Deeply Exploit Depth Information for Object Detection

Ruyue Han

August 21, 2018

Abstract

This paper addresses the issue on how to more effectively coordinate the depth with RGB aiming at boosting the performance of RGB-D object detection.At the very first, author investigate two ideas under the CNN model: property derivation and property fusion.Firstly,author proposed that the depth can be utilized not only as a type of extra information besides RGB but also to derive more visual properties ,author constructed a two-stage learning framework.Secondly,they explore the fusion method of different properties in feature learning,from which layer the properties should be fused together.The analysis shows that different semantic properties should be learned separately and combined before passing into the final classifier.

1. Introduction

Actually, the depth has some profitable attributes for visual analysis, e.g. being invariant to lighting or color variations, and providing geometrical cues for image structures [3].In this paper, author adopt CNN to extract rich features from the RGB-D images, i.e. author are under the CNN model to investigate the exploitation of the depth information.For the RGB-D object detection with CNN, the key is how to elegantly coordinate the RGB with depth information in feature learning. In the previous literatures, some intuitive methods have been proposed [1, 2].Roughly, previous literatures can be divided into two broad categories according to the strategy the depth is treated.The first one is to straightforwardly add the depth map to CNN as the fourth channel along with the RGB.The second is to process the color and depth separately, and they are combined before being fed into the final classifier.Author proposed a novel method to deeply exploit the depth information for object detection. Figure. 1 illustrates the main ideas of author's method. Firstly, various visual property maps are derived through analyzing the provided color and depth pairs.It is believed that more properties can contribute to the accurate description of the object and thus help boost the detection performance. Specifically, the derived properties include the contour, height, and angle maps 1 . Secondly, author systematically investigate the method to fuse different visual properties under the CNN model, i.e. how to represent a property, and from which layer the properties need to be fused together.

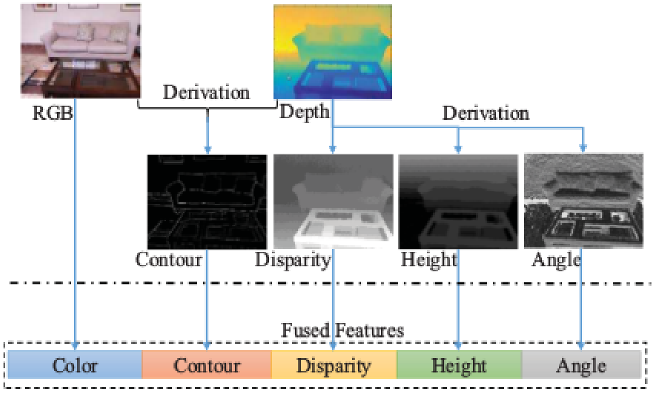

Figure 1. Illustration of learning rich features for RGB-D object detection. Various property maps are derived to describe the object from different perspectives. The features for these maps are learned independently and then fused for the final classification. Specifically, the derived maps include geometry contour from the color/depth pairs, and horizontal disparity, height above ground, angle with gravity from the depth data. These maps, as well as the RGB image, are sent into different CNNs for feature learning. And the features are joint before being fed into the classifier.

References

[1] C. Couprie, C. Farabet, L. Najman, and Y. LeCun. Indoor semantic segmentation using depth information. *CoRR*, pages 1301–3572, 2013. 1

[2] S. Gupta, R. Girshick, P. Arbelez, and J. Malik. Learning rich features from rgb-d images for object detection and segmentation. *ECCV*, 2014. 1

[3] R. Socher, B. Huval, B. Batha, C. D. Manning, and A. Y. Ng. Convolutional-recursive deep learning for 3d object classification. *NIPS*. 1