ScarceNet: Animal Pose Estimation with Scarce Annotations

Chen Li, Gim Hee Lee

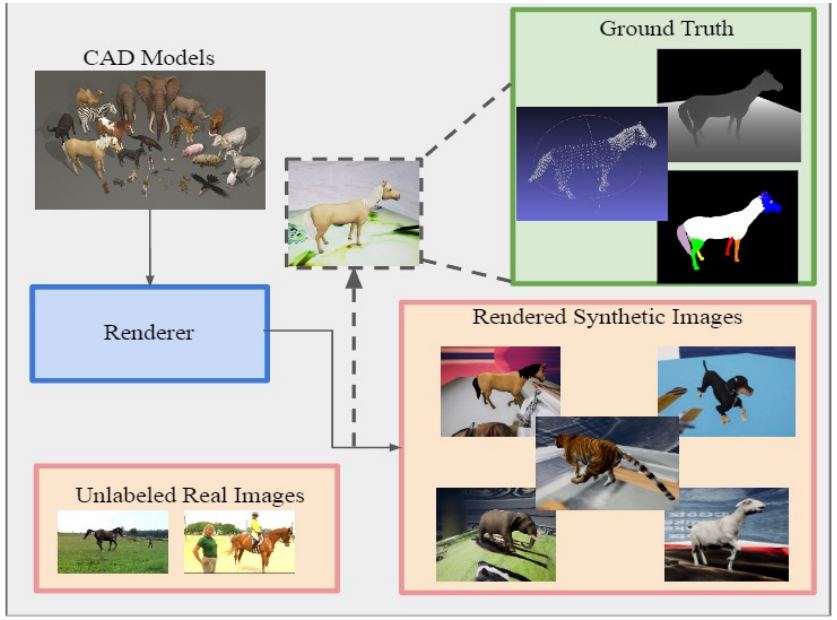
National University of Singapore

Poster Session: THU-AM-066

Animal pose estimation: Challenge

- Objective : estimate the semantic keypoints of animals

- Challenge: lack of animal pose data

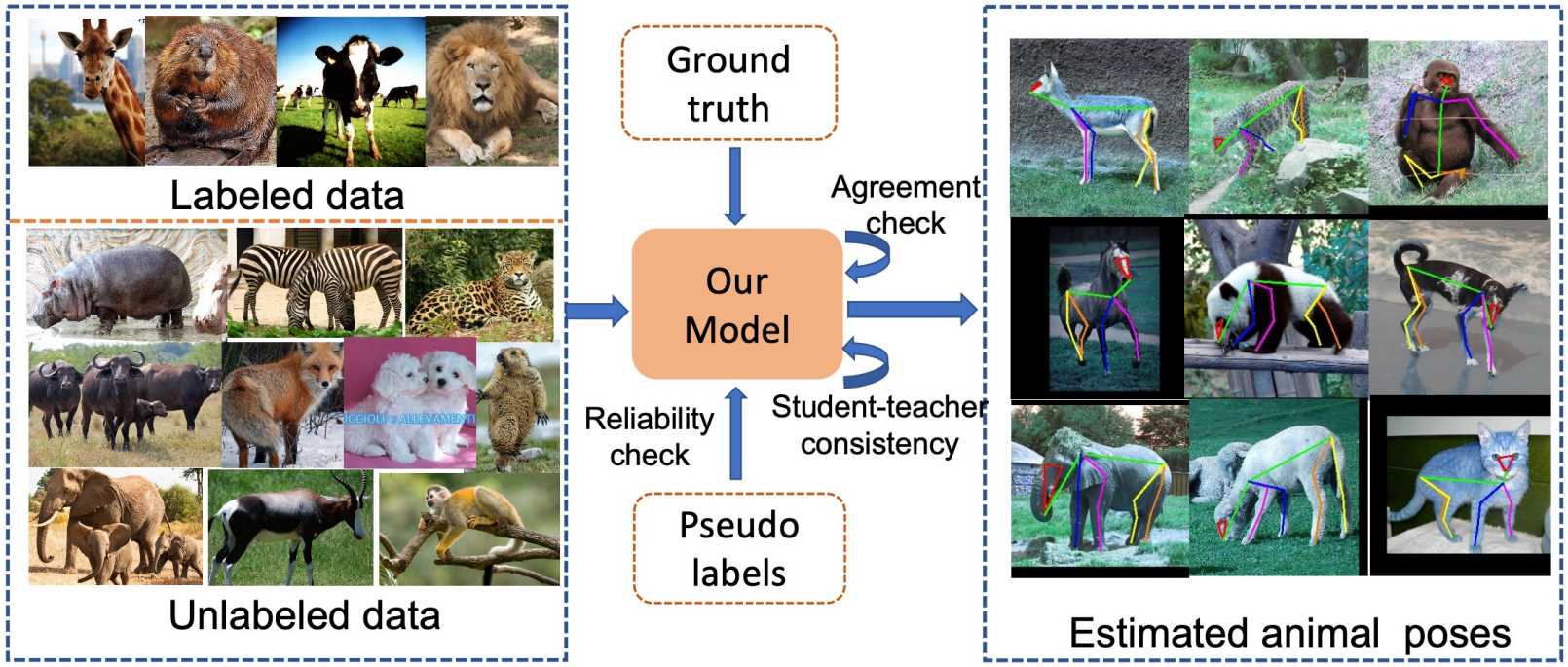


✗ Synthesize images is tedious

✗ large domain gap

Animal pose estimation: goal

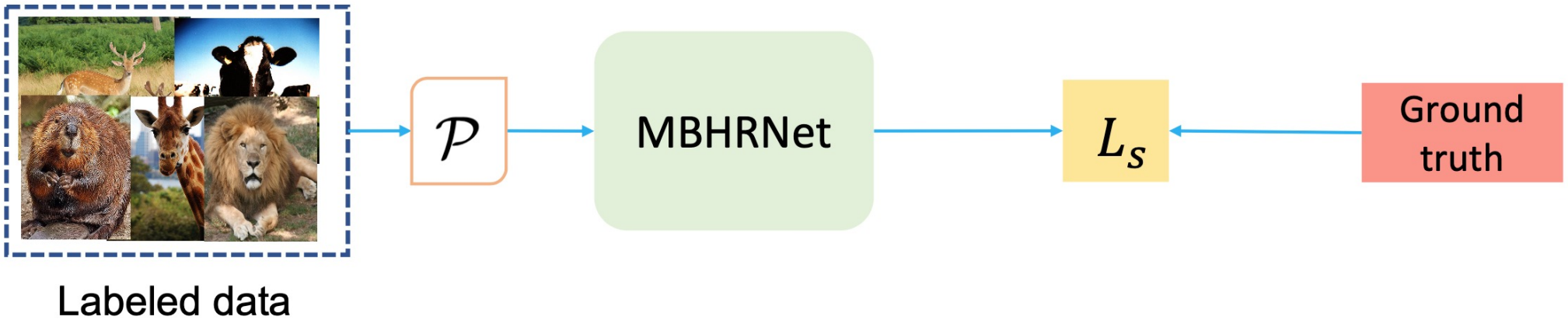
- Goal: achieve accurate animal pose estimation with minimal effort for annotating



A pseudo label based approach for semi-supervised animal pose estimation

Our method : ScarceNet

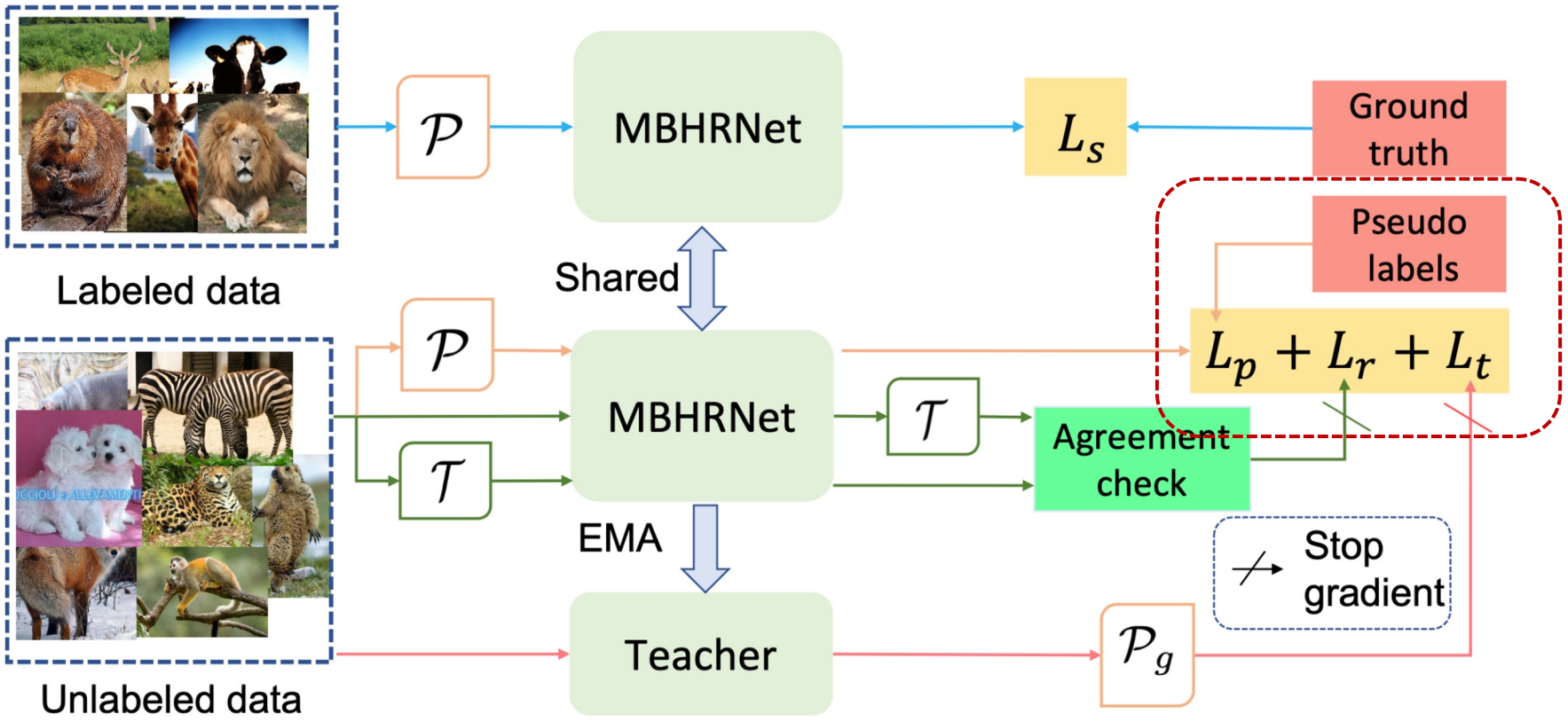
- Supervised pretrain: train a pose estimation network with $\mathcal{D}_l = \{\mathbf{x}_i^l, \mathbf{y}_i^l\}_{i=1}^{N_l}$



- Pseudo labeling: generate pseudo labels for the unlabeled data $\mathcal{D}_u = \{\mathbf{x}_i^u\}_{i=1}^{N_u}$
- Problem: the generated pseudo label is noisy, leading to degraded performance when applied directly.

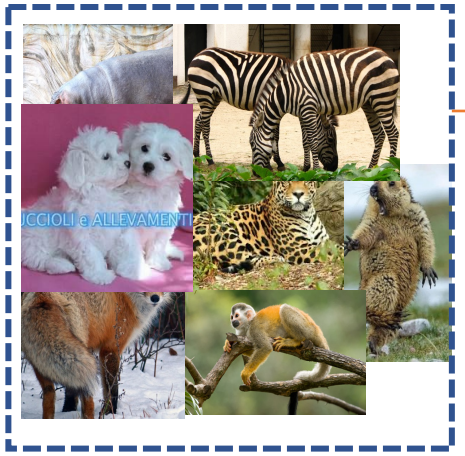
Our method : ScarceNet

Mitigate the negative effect of noisy pseudo label



Our method : ScarceNet

Mitigate the effect of noisy pseudo label: reliable pseudo label selection



Unlabeled data

$\mathcal{P}$

MBHRNet

Pseudo labels

$L_p + L_r + L_t$

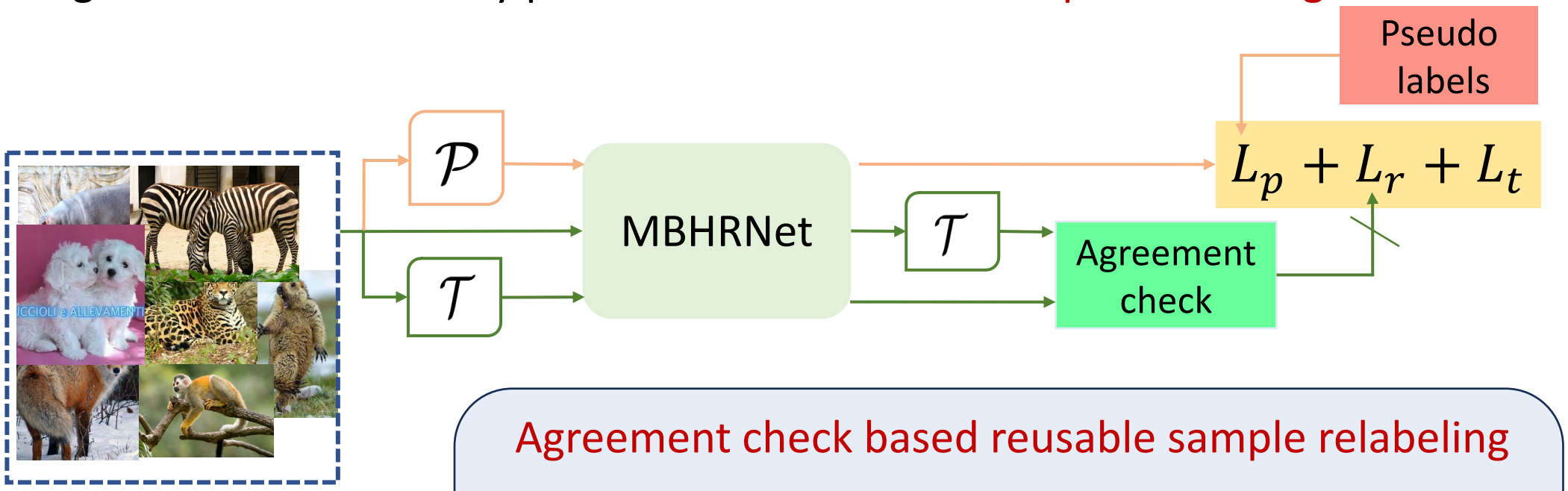
Small loss based pseudo label selection

$$\mathcal{L}_{\mathrm{p}} = \frac{\sum_{i=1}^{J} \{\mathcal{L}_{\mathrm{p}}^{i} < l_{\mathrm{r}}\}_1 \mathcal{L}_{\mathrm{p}}^{i}}{\sum_{i=1}^{J} \{\mathcal{L}_{\mathrm{p}}^{i} < l_{\mathrm{r}}\}_1}, \text{ where } \mathcal{L}_{\mathrm{p}}^{i} = \|\tilde{H}_i - \hat{H}_i\|^2$$

Samples with high loss are discarded!

Our method : ScarceNet

Mitigate the effect of noisy pseudo label: Reusable sample relabeling



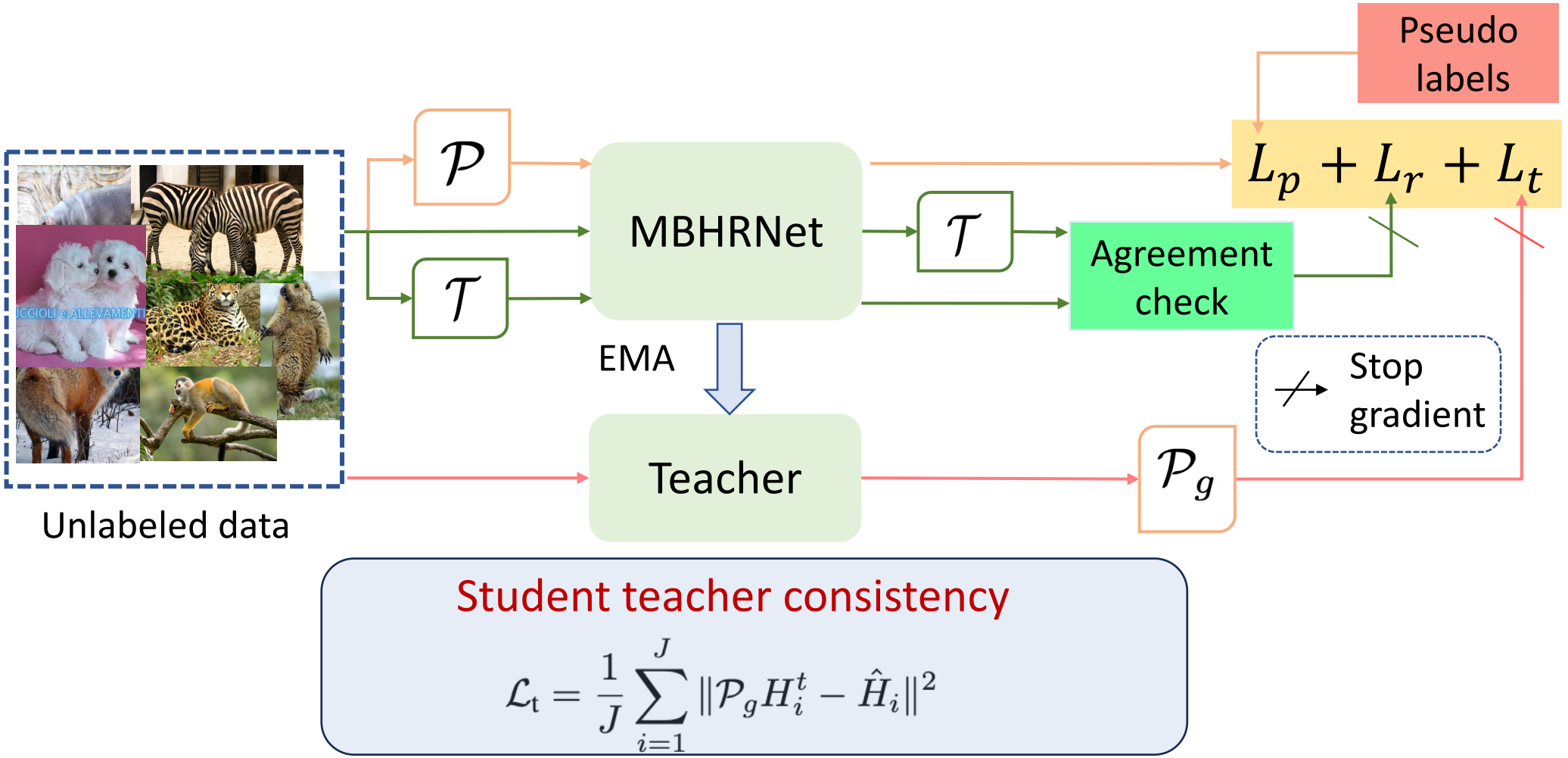
Agreement check based reusable sample relabeling

$$d_i = \|\mathcal{T}\mathbf{h}_i^{\mathrm{v}1} - \mathbf{h}_i^{\mathrm{v}2}\|^2$$

$$\mathcal{L}_{\mathrm{r}} = \frac{\sum_{i=1}^{J}\{d^i < d_{\mathrm{r}}\}_1 \mathcal{L}_{\mathrm{r}}^i}{\sum_{i=1}^{J}\{d^i < d_{\mathrm{r}}\}_1}, \text{ where } \mathcal{L}_{\mathrm{r}}^i = \|\bar{H}_i - \hat{H}_i\|^2$$

Our method : ScarceNet

Mitigate the effect of noisy pseudo label: student teacher consistency

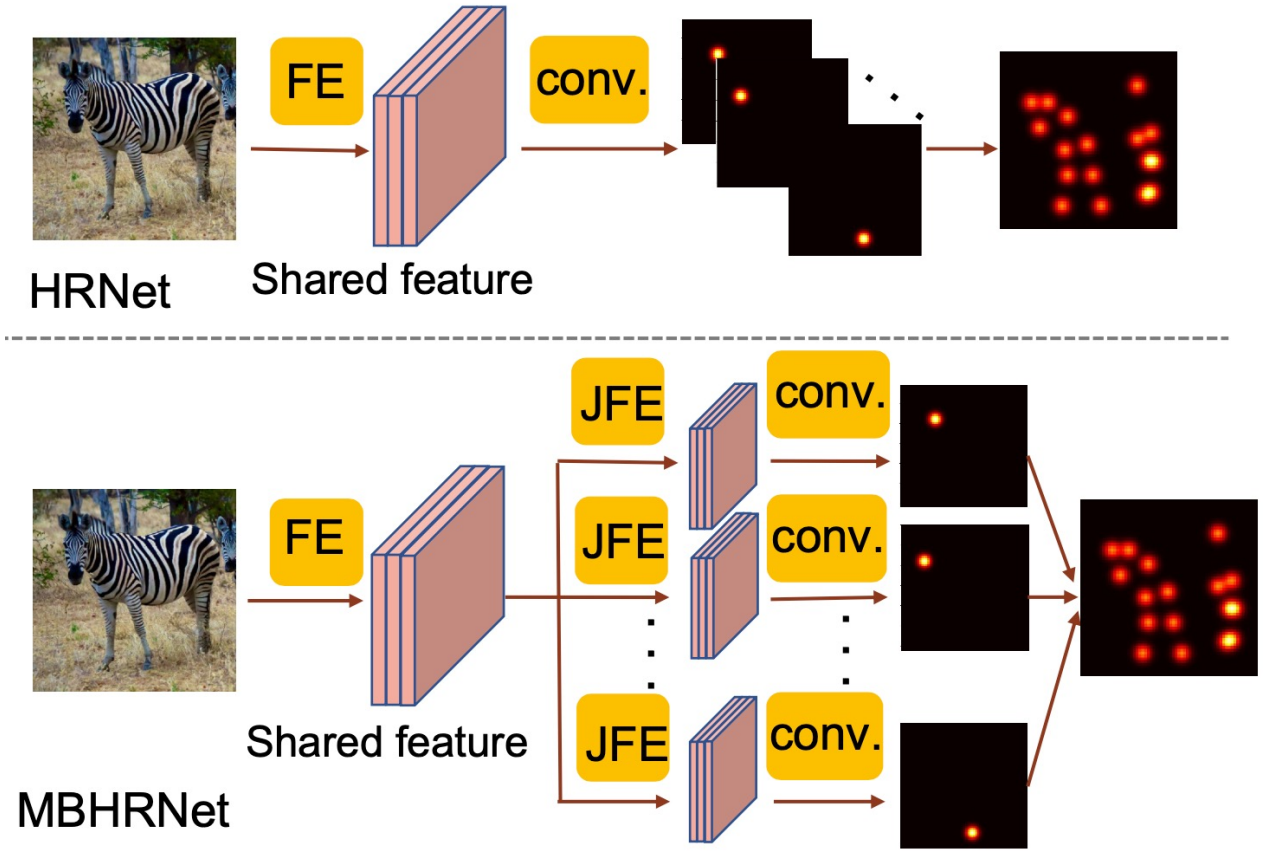


Student teacher consistency

$$\mathcal{L}_{\mathrm{t}} = \frac{1}{J}\sum_{i=1}^{J} \|\mathcal{P}_g H_i^t - \hat{H}_i\|^2$$

Our method : ScarceNet

Mitigate the effect of noisy pseudo label: multi-branch HRNet (MBHRNet)



Learn joint-specific feature to avoid the negative effect of noisy joints to others.

Experiments: quantitative results

- Comparison with existing semi-supervised approach

	5	10	15	20	25
HRNet [27]	0.360	0.463	0.511	0.547	0.588
UDA [30]	0.429	0.519	0.566	0.580	0.628
FixMatch [25]	0.478	0.544	0.589	0.601	0.631
FlexMatch [34]	0.466	0.555	0.596	0.618	0.646
Ours	**0.533**	**0.597**	**0.632**	**0.654**	**0.681**

- Comparison with domain-adaption based approach

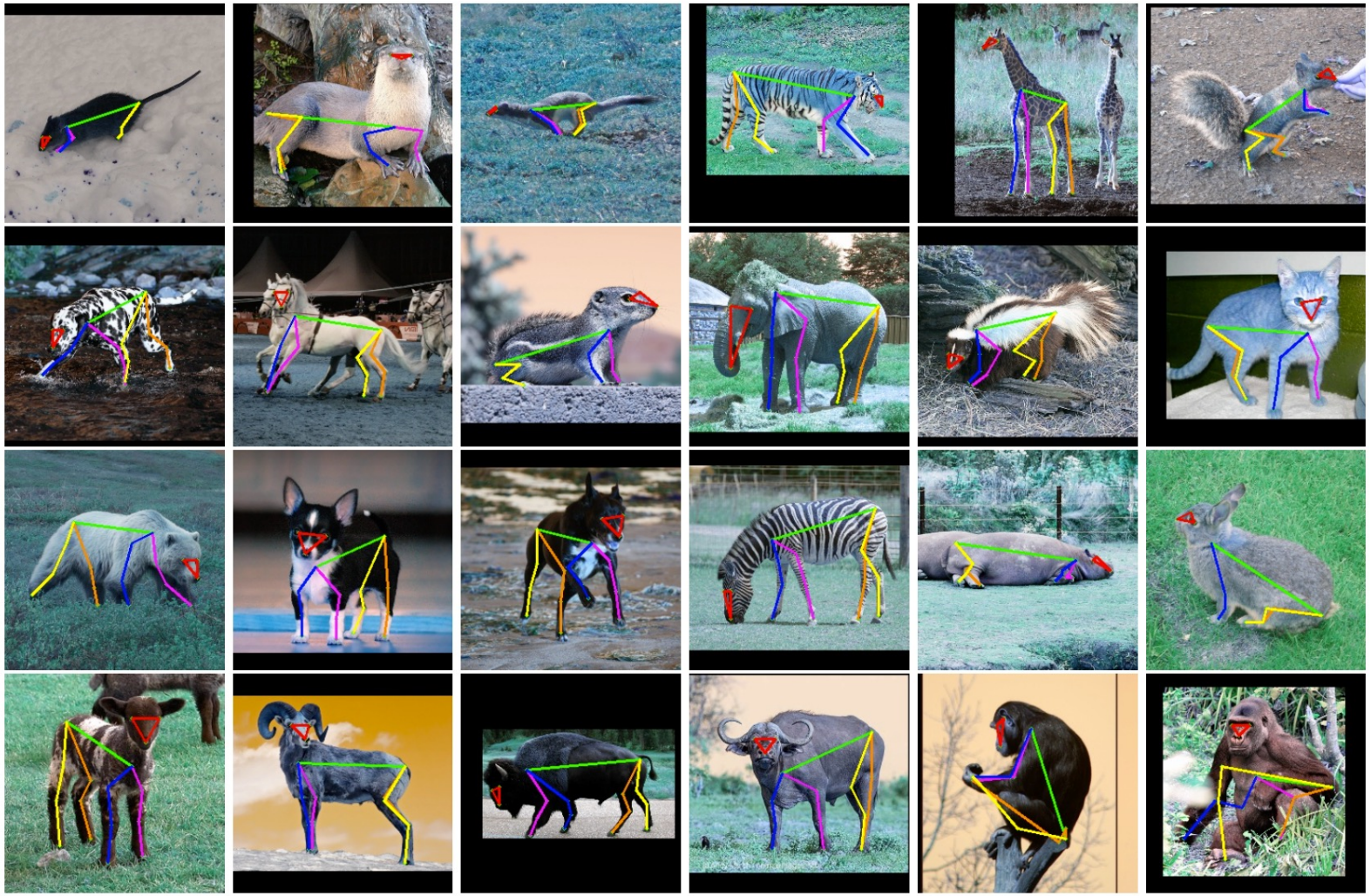
	Real	Cycgan [36]	BDL [17]	Cycada [10]	CC-SSL [20]	MDAM-MT [16]	Ours
Horse	78.98	51.86	62.33	55.57	70.77	**79.50**	73.05
Tiger	81.99	46.47	52.26	51.48	64.14	67.76	**74.88**
Average	80.48	49.17	57.30	53.53	67.52	73.66	**73.83**

Experiments: Ablation study

- Ablation study on the AP-10K dataset when 5, 10, 15, 20 and 25 images per animal species are labeled.

	5	10	15	20	25
Full	0.533	0.597	0.632	0.655	0.681
- RSR	0.521	0.586	0.621	0.633	0.668
- MT	0.487	0.568	0.614	0.628	0.659
- MB	0.520	0.585	0.624	0.645	0.670
- AUG	0.487	0.565	0.608	0.625	0.654

Experiments: qualitative results



Thank you !

ScarceNet: Animal Pose Estimation with Scarce Annotations

Chen Li Gim Hee Lee
Department of Computer Science, National University of Singapore
lichen@u.nus.edu gimhee.lee@comp.nus.edu.sg

Abstract

Animal pose estimation is an important but under-explored task due to the lack of labeled data. In this paper, we tackle the task of animal pose estimation with scarce annotations, where only a small set of labeled data and unlabeled images are available. At the core of the solution to this problem setting is the use of the unlabeled data to compensate for the lack of well-labeled animal pose data. To this end, we propose the ScarceNet, a pseudo label-based approach to generate artificial labels for the unlabeled images. The pseudo labels, which are generated with a model trained with the small set of labeled images, are generally noisy and can hurt the performance when directly used for training. To solve this problem, we first use a small-loss trick to select reliable pseudo labels. Although effective, the selection process is improvident since numerous high-loss samples are left unused. We further propose to identify reusable samples from the high-loss samples based on an agreement check. Pseudo labels are re-generated to provide supervision for those reusable samples. Lastly, we introduce a student-teacher framework to enforce a consistency constraint since there are still samples that are neither reliable nor reusable. By combining the reliable pseudo label selection with the reusable sample re-labeling and the consistency constraint, we can make full use of the unlabeled data. We evaluate our approach on the challenging AP-10K dataset, where our approach outperforms existing semi-supervised approaches by a large margin. We also test on the TigDog dataset, where our approach can achieve better performance than domain adaptation based approaches when only very few annotations are available. Our code is available at the project website [1].

1. Introduction

The ability to understand animal behavior is fundamental to many applications, such as farming, ecology and surveillance. Animal pose estimation is a key step to understand animal behavior and has attracted increasing attention in recent years [5, 16, 20, 23, 33]. Although great success is achieved for human pose estimation with the development of sophisticated deep learning models, these techniques cannot be directly used for animals due to the lack of labeled animal pose data. Existing works overcome this problem by learning from human pose data [5] or synthetic animal images [16, 20, 23]. However, there is a large domain gap between the real and synthetic (human) data. For example, the synthetic animal images in [20], which are generated from CAD models, only exhibit limited pose, appearance and background variations. As a result, the model trained with the synthetic data may not adapt well to real images, especially for images with crowded scene or self-occlusion. Moreover, the generation of synthetic images is a tedious process that also requires expert knowledge. The above-mentioned problems lead us to ask whether we can achieve accurate animal pose estimation with minimal effort for annotating? To answer this question, we focus on how to learn from scarce labeled data for animal pose estimation. As shown in Fig. 1, we aim to achieve accurate animal pose estimation with only a small set of labeled images and unlabeled images.

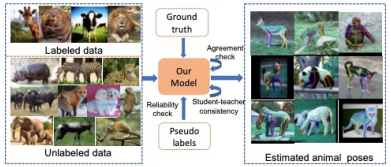


Figure 1. An illustration of our task. We aim to achieve accurate animal pose estimation with a small set of labeled images as well as unlabeled images.

The data scarcity problem is solved by semi-supervised learning (SSL) for the classification task [3, 25, 28, 34]. One powerful class of SSL is pseudo labeling (PL), where artificial labels generated from a pretrained model are used together with labeled data to train a model. Impressive

[1] https://github.com/chaneyddtt/ScarceNet