Appendix

A Notation Table

We provide the important notations and their descriptions for clarification on Table.3.

Symbol	Description
$\boldsymbol{X} \in \mathbb{R}^{M\times D}$	Source domain data matrix
$\boldsymbol{Z} \in \mathbb{R}^{N\times D}$	Target domain data matrix
M	Number of source samples
N	Number of target samples
D	Data dimension
$\boldsymbol{a} \in \mathbb{R}^{M}$	Source marginal probability vector
$\boldsymbol{b} \in \mathbb{R}^{N}$	Target marginal probability vector
$\boldsymbol{\pi} \in \mathbb{R}^{M\times N}$	Coupling matching matrix (transport plan)
$\boldsymbol{C} \in \mathbb{R}^{M\times N}$	Cost (distance) matrix
τ	KL divergence coefficient for SemiUOT
τ_a	KL divergence coefficient for source (UOT)
τ_b	KL divergence coefficient for target (UOT)
$\boldsymbol{\alpha}$	Adjusted source marginal (via ETM)
$\boldsymbol{\beta}$	Adjusted target marginal (via ETM)
$\boldsymbol{f}$	Dual variable (SemiUOT, source)
$\boldsymbol{g}$	Dual variable (SemiUOT, target)
$\boldsymbol{u}$	Dual variable (UOT, source)
$\boldsymbol{v}$	Dual variable (UOT, target)
ζ	Scalar dual variable (shared)
$\boldsymbol{s}$	KKT multiplier variable
ω	Scaling factor: $\omega = \frac{\langle \boldsymbol{b}, \boldsymbol{1}\rangle}{\langle \boldsymbol{a}, \boldsymbol{1}\rangle}$
ϵ	LogSumExp smoothing parameter
ν	Update step: $\nu = \frac{\tau\epsilon}{\tau+\epsilon}$
J_{SemiUOT}	Objective function of SemiUOT
J_{UOT}	Objective function of UOT
L_P	Exact SemiUOT Equation
$\hat{L}_P$	Approximate SemiUOT Equation
L_U	Exact UOT Equation
$\hat{L}_U$	Approximate UOT Equation
$G(\boldsymbol{\pi}, \boldsymbol{s})$	KKT-multiplier regularization term: $\langle \boldsymbol{\pi}, \boldsymbol{s}\rangle$
$L_{\text{Reg}}(\boldsymbol{\pi})$	Regularization term for OT (e.g. entropy or ℓ_2)
η_G	Regularization weight for multiplier term
η_{Reg}	Regularization weight (entropy or ℓ_2)
$\tilde{C}_{ij}$	Adjusted cost: $\tilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$
$\boldsymbol{\psi}$	Dual variable in MROT (source side)
$\boldsymbol{\phi}$	Dual variable in MROT (target side)

Table 3: Important notations

B Proof of Proposition 1

Proposition 1. (Principles of Equivalent Transformation Mechanism for SemiUOT) *Given SemiUOT with KL-Divergence J_{SemiUOT}, one can obtain its Fenchel-Lagrange multipliers form as:*

$$\begin{aligned} &\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \left[\tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i+\zeta}{\tau}\right) - \sum_{j=1}^{N} b_j (g_j - \zeta) \right] \\ &s.t.\ f_i + g_j + s_{ij} = C_{ij}, \quad s_{ij} \geq 0. \end{aligned} \tag{16}$$

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ denote Lagrange multipliers. Moreover, SemiUOT problem can be further transformed into the form of classic optimal transport as follows:

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{P}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle \\ &s.t. \begin{cases} \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{f}^* + \zeta^*}{\tau}\right) = \boldsymbol{\alpha} \\ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \end{cases}. \end{aligned} \tag{17}$$

When $\tau \to \infty$, the source marginal probability is given as $\boldsymbol{\pi} \mathbf{1}_N = \omega \boldsymbol{a}$ where $\omega = \langle \boldsymbol{b}, \mathbf{1}_N \rangle / \langle \boldsymbol{a}, \mathbf{1}_M \rangle$.

Proof. To start with, we first review the definition of SemiUOT as shown below:

$$\begin{aligned} &\min_{\pi_{ij} \geq 0} J_{\mathrm{SemiUOT}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) \\ &\text{s.t. } \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}. \end{aligned} \tag{18}$$

Then we can rewrite the optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) \\ &s.t. \begin{cases} (\text{Constraint}): \ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \\ (\text{Optional}): \ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha} \end{cases}. \end{aligned} \tag{19}$$

Note that we do not need to know the exact value of $\boldsymbol{\alpha}$ beforehand. We adopt this optional constraint only for simplifying the following deduction. The Lagrange multipliers of SemiUOT with KL-Divergence is given as:

$$\begin{aligned} \max_{\boldsymbol{s} \geq 0, \boldsymbol{f}, \boldsymbol{g}, \zeta} \min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{SemiUOT}} = \tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{f} + \zeta, \boldsymbol{\pi} \mathbf{1}_N \rangle + \langle \boldsymbol{g} - \zeta, \boldsymbol{b} \rangle + \\ \langle \boldsymbol{C} - \boldsymbol{u} \otimes \mathbf{1}_N^\top - \mathbf{1}_M \otimes \boldsymbol{v}^\top - \boldsymbol{s}, \boldsymbol{\pi} \rangle, \end{aligned} \tag{20}$$

where $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{s}$ and ζ are dual variables. By taking the differentiation on π_{ij} we have:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\mathrm{SemiUOT}}}{\partial \pi_{ij}} &= \left[\tau \log \frac{\sum_{j=1}^{N} \pi_{ij}}{a_i} + f_i + \zeta \right] + \left(C_{ij} - f_i - g_j - s_{ij}\right) \\ &= C_{ij} + \tau \log \frac{\sum_{j=1}^{N} \pi_{ij}}{a_i} + \zeta - g_j - s_{ij} \\ &= 0. \end{aligned} \tag{21}$$

Therefore, we can obtain the results as:

$$\begin{cases} \displaystyle\sum_{j=1}^{N} \pi_{ij} = a_i \exp\left(-\frac{f_i + \zeta}{\tau}\right) \\ \displaystyle\sum_{i=1}^{M} \pi_{ij} = b_j \\ C_{ij} - f_i - g_j - s_{ij} = 0 \end{cases}. \tag{22}$$

After that, we can take these back into KL-Divergence to simplify the calculation:

$$
\begin{aligned}
&\tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right)+\left\langle\boldsymbol{f}+\zeta, \boldsymbol{\pi} \mathbf{1}_N\right\rangle \\
=&\tau \mathrm{KL}\left(\left\langle\boldsymbol{a}, \exp \left(-\frac{\boldsymbol{f}+\zeta}{\tau}\right)\right\rangle \| \boldsymbol{a}\right)+\left\langle\boldsymbol{f}+\zeta, \boldsymbol{a} \exp \left(-\frac{\boldsymbol{f}+\zeta}{\tau}\right)\right\rangle \\
=&\tau \sum_{i=1}^M\left[a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right) \log \frac{a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right)}{a_i}-a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right)+a_i\right]+\sum_{i=1}^M\left(f_i+\zeta\right) a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right) \\
=&\sum_{i=1}^M\left[-\tau a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right)+\tau a_i\right] .
\end{aligned}
\tag{23}
$$

Therefore we can obtain its Fenchel-Lagrange multipliers form of SemiUOT as:

$$
\begin{aligned}
\min_{\boldsymbol{f}, \boldsymbol{g}, \zeta} \mathcal{J}_{\mathrm{SemiUOT}} &=-\tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right)-\left\langle\boldsymbol{f}+\zeta, \boldsymbol{\pi} \mathbf{1}_N\right\rangle-\left\langle\boldsymbol{g}-\zeta, \boldsymbol{\pi}^{\top} \mathbf{1}_M\right\rangle \\
&=\tau \exp \left(-\frac{\zeta}{\tau}\right)\left\langle\boldsymbol{a}, \exp \left(-\frac{\boldsymbol{f}}{\tau}\right)\right\rangle-\langle\boldsymbol{g}-\zeta, \boldsymbol{b}\rangle+\mathcal{O}_{\mathrm{Const}} \\
s.t.\ & f_i+g_j \leq C_{ij},
\end{aligned}
\tag{24}
$$

where $\mathcal{O}_{\mathrm{Const}}=-\sum_{i=1}^M \tau a_i$ and we can neglect it during the following calculation. Once we obtain the optimal solution on $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and ζ^*, we will discover that:

$$
\tau \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right)=\tau \mathrm{KL}\left(\left\langle\boldsymbol{a}, \exp \left(-\frac{\boldsymbol{f}^*+\zeta^*}{\tau}\right)\right\rangle \| \boldsymbol{a}\right)=\text{Const.}
\tag{25}
$$

Hence SemiUOT problem can be transformed into classic optimal transport problem accordingly. Finally we can obtain the optimal solution on ζ by considering $\frac{\partial \mathcal{J}_{\mathrm{SemiUOT}}}{\partial \zeta}=0$ as below:

$$
\zeta=\tau\left[\log \left(\sum_{i=1}^M a_i \exp \left(-\frac{f_i}{\tau}\right)\right)-\log \left(\sum_{j=1}^N b_j\right)\right].
\tag{26}
$$

Once we set $\tau \rightarrow \infty$, the results of the limitation will be shown as:

$$
\lim_{\tau \rightarrow+\infty} a_i \exp \left(-\frac{f_i+\zeta}{\tau}\right)=\lim_{\tau \rightarrow+\infty} a_i \exp \left(-\frac{\zeta}{\tau}\right)=a_i \frac{\left\langle\boldsymbol{b}, \mathbf{1}_N\right\rangle}{\left\langle\boldsymbol{a}, \mathbf{1}_M\right\rangle}=\omega a_i .
\tag{27}
$$

Therefore we conclude the proof of the Proposition 1. □

C Proof of Proposition 2

Proposition 2. (Calculation for Approximate SemiUOT Equation) *Given Approximate SemiUOT equation $\widehat{L}_{\mathrm{P}}$, it can be optimized via Equivalent Transformation Mechanism with Approximation (ETM-Approx). That is, ETM-Approx aims to solve the following equation for each $\widehat{f}_s$:*

$$
\frac{\partial \widehat{L}_{\mathrm{P}}}{\partial \widehat{f}_s}=-a_s \exp \left(-\frac{\widehat{f}_s+\zeta}{\tau}\right)+\exp \left(\frac{\widehat{f}_s}{\epsilon}\right) \sum_{j=1}^N\left[\frac{b_j \exp \left(-\frac{C_{sj}}{\epsilon}\right)}{\sum_{k=1}^M \exp \left(\frac{\widehat{f}_k-C_{kj}}{\epsilon}\right)}\right]=0 .
\tag{28}
$$

Specifically, we can adopt fixed-point iteration method for solving Eq.(28) *at the ℓ-th iteration as follows:*

$$
\begin{cases}
\widehat{f}_1^{\ell+1}=\nu\left[\log \left(a_1 \exp \left(-\frac{\zeta}{\tau}\right)\right)-\log \left[\sum_{j=1}^N\left(\frac{b_j}{\mathscr{W}_{\epsilon, j}(\widehat{\boldsymbol{f}}^{\ell})} \exp \left(-\frac{C_{1j}}{\epsilon}\right)\right)\right]\right] \\
\vdots \\
\widehat{f}_M^{\ell+1}=\nu\left[\log \left(a_M \exp \left(-\frac{\zeta}{\tau}\right)\right)-\log \left[\sum_{j=1}^N\left(\frac{b_j}{\mathscr{W}_{\epsilon, j}(\widehat{\boldsymbol{f}}^{\ell})} \exp \left(-\frac{C_{Mj}}{\epsilon}\right)\right)\right]\right]
\end{cases},
\tag{29}
$$

where $\nu = \tau\epsilon/(\tau+\epsilon)$ for simplification and $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)$ denotes the corresponding calculation:

$$\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell) = \sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k^\ell - C_{kj}}{\epsilon}\right). \tag{30}$$

The proposed procedure can be convergent with a theoretical guarantee. Finally, updating the Lagrange multiplier ζ by further considering $\nabla_\zeta \widehat{L}_{\mathrm{P}} = 0$ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{f}_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$. One can achieve the optimal results on $\widehat{\boldsymbol{f}}^*$ and ζ^* via iterative computing accordingly.

Proof. We first review the proposed Approximate SemiUOT Equation $\widehat{L}_{\mathrm{P}}$ as below:

$$\min_{\widehat{\boldsymbol{f}},\zeta} \widehat{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i e^{-\frac{\widehat{f}_i+\zeta}{\tau}} + \sum_{j=1}^{N} b_j \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{f}_k - C_{kj}}{\epsilon}}\right] + \zeta\right]. \tag{31}$$

Then we consider optimizing $\widehat{f}_s$ as follows:

$$\frac{\partial \widehat{L}_{\mathrm{P}}}{\partial \widehat{f}_s} = 0 \quad \Rightarrow \quad \exp\left(\frac{\tau+\epsilon}{\tau\epsilon}\widehat{f}_s\right) = \frac{a_s e^{-\frac{\zeta}{\tau}}}{\sum\limits_{j=1}^{N}\left(\frac{b_j \exp(-C_{sj}/\epsilon)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}})}\right)}, \tag{32}$$

where $\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}})$ is defined as Eq.(30). At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{f}}$ accordingly:

$$\begin{cases} \widehat{f}_1^{\ell+1} = \nu\left[\log\left(a_1 e^{-\frac{\zeta}{\tau}}\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j e^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right)\right]\right] = \mathcal{F}_1\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \\ \vdots \\ \widehat{f}_s^{\ell+1} = \nu\left[\log\left(a_s e^{-\frac{\zeta}{\tau}}\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j e^{-C_{sj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right)\right]\right] = \mathcal{F}_s\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \\ \vdots \\ \widehat{f}_M^{\ell+1} = \nu\left[\log\left(a_M e^{-\frac{\zeta}{\tau}}\right) - \log\left[\sum_{j=1}^{N}\left(\frac{b_j e^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right)\right]\right] = \mathcal{F}_M\left(\widehat{f}_1^\ell, \cdots, \widehat{f}_M^\ell\right) \end{cases}, \tag{33}$$

where $\nu = \frac{\tau\epsilon}{\tau+\epsilon}$. By taking the gradient on $\mathcal{F}_s(\widehat{f}_s^\ell)$ w.r.t $\widehat{f}_s^\ell$, we can observe that:

$$\begin{aligned} \frac{\partial \mathcal{F}_s(\widehat{f}_s^\ell)}{\partial \widehat{f}_s^\ell} &= -\frac{\tau\epsilon}{\tau+\epsilon} \frac{1}{\sum\limits_{j=1}^{N}\left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right] b_j} \frac{\partial}{\partial \widehat{f}_s^\ell}\left(\sum_{j=1}^{N}\left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right] b_j\right) \\ &= \frac{\tau}{\tau+\epsilon} \underbrace{\frac{1}{\sum\limits_{j=1}^{N}\left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right] b_j} \sum_{j=1}^{N}\left[\frac{b_j \exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)} \cdot \frac{\exp\left(\frac{\widehat{f}_s^\ell - C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^\ell)}\right]}_{<1} \\ &< 1. \end{aligned} \tag{34}$$

Likewise we can obtain the result:

$$\begin{aligned}\mathscr{F}_s\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) &= \left|\frac{\partial\mathcal{F}_s(\widehat{f}_1^{\ell})}{\partial\widehat{f}_1^{\ell}}\right| + \cdots + \left|\frac{\partial\mathcal{F}_s(\widehat{f}_s^{\ell})}{\partial\widehat{f}_s^{\ell}}\right| + \cdots + \left|\frac{\partial\mathcal{F}_s(\widehat{f}_M^{\ell})}{\partial\widehat{f}_M^{\ell}}\right| \\ &= \frac{\tau}{\tau+\epsilon}\frac{1}{\sum_{j=1}^{N}\left[\frac{\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})}\right]b_j}\sum_{j=1}^{N}\left[\frac{b_j\exp\left(-\frac{C_{sj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})}\cdot\sum_{u=1}^{M}\left(\frac{\exp\left(\frac{\widehat{f}_u^{\ell}-C_{uj}}{\epsilon}\right)}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{f}}^{\ell})}\right)\right] \\ &= \frac{\tau}{\tau+\epsilon} < 1.\end{aligned} \tag{35}$$

We can easily conclude that:

$$\begin{cases}\mathscr{F}_1\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) < 1 \\ \vdots \\ \mathscr{F}_s\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) < 1 \\ \vdots \\ \mathscr{F}_M\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) < 1\end{cases}. \tag{36}$$

Therefore, we can conclude that the proposed ETM-Approx method guarantees convergence according to Theorem 2.9 in [61]. □

Remark 1. *ETM-Approx can reach the linear convergence rate via the fixed-point iteration shown as* $\mathcal{O}(NM\log(1/\varepsilon_{\text{err}}))$ *where* $\varepsilon_{\text{err}} = ||\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{f}}^*||_\infty$ *and* $\widehat{\boldsymbol{f}}^*$ *denotes the optimal solution.*

Proof. We can formulate the whole optimization process for Proposition 2 as below :

$$\begin{cases}\widehat{f}_1^{\ell+1} = \mathcal{F}_1\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) \\ \vdots \\ \widehat{f}_s^{\ell+1} = \mathcal{F}_s\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right) \\ \vdots \\ \widehat{f}_M^{\ell+1} = \mathcal{F}_M\left(\widehat{f}_1^{\ell},\cdots,\widehat{f}_M^{\ell}\right)\end{cases} \Rightarrow \left(\widehat{f}_1^{(\ell+1)},\cdots,\widehat{f}_M^{(\ell+1)}\right) = \mathcal{F}_{\text{update}}\left(\widehat{f}_1^{(\ell)},\cdots,\widehat{f}_M^{(\ell)}\right). \tag{37}$$

According to the above discussion, we have the following results:

$$\begin{aligned}\left\|\widehat{\boldsymbol{f}}^{(\ell+1)} - \widehat{\boldsymbol{f}}^*\right\|_\infty &= \left\|\mathcal{F}_{\text{update}}\left(\widehat{\boldsymbol{f}}^{(\ell)}\right) - \mathcal{F}_{\text{update}}\left(\widehat{\boldsymbol{f}}^*\right)\right\|_\infty \\ &\leq \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^{(\ell)} - \widehat{\boldsymbol{f}}^*\right\|_\infty \\ &\leq \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^{(\ell+1)} - \widehat{\boldsymbol{f}}^*\right\|_\infty + \frac{\tau}{\tau+\epsilon}\left\|\widehat{\boldsymbol{f}}^{(\ell+1)} - \widehat{\boldsymbol{f}}^{(\ell)}\right\|_\infty.\end{aligned} \tag{38}$$

Therefore the error between the solution $\widehat{\boldsymbol{f}}^{(\ell+1)}$ at the $(\ell+1)$ iteration and the optimal results $\widehat{\boldsymbol{f}}^*$ is given as:

$$\left\|\widehat{\boldsymbol{f}}^{(\ell+1)} - \widehat{\boldsymbol{f}}^*\right\|_\infty \leq \frac{\tau+\epsilon}{\epsilon}\left(\frac{\tau}{\tau+\epsilon}\right)^{\ell}\left\|\widehat{\boldsymbol{f}}^{(1)} - \widehat{\boldsymbol{f}}^{(0)}\right\|_\infty. \tag{39}$$

Hence ETM-Approx can be linear convergence via the fixed-point iteration shown as $\mathcal{O}(NM\log(1/\varepsilon_{\text{err}}))$ where $\varepsilon_{\text{err}} = ||\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{f}}^*||_\infty$ and $\widehat{\boldsymbol{f}}^*$ denotes the optimal solution. □

D Algorithm for ETM-Based Method on SemiUOT

We also provide the pseudo algorithm of the proposed ETM-Based approachs (e.g., ETM-Exact, ETM-Approx, and ETM-Refine) for solving SemiUOT in Alg.1 to make a clearer illustration.

Algorithm 1 The algorithm of ETM-Based method on SemiUOT

Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; τ, ϵ: Hyper parameters.
Randomly initialize the value of $\boldsymbol{f}^{\text{init}}$.
Choose ETM-Exact, ETM-Approx or ETM-Refine on SemiUOT for optimization.
(1) Function: ETM-Exact on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \boldsymbol{f}^{t=0} = \boldsymbol{f}^{\text{init}}$)
Optimize $\boldsymbol{f}$ via L-BFGS algorithm on L_{P} as:

$$\min_{\boldsymbol{f}} L_{\text{P}} = \tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i + \zeta}{\tau}\right) - \sum_{j=1}^{N} \left[\inf_{k\in[M]} [C_{kj} - f_k] - \zeta\right] b_j,$$

Optimize $\boldsymbol{g}$ via $g_j = \inf_{k\in[M]} (C_{kj} - f_k^t)$.
Optimize ζ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-f_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$ as shown in Eq.(26).
Return: The optimal solutions of $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and ζ^*.
(2) Function: ETM-Approx on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$)
Optimize $\widehat{\boldsymbol{f}}$ via Proposition 2 on $\widehat{L}_{\text{P}}$ as:

$$\min_{\widehat{\boldsymbol{f}}} \widehat{L}_{\text{P}} = \tau \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{f}_i + \zeta}{\tau}\right) + \sum_{j=1}^{N} b_j \left[\epsilon \log\left[\sum_{k=1}^{M} \exp\left(\frac{\widehat{f}_k - C_{kj}}{\epsilon}\right)\right] + \zeta\right]$$

Optimize $\widehat{\boldsymbol{g}}$ via $\widehat{g}_j = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{f}_k - C_{kj})/\epsilon)]$.
Optimize ζ via $\zeta = \tau[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{f}_i/\tau)) - \log(\sum_{j=1}^{N} b_j)]$ as shown in Eq.(26).
Return: The optimal solutions of $\widehat{\boldsymbol{f}}^*$, $\widehat{\boldsymbol{g}}^*$ and ζ^*.
(3) Function: ETM-Refine on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$)
Obtain $\widehat{\boldsymbol{f}}^* =$ ETM-Approx on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \widehat{\boldsymbol{f}}^{t=0} = \boldsymbol{f}^{\text{init}}$).
Obtain $\boldsymbol{f}^* =$ ETM-Exact on SemiUOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau, \boldsymbol{f}^{t=0} = \widehat{\boldsymbol{f}}^*$).
Return: The optimal solutions of $\boldsymbol{f}^*$, $\boldsymbol{g}^*$ and ζ^*.

E Proof of Proposition 3

Proposition 3. (Principles of Equivalent Transformation Mechanism for UOT) *Given UOT with KL-Divergence J_{UOT}, its Fenchel-Lagrange multipliers form is given:*

$$\begin{aligned} &\min_{\boldsymbol{u},\boldsymbol{v},\zeta} \left[\tau_a \sum_{i=1}^{M} a_i e^{-\frac{u_i+\zeta}{\tau_a}} + \tau_b \sum_{j=1}^{N} b_j e^{-\frac{v_j-\zeta}{\tau_b}}\right] \\ &s.t.\ u_i + v_j + s_{ij} = C_{ij}, \quad s_{ij} \geq 0, \end{aligned} \tag{40}$$

where $\mathbf{u}$, $\mathbf{v}$, $\boldsymbol{s}$ and ζ denote Lagrange multipliers. Moreover, UOT problem can also be transformed into classic optimal transport as follows:

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{U}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle \\ &s.t. \begin{cases} \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{u}^* + \zeta^*}{\tau_a}\right) = \boldsymbol{\alpha} \\ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\dfrac{\boldsymbol{v}^* - \zeta^*}{\tau_b}\right) = \boldsymbol{\beta} \end{cases}. \end{aligned} \tag{41}$$

Note that when $\tau_a, \tau_b \to \infty$, the source and target marginal probabilities can be determined as $\boldsymbol{\pi}\mathbf{1}_N = \sqrt{\omega}\boldsymbol{a}$ and $\boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}/\sqrt{\omega}$ where $\omega = \langle \boldsymbol{b}, \mathbf{1}_N \rangle / \langle \boldsymbol{a}, \mathbf{1}_M \rangle$ respectively.

Proof. To start with, we first rewrite the optimization problem as below:

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) \\ &s.t.\ (\text{Optional}):\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}. \end{aligned} \tag{42}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the marginal probabilities for source and target domains respectively. Note that we do not need the true value fo $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ beforehand. That is, the constraints here are optional for the following UOT deduction. The Lagrange multipliers of UOT with KL-Divergence is given as:

$$\max_{\boldsymbol{s} \geq 0, \boldsymbol{u}, \boldsymbol{v}, \zeta} \min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{UOT}} = \tau_a \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi} \mathbf{1}_N \rangle + \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top \mathbf{1}_M \rangle + \mathscr{C}_{\mathrm{UOT}}, \tag{43}$$

where $\mathscr{C}_{\mathrm{UOT}} = \sum_{i,j} (C_{ij} - u_i - v_j - s_{ij}) \pi_{ij} = \langle \boldsymbol{C} - \boldsymbol{u} \otimes \mathbf{1}_N^\top - \mathbf{1}_M \otimes \boldsymbol{v}^\top - \boldsymbol{s}, \boldsymbol{\pi} \rangle$ and $\boldsymbol{u}, \boldsymbol{v}$ and ζ are dual variables. By taking the differentiation on π_{ij} we have:

$$\begin{aligned} \frac{\partial \mathcal{J}_{\mathrm{UOT}}}{\partial \pi_{ij}} &= \left[\tau_a \log \frac{\sum_{j=1}^{N} \pi_{ij}}{a_i} + u_i + \zeta \right] + \left[\tau_b \log \frac{\sum_{i=1}^{M} \pi_{ij}}{b_j} + v_j - \zeta \right] + (C_{ij} - u_i - v_j - s_{ij}) \\ &= C_{ij} + \tau_a \log \frac{\sum_{j=1}^{N} \pi_{ij}}{a_i} + \tau_b \log \frac{\sum_{i=1}^{M} \pi_{ij}}{b_j} - s_{ij} = 0. \end{aligned} \tag{44}$$

Then we can obtain the results:

$$\begin{cases} \sum_{j=1}^{N} \pi_{ij} = a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) \\ \sum_{i=1}^{M} \pi_{ij} = b_j \exp\left(-\frac{v_j - \zeta}{\tau_b}\right) \\ C_{ij} - u_i - v_j - s_{ij} = 0 \end{cases}. \tag{45}$$

By taking the above results into KL-Divergence, we can further simplify the results:

$$\begin{cases} \tau_a \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi} \mathbf{1}_N \rangle = \sum_{i=1}^{M} \left[-\tau_a a_i \exp\left(-\frac{f_i + \zeta}{\tau_a}\right) + \tau_a a_i \right] \\ \tau_b \mathrm{KL}\left(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}\right) + \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top \mathbf{1}_M \rangle = \sum_{j=1}^{N} \left[-\tau_b b_j \exp\left(-\frac{g_j - \zeta}{\tau_b}\right) + \tau_b b_j \right] \end{cases}. \tag{46}$$

Therefore we can obtain its Fenchel-Lagrange multipliers form of UOT as:

$$\begin{aligned} \min_{\boldsymbol{u}, \boldsymbol{v}, \zeta} \mathcal{J}_{\mathrm{UOT}} &= -\tau_a \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right) - \langle \boldsymbol{u} + \zeta, \boldsymbol{\pi} \mathbf{1}_N \rangle - \tau_b \mathrm{KL}(\boldsymbol{\pi}^\top \mathbf{1}_M \| \boldsymbol{b}) - \langle \boldsymbol{v} - \zeta, \boldsymbol{\pi}^\top \mathbf{1}_M \rangle \\ &= \tau_a \exp\left(-\frac{\zeta}{\tau_a}\right) \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right) \right\rangle + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right) \right\rangle + \mathcal{O}_{\mathrm{Const}} \\ s.t.\ & u_i + v_j \leq C_{ij}, \end{aligned} \tag{47}$$

where $\mathcal{O}_{\mathrm{Const}} = -\sum_{i=1}^{M} \tau_a a_i - \sum_{j=1}^{N} \tau_b b_j$, and we can neglect it during the following calculation. Once we obtain the optimal solution on $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^*, the KL-Divergence will turn out to be constants and therefore the original optimization problem can be transformed into classic optimal transport. Finally we can obtain the optimal solution on ζ by considering $\frac{\partial \mathcal{J}_{\mathrm{UOT}}}{\partial \zeta} = 0$ as below:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b} \left[\log \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right) \right\rangle - \log \left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right) \right\rangle \right]. \tag{48}$$

Once we set $\tau_a \to \infty$ and $\tau_b \to \infty$, the results of the limitation will be shown as:

$$\begin{aligned}
&\lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) = \lim_{\tau_a \to +\infty, \tau_b \to +\infty} a_i \exp\left(-\frac{\zeta}{\tau_a}\right) = a_i \sqrt{\frac{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}} = \sqrt{\omega} a_i, \\
&\lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{v_j - \zeta}{\tau_b}\right) = \lim_{\tau_a \to +\infty, \tau_b \to +\infty} b_j \exp\left(-\frac{\zeta}{\tau_b}\right) = b_j \sqrt{\frac{\langle \boldsymbol{a}, \mathbf{1}_M \rangle}{\langle \boldsymbol{b}, \mathbf{1}_N \rangle}} = \frac{1}{\sqrt{\omega}} b_j.
\end{aligned} \tag{49}$$

Therefore we conclude the proof of the Proposition 3. □

F Illustrations of Optimization 1

Optimization 1. (Calculation of ETM-Approx approach for UOT) To start with, we first review the Exact UOT Equation is defined as:

$$\begin{aligned}
\min_{\boldsymbol{u}, \zeta} L_{\mathrm{U}} &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(-\frac{v_j}{\tau_b}\right) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k \in [M]} (u_k - C_{kj})}{\tau_b}\right),
\end{aligned} \tag{50}$$

where $v_j = -\sup_{k \in [M]} (u_k - C_{kj})$ meanwhile the marginal probabilities are set as $\boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-(\boldsymbol{u} + \zeta)/\tau_a\right) = \boldsymbol{\alpha}$ and $\boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-(\boldsymbol{v} - \zeta)/\tau_b\right) = \boldsymbol{\beta}$. Since the optimization problem in Eq.(9) is convex, we can also utilize block gradient descent to optimize the problem. Specifically, we first fix v^l and optimize variable u^l at the l-th iteration by replacing the original marginal probability $\boldsymbol{b}$ in Eq.(6) with $\boldsymbol{\beta}$ accordingly to transform UOT into SemiUOT problem:

$$\begin{aligned}
&\min_{\boldsymbol{\pi} \geq 0} J_{\mathrm{U}}^u = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \mathrm{KL}\left(\boldsymbol{\pi} \mathbf{1}_N \| \boldsymbol{a}\right), \\
&s.t. \begin{cases} (\text{Constraint}): \ \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b} \odot \exp\left(-\dfrac{\boldsymbol{v} - \zeta}{\tau_b}\right) = \boldsymbol{\beta} \\ (\text{Optional}): \ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\dfrac{\boldsymbol{u} + \zeta}{\tau_a}\right) = \boldsymbol{\alpha} \end{cases}.
\end{aligned} \tag{51}$$

At that time, the Fenchel-Lagrange multipliers form of Eq.(51) is given via the Proposition 1:

$$\begin{aligned}
\min_{\boldsymbol{u}} L_{\mathrm{U}}^u &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widetilde{u}_i + \zeta}{\tau_a}\right) - \sum_{j=1}^{N} \beta_j (\widetilde{v}_j - \zeta) \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) - \sum_{j=1}^{N} \left(\inf_{k \in [M]} [C_{kj} - u_k] - \zeta\right) \beta_j.
\end{aligned} \tag{52}$$

Note that $\widetilde{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{v}}$ denote the Lagrange multiplier for Eq.(51) while we have $\widetilde{v}_j = \inf_{k \in [M]} [C_{kj} - u_k] = v_j$ and $\widetilde{\boldsymbol{u}} = \boldsymbol{u}$. To further accelerate the optimization process, we consider to make a smooth approximation on replacing $\inf(\cdot)$ as $\inf_{k \in [M]} [C_{kj} - u_k] \approx -\epsilon \log[\sum_{k=1}^{M} e^{\frac{u_k - C_{kj}}{\epsilon}}] = \widehat{v}_j$. Therefore, we first fix $\widehat{v}^l$ and optimize variable $\widehat{u}^l$ at the l-th iteration to solve the following equation on $\widehat{L}_{\mathrm{U}}^u$ accordingly:

$$\begin{aligned}
\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\mathrm{U}}^u &= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} \beta_j \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right] \\
&= \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} b_j \exp\left(-\frac{\widehat{v}_j - \zeta}{\tau_b}\right) \left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right].
\end{aligned} \tag{53}$$

The optimization objective shares a similar formulation as Eq.31. At that time we adopt fixed-point iteration method to optimize $\widehat{\boldsymbol{u}}$ accordingly based on the Proposition 2:

$$\begin{cases} \widehat{u}_1^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon} \left[\log\left(a_1 e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N} \left(\frac{\beta_j e^{-C_{1j}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_1\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right) \\ \vdots \\ \widehat{u}_s^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon} \left[\log\left(a_s e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N} \left(\frac{\beta_j e^{-C_{sj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_s\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right) \\ \vdots \\ \widehat{u}_M^{\ell+1} = \dfrac{\tau_a \epsilon}{\tau_a + \epsilon} \left[\log\left(a_M e^{-\frac{\zeta}{\tau_a}}\right) - \log\left[\displaystyle\sum_{j=1}^{N} \left(\frac{\beta_j e^{-C_{Mj}/\epsilon}}{\mathscr{W}_{\epsilon,j}(\widehat{\boldsymbol{u}}^\ell)}\right)\right]\right] = \mathcal{U}_M\left(\widehat{u}_1^\ell, \cdots, \widehat{u}_M^\ell\right) \end{cases} \quad . \tag{54}$$

The iteration process can be shown to converge efficiently based on Proposition 2. After that we fix $\widehat{u}$ and optimize variable $\widehat{v}$ via $\widehat{v}_j = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{u}_k - C_{kj})/\epsilon)]$. We can achieve the optimal solution on $\widehat{\boldsymbol{u}}^*$ and $\widehat{\boldsymbol{v}}^*$ via iteratively computing via the above procedure accordingly. Finally, we update ζ via $\zeta = (\tau_a \tau_b/(\tau_a + \tau_b))[\log(\sum_{i=1}^{M} a_i \exp(-\widehat{u}_i^*/\tau_a)) - \log(\sum_{j=1}^{N} b_j \exp(-\widehat{v}_j^*/\tau_b))]$.

G Algorithm for ETM-Based Method on UOT

We also provide the pseudo algorithm of the proposed ETM-Based approachs (e.g., ETM-Exact, ETM-Approx and ETM-Refine) for solving UOT in Alg.2 to make a more clear illustration.

H Proof of Proposition 4

Proposition 4. (The Definition and Usage of KKT-Multiplier Regularization) *Given any OT with multiplier $\boldsymbol{s}$, one can obtain accurate solution $\boldsymbol{\pi}^*$ via proposed KKT-multiplier regularization term $\mathcal{G}(\boldsymbol{\pi}, \boldsymbol{s}) = \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle$, which formulates Multiplier Regularized Optimal Transport (MROT):*

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq 0} \mathcal{J}_{\mathrm{G}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle + \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi}) \\ &s.t.\ \boldsymbol{\pi} \mathbf{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{\beta}, \end{aligned} \tag{55}$$

where $\mathcal{L}_{\mathrm{Reg}}(\boldsymbol{\pi})$ denotes the regularization term on $\boldsymbol{\pi}$. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ denote the final marginal probabilities obtained by ETM-based method, while η_{Reg}, η_G denotes the hyper parameter. Ideally, η_G should be set as a relatively large number. Meanwhile the dual form of MROT is given as:

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} L_{\mathrm{G}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\mathrm{Reg}} \mathcal{L}_{\mathrm{Reg}}^* \left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\mathrm{Reg}}} \right), \tag{56}$$

where $\widetilde{C}_{ij} = C_{ij} + \eta_G s_{ij}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ denote the Lagrange multipliers for MROT. $\mathcal{L}_{\mathrm{Reg}}^*(\cdot)$ denotes the conjugate function of $\mathcal{L}_{\mathrm{Reg}}(\cdot)$ and one can figure out the matching results of $\boldsymbol{\pi}$ via solving the following equation $\nabla_{\pi_{ij}} \mathcal{L}_{\mathrm{Reg}}(\pi_{ij}) = (\psi_i + \phi_j - \widetilde{C}_{ij})/\eta_{\mathrm{Reg}}$.

Algorithm 2 The algorithm of ETM-Based method on UOT

Input: $\boldsymbol{C}$: cost matrix; $\boldsymbol{a}, \boldsymbol{b}$: initial marginal probability; τ_a, τ_b, ϵ: Hyper parameters.

Randomly initialize the value of $\boldsymbol{u}^{\text{init}}$.

Choose ETM-Exact, ETM-Approx or ETM-Refine on UOT for optimization.

(1) Function: ETM-Exact on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \boldsymbol{u}^{t=0} = \boldsymbol{u}^{\text{init}}$)

Optimize $\boldsymbol{u}$ L-BFGS algorithm to optimize L_{U} as:

$$\min_{\boldsymbol{u}} L_{\text{U}} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i + \zeta}{\tau_a}\right) + \tau_b \exp\left(\frac{\zeta}{\tau_b}\right) \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k\in[M]}(u_k - C_{kj})}{\tau_b}\right)$$

Optimize $\boldsymbol{v}$ via $v_j = \inf_{k\in[M]} (C_{kj} - u_k)$.

Optimize ζ via $\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\right\rangle - \log\left\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\right\rangle\right]$ as shown in Eq.(48).

Return: The optimal solutions of $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^*.

(2) Function: ETM-Approx on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$)

Randomly initialize the value of $\widehat{\boldsymbol{v}}^{t'=1}$.

for $t' = 1$ to T' **do**

Optimize $\widehat{\boldsymbol{u}}^{t'}$ via Proposition 2 to optimize $\widehat{L}_{\text{U}}^{u}$ as:

$$\min_{\widehat{\boldsymbol{u}}} \widehat{L}_{\text{U}}^{u} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{\widehat{u}_i + \zeta}{\tau_a}\right) + \sum_{j=1}^{N} b_j \exp\left(-\frac{\widehat{v}_j - \zeta}{\tau_b}\right)\left[\epsilon \log\left[\sum_{k=1}^{M} e^{\frac{\widehat{u}_k - C_{kj}}{\epsilon}}\right] + \zeta\right]$$

Optimize $\widehat{\boldsymbol{v}}^{t'}$ via $\widehat{v}_j^{t'} = -\epsilon \log[\sum_{k=1}^{M} \exp((\widehat{u}_k^{t'} - C_{kj})/\epsilon)]$.

end for

Optimize ζ via $\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\left\langle \boldsymbol{a}, \exp\left(-\frac{\widehat{\boldsymbol{u}}}{\tau_a}\right)\right\rangle - \log\left\langle \boldsymbol{b}, \exp\left(-\frac{\widehat{\boldsymbol{v}}}{\tau_b}\right)\right\rangle\right]$ as shown in Eq.(48).

Return: The optimal solutions of $\widehat{\boldsymbol{u}}^*$, $\widehat{\boldsymbol{v}}^*$ and ζ^*.

(3) Function: ETM-Refine on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$)

Obtain $\widehat{\boldsymbol{u}}^* =$ ETM-Approx on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \widehat{\boldsymbol{u}}^{t=0} = \boldsymbol{u}^{\text{init}}$).

Obtain $\boldsymbol{u}^* =$ ETM-Exact on UOT($\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \tau_a, \tau_b, \boldsymbol{u}^{t=0} = \widehat{\boldsymbol{u}}^*$).

Return: The optimal solutions of $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and ζ^*.

Proof. We first provide the Lagrange multiplier of MROT as:

$$\begin{aligned}
\max_{\boldsymbol{\psi},\boldsymbol{\phi}} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{MROT}} &= \langle \boldsymbol{C}, \boldsymbol{\pi}\rangle + \eta_G \langle \boldsymbol{\pi}, \boldsymbol{s}\rangle + \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}(\boldsymbol{\pi}) - \langle \boldsymbol{\psi}, \boldsymbol{\pi}\mathbf{1}_N - \boldsymbol{\alpha}\rangle - \langle \boldsymbol{\phi}, \boldsymbol{\pi}^\top \mathbf{1}_M - \boldsymbol{\beta}\rangle \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle + \eta_{\text{Reg}} \inf_{\boldsymbol{\pi}} \left[\sum_{i,j} \left[\frac{C_{ij} + \eta_G s_{ij} - \psi_i - \phi_j}{\eta_{\text{Reg}}} \pi_{ij} + \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\text{Reg}} \sup_{\boldsymbol{\pi}} \left[\sum_{i,j} \left[\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \pi_{ij} - \mathcal{L}_{\text{Reg}}(\pi_{ij})\right]\right] \\
&= \langle \boldsymbol{\alpha}, \boldsymbol{\psi}\rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi}\rangle - \eta_{\text{Reg}} \mathcal{L}_{\text{Reg}}^* \left(\frac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right).
\end{aligned} \tag{57}$$

At that time we have the following results:

$$\begin{cases} \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \psi_i} = 0 \\ \dfrac{\partial \mathcal{J}_{\text{MROT}}}{\partial \phi_j} = 0 \end{cases} \Rightarrow \begin{cases} \nabla_{\psi_i} \mathcal{L}_{\text{Reg}}^* \left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \alpha_i \\ \nabla_{\phi_j} \mathcal{L}_{\text{Reg}}^* \left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}}\right) = \beta_j \end{cases}. \tag{58}$$

By taking the differentiation on π_{ij} we have:

$$\frac{\partial \mathcal{J}_{\text{MROT}}}{\partial \pi_{ij}} = \widetilde{C}_{ij} + \eta_{\text{Reg}} \nabla_{\pi_{ij}} \mathcal{L}_{\text{Reg}}(\pi_{ij}) - \psi_i - \phi_j = 0. \tag{59}$$

Table 4: Classification accuracy (%) on Digits (Source: LeNet) and VisDA dataset (Source:ResNet50) for UDA (unsupervised domain adaptation) task

Method	S→M	M→U	U→M	Avg	VisDA
Source	68.3±0.3	65.3±0.5	66.2±0.2	66.6	52.4
DeepJDOT [25]	95.4±0.1	95.6±0.4	96.4±0.3	95.8	68.0
JUMBOT [30]	98.9±0.1	96.7±0.5	98.2±0.1	97.9	72.5
JUMBOT + UOT(ETM-Refine + MROT-Ent)	99.4±0.1	98.7±0.3	99.2±0.1	99.1	73.6
JUMBOT + UOT(ETM-Refine + MROT-Norm)	99.7±0.1	99.3±0.2	99.6±0.1	**99.5**	**74.2**

For instance, when $\mathcal{L}_{\text{Reg}}(\boldsymbol{\pi}) = -\langle \boldsymbol{\pi}, \log(\boldsymbol{\pi}) - 1 \rangle$ denotes as the entropy regularization term, the dual form of MROT-Ent is shown as:

$$\begin{cases} \max\limits_{\boldsymbol{\psi}, \boldsymbol{\phi}} \mathcal{J}_{\text{MROT}-\text{Ent}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \eta_{\text{Reg}} \sum\limits_{i,j} \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \right) \\ \pi_{ij} = \exp\left(\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \right) \end{cases}. \tag{60}$$

When $\mathcal{L}_{\text{Reg}}(\boldsymbol{\pi}) = \langle \boldsymbol{\pi}, \boldsymbol{\pi} \rangle / 2$ denotes the square-norm regularization term, the dual form of MROT-Norm is shown as:

$$\begin{cases} \max\limits_{\boldsymbol{\psi}, \boldsymbol{\phi}} \mathcal{J}_{\text{MROT}-\text{Norm}} = \langle \boldsymbol{\alpha}, \boldsymbol{\psi} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{\phi} \rangle - \dfrac{\eta_{\text{Reg}}}{2} \sum\limits_{i,j} \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \right]_+^2 \\ \pi_{ij} = \left[\dfrac{\psi_i + \phi_j - \widetilde{C}_{ij}}{\eta_{\text{Reg}}} \right]_+ \end{cases}. \tag{61}$$

Therefore we conclude the proof of Proposition 4. □

In summary, the time complexity of the proposed ETM-Approx+MROT-Ent or ETM-Approx+MROT-Norm method is provided as $\mathcal{O}\left(NM \log\left(1/\varepsilon_{\text{err}}\right) + NMd_\pi\right)$. Meanwhile, the time complexity of the proposed ETM-Refine+MROT-Ent or ETM-Refine+MROT-Norm method is provided as $\mathcal{O}\left(NM \log\left(1/\varepsilon_{\text{err}}\right) + NMd_T + NMd_\pi\right)$.

I Experiments on Domain Adaptations

Datasets. We conduct the unsupervised domain adaptation tasks on *Digits*, *Office-Home*, and *VisDA*. *Digits* is the classic dataset for digit classification which contains three standard digit classification datasets: **MNIST** [48], **USPS**[40] and **SVHN** [67]. Each dataset consists of 10 classes of digits, ranging from 0 to 9. *Office-Home* [94] is a standard benchmark dataset which includes 15,500 images in 65 object classes in office and home settings, forming four dissimilar domains: Artistic images (**Ar**), Clip Art (**Cl**), Product images (**Pr**), and Real-World (**Rw**). *VisDA* [72] is a large-scale computer vision dataset on two domains, i.e., **Synthetic** and **Real** with 280K images in 12 classes.

Performance. We also conduct the UDA domain adaptation tasks on Digits and VisDA and the results are shown in Table.4. We can observe that the proposed ETM-Refine with MROT-Norm on SemiUOT achieves state-of-the-art performance on Digits and VisDA.

J Experiments on Partial Domain Adaptations

Datasets. We further conduct the domain adaptation tasks on new datasets, i.e., *Office-31* [79] and *ImageCLEF* [13]. **Office-31** is the commonly-used computer vision dataset for domain adaptation with 4,652 images from three different domains: *Amazon* (**A**), *Webcam* (**W**) and *DSLR* (**D**). Target domain has the first 10 classes (alphabetical order) following [11]. **ImageCLEF** contains 3 domains with 12 classes, i.e., *Caltech* (**C**), *ImageNet* (**I**) and *Pascal* (**P**). Target domain has the first 6 classes (alphabetical order) following [57].

Baselines. We involve **DeepJDOT** [25], **ROT** [5], **JUMBOT** [30], **ETN** [12], **AR** [38], **m-POT** [68], **MOT** [56], as the model baselines for the domain adaptation task. (1) **DeepJDOT** [25] first

Table 5: H-score (%) on *Office-Home* for universal unsupervised domain adaptation

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet [39]	44.65	48.04	50.13	46.64	46.91	48.96	47.47	43.17	50.23	48.45	44.76	48.43	47.32
OSBP [80]	39.59	45.09	46.17	45.70	45.24	46.75	45.26	40.54	45.75	45.08	41.64	46.90	44.48
UAN [100]	51.64	51.70	54.30	61.74	57.63	61.86	50.38	47.62	61.46	62.87	52.61	65.19	56.58
CMU [34]	56.02	56.93	59.15	66.95	64.27	67.82	54.72	51.09	66.39	68.24	57.89	69.73	61.60
DCC [49]	57.97	54.05	58.01	74.64	70.62	77.52	64.34	73.60	74.94	80.96	**75.12**	80.38	70.18
TNT [18]	61.90	74.60	80.20	73.50	71.40	79.60	74.20	**69.50**	82.70	77.30	70.10	81.20	74.70
UniOT [16]	67.27	80.54	86.03	73.51	77.33	84.28	75.54	63.33	85.99	77.77	65.37	81.92	76.57
UniOT + UOT(ETM-Refine + MROT-Ent)	68.63	81.72	87.94	75.88	79.03	86.21	77.29	68.77	87.14	78.59	73.62	82.83	78.97
UniOT + UOT(ETM-Refine + MROT-Norm)	**69.02**	**81.95**	**88.36**	**76.12**	**79.36**	**86.49**	**77.03**	69.25	**87.30**	**78.93**	74.18	**82.96**	**79.25**

Table 6: H-Score (%) on *Office-31* for universal unsupervised domain adaptation

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg
ResNet [39]	49.78	47.92	48.48	54.94	48.96	55.60	50.94
OSBP [80]	51.14	50.23	49.75	55.53	50.16	57.20	52.34
UAN [100]	59.68	58.61	60.11	70.62	60.34	71.42	63.46
CMU [34]	68.11	67.33	71.42	79.32	72.23	80.42	73.14
DCC [49]	88.50	78.54	70.18	79.29	75.87	88.58	80.16
TNT [18]	85.70	80.40	83.80	92.00	79.10	91.20	85.37
UniOT [16]	86.97	88.48	88.35	98.83	87.60	96.57	91.13
UniOT + UOT(ETM-Refine + MROT-Ent)	88.25	89.62	89.47	99.48	89.10	97.94	92.31
UniOT + UOT(ETM-Refine + MROT-Norm)	**88.67**	**90.14**	**90.03**	**99.58**	**89.42**	**98.46**	**92.72**

adopts optimal transport into solving domain adaptation problem with deep learning framework. (2) **ROT** [5] adopts robust optimal transport into adversarial training for domain adaptation. (3) **JUMBOT** [30] adopts mini-batch unbalanced optimal transport method for domain adaptation. (4) **ETN** [12] utilizes example transfer network to jointly learn domain-invariant representations and the progressive weighting scheme. (5) **AR** [38] adopts adversarial reweighting strategy on source domain data for alignment. (6) **m-POT** [68] adopts partial optimal transport method in the mini-batch settings for domain adaptation. (7) **MOT** [56] adopts masked unbalanced optimal transport technique on considering label information for PDA tasks.

K Experiments on Universal Domain Adaptations

We further conduct the experiments on universal domain adaptations. That is, there are shared labels between the source and target domains. Additionally, there are private labels specific to each domain [29, 103]. We conduct the universal domain adaptations on both *Office-31* and *Office-Home*. Specifically, we set the first 10 classes in alphabetical order as the common label set, the next 10 classes as source private label and the rest 11 classes as target private label for Office-31. Likewise, we set the first 10 classes in alphabetical order as the common label set, the next 5 classes as source private label and the rest 55 classes as target private label for Office-Home. We involve the following models as baselines: (1) **OSBP** [80] adopts domain adversarial learning for open-set domain adaptation, (2) **UAN** [100] utilizes transferability criterion for universal domain adaptation, (3) **CMU** [34] learns to detect open classes with uncertainty estimation, (4) **DCC** [49] adopts domain consensus clustering for adaptation, (5) **TNT** [18] adopts evidential neighborhood contrastive learning for adaptation, (6) **UniOT** [16] adopts unbalanced optimal transport with adaptive filtering for transferring.

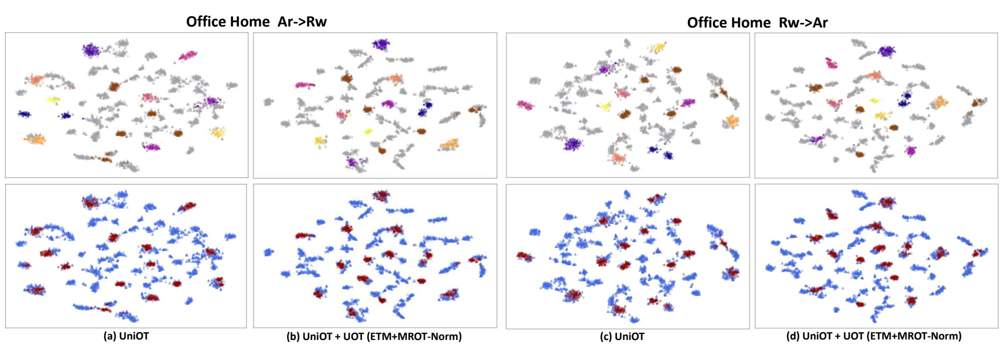


Figure 5: The T-SNE of data features on Ar → Rw (Office-Home) and Rw → Ar (Office-Home). The first row shows the original data sample distribution: The brown and gray colors denote the source and target private classes, respectively. The rest are common label set with different colors. The second row indicates the mapping between source and target domains: The red and blue points denote the source and target samples, respectively.

Table 7: Experimental results on Treatment Effect Estimation tasks.

	ACIC (PEHE)		ACIC (AUUC)		IHDP (PEHE)		IHDP (AUUC)	
	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample	In-Sample	Out-Sample
OLS [2]	3.749	4.340	0.843	0.496	3.856	5.674	0.652	0.492
TARNet [88]	3.236	3.254	**0.886**	0.662	0.749	1.788	0.654	0.711
PSM [78]	5.228	5.094	0.884	0.745	3.219	4.634	0.740	0.681
CFR-WASS [88]	3.128	3.207	0.873	0.669	0.657	1.704	0.656	0.715
ESCFR [96]	2.252	2.316	0.796	0.754	0.502	1.282	0.665	0.719
ESCFR + UOT(ETM-Refine + MROT-Ent)	2.327	2.261	0.839	0.814	0.497	1.275	0.769	0.763
ESCFR + UOT(ETM-Refine + MROT-Norm)	**2.104**	**2.216**	0.883	**0.839**	**0.475**	**1.146**	**0.798**	**0.802**

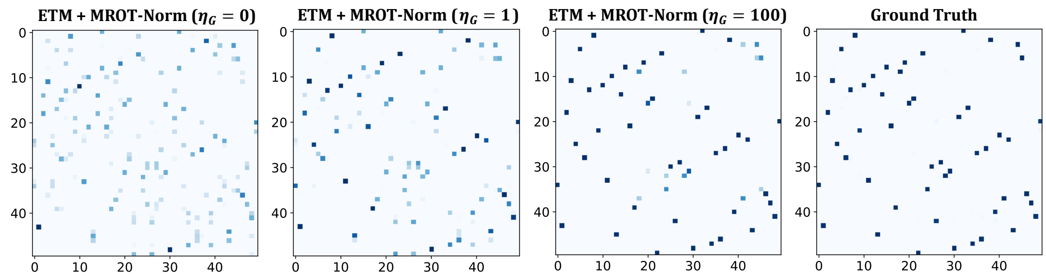


Figure 6: The matching results on ETM + MROT-Norm on SemiUOT with different values of $\eta_G = \{0, 1, 100\}$.

We adopt the same experimental settings as UniOT [16]. We utilize the commonly-used H-score [34] to validate the final results as shown in Table 5-6. Note that UniOT + UOT(ETM + MROT) only replaces the entropic UOT in UniOT with our proposed ETM-Refine method with MROT. From that, we can observe that UniOT + UOT(ETM-Refine + MROT-Norm) reaches the best performance, indicating that UOT with ETM + MROT can provide more accurate matching results. Moreover, we adopt T-SNE method [93] to plot the source and target data features in the latent space as shown in Fig.5(a)-(b). We can find that: (1) original UniOT could lead to rather scattered features in the latent space. That is because UniOT with entropic UOT could lead to dense and inaccurate matching solutions, which limits the model potentials. (2) Our proposed UniOT + UOT(ETM + MROT-Norm) can provide more compact features since it provides more accurate solutions. Thus it further illustrates the efficacy of our proposed ETM-Refine with MROT-Norm method.

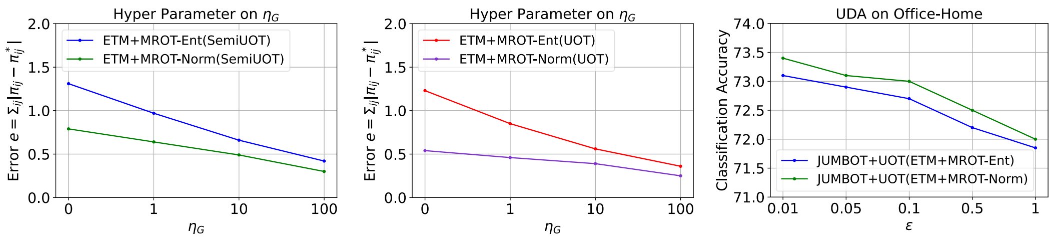
L Experiments on Treatment Effect Estimation

Datasets for Treatment Effect Estimation. We further conduct ETM-Refine on treatment effect estimation with two semi-synthetic datasets IHDP [88] and ACIC [99]. IHDP is set to estimate the effect of specialist home visits on infants' potential cognitive scores and it contains 747 observations and 25 covariates. ACIC includes 4802 observations and 58 covariates, which comes from the collaborative perinatal project.

Results. We involve the following models as baselines: (1) **OLS** [2] utilizes least square regression with treatment as covariates, (2) **TARNet** [88] adopts integral orobability metrics for adaptation, (3) **PSM** [78] adopts propensity score for causal effects, (4) **CFR-WASS** [78] utilizes standard optimal transport for adaptation, (5) **ESCFR** [96] further utilizes unbalanced optimal transport for adaptation. We adopt the same experimental settings as ESCFR [96]. We utilize Precision in Estimation of Heterogeneous Effect (PEHE) [88] and Area Under the Uplift Curve (AUUC) [7] for the evaluation. Note that ESCFR + UOT(ETM-Refine + MROT-Ent) only replaces the entropic UOT in ESCFR with our proposed approximate-to-exact ETM-Refine + MROT-Norm. The experimental results are shown in Table 7. From that, we can observe that ESCFR + UOT(ETM-Refine + MROT-Norm) achieves the best performance, indicating the efficacy of our proposed ETM-Refine method.

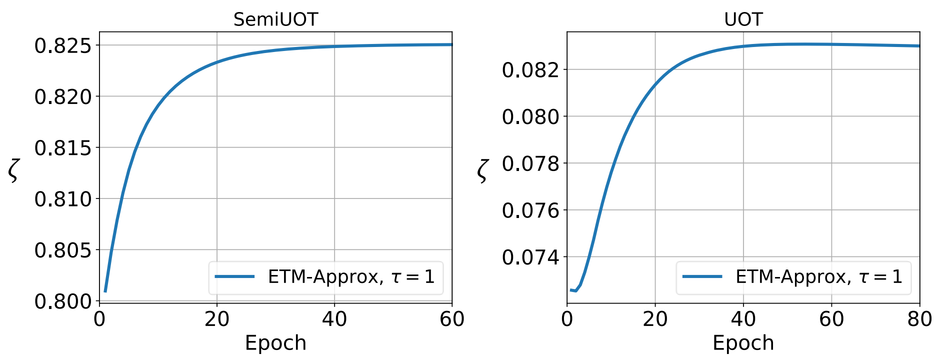
M More Experimental Results

Parameter sensitivity. We tune η_{G} on SemiUOT via ETM-Refine with MROT-Norm in range of $\eta_{\mathrm{G}} \in \{0, 1, 100\}$ using the same data samples shown in Fig.1 and show the results in Fig.6. We can observe that when η_{G} is smaller (e.g., $\eta_{\mathrm{G}} = 0$ or $\eta_{\mathrm{G}} = 1$), the proposed KKT-multiplier regularization term $\mathcal{G}(\boldsymbol{\pi}, \boldsymbol{s}) = \langle \boldsymbol{\pi}, \boldsymbol{s} \rangle$ may struggle to play a significant role during the optimization process. Meanwhile when $\eta_{\mathrm{G}} = 100$, ETM-Refine with MROT-Norm can achieve more accurate



(a) η_G on SemiUOT (Synthetic Dataset) **(b) η_G on UOT (Synthetic Dataset)** **(c) ϵ on UOT (UDA on Office-Home)**

Figure 7: The hyper parameters on η_G and ϵ.



(a) The result of ζ on SemiUOT **(b) The result of ζ on UOT**

Figure 8: The results of ζ and L_{P} on UOT and SemiUOT.

matching results. We can conclude that choosing a larger value of η_{G} can fully utilize the knowledge provided by KKT multiplier and enhance the final results. Moreover, we conduct the experiments for the absolute error when $\tau = 1$ with $N = 500$ synthetic data samples on both SemiUOT and UOT and report the results in Fig.7(a)-(b). Larger value on η_G can provide more useful KKT-multiplier information and boost the model performance and therefore we set $\eta_{\mathrm{G}} = 100$ empirically. Furthermore, we conduct the hyper parameter experiments by varying $\epsilon = \{0.01, 0.05, 0.1, 0.5, 1\}$ on UDA task in Office-Home and report the results in Fig.7(c). We can observe that smaller value of ϵ can provide a more accurate approximation with higher accuracy and thus we set $\epsilon = 0.01$.

N Miscellaneous Discussions

N.1 The role of ζ in ETM-based method

We first discuss why we should involve translation invariant ζ in both L_{U} and L_{P}. Specifically, we first analyze the case of SemiUOT. The Fenchel-Lagrange conjugate form of SemiUOT without translation invariant mechanism is given as:

$$\begin{aligned} &\min_{\boldsymbol{f},\boldsymbol{g},\zeta} \left[\tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) - \sum_{j=1}^{N} b_j g_j \right] \\ &s.t.\ f_i + g_j \leq C_{ij}. \end{aligned} \tag{62}$$

We can adopt c-transform on Eq.(62) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{f}} \widetilde{L}_{\mathrm{P}} = \tau \sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) - \sum_{j=1}^{N} \inf_{k \in [M]} [C_{kj} - f_k] b_j, \tag{63}$$

We adopt L-BFGS to optimize $\widetilde{L}_{\mathrm{P}}$ using the same data samples as shown in Fig.1 with $\tau = 1$. Meanwhile, the translation invariant term ζ in SemiUOT should be calculated as follows:

$$\zeta = \tau \log\left(\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right)\right) - \tau \log\left(\sum_{j=1}^{N} b_j\right). \tag{64}$$

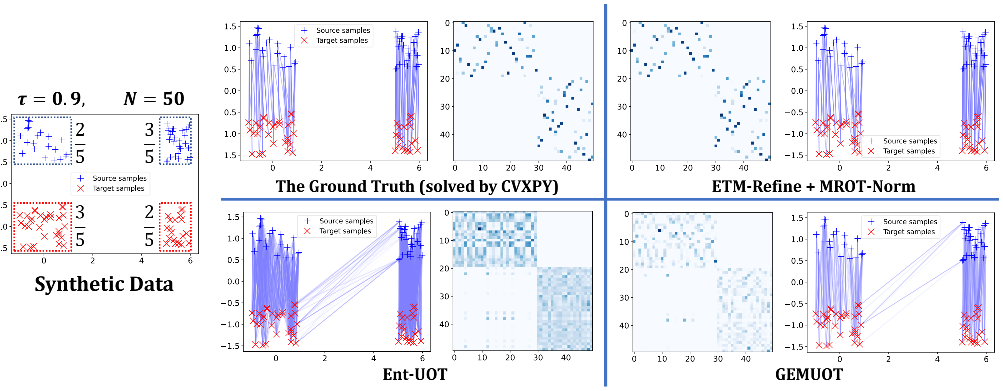


Figure 9: The UOT matching results on class imbalanced case. The blue '+' and red '×' denote the source and target samples, respectively.

Ideally, ζ should equals to 0 since $\sum_{i=1}^{M} a_i \exp\left(-\frac{f_i}{\tau}\right) = \sum_{j=1}^{N} b_j$. However, we can observe that $\zeta > 0$ during the iteration epoch on optimizing $\widetilde{L}_{\mathrm{P}}$ as shown in Fig.8(a). Therefore we can conclude that ζ is indispensable during the calculation on SemiUOT. Likewise, the Fenchel-Lagrange conjugate form of UOT without translation invariant mechanism is given as:

$$\min_{\boldsymbol{v},\boldsymbol{u}} \left[\tau_a \Big\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\Big\rangle + \tau_b \Big\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\Big\rangle\right] \quad s.t.\ u_i + v_j \le C_{ij}. \tag{65}$$

Here we can adopt c-transform on Eq.(65) to obtain the unconstrained optimization problem as:

$$\min_{\boldsymbol{u}} \widetilde{L}_{\mathrm{U}} = \tau_a \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i}{\tau_a}\right) + \tau_b \sum_{j=1}^{N} b_j \exp\left(\frac{\sup_{k=1}^{M}(u_k - C_{kj})}{\tau_b}\right). \tag{66}$$

We also adopt L-BFGS to optimize $\widetilde{L}_{\mathrm{U}}$ using the same data samples as shown in Fig.2 with $\tau_a = \tau_b = 1$. Meanwhile, the translation invariant term ζ in UOT should be calculated as follows:

$$\zeta = \frac{\tau_a \tau_b}{\tau_a + \tau_b}\left[\log\Big\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\Big\rangle - \log\Big\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\Big\rangle\right]. \tag{67}$$

Ideally, ζ should equals to 0 since $\Big\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}}{\tau_a}\right)\Big\rangle = \Big\langle \boldsymbol{b}, \exp\left(-\frac{\boldsymbol{v}}{\tau_b}\right)\Big\rangle$. However, we can observe that $\zeta > 0$ during the iteration epoch on optimizing $\widetilde{L}_{\mathrm{U}}$ as shown in Fig.8(b). Therefore we can conclude that ζ is indispensable during the calculation on UOT. In conclusion, the concept of translation invariant was first proposed in [87]. However, [87] only utilizes translation invariant for entropic UOT. **We highlight that, in this paper, we further extend translation invariant for standard UOT/SemiUOT scenario**. We illustrate that translation invariant is essential in solving UOT and SemiUOT problems.

N.2 UOT investigation on class imbalanced case

We also conduct the experiments for UOT on the class imbalanced scenario following the settings in [28] . Specifically, the source and target data distributions (Gaussian mixtures of two uniform distributions with different weights) are defined as $P_X = \frac{2}{5}U([-1,1]\times[0.5,1.5]) + \frac{3}{5}U([5,6]\times[0.5,1.5])$ and $P_Z = \frac{3}{5}U([-1,1]\times[-0.5,-1.5]) + \frac{2}{5}U([5,6]\times[-0.5,-1.5])$. We first sample $N = 50$ data for both P_X and P_Z with $\tau = 0.9$ and conduct the UOT experiments shown in Fig.9. Note that data No.1–No.20 in P_X are sampled from $U([-1,1]\times[0.5,1.5])$, and No.21–No.50 data in P_X are sampled from $U([5,6]\times[0.5,1.5])$. Meanwhile, No.1–No.30 data in P_Z are sampled from $U([-1,1]\times[-0.5,-1.5])$, and No.31–No.50 data in P_Z are sampled from $U([5,6]\times[-0.5,-1.5])$ to setup the synthetic data experiment. From that we can observe: (1) Ent-UOT could only provide coarse and inaccurate matching results. Moreover, Ent-UOT may lead to mismatch when $\tau = 0.9$. (2) GEMUOT with square norm regularization term can obtain more sparse matching results. However, the output results of GEMUOT are still far away from the ground truth. (3) Our proposed ETM-Refine + MROT-Norm can achieve more accurate results while avoiding mismatch compared with Ent-UOT and GEMUOT, which indicates the efficacy of the proposed method.