No Reason for No Supervision: Improved Generalization in Supervised Models

NAVER LABS Europe

Inria

Mert Bulent Sariyildiz
Yannis Kalantidis
Karteek Alahari
Diane Larlus

ICLR
May 2023, Kigali, Rwanda

Code and pretrained models
https://europe.naverlabs.com/t-rex

Transfer Learning

Train on **training task**

Model

Use for many **transfer tasks**


DTD


CIFAR10/100


FGVC Aircraft


Places205


MS-COCO


Pascal VOC 2007

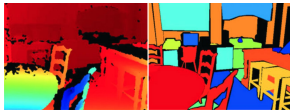

SUN397


Oxford-IIIT Pets


Birdsnap


Food-101


Stanford Cars


NYU-v2


Oxford 102 Flowers

...

Transfer Learning

Train on **training task**
- ➢ Classification (supervised)
- ➢ Self-supervised learning

Model

Use for many **transfer tasks**


DTD


CIFAR10/100


FGVC Aircraft


Places205


MS-COCO


Pascal VOC 2007

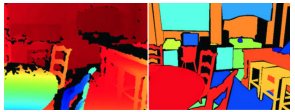

SUN397


Oxford-IIIT Pets


Birdsnap


Food-101


Stanford Cars


NYU-v2


Oxford 102 Flowers

...

Training vs Transfer performance

Transfer

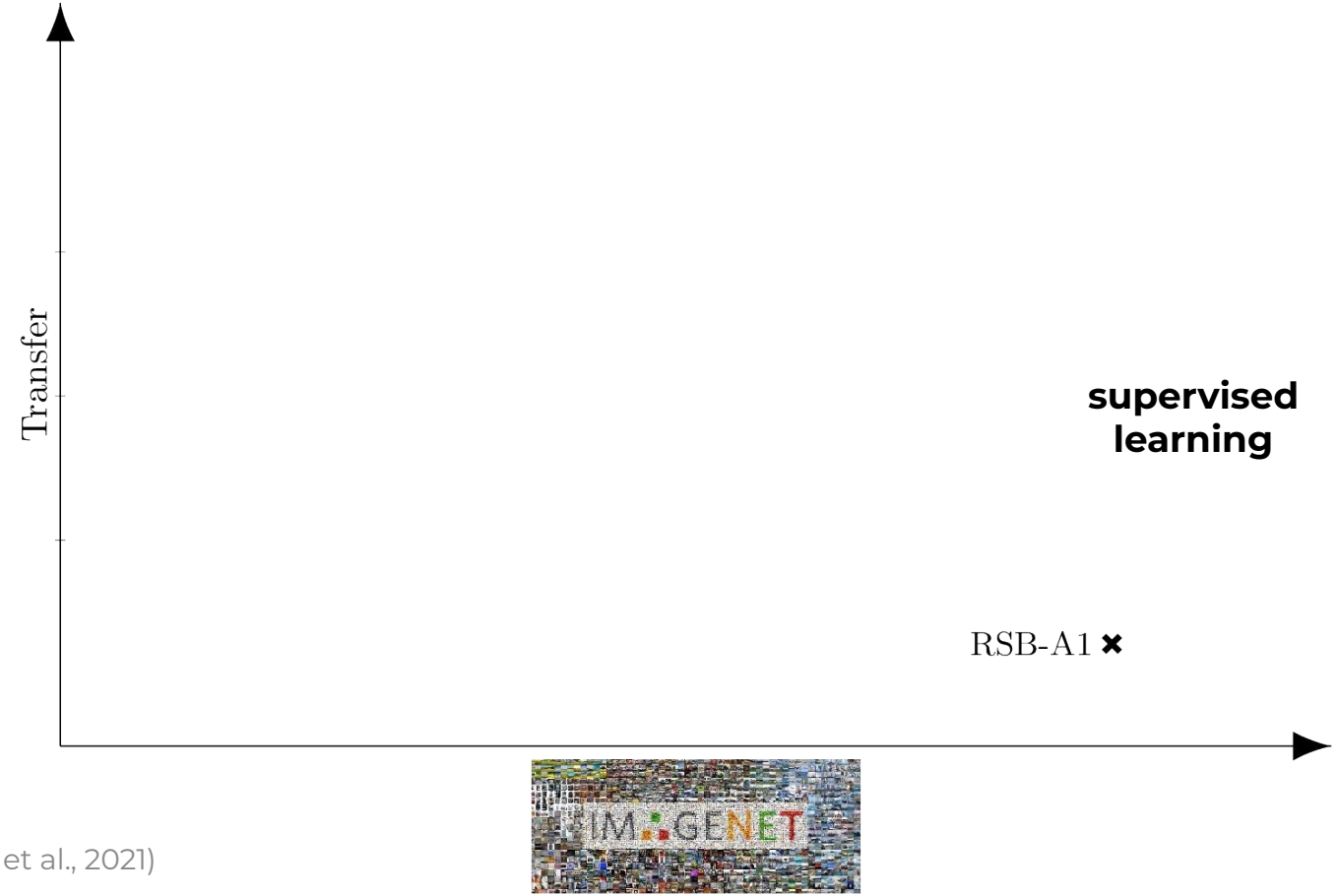
Transfer

Training

A good model should shine **both** on

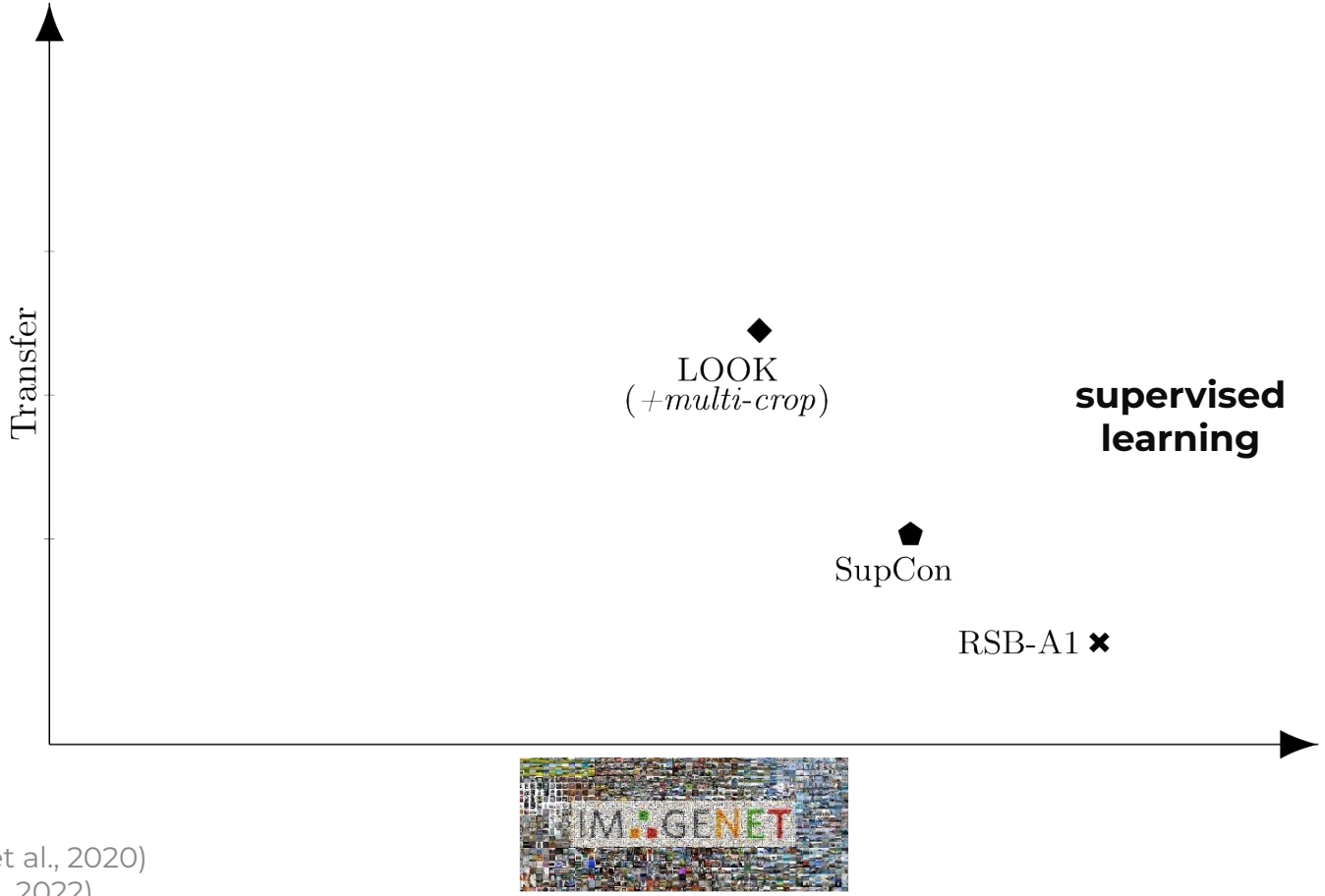
- The training task
- Transfer tasks

Training vs Transfer performance



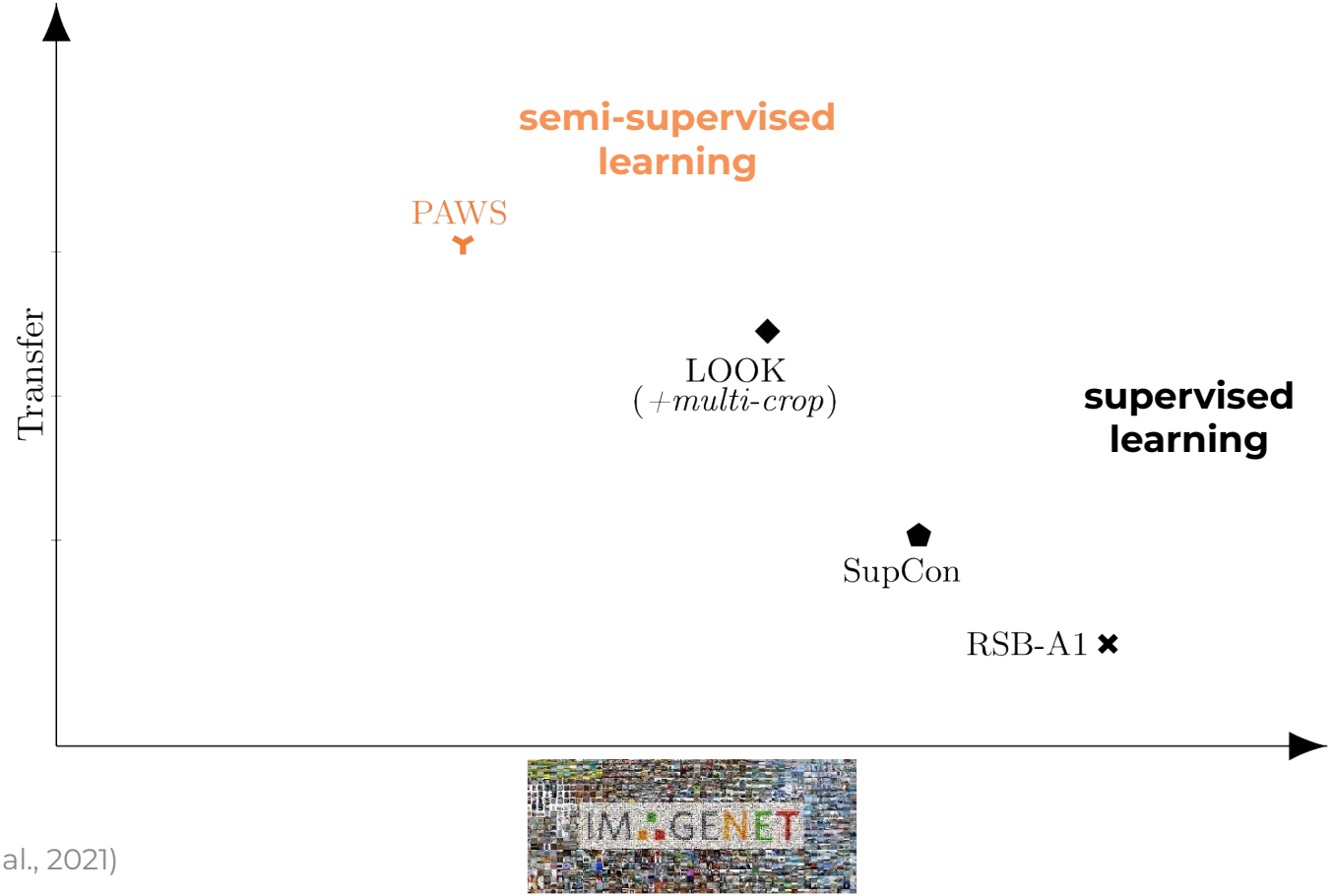
RSB (Wightman et al., 2021)

Training vs Transfer performance



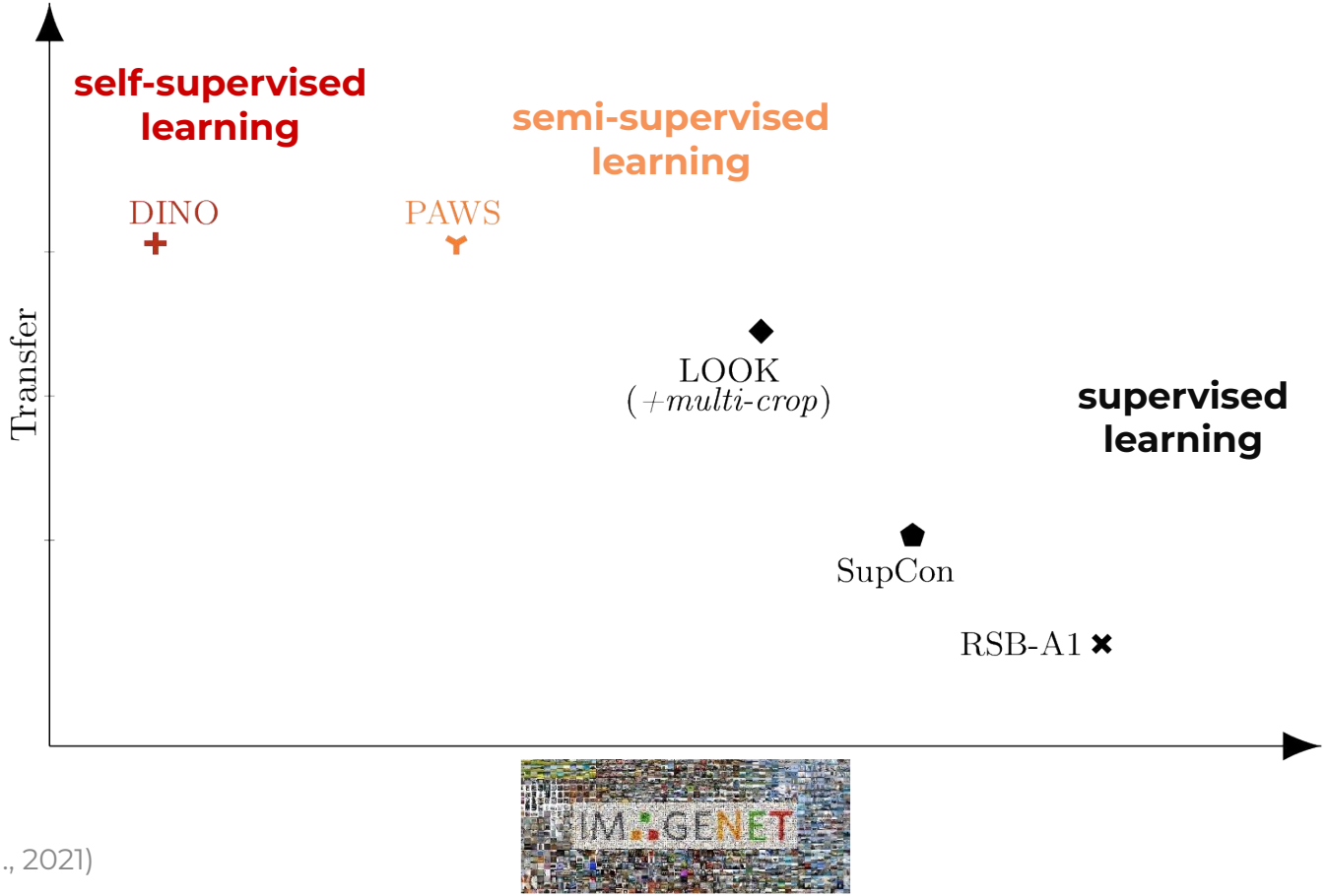
SupCon (Khosla et al., 2020)
LOOK (Feng et al., 2022)

Training vs Transfer performance



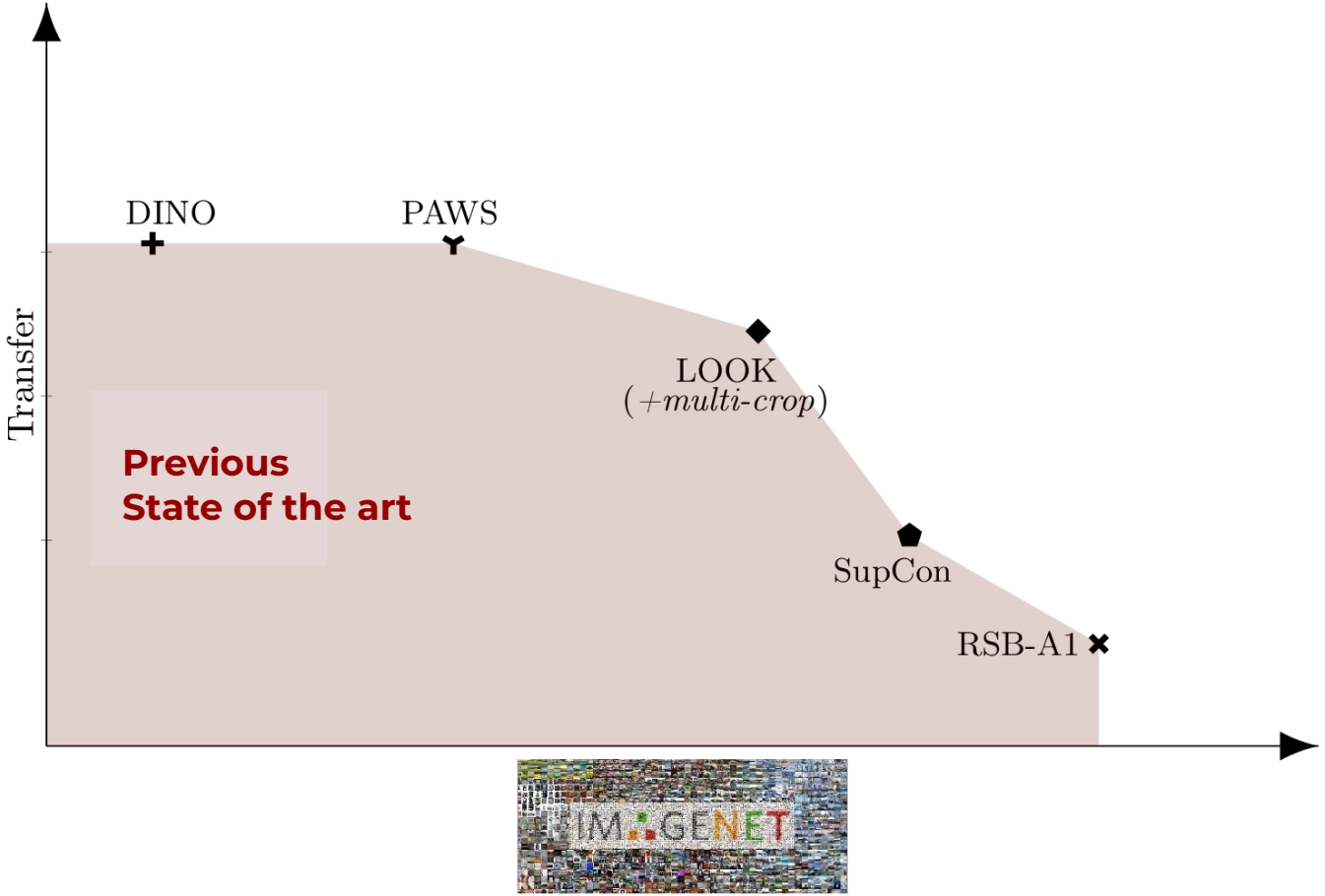
PAWS (Assran et al., 2021)

Training vs Transfer performance

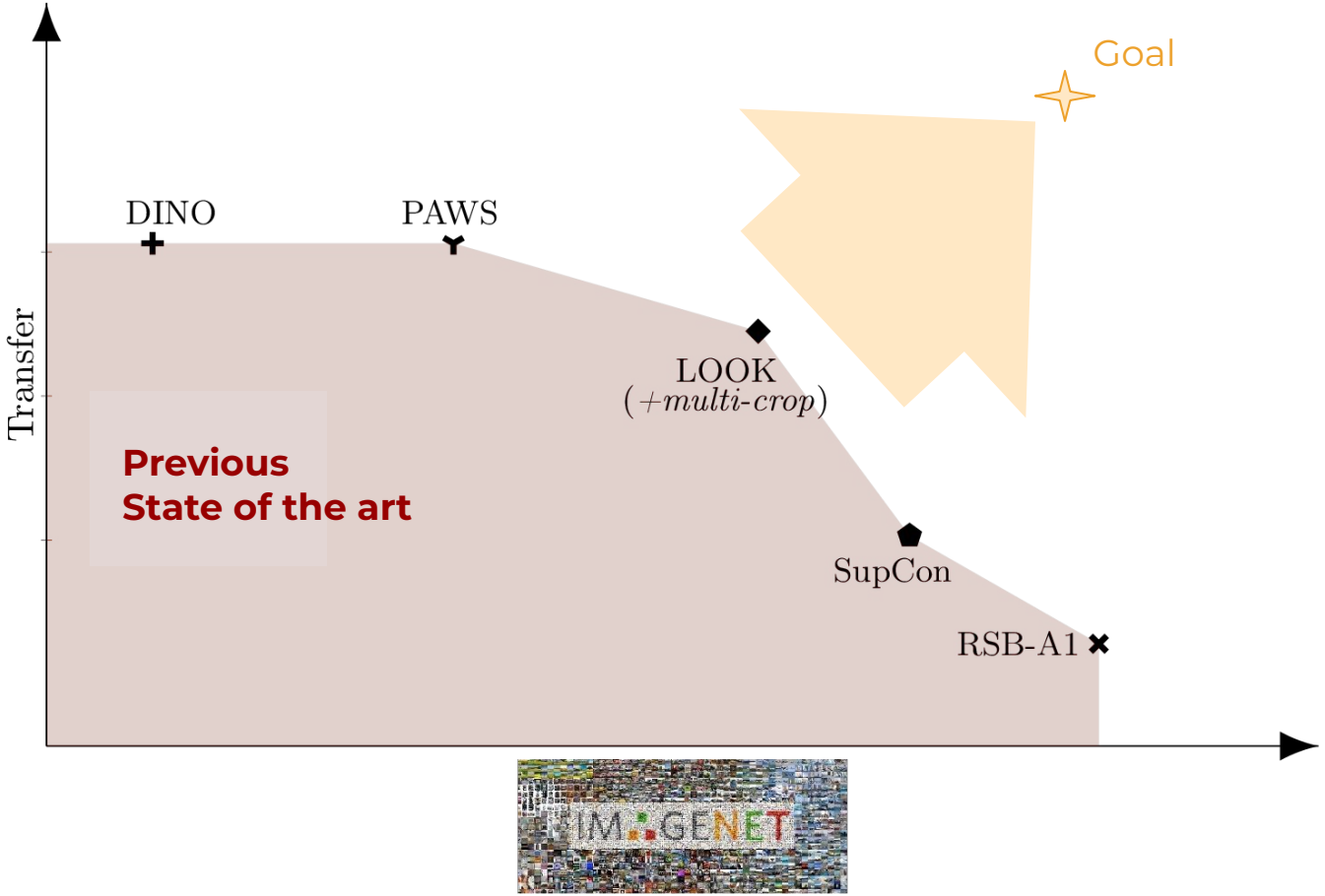


DINO (Caron et al., 2021)

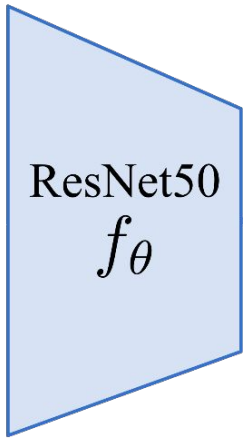
Training vs Transfer performance



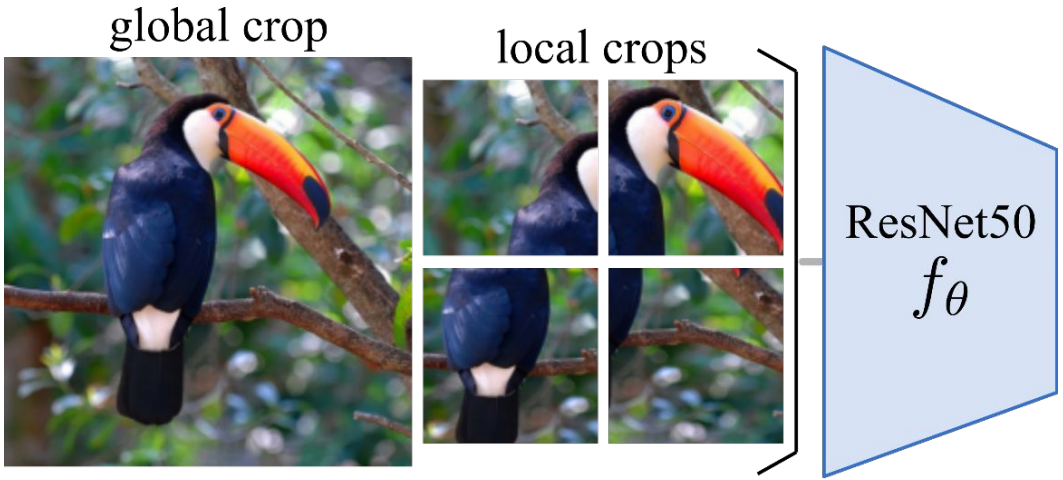
Training vs Transfer performance



An improved training setup for supervised learning

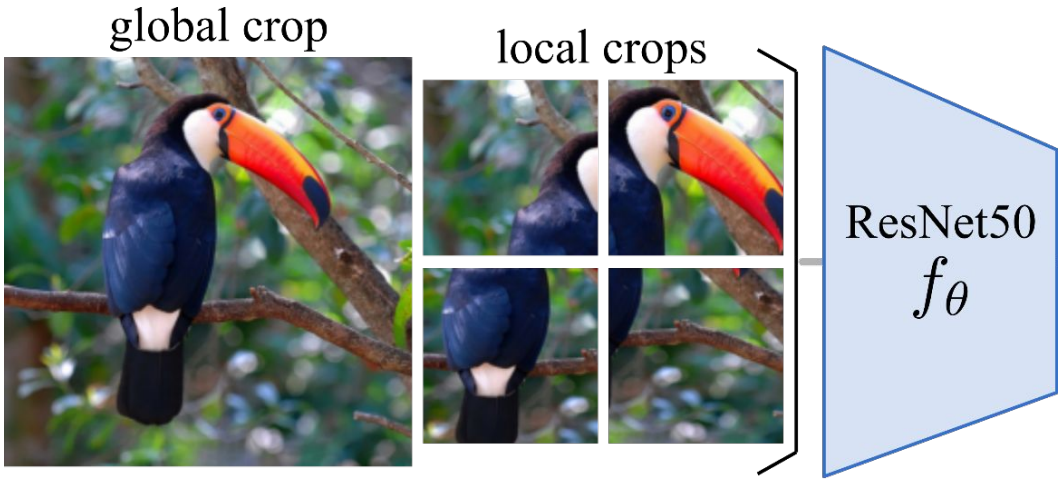


An improved training setup for supervised learning



1. Multi-crop data augmentation (Caron *et al.*, 2020)

An improved training setup for supervised learning

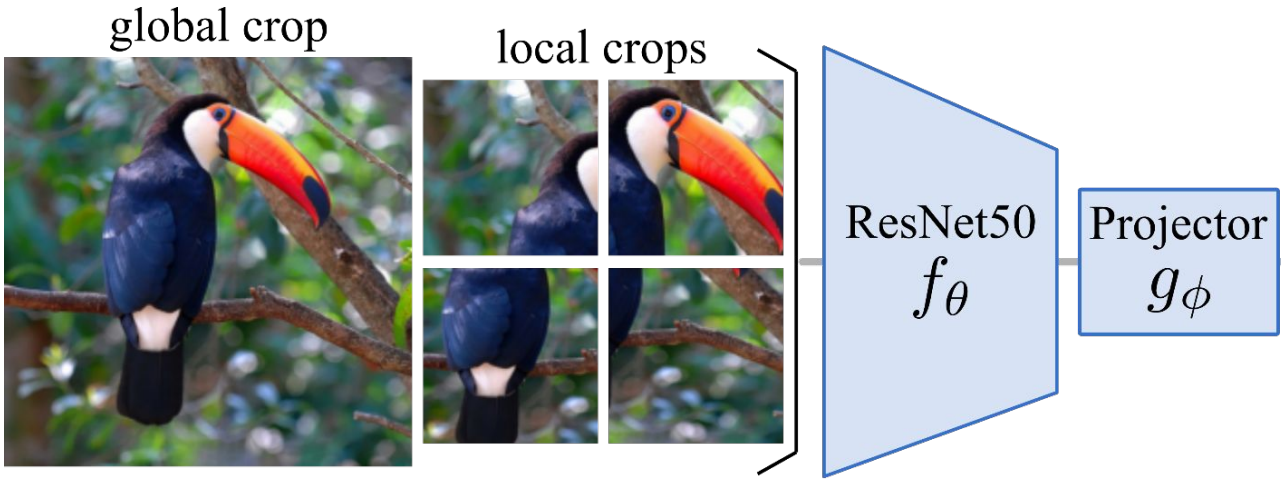


1. Multi-crop data augmentation

Multiple random crops per image

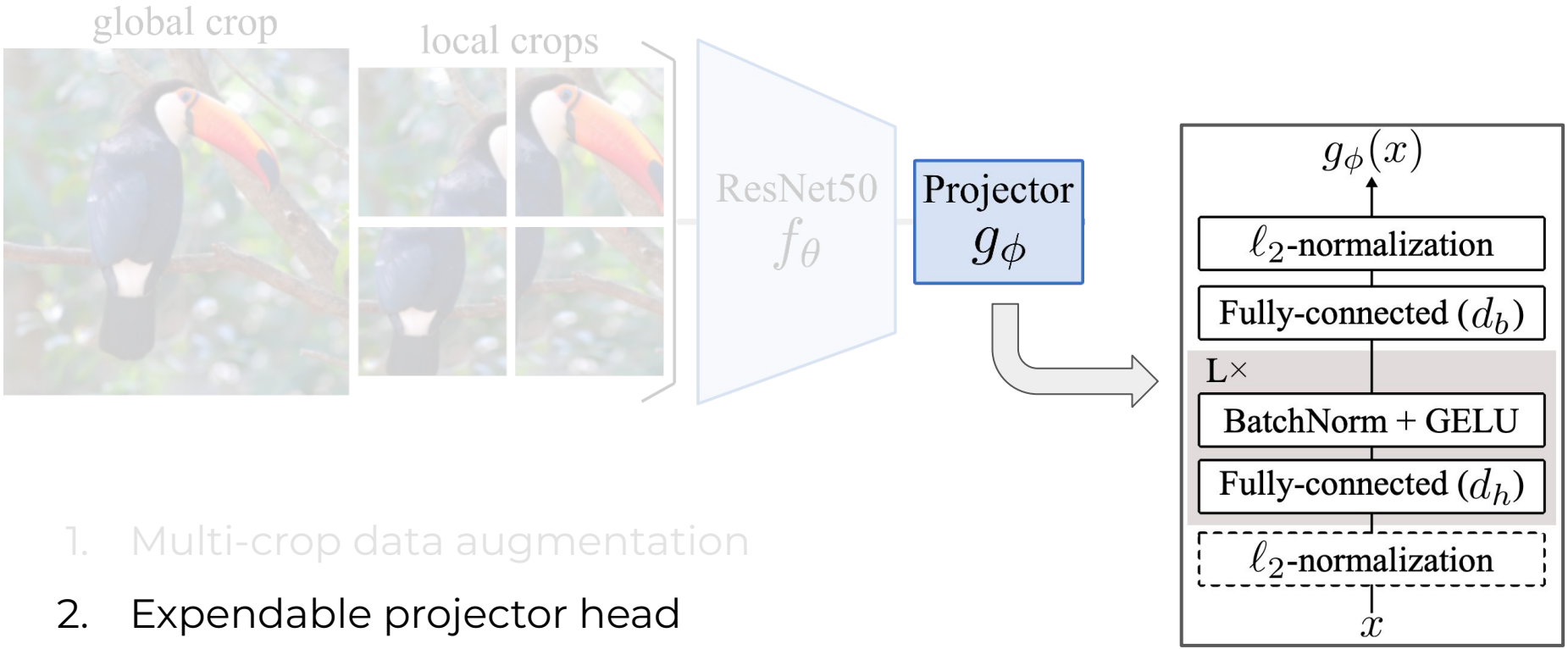
- Global crops @ 224 x 224
- Local crops @ 96 x 96

An improved training setup for supervised learning



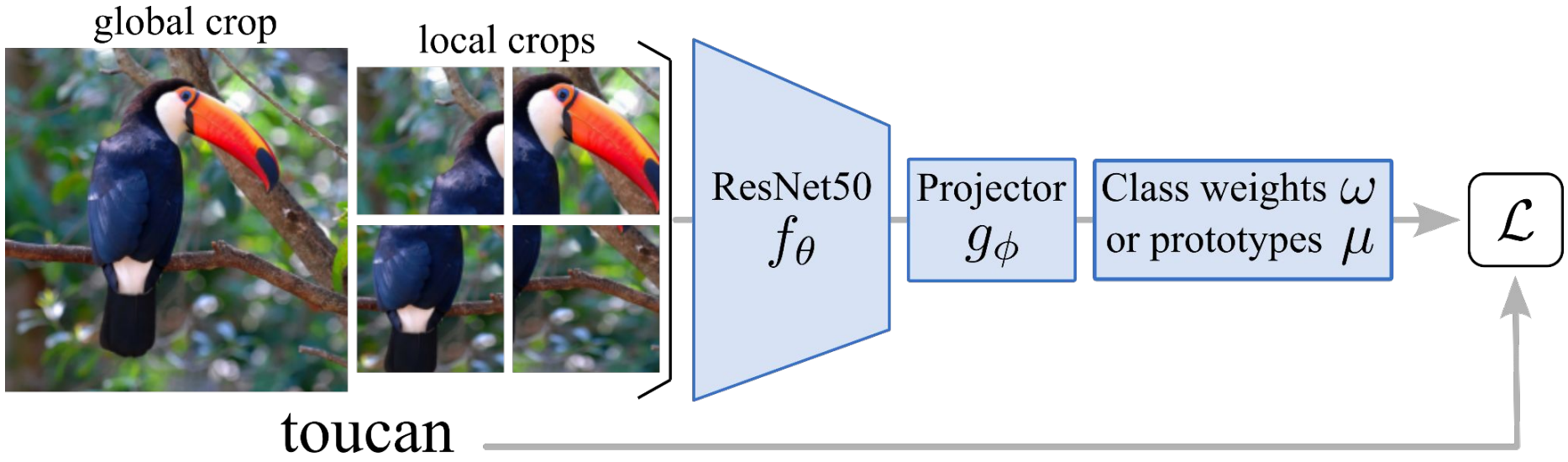
1. Multi-crop data augmentation
2. Expendable projector head (Chen *et al.*, 2020, Wang et al., 2022)

An improved training setup for supervised learning



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

An improved training setup for supervised learning



1. Multi-crop data augmentation
2. Expendable projector head
3. Cosine cross entropy loss (Kornblith *et al.*, 2021)

The loss function

Cosine cross entropy loss

$$\mathcal{L}^{*}_{\mathrm{CE}} = \sum_{c=1}^{C} y_{[c]} \log \frac{\exp(\quad x_j \quad ^{\top} \bar{\omega}_c / \tau)}{\sum_{k=1}^{C} \exp(\quad x_j \quad ^{\top} \bar{\omega}_k / \tau)}$$

The loss function

Cosine cross entropy loss

+ sum over all global & local crops

$$\mathcal{L}^{*}_{\mathrm{CE}} = -\frac{1}{M}\sum_{j=1}^{M}\sum_{c=1}^{C} y_{[c]} \log \frac{\exp(\quad x_j \ ^{\top}\bar{\omega}_c/\tau)}{\sum_{k=1}^{C}\exp(\quad x_j \ ^{\top}\bar{\omega}_k/\tau)}$$

The loss function

Cosine cross entropy loss

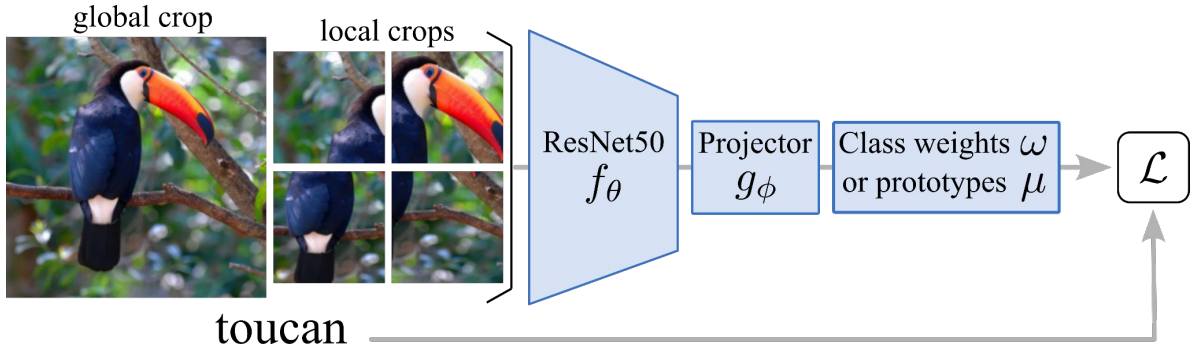
+ sum over all global & local crops
+ projector head g_ϕ

$$\mathcal{L}^*_{\mathrm{CE}} = -\frac{1}{M}\sum_{j=1}^{M}\sum_{c=1}^{C} y_{[c]} \log \frac{\exp(g_\phi(x_j)^\top \bar{\omega}_c/\tau)}{\sum_{k=1}^{C}\exp(g_\phi(x_j)^\top \bar{\omega}_k/\tau)}$$

An improved training setup for supervised learning

Changes over the basic supervised learning setup:

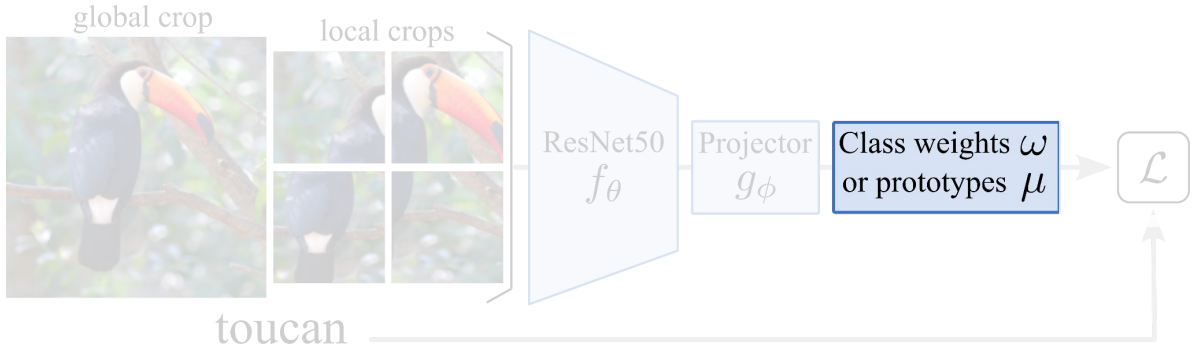
1. Multi-crop data augmentation
2. Expendable projector head
3. Cosine cross entropy loss



An improved training setup for supervised learning

Changes over the basic supervised learning setup:

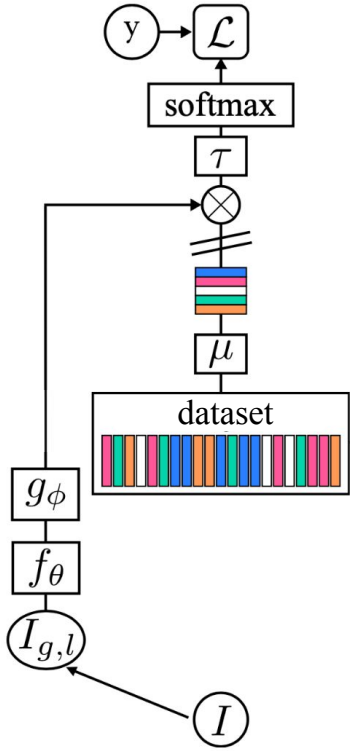
1. Multi-crop data augmentation
2. Expendable projector head
3. Cosine cross entropy loss
4. (*optional*) Replace class weights with class prototypes



Nearest Class Means (NCM)

(Mensink *et al.* 2012, Guerriero *et al.* 2018):

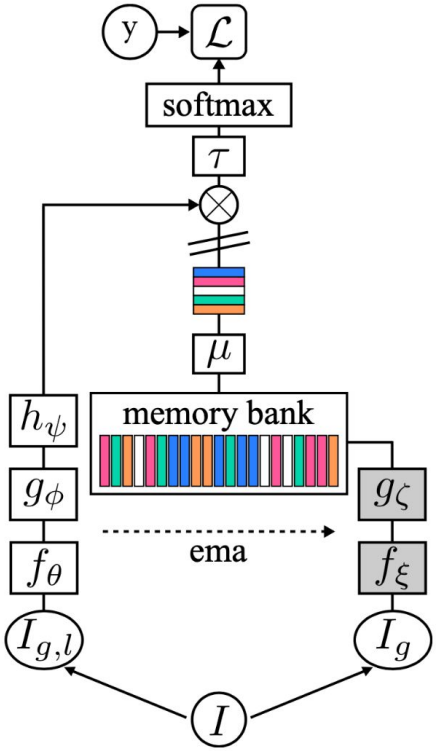
Replace the learnable **class weights**
with **class means**



Online Class Means (**OCM**)

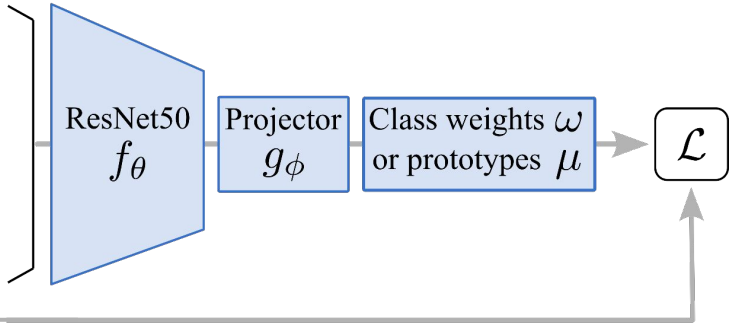
Replace the learnable **class weights**
with **class means**

computed over a **memory bank**
with features from a Momentum Encoder
(He *et al.*, 2020)



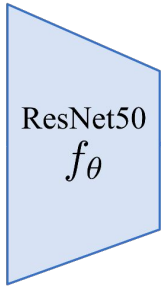
Experimental Setup

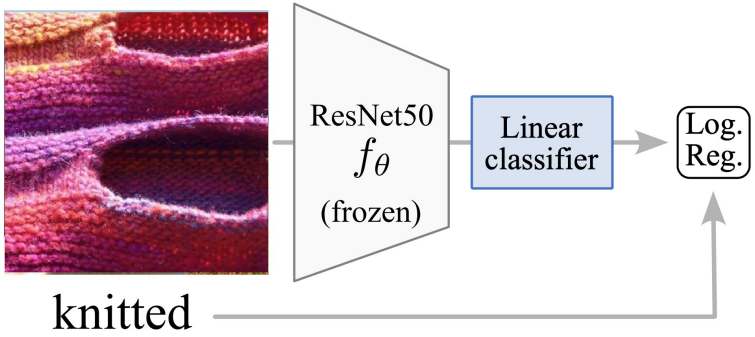
Train on ImageNet-1K



Experimental Setup

Train on ImageNet-1K





For every **transfer** task

Evaluation Protocol

Training task: results on the **ImageNet** dataset

Transfer tasks: A unifique transfer performance metric - aggregates **13 datasets**

“Log-odds” score averaged over

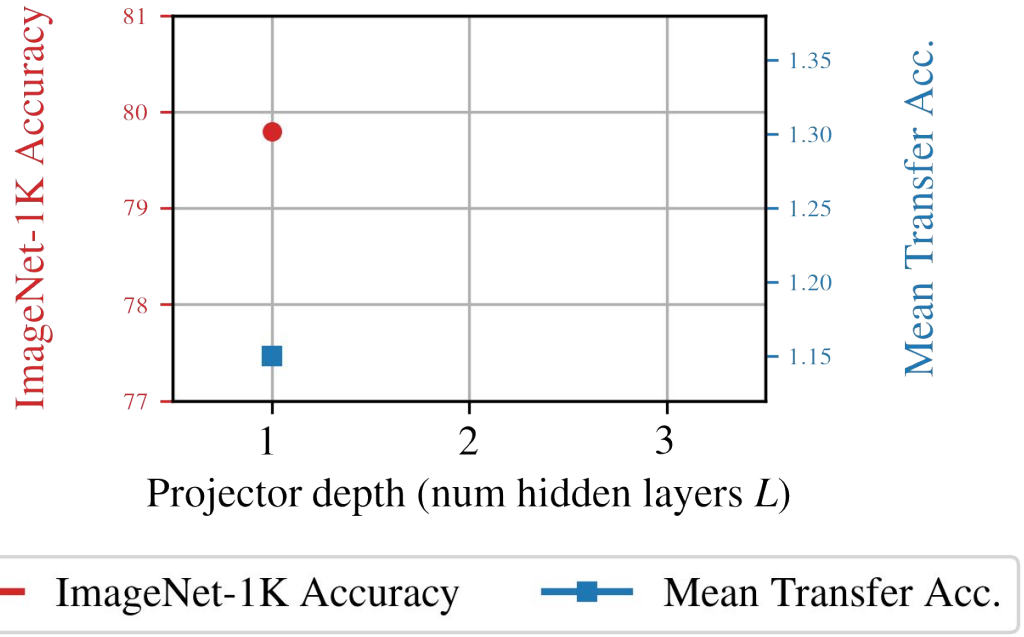
- 8 small **fine-grained classification** datasets
  Aircraft, Cars, DTD, EuroSAT, Flowers, Pets, Food101, SUN397

- 5 concept generalization benchmarks: **ImageNet-CoG** (Sariyildiz et al., 2021)
  CoG-L1, CoG-L2, CoG-L3, CoG-L4, CoG-L5

Impact of the projector

Observation 1

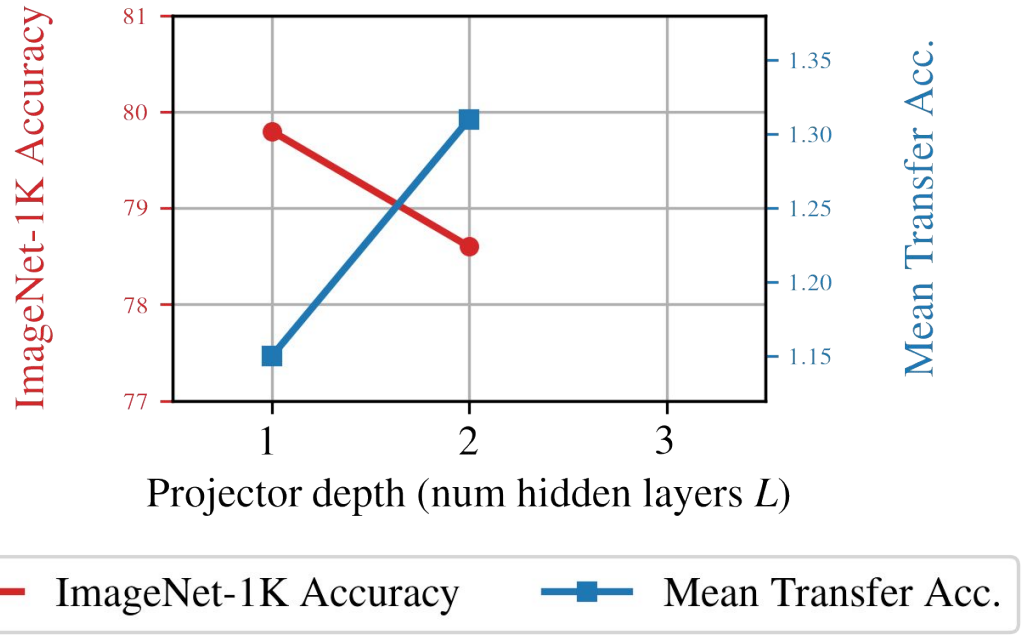
Projector depth controls the **trade-off** between Training & Transfer accuracy



Impact of the projector

Observation 1

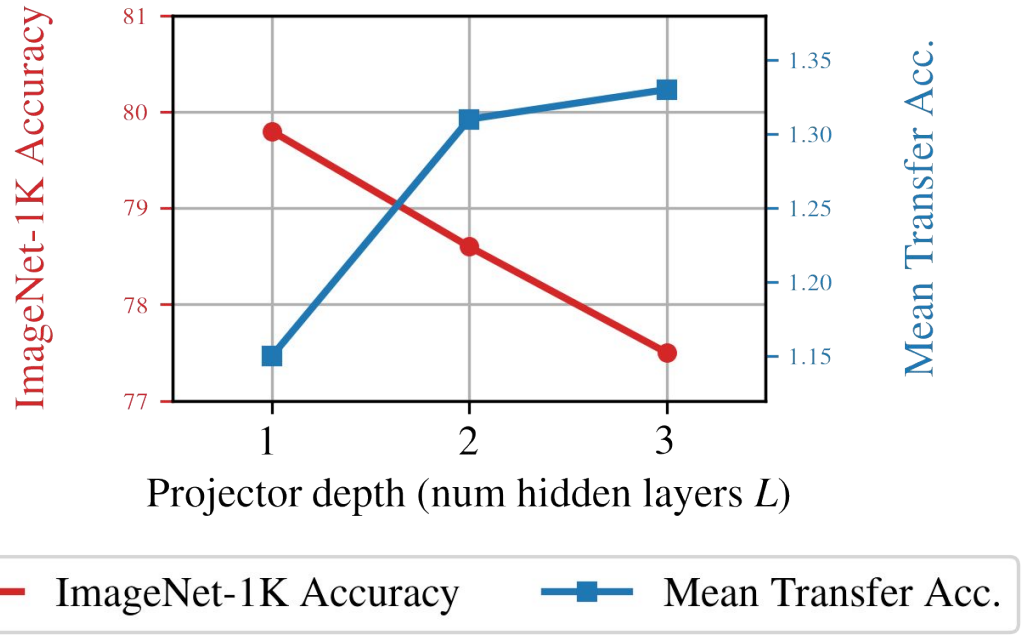
Projector depth controls the **trade-off** between Training & Transfer accuracy



Impact of the projector

Observation 1

Projector depth controls the **trade-off** between Training & Transfer accuracy



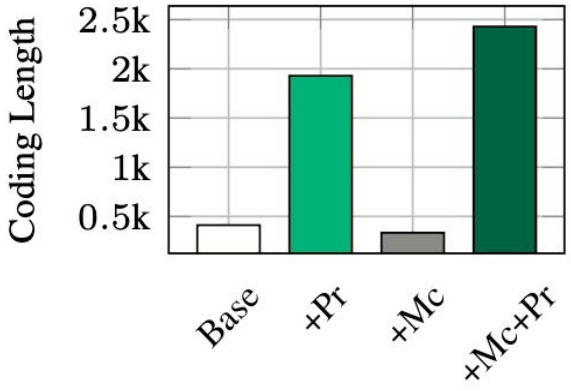
Impact of the projector

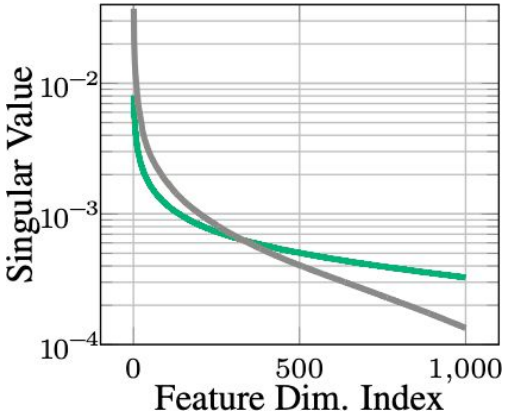
Observation 2

Projector also has an impact on the representations

Average coding length of representations and singular values

(computed over all transfer datasets)

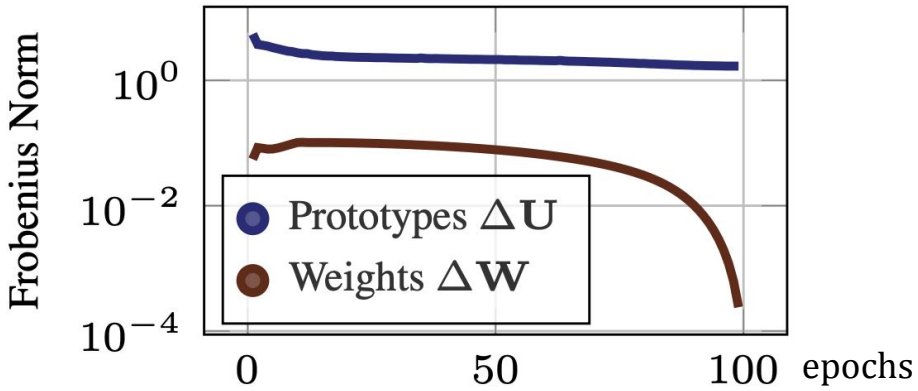




Observation 3

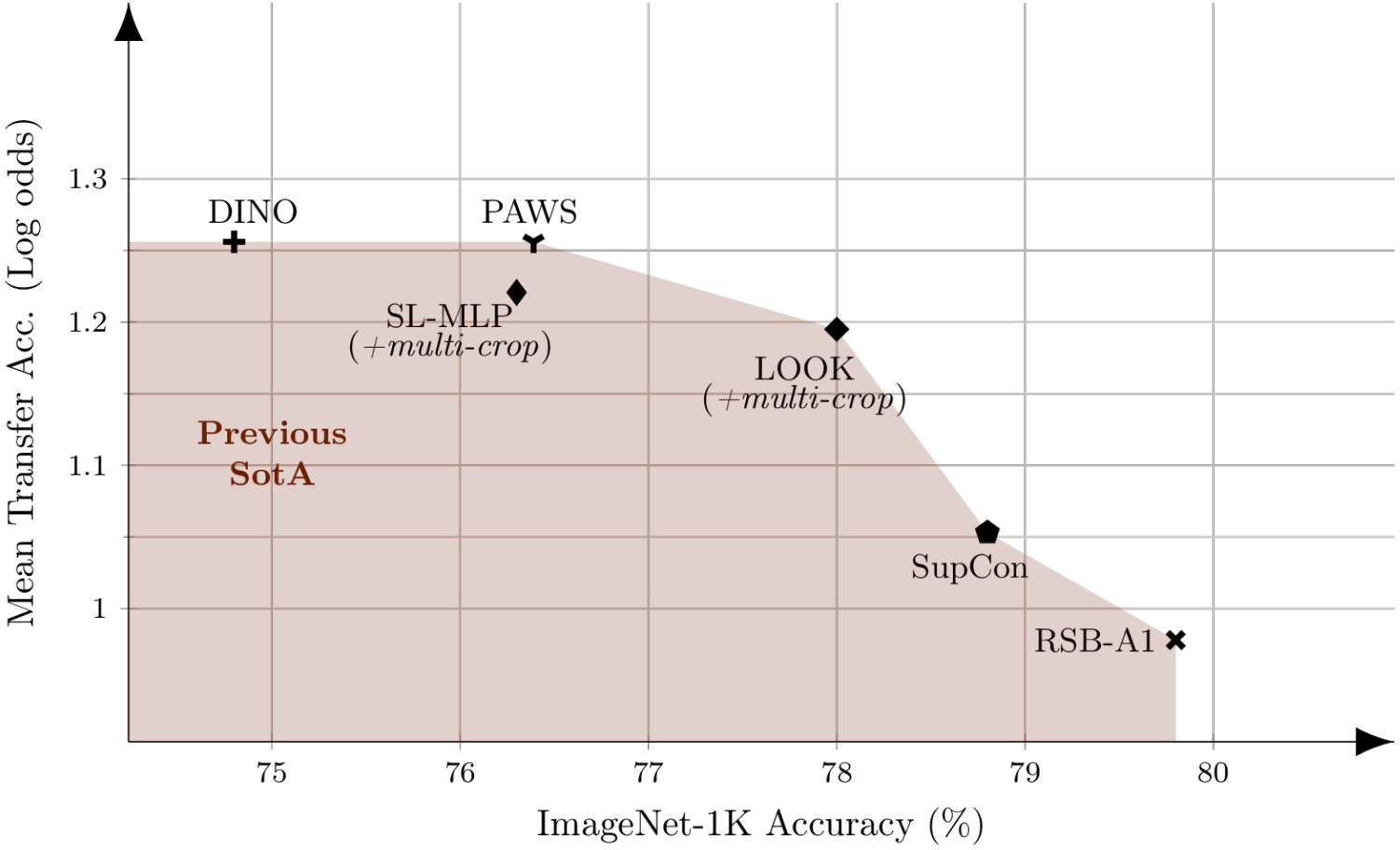
OCM decreases overfitting:

The prototypes change more than the learned weights

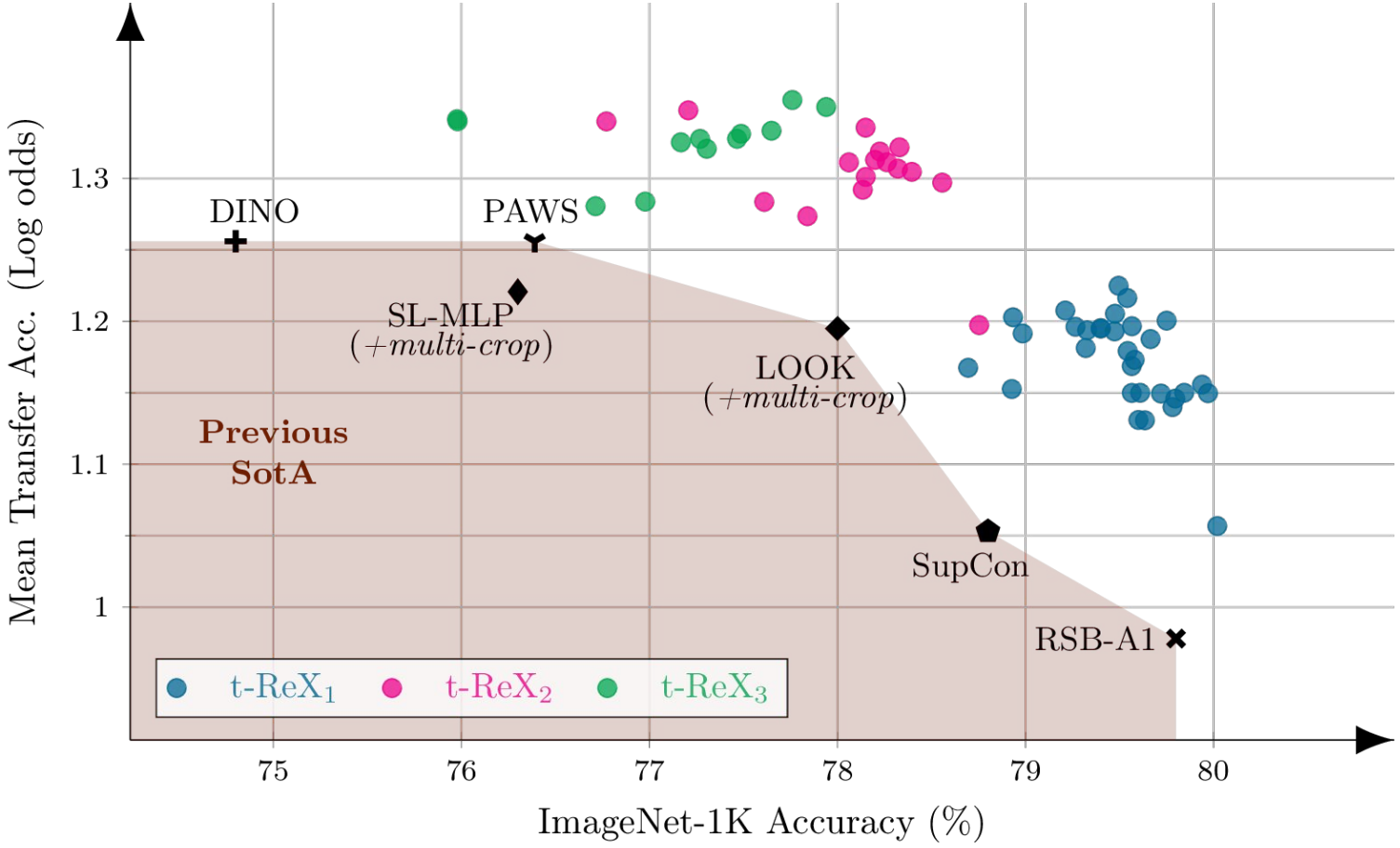


Change in class weights **W** and prototypes **U** at every iteration

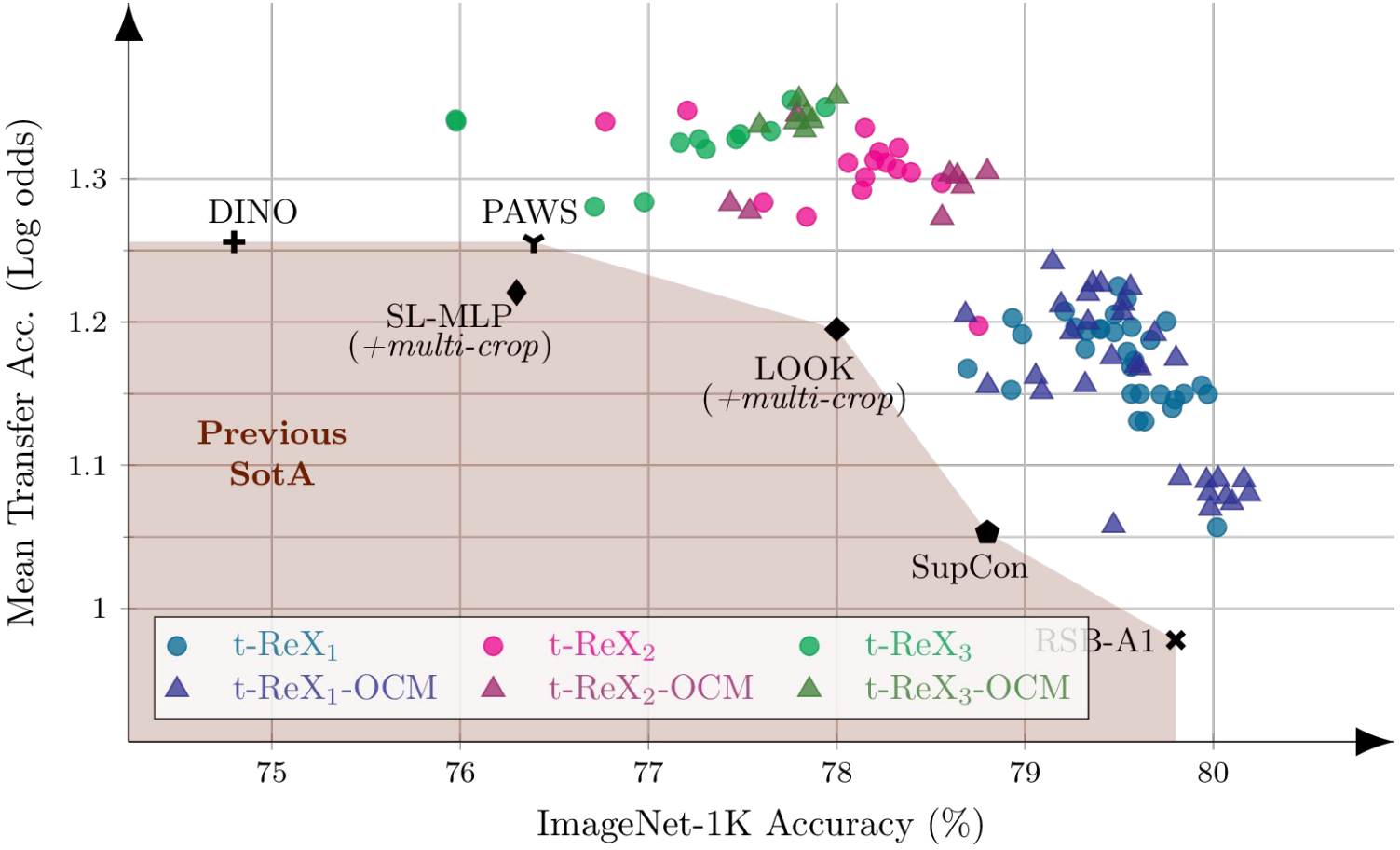
Comparison to the **state of the art**



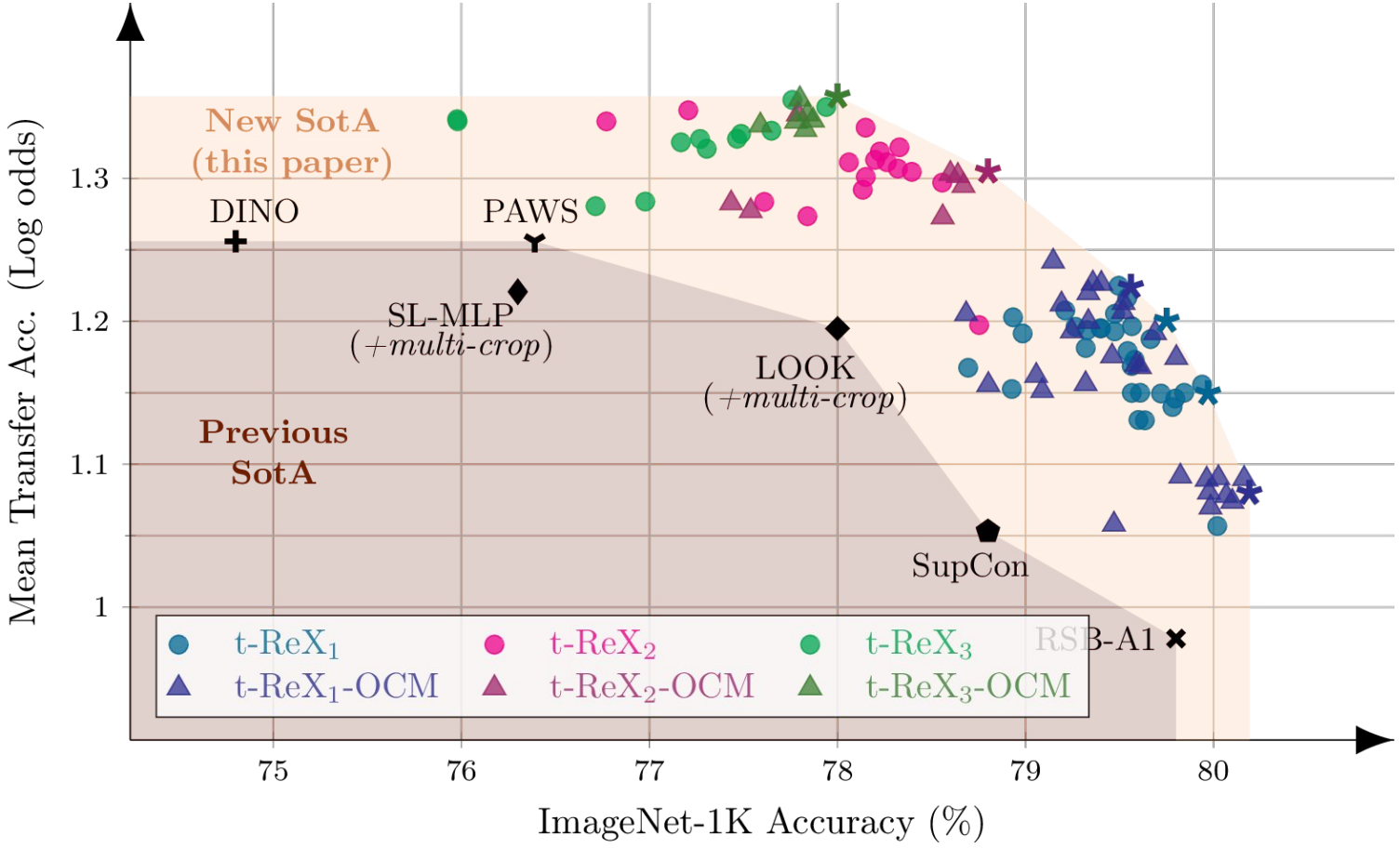
Comparison to the **state of the art**



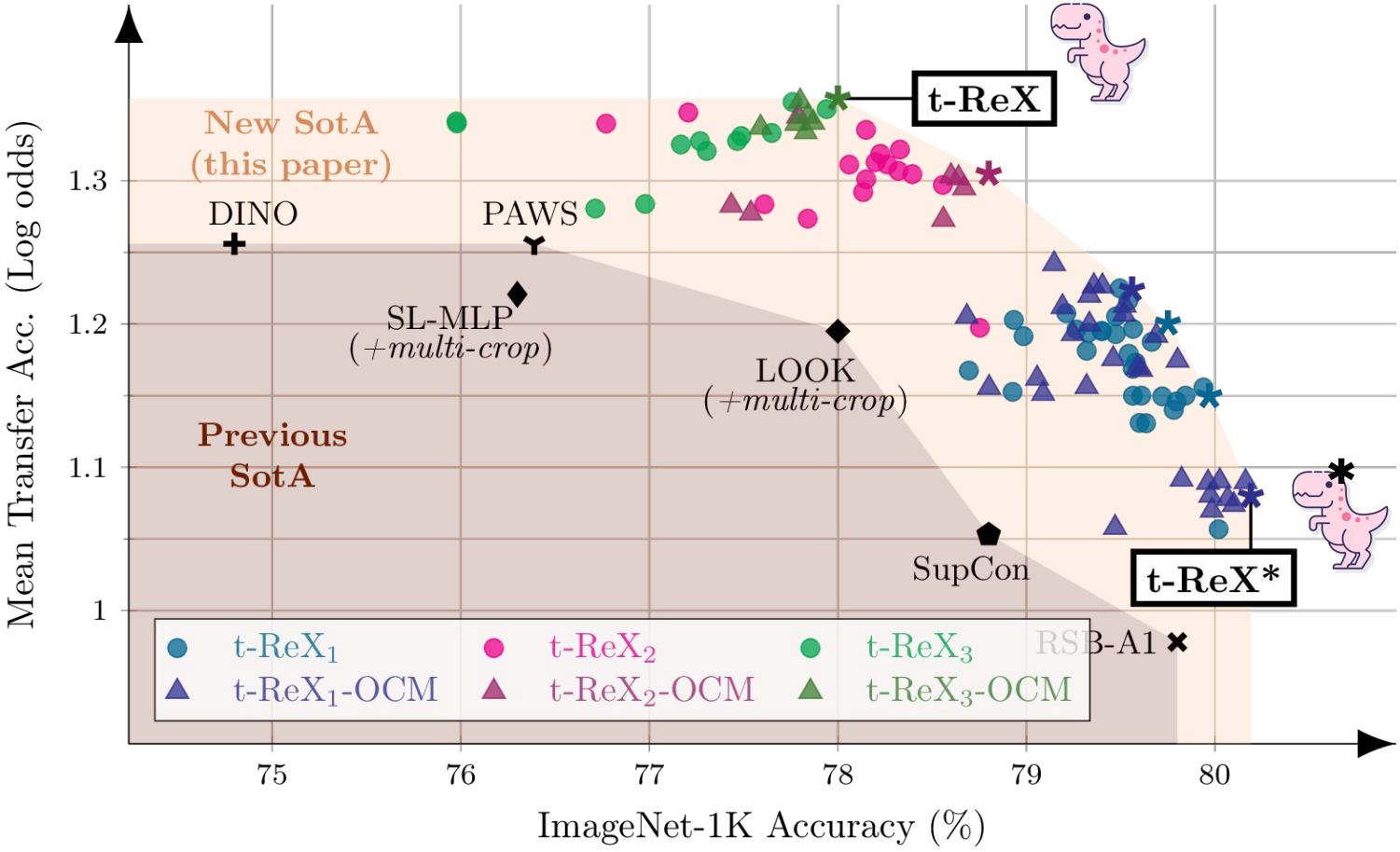
Comparison to the **state of the art**



Comparison to the **state of the art**



Comparison to the **state of the art**



Conclusions

No reason for no supervision!

- Our **t-Rex** models are **state of the art** for **transfer** 'despite' being supervised
  - Multi-crop
  - Expendable projector
  - Cosine Cross Entropy loss
  - (optional) Online Class Means
- **Training / Transfer** trade-off controlled via projector design
- **t-ReX** and **t-ReX*** ResNet50 models are available!

Code and pretrained models
https://europe.naverlabs.com/t-rex