Image Segmentation Based on Semantic Knowledge and Hierarchical Conditional Random Fields

Cao Qin[1], Yunzhou Zhang[1,2](✉), Meiyu Hu[1], Hao Chu[2], and Lei Wang[2]

[1] College of Information Science and Engineering, Northeastern University,
Shenyang, China
zhangyunzhou@mail.neu.edu.cn

[2] Faculty of Robot Science and Engineering, Northeastern University,
Shenyang, China

Abstract. Semantic segmentation is a fundamental and challenging task for semantic mapping. Most of the existing approaches focus on taking advantage of deep learning and conditional random fields (CRFs) based techniques to acquire pixel-level labeling. One major issue among these methods is the limited capacity of deep learning techniques on utilizing the obvious relationships among different objects which are specified as semantic knowledge. For CRFs, their basic low-order forms cannot bring substantial enhancement for labeling performance. To this end, we propose a novel approach that employs semantic knowledge to intensify the image segmentation capability. The semantic constraints are established by constructing an ontology-based knowledge network. In particular, hierarchical conditional random fields fused with semantic knowledge are used to infer and optimize the final segmentation. Experimental comparison with the state-of-the-art semantic segmentation methods has been carried out. Results reveal that our method improves the performance in terms of pixel and object-level.

Keywords: Image segmentation · Semantic knowledge · Ontology
Conditional random fields

1 Introduction

Mobile robots intended to perform in human environments need to access a world model that includes the representation of the surroundings. Since most people concentrate on the accurate geometry of the world, the semantic information arises and becomes a vital factor that assists the robot in executing tasks. Semantic segmentation can just provide this kind of information. Its purpose is

Research supported by National Natural Science Foundation of China (No. 61471110, 61733003), National Key R&D Program of China (No. 2017YFC0805000/5005), Fundamental Research Funds for the Central Universities (N172608005, N160413002).

J.-H. Lai et al. (Eds.): PRCV 2018, LNCS 11256, pp. 213–225, 2018.
https://doi.org/10.1007/978-3-030-03398-9_19

to divide the image into several groups of pixels with a certain meaning and to assign the corresponding label to each region. However, image semantic segmentation has become an intractable task due to the varieties of different objects, unconstrained layouts of indoor environments.

The seemingly complicated living environments for people possess a variety of repeated specific structures and spatial relations between different objects. For instance, a monitor is more likely found in a living room than in a kitchen. Also, a cup is more likely on the table than on the floor. Such kinds of specific objects and spatial relations can be defined as an alternative semantic knowledge which improves the quality of image segmentation and helps robots to recognize the interesting things.

Traditional image segmentation methods [5] take advantage of the low-level semantic information, including the color, texture, and shape of the image, to achieve the purpose of segmentation. But the result is not ideal enough in the case of complex scenes. In recent years, researchers have been committed to using convolution neural networks to enhance the segmentation of images. However, the method of deep learning to deal with the pixel tags only draws the outline of the objects coarsely. There also exists the problem that only local independent information is accessible and the deficiency of surrounding context constraints. [6] constructed the Conditional Random Fields model (CRF) [13] according to the pixel results produced by the neural network. This approach is designed to enhance the smoothness of the label, maintain the mask consistency of the adjacent pixels. Although the above-mentioned methods achieve remarkable pixel-level semantic segmentation, they only make use of the constrained relations among low-level features.

In this paper, we propose a semantic knowledge based hierarchical CRF approach to image semantic segmentation. Our method not only achieves better segmentation effect at pixel-level but also gets great improvements on the object-level. Figure 1 shows the overall framework of our method and the main contributions are summarized as follows:

- We construct an ontology-based knowledge network which is utilized to express the semantic constraints.
- We first propose an original hierarchical CRF model fused with semantic knowledge from the ontology.
- We make great progress in error classification at object-level by embedding the global observation of the image and using the high-level semantic concept correlation.

2 Related Works

2.1 Image Segmentation Based on CNNs and CRFs

Semantic image segmentation has been always a popular topic in the field of computer vision. In recent years, the methods of deep convolution neural network have made an unprecedented breakthrough in this field. [8] proposed an

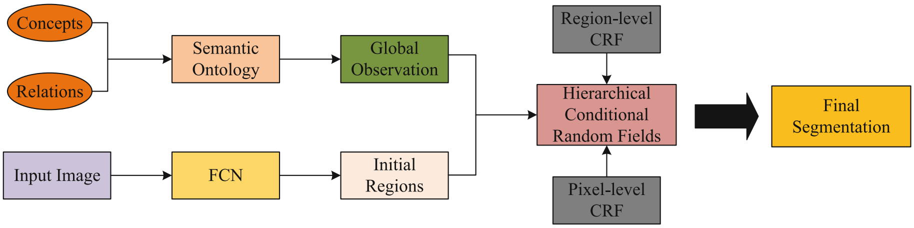

Fig. 1. Overall framework of our method. Concepts and relations are gathered from human's elicitation according to the image database. Global observation is derived from the semantic ontology network composed of the concepts and relations. FCN [15] accepts inputting image in any size and generates initial segmentation region which is utilized in both pixel-level CRF and region-level CRF. A hierarchical CRF model is constructed to combine two kinds of CRF models and produces the final segmentation.

R-CNN (regions with CNN features) method which combined region proposals with CNNs. It deals with the problem of object detection and semantic segmentation but needs a lot of storage and has limitation on efficiency. Prominent work FCN [15] designed a novel end-to-end fully convolutional network which accepted inputting image for any size and achieved pixel classification. Based on FCN, Vijay et al. [3] replicated the maximum pooling index and constructed an original and practical deep fully CNN architecture called SegNet. Although these methods have made good progress through CNNs, they lack the spatial consistency because of the neglect of the relationship between pixels.

On the basis of [15], Zheng et al. [22] modeled the conditional random fields as a recurrent neural network. This network utilized the back propagation algorithm for end-to-end training directly without the offline training on CNN and CRF models respectively. Lin et al. [14] introduced the contextual information into the semantic segmentation, and improved the rough prediction by capturing the semantic relations of the adjacent image. In contrast to the above methods, our method pays more attention to improve the segmentation of the region and object layer, which also help to promote the segmentation accuracy at the pixel level in a subtle way.

2.2 Semantic Knowledge

Semantics, as the carrier of knowledge information, transform the whole image content into intuitive and understandable semantic expression. Ontology has become a standard expressive form of relations between semantic concepts.

Wang et al. [20] constructed ontology network using the OWL DL language. Ontology network captures the hidden relationships between features in the feature diagrams precisely and helps to solve the task of feature modeling. An ontology-based approach to object recognition was presented in [7]. It endowed the object semantic meaning through the relations between the objects and the concepts in the ontology. Ruiz et al. [17] utilized the expert knowledge established

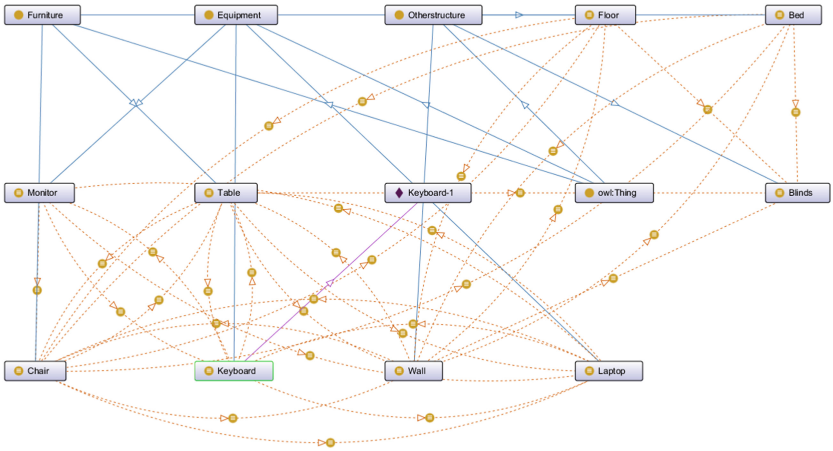

Fig. 2. A part of established ontology on the images of the NUY v2 dataset. The root concept is *Thing*. The blue, purple and brown lines represent the relation *has_subclass*, *has_individual* and *hasAppearedwith*, respectively. (Color figure online)

manually to extract semantic knowledge and trained probabilistic graph model. Subsequently, they proposed a hybrid system based on probabilistic graph model and semantic knowledge in [18]. The system makes full use of the context of the object in the image and shows excellent recognition effect even in complex or uncertain scenes. However, this method requires the laboriously manual design of the training data of the PGM model and only gets performance in the aspect of object recognition.

A related but very different work to our method is introduced in [21]. This work facilitated the semantic information to transform the low-level features of the image into the high-level feature space and assign the corresponding class labels to each object parts. In our work, we obtain the prediction directly from the FCN and utilize the combination of hierarchical CRFs and the ontology network to optimize the regional label. It has great advantages in efficiency because it does not need to train multiple CRF models.

2.3 Hierarchical CRFs

Primary CRF model only uses the local features of the images, such as pixel features and cannot utilize the high-level features, such as regional features and global features. [19] adopted the original potential energy function of CRF to define the constraint relation between the local feature and the high-level features and constructed the hierarchical CRF model. Huang et al. [10] established a hierarchical two-stage CRF model on the basis of the idea of parametric and nonparametric image labeling. Benjamin et al. [16] paid attention to both the pixel and object-level performance by merging region-based CRF model with dense pixel random fields in a hierarchical way. Compared with [16], our approach adds the global observation information from the ontology network into the hierarchical CRF which makes the system more robust in the global segmentation performance.

3 Approach

3.1 Semantic Knowledge Acquirement

3.1.1 Ontology Definition

Different semantic labels will appear in the same image. An image is usually labeled with a variety of semantic labels. The ontology is a clear and formal specification of shared concepts that is applied to define concepts and the relationships between concepts and concepts. In this work, we utilize the ontology as the carrier of semantic knowledge to form a reasoning engine for object labeling. Ontology is generated by human elicitation. For example, an indoor scene can be modeled by defining the types of objects that occur in the environment. E.g. Desk, Table, Bookshelf, etc.... In addition, the properties of the object and the contextual relations that exist between the objects should be formulated. As Fig. 2 illustrates, a multi-layer ontology-based structure is proposed to give the most understandable semantic representation of the image content. This graph is generated by using the software Protégé[11] based on the OWL DL language. The root concept is *Thing*, and its subordinate concept such as $furniture$, *equipment*, and *otherstructures* are easy to be found in a typical indoor environment. The ultimate goal of using ontology is to ensure that the labels of objects appearing in the image are consistent.

3.1.2 Semantic Constraints

The situation that objects contained in a specific scene owns certain probability of occurrence from the overall consideration. Therefore, each class that appears in the ontology should have a propriety which is defined as *has_Frequency* from the perspective of fuzzy description logics [2]. More importantly, what we should consider is how to generate the probability that two objects appear in one scene at the same time. We define the co-occurrence of the two objects by rule *hasAppearedwith* in the ontology.

As mentioned above, the context relations between objects are obtained by fuzzy description logics. The occurrence probability of a concept and the previous definition *has_Frequency* of each class are defined by the following formula:

$$has\_Frequency(C_i) = prob(C_i) = \frac{n_i}{N} \tag{1}$$

Where n_i refers to numbers of concept C_i appears in the image. N represents the number of images used in the dataset. Similarly, the probability of two objects appear in an image at the same time is formulated:

$$prob(C_i, C_j) = \frac{n_{i,j}}{N} \tag{2}$$

$n_{i,j}$ refers to the number of images in which concept C_i and C_j appear simultaneously in an image. On the basis of equation (2), we compute the Normalized Pointwise Mutual Information (NPMI) according to [4]:

$$p(C_i, C_j) = log\frac{prob(C_i, C_j)}{prob(C_i) * prob(C_j)} \tag{3}$$

If C_i and C_j are independent concepts mutually, it is easy to deduce that $prob(C_i, C_j) = 0$. In a word, $prob(C_i, C_j)$ measures the the degree of sharing information between concept C_i and C_j.

To normalize $prob(C_i, C_j)$ to the interval $[0, 1]$, we obtain the fuzzy representation of $hasAppearedwith$:

$$hasAppearedwith(C_i, C_j) = \frac{p(C_i, C_j)}{-log[max(prob(C_i), prob(C_j))]} \tag{4}$$

3.2 Hierarchical Conditional Random Fields

3.2.1 Pixel-Level CRFs

CRFs applied in semantic segmentation is a probabilistic model for the segmentation of class labels associated with given observation data. In CRF model, observation variable $Y = \{y_1, y_2, ..., y_N\}$ indicates the image pixel and the implicit random variable $X = \{x_1, x_2, ..., x_N\}$ refers to the labels of pixels. Given a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, $\boldsymbol{V} = \{1, 2, ..., N\}$. $e_{ij} \in \boldsymbol{E}$ means the collection of edges of adjacent variables x_i and x_j. Random variable x is defined over the set $L = \{l_1, l_2, ...l_K\}$. Under the premise of the given condition Y, the joint probability y distribution of the random variable X follows the Gibbs distribution:

$$P(X|y) = \frac{1}{Z}exp(-E(X|y)) \tag{5}$$

Energy function is defined by:

$$E(X|y) = \sum_{i \in \boldsymbol{V}} E_i(x_i) + \alpha \sum_{\{i,j\} \in \boldsymbol{E}} E_{ij}(x_i, x_j) \tag{6}$$

Where α is the weight coefficient, Z is the normalization factor. E_i is the unary potential, which includes the relationship between random variables and the observed values. Unary potential is usually deduced by some other classifiers that generate distributions over class labels. The unary potential used in this paper is produced by the FCN [15]. E_{ij} denotes the pairwise potentials, which represents the smoothness constraints on adjacent pixels for the same label and include the relationships between adjacent random variable nodes. According to [13], we model the pairwise potentials as follows:

$$E_{ij}(x_i, x_j) = u(x_i, x_j) \sum_{a=1}^{M} \omega^{(a)} k^{(a)}(f_i, f_j) \tag{7}$$

Where $k^{(a)}$ is a Gaussian kernel, $\omega^{(a)}$ is a weight parameter for kernel $k^{(a)}$ and f_i is a feature vector for pixel i. Function $u(.,.)$ is called the label compatibility function, which captures the compatibility between connected pairs of nodes that are assigned different labels. Since the above mentioned two kinds of energy items contain fewer hidden variables, they are also called low-order energy terms.

The main task of semantic segmentation is to select l_i from the set L and assign it to each random variable x_i. Thus, an energy expression is constructed to solve X which meets the maximum of a posteriori probability:

$$X^* = \underset{X}{\arg\max}\ P(X|y) = \underset{X}{\arg\min}\ E(X|y) \tag{8}$$

3.2.2 HCRF

As shown in Fig. 3, HCRF model consists of two layers: the pixel layer and the region layer. The pixel layer is composed of hidden random variable X, whose definition is consistent with the CRF model. The region layer is formed by the segmentation blocks obtained from FCN. $r = \{x_1, x_2, ...x_m\}$ represents a region block unit that is a set of the hidden random variables x. $R = \{r_1, r_2, ...r_p\}$ denotes a collection of all area blocks. According to the model described above, the energy expression for HCRF model is defined as follows:

$$\begin{aligned} E(X|y) = &\sum_{i \in \boldsymbol{V}} E_i(x_i) + \alpha \sum_{\{i,j\} \in \boldsymbol{E}} E_{ij}(x_i, x_j) \\ &+ \beta \sum_{m \in \boldsymbol{R}} E_m(r_m) + \gamma \sum_{\{m,n\} \in \boldsymbol{E'}} E_{mn}(r_m, r_n) \end{aligned} \tag{9}$$

The pixel layer corresponds to the CRF model uses pixels as the basic processing unit, including the low-order energy terms described above. The energy term reflects the constraints of the local texture feature for the pixel class and smoothness constraint between pixels. E_m depicts the unary potential defined in the region layer, which is the key to associating the pixel layer and the segmentation layer. It also reflects the constraints of the descriptive feature to the categories of segmentation region. β and γ are the weights of the corresponding energy function of the region.

The unary potential is divided into two parts in the regional energy function model. The one is the local observation part, which relates to the observation of the image region. The other one is the global observation part, which denotes the observation of relevant semantic label on the entire image dataset. In order to combine the pixel layer and the region layer, the region unary potential is formulated:

$$E_m(r_m) = -ln(f_i^r(x_i)) * occur(x_i) \tag{10}$$

Where $f_i^r(.)$ is the normalized region probability distribution of the region i as the local observation. It is computed from the implicit FCN pixel distribution. $occur(x_i) = prob(x_i)$ is the probability that the label of region r_m occurs in the whole image dataset as the global observation, which is calculated by the

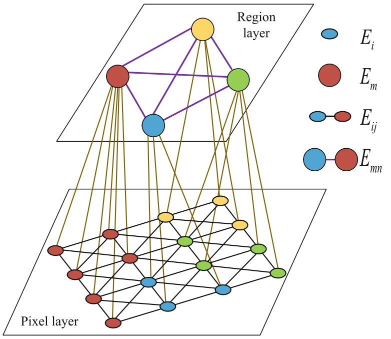

Fig. 3. Illustration of hierarchical conditional random fields. The smaller ellipses correspond to the unary potentials of the pixel, and the larger circles represent the unary potential defined in the region layer. Different colors mean different object labels.

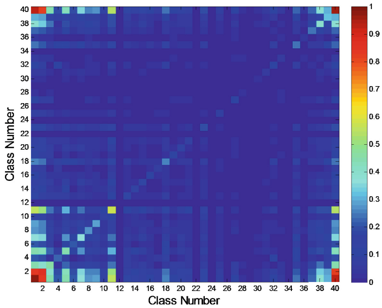

Fig. 4. Visualization of the occurrence probabilities of different classes. Off-diagonal entries are the probabilities of simultaneous occurrence of two concepts, while diagonal entries are the occurrence probabilities of the individual concepts. The class numbers correspond to the 40 different classes in the image dataset. (Color figure online)

$has\_Frequency$ in the last section. The global observation of the image is introduced to the unary potential function so that the unary potential is enhanced by the knowledge in a higher level. This is an effective complement to the limitations and deficiencies of the local observations and promotes the modeling ability of the unary potential function.

To take advantage of the context information, we utilize the pairwise potentials between the regions. The pairwise energy term is defined:

$$E_{mn}(r_m, r_n) = \begin{cases} 0 \text{ if } hasAppearedwith(x_m, x_n) \geq \tau \\ T \qquad\qquad \text{otherwise} \end{cases} \tag{11}$$

Where $hasAppearedwith(x_m, x_n)$ implies the probability that the labels of region r_m and r_n appear simultaneously in a picture. τ is a given threshold. T means the given penalty. Pairwise energy term of region E_{mn} is quite different from the pairwise energy term of pixel E_{ij}. E_{ij} encourages adjacent pixels to obtain the same class label. E_{mn} makes the label of the adjacent region in the semantic layer constrained and gives the mark of the irrelevant object in the adjacent area great punishment. Owing to the setting of the above parameters, our method has achieved excellent results in the experiment of misclassification at the object-level, as discussed in Sect. 4.2. As for calculating the weight parameters in the HCRF, we use the method of layer by layer weight parameter learning proposed by AHCRF [19].

The final semantic segmentation results are obtained by minimizing the energy function $E(X|y)$ as described in the formula (8). Because we introduce the potential energy function based on global observation, the graph cut based method proposed by Kahlil et al. [12] is used to complete the model inference.

4 Experiments and Analysis

4.1 Experimental Setup

4.1.1 Dataset

The semantic segmentation method we propose is evaluated by the dataset NYU v2. It contains 1449 images collected from 28 different indoor scenes. The whole dataset is divided into 795 training images and 654 test images. We exploit the 40-classes version provided by Gupta et al. [9]. As shown in Fig. 5, we can see the various objects marked with different colors in the image.

4.1.2 Implementation Details

In our approach, the highly expressive OWL DL language is employed to design and form the ontology of the dataset. In order to build the ontology model and obtain the data we need, we use the Protégé as our ontology editor. The semantic rules are applied on the dataset to construct the ontology. Figure 2 represents the generated ontology for the semantic classes of the NYU v2 dataset. It can be clearly seen that the degree of correlation between the two concepts which is also defined as the fuzzy rule *hasAppearedwith*. It cannot be ignored that $has\_Frequency$ has become the underlying properties of each concept. Figure 4 visualizes the occurrence probabilities of the concepts as a matrix representation. Element (i, j) of this matrix relates to $prob(C_i, C_j)$ and element (i, i) corresponds to $prob(C_i)$. There are obvious red areas in the lower left corner and the upper right corner of the picture, which indicates that these classes are more likely to appear. In more detail, the class 1 and 2 represent *wall* and $floor$ respectively and the class 40 means *otherprop*. These classes are extremely common and appear in almost every image of the dataset.

The semantic segmentation maps are generated by the up-to-date FCN network. In addition, the final result gets improvement by the optimization of backend hierarchical conditional random fields. Thus, our method will be compared to the effect of FCN only and the FCN with dense CRF [13]. We utilize the TensorFlow [1] to construct the deep CNN in Linux operation system. Our approach runs at 14 Hz on the TITAN-X GPU. Image segmentation is the most computationally intense task, taking 170 ms to segment an image of $480 * 640$ pixels.

4.1.3 Evaluation Metrics

The pixel accuracy (PA) is the ratio of correctly labeled pixels in an image to all pixels. It is specified by $\frac{\sum_i N_{ii}}{\sum_{i,j} N_{ij}}$, where N_{ij} represents the number of pixels of label i being labeled as j. Mean accuracy is defined as $\frac{1}{k} \frac{\sum_i N_{ii}}{\sum_j N_{ij} + \sum_j N_{ji} - \sum_i N_{ii}}$.

However, the mere use of the above three criteria at the pixel level is not sufficient to reflect the advantages of the method presented in this paper. Similar to [16], we calculate the number of object False Positives which represents the number of prediction regions that do not have any overlap with a ground truth instance of the same class. It is designed to evaluate the error-classification degree in order to reflect the excellent performance at the object-level.

4.2 Results and Analysis

For the sake of evaluating our method with existing approaches under the same circumstances, we conduct two series of experiments with NYU v2 dataset. First, we train our framework to distinguish between 40 semantic classes and compare our results to [15] directly. We can observe from the Table 1 that our method achieves the best results and outperforms the original FCN by more than 4% in pixel accuracy. Expectedly, we also get progress in Mean IU which achieves 33.4% and outperforms both of the compared methods.

Table 1. Quantitative results on NYU v2 dataset.

Algorithm	Performance			
	Pixel Acc.	Mean Acc.	Mean IU	False Positives
FCN [15]	60.0	42.2	29.2	43726
FCN + Dense CRF [13]	61.5	43.4	31.5	22350
Benjamin et al. [16]	63.4	-	32.5	17668
Ours	65.5	46.0	33.4	9813

In the aspect of object-level, the number of False Positives defined earlier is used to evaluate the performance. FCN results in 43726 False Positives which are much more than any other methods. This is because the initial result of the FCN is coarse, and it is full of false positive samples that have been misclassified as described in Fig. 5. Although Benjamin et al. [16] have made a great improvement on this value, our approach shows a strong dominance in this respect. In our experiments on the test set, we reduce the False Positives by almost 78% over FCN and nearly 50% over [16]. Apparently, it is beneficial to utilize the global observation and hierarchical random fields to optimize the results.

In Fig. 5, we further visually display the qualitative comparison with the other approaches. It shows that the contours of the objects in FCN results are not very clear. More importantly, there are more or less different classes with Ground Truth. From the result of FCN with Dense CRF, we can observe that the performance does not get significantly improved. In our case, our method considers the global observation jointly and leverages the benefit from the HCRF. Therefore, it can achieve more consistent performance with the Ground Truth.

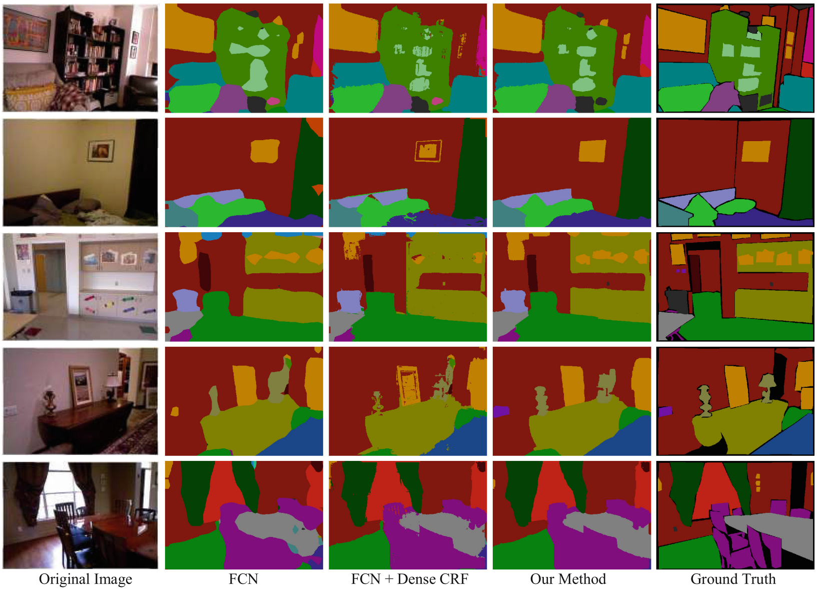

Fig. 5. Qualitative comparison with the other approaches. Left to right column: Original Image, FCN [15], FCN+ Dense CRF [13], Our Method and Ground Truth. Different colors indicate different classes.

5 Conclusion

We propose a novel approach that utilizes semantic knowledge to enhance the image segmentation performance. We formulate the problem in a hierarchical CRF integrated with the global observation. Our method achieves promising results in both pixel and object-level. However, the whole framework is not an end-to-end system and time-consuming. Future work includes replacing FCN with other approach which can achieve better performance on the initial segmentation. We will also improve the method by adding more semantic constrains rather than only using the pair-wise relation.

References

1. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
2. Baader, F.: The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, Cambridge (2003)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
4. Bannour, H., Hudelot, C.: Building and using fuzzy multimedia ontologies for semantic image annotation. Multimed. Tools Appl. **72**, 2107–2141 (2014)

5. Belongie, S., Carson, C., Greenspan, H., Malik, J.: Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval (1998)
6. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. Comput. Sci. **4**, 357–361 (2014)
7. Durand, N., et al.: Ontology-based object recognition for remote sensing image interpretation. In: 19th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2007, vol. 1, pp. 472–479. IEEE (2007)
8. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
9. Gupta, S., Arbelaez, P., Malik, J.: Perceptual organization and recognition of indoor scenes from RGB-D images. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 564–571. IEEE (2013)
10. Huang, Q., Han, M., Wu, B., Ioffe, S.: A hierarchical conditional random field model for labeling and segmenting images of street scenes. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1953–1960. IEEE (2011)
11. Knublauch, H., Fergerson, R.W., Noy, N.F., Musen, M.A.: The Protégé OWL plugin: an open development environment for semantic web applications. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 229–243. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30475-3_17
12. Kohli, P., Torr, P.H., et al.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vis. **82**(3), 302–324 (2009)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Advances in Neural Information Processing Systems, pp. 109–117 (2011)
14. Lin, G., Shen, C., Van Den Hengel, A., Reid, I.: Efficient piecewise training of deep structured models for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3194–3203 (2016)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
16. Meyer, B.J., Drummond, T.: Improved semantic segmentation for robotic applications with hierarchical conditional random fields. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 5258–5265. IEEE (2017)
17. Ruiz-Sarmiento, J.R., Galindo, C., Gonzalez-Jimenez, J.: Exploiting semantic knowledge for robot object recognition. Knowl. Based Syst. **86**, 131–142 (2015)
18. Ruiz-Sarmiento, J.R., Galindo, C., Gonzalez-Jimenez, J.: Scene object recognition for mobile robots through semantic knowledge and probabilistic graphical models. Expert. Syst. Appl. **42**(22), 8805–8816 (2015)
19. Russell, C., Kohli, P., Torr, P.H., et al.: Associative hierarchical CRFs for object class image segmentation. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 739–746. IEEE (2009)
20. Wang, H.H., Li, Y.F., Sun, J., Zhang, H., Pan, J.: Verifying feature models using owl. Web Semant. Sci., Serv. Agents World Wide Web **5**(2), 117–129 (2007)

21. Zand, M., Doraisamy, S., Halin, A.A., Mustaffa, M.R.: Ontology-based semantic image segmentation using mixture models and multiple CRFs. IEEE Trans. Image Process. **25**(7), 3233–3248 (2016)
22. Zheng, S., et al.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1529–1537 (2015)