The Curse of Low Task Diversity: On the Failure of Transfer Learning to Outperform MAML and Their Empirical Equivalence

Brando Miranda [1, 2] Patrick Yu [2] Yu-Xiong Wang [2] Sanmi Koyejo [1, 2]

[1]Computer Science, Stanford [2]University of Illinois Urbana-Champaign

Introduction and Motivation

Problem Statement : Recent work on meta-learning claims that transfer learning can beat most meta-learning algorithms. Without contextualizing claims, systematic comparisons, or data set analysis. Can we shed some light on this?

Goal :

- A **systematic comparison** of meta-learning and transfer learning
- A **fair comparison** of meta-learning and transfer learning
- **Contextualize claims** with an emphasis on a **data centric analysis** that **quantifies the intrinsic diversity of the data**

Our contributions are summarized as follows:

1. We propose a novel metric that quantifies the **intrinsic diversity** of the data of a few-shot learning benchmark – the *diversity coefficient*.
2. We show that two of the most prominent few-shot learning benchmarks – MiniImagenet and Cifar-fs – **have diversity is low**.
3. We contextualize and clarify past results and show that **Transfer Learning with USL does not outperform MAML** under a fair comparison

Background: MAML, Transfer Learning and Few-Shot-Learning

Model-Agnostic Meta-Learning (MAML) : attempts to meta-learn an initialization for a neural network that is primed for fast SGD adaptation:

- $f_{\hat{\theta}_{MAML}} = \min_\theta \sum_{\tau_i \in \mathcal{T}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta - \alpha \nabla \mathcal{L}_{\mathcal{T}_i(f_\theta)}})$

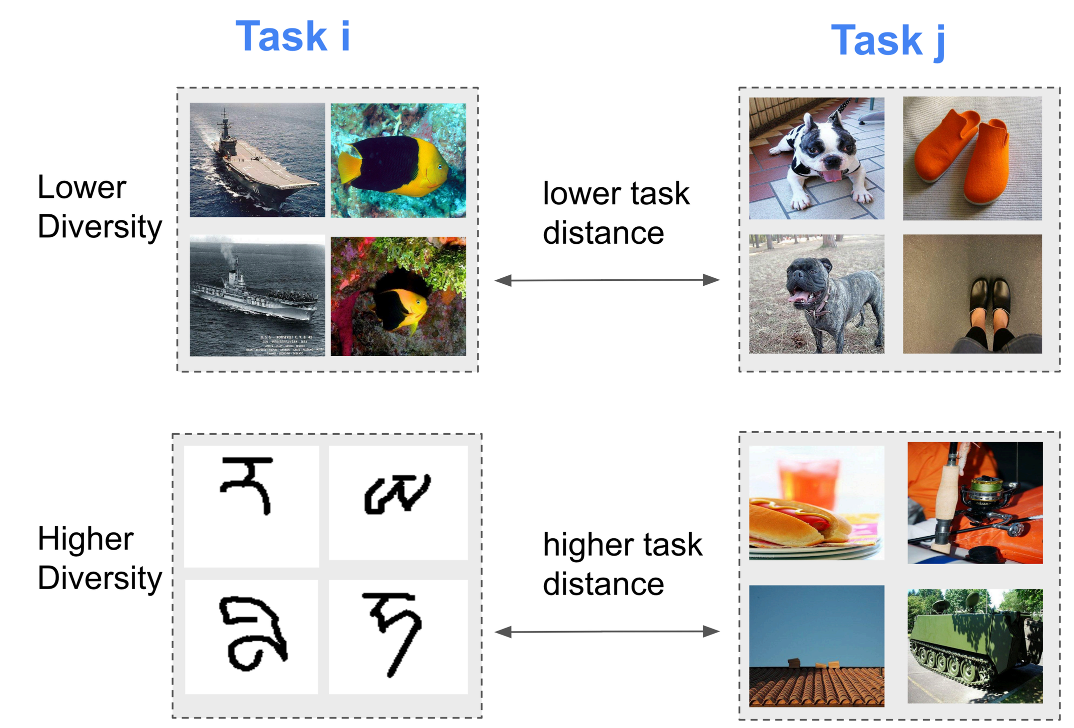
Transfer Learning with Union Supervised Learning (USL) :

1. Pre-train with a union of all classes: $f_{\hat{\theta}_{USL}} = \min_\theta \mathcal{L}_{USL}(\cup_{\tau_i \in \mathcal{T}} \tau_i, W_{cls} f_\theta)$ [USL]
2. At test tine fine-tune final layer: $f(x) = \hat{W}_{cls} f_{\theta_S L}(x)$ s.t. $\hat{W}_{cls} = \min_{W_{cls}} \mathcal{L}_{\mathcal{T}_i}(\tau_i, W_{cls} f_\theta)$ [USL]

Standard n-way, k-shot few-shot classification task:

Motivation for Diversity

Motivation: Intuitively, if a few-shot learning data set is not diverse (i.e. no large difference in tasks) – then there is little reason to adapt or perhaps meta-learn.



Formal Definition of Diversity

Definition: Therefore, the definition of few-shot learning data set captures some notion of "total" distance between distributions of tasks. Therefore the proposed **diversity coefficients**:

- Ground Truth Diversity Coefficient:

$$div(B) = \mathbb{E}_{\tau_1 \sim p(\tau|B), \tau_2 \sim p(\tau|B): \tau_1 \neq \tau_2}[d(p(x_1, y_1 \mid \tau_1), p(x_2, y_2 \mid \tau_2))]$$

- Diversity Coefficient on Real Data with Task Embeddings:

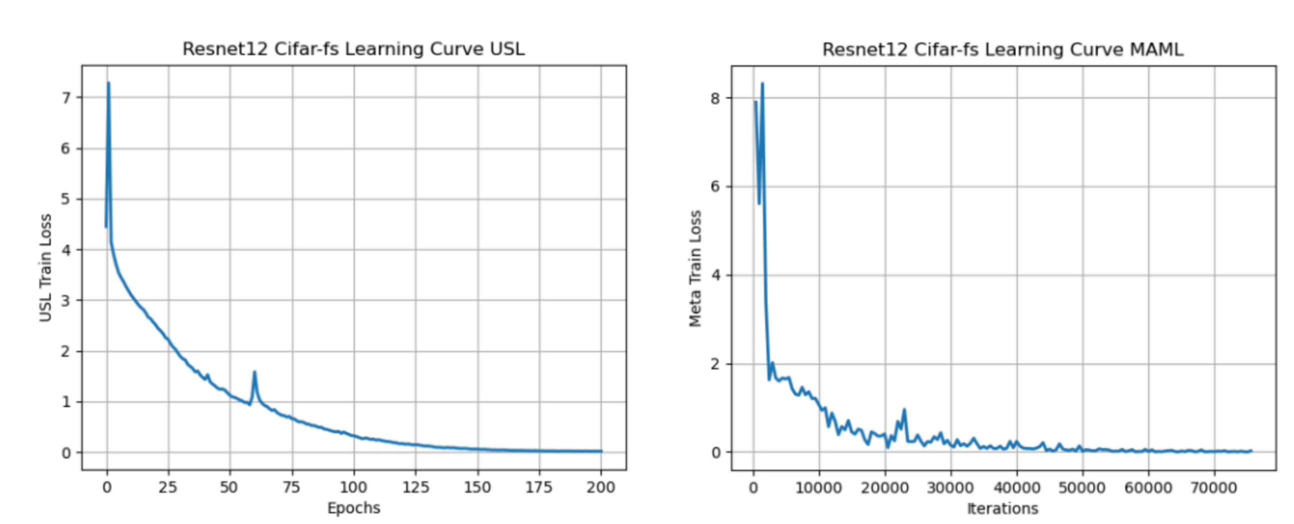
$$\hat{div}(B) = \mathbb{E}_{\tau_1 \sim \hat{p}(\tau|B), \tau_2 \sim \hat{p}(\tau|B): \tau_1 \neq \tau_2} \mathbb{E}_{D_1 \sim \hat{p}(x_1, y_1|\tau_1), D_2 \sim \hat{p}(x_2, y_2|\tau_2)} \left[d(\hat{F}_{D_1, f_w}, \hat{F}_{D_2, f_w})\right]$$

Where F_{D_τ, f_w} is the embedding of task τ with the Task2Vec method – which is the diagonal of the Fish Information Matrix (FIM) of the data set D from task τ with a fixed probe network f_w.

Method: Fair Comparison

Compute diversity, and compare performance (accuracy) fairly i.e.:

- Use same architecture
- Use same optimizer
- All models trained to convergence

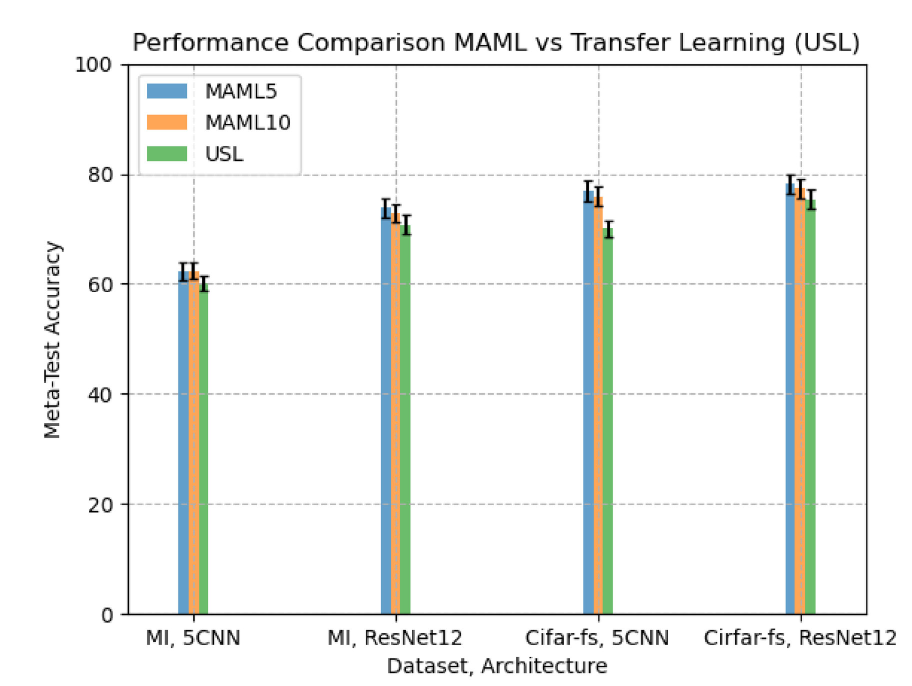


Results 1: Low Diversity Computations

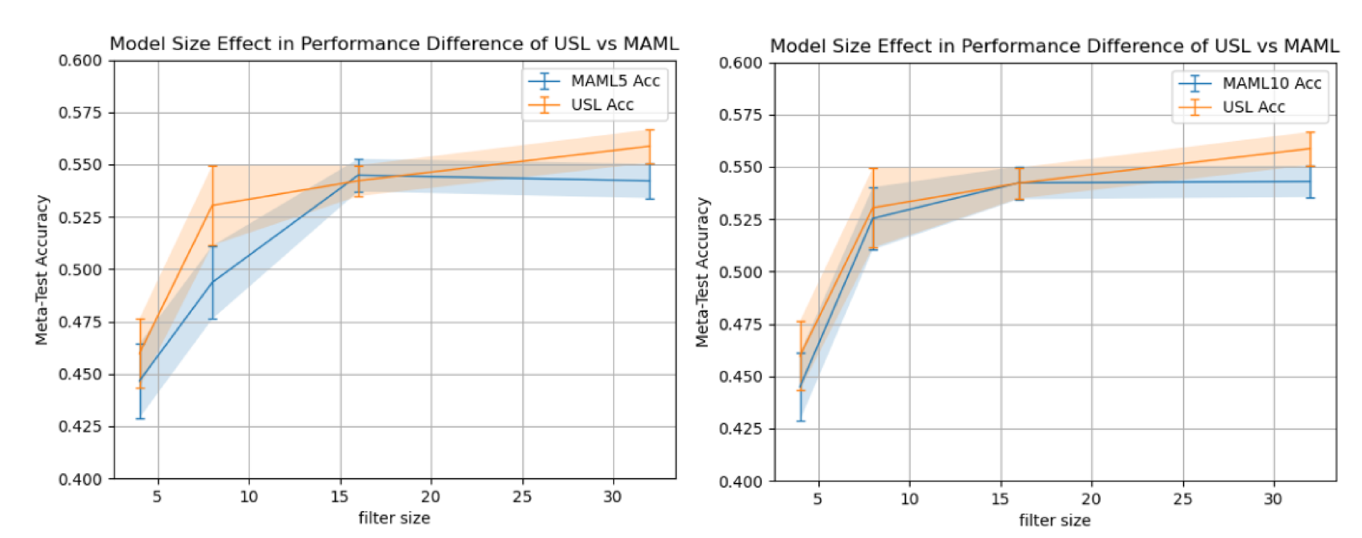
Probe Network	Diversity on MI	Diversity on Cifar-fs
Resnet18 (pt)	$0.117 \pm 2.098e\text{-}5$	$0.100 \pm 2.18e\text{-}5$
Resnet18 (rand)	$0.0955 \pm 1.29e\text{-}5$	$0.103 \pm 1.05e\text{-}5$
Resnet34 (pt)	$0.0999 \pm 1.95e\text{-}5$	$0.0847 \pm 3.06e\text{-}5$
Resnet34 (rand)	$0.0620 \pm 8.12e\text{-}6$	$0.0643 \pm 9.64e\text{-}6$

MI = "Mini-Imagenet"

Results 2: Transfer Learning with USL doesn't outperform MAML



Results 3: USL doesn't outperform MAML even as Model Size Changes



Conclusions

- Under a **fair comparison**
- And in the **low diversity regime**
- Transfer Learning with USL cannot outperform MAML