DOI:10.19651/j.cnki.emt.2107174

融合注意力机制的实时行人检测算法*

冯宇平 管玉宇 杨旭睿 刘 宁 王兆辉
(青岛科技大学 自动化与电子工程学院 青岛 266061)

摘 要: 为了提高 Tiny YOLOv3 目标检测算法在行人检测任务中的准确率,对该算法进行了研究改进。首先对 Tiny YOLOv3 的特征提取网络进行深化,增强网络特征提取能力;然后在预测网络的两个检测尺度分别加入通道域注意力机制,对特征图的不同通道赋予不同的权重,引导网络更多关注行人的可视区域;最后,改进激活函数和损失函数并采用 K-means 聚类算法重新选择初始候选框。实验结果表明,改进后 Tiny YOLOv3 算法的准确率在 VOC2007 行人子集上达到 77%,较 Tiny YOLOv3 提高 8.5%,在 INRIA 数据集上达到 92.7%,提高 2.5%,运行速度分别达到每秒 92.6 帧和 31.2 帧。该方法提高了行人的检测精度,保持了较快的检测速度,满足实时性运行需求。

关键词: 行人检测;Tiny YOLOv3;特征提取;通道注意力机制;实时检测
中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.60

Real-time pedestrian detection algorithm fused with attention mechanism

Feng Yuping Guan Yuyu Yang Xurui Liu Ning Wang Zhaohui
(College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: In order to improve the accuracy of the Tiny YOLOv3 target detection algorithm in pedestrian detection tasks, the algorithm is researched and improved. Firstly, deepen the feature extraction network of Tiny YOLOv3 to enhance the feature extraction capabilities of the network. Then, add the channel attention mechanism to the two detection scales of the prediction network, and assign different weights to different channels of the feature map to guide the network to pay more attention the visible area of pedestrians. Finally, the activation function and loss function are improved, and the K-means clustering algorithm is used to reselect the initial candidate frame. Experimental results show that the improved Tiny YOLOv3 algorithm has an average precision(AP) of 77% on the VOC2007 pedestrian subset and 92.7% on the INRIA data set, which is 8.5% and 2.5% higher than Tiny YOLOv3, and the running speed is 92.6 frame per second(FPS) and 31.2 FPS. The algorithm improves the accuracy of pedestrian detection, maintains a faster detection speed, and meets real-time operation requirements.

Keywords: pedestrian detection;Tiny YOLOv3;feature extraction;channel attention mechanism;real-time detection

0 引 言

近年来,行人检测技术在安防监控、智能预警等应用越来越广泛。由于含有行人的图像背景复杂,以及受姿态、穿着和遮挡问题的影响,大大增加了行人检测的难度,而在实际的行人检测系统中,不仅要求较高的准确性,还要求较高的实时性,因此对行人检测的研究具有非常重要的现实意义[1-2]。

传统的行人检测算法通常采用人工特征提取和分类的方法。例如,文献[3]通过提取正负样本的 HOG 特征,并采用支持向量机进行分类器训练,得到分类模型。文献[4]采用 SILTP 纹理特征和梯度方向直方图提取人体不同部位的特征,并通过 GPU 加速实现行人检测。而随着计算机算力的提升,基于卷积神经网络的目标检测算法被陆续提出。目前常用的方法有双阶段检测算法 R-CNN[5-7] 系列和一阶段检测算法 SSD[8]、YOLO[9-11] 系列。双阶段检测算法利用选择搜索或区域候选网络生成候选区域,再进一步对目标的种类和位置进行预测,提高了目标检测的精度。然而,由于候选区域生成和检测网络分开进行,难以实现实时目标检测。一阶段检测算法直接对目标的种类和位置进行回归,具有较快的检测速度。目前有诸多学者对行人检测展开研究。例如,文献[12]针对检测环境复杂和多尺度变化等问题,提出了自适应 RPN-Incep 网络,使行人检测精

收稿日期:2021-07-05
* 基金项目:国家自然科学基金项目(61971253)、青岛科技大学 2021 年大学生创新训练计划项目(S202110426006)资助

度进一步提高。文献[13]在 YOLOv2 的基础上改进特征提取网络，提出了一种基于密集连接和空间金字塔池化结构的 YOLO 目标检测算法，平衡了检测精度和速度。文献[14]利用视觉注意力机制和拉普拉斯金字塔融合的方法确定行人显著图，在 INRIA 数据集上取得了 92.78%检测精度。文献[15]提出了铁路周界入侵全天候综合检测方法，针对白天场景采用深度学习算法，针对夜晚场景采用三帧差法、混合高斯背景建模法等动态方式，提高了不同光照强度下的检测精度。以上方法有效提高了行人检测效果，但并不适合实际场景，对于一些实时性要求较高的场景而言，不仅要求较高的检测精度，还要求较快的检测速度。

YOLOv3 算法利用特征金字塔[16]（feature pyramid network，FPN）和残差网络[17]的结构设计有效提高了检测的精度。但该算法网络结构复杂，模型体积较大，难以在嵌入式设备上达到实时性要求。Tiny YOLOv3 是 YOLOv3 的简化版本，网络结构简单，模型体积小，检测速度较快，但是检测精度较低；同时，Tiny YOLOv3 利用 FPN 的结构设计对两个检测尺度的特征图进行融合，但这种方式仅仅是将不同通道的特征进行串联，不能反映出特征图通道之间的重要程度。针对以上问题，本文对 Tiny YOLOv3 算法进行优化改进。首先，采用 3×3 卷积对主干网络进行加深，增强网络的特征提取能力；接着，采用 1×1 卷积对特征图进行降维，降低模型参数量，并实现跨通道的信息交互；然后，在两个预测网络引入轻量级的通道域注意力机制，利用注意力机制融合不同尺度的信息，对特征图的不同通道赋予不同的权重，引导网络关注行人区域；最后，优化边界框回归损失函数和激活函数并采用 K-means 聚类算法，重新选择初始候选框。实验结果表明，改进后的 Tiny YOLOv3 具有更高的行人检测精度，并取得了较快的检测速度，模型参数少，体积小，适合实时和嵌入式应用。

1　研究方法

1.1　Tiny YOLOv3 算法

YOLO 系列算法是基于卷积神经网络的一阶段目标检测算法。该算法先将图像划分成 $S\times S$ 个网格，每个网格预测 B 个边界框和 C 个类别概率，得到每类目标的预测框和置信度，置信度公式为：

$$Confidence = P(object)\times IOU_{pre}^{truth} \tag{1}$$

其中，$P(object)$ 为网格中目标存在概率，IOU_{pre}^{truth} 为预测框和真实框的交并比。最后通过非极大抑制确定目标的坐标信息，得到最佳的预测框。

Tiny YOLOv3 是在 YOLOv3 基础上的简化版本。相比 YOLOv3 复杂的网络结构，Tiny YOLOv3 将特征提取网络缩减为 7 层卷积和 6 层最大池化（Maxpool），减小了模型尺寸，同时简化了 YOLOv3 的多尺度检测，采用 26×26 和 13×13 两种检测尺度对特征图进行预测输出，网络结构如图 1 所示。

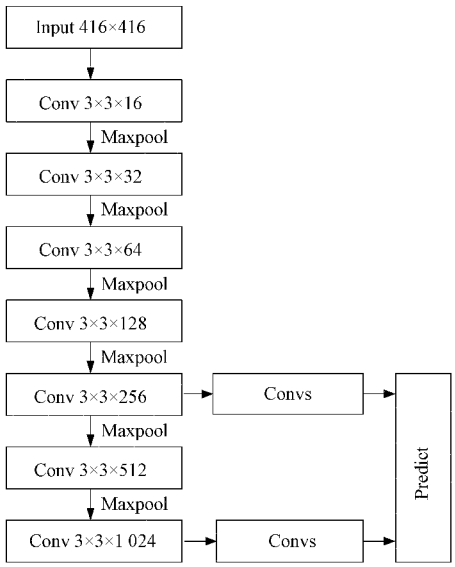


图 1　Tiny YOLOv3 模型结构

每个卷积层由卷积操作（Conv）、BN 归一化和激活函数组合而成。激活函数能够减少训练时的梯度消失现象，BN 归一化可以加快模型的收敛速度和防止过拟合。对于输入图像来说，首先是对图像的尺寸进行调整，之后图像经过卷积层进行特征提取，比如图像经过 3×3×32 的卷积核处理后，会产生尺寸为 $H\times W\times 32$ 的特征图（H，W 分别表示特征图的高和宽），图像经过 2×2 的最大池化层（其中 5 次步长为 2 的最大池化，1 次步长为 1 的最大池化）进行下采样，以输入图像尺寸为 416×416 为例，经过最大池化层的输出特征图尺寸依次为 208×208、104×104、52×52、26×26、13×13、13×13。

1.2　数据集聚类分析

在 Faster-RCNN 算法中提出了锚框机制，但锚框的大小通常是人工设置的，这将导致网络在训练过程中收敛缓慢，容易出现局部拟合。Tiny YOLOv3 利用了 Faster-RCNN 的锚框机制，在数据集上采用了 K-means 聚类的方法来寻找最优的初始预选框。Tiny YOLOv3 的初始预选框是通过对 COCO 数据集的聚类确定的，共有 6 个预选框。由于 COCO 数据集包含的目标种类众多，不同目标的横纵比差异较大，因此原有的预选框尺寸并不适合行人检测。为了提高模型的训练速度和行人检测的准确性，本文在 VOC2007 提取的行人数据集上采用 K-means 聚类算法，确定最优的初始预选框尺寸。传统的 K-means 聚类算法采用欧氏距离函数，但是较大边界框容易产生更大的误差，为此本文采用 IOU（候选框与标记框的交并比）来评价聚类结果。距离计算如式(2)所示。

$$D(box,centriod) = 1 - IOU(box,centriod) \tag{2}$$

其中，box 表示 K-means 聚类的结果，$centriod$ 表示聚类中心。

1.3　特征提取网络

Tiny YOLOv3 的特征提取网络较浅，难以提取深层的

特征,在行人目标检测上精度较低。为此,本文对特征提取网络进行加深,在原网络的基础上增加 4 个卷积核大小为 3×3 的卷积层,增强特征提取能力,提高检测精度。虽然增加卷积层能够提高行人检测的精度,但是随着卷积层的叠加,模型的参数量剧增,大大增加了计算量和内存资源的占用。

在计算量过大的情况下,本文借鉴残差网络的思想,在增加的 3×3 卷积层之前,引入卷积核大小为 1×1 的卷积层,降低通道维度,以减少网络的计算量。具体来讲,首先通过 3×3 卷积将通道数扩张到上一层的 2 倍,提取高维特征;然后通过 1×1 卷积,将通道数压缩为原来的 2 倍,降低通道维度,减少计算量同时实现信息的跨通道交互;最后再通过 3×3 卷积扩张通道,恢复原来的通道维度。改进后的模型结构如图 2 所示,其中左虚线框为改进后的特征提取网络。

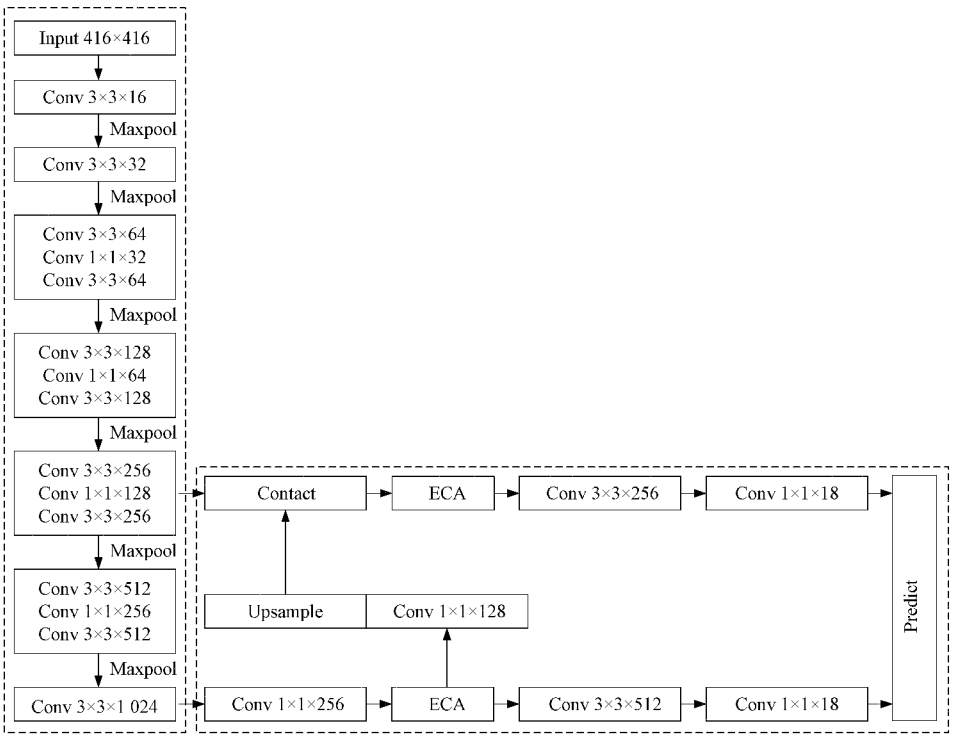


图 2 Tiny YOLOv3 改进模型

1.4 融合通道注意力的预测网络

注意力机制从本质上来说就是对信息资源的重新分配。例如当人们从图像中寻找行人信息时,会更多的把注意力集中在具有行人特征的图像区域,而忽略那些不具有行人特征的图像区域,这便是注意力机制的合理分配。对于卷积神经网络来说,注意力机制能够对图像中的重要信息赋予高权重,对不重要的信息赋予低权重,并且还可以根据不同的场景不断调整权重信息,因此可以提高模型的泛化性和鲁棒性。

在实际的行人检测场景中,背景信息的干扰和遮挡情况的存在,影响网络对行人特征的提取,进而影响行人检测精度。Tiny YOLOv3 的预测网络对两个尺度的特征图进行融合,这种融合方式仅仅在通道维度上对特征进行串联(contact),不能反映出行人特征在某些通道上的重要程度。为此,本文将注意力机制引入到 Tiny YOLOv3 的预测网络当中,利用注意力机制融合不同尺度的信息,对特征通道赋予不同的权重,引导网络关注行人特征,降低干扰信息的影响从而提高检测精度,图 2 中右虚线框为改进后的预测网络。为了使网络自动学习特征通道的权重,本文引入了无降维的轻量级通道域注意力机制 ECA-Net (efficient channel attention networks)[18],如图 3 所示。

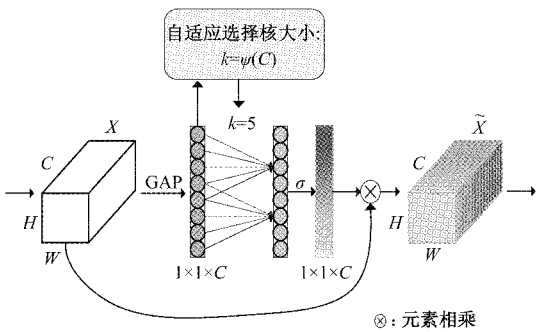


图 3 ECA 模型结构

图 3 中,输入特征图 $X \in R^{H\times W\times C}$,$X$ 有 C 个特征通道。一般来说,卷积神经网络只能学习局部感受野,不能利用该区域以外的上下文信息。为此,通过全局平均池化对全局空间信息进行压缩,即在空间维度 $H\times W$ 上进行压

缩，得到 1×1 的权重信息，全局平均池化[19]公式如下：

$$Y = F_{sq}(X_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_c(i,j) \tag{3}$$

其中，Y 为压缩之后得到的权重，$H \times W$ 为空间维度信息。

为了使网络自动学习不同通道的注意力权重，使用大小为 k 的一维卷积来完成跨通道的局部信息交互。一维卷积的大小由通道维数 C 的函数来自适应确定，计算参数 k 大小的公式为：

$$k = \psi(C) = \left|\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}\right|, \gamma = 2, b = 1 \tag{4}$$

通过大小为 k 的一维卷积并使用 Sigmoid 得到每个通道的权重。公式如下：

$$\omega_c = \sigma(C1D_k(y)) \tag{5}$$

其中，σ 是 Sigmoid 激活函数，ω_c 是生成的通道注意力权重，维度为 1×1×C。然后将注意力权重与输入特征图进行加权，实现对特征图通道的重要性表达，加权公式如下：

$$\widetilde{X}_c = X_c \otimes \omega_c \tag{6}$$

其中，$\otimes$表示逐元素相乘，$\widetilde{X}_c$ 表示通过注意力机制的输出结果。

如图 4 所示，两个检测尺度输出的特征图尺寸分别为 13×13 和 26×26，即输入图像被划分为 13×13 和 26×26 的网格，分别检测远距离和近距离的行人，每个网格与通道一一对应。每个网格预先设置 3 个预选框，在训练时不断调整，选择出最优的预选框作为输出结果。不同的通道代表每个网格的输出参数，以 13×13 的特征图为例，每个通道的参数包含预测框的中心坐标(b_x，b_y)、预测框的长宽(b_w，b_h)、预测框的置信度得分 p_0 以及行人的预测得分 s。每个网格包含 3 个预测框，每个网格包含以上 6 个参数，故输出特征图的通道维数均为 18。本文将 ECA 注意力模块与 Tiny YOLOv3 的预测网络相结合，分别加入到两个检测尺度当中。在输出 13×13 特征图的预测网络加入 ECA 注意力模块，将通过注意力模块后的特征图进行上采样与 26×26 特征图相串联，输出 384 维通道的特征图，再通过 ECA 注意力模块重新分配权重，最终的两个输出层将更多关注行人信息，有效降低了干扰信息和遮挡问题的影响。

1.5　改进损失函数和激活函数

在训练过程中，Tiny YOLOv3 的损失函数可分为 3 部分，分别为边界框回归损失、置信度损失和分类损失，总的 $Loss$ 可用式(7)表示：

$$Loss_{total} = \sum_{i=1}^{2} CoordLoss^i + ConfidenceLoss^i + ClassLoss^i \tag{7}$$

其中，i 表示尺度。行人检测的定位通常依赖于准确

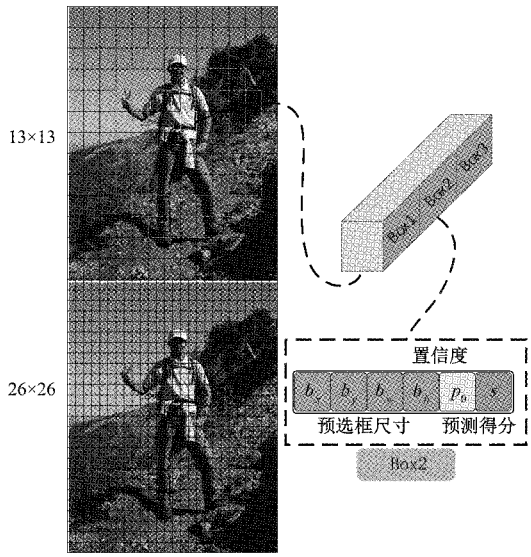


图 4　预测层结构

的边界框回归，为了提高定位的准确性和检测精度，本文对边界框回归损失进行优化改进。

边界框回归损失包括中心坐标损失和宽高损失。在 YOLOv1 算法中，直接对边界框的实际值进行预测，然而预测值的微小变化会扩展到整个图像范围，导致坐标波动大，预测不准确。YOLOv2 对这些问题进行了改进，可以表示为：

$$X = \sigma(b_x) + c_x \tag{8}$$

$$Y = \sigma(b_y) + c_y \tag{9}$$

$$W = p_w e^{b_w} \tag{10}$$

$$H = p_h e^{b_h} \tag{11}$$

其中，(c_x, c_y)是网格的坐标偏移量；(p_w, p_h)是预选框映射到特征图上的宽和高；(b_x, b_y, b_w, b_h)是网络输出的预测值；(X, Y, W, H)是预测的中心坐标和宽高；σ 是 Sigmoid 函数。

由式(8)、(9)可知，预测框的中心点坐标 X 和 Y 是由 Sigmoid 函数激活的，当神经网络的输出较大时，S 型函数的导数变得非常小。此时，利用平方误差得到的误差值非常小，导致网络收敛速度较慢。当输出值为 0 或 1 时，解决上述问题的常用方法是采用交叉熵损失函数，可以表示为：

$$Loss = -\frac{1}{n}\sum_{i-1}^{n}[a_i \times \log(\hat{a}_i) + (1-a_i) \times \log(1-\hat{a}_i)] \tag{12}$$

其中，a_i 为真实值，$\hat{a}_i$ 为经过 Sigmoid 函数后的输出值。当真实值 a_i 只能取 0 或 1 时，交叉熵损失函数满足要求。也就是说，当 a_i 和 $\hat{a}_i$ 等于 0 时，由式(12)可得，损失约为 0。同理，当 a_i 和 $\hat{a}_i$ 等于 1 时，损失也接近于 0。但是，对于预测框中心点坐标 XY，其真实值既不是 0 也不是 1，

而是介于0和1之间的值。例如,当$a_i=\hat{a}_i=0.7$时,交叉熵损失为$-0.7\times\log(0.7)-0.3\times\log(0.3)=0.265$,而不是0。因此,可以对中心点坐标$XY$的损失函数进行优化改进。

为了改进中心坐标的XY损失函数,本文采用了广义交并比(generalized intersection over union,GIOU)[20]作为回归损失。采用GIOU的原因有两个方面:1)当交并比(intersection over union,IOU)在真实框和预测框无交集的情况下,IOU无法进行评估度量;2)IOU无法精确反映真实框和预测框的重合度大小。IOU和GIOU的定义如下:

$$IOU=\frac{|B\cap B_{gt}|}{|B\cup B_{gt}|} \tag{13}$$

$$GIOU=IOU-\frac{|C\backslash(B\cap B_{gt})|}{|C|} \tag{14}$$

其中,B表示预测框,B_{gt}表示真实框,C表示包含真实框和预测框的最小封闭面。

激活函数是卷积神经网络的重要单元,随着网络模型的不断发展,各式的激活函数也被陆续提出,它可以使网络引入非线性因素,能够从输入输出之间生成非线性映射,有利于网络更好的学习。Tiny YOLOv3的特征提取网络采用Leaky ReLU激活函数,本文将其替换为Mish激活函数[21],函数公式如下:

$$Leaky\ ReLU=\begin{cases}x, & x\geqslant 0\\ 0.1x, & x<0\end{cases} \tag{15}$$

$$Mish=x\cdot\tanh(\ln(1+e^x)) \tag{16}$$

对于激活函数来说,无上界是一个理想属性,能够避免导致训练速度急剧下降的梯度饱和。无下界属性也是有利的,它会产生很强的正则化效应,有利于模型训练。Mish的非单调特性使得小的负输入保留为负输出,提高了表达性和梯度流。与Leaky ReLU相比,Mish的一阶导数是连续的,这说明Mish激活函数是连续可微的,而Leaky ReLU并不是连续可微的,这在基于梯度的优化中可能会造成负面影响。函数图像如图5所示。可以看出Mish激活函数更加平滑,这将使网络更好的学习行人信息,同时Mish激活函数允许较小的负梯度流入,保证信息不会中断,从而得到更好的准确性和泛化能力。

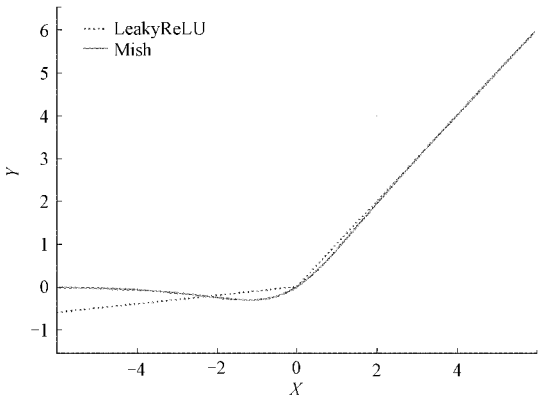


图5 Leaky ReLU和Mish激活函数

2 实验结果

本文实验环境配置如表1所示。实验采用python 3.6语言编写,深度学习框架为Pytorch 1.4。训练批次设置为300,小批量设置为16,初始学习率为0.01,权重衰减系数为0.000 5,动量系数为0.9。

表1 实验环境配置

名称	配置
操作系统	Ubuntu 16.04
CPU	Intel(R) Core(TM) i5-9400F
RAM	16 GB
GPU	NVIDIA GeForce GTX 1080Ti, 11 GB
GPU加速库	CUDA 9.2, CUDNN7.0

2.1 数据集

实验数据集使用VOC2007和INRIA数据集。VOC2007数据集包含20类目标,共计9 963张图像。本文从VOC2007数据集提取了所有行人图像,共计4 015张,数据集背景复杂,行人姿态变化较大,存在不同程度的遮挡,能够增强训练模型的泛化能力,数据集采用8∶2的比例划分训练集和测试集。INRIA数据集中行人大多呈站立姿势,接近真实道路场景,已划分训练集和测试集。数据集行人图像数量如表2所示。

表2 行人数据集图像数量

数据集	训练集/张	测试集/张
VOC2007	3 212	803
INRIA	614	288

由于图像的背景差异和光照强度的不同会影响模型对行人特征的提取,本文对行人图像进行HSV色域颜色空间变换。HSV是一种侧重于观感的颜色模型,本文在图像预处理阶段对行人图像的曝光度、饱和度和色调进行调节,增强行人图像的丰富度,以提高模型的泛化性。

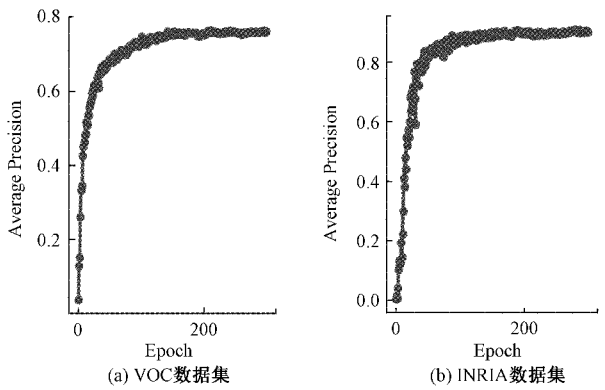
2.2 实验结果与分析

由于特征提取网络对特征图进行了32倍的下采样,所以网络的输入尺寸应为32的倍数。输入图像尺寸越大,网络提取的数据信息越丰富,但是训练和检测速度就会越慢。输入图像的尺寸过小,会导致信息的丢失,使模型无法学习到行人的重要特征,对模型的性能产生不利影响。在实际的行人图像中,行人的尺寸往往存在较大差异,为此本文采用多尺度训练方式,帮助模型适应各种大小的行人目标,提高模型的鲁棒性。在训练时每个批次的图像尺寸随机在(320,352,384,416,448,480,512,544,

576,608,640)中选择,以提高模型的泛化能力和鲁棒性。

为了评估改进算法的有效性,将 YOLOv3、Tiny YOLOv3、Tiny YOLOv4 和本文算法分别在 VOC2007 和 INRIA 数据集中进行训练并测试。训练之前,为了使前文提到的预选框更加贴合行人的形态,采用 K-means 聚类算法重新选择初始预选框,得到 6 个预选框尺寸,其中(38,97)、(81,202)、(126,386)对应 13×13 的预测层,(203,271)、(251,473)、(448,521)对应 26×26 的预测层。

测试指标包含精确率(Precision)、召回率(Recall),精确率指的是在检测出来的行人中,检测正确的行人所占的比例,召回率指的是所有的行人中,检测正确的行人所占的比例。采用综合指标精度均值(average precision, AP)来衡量检测算法的准确性,采用每秒帧数(frame per second, FPS)来衡量检测速度。为了得到训练的最佳模型,每个批次训练结束,使用测试集进行测试,保存 AP 最高的模型。图 6(a)和(b)为本文算法分别在 VOC2007 和 INRIA 数据集上训练的精度变化。可以看出,随着训练批次的增加,检测精度不断提高,最终达到一个相对稳定的阶段,表明网络模型训练效果良好。



(a) VOC数据集　(b) INRIA数据集

图 6　不同数据集下的 AP 变化

表 3 为不同算法的模型体积大小和参数量,本文算法的模型大小为 39.8 MB,与 Tiny YOLOv3 和 Tiny YOLOv4 相比模型大小有所增加,模型大小和参数量远小于 YOLOv3,在模型大小和参数量具有一定优势。

表 4 为各算法在两个数据集上的训练测试结果,与 Tiny YOLOv3 相比,本文算法的精确率和召回率均有提高。在 VOC 数据集上的行人检测准确率为 77%,较 Tiny YOLOv3 提高 8.5%,较 Tiny YOLOv4 提高 10.3%,虽然未达到 YOLOv3 的检测精度,但检测速度达到每秒 92.6 帧,相比 YOLOv3 提高 77.1%。在 INRIA 数据集上的准确率为 92.7%,较 Tiny YOLOv3 提高 2.5%,较 Tiny YOLOv4 提高 5.6%,比 YOLOv3 算法仅低 0.2%,本文算法的检测速度达到每秒 31.2 帧,满足实时性检测需求。

表 3　各算法的模型尺寸和参数量

算法名称	模型大小/MB	参数量×10^6
YOLOv3	235.0	61.50
Tiny YOLOv3	33.2	8.60
Tiny YOLOv4	30.9	8.02
本文算法	39.8	10.40

表 4　各算法的实验结果比较

数据集	算法名称	Precision/%	Recall/%	AP/%	FPS
VOC2007	YOLOv3	85.0	74.7	81.9	52.3
	Tiny YOLOv3	70.2	66.3	68.5	108.6
	Tiny YOLOv4	69.9	62.5	66.7	111.3
	本文算法	78.4	72.3	77.0	92.6
INRIA	YOLOv3	98.6	86.8	92.9	24.5
	Tiny YOLOv3	95.1	81.7	90.2	32.7
	Tiny YOLOv4	89.2	73.9	87.1	33.1
	本文算法	96.4	85.2	92.7	31.2

图 7 和 8 分别为 Tiny YOLOv3 和本文算法在不同场景下的检测结果对比。图 7(a)和 8(a)为侧面场景下的检测结果,Tiny YOLOv3 漏检了两个行人目标,本文算法无行人漏检;图 7(b)和 8(b)为拥挤场景下的检测结果,Tiny YOLOv3 行人漏检较为严重,本文算法得到明显改善;图 7(c)和 8(c)为正面场景下的检测结果,Tiny YOLOv3 漏检了左侧小尺寸的行人目标,本文算法中无行人漏检。可以看出,本文算法取得了更好的行人检测效果,并且在拥挤场景下和对小目标的检测中仍能取得良好的检测效果,这表明本文算法具有良好的泛化能力,能更准确地检测行人。

(a) 侧面场景　(b) 拥挤场景　(c) 正面场景

图 7　Tiny YOLOv3 检测结果

(a) 侧面场景 (b) 拥挤场景 (c) 正面场景

图8 本文算法检测结果

3 结 论

本文在Tiny YOLOv3的基础上,提出了一种融合注意力机制的行人检测算法,通过对网络的深化,提高了对行人信息的特征提取能力,通过1×1卷积降低了参数量和模型尺寸,保证了行人检测的速度。同时,在预测网络引入了一种无降维的轻量级通道注意力机制,对不同通道进行权重的再分配,使模型更加关注行人信息。并且,通过对边界框回归损失函数和激活函数的优化,进一步提高了检测精度。在VOC2007行人子集和INRIA数据集上取得了77%和92.7%的检测准确率,与Tiny YOLOv3相比精确率和召回率均有提高,检测速度分别达到每秒92.6帧和31.2帧,表明了该模型在不同数据集下具有良好的鲁棒性,且满足实时性检测需求。本文算法在保持较高检测准确率的同时具有速度优势,但是在面对行人姿态变化较大以及遮挡较为严重的情况,准确率与复杂的大型网络仍有差距,在接下来的工作中将考虑在满足实时检测的条件下,进一步提高检测精度。

参考文献

[1] 崔少华,汪徐德,王江涛,等.高斯建模和卷积神经网络联合的红外视频行人检测方法[J].电子测量与仪器学报,2020,34(5):140-148.

[2] 周志锋,万旺根,王旭智.基于YOLOv3框架改进的目标检测[J].电子测量技术,2020,43(18):102-106.

[3] 包本刚.融合多特征的目标检测与跟踪方法[J].电子测量与仪器学报,2019,33(9):93-99.

[4] 齐美彬,李佶,蒋建国,等.改进特征与GPU加速的行人检测[J].中国图象图形学报,2018,23(8):1171-1180.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, USA, 2014: 580-587.

[6] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440-1448.

[7] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. Advances in Neural InformationProcessing Systems,2015: 91-99.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceedings of 14th European Conference on Computer Vision, Amsterdam Netherlands,2016: 21-37.

[9] 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, Las Vegas, NV, USA, 2016: 779-788.

[10] REDMON J, FARHADI A. YOLO 9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 6517-6525.

[11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.

[12] 刘文强,辛大欣,华瑾,等.基于Faster RCNN的镁还原罐工人检测算法[J].国外电子测量技术,2019,38(4):12-17.

[13] HUANG Z C, WANG J L, FU X S, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection[J]. Information Sciences, 2020,522:241-258.

[14] XIAO F, LIU B T, LI R N. Pedestrian object detection with fusion of visual attention mechanism and semantic computation[J]. Multimedia Tools and Applications,2020,79(21-22):14593-14607.

[15] 王瑞,史天运,包云.一种基于视频的铁路周界入侵检测智能综合识别技术研究[J].仪器仪表学报,2020,41(9):188-195.

[16] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 2117-2125.

[17] HUANG G, LIU Z, LAURENS V, et al. Densely

connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017:2261-2269.

[18] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11531-11539.

[19] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks [J]. ArXiv Preprint, 2017, ArXiv: 1709.01507.

[20] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a Loss for bounding box regression[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 658-666.

[21] MISRA D. Mish: A self regularized non-monotonic neural activation function[J]. ArXiv Preprint, 2019, ArXiv: 1908.08681.

作者简介

冯宇平，博士，副教授，主要研究方向为图像处理、模式识别、人工智能。

E-mail:gjsfengyuping@163.com

管玉宇，硕士研究生，主要研究方向为模式识别与机器视觉。

E-mail:1095732969@qq.com

杨旭睿，硕士研究生，主要研究方向为目标检测。

E-mail:810218451@qq.com

刘宁，硕士研究生，主要研究方向为表情识别。

E-mail:1459226842@qq.com

王兆辉，本科生，主要研究方向为人工智能。

E-mail:1966945205@qq.com