Source-free Domain Adaptive Human Pose Estimation (*Supplementary Material*)

Qucheng Peng, Ce Zheng, Chen Chen
Center for Research in Computer Vision, University of Central Florida
{qucheng.peng,ce.zheng}@ucf.edu, chen.chen@crcv.ucf.edu

1. Overview

The supplementary material is organized into the following sections:

- Section 2: Additional qualitative results on FreiHand and Human3.6M datasets.
- Section 3: Additional ablation of framework on FreiHand and LSP datasets.
- Section 4: Additional ablation of losses on FreiHand and LSP datasets.
- Section 5: Domain generalization to unseen domains based on models trained on domain adaptation tasks.

2. Additional Qualitative Results

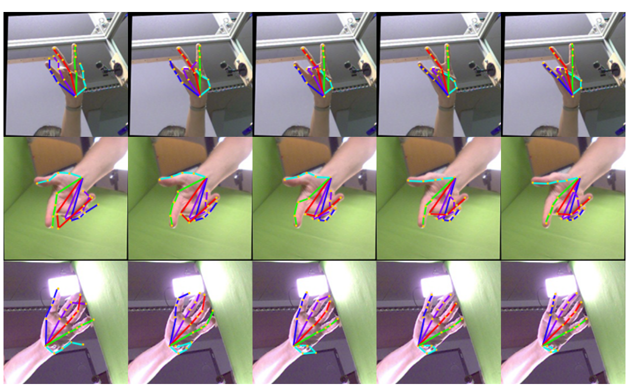

Figure 1: Qualitative results on FreiHand dataset (Best view with zoom in)

In this section, we offer additional qualitative results on the RHD→FreiHand task and the SURREAL→Human3.6M task. Results are exhibited in Fig. 1 and Fig. 2.These figures show that our method outperforms other competing methods, predicting more accurate poses in the target domain.

Figure 2: Qualitative results on Human3.6M dataset (Best view with zoom in)

3. Additional Ablation of Framework

Our method contains two modules in Step A & Step B as the source-protect module (**SP**) and the target-relevant module (**TR**) separately, and here we focus on their functions. Mutual Mean Teaching (**MMT**) [1] is a domain adaptation strategy that preserves source information. We utilize MMT as the baseline method to evaluate the effectiveness of TR. Furthermore, comparing MMT with SP enables us to determine if our SP module outperforms MMT. Table 1 and 2 show the ablation study of frameworks on RHD→ FreiHand and SURREAL→ LSP with different combinations of these modules.

Table 1: Ablation of Frameworks on RHD → FreiHand

Method	MCP	PIP	DIP	Fin	All
MMT [1]	39.6	60.4	60.0	57.8	52.6
MMT [1] +TR	41.5	62.1	63.9	60.4	55.9
SP+TR (Ours)	43.7	65.9	66.6	63.1	58.8

The results clearly demonstrate that both SP and TR contribute to improving the model's performance. Specifically, TR enhances the model's accuracy by 3.3% on RHD → FreiHand and 5.6% on SURREAL → LSP, while SP leads

Table 2: Ablation of Frameworks on SURREAL → LSP

Method	Sld	Elb	Wrist	Hip	Knee	Ankle	All
MMT [1]	60.9	70.9	70.3	81.1	79.3	72.8	71.5
MMT [1] +TR	65.2	79.6	81.4	82.3	82.8	79.7	77.1
SP+TR (Ours)	70.7	85.4	83.8	86.6	85.2	85.0	83.2

to an improvement of 2.9% on RHD → FreiHand and 5.1% on SURREAL → LSP compared to MMT. Notably, the two proposed modules provide similar levels of improvement.

4. Additional Ablation of Losses

We performed a detailed ablation study on the three proposed losses, namely $\mathcal{L}_{res}$, $\mathcal{L}_{cst}$, and $\mathcal{L}_{im}$, using the RHD→ FreiHand and SURREAL→ LSP tasks. Tables 3 and 4 present the results.

Table 3: Ablation of Losses on RHD → FreiHand

Method	MCP	PIP	DIP	Fin	All
Baseline	41.2	63.5	63.8	60.9	56.3
$\mathcal{L}_{res}$	41.6	64.0	64.4	61.5	56.9
$\mathcal{L}_{cst}$	41.8	64.7	65.0	62.2	57.9
$\mathcal{L}_{im}$	41.5	64.4	64.8	61.7	57.1
$\mathcal{L}_{cst}\&\mathcal{L}_{im}$	42.3	65.0	65.4	62.5	58.1
$\mathcal{L}_{res}\&\mathcal{L}_{cst}\&\mathcal{L}_{im}$	43.7	65.9	66.6	63.1	58.8

Table 4: Ablation of Losses on SURREAL → LSP

Method	Sld	Elb	Wrist	Hip	Knee	Ankle	All
Baseline	69.9	82.1	81.3	84.5	82.7	80.4	80.3
$\mathcal{L}_{res}$	70.2	82.8	81.6	85.0	83.2	80.5	80.9
$\mathcal{L}_{cst}$	70.5	83.5	82.0	85.8	84.0	83.7	82.0
$\mathcal{L}_{im}$	70.3	83.0	81.9	85.2	83.4	82.2	81.3
$\mathcal{L}_{cst}\&\mathcal{L}_{im}$	70.6	84.8	82.7	86.0	84.6	84.1	82.5
$\mathcal{L}_{res}\&\mathcal{L}_{cst}\&\mathcal{L}_{im}$	70.7	85.4	83.8	86.6	85.2	85.0	83.2

We observe that each loss is able to boost the model's performance. Simply applying $\mathcal{L}_{res}$ leads to an increase of 0.6% in RHD→ FreiHand and SURREAL→ LSP. $\mathcal{L}_{cst}$ causes an improvement of 1.6% in RHD→ FreiHand and 1.7% in SURREAL→ LSP. As for $\mathcal{L}_{im}$, adding it achieves a performance gain of 0.8% in RHD→ FreiHand and 1.2% in SURREAL→ LSP. In addition, we observe that $\mathcal{L}_{cst}$ has a more significant impact than $\mathcal{L}_{res}$ or $\mathcal{L}_{im}$, as it yields greater improvements when compared to the other two.

5. Generalization to Unseen Domains

Following prior works [4, 3], we also conduct experiments on the generalization to unseen domains. For hand pose estimation, we use models adapted in the RHD→H3D task and evaluate their performances on the validation set of FreiHand, as shown in Table 5. For human pose estimation, we use models adapted in the SURREAL→LSP task and evaluate their performance on Human3.6M, as shown in Table 6.

Table 5: Domain Generalization on FreiHand

Method	SF	MCP	PIP	DIP	Fin	All
Source-only	-	34.9	48.7	52.4	48.5	45.8
CC-SSL [7] (CVPR'20)	×	34.3	46.3	48.4	44.4	42.6
MDAM [4] (CVPR'21)	×	29.6	46.6	50.0	45.3	42.2
RegDA [2] (CVPR'21)	×	37.8	51.8	53.2	47.5	46.9
UniFrame [3] (ECCV'22)	×	35.6	52.3	55.4	50.6	47.1
RegDA-SF [2] (CVPR'21)	✓	30.5	47.6	50.6	44.9	42.5
SHOT [5] (ICML'20)	✓	32.0	48.1	49.9	42.4	41.8
UniFrame-SF [3] (ECCV'22)	✓	32.7	48.5	51.3	45.7	43.0
SHOT++ [6] (TPAMI'22)	✓	33.6	49.2	52.5	47.0	44.6
Ours	✓	**34.4**	**50.8**	**54.7**	**48.3**	**46.2**

Table 6: Domain Generalization on Human3.6M

Method	SF	Sld	Elb	Wrist	Hip	Knee	Ankle	All
Source-only	-	51.5	65.0	62.9	68.0	68.7	67.4	63.9
CC-SSL [7] (CVPR'20)	×	52.7	76.9	63.1	31.6	75.7	72.9	62.2
MDAM [4] (CVPR'21)	×	54.4	75.3	62.1	21.6	70.4	69.2	58.8
RegDA [2] (CVPR'21)	×	76.9	80.2	69.7	52.0	80.3	80.0	73.2
UniFrame [3] (ECCV'22)	×	77.0	85.9	73.8	47.6	80.7	80.6	74.3
RegDA-SF [2] (CVPR'21)	✓	67.4	74.1	65.8	47.4	71.8	74.0	65.6
SHOT [5] (ICML'20)	✓	68.6	75.8	67.0	48.1	72.4	74.4	66.2
UniFrame-SF [3] (ECCV'22)	✓	68.4	74.7	66.0	48.3	72.2	74.9	66.6
SHOT++ [6] (TPAMI'22)	✓	69.7	76.0	66.4	**48.8**	73.4	75.8	67.9
Ours	✓	**73.6**	**79.8**	**68.3**	48.0	**75.9**	**77.7**	**70.5**

From these two tables, we can see that our model outperforms the second-best source-free approach for a lead of 1.6% on FreiHand and 2.6% on Human3.6M.

References

[1] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations*, 2020. 1, 2

[2] Junguang Jiang, Yifei Ji, Ximei Wang, Yufeng Liu, Jianmin Wang, and Mingsheng Long. Regressive domain adaptation for unsupervised keypoint detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6780–6789, 2021. 2

[3] Donghyun Kim, Kaihong Wang, Kate Saenko, Margrit Betke, and Stan Sclaroff. A unified framework for domain adaptive pose estimation. *arXiv preprint arXiv:2204.00172*, 2022. 2

[4] Chen Li and Gim Hee Lee. From synthetic to real: Unsupervised domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1482–1491, 2021. 2

[5] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020. 2

[6] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8602–8617, 2022. 2

[7] Jiteng Mu, Weichao Qiu, Gregory D Hager, and Alan L Yuille. Learning from synthetic animals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12386–12395, 2020. 2