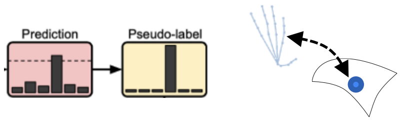
SemiHand: Semi-supervised Hand Pose Estimation with Consistency

Linlin Yang[1,2], Shicheng Chen[1], Angela Yao[1]
[1]National University of Singapore, [2]University of Bonn

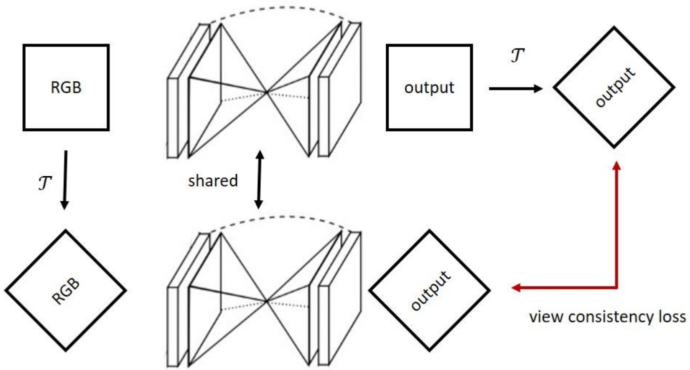
Motivation & Target

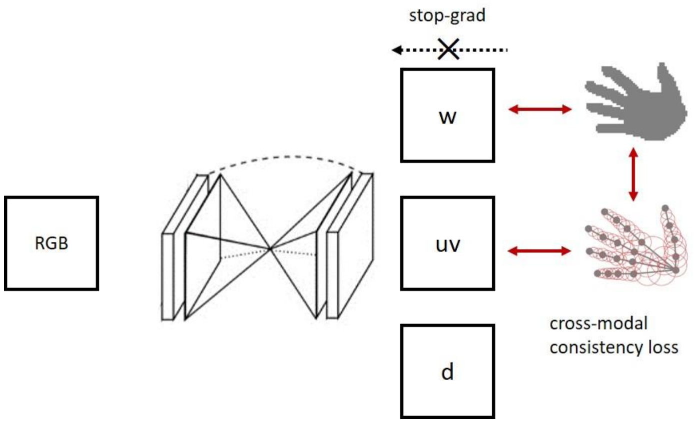
- **Synthesizing training data** is considered an easy alternative to get accurate labels.
- Yet there exists a significant **domain gap** between synthetic and real-world images.
- We propose a novel **RGB-based 3D hand pose** estimation framework using labelled synthetic data and unlabelled real-world data
- It is the first **semi-supervised** framework that combines **pseudo-labelling** with **consistency training** for 3D RGB-based hand pose

Challenge

	Classificaiton	Pose Estimation	Explanation
Perturbation	After adding perturbation like noise, rotation, translation,flip, the label remains the same	Labels change accordingly based on perturbation(spatial information)	When rotating the RGB images, the key point location in the image coordinate also changes
Modalities	One modality	Different modalities also changes accordingly (hand model)	When changing the poses, the depth maps, the mask, the RGB may also change.
Labels (pseudo-labels)	One-hot label	Pose as continuous labels should be biomechanically feasible	Bone length limits + joint angles limits
Correction	Threshold or sharpening is used to revise the label	Pose space is predefined to ensure the biomechanical feasibility	

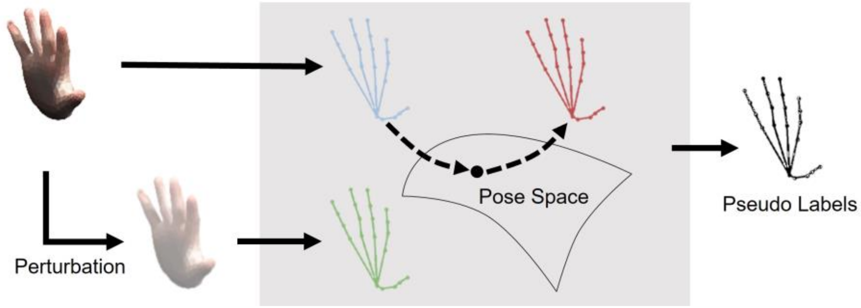
Modules(Consistency Training)





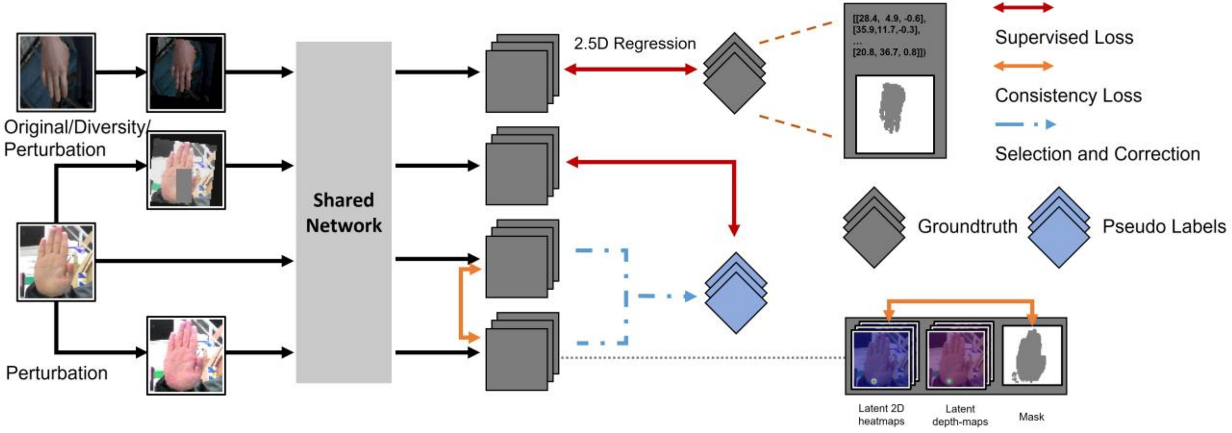
- Consistency for perturbations (geometric augmentation).
- Consistency for auxiliary modalities (hand mask).

Modules(Pseudo-labelling)



- Predicting the pseudo-labels (average).
- Estimating the confidence (variance).
- Based on original, perturbated and corrected predictions.

Pipeline



1. Weak-strong augmentation strategy.
2. Multi-task learning with consistency for perturbations and auxiliary modalities.
3. Label correction based on the biomechanical feasibility of hand pose.
4. Label confidence based on the feasibility and the consistency.

Results

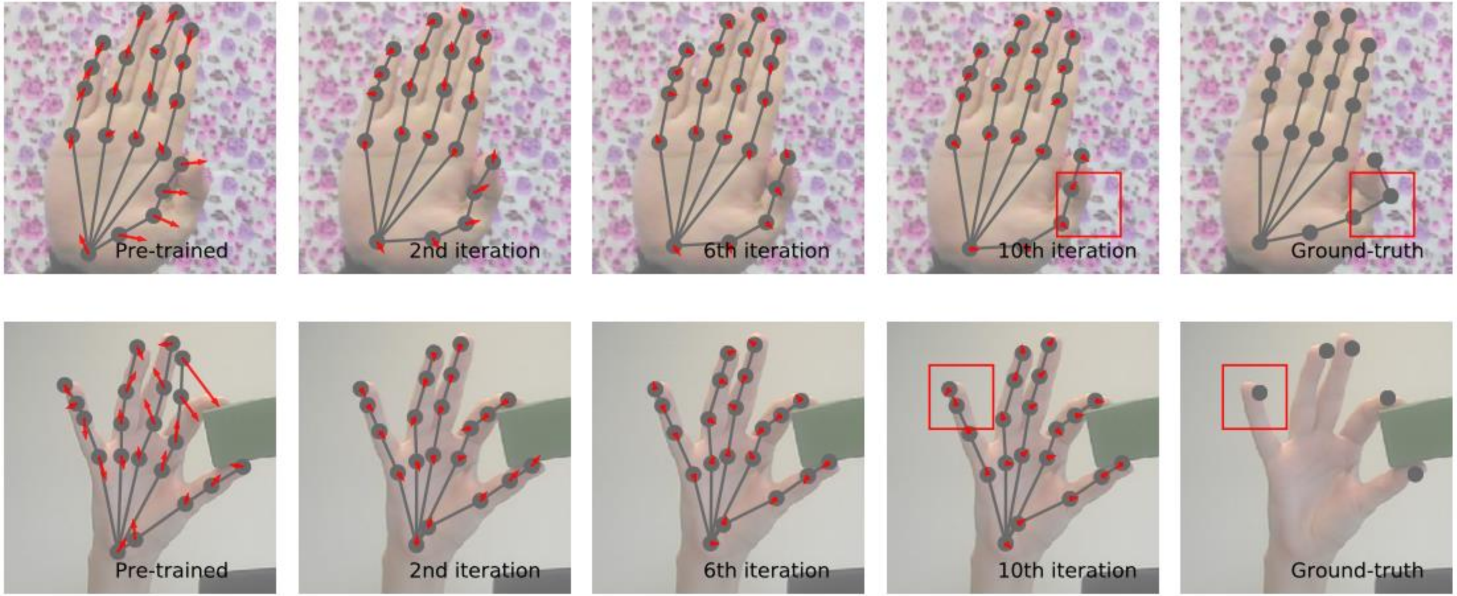


Figure 9: Gradual convergence from the prediction of pre-trained model to our final prediction. The arrows indicate the direction and distance of prediction movement during fine-tuning. For 10^{th} iteration, the optimization converges because the length of arrows become almost zeros. We highlight the differences between our stable predictions and the ground-truth poses with red boxes. Figure best viewed in colour.

Results

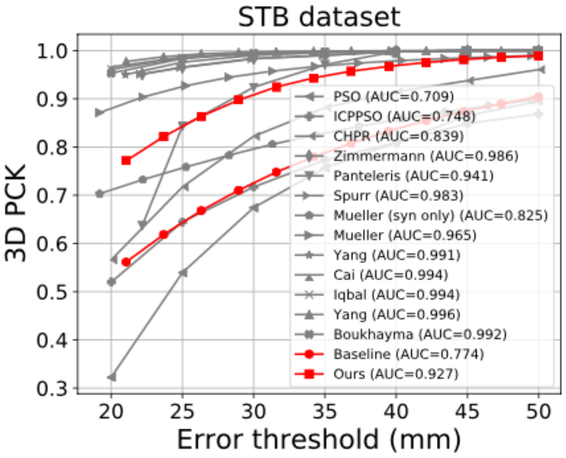


Figure 7: AUC: Comparison to state-of-the-art on STB.

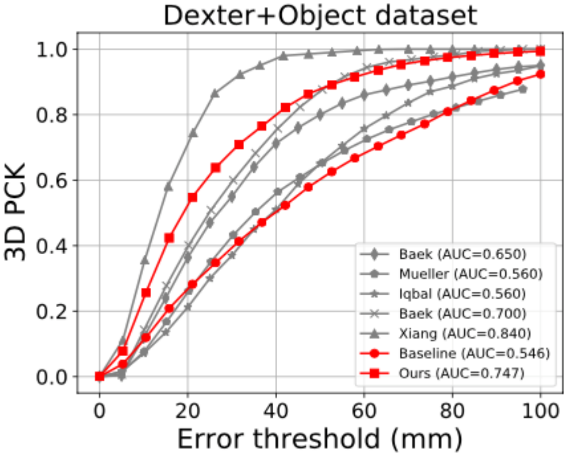
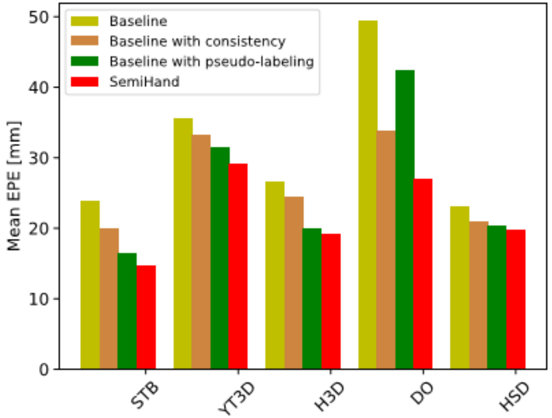


Figure 8: AUC: Comparison to state-of-the-art on DO.



Thanks