



University of Dundee

Multiple instance cancer detection by boosting regularised trees

Li, Wenqi; Zhang, Jianguo; McKenna, Stephen

Published in:

Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)

DOI:

[10.1007/978-3-319-24553-9_79](https://doi.org/10.1007/978-3-319-24553-9_79)

Publication date:

2015

Document Version

Early version, also known as pre-print

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Li, W., Zhang, J., & McKenna, S. (2015). Multiple instance cancer detection by boosting regularised trees. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015): 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part 1. (pp. 645-652). (Lecture notes in computer science ; Vol. 9349). Springer . 10.1007/978-3-319-24553-9_79

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The final publication is available at Springer via http://10.1007/978-3-319-24553-9_79

Multiple instance cancer detection by boosting regularised trees

Wenqi Li, Jianguo Zhang, and Stephen J. McKenna

CVIP, School of Computing, University of Dundee, Dundee, UK

Abstract. We propose a novel multiple instance learning algorithm for cancer detection in histopathology images. With images labelled at image-level, we first search a set of region-level prototypes by solving a submodular set cover problem. Regularised regression trees are then constructed and combined on the set of prototypes using a multiple instance boosting framework. The method compared favourably with competing methods in experiments on breast cancer tissue microarray images and optical tomographic images of colorectal polyps.

1 Introduction

Multiple instance learning (MIL) has recently been applied to histopathology image analysis for both segmentation and classification tasks [10, 16]. While training a tumour detector usually requires a large amount of high quality manual annotation [1], MIL methods can potentially infer tumour regions with image-level annotation, i.e., binary labels indicating whether tumour is present in the image. The MIL formulation is attractive as it does not require the effort of manually delineating image regions.

The general MIL inference rules are defined in the context of binary classification: a bag of instances is positive if at least one instance in the bag is positive, negative if all of the instances in the bag are negative. A common implementation of the rules in image classification treats each image as a bag, and regions in an image as instances. In terms of histopathology image analysis, an example application is to label an image as *cancer* if cancer is present in at least one region of the image, and as *non-cancer* otherwise.

In this paper, following the MIL setting, we propose a novel tree boosting algorithm for training a cancer detector. Our algorithm extends Multiple Instance Boosting (MILBoosting) [17] by boosting regularised trees with instance-to-prototype distances as features. The discriminative prototypes in our algorithm are searched by solving a submodular set cover problem. Our approach is validated on two types of histopathology images, namely, breast cancer tissue microarray (TMA) images and optical projection tomographic (OPT) images of colorectal polyps.

2 Related Work

Although MIL has been extensively studied since [12] and there exists a large literature (for a general review of MIL, see [2]), it was only recently applied to histopathology image analysis. Here we give a brief review of some of the most relevant work.

Zhao et al. [19] applied MILES [4] for 10 category histopathology image classification. Xu et al. [16] extended MILBoosting [17] to simultaneously detect and cluster multiple types of tissue region in TMA images. Kandemir et al. [9] evaluated MIL formulations on diagnosis of Barrett’s cancer with H&E images. Xu et al. [15] used MIL to classify colon cancer histopathology images with features extracted from convolutional neural networks.

Selecting instances as prototypes for bag classification was used previously with bags represented in terms of distances to prototypes [4, 7]. Our work extends MILBoosting to select prototypes with instance-to-prototype distances. We search a set of positive instance prototypes that is both discriminative and covers multiple modes of the appearance distribution. Instance-to-prototype distances are considered as features. A regularised regression tree boosting method is proposed to further select and combine the features.

Prototypes should satisfy three criteria; they should be 1) relevant: present in many positive images, 2) discriminative: dissimilar to negative instances, and 3) complementary: covering multiple types of positive instances. Song et al. [13] formalised these intuitions as a submodular set cover problem solved by a greedy algorithm. The set of prototypes was used as an initial training set for latent SVM. In this paper we adopt the ‘discriminateness’ of each prototype as a regularisation strength in a MILBoosting framework.

3 Method

3.1 Notation

Here we introduce notation adopted throughout the paper. We denote $\mathbf{x}_{ij} \in R^d$ as a d -dimensional feature representation of an instance (patch). Index ij represents the j^{th} instance in the i^{th} bag (image). $y_i \in \{0, 1\}$ represents the label of the bag, where 0 denotes non-cancer and 1 denotes cancer. The k^{th} prototype $\mathbf{p}_k \in R^d$ is an instance selected from the training instance set. In the following sections we introduce the two steps of our proposed method: searching for a set of discriminative prototypes and learning cancer detectors.

3.2 Discriminative prototypes

The discriminativeness of a prototype $g(\mathbf{p}_k)$ can be estimated as follows [13]: first find the m nearest neighbours of \mathbf{p}_k from the set of training instances $\{\mathbf{x}_{ij}\}$; then, counting the number of neighbours from the positive bags (denoted as m_{pos}), the ratio m_{pos}/m can be a measurement of discriminativeness. Greedy search

for a set of prototypes starts with an empty set of prototypes and a candidate set comprising all training instances. The most discriminative instance from the candidate set is then added to the prototype set, at the same time removing the prototype's m nearest neighbours from the candidate set. This is repeated until the candidate set is small enough (e.g., only 10% of initial candidates remain).

Song et al. [13] used the prototypes as an initialisation set for training latent SVM. We propose to combine the set of prototypes in a boosting framework. We further select prototype subsets and simultaneously learn an instance classifier by boosting regularised trees where we utilise $g(\mathbf{p}_k)$ as the regularisation strength.

3.3 Boosting with regularised regression trees

To learn detectors we adopt multiple instance boosting (MILBoosting) [17]. In MILBoosting, the instance classifier $F(\mathbf{x}_{ij})$ is formulated as a linear combination of T weak learners, i.e., $F(\mathbf{x}_{ij}) = \sum_{t=1}^T \alpha_t f_t(\mathbf{x}_{ij})$, where $f_t(\mathbf{x}_{ij})$ gives a score to each instance \mathbf{x}_{ij} ; α_t is the weight of f_t . The probabilities that cancer presents in an instance P_{ij} and in a bag P_i are respectively modelled as

$$P_{ij} = \frac{1}{1 + \exp(-F(\mathbf{x}_{ij}))} \quad \text{and} \quad P_i = 1 - \prod_j (1 - P_{ij}). \quad (1)$$

The instance classifier can be estimated by minimising the negative log-likelihood L of the bag labels:

$$L(y_i, F(\mathbf{x}_{ij})) = -\log \prod_i P_i^{y_i} (1 - P_i)^{(1-y_i)}. \quad (2)$$

Using the gradient boosting framework [6], L can be optimised by iteratively fitting weak learners f_t and optimising coefficients α_t . We adopt J -terminal regression trees as weak learners with a boosting shrinkage parameter ν [8]. However, instead of fitting regression trees to the feature set $\{\mathbf{x}_{ij}\}$, we first represent each instance in terms of distances to prototypes, i.e.,

$$\hat{\mathbf{x}}_{ij} = [d(\mathbf{x}_{ij}, \mathbf{p}_1), \dots, d(\mathbf{x}_{ij}, \mathbf{p}_k)], \quad (3)$$

where $d(.,.)$ is a distance measure, e.g., ℓ_2 -distance. Regression trees are then constructed on the new feature set $\{\hat{\mathbf{x}}_{ij}\}$. Each of the regression trees partitions the feature space into disjoint regions. The best variables to split and the optimal thresholds of the tree are searched by maximising information gain. In $\{\hat{\mathbf{x}}_{ij}\}$ each variable is associated with a prototype \mathbf{p}_k . Our method encourages trees constructed on $\{\hat{\mathbf{x}}_{ij}\}$ to split at those variables that are associated with large $g(\mathbf{p}_k)$. The motivation is to further select prototypes so that regression trees split at a few very discriminative prototypes, instead of splitting at many non-informative prototypes which could result in poor generalisation.

We introduce regularisation to properly control the variable to split in the tree construction process. The method we adopted is guided regularisation for

Algorithm 1 Summary of the proposed algorithm

```
1: procedure BOOSTING PROTOTYPES( $\{\mathbf{x}_{ij}\}, \{y_i\}, \nu, J, T$ )
2:    $\{\mathbf{p}_k\} \leftarrow$  Greedy search for prototypes (Section 3.2)
3:    $\{\hat{\mathbf{x}}_{ij}\} \leftarrow$  Transform  $\{\mathbf{x}_{ij}\}$  with  $\{\mathbf{p}_k\}$  (Formula (3))
4:   for  $t \leftarrow 1 \dots T$  do
5:     for all  $i, j$  do
6:       Compute  $r_{ij}^t \leftarrow -\frac{\partial L}{\partial f} \Big|_{f=f_{t-1}}$ 
7:     end for
8:     Fit a  $J$ -terminal regression tree  $f_t$  to  $r_{ij}^t$  (regularised by Formula (4))
9:     Line search:  $\alpha_t \leftarrow \arg \min_{\alpha} L(y_i, F_{t-1}(\hat{\mathbf{x}}_{ij}) + \alpha f_t)$ 
10:    Update classifier:  $F_t \leftarrow F_{t-1} + \nu \alpha_t f_t$  (shrinkage  $\nu$  is a fixed parameter)
11:  end for
12:  return instance classifier  $F_T$ 
13: end procedure
```

tree construction [5]. It was first proposed as a feature selection technique integrated in a random forest classifier; the selection of variables was guided by pre-computing variable importance from a preliminary random forest training. Here we combine the regularisation method with boosting trees. We utilise the discriminativeness $g(\mathbf{p}_k)$ as the regularisation strength instead of a preliminary random forest training.

Specifically given a set of K prototypes, we normalise $g(\mathbf{p}_k)$ as $\hat{g}(\mathbf{p}_k) = \frac{g(\mathbf{p}_k)}{\max_{k=1}^K g(\mathbf{p}_k)}$; when the tree chooses to split on the k^{th} feature of $\{\hat{\mathbf{x}}_{ij}\}$, the information gain is regularised by a function of $\hat{g}(\mathbf{p}_k)$:

$$\text{Gain}_R(k) = ((1 - \lambda) \cdot \gamma + \lambda \cdot \hat{g}(\mathbf{p}_k)) \cdot \text{Gain}(k), \quad (4)$$

where $\lambda \in [0, 1]$ is a free parameter to control the overall regularisation; $\gamma \in [0, 1]$ is a base regularisation coefficient. We calculate $\text{Gain}(k)$ as the reduction of variance at all leaf nodes when splitting at the k^{th} feature. The regularised regression tree can directly utilise discriminativeness to control the regularisation strength via Formula (4). The tree-based feature selection can capture non-linear variable interactions if $J > 2$. We set $J = 4$, $\gamma = 1$, and grid search for λ in our experiments. Algorithm 1 summarises the proposed procedure.

4 Evaluations

We evaluated the proposed method on cancer detection at 1) image level (predicting the presence of cancer in an unseen image) and 2) region level (localising the cancer region in an image). Two datasets were used in the experiments: a breast cancer TMA dataset and a colorectal polyp OPT dataset.

4.1 Breast cancer TMA images

The first dataset consists of 58 TMA breast cancer images stained with hematoxylin and eosin (H&E). 26 images are diagnosed as malignant, 32 as benign. For

Table 1. Cancer detection performance at image-level measured with AUC

Method	GPML [10]	RGPMIL [10]	Proposed
AUC	0.86	0.90	0.93

a fair comparison we used the feature sets made publicly available¹ by Kandemir et al. [10]. Each image was divided into 49 equally-sized instances. Each instance was further encoded with a 708-dimensional feature vector. This feature vector was comprised of SIFT descriptors, local binary patterns, colour histograms, and cell-level morphological features. Since instance location information is not available in the feature set, we focus on image-level performance evaluation.

We follow the 4-fold cross validation protocol used in [10]. For the proposed method we first applied the set cover search with $m = 20$. This usually selects 100 to 200 positive prototypes from a total of 1,500 instances. We set shrinkage parameter ν to 0.05, and the maximum number of iterations T to 300. The regularisation parameter λ was searched in the value set $\{0.1, 0.2, 0.3, \dots, 1.0\}$ with a 10-fold cross validation on the training folds. Averaged area under the ROC curve (AUC) was computed as the image classification performance measure (Table 1). The standard error of our method was 0.04. Equal Error Rate was 0.16 ± 0.03 . Note that Relational Gaussian Process MIL (RGPMIL) was designed for TMA images by explicitly modelling cells with a graph. Both GPML and RGPMIL outperformed widely-used MIL methods including EMDD [18], MILBoosting [17] and MI-SVM [3]. As shown in Table 1 our method achieved better image-level performance than the top-ranked methods.

4.2 OPT images of colorectal polyps

Dataset We evaluated both image- and instance-level cancer detection performance on 60 OPT images of colorectal polyps acquired using ultraviolet light and Cy3 dye [11]. Each 3-D image was of one colorectal polyp and consisted of 1024^3 voxels. 30 of the polyps were diagnosed as invasive cancer (ICA), and the other 30 as low-grade dysplasia (LGD). 3-D regions of cancer were annotated as sequences of 2-D slices by a trained pathologist. Our dataset for MIL evaluations consists of 200 2-D slices, with 100 slices randomly selected from ICA polyps and 100 from LGD. Fig. 2 shows some of the cancer slices and their annotations.

The pathologist was only asked to delineate major cancer regions with relatively high confidence rather than exhaustively trace all the cancer locations. Part of our motivation for applying MIL is that complete region-level annotations are difficult to obtain and validate. As a result, in ICA images, instance labels outside annotated regions were unknown. Since the instance annotations are not used in training MIL classifiers, quality of region annotation is not a concern in the training stage. We report instance-level test results based on the classifier output over all instances in LGD images, and all instances that have at least 50% overlap with ICA annotations.

¹ Link: <http://www.mipproblems.org/datasets/ucsb-breast/>

Table 2. Cancer detection at image-level and instance-level (with standard errors).

Method	MILBoosting [17]	Proposed	Inst-SVMs
AUC (image-level)	0.74 ± 0.04	0.79 ± 0.01	0.85 ± 0.03
F -measure (instance-level)	0.41 ± 0.01	0.45 ± 0.03	0.53 ± 0.05

Experiment protocol We treated each slice as a bag and densely extracted patches as instances. The size of each instance was 48×48 pixels. The sampling step size was 24 pixels in the training stage, and 12 in the test stage in both horizontal and vertical directions. We combined local binary patterns, SIFT features, and intensity histograms as instance features. The set cover search parameter was $m = 10$. Three-fold cross validation was conducted with the proposed method using the same grid search of parameters described in Section 4.1. Fig. 1(a) shows the image-level AUC of the proposed method plotted against the parameters T and ν on a validation set. The performance in terms of AUC is not very sensitive to T and ν . Choosing a large T and a small ν tends to give a high AUC. We set ν to 0.05, and T to 300.

We also implemented MILBoosting [17] as a baseline. Differences between the proposed method and MILBoosting are that the latter fits regression trees directly to $\{\mathbf{x}_{ij}\}$ without transformation and regularisation. In addition to the MIL methods we trained instance-level support vector machines (Inst-SVM) in a *fully supervised* setting as a comparison. In training Inst-SVM, instances with at least 50% overlap with the annotations were treated as cancer; the instances from LGD images were treated as non-cancer.

Results Table 2 compares image- and instance-level performance in terms of AUC and F -measure respectively. At image-level, Inst-SVM score was calculated as the maximum of instance scores in the image. At instance-level, score thresholds of each method were searched on the training set by maximising training F -measures. Our method outperformed MILBoosting in both image- and instance-level classification. However it was worse than fully supervised classification, as would be expected. Fig. 1(b) shows ROC curves at image-level. Fig. 2 shows a few cancer detection examples.

5 Conclusions

This paper introduced a novel multiple instance learning algorithm by combining discriminative prototype search with boosting of regularised regression trees. Empirical studies on two histopathology datasets showed that the method can achieve more accurate results than competing methods for both image- and instance-level classification. This work is based on an implementation of MIL inference rules in which instances are rectangular patches. In future we would consider generating instance regions using visual information [14].

Acknowledgement The authors would like to thank Ruixuan Wang and Siyamalan Manivannan for fruitful discussions.

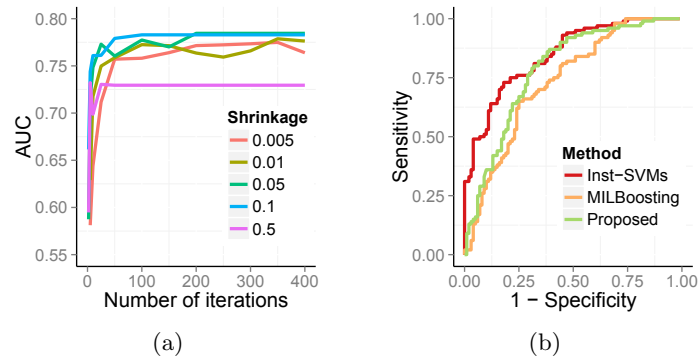


Fig. 1. Cancer detection at image-level (best viewed in color). (a) AUC of the proposed method against number of iterations T and shrinkage parameter ν . (b) ROC curves.

References

1. Akbar, S., Jordan, L., Thompson, A.M., McKenna, S.J.: Tumor localization in tissue microarrays using rotation invariant superpixel pyramids. In: ISBI (2015)
2. Amores, J.: Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence* 201, 81–105 (2013)
3. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: NIPS. pp. 561–568 (2002)
4. Chen, Y., Bi, J., Wang, J.Z.: MILES: Multiple-instance learning via embedded instance selection. *TPAMI* 28(12), 1931–1947 (2006)
5. Deng, H., Runger, G.: Gene selection with guided regularized random forest. *Pattern Recognition* 46(12), 3483–3489 (2013)
6. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. *Annals of Statistics* pp. 1189–1232 (2001)
7. Fu, Z., Robles-Kelly, A., Zhou, J.: MILIS: Multiple instance learning with instance selection. *TPAMI* 33(5), 958–977 (2011)
8. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media (2009)
9. Kandemir, M., Feuchtinger, A., Walch, A., Hamprecht, F.A.: Digital pathology: Multiple instance learning can detect Barrett’s cancer. In: ISBI (2014)
10. Kandemir, M., Zhang, C., Hamprecht, F.A.: Empowering multiple instance histopathology cancer diagnosis by cell graphs. In: MICCAI, pp. 228–235 (2014)
11. Li, W., Zhang, J., McKenna, S.J., Coats, M., Carey, F.A.: Classification of colorectal polyp regions in optical projection tomography. In: ISBI. pp. 736–739 (2013)
12. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: NIPS. pp. 570–576 (1998)
13. Song, H.O., Girshick, R., Jegelka, S., Mairal, J., Harchaoui, Z., Darrell, T., et al.: On learning to localize objects with minimal supervision. In: *International Conference on Machine Learning* (2014)
14. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M.: Selective search for object recognition. *IJCV* 104(2), 154–171 (2013)

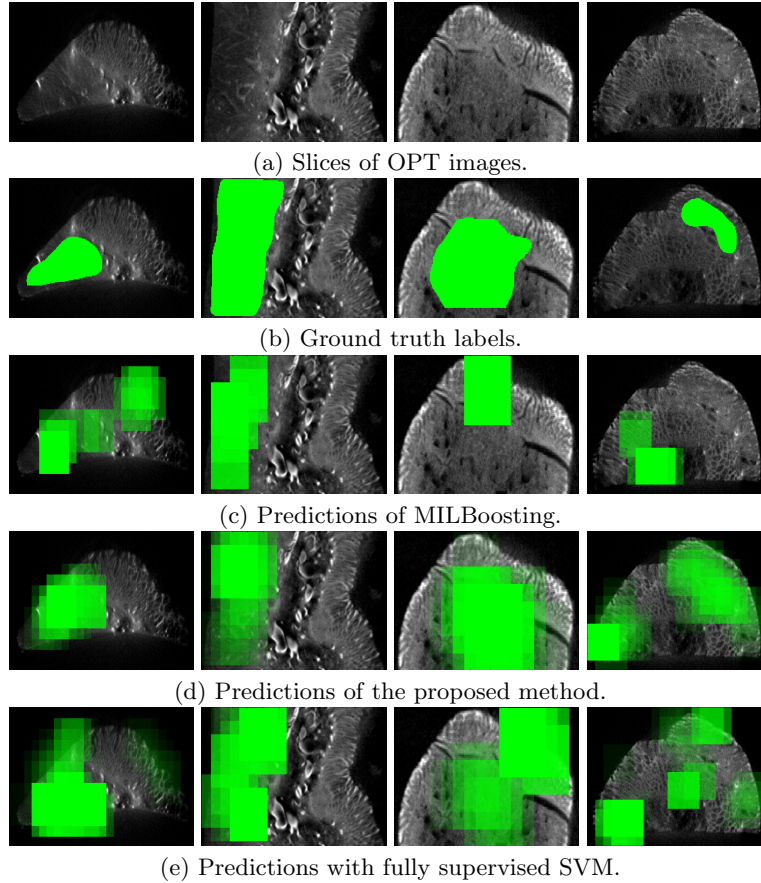


Fig. 2. Instance-level annotations and predictions. Green patches in third to fifth rows indicate scores of the instances are greater than the learned threshold. Instances with higher scores were mapped to higher opacity values (best viewed in colour).

15. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Chang, E.I.C., et al.: Deep learning of feature representation with multiple instance learning for medical image analysis. In: ICASSP. pp. 1626–1630 (2014)
16. Xu, Y., Zhu, J.Y., Chang, E.I.C., Lai, M., Tu, Z.: Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis* 18(3), 591–604 (2014)
17. Zhang, C., Platt, J.C., Viola, P.A.: Multiple instance boosting for object detection. In: NIPS. pp. 1417–1424 (2005)
18. Zhang, Q., Goldman, S.A.: EM-DD: An improved multiple-instance learning technique. In: NIPS. pp. 1073–1080 (2001)
19. Zhao, D., Chen, Y., Correa, N.: Automated classification of human histological images, a multiple-instance learning approach. In: Life Science Systems and Applications Workshop, 2006. IEEE/NLM. pp. 1–2. IEEE (2006)