Pattern Recognition Letters 98 (2017) 16–25

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Context-based abnormal object detection using the fully-connected conditional random fields

Jongsuk Oh, Hong-In Kim, Rae-Hong Park*

Department of Electronic Engineering, Sogang University, Seoul, 04107, Republic of Korea

ARTICLE INFO

Article history:
Received 28 December 2016
Available online 5 August 2017

Keyword:
Abnormal object detection
Contextual model
Conditional random field
Embedding location
Mean field approximation

ABSTRACT

The contextual information plays an important role in computer vision, particularly in object detection and scene understanding. The existing contextual models use only the relationship between normal objects and natural scenes, and thus there still remains a difficult problem in detection of abnormal objects. This paper proposes an abnormal object detection model using the fully-connected conditional random fields to integrate the contextual information such as the co-occurrence and geometric relationships between objects. With this formulation, the proposed model combines the co-occurrence, spatial interaction between objects, and scale information. To this end, we use a feature embedding technique to find a geometry that reflects the statistical relationship in the pairwise term. Abnormal object detection is solved by using probabilistic variational inference such as the mean field approximation. Experimental results show that the proposed abnormal object detection model achieves significant improvement over the state-of-the-art models on the out-of-context dataset and abnormal object dataset.

1. Introduction

Contextual information plays an important role in computer vision, especially, in object detection and semantic segmentation fields [1–4]. Using the contextual information, object detection and semantic segmentation have made a quantum leap. This is because the contextual information provides the information of whether a scene and an object are related to each other using the co-occurrence relationship between objects within a scene, or relative position and scale of objects. For example, there is a sofa in a living room in the indoor scene, but a car does not appear. Another advantage of using the contextual information such as the co-occurrence, relative position, and relative scale relationships between the objects helps to interpret a scene, and to remove false positives [5].

However, there still remains a difficulty, especially, in detecting and recognizing of *abnormal objects* that are unexpected objects in a scene. Since most of the existing object detection models focus on detecting normal objects to increase the performance of object detection by simply considering the normal contextual information, they cannot detect abnormal objects. Using the information that abnormal objects have small context scores, abnormal objects can be detected in a single image. The contextual information helps to detect abnormal objects.

Only a few papers on abnormal object detection have been proposed using the contextual information [5–8]. Choi et al. proposed a tree-based context model via latent co-occurrence and support tree structures [5,6]. The tree structure has a limit in expressing the fully-connected relationships between objects because of considering the parent-child relationships only. Park et al. proposed a generative model to generate both normal and abnormal objects using the contextual information of the canonical scene [7], which represents the configuration of the normal objects such as location distributions of objects in a scene. The drawback of the canonical scene takes account of only the location distribution of the object itself in the scene, without considering the relationships between objects. The previous work does not consider all the relationships between objects. Cao et al. proposed a high-order contextual descriptor that incorporates the contextual information such as semantic, spatial, and scale contexts [8]. Finding the fully-connected pairwise links between the detected objects, or high-order interaction, this method can provide dependencies among objects in an image and detect out-of-context objects. As in [5], they used the Gaussian distributions to take account of the relative position using the vertical position and depth information. However, the Gaussian distributions do not adequately describe the geometric relationships between objects. In contrast to Cao et al.'s method, we consider the geometric information using the support relationship as in [6].

* Corresponding author.
E-mail addresses: ojs8686gn@sogang.ac.kr (J. Oh), blacklleye@sogang.ac.kr (H.-I. Kim), rhpark@sogang.ac.kr (R.-H. Park).

http://dx.doi.org/10.1016/j.patrec.2017.08.003
0167-8655/

Other abnormal object detection methods have been proposed [9,10]. Saleh et al. proposed object-centric anomaly detection using attributes of the objects such as part attributes [11], shape, and color [9]. They extended attributed-based reasoning to scene-centric, context-centric, and object-centric reasoning for abnormality [10]. They also classified taxonomy depending on the reasons of abnormality in images such as scene-centric, context-centric, and object-centric reasoning. However, the proposed model does not focus on detecting or classifying object-centric abnormal objects.

Note that this paper is inspired by the fully-connected conditional random fields (CRFs) [12], which was used for object detection [13,14] and semantic segmentation [15–17]. Nematollahi et al. proposed a new context-based fully-connected CRF model by adding a hidden node that describes the overall context of an image to segment an image semantically [15]. Our goal is to label object candidates as normal or abnormal objects, rather than to label the pixel as in semantic segmentation.

In this paper, we propose a new approach to *abnormal object detection*. Unlike existing models, the proposed model takes account of the relationships between objects and object-scene as the fully-connected relationships for the co-occurrence and support relationships. The proposed model solves the abnormal object detection problem by inferring the labels of object candidates, e.g., normal or abnormal objects. To this end, the proposed model consists of two steps. First, we use an off-the-shelf detector such as a deformable part model (DPM) [1] to generate a pool of object candidates. Then, considering dependencies between objects and a scene, we build a fully-connected CRF for multi-class object labels [12], where nodes represent labels of the object candidates, and edges encode dependencies between object candidates. We also construct a fully-connected CRF to label abnormal objects. To detect abnormal objects, we extend the context-based fully-connected CRF [15]. Since we focus on detecting the context-violating abnormal objects, we need information on which objects violate the contextual information. Unlike the context-based fully-connected CRF with a context node, the proposed model constructs two fully-connected CRFs with the same number of nodes. In the proposed fully-connected CRFs, the object class nodes correspond to a context node, while abnormality nodes match object label nodes. Through variational inference such as the mean field approximation, we predict the labels of the object candidates and abnormal object candidates. Since the co-occurrence correlation and geometric information are represented using a statistical approach, it is difficult to directly use Gaussian kernels that are employed to efficiently infer the fully-connected CRFs. Therefore, we use an embedding technique to take account of dependencies between objects in the Euclidean feature space [18]. We embed the co-occurrence and support relationships between objects into the Euclidean feature space.

We use the SUN09 dataset [5] to train the contextual information of the normal objects. We also evaluate the proposed abnormal object detection model on three public datasets: the out-of-context dataset [6], the abnormal object dataset [7], and the previous out-of-context dataset [5]. Experimental results show that the proposed model outperforms the state-of-the-art models on three public datasets.

Two main contributions of this paper are described as follows:

1) To the best of our knowledge, we first apply the fully-connected CRFs to detect abnormal objects. The advantage of the fully-connected CRF is to consider the fully-connected relationships between nodes and to efficiently perform inference on the fully-connected graph structures using the variational inference method, e.g., mean field approximation.
2) We also use co-occurrence and support features in the Euclidean feature space. It means that we can use Gaussian kernels in pairwise terms and utilize the efficient mean field inference algorithm.

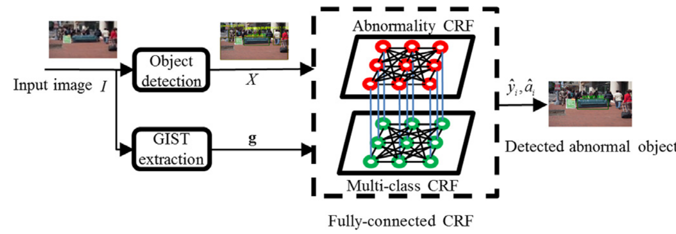

Fig. 1. Overview of the proposed model. The proposed model is composed of two steps. Given an input image I, we generate object candidate X using the DPM. Then, we build the fully-connected CRF models. Through inference, we detect abnormal objects.

This paper is organized as follows. Section 2 describes the proposed model. Experimental results and discussions are given in Section 3. Finally, Section 4 concludes the paper.

2. Proposed abnormal object detection

2.1. Generating object candidates

Fig. 1 shows an overview of the proposed model. Given an input image, the proposed model generates a pool of object candidates in the terms of bounding boxes by applying a pre-trained detector, DPM. A pool of object candidates is denoted by $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, where N is the number of object candidates. The ith object candidate is denoted by $\mathbf{x}_i = [c_{i,k}, \mathbf{p}_i, s_i]^T$, where $c_{i,k}$, $\mathbf{p}_i$, and s_i represent the object class, the position vector of a bounding box, and the score of the ith object candidate, respectively. Although the proposed model uses the DPMs to fairly compare the proposed model with the existing models, we can use the state-of-the-art models for object detection such as the region based convolutional neural network [19].

2.2. Multi-class CRF model for objects

If the object candidate pool X is given, our goal is to assign each object candidate to either object class or background. Co-occurrence and support configurations are related to dependencies between objects. To capture the dependencies between object candidates, a fully-connected CRF is defined over a set of the label variables $Y = \{y_i, \ldots, y_N\}$, where each variable takes a value from a set of object labels $C = \{c_1, \ldots c_K\}$ and 0 (background), i.e., $y_i \in \{\{0\} \cup C\}$, and K is the number of object classes. In addition, the proposed model introduces a set of abnormal indicator variables $A = \{a_1, \ldots, a_N\}$. The set A takes values from the abnormality label $a_i \in \{0, 1\}$, with a_i representing whether a given object candidate $\mathbf{x}_i$ is an abnormal object or not. If $a_i = 1$, the detected object is an abnormal object. The conditional probability of the joint distribution over object label variables Y and abnormal indicator variables A given object candidates X is expressed as

$$P(Y, A|X) = P(Y|X)P(A|Y, X) = \frac{1}{Z_{obj}(X)} \exp\left(-E_{obj}(Y|X)\right) \frac{1}{Z_{abn}(X)} \exp\left(-E_{abn}(A|Y, X)\right), \tag{1}$$

where $E_{obj}(Y|X)$ is a multi-class energy function that assigns object candidates to the object classes, and $E_{abn}(A|Y, X)$ represents an abnormality energy function. $Z_{obj}(X) = \sum_Y \exp(-E_{obj}(Y|X))$ and $Z_{abn}(Y, X) = \sum_A \exp(-E_{abn}(A|Y, X))$are partition functions for multi-class CRF and abnormality CRF, respectively.