Medical Image Analysis 27 (2016) 3–16

Contents lists available at ScienceDirect

Medical Image Analysis

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

Detecting overlapping instances in microscopy images using extremal region trees

Carlos Arteta[a,*], Victor Lempitsky[b], J. Alison Noble[a], Andrew Zisserman[a]

[a] Department of Engineering Science, University of Oxford, Oxford OX1 2JD, UK
[b] Skolkovo Institute of Science and Technology (Skoltech), Skolkovo 143025 Russia

ARTICLE INFO

Article history:
Received 2 June 2014
Revised 20 November 2014
Accepted 3 March 2015
Available online 14 April 2015

Keywords:
Cell detection
Microscopy image analysis
Overlapping object detection

ABSTRACT

In many microscopy applications the images may contain both regions of low and high cell densities corresponding to different tissues or colonies at different stages of growth. This poses a challenge to most previously developed automated cell detection and counting methods, which are designed to handle either the low-density scenario (through cell detection) or the high-density scenario (through density estimation or texture analysis).

The objective of this work is to detect all the instances of an object of interest in microscopy images. The instances may be partially overlapping and clustered. To this end we introduce a tree-structured discrete graphical model that is used to select and label a set of non-overlapping regions in the image by a global optimization of a classification score. Each region is labeled with the number of instances it contains – for example regions can be selected that contain two or three object instances, by defining separate classes for tuples of objects in the detection process.

We show that this formulation can be learned within the structured output SVM framework and that the inference in such a model can be accomplished using dynamic programming on a tree structured region graph. Furthermore, the learning only requires weak annotations – a dot on each instance. The candidate regions for the selection are obtained as extremal region of a surface computed from the microscopy image, and we show that the performance of the model can be improved by considering a proxy problem for learning the surface that allows better selection of the extremal regions. Furthermore, we consider a number of variations for the loss function used in the structured output learning.

The model is applied and evaluated over six quite disparate data sets of images covering: fluorescence microscopy, weak-fluorescence molecular images, phase contrast microscopy and histopathology images, and is shown to exceed the state of the art in performance.

1. Introduction

Automatic detection of objects (e.g. cell colonies, individual cells or nuclei) in microscopy images plays a crucial role in the analysis of microscopy-based experiments within a wide variety of microscopy applications for clinical and commercial settings. On its own, detection is able to determine the presence (and quantity) of an object of interest, such as cancer cells in a pathology image, but furthermore, it can also be the starting point for other objectives such as object segmentation or tracking. Among the challenges that characterize object detection in microscopy images, one that stands out is the necessity to deal with the presence of a large number of objects, often partially overlapping.

In many microscopy imaging modalities, objects of interest can often be identified as bright or dark blobs in one of the image channels. Such blobs correspond to *extremal regions* (Matas et al., 2004), and a natural approach to detection and understanding such images is (a) to consider the set of all extremal regions and (b) to identify those extremal regions that actually correspond to objects of interest. This is the approach that we pursue in this work. Several key challenges need to be addressed to make this approach successful, namely:

- Each object of interest typically corresponds to multiple, very similar and overlapping, extremal regions. The challenge then is to pick a subset of regions corresponding to objects of interest, so that each object of interest is represented by only

---

[*] Corresponding author.
E-mail address: carlos.arteta@eng.ox.ac.uk (C. Arteta).

http://dx.doi.org/10.1016/j.media.2015.03.002

one region. We show how this can be achieved via organizing extremal regions into a tree-shaped (or forest-shaped) discrete graphical model with binary labels. Message propagation (dynamic programming) in such a tree (forest) then produces a desired subset of *non-overlapping* extremal regions corresponding to objects.
- In more challenging images, it is often the case that groups of tightly overlapping objects (i.e. cells in a dense cluster) cannot be distinguished on the basis of extremal regions. In other words, for certain objects, there might not exist extremal regions that include one object but exclude others in the same group. We show that the model can be extended to handle such challenging situations. The extended model is able to identify the blobs (extremal regions) that correspond to *multiple* overlapping objects, and to label simultaneously the selected regions with labels that indicate the number of objects that each selected extremal region corresponds to (Fig. 1). This extension greatly widens the applicability of the approach without changing the topology of the underlying graphical model or increasing the complexity of the inference.
- Apart from the model and the inference in the model, a key question is one of machine learning, i.e. a method to identify which extremal regions correspond to objects and which do not (and in the extended case, identify the number of objects within certain regions). We demonstrate that all this can be done in a *weakly-supervised* learning setting, so that the method is trained on a set of *dotted* representative images, where each object is annotated only by a dot placed inside of it. The training is performed using latent structured output support vector machines (Yu & Joachims, 2009) with a specially designed *counting* loss-function.
- Finally, we address the task of automatically identifying a “good” image channel that contains extremal regions that are “good” for our approach. Specifically, we propose a method that automatically optimizes over a linear combination of input channels (where some input channels can actually correspond to filtered versions of other channels) to determine an input image. After such an optimization, the resulting image gives rise to extremal regions that allows the efficient identification of individual objects or small groups of overlapping objects. This procedure only requires the same dot annotated images as above.

We conduct a set of experiments with synthetic and real microscopy images and show that the proposed method achieves very good detection accuracies despite large amounts of overlap and very low effective spatial resolution. We assess the effect of the different elements of our detection system and show that the combination of them results in the highest accuracy. The resulting system outperforms other methods for instance detection in microscopy images and is comparable in counting accuracy with the methods that are trained to count (and do not perform detection). While microscopic image modalities form a natural domain for our method, the proposed approach is general and can be applied to macroscopic medical or non-medical images, as demonstrated in Arteta et al. (2013).

This paper extends the previous conference papers (Arteta et al., 2012; 2013) that developed the initial approach. In comparison to the more recent conference version (Arteta et al., 2013), this paper adds the following extensions: (i) it develops in more detail the inference procedure on the tree-structured graphical model (Section 5); (ii) it provides further evaluation and insights into the loss function, along with a new variant of it (Section 6.1); (iii) it proposes a method for picking a linear combination of input channels that optimizes the method’s performance (Section 7); and (iv) it provides additional experiments with challenging microscopy images (Section 8).

2. Background

2.1. Instance detection in crowded scenes

Most computer vision methods that address the task of understanding images with multiple overlapping objects fall into two classes. The first is based on individual object detection. Such detection can be based on a sliding window or the Hough transform, followed by an appropriate non-maxima suppression procedure (Barinova et al., 2010; Desai et al., 2009; Leibe et al., 2008), stochastic fitting of interacting particles or object models (Descombes et al., 2009; Dong et al., 2007; Wu & Nevatia, 2009), or region-based detection (Arteta et al., 2012; Matas & Zimmermann, 2005; Neumann & Matas, 2011). The second class contains the methods that avoid the detection of individual instances but instead perform analysis based on local or global texture and appearance descriptors, e.g. by recovering the overall real-valued count of objects in the scene (Chan & Vasconcelos, 2009; Kong et al., 2006; Marana et al., 1997; Ryan et al., 2009) or by estimating the local real-valued density of the objects in each location of interest (Fiaschi et al., 2012; Lempitsky & Zisserman, 2010).

Depending on the degree of overlap between objects, the first or the second class of methods might be more appropriate. For low object-density images with infrequent overlaps between them, detection methods may perform very well, while regression/density estimation methods can e.g. hallucinate small but non-zero object density/object count spread across the parts of images that do not in fact contain any objects. Furthermore, the localization of individual objects in the detection-based approaches facilitates more intricate analysis by revealing patterns of co-location, providing the possibility for shape and size estimation of individual instances, and allowing the linkage of individual detections through time for video analysis.

However, for the high-density images detection-based analysis may fail badly, especially when the amount of overlap and interocclusion between objects makes the detection of individual instances hard or impossible even for human experts. In such situations, the performance of density/global count estimation methods degrade more gracefully than detection-based methods. The analysis in this case is essentially reduced to texture matching between the test image and the training set, which may be feasible even when individual instances are not distinguishable.

In real life, many applications require the processing algorithm to handle both the high and the low-density scenarios. Furthermore, the two cases may co-exist within the same image. For example, a microscopy image may contain both regions of low and high cell densities corresponding to different tissues or colonies at different stages of growth. Our proposed method aims to perform within such scenario, making the link between individual instance detection and instance counting in dense scenes.

2.2. Biological object detection in microscopy images

It is often convenient to model biological objects (i.e. cells) in microscopy images as regions of local minima or maxima in the intensity channels, thus common blob detectors (or related custom algorithms) have commonly been applied for this instance detection task (Al-Kofahi et al., 2010; Bernardis & Yu, 2011; Kuse et al., 2011; Peng et al., 2009; Smith & Lepetit, 2008; Wienert et al., 2012). A blob detector approach within the microscopy image analysis scenario also benefits from the fact that the common task can be seen as detection of multiple instance of the same object (i.e. all cells in the image of a cell culture will look similar). However, in non-trivial scenarios, such as cluttered images with cell overlap, or cases where cell discrimination is required, blob detectors do not have the flexibility to capture complex cell models, and thus, fail to achieve a good balance between sensitivity and specificity. Nevertheless, if they are sensitive