Supplemental Material

Bilinear Heterogeneous Information Machine for RGB-D Action Recognition

Anonymous CVPR submission

Paper ID 129

1. Outline

In this supplemental material, we provide the proof for the deduction on Page 5.

**Proposition.** Denote $\tilde{X}^m_i = W^m_i B^{-\frac{1}{2}}$, $\tilde{X}^m = X^m_i W^m_i B^{-\frac{1}{2}}$. If $p(W, B^{-\frac{1}{2}} | X^m, O)$ is a uniform distribution, then $I(X^m, O) = I(\tilde{X}^m, \tilde{O})$.

**Proof.** According to the definition of mutual information, we have

$$ I(\tilde{X}^m, \tilde{O}) = \sum_{\tilde{X}^m} \sum_{\tilde{O}} p(\tilde{X}^m, \tilde{O}) \log \frac{p(\tilde{X}^m, \tilde{O})}{p(\tilde{X}^m)p(\tilde{O})}. \quad (1) $$

Note that $\tilde{X}^m_i = X^m_i W^m_i B^{-\frac{1}{2}}$ and $\tilde{O}^m_i = O^T W^m_i B^{-\frac{1}{2}}$. Therefore, Eq. (1) can be rewritten as

$$ I(\tilde{X}^m_i, \tilde{O}^m_i) = \sum_{X^m} \sum_{O} p(X^m, O) p(W, B^{-\frac{1}{2}} | X, O) \log \frac{p(X^m, O) p(W, B^{-\frac{1}{2}} | X, O)}{p(X^m)p(O)}. \quad (2) $$

We assume that both $p(\tilde{X}^m_i)$ and $p(\tilde{O}^m_i)$ are uniform distributions, same as $p(X^m)$ and $p(O)$. Then Eq. (2) can be rewritten as

$$ I(\tilde{X}^m_i, \tilde{O}^m_i) = \sum_{X^m} \sum_{O} p(X^m, O) p(W, B^{-\frac{1}{2}} | X, O) \log \frac{p(X^m, O) p(W, B^{-\frac{1}{2}} | X, O)}{p(X^m)p(O)}. \quad (3) $$

If $p(W, B^{-\frac{1}{2}} | X, O)$ is a uniform distribution, then the following equation can be derived:

$$ I(\tilde{X}^m, \tilde{O}) = \sum_{X^m} \sum_{O} p(X^m, O) \log \frac{p(X^m, O)}{p(X^m)p(O)} \quad (4) $$

$$ = I(X^m, O). $$