

# Picking deep filter responses for fine-grained image recognition

Shuang Sha

June 8, 2018

## 1. Related Works

Fine-grained image recognition is a challenging problem and has recently emerged as a hot topic. In this paper, the chief tasks discussed are part localization and feature representation.

### 1.1. Part Localization

As fine-grained datasets are often provided with extra annotations of bounding box and part landmarks [2, 3], most works rely on these annotations more or less. In early work, it is usually assumed that annotations are available at both training and testing time. And in later works, annotations are required only during training, and are not required at testing time. Recently, in some of new works, the conditions are more relaxed, and no annotations are needed at training and testing time. The approach of this paper belongs to the last research, which is free of any object/part level annotation at both training and testing stages. Different from previous works [4], researchers learn a set of discriminative detectors via elaborately selecting positive samples and iteratively updating part models.

### 1.2. Feature Representation

For the description of image, CNN features have achieved breakthrough on a large number of benchmarks [5]. Different from previous works which encode CNN descriptors globally [1], researchers project each response back to the original image and encode each part separately. Most importantly, they propose a picking strategy which conditionally selects responses based on their importance for recognition, and encodes them via spatially weighted combination of Fisher Vectors.

## 2. Learning Part Detectors

In this section, the goal is to learn a collection of discriminative detectors that automatically discover discriminative object/parts. And in order to realize this goal, this work consists of three modules: positive sample initialization, regularized detector training, and detector selection.

### 2.1. Picking Filters: Positive Sample Initialization

Different from previous works, researchers propose a picking strategy which elaborately selects distinctive and consistent patches based on the responses of convolutional neural network (CNN) filter banks [6]. In order to find which filters are distinctive for part discovery, they first generate a large pool of region proposals with selective search [25], and randomly sample a subset of one million patches. As shown in Figure 1, the response distributions are sparse, with most responses focusing on only a few channels (e.g., for CUB-200-2011, over 90% responses focus on the top 5% channels). They refer to these channels as distinctive filters, which respond to specific patterns significantly. In their experiment, they select channels which include the top 90% responses as distinctive.

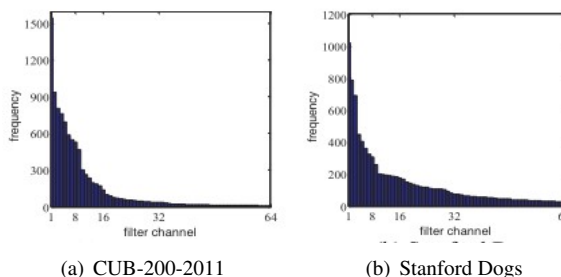


Figure 1. Response distributions of the top scored 10K patches on VGG-M (512 channels). The top scored responses only focus on a few channels. We remove the channels with lower response frequency for better visualization.

## References

- [1] M. Cimpoi, S. Maji, and A. Vedaldi. Deep filter banks for texture recognition and segmentation. In *CVPR*, pages 3828–3836, 2015. 1
- [2] O. M. Parkhi, A. Vedaldi, C. Jawahar, and A. Zisserman. The truth about cats and dogs. In *ICCV*, pages 1427–1434, 2011. 1
- [3] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. Jawahar. Cats and dogs. In *CVPR*, pages 3498–3505, 2012. 1

- [4] M. Simon and E. Rodner. Neural activation constellations: Unsupervised part model discovery with convolutional networks. In *ICCV*, pages 1143–1151, 2015. [1](#)
- [5] B. P. Yao, G. Bradski, and F. F. Li. A codebook-free and annotation-free approach for fine-grained image categorization. In *CVPR*, pages 3466–3473, 2012. [1](#)
- [6] X. P. Zhang, H. K. Xiong, W. G. Zhou, W. Y. Lin, and Q. Tian. Picking deep filter responses for fine-grained image recognition. In *CVPR*, pages 1134–1142, 2016. [1](#)