

# Heterogeneous Conditional Random Field: Realizing Joint Detection and Segmentation of Cell Regions in Microscopic Images

Jiyan Pan

Robotics Institute

Carnegie Mellon University, USA

jiyanpan@cs.cmu.edu

Takeo Kanade

Robotics Institute

Carnegie Mellon University, USA

Takeo.Kanade@cs.cmu.edu

Mei Chen

Intel Labs Pittsburgh, USA

mei.chen@intel.com

## Abstract

*Detecting and segmenting cell regions in microscopic images is a challenging task, because cells typically do not have rich features, and their shapes and appearances are highly irregular and flexible. Furthermore, cells often form clusters, rendering the existing joint detection and segmentation algorithms unable to segment out individual cells. We address these difficulties by proposing a Heterogeneous Conditional Random Field (HCRF), in which different nodes have different state sets. The state sets are designed in such a way that the resulting HCRF model could encode all possible detection/segmentation cases while keeping its identifiability and compactness. Attributed to the provably optimal design of the state sets, the proposed model successfully realizes joint detection and segmentation of the cell regions into individual cells whether the cells are separate or touch one another. Experiments on two different types of cell images show that the HCRF outperforms several recently proposed methods.*

## 1. Introduction

Reliable detection and segmentation of individual cells in microscopic images is one of the most important components of vision-based automated systems for cell behavior analysis [18]. Unfortunately, such a task is highly non-trivial even when images are taken with advanced imaging systems such as phase-contrast microscopy. The difficulties lie in several aspects. Firstly, unlike many objects in natural images, cells under phase-contrast microscopy do not have distinctive textures. As a result, detection algorithms such as SIFT [11] that heavily rely on textures are not applicable. Secondly, cells are highly morphable. As is shown in Figure 1, all the cells are of the same type in a single image taken by a phase-contrast microscope. However, they have such a wide range of shapes and appearances that many of them cannot be nicely isolated by rectangu-

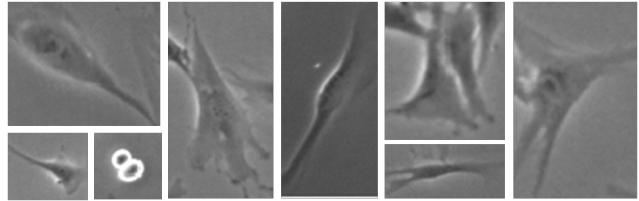


Figure 1. Examples of the same type of cells (C2C12 muscle stem cells) with disparate shapes and appearances. Note that a rectangular window is often not able to isolate a single cell without including other cells and/or the background.

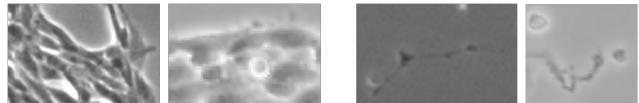


Figure 2. The left two images show clustered cells of two types (C2C12 muscle stem cells and bovine aortic endothelial cells). The right two images show background distractions for the two types of cells.

lar windows without including a large portion of other cells and/or the background. Consequently, detection methods based on sliding windows [6, 4, 7] are not effective. Thirdly, as we could observe in Figure 1, boundary-like structures within individual cells would fool vision algorithms into over-segmenting those cells. Fourthly, cells often form clusters, generating blurred boundaries and causing additional difficulties in segmenting out individual cells. Please see examples given in the left two images of Figure 2. Lastly, debris and other particles become strong background distractions in many cases, as is shown in the right two images of Figure 2.

Due to the high complexity of detecting individual cells as a whole, it would be more desirable to first make predictions on a large number of interest points as to how likely they belong to cell regions, as is proposed by Pan *et al.* [14]. Similar to object parts, interest points have more regularity than individual cells, and are therefore easier to classify. This is to some extent similar to part-based object detection [5, 1, 3], with the exception that the spatial relations of

the "parts" is highly flexible. In [14], interest points within cell regions are detected by thresholding the probability of each point belonging to a cell given local image features.

To obtain cell-level detection, we could reason over the interest points and segment the cell regions into individual cells by grouping those interest points based on how likely adjacent points come from the same cell. The method proposed in [14] transitively group the points whose pairwise conditional probability of being within the same cell is higher than a threshold. The total number of cells is automatically obtained at the end of the process. Alternatively, other segmentation techniques such as Normalized Cuts [9, 19, 12, 13] could be employed to segment the cell regions, yet the total number of cells, which is part of the information we want to infer, must be known in advance.

A major weakness of the approach in [14] is that detection and segmentation of the cell regions are performed separately, where each component requires a critical threshold that trades off between precision and recall. (Applying Normalized Cuts to the entire image does not circumvent this problem, because spectral clustering itself does not tell whether a segment belongs to a cell or background.) In fact, jointly dealing with detection and segmentation could enhance the performance for both of them. For instance, knowing that an ambiguous point lies close to an obvious cell point within the same plain region makes it more certain that the ambiguous point probably belongs to a cell. In other words, we need to enable the local detectors and segmenters to share information. In this process, a globally optimal solution could be reached, and critical thresholds could be reduced or even eliminated.

Conditional Random Field (CRF) [10] is a powerful tool to achieve joint detection and segmentation. Many algorithms based on CRF have been proposed to perform object detection and segmentation in natural images [17, 16, 15]. However, those algorithms are not readily applicable to the detection and segmentation of cell regions, because those algorithms perform *region level* rather than *object level* segmentation. In other words, if two objects of the same class are not completely separated in space, they will be regarded as a single region. This would cause a huge problem in segmenting cell images: since cells often form clusters, those algorithms would regard the entire cell cluster as a single segment.

Murphy *et al.* proposed iterative inference in CRF to tackle this difficulty [2]. In this algorithm, the state set of each node includes three states: background, the current cell, and all the other cells. Each round of inference segments out a single cell, and this process repeats until the segmented region is too small. However, there is an inherent contradiction in such a design of state set: the current cell and an adjacent cell belong to different classes, while two adjacent cells other than the current cell belong to the

same class. In fact, these two cases are statistically identical in terms of appearance in the image.

Ideally, we could design the state set of the CRF in such a way that each state is a unique cell ID, and then infer the optimal state (i.e. cell ID) for each node. Nevertheless, it is difficult to determine the size of the state set since we do not know the total number of cells. Even if we know the total number, optimal state assignment is still unidentifiable, because the IDs of different cells are interchangeable.

The main contribution of this paper is that we design a novel Heterogeneous CRF (HCRF) that jointly detects and segments all the cells in a single round of inference. The total number of cells is part of the optimal solution, and no critical parameter is required. We achieve this by defining *heterogeneous* state sets over the CRF: each node is associated with a state set *specific* to that node, which includes the background label and the cell IDs *allowed for that node*. The state sets are generated by a *restricted propagation rule*, and the inference over the HCRF is subject to a *non-maxima suppression rule*. Together, the resulting model is provably complete, sound, unique and irreducible in terms of encoding all possible detection/segmentation scenarios (please see definitions in Section 2).

The remainder of this paper is organized as follows. Section 2 describes the construction and inference over the Heterogeneous CRF (HCRF). The application of this model to the detection and segmentation of cell regions in phase-contrast microscopic images is described in Section 3. Experimental results are presented in Section 4, and we conclude the paper in Section 5.

## 2. Heterogeneous Conditional Random Field

### 2.1. Problem Reformulation

Joint detection and segmentation of cell regions in a microscopic image can be abstracted as the following problem.

*Given an image containing some unknown number of foreground objects, and a set of  $N$  interest points located either in the background or in one of the foreground objects, we want to detect all those interest points belonging to the foreground objects, and meanwhile segment those foreground points into individual sets corresponding to individual foreground objects.*

In what follows, we refer to cells as (*foreground*) *objects*, and interest points simply as *points*.

We construct an undirected graphical model to capture the probabilistic attributes of the points. Each point corresponds to a node in the graph. The structure of the graph is highly irregular, as the locations of the points could be arbitrary. We form the edges of the graph by linking each node with up to  $k$  nearest neighbors within radius  $r$ . The parameters  $k$  and  $r$  determine the complexity of the network, and the performance is insensitive to their specific values as

long as they are reasonably large.

The *state set* of each node (*i.e.* the set of states each node can take) consists of the background ID and object IDs. As each point can form a separate object, the number of object IDs is equal to the total number of points. The task of joint detection and segmentation is equivalent to assigning to each node a state from its state set. The log conditional distribution of such a joint assignment given the image  $X$  is

$$\log P(\mathbf{S}|X) = \sum_i \omega_i(S_i|X_i) + \sum_{i,j} \varphi_{ij}(S_i, S_j|X_{ij}) - \log Z(X), \quad (1)$$

where  $S_i$  is the state assigned to node  $i$ ,  $\mathbf{S} = \{S_i\}_{i=1}^N$  is the set of states assigned to all the  $N$  nodes,  $\omega_i(S_i|X_i)$  is the nodal potential of node  $i$  taking state  $S_i$  given local image features  $X_i$ , and  $\varphi_{ij}(S_i, S_j|X_{ij})$  is the pairwise potential of nodes  $i$  and  $j$  taking states  $S_i$  and  $S_j$  respectively, given pairwise image features  $X_{ij}$ .  $Z(X)$  is the partition function which is a constant during inference.

The nodal potentials can be directly computed from the probability of each node belonging to a foreground object given local image features:

$$\omega_i(S_i|X_i) = \begin{cases} 1 - P(n_i \in \mathcal{F}|X_i) & \text{if } S_i = 0 \\ P(n_i \in \mathcal{F}|X_i) & \text{if } S_i \neq 0 \end{cases} \quad (2)$$

where  $P(n_i \in \mathcal{F}|X_i)$  denotes the probability of node  $i$  belonging to the foreground region  $\mathcal{F}$  given local image features  $X_i$ . Note that  $S_i \in \mathbb{S}_i$ .

The pairwise potential can be computed from the conditional probability of an ordered pair of nodes having one of the five possible relations: background-background, same foreground object, background-foreground, foreground-background, and different foreground objects. Let  $P_{bb}$ ,  $P_{sf}$ ,  $P_{bf}$ ,  $P_{fb}$ ,  $P_{df}$  be the conditional probabilities of the five pairwise relations respectively, then we could define the pairwise potential as

$$\varphi_{ij}(S_i, S_j|X_{ij}) = \begin{cases} P_{bb} & \text{if } S_i = S_j = 0 \\ P_{sf} & \text{if } S_i = S_j \neq 0 \\ P_{bf} & \text{if } S_i = 0 \text{ and } S_j \neq 0 \\ P_{fb} & \text{if } S_i \neq 0 \text{ and } S_j = 0 \\ P_{df} & \text{if } S_i \neq 0 \text{ and } S_j \neq 0 \text{ and } S_i \neq S_j \end{cases} \quad (3)$$

Having defined the potentials, we perform the *most probable explanation (MPE)* inference over the graphical model using *max-product* belief propagation [8, 20] to obtain the optimal detection/segmentation solution  $\hat{\mathbf{S}}$ :

$$\hat{\mathbf{S}} = \arg \max_{\mathbf{S}} \log P(\mathbf{S}|X). \quad (4)$$

The problem seems to have been solved. However, as we will see soon, a lot of difficulties still need to be addressed.

## 2.2. Heterogeneous Conditional Random Field

If we simply make the state set of every node include the background ID and all possible object IDs, then this is a conventional CRF. However, such a state set design is highly redundant, although it could represent all the detection/segmentation cases. To see why, just assume that all the points belong to the same object, and then we have  $N$  different state assignments to represent this single case. Such redundancy renders the optimal state assignment unidentifiable. In addition, it would significantly increase computational complexity.

Therefore, we need to design a CRF with *heterogeneous* state sets that change size and content over different nodes. We also need to specify some rules when generating the state sets and doing inference, so that the resulting model is complete, irreducible, unique, and sound. The definitions of these properties, along with some introductory definitions, are given below.

**Definition 1** For a given graph  $G_N$  with  $N$  nodes, a Heterogeneous Conditional Random Field (HCRF) is a Conditional Random Field (CRF) which has a heterogeneous set of state sets  $\{\mathbb{S}_i\}_{i=1}^N$  associated with the nodes  $\{n_i\}_{i=1}^N$  of the graph, and which has a set of rules governing state set generation and inference.

**Definition 2** For a given graph  $G_N$  with  $N$  nodes and an HCRF with heterogeneous state sets  $\{\mathbb{S}_i\}_{i=1}^N$  and a set of rules, a *state assignment combination* is an ordered set of states  $\{S_i\}_{i=1}^N$  allowed by the rules of the HCRF such that each node  $n_i$  is assigned state  $S_i$  where  $S_i \in \mathbb{S}_i$ .

**Definition 3** Given graph  $G_N$ , we say an object in a detection/segmentation case is *legal* if, for any two nodes within the object, there exists at least one path connecting them along which every node belongs to that object. In other words, a *legal object* must be a connected region.

**Definition 4** A *legal detection/segmentation case* allowed by the structure of a graph  $G_N$  only consists of legal objects and/or the background.

**Definition 5 (completeness)** An HCRF is *complete* with respect to  $G_N$  if, for any legal detection/segmentation case allowed by the structure of  $G_N$ , there exists at least one state assignment combination that can realize such a case.

**Definition 6 (irreducibility)** An HCRF is *irreducible* with respect to  $G_N$  if removing any state from any individual state set  $\mathbb{S}_i$  would make the HCRF no longer complete.

**Definition 7 (uniqueness)** An HCRF is *unique* with respect to  $G_N$  if there exists at most one state assignment combination for the same legal detection/segmentation case.

**Definition 8 (soundness)** An HCRF is *sound* with respect to  $G_N$  if every state assignment combination represents a legal detection/segmentation case.

Now we show how to construct an HCRF (*i.e.* the heterogeneous state sets and the set of rules) that possesses all

the four properties.

Each node is associated with a specific state set which specifies all the eligible states the node could take. Suppose there are  $N$  nodes, and they are given an arbitrary order. For the  $m^{th}$  node, its state set consists of three types of states. The first type is called *background state* which takes on value 0, meaning the node could be in the background. The second type is called *initiating state* which takes on value  $m$ , meaning the node could initiate a new object indexed  $m$ . This could happen if all the  $m - 1$  lower-indexed nodes form  $m - 1$  different foreground objects. The third type is called *inherited states*. This type includes the initiating states of *other* nodes propagated to the  $m^{th}$  node via edges. For example, if the  $m^{th}$  node has an inherited state  $j$ , it means the  $m^{th}$  node could belong to object  $j$  initiated by the  $j^{th}$  node.

However, if we allow the initiating state of a node to propagate freely to all the other nodes, then all the nodes would have exactly the same state set which includes states 0 through  $N$ . As is previously discussed, this set of state sets is redundant, or, equivalently, not unique. Redundancy arises when the state sets of nodes belonging to the same object share states greater than the smallest initiating state among them. (E.g., for the upper object in Figure 3, the smallest initiating state among nodes 2 and 3 is 2, yet they both have states 3,4,5 in their state sets.) To remove this type of redundancy, we introduce the *restricted propagation rule*: the  $m^{th}$  node neither accepts nor passes on any propagated state greater than its own initiating state  $m$ . To see why, suppose an object contains  $M$  nodes, and node  $m$  has the smallest initiating state  $m$  among them. To represent the case that the  $M$  nodes belong to this object, they should take the same state, denoted as  $S$ . The restricted propagation rule prohibits node  $m$  from having any state greater than  $m$  in its state set. Hence,  $S$  cannot be greater than  $m$ , eliminating the type of redundancy exemplified in the right-most image of Figure 3. An example of the heterogeneous state sets generated under the restricted propagation rule is shown in the leftmost image of Figure 3.

Nevertheless, as  $S$  could be less than  $m$ , the restricted propagation rule alone does not guarantee the uniqueness of the HCRF. Similarly, the soundness of the HCRF may not be satisfied either. An example is illustrated in Figure 4. However, we solve all these problems by introducing a simple rule *after* the MPE inference over the graphical model. It is not hard to see that the redundant (or illegal) state assignment combinations have exactly the same conditional probability as the desirable state assignment combination. This (easy) unidentifiability can be directly resolved by enforcing the *non-maxima suppression rule*, which specifies that in case the MPE probability of a node is shared by several states, the node should take the largest state among them. The non-maxima suppression rule, along with the re-

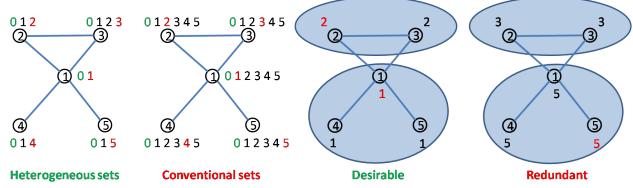


Figure 3. **Left:** The heterogeneous state sets generated under the restricted propagation rule. Here, the circled numbers are node indices, the green numbers indicate the ground state, the red numbers represent the initiating states, and the black numbers show the inherited states. **Middle left:** The state sets resulting from free propagation of the initiating states. **Middle right:** The desirable state assignment combination to represent a detection/segmentation case where blue ellipses indicate objects. **Right:** A redundant state assignment combination that represents the same detection/segmentation case as in the middle right image.

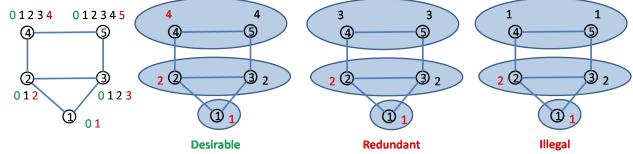


Figure 4. Violation of uniqueness and soundness when *only* the restricted propagation rule is enforced. **Left:** The heterogeneous state sets generated under the restricted propagation rule. **Middle left:** Desirable state assignment combination. **Middle right:** A redundant state assignment combination. **Right:** The state assignment combination represents an illegal detection/segmentation case in which object 1 does not form a connected region.

stricted propagation rule, ensures that the nodes belonging to the same object can *only* take a common state equal to the smallest initiating state among the nodes. In the example given in Figure 4, the MPE probability over the state set of node 4 is shared by states 1, 3, and 4. The non-maxima suppression rule forces node 4 to select state 4 as the final solution. The same happens to nodes 5. This way, both uniqueness violation and soundness violation are avoided.

Now we formally prove that the HCRF defined by the proposed heterogeneous state sets and the rule set consisting of the restricted propagation rule and the non-maxima suppression rule, is indeed complete, irreducible, unique, and sound.

**Theorem 1 (completeness)** For any given graph  $G_N$ , the HCRF defined by the proposed heterogeneous state sets and the rule set is complete.

**Proof** Any legal detection/segmentation case is composed of at most three types of segments: background region, single-node object, and multi-node object. For any background region formed by a set of nodes  $\{n_{b_i}\}_{i=1}^K$ , it can be realized by setting  $S_{b_i} = 0, i = 1, \dots, K$ , since 0 is always in the state set for any node. For any single-node object formed by node  $n_i$ , it can be realized by setting  $S_i = i$ , since the initiating state  $i$  is always in the

state set of node  $n_i$ . For any multi-node object formed by a set of nodes  $\{n_{f_i}\}_{i=1}^M$ , it can be realized by setting  $S_{f_i} = \min_i f_i, i = 1, \dots, M$ , where  $\min_i f_i$  is the smallest initiating state among the  $M$  nodes. This is realizable because every other node in the set is somehow connected to the node with the smallest initiating state (by the definition of legal detection/segmentation), and such a state is accepted by every other node (by the restricted propagation rule). End of proof.

**Theorem 2 (irreducibility)** For any given graph  $G_N$ , the HCRF defined by the proposed heterogeneous state sets and the rule set is irreducible.

**Proof** For any node  $i$  with a state set  $\mathbb{S}_i$ , removing the background state 0 from  $\mathbb{S}_i$  makes node  $i$  unable to be in the background; removing the initiating state  $i$  from  $\mathbb{S}_i$  makes it impossible to represent the case in which each node forms a different object; removing any inherited state  $k < i$  makes it impossible to represent the case in which nodes 1 through  $k - 1$  form  $k - 1$  different objects, and nodes  $k$  and  $i$  belong to the same object other than the previous  $k - 1$  objects. End of proof.

**Theorem 3 (uniqueness)** For any given graph  $G_N$ , the HCRF defined by the proposed heterogeneous state sets and the rule set is unique.

**Proof** Evidently, if there exist more than one state assignment combinations to represent a legal clustering case, then all those state assignment combinations would have exactly the same conditional probabilities. This means that, on the node level, the MPE probability of each node would be shared by several states corresponding to those state assignment combinations. The non-maxima suppression rule forces each node to take the largest state among those states. This ensures that the optimal solution given by the HCRF is unique, and that the optimal solution always satisfies, 1) for any single-node object, the node must take its initiating state; and 2) for any multi-node object, the nodes within the object must take a common state equal to the smallest initiating state among the nodes. (Note that the restricted propagation rule has already ruled out the possibility for the nodes taking a common state greater than the smallest initiating state among them.) According to the proof in Theorem 1, this optimal solution does represent the given detection/segmentation case. Therefore, the optimal solution is valid and unique. End of proof.

**Theorem 4 (soundness)** For any given graph  $G_N$ , the HCRF defined by the proposed heterogeneous state sets and the rule set is sound.

**Proof** If the HCRF is not sound, then there exists a solution of the HCRF that corresponds to an illegal detection/segmentation case in which a set of nodes taking a common state do not form a legal (*i.e.* connected) object. This set of nodes can always be decomposed into several groups of nodes where each group forms a connected ob-

ject. We could perform such a decomposition on every illegal object until all the objects are legal. Then these objects, along with the background, form a legal detection/segmentation case. As we know from the proof of Theorem 3, for any legal detection/segmentation case, the optimal solution given by the HCRF always assigns the smallest initiating state to the nodes in an object. As different legal objects do not overlap, they have different smallest initiating states. As a result, any unconnected set of nodes that compose an illegal object could never take any common state in the optimal solution given by the HCRF. This contradicts the assumption that the unconnected set of nodes form an illegal object by sharing a common state. End of proof.

Now we have proved that the HCRF possesses all the four properties. One may still wonder if the ordering of the nodes matters. As the proofs above do not assume any specific node ordering, the four properties are satisfied for any node ordering. However, the node ordering *does* affect the total number of states over all the nodes. (Note that irreducibility still holds for the heterogeneous state sets generated under any node ordering.) In order to reduce the total number of states as much as possible, a node with more neighbors should be given a lower index, because such a node would block more propagation paths of states greater than its index (which is equal to its initiating state) due to the restricted propagation rule.

The final detection/segmentation result can be obtained as follows: the nodes assigned with a non-zero state are detected as belonging to the foreground region, and individual foreground objects are segmented out by associating the nodes assigned with the same non-zero state. The total number of foreground objects is the total number of non-zero states assigned.

### 3. Cell Detection and Segmentation with the HCRF

In this section, we briefly describe detecting and segmenting cell regions in microscopic images using the HCRF detailed in Section 2.

We follow the approach proposed in [14] to obtain interest points. As each cell contains at least one local minima of intensity under phase-contrast microscopy, the interest points are selected as local minima of intensity. The locations of those local minima are further refined by a mean-shift process to provide more regularity. For the sake of comparison, we also extract the same unary features and pairwise features as in [14]. The unary features are used to train a *cell/background classifier* which is an SVM with a Gaussian kernel. In testing, the probabilistic output of the cell/background classifier gives the value of  $P(n_i \in \mathcal{F}|X_i)$  which is then used to compute the nodal potentials accord-

ing to Equation 2. Unlike the approach in [14], the pairwise features are used to train a *multi-relation classifier* which is a multi-class logistic regression classifier. In testing, the probabilistic output of the multi-relation classifier provides the values of  $P_{bb}$ ,  $P_{sf}$ ,  $P_{bf}$ ,  $P_{fb}$ , and  $P_{df}$  in Equation 3. Inference over the HCRF gives the detection/segmentation result over the interest points.

As cells cannot extend a long range, two nodes that are far apart cannot belong to the same cell. Therefore, the propagation of the initiating state of each node is conducted only within a certain radius centered at the node. This radius is set as 150 pixels, which is large enough. When constructing the graph, the radius to search for the nearest neighbors is set as  $r = 30$ , and the maximum number of the nearest neighbors is set as  $k = 5$ .

After classifying all the interest points, we need to classify the remaining pixels as well, so that cell boundaries can be obtained. To achieve this efficiently, we associate each remaining pixel with one of the interest points in its vicinity (within 30 pixels), and assign the label of that interest point to the pixel. To determine which interest point to choose, we compute, for each interest point nearby, the maximal intensity along the line connecting the pixel to the interest point, and the interest point with the lowest such intensity is associated with the pixel.

## 4. Experiments

In this section, we evaluate the performance of our proposed HCRF in cell detection/segmentation and compare it with the method proposed by Pan *et al.* [14] which treats detection and segmentation separately, as well as the approach proposed by Chen *et al.* [2] which employs a conventional CRF. Pan’s method could be used as is. To adapt Chen’s approach to our experiments and allow for a fair comparison, we set the “DNA potential” (i.e. nodal potential) to be the probabilistic output of the cell/background classifier, and the “boundary potential” (i.e. pairwise potential) to be the sum of  $P_{bf}$ ,  $P_{fb}$ , and  $P_{df}$  obtained from the multi-relation classifier.

We perform experiments on two different types of cells: bovine aortic endothelial cells and C2C12 muscle stem cells. For each type of cell, 10 images are used for training and 10 images for testing. Although the number of images is relatively small, yet each image contains a large number of cells, and the algorithms need to classify each individual interest point and segment out each individual cell. For the first type of cells, 20 images in total contain about 4,900 cells, 15,200 interest points, and 41,800 pairs of points. The statistics for the second type of cells are 9,800 cells, 94,100 interest points, and 260,840 pairs of points. The results obtained in our experiments are therefore statistically significant.

The cell/background classifier gives an initial probability

Point-level	Pan[14]	HCRF	Chen[2]
Bovine	P $92.8\pm1.5\%$	<b>97.2±0.7%</b>	$95.4\pm0.8\%$
	R $90.5\pm2.7\%$	<b>92.2±2.1%</b>	$85.3\pm2.2\%$
	F $91.6\pm1.0\%$	<b>94.6±1.0%</b>	$90.0\pm1.5\%$
C2C12	P $92.9\pm2.0\%$	<b>96.5±1.4%</b>	$95.2\pm1.5\%$
	R <b>85.3±3.6%</b>	$85.2\pm3.4\%$	$85.5\pm3.1\%$
	F $88.8\pm1.2\%$	<b>90.4±1.7%</b>	$90.1\pm1.4\%$

Figure 5. Quantitative comparison of point-level detection performance. “P”, “R”, and “F” represent precision, recall and F-measure, respectively.

Cell-level	Pan[14]	HCRF	Chen[2]
Bovine	P $95.1\pm1.0\%$	<b>96.1±1.5%</b>	$94.2\pm1.9\%$
	R $89.1\pm2.4\%$	<b>92.1±1.5%</b>	$44.5\pm1.4\%$
	F $91.9\pm1.2\%$	<b>94.0±0.7%</b>	$60.4\pm1.5\%$
C2C12	P <b>96.4±0.6%</b>	$93.8\pm1.0\%$	$90.5\pm1.4\%$
	R $83.7\pm1.5\%$	<b>91.4±1.7%</b>	$72.0\pm2.1\%$
	F $89.6\pm1.0\%$	<b>92.5±1.2%</b>	$80.1\pm1.1\%$

Figure 6. Quantitative comparison of cell-level detection/segmentation performance. “P”, “R”, and “F” represent precision, recall and F-measure, respectively.

Shape Quality	Pan[14]	HCRF	Chen[2]
Bovine	$0.63\pm0.04$	<b>0.64±0.04</b>	$0.58\pm0.03$
C2C12	$0.59\pm0.02$	<b>0.62±0.02</b>	$0.57\pm0.02$

Figure 7. Quantitative comparison of the shape quality for correctly detected cells. Please see the text for details.

of each point belonging to cell regions. The method in [14] sets a threshold on this probability to obtain the detection result on a point level. When the threshold is set as 0.5, the correctness of such prediction on an example testing image for the two types of cells is shown in the second column from the left in Figure 9. We could see that detection using unary features alone tends to make mistakes where cells form clusters and/or have non-typical appearances. Mistakes are also likely in face of background distractions.

The quantitative evaluation of point-level detection is summarized in Figure 5, where the mean and standard deviation of precision, recall and F-measure are computed over all the testing images for each type of cells. The performance is enhanced by up to 3% over [14] when detection is performed jointly with segmentation using the HCRF. Comparing the third column with the second column in Figure 9, we observe that many mistakes made by the cell/background classifier are corrected by the HCRF inference. Although the method in [2] also jointly performs detection and segmentation, the use of a conventional CRF with suboptimal state sets leads to less improvement, or even some deterioration, in the detection performance. This is illustrated both qualitatively in the rightmost column in Figure 9 and quantitatively in Figure 5.

As for cell-level detection/segmentation performance, the HCRF still outperforms the other algorithms. Figure 10 displays the final detection/segmentation results on an example testing image for the two types of cells, and quantitative results are listed in Figure 6. Compared with the

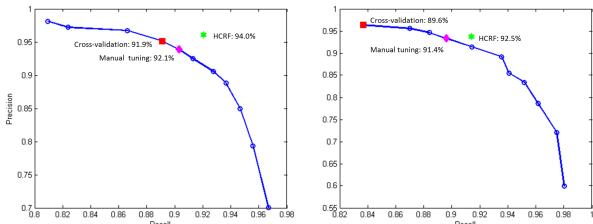


Figure 8. Comparison of the HCRF with the best possible performance of the method in [14]. The left plot is for bovine cells, and the right plot for C2C12 cells. The blue curve is the precision-recall curve obtained by setting different values of the two critical parameters in the method of [14]. The red square indicates the performance of using cross-validation to determine the parameters. The purple diamond indicates the best possible performance by manually tuning the parameters. The green star indicates the performance of the HCRF. Corresponding F-measures are shown next to the symbols.

algorithm of [14], the HCRF achieves an F-measure about 3% higher for C2C12 cells and 2% higher for bovine cells. Such an improvement is significant, because unlike the algorithm in [14] which has two critical parameters to tune, the HCRF does not require any critical parameters. The performance of the method in [2] is rather poor due to the inherent contradiction of its state sets as is mentioned in the Introduction. It is observed that the method tends to mistakenly associate points belonging to different cells. This leads to a large number of missed cells. Such a behavior is understandable, because the CRF in [2] is asked to classify all the other individual cells as the same class, which is not reasonable. The HCRF circumvents this problem by employing the heterogeneous state sets that could represent every individual cell.

In addition to precision and recall, we also report the quality of shapes for the cells correctly detected. For an algorithm-generated cell and a ground-truth cell that form a match, the shape quality of the algorithm-generated cell is measured as the ratio of the area of the intersection to the area of the union of the two cell regions. The average shape quality of all the correctly detected cells is listed in Figure 7 for the two types of cells. We can see that the HCRF also achieves the highest shape quality for those correctly detected cells.

Finally, we show in Figure 8 that the optimal detection/segmentation solution obtained by the HCRF is even better than the best performance achievable by tuning the parameters along the precision-recall curve generated by the approach in [14]. This is not surprising: as long as we have an optimal design of the state sets, the joint detection and segmentation in the HCRF provides additional information that is not available when detection and segmentation are performed separately.

## 5. Conclusion

In this paper, we propose a Heterogeneous Conditional Random Field (HCRF) to perform joint detection and segmentation of cell regions in microscopic images. In order to make the graphical model expressive enough to encode all detection/segmentation cases while avoiding unidentifiability and excessive complexity, we carefully design a heterogeneous set of state sets and a rule set, such that the model is provably optimal. Such a model design enables our algorithm to effectively incorporate detection and segmentation into a unified probabilistic framework and achieve an optimal solution without the need of tuning any critical parameters.

## References

- [1] Y. Amit and A. Trouve. Pop: Patchwork of parts models for object recognition. *IJCV*, 75(2):267–282, 2007.
- [2] S.-C. Chen, T. Zhao, G. J. Gordon, and R. F. Murphy. A novel graphical model approach to segmenting cell images. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2006.
- [3] D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatial priors for part-based recognition using statistical models. *CVPR*, 2005.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *CVPR*, 1:886–893, 2005.
- [5] B. Epshtain and S. Ullman. Semantic hierarchies for recognizing objects and parts. *CVPR*, 2007.
- [6] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. *CVPR*, 2008.
- [7] N. He, J. Cao, and L. Song. Scale space histogram of oriented gradients for human detection. *International Symposium on Information Science and Engineering*, 2:167–170, 2008.
- [8] T. Heskes. Stable fixed points of loopy belief propagation are minima of the bethe free energy. *Advances in NIPS*, 2003.
- [9] S. Jianbo and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on PAMI*, 22(8):888–905, 2000.
- [10] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML*, 2001.
- [11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [12] M. Meila and J. Shi. Learning segmentation with random walk. *NIPS*, 2001.
- [13] A. Y. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *NIPS 14*, 2002.
- [14] J. Pan, T. Kanade, and M. Chen. Learning to detect different types of cells under phase contrast microscopy. *Microscopic Image Analysis with Applications in Biology*, 2009.
- [15] X. Ren, C. C. Fowlkes, and J. Malik. Figure/ground assignment in natural images. *ECCV*, 2006.

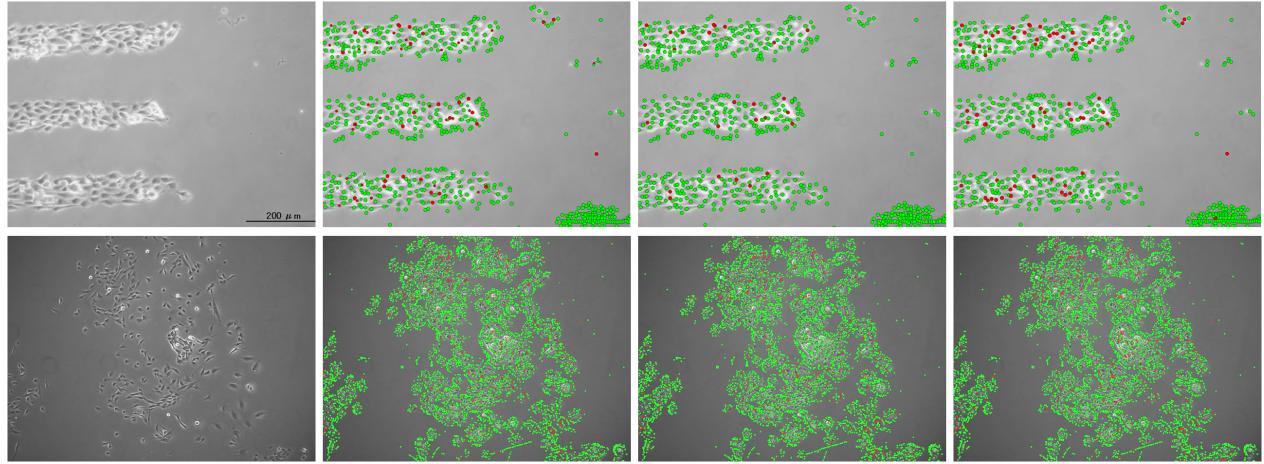


Figure 9. Qualitative comparison of point-level detection performance. The two rows from the top show bovine aortic endothelial cells and C2C12 muscle stem cells, respectively. The four columns from the left display the original image and the results of [14], the HCRF and [2], respectively. The green and red dots indicate correct and wrong detections, respectively.

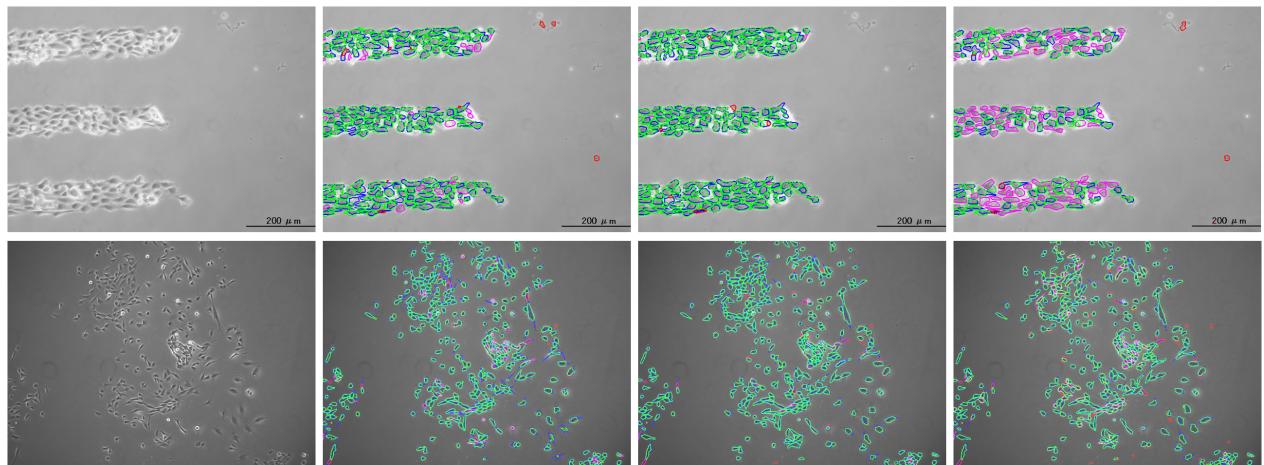


Figure 10. Qualitative comparison of cell-level detection performance. The two rows from the top show bovine aortic endothelial cells and C2C12 muscle stem cells, respectively. The four columns from the left display the original image and the results of [14], the HCRF and [2], respectively. The green and red boundaries show the algorithm-generated cells that can and cannot find a match with ground-truth cells, respectively. (*I.e.*, the red boundaries are false detections.) The blue and magenta boundaries show the ground-truth cells that can and cannot find a match with algorithm-generated cells, respectively. (*I.e.*, the magenta boundaries are missed cells.) An algorithm-generated cell can only be matched with at most one ground-truth cell, and vice versa. The optimal match is obtained by maximizing the overlapping area of all the algorithm-generated and ground-truth cells, subject to the constraint just mentioned. The optimization is realized using integer programming.

- [16] J. Shotton, M. Johnson, and R. Cipolla. Semantic textron forests for image categorization and segmentation. *CVPR*, 2008.
- [17] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textron-boost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. *ECCV*, 2006.
- [18] M. Usaj, D. Torkar, and D. Miklavcic. Automatic cell detection in phase-contrast images for evaluation of electroporation efficiency in vitro. *Mediterranean Conf. on Medical and Biomedical Engineering and Computing*, pages 851–855, 2007.
- [19] L. Xu, W. Li, and D. Schuurmans. Fast normalized cut with linear constraints. *CVPR*, pages 2866–2873, 2009.
- [20] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *IJCAI*, 2001.