# Boosting Semi-Supervised Semantic Segmentation with Probabilistic Representations

Haoyu Xie, Changqi Wang, Mingkai Zheng, Minjing Dong, Shan You, Chong Fu, Chang Xu

School of Computer Science and Engineering, Northeastern University, Shenyang, China
School of Computer Science, Faculty of Engineer, The University of Sydney, Sydney, Australia
SenseTime Research, Beijing, China
Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, China

## Introduction

TASK:

- Pixel-wise contrastive learning in Computer vision.
- Semi-supervised semantic segmentation.

MOTIVATION:

- Improve the quality of pixel-wise representation considering the probability, allow them to perform better under the inaccutate pseudo labels

CONTRIBUTION:

- Define pixel-wise representations from a new perspective of probability theory.
- Through modelling the mapping from pixels to representations as the probability via multivariate Gaussian distributions, tune the contribution of the ambiguous representations to tolerate the risk of inaccurate pseudo-labels.
- Define prototypes in the form of distributions, which indicates the confidence of a class, while the point prototype cannot.
- Propose a Probabilistic Representation Contrastive Learning (PRCL) framework that improves representation quality by taking its probability into consideration.

## Methodology

PROBABILISTIC REPRESENTATION


- We denote the probability of mapping a pixel $\boldsymbol{x}_i$ to a representation $\boldsymbol{z}_i$ as $p(\boldsymbol{z}_i|\boldsymbol{x}_i)$ and define the representation as a random variable following it. For simplicity, we take the form of multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\boldsymbol{I})$ as:

$$p(\boldsymbol{z}_i|\boldsymbol{x}_i) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2\boldsymbol{I}). \tag{1}$$

DISTRIBUTION PROTOTYPE

- The prototype is the posterior distribution after the $n^{th}$ observations $\{\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_n\}$. Under the assumption that all the observations are conditionally independent, the distribution prototype can be derived as:

$$p(\boldsymbol{\rho}|\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_{n+1}) = \alpha \frac{p(\boldsymbol{\rho}|\boldsymbol{z}_{n+1})}{p(\boldsymbol{\rho})} p(\boldsymbol{\rho}|\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_n), \tag{2}$$

where $\alpha$ is a normalization factor. In addition to Equation 1, we can rewrite the prototype as

$$\boldsymbol{\rho} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2\boldsymbol{I}), \tag{3}$$

where

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{n} \frac{\hat{\boldsymbol{\sigma}}^2}{\boldsymbol{\sigma}_i^2}\boldsymbol{\mu}_i \tag{4}$$

$$\frac{1}{\hat{\boldsymbol{\sigma}}^2} = \sum_{i=1}^{n} \frac{1}{\boldsymbol{\sigma}_i^2}. \tag{5}$$

SIMILARITY

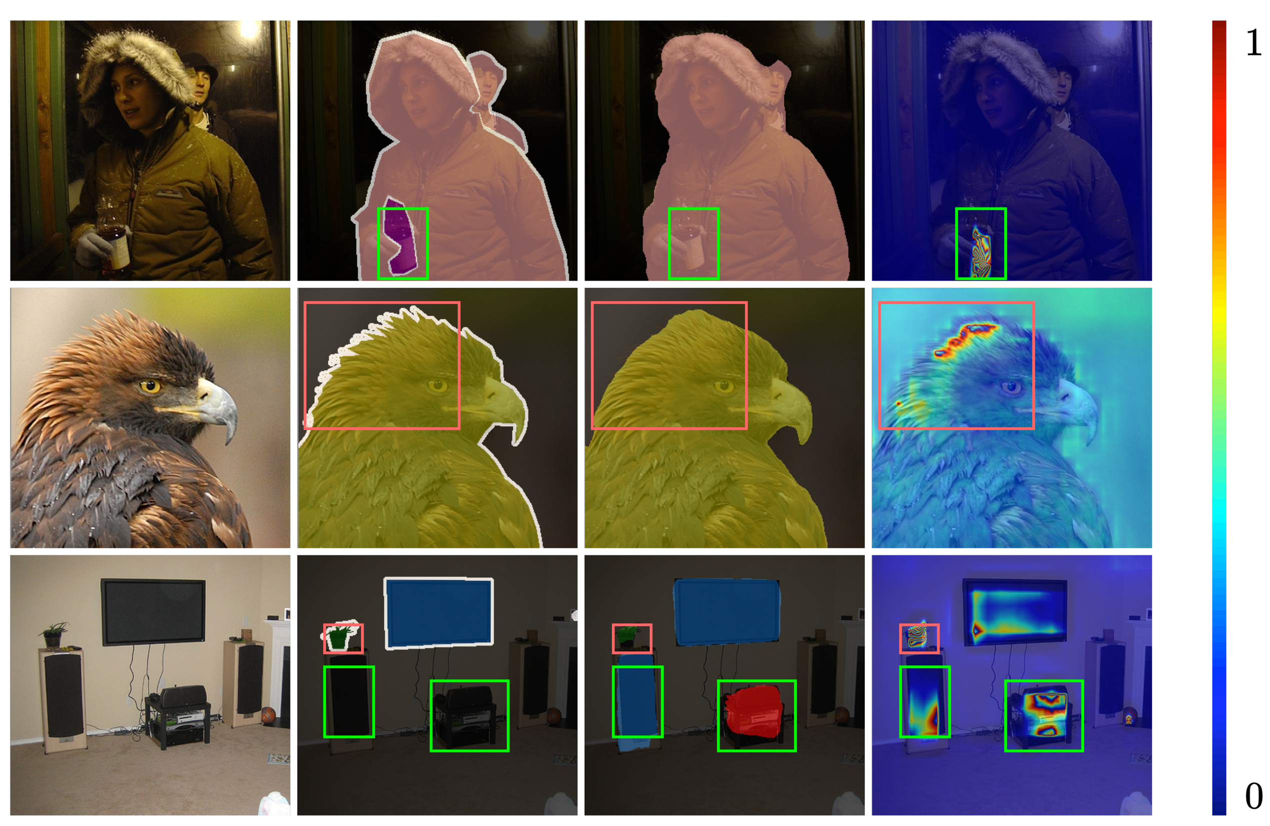
- We leverage Mutual Likelihood Score (MLS) to measure the similarity between two distributions $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$, as follows:

$$\begin{aligned} MLS(\boldsymbol{z}_i, \boldsymbol{z}_j) =& log(p(\boldsymbol{z_i} = \boldsymbol{z}_j)) \\ =& -\frac{1}{2}\sum_{l=1}^{D}\left(\frac{(\mu_i^{(l)} - \mu_j^{(l)})^2}{\sigma_i^{2(l)} + \sigma_j^{2(l)}} + log(\sigma_i^{2(l)} + \sigma_j^{2(l)})\right) \\ & -\frac{D}{2}log2\pi, \end{aligned} \tag{6}$$

- In the first term, the weight of $\ell_2$ distance is small when the $\boldsymbol{\sigma}^2$ is large, which indicates that the similarity between $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ becomes lower due to the low probabilities, even if they are very similar in the view of $\ell_2$ distance. The probability has been taken into consideration besides the simple similarity measure for representations learning.
- In the second term, $\boldsymbol{\sigma}^2$ is penalized for the low probability representations, which makes all the representations more reliable.
- Besides, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\mu}$ can interact with each other. The learnable $\boldsymbol{\sigma}^2$ is associated with $\ell_2$ distance. This means that $\boldsymbol{\sigma}^2$ can be learned via the relations among representations. On the other hand, the $\boldsymbol{\mu}$ can also be optimized via the $\boldsymbol{\sigma}^2$. This is consistent with intuitive cognition.

## Results

In Figure , columns from left to right represent input image, ground-truth, pseudo-label, and probability, respectively. For the fourth column, the red color represents the large $\boldsymbol{\sigma}^2$ (low probability). The green boxes mark the mismatches caused by inaccurate pseudo-labels (e.g., person and bottle) and the red boxes mark the fuzzy pixels (e.g., furry edge of the bird). These two cases are highlighted by $\boldsymbol{\sigma}^2$ and make low contribution in training process.




(a) Image (b) Ground-Truth (c) Pseudo-label (d) Probability