Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 52 (2018) 144–154

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

SR-POD: Sample rotation based on principal-axis orientation distribution for data augmentation in deep object detection

Yue Xi [a], Jiangbin Zheng [a,*], Xiuxiu Li [b], Xinying Xu [c], Jinchang Ren [c,d,*], Gang Xie [e]

[a] School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[b] School of Computer Science, Xi'an University of Technology, Xi'an, China
[c] College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China
[d] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[e] College of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan, China

Received 1 April 2018; received in revised form 13 May 2018; accepted 25 June 2018
Available online 4 July 2018

Abstract

Convolutional neural networks (CNNs) have outperformed most state-of-the-art methods in object detection. However, CNNs suffer the difficulty of detecting objects with rotation, because the dataset used to train the CCNs often does not contain sufficient samples with various angles of orientation. In this paper, we propose a novel data-augmentation approach to handle samples with rotation, which utilizes the distribution of the object's orientation without the time-consuming process of rotating the sample images. Firstly, we present an orientation descriptor, named as "principal-axis orientation", to describe the orientation of the object's principal axis in an image and estimate the distribution of objects' principal-axis orientations (PODs) of the whole dataset. Secondly, we define a similarity metric to calculate the POD similarity between the training set and an additional dataset, which is built by randomly selecting images from the *benchmark ImageNet ILSVRC2012* dataset. Finally, we optimize a cost function to obtain an optimal rotation angle, which indicates the highest POD similarity between the two aforementioned data sets. In order to evaluate our data augmentation method for object detection, experiments, conducted on the *benchmark PASCAL VOC2007* dataset, show that with the training set augmented using our method, the average precision (AP) of the Faster RCNN in the TV-monitor is improved by 7.5%. In addition, our experimental results also demonstrate that new samples generated by random rotation are more likely to result in poor performance of object detection.

Keywords: Data augmentation; Deep object detection; Deep learning; Object rotation

1. Introduction

Convolutional neural networks (CNNs), typically trained on large scale of data, have significantly advanced the performance of the solutions to various vision problems such as object detection (Cheng and Han, 2016; Girshick et al., 2014; Girshick, 2015; Ning et al., 2018; Zhang et al., 2013), salient object detection (Cao, Tao,

* Corresponding authors at: School of Computer Science, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an 710072, China (J. Zheng). Department of Electronic and Electrical Engineering, University of Strathclyde, 204 George Street, Glasgow G1 1XW, United Kingdom (J. Ren).

E-mail addresses: zhengjb@nwpu.edu.cn (J. Zheng), jinchang.ren@strath.ac.uk (J. Ren).

https://doi.org/10.1016/j.cogsys.2018.06.014
1389-0417/

Zhang, Fu, & Feng, 2014; Han et al., 2015; Huang, Feng, & Sun, 2017; Wang et al., 2018; Yan et al., 2018) and object recognition (Dixit et al., 2017; Ren et al., 2015; Wen et al., 2017; Zheng et al., 2016). Performance of CNNs on object detection is largely attributed to networks with millions of parameters, which have a strong ability to extract rich, high-level object representation features. In order to ensure good performance, CNNs must be trained on large scale of data. Otherwise, if being trained on limited data they tend to lead into overfitting. Unfortunately, labelling a large number of images is not only time consuming and tedious, but also requires professional knowledge and skills (Zhou et al., 2017). Moreover, there are few (or no) training instances in some situations, such as a few-shot or zero-shot learning scenario (Zhang et al., 2018). To mitigate the issue to some extent, an effective technique called data augmentation, which generates real samples, is widely adopted to extend the training data (Lv et al., 2017).

Research has been devoted to data augmentation. Existing methods can be grouped into two categories, i.e., the guided methods and non-guided methods. In the guided methods such as in Kittler, Huber, Feng, Hu, and Christmas , Huber, Feng, Hu, and Christmas (2016), Seyyedsalehi and Seyyedsalehi (2014) and Szegedy(2015), external information (e.g., background, texture, noise) is introduced to build crafted feature descriptors or 3D models. Whereas the non-guided methods, such as (Krizhevsky et al., 2012; Zeiler and Fergus, 2014) perform affine or non-linear transformation (e.g. flipping, cropping, rotating, adding noise) on existing samples. Compared with the guided methods, non-guided methods do not need to build complex data models, so they are easy to implement and also computationally less expensive. However, the non-guided approaches require certain prior knowledge of data; otherwise, they tend to produce fake samples, which do not exist in the real world. These samples, instead of enhancing the training, would slow down the training and lead to samples misclassified during training and testing.

Due to varying viewpoints and object orientations in 3D world, changes in object orientation of 2D image are ubiquitous (Jaderberg et al., 2015). In order to deal with object rotation, we propose a novel data augmentation method guided by the distribution of object orientation. Our work is based on a simple observation, i.e., when the distribution of object orientation of the training set is similar to that of the testing set, the training set can sufficiently cover the variability (i.e. object orientation) in the testing set. In addition, we integrate guided and non-guided methods into our framework, as shown in Fig. 1. Specifically, firstly, we present an orientation descriptor to estimate the principal-axis orientation distribution (POD) from dataset. Secondly, we design a similarity function to quantify the similarity of the POD between the training set and the testing set. Finally, a cost function is designed to enforce the POD of the training samples to be similar to that of the testing set after rotating. Moreover, we optimize the cost function to obtain an optimal rotation angle, at which the highest similarity is achieved between the two sets in terms of POD.

Our major contributions presented in this paper are in four folds: (1) We explore a novel data augmentation framework, which extending training set achieves a better coverage of objects' varying orientations in testing data, to improve the performance of CNNs on object detection; (2) We design a principal-axis orientation descriptor based on superpixel segmentation to represent the orientation of an object in an image; (3) We propose a similarity measure method of two datasets based on principal-axis orientation distribution; and (4) We evaluate the performance of CNNs on object detection with and without rotating images in testing set.

2. Related works

The data augmentation technique extends training set through generating new samples with better coverage of variability in testing set. According to whether there is external data introduced, existing approaches can be grouped into non-guided methods and guided methods.

Non-guided methods mean that no external data is introduced during the generation of synthetic samples. Krizhevsky was the first to employ affine transformations (flipping, rotating and PCA-based intensity transformation) to increase the size of training set for learning deep models in order to avoid overfitting caused by a small set of training samples (Krizhevsky et al., 2012). This method could effectively increase the variability in orientation and images' colour space in the training set. Zeiler and Fergus (2014) adopted it due to the low computational cost. Vincent et al. added Gaussian noise to original images

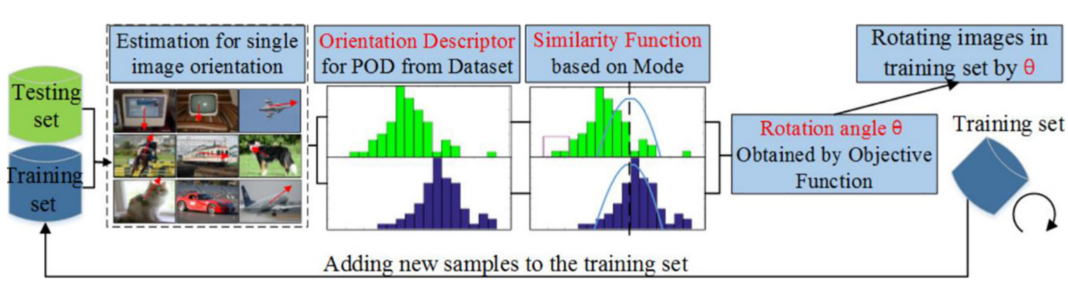

Fig. 1. Illustration of our proposed data augmentation algorithm.