The Pursuit of Human Labeling
A New Perspective on Unsupervised Learning

EPFL

Artyom Gadetsky & Maria Brbić

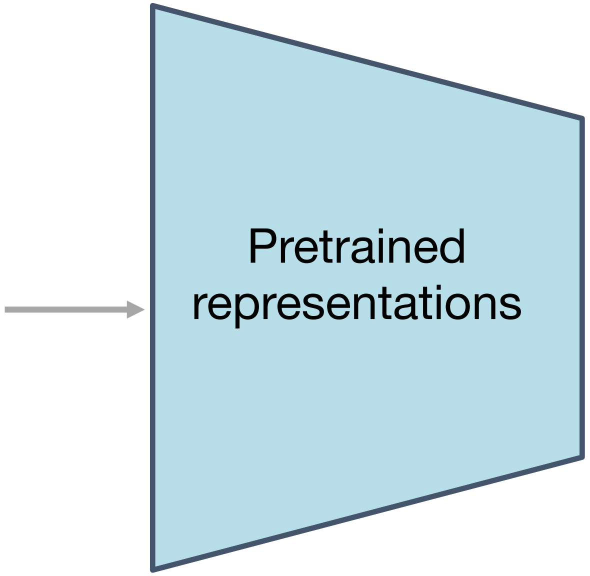
NEURAL INFORMATION PROCESSING SYSTEMS

Supervised Fine-tuning

Labels are given

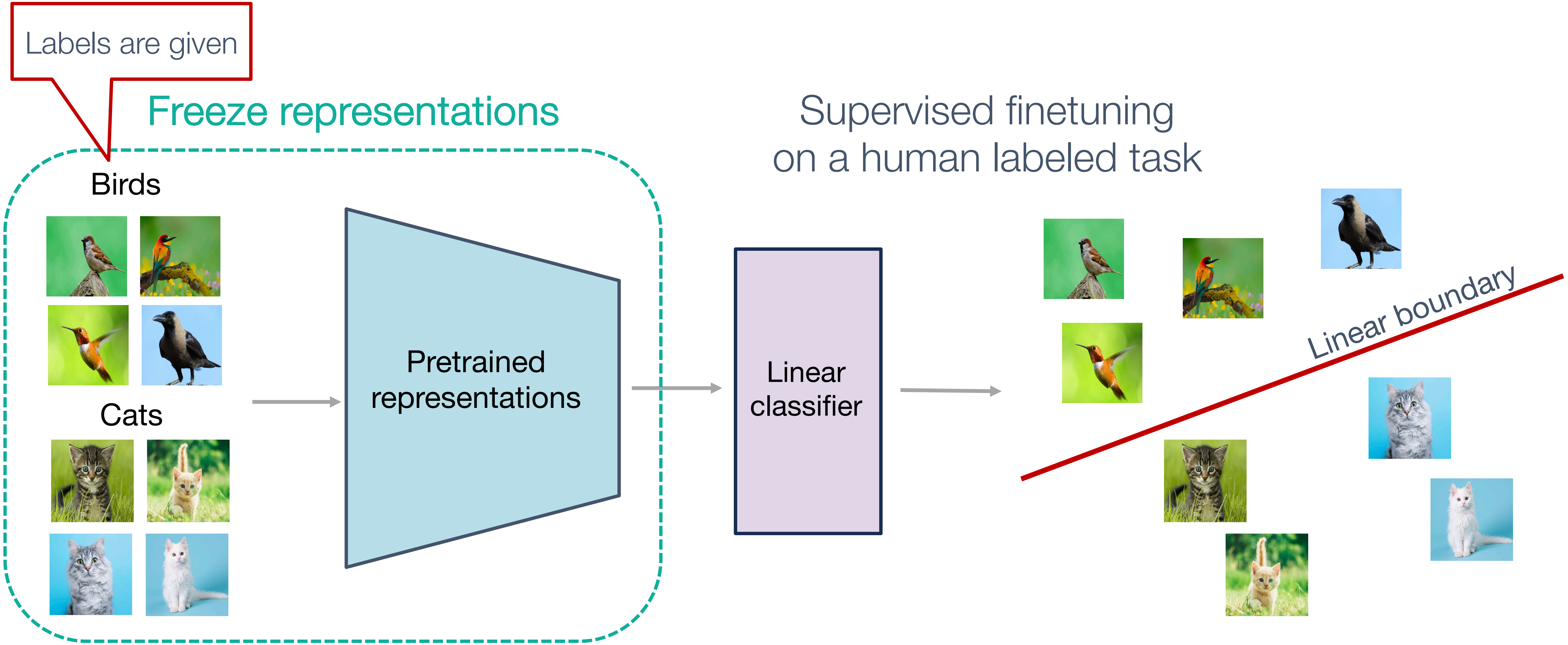
Birds

Cats



Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. *ICML 2020.*

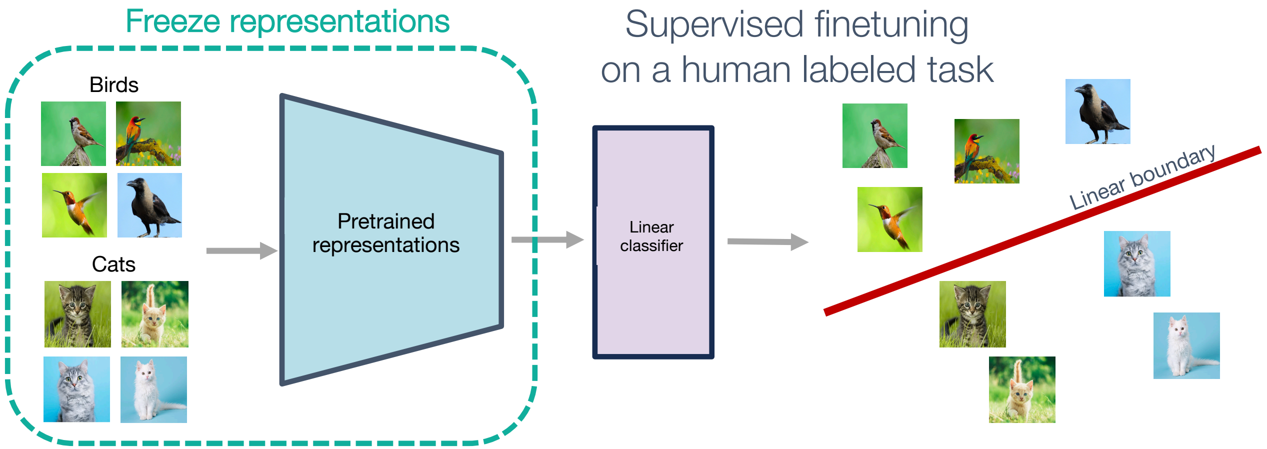
Supervised Fine-tuning



Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. *ICML 2020.*

Assessing Generalization of Supervised Fine-tuning

Train on the training split



Assess generalization on held-out data

Fine-tuning linear classifiers in modern representation spaces achieves great generalization, but requires supervision

Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. *ICML 2020.*

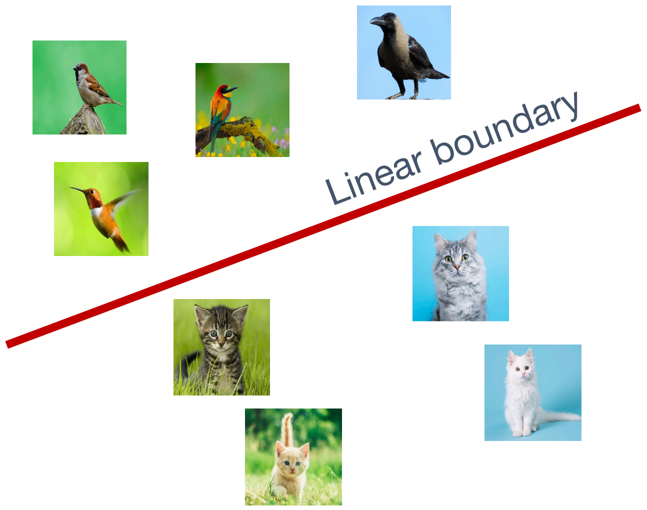
Can we use this paradigm for
unsupervised inference
of human labelings?

What Makes Human Labeled Tasks Special?

- **Observation 1:**

  Many human labeled tasks are linearly separable in a sufficiently strong representation space, e.g., CLIP, DINO and other spaces of foundation models

  Can we just search for a linearly separable task to recover underlying human labeling?



Oquab et al. DINOv2: Learning Robust Visual Features without Supervision. *TMLR 2023 (under review).*

Radford et al. Learning Transferable Visual Models from Natural Language Supervision. *ICML 2021.*

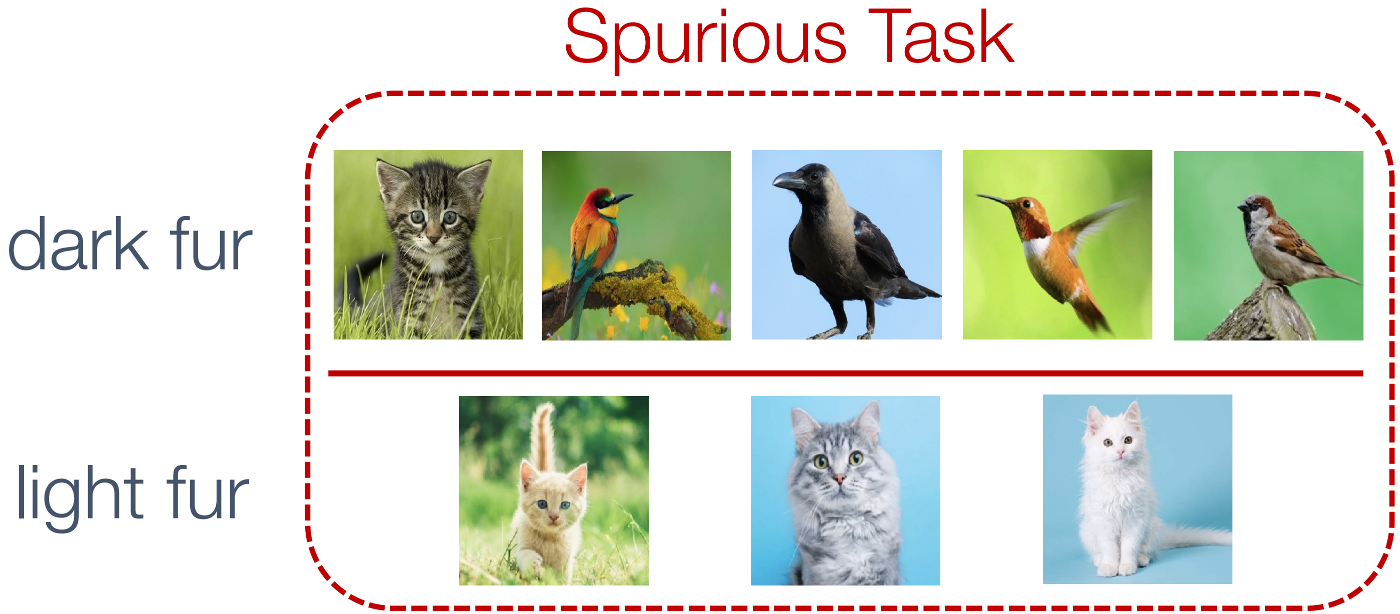
Inductive Biases of Representations

However, one dataset allows for many generalizable tasks which reflect the inductive biases of representations used to represent the dataset

Atanov et al. Task Discovery: Finding the Tasks that Neural Networks Generalize on. *NeurIPS 2022.*

Inductive Biases of Representations

However, one dataset allows for many generalizable tasks which reflect the inductive biases of representations used to represent the dataset



Atanov et al. Task Discovery: Finding the Tasks that Neural Networks Generalize on. *NeurIPS 2022.*

Inductive Biases of Representations

However, one dataset allows for <u>many generalizable tasks</u> which reflect the <u>inductive biases of representations</u> used to represent the dataset

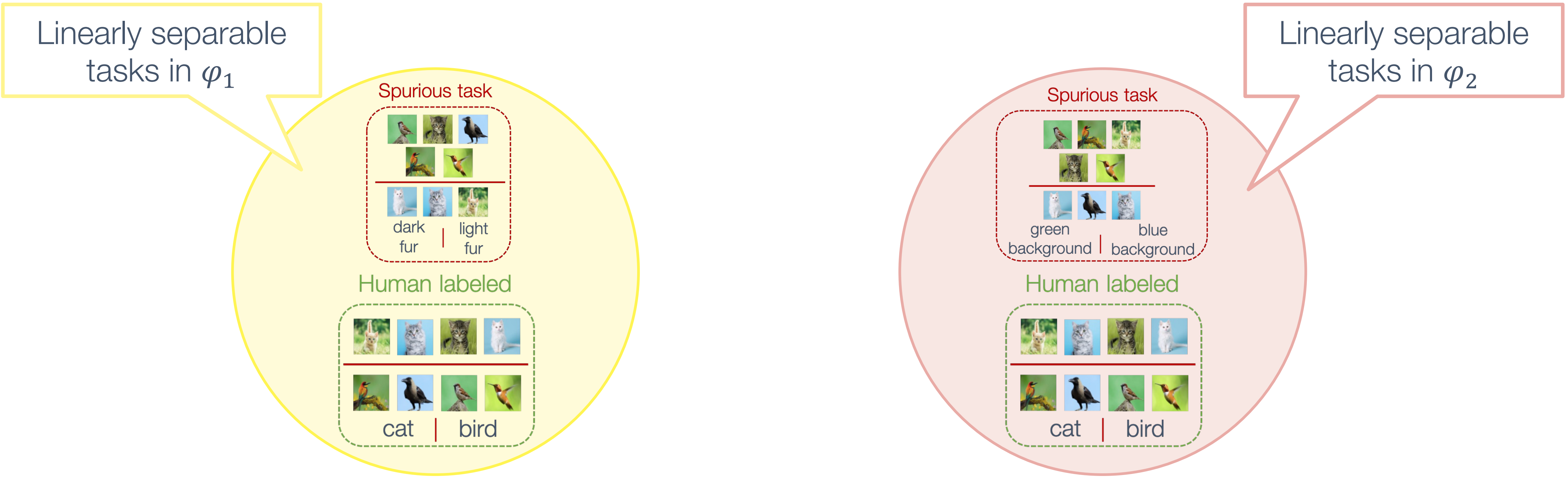
Spurious Task

Atanov et al. Task Discovery: Finding the Tasks that Neural Networks Generalize on. *NeurIPS 2022.*

What Makes Human Labeled Tasks Special?

- **Observation 2:**

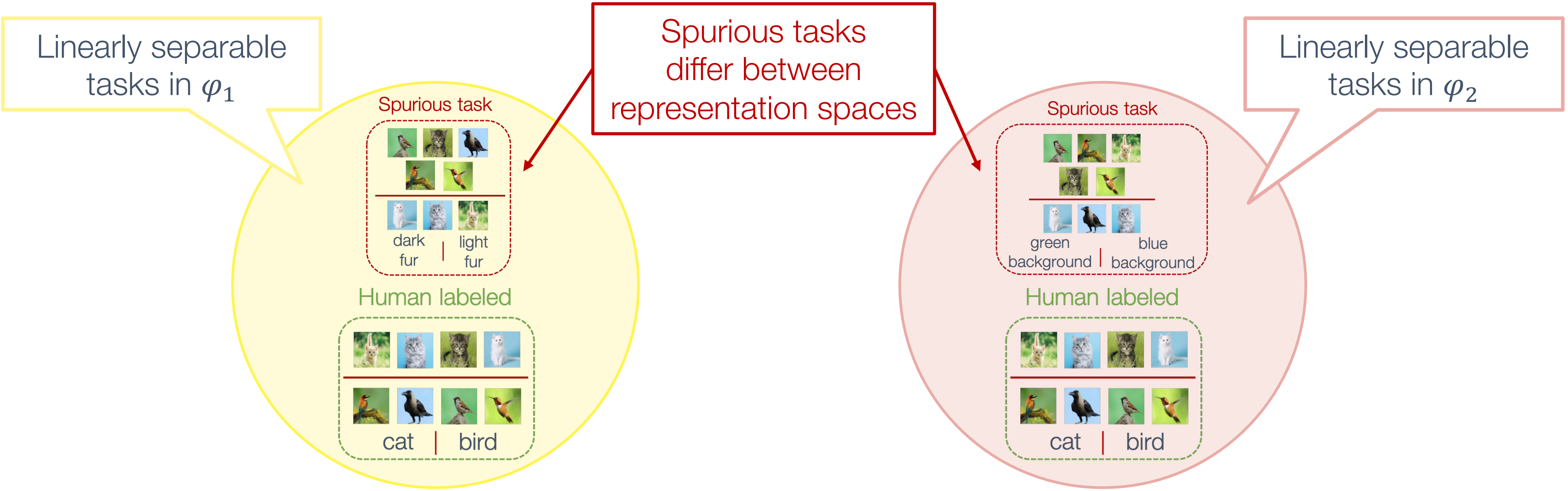
  Despite each representation space has its own inductive biases, human labeled tasks are invariant to the underlying representation space



What Makes Human Labeled Tasks Special?

- Observation 2:

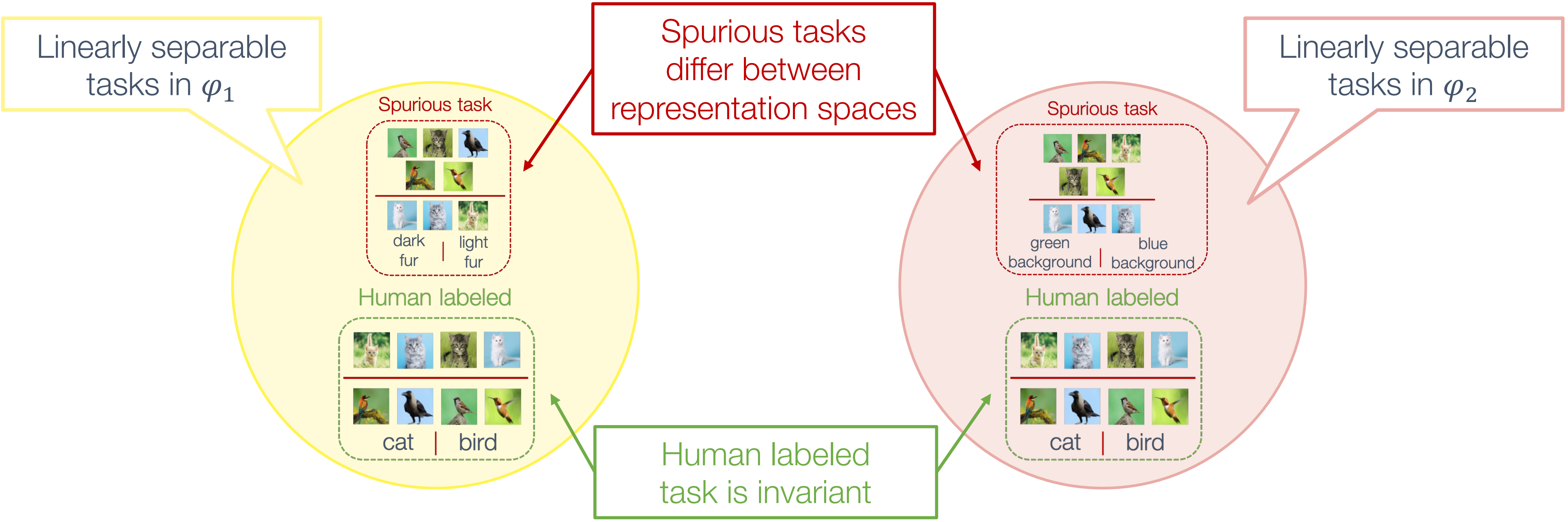
  Despite each representation space has its own inductive biases, human labeled tasks are invariant to the underlying representation space



What Makes Human Labeled Tasks Special?

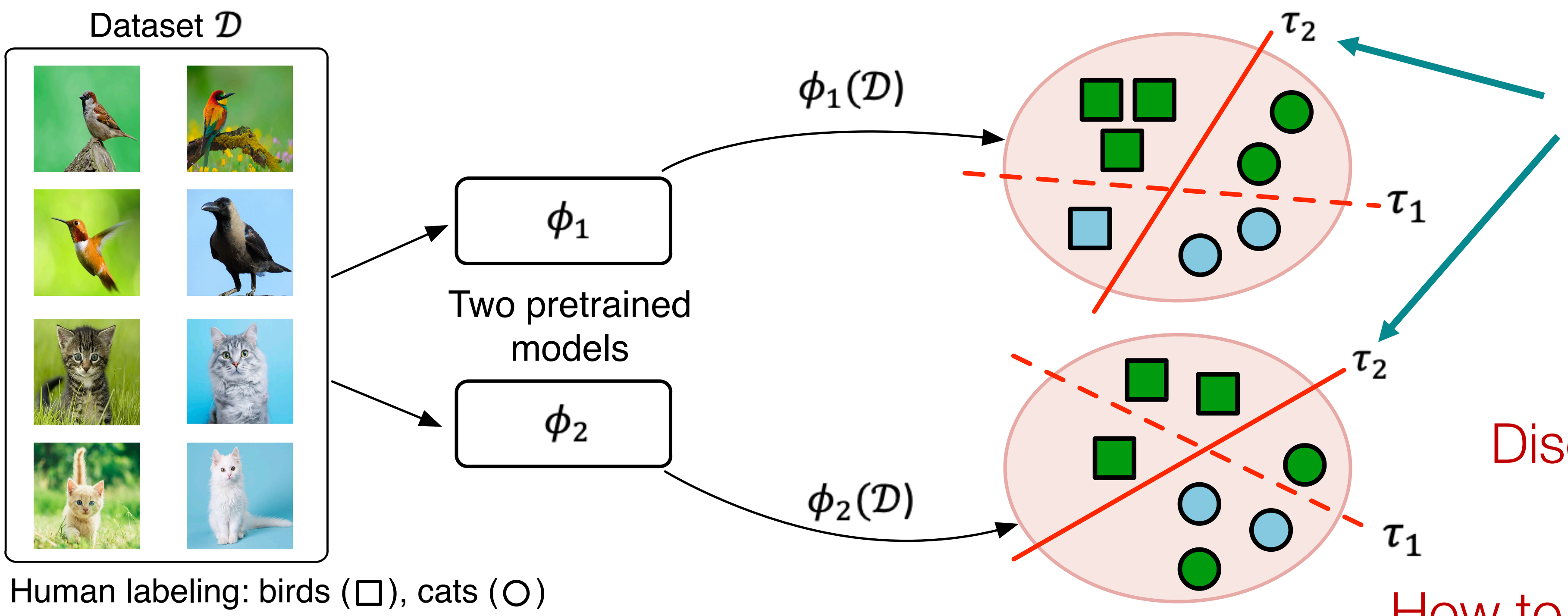
- Observation 2:

  Despite each representation space has its own inductive biases, human labeled tasks are invariant to the underlying representation space



HUME: Discovering Human Labeled Tasks

Key Idea: Search for the task which attains low generalization error <u>simultaneously</u> in different representation spaces

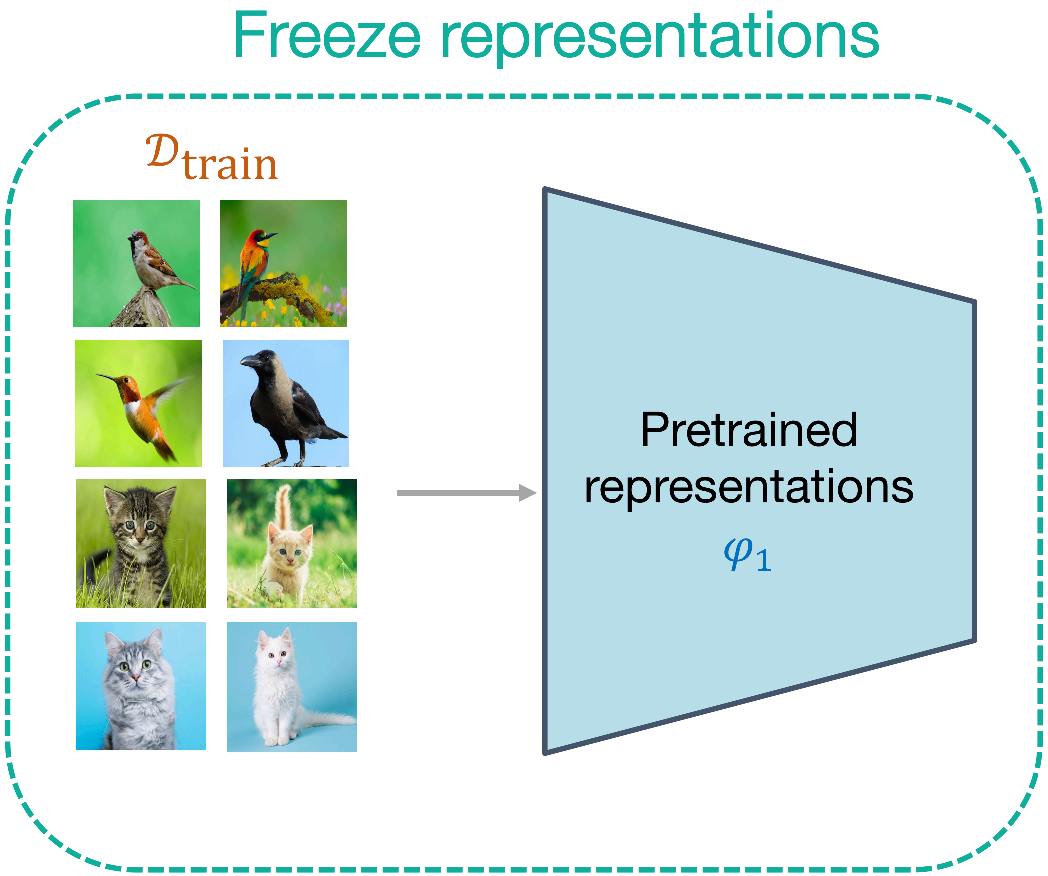


Task with low generalization error in both representation spaces

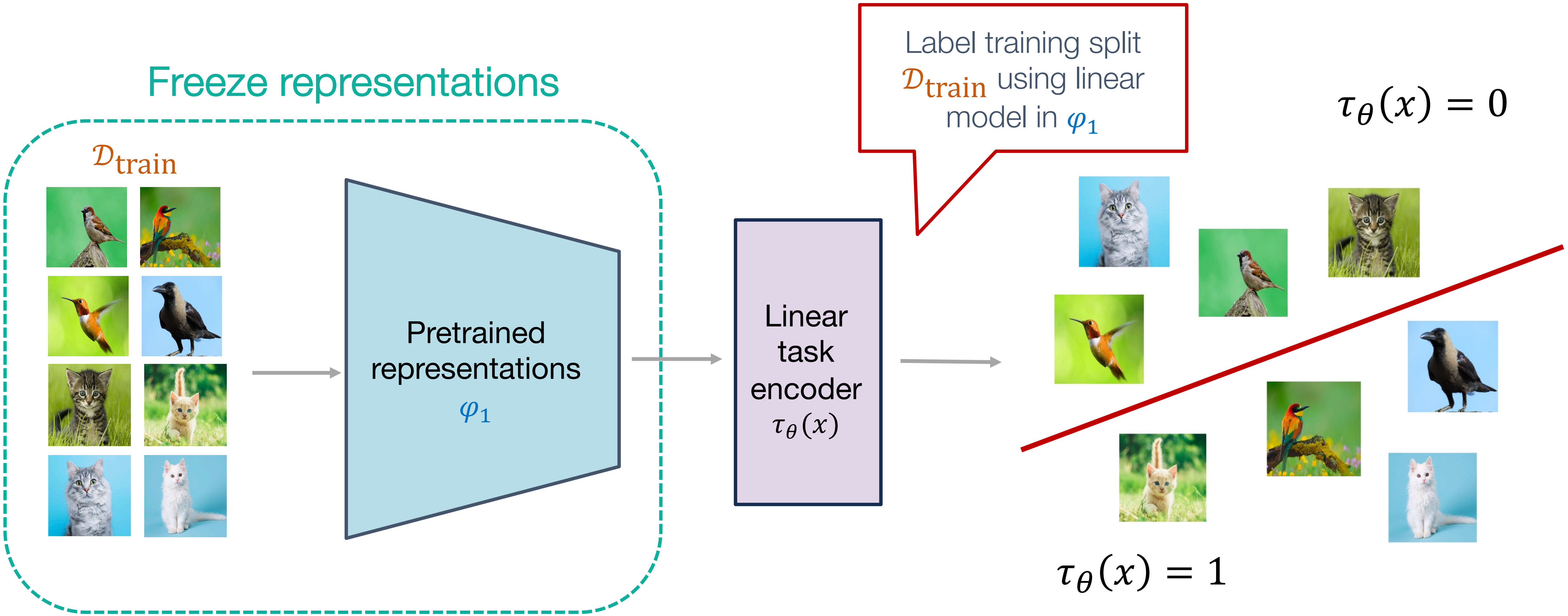
Discrete optimization problem 😱

How to solve it in practice?

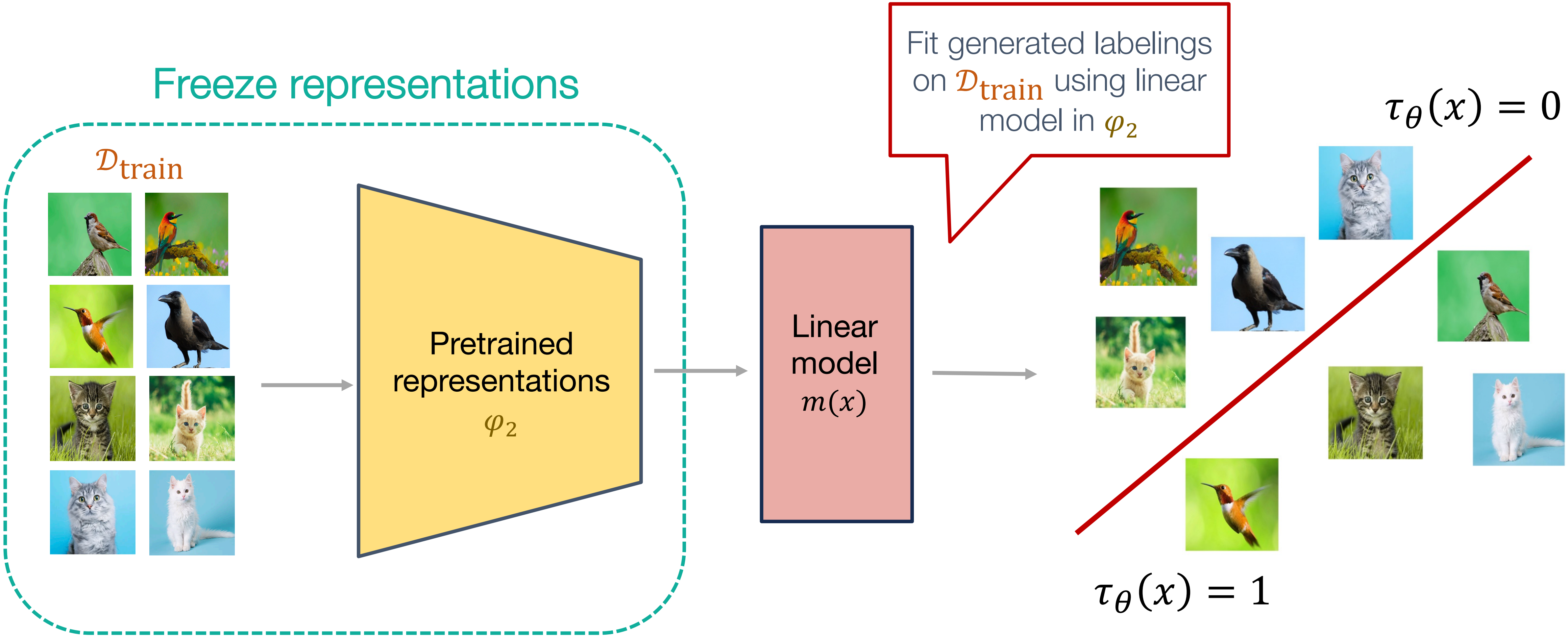
HUME: From Idea to Method



HUME: From Idea to Method

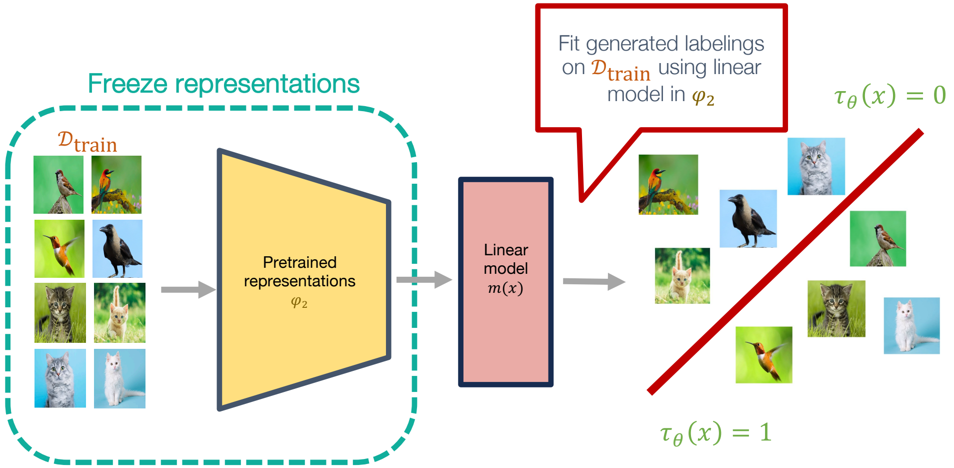


HUME: From Idea to Method



HUME: From Idea to Method

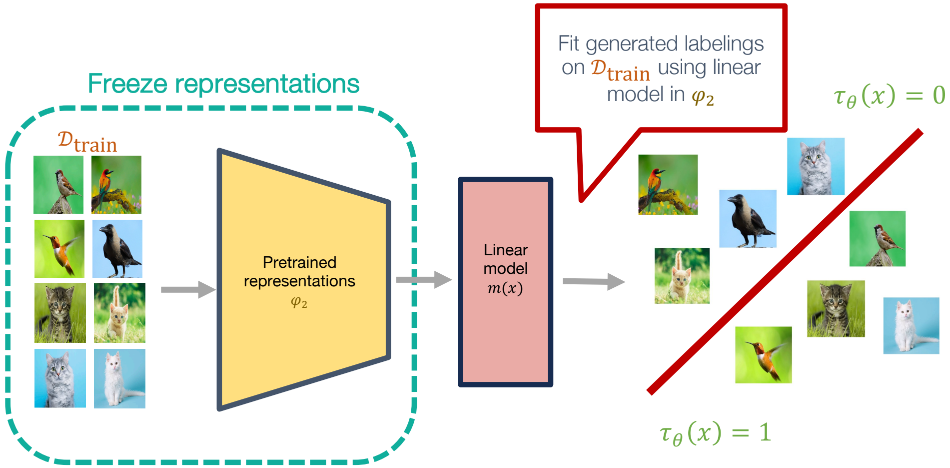
Train on the training split $\mathcal{D}_{\text{train}}$ with labeling $\tau_\theta(x)$ to get $m^*(x)$



HUME: From Idea to Method

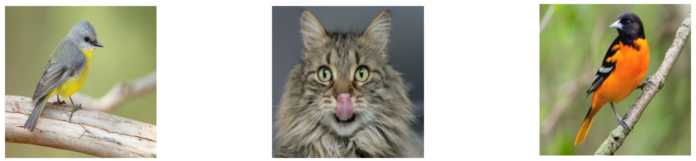
Train on the training split $\mathcal{D}_{\text{train}}$ with labeling $\tau_\theta(x)$ to get $m^*(x)$

Minimize generalization error of $m^*(x)$ w.r.t. labeling $\tau_\theta(x)$ on held-out $\mathcal{D}_{\text{test}}$



$m^*(x)$ might disagree with $\tau_\theta(x)$

$\mathcal{D}_{\text{test}}$

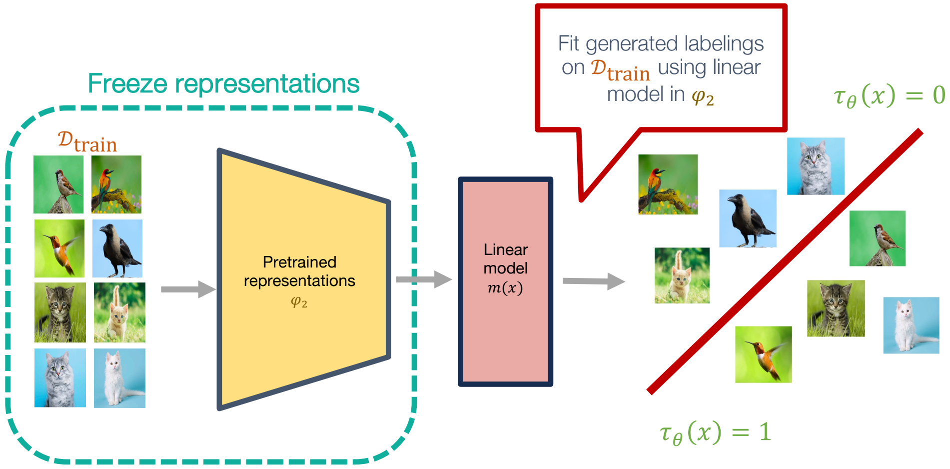


$m^*(x) = 0$	$m^*(x) = 1$	$m^*(x) = 1$
$\tau_\theta(x) = 1$	$\tau_\theta(x) = 1$	$\tau_\theta(x) = 0$

HUME: From Idea to Method

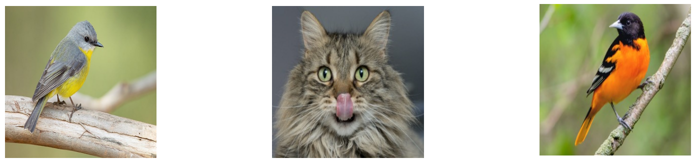
Train on the training split $\mathcal{D}_{\text{train}}$ with labeling $\tau_\theta(x)$ to get $m^*(x)$

Minimize generalization error of $m^*(x)$ w.r.t. labeling $\tau_\theta(x)$ on held-out $\mathcal{D}_{\text{test}}$



$m^*(x)$ might disagree with $\tau_\theta(x)$

$\mathcal{D}_{\text{test}}$



$m^*(x) = 0$	$m^*(x) = 1$	$m^*(x) = 1$
$\tau_\theta(x) = 1$	$\tau_\theta(x) = 1$	$\tau_\theta(x) = 0$

$$\min_{\tau_\theta} \mathcal{L}_{ce}(m^*(x);\ \tau_\theta(x))$$

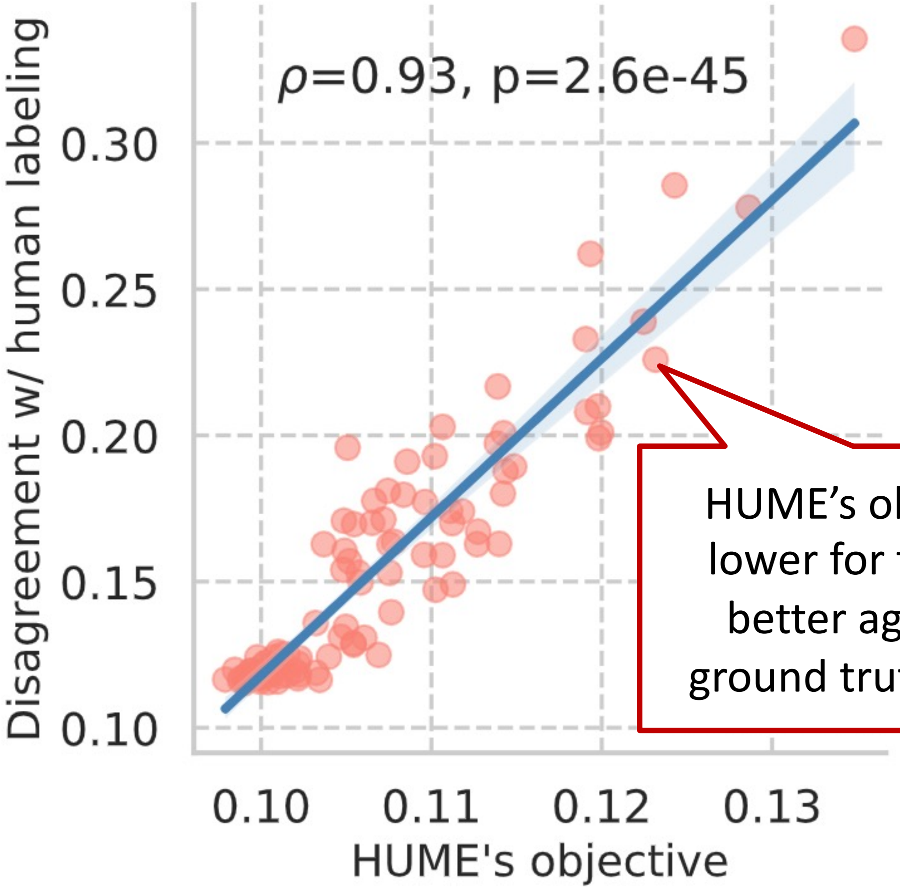
Standard cross-entropy

HUME: Agreement with Human Labeling

HUME's objective: generalization error of linear classifiers in different representation spaces

HUME only trains linear classifiers on top of pretrained models!

HUME's objective is strikingly well-correlated with human labeling



HUME's objective is lower for tasks that better agree with ground truth labeling

HUME Matches Supervised Learning

Supervised Linear Probe in φ_1 - MOCO trained on the target dataset

Method	STL-10 ACC	STL-10 ARI	CIFAR-10 ACC	CIFAR-10 ARI	CIFAR-100-20 ACC	CIFAR-100-20 ARI
MOCO Supervised Linear	88.9	77.7	**89.5**	79.0	**72.5**	**52.6**
HUME MOCO + BiT	90.3	80.5	86.6	75.0	48.8	34.5
HUME MOCO + CLIP	92.2	84.1	88.9	78.3	50.1	34.8
HUME MOCO + DINO	**93.2**	**86.0**	**89.2**	**79.2**	**56.7**	**39.6**

HUME:

φ_1 - MOCO trained on the target dataset
φ_2 - BiT, CLIP, DINO large foundation models

HUME matches the performance of supervised model while being fully-unsupervised!

HUME Outperforms Unsupervised Baselines

State-of-the-art Unsupervised Baselines in φ_1 - MOCO trained on the target dataset

Method	STL-10 ACC	STL-10 ARI	CIFAR-10 ACC	CIFAR-10 ARI	CIFAR-100-20 ACC	CIFAR-100-20 ARI
SCAN	77.8	61.3	83.3	70.5	45.4	29.7
SPICE	86.2	73.2	84.5	70.9	46.8	32.1
HUME	**90.8**	**81.2**	**88.4**	**77.6**	**55.5**	**37.7**
	+5%	+11%	+5%	+10%	+19%	+18%

HUME:

φ_1 - MOCO trained on the target dataset
φ_2 - DINO large foundation model

HUME outperforms existing unsupervised baselines by a large margin!

HUME Scales to Large Fine-grained Datasets

State-of-the-art Unsupervised Baselines in φ_1 - MOCO trained on the ImageNet-1000

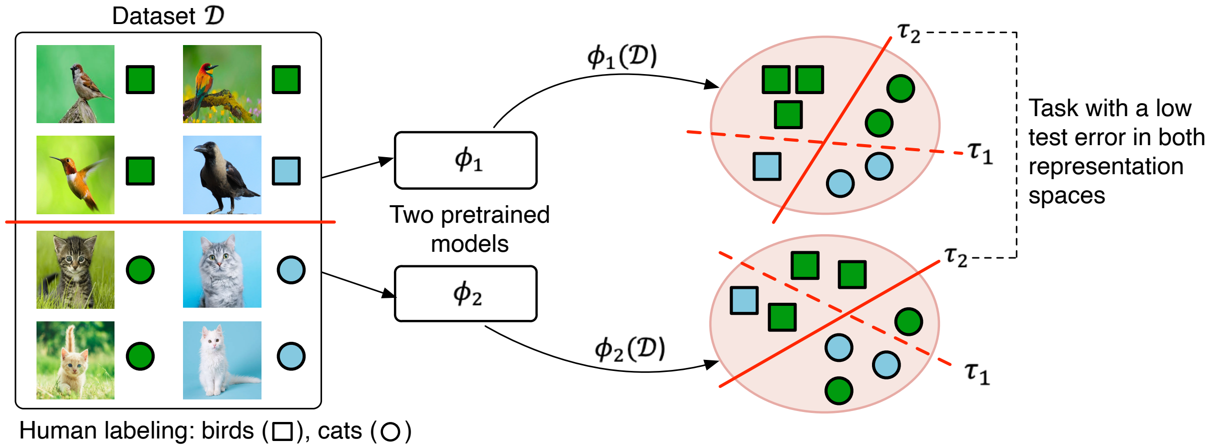
Method	ACC	ARI
SCAN	39.7	27.9
TWIST	40.6	30.0
Self-classifier	41.1	29.5
HUME	**51.1**	**38.1**
	+24%	+27%

ImageNet-1000:
- 1000 classes
- 1,281,167 training samples

HUME:
φ_1 - MOCO trained on the ImageNet-1000
φ_2 - DINO large foundation models

HUME achieves remarkable improvement on large-scale ImageNet-1k!

HUME Framework

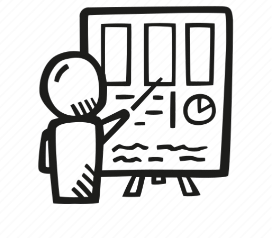


Check our paper and code for more details!

HUME:

- Provides a new view to tackle unsupervised learning
- Matches performance of supervised linear probe on the STL-10 and CIFAR-10 datasets
- Achieves state-of-the-art unsupervised performance and more…

Come to our poster to chat about HUME!



Tue 12 Dec 3:15 p.m. PST — 5:15 p.m. PST
Great Hall & Hall B1+B2 #1012