A Survey of Modern Deep Learning based Object Detection Models

Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam,
Nadia Kanwal, Mamoona Asghar, and Brian Lee

***Abstract*—Object Detection is the task of classification and localization of objects in an image or video. It has gained prominence in recent years due to its widespread applications. This article surveys recent developments in deep learning based object detectors. Concise overview of benchmark datasets and evaluation metrics used in detection is also provided along with some of the prominent backbone architectures used in recognition tasks. It also covers contemporary lightweight classification models used on edge devices. Lastly, we compare the performances of these architectures on multiple metrics.**

***Index Terms*—Object detection and recognition, convolutional neural networks (CNN), lightweight networks, deep learning**

I. Introduction

Object detection is a trivial task for humans. A few months old child can start recognizing common objects, however teaching it to the computer has been an uphill task until the turn of the last decade. It entails identifying and localizing all instances of an object (like cars, humans, street signs, etc.) within the field of view. Similarly, other tasks like classification, segmentation, motion estimation, scene understanding, etc, have been the fundamental problems in computer vision.

Early object detection models were built as an ensemble of hand-crafted feature extractors such as Viola-Jones detector [1], Histogram of Oriented Gradients (HOG) [2] etc. These models were slow, inaccurate and performed poorly on unfamiliar datasets. The re-introduction of convolutional neural network (CNNs) and deep learning for image classification changed the landscape of visual perception. Its use in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 challenge by AlexNet [3] inspired further research of its application in computer vision. Today, object detection finds application from self-driving cars and identity detection to security and medical uses. In recent years, it has seen exponential growth with rapid development of new tools and techniques.

This survey provides a comprehensive review of deep learning based object detectors and lightweight classification architectures. While existing reviews are quite thorough [4]–[7], most of them lack new developments in the domain. The main contributions of this paper are as follows:

S.S.A. Zaidi, N. Kanwal, M Asghar and B. Lee are with the Athlone Institute of Technology, Ireland. M.S. Ansari is with the Aligarh Muslim University, India. A. Aslam is with the Insight Center for Data Analytics, National University of Ireland, Galway. (Emails: sahilzaidi78@gmail.com, samar.ansari@zhect.ac.in, asra.aslam@insight-centre.org, nkanwal@ait.ie, masghar@ait.ie, blee@ait.ie)

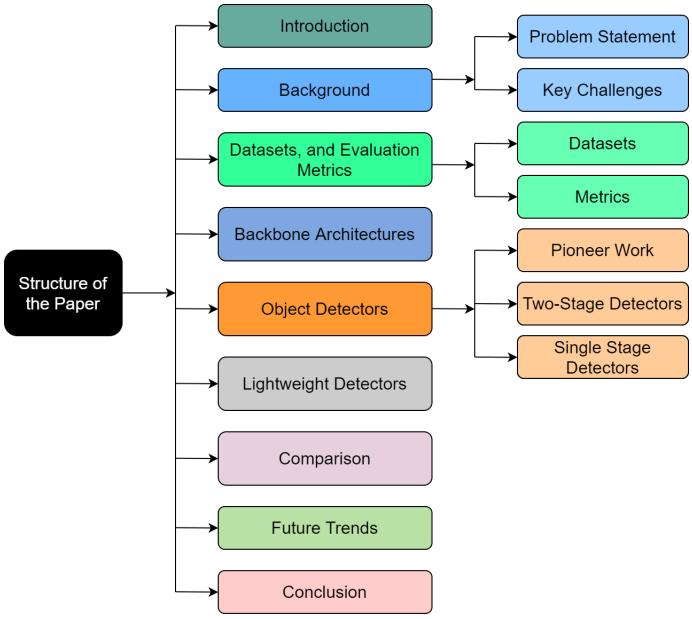

Fig. 1: Structure of the paper.

1) This paper provides an in-depth analysis of major object detectors in both categories – single and two stage detectors. Furthermore, we take historic look at the evolution of these methods.
2) We present a detailed evaluation of the landmark backbone architectures and lightweight models. We could not find any paper which provides a broad overview of both these topics.

In this paper, we have systematically reviewed various object detection architectures and its associated technologies, as illustrated in figure 1. Rest of this paper is organized as follows. In section II, the problem of object detection and its associated challenges are discussed. Various benchmark datasets and evaluation metrics are listed in Section III. In Section IV, several milestone backbone architectures used in modern object detectors are examined. Section V is divided into three major sub-section, each studying a different category of object detectors. This is followed by the analysis of a special classification of object detectors, called lightweight networks in section VI and a comparative analysis in Section VII. The future trends are mentioned in Section VIII while the paper is concluded in Section IX.