Supplementary Materials of "Regularized Contrastive Partial Multi-view Outlier Detection"

Anonymous Authors

1 DETAILED PROOF OF PROPOSITION 3.1

Proof. Note that the negative samples could be chosen from both inliers and outliers, the loss values of class-related outliers can be formulated as:

$$\begin{aligned} \mathcal{L}^o_{con} &= -\frac{1}{2}\sum_{m=1}^{2}\mathbb{E}_{z_o\in\mathcal{Z}_o}\log\left(\frac{f(z_o^{(m)}, z_o^{(m')})}{S_{in}+S_{out}}\right), \\ S_{in} &= \sum_{z_{in}\in\mathcal{Z}_{in}}\sum_{v=1}^{2} f(z_o^{(m)}, z_{in}^{(v)}), \\ S_{out} &= \sum_{z_{out}\in\mathcal{Z}_o}\sum_{v=1}^{2} f(z_o^{(m)}, z_{out}^{(v)}), \end{aligned} \tag{1}$$

where $\mathcal{Z}_o, \mathcal{Z}_{in}$ denote the latent representation sets of class-related outlier and inlier instances, respectively; $f(x, y)$ is defined as $e^{s(x,y)}/\tau_F$ in practice. According to [1], the optimal $f(x, y)$ satisfies:

$$f(x,y) = \frac{P(y|x)}{P(y)}. \tag{2}$$

Then, we have:

$$\begin{aligned} S_{in} &= \sum_{v=1}^{2}\sum_{z_{in}\in\mathcal{Z}_{in}}\frac{P(z_{in}^{(v)}|z_o^{(m)})}{P(z_{in}^{(v)})} \\ &\approx \sum_{v=1}^{2}|\mathcal{Z}_{in}|\cdot E_{z_{in}\in\mathcal{Z}_{in}}\frac{P(z_{in}^{(v)}|z_o^{(m)})}{P(z_{in}^{(v)})} \\ &= 2|\mathcal{Z}_{in}|, \end{aligned} \tag{3}$$

$$\begin{aligned} S_{out} =& f(z_o^{(m)}, z_o^{(m')}) + f(z_o^{(m)}, z_o^{(m)}) \\ &+ \sum_{v=1}^{2}\sum_{z_{out}\in\mathcal{Z}_o/\{z_o\}}\frac{P(z_{out}^{(v)}|z_o^{(m)})}{P(z_{out}^{(v)})} \\ \approx& f(z_o^{(m)}, z_o^{(m')}) + 2|\mathcal{Z}_o| - 1. \end{aligned} \tag{4}$$

When the size of $\mathcal{Z}_o$ and $\mathcal{Z}_{in}$ is sufficiently large, the approximation error approaches zero. Thus,

$$\begin{aligned} \mathcal{L}^o_{con} &= \frac{1}{2}\sum_{m=1}^{2}\mathbb{E}_{z_o\in\mathcal{Z}_o}\log\left(\frac{S_{in}+S_{out}}{f(z_o^{(m)}, z_o^{(m')})}\right) \\ &= \frac{1}{2}\sum_{m=1}^{2}\mathbb{E}_{z_o\in\mathcal{Z}_o}\log\left(1+\frac{P(z_o^{(m')})}{P(z_o^{(m')}|z_o^{(m)})}(2|\mathcal{Z}_{in}|+2|\mathcal{Z}_o|-1)\right) \\ &= \frac{1}{2}\sum_{m=1}^{2}\mathbb{E}_{z_o\in\mathcal{Z}_o}\log\left(1+\frac{P(z_o^{(m')})}{P(z_o^{(m')}|z_o^{(m)})}(2N-1)\right) \\ &\geq \frac{1}{2}\sum_{m=1}^{2}\mathbb{E}_{z_o\in\mathcal{Z}_o}\log\left(\frac{P(z_o^{(m')})}{P(z_o^{(m')}|z_o^{(m)})}\cdot 2N\right) \\ &= -I\left(z_o^{(1)}, z_o^{(2)}\right) + \log 2N, \end{aligned} \tag{5}$$

where the last line uses the fact $I(z_o^{(1)}, z_o^{(2)}) = I(z_o^{(2)}, z_o^{(1)})$.

According to the assumption made in the main paper:

$$I(\boldsymbol{x}_o^{(1)}, \boldsymbol{x}_o^{(2)}) \leq \varepsilon. \tag{6}$$

Meanwhile, by the data processing inequality, we have:

$$I(z_o^{(1)}, z_o^{(2)}) \leq I(\boldsymbol{x}_o^{(1)}, \boldsymbol{x}_o^{(2)}) \leq \varepsilon. \tag{7}$$

Combining the above results, we can obtain:

$$\mathcal{L}^o_{con} \geq \log(2N) - I(z_o^{(1)}, z_o^{(2)}) \geq \log(2N) - \varepsilon. \tag{8}$$

□

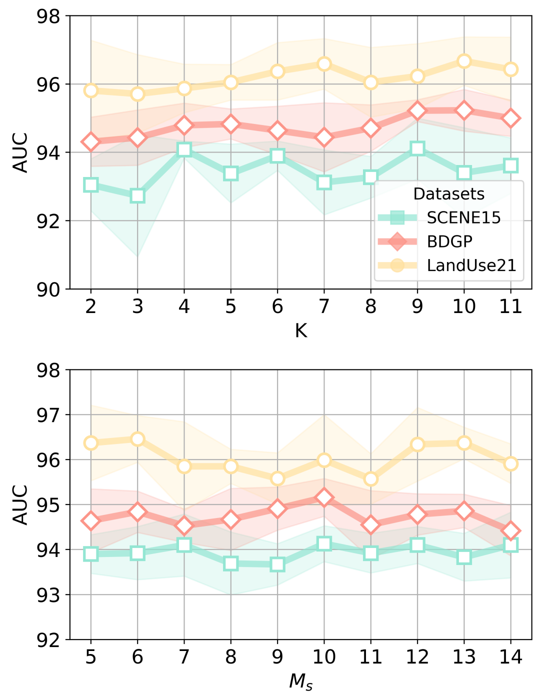

Figure 1: Sensitivity analysis on K and M_s.

2 ADDITIONAL SENSITIVITY ANALYSIS

There are some other less important hyperparameters in the proposed framework, *e.g.*, the number of neighbors K used in K-NN and the size of the memory bank N_M. As the memory bank is updated by adding potential outliers in the current batch and removing the potential outliers in the oldest batch in the memory bank in practice, we analyze M_s, the number of batches restored in the memory bank, instead. We fix the missing rate as 0.3 and the outlier ratio of all types of outliers as 0.05, and then evaluate RCPMOD with different values of K and M_s. As shown in Figure 1, RCPMOD is not very sensitive to these two hyperparameters.

REFERENCES

[1] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2019. Representation Learning with Contrastive Predictive Coding. arXiv:1807.03748