Feature Fusion for Efficient Object Classification Using Deep and Shallow Learning

T. Janani and A. Ramanan
Department of Computer Science, Faculty of Science, University of Jaffna
janani22thangavel@gmail.com, a.ramanan@jfn.ac.lk

Introduction

Shallow methods start by extracting a representation of the image using handcrafted local image descriptors (e.g., SIFT) and then aggregates such local descriptors into an overall image descriptor by using a pooling mechanism. The emphasis in *shallow learning* is often on feature engineering and selection while in *deep learning* the emphasis is on defining the most useful computational graph topology and optimizing hyper parameters correctly.

In this work we propose to construct a new feature set by processing CNN [1] activations from convolutional layers fused with the traditional BoF [2] representation for efficient object classification using SVMs [3]. The dimension of convolutional features were reduced using PCA technique and the bag-of-features representation was reduced by tailoring the visual codebook using a statistical codeword selection method, in order to obtain a compact representation of the new feature set which achieves increased classification rate while requiring less storage.

Objectives

To improve the overall classification rate of the BoF approach by fusing the convolutional features that are classified using a standard classifier with a relatively low-dimensional representation of both BoF and CNN features.

Methodology

- Connecting method is used in feature fusion. In the training stage,
- Fine-tuned CNN is used to extract the pooled Conv5 image features from the training set. Then principal components analysis (PCA) is performed to reduce the dimension of the features extracted by the pre-tuned CNN.
- For constructing category-specific visual codebooks, SIFT [4] features were extracted from each of the training set of a category and clustered independently using K-means algorithm. The category-specific codebooks were then concatenated into a global codebook. The codewords of the learnt global codebook then serves to construct a histogram for representing an image.
- In order to build a compact and discriminant codebook from an initially build larger codebook we reformulate it using a within-category confidence.
- In the testing stage, the testing images will also go through feature extraction and dimensionality reduction to get the ultimate feature. Then, these features: CNN and BoF are fused using the connecting method and put into the trained SVM classifier to generate the object class label of the testing images.

Methodology …

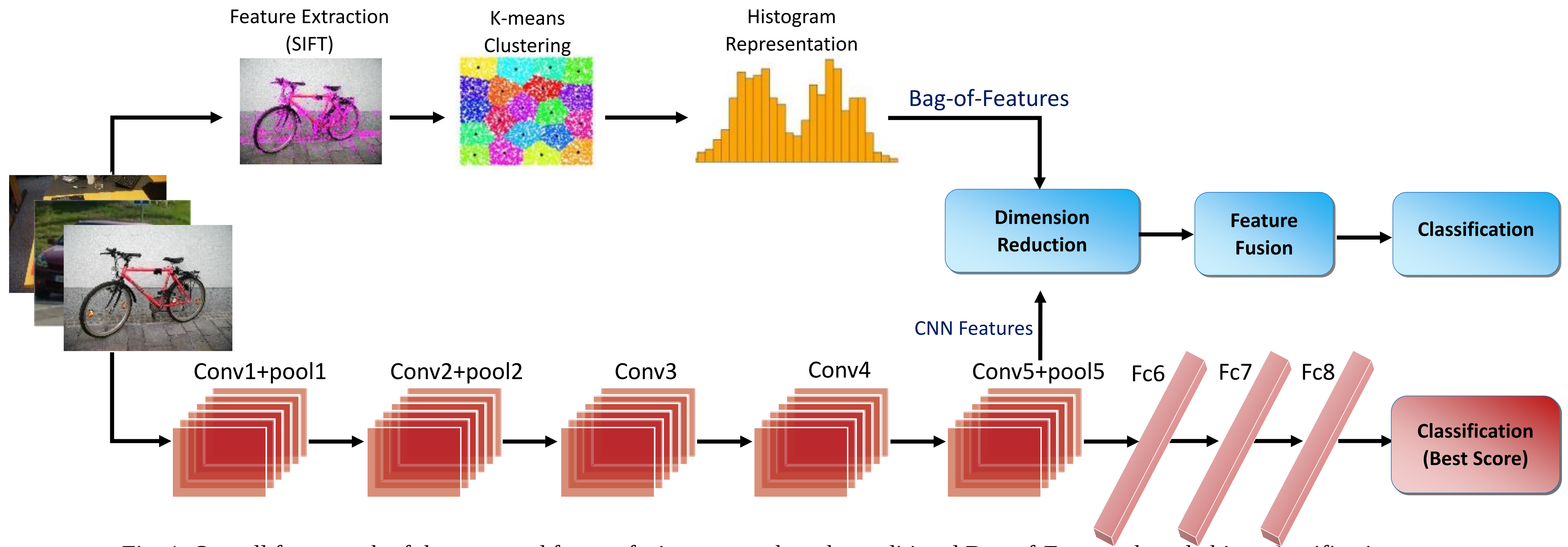
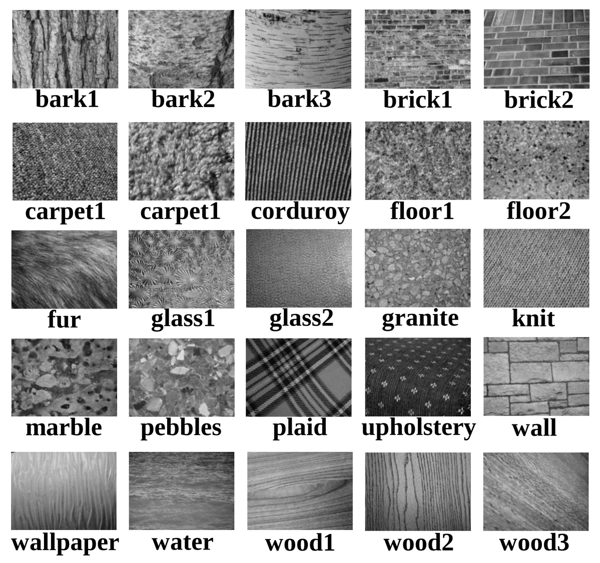


Fig. 1. Overall framework of the proposed feature fusion approach to the traditional Bag-of-Features based object classification

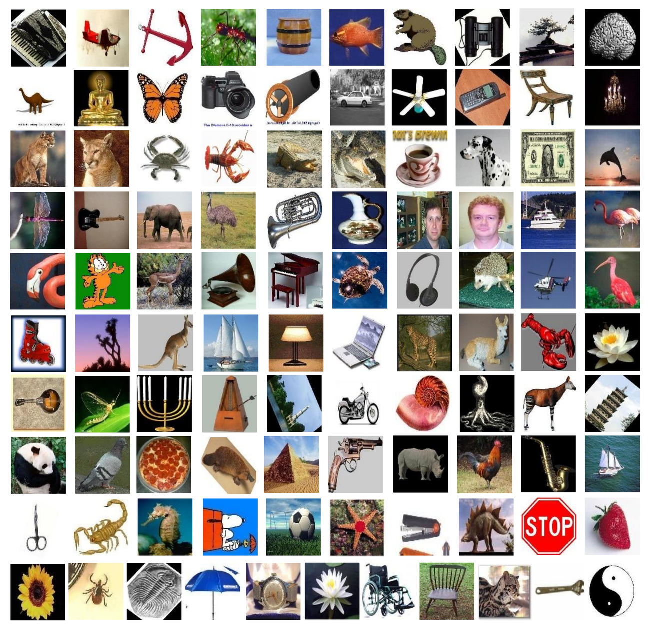
Experimental Setup

The proposed method was tested on three benchmark imagesets: Xerox7, UIUCTex, and Caltech-101.

Xerox7: 7 classes, 1776 images



UIUCTex: 25 classes, 1000 images



Caltech-101: 101 classes, 9146 images

- All image sets: 70-30% training-testing.
- Features: SIFT descriptors
- Codebook Construction: K-means
- Classifiers: Linear OVA-SVMs and CNN
- Fine tune CNN [5] : VGG-M

Testing Results

Table I: Classification rate of the proposed method compared to the standard bag-of-features approach without dimensionality reduction of features

Image Set	BoF	BoF+CNN
Xerox7	78.13	90.28
UIUCTex	98.86	99.17
Caltech-101	76.24	97.65

Table II: Classification rate of the proposed method with dimensionality reduction of features

Image Set	Dimension					Rate
	BoF	C_{within}	CNN	PCA	Total	
Xerox7	700	506	18432	1242	1748	91.67
UIUCTex	1000	852	18432	699	1551	99.17
Caltech-101	2020	1011	18432	6075	7086	97.43

Compact Representation

1. PCA

PCA is used to reduce the dimension of the CNN features. It is a linear dimensionality reduction technique that transforms a number of correlated variables into un-correlated variables called principal components. The goal of principal component analysis is to embed the data into a linear subspace of lower dimensionality describing as much of the variance in the data as possible.

2. Statistical codeword selection

Statistical codeword selection technique is used on BoF to reduce the dimention. Statistical codeword selection technique is based on within-category confidences.It is calculated by analysing a within-category variance of i^{th} codeword. The within-category confidence of the i^{th} codeword is represented as follows:

$$C_{within,\ i} = \frac{1}{\sum_{j=1}^{n} var(hij)}$$

where h_{ij} is the i^{th} codeword value of each image belonging to the j^{th} object in the BoF histogram domain, $var(h_{ij})$ is a variance of the h_{ij} and n is the number of object categories.

Conclusion

- The proposed method has significantly improved the classification accuracy compared to the traditional bag-of-features approach.
- The proposed method can be further improved by considering
  - the fully convolutional layers' CNN features
  - other pre-trained model such as VGG-S, VGG-16, GoogleNet, ect.

References

[1] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the Devil in the Details: Delving Deep into Convolutional Nets", In BMVC, 2014.
[2] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints", In Workshop, ECCV, pp.1–22, 2004.
[3] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", In ECML, pp.137–142, 1998.
[4] D. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints",In IJCV, 60(2), pp.91–110, 2004
[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik., "Rich feature hierarchies for accurate object detection and semantic segmentation", In CVPR,pp. 580–587, 2014.