Supplimentary Material of
Learning Loss for Active Learning

Donggeun Yoo[1,2] and In So Kweon[2]

[1]Lunit Inc., Seoul, South Korea.
[2]KAIST, Daejeon, South Korea.

dgyoo@lunit.io iskweon77@kaist.ac.kr

In this supplementary material, we provide numbers used for plots in the paper and examine performance changes according to several options for designing and learning the loss prediction module. The material is composed as follows.

Table of Contents

1. Values in the Plots

In the paper, the experimental results have been provided as plots due to the limited space. We provide a table for each plot in this supplementary material. Table 1,2,3,4 correspond to Figure 4,6,7,5 of the paper, respectively.

Number of labeled images	Random	Entropy [4]	Core-set [6]	Learn loss (MSE)	Learn loss
1k	0.5122±0.0100	0.5122±0.0100	0.5122±0.0100	0.5102±0.0125	0.5103±0.0083
2k	0.6225±0.0116	0.6365±0.0123	0.6333±0.0108	0.6236±0.0080	0.6513±0.0168
3k	0.7032±0.0115	0.7227±0.0126	0.7157±0.0159	0.7046±0.0103	0.7650±0.0087
4k	0.7602±0.0106	0.7845±0.0141	0.7850±0.0093	0.7747±0.0066	0.8193±0.0051
5k	0.8046±0.0060	0.8294±0.0098	0.8233±0.0039	0.8177±0.0040	0.8496±0.0017
6k	0.8276±0.0032	0.8534±0.0037	0.8495±0.0025	0.8432±0.0035	0.8683±0.0018
7k	0.8427±0.0036	0.8721±0.0034	0.8662±0.0037	0.8623±0.0037	0.8799±0.0025
8k	0.8555±0.0010	0.8848±0.0032	0.8801±0.0029	0.8770±0.0053	0.8918±0.0035
9k	0.8653±0.0017	0.8962±0.0035	0.8913±0.0005	0.8886±0.0041	0.9016±0.0011
10k	0.8764±0.0027	0.9059±0.0014	0.9010±0.0021	0.9009±0.0024	0.9101±0.0029

Table 1. Image classification results (Accuracy) on CIFAR-10 [5]. The performance mean and standard deviation from 5 trials are reported.

Number of labeled images	Random	Entropy [4]	Core-set [6]	Learn loss
1k	0.5262±0.0062	0.5262±0.0062	0.5262±0.0062	0.5238±0.0028
2k	0.6082±0.0019	0.6123±0.0081	0.6236±0.0052	0.6095±0.0042
3k	0.6423±0.0022	0.6357±0.0091	0.6590±0.0043	0.6491±0.0047
4k	0.6633±0.0018	0.6694±0.0021	0.6763±0.0021	0.6690±0.0028
5k	0.6751±0.0017	0.6870±0.0015	0.6888±0.0048	0.6905±0.0045
6k	0.6860±0.0050	0.6982±0.0011	0.6944±0.0032	0.7035±0.0055
7k	0.6927±0.0016	0.7018±0.0027	0.7016±0.0013	0.7149±0.0066
8k	0.7010±0.0017	0.7112±0.0012	0.7083±0.0012	0.7213±0.0060
9k	0.7044±0.0047	0.7166±0.0031	0.7115±0.0016	0.7273±0.0030
10k	0.7117±0.0016	0.7222±0.0024	0.7171±0.0025	0.7338±0.0028

Table 2. Object detection results (mAP) on PASCAL VOC 2007+2012 [2]. The performance mean and standard deviation from 3 trials are reported.

Number of labeled poses	Random	Entropy [4]	Core-set [6]	Learn loss
1k	0.6143±0.0020	0.6143±0.0020	0.6143±0.0020	0.5839±0.0011
2k	0.6923±0.0028	0.6874±0.0069	0.6859±0.0047	0.6781±0.0037
3k	0.7222±0.0046	0.7185±0.0035	0.7214±0.0025	0.7174±0.0040
4k	0.7385±0.0056	0.7380±0.0035	0.7392±0.0036	0.7400±0.0034
5k	0.7534±0.0044	0.7522±0.0047	0.7542±0.0021	0.7589±0.0032
6k	0.7622±0.0038	0.7624±0.0013	0.7646±0.0041	0.7728±0.0037
7k	0.7719±0.0048	0.7688±0.0011	0.7769±0.0029	0.7840±0.0029
8k	0.7765±0.0043	0.7788±0.0009	0.7856±0.0023	0.7918±0.0024
9k	0.7817±0.0030	0.7844±0.0021	0.7902±0.0013	0.7965±0.0057
10k	0.7862±0.0023	0.7899±0.0019	0.7985±0.0031	0.8046±0.0058

Table 3. Human pose estimation results (PCKh@0.5) on MPII [1]. The performance mean and standard deviation from 3 trials are reported.

Number of labeled images or poses	CIFAR-10	CIFAR-10 (MSE)	PASCAL VOC 2007+2012	MPII
1k	0.6545±0.0071	0.6365±0.0078	0.6916±0.0065	0.7600±0.0082
2k	0.7774±0.0132	0.7196±0.0114	0.6854±0.0007	0.7573±0.0069
3k	0.8682±0.0078	0.7954±0.0141	0.6829±0.0147	0.7575±0.0061
4k	0.8972±0.0021	0.8436±0.0080	0.6868±0.0049	0.7431±0.0104
5k	0.9146±0.0046	0.8524±0.0077	0.6944±0.0043	0.7471±0.0077
6k	0.9220±0.0039	0.8660±0.0109	0.7071±0.0013	0.7611±0.0142
7k	0.9171±0.0043	0.8726±0.0110	0.7080±0.0052	0.7534±0.0087
8k	0.9097±0.0060	0.8714±0.0076	0.7048±0.0088	0.7478±0.0086
9k	0.9070±0.0070	0.8701±0.0066	0.7103±0.0092	0.7656±0.0069
10k	0.9074±0.0120	0.8771±0.0095	0.7133±0.0076	0.7458±0.0089

Table 4. Ranking accuracy of the loss prediction module for each dataset. The performance mean and standard deviation are reported.

2. Single-Level VS. Multi-Level

We have used 4 mid-level feature maps as inputs to the loss prediction module in CIFAR-10 [5]. In this section, we examine how performance changes if only the last feature map from "conv5_2" of ResNet-18 [3] is given to the module. Figure 1 shows the comparison results. The use of a single feature map yields slightly lower active learning performances, as it has lower ranking accuracy.

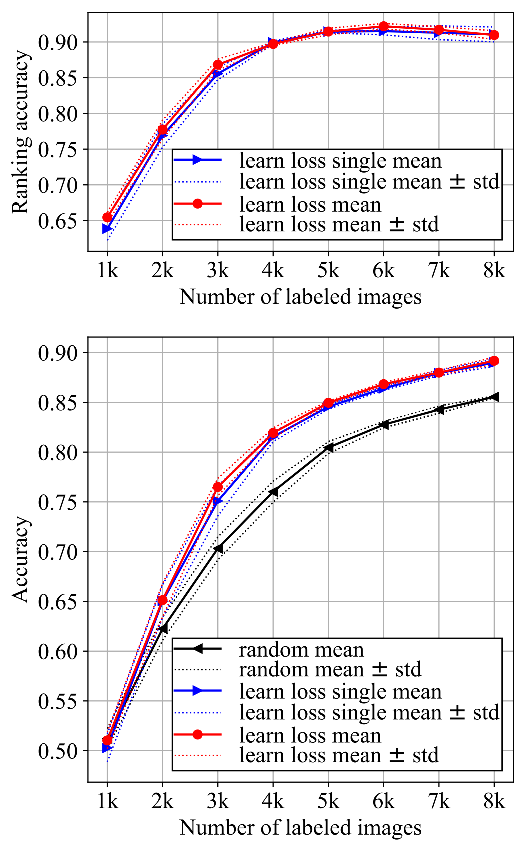


Figure 1. Comparison results on CIFAR-10 [5]. Top: Ranking accuracy. Bottom: Classification accuracy.

3. Wider Loss Prediction Module

We have designed each fully connected layer (FC) after a global average pooling layer (Figure 2 in the paper) to produce a 128-dimensional feature. In this section, we examine how performance changes if we add more parameters to the FC to produce higher dimensional features such as 256 and 512. Figure 3 shows the comparison results. The loss prediction modules show similar ranking accuracy regardless of the FC sizes and thus achieve similar performances in active learning.

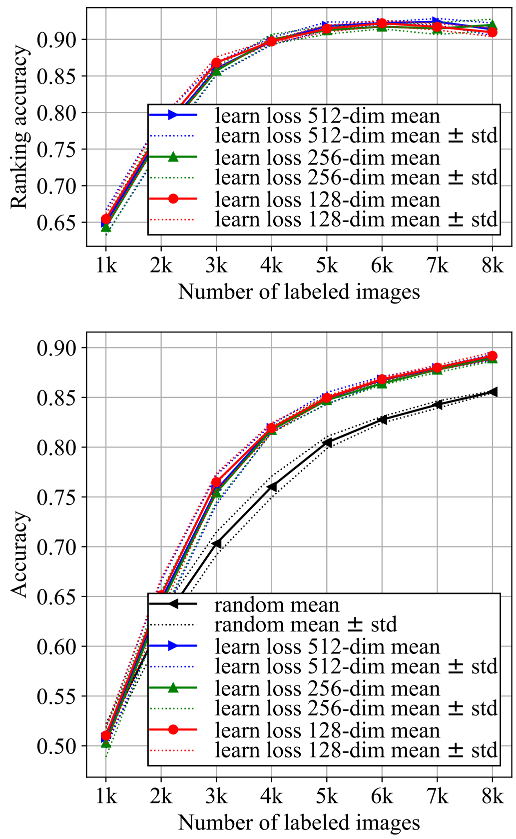


Figure 2. Comparison results on CIFAR-10 [5]. Top: Ranking accuracy. Bottom: Classification accuracy.

4. Deeper Loss Prediction module

Our loss prediction module illustrated in Figure 2 of the paper is shallow as it contains only two-story fully connected layers (FC). This module does not need to be deeper since the convolution filters contained in the target model can be tuned for loss prediction. In this section, we design a deeper loss prediction module by adding a convolution layer (128 kernels of 3×3 size with a stride of 1 and padding of 1) with a batch normalization layer as shown in Figure 4 to examine the performance changes. Figure 4 shows the comparison results. The loss prediction modules show similar ranking accuracy regardless of the depth and thus achieve similar performances in active learning.

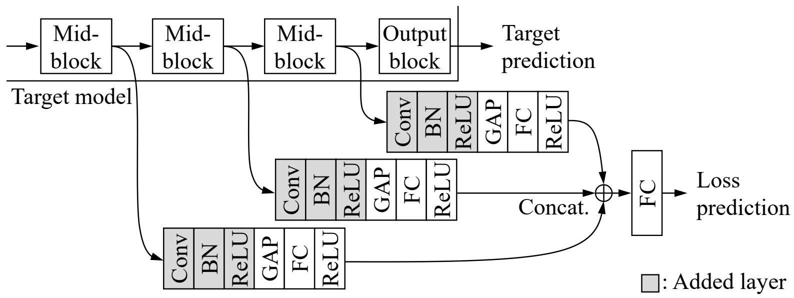


Figure 3. A deeper loss prediction module with convolution and batch normalization layers added.

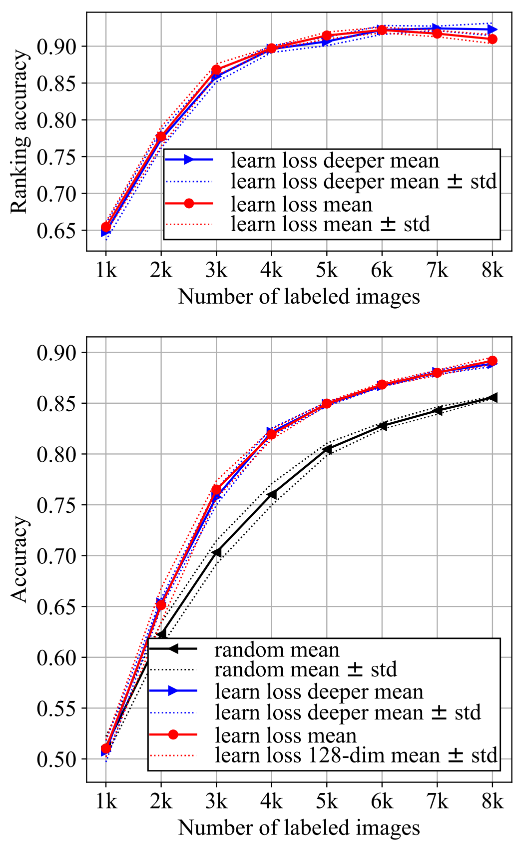
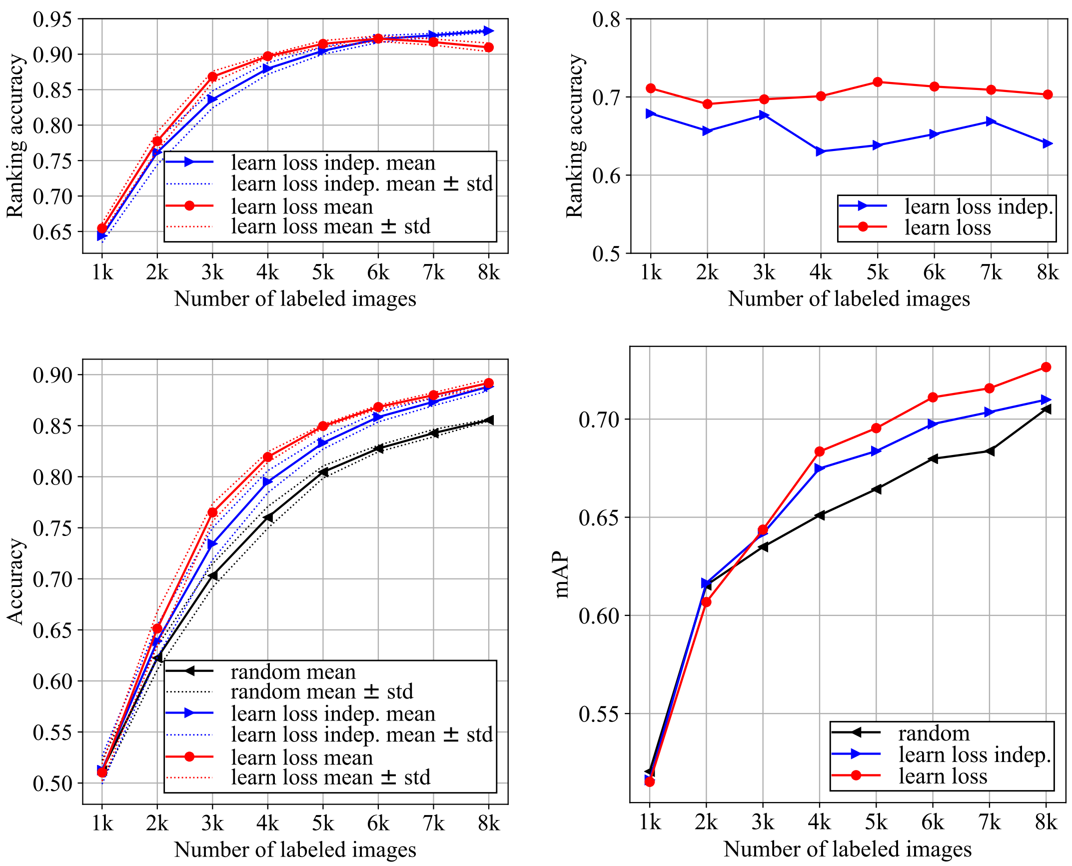


Figure 4. Comparison results on CIFAR-10 [5]. Top: Ranking accuracy. Bottom: Classification accuracy.

5. Independent Learning

Given the total loss L defined in Equation (3) in the paper, we update the model set $(\Theta_{\text{target}}, \Theta_{\text{loss}})$ by the back-propagation algorithm. A set of gradients from the loss prediction module $\partial L/\partial h$, where h is a set of multi-level feature maps as inputs to the loss prediction module, is also propagated to the target model to update Θ_{target}. We examine how performance changes if $\partial L/\partial h$ does *not* propagate to the target model. In this configuration, the loss prediction module cannot tune Θ_{target}.

Figure 5 shows the comparison results. "learn loss indep." represents the independent learning. In Figure 5 (a) CIFAR-10 [5], the loss prediction performance of this independent learning is significantly lower at early stages and reversed at the later stages. Therefore, the active learning performance of this method is low at early stages but close to that of the original method at later stages. To see if this phenomenon also occurs in other visual recognition tasks, we have attempted the same experiment for object detection over PASCAL VOC 2007+2012 [2] as shown in Figure 5 (b). However, in this task, the loss prediction performance of independent learning is always significantly lower than the original method. As a result, less informative data is accumulated in the labeled pool and the active learning performance gap to the original method has continued to widen.



(a) CIFAR-10 [5]

(b) PASCAL VOC 2007+2012 [2]

Figure 5. Comparison results. Top: Ranking accuracy. Bottom: Active learning performances.

References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014.

[2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] A. JOSHI. Multi-class active learning for image classification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2372–2379, 2009.

[5] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

[6] O. Sener and S. Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018.