New Convex Relaxations for MRF Inference with Unknown Graphs (Supplementary Material)

Zhenhua Wang †*, Tong Liu †, Qinfeng Shi ‡, M. Pawan Kumar ♯, Jianhua Zhang †
† Zhejiang University of Technology, ‡ The University of Adelaide, ♯ The University of Oxford

This supplementary provides:

- Motivations for estimating graph structures within the inference problem.
- The maximum likelihood interpretation of inferring MRF graphs and labels simultaneously.
- Detailed exposition of the matrix Θ.
- The proofs of Propositions 1-3.
- The pseudocode of our QP+CCCP algorithm.
- Additional experimental results.

1. Motivations for estimating graph structures within the inference problem

In some problems it is possible to learn the graph structure instead of inferring it. This approach is well suited to problems where all the instances share the same structure (that is, the structure is "homogeneous"). However, as noted by Lan *et al*. 2010, who introduced the problem of simultaneous estimation of MRF labels and structures, some problems are characterized by the lack of a common basic structure across its instances. We refer to these as "heterogeneous" problems. In other words, each instance of the problem has a different graph structure. One such problem, group activity recognition, is studied in our paper, where we do not even know the number of persons (that is, the number of random variables of the MRF) in a scene. Furthermore, the persons can enter and exit the scenes at different time frames thereby changing the number of variables. Treating the structure as an unknown and inferring it for each instance is an intuitive solution to address this problem.

2. Maximum Likelihood Interpretation

Because the underlying structure of heterogeneous graphs is changing and there are no multiple observations (or not sufficient many) for a fixed structure to draw meaningful statistical estimation, learning graph structures in this case is unachievable. Thus we would like to find the best graph structure and label jointly for each instance $\mathbf{x}$. Let $\theta_i(y_i)$ denote the unary potential when the node i takes a label y_i, and let $\theta_{s,t}(y_s, y_t)$ denote the pairwise potential when nodes s, t take y_s, y_s at the same time. We can define the joint likelihood $\mathrm{P}(\mathbf{y}, G|\,\boldsymbol{\theta})$ as

$$\mathrm{P}(\mathbf{y}, G|\,\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\big(-\sum_{i \in V} \theta_i(y_i) - \sum_{(s,t)\in E} \theta_{st}(y_s, y_t)\big). \tag{1}$$

Note in $G = (V, E)$, the node set V is given and the edge set E is the unknown to be estimated. The normalisation constant $Z(\boldsymbol{\theta})$, or called partition function is

$$Z(\boldsymbol{\theta}) = \sum_{\hat{\mathbf{y}} \in \mathcal{Y}} \sum_{\hat{G} \in \mathcal{G}} \exp\big(-\sum_{i \in \hat{V}} \theta_i(\hat{y}_i) - \sum_{(s,t)\in \hat{E}} \theta_{st}(\hat{y}_s, \hat{y}_t)\big). \tag{2}$$

Here $\hat{\mathbf{y}} = (\hat{y}_i)_{i=1,2,\ldots,n}$, $\hat{G} = (\hat{V}, \hat{E})$. The following equivalence holds:

$$\operatorname*{argmax}_{\mathbf{y},G} \mathrm{P}(\mathbf{y}, G|\,\boldsymbol{\theta}) = \operatorname*{argmax}_{\mathbf{y},G} \ln \mathrm{P}(\mathbf{y}, G|\,\boldsymbol{\theta}) = \operatorname*{argmin}_{\mathbf{y},G} \Big\{ \sum_{i \in V} \theta_i(y_i) + \sum_{(s,t)\in E} \theta_{st}(y_s, y_t) \Big\}. \tag{3}$$

This means that inference with unknown graphs can be done by maximising the potential function (or minimising the energy) jointly for both graph and label.

3. The Detailed Structure of Θ Matrix

Θ is a block diagonal matrix has the following form:

$$\Theta = \begin{pmatrix}
\theta_{11}(1,1) & 0 & \ldots & 0 & 0 \\
\theta_{11}(1,2) & \vdots & & \vdots & \vdots \\
\vdots & & & & \\
\theta_{11}(c,c) & 0 & & & \\
0 & \theta_{12}(1,1) & & 0 & 0 \\
\vdots & \theta_{12}(1,2) & & \vdots & \vdots \\
 & \vdots & & & \\
 & \theta_{12}(c,c) & & 0 & 0 \\
 & 0 & & \vdots & \vdots \\
 & \vdots & & & \\
 & & & \theta_{n-1n}(1,1) & 0 \\
 & & & \theta_{n-1n}(1,2) & \vdots \\
 & & & \vdots & \\
 & & & \theta_{n-1n}(c,c) & \\
 & & & 0 & \theta_{nn}(1,1) \\
 & & & \vdots & \theta_{nn}(1,2) \\
 & & & & \vdots \\
 & & & & \theta_{nn}(c,c)
\end{pmatrix},$$

here we let $\theta_{ij}(a,b) = 0 \ \forall i = j, a, b \in \mathcal{Y}$. As in the paper, n, c denote the number of nodes and the cardinality of the label set respectively.

4. Proofs

4.1. Proof of Proposition 1

Proof According to the first two groups of constraints in problem (6), $\lambda_{ij}(y_i, y_j) \leq \min\{z_{ij}, \mu_{ij}(y_i, y_j)\}$.

We define $\delta(.,.) = 1$ if both its arguments are true, and 0 otherwise.

$$\begin{aligned}
& & & \sum_{a,b} \lambda_{ij}(a,b) = z_{ij}, \\
& \Longrightarrow & & \sum_{a,b} (1 - \delta(a = y_i, b = y_j))\lambda_{ij}(a,b) + \lambda_{ij}(y_i, y_j) = z_{ij}, \\
& \Longrightarrow & & \sum_{a,b} (1 - \delta(a = y_i, b = y_j))\mu_{ij}(a,b) + \lambda_{ij}(y_i, y_j) \geq z_{ij}, \\
& \Longrightarrow & & 1 - \mu_{ij}(y_i, y_j) + \lambda_{ij}(y_i, y_j) \geq z_{ij}, \\
& \Longrightarrow & & \lambda_{ij}(y_i, y_j) \geq z_{ij} + \mu_{ij}(y_i, y_j) - 1.
\end{aligned}$$

Since $\lambda_{ij}(y_i, y_j) \geq 0$, we have $\lambda_{ij}(y_i, y_j) \geq \max\{0, z_{ij} + \mu_{ij}(y_i, y_j) - 1\}$. That is to say, the feasibility set of LP-M is a subset of LP-W. Moreover, it is easy to find a feasible point of LP-W, which is infeasible to LP-M. The proof is complete. ■

4.2. Proof of Proposition 2

Proof First $\mathbf{d}^*$ is a feasible point of problem (12). This is because $\hat{\Theta} + \mathbf{diag}(\mathbf{d})$ is a Hermitian diagonally dominant matrix with non-negative diagonal entries, thus positive semidefinite. Second we show $\mathbf{d}^*$ is optimal using proof by contradiction below. Suppose we can find a $\hat{\mathbf{d}} \neq \mathbf{d}^*$ such that $\|\hat{\mathbf{d}}\|_1 < \| \mathbf{d}^* \|_1$ and $\hat{\Theta} + \mathbf{diag}(\hat{\mathbf{d}}) \succeq 0$. The $\hat{\mathbf{d}}$ has the following structure:

$$\hat{\mathbf{d}} = \big[[\boldsymbol{\alpha}_{ij}]_{i,j\in V, i\leq j}, [\gamma_{ij}]_{i,j\in V, i\leq j}\big],$$

where $[\boldsymbol{\alpha}_{ij}]_{i,j\in V, i\leq j}$ is a concatenation of vectors $\boldsymbol{\alpha}_{ij}$ in an order formed by enumerating all possible indices in turn; $\boldsymbol{\alpha}_{ij} = [\alpha_{ij}(y_i, y_j)]_{y_i, y_j}$ is a vector formed by enumerating all possible labels in turn; similarly, $[\gamma_{ij}]_{i,j\in V, i\leq j}$ is another vector formed by enumerating all possible indices. It follows from $\|\hat{\mathbf{d}}\|_1 < \| \mathbf{d}^* \|_1$ that

$$\sum_{i,j\in V, i\leq j} \sum_{y_i, y_j} |\alpha_{ij}(y_i, y_j)| + \sum_{i,j\in V, i\leq j} |\gamma_{ij}| < \sum_{i,j\in V, i\leq j} \sum_{y_i, y_j} |\theta_{ij}(y_i, y_j)|.$$

Let $\alpha_{ij}(y_i, y_j) = \frac{1}{2}|\theta_{ij}(y_i, y_j)| + \rho_{ij}(y_i, y_j)$, $\gamma_{ij} = \sum_{y_i, y_j} \frac{1}{2}|\theta_{ij}(y_i, y_j)| + \sigma_{ij}(y_i, y_j)$, $\rho_{ij}(y_i, y_j), \sigma_{ij}(y_i, y_j) \in \mathbb{R}$. The following inequality holds:

$$\begin{aligned}
& \Big|\sum_{i,j\in V, i\leq j} \sum_{y_i, y_j} \alpha_{ij}(y_i, y_j) + \sum_{i,j\in V, i\leq j} \gamma_{ij}\Big| < \sum_{i,j\in V, i\leq j} \sum_{y_i, y_j} |\theta_{ij}(y_i, y_j)|, \\
\Longrightarrow \quad & \Big|\sum_{i\leq j} \sum_{y_i, y_j} |\theta_{ij}(y_i, y_j)| + \sum_{i\leq j} \sum_{y_i, y_j} \rho_{ij}(y_i, y_j) + \sigma_{ij}(y_i, y_j)\Big| < \sum_{i\leq j} \sum_{y_i, y_j} |\theta_{ij}(y_i, y_j)|. \\
\Longrightarrow \quad & \sum_{i,j\in V, i\leq j} \sum_{y_i, y_j} \rho_{ij}(y_i, y_j) + \sigma_{ij}(y_i, y_j) < 0.
\end{aligned}$$

Let $\mathbf{Q}' = \hat{\Theta} + \mathbf{diag}(\hat{\mathbf{d}})$. Defining $\mathbf{v}$ and $\mathbf{u}$ as

$$\mathbf{v} = [\mathbf{v}_{ij}]_{i,j\in V, i\leq j}, \text{ where } \mathbf{v}_{ij} = [v_{ij}(y_i, y_j)]_{y_i, y_j}, \mathbf{u} = [u_{ij}]_{i,j\in V, i\leq j}.$$

Here $\mathbf{u}, \mathbf{v}$ share the same structures as $\boldsymbol{\alpha}$ and $[\gamma_{ij}]_{i,j\in V, i<j}$ respectively. Since $\hat{\Theta} + \mathbf{diag}(\hat{\mathbf{d}}) \succeq 0$, $[\mathbf{v}, \mathbf{u}]^\top \mathbf{Q}'[\mathbf{v}, \mathbf{u}] \geq 0$ should always hold. According to (5), we can see

$$[\mathbf{v}, \mathbf{u}]^\top \mathbf{Q}'[\mathbf{v}, \mathbf{u}] = \frac{1}{2} \sum_{i\leq j} \sum_{y_i, y_j} \pi_{ij}(y_i, y_j) + \sum_{i\leq j} \sum_{y_i, y_j} (v_{ij}^2(y_i, y_j)\rho_{ij}(y_i, y_j) + u_{ij}^2 \sigma_{ij}(y_i, y_j)).$$

Here

$$\pi_{ij}(y_i, y_j) = \begin{cases} |\theta_{ij}(y_i, y_j)|(v_{ij}(y_i, y_j) + u_{ij})^2 & \text{if } \theta_{ij}(y_i, y_j) > 0, \\ |\theta_{ij}(y_i, y_j)|(v_{ij}(y_i, y_j) - u_{ij})^2 & \text{otherwise.} \end{cases}$$

Let $u_{ij} = 1$. Let $v_{ij}(y_i, y_j) = -u_{ij}$ if $\theta_{ij}(y_i, y_j) > 0$, and $v_{ij}(y_i, y_j) = u_{ij}$ otherwise. Then $[\mathbf{v}, \mathbf{u}]^\top \mathbf{Q}'[\mathbf{v}, \mathbf{u}] < 0$, which contradicts $\mathbf{Q}' \succeq 0$. The proof is complete. ■

4.3. Proof of Proposition 3

Lemma 1 *The convex QP problem (14) is equivalent to the following*

$$\begin{aligned}
\min \quad & \sum_{i\in V}\sum_{y_i}\mu_i(y_i)\theta_i(y_i) + \sum_{i,j\in V, i<j}\sum_{y_i,y_j}\theta_{ij}(y_i,y_j)\lambda_{ij}(y_i,y_j) \\
\text{s.t.} \quad & \lambda_{ij}(y_i,y_j) \le \frac{(z_{ij}+\mu_{ij}(y_i,y_j)) - (z_{ij}-\mu_{ij}(y_i,y_j))^2}{2} \quad \textit{if} \quad \theta_{ij}(y_i,y_j) < 0, \\
& \lambda_{ij}(y_i,y_j) \ge \frac{(z_{ij}+\mu_{ij}(y_i,y_j))^2 - (z_{ij}+\mu_{ij}(y_i,y_j))}{2} \quad \textit{if} \quad \theta_{ij}(y_i,y_j) \ge 0, \\
& \mu, z \in \mathcal{O}.
\end{aligned}$$

Proof Expanding the objective function in (14):

$$\begin{aligned}
\chi^\top \mathbf{q} + \chi^\top \mathbf{Q}\,\chi \quad & = \chi^\top(\hat{\boldsymbol{\theta}} - \mathbf{d}^*) + \chi^\top(\hat{\Theta} + \mathbf{diag}(\mathbf{d}^*))\chi \\
& = \sum_{i\in V}\sum_{y_i}\mu_i(y_i)\theta_i(y_i) + \sum_{i<j}\sum_{y_i,y_j}\theta_{ij}(y_i,y_j)\beta_{ij}(y_i,y_j),
\end{aligned}$$

where

$$\beta_{ij}(y_i,y_j) = \begin{cases} \frac{1}{2}(z_{ij}+\mu_{ij}(y_i,y_j) - \frac{1}{2}(z_{ij}-\mu_{ij}(y_i,y_j))^2 & \text{if} \quad \theta_{ij}(y_i,y_j) < 0, \\ \frac{1}{2}(z_{ij}+\mu_{ij}(y_i,y_j)^2 - \frac{1}{2}(z_{ij}+\mu_{ij}(y_i,y_j)) & \text{otherwise.} \end{cases}$$

Here $\beta_{ij}(y_i,y_j)$ has two cases because of the absolute value used in the definition of $\mathbf{d}^*$ in (12). Now we can see that the convex QP problem (14) is equivalent to the following optimization problem:

$$\begin{aligned}
\min \quad & \sum_{i\in V}\sum_{y_i}\mu_i(y_i)\theta_i(y_i) + \sum_{i,j\in V, i<j}\sum_{y_i,y_j}\theta_{ij}(y_i,y_j)\lambda_{ij}(y_i,y_j), \\
\text{s.t.} \quad & \theta_{ij}(y_i,y_j)\lambda_{ij}(y_i,y_j) \ge \theta_{ij}(y_i,y_j)\beta_{ij}(y_i,y_j), \quad \mu, z \in \mathcal{O}.
\end{aligned}$$

Clearly $\theta_{ij}(y_i,y_j)\lambda_{ij}(y_i,y_j) \ge \theta_{ij}(y_i,y_j)\beta_{ij}(y_i,y_j)$ is equivalent to the following:

$$\begin{aligned}
\lambda_{ij}(y_i,y_j) & \le \frac{(z_{ij}+\mu_{ij}(y_i,y_j)) - (z_{ij}-\mu_{ij}(y_i,y_j))^2}{2} \quad \text{if} \quad \theta_{ij}(y_i,y_j) < 0, \\
\lambda_{ij}(y_i,y_j) & \ge \frac{(z_{ij}+\mu_{ij}(y_i,y_j))^2 - (z_{ij}+\mu_{ij}(y_i,y_j))}{2} \quad \text{if} \quad \theta_{ij}(y_i,y_j) \ge 0.
\end{aligned}$$

The proof is complete. ■

With Lemma 1, now we can prove Proposition 3.

Proof According to the constraints $\lambda_{ij}(y_i,y_j) \ge \max\{0, z_{ij}+\mu_{ij}(y_i,y_j)-1\}$ in problem (3), we can get $\lambda_{ij}(y_i,y_j) \ge \frac{(z_{ij}+\mu_{ij}(y_i,y_j))^2-(z_{ij}+\mu_{ij}(y_i,y_j))}{2}$.

According to the constraints $\lambda_{ij}(y_i,y_j) \le \min\{\mu_{ij}(y_i,y_j), z_{ij}\}$. we have

$$\begin{aligned}
& |z_{ij}-\mu_{ij}(y_i,y_j)| = |(z_{ij}-\lambda_{ij}(y_i,y_j)) - (\mu_{ij}(y_i,y_j) - \lambda_{ij}(y_i,y_j))|, \\
\Longrightarrow \quad & |z_{ij}-\mu_{ij}(y_i,y_j)| \le (z_{ij}-\lambda_{ij}(y_i,y_j)) + (\mu_{ij}(y_i,y_j) - \lambda_{ij}(y_i,y_j)), \\
\Longrightarrow \quad & \lambda_{ij}(y_i,y_j) \le \frac{z_{ij}+\mu_{ij}(y_i,y_j) - |z_{ij}-\mu_{ij}(y_i,y_j)|}{2}.
\end{aligned}$$

Since $|z_{ij}-\mu_{ij}(y_i,y_j)| \le 1$, $\frac{z_{ij}+\mu_{ij}(y_i,y_j)-|z_{ij}-\mu_{ij}(y_i,y_j)|}{2} \le \frac{(z_{ij}+\mu_{ij}(y_i,y_j))-(z_{ij}-\mu_{ij}(y_i,y_j))^2}{2}$. It follows that $\lambda_{ij}(y_i,y_j) \le \frac{(z_{ij}+\mu_{ij}(y_i,y_j))-(z_{ij}-\mu_{ij}(y_i,y_j))^2}{2}$.

In summary, from the constriants of problem (3), we can get

$$\frac{(z_{ij}+\mu_{ij}(y_i,y_j)) - (z_{ij}-\mu_{ij}(y_i,y_j))^2}{2} \ge \lambda_{ij}(y_i,y_j) \ge \frac{(z_{ij}+\mu_{ij}(y_i,y_j))^2 - (z_{ij}+\mu_{ij}(y_i,y_j))}{2},$$

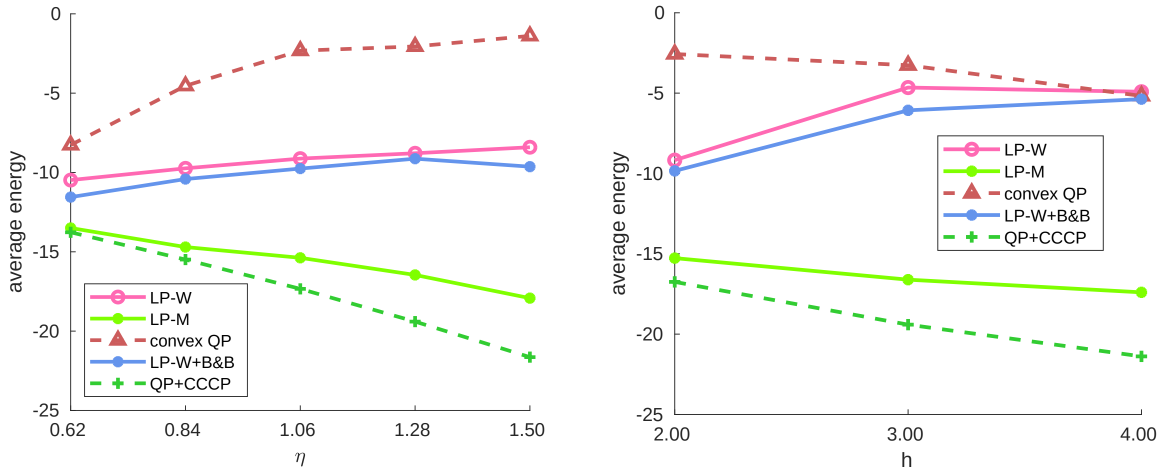

Figure 1. Estimated energy (the objective in (1)) as the increase of the coupling strength parameter η (left diagram) and the degree parameter h (right diagram). Lower is better.

consequently the constraints in the convex QP relaxation (14) using Lemma 1. Hence LP-W is tighter than the convex QP relaxation. With Proposition 1, we know that LP-M is tighter than LP-W, thus the proof is complete. ■

5. The QP+CCCP Algorithm

Algorithm 1 QP+CCCP

Require: node set V, potentials $\theta_i(y_i), \forall i \in V$, $\theta_{ij}(y_i, y_j), \forall i < j, y_i, y_j$, ϵ. t_{max}

Output: estimated labels Y^* and graph structure Z^*.

Solve convex QP (14), let $(\boldsymbol{\mu}^{(0)}, \mathbf{z}^{(0)})$ be the solution. $\chi^{(0)} \leftarrow [\boldsymbol{\mu}^{(0)}, \mathbf{z}^{(0)}]$.
Repeat solving convex QP (14) till $F(\chi^{(t-1)}) - F(\chi^{(t)}) < \epsilon$ or $t \geq t_{max}$.
Decode $Y^* \leftarrow \{y_i^*\}$: $y_i^* \leftarrow \operatorname{argmin}_{y_i} \mu_{ii}^{(t)}(y_i, y_i)$.
Decode $Z^* \leftarrow \{z_{ij}^*\}$: $z_{ij}^* \leftarrow 1$ if $z_{ij}^{(t)} \geq 0.5$, otherwise $z_{ij}^* \leftarrow 0$.
Return Y^*, Z^*.

6. Additional Experimental Results and Interpretations

To further show the benefits of the proposed LP-W and QP+CCCP algorithms, here we provide more experimental results.

6.1. Estimated Energy as η and h Changing

Note η is the coupling strength parameter used to create synthetic data (see Section 5 in the main text), and h controls the maximum degree of the estimated graph. For each η or h, we create synthetic data (20 examples) using the Potts distance. We report average energy of these 20 examples in Figure 1. We can see that our QP+CCCP always performs best with various h and η parameters among all methods. Lan performs second best among the methods that estimate the graph repetitively (including QP+CCCP, LP-W+B&B and Lan). Among methods that estimate the graph only once (including QP, LP-W and LP-M), the proposed LP-M performs best. LP-M even outperforms LP-W+B&B for all cases, and outperforms Lan on small η, despite the latter two estimating the graph repetitively.

6.2. Visualization of The Learned Parameters for Human Interaction Recognition

In order to model human interactions, within our MRF model, we include the relative spatial positions of one anchored person to another, which is typically utilized as a high-level contextual cue in human interaction recognition task.

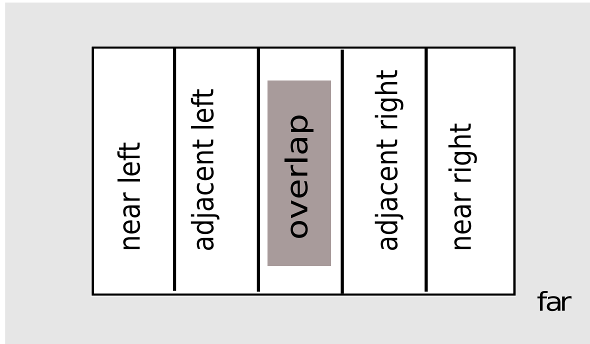

Figure 2. The discretized relative spatial positions of one anchored person (the shaded rectangle) to another (each cell with white background).

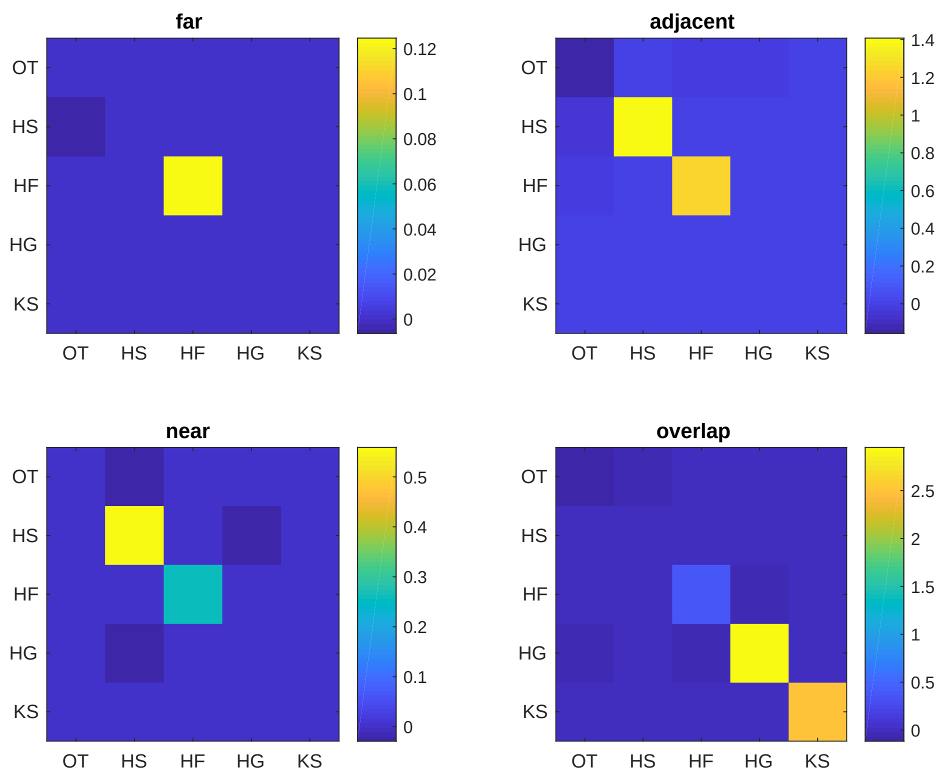

Figure 3. Visualization of the learned parameters for human interaction recognition on TVHI. Here OT, HS, HF, HG, KS means no-interaction, handshake, highfive, hug and kiss respectively. See text below for interpretation.

To illustrate the effectiveness of the learned model for human interaction recognition, we visualize the parameters that encodes the confidence (brighter indicates more confident) of assigning a particular pair of action labels to two individuals when they have interactions with other, see Figure 3. We can conclude that when persons are nearly overlapping with respect to their relative distance, it is confident for our model to recognize their interaction as hugging or kissing (the *overlap* diagram). When persons are adjacent or near to each other, our model is likely to recognize their interaction as handshake or highfive (the *adjacent* and the *near* diagrams). When people are far alway from each other, our model favors the highfive class (the *far* diagram), because the training set contains highfive-examples performed by distant persons.

6.3. Visualization of Group Activity Recognition Results on CAD

The results are shown in Figure 4.

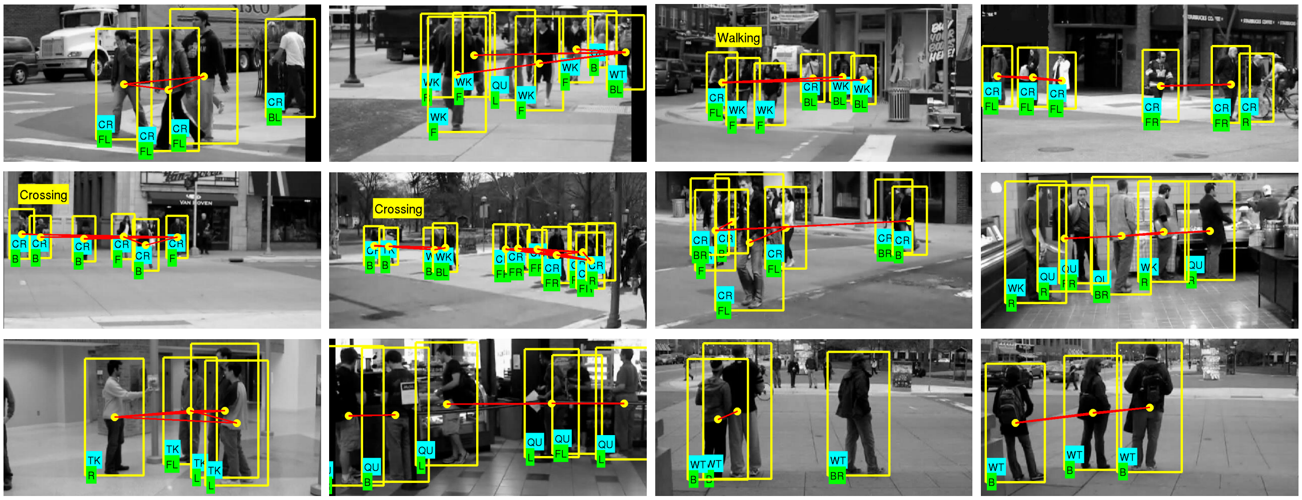

Figure 4. Visualisation of recognition results on CAD using our QP+CCCP algorithm. The predicted action and pose labels are shown in cyan and green boxes. The red edges represent the learned graph structures within the action layer. For action names, *CR, WK, QU, WT* indicate *cross, walk, queue* and *wait*. For poses, *B, L, R, F, BL, BR, FR, FL* denote *back, left, right, front, back-left, back-right, front-right* and *front-left* respectively.

6.4. Visualization of Human Interaction Recognition Results on TVHI

The results are shown in Figure 5.

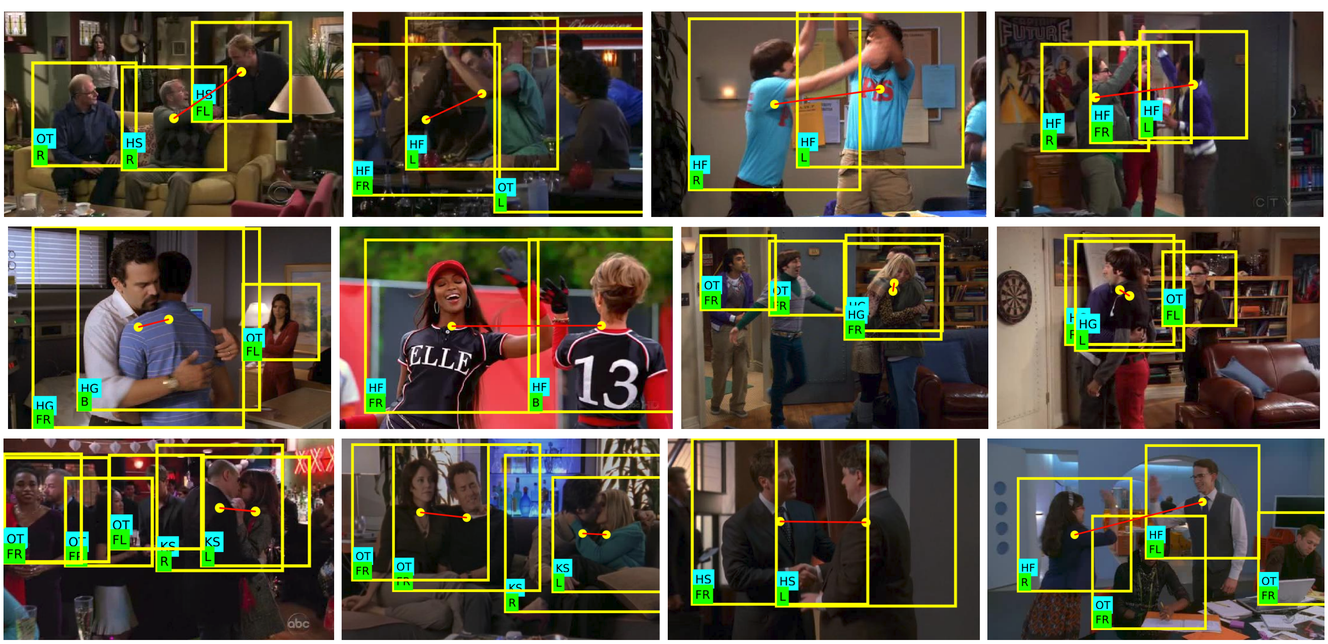

Figure 5. Visualisation of recognition results on TVHI using our QP+CCCP algorithm. The predicted action and pose labels are shown in cyan and green boxes. The red edges represent the learned graph structures. For interaction names, *OT, HS, HF, HG, KS* indicate *no-interaction, handshake, highfive, hug* and *kiss*. For poses, *B, L, R, FL, FR* denote *backwards, left, right, frontal-left, frontal-right* respectively.