

《计算机视觉》期末报告

201250182 郑义

1. Abstract

本文将简要介绍 Object Detection（物体检测）领域方面的问题描述、发展趋势，以及介绍目标检测领域的几种代表性传统方法，最后对总结文章中的相关方法介绍，并对物体检测领域的未来发展提出期望和看法。

关键词：Object Detection、Computer Vision

2. 问题描述

Object Detection（物体检测）是计算机视觉领域里一个非常重要的任务，其主要工作是将一张图片中的物体通过一个框（一般是一个矩形框）标出来，并且能够识别出框中物体的所属类别。其主要目标是为计算机视觉的下一步工作进行预处理，识别出图片中的目标物体，减少甚至排除掉其他无关物体对目标物体的影响。对于 Object Detection 来说，其主要任务就是：图片中有哪些对象？这些对象在哪里^[1]。

物体检测在 90 年代开始发展，受限于当时的计算能力和准确率，物体检测的应用场景十分受限，基本无法应用到实际场景当中，随着 2001 年 VJ Det 的出现，相比之前较快的计算速度和较高的准确率标志着物体检测正式投入应用正式开启。后续的发展主要围绕在如何在提高检测速度的基础上，提高检测准确率。

3. 发展趋势

Object Detection 在过去二十年里发展迅速，在准确率和检测速度上都有很大的提升与进步。根据使用的核心方法不同，其历史发展主要分为两个阶段，分别是：

- 1) 2014 年以前的以传统物体检测方法为代表的年代
- 2) 2014 年之后的以深度学习方法为代表的年代。

在 2014 年以前，传统的物体检测检测方法通过人工标注特征来获取数据，并通过多种出色有效的特征计算方式和算法来优化目标检测速度并提高检测的准确度，使其能够以实时的速度得出结果。传统物体检测方法的发展主要受限于图片数量、质量的缺乏以及算力有限导致的人工标注特征困难。后续提出的深度学习方法则在算力发展迅速、深度学习模型发展迅速的基础上提供了高效、准确的特征标注和目标检测方法，当下的实际应用中更多是以更准确、更迅速的深度学习方法作为目标检测的核心方法^[1]。

在本文作者看来，目标检测领域的未来更多会集成在深度学习和大模型基础上，探究如何以更快速的方法去获得更准确的目标检测结果；在 IoT 领域，传统学习方法也许还会成为选择，以较小的成本开销来获得可以接受的识别结果会是传统学习方法在 IoT 领域获得

应用的优势之一；但无论如何，以深度学习为基础的物体检测算法

4. 代表性方法

下面对 90 年代物体检测领域出现开始的一些非深度学习的传统算法进行简单介绍，主要包括 VJ Det、HOG Det、DPM 三种代表性的传统方法。

4.1. VJ Det

VJ Det 是 P. Viola 和 M. Jones 于 2001 年提出的一种基于滑动窗口、特征计算和机器学习的物体检测算法。在没有任何约束的情况，该算法在 700MHz 的 Intel Pentium III 处理器上可以以大约 15 张图片每秒的速度进行人脸检测^[2]。由于其对于目标检测的速度提升和准确率提升，该方法的提出标志着目标检测领域可以在实际应用场景落地使用。

滑动窗口指的是通过一个矩形窗口来遍历图像的每一个位置，并标记可能为目标物体的位置，使用滑动窗口的难点主要在于窗口大小的确定和遍历图像的时间开销该方法在滑动窗口的基础上提出了三个核心方面^[2]来优化目标检测的计算速度和计算准确率，分别是：

- 1) Integral Image，该方法是 VJ Det 提出的一种加快特征提取的方法，原先的 Haar 特征提取需要计算区域内灰度值之和，并通过相减的方式得出最后的结果，从而将明暗特征反应出来，该方法的计算量较大。Integral Image 则提出了计算图片左上区域像素灰度值之和的方法来提高计算速度。图 1 展示了一个 Integral Image 的计算示例。
- 2) 以 Adaboost 为基础的特征选择分类器，通过将多个简单的分类器聚合起来，使其在拥有较快计算速度的同时，能够有较好的分类能力。
- 3) Cascade（级联结构），在原先图片输入规模大，滑动窗口多造成计算时间长的背景下，提出了通过级联的结构来减少计算的时间和开销，先通过初步粗筛筛除掉大部分不符合目标的窗口，剩下只需要对少部分的窗口进行更加细致的判断就可以得出结果，这样省去了大部分不必要的检测时间。

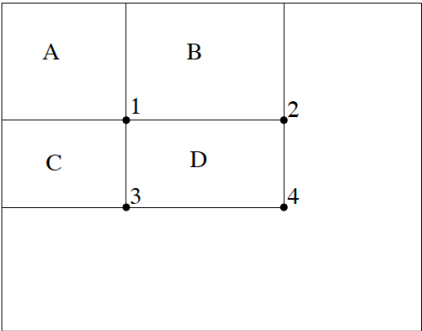


图 1 区域 D 的像素灰度值之和计算可以被划分为 4 个子区域，点 1 的值为区域 A 的像素灰度值之和，点 2 的值为区域 A + B 之和，点 3 的值为区域 A + C 之和，点 4 的值为区域 A + B + C + D 之和，计算区域 D 的像素灰度值之和可以通过 $4 + 1 - 2 - 3$ 得到^[2]

4.2. HOG Det

HOG 最早被提出用于行人检测领域^[4]，该方法主要是在 scale-invariant feature transform 和 shape contexts 上提出的改进。HOG 描述了一张有关图像梯度方向的直方图，通过将图像分割为相同大小的网格 (cell)，使用重叠的局部对比度归一化，最终得到一张图像梯度方向的直方图。

HOG 主要描述的是关于图片中物体边缘特征的信息。在本文作者看来，相比 VJ Det 描述图片明暗对比信息，物体边缘特征信息是会更加明显且更容易进行区分的，明暗特征可能会由于图片变为灰度图而被削弱，但 HOG 特征则平衡了在不同情况下（比如平移、比例缩放、光照等）的特征不变性，这使其对于不同场景下的图像输入都有着较好的特征获取能力。虽然 HOG 是在行人检测领域被提出，但如今也被广泛用于各个领域，可以称得上是物体检测领域的一种基础特征描述算子。

由于 HOG 是通过边缘特征来获取信息，因此在遇到物体遮挡重叠、噪声严重的情况下，图像中的物体边缘特征划分可能不够清晰，会影响到 HOG 的特征提取；除此之外，由于 HOG 的特征提取计算维度较高，因此需要耗费更多的时间来获取，在算力不足的年代无法较好的实时计算出数量较多图片的特征。

4.3. DPM

DPM 是 HOG 的一种拓展版本，最早在 2008 年被 P. Felzenszwalb 提出^[5]。其获得了 VOC-07、-08、-09 三年的冠军。DPM 采取了分而治之的思想，物体检测可以被看作是在分解物体，通过检测某个物体的组成来检测出该物体。比如要检测一辆车，可以看作是去检测这辆车的窗户、轮子、车身等。

5. 总结

物体检测领域的发展随着深度学习的出现正在快速前进着，早期的传统方法在数据量增长迅速、计算速度和准确率要求高的场景下大部分已经不再采用，但在早期算力有限的背景下，能够提出这些优秀的特征计算方法和结构来进行物体检测无疑是令人称赞的，从 VJ Det 的 Integral Image 和级联结构到 HOG 的梯度直方图再到 DPM 的分而治之思想，都是物体检测领域值得学习和借鉴的思想。

物体检测领域的应用场景在当下非常广泛，了解非深度学习的传统物体检测方法是奠定基础，希望本报告能够对想要在物体检测领域深入学习的读者有一定的影响。

致谢

感谢 Z. Zou, K. Chen 等编写的 *Object Detection in 20 Years: A Survey*。这篇综述很好的带我领略了物体检测领域的过去 20 年的发展，带我回顾了 90 年代到 14 年的传统物体检测方法打下的坚实基础，又带我领略了 14 年之后基于深度学习方法的物体

检测领域的新进展。

感谢任桐炜教授，他在《计算机视觉》这门课程上的生动讲授带我领略了计算机视觉领域的独特美景，带我窥探了计算机视觉领域的各个子任务、相关经典方法和进展，让我对计算机领域有了一定的了解。

感谢《计算机视觉》这门课程的助教，感谢和我一起组队的母舰同学，在他们的帮助下我才能够顺利完成每一次作业，完成这一次的报告编写。

愿《计算机视觉》课程能够越来越好！

参考文献

- [1] Z. Zou, K. Chen, Z. Shi, Y. Guo and J. Ye, "Object Detection in 20 Years: A Survey," in Proceedings of the IEEE, vol. 111, no. 3, pp. 257-276, March 2023, doi: 10.1109/JPROC.2023.3238524.
- [2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in CVPR, vol. 1.IEEE, 2001, pp. I-I.
- [3] P. Viola and M. J. Jones, "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137-154, 2004.
- [4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
- [5] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in CVPR. IEEE, 2008, pp. 1-8.
- [6] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in CVPR. IEEE, 2010, pp. 2241-2248.