Plant phenotyping using deep learning methods in a data-poor environment

Student: *Liliya Lemikhova*
Research Advisor: *Andrey Somov*

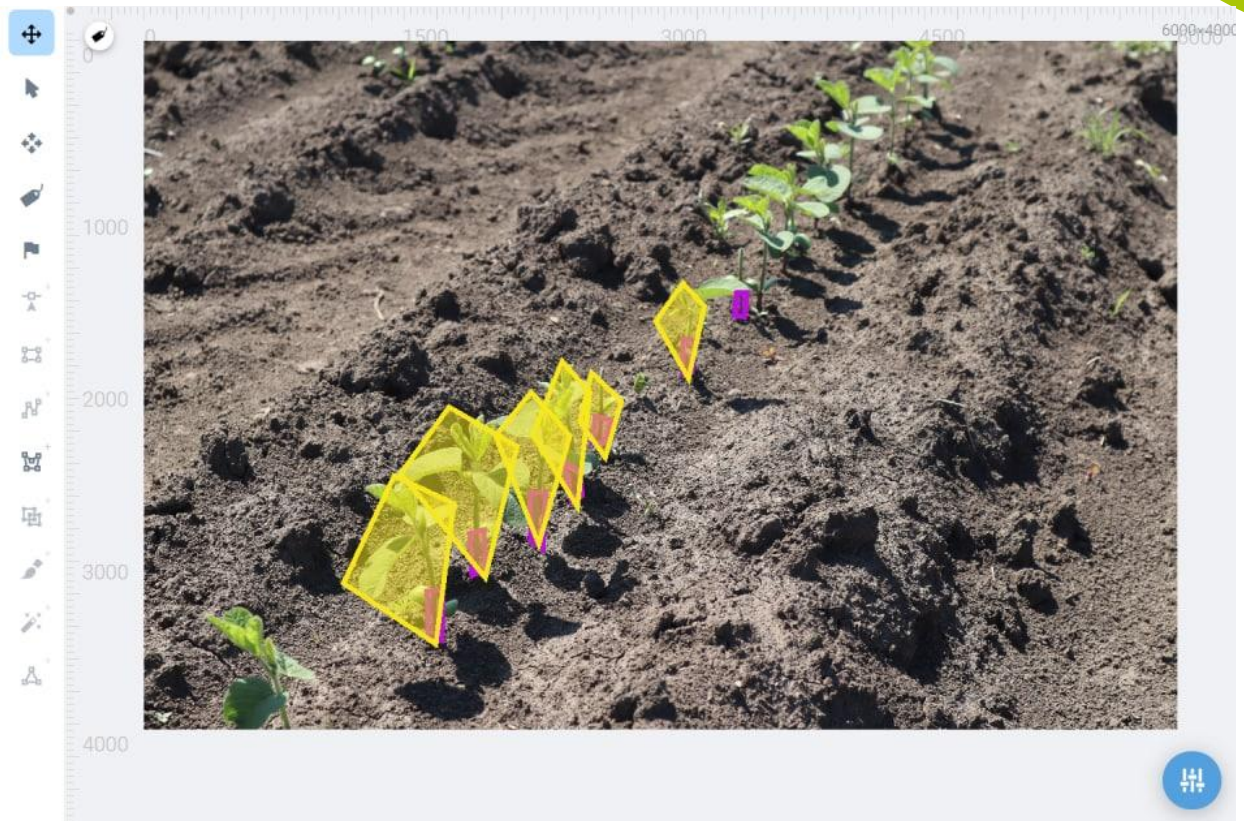
June, 2022

Skoltech

Background and problem statement

- summer experience with soy dataset
- data in agro is problematic to collect:
  - season dependent
  - places hard to reach
- and problematic to label:
  - small and similar objects
  - require expertise
- **need data-efficient strategies**

- results reported for datasets like Imagenet and Pascal VOC
- images less fine-grained
- **how new methods work in agro?**



Skoltech

Aim

To develop efficient methods for plant phenotyping in a data-poor environment for classification and object detection tasks.

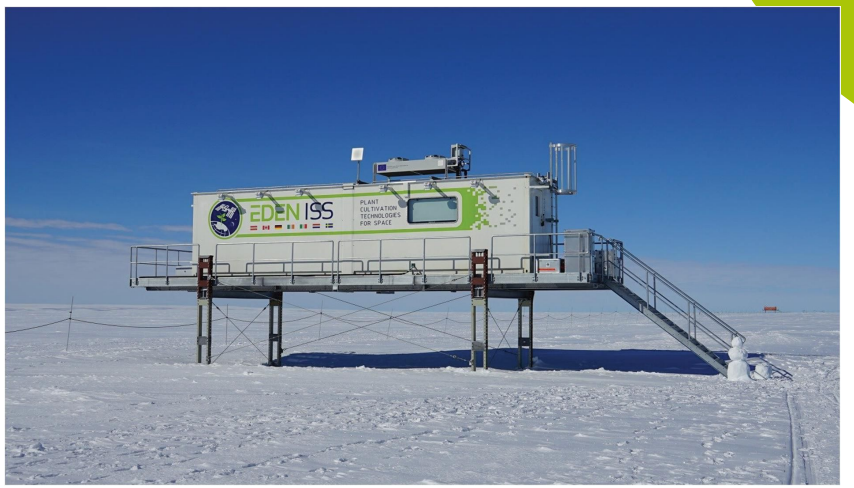
Objectives

Classification	Object Detection:

● zero-shot:
  ○ to run CLIP in zero-shot mode;
  ○ to find additional data;
● few-shot:
  ○ to test several pretraining methods.
	● to form dataset;
	● to run baseline experiment;
	● to run VOS experiment.

Skoltech

Classification



- Arctic station, almost no human control
- 16 plant classes and 1 class for empty slots
- 228 class-balanced images

Credits to German Aerospace Center

Results: zero-shot

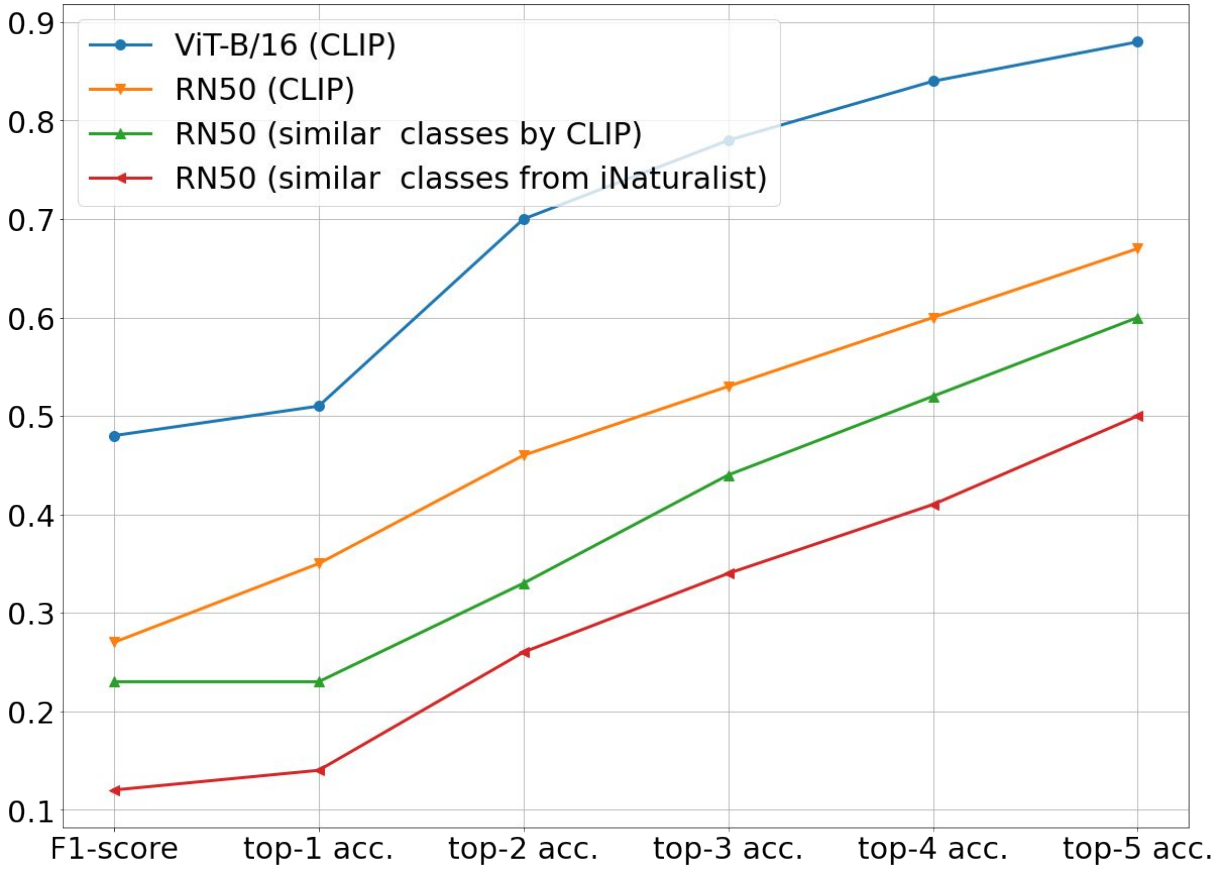
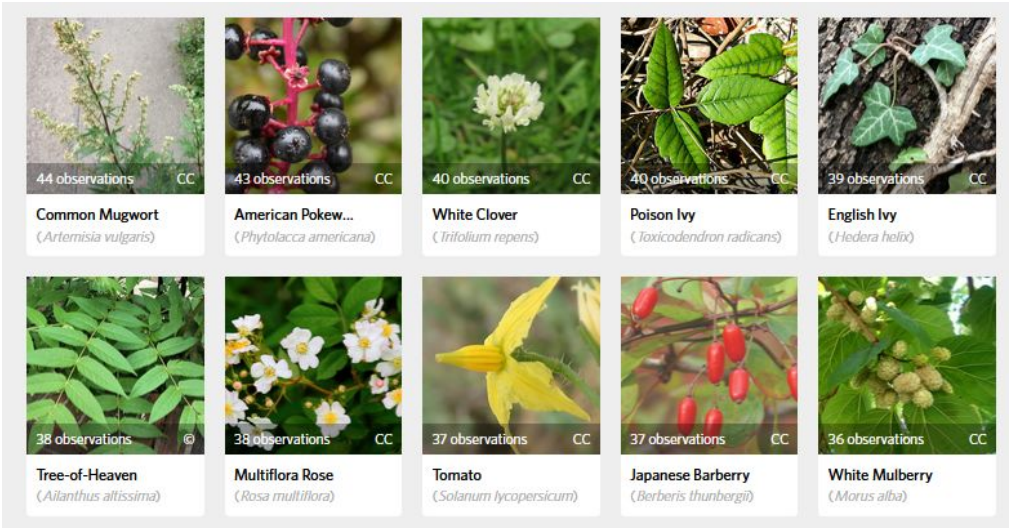


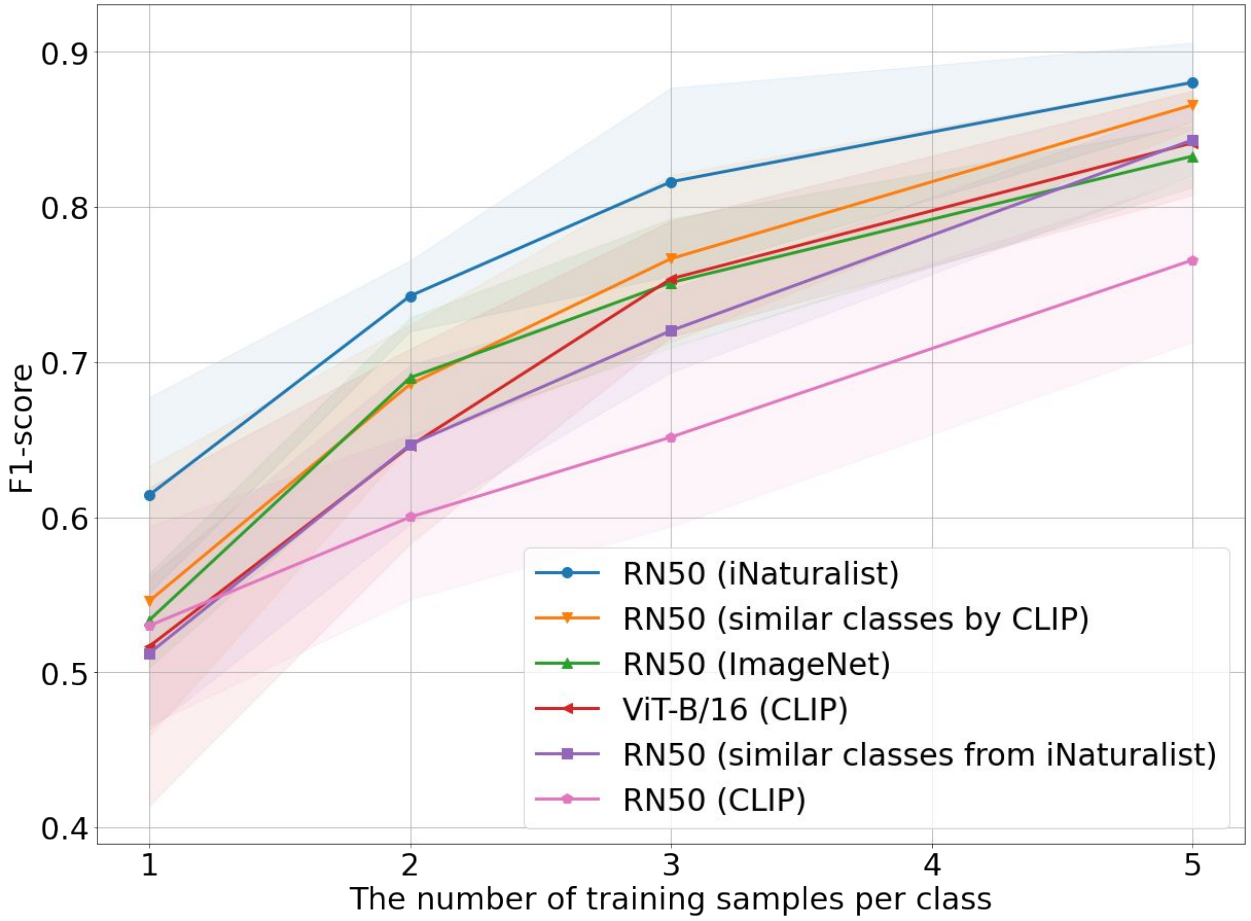
Image description for zero-shot CLIP:
“This is a photo of {name of plant} plants in greenhouse”

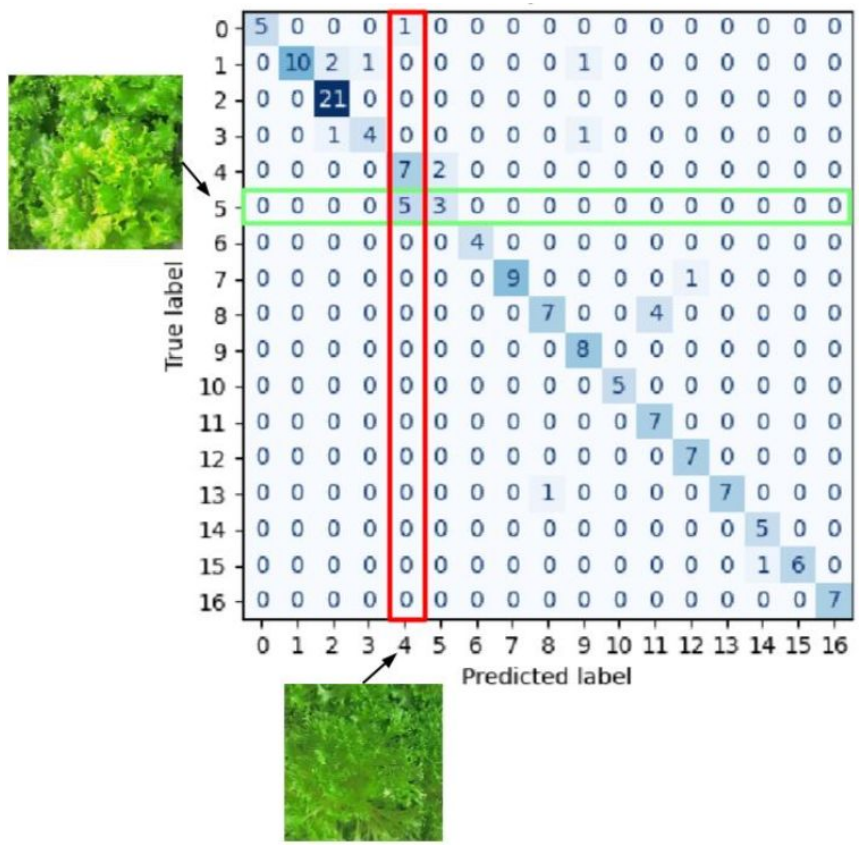


Examples from iNaturalist dataset

Skoltech

Results: few-shot





Confusion matrix of predictions by 5-shot CLIP ViT/16. The model confuses loose-leaf lettuce (label ”4”) and mizuna lettuce (label ”5”)

Skoltech

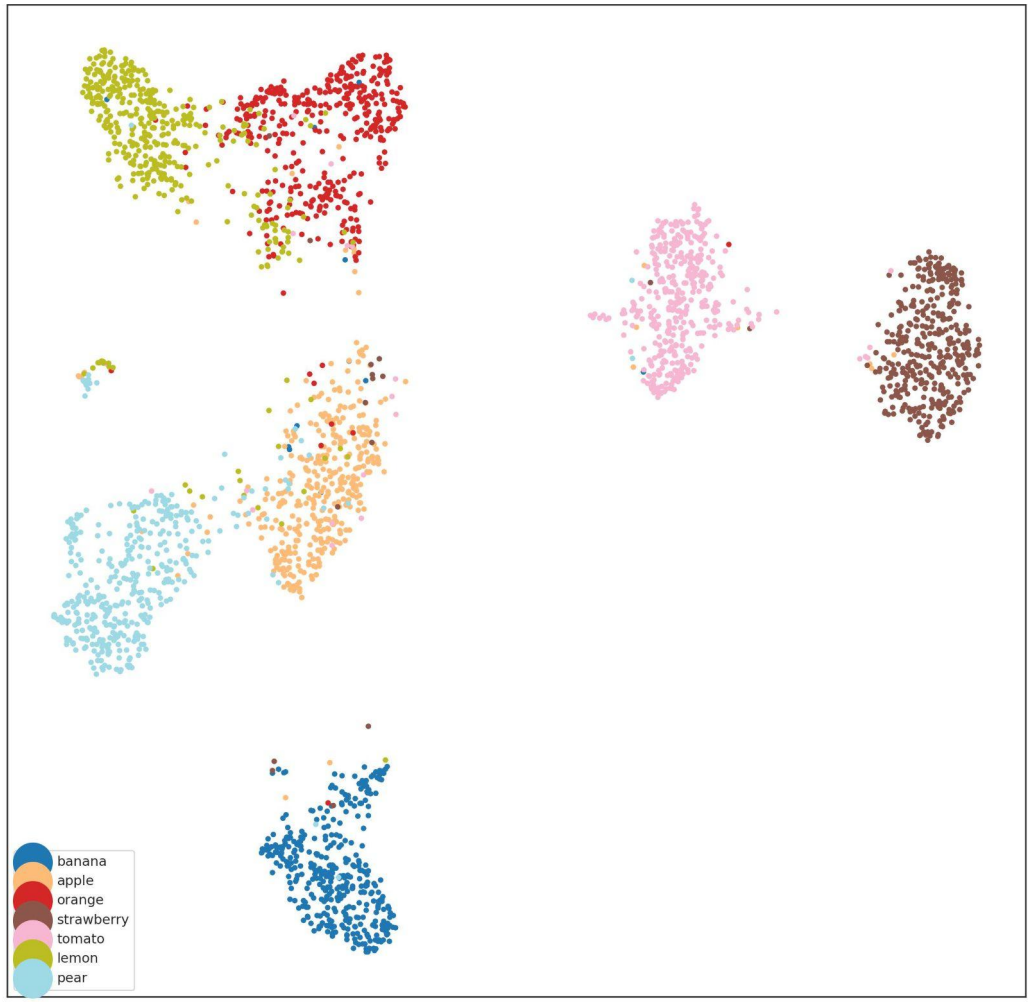
Discussion of results

Zero-shot:	**Few-shot:**
● CLIP ○ better than adding data; ○ worse for agro than for ImageNet; ○ sophisticated backbone helps; ○ needs accurate class names; ● Additional data ○ different domain is not helpful; ○ requires time to find relevant images.	● iNaturalist is better than ImageNet for pretraining; ● CLIP is worse, even with sophisticated backbone; ● two-stage finetune helps if similar domain.

Skoltech

Object detection

- Train dataset: fruits from COCO & Open Images
- chose 7/15 classes as ID, rest as OOD
- Target dataset: DeepFruits



Umap visualisation of embeddings of 400 samples for each ID class

Example images from Deep Fruits dataset

Skoltech

Methods

VOS (Virtual Outlier Synthesis)

- idea - sample outliers in embedding space

$$p_\theta(h(\mathbf{x},\mathbf{b})|y=k)=\mathcal{N}(\boldsymbol{\mu}_k,\boldsymbol{\Sigma})$$

- estimate mean and covariance during training

$$\widehat{\boldsymbol{\mu}}_k=\frac{1}{N_k}\sum_{i:y_i=k}h(\mathbf{x}_i,\mathbf{b}_i)$$

$$\widehat{\boldsymbol{\Sigma}}=\frac{1}{N}\sum_{k}\sum_{i:y_i=k}\left(h(\mathbf{x}_i,\mathbf{b}_i)-\widehat{\boldsymbol{\mu}}_k\right)\left(h(\mathbf{x}_i,\mathbf{b}_i)-\widehat{\boldsymbol{\mu}}_k\right)^\top$$

- sample outliers $\mathbf{v_k}$

$$\mathcal{V}_k=\{\mathbf{v}_k|\frac{1}{(2\pi)^{m/2}|\widehat{\boldsymbol{\Sigma}}|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{v}_k-\widehat{\boldsymbol{\mu}}_k)^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{v}_k-\widehat{\boldsymbol{\mu}}_k)\right)<\epsilon\}$$

$f(\mathbf{v};\theta)=W_{\text{cls}}^\top\mathbf{v}$ - output of classification branch

- log partition function:

$$F((\mathbf{x},\mathbf{b});\theta):=\log\sum_{k=1}^{K}e^{f_k((\mathbf{x},\mathbf{b});\theta)}$$

- loss:

$$\mathcal{L}_{\text{uncertainty}}=\mathbb{E}_{\mathbf{v}\sim\mathcal{V}}\mathbb{I}\{F(\mathbf{v};\theta)>0\}+\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}\mathbb{I}\{F(\mathbf{x};\theta)\leq 0\}$$

$$\min_\theta\mathbb{E}_{(\mathbf{x},\mathbf{b},y)\sim\mathcal{D}}\left[\mathcal{L}_{\text{cls}}+\mathcal{L}_{\text{loc}}\right]+\beta\cdot\mathcal{L}_{\text{uncertainty}}$$

- inference:

$$G\left(\mathbf{x}^*,\mathbf{b}^*\right)=\begin{cases}\text{ID}, & \text{if } F\geq\gamma\\ \text{OOD}, & \text{if } F<\gamma\end{cases}$$

Skoltech

ID

OOD

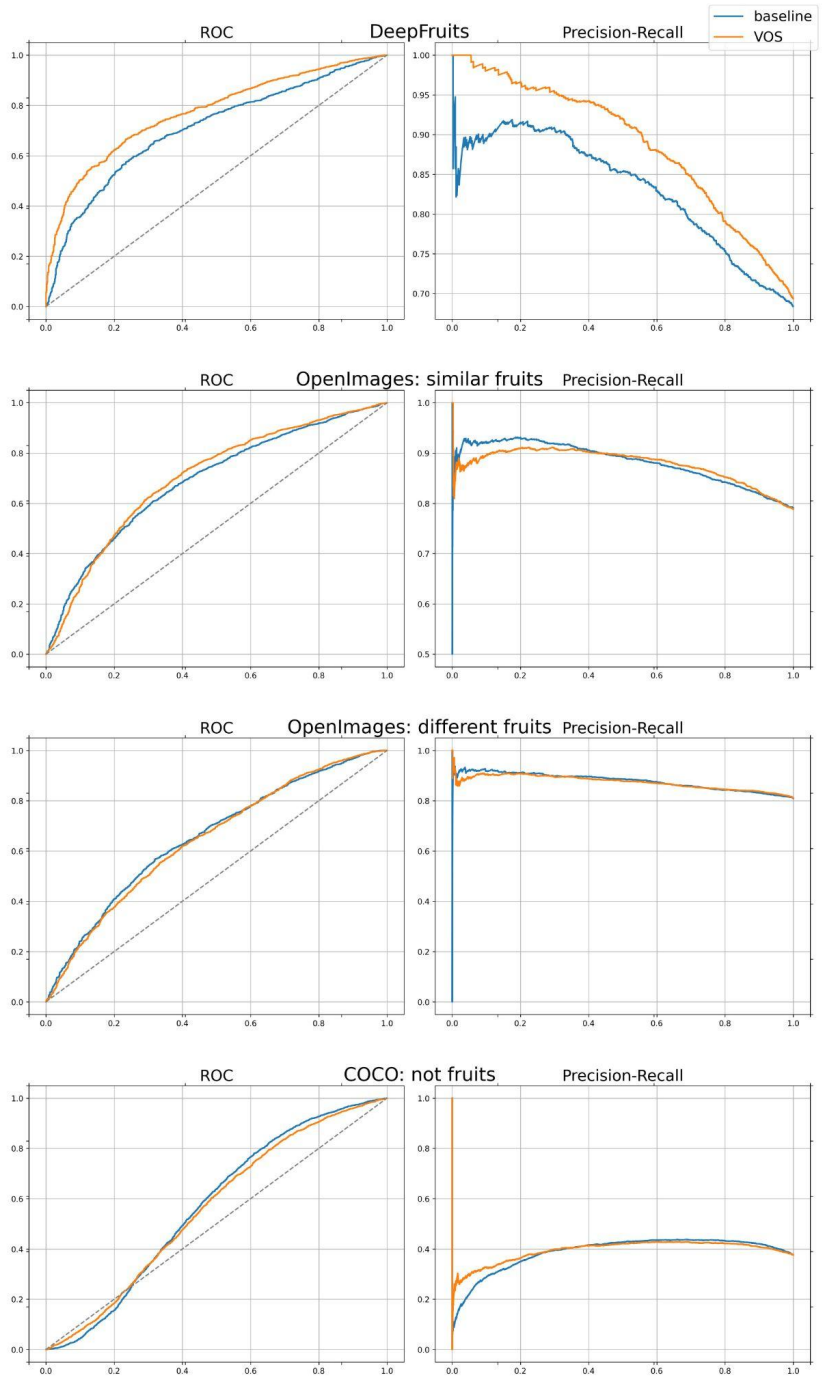
Comparison of baseline (left in pair of images) with VOS (right in pair) for ID and OOD example images

Skoltech

Results

base dataset	mAP_ID (w/o)/with VOS↑, %	OOD dataset	VOS	AUROC↑, %	AUPRC↑, %
DeepFruits	**78.07 / 78.07**	fruits_ood	w/o	70.76	83.22
			with	**77.20**	**88.83**
Open Images	57.30 / **58.14**	fruits_ood_sim	w/o	68.95	**88.18**
			with	**70.18**	87.79
		fruits_ood_diff	w/o	**65.57**	**87.79**
			with	64.72	87.36
COCO	**48.17** / 47.12	no_fruits	w/o	**56.96**	38.64
			with	56.28	**39.27**

Table 4.3: Comparative results of the models trained with and without VOS method for test fruits datasets made from COCO, Open Images and Deep Fruits images.



- improvement on DeepFruits
- no improvement on COCO and Open Images

Skoltech

Metrics limitations

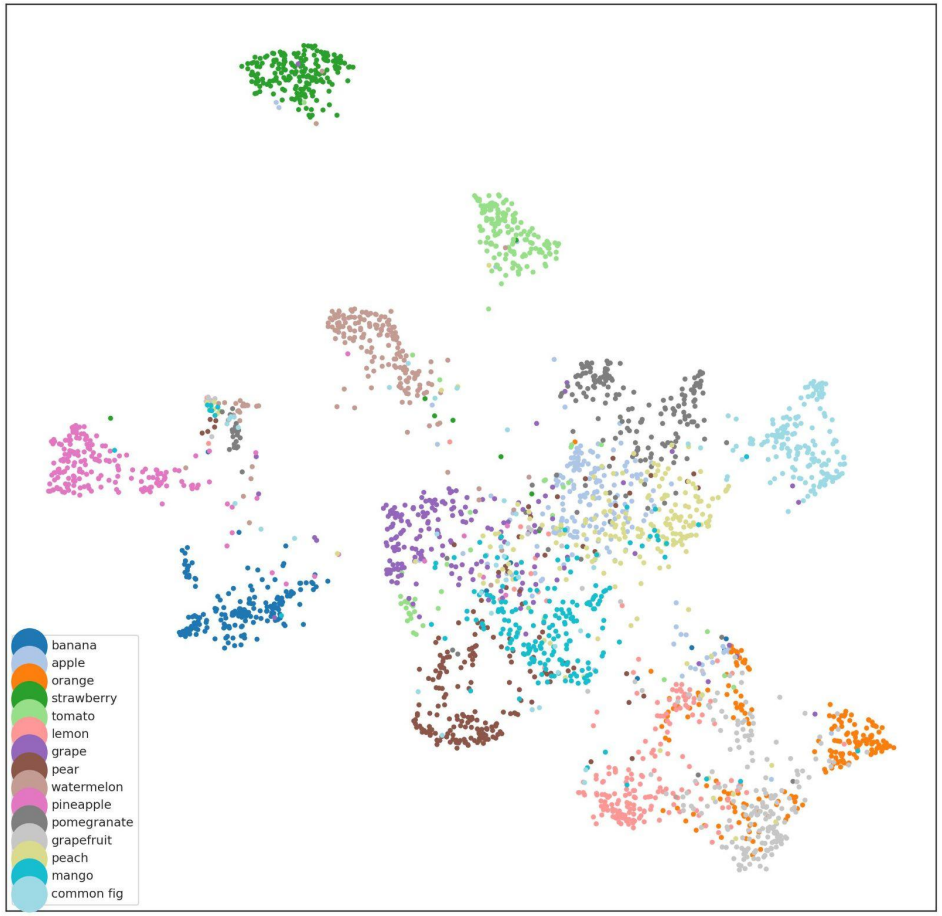
- “Preds from ID dataset - ID, from OOD - OOD”
- Problems when:
  - bad annotation; thus, ID objects in OOD;
  - FP for ID dataset.
- Especially relevant for small objects like in COCO
- Suggestions:
  - remove ID objects from OOD dataset;
  - reduce number of FPs;
  - ID are only TP boxes
    (need accurate annotation).

- suitable for DeepFruits
- not suitable for COCO

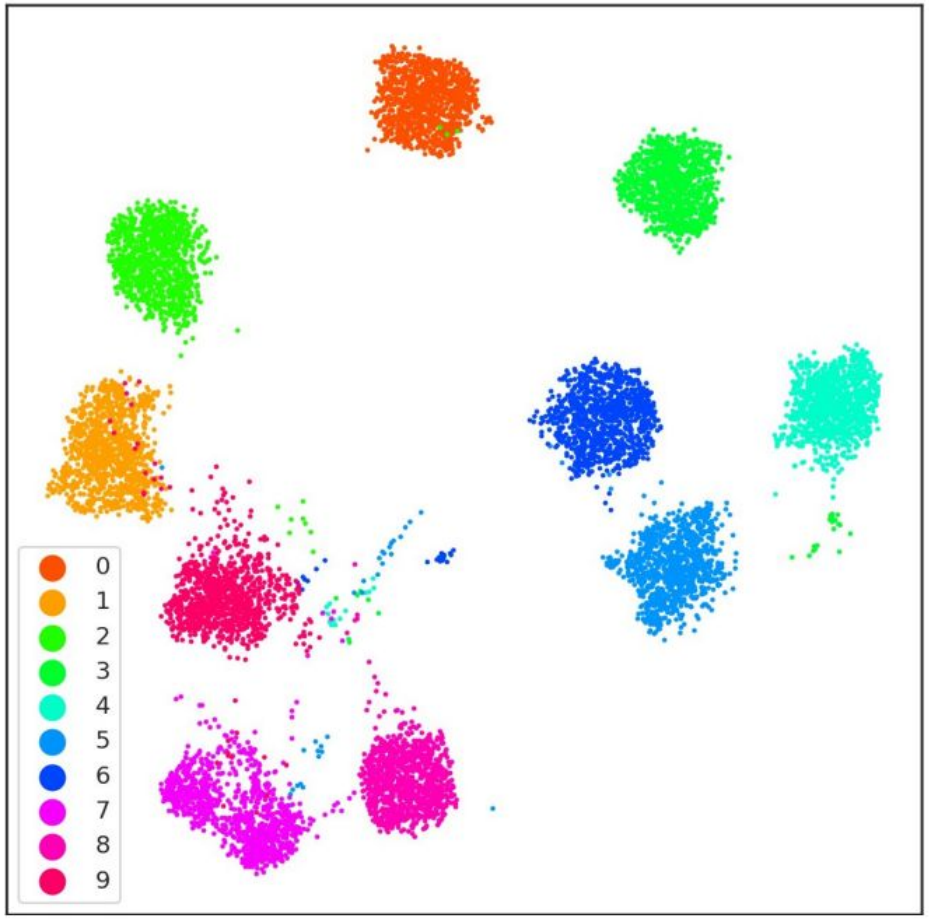
Examples from COCO ID and OOD datasets

Skoltech

Blended embedding distributions



Embeddings of fruit objects from Open Images



Embeddings of 10 classes from Pascal VOC [4]

Skoltech

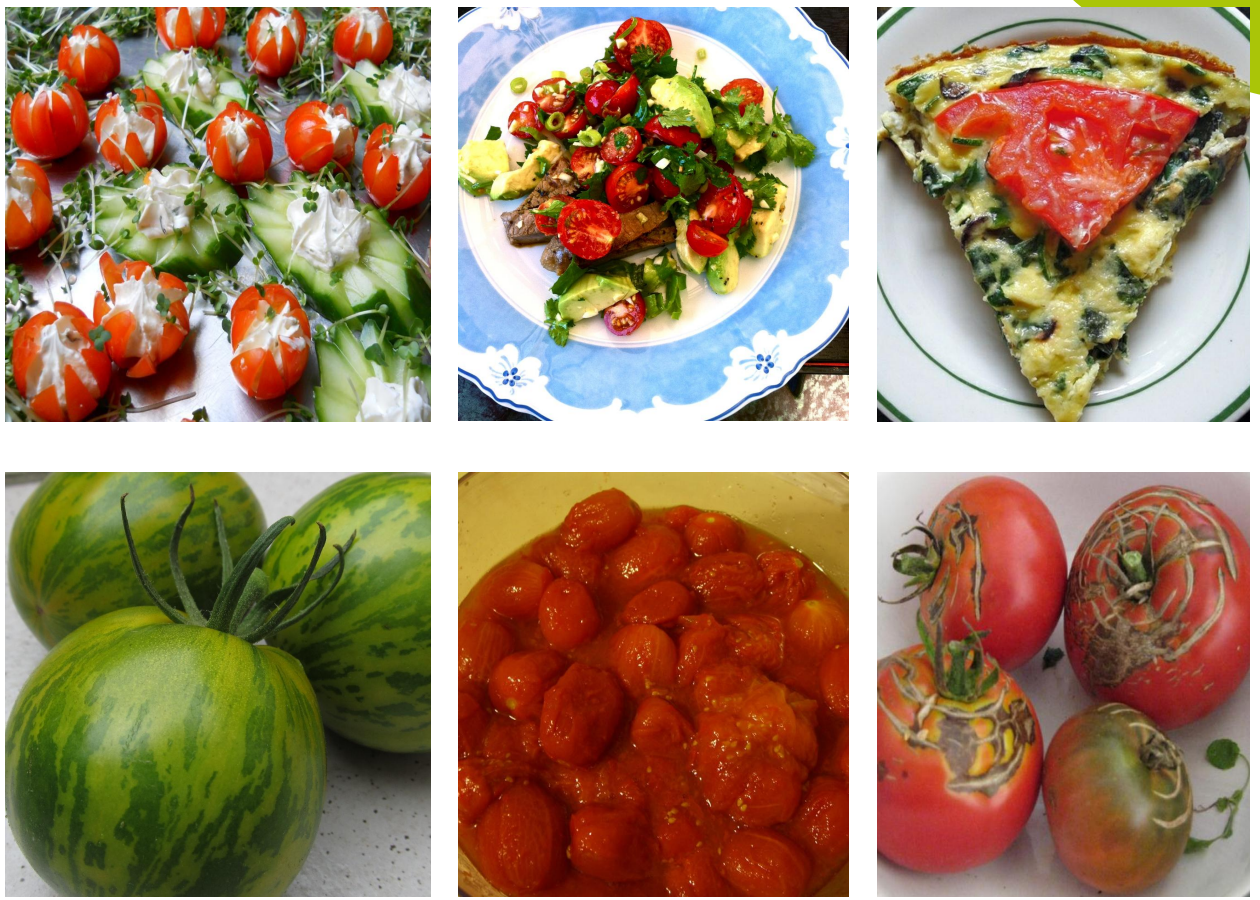
Various forms and little data

base dataset	type	train	val	test
Deep Fruits	ID	-	-	163
	OOD	-	-	166
Open Images	ID	3500	617	346
	OOD (sim)	-	-	166
	OOD (diff)	-	-	144
COCO	ID	2218	389	173
	OOD (not fruits)	-	-	4512

Table 4.2: Number of images in used datasets.

	Task 1	Task 2
ID train dataset	VOC train	BDD train
ID val dataset	VOC val	BDD val
OOD dataset	COCO and OpenImages val	COCO and OpenImages val
#ID train images	16,551	69,853
#ID val images	4,952	10,000
#OOD images for COCO	930	1,880
#OOD images for OpenImages	1,761	1,761

Data from paper introducing VOS method [4]



Example images of class “tomato” in Open Images testset

Skoltech

Scientific novelty

Classification:	Object detection:
• zero-shot Clip in agro; • transfer ability of iNaturalist in agro.	• VOS for similar objects and domain; • methods limitations.

Skoltech

Innovation

Possible industrial application:

- German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt e.V.)
- SOCO - UAV for agriculture company

Skoltech

Conclusions

Classification:	Object detection:
Zero-shot: ● CLIP in zero-shot mode; ● relative data from iNaturalist; and from CLIP training set. Few-shot: ● ResNet50 pretrained: ○ on ImageNet; ○ on iNaturalist; ○ with CLIP; ● ViT-B pretrained with CLIP; ● ResNet50 fine-tuned in two stages.	● formed training dataset; ● ran baseline experiment; ● ran VOS experiment; ● analyzed limitations: ○ metrics; ○ blended embeddings; ○ various forms; ○ little data.

Skoltech

Outcomes

Liliya Lemikhova, Sergey Nesteruk, and Andrey Somov. Transfer learning for few-shot plants recognition: Antarctic station greenhouse use-case. In *International Symposium on Industrial Electronics*, Anchorage, Alaska, USA, June 2022.

Skoltech

Outlook

Classification:	Object detection:

● zero-shot:
  ○ train CLIP on a bigger dataset.
● few-shot
  ○ other pretraining (SSL);
  ○ bigger dataset.
	● remove ID objects from OOD dataset; with neural network;
	● train with diverse negative examples;
	● use OOD classes in train;
	● zero-shot CLIP on OOD crops.

Skoltech

Acknowledgements

- Sergey Nesteruk, grad student at Skoltech;
- Andrey Somov, research advisor;
- German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt e.V.).

References

[1] EDEN ISS
[2] CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture
[3] Big Transfer (BiT): General Visual Representation Learning
[4] VOS: Learning What You Don't Know by Virtual Outlier Synthesis
[5]S3FD: Single Shot Scale-invariant Face Detector

Skoltech

thx.

Skoltech

External reviewer’s questions

1/3. As for few-shot learning case it is not clear how 228 pics from greenhouse station were used: what is the strategy for splitting data for train/test sets, whether different exps uses k different pics, whether testset is always fixed, what's its size?

- For each k in {1, 2, 3, 5}, we ran 5 experiments with different train/test splits to estimated mean and variance.
- Chose randomly k pictures of each class for train set and used the rest of data as test set.

Skoltech

External reviewer’s questions

2/3. You found out that in OOD COCO and OpenImages datasets many ID objects are present, which decrease target metrics. You propose to examine it manually, what are other possible solutions to that problem?

- Use more accurate annotation like LVIS for COCO dataset;
- Run SOTA model pretrained on dataset with these ID objects and examine manually only images with predictions of ID classes;
- Change target metrics:
  - ID == TP (IoU with gt > t) and OOD == FP (IoU with gt < t);
  - additional OOD predictions from ID dataset; so, you need less data to examine in OOD dataset;
  - need accurate annotation for ID dataset.

Skoltech

External reviewer’s questions

3/3. Also, as we have FP not only on ID objects, you explain it by lack of common sense, is it possible to overcome this problem too?

- Add images from a big diverse dataset such as COCO with no ID or OOD fruits objects during train;
- As there are many FP on small objects, use techniques to deal with small objects:
  - images during train of size smaller than during test;
  - specific framework [5];
  - etc.

Methods: zero-shot

zero-shot - no data from target dataset during training stage

CLIP

- 400 million (image, text) pairs
- trained to predict caption to image
- can predict the most relevant text snippet, given an image
- without optimization for the task

Image description for zero-shot CLIP:
“This is a photo of {name of plant} plants in greenhouse”

Skoltech

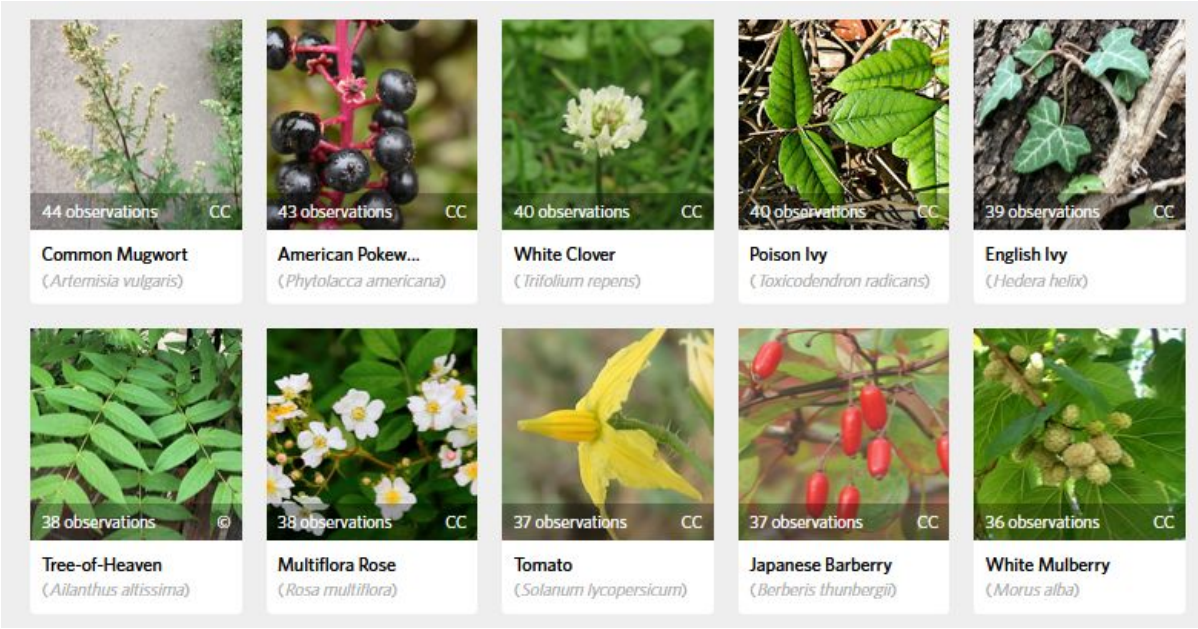
Methods: zero-shot

Additional data of same classes

- images by caption from CLIP training dataset
  - "{name of class} plants in greenhouse"
  - 100 pics/class closest to description
- images from iNaturalist
  - same plant family
  - 50 pics/class

Training:

- freeze parameters of ResNet50
- change number of output classes
- fine-tune last 2 FC layers



Examples from iNaturalist dataset

Skoltech

Methods: few-shot

few-shot - few samples of each class during training stage

Models

- ResNet50 pretrained on ImageNet
- ResNet50 pretrained on iNaturalist
- ResNet50 pretrained with CLIP
- ViT-B pretrained with CLIP
- ResNet50 fine-tuned for zero-shot
  - data by CLIP
  - data from iNaturalist

Training:

- freeze parameters
- change number of output classes
- fine-tune last FC layer
- k in {1, 2, 3, 5}
- 5 different dataset splits

Skoltech

Framework of VOS

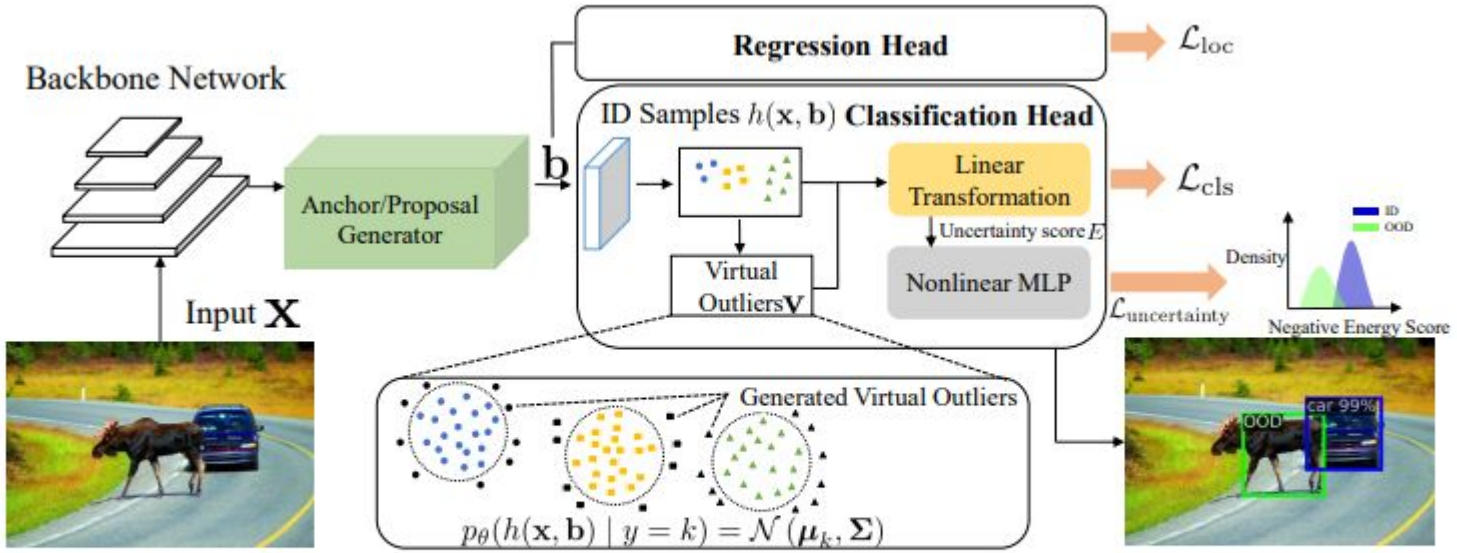
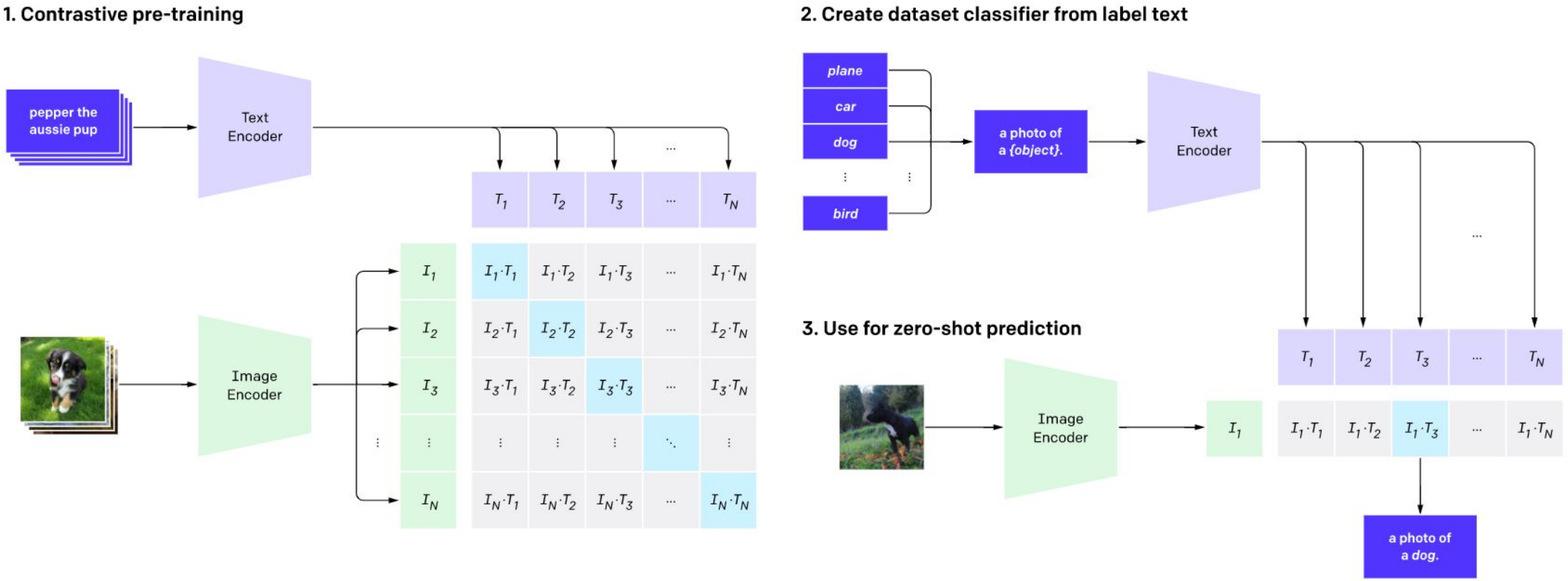


Figure 2: The framework of VOS. We model the feature representation of ID objects as class-conditional Gaussians, and sample virtual outliers $\mathbf{v}$ from the low-likelihood region. The virtual outliers, along with the ID objects, are used to produce the uncertainty loss for regularization. The uncertainty estimation branch ($\mathcal{L}_{\text{uncertainty}}$) is jointly trained with the object detection loss ($\mathcal{L}_{\text{loc}}, \mathcal{L}_{\text{cls}}$).

Skoltech

Framework of CLIP



CLIP pre-trains an image encoder and a text encoder to predict which images were paired with which texts in our dataset. We then use this behavior to turn CLIP into a zero-shot classifier. We convert all of a dataset's classes into captions such as "a photo of a *dog*" and predict the class of the caption CLIP estimates best pairs with a given image.

Skoltech

Data

base dataset	type	train	val	test
Deep Fruits	ID	-	-	163
	OOD	-	-	166
Open Images	ID	3500	617	346
	OOD (sim)	-	-	166
	OOD (diff)	-	-	144
COCO	ID	2218	389	173
	OOD (not fruits)	-	-	4512

Table 4.2: Number of images in used datasets.

	Task 1	Task 2
ID train dataset	VOC train	BDD train
ID val dataset	VOC val	BDD val
OOD dataset	COCO and OpenImages val	COCO and OpenImages val
#ID train images	16,551	69,853
#ID val images	4,952	10,000
#OOD images for COCO	930	1,880
#OOD images for OpenImages	1,761	1,761

classname	openim train	openim test	coco train	coco test	deep fruits	type
orange	4217	139	11477	1116	274	ID
banana	714	11	39299	4619	-	
apple	2023	87	15329	1701	359	
strawberry	6251	253	3553	484	348	
tomato	4263	315	10874	1312	-	
lemon	1357	127	1886	235	-	
pear	678	10	988	148	-	
grape	198	70	4269	216	-	OOD (sim)
grapefruit	311	14	-	-	-	
peach	289	35	889	155	-	
mango	211	11	-	-	401	
watermelon	611	26	704	52	-	OOD (diff)
pineapple	532	17	1530	169	-	
pomegranate	341	16	-	-	-	
common fig	208	22	-	-	-	
cantaloupe	108	13	-	-	-	
avocado	-	-	986	123	178	
rockmelon	-	-	-	-	137	

Table 4.1: Number of bounding boxes and its type for each fruits class fruits datasets.

Data from paper introducing VOS method

Skoltech

Strategy suggestions

Classification:

Zero-shot:

- if no features - Clip works best
- use Clip trained on bigger dataset
- adding close images from wild domain not effective

Few-shot:

- if k small - use iNat as pretrain
- if have resources - try iNat21 with sophisticated backbone

Object detection:

- if have no data for OOD classes - try VOS
- VOS is not enough for common sense
- try adding open dataset like COCO during train
- use accurate annotation like LVIS for COCO
- or manually remove OOD test images with ID objects
- test for a normal distribution
- correctly define threshold:
  - TPs - ID boxes,
    use IoA instead of IoU because of group ann.
  - FPs - OOD boxes

Skoltech