Robust Object Representation by Boosting-like Deep Learning Architecture

Lei Wang, Baochang Zhang, Jungong Han, Linlin Shen, Cheng-shan Qian

www.elsevier.com/locate/image

PII: S0923-5965(16)30088-1
DOI: http://dx.doi.org/10.1016/j.image.2016.06.002
Reference: IMAGE15101

To appear in: *Signal Processing : Image Communication*

Received date: 10 December 2015
Revised date: 11 May 2016
Accepted date: 2 June 2016

Cite this article as: Lei Wang, Baochang Zhang, Jungong Han, Linlin Shen and Cheng-shan Qian, Robust Object Representation by Boosting-like Deep Learning Architecture, *Signal Processing : Image Communication* http://dx.doi.org/10.1016/j.image.2016.06.002

Robust Object Representation by Boosting-like Deep Learning Architecture

Lei Wang[1], Baochang Zhang[1], Jungong Han[2], Linlin Shen[3], Cheng-shan Qian[4*]

[1]School of Automation Science and Electrical Engineering Beihang University, Beijing, China
[2]Computer Science and Digital Technologies Department, Northumbria University, Newcastle, UK
[3]Computer Vision Institute, School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[4]School of Computer & Software Nanjing University of Information Science & Technology

[*]qianchengshan@163.com

Abstract

This paper presents a new deep learning architecture for robust object representation, aiming at efficiently combining the proposed synchronized multi-stage feature (SMF) and a boosting-like algorithm. The SMF structure can capture a variety of characteristics from the inputting object based on the fusion of the handcraft features and deep learned features. With the proposed boosting-like algorithm, we can obtain more convergence stability on training multi-layer network by using the boosted samples. We show the generalization of our object representation architecture by applying it to undertake various tasks, i.e. pedestrian detection and action recognition. Our approach achieves 15.89% and 3.85% reduction in the average miss rate compared with ACF and JointDeep on the largest Caltech dataset, and acquires competitive results on the MSRAction3D dataset.

Key words: boosting, deep learning, object representation, synchronized feature

1 Introduction

Existing pattern recognition methods generally include two major steps: feature extraction and classifier design[1,2,40,39]. The quality of visual features is crucial for a wide range of computer vision topics, e.g., scene classification, object detection and human action recognition[44]. Handcrafted feature and learning based feature are two commonly used features feeding into a decision-making algorithm. Handcrafted features are usually inspired by the domain and the particular application, intending to capture certain morphological, statistical or texture attributes of objects. Therefore, the rules of handcrafted feature extraction are varied according to different application fields. Many famous handcrafted features, such as HOG [1], Haar-like [2], SIFT [3], covariance descriptors [4], integral channel features [5] and 3D geometric characteristic feature [6], have been successfully applied in pedestrian detection. Moreover, the spatio-temporal features [7, 8, 9, 10] and depth