Proofs and Details for the paper "A Generative-Discriminative Hybrid Method for Multi-View Object Detection"

1 Proof of Theorem 1

Proof. We start from the posterior probability $p(X|O, H = 1)$. According to the Bayes rule

$$p(X|O, H = 1) = \frac{1}{C} p(O|X, H = 1)p(X|H = 1)$$

where C is the normalization term, which is the positive likelihood $p(O|H = 1)$:

$$C = p(O|H = 1) = \sum_X p(O|X, H = 1)P(X|H = 1) \tag{1}$$

Next, let us rewrite the posterior probability $p(X|O, H = 1)$ as the following

$$p(X|O, H = 1) = \frac{\prod_u f^+_{B_1}(y_u) \prod_{uv} f^+_{B_2}(y_{uv})}{CZ} \frac{p(O|X, H = 1)p(X|H = 1)Z}{\prod_u f^+_{B_1}(y_u) \prod_{uv} f^+_{B_2}(y_{uv})} \tag{2}$$

Using the independence assumption

$$p(O|X, H = 1) = \prod_{uv} p(y_{uv}|x_{1u}, x_{1v}, ..., x_{Nu}, x_{Nv}, H = 1) \prod_u p(y_u|x_{1u}, ..., x_{Nu}, H = 1)$$

and plugging in the parameter mapping equations (also in the page 4 section 2.2 of the main paper)

$$\begin{aligned} p(y_u|x_{11} = 0, ..., x_{iu} = 1, ..., H = 1) &= f_i(y_u) \\ p(y_{uv}|x_{11} = 0, ..., x_{iu} = 1, x_{jv} = 1, ..., H = 1) &= f_{ij}(y_{uv}) \end{aligned}$$

and

$$\begin{aligned} p(y_u|x_{11} = 0, x_{iu} = 0, ..., x_{NM} = 0, H = 1) &= f^+_{B_1}(y_u) \\ p(y_{uv}|x_{11} = 0, x_{iu} = 0, ..., x_{NM} = 0, H = 1) &= f^+_{B_2}(y_{uv}) \end{aligned}$$

Comparing to the term in the Gibbs distribution in Eq.(8)(main paper), we note for every matching X, we have

$$\frac{p(O|X,H=1)p(X|H=1)Z}{\prod_u f^+_{B_1}(y_u)\prod_{uv} f^+_{B_2}(y_{uv})} = \prod_{iu,jv} \varsigma_{iu,jv}(x_{iu},x_{jv}) \prod_{iu} \eta_{iu}(x_{iu})$$

Since the spaces of all matchings of $p(X|O,H=1)$ and the Gibbs distribution in Eq.(7)(main paper) are the same, the normalization constant should be also equal, i.e.

$$\frac{CZ}{\prod_u f^+_{B_1}(y_u)\prod_{uv} f^+_{B_2}(y_{uv})} = Z'$$

Therefore the positive likelihood is

$$p(O|H=1) = C = \frac{Z'}{Z}\prod_u f^+_{B_1}(y_u)\prod_{uv} f^+_{B_2} \tag{3}$$

and the likelihood ratio is

$$\frac{p(O|H=1)}{p(O|H=0)} = \frac{\prod_u f^+_{B_1}(y_u)\prod_{uv} f^+_{B_2}}{\prod_u f^-_{B_1}(y_u)\prod_{uv} f^-_{B_2}}\frac{Z'}{Z} = \sigma\frac{Z'}{Z} \tag{4}$$

It would be also easy to show that for the integer-based representation of matching, we would have the same result. □

2 Proof of Lemma 1

Proof. We first prove a special case when $N \le M$ for the unpruned MRF. We first enumerate the matchings where there are i nodes $n_{I_1}, n_{I_2}...n_{I_i}$ in the RARG being matched to the nodes in ARG, where $1 \le i \le N$,and $I_1, I_2...I_i$ is the index of the RARG node. The corresponding summation is

$$M(M-1)(M-2)...(M-i+1)z_{I_1}z_{I_2}...z_{I_i} = \binom{M}{i} i! z_{I_1}z_{I_2}...z_{I_i}$$

For all matchings where there are i nodes being matched to RARG, the summation becomes

$$\binom{M}{i} i! \sum_{1\le I_1<I_2<...<I_i\le N} z_{I_1}z_{I_2}...z_{I_i} = \binom{M}{i} i! \Pi_i(z_1,z_2,...,z_N)$$

Where

$$\Pi_i(z_1,z_2,...,z_N) = \sum_{1\le I_1<I_2<...<I_i\le N} z_{I_1}z_{I_2}...z_{I_i}$$

is a shorthand notation known as *Elementary Symmetric Polynomial*. Therefore Z can be written as the following form and satisfies the inequality

$$Z = \sum_{i=0}^{N} M(M-1)...(M-i+1)\Pi_i(z_1, z_2, ..., z_N) \leq \sum_{i=0}^{N} M^i \Pi_i(z_1, z_2, ..., z_N) \quad (5)$$

The equality holds when N/M tends to zero. And we have the following relationship

$$\sum_{i=0}^{N} \Pi_i(z_1, z_2, ..., z_N) = 1+z_1+z_2+...+z_N+z_1z_2+...+z_{N-1}z_N+... = \prod_{i=1}^{N}(1+z_i)$$

and

$$M^i \Pi_i(z_1, z_2, ..., z_N) = \Pi_i(Mz_1, Mz_2, ..., Mz_N)$$

Therefore, the RHS in equation 5 can be simplified as the following

$$\sum_{i=0}^{N} M^i \Pi_i(z_1, z_2, ..., z_N) = \sum_{i=0}^{N} \Pi_i(Mz_1, Mz_2, ..., Mz_N) = \prod_{i=1}^{N}(1+Mz_i)$$

The above function in fact is the partition function of the Gibbs distribution if we remove the one-to-one constraints. Likewise, for the general case and unpruned MRF, the partition function is upper-bounded by the partition function of the Gibbs distribution if we remove the one-to-one constraints, which for pruned MRF, by enumerating the matchings, can be written as

$$1 + d_1z_1 + d_2z_2 + ... + d_Nz_N + d_1d_2z_1z_2 + ... = \prod_{i=1}^{N}(1+d_iz_i)$$

where d_i is the number of the nodes in the ARG that are allowed to match to the node i in the RARG after pruning the Association Graph. Therefore we have

$$\ln Z \leq \prod_{i=1}^{N}(1+d_iz_i)$$

□

3 Proof of Lemma 2

Proof. The partition function can be calculated by enumerating the admissible matching (matching that does not violate the one-to-one constraint) as the following

$$Z(N; M; z_1, z_2, ..., z_N) = \sum_{X} \prod_{iu,jv} \psi_{iu,jv}(x_{iu}, x_{jv}) \prod_{iu} \phi_{iu}(x_{iu}) = \sum_{admissible\ X} \prod_{iu} z_i$$

where $z_i = \phi_{iu}(1)$ is defined in the main paper section 2.2. Therefore, the partition function is the summation of monomials whose variables have maximum power of 1. And the fact is true for both pruned and unpruned MRF. We can separate the above monomial summation into two polynomials, the polynomial containing z_i and the polynomial not

$$Z(N; M; z_1, z_2, ..., z_N) = V_1(z_1, z_2, ..., z_i, ...z_N) + V_2(z_1, z_2, ..., z_{i-1}, z_{i+1}...z_N)$$

Then the *occurrence probability* r_i, which is defined as the prior probability of the occurrence of the node i in the generated ARG, should be

$$r_i = \frac{V_1}{Z} = \frac{z_i \frac{\partial V_1}{\partial z_i}}{Z} = \frac{z_i \frac{\partial Z}{\partial z_i}}{Z} = z_i \frac{\partial \ln Z}{\partial z_i}$$

Where we have used the fact that V_1 is the summation of the monomials in the form of $z_{I_1} z_{I_2} ... z_{I_L}$, which has the following invariant property

$$z_{I_1} z_{I_2} ... z_{I_L} = z_{I_k} \frac{\partial}{\partial z_{I_k}} (z_{I_1} z_{I_2} ... z_{I_L}), \qquad \forall I_k \in \{I_1, I_2, ..., I_L\}$$

□

4 Proof of the Equation (10) in the main paper

We need to estimate the occurrence probability r_i. The overall likelihood for positive training data is defined as

$$L = \sum_{k=1}^{K} \ln p(O_k | H = 1) \tag{6}$$

where K is the number of the positive training instances. We have the variational approximation of the overall log-likelihood

$$L \approx \sum_{k=1}^{K} \sum_{iu} \hat{q}(x_{iu}^k = 1) \ln z_i - K \ln Z(N; M; z_1, z_2, ..., z_N) + \alpha \tag{7}$$

where α is a term independent on the *occurrence probability* $r_1, r_2, ..., r_N$. To maximize the approximated likelihood with respect to z_i, we compute the partial derivative of the Eq.(7) to z_i, and equates it to zero. With the help of lemma 2, we obtain

$$\frac{\partial L}{\partial z_i} = \frac{\partial}{\partial z_i}\Big[\sum_{k=1}^{K}\sum_{iu}\hat{q}(x_{iu}^k = 1)\ln z_i\Big] - K\frac{\partial}{\partial z_i}\ln Z(N; M; z_1, z_2, ..., z_N)$$
$$= \sum_{k=1}^{K}\sum_{iu}\hat{q}(x_{iu}^k = 1)\frac{1}{z_i} - K\frac{r_i}{z_i} = 0 \tag{8}$$

Since z_i is assumed to be non-zero, the above equation leads to the definition equation of r_i

$$r_i = \frac{1}{K}\sum_k\sum_u \hat{q}(x_{iu}^k = 1) \tag{9}$$

5 The Update Equations for the Gaussian Density Functions at Nodes and Edges of the RARG

Refered by the E-M step in Section 2.4 (main paper page 5), the E-M update equations are listed as the following:

$$\xi_{iu}^k = \hat{q}(x_{iu}^k = 1), \quad \xi_{iu,jv}^k = \hat{q}(x_{iu}^k = 1, x_{jv}^k = 1); \qquad \bar{\xi}_{iu}^k = 1 - \xi_{iu}^k, \quad \bar{\xi}_{iu,jv}^k = 1 - \xi_{iu,jv}^k$$
$$\mu_i = \frac{\sum_k\sum_u \xi_{iu}^k y_u^k}{\sum_k\sum_u \xi_{iu}^k} \qquad \Sigma_i = \frac{\sum_k\sum_u \xi_{iu}^k (y_u^k - \mu_i)(y_u^k - \mu_i)^T}{\sum_k\sum_u \xi_{iu}^k}$$
$$\mu_{ij} = \frac{\sum_k\sum_{uv} \xi_{iu,jv}^k y_{uv}^k}{\sum_k\sum_{uv} \xi_{iu,jv}^k} \qquad \Sigma_{ij} = \frac{\sum_k\sum_{uv} \xi_{iu,jv}^k (y_{uv}^k - \mu_{ij})(y_{uv}^k - \mu_{ij})^T}{\sum_k\sum_{uv} \xi_{iu,jv}^k}$$
$$\mu_{B_1}^+ = \frac{\sum_k\sum_u \bar{\xi}_{iu}^k y_u^k}{\sum_k\sum_u \bar{\xi}_{iu}^k} \qquad \Sigma_{B_1}^+ = \frac{\sum_k\sum_u \bar{\xi}_{iu}^k (y_u^k - \mu_i)(y_u^k - \mu_i)^T}{\sum_k\sum_u \bar{\xi}_{iu}^k}$$
$$\mu_{B_2}^+ = \frac{\sum_k\sum_{uv} \bar{\xi}_{iu,jv}^k y_{uv}^k}{\sum_k\sum_{uv} \bar{\xi}_{iu,jv}^k} \qquad \Sigma_{B_2}^+ = \frac{\sum_k\sum_{uv} \bar{\xi}_{iu,jv}^k (y_{uv}^k - \mu_{ij})(y_{uv}^k - \mu_{ij})^T}{\sum_k\sum_{uv} \bar{\xi}_{iu,jv}^k}$$

where k is the iteration index, and μ_i and Σ_i are mean and covariance matrix of the Gaussian R.V. at the nodes of the Random ARG, μ_{ij} and Σ_{ij} are mean and covariance matrix of the Gassian R.V. at the edges of the Random ARG. $\mu_{B_1}^+$,$\mu_{B_2}^+$,$\Sigma_{B_1}^+$,$\Sigma_{B_2}^+$ are the corresponding parameters of the Gaussian R.V. for modelling background.

The above equations are obtained by maximizing the variational approximation of the overall likelihood of the positive training data, as defined in equation (6).