VNU Journal of Science: Comp. Science & Com. Eng., Vol. 40, No. 2 (2024) 12–21

VNU Journal of Science:
Computer Science and Communication Engineering

Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

Incremental Learning with Adaptive Augmentation for Image-based Active Learning

Nguyen Huynh Tra My, Viet Cuong Ta*

VNU University of Engineering and Technology, Hanoi, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

Received 8 Mar 2024;
Revised 17 May 2024; Accepted 4 June 2024

Abstract: Due to the increasing amount of unlabeled data, a more flexible approach is required to label data efficiently. Active learning aims to identify which data samples are the most valuable for learning with the dataset, thus achieving better performance with much fewer samples. Recent works show that although the data augmentation strategies are simple, they have the potential to improve active learning by expanding the input space's exploration and assisting in the discovery of more informative samples. By effectively controlling a set of augment operators on each active learning cycle, one could choose promising candidates from the set of unlabeled data for each iteration step of active learning. However, the scoring model is built on a hard reset at each data acquisition cycle, which is time-consuming and missing important information from previous cycles. To address the issues, we propose an incremental training procedure for active learning that avoids retraining the scoring model at each updating cycle. By relying on an augmentation strategy, the model can be used to derive a new score based on the combination between the lowest confidence score with its variance in previous cycles. Thus, the resulting scores give a better approximation of the uncertainty of the samples. We evaluate our proposed algorithms on three popular benchmarks, FASHION-MNIST, CIFAR-10, and SVHN, and the results highlight that our method can improve the accuracy from 2% to 4% in comparison with the other baselines.

Keywords: incremental training, data augmentation, active learning.

1. Introduction

Deep learning models have proven as the most successful approach for a wide range of applications [1, 2]. However, these successes often come with the requirement of enough labeled data, thus resulting in high labor and financial cost. Active learning which integrates the activity of acquiring data into the learning process of deep models, can

*Corresponding Author.
E-mail address: cuongtv@vnu.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.2396

be considered as an alternative approach to reduce the required resources for creating such a training dataset. Applications of active learning are employed in the domains of biology [3, 4] and chemistry [5, 6], autonomous driving [7, 8], medical research [9, 10], and remote sensing [11, 12]. To effectively reduce training costs and improve model performance, a selection mechanism of active learning approach is employed to decide which images should be labeled next. Current active learning methods can be divided into two large branches which are uncertainty-based methods [13, 14] and representation-based methods [15]. Compared to the latter, the former is preferred due to the comfort of designing a scoring mechanism for unlabelled data. It is worth noting that without enough training data, it is not easy to learn a good latent representation for the representation-based methods. The main intuition of uncertainty-based approaches is roughly to select the most useful samples from the available data pool which can help to increase the training efficiency of the deep model. In [16], the authors propose a sampling strategy that combines both the predictive uncertainty and the sample diversity into one framework.

On the other hand, one can increase the available training data by employing data augmentation techniques, especially for image-based active learning. Works in [17, 18] employ augmentation techniques on image data to provide virtual data instances for active learning from early iteration. However, these works tend to rely on a fixed set of data augmentation, which makes it difficult to select a proper set of operators. To make the augmentation operator more flexible, the authors in [19] introduce a controllable data augmentation process, namely CAMPAL, to enrich the pooling data. Generally speaking, CAMPAL learns a useful set of data augmentation automatically and based on the learned augmentation decides which are the most prominent samples to be labeled. The next selected samples are decided by a scoring model which is re-trained for each learning cycle. The usage of this scoring model is similar to other approaches such as in [20]. However, the re-training process, results in a more time-consuming process and missing knowledge from previous cycles.

In our work, we address the issue of CAMPAL by applying an incremental learning process for the scoring model. Instead of retraining the model at each learning cycle, we reuse the previous scoring model and adapt it with the new labeled data. The resulting model is then used to sample the next candidates for the next batch. Due to the continuing training process, it requires fewer training epochs in each active learning cycle. Moreover, the resulting model is more effective in identifying the most uncertain samples for the sampling process as it contains the past knowledge of the available data. We further derive a scoring mechanism that takes into account the historical scoring of each sample in the unlabelled data pool. The scoring mechanism is built on the assumption that if a sample has a large fluctuation in its score, it contains more uncertainty than the others. We evaluate our algorithm in three popular benchmarks FASHION-MNIST, CIFAR-10, and SVHN with different active learning settings. Our proposed methods can improve the accuracy by about 2% in the FASHION-MNIST benchmark, 2.2% in the SVHN, and 4% in the CIFAR-10 benchmark, in comparison with the standard augment-based active learning approach.

2. Related Works

Active learning is a branch of machine learning that aims to reduce the amount of labeled data required for training a model by actively selecting the most informative samples for annotation [21, 22]. Recent works on active learning can be divided into uncertainty-based sampling and representation-based sampling. Representation-based approaches choose unlabeled instances, which are considered to contain most of the representations of all classes [15, 23, 24]. On the other hand, uncertainty-based methods select samples that can minimize the uncertainty of the training classifier on certain datasets [25, 26]. Traditional active learning strategies compute the uncertainty of data rely on: entropy [27], query-by-committee [27], maximizing the error reduction [28], disagreement between experts [29] or Bayesian methods [14]. Recent approaches combine different techniques to improve performance such as [13] suggests characterizing the precise behavior of uncertainty sampling for high-dimensional Gaussian mixture data in a contemporary big data regime where the numbers of samples and features are correspondingly large. Another problem in the previous strategy is after each cycle, just one sample is chosen, which consumes a lot of training time. To avoid this, BADGE [16] samples groups of points that are high-magnitude and divergent when represented in a hallucinated gradient space to add sample diversity and predictive uncertainty into each batch that is chosen.

Several works suggest that the size of the data pool has a strong connection to the performance of general active learning framework [30–32]. In general, the small size of the labeled pool in early cycles could lead to a reduction in performance due to the cold-start issue of active learning. To tackle this issue, data augmentation techniques can be employed to enlarge the training pool of active learning methods. [33] propose a way to apply various transformations or modifications to existing image samples to create new synthetic data that is similar but not identical to the original ones. The augmentation operators include basic image transformations or more advanced ones such as picture mixing [34, 35], and geometric alterations [36]. A more recent work [37] proposes to optimize the strength of an augmentation group rather than creating new types of augmentations. To decide which samples are chosen to add to the labeled pool, the value of each sample needs to be evaluated. Some commonly used strategies that rely on uncertainty information include: entropy-based, least confidence, and margin sampling [19]. These methods, however, rely on a set of augmentation operators that have a predetermined strength. This augmentation technique may be feasible with a labeled pool at the current cycle but not with all cycles. To determine various possible strengths with updated training labeled pools in active learning, CAMPAL[19] is the first method which varies the strength of presented augmentation sets, thus adding more flexible to the augmentation operators.

3. Method

3.1. Definitions and Preliminaries

Let $\mathcal{D}$ be the underlying dataset, which is divided into two parts: one with labels $\mathcal{D}_{\mathcal{L}}$ and the other without label $\mathcal{D}_{\mathcal{U}}$, with $|\mathcal{D}_{\mathcal{U}}| \gg |\mathcal{D}_{\mathcal{L}}|$. The uncertainty-based approaches of the active learning framework attempt to extend the set $\mathcal{D}_{\mathcal{L}}$ by selecting the most informative samples from $\mathcal{D}_{\mathcal{U}}$. The uncertainty of a sample is measured by a fully-trained classifier f_θ. One could define a data acquisition function h_{acq} as follows:

$$h_{acq}(x, f_\theta) : \mathcal{D}_U \to \mathbb{R}, \tag{1}$$

which calculates the score for each data instance. Then as with a lot of active learning strategies, the most valuable sample batch will be selected and added to the label pool, which is used for training f_θ.

To leverage the lacking issue of label data samples, the authors of CAMPAL [19] rely on image augmentations such as translation, and histogram equalization to generate more training data. These basic operators for augmentation are denoted as $\mathcal{T}$. For a data point x_U, which

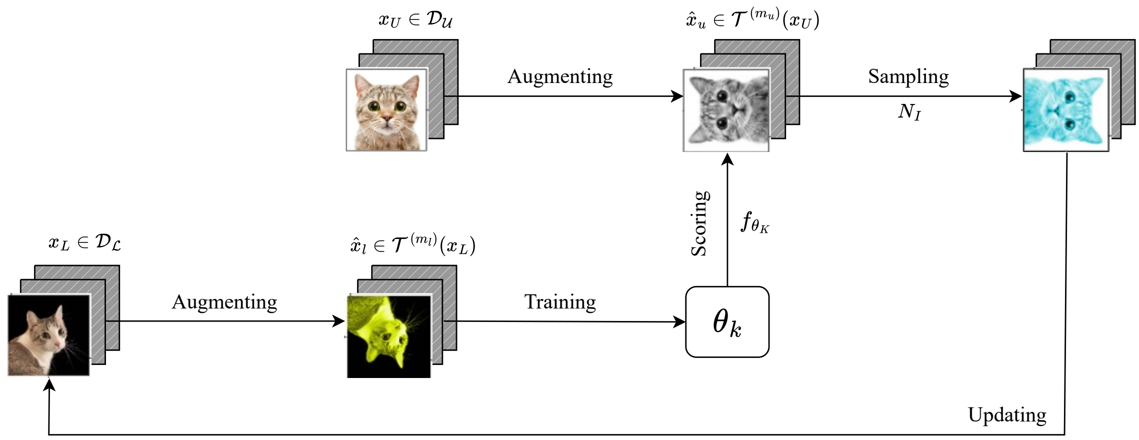

Figure 1. Our proposed incremental in active learning loop with the updated classifier. The $\hat{x}_l$ and $\hat{x}_u$ represent the augmentation set from labelled images and unlabelled images, respectively. The function f_θ denotes a short-form of a parameterized classifier with θ and f_{θ_K} is a specific classifier, which is updated from the previous cycle and re-used for training at cycle k^{th}. In each cycle, N_I unlabelled data is sampled for labelling. Labeled data after being augmented with appropriate operators and strength is used for training θ_k.

is sampled from $\mathcal{D}_\mathcal{U}$, $\mathcal{T}^{(ms)}(x)$ indicates all of its strengthened viewpoints with m. Given this setup of CAMPAL, the scoring function f_θ is re-init after each update on the labeled set $\mathcal{D}_\mathcal{L}$. For selecting the new candidate, several scoring strategies can be used [19]. For example, the margin sampling methods select the data samples that have the lower margin score between the two most confident classes. However, the design choice of re-training $\mathcal{D}_\mathcal{U}$ is mostly a time-consuming process. It usually requires more training epochs with the new samples in $\mathcal{D}_\mathcal{L}$ to arrive at a stable classifier model $\mathcal{D}_\mathcal{U}$. In addition to that, the newly trained classifier of each loop is independent, which results in the missing historical information of the scoring strategy. Therefore, we propose an incremental method to re-use knowledge from previous f_θ to address these issues.

3.2. Incremental active learning with adaptive data augmentation

To train with an incremental learning scoring function f_θ, we rely on the adaptive data augmentation in [19] with a margin sampling approach. The overview of a training cycle is illustrated in Figure 1. Starting with x_U in the unlabeled pool $\mathcal{D}_\mathcal{U}$, the augmenting operator with a strengthened viewpoint $\mathcal{T}^{(m)}$ is employed to generate the augmented $\hat{x}$. The augmented data $\hat{x}$ is then scored with the scoring function f_θ, which is parameterized by a neural network with θ. After selecting the highest score samples, the label of these samples is acquired and then they are added to the labeled pool $\mathcal{D}_\mathcal{L}$. The augmenting operator $\mathcal{T}^{(m)}$ is further applied to $\mathcal{D}_\mathcal{L}$ for generating the training samples to the f_θ. The scoring function f_θ is training by the following equation:

$$\theta \leftarrow \operatorname*{argmin}_{\theta} \frac{1}{|\mathcal{D}_L|} \sum_{x \in \mathcal{D}_L} \mathcal{L}(f_\theta(x), y) \qquad (2)$$

Then, the details of each active learning cycle are illustrated in the Algorithm 1. To provide appropriate control over labeled augmentations without incurring additional training costs, a virtual loss term $\mathcal{L}_f$ is incorporated to find the appropriate strength m_l for labeled samples as

Algorithm 1 Pseudocode of our proposed incremental training process with controllable augmentation at cycle K

1: **Input:** Labeled data pool $\mathcal{D}_L$, unlabeled data pool $\mathcal{D}_U$, classifier parameterized as θ_0
2: $m_l = \text{argmin}_m \frac{1}{|\mathcal{D}_L|} \sum_{x_L \in \mathcal{D}_L} \mathcal{L}_f(x_L, m)$;
3: Generate an augmentation set $\mathcal{T}^{(m_l)}$ with the strength m_l;
4: $\theta_K \leftarrow \text{argmin}_{\theta_{K-1}} \mathcal{L}_{\theta_K}$;
5: $m_u = \text{argmax}_m \sum_{x_U \in \mathcal{D}_U} \min \{\mathbb{H}(\tilde{x}_U) | (\tilde{x}_U) \in \mathcal{T}^{(m)}(x_U), f_\theta(\tilde{x}_U) = f_\theta(x_U)\}$;
6: Generate an augmentation set $\mathcal{T}^{(m_u)}$ with the strength m_u;
7: $h_{temp}(x) = \min_{\tilde{x} \in \mathcal{T}_u(x)} h_{base}(\tilde{x})$
8: $h_{acq}(x) = h_{temp}(x) + var(h_{acq}(x))$
9: Select N_I samples according to h_{acq};
10: $\mathcal{D}_U \leftarrow \mathcal{D}_U - N_I$
11: $\mathcal{D}_L \leftarrow \mathcal{D}_L \cup N_I$

shown in Step 2. The strength m_l is computed by minimizing over the loss $\mathcal{L}_f$ [19].

Classifier f_{θ_K} reuse the historical sample scoring through the load network from the last epoch of the previous active cycle and continue training with $\mathcal{T}^{(m_l)}$, where $\mathcal{L}_{\theta_K}$ is defined as follow:

$$\mathcal{L}_{\theta_K} = \frac{1}{|\mathcal{T}^{(m_{lab})}(\mathcal{D}_L)|} \sum_{x \in \mathcal{T}^{(m_l)}(\mathcal{D}_L)} \mathcal{L}(f_{\theta_{K-1}}(x), y) \quad (3)$$

Then, f_{θ_K} score augmented set $\mathcal{T}^{(m_u)}$, where m_u is a suitable strength that optimizes the unlabeled pool's total informativeness.

With the incremental setup of scoring parameterized by θ_k, it is able to identify the variance of sample scores across different training cycles. The standard margin score for each counterpart in augmented pool $\mathcal{T}_U(x)$ only contains the uncertainty under the specific scoring model at step k. However, when more training data is added to the scoring model, the uncertainty of a sample can be changed in an unpredictable way. Therefore, it is able to reuse the old scores of a sample to approximate the uncertainty of a sample. After generating the augmented set $\mathcal{T}_U(x)$, the temporal score is extracted based on the margin sampling of each augmented image $\hat{x}$

$$h_{temp}(x) = \min_{\tilde{x} \in \mathcal{T}_u(x)} h_{base}(\tilde{x}) \quad (4)$$

We compute variance of k step score and combine the temporal score $h_{temp}(x)$ with the variance over the historical scores:

$$h_{acq}(x) = h_{temp}(x) + var(h_{acq}(x)) \quad (5)$$

Our main intuition of samples that have higher variance scores are more difficult to learn than the other samples. In addition to that, if a sample's score fluctuates much across different training, with different augmentation strengths, the sample may come with much noise. Thus, the corresponding label of the samples is expected to add more information to the learning process. We illustrate our scoring mechanism in Fig. 2 in detail.

4. Experiment Results and Discussions

4.1. Experimental setup

To evaluate our method, we employ the three image datasets: CIFAR-10, SHVN, and FASHION-MNIST for comparison with other baselines. We give the details settings for each dataset as follows:

CIFAR-10[38]: There are 60,000 images in total. Images are 32x32 color images, which are divided into 10 classes, with 6,000 images in each class, consisting of the CIFAR-10 dataset. A total of 50,000 training and 10,000 test images are available.

FASHION-MNIST[39]: The dataset contains a training set with 60,000 examples and a test set with 10,000 examples. Every example consists of a 28 x 28 grayscale picture with a label from one to ten classes.

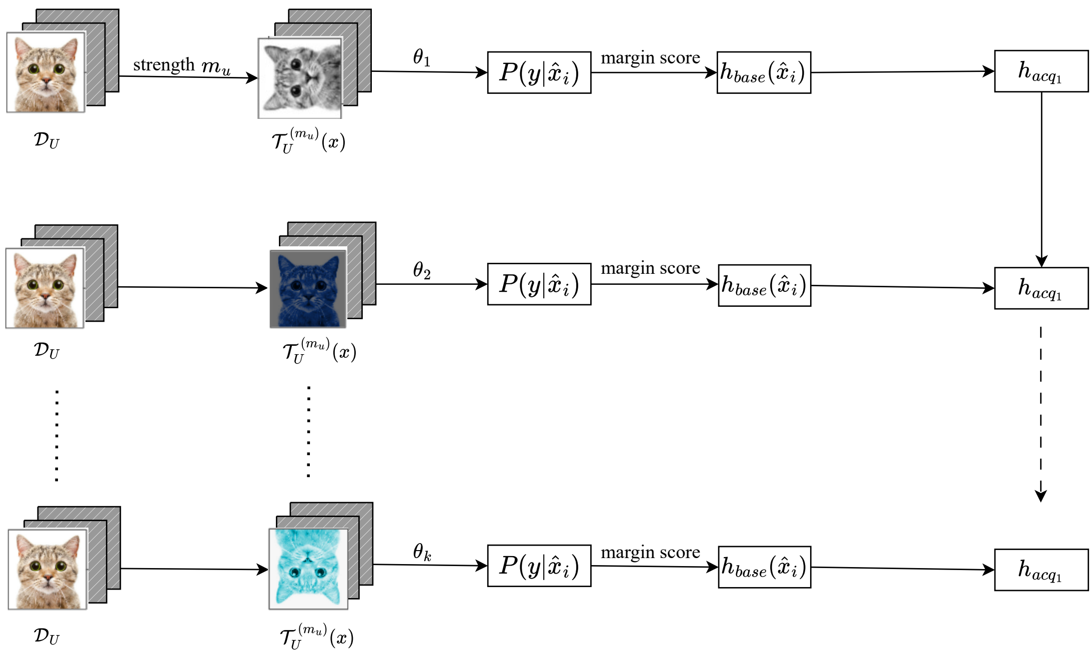

Figure 2. Illustration of our scoring method with scoring function parameterized by θ_k. Each sample $x \in \mathcal{D}_U$ is augmented with strength m_u as in Step 6 from Algorithm 1. Then, the classifier f_θ is used to label for each counterparts $\hat{x}$ in $\mathcal{T}^{(m_u)}(x)$. Based on the margin sampling strategy, margin score h_{base} is measured for each $\hat{x}$, and the minimum score h_{temp} is selected in all the augmented counterparts. Then, all the calculated scores are stored after each learning cycle and used to compute the next sampling score following Eq. 5.

SVHN [40]: A benchmark dataset for digit classification, Street View House Numbers (SVHN), comprises 600,000 32x32 RGB images of printed digits (numbered 0–9) that have been clipped from photos of house number plates. The reduced images retain surrounding numbers and other distractions, but they are centered around the digit of interest. Three sets comprise SVHN: training, testing, and an additional set of 530,000 easier-to-understand images that can be utilized to aid in the training process.

To train with an active learning framework, we first construct a random initial dataset with N_I labeled samples, then when training fully-classifier f_θ, we add N_I instances into the labeled set after each cycle. In this section, we choose N_I = {500, 1000} in both datasets to conduct experiment. The training cycle with active learning is 5 cycles. The set of augmentation operators uses conventional augmentation pools such as random crop, flip, and rotate which are predefined and fixed for each dataset. The ResNet-18 architecture [41] is used as the backbone architecture for both the base model θ_k and the classification model. Each update of the base model θ_k is 5 epochs with an SGD optimizer of learning rate 0.01, momentum 0.9, and weight decay 5e-4.

We compare our proposed method with other scoring-based systems including the three sampling methods **Entropy**, Least Confidence (**LC**), **Margin**, which are mentioned in the implementation of CAMPAL [19]. In addition to that, we also employ the **BADGE** [16] approach which is the standard baseline for batch active learning. To highlight the effectiveness of

Dataset	Method	CIFAR-10	SVHN	FASHION-MNIST
$N_I = 500$	BADGE	60.87 ± 2.38	82.63 ± 1.80	<u>97.57 ± 0.14</u>
	Entropy	62.11 ± 1.29	81.12 ± 2.88	97.03 ± 0.84
	LC	59.50 ± 2.89	81.42 ± 1.46	96.88 ± 0.72
	Margin	61.52 ± 2.46	80.91 ± 0.49	97.37 ± 0.31
	Incre	<u>62.60 ± 3.29</u>	<u>82.73 ± 1.43</u>	96.86 ± 0.36
	IncreVR	**66.07 ± 1.45**	**84.31 ± 0.53**	**97.65 ± 0.17**
$N_I = 1000$	BADGE	70.88 ± 1.78	<u>85.13 ± 0.53</u>	97.66 ± 0.28
	Entropy	70.67 ± 0.96	85.11 ± 2.00	98.15 ± 0.36
	LC	<u>71.16 ± 2.49</u>	85.10 ± 0.21	<u>98.28 ± 0.16</u>
	Margin	70.53 ± 1.12	83.95 ± 0.56	98.19 ± 0.46
	Incre	69.67 ± 5.40	84.08 ± 0.48	97.56 ± 0.61
	IncreVR	**72.15 ± 0.87**	**87.30 ± 0.33**	**98.29 ± 0.09**

Table 1. Comparison of the averaged test accuracy on three benchmarks and different active learning strategies. The best performance in each dataset is indicated in **bold** while the second performance is <u>underlined</u>. N_I denotes the number of initial labeled samples as well as the size of the added labeled pool.

our proposed method, we further implement an incremental version of CAMPAL (**Incre**) which replaces the initial classifier after each cycle with an incremental training model from the last epoch of the previous cycle. Our work is denoted as **IncreVR**, which adds the variance into the incremental training of θ_k.

4.2. Main Empirical Results

Table 1 shows the overall performance of our proposed method and the above baselines with two datasets, CIFAR-10 and FASHION-MNIST. Our proposed method **IncreVR**, which combines with a new scoring strategy to choose which samples are selected, has the most competitive performance against the other baselines. The largest increase is in the settings of $N_I = 500$ on the CIFAR-10 dataset, which has an improvement of about **3.47%**. In SVHN, our proposed method has a stable improvement, about **2.2%** in both settings of $N_I = 500$ and $N_I = 1000$. The other baselines are not as stable as our methods across the four settings. Our method shows the most improvements in the CIFAR-10, in comparison to the FASHION-MNIST. This can be explained by the latter being more trivial to learn than the former. The non-variance addition version of our approach, namely **Incre**, only has a slight advantage against the other baseline. Among tested baselines, the BADGE method underperformed in comparison with other augmentation-based approaches.

Table 2 shows the comparison of the accuracy and total training time across the Margin baseline and our IncreVR with the same settings. We set up this experiment on CIFAR-10 with 5 cycles while changing the number of epochs of each cycle. With the same epoch setup, our proposed method has the same training time as the baseline; however, our IncreVR outperforms the Margin baseline. Moreover, our **IncreVR** with 3 epochs has a smaller training time while outperforming the baseline with a setup of 4 epochs, which is similar when comparing 4 epochs **IncreVR** with 5 epochs baseline.

Figure 3 highlights the effectiveness of our **IncreVR** with different numbers of added samples in the CIFAR-10 dataset. We keep the same setting as the above except that we vary the sampling added size at each active cycle

Epoch	Margin		IncreVR	
	Acc.	**Time**	**Acc.**	**Time**
3	55.01	$5945s$	58.00	$5940s$
4	57.32	$6255s$	62.66	$6250s$
5	61.52	$6553s$	66.07	$6550s$

Table 2. Comparison of the averaged accuracy on test data and total training time on CIFAR-10 between Margin and our IncreVR.

$N_I = \{100, 200, 300, 400, 500\}$. As can be seen from Figure 3, with $N_I = 100$, both incremental learning methods, **IncreVR** and **Incre**, achieve the highest performance. The variance addition scoring strategy has better results when N_I increases to 500. Meanwhile, the incremental version **Incre** works best with a smaller number of added samples.

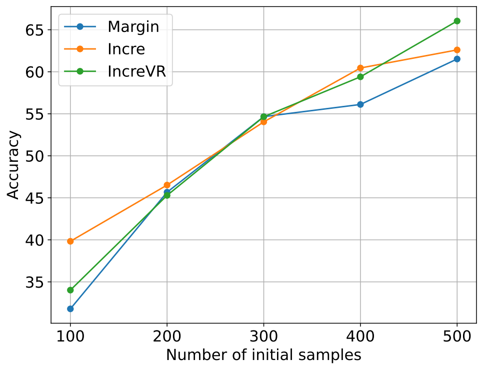

Figure 3. Experiment on different N_I with three scoring strategies in CIFAR-10. The number of N_I is varied from 100 to 500.

Figure 4 highlights the improvement in performance with the change in the number of training cycles. We apply reduced settings, which reduce the added size N_I and reduce the updating training epochs to 1. This setting is used to illustrate our efficiency in a limited learning setting. Our method **IncreVR** completely outperforms the **Margin** approach and the other version **Incre**. When the number of training cycles increases, the effect of variance addition is demonstrated clearly. With 20 updating cycles, **IncreVR** outperforms the other methods 3% approximately. Note that, because we only use $N_I = 100$ with 1 training epoch, the result accuracy is not as high as the other settings.

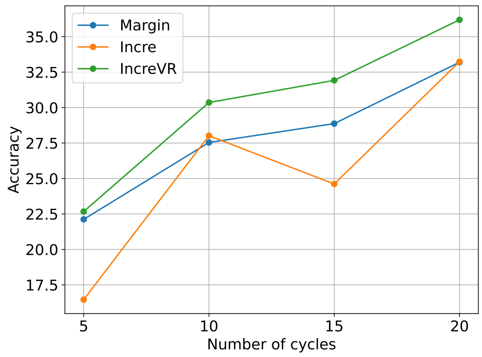

Figure 4. Experiment on different updating *cycle* for selecting new samples in CIFAR-10; the number of epochs is decreased to 1 and N_I is set to 100 samples.

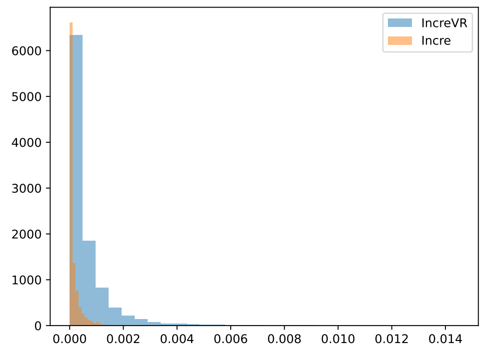

Figure 5. The fitting distribution of sample scores by sorting increases after 5 updating cycles on the CIFAR-10 with $N_I = 500$.

Figures 5 also plot the distribution of scores, which is used to evaluate the uncertainty of each data point in the augmented pool. To fairly compare, we choose a fixed setup with $N_I = 500$, the active learning cycle is 5 with each loop having 5 epochs and the same seed. It can be seen from the plot that the score range of **Incre** is narrower than the range of score with variance in the plot of **IncreVR**. By including the high variance sample across each updating cycle, our proposed algorithm can spread a range of uncertainty samples, thus we can discover more valuable newly labeled data.

5. Conclusions

In this paper, we present an augmented-based active learning method that employs incremental training in the score function. While re-training the model after each active learning cycle is time-consuming and misses important information from the previous cycle, our incremental training approach can be used to extract such information for evaluating the uncertainty of a sample. Based on this observation, we propose the IncreVR approach which combines the variance of sample scores for selecting the next added samples. We experiment on three benchmarks CIFAR-10, SVHN, and FASHION-MNIST to prove the effectiveness of our model. With CIFAR-10 and SVHN, the result shows that our proposed method outperforms all baselines from 2% to 4% accuracy while providing competitive performance in the FASHION-MNIST dataset. Given the flexibility of the proposed method, our work can be extended to include more flexible sampling techniques such as based on latent representation or reinforcement learning settings.

Acknowledgements

This work is funded by the Ministry of Science and Technology of Vietnam under Program KC 4.0, project title: "Research on the application of artificial intelligence in lung cancer diagnosis through analysis of chest CT images, flexible bronchoscopy images and histopathology images", No. KC-4.0-40/19-25.

References

[1] S. Dong, P. Wang, K. Abbas, A Survey on Deep Learning and Its Applications, Computer Science Review, Vol. 40, 2021, pp. 100379.

[2] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, S. S. Iyengar, A Survey on Deep Learning: Algorithms, Techniques, and Applications, ACM Computing Surveys (CSUR), Vol. 51, No. 5, 2018, pp. 1–36.

[3] X. Chen, S. Chen, M. Thomson, Active Feature Selection Discovers Minimal Gene Sets for Classifying Cell Types and Disease States with Single-cell mRNA-seq Data (2022). arXiv:2106.08317.

[4] X. Guan, Z. Li, Y. Zhou, W. Shao, D. Zhang, Active Learning for Efficient Analysis of High-throughput Nanopore Data, Bioinformatics, Vol. 39, No. 1, 2022, pp. btac764. doi:10.1093/bioinformatics/btac764.

[5] R. Tchoua, A. Ajith, Z. Hong, L. Ward, K. Chard, D. Audus, S. Patel, J. de Pablo, I. Foster, Active Learning Yields Better Training Data for Scientific Named Entity Recognition, in: 2019 15th International Conference on eScience (eScience), 2019, pp. 126–135. doi:10.1109/eScience.2019.00021.

[6] S. J. Ang, W. Wang, D. Schwalbe-Koda, S. Axelrod, R. Gómez-Bombarelli, Active Learning Accelerates ab Initio Molecular Dynamics on Reactive Energy Surfaces, Chem, Vol. 7, No. 3, 2021, pp. 738–751. doi:10.1016/j.chempr.2020.12.009.

[7] F. Peng, C. Wang, J. Liu, Z. Yang, Active Learning for Lane Detection: A Knowledge Distillation Approach, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 15132–15141. doi:10.1109/ICCV48922.2021.01487.

[8] O. Akcin, O. Unuvar, O. Ure, S. P. Chinchali, Fleet Active Learning: A Submodular Maximization Approach, in: 7th Annual Conference on Robot Learning, 2023.

[9] S. Nakarmi, K. Santosh, Active Learning to Minimize the Risk from Future Epidemics, in: 2023 IEEE Conference on Artificial Intelligence (CAI), 2023, pp. 329–330. doi:10.1109/CAI54212.2023.00145.

[10] M. A. Kadir, H. M. T. Alam, D. Sonntag, EdgeAL: An Edge Estimation Based Active Learning Approach for OCT Segmentation (2023). arXiv:2307.10745.

[11] X. Liao, B. Tu, J. Li, A. Plaza, Class-wise Graph Embedding-Based Active Learning for Hyperspectral Image Classification, IEEE Transactions on Geoscience and Remote Sensing, Vol. 61, 2023, pp. 1–13. doi:10.1109/TGRS.2023.3309032.

[12] G. Li, Q. Gao, M. Yang, X. Gao, Active Learning Based on Similarity Level Histogram and Adaptive-scale Sampling for Very High Resolution Image Classification, Neural Networks, Vol. 167, 2023, pp. 22–35. doi:10.1016/j.neunet.2023.08.012.

[13] X. Mai, S. Avestimehr, A. Ortega, M. Soltanolkotabi, On The Effectiveness of Active Learning by Uncertainty Sampling in Classification of High-Dimensional Gaussian Mixture Data, in: ICASSP 2022 - 2022 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 4238–4242. doi:10.1109/ICASSP43922.2022.9747685.

[14] Y. Gal, R. Islam, Z. Ghahramani, Deep Bayesian Active Learning with Image Data (2017). arXiv:1703.02910.

[15] O. Sener, S. Savarese, Active Learning for Convolutional Neural Networks: A Core-Set Approach (2018). arXiv:1708.00489.

[16] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, A. Agarwal, Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds (2020). arXiv:1906.03671.

[17] T. Tran, T.-T. Do, I. Reid, G. Carneiro, Bayesian Generative Active Deep Learning (2019). arXiv:1904.11643.

[18] Y.-Y. Kim, K. Song, J. Jang, I.-C. Moon, LADA: Look-Ahead Data Acquisition via Augmentation for Active Learning (2020). arXiv:2011.04194.

[19] J. Yang, H. Wang, S. Wu, G. Chen, J. Zhao, Towards Controlled Data Augmentations for Active Learning, in: A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, J. Scarlett (Eds.), Proceedings of the 40th International Conference on Machine Learning, Vol. 202 of Proceedings of Machine Learning Research, PMLR, 2023, pp. 39524–39542.

[20] J. T. Ash, R. P. Adams, On Warm-Starting Neural Network Training (2020). arXiv:1910.08475.

[21] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, X. Wang, A Survey of Deep Active Learning (2021). arXiv:2009.00236.

[22] S. Dasgupta, Two Faces of Active Learning, Theoretical Computer Science, Vol. 412, No. 19, 2011, pp. 1767–1781, algorithmic Learning Theory (ALT 2009).

[23] Y. Geifman, R. El-Yaniv, Deep Active Learning over the Long Tail (2017). arXiv:1711.00941.

[24] D. Gissin, S. Shalev-Shwartz, Discriminative Active Learning (2019). arXiv:1907.06347.

[25] S. Hanneke, et al., Theory of Disagreement-based Active Learning, Foundations and Trends® in Machine Learning, Vol. 7, No. 2-3, 2014, pp. 131–309. doi:10.1561/2200000037.

[26] W. H. Beluch, T. Genewein, A. Nurnberger, J. M. Kohler, The Power of Ensembles for Active Learning in Image Classification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 9368–9377. doi:10.1109/CVPR.2018.00976.

[27] C. E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, 1948, pp. 379–423.

[28] N. Roy, A. McCallum, Toward Optimal Active Learning through Monte Carlo Estimation of Error Reduction, in: International Conference on Machine Learning, 2001.

[29] S. Argamon-Engelson, I. Dagan, Committee-Based Sample Selection for Probabilistic Classifiers, Journal of Artificial Intelligence Research, Vol. 11, 1999, pp. 335–360. doi:10.1613/jair.612.

[30] A. Beygelzimer, S. Dasgupta, J. Langford, Importance Weighted Active Learning (2009). arXiv:0812.4952.

[31] D. Roth, K. Small, Margin-Based Active Learning for Structured Output Spaces, in: Proceedings of the 17th European Conference on Machine Learning, ECML'06, Springer-Verlag, Berlin, Heidelberg, 2006, p. 413–424.

[32] A. Frieze, R. Kannan, S. Vempala, Fast Monte-carlo Algorithms for Finding Low-rank Approximations, J. ACM, Vol. 51, No. 6, 2004, pp. 1025–1041.

[33] M. Xu, S. Yoon, A. Fuentes, D. S. Park, A Comprehensive Survey of Image Augmentation Techniques for Deep Learning, Pattern Recognition, Vol. 137, 2023, pp. 109347. doi:10.1016/j.patcog.2023.109347.

[34] B. Zhang, L. Li, S. Yang, S. Wang, Z.-J. Zha, Q. Huang, State-Relabeling Adversarial Active Learning (2020). arXiv:2004.04943.

[35] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, D. Rueckert, GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks (2018). arXiv:1810.10863.

[36] C. Shorten, T. M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, J. Big Data, Vol. 6, 2019, pp. 60. doi:10.1186/s40537-019-0197-0.

[37] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, RandAugment: Practical Automated Data Augmentation with a Reduced Search Space (2019). arXiv:1909.13719.

[38] A. Krizhevsky, G. Hinton, et al., Learning Multiple Layers of Features from Tiny Images (2009).

[39] H. Xiao, K. Rasul, R. Vollgraf, Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms (2017). arXiv:1708.07747.

[40] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A. Y. Ng, et al., Reading Digits in Natural Images with Unsupervised Feature Learning, in: NIPS Workshop on Deep Learning and Unsupervised Feature Learning, Granada, Spain, 2011, p. 7.

[41] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, CoRR, Vol. abs/1512.03385, (2015). arXiv:1512.03385.
URL `http://arxiv.org/abs/1512.03385`