

Boosted Projection: An Ensemble of Transformation Models

Ricardo Barbosa Kloss, Artur Jordão, and William Robson Schwartz

Smart Surveillance Interest Group, Department of Computer Science
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{rbk, arturjordao, william}@dcc.ufmg.br,

Abstract. Computer vision problems usually suffer from a very high dimensionality, which can make it hard to learn classifiers. A way to overcome this problem is to reduce the dimensionality of the input. This work presents a novel method for tackling this problem, referred to as Boosted Projection. It relies on the use of several projection models based on Principal Component Analysis or Partial Least Squares to build a more compact and richer data representation. We conducted experiments in two important computer vision tasks: pedestrian detection and image classification. Our experimental results demonstrate that the proposed approach outperforms many baselines and provides better results when compared to the original dimensionality reduction techniques of partial least squares.

Keywords: Dimensionality Reduction, Machine Learning, Ensemble, Partial Least Squares, Computer Vision.

1 Introduction

Problems such as pedestrian detection and image classification are active research topics in computer vision due to their applications in surveillance and robotics. However, these problems usually suffer from very high dimensional data. Therefore, the employment of dimensionality reduction techniques is common to achieve reduction in training and testing cost and, most importantly, as a workaround to the curse of dimensionality, as can be seen in the work of Guyon et al. [10].

It has been shown that performing dimensionality reduction on the raw features improves data representation while also reduces computational cost [17],[16],[14],[11]. The idea is to use a projection on the raw data to find a new space that uses a smaller number of features, improving the classification performance. For instance, the Partial Least Squares (PLS) technique [15] was employed to project to a low dimensionality space, achieving better results than the state-of-the-art on the task of hyperspectral face recognition with spatio-spectral information by Uzair et al. [20]. In another case, Akata et al. [1] employed Principal Component Analysis (PCA) to reduce the dimensionality

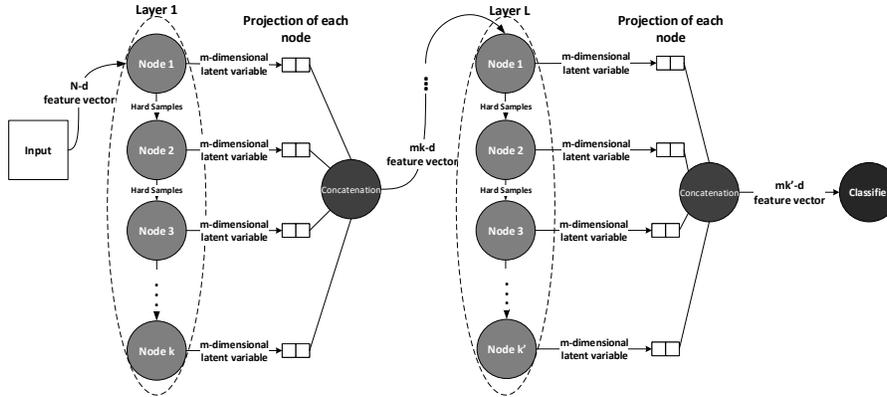


Fig. 1. General structure of the proposed framework. A node learns a projection using the hard samples found by the previous node in the same layer. Then, new layers can be created using the concatenation of the projections of each node in the previous layer as input.

of their SIFT [13] descriptors to circumvent the curse of dimensionality. It is important to note that such approaches could also be used together with deep learning techniques to potentially compact and improve deep features.

Besides the use of projection methods, another way of improving the classification is the employment of ensemble classifiers. These classifiers have been widely employed in several machine learning tasks [2],[19], [6], [11] due to their simplicity and improved results. In general, ensemble classifiers are composed of weak classifiers (e.g., decision stumps) that are able to achieve a powerful classification when combined.

Inspired by the aforementioned studies, this work proposes a novel method that combines dimensionality reduction and ensemble techniques to find richer variables than those estimated by a single projection model. As illustrated in Figure 1, the units of our ensemble are modeled as nodes built iteratively. Each node is composed of a projection model and a weak classifier learned with a distinct subset of samples hard to classify (referred to as *hard samples*¹) by the previous node. The idea behind this procedure is to create projection models yielding rich and discriminative features for different categories of samples (from samples easier to harder to classify). It is also possible to cascade the projection obtained by a set of nodes to a new set of nodes, this is represented by multiple layers, as seen in Figure 1. The goal of using multiple layers is to reduce possible redundancy among nodes.

Our proposed approach is similar to the Adaptive Boosting [9] in the sense that it reweighs samples according to their classification score. However, while the adaptive boosting focuses on grouping classifiers, we use the classifier as a

¹ Hard samples are samples that were misclassified when presented to a classifier.

tool to find hard samples. This is also similar to the hard mining employed by Dalal and Triggs [7], except that we do not search for these samples on a validation set but on the training set itself. Another difference is that we iteratively search for hard samples to construct multiple models, whereas Dalal and Triggs employ this operation to construct a single model. Since we use the boosting-based idea of reweighing hard samples applied to a projection context, we refer to our method as *Boosted Projection*. The proposed method takes advantage of the characteristics of the underlying projection method, for instance reducing redundancy, and it might also present richer features due to the employment of the ensemble scheme to find hard samples.

The main contribution of this paper is the proposal of a novel method for feature projection, which demonstrates the possibility to enhance data representation exploiting the advantages provided by dimensionality reduction techniques and ensemble methods. In addition, the main advantage compared to other dimensionality reduction methods is that the nodes in our method explore different subsets of the data, where the deeper the node, the harder it is to classify its subset of samples, which can better convey discriminative information.

Our experimental results analyze the impact of the proposed Boosted Projection coupled with PLS [22], a method for dimensionality reduction that has been extensively explored in many tasks across the literature. To evaluate the proposed method, we test it in the pedestrian detection, using the INRIA pedestrian detection benchmark [7], and in the image classification task, using the CIFAR-10 [12] benchmark. On the former task, our method achieves a log-average miss-rate of 33.92, an improvement of 3.34 percentage points (p.p.) over the work of Jordão and Schwartz [11], a PLS based random forest, which to the best of our knowledge is the most recent PLS-based approach for pedestrian detection. On the latter task, we achieve an average accuracy of 76.25%, which is 1.01 p.p. better than the single PLS approach with the same number of dimensions.

2 Proposed Approach

In this section, we explain the proposed boosted projection method. We aim at creating an ensemble of projection methods that can yield richer and more discriminative features than a single projection method. To achieve such goal, we propose a framework for dimensionality reduction that explores discriminative information by employing a classifier to find samples that are hard to classify (*hard samples*) and using this information to learn new projections that prioritize the fitting of such samples.

The first unit of our framework is the *node*. It receives features of samples as input and learns a projection model that generates latent variables, used as new features. This projection is performed using a traditional projection method such as PCA [23] or PLS [22] (while the former estimates projections that maximize the variance of the data, the latter finds projection that maximizes the covariance between the data and class labels). The latent variables are used by a classifier to learn a discriminative model used to find which samples have been misclassified

according to a threshold applied to the classifier responses. The threshold is initialized with an arbitrary value and is decreased for each new node to make samples easier to classify and guarantee that the nodes have more diversity. This is important since in ensemble methods, a high diversity is desirable [24][4].

After the hard samples have been identified for a given node, a subsequent node is created by learning a projection and a new classifier considering such samples, which will lead to a new subset of hard samples. This procedure continues until convergence, which can be met either by reaching a maximum number of nodes or by having no more misclassified samples. Thus, each node learns a model based on different subsets of samples, where the deeper the node, the harder it is to classify its subset of samples.

We refer to the set of learned nodes built until the convergence as a *layer*, as illustrated in Figure 1. The output of a layer is the concatenation of the latent variables of each node of that layer. After all the nodes of a layer have been built, their output can be used as the input to create a new layer. The benefit of using additional layers is that if there is redundancy in some nodes, probably due to a lack of diversity, the nodes of the next layer would be able to filter the redundancy out.

The output of the last layer is used to learn a final classifier, which might be different than the one used to find hard samples. It is important to emphasize that the output of the last layer is composed of projections learned for different subsets of samples and based on different features, causing each node’s output to be an expert on that subset. This is an advantage of the proposed approach over a single projection method, such as PLS. On the other hand, a single projection-based approach would find the projection that is best for the average of the whole data, which would fail to model multi-modal distributions, for instance.

3 Experimental Results

In this section, we present our experiments and results achieved. First, we evaluate our method on the INRIA pedestrian detection benchmark, using log-average miss rate and compare it with other baseline approaches. Then, we evaluate our method on the task of image classification with the CIFAR-10 benchmark, using the average accuracy to measure the performance of the proposed approach against a single projection method.

3.1 Pedestrian Detection

The studies of Schwartz et al. [17] and Jordão et al. [11] are the main works concerning PLS based dimensionality reduction applied to pedestrian detection. Therefore, to show that our method achieves better results than traditional projection-based approaches, we employed it to the pedestrian detection problem and compare with other methods that use comparable features (HOG and LBP).

To validate our results, we adopt the evaluation protocol used by state-of-the-art works, which is called *reasonable set* [8] (a detailed discussion regarding

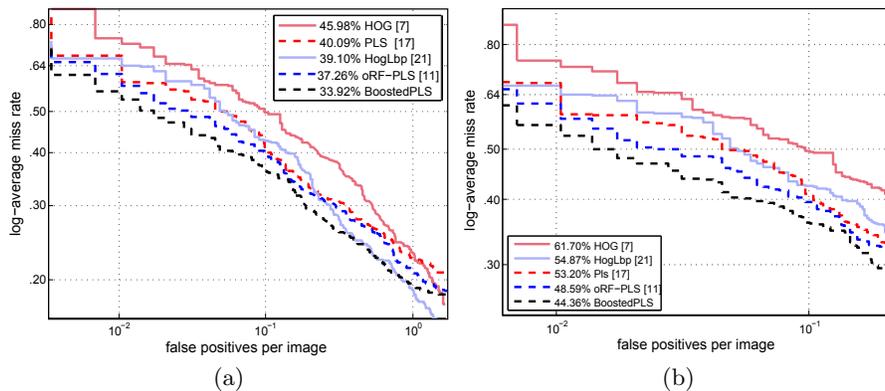


Fig. 2. Comparison of our boosted PLS projection with other approaches found in the literature. (a) results using the log-average miss-rate (lower values are better) of 10^{-2} to 10^0 (standard protocol); (b) results using the area from 10^{-2} to 10^{-1} .

this protocol can be found in [8,3]), where the results are reported on the log-average miss rate. In addition, similarly to Jordão et al. [11], we used the TUD pedestrian dataset as validation set to calibrate some parameters of our method.

Our first experiment evaluates the impact of the number of nodes and layers to compose the boosted projection. On the validation set, the best values for these parameters were 20 and 3, respectively. The remaining parameter to be defined in our method is the number of components (latent variables), for each PLS that will compose a node of a layer (see Section 2). Similarly to Jordao and Schwartz [11] and according to the results achieved on the validation set, we note that this parameter is the most important parameter for PLS-based methods since varying this parameter from 3 to 6, the log-average miss rate decreases from 53.20 to 50.07, respectively. Based on the results obtained in the validation, our final PLS based boosted projection was set to 3 layers, where each layer is composed of 20 nodes that project the data to a 6-dimensional feature space using the PLS.

In our next experiment, we compare the results of the proposed Boosted Projection with other baseline methods. To provide a fair comparison, we considered the results reported by the authors in their works. Figure 2(a) shows the log-average miss rate (on the standard protocol) achieved by the methods on the INRIA person dataset. Our method achieved a log-average miss rate of 33.92, outperforming the works of Schwartz et al. [17] and Jordão and Schwartz [11] in 6.17 and 3.34 percentage points, respectively. On the area on the curve from 10^{-2} to 10^{-1} (Figure 2(b)), which represents a very low false positive rate, our Boosted PLS still outperforms the other methods, a evidence of being robust to hard detections. Moreover, our proposed method uses around 30 times fewer features, compared to the work of Schwartz et al. [17], and achieves more accurate detections. It is also important to reinforce that our method is able to outper-

Table 1. Results on the CIFAR-10 test dataset.

Method	Average Accuracy	Dimensionality
SVM	71.62%	4096
PLS + SVM	69.10%	10
PLS + SVM	74.52%	30
PLS + SVM	75.24%	200
PLS + SVM	75.39%	500
Boosted PLS (1 layer)	75.94%	$20 \times 10 = 200$
Boosted PLS (2 layers)	76.25%	$20 \times 10 = 200$

form detectors that employ more complex features, for instance, HOG-LBP [21], while using simple HOG-features (the same setup as proposed by Dalal and Triggs [7]).

Our last experiment regarding pedestrian detection evaluates the computational cost compared to the work of Jordão and Schwartz [11], which, to the best of our knowledge, is the most recent PLS-based method. We execute, on the same hardware configuration of [11], the detection on a 640×480 image for 10 times, as is done by Jordão et al. [11], and we compute its confidence interval using 95% of confidence. Our method obtained a confidence interval of $[67.02s, 67.77s]$ against $[270.20s, 272.44s]$ (results reported by the authors) of [11], showing that the methods present statistical differences at the execution time. Therefore, according to the results, our method is four times faster than [11]. This reduction in computational cost is possible due to the use of fewer PLS projections than [11] (even if we consider the optimal case, where each oblique decision tree consists of only one PLS projection, see [11] for details).

3.2 Image Classification

To evaluate our method on the image classification task, we used the CIFAR-10 [12], a dataset with 32×32 pixels RGB colored images of ten classes. We extract deep features without data augmentation using the Keras Framework [5] with a VGG16 [18] based network. We tune the C parameter of the Support Vector Machine (SVM), with linear kernel, by employing stratified cross validation on the training set. The parameter was found to be optimal as $C = 0.1$, and we used this parameter for all models. Finally, we use the average of the accuracy of each class as the evaluation metric, which is the number of true positives over the sum of the number of true positives and false negatives (higher values are better).

The first column of Table 1 specifies the method used, which is either a SVM, a PLS transformation followed by a SVM, or the boosted version of PLS. The boosted version also uses a SVM to mine for hard samples and as the final classifier. The second and third columns show the results and the number of dimension for each method, respectively. Regarding the dimensionality (third column of Table 1), in the case of the SVM, the dimensionality is the original dimensionality of the features, 4096, however, with the dimensionality reduction

method, PLS, the reported dimensionality is the number of components used. Furthermore, with the Boosted methods, we report the dimensionality as the product of the number of nodes and the number of components of the underlying projection method.

According to Table 1, we find the impact of the number of components in Table 1, the more components the models have, the better its representation. We can also see a correlation between the accuracy and the number of components. Employing the Boosted PLS transformation, with 10 components, and classifying with SVM (Boosted PLS) also achieved better results than with just the PLS transformation of the same model (PLS + SVM). In this case, the improvement is likely due to the fact that while PLS transformations minimize an average error, but thanks to the specialization of each node, our framework can preserve information of small distributions in the data that falls far from the average.

Regarding the PLS transformation (rows 2 to 5 in Table 1), we can see that from 10 components to 30 components there is an improvement of 5.42 percentage points, between 200 components and 500 there is only an improvement of a 0.15, illustrating the saturation of the components learned. Our method, however, with 200 dimensions, was able to improve over the PLS model with 500 components by 0.86 percentage points.

4 Conclusions

We presented a novel method for dimensionality reduction, the *Boosted Projection*. The method focuses on the idea of using several projection models (e.g., Partial Least Squares), to build an ensemble with more compact and richer representation of the raw features. We conducted experiments on two important computer vision tasks: pedestrian detection and image classification. In the first, we demonstrate that the proposed method outperforms many pedestrian detectors, using simple features. In addition, our method is more accurate and faster than [11], one of the most recent detector based on PLS. In the second task, we demonstrate that the proposed method is able to compute features richer than a single dimensionality reduction method.

Acknowledgments

The authors would like to thank the Brazilian National Research Council – CNPq, the Minas Gerais Research Foundation – FAPEMIG (Grants APQ-00567-14 and PPM-00540-17) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project). The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the GeForce Titan X GPU used for this research.

References

1. Akata, Z., Perronnin, F., Harchaoui, Z., Schmid, C.: Good practice in large-scale learning for image classification. PAMI 36(3), 507–520 (2014)

2. Avidan, S.: Ensemble tracking. *PAMI* 29(2), 261–271 (2007)
3. Benenson, R., Omran, M., Hosang, J., Schiele, B.: Ten years of pedestrian detection, what have we learned? In: *ECCV*. pp. 613–627. Springer International Publishing, Cham (2015)
4. Brown, G., Kuncheva, L.I.: good and bad diversity in majority vote ensembles. In: *International Workshop on Multiple Classifier Systems*. pp. 124–133. Springer (2010)
5. Chollet, F.: Keras. <https://github.com/fchollet/keras> (2015)
6. Coganne, R., Fridrich, J.: Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory. *IEEE Transactions on Information Forensics and Security* 10(12), 2627–2642 (2015)
7. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *CVPR* (2005)
8. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. *PAMI* 34 (2012)
9. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: *European conference on computational learning theory*. pp. 23–37. Springer (1995)
10. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *Journal of machine learning research* 3(Mar), 1157–1182 (2003)
11. Jordao, A., Schwartz, W.R.: Oblique random forest based on partial least squares applied to pedestrian detection. In: *ICIP* (2016)
12. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images (2009)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), 91–110 (2004)
14. Martis, R.J., Acharya, U.R., Min, L.C.: Ecg beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control* 8(5), 437–448 (2013)
15. Rosipal, R., Krämer, N.: Overview and recent advances in partial least squares. In: *in Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science*. pp. 34–51. Springer (2006)
16. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision* 105(3), 222–245 (2013)
17. Schwartz, W., Kembhavi, A., Harwood, D., Davis, L.: Human Detection Using Partial Least Squares Analysis. In: *ICCV* (2009)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
19. Takemura, A., Shimizu, A., Hamamoto, K.: Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the adaboost algorithm with feature selection. *IEEE transactions on medical imaging* 29(3), 598–609 (2010)
20. Uzair, M., Mahmood, A., Mian, A.: Hyperspectral face recognition with spatio-spectral information fusion and pls regression. *TIP* 24(3), 1127–1137 (2015)
21. Wang, X., Han, T.X., Yan, S.: An HOG-LBP human detector with partial occlusion handling. In: *ICCV* (2009)
22. Wold, H.: Partial least squares. *Encyclopedia of statistical sciences* (1985)
23. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3), 37–52 (1987)
24. Woźniak, M., Graña, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. *Information Fusion* 16, 3–17 (2014)