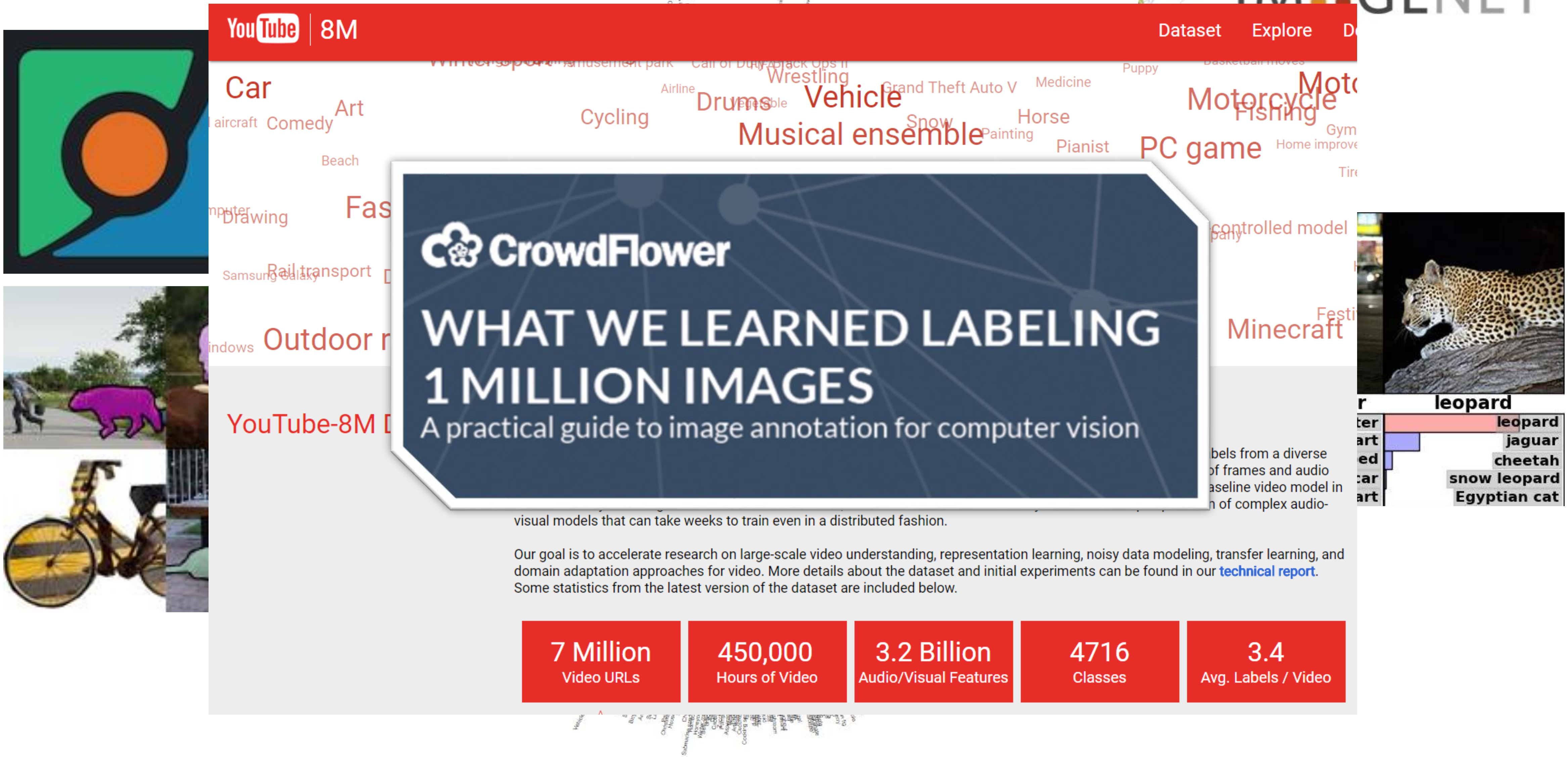
One-shot object X

Leonid Karlinsky

ORDL @ CVAR @ HRL @ IBM Research AI

Haifa Research Labs

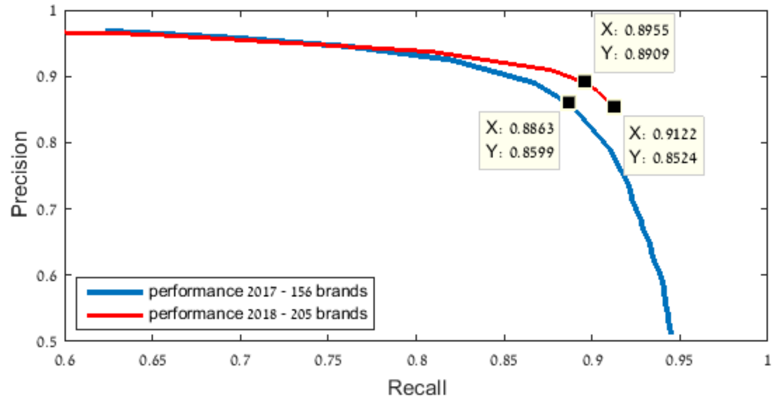
The age of big data(sets)

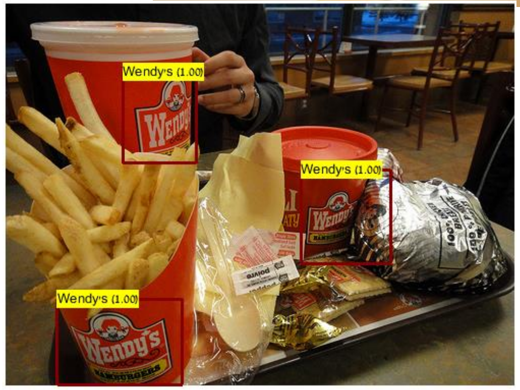


In the land of big data?

Logo Localization

Over **90% Precision** at **90% Recall** on 204 brands

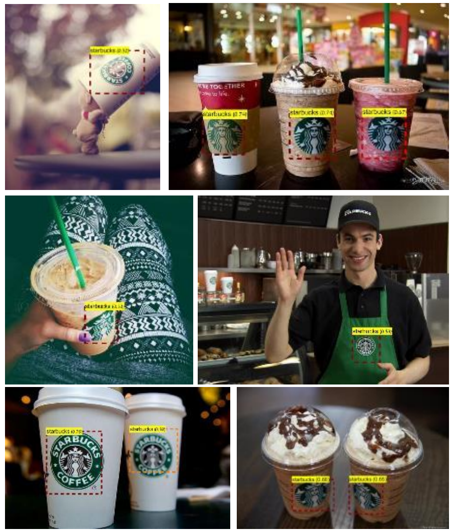




Land of little data – use cases …

Brand Logos

new brands can be added on the fly

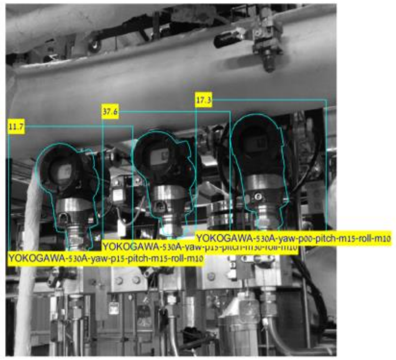


with just one or two examples

Food

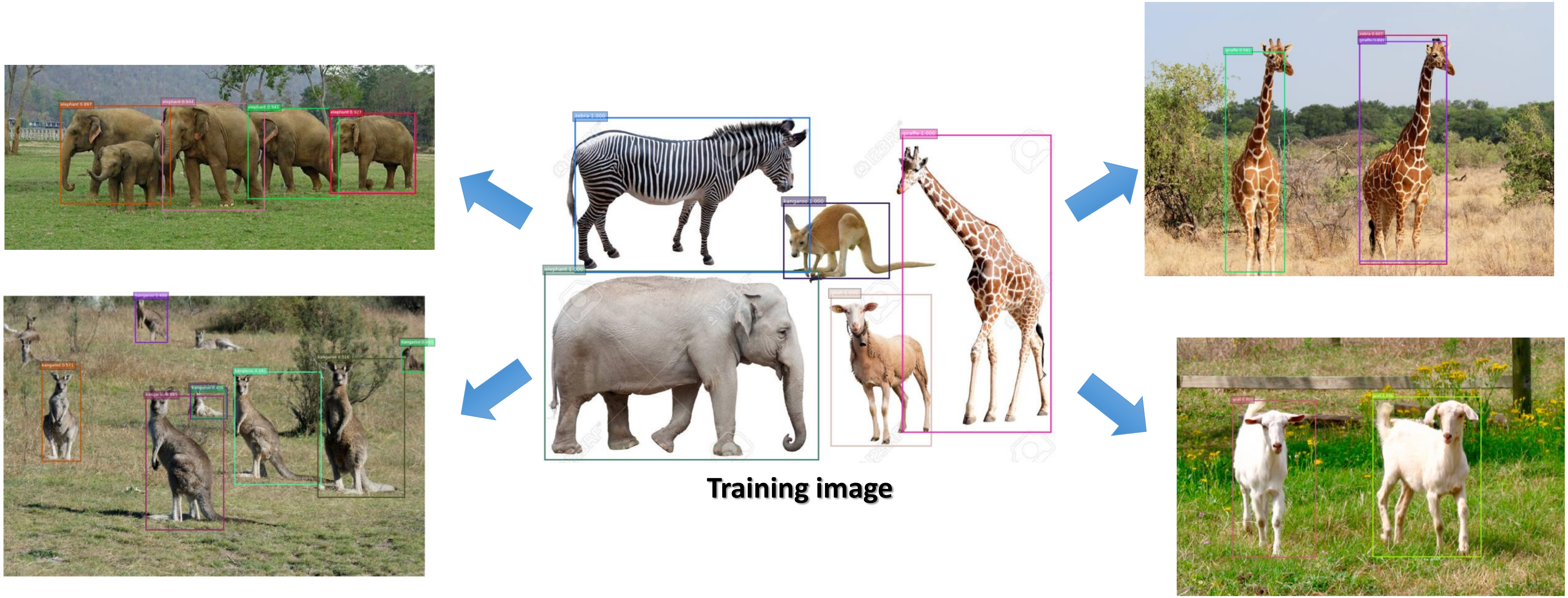
Retail products

Industrial



In the land of little data

- **Cool goal:** to be able to train visual object X with one or few samples (and online)

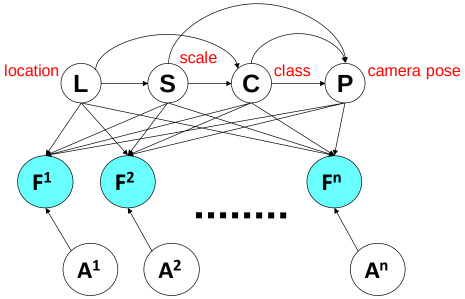


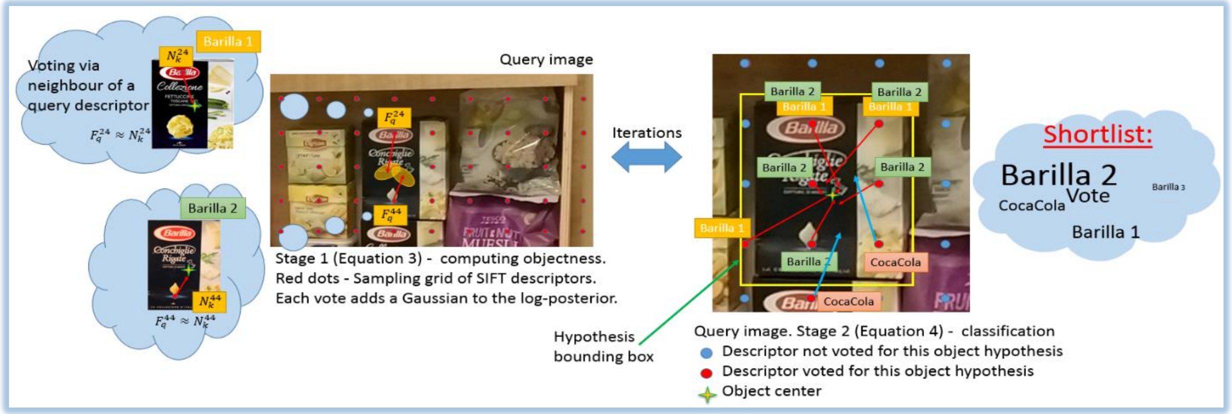
Training image

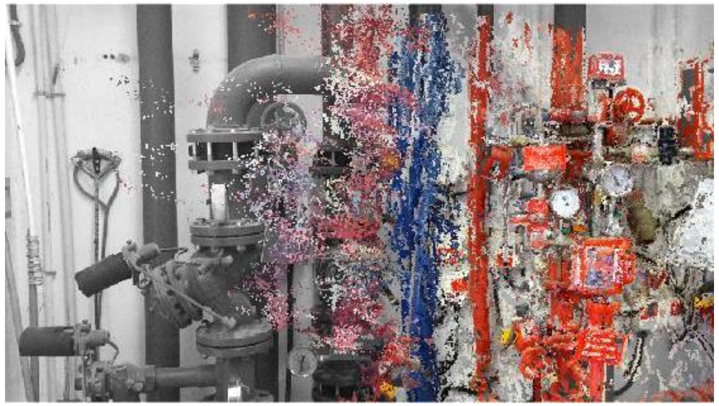
How we used to do it few years ago…

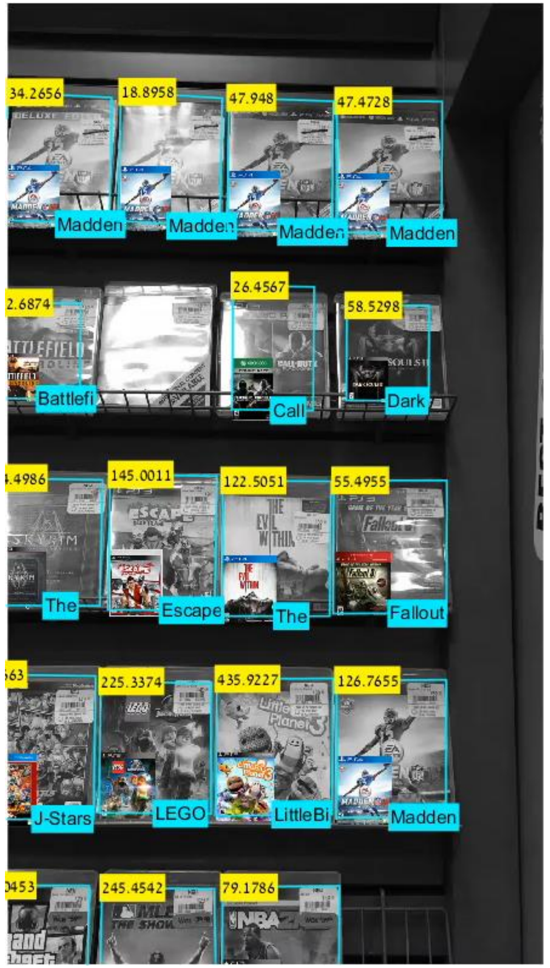
Fine-Grained Recognition of Thousands of Object Categories with Single-Example Training, CVPR 2017

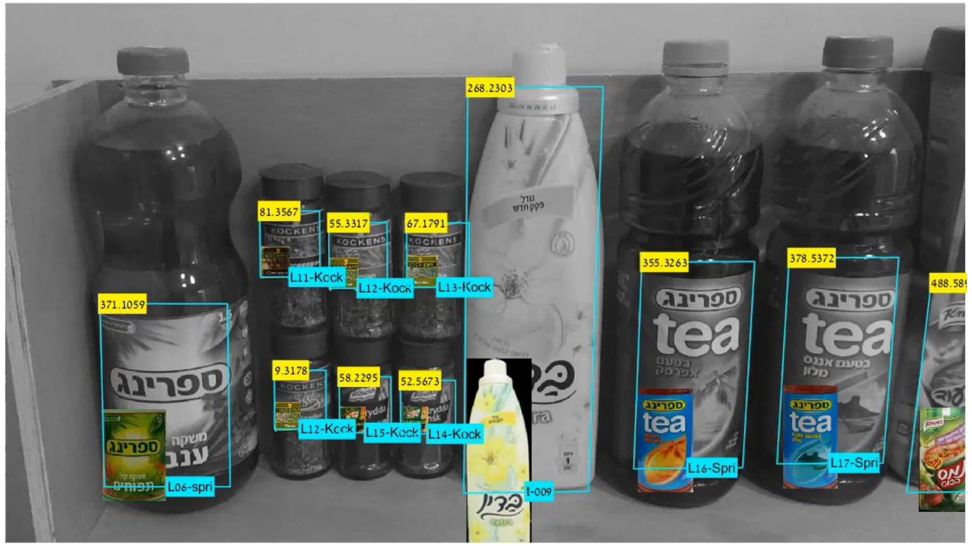
Leonid Karlinsky, Joseph Shtok, Yochay Tzur, Asaf Tzadok











and this is how we do it now…

RepMet: Representative-based metric learning for classification and one-shot object detection

Leonid Karlinsky*, Joseph Shtok*, Sivan Harary*, Eli Schwartz*,
Amit Aides, Rogerio Feris,
Raja Giryes, Alex M. Bronstein

CVPR 2019

LaSO: Label-Set Operations network for multi-label few-shot classification

Amit Alfassy*, Leonid Karlinsky*, Amit Aides*,
Joseph Shtok, Sivan Harary
Rogerio Feris, Raja Giryes, Alex M. Bronstein

CVPR 2019

Δ-encoder: an effective sample synthesis method for few-shot object recognition

Eli Schwartz*, Leonid Karlinsky*,
Joseph Shtok, Sivan Harary, Mattias Marder, Abhishek Kumar,
Rogerio Feris, Raja Giryes, Alex M. Bronstein

NeurIPS 2018

Δ-encoder: an effective sample synthesis method for few-shot object recognition

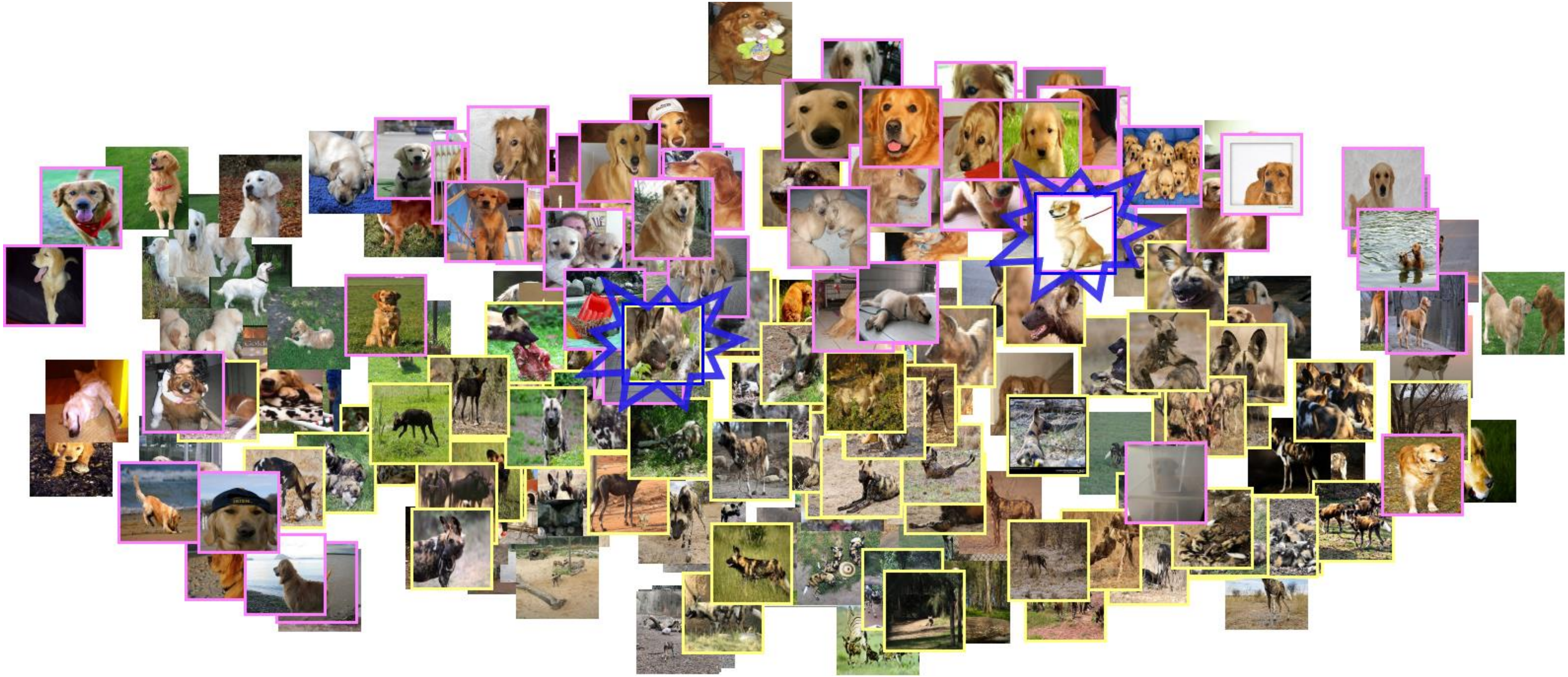
Eli Schwartz*, Leonid Karlinsky*,

Joseph Shtok, Sivan Harary, Mattias Marder, Abhishek Kumar,

Rogerio Feris, Raja Giryes, Alex M. Bronstein

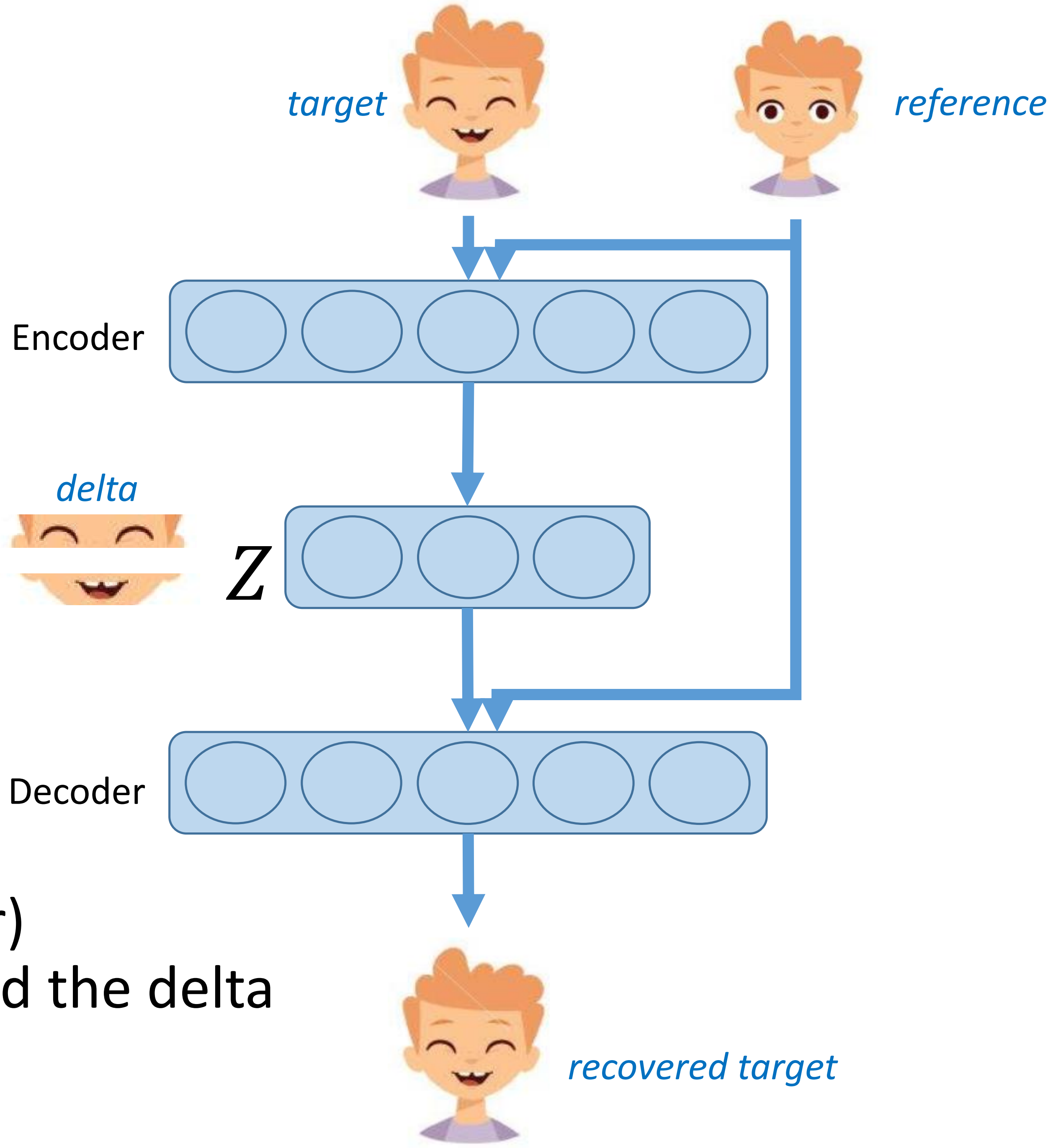
NeurIPS 2018

Who’s that dog?



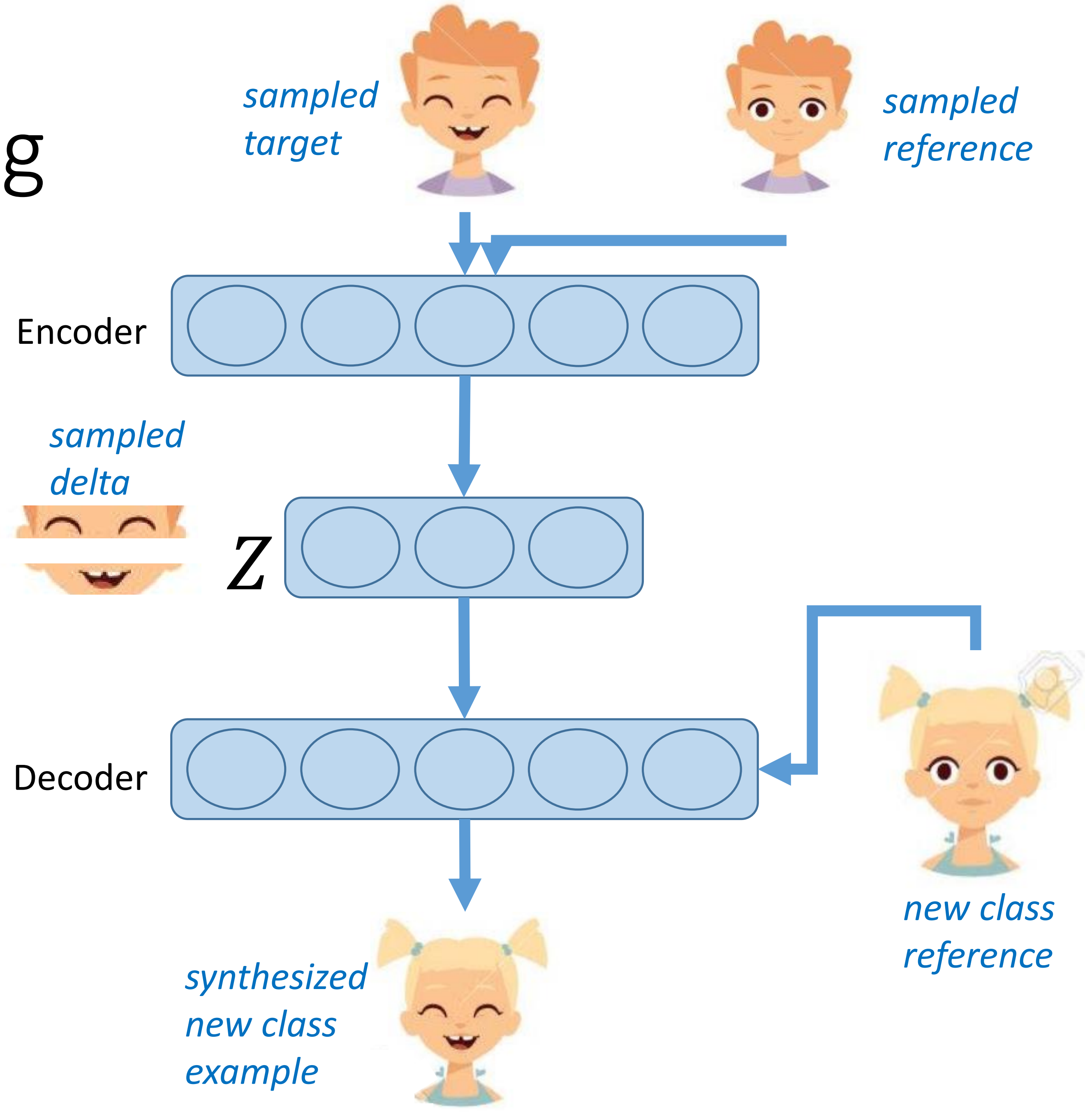
Key idea – training

- The model is a variant of an auto-encoder operating in feature space
- The network learns to encode the delta between the reference and the target image
- This delta is used to recover the target image as a (non-linear) combination of the reference and the delta

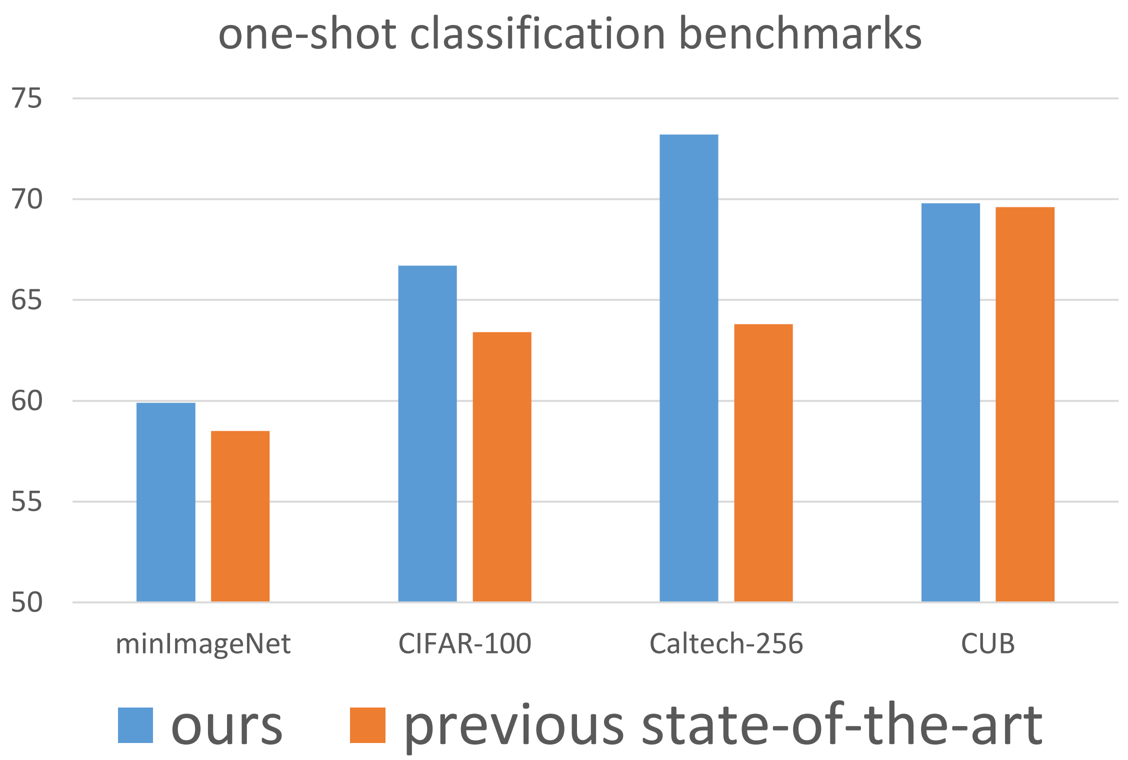


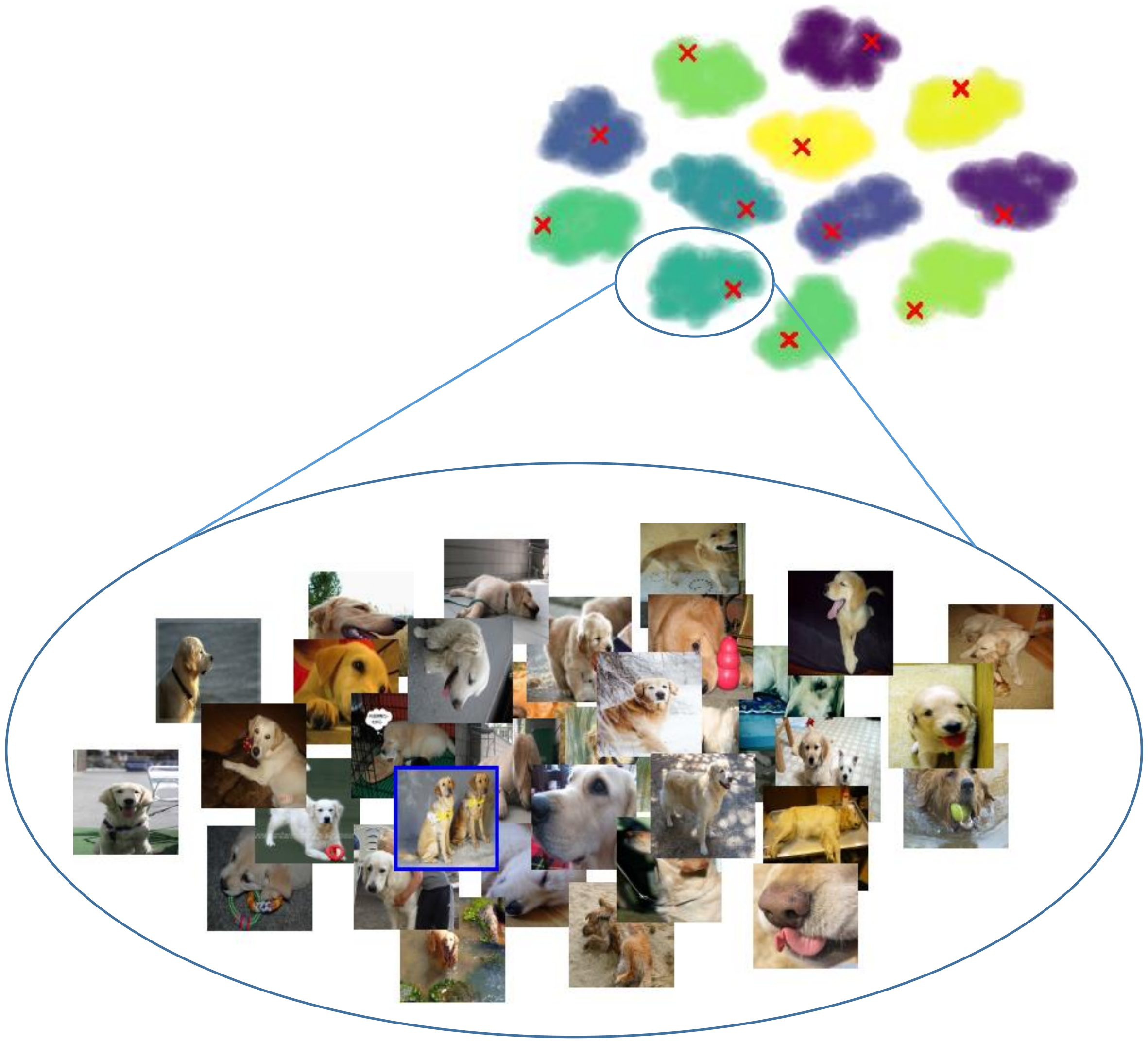
Key idea – synthesizing

- At test time we sample encoded deltas from random training image pairs
- The sampled deltas are used to create samples for new classes by combining them with the new class reference examples



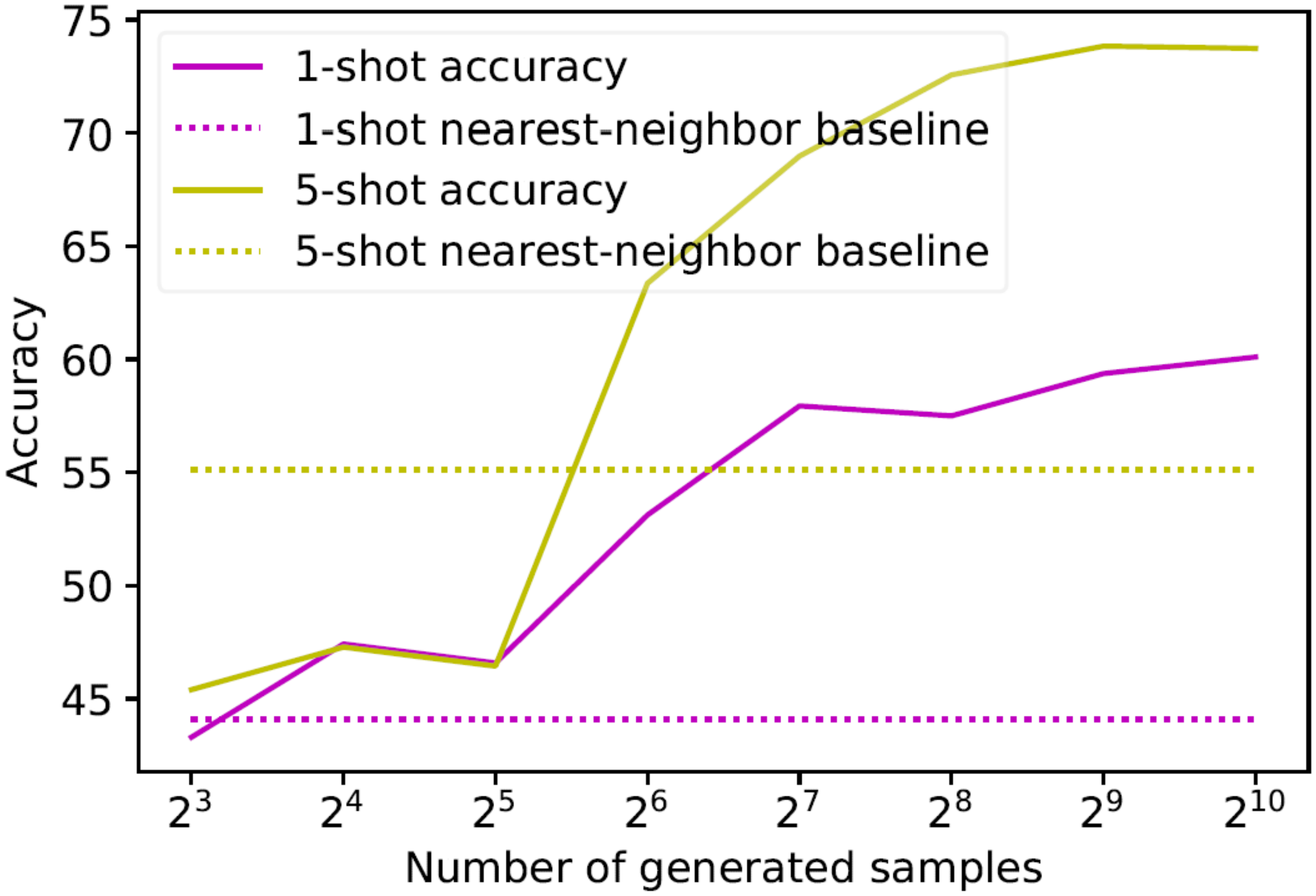
Few-shot classification experiments





miniImageNet: 58.5 (previous SOA) → 59.9 (ours)
CIFAR-100: 63.4 (previous SOA) → 66.7 (ours)
Caltech-256: 63.8 (previous SOA) → 73.2 (ours)
CUB: 69.6 (previous SOA) → 69.8 (ours)

Real vs synthetic examples ablation study



RepMet: Representative-based metric learning for classification and one-shot object detection

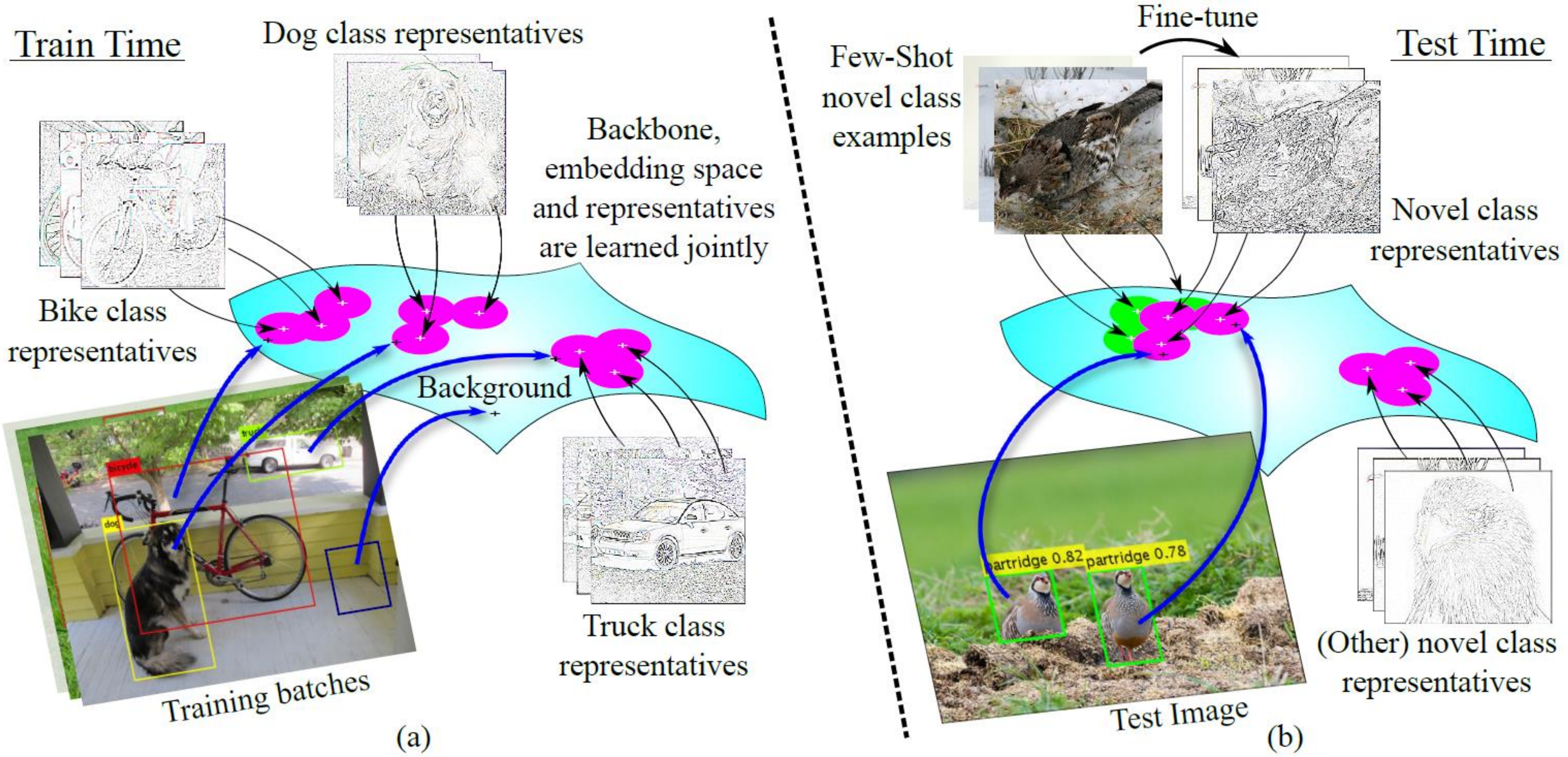
Leonid Karlinsky*, Joseph Shtok*, Sivan Harary*, Eli Schwartz*,

Amit Aides, Rogerio Feris,

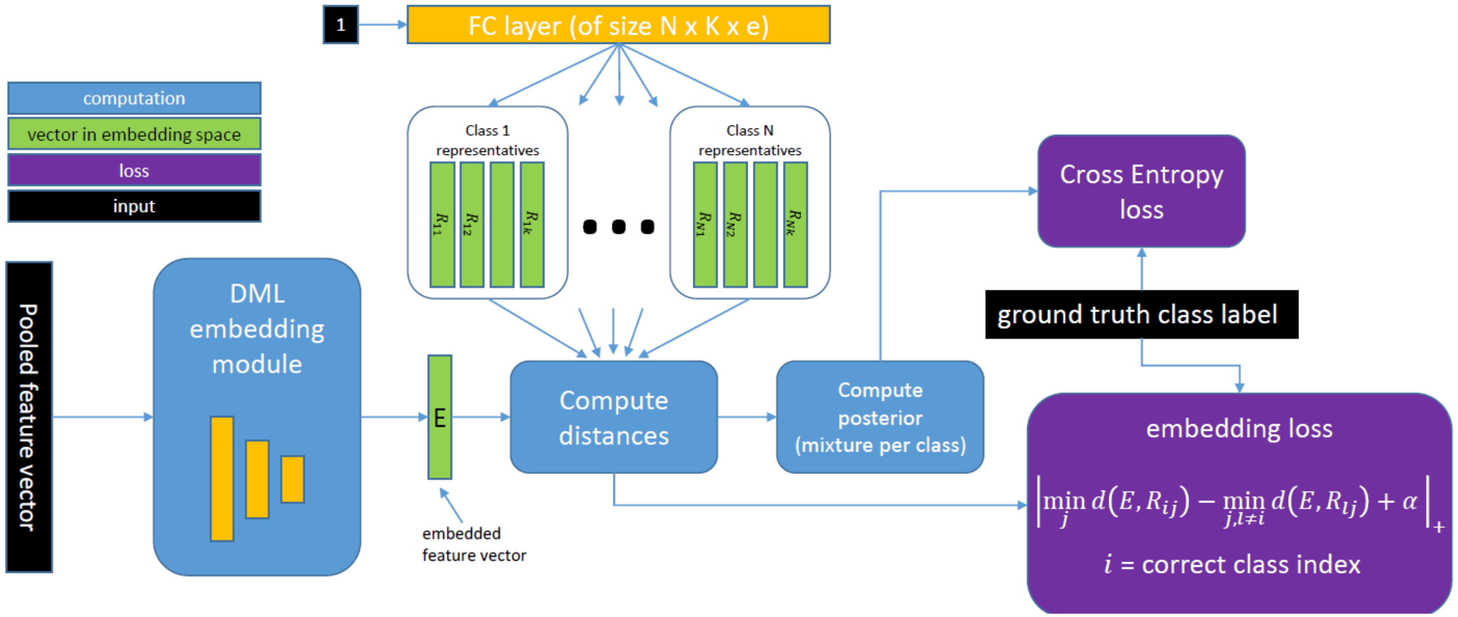
Raja Giryes, Alex M. Bronstein

CVPR 2019

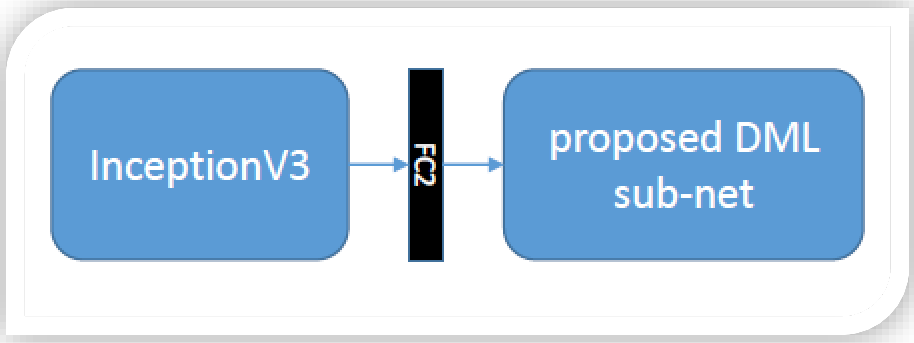
RepMet: joint training of the metric (embedding) and the class mixtures for effective DML based CLS/DET

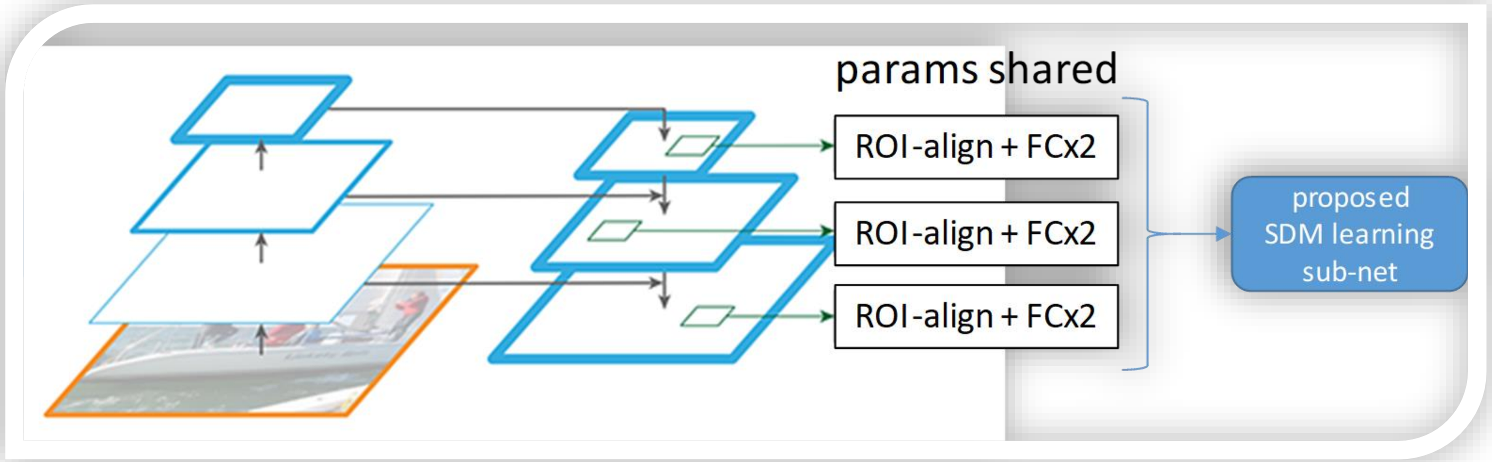


RepMet: the way it works



on top of a classifier→

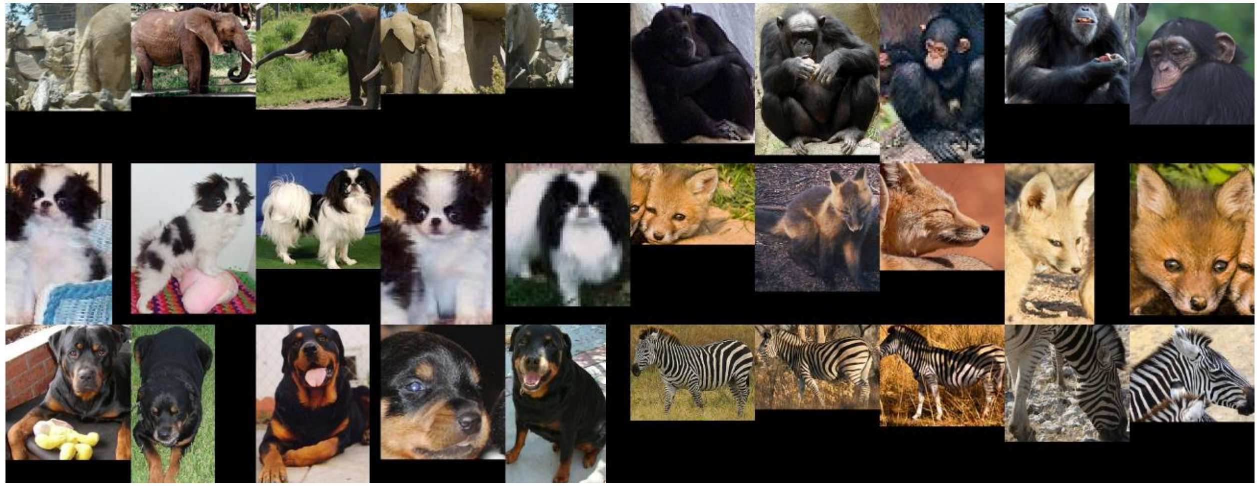


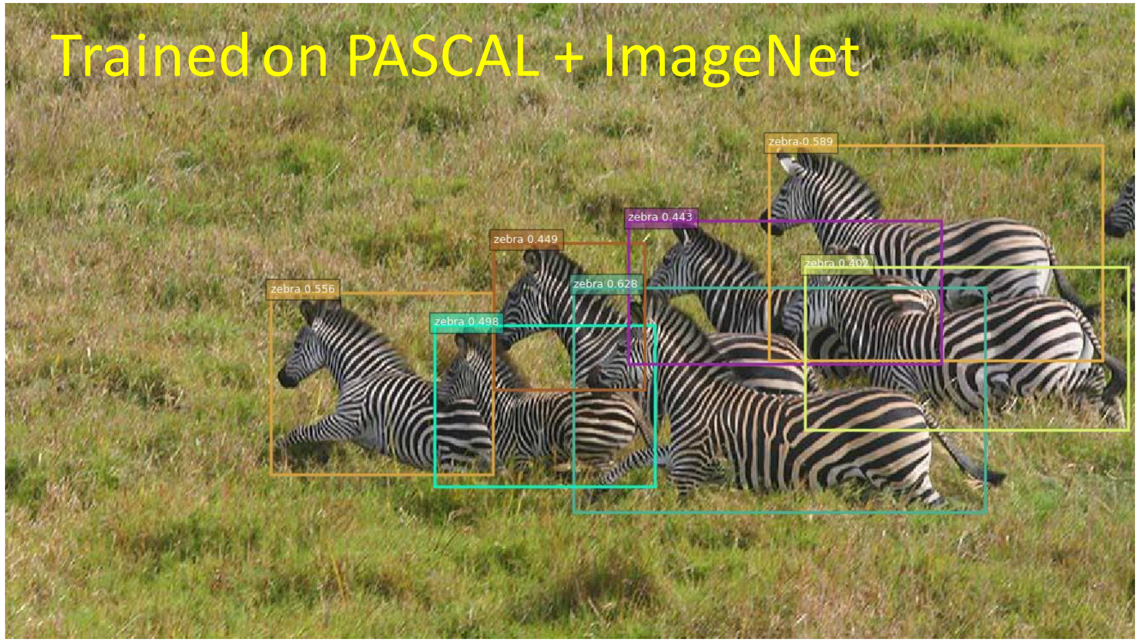


← on top of an FPN detector

“Regular” detection performance

Representatives learned, shown for some of the categories



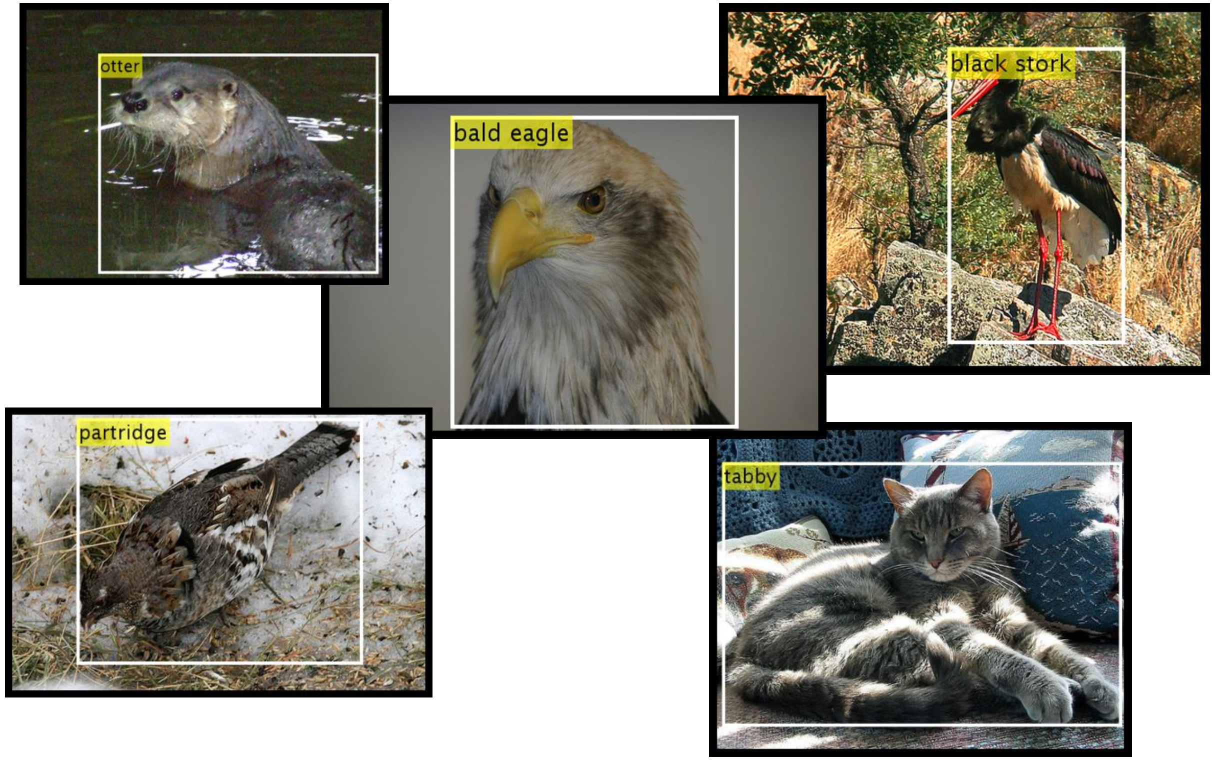


Regular detection performance, mAP(%). FPN-DCN evaluated using their original code.

	PASCAL VOC			ImageNet (LOC)		
acceptance IoU	0.7	0.5	0.3	0.7	0.5	0.3
FPN-DCN [6]	**74.6**	**83.5**	**85.3**	46.9	55.2	60.2
ours	73.7	82.9	84.9	**60.7**	**61.7**	**70.7**

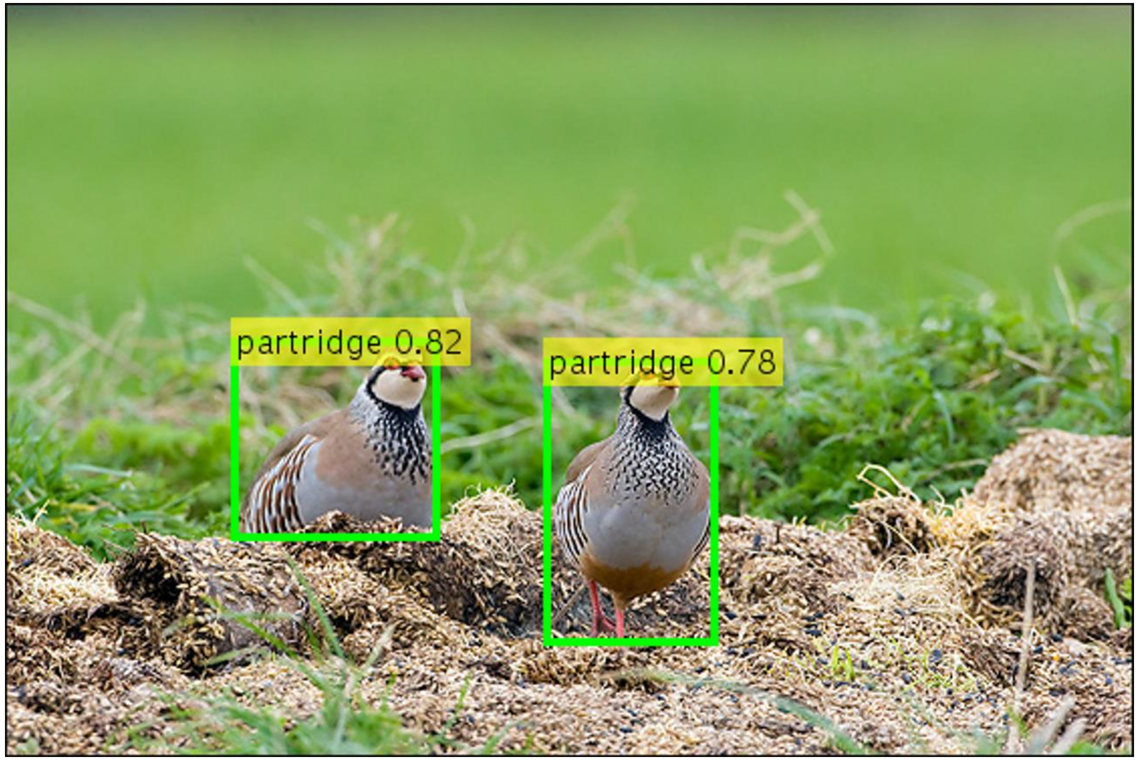
Few-shot Detection - experimental setup

Train



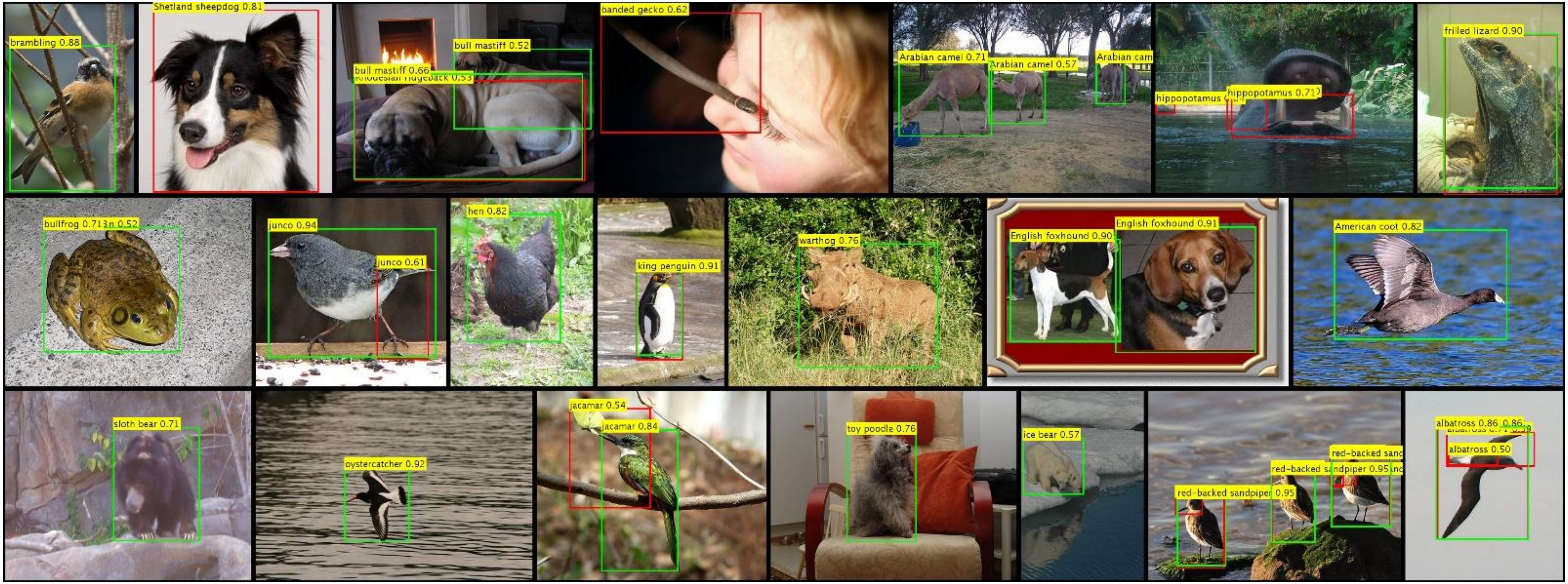
At test time – replace the “known” classes representatives with embedding of the strongly overlapping proposals from the episode training images

Test



1-shot, 5-way

Some qualitative results



few-shot detection performance

dataset	method	no episode fine-tuning			with episode fine-tuning		
		1-shot	5-shot	10-shot	1-shot	5-shot	10-shot
ImageNet-LOC (214 unseen animal classes)	baseline-FT (FPN-DCN [7])	—	—	—	35.0	51.0	59.7
	baseline-DML	41.3	58.2	61.6	41.3	59.7	66.5
	baseline-DML-external	19.0	30.2	30.4	32.1	37.2	38.1
	Ours	**56.9**	**68.8**	**71.5**	**59.2**	**73.9**	**79.2**
ImageNet-LOC (100 seen animal classes)	Ours - trained representatives	—	86.3	—	—	—	—
	Ours - episode representatives	64.5	79.4	82.6	—	—	—

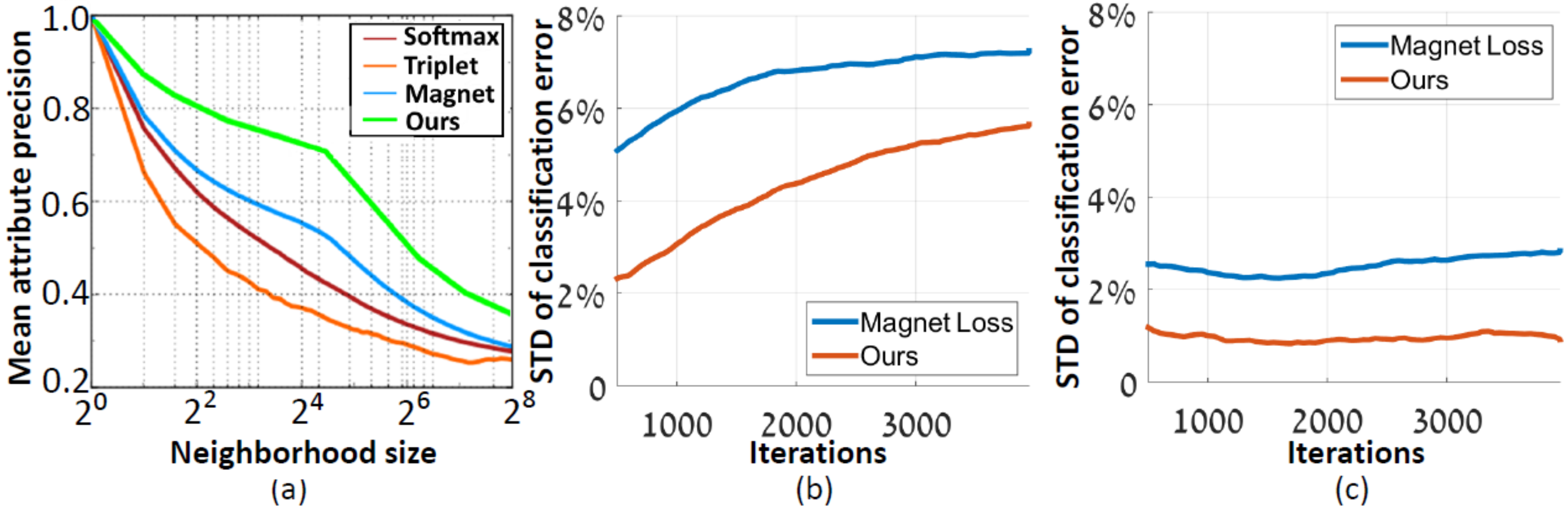
Table 3. Few-shot 5-way detection test performance on ImageNet-LOC. Reported as mAP in %.

	1-shot	5-shot	10-shot
LSTD [5]	19.2	37.4	44.3
ours	**24.1**	**39.6**	**49.2**

Metric learning classification results

Method	MsML [21]	Magnet [24]	VMF [37]	Ours
Stanford Dogs	29.7	24.9	24.0	**14.2**
Oxford Flowers	10.5	8.6	**4.4**	11.2
Oxford Pet	18.8	10.6	9.9	**6.9**
ImageNet Attributes	–	15.9	—	**13.2**

Table 1: Comparison of test error with state-of-the-art DML classifier approaches on different fine-grained classification datasets. Lower is better.



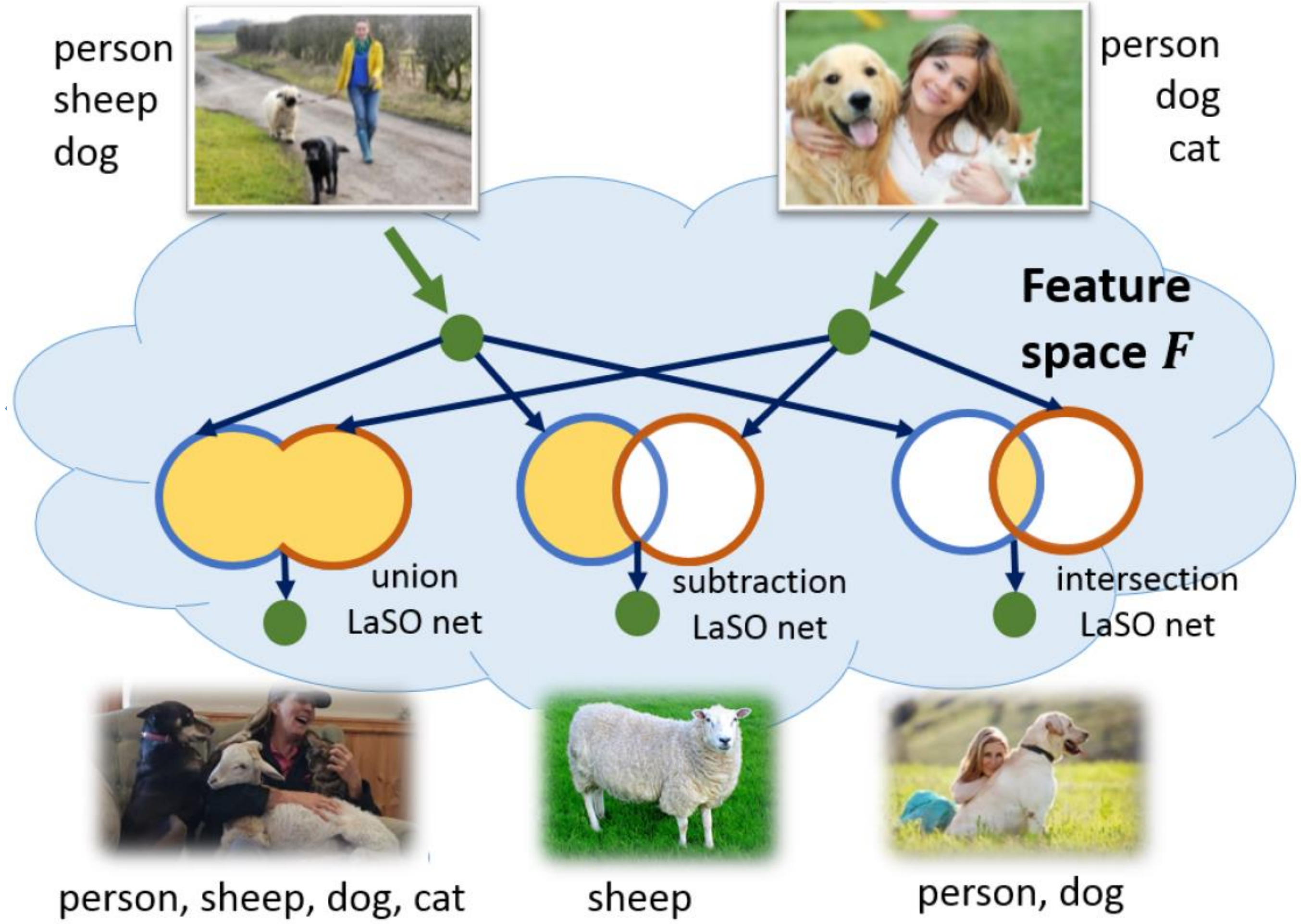
LaSO: Label-Set Operations network for multi-label few-shot classification

Amit Alfassy*, Leonid Karlinsky*, Amit Aides*,
Joseph Shtok, Sivan Harary
Rogerio Feris, Raja Giryes, Alex M. Bronstein

CVPR 2019

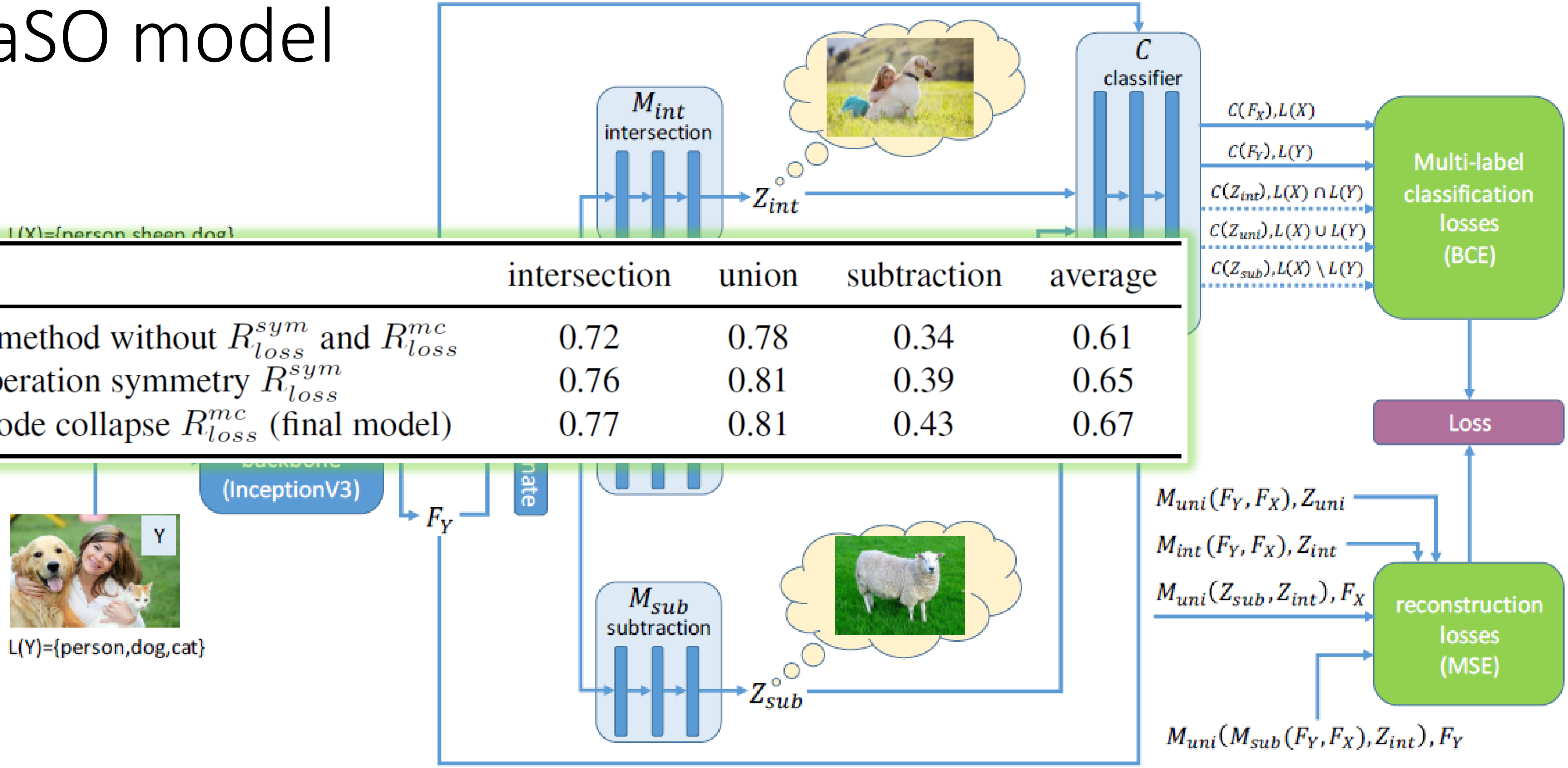
LaSO concept

Learning generic (label agnostic) operators for manipulating semantic content



=

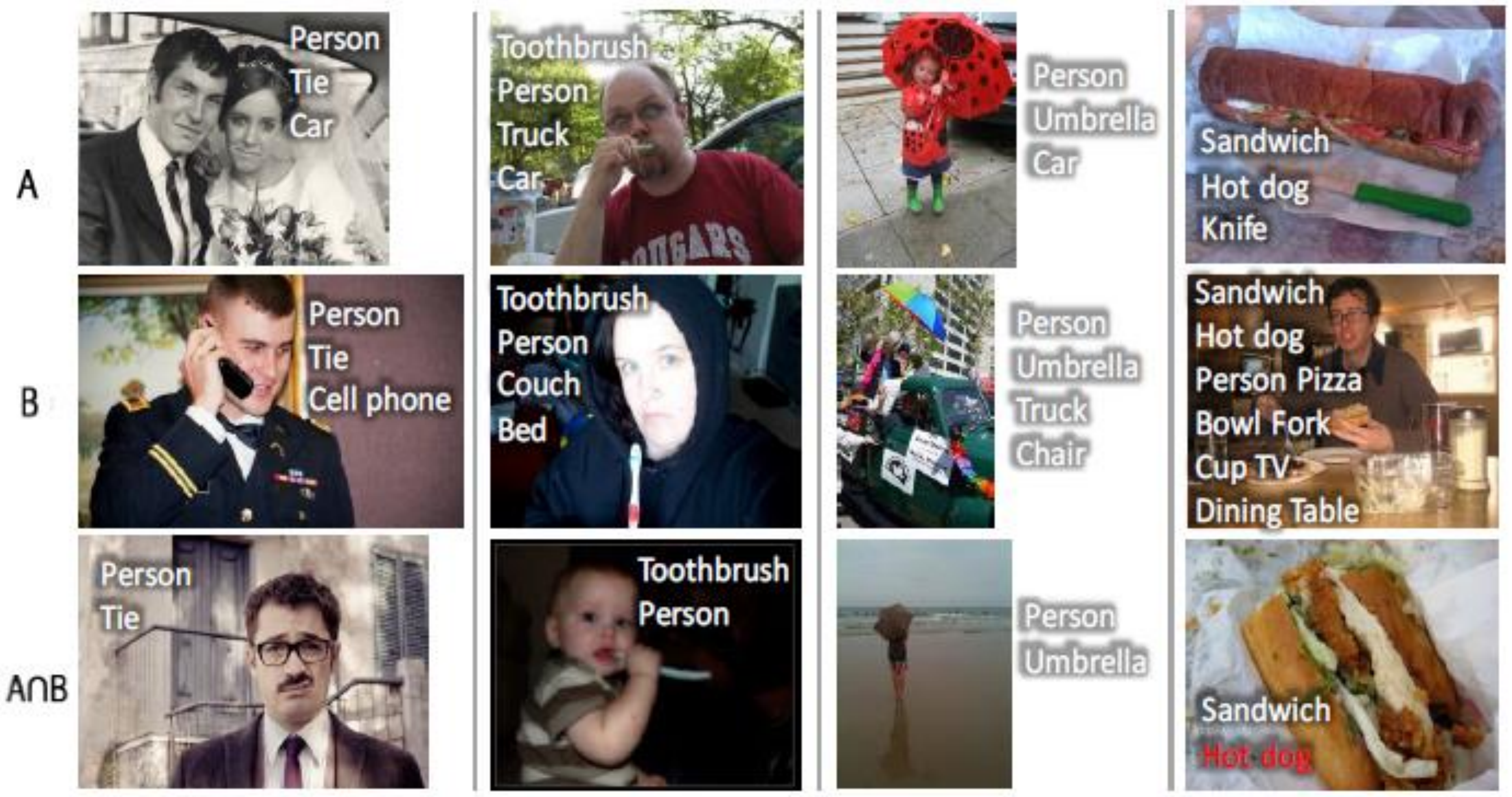
LaSO model



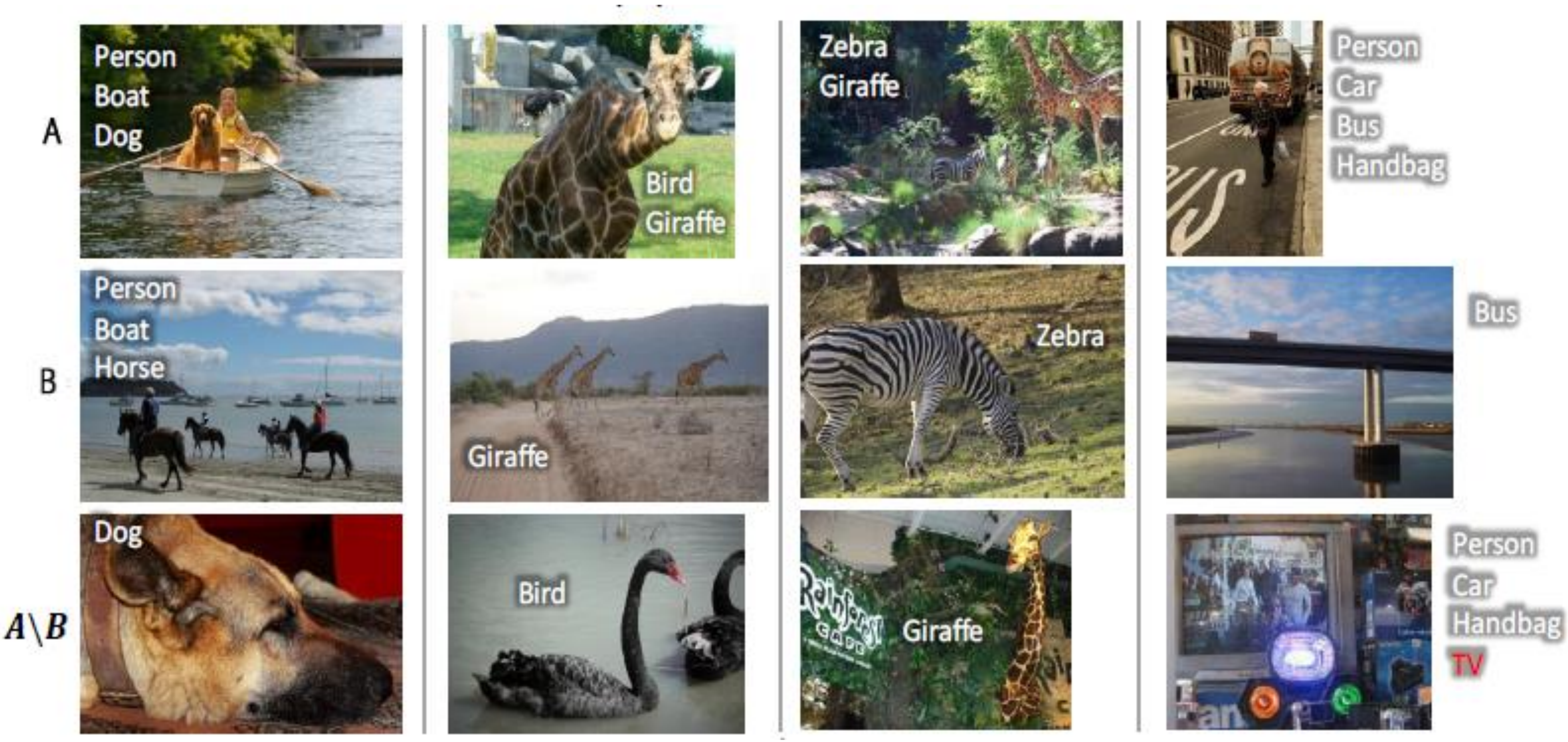
	intersection	union	subtraction	average
our method without R^{sym}_{loss} and R^{mc}_{loss}	0.72	0.78	0.34	0.61
+ operation symmetry R^{sym}_{loss}	0.76	0.81	0.39	0.65
+ mode collapse R^{mc}_{loss} (final model)	0.77	0.81	0.43	0.67

LaSO model: schematic illustration of all the components of the proposed approach (including training losses).

Some qualitative examples – intersection

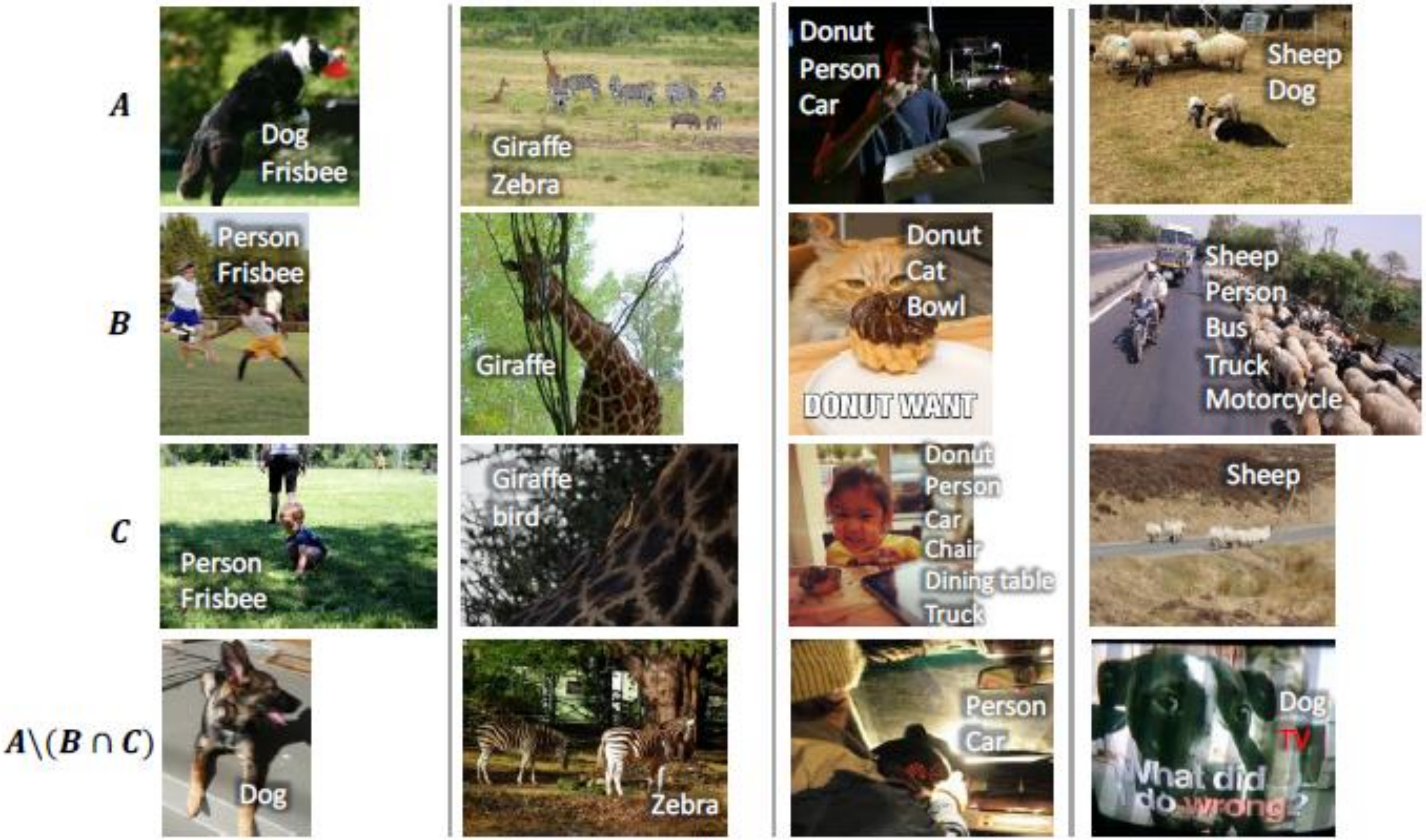


Some qualitative examples – subtraction



Some qualitative examples – union

Some qualitative examples – “tiger”



Quantitative results

Classification accuracy COCO

	64 seen classes	16 unseen classes
intersection	77	48
union	80	61
subtraction	43	14
upper bound	75	79

Classification accuracy CelebA

	40 facial attributes
intersection	48
union	75
subtraction	69
upper bound	79

Retrieval accuracy COCO

	64 seen classes			16 unseen classes		
	top-1	top-3	top-5	top-1	top-3	top-5
intersection	0.7	0.79	0.82	0.47	0.71	0.78
union	0.61	0.71	0.74	0.44	0.64	0.71
subtraction	0.19	0.32	0.4	0.21	0.4	0.51
upper bound	0.56	0.72	0.76	0.56	0.75	0.81

Learning to augment for multi-label few-shot classification

	1-shot	5-shot
B1: no augmentation	39.2	49.4
B2: basic aug.	39.2	52.7
B3: mixUP aug.	40.2	54.0
analytic intersection aug.	40.7	55.4
analytic union aug.	44.5	55.6
learned intersection aug.	40.5	57.2
learned union aug.	**45.3**	**58.1**

The team & collaborations

Team members

LEONID KARLINSKY
Team lead

JOSEPH SHTOK
Vision, Deep learning

SIVAN HARARY
Vision, Deep learning

MATTIAS MARDER
Vision, Deep learning

ELI SCHWARTZ
Vision, Deep learning

AMIT AIDES
Vision, Deep learning

AMIT ALFASSY
Vision, Deep learning

Collaborating with Rogerio Feris (IBM Research AI), Prof. Raja Giryes (TAU) and Prof. Alex Bronstein (Technion)

Thank you for listening!

13:50 - 14:20 - Advanced Deep Learning Tutorials
Few-shot Learning – State of the Art
Dr. Joseph Shtok, *IBM*

All day – "LaSO: Label-Set Operations network for multi-label few-shot classification" poster
Amit Alfassy, *IBM*