Task-Aware Variational Adversarial Active Learning

Kwanyoung Kim, Dongwon Park,
Kwang In Kim, Se Young Chun

CVPR - 2021

Bardia Mojra
February 2, 2022
Robotic Vision Lab
University of Texas at Arlington

Introduction

- Active learning (AL)
- Task-agnostic vs task-aware
- Learning Loss Ranker
- TA-VAAL Architecture
- Latent space with global representation
- Experiments

Active Learning

- In deep learning, labeling large amount of data is challenging due to high labeling cost and time required for training
- Active learning (AL) seeks to improve training rate by asking the most informative questions
- Two main directions:
  - Task-Agnostic or distribution-based
  - Task-Aware or model uncertainty-based

Task-Agnostic AL

- Based on distribution of labeled data
- Good for finding clusters and influential points
- Gaussian processes

Task-Aware AL

- Based on conditional distribution of output given labeled data
- Good for finding difficult points
- Good for finding probability in overlapping regions in classification tasks
- Bayesian processes

Learning Loss Ranker

- Predicts relative rankings of losses
- Loss rankings are embedded into the latent space of VAAL with the conditional latent variable r
- Smooth differentiable function with nice convergence properties for optimization
- Motivation:
  - Easier to learned and predict ranked relative attributes than the absolute attribute values
  - For AL, ranking the data points to labels is often sufficient

Learning Loss Ranker

$$L_R(\hat{l}_P, l_P) = -(2/B)\sum_{i=1}^{B/2}\{I_i \log[\sigma(\hat{l}_i - \hat{l}_j)] + (1 - I_i)\log[1 - \sigma(\hat{l}_i - \hat{l}_j)]\}$$

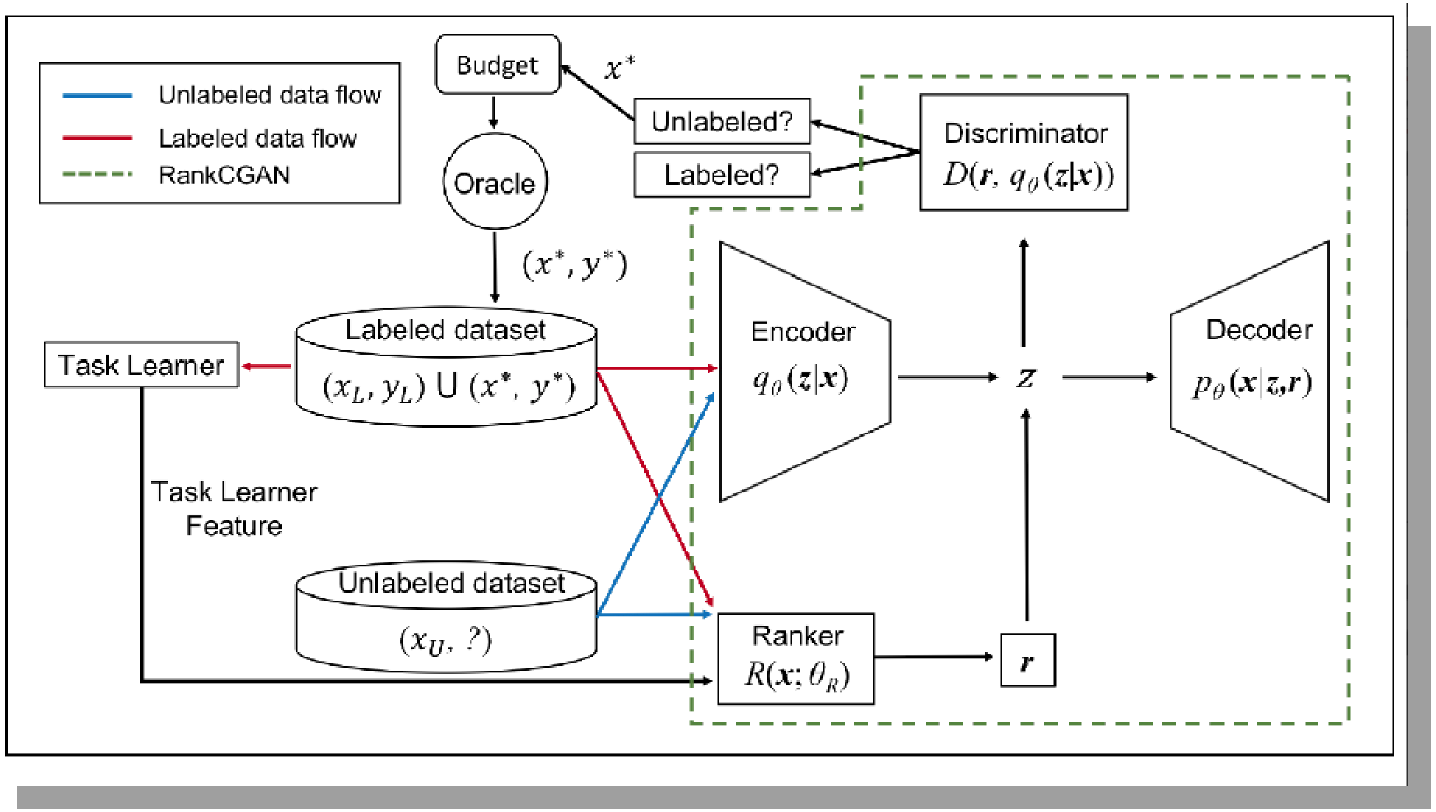
(eq. 1)

$$L_{total} = L_T(\hat{y}_L, y_L) + \eta L_R(R(x_P), l_P)$$

(eq. 2)

TA-VAAL

- R: Ranker
  - assigns r to x_u & x_l
- Z: Latent space var.
  - z_l ~ p_x_l
  - Forms global distribution



Latent Space Global Dist.

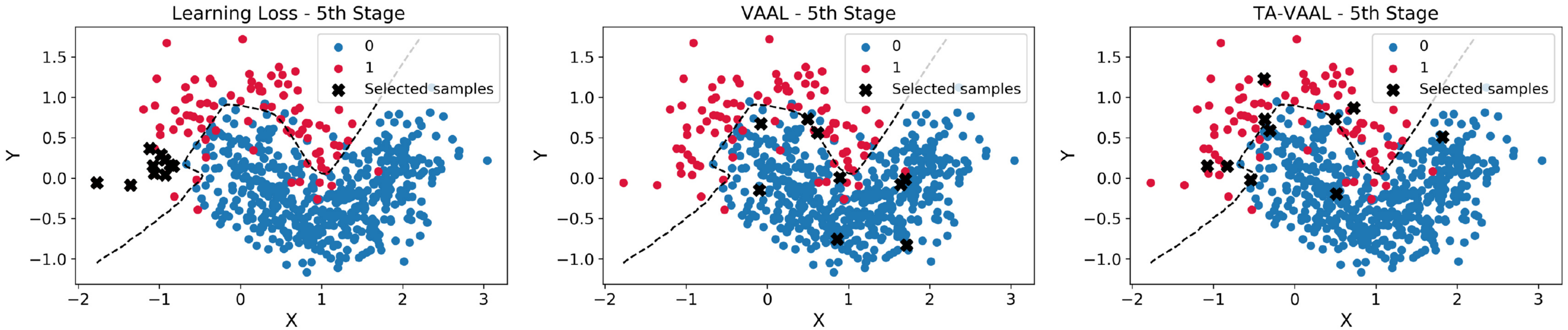
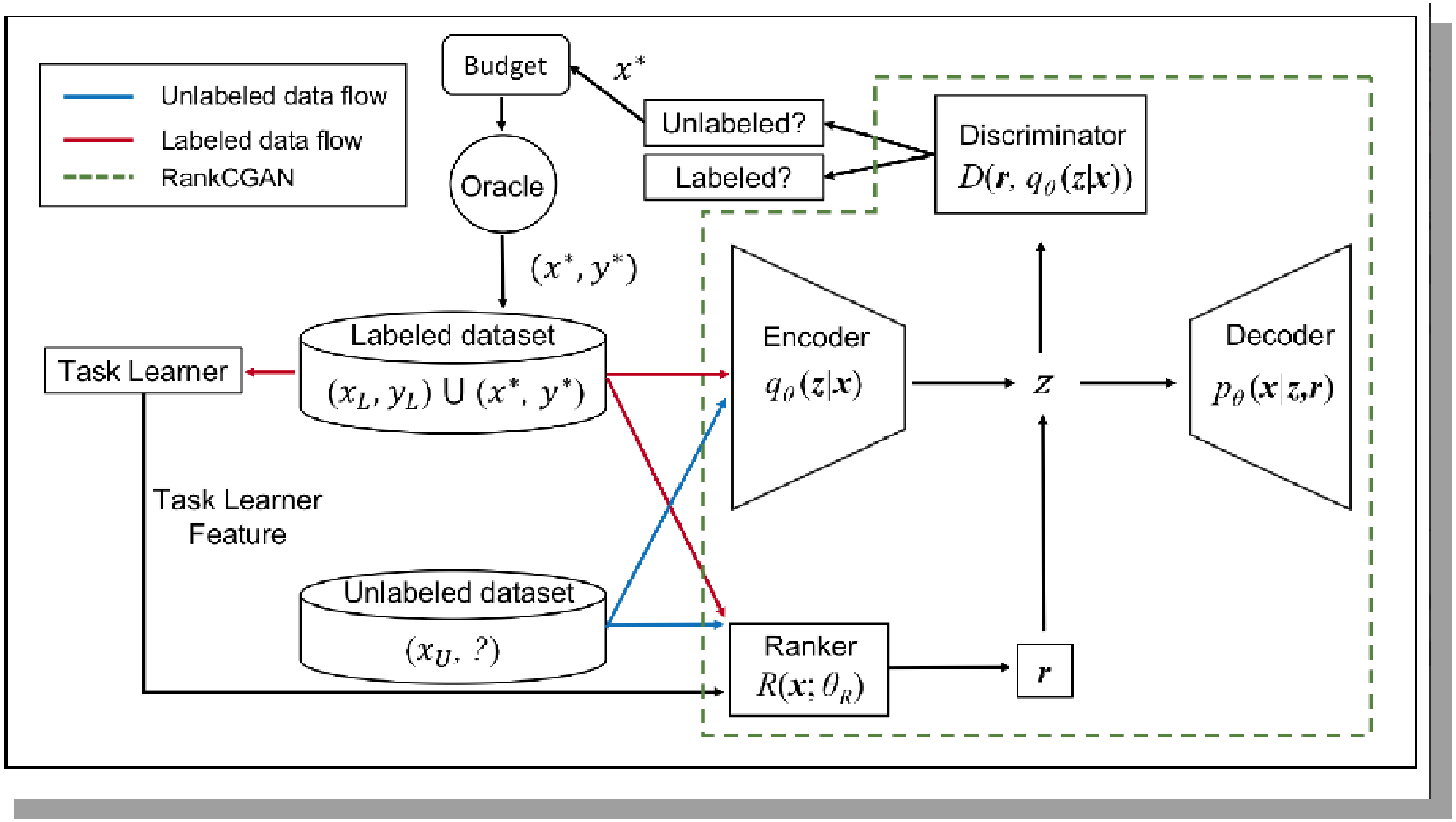


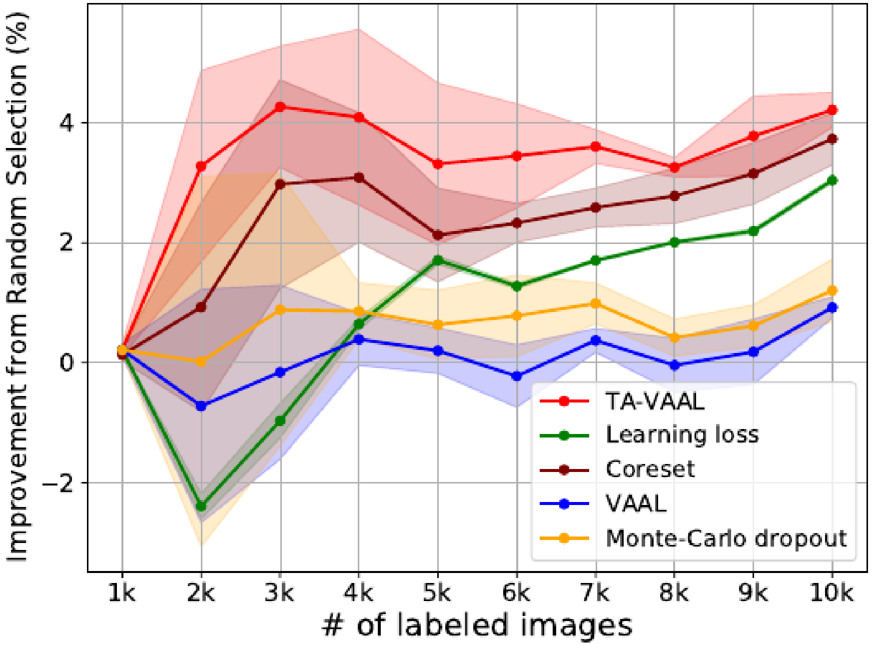
Figure 1: Visual results of active learning methods (Learning loss [40], VAAL [31], our TA-VAAL) on imbalanced toy example at the 5th stage. *Red* and *blue* dots indicate samples assigned to class 0 and 1, respectively. Ten samples at that stage (denoted by *black* cross) were selected using each method. The oracle decision boundary of the model is shown as a black dash line. Learning loss identified difficult samples near the decision boundary and VAAL found influential samples over the entire set. Our TA-VAAL selected samples that are both difficult (near decision boundary) and influential (over the entire set).

TA-VAAL

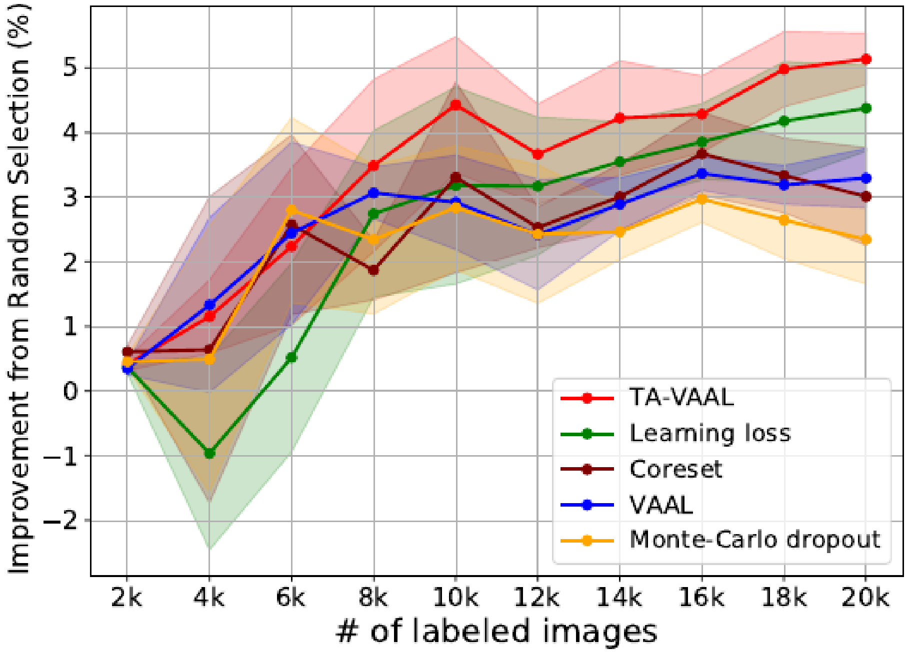


Unlabeled data flow
Labeled data flow
RankCGAN
Budget
x^*
Unlabeled?
Labeled?
Discriminator
$D(\boldsymbol{r}, q_\phi(\boldsymbol{z}|\boldsymbol{x}))$
Oracle
(x^*, y^*)
Labeled dataset
$(x_L, y_L) \cup (x^*, y^*)$
Task Learner
Encoder
$q_\phi(\boldsymbol{z}|\boldsymbol{x})$
z
Decoder
$p_\theta(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{r})$
Task Learner
Feature
Unlabeled dataset
$(x_U, ?)$
Ranker
$R(\boldsymbol{x}; \theta_R)$
$\boldsymbol{r}$

Experiments



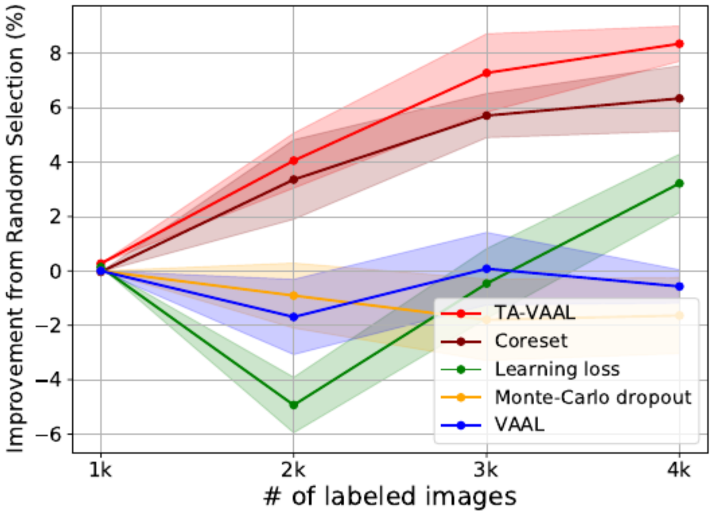
(a) CIFAR10



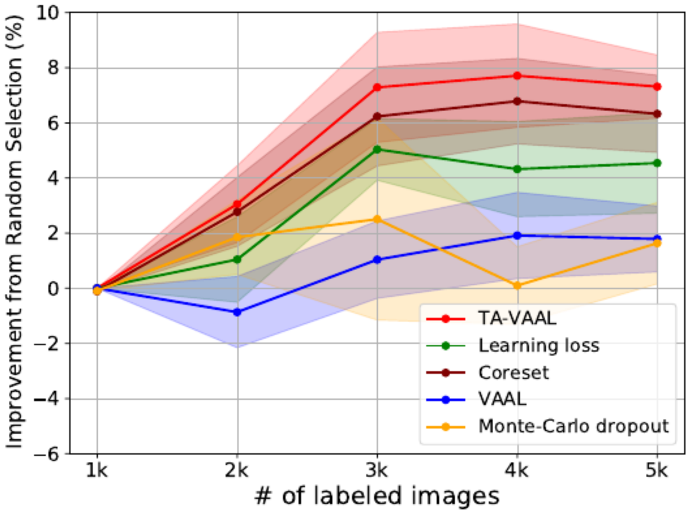
(b) CIFAR100

Figure 3: Mean accuracy improvements with standard deviation (shaded) of AL methods from random sampling baseline over the number of labeled samples. The absolute accuracy values are provided in the supplemental material. Our TA-VAAL outperformed others on (balanced) CIFAR10 in all stages and on (balanced) CIFAR100 after a few stages.

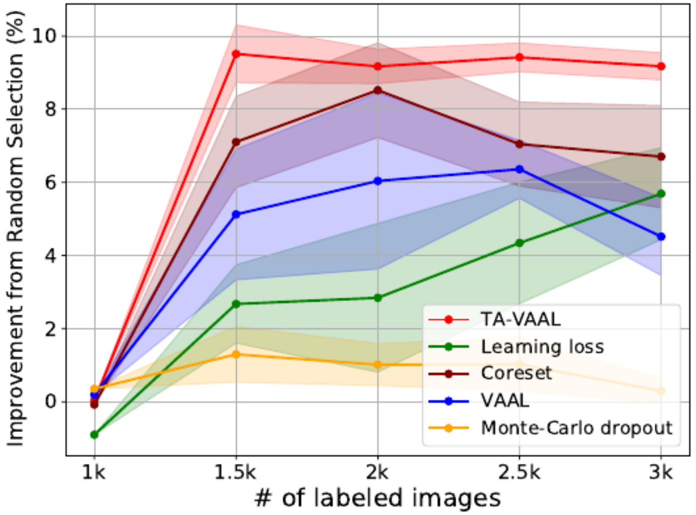
Experiments



(a) Modified CIFAR10 with imbalance ratio ×100



(b) Modified CIFAR10 with imbalance ratio ×10



(c) Caltech101

Figure 4: Mean accuracy improvements with standard deviation (shaded) of AL methods from random sampling baseline over the number of labeled samples on imbalanced datasets. Our TA-VAAL outperformed others on modified CIFAR10 with different imbalance ratios and Caltech101 in all stages.

Experiments

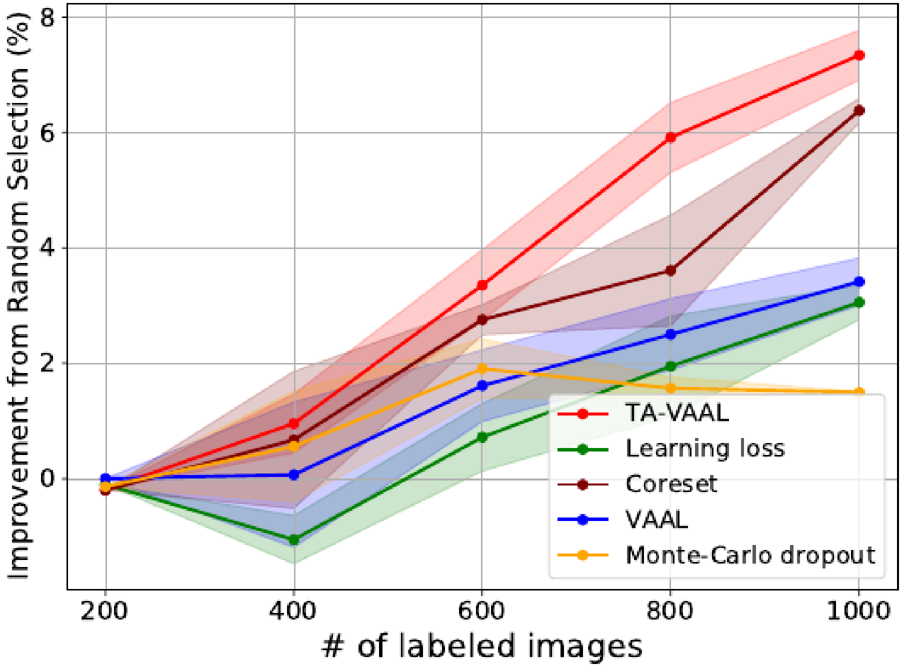


Figure 5: Relative accuracy improvements from random selection for semantic segmentation on Cityscape dataset.

Experiments

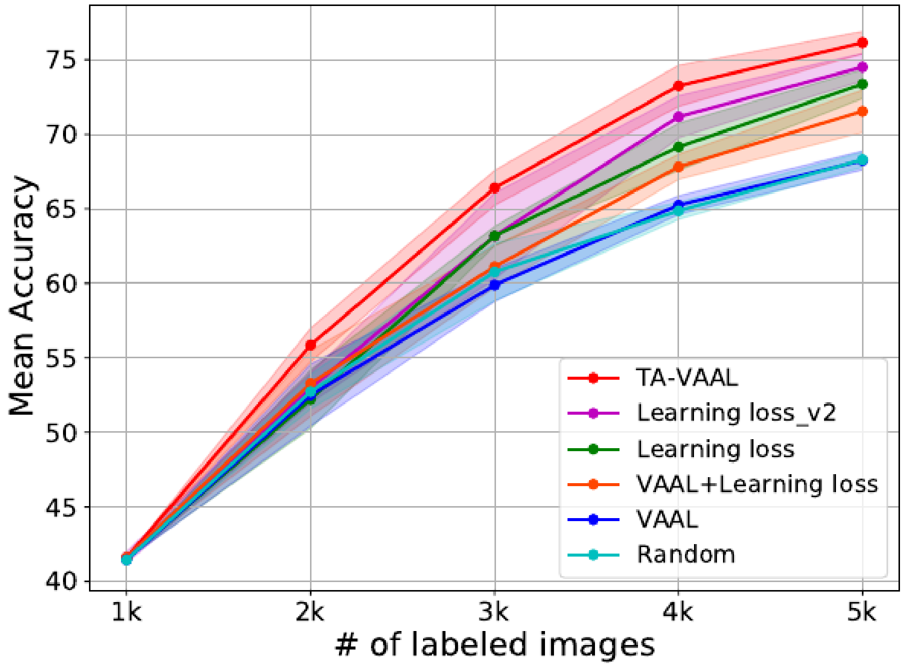
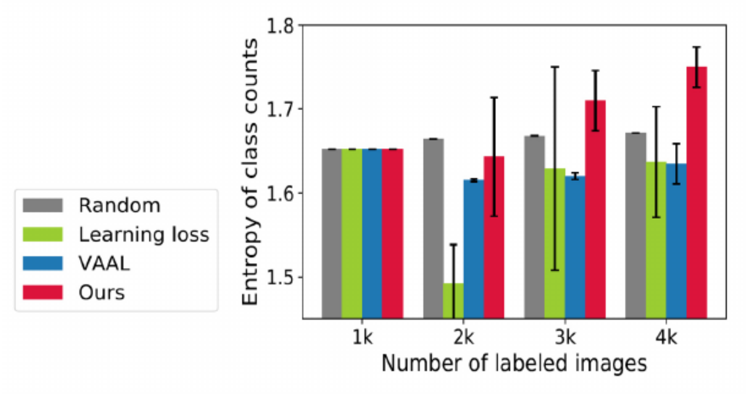
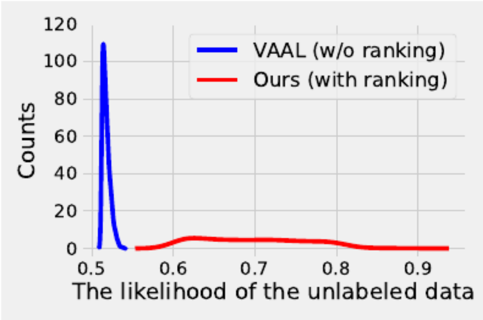


Figure 6: The results of ablation study by selectively removing core components (modified CIFAR10 with imbalanced ratio $\times 10$): Learning loss_v2 is ours without VAAL. VAAL+learning loss is ours with the original learning loss.

Experiments



(a)



(b)

Figure 7: For the modified CIFAR10 with imbalance ratio $\times 100$, (a) Bar graphs of number of labeled images vs. data class count entropy. (b) Likelihood of unlabeled data vs. number of samples at the last stage.

Experiments

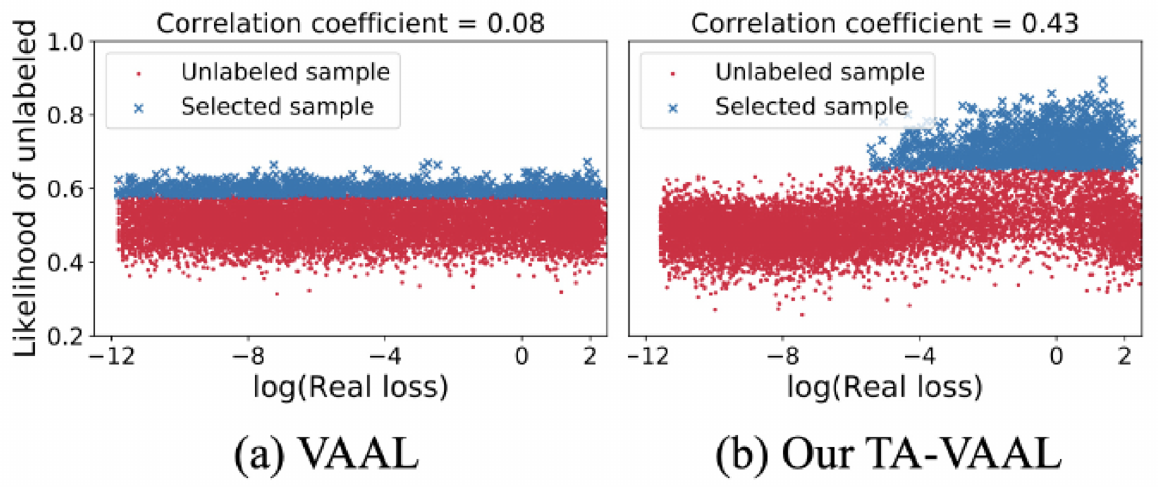


Figure 8: Relationships between real loss (task model uncertainty) and likelihood of data remaining unlabeled (task-agnostic data distribution) in (a) VAAL and (b) our TA-VAAL. We use the model from the last AL stage on imbalanced CIFAR-10. While task-agnostic VAAL selected samples with a wide range of real loss values, our TA-VAAL chose samples with relatively high real loss values.

- Questions!

- Thank you!