Supervised Kernel Descriptors for Visual Recognition

Ruyue Han

July 19,2018

Abstract

In visual recognition tasks,designing low level image features representation is fundamental work.The advent of local patch features from pixel attributes such as SIFT and LBP, has dramatic progressed.Recently,kernel descriptors(KDES) [5],a kernel view of these feature ,generalizes the feature design in an unsupervised way and yields impressive results.

In this paper, author give a supervised framework to embed the image level label information into the design of patch level kernel descriptors, which author call supervised kernel descriptors (SKDES).Specifically, researchers use the bag-of-words (BOW) image classification pipeline and a large margin criterion to learn the lowlevel patch representation, which makes the patch features much more compact and achieve better discriminative ability than KDES. With that method, researchers achieve competitive results over several public datasets comparing with state-of-the-art methods.

1. Introduction

For many visual recognition tasks, one critical problem is to discover robust image representations.The feature design is very challenging,because,on the one hand, image features should be invariant to the inner-class variation,on the other hand,image features also need to be discriminative regarding the inter-class differences for separating confusing classes.

To handle these challenges, current state-of-the-art image classification algorithms use the bag-of-words pipeline,which firstly extracts low-level patch based descriptors, then encodes them into a middle level representation through an over-complete dictionary, and finally obtains image features by a spatial pooling strategy [2–4, 8, 9].

Nevertheless, most work keeps the low level descriptors as hand-crafted features, such as HOG [7] or SIFT [8]. Aselaborated by [1, 6], the selection of raw descriptors is also an essential factor for achieving good performance in recognition tasks as the error at the beginning may propagate to latter stages. In author's work, they focus on learning discriminative patch descriptors by exploiting image label information for improving racognition accuracy.

References

[1] Q. Y. J. O. A. H. a. H. A.Tamrakar, S.Ali. Evaluation of low-level features and their combinations for complex event detection in open source videos. *CVPR*, 60(2):91–110, 2004.

[2] Y. a. T. J.Yang, K.Yu. Linear spatial pyramid matching using sparse coding for image classification. *CVPR*, pages 1794–1801, 2009.

[3] L.Bo and C.Sminchisescu. Efficient match kernel between sets of features for visual recognition. *NIPS*, pages 135–143, 2009.

[4] L.Bo and C.Sminchisescu. Geometric p-norm feature pooling for image classification. *CVPR*, pages 2697–2704, 2011.

[5] L.Bo, X.Ren, and D.Fox. Kernel descriptors for visual recognition. *NIPS*, pages 244–252, 2010.

[6] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell*, 27(10):91–110, 2005.

[7] L. a. P. S.Gao, I.W.H.Tsang. Histograms of oriented gradients for human detection. *CVPR*, pages 886–893, 2005.

[8] L. a. P. S.Gao, I.W.H.Tsang. Local features are not lonely laplacian sparse coding for image classification. *CVPR(2)*, pages 2169–2178, 2006.

[9] a. J. S.Lazebnik, C.Schmid. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *CVPR(2)*, pages 2169–2178, 2006.