Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification

Authors: Jianchao Yang , Kai Yu , Yihong Gong , Thomas Huang
Year: 2009
Journal: IEEE Conference on Computer Vision and Pattern Recognition

Report:

1. **Abstract:**

"Recently SVMs using spatial pyramid matching (SPM) kernel have been highly successful in image classification. Despite its popularity, these nonlinear SVMs have a complexity O(n 2 ~ n 3) in training and O(n) in testing, where n is the training size, implying that it is nontrivial to scale up the algorithms to handle more than thousands of training images. In this paper we develop an extension of the SPM method, by generalizing vector quantization to sparse coding followed by multi-scale spatial max pooling, and propose a linear SPM kernel based on SIFT sparse codes. This new approach remarkably reduces the complexity of SVMs to O(n) in training and a constant in testing. In a number of image categorization experiments, we find that, in terms of classification accuracy, the suggested linear SPM based on sparse coding of SIFT descriptors always significantly outperforms the linear SPM kernel on histograms, and is even better than the nonlinear SPM kernels, leading to state-of-the-art performance on several benchmarks by using a single type of descriptors."

2. **Introduction:**

The objective of the paper is to decrease the complexity of training and testing on non-linear SVMs to O(n) and constant respectively. For that reason, they have calculated the extension of SPM method by Vector Quantization to Sparse coding superseded by multi-scale spatial max-pooling. The approach and the objective is essential due to the day to day increase in the computer vision applications.

3. **Previous Works:**

- Most of the works have been going on for the advancement of Bag of Features, K-means clustering and the Spatial Pyramid Matching(SPM), in which SPM seems to be very effective of all.
- It is also seen that linear SVMs are used in the most of the applications as the data can be trained through that method fast.
- The article differed in implementation of the Sparse Coding on SIFT-features.

4. **Content and Approach:**

- For the different features X of the images various cluster points are generated as a matrix V, called as code-book, in which L_2 norm is found for the both V and X.
- The other matrix for indicating cluster membership is defined with U. This helps in finding the index of only non zero elements of matrix U indicating for which the cluster vector x_m in X belongs to U.
- Sparse Coding also contains training and coding phases as in Vector Quantization.
- The Sparse Coding is selected because of following reasons:
    - SC code will be having less reconstruction error than Vector Quantization.
    - Sparsity allows to acquire the salient features of the images.
    - As the image patches are sparse signals, it is easy to implement Sparse Coding.
- Generally for every image a single histogram is found. But in this case all the local histograms for the various parts of the image are calculated, summed and normalized.
- It is said that the linear Kernel on histograms will result in worse results due to high quantization of errors accumulated due to Vector Quantization.
- For this reason, the approach of using linear SVMs based SC of SIFT is implemented.
- The histograms were taken as some fooling function of the vectors u. They even found that max pooling is better than the alternative pooling approaches.
- These features of pooling of various locations and scales of the image are then summed up for a spatial pyramid representation of the image.
- Then, they have derived the primal formulation which decreased the training cost to O(n).
- The SPM based kernel also achieved excellent accuracy may be because of the following reasons:
    - As Sparse Coding has very less quantization errors than Vector Quantization.
    - Sparse Coding is suitable because of image patches being sparse.
    - The max pooling computing is more robust and salient for the local translations.
- In the implementation, the authors had iterated over the 50,000 SIFT descriptors by making the vectors u & v constant respectively one after the other.
- The experiments are done on the three classes (KSPM, LSPM, ScSPM) and four datasets (Caltech-101 Dataset, Caltech-256 Dataset, 15 Scenes Categorization, TRECVID 2008 Surveillance Video.
- The authors also noted the effect of codebook size, for which the lower codebook size, the discriminant power is lost by histogram feature. If the size is large, there is high risk that the histograms of same class of images will not match.

- The codebooks size is fixed for 512 and 1024 for LSPM and ScSPM respectively to achieve excellent results.

5. Results:

- The computation of the 200,000 examples with 5376 dimensional features is done in 5 minutes.
- Through KSPM results are not reproduced due to the dissimilarity in the contrast patches of the images.
- The Dataset Caltech -256 with ScSPM had outperformed the LSPM and KSPM by 15% and 4% respectively.
- In 15 Scene Categorization dataset, implementation of KSPM has not produced great results, for which the reason may be due to descriptor extraction and the normalization process.

6. Discussion and Thoughts:

- There is no significant improvement through the pooling over multiple patch scales, which is reasoned because of the ability of max-pooling to capture the salient properties of the local regions inappropriate to the scale of local patches.
- The max-pooling has produced the best performances due to its robustness and salient features.
- The linear model of the SIFT patterns with sparse features of the images are linearly separable analogous to the text classification.
- As the future work, they advised that the use of feed-forward networks helps in increasing the sparse of the sparse coding.

7. Conclusion:

- The article has resolved the problem of training complexity from $O(n^3)$ to $O(n)$ with the help of non-linear SPM.
- Moreover, the accuracy of testing has not affected by the approach which is very appropriate.
- Thus it created the accessibility of using training methods using SVM for the more complex and huge data.
- The implementation of the max-pooling and it's robustness is well explained by the series of experiments compared to other pooling techniques, in which max-pooling has shown a great performance comparatively.