Image Labeling with Markov Random Fields and Conditional Random Fields

Shangxuan Wu
Robotics Institute
Carnegie Mellon University
shangxuw@andrew.cmu.edu

Xinshuo Weng
Robotics Institute
Carnegie Mellon University
xinshuow@andrew.cmu.edu

Abstract

Most existing methods for object segmentation in computer vision are formulated as a labeling task. This, in general, could be transferred to a pixel-wise label assignment task, which is quite similar to the structure of hidden Markov random field. In terms of Markov random field, each pixel can be regarded as a state and has a transition probability to its neighbor pixel, the label behind each pixel is a latent variable and has an emission probability from its corresponding state. In this paper, we reviewed several modern image labeling methods based on Markov random field and conditional random Field. And we compare the result of these methods with some classical image labeling methods. The experiment demonstrates that the introduction of Markov random field and conditional random field make a big difference in the segmentation result.

1 Introduction

In the field of computer vision, segmentation [14, 2, 28, 15], i.e. image labeling often plays an intermediate but important role for some higher level image understanding and recognition. To some extent, the precision and accuracy of image labeling could be a bottleneck of these future work. So the challenge of achieving more find-grained segmentation result always attracts lots of research interest in the field.

In terms of image labeling task, each individual pixel of a given image needs to be assigned one specific label from a predefined discrete label set. But it's not possible to figure out what the best label is for each pixel by just looking at its pixel value. So the context information is needed for the labeling task. Some low level information and pattern in the image is also quite useful. For example, the color and texture might give a good sense of what the object class is, because pixel with the same color and texture might has higher possibility to be assigned a same label. But the challenge is some different classes (sky and water) share same color or texture and some objects involve multiple colors (human clothing). For disambiguating these confusion, more attempts in the field has incorporated additional information such as localization property, which means pixels with the same color and also close to each other should be assigned a same label. Also many graphical models have been used for modeling this task.

In this paper, we mainly focus on modeling the image labeling task by exploring the Markov property. One could always formulate all pixels in the image as an undirected graphical model. More specifically, to limit the scope of this paper, we only consider graphical model with 2D Markov property. That means the state of each pixel can only be determined by the state from a pre-defined set of pixel, such as four immediately connected neighbor pixels. In terms of labeling task, each pixel needs to be associated with one specific label from label classes. This refers to emission probability from the perspective of hidden Markov random field. Because of this similarity between Markov random field and image labeling, we could simply transfer the image labeling task to model the full

CMU 16811 Math Fundamentals for Robotics Course Project, (2016).

joint probability of the image and corresponding labels. In addition to the basic review of MRFs method, we focus more on conditional random field, which only models the conditional probability of labels given the image. This decreases a huge mount of parameter to be estimated in the model. Our comparative experimental results demonstrate that CRFs allow us to get more robust segmentation result with easier and faster inference process. Compared to the classical methods, the experiments validate our claim that model imposing the Markov property is superior to the traditional methods using only low level image information.

2 Related Works

One response to image labeling task is traditional unsupervised methods involving all kinds of low level information. By measuring color [6] and texture similarity, lots of attempts involving clustering methods such as k-means [24] and mean-shift [5] try to group the local region with higher similarity and assign them with a same label. For computational convenience, superpixel [17, 26] approach is often used as pre-processing step to group the potential similar pixel together. Then all previous approach works on the superpixel group instead of individual pixel. But the problem of these approach occurs due to the use of only local information. So research work starts to combine global and local context to get a better sense of semantic information in the image. In [22] a multi-scale structure is used for segmenting the image based on edge detection with different scale level. In [20, 21], combination of global context and local informationis proved to be very useful in segmentation task.

Another set of research work for solving image labeling task in vision is to couple one or more other high-level and related tasks with segmentation. In other words, additional high-level information is used for solving labeling task simultaneously. For example, in [19, 8, 11, 12, 13] multi-task learning has been explored to solve object detection and segmentation simultaneously. In [4] attention model is used to improve the performance of segmentation. Other related tasks such as edge detection, depth estimation and surface normal are also leveraged to couple with segmentation and achieve better result in [9, 3]. Even human interaction [1] with the input is very useful to provide additional information for better segmentation result.

The first time that Markov random field is introduced for segmentation is in [27], where it's only for the application of medical image. Then the attempt of modeling segmentation task with Markov random field is largely growing due to the similarity of their representation. But learning and inference for Markov random field is quite computationally expensive so that lots of approximation and learning scheme is invented for obtaining the robust segmentation result. In addition to Markov random field, one point worth to highlight is conditional random field, which in [14, 23, 16] is heavily used for segmentation in the recent decades. The benefit of conditional random field is that, for image labeling task, we actually don't care about the joint distribution of image pixel. So modeling only the label conditioned on image observation is enough and will decrease huge mount of parameters to estimate. Most recently, with the revolution of deep learning, conditional random field is usually used as a post-precessing step in [2] to refine the segmentation result produced by deep convolutional neural network. More attempt [28] etries to incorporate conditional random field into network architecture instead of post processing and achieve better result.

3 Random Field for Image Labeling

In this section, we review basics about how to model the segmentation task using Markov random field and conditional random field respectively. Obviously, image labeling needs information across many neighboring pixels, which we called contextual or local information. We can reasonable assume that the nearby pixels are more likely to be assigned a same label. With the use of Markov random fields, we typically formulated a probabilistic generative framework modeling the joint probability of the image and its corresponding labels. This allows our model to locally smooth the assigned labels. But the underlying generative model by MRFs is actually more complicated than what we need, because we are only interested in the posterior labels given the observed image instead of full joint distribution over all pixel values and its corresponding labels. Parameter estimation for the full joint generative model based on MRFs is very hard due to huge mount of parameters. So modeling only the conditional probability of the labels, namely conditional random fields, will significantly boost the speed. Actually, CRF models labels as random variables that forms Markov random field when

conditioned upon a global observation. So MRF is an generative model but CRF is a discriminative model.

3.1 Modeling

Concretely, in the fully connected pairwise CRF model, we will define the energy function for label assignment of each pixel here as follows,

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi(x_i, x_j) \tag{1}$$

where ψ_u is the unary energy component and ψ_p represents the neighborhood pairwise energy component,

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{M} w^{(m)} k_G^{(m)}(f_i, f_j) \tag{2}$$

Minimizing the CRF energy $E(x)$ above yields the most probable label assignment x for the given image.

3.2 Inference

In this section, we introduce several algorithms for solving the inference problem given a conditional random field model with known parameters. These are all iterative models, and one can either use unary potential or use other prior knowledge to initialize the inference.

3.2.1 Iterated Conditional Modes (ICM)

This is simply a greedy algorithm, but when unary clique potential dominates the potential, ICM could reach global optimum. It optimizes the potential by iteratively maximizing the conditional probability of $P(onevariable|othervariables)$.

3.2.2 Simulated Annealing

Simulated Annealing is a classical technique for optimization. It simulates the process of annealing to find global minimals or maximals from local minimals or maximals.

Algorithm 1 Simulated Annealing algorithm

1: **procedure** INITIALIZATION
2: Initialize the state s to s_0
3: **procedure** LOOP FOR $n = N$ TIMES
4: $T \leftarrow temparature(\frac{T}{T_{max}})$
5: Get a new state s_{new} from temperature T
6: If $f(s_{new}) > f(s_{old})$, $s \leftarrow s_{new}$
7: **procedure** OUTPUT
8: Output the state $s_{n=N}$

The random process of getting temperature could help avoiding stuck of the local minimal or maximal, and start finding the new and global extrema. Also, the temperature would be lower with the time increasing, and the state would be more and more stable.

3.2.3 Belief Propagation

Belief propagation, also known as "sum-product message passing", provides exact solution when there are no loop in the graph. Otherwise, belief propagation provides approximate (but often good) solution [7].

In BP, the estimated marginal probabilities are called beliefs. It updates message until convergence, and then calculates the beliefs. Here, we illustrate the message update of pairwise MRF in Equation 1 as the following schematic diagram:

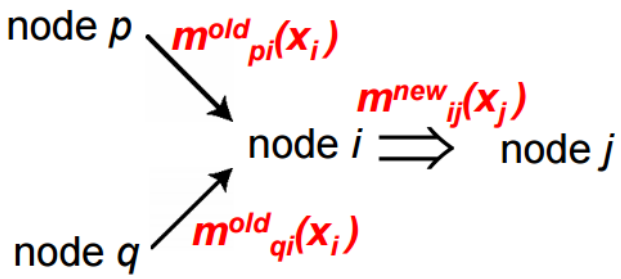


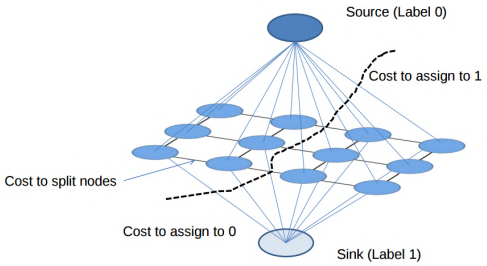
Figure 1: Illustration of message passing step in BP.

where $m_{ij}(x_j)$ represents message from node i to node j. And our belief could be represented as

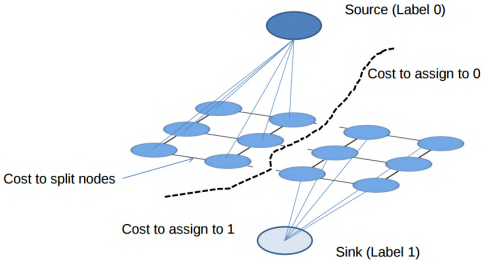
$$b_i(x_i) \propto g_i(x_i) \prod_{k \in Nbd(i)} m_{ki}x(i) \tag{3}$$

when the message converges.

3.2.4 Graph Cuts



Graph Cut Initial State.



Graph Cut Ending State.

[10] shows that, if the pairwise potentials of a **two-label** pairwise MRF could be defined as an Ising model, then we could get an accurate MAP solution by solving a mincut problem. By this theorem, the total energy $C(x)$ could be written as

$$C(x) = \sum_{i=1}^{n} x_i max(0, 1-\lambda_i) + \sum_{i=1}^{n}(1-x_i)max(0,\lambda_i) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}(x_i - x_j)^2 \tag{4}$$

where $x = (x_1, ..., x_n)$ means a binary image and capacity $c_{si} = \lambda_i$.

3.2.5 Mean-Field Algorithm

Mean-field algorithm is used to approximate the maximum a posterior marginal inference. It uses Gibbs distribution to initialize, and performs loop to minimize the CRF total energy Q. Message passing step computes the weighted Gaussian, and the following two steps converts it to the binary term of equation 2. The algorithm is summarized in 2. In [28], mean-field algorithm is proved to be able to be fitted as layers of recurrent neural network.

Algorithm 2 Mean-field inference algorithm [28]

1: **procedure** INITIALIZATION
2: $Q_i(l) \leftarrow \frac{1}{Z_i} exp(U_i(l))$ for all i
3: **procedure** LOOP UNTIL CONVERGE
4: $Q_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(f_i, f_j) Q_j^{(l)}$ for all m (Message Passing)
5: $Q_i^{(l)} \leftarrow \sum_m w^{(m)} Q_i(l)$(Weighting Filter Outputs)
6: $Q_i^{(l)} \leftarrow \sum_{l' \in \mathbf{L}} \mu(l, l') Q_i(l')$ (Compatibility Transform)
7: $Q_i^{(l)} \leftarrow U_i(l) - Q_i(l)$ (Adding Unary Potentials)
8: $Q_i^{(m)} \leftarrow \frac{1}{Z_i} exp(Q_i(l))$ for all m (Normalizing)

3.3 Parameter Estimation

Given the model, we need to find out the appropriate parameters for future inference. But the learning process of CRFs is very complicated. Given a $N \times N$ pixel picture, we can get a N^2-D space. How do we model the internal parameters? Data-driven optimization algorithms are entailed in this learning process. Here we won't cover much of the learning process. Please see [25] for reference of one learning algorithm called Markov Chain Monte Carlo (MCMC) method.

4 Experimental Results

4.1 Introduction of MRF/CRF-based Methods

In section 2 we introduced many image labeling methods. Here we make a detailed introduction to three of them: Hidden MRF and its Expectation-Maximization Algorithm [27], CRF-as-RNN [28] and DeepLab [2].

4.1.1 Hidden MRF and its Expectation-Maximization Algorithm

In [27]. A method for the segmentation of magnetic resonance images is proposed. The main algorithm for this paper is shown below:

According to MAP criterion, we seek the labeling **x*** that satisfy the following rule:

$$x^* = arg \max_x \{P(y|x, \Theta) P(x)\} \tag{5}$$

Therefore, [27] uses EM algorithm to estimate the parameter set $\Theta = \{\theta | l \in L\}$.

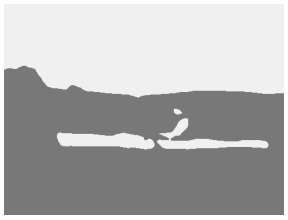
Algorithm 3 Hidden MRF-EM algorithm

1: **procedure** START
2: Initialize parameter set $\Theta \leftarrow \Theta^{(0)}$
3: **procedure** LOOP
4: **E step**: Calculate the conditional expectation $Q(\Theta|\Theta^t)$.
5: **M step**: Maximize $Q(\Theta|\Theta^t)$ and get Θ^{t+1}.

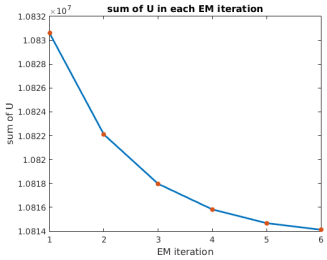
Then an iterative algorithm is used to minimize the total posterior energy:

$$x^* = arg \min_{x \in X} \{U(y|x, \Theta) + U(x)\} \tag{6}$$

Original image.



HMRF-EM algorithm result.



Sum of energy in each EM iteration.

Figure 2: HMRF-EM algorithm results.

4.1.2 CRF as RNN

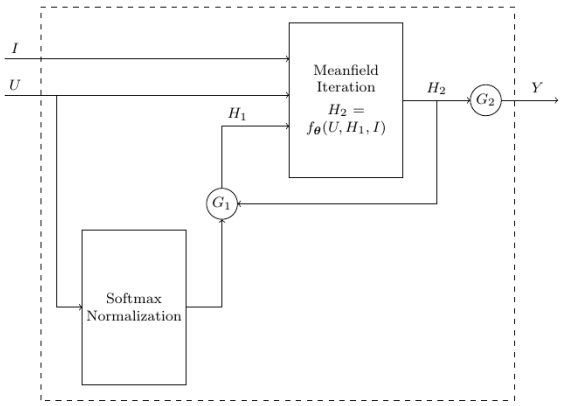


Figure 3: CRF facilitated as a RNN neuron.[28]

In [15], the message passing step is computed by a stack of convolutions. Thus in [28], these computation is then transferred to a stack of layers in the recurrent neural network by decomposing equation 1. These neurons are designed to perform the following steps:

- Initialization
- loop
  - Message Passing
  - Weighting Filter Outputs
  - Compatibility Transform
  - Adding Unary Potentials
  - Normalizing

These steps composes to the following equation:

$$U(X, I) = \sum_i \psi_u(x_i, I_i) + \sum_{i<j} \mu_p(x_i, x_j) \sum_{m=1}^{M} w^{(m)} k_G^{(m)}(f_i, f_j) \tag{7}$$

Denote the unary energy as U, parameters $\{w^{(m)}, \mu(l, l')\}$ as $\{\theta\}$, then the RNN cell follows these equations could be denoted as

$$H_1(t) = \begin{cases} softmax(U), t = 0 \\ H_2(t-1), 0 < t \leq T \end{cases} \tag{8}$$

$$f_\theta(U, H_1(t), I), 0 \leq t < T, \tag{9}$$

$$Y(t) = \begin{cases} 0, 0 \leq t < T \\ H_2(t), t = T \end{cases} \tag{10}$$

4.1.3 DeepLab

Proposed in 2015, DeepLab [2] should be the most famous CNN-based image labeling algorithm that incorporates conditional random fields. As a complementary component of CNN labeling results, this algorithm chose fully connected CRFs to enhance the segmentation result. The basic work flow of this algorithm is as following. We are not going to discuss the implementation details of CNN, but only focusing on the CRFs part.

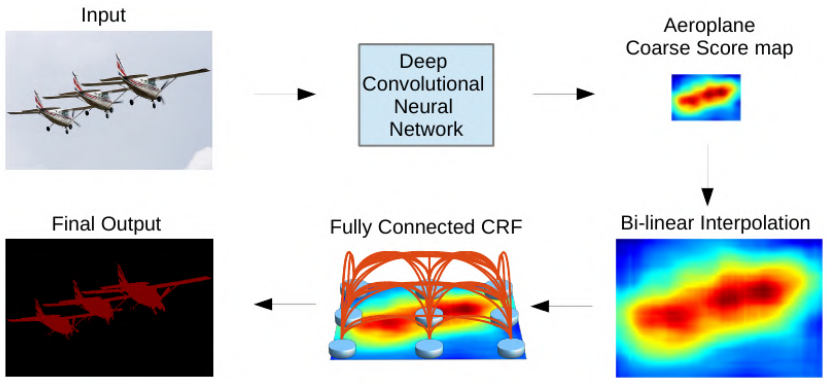


Figure 4: The basic work flow of DeepLab algorithm.[2]

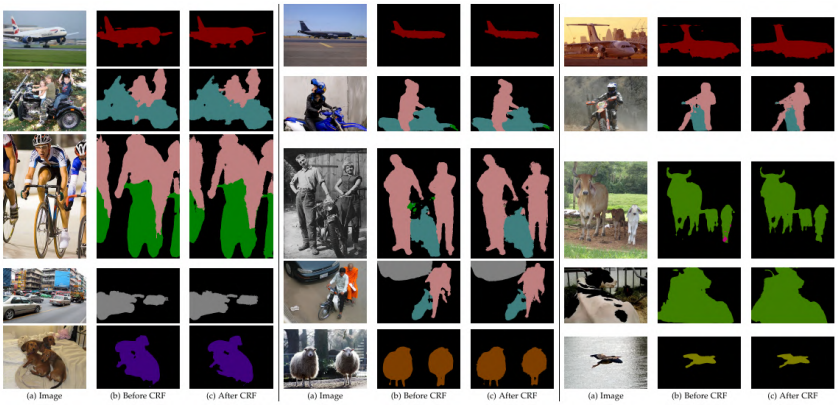


Figure 5: CNN outputs before and after CRF.

4.2 Evaluations

In this section, we evaluate the different usage of Markov random field and conditional random field by several experiments.

4.2.1 MRF as Edge Detection

Comparing to image labeling, edge detection is an easier task. In the following experiment, we show the importance of MRFs in edge detection.

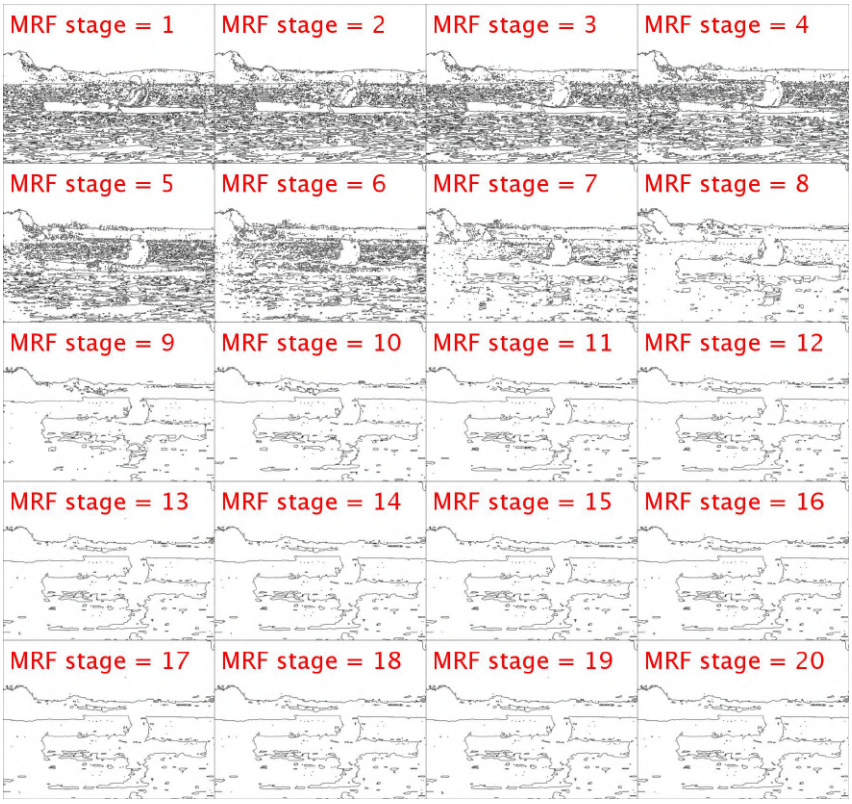


Figure 6: MRF iterations could improve the accuracy of edge detection.

This is an illustration of how MRF could improve the accuracy of edge detection by iterations. In the first row, we could see that the waves on the lake are separated into different small pieces, which is not what we wan. After 10+ iterations, noise are reduced and "semantic" edges are preserved. It is easy to see that MRF could make the edge a more semantically coherent.

4.2.2 DenseCRF for improving unsupervised segmentation

As we all know, K-Means is one of the classical (and best-known) image labeling solutions. However, when given complex images, k-means clustering could not reach an accurate synthetic segmentation result. We compare the segmentations results from k-means with the result after CRFs post processing. Result is summarized in Figure 7.

K-means initialization. 5 iterations. 10 iterations. 100 iterations.

Figure 7: Result of K-Means segmentation and result after CRFs post processing.

In the first figure, original K-Means segmentation ($K = 5$) result is shown. Different color means different categories. You can see that the horse is seperated into different classes, which is not acceptable. But after several iterations of CRF post-processing, the several classes on horse is combined together and formed a whole segmentation.

Also notice that, the front foot of horse is discarded by K-Means, but reconstructed by CRF method. (Best viewed in PDF)

4.2.3 DenseCRF for improving supervised segmentation algorithms

We use [28] to show that CRF method could significantly improve the result of supervised learning. We get the medium result of CRF-as-RNN network and pass it to a CRF algorithm. Clearly, DenseCRF could tighten the image boundary and shrink the wrong segmentation to a proper position. (Notice how hair segmentation was shrinking!)

Original image.

Medium result of [28] before RNN layer.

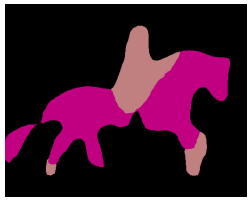
With DenseCRF [15].

4.2.4 Comparison of MRF and CRF methods

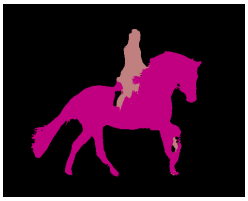
Comparing to MRFs, CRFs models the conditional distribution instead of joint distribution, since it is very difficult and unnecessary to get the full joint distribution over P(Labels, Image). CRFs reduce the time complexity from $O(n^k)$ to $O(n)$.

Original image

MRF



CRFasRNN [2].



DeepLab(CRF) [28].

Figure 8: Comparison between CRF and MRF methods

In Microsoft COCO Image Segmentation Challenge [18], both [28] and [2] have achieved good rankings. Here we use these two methods as a representation of CRF image labeling methods.

In Figure 8, we showed a comparison of how CRFs Methods outperforms MRFs in image labeling tasks. Result from CRFs are much more complete and smoother for semantic information. Also, it is far better than MRF in terms of time complexity.

5 Conclusion

In this paper, we review the basic concept of Markov random field and conditional random field. Also, we inspected some details in the CRFs optimization. And comparative and ablative analysis is conducted to prove the benefit of conditional random field.

In conclusion, MRFs and CRFs could significantly enhance the accuracy of image labeling in that it could make the edges smoother and more coherent to semantic objects. CRFs performs better because it models the conditional distribution instead of joint distribution. It would also reduce the time complexity for computing and make the inference more robust.

References

[1] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *ICCV*, 2001. 2

[2] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In *arXiv preprint*, 2016. 1, 2, 5, 7, 9

[3] L.-C. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille. Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform. In *CVPR*, 2015. 2

[4] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille. Attention to Scale: Scale-aware Semantic Image Segmentation. In *Cvpr*, 2016. 2

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. In *PAMI*, 2002. 2

[6] M. Coombes, W. Eaton, and W. Chen. Colour based semantic image segmentation and classification for unmanned ground operations. In *ICUAS*, 2016. 2

[7] J. Coughlan. A tutorial introduction to belief propagation. In *CRV*, 2009. 3

[8] J. Dong, Q. Chen, S. Yan, and A. Yuille. Towards Unified Object Detection and Semantic Segmentation. In *ICCV*, 2014. 2

[9] D. Eigen and R. Fergus. Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. In *ICCV*, 2014. 2

[10] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. In *Journal of the Royal Statistical Society. Series B (Methodological)*, 1989. 4

[11] S. Gupta, R. Girshick, P. Arbel??ez, and J. Malik. Learning rich features from RGB-D images for object detection and segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014. 2

[12] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Simultaneous Detection and Segmentation. *arXiv*, 2014. 2

[13] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In *CVPR*, 2015. 2

[14] X. He, R. Zemel, and M. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *CVPR*, 2004. 1, 2

[15] P. Krähenbühl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *NIPS*, 2011. 1, 6, 9

[16] J. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001. 2

[17] Z. Li, X. M. Wu, and S. F. Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *CVPR*, 2012. 2

[18] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 9

[19] S. Liu, X. Qi, J. Shi, H. Zhang, and J. Jia. Multi-scale Patch Aggregation for Simultaneous Detection and Segmentation. In *CVPR*, 2016. 2

[20] W. Liu, A. Rabinovich, and A. Berg. ParseNet: Looking Wider to See Better. *arXiv preprint*, 2015. 2

[21] A. Rabinovich and S. Belongie. Objects in Context. In *ICCV*, 2007. 2

[22] B. Sumengen and B. S. Manjunath. Multi-scale edge detection and image segmentation. In *EUSIPCO*, 2005. 2

[23] C. Sutton, K. Rohanimanesh, and A. McCallum. Dynamic Conditional Random Fields : Factorized Probabilistic Models for Labeling and Segmenting Sequence Data. In *ICML*, 2004. 2

[24] S. Tatiraju and A. Mehta. Image segmentation using k-means clustering, em and normalized cuts. Technical report, University Of California Irvine, 2008. 2

[25] L. Wang, J. Liu, and S. Z. Li. Mrf parameter estimation by mcmc method. *Pattern recognition*, 33(11):1919–1925, 2000. 5

[26] X. Wang and X.-P. Zhang. a New Localized Superpixel Markov Random Field for Image Segmentation. In *ICME*, 2009. 2

[27] Y. Zhang, , M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. In *IEEE transactions on medical imaging*, 2001. 2, 5

[28] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. Conditional random fields as recurrent neural networks. In *CVPR*, 2015. 1, 2, 4, 5, 6, 9