

Enhanced Hierarchical Conditional Random Field Model for Semantic Image Segmentation

Li-Li Wang, Shan-Shan Zhu and Nelson H. C. Yung

*Laboratory for Intelligent Transportation Systems Research, Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong, SAR, China*

Keywords: Conditional Random Field, Semantic Segmentation, Image Segmentation, Pairwise Potential, Higher Order Potential.

Abstract: Pairwise and higher order potentials in the Hierarchical Conditional Random Field (HCRF) model play a vital role in smoothing region boundary and extracting actual object contour in the labeling space. However, pairwise potential evaluated by color information has the tendency to over-smooth small regions which are similar to their neighbors in the color space; and the higher order potential associated with multiple segments is prone to produce incorrect guidance to inference, especially for objects having similar features to the background. To overcome these problems, this paper proposes two enhanced potentials in the HCRF model that is capable to abate the over smoothness by propagating the believed labeling from the unary potential and to perform coherent inference by ensuring reliable segment consistency. Experimental results on the MSRC-21 data set demonstrate that the enhanced HCRF model achieves pleasant visual results, as well as significant improvement in terms of both global accuracy of 87.52% and average accuracy of 80.18%, which outperforms other algorithms reported in the literature so far.

1 INTRODUCTION

Semantic image segmentation can essentially be formulated as a labeling problem that attempts to assign a class label from a predefined label set to each pixel or super pixel in a given image (Boix et al., 2012); (Kohli and Torr, 2009); (Ladicky et al., 2009). Over the years, many assignment approaches have been explored with varying degree of success. One of the popular ideas is perhaps the use of Conditional Random Field (CRF) (Lafferty et al., 2001) combined with various potentials. The CRF is a discriminative model (Kumar and Hebert, 2006) that focuses on searching the optimal hyperplane for different classes. The labeling problem is thus solved by minimizing an energy function defined in the conditional random field over pixels or patches in the image (He et al., 2004); (Kohli and Torr, 2009); (Kumar and Hebert, 2005); (Ladicky et al., 2009); (Shotton et al., 2006), which can be quite effective in semantic image segmentation. For instance, one simple CRF model was described in (Boykov and Jolly, 2001) for object and background segmentation. In this model, only two potentials, unary potential and pairwise potential, are defined in the energy function. It achieved good performance

for a two-class segmentation on grey images. However, this model treats all the random variables on the same layer, which does not capture high level contextual information. Plath et al., (2009) added a global node over the basic layer for multi-class image segmentation. A consistency potential is then defined as a Potts model to penalize each local node which is different from the global one. As a result, it enforces all the local nodes in a region are assigned the same labels as the global node. This might not be capable to interpret large regions including multiple classes. Given these problems of the simple CRF model, complex CRF models, such as the hierarchical CRF (HCRF) models as described in (He et al., 2004); (Kohli and Torr, 2009); (Kumar and Hebert, 2005); (Ladicky et al., 2009), are then being proposed. The HCRF models fuse different scales of contextual information together to jointly perform labeling inference. The most representative of all the HCRF models is probably the one outlined in (Ladicky et al., 2009). Mathematically, the HCRF model is characterized by an energy function defined over the unary, pairwise, higher order and co-occurrence potentials. The first three potentials consider local interactions. Specifically, the unary potential is given by the observation of each pixel

from low level cues. The pairwise potential expresses the dependencies of neighboring pairwise pixels based on the difference in colors. The higher order potential encodes the interaction of long range pixels in super pixels or segments, while relationships between objects are captured by the co-occurrence potential based on global statistics. Note that the pairwise potential used in (Ladicky et al., 2009) is evaluated only on the basis of color differences to enforce a smooth labeling, it is not always a rational decision. For example, if the neighboring pixels have similar color features but belong to different objects, the pairwise potential could result in over smoothness. Another problem arises from higher order potential. Note that the higher order potential is guided by segments. While segmentation methods (Comaniciu and Meer, 2002); (Felzenszwalb and Huttenlocher, 2004); (MacQueen, 1967); (Shi and Malik, 2000); (Tan and Yung, 2008); (Zhu and Yung, 2011) are plentiful, different qualities of segments from being over-segmented to under-segmented are obtained. If a fine segment is used, better inference results are usually produced. In contrast, if a coarse segment is used, inappropriate guidance would result in mis-classifications.

To solve the above issues in the HCRF model (Ladicky et al., 2009), one contribution of this paper is to develop an enhanced model for pairwise potential. Considering the pairwise model itself may not incorporate enough information for an efficient inference, the newly constructed model depends not only on the contrast in the color space but also on the differences in the Laplacian space for an efficient inference. The believed labeling from unary potential is propagated to reduce the side effect of the pairwise model. Another contribution is to establish a discriminative model for the higher order potential. The discriminative model has the capability to select fine segments that involve in the inference process. Therefore, the higher order potential can also be called a segment-reliable consistency potential. Consequently, coherent classification results are obtained. Experimental results show that the enhanced HCRF model achieves significant improvement in terms of both global accuracy and average accuracy, as compared to other models in the literature.

In Section 2, we review the HCRF based method and its shortcoming for semantic image segmentation. In Section 3, we describe the details of the proposed method. Experimental results are given in Section 4, and the paper is concluded in Section 5.

2 CONDITIONAL RANDOM FIELD BASED METHOD FOR SEMANTIC IMAGE SEGMENTATION

2.1 Conditional Random Field for Semantic Image Segmentation

The aim of the CRF approach is to minimize an energy function $E(x)$ defined on a discrete random field X . Each random variable $X_i \in X$ corresponds to a node in the graphical model. The indexes of all basic nodes consist of a set of $V = \{1, 2, \dots, N\}$. The value x_i of each random variable X_i (or each node) represents the class label which takes a value from the label set $L = \{l_1, l_2, \dots, l_k\}$. Thus the labeling problem is to find a label for each node in the graphical model from the label set.

The energy function in the HCRF model is defined on unary, pairwise, higher order and co-occurrence potentials (Ladicky et al., 2009) as

$$E(x) = \sum_{i \in V} \phi_i(x_i) + \sum_{i \in V, j \in N_i} \psi_{ij}(x_i, x_j) + \sum_{c \in S} \psi_c^h(x_c) + C(L) \quad (1)$$

where V corresponds to the set of all pixels in an image, N_i is the set of neighboring pixels of pixel i . S is a set of cliques (super pixels or segments).

In Equation (1), the unary potential $\phi_i(x_i)$ is defined on a pixel i . It can be calculated as the negative log of the likelihood that pixel i is labeled as x_i . The likelihood can be obtained from the output of an adaptive boosted classifier (Ladicky et al., 2009; Torralba, Murphy, & Freeman, 2004) based on low level features (such as texton (Shotton et al., 2006), scale invariant feature transform (SIFT), color SIFT and local binary pattern (LBP)) of each pixel in an image.

The pairwise potential $\Psi_{ij}(x_i, x_j)$ encodes a smoothness prior between the neighboring random variables X_i and X_j . In (Ladicky et al., 2009), this potential is typically calculated as

$$\psi_{ij}(x_i, x_j) = w_0 + w_1 \exp\left(-\frac{W \|I_i - I_j\|^2}{w_2 * d_{ij}}\right), \quad j \in N_i, \quad (2)$$

where w_0 , w_1 and w_2 are model parameters whose values are learned based on the training data. The parameter d_{ij} denotes the distance between pixel i and pixel j . I_i is the color vector of pixel i , and W is the weight vector corresponding to three color components.

In Equation (1), $\psi_c^h(x_c)$ denotes the higher order

potential defined over a set of pixels (super pixel/segment) which is often generated from one of many unsupervised image segmentation methods. It is adopted to capture long range pixel interactions (region continuity), that is powerful in interpreting middle level structural dependencies between pixels in regions. $C(L)$ denotes the co-occurrence potential based on high level statistics for encoding the relationships between objects.

To perform inference, graph cuts-based method (Boykov et al., 2001); (Boykov and Jolly, 2001); (Kohli and Torr, 2009); (Ladicky et al., 2009); (Ladický et al., 2012); (Szumner et al., 2008) is used in the HCRF model to minimize the energy function in Equation (1).

2.2 Problems with Pairwise and Higher Order Potentials for Semantic Segmentation

In essence, the pairwise potential encodes a smoothness prior over neighboring variables. It penalizes two neighboring pixels which are labeled as different classes. In such a way, it is capable of smoothing the boundary of regions achieved by inferring unary potentials in the label space. However, it also results in an undesirable side effect. As depicted in Fig. 1, the boat is smoothed out when pairwise potential is added. One reason is that the boat and the water have similar color. By evaluating the pairwise potential in the color space, a larger penalty to force neighboring variables to adopt the same label is assigned by the graph cuts inference. In such case, pairwise potential results in over-smoothness of some regions.

In order to capture the fine contours of objects, higher order potential defined over a set of segments is incorporated into the HCRF model in Equation (1) by Ladicky et al (Ladicky et al., 2009). In (Ladicky et al., 2009), six layers of image segment are extracted based on two methods. Three layers of segment are generated by the K-Means clustering method (MacQueen, 1967), and the other three layers of segment are obtained by the Meanshift clustering method (Comaniciu and Meer, 2002). Fig. 2 (a2)-(f2) depicts the segmentation results using different parameters. In this example, only three potentials including unary potential, segment consistency potential and co-occurrence potential are considered instead of four potentials in Equation (1), to eliminate possible side effects from pairwise potential as discussed above. Generally speaking, unsupervised segmentation methods can extract more accurate contour of objects when the parameter

values of K-Means and Meanshift are increased. Segment consistency potential has the capability to integrate the same object under the guidance of unsupervised segmentations. From the inference point of view, it helps the labeling process recover from false unary predictions. However, if unsupervised segments are too coarse, such as the results for the cat scene as shown in Fig. 2 (e2) and (f2), the inferred boundaries are not reliable under the guidance of inaccurate segments. As a result, it results in false labeling, such as the labeling results in Fig. 2 (e1), (f1), (e3) and (f3).

3 ENHANCED PAIRWISE AND HIGHER ORDER POTENTIALS FOR SEMANTIC IMAGE LABELING

3.1 Enhanced Pairwise Potential

Note that the pairwise potential in Equation (2) is evaluated solely based on the color space. From Equation (2), we can see that a larger penalty is given if two pixels have similar color. As a result, neighboring pixels tend to have the same labeling under the pairwise smoothing constraints, which it is not always the best decision. It is evident there are significant between-class overlaps in terms of color only. Especially, when the size of the object is small, the resolution of an image is low or the image is blurred. In one of these scenarios, the pairwise potential tends to result in over-smoothing as shown in Fig. 1. In order to incorporate sufficient information to express the relationship of neighboring pixels, an extra term based on the edge space is added to calculate the smoothing constraints. As such, a second order derivative operator, the Laplacian operator, may be used to convolve with an image. It can extract detailed edge information of an image and is isotropic. In considering these advantages, we formulate an enhanced pairwise potential evaluation method, which is defined in both the color space and Laplacian space as given in Equation (3).

$$\psi_{ij}^e(x_i, x_j) = \begin{cases} 0, & \text{if } x_j = x_i^{unary}, \forall j \in N(x_i), \\ w_0 + w_1 K_1(I_i - I_j) K_2(G_i - G_j), & \text{otherwise} \end{cases} \quad (3)$$

where

$$K_1(I_i - I_j) = \exp\left(-\frac{W \|I_i - I_j\|^2}{\beta_1 * d_{ij}}\right), \quad (4)$$

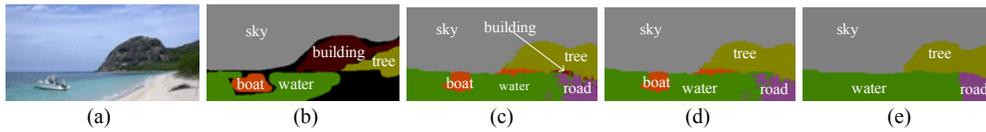


Figure 1: Impact of the pairwise potential: (a) Original image, (b) Groundtruth, (c) unary potential, (d) unary and co-occurrence potentials, (e) unary, pairwise and co-occurrence potentials.

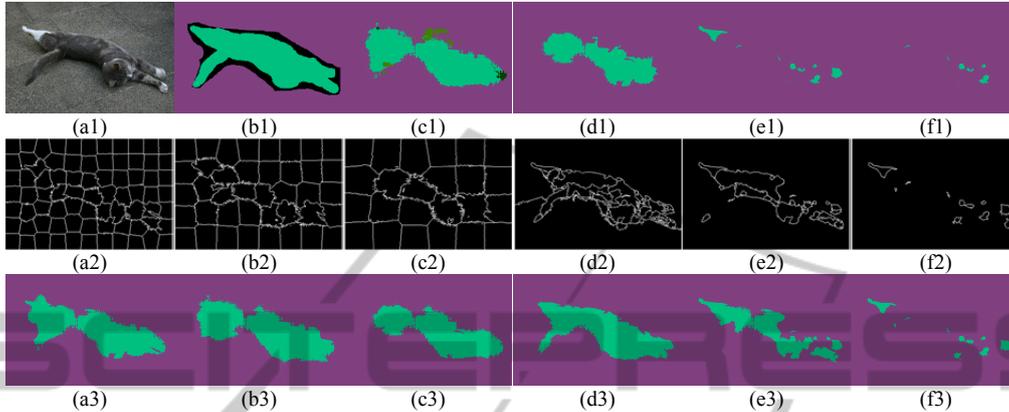


Figure 2: Unsupervised segmentation results and their semantic labeling results. (a1) Original image, (b1) Groundtruth, (c1) labeling result on unary potential, (d1) labeling result using three-layer K-Means segments, (e1) labeling result using three-layer MeanShift segments, (f1) labeling result using all six-layer segments; (a2)-(f2) Unsupervised segmentation results on (a2) K-Means(30), (b2) K-Means(40), (c2) K-Means(50), (d2) MeanShift(7.0x6.5), (e2) MeanShift(7.0x9.5), (f2) MeanShift(7.0x14.5); (a3)-(f3) Labeling results on (1) by using one-layer segments from (a2) to (f2), respectively.

and

$$K_2(G_i - G_j) = \exp\left(-\frac{W\|G_i - G_j\|^2}{\beta_2 * d_{ij}}\right) \quad (5)$$

In Equation (3), $N(x_i)$ denotes the indices of the set of pixel i (corresponding to the random variable X_i in the graphical model) and its neighboring eight pixels, and x_i^{unary} denotes the labeling of pixel i determined by the unary potential. If the random variable associated with its surrounding eight neighbors have the same labeling based on the minimization of unary potential function, the class label of this random variable is believed and propagated even if the pairwise potential is included in the energy function. In other words, the pairwise potential does not work when the class label of a random variable is propagated. Based on this criterion, the object classes with smaller sizes and similar color information to its adjacent objects are preserved. In Equation (3), $K_1(\cdot)$ and $K_2(\cdot)$ are two kernels defined in the color space and Laplacian space, respectively. They take forms as shown in Equations (4) and (5). In Equation (5), G_i is equal to the convolution between an image and Laplacian operator. w_0 , w_1 , β_0 and β_1 are model parameters, whose values are learned based on the training

dataset. By doing this, the pairwise potential is sensitive to contrast in both color and edge magnitude. To some extent, it suppresses the side effect of the original model.

3.2 Enhanced Higher Order Potential

Note that the higher order potential is defined over a set of segments. In the higher order term of Equation (1), the set S includes all segments from multiple-layer segmentations of an image by using two unsupervised segmentation algorithms. In (Kohli & Torr, 2009; Ladicky et al., 2009), the higher order potential takes the form of a robust P^n Potts model as

$$\psi_c^h(x_c) = \min_{l \in L} \left(\gamma_c^{\max}, \gamma_c^l + \sum_{i \in c} w_i k_c^l \Delta(x_i \neq l) \right), \quad (6)$$

where γ_c^{\max} denotes the maximum cost of the potential for segment c , γ_c^l represents the potential cost if the segment c takes a dominant label $l \in L$. $w_i k_c^l$ is used for calculating an additional penalty to each pixel in segment c without taking the label l . From Equation (6), we can see that the higher order potential encourages more pixels in segment c to take the dominant label l . This may result in over-

integrating some segments ((e3) and (f3) in Fig. 2) including more than one class label in the under segmentation situation (such as (e2) and (f2) in Fig. 2). To resolve this problem, we propose a segment-reliable consistency potential taking the form of

$$\psi_c^r(x_c) = \min_{l \in L} \left(\gamma_c^{\max}, \gamma_c^l + \sum_{i \in c} w_i k_c^l \Delta(x_i \neq l) \right) T[c \in S_r], \quad (7)$$

where $T[\cdot]$ is an indicator function, and S_r denotes the set of segments that provides more reliable guidance to an efficient inference. The minimization of the higher order potential can be solved by transforming it into an equivalent pairwise potential (Boros and Hammer, 2002); (Kohli et al., 2009); (Kohli and Torr, 2009); (Ladicky et al., 2009); (Rother et al., 2009). The critical problem is to determine which segments are reliable. In this paper, $T[c]$ is defined by Equation (8). When $T[c]$ is equal to one, it means that the segment c is reliable, and takes part in the inference process. Otherwise, the segment is excluded from the set S in Equation (1). In other words, the unreliable segments are not included in the energy minimization. As a result, the inference is not influenced by the unreliable segments any more, but decided by the other three potentials and the segment-reliable consistency potential.

$$T[c] = \begin{cases} 1, & \text{if } |c| > \alpha |I| \cap |c| - n_i(x_c) > \beta |c|, \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where parameters α (0.4) and β (0.1) can be learned from validation set. Consequently, the energy function is formulated in (9) for the enhanced HCRF model. The graph cuts algorithm proposed in (Ladicky et al., 2009) is then used to perform inference.

$$E(x) = \sum_{i \in V} \phi_i(x_i) + \sum_{i \in V, j \in N_i} \psi_{ij}^e(x_i, x_j) + \sum_{c \in S_r, c \subseteq S} \psi_c^r(x_c) + C(L) \quad (9)$$

4 EXPERIMENTAL RESULTS

Both the enhanced pairwise and higher order potentials have been tested on the MSRC-21 dataset (Shotton et al., 2006). They are evaluated based on the global and average-per-class recall criteria defined in (Ladicky et al., 2009). The MSRC-21 dataset includes 591 images with the resolution of 320×213 or 162×320 pixels, and 21 object classes. In our experiments, the dataset is typically partitioned into three sets including 45% for training, 45% for testing, and 10% for validation as

in (Ladicky et al., 2009); (Shotton et al., 2006). Each image has six-layer segments. Parameters for these six-layer unsupervised segmentations are set to the same as in (Ladicky et al., 2009).

To have a better understanding of the classification effects by adopting different potentials for these 21 object classes, four groups of semantic segmentation results have been generated and depicted in Table 1 based on the source code (automatic labeling environment, abbreviated to ALE) of the method in (Ladicky et al., 2009). “M0 (1 P)” denotes the classification based on the unary potential, which is also called the pixel-based random field (RF) method in (Ladicky et al., 2009). “M1 (2 Ps)” denotes the classification based on two potentials (unary and co-occurrence potentials). “M2 (3 Ps)” denotes the classification based on unary, pairwise and co-occurrence potentials. “M3 (4 Ps)” denotes the classification based on all four potentials. From the results in Table 1, it can be seen that the unary potential based inference has provided significant classification accuracy in both the overall (up to 83.56%) and average (up to 76.72%) categories. By fusing one more potential, further improvement in both categories is observed. When all four potentials are considered, roughly 3% and 1% increases as compared with the pixel-based RF method are achieved for the overall and average accuracy, respectively. This means that the HCRF model with higher order potentials (segment consistency potentials) is feasible by taking into account the interactions between different levels, and it is significantly superior to the one-layer CRF. However, it should also be noted that, compared with the pixel-based RF method, classification accuracies of some object classes, such as cow, cat, and boat, are substantially reduced when the segment consistency potential is included.

In Table 1, the experimental results of our proposed enhanced model are also presented. “iM2 (3 Ps)” denotes the classification based on three potentials which are similar to “M2 (3Ps)” but with the enhanced pairwise potential. “iM3 (4 Ps)” denotes the classification based on Equation (9). By substituting the pairwise potential in iM2 (3 Ps) with the enhanced version, the average classification accuracy are improved when compared with M2 (3 Ps). When both the enhanced pairwise and the segment-reliable consistency potentials are included, iM3 (4 Ps) achieves the best performance of 87.52% and 80.18% for global and average classification, respectively, which is slightly less than 1% of increase overall when compared with M3 (4 Ps), but close to 3% of increase in average

accuracy. As average accuracy is more representative in how well the method classifies, this percentage is clearly more significant. In terms of individual classes, the proposed model (iM3 (4 Ps)) performs equal or better in 18 classes (indicated by the bold font) when compared with M3 (4 Ps). When compared with the pixel-based RF method, the proposed model is superior in 17 classes. We also tried other datasets, such as Corel, Sowerby, Stanford used in (Ladicky et al., 2009), and the proposed algorithm still show better results than those in (Ladicky et al., 2009) in terms of both measurements.

Some of the successful classification results are depicted in Fig. 3 for visual evaluation. For objects (such as face and boat in Fig. 3) with smaller sizes in an image, they are often not discerned by the algorithm in ALE. By contrast, the enhanced HCRF model produces more pleasant results. Note that the appearance between different object classes may be

similar, such as cat and road in the third row of Fig. 3. Moreover, intra-class appearances are often not uniform, such as the cat in the fourth row of Fig. 3. By using the enhanced HCRF model, objects can be successfully segmented while the algorithm in ALE can only produce broken fragments.

To have a more comprehensive understanding of the failure cases, we focus on investigating the boat class, which has the lowest classification accuracy as shown in Table 1. From the confusion matrix, we note that boat is often mis-classified as building (20.5%), water (30.9%) or bicycle (17.5%).

Fig. 4 presents some of these cases for visual evaluation. It can be seen that the major reason for failure comes from the pixel-based RF classification.

In the pixel-based RF, low-level appearance features over a region about each pixel are adopted as the input to a boosted classifier (Ladicky et al., 2009); (Shotton et al., 2006) to determine its class label. However, overlaps in appearance features

Table 1: Classification accuracy on the MSRC-21 dataset in terms of percentage.

		Global	Average	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat
ALE	M0 (1 P)	83.56	76.72	67	96	90	87	88	93	84	81	89	76	90	80	59	40	93	61	87	82	52	80	34
	M1 (2 Ps)	84.03	77.19	69	96	91	88	90	93	85	81	89	77	91	81	59	41	93	59	87	84	52	82	34
	M2 (3 Ps)	84.43	77.42	70	97	91	88	91	94	83	82	89	77	92	82	60	40	94	60	88	85	53	81	32
	M3 (4 Ps)	86.87	77.67	76	99	91	75	86	99	78	88	87	76	88	93	76	51	95	65	92	68	52	79	18
Proposed	iM2 (3 Ps)	84.15	78.59	66	95	90	92	92	93	88	81	88	77	91	82	62	42	93	64	89	86	59	83	38
	iM3 (4 Ps)	87.52	80.18	76	99	90	86	94	96	84	88	88	79	89	93	74	51	95	68	92	83	56	81	24

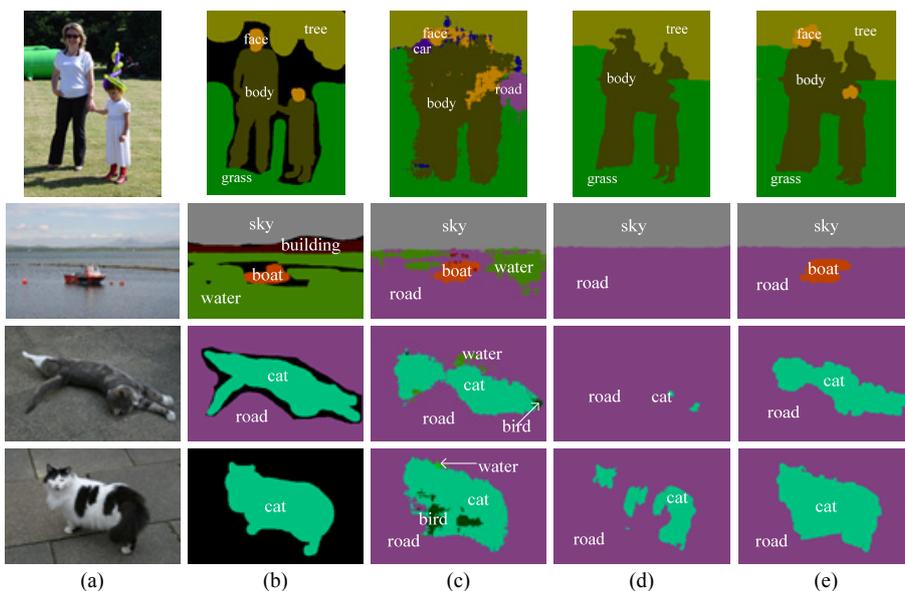


Figure 3: Some successful cases. (a) original image, (b) ground truth, (c) labeling result on unary potential, (d) labeling result based on ALE, and (e) labeling result based on the enhanced HCRF model.

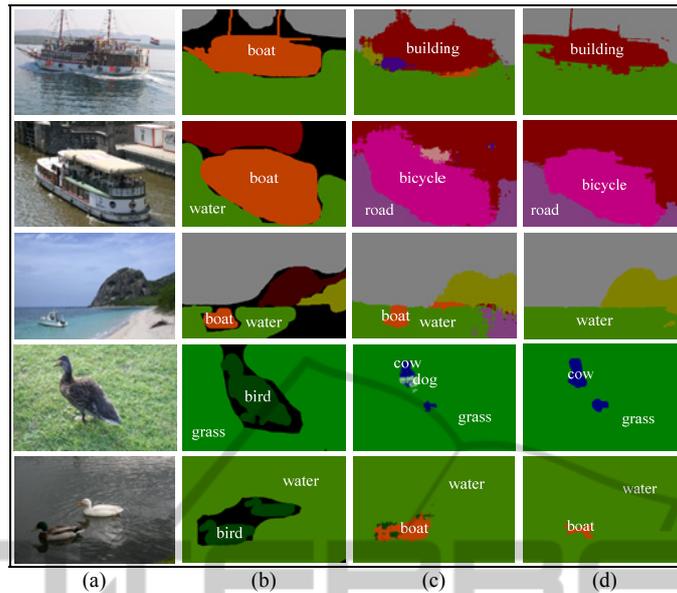


Figure 4: Some failure cases (a) Original image, (b) ground truth, (c) pixel-based RF and (d) enhanced HCRF.

(such as building and boat) between different classes confuse the inference. Furthermore, in under-segmented cases, objects with smaller block sizes (such as boat) are often merged with the background. As a result, the object classes cannot be inferred properly by the proposed HCRF model. For the bird class, which also has low classification accuracy, similar observations can be made as depicted in Fig. 4. Generally speaking, the major misclassifications are ascribed to two aspects for the pixel-based CRF model. One is being mis-classified as one of the adjacent object classes, and the other is mistakenly classified as the class with similar appearance. Both problems are propagated to the original HCRF model and the enhanced HCRF model, which eventually have limited their classification performance. If these problems can be resolved, higher classification success is expected for both models.

5 CONCLUSIONS

In conclusion, we have proposed an enhanced HCRF model for semantic image segmentation in this paper that performs significantly better in average classification accuracy than existing similar models. The proposed HCRF model consists of two enhanced potentials. The new pairwise potential comprises an additional Laplacian edge magnitude together with the original color differences. Moreover, it also propagates the believed labeling

determined by the unary potential to abate the over smoothness effect that the pairwise potential constraints lead to. The new segment-reliable consistency potential on the other hand is capable of selecting reliable segments to guide the inference. We have evaluated the enhanced HCRF model on the MSRC-21 data set, and the results show that the proposed model has achieved notable improvements in terms of both overall and average accuracy, when compared with other HCRF models. With regard to future research, focus will be placed on improving the performance of the unary potential by considering more discriminative features for object classes such as boat, bird, dog and chair.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, under Project HKU718912E.

REFERENCES

- Boix, Xavier, Gonfaus, Josep M, van de Weijer, Joost, Bagdanov, Andrew D, Serrat, Joan, & González, Jordi. (2012). Harmony potentials. *International journal of computer vision*, 96(1), 83-102.
- Boros, Endre, & Hammer, Peter L. (2002). Pseudo-boolean optimization. *Discrete applied mathematics*, 123(1), 155-225.

- Boykov, Yuri, Veksler, Olga, & Zabih, Ramin. (2001). Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11), 1222-1239.
- Boykov, Yuri Y, & Jolly, M-P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. Paper presented at the *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*.
- Comaniciu, Dorin, & Meer, Peter. (2002). Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5), 603-619.
- Felzenszwalb, Pedro F, & Huttenlocher, Daniel P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2), 167-181.
- He, Xuming, Zemel, Richard S, & Carreira-Perpinán, Miguel A. (2004). Multiscale conditional random fields for image labeling. Paper presented at the *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*.
- Kohli, Pushmeet, Kumar, M Pawan, & Torr, Philip HS. (2009). P³ & Beyond: Move Making Algorithms for Solving Higher Order Functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(9), 1645-1656.
- Kohli, Pushmeet, & Torr, Philip HS. (2009). Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3), 302-324.
- Kumar, Sanjiv, & Hebert, Martial. (2005). A hierarchical field framework for unified context-based classification. Paper presented at the *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*.
- Kumar, Sanjiv, & Hebert, Martial. (2006). Discriminative random fields. *International Journal of Computer Vision*, 68(2), 179-201.
- Ladicky, Lubor, Russell, Chris, Kohli, Pushmeet, & Torr, Philip HS. (2009). Associative hierarchical crfs for object class image segmentation. Paper presented at the *Computer Vision, 2009 IEEE 12th International Conference on*.
- Ladický, Lubor, Russell, Chris, Kohli, Pushmeet, & Torr, Philip HS. (2012). Inference Methods for CRFs with Co-occurrence Statistics. *International Journal of Computer Vision*, 1-13.
- Lafferty, John, McCallum, Andrew, & Pereira, Fernando CN. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Paper presented at the *Proceedings of Machine Learning*.
- MacQueen, James. (1967). Some methods for classification and analysis of multivariate observations. Paper presented at the *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*.
- Plath, Nils, Toussaint, Marc, & Nakajima, Shinichi. (2009). Multi-class image segmentation using conditional random fields and global classification. Paper presented at the *Proceedings of the 26th Annual International Conference on Machine Learning*.
- Rother, Carsten, Kohli, Pushmeet, Feng, Wei, & Jia, Jiaya. (2009). Minimizing sparse higher order energy functions of discrete variables. Paper presented at the *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*.
- Shi, Jianbo, & Malik, Jitendra. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8), 888-905.
- Shotton, Jamie, Winn, John, Rother, Carsten, & Criminisi, Antonio. (2006). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. Paper presented at the *Computer Vision–ECCV 2006*.
- Szummer, Martin, Kohli, Pushmeet, & Hoiem, Derek. (2008). Learning CRFs using graph cuts. Paper presented at the *Computer Vision–ECCV 2008*.
- Tan, Zhigang, & Yung, Nelson HC. (2008). Image segmentation towards natural clusters. Paper presented at the *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*.
- Torr, Philip HS, Murphy, Kevin P, & Freeman, William T. (2004). Sharing features: efficient boosting procedures for multiclass object detection. Paper presented at the *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*.
- Zhu, Shan-shan, & Yung, Nelson HC. (2011). Sub-scene generation: A step towards complex scene understanding. Paper presented at the *Multimedia and Expo (ICME), 2011 IEEE International Conference on*.