Neurocomputing 215 (2016) 225–231

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust vehicle detection by combining deep features with exemplar classification

Liujuan Cao [a,b], Qilin Jiang [a,b], Ming Cheng [a,b,*], Cheng Wang [a,b]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, Fujian, China
[b] School of Information Science and Engineering, Xiamen University, Xiamen, China

ARTICLE INFO

Article history:
Received 15 February 2015
Received in revised form
13 March 2016
Accepted 23 March 2016
Available online 23 June 2016

Keywords:
Superpixel segmentation
SLIC
Deep Neural Network
Exemplar SVMs
Vehicle detection
Robust classification

ABSTRACT

Very recently, vehicle detection in satellite images has become an emerging research topic with various applications ranging from military to commercial systems. However, it retains as an open problem, mainly due to the complex variations in imaging conditions, object intra-class changes, as well as due to its low-resolution. Coming with the rapid advances in deep learning for feature representation, in this paper we investigate the possibility to exploit deep neural features towards robust vehicle detection. In addition, along with the rapid growth in the data volume, new classification methodology is also demanded to explicitly handle the intra-class variations. In this paper, we propose a vehicle detection framework, which combines Deep Convolutional Neural Network (DNN) based feature learning with Exemplar-SVMs (E-SVMS) based, robust instance classifier to achieve robust vehicle detection in satellite images. In particular, we adopt DNN to learn discriminative image features, which has a high learning capacity. In our practice, the leverage of DNN has achieve significant performance boost by comparing to a serial of handcraft designed features. In addition, we adopt E-SVMs based robust classifier to further improve the classification robustness, which can be considered as an instance-specific metric learning scheme. By conducting extensive experiments with comparisons to a serial of state-of-the-art and alternative works, we further show that the combination of both schemes can benefit from each other to jointly improve the detection accuracy and effectiveness.

1. Introduction

Detecting vehicles in satellite image is an important task, with various applications ranging from urban vehicle management to intelligent transportation system and traffic surveillance. However, it retains as an open problem so far, mainly due to the fact that vehicles in satellite images occupy less pixels, while usually have been occluded by backgrounds like trees and buildings. Especially, the background might be extremely complicated, such as in regions like cities or highways. In addition, objects occurring on the top of the buildings might also cause false alarms during vehicle detection.

To tackle this problem, learning based methods have been very popular in recent years for vehicle detection, in either aerial images or satellite images. In general, such methods first extract features from aerial or satellite images, and then train classifiers to change the detection as a binary classification problem. In the literature, there are lots of related works that have been done for the aforementioned classification-based vehicle detection. And many different features, as well as their combinations, are used in these works, For example, Liang et al. [1] adopted generalized multiple kernel learning to deal with vehicle detection, which combines HOG feature with selected Haar features extracted via cascade boosting. For another example, Kembhavi et al. [2] combined HOG features, color probability maps and pairs of pixel together. Then they used Ordered Predictors Selection to select features. Finally, the vehicle detection is done by building a partial least square model. In [3], Cheng et al. proposed a pixel based vehicle detection method. The features used in [3] contain edge, corner, they transform the color space and then do color classification. Such features are sent to a Dynamic Bayesian Network to train vehicle detector. In [4], Tuermer et al. used HOG features and Disparity Maps, which is combined with global Digital Elevation Model (DEM) to precisely extract road segments for detection. Tan et al. [5] and Eikvil et al. [6] also used the road information to assist vehicle detection. And Leitloff et al. [7] proposed to detect vehicle queue, they considered most vehicles line in a queue in city satellite image (especially vehicle on road) to assist detection.

* Corresponding author at: Fujian Key Laboratory of Sensing and Computing for Smart City, Fujian, China.
E-mail address: chm99@xmu.edu.cn (M. Cheng).

http://dx.doi.org/10.1016/j.neucom.2016.03.094
0925-2312/

Then, such methods used adaptive boosting in combination with Haar-like features to detect vehicles.

Different from the above handcraft designed features, more recently in the literature of general object detection, learning based features have shown extremely promising performance. However, such features are not adopted in the problem of vehicle detection so far. especially in the satellite image domain. For instance, Krizhevsky et al. [8] proposed to train a deep convolutional neural networks for image classification task of ImageNet [9]. In [8], the features extracted from images are based on network outputs, which is also fused with the step of classification training to reinforce each other. And Chen et al. [10] proposed to train a deep convolutional neural networks. In this network, each image patch is an input for the network. The last layer in this network is then changed as features at different scales. Then, sliding windows based approach is further adopted to locate the vehicle. In particular, the deep convolutional neural networks can be regarded as two parts: The fist can be considered as a feature extractor, while the second can be considered as a classifier. As an improved version, in [11], Chen et al. further proposed to train a parallel version of the proposed deep convolutional neural network. For another instance, Malisiewicz et al. [12] proposed to use Exemplar-SVMs (ESVM) for object detection. In this method, exemplar SVM is trained for each positive example and millions of negative examples. Then, such SVMs are ensembled to do the object detection. In additional, hypergraph learning scheme is proposed in [13] for the task of semantic segmentation of potential objects from hyperspectral images.

In this paper, we proposed a vehicle detection scheme that combines DNN based features with Exemplar SVMs based classifier. In our consideration, the DNN based feature provides strong and discriminative feature representation, while Exemplar SVMs based classifier provides robust classification.

The rest of this paper is organized as follows: In Section 2, we introduce our detection method, including both DNN based features and Exemplar SVMs based classifier. Experimental settings and implementation details are shown in Section 3. Extensive experiments are given in Section 4. In Section 5, we summarize our conclusion and discuss our future work.

2. The proposed method

In the following subsections, we describe the main procedures of our vehicle detection system. Fig. 1 presents the pipeline of the proposed vehicle detection framework. In principle, the detection system works as follows: When a new satellite image coming in, we first use SLIC to over-segment this image into a set of superpixel patches. Then, such patches are sent to DNN based network to extract learning-based feature representation. Such features are subsequently sent to E-SVMs based classifiers to classify whether a given superpixel region is vehicle or not. Finally, a Non-Maximum Suppression operation is done to eliminate overlapped detection regions.

2.1. SLIC-based superpixel segmentation

Traditional schemes in vehicle detection typically resort to a sliding windows based setting, which is indeed very time-consuming. In this paper, we instead adopt a superpixel-level scanning to significantly reduce the time cost. In our setting, we notice that most vehicles in satellite images have strong edge responses. Therefore, it is more likely that the vehicles occur in homogeneous regions, which indicate the possibility of using superpixel as the basic unit in vehicle detection. In particular, we adopt SLIC based superpixel segmentation scheme proposed by Achanta et al. [14]. SLIC converts the target image into a large amount of superpixels.

Note that such superpixels are typically irregular. Subsequently, we further convert and adjust the shapes of such superpixels into regular ones. This is done by calculating the center of each superpixel and then extracting a fixed size of patches based on this center. The results after this step consists of the basic unit for the subsequent DNN+E-SVMs based vehicle detection.

2.2. DNN-based feature extraction

The Deep Convolutional Neural Networks is among one of the most popular Deep Learning models. The Convolutional Neural Networks (CNNs) can be considered as a special kind of multi-layer neural networks. CNN is trained by using an improved version of back propagation. Convolutional Neural Networks can be designed to recognize visual patterns directly from image pixels. It extracts image features while the pixels forward the network.

Among various implementations and variations of CNN features, LeNet5 proposed by LeCun et al. [15] is among the most popular and widely used ones, which was previously designed for the task of document recognition [16,17]. The learning capacity of LeNet5 and other CNN models are very flexible by adapting the depth and breadth of the network structure. For instance, Krizhevsky et al. [8] used a Deep Convolutional Neural Network to achieve the best result in ImageNet classification challenge.

In particular, we adopt the CAFFE implementation provided by Jia et al. [18] in this paper, whose structure is shown in Fig. 2. More specially, the CAFFE model contains three convolutional layers and three max-pooling layers. It adopts ReLU and Local Response Normalization (LRN) mentioned in [8] in the network design, which provide a faster learning ability than others.

We train the DNN model by providing a set of training instances. Each training instance include a vehicle/non-vehicle patch together with its label (0/1). We randomize the parameters in DNN model at first. Then, the labeled data will update the networks time by time until the termination condition is satisfied. In each

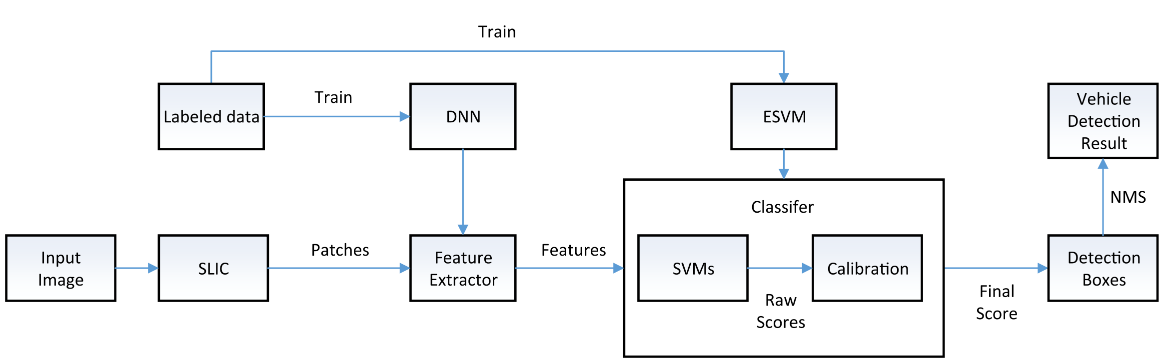

Fig. 1. This whole flowchart of the proposed DNN+E-SVMs based vehicle detection system. This system contains two parts, i.e., the training part and the detection part.