NEURAL INFORMATION PROCESSING SYSTEMS

NANJING UNIVERSITY LAMDA Learning And Mining from DatA ZHEJIANG UNIVERSITY

Model Spider: Learning to Rank Pre-Trained Models Efficiently

Yi-Kai Zhang[1], **Ting-Ji Huang**[1], **Yao-Xiang Ding**[2], **De-Chuan Zhan**[1], **Han-Jia Ye**[1,✉]

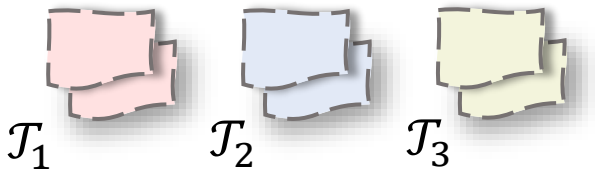
zhangyk@lamda.nju.edu.cn

NeurIPS 2023 [Spotlight]

We introduces Model Spider, a novel approach for efficiently selecting **the most suitable Pre-Trained Model (PTM)** for a given task. It tokenizes both **PTMs** and **tasks** into vector-**representations** and uses approximated performance on historical tasks to **rank PTMs**.

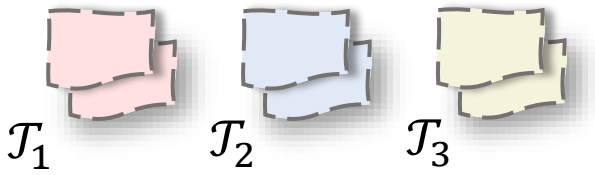
Motivation

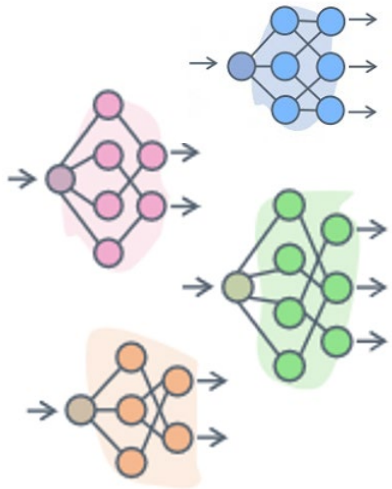
Target Tasks



Motivation

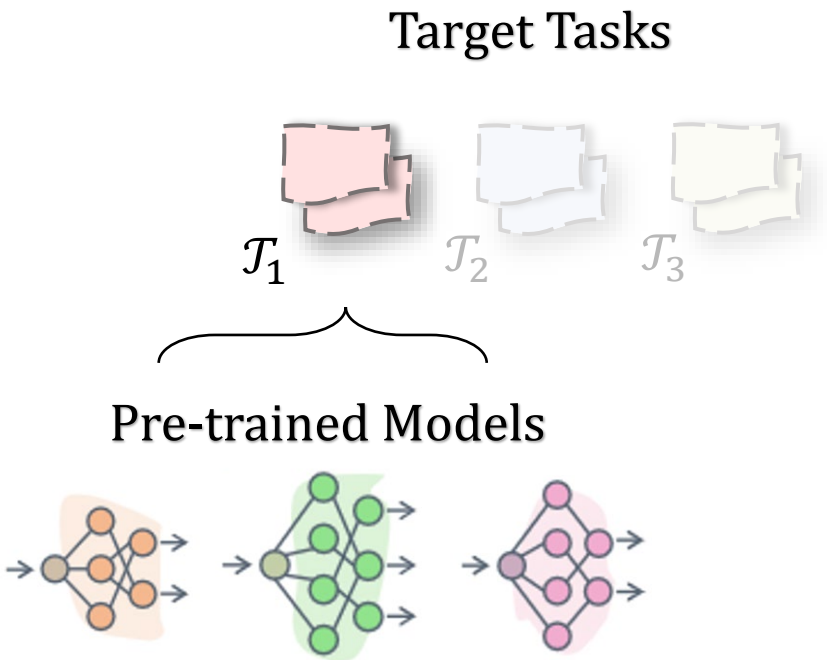
Target Tasks





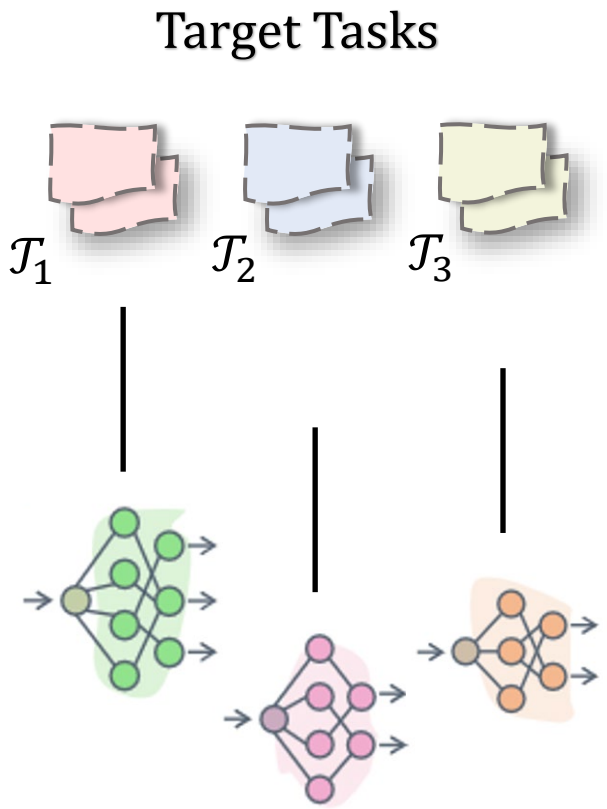
Pre-trained Models

Motivation



Fine-tuned performance: 85% **93%** 73%

Motivation

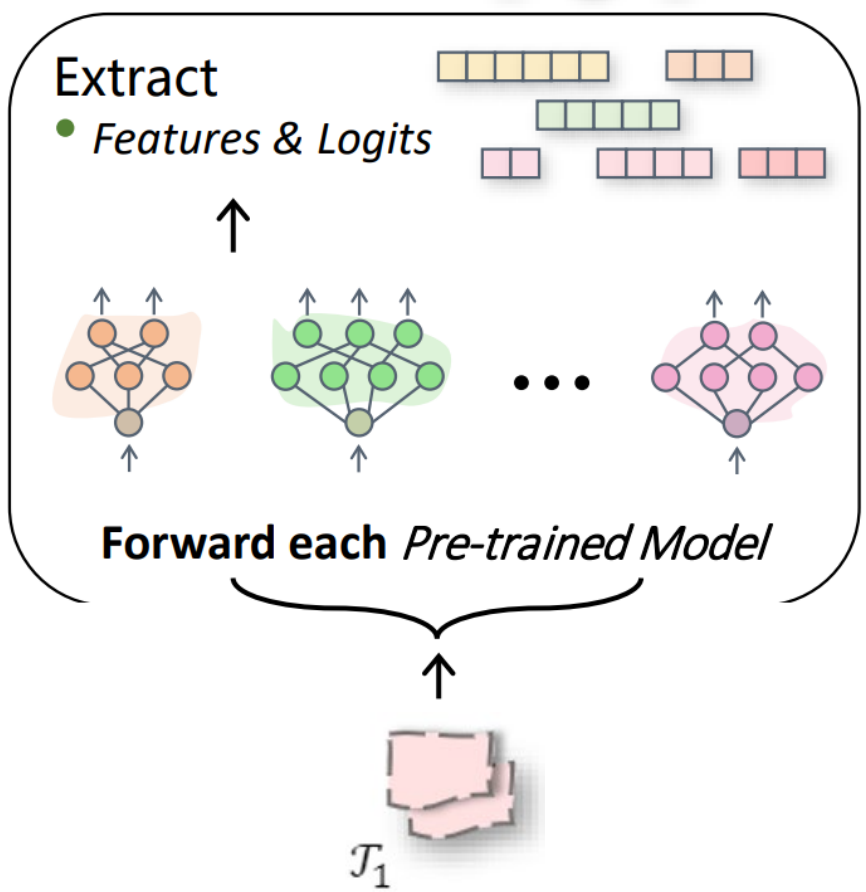


Pre-trained Model **Selection**

selecting **suitable** pre-trained models for target tasks

Motivation

Previous Works



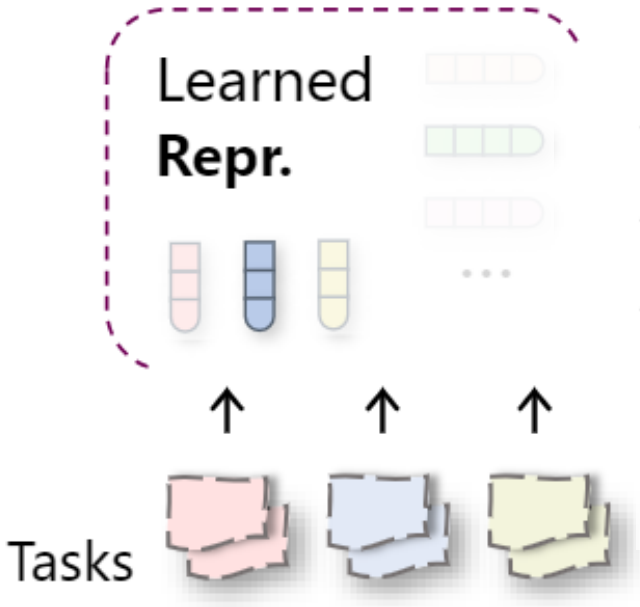
To select on M pre-trained models

1 task, but $\boldsymbol{M}$ forwards

If $\boldsymbol{M}$ is large?

Motivation

Our Solution



Task Representation

← **Just need one-time forward**

← of a ***general*** **encoder** ψ

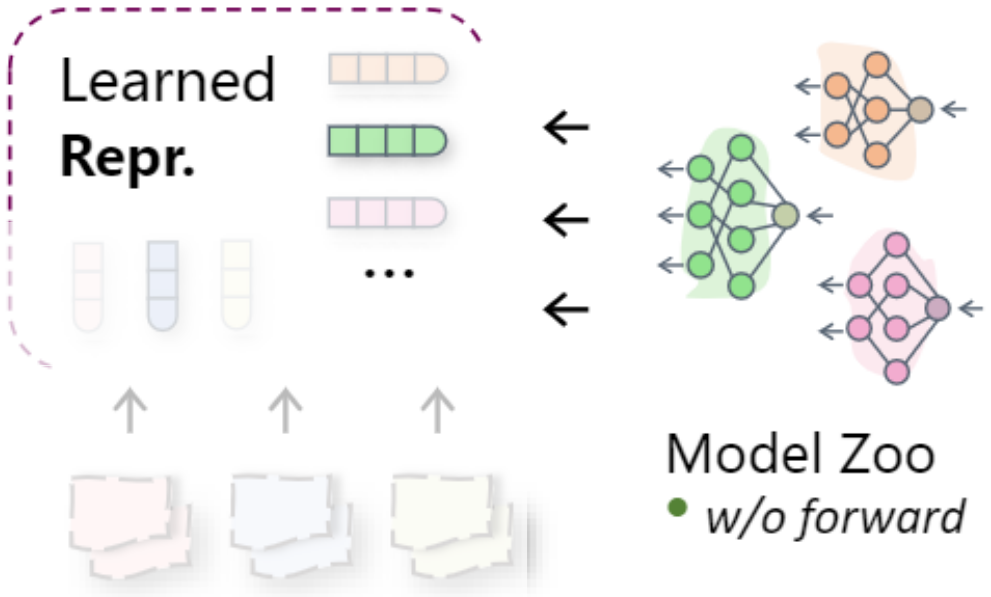
←

Model Zoo

• *w/o forward*

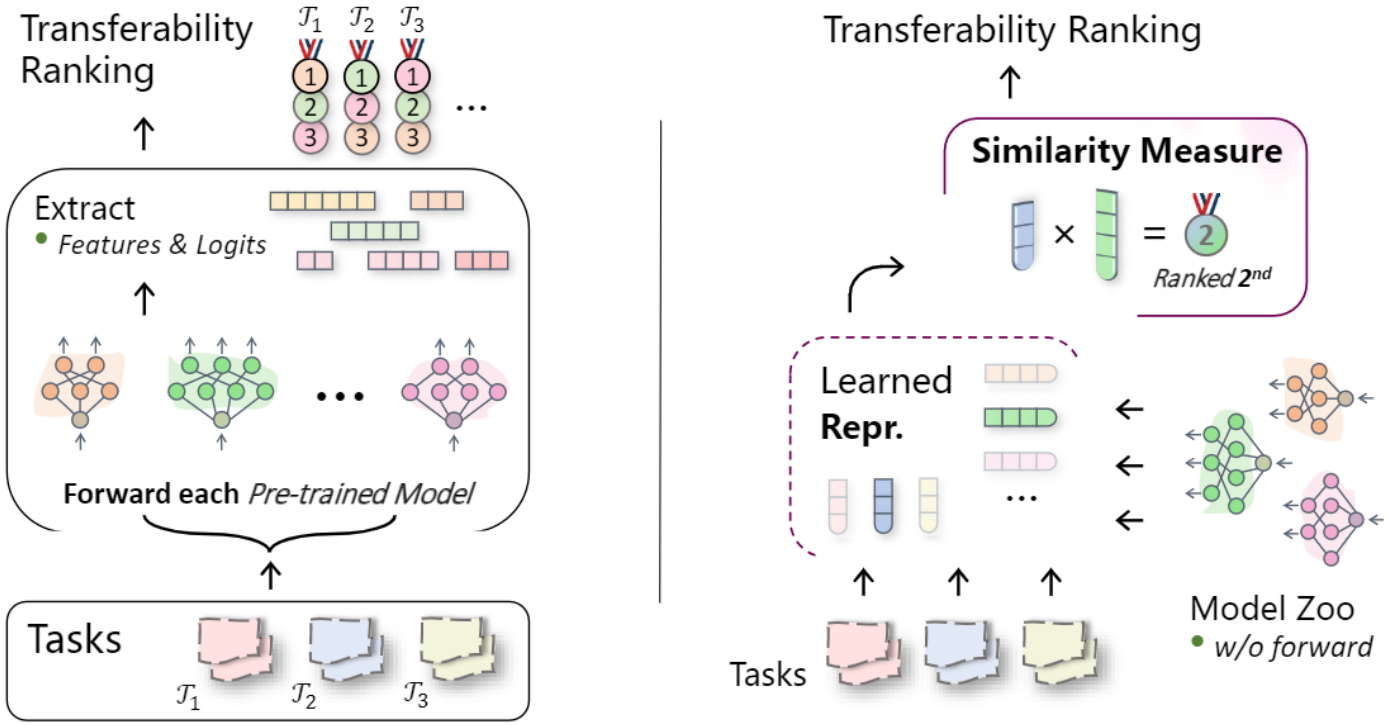
Motivation

Our Solution



Model Representation

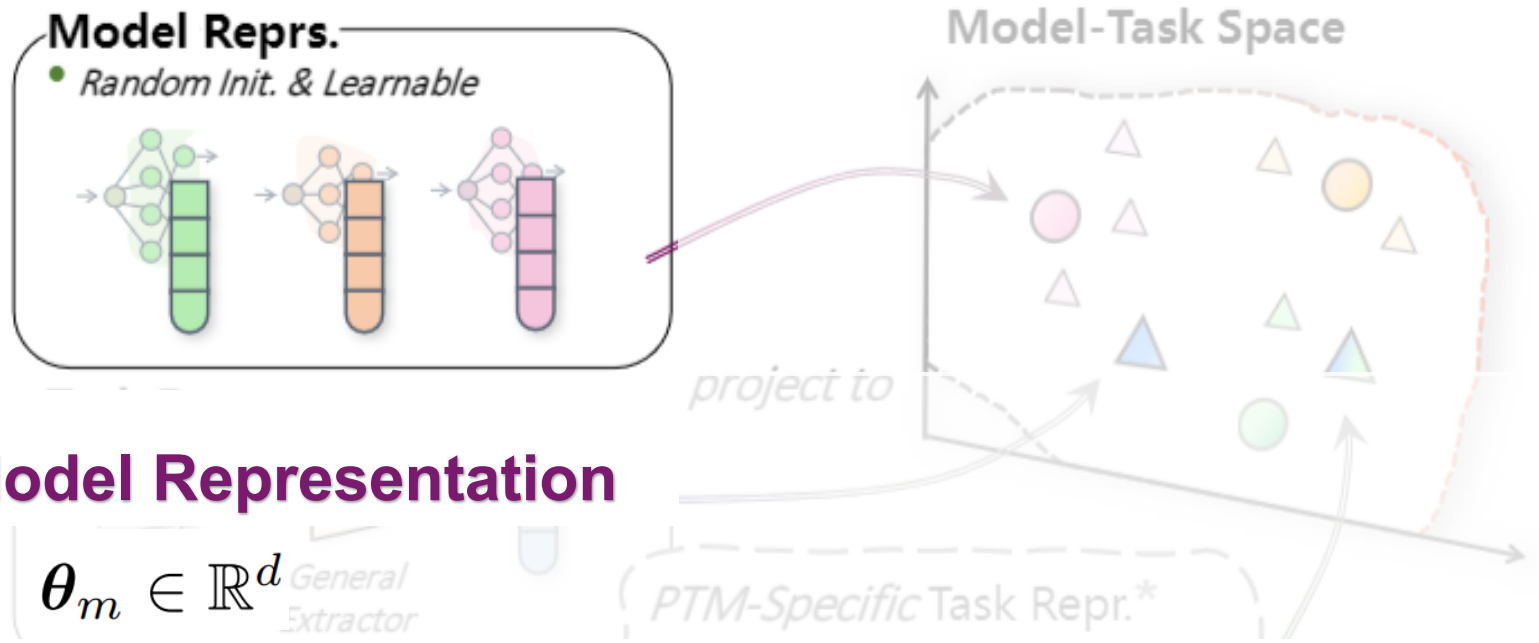
Learn model representation
through ***historical*** **task performance**



Forward-Based and **Representation/Specification**-Based Model Selection

Forward-based methods require forwarding **for each pre-trained model** which still incurs significant overhead!

Method



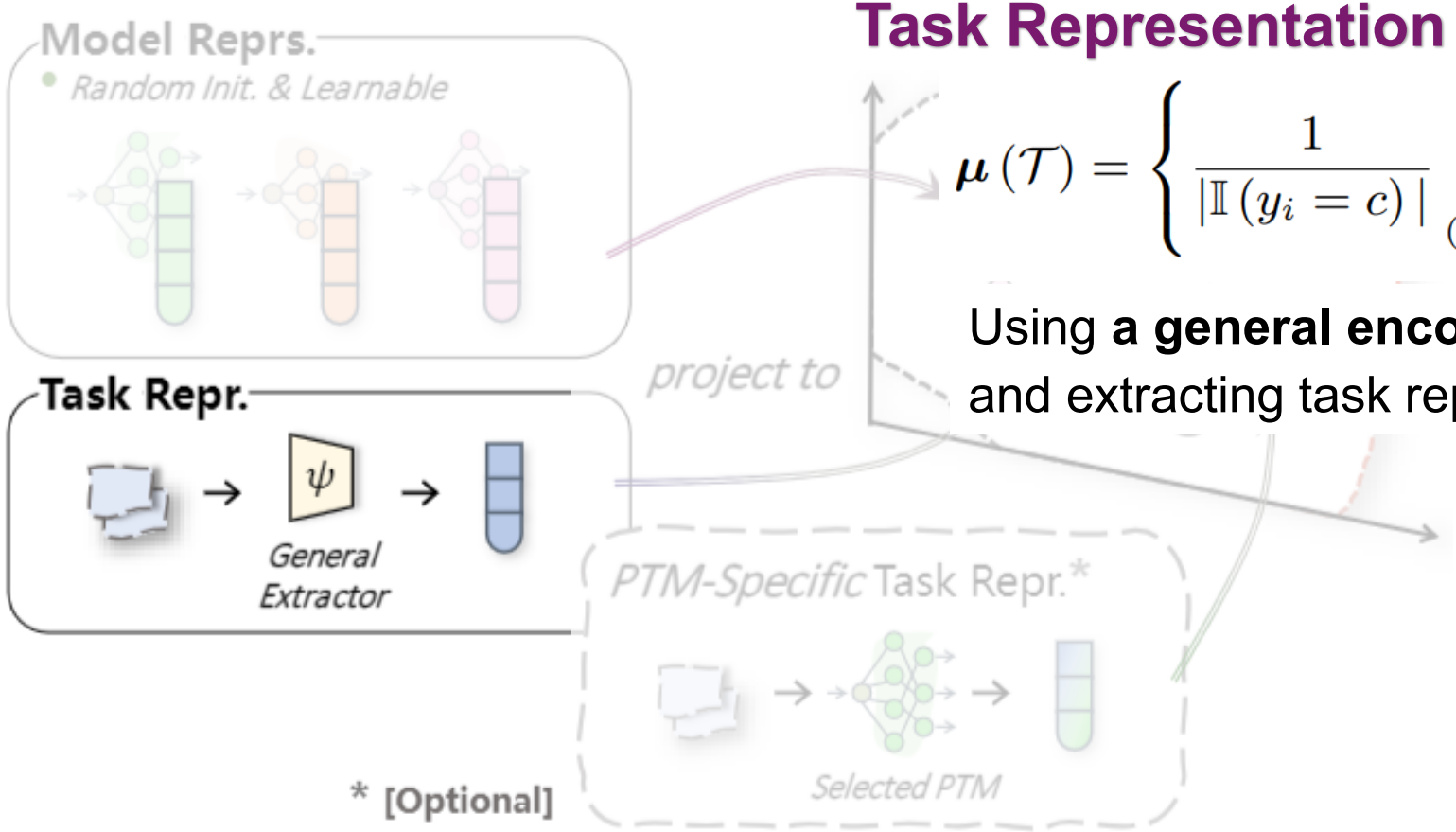
Model Representation

$$\boldsymbol{\theta}_m \in \mathbb{R}^d$$

Learning model representation in a **supervised** learning manner

learnable parameters randomly initialized and optimized through the training process.

Method

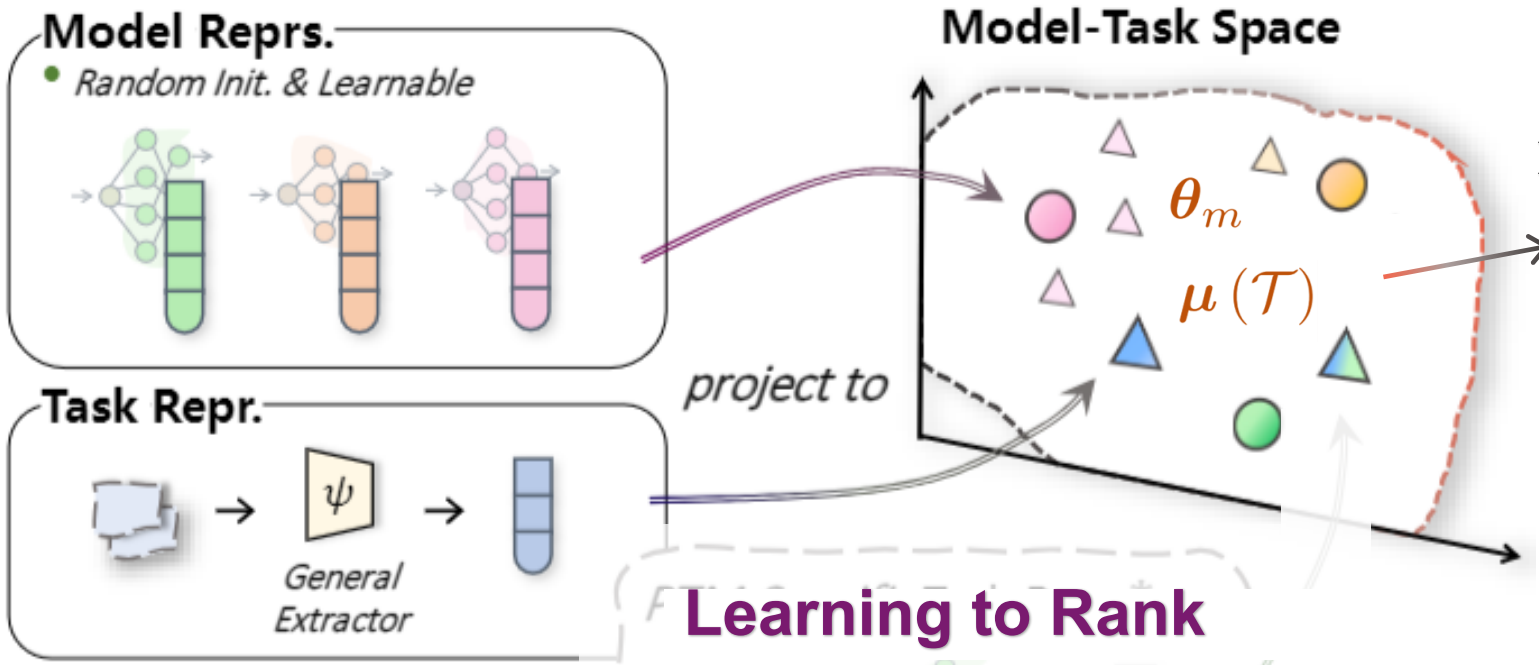


Task Representation

$$\boldsymbol{\mu}(\mathcal{T}) = \left\{ \frac{1}{|\mathbb{I}(y_i = c)|} \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{T}} \left[\psi(\boldsymbol{x}_i) \cdot \mathbb{I}(y_i = c) \right] \right\}_{c \in [C]}$$

Using **a general encoder** ψ
and extracting task repr. with **just one** forward pass.

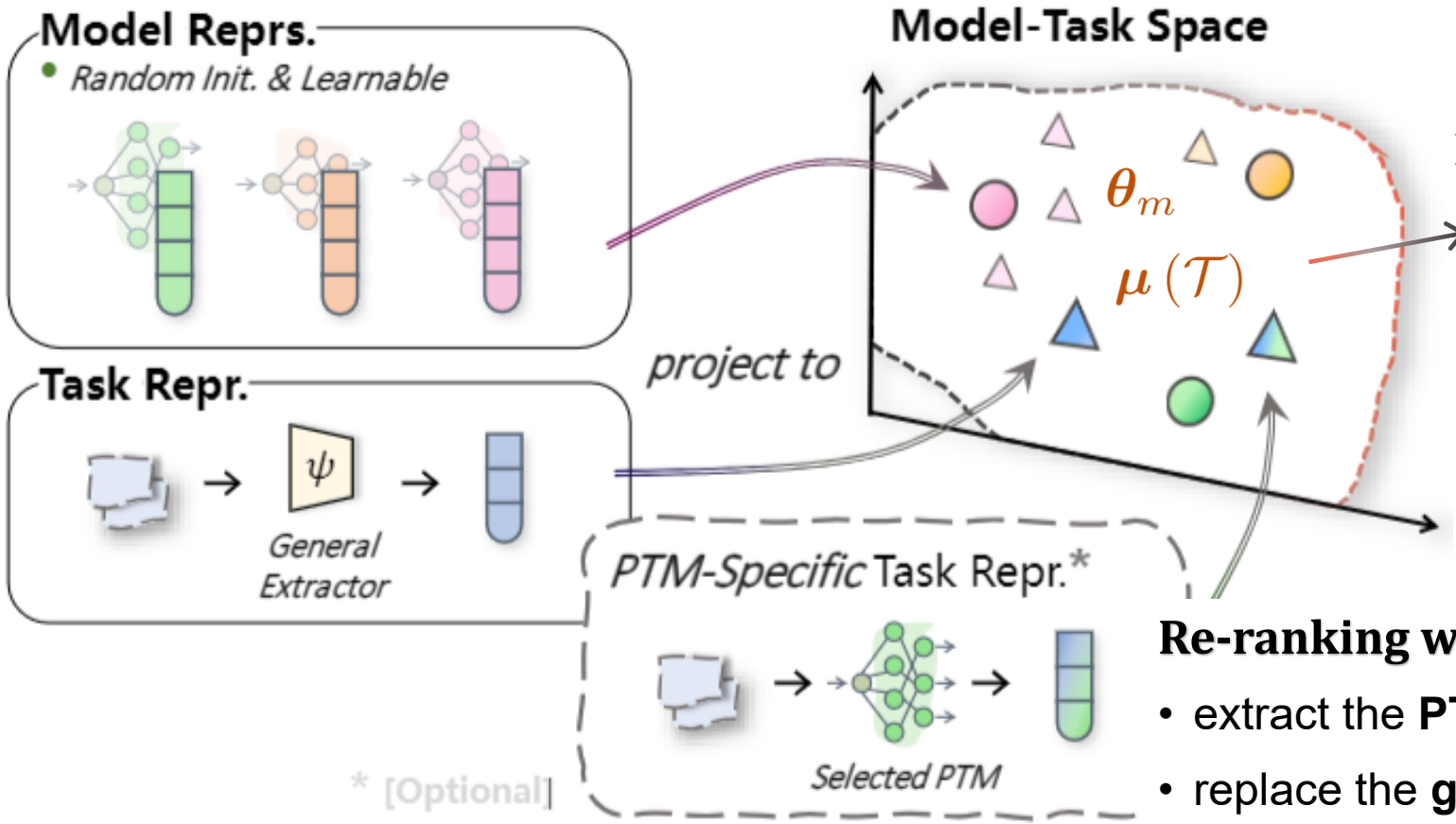
Method



$$\mathrm{sim}(\boldsymbol{\theta}_m, \boldsymbol{\mu}(\mathcal{T})) = \mathrm{FC}(\mathrm{transformer}(\boldsymbol{z})[0])$$

$$\min_{\Theta} \sum_{\mathcal{T}_i \sim \mathcal{D}} \ell_{\mathrm{rank}}\left(\hat{\boldsymbol{t}}_{\mathcal{T}_i} = \{\mathrm{sim}(\boldsymbol{\theta}_m, \boldsymbol{\mu}(\mathcal{T}_i))\}_{m=1}^{M}, \boldsymbol{t}_{\mathcal{T}_i}\right)$$

Method



Model-task Representation

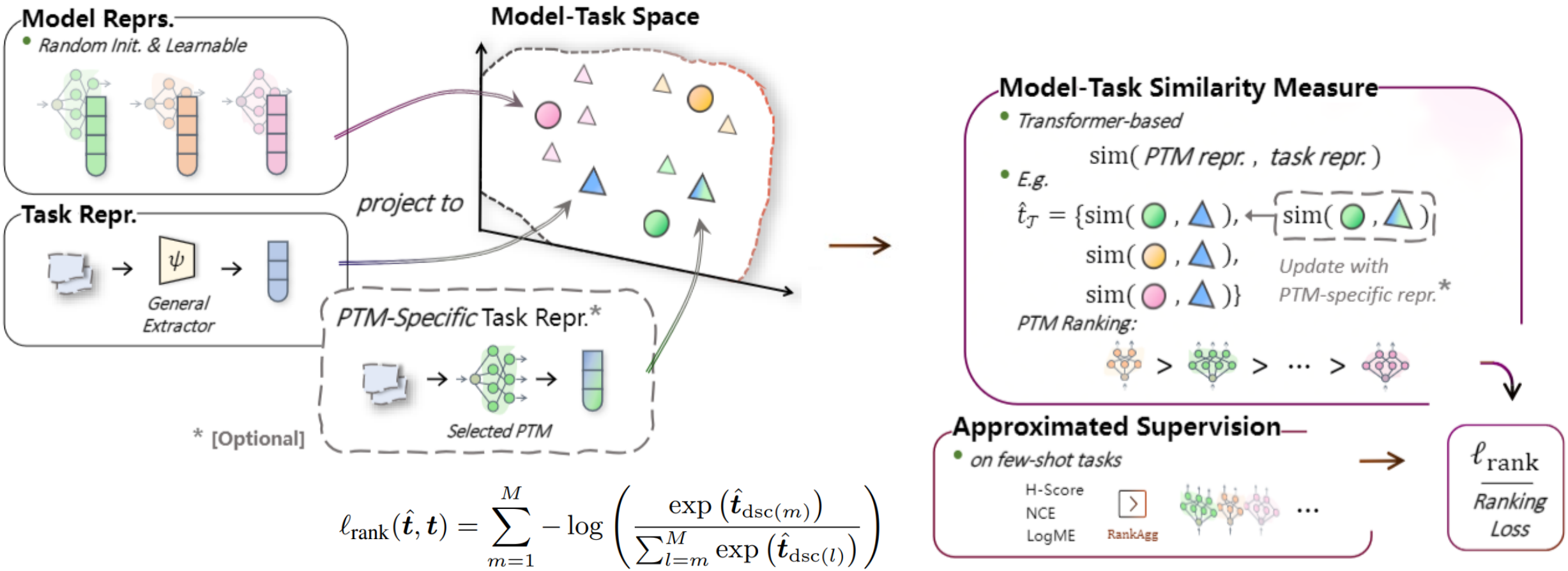
$$\boldsymbol{z} = [\boldsymbol{\theta}_m, \boldsymbol{\mu}(\mathcal{T})]$$

Re-ranking w/ Efficiency-Accuracy Trade-off

- extract the **PTM-specific** task representation
- replace the **general** task repr. $\boldsymbol{\mu}(\mathcal{T})$ via the **specific** one $\boldsymbol{\mu}_m(\mathcal{T})$ when calculating the **similarity** with the θ_m of **the mth PTM**

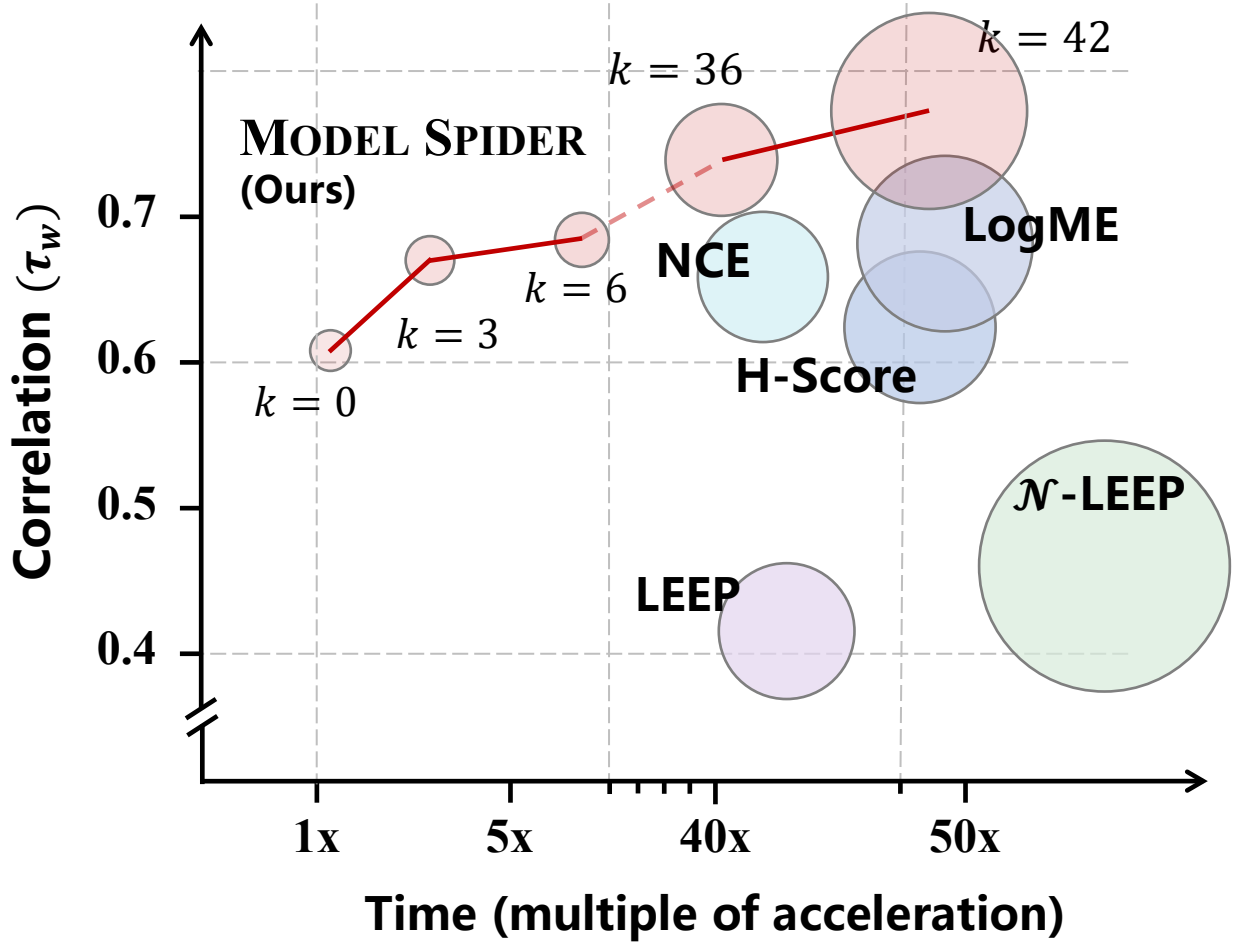
Method Overview

Model Spider: ***Learning to rank*** pre-trained models with **model** and **task representation**



$$\ell_{\text{rank}}(\hat{\boldsymbol{t}}, \boldsymbol{t}) = \sum_{m=1}^{M} -\log\left(\frac{\exp\left(\hat{\boldsymbol{t}}_{\text{dsc}(m)}\right)}{\sum_{l=m}^{M}\exp\left(\hat{\boldsymbol{t}}_{\text{dsc}(l)}\right)}\right)$$

Result



Model Spider:

Efficiency-Accuracy Trade-off

The circle sizes indicate the memory footprint. **Red circles** are our method with different values of the number of **PTM-specific features** k

Result

Method	Downstream Target Dataset								Mean
	Aircraft	Caltech101	Cars	CIFAR10	CIFAR100	DTD	Pets	SUN397	
Standard Evaluation									
H-Score [8]	0.328	0.738	0.616	0.797	0.784	0.395	0.610	0.918	0.648
NCE [73]	0.501	0.752	0.771	0.694	0.617	0.403	0.696	0.892	0.666
LEEP [53]	0.244	0.014	0.704	0.601	0.620	-0.111	0.680	0.509	0.408
$\mathcal{N}$-LEEP [45]	-0.725	0.599	0.622	0.768	0.776	0.074	0.787	0.730	0.454
LogME [83]	**0.540**	0.666	0.677	0.802	0.798	0.429	0.628	0.870	0.676
PACTran [21]	0.031	0.200	0.665	0.717	0.620	-0.236	0.616	0.565	0.397
OTCE [72]	-0.241	-0.011	-0.157	0.569	0.573	-0.165	0.402	0.218	0.149
LFC [19]	0.279	-0.165	0.243	0.346	0.418	-0.722	0.215	-0.344	0.034
GBC [56]	-0.744	-0.055	-0.265	0.758	0.544	-0.102	0.163	0.457	0.095
MODEL SPIDER	0.506	**0.761**	**0.785**	**0.909**	**1.000**	**0.695**	**0.788**	**0.954**	**0.800**
Few-Shot Evaluation (10-example per class)									
H-Score [8]	-0.014	0.078	0.375	0.018	0.005	-0.028	-0.006	0.853	0.160
NCE [73]	0.273	0.534	0.597	0.267	0.232	0.362	0.352	0.793	0.426
LEEP [53]	0.069	-0.038	0.476	0.530	0.471	-0.111	0.567	0.468	0.304
$\mathcal{N}$-LEEP [45]	-0.559	0.476	0.743	0.515	0.707	0.027	0.713	0.812	0.429
LogME [83]	0.341	0.453	0.497	0.718	0.698	0.407	0.657	0.817	0.574
PACTran [21]	0.136	0.262	0.484	0.631	0.614	-0.227	0.701	0.477	0.385
OTCE [72]	-0.316	-0.050	-0.127	0.515	0.505	-0.168	0.406	0.210	0.123
LFC [19]	0.226	-0.226	-0.235	0.330	0.271	-0.669	-0.059	-0.151	-0.064
MODEL SPIDER	**0.382**	**0.711**	**0.727**	**0.870**	**0.977**	**0.686**	**0.717**	**0.933**	**0.750**

Dataset	Methods for Regression Tasks			
	H-Score	LogME	GBC	**Ours**
dSprites	0.106	0.612	-0.283	**0.679**
UTKFace	0.075	-0.156	0.052	**0.364**

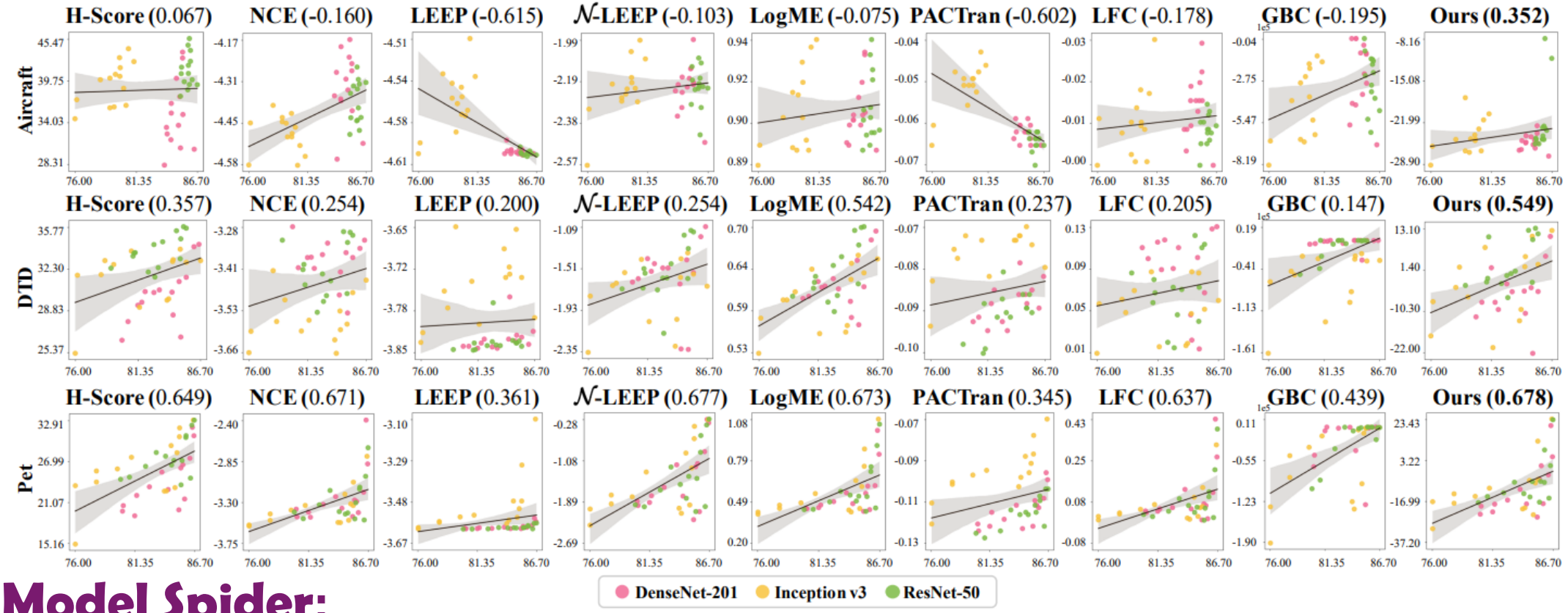
Model Spider:

State-of-the-art on

single-source heterogeneous pre-trained model zoo

10 PTMs pre-trained on ImageNet across five architecture families

Result



Model Spider:

***State-of-the-art on* multi-source heterogeneous pre-trained model zoo**

42 PTMs pre-trained on **14** datasets including animals, general and 3D objects, plants, scene-based, remote sensing and multi-domain recognition.

Result on Ranking LLMs

Method	Downstream Target Dataset					Mean
	Exam.	Language	Knowledge	Understand.	Reason.	
LLM Evaluations						
Alpaca-7B [93]	24.30	67.20	41.95	33.30	51.70	43.69
ChatGLM2-6B [115]	39.00	67.30	44.35	40.25	68.67	51.91
LLaMA2-7B [95]	31.30	67.40	55.90	40.30	52.93	49.57
Vicuna-7B [118]	29.10	66.70	49.45	34.70	52.67	46.52
ChatGPT [68]	39.90	60.90	**57.10**	55.40	69.90	56.64
Top-1 Results of LLM Ranking Methods, *Selected by*						
Self-assessed Confidence	34.60	67.40	45.10	37.45	62.60	49.43
Perf. on Similar Tasks	29.10	67.20	44.35	53.45	63.03	51.43
MODEL SPIDER	**41.30**	**67.65**	55.90	**56.80**	**70.07**	**58.34**

Model Spider:

State-of-the-art on a pre-trained model zoo **of 9 Large Language Models**

Top-1 ranked LLM performance comparisons against LLM evaluation results

including **2** directly baselines, as ranking by self-assessed and performance on similar task

10 downstream tasks are constructed based on the **OpenCompass*** benchmark

*: Opencompass: A universal evaluation platform for foundation models

南京大学 NANJING UNIVERSITY

南京大学 NANJING UNIVERSITY | LAMDA Learning And Mining from DatA | 浙江大学 ZHEJIANG UNIVERSITY

Thanks!

Yi-Kai Zhang's Homepage
lamda.nju.edu.cn/zhangyk

GitHub Page
github.com/zhangyikaii/Model-Spider