SUPPLEMENTAL MATERIAL TO "PRIVACY PROTECTED MULTI-DOMAIN COLLABORATIVE LEARNING"

Anonymous authors
Paper under double-blind review

In this supplement, we first summarize the experimental results of our proposed MDFNet on solving privacy protected multi-domain collaborative learning (P^{2}MDCL) with multiple clients, and then report the comparison on two tasks under the original protocol of SHOT, where all source samples are used for training and the evaluation focuses on target domain. Finally, we provide the formal proof for the Lemma 1, Lemma 2 and Theorem 1 mentioned in our main manuscript.

1 EXPERIMENTS

In the main manuscript, we focus on P^{2}MDCL with two clients (source and target domains). The corresponding results illustrate our learning strategy effectively eliminates the domain discrepancy and assists each client to gain the additional benefits.

To further evaluate our MDFNet, we execute the experiments with four clients on office-home benchmark including four domains: Artistic images (**Ar**), Clip Art (**Cl**), Product images (**Pr**) and Real-World images (**Rw**). Four subsets with 15,500 images share the identical label space of 65 categories. For the P^{2}MDCL with multi-clients, arbitrary two domains are considered as the labeled source clients and the remaining ones are unlabeled target clients. And each source domain is divided into one training set and one test set which involve the same number of samples. Under this condition, our MDFNet attempts to learn the consensus model with four clients and one global server. For source-only method (Src-only), we train the source model with all annotated samples and directly apply it to identify instances on the test set of each client. In terms of the source-free based solution SHOT, the source model trained with all available labeled images is adapted into the integration of two unlabeled domains, yielding the final model tested on each client. The comparisons of object recognition are reported in Table 1. Different from the SHOT and Src-only, our MDFNet not only gradually promotes the performance of model on target domains but also achieves a slight improvement on the source domains in most cases. It demonstrates our learning strategy effectively solves P^{2}MDCL with multiple clients.

Table 1: Comparisons of Object Recognition Accuracy (%) for P^{2}MDCL on Office-Home benchmark with multi-clients, where S denotes the labeled source clients while T means the unlabeled target clients. For each task, we adopt **bold** to highlight the best performance and underline to emphasis the second highest result. **HM** is the harmonic mean defined as $\frac{4}{(1/\mathbf{ACC}_{Ar})+(1/\mathbf{ACC}_{Cl})+(1/\mathbf{ACC}_{Pr})+(1/\mathbf{ACC}_{Rw})}$, where $\mathbf{ACC}_{Ar}$ is the accuracy on test set of **Ar** domain.

Method	Training \ Test	$\mathbf{ACC}_{Ar}$	$\mathbf{ACC}_{Cl}$	$\mathbf{ACC}_{Pr}$	$\mathbf{ACC}_{Rw}$	**HM**
Src-only	S:**Ar& Cl** T:**Pr& Rw**	76.52	**84.52**	69.29	74.71	75.88
SHOT		73.89	75.45	76.44	78.81	76.11
Ours		**77.35**	83.96	**78.37**	**80.81**	**80.04**
Src-only	S:**Ar& Pr** T:**Cl& Rw**	77.57	49.53	93.55	77.43	70.56
SHOT		74.05	56.17	88.06	79.84	72.48
Ours		**78.18**	**57.85**	**93.61**	**81.01**	**75.32**
Src-only	S:**Ar& Rw** T:**Pr& Rw**	77.08	50.26	73.55	88.09	69.18
SHOT		72.98	54.41	78.08	86.32	70.83
Ours		**77.76**	**55.9**	**78.49**	**89.12**	**73.11**
Src-only	S:**Cl& Pr** T:**Ar& Rw**	52.04	84.35	**93.14**	69.36	71.14
SHOT		61.43	79.84	88.92	73.81	74.63
Ours		**61.68**	**85.43**	92.93	**75.26**	**76.97**
Src-only	S:**Cl& Rw** T:**Ar& Pr**	55.79	85.1	69.52	88.23	72.21
SHOT		**61.72**	78.29	74.89	86.09	73.94
Ours		61.13	**86.21**	**76.44**	**89.03**	**76.75**
Src-only	S:**Pr& Rw** T:**Ar& Cl**	58.26	45.57	92.23	**90.07**	65.52
SHOT		62.88	50.51	87.66	85.13	67.96
Ours		**64.03**	**51.91**	**93.6**	89.44	**70.49**

Moreover, we follow the original protocol of source-free where all source instances are used for training and the evaluation focuses on target domain to re-implement SHOT on tasks Ar→Rw and

Pr→Cl of Office-Home datasets and achieve the performance as 80.9% and 54.5% which are very similar with their reported results. In addition, under the same setting, we evaluate our proposed method (MDFNet) on these two tasks and achieve the 82.5% and and 56.5%. Compared to the results in our manuscript, we can achieve two important conclusions. Firstly, the considerable reduction of source training samples negatively affects the model performance in the mentioned methods. Second, our MDFNet can achieve better results than other baselines with insufficient source samples.

2 THEORETICAL ANALYSIS

Lemma 1. Suppose the h is a hypothesis of class $\mathcal{H}$, for each unlabeled client, we then achieve:

$$|\epsilon_\alpha(h) - \epsilon_{u_j}(h)| \leq \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}),$$

where $\lambda_{l_i} := \epsilon_{l_i}(h^*) + \epsilon_{u_j}(h^*)$ and h^* is the hypothesis which achieves the minimum risk on $\mathcal{D}_{l_i}$ and $\mathcal{D}_{u_j}$, and λ_{u_i} similarly means the risk of optimal hypothesis on the mixture of $\mathcal{D}_{u_i}$ and $\mathcal{D}_{u_j}$. Akin to unlabeled clients, we also derive the analogous inequality in clients with ground-truth as:

$$|\epsilon_\alpha(h) - \epsilon_{l_j}(h)| \leq \sum_{i=1, i\neq j}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{l_j}) + \lambda_{l_i}) + \sum_{i=L+1}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{l_j}) + \lambda_{u_i}),$$

where λ_{l_i} is the risk of optimal hypothesis of $\mathcal{D}_{l_i}$ and $\mathcal{D}_{l_j}$, and $\lambda_{u_i} := \epsilon_{u_i}(h^*) + \epsilon_{l_j}(h^*)$.

Proof.

$$\begin{aligned}
|\epsilon_\alpha(h) - \epsilon_{u_j}(h)| &= |\sum_{i=1}^{L} \alpha_i \epsilon_{l_i}(h) + \sum_{i=L+1}^{L+U} \alpha_i \epsilon_{u_i}(h) - \sum_{i=1}^{L+U} \alpha_i \epsilon_{u_j}(h)| \\
&\leq \sum_{i=1}^{L} \alpha_i |\epsilon_{l_i}(h) - \epsilon_{u_j}(h)| + \sum_{i=L+1}^{L+U} \alpha_i |\epsilon_{u_i}(h) - \epsilon_{u_j}(h)| \\
&\leq \sum_{i=1}^{L} \alpha_i [|\epsilon_{l_i}(h) - \epsilon_{l_i}(h, h^*)| + |\epsilon_{l_i}(h, h^*) - \epsilon_{u_j}(h, h^*)| + |\epsilon_{u_j}(h, h^*) - \epsilon_{u_j}(h)|] \\
&\quad + \sum_{i=L}^{L+U} \alpha_i [|\epsilon_{u_i}(h) - \epsilon_{u_i}(h, h^*)| + |\epsilon_{u_i}(h, h^*) - \epsilon_{u_j}(h, h^*)| + |\epsilon_{u_j}(h, h^*) - \epsilon_{u_j}(h)|] \\
&\leq \sum_{i=1}^{L} \alpha_i [\epsilon_{l_i}(h^*) + |\epsilon_{l_i}(h, h^*) - \epsilon_{u_j}(h, h^*)| + \epsilon_{u_j}(h^*)] \\
&\quad + \sum_{i=L}^{L+U} \alpha_i [\epsilon_{u_i}(h^*) + |\epsilon_{u_i}(h, h^*) - \epsilon_{u_j}(h, h^*)| + \epsilon_{u_j}(h^*)] \\
&\leq \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i})
\end{aligned} \tag{1}$$

For the situation of labeled clients, the proof procedure is similar with the above.

Lemma 2. Given a hypothesis space $\mathcal{H}$ of VC-dimension d, if a random sample of size n is generated by selecting $n\beta_j$ data points from $\mathcal{D}_{l_j}$ or $\mathcal{D}_{u_j}$, and annotating them through f_{l_j} and f_{u_j}, then with probability at least $1-\delta$, $\forall h \in \mathcal{H}$, we have:

$$|\hat{\epsilon}_\alpha(h) - \epsilon_\alpha(h)| \leq \sqrt{\sum_{j=1}^{L+U} \frac{\alpha_j^2}{\beta_j}} \sqrt{\frac{d\log(2n) - \log\delta}{2n}}.$$

Proof. For the $n\beta_j$ samples x from the labeled client $\mathcal{D}_{l_j}$, we let $X_{n\sum_{k=1}^{j-1}\beta_k+1}, ..., X_{n\sum_{k=1}^{j}\beta_k}$ be the random variables with the value $\frac{\alpha_j}{\beta_j}|h(x) - f_{l_j}(x)|$. Similarly, for the $n\beta_j$ samples x from the

unlabeled client $\mathcal{D}_{u_j}$, we let $X_{n\sum_{k=1}^{j-1}\beta_k+1}, ..., X_{n\sum_{k=1}^{j}\beta_k}$ be the random variables with the value $\frac{\alpha_j}{\beta_j}|h(x) - f_{u_j}(x)|$. Note that for the labeled or unlabeled clients, $X_{n\sum_{k=1}^{j-1}\beta_k+1}, ..., X_{n\sum_{k=1}^{j}\beta_k} \in [0, \alpha_j/\beta_j]$. And then, we have

$$\begin{aligned}\hat{\epsilon}_\alpha(h) &= \sum_{j=1}^{L}\alpha_j\hat{\epsilon}_{l_j}(h) + \sum_{j=L+1}^{L+U}\alpha_j\hat{\epsilon}_{u_j}(h) \\ &= \sum_{j=1}^{L}\frac{\alpha_j}{\beta_j n}\sum_{x\in\mathcal{D}_{l_j}}|h(x) - f_{l_j}(x)| + \sum_{j=L+1}^{L+U}\frac{\alpha_j}{\beta_j n}\sum_{x\in\mathcal{D}_{u_j}}|h(x) - f_{u_j}(x)|\end{aligned} \tag{2}$$

Due to the linearity of expectations, we can achieve

$$\begin{aligned}E[\epsilon_\alpha(\hat{h})] &= \frac{1}{n}(\sum_{j=1}^{L}\beta_j n\frac{\alpha_j}{\beta_j}\epsilon_{l_j}(h) + \sum_{j=L+1}^{L+U}\beta_j n\frac{\alpha_j}{\beta_j}\epsilon_{u_j}(h)) \\ &= \sum_{j=1}^{L}\alpha_j\epsilon_{l_j}(h) + \sum_{j=L+1}^{L+U}\alpha_j\epsilon_{u_j}(h) = \epsilon_\alpha(h)\end{aligned} \tag{3}$$

With the Hoeffding's inequality Hoeffding (1994), the following holds for each h.

$$Pr[|\hat{\epsilon}_\alpha(h) - \epsilon_\alpha(h)| \geq \epsilon] \leq 2\exp(\frac{-2n\epsilon^2}{\sum_{j=1}^{L+U}\frac{\alpha_j^2}{\beta_j}}) \tag{4}$$

Finally, we substitute the probability with δ and achieve the following.

$$|\hat{\epsilon}_\alpha(h) - \epsilon_\alpha(h)| \leq \epsilon = \sqrt{\sum_{j=1}^{L+U}\frac{\alpha_j^2}{\beta_j}}\sqrt{\frac{d\log(2n) - \log\delta}{2n}} \tag{5}$$

Theorem 1. Suppose that we are given $n\beta_i$ labeled instances from client $\mathcal{D}_{l_i}$ for $i = 1...L$, and $n\beta_j$ unlabeled instances from client $\mathcal{D}_{u_j}$ in a federated learning system. Let $\hat{h} = \arg\min_{h\in\mathcal{H}}\hat{\epsilon}_\alpha(h)$, and $h^*_{l_i} := \arg\min_{h\in\mathcal{H}}\epsilon_{l_i}(h)$ and $h^*_{u_j} := \arg\min_{h\in\mathcal{H}}\epsilon_{u_j}(h)$. Then, $\forall\alpha_i \in \mathbb{R}_+, \sum_{i=1}^{L+U}\alpha_i = 1$, with probability at least $1-\delta$ over the choice of samples from each client,

$$\begin{aligned}\epsilon_{u_j}(\hat{h}) \leq& \epsilon_{u_j}(h^*_{u_j}) + 2\sqrt{\sum_{j=1}^{L+U}\frac{\alpha_j^2}{\beta_j}}\sqrt{\frac{d\log(2n) - \log\delta}{2n}} \\ &+ 2\big(\sum_{i=1}^{L}\alpha_i(\frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U}\alpha_i(\frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i})\big).\end{aligned}$$

Proof. Combining the conclusions from **Lemma 1** and **Lemma 2**, we have

$$
\begin{aligned}
\epsilon_{u_j}(\hat{h}) &\leq \epsilon_{\alpha}(\hat{h}) + \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}) \\
&\leq \hat{\epsilon}_{\alpha}(\hat{h}) + \sqrt{\sum_{j=1}^{L+U} \frac{\alpha_j^2}{\beta_j}} \sqrt{\frac{d\log(2n) - \log\delta}{2n}} \\
&+ \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}) \\
&\leq \hat{\epsilon}_{\alpha}(h^*_{u_j}) + \sqrt{\sum_{j=1}^{L+U} \frac{\alpha_j^2}{\beta_j}} \sqrt{\frac{d\log(2n) - \log\delta}{2n}} \\
&+ \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}) \\
&\leq \epsilon_{\alpha}(h^*_{u_j}) + 2\sqrt{\sum_{j=1}^{L+U} \frac{\alpha_j^2}{\beta_j}} \sqrt{\frac{d\log(2n) - \log\delta}{2n}} \\
&+ \sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}) \\
&\leq \epsilon_{u_j}(h^*_{u_j}) + +2\sqrt{\sum_{j=1}^{L+U} \frac{\alpha_j^2}{\beta_j}} \sqrt{\frac{d\log(2n) - \log\delta}{2n}} \\
&+ 2\Big(\sum_{i=1}^{L} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{l_i}, \mathcal{D}_{u_j}) + \lambda_{l_i}) + \sum_{i=L+1, i\neq j}^{L+U} \alpha_i (\frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_{u_i}, \mathcal{D}_{u_j}) + \lambda_{u_i}) \Big)
\end{aligned}
\tag{6}
$$

Similarly, for the labeled clients, we follow the above proof and can derive the similar conclusion.

REFERENCES

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pp. 409–426. Springer, 1994.