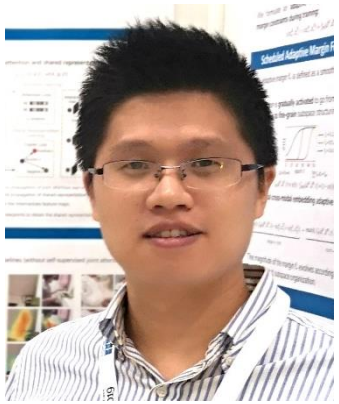
Generalizing Hand Segmentation in Egocentric Videos with Uncertainty-Guided Model Adaptation



Minjie Cai[1], Feng Lu[2], Yoichi Sato[3]

[1]Hunan University, [2]Beihang University, [3]University of Tokyo

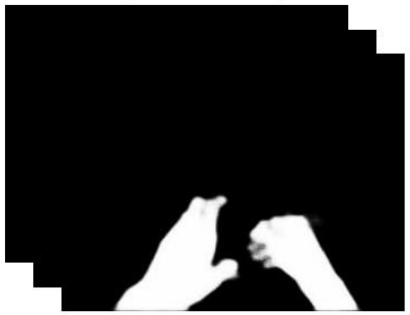
caiminjie@hnu.edu.cn, lufeng@buaa.edu.cn, ysato@iis.u-Tokyo.ac.jp

Problem definition

Goal:

Adapt a hand segmentation model pre-trained on a source domain to a new target domain without labels.

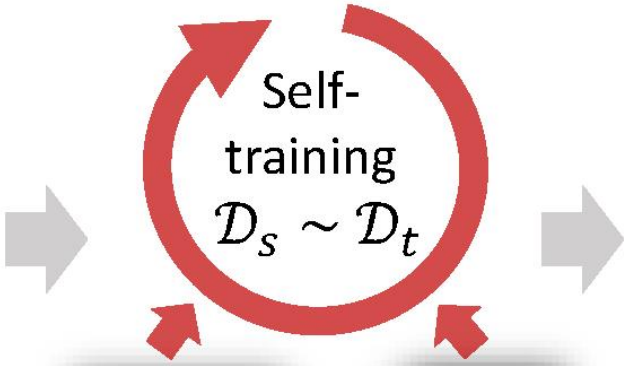
Input images



Hand labels

Source domain

Model adaptation



Model uncertainty

Hand shape prior

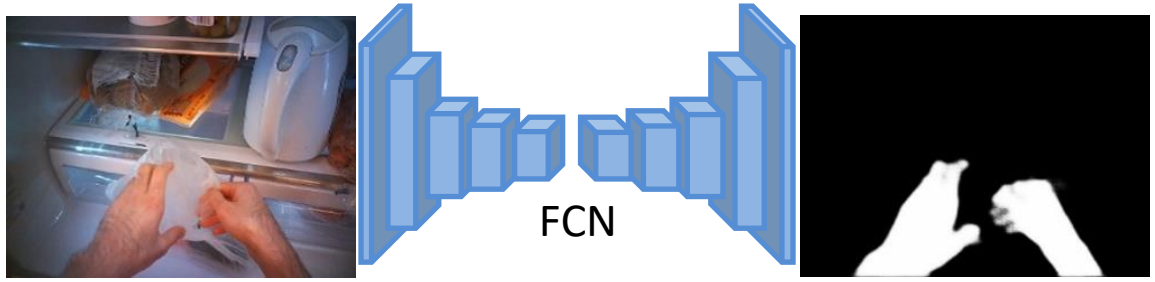
Before adaptation

After adaptation

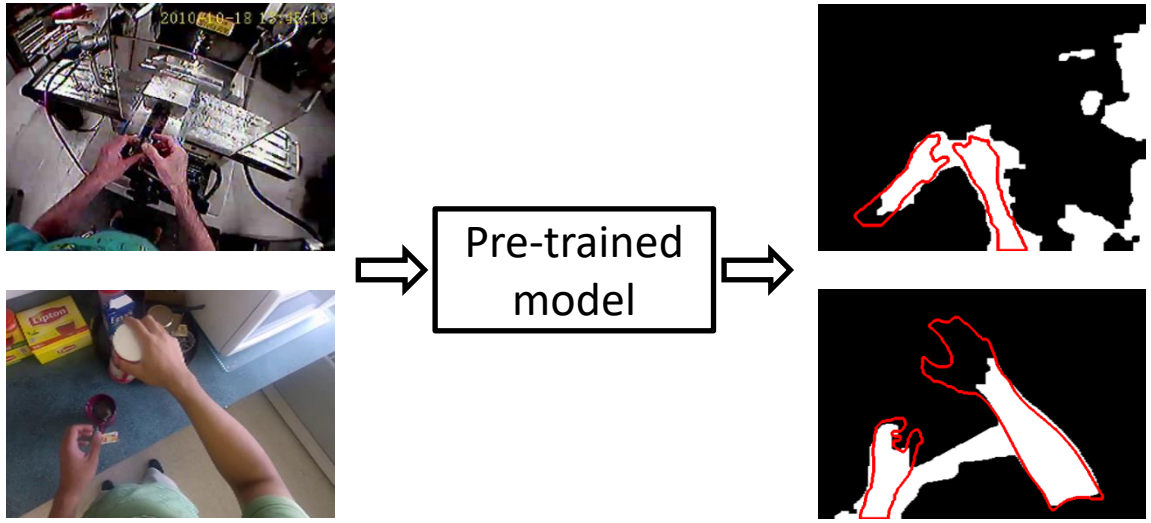
Target domain

Related work – Egocentric hand segmentation

- State-of-the-art performance with Fully Convolutional Network (FCN):



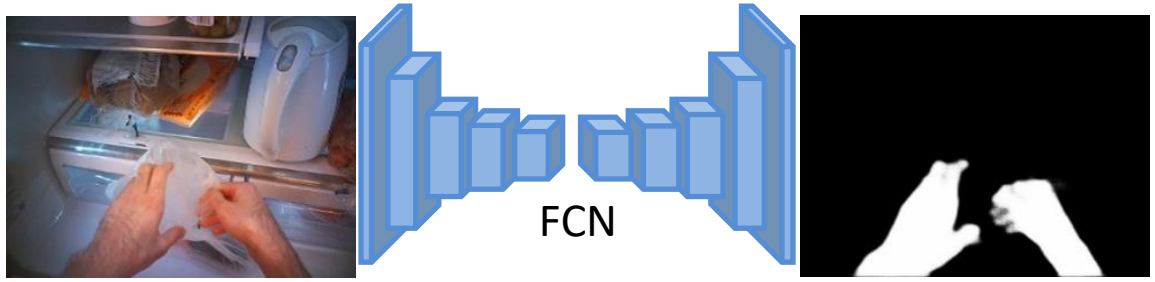
- However, it may fail in a new environment:



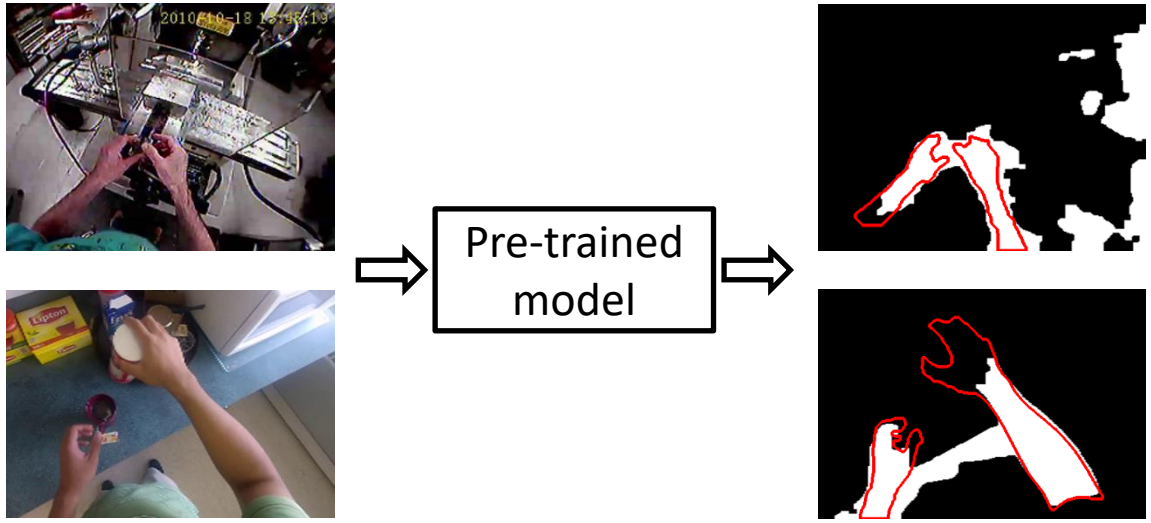
Urooj et al., “Analysis of hand segmentation in the wild”, CVPR2018

Related work – Egocentric hand segmentation

- State-of-the-art performance with Fully Convolutional Network (FCN):



- However, it may fail in a new environment:

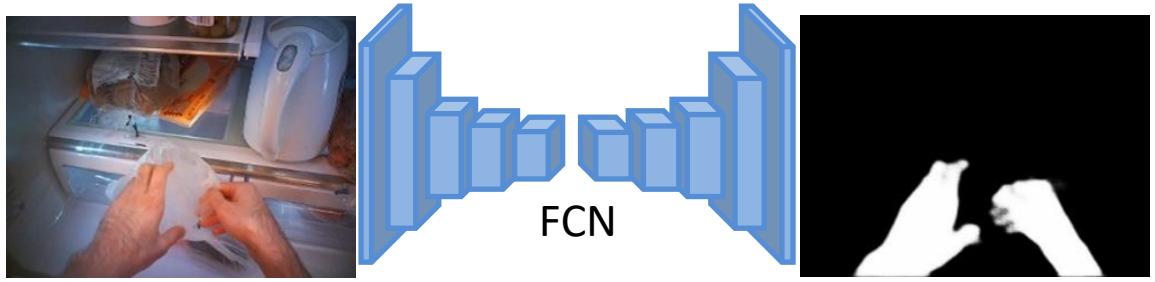


Seen environment

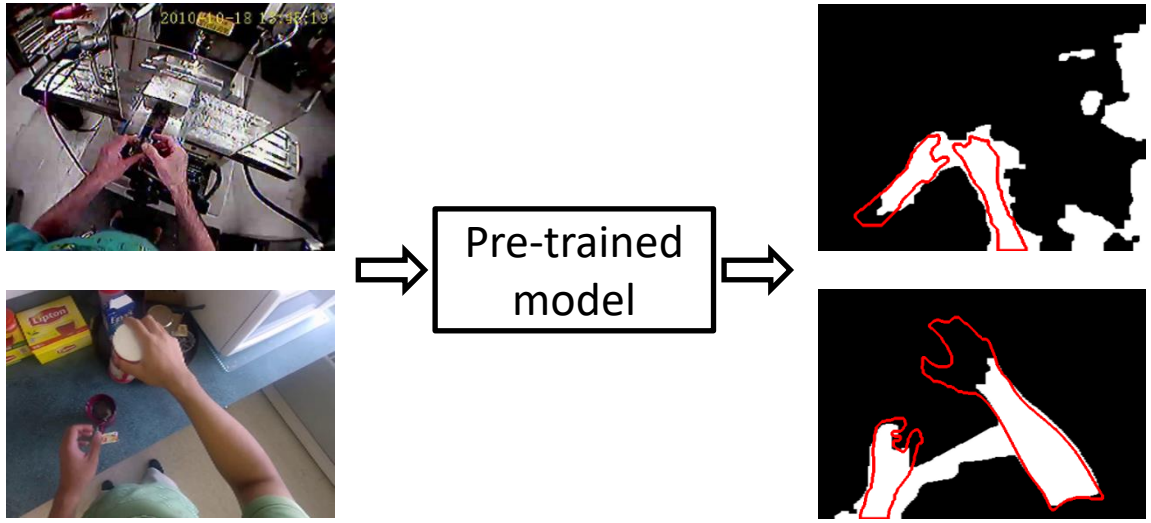
Urooj et al., “Analysis of hand segmentation in the wild”, CVPR2018

Related work – Egocentric hand segmentation

- State-of-the-art performance with Fully Convolutional Network (FCN):



- However, it may fail in a new environment:

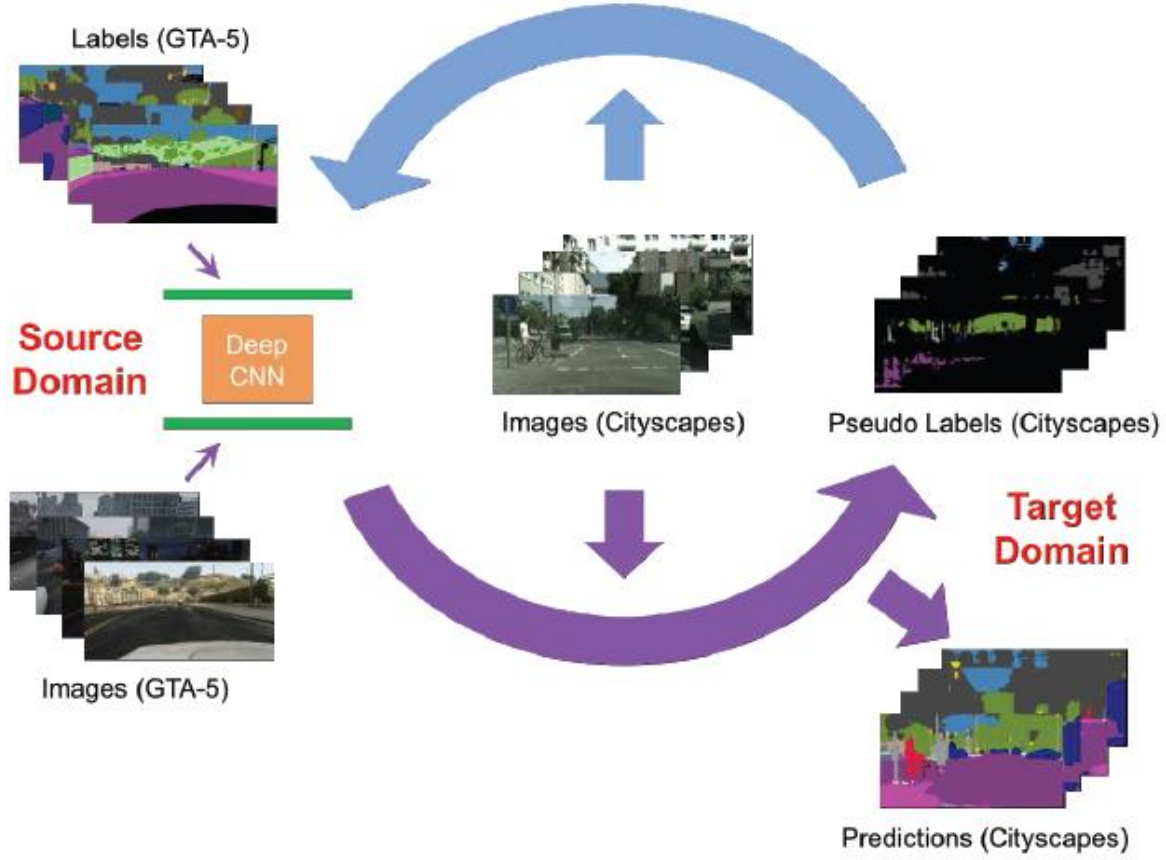


Unseen environment

Urooj et al., “Analysis of hand segmentation in the wild”, CVPR2018

Related work – Unsupervised domain adaptation

- Pseudo-labels based self-training (others include adversarial learning, image translation et al.)



Motivation:

- A unified and efficient framework for domain adaptation.
- Use predictions with high confidence (or low uncertainty) as labels.

Drawback:

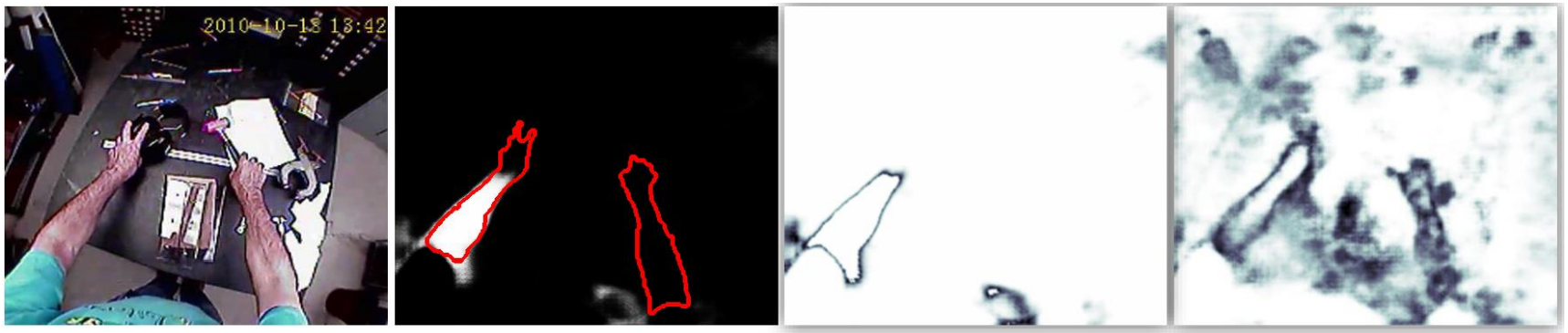
- Existing approaches didn't consider deeply about the real model uncertainty.

Zou et al., "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training", ECCV2018

Zou et al., "Confidence-regularized self-training", ICCV2019

Uncertainty

Ideal uncertainty for domain adaptation should reflect the model's "real" confidence/uncertainty about its predictions.



(a) Target image

(b) Normalized prediction score

(c) Uncertainty based on prediction score

(d) Uncertainty based on Bayesian CNN

Uncertainty estimation from Bayesian CNN

Approximate Bayesian inference:

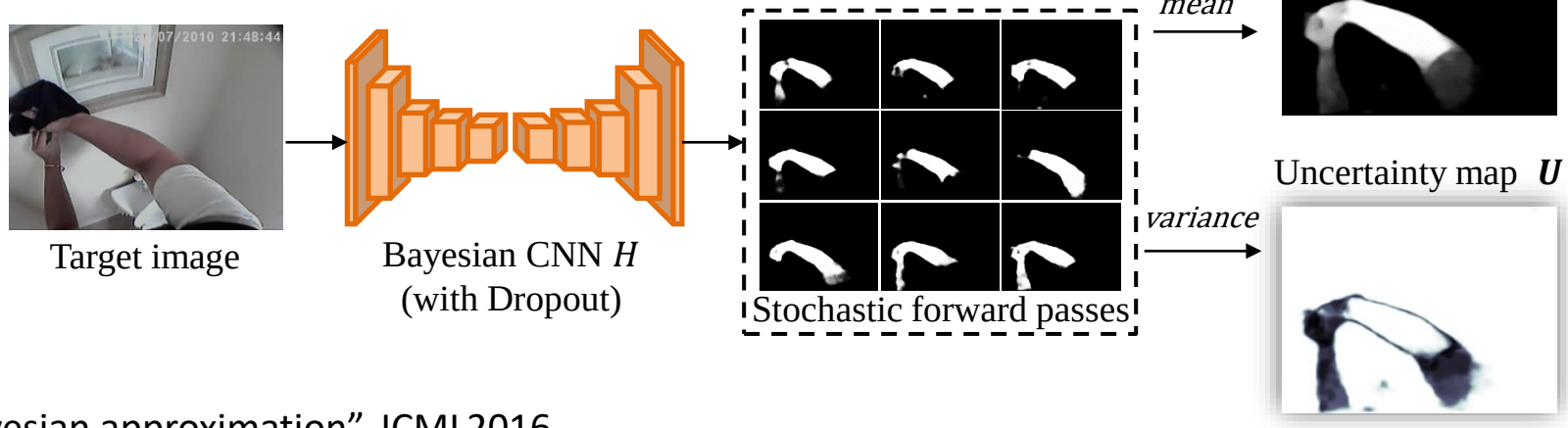
$$p(y|x) = \int p(y|x,w)q(w)\,dw$$

$$\approx \frac{1}{T}\sum_{i=1}^{T} p(y|x,w_i), \quad w_i \sim q(w)$$

Uncertainty estimation from Bayesian CNN (with Dropout):

$$\bar{\mathbf{P}} = \frac{1}{T}\sum_{i=1}^{T} H(\mathbf{I}, w_i), \quad w_i \sim dropout(w)$$

$$\mathbf{U} = \frac{1}{T}\sum_{i=1}^{T} \mathbf{P}_i^2 - \bar{\mathbf{P}}^2$$

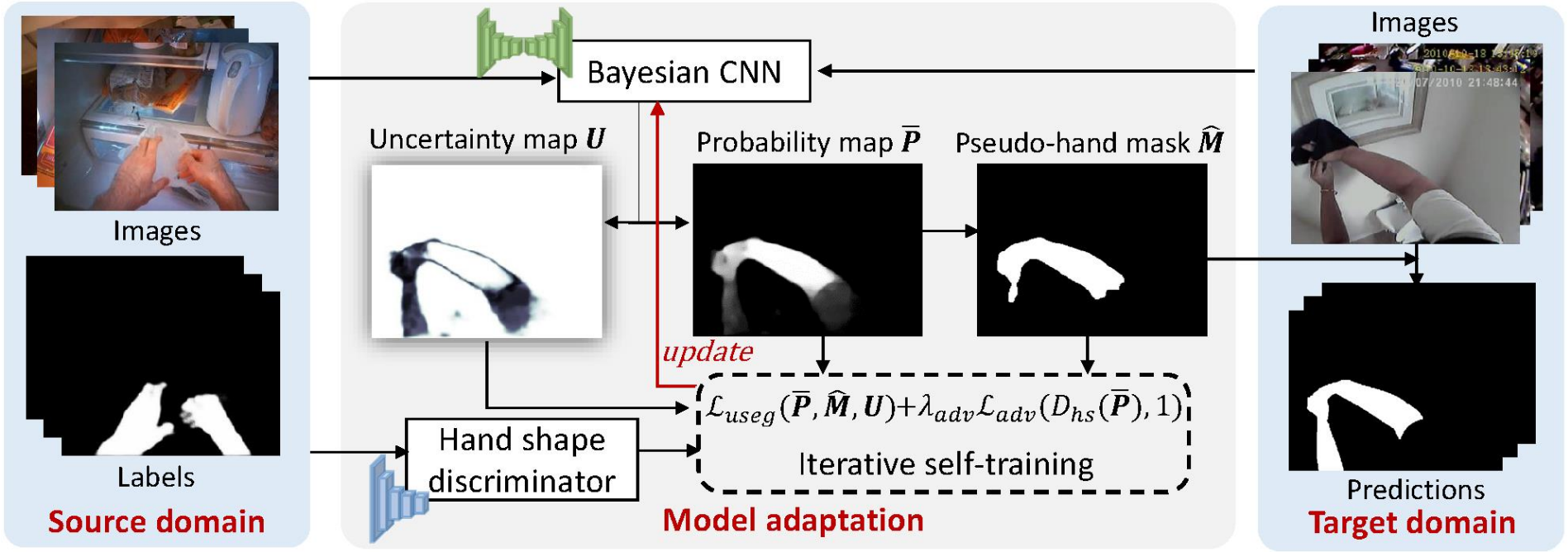


Gal et al., “Dropout as a Bayesian approximation”, ICML2016

Method overview

Key idea:

- Use uncertainty to guide self-training with pseudo-labels in the target domain.
- Use pre-trained hand discriminator to enforce hand shape consistency.



$$\mathcal{L}_{useg}(\bar{\mathbf{P}}, \hat{\mathbf{M}}, \mathbf{U}) = -\frac{1}{M}\sum_{m=1}^{M}(1-\mathrm{U}_m)\big(\hat{\mathrm{M}}_m \log \bar{\mathrm{P}}_m + (1-\hat{\mathrm{M}}_m)\log(1-\bar{\mathrm{P}}_m)\big)$$

Datasets and experimental setting

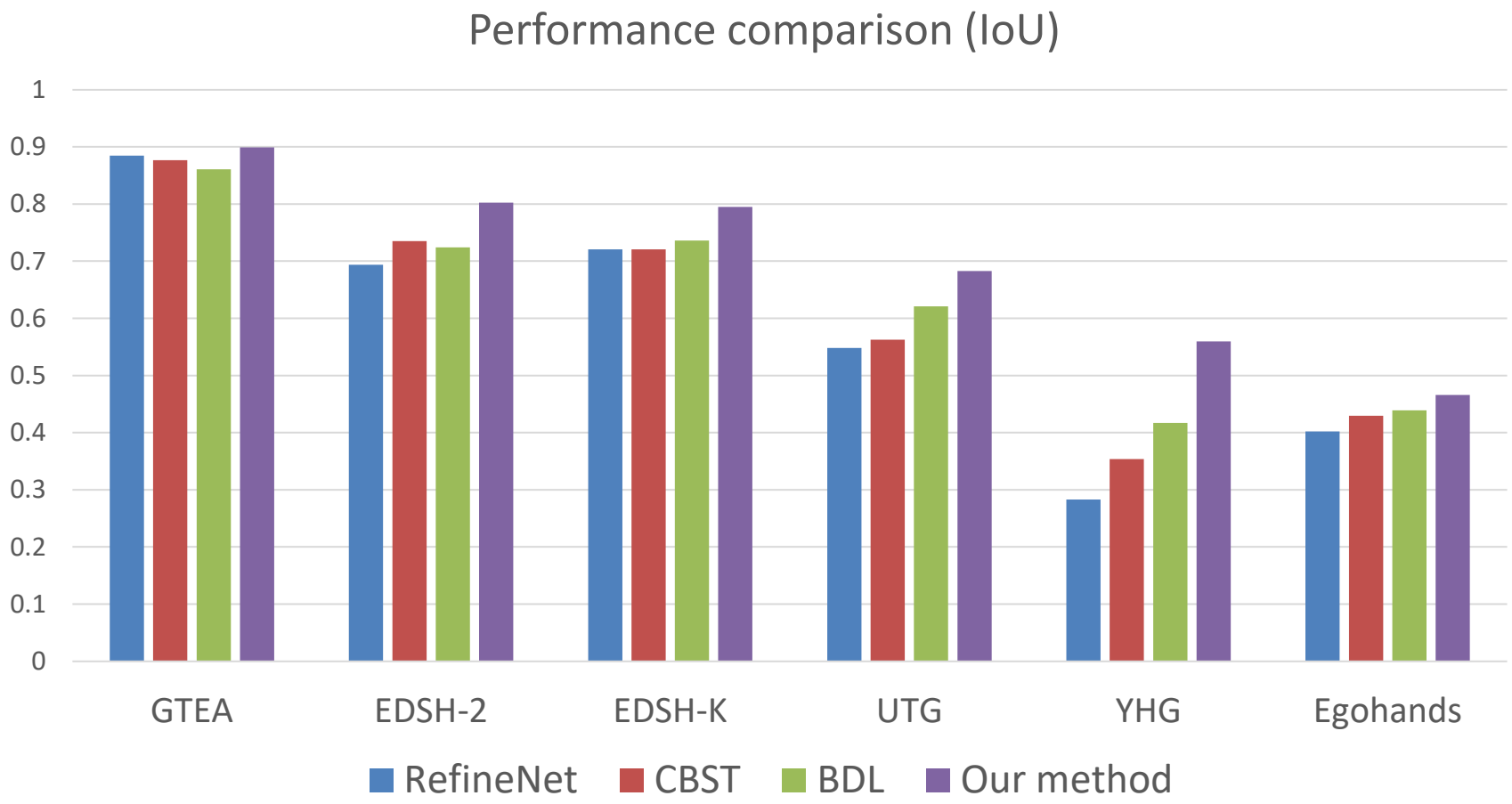
Train/test split:

- EGTEA is used as training data (source domain)
- The rest datasets are used as test data (target domain)

Evaluation metric:

- Intersection over Union (IoU)

Comparison with state-of-the-art



RefineNet: Analysis of hand segmentation in the wild, CVPR2018

CBST: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training, ECCV2018

BDL: Bidirectional learning for domain adaptation of semantic segmentation, CVPR2019

Ablation study

- Upper bound performance: Fine-tuning CNN with target labels

	GTEA	EDSH-2	EDSH-K	UTG	YHG	Egohands
Finetuning	0.9254	0.8448	0.7802	0.8495	0.8224	0.8463

- Performance of different components

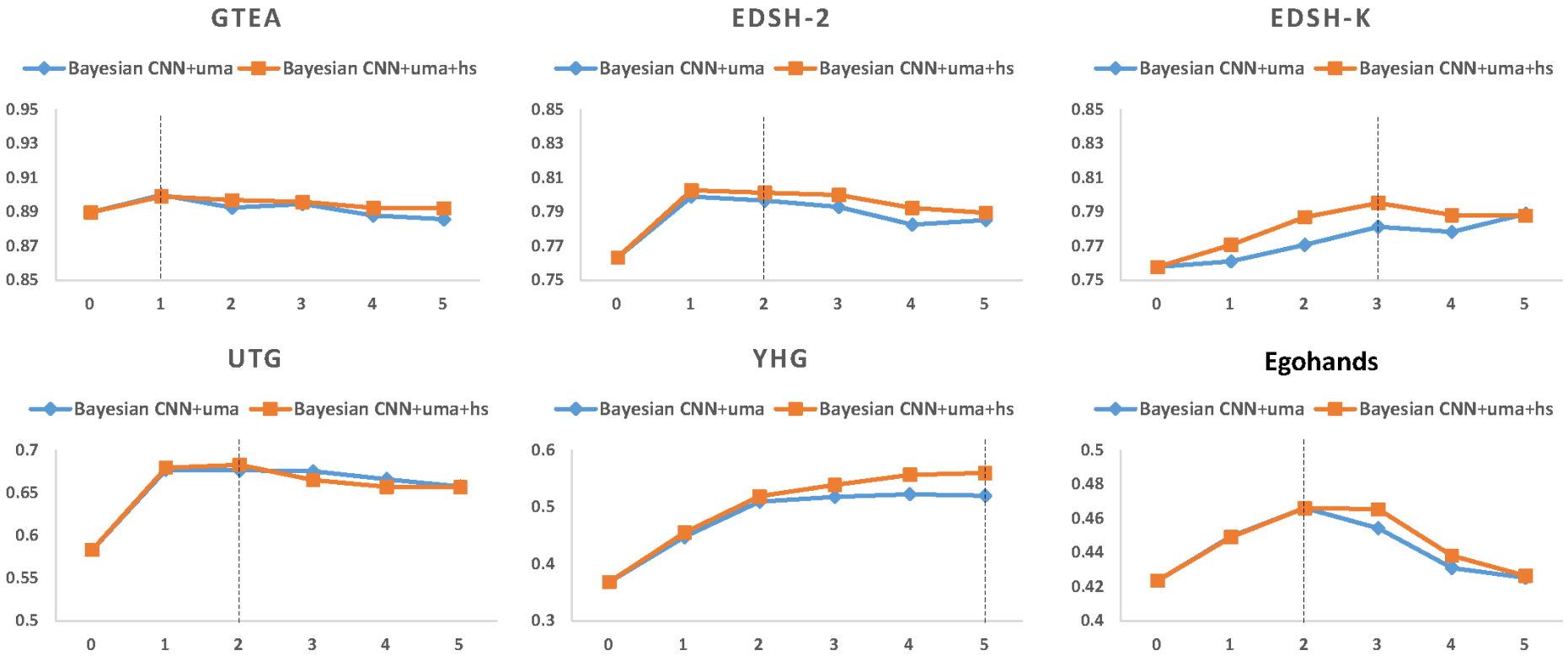
Method	GTEA		EDSH-2		EDSH-K		UTG		YHG		Egohands	
	mIoU	△mIoU	mIoU	△mIoU	mIoU	△mIoU	mIoU	△mIoU	mIoU	△mIoU	mIoU	△mIoU
CNN	0.8845	-0.0409	0.6936	-0.1512	0.7205	-0.0594	0.5481	-0.3014	0.2831	-0.5393	0.4019	-0.4444
CNN+uma	0.8766	-0.0488	0.7141	-0.1307	0.7723	-0.0079	0.6089	-0.2406	0.3159	-0.5065	0.4252	-0.4211
Bayesian CNN	0.8896	-0.0358	0.7632	-0.0816	0.7576	-0.0226	0.5832	-0.2663	0.3619	-0.4605	0.4235	-0.4228
Bayesian CNN+uma	0.8945	-0.0300	0.7965	-0.0483	0.7812	+0.0010	0.6762	-0.1733	0.5223	-0.3001	**0.4665**	**-0.3798**
Bayesian CNN+uma+hs	**0.8990**	**-0.0255**	**0.8025**	**-0.0423**	**0.7951**	**+0.0149**	**0.6827**	**-0.1668**	**0.5596**	**-0.2628**	0.4660	-0.3803

Δ Shows the gap between performance of fine-tuning.

- uma: uncertainty-guided model adaptation.
- hs: hand shape constraint
- Bayesian CNN+uma+hs: our full model

Model convergence

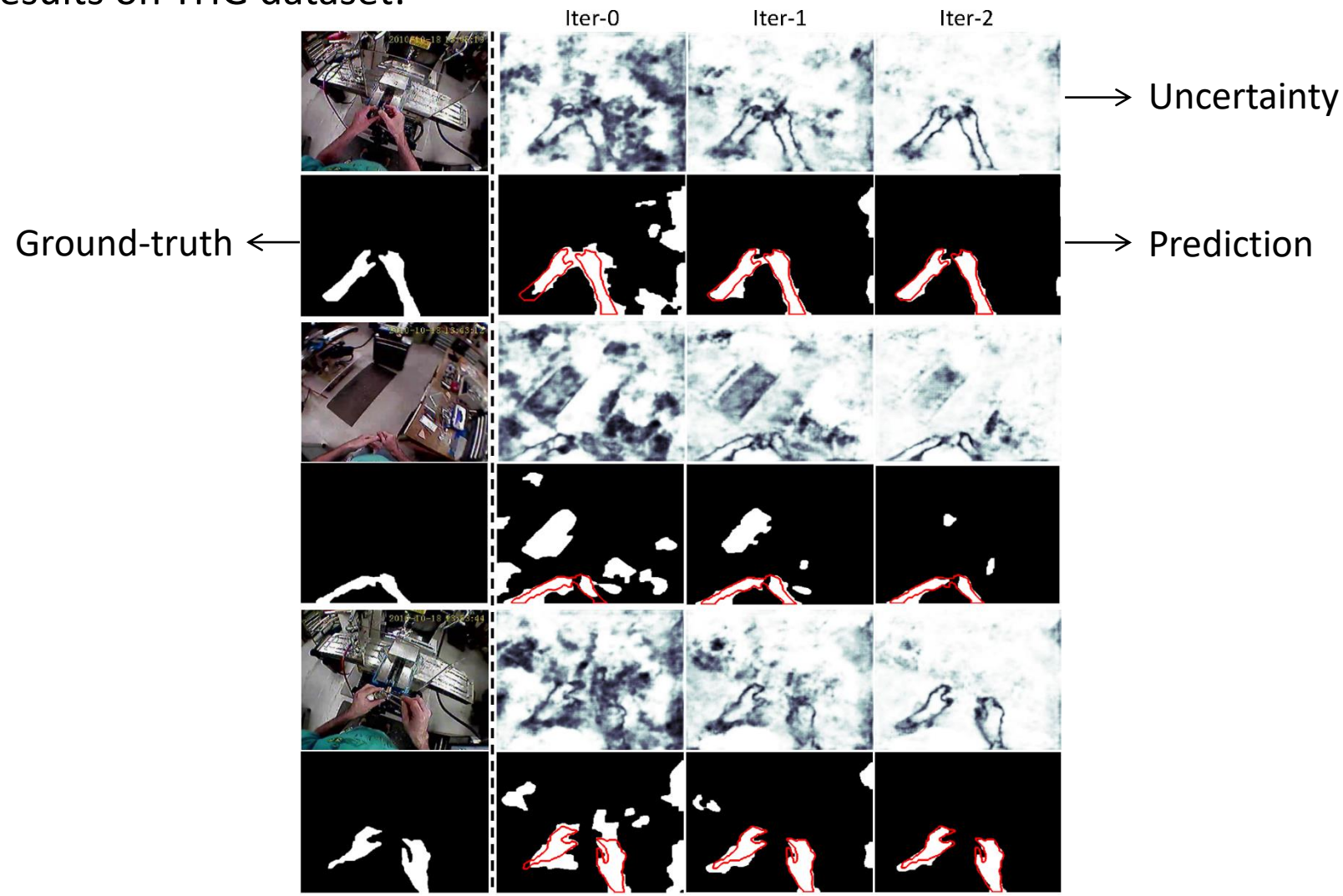
- Our method converges quickly to the target domain within a few iterations.



Performance variation with number of iterations

Qualitative results

Results on YHG dataset:



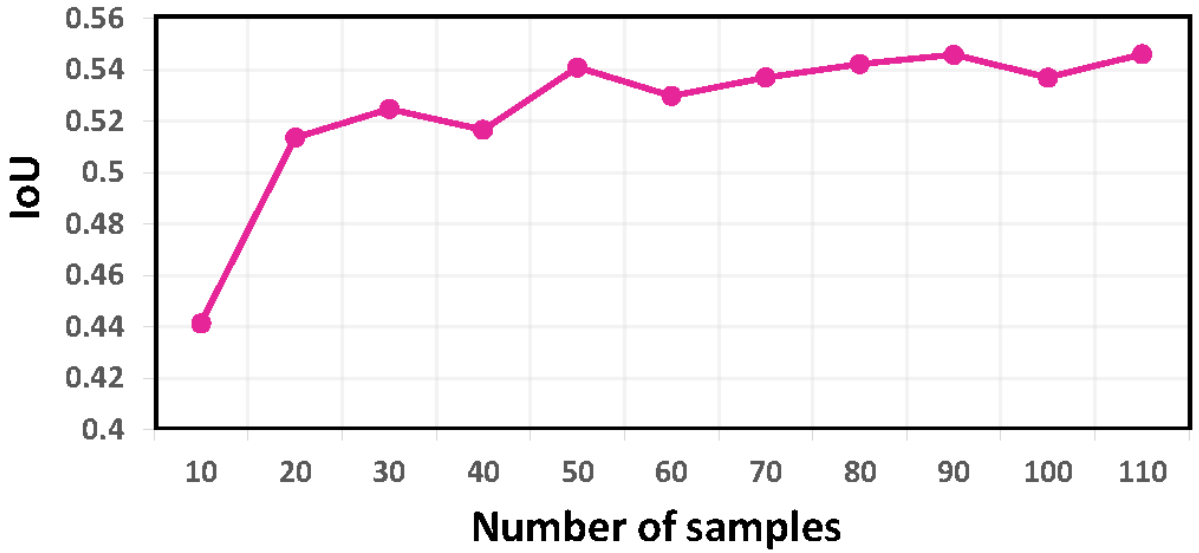
Qualitative comparison (YHG dataset)

Before adaptation

After adaptation

Online evaluation

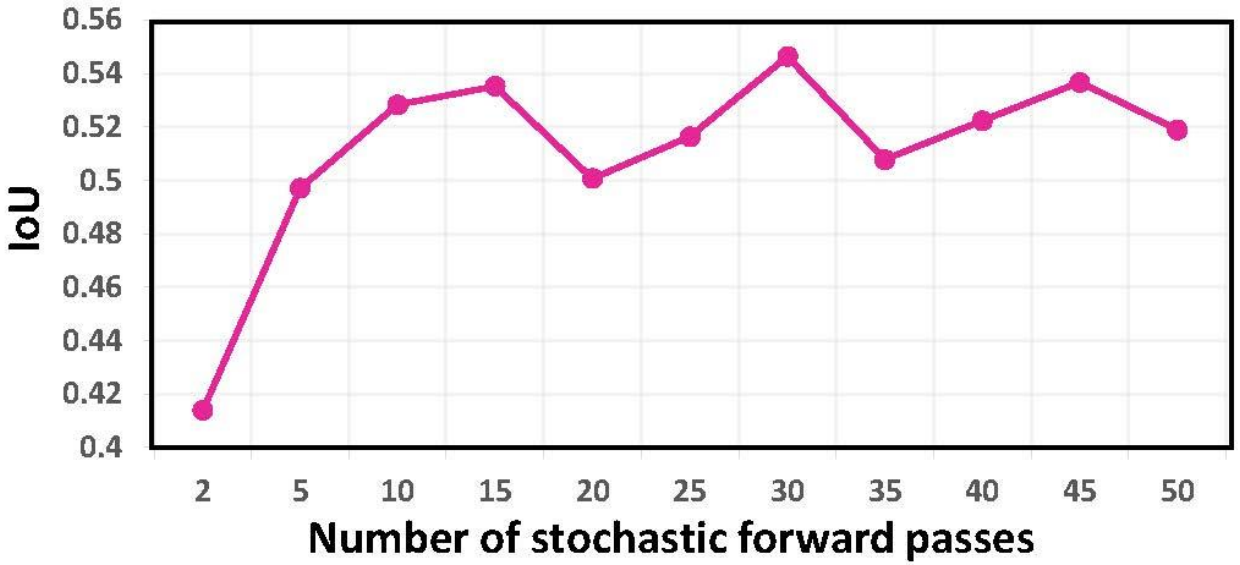
- Our method adapts quickly to the target domain with a few unlabeled samples.



Performance variation with number of target samples on YHG dataset

Existing risk: fluctuating performance

- The adaptation performance fluctuates with different number of sampling.



Performance variation with number of stochastic forward passes on YHG dataset

Conclusions

- We propose a novel and efficient approach for generalizing hand segmentation.
- The key idea is uncertainty-guided model adaptation, which can be extended to various domain adaptation tasks.
- The impact and mechanism of different sampling strategies for uncertainty estimation needs further study (to be done).

Code: https://github.com/cai-mj/UMA