Towards Intelligent Visual Understanding under Minimal Supervision

Dingwen Zhang
School of Automation, Northwestern Polytechnical University
zhangdingwen2006yyy@gmail.com

Abstract

Because of playing one of the most important roles in the artificial intelligent systems like robots, visual understanding has gained vast interests in the past few decades. Most of the existing approaches need human labelled training data to train the learning models for visual understanding and in the most recent years, significant performance gain was obtained relying on unparalleled tremendous amount of human labelled training data. Under this circumstance, people are endowed with great burden to cost energy and time on the tedious data annotation for the traditional visual understanding approaches. To alleviate this problem, we propose to develop novel visual understanding algorithms which can learn informative visual patterns under minimal (none or very weak) supervision and thus facilitate higher-level intelligence of the visual understanding systems. Specifically, we focus on three subtopics, i.e., saliency detection, co-saliency detection, and weakly supervised learning based object detection, which can be used in both the image and video understanding systems. The experimental results have demonstrated the effectiveness of the proposed algorithms.

1 Introduction

The goal of AI in vision is to endow the machines the capability of understanding the content of visual stimulus like images and videos. Despite the success of some recent machine learning techniques, e.g., deep learning, the problems in visual understanding systems are still largely under-addressed in practice due to the heavy burden of labeling the training samples. Take object detection as an example. People need to take about 15s to draw the bounding box annotations that can enclose the objects of interest properly with adopting some assistive tools. The annotations of more delicate object boundaries certainly cost more time. Essentially, in such big data era, high-level intelligent visual understanding systems are desired to be capable of autonomously discovering the intrinsic patterns from the cheaply and massively collected visual data, rather than having great requirements of huge amount of finely demarcated manual annotations. Thus it motivates us to develop more intelligent visual understanding systems which can work under minimal (none or very weak) supervision and thus largely alleviate the human labors.

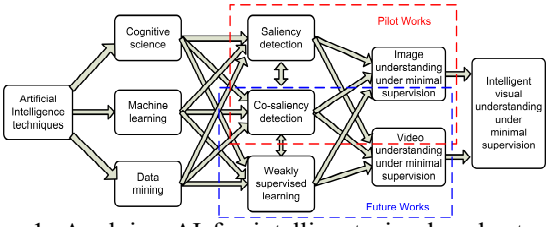

Figure 1: Applying AI for intelligent visual understanding under minimal supervision through saliency detection, co-saliency detection, and weakly supervised learning.

As shown in Fig. 1, we focus on three subtopics, i.e., saliency detection, co-saliency detection, and weakly supervised learning based object detection. On the one hand, we designed algorithms based on the AI techniques related to cognitive science, machine learning, and data mining to achieve good performances of saliency detection, co-saliency detection, and weakly supervised leraning based object detection. On the other hand, we applied the proposed saliency detection, co-saliency detection, and weakly supervised learning algorithms for realizing image and video understanding under minimal supervision. Finally, the higher-level visual understanding system can be established.

2 Pilot Methods

Saliency Detection: As shown in Fig. 2, the basic idea is to simulate the human visual attention mechanism to endow machine vision system the capability of predicting interesting regions from each single image autonomously. To solve this problem, in [Han et al., 2015], we proposed to apply stacked denoising autoencoders with deep learning architectures to model the background where latent patterns are explored and powerful representations of data are learned in an unsupervised and bottom-up manner. Afterward, we formulated the separation of salient objects from the background as a problem of measuring reconstruction residuals of the deep autoencoders. Comprehensive evaluations of three benchmark datasets and comparisons with nine state-of-the-art algorithms demonstrate the superiority of this work.

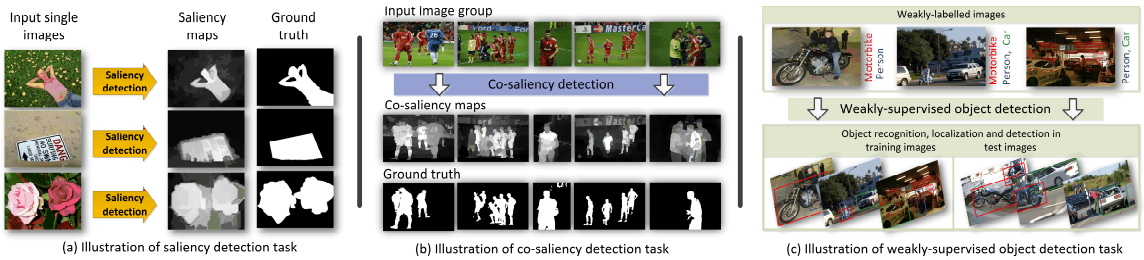

Figure 2: Illustration of the three investigated tasks for achieving visual understanding under minimal supervision.

Co-saliency Detection: By simulating the human attention ability to effectively identify common and salient objects among multiple related images, co-saliency detection has emerged to be an interesting research topic in recent years. As shown in Fig. 2, compared with the traditional saliency detection, co-saliency detection additionally explores the mutual information among multiple images/frames. Thus it can provide more useful information and generate more precise prediction for real-world applications.

To achieve this goal, we have addressed three critical problems in co-saliency detection, i.e., "what are the useful cues for co-saliency detection?", "how to explore such cues more effectively?", and "how to guide a robust unsupervised learning procedure?". Specifically, in [Zhang et al., 2015 a], we introduced two novel useful cues, i.e., deep and wide cues, for co-saliency detection. In [Zhang et al., 2015 b], the proposed intra-saliency prior transfer and deep inter-saliency mining are demonstrated to be powerful approaches to explore the information cues for co-saliency detection. In [Zhang et al., 2015 c], we proposed a novel framework based on the self-paced multiple instance learning regime, which was capable of fitting insightful metric measurements and discovering common patterns under co-salient regions in a robust self-learning way.

Weakly Supervised Learning: In the issue of weakly-supervised object detection (WOD), the key problem is to simultaneously infer the exact object locations in the training images and train the object detectors, given only the training images with weak image-level labels. To address this problem, we have proposed effective algorithms based on saliency detection and unsupervised feature learning technique. Specifically, [Zhang et al., 2015 d] proposed a saliency-based self-adaptive segmentation scheme to generate object candidates effectively, while [Han et al., 2015 b] extracted high-level feature representation via an unsupervised deep model and train weakly-supervised object detector via fusing saliency, intra-class compactness, and inter-class separability in a Bayesian framework. These algorithms have demonstrated to be effective in object detection systems for optical remote sensing images.

3 Future Work

Unconstrained Lare-scale Co-saliency Detection: One of our future works (submitted to IJCAI 2016) aims to design more intelligent algorithms to address the problems under much weaker assumption to simultaneously detect multi-class co-salient objects from such practical and cluttered image sets.

Bridge Saliency to Weakly Supervised Learning Based on Self-paced Curriculum: Saliency detection selects attractive objects in scenes and thus can provide useful priors for WOD. However, the way to adopt saliency detection in WOD is not trivial since the detected saliency region might be possibly highly ambiguous. To this end, another future work (submitted to IJCAI 2016) is to bridge saliency detection to WOD via the self-paced curriculum learning, which can guide the learning procedure to gradually achieve faithful knowledge of multi-class objects from easy to hard.

Video Understanding with Minimal Supervision: In the near future, we will make efforts on applying the proposed saliency, co-saliency, and WOD techniques to understanding the video content with minimal supervision.

Explore the nature relationship between saliency detection, co-saliency detection, and WOD: Finally, we will comprehensively analyze the relationships between the three investigated topics and propose a unified framework for minimizing the supervision in visual understanding.

References

[Han et al., 2015 a] J. Han, D. Zhang, X. Hu, L. Guo, F. Wu. Background Prior-Based Salient Object Detection via Deep Reconstruction Residual. *TCSVT*, 25(8): 1309-1321, 2015.

[Han et al., 2015 b] J. Han, D. Zhang, G. Cheng, et al. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. *TGRS*, 53(6): 3325-3337, 2015.

[Zhang et al., 2015 a] D. Zhang, J. Han, C. Li, J. Wang. Co-saliency detection via looking deep and wide. In *CVPR*, pages 2994-3002, 2015.

[Zhang et al., 2015 b] D. Zhang, J. Han, J. Han, L. Shao. Co-saliency Detection Based on Intra-saliency Prior Transfer and Deep Inter-saliency Mining. *TNNLS*, 2015.

[Zhang et al., 2015 c] D. Zhang, D. Meng, C. Li, et al. A self-paced multiple-instance learning framework for co-saliency detection. In *ICCV*, pages 594-602, 2015.

[Zhang et al., 2015 d] D. Zhang, J. Han, G. Cheng, et al. Weakly Supervised Learning for Target Detection in Remote Sensing Images. *GRSL*, 12(4): 701-705 2015.