Supplementary Material for `SHIFT`

1. Training Details and Adaptation Results

Prior training paradigm. We adopt two approaches for training our infant pose prior: the first approach includes training directly on the target agnostic dataset and the second approach includes training the prior on the source dataset and then fine-tuning (FT) on the target agnostic set. The results are as below:-

Table 1. **Quantitative Results (PCK@0.05)** for `SHIFT` against FiDIP [2].

Algorithm	SURREAL → MINI-RGBD							
	Head	Sld.	Elb.	Wrist	Hip	Knee	Ankle	Avg.
`SHIFT` w/o FT	96.00	29.20	48.90	34.40	86.10	43.50	75.00	52.80
`SHIFT`	100.00	14.90	68.80	45.20	96.50	40.60	72.70	56.40

Table 2. **Quantitative Results (PCK@0.05)** for `SHIFT` against FiDIP [2].

Algorithm	SURREAL → SyRIP							
	Head	Sld.	Elb.	Wrist	Hip	Knee	Ankle	Avg.
`SHIFT` w/o FT	43.40	40.20	35.20	38.40	49.20	29.20	36.80	38.10
`SHIFT`	45.60	45.00	35.90	38.00	51.40	31.40	32.00	39.00

Fine-tuning directly in a target agnostic setting provides better results than pre-training on source and fine-tuning on the target agnostic set. This suggests that our pre-training regimen is crucial towards preventing source knowledge forgetting; hence re-training the prior on the source dataset is not necessary.

Synthetic Infant to Real Data Adaptation. Using MINI-RGBD[1] as the source dataset results in unsatisfactory performance for both our method and the baseline. This is likely due to its limited diversity in infant poses and minimal inter-frame motion, which hinders effective pre-training for real images with high self-occlusion, as seen in SyRIP [2]. Despite SyRIP having fewer images, its diverse poses and scenarios make it a superior pre-training source.

Table 3. **Quantitative Results (PCK@0.05)** for **SyRIP** [2]→ **MINI-RGBD** [1]. The best accuracies are highlighted in red and the second best accuracies are highlighted in blue.

Algorithm	Unsup	SyRIP → MINI-RGBD							
		Head	Sld.	Elb.	Wrist	Hip	Knee	Ankle	Avg.
Oracle	-	89.40	82.10	65.70	66.10	64.10	50.70	54.50	63.80
FiDIP [2]	✗	52.20	21.30	22.40	14.40	33.20	26.00	23.90	27.55
`SHIFT`	✓	61.80	61.00	41.40	40.40	42.50	33.90	34.70	42.30

2. Additional Ablation Results

Effect of Loss Terms. We ablate each of the loss terms on the SyRIP [2] dataset. The strong role of Kp2Seg ($\mathcal{G}(\cdot)$) is seen in dealing with self-occlusions.

Table 4. We analyse the effects of each loss term and module in this table for **SURREAL** [4] → **SyRIP** [2].

Module	**Loss Terms**				**PCK@0.05**
	$\mathcal{L}_{\text{sup}}$	$\mathcal{L}_{\text{cons}}$	$\mathcal{L}_p$	$\mathcal{L}_{\text{ctx}}$	
Pre-Training	✓	✗	✗	✗	26.30
UDA [3]	✓	✓	✗	✗	34.20
UDA + Prior	✓	✓	✓	✗	35.90
`SHIFT`	✓	✓	✓	✓	**39.80**

References

[1] Nikolas Hesse, Christoph Bodensteiner, Michael Arens, Ulrich G Hofmann, Raphael Weinberger, and A Sebastian Schroeder. Computer vision for medical infant motion analysis: State of the art and rgb-d data set. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 1

[2] Xiaofei Huang, Nihang Fu, Shuangjun Liu, and Sarah Ostadabbas. Invariant representation learning for infant pose estimation with small data. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, 2021. 1

[3] Donghyun Kim, Kaihong Wang, Kate Saenko, Margrit Betke, and Stan Sclaroff. A unified framework for domain adaptive pose estimation. In *European Conference on Computer Vision*, pages 603–620. Springer, 2022. 1

[4] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 109–117, 2017. 1