A Proofs of Theoretical Results

A.1 Details of Eq. (10)

Eq. (10). Let network $g_\phi(\cdot)$ to be Lipschitz continuity mapping. $d^z_{ij} = k^* d^y_{ij}, k^* \in [1/K, K]$, where K is a Lipschitz constant. For neighborhoods points i and j, The difference between DLME loss L_{D} and contrastive learning loss L_c is

$$|L_{\mathrm{D}} - L_{\mathrm{c}}| = \mathbb{E}_{x_j,x_j}\left[A_{ij} - \kappa\left((1+(\alpha-1)A_{ij})k^* d^z_{ij}\right)\log\left(\frac{1}{\kappa(d^z_{ij})} - 1\right)\right], \tag{11}$$

Detail. The contrastive learning loss is written as, For notational simplicity, we omit redundant symbols:

$$\begin{aligned} L_c = & \mathbb{E}_{x_j,x_j}\left[B_{ij}\log\kappa\left(d^e_{ij},\nu\right) + (1-B_{ij})\log\left(1-\kappa\left(d^e_{ij},\nu\right)\right)\right] \\ = & \mathbb{E}_{x_j,x_j}\left[B_{ij}\log\kappa\left(d^e_{ij}\right) + (1-B_{ij})\log\left(1-\kappa\left(d^e_{ij}\right)\right)\right] \end{aligned} \tag{12}$$

where $B_{ij} = \pi[(x_i, x_j) \in E]$ shows weath i and j are neighborhoods in graph $G(X, E)$.

the DLME loss is written as, we :

$$\begin{aligned} L_{\mathrm{D}} = & E_{x_i,x_j}\left[\mathcal{D}\left(\kappa\left(R_{ij}d^y_{ij},\nu\right), \kappa\left(d^z_{ij},\nu\right)\right)\right] \\ = & E_{x_i,x_j}\left[\mathcal{D}\left(\kappa\left(R_{ij}d^y_{ij}\right), \kappa\left(d^z_{ij}\right)\right)\right] \end{aligned} \tag{13}$$

We substitute Eq. (5) into Eq. (14).

$$L_{\mathrm{D}} = E_{x_i,x_j}\left[\kappa\left(R_{ij}d^y_{ij}\right)\log\kappa\left(d^z_{ij}\right) + \left(1-\kappa\left(R_{ij}d^y_{ij}\right)\right)\log\left(1-\kappa\left(d^z_{ij}\right)\right)\right] \tag{14}$$

The difference between the two loss functions is:

$$
\begin{aligned}
&L_c - L_{\mathrm{D}} \\
=&\mathbb{E}_{x_j,x_j}\left[B_{ij}\log\kappa\left(d_{ij}^z\right)+(1-B_{ij})\log\left(1-\kappa\left(d_{ij}^z\right)\right)\right]- \\
&\mathbb{E}_{x_j,x_j}\left[\kappa\left(R_{ij}d_{ij}^y\right)\log\kappa\left(d_{ij}^z\right)+\left(1-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\left(1-\kappa\left(d_{ij}^z\right)\right)\right] \\
=&\mathbb{E}_{x_j,x_j}\left[B_{ij}\log\kappa(d_{ij}^z)+(1-B_{ij})\log\left(1-\kappa\left(d_{ij}^z\right)\right)-\kappa\left(R_{ij}d_{ij}^y\right)\log\kappa\left(d_{ij}^z\right)-\left(1-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\left(1-\kappa\left(d_{ij}^z\right)\right)\right] \\
=&\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\kappa(d_{ij}^z)+\left(1-B_{ij}-1+\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\left(1-\kappa\left(R_{ij}d_{ij}^z\right)\right)\right] \\
=&\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\kappa(d_{ij}^z)+\left(\kappa\left(R_{ij}d_{ij}^y\right)-B_{ij}\right)\log\left(1-\kappa\left(d_{ij}^z\right)\right)\right] \\
=&\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left(R_{ij}d_{ij}^y\right)\right)\left(\log\kappa(d_{ij}^z)-\log\left(1-\kappa\left(d_{ij}^z\right)\right)\right)\right] \\
=&-\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\left(\frac{1-\kappa\left(d_{ij}^z\right)}{\kappa\left(d_{ij}^z\right)}\right)\right] \\
=&-\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left(R_{ij}d_{ij}^y\right)\right)\log\left(\frac{1}{\kappa\left(d_{ij}^z\right)}-1\right)\right]
\end{aligned}
\tag{15}
$$

Substituting the relationship between B_{ij} and R_{ij}, $R_{ij}=1+(\alpha-1)B_{ij}$, we have

$$
L_c - L_{\mathrm{D}} = -\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left((1+(\alpha-1)B_{ij})d_{ij}^y\right)\right)\log\left(\frac{1}{\kappa\left(d_{ij}^z\right)}-1\right)\right] \tag{16}
$$

We assume that network $g_\phi(\cdot)$ to be a Lipschitz continuity function, then

$$
\begin{aligned}
\frac{1}{K}g_\phi(d_{ij}^y) \le d_{ij}^y \le K g_\phi(d_{ij}^y) \quad \forall i,j \in \{1,2,\cdots,N\} \\
\frac{1}{K}d_{ij}^z \le d_{ij}^y \le K d_{ij}^z \quad \forall i,j \in \{1,2,\cdots,N\}
\end{aligned}
\tag{17}
$$

We construct the inverse mapping of $g_\phi(\cdot) : g_\phi^{-1}(\cdot)$

$$
\frac{1}{K}d_{ij}^z \le d_{ij}^y \le K d_{ij}^z \quad \forall i,j \in \{1,2,\cdots,N\} \tag{18}
$$

then exit k^*, let:

$$
d_{ij}^y = k^* d_{ij}^z \quad k^* \in [1/K, K] \quad \forall i,j \in \{1,2,\cdots,N\} \tag{19}
$$

Substituting the Eq.(19) into Eq.(16).

$$
L_c - L_{\mathrm{D}} = -\mathbb{E}_{x_j,x_j}\left[\left(B_{ij}-\kappa\left((1+(\alpha-1)B_{ij})k^* d_{ij}^z\right)\right)\log\left(\frac{1}{\kappa\left(d_{ij}^z\right)}-1\right)\right] \tag{20}
$$

if $B_{ij}=1$, we have:

$$L_c - L_{\mathrm{D}} \quad |_{B_{ij}=1} = -\mathbb{E}_{x_j,x_j}\left[\left(1-\kappa\left(\alpha k^* d_{ij}^z\right)\right)\log\left(\frac{1}{\kappa\left(d_{ij}^z\right)}-1\right)\right] \tag{21}$$

then:

$$\begin{aligned}&\lim_{\alpha\to 0} L_c - L_{\mathrm{D}} \quad |_{B_{ij}=1}\\ =& -\mathbb{E}_{x_j,x_j}\left[\left(1-\kappa\left(\alpha k^* d_{ij}^z\right)\right)\log\left(\frac{1}{\kappa\left(d_{ij}^z\right)}-1\right)\right]\\ =&0\end{aligned} \tag{22}$$

Based on Eq.(22), we find that if i,j is neighbor and $\alpha \to 0$, there is no difference between the two loss functions L_c and L_{D}. When $\alpha \nrightarrow 0$, the difference between the loss functions will be the function of d_{ij}^z. Because the contrastive learning loss L_c only minimizes the distance between adjacent nodes and does not maintain any structural information. We believe that the loss of DLME will better preserve the structural information based on contrastive loss.

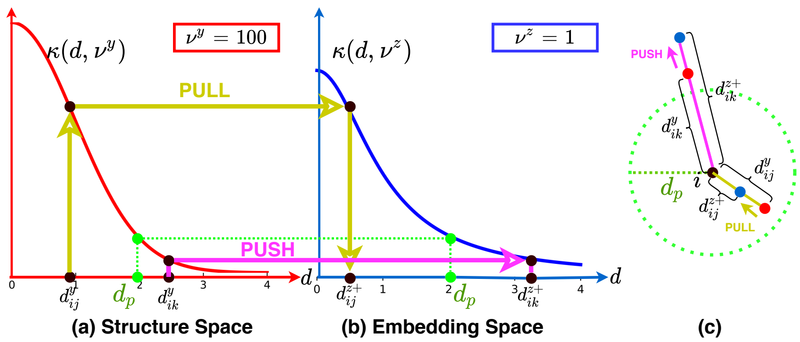

Fig. 8. Proof. of Lemma 2

A.2 Proof of Lemma 1

Lemma 1 (Push-pull property). let $\nu^y < \nu^z$ and let $d^{z+} = \kappa^{-1}(\kappa(d,\nu^y),\nu^z)$ be the solution of minimizing L_{D}. Then exists d_p so that $(d^y - d_p)(d^{z+} - d^y) > 0$.

Proof.

$$L_{\mathrm{D}} = E_{x_i,x_j}\left[\kappa\left(d^y,\nu^y\right)\log\kappa\left(d^z_{ij},\nu^z\right) + \left(1-\kappa\left(R_{ij}d^y_{ij},\nu^y\right)\right)\log\left(1-\kappa\left(d^z_{ij},\nu^z\right)\right)\right] \tag{23}$$

let $d^y = R_{ij}d^y_{ij}$ and let $d^z = d^z_{ij}$, we have,

$$L_{\mathrm{D}} = E_{x_i,x_j}\left[\kappa\left(d^y,\nu^y\right)\log\kappa\left(d^z,\nu^z\right) + \left(1-\kappa\left(d^y,\nu^y\right)\right)\log\left(1-\kappa\left(d^z,\nu^z\right)\right)\right] \tag{24}$$

then:

$$
\begin{aligned}
\frac{\partial L_{\mathrm{D}}}{\partial d^{z}} &= E_{x_i,x_j}\left[\kappa\left(d^{y},\nu^{y}\right)\frac{1}{\kappa\left(d_{ij}^{z},\nu^{z}\right)}\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}+\left(1-\kappa\left(d^{y},\nu^{y}\right)\right)\frac{-1}{1-\kappa\left(d_{ij}^{z},\nu^{z}\right)}\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}\right] \\
&= E_{x_i,x_j}\left[\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}\left(\frac{\kappa\left(d^{y},\nu^{y}\right)}{\kappa\left(d_{ij}^{z},\nu^{z}\right)}-\frac{1-\kappa\left(d^{y},\nu^{y}\right)}{1-\kappa\left(d_{ij}^{z},\nu^{z}\right)}\right)\right] \\
&= E_{x_i,x_j}\left[\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}\left(\frac{\kappa\left(d^{y},\nu^{y}\right)\left(1-\kappa\left(d_{ij}^{z},\nu^{z}\right)\right)-\kappa\left(d_{ij}^{z},\nu^{z}\right)\left(1-\kappa\left(d^{y},\nu^{y}\right)\right)}{\kappa\left(d_{ij}^{z},\nu^{z}\right)\left(1-\kappa\left(d_{ij}^{z},\nu^{z}\right)\right)}\right)\right] \\
&= E_{x_i,x_j}\left[\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}\left(\frac{\kappa\left(d^{y},\nu^{y}\right)-\kappa\left(d^{y},\nu^{y}\right)\kappa\left(d_{ij}^{z},\nu^{z}\right)-\kappa\left(d_{ij}^{z},\nu^{z}\right)+\kappa\left(d_{ij}^{z},\nu^{z}\right)\kappa\left(d^{y},\nu^{y}\right)}{\kappa\left(d_{ij}^{z},\nu^{z}\right)\left(1-\kappa\left(d_{ij}^{z},\nu^{z}\right)\right)}\right)\right] \\
&= E_{x_i,x_j}\left[\frac{\partial\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\partial d^{z}}\left(\frac{\kappa\left(d^{y},\nu^{y}\right)-\kappa\left(d_{ij}^{z},\nu^{z}\right)}{\kappa\left(d_{ij}^{z},\nu^{z}\right)\left(1-\kappa\left(d_{ij}^{z},\nu^{z}\right)\right)}\right)\right]
\end{aligned}
\tag{25}
$$

Because $\kappa\left(d,\nu^{y}\right)\in[0,1]$ is the kernel function of t-distribution. then they are Monotonically decreasing function in $d\in[0,+\infty]$ Then $\partial L_{\mathrm{D}}/\partial d^{z}$ is the Monotonically decreasing function and when $\kappa\left(d^{y},\nu^{y}\right)-\kappa\left(d_{ij}^{z},\nu^{z}\right)=0$ the $\partial L_{\mathrm{D}}/\partial d^{z}=0$.

We take the optimal solution as:

$$
d^{z+}=\kappa^{-1}\left(\kappa(d^{y},\nu^{y}),\nu^{z}\right) \tag{26}
$$

where $\kappa^{-1}(\cdot)$ is a inverse function of $\kappa(\cdot)$, and because $\kappa()$ is monotonically decreasing, the $\kappa^{-1}(\cdot)$ is monotonically decreasing.

Let $\Phi(d)=\kappa\left(d,\nu^{y}\right)-\kappa\left(d,\nu^{z}\right)$, we have

$$
\begin{aligned}
\lim_{d\to+\infty}\Phi(d) &= \lim_{d\to+\infty}\kappa\left(d,\nu^{y}\right)-\kappa\left(d,\nu^{z}\right) \\
&= \lim_{d\to+\infty}\frac{\operatorname{Gam}\left(\frac{\nu^{y}+1}{2}\right)}{\sqrt{\nu^{y}\pi}\operatorname{Gam}\left(\frac{\nu^{y}}{2}\right)}\left(1+\frac{d_{ij}^{2}}{\nu^{y}}\right)^{-\frac{\nu^{y}+1}{2}}-\lim_{d\to+\infty}\frac{\operatorname{Gam}\left(\frac{\nu^{z}+1}{2}\right)}{\sqrt{\nu^{z}\pi}\operatorname{Gam}\left(\frac{\nu^{z}}{2}\right)}\left(1+\frac{d_{ij}^{2}}{\nu^{z}}\right)^{-\frac{\nu^{z}+1}{2}} \\
&= 0^{-}<0
\end{aligned}
\tag{27}
$$

and

$$\begin{aligned}\lim_{d\to 0^+}\Phi(d) &= \lim_{d\to+\infty}\kappa\left(d,\nu^y\right)-\kappa\left(d,\nu^z\right)\\ &= \lim_{d\to 0^+}\frac{\operatorname{Gam}\left(\frac{\nu^y+1}{2}\right)}{\sqrt{\nu^y\pi}\operatorname{Gam}\left(\frac{\nu^y}{2}\right)}\left(1+\frac{d_{ij}^2}{\nu^y}\right)^{-\frac{\nu^y+1}{2}} - \lim_{d\to 0^+}\frac{\operatorname{Gam}\left(\frac{\nu^z+1}{2}\right)}{\sqrt{\nu^z\pi}\operatorname{Gam}\left(\frac{\nu^z}{2}\right)}\left(1+\frac{d_{ij}^2}{\nu^z}\right)^{-\frac{\nu^z+1}{2}}\\ &> 0\end{aligned} \tag{28}$$

And because the $\Phi(d)$ is a continuous function.

There must be d_p let $\Phi(d)=0$, that is

$$\begin{aligned}\kappa\left(d_p,\nu^y\right)-\kappa\left(d_p,\nu^z\right)&=0\\ d_p&=\kappa^{-1}\left(\kappa(d_p,\nu^y),\nu^z\right)\end{aligned} \tag{29}$$

Then we solve for this particular point:

$$\begin{aligned}\frac{\partial\Phi(d)}{\partial d^2} &= \frac{\operatorname{Gam}\left(\frac{\nu^y+1}{2}\right)}{\sqrt{\nu^y\pi}\operatorname{Gam}\left(\frac{\nu^y}{2}\right)}(-\frac{\nu^y+1}{2})\left(1+\frac{d_{ij}^2}{\nu^y}\right)^{-\frac{\nu^y+3}{2}}\frac{1}{v^y} - \frac{\operatorname{Gam}\left(\frac{\nu^z+1}{2}\right)}{\sqrt{\nu^z\pi}\operatorname{Gam}\left(\frac{\nu^z}{2}\right)}(-\frac{\nu^y+1}{2})\left(1+\frac{d_{ij}^2}{\nu^z}\right)^{-\frac{\nu^z+3}{2}}\frac{1}{v^z}\\ \frac{\partial\Phi(d)}{\partial d^2} &= \frac{1}{2\sqrt{\pi}}\left(\frac{\operatorname{Gam}\left(\frac{\nu^z+1}{2}\right)}{\nu^{z\,1.5}\operatorname{Gam}\left(\frac{\nu^z}{2}\right)}(\nu^z+1)\left(1+\frac{d_{ij}^2}{\nu^z}\right)^{-\frac{\nu^z+3}{2}} - \frac{\operatorname{Gam}\left(\frac{\nu^y+1}{2}\right)}{\nu^{y\,1.5}\operatorname{Gam}\left(\frac{\nu^y}{2}\right)}(\nu^y+1)\left(1+\frac{d_{ij}^2}{\nu^y}\right)^{-\frac{\nu^y+3}{2}}\right)\end{aligned} \tag{30}$$

$\frac{\partial\Phi(d)}{\partial d^2}=0$ has a unique solution d^*, and $\frac{\partial\Phi(d^*)}{\partial d^2}<0$.

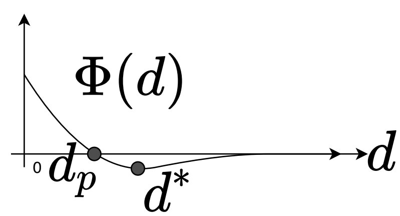

Fig. 9. Proof. of Lemma 2: $\Phi(d)$

Then d_p is the only solution of $\Phi(d)=0$, and if $d>d_p$ then $\Phi(d)<0$, and if $d>d_p$ then $\Phi(d)<0$, and then $(d-d_p)\Phi(d)<0$. Next we construct $\Psi(d)=\kappa^{-1}(\kappa\left(d,\nu^y\right),\nu^z)-\kappa^{-1}(\kappa\left(d,\nu^z\right),\nu^z)=\kappa^{-1}(\kappa\left(d,\nu^y\right),\nu^z)-d$.

Because the $\kappa^{-1}(\cdot)$ is monotone continuity, if

$$\kappa\left(d,\nu^y\right)>\kappa\left(d,\nu^z\right)$$

then

$$\kappa^{-1}(\kappa\left(d,\nu^y\right),\nu^z)<\kappa^{-1}(\kappa\left(d,\nu^z\right),\nu^z)=y$$

Substitute $\Phi(d)=\kappa\left(d,\nu^y\right)-\kappa\left(d,\nu^z\right)$ to $(d-d_p)\Psi(d)>0$, we have $(d^y-d_p)(d^{z+}-d^y)>0$.

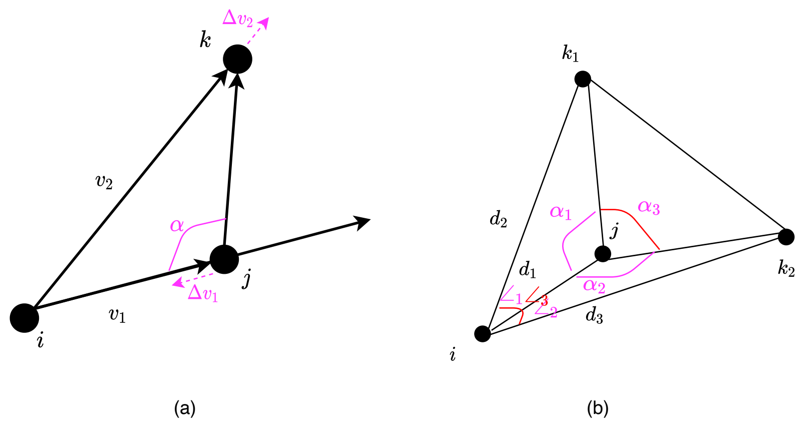

Fig. 10. Proof of Lemma 2

A.3 A.3 Proof of Lemma 2

Lemma 2. Assume f_θ satisfies HOP1-2 order preserving:

$$\max(\{d_{ij}^y\}_{j\in \mathrm{HOP}_1(x_i)}) < \min(\{d_{ij}^y\}_{j\in \mathrm{HOP}_2(x_i)}) \tag{31}$$

Then $\overline{K}_{\mathcal{M}}^{z+} < \overline{K}_{\mathcal{M}}^{y}$ where $\overline{K}_{\mathcal{M}}^{y}$ is the mean curvature in the structure space, and $\overline{K}_{\mathcal{M}}^{z+}$ is the mean curvature optimization results of L_{D} in the embedding space.

We randomly select a point j in the data set X, and then select 3 neighbor nodes i, k_1, k_2, while ensuring k_1 k_2 are the HOP2 of i.

As shown in Figure 10, Let d_{ij}^y, $d_{ik_1}^y$, $d_{ik_2}^y$ be distance in structure space, and let $\angle 1 = \angle jik_1$, $\angle 2 = \angle jik_2$, $\angle 3 = \angle k_1ik_2$. Because $j \in \mathrm{HOP}_1(i)$ and $k_1, k_2 \in \mathrm{HOP}_2(j)$, we have $d_{ik_1}^y > d_{ij}^y$ and $d_{ik_2}^y > d_{ij}^y$.

We can find a suitable α to make $d^y(ij)^y\alpha < d_p$ and $d^y(ik)^y\alpha > d_p$. and let $d_1 = d_{ij} = \alpha d^y(ij)^y$,$d_2 = d_{ik_1} = \alpha d^y(ik_1)^y$,$d_3 = d_{ik_2} = \alpha d^y(ik_2)^y$. Then according to Lemma 1, in the embedding sapce, $d^{z+}(ij) < d^y(ij)$ and $d(ik)^{z+} > d(ik)$ and.

First, according to Law of cosines[4], we have:

$$\begin{aligned} d_{jk_1} &= (d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}} \\ d_{jk_2} &= (d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2)^{\frac{1}{2}} \\ d_{k_1k_2} &= (d_2^2 + d_3^2 - 2d_2d_3\cos\angle 3)^{\frac{1}{2}} \end{aligned} \tag{32}$$

According to discrete Gauss-Bonnet theorem, the discrete curvature of point j is a function of $\angle ijk_1$, $\angle ijk_2$, $\angle k_1jk_2$. Let's write the three angle separately.

[4] https://en.wikipedia.org/wiki/Law_of_cosines

$$
\begin{aligned}
\cos(\angle k_1 j k_2) &= \frac{d_{jk_1}^2 + d_{jk_2}^2 - d_{k_1k_2}^2}{2 d_{jk_1} d_{jk_2}} \\
&= \frac{{(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}}^2 + {(d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2)^{\frac{1}{2}}}^2 - {(d_2^2 + d_3^2 - 2d_2d_3\cos\angle 3)^{\frac{1}{2}}}^2}{2(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}(d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2)^{\frac{1}{2}}} \\
&= \frac{(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1) + (d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2) - (d_2^2 + d_3^2 - 2d_2d_3\cos\angle 3)}{2(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}(d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2)^{\frac{1}{2}}} \\
&= \frac{d_1^2 - d_1d_2\cos\angle 1 - d_1d_3\cos\angle 2 + d_2d_3\cos\angle 3}{(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}(d_1^2 + d_3^2 - 2d_1d_3\cos\angle 2)^{\frac{1}{2}}}
\end{aligned}
\tag{33}
$$

$$
\begin{aligned}
\cos(\angle ijk_1) &= \frac{d_1^2 + d_{jk_1}^2 - d_2^2}{2 d_1 d_{jk_1}} \\
&= \frac{d_1^2 + {(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}}^2 - d_2^2}{2d_1(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}} \\
&= \frac{d_1^2 + d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1 - d_2^2}{2d_1(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}} \\
&= \frac{d_1^2 - d_1d_2\cos\angle 1}{d_1(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}} \\
&= \frac{d_1 - d_2\cos\angle 1}{(d_1^2 + d_2^2 - 2d_1d_2\cos\angle 1)^{\frac{1}{2}}}
\end{aligned}
\tag{34}
$$

The same as Eq. (34):

$$
\cos(\angle ijk_2) = \frac{d_1 - d_3\cos\angle 1}{(d_1^2 + d_3^2 - 2d_1d_3\cos\angle 1)^{\frac{1}{2}}}
\tag{35}
$$

Let $v_1 = \frac{d_2}{d_1}$, and $v_2 = \frac{d_3}{d_1}$, we directly represent three angles according to the inverse cosine function.

$$
\begin{aligned}
\angle ijk_1 &= \arccos\Big(\frac{1 - v_2\cos\angle 1}{(1 + v_2^2 - 2v_2\cos\angle 1)^{\frac{1}{2}}}\Big) \\
\angle ijk_2 &= \arccos\Big(\frac{1 - v_3\cos\angle 2}{(1 + v_3^2 - 2v_3\cos\angle 2)^{\frac{1}{2}}}\Big) \\
\angle k_1jk_2 &= \arccos\Big(\frac{1 - v_1\cos\angle 1 - v_2\cos\angle 2 + v_1v_2\cos\angle 3}{(1 + v_1^2 - 2v_1\cos\angle 1)^{\frac{1}{2}}(1 + v_2^2 - 2v_2\cos\angle 2)^{\frac{1}{2}}}\Big)
\end{aligned}
\tag{36}
$$

According to Lemma 1, L_{D} will increase d_2 and d_3, and decreases d_1. so L_{D} will increase v_1 amd v_2.

$$
\begin{aligned}
&\frac{\partial \angle ijk_1}{\partial v_1} \\
=&\frac{\partial}{\partial v_1}\arccos\left(\frac{1-v_1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}}\right) \\
=&\frac{-1}{1+\frac{(1-v_1\cos\angle 1)^2}{(1+v_1^2-2v_1\cos\angle 1)}}\frac{\partial}{\partial v_1}\frac{1-v_1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}} \\
=&\frac{-(1+v_1^2-2v_1\cos\angle 1)}{(1+v_1^2-2v_1\cos\angle 1)+(1-v_1\cos\angle 1)^2}\frac{\partial}{\partial v_1}\frac{1-v_1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}} \\
=&\frac{-(1+v_1^2-2v_1\cos\angle 1)}{(1+v_1^2-2v_1\cos\angle 1)+(1-v_1\cos\angle 1)^2}\left(\frac{-\cos\angle 1(1+v_1^2-2v_1\cos\angle 1)-(1-v_1\cos\angle 1)(v_1-\cos\angle 1)}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{3}{2}}}\right) \\
=&\frac{(\cos\angle 1)(1+v_1^2-2v_1\cos\angle 1)+(1-v_1\cos\angle 1)(v_1-\cos\angle 1)}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}(1+v_1^2-2v_1\cos\angle 1)+(1-v_1\cos\angle 1)^2} \\
=&\frac{v_1-v_1\cos\angle 1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}\left((1+v_1^2-2v_1\cos\angle 1)+(1-v_1\cos\angle 1)^2\right)} \\
=&\frac{\sin\angle 1}{\left(v_1^2-2v_1\cos\angle 1+1\right)}
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
&\frac{\partial \angle ijk_2}{\partial v_1} \\
=&\frac{\partial}{\partial v_1}\arccos\left(\frac{1-v_2\cos\angle 2}{(1+v_2^2-2v_2\cos\angle 1)^{\frac{1}{2}}}\right) \\
=&0
\end{aligned}
\tag{38}
$$

$$
\begin{aligned}
&\frac{\partial \angle k_1jk_2}{\partial v_1} \\
=&\frac{\partial}{\partial v_1}\arccos\left(\frac{1-v_1\cos\angle 1-v_2\cos\angle 2+v_1v_2\cos\angle 3}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}(1+v_2^2-2v_2\cos\angle 2)^{\frac{1}{2}}}\right) \\
=&\frac{-1}{1+\frac{(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2)^2}{(v_1^2-2v_1cos\angle 1)(v_2^2-2v_2\cos\angle 2)}}\frac{\partial}{\partial v_1}\frac{v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2}{(v_1^2-2v_1cos\angle 1)^{\frac{1}{2}}(v_2^2-2v_2\cos\angle 2)^{\frac{1}{2}}} \\
=&\frac{-(v_1^2-2v_1cos\angle 1)(v_2^2-2v_2\cos\angle 2)}{(v_1^2-2v_1cos\angle 1)(v_2^2-2v_2\cos\angle 2)+(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2)^2}\frac{\partial}{\partial v_1}\frac{v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2}{(v_1^2-2v_1cos\angle 1)^{\frac{1}{2}}(v_2^2-2v_2\cos\angle 2)^{\frac{1}{2}}} \\
=&\frac{-(v_1^2-2v_1cos\angle 1)(v_2^2-2v_2\cos\angle 2)}{(v_1^2-2v_1cos\angle 1)(v_2^2-2v_2\cos\angle 2)+(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2)^2}\frac{\partial}{\partial v_1}\frac{v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2}{(v_1^2-2v_1cos\angle 1)^{\frac{1}{2}}(v_2^2-2v_2\cos\angle 2)^{\frac{1}{2}}} \\
=&\frac{\left(v_1v_2\cos\angle 1\cos\angle 3-v_1v_2\cos\angle 2+v_1\sin^2\angle 1+v_2\cos\angle 1\cos\angle 2-v_2\cos\angle 3\right)}{\sqrt{\left(v_1^2-2v_1\cos\angle 1+1\right)\left(v_2^2-2v_2\cos\angle 2+1\right)-\left(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2+1\right)^2}\left(v_1^2-2v_1\cos\angle 1+1\right)}
\end{aligned}
\tag{39}
$$

Then, we sum Eq.(37) Eq.(38) Eq.(39) together.

$$
\begin{aligned}
&\frac{\partial \angle ijk_1+\angle ijk_2+\angle k_1jk_2}{\partial v_1}\\
=&\frac{\partial}{\partial v_1}[\arccos(\frac{1-v_2\cos\angle 2}{(1+v_2^2-2v_2\cos\angle 2)^{\frac{1}{2}}})+\arccos(\frac{1-v_1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}})+\frac{1-v_1\cos\angle 1}{(1+v_1^2-2v_1\cos\angle 1)^{\frac{1}{2}}}]\\
=&\frac{\Big(v_1v_2\cos\angle 1\cos\angle 3-v_1v_2\cos\angle 2+v_1\sin^2\angle 1+v_2\cos\angle 1\cos\angle 2-v_2\cos\angle 3\Big)}{\sqrt{\left(v_1^2-2v_1\cos\angle 1+1\right)\left(v_2^2-2v_2\cos\angle 2+1\right)-\left(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2+1\right)^2}\left(v_1^2-2v_1\cos\angle 1+1\right)}\\
&+\frac{\sin\angle 1}{\left(v_1^2-2v_1\cos\angle 1+1\right)}+0\\
=&\frac{\Big(v_1v_2\cos\angle 1\cos\angle 3-v_1v_2\cos\angle 2+v_1\sin^2\angle 1+v_2\cos\angle 1\cos\angle 2-v_2\cos\angle 3\Big)+\sin\angle 1M_1}{M_1\left(v_1^2-2v_1\cos\angle 1+1\right)}
\end{aligned}
\tag{40}
$$

where $M_1=\sqrt{\left(v_1^2-2v_1\cos\angle 1+1\right)\left(v_2^2-2v_2\cos\angle 2+1\right)-\left(v_1v_2\cos\angle 3-v_1\cos\angle 1-v_2\cos\angle 2+1\right)^2}$

Let $\phi=\Big(v_1v_2\cos\angle 1\cos\angle 3-v_1v_2\cos\angle 2+v_1\sin^2\angle 1+v_2\cos\angle 1\cos\angle 2-v_2\cos\angle 3\Big)+$ $\sin\angle 1M_1$

$$
\frac{\partial\phi}{\partial v_1}=\Big(v_2\cos\angle 1\cos\angle 3-v_2\cos\angle 2+\sin^2\angle 1\Big)+\sin\angle 1\frac{\partial M_1}{\partial v_1}
\tag{41}
$$

$$
\frac{\partial M_1}{\partial v_1}=\frac{\left(v_2-\cos\left(a_1\right)\right)\Big(v_3^2-2v_3\cos\left(a_2\right)+1\Big)-\Big(v_3\cos\left(a_3\right)-\cos\left(a_1\right)\Big)\Big(v_2v_3\cos\left(a_3\right)-v_2\cos\left(a_1\right)-v_3\cos\left(a_2\right)+1\Big)}{\sqrt{\left(v_2^2-2v_2\cos\left(a_1\right)+1\right)\left(v_3^2-2v_3\cos\left(a_2\right)+1\right)-\left(v_2v_3\cos\left(a_3\right)-v_2\cos\left(a_1\right)-v_3\cos\left(a_2\right)+1\right)^2}}
\tag{42}
$$

Let $M_2=\left(v_2-\cos\left(a_1\right)\right)\Big(v_3^2-2v_3\cos\left(a_2\right)+1\Big)-\Big(v_3\cos\left(a_3\right)-\cos\left(a_1\right)\Big)\Big(v_2v_3\cos\left(a_3\right)-v_2\cos\left(a_1\right)-v_3\cos\left(a_2\right)+$

$$
\begin{aligned}
\frac{\partial M2}{\partial v_1}&=v_2^2-2v_2\cos\left(\angle 2\right)+\Big(-v_2\cos\left(\angle 3\right)+\cos\angle 1\Big)\Big(v_2\cos\left(\angle 3\right)-\cos\angle 1\Big)+1\\
&=v_2^2-2v_2\cos\left(\angle 2\right)+\Big(-v_2\cos\left(\angle 3\right)+\cos\angle 1\Big)\Big(v_2\cos\left(\angle 3\right)-\cos\angle 1\Big)+1\\
&=v_2^2\sin^2\left(\angle 3\right)+2v_2\Big(\cos\angle 1\cos\left(\angle 3\right)-\cos\left(\angle 2\right)\Big)+\sin^2\angle 1\\
&>v_2^2\sin^2\left(\angle 3\right)-2v_2\Big(\sin\angle 1\sin\left(\angle 3\right)\Big)+\sin^2\angle 1\\
&>\Big(v_2^2\sin\left(\angle 3\right)-\sin\angle 1\Big)^2>0
\end{aligned}
\tag{43}
$$

So

$$
\begin{aligned}
M_1&>v_2^2-2v_2\cos\left(\angle 2\right)-\Big(-v_2\cos\left(\angle 2\right)+1\Big)^2+1\\
&>-v_3^2\cos^2\left(\angle_2\right)+v_3^2\\
&>v_3^2\sin^2\left(\angle_2\right)\\
&>0
\end{aligned}
\tag{44}
$$

So $\frac{\partial M2}{\partial v_1}>\frac{\partial M2}{\partial v_1}|_{d2=1}=v_2\sin\angle 3-v_2\sin\angle 1$, So

So

$$\begin{aligned}
\frac{\partial \phi}{\partial v_1} &= \left(v_2 \cos\angle 1 \cos\angle 3 - v_2 \cos\angle 2 + \sin^2\angle 1\right) + \sin\angle 1 \frac{\partial M_1}{\partial v_1} \\
&> v_2 \left(\cos\angle 1 \cos\angle 3 - \cos\angle 2\right) + \sin^2\angle 1 + \sin\angle 1 \frac{\partial M_1}{\partial v_1}_{d2=1} \\
&> v_2 \left(\sin\angle 1 \sin\angle 3 \cos P + \sin^2\angle 1\right) + \sin\angle 1 (v_2 \sin\angle 3 - v_2 \sin\angle 1) \\
&> \sin\angle 1 \left(v_2 \left(-1 \sin\angle 3 \cos P + \sin\angle 1\right) + v_2 \sin\angle 3 - v_2 \sin\angle 1\right) \\
&> 0
\end{aligned} \tag{45}$$

So $\frac{\partial \angle A}{\partial v_1} > 0$, and because $v_1^{z+} > v_1^{y}$ so $K_j^{z+} = K_j^{y} - \int_{v_1^{y}}^{v_1^{z+}} \frac{\partial \angle A}{\partial v_1} \, dv_1$, and $K_j^{z+} < K_j^{y}$.

B Detailed experiments settings: Easy Dataset

B.1 Experimental Setups and Datasets Information

The compared methods include two manifold learning methods (UMAP [28] [5], t-SNE [15] [6]) and three deep manifold learning methods (PHATE [30] [7], ivis [?] [8] and parametric UMAP(P-UMAP) [9] [35].) The datasets include six simple image datasets (Digits [10], Coil20 [11], Coil100 [12], Mnist [13], EMnist [14], and KMnist [15]) and six biological datasets (Colon [16], Activity [17] , MCA [18], Gast10k [19] , SAMUSIK [20] , and HCL).

For a fair comparison, we embed the data into a 2-dimensional latent space using the method to be evaluated and then evaluate the method performance by 10-fold cross-validation. We obtain classification accuracy by applying a linear SVM classifier in the latent space and clustering accuracy by using a kmeans cluster in the latent space. the classification accuracy and the clustering accuracy are shown in Table 1 Details of datasets, baseline methods, and evaluation metrics are in the Table 4.

Table 4. Datasets information of simple manifold embedding task

Dataset	Point Number	Dimension Number	Dataset	Point Number	Dimension Number
Digits	10,000	8×8×1	Colon	1,117	500
Coil20	1,440	128×128×1	Activity	10,299	561
Coil100	7,200	128×128×3	MCA	30,000	34947
Mnist	60000	28×28×1	Gast10k	10,638	1457
EMnist	60000	28×28×1	SAMUSIK	86,864	38
KMnist	60000	28×28×1	HCL	60,000	27341

B.2 Experimental Parameters

We use MLP as the network f_θ and network g_ϕ and use the AdamW optimizer with learning rate 0.001 and weight decay 1e-6. DLME are trained for 1500 epochs. The

[5] https://github.com/lmcinnes/umap
[6] https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
[7] https://github.com/KrishnaswamyLab/PHATE
[8] https://github.com/beringresearch/ivis
[9] https://github.com/lmcinnes/umap
[10] https://scikit-learn.org/stable/auto_examples/datasets/plot_digits_last_image.html
[11] https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[12] https://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[13] https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits
[14] https://www.tensorflow.org/datasets/catalog/emnist
[15] https://www.tensorflow.org/datasets/catalog/kmnist
[16] https://figshare.com/articles/dataset/The_microarray_dataset_of_colon_cancer_in_csv_format_/13658790/1
[17] https://www.kaggle.com/uciml/human-activity-recognition-with-smartphones
[18] https://figshare.com/articles/dataset/MCA_DGE_Data/5435866
[19] http://biogps.org/dataset/tag/gastric%20carcinoma/
[20] https://github.com/abbioinfo/CyAnno

architecture of f_θ is [-1,500,300,80], where -1 mean the dimension of input data. The architecture of g_ϕ is [-1,500,80,2]. $\nu_y = 100$, and $\alpha = 0.01$.

C Detailed experiments settings: CV Dataset

C.1 Experimental setups:

The compared methods include contrastive learning methods (NPID [41] [21] , ODC [44] [22], SimCLR [3] [23], MOCO.v2 [14] [24] and BYOL [11] [25]) and contrastive clustering methods (DAC [12] [26], DCCM [40] [27], PICA [17] [28] CC [23] [29], and CRLC [5]). The datasets include four image datasets (CIFAR10[30], CIFAR100[31], STL10[32], and tinyImageNet [33])

To compare with the two different baseline methods, the segmentation of the dataset used by the two subtasks is various.

Table 5. Dataset segmentation of linear-test sub-task

Dataset	Train data	Test data	Train Samples	Test Samples	Classes
CIFAR-10	Train	Test	50,000	10,000	10
CIFAR-100	Train	Test	50,000	10,000	100
STL-10	Train + Unlabeled	Test	5,000+100,000	8,000	10
Tiny-ImageNet	Train	Test	100,000	100,000	200

Table 6. Dataset segmentation of clustering sub-task

Dataset	Split	Samples	Classes
CIFAR-10	Train+Test	60,000	10
CIFAR-100	Train+Test	60,000	20
STL-10	Train+Test	13,000	10
Tiny-ImageNet	Train	100,000	200

C.2 Experimental parameters for linear-test sub-task

For fire compared with BYOL [11] and other methods, we use the same ResNet-50 architecture as the f_θ and use MLP as g_ϕ. The performance of two downstream tasks

[21] https://github.com/zhirongw/lemniscate.pytorch
[22] https://github.com/open-mmlab/OpenSelfSup
[23] https://github.com/google-research/simclr
[24] https://github.com/facebookresearch/moco
[25] https://github.com/deepmind/deepmind-research/tree/master/byol
[26] https://github.com/jplapp/associative_deep_clustering
[27] https://github.com/Cory-M/DCCM
[28] https://github.com/Raymond-sci/PICA
[29] https://github.com/Yunfan-Li/Contrastive-Clustering
[30] https://www.cs.toronto.edu/ kriz/cifar.html
[31] https://www.cs.toronto.edu/ kriz/cifar.html
[32] https://cs.stanford.edu/ acoates/stl10/
[33] https://www.kaggle.com/c/tiny-imagenet

is evaluated in the embedding space. we use the Adamw optimizer and weight decay is 1e-6. $\nu_y = 100$, $\nu_z = 10$, We use the same data augmentation method as BYOL [11].

We use the grid search method to determine the best super parameters. The super parameters adjusted in the grid search method are as follows.

Table 7. Hyperparameter search space for linear-test sub-task

Hyperparameters	Search Space
batch size	[256, 512]
learning rate	[2e-4, 5e-4, 6e-4]

C.3 Experimental parameters for clustering sub-task

For fire compared with CC [23] and other methods, we use the same ResNet-50 architecture as the f_θ and use MLP as g_ϕ. The performance of two downstream tasks is evaluated in the embedding space. we use the AdamW optimizer and weight decay is 1e-6. $\nu_y = 100$, We use the same data augmentation method as BYOL [11].

Table 8. Hyperparameter search space for clustering sub-task

Hyperparameters	Search Space
degree of freedom in embedding space ν^z	[1, 100, 100]
batch size	[128, 256, 512]
learning rate	[1e-4, 2e-4, 1e-3]

D Detailed experiments Results: Toy datasets

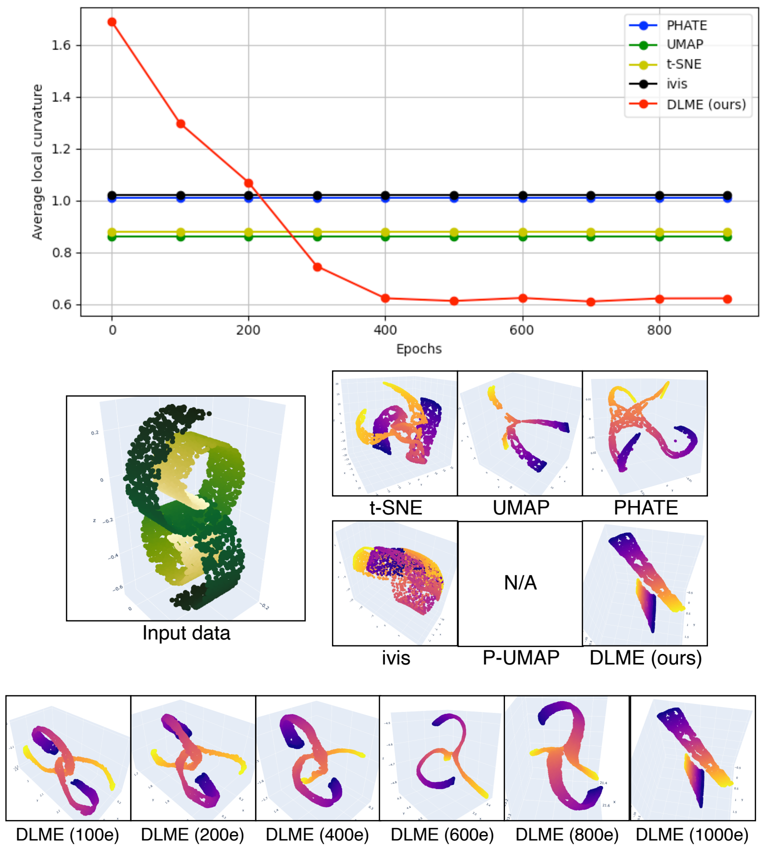

Fig. 11. Training details of TweinSwissRoll dataset

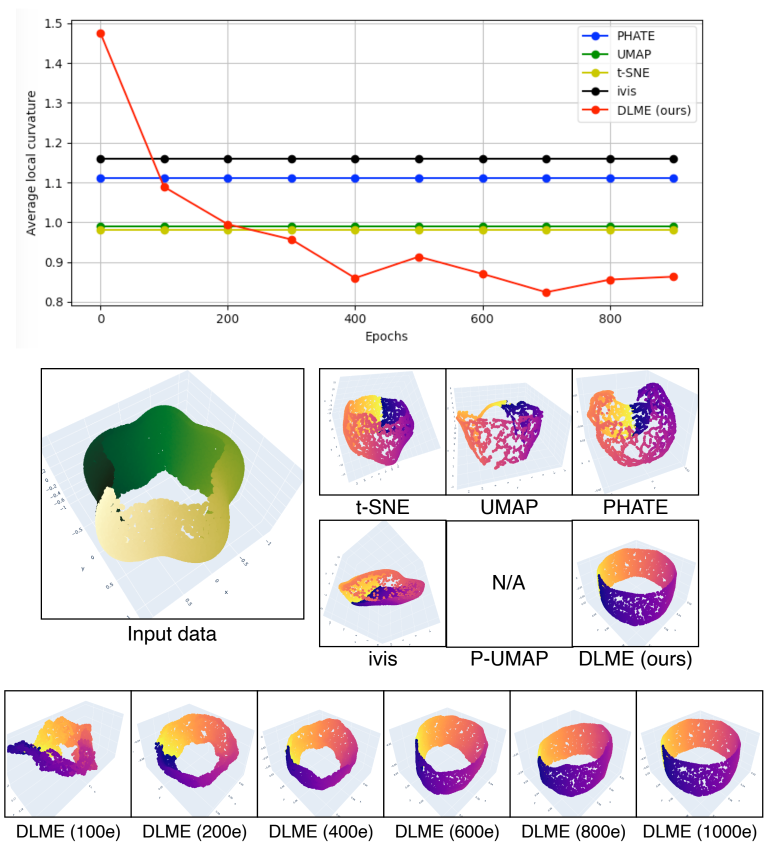

Fig. 12. Training details of StarFrut dataset

E Detailed experiments Results: Visualization

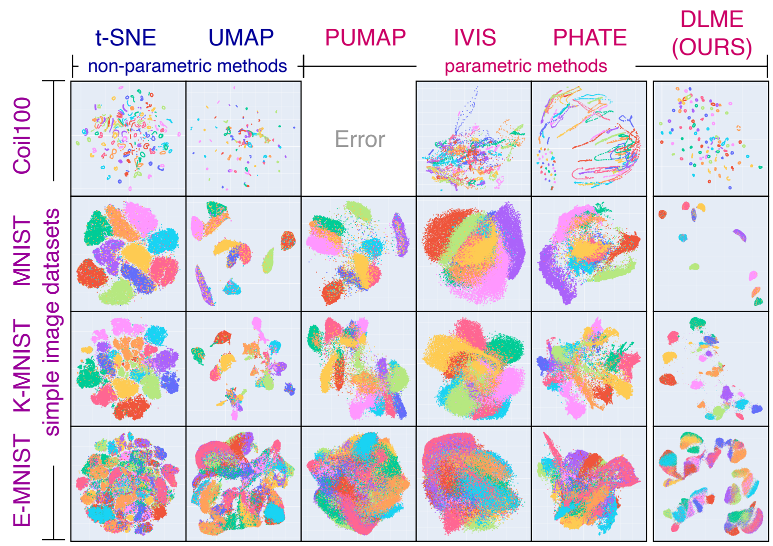

Fig. 13. More detailed visualization results of simple image data

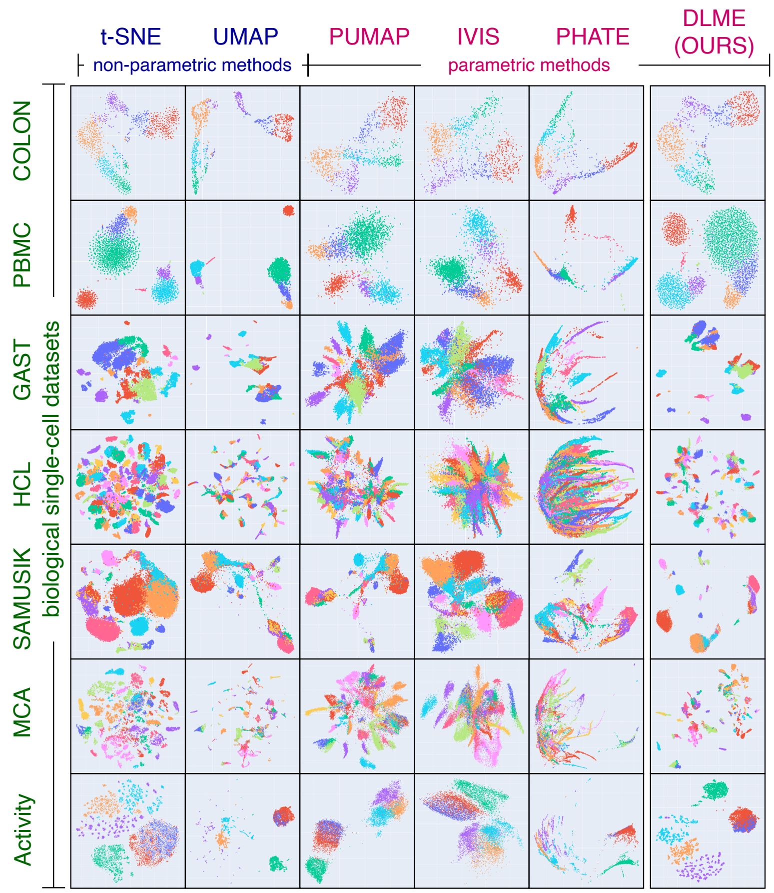

Fig. 14. More detailed visualization results of visualization data