Supplementary: Uncertainty and Energy based Loss Guided Semi-Supervised Semantic Segmentation

Rini Smita Thakur, Vinod K Kurmi
Indian Institute of Science Education and Research, Bhopal, India
{rinithakur, vinodkk}@iiserb.ac.in

In this supplementary, we present the experimental details using algorithm flow. We also provide additional results on PASCAL-VOC dataset and compare with other methods on ResNet101 on 1/8 partition. We provide a segmentation visualization and comparison with baseline methods.

1. Algorithm

In this section, we provide a comprehensive explanation of the algorithms employed to train our segmentation model using uncertainty and energy-based loss. While the main paper discusses the details of the loss function, this entire algorithmic process is given in Algorithm 1.

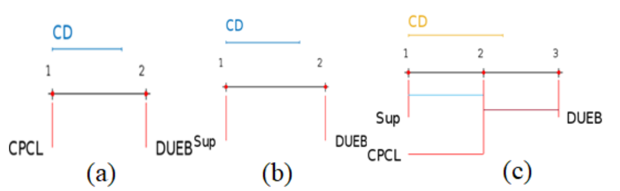


Figure 1. Analysis of statistical significant difference for Supervised, CPCL and DUEB model on CityScapes Dataset at 1/8 partition (a) CPCL and DUEB (b) Sup and DUEB (c) Sup, CPCL and DUEB

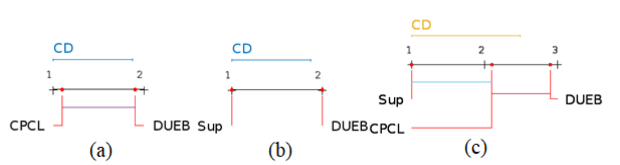


Figure 2. Analysis of statistical significant difference for Supervised, CPCL and DUEB model on PASCAL VOC Dataset at 1/8 partition (a) CPCL and DUEB (b) Sup and DUEB (c) Sup, CPCL and DUEB

2. Visualization

Furthermore, we provide visualizations of the segmentation outputs and compare them with baseline methods. These visualizations allow for a qualitative assessment of the segmentation quality achieved by our proposed method. Figure 3 displays the segmentation results obtained from the Cityscapes dataset, using a partition protocol of 1/8. We compare our results with the state-of-the-art method CPCL. Upon observing the figure, it becomes evident that the CPCL method exhibits false positives for the "Pole" class. Specifically, in the bottom-right image, the right portion mistakenly identifies a pole, whereas the ground truth does not include a pole at that location. In contrast, our proposed method's predictions align more accurately with the ground truth.

3. Statistical Significance Analysis

Figure 1 and 2 show the statistical significance [1] of the proposed method (DUEB) against supervised and CPCL baseline. The Nemenyi test is a post hoc test that is often used following a significant result in an analysis of variance (ANOVA) or a Friedman test, to determine which groups or treatments differ significantly from each other. The critical difference(CD) of the test is dependent on the confidence level (set to 0.05 for this exp) and average ranks for a number of datasets. If the difference between the the rank of two methods lies beyond the CD, then the methods are significantly different. It has been observed that DUEB is significantly different from CPCL and Supervised.

4. Class Disagreement Indicator

The detailed explanation of class disagreement indicator (I) utilized for union pseudolabels is given in this section. When there is disagreement between class prediction of two branches, the conventional approach selects pixel with high confidence as pseudolabel. However, this approach can be erroneous as the prediction with higher confidence can be wrong. Therefore, disagreement indicator selects the difficult class which is likely to create more confusion. Class disagreement indicator obtained by agreement matrix ($M \in \mathbb{R}^{C \times C}$) is used for class selection. The agreement matrix is $C \times C$ matrix with the entries $m_{j,k}$, which

Algorithm 1: Semi-Supervised Semantic Segmentation using Uncertainty and Energy-Based Loss in Pseudo Labels

Input: Training dataset labeled data $D_l = \left\{(X_l^1, G^1), ..., (X_l^N, G^N)\right\}$ and unlabeled data $D_u = \left\{X_u^1, ..., X_u^Q\right\}$

Output: Trained segmentation model $f(;\theta)$

for *mini batch of labeled samples* $X_l^i, G^i \in D_l$ *or* $X^j, X^k \in D_u$ **do**

$X_u^{jk} \leftarrow \text{mix}(X^j, X^k, mask)$;

Supervised loss:

$y_{lc}^i \leftarrow f(X_l^i;\theta_c)$; $\mathcal{L}_{sup} \leftarrow \text{CE}(y_{lc}^i; G_l^i)$

Unsupervised loss:

For input X_i^{jk} obtained the y^i_{inter} , y^i_{union} and w_{u}^i using [2]

$y_{uc}^i, \sigma_{uc}^i \leftarrow f(X_u^{jk};\theta_c)$; and $y_{up}^i, \sigma_{up}^i \leftarrow f(X_u^{jk};\theta_p)$

$\mathcal{L}_{int} \leftarrow w_{\text{u}}^i \text{CE}(y_{uc}^i; y^i_{\text{inter}})$; and $\mathcal{L}_{uni} \leftarrow w_{\text{u}}^i \text{CE}(y_{up}^i; y^i_{\text{union}})$

$\mathcal{L}_{det} \leftarrow \mathcal{L}_{sup} + \gamma_{int}\mathcal{L}_{int} + \gamma_{uni}\mathcal{L}_{uni}$

Uncertainty estimation:

$\mathit{diff}_{\text{p}} = \text{CE}(y_{up}^i; y^i_{\text{union}}) - \text{CE}(y_{up}^i; y^i_{\text{union}})$; $\epsilon_{\text{t}}^p \sim \mathcal{N}(0, \sigma_p)$

$\mathit{diff}_{\text{c}} = \text{CE}(y_{uc}^i; y^i_{\text{inter}}) - \text{CE}(y_{up}^i; y^i_{\text{inter}})$; $\epsilon_{\text{t}}^c \sim \mathcal{N}(0, \sigma_c)$

Using Eq.(12) from main paper, obtained $\mathcal{L}^{\text{c}}_{\text{ale}}$ and $\mathcal{L}^{\text{p}}_{\text{ale}}$

Energy loss:

$\mathcal{L}_e^c \leftarrow LogSumExp_{y^i_{\text{inter}}}(f(X_u^{jk};\theta_c)|y^i_{\text{inter}})$

$\mathcal{L}_e^p \leftarrow LogSumExp_{y^i_{\text{union}}}(f(X_u^{jk};\theta_p)|y^i_{\text{union}})$

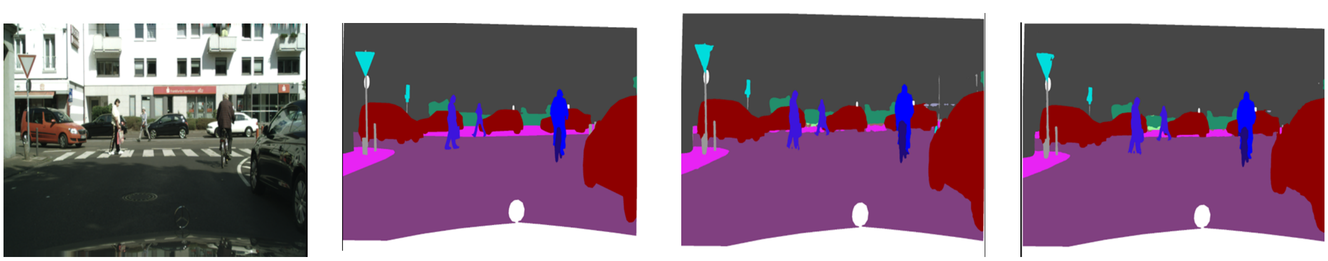
Total loss: $\mathcal{L}_{total} \leftarrow \mathcal{L}_{det} + \gamma_{ale}(\mathcal{L}^c_{ale} + \mathcal{L}^p_{ale}) + \gamma_e(\mathcal{L}_e^c + \mathcal{L}_e^p)$

Update parameters:

$\hat{\theta}_c \leftarrow \theta_c$ -$\lambda \frac{\partial \mathcal{L}_{total}}{\partial \theta_c}$; and $\hat{\theta}_p \leftarrow \theta_p$ -$\lambda \frac{\partial \mathcal{L}_{total}}{\partial \theta_p}$

end

Output trained segmentation parameters $\hat{\theta}_c$



(a) Input (b) Ground-truth (c) CPCL (d) DUEB

Figure 3. Segmentation Results of Cityscapes dataset (partition protocol: 1/8). The missegmented pixels in right part of image (gray) in CPCL are rectified on DUEB.

denotes that number of pixels where Y_{cw} is from class (j) and Y_{pw} is from class (k). The steps of disagreement indicator calculation are:

1. Calculation of agreement matrix ($M \in \mathbb{R}^{C \times C}$), where C is total number of classes. $m_{j,k}$ denotes number of pixels where Y_{cw} is from class (j) and Y_{pw} is from class (k) and $j, k \in [1, C]$.

2. Calculation of class disagreement indicator

$$I_j = 2 - \frac{m_{j,j}}{\sum_{k=1}^{C} m_{j,k}} - \frac{m_{j,j}}{\sum_{k=1}^{C} m_{k,j}} \tag{1}$$

$$I_k = 2 - \frac{m_{k,k}}{\sum_{j=1}^{C} m_{k,j}} - \frac{m_{k,k}}{\sum_{j=1}^{C} m_{j,k}} \tag{2}$$

where $j, k \in [1, C]$ are the indices of the classes.

3. Calculation of pseudolabel for disagreement part l_d^i:

$$l_d^i = c_j \quad \text{if } I_j \geq I_k,\ j \neq k \tag{3}$$

$$l_d^i = c_k \quad \text{if } I_k \geq I_j,\ j \neq k \tag{4}$$

The union pseudolabel is combination of agreement and disagreement parts of two branches. The agreement part implies that both branches give same class prediction. The disagreement part pseudolabels are given with l_d^i.

5. Confidence based dynamic loss

The presence of noise in pseudolabels is inevitable. The confidence based loss can further aid in network training with unreliable pseudolabels. It performs loss re-weighting based on confidence. The maximum softmax probability denotes class wise confidence. Let b_c^i is the prediction confidence of the conservative branch at the i^{th} pixel, similarly, b_p^i is the prediction confidence of the conservative branch at the i^{th} pixel. The confidence based loss is given by:

$$\omega_u^i = \begin{cases} \frac{1}{2}\left(b_c^i + b_p^i\right) & \text{if } Y_{cw}^i = Y_{pw}^i \\ b_c^i & \text{if } Y_{cw}^i \neq Y_{pw}^i,\, l_d^i \leftarrow Y_{cw}^i \\ b_p^i & \text{if } Y_{cw}^i \neq Y_{pw}^i,\, l_d^i \leftarrow Y_{pw}^i \end{cases} \quad (5)$$

Therefore, confidence based loss reduces the impact of unreliable pseudolabel with low confidence.

References

[1] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7:1–30, 2006. 1

[2] Siqi Fan, Fenghua Zhu, Zunlei Feng, Yisheng Lv, Mingli Song, and Fei-Yue Wang. Conservative-progressive collaborative learning for semi-supervised semantic segmentation. *IEEE Transactions on Image Processing*, 2023. 2