ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON COMPUTER VISION

WWW.ECCV2020.EU

Contextual Diversity for Active Learning

Sharat Agarwal*, Himanshu Arora*, Saket Anand, Chetan Arora




INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

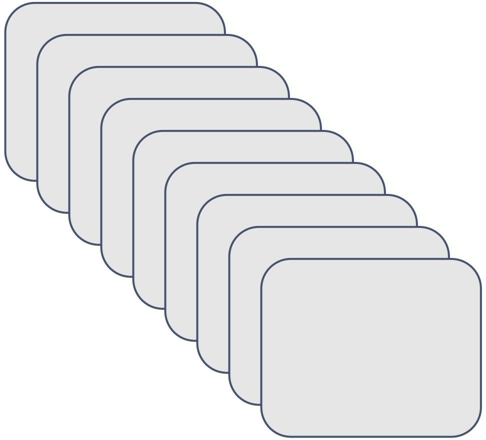
INDIAN INSTITUTE OF TECHNOLOGY DELHI

* equal contribution

Introduction

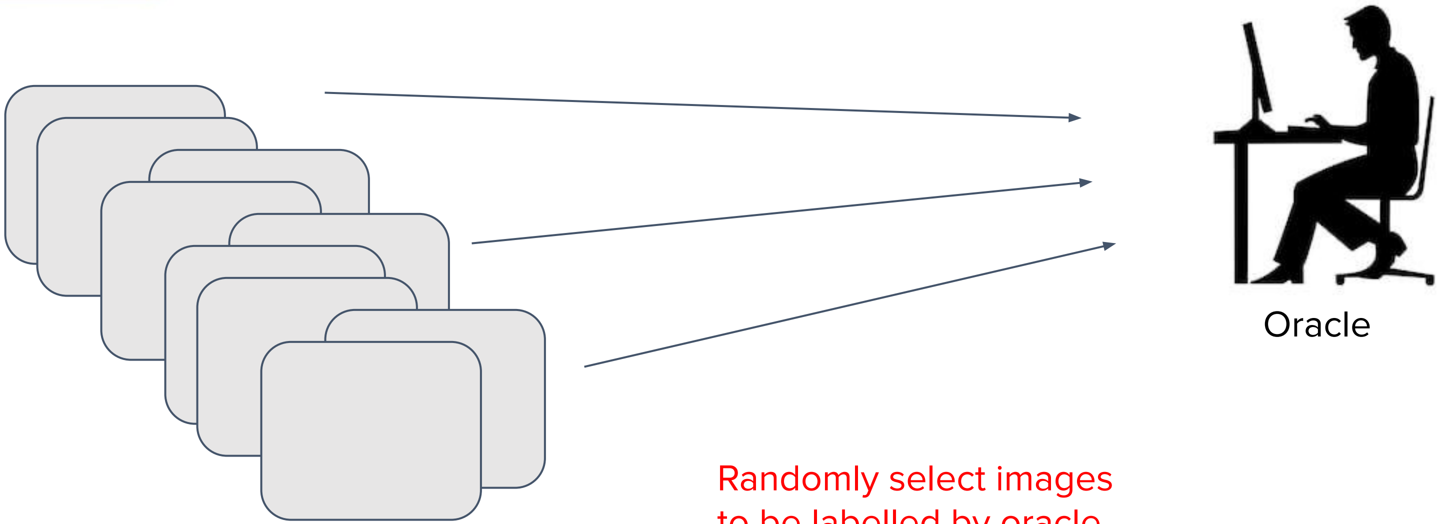
Introduction

Step by Step Active Learning



Unlabelled pool
of data

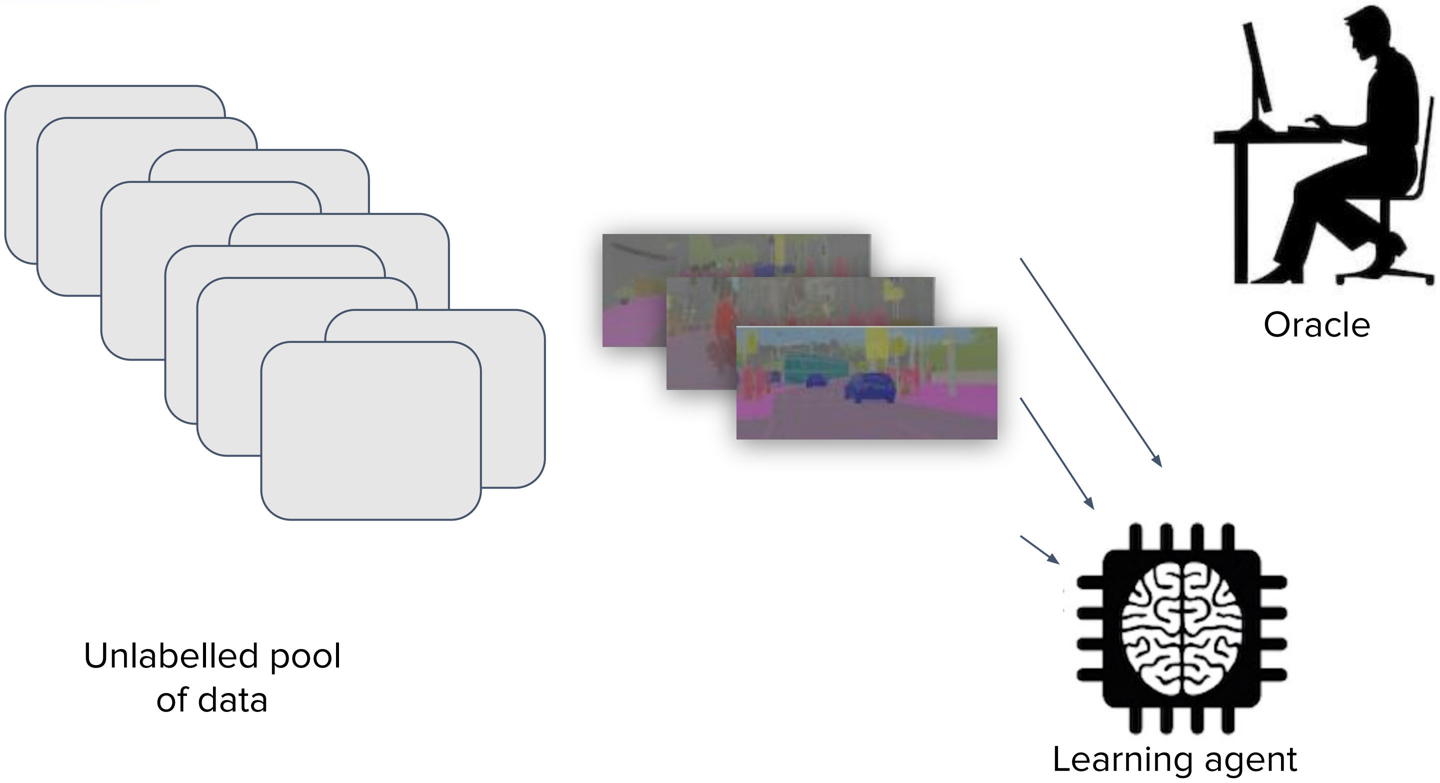
Step by Step Active Learning



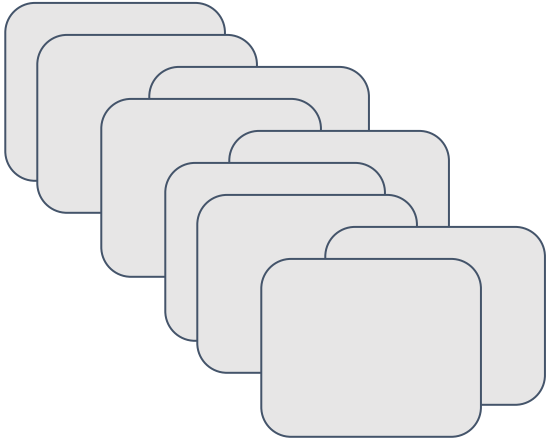
Randomly select images
to be labelled by oracle.

Unlabelled pool
of data

Step by Step Active Learning



Step by Step Active Learning



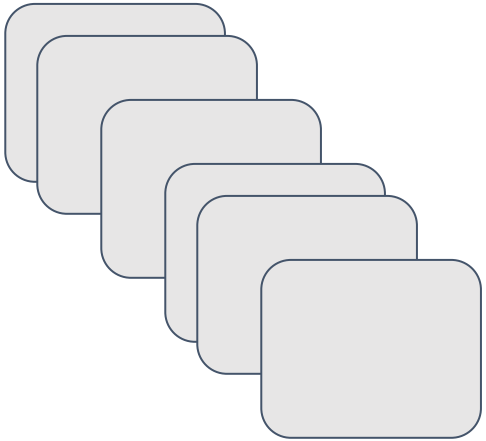
Unlabelled pool
of data

Oracle

Learning agent

60% accuracy

Step by Step Active Learning



Oracle

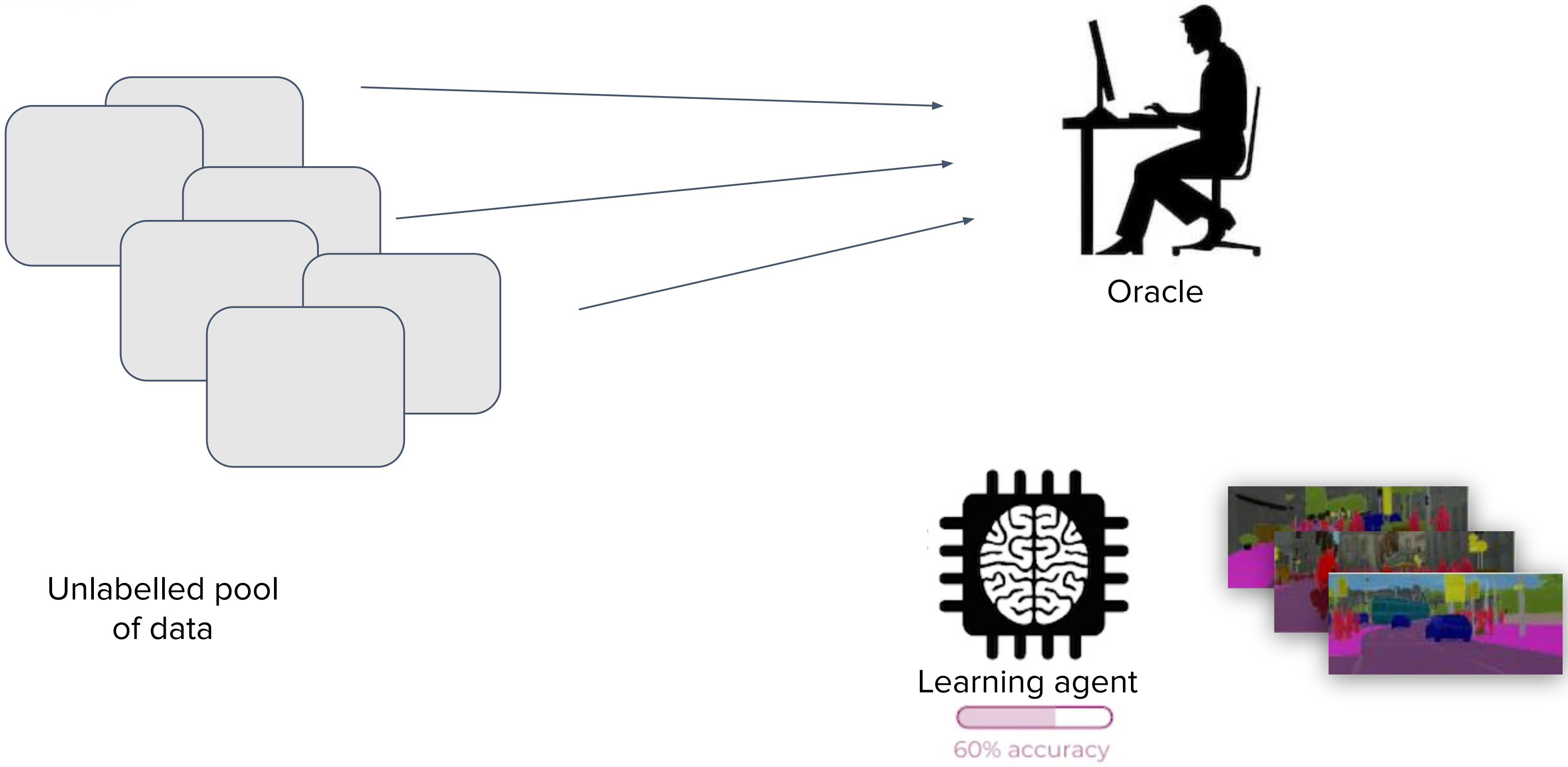
Unlabelled pool
of data

Infer labels to the
remaining data

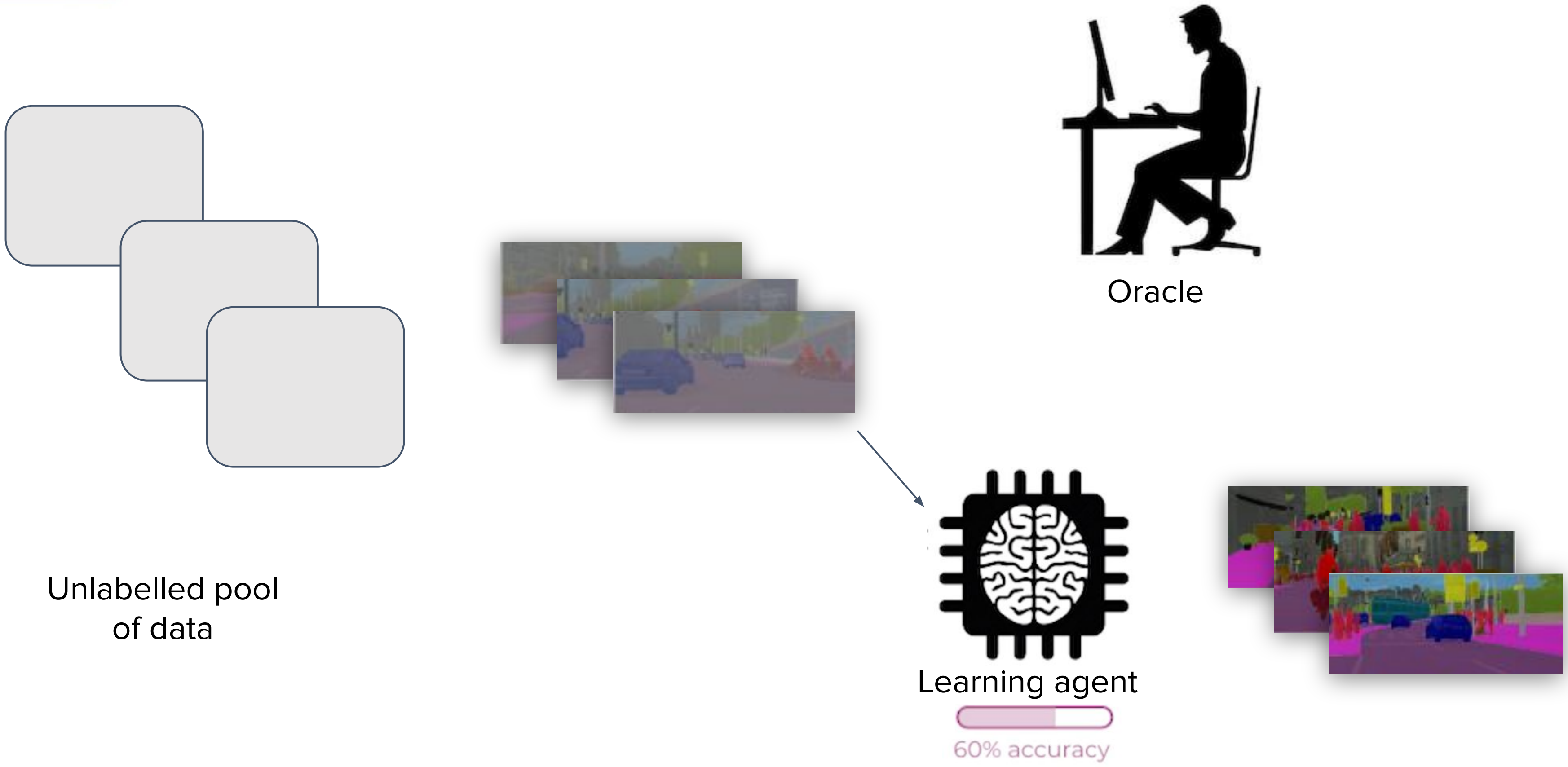
Learning agent

60% accuracy

Step by Step Active Learning



Step by Step Active Learning



Step by Step Active Learning

Oracle

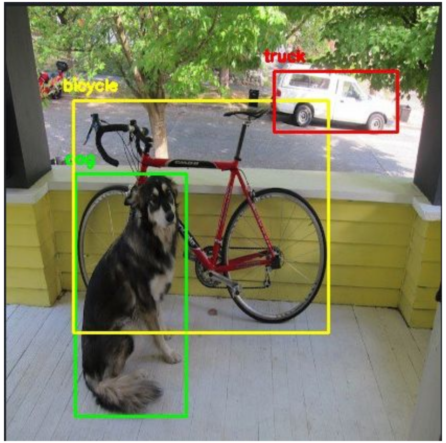
Unlabelled pool
of data

Learning agent

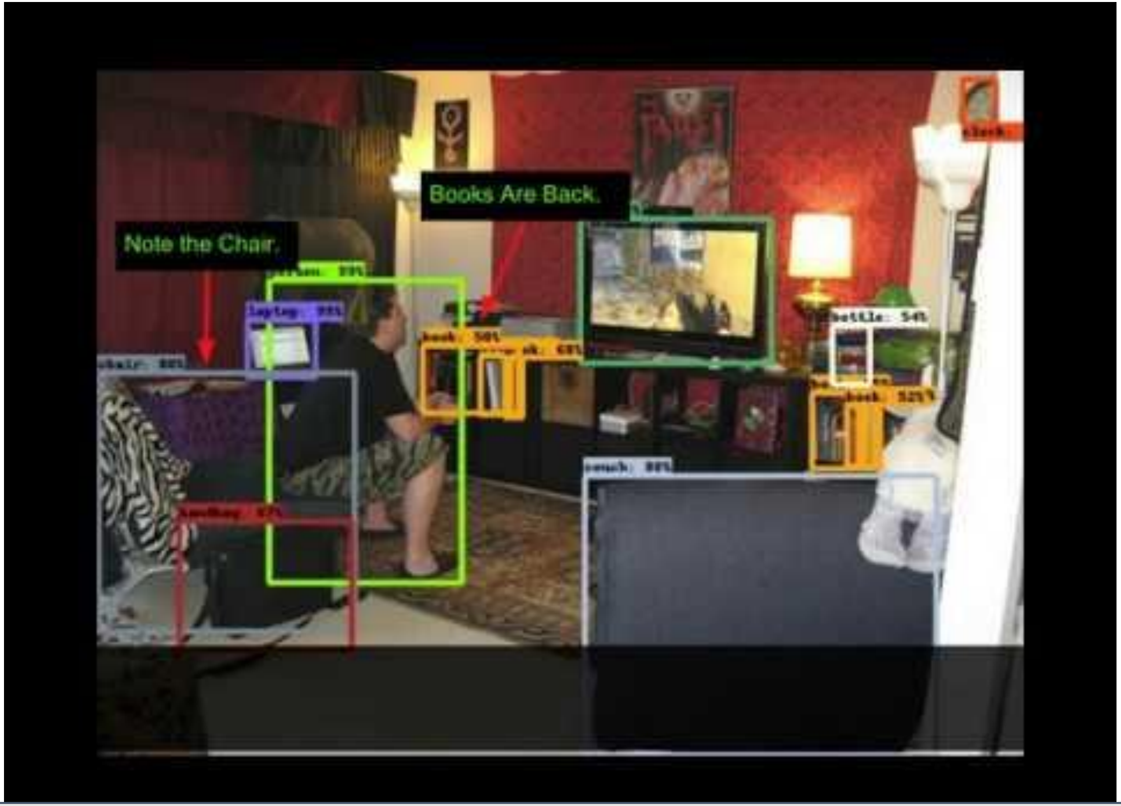
90% accuracy

Motivation

- Deep CNN models have large receptive fields
  - Enables learning semantically discriminative representations.
  - Leads to noisy predictions due to interference caused by spatially co-occuring objects.



Motivation



Rosenfeld, Amir and Zemel, Richard and Tsotsos, John K, The elephant in the room, arXiv preprint arXiv:1808.03305, 2018

Motivation

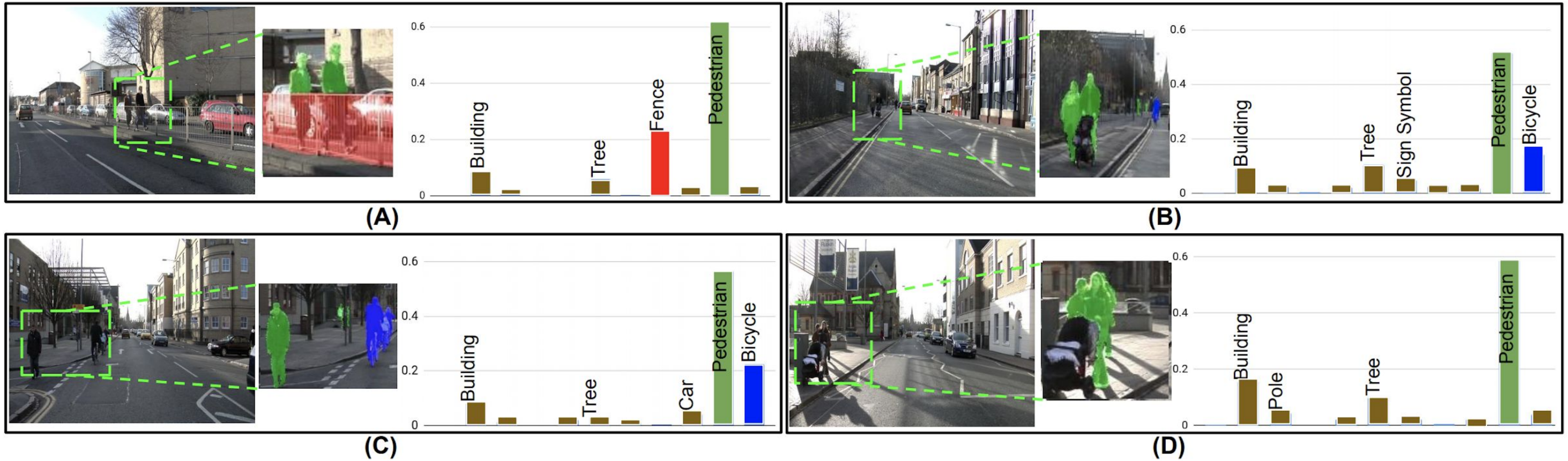
The objective is to select a set of training images that contain a diverse set of spatially co-occurring object classes.

Rosenfeld, Amir and Zemel, Richard and Tsotsos, John K, The elephant in the room, arXiv preprint arXiv:1808.03305, 2018

Key Contributions

- Novel information-theoretic distance-like measure, Contextual Diversity (CD).
- CD captures diversity in spatial and semantic context of various object categories.
- Two Active Learning (AL) approaches:
  - CD with core-set based active learning (CDAL-CS).
  - CD as a reward function in an RL framework (CDAL-RL).
- Experiments across three visual recognition tasks: semantic segmentation, object detection and image classification

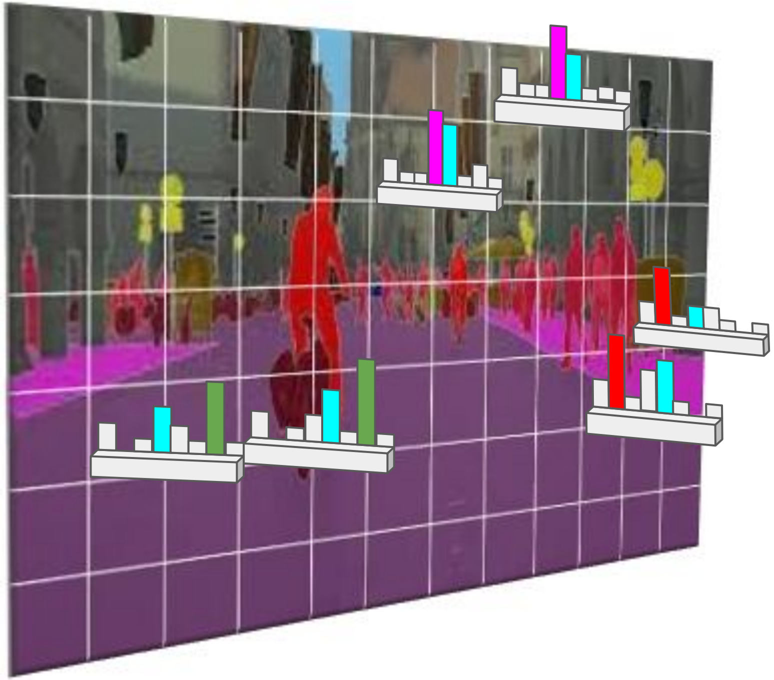
Contextual Diversity



- The softmax probabilities averaged over the set of pixels pseudo-labeled as '*Pedestrian*' show the confusion between spatially co-occurring classes.

- Contextual diversity based selection picks {(A), (C), (D)} as opposed to the set {(B),(C),(D)} picked by a maximum entropy based strategy.

Contextual Diversity



Inference for images in unlabeled pool using current model

Obtain probability vectors and *pseudo labels* for every pixel in image.

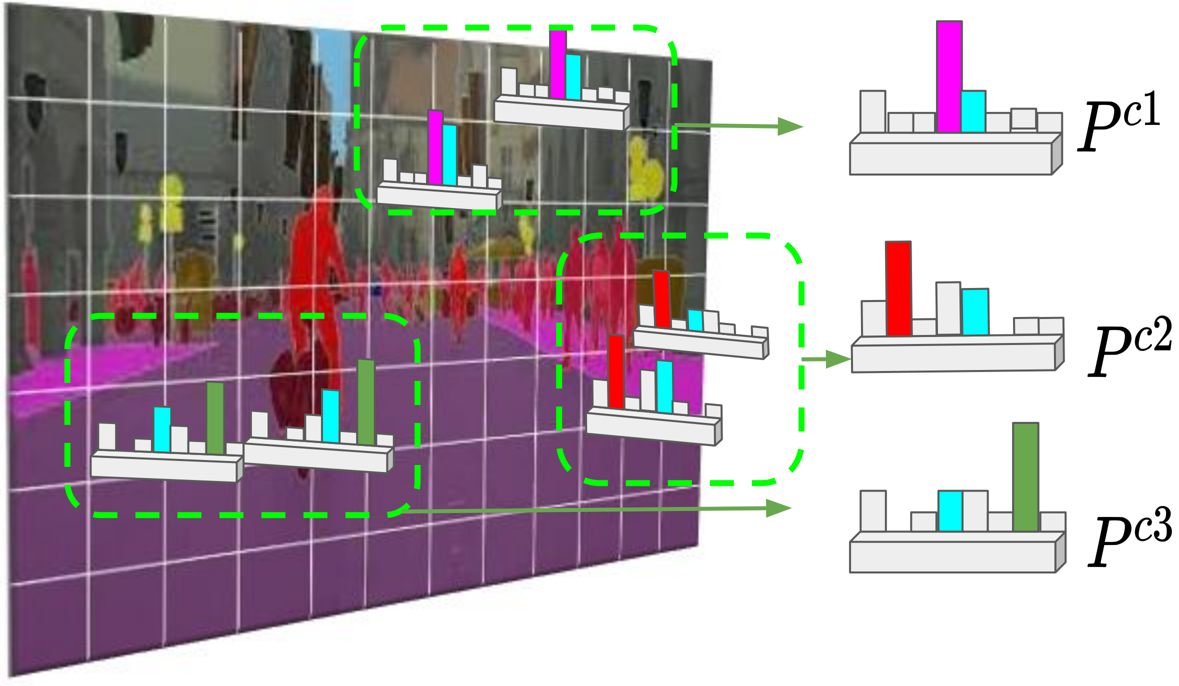
c1

c2

c3

pseudo labels

Contextual Diversity



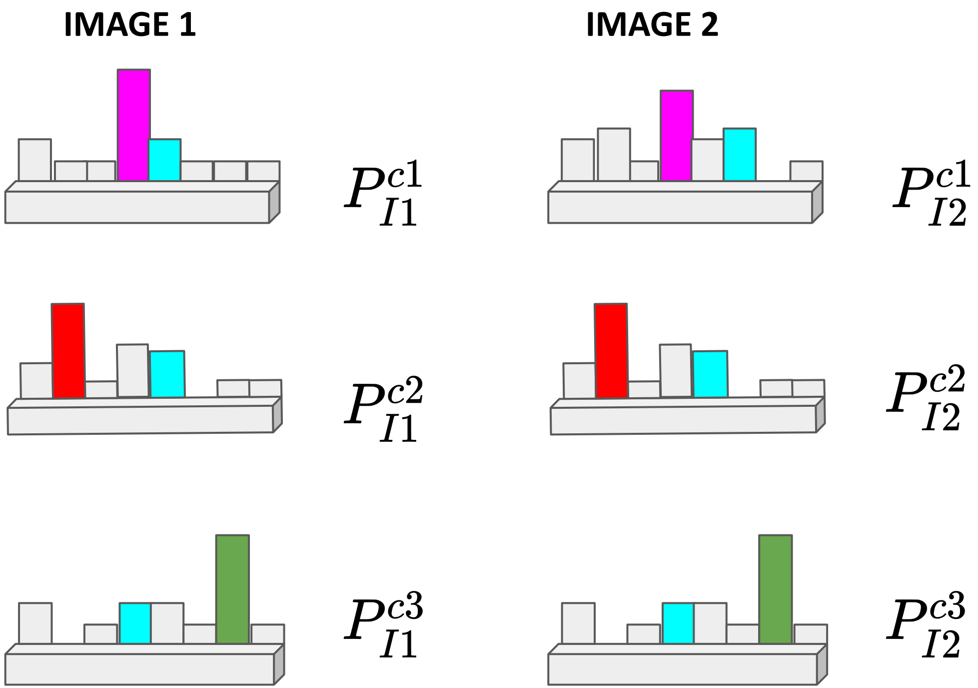
Compute a mixture distribution using probability vectors for each class using the pseudo labeled pixels.

$$P_I^c = \frac{1}{|I^c|}\sum_{\mathbf{I}\epsilon I^c} \frac{\sum_{r\epsilon R_{\mathbf{I}}^c} w_r \mathbf{P}_r(\hat{y}|\mathbf{I};\theta)}{\sum_{r\epsilon R_{\mathbf{I}}^c} w_r}$$

where the non-negative weights of the mixture is

$$w_r = -\sum_{j\epsilon C} P_r[j] log_2 P_r[j] + \epsilon, \epsilon > 0$$

Contextual Diversity



Compute these mixture distributions for all classes for all images in unlabelled set.

Compute pairwise contextual diversity for a pair of image using

$$d_{[\mathbf{I}_1,\mathbf{I}_2]} = \sum_{c\in C} \mathbb{1}^c(\mathbf{I}_1,\mathbf{I}_2)\left(0.5 * \mathrm{KL}(\boldsymbol{P}^c_{\mathbf{I}_1} \parallel \boldsymbol{P}^c_{\mathbf{I}_2}) + 0.5 * \mathrm{KL}(\boldsymbol{P}^c_{\mathbf{I}_2} \parallel \boldsymbol{P}^c_{\mathbf{I}_1})\right).$$

Finally, we add this pairwise measure over the selected batch to compute the aggregated contextual diversity

$$d_{\mathcal{I}_b} = \sum_{\mathbf{I}_m,\mathbf{I}_n \in \mathcal{I}_b} d_{[\mathbf{I}_m,\mathbf{I}_n]}.$$

CDAL-CS

- CDAL-CS contextual diversity based active learning using core-set.
- Inspired by the core-set approach for Active Learning.
- We simply replace the Euclidean distance with the pairwise contextual diversity and use it in the K-Center-Greedy algorithm.

Algorithm 1 CDAL-CS

Input: Unlabelled pool features X_L, Budget b, selected pool s

1: Add randomly selected data point d_0 to s
2: Initialize a min distance matrix D using Eq.(2) as distance metric
3: **repeat**
4: select new centre using $u = argmax(D)$
5: add u to selected pool s
6: update D
7: **until** $|s| = |b|$
8: **return** s

Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)

CDAL-RL

- Contextual Diversity R_{cd}

  This is simply the aggregated contextual diversity as given in Eq. 3 over the selected subset of images. I_b

- Semantic representation $R_{sr} = \sum_{c \epsilon C} log(\frac{N_c}{\lambda})$

  This term in the reward is to ensure that each class is sufficiently represented in the set of selected frames.
  N_c are the number of pixels classified as class c and λ is a hyperparameter. $(N_c << \lambda)$

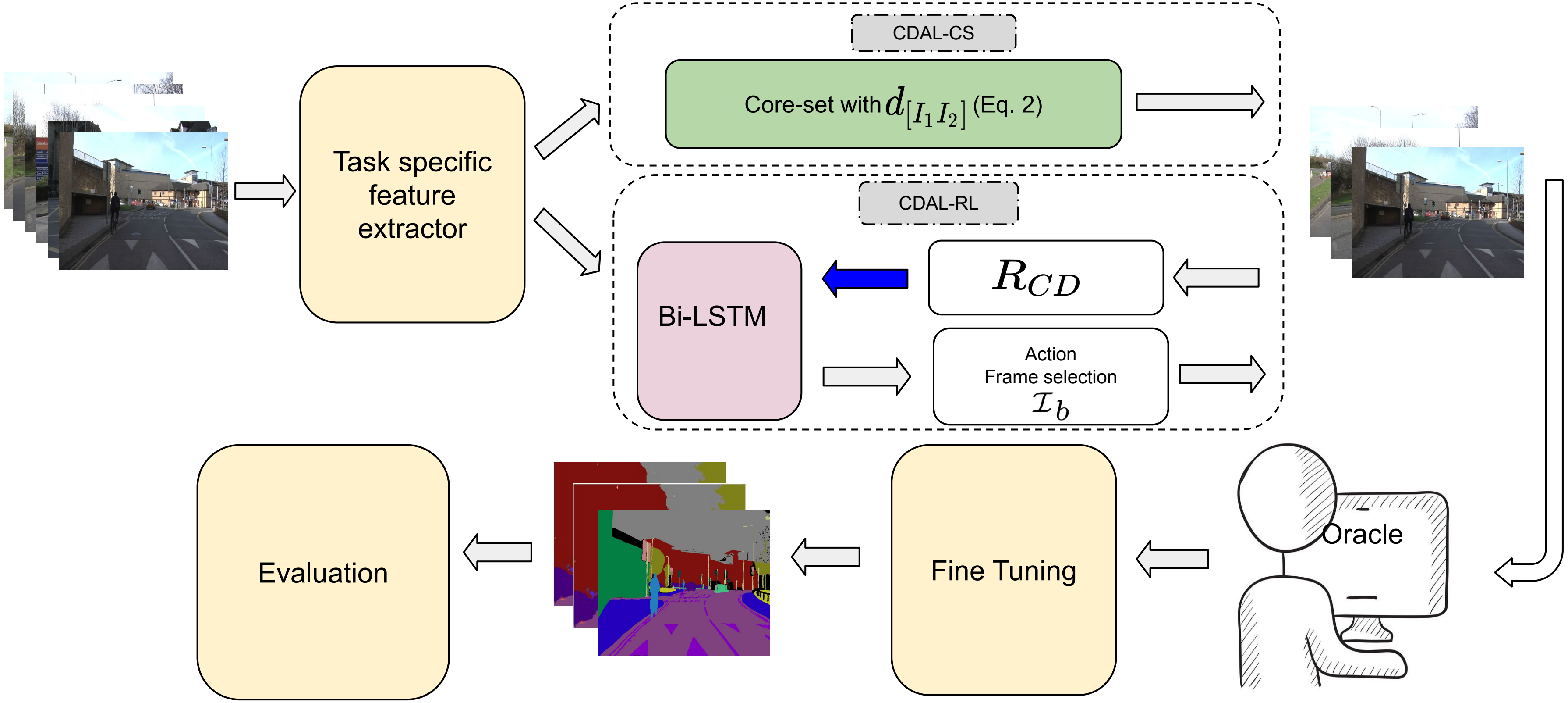
- Visual representation $R_{vr} = exp(\frac{-1}{|V|} \sum_{i=1}^{V} min_{j \epsilon S, j \neq i}(||x_i - x_j||_2))$

  This ensures that we have sufficient diverse set of frames that capture the visual dynamics in the videos. This term is necessary to have a good representation of visually diverse semantic classes.

$$\boxed{R = R_{cd} + R_{vr} + Rsr}$$

We define the total reward as and use it to train our LSTM based policy network.

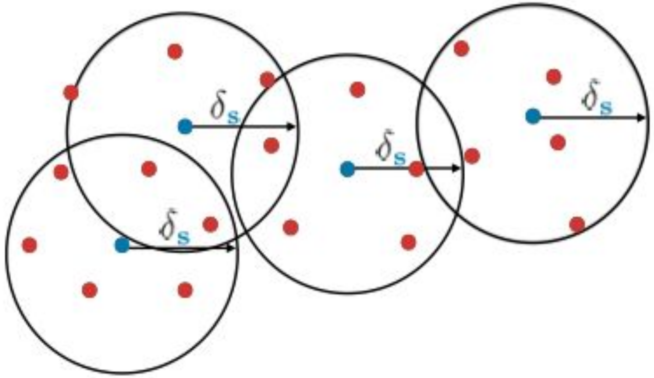
Proposed CDAL Architecture



Core-Set for CNNs (ICLR-18)

- The Core-set approach for active learning works on the set cover principle.

- It selects a subset of points in the CNN's feature space such that the union of $\mathbb{R}^n$ balls of radius δ around these points contain all the remaining unlabeled points.



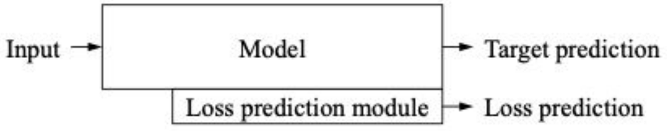
Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)

Learning Loss for Active Learning (CVPR-19)

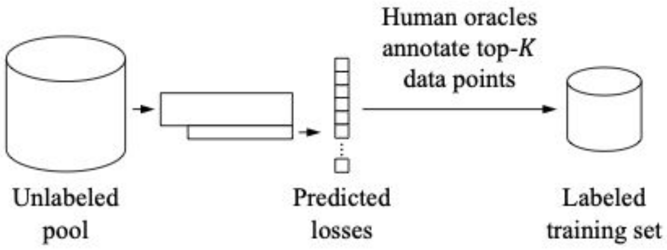
- Novel measure of uncertainty: a neural-net module learns to predict the loss value of an unlabeled data sample.

- Sampled data is ranked on the basis of predicted loss value.

- Top-k samples are selected for annotation.



(a) A model with a loss prediction module



(b) Active learning with a loss prediction module

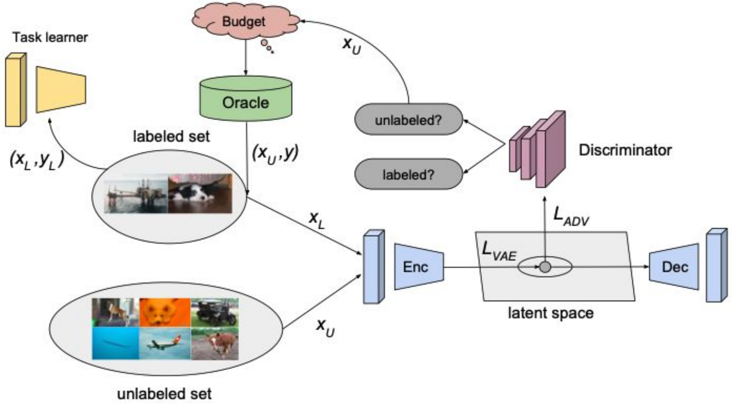
Yoo, D., Kweon, I.S.: Learning loss for active learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

Variational Adversarial Active Learning (VAAL)

(ICCV-19)

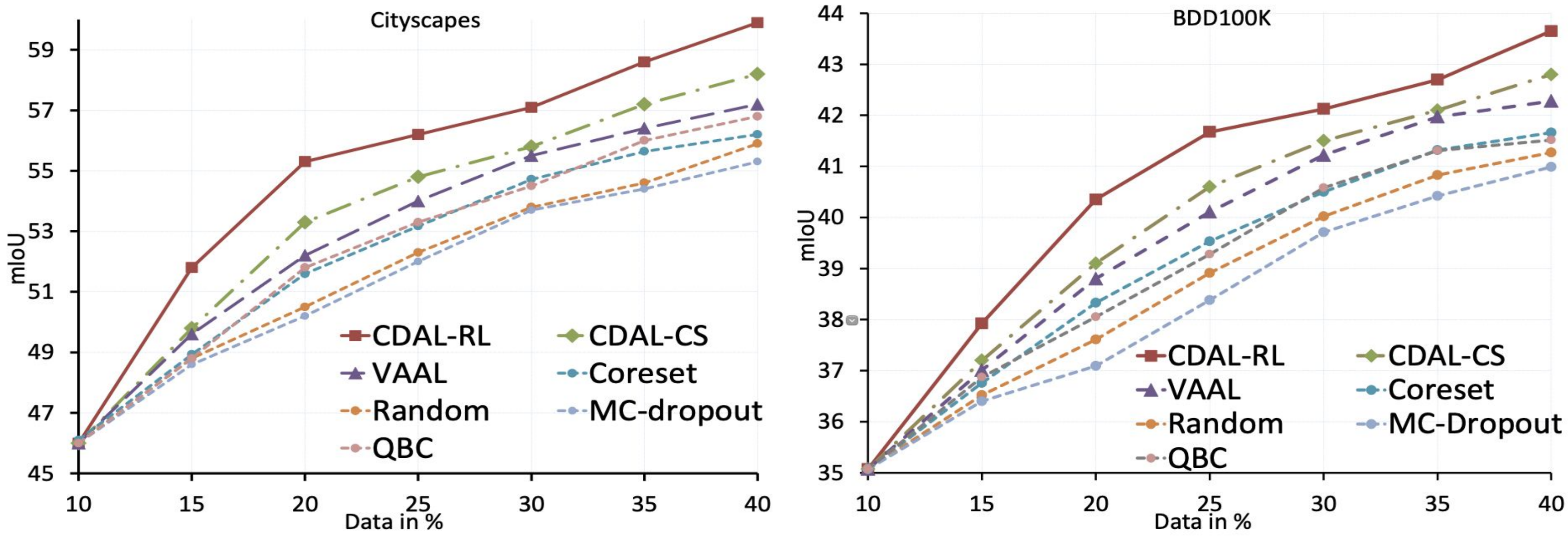
- Trains a VAE to map both *labeled* and *unlabeled* data into a common latent space, and a discriminator to distinguish between the two.

- Sample selection is performed based on the discriminator's prediction probability.



Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019)

Semantic Segmentation



- Annotation budget is set to 150 and 400 for Cityscapes and BDD100k respectively.
- CDAL-RL can achieve SOTA performance by reducing the labeling effort 300 and 800 frames on Cityscapes and BDD100k respectively.
- CD effectively captures the spatial and semantic context and selects the most informative samples.

Object Detection

- Comparing with Learning loss and following experimental setup.
- SSD as base detector network with VGG-16 backbone
- Annotation budget is set 1k samples.
- After 5k CDAL outperforms all the approaches.
- CDAL-RL achieved 73.3 mAP using 8k data where learning loss achieved it by 10K data hence reducing annotation cost by 2k samples.

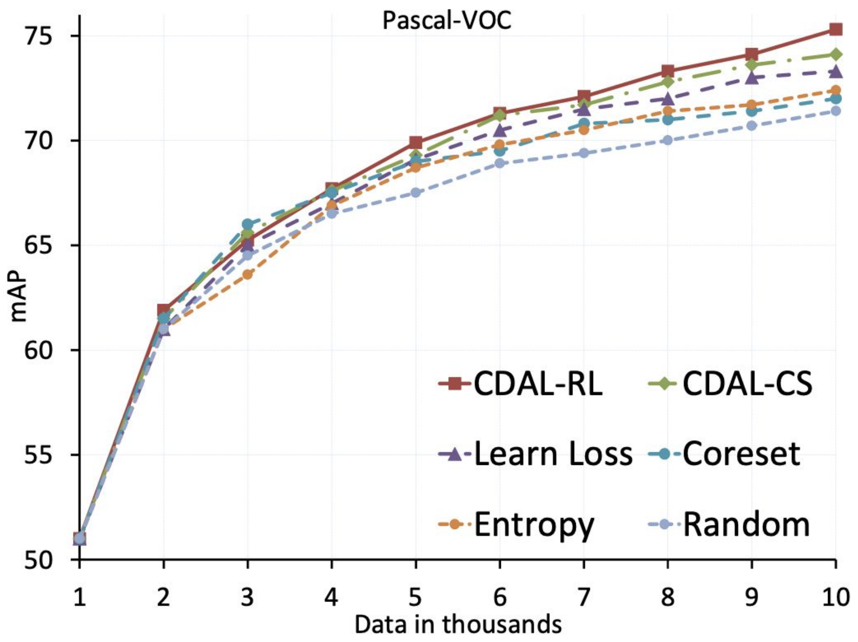
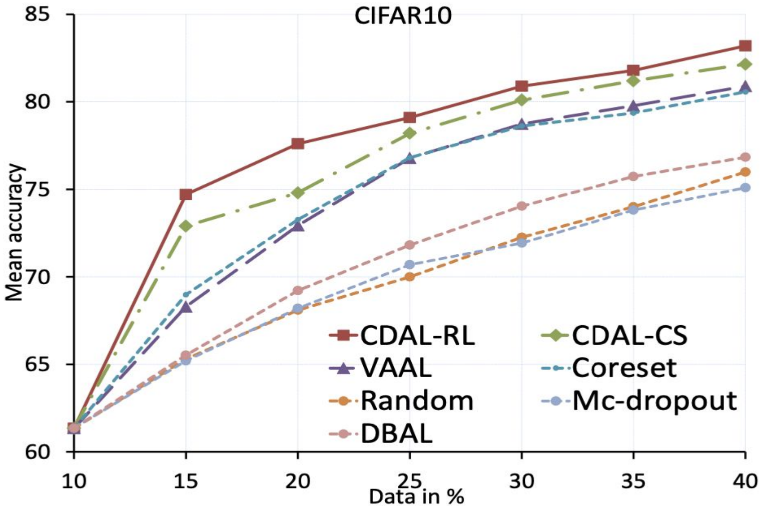
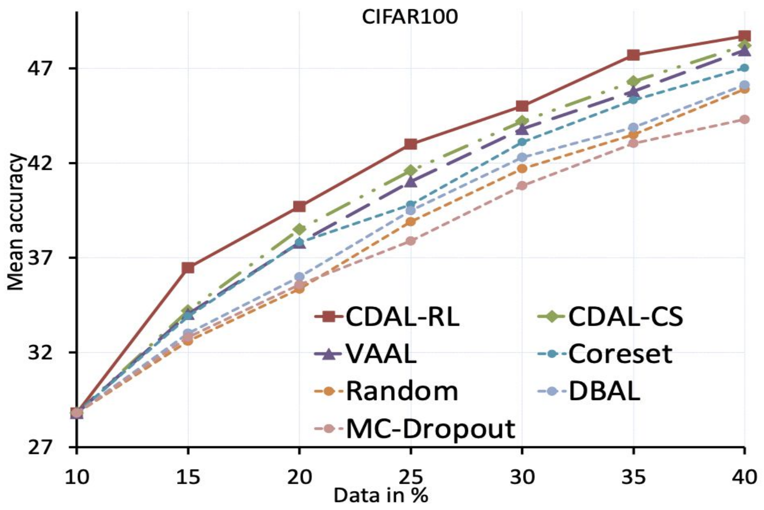


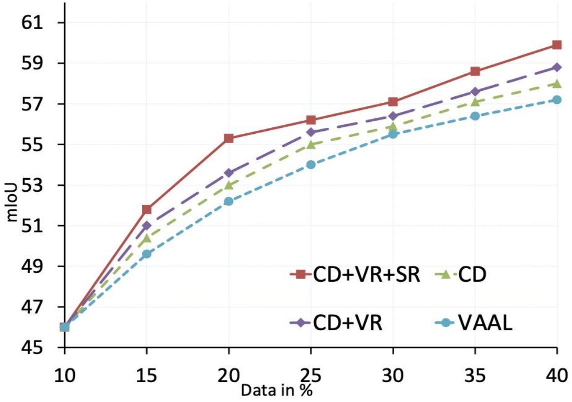
Image Classification



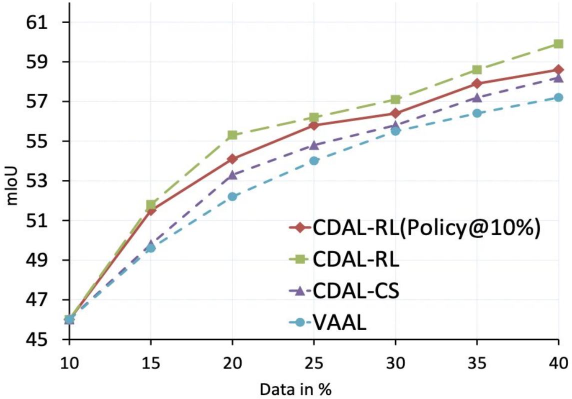


- CoreSet does not scale well for large number of classes, to demonstrate we have done classification on CIFAR100 as well with 100 classes.
- CDAL-RL can achieve 81% accuracy on CIFAR10 by using 5000 samples less than VAAL, and 47.95% accuracy by 2500 less samples on CIFAR100.
- KL divergence scales well with high dimensions unlike other metrics such as Euclidean Distance.

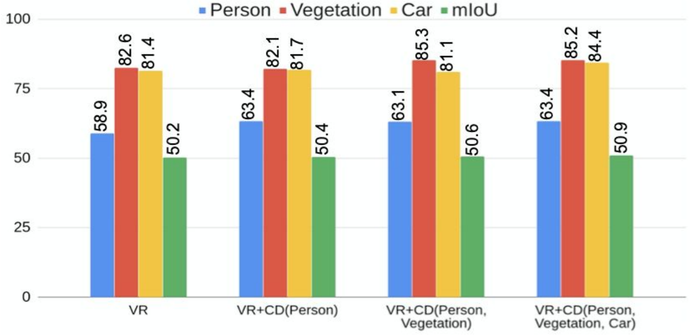
Analysis and Ablation experiments



(a)



(b)

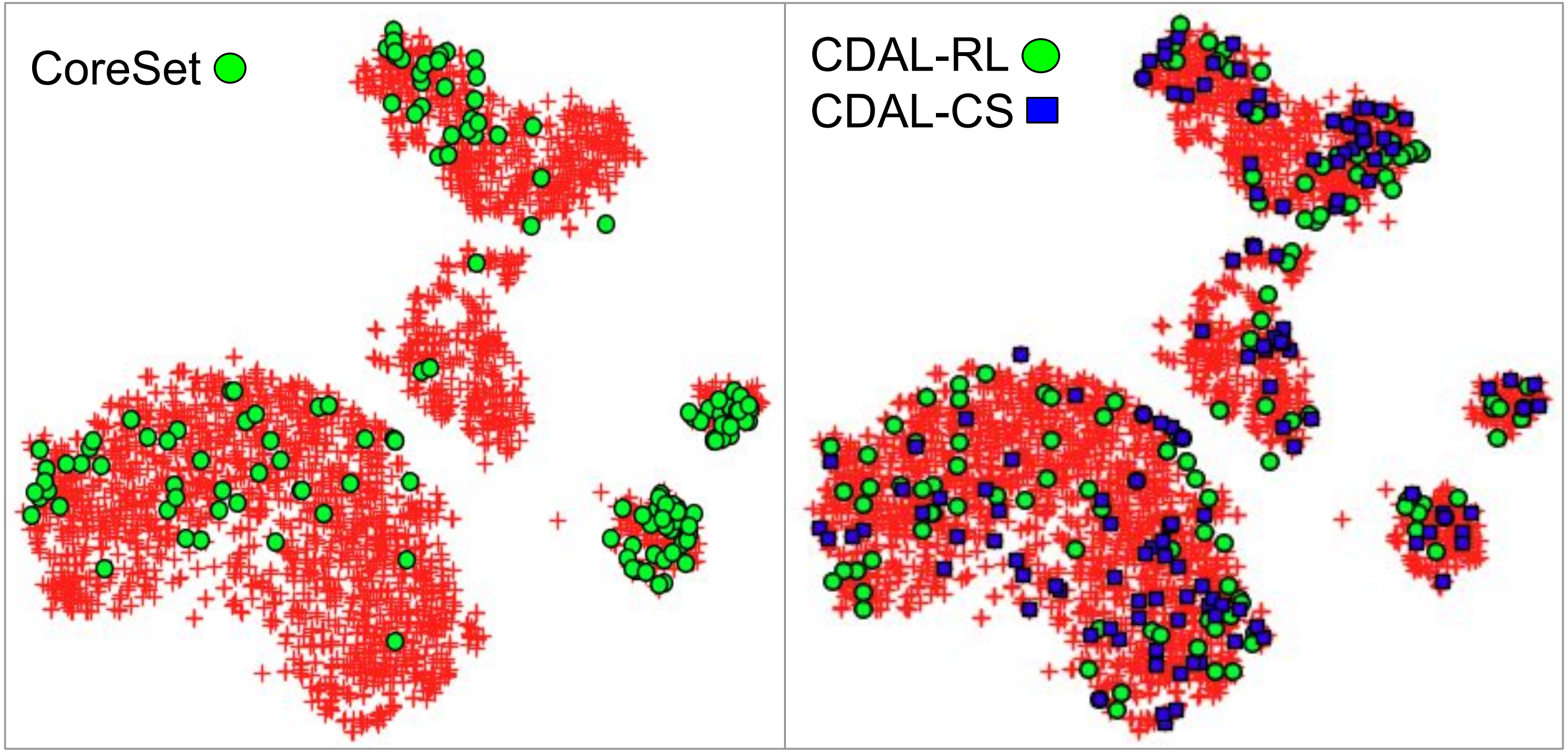


(c)

- **(a) Reward Component Ablation:** shows the performance of CDAL in three different reward settings.
- **(b) Policy Training Analysis:** train the policy network using the randomly selected 10% and use it in each of the AL iterations for frame selection without further fine-tuning.
- **(c) Class wise contextual diversity Reward:** initial model is trained using only the visual representation reward (leftmost group). As we include the Rcd term in the reward with the CD only being computed for the person class we see a substantial rise in IoU score. Similarly when person and Vegetation is included there is improvement in IoU.

Results on Cityscapes dataset

Visualisation of CDAL selection



Results on Cityscapes dataset

Conclusion

- Introduced contextual diversity based measure for the active frame selection problem.
- Experiments on three visual recognition task semantic segmentation, object detection and image classification.
- CD used as a distance metric with core-set or as a reward function for RL is a befitting choice.
- CD is designed using an information theoretic distance-like measure computed over the mixture of distributions of pseudo labeled samples which captures the model's predictive uncertainty as well as confusion across classes.