JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Architecture, Dataset and Model-Scale Agnostic Data-free Meta-Learning

Zixuan Hu[1], Li Shen[2], Zhenyi Wang[3], Tongliang Liu[4], Chun Yuan[1], Dacheng Tao[2]

1Tsinghua Shenzhen International Graduate School, China; 2 JD Explore Academy, China
3 State University of New York at Buffalo, USA; 4 The University of Sydney, Australia

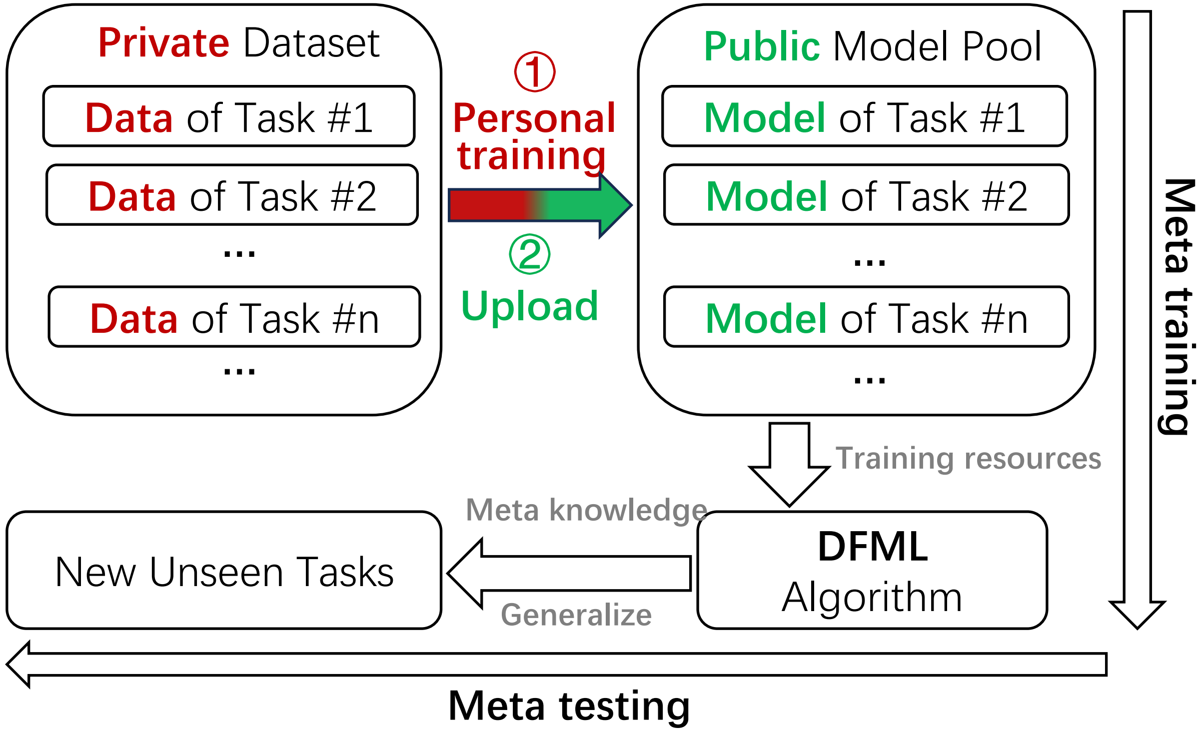
Poster: TUE-PM-345

清華大學 Tsinghua University 京东探索研究院 JD EXPLORE ACADEMY University at Buffalo The State University of New York THE UNIVERSITY OF SYDNEY

Overview

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem definition of data-free meta-learning (DFML):



Our contributions:

- Propose a united framework **PURER**, which is architecture, dataset and model-scale agnostic.
- Propose two techniques:
  - **Episode Curriculum Inversion** → Efficient meta training with pseudo tasks of adaptively increasing difficulty level
  - **Inversion Calibration following Inner Loop** → Alleviate task-distribution shift
- Evaluate DFML on three real-world scenarios with superior gains:
  - **SS**: modes pre-trained on **S**ame dataset with **S**ame architecture
  - **SH**: modes pre-trained on **S**ame dataset with **H**eterogeneous architecture
  - **MH**: modes pre-trained on **M**ultiple dataset with **H**eterogeneous architecture

Goal: DFML aims to enable efficient learning of new unseen tasks by meta-learning the meta-knowledge from a collection of public pre-trained models without access to their private training data.

Motivation

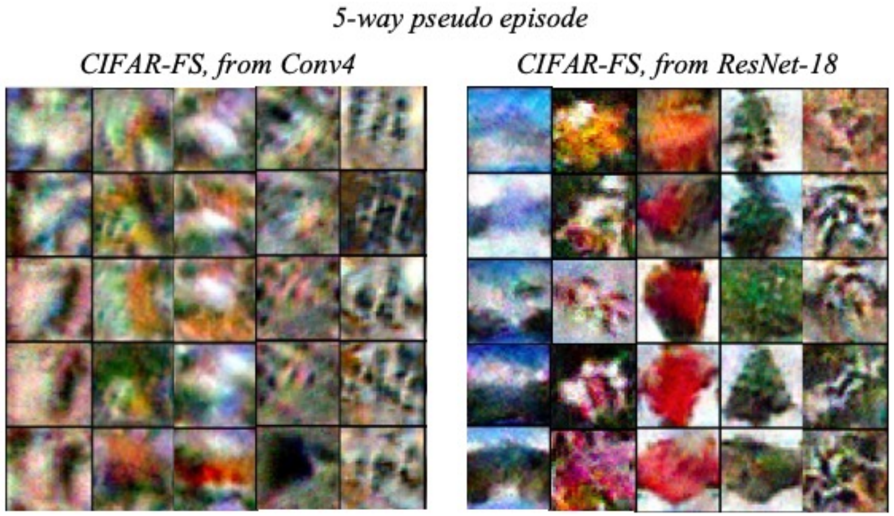
What information can we obtain from numerous pre-trained models of different tasks?

- Model parameters
- Model architectures
- …
- What else?

Model parameters / Model architectures > Prior work DRO*(UAI, 2022) —limitations→

① limited information with unsatisfactory performance
② only applicable to models with the same architecture
③ can not deal with large-scale models

Explore underlying data knowledge contained in models:



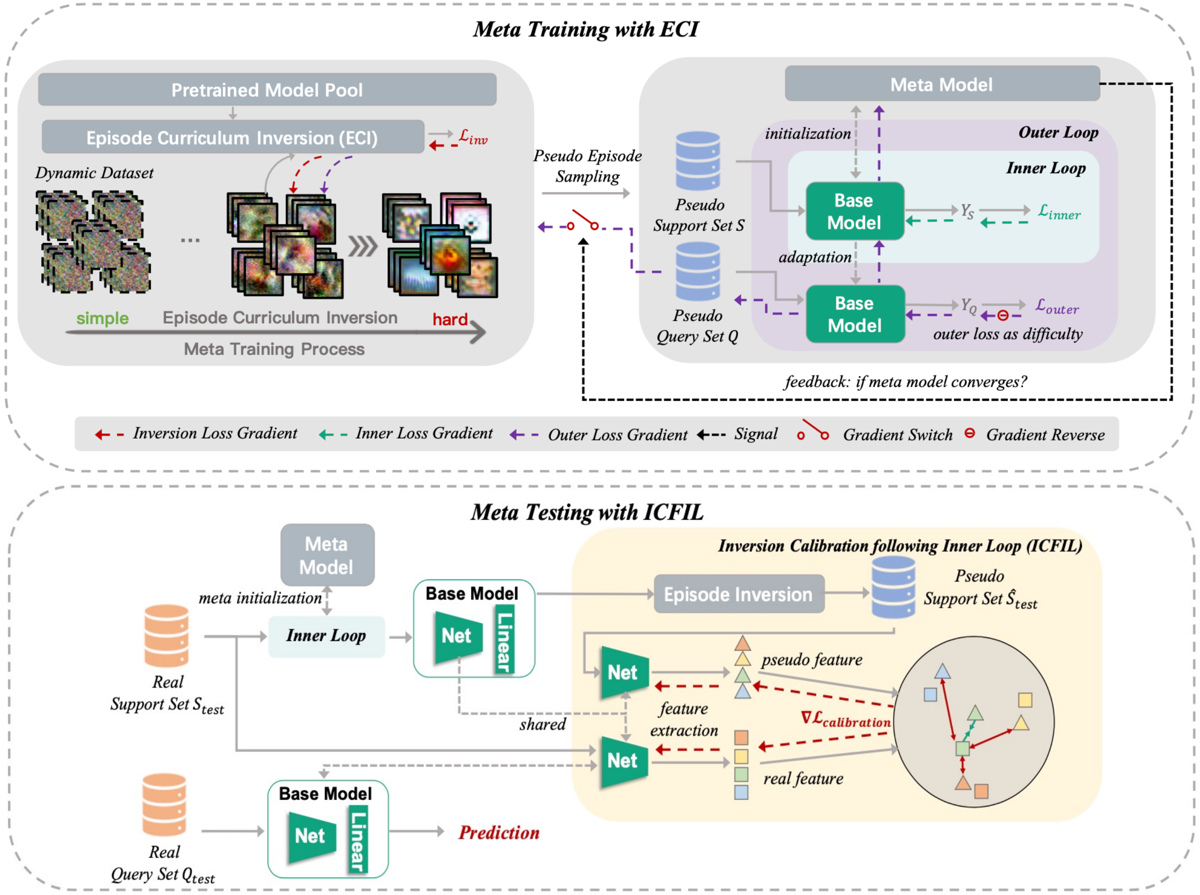
Pseudo Data $\xleftarrow{\text{Inverse}}$ Pre-trained Models

* Zhenyi Wang, et al. Meta-learning without data via Wasserstein distributionally-robust model fusion. UAI 2022.

Methodology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Overall framework of **PURER**:



- Meta-training with **Episode Curriculum Inversion (ECI)**:
  Distill a sequence of pseudo tasks from models of adaptively increasing difficulty level according to the real-time feedback of the meta model.

- Meta-testing with **Inversion Calibration following Inner Loop (ICFIL)**:
  A plug-and-play supplement of mete testing to alleviate the task-distribution shift issue caused by the gap between pseudo tasks during meta training and real tasks during meta testing.

Methodology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Adversarial learning objective of PURER:

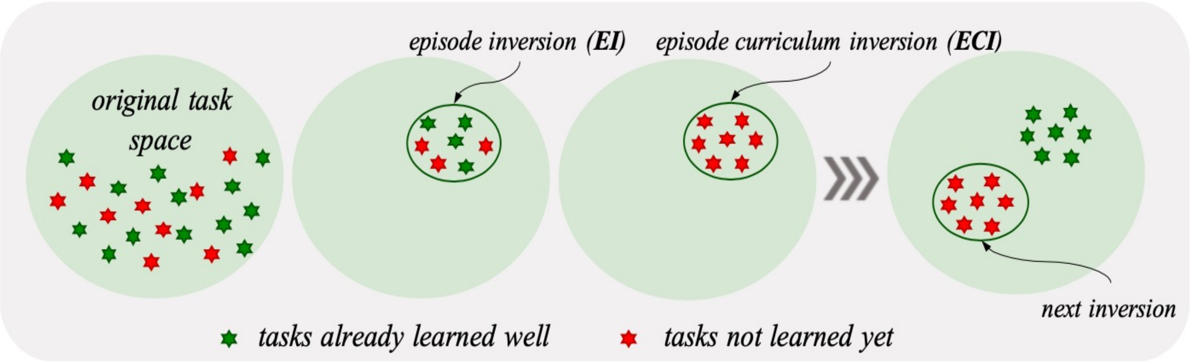
Adversarially update pseudo dataset $\mathcal{D}$ and meta-learner θ:

$$\min_{\boldsymbol{\theta}} \max_{\mathcal{D}} \mathop{\mathbb{E}}_{\mathcal{T}\in\mathcal{D}} [\underbrace{-\mathcal{L}_{inv}(\mathcal{D})}_{\text{data-generation loss}} + \mathbb{I}(\Omega) * \underbrace{\mathcal{L}_{outer}(\mathcal{T};\boldsymbol{\theta})}_{\text{meta-learning loss}}].$$

Methodology

Episode Curriculum Inversion (ECI):

Goal: Distill a sequence of pseudo tasks with adaptively increasing difficulty level to achieve efficient meta training.



At each episode, EI may repeatedly synthesize the tasks already learned well, while **ECI** only synthesizes **harder tasks not learned yet**.

Dataset generation:

$$\min_{\mathcal{D}} \mathcal{L}_{inv}(\mathcal{D}) = \sum_{(\hat{\boldsymbol{x}},y)\in\mathcal{D}} \underbrace{l(\hat{\boldsymbol{x}}, y; \boldsymbol{\psi})}_{\text{classification loss}} + \underbrace{\mathcal{R}_{prior}(\hat{\boldsymbol{x}}) + \mathcal{R}_{feature}(\hat{\boldsymbol{x}})}_{\text{regularization terms}}$$

Curriculum mechanism:

$$\max_{\mathcal{D}} \boxed{\mathbb{I}(\Omega)} * \mathcal{L}_{outer}(\mathcal{T}; \boldsymbol{\theta})]$$

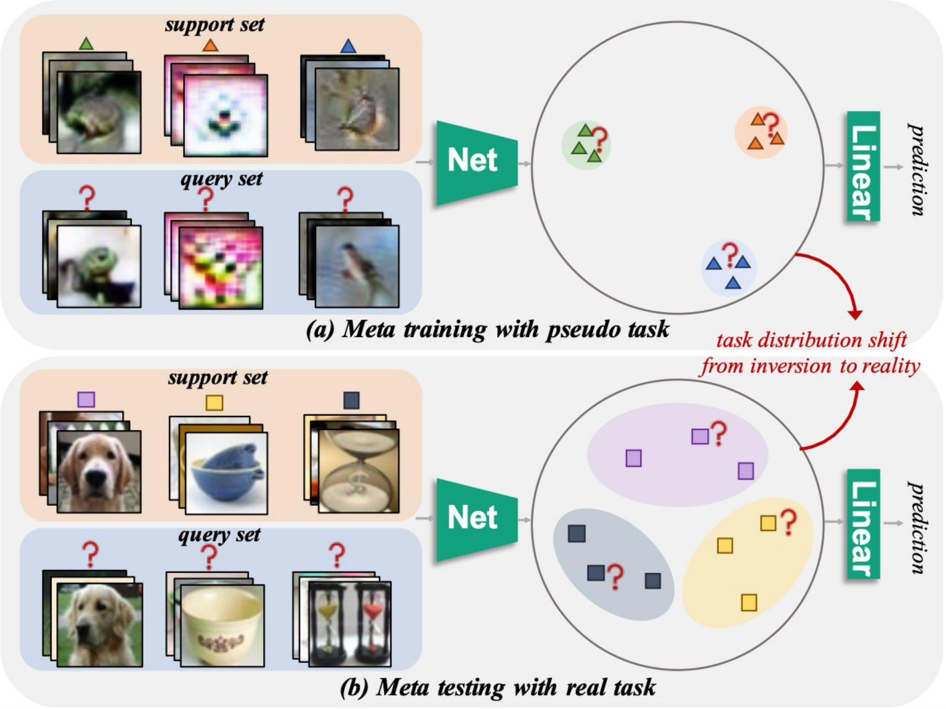
feedback from meta model indicating when harder tasks are needed

Methodology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Inversion Calibration following Inner Loop (ICFIL)

Goal: A plug-and-play supplement of meta testing to alleviate the task-distribution shift issue caused by the gap between pseudo tasks of meta training and real tasks of meta testing.



Task-distribution shift between meta training and testing.

Calibrate the meta-learned meta model via contrastive learning:

$$\mathcal{L}_{calibration}(\phi) = -\sum_{\boldsymbol{x}\in\mathcal{S}_{test}} \sum_{\hat{\boldsymbol{x}}^+} \log \frac{\exp\left[\phi(\boldsymbol{x})^T\phi(\hat{\boldsymbol{x}}^+)/\tau\right]}{\sum_{\hat{\boldsymbol{x}}} \exp\left[\phi(\boldsymbol{x})^T\phi(\hat{\boldsymbol{x}})/\tau\right]}$$

Enforce the backbone to extract similar features for both pseudo and real data, thus alleviating task-distribution shift.

- $\hat{x}^+$: pseudo data with the same label of real data x
- $\hat{x}^-$: pseudo data with different label of real data x

Results

Three real-world scenarios:

- **SS**: modes pre-trained on **S**ame dataset with **S**ame architecture
- **SH**: modes pre-trained on **S**ame dataset with **H**eterogeneous architecture
- **MH**: modes pre-trained on **M**ultiple dataset with **H**eterogeneous architecture

Quantitative results:

SS	Method	1-shot	5-shot
CIFAR-FS 5-way	Random	21.65 ± 0.45	21.59 ± 0.45
	Average	28.12 ± 0.62	32.15 ± 0.64
	OTA	29.10 ± 0.65	34.33 ± 0.67
	DRO	23.92 ± 0.49	24.34 ± 0.49
	Ours	$\mathbf{38.66 \pm 0.78}$	$\mathbf{51.95 \pm 0.79}$
MiniImageNet 5-way	Random	22.45 ± 0.41	23.48 ± 0.45
	Average	22.87 ± 0.39	26.13 ± 0.43
	OTA	24.22 ± 0.53	27.22 ± 0.59
	DRO	23.96 ± 0.42	25.81 ± 0.41
	Ours	$\mathbf{31.14 \pm 0.63}$	$\mathbf{40.86 \pm 0.64}$

SH	Method	1-shot	5-shot
CIFAR-FS 5-way	Random	21.65 ± 0.45	21.59 ± 0.45
	Ours	$\mathbf{39.15 \pm 0.70}$	$\mathbf{49.08 \pm 0.74}$
MiniImageNet 5-way	Random	22.45 ± 0.41	23.48 ± 0.45
	Ours	$\mathbf{28.76 \pm 0.60}$	$\mathbf{35.19 \pm 0.64}$

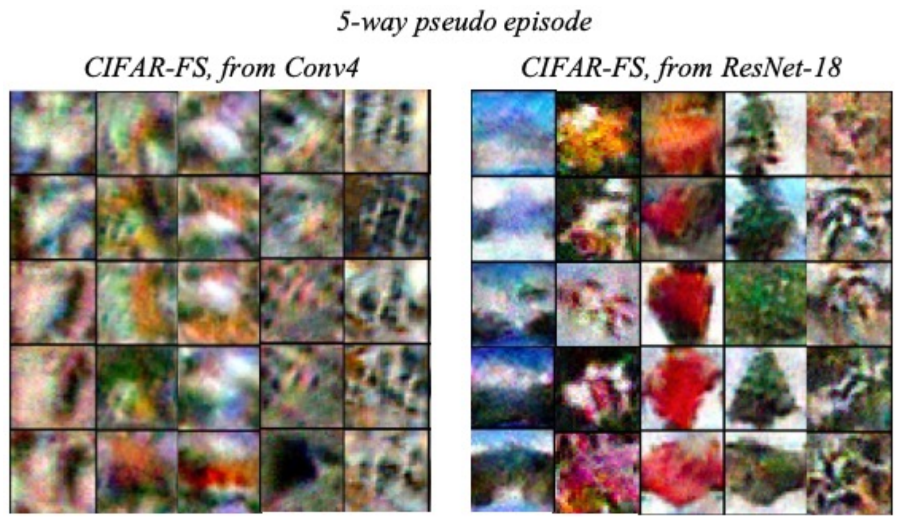
MH	Method	1-shot	5-shot
5-way	Random	21.11 ± 0.41	21.34 ± 0.40
	Ours	$\mathbf{28.50 \pm 0.63}$	$\mathbf{33.10 \pm 0.69}$

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

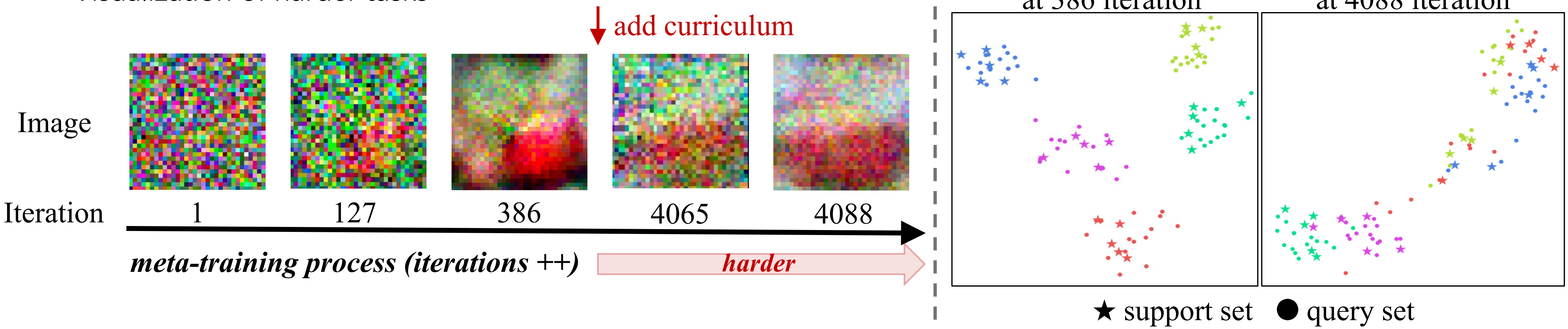
Results

Qualitative results:

Visualization of pseudo tasks



Visualization of harder tasks



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Architecture, Dataset and Model-Scale Agnostic Data-free Meta-Learning

Zixuan Hu[1], Li Shen[2], Zhenyi Wang[3], Tongliang Liu[4], Chun Yuan[1], Dacheng Tao[2]

1Tsinghua Shenzhen International Graduate School, China; 2 JD Explore Academy, China
3 State University of New York at Buffalo, USA; 4 The University of Sydney, Australia

Poster: TUE-PM-345

清華大學 Tsinghua University
京东探索研究院 JD EXPLORE ACADEMY
University at Buffalo The State University of New York
THE UNIVERSITY OF SYDNEY