Reconstruction Weighting Principal Component Analysis with Fusion Contrastive Learning

(Supplementary Material)

1 Optimization

Step1: Fix m and W, and update r:

Because $\mathbf{m}$ and $\mathbf{W}$ are fixed, we assume $l_i = ||\mathbf{x}_i - \mathbf{m} - \mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{m})||_2^2$, then the objective function can be rewritten as:

$$arg\min_{r_i} \sum_{i=1}^{N} \frac{l_i}{r_i}, s.t. \sum_{i=1}^{N} r_i = 1, 0 \leqslant r_i \leqslant 1 \tag{1}$$

The Lagrangian function of Eq.(1) is below:

$$\mathcal{L} = \sum_{i=1}^{N} \frac{l_i}{r_i} + \beta(1 - \sum_{i=1}^{N} r_i) + \sum_{i=1}^{N} \gamma_i(-r_i) \tag{2}$$

where β and γ_i are Langrange multipliers. The optimal solution of r_i can be obtained as:

$$r_i = \frac{||\mathbf{x}_i - \mathbf{m} - \mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{m})||_2}{\sum\limits_{i=1}^{N} ||\mathbf{x}_i - \mathbf{m} - \mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{m})||_2} \tag{3}$$

Step2: Fix r and W, and update m:

The optimized objective function is:

$$\begin{aligned} m^* &= arg\min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} ||\mathbf{x}_i - \mathbf{m} - \mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{m})||_2^2 \\ &= arg\min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} ||\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{m}||_2^2 \end{aligned} \tag{4}$$

Where $\mathbf{A} = \mathbf{I} - \mathbf{W}\mathbf{W}^\mathbf{T}$.

Eq.(4) can be further simplified as:

$$\begin{aligned} &\min \sum_{i=1}^{N} \frac{1}{r_i} ||\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{m}||_2^2 \\ &= \min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} tr(\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{m})^T(\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{m}) \\ &= \min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} \{tr(\mathbf{m}^T\mathbf{A}^T\mathbf{A}\mathbf{m}) - 2tr(\mathbf{m}^T\mathbf{A}^T\mathbf{A}\mathbf{x}_i) \\ &+ tr({\mathbf{x}_i}^T\mathbf{A}^T\mathbf{A}\mathbf{x}_i)\} \end{aligned} \tag{5}$$

Therefore, the optimized objective function is further changed to:

$$\mathbf{m}^* = arg\min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} \{tr(\mathbf{m}^T\mathbf{A}^T\mathbf{A}\mathbf{m}) - 2tr(\mathbf{m}^T\mathbf{A}^T\mathbf{A}\mathbf{x}_i)\} \tag{6}$$

We find that:

$$\begin{aligned} &\mathbf{A}^T\mathbf{A} = (\mathbf{I} - \mathbf{W}\mathbf{W}^T)^T(\mathbf{I} - \mathbf{W}\mathbf{W}^T) \\ &= \mathbf{I} - \mathbf{W}\mathbf{W}^T \\ &= \mathbf{A} \end{aligned} \tag{7}$$

Then, the final objective function is:

$$\mathbf{m}^* = arg\min_{\mathbf{m}} \sum_{i=1}^{N} \frac{1}{r_i} \{tr(\mathbf{m}^T\mathbf{A}\mathbf{m}) - 2tr(\mathbf{m}^T\mathbf{A}\mathbf{x}_i)\} \tag{8}$$

Let the derivative of Eq.(8) with respect to m equals 0:

$$\mathbf{m}^* = \frac{\sum\limits_{i=1}^{N} \frac{1}{r_i}\mathbf{x}_i}{\sum\limits_{i=1}^{N} \frac{1}{r_i}} \tag{9}$$

Step3: Fix r and m, update W:

To facilitate the solution, we divide the objective function into two parts. $\mathbf{R}$ represents the reconstruction error, $\mathbf{L}$ represents the contrastive loss, and the total loss is $\mathbf{A} = \mathbf{R} + \mathbf{L}$.

Calculate the derivative of R with respect to W:

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{W}} \left(||\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}||_F^2 \right) \\ &= \frac{\partial}{\partial \mathbf{W}} \left(tr(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X})^T(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}) \right) \\ &= \frac{\partial}{\partial \mathbf{W}} \left(tr(\mathbf{X}\mathbf{X}^T - 2\mathbf{X}^T\mathbf{W}\mathbf{W}^T\mathbf{X} + \mathbf{X}^T\mathbf{W}\mathbf{W}^T\mathbf{W}\mathbf{W}^T\mathbf{X}) \right) \\ &= \frac{\partial}{\partial \mathbf{W}} \left(tr(-2\mathbf{X}^T\mathbf{W}\mathbf{W}^T\mathbf{X} + \mathbf{X}^T\mathbf{W}\mathbf{W}^T\mathbf{X}) \right) \\ &= \frac{\partial}{\partial \mathbf{W}} \left(tr(-\mathbf{X}^T\mathbf{W}\mathbf{W}^T\mathbf{X}) \right) \\ &= -2\mathbf{X}\mathbf{X}^T\mathbf{W} \end{aligned} \tag{10}$$

Calculate the derivative of L with respect to W. Review the objective function of contrastive loss:

$$L=\frac{1}{N}\sum_{i=1}^{N}-\log\frac{\exp(s_1/\tau)}{\sum\limits_{j=1,j\neq k}^{N}[\exp(s_2/\tau)+\exp(s_3/\tau)]} \tag{11}$$

s is the cosine similarity function. s_1, s_2, and s_3 are as follows:

$$\begin{aligned} s_1 &= s\left(\mathbf{b}_i,\mathbf{a}_i\right)=s((\mathbf{W}\mathbf{W}^T\mathbf{x}_i)^T,{\mathbf{x}_i}^T) \\ &= \frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{x}_i\|} \end{aligned} \tag{12}$$

$$\begin{aligned} s_2 &= s\left(\mathbf{b}_i,\mathbf{b}_j\right)=s((\mathbf{W}\mathbf{W}^T\mathbf{x}_i)^T,(\mathbf{W}\mathbf{W}^T\mathbf{x}_j)^T) \\ &= \frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|} \end{aligned} \tag{13}$$

$$\begin{aligned} s_3 &= s\left(\mathbf{b}_i,\mathbf{a}_j\right)=s((\mathbf{W}\mathbf{W}^T\mathbf{x}_i)^T,{\mathbf{x}_j}^T) \\ &= \frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{x}_j\|} \end{aligned} \tag{14}$$

Calculate the derivatives of s_1, s_2, s_3 with respect to W:

$$\begin{aligned} \frac{\partial s_1}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}}\left\{\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\|\mathbf{x}_i\|}\right\} \\ &= \frac{1}{\|\mathbf{x}_i\|}\frac{\partial}{\partial \mathbf{W}}\left(\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\right) \\ &= \frac{1}{\|\mathbf{x}_i\|}\frac{\partial({\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i)}{\partial \mathbf{W}}\cdot\frac{1}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|} \\ &+ {\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i\cdot\frac{\partial}{\partial \mathbf{W}}\left(\frac{1}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\right) \\ &= \frac{1}{\|\mathbf{x}_i\|}\left(\frac{2\mathbf{x}_i{\mathbf{x}_i}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}-\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i\cdot\mathbf{x}_i{\mathbf{x}_i}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|^3}\right) \\ &= \frac{1}{\|\mathbf{x}_i\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\left(2\mathbf{I}-\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_i}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|^3}\right)\mathbf{x}_i{\mathbf{x}_i}^T\mathbf{W} \end{aligned} \tag{15}$$

$$\begin{aligned} \frac{\partial s_2}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}}\left\{\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|}\right\} \\ &= \frac{\partial({\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j)}{\partial \mathbf{W}}\cdot\frac{1}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|} \\ &+ {\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j\cdot\frac{\partial}{\partial \mathbf{W}}\left(\frac{1}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|}\right) \\ &= \frac{\mathbf{x}_i{\mathbf{x}_j}^T\mathbf{W}+\mathbf{x}_j{\mathbf{x}_i}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|} \\ &- {\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j\cdot\frac{\mathbf{x}_i{\mathbf{x}_i}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|^3\|\mathbf{W}\mathbf{W}^T\mathbf{x}_j\|} \\ &- {\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j\cdot\frac{\mathbf{x}_j{\mathbf{x}_j}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|^3} \end{aligned} \tag{16}$$

Algorithm 1 The optimization algorithm of our model

Input: Image dataset $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N]$, the dimension of projection c;
Parameter: Projection matrix $\mathbf{W}\in\mathbb{R}^{d\times c}$, mean matrix $\mathbf{m}$, adaptive weight $\mathbf{r}$, contrastive loss weight λ, temperature τ, pseudo negative sample rate η, index of pseudo negative samples $\mathbf{k}$;
Output: Optimal projection matrix$\mathbf{W}^*\in\mathbb{R}^{d\times c}$;

Let $\lambda=0.0001$, $\tau=0.1$, $\eta=0.7$;
while not converge **do**
  Update the $\mathbf{r}$, $\mathbf{m}$, $\mathbf{k}$;
  Calculate the derivative of s_1, s_2, s_3 according to Eq. 15, 16, 17;
  Calculate the derivative of contrastive learning loss L with Eq.18;
  Calculate the derivative of reconstruction error R with Eq. 10;
  Optimize the projection matrix $\mathbf{W}^*$ with Eq. 19.
end while
return $\mathbf{W}^*$

$$\begin{aligned} \frac{\partial s_3}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}}\left\{\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|\|\mathbf{x}_j\|}\right\} \\ &= \frac{1}{\|\mathbf{x}_j\|}\frac{\partial}{\partial \mathbf{W}}\left\{\frac{{\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\right\} \\ &= \frac{1}{\|\mathbf{x}_j\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\left(\mathbf{x}_i{\mathbf{x}_j}^T\mathbf{W}+\mathbf{x}_j{\mathbf{x}_i}^T\mathbf{W}\right) \\ &- \frac{1}{\|\mathbf{x}_j\|\,\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|}\left({\mathbf{x}_i}^T\mathbf{W}\mathbf{W}^T\mathbf{x}_j\cdot\frac{\mathbf{x}_i{\mathbf{x}_i}^T\mathbf{W}}{\|\mathbf{W}\mathbf{W}^T\mathbf{x}_i\|^2}\right) \end{aligned} \tag{17}$$

Then the derivative of L with respect to W is:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{W}} &= \frac{1}{N}\sum_{i=1}^{N}\frac{1/\tau\sum\limits_{j=1}^{N}\left(\exp(s_2/\tau)\dfrac{\partial s_2}{\partial \mathbf{W}}+\exp(s_3/\tau)\dfrac{\partial s_3}{\partial \mathbf{W}}\right)}{\sum\limits_{j=1}^{N}(\exp(s_2/\tau)+\exp(s_3/\tau))} \\ &- \frac{1}{N}\sum_{i=1}^{N}1/\tau\frac{\partial s_1}{\partial \mathbf{W}} \end{aligned} \tag{18}$$

Calculate the derivative of A with respect to W. Combine Eq.(10) and Eq.(18), the derivative of total loss to W is:

$$\begin{aligned} \frac{\partial A}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}}\left(||\mathbf{X}-\mathbf{W}\mathbf{W}^T\mathbf{X}||_F^2\right)+\lambda\frac{\partial L}{\partial \mathbf{W}} \\ &= -2\mathbf{X}\mathbf{X}^T\mathbf{W}+\lambda\frac{\partial L}{\partial \mathbf{W}} \end{aligned} \tag{19}$$

Algorithm 1 provides pseudocode for solving the objective function.