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

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

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

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

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 \(h^i\) of the \(i\)-th component is

\begin{align}
P(h^i|x, y, s, m) &= \frac{P(h^i, x, y, s, m)}{P(x, y, s, m)} \\
&= \frac{\sum_{\tilde{h}, h^i} P(\tilde{h}, x, y, s, m)}{\sum_{\tilde{h}} P(\tilde{h}, x, y, s, m)} \\
&= \frac{1}{Z} \sum_{\tilde{h}, h^i} \exp(-E(\tilde{h}, x, y, s, m)) \\
&= \frac{1}{Z} \sum_{\tilde{h}} \exp(-E(\tilde{h}, x, y, s, m)) \\
&= \frac{\exp(\sum_{k=1}^{K} s^k c^{kT}(x \circ m^k) + d^Ty) \times \\
\sum_{\tilde{h}, h^i} \exp(\sum_{k=1}^{K} s^k h^{kT}(W^k(x \circ m^k) + b^k) + y^T U \sum_{k=1}^{K} s^k \tilde{h}^k)}{\sum_{\tilde{h}} \exp(\sum_{k=1}^{K} s^k h^{kT}(W^k(x \circ m^k) + b^k) + y^T U \sum_{k=1}^{K} s^k h^k)}
\end{align}
Then, we rewrite the conditional probability in a matrix form:
\[
(\mathbf{x} \circ \mathbf{m}) \frac{P(\mathbf{h}^T | \mathbf{x}, \mathbf{y}, \mathbf{s}, \mathbf{m})}{1 + \exp(-\mathbf{x})} = \frac{1}{1 + \exp(-\mathbf{x})},
\]
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 $x$.

$$P(x|h, s, m) = \frac{\sum_{\tilde{y}} P(x, \tilde{y}, h, s, m)}{\sum_{\tilde{y}} \sum_{\tilde{x}} P(\tilde{x}, \tilde{y}, h, s, m)}$$  \hspace{1cm} (20)

$$= \frac{1}{Z} \sum_{\tilde{y}} \exp(-E(x, \tilde{y}, h, s, m))$$  \hspace{1cm} (21)

$$= \frac{\exp(\sum_{k=1}^{K} s^k h^T h^k)}{\exp(\sum_{k=1}^{K} s^k h^T h^k) \times \sum_{y} \exp(\tilde{y}^T U \sum_{k=1}^{K} s^k h^k + d^T \tilde{y})}$$  \hspace{1cm} (22)

$$= \frac{\sum_{x} \sum_{\tilde{y}} \exp(\sum_{k=1}^{K} s^k (h^k W^k + c^k)(x \circ m^k))}{\sum_{y} \sum_{x} \sum_{\tilde{y}} \exp(\tilde{y}^T U \sum_{k=1}^{K} s^k h^k + d^T \tilde{y})}$$  \hspace{1cm} (23)

$$= \frac{\prod_{i=1}^{N} P(x_i|h, s, m)}{\prod_{i=1}^{N} 1 + e^{\sum_{k=1}^{K} s^k m^k_i (W^k_i h^k + c^k_i)}}$$  \hspace{1cm} (24)

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

$$P(x_i = 1|h, s, m) = \frac{e^{\sum_{k=1}^{K} s^k m^k_i (W^k_i h^k + c^k_i)}}{1 + e^{\sum_{k=1}^{K} s^k m^k_i (W^k_i h^k + c^k_i)}}$$  \hspace{1cm} (25)

The conditional probability of $x$ is written in matrix form as

$$P(x = 1|h, s, m) = \tau(\sum_{k=1}^{K} s^k m^k(W^k h^k + c^k)).$$  \hspace{1cm} (30)
1.3. Conditional Probability for Sampling Labels

The conditional probability of \( y \) is

\[
P(y|h, s) = \frac{\sum_{\tilde{m}} \sum_{\tilde{x}} P(\tilde{x}, y, h, s, \tilde{m})}{\sum_{\tilde{m}} \sum_{\tilde{x}} \sum_{\tilde{y}} P(\tilde{x}, \tilde{y}, h, s, \tilde{m})} \tag{31}
\]

\[
= \frac{1}{Z} \sum_{\tilde{m}} \sum_{\tilde{x}} \exp(-E(\tilde{x}, y, h, s, \tilde{m}))
\]

\[
= \frac{1}{Z} \sum_{\tilde{m}} \sum_{\tilde{x}} \exp(\sum_{k=1}^{K} s_k h^T (W^k (\tilde{x} \circ \tilde{m}^k) + b^k) + \sum_{k=1}^{K} s_k e^{kT} (\tilde{x} \circ \tilde{m}^k))
\]

\[
= \frac{1}{Z} \sum_{\tilde{m}} \sum_{\tilde{x}} \exp(\sum_{k=1}^{K} s_k h^T (W^k (\tilde{x} \circ \tilde{m}^k) + b^k) + \sum_{k=1}^{K} s_k e^{kT} (\tilde{x} \circ \tilde{m}^k))
\]

\[
\times \exp(y^T U \sum_{k=1}^{K} s_k h^k + d^T y)
\]

\[
= \frac{\exp(\sum_{l=1}^{L} y_l (U_{l*} \sum_{k=1}^{K} s_k h^k + d_l)) \sum_{\tilde{y}} \exp(\sum_{l=1}^{L} y_l (U_{l*} \sum_{k=1}^{K} s_k h^k + d_l))}{\sum_{\tilde{y}} \sum_{l=1}^{L} e^{y_l (U_{l*} \sum_{k=1}^{K} s_k h^k + d_l)} \times \ldots \times \sum_{\tilde{y} L} e^{y_L (U_{L*} \sum_{k=1}^{K} s_k h^k + d_L)}}
\]

\[
= \prod_{l=1}^{L} \frac{e^{y_l (U_{l*} \sum_{k=1}^{K} s_k h^k + d_l)}}{1 + e^{U_{l*} \sum_{k=1}^{K} s_k h^k + d_l}}
\]

\[
= \prod_{l=1}^{L} P(y_l|h, s).
\] \tag{38}

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

\[
P(y_l = 1|h, s) = \frac{e^{U_{l*} \sum_{k=1}^{K} s_k h^k + d_l}}{1 + e^{U_{l*} \sum_{k=1}^{K} s_k h^k + d_l}} \tag{39}
\]

\[
= \frac{1}{1 + e^{-(U_{l*} \sum_{k=1}^{K} s_k h^k + d_l)}}.
\] \tag{40}

The conditional probability of \( y \) is written as

\[
P(y = 1|h, s) = \tau(U(\sum_{k=1}^{K} s_k h^k + d)).
\] \tag{41}
1.4. Conditional Probability for Sampling Saliency Maps

The conditional probability of the mask $m^i$ of the $i$-th component is

$$P(m^i | x, h, s) = \frac{\sum \tilde{y} \sum \tilde{m}(m^i) P(x, \tilde{y}, h, s, \tilde{m})}{\sum \tilde{y} \sum \tilde{m} P(x, \tilde{y}, h, s, \tilde{m})}$$

$$= \frac{1}{Z} \sum \tilde{y} \sum \tilde{m}(m^i) \exp(-E(x, \tilde{y}, h, s, \tilde{m}))$$

$$= \frac{1}{Z} \sum \tilde{y} \sum \tilde{m} \exp(-E(x, \tilde{y}, h, s, \tilde{m}))$$

$$= \frac{\exp(\sum_{k=1}^{K} s_k h^T b^k) \times \sum \tilde{y} \exp(\tilde{y}^T U \sum_{k=1}^{K} s_k h^k + d^T \tilde{y})}{\exp(\sum_{k=1}^{K} s_k h^T b^k) \times \sum \tilde{y} \exp(\tilde{y}^T U \sum_{k=1}^{K} s_k h^k + d^T \tilde{y})} \times \sum \tilde{m}(m^i) \exp(\sum_{k=1}^{K} s_k h^k (W_k^T x + b_k) + \sum_{k=1}^{K} c_k^k (x \circ \tilde{m}_k^i))$$

$$= \sum \tilde{m} \exp(\sum_{k=1}^{K} s_k (W_k^T h^k + c_k^1) x_i \tilde{m}_k^i)$$

$$= \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i) \times \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i)$$

$$= \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i) \times \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i)$$

$$= e^{s^i x_i (W_i^T h^i + c_i^1)} \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i)$$

$$= e^{s^i x_i (W_i^T h^i + c_i^1)} \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i)$$

$$= e^{s^i x_i (W_i^T h^i + c_i^1)} \sum \tilde{m} \exp(\sum_{i=1}^{N} s^i (W_i^T h^i + c_i^1) x_i \tilde{m}_i^i)$$

$$= \prod_{i=1}^{N} P(m_i^i | x, h, s).$$

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

$$P(m_i^i = 1 | x, h, s) = \frac{1}{1 + e^{s^i x_i (W_i^T h^i + c_i^1)}}$$

Then, we rewrite the conditional probability of $m_i$ as

$$P(m_i = 1 | x, h, s) = \tau(s^i x(W_i^T h + c^i)).$$
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 \( s^t \) is derived as

\[
P(s^t = 1 | x, y, h, m) = \frac{P(s^t = 1, x, y, h, m)}{\sum_{k=1}^{K} P(s^t = 1, x, y, h, m)}
\]

\[
= \frac{\exp(\mathbf{h}^T (\mathbf{W}^k (x \circ m^t) + b^t) + c^T (x \circ m^t) + y^T \mathbf{U}h^t)}{\sum_{k=1}^{K} \exp(\mathbf{h}^k (\mathbf{W}^k (x \circ m^t) + b^t) + c^T (x \circ m^t) + y^T \mathbf{U}h^t)}
\]

\[
= \frac{1}{Z_0} \exp(\mathbf{h}^T (\mathbf{W}^t (x \circ m^t) + b^t) + c^T (x \circ m^t) + y^T \mathbf{U}h^t).
\]

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

2. Proof of Eq.(10)

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

\[
P(x, y|s) \propto \frac{1}{Z} \sum_{h, m} e^{-E(x, y, h, s, m)}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \sum_{h, m} e^{\sum_{k=1}^{K} s^k (\mathbf{h}^k (\mathbf{W}^k (x \circ m^k) + b^k) + y^T \mathbf{U}h^k + c^T (x \circ m^k))}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \sum_{h, m} \prod_{k=1}^{K} e^{s^k (\mathbf{h}^k (\mathbf{W}^k (x \circ m^k) + b^k) + y^T \mathbf{U}h^k + c^T (x \circ m^k))}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \prod_{k=1}^{K} \sum_{h^k} e^{s^k (y^T U + b^T)h^k} \sum_{m^k} e^{s^k (\mathbf{h}^k \mathbf{W}^k + c^T) (x \circ m^k)}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \prod_{k=1}^{K} \sum_{h^k} e^{s^k (y^T U + b^T)h^k} \sum_{m^k} e^{s^k \sum_{i=1}^{N} (\mathbf{W}^k_i \mathbf{h}^k + c_i^k) (x_i \circ m_{i}^k)}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \prod_{k=1}^{K} \sum_{h^k} e^{s^k (y^T U + b^T)h^k} \prod_{i=1}^{N} e^{s^k (\mathbf{W}^k_i \mathbf{h}^k + c_i^k) (x_i \circ m_{i}^k)}
\]

\[
= \frac{1}{Z} e^{\mathbf{d}^T y} \prod_{k=1}^{K} \sum_{h^k} e^{s^k (y^T U + b^T)h^k} \prod_{i=1}^{N} (1 + e^{s^k (\mathbf{W}^k_i \mathbf{h}^k + c_i^k) x_i})
\]

In order to simplify the expression, we firstly expand the term \( \prod_{i=1}^{N} (1 + e^{s^k (\mathbf{W}^k_i \mathbf{h}^k + c_i^k) x_i}) \), which is similar to binomial expression, and then define a series of \( G_{ik} (l = 0, 1, 2, ..., N) \), where \( G_{ik} \) indicates that we choose 1 term \( N - k \) times and exponent term \( k \) times generally. Extremely, \( G_{i0} \) means we choose 1 term \( N \) times, exponent term 0 time, and \( G_{iN} \) means we choose 1 term 0 time, exponent term \( N \) times. Taking the common multiplication factor \( e^{s^k (y^T U + b^T)h^k} \) into account, we are able to write \( G_{ik} \) as follows.
\[ G_{0k} = e^{sk}(y^T U + b^T)\tilde{h}^k \cdot 1^N \] (69)

\[ G_{1k} = e^{sk}(y^T U + b^T)\tilde{h}^k \cdot 1^{N-1} \cdot (e^{sk}(W_{12}^k T \tilde{h}^k + c_k^1)x_1 + e^{sk}(W_{22}^k T \tilde{h}^k + c_k^2)x_2 + \ldots + e^{sk}(W_{N2}^k T \tilde{h}^k + c_k^N)x_N) \] (70)

\[ G_{2k} = e^{sk}(y^T U + b^T)\tilde{h}^k \cdot 1^{N-2} (e^{sk}(W_{12}^k T \tilde{h}^k + c_k^1)x_1 + e^{sk}(W_{22}^k T \tilde{h}^k + c_k^2)x_2 + \ldots + e^{sk}(W_{N2}^k T \tilde{h}^k + c_k^N)x_N) \] (71)

\ldots

\[ G_{Nk} = e^{sk}(y^T U + b^T)\tilde{h}^k \cdot 1^0 \cdot e^{s\sum_{i=1}^{N} (e^{sk}(W_{i2}^k T \tilde{h}^k + c_k^i)x_i} \] (74)

Then

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

\[ \propto \frac{1}{Z} e^{d^T y} \prod_{k=1}^{K} \sum_{l=0}^{N} G_{lk} \] (76)

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

\[ Q_{0k} = \sum_{\tilde{h}^k} G_{0k} \] (77)

\[ = \sum_{\tilde{h}^k} e^{s\sum_{i=1}^{N} (e^{sk}(W_{i2}^k T \tilde{h}^k + c_k^i)x_i} \] (78)

\[ = \sum_{\tilde{h}^k} M \prod_{j=1}^{M} (1 + e^{s\sum_{i=1}^{N} (e^{sk}(W_{i2}^k T \tilde{h}^k + c_k^i)x_i}) \] (82)
Similarly, when \( l = 1, 2, \ldots, N \),

\[
Q_{1k} = e^{skx_1c_1^k} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_1 W_{1j}^{kT})} \right) + e^{skx_2c_2^k} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_2 W_{2j}^{kT})} \right) + \ldots \tag{83}
\]

\[
+ e^{skx_N c_N^k} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_N W_{Nj}^{kT})} \right) \tag{84}
\]

\[
Q_{2k} = e^{sk(x_1 c_1^k + x_2 c_2^k)} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_1 W_{1j}^{kT} + x_2 W_{2j}^{kT})} \right) + \ldots \tag{85}
\]

\[
e^{sk(x_1 c_1^k + x_3 c_3^k)} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_1 W_{1j}^{kT} + x_3 W_{3j}^{kT})} \right) + \ldots \tag{86}
\]

\[
e^{sk(x_{N-1} c_{N-1}^k + x_N c_N^k)} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + x_{N-1} W_{(N-1)j}^{kT} + x_N W_{Nj}^{kT})} \right) \tag{87}
\]

\[
\ldots \tag{88}
\]

\[
Q_{Nk} = e^{\sum_{i=1}^{N} sk x_i c_i^k} \prod_{j=1}^{M} \left( 1 + e^{sk(y^T U_{sj} + b_j^T + \sum_{i=1}^{N} x_i W_{ij}^{kT})} \right) \tag{89}
\]

Therefore, we could rewrite \( P(x, y | s) \) as:

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