Let Go of Your Labels with Unsupervised Transfer

Artyom Gadetsky*, Yulun Jiang*, Maria Brbić

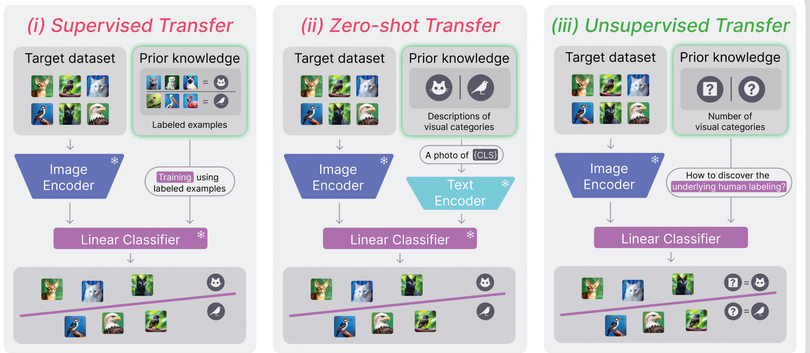
ICML International Conference On Machine Learning

EPFL

Project page

Motivation

Given pre-trained foundation models, current downstream transfer settings require supervision.



How can we enable unsupervised transfer to solve a new task in a fully unsupervised manner?

Theoretical insight

Generalization-based learning of human labelings (Gadetsky and Brbić. *NeurIPS* '23)

Key observation:

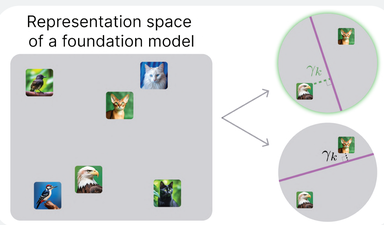
Generalization ability of a linear model on top of pre-trained representations provides a measure of disagreement between any given labeling and the underlying human labeling.

$$\min_{\tau} \sum_{x \in \mathcal{D}_{te}} \mathcal{L}(\hat{f}(x), \tau(x))$$
$$\text{s.t. } \hat{f} = \arg\min_{f} \sum_{x \in \mathcal{D}_{tr}} \mathcal{L}(f(x), \tau(x))$$

Emergence of Unsupervised Transfer

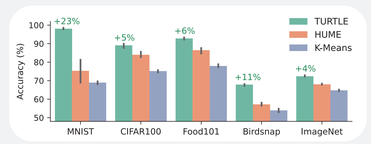
Our key result:

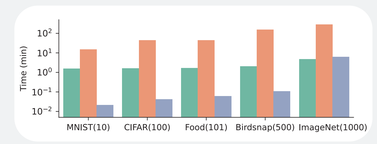
Unsupervised transfer emerges when searching for labelings that induce maximum margin classifiers in representation spaces of foundation models.



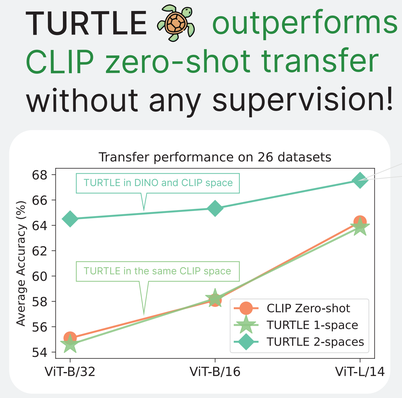
Experimental Results

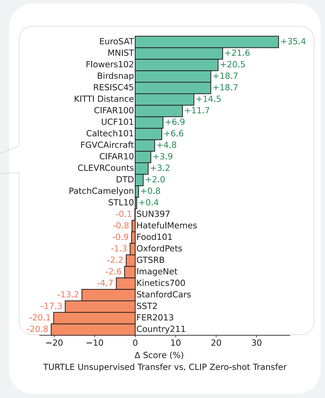
TURTLE achieves state-of-the-art unsupervised performance and is highly scalable!





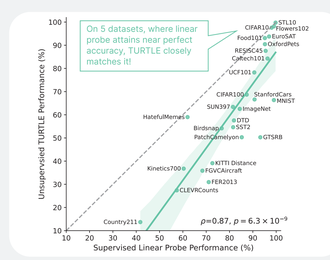
TURTLE outperforms CLIP zero-shot transfer without any supervision!





TURTLE outperforms unsupervised prompt tuning methods that aim to enhance CLIP zero-shot transfer!

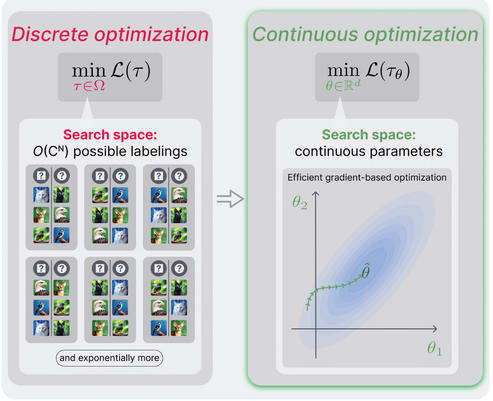
Method	ZS	Pets	Flowers	FGVC	DTD	EuroSAT	Cars	Food	SUN	Caltech	UCF	ImageNet	Avg.
POUF	✓	88.0	66.7	16.7	41.5	42.1	57.4	74.7	58.6	86.9	61.1	55.2	59.0
UPL	✓	88.3	68.9	17.3	46.6	54.8	**62.1**	77.6	64.0	**89.9**	67.2	60.5	63.4
GDA	✓	89.9	72.7	18.7	46.8	49.9	60.8	78.3	63.6	87.5	68.7	61.2	63.5
TURTLE	✗	**90.9**	**99.7**	**25.3**	**57.0**	**95.5**	32.6	**84.1**	**65.7**	88.6	**77.7**	**66.3**	**71.2**

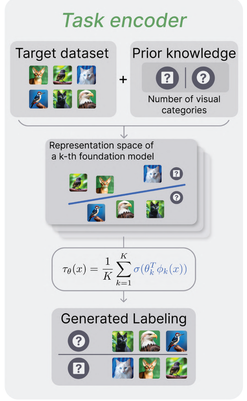


TURTLE can infer "optimal" classifier without supervision given high-quality representations!

Our approach: TURTLE

To avoid combinatorial search over all possible labelings of a dataset we define a task encoder and optimize over its continuous parameters.





Optimization procedure alternates between
(i) proposing data labeling via task encoder
(ii) assessing data labeling by training a linear classifier and evaluating its margin size