

基于 Grad-CAM 与 KL 损失的 SSD 目标检测算法

侯庆山, 邢进生

(山西师范大学数学与计算机科学学院, 山西临汾 041004)

摘要: 鉴于 Single Shot Multibox Detector (SSD) 算法对中小目标检测时会出现漏检甚至错检的情况, 提出一种改进的 SSD 目标检测算法, 以提高中小目标检测的准确性. 运用 Gradient-weighted Class Activation Mapping (Grad-CAM) 技术对检测过程中的细节作可视化处理, 并以类激活图的形式呈现各检测层细节, 分析各检测层的类激活图发现 SSD 算法中待检测目标的错检以及中小目标的漏检现象与回归损失函数相关. 据此, 采用 Kullback-Leibler (KL) 边框回归损失策略, 利用 Non Maximum Suppression (NMS) 算法输出最终预测框. 实验结果表明, 改进算法相较于已有检测算法具有更高的准确率以及稳定性.

关键词: 目标检测; 可视化; 类激活图; Grad-CAM; SSD; KL 损失

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2020)12-2409-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.12.017

SSD Object Detection Algorithm Based on KL Loss and Grad-CAM

HOU Qing-shan, XING Jin-sheng

(School of Mathematics and Computer Science, Shanxi Normal University, Linfen, Shanxi 041004, China)

Abstract: Considering that the single shot multibox detector (SSD) algorithm will be missed or even false when it is used to detect the small and medium-sized objects, an improved SSD object detection algorithm is proposed to improve the accuracy of small and medium-sized objects detection. The details in the detection process are visualized with gradient-weighted class activation mapping (Grad-CAM) technology, and the details of each detection layer are shown in the form of class activation maps. Then it is noted that the phenomenon of the false or missed detection of the objects to be detected on small and medium-sized objects in the SSD algorithm is related to the regression loss function. Accordingly, Kullback-Leibler (KL) border regression loss strategy is adopted and non maximum suppression (NMS) algorithm is used to output the final prediction boxes. Experimental results show that compared with the existing detection algorithms, the improved algorithm in this paper has higher accuracy and stability.

Key words: object detection; visualization; class activation maps; Grad-CAM; SSD; KL loss

1 引言

目标检测技术的不断完善和发展, 引起智能化交通系统^[1]、智能监控系统^[2]、军事目标检测^[3]及医学导航手术中手术器械定位^[4]等领域的广泛关注, 是计算机视觉领域的核心技术之一. 在国内外相关研究机构的努力下, 高效的目标检测算法如雨后春笋般层出不穷, 如基于单阶段过程的 You Only Look Once (YOLO) 算法^[5]、SSD 算法^[6]、基于双阶段过程的 Region with CNN features (R-CNN) 算法^[7]等.

针对 SSD 算法在中小目标检测中漏检甚至错检的

问题, Wen, Zhan^[8]等人改进原 SSD 算法的激活函数, 并引入 Atrous 滤波器. Wu^[9]等人在传统的 SSD 检测算法中加入转置卷积结构, 将低分辨率、高语义与高分辨率、低语义的输出特征图进行融合. 针对图像中某些待检测目标被遮挡的情况, Xing^[10]等人根据检测目标的分布规律调整了输出特征层的缩放系数, 一定程度上解决了漏检问题. Tang^[11]等人提出了多视窗的方法, 但由于该方法存在目标被拆分检测等问题, 使得其准确性及稳定性受到影响. 针对 SSD 模型的低特征层, Chen^[12]等人采用区域放大提取的方法提升了算法的中小目标检测能力. Fu^[13]等人提出 Deconvolutional Single Shot

Detector (DSSD)模型,但其网络结构也随之变得更复杂,导致模型的检测实时性下降. Jeong^[14]等人提出的 Rainbow Single Shot Multibox Detector (RSSD)模型改进了特征融合的方式,对各输出层特征更加充分的利用,提升算法对中小目标的检测效果. Li^[15]等人通过特征融合与下采样的方法获取更多输出特征层细节信息,从而提出 Feature fusion Single Shot multibox Detector (FSSD)目标检测模型.

为了方便对 SSD 算法在中小目标检测方面性能的提高,本文引入可视化技术^[16]对原算法的检测过程进行分析,优化了原算法的损失函数. 改进后的检测算法对中小目标的检测能力更强,总体检测效果更好,并且目标漏检、错检问题也得到相应改善.

2 相关研究

2.1 SSD 模型

SSD 算法使用多尺度特征层进行目标检测,是一种直接预测目标类别和边界框的多目标检测算法. 与双阶段过程的目标检测算法相比 SSD 算法将所有计算在单阶段过程中进行,没有候选框区域的生成和后续特征重采样阶段的过程,提升了 SSD 算法的检测速度. 该算法以 Visual Geometry Group 16 (VGG-16)^[17]为基础网络,中上部分采用采样金字塔结构^[18],将最后两个全连接层替换为卷积层,并在其后增加四个卷积层.

2.2 KL 损失函数

现有的目标检测算法在大规模数据集上进行某些边界框标定时存在歧义. KL 损失^[19]针对边界框的移动与位置间的方差进行学习,在同等计算量下提升不同结构定位的准确率. KL 边界框回归损失可以获取到数据集中的模糊,从而降低预测边界框的回归损失.

基于对 Faster R-CNN^[20], Mask R-CNN^[21]检测网络的认知,KL 损失将原本的回归框参数由 (x, y, w, h) 替换为 (x_1, y_1, x_2, y_2) , 这样便于在每个维度上单独优化,预测边界框与真实边界框的偏差可由式(1)表达.

$$\begin{aligned} t_{x_1} &= \frac{x_1 - x_{1a}}{w_a}, & t_{x_2} &= \frac{x_2 - x_{2a}}{w_a} \\ t_{y_1} &= \frac{y_1 - y_{1a}}{h_a}, & t_{y_2} &= \frac{y_2 - y_{2a}}{h_a} \\ t_{x_1}^* &= \frac{x_1^* - x_{1a}}{w_a}, & t_{x_2}^* &= \frac{x_2^* - x_{2a}}{w_a} \\ t_{y_1}^* &= \frac{y_1^* - y_{1a}}{h_a}, & t_{y_2}^* &= \frac{y_2^* - y_{2a}}{h_a} \end{aligned} \quad (1)$$

(x_1, y_1, x_2, y_2) 表示预测边界框左上角和右下角的坐标, $(x_1^*, y_1^*, x_2^*, y_2^*)$ 表示真实边界框左上角以及右

下角的坐标. $(x_{1a}, x_{2a}, y_{1a}, y_{2a}, w_a, h_a)$ 是由真实边界框聚类产生的锚框.

KL 损失重点评估位置的置信度,在预测位置的同时给出置信度. 假设坐标间互相独立并使用单变量高斯模型,将边界框坐标表示为 x , 单独对每个坐标进行独立优化,可由式(2)表达.

$$P_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_e)^2}{2\sigma^2}} \quad (2)$$

式(2)中 θ 表示一组要学习的参数, x_e 表示预测边界框的位置, $x_e \in (x_1, y_1, x_2, y_2)$, σ 表示位置估计的置信度.

真实边界框也可以表示为 $\sigma \rightarrow 0$ 的高斯分布,如式(3)所示. 式(3)中的 x_g 表示真实边界框的位置.

$$P_D(x) = \delta(x - x_g) \quad (3)$$

选取 N 个样本,将这些样本的预测边界框分布 $P_{\theta}(x)$ 和真实边界框分布 $P_D(x)$ 之间的误差最小化,进而对 $\hat{\theta}$ 进行评估,获取位置定位的目标,可由式(4)表达.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum D_{\text{KL}}(P_D(x) \| P_{\theta}(x)) \quad (4)$$

基于 KL 损失的检测算法中,边界框回归的损失函数采用 KL 损失,对于单个样本的边界框回归损失由式(5)定义.

$$\begin{aligned} L_{\text{reg}} &= D_{\text{KL}}(P_D(x) \| P_{\theta}(x)) \\ &= \int P_D(x) \log P_D(x) dx - \int P_D(x) \log P_{\theta}(x) dx \\ &= \int P_D(x) \log P_D(x) dx - \int P_D(x) \log P_{\theta}(x) dx \\ &= \frac{(x_g - x_e)^2}{2\sigma^2} + \frac{\log(\sigma^2)}{2} + \frac{\log(2\pi)}{2} - H(P_D(x)) \end{aligned} \quad (5)$$

x_e 表示预测边界框的位置,当预测边界框的坐标位置不准确时,网络通过预测更大的方差 σ^2 ,使回归损失降低,由于式(5)中的后两部分与参数 θ 无关,得式(6).

$$L_{\text{reg}} \propto \frac{(x_g - x_e)^2}{2\sigma^2} + \frac{\log(\sigma^2)}{2} \quad (6)$$

当 $\sigma = 1$ 时,损失变为欧氏距离,如式(7).

$$L_{\text{reg}} \propto \frac{(x_g - x_e)^2}{2} \quad (7)$$

对式(6)中预测边界框的位置估计和定位标准偏差 σ 分别求偏导,如式(8)、(9)所示.

$$\frac{d}{dx_e} L_{\text{reg}} = \frac{x_e - x_g}{\sigma^2} \quad (8)$$

$$\frac{d}{d\sigma} L_{\text{reg}} = -\frac{(x_e - x_g)}{\sigma^3} - \frac{1}{\sigma} \quad (9)$$

为了防止在训练开始时发生梯度爆炸,令 $\alpha = \log(\sigma^2)$,在进行测试时,将 α 重新变回为 σ ,变换后的损失可由式(10)表达.

$$L_{\text{reg}} \propto \frac{e^{-\alpha}}{2} (x_g - x_e)^2 + \frac{1}{2} \alpha \quad (10)$$

用随机高斯分布对全连接层初始化,KL 损失在开始训练时类似于 smooth L_1 损失,当 $|x_g - x_e| > 1$ 时,损失可定义为式(11).

$$L_{\text{reg}} = e^{-\alpha} \left(|x_g - x_e| - \frac{1}{2} \right) + \frac{1}{2} \alpha \quad (11)$$

2.3 类激活图

为解决卷积神经网络模型中的不可见问题,通过 Grad-CAM 技术^[22]可将网络模型对输入图像中感兴趣的区域表示出来,将网络模型可视化,直观的表达了算法得出分类结果的原因.在 Grad-CAM 技术中,利用类别输出结果对卷积层的输出特征图求导获取特征激活图,由式(12)计算得到权重值 ω_k^c .将 ω_k^c 与特征图累加得到可视化结果,权重 ω_k^c 说明了类别 c 对特征映射的重要度.

$$\omega_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (12)$$

通过 ReLU 函数除去特征图上负值对分类结果的影响,最终的分类任务由式(13)、(14)表达.

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \omega_k^c A_{ij}^k \right) \quad (13)$$

$$y^c = \frac{1}{Z} \sum_i \sum_j L_{\text{Grad-CAM}}^c \quad (14)$$

3 改进框架设计

3.1 SSD 检测算法的可视化

Grad-CAM 技术根据输入图像中各目标的类别标签判别目标区域.将该技术运用到 SSD 检测算法,使该算法的执行过程可视化,针对用于目标检测的各卷积层,对这些层的特征输出层进行分析,构建各检测层的类激活图.

SSD 检测算法的可视化,能够更好的理解网络的目标检测过程,通过分析类激活图,得出中小目标检测不准确或漏检的原因,针对该原因做出相应的改进策略,提高检测效率.SSD 300 目标检测算法的可视化过程如图 1 所示.

3.2 基于 KL 损失的 SSD 算法

3.2.1 默认框设定

检测算法沿用了原 SSD 算法中默认框的设定方法,针对多个目标检测层的输出特征图进行设定,对于特征图中的每个像素点生成大小不同的两个正方形及多个长方形默认框.比例值 a_r 的数目决定了特征输出图中某一像素点下默认框的数目,默认框的大小取决于

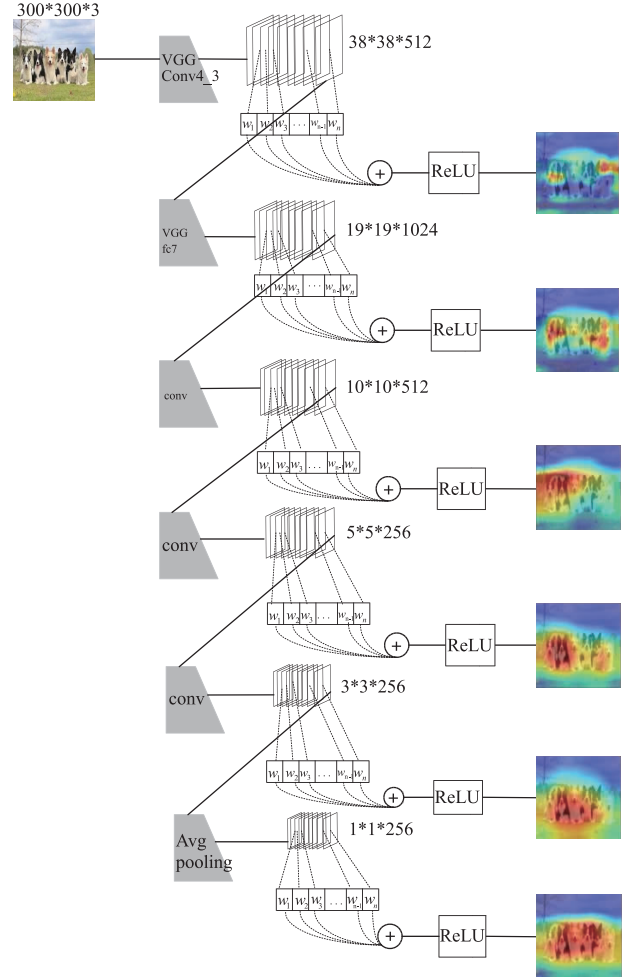


图1 SSD 300检测框架的可视化

S_k 值的设定,默认框的长和宽可由式(15)、(16)进行表达.

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1} (k - 1), k \in [1, m] \quad (15)$$

其中 m 为特征图的个数, $S_{\min} = 0.2$, $S_{\max} = 0.9$.

$$\text{width} = S_k \sqrt{a_r}, \text{height} = \frac{S_k}{\sqrt{a_r}} \quad (16)$$

其中默认框长宽比 $a_r \in (1, 2, 3, 1/2, 1/3)$,另外设置一个 $a_r = 1$, $S'_k = \sqrt{S_k \cdot S_{k+1}}$ 的默认框.默认框中心点的位置可由式(17)计算.

$$\left(\frac{i + 0.5}{|f_k|}, \frac{j + 0.5}{|f_k|} \right) \quad (17)$$

其中 $|f_k|$ 为第 k 个特征图的大小, $i, j \in [0, |f_k|]$.

3.2.2 正负样本标定

利用检测算法对输入图像中的目标进行检测时,需要对默认边界框进行类别标定.通过 IoU 值进行匹配, IoU 值由式(18)定义.对于每一个待检测目标的真实边界框,对应默认边界框的匹配规则如下:设定 IoU 的阈值为 0.5,当默认边界框与真实边界框的 IoU 值大

于 0.5 时,将此默认边界框标定为正例样本,否则标定为负例样本.

$$\text{IoU}(\text{Area}_{\text{db}}, \text{Area}_{\text{gt}}) = \frac{|\text{Area}_{\text{db}} \cap \text{Area}_{\text{gt}}|}{|\text{Area}_{\text{db}} \cup \text{Area}_{\text{gt}}|} \quad (18)$$

其中 $\text{Area}_{\text{db}}, \text{Area}_{\text{gt}}$ 分别表示默认边界框与真实边界框的面积.

3.2.3 损失策略

改进的 SSD 目标检测算法的训练损失由目标分类损失和预测边界框的回归损失两部分组成,改进算法训练损失可由式(19)表达.

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (19)$$

式中 N 为正例样本数, α 表示用于控制分类损失和回归损失的比重.

目标分类损失采用交叉熵损失^[23], 损失由式(20)表达.

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \quad (20)$$

x_{ij}^p 表示的是第 i 个默认边界框与第 j 个真实边界框关于目标类别 p 是否匹配, 当 $x_{ij}^p = 0$ 时不匹配, $x_{ij}^p = 1$ 时匹配.

预测边界框的回归损失采用 KL 损失, 回归损失由式(21)定义.

$$L_{\text{loc}}(x, p, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{cx_1, cy_1, cx_2, cy_2, s, \sigma^x, \sigma^y, \sigma^x, \sigma^y\}} x_{ij}^k \text{KL}_{\text{loss}}(p_i^m - \hat{g}_i^m) \quad (21)$$

$$\hat{g}_j^{cx_1} = (g_j^{cx_1} - d_i^{cx_1}) / d_i^w, \hat{g}_j^{cy_1} = (g_j^{cy_1} - d_i^{cy_1}) / d_i^h$$

$$\hat{g}_j^{cx_2} = (g_j^{cx_2} - d_i^{cx_2}) / d_i^w, \hat{g}_j^{cy_2} = (g_j^{cy_2} - d_i^{cy_2}) / d_i^h$$

$$\hat{p}_j^{cx_1} = (p_j^{cx_1} - d_i^{cx_1}) / d_i^w, \hat{p}_j^{cy_1} = (p_j^{cy_1} - d_i^{cy_1}) / d_i^h$$

$$\hat{p}_j^{cx_2} = (p_j^{cx_2} - d_i^{cx_2}) / d_i^w, \hat{p}_j^{cy_2} = (p_j^{cy_2} - d_i^{cy_2}) / d_i^h$$

式中 p, g 分别表示预测边界框和真实边界框与默认边界框之间的偏移量, 在边界框的回归损失中, 通常期望 p 和 g 差距较小.

对经过 NMS 算法处理后所选中的边界框位置投票, 选择最大得分边界框 $M\{x_1, y_1, x_2, y_2, s, \sigma^{x1}, \sigma^{y1}, \sigma^{x2}, \sigma^{y2}\}$, 根据自身和相邻边界框的参数更新坐标位置, 邻近边界框权重值 P_i 和新坐标的计算式由式(22)定义.

$$p_i = e^{-(1 - \text{IoU}(b, b)) / \sigma_i}$$

$$x = \frac{\sum_i p_i x_i / \sigma_{x,i}^2}{\sum_i p_i / \sigma_{x,i}^2} \quad (22)$$

式中 σ_i 表示变量投票表决时的可调参数.

4 图像数据预处理

为增强检测算法的通用性和稳定性, 采取数据增广的方法对数据预处理. 通常将图像目标区域分割掩

码中像素数量小于 1024 的目标定义为小目标, 大于 1024 且小于 9216 的目标定义为中等目标. 为此, 针对改进的 SSD 算法制定如下数据增广规则:

(a) 对原始的输入图像进行随机位置的采样, 采样图像与原始图像的比例取 1/64、1/32、1/16、1/8、1/4、1/2、1.

(b) 对采样图像以一定的概率作翻转和裁剪处理.

(c) 对采样图像以一定的概率添加色彩变换.

与原 SSD 算法的采样比例相比, 更小的采样比例可以使得由随机采样得到的小目标经过图像尺寸恢复后被更明显的表示出来.

5 实验设计

5.1 实验数据集说明

以 PASCAL VOC 和 MS COCO 2017 三种数据集为基础, 设计相关实验数据集, 数据集的说明如下:

选取 MS COCO 数据集中的 30 类目标作为检测类别. 分别从 MS COCO 和 PASCAL VOC 数据集的训练集中选取 3000 张和 2000 张带有检测类别的图像, 将选取的 5000 张图像作为训练集; 从 MS COCO 数据集的验证集中选取 1000 张图像作为验证集, 用于调控模型的超参数, 监测过拟合现象; 分别从 MS COCO 和 PASCAL VOC 数据集的测试集中选取 800 张图像作为检测算法的测试集.

5.2 实验相关过程描述

基于设计好的实验数据集, 对改进的 SSD 目标检测算法进行训练测试. 改进的 SSD 目标检测算法采用了随机梯度下降的方法, 对改进算法的初始学习率、动量、权重大小、批量等参数进行设定. 检测算法在训练数据时将初始学习率设定为 0.001, 对训练集中的图像进行 30000 次的迭代训练; 进一步将学习率设定为 0.0001, 对训练集中的图像进行 10000 次的迭代训练; 最后将学习率下降至 0.00005, 对训练集中的图像进行 10000 次的迭代训练. 此外, 设置 IoU 阈值为 0.5, 利用验证集对超参数进行调节, 监测过拟合现象.

利用测试集对训练好的改进目标检测算法进行测试, 与原算法的 mPA 及 fps 值作对比, 体现改进检测算法的优势, 验证算法改进的合理性和有效性.

5.3 实验结果分析

5.3.1 各检测层结果对比

检测算法的低检测层对小目标特征进行提取, 对原 SSD 目标检测算法中各检测层的检测效果进行测试, 与原检测算法相比, 改进算法对小目标检测的准确度明显提升, 检测精度提升约 12%. 计算原算法和改进算法各检测层的 mPA 值和 fps 值, 如表 1 所示.

表 1 目标检测算法各检测层 mPA 值和 fps 值对比

用于目标检测的特征层	传统 SSD mPA/%	改进 SSD mPA/%	传统 SSD fps	改进 SSD fps
Conv4_3	65.4	76.3	56	54
fc7	63.7	74.6	56	54
Conv8_2	64.2	75.2	56	54
Conv9_2	59.4	68.7	56	54
Conv10_2	61.5	70.8	56	54
Conv11_2	60.5	69.7	56	54

由表 1 数据可知,改进后的低特征层对小目标的检

表 2 SSD 相关改进算法 mPA 和 fps 对比

数据集	目标检测算法	训练 mAP/%	预测 mAP/%	fps
PASCAL VOC 2012	Wen, et al. 's method ^[8]	74.9	73.6	54
	Chen, et al. 's method ^[12]	73.5	72.7	58
	DSSD ^[13]	72.9	71.4	57
	RSSD ^[14]	75.1	73.8	58
	FSSD ^[15]	73.1	72.5	56
	Our SSD300	77.2	75.3	58
MS COCO	Wen, et al. 's method ^[8]	72.6	71.4	53
	Chen, et al. 's method ^[12]	71.5	70.6	56
	DSSD ^[13]	69.8	68.5	54
	RSSD ^[14]	73.1	72.4	56
	FSSD ^[15]	70.9	69.4	54
	Our SSD 300	74.3	73.8	55
实验设计数据集	Wen, et al. 's method ^[8]	73.7	72.6	52
	Chen, et al. 's method ^[12]	73.2	72.3	56
	DSSD ^[13]	72.6	71.4	55
	RSSD ^[14]	74.1	73.2	56
	FSSD ^[15]	72.9	71.8	55
	Our SSD 300	75.5	74.6	56

5.3.3 常见目标检测算法框架对比

对现有的一些目标检测算法在实验设计数据集上训练和预测,将训练和预测结果与改进的 SSD 目标检测算法做对比,对比结果如表 3 所示。

将改进算法训练过程的 mPA 值与 YOLO、YOLOv2 算法训练过程的 mPA 值对比,对比结果如图 2 所示。另外选取基于两阶段过程的 Fast R-CNN、Faster R-CNN 算法,训练过程的 mPA 值对比结果如图 3 所示。

对比改进算法与原 SSD 算法训练过程的 mPA 值,对比结果如图 4 所示。

与现有的目标检测算法对比,改进后的 SSD 目标检测算法检测精度有较明显的提高,算法的检测实时

测效果提升明显,能够收集到更丰富的语义信息。通过 fps 值的对比,改进后的 SSD 目标检测算法保持了良好的检测实时性,SSD 算法得到明显改进。

5.3.2 SSD 相关改进算法对比

对现有的一些 SSD 改进算法在相关数据集上训练和预测,将训练和预测结果与改进的 SSD 300 目标检测算法做对比,对比结果如表 2 所示。

分析表 2 可知,与选取的 SSD 改进算法相比,本文提出的改进算法检测精度有较明显的提高,算法的检测实时性得到较好保持。

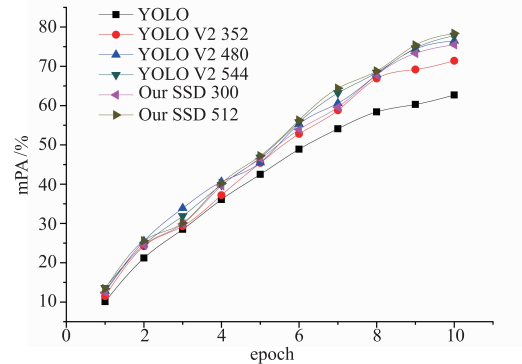


图2 改进算法与YOLO、YOLOv2算法mPA值对比
性较好。一方面,对原 SSD 目标检测算法的损失策略进

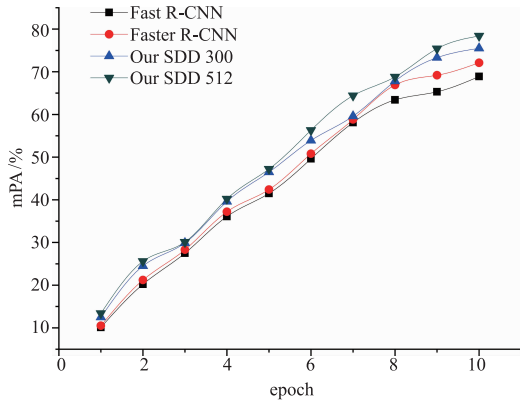


图3 改进算法与Fast R-CNN、Faster R-CNN算法mPA值对比

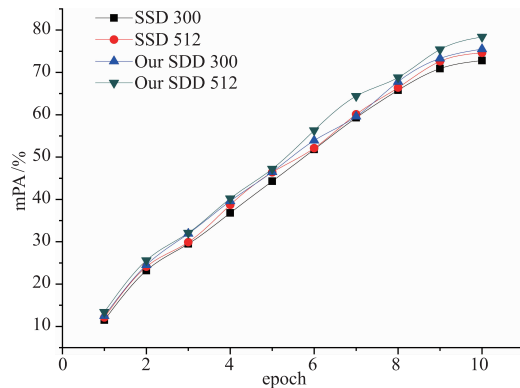


图4 改进算法与原算法mPA值对比

行改进,进一步缩小了检测算法训练时的损失;另一方面,针对图像中的中小目标做出相应的图像增广策略,

表3 目标检测算法 mPA 和 fps 对比

目标检测算法	训练 mAP/%	预测 mAP/%	fps
YOLO ^[5]	62.7	61.5	48
YOLOv2 352 ^[24]	71.4	70.6	84
YOLOv2 480 ^[24]	76.5	74.3	60
YOLOv2 544 ^[24]	77.8	75.1	43
SSD 300 ^[6]	72.8	71.2	58
SSD 512 ^[6]	74.6	73.8	41
FastR-CNN ^[25]	68.9	66.4	2
FasterR-CNN ^[20]	72.1	70.6	9
Ours 300	75.5	74.6	56
Ours 512	78.4	77.2	39

使得检测算法对中小目标特征的收集更加全面,提高了对输入图像上中小目标的检测精度.

5.3.4 效果图对比

将改进算法的检测效果图与原算法的检测效果图进行对比,对比效果图如图5、6所示.改进后的SSD目标检测算法对输入图像中的中小目标能够更加精准的检测和分类,对由于遮挡现象造成的目标漏检问题有良好改善.在复杂场景中,与原SSD检测算法相比,改进的SSD目标检测算法能够检测到更多的目标;在简单场景中,改进算法对待检测目标类型的判别更加精确,目标的正确分类得分更高.此外,在处理模糊场景及小目标场景时,改进算法能够检测到更多的模糊目标与小目标.

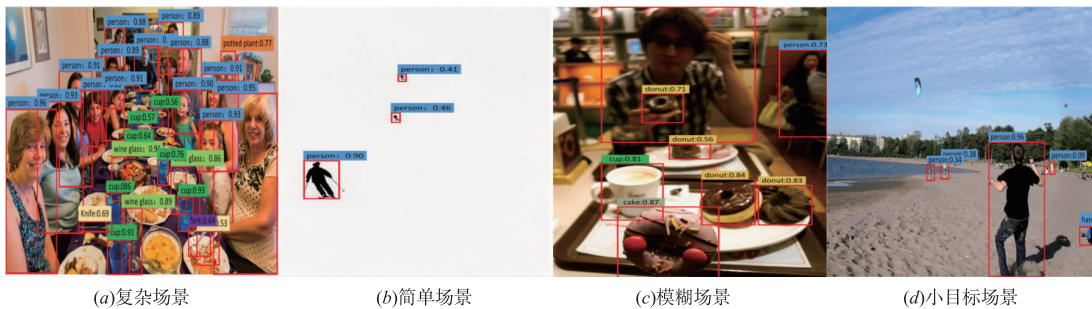


图5 原SSD目标检测算法的检测效果

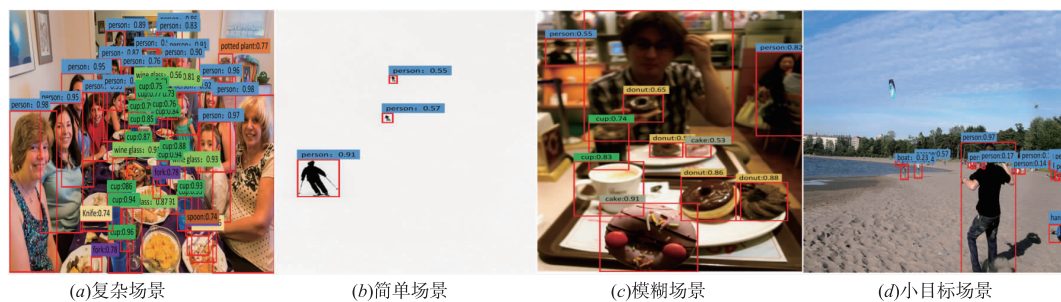


图6 改进后的SSD目标检测算法的检测效果

6 结束语

改进的 SSD 目标检测算法提高了原 SSD 目标检测算法的稳定性和准确率,改善了在遮挡情况下造成的目标错检现象,对中小目标的检测效果有明显的提升.改进算法的贡献主要体现在以下方面:

(1) 对 SSD 目标检测算法进行可视化,分析 SSD 目标检测算法的检测过程.

(2) 针对原算法的不足,改进了输入图像的增广策略.

(3) 基于对可视化结果的分析,优化了传统 SSD 目标检测算法的边界框回归损失.

(4) 利用改进的 NMS 算法输出目标的预测边界框.

参考文献

- [1] Gomaa Ahmed, Abdelwahab Moataz M, et al. Robust vehicle detection and counting algorithm employing a convolution neural network and optical flow[J]. *Sensors (Basel, Switzerland)*, 2019, 19(20):4588 – 4600.
- [2] Ju J, Xing J. Moving object detection based on smoothing three frame difference method fused with RPCA[J]. *Multimed Tools Appl*, 2019, 78(21):29937 – 29951.
- [3] You Zhou, Ting Rui, Yiran Li, Xuegang Zuo. A UAV patrol system using panoramic stitching and object detection[J]. *Computers and Electrical Engineering*, 2019, 80(17):106473 – 106481.
- [4] Shen Zhao, Xi Wu, Bo Chen, Shuo Li. Automatic spondylolisthesis grading from MRIs across modalities using faster adversarial recognition network[J]. *Medical Image Analysis*, 2019, 58(12):101533 – 101572.
- [5] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas:IEEE, 2016. 779 – 788.
- [6] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[A]. *Computer Vision-ECCV 2016-14th European Conference[C]*. Amsterdam:IEEE, 2016. 21 – 37.
- [7] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation[A]. 2014 IEEE Conference on Computer Vision and Pattern Recognition[C]. Columbus:IEEE, 2014. 580 – 587.
- [8] 温捷文, 战荫伟, 李楚宏. 一种加强 SSD 小目标检测能力的 Atrous 滤波器设计[J]. *计算机应用研究*, 2019, 36(3):861 – 865.
Wen J W, Zhan Y W, Li C H, et al. Design of Atrous filter to strengthen small object detection capability of SSD[J]. *The Pascal Application Research of Computers*, 2019, 36(3):861 – 865.
- [9] Wu Tianshu, Zhang Zhijia, Liu Yunpeng, et al. A lightweight small object detection algorithm based on improved SSD[J]. *Infrared and Laser Engineering*, 2018, 47(7):703005.
- [10] 邢浩强, 杜志岐, 苏波. 基于改进 SSD 行人检测方法[J]. *计算机工程*, 2018, 44(11):228 – 233.
Xing H Q, Du Z Q, Su B. Pedestrian detection method based on modified SSD[J]. *Computer Engineering*, 2018, 44(11):228 – 233. (in Chinese)
- [11] Tang C, Ling Y S, Zheng K D, et al. Object detection method of multi-view SSD based on deep learning[J]. *Infrared and Laser Engineering*, 2018, 47(1):302 – 310.
- [12] Chen H J, Wang Q Q, Yang G W, et al. SSD object detection algorithm with multi-scale convolution feature fusion[J]. *Journal of Frontiers of Computer Science and Technology*, 2019, 13(6):1049 – 1061.
- [13] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional Single Shot Detector[EB/OL]. <https://arxiv.org/abs/1701.06659>, 2017-01-23.
- [14] Jeong J, Park H, Kwak N. Enhancement of SSD by Concatenating Feature Maps for Object Detection[EB/OL]. <https://arxiv.org/abs/1705.09587>, 2017-05-26.
- [15] Li Z X, Zhou F Q. FSSD: Feature Fusion Single Shot Multibox Detector[EB/OL]. <https://arxiv.org/abs/1712.00960>, 2018-05-17.
- [16] Zeiler M D, Fergus R. Visualizing and understanding convolutional neural networks[A]. *Computer Vision-ECCV 2014-13th European Conference[C]*. Zurich:IEEE, 2014. 818 – 833.
- [17] K Simonyan, A Zisserman. Very Deep Convolutional Networks for Large-scale Image Recognition[EB/OL]. <https://arxiv.org/abs/1409.1556>, 2015-04-10.
- [18] Lin T Y, Dollár, Piotr, Girshick R, et al. Feature pyramid networks for object detection[A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition[C]. Honolulu:IEEE, 2017. 2117 – 2125.
- [19] Y He, C Zhu, J Wang, M Savvides, X Zhang. Bounding box regression with uncertainty for accurate object detection[A]. 2019 IEEE Conference on Computer Vision and Pattern Recognition[C]. Long Beach:IEEE, 2019. 2888 – 2897.
- [20] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 39(6):1137 – 1149.
- [21] K He, G Gkioxari, P Dollar, R Girshick. Mask R-CNN[EB/OL]. <https://arxiv.org/abs/1703.06870>, 2018-01-24.
- [22] Ramprasaath R, Abhishek D, Ramakrishna V, Michael C,

Devi P, Dhruv B. Grad-CAM: Why did you say that? visual explanations from deep networks via gradient-based localization [A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu: IEEE, 2017. 618 – 626.

- [23] Farahnak-Ghazani F, Baghshah M S. Multi-label classification with feature-aware implicit encoding and generalized cross-entropy loss [A]. 24th Iranian Conference on

Electrical Engineering [C]. Persia: IEEE, 2016. 7585772.

- [24] J Redmon, A Farhadi. YOLO9000: better, faster, stronger [A]. 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu: IEEE, 2017. 7263 – 7271.
- [25] Girshick R. Fast R-CNN [A]. 2015 IEEE International Conference on Computer Vision [C]. Santiago: IEEE, 2015. 1440 – 1448.

作者简介



侯庆山 男, 1995 年出生于山西阳泉, 现为山西师范大学数学与计算机科学学院在读硕士研究生. 主要研究方向: 计算机视觉.
E-mail: 18435202842@163.com



邢进生(通信作者) 男, 1964 年出生于山西太原, 现为山西师范大学数学与计算机科学学院教授、硕士生导师. 研究方向: 计算智能、数据挖掘、计算机视觉.
E-mail: xjs19640408@163.com