

基于改进 YOLOv 网络的外观检测研究

李莉, 黄承宁

(南京工业大学 浦江学院, 南京 210000)

摘要: 外观检测涉及对图像或视频中的物体进行准确和高效的识别和定位, 为了解决物体表面小尺寸目标检测的问题, 研究通过优化 YOLOv3 网络模型, 引入多尺度检测和深度可分离卷积技术来提高检测精度和模型效率, 以增强对小尺寸目标的识别能力, 再采用深度可分离卷积技术来减少计算量, 并提高模型的训练效果; 实验结果表明, 研究模型在物体表面小尺寸检测方面取得显著提升; 与其他金属表面损伤检测算法相比, 优化后的 YOLOv3 实现了 71.52% 的检测精度, 超越 Faster R-CNN 6.83%; 尽管 Faster R-CNN 在准确性方面优异但速度慢, SSD 速度较快但不及 YOLOv2; 而 YOLOv2 虽速度快但精度稍低; 相对于原始模型, 研究算法的平均精度提升了 7.77 个百分点, 达到了 79.21%; 虽然网络深度的提升稍增计算量, 略有检测速率下降, 但引入深度可分离卷积后, 检测速度达到 36.2 帧/秒, 仅较原模型稍低 2.4 帧/秒; 研究可以优化算法, 提高小尺寸目标检测的准确性和鲁棒性, 推动其在计算机视觉领域的广泛应用。

关键词: 外观检测; 深度学习; yolov; 多尺度融合; 聚类算法

Research on Appearance Detection Based on Improved YOLOv Network

LI Li, HUANG Chengning

(College of Pujiang, Nanjing University of Technology, Nanjing 210000, China)

Abstract: Appearance detection involves the accurate and efficient recognition and positioning of objects in images or videos. In order to solve the small-sized target detection on the surface of objects, the paper optimizes the YOLOv3 network model, introduces the multi-scale detection and depth separable convolution technology to improve the detection accuracy and efficiency of the model to enhance the recognition ability of small-sized targets, and then the depth separable convolution technology is used to reduce the amount of calculation and improve the training effect of the model. The experimental results show that the research model has achieved a significant improvement in the detection of small-scale objects on the surface. Compared with other metal surface damage detection algorithms, the optimized YOLOv3 achieves a detection accuracy of 71.52%, surpassing that of the Faster R-CNN by 6.83%. Although the Faster R-CNN is excellent in accuracy but slow in speed, the seed of SSD is faster but not as good as that of YOLOv2. While the speed of YOLOv2 is fast but low in accuracy. Compared with the accuracy of the original model, the average accuracy of the research algorithm has increased by 7.77%, reaching 79.21%. Although the increase in network depth slightly increases the amount of calculations and slightly reduces the detection rate, but the introduction of depth-separable convolution, the detection speed reaches 36.2 fps, which is only slightly lower than the original model by 2.4 fps. The optimized algorithm improves the accuracy and robustness of small-sized object detection, and promotes its wide application in the field of computer vision.

Keywords: appearance detection; deep learning; Yolov; multi-scale fusion; clustering algorithm

0 引言

在计算机视觉领域, 目标检测是一项重要的任务, 其在诸多应用中发挥着关键作用。然而, 针对物体表面小尺寸目标的准确检测一直是一个具有挑战性的问题, Faster R-CNN (Region-based Convolutional Neural Networks) 和 YOLOv3 (You Only Look Once) 是目前较为流行和广泛应用的两种目标检测模型。Faster R-CNN 是一种基于深度学习的目标检测算法, 它引入了候选区域提取网络和区域分类网络, 通过联合训练两者来实现目标检测^[1]。YOLOv3 是另一种颇具代表性的目标检测模型, 它将图像分割成网格, 并在每个网格单元中预测目标的类别和边界框, 并通

过多个尺度的特征图来检测不同大小的目标来增加检测的准确性^[2]。研究将探索对现有模型的网络结构进行调整和增强, 以更好地适应小尺寸目标的特点。比如包括增加或调整网络层, 引入更适合捕捉小尺寸目标特征的卷积结构等, 同时借鉴 YOLOv3 的多尺度特征图思想, 在模型中引入针对不同尺寸目标的多尺度特征图层, 最后采用深度可分离卷积技术, 将标准卷积操作拆分为深度卷积和逐点卷积。研究的创新点在于对网络结构进行调整和增强, 引入多尺度特征图层, 采用了深度可分离卷积技术, 将卷积操作拆分为深度卷积和逐点卷积, 从而减少计算量并加快模型的推理速度, 期望能够将这项技术应用于各种需要对小

收稿日期: 2023-07-25; 修回日期: 2023-09-04。

作者简介: 李莉 (1978-), 女, 硕士, 高级工程师。

黄承宁 (1985-), 男, 博士, 副教授。

引用格式: 李莉, 黄承宁. 基于改进 YOLOv 网络的外观检测研究[J]. 计算机测量与控制, 2024, 32(3): 92-98, 105.

尺寸目标进行准确检测的领域。

1 基于深度学习的目标外观特征提取方法

1.1 Faster R-CNN 算法

Faster R-CNN 算法是一种用于目标检测的深度学习算法。它的核心思想是通过集成候选区域操作、特征提取操作、分类器操作和定位操作, 利用一个统一的深度网络进行模型训练, 从而实现对目标的准确检测和定位^[3]。Faster R-CNN 算法结构如图 1 所示。

Faster R-CNN 算法的关键步骤分为 4 个部分, 首先是生成候选区域, 使用候选区域操作 (RPN, region proposal network) 来生成候选目标区域, RPN 通过在输入图像上滑动一个小窗口, 并预测窗口中是否包含目标。然后进行特征提取, 通过卷积神经网络来提取候选区域中的特征表示, 这些网络将候选区域映射为固定长度的特征向量, 用于后续的分类和定位操作^[4]。之后再提取目标, 使用分类器操作对提取的特征进行目标分类, 并使用定位操作对目标的位置进行回归, 分类器可以是全连接层或软最大化层, 用于判断目标类别的概率。最后, 再对损失函数进行训练, Faster R-CNN 通过定义合适的损失函数来优化模型, 损失函数包括目标分类损失、边界框回归损失等等, 通过反向传播算法和梯度下降优化方法, 训练网络参数以最小化损失函数^[5]。Faster R-CNN 算法的主要优势在于其端到端的训练过程, 通过共享特征提取操作, 可以提高检测的效率和准确性。此外, Faster R-CNN 还具备较强的泛化能力, 在不同的数据集和场景下都能表现出良好的检测结果。

1.2 YOLOv 检测模型

YOLOv1 是 YOLO 系列的第一代模型, 它将目标检测问题转化为一个回归问题, 通过将图像划分为网格单元, 每个单元预测固定数量的边界框和类别概率。YOLOv1 通过 CNN 提取特征, 并使用全连接层进行目标分类和边界框回归^[6]。YOLOv1 的思路是在输入目标图像后, 通过一次

前向传播, 直接回归目标边界框的位置信息和分类信息。网络的基础架构由 24 个卷积层和 2 个全连接层构成, 借鉴了 GoogleNet 网络的结构。不同于 Inception 模块, YOLOv1 使用了 1×1 卷积和 3×3 卷积替代。最终输出的检测值是一个 $7 \times 7 \times 30$ 的张量。假设输入图像被分成 $S \times S$ 个网格块, 每个网格块预测 M 个边界框, 每个边界框由坐标信息 (x, y, w, h) 和置信度 (Confidence) 组成。置信度用于判断该网格块中是否存在目标。置信度的计算公式如公式 (1) 所示:

$$confidence = PR(Object) \times IoU_{pred}^{truth} \quad (1)$$

如式 (1) 所示, 在每个网格中, 如果存在目标, 则将置信度的值设置为 1, 否则为 0。为了衡量边界框和真实框之间的重叠程度, 采用了交并比 (IoU, intersection over union) 作为关联值。此外, 每个网格还需要预测 N 个类别, 因此最终输出的张量的大小为 $S \times S \times (5 \times M \times N)$, 其中, S 表示网格的数量, M 表示每个网格预测的边界框数量, N 表示类别的数量。YOLOv2 是 YOLO 的第二代模型。它通过引入 Darknet-19 网络作为特征提取器, 并使用卷积层输出多尺度特征图。相比于 YOLOv1, 它引入批归一化层 (BN, batch normalization), 在 YOLOv2 中, 每个卷积层后面都加入了 BN 层^[7]。同时, BN 层还能加速训练模型的收敛速度。YOLOv2 直接将分类模型分为两个阶段, 在前 160 个 epoch 中, 输入图像分辨率为 224×224 ; 而后 10 个 epoch 中, 将分辨率提升到 448×448 。通过这种方式, 可以顺利过渡到检测模型, 使得模型能够更好地处理不同分辨率的输入图像。YOLOv2 还借鉴了 Faster R-CNN 中的 Anchor Box 思想, 引入 Anchor Box 后, 模型的检测精度可能会稍有损失, 但是召回率却有显著提高, 这使得模型更加注重整体检测质量的提升。YOLOv2 采用了 K-means 聚类算法来自动计算出更优的 Anchor Box 值。通过聚类算法, 可以得到多个合适的 Anchor Box 尺寸。为了评估聚类结果的质量, 还使用边界框之间的 IoU 值作为评判指标。为了

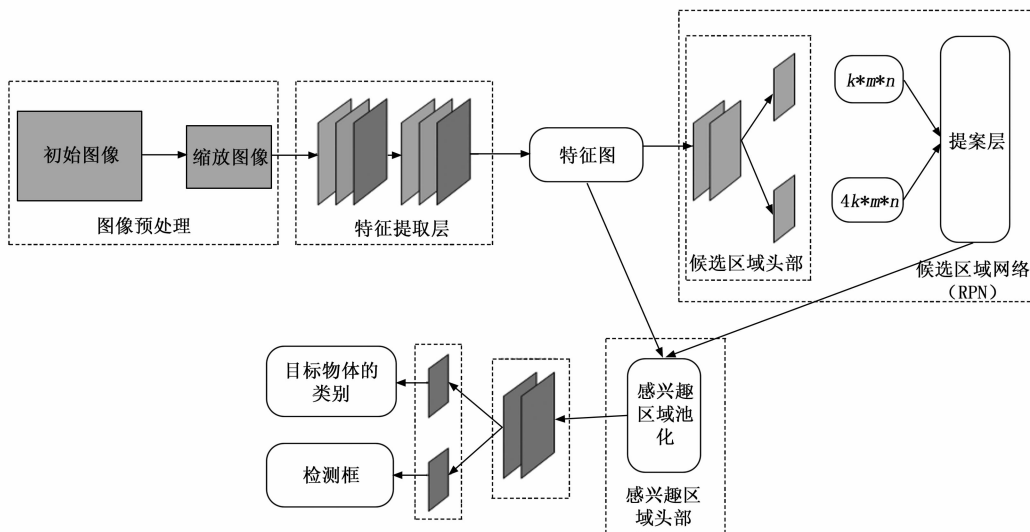


图 1 Faster R-CNN 结构图

提高模型的稳定性, YOLOv2 采用了直接位置预测的方法。在每个网格中, 预测参数包括 t_x 、 t_y 、 t_w 、 t_h 和 confidence。其中, t_x 和 t_y 经过 Sigmoid 函数的归一化处理。每个网格的宽度和高度为 c_x 和 c_y , Anchor Box 的宽度和高度为 p_w 和 p_h 。预测框的中心坐标为 b_x 、 b_y , 宽度和高度为 b_w 和 b_h 。通过计算可以得到每个参数的预测值, 采用直接位置预测的方法使得 YOLOv2 能够更准确地预测目标的位置和尺寸。计算公式如式 (2) 所示:

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

YOLOv2 引入多尺度检测可以增强训练出的模型的鲁棒性, 在训练过程中, 动态调整输入图像的尺寸, 并将连续的 10 个 epoch 分为一组进行批次训练, 这样做的好处是可以随机选择不同尺寸的图像, 使得 YOLOv2 能够检测不同大小的目标, 提高了模型的适应能力。YOLOv2 采用了 DarkNet-19 作为主干网络^[8]。DarkNet-19 网络由 19 个卷积层和 5 个池化层组成, 相较于 YOLOv1 的网络结构, DarkNet-19 减少了 5 层, 从而在一定程度上减少了计算量, DarkNet-19 网络具有较强的特征提取能力, 能够有效地捕捉目标的细节和上下文信息, 从而提升了 YOLOv2 的检测性能。

1.3 面向目标外观检测的识别精确度提升方法

在目标检测项目中, IoU 通常被用作评估目标检测的精度^[9]。 IoU 值实质上是真实框和检测框的交集与并集之间的比值。通过计算交集的面积除以并集的面积, 可以得到 IoU 值。 IoU 值的计算公式如公式 (3) 所示:

$$IoU = \frac{area(C) \cap area(G)}{area(C) \cup area(G)} \quad (3)$$

IoU 值的范围在 0 到 1 之间, 值越接近 1 表示检测结果与真实目标的重叠程度越高, 即检测精度越高。通常情况下, 当 IoU 值大于设定的阈值时, 将检测结果判定为正确检测, 否则判定为错误检测^[10]。第二种评估指标为 $P-R$ 曲线, 在 $P-R$ (Precision-Recall) 曲线的二维坐标图中, 横轴表示召回率 (Recall), 竖轴表示精确度 (Precision), $P-R$ 曲线的思路是在不同的阈值下, 计算对应的精确度和召回率, 每个阈值所对应的精确度和召回率可以看作是二维坐标中的一个点, 通过采样多个阈值, 可以得到多个点, 并将这些点连接在一起形成 $P-R$ 曲线。在计算 $P-R$ 曲线时, 首先根据不同的阈值对目标检测结果进行分类, 将检测框判定为正例或负例, 然后, 通过比较分类结果和真实标签, 计算出在每个阈值下的精确度和召回率。精确度表示被分类为正例的样本中真正为正例的比例, 而召回率表示真正为正例的样本被正确分类为正例的比例。通过取得足够密集的阈值, 可以获得多个精确度和召回率的点, 这些点可以在二维坐标上连接成一条 $P-R$ 曲线^[11]。 $P-R$ 曲线能够直观地反映出在不同阈值下精确度和召回率之间的关系。通

常情况下, $P-R$ 曲线越靠近图像的左上角, 说明模型的性能越好, 同时具有较高的精确度和召回率。计算精确度的公式如式 (4) 所示:

$$P = \frac{TP}{TP + FP} \quad (4)$$

式 (4) 中, TP 表示正确分类为正例的样本数量, 即模型将正例正确地判定为正例的数量; FP 表示错误分类为正例的样本数量。即模型将负例错误地判定为正例的数量; 召回率公式如式 (5) 所示:

$$R = \frac{TP}{TP + FN} \quad (5)$$

式 (5) 中, FN 代表错误分类为负例的样本数量。即模型将正例错误地判定为负例的数量。第三种评估指标为平均精度 (AP, average precision), 用于衡量各个类别的检测精度。 AP 的计算方法是对不同召回率点上的精确度进行平均, 在 $P-R$ 曲线上, 召回率通常位于横轴, 精确度位于纵轴。通过计算 $P-R$ 曲线下的面积, 可以得到该类别的平均精确度, AP 值越高, 表示模型在该类别上的检测性能越好。具体计算公式如式 (6) 所示:

$$AP = \int_0^1 P(R) dR \quad (6)$$

通过计算每个类别的 AP 值, 可以对目标检测模型在不同类别上的检测性能进行评估。 AP 值能够综合考虑精确度和召回率之间的权衡, 提供了对目标检测模型整体性能的评价。同时, 可以使用平均精度均值 (mAP, mean average precision) 来表示所有类别 AP 的平均值, 进一步评估整个网络的检测性能。 mAP 计算公式如式 (7) 所示:

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (7)$$

mAP 值的范围通常是 0 到 1 之间, 数值越高表示模型的检测性能越好。对于大规模的目标检测数据集, 常见的评估阈值通常是 0.5, 即当 IoU 大于 0.5 时, 将目标视为检测正确。当 mAP 值越接近 1 时, 表示模型在不同类别上的检测精度更高。除了检测精度, 每秒帧率 (FPS, frames per second) 是另一个重要的评价指标, 用于评估目标检测模型的检测速度。 FPS 表示每秒钟可以处理的图像帧数, 它反映了模型在单位时间内的处理能力^[12]。对于目标检测算法来说, 实时性是一个关键的考量因素, 尤其在需要进行实时监控、视频分析和自动驾驶等领域。检测速度越快, 模型能够在更短的时间内处理更多的图像帧, 从而实现实时检测的要求。

2 基于视觉算法的 YOLOv 网络改进策略

2.1 优化聚类算法下 YOLOv 网络算法的改进

聚类算法是一种在无先验知识的情况下, 通过发现数据对象之间的相似性来进行分组的方法。它利用距离计算来衡量数据对象之间的相似程度, 并根据相似性将数据划分为不同的簇。一个好的聚类结果应该是簇内相似性高、簇间差异大。YOLOv3 采用的 K-means 聚类算法是一种常

用的基于距离计算的聚类方法,它将距离作为衡量相似性的指标^[13]。具体操作是假设两个对象之间的距离越近,它们的相似性就越高。该算法的特点是将一组样本数据划分为 k 个不相交的簇,同一簇内的对象相似度较高,而不同簇之间的对象相似度较低。假设给定数据样本 X 包含 n 个对象,表示为 $X = \{X_1, X_2, \dots, X_n\}$,每个对象具有 m 个属性维度。K-means算法的目标是根据对象之间的相似性将它们聚集到指定的 k 个类簇中。算法首先初始化 k 个聚类中心 $\{C_1, C_2, \dots, C_k\}$,然后计算每个数据对象到每个聚类中心的欧几里得距离,距离计算公式如式(8)所示:

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2} \quad (8)$$

式(8)中 i 的取值范围是 $[1, n]$, t 表示属性的索引,取值范围是 $[1, m]$, j 的取值范围是 $[1, k]$ 。每个聚类中心也有 m 个属性。曼哈顿距离公式如式(9)所示:

$$dis(X_i, C_j) = \sum_{t=1}^m (|X_{it} - C_{jt}|) \quad (9)$$

曼哈顿距离是计算两个点之间的距离的一种度量方法。它基于点在坐标系中沿着轴的水平垂直移动的总距离。余弦距离公式如式(10)所示:

$$\cos\theta = \frac{\sum_{i=1}^m (X_{it} * C_{jt})}{\sqrt{\sum_{i=1}^m (X_{it})^2} * \sqrt{\sum_{i=1}^m (C_{jt})^2}} \quad (10)$$

在K-means算法中,为了确定每个类簇的中心点,需要计算每个维度上所有对象值的平均值,作为该维度上类簇中心的定义。平均值计算的公式如式(11)所示:

$$C_l = \frac{\sum_{x_i \in S_l} X_i}{|S_l|} \quad (11)$$

式(11)中, C_l 表示第 l 个类簇的中心, l 的取值范围为 $[1, k]$, $|S_l|$ 表示第 l 个类簇中的对象数量,即类簇中包含的对象的个数。 X 表示第一个类簇中的第 i 个对象,其中 i 的取值范围是 $[1, |S_l|]$,表示第 l 个类簇中的对象的编号。K-means聚类是一种迭代算法,通过计算数据点之间的距离和聚类中心之间的距离来对数据进行分组。算法通过不断地循环计算来优化聚类结果,直到达到预设的迭代次数。不同的距离度量方法对应不同的目标函数,当算法的迭代次数达到最大值时,循环计算停止并输出聚类结果,例如,当使用欧几里得距离时,目标函数通常是最小化每个数据点与其所属聚类中心之间距离的平方和。具体公式如式(12)所示:

$$\min \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2 \quad (12)$$

当使用余弦距离作为距离度量时,K-means聚类的目标函数通常是最大化每个数据点与其所属聚类中心之间余弦距离的和。具体如式(13)所示:

$$\max \sum_{i=1}^K \sum_{x \in C_i} \cos(c_i, x) \quad (13)$$

当簇之间的差异明显且簇内的数据点相似度较高时,

K-means算法可以得到最理想的聚类结果。K-means++算法是一种改进的K-means算法,它与K-means在计算过程上非常相似,但在初始聚类中心的选取上有所不同。K-means++通过一种策略来选择初始聚类中心,该策略旨在使初始聚类中心之间的距离较远,具体而言,它遍历整个数据集,并根据一定的概率分布选择下一个初始聚类中心,以确保选择的中心点之间的距离较大^[14]。通过这种方式,K-means++能够减少初始聚类中心选择对最终聚类结果的影响,K-means++的思路是先随机选择一个数据点作为第一个初始聚类中心,然后根据每个数据点与已选取的聚类中心的距离来计算选择下一个初始聚类中心的概率。距离较远的数据点更有可能被选为下一个初始聚类中心^[15]。重复这个过程,直到选择了所有的初始聚类中心。选择概率的公式如式(14)所示:

$$p = \frac{D(x)^2}{\sum_{i=1}^n D(x_i)^2} \quad (14)$$

K-means++算法通过改进初始聚类中心的选择,能够显著提高聚类结果的准确性,它的改进点相当简单直观,但非常有效。该算法的关键在于在选择初始聚类中心时,通过计算对象与已选中聚类中心的距离,选择距离较远的对象作为下一个聚类中心,这样可以确保初始聚类中心之间的距离较大,有利于聚类的精确性。

2.2 基于多尺度融合的目标外观检测优化

YOLOv3是在YOLOv2基础上改进和提升的目标检测算法,YOLOv3的基本思想可以分为两个部分,首先,根据一定规则在输入图像上生成一系列的候选框,这些候选框是可能包含目标的区域,通过与真实框进行标注,将其分为正样本和负样本,正样本是指与真实框完全重合的候选框,而负样本则是与真实框有一定偏离的候选框^[16]。其次,利用卷积神经网络对候选区域进行特征提取,并进行位置定位和类型识别,将候选区域输入到卷积神经网络中,获取与目标相关的特征表示,然后,通过位置定位和类型识别,最终得到检测结果。将这些检测结果与真实框的标签进行比较,判断是否正确检测出目标。YOLOv3在特征提取网络方面采用了DarkNet-53替代了YOLOv2中的DarkNet-19。DarkNet-53是一种全卷积网络结构,它由多个 1×1 和 3×3 的卷积层组成,每个卷积层后面跟着批归一化层和激活层。与YOLOv2不同的是,DarkNet-53网络中没有池化层和全连接层,而是通过步长为2的卷积进行下采样操作。经过5次下采样,特征图的尺寸缩小为原始图像的 $1/32$ 。DarkNet-53网络还引入了残差块结构(Residual Network, ResNet),在卷积层之间设置了快捷连接。这种结构有效地降低了训练深层网络的难度,使得网络能够更好地收敛。Darknet3网络结构图如图2所示。

在DarkNet-53网络中,通过将第2个残差块的浅层输出与经过2倍上采样后的网络深层输出进行融合,再经过 1×1 卷积层生成一个新的特征图层 104×104 ,这样的新增

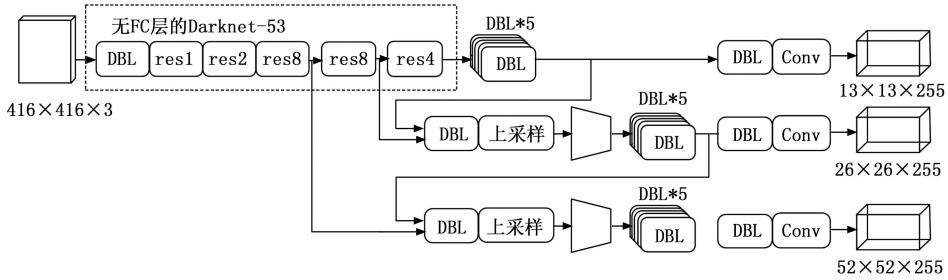


图 2 Darknet3 网络结构图

特征图层具有更强的特征提取能力。它的尺寸是输入图像大小的 1/4，将输入图像划分为更小的 4 × 4 网格，增强了网络对小目标的敏感性。该融合操作既继承了深层特征的信息，又充分利用了网络浅层特征，提高了模型对小目标特征的提取能力，降低了漏检小损伤目标的概率，从而提高了检测精度^[18]。网络中每个特征图层仍然使用 3 个 Anchor Box，但新增的特征图层使得总的 Anchor Box 数量从 9 个增加到 12 个，增强了检测密度。通过叠加多个特征图层，网络可以划分不同尺寸层次的目标，提高了对不同尺寸目标的综合检测能力^[19]。

在传统的标准卷积中，同一个卷积核被用于处理图像的所有通道。这意味着在进行卷积操作时，每个卷积核都负责处理所有通道的特征，如果需要检测更多的属性，就需要增加更多的卷积核。然而，深度可分离卷积引入了一种不同的思路，它将卷积操作分解为两个独立的计算过程。首先是深度卷积，它将每个通道单独进行卷积操作，而通道数保持不变。接下来是逐点卷积，它使用 1 × 1 × 3 的卷积核来扩充通道数。逐点卷积的输出结果与传统卷积相同^[20]。通过这种分解方式，研究可以灵活地增加更多的 1 × 1 × 3 卷积核来提取更多的特征。深度可分离卷积不仅减少了参数数量和计算量，还提高了模型的表达能力。因此，在 YOLOv3 中采用深度可分离卷积对模型进行优化，以提高检测性能和效率。深度可分离卷积运算量公式如式 (15) 所示：

$$Q_{dw} = w \times w \times d \times c \times c + w \times w \times d \times k = w \times w \times d \times (c \times c + k) \quad (15)$$

深度可分离卷积本质上是将 3D 卷积的通道分解，以降低计算成本。根据式 (15)，在这种卷积中，对于一个输入大小为 $w \times w \times d$ 的图像，使用 k 个大小为 $c \times c$ 的卷积核，步长为 1，填充为 1。与传统的卷积操作不同，深度可分离卷积首先进行深度卷积，将每个通道独立进行卷积操作。然后，再进行逐点卷积，利用 1 × 1 的卷积核进行通道数的扩展。比值 I 表示深度可分离卷积的运算量相对于普通卷积的运算量的减少程度，如公式 (16) 所示：

$$I = \frac{Q_{dw}}{Q} = \frac{1}{k} + \frac{1}{c \times c} \quad (16)$$

为了降低 YOLOv3 网络中的计算时间，研究采用了深度可分离卷积方法来改进残差块结构。这种方法通过在卷

积操作中减少参数数量，从而降低了计算复杂度，研究在残差块中引入了 1 × 1 卷积结构进行点卷积，并添加了 1 × 1 卷积层来平滑特征。这样的结构既能有效减少模型的计算量，又能够提取更多目标特征信息，从而提高了检测精度，通过这种优化，能够在保持准确性的同时提高算法的运行效率。

3 实验和结果分析

3.1 优化聚类算法的精确度验证

使用 K-means 算法和 K-means++ 算法对数据进行聚类后，研究可以通过绘制二维坐标图来展示它们的聚类结果。在图中，每个数据点表示一个样本对象，而不同的颜色表示不同的聚类簇。两种算法聚类精确度对比如图 3 所示。

根据图 3 所示，可以发现在 K-means++ 算法下，聚类结果展现出更高的识别精确度。这是因为 K-means++ 算法在选择初始聚类中心时更加巧妙，通过设置初始聚类中心之间的距离较远，有效地避免了初始聚类中心对聚类结果的影响。因此，K-means++ 算法能够更好地捕捉到数据的内在结构，使得聚类结果更加准确。

聚类中心个数为 9 时，对比了 K-means 算法和 K-means++ 算法生成的 Anchor Box 的宽度、高度和面积，详见表 1。将聚类结果分配给 3 个特征图层后，比较了每个尺寸特征图中 Anchor Box 的差异。可以观察到，K-means++ 算法生成的 Anchor Box 在不同尺寸特征图层之间的差异更大。

表 1 算法聚类差异

Anchor Box	特征图尺寸	32 × 32	16 × 16	8 × 8
平均宽度	K-means	33	53	127
	K-means++	29	86	118
平均高度	K-means	48	114	126
	K-means++	49	107	131
平均面积	K-means	1 471	5 364	15 987
	K-means++	1 444	5 892	17 062

根据表 1，可以明显地看到 K-means++ 算法生成的 Anchor Box 在不同尺寸特征图层之间的差异更大。这意味着 K-means++ 算法生成的 Anchor Box 在不同尺寸的目标检测上具有更好的适应性和灵活性。与此相比，K-means

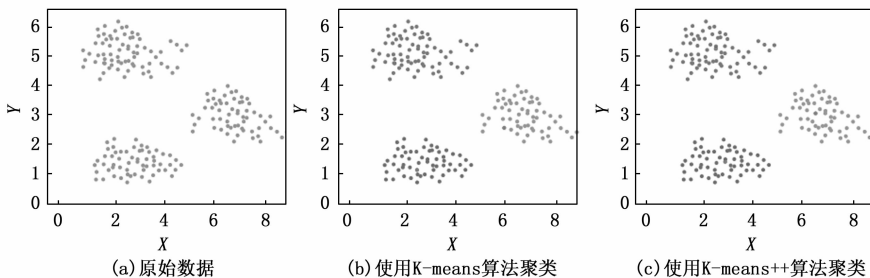


图 3 算法聚类精确度对比

算法生成的 Anchor Box 在不同尺寸特征图层之间的差异相对较小。因为不同尺寸的目标可能需要不同尺寸的 Anchor Box 来进行准确的检测。通过使用 K-means++ 算法生成的 Anchor Box, 能够更好地满足不同尺寸目标的检测需求, 从而提高目标检测算法的性能和准确性。因此, K-means++ 算法在 YOLOv3 中的优化中起到了关键作用。

为了进一步详细分析本章算法与 YOLOv3 在物体表面检测方面的性能, 如图 4 所示, 绘制出各类算法在 NEU-DET 数据集上部分损伤类型的 P-R 曲线。P-R 曲线的特性是区域的面积越大, 算法的检测能力越强。在图 4 中, 可以看到三条曲线, 其中第三条曲线下的面积最大, 第二条曲线下的面积最小。这意味着优化后的算法对于金属表面不同类型的损伤具有更强的检测能力, 第三条曲线所代表的损伤类型, 在不同召回率和精确率之间达到了最优的平衡, 即在保持高精确率的同时, 能够实现较高的召回率。相比之下, 第二条曲线所代表的损伤类型的检测能力较弱, 即使在较高的召回率下, 精确率相对较低。

3.2 多尺度融合的目标外观检测模型对比分析

为验证改进算法对金属表面损伤检测的有效性, 研究在 NEUDET 数据集的基础上进行了一系列实验, 将 Faster R-CNN、SSD、YOLOv2 以及本文算法进行了详细对比。为确保实验公平, 研究在相同的实验环境下进行了训练和评估。在实验设置中, 将模型的学习率初始值设定为 0.001, 并将训练进行至少 10 800 轮。为了更全面地评估算法性能, 研究采用了两项主要评估指标, FPS (每秒帧数) 和 *mAP* (平均精度均值)。为了保持一致性, 研究将本文算法中的 *DIoU* 阈值设置为 0.45, 而将其他算法模型中的 *IoU* 阈值设置为 0.5。同时, 我们保持了其他参数的一致

性。在实验结果方面, 对算法性能进行了对比, 并将对比结果列于表 2 中。

表 2 不同算法性能对比

网络模型	<i>mAP</i> /%	<i>FPS</i> /(f/s)
原 YOLOv3	60.15	36.3
Faster RCNN	64.69	13.0
改进 Faster RCNN ^[21]	65.26	34.7
YOLOv3 ^[22]	70.02	15.9
SSD	58.45	29.7
YOLOv2	52.02	39.2
优化后的 YOLOv3	71.52	35.4

在探究各深度学习算法在金属表面损伤检测中的效能时, 研究提出的优化后的 YOLOv3 算法以 71.52% 检测精度在五种模型中较为突出。具体而言, 与 Faster R-CNN 相比, 研究提出的算法实现了高达 6.83% 的精度优势, 虽然 Faster R-CNN 在准确性上优于其他 3 种深度学习算法, 但因其分阶段的检测方法, 导致检测时间较长, 从而成为速度最慢的选项。SSD 在检测速度方面明显超越 Faster R-CNN, 但相对于 YOLOv2 而言, 速度仍有所不及。YOLOv2 以一种单阶段算法的身份, 极大地提高了检测速度, 然而, 其相对较少的网络层数限制了其对有用目标特征信息的提取, 导致检测精度略有下降。此外, Mask R-CNN 改进算法在检测精度上表现出色, 但在速度上仍不如单阶段算法。同样地, 基于 YOLOv3 的改进算法尽管使用了类似的基础模型, 但其检测精度和速度均不及研究提出的算法, 这可以归因于本文算法引入了全新的特征图层, 并通过深度可分离卷积技术精简了模型的计算复杂度。研

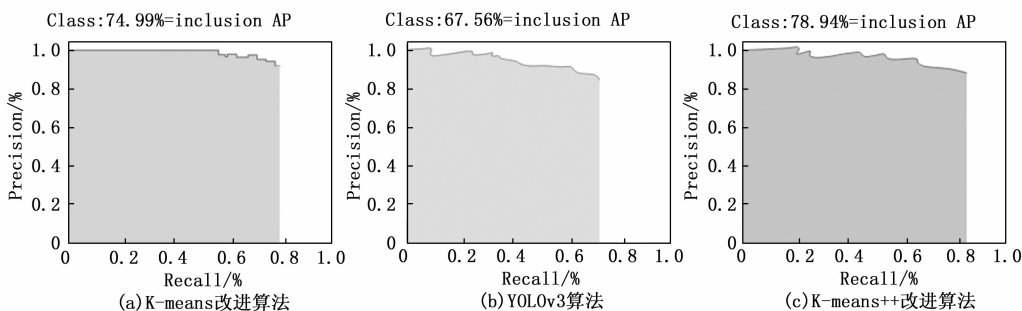


图 4 各类算法 P-R 曲线

究通过与 YOLOv3 算法的对比分析,从 mAP 和 FPS 两个方面评估了研究算法的性能,具体如表 3 所示。

表 3 多尺度融合后的算法与 YOLOv3 算法对比

网络模型	$mAP/\%$	$FPS/(f/s)$
原 YOLOv3	71.44	38.6
多尺寸融合优化后的 YOLOv3	79.21	36.2

根据表 3 的结果显示,研究算法的平均检测精度为 79.21%,相较于 YOLOv3 算法提高了 7.77 个百分点。这一改进主要体现在网络结构的优化,通过增加一个尺度更大的输出来提升检测能力,尤其是对小尺寸目标损伤的检测能力,减少了漏检的概率。另外,对比两种算法的 FPS ,由于研究算法增加了网络层数,模型参数增加,导致计算量增加,因此检测速度略有下降。然而,引入深度可分离卷积后,模型的计算量仅为原计算量的九分之一,大大提高了检测效率。具体来说,研究算法的检测速度为 36.2 f/s ,仅比 YOLOv3 算法低 2.4 f/s 。

图 5 呈现了各个损伤类型的检测精度对比结果。从图中数据可以观察到,改进后的网络结构在所有类别的检测精度上都有显著提升,特别是对于小尺寸类型目标的检测,改进后的网络结构展现出明显优势。如图 5 中,裂纹类别的检测精度提高了 6.35%,鳞片痕类别的检测精度提高了 6.40%。这表明新增的特征图层能够有效解决原网络结构在物体表面小尺寸目标的特征提取和检测方面的不足。因此,该改进方法对于小尺寸目标的优化是具有可行性的。综合来看,研究优化后的算法在检测性能上明显优于 YOLOv3 算法,适用于物体表面小尺寸检测。

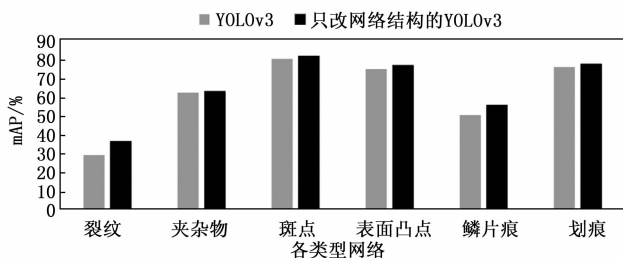


图 5 改进网络前后各类型网络的 mAP 对比

4 结束语

研究旨在解决物体目标检测中小尺寸目标的不敏感性和检测精度不足的问题。为此,研究对 YOLOv3 网络模型进行了结构上的改进和实验分析。首先,通过融合一个新的特征图层,实现了多尺度检测,以增强对小损伤目标的识别能力。其次,研究采用了深度可分离卷积技术来替代标准卷积,大幅降低了模型的计算量和大小。实验结果表明,经过一系列的改进,研究的模型在物体表面小尺寸检测方面取得了显著的性能提升,比较金属表面损伤检测算法,优化后的 YOLOv3 精度达 71.52%,高于 Faster R-CNN 6.83%。Faster R-CNN 准确但慢,SSD 快于它,但不

及 YOLOv2。YOLOv2 速度快但精度略低。Mask R-CNN 精度好但慢,基于 YOLOv3 的改进算法则综合优势。相对于原始模型,研究算法的平均精度达到了 79.21%,比原模型提高了 7.77 个百分点。尽管研究算法在网络深度上增加了计算量,导致模型的检测速率略有下降,但引入深度可分离卷积后,模型的检测速度明显提高,达到了 36.2 f/s ,仅比原模型略低 2.4 f/s 。研究仍存在一些不足之处,在算法的训练过程中,仍需注意如何平衡模型的精度和速度之间的权衡,以达到更好的检测效果和实时性能。

参考文献:

- [1] 范加利,田少兵,黄葵,等.基于 Faster R-CNN 的航母舰面多尺度目标检测算法[J].系统工程与电子技术,2022,44(1):40-46.
- [2] 胡根生,吴继甜,鲍文霞,等.基于改进 YOLOv5 网络的复杂背景图像中茶尺蠖检测[J].农业工程学报,2021,37(21):191-198.
- [3] XU X, ZHAO M, SHI P. Crack detection and comparison study based on faster R-CNN and mask R-CNN [J]. Sensors, 2022, 22(3): 1215-1231.
- [4] 王金,李颜娥,冯海林,等.基于改进的 Faster R-CNN 的小目标储粮害虫检测研究[J].中国粮油学报,2021,36(9):164-171.
- [5] 马佳良,陈斌,孙晓飞.基于改进的 Faster R-CNN 的通用目标检测框架[J].计算机应用,2021,41(9):2712-2719.
- [6] 宋谱怡,陈红,苟浩波.改进 YOLOv5s 的无人机目标检测算法[J].计算机工程与应用,2023,59(1):108-116.
- [7] 徐武,陈盈君,汤弘毅,等.局部特征映射与融合网络的人脸识别优化算法[J].河南科技大学学报(自然科学版),2023,44(2):59-64.
- [8] 邵延华,张铎,楚红雨,等.基于深度学习的 YOLO 目标检测综述[J].电子与信息学报,2022,44(10):3697-3708.
- [9] 张高青,吕晨,游吉,等.基于预测 IoU 的目标跟踪[J].机器人产业,2021(6):30-33.
- [10] 周丽芳,刘金兰,李伟生,等.基于 IoU 约束的孪生网络目标跟踪方法[J].北京航空航天大学学报,2022,48(8):1390-1398.
- [11] 陈兆凡,赵春阳,李博.一种改进 IoU 损失的边框回归损失函数[J].计算机应用研究,2020,37(s2):293-296.
- [12] 盛恒,黄铭,杨晶晶.基于 Faster R-CNN 和 IoU 优化的实验室人数统计与管理系统的[J].计算机应用,2019,39(6):1669-1674.
- [13] 罗鑫帅,高洋.基于改进型 K-means 算法的高校研究生成绩画像研究[J].陕西教育(高教),2023(7):49-51.
- [14] 戴响.基于大数据的 K-means 聚类算法的网络安全监测应用研究[J].信息记录材料,2023,24(4):140-142.
- [15] 赵春艳.基于改进 K-means 聚类分割方法的桥梁裂缝识别研究[J].黑龙江交通科技,2023,46(7):122-124.
- [16] 马纪颖,罗星,王一早等.基于改进 YOLOv3 的杂草检测研究[J].物联网技术,2023,13(6):11-16.

(下转第 105 页)