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

Peng Wang1, Chunhua Shen1,2, Anton van den Hengel1,2
1University of Adelaide, Australia. 2Australian 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 $x = [x_1, \ldots, x_N]^T$ and $L$ states per variable, the associated MAP inference problem can be expressed as the following energy minimization problem:

$$
\min_{x \in \mathcal{X}} \sum_{i \in \mathcal{N}} \psi_i(x_i) + \sum_{i,j \in \mathcal{N}, i \neq j} \delta(x_i \neq x_j) \sum_{m=1}^{M} w(m)k(m)(f_i, f_j),
$$

(1)

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

By defining $X \in \{0, 1\}^{N \times L}$, $H \in \mathbb{R}^{N \times L}$ and $K \in S_+^N$ as $X_{i,j} = \psi_i(l)$ and $K_{i,j} = \sum_{m=1}^{M} w(m)k(m)(f_i, f_j)$, the problem (1) is equivalent to the following binary quadratic problem (BQP):

$$
\min_{x \in \{0, 1\}^N} \frac{1}{2} (HX) \cdot \begin{bmatrix} H & K \end{bmatrix}, \quad \text{s.t.} \sum_{i=1}^{N} X_{i,j} = 1, \quad \forall i \in \mathcal{N},
$$

(2)

which can be further relaxed by introducing $Y := \begin{bmatrix} 0 & K \end{bmatrix} \begin{bmatrix} 0 \ 1 \ -K \end{bmatrix}^T$

$$
\min_{Y \in \mathbb{S}_{+}^{N \times L}} \langle Y, \begin{bmatrix} 1 & 0 \\ 0 & H' - K \end{bmatrix} \rangle, \quad \text{s.t.} \quad Y_{i,l} = 1, \quad i \in \mathcal{L}, \quad l \leq l', \quad l, l' \in \mathcal{L},
$$

(3a)

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

(3b)

$$
\frac{1}{2} \sum_{l=1}^{L} Y_{l,l} = 1, \quad i \in \mathcal{N},
$$

(3c)

$$
Y_{i,l+i+L} = 1, \quad i \in \mathcal{N},
$$

(3d)

where $I_L$ denotes an $L \times L$ identity matrix and $0$ denotes an all-zero matrix.

Solving the above SDP problem using conventional interior-point methods is typically computationally inefficient, which in general needs $O(N^3)$ 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 $(\|Y\|_F^2 - (N + L)^2)/2\gamma$ is added to the objective function (3a) and results in an accurate approximation to (3) given a sufficiently large constant $\gamma$. This approximated problem has a much simpler Lagrangian dual than the original problem (3), which can be solved using quasi-Newton methods in $O(N^3)$ 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 $K$ 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 $\#\text{states} \times \#\text{variables}$ up to $10^5$. 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).

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