Regressive Domain Adaptation for Unsupervised Keypoint Detection

Junguang Jiang[1], Yifei Ji[1], Ximei Wang[1], Yufeng Liu[2], Jianmin Wang[1], Mingsheng Long[1]

[1]Tsinghua University, [2]Kuaishou Technology

Problem

- The annotations of 2D keypoints on real images are expensive and time-consuming to collect while that on synthetic images can be obtained in abundance by CG at a low cost.
- Domain shifts between virtual and real domains will cause significant performance drop, thus domain adaptation is important for this problem.

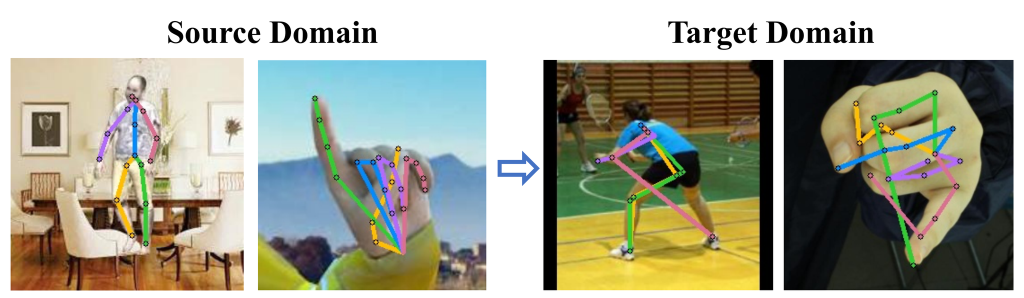

Figure 1:Confusion of keypoints on the target domain.

- No clear decision boundary exists in regression, thus feature alignment, such as DAN and DANN, cannot enlarge the margins of boundaries to generalize the model as done in classification.

Architecture

- Train the adversarial regressor f' to predict correctly on the source domain, while differ from f as much as possible on the target domain.
- Encourage the feature extractor ψ to output domain-invariant features and deceive f'.

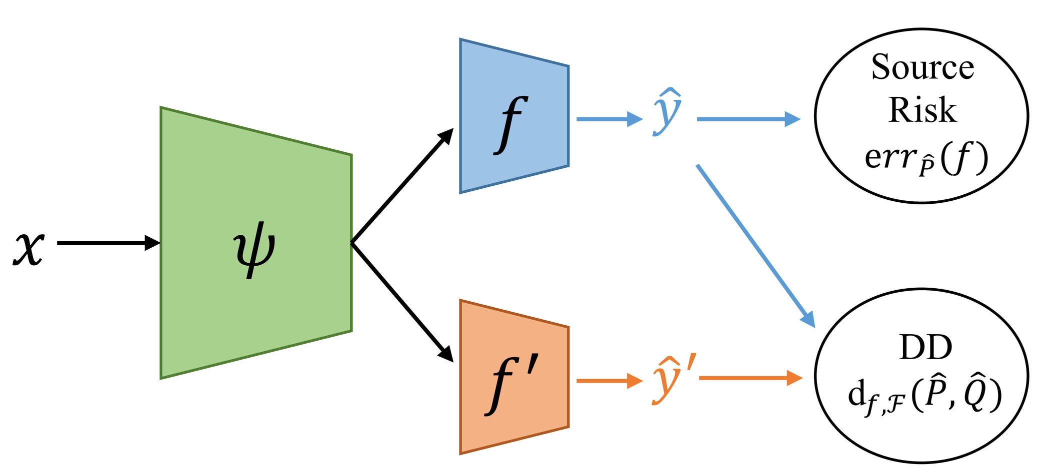

Figure 2:Architecture. ψ: feature generator, f: regressor head, f': adversarial regressor head.

- In keypoint detection where output space is large, it's hard to find the adversarial regressor f' that does poorly *only* on the target domain.

Sparsity of the Spatial Density

- When the position of the right ankle is mistaken, most likely the left ankle is predicted, occasionally other keypoints predicted, and rarely positions on the background are predicted.

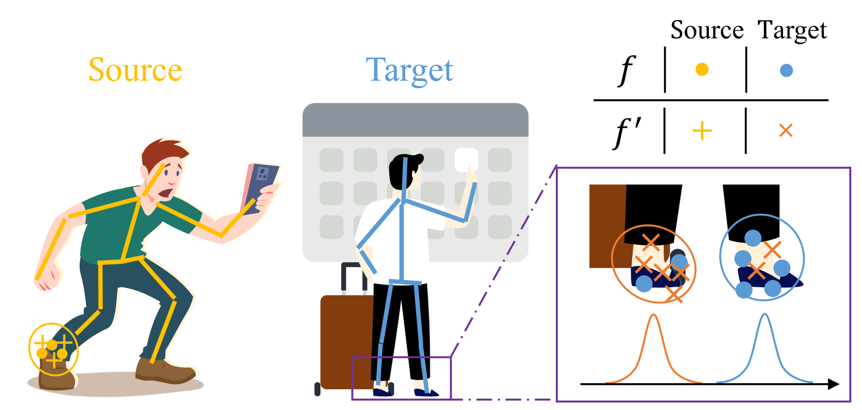

Figure 3:The probability density of the output space is sparse

- Define *ground false* prediction.

$$\mathcal{H}_{\mathrm{F}}(\widehat{\boldsymbol{y}}_k)_{h,w} = \sum_{k' \neq k} \mathcal{H}(\widehat{\boldsymbol{y}}_{k'})_{h,w}, \tag{1}$$

$$\mathcal{P}_{\mathrm{F}}(\widehat{\boldsymbol{y}}_k)_{h,w} = \frac{\mathcal{H}_{\mathrm{F}}(\widehat{\boldsymbol{y}}_k)_{h,w}}{\sum_{h'=1}^{H'} \sum_{w'=1}^{W'} \mathcal{H}_{\mathrm{F}}(\widehat{\boldsymbol{y}}_k)_{h',w'}}, \tag{2}$$

where $\mathcal{H}(\boldsymbol{y}_k)$ generate a ground truth heatmap for $\boldsymbol{y}_k$.

Minimax of Target Disparity

- When maximizing the discrepancy between f' and f, we expect to maximize the mean difference, but often the variance is changed.

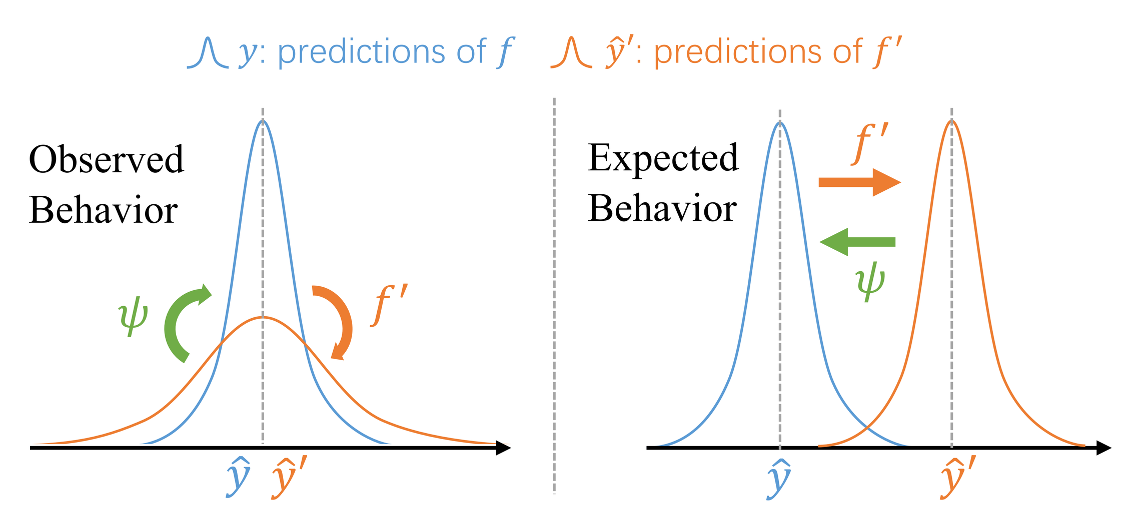

Figure 4:Maximizing the distance between two predictions.

- Convert the minimax game to the minimization of two opposite goals.

Overall Objectives

- Supervised training on the source domain.
- Update f' to minimize its KL with *ground false* predictions of f.
- Update ψ to minimize KL between prediction of f' with *ground truth* prediction of f.

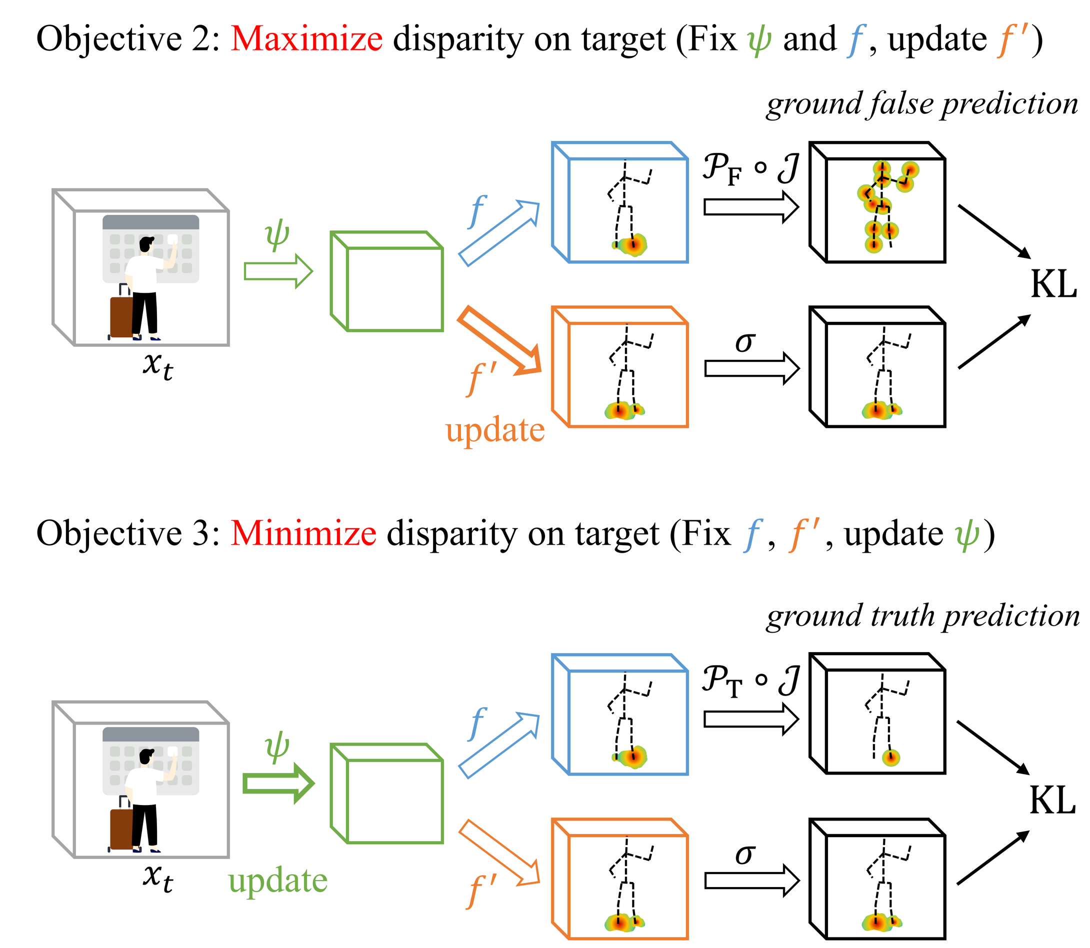

Figure 5:Adversarial training objectives

Quantitative Results

Method	MCP	PIP	DIP	Fingertip	Avg
ResNet101	67.4	64.2	63.3	54.8	61.8
DAN	59.0	57.0	56.3	48.4	55.1
DANN	67.3	62.6	60.9	51.2	60.6
MCD	59.1	56.1	54.7	46.9	54.6
DD	72.7	69.6	66.2	54.4	65.2
RegDA	**79.6**	**74.4**	**71.2**	**62.9**	**72.5**
Oracle	97.7	97.2	95.7	92.5	95.8

Table 1:PCK on task *RHD→H3D*.

Method	f'	ψ	MCP	PIP	DIP	Fingertip	Avg
DD	max L_{T}	min L_{T}	72.7	69.6	66.2	54.4	65.2
	min L_{F}	max L_{F}	74.4	71.1	66.9	56.4	66.5
RegDA	min L_{F}	min L_{T}	**79.6**	**74.4**	**71.2**	**62.9**	**72.5**

Table 2:Ablation on the minimax of target disparity.

Qualitative results

Source only often confuses different key points, resulting in the predicted skeleton not look real. In contrast, the outputs of **RegDA** look more like a human hand or body.

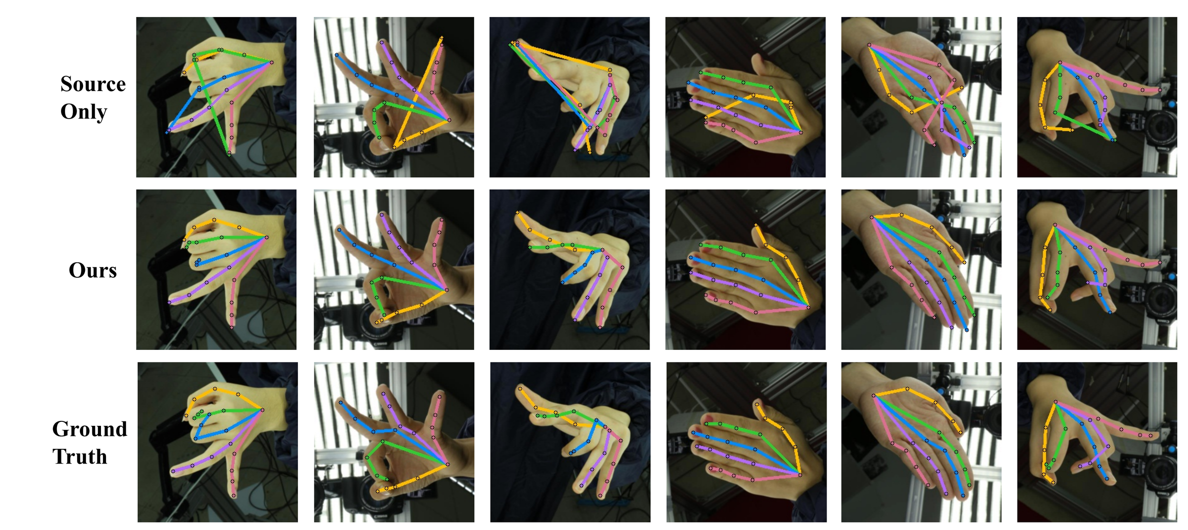

Figure 6:*RHD* → *H3D* dataset.

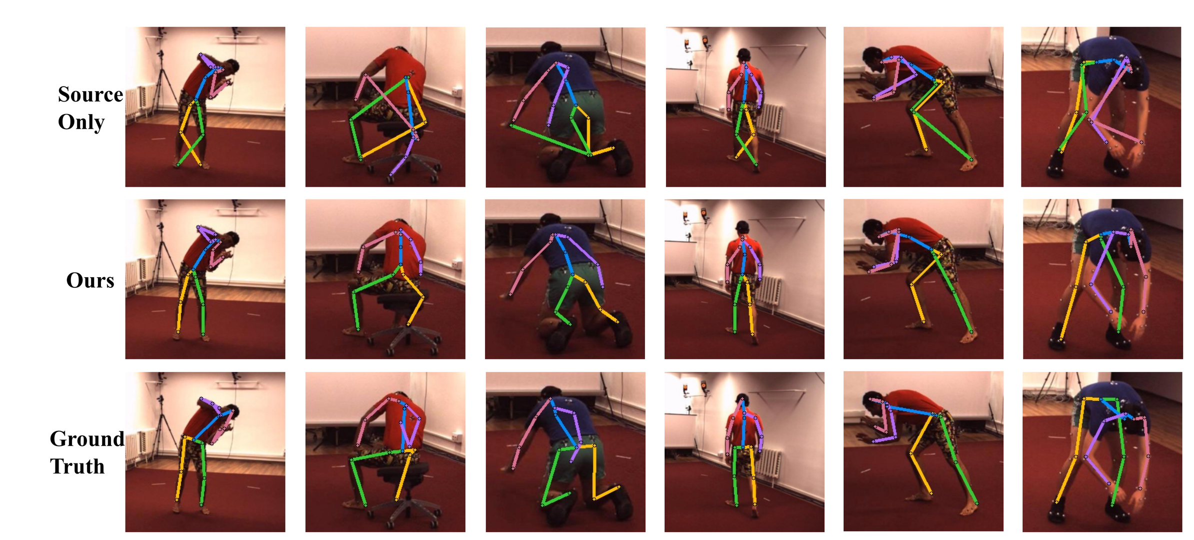

Figure 7:*SURREAL* → *Human3.6M*.

Figure 8:*SURREAL* → *LSP*.

Contact Information

- Paper: Regressive Domain Adaptation for Unsupervised Keypoint Detection
- Code: Transfer-Learning-Library
- Email: JiangJunguang1123@outlook.com