Batch Mode Active Learning for Object Detection Based on Maximum Mean Discrepancy

Yingying Liu
School of Computer Science and Engineering
University of New South Wales
Kensington NSW 2052, Australia
Email: liuy@cse.unsw.edu.au

Yang Wang
National ICT Australia
Australian Technology Park
Eveleigh NSW 2015, Australia
Email: Yang.Wang@nicta.com.au

Arcot Sowmya
School of Computer Science and Engineering
University of New South Wales
Kensington NSW 2052, Australia
Email: sowmya@cse.unsw.edu.au

***Abstract*—Various active learning methods have been proposed for image classification problems, while very little work addresses object detection. Measuring the informativeness of an image based on its object windows is a key problem in active learning for object detection. In this paper, an image selection method to select the most representative images is proposed based on measuring their object window distributions by Maximum Mean Discrepancy (MMD). Then an active learning method for object detection is introduced based on MMD-based image selection. Experimental results show that MMD-based image selection can improve object detection performance compared to random image selection. The proposed active learning method based on MMD image selection also outperforms a classical active learning method and passive learning method.**

I. Introduction

Traditional supervised learning methods require large amounts of training data. Various machine learning techniques have been used to reduce this requirement. Active learning is a well-known technique that provides feasible solutions when the training data is large and annotation is expensive.

The key concept of active learning is to select the most informative data in each iteration. A literature survey of active learning by Settles [1] discusses six active learning query strategies. In all the six strategies, data "uncertainty" and "representativeness" are the main factors that determine data informativeness.

Early active learning methods selected one data item in each iteration, which is quite time-consuming. In recent years, batch mode active learning methods that select a set of data points has drawn lots of attention, as they can reduce the learning time. Batch mode active learning has been applied in different areas, for example, text categorization [2], medical image classification [2], and object detection [3].

Object detection is a fundamental problem in computer vision. Many learning techniques have been applied on object detection, for example boosting [4], SVM [5], [6] and random forest [7]. Among object detection methods, deformable part models (DPM) [6] is a well-known SVM based object detector which has achieved outstanding performance.

In this work, we first introduce an image selection method based on Maximum Mean Discrepancy (MMD) [8] to select the most representative images. MMD is utilized to select the most representative data by measuring data distributions. As an image may contain more than one object window, we propose to describe an image by its object windows. Then we present a batch mode MMD-based active learning method for object detection by utilizing our proposed MMD-based image selection method. An SVM based state-of-art object detector DPM [6] is used as the object detector in this work.

A. Related Work

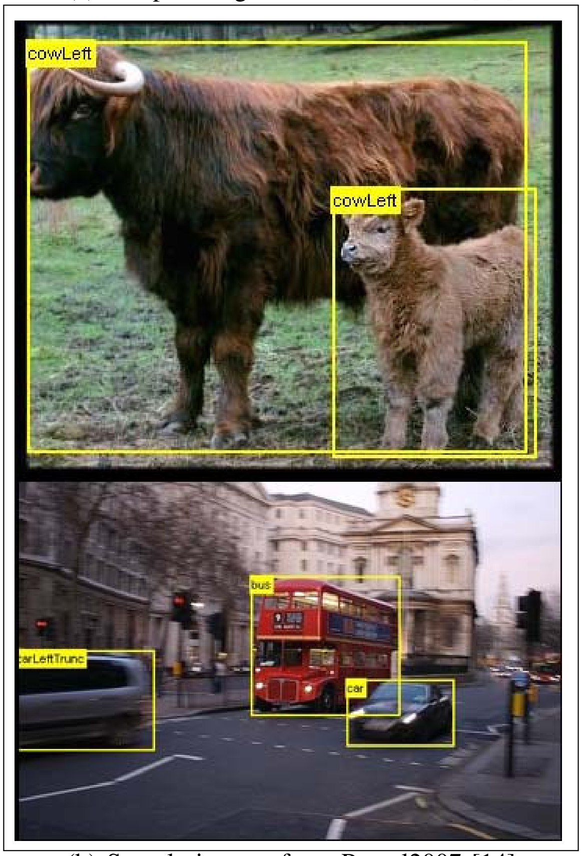
Three SVM based active learning methods "Simple Margin", "MaxMin Margin" and "Ratio Margin", have been proposed [9]. "Uncertainty" of data is used to determine the data informativeness. The most uncertain data in the current model are considered to reduce the version space the most. Although the last two methods are more robust, "Simple Margin" is more applicable in practice. For the active learning based object detection problem, Vijayanarasimhan and Grauman [3] proposed a method based on "Simple Margin". In their work, images that contain the top ranked object windows by the "Simple Margin" method were selected in each active learning iteration.

In recent years, active learning methods based on data "representativeness" measured by MMD have drawn attention. MMD was first introduced by Gretton, Borgwardt, Rasch, Schölkopf and Smola [8] to analyse and compare distributions in two sample test problems. It can be applied to arbitrary data, for example neuroscience, bioinformatics, and attribute matching. It works well on high dimensional data with low sample size. This makes it suitable for computer vision problems, where high-dimensional features are often involved. Various MMD based techniques have been proposed recently. For example, Baktashmotlagh, Harandi, Lovell and Salzmann [10] applied MMD on the transfer learning problem by measuring the dissimilarity between the empirical distributions of source and target data. Iyer, Nath and Sarawagi [11] proposed a method to estimate the class ratio in unlabelled data based on MMD. Chattopadhyay *et al.* [12] proposed an active learning method for image classification by minimizing the MMD between the newly formed training data (selected data in current iteration with labelled data in previous iterations) and unlabelled data. However, their method is not suitable for object detection problems. This is because multiple object windows can exist in one image in object detection problems, while in image classification problems each image represents one

978-1-4673-6795-0/15/$31.00 ©2015 IEEE

object. In this work, we first propose an MMD-based image selection method to select the most representative images based on their object windows. Then an active learning method for object detection using DPM detector is introduced by applying the proposed MMD-based image selection method.

(a) Sample images from Caltech101 [13]



(b) Sample images from Pascal2007 [14]

Fig. 1. (a) sample images for image classification problem (b) sample images for object detection problem.

B. Contribution

Our contributions are two fold:

(i) an MMD-based image selection method to select the most representative images for object detection is proposed. An occurrence matrix is introduced to select the most representative images by representing them using their object windows.

(ii) an MMD-based batch mode active learning method for object detection based on the proposed MMD-based image selection is introduced. To our knowledge, this is the first work applying MMD based active learning to object detection.

The paper is organized as follows: in section 2, an MMD-based image selection method is proposed, which aims to select the most representative images based on the distribution of their object windows; section 3 describes our MMD-based active learning method for object detection; experimental results and analysis are given in section 4; in section 5, a discussion on the proposed method is presented.

II. MMD-BASED IMAGE SELECTION

MMD was first described to analyse and compare distributions [8]. Given two datasets $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$, which are independently and identically sampled from two distributions p and q, the similarity of p and q can be measured by MMD:

$$MMD[\mathcal{F}, p, q] = \sup_{f \in \mathcal{F}} (E_x(f(x)) - E_y(f(y))) \quad (1)$$

where function class $\mathcal{F}$ is the unit ball of a reproducing kernel Hilbert space $\mathcal{H}$. The MMD of two distributions p and q is the largest expectation difference of their samples X and Y over the functions in $\mathcal{F}$.

In this work, we propose a method to select the most representative images from a dataset. Given a dataset T and the labelled data L in it, a dataset S can be selected from the unlabelled data in T according to equation 2. Therefore, the newly formed data combining S and L, can maximally represent the distribution of the whole dataset T.

$$MMD[\mathcal{F}, S \cup L, T] = || \frac{\sum\limits_{i \in S \cup L} \Phi(x_i)}{N_s + N_l} - \frac{\sum\limits_{i \in T} \Phi(x_i)}{N_t} ||_H^2 \quad (2)$$

where N_t, N_l and N_s are the numbers of the data items in the three sets T, L and S. Φ is a function in the unit ball of the reproducing kernel Hilbert space.

In image classification problems, each image contains one object and usually the image background is quite simple (see Fig. 1 (a)). Therefore, the image features correspond well to the object in it. However in object detection problems, more than one object window may exist in a single image (see Fig. 1 (b)). Therefore, an occurrence matrix M is introduced in this work to measure the distribution of the image over its object windows.

M is a $N_i * N_w$ matrix composed of 0's and 1's, where N_i is the image number and N_w the number of object windows in all images. $M(i,j) = 1$ $(s.t.\ i \in \{1, \cdots, N_i\}, j \in \{1, \cdots, N_w\})$ means the j^{th} object window belongs to the i^{th} image. In the following $m_{i,j} = M(i,j)$. M shows the occurrence relationship between the image and its object windows.

By utilizing M, the most representative images can be selected by minimizing the cost function as below:

$$\arg\min_{\alpha} \Big\| \frac{\sum_{i=1}^{N_i} \alpha_i * \sum_{j=1}^{N_w} m_{i,j} * \Phi(x_j) + \sum_{i=1}^{N_l} \Phi(x_i)}{N_s + N_l} - \frac{\sum_{i=1}^{N_t} \Phi(x_i)}{N_t} \Big\|_H^2$$
$$s.t. \quad \alpha \in \{0,1\}, \quad \alpha^T \mathbf{1}_I = N_{si}. \tag{3}$$

where α is a vector composed of 0 and 1, which indicates whether the image is selected or not. When $\alpha_i = 1$, the i^{th} image is selected. N_s is the number of object windows in the selected images. N_{si} is the number of selected images. N_i is the number of the unlabelled images. N_w is the number of the object windows in unlabelled images. N_l is the number of object windows in labelled images. N_t is the number of object windows in all images. $\mathbf{1}_I$ is a $N_i * 1$ vector, all of whose elements are 1. The first part of the cost function is the distribution of selected images and already labelled images. The second part of the cost function is the distribution of all the images.

Cost function (3) can be written in the following form:

$$\begin{aligned} \min \quad & \frac{\alpha^T M K_{UU} M^T \alpha}{(N_s + N_l)^2} + \frac{2\mathbf{1}_L^T K_{LU} M^T \alpha}{(N_s + N_l)^2} \\ & - \frac{2\mathbf{1}_T^T K_{TU} M^T \alpha}{N_t * (N_s + N_l)} + const \\ & s.t. \quad \alpha \in \{0,1\}, \quad \alpha^T \mathbf{1}_I = N_{si}. \end{aligned} \tag{4}$$

where U is the unlabelled object windows, L is the labelled object windows, and T is the object windows in all images. K_{UU}, K_{LU}, K_{TU} are three kernel Gram matrices between object windows datasets U and U, L and U, T and U. Gaussian kernel is used in this work. Similar to others [8], the Gaussian kernel bandwidth is set to be the median value of the distance between the points in the aggregate data used in definition (4). $\mathbf{1}_L$ is a $N_l * 1$ vector and $\mathbf{1}_T$ is a $N_t * 1$ vector, all of their elements are 1.

By relaxing α to be composed of continuous values in $[0, 1]$, the problem can be solved using quadratic programming:

$$\begin{aligned} \min \quad & 0.5\alpha^T H \alpha + f^T \alpha \\ & H = \frac{M K_{UU} M^T}{(N_s + N_l)^2}, \\ & f = [\frac{\mathbf{1}_L^T K_{LU} M^T}{(N_s + N_l)^2} - \frac{\mathbf{1}_T^T K_{TU} M^T}{N_t * (N_s + N_l)}]^T, \\ & s.t. \quad \alpha \in [0,1], \alpha^T \mathbf{1}_I = N_{si}. \end{aligned} \tag{5}$$

In (5), while N_{si}, N_l and N_t are given, N_s is unknown. Ideally, N_s is the number of object windows in the selected images. However, it is quite difficult to obtain its value, since the number of object windows in each image may be different. Therefore in practice, an estimated value is used: $N_s = N_{avg} * N_{si}$. N_{avg} is the average number of object windows in one image. Many quadratic programming solvers have been developed in several optimization software, for example, CVX [15], [16], Matlab optimization toolbox [17]. In this work, Matlab optimization toolbox is used to solve the quadratic programming problem, as the scale of our problem is not quite large.

III. MMD-based Active Learning for Object Detection

In section 2, an MMD-based image selection method based on object windows was proposed. However, in active learning object detection problems, object windows are not given. Therefore, in this section, we propose an MMD-based active learning object detection method based on the distribution of reliably detected windows in images. We assume that detected object windows with high scores are more reliable than ones with low scores. Thus for active learning object detection, the reliably detected windows are the top scoring detections in each image in the previous iteration. To test the performance of the proposed method, a dataset is evenly separated into training and test datasets. The active learning methods are trained on the training dataset, and evaluated on the test dataset.

Algorithm 1 MMD-based active learning for object detection

Require: Initial detector: D_1, training image set $IMG = \{img_1, img_2, \cdots, img_n\}$, active learning step N_{si}.
Ensure: Object detector D_{target}.
Create initial windows proposal set WIN_1 by applying D_1 to training image set IMG:
    $WIN_1 = \{w_1, \cdots, w_n\}$.
  $IMG_L = \emptyset$: the labelled images in the last iteration.
  $NEG_L = \emptyset$: the negative windows obtained from labelled images IMG_L.
for $Iteration : index = 1 \rightarrow m$ **do**
  Image selection:
  Select N_{si} images $IMG_S = \{img_i, img_j, \cdots, img_k\}$ from the unlabelled images, based on the proposed MMD-based image selection method according to the distribution of detected windows WIN_{index}.
  Negative data creation:
  Create negative data NEG_S from IMG_S.
  Update IMG_L and NEG_L:
  $IMG_L = IMG_L \cup IMG_S$, $NEG_L = NEG_L \cup NEG_S$.
  Train object detector $D_{index+1}$ on IMG_L and NEG_L.
  Created new windows proposal set $WIN_{index+1}$ by applying $D_{index+1}$ on IMG.
end for
$D_{target} = D_{m+1}$

The framework of the proposed active learning object detection method is described in Algorithm 1. In the first step, an initial object windows set WIN_1 is created by applying an initial detector D_1 to the training image set IMG. WIN_1 is composed of the detected windows in each image, i.e., $WIN_1 = \{w_1, \cdots, w_n\}$. w_i is all the reliably detected windows in the i^{th} training image. Then the MMD-based image selection method is applied to select the most representative images by estimating the distribution of the object windows in WIN_1. After that, object detector D_2 is trained on all labelled images IMG_L and their negative windows NEG_L. Note that $IMG_L = \emptyset$ and $NEG_L = \emptyset$ at the beginning.

Then, detected windows set WIN_2 is created by applying D_2 to the training dataset, which will then be used in the second iteration of active learning. The labelled windows L used in the proposed image selection method include the annotated positive windows and negative windows created from the same image. Finally, the target detector D_{target} is the detector trained in the last iteration.

In object detection, negative data are not always available. Hard negatives and random negatives are two main negative data sampling methods. Hard negatives are false positive windows with high detection scores. It was introduced by Felzenszwalb, Girshick, McAllester and Ramanan [6], and used by other researchers [3]. Random negatives are the negative windows sampled in the background of each image frame. In this work, we explore the effect of using both hard negatives and random negatives.

IV. Experimental Results

We first evaluate the performance of the MMD-based image selection method. Then experiments on MMD-based active learning for object detection are conducted. TUD-pedestrian dataset [18] is used as the experimental dataset. It contains 250 pedestrian detection images. We evenly split it into two datasets, one for training, and the other for testing. A state-of-art detector 1-component DPM [6] is used in the experiments. HOG feature pyramid [6] is used as the window descriptor. Detections that have more than 50% overlap with the ground truth bounding boxes are correct detections. Multiple detections are penalized. Average precision and ROC (FPPI-TPR) curves are reported for each method.

A. MMD-based Image Selection

To evaluate the capability of the proposed MMD-based image selection method, we first apply it on ground truth windows of the training images. The performance of image selection task is given in terms of the accuracy of the object detectors trained on the selected images. The most representative images are selected based on definition (3). Object detectors are trained on the selected images to test their ability to represent the whole dataset and tested on the test set. The proposed image selection method is compared with a random selection method.

Six experiments (with 20, 40, 60, 80, 100, 125 images selected) were conducted. Note that in all the experiments, $N_l = 0$ in definition (3). In each experiment, the same number of images are selected in both proposed image selection method and random selection method. Then the groundtruth data of the selected images are used to train the detectors in both methods. Negative data is created by randomly sampling 5 background windows in each image. In each experiment, the same randomly sampled negative dataset is used in both methods. The number of the negative windows is 5 times the the number of selected images all through the experiments. The experiments of random selection method are repeated 5 times and the average performance are reported. Since function (5) has unique solution, the proposed method has the same image selection results in the same iteration no matter how many times we run the experiments.

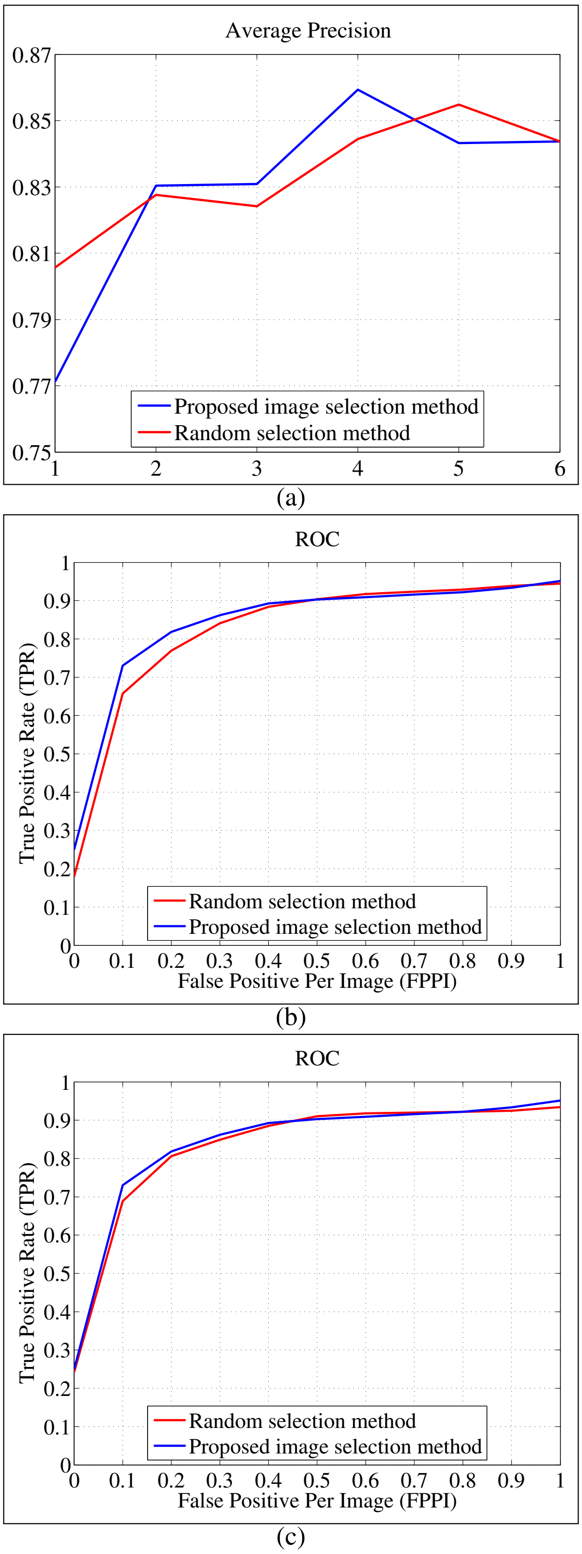


Fig. 2. (a) Average precision of image selection methods. (b) ROC curves at the 4^{th} experiment. (c) ROC curves of the highest performances: the 4^{th} experiment of the proposed method and the 5^{th} experiment of random selection method.

In Fig. 2 (a), the average precision curves of the two methods is shown. In the 2^{nd}, 3^{rd} and 4^{th} experiments, the proposed image selection method has better performance than the random selection method. Moreover, the proposed image selection method has the highest performance among all experiments. This shows that the proposed method is able to select more representative images than random selection alone.

It can also be seen that the proposed image selection method is not as good as random selection when selecting the first 20 images. This is because a small number of selected images is not sufficient to represent the original training data distribution. Although the proposed method is based on selecting the most representative data from the training data, it cannot guarantee good performance when the distribution of the selected data is not close enough to the original distribution. It can be noticed that both methods have their highest performance using subsets of training images, which shows that training on the whole dataset is not necessary for both methods. In Fig. 2 (b), the ROC curves of both methods at the 4^{th} experiment are shown. It can be seen that the proposed method outperforms the random selection method. In Fig. 2 (c), the ROC curves of the two methods when they attain their best performances are shown. It can be seen that the best performance of the proposed method outperforms the best performance of random selection method.

B. *MMD-based Active Learning for Object Detection*

The proposed MMD-based active learning object detection method is compared with two other methods, namely simple margin and passive learning. We set up the iteration number to be 10. In each iteration, 10 images are selected in all three methods. To compare the performance of the different methods with respect to different negative data, two experiments were conducted, one each with hard negatives and random negatives.

For the proposed MMD-based active learning object detection method, in each iteration, images are selected according to definition (3). The first 5 highest scored detections in each image are used as reliable object windows. An initial detector is required in the proposed method to create the object windows for the first iteration. In this paper, we use a DPM detector [6] trained on PASCAL2007 [14].

We implement the simple margin active learning method following a published work [3]. Images that contain the most uncertain object windows are selected. The most uncertain object windows are the ones closet to the separating hyperplane:

$$O^* = \arg\min_{O_i \in U} ||w^T * \phi(O_i)|| \tag{6}$$

where w is the separating hyperplane normal, and U are the object windows. The experimental settings are different between the proposed method and [3]. In [3], active learning is applied to select the most informative images from both positive images and negative images, while the proposed method aims to select the most informative images from positive images. Note that positive images in object detection are the ones that contain the object, while negative images are the ones that do not contain the object.

For the passive learning method, in each iteration, a set of images are randomly sampled from the unlabelled images and added into the training dataset. The experiments of passive learning method are repeated 5 times and the average performance are reported.

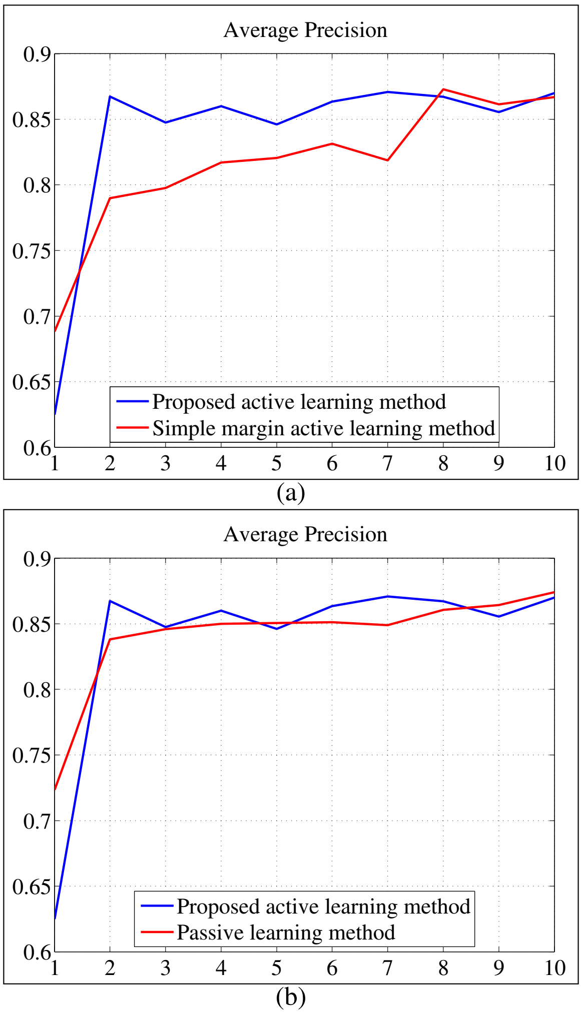


Fig. 3. Average precision curves when using hard negatives. (a) shows the comparison between the proposed method and simple margin method, (b) shows the comparison between the proposed method and passive learning method.

From Fig. 3, it can be concluded that the proposed method outperforms both the simple margin method and random selection method from the 2^{nd} iteration when using hard negatives. Although in the first iteration, the performance of the proposed method is not as good as the comparison methods, it outperforms them in later iterations. The low start of the proposed method is due to the small number of selected training examples, which is not sufficient to reliably construct the distribution of the whole dataset.

Similar conclusions can be made when using random negatives in Fig. 4. However, it can be seen that when using random negatives, the proposed method grows more slowly than when using hard negatives. For example, the proposed method with hard negative dataset exceeds 0.85 average precision in the 2^{nd} iteration (see Fig. 3), while it exceeds 0.85 average precision only in the 5^{th} iteration when use random negatives (see Fig. 4). Moreover, when using random negatives, passive learning has higher performance than the proposed method in

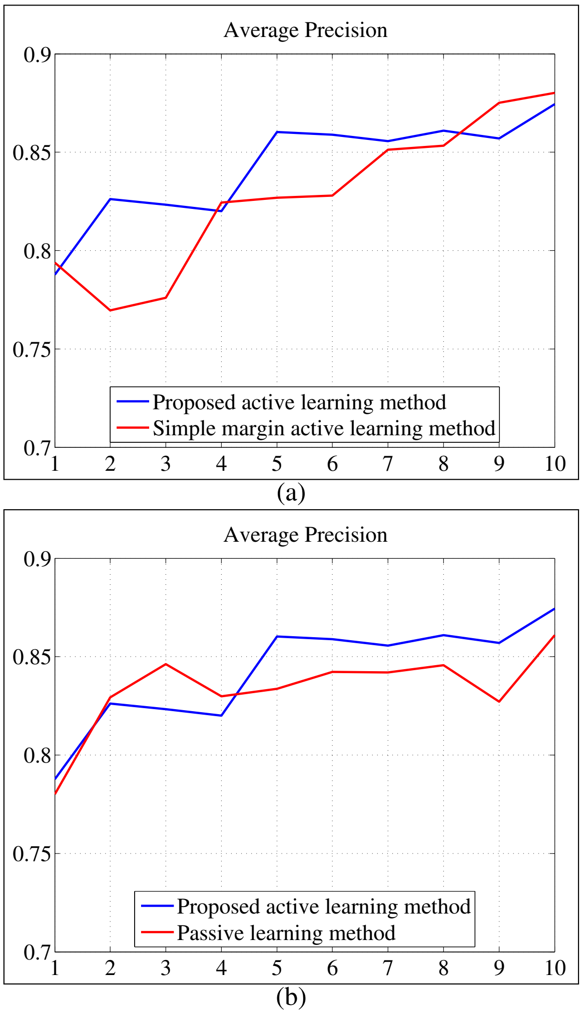


Fig. 4. Average precision curves when using random negatives. (a) shows the comparison between the proposed method and simple margin method, (b) shows comparison between the proposed method and passive learning method.

the first few iterations. This is because the proposed method does not consider the negative data distribution. Thus it cannot guarantee selection of images that contain the best negative data. However, we can also see that the proposed method outperforms the passive learning method after more negative data is selected in the following iterations. This is because when the negative data size increases, its distribution will be closer to the real distribution of the negative data in all images. Thus the difference between the two negative data subsets used in passive learning and the proposed method will be smaller.

In Table I, the highest average precision obtained in each experiment is shown. It can be seen that with hard negative data, the proposed method attains the highest average precision faster than the other two methods. When using random negatives, all three methods attain their highest average precisions at the last iteration. The ROC curves in Fig. 5 show the performances when the proposed method attains its highest average precision. As we can see, the proposed method outperforms passive learning method irrespective of hard negatives or random negatives. When using random negatives, the proposed method attains similar performance to simple margin method.

TABLE I. HIGHEST AVERAGE PRECISIONS OBTAINED BY DIFFERENT METHODS.

Method	Negative type	Average precision	Iteration
Proposed method	Hard negative	0.87	2
Simple Margin	Hard negative	0.87	8
Passive learning	Hard negative	0.87	10
Proposed method	Random negative	0.87	10
Simple Margin	Random negative	0.88	10
Passive learning	Random negative	0.86	10

It can be concluded from the experiments that the proposed method outperforms the two comparison methods, although it has a low start when the training data is small.

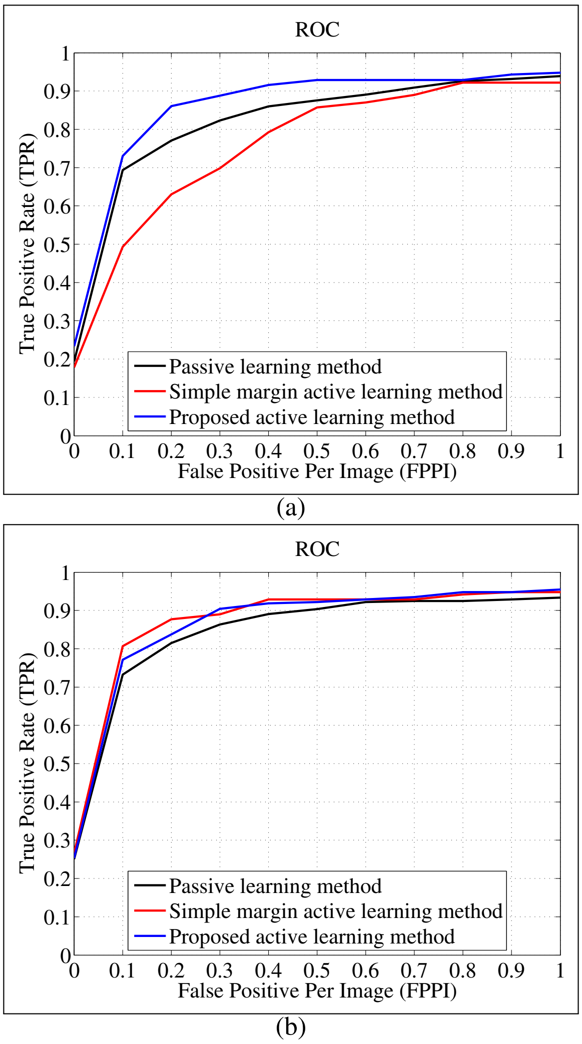


Fig. 5. ROC curves. (a) shows the ROC curves at the 2^{nd} iteration when using hard negatives. (b) shows the ROC curves at the 10^{th} iteration when using random negatives.

V. Discussion

In this paper, we present an MMD-based image selection method to select the most representative images based on the distribution of object windows. Experimental results show that the proposed image selection method outperforms the random selection method and attains the highest average precision.

Based on the proposed image selection method, an MMD-based active learning based object detection method is introduced in this paper. Experimental results show that the proposed method outperforms simple margin, which is a classical method in SVM based active learning. It also outperforms the passive learning method. However, the proposed method has some drawbacks. First, it has a low start when the number of selected data is small. The reason is that a small number of data samples cannot reliably reconstruct the distribution of the original data. Second, the proposed method relies on the detected windows in the previous iteration. Therefore, if an object window is not detected in the last iteration, it has no effect on the image selection.

In future, more research can be done on the proposed active learning object detection method. A method to select a combination of random selection and the proposed image selection method can be designed to address the low start problem of the proposed method. Unsupervised window proposal generation methods [19]–[21] can be utilized to replace detected windows in the proposed active learning object detection method. This can yield a weakly supervised object detection method. Moreover, selection of positive and negative data according to their distributions is worthy of more research. Finally, current active learning methods retrain the detector model at each iteration, which is quite time-consuming. Model updating strategies can be explored to improve the model training efficiency in the active learning iterations.

REFERENCES

[1] B. Settles, "Active learning literature survey," University of Wisconsin-Madison, Tech. Rep., 2010.

[2] S. C. H. Hoi, R. Jin, and M. R. Lyu, "Large-scale text categorization by batch mode active learning," in *Proceedings of the 15th International Conference on World Wide Web*, ser. WWW '06. New York, NY, USA: ACM, 2006, pp. 633–642. [Online]. Available: http://doi.acm.org/10.1145/1135777.1135870

[3] S. Vijayanarasimhan and K. Grauman, "Large-scale live active learning: Training object detectors with crawled data and crowds," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1449–1456. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2011.5995430

[4] P. Dollar, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2009, pp. 91.1–91.11, doi:10.5244/C.23.91.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, June 2005, pp. 886–893 vol. 1.

[6] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, Sept 2010.

[7] J. Gall, A. Yao, N. Razavi, L. V. Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2188–2202, November 2011.

[8] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," vol. 13. JMLR.org, Mar. 2012, pp. 723–773. [Online]. Available: http://dl.acm.org/citation.cfm?id=2188385.2188410

[9] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1162/153244302760185243

[10] M. Baktashmotlagh, M. Harandi, B. Lovell, and M. Salzmann, "Unsupervised domain adaptation by domain invariant projection," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 769–776.

[11] A. Iyer, S. Nath, and S. Sarawagi, "Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, T. Jebara and E. P. Xing, Eds. JMLR Workshop and Conference Proceedings, 2014, pp. 530–538. [Online]. Available: http://jmlr.org/proceedings/papers/v32/iyer14.pdf

[12] R. Chattopadhyay, Z. Wang, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, "Batch mode active sampling based on marginal probability distribution matching," *ACM Trans. Knowl. Discov. Data*, vol. 7, no. 3, pp. 13:1–13:25, Sep. 2013. [Online]. Available: http://doi.acm.org/10.1145/2513092.2513094

[13] R. F. L. Fei-Fei and P. Perona, "Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories," 2004, iEEE. CVPR 2004, Workshop on Generative-Model Based Vision.

[14] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html, 2007.

[15] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[16] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[17] "Matlab and optimization toolbox release 2014a," The MathWorks Inc., 2014.

[18] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[19] I. Endres and D. Hoiem, "Category-independent object proposals with diverse ranking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 2, pp. 222–234, Feb 2014.

[20] K. E. A. van de Sande, J. R. R. Uijlings, T. Gevers, and A. W. M. Smeulders, "Segmentation as selective search for object recognition," in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1879–1886. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2011.6126456

[21] S. Manen, M. Guillaumin, and L. V. Gool, "Prime object proposals with randomized prim's algorithm," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 2536–2543. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2013.315