

文章编号:1007-2780(2019)12-1191-11

基于深度可分离卷积的交通标志识别算法

杨晋生, 杨雁南*, 李天骄

(天津大学 微电子学院, 天津 030072)

摘要:针对目前实时交通标志识别中出现的对于中小型目标检测精度低的问题,本文提出了一种基于 YOLOv3-tiny 的深度可分离卷积(Depthwise Separable Convolutions)的轻量级交通标志检测网络。采用深度可分离卷积构建深度可分离卷积模块代替普通卷积搭建特征提取网络,在保证计算量的前提下更好地提取中小型目标的特征信息。同时,改进多尺度特征融合网络,提高对中小型交通标志的检测精度,使用 h-swish 激活函数减少因为网络层数增加而丧失的图像特征,实现对多类交通标志的检测。实验结果表明:本算法有效的提高了对中小型交通标志的检测,在验证集上对警告标志(Warining)指示标志(Mandatory)、禁止标志(Prohibitory)3 类交通标志进行检测,检测精度(AP)结果分别为 98.57%,96.03%,98.04%。检测平均精度(mAP)97.54%、检测速度为 201.5 f/s,平均精度较 YOLOv3-tiny 提高了 14.01%。在保证轻型网络的计算量低、检测时效性好的前提下,有效地提升了检测精度。

关键词:图像处理;目标检测;交通标志;深度可分离卷积;YOLOv3;h-swish

中图分类号:TP391 **文献标识码:**A **doi:**10.3788/YJYXS20193412.1191

Traffic sign recognition algorithm based on depthwise separable convolutions

YANG Jin-sheng, YANG Yan-nan*, LI Tian-jiao

(School of Microelectronics, Tianjin University, Tianjin 030072, China)

Abstract: In order to solve the problem of low precision detection for small and medium-sized targets in the real-time traffic sign recognition, a lightweight traffic sign detection network based on YOLOv3-tiny depthwise separable convolution is proposed. In this paper, a deep separable convolution module is constructed by using deep separable convolution instead of the common convolution to construction feature extraction network, the feature information of small and medium-sized targets is better extracted under the premise of ensuring the calculation amount. At the same time, the multi-scale feature fusion network is improved to improve the detection accuracy of small and medium-sized traffic signs. The h-swish activation function is used to reduce the image features lost due to the increase of the number of network layers, and the detection of multiple types of traffic signs is realized. The experimental results show that the algorithm effectively improves the detection of small and medium-sized traffic signs. Warining, mandatory and prohibitory traffic signs are detected on the verification set. The detection accuracy (AP) is 98.57%, 96.03% and 98.04% respectively. The average

收稿日期:2019-08-26;修订日期:2019-09-26.

* 通信联系人, E-mail:1191547337@qq.com

detection accuracy (mAP) was 97.54% and the detection speed was 201.5 f/s. The average accuracy was 14.01% higher than that of YOLOv3-tiny. The algorithm effectively improves the detection accuracy under the premise of ensuring low calculation of light network and good timeliness of detection.

Key words: image processing; target detection; traffic signs; depth separable convolution; YOLOv3; h-swish

1 引言

随着信息时代的到来,在科学研究、互联网应用、电子商务、自动驾驶等诸多应用领域对于人工智能的需求日益迫切^[1]。在这种情况下,基于机器学习和人工智能的研究必将成为今后计算机和通信技术发展中最重要的一部分^[2]。

交通标志是现代公路交通环境中最重要的组成部分,随着社会的进步,路面交通变得多样化、立体化,交通标志也因此变得多种多样。自动驾驶作为时下最热门的内容,包括路面感知、路径规划、决策等技术。其中,作为车辆行驶的指挥者,交通标志的识别是自动驾驶和车辆辅助驾驶技术中最核心的技术之一。受天气状况、光照变化和硬件设施的限制,在车辆行驶过程中所捕获到的图像常常受到图像模糊、路标褪色、背景掩盖等问题的影响,如何快速准确地识别交通标志成为了当前研究的一个重要方向。

目前对于交通标志的识别主要可分为传统图像处理方法和基于卷积神经网络两个大的研究方向。采用传统图像处理方法进行目标识别多采用经典的基于 HOG-SVM^[3]的方法进行目标识别。卷积神经网络(Convolutional Neural Networks, CNN)的出现为目标识别提供了一种全新的处理方法。采用深度学习进行目标识别大致分为两类:基于候选区域目标检测器的网络如 AlexNet^[4]、R-CNN^[5]、Fast R-CNN^[6]、Faster R-CNN^[7]、R-FCN^[8]、FPN^[9]等;基于端到端的不需要提供候选区域的网络如 SSD^[10]、YOLO^[11]等。基于候选区域目标检测采用选择性搜索算法进行评估相邻图像间的相似度,将相似度高的合并,进而选出合适的子图进行特征提取。基于候选区域目标检测的算法具有较高的精度,但同时由于其庞大的计算量,其检测速度并不能满足交通标志识别所需要的时效性。如 Ren 等人提出的 Faster R-CNN,其在 VOC2012^[12]数据集上的检

测速度只有 5 f/s。由于基于候选区域的检测算法普遍存在运算速度慢、训练复杂的特点,YOLO 创造性地提出了将物体定位和物体分类合二为一的算法,大大提高了对物体识别速度的要求,使识别速度提升到了 45 f/s,满足了对检测实效性的要求。但同时 YOLO 系列算法也存在识别精度较差的缺陷,在对 COCO 数据集中 20 类物体的检测中仅获得平均精度(mAP)63%的结果,其在小目标检测上更有着明显的缺陷。尽管如此,其在检测速度上的优势,使其在移动端部署有着得天独厚的优势。

近年来,在保证检测速度的同时提升检测精度已经成为了当下目标检测算法中最热门的研究领域。SSD 综合了检测精度最高的 Faster-RCNN 和检测速度最快的 YOLO 的优势,在识别精度和速度上进行了平衡。2017 年,Redmon^[13]在 YOLO 的基础上进行了优化形成了 YOLOv2,全新的 YOLOv2 在 VOC2017 数据集上获得了 76.8 mAP,优于使用 ResNet 的 SSD 和 Faster-RCNN,2018 年 Redmon 在 YOLOv2 的基础上引入了多尺度预测产生了 YOLOv3,对小目标的检测进行了优化,同时基于 YOLOv3 也设计了更简化的版本 YOLOv3-tiny,有利于其在移动端上的应用。

本文所做的贡献有:

(1)针对 YOLOv3-tiny 检测精度低的问题,采用深度可分离卷积(Depthwise Separable Convolution^[14])搭建主干网络,相较于普通的卷积运算,大量减少计算量,从而在相同的计算量条件下搭建更多层次的网络,提取更高层次的语义特征。

(2)优化深度可分离卷积计算模块,采用 swish^[15]函数的简化版本 h-swish 做为激活函数,在保证激活函数计算量小的同时保留更多的特征,减少因网络层数增加而造成的随机梯度消失,进而提高检测的精度。

(3)在 YOLOv3-tiny 的多尺度特征融合网络基础上,增加 8 倍下采样融合,构建更多尺度的特

征融合网络,提高对小目标的检测精度。

2 目标检测原理

2.1 YOLO 目标检测原理

YOLO(You Only Look Once)是目前最先进的端对端目标检测的算法之一。YOLO 将多目标检测转换成一个回归问题,去除了以往结构中提取待检测区域(Proposal)的步骤,通过一次回归运算完成目标位置、目标种类及其目标置信度的工作。通过调整网络最终输出种类的数量控制最后输出的检测矢量大小,进而实现对多类物体的检测。

YOLOv3 作为目前 YOLO 系列最新的一代算法,是在吸取了 Faster R-CNN 采用预测 Anchor box 偏移值代替直接预测目标框中心位置坐标值、借鉴 Resnet 加入残差模块、增加 FPN^[16] (Feature pyramid networks)等方法后提出的最新的网络结构。YOLOv3 网络由骨干的 Darknet-53 网络和基于 FPN 的多尺度融合的检测网络构成。通过在 13×13 、 26×26 、 52×52 三种不同尺度的 YOLO 层上进行预测,极大的改善了对不同尺度大小问题和被遮挡物体的效果。在针对多目标检测问题上,使用 Logistic 分类器取代了 Softmax 分类器,采用多尺度随机训练来增强网络的鲁棒性。

在 YOLO 中,将整个图片划分成 $S \times S$ 个格子,每个格子包含 B 个预测框(Bounding Boxes)和 C 个物体类别的预测概率 $P_r(\text{class}_i | \text{object})$,其中每个预测框返回 5 个参数 $(x, y, w, h, \text{object.conf})$,其中 (x, y, w, h) 为预测框的中心位置, (w, h) 为预测框的长宽值, object.conf 为该预测框的置信度(C Confidence score), $\text{IOU}_{\text{predict}}^{\text{truth}}$ 为预测边框和真实边框的交集比例(Intersection over union)。该预测框内物体为每一种物体的概率为预测框置信度和该预测框中 C 类物体对应的概率 $P_r(\text{class} | \text{object})$ 的乘积,也等于 C 类物体中每类的概率 $P_r(\text{class}_i)$ 与如公式(1)所示:

$$\begin{aligned} \text{object.conf} &= P_r(\text{object}) \cdot \text{IOU}_{\text{predict}}^{\text{truth}} \cdot I_{\text{conf}} = \\ &P_r(\text{class} | \text{object}) \cdot \text{object.conf} = \\ &P_r(\text{class}_i) \cdot \text{IOU}_{\text{predict}}^{\text{truth}}. \end{aligned} \quad (1)$$

YOLO 系列从 YOLOv2 开始使用 Anchor box 的思想进行预测框的迭代逼近,对预测框的

逼近算法进行了更新,每一个预测框返回值参数为 (b_x, b_y, b_w, b_h) ,其具体计算方式如公式(2)所示:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases}. \quad (2)$$

YOLOv3 的损失函数如公式(3)所示:

$$\begin{aligned} K_{\text{loss}} &= \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + \\ &(y_i - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \\ &[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \\ &\sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \\ &\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 + \\ &\sum_{i=0}^{S^2} 1_{ij}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned}, \quad (3)$$

其中: λ_{coord} 为是增强定位误差(Localization error)在计算中的重要性参数, λ_{coord} 是减少定位误差在计算中重要性的参数, C_i 是对应目标框的可信度, 1_{ij}^{obj} 为目标在网格 i 预测框 j 中的取值,当目标框中存在物体是值为 1,不存在物体时值为 0。

YOLOv3-Tiny 作为 YOLOv3 的简化版本,在牺牲了一部分检测精度的情况下将检测速度的优势进一步发挥。YOLOv3-Tiny 的主干网络由 7 层的类 Darknet-19 网络构成,输出 13×13 和 26×26 两种不同尺度的特征进行检测特征回归,预测出多个预测框和其相应的置信度得分,通过使用非极大抑制(Non-maximum suppression, NMS)算法剔除掉低置信度的预测结果,保留高置信度的预测框输出。

2.2 深度可分离网络

对于传统的卷积计算,在进行卷积运算时需要同时将对应图像区域中的所有通道进行考虑。深度可分离卷积提出了一种将图像的对应区域和通道分开进行处理,先考虑区域特征,然后再考虑通道特征,将传统的卷积运算分解成 Depthwise

和 Pointwise 两个过程。

对于传统的卷积运算,有 $N \times H \times W \times C$ 的输入与 k 个 $n \times n$ 的卷积核进行卷积运算,在 $pad=1$ 和 $stride=1$ 的情况下输出为 $N \times H \times W \times k$ 。

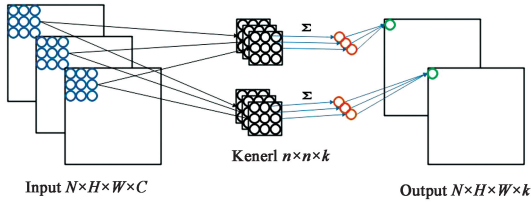


图 1 标准卷积过程

Fig.1 Standard convolution process

对于深度可分离卷积,在 Depthwise 过程中,将输入的 $N \times H \times W \times C$ 划分成 C 组,对于每一组做 $n \times n$ 的卷积运算,收集卷积运算中的每个 channel 的空间特征。

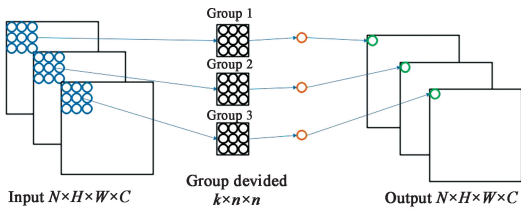


图 2 Depthwise 卷积过程

Fig.2 Depthwise convolution process

在 Pointwise 过程中,对 Depthwise 中的结果进行 k 个 1×1 的卷积运算,收集每个 point 上的特征,也就是卷积运算中的位置特征,经过 Pointwise 和 Depthwise 两个步骤,最终输出的结果也是 $N \times H \times W \times k$ 。

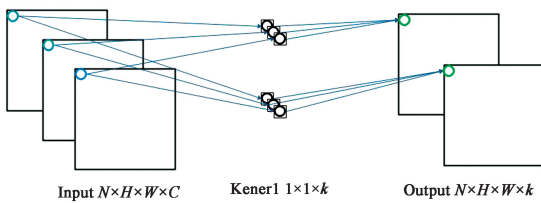


图 3 Pointwise 卷积过程

Fig.3 Pointwise convolution process

对于 $N \times H \times W \times C$ 和 $k \times n \times n$ 的卷积运算,普通的卷积运算运算量为 $H \times W \times C \times k \times n \times n$,对于深度可分离卷积运算,在 Depthwise 阶段计算量为 $H \times W \times C \times n \times n$;在 Pointwise 阶段计算量为 $H \times W \times C \times k$,总的计算量为 $H \times$

$W \times C \times (k + n \times n)$ 。

相较于普通的卷积计算,其计算量压缩 P 如公式(4)所示:

$$P = \frac{\text{depthwise} + \text{pointwise}}{\text{convolution}} = \frac{H \times W \times C \times n \times n + H \times W \times C \times k}{H \times W \times C \times k \times n \times n} = \frac{1}{k} + \frac{1}{n \times n} \approx \frac{1}{n \times n} \quad (4)$$

在一般的卷积运算中, k 的值远大于 $n \times n$,计算压缩量 P 近似于 $\frac{1}{n \times n}$,以常用的 3×3 的卷积核来说,其计算量压缩约为之前的 $\frac{1}{9}$,相较于普通的卷积运算,基于深度可分离卷积的运算大大节省了卷积计算的运算量。

3 改进的网络模型

3.1 优化的深度可分离卷积的特征提取主体网络

为了实现对交通标志的动态识别,解决现在 YOLOv3-tiny 等 YOLO 系列轻型网络检测精度低的问题,采用可分离深度卷积重新构建了特征提取网络,代替 YOLOv3-tiny 中的类 Darknet-19 网络,提取更高层次的图像特征信息。

相较于普通的卷积运算模块,基础的深度可分离卷积模块如图 4 所示。

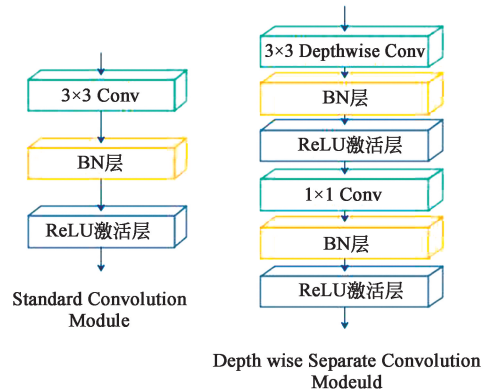


图 4 标准卷积与深度可分离卷积模块

Fig.4 Standard convolution and depth separable convolution module

与普通的卷积模块相似,深度可分离卷积模块在 Depthwise 和 Pointwise 层后增加 BN 层和激活函数层。但是采用 ReLU 作为激活函数,因

为其对负值取 0 的特性,会丢失掉部分的特征信息的问题,所以在 Depthwise Conv 模块后选择 Leaky ReLU 代替 ReLU 作为激活函数。同时在 1×1 的 Conv 层中,选择使用 h-swish 函数来代替 Relu 作为激活函数,来减少由于网络层数增加造成的网络梯度消失的影响。

h-swish 函数是 swish 激活函数的简化版,swish 的计算方法如公式(5)所示:

$$\text{Swish}(x) = x \cdot \text{sigmond}(\beta \cdot x)$$

$$\text{sigmond} = \frac{1}{1 + e^{-x}}, \quad (5)$$

其中: β 是个常数或可训练的参数。swish 具备无上界有下界、平滑、非单调的特性。

相较于 ReLU 激活函数及其衍生激活函数,在深层网络中,swish 激活函数可以很好的提高神经网络的准确性。h-swish 函数的计算方式如公式(6)所示:

$$h\text{-Swish}[x] = x \cdot \frac{\text{ReLU6}(x + 3)}{6} \quad (6)$$

$$\text{ReLU6}[x] = \min(\max(0, x), 6)$$

两种激活函数图像对比如图 5 所示。相较于 swish 函数需要计算 Sigmoid 从而计算时间长的特点,在图像上 h-swish 函数在保持 swish 激活

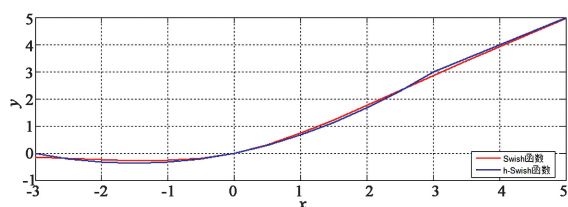


图 5 h-swish 函数与 swish 函数对比图

Fig.5 Comparison of h-swish function and swish function

函数具备无上界有下界、平滑、非单调的特点外,其函数只由 ReLU 函数构成外,极大的减少了 swish 函数的运算时间,实现了精度和运算速度上的兼容。

优化后的基础深度可分离卷积模块结构如下图 6 所示。

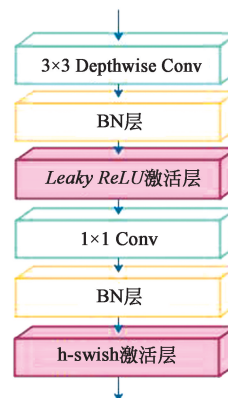


图 6 优化的深度可分离卷积模块

Fig.6 Optimized deep separable convolution module

3.2 优化的多尺度融合网络

采用优化后的深度可分离卷积模块搭建主体网络提取图像特征。对于实时的交通路标检测,其重点是对中远距离交通路标的检测,也就是对图像中小目标的检测,为了提高对中远距离交通标志的检测,给予驾驶员足够的反应时间。为减少中小尺度的目标语义损失、更好的提取中小目标的特征,同时避免高分辨率特征提取增加较大的计算量。在网络设置中减少对 32 倍下采样的特征提取层设置,在 8 倍下采样和 16 倍下采样设置更多层次网络进行的特征提取。特征提取主体网络结构如表 1 所示。

表 1 特征提取主体网络

Tab.1 Feature extraction subject network

Type	Filters	Input	Output
Conv	32	$416 \times 416 \times 3$	$208 \times 208 \times 32$
Depthwise Conv	32	$208 \times 208 \times 32$	$208 \times 208 \times 32$
Conv	64	$208 \times 208 \times 32$	$208 \times 208 \times 64$
Depthwise Conv	64	$208 \times 208 \times 64$	$104 \times 104 \times 64$
Conv	128	$104 \times 104 \times 64$	$104 \times 104 \times 128$
Depthwise Conv	128	$104 \times 104 \times 128$	$104 \times 104 \times 128$

续 表

	Type	Filters	Input	Output
	Conv	128	104×104×128	104×104×128
	Depthwise Conv	128	104×104×128	52×52×128
	Conv	256	52×52×128	52×52×256
3×	Depthwise Conv	256	52×52×256	52×52×256
	Conv	256	52×52×256	52×52×256
	Depthwise Conv	256	52×52×256	26×26×256
	Conv	512	26×26×256	26×26×512
3×	Depthwise Conv	512	26×26×512	26×26×512
	Conv	512	26×26×512	26×26×512
	Depthwise Conv	512	26×26×512	13×13×512
	Conv	1 024	13×13×512	13×13×1 024
1×	Depthwise Conv	1 024	13×13×1 024	13×13×1 024
	Conv	1 024	13×13×1 024	13×13×1 024

YOLOv3-tiny 采用 16 倍下采样和 32 倍下采样的两个尺度的特征融合,这意味着当输入的检测目标小于 16 pixel×16 pixel 时,就丧失了对该目标的检测能力。为提高网络对小目标的检测精度,提高对目标位置信息的判断,使得网络获得更多小目标的特征信息。在 YOLOv3-tiny 的基础上增加对网络 8 倍下采样的尺度融合,将原本 16 倍下采样融合后的输出结果进行上采样,与主体网络中输出的 8 倍下采样结果进行融合,建立 8 倍下采样的融合特征检测层,收集 8 pixel×8 pixel 尺度的特征,提高对小目标的检测和位置信息判断。

网络结构示意图如图 7 所示。

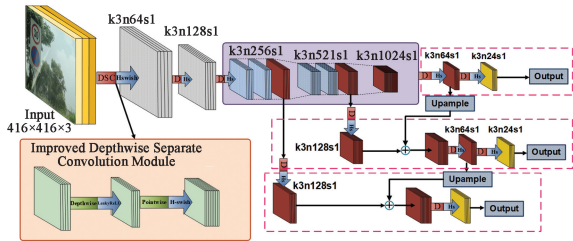


图 7 网络结构示意图

Fig.7 Schematic diagram of the network structure

4 结果分析

本文采用中国交通标志检测数据集(CCTS-

DB^[17])作为实验数据集,实验使用 Ubuntu 16.04 操作系统、Darknet 深度学习框架、CUDA9.0 进行加速,硬件配置为:CPU Intel Core i7-9750H、GPU Nvidia GTX 1080。本次实验中,我们将从网络计算量大小、模型检测速度、检测准确度和召回率等多个指标与将本文算法与目前最主流的轻型网络 YOLOv3-tiny、YOLOv2-tiny 以及 YOLO 系列的最新版本 YOLOv3 进行对比。

4.1 数据集

中国交通标志检测数据集(CCTSDB)是在原中国交通标志数据集(CTSD)数据集的基础上扩充得到的。数据集包含 15 734 张由摄像机拍摄而成的现实路况图片。数据集由正样本和负样本构成,正样本中包含一个或多个交通标志,负样本不包含交通标志,所有图像横纵尺寸均大于 416 pixel。数据集采用 PASCAL VOC 格式进行标注,标注数据依据交通标志的作用分为警告标志(Warining)指示标志(Mandatory)、禁止标志(Prohibitory)3 类。本次训练选取前 2 000 张为验证集,剩余图片作为训练集对数据集中标注的 3 类交通标志进行检测,训练集与验证集比例约为 7 : 1。

4.2 训练参数设置

本次实验训练阶段初始学习率为 0.001,衰

减系数为 0.000 5,在训练迭代次数为 40 000 次和 45 000 次时,分别将学习率降低为 0.000 1 和 0.000 01。使损失函数进一步收敛,在训练阶段利用增加对比度、调整曝光量和旋转图像方法对训练集中图像进行扩充。

从 YOLOv2 开始引入 anchor box 的思想,YOLO 系列中设置的初始 anchor box 是作者使用 K-means 聚类方法在 COCO 数据集上聚类的结果,对本次训练所检测的交通标志并不适用,因此需要重新设置 anchor box。在本次实验中使用平均重叠度(Avg IOU)这一指标作为初始 anchor box 的聚类指标,其计算公式如公式(7)所示:

$$\operatorname{argmax} \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} I_{\text{IOU}}(B, C)}{n}, \quad (7)$$

其中: B 表示 ground truth 中标注好的目标框位置。 C 表示设置 anchor 的中心, n_k 表示第 k 个

anchor box 样本的个数, n 表示样本的总个数, k 表示 anchor box 的个数, $I_{\text{IOU}}(B, C)$ 表示簇的中心框和聚类框的交并比。

对本文中数据进行 Avg IOU 聚类,综合网络结构和 Avg IOU 后,选择 $k=6$,设置 6 个初始 anchor box,各个 anchor 对应长宽分别为:(7, 9),(14, 18),(23, 30),(26, 41),(41, 62),(74, 106)。

本文算法模型及各对比算法模型的损失函数下降图如图 8 所示。

由损失函数曲线可以看出,对比 4 种算法,YOLOv3-tiny 收敛最快,在 40 000 次迭代后趋于稳定,但损失函数下降波动相较本文算法波动较大,本文算法损失函数由于采用深度可分离卷积搭建特征提取网络,因此收敛速度较普通卷积相比较慢在 70 000 次左右损失函数趋于平稳,但由于增加了更多层次的特征提取网络和三尺度的特

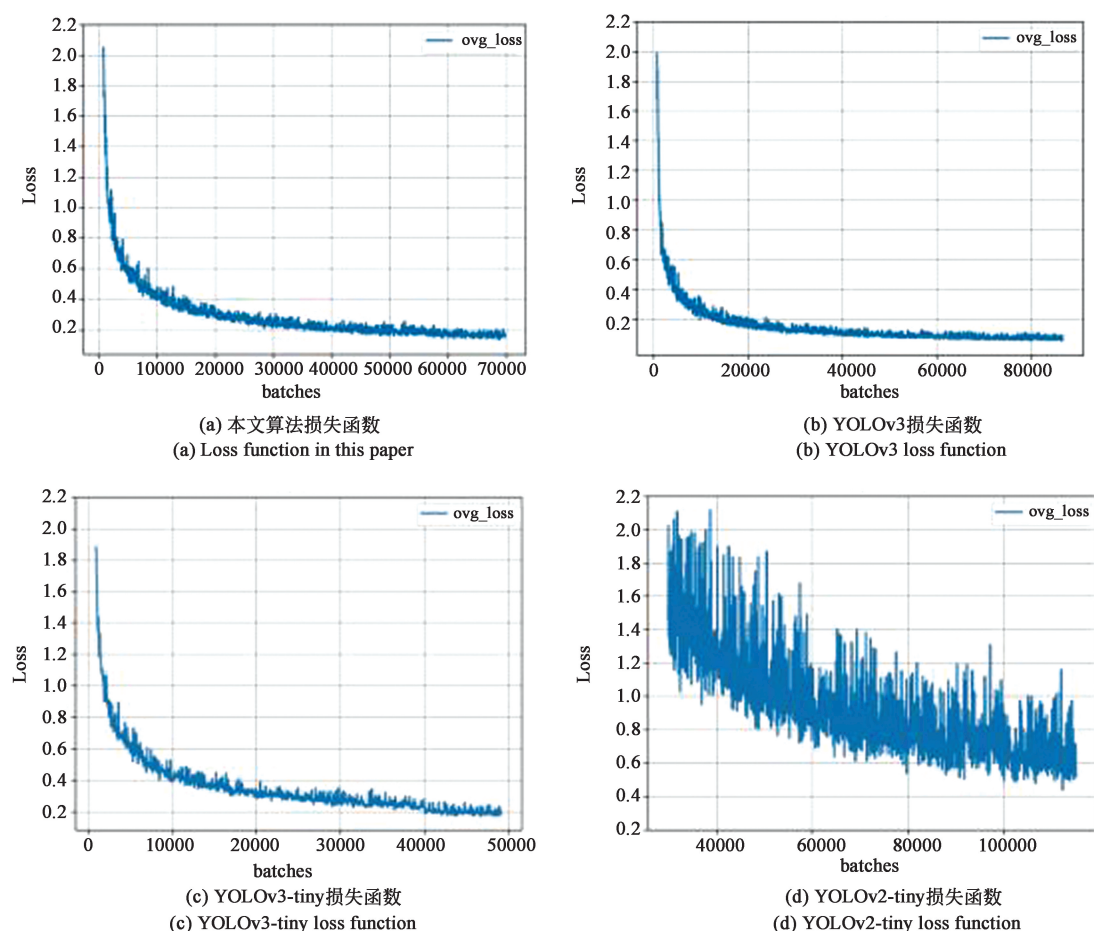


图 8 损失函数下降图

Fig.8 Loss function decline graph

征融合,训练平稳后的损失函数更小,精度更高。YOLOv3 作为大型网络,其下降趋势平稳且波动最小,但由于 YOLOv3 网络层数多且训练参数复杂,因此迭代收敛速度较慢,在 80 000 次迭代后才趋于平稳,但训练平稳后损失函数最小,精度也最高。YOLOv2-tiny 虽然网络结构最简单,训练速度快,但其收敛最慢,在 100 000 次训练之后才趋于稳定,且稳定后的收敛函数数值最大,精度最低。

4.3 实验结果与分析

为进一步验证本模型能有效的降低了卷积运算中的计算量和参数数量,本文采用浮点数计算量 BFLOPs (Billion Floating Point Operations) 对传统的卷积模型和本文所提出的基于深度可分离卷积的计算量分析。在本文中 52×52 , 26×26 , 13×13 三个尺度上的基于传统卷积和基于深度可分离卷积的运算次数对比如表 2 所示。

表 2 浮点数计算量对比表

Tab.2 Comparison of floating point calculation amount

Input	Output	Depthwise Separable Conv			Standard Conv	Compression Ratio
		Depthwise Conv	Pointwise Conv	Total		
$52 \times 52 \times 64$	$52 \times 52 \times 128$	0.006	0.044	0.050	0.399	0.125
$26 \times 26 \times 128$	$26 \times 26 \times 256$	0.003	0.044	0.047	0.399	0.118
$13 \times 13 \times 256$	$13 \times 13 \times 512$	0.002	0.044	0.046	0.399	0.115

由此可见,对于 3 个尺度上的计算,相较于传统的卷积运算,基于深度可分离卷积的运算模块均有效的减少了运算的计算量。

为进一步评价本算法效果,类别检测平均精度度 (Average Precision, AP)、平均精度 (Mean Average Precision, mAP)、每秒帧数 (Frames Per Second, fps) 作为评价标准,将本文网络结构与 YOLOv3-tiny、YOLOv2-tiny 和 YOLOv3 进行对比。

平均精度度及平均精度计算方式如公式(8)所示:

$$\begin{cases} AP_c = \frac{1}{11} \sum_{R \in \{0, 0.1, \dots, 1\}} P(R_c) \\ P(R_c) = \max_{\tilde{R}, \tilde{R} \geq R} p(\tilde{R}_c) \\ mAP = \frac{1}{N} \sum AP_c \end{cases}, \quad (8)$$

其中: AP_c 采用 VOC2007^[18] 的计算方式进行计

算,将召回率曲线中划分共 11 个等距点(0, 0.1, 0.2, ..., 0.9, 1)后,求其最大精度的平均值可获得平均精度 AP,即各类物体的平均准确率。N 为网络中物体类别的数量,mAP 为各类物体 AP 的平均数。

本文使用法每秒处理的图像数量 f/s 作为评价网络检测速度的指标。

将本文算法与 YOLOv3-tiny、YOLOv2-tiny、YOLOv3 的结果进行对比,所有算法均采用上文设置的 anchor box,将其在 2000 张验证集上检测,其各项性能对比如表 3 所示。

从表 3 可以看出,本文提出的算法在验证集上获得 mAP 为 97.54%、Waring 类准确率 98.57%、Mandatory 类准确率 96.03%、Prohibitory 类准确率 98.17%、检测速度为 201.5 f/s,相较于 YOLOv3-tiny 分别提高了 14.01%, 9.74%,12.43%、

表 3 不同算法性能对比表

Tab.3 Performance comparison of different algorithms

	mAP/%	Waring/%	Mandatory/%	Prohibitory/%	检测速度/(f · s ⁻¹)
YOLOv3	98.87	99.48	98.97	98.17	33.8
YOLOv3-tiny	83.53	88.83	83.60	78.18	208.3

续 表

	mAP/%	Warning/%	Mandatory/%	Prohibitory/%	检测速度/(f · s ⁻¹)
YOLOv2-tiny	58.39	67.95	59.04	48.20	214.7
Article Method	97.54	98.57	96.03	98.04	201.5

19.86%, 相较于 YOLOv2-tiny 分别提高了 39.15%, 30.62%, 36.99%, 49.84%, 检测速度指标上 3 种算法基本持平, 均稳定在 200 f/s 以上。其主要原因是本文方法和 YOLOv3-tiny 均采用特征图像金字塔搭建多尺度特征融合, 将多个尺度的特征融合进行检测, 因此在检测精度上较 YOLOv2-tiny 有了较大提升。本文算法较 YOLOv3-tiny 使用优化后的深度可分离卷积搭建了 27 层特征提取网络, 较 YOLOv3-tiny 原

本的 7 层的特征提取网络有了明显的提升, 同时加入了 8 倍的下采样, 便于采集中小型目标的特征, 使用 h-swish 函数作为激活函数有利于保留更多的特征信息, 提高了检测精度。本文算法作为轻型网络, 在 mAP 值较 YOLOv3 上较 YOLOv3 这样的大型网络下降了 1.33%, 但在检测速度上提高了 167.7 f/s。同时相较 YOLOv3 在检测过程中参数多、泛化能力弱的缺点有了长足的进步。

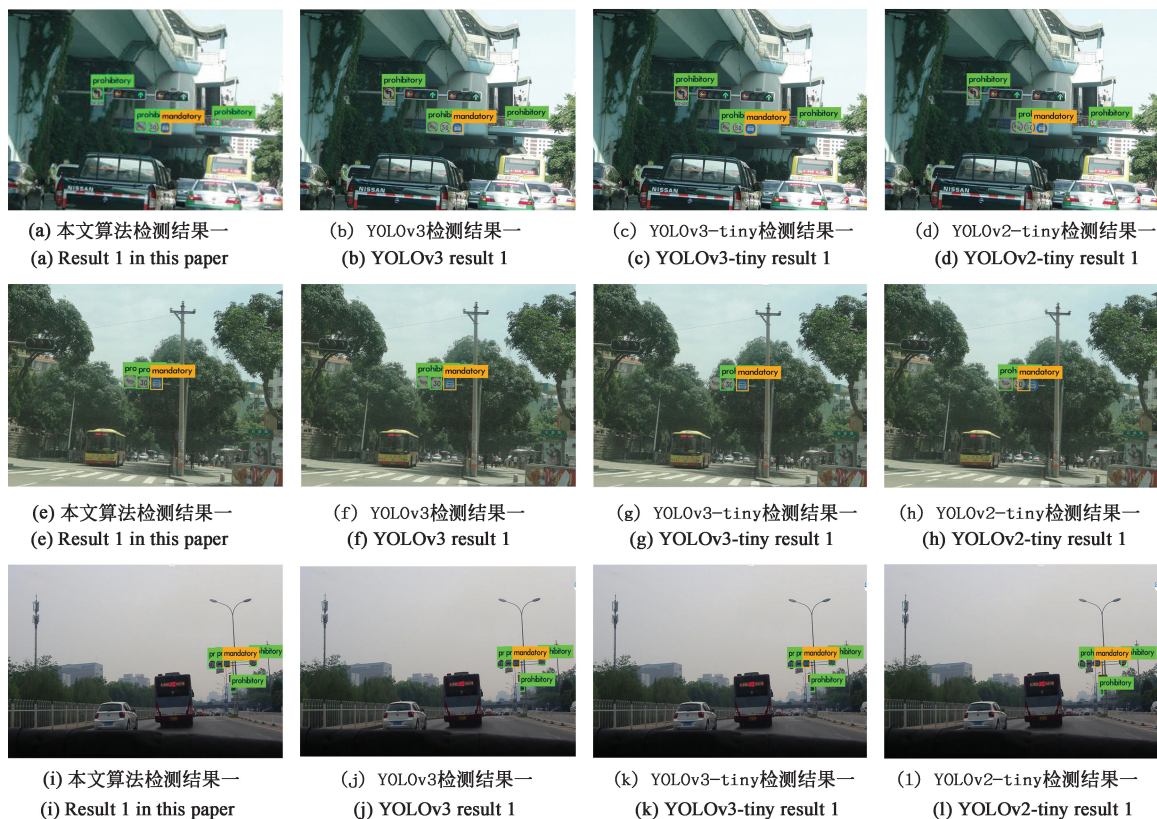


图 9 各种算法的检测效果图

Fig.9 Detection effect chart of various algorithms

图 9 为本文算法与 YOLOv3、YOLOv3-tiny、YOLOv2-tiny 对交通标志的检测效果对比图。其中图 (a)、(b)、(c)、(d) 中, 本文算法、YOLOv3 和 YOLOv3-tiny 均可有效对交通标志进行检测, 但 YOLOv2 因为没有采用多尺度检

测, 所以检测精度最差, 对部分交通标志没有办法进行识别。图 (e)、(f)、(g)、(h) 中, 面对交通标志褪色的现象, 本文算法、YOLOv3 均可正确识别交通标志, 但 YOLOv3-tiny 因为只有两个层次的检测, 没有进行小目标的融合, 因此没有办法很好

地识别交通标志, YOLOv2-tiny 除了没有办法识别交通标志外还出现了明显的预测位置偏移。图(h)、(i)、(j)、(k)中当面对天气变暗的条件下, 本文算法、YOLOv3 和 YOLOv3-tiny 均能准确识别交通标志, YOLOv2-tiny 并不能准确检测。

5 结 论

本文根据目前针对交通标志识别的算法难以兼顾检测速度和检测精度的问题, 提出了一种基于深度可分离卷积的轻量级交通标志识别网络。分析了目标识别算法的发展过程, 对

YOLO 系列算法和深度可分离卷积思想进行了介绍。使用深度可分离卷积搭建了特征提取网络代替 YOLOv3-tiny 的特征提取网络, 同时为更好地识别中小目标, 增加了 8 倍下采样的尺度融合, 在搭建网络过程中使用 h-swish 激活函数代替 ReLU 激活函数, 保留了更多的特征信息。实验结果表明, 本文提出的算法相较于 YOLOv3-tiny 有了较大的提升, 交通标志的识别平均精度由 83.53% 提高到了 97.54%, 检测速度达到了 201.5 f/s, 同时相较 YOLO 系列网络具有更好的泛化能力, 有效满足了自动驾驶中对于交通标志检测准确性和时效性的需求。

参 考 文 献:

- [1] MUKHTAR A, XIA L K, TANG T B. Vehicle detection techniques for collision avoidance systems: a review [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(5): 2318-2338.
- [2] 马永杰, 宋晓凤. 基于 YOLO 和嵌入式系统的车流量检测[J]. 液晶与显示, 2019, 34(6): 613-618.
MA Y J, SONG X F. Vehicle flow detection based on YOLO and embedded system [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(6): 613-618. (in Chinese)
- [3] DENG Z D, ZHOU L P. Detection and recognition of traffic planar objects using colorized laser scan and perspective distortion rectification [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(5): 1485-1495.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada: Curran Associates Inc, 2012.
- [5] GIRSHICK R, DONAHUE J, DARRELLT, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014.
- [6] GIRSHICK R. Fast R-CNN [C]//*IEEE International Conference on Computer Vision* Santiago, Chile: IEEE, 2015: 1440-1448.
- [7] 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.
- [8] 王丹丹, 何东健. 基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别[J]. 农业工程学报, 2019, 35(3): 156-163.
WANG D D, HE D J. Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(3): 156-163. (in Chinese)
- [9] HUANG S Z, LI X Y, JIANG Z Q, et al. Hyper feature fusion pyramid network for object detection [C]//*Proceedings of 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. San Diego, CA, USA: IEEE, 2018.
- [10] LIU W, ANGUELOV D, ERHAND, et al. SSD: single shot MultiBox detector [C]//*Proceedings of the 14th European Conference Computer Vision ECCV 2016*. The Netherlands: Springer, 2016.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016.
- [12] VISHWAKARMA A, BHUYAN M K, IWAHORIY. An efficient algorithm for medical image fusion using non-

- subsampled shearlet transform[C]//CHAUDHURI B B, KANKANHALLI M S, RAMAN B. *Proceedings of 2nd International Conference on Computer Vision & Image Processing: CVIP 2017, Volume 1*. Singapore: Springer, 2018.
- [13] JO K U, IM J H, KIM J, *et al.* A real-time multi-class multi-object tracker using YOLOv2 [C]//*Proceedings of 2017 IEEE International Conference on Signal and Image Processing Applications*. Kuching, Malaysia: IEEE, 2017: 507-511.
- [14] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017.
- [15] CHIENG H H, WAHID N, ONG P, *et al.* Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning [J]. *International Journal of Advances in Intelligent Informatics*, 2018, 4(2): 76-86.
- [16] HUANG S Z, LI X Y, JIANG Z Q, *et al.* Hyper feature fusion pyramid network for object detection [C]//*Proceedings of 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. San Diego, CA, USA: IEEE, 2018.
- [17] ZHANG J M, HUANG Q Q, WU H L, *et al.* Effective traffic signs recognition via kernel PCA network [J]. *International Journal of Embedded Systems*, 2018, 10(2): 120-125.
- [18] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, *et al.* The PASCAL visual object classes challenge 2007 (VOC2007) results [EB/OL]. <http://www.pascal-network.org/challenges/VOC/voc2007/index.html>.

作者简介:



杨晋生(1965—),男,河北邢台人,副教授,硕士生导师,1999 年于天津大学获得硕士学位,主要从事无线传播理论与技术的研究。E-mail:jsyang@tju.edu.cn



杨雁南(1993—),男,山西阳高人,硕士研究生,2016 年于武汉理工大学获得学士学位,主要研究方向为计算机视觉。E-mail: 1191547337@qq.com