Rethinking the Bias of Foundation Model under Long-tailed Distribution

Jiahao Chen, Bin Qin*, Jiangmeng Li*, Hao Chen, Bing Su†*
Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
{nicelemon666, subingats}@gmail.com

Introduction

As shown in Fig.1 the fine-tuned model influenced by the bias of pre-training dataset as well as downstream dataset

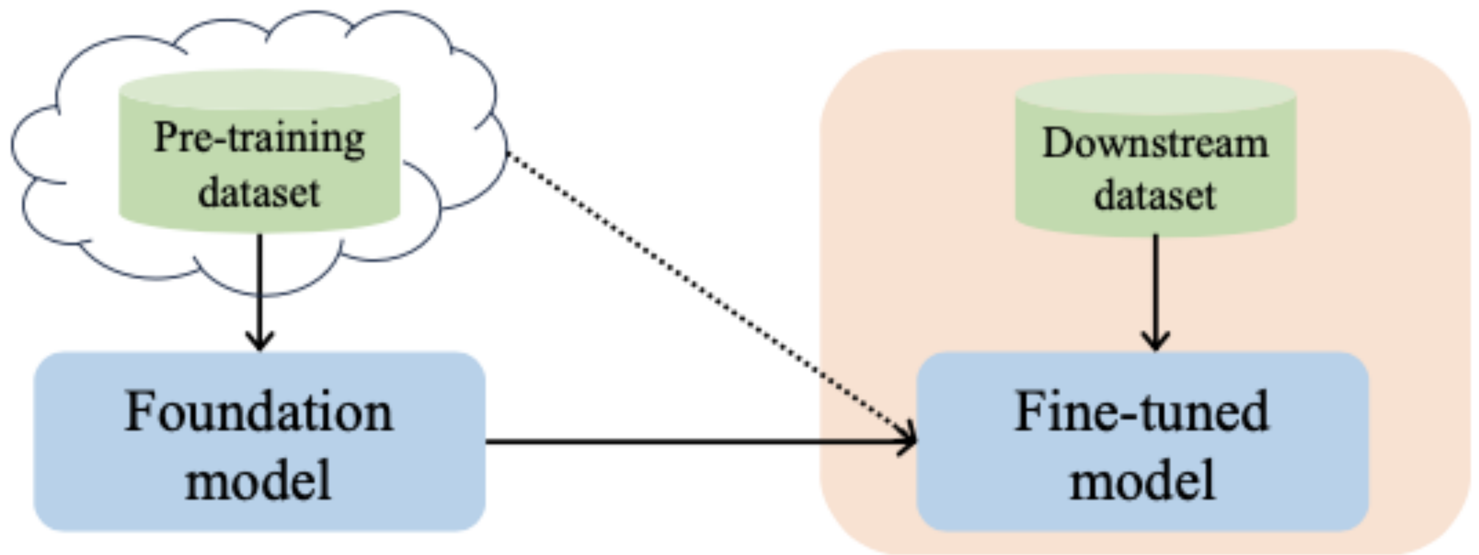
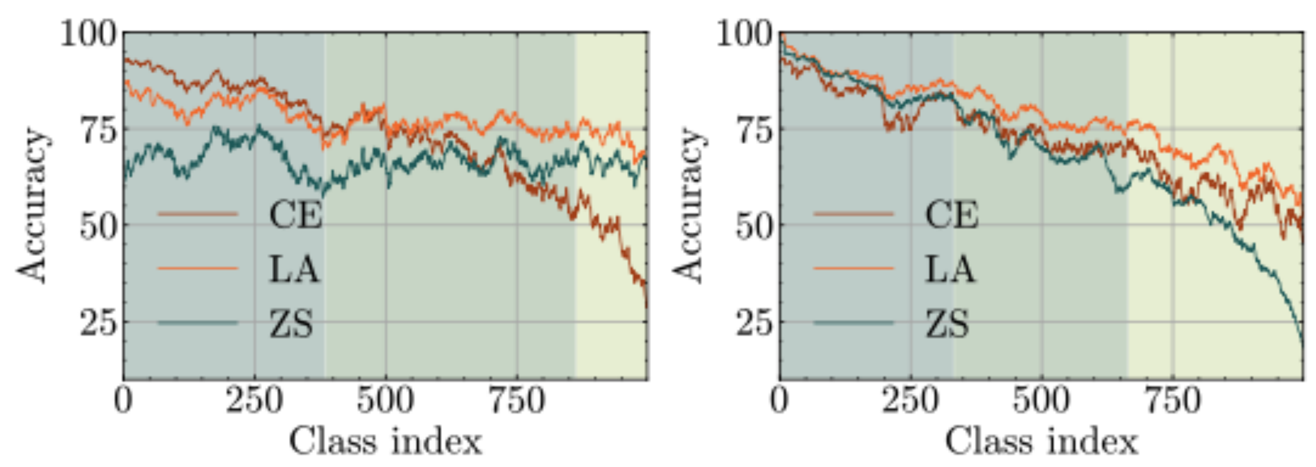


Figure 1. Previous methods focus on using the downstream data to fine-tune the foundation model while ignoring that the pre-training data has a potential influence of bias (dashed line).

Parameter imbalance:
Since the pre-training data is often inaccessible, its influence is primarily reflected in the pre-trained weights, or parameters, which we refer to as parameter imbalance.

Data imbalance:
downstream data is accessible and directly affects the downstream task, which we define as data imbalance.

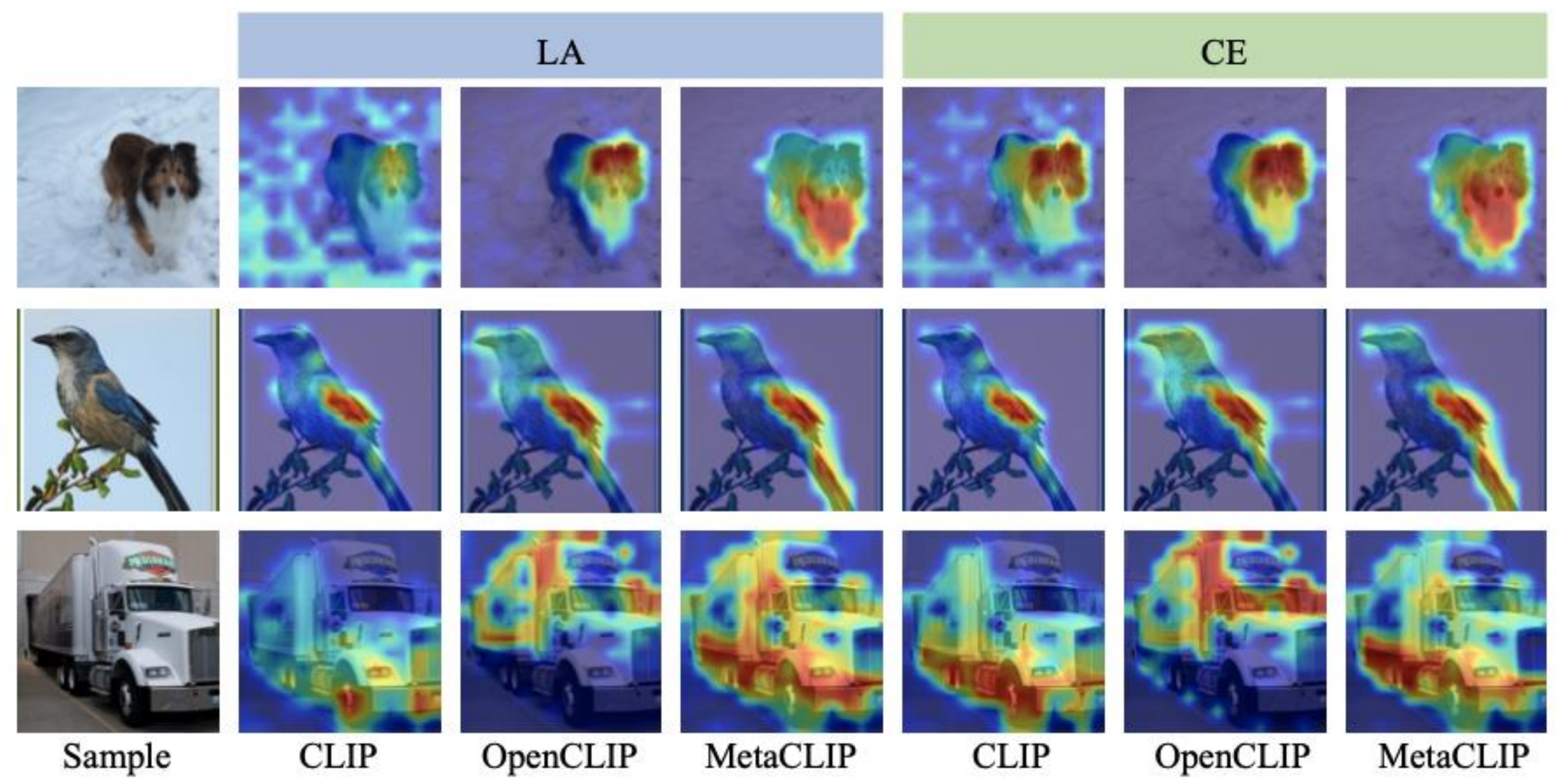
Observation

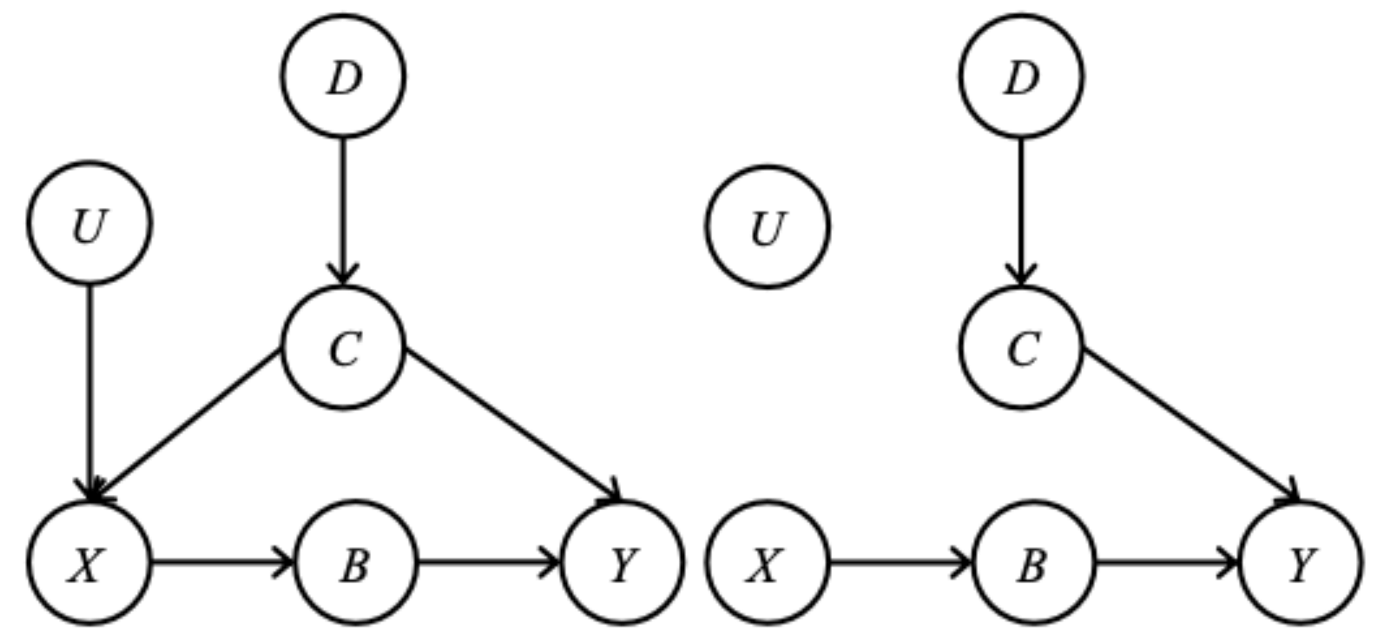


(a) Data and parameter imbalance on ImageNet-LT

1. Parameter imbalance is more important
2. Foundation models serve different parameter imbalance.
3. The parameter imbalance occupy
4. Tail-Tail hurts

The Proposed Method





(a) Confounded (b) Deconfounded

Figure 5. The framework of our proposed method. (a) In the confounded setting, C influences both X and Y, leading to the backdoor path $X \leftarrow C \rightarrow Y$, thereby introducing confounding bias. (b) After intervening on X, its parent nodes C and U are severed, eliminating the unstable backdoor path and leading to a more reliable estimation of the causal effect between X and Y.

Our solution (Backdoor adjustment):

$$\mathbb{P}(Y=y \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{b}\in B}\sum_{c\in C}\mathbb{P}(Y=y \mid \boldsymbol{b}, c)\mathbb{P}(c)\mathbb{P}(\boldsymbol{b} \mid \boldsymbol{x})$$
$$= \sum_{c\in C}\mathbb{P}(Y=y \mid \boldsymbol{b}, c)\mathbb{P}(c)$$

Ensemble the output logits of different foundation models!

Experiments

Dataset: Places365-LT, ImageNet-LT, iNaturalist2018
Evaluation: Many, Medium, Few, All accuracy

Table 6. Results on ImageNet-LT

Method	Epochs	D-Many	D-Medium	D-Few	All
LiVT	100	73.6	56.4	41.0	60.9
VL-LTR	100	**84.5**	74.6	59.3	77.2
BALLAD	60	79.1	74.5	69.8	75.7
Decoder	18	-	-	-	73.2
GML	100	-	-	-	78.0
LIFT	10	80.2	76.1	71.5	77.0
Ours	10	82.21	**78.75**	**74.99**	**79.57**

Table 8. The ablation study with the number of semantic factors M. Open and Meta denote the OpenCLIP and MetaCLIP, respectively.

	CLIP	Open	Meta	ImageNet-LT D-Many	ImageNet-LT D-Medium	ImageNet-LT D-Few
$M=1$	✔			80.25	76.05	71.53
		✔		79.94	76.17	71.70
			✔	80.45	77.06	72.64
$M=2$	✔	✔		81.65	77.99	73.91
	✔		✔	81.86	78.26	74.33
		✔	✔	81.64	78.02	74.08
$M=3$	✔	✔	✔	**82.21**	**78.75**	**74.99**

Conclusion

This paper investigates how foundation model bias impacts downstream imbalanced tasks. We formally define parameter and data imbalance and show that while fine-tuning mitigates data imbalance, parameter imbalance remains a challenge. Re-balancing methods offer limited improvement in feature representation. To address this, we construct a causal structure graph to identify confounding semantic factors and propose a backdoor adjustment method. Experiments confirm its effectiveness. Future work will extend the approach to tasks like object detection and segmentation under long-tailed distributions.