

StyP-Boost: A Bilinear Boosting Algorithm for Learning Style-Parameterized Classifiers

Jonathan Warrell¹
 jwarrell@brookes.ac.uk
 Simon Prince²
 s.prince@ucl.ac.uk
 Philip Torr¹
 philiporr@brookes.ac.uk

¹ Oxford Brookes University
 Oxford, UK
<http://cms.brookes.ac.uk/research/visiongroup>
² University College London
 London, UK

We introduce a novel bilinear boosting algorithm, which extends the multi-class boosting framework of JointBoost [7] to optimize a bilinear objective function. This allows style parameters to be introduced to aid classification, where *style* is any factor which the classes vary with systematically, modeled by a vector quantity. The algorithm allows learning to take place across different styles. We apply this *Style Parameterized Boosting* framework (StyP-Boost) to two object class segmentation tasks: road surface segmentation and general scene parsing. In the former the style parameters represent global surface appearance, and in the latter the probability of belonging to a scene-class. We show how our framework improves on 1) learning without style, and 2) learning independent classifiers within each style. Further, we achieve state-of-the-art results on the Corel database for scene parsing.

Traditionally, two-class AdaBoost has been derived from the exponential loss function, $J = \sum_i \exp(-z_i H(i))$, where i runs across the training instances, $z_i \in \{1, -1\}$ indicates if the example is in the positive or negative class, and $H(i)$ is the *strong learner* to be learnt. $H(i)$ is composed of *weak learners* $h(i)$, $H(i) = \sum_m h_m(i)$, which are selected successively over m rounds of boosting based on how well they minimize J . A number of formulations for multi-class boosting have been proposed. In AdaBoost.MH [2], the exponential loss above is expanded in the following way to $k = 1 \dots K$ classes:

$$J = \sum_{k,i} \exp(-z_{k,i} H(k,i)) \quad (1)$$

$$H(k,i) = \sum_m h_m(k,i) \quad (2)$$

where, $z_{k,i}$ is 1 if the instance is in class k and -1 otherwise. JointBoost [7] is a particular algorithm for minimizing (1) where weak-learners are *shared* between output classes.

In style-parameterized boosting (StyP-Boost), we replace the linear combination of weak-learners in (2) with a bilinear combination involving the style-parameters. This can be expressed as:

$$H(k,i) = \sum_m [h_m(k,i) \cdot \sum_{c \in T_m} p_{c,i}] \quad (3)$$

Here, style-parameters $p_{c,i}$ represent the probability instance i belongs to style category c (or its affinity with that style), and $T_m \subset \{1 \dots C\}$ is the set of styles shared by weak learner m . Assuming the p_i 's to be normalized probabilities, we are thus weighting the weak learner by the combined probability the instance is in any of the shared styles. T_m is thus an additional parameter which needs to be set for each new weak-learner. We directly adapt JointBoost to allow this, resulting in weak-learners being learnt which are shared between both output classes and styles.

As an intuition, consider we are to recognize 'grass' at several different times of day and night (the 'style categories'). We might expect that weak learners based on the feature green should be shared by the day styles but not the night styles. Texture-based learners on the other hand might best be shared by all style categories.

We show how introducing style parameters helps in two object-class segmentation tasks, scene parsing, and road surface segmentation. For scene parsing, we use the Corel database, and compare with DenseBoost [5], a variant of TextonBoost and JointBoost [6], using Colour-HOG, HOG and textons as local features. We based the style parameters on semantic scene-classes, found by clustering the object label histograms of the training images. These are combined with local features during the StyP-Boost training. Global feature histograms are then used to predict the style parameters for test images. Figures 1a-c show the improvement achieved by using 4 style categories to help with segmentation, the 4 corresponding scene-classes being shown in 1d. As shown in table 1 we

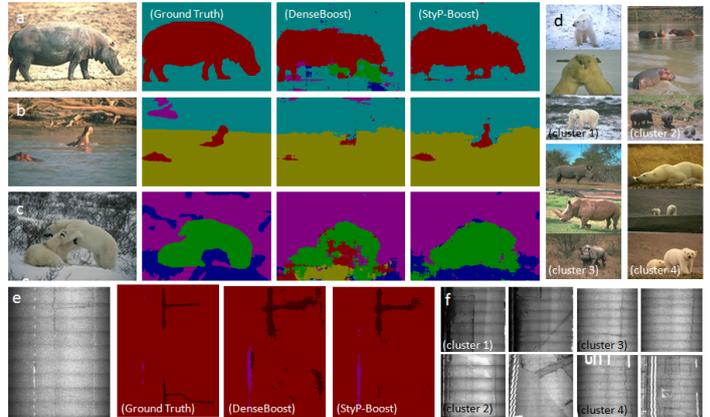


Figure 1: Results for scene parsing using the Corel database. (a)-(c) juxtapose original image, ground truth, DenseBoost (a TextonBoost variant) and StyP-Boost labelings. Notice that StyP-Boost both eliminates inappropriate classes (a)(c), and gives a stronger signal for true classes (b). (d) shows images from the 4 scene-classes found through clustering, which provide the 'style parameters' for the algorithm. (e) shows how a similar approach can be used for road surface segmentation to discover areas of cracking, where global surface appearance (lighting, markings etc.) provides the style parameters (f).

	D*Boost	StyP (2)	StyP (2-4)	Shotton [6]	He [3]	He [4]	Batra [1]
unary	0.767	0.822	0.845	0.684	0.669	0.710	0.632
un+CRF	0.787	0.837	0.860	0.746	0.800	0.810	0.828

Table 1: Comparing performance against previous results on the Corel database, unary and full CRF results given. Performance is given as the overall proportion of pixels correctly labeled. Incorporating style parameters both improves on DenseBoost, and previous results on the dataset. Using up to 4 styles gives the best performance.

achieve a notable improvement on the state-of-the-art for Corel, achieving best performance by combining 2-4 style categories in a cascade framework (see paper) and placing the unary estimates in a CRF with pairwise smoothing term. Further results are given on road surface data, where global style features help distinguish local cracking for different lighting, marking conditions etc (1e-f).

Future possibilities include 1) adapting the approach to learn the most discriminative style parameters in an unsupervised manner, rather than fixing these before training, and 2) using the bilinear framework as a way of embedding latent scene-class variables in larger probabilistic models (e.g. hierarchical CRFs [5]) in a principled fashion.

- [1] D. Batra, R. Sukthankar, and T. Chen. Learning class-specific affinities for image labelling. In *CVPR*, 2008.
- [2] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. In *Annals of Statistics*, 2000.
- [3] X. He, R. Zemel, and M. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *CVPR*, 2004.
- [4] X. He, R.S. Zemel, and D. Ray. Learning and incorporating top-down cues in image segmentation. In *Lecture Notes in Computer Science*, 2006.
- [5] L. Ladicky, C. Russell, P. Kohli, and P.H.S. Torr. Associative hierarchical crfs for object class image segmentation. In *ICCV*, 2009.
- [6] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *IJCV*, 2009.
- [7] A. Torralba, K. P. Murphy, and W.T. Freeman. Sharing visual features for multiclass and multiview object detection. *PAMI*, 29(5):854–869, 2007.