# Understanding the Transferability of Representations via Task-Relatedness

Akshay Mehra, Yunbei Zhang, and Jihun Hamm

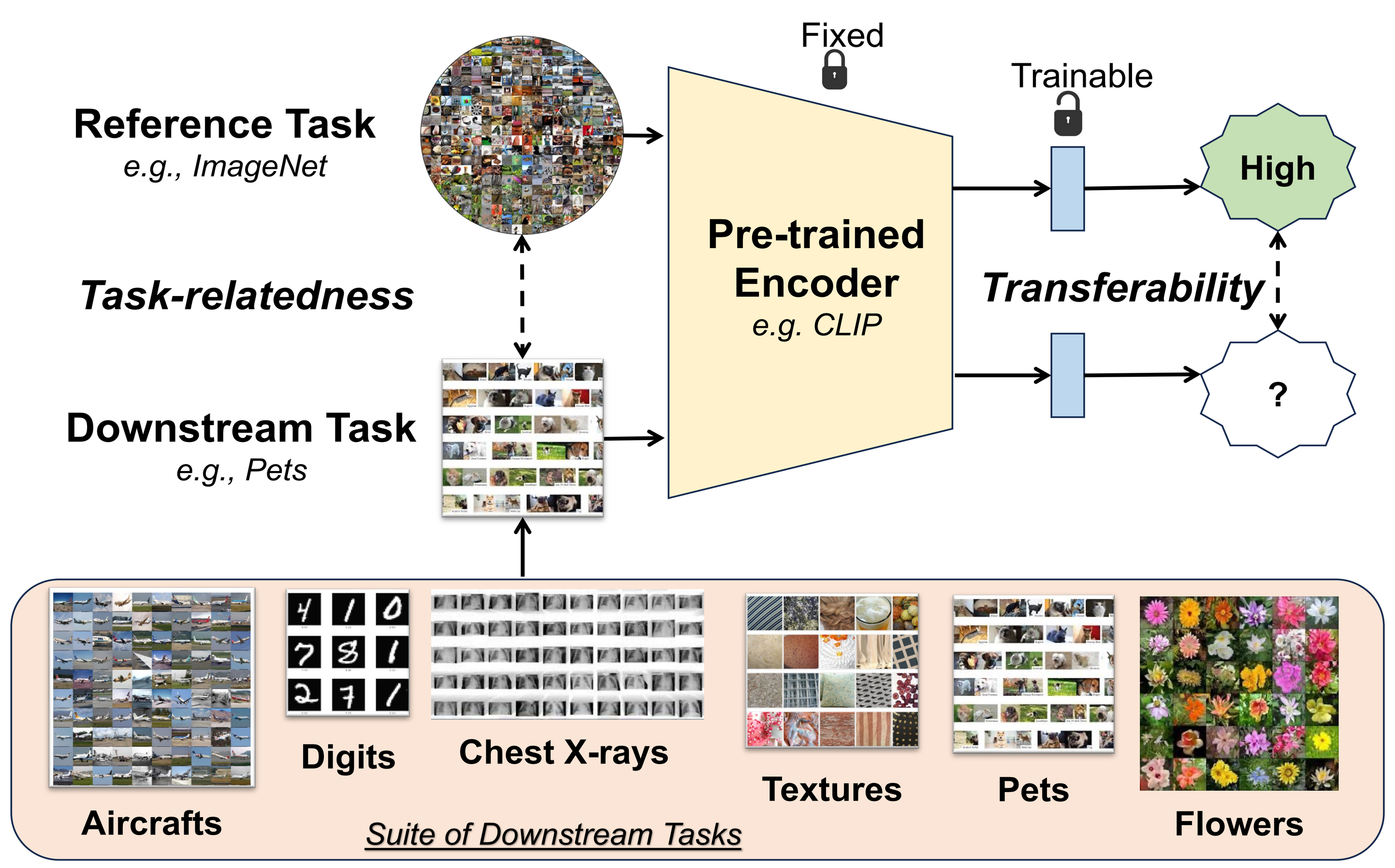
Tulane University

## Motivation

- Transfer learning enables using representations from large pre-trained models to learn high-performing ML models with limited data/compute.
- However, existing analysis are insufficient in explaining the conditions for achieving high transferability in cross-domain cross-task setting.

## Key Question

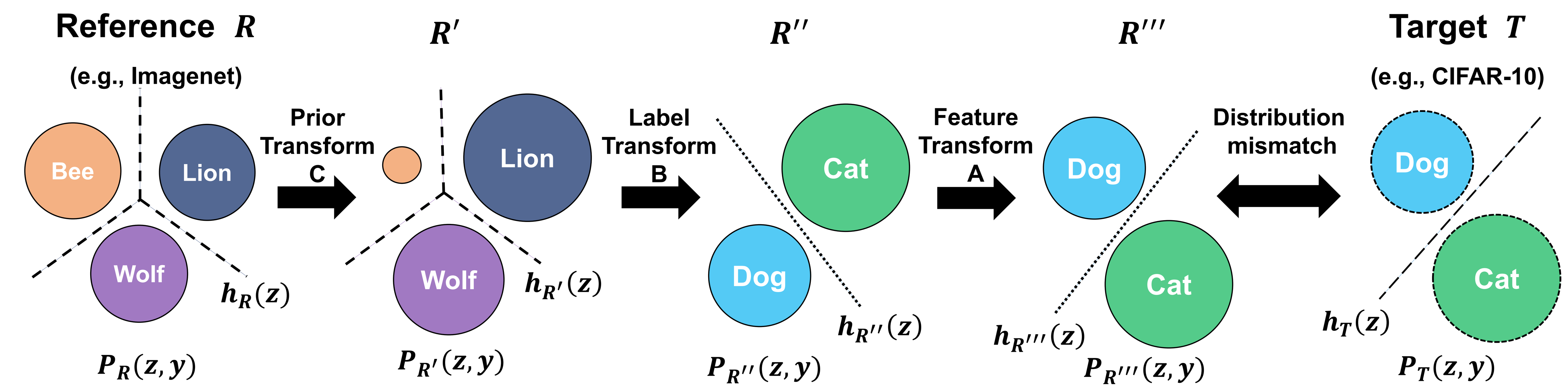
- Given a pre-trained encoder (e.g., CLIP), how does the performance after fine-tuning it on a reference task (e.g., ImageNet) relate to the performance after fine-tuning it on other downstream tasks?




## Contributions

- We propose an analysis that explains transferability in terms of relatedness between the reference and target tasks.
- We show that task-relatedness is efficiently computable with limited target data (even without target labels) and is predictive of the accuracy after end-to-end fine-tuning of pre-trained encoder on target tasks.

## Task Transfer Analysis

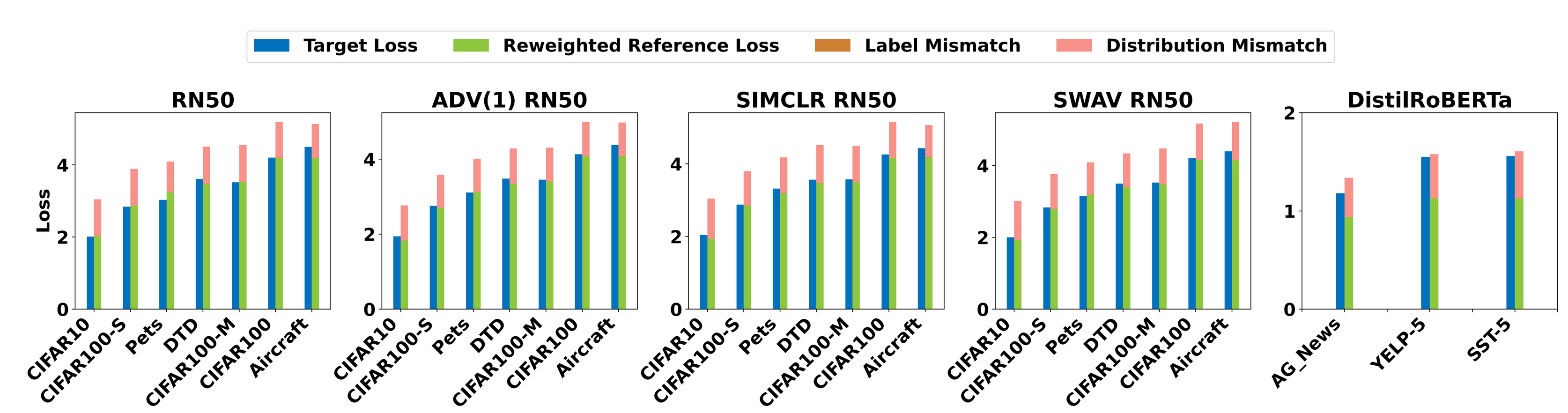



- A series of transformations (parametrized by $A, B, C$) are applied to the reference task distribution $P_R(z,y)$ and its classifier $h_R$ to produce the transformed distribution $P_{R'''}(z,y)$ and classifier $h_{R'''}$ to explain transferability to the downstream target task with distribution $P_T(z,y)$ and classifier $h_T$.
- Let $A: Z \to Z$ be an invertible linear map, $B$ be a $|Y_T| \times |Y_R|$ matrix such that $B_{ij} = P(y_{R''} = i | y_{R'} = j)$, $C := \left[\frac{P_{R'}(y)}{P_R(y)}\right]_{y=1}^{|Y_R|}$ be a vector of probability ratios, and assumption 1 (omitted here) holds then transferability (LHS) is provably explained by task-relatedness (RHS). Here $H$ is the conditional entropy, $W_d$ is the Wasserstein distance, $d\big((z,y),(z',y')\big) := \|z - z'\|_2^2 + \infty \cdot 1_{y \neq y'}$ is the base distance, $\tau$ is the Lipschitz constant, and $\ell$ is the cross-entropy loss.

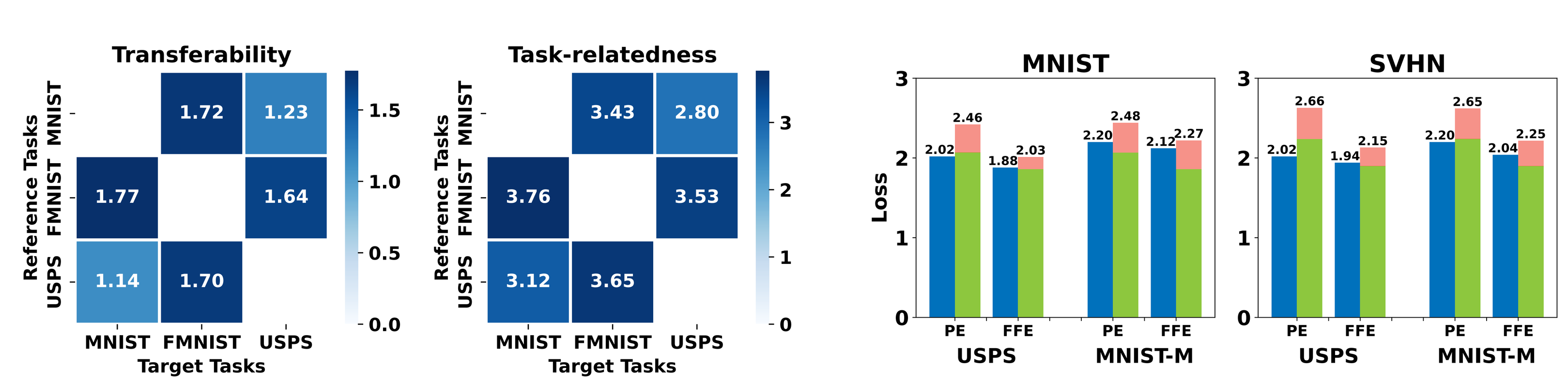
$$\underbrace{\mathbb{E}_{P_T}[\ell(h_T(z),y)]}_{\text{Transferability}} \leq \underbrace{\mathbb{E}_{P_R}[C(y)\ell(h_R(z),y)]}_{\text{Re-weighted reference loss}} + \underbrace{H(Y_{R''}|Y_{R'})}_{\text{Label Mismatch}} + \underbrace{\tau \cdot W_d(P_{R'''},P_T)}_{\text{Distribution Mismatch}}.$$

## Empirical Analysis

- Task-relatedness produces a small gap to transferability.



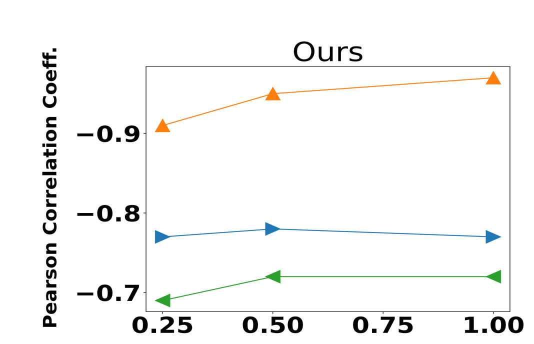

- (a) Task-relatedness and transferability are highly correlated across various reference-target pairs.
  (b) Improving the transferability of an encoder on a reference task leads to improved transferability of all related target tasks.




- Task-relatedness achieves high (negative) Pearson correlation to the accuracy after end-to-end fine-tuning for various tasks.

| Target task | LogMe | Leep | NCE | PACTran | SFDA | H-Score | OT-NCE | OTCE | Ours |
|---|---|---|---|---|---|---|---|---|---|
| Pets | 0.82 | 0.80 | 0.73 | -0.82 | 0.57 | 0.77 | 0.88 | 0.86 | -0.77 |
| DTD | 0.88 | 0.96 | -0.19 | -0.85 | 0.90 | 0.89 | 0.84 | 0.82 | -0.97 |
| Aircraft | -0.60 | 0.92 | 0.97 | 0.11 | 0.72 | -0.80 | 0.56 | 0.60 | -0.72 |
| **Average** | 0.37 | 0.90 | 0.50 | -0.52 | 0.73 | 0.29 | 0.76 | 0.76 | -0.82 |

- (a) Task-relatedness remains highly correlated with accuracy after end-to-end fine-tuning on a target task even with using (a) a small percentage of target data,
  (b) no target labels ($y_T^{pseudo} = \arg\max_{y \in \mathcal{Y}_T} Bh_R(A^{-1}(z_T))$).




| Target | True labels | Pseudo labels |
|---|---|---|
| Pets | -0.77 | -0.76 |
| DTD | -0.97 | -0.91 |
| Aircraft | -0.72 | -0.16 |

This work was supported by the NSF EPSCoR-Louisiana Materials Design Alliance (LAMDA) program #OIA1946231.