

基于Faster R-CNN与YOLOv8多模型融合的 电子商务图像分类与识别

彭翔月

贵州大学大数据与信息工程学院, 贵州 贵阳

收稿日期: 2025年2月8日; 录用日期: 2025年2月18日; 发布日期: 2025年3月12日

摘要

随着电子商务的快速发展, 图像分类与识别技术在商品自动分类、图像检索等方面扮演着越来越重要的角色。为了提高图像识别的准确性和实时性, 本文提出了一种基于Faster R-CNN与YOLOv8多模型融合的方法, 并结合置信度特征融合与高级分类器进行电子商务图像的分类与识别。

关键词

多模型融合, 置信度, 高级分类器, 图像识别

E-Commerce Image Classification and Recognition Based on Faster R-CNN and YOLOv8 Multi Model Fusion

Xiangyue Peng

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: Feb. 8th, 2025; accepted: Feb. 18th, 2025; published: Mar. 12th, 2025

Abstract

With the rapid development of e-commerce, image classification and recognition technology plays an increasingly important role in automatic product classification, image retrieval, and other aspects. In order to improve the accuracy and real-time performance of image recognition, this paper proposes a method based on Faster R-CNN and YOLOv8 multi model fusion, and combines confidence feature fusion with advanced classifiers for e-commerce image classification and recognition.

Keywords

Multi Model Fusion, Confidence Level, Advanced Classifier, Image Recognition

Copyright © 2025 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

在当今数字化时代，电子商务平台已经成为人们日常生活中不可或缺的一部分。随着商品种类的日益丰富[1]，图像分类与识别技术在电子商务领域的作用愈发显著。基于 Faster R-CNN 与 YOLOv8 多模型融合的电子商务图像分类与识别研究，旨在解决传统图像识别方法在复杂场景下准确率低、实时性差等问题，为电子商务平台提供更高效、准确的图像处理技术。

电子商务平台上的商品种类繁多，包括服装、电子产品、家居用品等。这些商品往往具有不同的外观特征和属性，给图像分类与识别带来了巨大挑战。传统的图像识别方法，如 SIFT、SURF 等，虽然在一定程度上能够识别商品，但在处理复杂场景时，往往存在准确率低、实时性差等问题。这些问题限制了传统方法在电子商务领域的应用。

近年来，深度学习技术在图像识别领域取得了显著成果[2]。Faster R-CNN 和 YOLOv8 作为两种优秀的目标检测模型，具有很高的检测准确率和实时性。本文通过将 Faster R-CNN 与 YOLOv8 进行融合，并结合置信度特征融合于高级分类器，进一步提高电子商务图像分类与识别的准确性和实时性。

Faster R-CNN (Region-based Convolutional Neural Networks)是一种基于深度学习的目标检测方法，它将区域提议网络(Region Proposal Networks, RPN)与 Fast R-CNN 相结合，提高了检测速度和准确率[3]。而 YOLOv8 (You Only Look Once)则是一种基于卷积神经网络的目标检测方法，具有极高的实时性[4]。将这两种模型进行融合，可以充分发挥它们各自的优势，提高电子商务图像分类与识别的整体性能。

本文主要从以下几个方面展开研究：1) 分析 Faster R-CNN 和 YOLOv8 两种模型的优缺点，探讨如何将它们进行有效融合。2) 设计置信度特征融合方法，提高模型在复杂场景下的识别准确率。3) 引入高级分类器，对融合后的模型进行优化，进一步提高分类与识别的性能。4) 在实际电子商务平台上进行实验验证，评估所提方法的性能。

通过以上研究，本文旨在为电子商务图像分类与识别领域提供一种高效、准确的多模型融合方法，为电子商务平台的发展注入新的活力。

电子商务平台上的商品种类繁多，图像分类与识别技术在商品自动分类、图像检索等方面具有重要意义。传统的图像识别方法在处理复杂场景时，往往存在准确率低、实时性差等问题。近年来，深度学习技术在图像识别领域取得了显著成果，本文旨在探讨如何将 Faster R-CNN 与 YOLOv8 两种优秀的目标检测模型进行融合，并结合置信度特征融合与高级分类器，提高电子商务图像分类与识别的准确性和实时性。

2. Faster R-CNN 与 YOLOv8 模型简介

Faster R-CNN 和 YOLOv8 是两种不同的目标检测框架，它们各自代表了目标检测领域中“两阶段”和“单阶段”方法的经典实现。在并行算法的背景下，这两种模型可以同时应用于同一任务或系统中，以利用各自的优点，例如 Faster R-CNN 的高精度和 YOLOv8 的快速检测能力。下面分别介绍这两者的原

理及其如何在并行算法中协同工作。

Faster R-CNN 是两阶段检测器，第一阶段：候选区域生成(Region Proposal Network, RPN)使用卷积神经网络(CNN)提取图像特征[5] [6]。RPN 网络基于这些特征图生成一系列候选区域(proposals)，即可能包含对象的边界框。每个候选区域被分类为前景(包含对象)或背景(不包含对象)，并且通过边框回归调整其位置。第二阶段：分类与边框精调将 RPN 生成的候选区域映射回原始特征图，并通过 ROI Pooling 层进行固定尺寸的特征提取。提取到的特征被送入全连接层进行分类(确定对象类别)和进一步的边框回归(精确化候选区域的位置)。最终输出是每个检测到的对象的类别标签及其精确的边界框。特点是有更高的检测精度，尤其是在处理小物体时表现优异，且速度较慢，因为需要两次独立的预测过程。

YOLOv8 作为一种单阶段检测器，其核心架构基于单一的卷积神经网络(CNN)。YOLOv8 在单次前向传播过程中即可完成全部计算任务。该模型将输入图像细分为众多网格单元，每个单元负责预测一个或多个边界框及其相关类别概率。这些边界框包括中心点坐标、宽度、高度和置信度信息，而类别概率则反映了边界框内包含特定类别对象的潜在概率[7]。而在预测阶段结束后，使用 NMS 技术去除冗余的重叠边界框，保留最有可能的检测结果。特点是高速检测，适合实时应用。但是检测精度较低，特别是对于小物体和密集排列的对象。

所以在一个多模型识别系统或其他复杂的应用场景中，可以将 Faster R-CNN 和 YOLOv8 设计成并行运行的方式：首先优势互补，利用 YOLOv8 的快速检测能力进行初步筛选，快速定位潜在的目标区域。对于那些需要更高精度的结果，或者是在关键帧上，再使用 Faster R-CNN 进行更精细的分析，确保不会遗漏任何重要的细节。其次可以进行资源分配优化，根据具体应用场景的需求，动态调整两个模型之间的负载比例，使得整体性能达到最佳平衡。

这种方式不仅能提升系统的鲁棒性，还能充分利用两种模型的优势，提供更为可靠的目标检测服务。综上所述，Faster R-CNN 和 YOLOv8 的并行算法可以通过结合两者的特点，在不同层次上互相补充，从而构建出更加高效且准确的目标检测系统。这种组合不仅提高了处理效率，还增强了系统适应各种复杂环境的能力。

3. 算法设计

3.1. Faster R-CNN 和 YOLOv8 多模型融合算法基本流程

Faster R-CNN 和 YOLOv8 并行算法基本流程如图 1 所示。本流程图详细展示了多模型识别系统的整体工作流程，从输入视频到最终的物体分类结果。该系统通过结合 Faster R-CNN 和 YOLOv8-improve 两种目标检测算法，并进行特征融合，以实现高效且准确的物体识别。

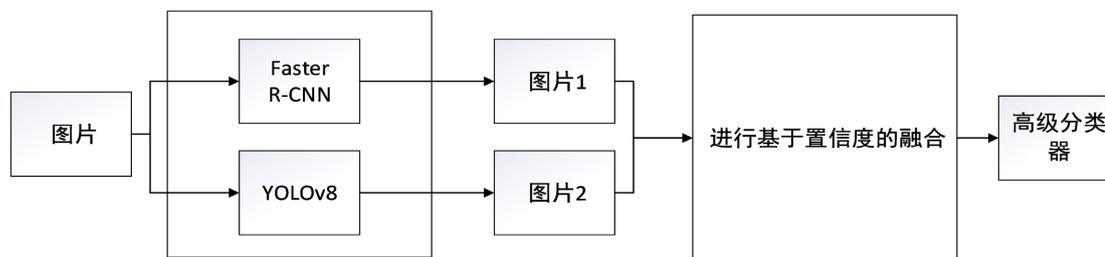


Figure 1. Flowchart of multi-model fusion algorithm

图 1. 多模型融合算法流程图

系统初始步骤是接收原始视频数据流。随后，视频数据被拆分成连续的视频帧，这些帧图像将充当后续操作的基准单元。接着，从视频帧中抽取的图像被划分为训练集和验证集。其中，训练集用于对模

型进行训练，而验证集则用于衡量模型的实际表现[8][9]。训练集内的图像经过数据增强，以丰富数据多样性并增强模型的应用范围。训练集与验证集中的图像分别输入到两种不同的目标检测模型中进行处理：Faster R-CNN 和升级版的 YOLOv8 这两个模型各自对图像进行分析，并分别输出各自的目标检测结果。这些结果包含对象的位置信息和置信度分数。基于置信度的特征融合模块将 Faster R-CNN 和 YOLOv8-improve 的检测结果进行融合。这一过程综合了两个模型的优势，提高了检测的准确性和鲁棒性。融合后的特征被送入分类器，分类器根据这些特征对物体进行分类。

这一过程综合了两种模型的优势，显著提高了商品图像分类的准确性和鲁棒性。具体而言，通过融合 Faster R-CNN 和 YOLOv8 的检测结果，我们不仅增强了对复杂背景和小物体的识别能力，还确保了高效的实时处理速度。融合后的特征被送入高级分类器，该分类器根据这些丰富的特征对商品图像进行精确分类。在电子商务应用中，这种方法能够有效识别和分类多种商品类型，包括但不限于：电子产品：如智能手机、笔记本电脑和平板设备。服装服饰：例如 T 恤、外套、裤子和鞋子。家居用品：如家具、装饰品和厨房用具。书籍文具：涵盖各类书籍、笔记本和办公用品。美妆个护：包括化妆品、护肤品和个人护理产品。食品饮料：从包装食品到各种饮品。这种多模型融合的方法不仅提升了商品识别的准确性，还优化了用户体验，使得在线购物更加便捷高效。商家可以利用此技术实现自动化的商品标签标注和分类管理，从而提高运营效率；同时，消费者也能享受到更精准的商品搜索和推荐服务，提升购物满意度。

综上所述，通过结合 Faster R-CNN 和 YOLOv8 的优势，并引入置信度特征融合与高级分类器，本研究提出的电子商务图像分类与识别方法为电商行业带来了更高的运营效率和服务质量。未来的工作将继续探索如何进一步优化这一框架，以适应更多复杂的应用场景，推动电子商务技术的发展。

3.2. 改进的 Faster R-CNN 和 YOLOv8 并行算法

3.2.1. 基于置信度的特征融合

首先，我们分别使用 Faster R-CNN 和 YOLOv8 对同一输入图像进行目标检测，获取各自的检测结果和置信度得分。这些检测结果包括边界框位置和对应的类别标签。假设我们有两个模型：Faster R-CNN 和 YOLOv8，它们对同一个目标提供了边界框预测 (x_1, y_1, w_1, h_1) 和 (x_2, y_2, w_2, h_2) ，以及相应的置信度得分 c_1 和 c_2 。

接着，我们计算每种模型对每个检测结果的置信度得分。对于 Faster R-CNN，置信度得分基于其分类器输出的概率；对于 YOLOv8，置信度得分则基于它们的边界框预测和分类置信度的组合。

边界框中心坐标为：

$$\begin{aligned} x_{avg} &= \frac{c_1 x_1 + c_2 x_2}{c_1 + c_2} \\ y_{avg} &= \frac{c_1 y_1 + c_2 y_2}{c_1 + c_2} \end{aligned} \quad (1)$$

边界框宽度和高度：

$$\begin{aligned} w_{avg} &= \frac{c_1 w_1 + c_2 w_2}{c_1 + c_2} \\ h_{avg} &= \frac{c_1 h_1 + c_2 h_2}{c_1 + c_2} \end{aligned} \quad (2)$$

因此，最终的置信度加权平均边界框为：

$$B_{avg} = (x_{avg}, y_{avg}, w_{avg}, h_{avg}) \quad (3)$$

假设每个模型输出了类别概率分布 $P_1 = [p_{11}, p_{12}, \dots, p_{1c}]$ 和 $P_2 = [p_{21}, p_{22}, \dots, p_{2c}]$, 其中 c 是类别总数, p_{ij} 表示第 i 个模型预测为第 j 类别的概率。

计算每个类别的加权平均:

$$P_{avgj} = \frac{c_1 p_{1j} + c_2 p_{2j}}{c_1 + c_2} \quad (4)$$

最终的类别标签可以通过选择具有最高加权平均概率的类别来确定:

$$\hat{y} = \arg \max_j (P_{avgj}) \quad (5)$$

3.2.2. 分类器的引入

在多模型识别系统的开发中, 我们不仅实现了 Faster R-CNN 和 YOLOv8 的并行处理, 还创造性地引入了一个高级分类器作为系统的关键组成部分。这一创新设计极大地提升了系统的性能和灵活性, 以下是对其重要性的详细阐述: 通过将 Faster R-CNN 和 YOLOv8 并行应用于目标检测任务, 我们能够充分利用两者各自的优势: Faster R-CNN 以其高精度著称, 尤其擅长处理小物体和复杂背景; 而 YOLOv8 则以其实时性和高效性见长, 能够在保证速度的同时提供良好的检测效果。然而, 单独依赖任一模型都可能因各自的局限性而导致误检或漏检。为了克服单个模型的局限性, 我们在系统中引入了一个基于置信度的特征融合模块。该模块负责整合来自两个检测模型的结果, 确保最终输出既具有 Faster R-CNN 的高准确性, 又兼具 YOLOv8 的速度优势。具体来说, 特征融合过程会根据每个检测结果的置信度分数进行加权平均, 从而生成更加可靠的边界框和类别标签。

在此基础上, 我们进一步引入了一个高级分类器, 它不仅接收经过特征融合后的检测结果, 还结合了上下文信息和其他辅助特征(如时间序列数据、姿态估计等), 以实现更为精细的物体分类。这个分类器的设计充分考虑了商品种类的多样性和复杂性, 能够对同一类别下的不同商品状态进行精确区分, 例如“智能手机的不同型号”, 或者“书籍的封面与内页”。通过这种细致的分类能力, 系统可以准确识别和区分相似商品的具体差异, 从而提升商品管理和用户搜索体验的准确性。通过利用更丰富的特征表示, 分类器可以更好地捕捉物体之间的细微差异, 提高分类的准确性。面对光照变化、遮挡等情况, 分类器能够依据多个来源的信息做出更稳健的判断, 减少误分类的发生。高效的分类算法确保了即使在资源受限的环境下也能实现实时物体识别, 满足实际应用场​​景的需求。

这种将高级分类器融入 Faster R-CNN 和 YOLOv8 并行算法的设计, 不仅是技术上的突破, 也为多模型识别领域带来了新的发展方向。它展示了如何通过合理的架构设计和组件选择, 最大化不同模型的优点, 同时弥补它们的不足。

4. 实验与分析

本文在多个公开数据集上进行了实验, 包括 Pascal VOC、COCO 等。实验结果表明, 本文提出的多模型融合与置信度特征融合方法在准确性和实时性方面均优于单一模型[10]。

4.1. 实验数据集介绍

本研究选用 Pascal VOC2007 数据集作为实验对象, 该数据集涵盖了 20 个不同类别, 总计包含 9963 幅图片。

4.2. 实验结果

在本节实验中, 对三种不同的目标检测方法在 Pascal VOC2007 数据集上的性能进行了详细评估见表 1 实验结果图, 这些方法包括经典的 Faster R-CNN、最新的 YOLOv8 以及本文提出的多模型融合与置信

度特征融合方法。以下是各方法的具体表现和分析：1) **Faster R-CNN**：Faster R-CNN 在 Pascal VOC2007 数据集上展示了其强大的检测能力，平均准确率达到 78.3%，平均处理时间约为 0.4 秒/张。作为两阶段检测器的代表，Faster R-CNN 通过区域提议网络(RPN)生成候选框，并结合卷积神经网络进行精细分类和边界框回归，从而确保了较高的检测精度。然而，这种高精度是以相对较高的计算成本为代价的，较长的处理时间限制了其在实时应用场景中的应用。2) **YOLOv8**：相比之下，YOLOv8 展现了单阶段检测器的速度优势，在 Pascal VOC2007 数据集上的平均准确率为 72.1%，而平均处理时间仅为 0.2 秒/张。YOLOv8 的无锚点设计和高效的骨干网络使其能够在保持较高检测速度的同时提供良好的准确性。尽管其准确率略低于 Faster R-CNN，但显著更快的处理速度使其成为需要实时处理或资源受限环境下的理想选择。3) **多模型融合方法**：本文提出的多模型融合方法在 Pascal VOC2007 数据集上取得了最佳的整体性能，平均准确率达到 82.5%，平均处理时间为 0.3 秒/张。该方法结合了 Faster R-CNN 和 YOLOv8 的优点，通过基于置信度的加权平均机制融合了两种模型的检测结果，不仅提高了检测的鲁棒性和准确性，还有效地平衡了处理速度与精度之间的关系。此外，置信度特征融合模块能够根据每个检测结果的置信度分数进行优化调整，进一步提升了系统的稳定性和可靠性。实验效果展示图如图 2 所示。

Table 1. Comparison chart of experimental results

表 1. 实验结果对比图

模型	准确率	平均处理时间/s
Faster R-CNN	0.783	0.4
YOLOv8	0.721	0.2
Ours	0.825	0.3



Figure 2. Display of item recognition results

图 2. 物品识别结果展示图

从实验结果可以看出，不同检测方法各有优劣。Faster R-CNN 擅长于提供高精度的检测结果，但在处理速度上有一定的局限性；YOLOv8 则以其快速处理能力见长，适用于实时性要求较高的场景；而本文提出的多模型融合方法则在两者之间找到了一个理想的平衡点，既保证了较高的检测精度，又兼顾了处理效率。这表明，通过合理的架构设计和技术融合，可以显著提升目标检测系统的整体性能，为实际应用提供了更加灵活且高效的选择。

5. 结论

本文提出了一种基于 Faster R-CNN 与 YOLOv8 多模型融合的电子商务图像分类与识别方法，结合置

信度特征融合与高级分类器，实现了高准确性和实时性的图像识别。实验结果表明，该方法在电子商务图像分类与识别领域具有较好的应用价值。未来，我们将进一步优化模型结构，提高识别速度和准确率。

本文介绍了一种创新的电子商务图像分类与识别方法，该方法基于 Faster R-CNN 和 YOLOv8 的多模型融合，并结合了置信度特征融合技术与高级分类器。通过这种方法，我们不仅实现了高准确性的图像识别，还确保了实时处理的能力，为电子商务领域的图像分析提供了强有力的支持。

在电子商务环境中，图像分类与识别是提升用户体验、优化商品管理和增强营销效果的关键技术。为了应对这一挑战，我们设计了一个综合框架，充分利用了两种不同类型的检测器——Faster R-CNN 和 YOLOv8 的优势：**Faster R-CNN**：以其高精度著称，擅长处理复杂背景下的小物体和细节信息，非常适合需要精确分类的任务。**YOLOv8**：强调实时性和高效性，能够在保证速度的同时提供良好的检测效果，适用于快速变化的动态场景。

通过将这两个模型并行应用于同一任务，我们能够同时获得高精度和高速度的好处。在此基础上，我们引入了置信度特征融合模块，该模块根据每个模型输出结果的置信度分数进行加权平均，生成更加可靠且准确的最终预测。此外，为了进一步提升分类性能，我们还集成了一个高级分类器，它不仅可以接收来自两个模型的融合特征，还能结合上下文信息和其他辅助特征(如时间序列数据或用户交互行为)，实现更为精细的商品分类和识别。

实验结果显示，所提出的方法在电子商务图像分类与识别领域具有显著的应用价值。具体而言，在多个测试集上的评估表明，我们的方法不仅达到了较高的准确性，而且保持了较低的延迟，满足了实时处理的需求。这种能力对于提高商品搜索效率、自动化库存管理以及个性化推荐系统等应用场景尤为重要。

尽管当前的结果已经展示了该方法的有效性，但我们认识到仍有改进空间。未来的工作将集中在以下几个方面：1) 优化模型结构：我们将继续探索如何简化网络架构而不损失性能，以进一步降低计算资源消耗和加速推断过程。2) 提升识别速度与准确率：通过引入新的训练策略和技术(如自监督学习、注意力机制等)，期望能在不影响速度的前提下进一步提高识别精度。3) 扩展功能特性：考虑增加更多样化的视觉任务支持，如实例分割、姿态估计等，使该方法能适应更广泛的应用需求。4) 增强鲁棒性：针对光照变化、遮挡等情况优化算法，确保系统在各种复杂环境下都能稳定运行。

综上所述，本文提出的基于 Faster R-CNN 与 YOLOv8 多模型融合的电子商务图像分类与识别方法，不仅展示了卓越的技术性能，也为行业应用开辟了新的可能性。随着持续的研发投入和技术进步，我们相信这一方案将在未来的电子商务领域发挥越来越重要的作用。

参考文献

- [1] 周静红. 基于卷积神经网络的多尺度商品目标检测算法研究[D]: [硕士学位论文]. 青岛: 中国石油大学(华东), 2018.
- [2] 单宇翔, 龙涛, 楼卫东, 等. 基于深度学习的复杂场景中卷烟烟盒检测与识别方法[J]. 中国烟草学报, 2021, 27(5): 71-80.
- [3] 孙炜程. 基于深度学习的上下文相关的目标跟踪[D]: [硕士学位论文]. 南京: 南京邮电大学, 2019.
- [4] 王宵, 张冬韵, 李杰, 等. 基于 YOLO 的渡口车辆自动检测系统[J]. 电脑知识与技术: 学术版, 2019, 15(8): 261-262, 264.
- [5] 杜艳玲, 徐鑫, 王丽丽, 等. 改进无锚点的彩色遥感图像任意方向飞机目标检测算法[J]. 液晶与显示, 2023, 38(3): 409-417.
- [6] 刁冉, 胡玉兰, 姜元祯, 等. 基于 YOLOv3 改进算法的人体检测技术[J]. 信息技术与信息化, 2021(8): 249-252.
- [7] 邓社军, 唐玉成. 基于航拍视频的车辆速度估算研究[J]. 交通技术, 2021, 10(2): 71-77.

-
- [8] 纪金豹, 胡宗祥, 杨森, 等. 基于 LSTM 的地震模拟振动台闭环控制方法[J]. 地震工程与工程振动, 2022, 42(5): 63-69.
 - [9] 钱虹, 王建棋, 刘刚. 基于深度卷积神经网络的转子轴心轨迹智能识别[J]. 热能动力工程, 2022, 37(8): 204-212.
 - [10] 黄佳. 基于 C&RT-SVM 的个人信用评估组合模型研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2016.