Samsung Research

Online Continual Learning for Robust Indoor Object Recognition

Umberto Michieli, Mete Ozay
Samsung Research UK - n.surname@samsung.com

Introduction

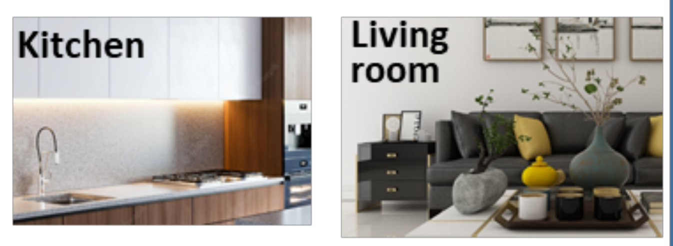
TASK:
Class-Incremental Online Continual Learning

DESIDERATA:
- Robust to Test Time Variations
- Data-Efficient (Few-Shot Training)
- Targeting Limited-Resource Devices

1. PRE-TRAINING IN FACTORY

- General-purpose datasets
  - Pre-defined **classes**

  - Limited **domains**



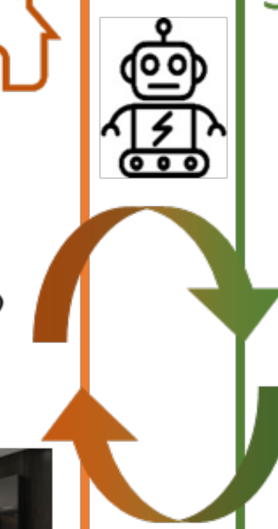
- Large batch size
- Update whole model

2. PERSONALIZATION TO USER ENVIRONMENT

- Continual Learning
  - Class-incremental

  - Domain changes

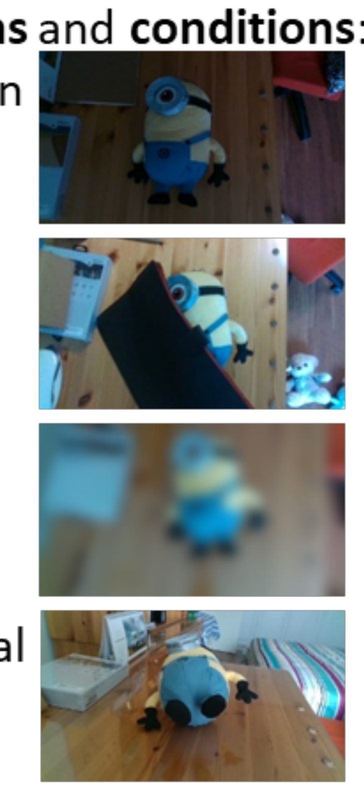
- Robot has limited resources
  - **No data storage** → no large batch size
  - **No update of whole model**
  - **Few labelled samples**



3. TEST-TIME USAGE IN USER ENVIRONMENT

Need to recognize **new classes** in **new domains** and **conditions**:

- Illumination
- Occlusion
- Blur
- Geometrical distortion



Method

Three Main Components

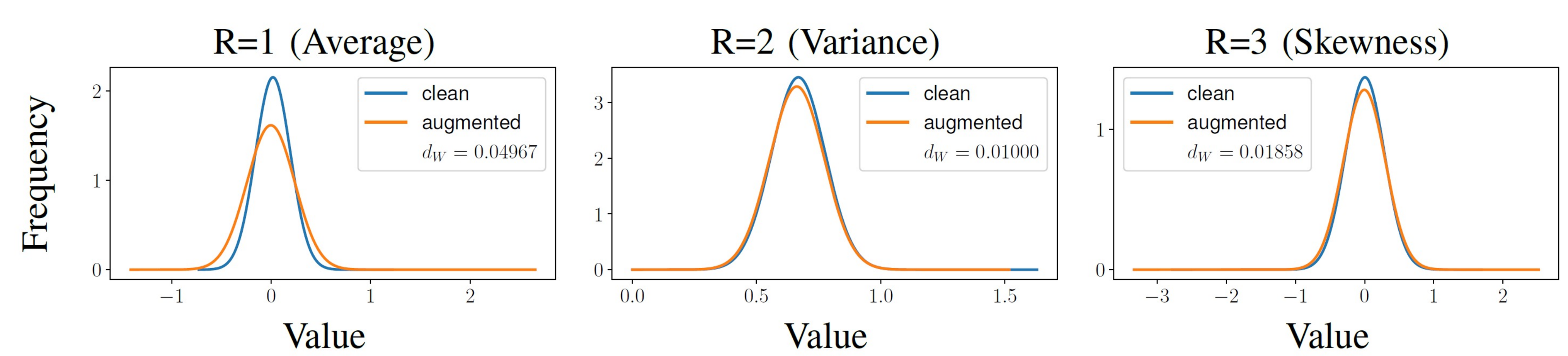
1) Feature Extractor
- Pre-trained on server on public data
- Frozen during on-device adaptation

2) Pooling Scheme
- Concatenation of the first R statistical moments (e.g., R=3)
- Richer feature space to extract cues during a single-epoch training.

3) Classifier
- Lightweight Online Continual Learning Method
- We use SLDA [1] on the enriched feature space.
- SLDA estimates a Gaussian model for each class over the feature space with a class-wise mean (prototype) and shared-across-classes variance.

Intuition

We plot the distribution of statistical moments of features extracted from **clean** or **augmented** samples and we compute their Wasserstein distance (d_W) :



First statistical moment suffers from more variability than second and third ones.

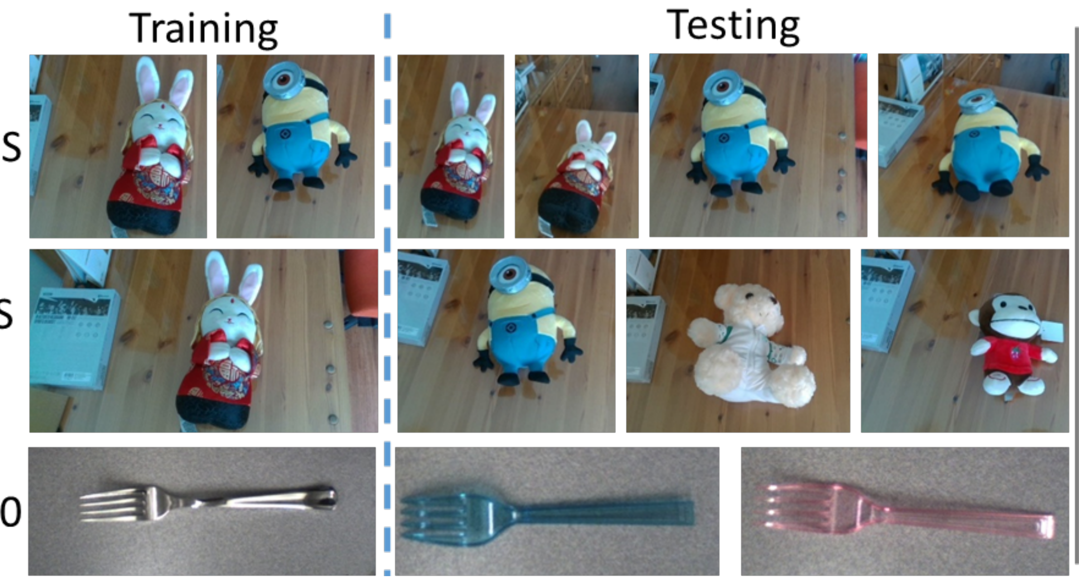
→ Using higher order moments is beneficial for robustness/resilience and invariance to domain shift.

Dataset

Datasets employed in our work

Right-hand side table summarizes differences between train and test sets.

[Text within squared brackets]: properties change when our augmentations are applied



	Different Objects	Different Conditions	Few Shot	Augs Type
OpenLORIS	✗	partial [✓]	✗	Real [Mix]
OpenLORIS small	✓	partial [✓]	partial	Real [Mix]
F-SIOL-310	✓	✗ [✓]	✓	Real [Mix]

[1] Hayes et al., "Lifelong Machine Learning with Deep Streaming Linear Discriminant Analysis", CVPRW (2020).

Results

3 USE CASES:

1. Same-Domain

OpenLORIS dataset.
Averaged accuracy over 16 backbones (9 CNNs and 7 transformers)

CL method	Averaged Accuracy
FT	87.0
Perceptron	82.4
SNB	34.4
SOvR	54.2
NCM/CBCL	80.1
SQDA	36.4
iCARL	94.7
iCARL (2pc)	92.5
SLDA	97.7
RobOCLe$_{SLDA}$ (ours)	**99.2**
(Δ_R [%])	(+65.8)

2. Real Other-Domain Few-Shot

Averaged accuracy over 7 backbones (3 CNNs and 4 transformers)

CL method	OpenLORIS-small	F-SIOL-310 (5-shots)	F-SIOL-310 (10-shots)
FT	7.6	21.0	17.7
Perceptron	11.9	16.2	17.2
SNB	5.3	9.7	8.3
SOvR	28.9	43.5	41.8
NCM/CBCL	41.2	93.3	94.7
SQDA	40.3	94.1	96.0
iCARL	42.2	47.8	59.2
iCARL (2pc)	39.7	48.0	59.4
SLDA	46.4	95.0	97.9
RobOCLe$_{SLDA}$ (ours)	**47.3**	**96.0**	**98.3**
(Δ_R [%])	(+1.8)	(+19.5)	(+17.5)

3. Synthetic Other-Domain Few-Shot

OpenLORIS with ViT-L16

		NCM						RobOCLe$_{NCM}$ (ours)						
	tr↓ te→	clean	illum	geom	noise	all	Avg OD	clean	illum	geom	noise	all	Avg OD	Δ_R^{OD}
ViT-L16	clean	79.47	35.11	73.43	66.57	19.49	48.65	**80.31**	33.13	73.69	68.38	19.74	**48.73**	(+0.2)
	illum	30.78	40.58	33.00	29.12	30.00	30.73	30.82	**35.00**	35.17	30.65	31.22	**31.96**	(+1.8)
	geom	78.01	36.62	79.19	68.73	23.31	51.67	78.70	37.05	**78.64**	68.87	24.32	**52.24**	(+1.2)
	noise	73.67	36.37	66.27	75.64	25.41	50.43	73.84	36.58	68.55	**76.51**	25.56	**51.13**	(+1.4)
	all	15.32	32.27	21.58	15.25	31.79	21.11	17.05	32.26	20.98	18.21	**24.77**	**22.13**	(+1.3)
		SLDA						RobOCLe$_{SLDA}$ (ours)						
	clean	98.69	38.30	92.94	90.09	21.49	60.71	**99.73**	40.69	95.04	93.03	22.53	**62.82**	(+5.4)
	illum	83.68	77.13	72.86	72.39	50.58	69.88	86.13	**80.65**	77.00	76.19	52.56	**72.97**	(+10.3)
	geom	97.72	46.72	96.86	89.44	27.98	65.46	98.74	45.87	**98.10**	92.25	28.08	**66.23**	(+2.2)
	noise	97.42	40.08	89.87	97.19	26.21	63.40	98.82	41.90	93.41	**98.68**	26.69	**65.21**	(+4.9)
	all	71.27	68.15	66.03	68.67	65.09	68.53	72.81	71.36	71.58	71.59	**68.43**	**71.84**	(+10.5)

Conclusions

New method: RobOCLe, a data- and parameter-efficient online continual learning method with robust performance under test-time corruptions.

RobOCLe features:
- *Lightweight* solution: frozen feature extractor + class-conditional Gaussian modelling of feature space
- *High-order statistical moments* of embedded features from inputs
- *Robust* recognition in a variety of scenarios, using several backbones, low-shot setups, per-step accuracy, and controlled train/test augmentation on both same-domain and other-domain data