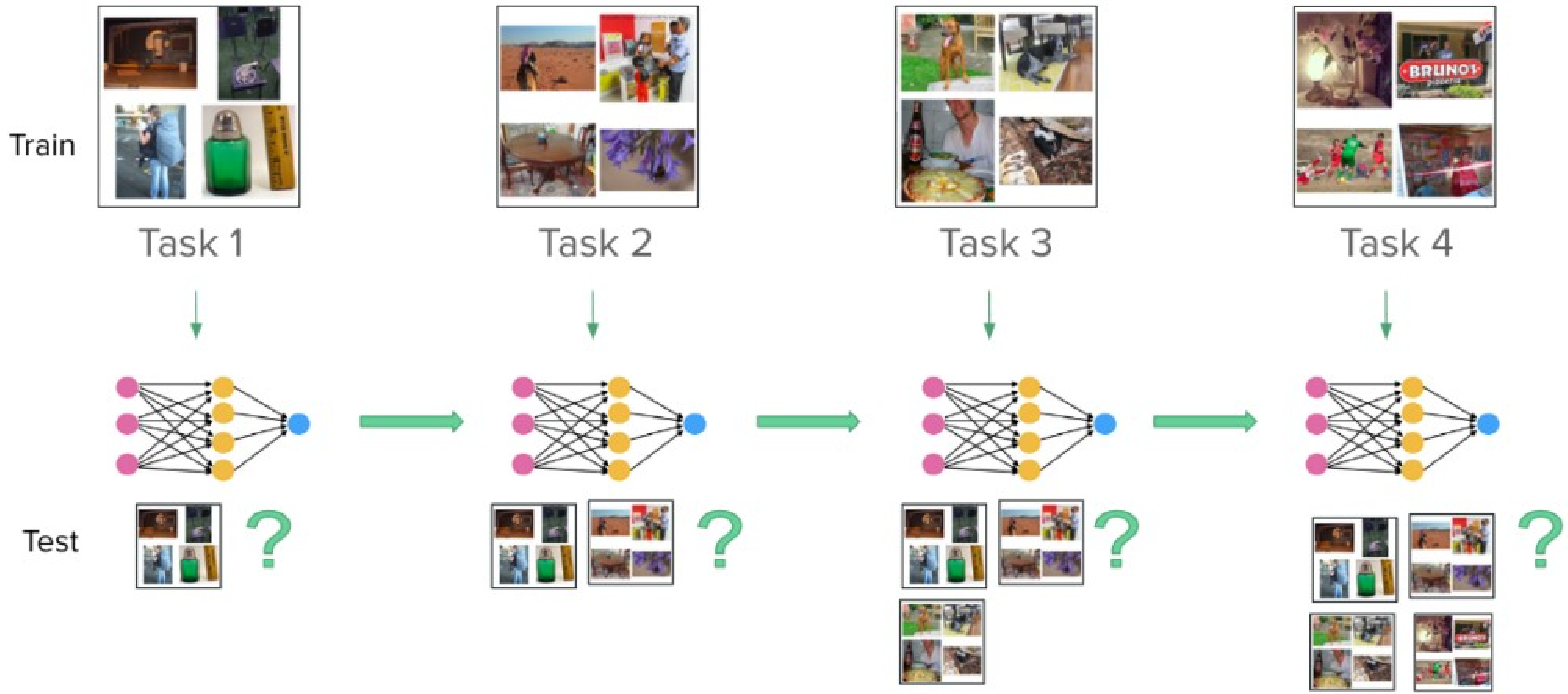
IDEAS NCBR

Task-recency bias strikes back: Adapting covariances in Exemplar-Free Class Incremental Learning

Grzegorz Rypeść[1,2], Sebastian Cygert[1,3], Tomasz Trzciński[1,2,4], Bartłomiej Twardowski[1,5.7]

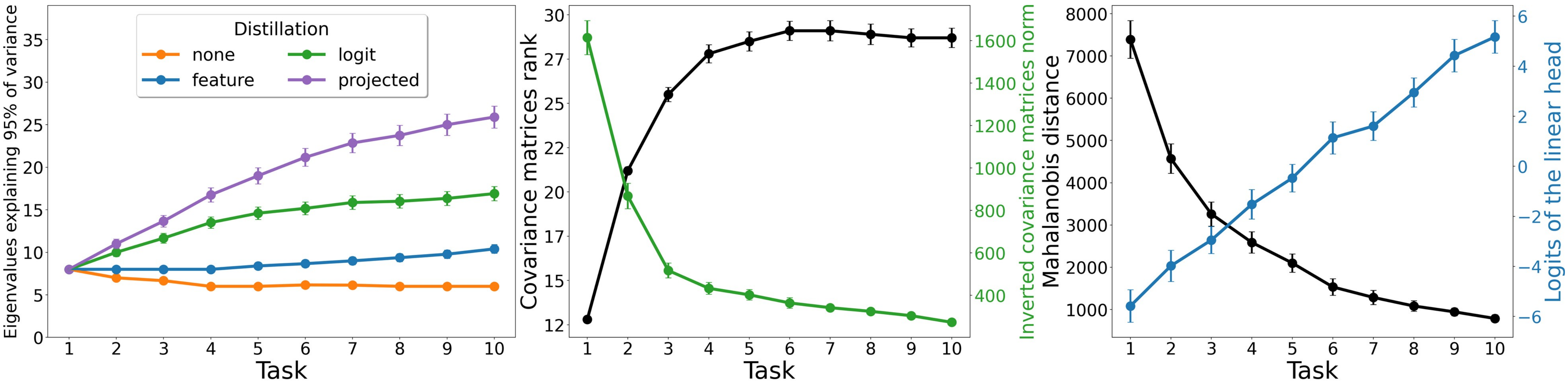
[1]IDEAS-NCBR, [2]Warsaw University of Technology, [3]Gdańsk University of Technology, [4]Tooploox, [5]Computer Vision Center, Barcelona

Exemplar-Free Class Incremental Learning



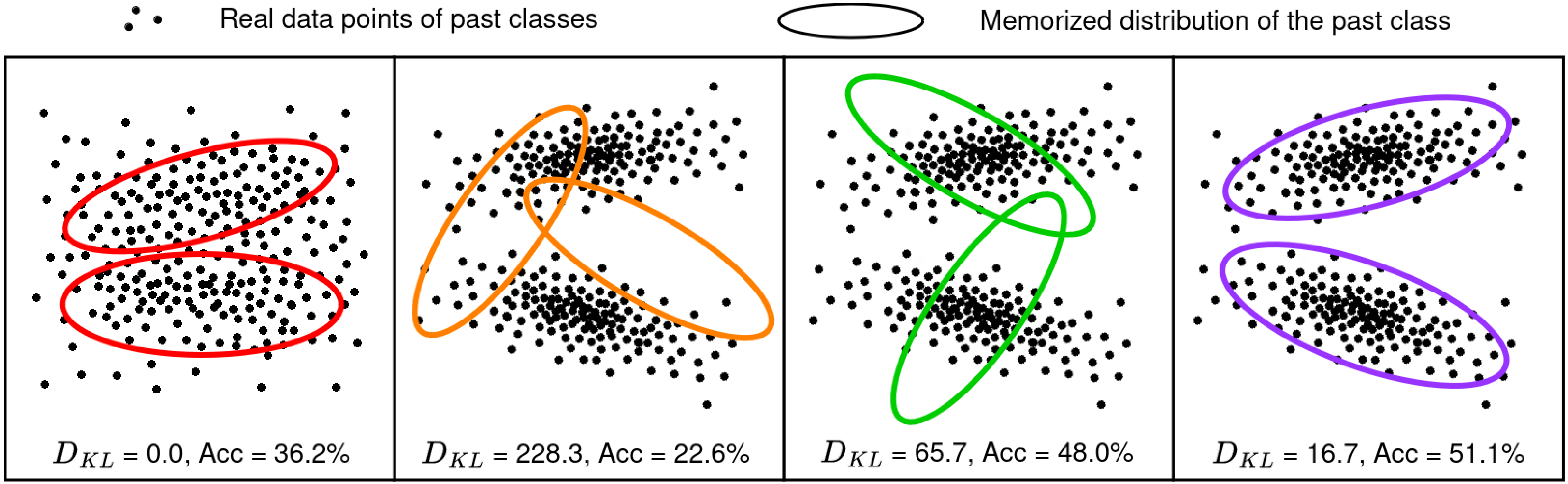
van de Ven, Gido M., and Andreas S. Tolias. "Three scenarios for continual learning." arXiv e-prints (2019): arXiv-1904.

Task recency bias in the latent space



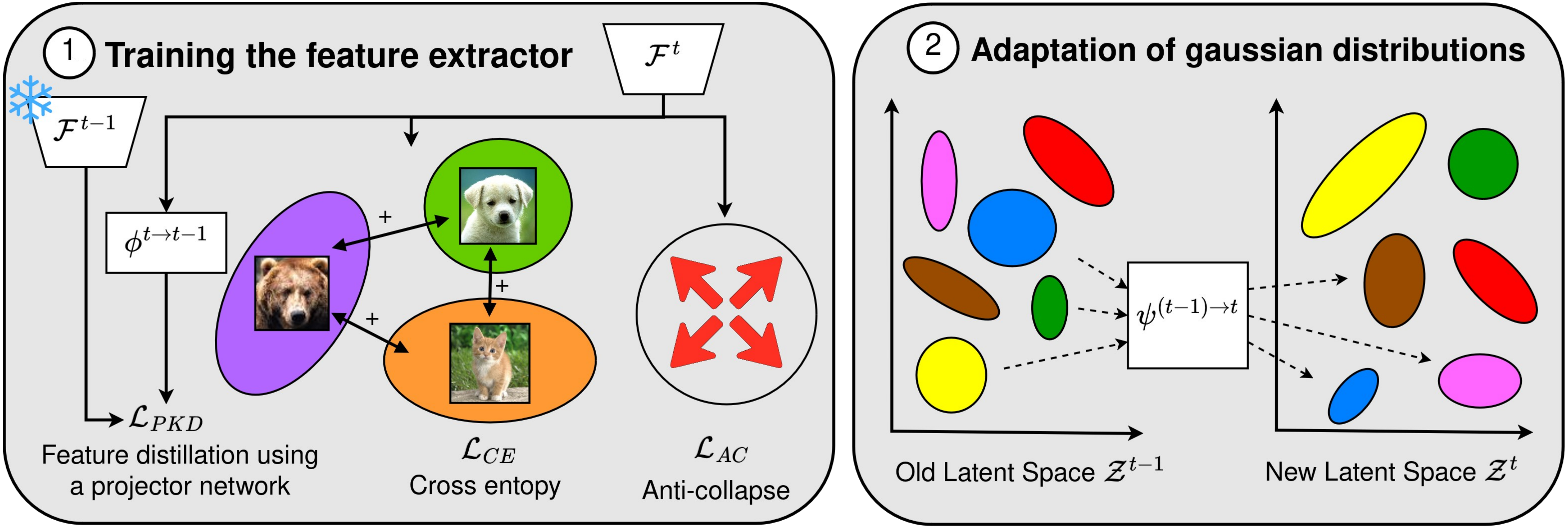
The representation strength of the feature extractor grows with each task, which makes the rank of covariance of new classes higher than that of old classes. That causes the norm of the inverse of covariance matrices to be lower for later tasks. It causes task recency bias because the inverse is utilized to sample from memorized distributions or to calculate Mahalanobis distance when classifying.

Distributions of old classes must be adapted!



Frozen feature extractor | Unfrozen, no adaptation | Unfrozen, mean adaptation | Mean and covariance adaptation

Our method: AdaGauss



Training from scratch

Table 1: Average incremental and last accuracy in EFCIL when training the feature extractor from scratch. The mean of 5 runs is reported. Full results are in Tab. 5. We denote the best results **in bold**.

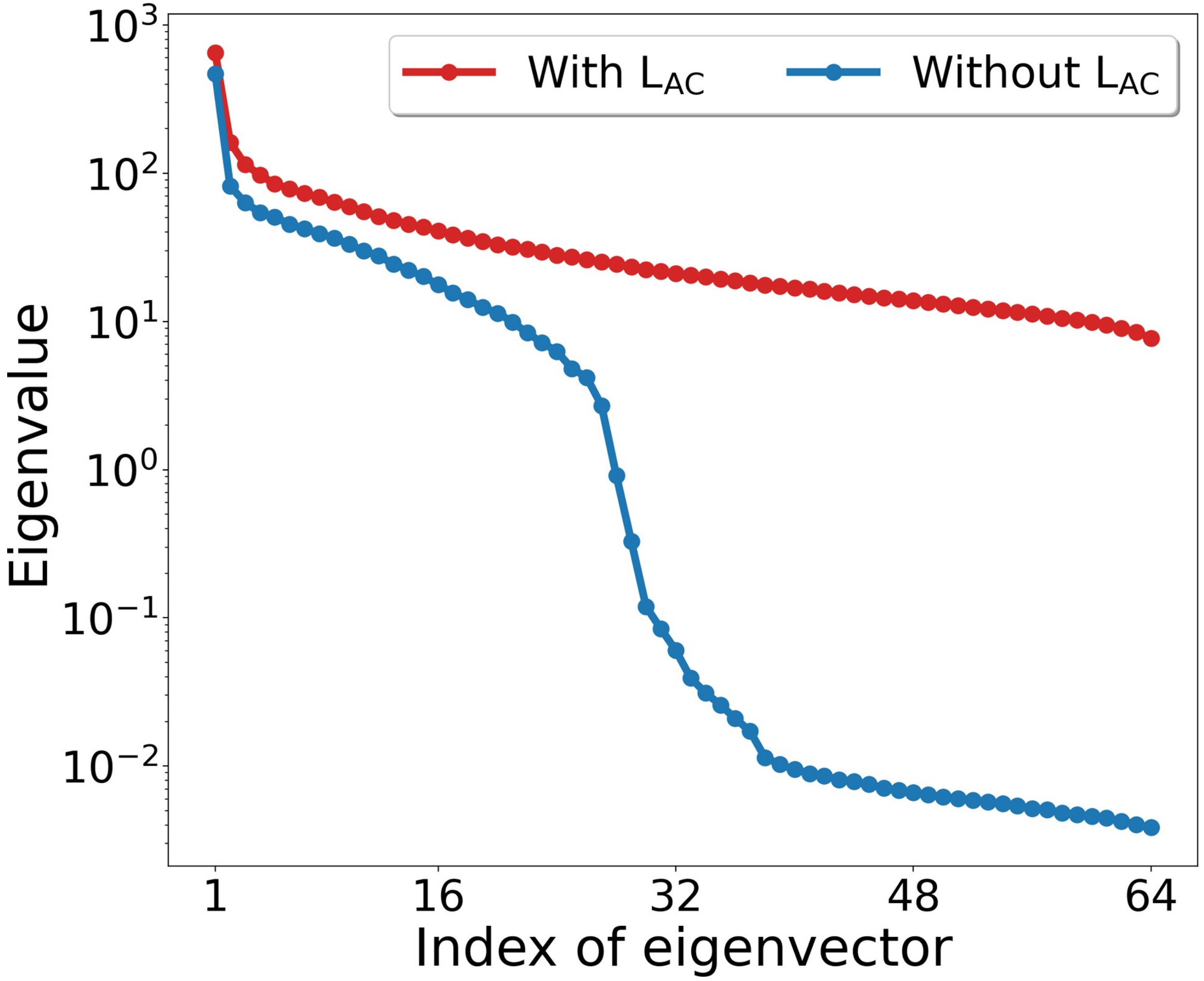
Method	CIFAR-100				TinyImageNet				ImagenetSubset			
	T=10		T=20		T=10		T=20		T=10		T=20	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
EWC [17]	31.2	49.1	17.4	31.0	17.6	32.6	11.3	26.8	24.6	39.4	12.8	27.0
LwF [21]	32.8	53.9	17.4	38.4	26.1	45.1	15.0	32.9	37.7	56.4	18.6	40.2
PASS [52]	30.5	47.9	17.4	32.9	24.1	39.3	18.7	32.0	26.4	45.7	14.4	31.7
IL2A [51]	31.7	48.4	23.0	40.2	25.3	42.0	19.8	35.5	27.7	48.4	17.5	34.9
SSRE [53]	30.4	47.3	17.5	32.5	22.9	38.8	17.3	30.6	25.4	43.8	16.3	31.2
FeTrIL [31]	34.9	51.2	23.3	38.5	31.0	45.6	25.7	39.5	36.2	52.6	26.6	42.4
FeCAM [10]	32.4	48.3	20.6	34.1	30.8	44.5	25.2	38.3	38.7	54.8	29.0	44.6
DS-AL [54]	40.8	54.9	31.7	43.2	33.6	47.2	26.5	41.6	46.8	58.6	36.7	48.5
EFC [24]	43.6	58.6	32.2	47.3	34.1	48.0	28.7	42.1	47.4	59.9	35.8	49.9
AdaGauss	**46.1**	**60.2**	**37.8**	**52.4**	**36.5**	**50.6**	**31.3**	**45.1**	**51.1**	**65.0**	**42.6**	**57.4**

Starting from a pretrained model

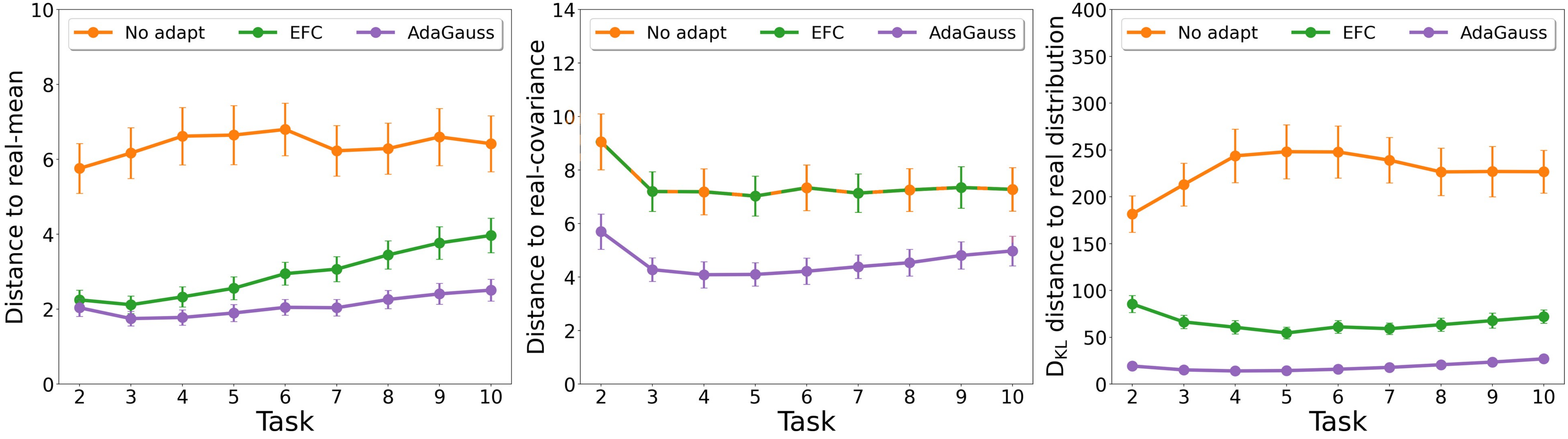
Table 2: Average incremental and last accuracy in EFCIL fine-grained scenarios when utilizing a pre-trained feature extractor. We report the mean of 5 runs, while variances are reported in Tab. 6.

Method	CUB200						FGVCAircraft					
	T=5		T=10		T=20		T=5		T=10		T=20	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
EWC [17]	21.6	38.2	15.8	32.6	12.3	27.2	24.3	44.0	14.3	34.5	10.9	27.9
LwF [21]	44.3	57.7	30.4	46.1	19.4	34.7	39.0	55.2	28.0	46.5	14.7	30.5
PASS [52]	34.5	48.6	27.0	42.3	18.1	36.9	33.3	48.9	26.4	41.0	13.9	28.1
IL2A [51]	36.9	51.3	29.4	45.5	20.8	35.1	39.4	49.1	27.3	45.1	14.2	28.7
FeTrIL [31]	41.9	53.2	36.9	48.2	34.6	45.3	46.0	58.5	40.5	53.4	32.5	43.3
FeCAM [10]	43.5	56.0	40.2	54.9	36.2	48.9	45.3	58.0	41.4	55.2	34.0	46.0
DS-AL [54]	49.4	61.9	45.8	59.1	41.4	53.8	50.6	62.7	42.6	56.4	34.2	46.7
EFC [24]	58.3	68.9	51.0	63.3	46.1	59.3	50.1	63.2	43.1	57.6	28.1	48.2
AdaGauss	**60.4**	**69.2**	**55.8**	**66.2**	**47.4**	**60.6**	**53.3**	**64.0**	**47.5**	**58.5**	**34.8**	**48.6**

Anti-collapse helps to increase strength of representations



Adaptation results



Find out more!

Code

Arxiv

Grzegorz Rypeść
LinkedIn
grzegorz.rypesc@ideas-ncbr.pl