Lifelong Learning for Visual Representations

Diane Larlus

February 2025 – VISIGRAPP 2025
20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications

Motivation for lifelong learning

The brain learns **incrementally**

.. and **retains** acquired skills

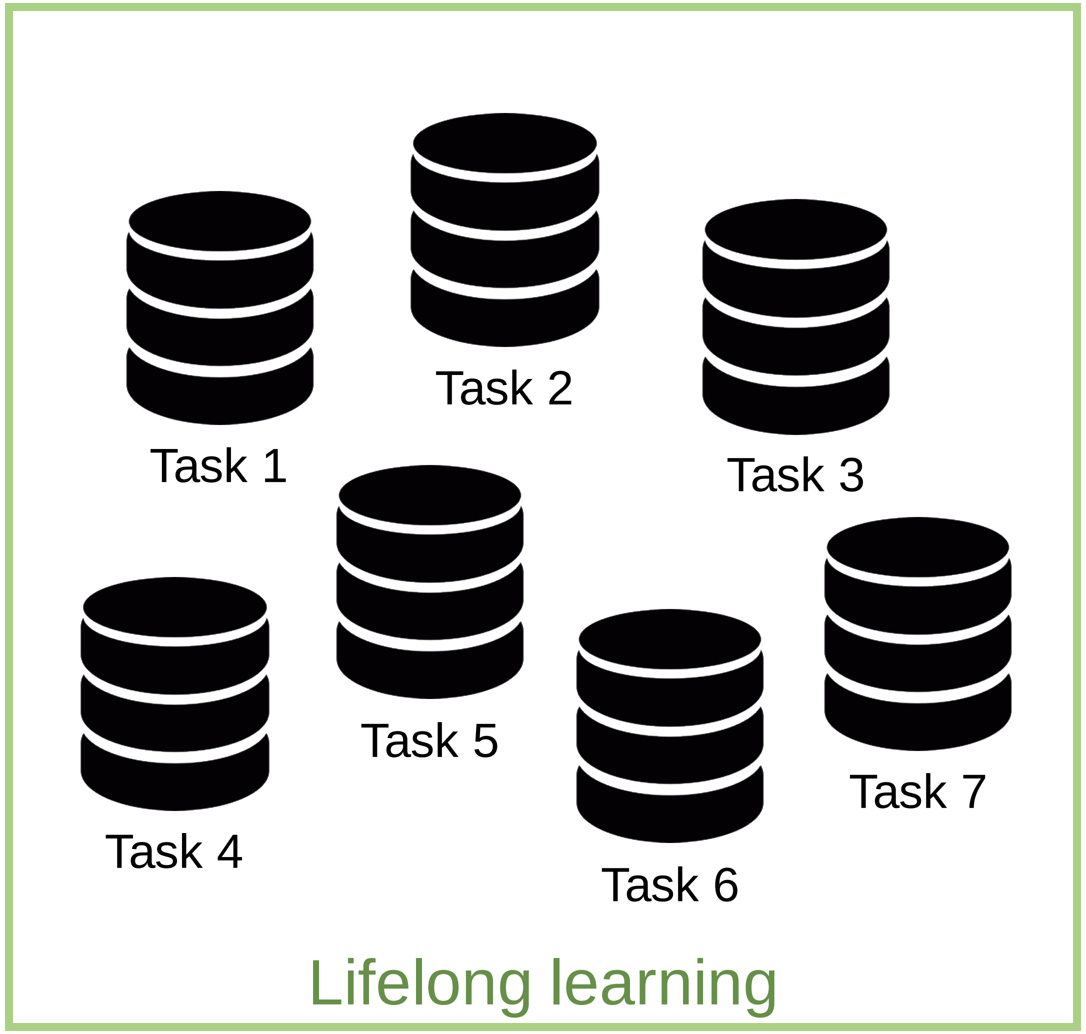
We can **limit compute** ..

.. by **reusing**, **adapting**, **transferring**


NAVER
NAVER

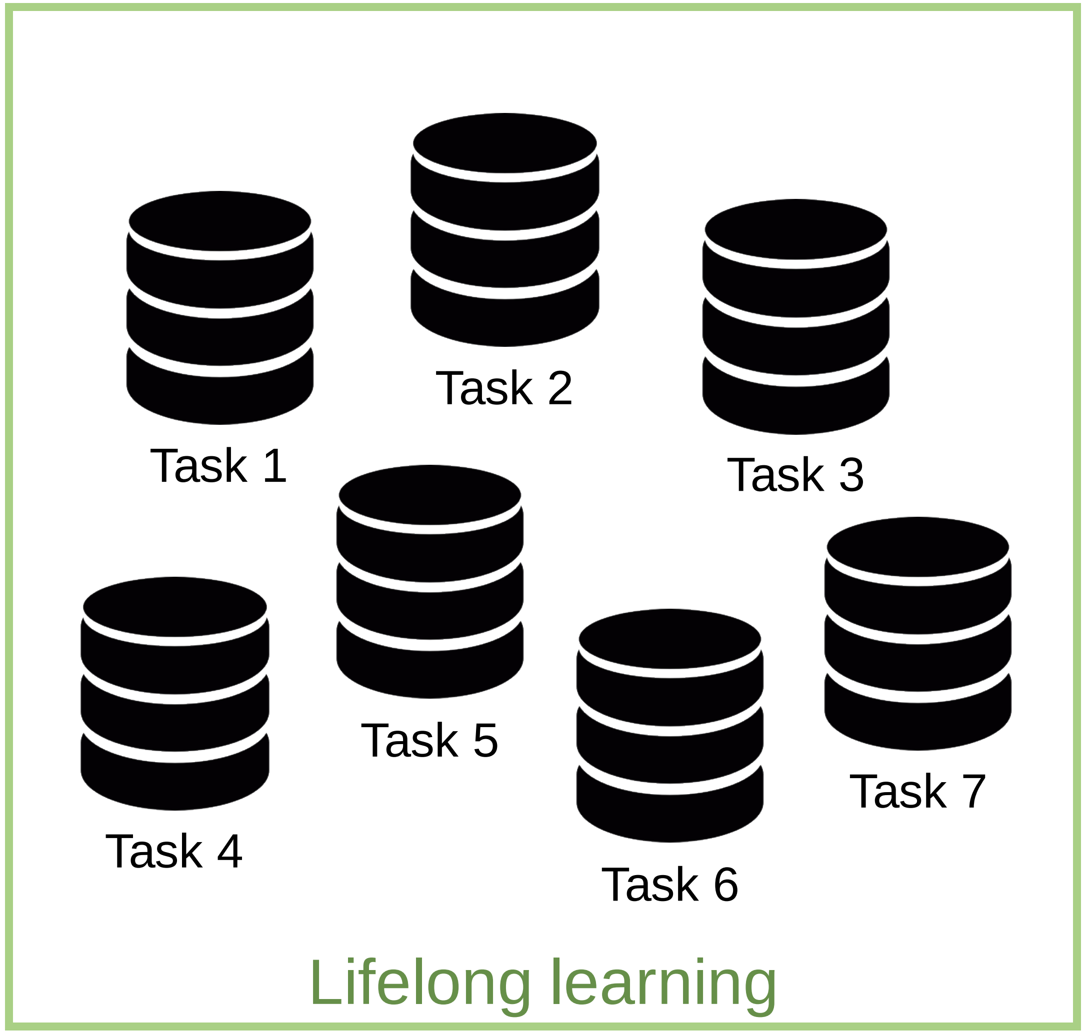
Unknown environments require **generalization** and **adaptation**

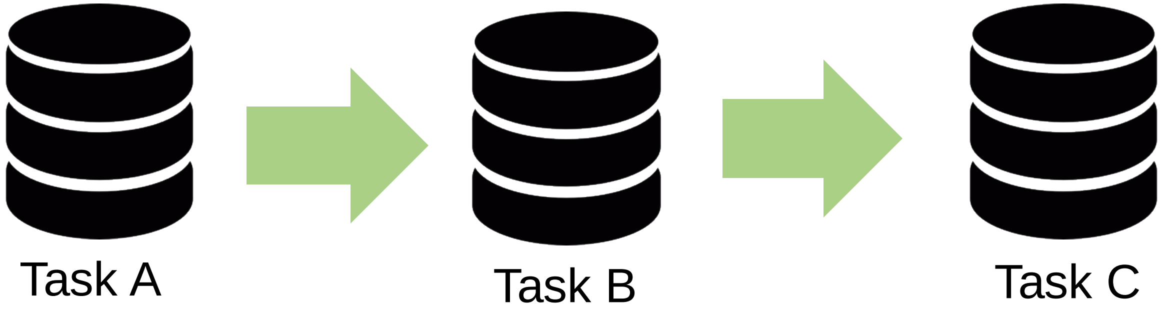
A broad definition



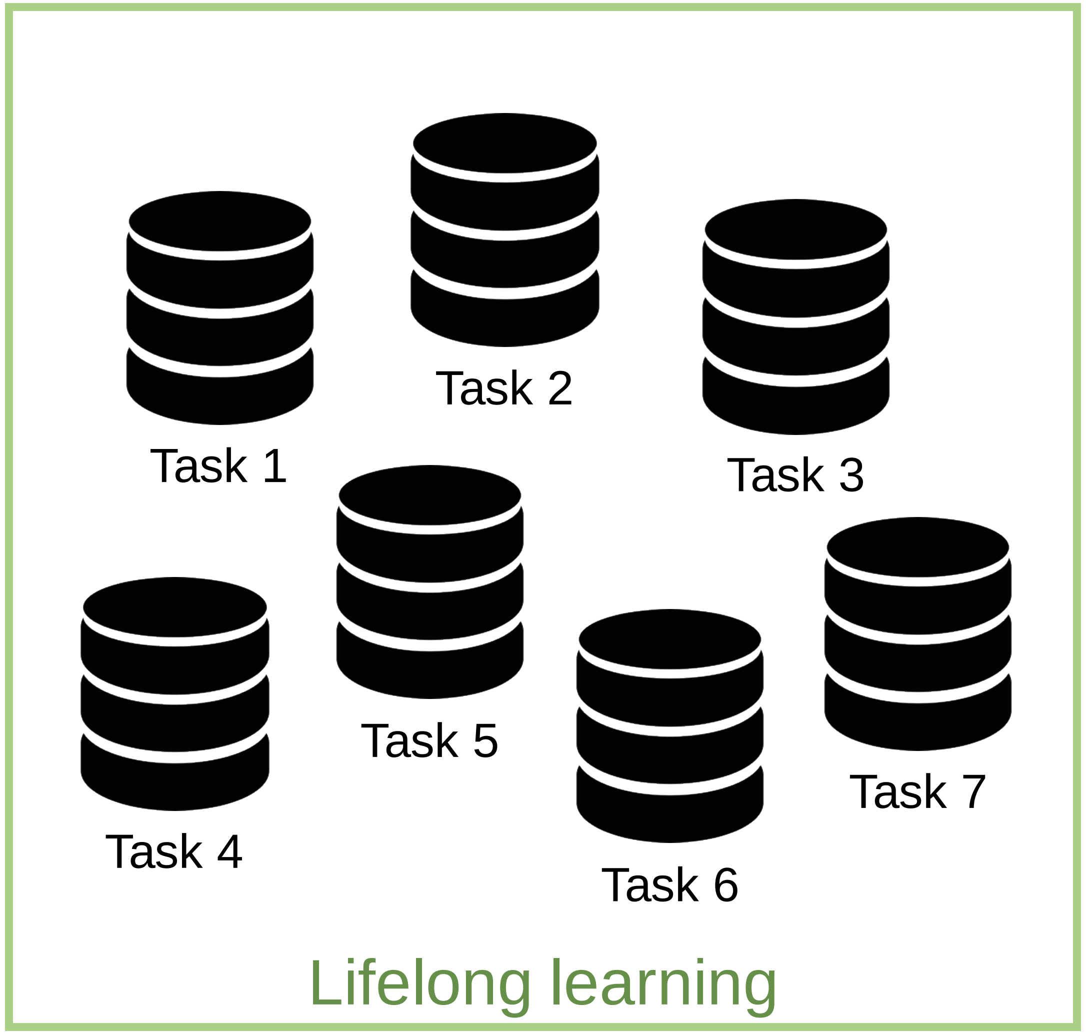
- Learn incrementally
- Retain acquired skills
- Reuse / recycle / extend skills

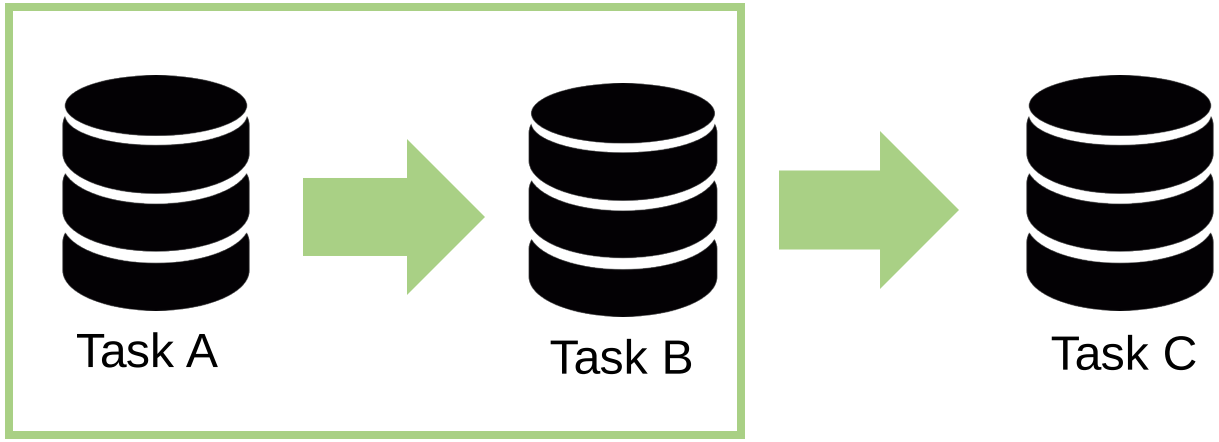
Lifelong learning / Continual learning





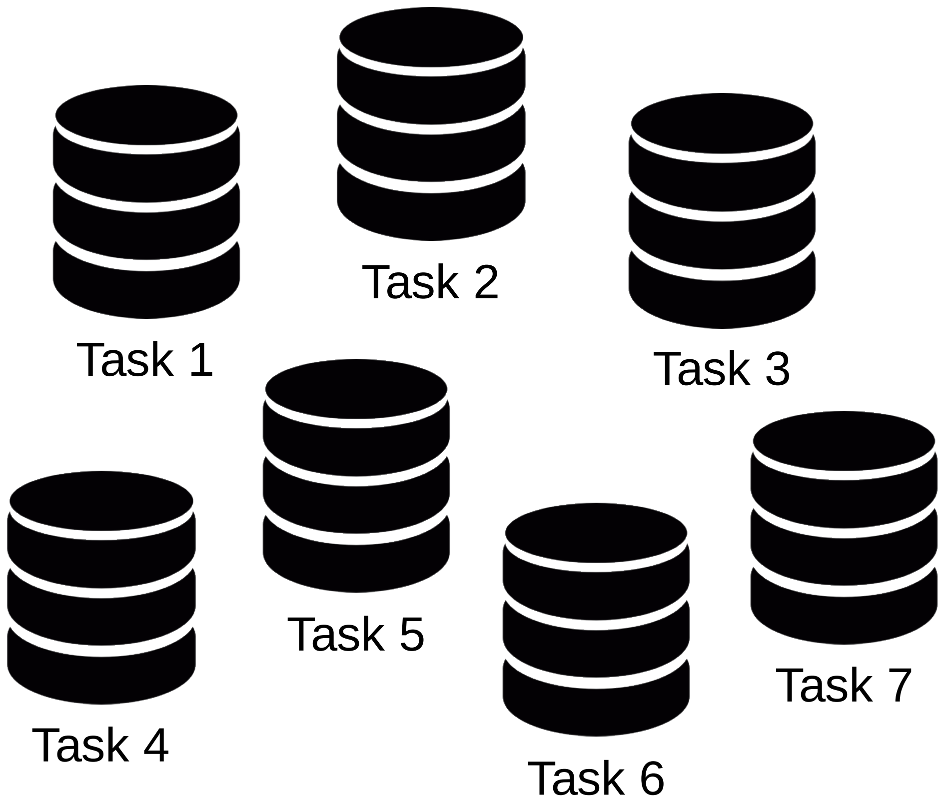
Lifelong learning / Continual learning



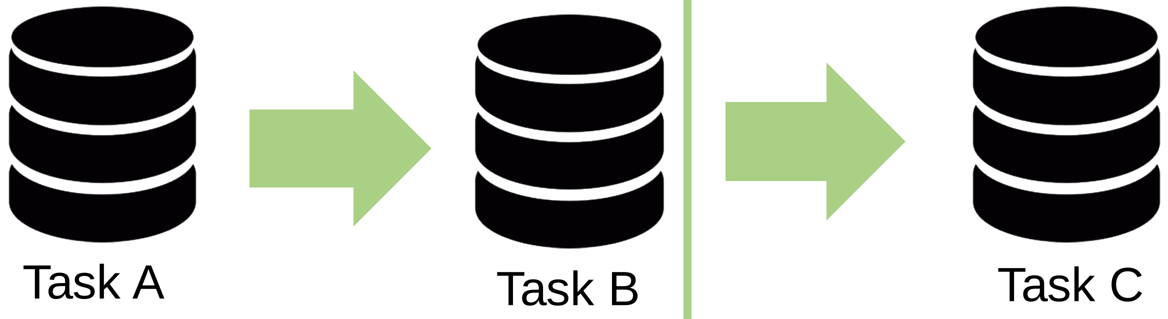


Option 1:
Task A is **only** a **starting point** for Task B

Lifelong learning / Continual learning



Lifelong learning

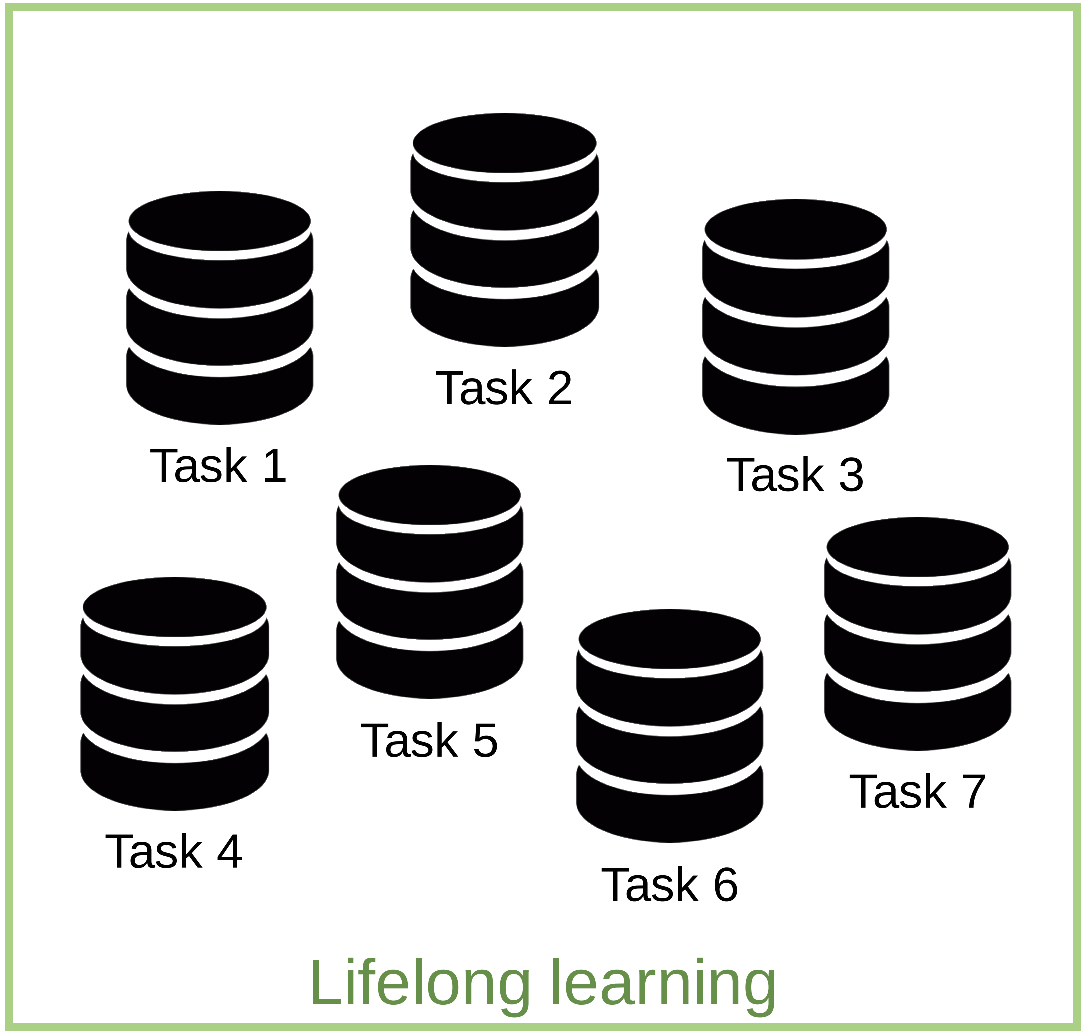


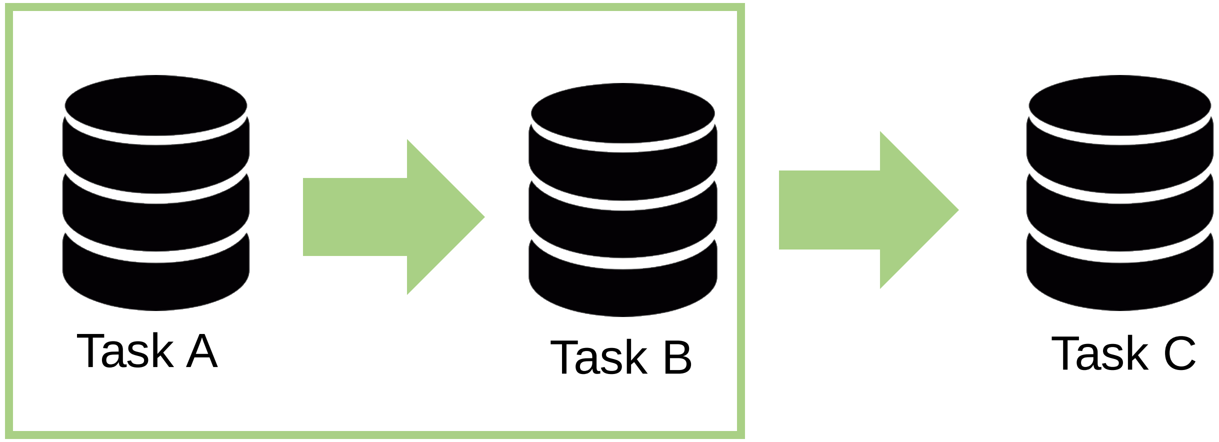
Option 2:
Task A is **still relevant**

Main challenge
Catastrophic forgetting

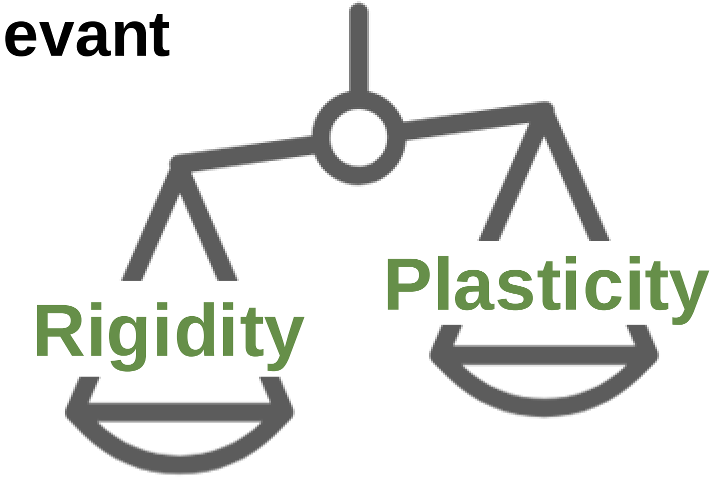
[French@CognitiveScience99]

Lifelong learning / Continual learning

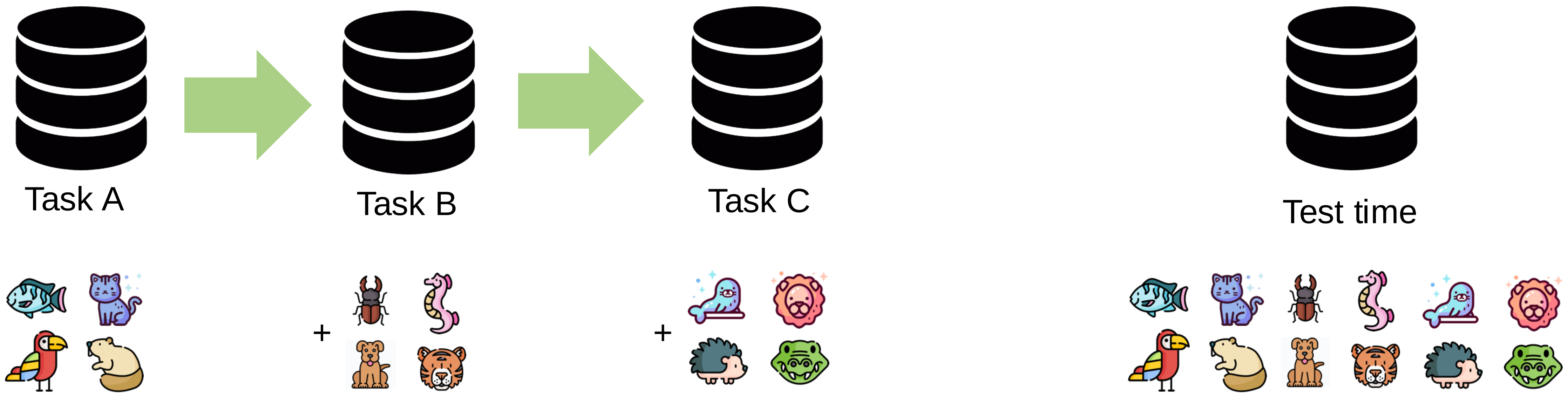




Option 2:
Task A is **still relevant**

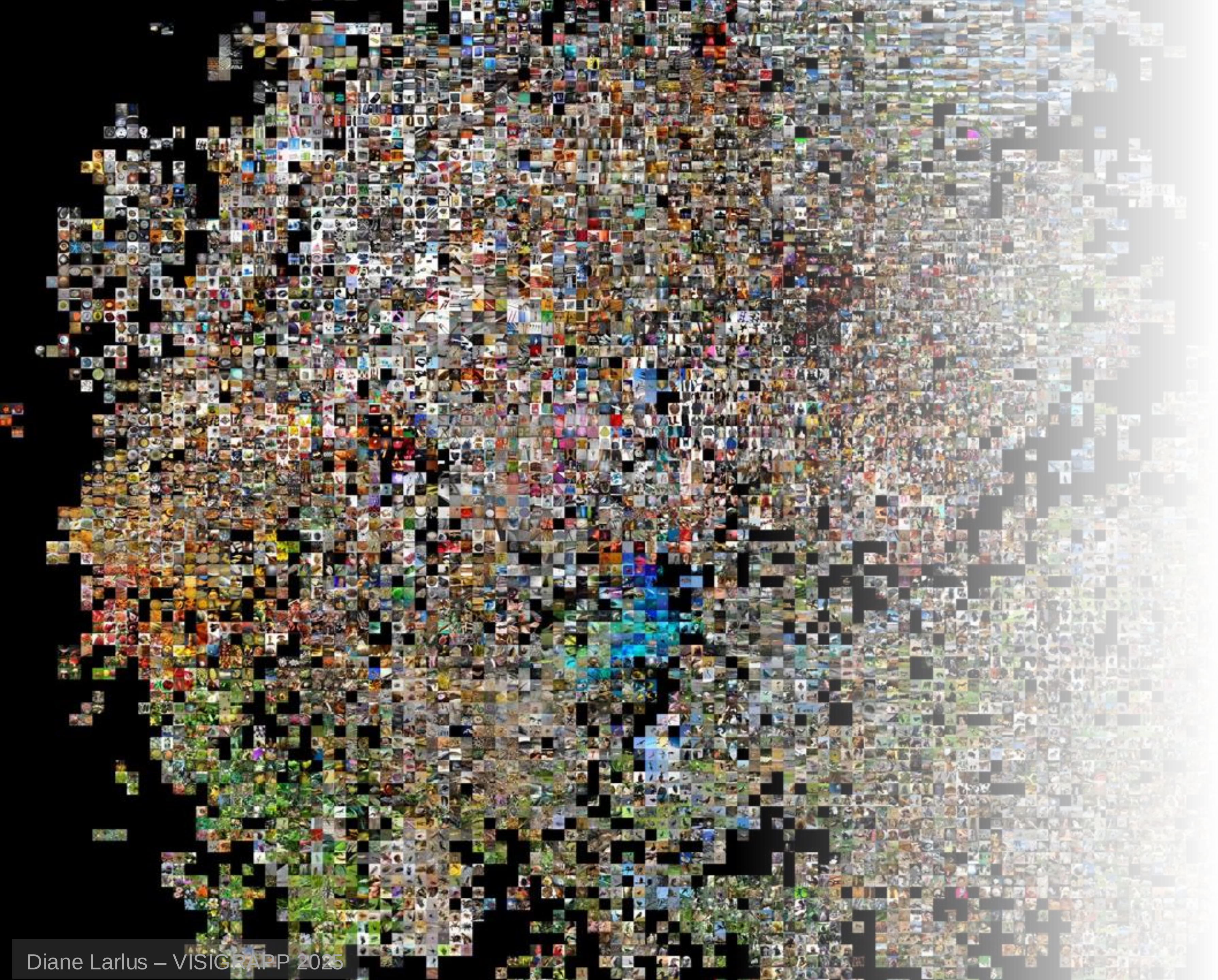


Lifelong learning / Continual learning



Continual learning

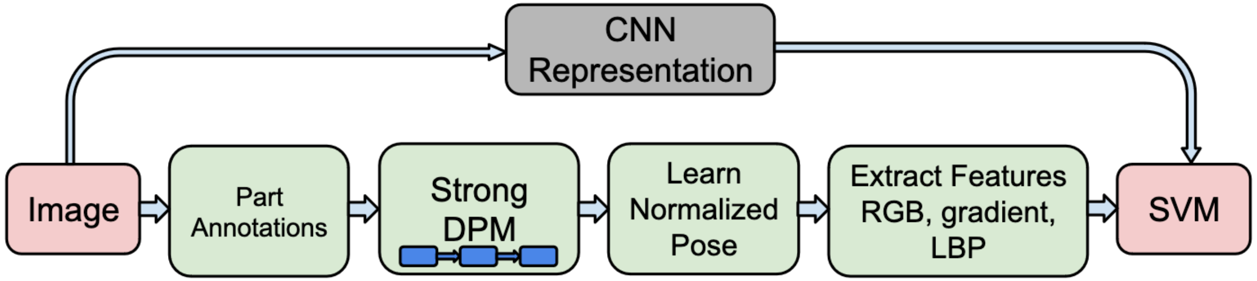
in the age of large pretrained models

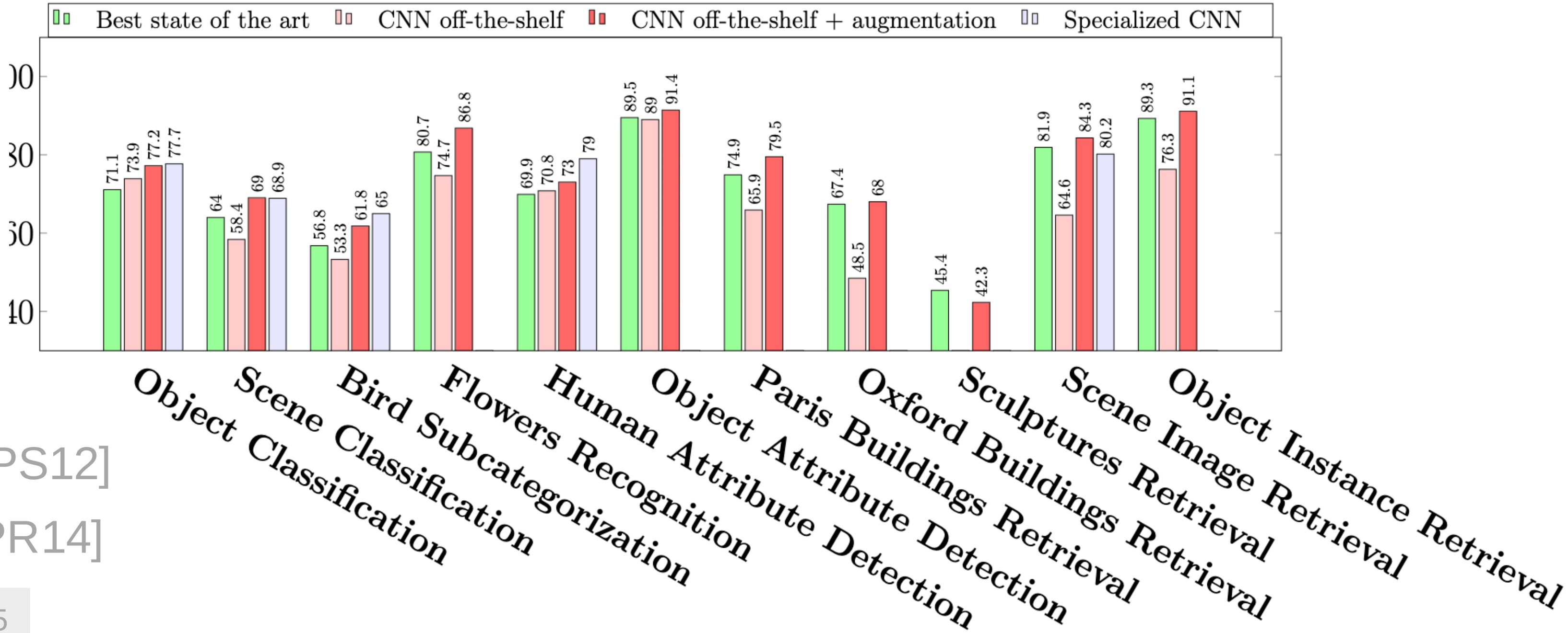


IMAGENET

[Deng@CVPR09]

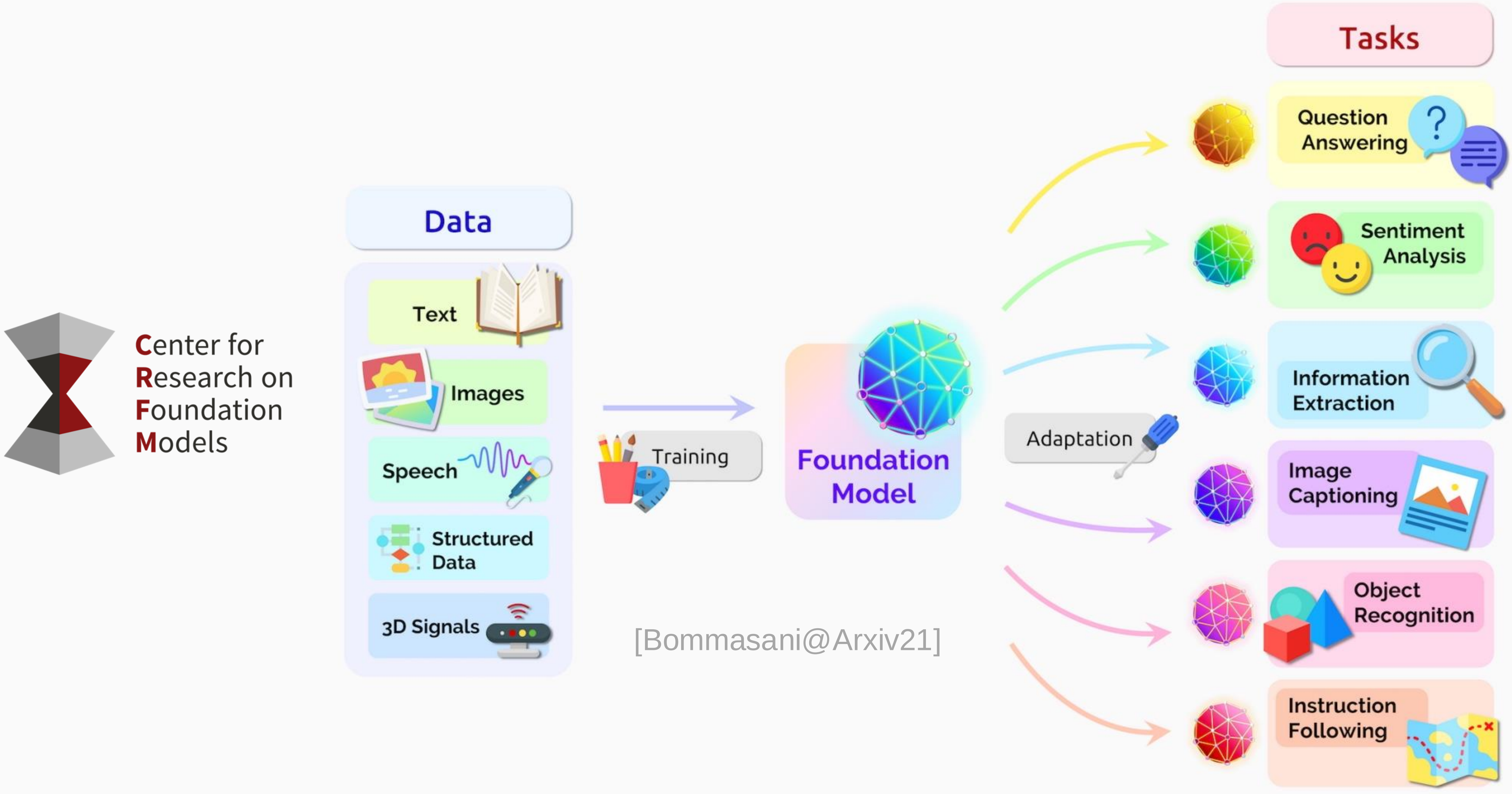
CNN Features off-the-shelf: an Astounding Baseline for Recognition



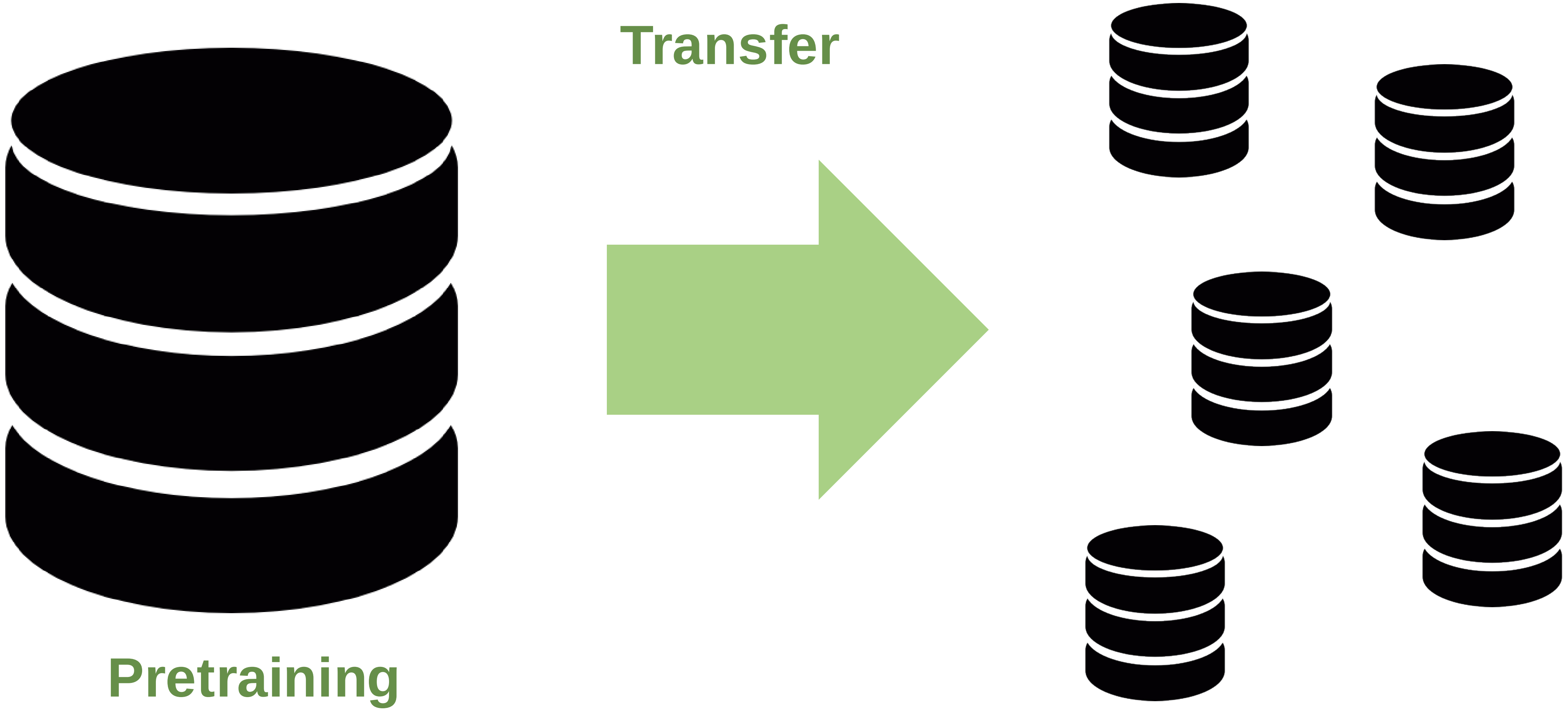


[Krizhevsky@NeurIPS12]

[Razavian@W_CVPR14]



[Bommasani@Arxiv21]


Transfer
Pretraining

Let’s assume two phases

- **Pretraining** & **Transfer**

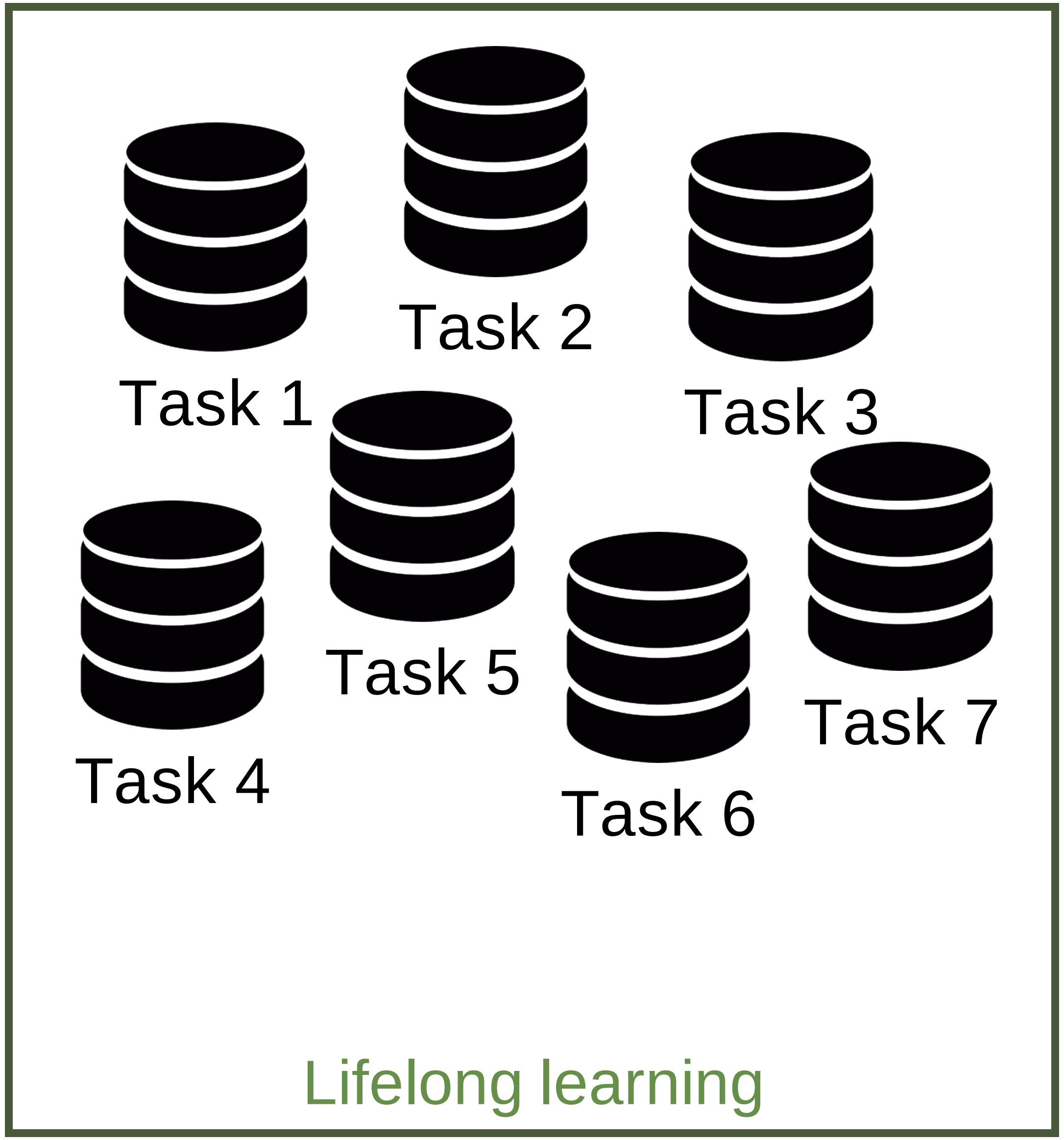
What comes next

- *How can we **pretrain** a good model?*
  - Evaluating generalization capabilities
  - Training visual features using data beyond images
- *How should we **transfer***?
  - From one pretrained model to a specific task
  - From multiple pretrained models

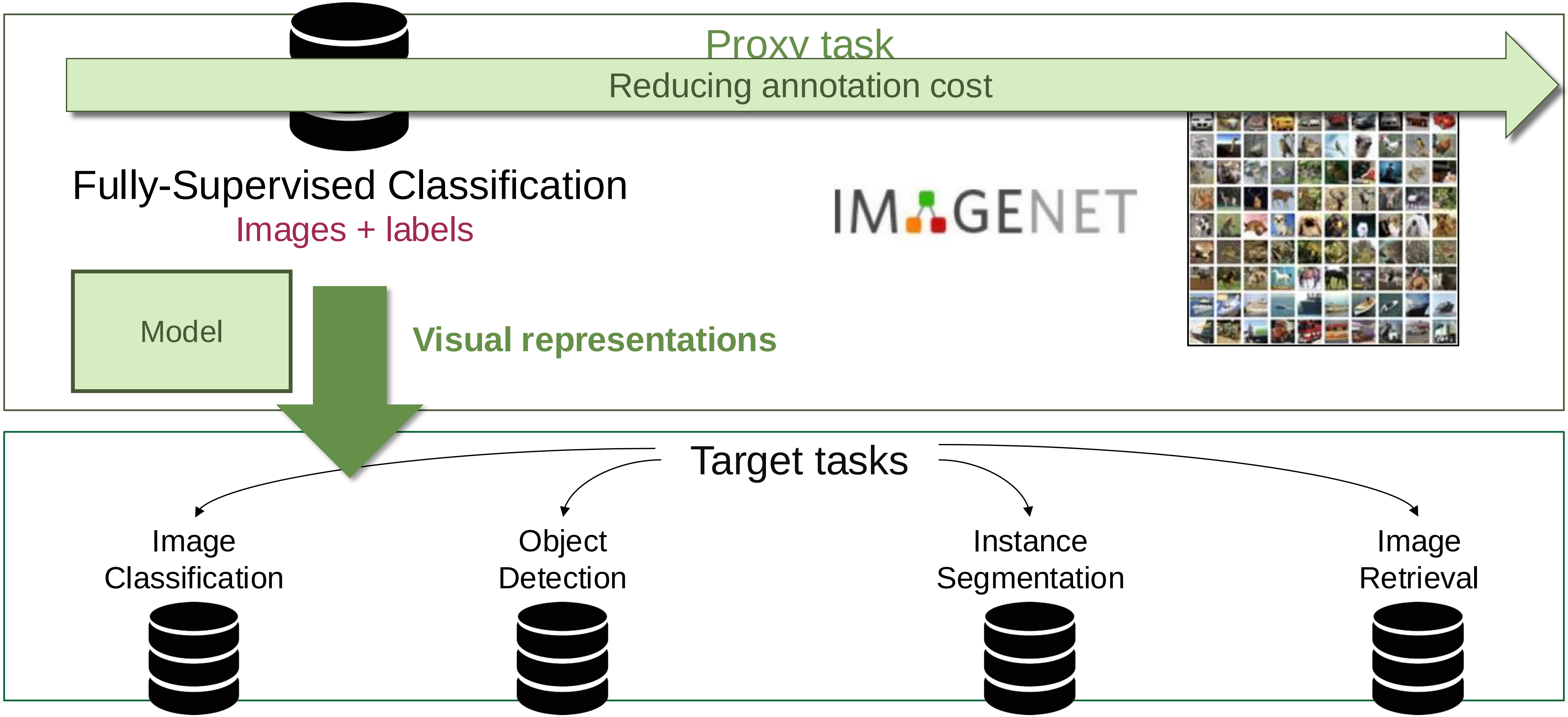
Pretraining strong models

A “good” pretrained model

- **Broad** knowledge
- Robust to **concept** shifts
- Easily **adapts** to new tasks



We don’t know the target task


Proxy task
Reducing annotation cost
Fully-Supervised Classification
Images + labels
IMAGENET
Model
Visual representations
Target tasks
Image Classification
Object Detection
Instance Segmentation
Image Retrieval

Reducing annotation cost

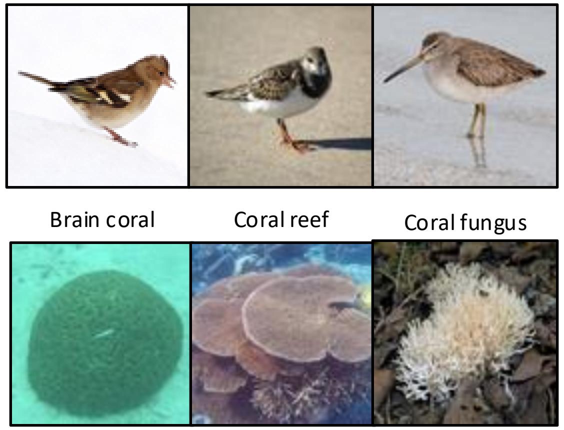
Fully-Supervised

fine-grained annotations

expert knowledge

Eurasian finch | Ruddy turnstone | Dowitcher

Brain coral | Coral reef | Coral fungus



Self-supervised

annotation-free images

No supervision

Reducing annotation cost

Fully-Supervised

fine-grained annotations

expert knowledge

Self-supervised

annotation-free images

~~labels~~

No supervision

Reducing annotation cost

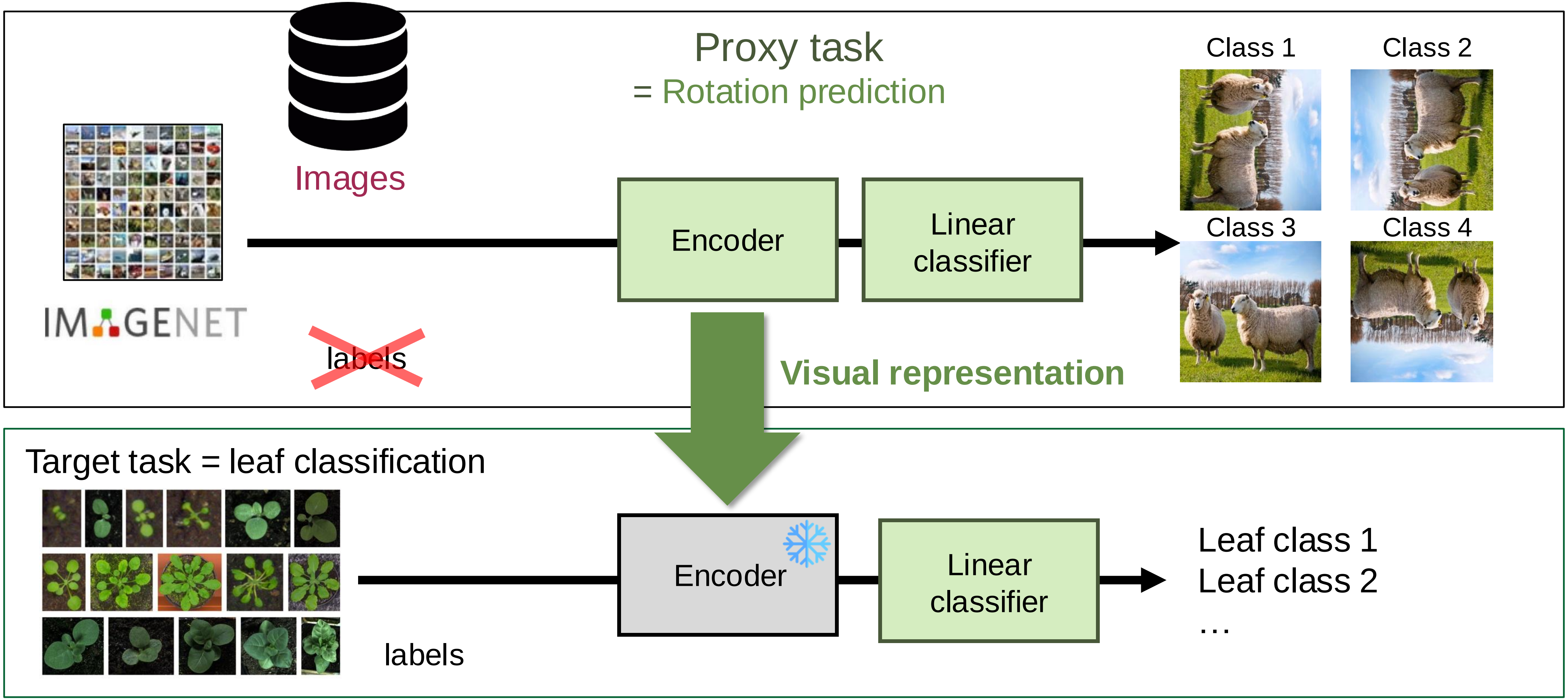
Fully-Supervised
fine-grained annotations
expert knowledge

Self-supervised
annotation-free images

DINOv2

[Oquab@TMLR24]

~~labels~~



RotNet [Gidaris@ICRL18]

How well do those pretrained models generalize?

Proxy task

Fully-supervised classification or
Self-supervised approaches, etc.

IMAGENET

Model

Visual representations

Target tasks

Image Classification

Object Detection

Instance Segmentation

Image Retrieval

Proxy task

Fully-supervised classification or
Self-supervised approaches, etc.

IMAGENET

Model

Visual representations

Target tasks

??

How well does the produced visual representation transfer?

Proxy task

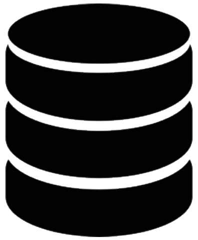
Fully-supervised classification or
Self-supervised approaches, etc.

IMAGENET

Model

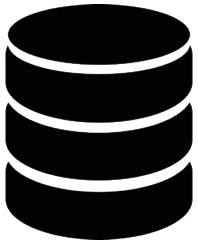
Visual representations

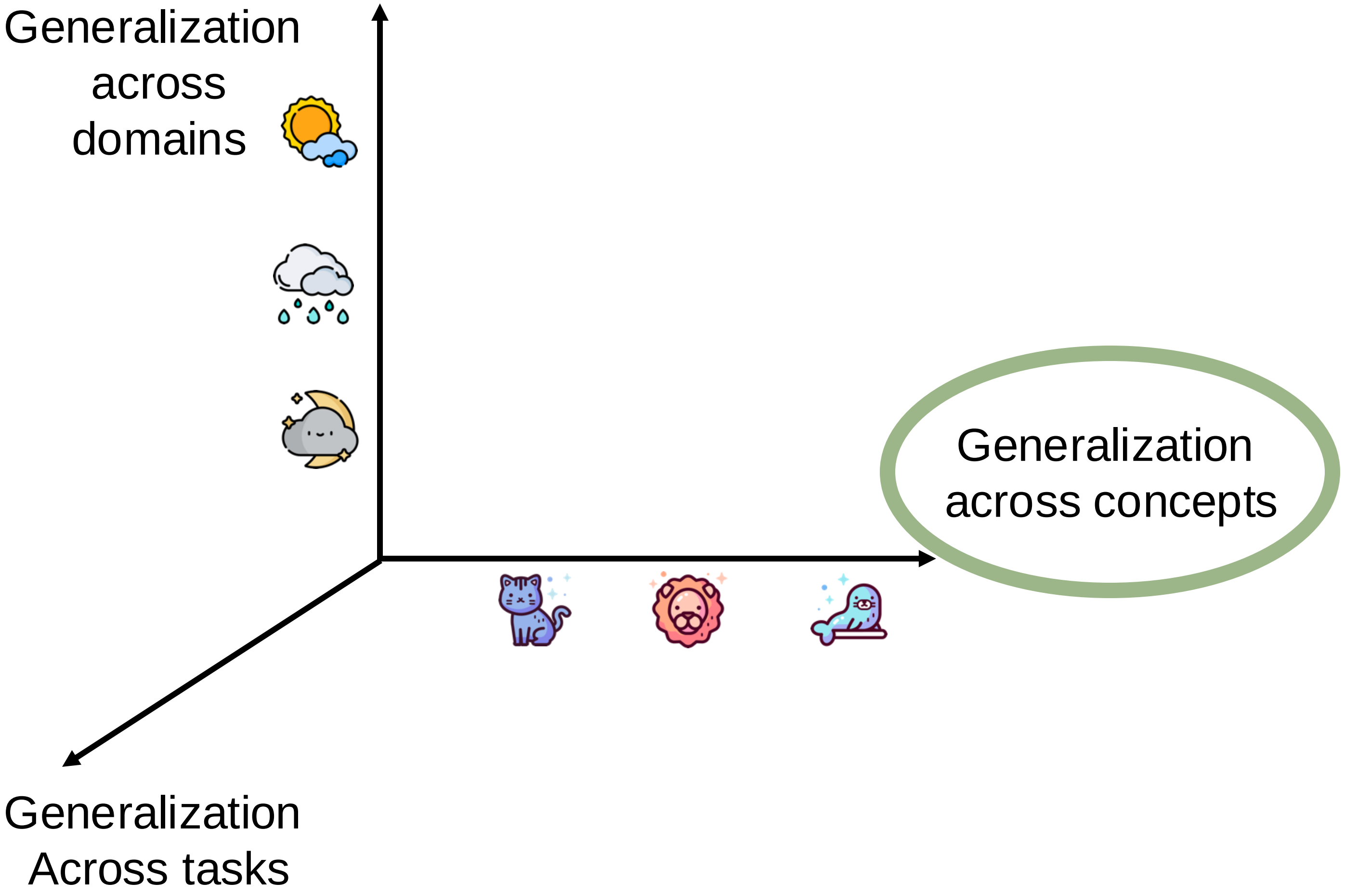
Measure performance on (many) other datasets



Proxy task

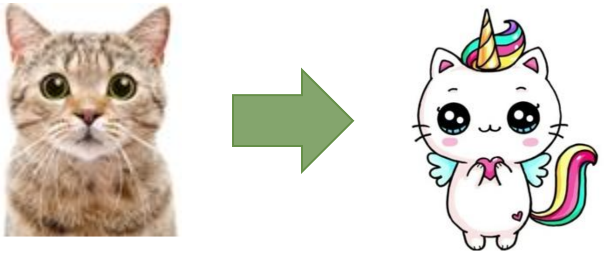
Target task





Evaluation of visual representations – concept generalization

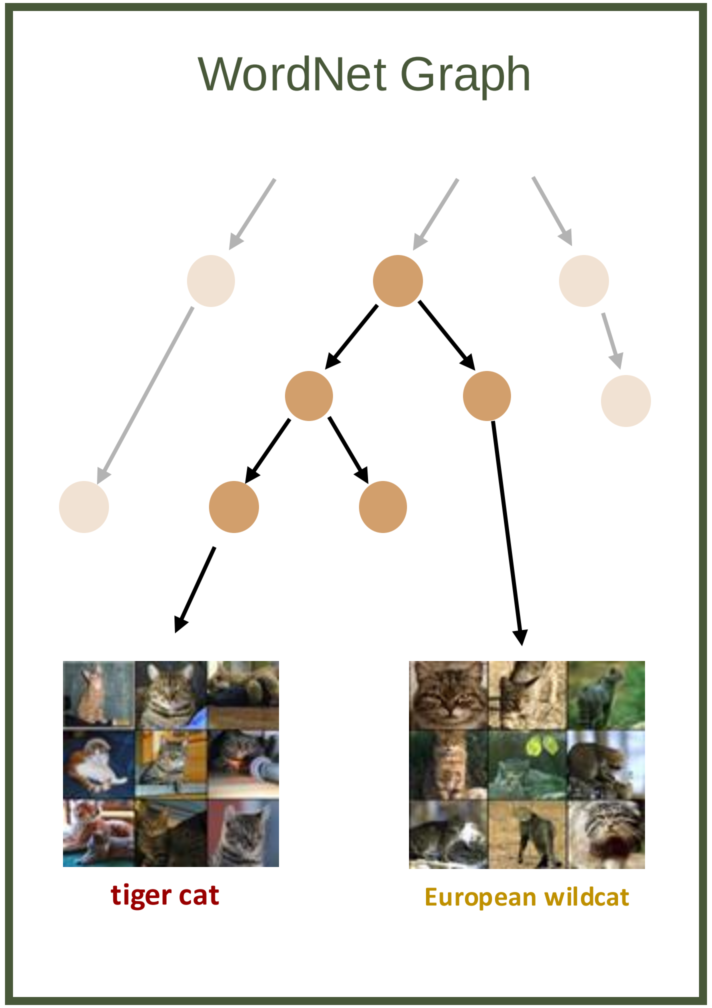
When training a model on a set of ***seen*** *concepts,*
how well does it ***generalize*** *to* ***new, unseen*** *concepts ?*



Seen: Cats
Unseen: Caticorns

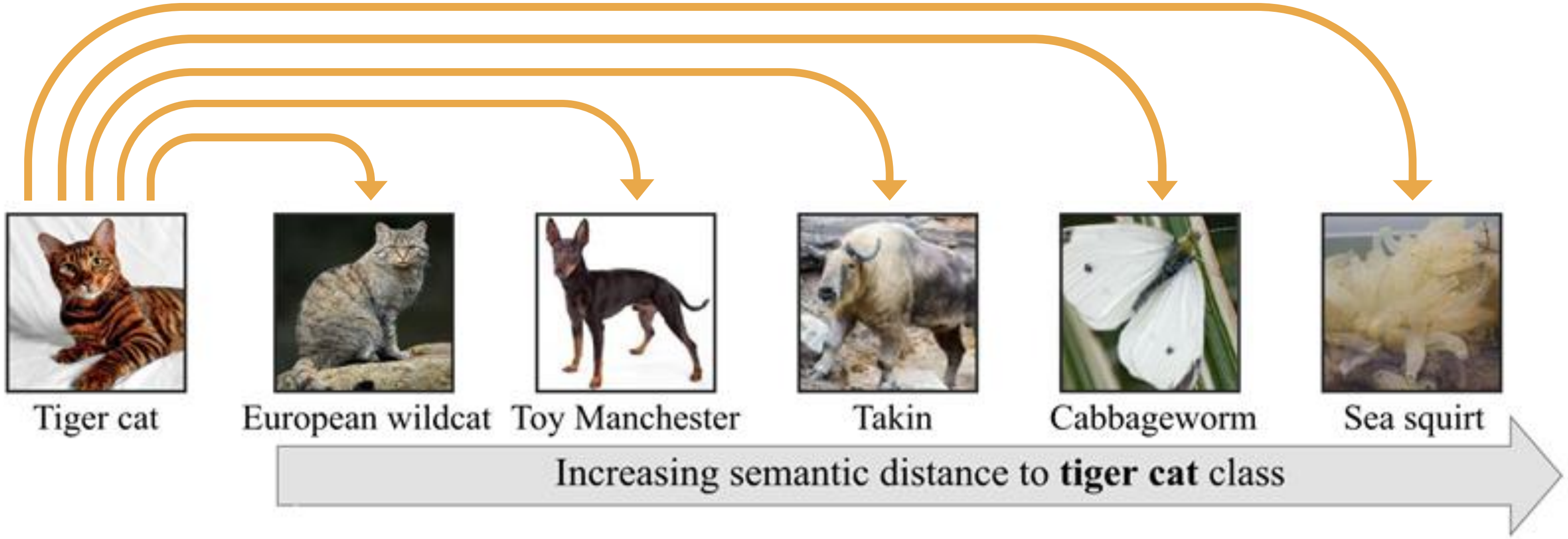
Hypothesis: **Semantic similarity** between **seen** and **unseen** concepts matters for generalization

Semantic distance between concepts



[**Lin**: Lin@ICML1998]

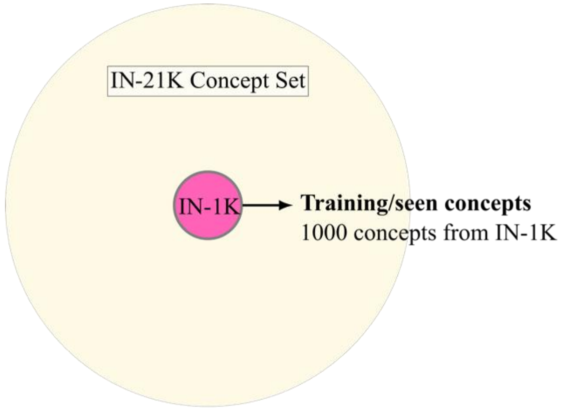
Semantic distance between concepts

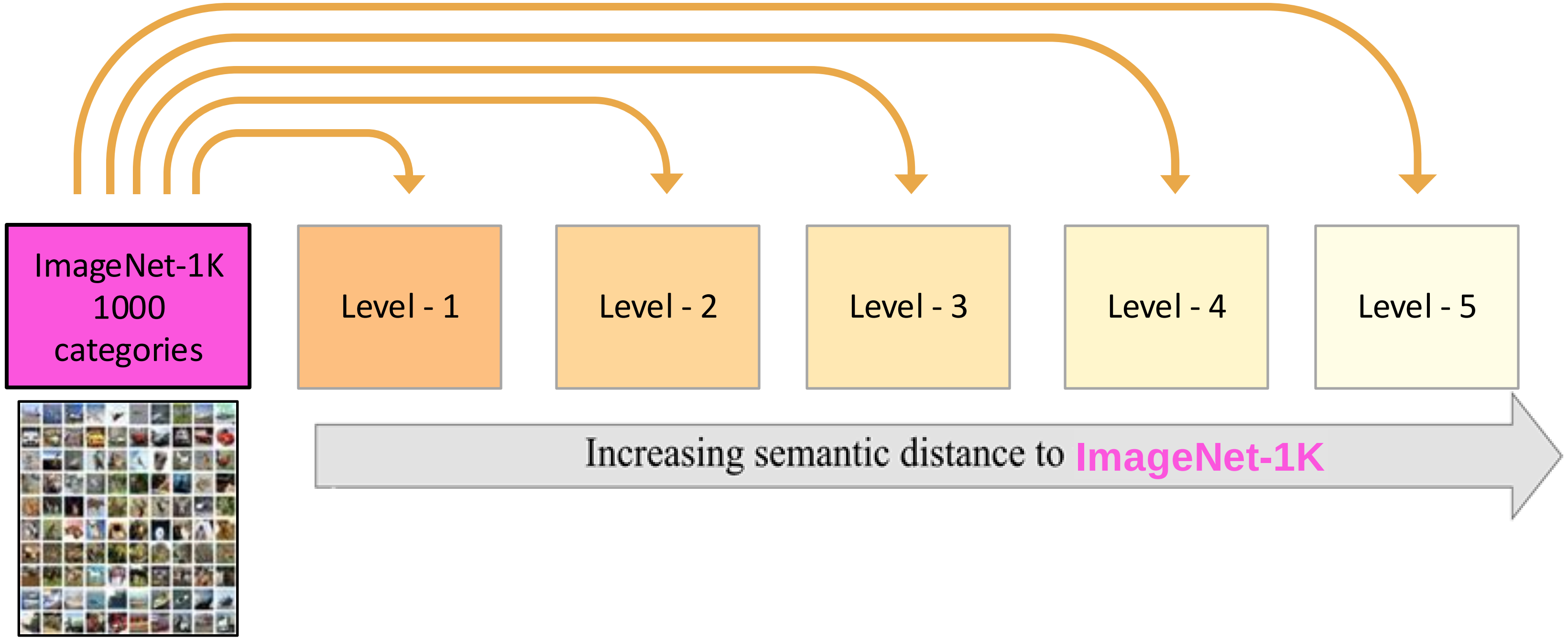


[**Lin**: Lin@ICML1998]

Semantic distance between sets of concepts

[**ImageNet**: Deng@CVPR2009]

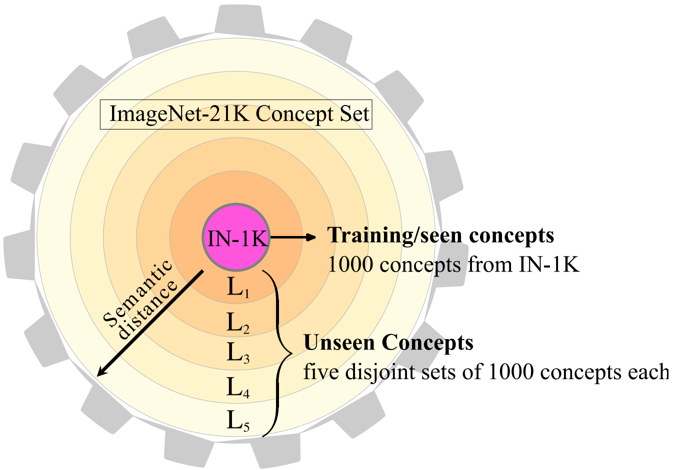




[**CoG**: Sariyildiz@ICCV21]

The Concept Generalization (CoG) benchmark

[**ImageNet**: Deng@CVPR2009]



The CoG benchmark

ImageNet-1K
1000
categories

Training set

Level - 1

Level - 2

Level - 3

Level - 4

Level - 5

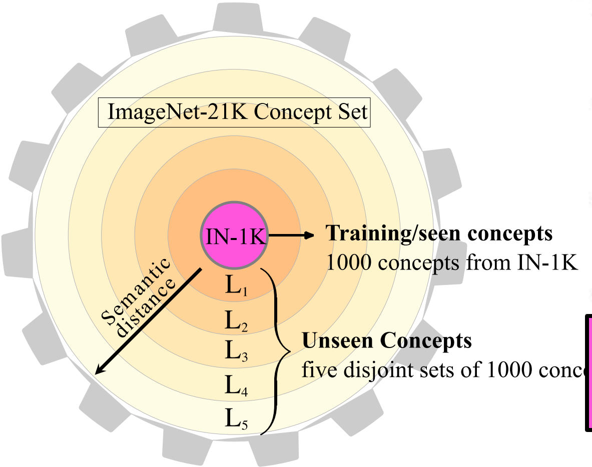
Increasing semantic distance to ImageNet-1K

Increasingly difficult **test sets** (5 concept generalization **levels**)

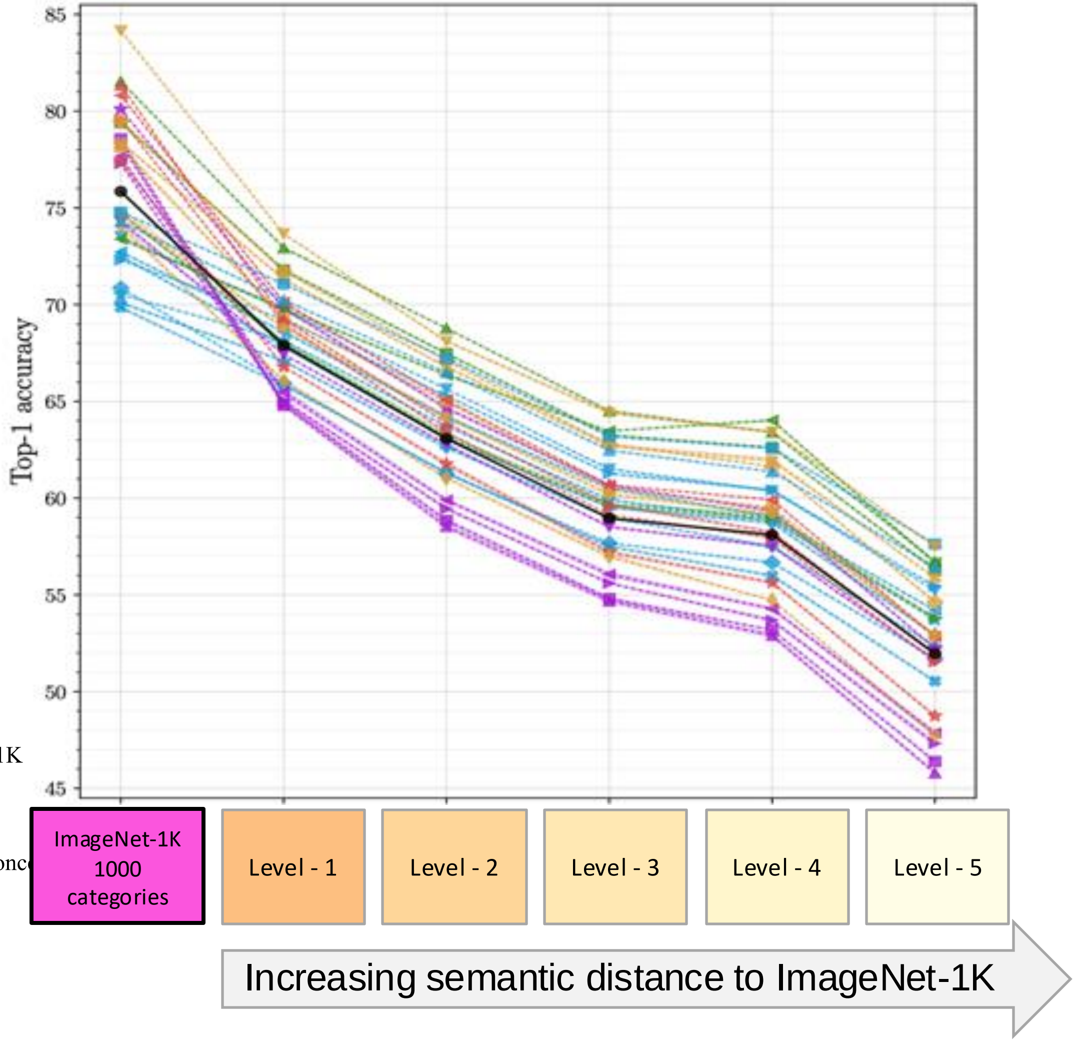
[**CoG**: Sariyildiz@ICCV21]

Observations

- It is harder to generalize to semantically distant concepts



The **CoG** benchmark



Architecture: *Models with different backbone*	
a-T2T-ViT-t-14	Visual transformer (21.5M)
a-DeiT-S	Visual transformer (22M)
a-DeiT-S-distilled	Distilled *a*-DeiT-S (22M)
a-Inception-v3	CNN with inception modules (27.2M)
a-NAT-M4	Neural architecture search model (7.6M)
a-EfficientNet-B1	Neural architecture search model (7.8M)
a-DeiT-B-distilled	Bigger version of *a*-DeiT-S-distilled (87.6M)
a-ResNet152	Bigger version of ResNet50 (60.2M)
a-VGG19	Simple CNN architecture (143.5M)

Self-supervision: *ResNet50 models trained in this framework*	
s-SimCLR-v2	Online instance discrimination (ID)
s-MoCo-v2	ID with momentum encoder and memory bank
s-SwAV	Online clustering
s-BYOL	Negative-free ID with momentum encoder
s-MoCHi	ID with negative pair mining
s-InfoMin	ID with careful positive pair selection
s-OBoW	Online bag-of-visual-words prediction
s-CompReSS	Distilled from SimCLR-v1 (with ResNet50x4)

Regularization: *ResNet50 models with additional regularization*	
r-MixUp	Label-associated data augmentation
r-Manifold-MixUp	Label-associated data augmentation
r-CutMix	Label-associated data augmentation
r-ReLabel	Trained on a "multi-label" version of IN-1K
r-Adv-Robust	Adversarially robust model
r-MEAL-v2	Distilled ResNet50

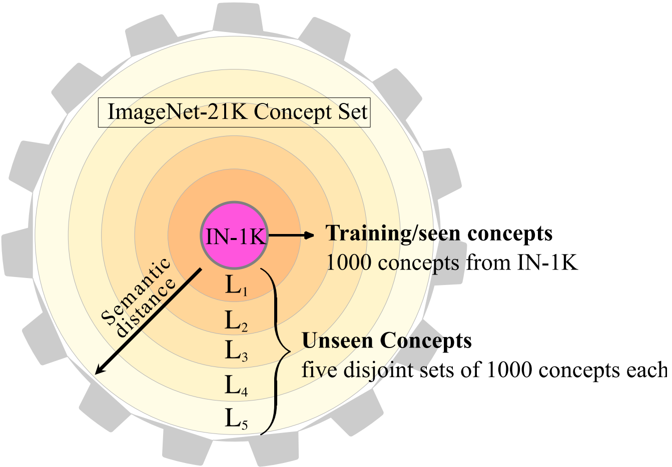
Use of web data: *ResNet50 models using additional data*	
d-MoPro	Trained on WebVision-V1 ($\sim 2\times$)
d-Semi-Sup	Pretrained on YFCC-100M ($\sim 100\times$), then fine-tuned on IN-1K
d-Semi-Weakly-Sup	Pretrained on IG-1B ($\sim 1000\times$), then fine-tuned on IN-1K
d-CLIP	Trained on WebImageText ($\sim 400\times$)

[**CoG**: Sariyildiz@ICCV21]

Observations

- It is harder to generalize to semantically distant concepts
- Recent **self-supervised** approaches generalize better
- Label-based augmentations hurt concept generalization

Reference
Concept generalization in visual representation learning
Mert Bulent Sariyildiz, Yannis Kalantidis, Diane Larlus, Karteek Alahari
ICCV 2021



The **CoG** benchmark

Architecture: *Models with different backbone*	
a-T2T-ViT-t-14	Visual transformer (21.5M)
a-DeiT-S	Visual transformer (22M)
a-DeiT-S-distilled	Distilled *a*-DeiT-S (22M)
a-Inception-v3	CNN with inception modules (27.2M)
a-NAT-M4	Neural architecture search model (7.6M)
a-EfficientNet-B1	Neural architecture search model (7.8M)
a-DeiT-B-distilled	Bigger version of *a*-DeiT-S-distilled (87.6M)
a-ResNet152	Bigger version of ResNet50 (60.2M)
a-VGG19	Simple CNN architecture (143.5M)

Self-supervision: *ResNet50 models trained in this framework*	
s-SimCLR-v2	Online instance discrimination (ID)
s-MoCo-v2	ID with momentum encoder and memory bank
s-SwAV	Online clustering
s-BYOL	Negative-free ID with momentum encoder
s-MoCHi	ID with negative pair mining
s-InfoMin	ID with careful positive pair selection
s-OBoW	Online bag-of-visual-words prediction
s-CompReSS	Distilled from SimCLR-v1 (with ResNet50x4)

Regularization: *ResNet50 models with additional regularization*	
r-MixUp	Label-associated data augmentation
r-Manifold-MixUp	Label-associated data augmentation
r-CutMix	Label-associated data augmentation
r-ReLabel	Trained on a "multi-label" version of IN-1K
r-Adv-Robust	Adversarially robust model
r-MEAL-v2	Distilled ResNet50

Use of web data: *ResNet50 models using additional data*	
d-MoPro	Trained on WebVision-V1 ($\sim 2\times$)
d-Semi-Sup	Pretrained on YFCC-100M ($\sim 100\times$), then fine-tuned on IN-1K
d-Semi-Weakly-Sup	Pretrained on IG-1B ($\sim 1000\times$), then fine-tuned on IN-1K
d-CLIP	Trained on WebImageText ($\sim 400\times$)

[**CoG**: Sariyildiz@ICCV21]

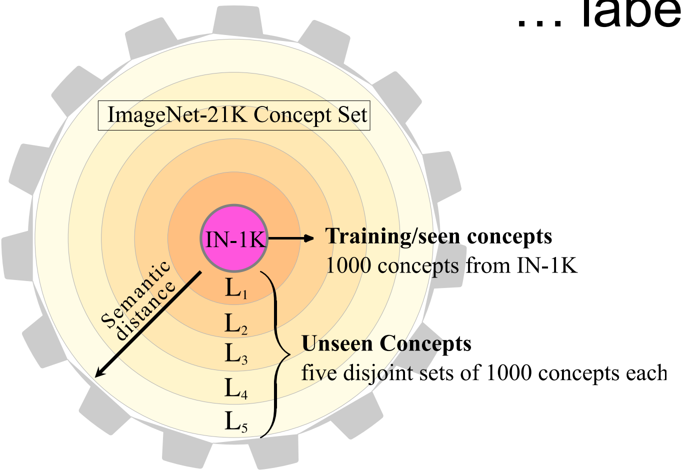
Observations

- Recent **self-supervised** approaches generalize better

Yes, but ..

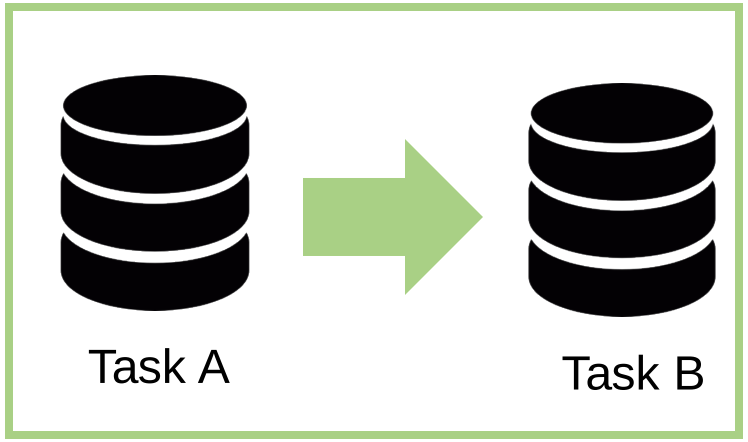
.. a good model should shine **both** on
- **Training** task
- **Transfer** tasks

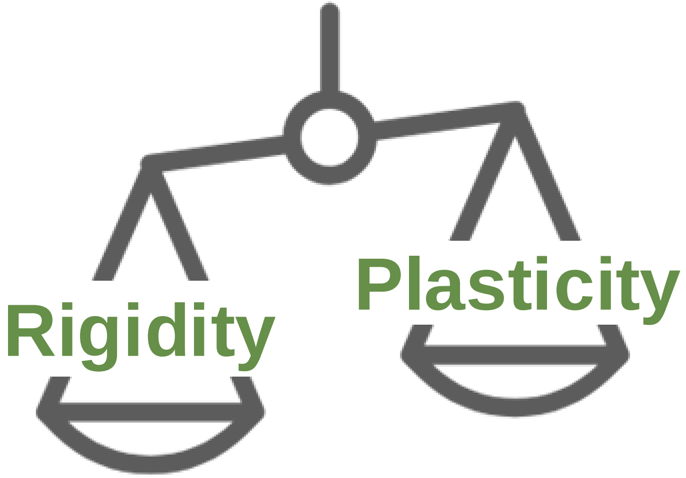
… labels shouldn't hurt

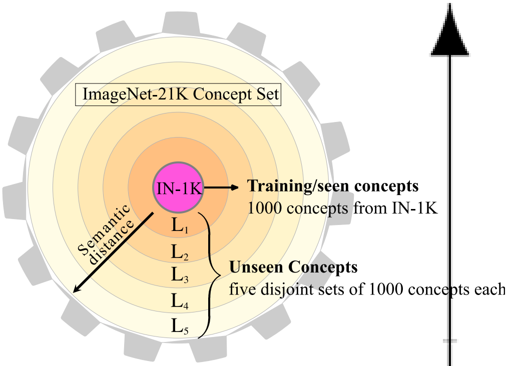


The **CoG** benchmark

Option 2:
Task A is **still relevant**







CoG
+
8 datasets

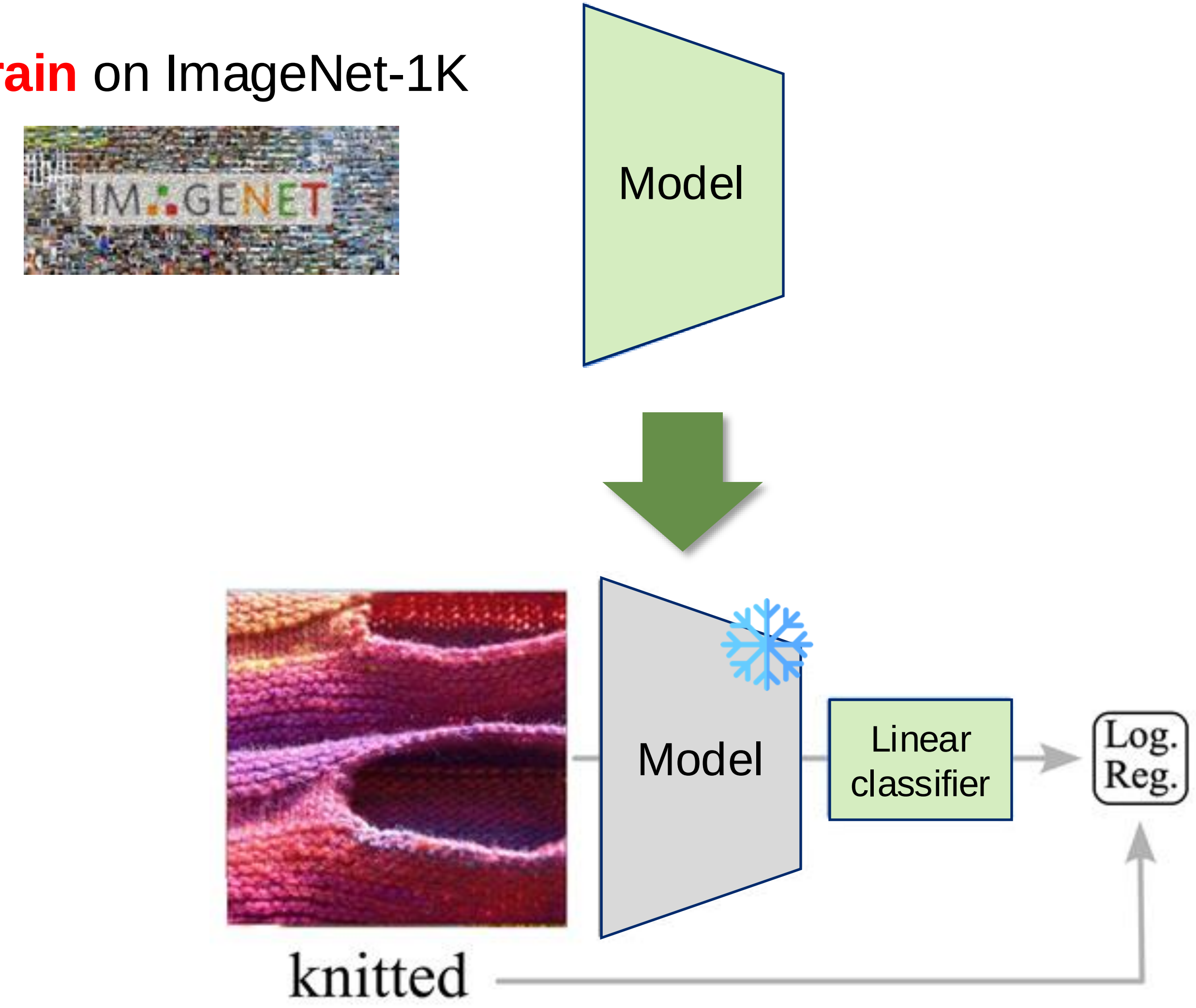
Transfer

.. a good model should shine **both** on
- **Training** task
- **Transfer** tasks

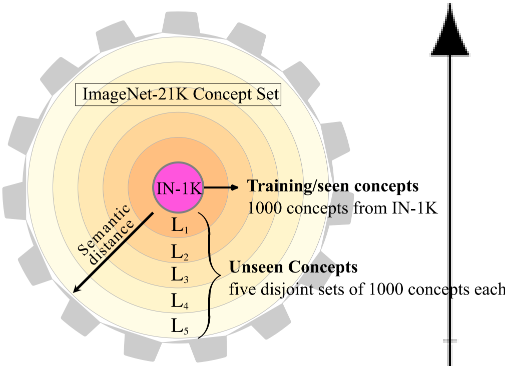
… labels shouldn't hurt

Training

Train on ImageNet-1K



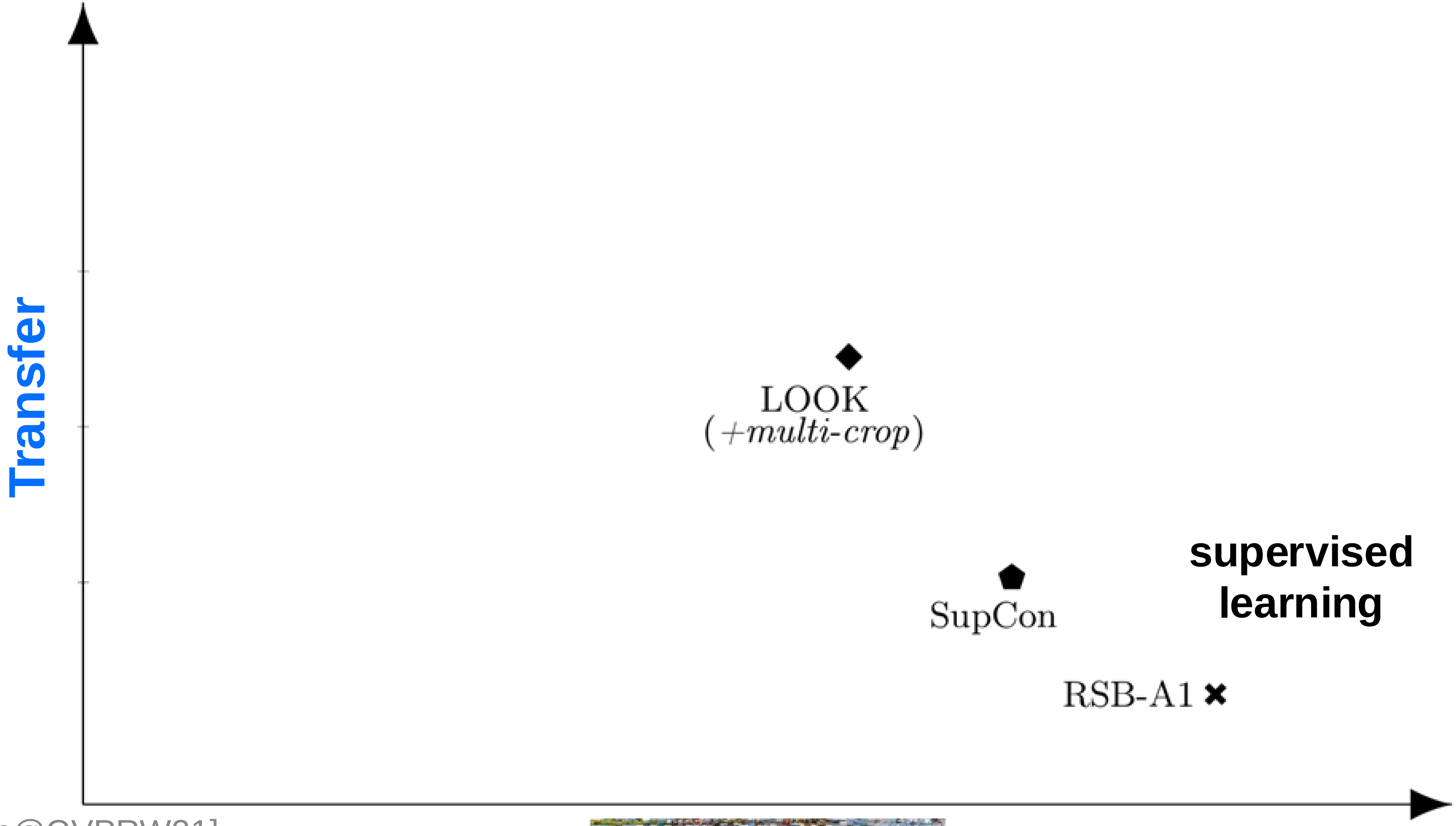
For the **Training** task + every **Transfer** task



CoG
+
8 datasets

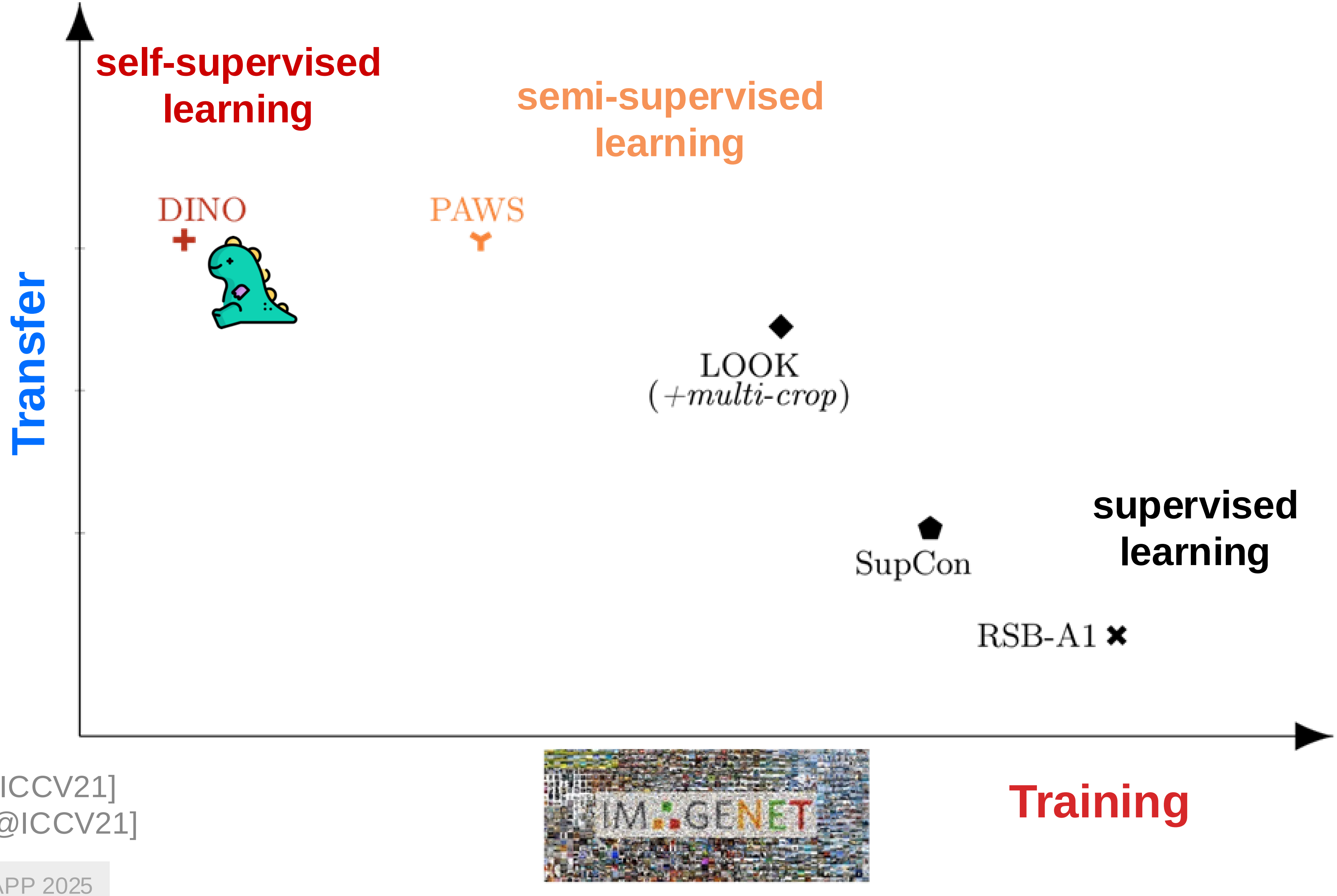
Transfer

Training

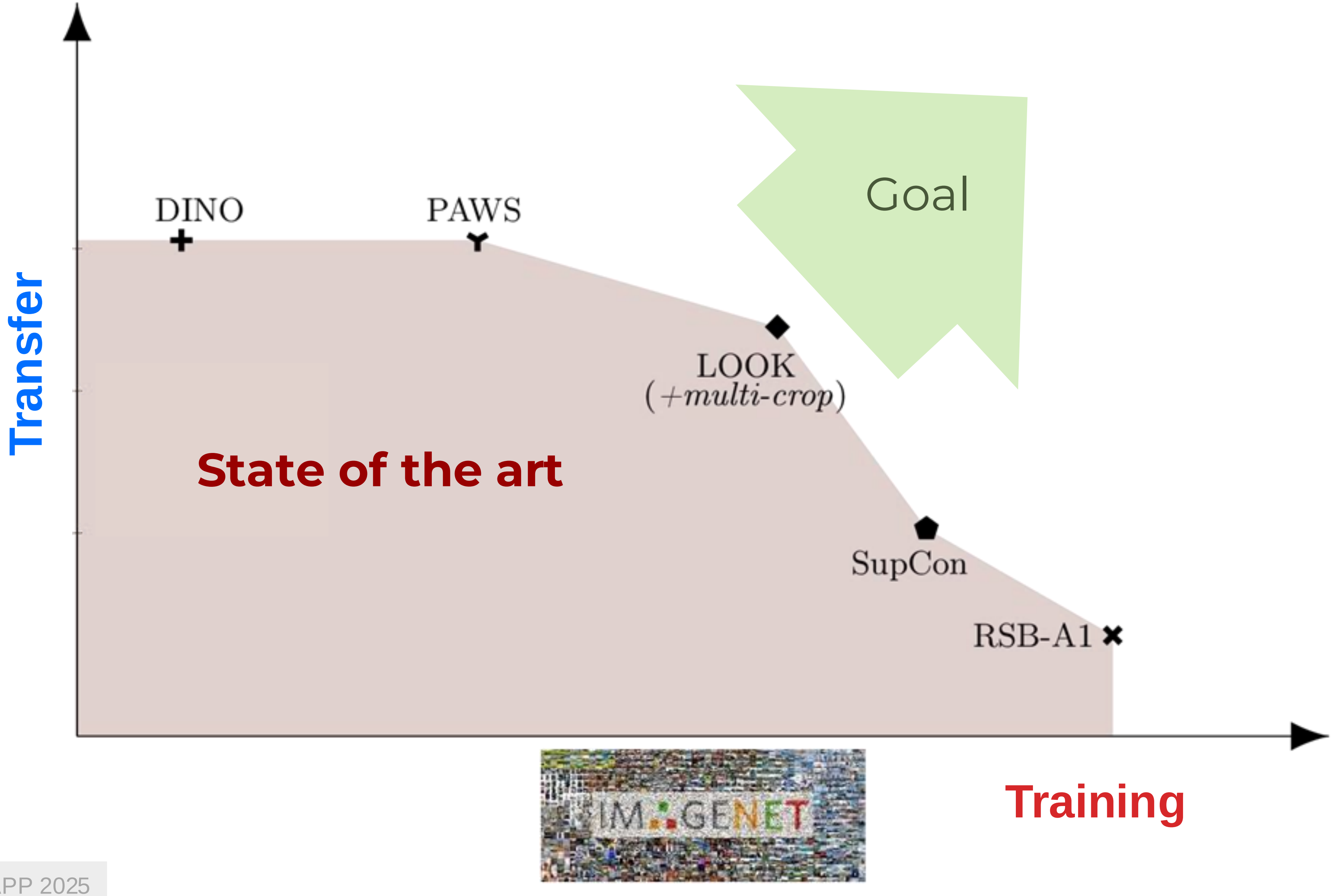


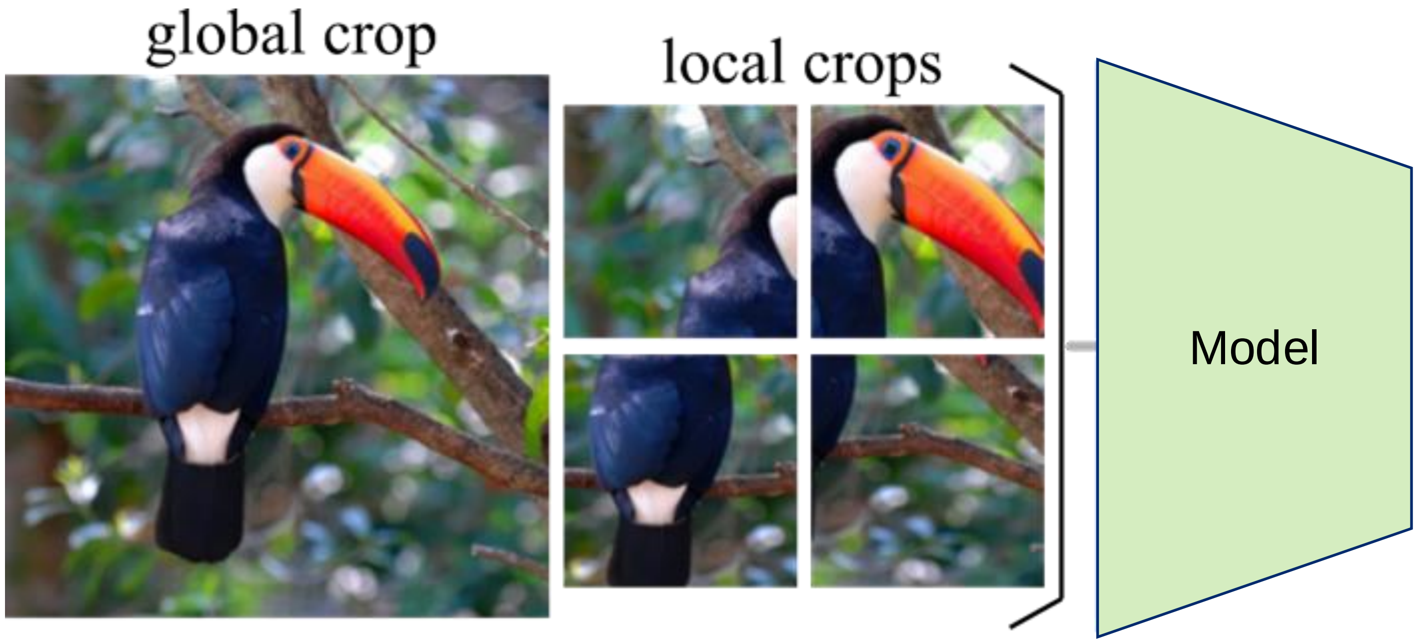
Training

[**RSB** = Wightman@CVPRW21]
[**SupCon** = Khosla@NeurIPS20]
[**Look** = Feng@ICLR22]



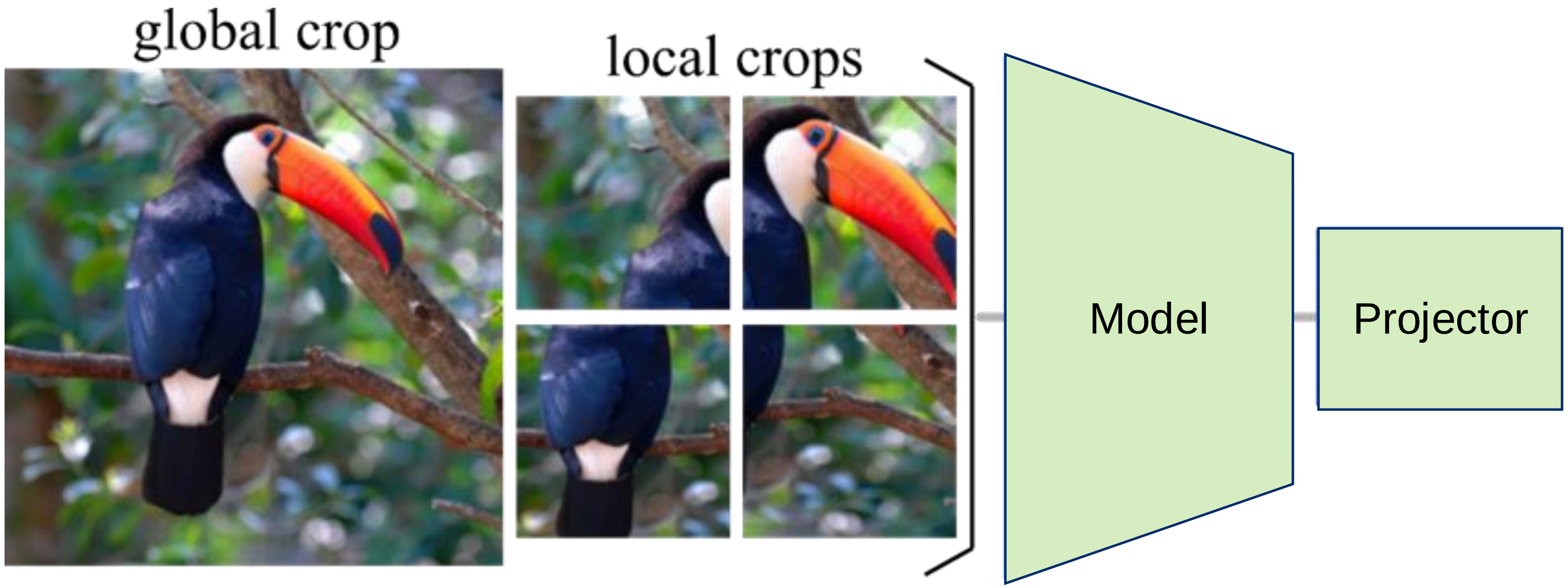
[**DINO** = Caron@ICCV21]
[**PAWS** = Assran@ICCV21]


Transfer
DINO
PAWS
Goal
LOOK
(+multi-crop)
State of the art
SupCon
RSB-A1
Training



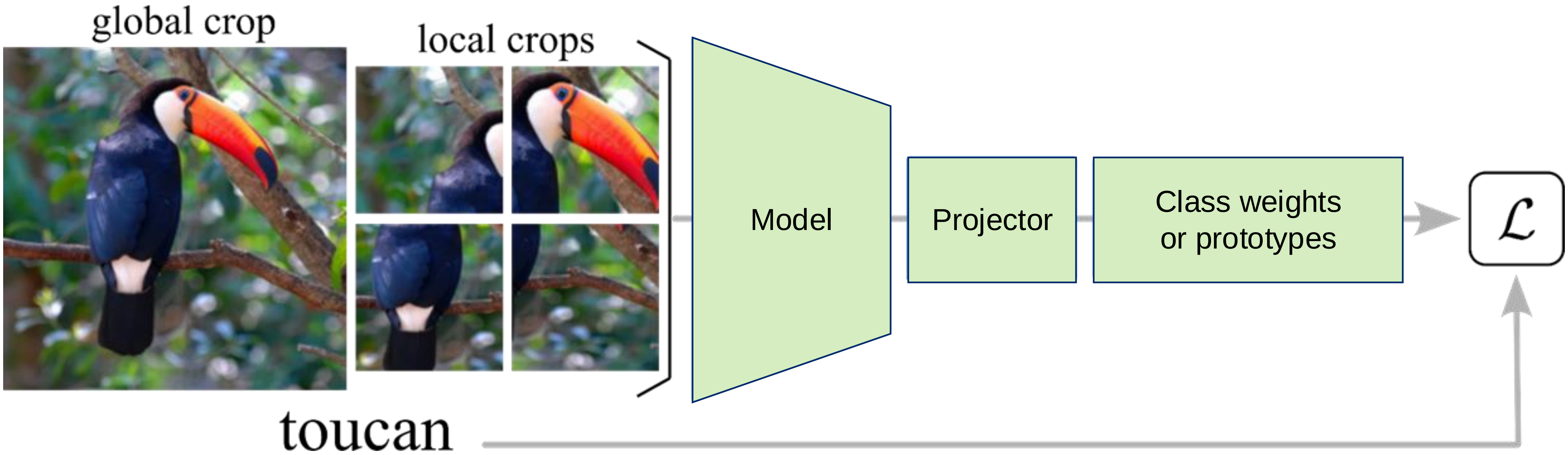
1. Multi-crop data augmentation

[**SWAV** = Caron@NeurIPS20]
[**DINO** = Caron@ICCV21]



1. Multi-crop data augmentation
2. Expendable projector head

[**MoCo** = He@CVPR20]
[**SimCLR** = Chen@ICML20]
[**MoChi** = Kalantidis@NeurIPS20]
[**DiNO** = Caron@NeurIPS21]
[Wang@CVPR22]

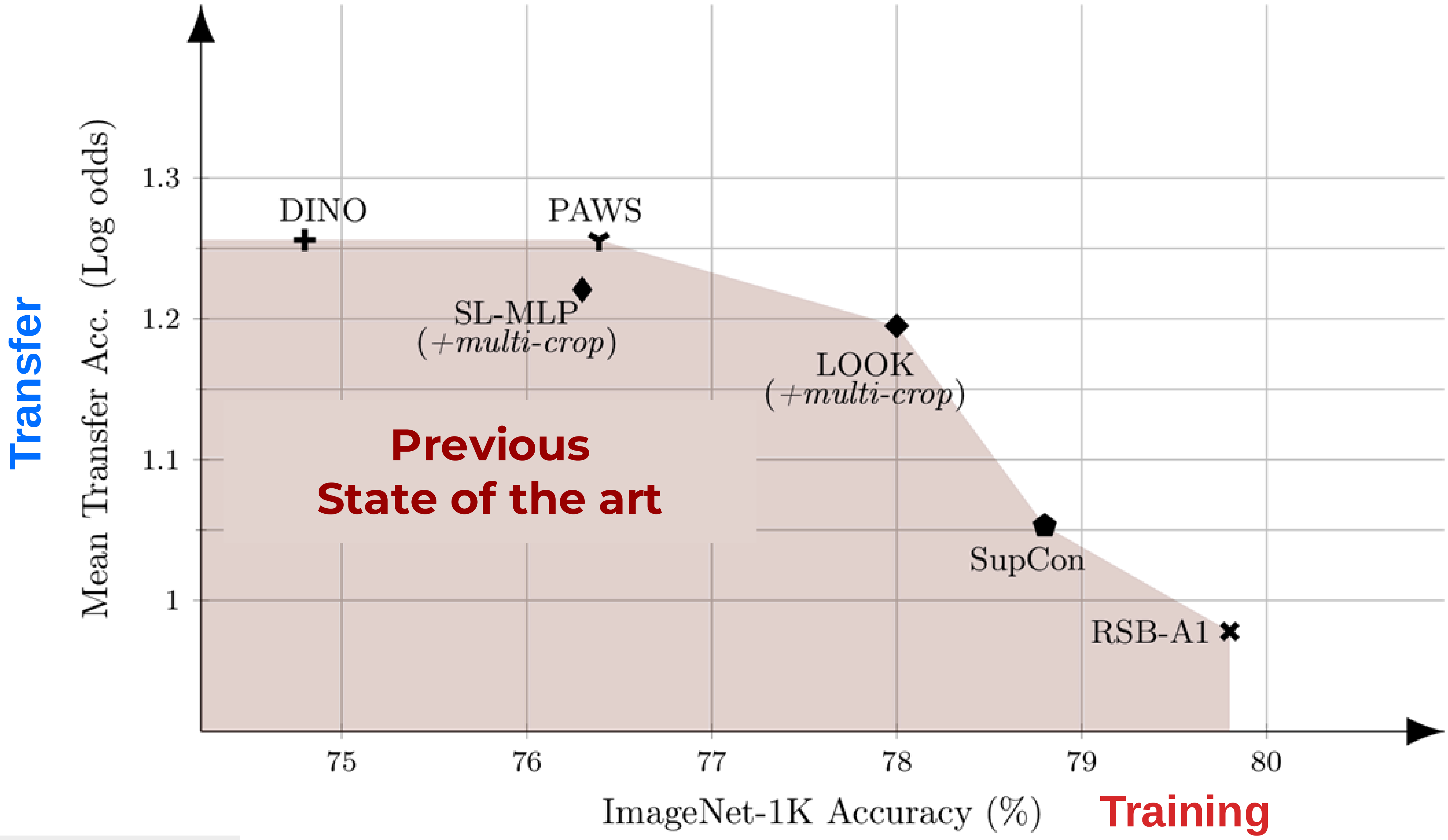


1. Multi-crop data augmentation
2. Expendable projector head
3. (*optional*) Replace class weights with class prototypes

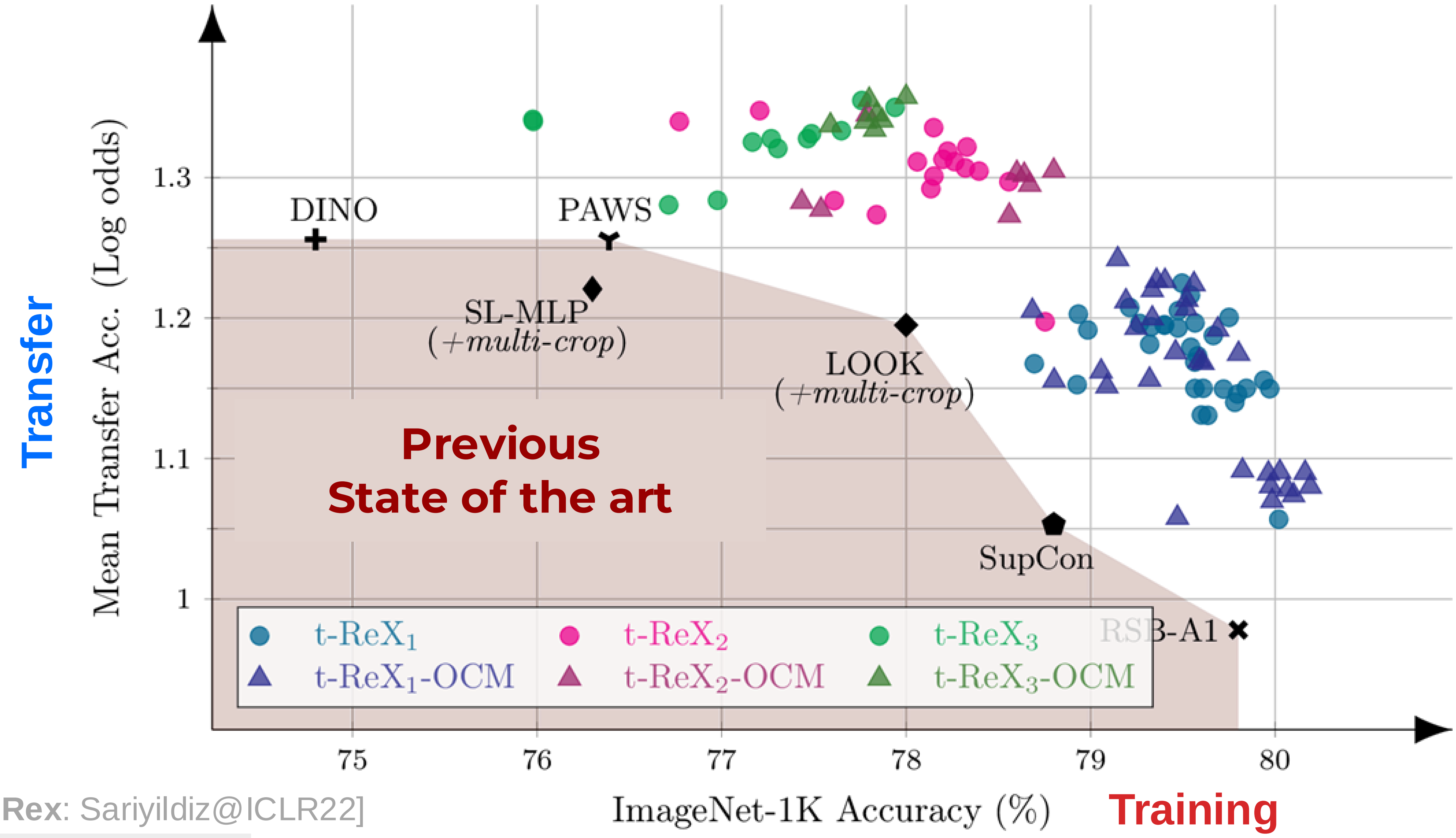
Nearest Class Means (NCM)

[**NCM** = Mensink@ECCV12]
[**DeepNCM** = Guerriero@W-ICLR18]

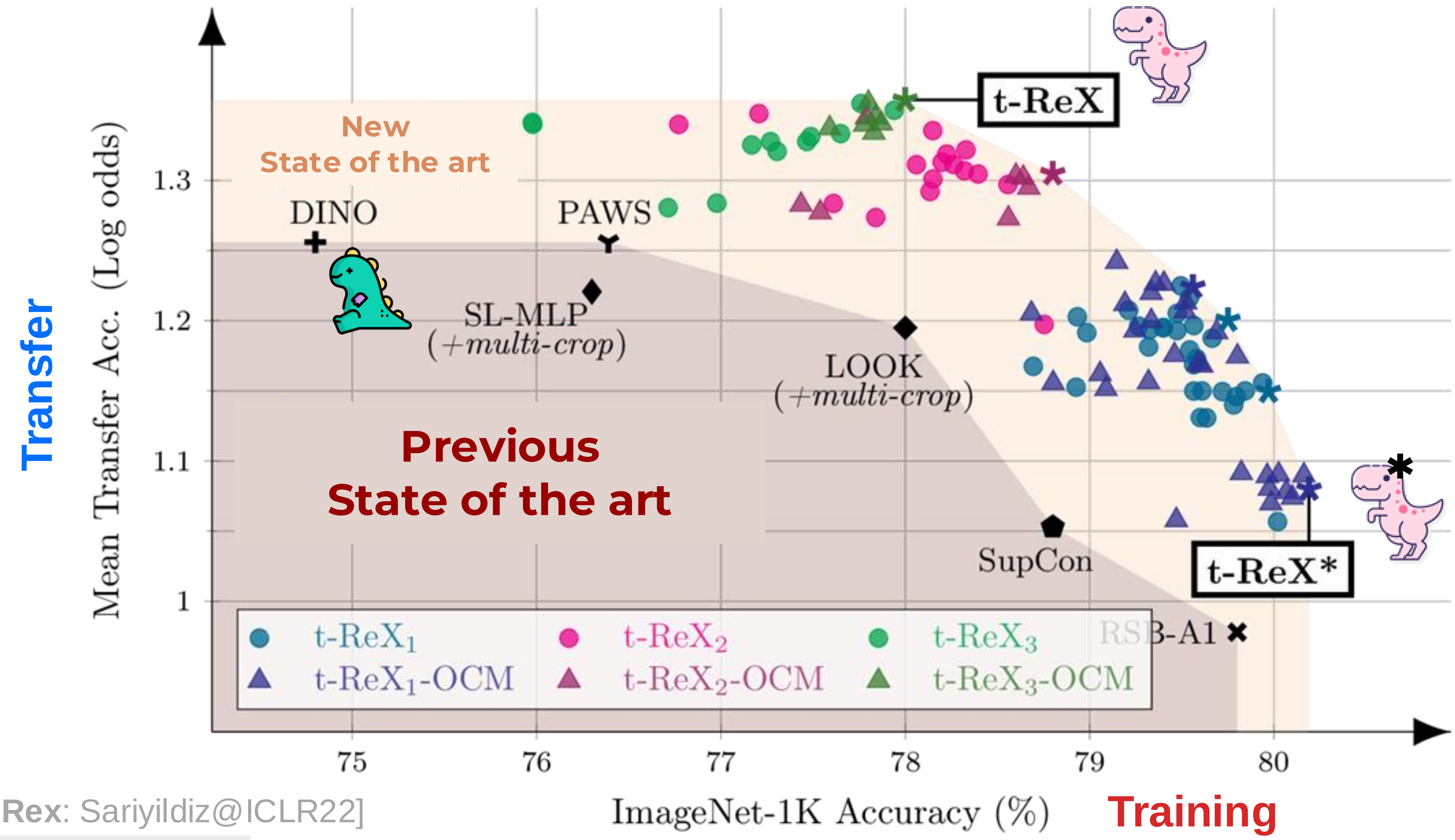
[**T-Rex**: Sariyildiz@ICLR22]



Transfer
Mean Transfer Acc. (Log odds)
1.3
1.2
1.1
1
DINO
PAWS
SL-MLP
(+multi-crop)
LOOK
(+multi-crop)
Previous
State of the art
SupCon
RSB-A1
75
76
77
78
79
80
ImageNet-1K Accuracy (%)
Training



[**T-Rex**: Sariyildiz@ICLR22]



[**T-Rex**: Sariyildiz@ICLR22]

Take home message

T-ReX

There is **no reason for no supervision**!

- Multi-crop data augmentation helps
- Expendable projector controls **Training / Transfer** trade-off

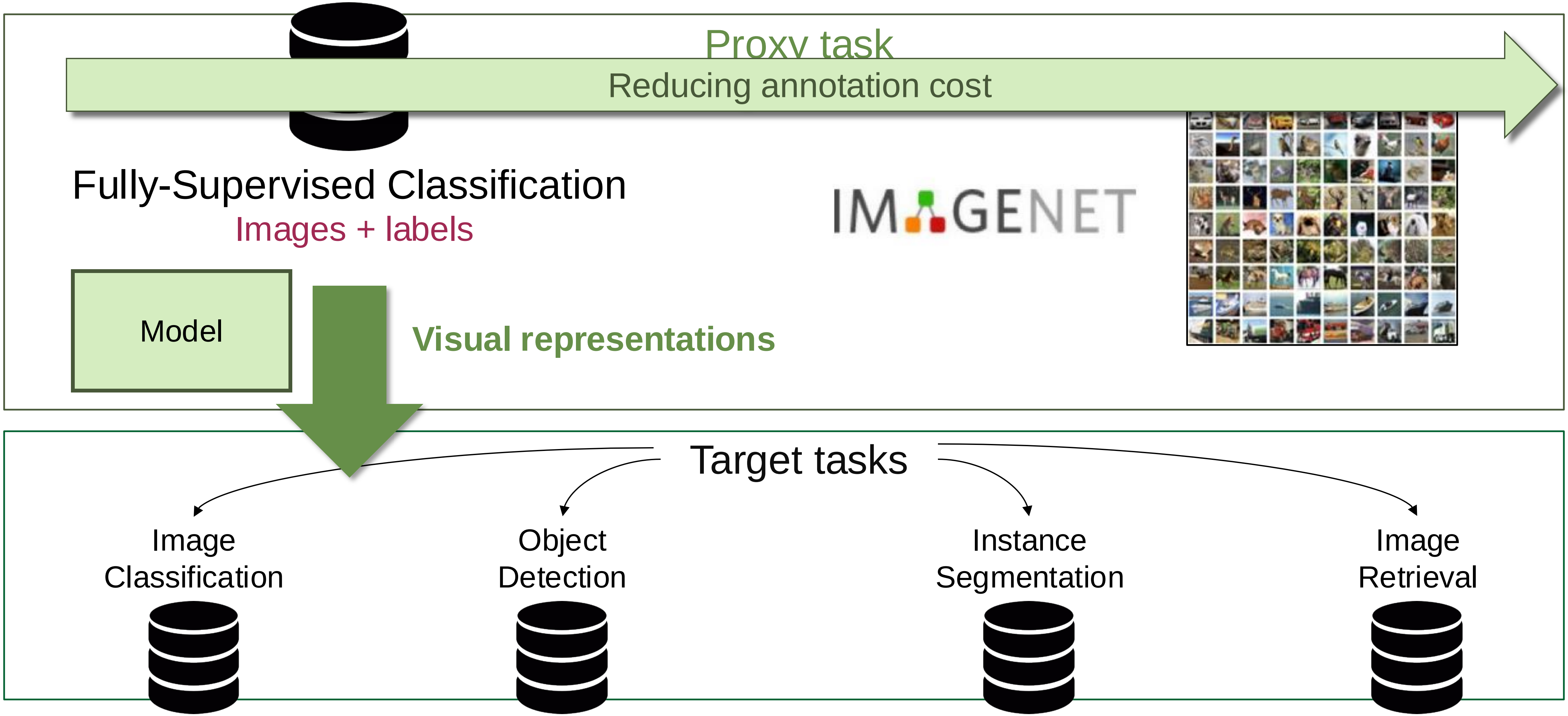
Reference

No Reason for No Supervision: Improved Generalization in Supervised Models

Mert Bulent Sariyildiz, Yannis Kalantidis, Karteek Alahari, Diane Larlus

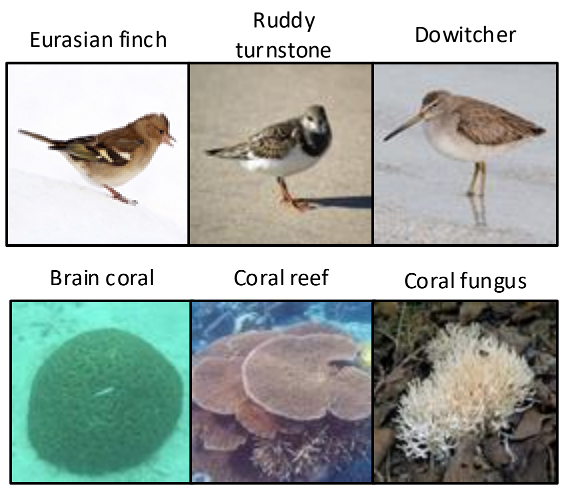
ICLR 2023

Pretraining visual representations from multimodal data .. or models


Proxy task
Reducing annotation cost
Fully-Supervised Classification
Images + labels
Model
Visual representations
IMAGENET
Target tasks
Image Classification
Object Detection
Instance Segmentation
Image Retrieval

Reducing annotation cost

Fully-Supervised
fine-grained annotations

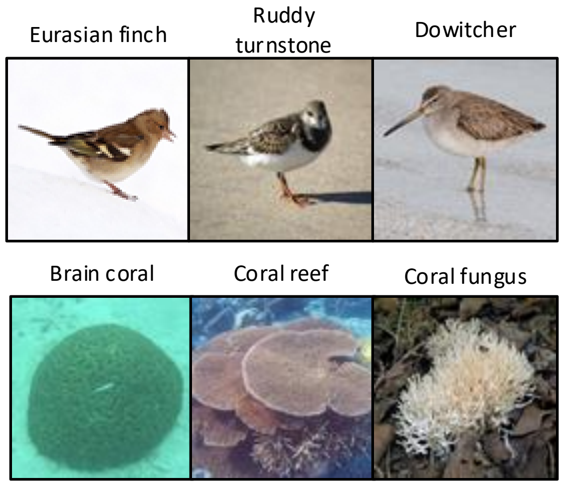


Self-supervised
annotation-free images

Reducing annotation cost

Fully-Supervised

fine-grained annotations



Caption-supervised

side information

a statue of a man stands in front of an old red bus.
a big and red bus with many displays for people to watch.
a red double decker bus parked next to a statue.
the double decker bus is beside a statue near restaurant tables.
a view of a bus sitting in front a small wooden statue.

a busy street with cars and trucks down it
an intersection with a view that looks towards a small downtown area.
cars parked on the side of the street and traveling down the road
an intersection with a stop light on a city street.
a street filled with lots of traffic under a traffic light.

Self-supervised

annotation-free images

Weak annotations

Reducing annotation cost

Fully-Supervised
fine-grained annotations

Eurasian finch | Ruddy turnstone | Dowitcher

Brain coral | Coral reef | Coral fungus

Caption-supervised
side information

a statue of a man stands in front of an old red bus.
a big and red bus with many displays for people to watch.
a red double decker bus parked next to a statue.
the double decker bus is beside a statue near restaurant tables.
a view of a bus sitting in front a small wooden statue.

a busy street with cars and trucks down it
an intersection with a view that looks towards a small downtown area.
cars parked on the side of the street and traveling down the road
an intersection with a stop light on a city street.
a street filled with lots of traffic under a traffic light.

Self-supervised
annotation-free images

Text

"Little girl holding red umbrella"

↓

Mask a token

↓

"Little girl holding red [MASK]"

↓

Language Model

⇩

[MASK] = Umbrella

BERT model
[Delvin *et al.* 2018]

Input: Image

Caption

“Little girl holding red umbrella”

Mask a token

“Little girl holding red [MASK]”

[MASK] = Umbrella

Input: Image

Caption

"Little girl holding red umbrella"

Mask a token

"Little girl holding red [MASK]"

Visual representation
(learnt from scratch)

Textual representation
(frozen)

Multi-modal network =
Auxiliary modules

[MASK] = Umbrella

Image-Conditioned **Masked** Language Modeling Task (ICMLM)

[**ICMLM** = Sariyildiz@ECCV20]

[**VirTex** = Desai@CVPR21]

Weak annotations

Reducing annotation cost

[**ICMLM** = Sariyildiz@ECCV20]

[**VirTex** = Desai@CVPR21]

Caption-supervised

side information

smaller sets

a statue of a man stands in front of an old red bus.
a big and red bus with many displays for people to watch.
a red double decker bus parked next to a statue.
the double decker bus is beside a statue near restaurant tables.
a view of a bus sitting in front a small wooden statue.

Weak annotations

Reducing annotation cost

Dataset scale

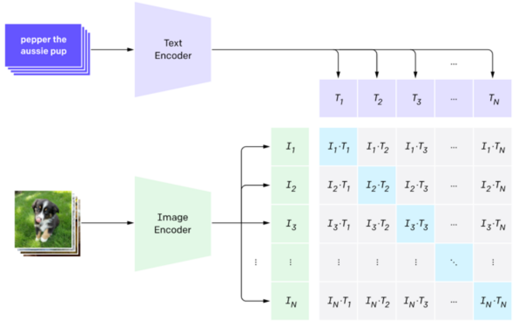
[**ICMLM** = Sariyildiz@ECCV20]

[**VirTex** = Desai@CVPR21]

Caption-supervised
side information
smaller sets

a statue of a man stands in front of an old red bus.
a big and red bus with many displays for people to watch.
a red double decker bus parked next to a statue.
the double decker bus is beside a statue near restaurant tables.
a view of a bus sitting in front a small wooden statue.

[**CLIP** = Radford@ICLM21]



Unfiltered
Image + Text
large scale

Text-to-image generation

[**DALL-E** = Ramesh@ICML21]

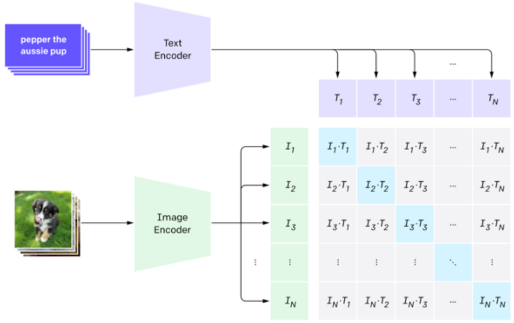
[**DALL-E2** = Saharia@NeurIPS21]

[**DALL-E3** = Betker@TechReport23]

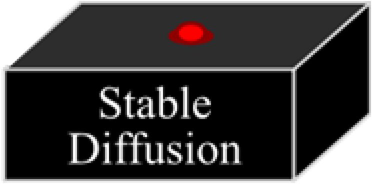
[**Stable diffusion** = Rombach@CVPR22]

[**Stable diffusion 3** = Esser@Arxiv24]

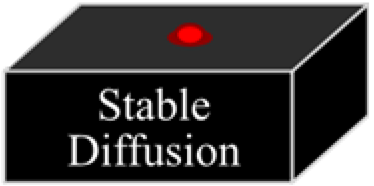
[**CLIP** = Radford@ICLM21]



Unfiltered
Image + Text
large scale

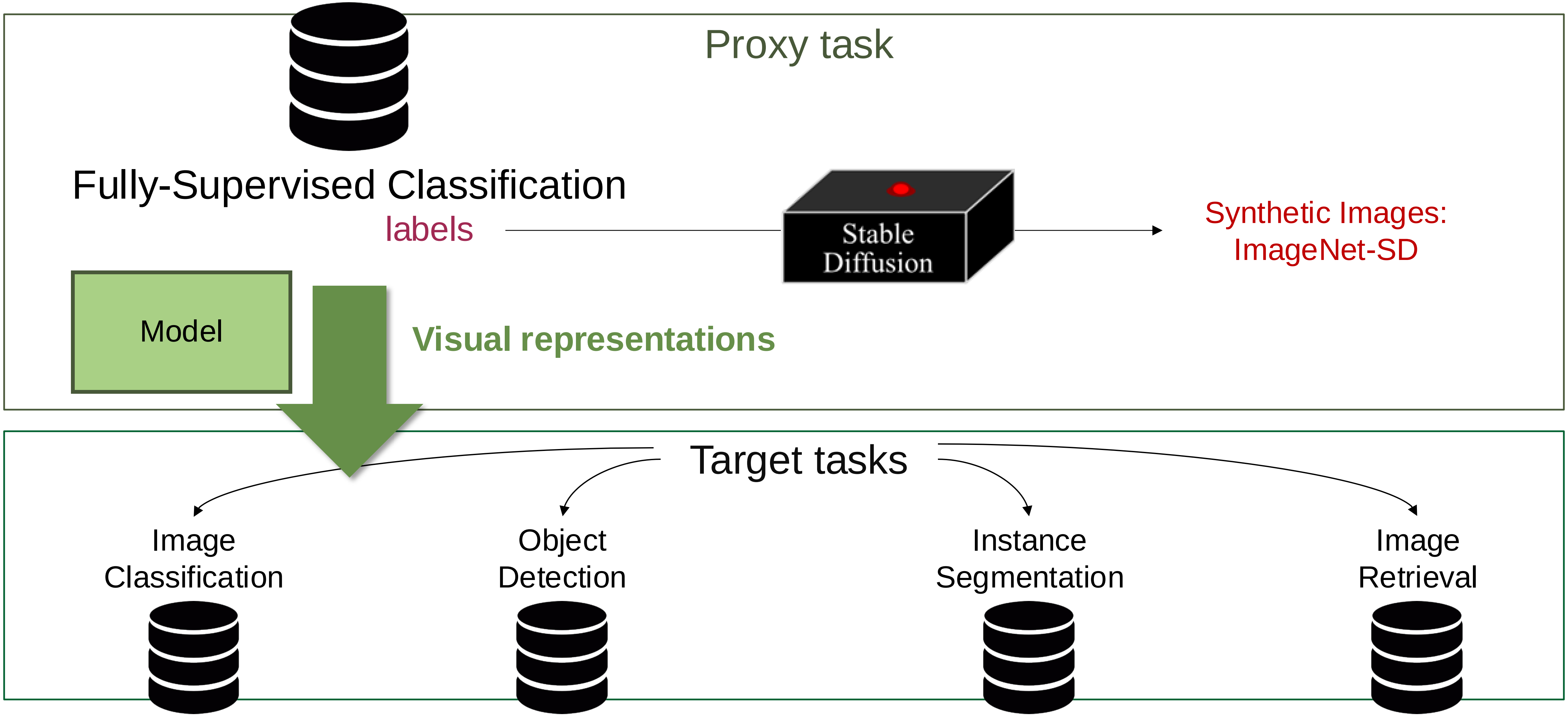


cat

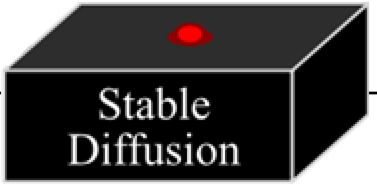


[**Stable Diffusion** = Rombach@CVPR22]

***Do we still need actual images
to (pre-)train visual representations?***


Proxy task
Fully-Supervised Classification
labels
Stable Diffusion
Synthetic Images: ImageNet-SD
Model
Visual representations
Target tasks
Image Classification
Object Detection
Instance Segmentation
Image Retrieval

prompt = class name



Synthetic Image

papillon

lorikeet

pirate ship

Semantic errors

Lack of diversity

Domain issues

“papillon” class in ImageNet

“pirate ship” class in ImageNet

prompt = class name

prompt = class name, hypernym*

prompt = class name, description*

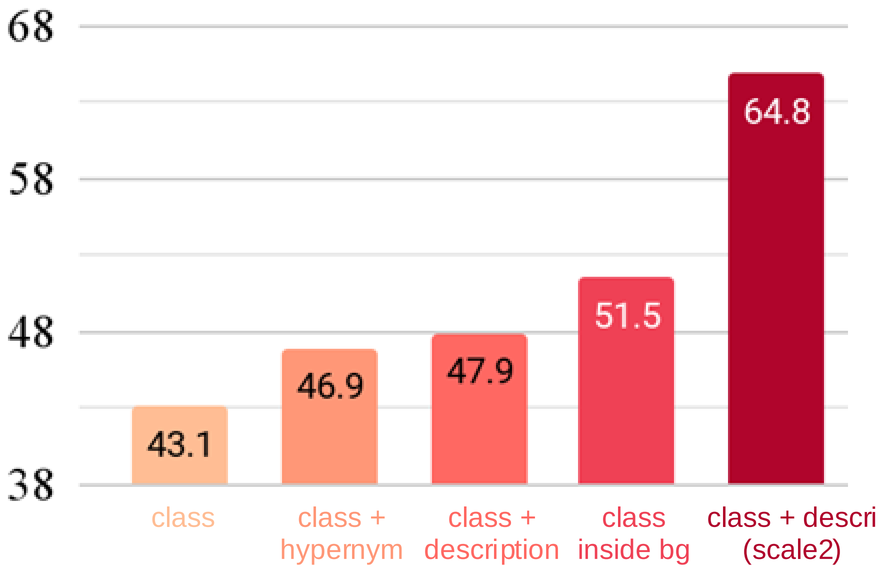
prompt = class name, hypernym inside background**

prompt = class name, description (+ reduce guidance scale)

* from **Wordnet** lexical database

** from **Places 365** dataset

Performance on ImageNet-100-Val (**Top-1** acc - **real images**)



How well does each model perform when classiying ***real images****?*

[**ImageNetSD** = Sariyildiz@CVPR23]

prompt = class name

prompt = class name, hypernym*

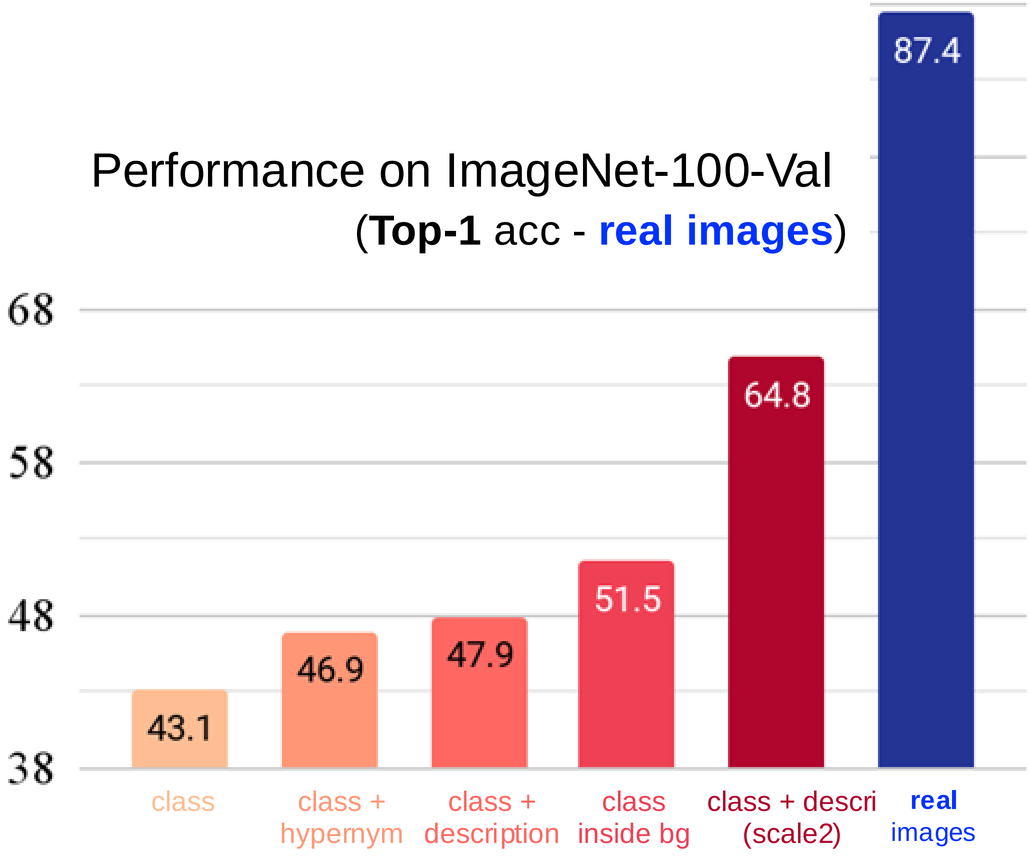
prompt = class name, description*

prompt = class name, hypernym inside background**

prompt = class name, description (+ reduce guidance scale)

* from **Wordnet** lexical database

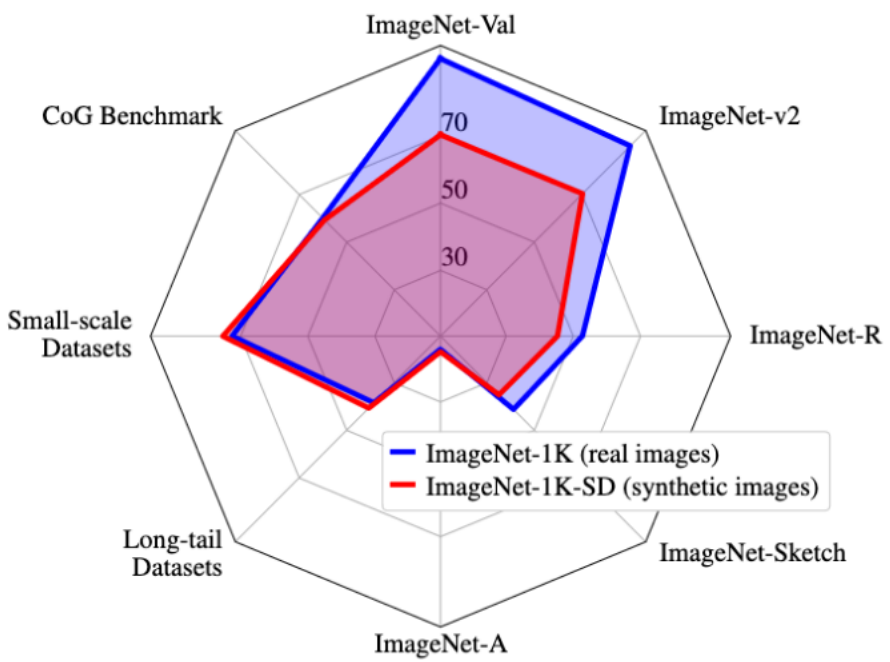
** from **Places 365** dataset



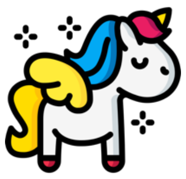
[**ImageNetSD** = Sariyildiz@CVPR23]

Do we still need actual images to pretrain visual representations?

- Promising results on the ImageNet variants
- Strong transfer results

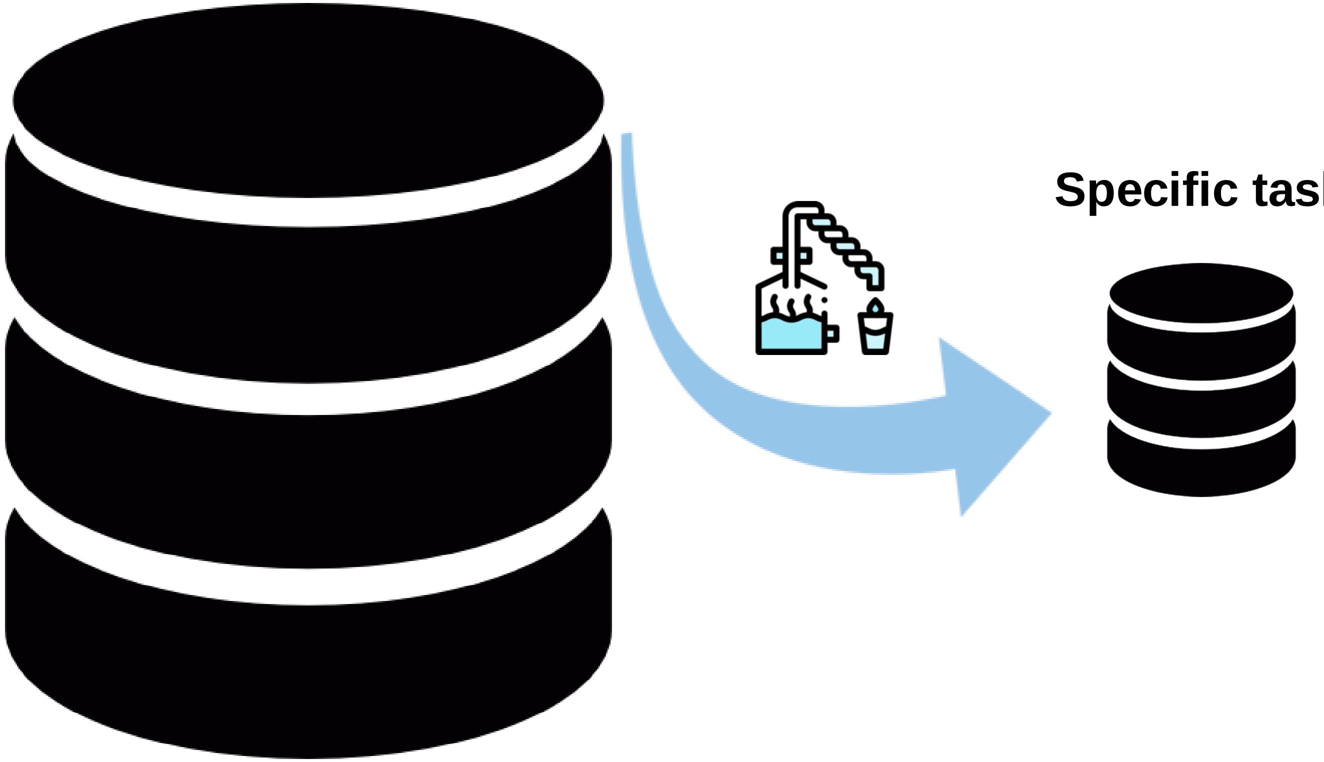


Reference
Fake it till you make it: Learning transferable representations from synthetic ImageNet clones
Mert Bulent Sariyildiz, Karteek Alahari, Diane Larlus, Yannis Kalantidis
CVPR 2023



Once we've pretrained, what do we do?

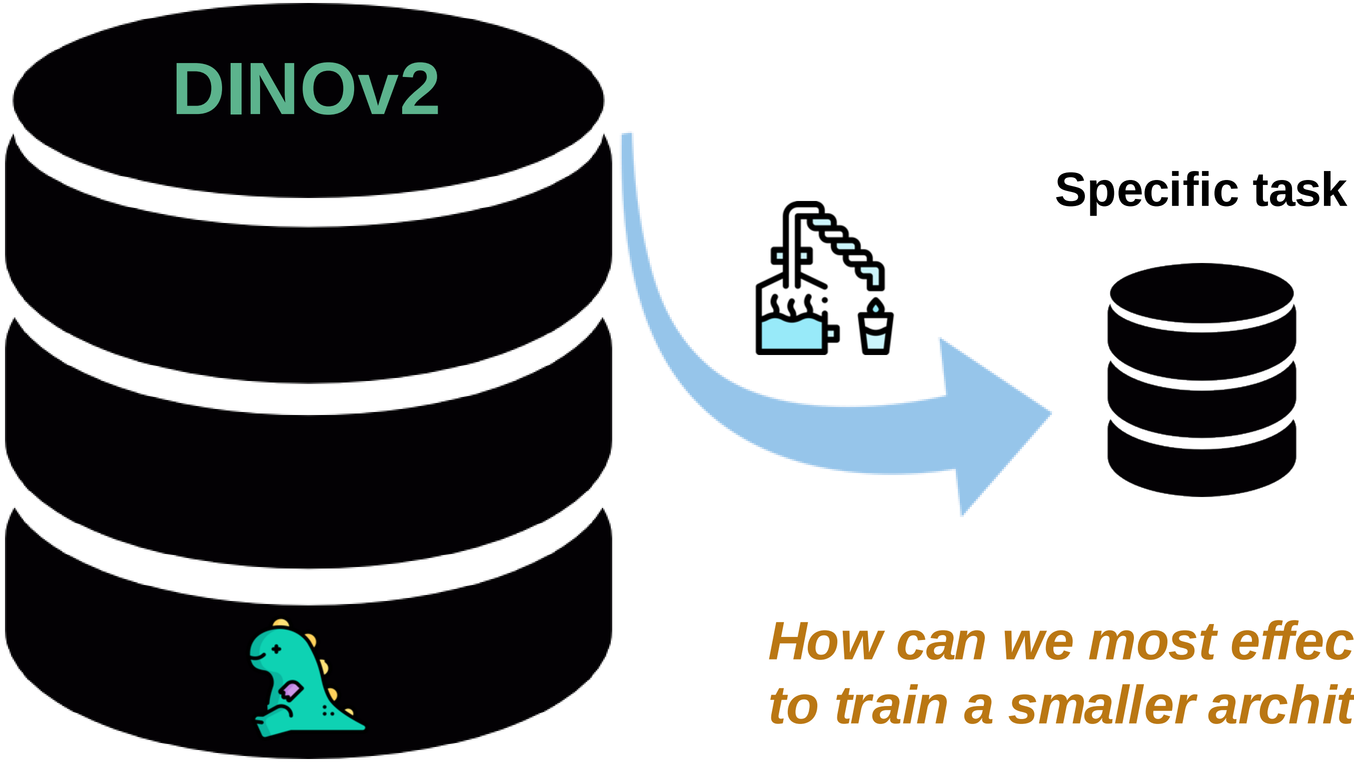
Adapting to new tasks



Distillation

[Hinton@W_NeurIPS15]

Adapting to new tasks from DINOv2



Distillation

How can we most effectively leverage those large models to train a smaller architecture, for a specialized task?

Task-agnostic distillation

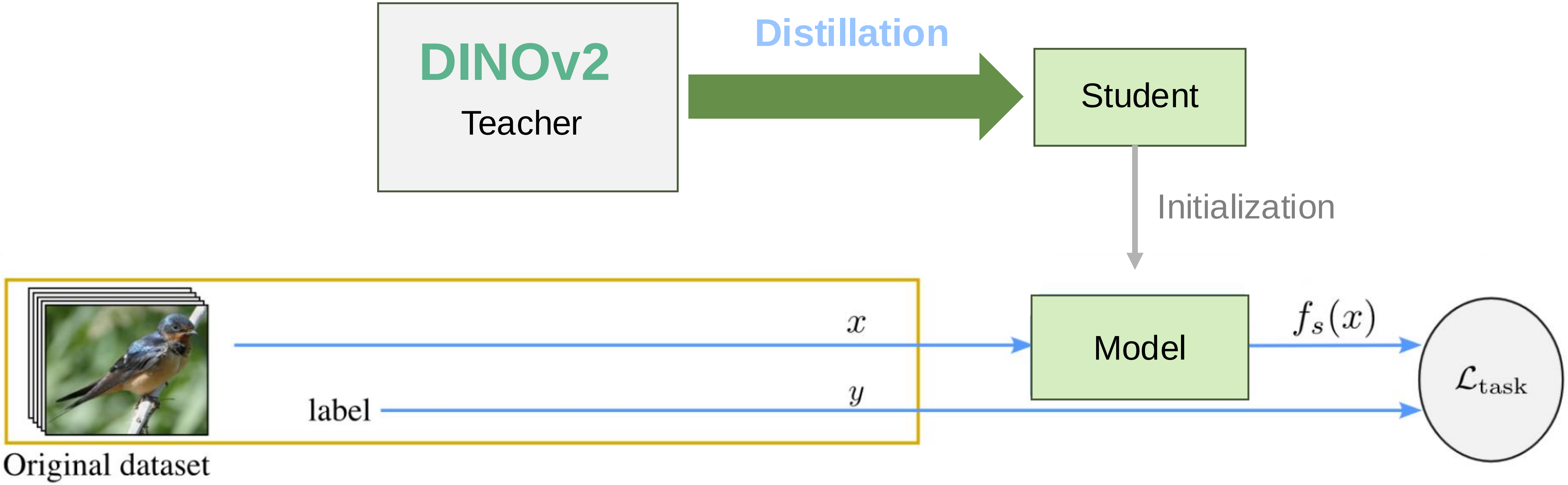
Step 1: reduce the teacher to the target architecture using distillation

Step 2: use this model to initialize the student

[Sun@EMNLP19]

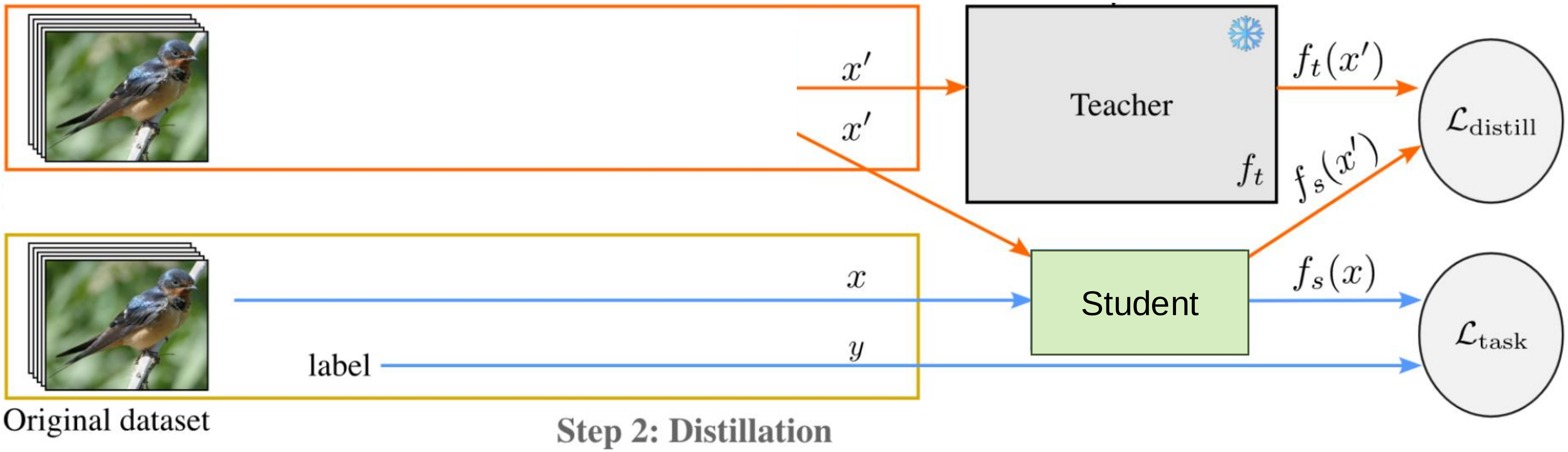
[Touvron@ICML21]

[Beyer@CVPR22]



Task-specific distillation

What should the teacher look like?

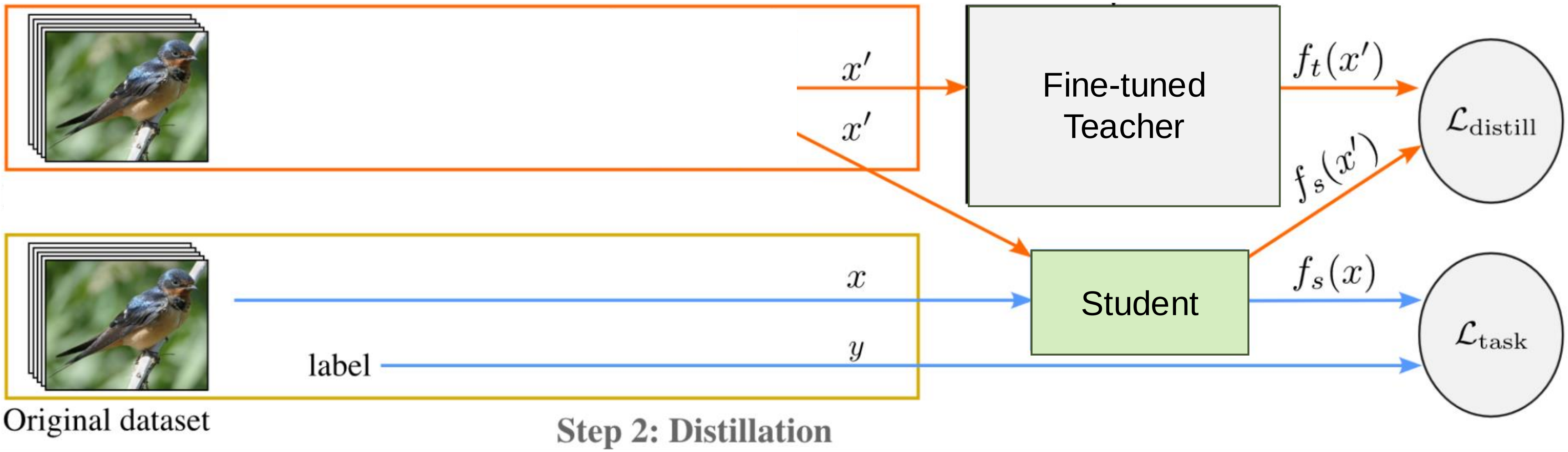


Task-specific distillation

Standard strategy: **fine-tune** the teacher for the task

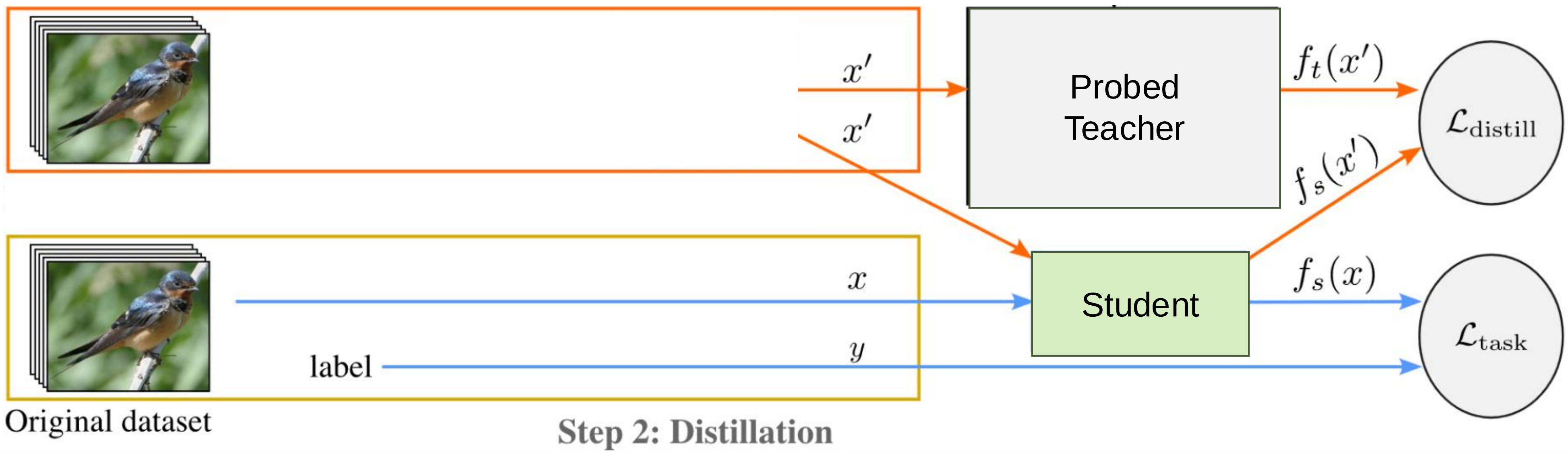
[Huang@CVPR23]

Issues
- Cost
- Not necessarily optimal

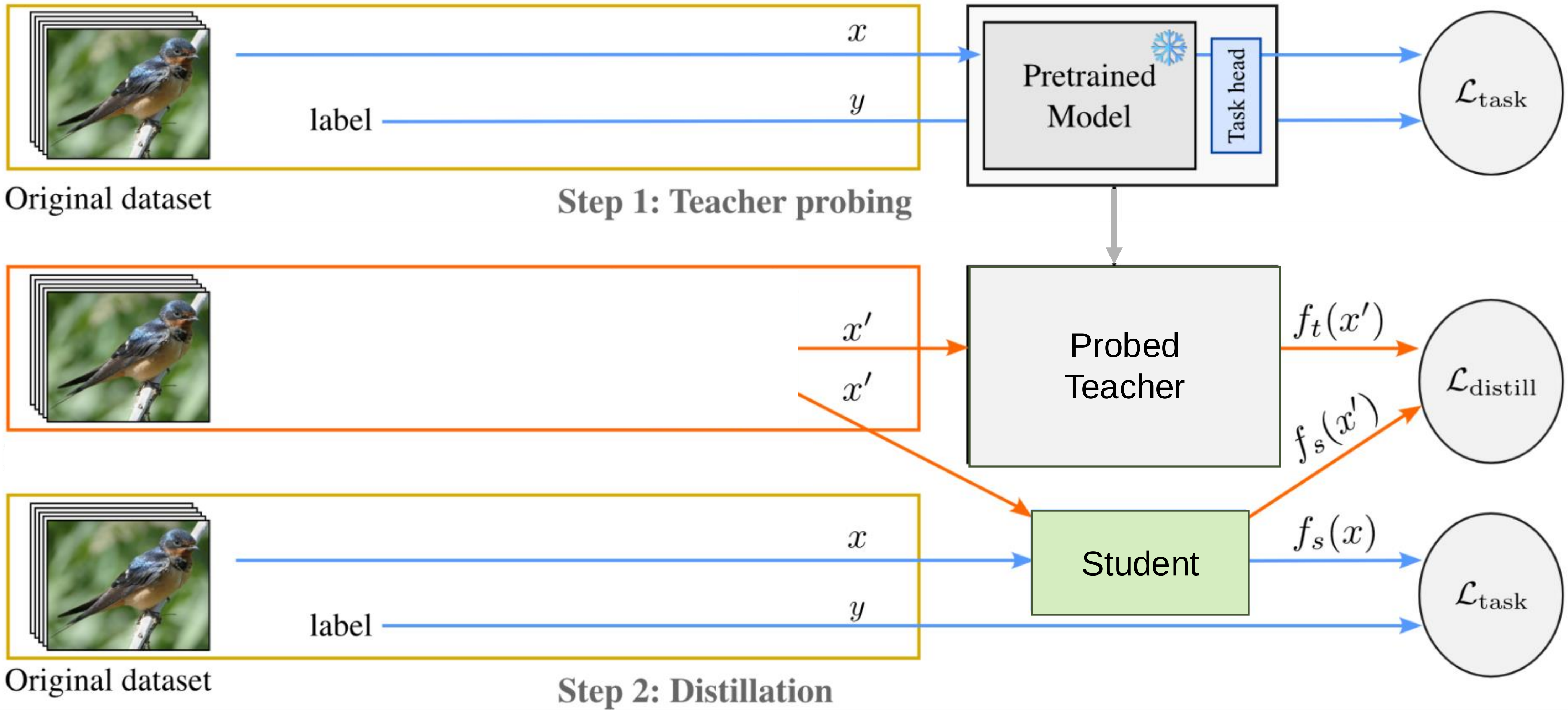


Task-specific distillation

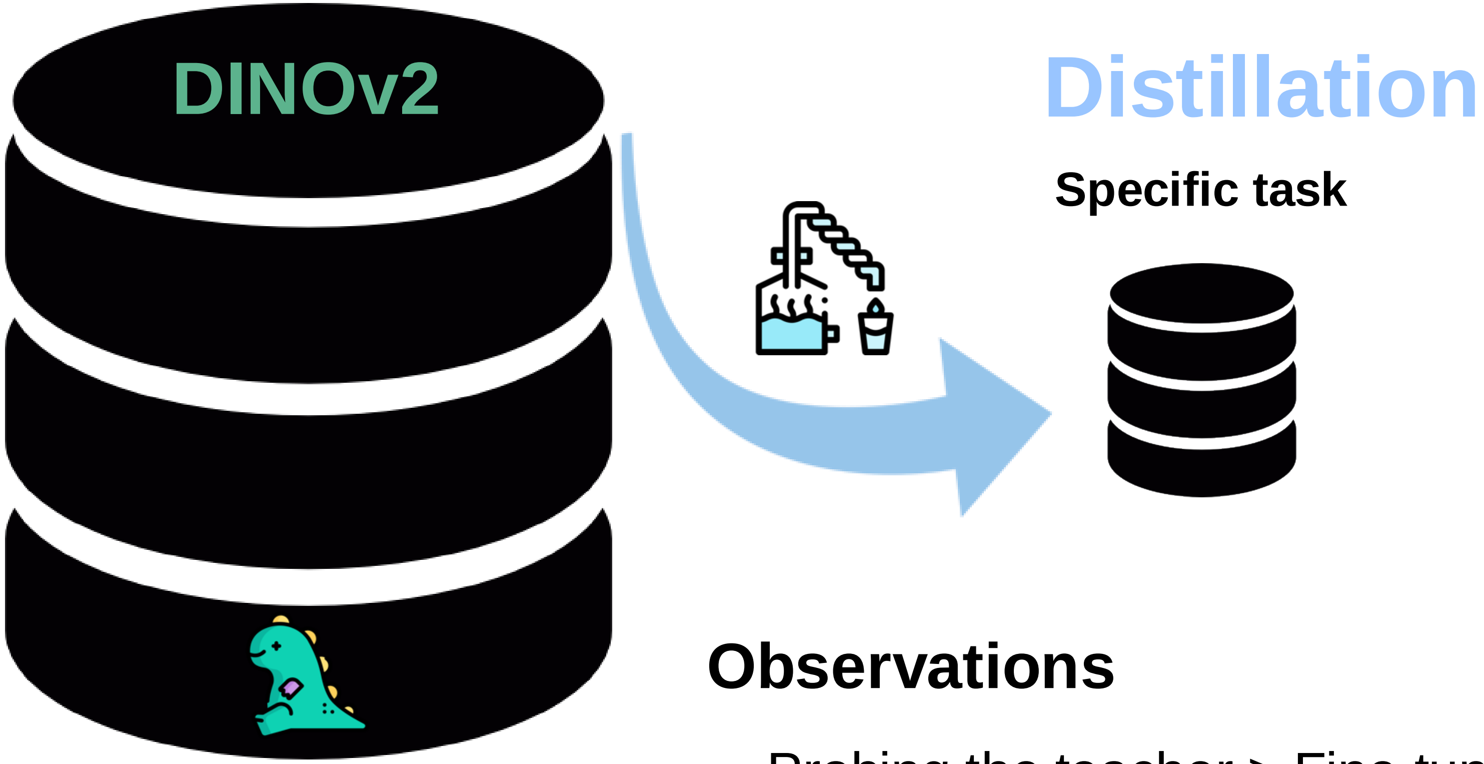
Proposed strategy: **probe** the teacher to the task



Task-specific distillation



Adapting to new tasks from DINOv2

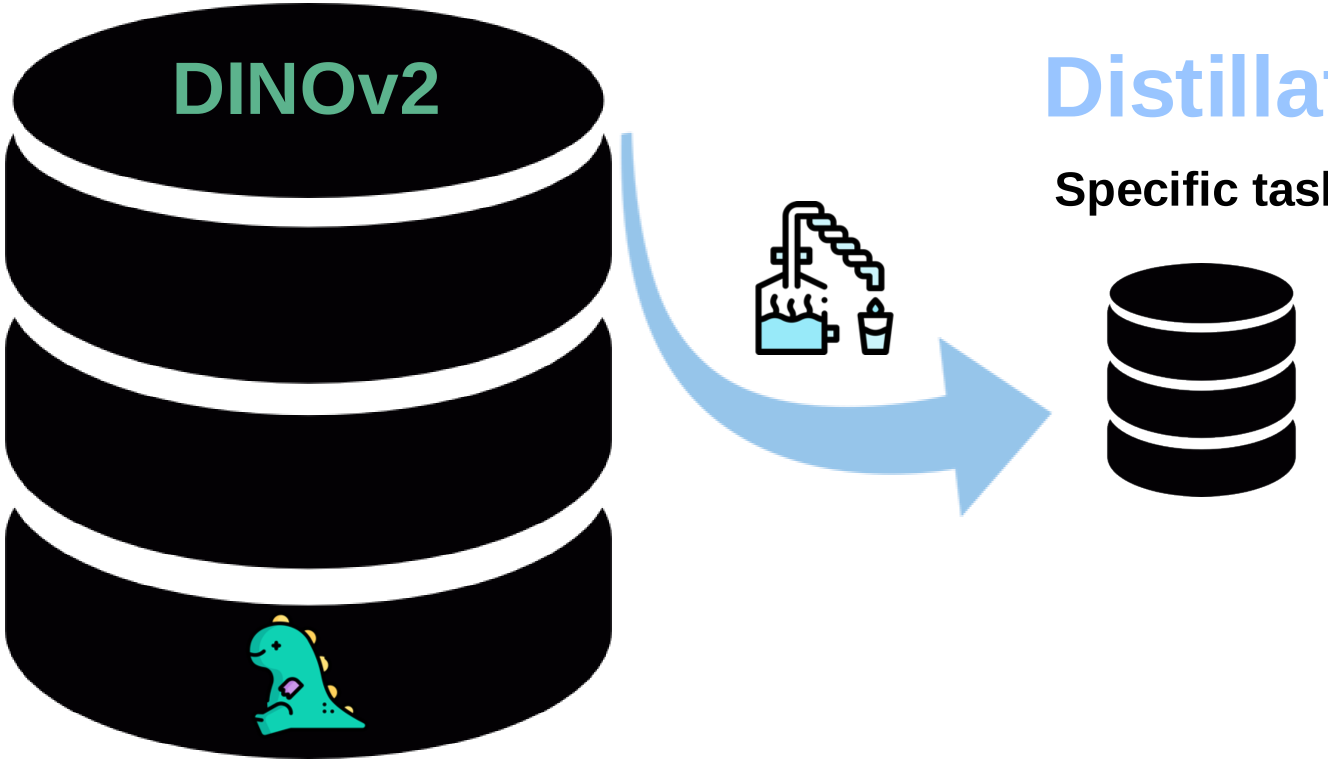


Observations

- Probing the teacher > Fine-tuning it
- Task-specific distillation complements Task-agnostic distillation
- Drastic model size changes are possible

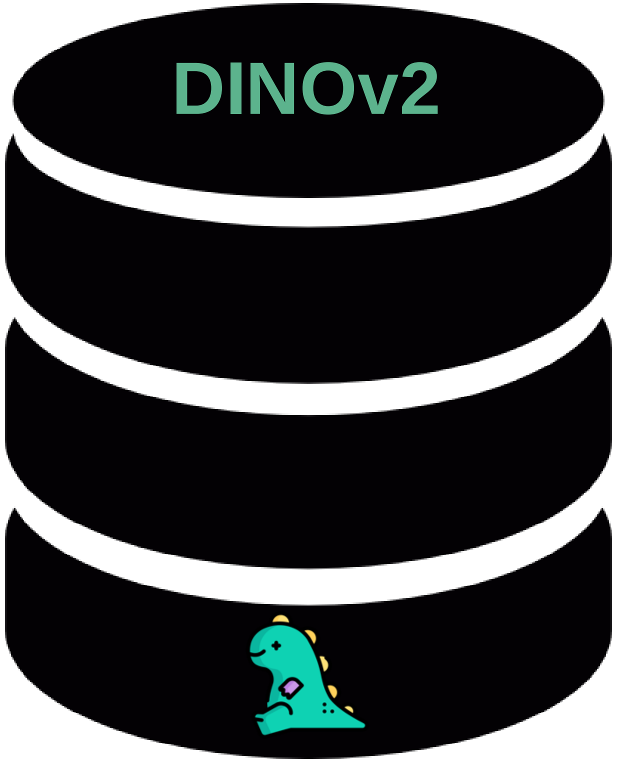
[Marrie@TMLR24]

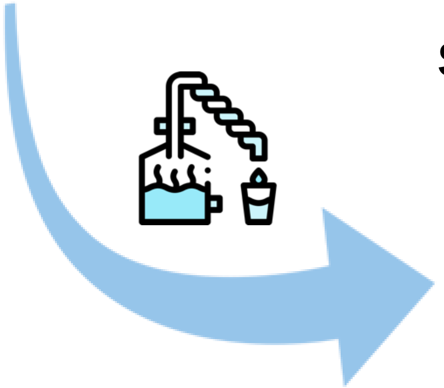
Adapting to new tasks from DINOv2



Do we have enough data to distill from?

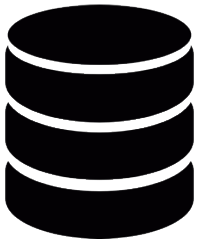
Adapting to new tasks from DINOv2





Distillation

Specific task



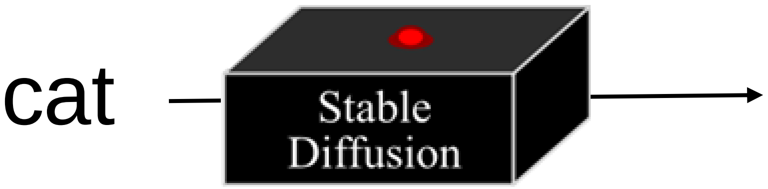
Potential solution: synthetic data

- Use text-to-image generation

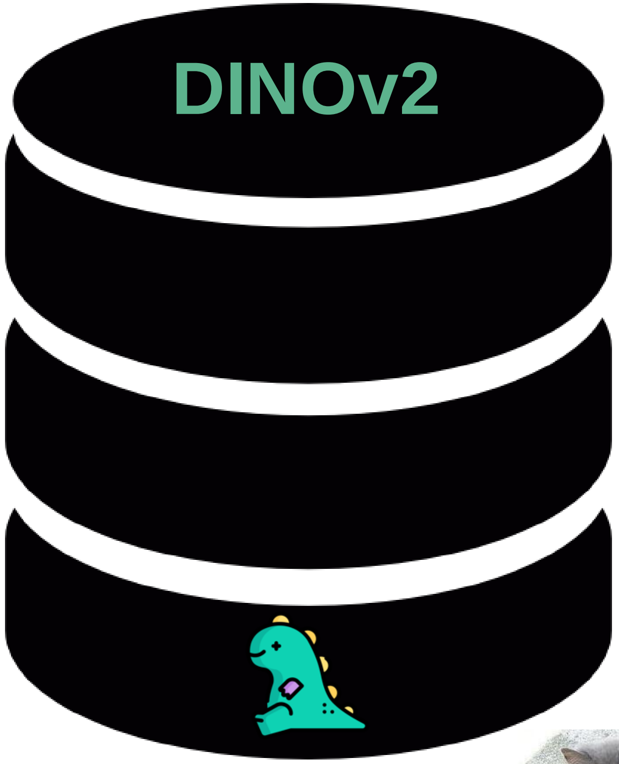
  Stable Diffusion [Rombach@CVPR22]

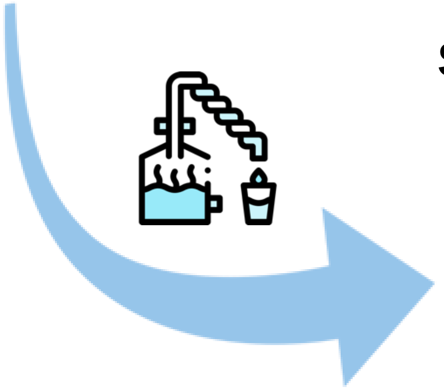
Issues

- Requires to know the class names
- Challenging beyond classification



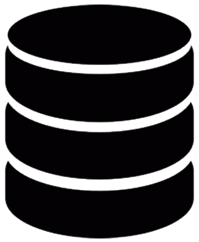
Adapting to new tasks from DINOv2





Distillation

Specific task



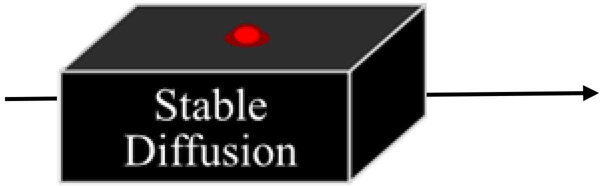
Potential solution: synthetic data

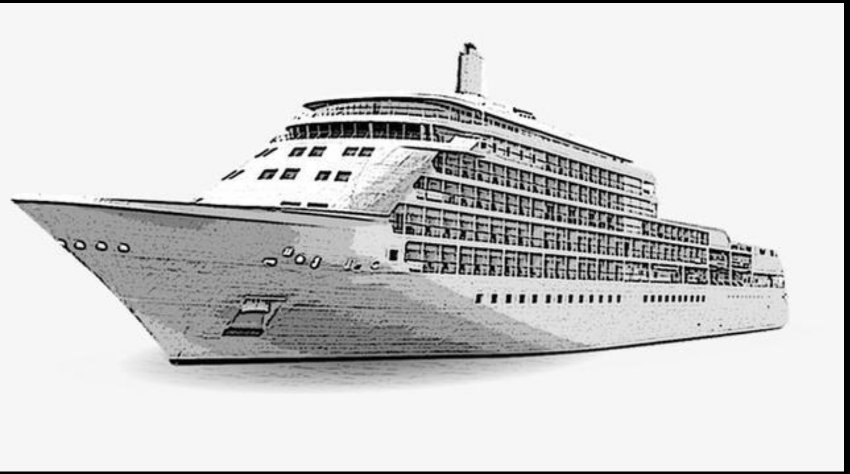
- Simply combining existing data

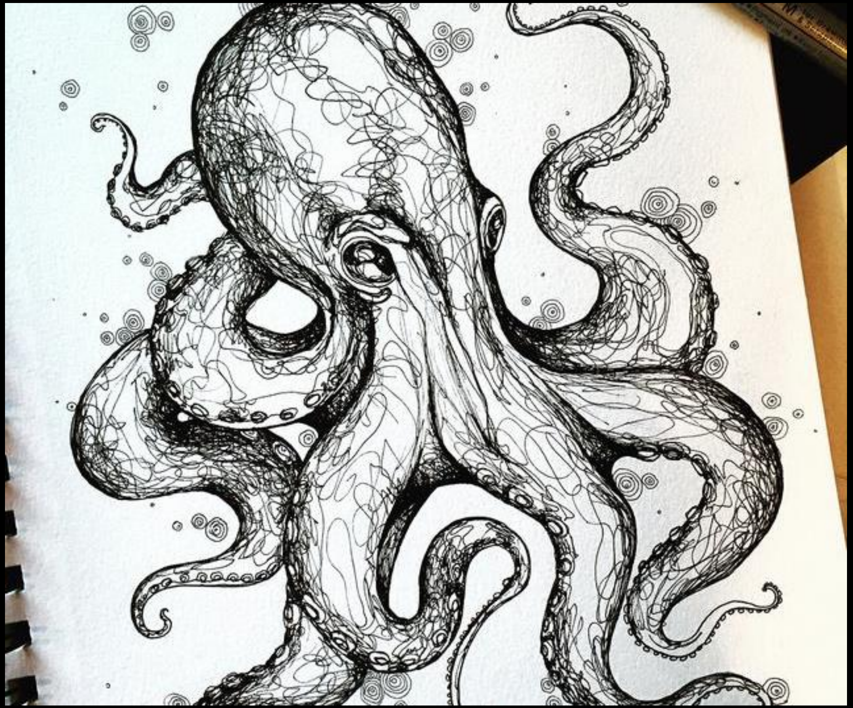
ImageMixer [Pinkney22]

Advantages

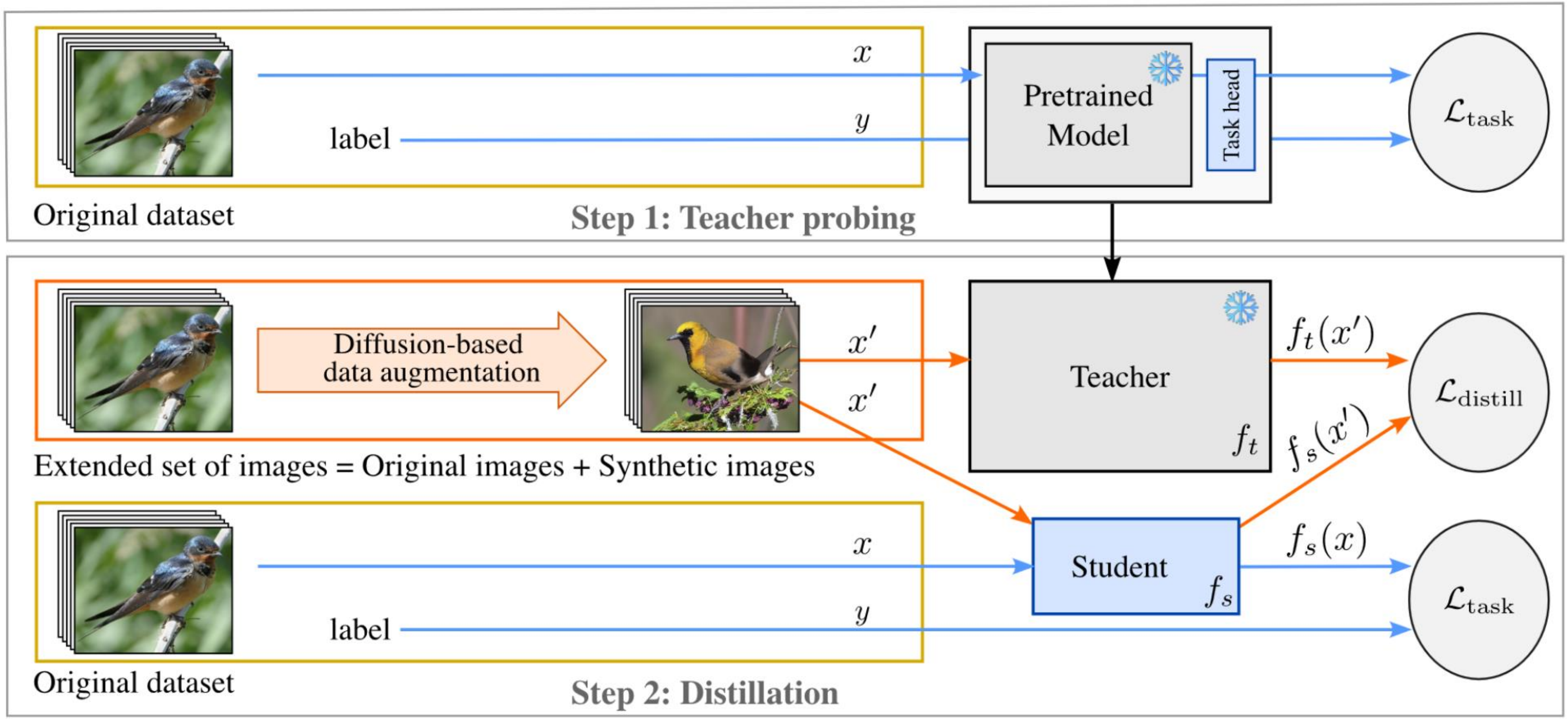
- Class-agnostic
- Can merge across classes







Final distillation pipeline



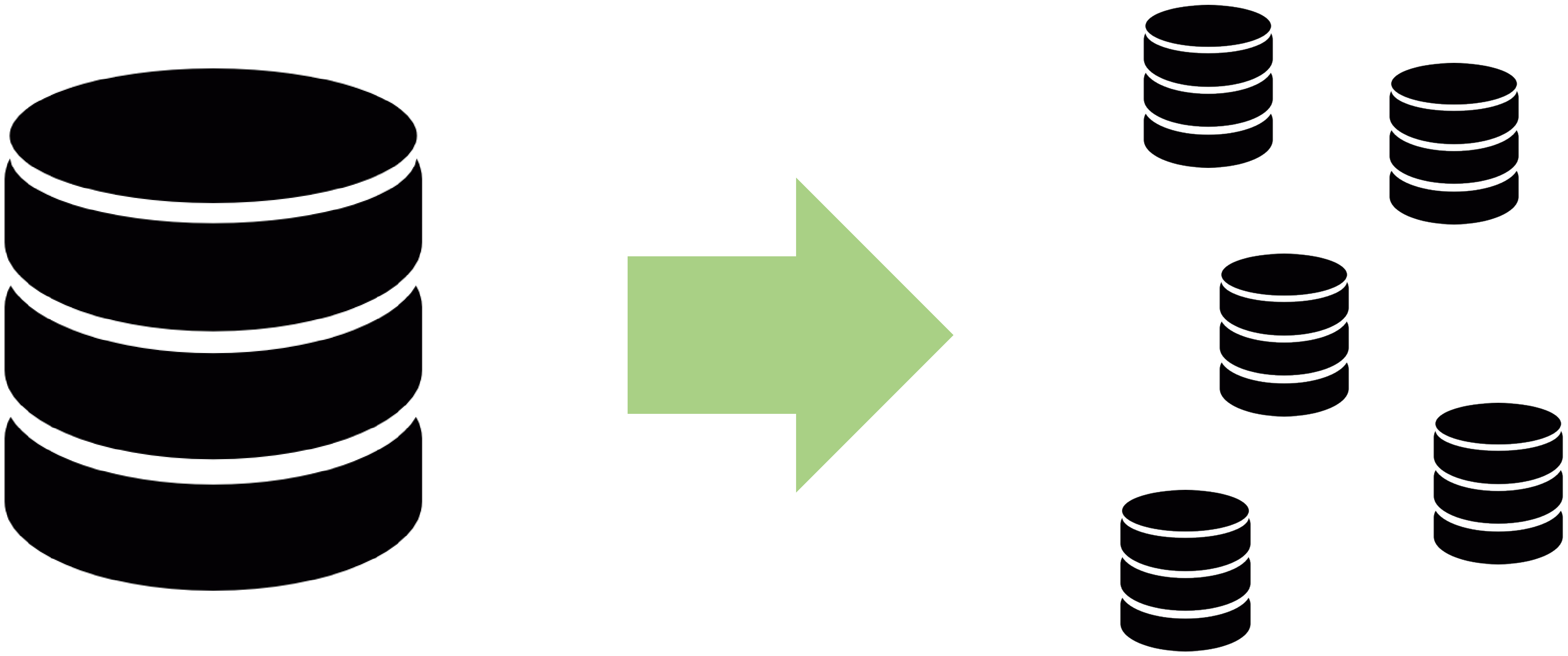
Reference

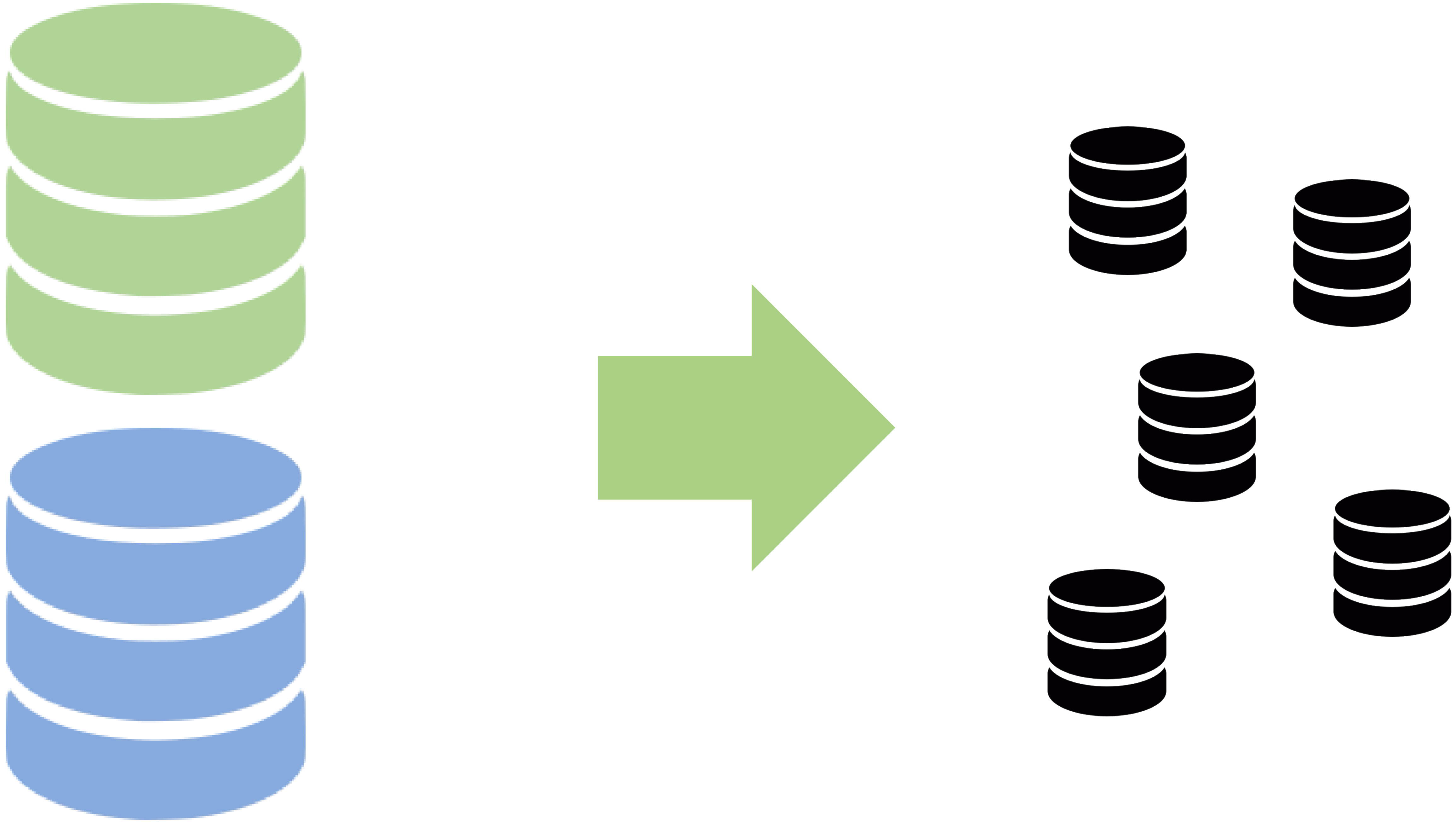
On Good Practices for Task-Specific Distillation of Large Pretrained Visual Models

Juliette Marrie, Michael Arbel, Julien Mairal, Diane Larlus

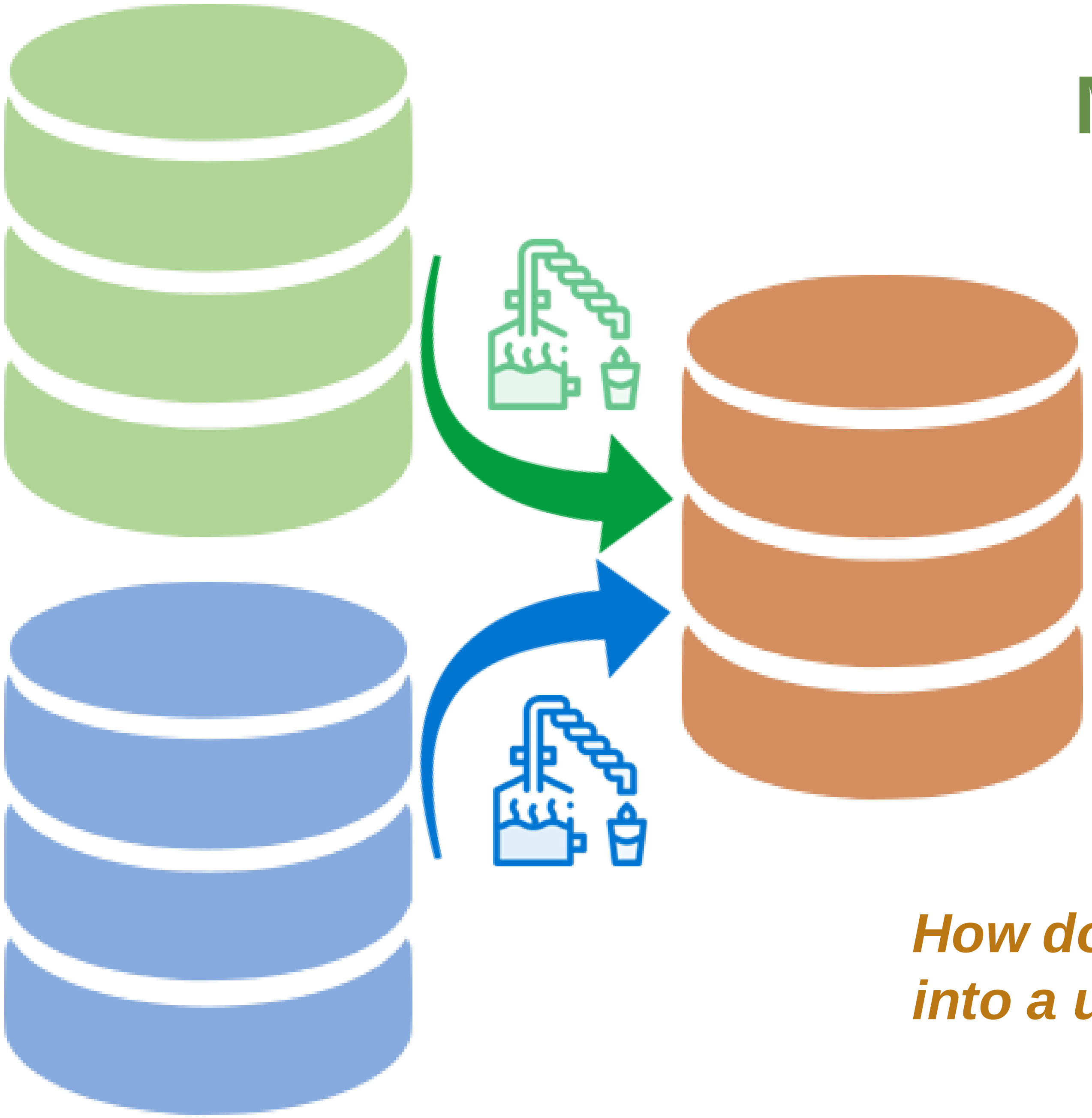
TMLR 2024

What if there are several complementary pretrained models to start from?





Multi-teacher distillation

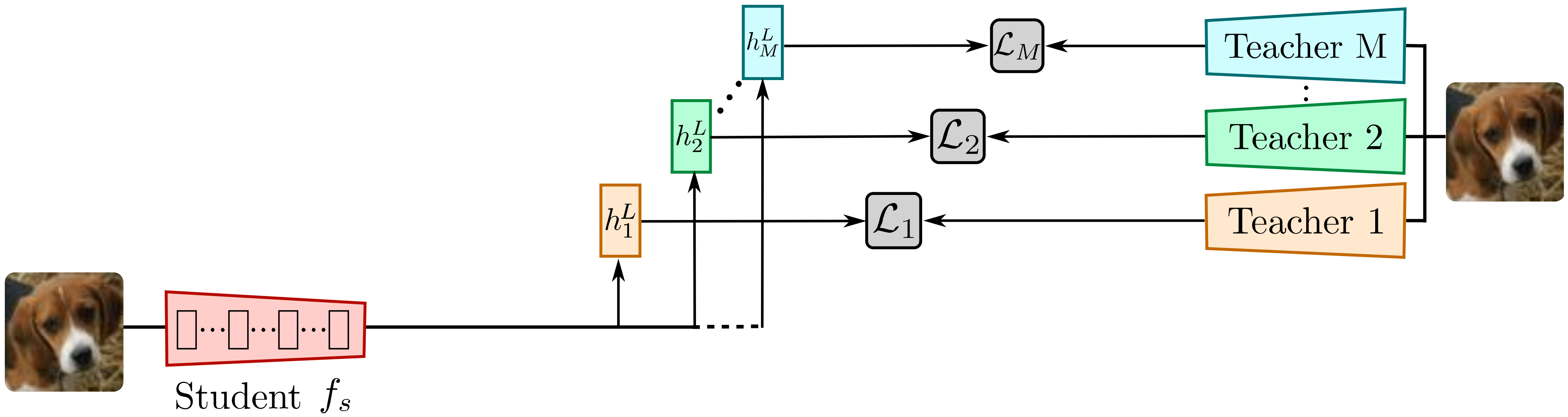


How do we merge models
into a unified pretrained model?

Multi-teacher distillation

A basic setup

- Sum across teacher losses
- Teacher-specific expendable projectors

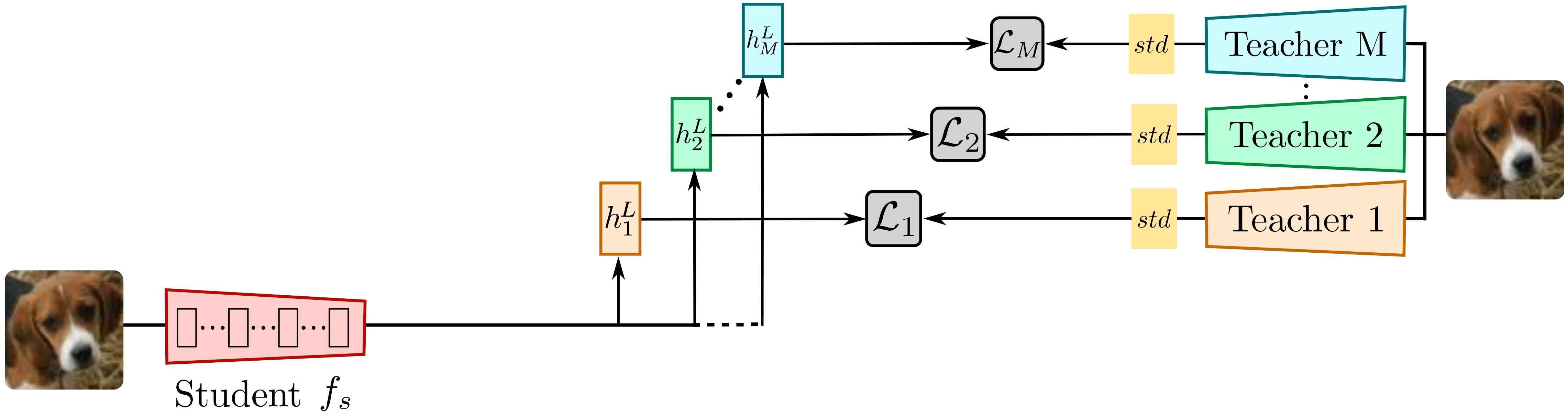


AM-RADIO [Ranzinger@CVPR24]
UNIC [Sariyildiz@ECCV24]

Multi-teacher distillation

A basic setup
- Sum across teacher losses
- Teacher-specific expendable projectors

Improvements
- Feature standardization across teachers



AM-RADIO [Ranzinger@CVPR24]
UNIC [Sariyildiz@ECCV24]

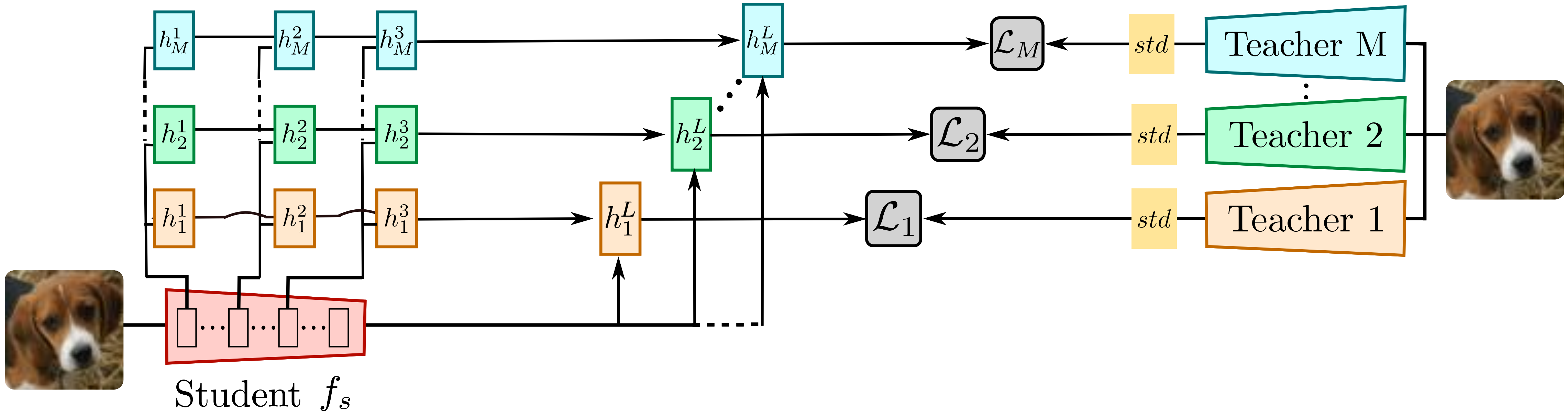
A basic setup

- Sum across teacher losses
- Teacher-specific expendable projectors

Improvements

- Feature standardization across teachers
- **Ladder of projectors**: get input from intermediate layers

Multi-teacher distillation



UNIC [Sariyildiz@ECCV24]

Multi-teacher distillation

A basic setup

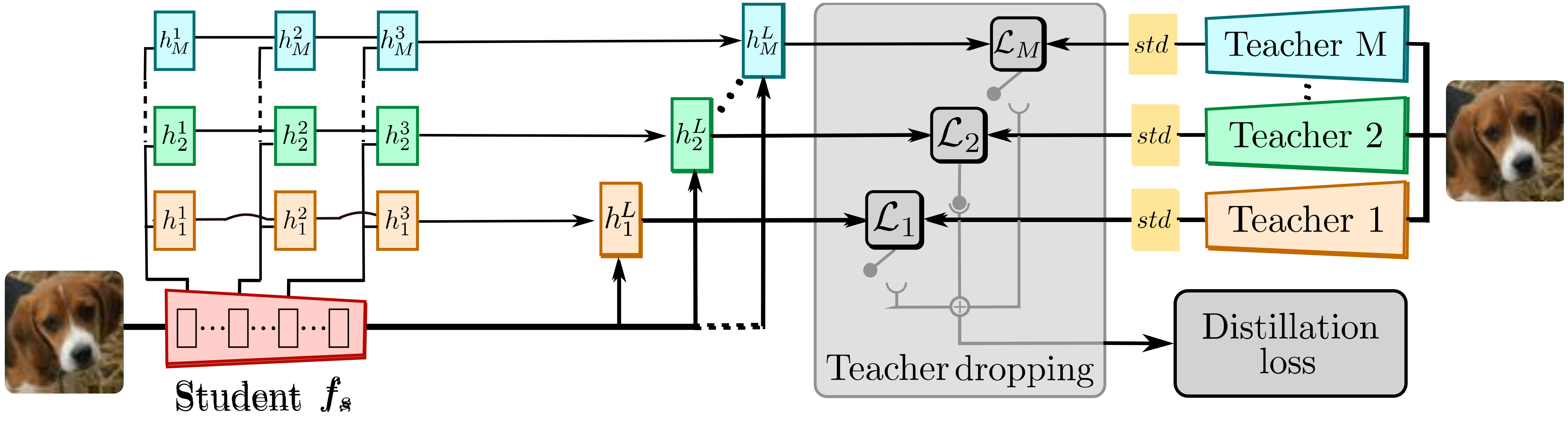
- Sum across teacher losses
- Teacher-specific expendable projectors

Improvements

- Feature standardization across teachers
- **Ladder of projectors**: get input from intermediate layers
- Loss-based **teacher dropping**

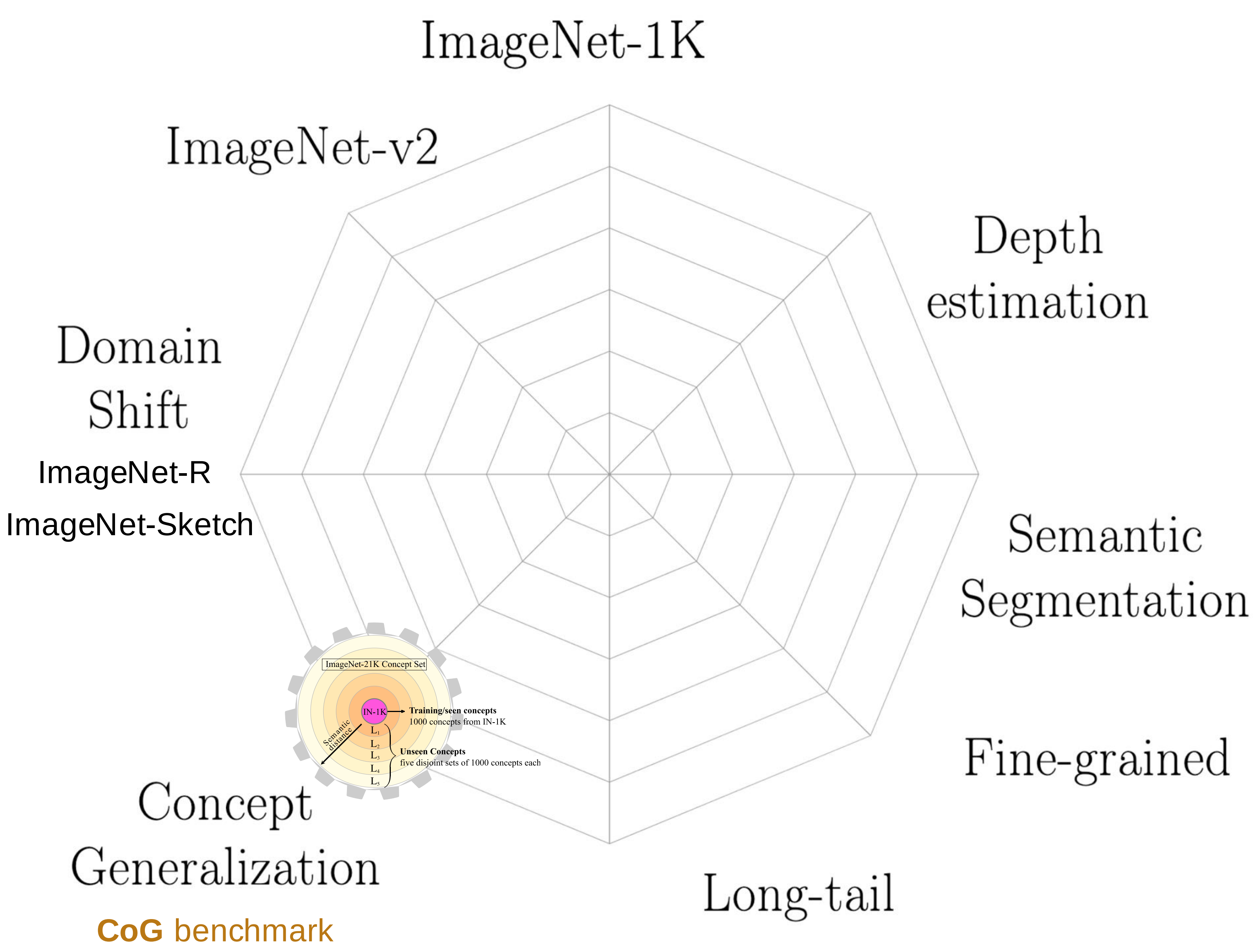
UNIC

A **UNI**versal model for **C**lassification



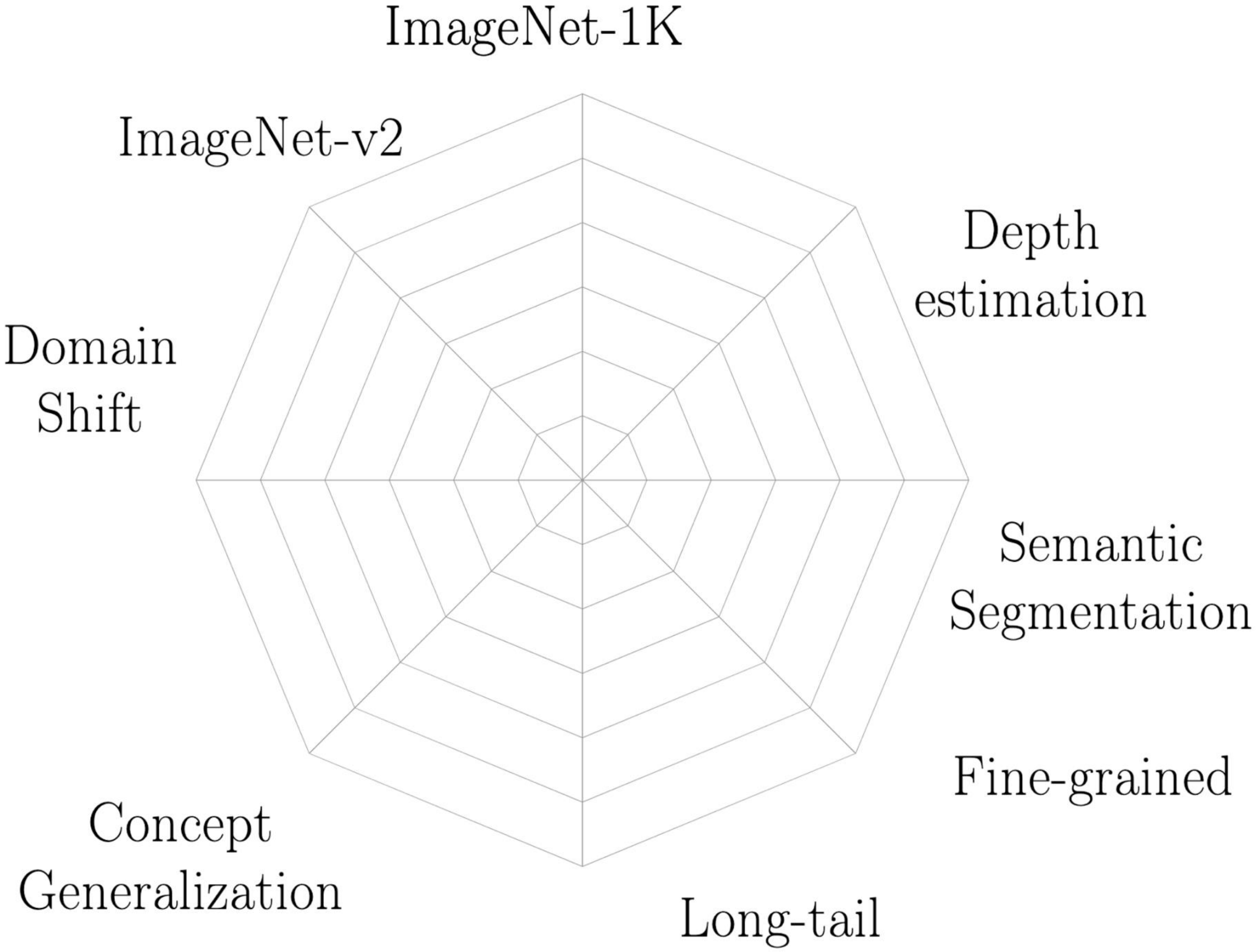
UNIC [Sariyildiz@ECCV24]

Experiments



UNIC [Sariyildiz@ECCV24]

Experiments



4 Teachers

- DINO [Caron@ICCV21]
- iBoT [Shou@ICLR22]
- DeiT-III [Touvron@ECCV22]
- dBoT-ft [Liu@ICLR22]

Setup

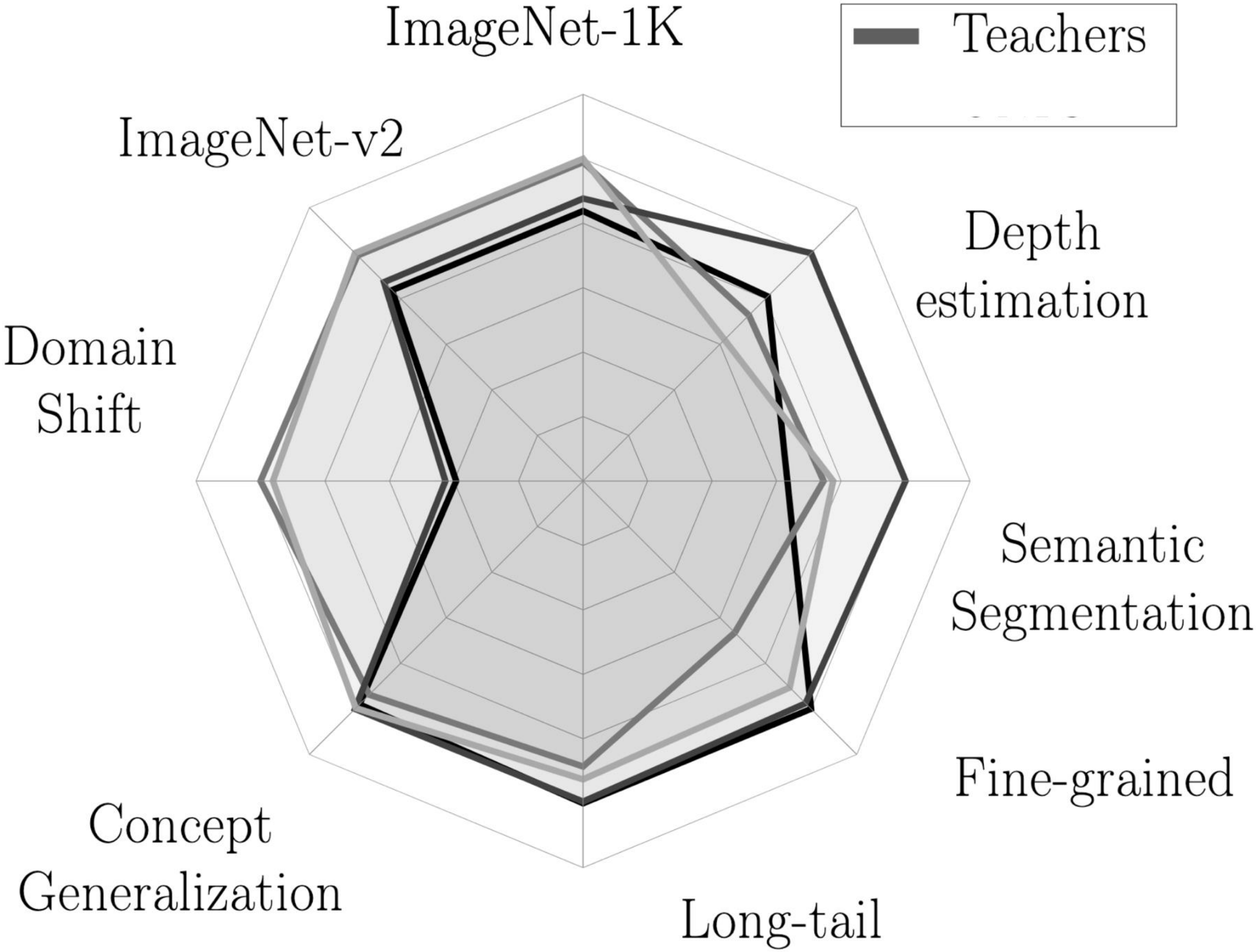
- ImageNet-1K
- ViT-Base + linear probing

UNIC [Sariyildiz@ECCV24]

Experiments

4 Teachers

- DINO [Caron@ICCV21]
- iBoT [Shou@ICLR22]
- DeiT-III [Touvron@ECCV22]
- dBoT-ft [Liu@ICLR22]

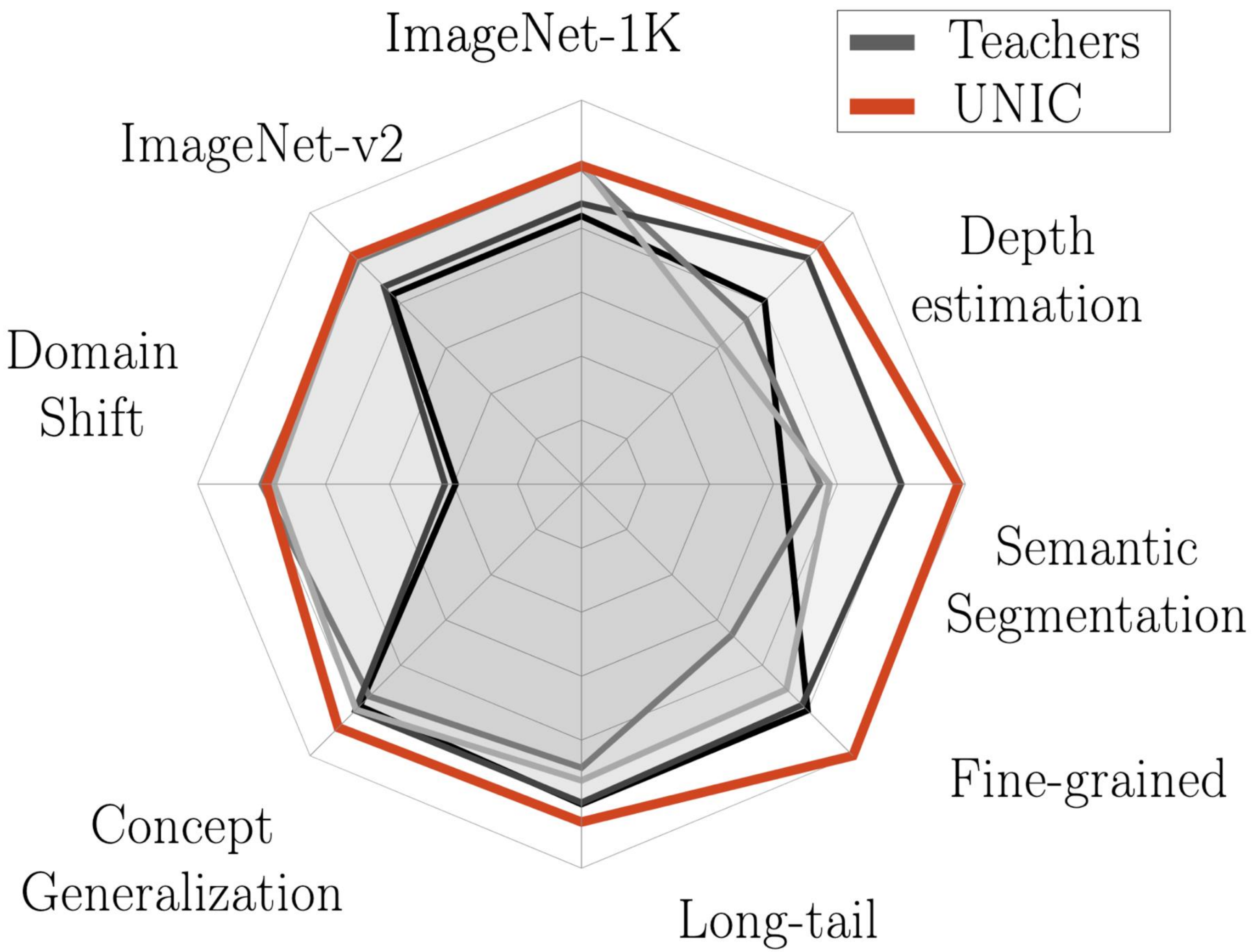


UNIC [Sariyildiz@ECCV24]

Experiments

4 Teachers

- DINO [Caron@ICCV21]
- iBoT [Shou@ICLR22]
- DeiT-III [Touvron@ECCV22]
- dBoT-ft [Liu@ICLR22]

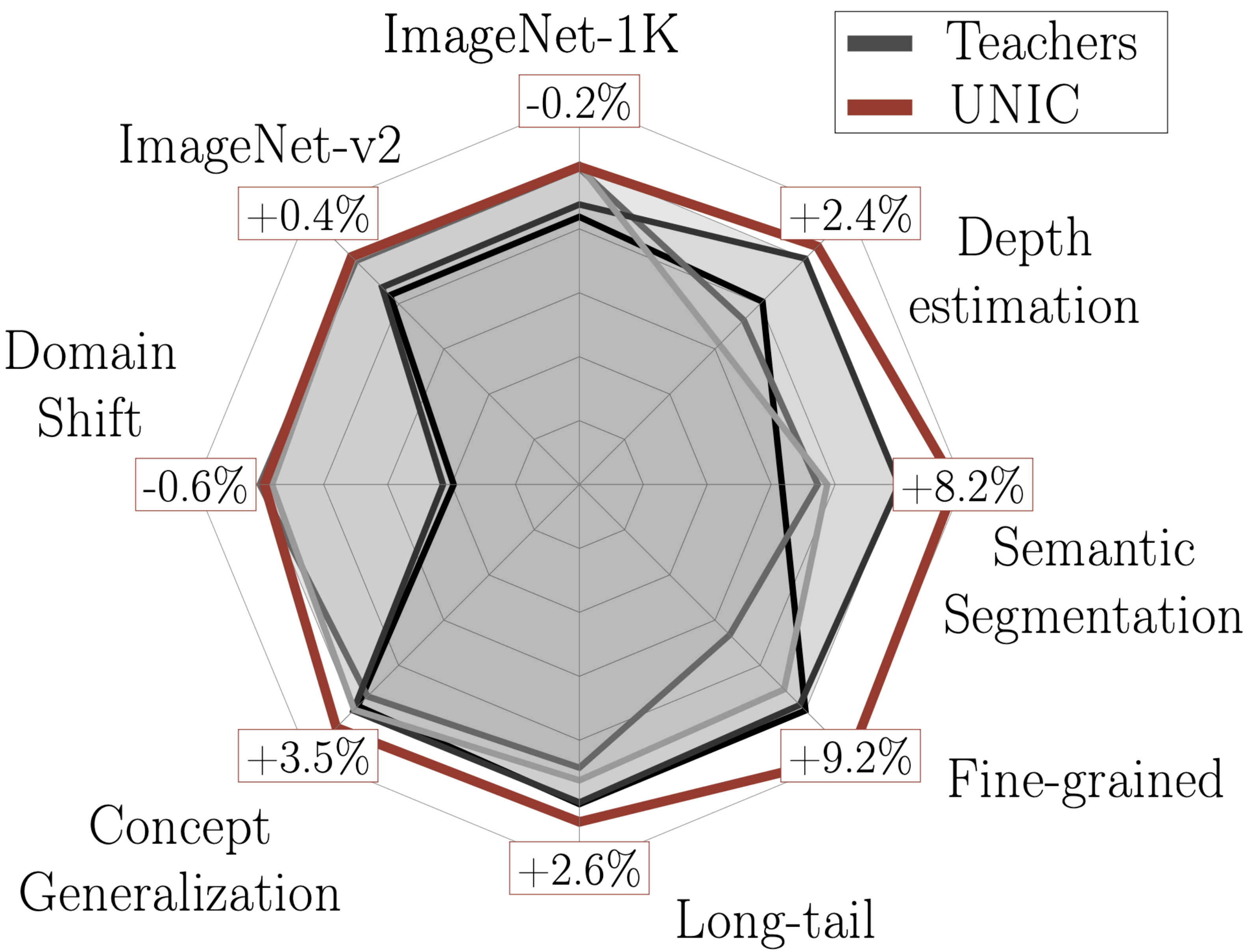


UNIC [Sariyildiz@ECCV24]

Experiments

4 Teachers

- DINO [Caron@ICCV21]
- iBoT [Shou@ICLR22]
- DeiT-III [Touvron@ECCV22]
- dBoT-ft [Liu@ICLR22]

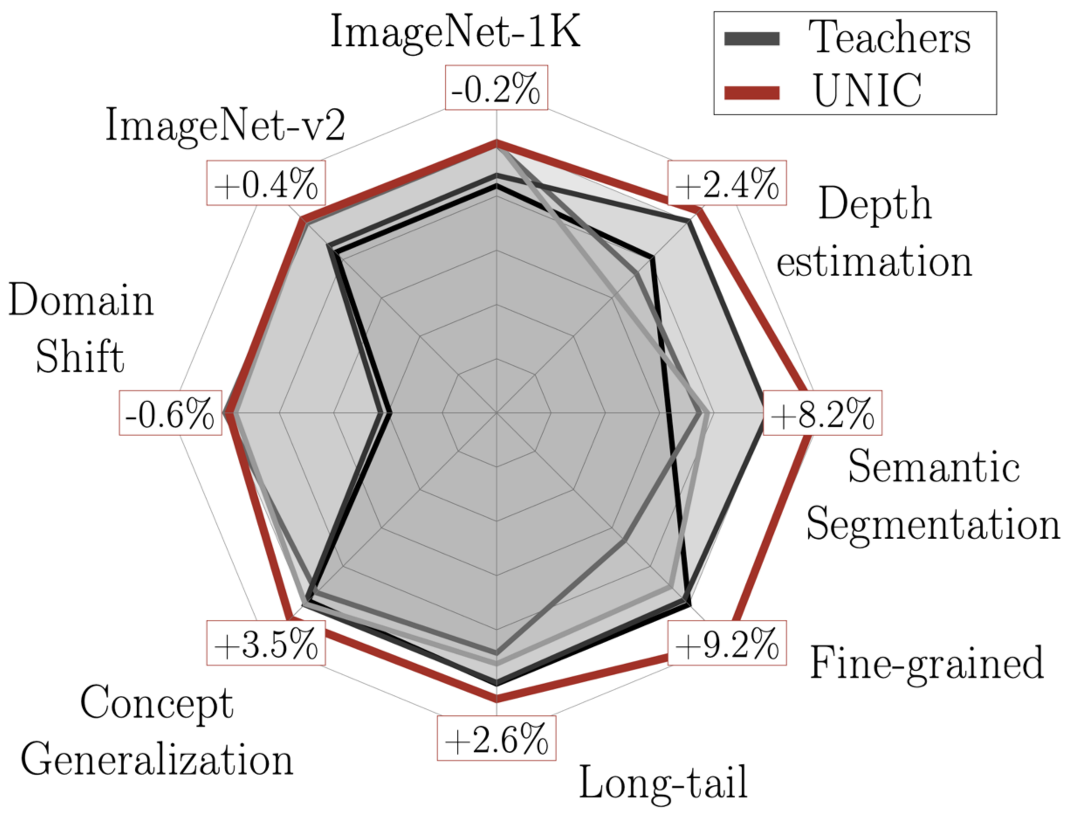


UNIC [Sariyildiz@ECCV24]

Take home message

Multi-teacher distillation

- combines models with complementary strengths



UNIC is strong at image-level classification

Reference
UNIC: Universal Classification Models via Multi-Teacher Distillation
Mert Bülent Sariyildiz, Philippe Weinzaepfel, Thomas Lucas, Diane Larlus, Yannis Kalantidis
ECCV 2024

Conclusion & References

A few ideas to bring home

Lifelong learning is extremely relevant in computer vision …

... and most likely beyond as well

Yet, it should be revisited in the light of large pretrained models

Large pretrained models

- If you would like to train one from scratch
  - Use everything you can (labels, text, etc.)
  - Beyond vision and language, more modalities could play a role
- If you would rather not
  - Mix, match, reuse existing model
  - Distillation is a powerful tool

Thanks!

Joint work with ..

Bülent Sariyildiz

Karteek Alahari

Philippe Weinzaepfel

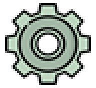
Yannis Kalantidis

Juliette Marrie

Michael Arbel

Thomas Lucas

Julien Mairal



Concept generalization in visual representation learning
Mert Bülent Sariyildiz, Yannis Kalantidis, Diane Larlus, Karteek Alahari
International Conference in Computer Vision (**ICCV**) 2021

No Reason for No Supervision: Improved Generalization in Supervised Models
Mert Bülent Sariyildiz, Yannis Kalantidis, Karteek Alahari, Diane Larlus
International Conference in Representation Learning (**ICLR**) 2023

Fake it till you make it: Learning transferable representations from synthetic ImageNet clones
Mert Bülent Sariyildiz, Karteek Alahari, Diane Larlus, Yannis Kalantidis
Conference in Computer Vision and Pattern Recognition (**CVPR**) 2023

On Good Practices for Task-Specific Distillation of Large Pretrained Visual Models
Juliette Marrie, Michael Arbel, Julien Mairal, Diane Larlus
Transactions on Machine Learning Research (**TMLR**) 2024

UNIC: Universal Classification Models via Multi-Teacher Distillation
Mert Bülent Sariyildiz, Philippe Weinzaepfel, Thomas Lucas, Diane Larlus, Yannis Kalantidis
European Conference on Computer Vision (**ECCV**) 2024

Credit icons: https://www.flaticon.com/free-icons

Thanks!

NAVER LABS
Europe