Neurocomputing 287 (2018) 221–231

Contents lists available at ScienceDirect

Neurocomputing

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

Noise-aware co-segmentation with local and global priors

Qingqun Ning[a], Zhao Liu[a], Jianke Zhu[a,b,*], Mingli Song[a,b], Jiajun Bu[a,b], Chun Chen[a]

[a] College of Computer Science and Technology, Zhejiang University, 38 Zheda Road, Hangzhou, Zhejiang Province, PR China
[b] Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, 38 Zheda Road, Hangzhou, Zhejiang Province, PR China

ARTICLE INFO

Article history:
Received 21 November 2016
Revised 11 November 2017
Accepted 1 February 2018
Available online 8 February 2018

Communicated by Dr. Kaizhu Huang

Keywords:
Co-segmentation
Semantic proposal
Attentiveness
Local/global prior
Dense correspondence mapping

ABSTRACT

Image segmentation is a long-standing challenge in image and video processing. The method of co-segmentation aims at discovering common foreground object shared in image set. The traditional co-segmentation methods usually assume that all images should contain the target object. In this paper, we perform co-segmentation by first refining the image set. To this end, we propose to use attentiveness score, which is built upon the semantic proposals to identify the target object. We further filter out the noisy images using affinity propagation clustering. Then, both local and global shape priors are computed from the cleaned image set. The local prior can accurately estimate the foreground boundary, and the global prior supervises the pose and viewpoint of target object. These priors are optimized via dense correspondence mapping. Finally, we perform co-segmentation by minimizing an energy function. Experiments on three testbeds including Graz02, Internet images and MSRC object dataset, demonstrate that the proposed method outperforms the state-of-the-art co-segmentation methods.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Image segmentation has been a fundamental problem in computer vision. Many approaches [1–3] have been proposed to deal with this problem. Due to the interior complexity in both object and context world, it is still now a challenging problem. One of these methods is to simultaneously segment common objects from a collection of images, which is well-known as co-segmentation.

The early image co-segmentation methods [3–7] mainly focus on segmenting image pair given by the user. Many existing methods [8–13] for multiple image co-segmentation usually assume that the input images are clean, i.e., the co-segmentation of multiple photos taken at a certain period. In most of cases, we suppose to perform co-segmentation on ordinary image sets, where the common objects often vary in shape, pose, color and viewpoint. Due to the semantic similarity between the foreground and common object, noisy images without the region of interest are inevitably involved. Fig. 1 shows an example of image of the automobile interior. To deal with this problem, we propose to first find out the clean images before co-segmenting the common objects.

Recently, some methods have been proposed to solve the co-segmentation with noisy images. Rubinstein et al. [14] tried to establish a shape prior based on the saliency of each image. Chen et al. [15] divided a large set into many small subcategories, and build a shape template based on the latent-SVM detector for each subcategory. These methods are not scalable for a large set of input images. As in [14], their method requires a cluster with 36 CPU cores, which greatly limits its application domain. Since neither the saliency nor the bounding box detected by latent-SVM is able to approximate the true boundary of the foreground object, it may lead to the overfitting issue. Moreover, the salient part of an image may not always reflect the desired object region. Alternatively, Wang et al. [16] directly estimated the common context of the input images by training an auto-context model. As the context of noisy image set is usually more complex than the foreground, it involves with a time-consuming iterative scheme.

We obtain a clean image collection by taking advantage of a two-step scheme: (1) to ensure that an image collection contains common foreground object, we need to find it out first. To this end, we introduce a measure called "attentiveness" on the object. We compute the attentiveness scores of different categories in the image collection so as to identify the target object; (2) affinity propagation clustering is employed to further purify the collection. The second step aims to cluster foreground objects into similar shape or viewpoint. This step is important due to the difference of the difficulty between these two cases: co-segmentation on images with front and side view of planes and co-segmentation on images with only side view of planes. One merit of our approach is that this process on image set is co-segmentation scheme-agnostic. It

* Corresponding author at: College of Computer Science and Technology, Zhejiang University, 38 Zheda Road, Hangzhou, Zhejiang Province, PR China.

E-mail addresses: ningqingqun@zju.edu.cn (Q. Ning), liuzhao@zju.edu.cn (Z. Liu), jkzhu@zju.edu.cn (J. Zhu), brooksong@zju.edu.cn (M. Song), bjj@zju.edu.cn (J. Bu), chenc@zju.edu.cn (C. Chen).

https://doi.org/10.1016/j.neucom.2018.02.018
0925-2312/© 2018 Elsevier B.V. All rights reserved.

Fig. 1. Our approach filters out noisy images based on the attentiveness of semantic proposals. We provide some examples of the prefiltering result on Internet image dataset. It can be seen that those images without target category have been successfully swept out.

implies that many previous methods can easily make use of this process to perform co-segment on ordinary image set.

Given a clean image set, the prevailing co-segmentation approaches [3,4,9] depend on the user-interaction. They require user scribbles to indicate the foreground and background. Since it is inefficient to manually label each image for performing co-segmentation, this limits the applying of these methods on large scale image set. Different from these methods, we propose to use shape priors to replace user scribbles as initialization. In contrast to the conventional methods using soft-boundary [17,18], transferred maps of nearest neighbours [19,20], and exemplar-based detectors [15,21] to model the shape prior, we try to simultaneously build both local and global priors to model the common object on the a whole input image set.

To this end, we take advantage of the semantic proposals from Simultaneous Detection and Segmentation (SDS) framework [22]. A local shape prior is constructed upon those semantic proposals from a single image, while a global one is built on the whole image set. Both two priors are conducted on uniform spatial proposals. Since these priors are relatively coarse representation of the common object, we employ dense correspondence mapping to further optimize them. Finally, a Markov Random Field (MRF) based energy functions is constructed using these two priors. The global prior can approximately estimate the viewpoint the common object, and the local prior captures foreground boundary. Our segmentation results are boosted via combining the information of these two priors. The co-segmentation is performed by minimizing the energy function of the graph cut [23–25]. The whole process of our proposed approach is fully automatic without human supervision. Thus, it is easy to use in practice.

We aim at simultaneously selecting the clean image from the noisy image set and segmenting the object region from them with local and global priors. In summary, the main contributions of this paper are: (1) a new approach to refine image set by first identifying the target object using a measure of attention. Thus, the noisy images are filtered out by the target object category, and then by clustering images into similar shape or viewpoint through affinity propagation; (2) a novel object-level prior that estimates the shape of object region, which is further optimized by the dense correspondence mapping. It is able to outperform the conventional priors using saliency or bounding box; (3) an energy-based formulation to co-segment the target object that takes advantage of both local and global shape prior information.

2. Related work

Co-segmentation is an interesting problem in computer vision [5]. The early methods [3,4] mainly focus on finding the segmentations with the similar foreground in image pairs, where the user interaction is typically used. Recently, Wang and Shen [9] employed higher-order energy based function for co-segmentation on multiple images, which still relies on user-interaction.

In recent years, some methods that do not need user interaction are proposed. Firstly, researchers focus on the problem with multiple images rather than image pairs. Kim et al. [26] proposed an optimization scheme for a relatively large dataset. Kuettel and Ferrari [20] transferred the object-like area to the unknown images, which is further evaluated on ImageNet dataset [19]. Since the user interaction is not always available, some methods pay attention to the unsupervised setting. For instance, Joulin et al. [27] combined the techniques used in over-segmentation and kernel methods in a discriminative clustering framework, which is able to handle multiple images. Moreover, they extend this framework by an efficient EM optimization process. Thus, it can segment multiple instances in images [28]. Meng et al. [8] casted co-segmentation as the shortest path problem. Rubio et al. [10] presented an unsupervised method for segmenting object from multiple images. The main idea of their work is to build the correspondence between the similar superpixel regions in images, in which Gaussian Mixture Model (GMM) is employed to predict the pixel distribution. Similarly, Faktor and Irani [29] found the relationship between various compositions and estimate the shape prior based on the high relevant compositions. Most of the above methods assume that the foreground regions are similar, and can only deal with the clean data, i.e., all the images in the dataset must contain the target objects.

Recently, Kim and Xing [30] try to tackle the co-segmentation problem on the noisy web images with user-interactions, which is not fully unsupervised. Rubinstein et al. [14] use the pixel-wise SIFT-Flow to achieve the pixel-level correspondence between the normalized saliency to discover the object category, which is both memory and computational demanding. Instead of relying on the visual prior on the whole image set, Chen et al. [15] divide the image set into several subcategories, and model the visual priors for all of them. The shape priors are learned through a part-free latent SVM model. In a recent work, Meng et al. [31] design constraints to direct graph clustering so that the proposed method is able to filter the noisy data and focus on clustering the target images. Tang et al. [11] locate the object instances in each clean image, and then relax the co-segmentation problem into a quadratic program. Wang et al. [16] tried to solve object recognition and co-segmentation with noisy images in an iterative framework through an autocontext model, where training the auto-context model involves a time-consuming iterative process.

Currently, deep learning features are widely used in image classification. In contrast to the hand-crafted features, researchers are able to add semantic information to object segmentation instances. Girshick et al. [32] propose a semantic segmentation framework by training multi-way SVMs on the selective search windows and the selected CNN features. Hariharan et al. [22] extend it by replacing the selective search window with the grouped hierarchical segmentation [33]. These methods are able to find several foreground proposals in each image. In our proposed approach, we focus on predicting the common shape shared in a large scale image set. By building visual priors of different object categories in a dataset and choosing the most important object category, we can accurately filter out the noisy images and segment the target objects in the clean images.

3. Preprocessing with attentiveness

In this section, we describe in details the preprocessing stage of our co-segmentation method. We start by introducing the definition of attentiveness. Then, we describe how to obtain clean

Download English Version:

https://daneshyari.com/en/article/6864390

Download Persian Version:

https://daneshyari.com/article/6864390

Daneshyari.com