A Principled Deep Random Field Model for Image Segmentation

Pushmeet Kohli
Microsoft Research
Cambridge, UK

Anton Osokin
Moscow State University
Moscow, Russia

Stefanie Jegelka
UC Berkeley
Berkeley, CA, USA

Abstract

We discuss a model for image segmentation that is able to overcome the short-boundary bias observed in standard pairwise random field based approaches. To wit, we show that a random field with multi-layered hidden units can encode boundary preserving higher order potentials such as the ones used in the cooperative cuts model of [11] while still allowing for fast and exact MAP inference. Exact inference allows our model to outperform previous image segmentation methods, and to see the true effect of coupling graph edges. Finally, our model can be easily extended to handle segmentation instances with multiple labels, for which it yields promising results.

1. Introduction

Interactive figure-ground segmentation is an important problem in computer vision and image processing. Given some user input, which typically takes the form of marking some pixels that belong to the figure or background, the system is required to find the set of pixels that belong to the figure. Like many other image labeling problems, interactive segmentation is commonly modelled using pairwise Markov Random fields that incorporate priors on labels of pairs of neighbouring pixels. These models allow efficient inference of the Maximum a Posteriori (MAP) solution using algorithms such graph cuts [4, 28] but have restricted expressive power [12].

One of the major side-effects of using pairwise MRFs for segmentation is the short-boundary bias [20], illustrated in Figure 1(c). It results from the fact that the standard pairwise model encourages smooth segmentations by penalizing the assignment of different labels to neighbouring pixels [1, 4]. This penalty equivalently discourages long object boundaries. Several papers have tried to address this problem, e.g. via topology constraints [21, 30].

Jegelka & Bilmes [11] recently proposed a different approach – a *cooperative graph cut* model – to overcome the short-boundary bias. Instead of favoring short object boundaries by penalizing the *number* of label discontinuities, their model favors "congruous" boundaries by penalizing the number of *types* of label discontinuities. This new term is a submodular function on pairs of variables, a *cooperative cut potential*. The cost of label discontinuities does not grow linearly with the number of discontinuities in the labeling, as is the case in the standard pairwise model. Instead, the penalty is a concave function of the number of discontinuous pairs of the particular type[1]. Cooperative cut potentials can make MAP inference an NP-hard problem, and therefore [11] propose an approximation algorithm that still outperforms previous segmentation methods. Nevertheless, given that the algorithm yields an approximate solution, it is not entirely clear whether the improvement is an artefact of the algorithm or actually induced by a better model. In the sequel, we will confirm the latter.

In this work, we rephrase the model of [11] and construct an equivalent hierarchical model by using a transformation inspired by a lower envelope representation of higher-order potentials [13]. The model introduced here has a number of benefits: (i) It enables us to derive an *exact* yet practical algorithm for MAP inference. In Section 5, we will see that the exact results outperform results of state-of-the-art image segmentation methods. In addition, we propose efficient heuristics that, in our experiments, always return the optimal solution. (ii) It yields a refined complexity analysis and, in the language of parameterized complexity [8, 23], shows that a practically useful subset of cooperative cut problems is *fixed-parameter tractable*. Furthermore, an FPTAS is possible for a wider class of functions. (iii) The model is well extensible to handle segmentation instances with multiple labels, and we show promising theoretical and empirical results. (iv) The model structure shows an explicit connection between cooperative cuts and hierarchical models, hinting at the potential representational power of deep models such as Deep Boltzmann Machines (DBM) [27].

2. Coupling Edges for Image Segmentation

We first introduce the standard pairwise Markov random field (MRF) model for interactive segmentation and the extension proposed in [11]. Let $\mathcal{V}$ be the set of pixels in the image I we want to segment, and let the set $\mathcal{E}$ contain all pairs of neighbouring pixels – these are the

[1] We will describe the model in detail in section 2.1.

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.257

IEEE computer society

(a) (b) (c) (d) (e)

Figure 1: (a) User-specified seeds; (b) Ground truth segmentation; (c) MAP solution of standard pairwise MRF model; (d) Global minimum of cooperative cut energy obtained by our method. The (summed) weight of the boundary is 4849 in (c) and 11273 in (d). 95% of the 1546 edges cut in (c) belong to 3 out of 11 types, while for (d) the 3 types include 99% of the 3932 edges cut. (e) type ID of cut edges: colour channel C (R, G, or B) of the pixel is set to 255 if this pixel is incident to a cut edge whose type is C. The coupling potentials Ψ_g for the three main edge types contribute a total penalty that is 0.35 times the pairwise penalty in (c) and only 0.2 times the pairwise penalty in (d), and hence solutions like (d) are favored.

edges in the MRF. The label of each pixel $i \in \mathcal{V}$ is given by a binary random variable x_i whose labels (0/1) denote the label "figure"/"ground".The set of all variables x_i for $i \in \mathcal{V}$ is denoted by $\mathbf{x}$. The posterior distribution $P(\mathbf{x}|I) = (1/Z)\exp(-E(\mathbf{x}))$ of the pairwise model factorizes into unary potentials $\phi_i(x_i)$ and pairwise potentials $\psi_{ij}(x_i, x_j)$. The corresponding Gibbs energy E is

$$E(\mathbf{x}) = \sum\nolimits_{i\in\mathcal{V}} \phi_i(x_i) + \sum\nolimits_{ij\in\mathcal{E}} \psi_{ij}(x_i, x_j). \quad (1)$$

The functions ϕ_i encode the likelihood of pixel i belonging to figure or ground, while ψ_{ij} is a contrast sensitive prior

$$\psi_{ij}(x_i, x_j) = \theta(I_i, I_j)|x_i - x_j| \quad (2)$$
$$= \theta(I_i, I_j)(x_i + x_j - 2x_i x_j) \quad (3)$$

The function $\theta(I_i, I_j)$ modulates the penalty for assigning different labels to pixels i and j based on their respective appearances (colours) I_i and I_j.

2.1. The Short-Boundary Bias and Coupling Edges

The above pairwise model can be illustrated by a grid-structured graph, where each potential ψ_{ij} corresponds to an edge. A labeling corresponds to a partition of this graph, and label discontinuities between neighboring pixels correspond to cut edges. The potential (1) grows linearly with the weights θ_{ij} of cut edges or neighboring pixels with differing labels. In particular, in the special case of θ being constant, the contribution of the pairwise potentials to the energy is a linear function of the perimeter of the figure. This "short-boundary" prior leads to erroneous segmentations of objects with fine structures, as illustrated in Figure 1.

Although a large number of neighbouring pixels in such images may take different labels, the majority of these pixels has a consistent appearance. For instance, in Figure 1, most pixel pairs along the object boundary have a consistent (brown-white) transition. Therefore, [11] proposed a prior that penalizes not the length but the *diversity* of the object boundary, i.e., the number of *types* of transitions. This new potential does not suffer from the short-boundary bias.

"Diversity" is defined by partitioning the set $\mathcal{E}$ of pixel pairs into groups (types) $\mathcal{E}_g$ ($g \in \mathcal{G}$) of similar pairs, and a congruous boundary uses few types. The new energy considers all pairs of a type simultaneously and gives a "discount" for using edges (pairs) of the same type. The discount results from a monotonically increasing concave function F, and the resulting energy function is

$$E(\mathbf{x}) = \sum_{i\in\mathcal{V}} \phi_i(x_i) + \sum_{g\in\mathcal{G}} \Psi_g(\mathbf{x}), \quad \text{where} \quad (4)$$

$$\Psi_g(\mathbf{x}) = F\Big(\sum\nolimits_{ij\in\mathcal{E}_g} \psi_{ij}(x_i, x_j)\Big), \quad (5)$$

As the values of the pairwise penalty terms are grouped together using a higher-order potential, [11] referred to this model as *coupling edges*, or cooperative cuts. It is easy to see that the above formulation is a generalization of the standard pairwise MRF model. If F is the identity, then the energy in Equation (4) is equal to the standard energy (1).

In principle, unless the edge groups g have a specific structure, the algorithms for MAP inference in the model (4) will be indifferent and apply to any set $\mathcal{G}$ of groups. Edge groups that lead to a bias for congruous boundaries can be computed by clustering graph edges [11].

MAP Inference. Inferring the Maximum a Posteriori (MAP) solution of the models above corresponds to minimizing their respective energy functions. It is well known that the energy function (1) is submodular if all $\theta(I_i, I_j) \geq 0$ and can then be minimized in polynomial time by solving an (s,t)-mincut problem [4]. In contrast, the higher-order potential (5) makes MAP inference in general NP-hard, and therefore [11] proposed an iterative bound minimization algorithm for approximate inference. We show

next that higher order potentials of the form (5) can be converted into a pairwise model by the addition of some binary auxiliary variables.

3. Lower Envelope Representations for Edge Coupling Potentials

One of the key challenges posed by higher-order models is efficient MAP inference. Since inference in pairwise models is very well studied, one popular technique is to transform the higher-order energy function into that of a pairwise random field. In fact, any higher-order pseudo-boolean function can be converted to a pairwise one, by introducing additional auxiliary random variables [2, 10, 13]. Unfortunately, the number of auxiliary variables grows exponentially with the arity of the function, and in practice this approach is only feasible for higher-order functions with few variables. If however the higher-order function contains inherent "structure", then MAP inference can be practically feasible even with terms that act on thousands of variables [14, 12, 25, 29]. This is the case for the edge-coupling potentials (4).

We explain our transformation using the example of a special cooperative cut potential:

$$\Psi_g(\mathbf{x}) = \min\Big\{\sum\nolimits_{ij\in\mathcal{E}_g} \theta_{ij}|x_i - x_j|, T\Big\}, \qquad (6)$$

where T is a truncation parameter. If $\theta_{ij} = T$, then this potential indicates whether any pixel pair corresponding to a group g has been assigned different labels. Incorporated into energy (4), this potential penalizes the number of types of transitions in the boundary.

We apply the lower-envelope representation for higher-order functions proposed by Kohli & Kumar [13] to the function (6). Instead of using lower envelopes of linear (modular) functions as they did, our transformation will employ lower envelopes of general second order functions. A similar application of lower envelopes was used in [9], but for different potentials in a different application.

Theorem 1. *The problem of minimizing the cooperative cut potential (6) is equivalent to minimizing a pairwise function with additional $|\mathcal{E}_g| + 1$ auxiliary variables.*

Proof. Our proof is constructive. First, we rewrite the potential in Equation (6) as a third order pseudo-boolean function by representing it as a lower envelope of the two functions $f_1(\mathbf{x}) = \sum_{ij\in\mathcal{E}_g} \theta_{ij}|x_i - x_j|$ and $f_2(\mathbf{x}) = T$. The transformation relies on a switching variable $h_g \in \{0, 1\}$:

$$\begin{aligned}
\Psi_g(\mathbf{x}) &= \min_{h_g}\Big(\sum_{ij\in\mathcal{E}_g} \theta_{ij}|x_i - x_j|\Big)h_g + T(1 - h_g) \qquad (7)\\
&= \min_{h_g}\ T + \Big(\sum_{ij\in\mathcal{E}_g} \theta_{ij}|x_i - x_j| - T\Big)h_g\\
&= \min_{h_g}\ T + \Big(\sum_{ij\in\mathcal{E}_g} \theta_{ij}(x_i + x_j - 2x_ix_j) - T\Big)h_g\\
&= \min_{h_g}\ T + \sum_{ij\in\mathcal{E}_g} \theta_{ij}(x_ih_g + x_jh_g - 2x_ix_jh_g) - Th_g
\end{aligned}$$

Each third-order term $-x_ix_jh_g$ can be rephrased as a pairwise energy via an auxiliary boolean variable z_{ij}:

$$-x_ix_jh_g = \min_{z_{ij}\in\{0,1\}} -z_{ij}(x_i + x_j + h_g - 2). \qquad (8)$$

Thus, by adding the set of binary variables $\mathbf{z} = \{z_{ij} \mid (i, j) \in \mathcal{E}_g\}$ and the variable h_g, we have rewritten the potential (6) as the quadratic pseudo-boolean function (QPBF)

$$\begin{aligned}
\Psi_g(\mathbf{x}) \ =& T + \min_{h_g,\mathbf{z}}\Big\{\sum_{ij\in\mathcal{E}_g} \theta_{ij}((x_i + x_j - 2z_{ij})h_g \qquad (9)\\
&- 2(x_i + x_j)z_{ij} + 4z_{ij}) - Th_g\Big\}. \qquad \square
\end{aligned}$$

The factor graph induced by the above transformation is illustrated in Figure 2. This factor graph has a multi-layer structure similar to a Deep Boltzmann Machine [27]. In total, we need to add $|G| + \sum_{g\in\mathcal{G}} |\mathcal{E}_g|$ auxiliary variables.

Representation of arbitrary concave functions. Above, we have transformed the truncation potential (6) into a quadratic pseudo-boolean function. This technique applies to any cooperative cut potential of the form (5) for a nondecreasing concave function F: we express (or approximate) F as the lower envelope of multiple linear functions, and this envelope is represented by adding multiple potentials of the form (6) with different values of θ_{ij} and T [13, 15]. This strategy has been used to express higher order potentials that are concave functions of the value of modular function [15]. If the linearizations are chosen appropriately, then it can be shown that our algorithm computes an FPTAS (fully polynomial-time approximation scheme). We prove Lemma 1 [16], using results from [7, 22].

Lemma 1. *If $|\mathcal{G}|$ is fixed, there is an FPTAS for minimizing energy* (5) *with any nondecreasing, nonnegative concave function F,* i.e.*, there is an algorithm that, for any $\epsilon > 0$, runs in time polynomial in n and $1/\epsilon$ and returns a solution $\mathbf{x}$ with $E(\mathbf{x}) \le (1 + \epsilon)E(\mathbf{x}^*)$.*

Asymmetric potentials. In the derivation above, we assumed that $\psi_{ij}(x_i, x_j)$ is symmetric. Alternatively, we may

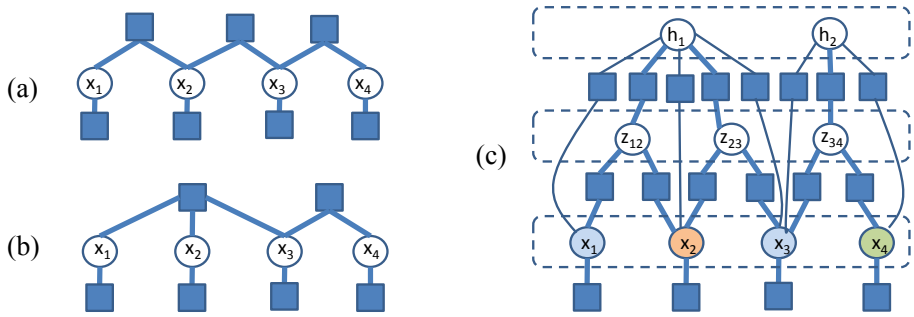

Figure 2: (a) The Markov Random Field model for image segmentation. (b) The cooperative cut model [11]. (c) The transformed cooperative cut model.

assume that the pairwise potentials are directed, *i.e.*, that $\psi_{ij}(x_i, x_j) = \theta(I_i, I_j)(x_i - x_j)_+ = \theta(I_i, I_j)(x_i - x_i x_j)$, and we have one potential $\psi_{ij} \neq \psi_{ji}$ for each direction. These correspond to directed edges in the graph. The potentials (6) (and equally (5)) extend to directed edges too:

$$\Psi_g(\mathbf{x}) = \min \Big\{ \sum\nolimits_{ij \in \mathcal{E}_g} \theta_{ij}(x_i - x_j)_+, T \Big\}. \qquad (10)$$

The two potentials ψ_{ij} and ψ_{ji} may belong to different groups g; in that case they are discounted differently and independently. Theorem 1 can still be transferred to directed edges too, using the same approach:

Corollary 1. *The cooperative cut potential* (10) *is equivalent to a pairwise potential with* $|\mathcal{E}_g|+1$ *auxiliary variables.*

4. Exact MAP Inference in Deep Fields

To perform MAP inference, we now aim to minimize the derived QPBF and compute

$$\arg\min_{\mathbf{x}} E(\mathbf{x}) = \arg\min_{\mathbf{x}} \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{g \in \mathcal{G}} \Psi_g(\mathbf{x}), \quad (11)$$

$$\text{or} \quad \arg\min_{\mathbf{x},\mathbf{h},\mathbf{z}} \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{g \in \mathcal{G}} \nu_g(\mathbf{x}, h_g, \mathbf{z})$$

where $\nu_g(\mathbf{x}, h_g, \mathbf{z}) =$

$$T + \sum_{ij \in \mathcal{E}_g} \theta_{ij}((x_i + x_j - 2z_{ij})h_g - 2(x_i + x_j)z_{ij} + 4z_{ij}) - T h_g$$

It is well known that submodular QPBFs can be minimized exactly in polynomial time by transforming them to an equivalent (s,t)-mincut problem [18, 2]. However, these results do not apply to our energy, because it contains non-submodular terms that make QPBFs NP-hard to minimize in general. That said, a number of methods have been proposed for minimizing such functions approximately [2, 3, 26]. But the problem (11) contains one additional binary variable per edge z_{ij} and per group of edges h_g, and this large number makes any such direct relaxation approach computationally very expensive.

However, only the terms $(x_i + x_j)h_g$ make the energy (11) non-submodular; all other terms of the function are submodular. If we fix the value of the variables h_g, then the non-submodular pairwise terms $(x_i + x_j)h_g$ become linear, and the resulting submodular energy function can be minimized exactly and efficiently. As there are only few h_g, we compute the MAP solution of the overall energy by finding the global minimum for each possible assignment of the h_g using a graph cut algorithm [5]. Even more, if we fix the h_g, we can directly use Equation (7) and do not need to explicitly implement the z_{ij} variables. The complexity of this algorithms is $O(\tau(|\mathcal{V}|, |\mathcal{E}|)2^{|\mathcal{G}|})$ where $\tau(n, m)$ is the complexity of computing a single st-mincut in a graph with n nodes and m edges, and $|\mathcal{G}|$ is the number of edge groups.

Efficient computation using dynamic graph cuts. To speed up the search over all possible assignments of h_g we use a dynamic max-flow algorithm [17] that rapidly solves a set of related max-flow problems. The dynamic algorithm is particularly effective when the max-flow problems are similar to each other, and therefore we order the subproblems of minimizing the projections so that they only differ in the value of one h_g. Moreover, we sort the variables h_g in decreasing order with respect to the total weight of edges that correspond to type g and make variables with lower weight change their value more often than those with higher weight. These heuristics accelerate the exhaustive search by a factor of 5 to 100 depending on the problem instance.

In addition, we investigate three greedy heuristics that are polynomial in the number $|\mathcal{G}|$ of types:

Greedy: Start with all $h_g = 1$, and iteratively switch the label of that h_g whose switch to zero decreases the energy most; repeat until there is no more improvement.

Descent: Like *Greedy*, but applying the switch whenever it improves the energy without searching for the best switch.

1-pass: Pass over all variables h_g only once, applying all

switches that decrease the energy.

All of these heuristics require testing whether changing the value of a particular variable h_g will reduce the energy. This test can be performed efficiently by computing min-marginals via dynamic graph cuts [17].

Inference by Truncation. If we want to avoid searching over all assignments of the group indicator variables h_g, then we can alternatively employ the common heuristic of simply pruning out the non-submodular terms, here the terms $(x_i + x_j)h_g$, from the energy [19, 24]. This truncation results in an under-approximation of the energy since the deleted terms contributed a positive cost. Minimizing the remaining energy using graph cuts, we obtain solutions $\mathbf{h}^*$ and $\mathbf{x}^*$. If $h_g^* = 0$ or $x_i^* = x_j^* = 0$ for all $(i, j) \in g$, then $\mathbf{x}^*$ is the MAP solution since no additional cost is added by the missing terms. For our energies here, pruning is not necessary since the exact algorithm is fast enough, and we do not truncate in the sequel.

4.1. Multiple labels

Equivalently to binary MRFs, our higher-order models can be extended to variables $x_i \in \mathcal{L}$ that take labels in a discrete space $\mathcal{L}$. One way to generalize the potentials (5) to multiple labels is to define *label-sensitive* couplings $\Psi_g(\mathbf{x}) = \sum_{\ell \in \mathcal{L}} F_\ell(\sum_{ij \in \mathcal{E}_g} \psi_{ij}(x_i, x_j) 1_{[x_i = \ell]})$. Using an analogous approach as in the binary case, we introduce auxiliary variables $h_{g,\ell}$. As in the binary case, fixing all variables $h_{g,\ell}$ makes the energy pairwise and the potentials metric, and techniques such as the α-expansion algorithm [6] apply. Again, we iterate over the $h_{g,\ell}$, and the same heuristics for speed-ups as above can be used.

Lemma 2. *The multi-label model can be reduced to a non-submodular pairwise model analogous to the binary model. If $|\mathcal{L}|$ and $|\mathcal{G}|$ are constants, then, with the help of $|\mathcal{L}||\mathcal{G}|$ auxiliary variables, we can compute an exact expansion move in polynomial time.*

Theorem 2. *The expansion move algorithm for co-operative multiple-label potentials of the form* (5) *returns a solution* $\mathbf{x}$ *that is a* $2c$*-approximation to the optimal solution* $\mathbf{x}^*$*: $E(\mathbf{x}) \leq 2cE(\mathbf{x}^*)$, for $c = \max_{\ell_1,\ell_2,g} F'_{\ell_1}(0)/F'_{\ell_2}(\sum_{ij \in \mathcal{E}_g} \theta_{ij})$.*

Theorem 2 is proved in [16]. We tested the algorithm on the MSRC data set. The example results in Figure 5 show that label-sensitive coupling improves detailed segmentations in the multi-label case too.

5. Experiments

We compare the performance of the proposed model and algorithms to state-of-the-art methods for image segmentation. Following [11], we use the problem of interactive binary image segmentation and their "twigs and legs" dataset[2]. In the experiments, we use piece-wise linear coupling functions F with one breakpoint of the form

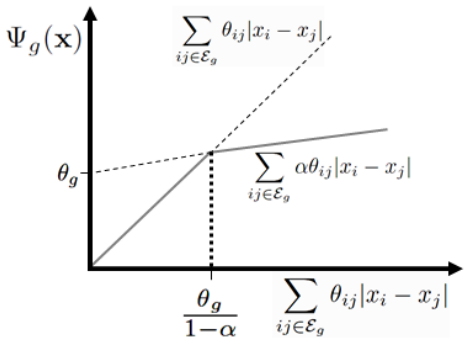

Figure 3: Potential used in the experiments (Eqn. (12)).

$$\Psi_g(\mathbf{x}) = \lambda \min \left\{ \sum_{ij \in \mathcal{E}_g} \theta_{ij}(x_i - x_j)_+, \sum_{ij \in \mathcal{E}_g} \alpha\theta_{ij}(x_i - x_j)_+ + \theta_g \right\} \tag{12}$$

where $\theta_g = \vartheta \sum_{ij \in \mathcal{E}_g} \theta_{ij}$, and $\vartheta, \alpha \in (0, 1]$. The three parameters have the following effect: λ controls the weight of the high-order potentials relative to the unary ones, α defines the slope after the breakpoint, and ϑ controls the position of the breakpoint which is located at $\frac{\theta_g}{1-\alpha}$, depicted in Figure 3. Like [11], we use the directed potentials (10).

The unary potentials are computed by fitting a Gaussian mixture model with 5 components to pixels of seed regions. We use an 8-neighbor graph structure and contrast-dependent Potts pairwise potentials $\theta_{ij} = 2.5 + 47.5 \exp(-0.5\|I_i - I_j\|^2/\sigma)$, where σ is the mean of the color gradients in the image. Edge types are obtained by k-means clustering of graph edges. The feature of an edge (i, j) used for clustering is the three-dimensional vector $I_i - I_j$ of RGB differences. Edges between identically colored pixels form an extra type in which discontinuities are penalized linearly, *i.e.*, $\theta_{g_0} = \sum_{ij \in \mathcal{E}_{g_0}} \theta_{ij}$. We compare the standard MRF approach, the iterative algorithm in [11], the new exact algorithm and the three heuristics.

Qualitative and Quantitative Evaluation. Figures 4 and 5 show qualitative results of different algorithms on the binary and multi-label image segmentation problems. The exact solutions of the cooperative cut potentials extract fine structures much better than the standard model, and slightly improve on the approximate solutions. Table 1 displays a quantitative comparison. We use the following criteria:

Energy: energy (4) averaged over the data. The energy for each image is normalized such that 0.0 is the global minimum and 1.0 is the energy for the MAP labeling of the standard pairwise MRF.

[2] http://ssli.ee.washington.edu/~jegelka/cc/

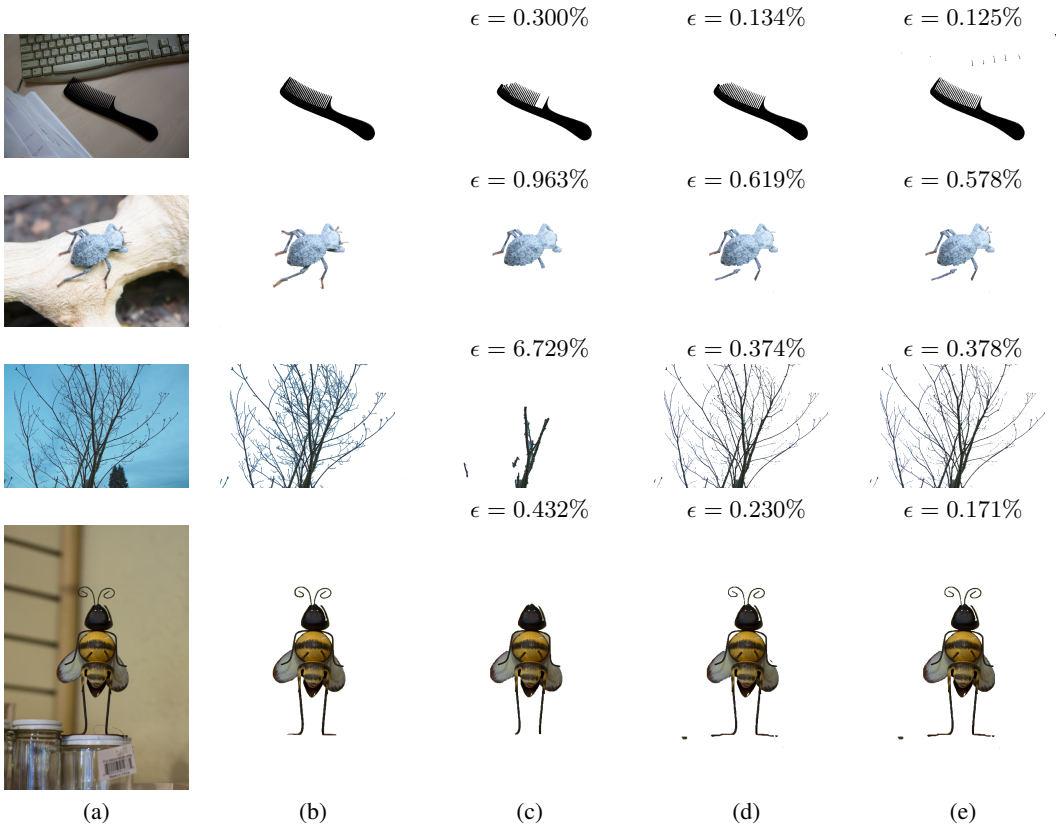

Figure 4: Qualitative results of different algorithms and models. (a) Images from the dataset; (b) Ground truth segmentation;(c) MAP solution from a standard pairwise MRF model; (d) result of the iterative approximation from [11]; (e) Global minimum of cooperative cut energy (4) obtained by our method. ϵ denotes the percentage of incorrectly labeled pixels.

Time: average running time, including time for building the required data structures, executing the algorithm and excluding I/O operations. All algorithms are implemented in C++, using the min-cut/max-flow algorithm from [5][3].

Hamming: pixel-wise Hamming distance to the ground truth. The distance is divided by the number of labeled pixels in the image and then averaged over the data.

Twig: Hamming distance computed on the hand-labeled thin structure regions from [11]

HAC: Hamming distance Averaged over Class size, where the penalty for mislabeling a pixel is inversely proportional to the number of pixels of that class in the ground truth labeling.

JD: Jaccard distance between the set of foreground pixels in the current segmentation and the corresponding set in the ground truth; also known as the intersection/union metric and used for evaluation e.g. in the PASCAL VOC segmentation challenge.

The comparison shows a favorable performance of the exact algorithm and its heuristic variations. Moreover, these results demonstrate that not only an approximate minimum as in [11], but in particular the global minimum of the cooperative cut energy results in very low errors, especially compared to the standard model. This suggests that cooperative cut energies very well capture the particularities of segmentation problems and settles the question raised in the introduction.

[3] http://pub.ist.ac.at/~vnk/software/maxflow-v3.02.src.tar.gz

Multiple labels. We also tested the multi-label potential that is sketched in Section 4.1 on the MSRC data. The edge groups and function F were defined analogous to the binary case, and we compute exact expansion moves, and the unary potentials were learned as GMMs as in the binary case (details in [16]). However, we found that the ground truth labeling on the MSRC data is not very detailed on fine structures such as trees, and too coarse to be used for qualitatively evaluating the details of segmentation results. Therefore, Figure 5 only shows example quantitative results that nevertheless suggest beneficial effects of edge coupling for detailed multi-label segmentations too.

6. Conclusions and Discussion

In this paper, we explored a hierarchical model that is equivalent to the cooperative cut model of [11]. The associated exact algorithm and the proposed speed-ups have a number of implications.

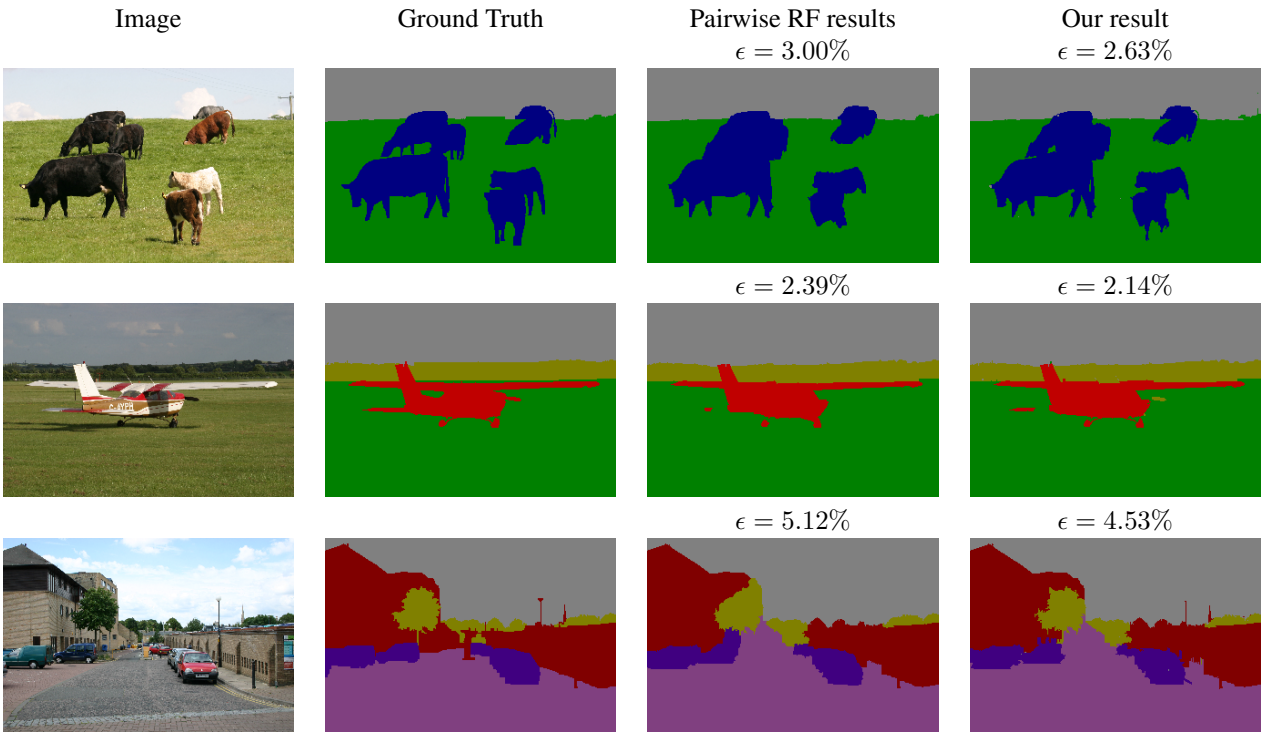

Figure 5: Qualitative results with multiple labels. (Column 1) Images from the dataset; (Column 2) Ground truth segmentation;(Column 3) MAP solution from a standard pairwise MRF model; (Column 4) result of our hierarchical model using the inference method described in section 4.1. The Error ϵ denotes the percentage of incorrectly labeled pixels, keeping in mind that he ground truth labeling is not very accurate at the level of details.

Table 1: Quantitative results for the "twig&legs" dataset [11] with 10 edge classes. Parameters are chosen in such a way that the global minimum of the energy has maximum Hamming accuracy: $\lambda = 1.5$, $10^4\vartheta = 50$, $\alpha = 0$.

Method	Energy	Hamming	Twig	HAC	Jaccard (JD)	Time
GraphCut	1.00	1.61	39.10	9.96	23.34	**0.19**
It. bound min. [11]	0.39	0.77	26.93	4.98	13.27	0.47
Global Minimum	**0.00**	**0.73**	**21.82**	**4.45**	**12.37**	14.32
Greedy	**0.00**	**0.73**	**21.82**	**4.45**	**12.37**	5.05
Descent	**0.00**	**0.73**	**21.82**	**4.45**	**12.37**	2.37
1 pass of descent	0.03	0.87	22.19	5.20	13.88	1.23

Implications for Cooperative Cuts. The algorithm solves the class of cooperative cut potentials proposed in [11] exactly or within a factor of $(1 + \epsilon)$ and thereby admits to show that not only an approximate solution (as in [11]), but also the true MAP solution of those potentials overcomes the short-boundary bias that poses severe problems to the standard pairwise model for image segmentation. The exact solutions improve over state-of-the-art results for image segmentation.

NP-hardness. The exact algorithm does not violate the NP-hardness of cooperative cuts. Instead, it demonstrates that the potentials (5) are fixed-parameter tractable: if the number of edge groups is assumed to be constant (and in practice it is small), then the algorithm runs in polynomial-time. The proposed heuristics make our approach a practical alternative and in the experiments always find the optimal solution. If $|\mathcal{G}|$ is not used explicitly in the analysis and is not fixed, then no polynomial-time algorithm exists, in fact, then not even a constant-factor approximation for the class of potentials (6) is possible [31].

Implications beyond segmentation. In general, the proposed algorithms are an effective and very practical method for a sub-class of minimum cuts with submodular costs [11] – the images in the experiments have up to $6 \cdot 10^5$ x-variables. By defining different groups of coupled edges,

this class of functions is applicable beyond fine-structured segmentation, and the results in this work extend to all these applications. In particular, the accelerated algorithms and theoretical results also apply to models in [12, 9, 31] that all have the form (5) or (6). The P^n *functions* defined in Equations (14) and (15) in [12] are of the form (5) here, and shown to be amenable to expansion moves if the pairwise potentials form a semi-metric. By Corollary 1, the condition of a semi-metric can be relaxed to asymmetric potentials. In addition, the formulation here does not require general submodular minimization (which can be impractical and also does not in general apply if asymmetric potentials are used). Beyond computer vision, functions of the form (6) also include *label costs* [31] that arise in computer security.

Deep models. The explicit hierarchical formulation in Figure 2 shows that the higher-order potentials for improved segmentations can be expressed as a deep multi-layer pairwise model with additional hidden (auxiliary) variables. This viewpoint suggests a comparison to previously proposed deep models for image labeling. Compared to those, our model has a large number of hidden units representing the useful higher-order interactions. Furthermore, its connectivity structure is well-specified and sparse. Considering the expressive power of cooperative cut models [11, 9, 12, 31], this connection highlights the representational power of multi-layer random fields.

Acknowledgements A. Osokin was supported by RFBR grants 12-01-31254, 12-01-33085.

References

[1] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *ECCV*, pages I: 428–441, 2004.

[2] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.

[3] E. Boros, P. Hammer, and G. Tavares. Local search heuristics for quadratic unconstrained binary optimization (QUBO). *J. Heuristics*, 13(2):99–132, 2007.

[4] Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV*, pages I: 105–112, 2001.

[5] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *PAMI*, 26(9):1124–1137, 2004.

[6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *EEE Trans. on Pattern Analysis and Machine Intelligence*, 23, 2001.

[7] I. Diakonikolas and M. Yannakakis. Succinct approximate convex pareto curves. In *SODA*, 2008.

[8] R. G. Downsey and M. R. Fellows. *Parameterized complexity*. Springer, 1999.

[9] S. Gould. Multi-class labeling with non-local matching constraints. In *CVPR*, 2012.

[10] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *CVPR*, pages 2993–3000, 2009.

[11] S. Jegelka and J. Bilmes. Submodularity beyond submodular energies: Coupling edges in graph cuts. In *CVPR*, pages 1897–1904, 2011.

[12] P. Kohli, M. Kumar, and P. Torr. P^3 and beyond: Solving energies with higher order cliques. In *CVPR*, 2007.

[13] P. Kohli and M. P. Kumar. Energy minimization for linear envelope MRFs. In *CVPR*, pages 1863–1870, 2010.

[14] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.

[15] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. *IJCV*, 82(3):302–324, 2009.

[16] P. Kohli, A. Osokin, and S. Jegelka. Supplementary material.

[17] P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV*, volume II, pages 922–929, 2005.

[18] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV*, 2002.

[19] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *ICCV*, 2005.

[20] V. S. Lempitsky and Y. Boykov. Global optimization for shape fitting. In *CVPR*, 2007.

[21] V. S. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *ICCV*, pages 277–284, 2009.

[22] S. Mittal and A. Schulz. An FPTAS for optimizing a class of low-rank functions over a polytope. *Optimization online*, 2011.

[23] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.

[24] A. Raj and R. Zabih. A graph cut algorithm for generalized image deconvolution. In *ICCV*, pages 1048–1054, 2005.

[25] C. Rother, P. Kohli, W. Feng, and J. Jia. Minimizing sparse higher order energy functions of discrete variables. In *CVPR*, pages 1382–1389, 2009.

[26] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.

[27] R. Salakhutdinov and G. E. Hinton. Deep Boltzmann machines. In *AISTATS*, 2009.

[28] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields. In *ECCV*, pages 16–29, 2006.

[29] D. Tarlow, I. E. Givoni, and R. S. Zemel. Hop-map: Efficient message passing with high order potentials. In *AISTATS*, 2010.

[30] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *CVPR*, 2008.

[31] P. Zhang, C. J.-Y, L.-Q. Tang, and W.-B. Zhao. Approximation and hardness results for label cut and related problems. *J. Comb. Optim.*, 2011.