Region-Based Active Learning for Efficient Labelling in Semantic Segmentation

Tejaswi Kasarla[1], G Nagendar[1], Guruprasad Hegde[2], Vineeth N. Balasubramanian[3], C.V. Jawahar[1]

[1] Center for Visual Information Technology, IIIT Hyderabad, India
[2] Bosch Research and Technology Centre, India
[3] IIT Hyderabad, India

Annotations for Segmentation

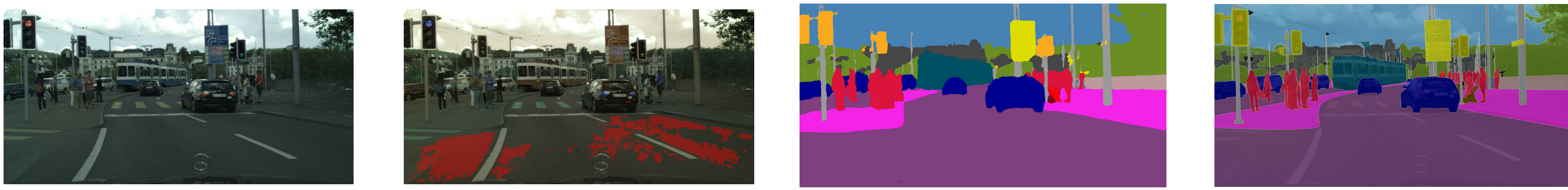
1024 x 2048

~ 1.5 to 2 hours for fine annotation

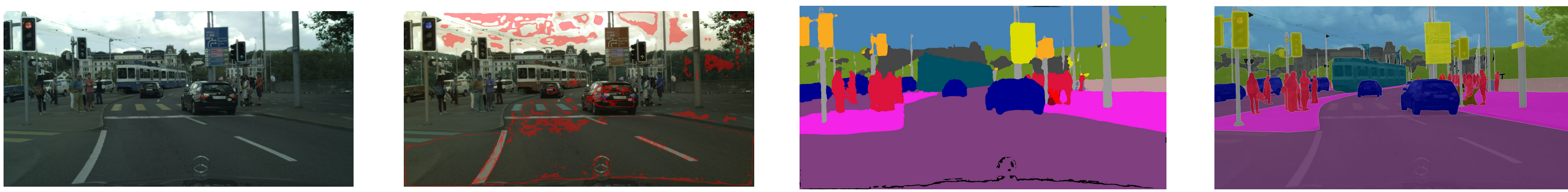
Expensive to obtain!

[1] Cordts el. al, Cityscapes Dataset, 2016

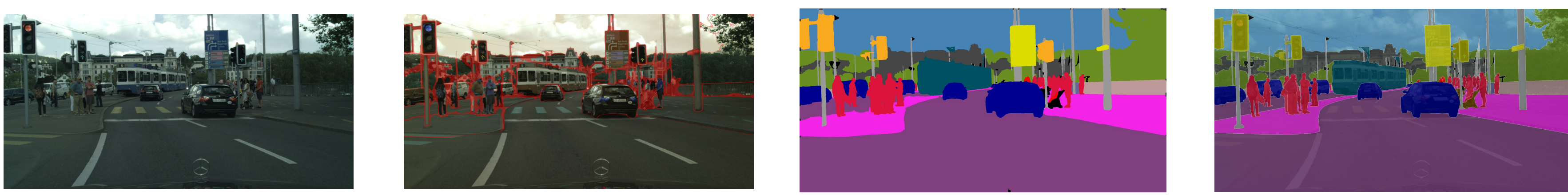
Getting less expensive annotation



Sampling Strategy 1



Sampling Strategy 2

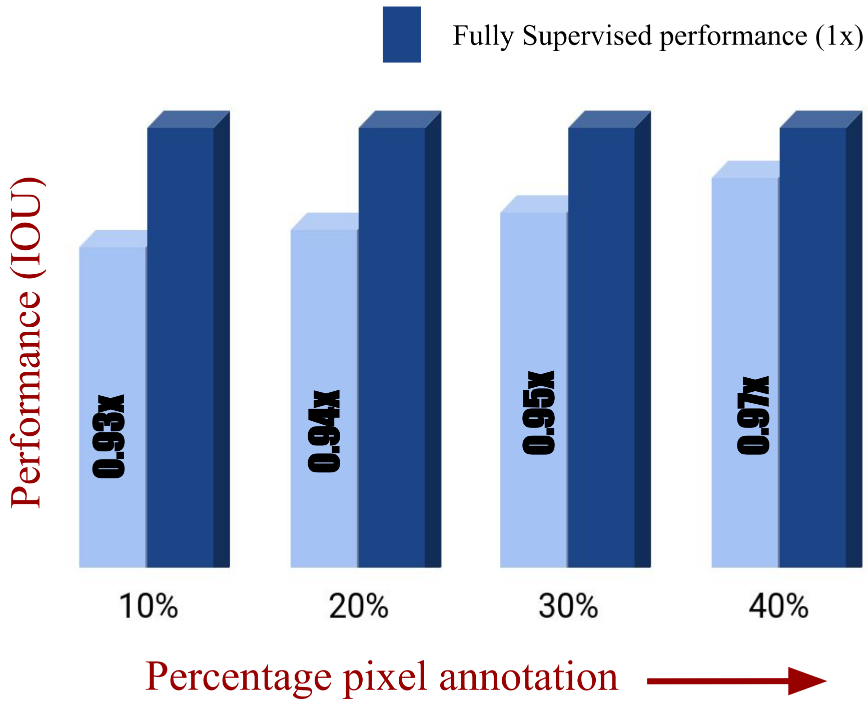


Our sampling strategy

The way we annotate matters!

Getting less expensive annotation

What if we have a way to get similar performance of full annotations with intelligently selecting the data points?



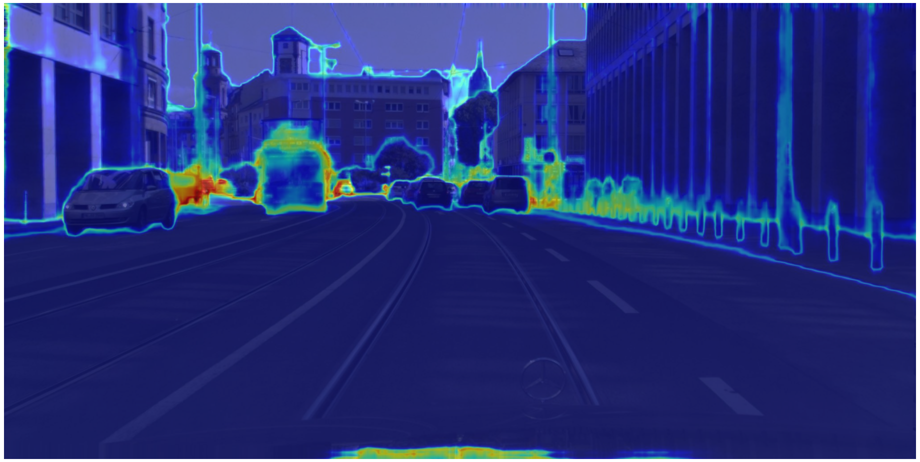
Active Learning is the answer!

[2] Burr Settles, Active Learning Literature Survey, 2009

Method

But how to intelligently select the data points?

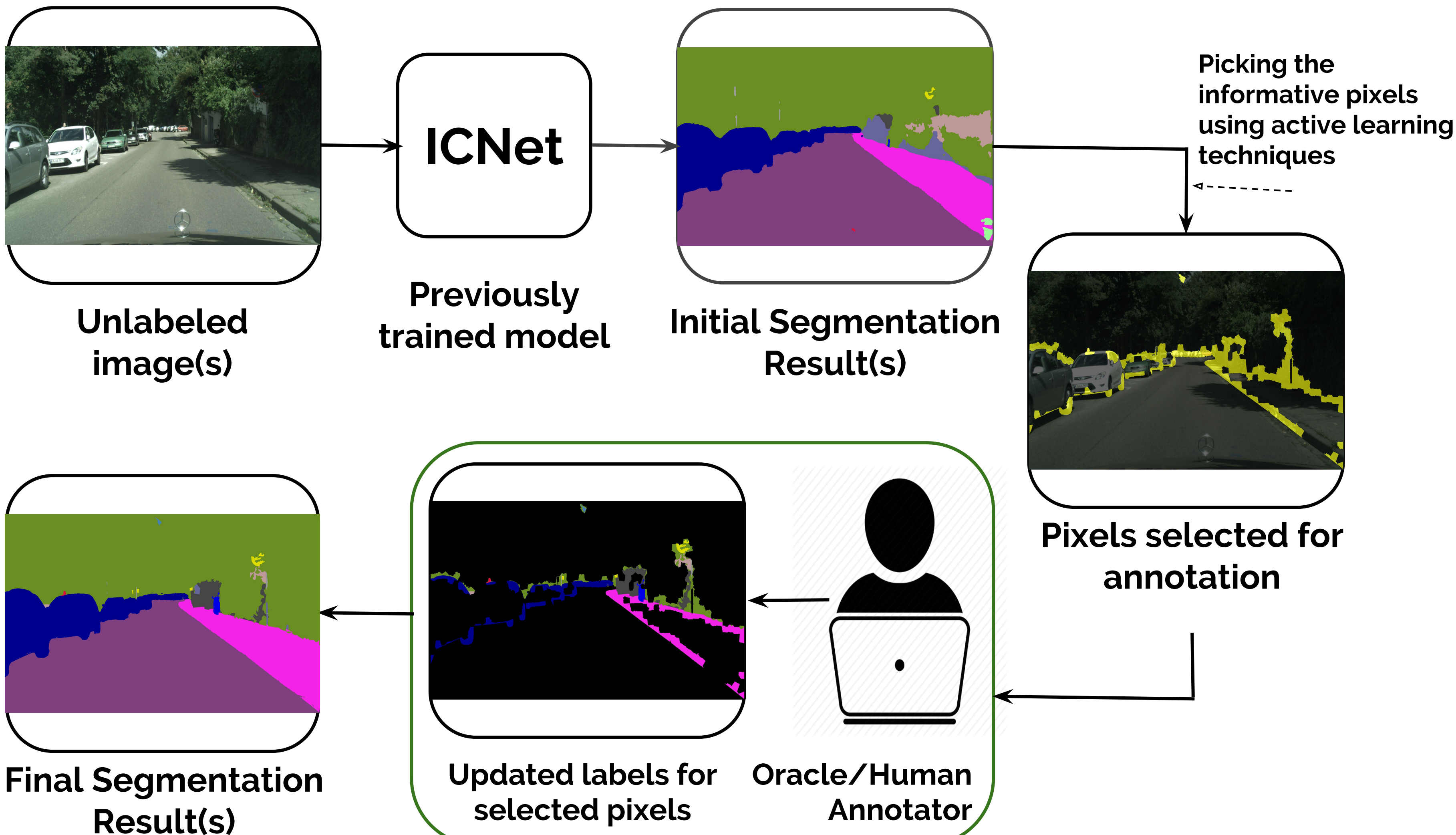
By finding the uncertain regions in the image and providing annotation for it.



Entropy:

$$H_i^j = \sum_{k=1}^{C} p(c_k | x_i^j, \Theta) \log(p(c_k | x_i^j, \Theta))$$

Pipeline


ICNet
Unlabeled image(s)
Previously trained model
Initial Segmentation Result(s)
Picking the informative pixels using active learning techniques
Pixels selected for annotation
Oracle/Human Annotator
Updated labels for selected pixels
Final Segmentation Result(s)

Proposed Active Learning Method

- **Pixel -** Obtain pixel entropies to pick the most m% uncertain pixels to query for annotation.
- **Edge + Pixel** - Gives more weightage to pick edge pixels and then picks most uncertain pixels
- **SP** - region based method – superpixels are annotated (instead of pixels)
- **SP + CRF** - CRF post-processing after annotating superpixel
- **Class-specific SP + CRF** - SP+CRF for each class separately

Results

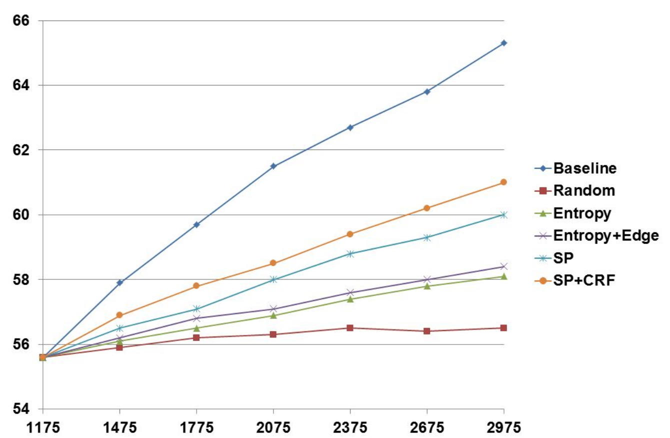
Datasets for evaluation: Cityscapes, Mapillary

DNN for the training: ICNet

Experimental Setting:
Cityscapes - 1175 fully annotated images (to train initial model) + 1800 unlabeled images (to be used for partial annotation with active learning)
Mapillary - 18000 unlabeled images

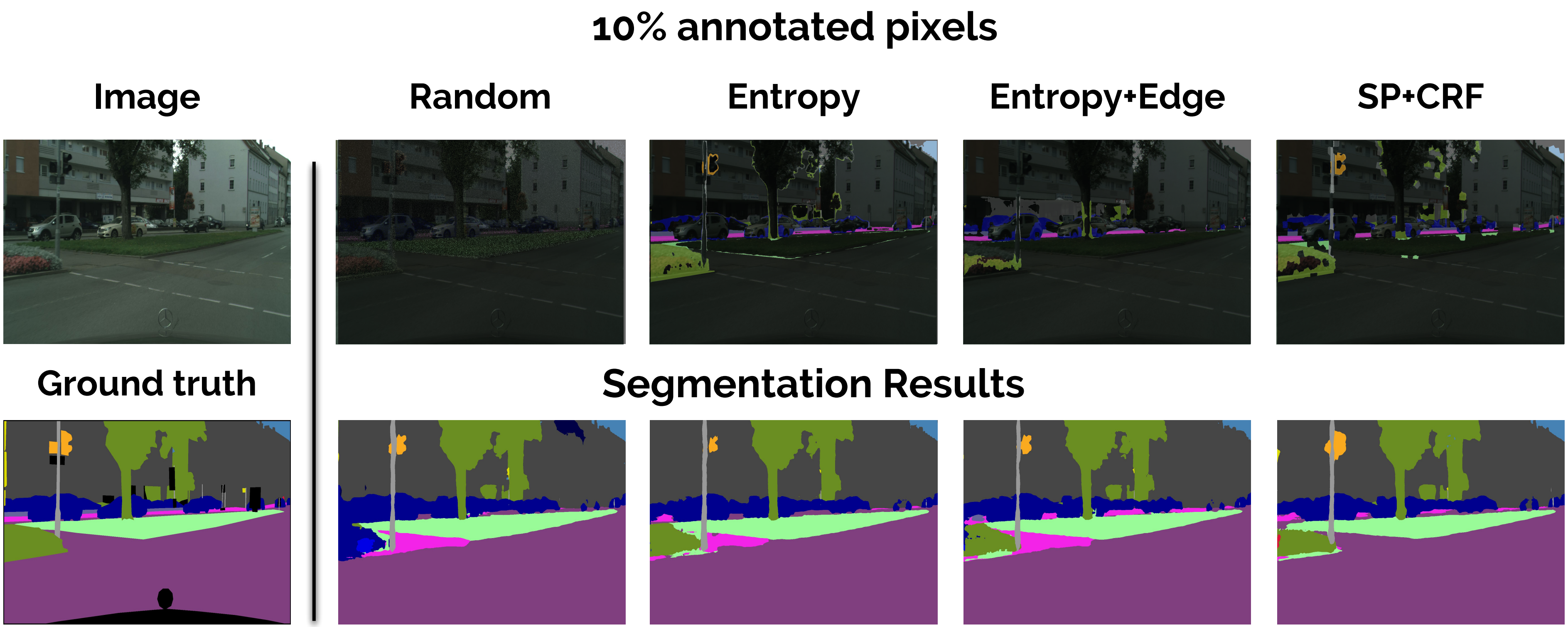
Results: Cityscapes

	Baseline	Random 10% GT	Entropy	Entropy + Edge pixels	SP	SP + CRF	Class-specific SP+CRF
# Training images	100% GT	10% GT					
1175	55.6	55.6	55.6	55.6	55.6	55.6	55.6
1475	57.9	55.9 (96.5%)	56.1 (96.8%)	56.4 (97.4%)	56.5 (97.5%)	56.9 (98.2%)	**57.0 (98.4%)**
1775	59.7	56.2 (94.1%)	56.5 (94.6%)	57.0 (95.4%)	57.1 (95.6%)	57.8 (96.8%)	**57.9 (96.9%)**
2075	61.5	56.3 (91.5%)	56.9 (92.5%)	57.9 (94.1%)	58.0 (94.3%)	58.5 (95.1%)	**58.7 (95.4%)**
2375	62.7	56.5 (90.1%)	57.4 (91.5%)	58.7 (93.6%)	58.8 (93.7%)	59.4 (94.7%)	**59.7 (95.2%)**
2675	63.8	56.4 (88.4%)	57.8 (90.5%)	59.4 (93.1%)	59.3 (92.9%)	60.2 (94.3%)	**60.4 (95.2%)**
2975	65.3	56.5 (86.5%)	58.1 (88.9%)	59.8 (91.5%)	60.0 (91.8%)	61.0 (93.4%)	**61.3 (93.8%)**



Performance of the proposed active learning methods over incremental selection of batches on cityscapes data.

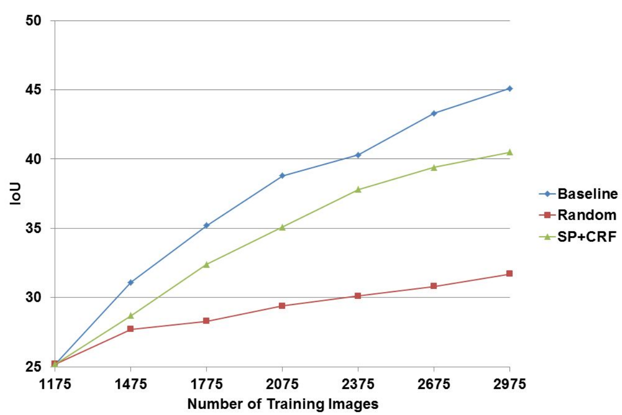
Results: Cityscapes



Sampled active learning (super) pixels and their corresponding segmentation results of various methods after training.

Results: Mapillary

# Training images	Baseline	Random	SP + CRF
	100% GT	10% GT	
Cityscapes - 2975	25.2	25.2	25.2
3000	31.1	27.7 (89.0%)	**28.7 (92.2%)**
6000	35.2	28.3 (80.3%)	**32.4 (92.0%)**
9000	38.8	29.0 (74.7%)	**35.1 (90.4%)**
12000	40.3	29.8 (73.9%)	**37.8 (93.7%)**
15000	43.3	30.3 (69.9%)	**39.4 (90.9%)**
18000	45.1	30.6 (67.8%)	**40.5 (89.8%)**

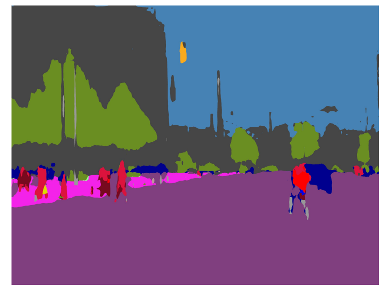


Performance of the proposed active learning methods over incremental selection of batches on mapillary data.

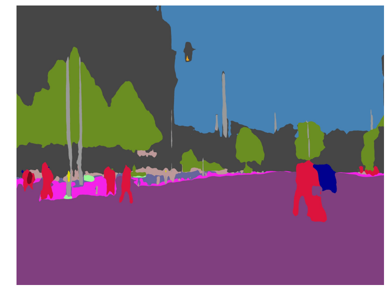
Results: Mapillary

Image

Groundtruth



No additional labels



Using 10% labelling

Segmentation results using transfer learning on mapillary data.

Qualitative Results

Proposed Method **Fully Supervised Method**

The results of the proposed region-based active learning method and the fully supervised method are visually almost similar.

Thank you.