Supplementary Material for
Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation

Yao-Hung Hubert Tsai[1], Yi-Ren Yeh[2], Yu-Chiang Frank Wang[1]
[1]Research Center for IT Innovation,Academia Sinica,Taipei, Taiwan
[2]Department of Mathematics, National Kaohsiung Normal University, Kaohsiung, Taiwan
y.h.huberttsai@gmail.com, yryeh@nknu.edu.tw, ycwang@citi.sinica.edu.tw

We now provide technical details on the derivations of the supervised and full versions of our proposed *Cross-Domain Landmark Selection (CDLS)* algorithm. For the sake of conciseness, we do not repeat the definitions for each notation in the Supplementary.

I. Optimization of CDLS_sup

Recall that in Section 3.2.1 of our manuscript, the objective function of CDLS for heterogeneous domain adaptation (HDA) (i.e., **CDLS**_sup) is:

$$\min_{\mathbf{A}} E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + \lambda \left\|\mathbf{A}\right\|^2, \quad \text{(i)}$$

where $E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) =$

$$\left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \mathbf{A}\mathbf{x}_s^i - \frac{1}{n_L} \sum_{i=1}^{n_L} \widehat{\mathbf{x}}_l^i \right\|^2, \quad \text{(ii)}$$

and $E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) =$

$$\sum_{c=1}^{C} \frac{1}{n_S^c n_L^c} \sum_{i=1}^{n_S^c} \sum_{j=1}^{n_L^c} \left\| \mathbf{A}\mathbf{x}_s^{i,c} - \widehat{\mathbf{x}}_l^{j,c} \right\|^2 + \left\| \frac{1}{n_S^c} \sum_{i=1}^{n_S^c} \mathbf{A}\mathbf{x}_s^{i,c} - \frac{1}{n_L^c} \sum_{i=1}^{n_L^c} \widehat{\mathbf{x}}_l^{i,c} \right\|^2. \quad \text{(iii)}$$

The minimization problems of (ii) and (iii) with respect to the transformation $\mathbf{A}$ can be rewritten as follows:

$$\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{SM\_sup} \mathbf{X}_S^\top \mathbf{A} - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{LM\_sup} \widehat{\mathbf{X}}_L^\top + \text{const}, \quad \text{(iv)}$$

and

$$\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{SC\_sup} \mathbf{X}_S^\top \mathbf{A} - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{LC\_sup} \widehat{\mathbf{X}}_L^\top + \text{const}, \quad \text{(v)}$$

where

$$\mathbf{H}_{SM\_sup}, \mathbf{H}_{SC\_sup} \in \mathbb{R}^{n_S \times n_S},$$
$$\mathbf{H}_{LM\_sup}, \mathbf{H}_{LC\_sup} \in \mathbb{R}^{n_S \times n_L},$$
with entries

$$(\mathbf{H}_{SM\_sup})_{ij} = \frac{1}{n_S n_S}, \ (\mathbf{H}_{LM\_sup})_{ij} = \frac{1}{n_S n_L}$$

$$(\mathbf{H}_{SC\_sup})_{ij} = \begin{cases} \frac{1+n_S^c}{n_S^c n_S^c} & \text{if } i,j \in \text{class } c \text{ and } i = j \\ \frac{1}{n_S^c n_S^c} & \text{if } i,j \in \text{class } c \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases},$$

$$(\mathbf{H}_{LC\_sup})_{ij} = \begin{cases} \frac{2}{n_S^c n_L^c} & \text{if } i,j \in \text{class } c \\ 0 & \text{otherwise} \end{cases}.$$

By taking the derivatives of (iv) and (v) in (i) with respect to $\mathbf{A}$ and setting it as 0, the closed-form of $\mathbf{A}$ can be derived as:

$$\mathbf{A} = \left(\lambda \mathbf{I}_{d_S} + \mathbf{X}_S \left(\mathbf{H}_{SM\_sup} + \mathbf{H}_{SC\_sup}\right) \mathbf{X}_S^\top\right)^{-1} \cdot \left(\mathbf{X}_S \left(\mathbf{H}_{LM\_sup} + \mathbf{H}_{LC\_sup}\right) \widehat{\mathbf{X}}_L^\top\right). \quad \text{(vi)}$$

From the above derivations, the optimal solution $\mathbf{A}$ for our **CDLS**_sup can be obtained.

II. Optimization of CDLS

We now detail the optimization process for the full version of CDLS, which can be applied to solve semi-supervised HDA problems. For simplicity, we have $\{\mathbf{X}_S, \mathbf{X}_T, \mathbf{X}_U\}$ denote source-domain data, labeled and unlabeled target-domain data, respectively.

II.1. Derivation of Transformation A

In our work, we apply the technique of alternative optimization for solving **CDLS**. As noted in our manuscript, with fixed landmark weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the objective function for solving $\mathbf{A}$ is:

$$\min_{\mathbf{A}} E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) + E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \lambda \|\mathbf{A}\|^2, \tag{vii}$$

where $E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) =$

$$\left\| \frac{1}{\delta n_S} \sum_{i=1}^{n_S} \alpha_i \mathbf{A}\mathbf{x}_s^i - \frac{1}{n_L + \delta n_U} \left(\sum_{i=1}^{n_L} \widehat{\mathbf{x}}_l^i + \sum_{i=1}^{n_U} \beta_i \widehat{\mathbf{x}}_u^i \right) \right\|^2, \tag{viii}$$

and $E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) =$

$$\sum_{c=1}^{C} E_{cond}^c + \frac{1}{e^c} E_{embed}^c. \tag{ix}$$

Recall that, we have

$E_{cond}^c =$

$$\left\| \frac{1}{\delta n_S^c} \sum_{i=1}^{n_S^c} \alpha_i \mathbf{A}\mathbf{x}_s^{i,c} - \frac{1}{n_L^c + \delta n_U^c} \left(\sum_{i=1}^{n_L^c} \widehat{\mathbf{x}}_l^{i,c} + \sum_{i=1}^{n_U^c} \beta_i \widehat{\mathbf{x}}_u^{i,c} \right) \right\|^2,$$

$E_{embed}^c =$

$$\sum_{i=1}^{n_S^c} \sum_{j=1}^{n_L^c} \left\| \alpha_i \mathbf{A}\mathbf{x}_s^{i,c} - \widehat{\mathbf{x}}_l^{j,c} \right\|^2 + \sum_{i=1}^{n_L^c} \sum_{j=1}^{n_U^c} \left\| \widehat{\mathbf{x}}_l^{i,c} - \beta_j \widehat{\mathbf{x}}_u^{j,c} \right\|^2 + \sum_{i=1}^{n_U^c} \sum_{j=1}^{n_S^c} \left\| \beta_i \widehat{\mathbf{x}}_u^{i,c} - \alpha_j \mathbf{A}\mathbf{x}_s^{j,c} \right\|^2,$$

and $e^c = \delta n_S^c n_L^c + \delta n_L^c n_U^c + \delta^2 n_U^c n_S^c$.

With the constraint of $\frac{\boldsymbol{\alpha}^{c\top}\mathbf{1}}{n_S^c} = \frac{\boldsymbol{\beta}^{c\top}\mathbf{1}}{n_U^c} = \delta$, we rewrite (viii) and (ix) into the following formulations:

$$\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{SM} \mathbf{X}_S^\top \mathbf{A} - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{LM} \widehat{\mathbf{X}}_L^\top - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{UM} \widehat{\mathbf{X}}_U^\top + \text{const}, \tag{x}$$

and

$$\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{SC} \mathbf{X}_S^\top \mathbf{A} - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{LC} \widehat{\mathbf{X}}_L^\top - 2\mathbf{A}^\top \mathbf{X}_S \mathbf{H}_{UC} \widehat{\mathbf{X}}_U^\top + \text{const}, \tag{xi}$$

where

$$\mathbf{H}_{SM} = \frac{1}{\delta^2 n_S n_S} \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}^\top,$$
$$\mathbf{H}_{LM} = \frac{1}{\delta n_S (n_L + \delta n_U)} \boldsymbol{\alpha} \cdot \mathbf{1}_{n_L}^\top,$$
$$\mathbf{H}_{UM} = \frac{1}{\delta n_S (n_L + \delta n_U)} \boldsymbol{\alpha} \cdot \boldsymbol{\beta}^\top,$$

$$\mathbf{H}_{SC} = \begin{bmatrix} \mathbf{H}_{SC}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{SC}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{SC}^C \end{bmatrix} \textit{with}$$

$$\mathbf{H}_{SC}^c = \frac{1}{\delta^2 n_S^c n_S^c} \boldsymbol{\alpha}^c \cdot \boldsymbol{\alpha}^{c\top} + \frac{n_L^c + n_U^c}{e^c} \mathbf{diag}(\boldsymbol{\alpha}^c \odot \boldsymbol{\alpha}^c),$$

$$\mathbf{H}_{LC} = \begin{bmatrix} \mathbf{H}_{LC}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{LC}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{LC}^C \end{bmatrix} \textit{with}$$

$$\mathbf{H}_{LC}^c = \left(\frac{1}{\delta n_S^c (n_L^c + \delta n_U^c)} + \frac{1}{e^c}\right) \boldsymbol{\alpha}^c \cdot \mathbf{1}_{n_L^c}^\top,$$

$$\mathbf{H}_{UC} = \begin{bmatrix} \mathbf{H}_{UC}^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{UC}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{UC}^C \end{bmatrix} \textit{with}$$

$$\mathbf{H}_{UC}^c = \left(\frac{1}{\delta n_S^c (n_L^c + \delta n_U^c)} + \frac{1}{e^c}\right) \boldsymbol{\alpha}^c \cdot \boldsymbol{\beta}^{c\top},$$

and $\odot$ denotes element $-$ wise multiplication.

Similar to the optimization of **CDLS**_*sup*, we take the derivatives of (x) and (xi) in (vii) with respect to $\mathbf{A}$ and set it as 0. Then, we obtain the optimal solution $\mathbf{A}$ as:

$$\mathbf{A} = \left(\lambda \mathbf{I}_{d_S} + \mathbf{X}_S \mathbf{H}_S \mathbf{X}_S{}^\top\right)^{-1} \left(\mathbf{X}_S \left(\mathbf{H}_L \widehat{\mathbf{X}}_L^\top + \mathbf{H}_U \widehat{\mathbf{X}}_U^\top\right)\right), \tag{xii}$$

where

$$\mathbf{H}_S = \mathbf{H}_{SM} + \mathbf{H}_{SC}$$
$$\mathbf{H}_L = \mathbf{H}_{LM} + \mathbf{H}_{LC}$$
$$\mathbf{H}_U = \mathbf{H}_{UM} + \mathbf{H}_{UC}.$$

II.2. Derivations of Landmark Weights

Now, we fix $\mathbf{A}$ for optimizing landmark weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The objective function can be formulated as:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \; & E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \\ & E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{s.t.} \; & \{\alpha_i^c, \beta_i^c\} \in [0, 1], \; \frac{\boldsymbol{\alpha}^{c\top}\mathbf{1}}{n_S^c} = \frac{\boldsymbol{\beta}^{c\top}\mathbf{1}}{n_U^c} = \delta, \end{aligned} \tag{xiii}$$

where E_M and E_C are defined in Section II.1.

To solve the above optimization problem, a Gram matrix $\mathbf{G}$ for describing cross-domain data is introduced:

$$\begin{aligned} \mathbf{G} &\in \mathbb{R}^{(n_S + n_L + n_U) \times (n_S + n_L + n_U)} \\ &= [\mathbf{X}_S, \mathbf{X}_T, \mathbf{X}_U]^\top [\mathbf{X}_S, \mathbf{X}_T, \mathbf{X}_U]. \end{aligned}$$

For the ease of the derivations, we define four scaling matrices $\Theta_{M,C1,C2,C3}$ as follows:

$$\Theta_M = \theta_M \theta_M^\top \; \textit{with} \; \theta_M = \left[\frac{1}{\delta n_S} \mathbf{1}_{n_S}; \frac{1}{n_L + \delta n_U} \mathbf{1}_{n_L + n_U}\right],$$

$$
\begin{aligned}
\Theta_{C1} &= \sum_{c=1}^{C} \theta_{C1}^{c} {\theta_{C1}^{c}}^{\top} \; with \\
\theta_{C1}^{c} &= [\mathbf{0};\cdots;\mathbf{0};\frac{1}{\delta n_S^c}\mathbf{1}_{n_S}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\frac{1}{n_L^c+\delta n_U^c}\mathbf{1}_{n_L}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\frac{1}{n_L^c+\delta n_U^c}\mathbf{1}_{n_U}^{c};\mathbf{0};\cdots;\mathbf{0}\;], \\
\Theta_{C2} &= \sum_{c=1}^{C} \frac{1}{e^c} \theta_{C2}^{c} {\theta_{C2}^{c}}^{\top} \; with \\
\theta_{C2}^{c} &= [\mathbf{0};\cdots;\mathbf{0};\mathbf{1}_{n_S}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\mathbf{1}_{n_L}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\mathbf{1}_{n_U}^{c};\mathbf{0};\cdots;\mathbf{0}\;],
\end{aligned}
$$

and

$$
\begin{aligned}
\Theta_{C3} &= \sum_{c=1}^{C} \mathbf{diag}(\theta_{C3}^{c}) \; with \\
\theta_{C3}^{c} &= [\mathbf{0};\cdots;\mathbf{0};\frac{n_L^c+n_U^c}{e^c}\mathbf{1}_{n_S}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\frac{n_S^c+n_U^c}{e^c}\mathbf{1}_{n_L}^{c};\mathbf{0};\cdots;\mathbf{0}; \\
&\quad \mathbf{0};\cdots;\mathbf{0};\frac{n_L^c+n_S^c}{e^c}\mathbf{1}_{n_U}^{c};\mathbf{0};\cdots;\mathbf{0}\;].
\end{aligned}
$$

By integrating the scaling matrix with the Gram matrix $\mathbf{G}$, we have the following Gram matrices $\mathbf{G}_{1\sim3}$ calculated as:

$$
\begin{aligned}
\mathbf{G}_1 &= \mathbf{G} \odot (\Theta_M + \Theta_{C1} + \Theta_{C3}), \\
\mathbf{G}_2 &= \mathbf{G} \odot (\Theta_M + \Theta_{C1} + \Theta_{C2}), \\
\mathbf{G}_3 &= \mathbf{G} \odot (\Theta_M + \Theta_{C1} - \Theta_{C2}).
\end{aligned} \tag{xiv}
$$

With (xiv), the original optimization problem of (xiii) becomes:

$$
\begin{aligned}
\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \; & \frac{1}{2}\boldsymbol{\alpha}^{\top}\mathbf{K}_{S,S}\boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\beta}^{\top}\mathbf{K}_{U,U}\boldsymbol{\beta} - \boldsymbol{\alpha}^{\top}\mathbf{K}_{S,U}\boldsymbol{\beta} \\
& - {\mathbf{k}_{S,L}}^{\top}\boldsymbol{\alpha} + {\mathbf{k}_{U,L}}^{\top}\boldsymbol{\beta} + \text{const.} \\
\text{s.t.} \; & \{\alpha_i^c, \beta_i^c\} \in [0,1], \; \frac{{\boldsymbol{\alpha}^c}^{\top}\mathbf{1}}{n_S^c} = \frac{{\boldsymbol{\beta}^c}^{\top}\mathbf{1}}{n_U^c} = \delta,
\end{aligned} \tag{xv}
$$

where

$$
\begin{aligned}
\mathbf{K}_{S,S} &= \mathbf{G}_1(1:n_S, 1:n_S), \\
\mathbf{K}_{U,U} &= \mathbf{G}_1(n_S+n_L+1:\text{end}, n_S+n_L+1:\text{end}), \\
\mathbf{K}_{S,U} &= \mathbf{G}_2(1:n_S, n_S+n_L+1:\text{end}), \\
\mathbf{k}_{S,L}(i) &= \sum_{j=1}^{n_L} \mathbf{G}_2(i, n_S+j), \\
\mathbf{k}_{U,L}(i) &= \sum_{j=1}^{n_L} \mathbf{G}_3(n_S+n_L+i, n_S+j).
\end{aligned}
$$

To solve (xv), one can apply existing *Quadratic Programming* (QP) solvers and tackle the following problem instead:

$$
\min_{z_i\in[0,1],\mathbf{Z}^{\top}\mathbf{V}=\mathbf{W}} \quad \frac{1}{2}\mathbf{Z}^{\top}\mathbf{B}\mathbf{Z} + \mathbf{b}^{\top}\mathbf{Z}, \tag{xvi}
$$

where

$$
\mathbf{Z} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{K}_{S,S} & -\mathbf{K}_{S,U} \\ -\mathbf{K}_{S,U}^{\top} & \mathbf{K}_{U,U} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -\mathbf{k}_{S,L} \\ \mathbf{k}_{U,L} \end{pmatrix},
$$

$$
\begin{aligned}
\mathbf{V} &= \begin{bmatrix} \mathbf{V}_S & \mathbf{0}_{n_S\times C} \\ \mathbf{0}_{n_U\times C} & \mathbf{V}_U \end{bmatrix} \in \mathbb{R}^{(n_S+n_U)\times 2C} \text{ with} \\
(\mathbf{V}_S)_{ij} &= \begin{cases} 1 & \text{if } \mathbf{x}_s^i \in \text{class } j \\ 0 & \text{otherwise} \end{cases} \\
(\mathbf{V}_U)_{ij} &= \begin{cases} 1 & \text{if } \mathbf{x}_u^i \text{ predicted as class } j \\ 0 & \text{otherwise} \end{cases}, \\
\mathbf{W} &\in \mathbb{R}^{1\times 2C} \text{ with } (\mathbf{W})_c = \begin{cases} \delta n_S^c & \text{if } c \le C \\ \delta n_U^{c-C} & \text{if } c > C \end{cases}.
\end{aligned}
$$