Comparison of Mid-Level Feature Coding Approaches And Pooling Strategies in Visual Concept Detection (Supplementary Material).

P. Koniusz, F. Yan, K. Mikolajczyk

Centre for Vision, Speech and Signal Processing, University of Surrey, GU2 7XH, Guildford, UK

Abstract

A number of techniques for generating mid-level features, including two variants of Soft Assignment, Locality-constrained Linear Coding, and Sparse Coding, are evaluated in the main document [1]. Pooling methods that aggregate mid-level features into vectors representing images like Average pooling, Max-pooling, and a family of likelihood inspired pooling strategies are scrutinised there. This supplementary material extends our evaluations to the PascalVOC07 dataset given Sparse Coding, as state-of-the-art classification performance it the main document is demonstrated thus far on Caltech101, Flower17, and ImageCLEF11 datasets.

Keywords: Bag-Of-Words, Mid-level features, Soft Assignment, Sparse Coding, Locality-constrained Linear Coding, Max-pooling, Analytical Pooling, Power Normalisation, Comparison

1. Experimental Arrangements

Sparse Coding [2, 3] (SC) is evaluated on the PascalVOC07 [4] dataset. Online Dictionary Learning is used to train dictionaries for this experiment [5]. The spatial relations in images are exploited by either Spatial Coordinate Coding [6, 1] (SCC) or Spatial Pyramid Matching [7] (SPM). Dominant Angle Pyramid Matching [6, 1] (DoPM) that exploits orientations of dominant edges from the local descriptors is also evaluated. SPM is set to 3 levels of coarseness with 1x1, 1x3, 3x1, and 2x2 grids. DoPM is set to 5 levels of coarseness with 1, 3, 6, 9, and 12 grids. Moreover, DoPM employs SCC by default.

The mid-level features are aggregated by Max-pooling [3] (Max), Power Normalisation [8] (Gamma), *theoretical expectation of Max-pooling* [9] (MaxExp), its linearised approximation [1] (AxMin), and the $@n$ scheme [1] combined with MaxExp (MaxExp$@n$). Note that the $@n$ scheme combined with AxMin (AxMin$@n$) is evaluated in the main document [1]. Moreover, linear kernels are used in the following experiments. Multi-label KDA [10] is applied on PascalVOC07, as it was previously found to be a robust performer on this set. Mean Average Precision [10] (MAP) is used to report the classification performance. Table 1 details the experimental parameters.

Email address: p.koniusz@surrey.ac.uk (P. Koniusz)

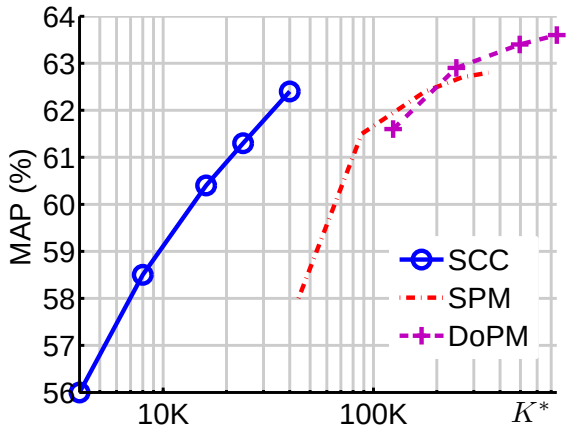

Figure 1: Evaluation of SCC, SPM, and DoPM approaches on the PascalVOC07 set. The overall signature length K^* is indicated. Linear kernels and MaxExp$@n=7$ are used for this experiment.

2. Evaluations on PascalVOC07

Figure 1 compares the classification performance of SCC, SPM, and DoPM approaches on the PascalVOC07 set given various dictionary sizes. Linear kernels and MaxExp$@n$ = 7 are used for this experiment. The dictionary size is varied from 4000 to 40000 atoms for SCC. The signature lengths K^* are the same as the dictionary sizes. The highest result attained by SCC amounts to 62.4% MAP. Moreover, we vary the dictionary size from 4000 to 32000 atoms for SPM. This results in the signature lengths between K^* = 44000 and K^* = 352000. The best

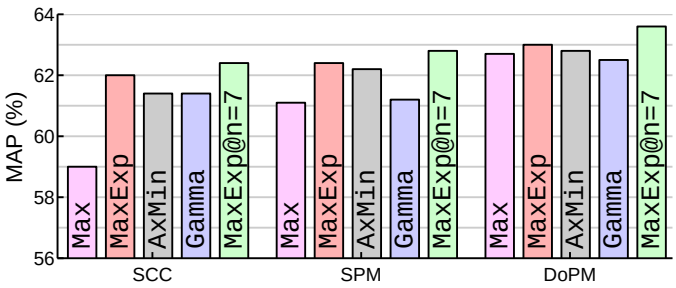

Figure 2: Evaluation of SCC, SPM, and DoPM schemes on the PascalVOC07 set given Max-pooling, MaxExp, AxMin, Gamma, and MaxExp$@n=7$. The dictionary sizes are 40000, 32000, and 24000 atoms for SCC, SPM, and DoPM.

Dataset	Splits no.	Train+Val. samples		Test samples	Total images	Dict. size	Descr. type/ dimensions
PascalVOC07	1x	2501+2510=5011		4952	9963	4K-40K	SIFT/128D
	Descr. interval (px)	Radii (px)	Descr. per img.	Spatial/other schemes		Kernel types	Classifier used
	4,6,8,10,12,14,16	12,16,24,32,40,48,56	19420	SCC/SPM/DOPM		linear	multilabel

Table 1: Summary of the descriptor parameters and various experimental details.

result attained by SPM amounts to 62.8% MAP. Lastly, the dictionary size is varied from 4000 to 24000 atoms for DoPM. The corresponding signature lengths are between $K^* = 124000$ and $K^* = 744000$. This method scores 63.6% MAP.

Figure 2 demonstrates various pooling strategies given dictionary sizes of 40000, 32000, and 24000 atoms for SCC, SPM, and DoPM approaches, respectively. Firstly, we discuss SCC approach. MaxExp@$n=7$ scores 62.4% MAP followed closely by MaxExp that yields 62.0% MAP. AxMin and Gamma attain the same score of 61.4% MAP followed by Max-pooling that yields 59.0% MAP only.

Next, we discuss SPM approach. MaxExp@$n = 7$ scores 62.8% MAP followed closely by MaxExp and AxMin that yield 62.4% and 62.2% MAP. Gamma and Max-pooling attain 61.2% and 61.1% MAP only.

Lastly, we discuss DoPM approach. MaxExp@$n=7$ scores 63.6% MAP followed by MaxExp and AxMin that yield 63.0% and 62.8% MAP. Max-pooling attains 62.7% MAP and outperforms Gamma that yields 62.5% MAP only.

3. Conclusions

SCC approach results in very competitive signature lengths. However, the coding step is computationally prohibitive for large visual dictionaries. It takes 815 and 3.6 seconds to code 1000 descriptors on a single 2.3GHz AMD Opteron core given $K = 40000$ and $K = 4000$ atoms, respectively. This may be partially addressed by Fast Hierarchical Nearest Neighbour Search (FHNNS) proposed in the main document [1]. SPM achieves a marginally better performance with somewhat smaller dictionaries at a price of larger image signatures. DoPM achieves the best performance at a price of sizeable image signatures.

Furthermore, we observe that the @n scheme (combined with MaxExp) attains the highest scores amongst the investigated pooling strategies. MaxExp and its approximation AxMin are also strong performers followed by Gamma and Max-pooling. These results are consistent with the main observations in [1].

References

[1] P. Koniusz, F. Yan, K. Mikolajczyk, Comparison of Mid-Level Feature Coding Approaches And Pooling Strategies in Visual Concept Detection, CVIU (2012).

[2] H. Lee, A. Battle, R. Raina, A. Y. Ng, Efficient Sparse Coding Algorithms, NIPS (2007) 801–808.

[3] J. Yang, K. Yu, Y. Gong, T. S. Huang, Linear Spatial Pyramid Matching using Sparse Coding for Image Classification, CVPR (2009) 1794–1801.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, http://pascallin.ecs.soton.ac.uk/challenges/VOC, 2007.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online Learning for Matrix Factorization and Sparse Coding, JMLR (2010).

[6] P. Koniusz, K. Mikolajczyk, Spatial Coordinate Coding to Reduce Histogram Representations, Dominant Angle and Colour Pyramid Match, ICIP (2011).

[7] S. Lazebnik, C. Schmid, J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2 (2006) 2169–2178.

[8] F. Perronnin, J. Sánchez, T. Mensink, Improving the Fisher Kernel for Large-Scale Image Classification, ECCV (2010) 143–156.

[9] Y. Boureau, J. Ponce, Y. LeCun, A Theoretical Analysis of Feature Pooling in Vision Algorithms, ICML (2010).

[10] M. Tahir, J. Kittler, K. Mikolajczyk, F. Yan, K. van de Sande, T. Gevers, Visual Category Recognition using Spectral Regression and Kernel Discriminant Analysis, ICCV Workshop on Subspace Methods (2009).