**Cost Effective REgion-based Active Learning for Semantic Segmentation (CEREALS)**

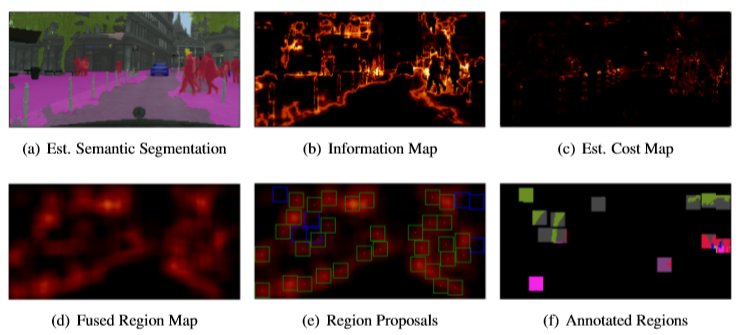
2b) Cost Extraction: It is the work of assumption some samples in an unlabeled pool are more costly to label by a human oracle than others. They are approximating costs by the number of clicks necessary to annotate an image. For cost extraction they perform a forward pass for each individual image within the current unlabeled pool through the cost model for retrieving an estimate about clicks. They denote the result given an image as cost map.

2c) Region Aggregation and Fusion: Some regions could be very costly to label while having only little positive impact on the models performance and vice-versa. They linearly scaled region information maps and region cost maps w.r.t. the whole dataset, such that all values are in [0;1].



They have evaluated the three simple fusion function with the region information map I and the region cost map C. The parameter **Alpha** in (5) allows to set a trade-off for linearly interpolating between both region maps.

After fusing the region information and the region cost map pairs for all images they performed non-maximum-suppression to retrieve ﬁxed-size region candidates for each individual image and store the region candidates for each individual image of the unlabeled pool within a region proposal



3) Acquisition: From the region proposal pool extracted as many top scoring regions as would correspond to extracting m images out of a pool of equally sized images regarding their amount of pixels for a fair comparison to the image-based acquisition of labels. They update the labeled and unlabeled pool and learn semantic segmentation model and cost model from scratch.

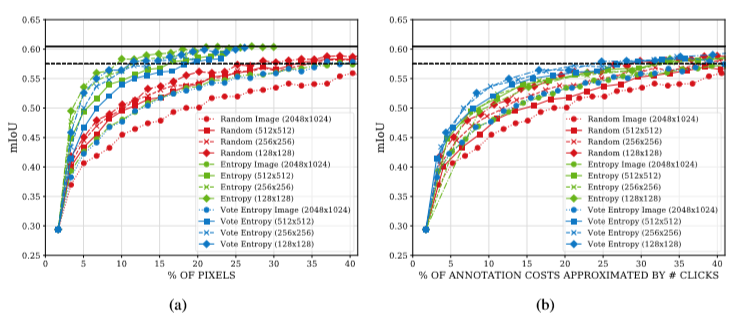
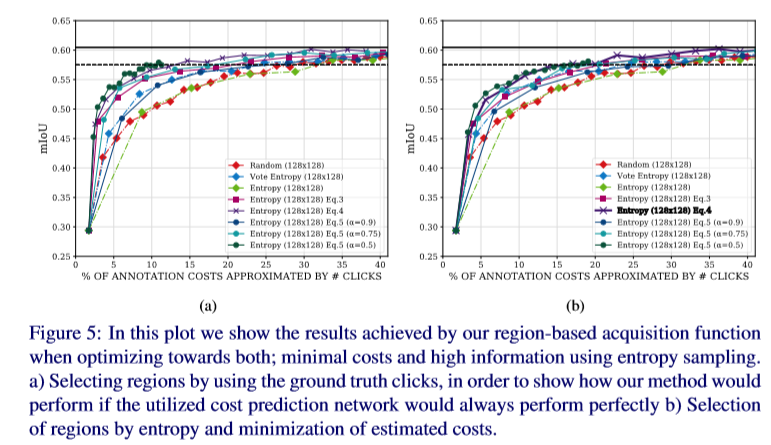


Figure: AL curves showing the relationship between pixels and annotation costs approximated by the number of clicks regarding different acquisition functions. The solid blackline shows the mloU achieved by training the model on the whole training set of Cityscapes. The dashed blackline marks 95% of the performance achieved by this model. **a)** Resulting mloU as function over the amount of labeled pixels queried from annotator. **b)** Same obtained results but plotted as a function over the annotation effort measured by the number of clicks.

4) Results: All processed experiments presented in this work are repeated five times and they report the average mean Intersection over Union (mIoU) calculated on the validation dataset of Cityscapes after training convergences.

After 21 acquisition steps corresponding to 35.29% of queried labels by using entropy sampling, they achieve 95% of the performance as compared to the obtained result of 0.605, when training on the full training set of Cityscapes.



Their proposed method for cost effective active learning for semantic segmentation tailored to fully convolutional neural networks. They demonstrated their framework’s performance on Cityscapes, a highly diverse high definition dataset consisting of images of urban scenes captured in the wild. They showed that combining information content and cost estimates is a powerful approach for cost-effectively building new training datasets from scratch. With only 17% of the effort measured by the number of clicks which were executed for annotating the Cityscapes training set, it was able to achieve 95% of the full training set’s performance.

Reference:

<https://www.researchgate.net/publication/328474791_CEREALS_-_Cost-Effective_REgion-based_Active_Learning_for_Semantic_Segmentation/link/5c48316c458515a4c73a05ce/download>