Supplementary Material for Shape-Erased Feature Learning for Visible-Infrared Person Re-Identification

A. On Minimizing $I(Z_{se}^{(i)}; Y; X^{(s)})$

In Appendix A, we will proof $I(Z_{se}^{(i)}; Y; X^{(s)})$ can be approximately upper-bounded by 0, so that $\max I(Z_{se}^{(i)}; Y|X^{(s)})$ can be lower-bounded by $\max I(Z_{se}^{(i)}; Y)$ (Eq. (4), (13) in the paper). We first enumerate the following four hypotheses for minimizing $I(Z_{se}^{(i)}; Y; X^{(s)})$.

Hypothesis:

1. Following [1], if $Z^{(s)}$ is a representation of $X^{(s)}$, then we state that $Z^{(s)}$ is conditionally independent from any other variable in the system once $X^{(s)}$ is observed (*e.g.* $Z^{(s)}$ can be a deterministic function of $X^{(s)}$):
$$\forall A, B, \quad I(A; Z^{(s)}|X^{(s)}, B) = 0.$$
2. (Eq. (6) in the paper) $Z^{(s)}$ is a *sufficient* representation of $X^{(s)}$ for Y, *i.e.*, $I(Y; X^{(s)}|Z^{(s)}) = 0$.
3. (Eq. (8) in the paper) $Z_{sr}^{(i)}$ can fully represent $Z^{(s)}$, *i.e.*, $Z^{(s)} \equiv Z_{sr}^{(i)}$.
4. (Eq. (1), (2) in the paper) The orthogonality between $z_{se}^{(i)}$ and $z_{sr}^{(i)}$ can be regarded as a relaxation of independence, *i.e.*,
$$\forall\, (z_{sr}^{(i)}, z_{se}^{(i)}) \sim (Z_{sr}^{(i)}, Z_{se}^{(i)}),\ z_{sr}^{(i)} \perp z_{se}^{(i)} \Longrightarrow I(Z_{sr}^{(i)}; Z_{se}^{(i)}) \approx 0.$$

Hypothesis 2 can be satisfied by **Proposition 1** in Appendix B. To approximate *Hypothesis* 3, we minimize element-wise mean squared error (MSE) between them (Eq. (9) in the paper), and it is to be noted that the gradient of $z^{(s)}$ is discarded.

Theorem 1. *If representation $Z^{(s)}$ of $X^{(s)}$ is **sufficient** for Y, then $I(Z_{se}^{(i)}; Y; X^{(s)}) = I(Z_{se}^{(i)}; Y; Z^{(s)})$.*

Proof. Obviously, $I(Z_{se}^{(i)}; Y; X^{(s)}) \geq I(Z_{se}^{(i)}; Y; Z^{(s)})$ holds due to data processing inequality ($X^{(s)} \rightarrow Z^{(s)}$). On the other side, $I(Z_{se}^{(i)}; Y; X^{(s)})$ can be factorized into two terms by introducing $Z^{(s)}$:

$$I(Z_{se}^{(i)}; Y; X^{(s)}) = I(Z_{se}^{(i)}; Y; X^{(s)}|Z^{(s)}) + I(Z_{se}^{(i)}; Y; X^{(s)}; Z^{(s)}). \tag{A.1}$$

For the first term of RHS in Eq. (A.1):

$$\begin{aligned} I(Z_{se}^{(i)}; Y; X^{(s)}|Z^{(s)}) =& I(Y; X^{(s)}|Z^{(s)}) - I(Y; X^{(s)}|Z_{se}^{(i)}, Z^{(s)}) \\ =& 0 - I(Y; X^{(s)}|Z_{se}^{(i)}, Z^{(s)}) \leq 0, \end{aligned} \tag{A.2}$$

where $I(Y; X^{(s)}|Z^{(s)}) = 0$ using the definition of *sufficiency*; For the second term of RHS in Eq. (A.1):

$$I(Z_{se}^{(i)}; Y; X^{(s)}; Z^{(s)}) = I(Z_{se}^{(i)}; Y; Z^{(s)}) - I(Z_{se}^{(i)}; Y; Z^{(s)}|X^{(s)}). \tag{A.3}$$

For the second term of RHS in Eq. (A.3):

$$\begin{aligned} I(Z_{se}^{(i)}; Y; Z^{(s)}|X^{(s)}) &= I(Y; Z^{(s)}|X^{(s)}) - I(Y; Z^{(s)}|X^{(s)}, Z_{se}^{(i)}) \\ &= 0 - 0 = 0, \end{aligned} \tag{A.4}$$

where $I(Y;Z^{(s)}|X^{(s)}) = I(Y;Z^{(s)}|X^{(s)},Z_{se}^{(i)}) = 0$ as $Z^{(s)}$ is a representation of $X^{(s)}$ using *Hypothesis* 1. Therefore, combining Eq. (A.1) - (A.4) concludes:

$$I(Z_{se}^{(i)};Y;X^{(s)}) = I(Z_{se}^{(i)};Y;Z^{(s)}). \tag{A.5}$$

□

Following **Theorem 1**, and using *Hypothesis* 3 and 4, we have:

$$I(Z_{se}^{(i)};Y;X^{(s)}) = I(Z_{se}^{(i)};Y;Z_{sr}^{(i)}) \leq I(Z_{se}^{(i)};Z_{sr}^{(i)}) \approx 0. \tag{A.6}$$

Based on the above analysis, it is concluded that $I(Z_{se}^{(i)};Y;X^{(s)})$ can be upper-bounded by $I(Z_{se}^{(i)};Z_{sr}^{(i)}) \approx 0$.

B. On Loss Functions

In Section 3, we maximize mutual information between representation and label by minimizing cross-entropy loss (Eq. (5), (7) in the paper). We formulate this approximation as the following **Proposition 1**.

Proposition 1. *Let X and Y be random variables with domain $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let Z be a representation of X. Then, maximizing $I(Z;Y)$ can be approximated by minimizing cross-entropy loss of $q(y|z)$ given observations from $P(X,Y)$ as $\{x_j,y_j\}_{j=1}^N$. $q(y|z)$ is regarded as classifier in practical.*

Proof. Using the definitions of mutual information and entropy:

$$\max I(Z;Y) = H(Y) - H(Y|Z), \tag{B.1}$$

and as $H(Y)$ will not change if domain $\mathcal{Y}$ does not change, maximizing $I(Z;Y)$ is equivalent to minimizing $H(Y|Z)$:

$$\begin{aligned} \min H(Y|Z) &= \int p(z)H(Y|Z=z)dz \\ &= -\iint p(z)p(y|z)\log p(y|z)\ dydz. \end{aligned} \tag{B.2}$$

As $D_{KL}(p(y|z)\|q(y|z)) = \int p(y|z)\log p(y|z) - p(y|z)\log q(y|z)dz \geq 0$ holds identically:

$$\begin{aligned} \min H(Y|Z) &= -\iint p(z)p(y|z)\log p(y|z)\ dydz \\ &\leq -\iint p(z)p(y|z)\log q(y|z)\ dydz \\ &= -\iint p(z,y)\log q(y|z)\ dydz \\ &= -\iiint p(y|x)p(z|x)p(x)\log q(y|z)\ dxdydz. \end{aligned} \tag{B.3}$$

The last equation holds for Z is conditional independent from Y given X based on the graphical model illustrated in Section 3 in the paper ($Y \to X \to Z$), *i.e.*, $p(y,z|x) = p(y|x)p(z|x)$. For specific observations $\{x_j,y_j\}_{j=1}^N$ (and note that $p(z|x)$ is usually represented as a deterministic function), we can approximate the upper bound of $H(Y|Z)$ by Monte Carlo sampling:

$$\begin{aligned} \min H(Y|Z) &\leq -\iiint p(y|x)p(z|x)p(x)\log q(y|z)\ dxdydz \\ &\approx -\frac{1}{N}\sum_{j=1}^{N}\log q(y_j|z_j), \end{aligned} \tag{B.4}$$

which is a typical form of cross-entropy loss. Therefore, **Proposition 1** holds. □

Remark. *For Hypothesis 1 in Appendix A, if the approximation in **Proposition 1** is close enough, then we can infer that $D_{KL}(p(y|x)\|q(y|z)) \to 0^+$, which indicates the sufficiency of $Z^{(s)}$ of $X^{(s)}$ for Y.*

In Section 3, we minimize cross-view conditional mutual information by minimizing cross-entropy loss (Eq. (10), (11), (14), (15), (17) in the paper). We formulate this approximation as the following **Proposition 2**.

Proposition 2. *Let $X^{(1)}$ and $X^{(2)}$ be random variables from visible modality and infrared modality (or generally two different views,* i.e.*, modality view and body shape view), Y be random variable of identity. Let $Z^{(1)}$ and $Z^{(2)}$ be representations of $X^{(1)}$ and $X^{(2)}$. Then minimizing $I(X^{(1)}; Z^{(1)}|X^{(2)})$ can be approximated by minimizing cross-entropy between $p(y|z^{(2)})$ and $p(y|z^{(1)})$.*

Proof.

$$
\begin{aligned}
I(X^{(1)}; Z^{(1)}|X^{(2)}) &= \iiint p(x^{(1)}, x^{(2)}, z^{(1)}) \log \frac{p(z^{(1)}, x^{(1)}|x^{(2)})}{p(z^{(1)}|x^{(2)})p(x^{(1)}|x^{(2)})} \, dx^{(1)}dx^{(2)}dz^{(1)} \\
&= \iiint p(x^{(1)}, x^{(2)}, z^{(1)}) \log \frac{p(z^{(1)}|x^{(1)}, x^{(2)})p(x^{(1)}|x^{(2)})}{p(z^{(1)}|x^{(2)})p(x^{(1)}|x^{(2)})} \, dx^{(1)}dx^{(2)}dz^{(1)} \\
&= \iiint p(x^{(1)}, x^{(2)}, z^{(1)}) \log \frac{p(z^{(1)}|x^{(1)})}{p(z^{(1)}|x^{(2)})} \, dx^{(1)}dx^{(2)}dz^{(1)} \\
&= \iiint p(x^{(1)}, x^{(2)}, z^{(1)}) \log \frac{p(z^{(1)}|x^{(1)})p(z^{(2)}|x^{(2)})}{p(z^{(1)}|x^{(2)})p(z^{(2)}|x^{(2)})} \, dx^{(1)}dx^{(2)}dz^{(1)} \\
&= \iint p(x^{(1)}, x^{(2)}) D_{KL}(p(z^{(1)}|x^{(1)}) \| p(z^{(2)}|x^{(2)})) \, dx^{(1)}dx^{(2)} \\
&\quad - \int p(x^{(2)}) D_{KL}(p(z^{(1)}|x^{(2)}) \| p(z^{(2)}|x^{(2)})) \, dx^{(2)} \\
&\leq \iint p(x^{(1)}, x^{(2)}) D_{KL}(p(z^{(1)}|x^{(1)}) \| p(z^{(2)}|x^{(2)})) \, dx^{(1)}dx^{(2)}.
\end{aligned}
\tag{B.5}
$$

Thus, $I(X^{(1)}; Z^{(1)}|X^{(2)})$ can be upper-bounded by $D_{KL}(p(z^{(1)}|x^{(1)}) \| p(z^{(2)}|x^{(2)}))$ integrated over $x^{(1)}, x^{(2)}$. We can approximate this KL divergence by:

$$
\begin{aligned}
D_{KL}(p(y|z^{(1)}) \| p(y|z^{(2)})) &= \int p(y|z^{(1)}) \log \frac{p(y|z^{(1)})}{p(y|z^{(2)})} \, dy \\
&= \int p(y|z^{(1)}) \log p(y|z^{(1)}) \, dy - \int p(y|z^{(1)}) \log p(y|z^{(2)}) \, dy,
\end{aligned}
\tag{B.6}
$$

where the first term of RHS of the last equation assumes to be constant, and the second term can be approximated by cross-entropy loss using Monte Carlo sampling similarly in **Proposition 1**. Therefore, **Proposition 2** holds. □

C. Comparison to MPANet Using the Same Baseline

We conduct an additional experiment to compare the performances of our method and others using our baseline. We choose MPANet [2], which performed the highest accuracy on SYSU-MM01 among current open-source works. We reproduce it on our baseline. It is demonstrated in Table S1 that our method achieves higher performances compared to MPANet using the same baseline.

Method	SYSU-MM01		HITSZ-VCM			
			Infrared-Visible		Visible-Infrared	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
MPANet	70.58	68.24	46.51	35.26	50.32	37.80
on our base	71.39	67.77	58.46	45.69	61.01	46.98
Ours	75.18	70.12	67.65	52.30	70.23	52.54

Table S1. Reproduce MPANet on our baseline. All Hyper-parameters have been carefully tuned.

References

[1] Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. In *International Conference on Learning Representations*, 2020. 1

[2] Qiong Wu, Pingyang Dai, Jie Chen, Chia-Wen Lin, Yongjian Wu, Feiyue Huang, Bineng Zhong, and Rongrong Ji. Discover cross-modality nuances for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4330–4339, June 2021. 3