Region-Based Active Learning for Efficient Labelling in Semantic Segmentation

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

[1] CVIT, IIIT Hyderabad, India [2] Bosch Research and Technology Centre, India [3] IIT Hyderabad, India

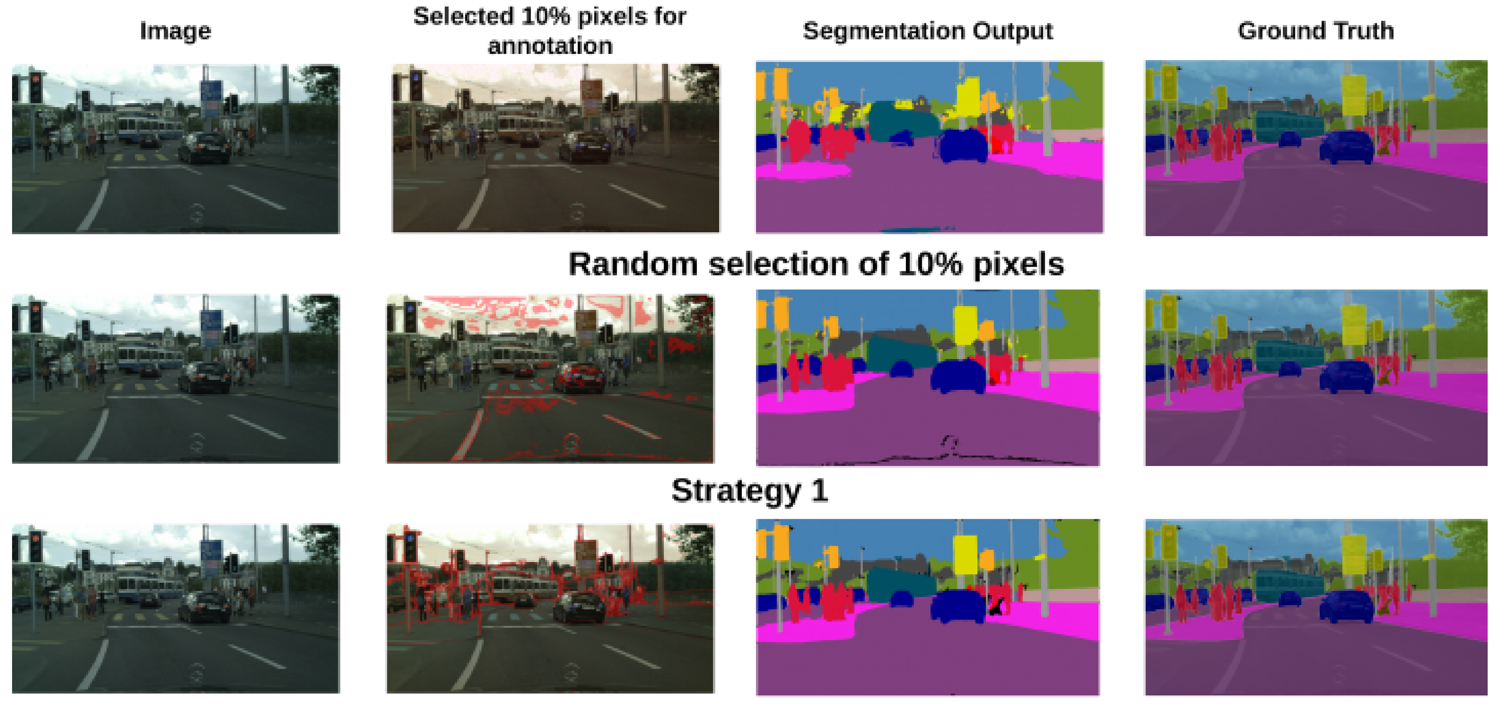
Motivation

Pixel level annotations in semantic segmentation

Annotations are expensive to obtain for large data.

Goal: To reduce annotation effort for semantic segmentation and get reasonable segmentation results

Challenges



Performance depends on the selected pixels for Annotation

Contributions

- ❖ Proposed an active learning strategy for selecting most uncertain pixels for annotation
- ❖ Uncertainty of images/pixels is computed using entropy

10% annotation of image pixels/superpixels gives **90-93%** of fully supervised performance

Email: kasarla.tejaswi@research.iiit.ac.in

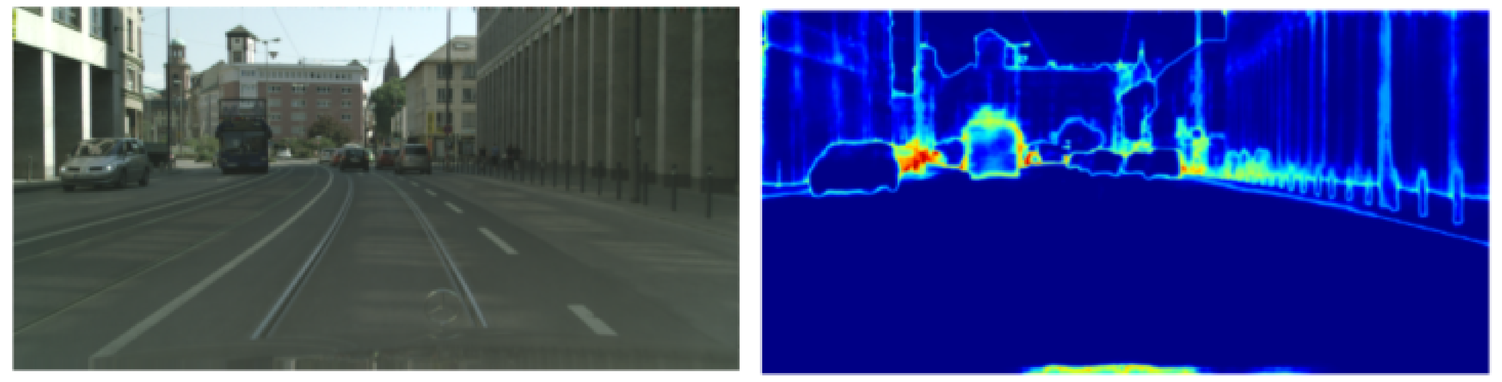
Proposed Method

Entropy : Gives the measure of uncertainty.

$$S = -\sum_i p_i \log(p_i)$$

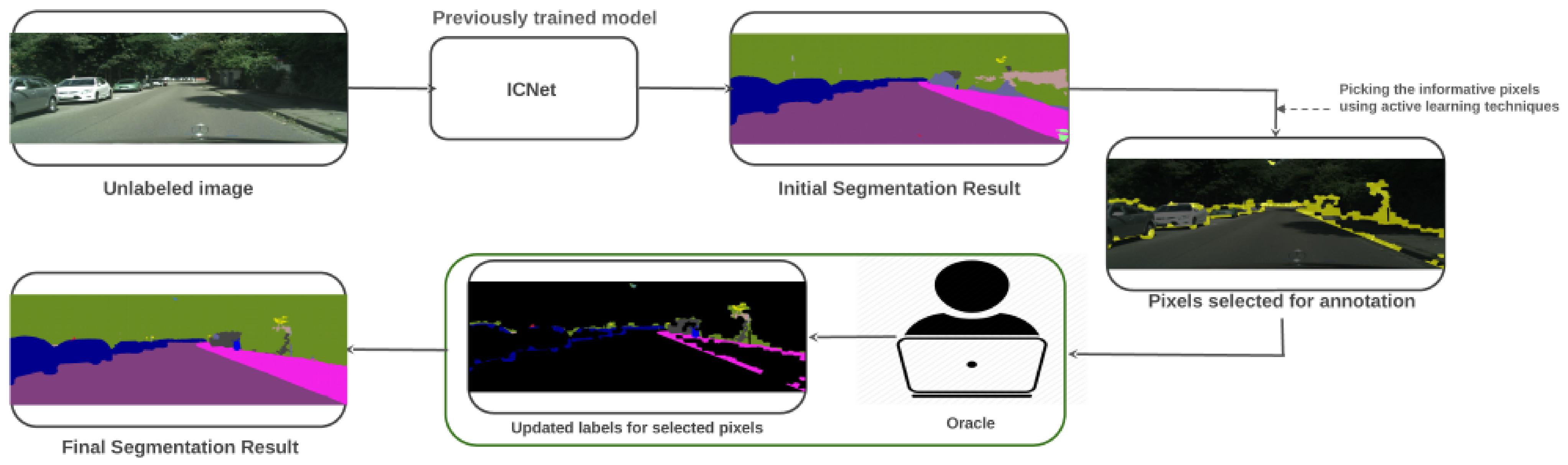
Entropy for a pixel x_i^j

$$H_i^j = \sum_{k=1}^{C} p(c_k|x_i^j,\theta)\log(p(c_k|x_i^j,\theta))$$ C is the number of classes



Pipeline

- Given data X=PUQ. P is labeled data and Q is unlabeled data
- Θ is the deep learning model trained on P



Input: X, Θ and oracle O
Output: Updated model Θ

- Given an unlabeled image, compute its probability score map using Θ
- Compute most uncertain pixels using entropy
- Annotate the uncertain pixels using Oracle O
- Retrain the network Θ using new annotations

Proposed Strategies for Computing the Uncertainty

- Entropy: Computes pixel level uncertainty
- Entropy + Edge: Gives higher weightage to edge pixels
- Superpixels (SP): Computes uncertainty at superpixel level
- SP + CRF: Improves the ‘SP ‘ method performance using CRF
- Class-Specific SP + CRF: Handles class imbalance issues

Results

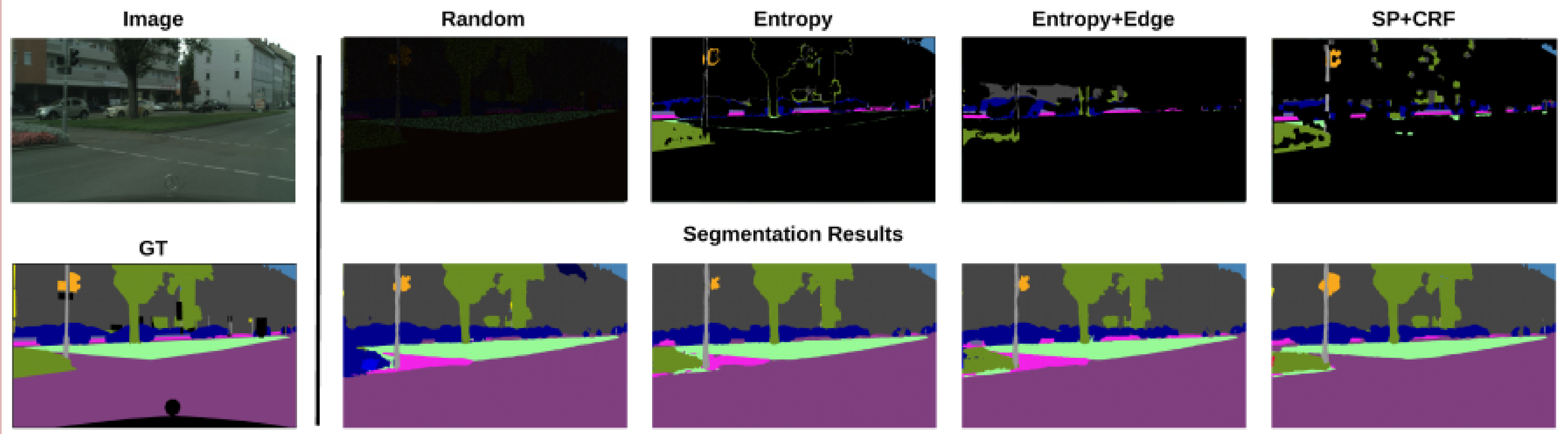
Datasets and Experimental Settings

- **CityScapes:** 2975 images (1175 - labeled + 1800 - unlabeled)
- **Mapillary:** 18000 images
- **Deep Learning Network:** ICNet

Results On Cityscapes Dataset

	Baseline	Random 10% GT	Entropy	Entropy + Edge pixels	SP	SP + CRF	Class level 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%)

Qualitative Results



Results On Mapillary Dataset

# 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%)**