SUPPLEMENTARY MATERIAL FOR MULTI-GRANULARITY SELF-LEARNING GUIDED DEEP MULTI-VIEW CLUSTERING

Zhikui Chen[†⋆] *Yuzhe Li*[†] *Yu Zhou*[†]

[†] School of Software Technology, Dalian University of Technology, Dalian, China

1. INTRODUCTION

In this supplementary material, we provide the detailed derivation of loss $\mathcal{L}_{sic}$, $\mathcal{L}_{rci}$, and $\mathcal{L}_{cca}$.

2. NOTATIONS AND DEFINITIONS

Table 1 presents the mathematical notations and the corresponding definitions.

Table 1: Notation and definition.

Notation	Definition
$\mathbf{X}^v$	set of samples in the v-th view
$\mathbf{Z}^v$	set of latent representations in the v-th view
$\mathbf{Z}$	latent fusion representation
$\mathbf{Q}^v$	cluster assignment in the v-th view
$\mathbf{P}$	unified cluster assignment
$\mathbf{p}_i$	the i-th row of $\mathbf{P}$
$\mathbf{p}_{:,i}$	the i-th column of $\mathbf{P}$

3. SAMPLE-GUIDED INFORMATION COMPRESSION LOSS

The definition of $D_{KL}\left(P(\mathbf{X}^v|\mathbf{Z}^v)||D(\mathbf{X}^v|\mathbf{Z}^v)\right)$ is as follow:

$$\begin{aligned}
&D_{KL}\left(P(\mathbf{X}^v|\mathbf{Z}^v)||D(\mathbf{X}^v|\mathbf{Z}^v)\right)\\
&=\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log\frac{p(\mathbf{x}^v|\mathbf{z}^v)}{q(\mathbf{x}^v|\mathbf{z}^v)}d\mathbf{x}^v\\
&=\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log p(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v-\\
&\qquad\qquad\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log q(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v
\end{aligned} \tag{1}$$

where q denotes the probability density function of distribution D. Since the non-negativity of the KL divergence, the following inequality holds:

$$\begin{aligned}
&\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log p(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v\\
&\geq \int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log q(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v
\end{aligned} \tag{2}$$

Therefore, the mutual information $I(\mathbf{X}^v;\mathbf{Z}^v)$ is:

$$\begin{aligned}
I(\mathbf{X}^v;\mathbf{Z}^v) &= \int_{\mathbf{Z}^v}\int_{\mathbf{X}^v} p\left(\mathbf{x}^v,\mathbf{y}^v\right)\log\frac{p(\mathbf{x}^v,\mathbf{z}^v)}{p(\mathbf{x}^v)p(\mathbf{z}^v)}d\mathbf{x}^v d\mathbf{z}^v\\
&= H(\mathbf{X}^v)+\int_{\mathbf{Z}^v} p(\mathbf{z}^v)\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log p(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v d\mathbf{z}^v\\
&\geq \int_{\mathbf{Z}^v} p(\mathbf{z}^v)\int_{\mathbf{X}^v} p(\mathbf{x}^v|\mathbf{z}^v)\log q(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{x}^v d\mathbf{z}^v\\
&= \int_{\mathbf{X}^v} p(\mathbf{x}^v)\int_{\mathbf{Z}^v} p(\mathbf{z}^v|\mathbf{x}^v)\log q(\mathbf{x}^v|\mathbf{z}^v)d\mathbf{z}^v d\mathbf{x}^v\\
&= \mathbb{E}_{P(\mathbf{x}^v,\mathbf{z}^v)}\log D(\mathbf{x}^v|\mathbf{z}^v)
\end{aligned} \tag{3}$$

where $H(\mathbf{X}^v)$ only depends on $P(\mathbf{X})$, which can be ignored. if we assume that $D(\mathbf{X}^v|\mathbf{Z}^v) = \mathcal{N}\left(\mu(\mathbf{X}^v|\mathbf{Z}^v),\sigma^2 I\right)$, we have:

$$\begin{aligned}
&I(\mathbf{X}^v;\mathbf{Z}^v)\\
&=\mathbb{E}_{P(\mathbf{x}^v,\mathbf{z}^v)}\left[\log\left(\frac{1}{(2\pi\sigma^2)^{d_x^v/2}}e^{-\frac{\|\mathbf{x}^v-\mu^v(\mathbf{z})\|^2}{2\sigma^2}}\right)\right]
\end{aligned} \tag{4}$$

where d_x^v is the dimension of $\mathbf{X}^v$.

4. REPRESENTATION-GUIDED COMPLEMENTARY INTEGRATION LOSS

Similarly, because of the non-negativity of KL divergence, we have:

$$\begin{aligned}
H(\mathbf{Z}^v|\mathbf{Z}) &= -\int_{\mathbf{Z}}\int_{\mathbf{Z}^v} p(\mathbf{z}^v,\mathbf{z})\log p(\mathbf{z}^v|\mathbf{z})d\mathbf{z}^v d\mathbf{z}\\
&= -\mathbb{E}_{p(\mathbf{z}^v,\mathbf{z})}[\log p(\mathbf{z}^v|\mathbf{z})]\\
&= -\mathbb{E}_{p(\mathbf{z}^v,\mathbf{z})}[\log R(\mathbf{z}^v|\mathbf{z})] + D_{KL}(\mathcal{P}(\mathbf{Z}^v|\mathbf{Z})||R(\mathbf{Z}^v|\mathbf{Z}))\\
&\leq -\mathbb{E}_{P(\mathbf{z}^v,\mathbf{z})}[\log R(\mathbf{z}^v|\mathbf{z})]\\
&= -\mathbb{E}_{P(\mathbf{z}^v,\mathbf{z})}\left[\log\left(\frac{1}{(2\pi\sigma^2)^{d_z^v/2}}e^{-\frac{\|\mathbf{z}^v-G^v(\mathbf{z})\|^2}{2\sigma^2}}\right)\right]
\end{aligned} \tag{5}$$

where $R(\mathbf{Z}^v|\mathbf{Z})$ is assumed follows a Gaussian distribution $\mathcal{N}(\mathbf{Z}^v|G^v(\mathbf{Z}),\sigma^2\mathbf{I})$.

5. CLUSTER-GUIDED CONSISTENT ALIGNMENT LOSS

We take the cluster-level alignment as an example to derive the solution of $I(\mathbf{P}^\top, (\mathbf{Q}^v)^\top)$. Since $\mathbf{p}_{:,i}$ and $\mathbf{q}^v_{:,i}$ are positive pair, $p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i}) > p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,i})$. Meanwhile, different samples are independent of each other, $p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,j}) = p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,j})$, when $i \neq j$. Since $p(\mathbf{q}^v_{:,i}|\mathbf{p}_{:,i}) > 0$, there exists a positive constant, such that $p(\mathbf{q}^v_{:,i}|\mathbf{p}_{:,i}) > \delta$. For simplicity, we take:

$$r_{ij} = \frac{p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,j})}{p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,j})} \tag{6}$$

Assuming a uniform prior over K prototypes, we have $p(\mathbf{p}_{:,i}) \approx 1/K$. Furthermore, according to previous work, we have $e^{sim(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i})/\tau} \propto r_{ij}$. Therefore, we have:

$$\begin{aligned}
&\frac{1}{V}\sum_{v=1}^{V} I(\mathbf{P}^\top, (\mathbf{Q}^v)^\top) \\
&= \frac{1}{V}\sum_{v=1}^{V}\sum_{i=1}^{K}\sum_{j=1}^{K} p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,j}) \log \frac{p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,j})}{p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,j})} \\
&= \frac{1}{V}\sum_{v=1}^{V}\sum_{i=1}^{K} p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i}) \log \frac{p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i})}{p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,i})} + 0 \\
&\approx \frac{1}{KV}\sum_{v=1}^{V}\sum_{i=1}^{K} p(\mathbf{q}^v_{:,i}|\mathbf{p}_{:,i}) \log \frac{p(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i})}{p(\mathbf{p}_{:,i})p(\mathbf{q}^v_{:,i})} \\
&\geq \frac{\delta}{KV}\sum_{v=1}^{V}\sum_{i=1}^{K} \log \frac{r_{ii}}{\sum_{j=1}^{K} r_{ij}} + \frac{\delta}{K}\sum_{v=1}^{V}\sum_{i=1}^{K} \log \sum_{j=1}^{K} r_{ij} \\
&= -\delta\mathcal{L} + \frac{\delta}{K}\sum_{v=1}^{V}\sum_{i=1}^{K} \log\left(K - 1 + r_{ii}\right) \\
&\geq -\delta\mathcal{L} + \delta \log K
\end{aligned} \tag{7}$$

where $\mathcal{L}$ is:

$$\mathcal{L} = -\frac{1}{KV}\sum_{v=1}^{V}\sum_{i=1}^{K} \log \frac{e^{sim(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,i})/\tau}}{\sum_{j=1}^{K} e^{sim(\mathbf{p}_{:,i}, \mathbf{q}^v_{:,j})/\tau}} \tag{8}$$

where τ is the temperature parameter.