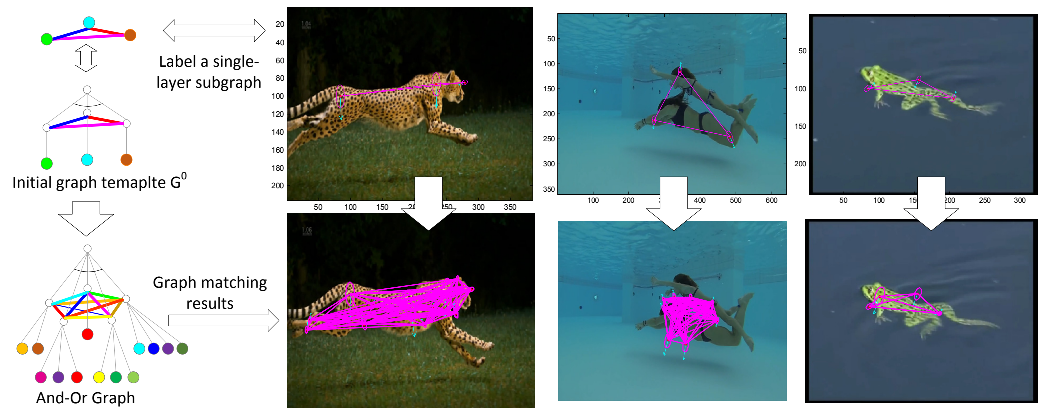
Mining And-Or Graphs for Graph Matching and Object Discovery

Supplementary Materials

1. Initial graph templates G^0 and matching results based on the final AoGs



The initial graph templates for the cheetah, swimming girls, and the frog in Experiment 4 were labeled in their first video frames. The initial graph template G^0 simply consisted of three OR nodes, each having a single terminal node. During the mining process, more OR nodes for the AoG were automatically discovered.

2. Proof of Operation 1, attribute estimation

We define $\Delta = \sum_{s\in V}(\mathcal{E}_s^- - \mathcal{E}_s^+ - \lambda|\Omega_s|)$. Thus, **Obj.**(b) is transformed to $(\mathbf{F}_V, \mathbf{F}_E) \leftarrow \text{argmax}_{\mathbf{F}_V,\mathbf{F}_E}\Delta$.

$$\frac{\partial\Delta}{\partial F_i^{\psi_s}} = \frac{\partial\mathcal{E}_s^-}{\partial F_i^{\psi_s}} - \frac{\partial\mathcal{E}_s^+}{\partial F_i^{\psi_s}} \qquad (\divideontimes.1)$$

Unlike the matches to positive ARGs that can be converged to true latent subgraph patterns within the ARGs, the matches to negative ARGs usually have no regulations. Therefore, theoretically, the average attributes among the matched nodes $\{\check{x}_s^l\}$ in the negative ARGs should be converged to the pattern's attributes, *i.e.* $\lim_{N^-\to+\infty} \text{mean}_{\substack{1\leq l\leq N^-:\\ \delta(\check{x}_s^l)=1,\check{\psi}_s^l=\psi_s}} \mathcal{F}_i^{\check{x}_s^l} = F_i^{\psi_s}$. Thus, we get

$$\lim_{N^-\to+\infty}\frac{\partial\mathcal{E}_s^-}{\partial F_i^{\psi_s}} = 2w_i^u\big(F_i^{\psi_s} - \lim_{N^-\to+\infty}\text{mean}_{\substack{1\leq l\leq N^-:\\ \delta(\check{x}_s^l)=1,\check{\psi}_s^l=\psi_s}}\mathcal{F}_i^{\check{x}_s^l}\big) = 0$$

where function $\delta(x)$ returns 0, if $x = none$; otherwise 1. In this case, the above derivative is independent with $F_i^{\psi_s}$[1]. Thus, we can make an approximation $\frac{\partial\mathcal{E}_s^-}{\partial F_i^{\psi_s}} \equiv 0$ to simplify the computation.

$$\frac{\partial\Delta}{\partial F_i^{\psi_s}} \approx -\frac{\partial\mathcal{E}_s^+}{\partial F_i^{\psi_s}} = -2w_i^u\big(F_i^{\psi_s} - \text{mean}_{k:\delta(\hat{x}_s^k)=1,\hat{\psi}_s^k=\psi_s}\mathcal{F}_i^{\hat{x}_s^k}\big)$$

Therefore, we can estimate attributes as follows.

$$F_i^{\psi_s} \leftarrow \text{mean}_{k:\delta(\hat{x}_s^k)=1,\hat{\psi}_s^k=\psi_s}\mathcal{F}_i^{\hat{x}_s^k}, \quad F_j^{st} \leftarrow \text{mean}_{k:\delta(\hat{x}_s^k)\delta(\hat{x}_t^k)=1}\mathcal{F}_j^{\hat{x}_s^k\hat{x}_t^k} \qquad (\divideontimes.2)$$

[1] In ($\divideontimes$.1), pattern attribute $F_i^{\psi_s}$ can determine the matching assignments $\{\check{x}_s^l\}$, which in return counteracts the effects of $F_i^{\psi_s}$.

3. Proof of Operation 3, object discovery

Given the current pattern G, **Objs.**(a–b) *w.r.t.* a potential new node y can be re-written as

$$\textbf{(a)}\;\; (\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}) \leftarrow \underset{\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}}{\operatorname{argmin}} \mathcal{E}_y^+,\;\; (\{\check{x}_y^l\},\{\check{\psi}_y^l\}) \leftarrow \underset{\{\check{x}_y^l\},\{\check{\psi}_y^l\}}{\operatorname{argmin}} \mathcal{E}_y^-$$

$$\textbf{(b)}\;\; (\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}) \leftarrow \max_{\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}} (\mathcal{E}_y^- - \mathcal{E}_y^+ - \lambda|\Omega_y|) \qquad (\divideontimes.3)$$

As the new node y should be well matched to most of the positive ARGs in Λ^+, we ignore the small possibility of y being matched to *none* to simplify the calculation. In other words, the range of matching assignments $\hat{x}_y^k$ is limited to $\mathcal{V}_k^+ \setminus \hat{\mathbf{x}}^k$ (*i.e.* $\delta(\hat{x}_y^k) \equiv 1$). Similar to ($\divideontimes$.2), we obtain

$$F_i^{\psi_y} \leftarrow \underset{1\leq k\leq N^+:\hat{\psi}_y^k=\psi_y}{\operatorname{mean}} \mathcal{F}_i^{\hat{x}_y^k}, \quad F_j^{yt} \leftarrow \underset{k:\delta(\hat{x}_t^k)=1}{\operatorname{mean}} \mathcal{F}_j^{\hat{x}_y^k \hat{x}_t^k} \qquad (\divideontimes.4)$$

3.1. formulation of $\mathcal{E}_y^+$

Based on the equation above, we can use $\{\hat{x}_y^k\}$ to represent $\{F_i^{\psi_y}\}, \{F_j^{yt}\}$ and reformulate $\mathcal{E}_y^+$ and $\mathcal{E}_y^-$.

$$\mathcal{E}_y^+ = \underset{1\leq k\leq N^+}{\operatorname{mean}} \Big[\mathcal{E}(\hat{\psi}_y^k \mapsto \hat{x}_y^k) + \sum_{t\in V} \mathcal{E}(\langle y,t\rangle \mapsto \langle \hat{x}_y^k, \hat{x}_t^k\rangle)\Big] = \frac{1}{N^+}\sum_{k=1}^{N^+} \Big\{\mathcal{E}(\hat{\psi}_y^k \mapsto \hat{x}_y^k) + \sum_{t\in V} \mathcal{E}(\langle y,t\rangle \mapsto \langle \hat{x}_y^k, \hat{x}_t^k\rangle)\Big\} \qquad (\divideontimes.5)$$

As mentioned in the manuscript, considering that new node y should be well matched to most positive ARGs, we just provide an approximate solution that ignores the possibility of matching y to *none* in this calculation, *i.e.* $\delta(\hat{x}_y^k) \equiv 1$. Therefore, we get

$$\begin{aligned}
\mathcal{E}(\hat{\psi}_y^k \mapsto \hat{x}_y^k) &= \begin{cases} \sum_{i=1}^{N_u} w_i^u \|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2, & \delta(\hat{x}_y^k)=1 \\ u_{none}, & \delta(\hat{x}_y^k)=0 \end{cases} \\
&= \sum_{i=1}^{N_u} w_i^u \|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2 \\
\mathcal{E}(\langle y,t\rangle \mapsto \langle \hat{x}_y^k, \hat{x}_t^k\rangle) &= \begin{cases} \frac{\sum_{j=1}^{N_p} w_j^p \|F_j^{yt} - \mathcal{F}_j^{\hat{x}_y^k \hat{x}_t^k}\|^2}{|V|}, & \delta(\hat{x}_y^k)\delta(\hat{x}_t^k)=1 \\ \frac{p_{none}}{|V|}, & \delta(\hat{x}_y^k)\delta(\hat{x}_t^k)=0 \end{cases} \\
&= \begin{cases} \frac{\sum_{j=1}^{N_p} w_j^p \|F_j^{yt} - \mathcal{F}_j^{\hat{x}_y^k \hat{x}_t^k}\|^2}{|V|}, & \delta(\hat{x}_t^k)=1 \\ \frac{p_{none}}{|V|}, & \delta(\hat{x}_t^k)=0 \end{cases}
\end{aligned}$$

Because y is a new node, y has a total of $|V|$ edges, and we normalize the pairwise energy $\mathcal{E}(\langle y,t\rangle \mapsto \langle \hat{x}_y^k, \hat{x}_t^k\rangle)$ by using $|V|$, rather than $|V|-1$.

We substitute these to ($\divideontimes$.5) and obtain

$$\begin{aligned}
\mathcal{E}_y^+ &= \frac{1}{N^+}\sum_{k=1}^{N^+} \Big\{\mathcal{E}(\hat{\psi}_y^k \mapsto \hat{x}_y^k) + \sum_{t\in V} \mathcal{E}(\langle y,t\rangle \mapsto \langle \hat{x}_y^k, \hat{x}_t^k\rangle)\Big\} \\
&= \frac{1}{N^+}\sum_{k=1}^{N^+} \Big\{ \Big[\sum_{i=1}^{N_u} w_i^u \|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2\Big] + \frac{1}{|V|}\Big[\sum_{\substack{t\in V\\ \delta(\hat{x}_t^k)=1}} \sum_{i=1}^{N_p} w_i^p \|F_i^{yt} - \mathcal{F}_i^{\hat{x}_y^k \hat{x}_t^k}\|^2 + \sum_{\substack{t\in V\\ \delta(\hat{x}_t^k)=0}} p_{none}\Big]\Big\} \qquad (\divideontimes.6) \\
&= \Big\{\frac{1}{N^+}\sum_{i=1}^{N_u} w_i^u \sum_{k=1}^{N^+} \|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2\Big\} + \Big\{\frac{1}{|V|(N^+)}\sum_{i=1}^{N_p} w_i^p \sum_{t\in V} \sum_{\substack{1\leq k\leq N^+:\\ \delta(\hat{x}_t^k)=1}} \|F_i^{yt} - \mathcal{F}_i^{\hat{x}_y^k \hat{x}_t^k}\|^2\Big\} + \Big\{\frac{1}{|V|(N^+)}\sum_{t\in V} \sum_{\substack{1\leq k\leq N^+:\\ \delta(\hat{x}_t^k)=0}} p_{none}\Big\}
\end{aligned}$$

Then, we focus on the terms of $\sum_{k=1}^{N^+}\|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2$ and $\sum_{\substack{1\le k\le N^+:\\ \delta(\hat{x}_t^k)=1}}\|F_i^{yt} - \mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2$ in Equation (⋇.6). We substitute (⋇.4) to these terms, and it is easy to prove these identical equations:

$$
\begin{aligned}
\sum_{k=1}^{N^+}\|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2 &= \sum_{\psi_y\in\Omega_y}\sum_{\substack{1\le k\le N^+:\\ \hat{\psi}_y^k=\psi_y}}\|F_i^{\psi_y} - \mathcal{F}_i^{\hat{x}_y^k}\|^2 \\
&= \sum_{\psi_y\in\Omega_y}\sum_{\substack{1\le k_1\le N^+:\\ \hat{\psi}_y^{k_1}=\psi_y}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \frac{1}{\sum_j \eta(\hat{\psi}_y^j=\psi_y)}\sum_{\substack{1\le k_2\le N^+:\\ \hat{\psi}_y^{k_2}=\psi_y}}\mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2 \\
&= \sum_{\psi_y\in\Omega_y}\frac{1}{2\sum_j \eta(\hat{\psi}_y^j=\psi_y)}\sum_{\substack{1\le k_1\le N^+:\\ \hat{\psi}_y^{k_1}=\psi_y}}\sum_{\substack{1\le k_2\le N^+:\\ \hat{\psi}_y^{k_2}=\psi_y}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2 \\
\sum_{\substack{1\le k\le N^+:\\ \delta(\hat{x}_t^k)=1}}\|F_i^{yt} - \mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2 &= \sum_{\substack{1\le k_1\le N^+:\\ \delta(\hat{x}_t^{k_1})=1}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \frac{1}{\sum_j \delta(\hat{x}_t^j)}\sum_{\substack{1\le k_2\le N^+:\\ \delta(\hat{x}_t^{k_2})=1}}\mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2 \\
&= \frac{1}{2\sum_j \delta(\hat{x}_t^j)}\sum_{\substack{1\le k_1\le N^+:\\ \delta(\hat{x}_t^{k_1})=1}}\sum_{\substack{1\le k_2\le N^+:\\ \delta(\hat{x}_t^{k_2})=1}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2
\end{aligned}
$$

We then substitute the above equations back to Equation (⋇.6) and obtain[2]

$$
\begin{aligned}
\mathcal{E}_y^+ &= \Big\{\frac{1}{N^+}\sum_{i=1}^{N_u} w_i^u \sum_{k=1}^{N^+}\|F_i^{\hat{\psi}_y^k} - \mathcal{F}_i^{\hat{x}_y^k}\|^2\Big\} + \Big\{\frac{1}{|V|(N^+)}\sum_{i=1}^{N_p} w_i^p \sum_{t\in V}\sum_{\substack{1\le k\le N^+:\\ \delta(\hat{x}_t^k)=1}}\|F_i^{yt} - \mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2\Big\} + \Big\{\frac{1}{|V|(N^+)}\sum_{t\in V}\sum_{\substack{1\le k\le N^+:\\ \delta(\hat{x}_t^k)=0}} p_{none}\Big\} \\
&= \Big\{\frac{1}{N^+}\sum_{i=1}^{N_u} w_i^u \sum_{\psi_y\in\Omega_y}\frac{1}{2\sum_j \eta(\hat{\psi}_y^j=\psi_y)}\sum_{\substack{1\le k_1\le N^+:\\ \hat{\psi}_y^{k_1}=\psi_y}}\sum_{\substack{1\le k_2\le N^+:\\ \hat{\psi}_y^{k_2}=\psi_y}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2\Big\} + \Big\{\frac{1}{|V|(N^+)}\sum_{i=1}^{N_p} w_i^p \sum_{t\in V}\frac{1}{2\sum_j \delta(\hat{x}_t^j)} \\
&\quad\cdot\sum_{\substack{1\le k_1\le N^+:\\ \delta(\hat{x}_t^{k_1})=1}}\sum_{\substack{1\le k_2\le N^+:\\ \delta(\hat{x}_t^{k_2})=1}}\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2\Big\} + \Big\{\sum_{t\in V}\frac{p_{none}(N^+ - \sum_j \delta(\hat{x}_t^j))}{|V|(N^+)}\Big\} \\
&= \sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}\Big\{\sum_{\psi_y\in\Omega_y}\frac{\eta(\hat{\psi}_y^{k_1}=\psi_y)\eta(\hat{\psi}_y^{k_2}=\psi_y)\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)\sum_j \eta(\hat{\psi}_y^j=\psi_y)} + \sum_{\substack{t\in V:\\ \delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})=1}}\frac{\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2|V|(N^+)\sum_j \delta(\hat{x}_t^j)} \\
&\quad + \sum_{\substack{t\in V:\\ \delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})=0}}\frac{p_{none}(N^+ - \sum_j \delta(\hat{x}_t^j))}{|V|(N^+)[(N^+)^2 - (\sum_j \delta(\hat{x}_t^j))^2]}\Big\} \\
&= \sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}\Big\{\sum_{\psi_y\in\Omega_y}\frac{\eta(\hat{\psi}_y^{k_1}=\psi_y)\eta(\hat{\psi}_y^{k_2}=\psi_y)\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)\sum_j \eta(\hat{\psi}_y^j=\psi_y)} + \sum_{t\in V}\Big[\frac{(1-\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2}))p_{none}}{|V|(N^+)(N^+ + \sum_j \delta(\hat{x}_t^j))} \\
&\quad + \frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2|V|(N^+)\sum_j \delta(\hat{x}_t^j)}\Big]\Big\}
\end{aligned}
$$

Therefore, we obtain[3]

$$
\mathcal{E}_y^+ = \sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}(\mathcal{M}_{k_1k_2} + \mathcal{N}_{k_1k_2}), \quad \text{where } \mathcal{N}_{k_1k_2} = \sum_{\psi_y\in\Omega_s}\frac{\eta(\psi_y=\hat{\psi}_y^{k_1})\eta(\psi_y=\hat{\psi}_y^{k_2})\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)\sum_j \eta(\psi_y=\hat{\psi}_y^j)}
$$

$$
\mathcal{M}_{k_1k_2} = \sum_{t\in V}\Big\{\frac{[1-\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})]p_{none}}{|V|(N^+)[(N^+) + \sum_j \delta(\hat{x}_t^j)]} + \frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}} - \mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2|V|(N^+)\sum_j \delta(\hat{x}_t^j)}\Big\}
$$

[2] We use function $\delta(\cdot)$ to avoid putting undefined pairwise attribute $\mathcal{F}_i^{x,none}$ into computation.
[3] We use function $\delta(\cdot)$ to avoid putting undefined pairwise attribute $\mathcal{F}_i^{x,none}$ into computation.

3.2. formulation of $\mathcal{E}_y^-$

Considering the following equation

$$\|a - \underset{1\leq k\leq N}{\text{mean}}\, b_k\|^2 = \frac{1}{N}\sum_{1\leq k\leq N}\|a-b_k\|^2 - \frac{1}{2N^2}\sum_{1\leq k\leq N}\sum_{1\leq l\leq N}\|b_k - b_l\|^2$$

we can obtain

$$\begin{aligned}\mathcal{E}_y^- &= \underset{1\leq l\leq N^-}{\text{mean}}\Big[\mathcal{E}(\check{\psi}_y^l \mapsto \check{x}_y^l) + \sum_{t\in V}\mathcal{E}(\langle y,t\rangle \mapsto \langle \check{x}_y^l, \check{x}_t^l\rangle)\Big]\\ &= \frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\sum_{i=1}^{N_u} w_i^u \|\mathcal{F}_i^{\check{x}_y^l} - F_i^{\check{\psi}_y^l}\|^2 + \frac{1}{|V|}\sum_{t\in V}\sum_{i=1}^{N_p} w_i^p \|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l} - F_i^{yt}\|^2\Big\}\end{aligned}$$

Actually, when we match G to negative ARGs in Λ^-, some nodes in G would be matched to $none$ with constant penalties u_{none} and p_{none}. However, in order to develop a direct solution to (⋇.3) without node enumeration, we ignore the matches to $none$[4], *i.e.* $\delta(\check{x}_s^l), \delta(\check{x}_y^l) \equiv 1$. In this way, the average penalty for matching y to negative ARGs, $\mathcal{E}_y^-$, can be approximately transformed to be a function *w.r.t.* $\{\check{x}_y^l\}$ (this will be proved later).

Note that 1) the negative ARG $\mathcal{G}_l^-$ does not contain the target subgraph pattern G, and 2) local variations of the terminal nodes within an OR node are usually much smaller than the variations between terminal nodes of different OR nodes. In most cases, there is usually no strong trend toward assigning a node in $\mathcal{G}_l^-$ to any particular terminal $\check{\psi}_y^l$ of the OR node y. Therefore, we can apply the following approximation: instead of assigning the matched node $\check{x}_y^l$ in $\mathcal{G}_l^-$ with a specific terminal $\check{\psi}_y^l$, we can try all the terminals to fit $\check{x}_y^l$.

$$\begin{aligned}\|\mathcal{F}_i^{\check{x}_y^l} - F_i^{\check{\psi}_y^l}\|^2 &\approx \sum_{\psi_y\in\Omega_y} Probability(\psi_y)\|\mathcal{F}_i^{\check{x}_y^l} - F_i^{\psi_y}\|^2\\ &= \sum_{\psi_y\in\Omega_y}\frac{\sum_j \eta(\hat{\psi}_y^j = \psi_y)}{N^+}\|\mathcal{F}_i^{\check{x}_y^l} - F_i^{\psi_y}\|^2\end{aligned}$$

where we use the distribution of terminal assignments in positive ARGs to estimate $Probability(\psi_y)$ for negative ARGs.

[4] We tentatively set u_{none} and p_{none} to be infinite to avoid matching to $none$.

Then, we substitute the approximation above and ($\circledast$.4) to $\mathcal{E}_y^-$ and get[5]

$$
\begin{aligned}
\mathcal{E}_y^- \approx & \frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\sum_{i=1}^{N_u} w_i^u \sum_{\psi_y\in\Omega_y}\frac{\sum_j \eta(\hat{\psi}_y^j=\psi_y)}{N^+}\|\mathcal{F}_i^{\check{x}_y^l}-F_i^{\psi_y}\|^2+\frac{1}{|V|}\sum_{t\in V}\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\underset{k:\delta(\hat{x}_t^k)=1}{\text{mean}}\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2\Big\}\\
= & \frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\sum_{i=1}^{N_u} w_i^u \sum_{\psi_y\in\Omega_y}\frac{\sum_j \eta(\hat{\psi}_y^j=\psi_y)}{N^+}\|\mathcal{F}_i^{\check{x}_y^l}-\underset{k:\hat{\psi}_y^k=\psi_y}{\text{mean}}\mathcal{F}_i^{\hat{x}_y^k}\|^2+\frac{1}{|V|}\sum_{t\in V}\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\underset{k:\delta(\hat{x}_t^k)=1}{\text{mean}}\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2\Big\}\\
= & \frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\sum_{i=1}^{N_u} w_i^u \sum_{\psi_y\in\Omega_y}\frac{\sum_j \eta(\hat{\psi}_y^j=\psi_y)}{N^+}\Big[\frac{\sum_{k:\hat{\psi}_y^k=\psi_y}\|\mathcal{F}_i^{\check{x}_y^l}-\mathcal{F}_i^{\hat{x}_y^k}\|^2}{\sum_j \eta(\hat{\psi}_y^j=\psi_y)}-\frac{\sum_{k_1:\hat{\psi}_y^{k_1}=\psi_y}\sum_{k_2:\hat{\psi}_y^{k_2}=\psi_y}\|\mathcal{F}_i^{\hat{x}_y^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2[\sum_j \eta(\hat{\psi}_y^j=\psi_y)]^2}\Big]\\
& +\frac{1}{|V|}\sum_{t\in V}\sum_{i=1}^{N_p} w_i^p\Big[\frac{1}{\sum_j\delta(\hat{x}_t^j)}\sum_{k:\delta(\hat{x}_t^k)=1}\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2-\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}\frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2[\sum_j\delta(\hat{x}_t^j)]^2}\Big]\Big\}\\
= & \sum_{k=1}^{N^+}\sum_{l=1}^{N^-}\Big\{\frac{\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\check{x}_y^l}-\mathcal{F}_i^{\hat{x}_y^k}\|^2}{(N^-)(N^+)}+\sum_{t\in V}\frac{\sum_{i=1}^{N_p} w_i^p\delta(\hat{x}_t^k)\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2}{|V|(N^-)\sum_j\delta(\hat{x}_t^j)}\Big\}-\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}\Big\{\\
& \sum_{\psi_y\in\Omega_y}\frac{\eta(\hat{\psi}_y^{k_1}=\psi_y)\eta(\hat{\psi}_y^{k_2}=\psi_y)\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)\sum_j\eta(\hat{\psi}_y^j=\psi_y)}+\sum_{t\in V}\frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2|V|[\sum_j\delta(\hat{x}_t^j)]^2}\Big\}\\
= & \sum_{k=1}^{N^+}\sum_{l=1}^{N^-}\mathcal{U}_{kl}-\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}(\mathcal{V}_{k_1k_2}+\mathcal{N}_{k_1k_2})
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{U}_{kl} = & \frac{\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\check{x}_y^l}-\mathcal{F}_i^{\hat{x}_y^k}\|^2}{(N^-)(N^+)}+\sum_{t\in V}\frac{\sum_{i=1}^{N_p} w_i^p\delta(\hat{x}_t^k)\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2}{|V|(N^-)\sum_j\delta(\hat{x}_t^j)}\\
\mathcal{V}_{k_1k_2} = & \sum_{t\in V}\frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})\sum_{i=1}^{N_p} w_i^p\|\mathcal{F}_i^{\hat{x}_y^{k_1}\hat{x}_t^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}\hat{x}_t^{k_2}}\|^2}{2|V|[\sum_j\delta(\hat{x}_t^j)]^2}\\
\mathcal{N}_{k_1k_2} = & \sum_{\psi_y\in\Omega_y}\frac{\eta(\hat{\psi}_y^{k_1}=\psi_y)\eta(\hat{\psi}_y^{k_2}=\psi_y)\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)\sum_j\eta(\hat{\psi}_y^j=\psi_y)}
\end{aligned}
$$

As mentioned in the manuscript, in order to develop a direct solution to mining AOGs without node enumeration, we ignore the matches to *none*[6], *i.e.* $\delta(\check{x}_s^l),\delta(\check{x}_y^l)\equiv 1$, and thus represent the average penalty for matching y to negative ARGs, $\mathcal{E}_y^-$, as function *w.r.t.* $\{\check{x}_y^l\}$ in the equation above.

3.3. Solution to $\max_{\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}}(\mathcal{E}_y^- - \mathcal{E}_y^+ - \lambda|\Omega_y|)$

Because 1) $\mathcal{E}_y^+$, $\mathcal{E}_y^-$, $\{F_i^{\psi_y}|1\leq i\leq N_u,\psi_y\in\Omega_y\}$, and $\{F_j^{yt}|1\leq j\leq N_p,t\in V\}$ are all represented as functions *w.r.t.* $\{\hat{x}_y^k|1\leq k\leq N^+\}$, and 2) $\mathcal{E}_y^-$ can be approximately formulated as a function independent to $\{\check{\psi}_y^l|1\leq l\leq N^-\}$, we can

[5]We use function $\delta(\cdot)$ to avoid putting undefined pairwise attribute $\mathcal{F}_i^{x,none}$ into computation.

[6]We tentatively set u_{none} and p_{none} to be infinite to avoid matching to $none$.

re-write (⁎.3) as

$$\begin{aligned}
&\max_{\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}}\big(\min_{\{\check{x}_y^l\},\{\check{\psi}_y^l\}}\mathcal{E}_y^- - \min_{\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}}\mathcal{E}_y^+ - \lambda|\Omega_y|\big)\\
=&\min_{\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}}\big(\min_{\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}}\mathcal{E}_y^+ - \min_{\{\check{x}_y^l\},\{\check{\psi}_y^l\}}\mathcal{E}_y^- + \lambda|\Omega_y|\big)\\
=&\min_{\Omega_y,\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}}\big(\mathcal{E}_y^+ - \min_{\{\check{x}_y^l\}}\mathcal{E}_y^- + \lambda|\Omega_y|\big)\\
=&\min_{\{\hat{x}_y^k\}}\Big\{-\sum_{l=1}^{N^-}\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\mathcal{U}_{kl} + \sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}(\mathcal{M}_{k_1k_2}+\mathcal{V}_{k_1k_2}) + \min_{\Omega_y,\{\hat{\psi}_y^k\}}\Big[\big(\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}2\mathcal{N}_{k_1k_2}\big)+\lambda|\Omega_y|\Big]\Big\} \qquad (⁎.7)
\end{aligned}$$

In the equation above, we need to determine four kinds of parameters, *i.e.* $\{\hat{x}_y^k\}$, $\{\check{x}_y^l\}$, $\{\hat{\psi}_y^k\}$, and Ω_y. The problem of energy minimization requires a huge amount of computation to enumerate all parameter values, which proposes great challenges to state-of-the-art optimization techniques.

Fortunately, our algorithm of model mining is performed iteratively, and inaccurate model parameters (including attributes, the node set, the terminal set for each node, and matching parameters) can be continuously refined in later iterations. Therefore, in this paper, we simply propose an efficient but approximate solution to node discovery and let the inaccuracy led by the approximations be corrected in later iterations, as follows.

We first approximate the values of $\{\check{x}_y^l\}$, $\{\hat{\psi}_y^k\}$, and Ω_y via local optimizations, so that we can provide a rough estimation of $\{\hat{x}_y^k\}$. Then, we use the values of $\{\hat{x}_y^k\}$ to further determine other parameters, such as Ω_y, $\{F_i^{\psi_y}\}$, and $\{F_j^{yt}\}$.

First, we focus on the term of $\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\mathcal{U}_{kl}$ in (⁎.7). According to our scenario of model mining, the model pattern should be contained and only contained by positive ARGs. In other words, we do not take two random sets of ARGs as the positive and negative ARGs, and mine meaningless patterns from such chaotic ARGs. We assume that the true patterns should exist physically. Therefore, as long as the mining process is not significantly biased, the attributes in the positive ARGs should be much more converged than those in the negative ARGs, *i.e.* $\underset{1\leq k\leq N^+}{variation}(\mathcal{F}_i^{\hat{x}_y^k}) \ll \underset{1\leq l\leq N^-}{variation}(\mathcal{F}_i^{\check{x}_y^l})$ and $\underset{1\leq k\leq N^+,\delta(\hat{x}_t^k)=1}{variation}(\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}) \ll \underset{1\leq l\leq N^-,\delta(\check{x}_t^l)=1}{variation}(\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l})$. In this case, we get

$$\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\|\mathcal{F}_i^{\check{x}_y^l}-\mathcal{F}_i^{\hat{x}_y^k}\|^2 \approx \sum_{k=1}^{N^+}\min_{\check{x}_y^l}\|\mathcal{F}_i^{\check{x}_y^l}-\mathcal{F}_i^{\hat{x}_y^k}\|^2$$
$$\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\delta(\hat{x}_t^k)\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2 \approx \sum_{k=1}^{N^+}\min_{\check{x}_y^l}\delta(\hat{x}_t^k)\|\mathcal{F}_i^{\check{x}_y^l\check{x}_t^l}-\mathcal{F}_i^{\hat{x}_y^k\hat{x}_t^k}\|^2$$

Using the equation above, it is easy to prove

$$\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\mathcal{U}_{kl} \approx \sum_{k=1}^{N^+}\min_{\check{x}_y^l}\mathcal{U}_{kl} \qquad (⁎.8)$$

Second, we focus on the term of $\min_{\Omega_y,\{\hat{\psi}_y^k\}}\Big[\big(\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}2\mathcal{N}_{k_1k_2}\big)+\lambda|\Omega_y|\Big]$ in (⁎.7). In order to simplify the computation, we tentatively set node y with a single terminal (*i.e.* $|\Omega_y|=1$) and produce a rough estimation of $\{\hat{x}_y^k\}$. Then, given $\{\hat{x}_y^k\}$, the true set of terminals of node y (Ω_y), as well as their assignments in matching ($\{\hat{\psi}_y^k\}$), will be determined later. Thus, the objective function for node discovery in (⋇.3) is transformed from (⋇.7) to

$$\max_{\Omega_y,\{F_i^{\psi_y}\},\{F_j^{yt}\}} \Big(\min_{\{\check{x}_y^l\},\{\check{\psi}_y^l\}} \mathcal{E}_y^- - \min_{\{\hat{x}_y^k\},\{\hat{\psi}_y^k\}} \mathcal{E}_y^+ - \lambda|\Omega_y| \Big)$$

$$=\min_{\{\hat{x}_y^k\}}\Big\{-\sum_{l=1}^{N^-}\min_{\check{x}_y^l}\sum_{k=1}^{N^+}\mathcal{U}_{kl}+\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}(\mathcal{M}_{k_1k_2}+\mathcal{V}_{k_1k_2})+\min_{\Omega_y,\{\hat{\psi}_y^k\}}\Big[\Big(\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}2\mathcal{N}_{k_1k_2}\Big)+\lambda|\Omega_y|\Big]\Big\} \tag{⋇.9}$$

$$\approx\min_{\{\hat{x}_y^k\}}\Big\{-\sum_{l=1}^{N^-}\sum_{k=1}^{N^+}\min_{\check{x}_y^l}\mathcal{U}_{kl}+\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}(\mathcal{M}_{k_1k_2}+\mathcal{V}_{k_1k_2}+2\mathcal{N}'_{k_1k_2})+\lambda\Big\}$$

where $\mathcal{N}'_{k_1k_2}=\frac{\sum_{i=1}^{N_u} w_i^u\|\mathcal{F}_i^{\hat{x}_y^{k_1}}-\mathcal{F}_i^{\hat{x}_y^{k_2}}\|^2}{2(N^+)^2}$

Therefore, we can rewrite the above conclusion as

$$\min_{\{\hat{x}_y^k\}}\Big\{\sum_{k=1}^{N^+}\Phi(\hat{x}_y^k)+\sum_{k_1=1}^{N^+}\sum_{k_2=1}^{N^+}\Phi(\hat{x}_y^{k_1},\hat{x}_y^{k_2})\Big\}$$

$$\Phi(\hat{x}_y^k)=-\sum_{l=1}^{N^-}\min_{\check{x}_y^l}\Big[\sum_{t\in V}\frac{\delta(\hat{x}_t^k)\mathcal{E}(\langle\hat{x}_y^k,\hat{x}_t^k\rangle\mapsto\langle\check{x}_y^l,\check{x}_t^l\rangle)}{|V|(N^-)\sum_j\delta(\hat{x}_t^j)}+\frac{\mathcal{E}(\hat{x}_y^k\mapsto\check{x}_y^l)}{(N^-)(N^+)}\Big]$$

$$\Phi(\hat{x}_y^{k_1},\hat{x}_y^{k_2})=\frac{\mathcal{E}(\hat{x}_y^{k_1}\mapsto\hat{x}_y^{k_2})}{(N^+)^2}+\sum_{t\in V}\Big(\frac{[1-\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})]p_{none}}{|V|(N^+)[(N^+)+\sum_j\delta(\hat{x}_t^j)]}+\frac{\delta(\hat{x}_t^{k_1})\delta(\hat{x}_t^{k_2})(\sum_j\delta(\hat{x}_t^j)+N^+)}{2|V|(N^+)[\sum_j\delta(\hat{x}_t^j)]^2}\mathcal{E}(\langle\hat{x}_y^{k_1},\hat{x}_t^{k_1}\rangle\mapsto\langle\hat{x}_y^{k_2},\hat{x}_t^{k_2}\rangle)\Big)$$

where $\delta(x)$ returns 0, if $x=none$; otherwise, it returns 1.

In the equation above, the range of matching assignments $\check{x}_y^l$ and $\hat{x}_y^k$ are the unmatched nodes in $\mathcal{G}_l^-$ and $\mathcal{G}_k^+$, respectively, *i.e.* $\check{x}_y^l\in\mathcal{V}_l^-\setminus\check{\mathbf{x}}^l$ and $\hat{x}_y^k\in\mathcal{V}_k^+\setminus\hat{\mathbf{x}}^k$. Parameters $\{\check{x}_y^l\}$ are determined in a local manner, and the estimation of $\mathbf{x}$ can be regarded as an energy minimization of an MRF *w.r.t* $\{\hat{x}_y^k\}$, which can be directly solved via global optimizations. In this study, we use the TRW-S [3] to compute $\{\hat{x}_y^k\}$.

4. Proof of Operation 4, terminal determination

Given matching assignments $\{\hat{x}_s^k\}$ of each OR node s, this operation determines the optimal number of terminals for node s and their matching assignments $\{\hat{\psi}_s^k\}$. Just as (⋇.3), requirements in **Objs.**(a–b) for each node s can be re-written as

$$\textbf{(a)}\ \ (\{\hat{x}_s^k\},\{\hat{\psi}_s^k\})\leftarrow \underset{\{\hat{x}_s^k\},\{\hat{\psi}_s^k\}}{\text{argmin}}\,\mathcal{E}_s^+,\ \ (\{\check{x}_s^l\},\{\check{\psi}_s^l\})\leftarrow \underset{\{\check{x}_s^l\},\{\check{\psi}_s^l\}}{\text{argmin}}\,\mathcal{E}_s^-$$

$$\textbf{(b)}\ \ (\Omega_s,\{F_i^{\psi_s}\},\{F_j^{st}\})\leftarrow \max_{\Omega_s,\{F_i^{\psi_s}\},\{F_j^{st}\}}(\mathcal{E}_s^- - \mathcal{E}_s^+ - \lambda|\Omega_s|) \qquad (⋇.10)$$

Different from the formulation of $\mathcal{E}_y^+$, $\mathcal{E}_s^+$ should be formulated in a more general form that does not ignore the probability of matching node s to $none$, as follows[7].

$$\begin{aligned}
\mathcal{E}_s^+ &= \underset{1\le k\le N^+}{\text{mean}}\Big[\mathcal{E}(\hat{\psi}_s^k\mapsto\hat{x}_s^k)+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\hat{x}_s^k,\hat{x}_t^k\rangle)\Big]\\
&=\frac{1}{N^+}\sum_{k=1}^{N^+}\Big\{\delta(\hat{x}_s^k)\Big[\sum_{i=1}^{N_u}w_i^u\|F_i^{\hat{\psi}_s^k}-\mathcal{F}_i^{\hat{x}_s^k}\|^2\Big]+\Big[1-\delta(\hat{x}_s^k)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\hat{x}_s^k,\hat{x}_t^k\rangle)\Big\}\\
&=\Big\{\frac{1}{N^+}\sum_{i=1}^{N_u}w_i^u\sum_{\substack{1\le k\le N^+:\\ \delta(\hat{x}_s^k)=1}}\|F_i^{\hat{\psi}_s^k}-\mathcal{F}_i^{\hat{x}_s^k}\|^2\Big\}+\frac{1}{N^+}\sum_{k=1}^{N^+}\Big\{\Big[1-\delta(\hat{x}_s^k)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\hat{x}_s^k,\hat{x}_t^k\rangle)\Big\}\\
&=\frac{1}{N^+}\sum_{i=1}^{N_u}w_i^u\Big\{\sum_{\psi_s\in\Omega_s}\sum_{\substack{1\le k_1\le N^+:\\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}}\|\mathcal{F}_i^{\hat{x}_s^{k_1}}-\underset{\substack{1\le k_2\le N^+:\\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}}{\text{mean}}\mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2\Big\}+\mathcal{C}_s^+\\
\text{where}\quad \mathcal{C}_s^+ &=\frac{1}{N^+}\sum_{k=1}^{N^+}\Big\{\Big[1-\delta(\hat{x}_s^k)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\hat{x}_s^k,\hat{x}_t^k\rangle)\Big\}
\end{aligned}$$

Then, we focus on the formulation of $\mathcal{E}_s^-$.

$$\begin{aligned}
\mathcal{E}_s^- &= \underset{1\le l\le N^-}{\text{mean}}\Big[\mathcal{E}(\check{\psi}_s^l\mapsto\check{x}_s^l)+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\check{x}_s^l,\check{x}_t^l\rangle)\Big]\\
&=\frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\delta(\check{x}_s^l)\Big[\sum_{i=1}^{N_u}w_i^u\|F_i^{\check{\psi}_s^l}-\mathcal{F}_i^{\check{x}_s^l}\|^2\Big]+\Big[1-\delta(\check{x}_s^l)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\check{x}_s^l,\check{x}_t^l\rangle)\Big\}\\
&=\Big\{\frac{1}{N^-}\sum_{i=1}^{N_u}w_i^u\sum_{\substack{1\le l\le N^-:\\ \delta(\check{x}_s^l)=1}}\|F_i^{\check{\psi}_s^l}-\mathcal{F}_i^{\check{x}_s^l}\|^2\Big\}+\frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\Big[1-\delta(\check{x}_s^l)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\check{x}_s^l,\check{x}_t^l\rangle)\Big\}\\
&=\Big\{\frac{1}{N^-}\sum_{i=1}^{N_u}w_i^u\sum_{\substack{1\le l\le N^-:\\ \delta(\check{x}_s^l)=1}}\|F_i^{\check{\psi}_s^l}-\mathcal{F}_i^{\check{x}_s^l}\|^2\Big\}+\mathcal{C}_s^-\\
\text{where}\quad \mathcal{C}_s^- &=\frac{1}{N^-}\sum_{l=1}^{N^-}\Big\{\Big[1-\delta(\check{x}_s^l)\Big]u_{none}+\sum_{t\in V}\mathcal{E}(\langle s,t\rangle\mapsto\langle\check{x}_s^l,\check{x}_t^l\rangle)\Big\}
\end{aligned}$$

Considering 1) negative ARGs $\{\mathcal{G}_l^-\}$ do not contain the target subgraph pattern G, and 2) local variations of the terminal nodes within an OR node are usually much smaller than the variations between terminal nodes of different OR nodes, we apply the same approximation that is used to formulate $\mathcal{E}_s^-$. In most cases, there is usually no strong regulations toward assigning a node in $\mathcal{G}_l^-$ to any particular terminal $\check{\psi}_s^l$ in a pattern node s in G. Therefore, we can apply the following

[7] We use function $\delta(\cdot)$ to avoid putting undefined pairwise attribute $\mathcal{F}_i^{none}$ into computation.

approximation: instead of assigning the matched node $\check{x}_s^l$ in $\mathcal{G}_l^-$ with a specific terminal $\check{\psi}_s^l$, we can try all the terminals to fit $\check{x}_s^l$:

$$\begin{aligned}\|\mathcal{F}_i^{\check{x}_s^l} - F_i^{\check{\psi}_s^l}\|^2 &\approx \sum_{\psi_s\in\Omega_s} Probability(\psi_s)\|\mathcal{F}_i^{\check{x}_s^l} - F_i^{\psi_s}\|^2 \\ &= \sum_{\psi_s\in\Omega_s} \frac{\sum_j \eta(\hat{\psi}_s^j = \psi_s)\delta(\hat{x}_s^j)}{\sum_j \delta(\hat{x}_s^j)}\|\mathcal{F}_i^{\check{x}_s^l} - F_i^{\psi_s}\|^2\end{aligned}$$

Thus, we get

$$\mathcal{E}_s^- \approx \frac{1}{N^-}\sum_{\psi_s\in\Omega_s}\frac{\sum_j \eta(\hat{\psi}_s^j = \psi_s)\delta(\hat{x}_s^j)}{\sum_j \delta(\hat{x}_s^j)}\sum_{\substack{1\le l\le N^-:\\ \delta(\check{x}_s^l)=1}}\sum_{i=1}^{N_u} w_i^u\|F_i^{\psi_s} - \mathcal{F}_i^{\check{x}_s^l}\|^2 + \mathcal{C}_s^-$$

We substitute ($\circledast$.2) to the equation above. Then, considering the following equation

$$\|a - \underset{1\le k\le N}{\text{mean}}\, b_k\|^2 = \frac{1}{N}\sum_{1\le k\le N}\|a-b_k\|^2 - \frac{1}{2N^2}\sum_{1\le k\le N}\sum_{1\le l\le N}\|b_k - b_l\|^2$$

$$\sum_{1\le k_1\le N}\|a_{k_1} - \underset{1\le k_2\le N}{\text{mean}}\, a_{k_2}\|^2 = \frac{1}{2N}\sum_{1\le k_1\le N}\sum_{1\le k_2\le N}\|a_{k_1} - a_{k_2}\|^2$$

we obtain

$$\begin{aligned}\mathcal{E}_s^- \approx &\frac{1}{N^-}\sum_{\psi_s\in\Omega_s}\frac{\sum_j \eta(\hat{\psi}_s^j = \psi_s)\delta(\hat{x}_s^j)}{\sum_j \delta(\hat{x}_s^j)}\sum_{\substack{1\le l\le N^-:\\ \delta(\check{x}_s^l)=1}}\sum_{i=1}^{N_u} w_i^u\Big\{\sum_{\substack{1\le k\le N^+:\\ \eta(\hat{\psi}_s^k=\psi_s)\delta(\hat{x}_s^k)=1}}\frac{\|\mathcal{F}_i^{\hat{x}_s^k} - \mathcal{F}_i^{\check{x}_s^l}\|^2}{\sum_j \eta(\hat{\psi}_s^j = \psi_s)\delta(\hat{x}_s^j)} \\ &- \sum_{\substack{1\le k_1\le N^+:\\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}}\sum_{\substack{1\le k_2\le N^+:\\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}}\frac{\|\mathcal{F}_i^{\hat{x}_s^{k_1}} - \mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2}{2[\sum_j \eta(\hat{\psi}_s^j = \psi_s)\delta(\hat{x}_s^j)]^2}\Big\} + \mathcal{C}_s^- \\ = &-\frac{\sum_j \delta(\check{x}_s^j)}{(N^-)\sum_j \delta(\hat{x}_s^j)}\sum_{i=1}^{N_u} w_i^u\sum_{\psi_s\in\Omega_s}\sum_{\substack{1\le k_1\le N^+:\\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}}\|\mathcal{F}_i^{\hat{x}_s^{k_1}} - \underset{\substack{1\le k_2\le N^+:\\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}}{\text{mean}}\mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2 + \mathcal{D}_s^- + \mathcal{C}_s^- \\ &\text{where}\quad \mathcal{D}_s^- = \frac{1}{N^-}\sum_{i=1}^{N_u} w_i^u\sum_{\substack{1\le l\le N^-:\\ \delta(\check{x}_s^l)=1}}\underset{\substack{1\le k\le N^+:\\ \delta(\hat{x}_s^k)=1}}{\text{mean}}\|\mathcal{F}_i^{\hat{x}_s^k} - \mathcal{F}_i^{\check{x}_s^l}\|^2\end{aligned}$$

Therefore we can get

$$\begin{aligned}&\underset{\Omega_s,\{F_i^{\psi_s}\}}{\text{argmax}}\,(\mathcal{E}_s^- - \mathcal{E}_s^+ - \lambda|\Omega_s|) \\ =&\underset{\Omega_s,\{F_i^{\psi_s}\}}{\text{argmin}}\,(\mathcal{E}_s^+ - \mathcal{E}_s^- + \lambda|\Omega_s|) \\ \approx&\underset{\Omega_s,\{F_i^{\psi_s}\}}{\text{argmin}}\Big\{\Big[\frac{1}{N^+} + \frac{\sum_j \delta(\check{x}_s^j)}{(N^-)\sum_j \delta(\hat{x}_s^j)}\Big]\sum_{i=1}^{N_u} w_i^u\sum_{\psi_s\in\Omega_s}\sum_{\substack{1\le k_1\le N^+:\\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}}\|\mathcal{F}_i^{\hat{x}_s^{k_1}} - \underset{\substack{1\le k_2\le N^+:\\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}}{\text{mean}}\mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2 + \mathcal{C}_s^+ - \mathcal{C}_s^- - \mathcal{D}_s^- + \lambda|\Omega_s|\Big\}\end{aligned}$$

Considering that $\mathcal{C}_s^+, \mathcal{C}_s^-$, and $\mathcal{D}_s^-$ are all independent with Ω_s and $\{F_i^{\psi_s}\}$, we get

$$\begin{aligned}&\underset{\Omega_s,\{F_i^{\psi_s}\}}{\text{argmax}}\,(\mathcal{E}_s^- - \mathcal{E}_s^+ - \lambda|\Omega_s|) \\ \approx&\underset{\Omega_s,\{F_i^{\psi_s}\}}{\text{argmin}}\Big\{\Big[\frac{1}{N^+} + \frac{\sum_j \delta(\check{x}_s^j)}{(N^-)\sum_j \delta(\hat{x}_s^j)}\Big]\sum_{i=1}^{N_u} w_i^u\sum_{\psi_s\in\Omega_s}\sum_{\substack{1\le k_1\le N^+:\\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}}\|\mathcal{F}_i^{\hat{x}_s^{k_1}} - \underset{\substack{1\le k_2\le N^+:\\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}}{\text{mean}}\mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2 + \lambda|\Omega_s|\Big\} \qquad (\circledast.11)\end{aligned}$$

Note that in the equation above, $\sum_{i=1}^{N_u} w_i^u \sum_{\psi_s \in \Omega_s} \sum_{\substack{1 \leq k_1 \leq N^+: \\ \eta(\hat{\psi}_s^{k_1}=\psi_s)\delta(\hat{x}_s^{k_1})=1}} \|\mathcal{F}_i^{\hat{x}_s^{k_1}} - \text{mean}_{\substack{1 \leq k_2 \leq N^+: \\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}} \mathcal{F}_i^{\hat{x}_s^{k_2}}\|^2$ can be regarded as the energy function for clustering, where a set of points $\{\hat{x}_s^k | 1 \leq k \leq N^+, \delta(\hat{x}_s^k) = 1\}$ are grouped into $|\Omega_s|$ clusters. The feature for each point $\hat{x}_s^k$ is $\mathbf{f}^{\hat{x}_s^k} = [\sqrt{w_1^u}(\mathcal{F}_1^{\hat{x}_s^k})^T, \sqrt{w_2^u}(\mathcal{F}_2^{\hat{x}_s^k})^T, \ldots, \sqrt{w_{N_u}^u}(\mathcal{F}_{N_u}^{\hat{x}_s^k})^T]^T$. The center of each cluster $\psi_s \in \Omega_s$ is $\sum_{\substack{1 \leq k_2 \leq N^+: \\ \eta(\hat{\psi}_s^{k_2}=\psi_s)\delta(\hat{x}_s^{k_2})=1}} \mathcal{F}_i^{\hat{x}_s^{k_2}}$.

$$\min_{\Omega_s, \{\hat{\psi}_s^k\}} \Big\{ C \sum_{\psi_s \in \Omega_s} \sum_{\substack{1 \leq k_1 \leq N^+: \\ \psi_s \mapsto \hat{x}_s^{k_1}, \delta(\hat{x}_s^{k_1})=1}} \|\mathbf{f}^{\hat{x}_s^{k_1}} - \text{mean}_{\substack{1 \leq k_2 \leq N^+: \\ \psi_s \mapsto \hat{x}_s^{k_2}, \delta(\hat{x}_s^{k_2})=1}} \mathbf{f}^{\hat{x}_s^{k_2}}\|^2 + \lambda |\Omega_s| \Big\}$$

where $C = 1/N^+ + \sum_j \delta(\check{x}_s^j)/[(N^-)\sum_j \delta(\hat{x}_s^j)]$.

Therefore, we can use a hierarchical clustering to solve ($\divideontimes$.11). In the beginning, each point in $\{\hat{x}_s^k\}$ is initialized as a cluster. In each following step, we merge two nearest clusters and use the cluster center to represent the new cluster. We keep merging the clusters, until the energy in ($\divideontimes$.11) is minimized. Thus, we can set terminal attributes $\{F_i^{\psi_s} | 1 \leq i \leq N_u\}$ as the cluster centers and assign different nodes $\{\hat{x}_s^k\}$ to different terminals.

5. Sub-objectives of Objs.(a–d)

Sub-objectives of **Objs.**(a–d) present an approximate solution to the overall objective of graph mining. The overall objective is defined in (9) as

$$G^* = \underset{G}{\text{argmax}} \log \Big\{ \frac{P(\Lambda^+|G)}{P(\Lambda^-|G)} \cdot e^{-\lambda \text{Complexity}(G)} \Big\} = \underset{G}{\text{argmin}} \{ \underbrace{\sum_{s \in V} (\mathcal{E}_s^+ - \mathcal{E}_s^-)}_{\text{generative loss; we hope } \mathcal{E}_s^+ \ll \mathcal{E}_s^-.} + \underbrace{\sum_{s \in V} \lambda |\Omega_s|}_{\text{complexity loss}} \}$$

where the generative loss corresponds the discriminative capability of the model, while the complexity loss controls the number of terminal nodes to avoid over-fitting.

Thus, we approximate this objective using **Objs.**(a–d). First, **Obj.**(a) is proposed according to the definition of $\mathcal{E}_s^+$ and $\mathcal{E}_s^-$. **Obj.**(a) computes the optimal matching assignments of the current AoG G, and $\mathcal{E}_s^+$ and $\mathcal{E}_s^-$ are defined on the basis of these matching assignments.

Obj.(a): $\underset{\hat{\mathbf{x}}^k, \hat{\Psi}^k}{\text{argmin}} \mathcal{E}(G \overset{\hat{\Psi}^k, \hat{\mathbf{x}}^k}{\longmapsto} \mathcal{G}_k^+)$, $\underset{\check{\mathbf{x}}^l, \check{\Psi}^l}{\text{argmin}} \mathcal{E}(G \overset{\check{\Psi}^l, \check{\mathbf{x}}^l}{\longmapsto} \mathcal{G}_l^-)$

Second, **Objs.**(b–c) are straightforward based on the above overall objective.

Obj.(b): $\text{argmin}_{\mathbf{\Omega}, \mathbf{F}_V, \mathbf{F}_E} \sum_{s \in V} (\mathcal{E}_s^+ - \mathcal{E}_s^- + \lambda |\Omega_s|)$

Obj.(c): $\text{argmax} |V|$ s.t. $\forall s \in V, \mathcal{E}_s^+ - \mathcal{E}_s^- + \lambda |\Omega_s| \leq \tau$

In **Obj.**(c), we use a threshold τ to control the discovery and elimination of OR nodes. Based on **Obj.**(c), we can grow the AoG to contain a maximum number of OR nodes. We use the threshold τ to approximate the overall objective.

In addition, we can understand **Obj.**(c) from another perspective. Without loss of generality, if we reformulate the model complexity as

$$\text{Complexity}(G) = |\mathbf{\Omega}| + \beta |V| = \sum_{s \in V} (|\Omega_s| + \beta)$$

where V is the set of OR nodes. Then, the overall objective corresponds to

$$G^* = \underset{G}{\text{argmax}} \log \Big\{ \frac{P(\Lambda^+|G)}{P(\Lambda^-|G)} \cdot e^{-\lambda \text{Complexity}(G)} \Big\} = \underset{G}{\text{argmin}} \{ \underbrace{\sum_{s \in V} (\mathcal{E}_s^+ - \mathcal{E}_s^-)}_{\text{generative loss; we hope } \mathcal{E}_s^+ \ll \mathcal{E}_s^-.} + \underbrace{\sum_{s \in V} \lambda (|\Omega_s| + \beta)}_{\text{complexity loss}} \}$$

Theoretically, this new objective does not change the formulations of **Objs.**(a,b,d) and the six operations for graph mining. Whereas, in terms of **Obj.**(c), it requires that

$$\text{argmax} |V| \quad \text{s.t. } \forall s \in V, \mathcal{E}_s^+ - \mathcal{E}_s^- + \lambda |\Omega_s| + \lambda \beta \leq 0$$

If we set $\tau = -\lambda\beta$, this equation is just equivalent to **Obj.**(c).

In fact, both the original objective in (9) and the above modified objective can explain **Objs.**(a–d) and the six operations for graph mining. We would like to use the original one in (9) in our article for clarity. In this case, **Obj.**(c) presents an approximate solution to (9).

Finally, we use a linear SVM to train the matching parameters $\mathbf{W}$, which is an approximate solution to the minimization of the generative loss.

Obj.(d): $\min_{\mathbf{W}} \|\mathbf{w}\|^2 + \frac{C}{N^+}\sum_{k=1}^{N^+} \xi_k^+ + \frac{C}{N^-}\sum_{l=1}^{N^-} \xi_l^-$,

$$\forall k = 1, 2, ..., N^+, -[\mathcal{E}(G \overset{\hat{\Psi}^k, \hat{\mathbf{x}}^k}{\longmapsto} \mathcal{G}_k^+) + b] \geq 1 - \xi_k^+, \quad \forall l = 1, 2, ..., N^-, \mathcal{E}(G \overset{\check{\Psi}^l, \check{\mathbf{x}}^l}{\longmapsto} \mathcal{G}_l^-) + b \geq 1 - \xi_l^-$$

6. About Figures 6, 7, and 8

Competing methods of *MA*, *MS*, *MT*, *LS*, *LT* cannot modify the pattern size. Thus, they simply correspond to five dots in Figure 7. The *SR* method can only delete redundant nodes from the pattern without the ability of discovering new nodes. Therefore, its corresponding curves in Figures 6 and 7 are shorter than curves of *SAP* and our method.

Note that it is not fair to directly compare the matching performance of two graph patterns produced by different methods, if the graph patterns do not have the same size. It is because 1) too small patterns may be lack of important nodes (object parts) and have bad matching performance; 2) large patterns may contain some correct but not so distinguishing node (object parts), which decreases the matching accuracy. For example, pedals are necessary parts in bicycles, but they are not powerful in object matching. Thus, if the bicycle pattern contains the pedals, this pattern will have relatively bad matching performance, although containing pedals makes this pattern more complete.

Therefore, in our experiments, we regarded the matching performance (such as the average precision and the error matching rate) as a function of the pattern size[4]. In Figures 6, 7, and 8, we compare the matching performance between the patterns with similar sizes.

7. About comparison between supervised/unsupervised learning for graph matching and graph mining

In fact, the main idea of this section has been introduced in the related-work section in the article. Here, we provide more details of the comparisons between supervised and unsupervised approaches for learning graph matching and graph-mining methods.

Methods for learning graph matching usually train parameters or refine the template to achieve stable matching performance. The key difference between supervised methods [2, 1, 4, 8] and unsupervised methods [5, 9] for learning graph matching is whether the method requires people to manually label the matching assignments between for training. From this perspective, unsupervised methods [5, 9] are more related to the spirit of knowledge mining.

Given a graph template (initial graph pattern), [5] learns the matching parameters, while [9] mainly refines the error parts in the graph pattern. Therefore, essentially, these methods are designed for improving matching performance, rather than recover object models from object fragments by discovering new object parts. In other words, their objectives are not to discover category models from unlabeled data. If people do not label all the possible object parts, these methods cannot automatically recover the whole object shape. In contrast, the core task of graph mining is to automatically discover new pattern nodes from the ARGs, so as to grow the pattern to a maximum size. The direct discovery of new nodes without node enumeration raises the biggest challenge of graph mining.

[4]For single-layer graph models used in baselines, this is the total node number. For our hierarchical AoG, this is the number of OR nodes.

8. About the general forms of graph matching

Different from graph-mining techniques for "labeled graphs", the graph-mining problem defined in graph domain of visual ARGs should be formulated on the basis of graph matching.

Actually, graph matching has two typical forms. One is the maximization of a compatibility function applied by [5, 1, 4]. The other is the minimization of an energy function, like [10, 6, 8] and ours. The first from of graph matching aims to maximize the average compatibility between each pair of corresponding attributes. Let us take a unary (or pairwise) attribute $f_{template}$ in the graph template G and its corresponding unary (or pairwise) attribute f_{graph} in the target ARG, for example. Their matching compatibility is usually defined as $\exp(-\|f_{template} - f_{graph}\|)$, $\exp(-\|f_{template} - f_{graph}\|^2)$, or $-\|f_{template} - f_{graph}\|^2$. Then, for the second form, the matching energy is mainly defined as $\|f_{template} - f_{graph}\|^2$ just like (3) in this study.

Actually, these two forms of graph matching is intrinsically equivalent to each other, to some extent. All of them are typical quadratic assignment problems, and can be solved by similar matching optimization techniques [7].

9. Computational cost

We now briefly analyze the computational cost of our method. The main computational task of graph mining is the energy minimization involved in both graph matching in *Operation 1* and node discovery in (10) during the mining process.

First, we focus on computational cost of graph matching. In each iteration, we match the current pattern to all N^+ positive ARGs and N^- negative ARGs. Let n_k^+ and n_l^- denote the node number in the k-th positive ARG $\mathcal{G}_k^+$ and the node number in the l-th negative ARG $\mathcal{G}_l^-$, respectively. We use $G^0, G^1, ..., G^M$ to denote the AoGs that are produced after $0, 1, ..., M$ iterations, respectively, where M is the total iteration number of graph mining. $n^0, n^1, ..., n^M$ denote the number of the OR nodes in $G^0, G^1, ..., G^M$. The matching between an AoG G^m and an ARG $\mathcal{G}_k^+$, *i.e.* $G^m \xmapsto{\Psi,\mathbf{x}} \mathcal{G}_k^+$, can be computed as a QAP that assigns each of the n^m OR nodes (more precisely, one terminal of this OR node) to one of $n_k^+ + 1$ labels[8]. Thus, we simply use $c(n^m \rightarrow n_k^+ + 1)$ to denote the computational cost of this QAP. Obviously, larger values of n^m and n_k^+ will result in higher cost. Therefore, the computational cost of graph matching during the mining process can be formulated as $\sum_{m=0}^{M-1}[\sum_{k=1}^{N^+} c(n^m \rightarrow n_k^+ + 1) + \sum_{l=1}^{N^-} c(n^m \rightarrow n_l^- + 1)]$.

Then, let us consider the computation cost of node discovery. Its computational cost in each iteration $m+1$ (in which G^m is used as the current AoG) can be formulated as a QAP that assigns each of the N^+ positive ARGs with one of $\max_k\{n_k^+ - n^m\}$ labels (*i.e.* determining the node corresponding to the new OR node y in each positive ARG). Hence, its computational cost is $c(N^+ \rightarrow \max_k\{n_k^+ - n^m\})$. Thus, we can summarize the overall computational cost as $\sum_{m=0}^{M-1} c(N^+ \rightarrow \max_k\{n_k^+ - n^m\})$.

Therefore, the overall computational cost can be summarized as $\sum_{m=0}^{M-1}[c(N^+ \rightarrow \max_k\{n_k^+ - n^m\}) + \sum_{k=1}^{N^+} c(n^m \rightarrow n_k^+ + 1) + \sum_{l=1}^{N^-} c(n^m \rightarrow n_l^- + 1)]$.

In fact, many techniques can be applied to these QAPs, and each of them has its own accuracy and computational cost. Please refer to [7] for detailed computational costs of the QAP $c(\cdot)$ based on different optimization methods.

[8] Note that we should consider the dummy matching choice of *none*.

10. Evaluation of object discovery

[11] has partitioned the twenty-five categories in the dataset into five sub-groups, named *SIVAL1* to *SIVAL5*, and measured average clustering purity within each of these groups. The division of the five groups are as follows.

Group name		Category	Number of images
SIVAL1	c1:	ajaxorange	60
	c2:	checkeredscarf	60
	c3:	bluescrunge	60
	c4:	glazedwoodpot	60
	c5:	juliespot	60
SIVAL2	c1:	dirtyworkgloves	60
	c2:	greenteabox	60
	c3:	goldmedal	60
	c4:	smileyfacedoll	60
	c5:	spritecan	60
SIVAL3	c1:	cardboardbox	60
	c2:	feltflowerrug	60
	c3:	stripednotebook	60
	c4:	wd40can	60
	c5:	woodrollingpin	60
SIVAL4	c1:	apple	60
	c2:	candlewithholder	60
	c3:	fabricsoftenerbox	60
	c4:	rapbook	60
	c5:	translucentbowl	60
SIVAL5	c1:	banana	60
	c2:	cokecan	60
	c3:	dataminingbook	60
	c4:	dirtyrunningshoe	60
	c5:	largespoon	60

In Experiment 3, we mined the AoGs for the 25 categories by setting $\tau = 1.1$. The following table shows the detailed purity of each category, where $c1$–$c5$ correspond to the five categories in each group as shown above.

	c1	c2	c3	c4	c5	**Avg.**
SIVAL1	100.	82.0	87.3	98.5	77.3	89.0
SIVAL2	80.0	86.0	100.	100.	100.	93.2
SIVAL3	80.0	83.6	100.	96.7	81.8	88.4
SIVAL4	94.3	87.7	91.7	89.7	75.8	87.8
SIVAL5	86.7	96.8	92.9	98.4	88.5	92.7

11. Difference between our ICCV submission and "visual graph mining" attached behind this page

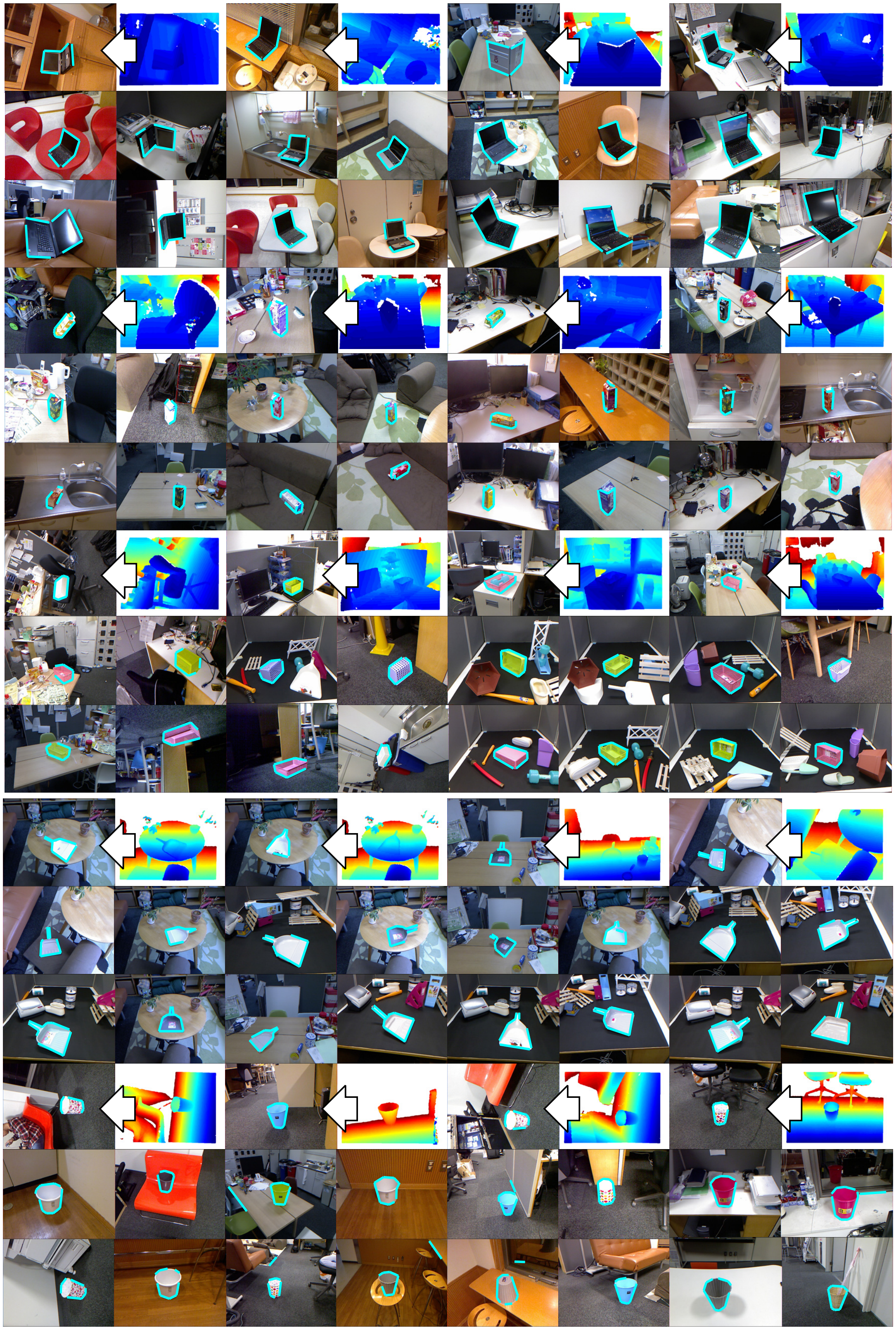
The paper of "visual graph mining" is a technical report in our reference. These two studies present different graph-mining techniques. In these papers, both the subgraph patterns and the objectives of graph mining are defined in two totally different ways:

1) In the ICCV article, we define a hierarchical And-Or Graph to represent the subgraph pattern, while, in the paper of "visual graph mining", the subgraph pattern is defined as a single-layer pattern. Obviously, the And-Or Graph has stronger expressive power than the single-layer pattern.

2) In the ICCV article, we use the energy gap between positive matches and negative matches (*i.e.* $\mathcal{E}_s^+ - \mathcal{E}_s^-$) as a generative loss to guide the mining of And-Or Graphs. This ensures that the And-Or Graph is exclusively contained by positive ARGs and not embedded in negative ARGs. However, the paper of "visual graph mining" does not has such design.

3) The techniques of graph mining are different. For example, compared to the single-layer pattern, the matching of the And-Or Graph should additionally consider the selection of terminal nodes. In our ICCV submission, we should discover both OR nodes and terminal nodes, which is more challenging.

12. Matching results on RGB-D images in Experiment 1



13. Matching results on RGB images in Experiment 3

14. Matching results on videos in Experiment 4

References

[1] T. S. Caetano, J. J. McAuley, L. Cheng, Q. V. Le, and A. J. Smola. Learning graph matching. *In* PAMI, 2009. 12, 13

[2] M. Cho, K. Alahari, and J. Ponce. Learning graphs to match. *In* ICCV, 2013. 12

[3] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *In* PAMI, 28(10):1568–1583, 2006. 8

[4] M. Leordeanu and M. Hebert. Smoothing-based optimization. *In* CVPR, 2008. 12, 13

[5] M. Leordeanu and M. Hebert. Unsupervised learning for graph matching. *In* CVPR, 2009. 12, 13

[6] A. Shekhovtsov, I. Kovtun, and V. Hlaváč. Efficient mrf deformation model for non-rigid image matching. *In* CVPR, 2007. 13

[7] 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*, 2006. 13

[8] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. *In* ECCV, 2008. 12, 13

[9] Q. Zhang, X. Song, X. Shao, H. Zhao, and R. Shibasaki. Learning graph matching for category modeling from cluttered scenes. *In* ICCV, 2013. 12

[10] Q. Zhang, X. Song, X. Shao, H. Zhao, and R. Shibasaki. Attributed graph mining and matching: An attempt to define and extract soft attributed patterns. *In* CVPR, 2014. 13

[11] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu. Unsupervised object class discovery via saliency-guided multiple class learning. *In* PAMI, 2014. 14