RADIAL: Random Sampling from Intelligent Pool for Active Learning

Shivangana Rawat cs20mtech12001@iith.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Sai Vikas Desai cs17mtech11011@iith.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Vineeth N. Balasubramanian vineethnb@cse.iith.ac.in
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Abstract

Training CNNs on dense pixel-level prediction tasks such as semantic segmentation requires huge amounts of pixel wise annotated data, which is typically costly and time consuming to obtain. To reduce the annotated data requirement, recent works on active learning for semantic segmentation have introduced region level sampling where only the most uncertain regions in the image are annotated. However, picking regions only based on probabilistic uncertainty reduces the diversity of the training set. To alleviate this issue, we propose *RADIAL: RAnDom Sampling from Intelligent pool for Active Learning*. In this work, we describe how RADIAL addresses the exploration-exploitation tradeoff in active learning by randomly sampling from an intelligently curated subset of data. We show promising results on Cityscapes, Camvid and Weed segmentation datasets.

Keywords: Active Learning, Semantic Segmentation, Exploration-Exploitation

1. Introduction

Semantic segmentation is one of the cornerstone tasks in computer vision. While deep neural networks have achieved state of the art performance in semantic segmentation, training such networks requires huge amounts of pixel-wise annotated data, which is time consuming and expensive to obtain. Active learning(AL) has emerged as one of the promising approaches for reducing the annotated data required for training machine learning models (Settles (2009)). Active Learning aims to iteratively correct the decision boundary using minimum number of samples by querying labels for only the uncertain samples i.e., samples which are deemed close to the model's current decision boundary. These uncertain samples are added to the training set and the model is retrained on the updated training set. This *select-label-train* cycle is repeated until either the desired model performance is achieved or the annotation budget is exhausted.

Active learning has been heavily studied for the task of image classification (Sener and Savarese (2017); Gal et al. (2017); Yoo and Kweon (2019); Sinha et al. (2019)). Contrary to image classification which requires image-level labels, semantic segmentation requires

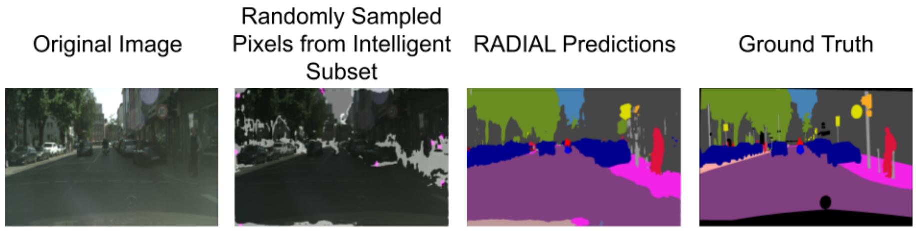

Figure 1: Output of the RADIAL method: RADIAL first intelligently obtains a subset of pixels (represented as white pixels in the second image) from an image using a measure of uncertainty. Then, a subset of pixels (shown in pink) are randomly sampled from those white pixels.

pixel-level labels for training. By assuming that not all pixel labels are equally valuable for training, recent works on active learning for semantic segmentation have attempted to identify and label only the most valuable regions in an image, further reducing the annotation requirement. This *region-level* active learning has been studied at three levels of granularity: 1) grid rectangles, 2) superpixels and 3) pixels. We focus on the lowest level of granularity i.e., identifying and labeling only the most valuable *pixels* for training.

In the pixel level active learning space, Kasarla et al. (2019) show that region level annotations for grid rectangles, superpixels and pixels can be used for obtaining 93.8% of the fully supervised performance while using only 10% of the pixels in the training set. Similarly, Shin et al. (2021) report that picking small number of pixels from more number of images is better than picking more pixels from less images. However, existing works on region based AL for semantic segmentation have two limitations: 1) The active learning methods only focus on uncertainty and not on diversity and 2) Shin et al. (2021) assumes that the entire unlabeled dataset is available before starting the active learning process. We address both these limitations in our work.

Most AL methods query labels for samples having the highest probabilistic uncertainty, which is a function of the model's output softmax vector. However, labeling only highly uncertain samples can introduce redundancy in the training set. To put this in the context of the well known exploration-exploitation trade-off, most AL methods favor exploitation of the search space of samples and ignore exploration. To promote exploration and introduce diversity in the labeled set, Sener and Savarese (2017) sample core-sets, which are highly diverse subsets of the dataset. However, sampling core-sets is NP-Hard and even the greedy approximation takes $O(N^2)$ time, where N is the size of the dataset. In our work, we aim to obtain diverse samples by a much simpler albeit faster method - random sampling. Since random sampling is a strong baseline for active learning methods, we combine the effectiveness of highly uncertain samples for their informativeness and randomly chosen samples for their diversity, we propose RADIAL - Random Sampling from Intelligent Pool for Active Learning. Instead of directly selecting a highly uncertain subset from the dataset, RADIAL samples a larger subset of samples based on uncertainty and then further selects a bunch of samples from this larger subset using random sampling. By combining both

Algorithm 1 Pixel based Active Learning Pipeline

Input: Training set T, Size of initial image pool k_i, AL batch size k, Number of labeled pixels per image p, Number of AL cycles N

Randomly split $T \rightarrow \{U, L\}$ where $|L| = k_i$.
Obtain labels $\forall x \in L$.
$M \leftarrow Train(L)$. {Until convergence}
for $i = 1...N$ **do**
  Select image set using acquisition function $B_i = Q_i(M, U)$ such that $|B_i| = k$.
  Select pixel set using acquisition function $B_p = Q_p(M, B_i)$ such that $|B_p| = p$.
  Obtain labels for B_p from the oracle.
  $L \leftarrow L \cup B_i$
  $U \leftarrow U \setminus B_i$
  $M \leftarrow Train(L)$ {Until convergence}
end for
return M

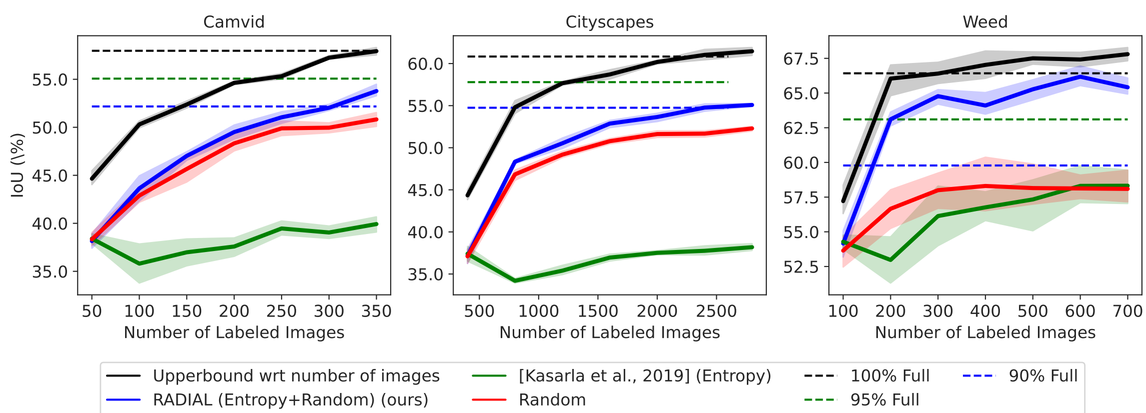

Figure 2: Active Learning performance for the Camvid, Cityscapes and the Weed dataset

informativeness and diversity, RADIAL aims to achieve a balance between exploration and exploitation.

2. RADIAL

2.1 Pixel based Active Learning for Semantic segmentation Pipeline

We consider the pool based active learning setting where we have a pool of images and the objective is to train a machine learning model with minimum labeling budget. We consider a training set T, which is a union of the labeled set of images L and an unlabeled set of images U. We start with a labeled initial pool, train a model on the initial pool, sample using the trained model and finally select samples based on the trained model. The entire pipeline is described in Algorithm 1.

2.1.1 Acquisition Functions

We go on to describe the different image level and pixel level acquisition functions for active learning. Given the pixel level acquisition function $S_{(i,j)}$, the image level acquisition function S_I can be computed as given below

$$S_I = \frac{1}{|H * W|} \sum_{i \leq H, j \leq W} S_{(i,j)} \tag{1}$$

Random image sampling and pixel sampling randomly picks images and pixels from the available set of unlabeled points respectively.

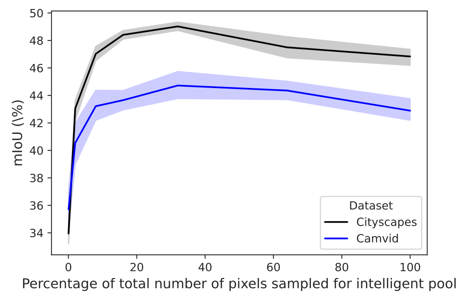

Figure 3: Effect of size of intelligent subset for the Camvid and Cityscapes dataset

Entropy computes uncertainty score by considering the entire output probability distribution. Pixel entropy can be computed using Equation 2.

$$S_{(i,j)} = - \sum_{c \in \{1 \ldots C\}} p_c \log p_c \tag{2}$$

Here p_c represents the output class probability of class c and C is the total number of classes in the dataset.

Margin computes uncertainty score by considering the top two highest conditional probabilities from the output probability distribution. Pixel entropy can be computed using Equation 2.

$$S_{(i,j)} = -(p_1 - p_2) \tag{3}$$

Here p_1 represents the probability of the class which has the highest probability and p_2 is the class which has the second highest probability.

2.2 Pixel Sampling

Images contain a large number of pixels and the model could be confused over a specific class and selecting the most uncertain samples would not account for the diversity in the pixels. In order to account for the diversity in the highly uncertain samples, we randomly select from the intelligently sampled uncertain subset. The intelligent sampled subset accounts for the exploration aspect whereas the random sampling accounts for the exploitation aspect. Ensuring the diversity aspect among the selected samples can be computationally expensive thus random sampling provides us with a constant time approach to select diverse samples.

3. Experiments

3.1 Datasets

Camvid dataset Brostow et al. (2009) is an autonomous driving segmentation dataset which contains images of size 360x480 with 11 classes.

Cityscapes dataset Cordts et al. (2016) is another standard autonomous driving segmentation dataset which contains images of size 1024x2048 with 19 classes.

Weed segmentation dataset Haug and Ostermann (2014) is a real world agricultural dataset. The weed dataset contains images of size 384x512 with 4 classes.

3.2 Parameter Settings

Following Xie et al. (2020), we use Deeplabv3+ as our segmentation model with a MobileNetv2 backbone and we train it for 50 epochs. We use an AL batch size of 50 for Camvid, 400 for Cityscapes and 100 for the Weed dataset. For every image we sample 50 pixels to be labeled by the oracle. For all our results we report the mean of the IoU for 5 runs.

3.3 Results

Figure 2 shows the results of our experiments. Upper bound refers to the maximum performance which can be obtained with a specific number of images. For the camvid dataset, we observe that entropy sampling achieves a performance of 39.92% which is worse than the performance achieved using randomly sampled pixels which is 50.82%. We also observe that randomly sampling from an intelligent subset achieves a performance of 53.77% which is better than both random and entropy sampling. We observe the same trend for the Cityscapes dataset where the number of images is much greater than the Camvid dataset. Random sampling achieves a performance of 52.29% and entropy sampling achieves a performance of 38.18% but randomly sampling from the intelligent subset chosen using entropy achieves a performance of 55.08% which is better than both the baseline methods. We also show results on a real world agricultural weed segmentation dataset since weed detection is a very important task for agriculture. We observe that random sampling achieves a performance of 58.35% and entropy sampling achieves a performance of 58.84%. For this dataset also, we observe that the random sampling from an intelligent pool achieves a mIoU of 65.08% which is better than both the other sampling techniques.

We observe that pixel selection based on uncertainty lacks diversity and leads to worse performance than random sampling. Random sampling from an intelligently chosen subset outperforms random sampling and entropy sampling since it contains diverse samples which have high uncertainty. The uncertainty sampling can be viewed as exploitation of the knowledge which the model processes. On the other hand, random sampling helps us to explore data points which the model may be uncertain on and introduce diversity. Figure 3 shows the exploration exploitation trade off associated with random sampling from an intelligent subset. It shows the mIoU with respect to the percentage of the total number of pixels which are chosen as the intelligent subset. The maximum performance gain is obtained in the first episode so we plot the mIoU after the first episode by varying the size of the intelligently sampled subset. When the size of the intelligent subset is small then in that case, the method performs similar to entropy sampling and the performance is low. As the size of the intelligent subset increases, diversity among samples is introduced and the performance of the model increases. The trade off is optimized at approximately 35% for our datasets at hand. The performance of the model starts to decrease after that point and drops to that of random sampling.

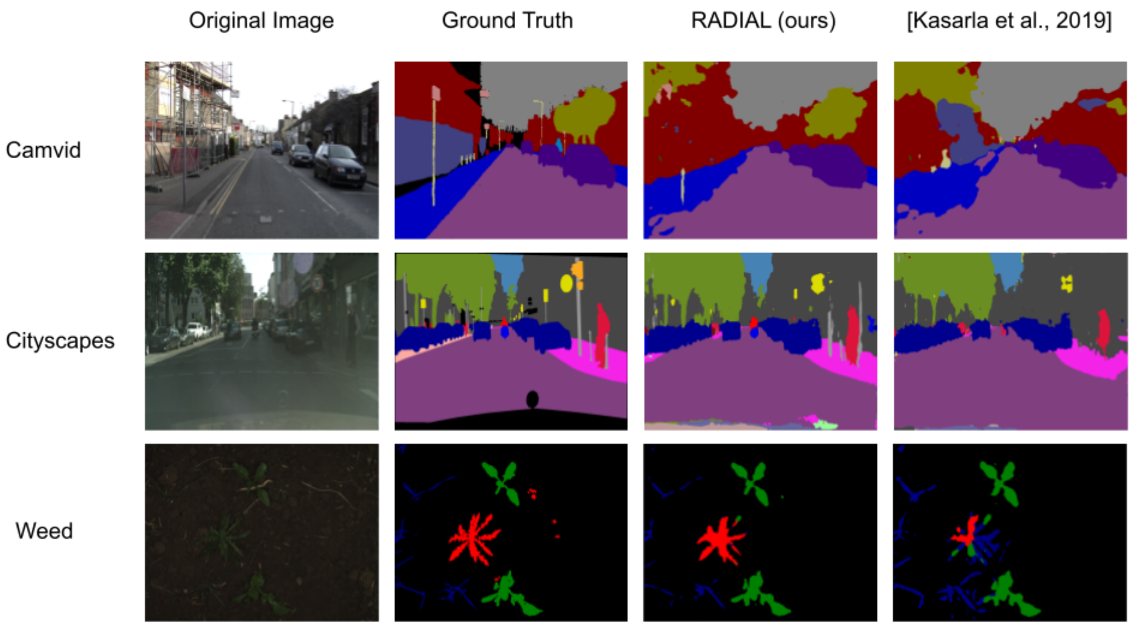

Figure 4: Qualitative results for our experiments on the Camvid, Cityscapes and the Weed datasets.

Figure 4 shows the qualitative results for our experiments. We can observe that the predictions made using our method capture a lot of detail and are also visually better than Kasarla et al. (2019) which are obtained without using an intelligent subset.

4. Conclusion

In this paper, we propose RADIAL: Random Sampling from Intelligent Pool for Active Learning. Our experiments have shown that RADIAL can effectively perform pixel level active learning on the semantic segmentation task, achieving 90% of full supervision performance with 0.04% of labeled data. Results on Camvid, Cityscapes and Weed datasets show that RADIAL outperforms both random sampling and entropy sampling. In addition, we have studied the exploration-exploitation tradeoff in the context of active learning by varying the size of intelligent pool from which random sampling is done. These results strongly suggest that random sampling from an intelligent labeled pool can help pick a subset of samples which are both informative and diverse, thus improving the model performance faster when compared to active learning methods which only pick informative samples.

References

Gabriel J Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2):88–97, 2009.

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, M. Enzweiler, Rodrigo Benenson, Uwe Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017.

Sebastian Haug and Jörn Ostermann. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In *European conference on computer vision*, pages 105–116. Springer, 2014.

Tejaswi Kasarla, Gattigorla Nagendar, Guruprasad M Hegde, Vineeth Balasubramanian, and CV Jawahar. Region-based active learning for efficient labeling in semantic segmentation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1109–1117. IEEE, 2019.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Burr Settles. Active learning literature survey. 2009.

Gyungin Shin, Weidi Xie, and Samuel Albanie. All you need are a few pixels: semantic segmentation with pixelpick. *arXiv preprint arXiv:2104.06394*, 2021.

Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5972–5981, 2019.

Shuai Xie, Zunlei Feng, Ying Chen, Songtao Sun, Chao Ma, and Mingli Song. Deal: Difficulty-aware active learning for semantic segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 93–102, 2019.