

# Balanced Active Learning Method for Image Classification

Dávid Papp<sup>a</sup> and Gábor Szűcs<sup>a</sup>

## Abstract

The manual labeling of natural images is and has always been painstaking and slow process, especially when large data sets are involved. Nowadays, many studies focus on solving this problem, and most of them use active learning, which offers a solution for reducing the number of images that need to be labeled. Active learning procedures usually select a subset of the whole data by iteratively querying the unlabeled instances based on their predicted informativeness. One way of estimating the information content of an image is by using uncertainty sampling as a query strategy. This basic technique can significantly reduce the number of label needed; e.g. to set up a good model for classification. Our goal was to improve this method by balancing the distribution of the already labeled images. This modification is based on a novel metric that we present in this paper. We conducted experiments on two popular data sets to demonstrate the efficiency of our proposed balanced active learning (BAL) approach, and the results showed that it outperforms the basic uncertainty sampling.

**Keywords:** active learning, image classification, uncertainty sampling, balanced method

## 1 Introduction

Nowadays, for categorizing the huge amounts of visual contents both online and offline, image classification is an indispensable tool. The number of images available online is increasing with the rapidly growing Internet usage. Besides this numerous electronic devices are capable of taking a digital picture (e.g. cameras, telephones and so on), furthermore, smart devices are a click away from uploading and sharing these pictures. This leads to massive data warehouses that need to be structured, i.e. categorized. The classification of images requires labeled instances, but usually the contents are unlabeled, and labeling them is a costly manual process. Active learning [2] is a way of addressing this problem since it selects a subset of the data

---

<sup>a</sup>Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., H-1117, Budapest, Hungary.  
E-mail: {pappd,szucs}@tmit.bme.hu

by iteratively querying the most informative image(s) from the unlabeled ones, and then it builds the classification model based on this subset instead of the whole data. In this way, the active learning algorithm seeks to label as few instances as possible (i.e., minimizing the labeling cost) while it attempts to retain the same level of accuracy that could be achievable by using the total dataset. The most important question is how to estimate the informativeness of unlabeled instances, since different query strategies may lead to a better or worse classification accuracy, compared to that using random sampling. There are many proposed strategies in the literature, e.g. uncertainty sampling [33], query-by-committee [31], expected model change [6], expected error reduction [20]. Uncertainty sampling is a widely used, and the simplest query strategy framework, which aims to query the instances with least certainty about their labels.

The main goal of our study was to improve the accuracy by enhancing it with a distribution analysis on the labeled dataset. We developed and implemented a solution to determine the distribution based on a novel penalty metric. We demonstrated the efficiency of the proposed approach on two large datasets; namely, the PASCAL VOC2010 [10] training and validation data, and the Caltech101 [17] image collection (see Section 5 for details and Figure 1 for sample images). Note that the queried images could be re-annotated by relying on human labor to deliver a more useful feedback on the effectiveness of the proposed approach in a production-like active learning environment. Of course, this approach would be resource-intensive, hence we decided to use a viable substitute to measure the benefits of the BAL algorithm (where we used the ground truth categorizations provided by the datasets).

## 2 Related Work

Now, we will briefly review the related studies on active learning. We will focus on computer vision problems (especially on images) [6, 13, 31, 33] that require substantial amounts of training data to perform accurate classification. Nowadays active learning is gaining increasing interest in the computer vision community. In the following, we review relevant work on active learning and image classification. Some of these studies are only theoretical ones [7] without any experimental part or results. The authors of [32] used simple margin selection method for SVM for the active selection of object windows in images. The system autonomously refines its models by actively requesting crowd-sourced annotations on images crawled from the Web. The authors of [24] suggested combining spectral and spatial information directly in the iterative process of sample selection, where the criterion involves the concept of spatial entropy.

As we already mentioned in the Introduction, uncertainty sampling is a frequently adopted strategy in active learning (because of its simplicity), which builds upon the presence of uncertainty in classification. Minakawa et al. [23] apply this uncertainty sampling method to image sequence recognition in an active learning scenario; a margin sampling criterion and entropy criterion were used in the condition part of this method. Many other works on uncertainty sampling methods are



Figure 1: Samples taken from the PASCAL 1 test set; here each column corresponds to a different category; namely, airplane, boat, cat, motorbike and sheep, respectively

based on the entropy notion. In a recent paper [33], the authors evaluate the uncertainty via random walks on a graph, and Shannon Entropy is used to measure the uncertainty of random variables in this random process. We compared our method with the most frequently used uncertainty sampling technique (based on entropy criterion) in this paper. Another solution called the DBALStream method [14] is based on uncertainty sampling as well for active learning over evolving stream data. This approach has to decide whether to label an instance or not, but the method works in only the kind of scenario where each instance in the stream is processed as soon as it arrives. In the uncertainty sampling topic, one of the most recent paper [28] attempts to distinguish between the two types of uncertainties (conflicting-evidence vs. insufficient-evidence), but it does not provide another alternative approach for improving uncertainty sampling. One possible drawback of the uncertainty sampling algorithms is that they ignore the output distributions for the remaining class labels [12]; however our solution attempts to overcome this problem by a proposed balancing extension.

### 3 Image Classification

In the case of active learning, we iteratively query new images to expand the training set, and build a new classification model every time the training set changes. The decision of which image is next in line to be queried depends on the current model,

therefore the model creation process is an important part of the classification and the quality of the classification is an important part of the active learning. Following the general trend, we applied the BoW (Bag-of-Words) model [11, 16, 18] for the mathematical representation of the images and SVMs (Support Vector Machines) [1, 8, 16] for the classification.

The key idea behind the BoW model is to represent an image (based on its visual content) with so-called visual code words while ignoring their spatial distribution. This technique consists of three steps, these being (i) feature detection, (ii) feature description and (iii) image description as usual phases in computer vision. For feature detection we utilized the Harris-Laplace corner detector [4, 22], and SIFT (Scale Invariant Feature Transform) [19] to describe them. We should add that we used the default parameterization of SIFT proposed by Lowe; hence we got descriptor vectors with 128 dimensions. To define the visual code words from the descriptor vectors, we used the GMM (Gaussian Mixture Model) [25, 30], which is a parametric probability density function represented as a weighted sum of (in our case 256) Gaussian component densities. As can be seen below,

$$p(X | \lambda) = \sum_{j=1}^K \omega_j g(X | \mu_j, \sigma_j), \quad (1)$$

where  $\mu_j$  and  $\sigma_j$  denote the expected value and the variance of the  $j^{\text{th}}$  Gaussian component, respectively, and here  $K = 256$ . We calculated the  $\lambda$  parameter with the ML (Maximum Likelihood) estimation using the iterative EM (Expectation Maximization) algorithm [9, 30]. We performed K-means clustering [21] over all the descriptors with 256 clusters to get the initial parameter model for the EM. Next, we had to create a descriptor that specifies the distribution of the visual code words in an image called the high-level descriptor. To represent an image with a high-level descriptor, the GMM-based Fisher vector was calculated, as can be seen in Equation 2. These vectors were the final representations (image descriptor) of the images.

$$F = \nabla_{\lambda} \log p(X | \lambda). \quad (2)$$

For the classification subtask, we used a variation of SVM called the C-SVC (C-support vector classification) [1, 8] with a RBF (Radial Basis Function) kernel. Furthermore we applied the one-against-all approach to extend the SVM to the multi-class classification case.

## 4 Proposed Approach

Uncertainty sampling is the most common active learning query strategy framework, which queries the instances with least certainty about their labels. There are numerous ways to measure the amount of uncertainty, the easiest one being to query the instance whose prediction is the least confident. The problem with this approach is that it just processes information about the most probable class,

and throws away information about the remaining label distribution. On the other hand, a different uncertainty sampling variant called margin-sampling takes the first and second most probable class labels, but this still ignores many information. More general and probably the most popular uncertainty sampling strategies use entropy as an uncertainty measure, and query the instance with the maximum one.

The basis of our approach is this variation of uncertainty sampling. As we mentioned previously, we used SVM for classification. According to the literature, querying the closest instance to the linear decision boundary is analogous to uncertainty sampling with a probabilistic binary linear classifier [3, 13, 27]. SVM is basically not a probabilistic classifier so we applied a variation of Platt's [15] approach as a probability estimator to get the probability values (confidence values) of the possible labels. This approach is included in LIBSVM [5, 29], and we calculated the entropy based on it by using

$$H_j = - \sum_{i=1}^m P(l_i | t_j) \times \log P(l_i | t_j), \quad (3)$$

where  $H_j$  denotes the entropy of the  $j^{th}$  unlabeled image,  $m$  is the number of categories and  $P(l_i | t_j)$  denotes the confidence value that  $l_i$  is the label of image  $t_j$ . Furthermore, we defined a novel so-called penalty metric as

$$Penalty_j = CTR_{j^*} \times \frac{1}{m}, \quad (4)$$

where  $j^*$  denotes the estimated category of the  $j^{th}$  unlabeled image, and  $CTR_{j^*}$  denotes a general counter whose value increases by 1 with each iteration of queries when the received category is other than  $j^*$ . Thus each unlabeled image possess a penalty value depending on its predicted category. Here, we normalized the  $CTR$  values by the number of classes ( $m$ ), hence the penalty metric also depends on the number of classes. When the learning system queries a particular category (i.e. an image from that category), the penalty value of each unlabeled image we predict to be in that category is reset to zero. Within each cycle, we merge the actual penalty values with the ones coming from uncertainty sampling, and these results will give the final decision scores. The fusion was made based on a  $\beta$  weighting coefficient in the following formula:

$$HP_j = (1 - \beta) \times H_j + \beta \times Penalty_j, \quad (5)$$

where  $HP_j$  denotes the informativeness of the  $j^{th}$  unlabeled image. The higher the  $HP$  score of an image is, the more likely it will query its label. Therefore, we sorted the images by their  $HP$  values in descending order and chose the first (few) candidate images for labeling. In order to keep the same value of  $\beta$  for different data sets or tasks (i.e. the different number of classes), we decided to use the kind of normalization we described in Equation 4. The advantage of our proposed approach is that it strikes a balance among the classes of labeled instances, especially when the number of classes is very high, perhaps a hundred or more (see Section 5). In the following, we present our balanced active learning (BAL) algorithm in a pseudo

code (see Algorithm 1); note that in the third line, the "terminating condition is false" statement should be replaced with the desired one (which in our case was  $size(I_L) \neq size(I_U)$ ).

---

**Algorithm 1** Balanced Active Learning Algorithm (BAL)
 

---

**input:** Unlabeled image set  $I_U$   
**initialize:** Create the initial labeled image set  $I_L$ .  $\forall C \in I_U$  : select an image  
**while** (terminating condition is false) **do**  
   Train SVM model  $M$  based on the actual  $I_L$   
   **for**  $\forall$  image  $j \in I_U$  **do**  
     Calculate class membership probabilities ( $P(C | j)$ ) using  $M$   
     Calculate the entropy ( $H_j$ ) using Equation 3  
     Calculate the penalty metric ( $Penalty_j$ ) using Equation 4  
     Merge the entropy with the penalty metric ( $HP_j$ ) using Equation 5  
   **end for**  
   **sort:**  $\forall$  image  $j \in I_U$  by  $HP_j$  in descending order  
   **query:** image  $j_q$  from top of the list:  $j_q = query(\argmax_{j \in I_U} HP_j)$   
   **insert:**  $j_q$  and its real label  $l$  into  $I_L$ :  $I_L = I_L \cup \{j_q, l\}$   
   **remove:**  $j_q$  and its real label  $l$  from  $I_U$ :  $I_U = I_U \setminus \{j_q, l\}$   
   **reset:** penalty metrics based on the learned real label:  $\forall j \in C_l : Penalty_j = 0$   
**end while**

---

## 5 Experimental Results

### 5.1 Experimental Environment

Next, we will present the experimental results of our proposed balanced active learning query strategy. We evaluated three strategies for comparison. These were

- Random sampling
- Uncertainty sampling using entropy as measure
- The proposed balanced active learning (BAL) approach. That is, uncertainty sampling using an entropy metric merged with a novel penalty metric calculated from the distribution of recently queried class labels (see Algorithm 1)

We used two large and popular image classification data sets in our experiments, namely the PASCAL VOC2010 [10] training and validation datasets, and the Caltech101 [17] image collection. The PASCAL VOC2010 consists of 20 classes and it has 10,103 images in total. We randomly selected 50 images from each category to form a subset of this data set. The Caltech101 collection contains 8677 images taken from 101 different (real) categories and an additional noise category. We

discarded the noise category to simplify the data set and randomly chose 30 images from each class to create our second subset. Afterwards, we randomly selected 10 categories from the PASCAL and 20 categories from the Caltech101 subsets to form two additional subsets. The following table presents the details concerning our chosen data sets; and in Figure 1 we can see some sample images from the PASCAL 1 test set.

Table 1: Details of the randomly selected subsets and imbalanced data sets

	Number of classes	Number of runs	Number of images
CALTECH101 1	20	10	600
CALTECH101 2	101	10	3030
CALTECH101 3	20	1	1961
CALTECH101 4	101	1	8677
PASCAL 1	10	10	500
PASCAL 2	20	10	1000

The distribution of images among the classes were even in the created subsets, but our approach is also capable of handling imbalanced data sets. Therefore we included two more test sets (CALTECH101 3 and CALTECH101 4) those consisted of the same categories as CALTECH101 1 and CALTECH101 2, but the number of images were not reduced (i.e. CALTECH101 4 was the whole Caltech101 dataset).

During the experiments, each queried image and its label was moved to the training set from the test set. For the initial training set, we randomly selected an image from each category. In the last iteration, the size of the training set and test set were the same; e.g. in the case of PASCAL 1 data set, the learning system stopped querying the unlabeled instances when the number of training images reached 250. As can be seen in the third column of Table 1, we performed the experiments several times with the same data set, and we averaged out the results of the separate runs.

## 5.2 Results

Now we will turn to the results. Within each cycle, we evaluated the accuracy and MAP (Mean Average Precision) [26] metrics on the actual test set. After a series of tests, we chose  $\beta$  to be 0.1 (see Equation 5); because this gave a sufficient weight for the penalty metric to positively influence the entropy. It should be added that the proposed approach is highly sensitive to the choice of  $\beta$  (see Figure 5 for details). The overall results of the generated subsets of the Caltech101 collection and the PASCAL data set can be seen in figures 2 and 3, respectively; and the results of the unbalanced test sets can be seen in Figure 4. As we mentioned previously, each data set was tested several times to be able to take the average of the separate runs and thus give a more trustworthy result. This was important because we commenced each of the tests with different initial training sets, since the first images were

randomly selected. In the graphs we can see the averages, and for our proposed approach, we also included error bars to show the minimum and maximum values at each sampling point. This presents the variance of the BAL algorithm, as can be seen in the following figures.

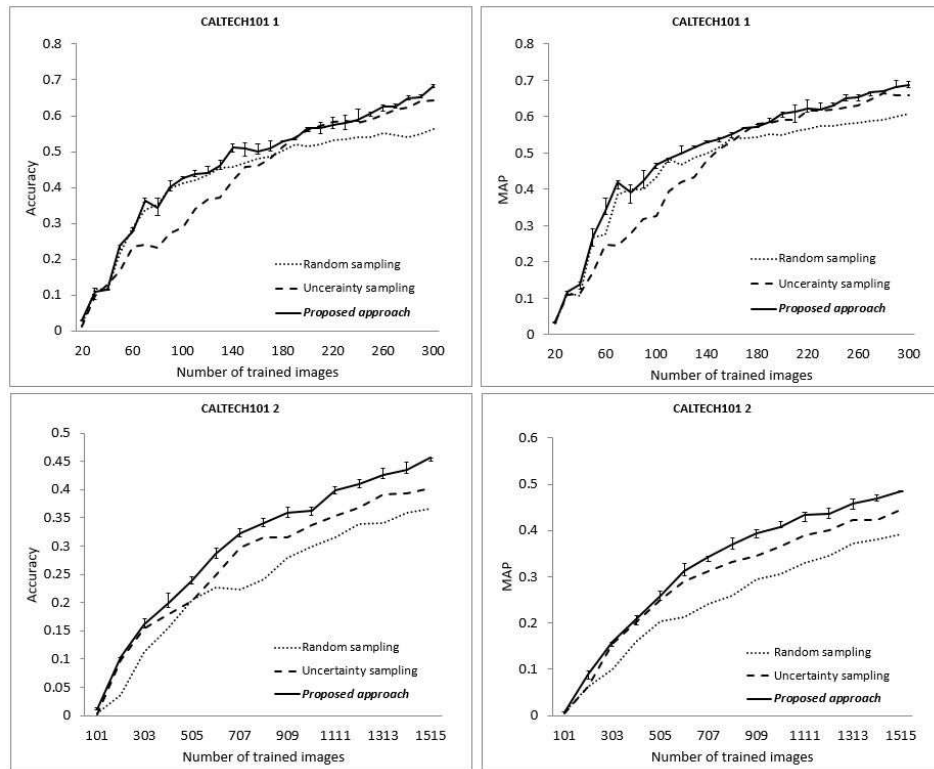


Figure 2: Evaluation of the results obtained on the CALTECH101 1 and CALTECH101 2 datasets. The accuracy and MAP metrics are on the y-axis, while the number of trained images is on the x-axis. The proposed approach, uncertainty sampling and random sampling are represented by solid, dashed and dotted lines, respectively.

As the results of the CALTECH101 1 test set in Figure 2 show, the distribution analysis has a beneficial effect. The basic uncertainty sampling with entropy (represented by a dashed line) yielded worse results at the first 150 iterations. This was because it queried many instances from the same category, but this could not happen with the proposed approach, since it balances the label distribution. At the last sampling point, our approach gave better results than the other competitor methods. In Figure 2, the results of the CALTECH101 2 test set are also presented. As can be seen, the balanced active learning algorithm consistently outperforms the



other two approaches. Based on this, we may conclude that the balancing technique is more advantageous when the number of classes is relatively high.

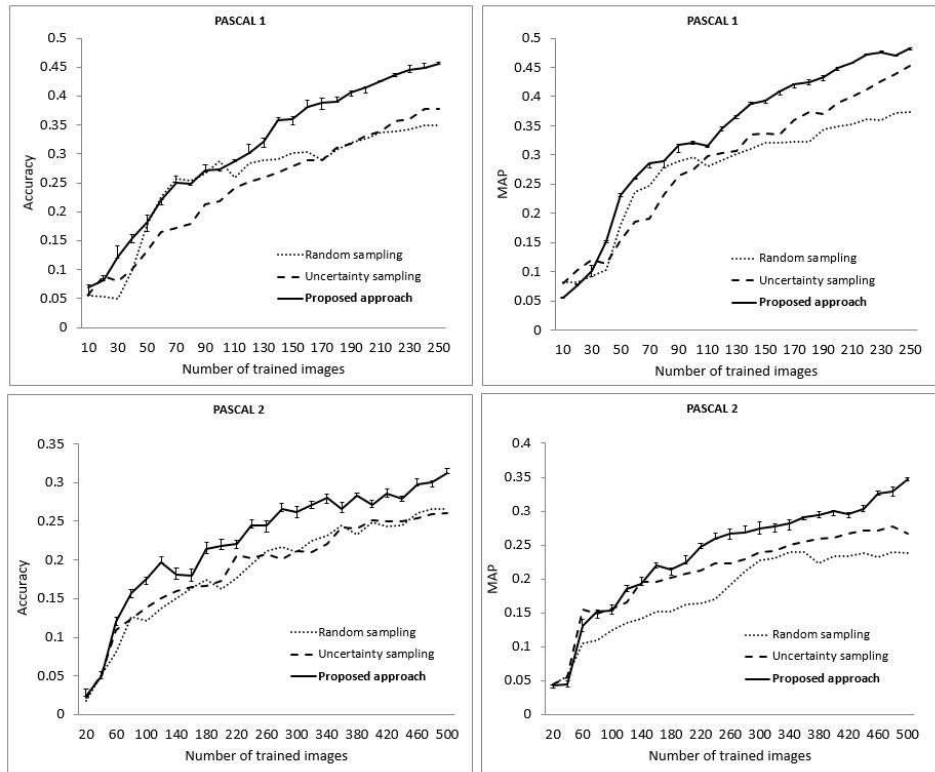


Figure 3: Evaluation of the results obtained on the PASCAL 1 and PASCAL 2 datasets. The accuracy and MAP metrics are on the y-axis, while the number of trained images is on the x-axis. The proposed approach, uncertainty sampling and random sampling are represented by solid, dashed and dotted lines, respectively.

The results of the PASCAL 1 and PASCAL 2 test sets are shown in Figure 3; in the graphs we can discern some similarities in the pattern of the curves to those in Figure 2. For example, the proposed approach gave the highest metrics among the three methods for almost each sampling point for both test sets; although if we look at the accuracy values of PASCAL 2, we see that the difference in the results is higher and less volatile between our BAL approach and the other two methods. This observation seems to support our hypothesis about the existence of a proportional relationship between the performance of BAL algorithm and the number of categories.

The results of CALTECH101 3 and CALTECH101 4 can be seen in Figure 4. In this case, the accuracy and MAP values were higher for the last sampling point,

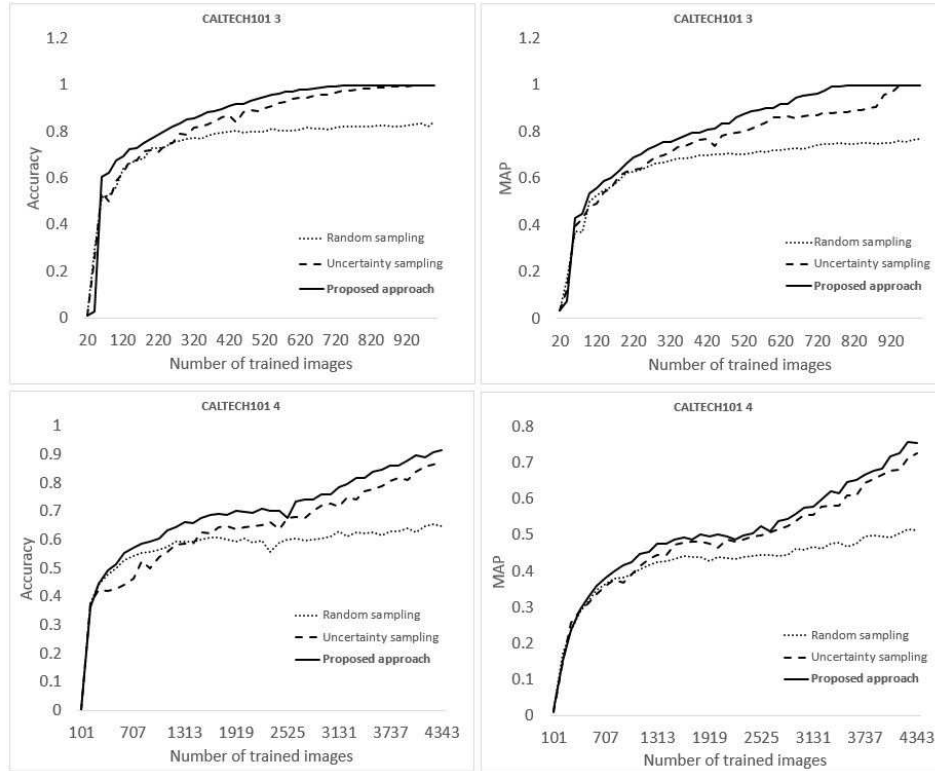


Figure 4: Evaluation of the results got on the CALTECH101 3 and CALTECH101 4 datasets. The accuracy and MAP metrics are on the y-axis, while the number of trained images is on the x-axis. The proposed approach, uncertainty sampling and random sampling are represented by solid, dashed and dotted lines, respectively.

because the learning system had more training images. Although this difference decreased, the proposed approach outperforms the other two competitor methods tested here almost at every sampling point (except the first few at CALTECH101 3). However, the BAL algorithm attains an accuracy of 1.0 and MAP of 1.0 faster (i.e. with fewer training images) than the general uncertainty sampling; therefore our proposed approach reduces the number of required training instances to achieve the maximum possible classification accuracy, which is the main goal of our study (and using active learning in general).

We summarized the result obtained of all our experiments in Table 2; the accuracy and MAP values shown in the table were measured for the last iteration. As can be seen, the proposed approach outperformed the other query strategies in every case here.

In the last experiment, we give a brief explanation of our choice of  $\beta$  (see

Table 2: Summary of the accuracy and MAP values got on the four test sets

	Rand. sampling		Unc. sampling		<b>BAL approach</b>	
	Acc.	MAP	Acc.	MAP	Acc.	MAP
CALTECH101 1	0.562	0.609	0.642	0.660	<b>0.683</b>	<b>0.688</b>
CALTECH101 2	0.366	0.390	0.401	0.445	<b>0.456</b>	<b>0.484</b>
CALTECH101 3	0.842	0.770	0.998	0.997	<b>1.000</b>	<b>1.000</b>
CALTECH101 4	0.648	0.514	0.876	0.725	<b>0.916</b>	<b>0.755</b>
PASCAL 1	0.350	0.374	0.378	0.453	<b>0.455</b>	<b>0.482</b>
PASCAL 2	0.266	0.238	0.260	0.266	<b>0.312</b>	<b>0.347</b>

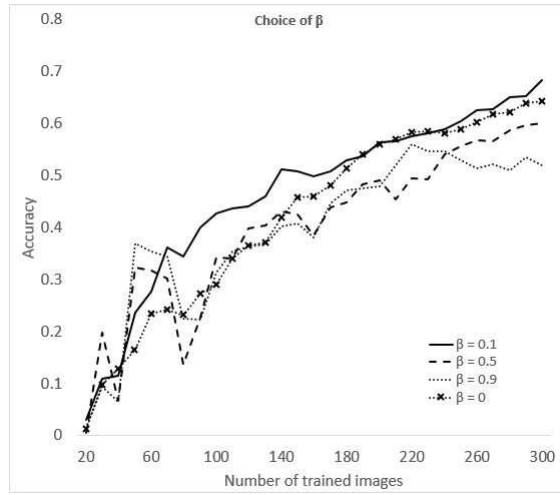


Figure 5: Sensitivity for the choice of  $\beta$  on the CALTECH101 1 dataset. Accuracy metric is on the y-axis, while the number of trained images is on the x-axis. The cases  $\beta = 0.1$ ,  $\beta = 0.5$  and  $\beta = 0.9$  are represented by solid, dashed and dotted lines, respectively.

Figure 5). We used the CALTECH101 1 data set for this test, and we evaluated the results for the following  $\beta$  values: 0.1, 0.5, 0.9 and 0. Here,  $\beta = 0$  is a special case, since it returns the general uncertainty sampling with an entropy measure (i.e. the dashed line in the upper left sub-figure of Figure 2). This figure tells us that the proposed approach is highly sensitive to the choice of  $\beta$ , because with the wrong values it is possible to achieve a worse performance than that using the simple entropy measure.

## 6 Conclusions

In this paper we presented a new query strategy for active learning, which is an improvement of the basic uncertainty sampling technique. The proposed approach combines the entropy measure with a novel penalty metric to balance the class distribution of labeled instances. This modification solves a possible deficiency of the uncertainty sampling, the skipping of certain categories. We employed two large data sets (PASCAL VOC2010 and Caltech101) to demonstrate the efficiency of our approach, and we evaluated the accuracy and MAP metrics. Our experiments demonstrated that the proposed balanced active learning (BAL) approach generally outperforms the random sampling and the basic uncertainty sampling methods.

## References

- [1] Boser, B., Guyon, I. and Vapnik, V. A Training Algorithm for Optimal Margin Classifier. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144-152, 1992.
- [2] Burr Settles *Active learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning, Vol. 6, No. 1, pp. 1-114, 2012.
- [3] C. Campbell, N. Cristianini and A. Smola Query learning with large margin classifiers. *Proceedings of the 17th International Conference on Machine Learning*, pp. 111-118, 2000.
- [4] C. Harris, M. Stephens A combined corner and edge detector. *Proceedings of the Alvey Vision Conference*, pp. 23.1-23.6, 1988.
- [5] C.-C. Chang, C.-J. Lin LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 27.1-27.27, 2011.
- [6] Cai, W., Zhang, Y. and Zhou, J. Maximizing expected model change for active learning in regression. *IEEE 13th International Conference on Data Mining*, pp. 51-60, 2013.
- [7] Chaudhuri, K., Kakade, S. M., Netrapalli, P. and Sanghavi, S. Convergence rates of active learning for maximum likelihood estimation. *Advances in Neural Information Processing Systems 28*, pp. 1090-1098, 2015.
- [8] Cortes, C., Vapnik, V. Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273-297, 1995.
- [9] Dempster, A., Laird, N. and Rubin, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Vol. 39, No. 1, pp. 1-38, 1977.

- [10] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 303-338, 2010.
- [11] Fei-Fei, L., Fergus, R. and Torralba, A. Recognizing and Learning Object Categories. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [12] Fu, Y., Zhu, X. and Li, B. A survey on instance selection for active learning. *Knowledge and Information Systems*, Vol. 35, No. 2, pp. 249-283, 2013.
- [13] G. Schohn, D. Cohn Less is more: Active learning with support vector machines. *Proceedings of the 17th International Conference on Machine Learning*, pp. 839-836, 2000.
- [14] Ienco, D., Zliobaite, I. and Pfahringer, B. High density-focused uncertainty sampling for active learning over evolving stream data. *Proceedings of the 3rd International Workshop on Big Data*, pp. 133-148, 2014.
- [15] J. Platt Probabilistic outputs for support vector machines and comparison to regularize likelihood methods. *Advances in Large Margin Classifiers*, pp. 61-74, 2000.
- [16] K. Chatfield, V. Lempitsky, A. Vedaldi and A. Zisserman The devil is in the details: an evaluation of recent feature encoding methods. *Proceedings of the 22nd British Machine Vision Conference*, pp. 76.1-76.12, 2011.
- [17] L. Fei-Fei, R. Fergus and P. Perona Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Workshop on Generative-Model Based Vision, 2004.
- [18] Lazebnik, S., Schmid, C. and Ponce, J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 2169-2178, 2006.
- [19] Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [20] Mac Aodha, O., Campbell, N., Kautz, J. and Brostow, G. Hierarchical subquery evaluation for active learning on a graph. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 564-571, 2014.
- [21] MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-297, 1967.
- [22] Mikolajczyk, K., Schmid, C. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, Vol. 60, No. 1, pp. 63-86, 2004.

- [23] Minakawa, M., Raytchev, B., Tamaki, T. and Kaneda, K. Image sequence recognition with active learning using uncertainty sampling. *Proceedings of the International Joint Conference on Neural Networks*, pp. 1-6, 2013.
- [24] Pasolli, E., Melgani, F., Tuia, D., Pacifici, F. and Emery, W. J. SVM active learning approach for image classification using spatial information. *Proceedings of the IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, No. 4, pp. 2217-2233, 2014.
- [25] Reynolds D. A. *Gaussian Mixture Models*. Encyclopedia of Biometric Recognition, pp. 659-663, 2009.
- [26] Robertson, S. A new interpretation of average precision. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 689-690, 2008.
- [27] S. Tong, D. Koller Support vector machine active learning with applications to text classification. *Proceedings of the 17th International Conference on Machine Learning*, pp. 9991006, 2000.
- [28] Sharma, M., Bilgic, M. *Evidence-based uncertainty sampling for active learning*. Data Mining and Knowledge Discovery, pp. 1-39, 2016.
- [29] T.-K. Huang, R. C. Weng and C.-J. Lin Generalized Bradley-Terry models and multi-class probability estimates. *Journal of Machine Learning Research*, Vol. 7, pp. 85-115, 2006.
- [30] Tomasi C. Estimating Gaussian mixture densities with EM - A Tutorial. Technical report, Duke University, 2004.
- [31] Tsai, Y. L., Tsai, R. T. H., Chueh, C. H. and Chang, S. C. Cross-Domain Opinion Word Identification with Query-By-Committee Active Learning. *Proceedings of the 19th International Conference on Technologies and Applications of Artificial Intelligence*, pp. 334-343, 2014.
- [32] Vijayanarasimhan, S., Grauman, K. Large-scale live active learning: Training object detectors with crawled data and crowds. *International Journal of Computer Vision*, Vol. 108, No. 1-2, pp. 97-114, 2014.
- [33] Yang, Y., Ma, Z., Nie, F., Chang, X. and Hauptmann, A. G. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, Vol. 113, No. 2, pp. 113-127, 2015.