Weakly Supervised RBM for Semantic Segmentation

Yong Li*, Jing Liu, Yuhang Wang, Hanqing Lu, Songde Ma

yong.li@nlpr.ia.ac.cn, NLPR, Institute of Automation, Chinese Academy of Sciences

Motivation

Semantic image segmentation is to assign semantic labels to image regions. It provides higher-level understanding about image contents and bridges high-level concept to low-level features. It becomes one of the core problems of computer vision. The motivation of this work is as follows,

- Semantic image segmentation with pixel-level labels is of high cost on the ground truth acquisition
- Semantic segmentation with image-level labels is meaningful, since large numbers of images with image-level labels are available.

Introduction

The weakly supervised semantic segmentation problem becomes very challenging due to the absence of pixel-level labels. However, image-level labels provide the important cue that there will be no mapping from the superpixels to the non-image-level image labels (labels not in S_i). We attempt to do semantic segmentation by such property and the main contributions are summarized as follows.

- We propose a RBM-based learning framework to learn mapping from the superpixels to the image-level labels without supervision of pixel-level labels.
- The non-image-level label suppression and the semantic graph propagation are employed together to make full use of the image-level labels and alleviate the effect of the noisy labels.
- The changeable block size are designed to handle the problem of label imbalance.

Notations

Given a set of images τ, each image i is over-segmented into N_i superpixels. x_{ij} denotes the feature of the j-th superpixel in image i. In addition, the image-level label set for the i-th image is denoted as S_i.

References

[1] X. Liu, B. Cheng, S. Yan, J. Tang, T. Chua and H. Jin Label to Region by Bi-layer Sparsity Priors In *MM 2009*

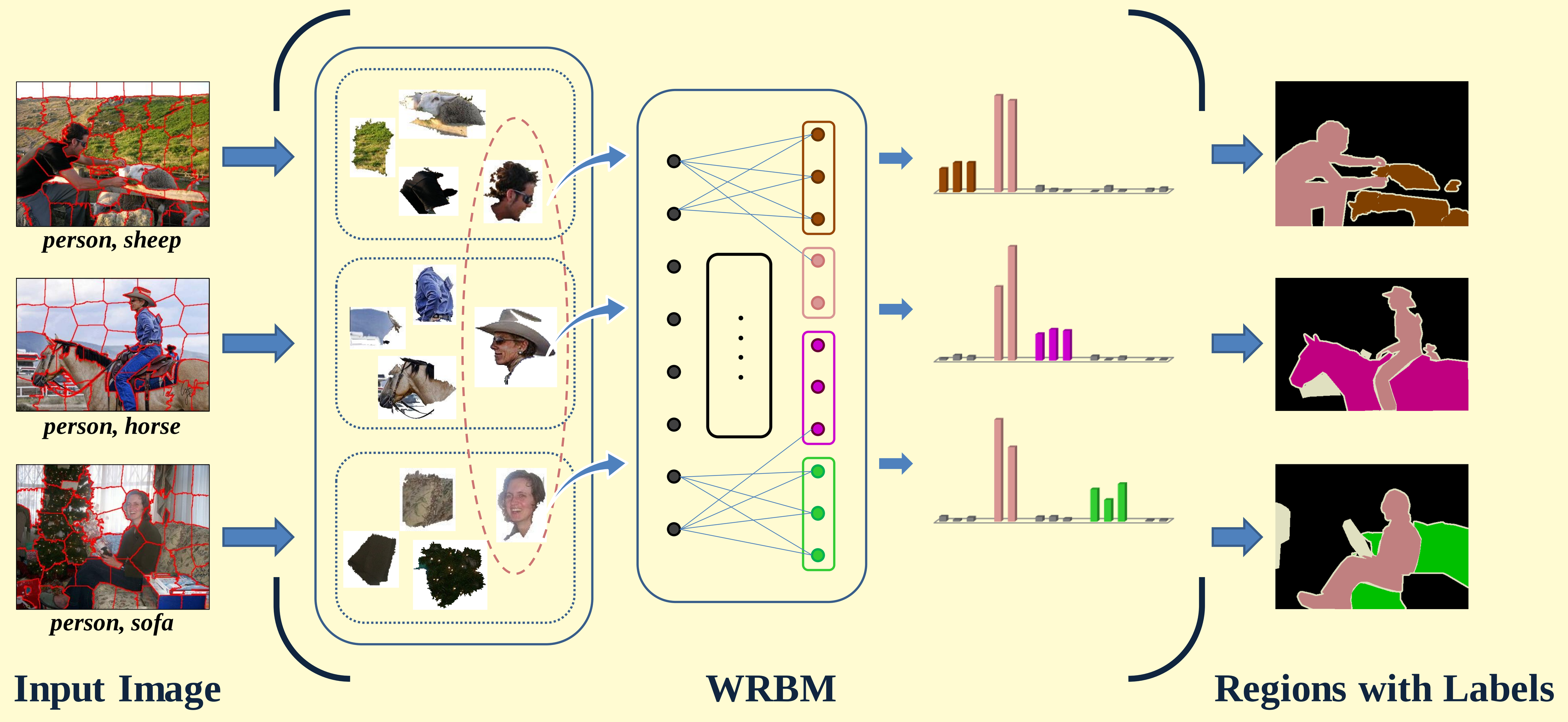
[2] S. Liu, S. Yan, T. Zhang, C. Xu, J. Liu and H. Lu Weakly Supervised Graph Propagation Towards Collective Image Parsing In *TMM 2012*

[3] K. Zhang, W. Zhang, Y. Zheng and X. Xue Sparse Reconstruction for Weakly Supervised Semantic Segmentation In *IJCAI 2013*

[4] K. Zhang, W. Zhang, S. Zeng and X. Xue Semantic Segmentation Using Multiple Graphs with Block-Diagonal Constraints In *AAAI 2014*

[5] W. Xie, Y. Peng and J. Xiao Semantic Graph Construction for Weakly-Supervised Image Parsing In *AAAI 2014*

Overview of The Proposed Approach WRBM



The Proposed Model WRBM

The proposed approach mainly consists of three parts: the energy term of standard RBM, the non-image-level label suppression term and the semantic graph propagation term.
The energy term of standard RBM with a joint configuration of the visible units **v** and hidden units **h** is as follows,

$$E_r(\mathbf{v}, \mathbf{h}) = -\mathbf{h}^T W \mathbf{v} - \mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h}. \tag{1}$$

The non-image-level label suppression term is imported to regularize the response of blocks corresponding to the non-image-level labels to be small. For a given block B_k, the response is defined as follows,

$$E_{B_k} = \sum_{m \in B_k} \mathbf{h}_m^2. \tag{2}$$

Then, the non-image-level label suppression term is formulated as follows,

$$E_s = \sum_{i \in \tau} \sum_{j \in N_i} E_{B_{k \notin S_i}}. \tag{3}$$

The semantic graph propagation term is proposed to make sure that similar superpixels sharing common image-level label have similar hidden response.

$$E_g = \sum_{i \in \tau} \sum_{j \in N_i} \sum_{l \in K(ij)} A_{ij,l} \|\mathbf{h}(x_{ij}) - \mathbf{h}(x_l)\|^2, \tag{4}$$

where $K(ij)$ denotes the index set of the K nearest neighbors.
The final energy function is as follows,

$$E = E_r + \alpha E_s + \beta E_g. \tag{5}$$

Results on The PASCAL Dataset

Method	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motorbike	person	plant	sheep	sofa	train	tv	bkgd	mean
[Liu et al., 2009]	24	25	40	25	32	35	27	45	16	49	24	32	13	25	56	28	17	16	33	18	**82**	32
[Liu et al., 2012]	28	20	52	28	46	41	39	60	25	68	25	35	17	35	56	36	46	17	31	20	65	38
[Zhang et al., 2013]	48	20	26	25	3	7	23	13	38	19	15	39	17	18	25	47	9	**41**	17	33	-	24
[Zhang et al., 2014]	65	25	39	8	17	38	17	26	25	17	**47**	**41**	**44**	32	59	34	36	23	35	31	-	33
[Xie et al., 2014]	**85**	**55**	**87**	45	42	31	34	57	21	81	23	16	6	11	42	31	**72**	24	**49**	40	41	42
Ours (equal block)	56	16	77	25	**62**	22	63	**66**	**42**	**83**	15	37	13	5	**81**	60	50	23	29	**72**	41	45
Ours (adaptive block)	33	50	72	**66**	46	**70**	**73**	43	30	78	29	31	16	**52**	33	**61**	41	38	47	48	50	**48**

- Our approach with adaptive block size outperforms state of the art [Xie et al., 2014] by 6%.
- Compared with the setting of equal block size, our approach with adaptive block size is more robust to class changes by taking data distribution into consideration.

Conclusions

- We propose a weakly supervised semantic segmentation method via non-image-level label suppression and semantic graph propagation.
- The changeable block size is able to handle the problem of label imbalance. While, multiple blocks are preferred to deal with diverse backgrounds by adding a random background label.