Weakly Supervised RBM for Semantic Segmentation

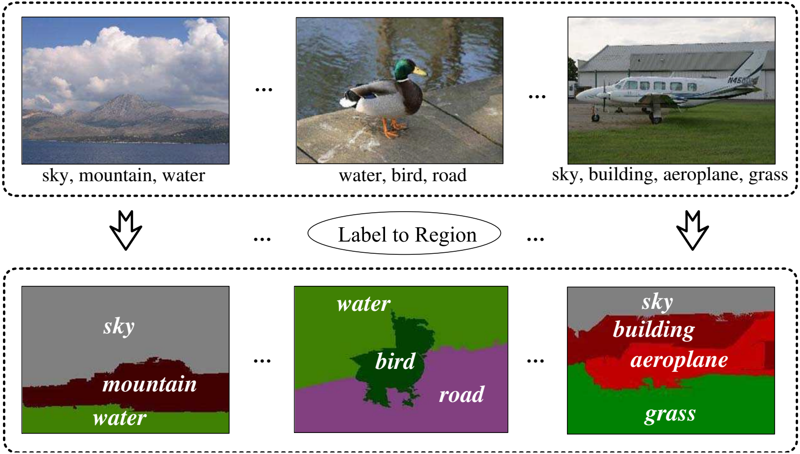
Yong Li, Jing Liu, Yuhang Wang, Hanqing lu, Songde Ma

Institute of Automation, Chinese Academy of Sciences

July 30, 2015

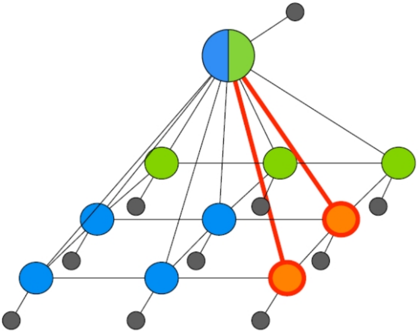
Outline

Introduction to semantic segmentation



[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

Semantic segmentation has achieved significant progress in the past years under fully supervised setting with pixel-level labels [1-2].



[1]:L. Ladicky, C. Russell, P. Kohli, and P.H.S. Torr. Associative hierarchical crfs for object class image segmentation.ICCV2009.

[2]:X. Boix J. M. Gonfaus F. S. Khan J. van de Weijer A. Bagdanov M. Pedersoli J. Gonzlez J. Serrat, Combining local and global Bag-of-Words representations for semantic segmentation. ICCV Workshop 2009.

Motivation

We focus on the weakly supervised semantic segmentation problem with only image-level labels.

- Semantic image segmentation with pixel-level labels is of high cost to acquire
- Large numbers of images with image-level labels are available (e.g., Flicker).
- Semantic segmentation with the weakly supervised setting is meaningful and scalable

Motivation

We focus on the weakly supervised semantic segmentation problem with only image-level labels.

- Semantic image segmentation with pixel-level labels is of high cost to acquire
- Large numbers of images with image-level labels are available (e.g., Flicker).
- Semantic segmentation with the weakly supervised setting is meaningful and scalable

Motivation

We focus on the weakly supervised semantic segmentation problem with only image-level labels.

- Semantic image segmentation with pixel-level labels is of high cost to acquire
- Large numbers of images with image-level labels are available (e.g., Flicker).
- Semantic segmentation with the weakly supervised setting is meaningful and scalable

The Proposed Approach WRBM

A novel way to leverage image-level labels.

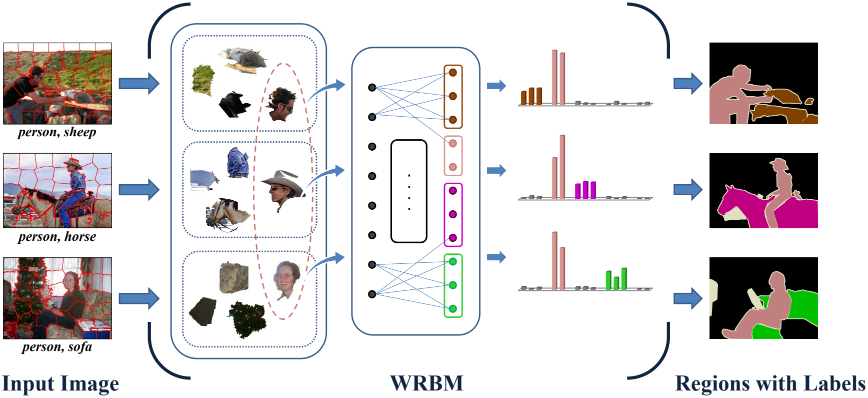
- Image-level labels provide the important cue that there will be no mapping from the superpixels to the non-image-level labels.
- We make semantic segmentation by catching such property with weakly supervised RBM.
- The hidden nodes of RBM are divided into several blocks, where each block corresponds to a specific label.

The Proposed Approach WRBM

The proposed approach mainly consists of three parts,

- The standard RBM term is to learn the hidden representation of input features.
- The non-image-level suppression term is imported to regularize the response of blocks corresponding to the non-image-level labels to be small.
- The semantic graph propagation term is proposed to make sure that similar superpixels sharing common image-level label have similar hidden response.

Overview of The Proposed Approach

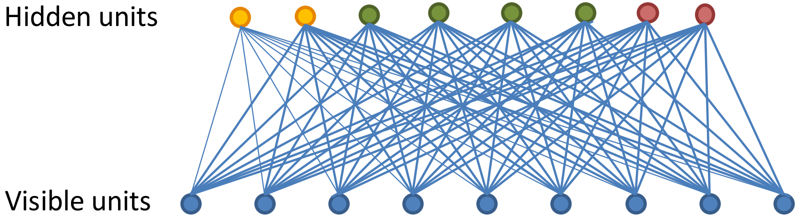


Details of The Proposed Approach

Standard RBM is to learn hidden representation of the input features in an unsupervised setting. The energy function is as follows,

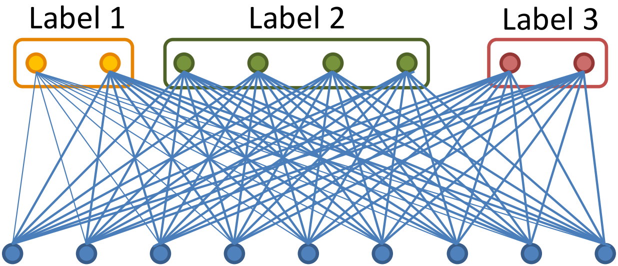
$$E_r(\mathbf{v}, \mathbf{h}) = -\mathbf{h}^T W\mathbf{v} - \mathbf{b}^T\mathbf{v} - \mathbf{c}^T\mathbf{h} \tag{1}$$

where $\mathbf{v}$ are the visible units while $\mathbf{h}$ are hidden units.



Details of The Proposed Approach

The hidden nodes of WRBM are divided into several blocks, where each block corresponds to a specific label.



Two ways to divide the hidden nodes

- The block size is of the same.
- The block size is adaptively adjusted based on the data distribution.

Non-image-level Label Suppression (NLS)

It is imported to leverage image-level labels and regularize the response corresponding to the non-image-level labels (labels not in S_i) to be small. The response function for each block B_k is defined as follows,

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

where m is the index for the hidden unit in block B_k.

The NLS term can be formulated as follows,

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

As a result, mapping to the image-level labels will be encouraged for each superpixel.

Non-image-level Label Suppression (NLS)

It is imported to leverage image-level labels and regularize the response corresponding to the non-image-level labels (labels not in S_i) to be small. The response function for each block B_k is defined as follows,

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

where m is the index for the hidden unit in block B_k.
The NLS term can be formulated as follows,

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

As a result, mapping to the image-level labels will be encouraged for each superpixel.

Semantic Graph Propagation

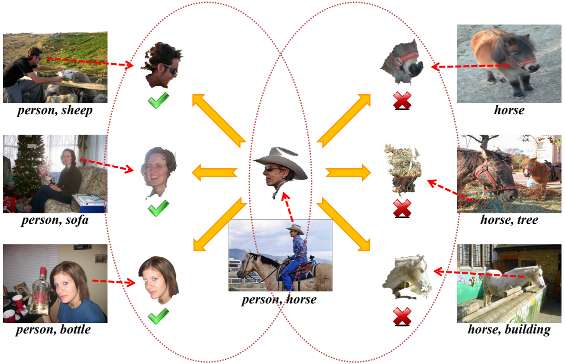
- It is proposed to make sure that similar superpixels have similar hidden response.
- If two similar local regions from different images share common label, then it is natural to tag these regions with the common label.
- In order to deal with high correlated concepts, like “grass” and “sheep”. More discriminative information can be embedded.

Semantic Graph Propagation

We seek to find the K nearest neighbors for each label L_i separately and optimize for the best semantic nearest neighbors.

$$\max_{L_i} \sum_{l \in K(ij)} A_{ij,l}^{L_i} \tag{4}$$

where $A_{ij,l}^{L_i}$ is the similarity measure for the superpixel pair with label L_i.



Target Function of The Proposed Approach WRBM

The semantic graph propagation term,

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

The standard RBM term,

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

The non-image-level label suppression term,

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

We get the final target function by combining the above three terms,

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

where α and β are the tradeoff parameters.

Target Function of The Proposed Approach WRBM

The semantic graph propagation term,

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

The standard RBM term,

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

The non-image-level label suppression term,

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

We get the final target function by combining the above three terms,

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

where α and β are the tradeoff parameters.

Target Function of The Proposed Approach WRBM

The semantic graph propagation term,

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

The standard RBM term,

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

The non-image-level label suppression term,

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

We get the final target function by combining the above three terms,

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

where α and β are the tradeoff parameters.

Datasets

- The PASCAL VOC 2007 dataset with 632 images of 20 categories.
- The LabelMe LMO dataset with 2688 images of 33 categories.

Comparison Methods

- X. Liu, B. Cheng, S. Yan, J. Tang, T. Chua and H. Jin, Label to Region by Bi-layer Sparsity Priors In MM 2009
- S. Liu, S. Yan, T. Zhang, C. Xu, J. Liu and H. Lu Weakly Supervised Graph Propagation Towards Collective Image Parsing In TMM 2012
- K. Zhang, W. Zhang, Y. Zheng and X. Xue Sparse Reconstruction for Weakly Supervised Semantic Segmentation In IJCAI 2013
- K. Zhang,W. Zhang, S. Zeng and X. Xue Semantic Segmentation Using Multiple Graphs with Block-Diagonal Constraints In AAAI 2014
- W. Xie, Y. Peng and J. Xiao Semantic Graph Construction for Weakly-Supervised Image Parsing In AAAI 2014

Experiments and Results

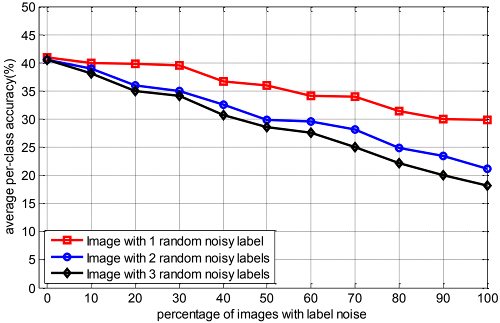
Table 1: Semantic segmentation results on PASCAL dataset.

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.*, 2009b]	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 and achieves better performance.

Robustness to Label Noise

To validate the robustness of our model to label noise, we conduct experiments with different numbers of noise labels and noise images.



- As the number of noise images increases, the average class accuracy decays linearly.
- The proposed approach achieves satisfactory performance even when every image is with a noise label.

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.
- The proposed approach is robust to label noise.
- It can be applied to more multi-instance multi-label learning problems.

Thank You!

More info: http://www.foreverlee.net