南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Agreement-Discrepancy-Selection: Active Learning with Progressive Distribution Alignment

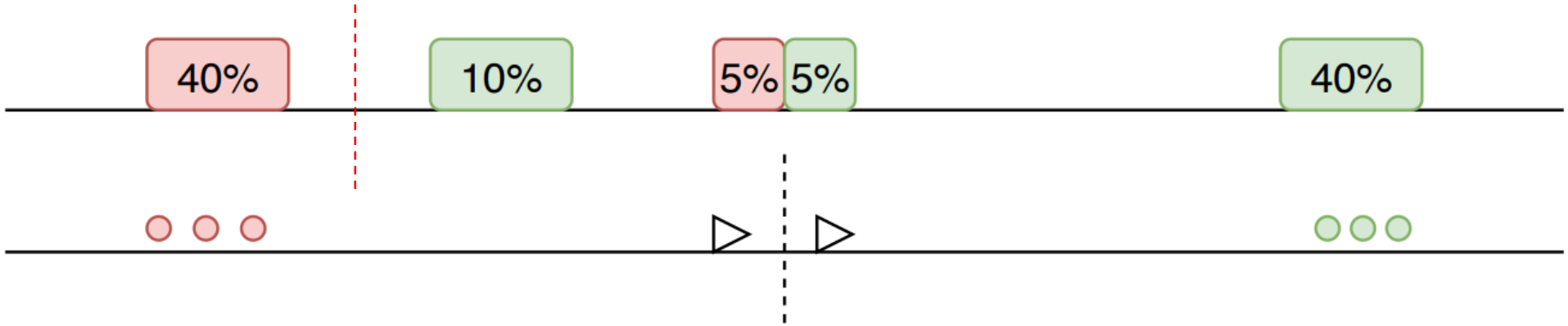
Mengying Fu[†], Tianning Yuan[†], Fang Wan[†*], Songcen Xu[‡], Qixiang Ye[†*]

[†] University of Chinese Academy of Sciences, Beijing, China
[‡] Noah's Ark Lab, Huawei Technologies, Shenzhen, China
{fumengying19, yuantianning19}@mails.ucas.ac.cn, xusongcen@huawei.com, {wanfang, qxye}@ucas.ac.cn

AAAI 2021

Motivation

Sampling bias: the current labeled points do not representative of the underlying distribution.



The initial observations:

Decision boundary:

Query sample:

Contribution

- Proposed an agreement-discrepancy-selection (ADS) approach.
  - Solving the active learning problem by aligning the distributions of unlabeled samples with those of labeled samples in a continuous and progressive fashion.
- Design an entropy-based metric to measure the distribution alignment and discrepancy.

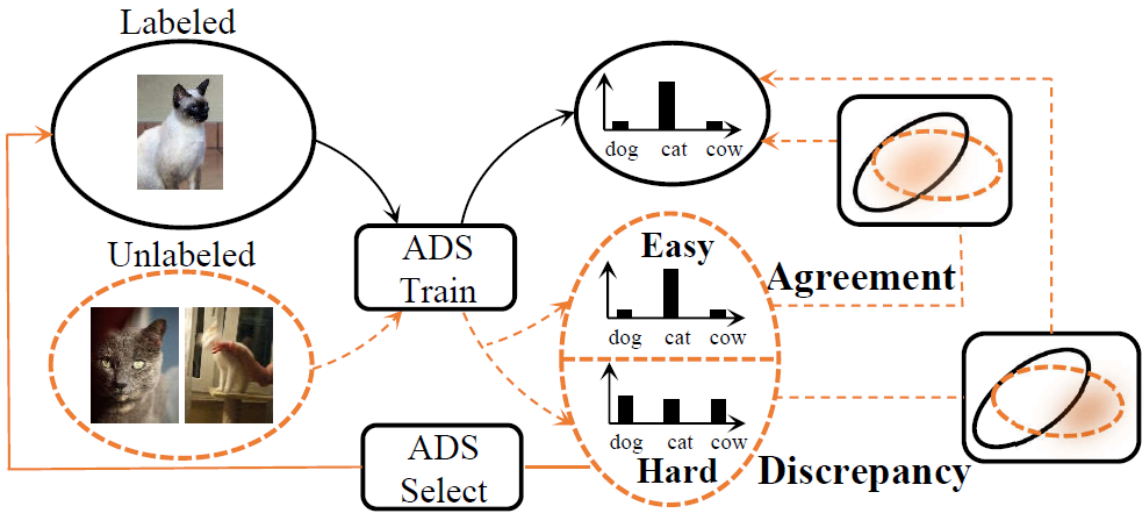
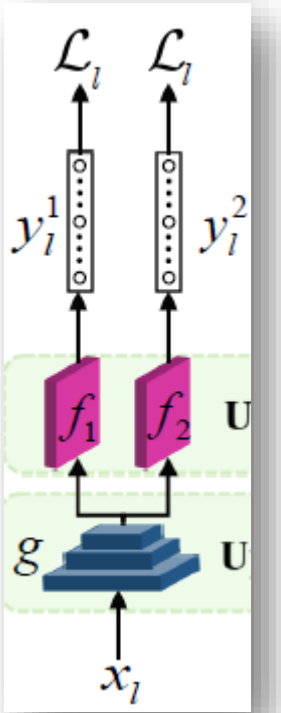


Figure 1: Overview of ADS, which leverages the prediction agreement and discrepancy to select informative unlabeled samples.

Method

➢ Agreement-discrepancy-selection (ADS) flowchart



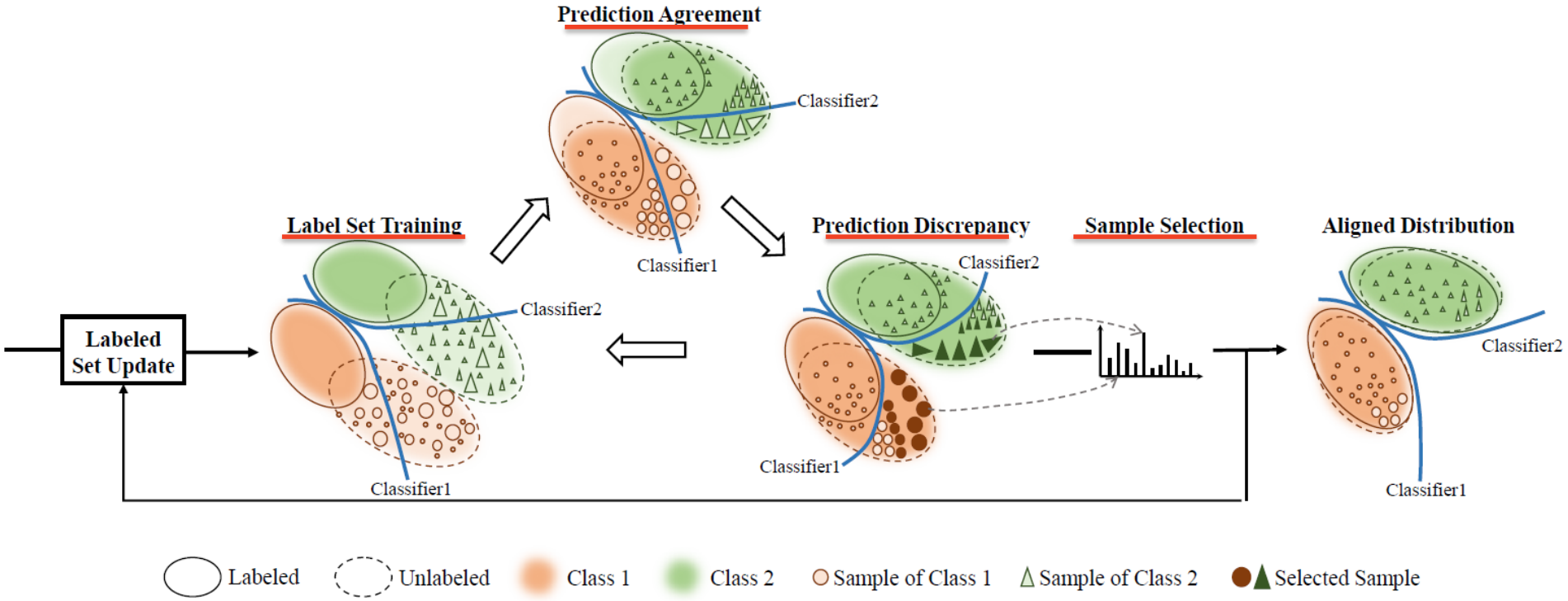
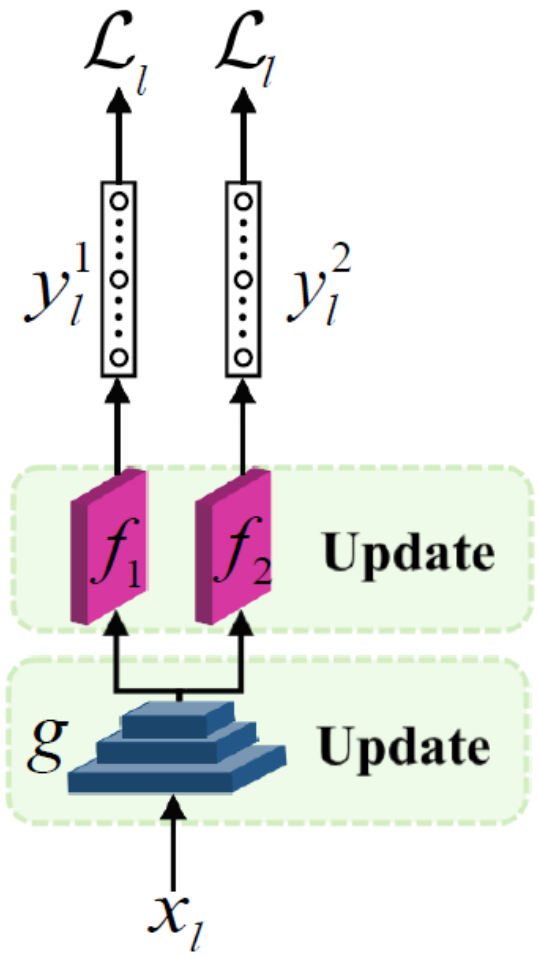


Figure 2: ADS flowchart.The prediction agreement step "pulls" the distributions of labeled and unlabeled samples with low and mid entropy together by updating features while the prediction discrepancy step "push" the distribution of unlabeled samples with high and middle entropy out of the alignment area by updating classifiers. Iterative agreement-discrepancy progressively aligns distributions of unlabeled samples with those of labeled samples. Larger circles/triangles denote more informative samples with larger entropy.

Method

- Label Set Training



$$\underset{\theta_g,\theta_{f_1},\theta_{f_2}}{\operatorname{argmin}} \mathcal{L}_l(\hat{y}_l^1, \hat{y}_l^2)$$

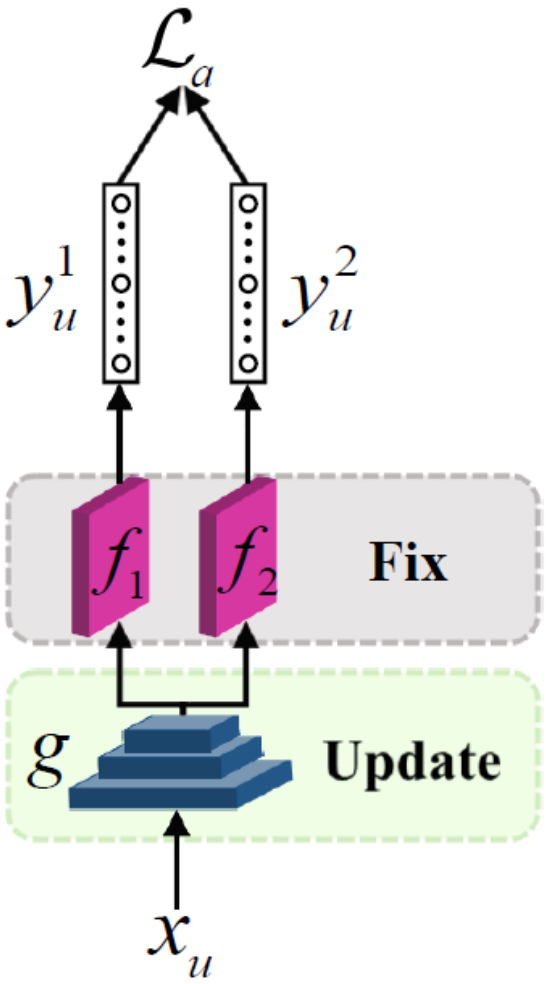
$$= -\sum_{c=1}^{C}(y_{l,c} log \hat{y}_{l,c}^1 + (1 - y_{l,c}) log(1 - \hat{y}_{l,c}^1)$$

$$+ y_{l,c} log \hat{y}_{l,c}^2 + (1 - y_{l,c}) log(1 - \hat{y}_{l,c}^2)).$$

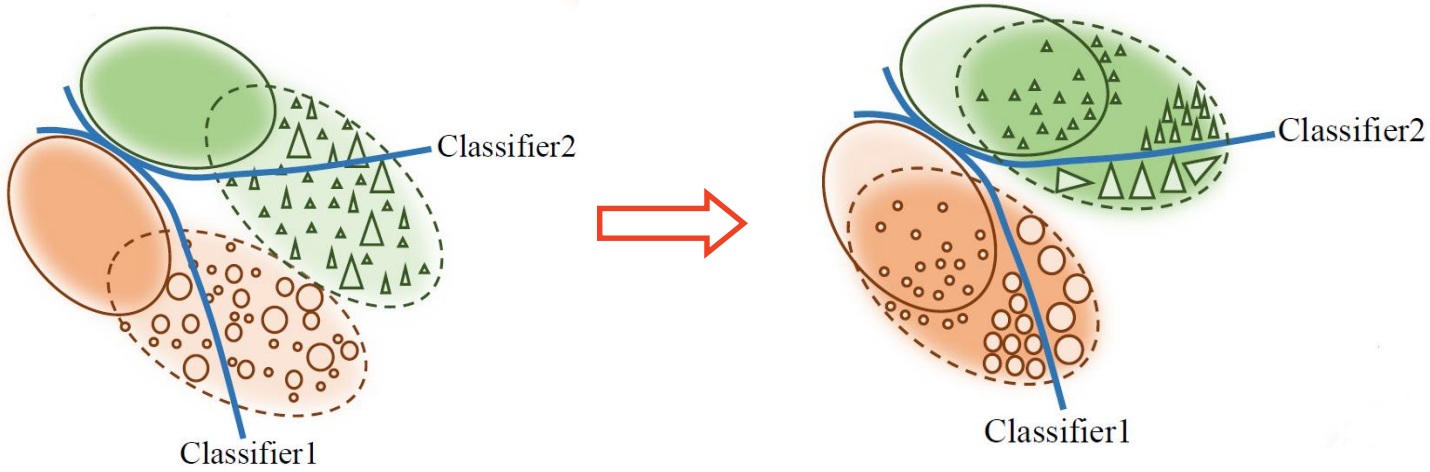
Method

南京航空航天大学 Nanjing University of Aeronautics and Astronautics

➢ Prediction Agreement: Distribution Alignment



By forcing the two classifiers to agree with each other on predictions, the feature representation is updated so that some unlabeled samples can "move" towards L in the feature space.



Method

➢ Prediction Agreement: Distribution Alignment

✓ An entropy-based metric is proposed to qualify the prediction alignment.

$$E(u) = -\frac{1}{2}(\sum_{c=1}^{C} \hat{y}_{u,c}^{1} log \hat{y}_{u,c}^{1} + \sum_{c=1}^{C} \hat{y}_{u,c}^{2} log \hat{y}_{u,c}^{2}),$$

✓ A calibration weight is designed to differentiate the samples with large entropy from those with small entropy.

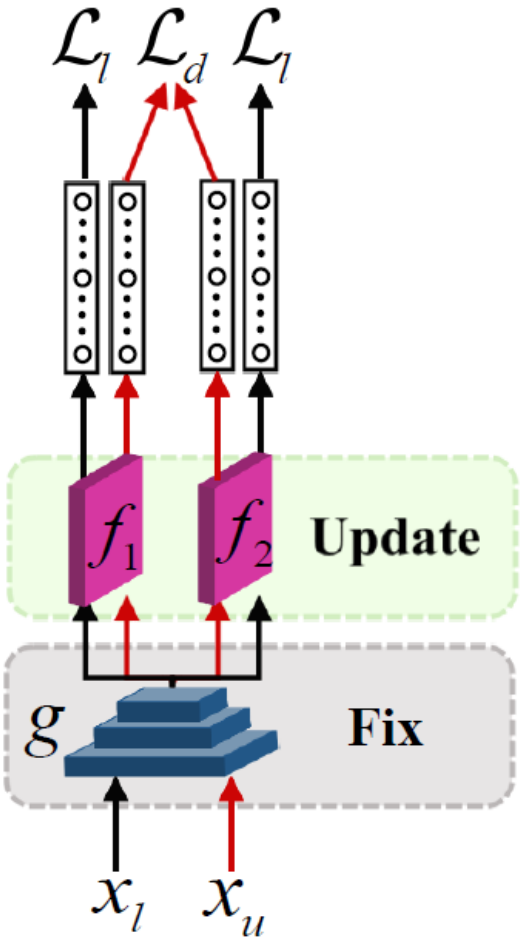
$$w_a = \frac{1}{n}\left(1 - \delta(E(u) - \tau)\right) \qquad \delta(x) = \frac{1}{1+e^{-x}}$$

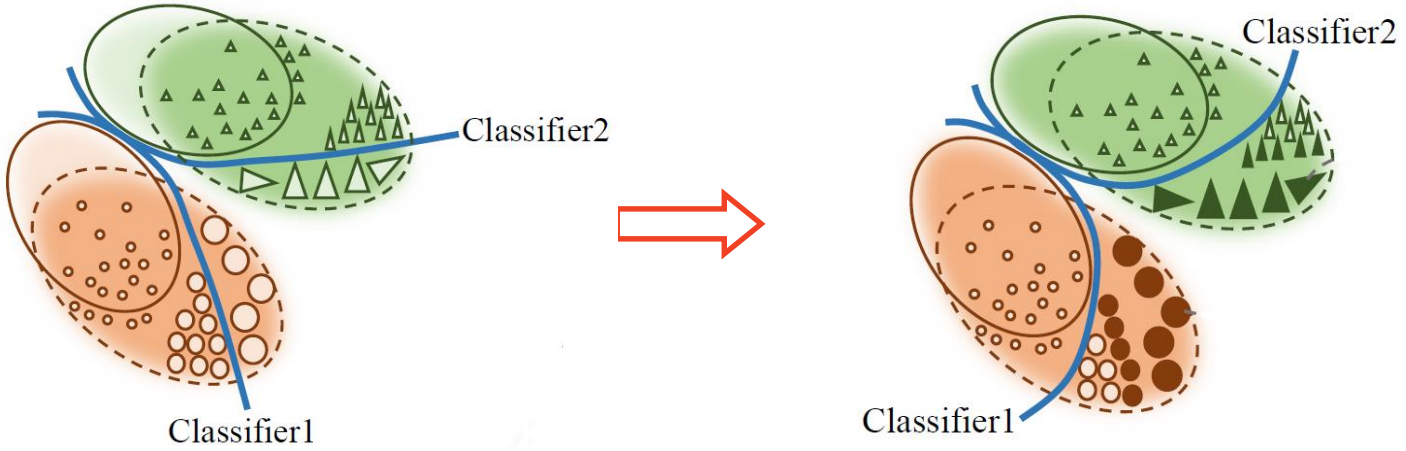
✓ Loss:

$$\underset{\theta_g}{\operatorname{argmin}}\, \mathcal{L}_a = w_a \mathcal{L}_{a'}(\hat{y}_u^1, \hat{y}_u^2). \qquad \underset{\theta_g}{\operatorname{argmin}}\, \mathcal{L}_{a'}(\hat{y}_u^1, \hat{y}_u^2) = \frac{1}{C}\sum_{c=1}^{C} \left|\hat{y}_{u,c}^1 - \hat{y}_{u,c}^2\right|$$

Method

➢ Prediction Discrepancy: Highlighting Informative Samples





$$w_d = \frac{1}{n}(\delta(E(u) - \tau)) \qquad \underset{\theta_{f_1},\theta_{f_2}}{\operatorname{argmin}} \mathcal{L}_{d'}(\hat{y}_u^1, \hat{y}_u^2) = 1 - \frac{1}{C}\sum_{c=1}^{C} |\hat{y}_{u,c}^1 - \hat{y}_{u,c}^2|$$

$$\underset{\theta_{f_1},\theta_{f_2}}{\operatorname{argmin}} \mathcal{L}_d = w_d \mathcal{L}_{d'}(\hat{y}_u^1, \hat{y}_u^2)$$

Method

➢ Entropy-based Sample Selection

$$E(u) = -\frac{1}{2}(\sum_{c=1}^{C} \hat{y}_{u,c}^{1} log \hat{y}_{u,c}^{1} + \sum_{c=1}^{C} \hat{y}_{u,c}^{2} log \hat{y}_{u,c}^{2}),$$

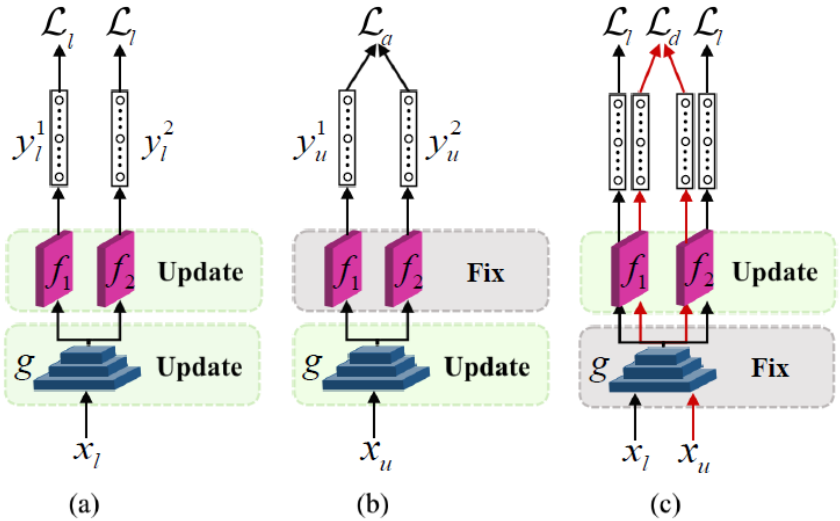


Figure 3: Network architectures. (a) Label set training. (b) Prediction agreement. (c) Prediction discrepancy.

Algorithm 1: ADS Training Procedure

Require: Network parameters θ_g, classifiers' parameters θ_{f_1} and θ_{f_2}, labeled set L and unlabeled set U.
for *iteration* **do**
  for *epoch* **do**
    if *epoch* == *0* **then**
      Training on L using Eq. 1;
    Compute the calibration weight w_a;
    Maximize prediction agreement upon U using Eq. 5;
    Compute the calibration weight w_d;
    Maximize prediction discrepancy on U and L using Eq. 8 and Eq. 1;
    Training on L using Eq. 1;
  Select samples using the entropy metric, Eq. 3;
  update L and U.

Experiment

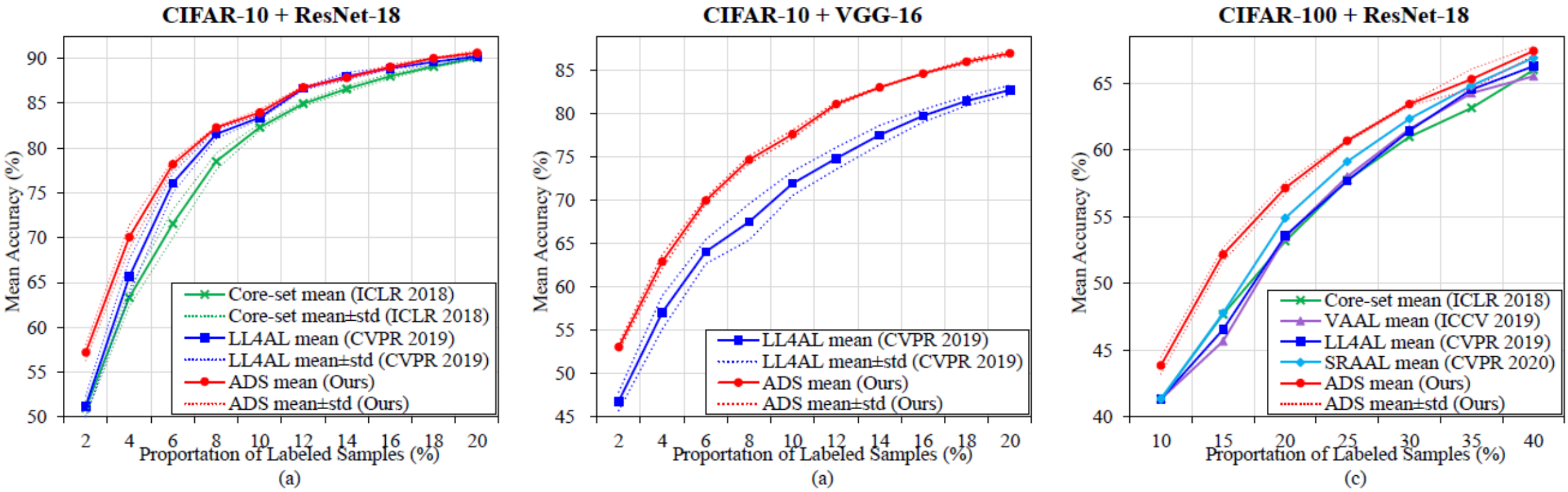


Figure 5: Comparison of ADS with Core-set (Sener and Savarese 2018), VAAL (Sinha, Ebrahimi, and Darrell 2019), LL4AL (Yoo and Kweon 2019) and SRAAL (Zhang et al. 2020): (a) on CIFAR-10 using the ResNet-18 backbone, (b) on CIFAR-10 using the VGG-16 backbone, (c) on CIFAR-100 using the ResNet-18 backbone.

LL4AL: Learning Loss for Active Learning
SRAAL: State-Relabeling Adversarial Active Learning

VAAL: Variational Adversarial Active Learning

Experiment

Table 1: Module evaluation on CIFAR-10 using ResNet-18. "Non", "Ent.(w.)" and "Cal." respectively denote ADS without entropy weight, with entropy weight and calibration weight. "Ent.(sel.)" denotes ADS using the entropy metric to select samples.

ADS				Accuracy (%) on Proportion (%) of Labeled Samples									
Non	Ent. (w.)	Cal.	Ent. (sel.)	2	4	6	8	10	12	14	16	18	20
				51.01	61.48	69.14	75.14	79.77	82.83	84.77	85.78	86.89	87.27
✓				**58.07**	67.75	74.91	78.88	80.96	83.23	84.66	85.29	86.50	87.24
✓			✓	54.28	66.23	74.61	80.18	82.89	85.99	**88.00**	88.86	89.86	90.41
	✓		✓	55.43	67.21	75.49	80.08	83.46	85.40	87.13	88.55	89.72	90.02
		✓	✓	**57.22**	**70.08**	**78.18**	**82.30**	**83.97**	**86.78**	87.82	**89.05**	**90.03**	**90.63**

Table 2: Comparison of prediction alignment metrics on CIFAR-10. "ADS(non/max/min/mean)" respectively denote ADS without entropy weight, with the max entropy weight, the min entropy weight, the mean entropy weight of the two classifiers.

Metric	Accuracy (%) on Proportion (%) of Labeled Samples									
	2	4	6	8	10	12	14	16	18	20
Baseline	51.01	61.48	69.14	75.14	79.77	82.83	84.77	85.78	86.89	87.27
ADS (non)	54.28	66.23	74.61	80.18	82.89	85.99	88.00	88.86	89.86	90.41
ADS (max)	54.73	66.51	74.7	79.89	83.13	85.03	86.56	88.29	89.31	89.87
ADS (min)	54.31	65.61	73.87	79.65	82.89	**85.49**	86.85	88.36	89.41	**90.04**
ADS (mean)	**55.43**	**67.21**	**75.49**	**80.08**	**83.46**	85.40	**87.13**	**88.55**	**89.72**	90.02

Experiment

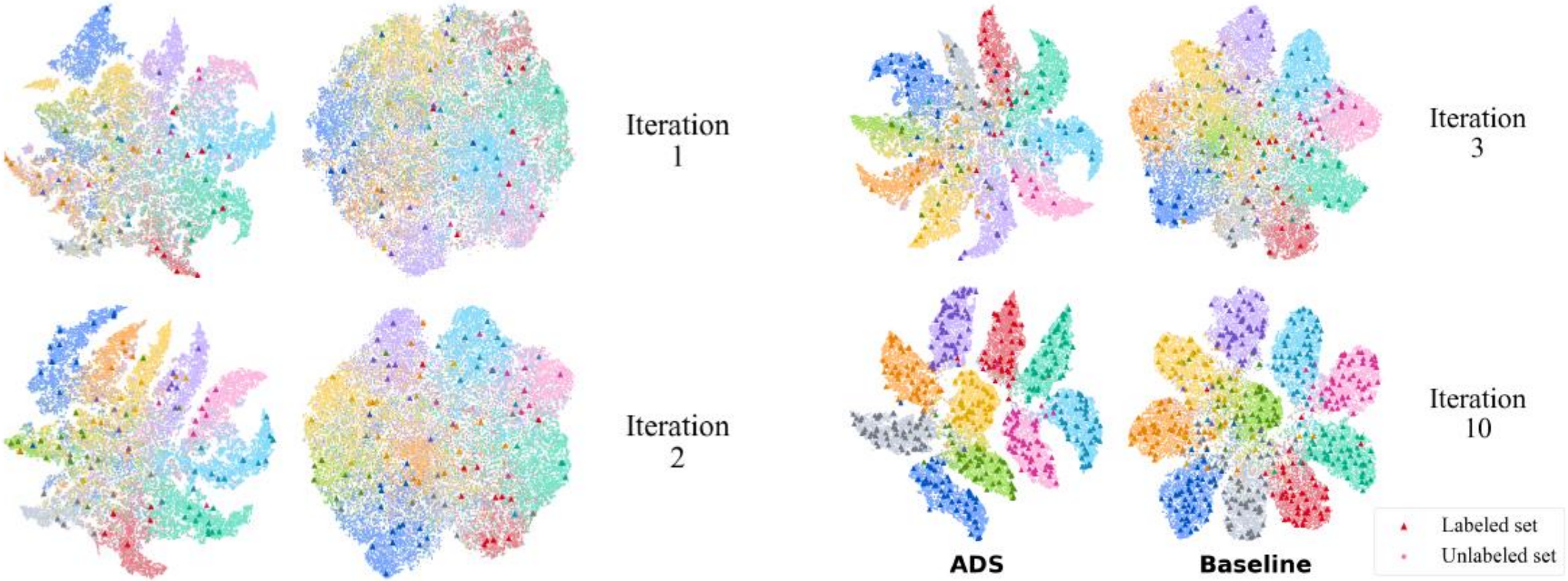


Figure 4: T-SNE visualization of sample distributions. "Triangles" and "dots" respectively denote labeled and unlabeled samples. The "triangles" are progressively aligned with "dots", showing that ADS aligns the distributions of unlabeled samples with those of labeled samples while training discriminative models. The baseline method uses a single classifier to randomly select samples. (Best viewed in color with zoom).

Experiment

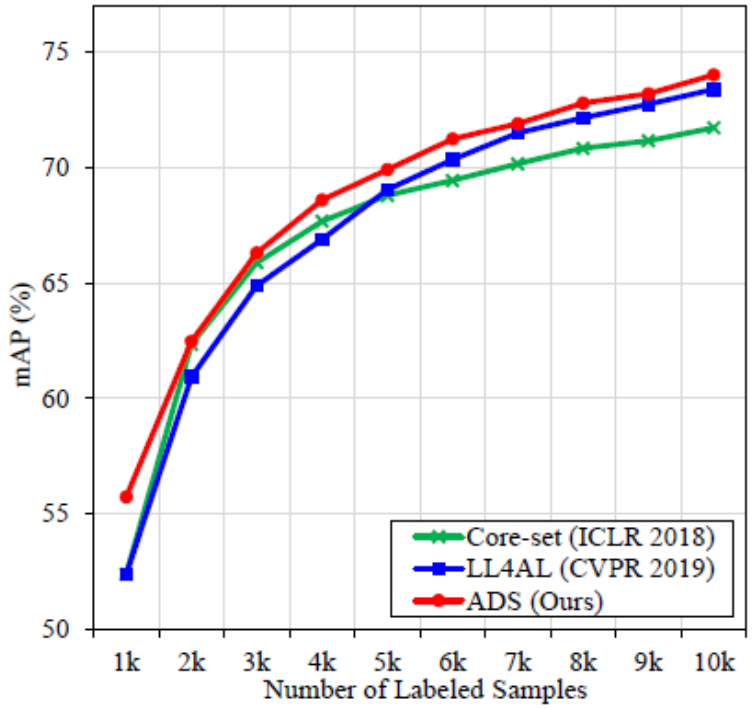


Figure 6: Comparison of ADS with Core-set (Sener and Savarese 2018) and LL4AL (Yoo and Kweon 2019) on PASCAL VOC using the VGG-16 backbone.

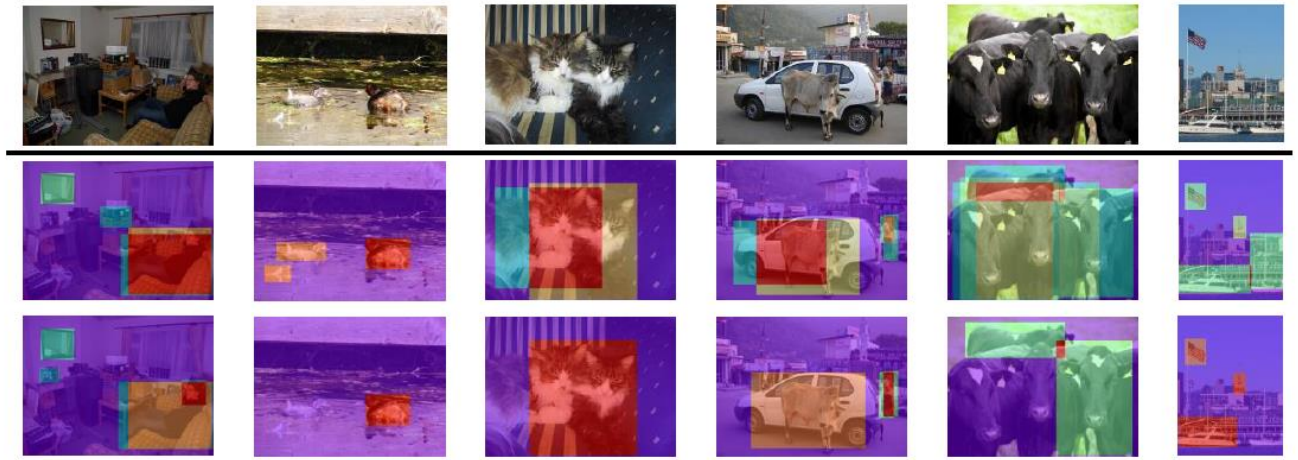


Figure 7: Visualization of object predictions of the two adversarial classifiers. The first row shows the original images, the second and the third rows show the predictions of classifier 1 and classifier 2 respectively. Redder colors indicate higher scores. (Best viewed in color with zoom)

THANKS