Contextual Diversity for Active Learning Supplementary Material

Sharat Agarwal*[1], Himanshu Arora*†[2], Saket Anand[1], and Chetan Arora[3]

[1] IIIT-Delhi, India, {sharata,anands}@iiitd.ac.in
[2] Flixstock Inc., himanshu@flixstock.com
[3] Indian Institute of Technology Delhi, India, chetan@cse.iitd.ac.in

1 Region Level Selection

We also compare with two approaches that are primarily considered as a region level selection approach: CEREALS [3] and RBAL [1]. CEREALS uses an FCN-8s [2] (with a 0.25 width multiplier) architecture, while RBAL uses the ICNet [5] architecture. We use the respective architectures to compute the contextual diversity (CD) for frame selection in our experimental comparisons. CEREALS operates in different modes, one of which uses the entire frame as selected region. This setting, while not the best performing version of CEREALS, is directly comparable with CDAL. We report this comparison in the first row of Table 1. For this comparison, we follow the same protocol as presented in [3] and take the initial seed set of 50 images and at each step another 50 samples are selected. We report the results when nearly 10% of the data ise selected, which amounts to about 300 frames in Cityscapes. We can see that for frame selection CDAL-RL outperforms CEREALS by about 3.4% mIoU. We also note that CDAL is complementary to the approach that CEREALS. The best configuration of CEREALS, which upon leveraging only small patches (128×128) within an image achieves an mIoU of 57.5% by annotating about 10% of the data.

In RBAL [1], the ICNet [5] model was pre-trained using 1175 frames, followed by selection of 10% of pixels across the remaining 1800 frames. For a fair comparison, we maintain the same annotation budget (1175 + 0.1*1800) of 1355 frames. We pre-train the ICNet model using 586 frames and select the remaining 769 frames using CDAL. The results after fine-tuning the model with the selected frames is compared with the mIoU reported in [1] in the second row of Table 1. We observe a reasonable improvement, even though CDAL only selects 1355 frames as opposed to RBAL accessing all 2975 frames.

Both of these results indicate that CDAL based frame selection complements the region based selection.

2 Qualitative Results for CDAL

In Fig. 1, we show the top 3 frames selected from three independent runs of CDAL-RL from the Class-wise Contextual Diversity Reward ablation described

* Equal contribution.
† Work done while the author was at IIIT-Delhi.

	Selected Data	Selected Frames	Base Model	mIoU (%)
CEREALS [3]	∼ 10%	300	FCN8s	48.2
CDAL-RL		300		**51.6**
RBAL [1]	45%	2975	ICNet	61.3
CDAL-RL		1355		**62.9**

Table 1. Comparisons with region-based active learning approaches on Cityscapes. CDAL-RL again outperforms both the methods with a significant margin.

in Sec. 5 of the main paper. The left most column shows images selected when CDAL-RL is only trained with a CD reward using the set of classes {Sidewalk}, the middle column with CD computed using {Sidewalk, Fence} and the right-most column with CD computed using {Sidewalk, Fence, Vegetation}. Two rows are shown for each selection. The first row shows the frame selected with the predictions for each class used for CD computation overlaid. The insets show the ground truth labels (top) and the predicted labels. The second row shows the class-specific confusion for the three classes considered: Sidewalk, Fence and Vegetation.

As we scan through the rows, we see that as the classes are included in the CD computation, the corresponding class-confusion reduces, which is evident from the reduced entropy (and increased *peakiness*) of each of the mixture distributions. As we see the selected images along the columns, we observe that the spatial neighborhood of the classes like Sidewalk contain different classes like Car, Motorcycle, and Person in the first column. Similarly, in the other two columns we see a different set of classes appearing in the selected frames in the spatial neighborhood of regions corresponding to the predictions of Sidewalk, Fence and Vegetation[1].

3 Ablation on Image Classification: CIFAR100

Biased Initial Pool. In the first case we check the robustness of CDAL in terms of biased initial labeled pool. We follow the exact experimental setup of VAAL [4] and at random exclude data for m=10 and m=20 classes from the initial labeled pool. From Figures 2(a), and 2(b), we can see that in both cases of m=10 and m=20 respectively the results of CDAL-RL are better than existing techniques.

Noisy Oracle. In the second set of experiments, we incorporated noisy oracle, similar to VAAL [4]. We replaced 10%, 20% and 30% of the selected labels with a random class from the same super-class. Figure 2(c) shows that CDAL-RL is substantially more robust to noisy labels as compared to other approaches. This is possibly due to the pairwise KL-divergence based selection of frames,

[1] More qualitative results are included in our project page at: https://github.com/sharat29ag/CDAL

which effectively captures the disparity of confusion between frames and is less sensitive to outliers.

Varying Budget. In the second set of experiments we changed the step size for varying budget experiments. In the main paper, all the experiments have been shown with a step size of 5% (e.g. 20%, 25%, 30% and so on). Figure 2(d) shows results on CIFAR 100 with budget steps of 10%. We can see that varying budget does not much effect the performance of CDAL-RL and performs better than VAAL [4] and its competitive approaches.

Change in network Architecture. As shown in the case of semantic segmentation CDAL-RL performs better irrespective of the network architecture, here we show results of CIFAR-100 on ResNet18, as shown in Figure 3 our CDAL-RLoutperforms all existing baselines by a substantial margin.

4 Ablation: Sensitivity analysis of α

As discussed in section 3.2 we have performed all the experiments with $\alpha = 0.75$. Here we justify that the selection of the weight for the reward was not very sensitive, but we focused on giving more weight to the CD component of the reward. The experiment was performed for CDAL-RL using the model trained at 20%, with the goal of selecting the next 5% samples to retrain the model at 25%. As we can see in the Table 2, we performed experiments with α taking values of 0.25, 0.5 and 0.75 and obtained the highest mIoU for 0.75. Even the smallest mIoU of 55.5% is still significantly higher than VAAL ($\sim$ 54%) and CDAL-CS ($\sim$ 54.9%). Therefore, while the performance may change by increasing α, it still is better than other competing methods.

α	0.25	0.50	0.75
mIoU	55.5	55.8	56.3

Table 2. Cityscapes ablation of α weighting factor for reward in CDAL-RL

4.1 Weights ablation of mixture distribution

Following the same experimental setup as above, we used the Cityscapes dataset and the model trained using 20% of teh data. With the goal of making the selection of the next 5% samples, we changed the mixture weights in Eq. (1) instead of Shannon's entropy. As we result, there was a deterioration in the 25% performance from 56.3% mIoU to 55.2% mIoU. This is in line with our expectation that the Shannon's entropy used as weight better captures the uncertainty in the predictions and therefore leads to better selections.

5 Algorithm

Algorithm 1 CDAL-CS

Input: Unlabelled pool features X^u, Budget b, selected pool s

Add randomly selected data point $x_0 \in X^u$ to s
Initialize a distance matrix D of size $|S| \times |X^u|$ using Eq.(2) as distance metric.
repeat
  compute $\widehat{D}$ as a $|X^u|$ dimensional vector of minimum distances from each centroid
    select new centroid using $u = \arg\max(\widehat{D})$
    add u to selected pool s
    update D
until $|s| = |b|$
return s

Algorithm 2 CDAL-RL

Input: Unlabelled pool features X^u, RL Model Parameters θ_{RL}

for e = 1 to epochs **do**
    Predict p_t for every data point
    sample $x^u \sim X^u$ using highest probabilities
    compute R_{cd} , R_{vr} , R_{sr}
    Using REINFORCE algorithm, calculate gradient $\nabla_\theta J(\theta)$
    $\nabla_\theta J(\theta) = \frac{1}{N} \sum \sum R_n$
    Update θ_{RL} using SGD
    $\theta_{RL} = \theta_{RL} - \alpha \nabla_\theta(-J)$
return trained θ_{RL}

References

1. Kasarla, T., Nagendar, G., Hegde, G., Balasubramanian, V., Jawahar, C.: Region-based active learning for efficient labeling in semantic segmentation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1109–1118 (Jan 2019). https://doi.org/10.1109/WACV.2019.00124 1, 2
2. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015) 1
3. Mackowiak, R., Lenz, P., Ghori, O., Diego, F., Lange, O., Rother, C.: CEREALS - cost-effective region-based active learning for semantic segmentation. In: British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3-6, 2018 (2018) 1, 2
4. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019) 2, 3

5. Zhao, H., Qi, X., Shen, X., Shi, J., Jia, J.: ICNet for real-time semantic segmentation on high-resolution images. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 405–420 (2018) 1

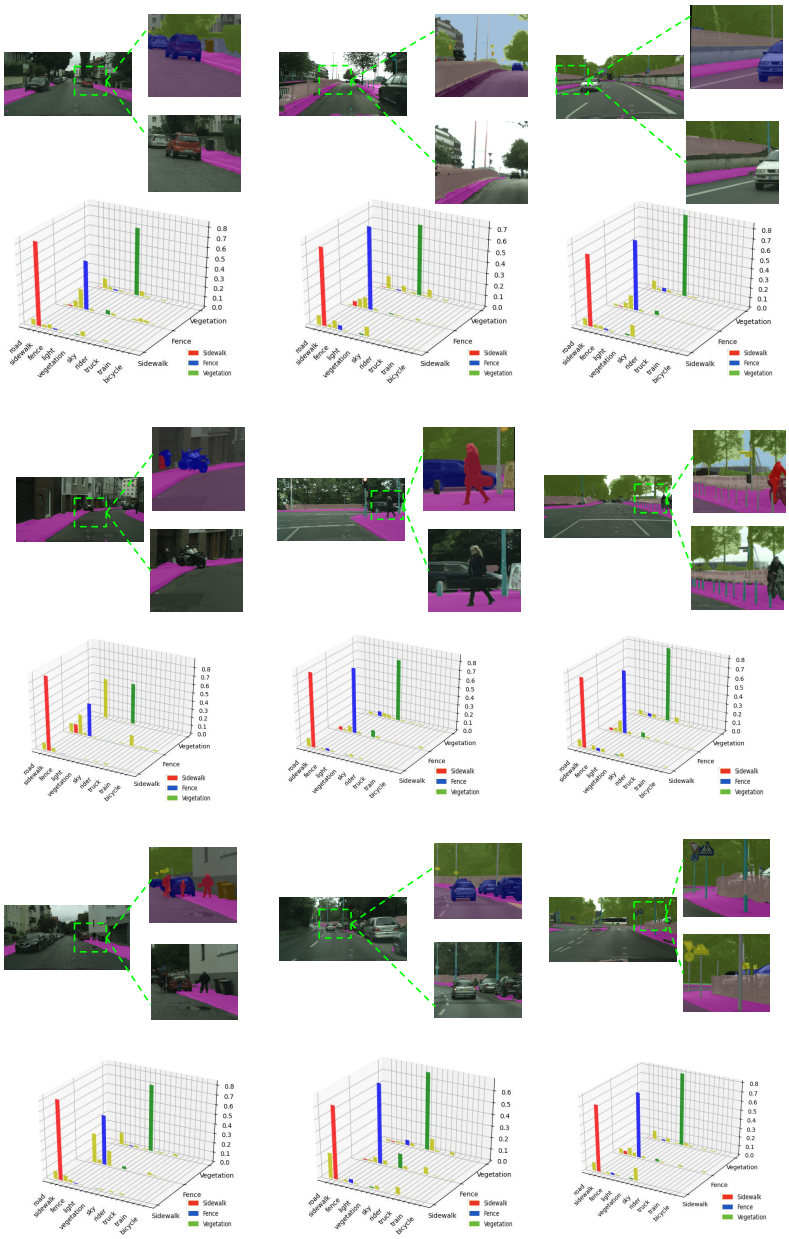

Fig. 1. Qualitative results for Class-wise Contextual Diversity Reward. Columns (Left to Right) show outcomes of experiments run when the Contextual Diversity (CD) is computed using the following sets of classes: {Sidewalk}, {Sidewalk and Fence}, {Sidewalk, Fence and Vegetation}. Pairs of rows show frames selected by CDAL when CD was computed using the aforementioned classes in the top row. The top and bottom insets show the ground truth and the predictions overlaid over the regions respectively. The bottom row, shows the class-specific confusion corresponding to the three classes used for computing CD. The mixture distribution depicting the class-specific confusion is computed over the selected image. As more classes are included in the CD computation, we see the confusion reducing.

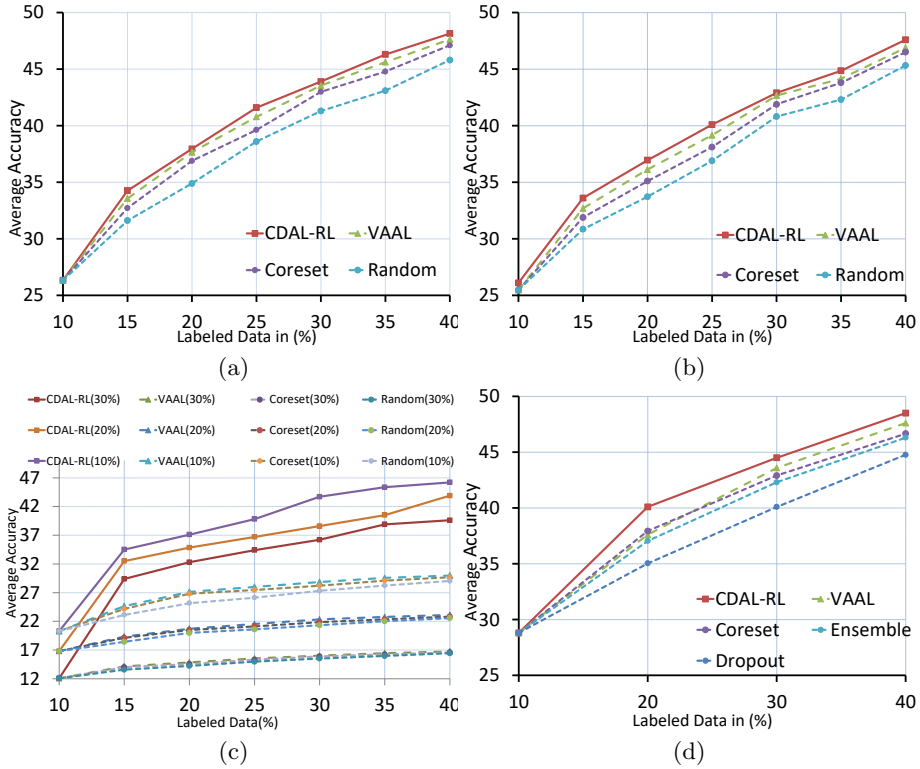

Fig. 2. Ablation of CDAL-RL on CIFAR-100. Biased Initial pool with with (a)m=10 and (b)m=20, (c) Noisy Oracle, (d) Varying budget with a step size of 10%

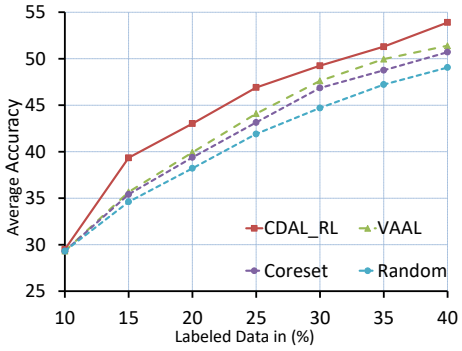

Fig. 3. Performance comparison of CDAL-RL using the ResNet-18 architecture on CIFAR-100.