# Supplementary Material: Domain-Sum Feature Transformation For Multi-Target Domain Adaptation

Takumi Kobayashi[1,2]
takumi.kobayashi@aist.go.jp
Lincon S. Souza[1]
lincon.souza@aist.go.jp
Kazuhiro Fukui[2]
kfukui@cs.tsukuba.ac.jp

[1] National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Japan
[2] University of Tsukuba
Tsukuba, Japan

## A Proofs

**Preliminary.** The geodesic path between $\boldsymbol{U}_{\mathcal{S}} \in \mathbb{R}^{d\times r}$ and $\boldsymbol{U}_{\mathcal{T}} \in \mathbb{R}^{d\times r}$ on a Grassmann manifold is described by

$$\boldsymbol{U}_{(t)} = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\boldsymbol{\Gamma}_{(t)} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\boldsymbol{\Sigma}_{(t)},\ t\in[0,1], \tag{A.0.1}$$

where $\bar{\boldsymbol{U}}_{\mathcal{S}} \in \mathbb{R}^{d-r\times r}$ is an orthogonal complement matrix to $\boldsymbol{U}_{\mathcal{S}}$ and $\boldsymbol{S}\in\mathbb{R}^{r\times r}$ and $\bar{\boldsymbol{S}}\in\mathbb{R}^{d-r\times r}$ are orthonormal matrices given by the (generalized) SVD of

$$\boldsymbol{U}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}} = \boldsymbol{S}\boldsymbol{\Gamma}\boldsymbol{T}^\top, \qquad \bar{\boldsymbol{U}}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}} = -\bar{\boldsymbol{S}}\boldsymbol{\Sigma}\boldsymbol{T}^\top, \tag{A.0.2}$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\{\cos\theta_k\}_{k=1}^r), \qquad \boldsymbol{\Sigma} = \mathrm{diag}(\{\sin\theta_k\}_{k=1}^r), \tag{A.0.3}$$

$$\boldsymbol{\Gamma}_{(t)} = \mathrm{diag}(\{\cos(t\theta_k)\}_{k=1}^r), \qquad \boldsymbol{\Sigma}_{(t)} = \mathrm{diag}(\{\sin(t\theta_k)\}_{k=1}^r), \tag{A.0.4}$$

which use an orthonormal matrix $\boldsymbol{T}\in\mathbb{R}^{r\times r}$ and the canonical angles $\{\theta_k\}_{k=1}^r$ between $\boldsymbol{U}_{\mathcal{S}}$ and $\boldsymbol{U}_{\mathcal{T}}$.

Based on the geodesic path, the gradient flow kernel (GFK) matrix [1] is formulated in

$$\boldsymbol{G} = [\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}, \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}] \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2 \\ \boldsymbol{\Lambda}_2 & \boldsymbol{\Lambda}_3 \end{bmatrix} [\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}, \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}]^\top, \tag{A.0.5}$$

where the diagonal matrices are

$$\boldsymbol{\Lambda}_1 = \mathrm{diag}\left(\left\{1+\frac{\sin(2\theta_k)}{2\theta_k}\right\}_{k=1}^r\right), \tag{A.0.6}$$

$$\boldsymbol{\Lambda}_2 = \mathrm{diag}\left(\left\{\frac{\cos(2\theta_k)-1}{2\theta_k}\right\}_{k=1}^r\right), \tag{A.0.7}$$

$$\boldsymbol{\Lambda}_3 = \mathrm{diag}\left(\left\{1-\frac{\sin(2\theta_k)}{2\theta_k}\right\}_{k=1}^r\right). \tag{A.0.8}$$

On the other hand, the proposed method is based on the domain-sum matrix of

$$\boldsymbol{H} = \sum_{t\in\{0,1\}} \boldsymbol{U}_{(t)}\boldsymbol{U}_{(t)}^\top = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{U}_{\mathcal{T}}^\top = \boldsymbol{P}_{\mathcal{S}} + \boldsymbol{P}_{\mathcal{T}}. \tag{A.0.9}$$

The main manuscript presents several relationships among these matrices which are proven as follows.

## A.1 Proof for Proposition 1

**Proposition 1.** *GFK matrix* $\boldsymbol{G}$ (A.0.5) *and DS matrix* $\boldsymbol{H}$ (A.0.9) *are similarly eigen-decomposed as*

$$\boldsymbol{G} = [\boldsymbol{U}_+, \boldsymbol{U}_-]\begin{bmatrix}\boldsymbol{\Psi}_+ & \\ & \boldsymbol{\Psi}_-\end{bmatrix}[\boldsymbol{U}_+, \boldsymbol{U}_-]^\top, \tag{A.1.1}$$

$$\boldsymbol{H} = [\boldsymbol{U}_+, \boldsymbol{U}_-]\begin{bmatrix}\boldsymbol{\Phi}_+ & \\ & \boldsymbol{\Phi}_-\end{bmatrix}[\boldsymbol{U}_+, \boldsymbol{U}_-]^\top, \tag{A.1.2}$$

*where*

$$\boldsymbol{U}_\pm = \mathsf{colnorm}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} \pm \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}) \in \mathbb{R}^{d\times r}, \tag{A.1.3}$$

$$\boldsymbol{\Psi}_\pm = \mathsf{diag}(\{1 \pm \mathrm{sinc}\,\theta_k\}_{k=1}^r), \tag{A.1.4}$$

$$\boldsymbol{\Phi}_\pm = \mathsf{diag}(\{1 \pm \cos\theta_k\}_{k=1}^r), \tag{A.1.5}$$

*and* $\mathsf{colnorm}$ *is a column-wise normalization operator.*

*Proof.* Based on the definition (A.0.9), $\boldsymbol{H}$ is written as

$$\boldsymbol{H} = [\boldsymbol{U}_{\mathcal{S}}, \boldsymbol{U}_{\mathcal{T}}][\boldsymbol{U}_{\mathcal{S}}, \boldsymbol{U}_{\mathcal{T}}]^\top = \frac{1}{2}[\boldsymbol{U}_{\mathcal{S}}, \boldsymbol{U}_{\mathcal{T}}]\begin{bmatrix}\boldsymbol{S} & \boldsymbol{S}\\ \boldsymbol{T} & -\boldsymbol{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{S} & \boldsymbol{S}\\ \boldsymbol{T} & -\boldsymbol{T}\end{bmatrix}^\top[\boldsymbol{U}_{\mathcal{S}}, \boldsymbol{U}_{\mathcal{T}}]^\top, \tag{A.1.6}$$

where we apply the orthonormal matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ in (A.0.2) to provide

$$\boldsymbol{I} = \frac{1}{2}\begin{bmatrix}\boldsymbol{S} & \boldsymbol{S}\\ \boldsymbol{T} & -\boldsymbol{T}\end{bmatrix}\begin{bmatrix}\boldsymbol{S} & \boldsymbol{S}\\ \boldsymbol{T} & -\boldsymbol{T}\end{bmatrix}^\top. \tag{A.1.7}$$

By using the following relationships,

$$\frac{1}{2}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})^\top(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}) = \boldsymbol{I} + \boldsymbol{\Gamma}, \tag{A.1.8}$$

$$\frac{1}{2}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})^\top(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}) = \boldsymbol{0}, \tag{A.1.9}$$

$$\frac{1}{2}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})^\top(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}) = \boldsymbol{I} - \boldsymbol{\Gamma}, \tag{A.1.10}$$

(A.1.6) is reduced into

$$\boldsymbol{H} = \frac{1}{2}[\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}, \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}][\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}, \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}]^\top \tag{A.1.11}$$

$$= [\boldsymbol{U}_+, \boldsymbol{U}_-]\begin{bmatrix}\boldsymbol{I} + \boldsymbol{\Gamma} & \\ & \boldsymbol{I} - \boldsymbol{\Gamma}\end{bmatrix}^\top[\boldsymbol{U}_+, \boldsymbol{U}_-]^\top, \tag{A.1.12}$$

which is the eigen-decomposition of $\boldsymbol{H}$ using $\boldsymbol{\Gamma} = \mathrm{diag}(\{\cos\theta_k\}_{k=1}^r)$ in (A.0.3) and

$$\boldsymbol{U}_{\pm} = \mathsf{colnorm}\{\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} \pm \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}\} = \frac{1}{\sqrt{2}}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} \pm \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})(\boldsymbol{I} \pm \boldsymbol{\Gamma})^{-\frac{1}{2}}. \tag{A.1.13}$$

As to $\boldsymbol{G}$, for simplicity, we focus on the $k$-th component in (A.0.5) as

$$\boldsymbol{G}_{[k]} = [\boldsymbol{U}_{\mathcal{S}}\boldsymbol{s}_k, \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{s}}_k] \begin{bmatrix} 1+\frac{\sin(2\theta_k)}{2\theta_k} & \frac{\cos(2\theta_k)-1}{2\theta_k} \\ \frac{\cos(2\theta_k)-1}{2\theta_k} & 1-\frac{\sin(2\theta_k)}{2\theta_k} \end{bmatrix} [\boldsymbol{U}_{\mathcal{S}}\boldsymbol{s}_k, \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{s}}_k]^\top, \tag{A.1.14}$$

and omit the subscript $k$ from now on. We can rewrite the parts based on the canonical angle $\theta$ by using $\mathrm{sinc}\,\theta = \frac{\sin\theta}{\theta}$ for $0 \leq \theta \leq \frac{\pi}{2}$, into

$$1+\frac{\sin(2\theta)}{2\theta} = 1+\mathrm{sinc}\,\theta\cos\theta, \tag{A.1.15}$$

$$\frac{\cos(2\theta)-1}{2\theta} = -\mathrm{sinc}\,\theta\sin\theta, \tag{A.1.16}$$

$$1-\frac{\sin(2\theta)}{2\theta} = 1-\mathrm{sinc}\,\theta\cos\theta. \tag{A.1.17}$$

Thereby, we have

$$\begin{bmatrix} 1+\frac{\sin(2\theta)}{2\theta} & \frac{\cos(2\theta)-1}{2\theta} \\ \frac{\cos(2\theta)-1}{2\theta} & 1-\frac{\sin(2\theta)}{2\theta} \end{bmatrix} \tag{A.1.18}$$

$$= \boldsymbol{I} + \mathrm{sinc}\,\theta \begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix} \tag{A.1.19}$$

$$= \boldsymbol{I} + \mathrm{sinc}\,\theta \begin{bmatrix} 2\cos^2\frac{\theta}{2}-1 & -2\cos\frac{\theta}{2}\sin\frac{\theta}{2} \\ -2\cos\frac{\theta}{2}\sin\frac{\theta}{2} & 2\sin^2\frac{\theta}{2}-1 \end{bmatrix} \tag{A.1.20}$$

$$= (1-\mathrm{sinc}\,\theta)\boldsymbol{I} + 2\mathrm{sinc}\,\theta \begin{bmatrix} \cos\frac{\theta}{2} \\ -\sin\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} \\ -\sin\frac{\theta}{2} \end{bmatrix}^\top \tag{A.1.21}$$

$$= \begin{bmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} 1+\mathrm{sinc}\theta & \\ & 1-\mathrm{sinc}\theta \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}^\top. \tag{A.1.22}$$

This is put into (A.1.14) to produce

$$\boldsymbol{G}_{[k]} = [\boldsymbol{u}_{+k}, \boldsymbol{u}_{-k}] \begin{bmatrix} 1+\mathrm{sinc}\,\theta_k & \\ & 1-\mathrm{sinc}\,\theta_k \end{bmatrix} [\boldsymbol{u}_{+k}, \boldsymbol{u}_{-k}]^\top, \tag{A.1.23}$$

where we apply (A.1.32) and (A.1.38) to provide

$$\boldsymbol{u}_{+k} = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{s}_k\cos\frac{\theta_k}{2} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{s}}_k\sin\frac{\theta_k}{2}, \quad \boldsymbol{u}_{-k} = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{s}_k\sin\frac{\theta_k}{2} + \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{s}}_k\cos\frac{\theta_k}{2}. \tag{A.1.24}$$

Thus, through aggregating (A.1.23) for all $k$, we can finally obtain the eigen-decomposition of $\boldsymbol{G}$ in (A.1.1). □

We here show the following lemma regarding the principal component subspace $\boldsymbol{U}_+$.

**Lemma 1.** *The principal component subspace $\boldsymbol{U}_+ = \mathsf{colnorm}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}+\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})$ is the center point on the geodesic path (A.0.1) as*

$$\boldsymbol{U}_+ = \boldsymbol{U}_{(t=\frac{1}{2})} = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\boldsymbol{\Gamma}_{(t=\frac{1}{2})} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\boldsymbol{\Sigma}_{(t=\frac{1}{2})}. \tag{A.1.25}$$

*Proof.* In (A.0.2), for clarity, we describe $\boldsymbol{\Gamma} = \cos\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma} = \sin\boldsymbol{\Theta}$ with $\boldsymbol{\Theta} = \mathsf{diag}(\{\theta_k\}_{k=1}^r)$, and thereby obtain

$$\boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T} + \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} = \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}(\cos\boldsymbol{\Theta}+\boldsymbol{I}) = 2\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\cos^2\frac{\boldsymbol{\Theta}}{2}, \tag{A.1.26}$$

$$\bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{U}}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T} = -\bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\sin\boldsymbol{\Theta} = -2\bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\sin\frac{\boldsymbol{\Theta}}{2}\cos\frac{\boldsymbol{\Theta}}{2}. \tag{A.1.27}$$

The center point on the geodesic path is thus written by

$$\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\boldsymbol{\Gamma}_{(t=\frac{1}{2})} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\boldsymbol{\Sigma}_{(t=\frac{1}{2})} \tag{A.1.28}$$

$$= \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\cos\frac{\boldsymbol{\Theta}}{2} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\sin\frac{\boldsymbol{\Theta}}{2} \tag{A.1.29}$$

$$= \frac{1}{2}\left\{\boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T} + \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{U}}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}\right\}\cos^{-1}\frac{\boldsymbol{\Theta}}{2} \tag{A.1.30}$$

$$= \frac{1}{2}\left\{\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + (\boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top + \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{U}}_{\mathcal{S}}^\top)\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}\right\}\cos^{-1}\frac{\boldsymbol{\Theta}}{2} \tag{A.1.31}$$

$$= \frac{1}{2}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})\cos^{-1}\frac{\boldsymbol{\Theta}}{2} = \boldsymbol{U}_+, \tag{A.1.32}$$

where we use

$$\boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top + \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{U}}_{\mathcal{S}}^\top = \boldsymbol{I}, \tag{A.1.33}$$

$$\boldsymbol{U}_+ = \mathsf{colnorm}\{\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} + \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}\}, \tag{A.1.34}$$

$$(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}+\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})^\top(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}+\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}) = 2\boldsymbol{I} + 2\cos\boldsymbol{\Theta} = 4\cos^2\frac{\boldsymbol{\Theta}}{2}. \tag{A.1.35}$$

Similarly, we can also obtain

$$\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S}\sin\frac{\boldsymbol{\Theta}}{2} + \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{S}}\cos\frac{\boldsymbol{\Theta}}{2} \tag{A.1.36}$$

$$= \frac{1}{2}\{-\boldsymbol{U}_{\mathcal{S}}\boldsymbol{U}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T} + \boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \bar{\boldsymbol{U}}_{\mathcal{S}}\bar{\boldsymbol{U}}_{\mathcal{S}}^\top\boldsymbol{U}_{\mathcal{T}}\boldsymbol{T}\}\sin^{-1}\frac{\boldsymbol{\Theta}}{2} \tag{A.1.37}$$

$$= \frac{1}{2}(\boldsymbol{U}_{\mathcal{S}}\boldsymbol{S} - \boldsymbol{U}_{\mathcal{T}}\boldsymbol{T})\sin^{-1}\frac{\boldsymbol{\Theta}}{2} = \boldsymbol{U}_-. \tag{A.1.38}$$

□

## A.2 Proof for Proposition 2

**Proposition 2.** *Suppose backbone DNN can flexibly produce a feature vector $\boldsymbol{x}$ which is normalized as $\|\boldsymbol{x}\|_2 = 1$; for a well separable classifier $\boldsymbol{W} = \{\boldsymbol{w}_c\}_{c=1}^C$, it can produce $\{\boldsymbol{x}_i\}_{i=1}^n$ such that $\boldsymbol{w}_{y_i}^\top(\boldsymbol{P}_{\mathcal{S}}+\boldsymbol{P}_{\mathcal{T}})\boldsymbol{x}_i \geq \boldsymbol{w}_c^\top(\boldsymbol{P}_{\mathcal{S}}+\boldsymbol{P}_{\mathcal{T}})\boldsymbol{x}_i\ \forall c$, where $\boldsymbol{P}_{\mathcal{S}}$ is a r-rank subspace projection matrix for $\boldsymbol{x}_i$, i.e., $\boldsymbol{x}_i = \boldsymbol{P}_{\mathcal{S}}\boldsymbol{x}_i$, and $\boldsymbol{P}_{\mathcal{T}}$ is an arbitrary subspace of rank r. We partition a feature space into $\mathbb{S}(\boldsymbol{\theta};\boldsymbol{P}_{\mathcal{T}}) = \{\boldsymbol{x}\in\boldsymbol{P}_{\mathcal{S}} \mid \|\boldsymbol{x}\|_2 = 1, \angle(\boldsymbol{P}_{\mathcal{S}},\boldsymbol{P}_{\mathcal{T}}) = \boldsymbol{\theta}\}$ where $\angle$ is an operator to* *measure canonical angles. We define a softmax loss $\ell_{CE}$ optimized w.r.t $\boldsymbol{x} \in \mathbb{S}(\boldsymbol{\theta};\boldsymbol{P}_\mathcal{T})$ and $\boldsymbol{W}$, given $\boldsymbol{\theta}$ and $\boldsymbol{P}_\mathcal{T}$ as*

$$\ell_{CE}(\boldsymbol{\theta};\boldsymbol{P}_\mathcal{T}) = \min_{\{\boldsymbol{x}_i\in\mathbb{S}(\boldsymbol{\theta};\boldsymbol{P}_\mathcal{T})\}_{i=1}^n,\boldsymbol{W}} -\sum_{i=1}^n \log\frac{\exp(\boldsymbol{w}_{y_i}^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i)}{\exp(\sum_c \boldsymbol{w}_c^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i)}. \tag{A.2.1}$$

*Then, we have the following relationship between the loss and the canonical angle $\boldsymbol{\theta}$;*

$$\boldsymbol{\theta}^* \leq \boldsymbol{\theta} \;\Rightarrow\; \ell_{CE}(\boldsymbol{\theta}^*;\boldsymbol{P}_\mathcal{T}) \leq \ell_{CE}(\boldsymbol{\theta};\boldsymbol{P}_\mathcal{T}). \tag{A.2.2}$$

*Proof.* Without loss of generality, such as by relying on representor theorem [6], the classifier weight vector $\boldsymbol{w}$ is described by

$$\boldsymbol{w} = \sum_i(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i\alpha_i = (\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{w}_\mathcal{S}, \quad \boldsymbol{w}_\mathcal{S}\in\boldsymbol{P}_\mathcal{S}. \tag{A.2.3}$$

Thus, the classifier logit is given by

$$\boldsymbol{w}^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x} = \boldsymbol{w}_\mathcal{S}^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})^2\boldsymbol{x} \tag{A.2.4}$$
$$= \boldsymbol{w}_\mathcal{S}^\top\boldsymbol{P}_\mathcal{S}(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})^2\boldsymbol{P}_\mathcal{S}\boldsymbol{x} = \boldsymbol{w}_\mathcal{S}^\top(\boldsymbol{P}_\mathcal{S}+3\boldsymbol{P}_\mathcal{S}\boldsymbol{P}_\mathcal{T}\boldsymbol{P}_\mathcal{S})\boldsymbol{x} \tag{A.2.5}$$
$$= \boldsymbol{w}_\mathcal{S}^\top\boldsymbol{U}_\mathcal{S}\boldsymbol{S}(\boldsymbol{I}+3\boldsymbol{\Gamma}^2)\boldsymbol{S}^\top\boldsymbol{U}_\mathcal{S}^\top\boldsymbol{x} \tag{A.2.6}$$
$$= \boldsymbol{w}_\mathcal{S}^\top\boldsymbol{U}_\mathcal{S}\boldsymbol{S}(\boldsymbol{I}+3\cos^2\boldsymbol{\Theta})\boldsymbol{S}^\top\boldsymbol{U}_\mathcal{S}^\top\boldsymbol{x}, \tag{A.2.7}$$

where we use (A.0.2) and

$$\boldsymbol{w}_\mathcal{S} = \boldsymbol{P}_\mathcal{S}\boldsymbol{w}_\mathcal{S}, \quad \boldsymbol{x} = \boldsymbol{P}_\mathcal{S}\boldsymbol{x}, \tag{A.2.8}$$
$$\boldsymbol{P}_\mathcal{S} = \boldsymbol{U}_\mathcal{S}\boldsymbol{U}_\mathcal{S}^\top, \quad \boldsymbol{P}_\mathcal{T} = \boldsymbol{U}_\mathcal{T}\boldsymbol{U}_\mathcal{T}^\top, \quad \boldsymbol{P}_\mathcal{S}^2 = \boldsymbol{P}_\mathcal{S}, \quad \boldsymbol{P}_\mathcal{T}^2 = \boldsymbol{P}_\mathcal{T}. \tag{A.2.9}$$

For $\boldsymbol{P}_\mathcal{S}^*$ which produces the canonical angles $\boldsymbol{\theta}^*$ with $\boldsymbol{\theta}^* \leq \boldsymbol{\theta}$, we can have $\boldsymbol{x}^*$ and $\boldsymbol{w}^*$ such that

$$s(\boldsymbol{I}+3\cos^2\boldsymbol{\Theta}^*)(\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*)^\top\boldsymbol{x}^* = (\boldsymbol{I}+3\cos^2\boldsymbol{\Theta})(\boldsymbol{U}_\mathcal{S}\boldsymbol{S})^\top\boldsymbol{x}, \tag{A.2.10}$$
$$(\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*)^\top\boldsymbol{w}^* = (\boldsymbol{U}_\mathcal{S}\boldsymbol{S})^\top\boldsymbol{w}, \tag{A.2.11}$$

where $s\leq 1$ since $\cos^2\boldsymbol{\Theta}^* \geq \cos^2\boldsymbol{\Theta}$ and $\|(\boldsymbol{U}_\mathcal{S}\boldsymbol{S})^\top\boldsymbol{x}\|_2 = \|(\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*)^\top\boldsymbol{x}^*\|_2 = 1$. Therefore, differences between class-wise logits are described by

$$0 < (\boldsymbol{w}_y-\boldsymbol{w}_c)^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x} \tag{A.2.12}$$
$$= (\boldsymbol{w}_{\mathcal{S}y}-\boldsymbol{w}_{\mathcal{S}c})^\top\boldsymbol{U}_\mathcal{S}\boldsymbol{S}(\boldsymbol{I}+3\cos^2\boldsymbol{\Theta})(\boldsymbol{U}_\mathcal{S}\boldsymbol{S})^\top\boldsymbol{x} \tag{A.2.13}$$
$$= s(\boldsymbol{w}_{\mathcal{S}y}^*-\boldsymbol{w}_{\mathcal{S}c}^*)^\top\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*(\boldsymbol{I}+3\cos^2\boldsymbol{\Theta}^*)(\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*)^\top\boldsymbol{x}^* \tag{A.2.14}$$
$$\leq (\boldsymbol{w}_{\mathcal{S}y}^*-\boldsymbol{w}_{\mathcal{S}c}^*)^\top\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*(\boldsymbol{I}+3\cos^2\boldsymbol{\Theta}^*)(\boldsymbol{U}_\mathcal{S}^*\boldsymbol{S}^*)^\top\boldsymbol{x}^* \tag{A.2.15}$$
$$= (\boldsymbol{w}_y^*-\boldsymbol{w}_c^*)^\top(\boldsymbol{P}_\mathcal{S}^*+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}^*, \tag{A.2.16}$$

which leads to

$$\ell_{CE}(\boldsymbol{\theta};\boldsymbol{P}_\mathcal{T}) = -\sum_{i=1}^n\log\frac{\exp(\boldsymbol{w}_{y_i}^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i)}{\exp(\sum_c\boldsymbol{w}_c^\top(\boldsymbol{P}_\mathcal{S}+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i)} \tag{A.2.17}$$
$$\geq -\sum_{i=1}^n\log\frac{\exp(\boldsymbol{w}_{y_i}^{*\top}(\boldsymbol{P}_\mathcal{S}^*+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i^*)}{\exp(\sum_c\boldsymbol{w}_c^{*\top}(\boldsymbol{P}_\mathcal{S}^*+\boldsymbol{P}_\mathcal{T})\boldsymbol{x}_i^*)} \tag{A.2.18}$$
$$\geq \ell_{CE}(\boldsymbol{\theta}^*;\boldsymbol{P}_\mathcal{T}). \tag{A.2.19}$$

□

**Algorithm 1** Subspace extraction module

**Input:** $\boldsymbol{X} \in \mathbb{R}^{d \times B}$: features on a mini-batch of size $B$,
$r$: subspace rank,
$\mathtt{M} \in \mathbb{R}^{d \times Q}$: memory bank to store samples,
**Output:** $\boldsymbol{P} \in \mathbb{R}^{d \times d}$: subspace projection matrix

$[\boldsymbol{X}, \mathtt{M}] = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ : SVD [8]
$\boldsymbol{P} = \boldsymbol{U}_{:r}\boldsymbol{U}_{:r}^\top$ : extract the first $r$-rank basis vectors
Enqueue $\boldsymbol{X}$ to $\mathtt{M}$ and dequeue old samples from $\mathtt{M}$
**return** $\boldsymbol{P}$

# B Subspace computation

As shown in Algorithm 1, a subspace is computed by applying stable SVD [8] to mini-batch samples and memory-bank samples. The memory bank which stores previous mini-batch samples is introduced in order to further stabilize the subspace computation by smoothly updating subspaces through mini-batch training; the performance is improved as shown in Table B.1. It should be noted that the back-propagation is not applied to the memory-bank samples, and this subspace computation is applied only on training as described in Sec. 2.3.

Table B.1: Performance results by memory bank.

| memory size | Office-31 | | | | | | |
|---|---|---|---|---|---|---|---|
| | A→ | | D→ | | W→ | | |
| | D | W | A | W | A | D | Avg. |
| $Q = 0$ | 86.35 | 87.55 | 69.65 | 97.74 | 68.44 | 100.00 | 84.95 |
| $Q = 32$ | 88.55 | 88.18 | 72.20 | 97.86 | 72.56 | 100.00 | 86.56 |

# C Training protocol in deep domain adaptation

The deep domain adaptation is applied to the datasets in Table C.1 and trained with the following procedure. $B$ image samples are randomly drawn from source and target domains, respectively, to simultaneously construct domain-specific mini-batches of size $B$. They are passed through the backbone of ResNet-50 [2] and the projection head to produce $d$-dimensional normalized features on which the domain adaptation methods work such as by means of transformation and/or regularization losses. A fully-connected classifier is finally applied to the features for constructing a primary loss based on softmax cross-entropy. The whole network is trained by SGD optimizer with momentum of 0.9, weight decay of 0.0005 and initial learning rate of 0.0003 for a backbone and 0.003 for other modules which are decayed by $(1 + 0.0003 \cdot t)^{-0.75}$ where $t$ is the optimization step; the ResNet-50 backbone is pre-trained on ImageNet and thus is subject to smaller learning rate. The other parameters for respective datasets are shown in Table C.2.

Table C.1: Datasets.

| Dataset | Domains | | Classes | Samples |
|---|---|---|---|---|
| Office-31 [5] | 3 | Amazon, DSLR, Webcam | 31 | 4,110 |
| Office-Home [7] | 4 | Artistic, Clipart, Product, Real-world | 65 | 15,500 |
| Adaptiope [4] | 3 | Product, Real, Synthetic | 123 | 36,900 |
| DomainNet [3] | 6 | Clipart, Infograph, Paint, Quickdraw, Real, Sketch | 345 | 569,010 |

Table C.2: Training parameters.

| Dataset | Feature dim. $d$ | Subspace rank $r$ | Batch size $B$ | Memory size $Q$ | Training steps |
|---|---|---|---|---|---|
| Office-31 | 256 | 32 | 32 | 32 | 25,000 |
| Office-Home | 256 | 64 | 64 | 64 | 40,000 |
| Adaptiope | 512 | 128 | 128 | 128 | 80,000 |
| DomainNet | 512 | 352 | 64 | 320 | 160,000 |

# D Detailed performance results

In Table 4, we showed performance comparison on the tasks of multi-target domain adaptation, though averaging classification accuracies over the multiple target domains. Tables D.1∼D.3 detail the performances by reporting accuracies on respective target domains.

# References


[1] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *CVPR*, pages 2066–2073, 2012.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[3] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, pages 1406–1415, 2019.

[4] Tobias Ringwald and Rainer Stiefelhagen. Adaptiope: A modern benchmark for unsupervised domain adaptation. In *WACV*, pages 101–110, 2021.

[5] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010.

[6] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[7] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, pages 5018–5027, 2017.

[8] Wei Wang, Zheng Dang, Yinlin Hu, Pascal Fua, and Mathieu Salzmann. Robust differentiable svd. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Table D.1: Detailed performance results (accuracy %) of domain adaptation on Office-31 dataset.

(a) Single-target domain adaptation (see Table 1 in the main manuscript)

| Method | Trans. matrix | A→D | A→W | D→A | D→W | W→A | W→D | Avg. |
|---|---|---|---|---|---|---|---|---|
| Raw | $\boldsymbol{I}$ | 82.93 | 78.99 | 66.13 | 97.86 | 65.74 | 100.00 | 81.94 |
| Auto-Corr. | $\boldsymbol{A}^{\frac{1}{2}}$ | 23.29 | 28.93 | 15.55 | 18.11 | 15.65 | 26.10 | 21.27 |
| CORAL | $\boldsymbol{C}_{\mathcal{T}}^{\frac{1}{2}}\boldsymbol{C}_{\mathcal{S}}^{-\frac{1}{2}}$ | 76.51 | 77.23 | 52.75 | 85.28 | 56.66 | 96.79 | 74.20 |
| Principal component | $\boldsymbol{U}_{+}\boldsymbol{U}_{+}^{\top}$ | 81.93 | 78.11 | 64.64 | 95.85 | 65.32 | 99.00 | 80.81 |
| GFK | $\boldsymbol{G}^{\frac{1}{2}}$ | 85.14 | 84.53 | 70.57 | 97.74 | 69.47 | 100.00 | 84.58 |
| Sum-of-subspaces | $\boldsymbol{H}^{\frac{1}{2}}$ | 84.74 | 85.03 | 71.00 | 97.74 | 69.86 | 100.00 | 84.73 |
| Ours | $\boldsymbol{H}$ | 86.55 | 85.91 | 71.21 | 97.74 | 70.04 | 100.00 | 85.24 |

| Trans. | Reg. | Cls. Sub. | A→D | A→W | D→A | D→W | W→A | W→D | Avg. |
|---|---|---|---|---|---|---|---|---|---|
| - | - | - | 82.93 | 78.99 | 66.13 | 97.86 | 65.74 | 100.00 | 81.94 |
| - | ✓ | - | 82.33 | 84.78 | 72.20 | 97.23 | 70.36 | 99.80 | 84.45 |
| ✓ | - | - | 86.55 | 85.91 | 71.21 | 97.74 | 70.04 | 100.00 | 85.24 |
| ✓ | ✓ | - | 87.15 | 85.79 | 72.10 | 96.98 | 71.32 | 100.00 | 85.56 |
| ✓ | ✓ $(\eta = 2)$ | - | 84.54 | 86.79 | 71.49 | 96.86 | 68.05 | 99.80 | 84.59 |
| ✓ | ✓ | ✓ | 87.15 | 87.92 | 72.60 | 98.11 | 73.02 | 100.00 | 86.47 |

(b) Multi-target domain adaptation (see Table 2 in the main manuscript)

| | | Office-31 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | A→ | | D→ | | W→ | | |
| | | D | W | A | W | A | D | Avg. |
| *Domain Concat.* | raw | 82.93 | 78.99 | 66.13 | 97.86 | 65.74 | 100 | 81.94 |
| | DANN | 80.93 | 82.26 | 63.08 | 96.98 | 65.74 | 99.60 | 81.43 |
| | BNM | 85.94 | 87.30 | 61.85 | 94.84 | 63.22 | 98.19 | 81.89 |
| | SCDA | 85.14 | 87.04 | 59.67 | 93.08 | 63.54 | 99.20 | 81.28 |
| | DSAN | 87.15 | 89.18 | 65.39 | 96.48 | 64.71 | 98.80 | 83.62 |
| | Ours-c | 88.96 | 87.80 | 71.92 | 97.86 | 71.35 | 100 | 86.31 |
| *Multi.* | DANN | 84.94 | 86.16 | 65.85 | 94.59 | 69.72 | 98.19 | 83.24 |
| | BNM | 84.94 | 87.17 | 60.67 | 90.31 | 65.81 | 98.19 | 81.18 |
| | SCDA | 85.94 | 87.67 | 60.28 | 92.08 | 64.50 | 99.60 | 81.68 |
| | DSAN | 84.74 | 86.92 | 61.24 | 92.70 | 65.07 | 98.39 | 81.51 |
| | Ours | 88.55 | 88.18 | 72.20 | 97.86 | 72.56 | 100 | 86.56 |
| *Joint method* | +DANN | 88.35 | 89.94 | 72.52 | 97.99 | 72.35 | 100 | 86.86 |
| | +BNM | 91.77 | 93.84 | 74.80 | 97.86 | 74.76 | 100 | 88.84 |
| | +SCDA | 88.96 | 90.94 | 72.38 | 97.99 | 72.02 | 100 | 87.05 |
| | +DSAN | 89.96 | 91.95 | 72.95 | 97.86 | 73.09 | 100 | 87.64 |

Table D.2: Detailed performance results (accuracy %) of multi-target domain adaptation on Office-Home dataset.

| | | Office-Home | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A→ | | | C→ | | | P→ | | | R→ | | | |
| | | C | P | R | A | P | R | A | C | R | A | C | P | Avg. |
| *Domain Concat.* | raw | 46.03 | 55.96 | 67.34 | 52.62 | 60.22 | 61.90 | 54.64 | 47.40 | 73.08 | 67.04 | 52.76 | 77.07 | 59.67 |
| | DANN | 39.06 | 44.51 | 57.54 | 47.14 | 51.50 | 52.54 | 52.45 | 47.38 | 65.14 | 67.28 | 50.49 | 73.03 | 54.00 |
| | BNM | 33.01 | 36.56 | 49.53 | 46.44 | 48.50 | 50.42 | 49.49 | 44.97 | 62.91 | 66.38 | 50.95 | 72.81 | 51.00 |
| | SCDA | 37.75 | 41.23 | 55.18 | 46.72 | 50.46 | 51.82 | 48.25 | 43.89 | 65.60 | 66.01 | 49.55 | 72.83 | 52.44 |
| | DSAN | 39.06 | 42.85 | 55.72 | 46.85 | 49.90 | 50.81 | 49.40 | 44.95 | 65.02 | 65.55 | 50.06 | 72.92 | 52.76 |
| | Ours-c | 56.22 | 70.24 | 76.04 | 60.65 | 68.75 | 71.72 | 61.39 | 50.42 | 77.85 | 69.80 | 55.72 | 80.33 | 66.59 |
| *Multi.* | DANN | 40.30 | 45.93 | 55.68 | 46.89 | 54.29 | 54.42 | 49.07 | 47.61 | 64.72 | 65.60 | 53.61 | 72.29 | 54.20 |
| | BNM | 34.69 | 38.59 | 51.50 | 45.57 | 49.47 | 50.77 | 47.30 | 45.20 | 62.68 | 65.22 | 51.18 | 73.35 | 51.29 |
| | SCDA | 38.58 | 43.73 | 56.55 | 47.14 | 52.58 | 53.34 | 49.03 | 44.83 | 65.71 | 66.30 | 52.19 | 74.93 | 53.74 |
| | DSAN | 31.07 | 34.83 | 46.39 | 44.05 | 46.14 | 47.90 | 46.11 | 44.08 | 61.79 | 64.94 | 50.36 | 72.43 | 49.17 |
| | Ours | 56.79 | 70.33 | 76.38 | 60.69 | 70.33 | 71.93 | 61.52 | 52.74 | 77.74 | 69.02 | 58.79 | 81.39 | 67.30 |
| *Joint method* | +DANN | 55.03 | 69.61 | 75.83 | 62.22 | 71.95 | 72.53 | 62.05 | 56.66 | 77.76 | 72.11 | 60.57 | 82.16 | 68.21 |
| | +BNM | 58.63 | 71.80 | 77.09 | 64.52 | 73.89 | 74.39 | 65.10 | 56.24 | 78.88 | 70.58 | 60.60 | 82.36 | 69.51 |
| | +SCDA | 56.98 | 70.44 | 76.29 | 61.31 | 70.56 | 71.93 | 61.76 | 53.29 | 77.90 | 69.26 | 58.81 | 81.44 | 67.50 |
| | +DSAN | 56.93 | 71.07 | 76.70 | 62.01 | 71.64 | 72.34 | 62.96 | 54.20 | 77.87 | 69.02 | 59.18 | 81.84 | 67.98 |

| | | Office-Home | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A,C→ | | P,R→ | | A,P→ | | C,R→ | | A,R→ | | C,P→ | | |
| | | P | R | A | C | C | R | A | P | C | P | A | R | Avg. |
| *Domain Concat.* | raw | 59.18 | 66.01 | 63.29 | 51.00 | 53.65 | 74.78 | 63.29 | 70.96 | 51.73 | 66.84 | 55.38 | 67.91 | 62.00 |
| | DANN | 56.23 | 63.94 | 65.27 | 53.88 | 55.03 | 74.23 | 63.37 | 67.06 | 49.46 | 65.62 | 52.00 | 65.21 | 60.94 |
| | BNM | 54.58 | 61.65 | 65.02 | 53.77 | 51.89 | 71.38 | 62.09 | 67.56 | 47.22 | 60.58 | 56.94 | 65.39 | 59.84 |
| | SCDA | 56.23 | 63.62 | 63.86 | 51.84 | 53.13 | 73.79 | 62.51 | 68.55 | 50.84 | 64.36 | 55.13 | 66.08 | 60.83 |
| | DSAN | 56.84 | 63.32 | 64.48 | 52.46 | 54.14 | 73.86 | 62.13 | 67.88 | 51.41 | 64.11 | 55.87 | 65.64 | 61.01 |
| | Ours-c | 76.05 | 79.39 | 70.21 | 56.88 | 59.54 | 82.05 | 70.70 | 81.19 | 61.67 | 80.56 | 66.87 | 80.08 | 72.10 |
| *Multi.* | DANN | 50.03 | 51.66 | 62.55 | 52.00 | 43.99 | 62.43 | 62.88 | 71.39 | 52.88 | 72.59 | 45.49 | 62.34 | 57.52 |
| | BNM | 50.01 | 51.00 | 64.85 | 51.71 | 45.11 | 62.22 | 65.72 | 73.46 | 51.43 | 73.39 | 49.20 | 62.86 | 58.41 |
| | SCDA | 50.53 | 51.85 | 64.69 | 50.97 | 45.02 | 64.72 | 65.27 | 74.03 | 51.41 | 73.98 | 49.16 | 64.86 | 58.87 |
| | DSAN | 37.17 | 41.15 | 63.58 | 48.43 | 42.93 | 60.55 | 65.06 | 72.65 | 49.64 | 72.88 | 45.57 | 60.82 | 55.04 |
| | Ours | 75.13 | 78.75 | 69.39 | 60.50 | 59.93 | 82.30 | 70.00 | 81.93 | 61.03 | 81.12 | 66.17 | 80.63 | 72.24 |
| *Joint method* | +DANN | 76.12 | 78.66 | 71.61 | 61.58 | 63.41 | 82.99 | 71.90 | 83.31 | 61.58 | 81.10 | 64.52 | 79.57 | 73.03 |
| | +BNM | 77.79 | 80.03 | 71.20 | 62.59 | 61.95 | 82.79 | 71.78 | 83.19 | 62.73 | 82.29 | 69.02 | 81.64 | 73.92 |
| | +SCDA | 75.49 | 78.77 | 69.55 | 61.01 | 60.21 | 82.30 | 70.42 | 82.14 | 61.19 | 81.32 | 66.63 | 80.81 | 72.49 |
| | +DSAN | 76.95 | 79.23 | 69.59 | 60.96 | 60.32 | 81.98 | 70.25 | 82.54 | 61.70 | 81.21 | 66.79 | 80.42 | 72.66 |

Table D.3: Detailed performance results (accuracy %) of multi-target domain adaptation on Adaptiope and DomainNet datasets.

| | | Adaptiope | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | P→ | | R→ | | S→ | | |
| | | R | S | P | S | P | R | Avg. |
| *Domain Concat.* | raw | 68.70 | 39.91 | 76.82 | 34.28 | 51.95 | 34.62 | 51.05 |
| | DANN | 70.89 | 46.62 | 75.24 | 37.02 | 56.37 | 33.67 | 53.30 |
| | BNM | 69.11 | 39.89 | 76.20 | 33.94 | 56.37 | 35.06 | 51.76 |
| | SCDA | 68.42 | 36.20 | 76.91 | 32.58 | 53.01 | 31.07 | 49.70 |
| | DSAN | 68.76 | 38.54 | 76.93 | 33.33 | 54.63 | 32.85 | 50.84 |
| | Ours-c | 73.59 | 41.94 | 88.70 | 34.07 | 64.44 | 41.84 | 57.43 |
| *Multi.* | Ours | 71.61 | 48.58 | 87.55 | 41.63 | 66.76 | 48.60 | 60.79 |
| *Joint method* | +DANN | 74.20 | 54.59 | 90.27 | 52.09 | 70.73 | 53.50 | 65.90 |
| | +BNM | 72.03 | 52.45 | 88.11 | 43.67 | 68.72 | 50.55 | 62.59 |
| | +SCDA | 71.77 | 49.31 | 87.80 | 41.42 | 66.20 | 48.21 | 60.78 |
| | +DSAN | 72.26 | 50.49 | 88.02 | 42.45 | 67.33 | 49.52 | 61.68 |

| | | DomainNet | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | C,I,P→ | | | Q,R,S→ | | | |
| | | Q | R | S | C | I | P | Avg. |
| *Domain Concat.* | raw | 3.37 | 11.65 | 10.28 | 16.53 | 5.04 | 14.34 | 10.20 |
| | DANN | 2.42 | 10.87 | 9.13 | 16.65 | 5.26 | 14.07 | 9.73 |
| | BNM | 3.16 | 11.33 | 9.94 | 16.75 | 5.25 | 14.44 | 10.14 |
| | SCDA | 3.28 | 11.51 | 10.10 | 16.31 | 5.00 | 14.19 | 10.07 |
| | DSAN | 2.59 | 10.26 | 8.27 | 16.21 | 4.42 | 13.01 | 9.13 |
| | Ours-c | 10.34 | 41.32 | 35.09 | 45.46 | 15.27 | 38.65 | 31.02 |
| *Multi.* | Ours | 12.92 | 63.48 | 50.28 | 61.84 | 20.56 | 50.50 | 43.26 |
| *Joint method* | +DANN | 12.53 | 61.75 | 48.77 | 60.52 | 20.66 | 50.04 | 42.38 |
| | +BNM | 12.67 | 63.49 | 50.39 | 61.57 | 20.38 | 50.46 | 43.16 |
| | +SCDA | 12.96 | 63.67 | 50.09 | 61.59 | 20.47 | 50.64 | 43.24 |
| | +DSAN | 12.90 | 63.69 | 50.10 | 61.47 | 20.50 | 50.78 | 43.24 |

| | | DomainNet | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | C,I→ | | | | P,Q→ | | | | R,S→ | | | | |
| | | P | Q | R | S | C | I | R | S | C | I | P | Q | Avg. |
| *Domain Concat.* | raw | 6.03 | 2.88 | 8.62 | 8.05 | 9.39 | 4.45 | 12.32 | 10.73 | 17.91 | 5.74 | 16.53 | 2.86 | 8.79 |
| | DANN | 4.95 | 2.53 | 7.14 | 6.77 | 7.96 | 3.44 | 11.39 | 8.45 | 17.76 | 5.33 | 16.52 | 2.67 | 7.91 |
| | BNM | 5.96 | 2.88 | 8.54 | 7.99 | 9.33 | 4.38 | 12.24 | 10.72 | 17.84 | 5.71 | 16.42 | 2.84 | 8.74 |
| | SCDA | 5.96 | 2.88 | 8.48 | 7.95 | 9.29 | 4.43 | 12.14 | 10.65 | 17.82 | 5.71 | 16.41 | 2.83 | 8.71 |
| | DSAN | 4.41 | 1.38 | 7.69 | 5.09 | 8.23 | 3.17 | 13.11 | 9.29 | 17.22 | 4.83 | 17.61 | 2.35 | 7.86 |
| | Ours-c | 28.20 | 8.51 | 36.94 | 31.22 | 30.57 | 13.44 | 37.50 | 29.70 | 48.49 | 18.78 | 44.26 | 8.45 | 28.00 |
| *Multi.* | Ours | 40.77 | 12.81 | 57.95 | 44.88 | 48.11 | 16.35 | 53.35 | 38.16 | 61.01 | 21.60 | 51.53 | 12.53 | 38.25 |
| *Joint method* | +DANN | 40.20 | 12.50 | 56.02 | 44.11 | 47.67 | 16.19 | 52.12 | 38.94 | 59.27 | 21.06 | 51.29 | 12.72 | 37.67 |
| | +BNM | 41.67 | 12.84 | 58.33 | 45.23 | 48.45 | 16.00 | 53.65 | 38.62 | 60.87 | 22.01 | 51.48 | 12.69 | 38.49 |
| | +SCDA | 41.67 | 12.82 | 58.25 | 45.15 | 48.25 | 16.14 | 53.59 | 38.88 | 60.90 | 21.92 | 51.48 | 12.62 | 38.47 |
| | +DSAN | 40.75 | 12.79 | 57.95 | 44.91 | 48.03 | 16.36 | 53.45 | 38.19 | 61.03 | 21.69 | 51.56 | 12.61 | 38.28 |