Learning to Rank for Active Learning: A Listwise Approach

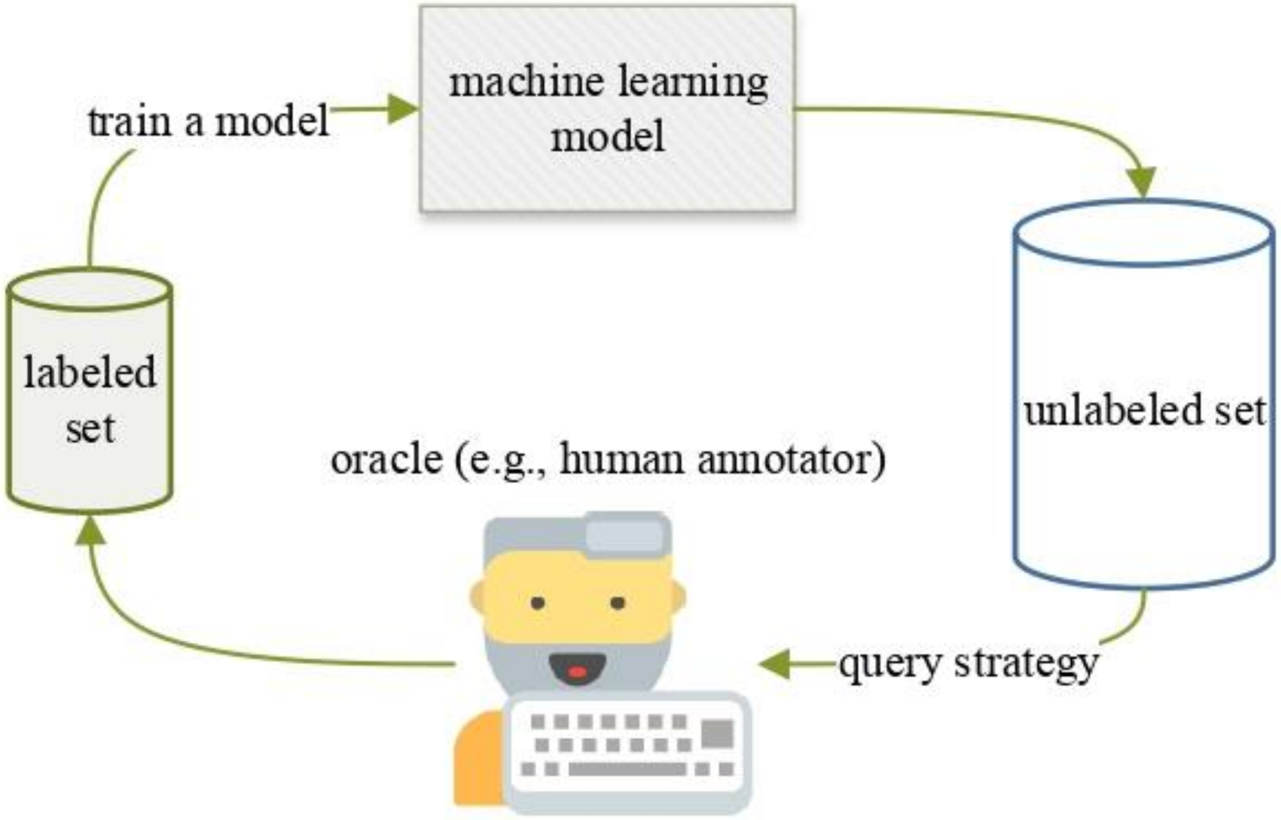
Minghan Li∗†, Xialei Liu∗ , Joost van de Weijer∗ , Bogdan Raducanu∗

∗Computer Vision Center, Universitat Autonoma de Barcelona, Spain

†Universite Grenoble Alpes, France

Active Learning

- Annotating data manually is time consuming and expensive
- The goal of active learning is to automatically select a number of unlabeled samples for annotation, based on an query function which indicates how valuable a sample is for training the model



Related SOTA Work

Learning loss for active learning (LL4AL):

- Task-agnostic and effective.
- This method learns to predict the loss of unlabeled input sample, and uses the predicted loss as a measure of uncertainty

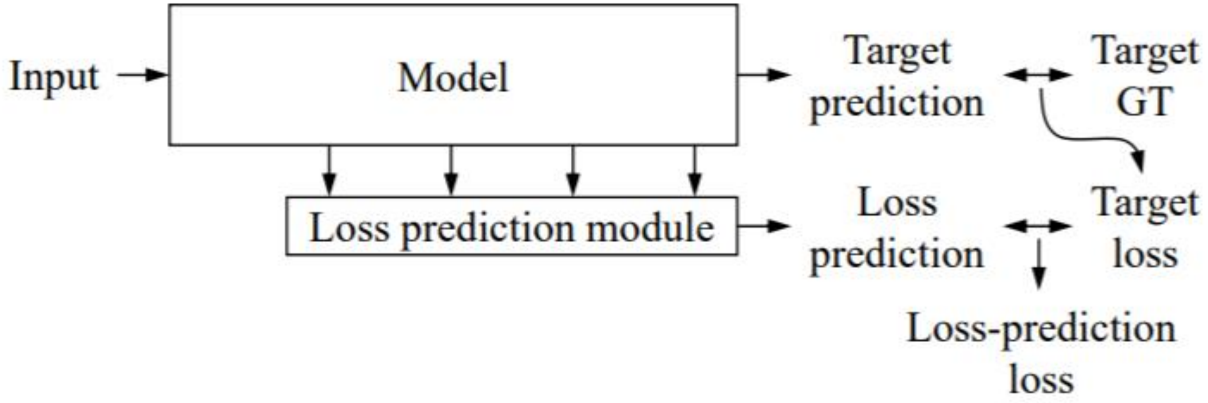


Figure from: Yoo, Donggeun, and In So Kweon. "Learning loss for active learning." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

The Problem of LL4AL

The Authors mentioned in their paper:

“Perhaps the simplest way to define the loss-prediction loss function is the mean square error (**MSE**). However, MSE is not a suitable choice for this problem since the scale of the real loss changes (**decreases** in overall) as learning of the target model progresses. Minimizing MSE would let the loss prediction module adapt roughly to the scale changes of the loss, rather than fitting to the exact value. We have tried to minimize MSE but **failed** to learn a good loss prediction module, and active learning with this module actually demonstrates performance worse than previous methods.”

The solution of LL4AL

Their solution is to compare a pair of samples.

In the mini-batch whose size is B, B/2 data pairs can be made.

The loss function for a pair is defined as:

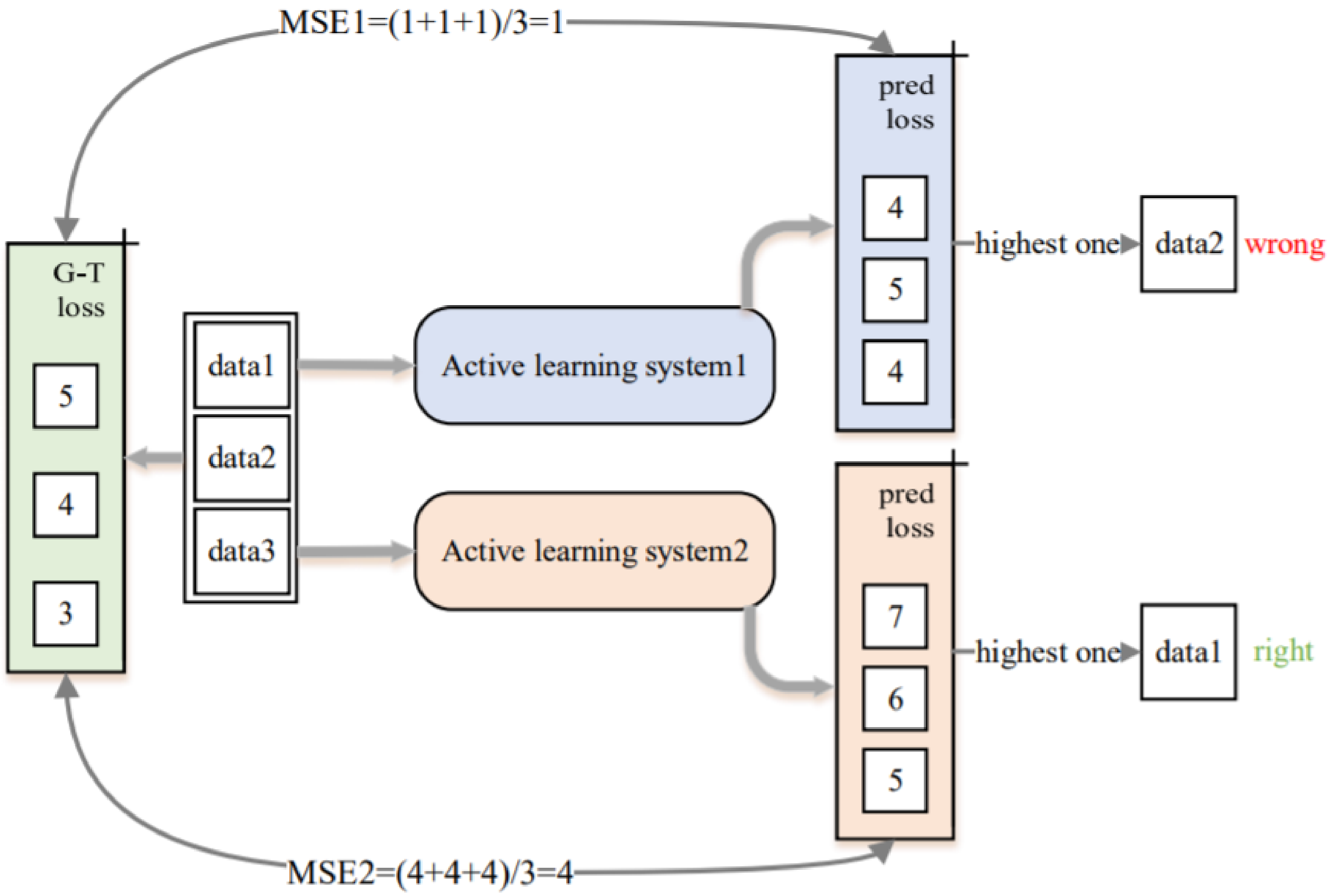
$$L_{\text{loss}}(\hat{l^p}, l^p) = \max\left(0, -\mathbb{1}(l_i, l_j) \cdot (\hat{l_i} - \hat{l_j}) + \xi\right)$$

$$\text{s.t.} \quad \mathbb{1}(l_i, l_j) = \begin{cases} +1, & \text{if} \quad l_i > l_j \\ -1, & \text{otherwise} \end{cases}$$

However, the learning loss based active learning problem is actually a **ranking problem**. We clarify this aspect in this paper and demonstrate that the loss prediction module should be **trained by minimizing the ranking error**.

Why Learning to Rank

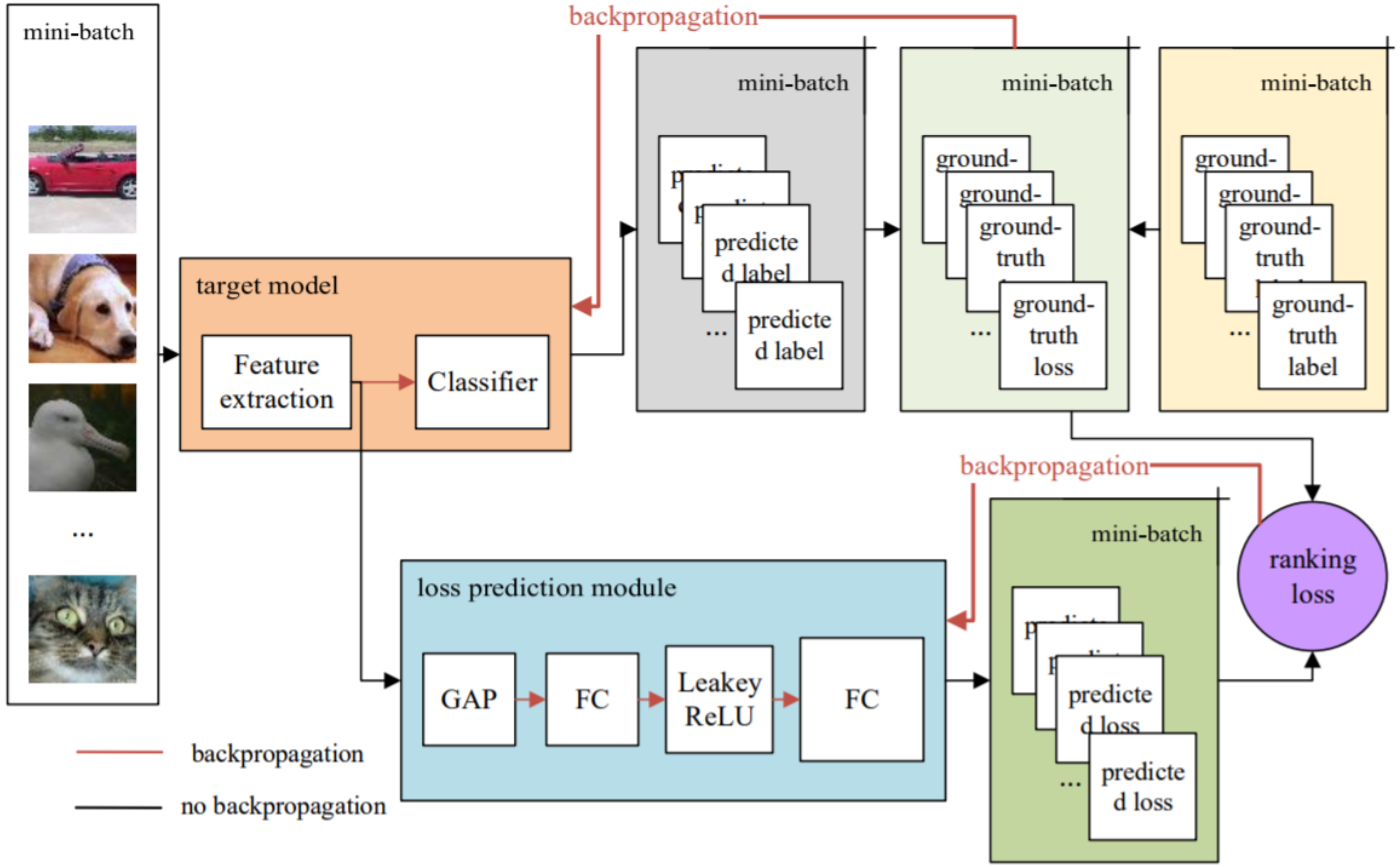
An example of why the module should be trained with rank.



The proposed active learning algorithm

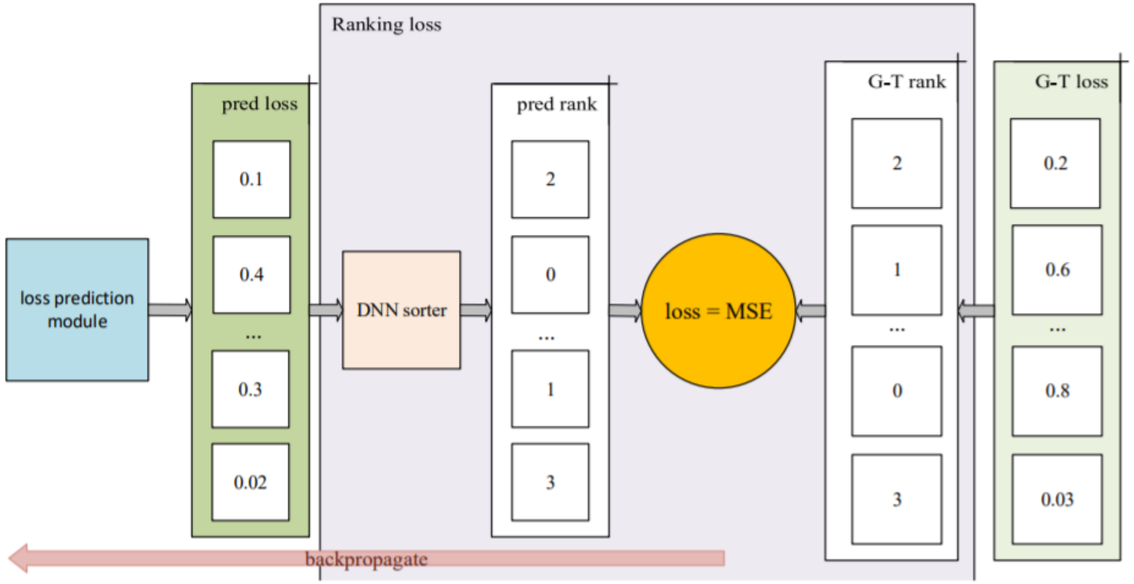
The improvement:

- Use the outputs of last block before the fully-connected layer as the feature
- Stop the ranking loss gradient to the target model and we separate the two losses
- Use a listwise learning to rank (LTR) approach to train the loss prediction module.



Differentiable Listwise LTR Loss

- The listwise approach is difficult in the context of deep learning end-to-end architecures because most of the metrics are not differentiable.
- A pretrained sorter (e.g. bidirectional GRU) is used to convert the predicted losses into predicted ranks. Thus the ranking loss is differentiable, which is equivalent to optimize Spearman's Rank correlation. (See *Engilberge, Martin, et al. "SoDeep: a Sorting Deep net to learn ranking loss surrogates." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019*.)



Differentiable Listwise LTR Loss

1. The Spearman's rank correlation [37] is defined as:

$$r_s = 1 - \frac{6\|rk(l_{\Theta_{tg}}) - rk(\hat{l}_{\Theta_{pre}})\|_2^2}{d(d^2-1)}$$

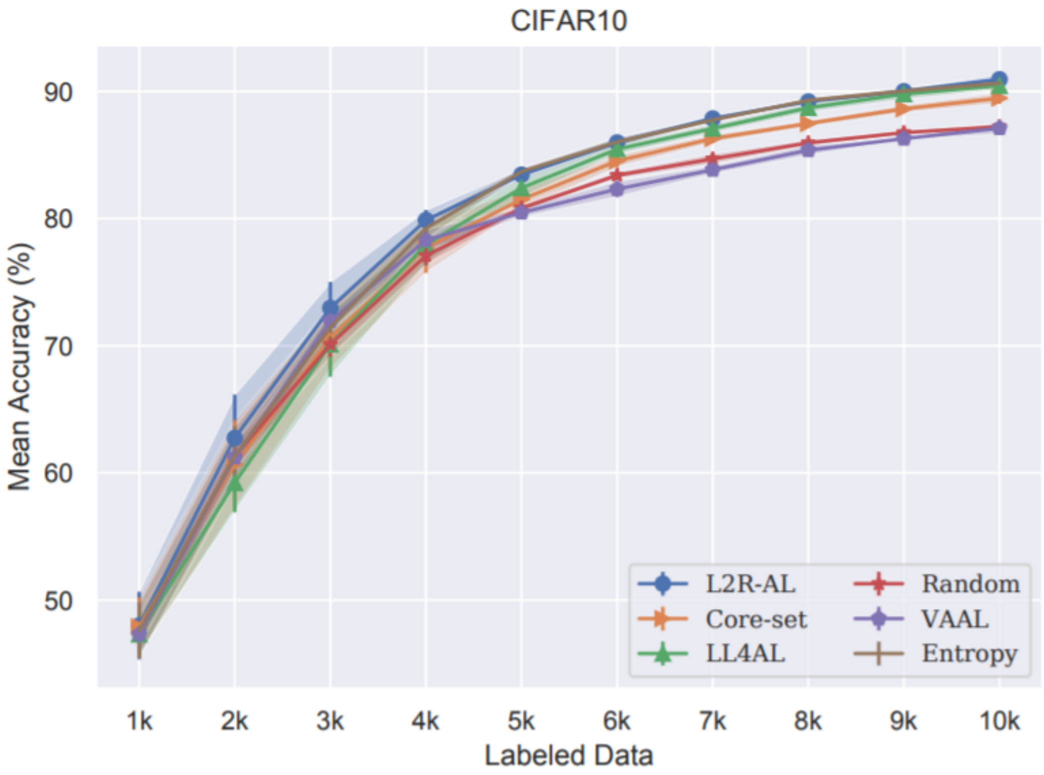
2. The range of this metric is from -1 to 1. If the predicted ranks are same as the ground-truth ranks in every dimension, the value is 1, otherwise -1. Our aim is to maximize the Spearman's rank correlation, this equals to:

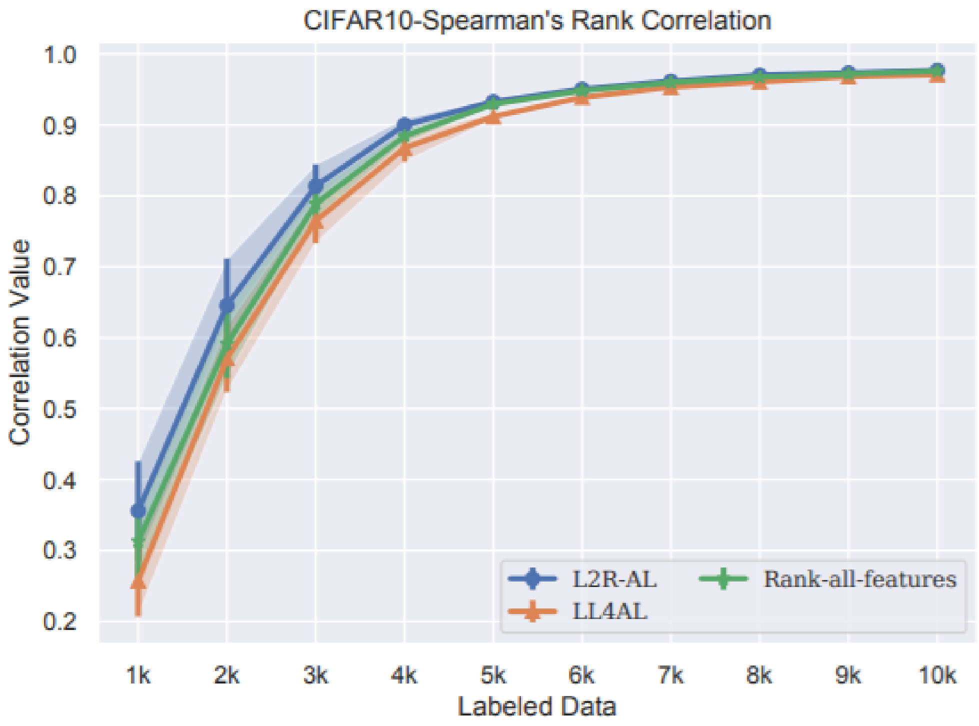
$$\min_{\Theta_{pre}} \|rk(l_{\Theta_{tg}}) - rk(\hat{l}_{\Theta_{pre}})\|_2^2$$

3. For a mini-batch with size d, then the ranking loss of this batch is:

$$loss_{\Theta_{pre}}^{bat} = \frac{1}{d}\sum_{i=1}^{d}(rk(l_i) - f_{\Theta_{sort}}(\hat{l}_i))^2$$

Experiment on CIFAR-10





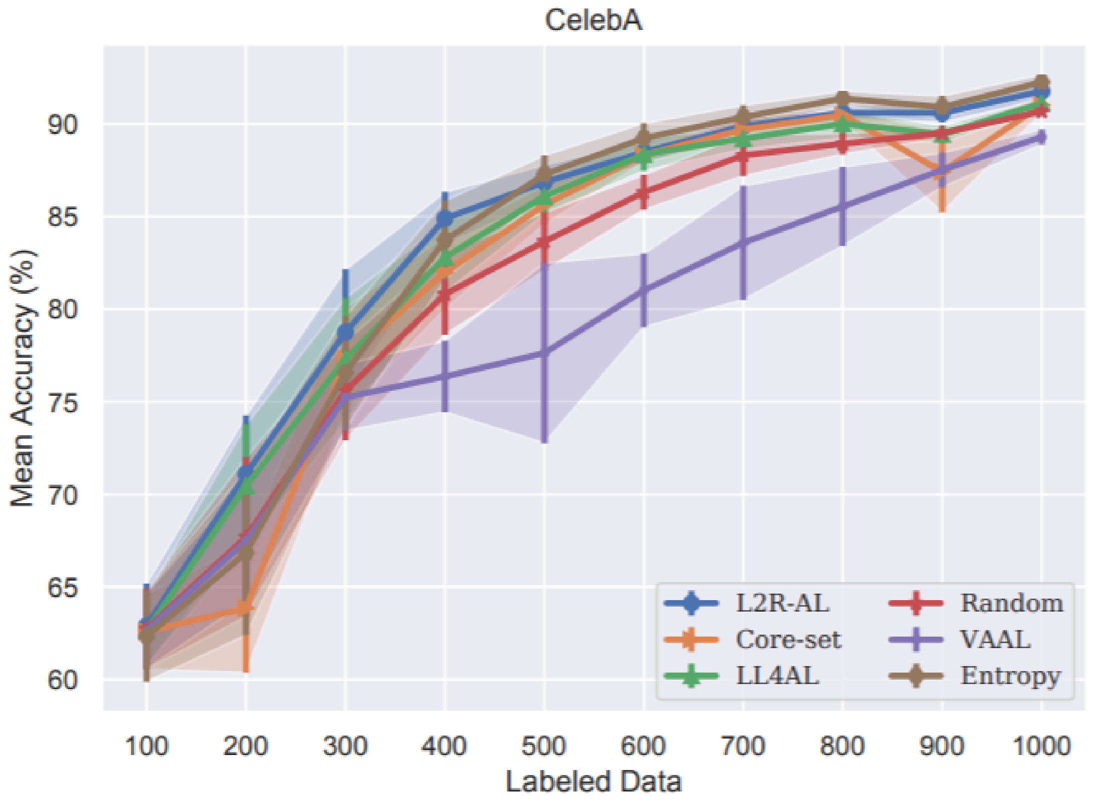
The comparison of Spearman's rank correlation on CIFAR-10

Active learning results of CIFAR-10 image classification

L2R-AL	Core-set	LL4AL	Random	VAAL	Entropy
90.95%	89.47%	90.45%	87.72%	87.11%	90.64%

Experiment on CelebA

We choose the gender classification problem as image classification task using the male/female attribute

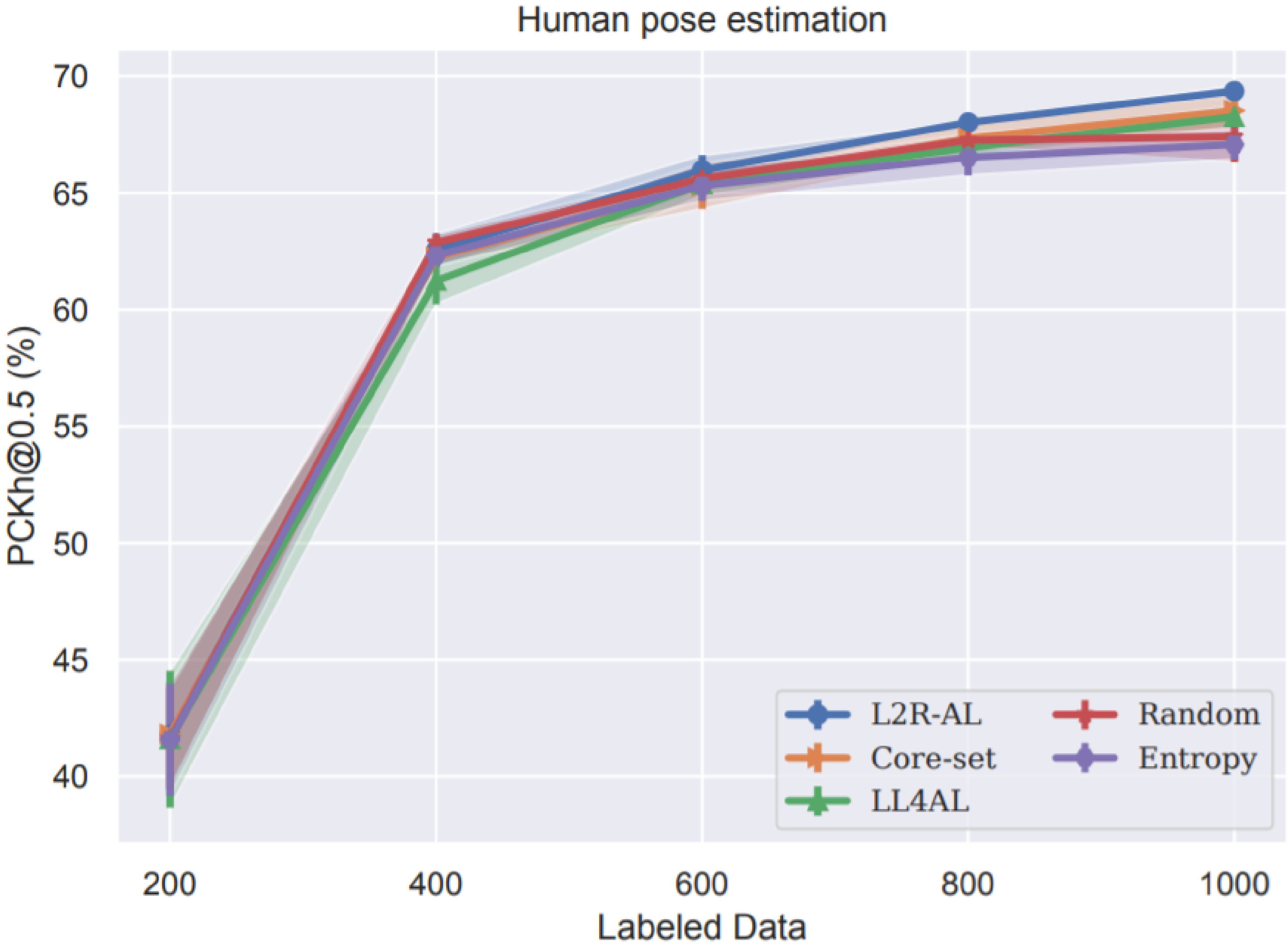


Active learning results of CelebA image classification

L2R-AL	Core-set	LL4AL	Random	VAAL	Entropy
91.76%	91.02%	91.11%	90.67%	89.29%	**92.25%**

Experiment on Human Pose Estimation

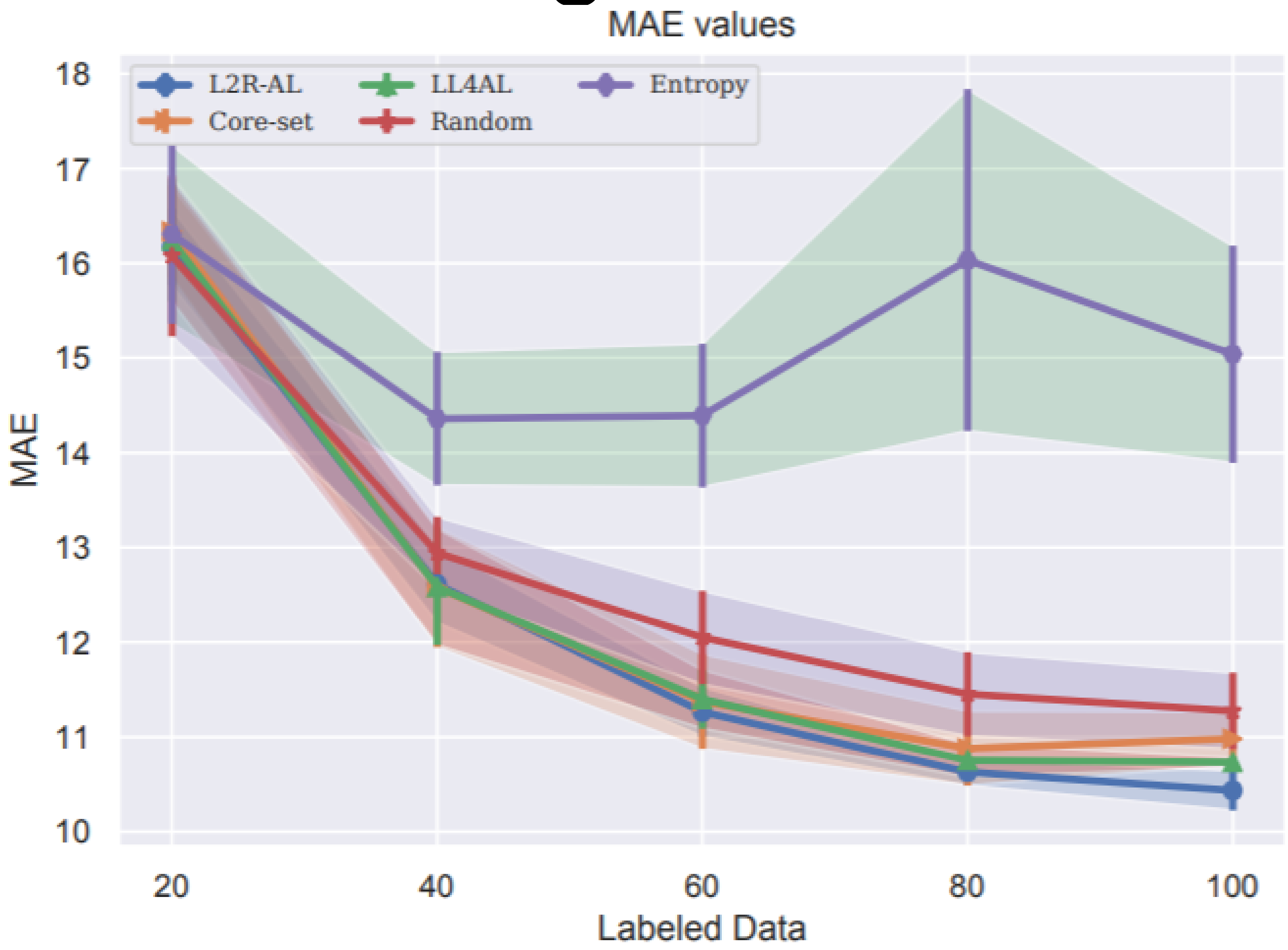
- Regression problem
- Use the MPII dataset



L2R-AL	Core-set	LL4AL	Random	Entropy
69.37%	68.53%	68.27%	67.41%	67.08%

Experiment on Crowd Counting

- Regression problem
- Use the ShanghaiTech Part B dataset



L2R-AL	Core-set	LL4AL	Random	Entropy
10.43	10.98	10.73	11.28	15.04

Conclusion

- This paper demonstrates that learning loss based active learning algorithm actually is a learning to rank problem.
- This paper uses a simple and effective listwise approach to train the loss prediction module by optimizing the Spearman's rank correlation metric.
- Validate the proposed approach on two tasks: image classification (CIFAR-10, CelebA) and regression (MPII, ShanghaiTech Part B). The experimental results show that the proposed algorithm outperforms recent state-of-the-art active learning algorithms.

Thanks for you listening