Journal of Visual Communication and Image Representation 55 (2018) 201–214

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Co-segmentation via visualization[☆]

Zahra Kamranian[a,b], Federico Tombari[b], Ahmad Reza Naghsh Nilchi[a,*], Amirhassan Monadjemi[a], Nassir Navab[b,c]

[a] Department of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Isfahan 8174673441, Iran
[b] Computer Aided Medical Procedures and Augmented Reality, Technische Universität München, Munich, Germany
[c] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Co-segmentation
Convolutional Neural Network (CNN)
Feature visualization
Occlusion sensitivity
Adaptive learning

ABSTRACT

This paper addresses the co-segmentation problem using feature visualization for CNNs. Visualization is exploited as an auxiliary information to discriminate salient image regions (dubbed as “heat-regions”) from non-salient ones. Region occlusion sensitivity is proposed for feature visualization. The co-segmentation problem is formulated via a convex quadratic optimization which is initialized by the heat-regions. The information obtained through the visualization is considered as an extra energy term in the cost function. The results of the visualization demonstrate that there exist some heat-regions which are not productive in the co-segmentation. To detect helpful regions among them, an adaptive strategy in the form of an iterative algorithm is proposed according to the consistency among all images. Comparison experiments conducted on two benchmark datasets, iCoseg and MSRC, illustrate the superior performance of the proposed approach over state-of-the-art algorithms.

1. Introduction and related works

The goal of *co-segmentation* [1] is to extract the foreground objects that are in common across a given set of images. In comparison with single image segmentation algorithms, co-segmentation relies on foreground similarity constraints, which have been shown to improve segmentation accuracy [1–3]. It can be applied to a variety of high-level vision applications such as image retrieval [1], medical imaging [4], image classification [5], object recognition [6], and 3D reconstruction [7]. In case of large variations of the common foreground and in the presence of complex background, co-segmentation is more challenging. This issue attracts particular attention in the computer vision community [8,9,3,10–12].

1.1. Motivation

In recent years, Convolutional Neural Networks (CNNs) have demonstrated impressive performance in a wide range of computer vision applications, including image classification, object detection, segmentation, object recognition, and depth estimation [13–17]. In the task of image segmentation, which aims at assigning a label to each pixel in a given image, the use of CNNs has brought in an improvement in segmentation accuracy of around 50% [15,16].

In this work, we examine whether CNNs can also be usefully employed for the task of co-segmentation. As a starting point, we evaluate the outputs of the pre-trained CNN proposed in [16] on some related images. Results are illustrated in Fig. 1. While all the images are related to one white/black dog in various situations, different segmentations are achieved in Fig. 1b.

This experiment illustrates that CNN has acceptable results in the segmentation task. However, since it does not exploit consistent information shared across the target images, there are some errors when the common object is not easy to localize within the image.

Co-segmentation datasets like iCoseg [9] only contain hundreds of images, which means that the pixel-wise annotations are too limited in number. Therefore, it is inappropriate to train a CNN from scratch on that data, given many parameters that have to be learned.

On the other hand, collecting a sufficient pixel-wise annotation for a set of images could be very costly [18,19]. In other words, creating an image dataset with class-level annotation involves much less effort than pixel-level annotation. For instance, there exist some handcrafted available image datasets, such as ImageNet [20], which include many images with class-level annotation. In this paper, the pre-trained CNN proposed in [14], which requires only class-level annotation, is considered for the co-segmentation task.

Our proposed algorithm relies on using class heat-maps of CNNs,

[☆] This paper has been recommended for acceptance by Hongliang Li.

[*] Corresponding author.
E-mail addresses: zahra.kamranian@eng.ui.ac.ir (Z. Kamranian), tombari@in.tum.de (F. Tombari), nilchi@eng.ui.ac.ir (A.R. Naghsh Nilchi), monadjemi@eng.ui.ac.ir (A. Monadjemi), navab@cs.tum.edu (N. Navab).

https://doi.org/10.1016/j.jvcir.2018.05.014
Received 13 September 2017; Received in revised form 6 May 2018; Accepted 19 May 2018
Available online 26 May 2018
1047-3203/

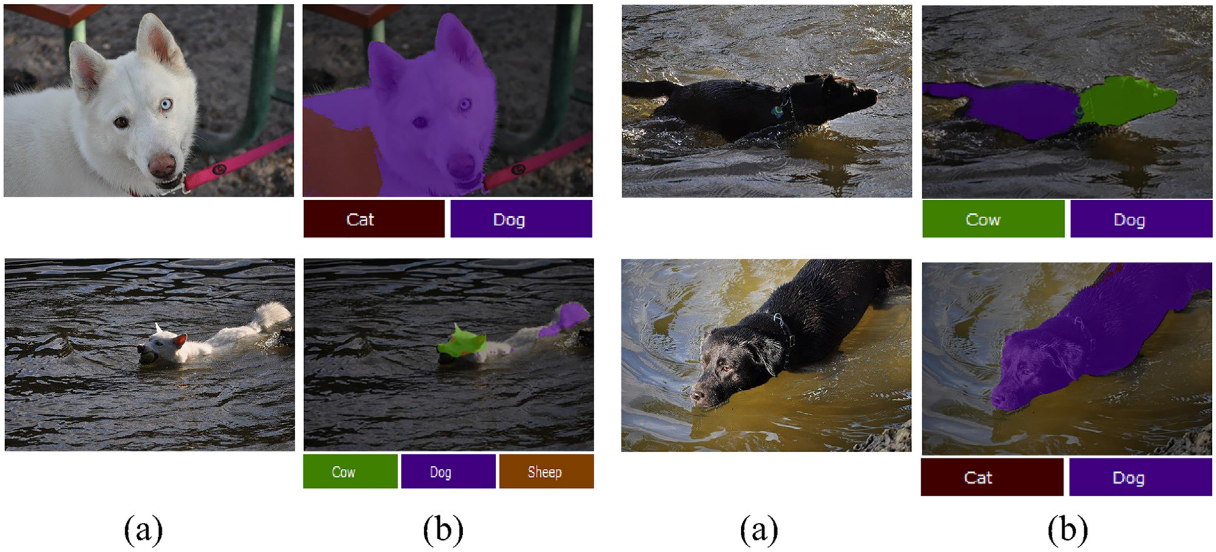

Fig. 1. CNN segmentation results. a) Original images. b) Segmentation results using the pre-trained model developed in [16].

(i.e., feature visualization for CNNs) to initialize co-segmentation. The intuitive purpose of using feature visualization is to utilize the class-level information to approximate the object location in each image of a group. This information can be effective to improve the co-segmentation. The task is performed by localizing the most relevant and distinctive image regions that drive the network to classify the input image into a certain class. To obtain such regions, we improve the method proposed by Zeiler & Fergus [21], where different portions of a given image are systematically occluded with a sliding gray square, and the output of the classifier is evaluated at every position of such square. It is shown that the probability of the correct class assignment drops significantly when the object is occluded. However, the result of the classifier depends on the size of the object in the input image, as well as on the size of the gray square. To address this issue, we propose a *region occlusion sensitivity* method, which can determine salient regions of the image more accurately without the needs to set the size of the gray square for each image.

The results of the feature visualization step are utilized as a prior information for the co-segmentation algorithm. That is, the important regions are located on the objects, and the areas which do not change the activations are on the background. However, there exist some regions which do not obey this rule. To handle the issue, we introduce an adaptive strategy to select the most effective regions. This strategy is an improvement over previous methods, such as [22,19], which use the areas obtained in the feature visualization step as raw information for the following steps.

As a result of our method, a highly accurate segmentation can be achieved via co-segmentation, which outperforms the state of the art on benchmark datasets.

1.2. Contributions

The main contributions of this study are summarized as follows:

1. A feature visualization method (by using CNN-based class heatmaps) is exploited to improve co-segmentation. It is used as a prior information as well as an extra energy for the co-segmentation cost function.
2. Region occlusion sensitivity is introduced for feature visualization. The model localizes the main regions in the images even with different object sizes.
3. An adaptive strategy is designed to detect correct foreground and background regions determined during the visualization step.

1.3. Related works

1.3.1. Co-segmentation

The current image co-segmentation algorithms can be categorized to unsupervised methods and interactive ones. The main idea of the most unsupervised co-segmentation techniques [1,4,8,6,23–25,12,26] is to formulate the problem as an energy optimization and binary labeling. For instance, Rother et al. [1] propose a cost function, including a Markov Random Field (MRF) term (as an intra-image energy) and a histogram matching term (as an inter-image energy) which penalizes the variation in the foreground histograms. Hochbaum and Singh [4] reward the consistency of two foreground histograms to simplify the energy optimization. Moreover, in addition to the intra-image and inter-image energies, Meng et al. [26] consider inter-group energy to solve the multiple image groups co-segmentation. Then, Expectation–Maximization algorithm is adapted to optimize the energy function.

Apart from MRF-based approaches, Joulin et al. [8] solve co-segmentation task via a combination of spectral algorithm and kernel method within a discriminative clustering framework. They extend their proposed approach to multi-class co-segmentation for a large number of images [27]. A diffusion-based optimization which exploits anisotropic heat diffusion method is proposed by Kim et al. [23]. Meng et al. [2] encode the co-segmentation as a shortest path problem, and design a di-graph to depict the local region similarities with respect to the feature distance and the saliency map. A multi-scale generative model proposed by Rubio et al. [24] estimates the distributions of foreground and background appearances, jointly. Rubinstein et al. [25] utilize SIFT flow to find dense correspondence among the images to capture the sparsity of a common foreground. In order to enhance the target consistency, Tao et al. [28] and Fu et al. [11] apply shape consistency for the common foreground and image depth, respectively. In [29], the foreground regions are identified by considering the intra-video foreground coherency and inter-video foreground consistency. Li et al. [12] and Quan et al. [30] introduce a multi-search strategy and a global loop-closure graph optimization approach, respectively. Han et al. [31] develop the method of [30] by proposing a new manifold ranking with the self-learned graph structure to improve the robustness in inferring the foreground and background probability of pixels.

In this study, to improve previous co-segmentation methods, we propose to exploit high-level information provided by deep feature visualization approaches. This information is used to initialize co-segmentation as well as an additional energy term in the definition of the