Increased Discrimination in Level Set Methods with Embedded Conditional Random Fields

Dana Cobzas
Department of Computing Science
University of Alberta

Mark Schmidt
Department of Computer Science
University of British Columbia

Abstract

We propose a novel approach for improving level set segmentation methods by embedding the potential functions from a discriminatively trained conditional random field (CRF) into a level set energy function. The CRF terms can be efficiently estimated and lead to both discriminative local potentials and edge regularizers that take into account interactions among the labels. Unlike discrete CRFs, the use of a continuous level set framework allows the natural use of flexible continuous regularizers such as shape priors. We show promising experimental results for the method on two difficult medical image segmentation tasks.

1. Introduction

We consider the task of image segmentation, a fundamental problem in the field of computer vision and medical image analysis. The goal of image segmentation is to partition the image into meaningful, consistent regions. For example, separating different objects from the background, or differentiating tumors from normal tissue in medical images.

There has recently been substantial advances in weakly supervised approaches to image segmentation, owing to models that take advantage of graph-cut algorithms [18] and variational level set methods [6, 7]. In many tasks, these methods can produce surprisingly good results with only a small amount of human guidance, typically by iterating between segmentation and re-estimation of model parameters.

In this paper we consider the problem of using supervised segmentation to build an automatic segmentation tool, where labeled training images are used to guide estimation towards parameters that reproduce the semantically meaningful labels in the training data. Discriminative models that address this problem, such as Vapnik's support vector machines (SVMs) [4] and boosting algorithms [9], have had a significant impact in recent vision work and form key components in many state of the art vision systems (see [8], for example). However, these methods typically assume that the labels of different pixels in an image are statistically independent, and this assumption is clearly violated in image data; neighboring pixels are likely to receive the same value. This drawback has motivated the exploration of structured discriminative models like conditional random fields (CRFs) [16, 15]. Unlike discriminative classifiers like SVMs that make independent decisions at each pixel location, CRFs model the joint distribution of discrete label configurations; CRFs represent the affinities between neighboring pixel's labels. However, for tractability, CRFs must define a set of Markov independences with respect to an underlying undirected graph. These required Markov independences make it difficult to model the complex label interactions present in image data, such as the shapes of labeled regions.

In contrast, continuous level set segmentation methods can model non-local dependencies; it is straightforward and computationally tractable to incorporate priors on the shapes of labeled regions [7]. Despite this added modeling power, level set methods often underperform state-of-the-art discriminative models such as SVMs that make much simpler assumptions about the labels.

We believe this is a result of the parameter estimation methods typically used by level set methods, and the use of image-based regularization. In particular, level set methods typically train a generative model that assumes independence among the labels of neighboring pixels. Further, the regularization functionals between neighboring pixels are often based on image gradients, but smoothness in the image gradient may not necessarily correspond to smoothness in the labels.

To alleviate these problems, we propose to embed CRF potential functions within a level set framework. This leads to a simple model that has several appealing properties:

- Correlations in the labels of neighboring pixels are taken into account during parameter estimation.
- The edge regularization term uses features to estimate a potential function that is related to the image labels.

- It does not require a generative model that accounts for the observed image (nor its features).
- It allows the incorporation of non-local continuous regularizers.

The next two Sections review level set methods and CRFs. We then show to how to embed an "associative" CRF within a level set framework in §4, and present experimental results in §5.

2. Level Set Segmentation

Level set methods consider segmentation in the continuous space of the image plane $\Omega \subset \Re^2$. The segmentation of an image $X : \Omega \rightarrow \Re$ is regarded as an infinite-dimensional optimization and is achieved by using variational methods to evolve a contour in the direction of the negative gradient energy. The implicit contour representation is the zero level set of an embedding function $\Phi : \Omega \rightarrow \Re$

$$C = \{x \in \Omega | \Phi(x) = 0\}, \tag{1}$$

where Φ is typically chosen as a signed distance function that has negative values inside the contour and positive values outside [7].

We consider an extension of the Chan-Vese active region method [6]. It partitions an image into two disjoint regions Ω_1 and Ω_2, based on their appearance. Similar to texture segmentation approaches [7, 17], appearance is characterized using a statistical model defined on a set of image features $\mathbf{f}(X)$ (e.g. image gradient, filter responses, etc.). The energy functional in the Chan-Vese model can be written as

$$\begin{aligned} E(\Phi, \mathbf{f}(X), \mathbf{w}) = \int_{\Omega} & -H(\Phi) \log p_1(\mathbf{f}(x), \mathbf{w}) \\ & - (1 - H(\Phi)) \log p_2(\mathbf{f}(x), \mathbf{w}) \\ & + v|\nabla H(\Phi)| g(X, \alpha) dx, \end{aligned} \tag{2}$$

where the *data energy* $\log p_i$ is the probability mass function (with parameters $\mathbf{w}$) of region Ω_i defined over features $\mathbf{f}(x)$ associated with each pixel location. The *regularization energy* $g(X, \alpha)$ is a non-decreasing function with parameter α that has low values at image edges to penalize less when the contour coincides with an image edge, such as $g(X, \alpha) = \frac{1}{1+\alpha|\nabla X|}$. This anisotropic curve length regularization can be interpreted as the length of the contour in a Riemannian space with a metric induced by the image intensity (through $g(X, \alpha)$) [17]. The parameter v controls the strength of the regularization energy relative to the data energy. Finally, $H(\Phi)$ denotes the regularized Heaviside function

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Minimizing the energy from Equation 2 is equivalent to solving the following Euler Lagrange equation (that defines the contour evolution):

$$\begin{aligned} \frac{\partial \Phi}{\partial t} = \delta(\Phi)(\quad & \log p_2(\mathbf{f}(x), \mathbf{w}) - \log p_1(\mathbf{f}(x), \mathbf{w}) \\ & + v \text{ div} \left(g(X, \alpha) \frac{\nabla \Phi}{|\nabla \Phi|} \right)), \end{aligned} \tag{3}$$

where $\delta(\Phi)$ is a regularized Dirac delta function.

A convenient and widely used strategy for estimating the parameters $\mathbf{w}$ is to build a generative model for each class, by holding the regions Ω_1 and Ω_2 fixed and assuming that the labels of neighboring pixels are independent [7]. However, this assumption is clearly violated in most image segmentation tasks. A further problem with this model is that the edge regularization is based exclusively on image gradients (depending on the scalar parameters v and α, that are typically manually tuned). In image segmentation problems where object boundaries are not clearly defined by image edges, the discontinuities in the labels do not directly correspond to image discontinuities, and this regularization may not be appropriate.

3. Conditional Random Fields

CRFs are probabilistic models for segmenting data with structured labels [16]. CRFs directly model the distribution $P(Y|X, \mathbf{w}, \mathbf{v})$ of the labels Y (indexed as y_i for element i) conditional on fixed features $\mathbf{f}(X)$ of the image data X (for node parameters $\mathbf{w}$ and edge parameters $\mathbf{v}$). In this work, we consider binary CRFs with weighted Ising potentials. In this model, the conditional probability is written as:

$$\begin{aligned} & P(Y|X, \mathbf{w}, \mathbf{v}) = \\ & \frac{1}{Z} \exp \left(\textstyle\sum_{i \in N} y_i \mathbf{w}^T \mathbf{f}_i(X) + \sum_{i,j \in E} y_{ij} \mathbf{v}^T \mathbf{f}_{ij}(X) \right) = \\ & \frac{1}{Z} \exp \left(\textstyle\sum_{i \in N} A_i + \sum_{i,j \in E} I_{ij} \right), \end{aligned} \tag{4}$$

where $\mathbf{f}_i(X)$ denotes features for node i, $f_{ij} = F((\mathbf{f}_i(X), \mathbf{f}_j(X))$ denote features for the edge between i and j, the binary node labels are $y_i = \pm 1$, and the binary edge labels are $y_{ij} = y_i y_j$. We have used N to denote the set of pixels, E to denote the set of edges between adjacent pixels, A_i to denote $y_i \mathbf{w}^T \mathbf{f}_i(X)$, and I_{ij} to denote $y_{ij} \mathbf{v}^T \mathbf{f}_{ij}(X)$.

The term A_i, sometimes referred as the *association* function, is a local discriminative model that outputs the association between a label y_i and the data X for site i; this term is analogous to the *data energy* in the active region model. The second term I_{ij}, referred as the *interaction* function, can be seen as a data dependent smoothing term that discriminates between discontinuity and uniformity in the labels for an edge (i, j); this term is analogous to the *regularization energy* in the active region model. The normalizing constant Z sums the numerator over all possible assignments of Y.

Equation 4 is jointly log-concave in $\mathbf{w}$ and $\mathbf{v}$. We can thus find the global maximum of (4) in terms of $\mathbf{w}$ and $\mathbf{v}$ using efficient iterative optimization algorithms. Unfortunately, computing the gradient of $\log(Z)$ requires inference (ie. computing marginal probabilities such as $P(y_i|X,\mathbf{w},\mathbf{v})$ and $P(y_i,y_j|X,\mathbf{w},\mathbf{v})$), and this is intractable. Although other approximations like loopy belief propagation and Monte Carlo methods are also applicable, in this work we used a conditional variant of Besag's pseudo-likelihood [1]:

$$P(y_i|X,y_{-i},\mathbf{w},\mathbf{v}) = \frac{1}{Z_i}\exp\left(y_i\mathbf{w}^T\mathbf{f}_i(X)+\sum_{j|(i,j)\in E} y_{ij}\mathbf{v}^T\mathbf{f}_{ij}(X)\right), \quad (5)$$

where y_{-i} is the labels of all nodes except i and is considered fixed in each of the conditionals. The local normalizing constant Z_i only sums over assignments to y_i, making this expression efficient to compute. This approximation remains log-concave in $\mathbf{w}$ and $\mathbf{v}$, and it is asymptotically consistent (it converges to the true parameters as the size of the training set increases).

Formulating parameter estimation as a joint optimization of $\mathbf{w}$ and $\mathbf{v}$ has several key advantages. First, the parameter estimation now takes into account that the pixel's labels are not independent, leading to a model where the node and edge parameters are appropriately scaled. Second, although the interaction term in the CRF serves a similar purpose to the regularization energy in the active region method, the parameters of the regularization are now learned from training data and therefore seek to define discontinuities directly connected to label discontinuities, rather than to image edges. This is an important difference in applications where the object boundary is not very well defined by image edges.

4. Active Regions with an Embedded CRF

This Section discusses embedding the potential functions from a trained CRF model into a level set segmentation framework, allowing the level set method to exploit CRF training methods. The first part of this Section discusses two minor modifications of the CRF model that are needed in order to embed it in the level set framework. We then present the new method, and finally discuss the incorporation of shape priors.

4.1. Associative CRF

To embed the node and edge potential functions learned by the CRF model into a level set segmentation framework, we require two modifications of the CRF model: (i) We must make the CRF labels consistent with the continuous labels given by the Heaviside function, and (ii) the interaction terms I_{ij} must be non-negative since they correspond to a curve length. We refer to a CRF with the second property as an associative CRF, in analogy with associative max-margin Markov networks [21].

We first consider changing the CRF labels from $y_i = \pm 1$ to correspond with the Heaviside labels that are in $\{0,1\}$. By multiplying $P(Y|X,\mathbf{w},\mathbf{v})$ by $\exp(\mathbf{w}^T\mathbf{f}_i(X))/\exp(\mathbf{w}^T\mathbf{f}_i(X))$ for each node i in N, we can write the association function concisely as $A_i = 2y_i\mathbf{w}^T\mathbf{f}_i(X)$ with y in $\{0,1\}$. By similarly multiplying $P(Y|X,\mathbf{w},\mathbf{v})$ by $\exp(\mathbf{v}^T\mathbf{f}_{ij}(X))/\exp(\mathbf{v}^T\mathbf{f}_{ij}(X))$ for each edge $\{i,j\}$ in E, the interaction function can be written in the $y_i \in \{0,1\}$ representation as $I_{ij} = 2(1-|y_i - y_j|)\mathbf{v}^T\mathbf{f}_{ij}(X)$.

We now consider enforcing the needed condition that the interaction function I_{ij} is always non-negative. In the $\{0,1\}$ representation, it is clear that I_{ij} will be non-negative as long as both $\mathbf{f}_{ij}(X)$ and $\mathbf{v}$ are non-negative. To ensure that $\mathbf{f}_{ij}(X) \geq 0$, we choose non-negative edge features that represent a measure of association between the two pixels involved in an edge. In particular, we used edge features of the form (for each node feature k):

$$f_{ijk}(X) \triangleq \frac{1}{1+|f_{ik}(X)-f_{jk}(X)|} \quad (6)$$

giving $F(\mathbf{f}_i(X),\mathbf{f}_j(X)) = \{f_{ijk}(X), k = 1\ldots K\}$. Since $\mathbf{v}$ is non-negative, these edge features have the intuitive effect that they encourage the most smoothing at locations where the features are most similar.

Ensuring that $\mathbf{v}$ is non-negative can be done by changing the optimization procedure. In particular, rather than optimizing the unconstrained pseudo-likelihood, we maximize the pseudo-likelihood subject to $\mathbf{v} \geq 0$. The global optimum under this constraint can still be computed efficiently even with a very large number of variables by using bound-constrained limited-memory quasi-Newton methods such as L-BFGS-B [5]. An interesting consequence of the form of the interaction function I_{ij} and the convex polytope that results from these constraints is that the model performs 'associative feature selection'. That is, it will set v_k to exactly 0 (effectively ignoring the feature) for any edge feature k that does not have an associative (smoothing) effect.

By putting everything together and dividing out the constant factor two from every term, the final formulation for the associative CRF with y_i in $\{0,1\}$ is the following:

$$p(Y|X) = \frac{1}{Z}\exp\left(\sum_i y_i\mathbf{w}^T\mathbf{f}_i + \sum_{ij}(1-|y_i-y_j|)\sum_k v_k f_{ijk}\right) \quad \text{subject to } \mathbf{v} \geq 0 \quad (7)$$

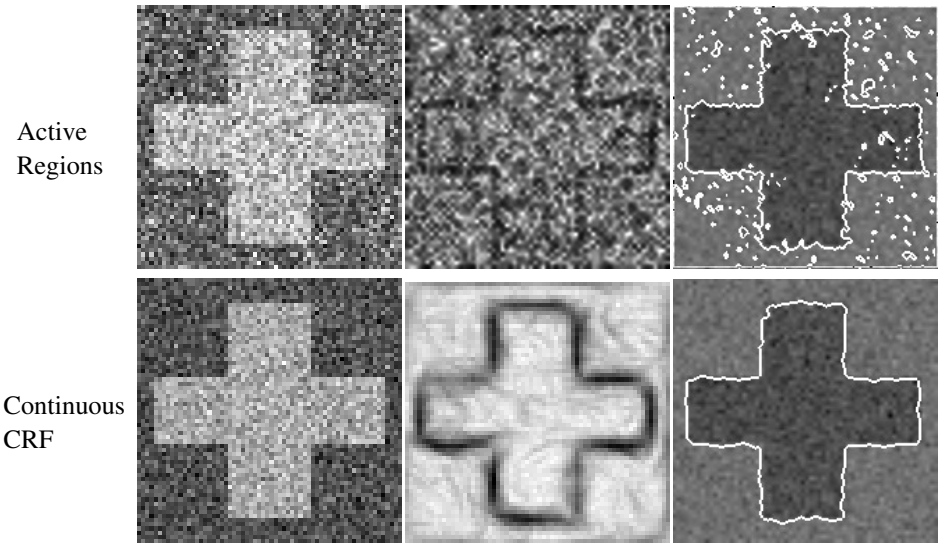


Figure 1. Example of corresponding data and regularization energies in the continuous CRF model (bottom) and the traditional active region model with Gaussian region statistics and the the usual edge based regularization $g = \frac{1}{1+\alpha|\nabla X|}$ (top). The task is to denoise a simple binary image corrupted with independent Gaussian noise. Notice how the CRF regularization energy picks the correct label discontinuities (bottom middle) as compared to the usual edge detection function (top middle).

	CRF	cont. CRF
node labels	y_i	$H(\Phi(x))$
edge labels	$1 - \|y_i - y_j\|$	$1 - \|\nabla H(\Phi(x))\|$
node features	$\mathbf{f}_i(X)$	$\mathbf{f}(x)$
edge features	$f_{ij} = F(\mathbf{f}_i(X), \mathbf{f}_j(X))$	$F(\nabla \mathbf{f}(x))$

Table 1. Corresponding terms between CRF and continuous CRF

4.2. Continuous Domain CRF

The proposed associative CRF can now be fully embedded into a continuous model that has the same energy:

$$E(\Phi) = \int_\Omega \begin{array}{l} -H(\Phi)(\mathbf{w}^T\mathbf{f}) + (1 - H(\Phi))(\mathbf{w}^T\mathbf{f}) \\ +|\nabla H(\Phi)| \sum_k v_k \frac{1}{1+|\nabla f_k|} dx \end{array} \quad (8)$$

Table 1 presents the corresponding entities between the two models, with x corresponding to the location i in the discrete representation.

The CRF *association term* models the class association potential and differentiates the two classes object(inside)/background(outside). This term corresponds to the level set *data energy* $\log p_1$, $\log p_2$ of a point being inside or outside the contour.

The *interaction term* corresponds to the *regularization energy* in the continuous formulation. The discrete CRF interaction is defined on pairs of neighboring locations and on the corresponding values for the data. The corresponding regularization in the continuous model is defined on $\nabla H(\Phi)$ and it depends on pairwise features (defined using the gradient of the features ∇f). In two-dimensions, we work with a 2 neighborhood system for the finite difference approximation, using differences between a pixel's northern and eastern neighbors in the image. Figure 1 illustrates the difference between the regular, edge based anisotropic regularization used in level set segmentation methods, and the proposed regularization based on the CRF interaction potential that encourages discontinuities at label discontinuities rather than image edges (as discussed in Section 3).

The Euler-Lagrange evolution equation corresponding to the continuous CRF energy from Equation 8 is:

$$\begin{array}{rl} \frac{\partial \Phi}{\partial t} & = -2\delta(\Phi)\mathbf{w}^T\mathbf{f} \\ & +\delta(\Phi)\text{div}\left(\left(\sum_k v_k \frac{1}{1+|\nabla f_k|}\right)\frac{\nabla\Phi}{|\nabla\Phi|}\right) \end{array} \quad (9)$$

An outline of the continuous CRF segmentation algorithm is:

Training:

Given: a set of images $X_1, X_2, \ldots, X_n$
Extract features $\mathbf{f}(X_1), \mathbf{f}(X_2), \ldots, \mathbf{f}(X_n)$
Compute optimal node and edge params $\mathbf{v}, \mathbf{w}$
by maximizing the constrained
pseudo-likelihood of the CRF (Equation 7)

Segmentation:

Given: one image X
Extract features $\mathbf{f}(X)$
Compute segmentation by evolving a curve
driven by Equation 9

4.3. Shape priors

As will see in Section 5.2 for the task of muscle segmentation in CT abdominal images, in some cases we can

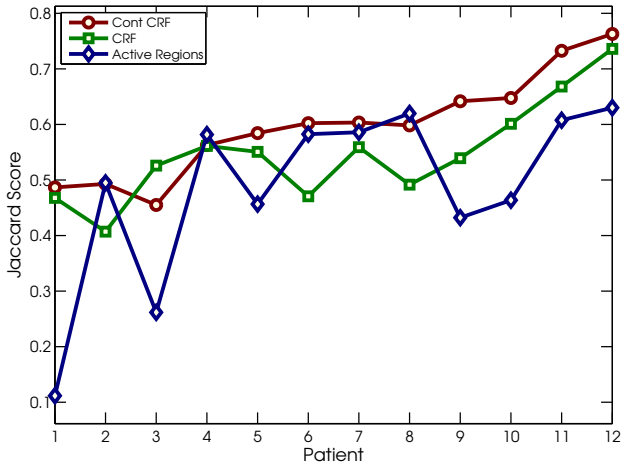


Figure 2. Test image Jaccard scores ($A \cap B/A \cup B$) for brain tumor segmentation in MR images with three different methods for 12 different patients. The patients are sorted by the maximum score achieved across the methods.

substantially improve performance by adding a shape prior to the model. Incorporating shape information in a discrete CRF model is not trivial, as either the graph structure has to be changed (to include factors on regions of nodes and suffer the corresponding computational expense), or shape information must be incorporated into the local potentials (see [14], for example).

As we show here, one of the advantages of embedding the CRF within a continuous model is that we can incorporate terms like continuous shape priors into the energy. In particular, we used a shape energy of the following form [20]

$$E_s(\Phi) = \int_\Omega \beta H(\Phi) \left(s\Phi - \Phi_s(\mathcal{A}(x))\right)^2 dx, \qquad (10)$$

where $\mathcal{A}(x)$ represents an affine transformation with scale s of the shape prior level set Φ_s, and β is the strength of the regularizer.

The shape prior is easily incorporated into the continuous CRF by adding (10) as an additional term in the energy function (8) during the segmentation phase of the algorithm. The scalar parameter β weighing the influence of the shape prior against the CRF energy terms is selected by cross-validation, and for our experiments we chose the shape prior as being the mean of the labels in the training set.

5. Experimental Results

We evaluated the proposed continuous CRF model on two difficult medical imaging problems: (1) brain tumor segmentation in MRI images (2) muscle segmentation in CT abdominal images. Our results are compared with a traditional level set active region [19] and a discrete CRF. Dataset (1) shows the superiority of the proposed conditional model over the generative active region model for data with a complex appearance. Dataset (2) shows the advantage of the continuous model over discrete CRFs when a shape prior is available. In our experiments, we compared the following three models:

Continuous CRF: We first train the discrete CRF (Equation 7) on a set of features extracted from the training set and then evolve the continuous CRF (Equation 8) with fixed parameters on the test images.

Active regions: We learned Parzen histograms for in/out regions from the training set using the same features as the CRF. Segmentation is done using traditional level set evolution [19] with fixed region statistics.

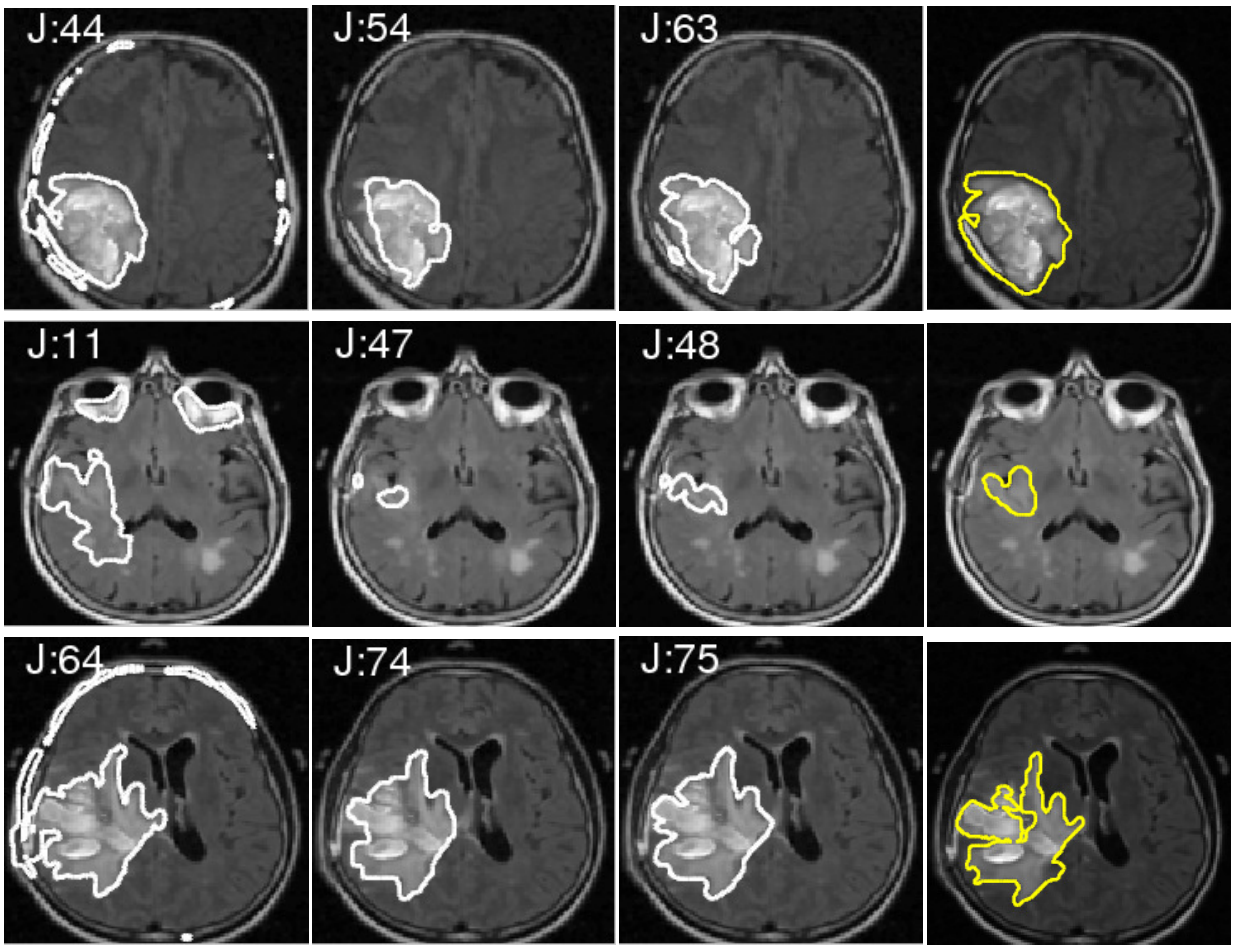
CRF: The CRF was trained in the same way as the continuous CRF (Equation 7). In this model we find the optimal segmentation using a graph cut algorithm [3].

5.1. Brain tumor segmentation

We consider the problem of brain tumor segmentation in 3D MRI images. The data consists of FLAIR and contrast-enhanced T1 MRI images of dimension $128 \times 128 \times 20$, from 12 patients having either a grade 2 astrocytoma, an anaplastic astrocytoma, or a glioblastoma multiforme. The tumor area was manually segmented slice-by-slice in each data set by an expert radiologist.

We used the following image features: the two MRI modalities, corresponding symmetry images (difference between the image and the vertically flipped image) and the normalized distance from the skull. Training was done on two folds (6 training images and 6 testing images), and the level set was initialized with the whole skull (for both the continuous CRF and the active regions). On this data set, we used a three-dimensional lattice graph structure and performed all operations (CRF training, level evolution, graph cut segmentation, and evaluation of performance) in three-dimensions.

The Jaccard scores achieved by the three methods across all patients are summarized in Figure 2. Three example segmentation are shown in Figure 3 (selected slices from the 3D segmentation). Corresponding Jaccard scores are displayed in the top left corner of each figure. The main difficulty in this dataset is the complex and varying appearance of the tumor. Our results show that for some patients the generative active region is less successful in capturing this appearance than the discriminative models. Although both CRF models have the same energy and the discrete CRF finds the global minimum of this energy, the continuous CRF tended to yield better solutions on this data set (obtaining the highest score among the three methods in 8 of the 12 volumes). We suspect that this is because the optimal decoding may be unstable (ie. due to its discrete nature



Active Regions CRF Cont CRF Expert Label

Figure 3. Comparative results for brain tumor segmentation. Selected slices are shown while the segmentation was done in 3D. Jaccard scored are displayed at the top left.

it can change drastically with minor changes in the model). A more robust estimate would be to use the most likely label for each pixel (rather than the jointly most likely assignment of labels), but currently we know of no exact methods for finding the marginally most likely labels for image-sized data (we tried to approximate this using loopy belief propagation, and found that in roughly half the patients it achieved a similar score to the continuous CRF). In contrast to the optimal discrete decoding that can potentially change drastically with small changes to the model, there must be continuity around a (local) minimum found by continuous energy minimization, indicating that the minimum is insensitive to minor perturbations.

5.2. Skeletal muscle segmentation

Our second medical imaging application was skeletal muscle segmentation in 37 CT abdominal scans of 18 cancer patients. Two consecutive 2D axial CT images at the level of L3 were selected for each patient and manually segmented by medical experts. These images have a resolution of 200×150.

We used two features: the original CT image and the normalized distance from the outer body boundary. Training was again done in two folds and the level set was initialized with the body boundary. Both continuous models incorporate the shape prior discussed in Section 4.3, while the CRF does not. The shape prior is shown in Figure 4 (Left).

A selection of results with corresponding Jaccard scores are displayed in Figure 4, while scores across the patients are summarized in Figure 5. The muscle tissue has a very similar appearance to the enclosed internal organs and therefore cannot be segmented entirely based on its appearance signature, as shown by the poor performance of the discrete CRF model. The shape prior used in the continuous models helps disambiguate muscle from organ tissue, while the discriminative model has a better appearance model and

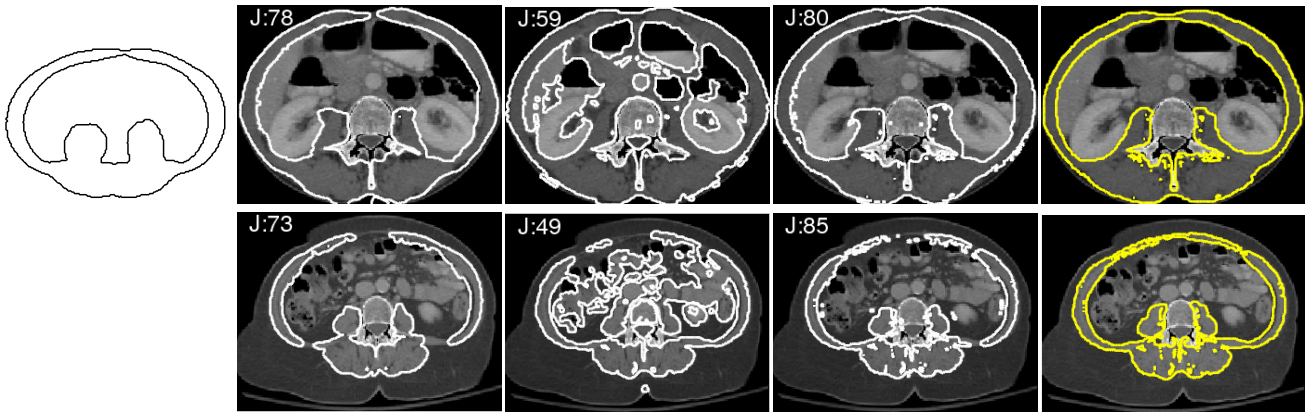


Figure 4. Comparative results for skeletal muscle segmentation. Left: Shape prior used by the continuous models. Jaccard scores are displayed on the top left.

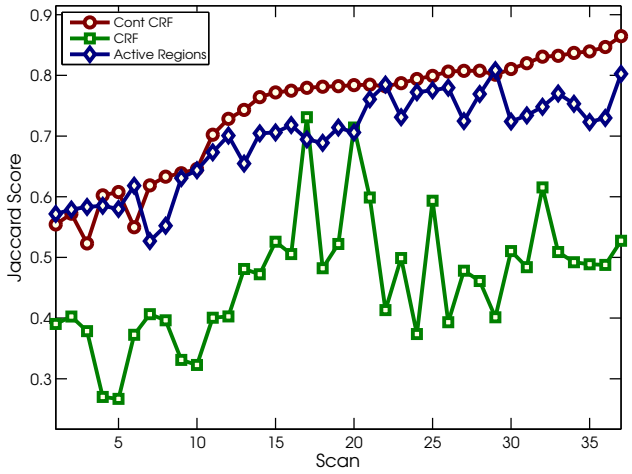


Figure 5. Test image Jaccard scores ($A \cap B / A \cup B$) for muscle segmentation in CT images with three different methods for 37 different scans across 18 patients. The scans are sorted by the maximum score achieved across the methods.

outperforms the generative one. For this data set, the continuous CRF obtained the highest score in 31 of the 37 volumes.

6. Discussion

There has been a previous attempt to couple a CRF with an active contour model [22], extending previous work on integrating probabilistic deformable models with Markov random fields [11]. The main difference between our work and this previous work is subtle but very important; in [22], the model requires an 'image prior' that represents the distribution over images, and due to the complexity of images this is extremely complicated to specify. In [22], they use an image prior that factorizes into independent Gaussians at each pixel (ie. an 'average' image is scaled white noise, and realistic images are extremely unlikely). The model in [11] similarly requires a distribution over images given the labels. The key feature of CRFs is that they condition on the image (treating it as a fixed observation), and do not need an image prior. Although [22] uses a CRF as part of a larger model, the model doesn't take advantage of this key feature because the full model still needs an image prior. This distinction is important when we want to enhance discrimination by using relevant 'features' instead of just pixel values. In [22] you would need a 'feature prior', specifying the distribution over the features, which might be even harder than specifying a realistic distribution over images. In our model, we can use arbitrary features without needing to account for their probability, and we obtain a standard CRF model (at the appropriate discretization level) in the special case where no additional continuous regularizer is incorporated.

We would like to note another important differences between our method and the previous work. We address the issue of joint parameter estimation from training data, formulating it as a convex optimization. The quantitative evaluation in [22] used manual initialization of the contour for each image to be segmented (similar to most previous work on level set methods), while our experiments tested on the arguably much more difficult task of segmenting completely new images with automatic initialization of the contour.

The edge potentials defined by the associative model in Section 4 obey the submodularity constraint $I(0,0) + I(1,1) \geq I(1,0) + I(0,1)$ (since $\mathbf{v}^T \mathbf{f}_{ij}(X) \geq 0$), and therefore the optimal decoding of this model can be formulated as a graph cut problem [13]. Our continuous interpretation of the CRF is therefore connected to the work of Boykov and Kolmogorov on geo-cuts [2, 12], graph cuts

that approximate continuous energies (the work in [10] is also closely related). They solved a related inverse problem and showed how to design a graph such that a cut approximates a Riemannian metric with edge weights $w_k = \frac{\rho^2 g}{2|e_k|}$, where ρ is the cell size and e_k is a vector associated with an edge. For a 2×2 neighborhood system with unit cells the edge weight is $w_k = g/2$. As we are using only 2 edges for each node the edge weight is therefore $w_k = g$. This corresponds to the edge potential in the CRF model that is correctly used as a metric for the continuous regularization.

We have presented a method to improve the robustness of level set models for supervised segmentation. In particular, we have proposed taking advantage of an embedded CRF model whose parameters are learned from an equivalent discrete CRF. The model takes advantage of several appealing aspects of CRFs; it does not require a generative model of the image, it does not assume label independence during training, and it estimates a conditional regularization term that is related to the labels rather than simply the image gradient. In contrast, the model is advantageous over discrete CRFs in cases where an additional continuous regularizer can incorporate non-Markov prior knowledge, which we demonstrated though the use of a shape prior.

There are numerous directions of future exploration. In particular, the use of a trivial shape prior is only a simple example of the additional flexibility afforded by augmenting a CRF model with a continuous regularizer. For example, we are exploring applications in medical imaging where a CRF model is augmented with not only a shape prior but a model of the continuous spatial intensity inhomogeneity field.

Acknowledgements

We would like to thank Dr. Albert Murtha from the Alberta Cross Cancer Institute for help in acquiring and segmenting the MRI brain images, Dr. Vickie Baracos for help in acquiring and segmenting the CT muscle images, and the anonymous reviewers for helpful comments that improved the paper.

References

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *RoyalStat*, B-36(2):192–236, 1974.

[2] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV*, pages 26–33, 2003.

[3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in computer vision. *IEE PAMI*, 26(9):1124–1137, 2004.

[4] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[5] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208, 1995.

[6] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[7] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *IJCV*, 72(2), April 2007.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge (VOC2007).

[9] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *EuroCOLT*, pages 23–37, 1995.

[10] L. Grady and C. Alvino. Reformulating and optimizing the mumford-shah functional on a graph - a faster, lower energy solution. In *ECCV*, pages 248–261, 2008.

[11] R. Huang, V. Pavlovic, and D. N. Metaxas. A graphical model framework for coupling mrfs and deformable models. In *CVPR*, pages 739–746, 2004.

[12] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *ICCV*, pages 564–571, 2005.

[13] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.

[14] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Obj cut. In *CVPR*, pages 18–25, 2005.

[15] S. Kumar and M. Hebert. Discriminative random fields. *Int. J. Comput. Vision*, 68(2):179–201, 2006.

[16] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, pages 282–289, 2001.

[17] N. Paragios and R. Deriche. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Visual Communication and Image Representation*, 13:249–268, 2002.

[18] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[19] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *CVPR*, 2003.

[20] M. Rousson and N. Paragios. Shape priors for level set representations. In *ECCV (2)*, pages 78–92, 2002.

[21] B. Taskar, V. Chatalbashev, and D. Koller. Learning associative markov networks. In *ICML*, page 102, 2004.

[22] G. Tsechpenakis and D. Metaxas. Crf-driven implicit deformable model. In *CVPR*, 2007.