Res. Lett. Inf. Math. Sci., 2010, Vol. **14**, pp. 17–24
Available online at http://iims.massey.ac.nz/research/letters/

A Novel Bootstrapping Method for Positive Datasets in Cascades of Boosted Ensembles

T. Susnjak, A. L. C. Barczak, K. A. Hawick
Computer Science
Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand
Email: {T.Susnjak | A.L.Barczak | k.a.hawick}@massey.ac.nz

We present a novel method for efficiently training a face detector using large positive datasets in a cascade of boosted ensembles. We extend the successful Viola-Jones [1] framework which achieved low false acceptance rates through bootstrapping negative samples with the capability to also bootstrap large positive datasets thereby capturing more in-class variation of the target object. We achieve this form of bootstrapping by way of an additional embedded cascade within each layer and term the new structure as the Bootstrapped Dual-Cascaded (BDC) framework. We demonstrate its ability to easily and efficiently train a classifier on large and complex face datasets which exhibit acute in-class variation.

Keywords: face detection, boosting, AdaBoost, CoBE, classifier, training.

1 Introduction

Face detection plays a vital role in an increasing number of applications. Considerable attention has been dedicated to enhancing face detection systems in recent years but a robust, general purpose detector is still an open problem.

Robust face detectors remain elusive due to significant challenges posed by extrinsic factors such as illumination changes in their operating environments and intrinsic factors defined by broad in-class variations of a human face which possesses a range of facial expressions.

Producing an effective, general purpose face detector therefore requires not only extensive negative training samples to ensure low false acceptance rates but also comprehensive positive datasets that encapsulate maximal environmental variation and complexity found in human faces.

Viola-Jones [1] developed the first real-time, accurate face detection system that has been the backbone of much subsequent research. Their groundbreaking work introduced the concept of cascades of boosted ensembles (CoBE), trained using AdaBoost [2] in conjunction with Haar-like features that could be rapidly calculated using integral images. Their framework facilitated the training of classifiers using negative bootstrapping [3] which enables learning with massive negative datasets. However, their framework provided no capability to bootstrap positive samples. Consequently, all positive sample training on this framework has to be conducted simultaneously. As a result of intense computational and memory demands of CoBE frameworks, the size of the positive datasets is limited and comprises a fraction of the size of the negative sets [4].

While the predominant body of research in the last decade has been devoted to improving various components of the CoBE architecture, only recently has research begun to surface with methods enabling face detector training on large positive datasets [4]. Xiao et al. [5] were first to propose a bootstrapping approach which worked by training directly only on a small subset of a massive positive dataset whose distribution was iteratively compared to that of the entire positive dataset and updated correspondingly. Their face detector displayed an impressive capability for handling large positive datasets and their framework was implemented for distributed learning on a number of machines. [4] proposed a simpler approach on a single machine in which training on a subset of an entire positive dataset was also utilized. Their methodology employed iterative training of each cascade layer using only a positive subset. During the layer construction of their approach, the intermediate sub-classifier undergoes periodic validation against the entire positive dataset. When the detection rate is below a designated target, a portion of misclassified positive samples from the massive positive set are added to the subset and the learning resumes until the targets are met.

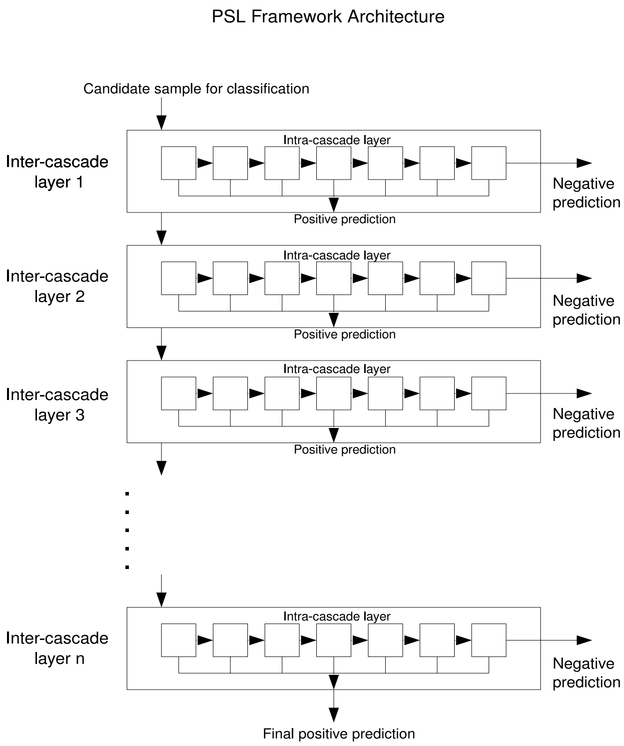


Figure 1: PSL architecture

In this paper, we present an alternative positive sample bootstrapping method for CoBEs that is robust for training on complex datasets which exhibit considerable in-class variations. It is capable of achieving 100% layer hit rates together with high negative rejection rates, without artificial layer thresholds and is versatile in that it can be calibrated to operate to required execution runtime requirements.

2 Bootstrapped Dual-Cascaded Framework

The BDC is an intuitive extension of the original non-bootstrapping PSL (*Parallel Strong* classifier within the same *Layer*) framework proposed by Barczak [6] seen in Figure 1. Unlike the positive sample bootstrapping approaches of [4, 5], the BDC training framework portrayed in Figure 2 implements positive dataset bootstrapping by constructing an additional cascade inside each layer of a strong classifier. Using this nested cascade, the BDC achieves further modularization and tractability of the training procedure through a divide and conquer strategy. While the standard *inter-layer* cascades of the Viola-Jones framework focus on rejecting negative samples that are subsequently replaced by new bootstrapped samples, our *intra-layer* cascades invert this goal and emphasize instead the correct prediction of positive samples. Following the construction of each BDC *intra-layer* cascade stage, the correctly predicted positive samples are removed and replaced by *unseen* bootstrapped positive samples. The process of positive bootstrapping continues until a sufficient number of BDC *intra-layer* stages have been constructed to correctly predict the entire positive dataset.

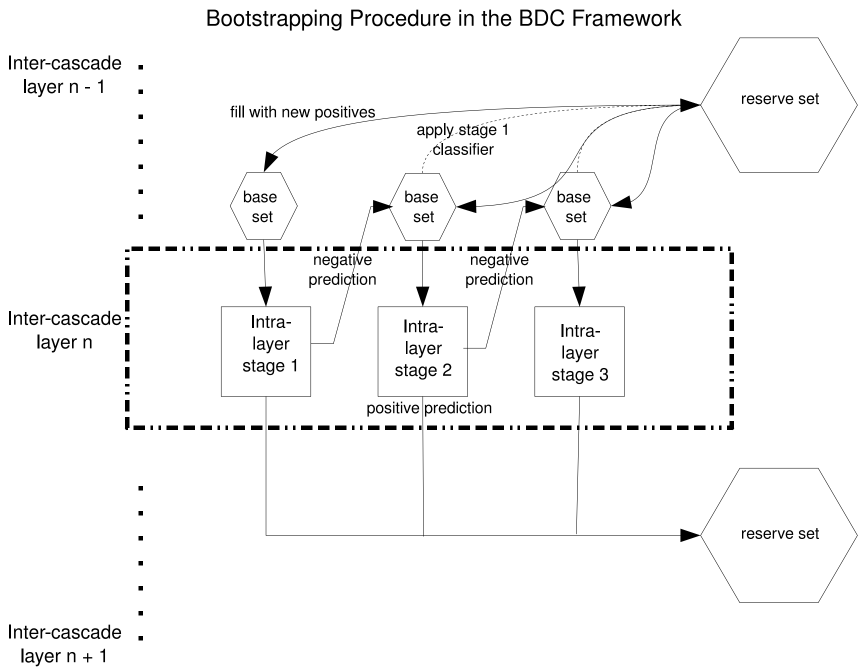


Figure 2: BDC bootstrapping method

The whole negative dataset and a subset of the entire positive dataset constitute the training sets used for each stage of a nested cascade. The positive sample subset which the learning algorithm *sees* and trains on explicitly we call the *base set.* The entire positive dataset from which new positive samples are bootstrapped is referred to as the *reserve set.*

The procedure for intra-layer cascade training can be seen in Algorithm 1. The training of an intra-layer cascade initiates with randomly selecting a comparably small subset of positive samples from the reserve set in order to construct the base set. The base set is then trained against the negative dataset to produce individual stages. Each stage of this nested cascade is trained with a target hit rate of 100% and a high rejection rate. However, the size of each nested stage is restricted by the maximum number of weak classifiers that can comprise it. Once this maximum number has been reached, the training for that stage ceases and a new intra-layer stage begins. The positive bootstrapping procedure is then initiated. The positive samples in the reserve set are validated against the resulting stage classifier and all correctly predicted samples are removed from training subsequent nested stages. The remaining positive samples are randomly selected to comprise the new *base set* for the next intra-layer stage together with all the incorrectly predicted positive samples from the previous stage's base set.

At detection time, the classification process also becomes modularized and more efficient. A candidate sample is predicted as a negative by a layer only if all nested stages within it classify it

Given:
$C_n = n_{th}$ inter-layer layer sub-classifier
$S_i = i_{th}$ intra-layer stage sub-classifier
PB_i = positive base set used on S_i
PR = positive reserve set
f_{min} = minimum false acceptance rate
d_{min} = minimum required hit rate set at 100%
WK_{max} = max number of weak classifiers

1 randomly select positive samples from PR to create PB_i
2 train C_nS_i against PB_i until f_{min} and d_{min} or WK_{max}
3 validate PR using C_nS_i and remove from it correctly classified samples
4 if all samples in PR have been correctly predicted then start a new layer C_{n+1} otherwise start new stage S_{i+1} repeat step 1

Algorithm 1: BDC bootstrapping method for each cascade layer

as a negative. A sample is accepted as a positive once any nested stage predicts it as a positive thereby not requiring the computation of the remaining internal stages.

3 Experiments

In order to demonstrate the ability of our bootstrapping method to effectively handle large and difficult datasets (example of which is shown in Figure 3), we compare it to the standard cascaded and the PSL training structures without a positive sample bootstrapping capability. We collected 15000 facial images from various publicly available datasets; FERET, Yale *Face Database B* [7] and the face database from the Vision Group of Essex University. We used these images for training without any modification to them or additional synthetic data for fine tuning classifiers for accuracy on our test dataset.

Figure 3: An example of faces used for the positive dataset.

3.1 Experimental Setup

Three positive datasets of different sizes were used to train classifiers. They consisted of 5000, 10000 and 15000 samples and were trained against a negative dataset of 2000 samples, bootstrapped from a pool of 2500 negative images. The size of the base sets was 500. The PSL and BDC structures produced two classifiers for each dataset with different parameters for stage sizes. One being a maximum of five weak classifiers per stage and the other ten. All classifiers were trained with 100%

Table 1: Training settings and dataset details.

Property	Attribute
Positive datasets	5000,10000,15000
Negative datasets	2000
PSL & BDC stage sizes	5,10
Haar-like feature types	8
Max Haar-like features	200000
Minimum pixel area size per Haar-like feature	16
Maximum available negative images	2500
Subwindow scale increase factor	1.2
Subwindow raster scan pixel increment	2
Positive sample kernel size	24x24
Initial subwindow size	24x24
Target training error	0%
Target layer hit rate	100%
Target layer false alarm rate	50%
Maximum stages per layer (PSL & BDC only)	10
Maximum layers	100

hit rate and 50% false alarm rate layer targets. A summary of the training parameters as well as dataset details can be seen in Table 1.

The classifiers were tested using the CMU MIT image dataset comprising of 130 images which contain 506 positive face images. The number of negative samples generated from the 130 images were 72,654,174. Each image was scanned by raster beginning with a 24x24 pixel kernel. After each calculation of a sub-window, the kernel shifted by two pixels until the entire image was scanned and thereafter increased by scale of 1.2. An error margin was given for each positive sample and calculated to be one third of image's length and hight in each direction on its axis.

3.2 Results

Figure 4 shows classifier training runtimes in which the standard cascaded classifiers of Viola-Jones and PSL classifier runtimes act as a baseline for comparisons against the proposed bootstrapping approach. We observe that the standard cascaded classifiers require significantly longer runtimes than those of the PSL and BDC classifiers. Furthermore, the figure demonstrates that the bootstrapping component of BDC effectively halves the runtimes of the PSL structure.

Importantly, the figure also points to the resistance of our bootstrapping method to proportional increases in training runtimes in respect to size increases of the positive training set. The figure shows the ability of small base sets in our bootstrapping method to accurately encapsulate enough information about the entire positive reserve dataset without being exposed to all the training samples. In turn, it testifies to the capability of the method to potentially handle even more massive positive datasets with only modest elevations in runtimes.

The ROC curves in Figure 5 show the classifier accuracy on the CMU MIT test dataset. We see that standard cascaded classifiers have clearly achieved stronger hit rates than the PSL and BDC classifiers on segments of the ROC curve that highlight low false acceptance rates. However, the comparable generalization patterns of PSL and BDC classifiers demonstrate that the additional bootstrapping component is indeed robust and effective.

Although the PSL classifiers have generalized better than BDC classifiers when using stage sizes comprising of a maximum of five weak classifiers, we see on the other hand that the BDC classifiers have performed stronger than PSL when stage sizes were raised to ten weak classifiers. A larger stage size is arguably a more important comparison since earlier experiments in [8] have shown

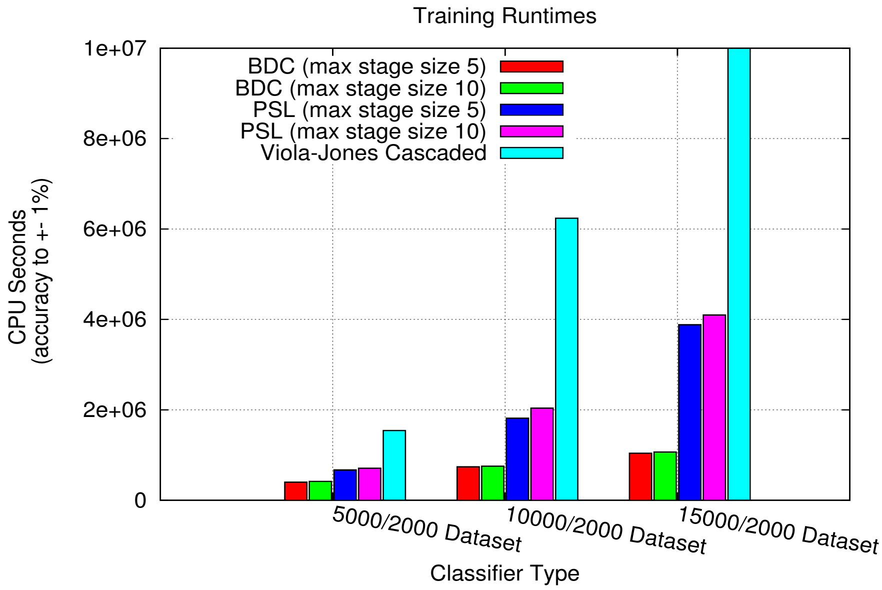


Figure 4: Training runtimes comparing Viola-Jones, PSL and BDC classifiers.

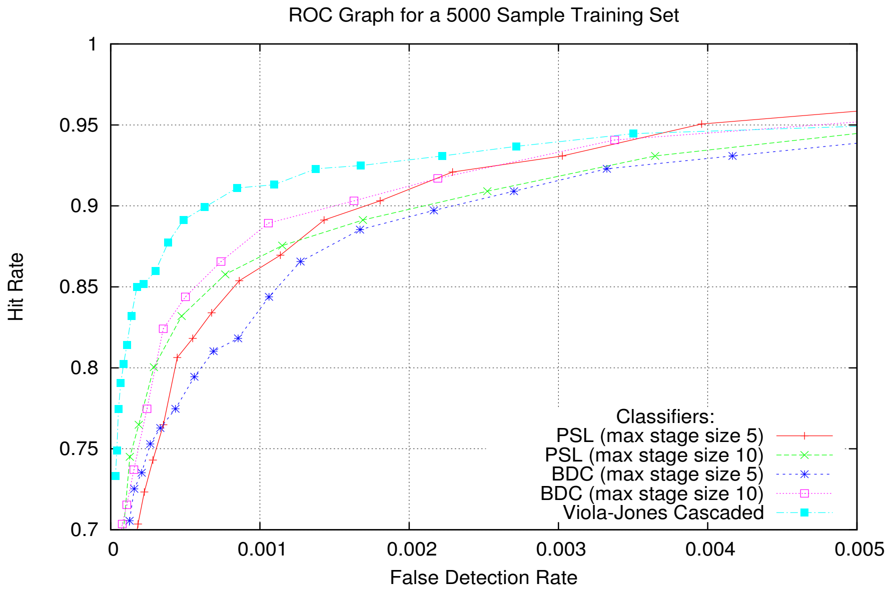
that face detectors trained using up to ten weak classifiers achieve more optimal generalizations while anything below this size becomes erratic.

3.3 Discussion

Since extremely low false positive rates are vitally important in rare-event operating domains such as face detection, the primary drawback of the PSL framework which serves as a foundation for the BDC approach, are its elevated false acceptance rates. This research has successfully shown that the proposed bootstrapping component itself is robust in terms of its generalization ability and that the problem must therefore lie with the underlying approach to training inherited from the PSL framework.

We believe that the elevated false detections are induced by the overfitting that stems from the manner in which some intra-layer stages are trained. We have observed that final stages inside each layer tend to be comprised of very few difficult positive samples which prior stages have misclassified. It is likely that the training taking place on these isolated samples results in overfitting. The consequence is that a cascade-of-stages becomes only as robust as the weakest stages. When the accuracy of an entire cascaded classifier is considered in the PSL and BDC architectures, the overall false detection rate becomes the combined accuracy of all the weakest stages from each layer.

Our solution is to augment trailing intra-layer stages with positive samples which have been correctly learnt in preceding stages. This would steer the learning process away from overfitting to small and unrepresentative patterns of a complete positive dataset. In order to ensure that the misclassified positive samples are still learnt and 100% layer hit rates are satisfied, the augmenting positive samples would be removed from stage training on the fly in a stepwise manner as the number of weak classifiers converge to the maximum allowable number.



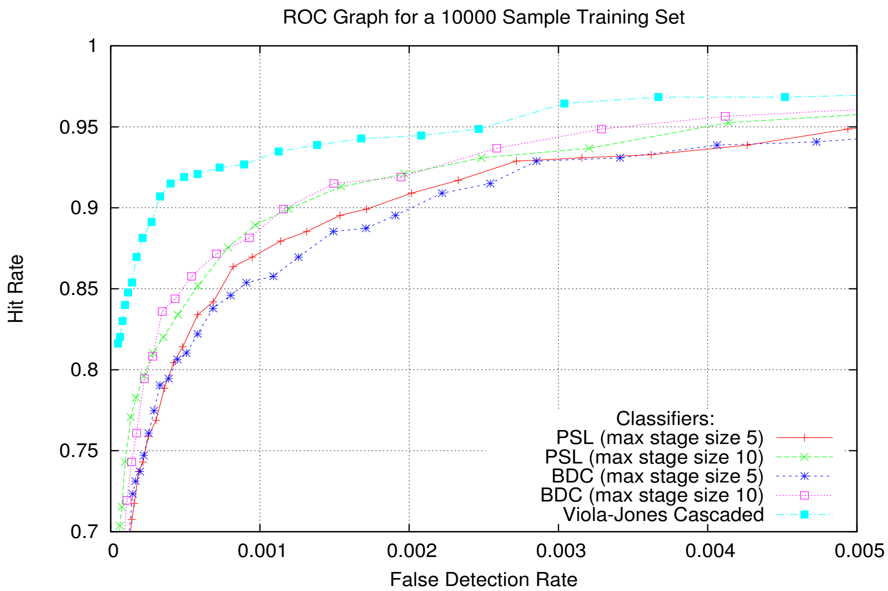


Figure 5: Viola-Jones, PSL and BDC classifier ROC graph curves on the CMU MIT dataset.

4 Conclusion

We have proposed a new framework for training classifiers within CoBE that has the capability to bootstrap positive samples. Our method potentially allows for usage of massive positive datasets with relatively small increases in training runtimes whilst maintaining 100% layer hit rates. We have trained face detectors on very difficult datasets and shown that the training runtimes using our bootstrapping method are a fraction of duration of standard training structures without this capability.

Our future work involves addressing the vulnerability of our framework to elevate false acceptance rates in order to make it more suitable for operating domains which require rare-event detection.

References

[1] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR01, Kauai, HI, IEEE (December 2001) I:511–518

[2] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences **55**(1) (1997) 119–139

[3] Sung, K., Poggio, T.: Example-based learning for view-based face detection. IEEE Patt. Anal. Mach. Intell. **20**(39-51) (1998)

[4] Yan, S., Shan, S., Chen, X., Gao, W., Chen, J.: Matrix-Structural Learning (MSL) of cascaded classifier from enormous training set. (2007)

[5] Xiao, R., Zhu, H., Sun, H., Tang, X.: Dynamic cascades for face detection. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (Oct. 2007) 1–8

[6] Barczak, A.L.C., Johnson, M.J., Messom, C.H.: Empirical evaluation of a new structure for adaboost. In: SAC '08: Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil, ACM (2008) 1764 – 1765

[7] Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence **23**(6) (2001) 643–660

[8] Susnjak, T., Barczak, A.L.C., Hawick, K.A.: Accelerated face detector training using the psl framework. Research Letters in the Information and Mathematical Sciences Volume 13, pp.68 - 80, ISSN 1175-2777, Institute of Information and Mathematical Sciences, Massey University Albany (2009)