

分布外汎化・較正性能向上のための事前学習済みモデルの 大規模な比較実験

理研AIP

幡谷龍一郎

チーム

長沼大樹 hiroki11x.github.io



博士学生@モントリオール大・MILA
研究の興味

深層モデルの分布外汎化， 較正
深層モデルの最適化

幡谷龍一郎 mosko.tokyo

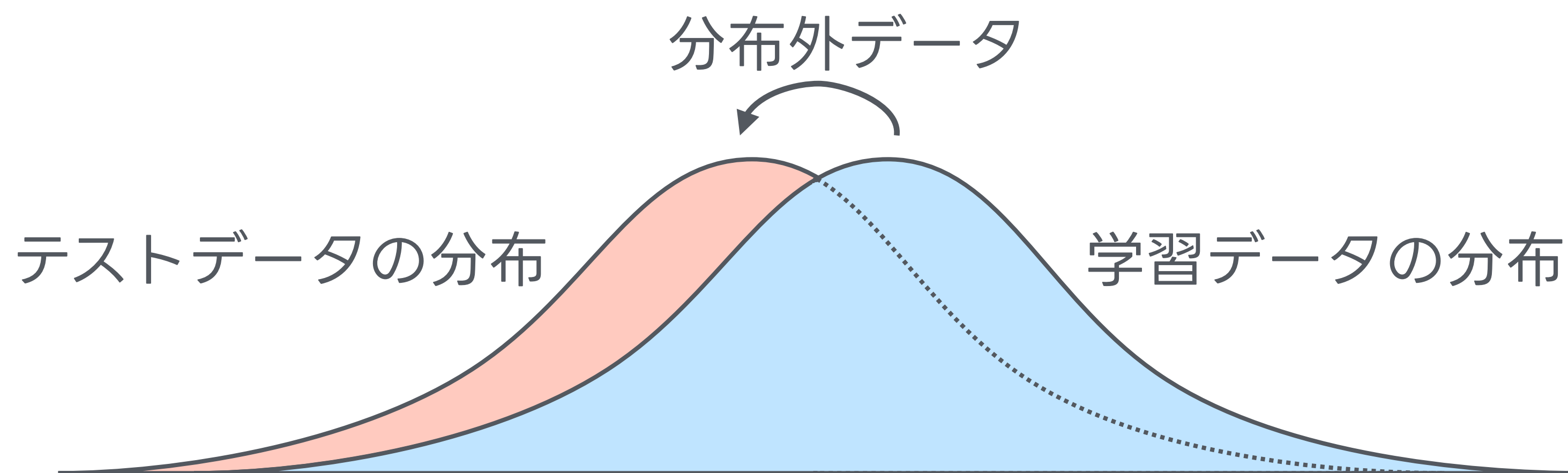


特別研究員@理研AIP
研究の興味

勾配法によるハイパーパラメータ最適化
生成モデルによるデータセット汚染
Transformerの理論解析

事前学習モデルの差異は**分布外汎化**にどのような影響を及ぼすのか？

→ Huggingface hubの97モデルを使って調査した



分布外汎化

分布外データ上での汎化能力

分布外データに対しても**正しく分類**できるか

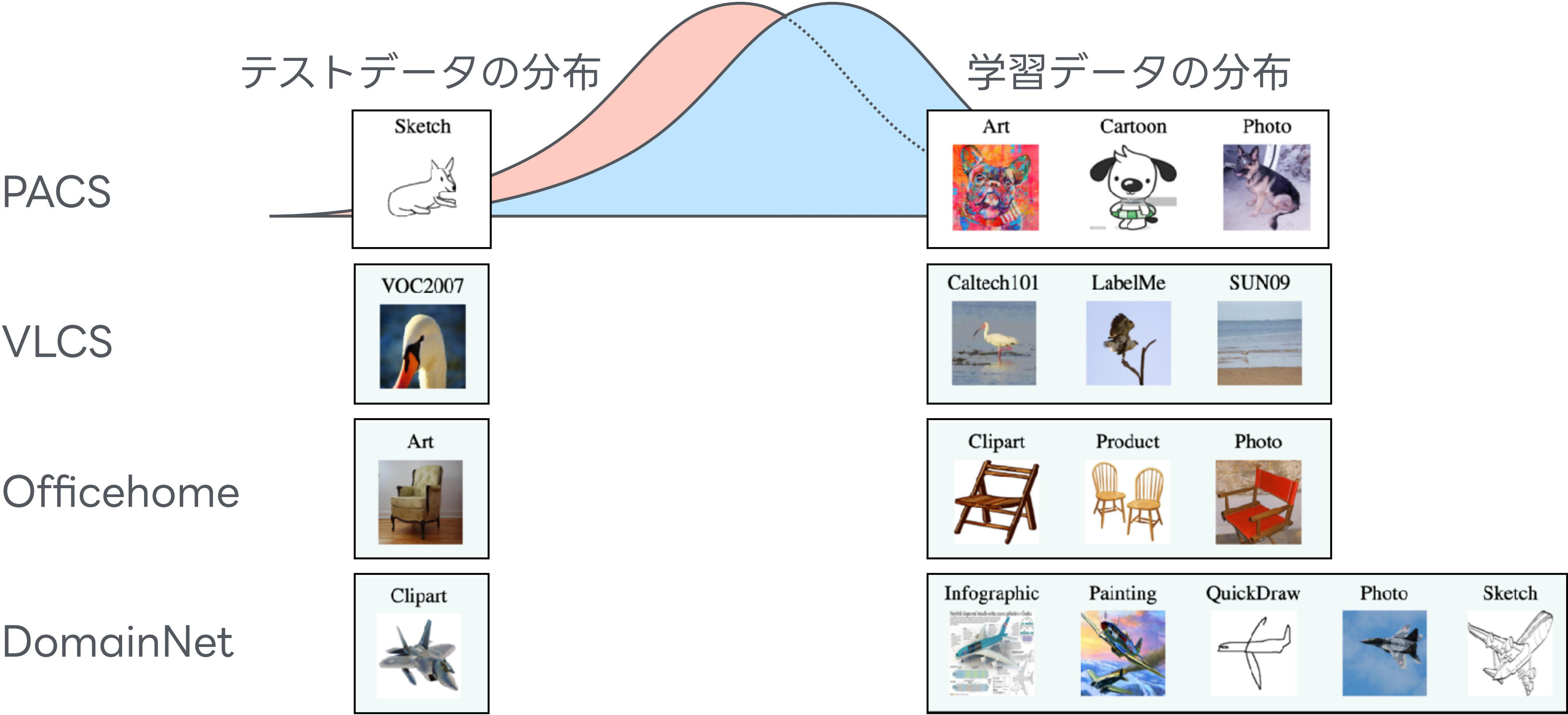
分布外データ上での校正能力

分布外データに対しては**不確実**であるか

本研究の設定（データ）

DomainBedベンチマーク中の4データセットを使用

[Gulrajani&Lopez-Paz, 20]



本来はImageNetで事前学習したResNet-50を学習データでファインチューニングする設定

本研究の設定（モデル）

Huggingface hub上のImageNet事前学習済み画像分類モデルを97個選択



モデルの種類

VGG
ResNet
RegNet
EfficientNet
ConvNeXt
MLP-mixer
Vision Transformer

事前学習データ

ImageNet-1k
ILSVRC-2012
ImageNet-12k
ImageNet-21k

事前学習方法

教師あり学習
教師あり学習+データ拡張
自己教示あり学習（MAE）
など

より大規模な事前学習データはDomainBedのデータを含む可能性があるので排除

本研究の設定（ファインチューニング・モデル選択）

ファインチューニング

最適化アルゴリズムのハイパーパラメータのみをグリッド探索

SGD

学習率： $\{1.0 \times 10^{-4}, 5.0 \times 10^{-4}, \dots, 5.0 \times 10^{-1}\}$

重み減衰： $\{1.0 \times 10^{-4}, 1.0 \times 10^{-3}, 1.0 \times 10^{-2}\}$

事前学習済みモデルの用いた最適化アルゴリズム（Adamなど）

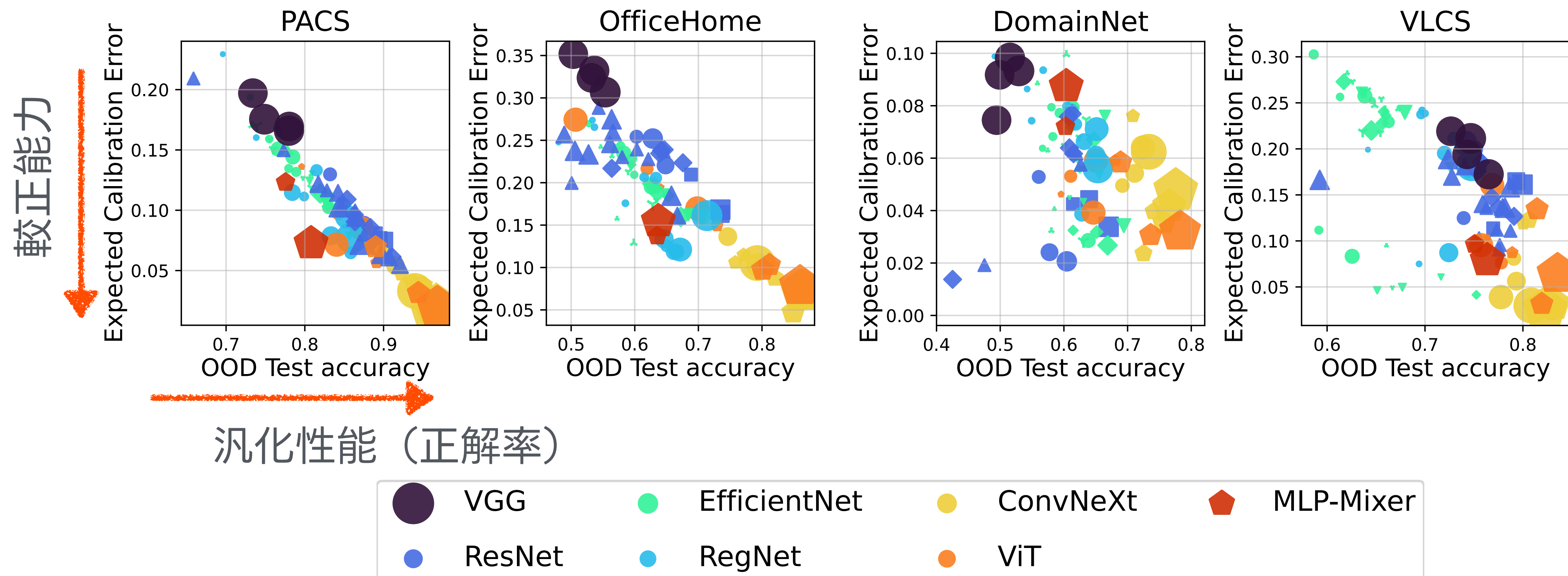
学習率： $\{1.0 \times 10^{-4}, 5.0 \times 10^{-4}, \dots, 5.0 \times 10^{-1}\}$

モデル選択

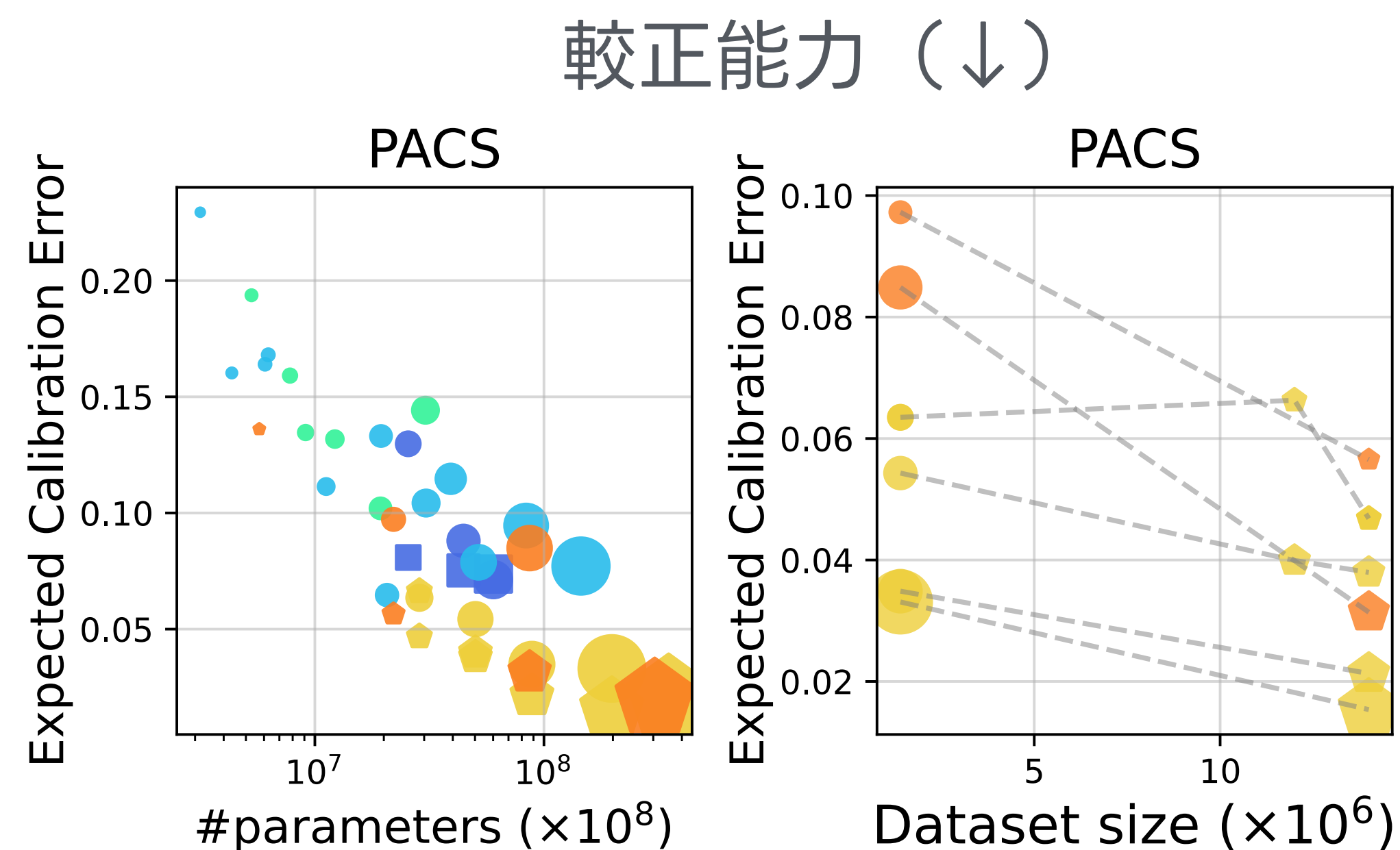
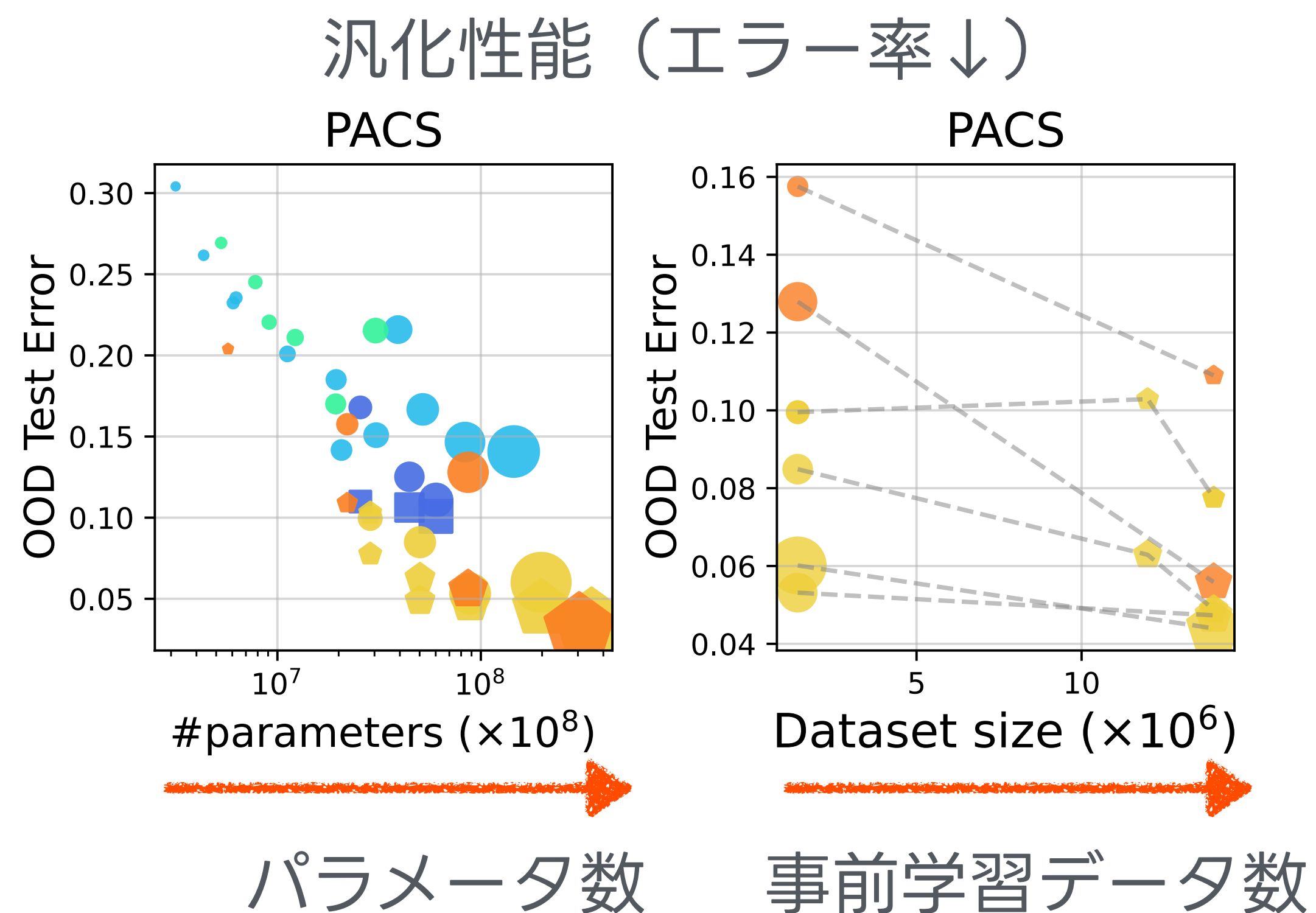
学習データの一部を評価データとして切り分けて

評価データ上での正解率を用いてモデルを選択

分布外データでの汎化性能と較正能力には相関

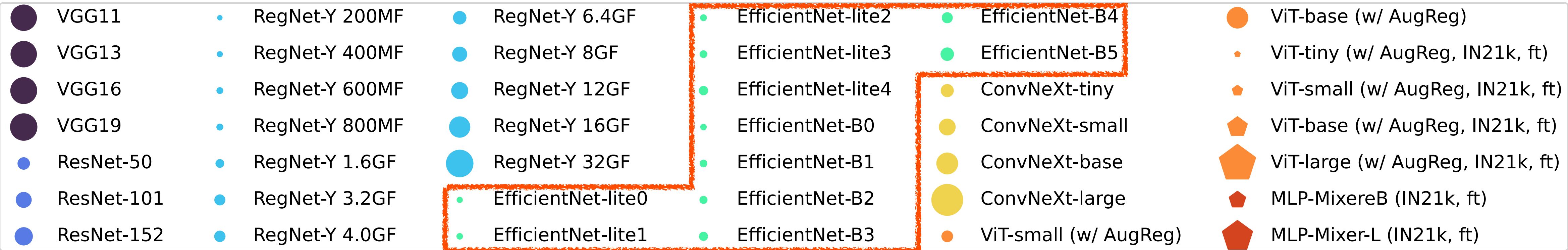
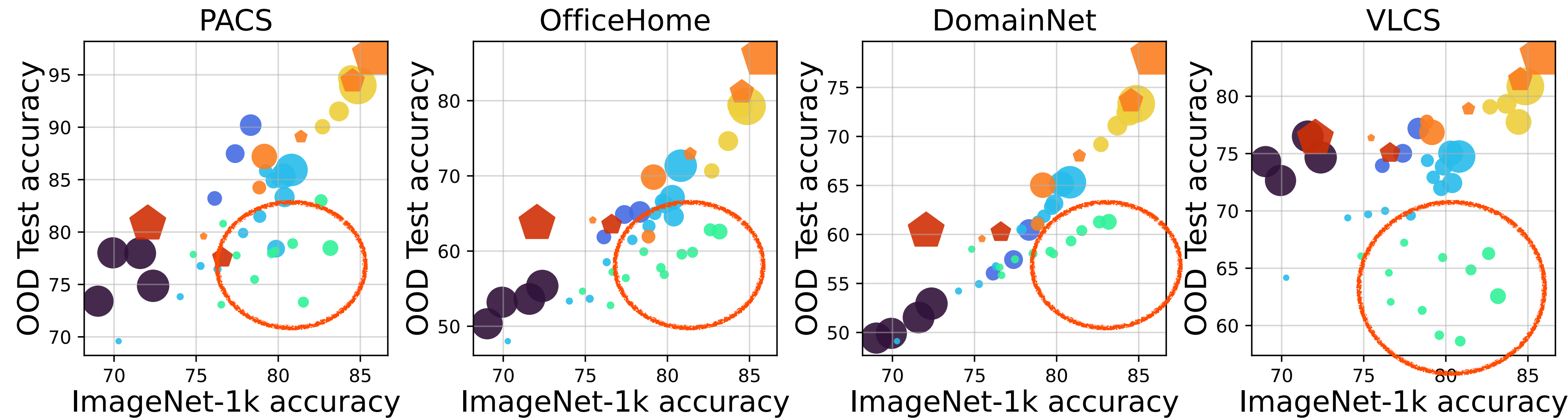


モデルのパラメータ数・事前学習データ数による「スケール則」

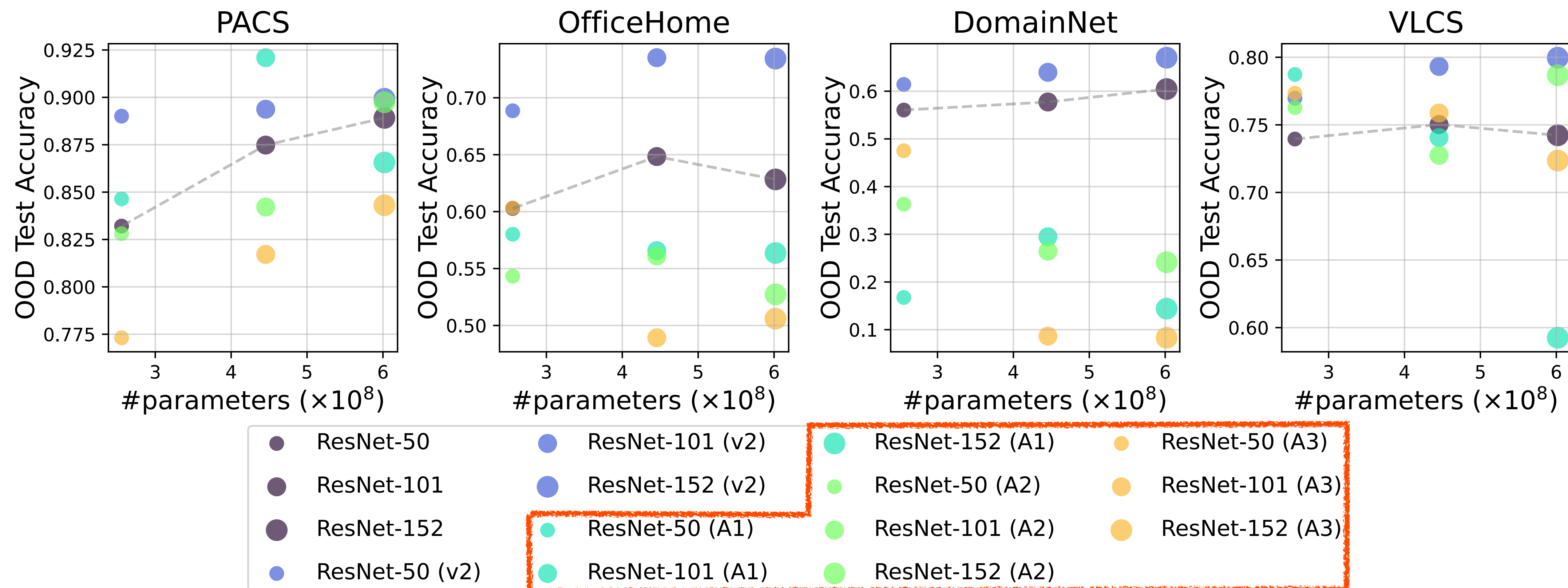


→ 大きなモデルを大きなデータセットで訓練することが重要

ImageNetに「過適合」したモデルが存在する



ImageNetに「過適合」した事前学習方法が存在する



ResNet Strikes-backの改良された学習方法

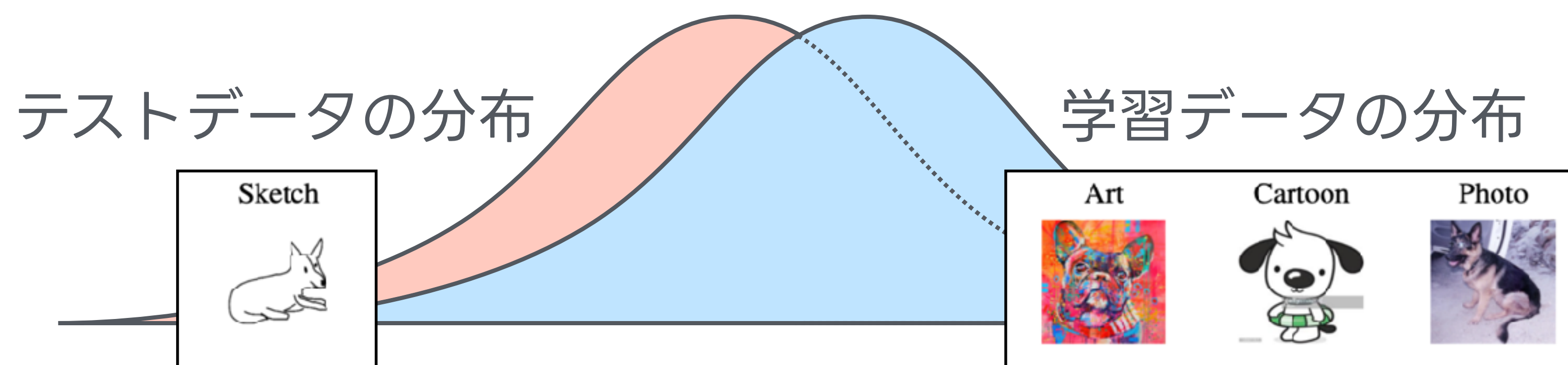
[Wightman et al. 21]

ImageNetでの性能向上が必ずしも分布外汎化能力に寄与しないことも

まとめ

<https://arxiv.org/abs/2307.08187>

事前学習モデルの差異は**分布外汎化**にどのような影響を及ぼすのかを調査



モデルのパラメータ数・事前学習データ数による「スケール則」が存在

大きなモデルを大きなデータセットで事前学習しておくことが重要

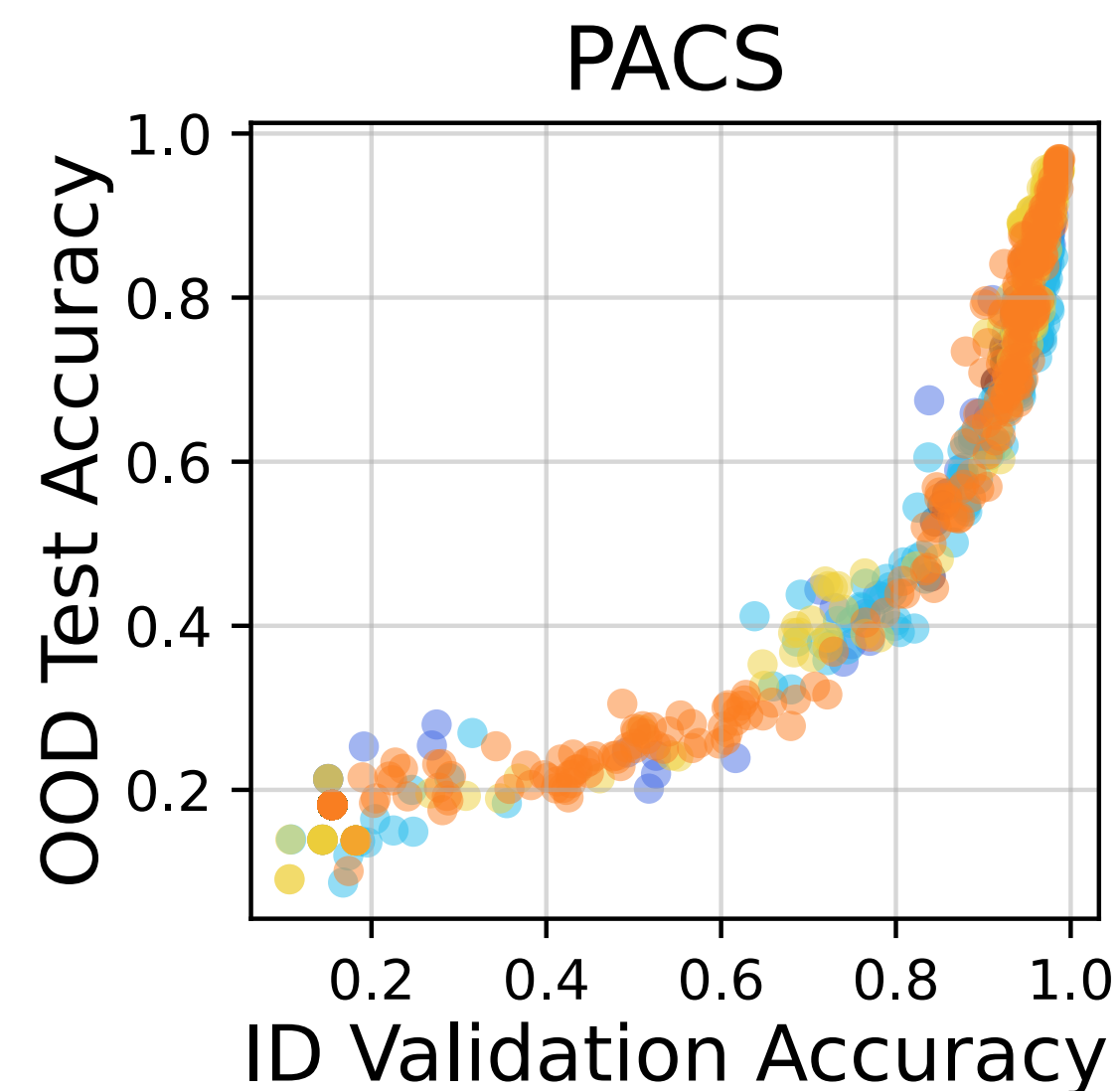
事前学習データへの「過適合」も見られる

「一般的」な事前学習済みモデルを使うことが重要

→ ファインチューニング後に焦点を当てたモデル開発も必要か

実験を進める上で有用だったこと

実験数が非常に多い



実験環境が各地にある

ABCI
MILA
理研AIP
⋮



wandbで実験結果を集約 → 実験結果をダウンロードして分析

実験終了後に結果をアップロード

pandas+matplotlib

ジョブの管理（長沼）と結果の分析（幡谷）で分担