

## Research Article

# A Deep Multiview Active Learning for Large-Scale Image Classification

Tuozhong Yao,<sup>1</sup> Wenfeng Wang ,<sup>1,2</sup> and Yuhong Gu<sup>3</sup>

<sup>1</sup>School of Electronic and Information Engineering, Ningbo University of Technology, Ningbo 315211, China

<sup>2</sup>School of Electronic and Electrical Engineering, Shanghai Institute of Technology, Shanghai 200235, China

<sup>3</sup>Shihezi Medical School, Shihezi 832000, China

Correspondence should be addressed to Wenfeng Wang; wangwenfeng@sit.edu.cn

Received 26 October 2020; Revised 12 November 2020; Accepted 30 November 2020; Published 15 December 2020

Academic Editor: Xingling Shao

Copyright © 2020 Tuozhong Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiview active learning (MAL) is a technique which can achieve a large decrease in the size of the version space than traditional active learning and has great potential applications in large-scale data analysis. In this paper, we present a new deep multiview active learning (DMAL) framework which is the first to combine multiview active learning and deep learning for annotation effort reduction. In this framework, our approach advances the existing active learning methods in two aspects. First, we incorporate two different deep convolutional neural networks into active learning which uses multiview complementary information to improve the feature learnings. Second, through the properly designed framework, the feature representation and the classifier can be simultaneously updated with progressively annotated informative samples. The experiments with two challenging image datasets demonstrate that our proposed DMAL algorithm can achieve promising results than several state-of-the-art active learning algorithms.

## 1. Introduction

Active learning was firstly proposed by Simon and Lea [1]. It ranks the unlabeled samples iteratively and only selects the samples with high uncertainty or which cause great ambiguity for the classifier. In PAC learning theory, compared with traditional passive learning, it can exponentially reduce its sample complexity to  $O(\log(1/\epsilon))$  in the feature space for learning a classifier with expectation classification error  $\epsilon$ . In the past twenty years, active learning has been developed rapidly and applied in many computer vision-related fields such as image classification [2], object segmentation [3], face recognition [4], object tracking [5], scene reconstruction [6], and activity recognition [7].

The key problem of active learning is selective sampling, and there are mainly two different solutions: pool-based sampling and stream-based one. The stream-based methods usually need manual thresholding for unlabeled sample evaluation and can't compare the input samples one by one which restricts its further application. There are also two

different pool-based methods based on the number of generated hypothesis. (a) Single-hypothesis-based methods only have one base classifier which uses uncertainty reduction [8] or expectation error minimization [9] as the sampling standard, and they usually perform better than random sampling. (b) Committee-based methods use two or more hypotheses to set up a committee, and the sample which reduces the size of the version space maximally is queried by this committee. In most cases, these methods such as Query by Committee [10], Query by Bagging [11], and Query by Boosting [12] can achieve a larger decrease in the size of the version space than single-hypothesis-based ones. However, in the committee-based methods, the generated hypotheses usually lack of diversity which affects their performance. Another special kind of committee-based methods is MAL such as Co-Testing [13] which uses the concept of redundant views from Co-Training [14]. They combine multiple conditionally independent hypotheses (view) to maximally improve the query performance of one of the views and query the sample which leads to the

classification error of at least one view because a classifier usually improves itself from the error. Theoretically, MAL can achieve a decrease in more than a half size of the version space which is more efficient than both single-hypothesis-based and committee-based active learning.

## 2. Related Work

The first MAL algorithm Co-Testing was proposed by Muslea in 2000 [13]. In the past fifteen years, MAL has been developed quickly. Later, they proposed the Co-EMT algorithm [15]. It used Co-Testing to query the unlabeled samples and multiview learning to improve the hypotheses, respectively, for text classification. Furthermore, Muslea et al. also presented another MAL algorithm called Aggressive Co-Testing and showed how to use the weak view to improve both the selective sampling and hypothesis generation [16]. Cheng and Wang proposed the Co-SVM algorithm which can be converged faster than traditional single-view active learning (SAL) through image retrieval experiments [17]. Wang and Zhou designed a semi-supervised-based MVAL learning algorithm and also proved that its sample complexity is reduced exponentially in nonrealized conditions while SAL can only achieve polynomial reduction at most in sample complexity [18] and also discussed MAL in the nonrealized case [19]. Di and Crawford presented a new coregularization-based active learning framework which combined both the consistency between views and local proximity in hypothesis to optimize the selective sampling [20]. Cai et al. also proposed a multiview active learning approach to fully exploit the visual view of videos, while querying as few annotations as possible from the text view for reducing the annotation cost [21]. MAL has been widely studied and applied to many applications.

Recently, incredible progress on visual recognition tasks has been made by deep learning (DL) approaches. With sufficient labeled data, deep convolutional neural networks are trained to directly learn features from raw pixels which have achieved the state-of-the-art performance for many visual-based applications such as image classification. However, in many real applications of large-scale image classification, the labeled data are far too enough since the tedious manual labeling process requires a lot of time and labor. Thus, it has great practical significance to develop a framework by combining DL and active learning. Some amazing works have revealed that such combination can achieve significant performance improvement [22–26]. Due to these successes, a new combination of MAL and DL is worth for comprehensive research.

## 3. DMAL Algorithm

To the best of our knowledge, in this paper, we are the first to incorporate MAL and DL into a unified framework for large-scale image classification. Our objective is to apply active learning to deep image classification tasks by progressively selecting complementary samples for model updating. The flowchart of our DMAL framework is illustrated in Figure 1.

Suppose we have a dataset of  $M$  categories and  $N$  samples denoted as  $D = \{x_i\}_{i=1}^N$ . We denote the currently annotated samples of  $D$  as  $D^L$  while the unlabeled ones as  $D^U$ . The label of  $x_i$  is denoted as  $y_i = j$ ,  $j \in \{1, \dots, M\}$ . In our dataset  $D$ , almost all data are unlabeled and  $D^U$  should be inputted into the DMAL framework in an incremental way.

The DMAL for image classification is formulated as follows:

$$\min_{\{W, y_i, i \in D^U\}} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M 1\{y_j = j\} \log p(y_i = j | x_i; W), \quad (1)$$

where  $1\{\cdot\}$  is the indicator function so that  $1\{\text{true}\} = 1$  and  $1\{\text{false}\} = 0$ .  $W$  denotes the network parameters of the CNN. In DMAL, we utilize two different CNNs for learning image features as different views. Thus,  $W \in W_1 \cup W_2$  where  $W_1$  and  $W_2$  denote the parameters of two CNNs, respectively. Inspired from Lin et al.'s work in deep multiview learning for clustering analysis [27], in this paper,  $p(y_i = j | x_i; W)$  denotes the mean of the softmax output of the two CNNs for the  $j$ th category which represents the probability of the sample  $x_i$  belonging to the  $j$ th class and is formulated as follows:

$$p(y_i = j | x_i; W) = \sum_{k=1}^2 \alpha_k \cdot p(y_i = j | x_i; W_k); \quad \sum_{k=1}^2 \alpha_k = 1, \quad (2)$$

where  $\alpha_k$  is the normalized importance weight which measures the importance of outputs of multiple views. Equation (2) suggests the multiview fusion may more accurately obtain the real class of  $x_i$ .

Specifically, the algorithm is designed by alternatively updating the labeled sample  $y_i \in D^L$  and the network parameters  $W$ . In the following, we introduce the details of the optimization steps and give their physical interpretations. The practical implementation of the DMAL will also be discussed in the end.

Algorithm 1 can be summarized as follows.

**3.1. Initialization.** In DMAL, two CNNs are utilized for generating different hypotheses. The labeled sample  $D^L$  is empty at the beginning. For each category, we randomly select a few training samples from  $D^U$  and manually annotate them for initializing the parameters  $W_k$ ,  $k = 1, 2$  of the two CNNs.

**3.2. Sample Selection.** By fixing the CNNs' parameters  $W_k$ ,  $k = 1, 2$ , we first rank all unlabeled samples in different views according to the common active learning criteria and then manually annotate those most uncertain samples and add them into  $D^L$ . Our DMAL selects  $K$  most informative/uncertain samples left in  $D^U$  by ranking all the unlabeled samples in a descending order according to their Entropy  $E_i$  [22] which is defined as follows:

$$E_i = - \sum_{j=1}^M p(y_j = j | x_i; W) \log p(y_j = j | x_i; W). \quad (3)$$

This measure takes all class label probabilities into consideration to measure the uncertainty. The higher the entropy value is, the more uncertain the sample is.

**3.3. CNN Parameter Update.** Most of active learning methods only focus on classifier retraining. Their strategies to select the most informative/uncertain samples are heavily dependent on the assumption that the feature representation is fixed. However, the feature learning and classifier training are jointly optimized in CNNs, and simply fine-tuning CNNs in the traditional active learning framework probably face the divergence problem.

To solve this problem, after sample selection, we employ the standard back propagation to update the CNNs' parameters  $W_k$ ,  $k = 1, 2$  mainly based on  $D^L$ . Specifically, let  $L$  denote the loss function of equation (4), then the partial derive of the network parameter  $W$  according to equation (4) is as follows:

$$\begin{aligned} \frac{\partial L}{\partial W} &= \frac{\partial \left[ -(1/C) \sum_{i=1}^C \sum_{j=1}^M 1\{y_j = j\} \log p(y_i = j | x_i; W) \right]}{\partial W} \\ &= \frac{1}{C} \sum_{i=1}^N \sum_{j=1}^M (1\{y_i = j\} - p(y_i = j | x_i; W)) \frac{\partial z_j(x_i; W)}{\partial W}, \end{aligned} \quad (4)$$

where  $\{z_j(x_i; W)\}_{j=1}^M$  denotes the activation for the  $i$ th sample of the last layer of the CNN model before feeding into the softmax classifier which is defined as follows:

$$p(y_i = j | x_i; W) = \frac{e^{z_j(x_i; W)}}{\sum_{l=1}^M e^{z_l(x_i; W)}}. \quad (5)$$

After CNN fine-tuning, we go to the next iteration until the maximum iteration  $T$  is reached.

## 4. Experiments

**4.1. Datasets and Experiment Settings.** In our experiment, we use two large-scale image datasets: CIFAR-10 [28] and ImageNet ILSVRC [29] datasets for testing our algorithm. The CIFAR-10 dataset consists of 60000  $32 \times 32$  color images in 10 different classes, with 6000 images per class. In the ImageNet ILSVRC dataset, there are more than 1.2 million training images and 100 thousand testing images in 1000 different classes. We use different CNNs for hypotheses generation in DMAL. In this paper, Alexnet [30] and VGGnet [31] are both utilized for multiview feature learning. For CIFAR-10, we resize all the images into  $200 \times 150$  and set the learning rates of all the layers as 0.01. For ImageNet ILSVRC, we resize all the images into  $256 \times 256$  and set the learning rates of all the layers as 0.001. We set the normalized importance weight  $\alpha_1 = \alpha_2 = 0.5$ .

**4.2. Algorithm Comparison.** We compare our DMAL algorithm with the following seven state-of-the-art active learning algorithms:

- (a) SVM\_RAN: During the training process, we randomly select samples to be annotated and use support vector machine (SVM) as the base classifier. This method discards all active learning techniques and can be considered as the lower bound.
- (b) SVM\_QP [32]: A single-view SVM active learning by QP-based selective sampling which is dependent on the sample uncertainty and redundancy.
- (c) MAL [16]: Co-Testing multiview active learning with conservative sampling strategy.
- (d) AL\_DIS [33]: A disagreement-based active learning from weak and strong labelers.
- (e) AL\_TC [34]: Triple criteria-based active learning is designed to jointly evaluate sample selection criteria including uncertainty, diversity, and density.
- (f) AL\_DL [26]: An active learning algorithm based on weighted incremental dictionary learning.
- (g) DAL\_ALEX: The single-view version of the deep active learning algorithm with only Alexnet for feature learning.

In SAL (a), (b), (d), and (e), Color SIFT and HOG descriptors are both densely sampled for feature representation. In MAL (c), Color SIFT and HOG descriptors are sampled for different hypotheses generations, respectively. We evaluate these algorithms by varying the following two parameters: (1) label size ( $D^L$ ): the number of initially sampled training set and (2) batch size ( $K$ ): the number of queried informative/uncertain samples in each iteration.

**4.3. Experiment 1: Fixed Label Size and Batch Size.** In Experiment 1, we set label size = 5000, batch size = 500, and iteration ( $T$ ) = 15. Figure 2 shows the comparison of different active learning algorithms by average precision (AP) in different relevance feedbacks. Here, we only show four intermediate results with iteration interval equal to 4 due to space restriction. A1–A6 denote the following six active learning algorithms: SVM\_RAN, SVM\_QP, MAL, AL\_DIS, AL\_TC, and AL\_DL which are drawn by dotted lines. A7 and A8 denote the following two deep active learning algorithms: DAL\_ALEX and DMAL which are drawn by solid lines.

In Figure 2, we can see that (1) all of seven algorithms achieve higher AP than SVM\_RAN which demonstrates that their sample selection strategies work and can query informative/uncertainty unlabeled samples, (2) both DAL\_ALEX and DMAL achieve higher AP than other six algorithms which suggest the deep CNN is more efficient in feature representation than traditional features, and (3) DAL\_ALEX achieves lower AP than DMAL after four iterations which demonstrates the DAL with complementary views can achieve better performance than DAL with a single view.

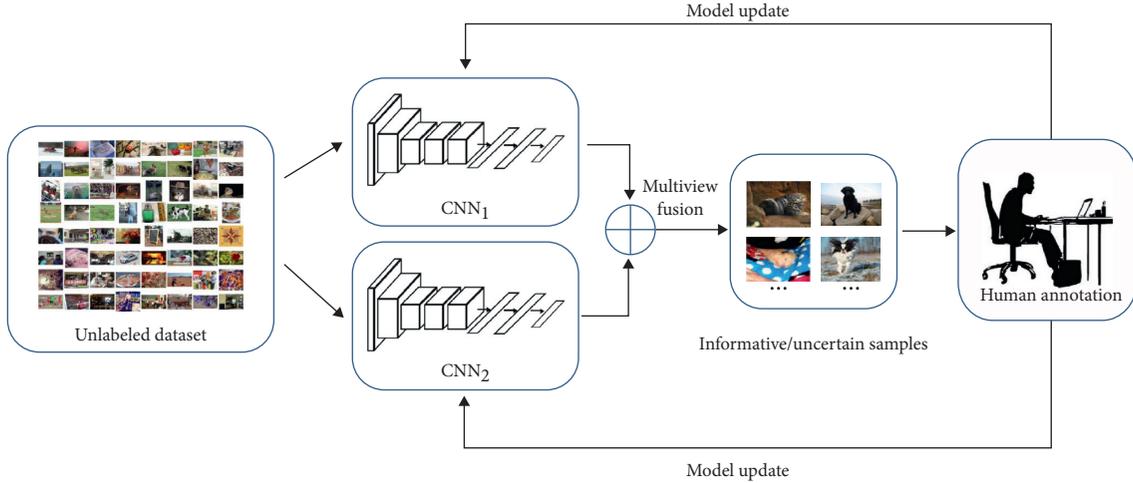


FIGURE 1: Illustration of our proposed DMAL framework.

**Input:**

Unlabeled samples  $D^U$ , initially labeled samples  $D^L$ , size of queried samples  $K$  in each iteration, maximum iteration number  $T$

**Output:**

CNN parameters  $W$

- (1) Initialize  $W$  with  $D^L$ ;
- (2) **while** not reach maximum iteration  $T$  **do**
- (3) utilize two CNNs to obtain  $p(y_i = j | x_i; W)$  based on equation (2);
- (4) Add uncertainty samples into  $D^L$  based on equation (3);
- (5) In iteration  $t$ :  
Update  $W$  via fine-tuning CNN according to equation (4) with updated  $D^L$ ;
- (6) **end while**
- (7) **return**  $W$

ALGORITHM 1: DMAL algorithm.

**4.4. Experiment 2: Variable Label Size.** In Experiment 2, we fix batch size = 500 and evaluate the performance of these algorithms by varying label size. We also use average precision (AP) and mean of average precision (MAP) as the measures.

Figure 3 shows AP variations of the first 200 feedback images based on different label size (label size = 1000, 2000, 3000, 4000, and 5000). We can see that both DAL\_ALEX and DMAL always achieve the highest AP compared with other algorithms. As label size increases, AP growth of both DAL\_ALEX and DMAL becomes slow which demonstrates that they have good convergence and also fit for small training sets.

Table 1 illustrates MAP comparisons of eight active learning algorithms with different label size for the first 500 feedback images. The value behind symbol  $\pm$  denotes the standard deviation of AP.

In Table 1, we can see that (1) compared with SVM\_RAN, other seven algorithms A2–A8 achieve higher MAP which demonstrates that their sampling strategies are more effective than random sampling, (2) MAP of DAL\_ALEX is 6.13% and 4.3% higher than that of AL\_DL in both datasets which demonstrates that the CNN is useful to improve the feature representation, and (3) MDAL

outperforms state-of-the-art algorithms AL\_DL, AL\_TC, and AL\_DIS in MAP which suggests the DL and MAL is the efficient combination.

**4.5. Experiment 3: Variable Batch Size.** In Experiment 3, we fix label size = 5000 and evaluate the performance of these algorithms by varying batch size. Figure 4 illustrates AP variations of the first 200 feedback images based on different label size (batch size = 100, 200, 300, 400, and 500).

Through analysis from Figure 4, we can make the following conclusion: our proposed algorithm MDAL outperforms all other algorithms in different batch size conditions. Furthermore, we can obtain an implicit regulation: As batch size increases, AP growth ratios of both DAL\_ALEX and MDAL are higher than algorithms such as AL\_TC and AL\_DL that do not use the CNN for feature representation. For example, in the CIFAR-10 dataset, when batch size increases from 200 to 300, AP growth ratios of MDAL, DAL\_ALEX, and AL\_DL are 7.52%, 7.83%, and 5.48%, respectively; when batch size increases from 300 to 400, AP growth ratios of the three algorithms are 3.25%, 3.17%, and 1.60%, respectively. In the ImageNet ILSVRC dataset, when batch size increases from 200 to 300, AP growth ratios of MDAL, DAL\_ALEX, and AL\_DL are 6.54% 6.78%,

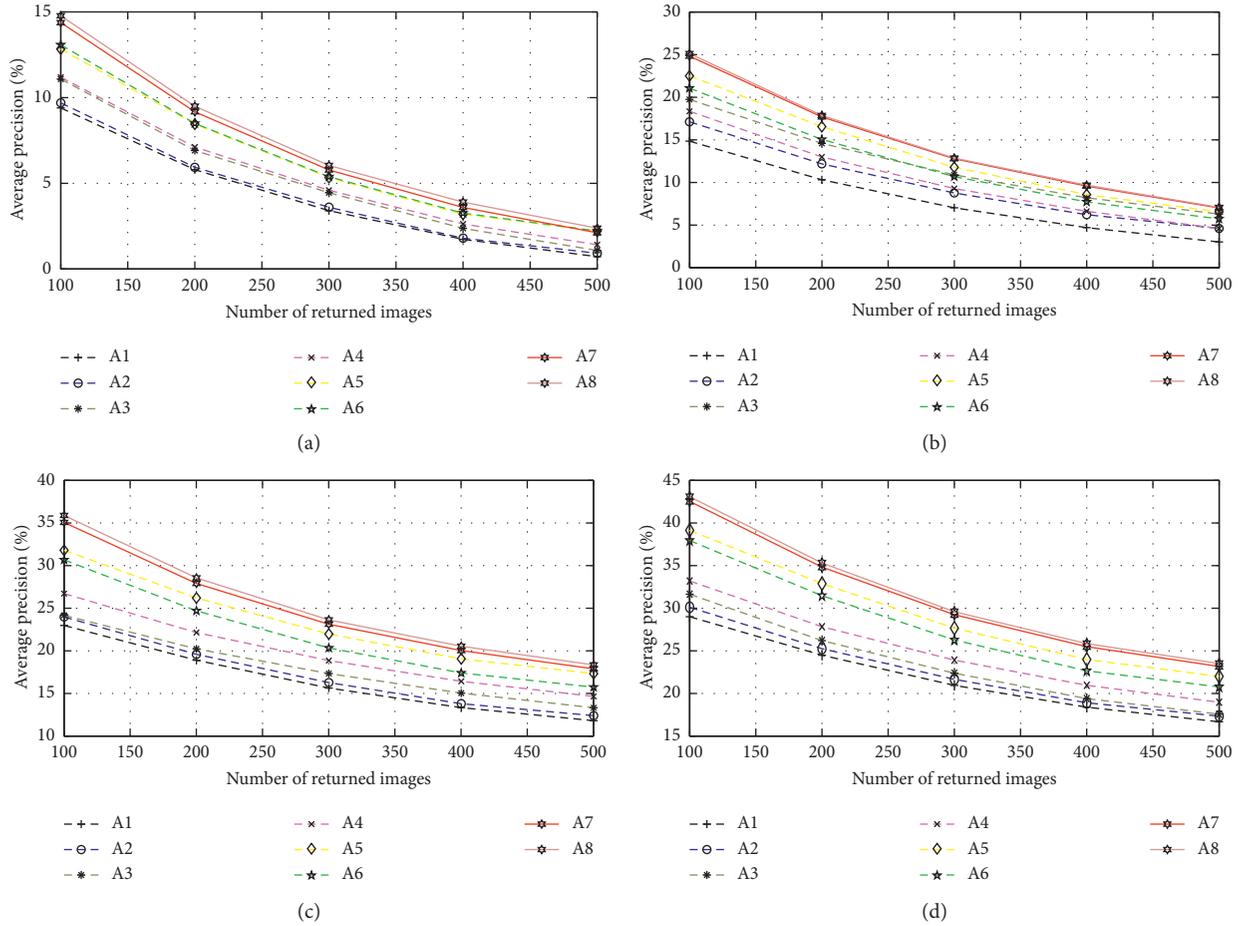


FIGURE 2: Comparison of eight different active learning algorithms in Iteration 2, 6, 10, and 14 (CIFAR-10): (a) iteration 4; (b) iteration 6; (c) iteration 8; (d) iteration 10.

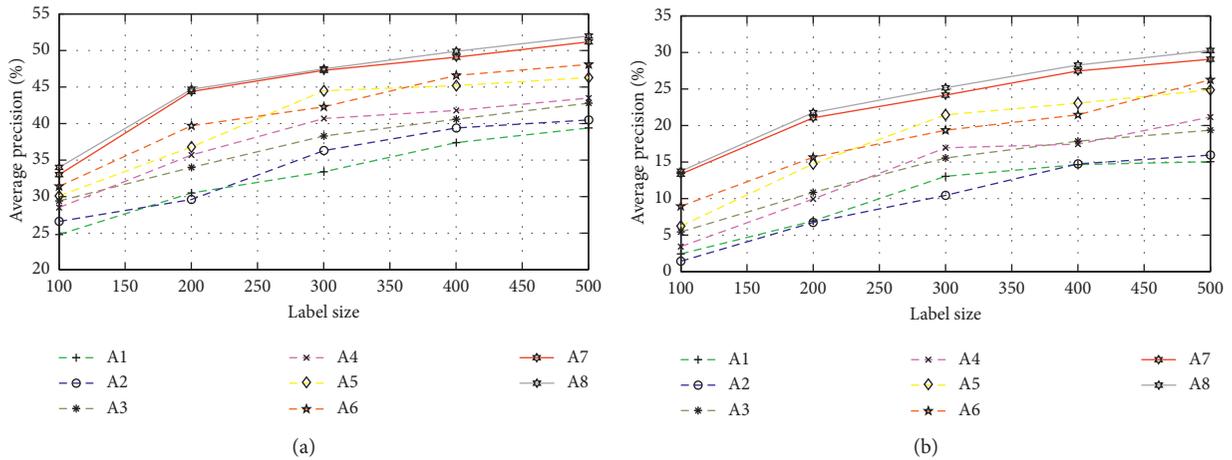


FIGURE 3: AP variations of the first 500 feedback images based on variable label size. (a) CIFAR-10. (b) ImageNet ILSVRC.

and 3.91%, respectively; when batch size increases from 300 to 400, AP growth ratios of the three algorithms are 5.45%, 5.29%, and 4.77%, respectively. These evidences demonstrate that deep learning-based and multiview-based active learning algorithms diverge faster than other algorithms which strongly suggests

that incorporating deep learning into the multiview active learning framework is an efficient way for active learning improvement.

Table 2 illustrates MAP comparison of all eight algorithms with different batch size for the first 500 feedback

TABLE 1: MAP comparisons of eight active learning algorithms with different label size.

MAP (%)	A1	A2	A3	A4	A5	A6	A7	A8
CIFAR-10	38.97 ± 0.36	40.30 ± 0.89	42.40 ± 0.93	43.05 ± 0.81	45.72 ± 0.07	47.62 ± 0.20	51.85 ± 0.68	<b>52.80 ± 0.55</b>
ImageNet ILSVRC	15.60 ± 0.84	17.66 ± 2.16	19.03 ± 1.97	22.17 ± 0.47	24.84 ± 1.00	26.58 ± 1.12	29.14 ± 1.22	<b>30.23 ± 0.67</b>

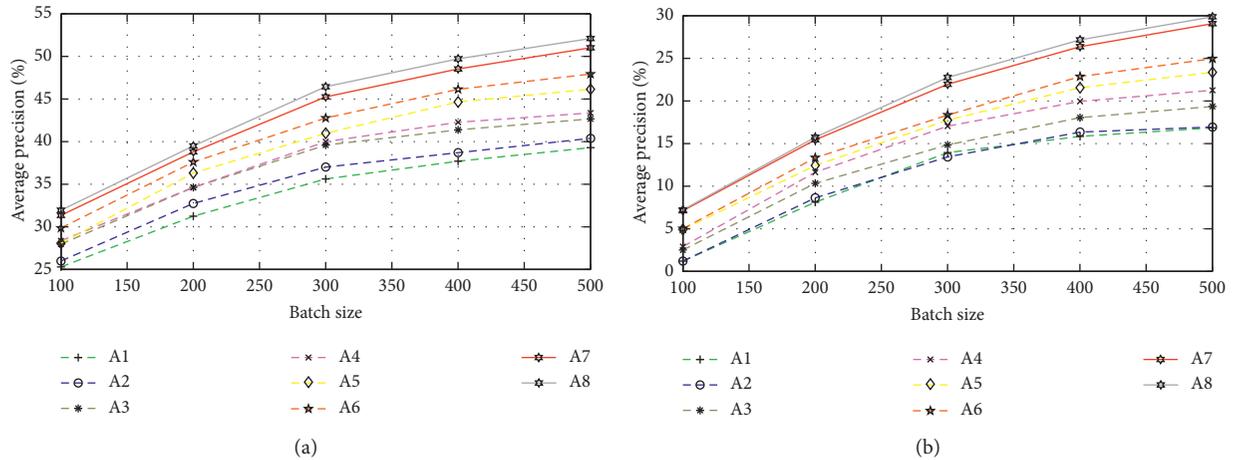


FIGURE 4: AP variations of the first 500 feedback images based on variable batch size are similar with the results based on variable label size in Figure 3. (a) CIFAR-10. (b) ImageNet ILSVRC.

TABLE 2: MAP comparison of eight active learning algorithms with different batch sizes.

MAP (%)	A1	A2	A3	A4	A5	A6	A7	A8
CIFAR-10	39.78 ± 1.35	40.51 ± 0.54	42.88 ± 0.91	43.75 ± 0.07	46.02 ± 0.93	47.52 ± 0.59	51.85 ± 0.70	<b>52.47 ± 0.49</b>
ImageNet ILSVRC	13.38 ± 0.23	13.62 ± 1.30	19.68 ± 0.43	22.07 ± 0.69	23.46 ± 1.26	24.82 ± 0.06	29.14 ± 1.00	<b>30.25 ± 0.73</b>

images. Obviously, we can make the similar conclusion as from Table 1.

## 5. Conclusions

In this paper, we propose a novel MDAL framework for large-scale image classification tasks. To the best of our knowledge, we are the first to incorporate deep learning into the multiview active learning framework. Furthermore, unlike traditional active learnings, both feature representation and classifier are jointly updated with progressively added informatively samples. The detailed experiment results on two public challenging benchmarks justify the effectiveness of our proposed MADL framework.

## Data Availability

All the data utilized in our research can be accessed from the following website: <http://www.cs.toronto.edu/~kriz/cifar.html?usg=alkjrjhqbhw2llxlo8emqns-tbk0at96jq>, <http://www.image-net.org/challenges/LSVRC/>

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was funded by the High-Level Base-Building Project for Industrial Technology Innovation (1021GN204005-A06), Key Research and Development Plan of Zhejiang Province (2018C01086), Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (ICT20047), Zhejiang Provincial Natural Science Foundation of China (LY18F020025), and National Thousand Talents Program (Y474161).

## References

- [1] H. A. Simon and G. Lea, "Problem solving and rule education: a unified view knowledge and organization," *Ernuam*, vol. 15, no. 2, pp. 63–73, 1974.
- [2] X. Li and Y. H. Guo, "Adaptive active learning for image classification," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, 2013.
- [3] K. Konyushkova, R. Sznitman, and P. Fua, "Introducing geometry in active learning for image segmentation," in *Proceedings of the 2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015.
- [4] C. C. Long and G. Hua, "Multi-class multi-annotator active learning with robust Gaussian process for visual recognition," in *Proceedings of the 2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015.

- [5] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267–276, 2010.
- [6] A. Kowdle, Y. J. Chang, A. Gallagher, and T. Chen, "Active learning for piecewise planar 3D reconstruction," in *Proceedings of the 2011 IEEE Computer Vision and Pattern Recognition*, Providence, RI, USA, 2011.
- [7] C. Long, G. Hua, and A. Kapoor, "A joint Gaussian process model for active visual recognition with expertise estimation in crowdsourcing," *International Journal of Computer Vision*, vol. 116, no. 2, pp. 136–160, 2016.
- [8] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, no. 2, pp. 201–221, 1994.
- [9] N. Roy and A. McCallum, "Toward optimal active learning through sampling estimation of error reduction," in *Proceedings of the 18th International Conference on Machine Learning*, Williamstown, MA, USA, 2001.
- [10] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, "Selective sampling using the query by committee algorithm," *Machine Learning*, vol. 28, no. 2-3, pp. 133–168, 1997.
- [11] N. Abe and H. Mamitsuka, "Query learning strategies using boosting and bagging," in *Proceedings of the 15th International Conference on Machine Learning*, Madison, WI, USA, 1998.
- [12] R. E. Schapire, "The boosting approach to machine learning," in *Proceedings of the 2002 MSRI Workshop on Nonlinear Estimation and Classification*, Berkeley, CA, USA, 2002.
- [13] I. Muslea, S. Minton, and C. A. Knoblock, "Selective sampling with redundant views," in *Proceedings of the 17th National Conference on Artificial Intelligence*, Austin, TX, USA, 2000.
- [14] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the 1998 Workshop on Computational Learning Theory*, pp. 92–100, Pittsburgh, PA, USA, 1998.
- [15] I. Muslea, S. Minton, and C. A. Knoblock, "Active semi-supervised learning = robust multi-view learning," in *Proceedings of the 19th International Conference on Machine Learning*, Sydney, Australia, 2002.
- [16] I. Muslea, S. Minton, and C. A. Knoblock, "Active learning with multiple views," *Journal of Artificial Intelligence Research*, vol. 27, pp. 203–233, 2006.
- [17] J. Cheng and K. Q. Wang, "Active learning for image retrieval with Co-SVM," *Pattern Recognition*, vol. 40, no. 1, pp. 330–334, 2006.
- [18] W. Wang and Z. H. Zhou, "On multi-view active learning and the combination with semi-supervised learning," in *Proceedings of the 2008 International Conference on Machine Learning*, Helsinki, Finland, 2008.
- [19] W. Wang and Z. H. Zhou, "Multi-view active learning in the non-realized case," in *Proceedings of the 2010 Conference on Advances in Neural Information Processing Systems*, Vancouver, Canada, 2010.
- [20] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 618–628, 2011.
- [21] J. J. Cai, J. Tang, Q. G. Chen, Y. Hu, X. B. Wang, and S. J. Huang, "Multi-view active learning for video recommendation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, China, 2019.
- [22] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-effective active learning for deep image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 12, pp. 2591–2600, 2017.
- [23] Z. W. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. M. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017.
- [24] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *Proceedings of the 2017 International Conference on Machine Learning*, Sydney, Australia, 2017.
- [25] M. Huijser and J. C. van Gemert, "Active decision boundary annotation with deep generative models," in *Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Italy, 2017.
- [26] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 2, pp. 712–724, 2017.
- [27] B. Q. Lin, Y. Xie, Y. Y. Qu, C. H. Li, and X. D. Liang, "Jointly deep multi-view learning for clustering analysis," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake, UT, USA, 2018.
- [28] A. Krizhevsky, "Learning multiple layers of features from tiny images," Technical report, University of Toronto, Toronto, Canada, 2009.
- [29] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal on Computer Vision*, vol. 115, pp. 211–252, 2015.
- [30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 2012 Conference on Advances in Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012.
- [31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014.
- [32] S. C. H. Hoi, R. Jin, J. K. Zhu, and M. R. Lyu, "Semi-supervised SVM batch mode active learning for image retrieval," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 2008.
- [33] C. C. Zhang and K. Chaudhuri, "Active learning from weak and strong labelers," in *Proceedings of the 2015 Conference on Advances in Neural Information Processing Systems*, Montreal, Canada, 2015.
- [34] B. Demir and L. Bruzzone, "A novel active learning method in relevance feedback for content-based remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2323–2334, 2015.