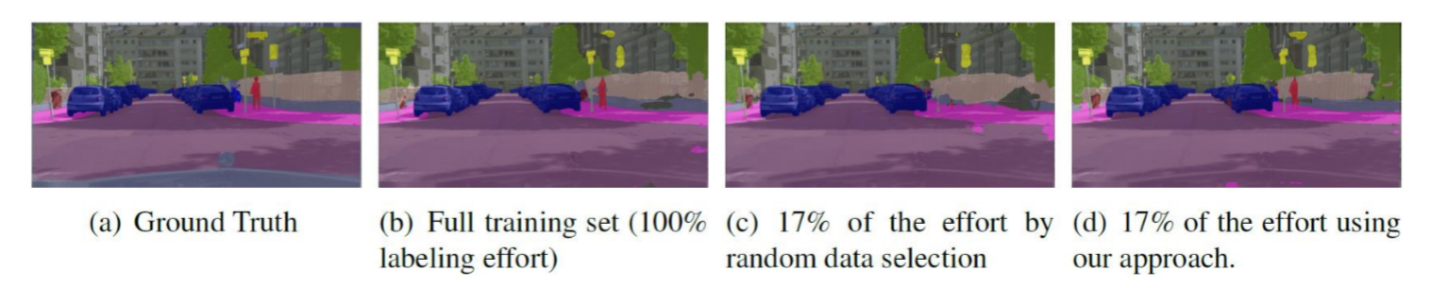
**Cost Effective REgion-based Active Learning for Semantic Segmentation (CEREALS)**

*Radek Mackowiak, Philip Lenz, Omair Ghori, Ferran Diego, Oliver Lange, Carsten Rother*

Another proposal on Active learning based on strategy called CEREALS where it minimizes human annotation effort while maximizing the performance of semantic image segmentation method. Semantic Image segmentation are trained in a supervised fashion using a large corpus (a collection of written texts) of fully labeled training images. However, gathering such a corpus is expensive, due to human annotation effort, in contrast to gathering unlabeled data.

The proposed framework reduces the labeling effort by (1) Utilizing spatial estimates about annotation costs inferred from a learned cost prediction CNN and (2) By focusing on image regions promising high information content and low annotation costs in a global context.

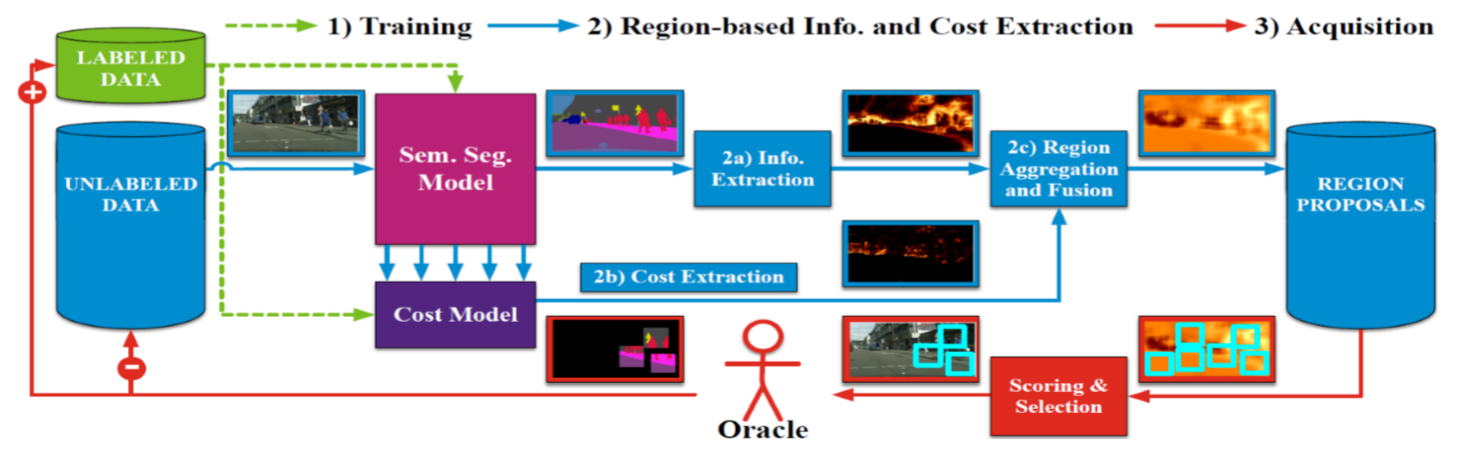
The performance of CEREALS is demonstrated on Cityscapes, where it was able to reduce the annotation effort to 17%, while keeping 95% of the mean Intersection over Union (mIoU) of a model that was trained with the fully annotated training set of Cityscapes.



In this proposal they tried to lose the burden of manual annotation and tried to apply 6 different strategies **(1) Pre-training (2) Weakly-supervised learning (3) Semi-supervised learning (4) Interactive segmentation (5) Active learning**.

Active learning is a form of semi supervised machine learning where the algorithm can you choose which data it wants to learn from. Pool-based active learning exploits the inequality of amount of information in an existing unlabeled pool and Angels to find the most valuable sample to be labelled by an Oracle able to reveal the ground truth semantics of interest given some data. Cost effective AL for CNNs has been recently proposed for image classification tasks, where highest confidence pseudo-annotated unlabeled samples are added to the training set with no human cost at all. Here they used CNN trained in a strongly supervised manner and a simple polygon-based annotation tool.

They consider a pool-based AL scenario running in batch-mode. In such a setting a large unlabeled pool of data exists from which a small, randomly sampled subset, called the seed set, is initially extracted and labeled by an oracle. Using this seed set the algorithm works as follows: First, a model is trained on the currently labeled pool. Secondly, some measure of information on each individual unlabeled sample is being computed. Thirdly, an acquisition function is applied. A subset of a pre-specified number of elements maximizing the acquisition function is annotated by an oracle. It is then added to the labeled pool. The process is repeated until either a desired performance or labeling budget is reached. the stopping criterion is satisfied whenever the unlabeled pool becomes exhausted which is indicated by no further improvements after several acquisition steps.

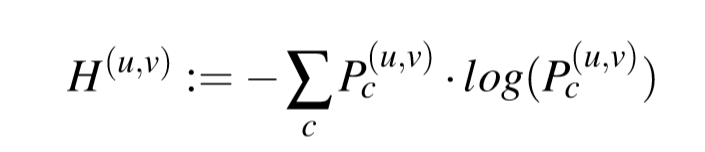


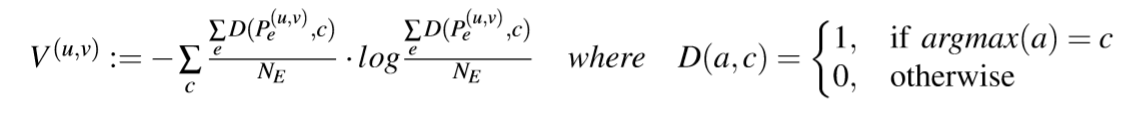
**Figure: Diagram for CEREALS**

In the method that they proposed to design acquisition to explicitly focus on image regions inside of the entire unlabeled pool of images and father to not only considered information during reason selection but also annotation costs. The proposed method works as follows:

**1) Training:** For seed construction they uniformly took N images and fully labeled them by an Oracle. Here they work on two deep constitutional neural network one is semantic segmentation model based on FCN8s architecture semantic segmentation which is based on Cost model.

**2a) Information Extraction:** When they raise awareness towards costs. They compared to classical heterogeneous information measures only. Entropy is the most widely used information measure seen in active learning literature. Entropy's value is maximized when the model assigns each considered class the same Probability and very small if the model is sure about its decision. Entropy computation for retrieving per-pixel information as follows:

The Vote Entropy information measure entails ﬁrst constructing a committee E of NE different classiﬁers that ideally are all consistent with the labeled pool. Each committee member *e* places a vote on vector Pe(u,v) . Then a disagreement factor among the members is calculated. We utilize vote entropy which we adapt for the semantic segmentation case as follows:



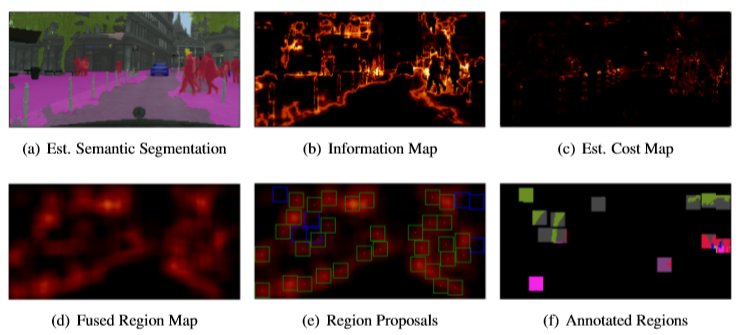
**2b) Cost Extraction:** It is the work of assumption some samples in an unlabeled pool are more costly to label by a human oracle than others. They are approximating costs by the number of clicks necessary to annotate an image. For cost extraction they perform a forward pass for each individual image 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 model’s 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].

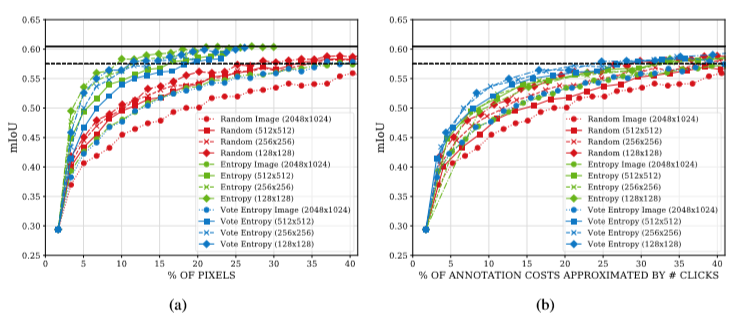


They have evaluated the three simple fusion function with the region information map I and the region cost map C. The parameter **Alpha** in (5) allows to set a trade-off for linearly interpolating between both region maps.

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



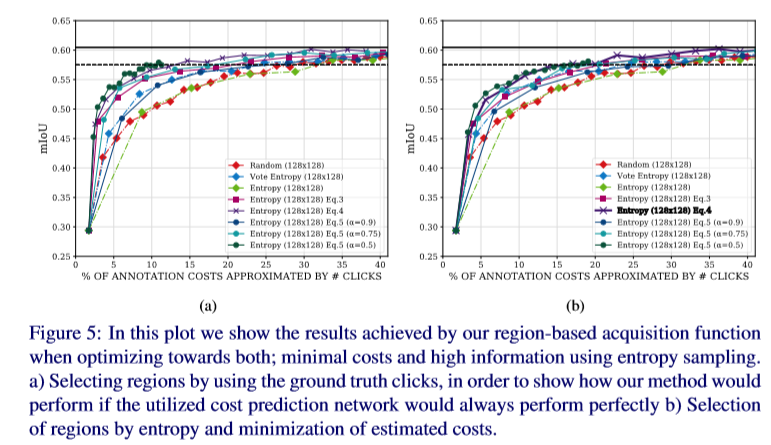
**3) Acquisition:** From the region proposal pool extracted as many top scoring regions as would correspond to extracting m images out of a pool of equally sized images regarding their amount of pixels for a fair comparison to the image-based acquisition of labels. They update the labeled and unlabeled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

**References:**

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