WASEDA University
Graduate School of Information, Production and Systems

Semantic Segmentation Refinement Using Entropy and Boundary-guided Monte Carlo Sampling and Directed Regional Search

Zitang Sun(1112), Image Media Lab, Waseda University.

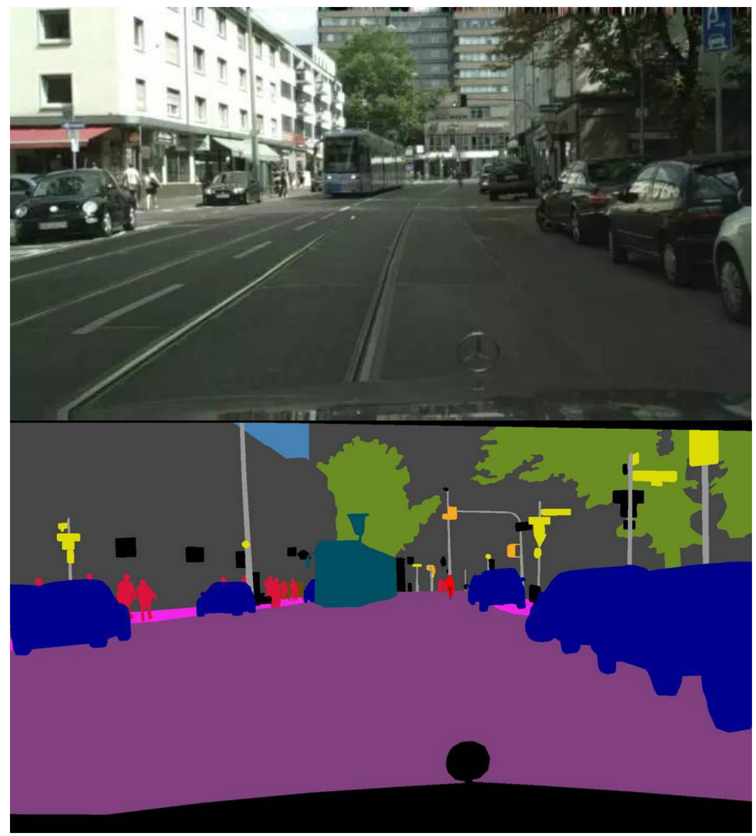
Outline

I. Introduction

Background

- Application: Autonomous Driving
- Segmentation network automatically analysis objects in input images to guide cars drive along the road and avoid obstacles such as pedestrians and vehicles.

Car-Robo course: autonomous car



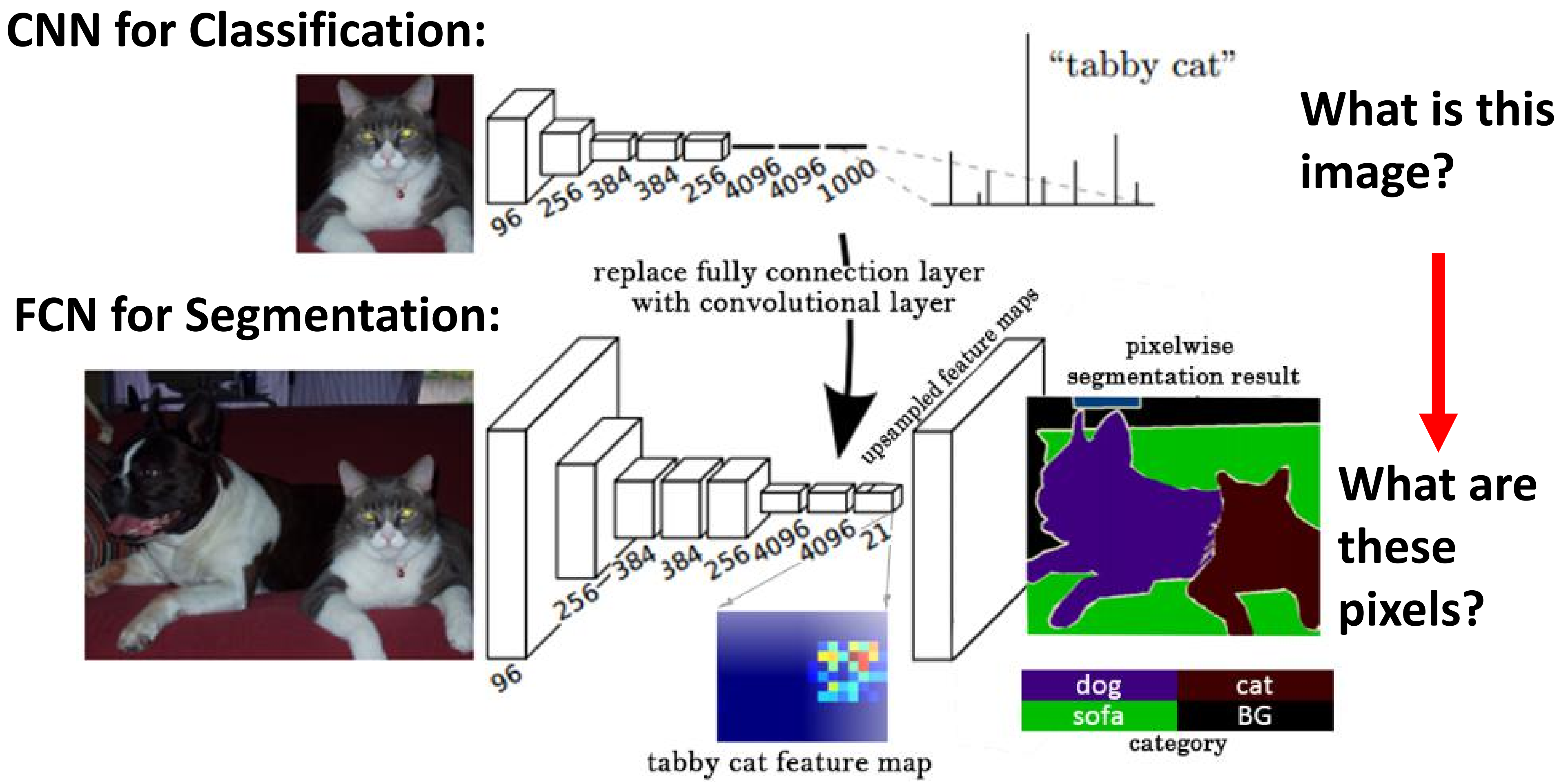
Real world RGB image from Cityscape[2]

Semantic segmentation

[2] M. Cordts.The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.

Existing works

In 2015, Fully Convolutional Network[3] first introduce convolutional neural network(CNN) for semantic segmentation.



Existing works

Paper	Proposal	Baseline	Remarks
FCN[3], 2015 CVPR	Fully convolutional	VGG19	Firstly introduce CNN for semantic segm
Deeplab[5], 2016 ICLR	Dilated convolution	VGG19	Dilat pooling
RefineNet[10], 2016 CVPR	Residual convolution	Resnet101	Long-range Residual Connection
PSPNet[6], 2017 CVPR	Pyramid pooling	Resnet101	Apply
G-SCNN[7], 2019 ICCV	Gated Conv.	W-Res101	Use double branches to learn both semantic edge and segmentation
DANet[9], 2019 CVPR	Attention mechanism	Res-101	Construct dual attention block for Deco
SpyGR[8] 2020 CVPR	Graph Conv.	Res-101	Establish spatial pyramid for graph reasoning for decoder

FCN-based Method

Standard Encoder (Resnet, VGG, etc.)

Improving Decoder

Problem

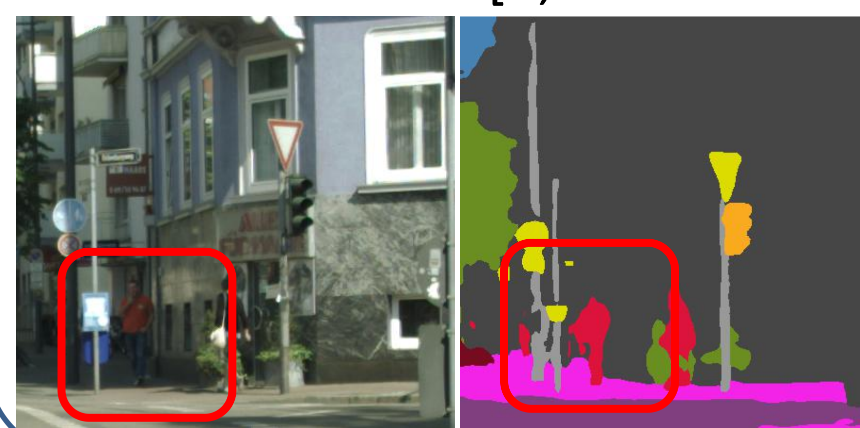
◆ Bottlenecks of existing work

The existing FCN backbone always ignores details such as contour and edges of the image.

- Small objects and details important but take small proportion in image such as : Pedestrian, rider, telephone pole, traffic sign pole
- Large inner region objects provide useless information to network, but it always takes up a lot of pixel proportions.

Result of DANet[9, 2019 CVPR]

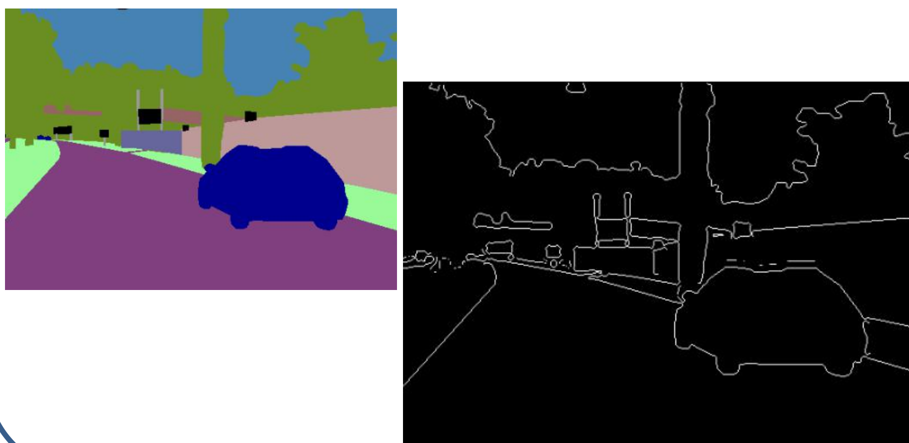
Result of PSPNet[6, 2017 CVPR]



Work Direction

- Design a novel **algorithm** that can be independently embedded into the backend of network to optimize results generate by FCN-based network.
- Focus on optimizing error-prone small objects and edge regions.
- Two important factors in this work:

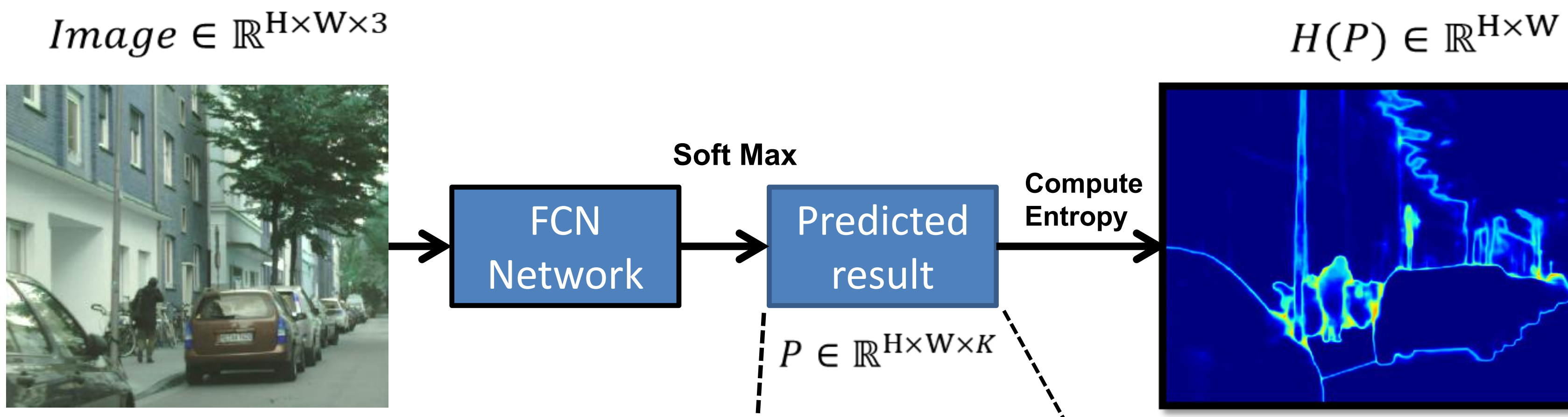
① Semantic boundary

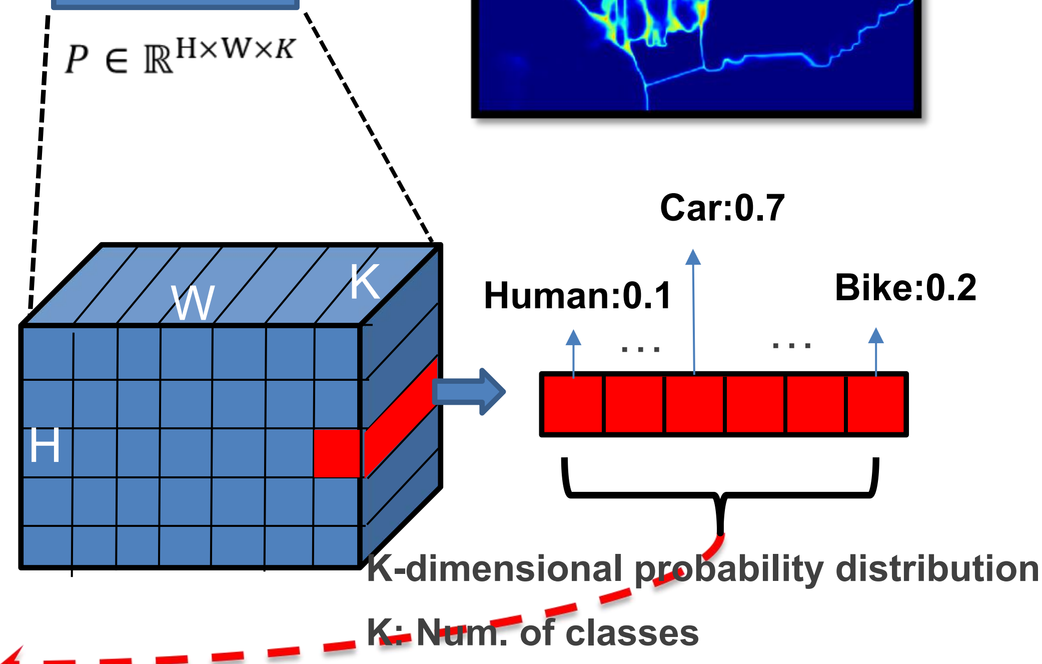


② Information entropy

Information Entropy

- Information entropy in Semantic segmentation.





FCN will generate a K-dimensional vector for each pixel describing the predicted probability.

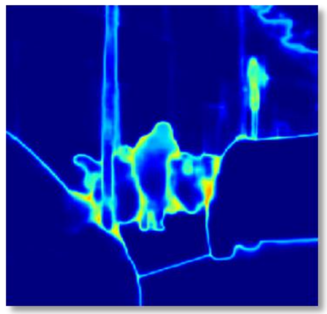
Calculate Entropy:

$$\mathrm{H}(p_{ij}) = -\sum_{k=1}^{K} \hat{p}_{ij}(k) \log \hat{p}_{ij}(k)$$

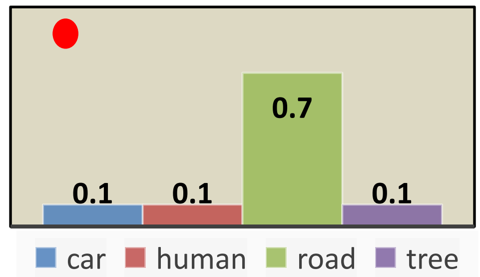
$\hat{p}_{ij}(k)$:The probability of the pixel P_{ij} in Kth category.

Information Entropy

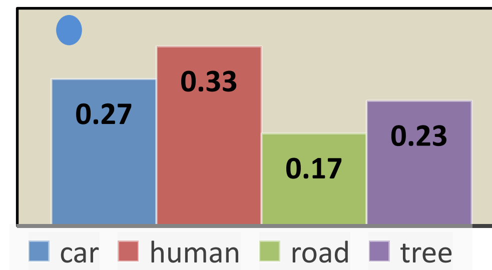
- Entropy can well reflect the confidence of each predicted result from the probability distribution generated by network.



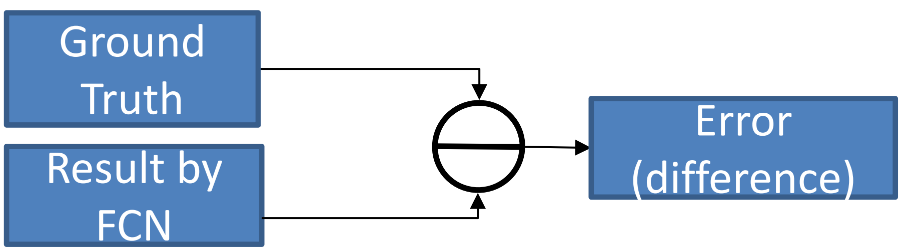
High-confidence (low-entropy)



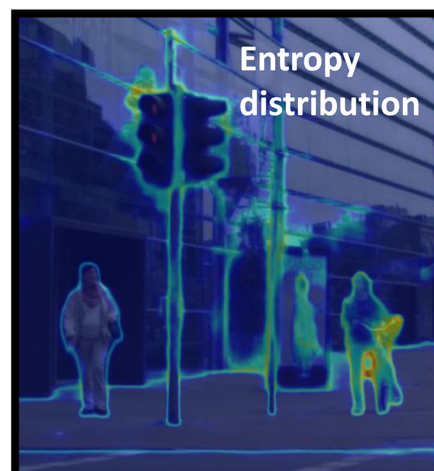
Low-confidence (high-entropy)

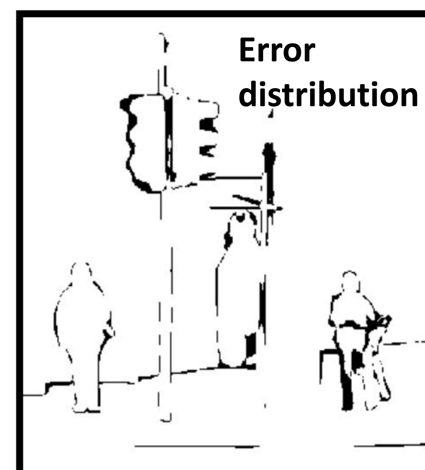


- It's easy to understand that the entropy energy always concentrate on the low-confidence regions, such as edge and regions with tiny objects.



- High-entropy (low-confidence) regions are always accompanied by false predicted results





II. Proposed Model

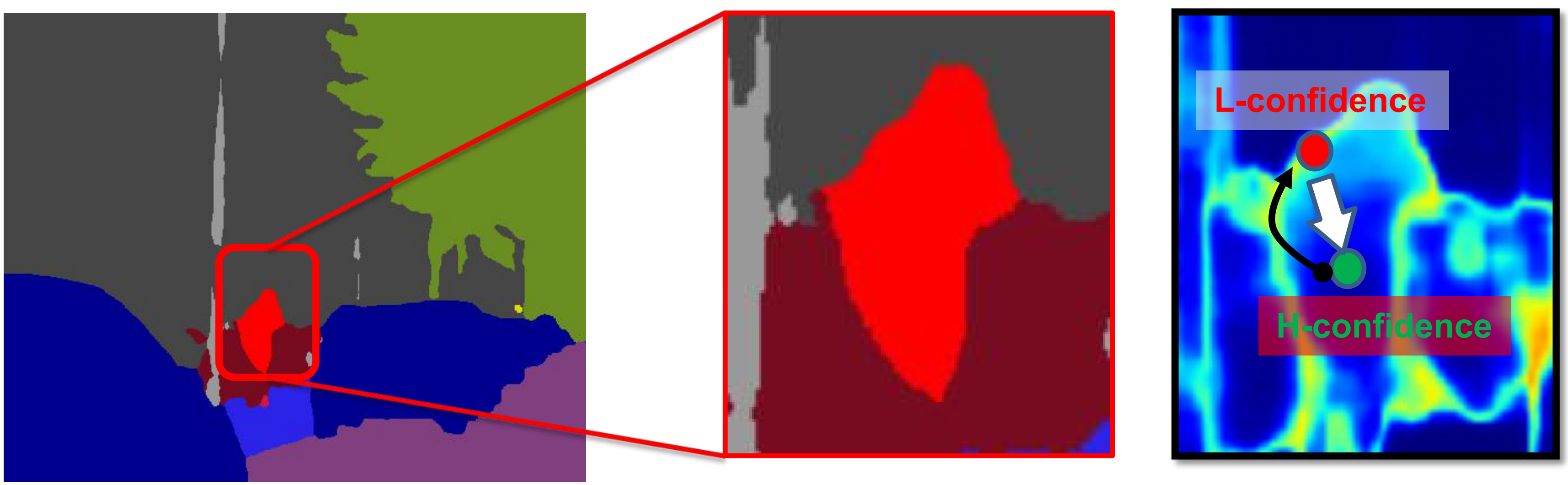
Method: Concept

- Capture and improve the low-confidence regions.

① Capture the error regions (edge and low-confidence regions)

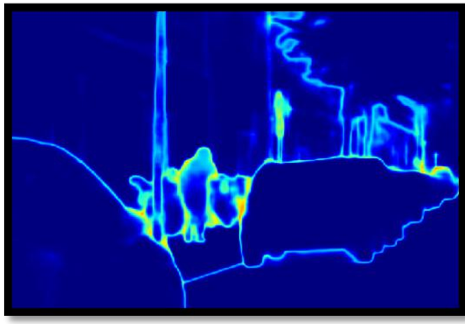
② Move inside and search for high-confidence regions nearby

③ Judgment and Determine whether the H-confidence results should be adopted to replace L-confidence results.



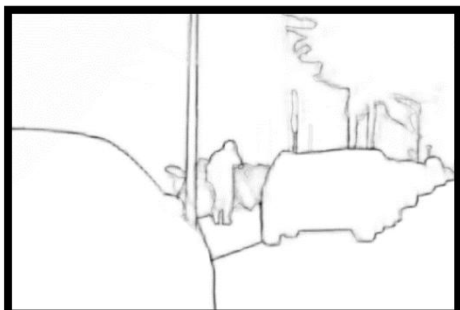
Prediction Entropy

Method: Capture low-confidence regions



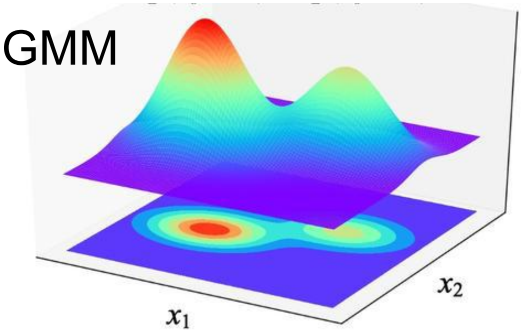
entropy

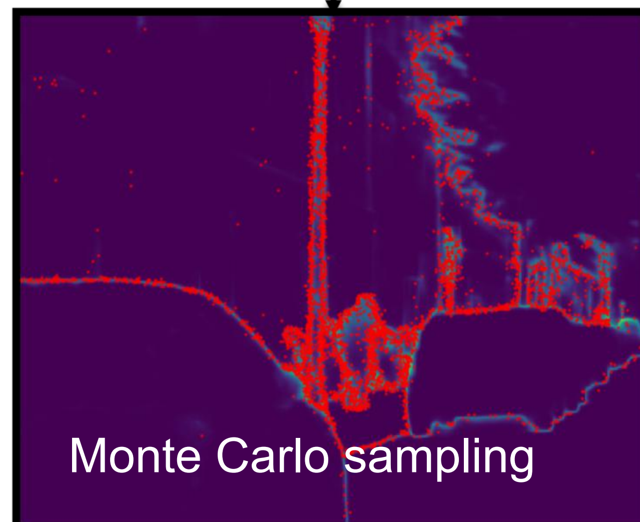
Weighted average



boundary

GMM





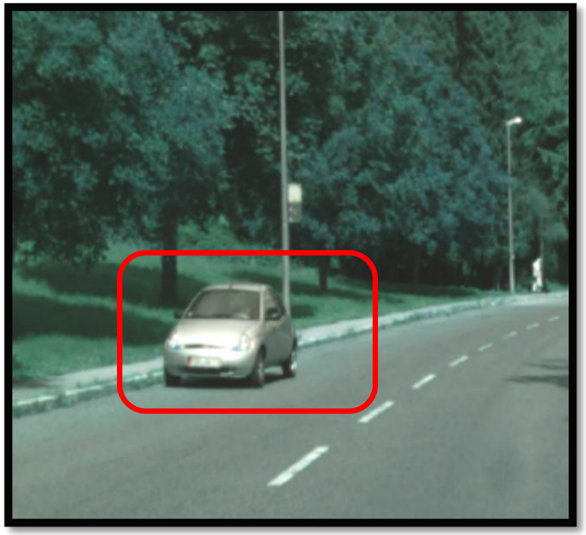
- Given the entropy distribution, we establish a Gaussian Mixture model (GMM) and Capture low-confidence pixels through Monte Carlo Sampling.

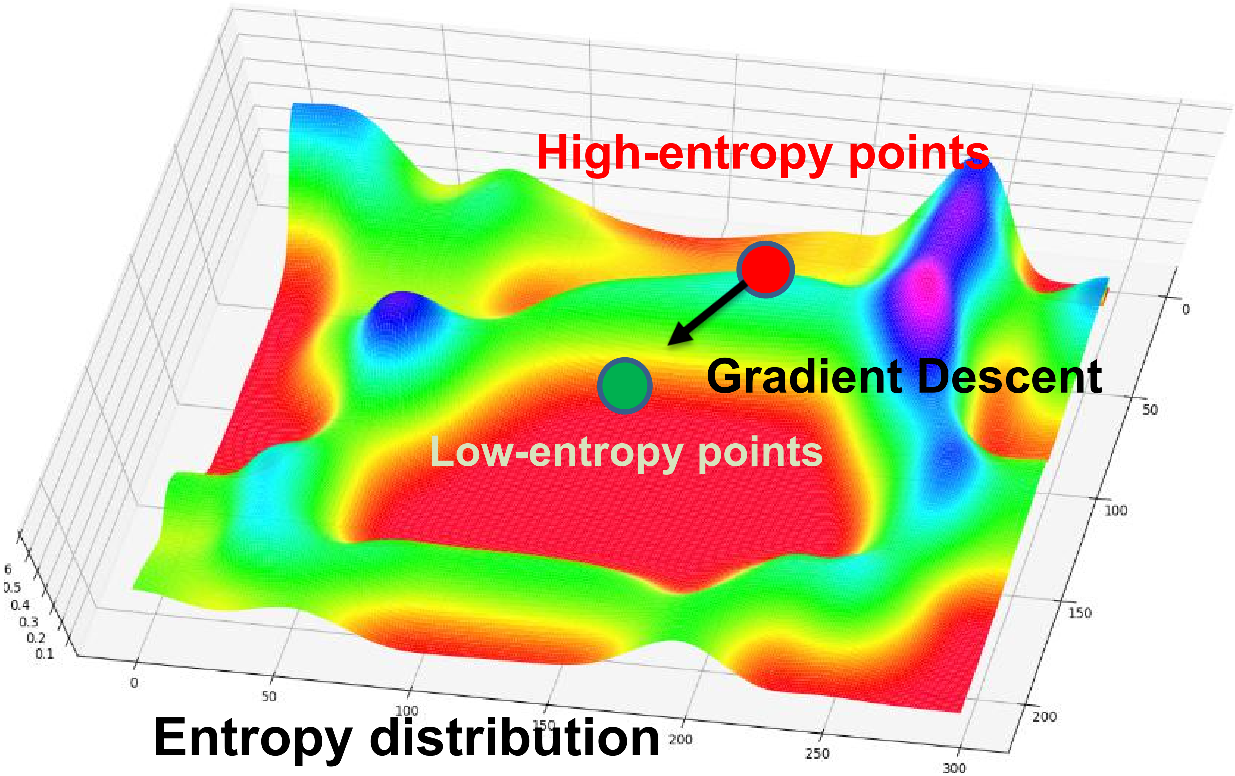
$$P_{GMM}(x)=\sum_{i=1}^{H}\sum_{j=1}^{W}\eta_{ij}N(\mu,\Sigma), s.t. \sum_{i=1}^{H}\sum_{j=1}^{W}\eta_{ij}=1$$

η_{ij}: entropy and boundary weighted value at each pixel *(i,j)*

Method: Search Strategy

- Using information entropy as clues to perform gradient descent.
- Find low-entropy regions and adopts their predicted labels.

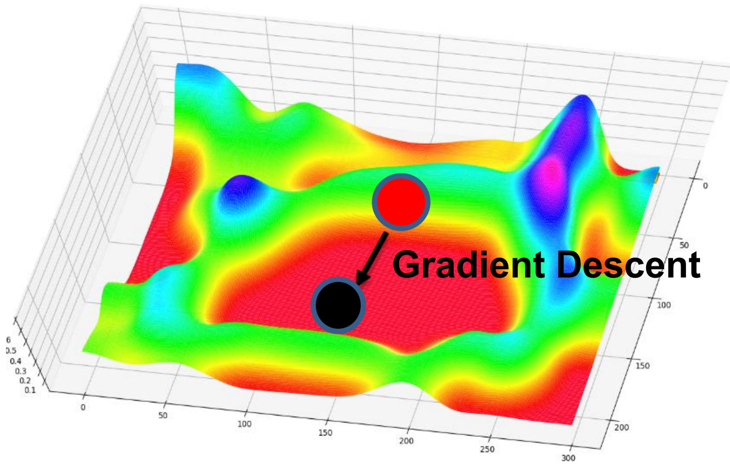




Entropy distribution

Search Strategy

- Using **semantic boundary** as clues to restrict search regions.



- If we far from the boundary region, using the gradient direction of information entropy (search for low-entropy regions)

- Alternate semantic boundary and entropy are used for gradient descent

① If we near the boundary region, using the gradient direction of **semantic boundary** **(** Prevent move into other region of different classes)

Majority voting and final classification

- Set reasonable judgement criteria based on **Majority voting** to avoid misclassification.

If we go through N points, Labels are replaced only if the following conditions are met:

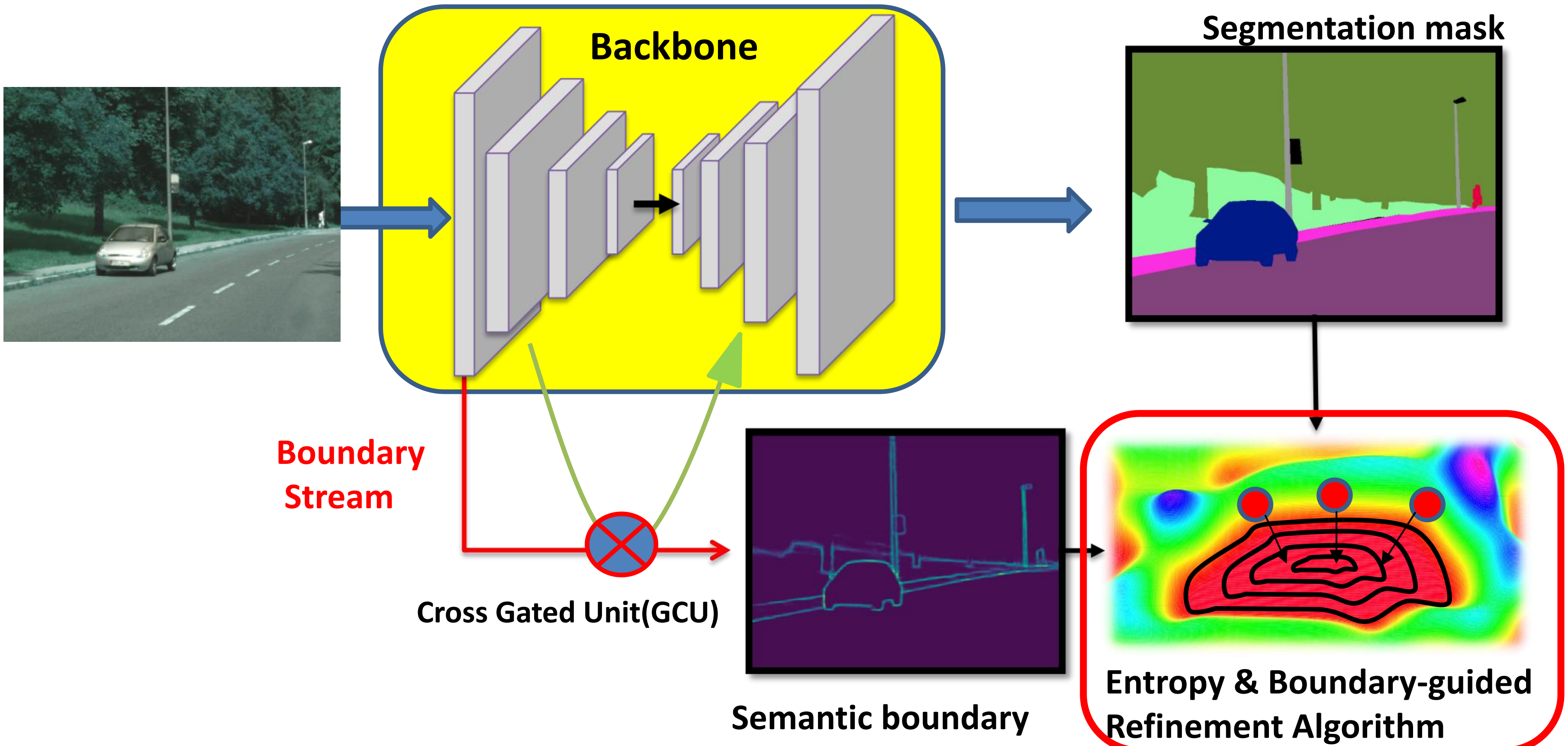
1. The first α% of the lowest entropy points have the same label ***A***
2. More than β% of the N points with label ***A***
3. $D_{KL}(\text{head}, end) < Threshold$

Then we will assgin N points the same label ***A***.

- The hyper-parameters of the above criteria are obtained after a large number of experiments on the specific dataset.

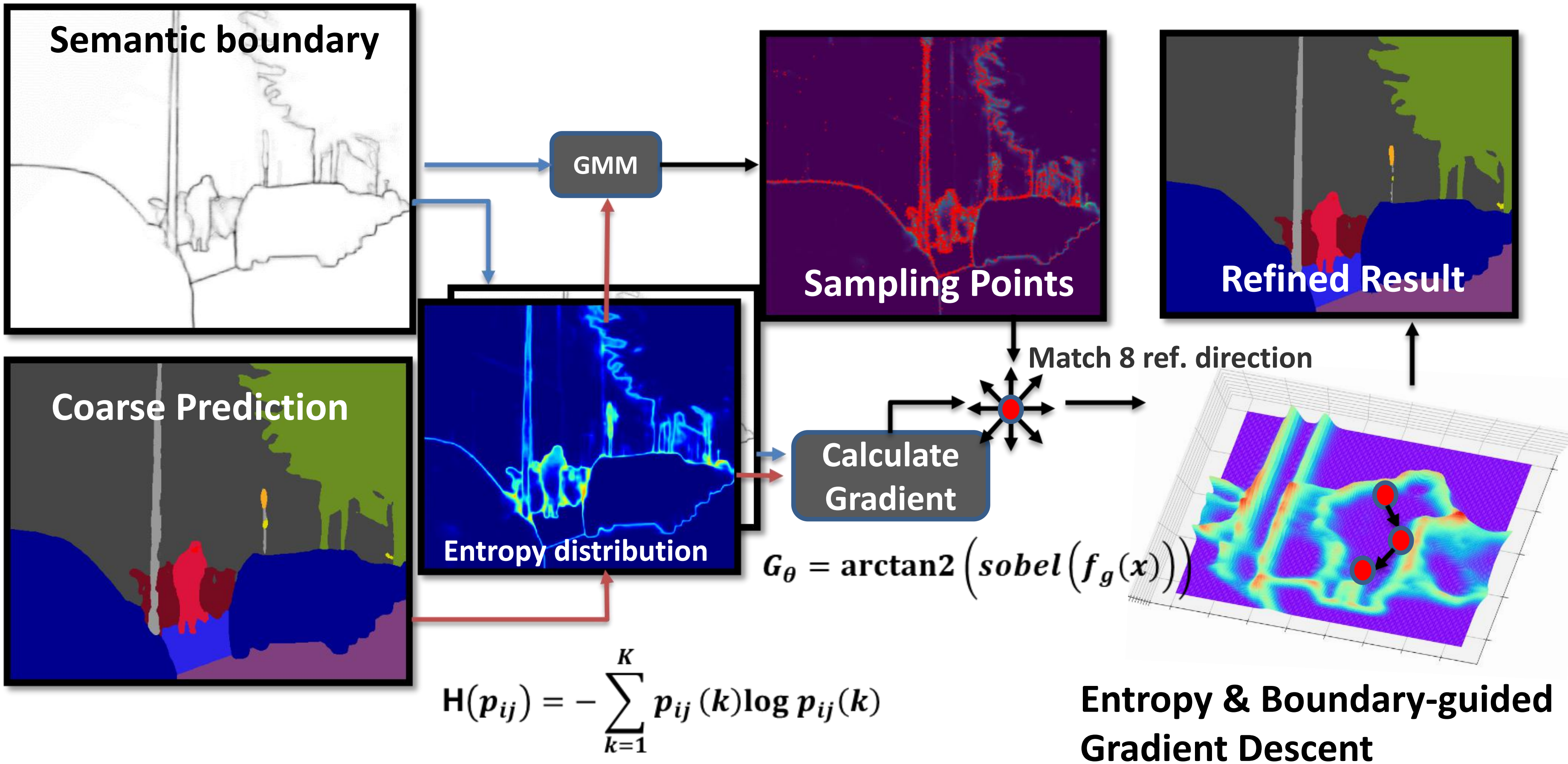
Generate Semantic boundary

- Inspired by Gated-SCNN[7], we added an extra branch to the original backbone to get the semantic boundary in advance



Whole Process

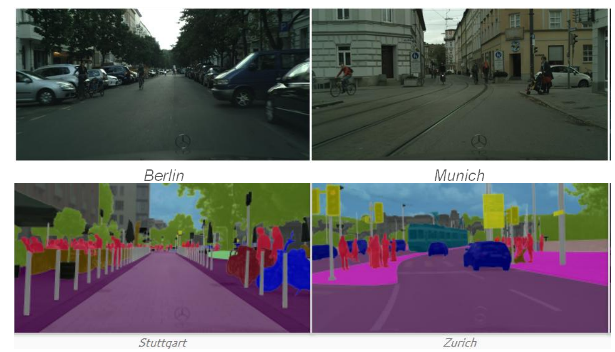
- The whole process about our ***Entropy and Boundary-guided Refinement*** (EBR) algorithm.



III. Experiments

Dataset

Cityscapes



◆ Cityscapes Dataset[2]

- a urban street scene dataset from cars perspective
- 19 semantic classes like road, person, vegetation, car.*etc*.

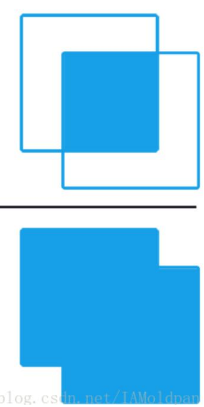
◆ PASCAL VOC 2012[1]

- A standardized dataset for image recognition, classification and segmentation.
- 4369 images with 20 semantic categories, including person, airplane, animal. *etc*.

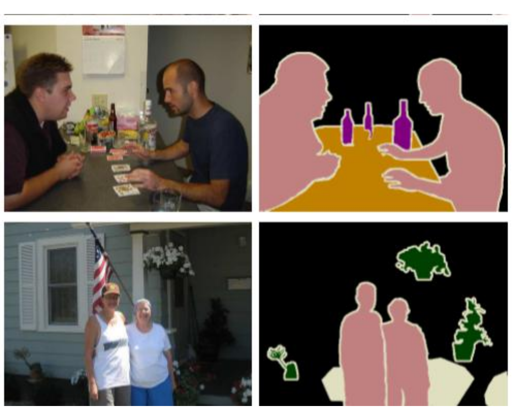
◆ Evaluate Criterion

- Intersection-over-Union (IoU)

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

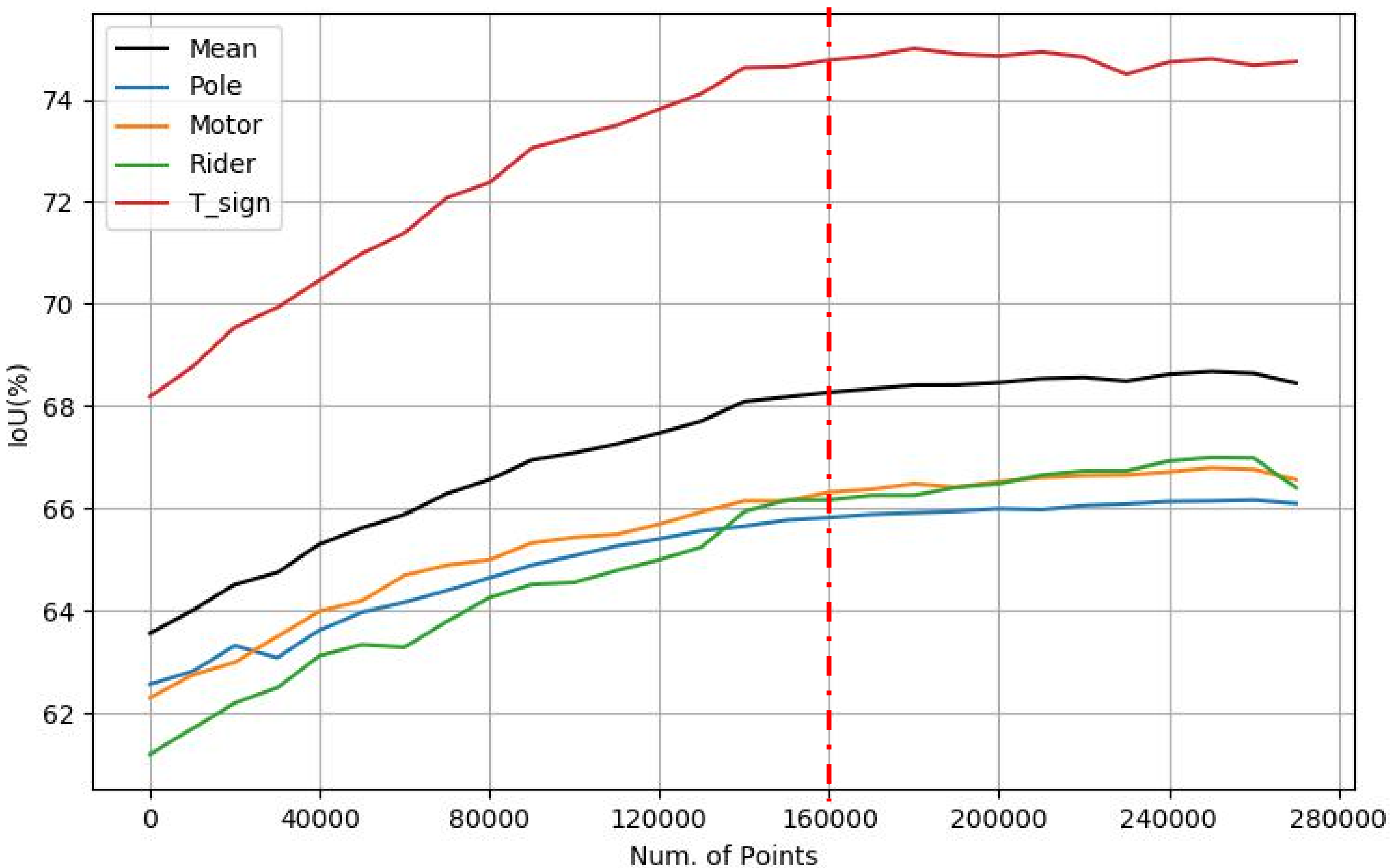


PASCAL VOC



Experiment Results

- Evaluation of our algorithm with increasing sampling points.
- (Backbone : lightweight-RefineNet[10] with ResNet50; Dataset: Cityscapes)



Ablation Study

- Baseline: lightweight-RefineNet with ResNet-50; Sample 6.5% points for refinement.
- Mean IoU evaluation on Cityscape val dataset.

Method	mIoU	Car	Road	motor	bike	pole	T-sign	fence
Baseline	74.3	94.3	98.2	62.3	71.4	60.6	68.2	61.6
+EBR	78.8	93.7	97.1	66.7	74.5	64.3	74.2	66.2
Difference	+4.5	-0.6	-1.1	+4.4	+3.1	+3.7	+6	+5.1

- Combination of the state-of-the-art methods on Cityscapes and VOC datasets.
- backbone: ResNet101, (SS: single-scale inference; MS: multi-scale inference)

Method	Cityscapes				VOC 2012			
	SS		MS		SS		MS	
	base	+EBR	base	+EBR	base	+EBR	base	+EBR
[11]D.L.V3+[,2018 ECCV]	81.0	**81.6**	82.1	82.4	78.6	**79.5**	79.5	**79.7**
[10] D.L.V3[,2017]	79.7	80.6	80.8	81.2	77.9	78.9	79.2	79.4
[6] PSPNet[,2017 CVPR]	79.7	81.2	80.9	82.0	77.7	78.2	79.1	78.9
[9] DANet[,2019 CVPR]	80.5	81.4	82.0	**82.5**	76.8	78.4	77.3	79.1

Experiment Results

- Visualization Result

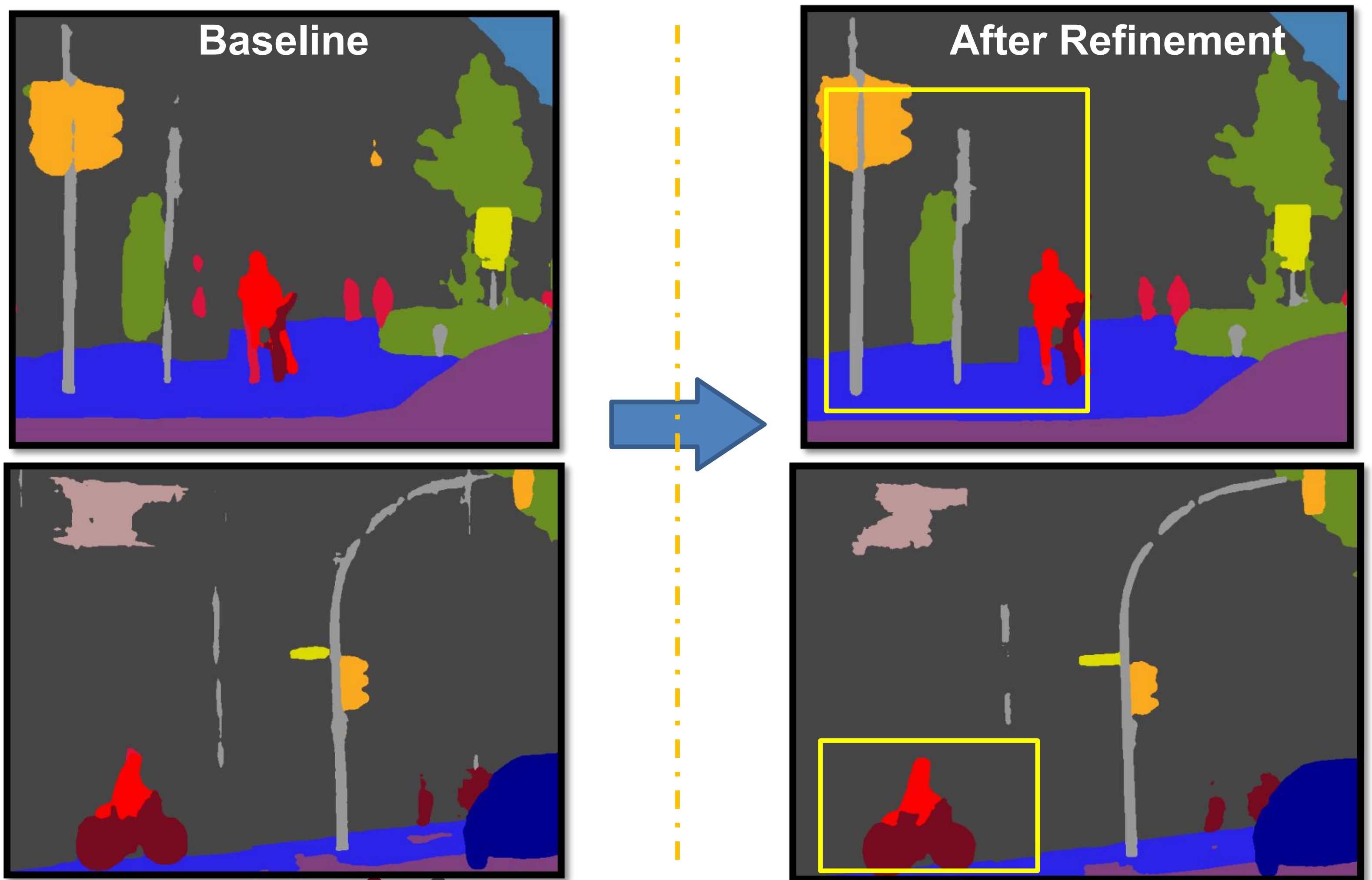
FCN output

Sampling 10K points

Sampling 80K points

Experiment Results

- Visualization Result



IV. Conclusion & Future works

Conclusion & Future Extension

◆ Conclusion

- We propose a novel ***Entropy and Boundary-guided Semantic Segmentation Refinement (EBR) Algorithm*** to improve the segmentation accuracy significantly on small objects.
- We analyze the meaning of ***Information Entropy*** in semantic segmentation task and skillfully apply it to our algorithm.
- In experiments, our method can be flexibly embedded in most existing networks to improve the performance on standard datasets. When combined with DANet, our method achieved a score of Mean IoU 82.5% at Cityscapes, ranking among the best in the existing methods

◆ Future Extension

- Try to combine it with deep learning to form a network that can be optimized end-to-end.

Partial Reference

[1] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. In IJCV, 2010.
[2] M. Cordts, M. Omran *et al.*, "The cityscapes dataset for semantic urban scene understanding," *CoRR*, vol. abs/1604.01685, 2016. 4
[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
[4] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. CoRR, 2015
[5]C. Liang-Chieh, G. Papandreou, I. Kokkinos, K. Murphy, Semantic image segmentation with deep convolutional nets and fully connected crfs, in ICLR, 2015.
[6] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network, Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[7] Towaki Takikawa, David Acuna, Varun Jampani, Sanja Fidler, Gated-SCNN: Gated Shape CNNs for Semantic Segmentation, in ICCV 2019.
[8] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, and H. Liu, "Spatial pyramid based graph reasoning for semantic segmentation," in CVPR 2020. 1, 2, 3.3, 4.4, 4, 5, 6, 4.5
[9] Jun Fu, Jing Liu, *et al.*, Dual Attention Network for Scene Segmentation, in CVPR 2019.
[10] Vladimir Nekrasov, Chunhua Shen, Ian Reid, Light-Weight RefineNet for Real-Time Semantic Segmentation, In BMVC 2018.
[11] Liang-Chieh Chen, Yukun Zhu, *et al*, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, In ECCV 2018.

Thank You！

Q & A

Appendix-EBR Algorithm

Algorithm 1 Entropy and boundary-guided refinement.

Input: The coarse segmentation label of CNN, P; The entropy distribution, H; The Semantic boundary, B;

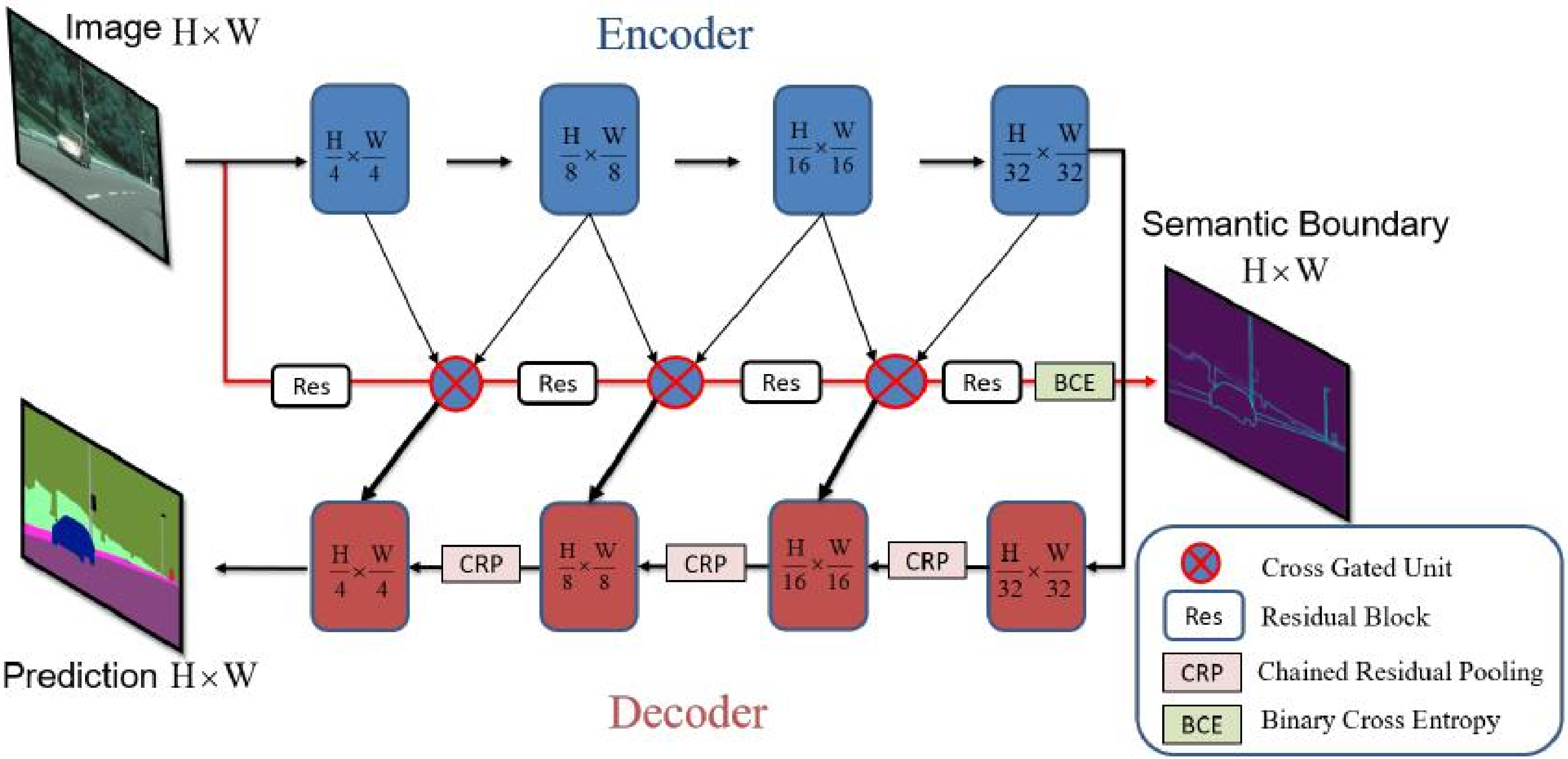
Output: Refined segmentation mask $\hat{P}$ of current input ;

Using B and H to establish the Gaussian mixture model and select a set of low-confidence pixels M through Monte Carlo sampling;
Compute gradient direction $\hat{G}_{\theta_H}$ of H and $\hat{G}_{\theta_B}$ of B ;
for each $m_i \in M$ **do**
    Initialize a empty list L_i;
    while $len(L_i) <= N$ **do**;
        if m_i near the boundary region **then**
            $\hat{G}_\theta = \hat{G}_{\theta_B}$;
        else
            $\hat{G}_\theta = \hat{G}_{\theta_H}$;
        end if
        Search next pixel $\hat{m}_i$ according to $\hat{G}_\theta$ of m_i;
        m_i=$\hat{m}_i$;
        Append the label and entropy value of m_i to L_i;
    end while
    if $D_{KL}(start, end) < Thres$ and $L_{15\%} == A$ and $L_{60\%} == A$ **then**
        $P_{L_i} = A$; $H_{L_i} = mean(H_{L_{15\%}})$
    end if
end for
return P

- It should be noted that the EBR algorithm does not require precise semantic boundaries due to the driving effect of entropy distribution itself and the three safeguards we have carefully set up.
- Even using a somewhat fuzzy semantic outline can help the algorithm refine small and slender objects' results. Although the generated semantic boundaries are not accurate, for complex boundary regions and tiny objects, utilizing boundary clues is often more efficient than direct segmentation.

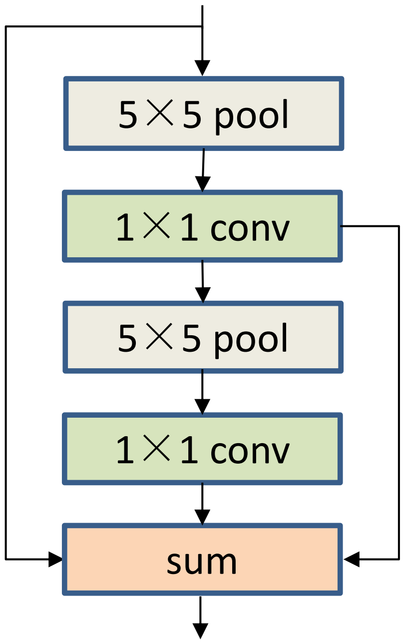
Appendix-Cross Gated Unit

- Building another branch in the encoder backbone to generate semantic boundaries[refer to Gated-SCNN]

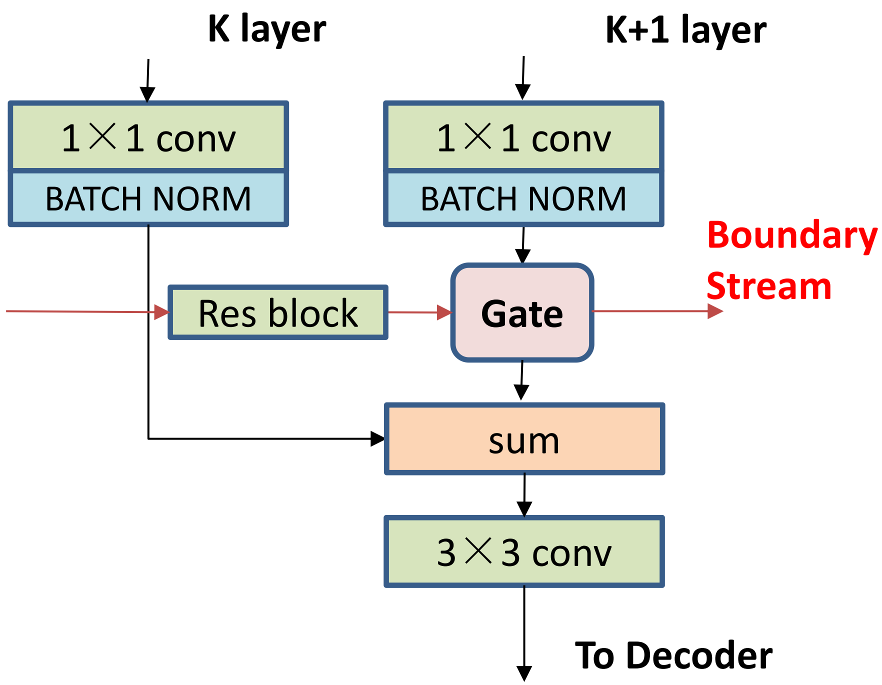


Appendix-Cross Gated Unit

- CPR block



- Cross Gated Unit



- Gated Operation

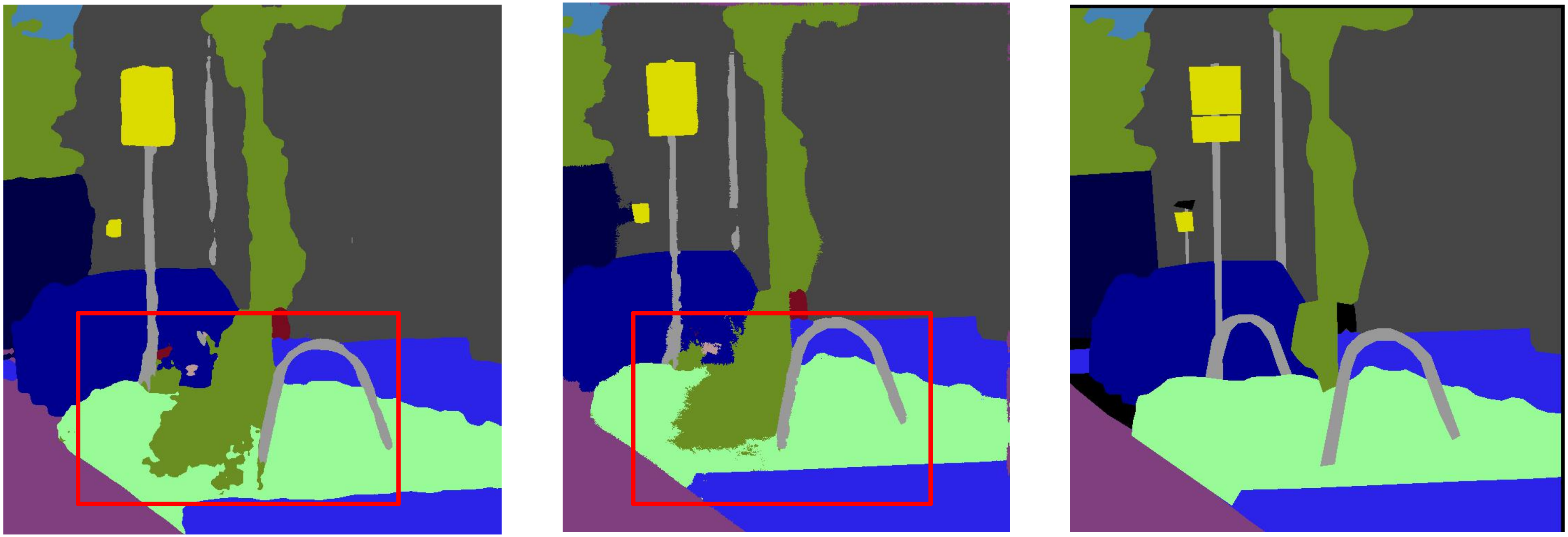
  1. Get attention map
  2. Gated Convolution

$$\alpha = \mathrm{sigmoid}(\mathrm{conv}1 \times 1(s_{k_i} || r_{k_{i+1}}))$$ Attention map

$$\hat{s}_{k_i} = \mathrm{conv}1 \times 1(\alpha_s \odot s_{k_i}) + bias$$ Gated Conv.

$$\hat{r}_{k_{i+1}} = \mathrm{conv}1 \times 1(\alpha_r \odot r_{k_{i+1}}) + bias$$

Appendix-Examples of failure



◆ Main Reason

- There are too many iterations in the algorithm.
- The quality of the predicted semantic boundaries is not good.
- The prior results of network prediction are poor and not credible.