# Efficient SDP Inference for Fully-connected CRFs Based on Low-rank Decomposition

Peng Wang[1], Chunhua Shen[1,2], Anton van den Hengel[1,2]
[1]University of Adelaide, Australia. [2]Australian Centre for Robotic Vision.

Conditional random fields (CRFs) have been one of the most successful approaches to semantic pixel labelling, which solves the problem as maximum a posteriori (MAP) estimation. Standard CRFs typically contain unary potentials defined on local features and edge potentials defined on 4- or 8-neighbouring pixels. Although these CRF models have achieved encouraging results for segmentation, they fail to model more complex priors such as long-range contextual relationships. In the literature, fully-connected CRFs have been proposed for this purpose. The main challenge for inference on fully-connected CRFs stems from the computational cost. Although there have been a variety of methods for MAP estimation [2, 4], they are usually computationally infeasible for such cases. The authors of [3] have proffered an efficient mean field approximation method for MAP inference of fully-connected pairwise CRFs. In their algorithms, the computational burden can be expressed as the products of kernel matrices and column vectors. A filter-based method [1] is used to accelerate the computation of such matrix-vector products. There are two major limitations to the method in [3]: 1) Mean field approximation may converge to local minimum and is often sensitive to initialization; 2) The pairwise terms are assumed to be in the form of a weighted mixture of Gaussian kernels such that the filter method [1] can be applied for fast computation.

In this work, an efficient, yet general SDP approach is proposed for MAP estimation in large-scale fully-connected CRFs. The core of the proposed algorithm is a tailored quasi-Newton method, which solves a specialized SDP dual problem and takes advantage of the low-rank matrix approximation for fast computation. The superiority of our approach is two-fold: 1) In contrast to mean field, our approach solves a *convex* problem and generally provides more stable and accurate solutions. 2) As alternatives to the filter-based methods, the use of low-rank approximation methods relax the limitation to the pairwise term from being (a mixture of) Gaussian kernels to all symmetric positive-semidefinite (SPSD) kernels.

Let us consider a pairwise fully-connected CRF with $N$ variables $\mathbf{x} = [x_1, \cdots, x_N]^\top$ and $L$ states per variable, the associated MAP inference problem can be expressed as the following energy minimization problem:

$$\min_{\mathbf{x}\in\mathcal{L}^N} \sum_{i\in\mathcal{N}} \psi_i(x_i) + \sum_{i,j\in\mathcal{N}, i<j} \delta(x_i \neq x_j) \sum_{m=1}^{M} w^{(m)} \mathrm{k}^{(m)}(\mathbf{f}_i, \mathbf{f}_j), \quad (1)$$

where $\mathcal{N} := \{1, \ldots, N\}$, $\mathcal{L} := \{1, \ldots, L\}$. $\psi_i : \mathcal{L} \to \mathbb{R}$ corresponds to the unary potentials and the pairwise potentials are assumed to be a linear combination of $M$ SPSD kernel functions. $\mathbf{f}_i, \mathbf{f}_j \in \mathbb{R}^D$ indicate $D$-dimensional feature vectors corresponding to variables $x_i$ and $x_j$ respectively. $\mathrm{k}^{(m)} : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ denotes the $m$-th SPSD kernel and $w^{(m)} > 0$ is the associated weight. $\delta(\cdot)$ denotes the indicator function.

By defining $\mathbf{X} \in \{0,1\}^{N\times L}$, $\mathbf{H} \in \mathbb{R}^{N\times L}$ and $\mathbf{K} \in \mathcal{S}_+^N$ as $X_{i,l} = \delta(x_i = l)$, $H_{i,l} = \psi_i(l)$ and $K_{i,j} = \sum_{m=1}^{M} w^{(m)} \mathrm{k}^{(m)}(\mathbf{f}_i, \mathbf{f}_j)$, the problem (1) is equivalent to the following binary quadratic problem (BQP):

$$\min_{\mathbf{X}\in\{0,1\}^{N\times L}} \langle \mathbf{H}, \mathbf{X} \rangle - \frac{1}{2} \langle \mathbf{X}\mathbf{X}^\top, \mathbf{K} \rangle, \quad \text{s.t. } \textstyle\sum_{l=1}^{L} X_{i,l} = 1, \forall i \in \mathcal{N}, \quad (2)$$

which can be further relaxed by introducing $\mathbf{Y} := \begin{bmatrix} \mathbf{I}_L \\ \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{X} \end{bmatrix}^\top$:

$$\min_{\mathbf{Y}\in\mathcal{S}_+^{N+L}} \quad \langle \mathbf{Y}, \frac{1}{2} \begin{bmatrix} \mathbf{0} & \mathbf{H}^\top \\ \mathbf{H} & -\mathbf{K} \end{bmatrix} \rangle, \quad (3a)$$

$$\text{s.t.} \quad Y_{l,l} = 1, \ l \in \mathcal{L}, \quad (3b)$$

$$\frac{1}{2}(Y_{l,l'} + Y_{l',l}) = 0, \ l \leq l', l, l' \in \mathcal{L}, \quad (3c)$$

$$\frac{1}{2} \textstyle\sum_{l=1}^{L} (Y_{i+L,l} + Y_{l,i+L}) = 1, \ i \in \mathcal{N}, \quad (3d)$$

$$Y_{i+L,i+L} = 1, \ i \in \mathcal{N}, \quad (3e)$$

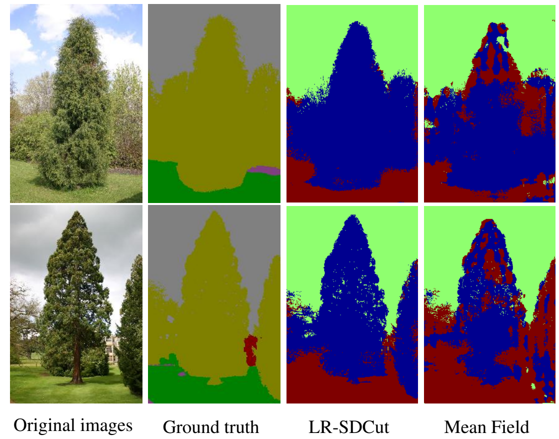


Figure 1: Results for image co-segmentation. Our method performs significantly better than mean field approximation.

where $\mathbf{I}_L$ denotes an $L \times L$ identity matrix and $\mathbf{0}$ denotes an all-zero matrix.

Solving the above SDP problem using conventional interior-point methods is typically computationally inefficient, which in general needs $\mathcal{O}(N^4)$ time. To this end, we propose to solve the problem (3) approximately based on the method in [5] (refer to as SDCut). In the formulation of SDCut, a Frobenius-norm term $(\|\mathbf{Y}\|_F^2 - (N+L)^2))/2\gamma$ is added to the objective function (3a) and results in an accurate approximation to (3) given a sufficiently large constant $\gamma$ [5]. This approximated problem has a much simpler Lagrangian dual than the original problem (3), which can be solved using quasi-Newton methods in $\mathcal{O}(N^3)$ time. Although the original SDCut algorithm is more scalable than interior-point methods, it still cannot be applied directly to large-scale fully-connected CRFs where $N$ can be more than $10^6$.

In this work, several significant improvements over SDCut are presented and result in an improved SDCut algorithm (refer to as LR-SDCut) whose computational complexity is further reduced to be *linear* in $N$. Similar to mean field approaches [3], the computationally intensive part in solving (3) is on the computation of the products of the positive semidefinite matrix $\mathbf{K}$ and random column vectors, which is accelerated using the low-rank approximation of kernel matrices, in particular Nyström methods [6], instead of the filter-based method used in [3].

The proposed SDP approach is much more general and scalable, and thus has a broader range of applications. Our method can handle fully-connected CRFs of #states $\times$ #variables up to $10^6$. In particular, we show that on an image co-segmentation application, the fast method of [3] is not applicable while our method achieves superior segmentation accuracy (see Figure 1).

[1] A. Adams, J. Baek, and M. A. Davis. Fast high-dimensional filtering using the permutohedral lattice. In *EUROGRAPHICS*, 2010.

[2] J. H. Kappes et al. A comparative study of modern inference techniques for discrete energy minimization problems. In *Proc. CVPR*, 2013.

[3] P. Krähenbühl and V. Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials. In *Proc. NIPS*, 2011.

[4] R. Szeliski et al. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE T-PAMI*, 30(6):1068–1080, 2008.

[5] P. Wang, C. Shen, and A. Hengel. A fast semidefinite approach to solving binary quadratic problems. In *Proc. CVPR*, 2013.

[6] C. Williams and M. Seeger. The effect of the input density distribution on kernel-based classifiers. In *Proc. ICML*, 2000.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.