

Picking deep filter responses for fine-grained image recognition

Shuang Sha

June 6, 2018

Abstract

Recognizing fine-grained sub-categories such as birds and dogs is extremely challenging due to the highly localized and subtle differences in some specific parts. This paper proposes an automatic fine-grained recognition approach which is free of any object/part annotation at both training and testing stages. This method consists of two steps of deep filter response picking. The first picking step is to find distinctive filters which respond to specific patterns significantly and consistently. The second picking step is to pool deep filter responses via spatially weighted combination of Fisher Vectors.

1. Introduction

As an emerging research topic, fine-grained recognition aims at discriminating usually hundreds of sub-categories belonging to the same basic-level category. It lies between the basic-level category classification (e.g. categorizing bikes, boats, cars, and so on in Pascal VOC [1]) and the identification of individual instances (e.g. face recognition). In fact, fine-grained sub-categories often share the same parts, and are often discriminated by the subtle differences in texture and color properties of these parts. For example, all birds should have wings, legs and breast etc. And when discriminating similar birds, we can count by only the breast color.

This paper proposes an automatic part detection strategy for fine-grained recognition, which is free of any object / part level annotation at both training and testing stages [3]. As shown in Figure 1, which shows some top responding patches of some filters on CUB200-2011. It can be found that some filters work as part detectors and respond to specific parts (e.g., the head of bird). The approach of this paper is to elaborately pick deep filters with significant and consistent responses. And an overview of this paper proposed framework is shown in Figure 2. We can see that the approach consists of two steps.



Figure 1. Figure 1. Illustration of filter selectivity for a typical network VGG-M on CUB-200-2011. researchers generate candidate patches with selective search [2] and compute response of each patch at conv4 layer. They show several top responding patches of some channels and observe that there exist some filters which respond to specific patterns (e.g., the head or leg of bird), while most of them respond chaotically. This paper proposes to pick deep filters with significant and consistent responses, and learn a set of discriminative detectors for recognition.

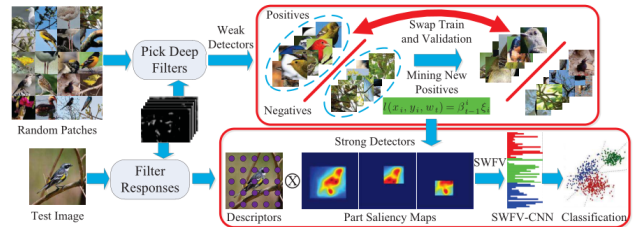


Figure 2. An overview of this paper proposed framework.

References

- [1] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *IJCV*, 88(2):303–338, 2010. 1
- [2] J. R. Uijlings, K. E. Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. *IJCV*, 104(2):154–171, 2013. 1
- [3] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian. Picking deep filter responses for fine-grained image recognition. In *CVPR*, pages 1134–1142, 2016. 1