Adaptive Compatibility Matrix for Superpixel-CRF

Boquan Zhou[1](✉) and Chunpeng Li[2]

[1] School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China
lovelyfrog@bupt.edu.cn

[2] Chinese Institute for Brain Research, Beijing 102206, China
lichunpeng@cibr.ac.cn

Abstract. Compatibility Matrix plays a very important role in Fully-connected pairwise Conditional Random Field (full-CRF). However, former studies often fix it on any dataset or make too strong assumptions, which may cause abnormal object co-occurrence in image segmentation. The reason lies on the fixed compatibility matrix will give both normal objects co-occurrence and abnormal objects co-occurrence the same penalty. In this paper, we propose an adaptive compatibility matrix to replace the fixed compatibility matrix in full-CRF. Based on a weaker assumption of local independence, we propose the algorithm for adaptive compatibility matrix to learn from the dataset. In order to decrease the high computational complexity of full-CRF and maintain the accuracy at the same time, we build superpixel-CRF with adaptive compatibility matrix and propose the corresponding method to solve it. Our experiments demonstrate that the adaptive compatibility matrix improves the accuracy of full-CRF. The expansion in superpixel-CRF not only reduces the complexity but also performs well on the results.

Keywords: CRF · Image segmentation · Adaptive compatibility matrix

1 Introduction

The goal of multi-class image segmentation is to classify each pixel in an image to a label and split the image into different semantic patches, which is one of the most challenging problems in computer vision. Multi-class image segmentation is widely used in many applications such as video surveillance [4], object recognition [6], autopilot [5], etc.

A commonly used approach to solve segmentation problem is modelling each pixel as a node and an image as a graph. The segmentation problem is converted to a maximum a posteriori (MAP) inference defined over pixels or image patches [9,11,14,18,20]. The potentials in CRF incorporate pairwise potentials that model contextual relationships between object classes. However, former

E. Pimenidis et al. (Eds.): ICANN 2022, LNCS 13532, pp. 199–211, 2022.
https://doi.org/10.1007/978-3-031-15937-4_17

studies [3,13] often fixed the compatibility matrix in pairwise potentials, where the penalties between normal co-occurrence objects and abnormal co-occurrence objects are equal. That will cause some abnormal objects to appear together (Fig. 1). In order to solve that problem, we propose an adaptive compatibility matrix learned from the dataset to substitute the fixed compatibility matrix. In [8,20] people build local connected CRF, however it can not model long-range connections. For modeling long-range connections, [13] first builds the full-CRF. For solving CRF, traditional discrete optimization methods such as graph cut: α-expansion [3] or tree reweighted message passing(TRW) [12] work well for local connected CRF, but are too expensive for full-CRF. In [13], they develop an approximate optimization algorithm that is sublinear in the number of pairwise potentials. It is based on mean field inference, which is a local technique and the solution can be arbitrarily far from the optimum [10]. As a result, in [22] they develop an efficient graph cut optimization for full-CRF, assuming that image pixels have been tessellated into superpixels, and the weight of an edge between two pixels depends only on the superpixels these pixels belong to. The new superpixel-CRF decreases the computational complexity greatly and guarantees that the result will have the approximation factor of two, however it still suffers from the abnormal objects co-occurrence problem. In order to solve that problem, we fuse adaptive compatibility matrix into superpixel-CRF. Our contributions lie on 2 aspects: 1. We propose an adaptive compatibility matrix that is learned from the dataset by variational method based on local-independence, which can relieve the abnormal object co-occurrence problem in pixel-CRF. 2. We introduce the adaptive compatibility matrix into superpixel-CRF, which decreases the computation time greatly than pixel-CRF and relieve the abnormal object co-occurrence problem of the original superpixel-CRF.

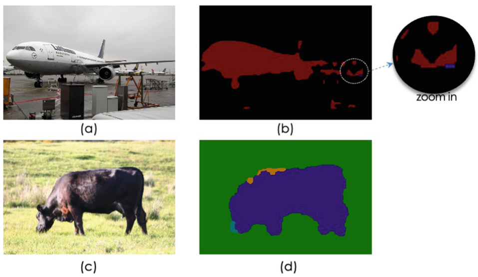

Fig. 1. Examples of abnormal objects co-occurrence. (a, c) is the original image, (b, d) is the corresponding segmentation result of potts model. (b) is the segmentation result of (a), black represents the background, red represents aeroplane, blue represents boat. (d) is the segmentation result of (c), green represents grass, dark blue represents cow, light blue represents sheep, orange represents face. (Color figure online)

2 Related Work

2.1 CRF and Superpixel-CRF

CRF is a kind of statistical modeling methods often applied in pattern recognition and machine learning. CRF models are composed of unary potentials and pairwise potentials on pixels or superpixels. There are many methods for solving sparsely connected CRF, such as iterative conditional modes(ICM) [21], TRW [12] and graph cut(expansion move, swap move) [3]. For densely connected CRF, mean field inference in [13] can solve it with fast speed; However, the result can be far from global minimum. [10] points that discrete optimization methods based on graph-cut can work better. To decrease computational complexity, in [2,17,23], they transform from pixel level to superpixel level. In [22], they consider that the weight of an edge between two pixels depends only on the superpixels they belong to, but the use of potts model [3] as compatibility function will cause the abnormal objects co-occurrence problem.

2.2 Compatibility Function

As an import part of overall potential, pairwise potential models relationship between two nodes. Pairwise potential consists of two parts: weight between two nodes and compatibility function $\mu(a, b)$ between two labels classified to these two nodes. In [15,20], they use potts model as compatibility matrix which is very simple: $\mu(a, b) = 1$ if $a \neq b$ else 0. In [21], they use truncated distance as compatibility value between two labels. No matter the potts model or truncated distance, they are all fixed values with no difference between different datasets. [13] makes an assumption that all nodes are independent and get their compatibility function, which is too strong because the pixels in an image are connected with each other and can not simply considered as independent. All the above compatibility function will cause the abnormal objects to appear. To be more precise, only two pixels are considered independent during the process of iteration in our study which maintain the connections between other pixels. In this article, we map compatibility function to a matrix M called compatibility matrix with $\mu(a, b)$ mapped to M_{ab}. With the compatibility matrix, the equations and the optimization problem will become more concise.

3 Preliminary

Consider an image $\mathbf{I}$ which is defined over all its pixels $\{I_1, I_2, ...I_N\}$, and a random field $\mathbf{X}$ which is defined over all its labels $\{X_1, X_2, ..., X_N\}$, where N is the number of pixels. I_i means the image vector of pixel i in $\mathbf{I}$, and X_i means the corresponding label of pixel i in $\mathbf{X}$. Based on the graph G constructed, the probability of $\mathbf{X}$ given $\mathbf{I}$ can be computed [16]:

$$P_G(\mathbf{X}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})} \cdot \exp(-E(\mathbf{X}|\mathbf{I})) \tag{1}$$

To simplify the notation, the subscript G and the condition $\mathbf{I}$ are omitted: $P(\mathbf{X}) = \frac{1}{Z} \cdot \exp(-E(\mathbf{X}))$. The $E(\mathbf{X})$ in P means the overall potentials:

$$E(\mathbf{X}) = \sum_{i} f_u(x_i) + \sum_{i,j} f_p(x_i, x_j) \tag{2}$$

$f_u(x_i)$ means the unary potential of pixel i, which is computed from an independent classifier. $f_p(x_i, x_j)$ means the pairwise potential between pixel i and pixel j, which is defined as [13]

$$f_p(x_i = a, x_j = b) = w_{ij} \cdot \mu(a, b) \tag{3}$$

where the w_{ij} means the weight between pixel i and pixel j, which is defined as

$$w_{ij} = \lambda_1 \cdot \exp\left(-\frac{||p_i - p_j||^2}{2\theta_1^2}\right) + \lambda_2 \cdot \exp\left(-\frac{||I_i - I_j||^2}{2\theta_2^2}\right) \tag{4}$$

where p_i means the position of pixel i, and I_i means the image vector or pixel i. $\mu(a, b)$ is the compatibility function that measures the penalty between any two pixels i and j classified as label a and b. In the potts model, if $a \neq b$, $\mu(a, b) = 1$ else 0.

4 Adaptive Compatibility Matrix

We substitute the compatibility function with compatibility matrix $\mathbf{M}$ for two reasons: one is the simplicity in notation, the other is that any value in the matrix is the penalty that two pixels classified to the two labels represented by the row and column of the value and you can observe the penalty visually in the heatmap of $\mathbf{M}$. Former studies often use the fixed compatibility matrix such as potts model [22]. However, using the fixed compatibility matrix will cause abnormal objects co-occurrence problem as Fig. 1 shows. In Fig. 1(a), some parts of aeroplane are classified as boat and in Fig. 1(d), some parts of cow are classified as sheep and face. The reason lies on two aspects. The first aspect is the unary classifier, which will give these places classified as boat low unary potentials. The second aspect is the compatibility matrix $\mathbf{M}$ of potts model, which considers the penalty that aeroplane-background co-occurrence and boat-background co-occurrence is the same, and the aeroplane-aeroplane co-occurrence is only tiny smaller than aeroplane-boat co-occurrence. As a result, some pixels will be classified as boats.

If the compatibility matrix is learned from the dataset, we will know which objects are less likely to appear together and give them bigger penalties. Then the abnormal objects co-occurrence problem can be get rid of. In order to learn the matrix, the following optimization problem needs to be focused on:

$$\arg\max_{\mathbf{M}} \prod_{i=1}^{K} P(\mathbf{T}^{(i)} | \mathbf{I}^{(i)}, \mathbf{M}) \tag{5}$$

where K is the size of our train dataset, $\mathbf{T}^{(i)}$ is the true label of image $\mathbf{I}^{(i)}$. After some transformations, we can get the loss $\mathcal{L}(\mathbf{M}) = \sum_{i=1}^{K} -log(P(\mathbf{T}^{(i)}|\mathbf{I}^{(i)}, \mathbf{M}))$ and the optimization problem becomes:

$$\arg\min_{\mathbf{M}} \mathcal{L}(\mathbf{M}) \tag{6}$$

However, it is impossible to get the analytic solution of M(as it's a transcendental equation). We need use stochastic gradient descent to optimize it gradually. For the sake of simplicity, we consider the loss on one image $\mathbf{I}$ and compute the partial derivative for M_{ab}:

$$\frac{\partial \mathcal{L}}{\partial M_{ab}} = \sum_{i,j} w_{ij} \cdot \left[1_{t_i=a} \cdot 1_{t_j=b} - \sum_{\mathbf{X}} P(\mathbf{X}) 1_{x_i=a} \cdot 1_{x_j=b} \right] \tag{7}$$

where $1_{x_i=a}$ means if $x_i = a$ it equals 1 otherwise 0.

However, the complexity of computing the $\sum_{\mathbf{X}} P(\mathbf{X}) 1_{x_i=a} \cdot 1_{x_j=b}$ is huge. If there are a total of B categories of labels, the time complexity will be $O(B^N)$. As a result, some approximations need to be considered to $P(\mathbf{X})$ and the computation need to be simplified. We can write $\sum_{\mathbf{X}} P(\mathbf{X}) 1_{x_i=a} \cdot 1_{x_j=b}$ as this form: $\sum_{\mathbf{X}} P(x_1, ...x_i = a, ..., x_j = b, ..., x_N)$, and then $P(x_i = a, x_j = b)$.

The overall energy $E(\mathbf{X})$ can be written as:

$$E(\mathbf{X}) = E(x_i, x_j) + E_{o.t}^{i,j}(\mathbf{X}) \tag{8}$$

where $E(x_i, x_j) = f_u(x_i) + f_u(x_j) + f_p(x_i, x_j)$, and $E_{o.t}^{i,j}(\mathbf{X})$ are all other terms in $E(\mathbf{X})$ except $E(x_i, x_j)$. Then we can get :

$$p(x_i = a, x_j = b) = \frac{\exp[-E(x_i = a, x_j = b)] \cdot \sum_{\mathbf{X}-x_i-x_j} \exp[-E_{o.t}^{i,j}(\mathbf{X})]}{Z} \tag{9}$$

In [13], they use mean field which considers that the classifications of all pixels are independent, and then they get the compatibility matrix. Obviously this assumption is too strong because in the real world, the classification is not independent at all. Instead we make a weaker assumption that the classifications of pixel i and j are independent, then $P(x_i = a, x_j = b) = P(x_i = a) \cdot P(x_j = b)$. We consider the approximation distribution as $Q(x_i, x_j)$. In order to get $Q(x_i, x_j)$, we minimize the KL divergence $D_{KL}(Q||P)$. Let $\sum_{\mathbf{X}-x_i-x_j} \exp[-E_{o.t}^{i,j}(\mathbf{X})]$ be $h(x_i = a, x_j = b)$. For solving this variational problem, we will get:

$$\begin{aligned} L(Q_i, Q_j) = & \sum_{x_i, x_j} Q(x_i, x_j)[E(x_i, x_j) - \ln h(x_i, x_j)] \\ & + \sum_{x_i=1}^{M} [Q_i(x_i) - \lambda_i] \ln Q_i(x_i) + \sum_{x_j=1}^{M} [Q_j(x_j) - \lambda_j] \ln Q_j(x_j) \end{aligned} \tag{10}$$

Using Euler-Lagrange equation we will have:

$$\frac{\partial L}{\partial Q_i} = \ln Q_i(x_i) + 1 - \lambda_i + \sum_{x_j} Q_j(x_j)[E(x_i, x_j) - \ln h(x_i, x_j)] \tag{11}$$

Then $Q_i(x_i)$ can be computed:

$$Q_i(x_i) = \frac{1}{Z} \cdot \exp[-\sum_{x_j} Q_j(x_j)E(x_i, x_j)] \cdot \exp[\sum_{x_j} Q_j(x_j) \ln h(x_i, x_j)] \quad (12)$$

For any label a, compared with the other terms in $h(x_i = a, x_j)$, the $\sum_n f_p(x_i = a, x_n)$ is much smaller for one order of magnitude. As a result, we can omit it. Then for any x_i, $\exp[\sum_{x_j} Q_j(x_j) \ln h(x_i, x_j)]$ will be the same. Finally we will get:

$$Q_i(x_i) = \frac{1}{Z} \cdot \exp[-\sum_{x_j} Q_j(x_j)E(x_i, x_j)] \quad (13)$$

where Z_i is the normalization factor. Then it needs to be solved iteratively. After getting the approximation distribution Q, we can compute the partial derivative with respect to M_{ab}. The details can be seen in Algorithm 1. Then the adaptive compatibility matrix $\mathbf{M}$ can be updated iteratively. It is easy to prove that the pairwise potential with M satisfies semi-metric [3], then we can use α-expansion to solve the CRF.

Algorithm 1 Variational Method: Calculate $\frac{\partial \mathcal{L}}{\partial \mathbf{M}}$

for each $i \in [1, N]$ **do**
    initialize $Q_i^{new}(x_i) = \frac{\exp[-f_u(x_i)]}{Z_i}$
end for
for each $a \in [1, B]$ **do**
    for each $b \in [a+1, B]$ **do**
        for each $i, j \in [1, N]$ **do**
            $Q_i^{old} = 0$;
            while $\max |Q_i^{new} - Q_i^{old}| > \epsilon$ **do**
                $Q_i^{old} = Q_i^{new}$
                for each $x_i \in [1, B]$ **do**
                    $Q_i^{new}(x_i)$
                    $= \frac{\exp\left[-f_u(x_i) - \sum_{j;j>i} \sum_{x_j} f_p(x_i, x_j) Q_j^{old}(x_j)\right]}{Z_i}$
                end for
            end while
        end for
        $\frac{\partial \mathcal{L}}{\partial M_{ab}} = \sum_{i,j} w_{ij} \left[1_{a=T_i} \cdot 1_{b=T_j} - Q_i(a) \cdot Q_j(b)\right]$
    end for
end for

5 Apply Adaptive Compatibility Matrix to Superpixel CRF

Solving full-CRF via graph-cut is very expensive especially when the resolution of the image is high. However, compared with mean field, graph cut works significantly better on the quality of solution [10]. In [22], they first try the superpixel-CRF, and use graph cut to solve it which decreases the computation time greatly. However, the accuracy also declines and it still suffers from the abnormal object co-occurrence problem. In order to relieve that problem, we introduce the adaptive compatibility matrix to superpixel-CRF. The pixels in the superpixel are similar to each other, as a result,the weights between any two pixels in one superpixel can be considered as the same which are called inner weights. Also, the weights between any pixel in one superpixel s and any pixel in another superpixel t can be considered as the same too, which are called outer weights. For two pixels i and j in the same superpixel, the inner weight is defined as

$$w_{ij} = \lambda_1 \exp(-\frac{\sigma_i^2}{2\beta_1^2})$$

where the σ_i^2 is the intensity variance of the superpixel that i belongs to. For two pixels i and j that belong to two different superpixels $S(i)$ and $S(j)$, the outer weight is defined as

$$w_{ij} = \lambda_1 \exp(-\frac{|d_i - d_j|^2}{2\beta_2^2}) + \lambda_2 \exp(-\frac{|\mu_i - \mu_j|^2}{2\beta_3^2})$$

where d_i and μ_i are the center and the intensity mean of the superpixel $S(i)$. We consider the binary class first, and then expand it to multi-class.

5.1 Binary Class

In this section, the energy function is assumed to contain two classes $\mathcal{L} = \{0, 1\}$. For the notation simplicity, we put the unary potential$f_u(x_i = 1) = f_u(x_i = 1) - f_u(x_i = 0)$, and $f_u(x_i = 0) = 0$ which is equivalent to the original problem.

The compatibility matrix for the binary class is:

$$\begin{bmatrix} 0 & M_{01} \\ M_{01} & 0 \end{bmatrix}$$

Consider the superpixel s and t, we assume that pixel $p, q \in P_s, a \in P_t$, where P_s, P_t is the pixel set of superpixel s and t. Assume that n_s is the number of pixels in P_s, y_s is the number of pixels that are assigned as 1 in P_s. Then the inner weight inside P_s is $M_{01} w_{pq}(n_s - y_s)$, the outer weight between P_s and P_t is $M_{01} w_{pa}[y_s(n_t - y_t) + y_t(n_s - y_s)]$. Let $g_s(y_s)$ be the unary potential of superpixel s which contains the unary potentials of all the pixels and all the pairwise potentials in s, $o(p)$ as the ranking of pixel p's original unary potential in the superpixel that it belongs to (if p has the least unary energy in the superpixel, then $o(p) = 1$), then:

$$g_s(y_s) = M_{01} w_{pq} y_s(n_s - y_s) + \sum_{p \in P_s, o(p) \leq y_s} f_u(x_p = 1) \tag{14}$$

The pairwise potential V_{st} between superpixel s and t which contains all the pairwise potentials between any pixels in s and t is:

$$V_{st} = M_{01} w_{pq} [y_s(n_t - y_t) + y_t(n_s - y_s)] \tag{15}$$

After $M_{01} w_{pq}(y_s n_t - y_s^2)$ transferred into $g_s(y_s)$ and $M_{01} w_{pq}(y_t n_s - y_t^2)$ transferred into $g_t(y_t)$, V_{st} equals $M_{01} w_{pq}(y_s - y_t)^2$ which satisfies semi-metric. Then the overall potential will be:

$$g_(y) = \sum_{s \in S} g_s(y_s) + \sum_{s,t \in S} V_{st}(y_s, y_t) \tag{16}$$

which can be solved easily via α-expansion.

5.2 Multi-class

Multi-class problem is our ultimate task and the knowledge in binary-class can be expanded to Multi-class. In each iteration, we only consider one label α. Assume that the result of the last iteration is $\mathbf{X}_{old}$, the pixels can be split into two parts: $x_i^{old} \neq \alpha$ and $x_i^{old} = \alpha$. For each pixel i, we define a binary-class classification z_i. The result after this iteration is defined as $\mathbf{X}$ and the relationship between $\mathbf{Z}$ and $\mathbf{X}$ is:

$$\begin{cases} z_i = 0, & \text{if } x_i = x_i^{old} \\ z_i = 1, & \text{if } x_i = \alpha \end{cases} \tag{17}$$

Then the new unary potential h_u of $\mathbf{Z}$ will be defined as: $h_u(z_i = 0) = f_u(x_i = x_i^{old}), h_u(z_i = 1) = f_u(x_i = \alpha)$. The new pairwise potential will be:

$$h_p(z_i = 1, z_j = 1) = 0 \tag{18}$$

$$h_p(z_i = 0, z_j = 0) = \begin{cases} 0, & \text{if } x_i^{old} = x_j^{old} \\ w_{ij} \cdot M_{x_i^{old} x_j^{old}}, & \text{if } x_i^{old} \neq x_j^{old} \end{cases} \tag{19}$$

$$h_p(z_i = 0, z_j = 1) = w_{ij} \cdot M_{x_i^{old} \alpha} \tag{20}$$

However, when $z_i = 0, z_j = 0$ and $x_i^{old} \neq x_j^{old}$, the pairwise potential is not 0, which does not satisfy the requirement of semi-metric. As a result, we need to do some transformations. After some terms are transferred from pairwise terms to unary terms, we get the new unary potential and new pairwise potential.

$$h_u(z_i = 0) = f_u(x_i = x_i^{old}) + \frac{1}{2} \sum_{j; x_i^{old} \neq x_j^{old}} w_{ij} \cdot (M_{x_i^{old} x_j^{old}} - 2M_{x_j^{old} \alpha}) \tag{21}$$

$$h_u(z_i = 1) = f_u(x_i = \alpha) \tag{22}$$

$$h_p(z_i, z_j | z_i = z_j) = 0 \tag{23}$$

$$h(z_i, z_j | z_i \neq z_j) = \begin{cases} w_{ij} \cdot M_{x_i^{old}\alpha}, & x_i^{old} = x_j^{old} \\ \frac{1}{2} w_{ij} \cdot [2M_{x_i^{old}\alpha} + 2M_{x_j^{old}\alpha} - M_{x_i^{old}x_j^{old}}], & x_i^{old} \neq x_j^{old} \end{cases} \tag{24}$$

After that, the new potentials are equivalent to the original ones and the current pairwise potentials satisfy semi-metric. Then the method in binary class can be used to transform the CRF from pixel level to superpixel level and the multi-class problem can be solved iteratively.

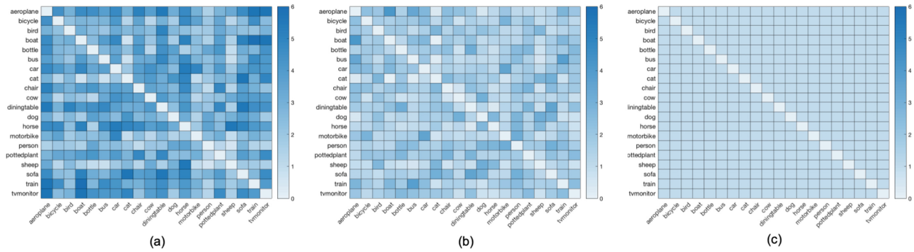

Fig. 2. Visualization of three compatibility matrices from VOC. (a) is the adaptive compatibility matrix using out method. (b) is the compatibility matrix using mean field. (c) is the compatibility matrix of potts model.

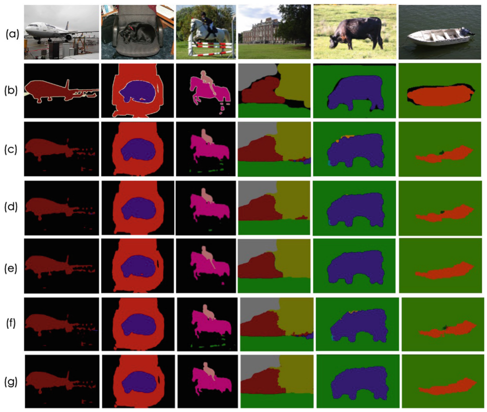

Fig. 3. The left 3 images are from VOC and the right 3 images are from MSRC-21. (a) Original images. (b) Ground truth. (c)–(e) Segmentation from fixed compatibility matrix, compatibility matrix computed by mean field and adaptive compatibility matrix of full-CRF. (f)–(g) Segmentation from fixed compatibility matrix and adaptive compatibility matrix of superpixel-CRF.

6 Experiments

We conduct many experiments on two standard benchmarks to see the improvement of the segmentation result. The first is PASCAL VOC 2012 (VOC) [7] which contains 1928 color images. The images in VOC are approximately 500×300 and there are 21 object class labels. The second is the MSRC-21 dataset [19], which consists of 591 color images of size 320×213 with corresponding ground truth labelings of 21 object classes. The experiment environment is person computer with Intel-i5 processor and 8G memory. The unary terms of VOC come from a pre-trained CNN classifier from [1] and the unary terms of MSRC-21 come from textonBoost [20]. We use SLIC [2] to compute superpixels, approximately 1000 per image for VOC and 800 per image for MSRC-21. To evaluate the performance, we compute the mean IOU for VOC and the pixel accuracy for MSRC-21.

First we learn the adaptive compatibility matrix of VOC and MSRC-21 respectively, using Algorithm 1. The heatmap of the three compatibility matrices from VOC using different methods can be seen in Fig. 2. The larger the penalty between the two labels is, the darker the square will be. In Fig. 2(a), it can be seen that the penalty of aeroplane-boat occurrence becomes more prominent than 1 which will solve the aeroplane-boat occurrence problem. Other than aeroplane-boat, the penalty of many abnormal objects co-occurrence also increase. Compared to our adaptive compatibility matrix, the matrix learned by mean field can not learn much useful information from the dataset. For many pairs of abnormal objects such as aeroplane-bicycle, boat-horse and bus-dining table, our adaptive compatibility matrix will give them large penalties while the matrix learned by mean field can not perform that well.

Then we substitute the fixed compatibility matrix with the learned matrix in full-CRF and superpixel-CRF with the use of α-expansion to get the segmentation result. In Table 1 and Table 2, potts model means using full-CRF with fixed compatibility matrix in [3], mean field means using full-CRF with compatibility matrix learned by mean field in [13], adaptive means using full-CRF with our adaptive compatibility matrix, sp+origin means using superpixel-CRF with fixed compatibility matrix in [22], sp+adaptive means using superpixel-CRF with our adaptive compatibility matrix. From Table 1 and Table 2, we can see that the adaptive compatibility matrix increases the mean IOU of VOC by 2 points and the pixel accuracy of MSRC-21 by 2 points than the fixed compatibility matrix in full-CRF. Moreover, the computation time of superpixel-CRF decreases greatly than full-CRF and superpixel-CRF with our adaptive compatibility matrix obtain good performance close to full-CRF. The visual result of three methods on full-CRF can be seen in Fig. 3(c, d, e) and two methods on superpixel-CRF can be seen in Fig. 3(f, g).

Table 1. The comparisons of mIoU and computation time of 5 methods on PASCAL VOC 2012.

	mIoU	Computation time
Potts model	71	4.1 min
Mean field	72	4.2 min
Adaptive	**73**	4.8 min
sp+origin	69	3.1 s
sp+adaptive	71	3.9 s

Table 2. The comparisons of pixel accuracy and computation time of 5 methods on MSRC-21.

	Pixel accuracy	Computation
Potts model	80	3.8 min
Mean field	81	3.9 min
Adaptive	**82**	4.6 min
sp+origin	77	2.8 s
sp+adaptive	80	3.7 s

7 Conclusions

In this paper, we propose the adaptive compatibility matrix of pairwise potential based on local-independence assumption and the corresponding algorithm to learn it from dataset. By introducing the adaptive compatibility matrix to full-CRF, we relieve the abnormal objects co-occurrence problem. However, the computation speed of solving full-CRF is too slow. Superpixel-CRF can be solved efficiently but due to the use of fixed compatibility matrix, it still suffers from the abnormal objects co-occurrence problem. As a result,we introduce the adaptive compatibility matrix to superpixel-CRF, and propose the corresponding method to solve it. The experiments show that superpixel-CRF with adaptive compatibility matrix not only decreases the computation time greatly but also alleviates the abnormal object co-occurrence problem and improves the performance.

References

1. Vedaldi, A., Gupta, K.L.: MatConvNet: CNNs for MATLAB (2016). http://www.vlfeat.org/matconvnet/pretrained/
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012). https://doi.org/10.1109/TPAMI.2012.120
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)

4. Cao, X., Gao, S., Chen, L., Wang, Y.: Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance. Multimedia Tools Appl. **79**(13), 9177–9192 (2020). https://doi.org/10.1007/s11042-018-7138-3
5. De Croon, G., De Wagter, C., Remes, B., Ruijsink, R.: Sky segmentation approach to obstacle avoidance. In: 2011 Aerospace Conference, pp. 1–16. IEEE (2011)
6. Delmerico, J.A., David, P., Corso, J.J.: Building facade detection, segmentation, and parameter estimation for mobile robot localization and guidance. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1632–1639. IEEE (2011)
7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results (2012). http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html
8. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 670–677. IEEE (2009)
9. He, X., Zemel, R., Carreira-Perpinan, M.: Multiscale conditional random fields for image labeling. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II-II (2004). https://doi.org/10.1109/CVPR.2004.1315232
10. Kappes, J., et al.: A comparative study of modern inference techniques for discrete energy minimization problems. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1328–1335 (2013)
11. Kohli, P., Torr, P.H., et al.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vis. **82**(3), 302–324 (2009). https://doi.org/10.1007/s11263-008-0202-0
12. Kolmogorov, V., Schoenemann, T.: Generalized sequential tree-reweighted message passing (2012)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with gaussian edge potentials. Adv. Neural Inf. Process. Syst. **24**, 109–117 (2011)
14. Kumar, S., Hebert, M.: A hierarchical field framework for unified context-based classification. In: Tenth IEEE International Conference on Computer Vision (ICCV 2005) Volume 1, vol. 2, pp. 1284–1291 (2005). https://doi.org/10.1109/ICCV.2005.9
15. Ladický, L., Russell, C., Kohli, P., Torr, P.H.: Associative hierarchical CRFs for object class image segmentation. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 739–746 (2009). https://doi.org/10.1109/ICCV.2009.5459248
16. Lafferty, J., McCallum, A., Pereira, F.C.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001)
17. Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K.: TurboPixels: fast superpixels using geometric flows. IEEE Trans. Pattern Anal. Mach. Intell. **31**(12), 2290–2297 (2009). https://doi.org/10.1109/TPAMI.2009.96
18. Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., Belongie, S.: Objects in context. In: 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007). https://doi.org/10.1109/ICCV.2007.4408986
19. Shotton, J., Winn, J., Rother, C., Criminisi, A.: *TextonBoost*: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_1

20. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context **81**, 2–23 (2009)
21. Szeliski, R., et al.: A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. IEEE Trans. Pattern Anal. Mach. Intell. **30**(6), 1068–1080 (2008)
22. Veksler, O.: Efficient graph cut optimization for full CRFs with quantized edges. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 1005–1012 (2020)
23. Veksler, O., Boykov, Y., Mehrani, P.: Superpixels and supervoxels in an energy optimization framework. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 211–224. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15555-0_16