Supplementary Material—Switchable Deep Network for Pedestrian Detection

Anonymous CVPR submission

Paper ID 938

1. Proof of Eq.(7) and Eq.(8)

The energy function of SRBM is formulated as

$$E(\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{m}) = -\sum_{k=1}^{K} s^k \mathbf{h}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{b}^k) - \sum_{k=1}^{K} s^k \mathbf{c}^{kT}(\mathbf{x}\circ\mathbf{m}^k) - \mathbf{y}^T\mathbf{U}\sum_{k=1}^{K} s^k\mathbf{h}^k - \mathbf{d}^T\mathbf{y}. \tag{1}$$

Substituting the energy function into the general joint distribution, we obtain the corresponding probability density function (PDF)

$$P(\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{m}) = \frac{1}{Z}\exp(-E(\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{m})) = \frac{1}{Z}\exp(\sum_{k=1}^{K} s^k \mathbf{h}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{b}^k) + \sum_{k=1}^{K} s^k \mathbf{c}^{kT}(\mathbf{x}\circ\mathbf{m}^k) + \mathbf{y}^T\mathbf{U}\sum_{k=1}^{K} s^k\mathbf{h}^k + \mathbf{d}^T\mathbf{y}). \tag{2}$$

For the convenience, we define the number of components, number of visible nodes, number of hidden nodes for each component, and the number of categories as K, N, M, and L, respectively.

1.1. Conditional Probability for Sampling Hidden Features

The conditional probability of the hidden features $\mathbf{h}^\iota$ of the ι-th component is

$$P(\mathbf{h}^\iota|\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}) = \frac{P(\mathbf{h}^\iota,\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m})}{P(\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m})} \tag{3}$$

$$= \frac{\sum_{\widetilde{\mathbf{h}}\backslash\mathbf{h}^\iota}^{\widetilde{\mathbf{H}}} P(\widetilde{\mathbf{h}},\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m})}{\sum_{\widetilde{\mathbf{h}}}^{\widetilde{\mathbf{H}}} P(\widetilde{\mathbf{h}},\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m})} \tag{4}$$

$$= \frac{\frac{1}{Z}\sum_{\widetilde{\mathbf{h}}\backslash\mathbf{h}^\iota}^{\widetilde{\mathbf{H}}} \exp(-E(\widetilde{\mathbf{h}},\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}))}{\frac{1}{Z}\sum_{\widetilde{\mathbf{h}}}^{\widetilde{\mathbf{H}}} \exp(-E(\widetilde{\mathbf{h}},\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}))} \tag{5}$$

$$= \frac{\exp(\sum_{k=1}^{K} s^k\mathbf{c}^{kT}(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{d}^T\mathbf{y})}{\exp(\sum_{k=1}^{K} s^k\mathbf{c}^{kT}(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{d}^T\mathbf{y})} \times \tag{6}$$

$$\frac{\sum_{\widetilde{\mathbf{h}}\backslash\mathbf{h}^\iota}^{\widetilde{\mathbf{H}}} \exp(\sum_{k=1}^{K} s^k\widetilde{\mathbf{h}}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{b}^k)+\mathbf{y}^T\mathbf{U}\sum_{k=1}^{K} s^k\widetilde{\mathbf{h}}^k)}{\sum_{\widetilde{\mathbf{h}}}^{\widetilde{\mathbf{H}}} \exp(\sum_{k=1}^{K} s^k\widetilde{\mathbf{h}}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{b}^k)+\mathbf{y}^T\mathbf{U}\sum_{k=1}^{K} s^k\widetilde{\mathbf{h}}^k)} \tag{7}$$

$$
\begin{aligned}
&= \frac{\sum_{\widetilde{\mathbf{h}}\backslash\mathbf{h}^\iota}^{\widetilde{\mathbf{H}}} \exp(\sum_{k=1}^{K}\sum_{j=1}^{M} s^k \widetilde{h}_j^k(\mathbf{w}_{j*}^k(\mathbf{x}\circ\mathbf{m}^k) + b_j^k + \mathbf{y}^T\mathbf{U}_{*j}))}{\sum_{\widetilde{\mathbf{h}}}^{\widetilde{\mathbf{H}}} \exp(\sum_{k=1}^{K}\sum_{j=1}^{M} s^k \widetilde{h}_j^k(\mathbf{w}_{j*}^k(\mathbf{x}\circ\mathbf{m}^k) + b_j^k + \mathbf{y}^T\mathbf{U}_{*j}))} && (8)\\
&= \frac{\sum_{\widetilde{\mathbf{h}}^1} \exp(\sum_{j=1}^{M} s^1 \widetilde{h}_j^1(\mathbf{w}_{j*}^1(\mathbf{x}\circ\mathbf{m}^1) + b_j^1 + \mathbf{y}^T\mathbf{U}_{*j}))}{\sum_{\widetilde{\mathbf{h}}^1} \exp(\sum_{j=1}^{M} s^1 \widetilde{h}_j^1(\mathbf{w}_{j*}^1(\mathbf{x}\circ\mathbf{m}^1) + b_j^1 + \mathbf{y}^T\mathbf{U}_{*j}))} \times && (9)\\
&\ldots \times \frac{\exp(\sum_{j=1}^{M} s^\iota h_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j}))}{\sum_{\widetilde{\mathbf{h}}^\iota} \exp(\sum_{j=1}^{M} s^\iota \widetilde{h}_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j}))} \times && (10)\\
&\ldots \times \frac{\sum_{\widetilde{\mathbf{h}}^K} \exp(\sum_{j=1}^{M} s^K \widetilde{h}_j^K(\mathbf{w}_{j*}^K(\mathbf{x}\circ\mathbf{m}^K) + b_j^K + \mathbf{y}^T\mathbf{U}_{*j}))}{\sum_{\widetilde{\mathbf{h}}^K} \exp(\sum_{j=1}^{M} s^K \widetilde{h}_j^K(\mathbf{w}_{j*}^K(\mathbf{x}\circ\mathbf{m}^K) + b_j^K + \mathbf{y}^T\mathbf{U}_{*j}))} && (11)\\
&= \frac{\exp(\sum_{j=1}^{M} s^\iota h_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j}))}{\sum_{\widetilde{\mathbf{h}}^\iota} \exp(\sum_{j=1}^{M} s^\iota \widetilde{h}_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j}))} && (12)\\
&= \frac{\prod_{j=1}^{M} e^{s^\iota h_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}}{\sum_{\widetilde{\mathbf{h}}^\iota}\prod_{j=1}^{M} e^{s^\iota \widetilde{h}_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}} && (13)\\
&= \frac{\prod_{j=1}^{M} e^{s^\iota h_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}}{\sum_{\widetilde{h}_1^\iota} e^{s^\iota \widetilde{h}_1^\iota(\mathbf{w}_{1*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_1^\iota + \mathbf{y}^T\mathbf{U}_{*1})} \times \ldots \times \sum_{\widetilde{h}_M^\iota} e^{s^\iota \widetilde{h}_M^\iota(\mathbf{w}_{M*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_M^\iota + \mathbf{y}^T\mathbf{U}_{*M})}} && (14)\\
&= \prod_{j=1}^{M} \frac{e^{s^\iota h_j^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}}{1 + e^{s^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}} && (15)\\
&= \prod_{j=1}^{M} P(h_j^\iota|\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}). && (16)
\end{aligned}
$$

We obtain the probability of a particular hidden unit $h_j^\iota \in \{0,1\}$ being active given $\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}$ by:

$$
\begin{aligned}
P(h_j^\iota = 1|\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}) &= \frac{e^{s^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}}{1 + e^{s^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}} && (17)\\
&= \frac{1}{1 + e^{-s^\iota(\mathbf{w}_{j*}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + b_j^\iota + \mathbf{y}^T\mathbf{U}_{*j})}}. && (18)
\end{aligned}
$$

Then, we rewrite the conditional probability in a matrix form:

$$
P(\mathbf{h}^\iota = 1|\mathbf{x},\mathbf{y},\mathbf{s},\mathbf{m}) = \tau(s^\iota(\mathbf{W}^\iota(\mathbf{x}\circ\mathbf{m}^\iota) + \mathbf{b}^\iota + \mathbf{U}^T\mathbf{y})), \quad (19)
$$

where $\tau(\mathbf{x}) = 1/(1+\exp(-\mathbf{x}))$ is the sigmoid function.

1.2. Conditional Probability for Sampling Visible Data

Here, we derive the conditional probability for $\mathbf{x}$.

$$P(\mathbf{x}|\mathbf{h},\mathbf{s},\mathbf{m}) = \frac{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} P(\mathbf{x},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\mathbf{m})}{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} P(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\mathbf{m})} \tag{20}$$

$$= \frac{\frac{1}{Z}\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(-E(\mathbf{x},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\mathbf{m}))}{\frac{1}{Z}\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(-E(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\mathbf{m}))} \tag{21}$$

$$= \frac{\exp(\sum_{k=1}^{K} s^k \mathbf{h}^{kT}\mathbf{b}^k)}{\exp(\sum_{k=1}^{K} s^k \mathbf{h}^{kT}\mathbf{b}^k)} \times \frac{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(\widetilde{\mathbf{y}}^T U \sum_{k=1}^{K} s^k \mathbf{h}^k + \mathbf{d}^T\widetilde{\mathbf{y}})}{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(\widetilde{\mathbf{y}}^T U \sum_{k=1}^{K} s^k \mathbf{h}^k + \mathbf{d}^T\widetilde{\mathbf{y}})} \tag{22}$$

$$\times \frac{\exp(\sum_{k=1}^{K} s^k (\mathbf{h}^{kT}\mathbf{W}^k + \mathbf{c}^{kT})(\mathbf{x}\circ\mathbf{m}^k))}{\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(\sum_{k=1}^{K} s^k (\mathbf{h}^{kT}\mathbf{W}^k + \mathbf{c}^{kT})(\widetilde{\mathbf{x}}\circ\mathbf{m}^k))} \tag{23}$$

$$= \frac{\exp(\sum_{k=1}^{K}\sum_{i=1}^{N} s^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k) x_i m_i^k)}{\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(\sum_{k=1}^{K}\sum_{i=1}^{N} s^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k) \widetilde{x}_i m_i^k)} \tag{24}$$

$$= \frac{\prod_{i=1}^{N} e^{x_i \sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}}{\sum_{\widetilde{x}_1} e^{\widetilde{x}_1 \sum_{k=1}^{K} s^k m_1^k (\mathbf{W}_{1*}^{kT}\mathbf{h}^k + c_1^k)} \times \ldots \times \sum_{\widetilde{x}_N} e^{\widetilde{x}_N \sum_{k=1}^{K} s^k m_N^k (\mathbf{W}_{N*}^{kT}\mathbf{h}^k + c_N^k)}} \tag{25}$$

$$= \prod_{i=1}^{N} \frac{e^{x_i \sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}}{1 + e^{\sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}} \tag{26}$$

$$= \prod_{i=1}^{N} P(x_i|\mathbf{h},\mathbf{s},\mathbf{m}). \tag{27}$$

We obtain the probability of a particular visible unit $x_i \in \{0,1\}$ being active given $\mathbf{h},\mathbf{s},\mathbf{m}$ by

$$P(x_i = 1|\mathbf{h},\mathbf{s},\mathbf{m}) = \frac{e^{\sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}}{1 + e^{\sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}} \tag{28}$$

$$= \frac{1}{1 + e^{-\sum_{k=1}^{K} s^k m_i^k (\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)}}. \tag{29}$$

The conditional probability of $\mathbf{x}$ is written in matrix form as

$$P(\mathbf{x} = 1|\mathbf{h},\mathbf{s},\mathbf{m}) = \tau(\sum_{k=1}^{K} s^k \mathbf{m}^k (\mathbf{W}^{kT}\mathbf{h}^k + \mathbf{c}^k)). \tag{30}$$

1.3. Conditional Probability for Sampling Labels

The conditional probability of $\mathbf{y}$ is

$$
\begin{aligned}
P(\mathbf{y}|\mathbf{h},\mathbf{s}) &= \frac{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} P(\widetilde{\mathbf{x}},\mathbf{y},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}})}{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}}\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} P(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}})} && (31)\\
&= \frac{\frac{1}{Z}\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(-E(\widetilde{\mathbf{x}},\mathbf{y},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}}))}{\frac{1}{Z}\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}}\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(-E(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}}))} && (32)\\
&= \frac{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(\sum_{k=1}^{K} s^k\mathbf{h}^{kT}(\mathbf{W}^k(\widetilde{\mathbf{x}}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k)+\sum_{k=1}^{K} s^k\mathbf{c}^{kT}(\widetilde{\mathbf{x}}\circ\widetilde{\mathbf{m}}^k))}{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}}\sum_{\widetilde{\mathbf{x}}}^{\widetilde{\mathbf{X}}} \exp(\sum_{k=1}^{K} s^k\mathbf{h}^{kT}(\mathbf{W}^k(\widetilde{\mathbf{x}}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k)+\sum_{k=1}^{K} s^k\mathbf{c}^{kT}(\widetilde{\mathbf{x}}\circ\widetilde{\mathbf{m}}^k))} && (33)\\
&\quad \times\frac{\exp(\mathbf{y}^T\mathbf{U}\sum_{k=1}^{K} s^k\mathbf{h}^k+\mathbf{d}^T\mathbf{y})}{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}}\exp(\widetilde{\mathbf{y}}^T\mathbf{U}\sum_{k=1}^{K} s^k\mathbf{h}^k+\mathbf{d}^T\widetilde{\mathbf{y}})} && (34)\\
&= \frac{\exp(\sum_{l=1}^{L} y_l(\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l))}{\sum_{\widetilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}}\exp(\sum_{l=1}^{L} \widetilde{y}_l(\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l))} && (35)\\
&= \frac{\prod_{l=1}^{L} e^{y_l(\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l)}}{\sum_{\widetilde{y}_1} e^{\widetilde{y}_1(\mathbf{U}_{1*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_1)}\times\ldots\times\sum_{\widetilde{y}_L} e^{\widetilde{y}_L(\mathbf{U}_{L*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_L)}} && (36)\\
&= \prod_{l=1}^{L}\frac{e^{y_l(\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l)}}{1+e^{\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l}} && (37)\\
&= \prod_{l=1}^{L} P(y_l|\mathbf{h},\mathbf{s}). && (38)
\end{aligned}
$$

We get the probability of a particular label $y_l \in \{0,1\}$ being active given $\mathbf{h},\mathbf{s}$ by

$$
\begin{aligned}
P(y_l=1|\mathbf{h},\mathbf{s}) &= \frac{e^{\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l}}{1+e^{\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l}} && (39)\\
&= \frac{1}{1+e^{-(\mathbf{U}_{l*}\sum_{k=1}^{K} s^k\mathbf{h}^k+d_l)}}. && (40)
\end{aligned}
$$

The conditional probability of $\mathbf{y}$ is written as

$$
P(\mathbf{y}=1|\mathbf{h},\mathbf{s}) = \tau(\mathbf{U}(\sum_{k=1}^{K} s^k\mathbf{h}^k)+\mathbf{d}). \quad (41)
$$

1.4. Conditional Probability for Sampling Saliency Maps

The conditional probability of the mask $\mathbf{m}^\iota$ of the ι-th component is

$$
\begin{aligned}
P(\mathbf{m}^\iota|\mathbf{x},\mathbf{h},\mathbf{s}) &= \frac{\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{m}}\backslash\mathbf{m}^\iota}^{\widetilde{\mathbf{M}}} P(\mathbf{x},\tilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}})}{\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}} P(\mathbf{x},\tilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}})} && (42)\\
&= \frac{\frac{1}{Z}\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{m}}\backslash\mathbf{m}^\iota}^{\widetilde{\mathbf{M}}} \exp(-E(\mathbf{x},\tilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}}))}{\frac{1}{Z}\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}} \exp(-E(\mathbf{x},\tilde{\mathbf{y}},\mathbf{h},\mathbf{s},\widetilde{\mathbf{m}}))} && (43)\\
&= \frac{\exp(\sum_{k=1}^K s^k \mathbf{h}^{kT}\mathbf{b}^k)}{\exp(\sum_{k=1}^K s^k \mathbf{h}^{kT}\mathbf{b}^k)} \times \frac{\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(\tilde{\mathbf{y}}^T\mathbf{U}\sum_{k=1}^K s^k\mathbf{h}^k + \mathbf{d}^T\tilde{\mathbf{y}})}{\sum_{\tilde{\mathbf{y}}}^{\widetilde{\mathbf{Y}}} \exp(\tilde{\mathbf{y}}^T\mathbf{U}\sum_{k=1}^K s^k\mathbf{h}^k + \mathbf{d}^T\tilde{\mathbf{y}})} \times && (44)\\
&\quad \frac{\sum_{\widetilde{\mathbf{m}}\backslash\mathbf{m}^\iota}^{\widetilde{\mathbf{M}}} \exp(\sum_{k=1}^K s^k\mathbf{h}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k) + \sum_{k=1}^K s^k\mathbf{c}^{kT}(\mathbf{x}\circ\widetilde{\mathbf{m}}^k))}{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}} \exp(\sum_{k=1}^K s^k\mathbf{h}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k) + \sum_{k=1}^K s^k\mathbf{c}^{kT}(\mathbf{x}\circ\widetilde{\mathbf{m}}^k))} && (45)\\
&= \frac{\sum_{\widetilde{\mathbf{m}}\backslash\mathbf{m}^\iota}^{\widetilde{\mathbf{M}}} \exp(\sum_{k=1}^K\sum_{i=1}^N s^k(\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)x_i\widetilde{m}_i^k)}{\sum_{\widetilde{\mathbf{m}}}^{\widetilde{\mathbf{M}}} \exp(\sum_{k=1}^K\sum_{i=1}^N s^k(\mathbf{W}_{i*}^{kT}\mathbf{h}^k + c_i^k)x_i\widetilde{m}_i^k)} && (46)\\
&= \frac{\sum_{\widetilde{\mathbf{m}}^1}\exp(\sum_{i=1}^N s^1(\mathbf{W}_{i*}^{1T}\mathbf{h}^1 + c_i^1)x_i\widetilde{m}_i^1)}{\sum_{\widetilde{\mathbf{m}}^1}\exp(\sum_{i=1}^N s^1(\mathbf{W}_{i*}^{1T}\mathbf{h}^1 + c_i^1)x_i\widetilde{m}_i^1)} \times \dots && (47)\\
&\quad \times\frac{\exp(\sum_{i=1}^N s^\iota(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)x_i m_i^\iota)}{\sum_{\widetilde{\mathbf{m}}^\iota}\exp(\sum_{i=1}^N s^\iota(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)x_i\widetilde{m}_i^\iota)} \times \dots && (48)\\
&\quad \times\frac{\sum_{\widetilde{\mathbf{m}}^K}\exp(\sum_{i=1}^N s^K(\mathbf{W}_{i*}^{KT}\mathbf{h}^K + c_i^K)x_i\widetilde{m}_i^K)}{\sum_{\widetilde{\mathbf{m}}^K}\exp(\sum_{i=1}^N s^K(\mathbf{W}_{i*}^{KT}\mathbf{h}^K + c_i^K)x_i\widetilde{m}_i^K)} && (49)\\
&= \frac{\exp(\sum_{i=1}^N s^\iota(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)x_i m_i^\iota)}{\sum_{\widetilde{\mathbf{m}}^\iota}\exp(\sum_{i=1}^N s^\iota(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)x_i\widetilde{m}_i^\iota)} && (50)\\
&= \frac{\prod_{i=1}^N e^{s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)m_i^\iota}}{\sum_{\widetilde{m}_1^\iota} e^{s^\iota x_1(\mathbf{W}_{1*}^{\iota T}\mathbf{h}^\iota + c_1^\iota)\widetilde{m}_1^\iota} \times \dots \times \sum_{\widetilde{m}_N^\iota} e^{s^\iota x_N(\mathbf{W}_{N*}^{\iota T}\mathbf{h}^\iota + c_N^\iota)\widetilde{m}_N^\iota}} && (51)\\
&= \prod_{i=1}^N \frac{e^{s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)m_i^\iota}}{1 + e^{s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)}} && (52)\\
&= \prod_{i=1}^N P(m_i^\iota|\mathbf{x},\mathbf{h},\mathbf{s}). && (53)
\end{aligned}
$$

We get the probability of a particular mask unit i of component ι, $m_i^\iota \in [0,1]$ being active given $\mathbf{h},\mathbf{s}$ by

$$
\begin{aligned}
P(m_i^\iota = 1|\mathbf{x},\mathbf{h},\mathbf{s}) &= \frac{e^{s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)}}{1 + e^{s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)}} && (54)\\
&= \frac{1}{1 + e^{-s^\iota x_i(\mathbf{W}_{i*}^{\iota T}\mathbf{h}^\iota + c_i^\iota)}}. && (55)
\end{aligned}
$$

Then, we rewrite the conditional probability of $\mathbf{m}_\iota$ as

$$
P(\mathbf{m}^\iota = 1|\mathbf{x},\mathbf{h},\mathbf{s}) = \tau(s^\iota\mathbf{x}(\mathbf{W}^{\iota T}\mathbf{h}^\iota + \mathbf{c}^\iota)). \quad (56)
$$

1.5. Conditional Probability for Sampling Switch Variables

Subject to the constraint $s^k \in [0,1], \sum_{k=1}^{K} s^k = 1$, the probability of sampling switch variable s^ι is derived as

$$P(s^\iota = 1|\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{m}) = \frac{P(s^\iota = 1,\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{m})}{\sum_{k=1}^{K} P(s^k = 1,\mathbf{x},\mathbf{y},\mathbf{h},\mathbf{m})} \tag{57}$$

$$= \frac{\exp(\mathbf{h}^{\iota T}(\mathbf{W}^\iota(\mathbf{x}\circ\mathbf{m}^\iota)+\mathbf{b}^\iota)+c^{\iota T}(\mathbf{x}\circ\mathbf{m}^\iota)+\mathbf{y}^T\mathbf{U}\mathbf{h}^\iota)}{\sum_{k=1}^{K}\exp(\mathbf{h}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{b}^k)+c^{kT}(\mathbf{x}\circ\mathbf{m}^k)+\mathbf{y}^T\mathbf{U}\mathbf{h}^k)} \tag{58}$$

$$= \frac{1}{Z_0}\exp(\mathbf{h}^{\iota T}(\mathbf{W}^\iota(\mathbf{x}\circ\mathbf{m}^\iota)+\mathbf{b}^\iota)+c^{\iota T}(\mathbf{x}\circ\mathbf{m}^\iota)+\mathbf{y}^T\mathbf{U}\mathbf{h}^\iota). \tag{59}$$

where Z_0 indicates a normalized factor which is equal to the sum of all components.

2. Proof of Eq.(10)

$P(\mathbf{x},\mathbf{y}|\mathbf{x})$ can be defined by integrating over $\mathbf{h}$ and $\mathbf{m}$. The detailed derivations are as follows:

$$P(\mathbf{x},\mathbf{y}|\mathbf{s}) \propto \frac{1}{Z}\sum_{\widetilde{\mathbf{h}},\widetilde{\mathbf{m}}} e^{-E(\mathbf{x},\mathbf{y},\widetilde{\mathbf{h}},\mathbf{s},\widetilde{\mathbf{m}})} \tag{60}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\sum_{\widetilde{\mathbf{h}},\widetilde{\mathbf{m}}} e^{\sum_{k=1}^{K} s^k(\widetilde{\mathbf{h}}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k)+\mathbf{y}^T\mathbf{U}\widetilde{\mathbf{h}}^k+\mathbf{c}^{kT}(\mathbf{x}\circ\widetilde{\mathbf{m}}^k))} \tag{61}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\sum_{\widetilde{\mathbf{h}},\widetilde{\mathbf{m}}}\prod_{k=1}^{K} e^{s^k(\widetilde{\mathbf{h}}^{kT}(\mathbf{W}^k(\mathbf{x}\circ\widetilde{\mathbf{m}}^k)+\mathbf{b}^k)+\mathbf{y}^T\mathbf{U}\widetilde{\mathbf{h}}^k+\mathbf{c}^{kT}(\mathbf{x}\circ\widetilde{\mathbf{m}}^k))} \tag{62}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^{K}\sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}\sum_{\widetilde{\mathbf{m}}^k} e^{s^k(\widetilde{\mathbf{h}}^{kT}\mathbf{W}^k+\mathbf{c}^{kT})(\mathbf{x}\circ\widetilde{\mathbf{m}}^k)} \tag{63}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^{K}\sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}\sum_{\widetilde{\mathbf{m}}^k} e^{s^k\sum_{i=1}^{N}(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)(x_i\widetilde{m}_i^k)} \tag{64}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^{K}\sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}\sum_{\widetilde{\mathbf{m}}^k}\prod_{i=1}^{N} e^{s^k(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)(x_i\widetilde{m}_i^k)} \tag{65}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^{K}\sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}\prod_{i=1}^{N}\sum_{\widetilde{m}_i^k} e^{s^k(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)(x_i\widetilde{m}_i^k)} \tag{66}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^{K}\sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}\prod_{i=1}^{N}(1+e^{s^k(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)x_i}) \tag{67}$$

$$\tag{68}$$

In order to simplify the expression, we firstly expand the term $\prod_{i=1}^{N}(1+e^{s^k(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)x_i})$, which is similar to binomial expression, and then define a series of $G_{lk}(l=0,1,2...,N)$, where G_{lk} indicates that we choose 1 term $N-k$ times and exponent term k times generally. Extremely, G_{l0} means we choose 1 term N times, exponent term 0 time, and G_{lN} means we choose 1 term 0 time, exponent term N times. Taking the common multiplication factor $e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k}$ into account, we are able to write G_{lk} as follows.

$$G_{0k} = e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k} * 1^N \tag{69}$$

$$G_{1k} = e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k} * 1^{N-1} * (e^{s^k(\mathbf{W}_{1*}^{kT}\widetilde{\mathbf{h}}^k+c_1^k)x_1} + e^{s^k(\mathbf{W}_{2*}^{kT}\widetilde{\mathbf{h}}^k+c_2^k)x_2} + ... + e^{s^k(\mathbf{W}_{N*}^{kT}\widetilde{\mathbf{h}}^k+c_N^k)x_N}) \tag{70}$$

$$G_{2k} = e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k} * 1^{N-2}(e^{s^k(\mathbf{W}_{1*}^{kT}\widetilde{\mathbf{h}}^k+c_1^k)x_1+s^k(\mathbf{W}_{2*}^{kT}\widetilde{\mathbf{h}}^k+c_2^k)x_2} + \tag{71}$$

$$e^{s^k(\mathbf{W}_{1*}^{kT}\widetilde{\mathbf{h}}^k+c_1^k)x_1+s^k(\mathbf{W}_{3*}^{kT}\widetilde{\mathbf{h}}^k+c_3^k)x_3} + ... + e^{s^k(\mathbf{W}_{N-1*}^{kT}\widetilde{\mathbf{h}}^k+c_{N-1}^k)x_{N-1}+s^k(\mathbf{W}_{N*}^{kT}\widetilde{\mathbf{h}}^k+c_N^k)x_N}) \tag{72}$$

$$... \tag{73}$$

$$G_{Nk} = e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k} * 1^0 * e^{\sum_{i=1}^N s^k(\mathbf{W}_{i*}^{kT}\widetilde{\mathbf{h}}^k+c_i^k)x_i} \tag{74}$$

Then

$$P(\mathbf{x},\mathbf{y}|\mathbf{s}) \propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^K\sum_{\widetilde{\mathbf{h}}^k}\sum_{l=0}^N G_{lk} \tag{75}$$

$$\propto \frac{1}{Z}e^{\mathbf{d}^T\mathbf{y}}\prod_{k=1}^K\sum_{l=0}^N\sum_{\widetilde{\mathbf{h}}^k} G_{lk} \tag{76}$$

Next we calculate $\sum_{\widetilde{\mathbf{h}}^k}$ operation for each G_{lk}, and could obtain a series of $Q_{lk}(l = 0, 1, 2..., N)$. For example, when $l = 0$, the derivation for Q_{l0} is,

$$Q_{0k} = \sum_{\widetilde{\mathbf{h}}^k} G_{0k} \tag{77}$$

$$= \sum_{\widetilde{\mathbf{h}}^k} e^{s^k(\mathbf{y}^T\mathbf{U}+\mathbf{b}^T)\widetilde{\mathbf{h}}^k} \tag{78}$$

$$= \sum_{\widetilde{\mathbf{h}}^k}\prod_{j=1}^M e^{s^k(\mathbf{y}^T\mathbf{U}_{*j}+\mathbf{b}_j^T)\widetilde{h}_j^k} \tag{79}$$

$$= \prod_{j=1}^M\sum_{\widetilde{h}_j^k} e^{s^k(\mathbf{y}^T\mathbf{U}_{*j}+\mathbf{b}_j^T)\widetilde{h}_j^k} \tag{80}$$

$$= \prod_{j=1}^M(1 + e^{s^k(\mathbf{y}^T\mathbf{U}_{*j}+\mathbf{b}_j^T)}) \tag{81}$$

$$\tag{82}$$

Similarly, when $l = 1, 2, ..., N$,

$$Q_{1k} = e^{s^k x_1 c_1^k} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_1 \mathbf{W}_{1j}^{kT})}) + e^{s^k x_2 c_2^k} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_2 \mathbf{W}_{2j}^{kT})}) + ... \quad (83)$$

$$+ e^{s^k x_N c_N^k} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_N \mathbf{W}_{Nj}^{kT})}) \quad (84)$$

$$Q_{2k} = e^{s^k(x_1 c_1^k + x_2 c_2^k)} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_1 \mathbf{W}_{1j}^{kT} + x_2 \mathbf{W}_{2j}^{kT})}) + \quad (85)$$

$$e^{s^k(x_1 c_1^k + x_3 c_3^k)} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_1 \mathbf{W}_{1j}^{kT} + x_3 \mathbf{W}_{3j}^{kT})}) + ... \quad (86)$$

$$e^{s^k(x_{N-1} c_{N-1}^k + x_N c_N^k)} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + x_{N-1} \mathbf{W}_{(N-1)j}^{kT} + x_N \mathbf{W}_{Nj}^{kT})}) \quad (87)$$

$$... \quad (88)$$

$$Q_{Nk} = e^{\sum_{i=1}^{N} s^k x_i c_i^k} \prod_{j=1}^{M} (1 + e^{s^k(\mathbf{y}^T \mathbf{U}_{*j} + \mathbf{b}_j^T + \sum_{i=1}^{N} x_i \mathbf{W}_{ij}^{kT})}) \quad (89)$$

Therefore, we could rewrite $P(\mathbf{x}, \mathbf{y}|\mathbf{s})$ as:

$$P(\mathbf{x}, \mathbf{y}|\mathbf{s}) \propto \frac{1}{Z} e^{\mathbf{d}^T \mathbf{y}} \prod_{k=1}^{K} \sum_{l=0}^{N} Q_{lk} \quad (90)$$