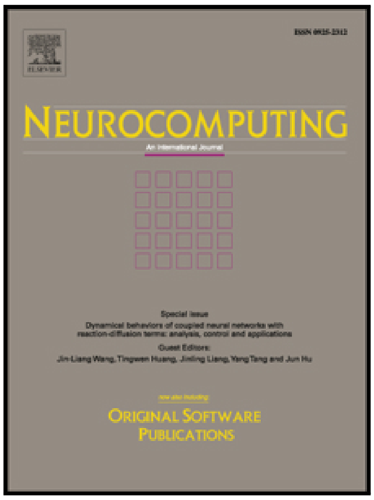
Accepted Manuscript



Efficient Deep Network for Vision-based Object Detection in Robotic Applications

Keyu Lu, Xiangjing An, Jian Li, Hangen He

PII: S0925-2312(17)30565-9
DOI: 10.1016/j.neucom.2017.03.050
Reference: NEUCOM 18278

To appear in: *Neurocomputing*

Received date: 17 June 2016
Revised date: 11 February 2017
Accepted date: 17 March 2017

Please cite this article as: Keyu Lu, Xiangjing An, Jian Li, Hangen He, Efficient Deep Network for Vision-based Object Detection in Robotic Applications, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.03.050

Efficient Deep Network for Vision-based Object Detection in Robotic Applications

Keyu Lu[a,*], Xiangjing An[a], Jian Li[a], Hangen He[a]

[a]*College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, Hunan, China*

Abstract

Vision-based object detection is essential for a multitude of robotic applications. However, it is also a challenging job due to the diversity of the environments in which such applications are required to operate, and the strict constraints that apply to many robot systems in terms of run-time, power and space. To meet these special requirements of robotic applications, we propose an efficient deep network for vision-based object detection. More specifically, for a given image captured by a robot mount camera, we first introduce a novel proposal layer to efficiently generate potential object bounding-boxes. The proposal layer consists of efficient on-line convolutions and effective off-line optimization. Afterwards, we construct a robust detection layer which contains a multiple population genetic algorithm-based convolutional neural network (MPGA-based CNN) module and a TLD-based multi-frame fusion procedure. Unlike most deep learning based approaches, which rely on GPU, all of the on-line processes in our system are able to run efficiently without GPU support. We perform several experiments to validate each component of our proposed object detection approach and compare the approach with some recently published state-of-the-art object detection algorithms on widely used datasets. The experimental results demonstrate that the proposed network exhibits high efficiency and robustness in object detection tasks.

Keywords: Deep network, Object detection, Computer vision, Robotic application, MPGA

1. Introduction

Object detection plays a critical role in a wide range of robotic applications such as service robot interaction [1], autonomous driving [2], and collision avoidance [3], which need to detect the presence of both stationary and moving objects in a specific area of interest around the host robots in order to perform corresponding actions such as interaction, braking and evading.

Vision sensors become particularly important in robot systems [4]. In the first place, compared with other popular object detection sensors such as LiDAR and millimeter-wave radar, vision sensors are able to obtain rich information from objects (such as luminance, color and texture) and to monitor wide areas [5], and these capabilities are of value to a range of robotic applications. Moreover, the cost of most vision sensors is much lower than that of LiDAR or millimeter-wave radar [6], and consequently vision sensors are a more feasible solution for robots employed for commercial use. Last but not least, vision sensors require much less power and space, and accordingly they are particularly suitable for integration into robot systems where power and space are limited.

Vision-based object detection has been addressed through a wide variety of methods in recent years. In the last decade, hand-craft solutions have proved extremely popular and have been applied successfully in several robotic applications. Early hand-craft solutions mainly focused on detecting objects using keypoints. Lowe proposed the SIFT algorithm [7] that is able to achieve scale and rotational invariance in keypoint detection, and this made keypoint-based approaches a popular solution to the object detection problem. Subsequently, numerous extensions (e.g., PCA-SIFT [8] and SUFR [9]) emerged that aimed at improving object detection performance in the context of robotic applications. In a different line of work, Viola and Jones [10] proposed a boosted object detection method for real-time object detection, using cascades strategy and Haar-like features to make efficient calculations through integral images. In the same vein, Dollár *et al.* introduced the Aggregated Channel Feature (ACF) [11], which obtains multiple channels of carefully tuned features through integral images and trains a robust object detector using AdaBoost [10] and efficient decision trees. Meanwhile, Dalal *et al.* [12] proposed an alternative hand-craft approach and popularized the use of the "HOG+SVM" paradigm in object detection. The main idea of this approach is to represent each image patch using histogram of oriented gradients and to classify the feature vector of each image patch using Support Vector Machine (SVM) [13]. Later, a breakthrough in the application of the "HOG+SVM" paradigm to object detection was achieved in the form of the DPM (Deformable Part Model) [14], which represents an image using a variant of

*Corresponding author

Email addresses: keyu.lu@nudt.edu.cn (Keyu Lu), anxiangjing@gmail.com (Xiangjing An), jianli@nudt.edu.cn (Jian Li), hangenhe@yahoo.com (Hangen He)