Supplement of Spherical Feature Transform for Deep Metric Learning

Yuke Zhu[1], Yan Bai[2], and Yichen Wei[1]

[1] MEGVII Inc. {zhuyuke, weiyichen}@megvii.com
[2] Tong Ji Unversity yan.bai@tongji.edu.cn

Proof of Proposition 1

Proposition 1. *Suppose $N_1(\boldsymbol{\mu_1}, \Sigma_1)$ and $N_2(\boldsymbol{\mu_2}, \Sigma_2)$ are two Gaussian approximations of spherical distributions. If they are spherical-homoscedastic, then the rotation matrix between them is spanned by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.*

Proof. From the Definition 2, the rotation matrix from N_1 to N_2 is $\mathbf{A}$ and is spanned by $\boldsymbol{\mu}_1$ and one eigenvector $\mathbf{v}_1$ of $\boldsymbol{\Sigma}_1$. Analogously, N_2 is also spherical-homoscedastic with N_1 and the rotation matrix from N_2 to N_1 is $\mathbf{A}^{-1}$, which is spanned by $\boldsymbol{\mu}_2$ and one eigenvector $\mathbf{v}_2$ of $\boldsymbol{\Sigma}_2$. As $\mathbf{A}$ is orthogonal and the subspace for the *rotation transform* and its reverse is identical, the subspace spanned by $\{\boldsymbol{\mu}_1, \mathbf{v}_1\}$ and the subspace spanned by $\{\boldsymbol{\mu}_2, \mathbf{v}_2\}$ are exactly the same. As $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are linearly independent, so the subspace is also spanned by $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2\}$. □

Proof of Proposition 3

Proposition 3. SFT degenerates to the *translation transform* iff for ∀ feature $\mathbf{x}$ with $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\sigma}$, $\boldsymbol{\mu} \perp \boldsymbol{\sigma}$

Proof of Sufficiency. Suppose $\mathbf{U} = \{\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_m}\}$ and $V = \{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}\}$ are two sets of basis for the two subspaces of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ respectively. Let $\mathbf{Q} = \mathbf{V}\mathbf{V}^T$. $\mathbf{Q}$ is the projection matrix of subspace $\mathbf{V}$. For any $\boldsymbol{\sigma}$ and any $\boldsymbol{\mu}$, the following relation holds:

$$\begin{aligned} \mathbf{Q}\boldsymbol{\sigma} &= \boldsymbol{\sigma} \\ \mathbf{Q}\boldsymbol{\mu} &= \mathbf{0} \end{aligned} \tag{1}$$

We use this relation to evaluate the SFT defined in Eq. 3, $\tilde{\mathbf{x}}_2 = \mathbf{A}\mathbf{x}_1$.

$$\tilde{\mathbf{x}}_2 = \mathbf{A}\mathbf{x}_1 = \mathbf{A}(\boldsymbol{\mu}_1 + \boldsymbol{\sigma}_1) = \mathbf{A}(\boldsymbol{\mu}_1 + \mathbf{Q}\boldsymbol{\sigma}_1) = \boldsymbol{\mu}_2 + \mathbf{A}\mathbf{Q}\boldsymbol{\sigma}_1. \tag{2}$$

We further use the calculation of matrix $\mathbf{A}$ defined in Eq. 4 to evaluate the right part of the above expression.

$$\begin{aligned} \mathbf{A}\mathbf{Q} = \mathbf{Q} &+ (\mathbf{n}_2\mathbf{n}_1^T\mathbf{Q} - \mathbf{n}_1\mathbf{n}_2^T\mathbf{Q})\sin(\alpha) \\ &+ (\mathbf{n}_1\mathbf{n}_1^T\mathbf{Q} + \mathbf{n}_2\mathbf{n}_2^T\mathbf{Q})(\cos(\alpha) - 1), \end{aligned} \tag{3}$$

As $\mathbf{n_1}$ and $\mathbf{n_2}$ used in Eq. 4 are calculated by linear combination of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, they are also projected to null space by $\mathbf{Q}$, which results in $\mathbf{A}\mathbf{Q} = \mathbf{Q}$. Then Eq. 2 can further be written as:

$$\tilde{\mathbf{x}}_2 = \boldsymbol{\mu}_2 + \mathbf{A}\mathbf{Q}\boldsymbol{\sigma}_1 = \mathbf{x_1} + \mathbf{Q}\boldsymbol{\sigma}_1 = \mathbf{x_1} + \boldsymbol{\sigma}_1, \tag{4}$$

which is just the *translation transform.* □

Proof of Necessity. Suppose for $\forall \mathbf{x}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\sigma}_1$ and any target class $\boldsymbol{\mu}_2$, the result of *rotation transform* is equivalent with the that of *translation transform*: $\mathbf{A}\mathbf{x}_1 = \boldsymbol{\mu}_2 + \boldsymbol{\sigma}_1$. We have $\mathbf{A}\mathbf{x}_1 = \mathbf{A}(\boldsymbol{\mu}_1 + \boldsymbol{\sigma}_1) = \boldsymbol{\mu}_2 + \mathbf{A}\boldsymbol{\sigma}_1$. So $\mathbf{A}\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_1$. It means that $\boldsymbol{\sigma}_1$ lies in the invariant subspace of $\mathbf{A}$. As $\mathbf{A}$ is spanned by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, $\boldsymbol{\sigma}_1 \perp \boldsymbol{\mu}_1, \boldsymbol{\sigma}_1 \perp \boldsymbol{\mu}_2$. As the degeneration happens for any $\boldsymbol{\sigma}$ and any $\boldsymbol{\mu}$, so $\boldsymbol{\mu} \perp \boldsymbol{\sigma}$ must be always hold. □