

基于改进YOLOv5的密集人群检测算法研究

袁谷长

天津职业技术师范大学理学院, 天津

收稿日期: 2024年9月23日; 录用日期: 2024年10月15日; 发布日期: 2024年10月25日

摘要

密集人群检测是计算机视觉领域的重要任务, 在监控、防踩踏、公共安全、交通监控等应用中具有广泛的应用价值。尽管基于深度学习的人群检测方法取得了显著进展, 但在处理密集人群、小目标等复杂场景时存在不足, 在涉及不同尺度的目标时无法准确地对其进行识别。本文提出了一种改进的YOLOv5密集人群检测算法, 采用多尺度检测和图像分割技术提升算法在复杂场景下的检测能力。实验数据表明改进后的算法在密集多尺度场景下的效果优于传统的YOLOv5和YOLOv10, 改进后的算法为密集人群检测任务提供了更全面有效的解决方案, 有望在实际应用中更好地满足公共安全领域、交通领域等对密集人群检测的需求, 对于预防踩踏等安全事故也具有重要意义。

关键词

密集人群检测, 深度学习, NMS, YOLOv5, YOLOv10, 分割, NWPU

Research on Dense Crowd Detection Algorithm Based on Improved YOLOv5

Guchang Yuan

School of Science, Tianjin University of Technology and Education, Tianjin

Received: Sep. 23rd, 2024; accepted: Oct. 15th, 2024; published: Oct. 25th, 2024

Abstract

Dense crowd detection is an important task in the field of computer vision and has extensive application value in applications such as monitoring, anti-stampede, public safety, and traffic monitoring. Although crowd detection methods based on deep learning have made remarkable progress, there are still deficiencies when dealing with complex scenes such as dense crowds and small targets, and they cannot accurately identify targets of different scales. This paper proposes a dense crowd detection algorithm based on improved YOLOv5, which uses multi-scale detection and image segmentation

technology to improve the detection ability of the algorithm in complex scenes. Experimental data shows that the improved algorithm is better than the traditional YOLOv5 and YOLOv10 in dense multi-scale scenes. The improved algorithm provides a more comprehensive and effective solution for dense crowd detection tasks, and is expected to better meet the needs of dense crowd detection in the fields of public safety and transportation in practical applications, and is also of great significance for preventing safety accidents such as stampedes.

Keywords

Dense Crowd Detection, Deep Learning, NMS, YOLOv5, YOLOv10, Segmentation, NWPU

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

密集人群检测是计算机视觉领域的一个重要任务，其目的是从图像或视频中准确地检测和定位人群。密集人群检测在许多应用中具有重要价值，例如智能交通、视频监控和自动驾驶等[1]。2015年，William Wang 和 Shaoqing Ren 提出了 SSD，该模型通过在卷积神经网络(CNN)的基础上引入辅助卷积层进行多尺度特征提取，并使用多个 default box (默认框)负责检测不同尺寸和长宽比的目标。同年，Joseph Redmon 等人首次提出了 YOLO [2]。2020年，Glen J. Braden 在 YOLOv4 [3]的基础上提出了 YOLOv5 检测模型。俞军[4]等在 YOLOv5 中设计了上下文提取模块来抑制多尺度特征融合的冲突，实现小目标人群检测。2024年，YOLO 系列的最新版本 YOLOv10 [5]进一步优化了检测精度和速度。虽然上述算法能在一定程度上改善小目标的检测效果，但仍不能聚焦密集人群目标区域。

2. 改进的 YOLOv5 算法

2.1. YOLOv5 网络结构

2020年 Ultralytics LLC 公司发布了 YOLOv5 算法。YOLOv5 的结构分为图像输入、骨干网络、颈部网络、输出端。输入端对图片数据进行一系列增强，例如图片缩放、Anchor 自适应计算等。骨干部分由 CBS、CSP 与空间金字塔(SPPF)组成。YOLOv5 的颈部结构根据特征金字塔结构和通道聚合结构设计而成。最后 Prediction 得到颈部部分传入的向量，输出目标的位置、类别以及置信度，图 1 为 YOLOv5 算法结构图。

2.2. 改进的 YOLOv5 算法

YOLOv5 算法在处理密集、小目标等复杂场景时，在多尺度检测的情况下对于小目标检测准确性欠佳且易出现漏检。为了解决多尺度问题，并在检测速度与精度之间找到平衡点，本文提出了一种改进的 YOLOv5 密集人群检测算法。改进的 YOLOv5 密集人群检测算法是一种自近及远的多尺度检测方法，具备感知不同尺度目标的能力，适用于各种复杂场景。图 2 是改进的 YOLOv5 算法的流程图。

首先进行初步检测，将输入图像直接送入 YOLOv5 网络进行粗略的检测，主要检测近距离的大目标人群。近距离人群的细节丰富，模型能够获取更多的纹理和形状信息。相比之下，远距离人群的细节模糊，在图像中占据的像素较少，导致纹理和形状信息丢失，影响识别精度。之后为了识别远距离人群，

进行初次分割检测, 将原图分割成 9 个子块。通过注意力分数筛选需要检测的子块, 式(1)是注意力分数的计算公式, 其中 edge_density (边缘密度)是指图像中边缘像素的总数与图像总像素数的比例, color_contrast (颜色对比度)是通过计算彩色图像 crop 在 RGB 三个通道上的方差来计算的。

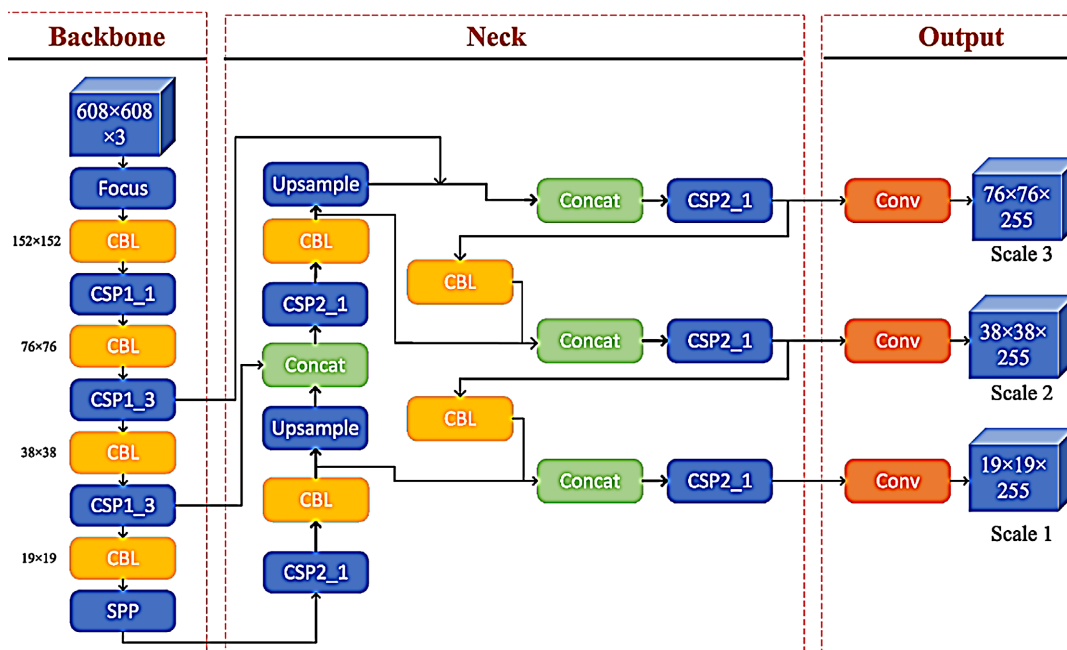


Figure 1. Structural diagram of YOLOv5 algorithm

图 1. YOLOv5 算法结构图

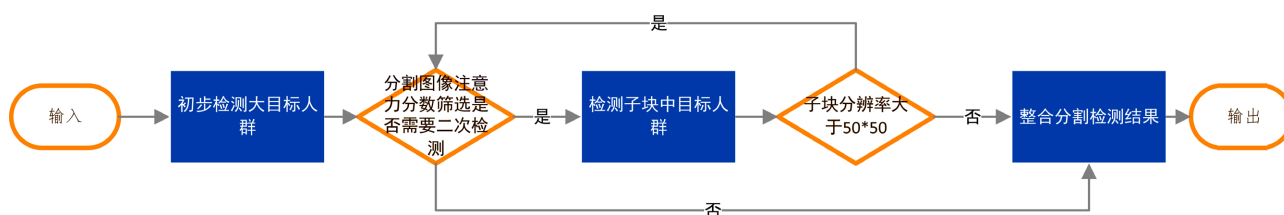


Figure 2. Flow chart of the improved YOLOv5 algorithm

图 2. 改进的 YOLOv5 算法流程图

$$\text{score} = 0.1 * \text{edge_density} + 0.9 * \text{color_contrast} \quad (1)$$

图 3 是注意力分数计算的策略。将选出来的子块送入 YOLOv5 网络进行检测。通过这种方式, 可以检测到中等大小的人群。接着使用非极大值抑制(NMS)算法去除重复的检测框, 以确保每个目标仅保留一个检测框。这个时候当子块的分辨率大于 $50 * 50$ 则对之前分割得到的每个子块再次进行分割, 将每个子块分割成 9 个更小的子块。同样通过注意力分数筛选需要检测的子块, 将选出来的子块再送入 YOLOv5 网络中进行检测。这一阶段主要针对更小的目标进行检测, 再次使用 NMS 算法去除重复的检测框, 以确保小目标检测的准确性, 一直到子块的分辨率小于 $50 * 50$ 停止分割。最后将所有分割检测的结果还原到初始图像, 得到最终的检测结果。通过对分割块检测结果的整合, 可以全面覆盖图像中的各种目标, 提高整体检测效果。

通过上述改进方法, YOLOv5 在处理密集、小目标等复杂场景时, 能够显著提升检测效果。改进后的 YOLOv5 算法不仅在正常场景下表现优异, 还能够有效应对复杂场景中的小目标和密集人群的检测需

求，为密集人群检测提供了更为全面和准确的解决方案。

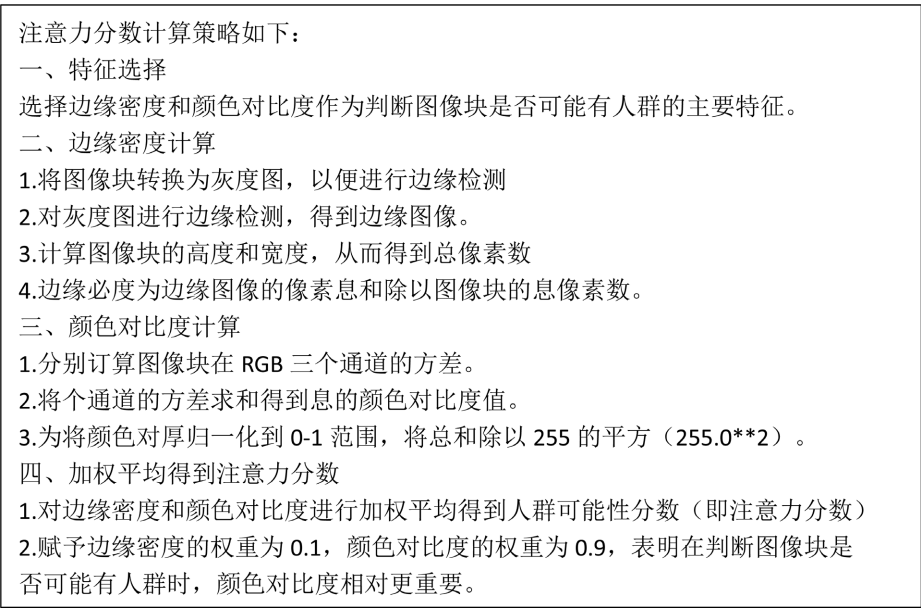


Figure 3. Calculation strategy of attention score
图 3. 注意力分数计算策略

3. 实验与分析

3.1. 实验数据集

NWPU [6]数据集全称为西北工业大学(Northwestern Polytechnical University)数据集，是在计算机视觉领域中广泛使用的一个数据集。NWPU 数据集通常包含大量的图像，涵盖了各种各样的场景和对象类别。NWPU 数据集在目标检测、图像分类、场景理解等计算机视觉任务中被广泛应用。

3.2. 实验过程

在 coco128 数据集的基础上只保留类别为 0 (person)的这一类，并调整超参数进行迭代次数为 90 次的训练，同时记录训练集与验证集上面的 3 种 loss，使用的评价指标为 precision，recall，以及 mAP_0.5，以及 mAP_0.5:0.95，图 4 是 3 种 loss 与评价指标的曲线图，横坐标的含义是迭代次数。从“tain/box_loss”、“tain/obj_loss”、“tain/cls_loss”可以看出，随着训练的进行，三种损失值在不断变化。这些损失值总体呈下降趋势，说明模型在不断学习和优化，逐渐提高对目标框位置、目标存在与否以及目标分类的准确性。precision 和 recall 这两个指标反映了模型在检测人群时的准确性和完整性。结果显示精确率较高，说明模型在判断一个检测结果为正例时的准确性较高，误报较少；同时它的召回率也较高，说明模型能够检测出更多的目标，漏报较少。

3.3. 实验结果

NWPU 数据集的实验结果如图 5 所示，本文将所述模型与 YOLOv5，YOLOv10 方法进行了比较。在图 5 中 a 是原图，b、c、d 是置信度等于 0.1 的条件下对应模型检测结果。在图 5 中 YOLOv10 分别检测出 11 与 12 人，而 YOLOv5 分别检测出超过 30 人与 16 人，改进的 YOLOv5 都检测出超过 60 人，我们可以发现 YOLOv10 的检测效果不如 YOLOv5，这是因为 YOLOv10 适用于检测较小的物体或远处的物体，

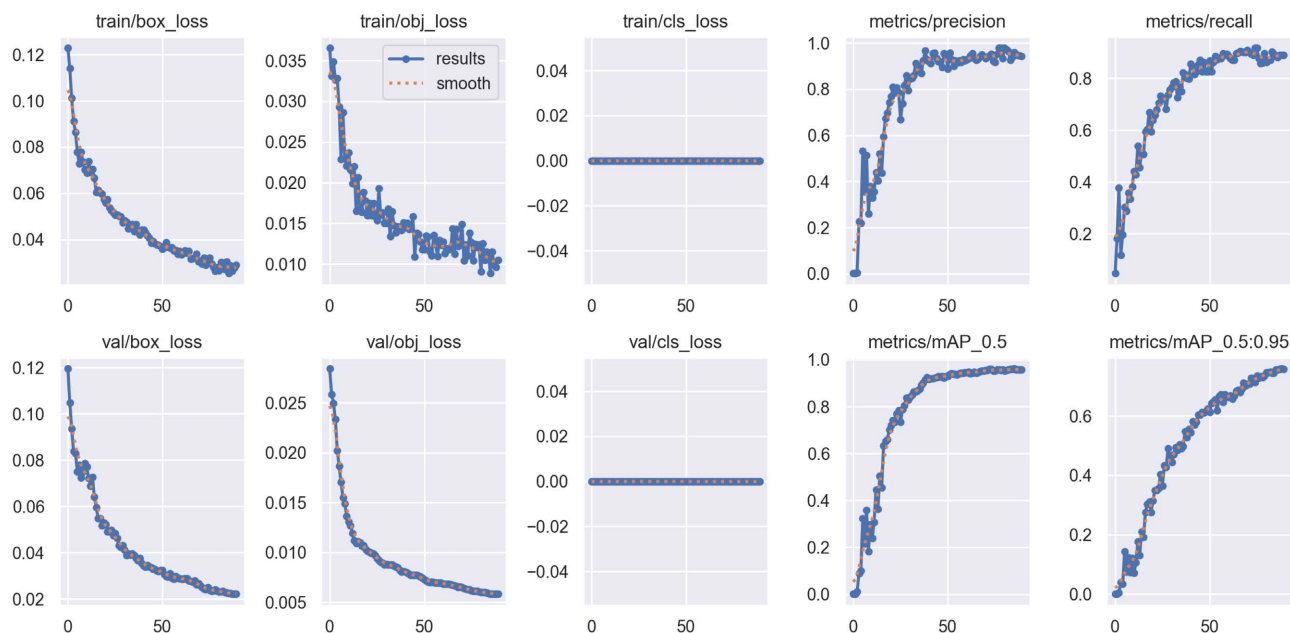


Figure 4. Curve graphs of three losses and evaluation indicators

图 4. 3 种 loss 与评价指标的曲线图



Figure 5. Original image and detection results of three models

图 5. 原图与 3 种模型检测结果

而不适用于检测尺度变化大的情况。本文提出的改进 YOLOv5 检测效果明显优于 YOLOv5 与 YOLOv10。在图 5 中原图总人数均在 200 到 300 人之间，从检测人数的准确率来看，YOLOv10 的检测结果准确率较低不到 10%，YOLOv5 检测结果准确率也较低不到 20%，改进的 YOLOv5 检测结果准确率较高在 70%到 80%之间。这验证了我们提出的方法适用于检测人群尺度变化大并且人比较小的情况。

4. 结束语

本文针对现有密集人群检测方法在处理密集、小目标等复杂场景时的不足，提出了一种改进的 YOLOv5 密集人群检测算法。该算法通过多尺度检测方法和分割输入图像进行检测，有效提高了复杂场景下的检测精度。实验结果表明，改进后的 YOLOv5 算法在处理人群尺度变化大和目标较小的情况时，

检测效果优于 YOLOv5 和 YOLOv10, 为密集人群检测提供了一种更为准确的解决方案。

基金项目

本研究由“山西省基础研究计划项目(202303021211245)”资助。

参考文献

- [1] Zhao, C.R., Qi, D., Dou, S.G., *et al.* (2021) Key Technologies of Intelligent Video Surveillance: A Review of Pedestrian Reidentification Research. *Chinese Science: Information Science*, **51**, 1979-2015.
- [2] Qiu, X., Wang, G.S., Zhao, Y.Y., *et al.* (2021) Multi-Box Pose Estimation Based on YOLOv3 and EPNP Algorithm. *Computer Measurement and Control*, **29**, 126-131.
- [3] Bochkovskiy, A., Wang, C. and Liao, H.Y. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. <https://doi.org/10.48550/arXiv.2004.10934>
- [4] 俞军, 贾银山. 改进 YOLOv5 的小目标检测算法[J]. 计算机工程与应用, 2023, 59(12): 201-207.
- [5] Wang, A., Chen, H., Liu, L.H., *et al.* (2024) YOLOv10: Real-Time End-to-End Object Detection. <https://doi.org/10.48550/arXiv.2405.14458>
- [6] Wang, Q., Gao, J., Lin, W. and Li, X. (2021) NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 2141-2149. <https://doi.org/10.1109/tpami.2020.3013269>