Semantic Segmentation for Real-World Data by Jointly Exploiting Supervised and Transferrable Knowledge

Li-Hsien Lu
s103062588@m103.nthu.edu.tw

Chiou-Ting Hsu
cthsu@cs.nthu.edu.tw

Dept. of Computer Science
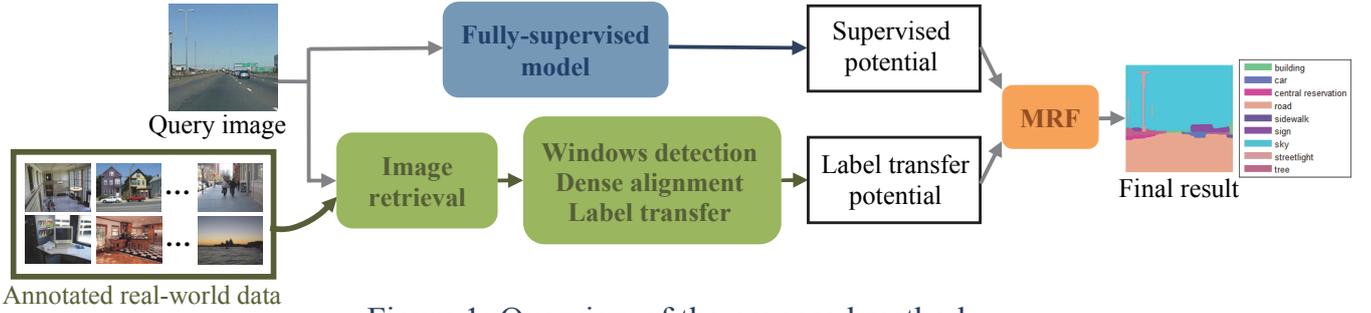National Tsing Hua University
Hsinchu, Taiwan

Figure 1: Overview of the proposed method.

This paper addresses two major challenges in semantic segmentation for real-world data. First, with ever-increasing semantic labels, we need a more pragmatic approach other than existing fully-supervised methods. Second, semantic segmentation for rarely-appeared objects are still challenging for existing methods.

We assume that there exist one supervised model and one annotated real-world dataset. Our goal is to leverage the well-learned knowledge from the existing model and infer new labels via label transfer from the real-world dataset. We propose a "content-adaptive" and "label-aware" MRF framework to jointly exploiting both the supervised and transferrable knowledge. The proposed method needs no off-line training and can easily adapt to real-world data.

Assume that the existing supervised model is trained with the label set C_{fs}, and the real-world dataset is annotated with the label set C_r, where $C_{fs} \neq C_r$. Note that the supervised model is unaware of unknown labels $c \in C_r \backslash C_{fs}$. We formulate the MRF energy function as:

$$E(\mathbf{c}) = -\sum_{\mathbf{p} \in I}[(1 - \alpha(I)) \cdot \psi_{fs}(c, \mathbf{p})$$
$$+ \alpha(I)\beta(c, I) \cdot \psi_{trans}(c, \mathbf{p})]$$
$$+ \lambda \sum_{(\mathbf{p}, \mathbf{q}) \in \varepsilon} \theta(c(\mathbf{p}), c(\mathbf{q})), \quad (1)$$

where $\psi_{fs}(c, \mathbf{p})$ is the supervised potential derived by FCN [1]; $\psi_{trans}(c, \mathbf{p})$ is the label transfer potential obtained by the modified nonparametric method [2]; and $\theta(\cdot, \cdot)$ is the pairwise potential term and λ is a smoothing constant. $\alpha(I)$ is adaptive to different query image I to dynamically combine the two potentials. $\beta(c, I)$ is a label-aware parameter for balancing the priority of rare labels in the query.

In Table 1, we compare our method with existing methods on SIFT Flow and LMSun dataset, and the results demonstrate the effectiveness of the proposed method.

Table 1: Comparison with existing methods.

Method	Per-pixel (%)	Per-class (%)
SIFT Flow dataset $\|C_r\| = 33$		
C. H. Ma et al. [2]	78.3	46.1
M. George [3]	81.7	50.1
Ours ($\|C_{fs} \cap C_r\| = 7$)	81.7	50.0
J. Long et al. [1]	**85.6**	50.1
Ours ($C_{fs} = C_r$)	85.2	**52.0**
LMSun dataset $\|C_r\| = 232$		
M. George [3]	61.2	16.0
J. Yang et al. [4]	60.6	**18.0**
Ours ($\|C_{fs} \cap C_r\| = 55$)	**65.4**	16.5

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.

[2] C. H. Ma, C. T. Hsu and B. Huet. Nonparametric scene parsing with deep convolutional features and dense alignment. In *ICIP*, 2015.

[3] M. George. Image parsing with a wide range of classes and scene-level context. In *CVPR*, 2015.

[4] J. Yang, B. Price, S. Cohen and M. Yang. Context driven scene parsing with attention to rare classes. In *CVPR*, 2014.