valeo.ai

Efficient construction of training datasets for 2D & 3D data

by Oriane Siméoni

Efficient construction of training datasets

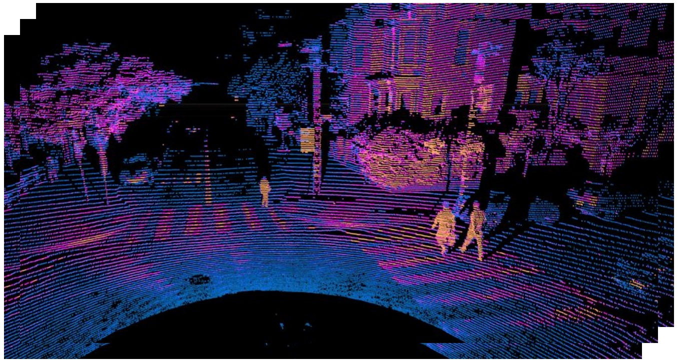
- Best perception models are trained in a **fully-supervised** fashion
- Require **large amount of annotated** data
- Data curation and manual annotation is **time-consuming** and **expensive** (eg. 35 secs per bounding box on an image)

Questions

- How to mitigate costs?
- What data to **annotate** ?

2D data

3D data

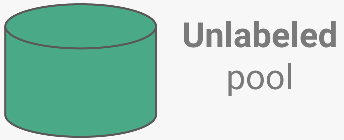


Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**

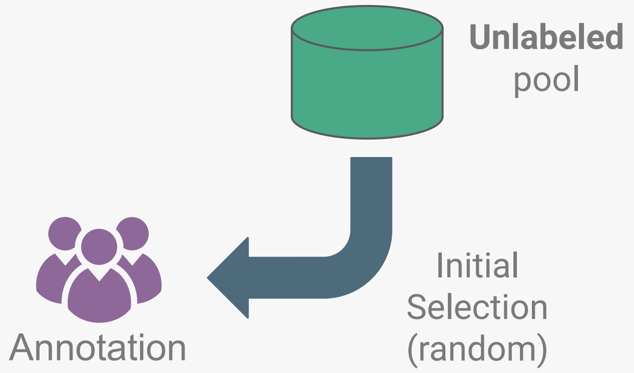
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



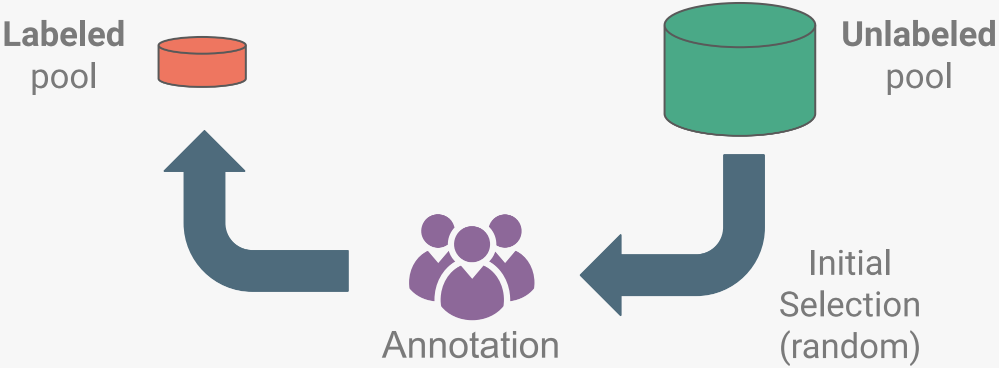
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



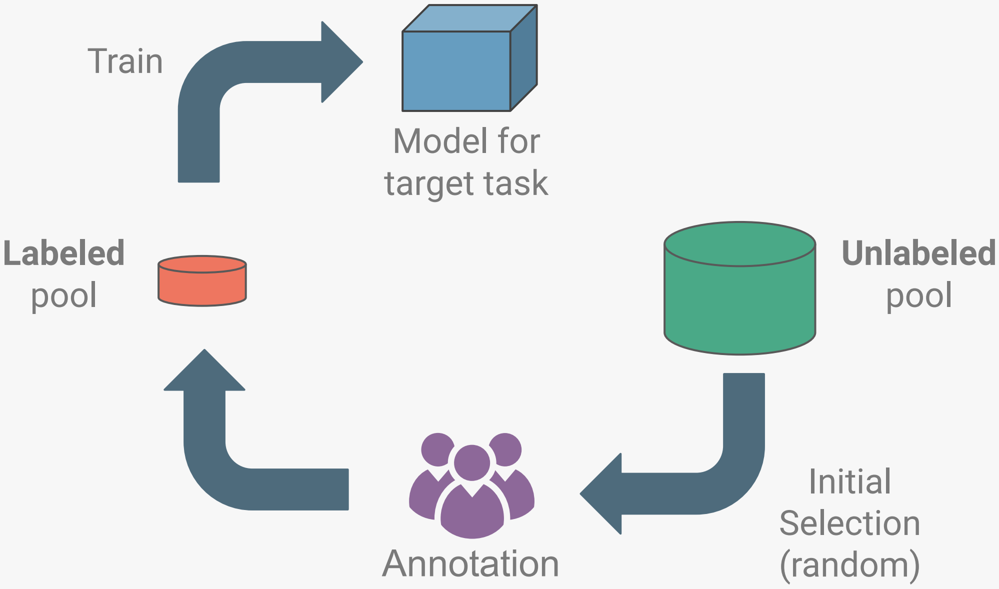
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



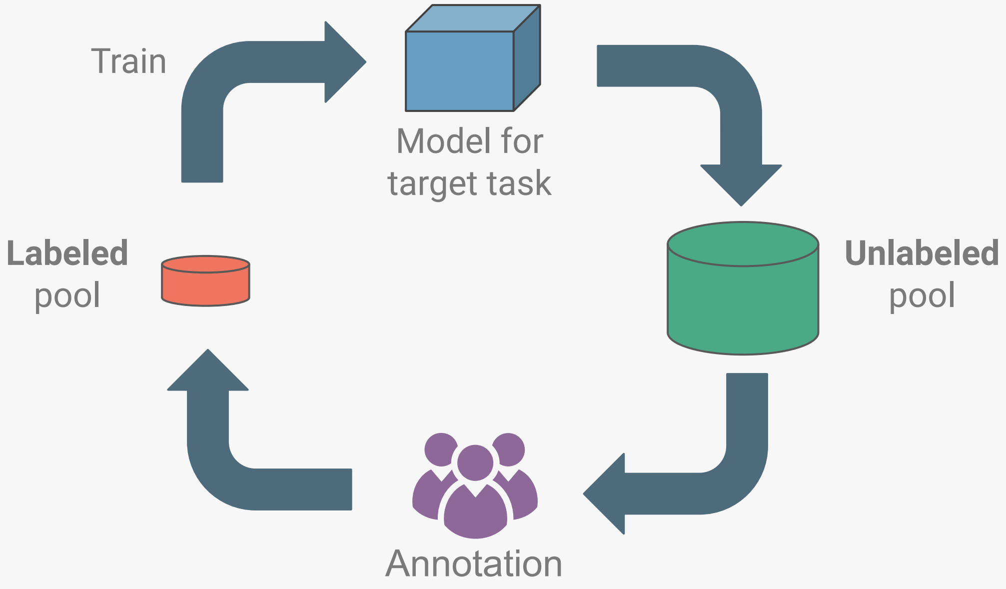
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



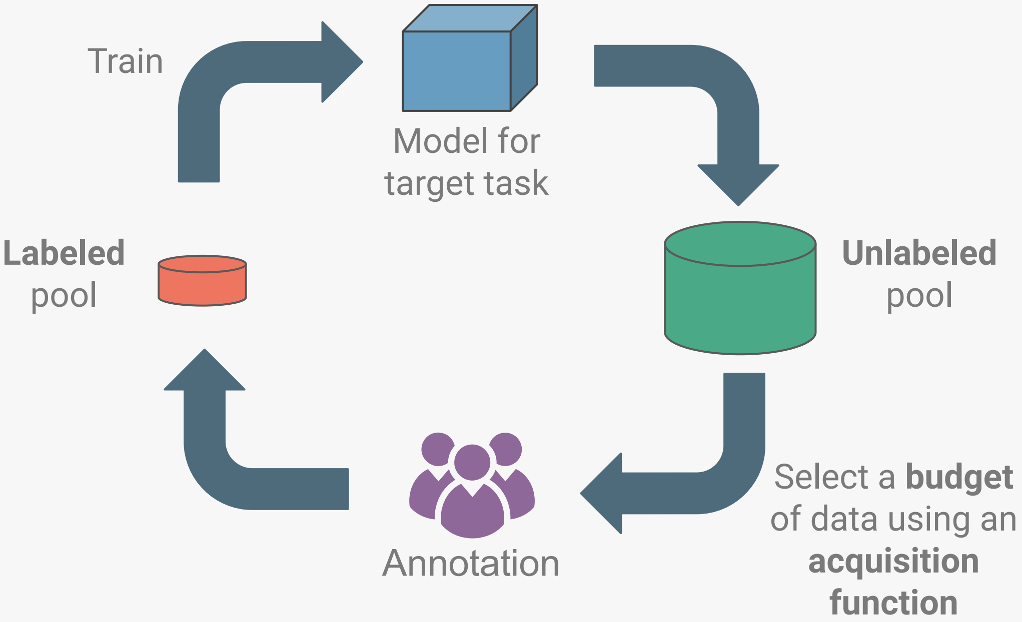
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



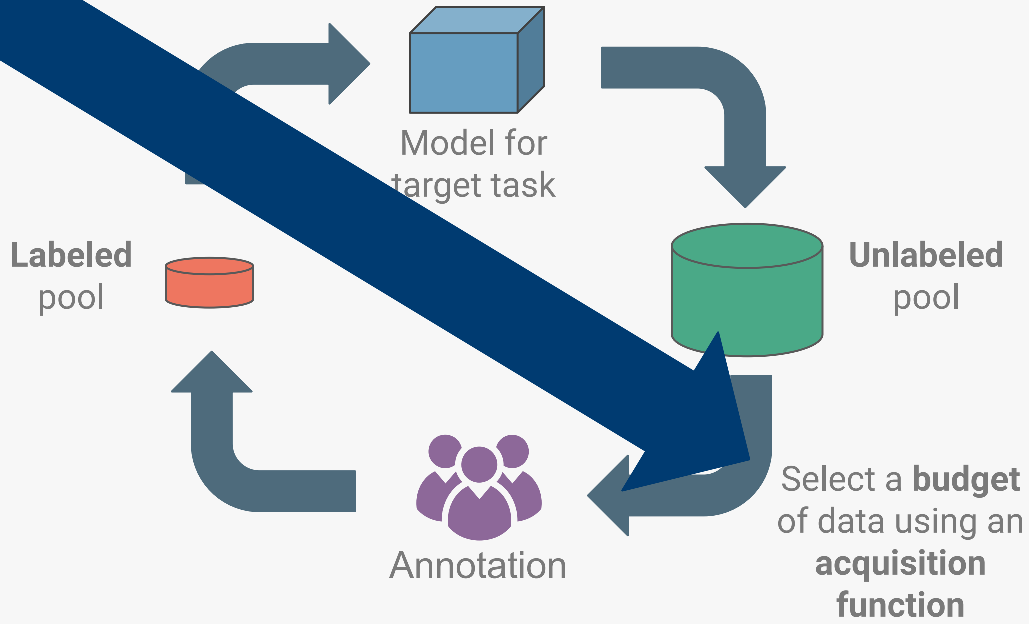
Active Learning
Core Concept

- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**

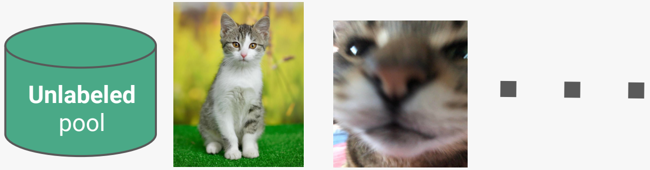


Active Learning
Core Concept

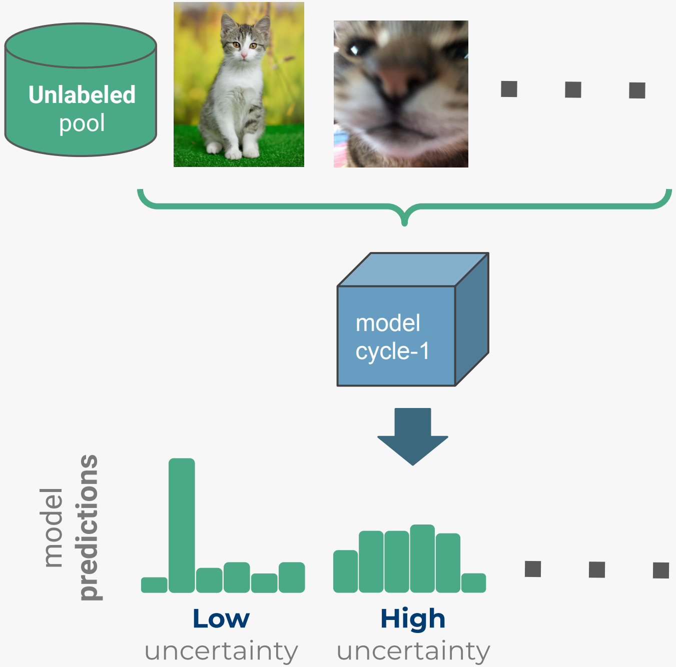
- Select the **best** images to be **annotated** for a **model** trained for a target task in **cycles**



Let's start the discussion with image classification



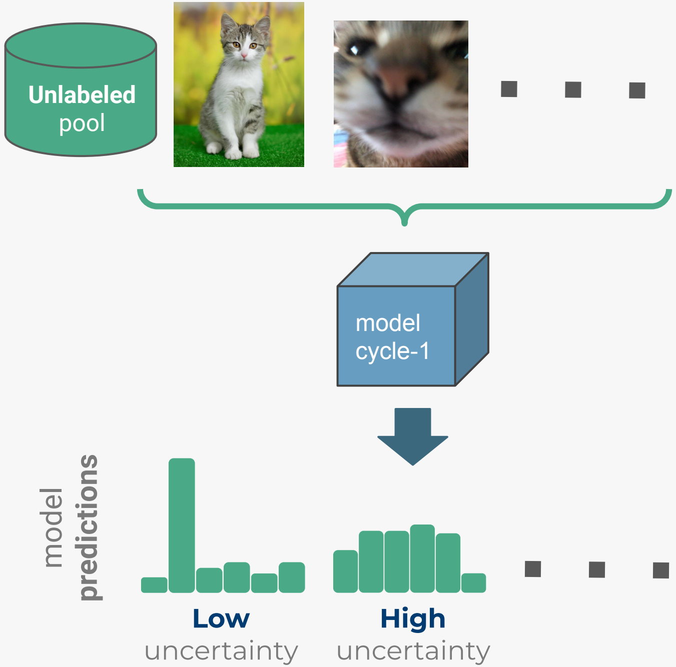
Typical Active Learning selection based on



Typical Active Learning selection based on

Data informativeness - model uncertainty

*“The ability to reduce the **generalization error** of the classification model”*



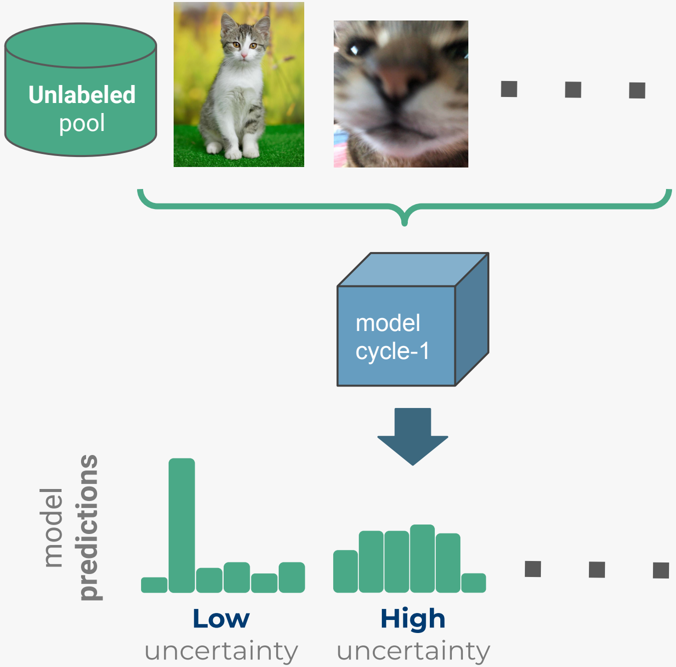
Typical Active Learning selection based on

Data informativeness - model uncertainty

“The ability to reduce the ***generalization error*** *of the classification model”*

Based on the

- **probabilities** outputted by the model [Settles Tech Rep’09, Wang et al. TCSVT’16],
  - Softmax **Confidence**, Softmax **Margin**, Softmax **Entropy**



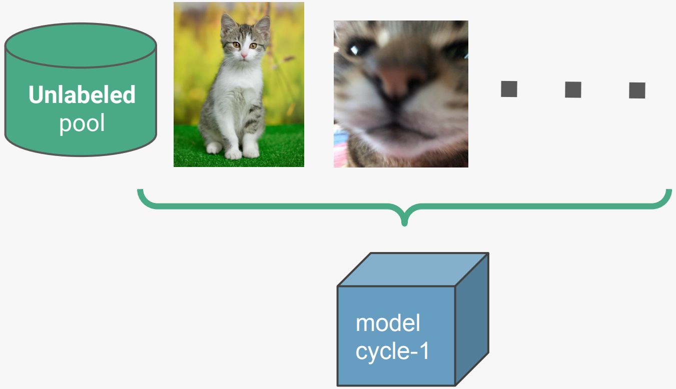
Typical Active Learning selection based on

Data informativeness - model uncertainty

*"The ability to reduce the **generalization error** of the classification model"*

Based on the

- **probabilities** outputted by the model [Settles Tech Rep'09, Wang et al. TCSVT'16],
  - Softmax **Confidence**, Softmax **Margin**, Softmax **Entropy**
- uncertainty between **outputs of several models**
  - MC Dropout [Gal et al. ICML'17]
  - Ensembles [Yang et al. Springer'17, Beluch et al. CVPR'18]
- the **impact** of the sample on the model [Ash et al. ICLR'20, Yoo1 & Kweon CVPR'19]



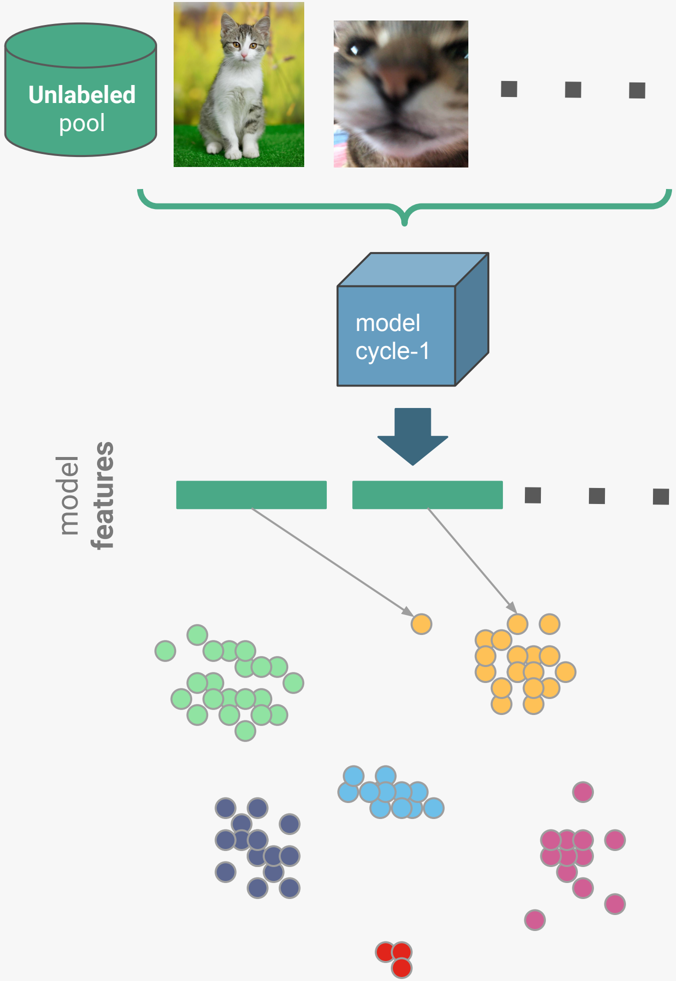
Typical Active Learning selection based on

Data representativeness - data geometry

*How much are selected data **representative** of the dataset?*

Selected images should be

- **Diverse** & represent the **whole** dataset
- **Feature**-based methods
  - **core-sets** [Sener & Savarese ICLR'18, Geifman & El-Yaniv arxiv'17]
  - k-means clustering [Zhdanov arxiv'19]



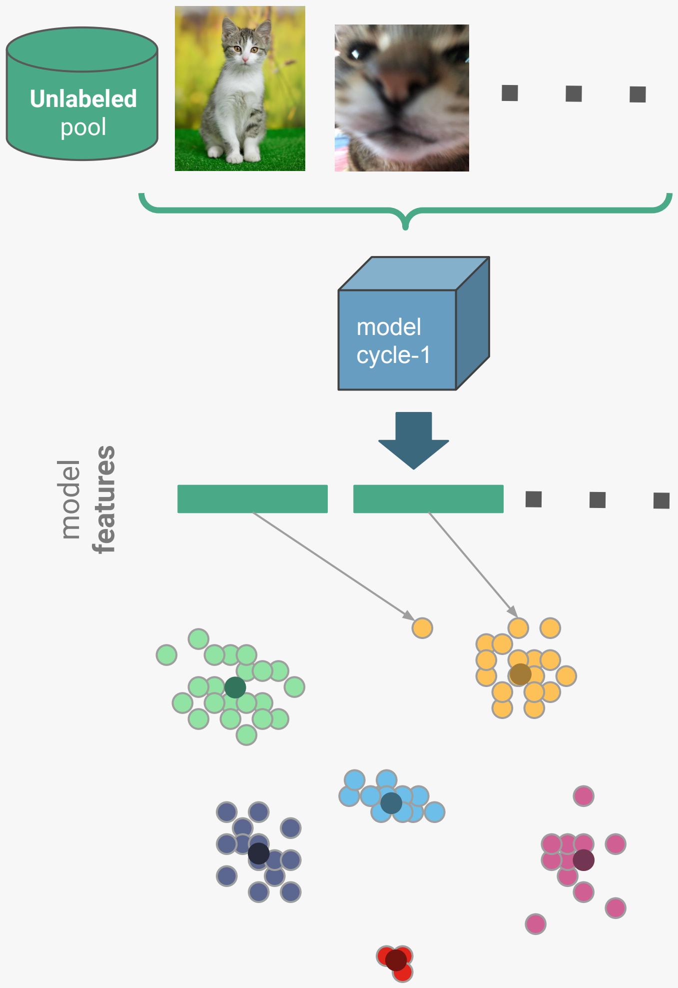
Typical Active Learning selection based on

Data representativeness - data geometry

*How much are selected data **representative** of the dataset?*

Selected images should be

- **Diverse** & represent the **whole** dataset
- **Feature**-based methods
  - **core-sets** [Sener & Savarese ICLR'18, Geifman & El-Yaniv arxiv'17]
  - k-means clustering [Zhdanov arxiv'19]



Typical Active Learning selection based on

Data representativeness - data geometry

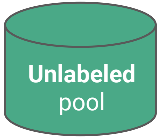
How much are selected data ***representative*** *of the dataset?*

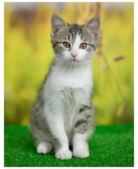
Selected images should be

- **Diverse** & represent the **whole** dataset
- **Feature**-based methods
    - **core-sets** [Sener & Savarese ICLR'18, Geifman & El-Yaniv arxiv'17]
    - k-means clustering [Zhdanov arxiv'19]

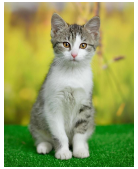
From image classification to more complex tasks

The data





The task



cat

dog
glasses

Image **classification**

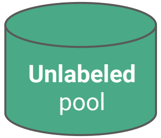
Active learning

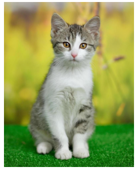
Select the best **images**

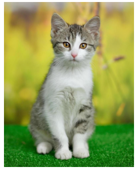
→ **Image**-level uncertainty / diversity

From image classification to more complex tasks

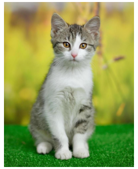
The data







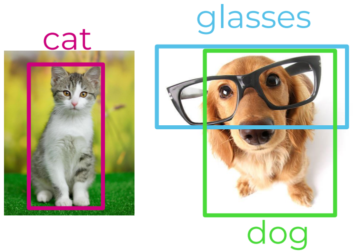
The task



cat

dog
glasses

Image **classification**



Object **detection**

Active learning

Select the best **images**

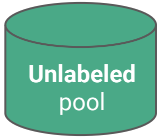
→ **Image**-level
uncertainty / diversity

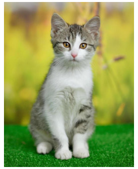
Select the best **images**

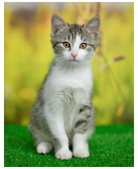
?

From image classification to more complex tasks

The data

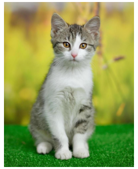






3D point clouds

The task



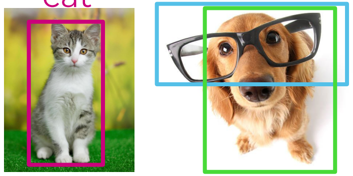
cat

dog
glasses

Image **classification**

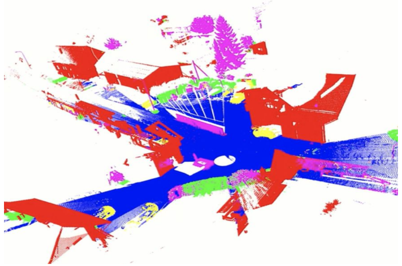
cat

glasses



dog

Object **detection**



Semantic **segmentation**

Active learning

Select the best **images**

➔ **Image**-level
uncertainty / diversity

Select the best **images**

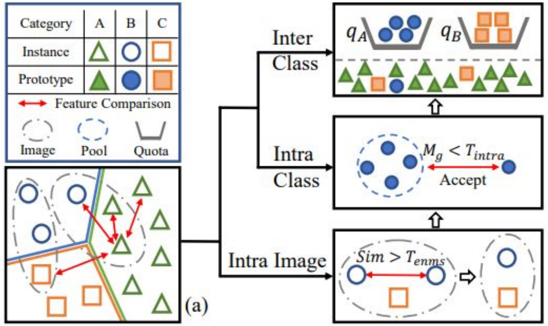
?

Select the best **scenes**

?

Active learning for object detection

- **Box localization** uncertainty
  - before / after localization refinement step [Kao et al, ACCV'18]
- **Instance-level** uncertainty
  - given different views [Elezi et al. CVPR'22]
  - using ensembling-like approach [Choi et al, ICCV'21]
- Combine **instance** & **image**-level diversity [Wu et al. CVPR'22]



Adapting from image classification to object detection

ECCV TEL AVIV 2022

Active Learning Strategies for Weakly-Supervised Object Detection

Huy V. Vo[1,2], Oriane Siméoni[2], Spyros Gidaris[2], Andrei Bursuc[2], Patrick Pérez[2], Jean Ponce[1,3]

[1] Inria and DI/ENS (ENS-PSL, CNRS, Inria), [2] Valeo.ai, [3] Center for Data Science






Weakly-Supervised Object Detection

Require **only tag** annotation

- Costs **1 sec** per class per image
- Tags could also be obtained **automatically**

Weakly-Supervised Object Detection

Require **only tag** annotation

- Costs **1 sec** per class per image
- Tags could also be obtained **automatically**

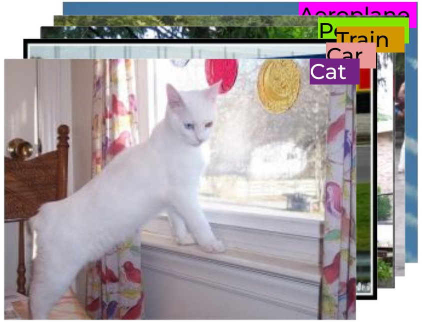
BUT

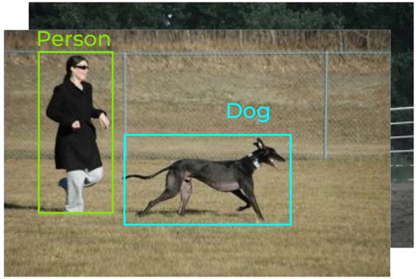
- **Lower** performances
- **Recurrent** type of **mistakes**

Example of weakly-supervised **predictions**

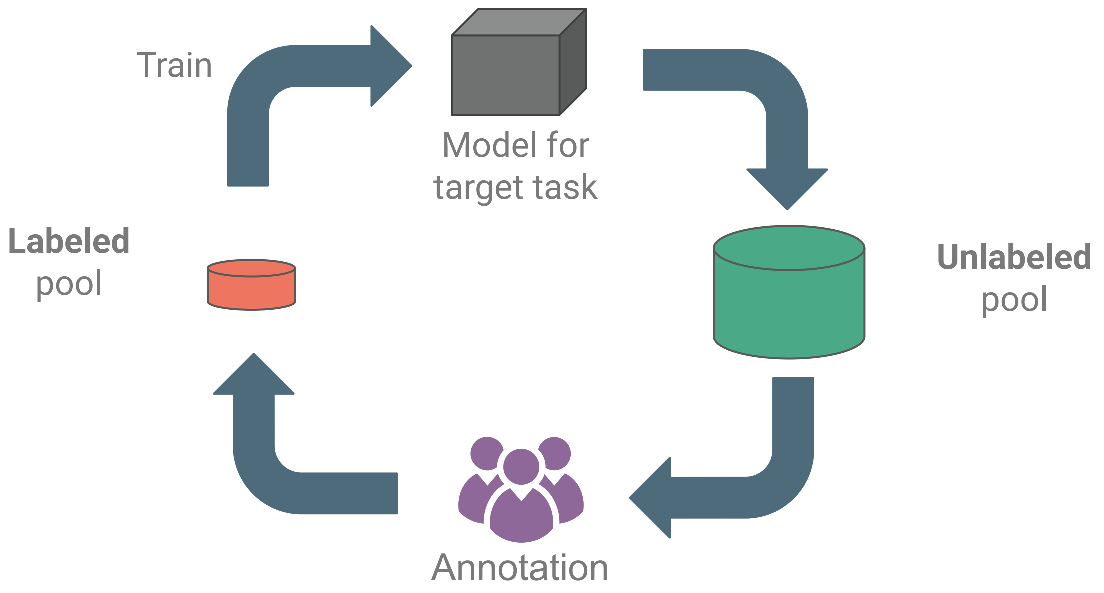
Weakly-Supervised Object Detection

Idea :
Improve weakly-supervised detectors with a **few fully-annotated** images

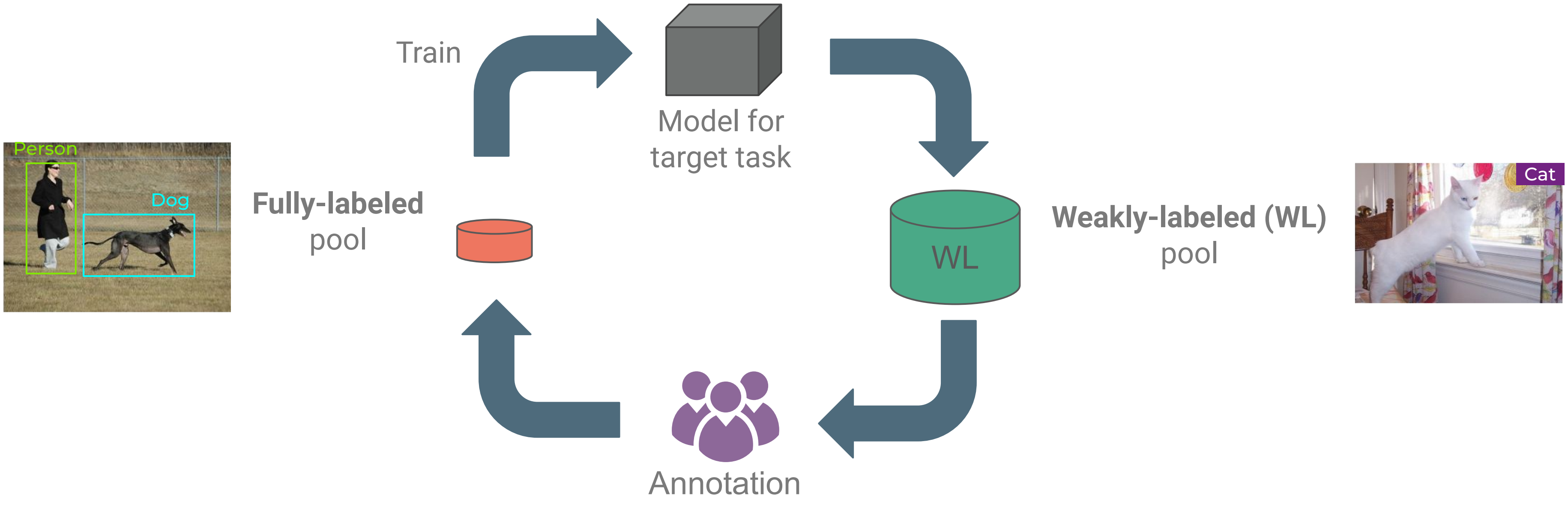




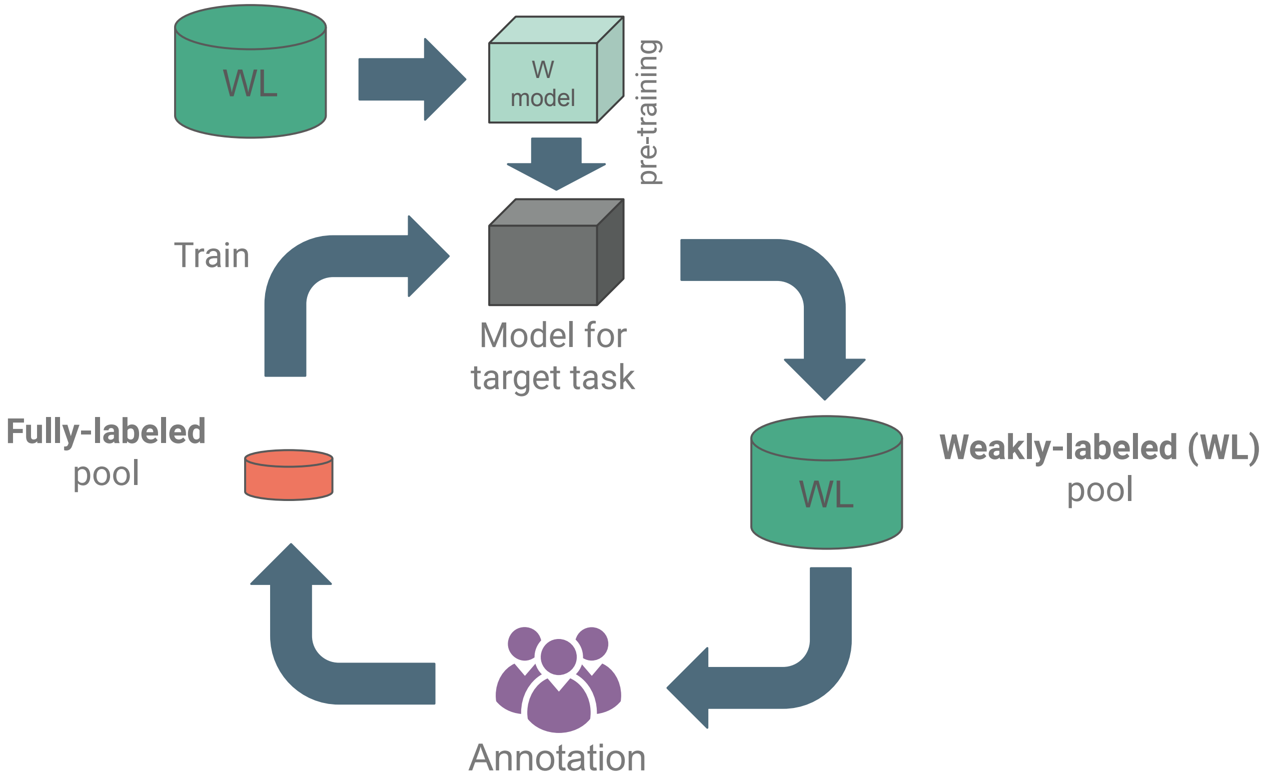
The proposed pipeline



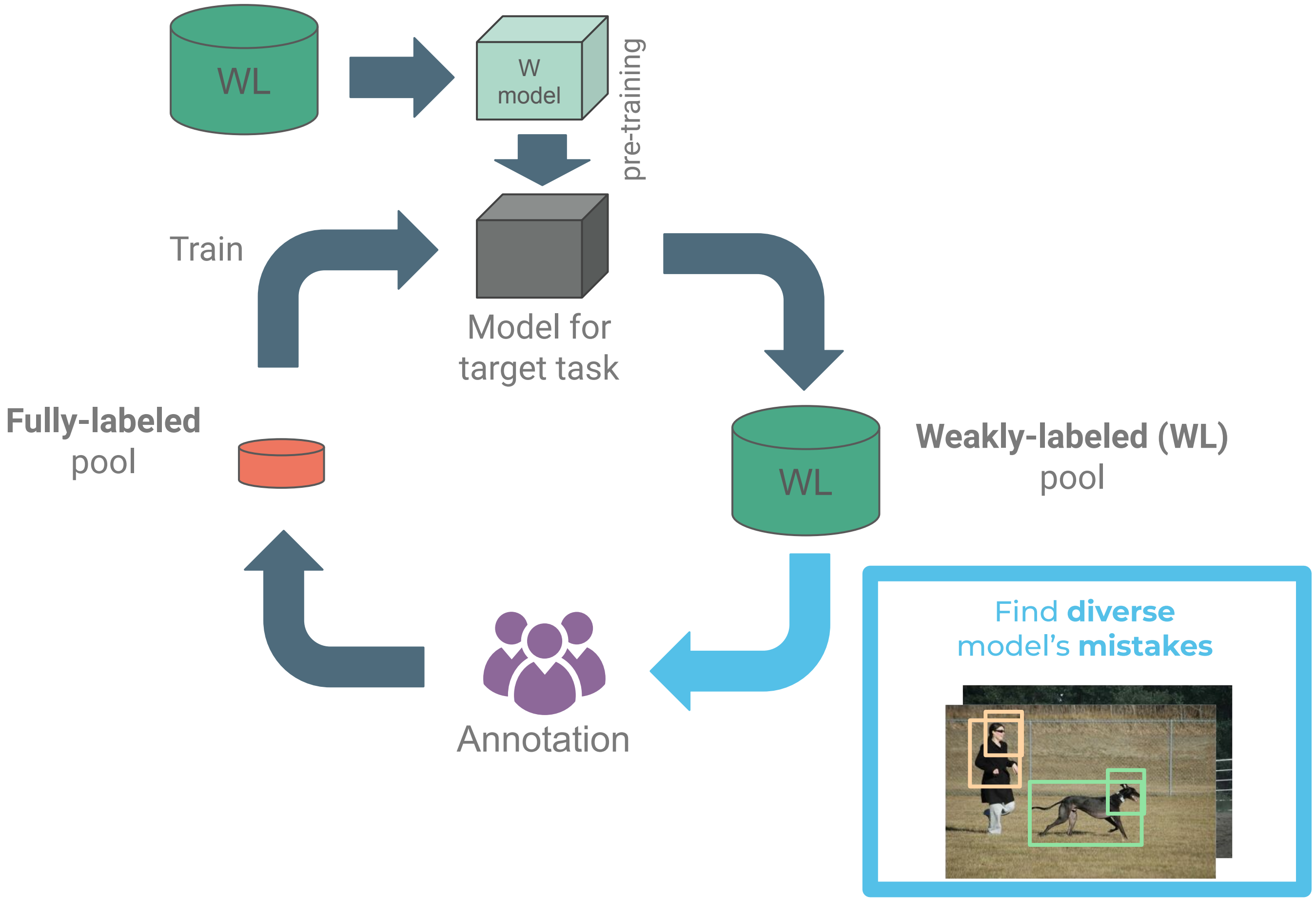
The proposed pipeline



The proposed pipeline



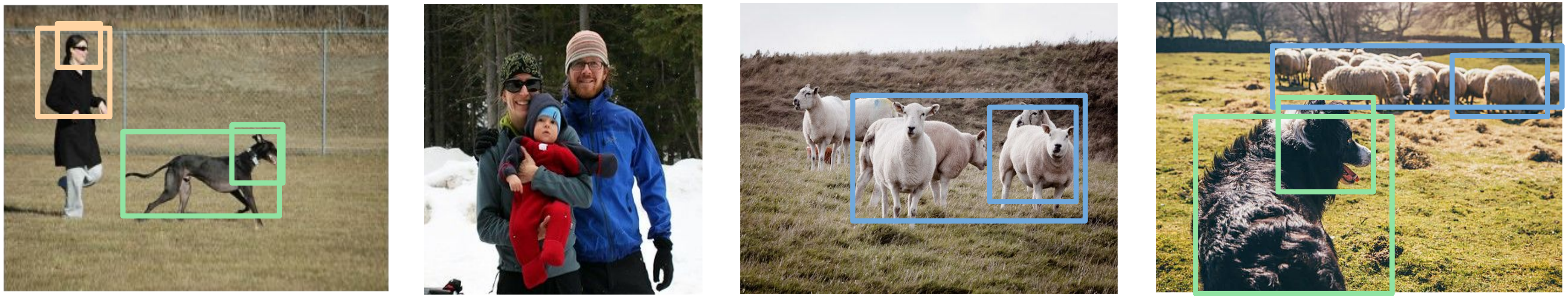
The proposed pipeline



Box-in-Box AL strategy

Step 1: find **Box-in-Box (BiB)** pairs

- Diverse pairs over the dataset of the **same class**
- with a **box "contained"** in the other

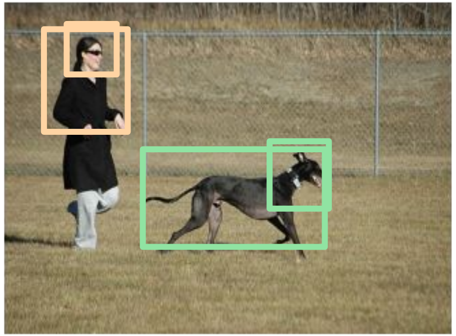


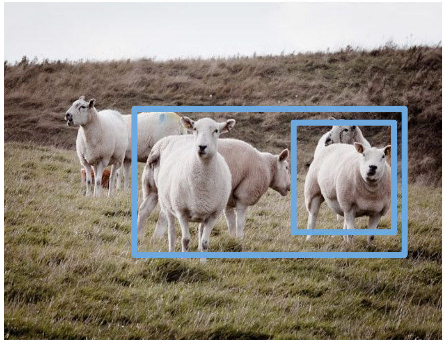
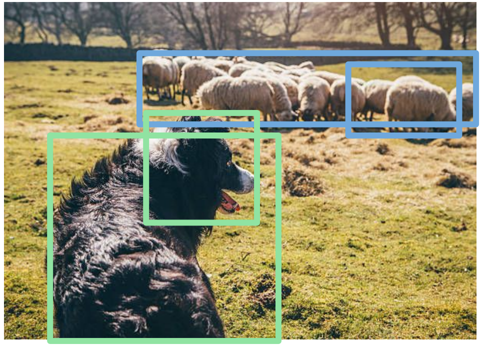
Examples of **BiB** pairs

Box-in-Box AL strategy

Step 1: find **Box-in-Box** (**BiB**) pairs

- Diverse pairs over the dataset of the **same class**
- with a **box "contained"** in the other



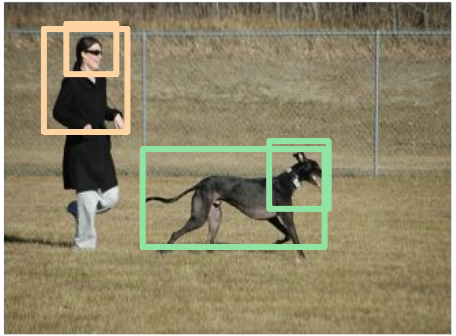



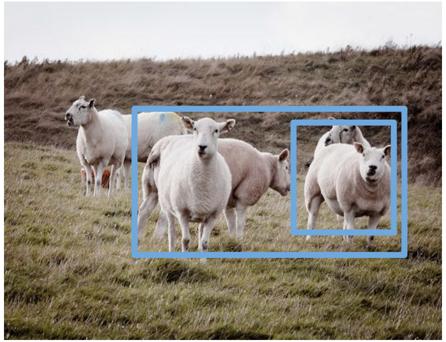
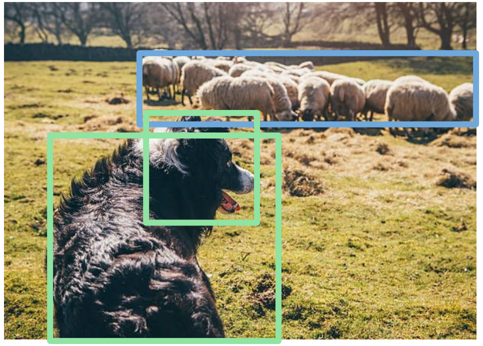
Examples of **BiB** pairs

Box-in-Box AL strategy

Step 1: find **Box-in-Box (BiB)** pairs
- Diverse pairs over the dataset of the **same class**
- with a **box "contained"** in the other

Step 2: Select **diverse mistakes** over the dataset
- Apply a clustering method



Examples of **BiB** pairs

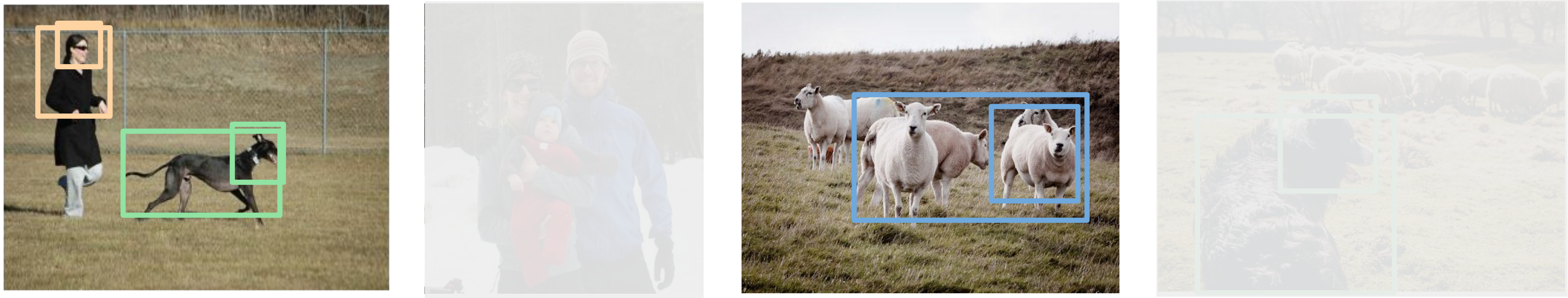
Box-in-Box AL strategy

Step 1: find **Box-in-Box (BiB)** pairs

- Diverse pairs over the dataset of the **same class**
- with a **box "contained"** in the other

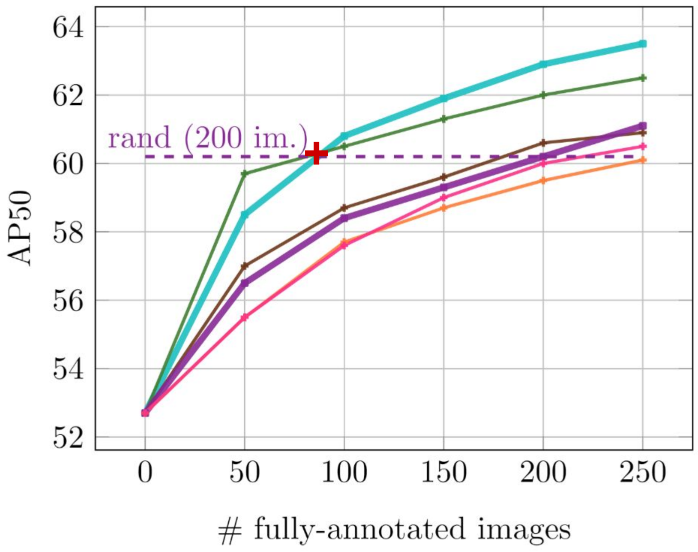
Step 2: Select **diverse mistakes** over the dataset

- Apply a clustering method

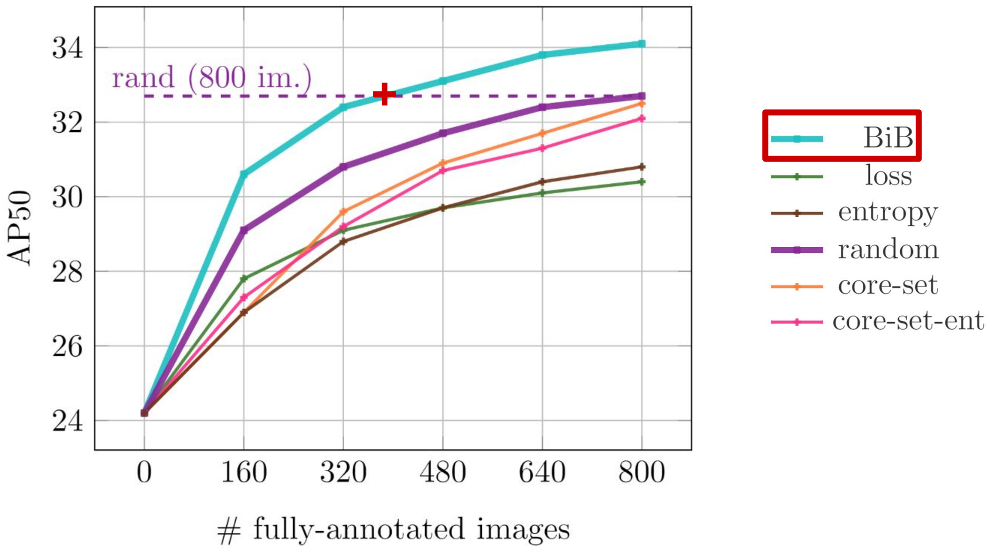


Diverse examples of **BiB** pairs

Active learning for weakly-sup detector results



VOC07 results



COCO results

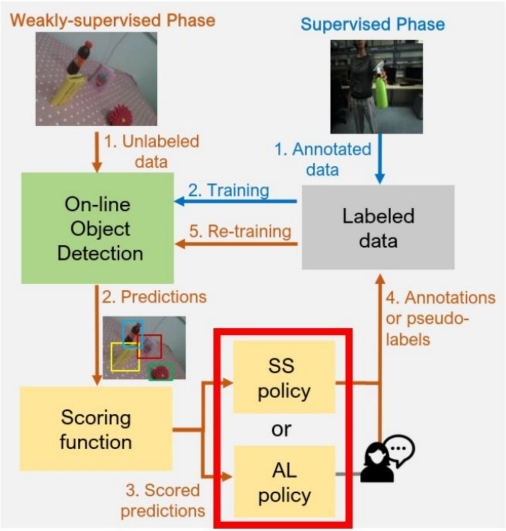
- **BiB** requires **half** the number of fully annotated images to obtain the same as random
- with **only 1%** (800 images) fully annotated get **91% of fully-supervised** model's score on COCO

With a robotic perspective @RO-MAN’22

From Handheld to Unconstrained Object Detection: a Weakly-supervised On-line Learning Approach

Elisa Maiettini*[1], Andrea Maracani*[1,2,3] Raffaello Camoriano[2], Giulia Pasquale[1], Vadim Tikhanoff[4], Lorenzo Rosasco[2,3,5] and Lorenzo Natale[1]

- With a robot perspective
- Also starts with a **weakly**-supervised phase
- Let the robot explore the environment with a **limited** human labeling budget
- Apply
  - Human-made annotation or
  - Semi-supervised labels



Follow up work @ ICCV'23

ALWOD [Wang et al. ICCV'23]

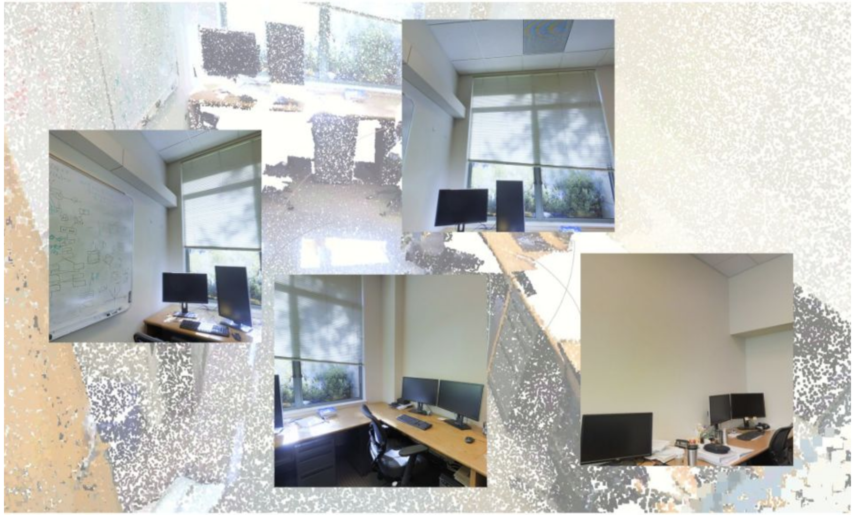
ALWOD: Active Learning for Weakly-Supervised Object Detection

Yuting Wang[1], Velibor Ilic[2], Jiatong Li[1], Branislav Kisačanin[3,2], and Vladimir Pavlovic[1]

[1]Rutgers University, NJ, USA
[2]The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia
[3]Nvidia Corporation, TX, USA
yw632@rutgers.edu, velibor.ilic@ivi.ac.rs, jiatong.li@rutgers.edu, b.kisacanin@ieee.org, vladimir@cs.rutgers.edu

Application to 3D data

Point Cloud scenes



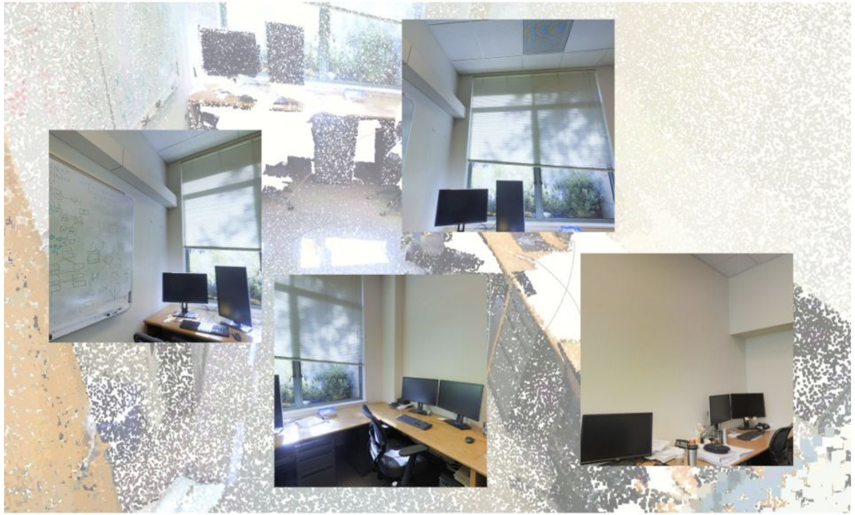
Example of **Point Cloud (PC)** scenes

- 3D PC data
- Temporally **aligned** 2D **images**
- **No annotation** yet available

The task tackled:

- **Semantic PC segmentation**

Point Cloud scenes



Example of **Point Cloud (PC)** scenes

- 3D PC data
- Temporally **aligned** 2D **images**
- **No annotation** yet available

The task tackled:

- **Semantic PC segmentation**

Annotation of 3D PC **very expensive**

- Require **a label per point**
- A single scene in the order of **50-200k points**

Active learning for 3D data

What data to select ?

Select **full scenes** to be annotated based on

- The model **uncertainty**
  - Softmax Confidence [Wang et al. TODO'14]

    Softmax Margin [Wang et al. TODO'14]

    Softmax Entropy [Wang et al. TODO'14]
  - Segment Entropy [Lin et al. '20]
  - MC-Dropout [Gal et al. ICML'17]
- The scene **diversity**
  - CoreSet [Sener & Savarese ICLR'18]

Active learning for 3D data

What data to select ?

Select **full scenes** to be annotated based on
- The model **uncertainty**
  - Softmax Confidence [Wang et al. IJCNN'14]
    Softmax Margin [Wang et al. IJCNN'14]
    Softmax Entropy [Wang et al. IJCNN'14]
  - Segment Entropy [Lin et al. '20]
  - MC-Dropout [Gal et al. ICML'17]
- The scene **diversity**
  - CoreSet [Sener & Savarese ICLR'18]

Select **regions** (ensemble of close points) to be annotated
- ReDAL [Wu et al. ICCV'21]
  - Mix regions diversity / informativness
- LiDAL [Hu et al. ECCV'22]
  - Additionally integrates self-training and pseudo-labeling

Leveraging 2D data to boost 3D Active Learning

École des Ponts ParisTech

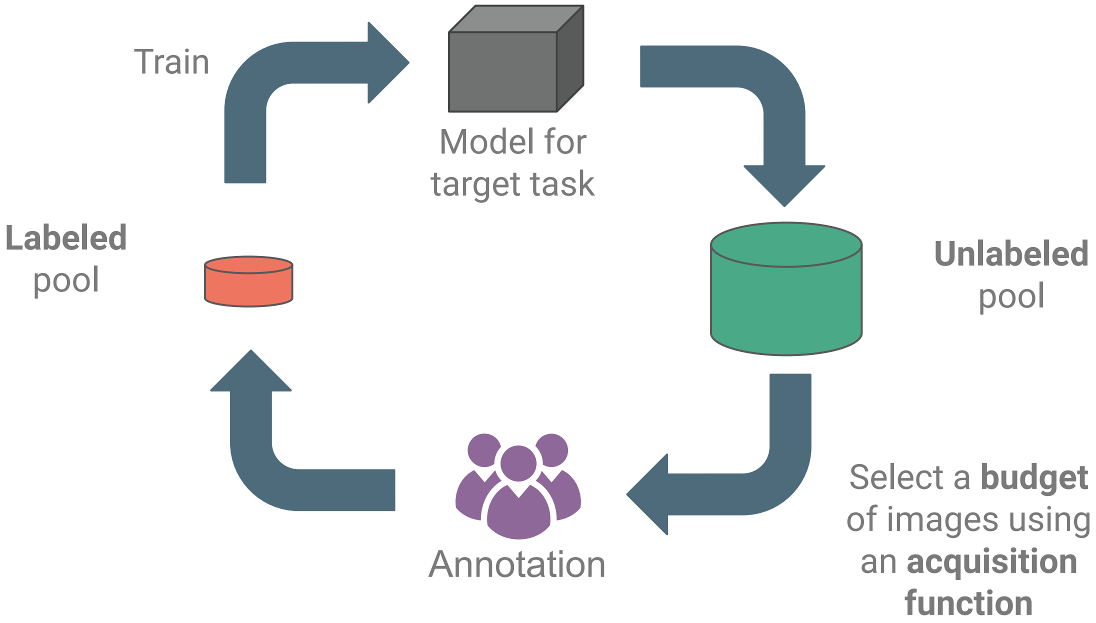
valeo.ai

Seeding Active Learning for 3D Semantic Segmentation

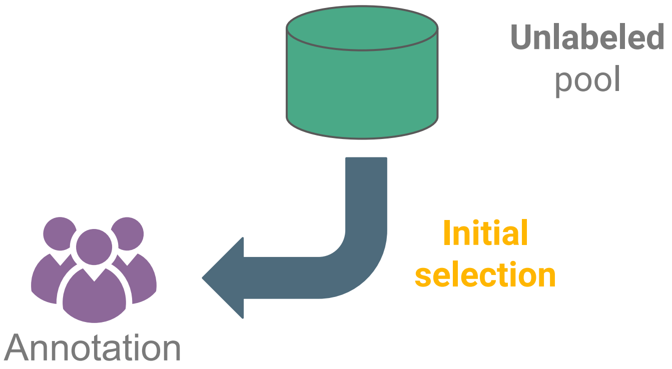
Nermin Samet[1], Oriane Siméoni[2], Gilles Puy[2], Renaud Marlet[1], Vincent Lepetit[1]

[1] LIGM, Ecole des Ponts, Univ Gustave Eiffel, [2] Valeo.ai

Where do we start from ?

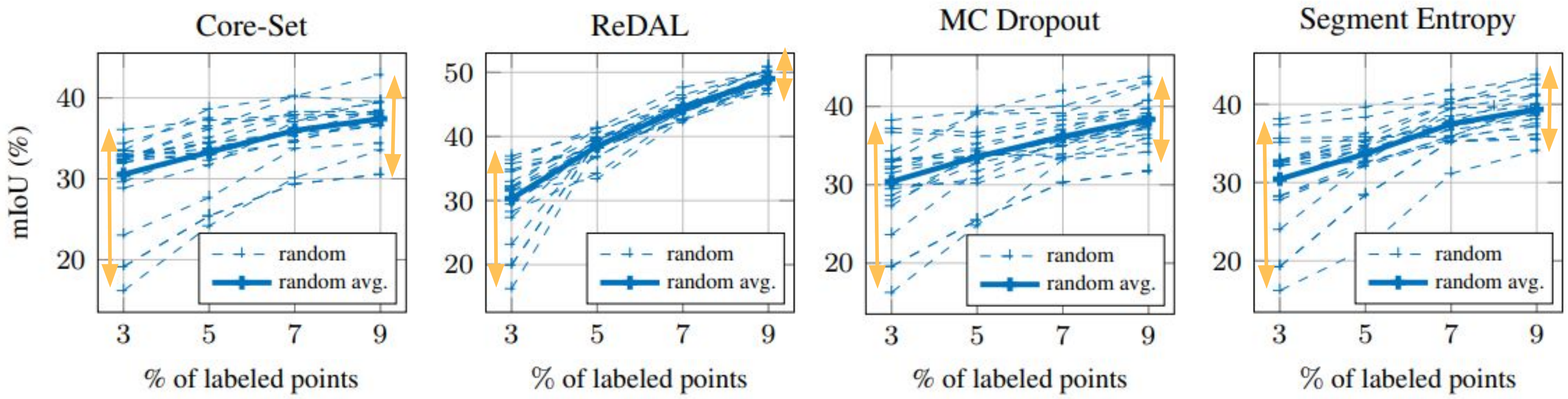


Where do we start from ?



- Initial selection typically is **randomly** selected
- Could we make the **initial label** smarter ?
- What impact ?

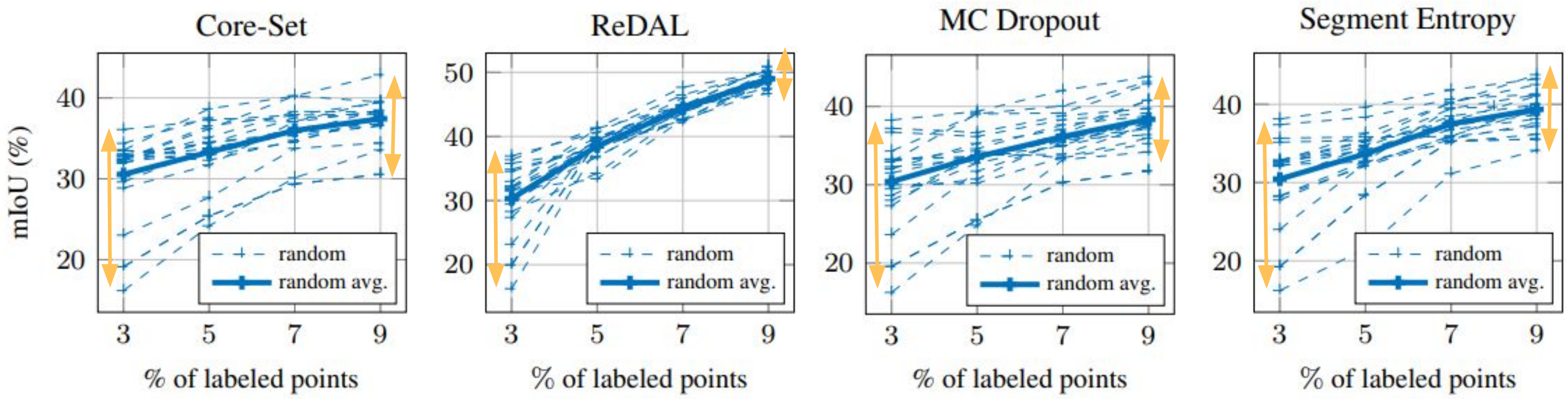
The effect of a random selection



- High **variability** between different **draws**
- Variability with all SoTA AL methods
- The larger variability at first cycles

Could we design the selection of a **lucky seed** for 3D data ?

The effect of a random selection



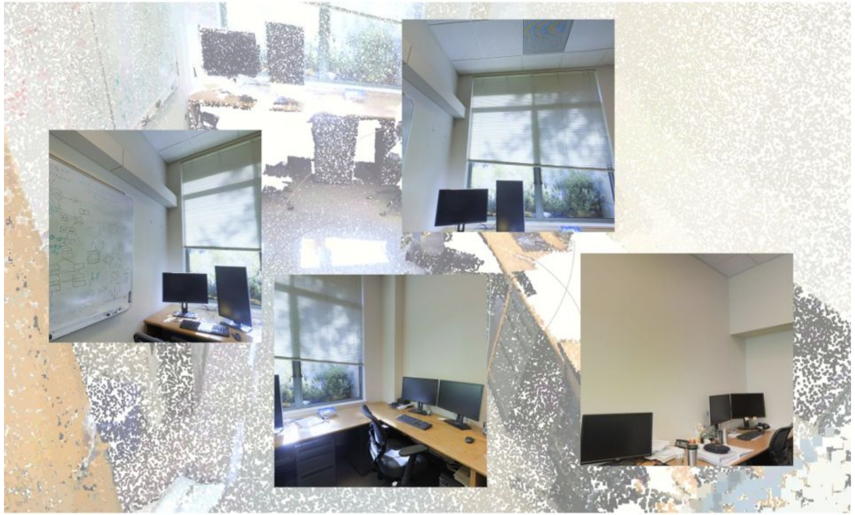
- High **variability** between different **draws**
- Variability with all SoTA AL methods
- The larger variability at first cycles

Has been **explored in the 2D** space

- Diversity-based strategy using K-means [Pourahmadi et al. WACV'23, Chen et al. NeurIPS'22] or core-set [Mahmood et al. ICLR'22] on self-sup. features
- Generation of pseudo-labels via proxy task [Nath at al. MICCAI'22]

Could we design the selection of a **lucky seed** for 3D data ?

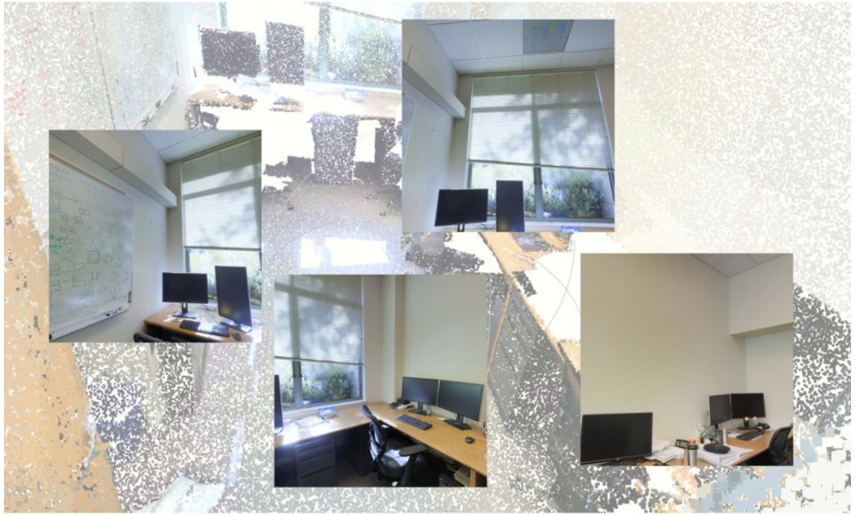
Point Cloud scenes



Example of **point cloud (PC)** scenes

- 3D PC data
- Temporally **aligned** 2D **images**
- **No annotation** yet available

Point Cloud scenes



Example of **point cloud (PC)** scenes

- 3D PC data
- Temporally **aligned** 2D **images**
- **No annotation** yet available

High **redundancy** in data → **diversity** important

Diversity-based selection: what features ?

But **what features to use**?

- No label → **self-supervised** features

RotNet [Gidaris et al. ICLR'18]

Diversity-based selection: what features ?

But **what features to use**?

- No label → **self-supervised** features
- Self-supervised **3D PC feature**, available **but sensitive** to:
  - scene type (indoors vs outdoors)
  - sensor type (photogrammetry, depth cameras, lidars)

RotNet [Gidaris et al. ICLR'18]

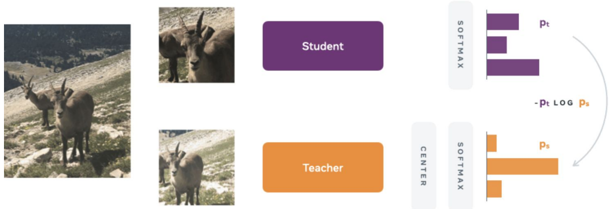
Diversity-based selection: what features ?

But **what features to use**?

- No label → **self-supervised** features
- Self-supervised **3D PC feature**, available **but sensitive** to:
    - scene type (indoors vs outdoors)
    - sensor type (photogrammetry, depth cameras, lidars)
- Self-supervised **2D feature**:
    - Trained on larger dataset → **generalize better**
    - Less sensitive to data specificity (eg. both indoor/outdoor)
    - Good **discrimilality** properties

RotNet [Gidaris et al. ICLR’18]

DeepCluster [Caron et al. ECCV’18]
MOCO [He et al. CVPR’20]
MAE [He et al. CVPR’22]



DINO [Caron et al. ICCV’21]

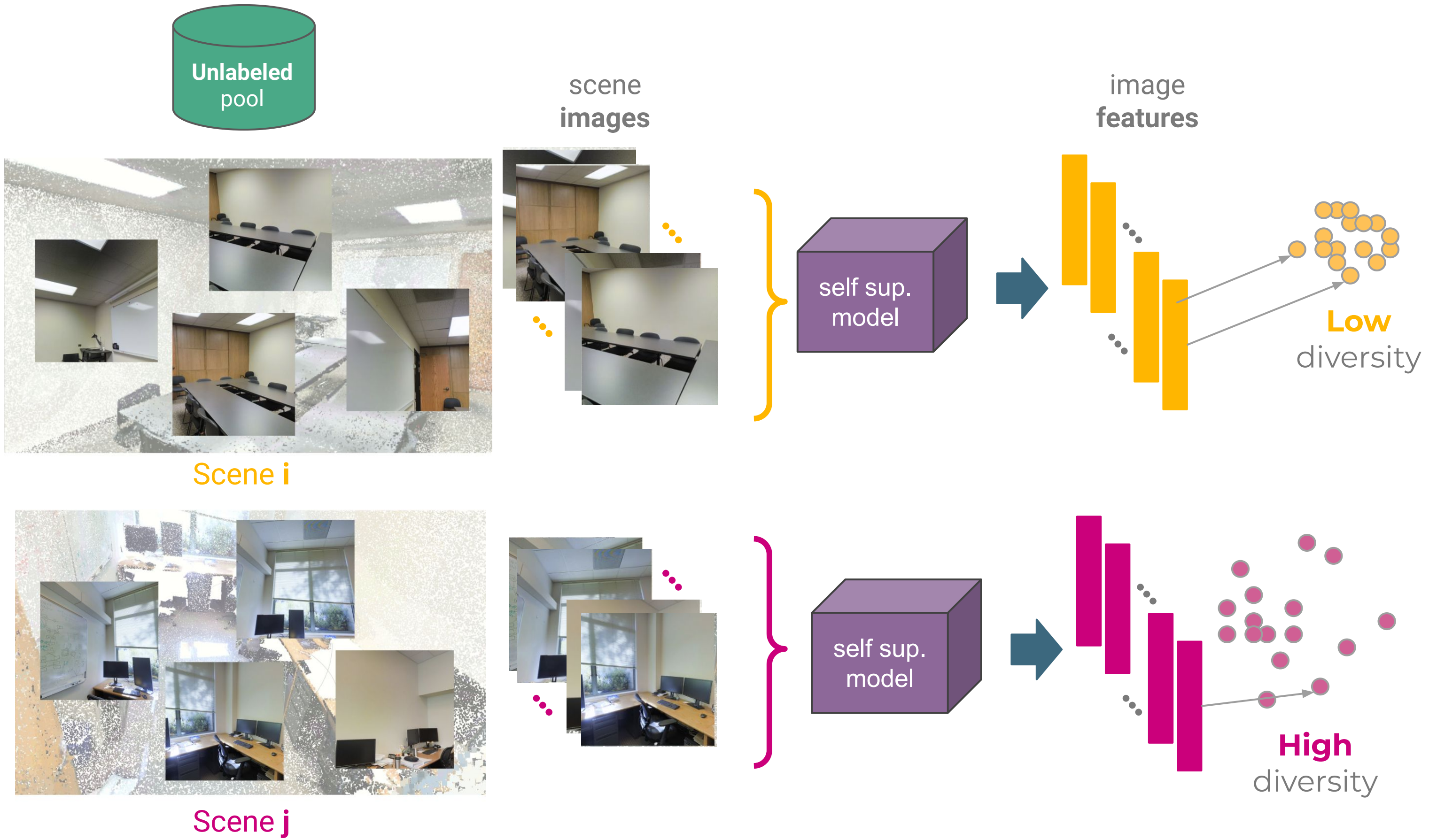
What we propose

- Select the **most diverse** scenes with
  - high **intra**-scene diversity
    - Eg. high **variety of objects** depicted
    - Not uniform room
  - high **inter**-scene diversity
    → Select **different type** of scenes
  - use **image features** to evaluate the scenes diversity

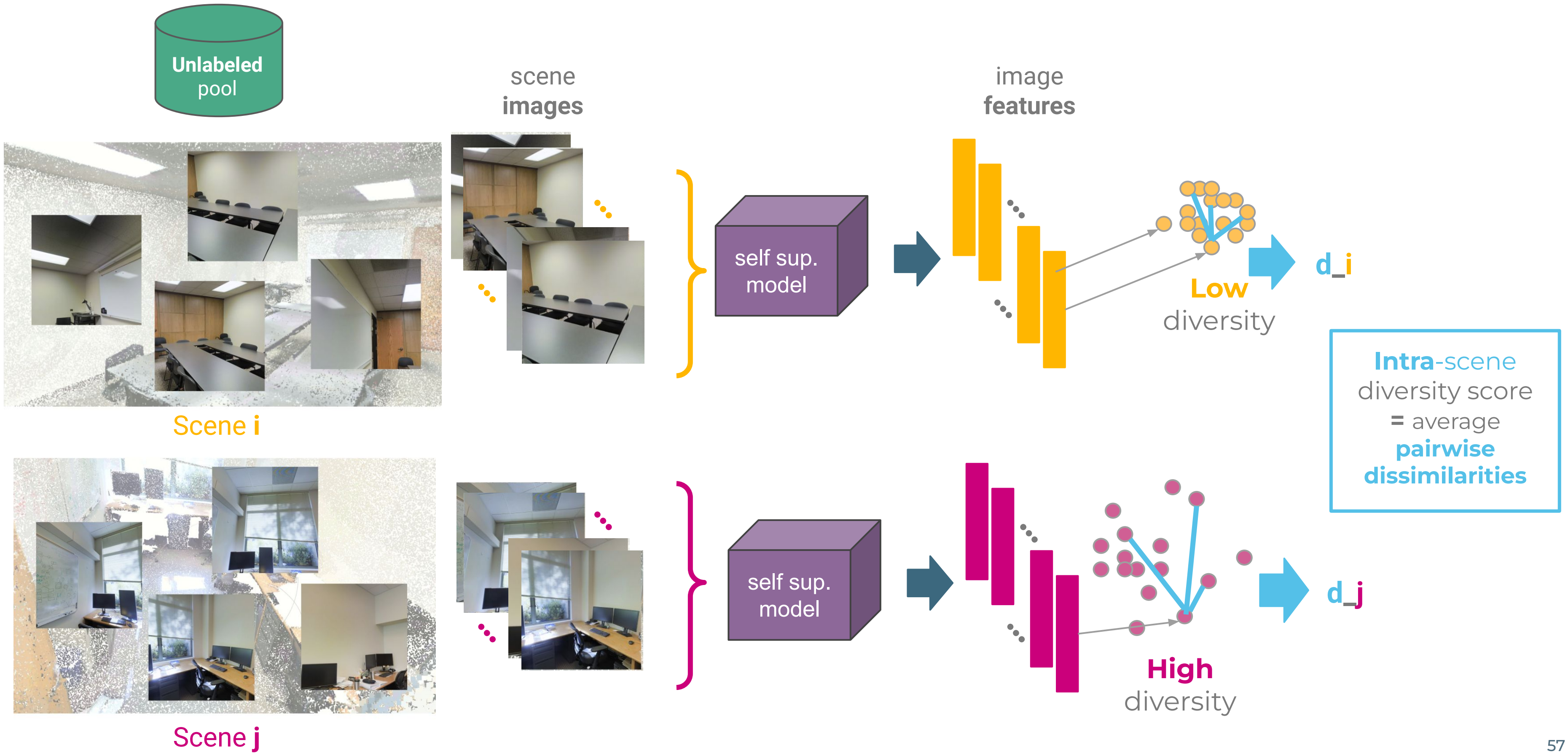
Intra-scene diversity score

Inter-scenes diversity score

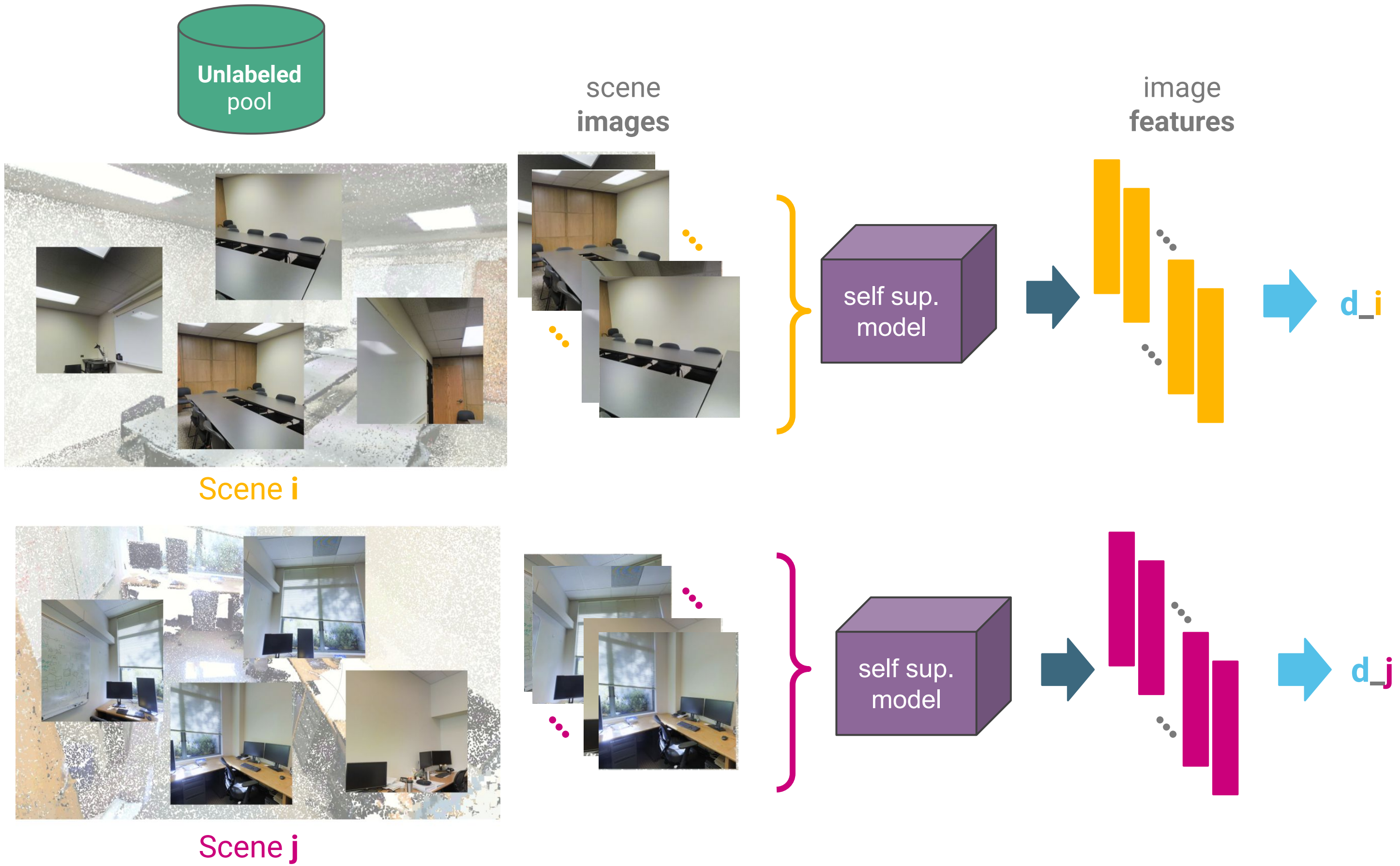
Modelling intra-scene diversity



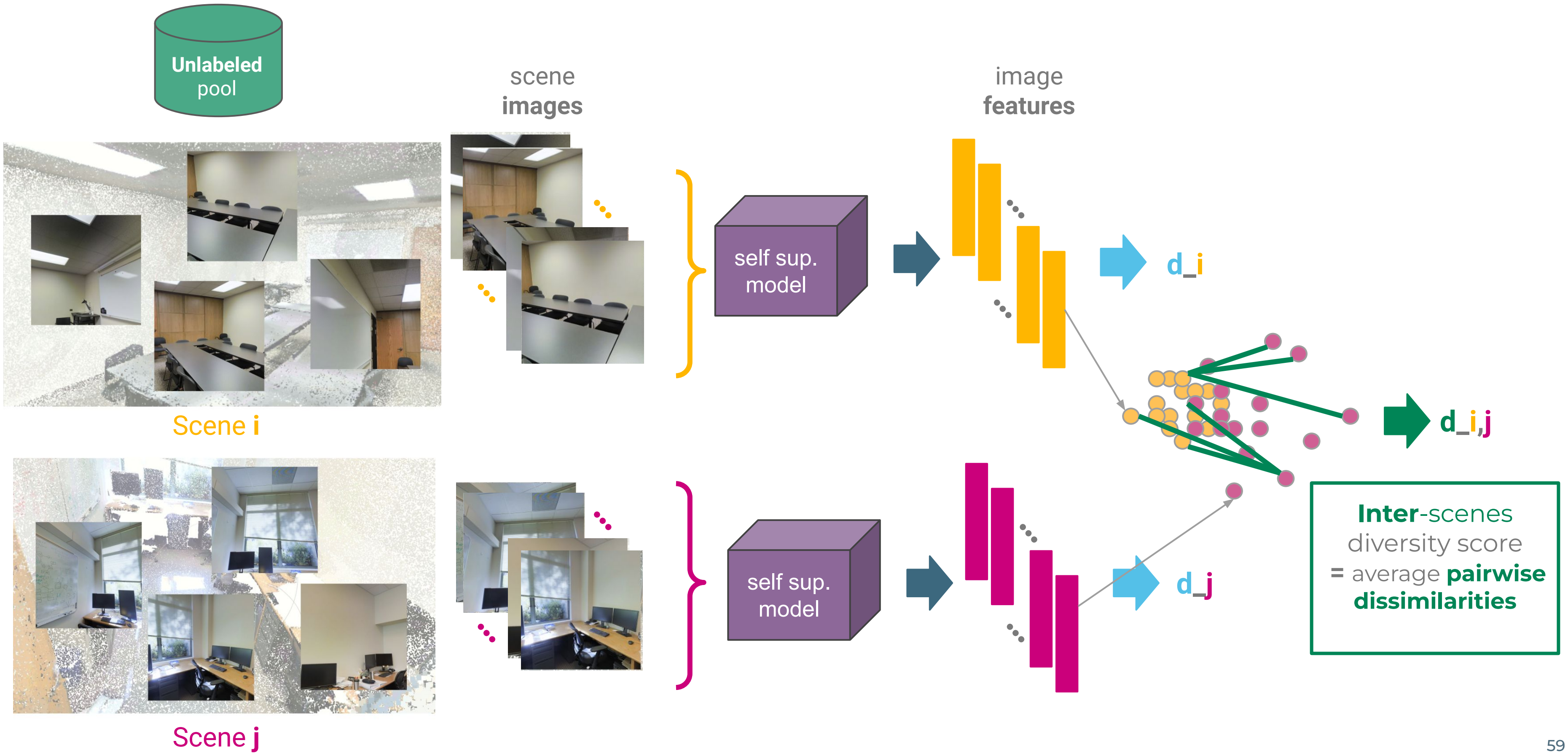
Modelling intra-scene diversity



Modelling intra-scene diversity



Modelling intra-scene diversity



Modelling intra-scene diversity

Modelling intra-scene diversity

Modelling inter-scene diversity

Modelling inter-scene diversity

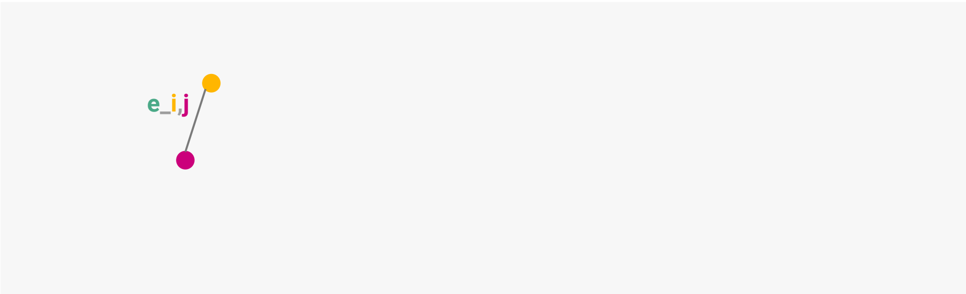
Unlabeled pool

Scene i

Scene j

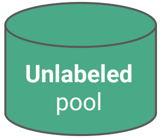
...

e_i,j

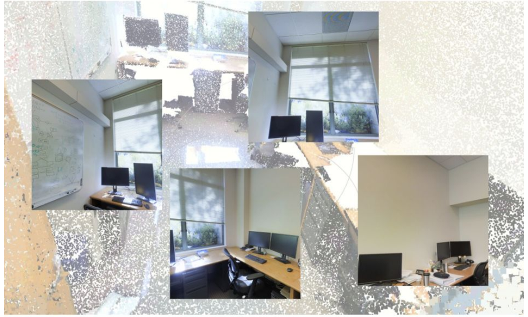


Modelling inter-scene diversity

Modelling inter-scene diversity



Scene **i**



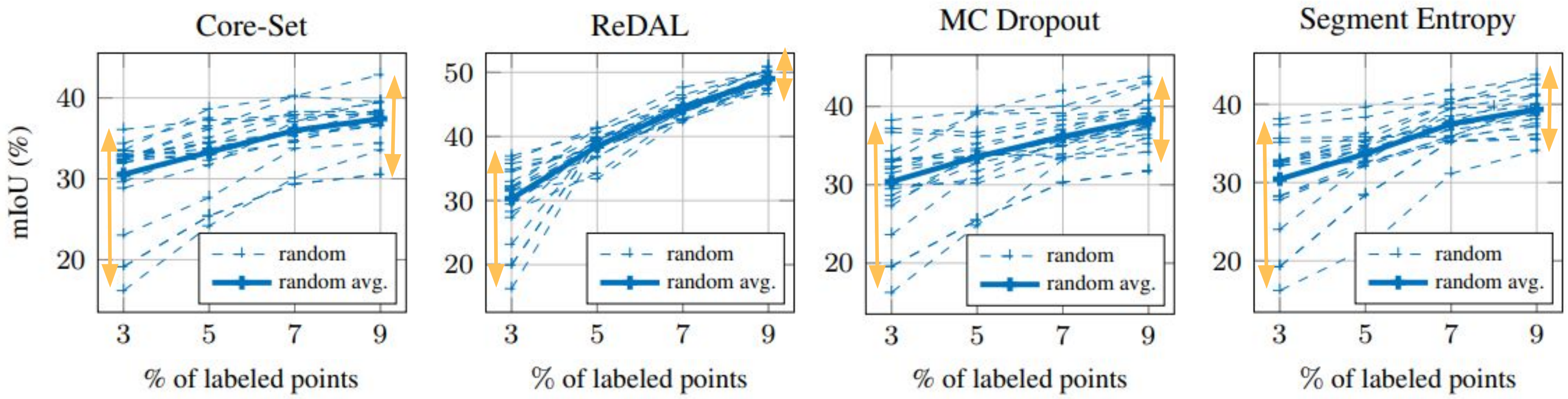
Scene **j**

...

e_i,j

Optimization to select data implying the **most diversity** within a budget

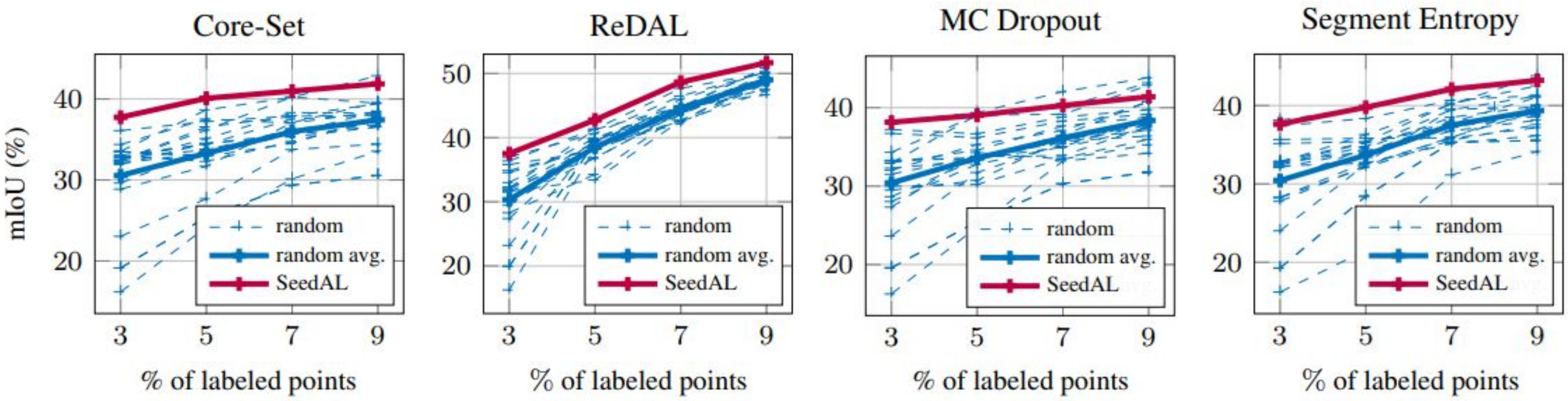
The effect of a random selection



- High **variability** between different **draws**
- Variability with all SoTA AL methods
- The larger variability at first cycles

Could we design the selection of a **lucky seed** ?

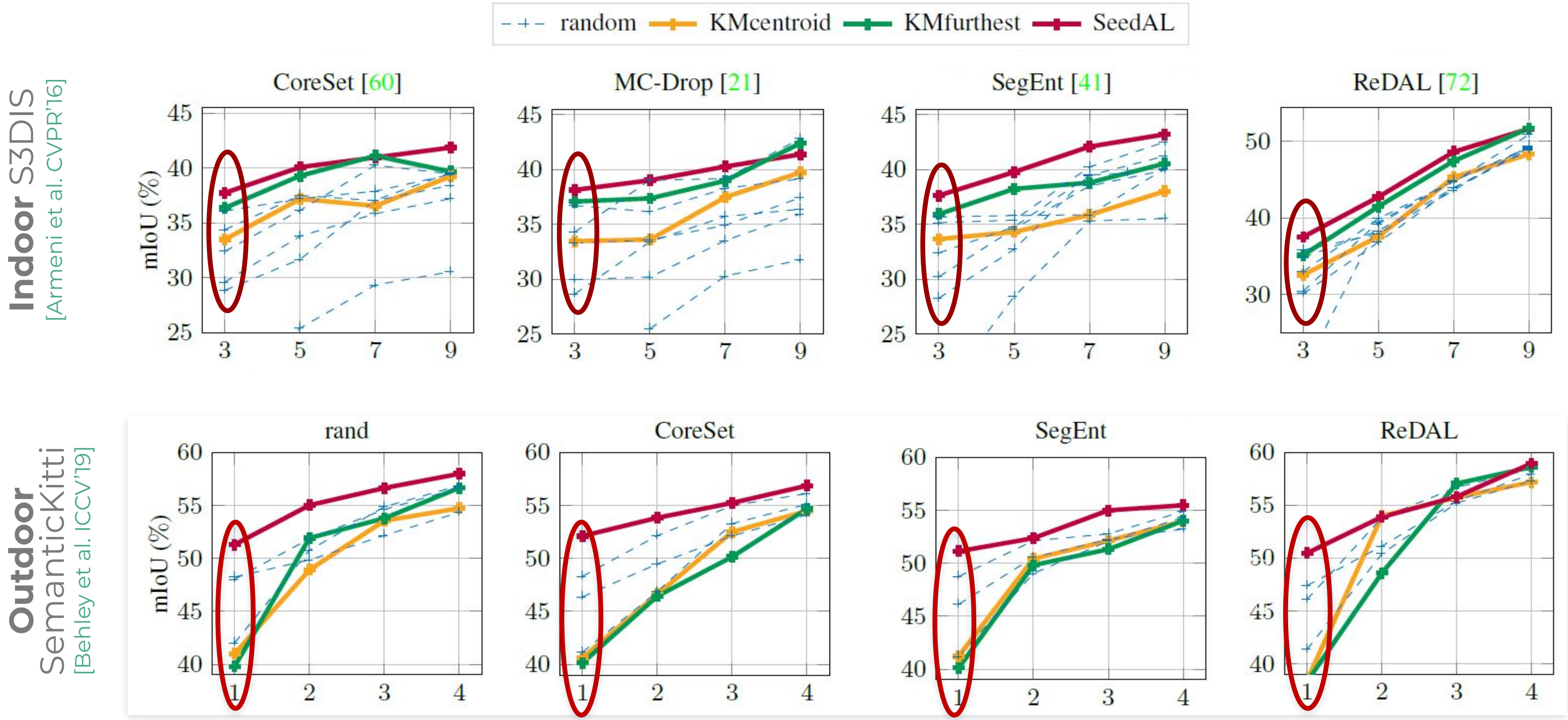
SeedAL: a *lucky* seed



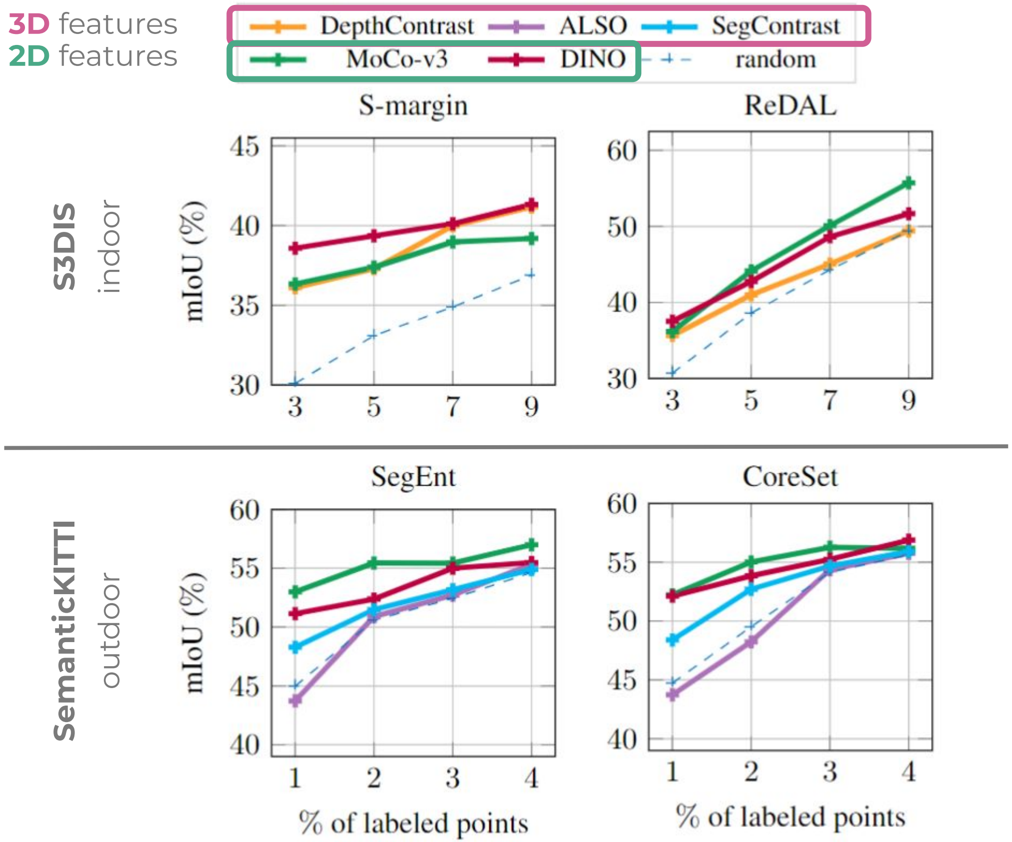
- High **variability** between different **draws**
- Variability with all SoTA AL methods
- The larger variability at first cycles

We **can** design the selection of a **lucky seed**

Comparison to k-means based baselines



Different features, same conclusions



Conclusions

- Active Learning methods mix data with high **informativeness** & **representativeness**
- AL strategy needs to be **adapted** depending on the task, e.g.
  - Finding a model's typical **mistake**
  - Adapting to the **structure** of the **data/task** (boxes / 3D data)
  - Possible to leverage 2D & 3D features
- The **first selection matters !**

Conclusions

- Active Learning methods mix data with high **informativeness** & **representativeness**
- AL strategy needs to be **adapted** depending on the task, e.g.
    - Finding a model's typical **mistake**
    - Adapting to the **structure** of the **data/task** (boxes / 3D data)
    - Possible to leverage 2D & 3D features
- The **first selection matters !**

Questions ?