%，而 *FPS* 仅损失 3 左右。相较于 YOLOv5s 在 *AP*@0.5 和 *AP*@0.5∶0.95 分别提高了 4.51%和 4.53%，但在 *FPS* 上表现远不如 YOLOv5。与文献 [5] 和文献 [6] 中所提模型相比在精度上具备较大优势，而在 *FPS* 表现上不如文献 [6] 中模型。综上说明改进后的模型在检测精度上得到较为明显的提升，检测速度上也基本满足实时检测的要求，较好地平衡了检测精度与速度。

4 结束语

针对吸烟目标较小，且存在遮挡等问题，本文对 YOLOv3 算法进行了改进。主要贡献如下：设计一种卷积－转置卷积替换上采样模块，该模块虽引入额外参数但性能较传统上采样更优。在 FPN 结构底层加入 CA 注意力机制，使得模型更加关注目标，可视化效果也优于其他经典注意力机制。采用更适合数据集的锚框尺寸以及损失函数。最终改进的模型相比原 YOLOv3 仅损失较少的 FPS 提升了较大的精度。再与当前主流的算法进行对比，表明改进算法仍有较好的综合性能，但是在实际应用中可能仍然存在部署困难和模型泛化性不够等问题。为了进一步深入研究本课题，接下来的工作主要围绕两个方面进行：一是扩充数据集，使得模型能够在更多复杂场景下进行检测；二是设计轻量化网络结构减少模型参数，以便部署使用。

参考文献：

[1] 许德刚，王 露，李 凡. 深度学习的典型目标检测算法研究综述 [J]. 计算机工程与应用，2021，57 (8)：10－25.

[2] LIANG F，ZHOU Y，CHEN X，et al. Review of target detection technology based on deep learning [C] //Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence，2021：132－135.

[3] DUBEY A，FUCHS J，MADHAVAN V，et al. Region based single-stage interference mitigation and target detection [C] //2020 IEEE Radar Conference (RadarConf20). IEEE，2020：1－5.

[4] 韩贵金，李 倩. 基于 Faster R-CNN 的吸烟快速检测算法 [J]. 西安邮电大学学报，2020，25 (2)：85－91.

[5] 徐望明，徐天赐，李传东，等. 基于深度学习与行为先验的吸烟和打电话检测方法 [J]. 计算机应用与软件，2022，39 (4)：199－204.

[6] 陈睿龙，罗 磊，蔡志平，等. 基于深度学习的实时吸烟检测算法 [J]. 计算机科学与探索，2021，15 (2)：327－337.

[7] FARHADI A，REDMON J. Yolov3：An incremental improvement [C] //Computer Vision and Pattern Recognition. Berlin/Heidelberg，Germany：Springer，2018，1804：1－6.

[8] 董 刚，谢维成，黄小龙，等. 深度学习小目标检测算法综述 [J]. 计算机工程与应用，2023，59 (11)：16－27.

[9] 李炳臻，姜文志，顾佼佼，等. 基于 SSD 的小目标特征强化检测算法 [J]. 兵工自动化，2021，40 (2)：32－37.

[10] 牛浩青，欧 鸥，饶姗姗，等. 改进 YOLOv3 的遥感影像小目标检测方法 [J]. 计算机工程与应用，2022，58 (13)：241－248.

[11] APICELLA A，DONNARUMMA F，ISGRò F，et al. A survey on modern trainable activation functions [J]. Neural Networks，2021，138：14－32.

[12] 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.

[13] 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.

(下转第 54 页)