模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Learning Equi-angular Representations for Online Continual Learning

Minhyuk Seo[1,†] Hyunseo Koh[1] Wonje Jeung[1] Minjae Lee[1] San Kim[1] Hankook Lee[2,3]
Sungjun Cho[2] Sungik Choi[2] Hyunwoo Kim[4,*] Jonghyun Choi[5,*]

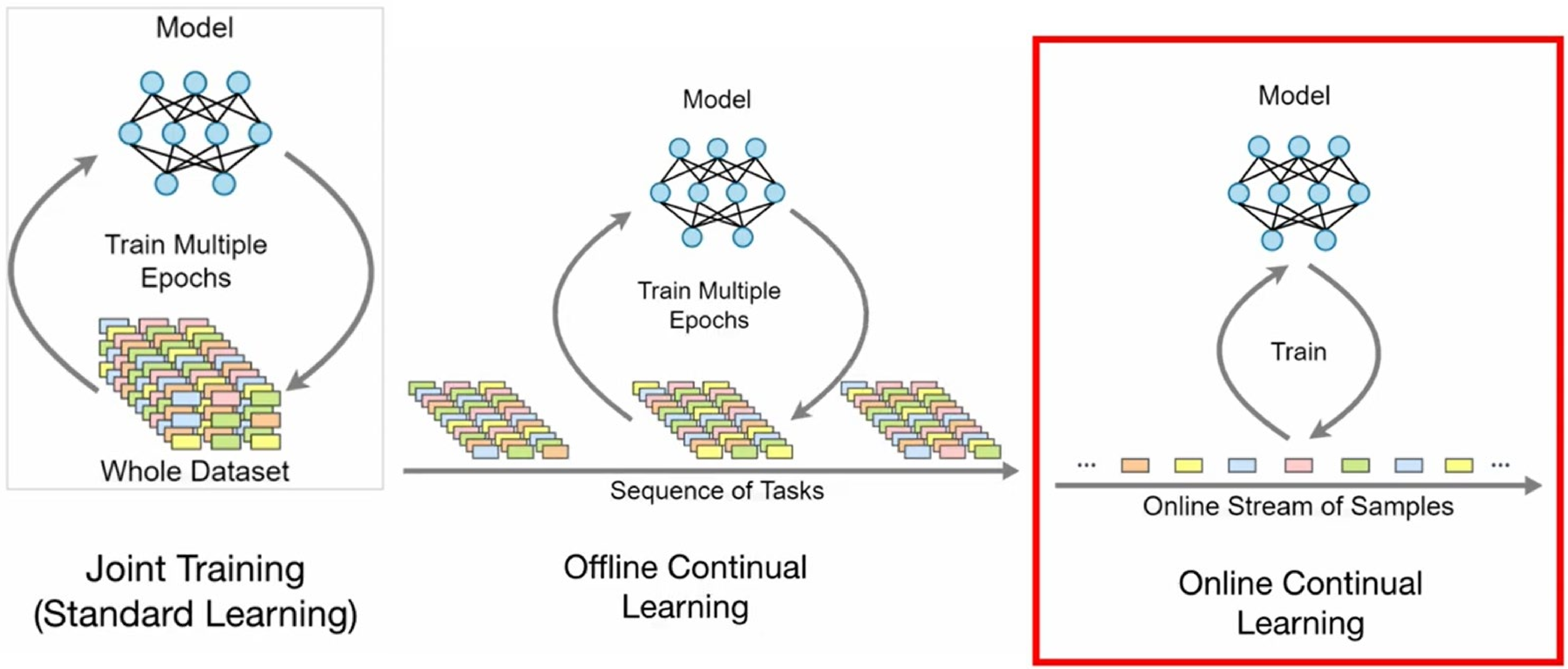
[1]Yonsei Univ. [2]LG AI Research [3]Sungkyunkwan Univ. [4]Zhejiang Lab [5]Seoul National Univ.

{dbd0508, specific0924, reccos1020, nasmik419}@yonsei.ac.kr
khs8157@gmail.com, {sungik.choi,sungjun.cho}@lgresearch.ai
hankook.lee@skku.edu, hwkim@zhejianglab.com, jonghyunchoi@snu.ac.kr

CVPR 2024

Preliminary: Standard Learning vs Continual Learning

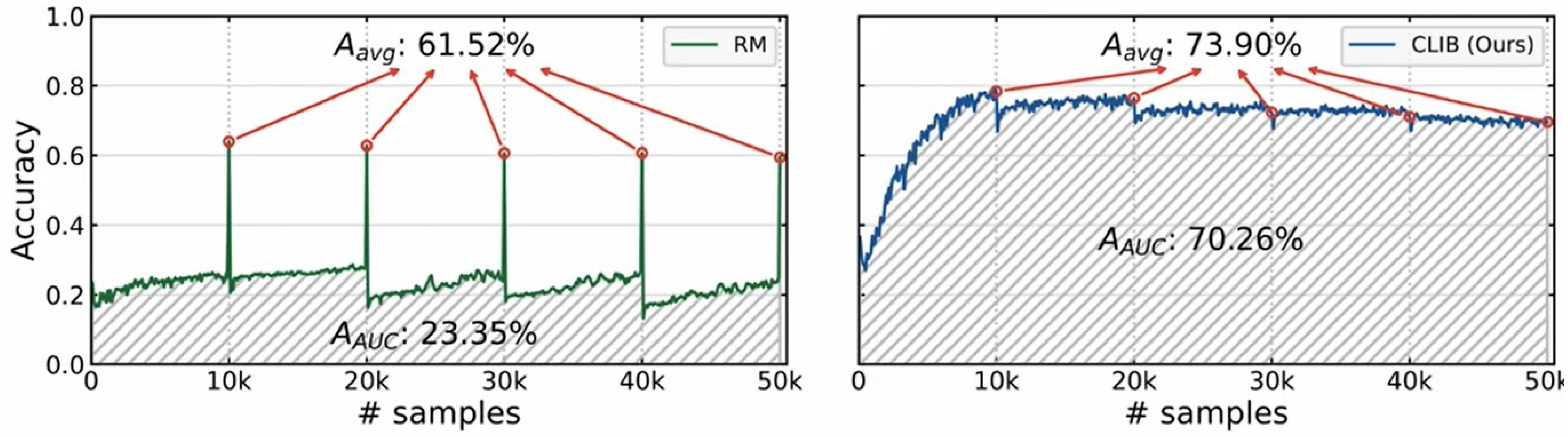
模式识别与神经计算研究组
PAttern Recognition and NEural Computing



Preliminary: Anytime Inference Performance [1, 2]

- Continual learning in the real-world scenario is endless in learning

  => Must be available for inference at any time



Rainbow Memory [1]   CLIB [3]

[1] Bang et al., Rainbow Memory: Continual Learning with a Memory of Diverse Samples, CVPR 2021
[2] Caccia et al., On Anytime Learning at Macroscale, CoLLAs 2022
[3] Koh et al., Online continual learning on class incremental blurry ta:sk configuration with anytime inferenceICLR 2022

Imbalanced Distribution in Continual Learning

- Continual Learning has an imbalanced distribution at each intermediate time point, even if the overall dataset is balanced since new data are encountered continuously

Problem in Imbalanced Distribution) Minority Collapse [1]

ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

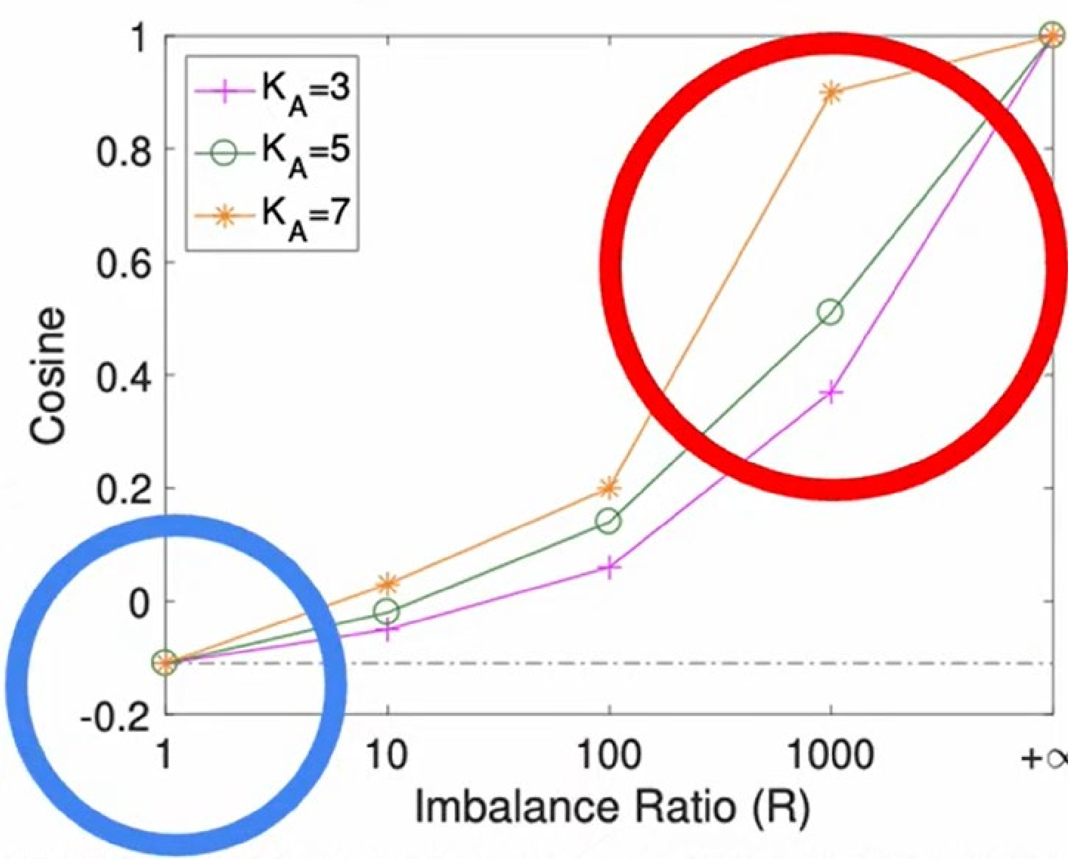
- Recently, minority collapse, the phenomenon in which angles between classifier vectors for minor classes become narrow, has been proposed as a fundamental issue in training with imbalanced data

  (minor class: the class that have relatively small # of samples)

- Imbalanced Ratio (R)

$$= \frac{\#\ of\ samples\ for\ Major\ classes}{\#\ of\ samples\ for\ Minor\ classes}$$

Avg Cosine similarity between classifier vectors of minor classes



[1] Fang et al., Exploring Deep Neural Networks via Layer-Pedeled Model: MinorityCollapse in Imbalanced Training, PNAS 2021

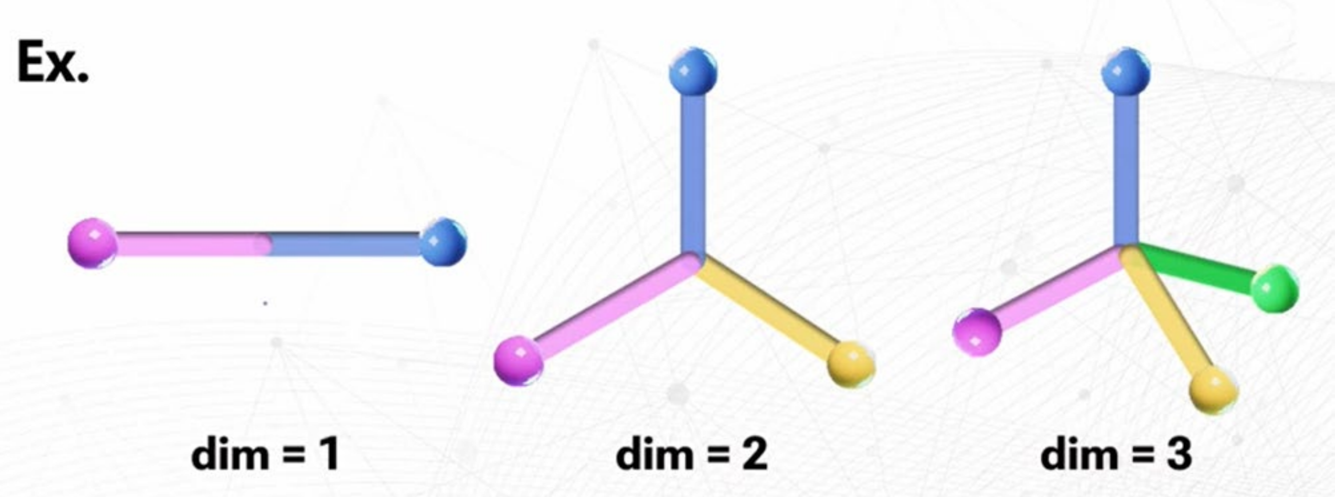
Neural Collapse [1] in Balanced Distribution

- Sufficient training on a balanced dataset leads to Neural Collapse.

➢ A phenomenon that classifier vectors and the last layer actiivations for all classesconverge into an optimal geometric structure, named thesimplex equiangular tightframe (ETF)

Neural Collapse [1] in Balanced Distribution

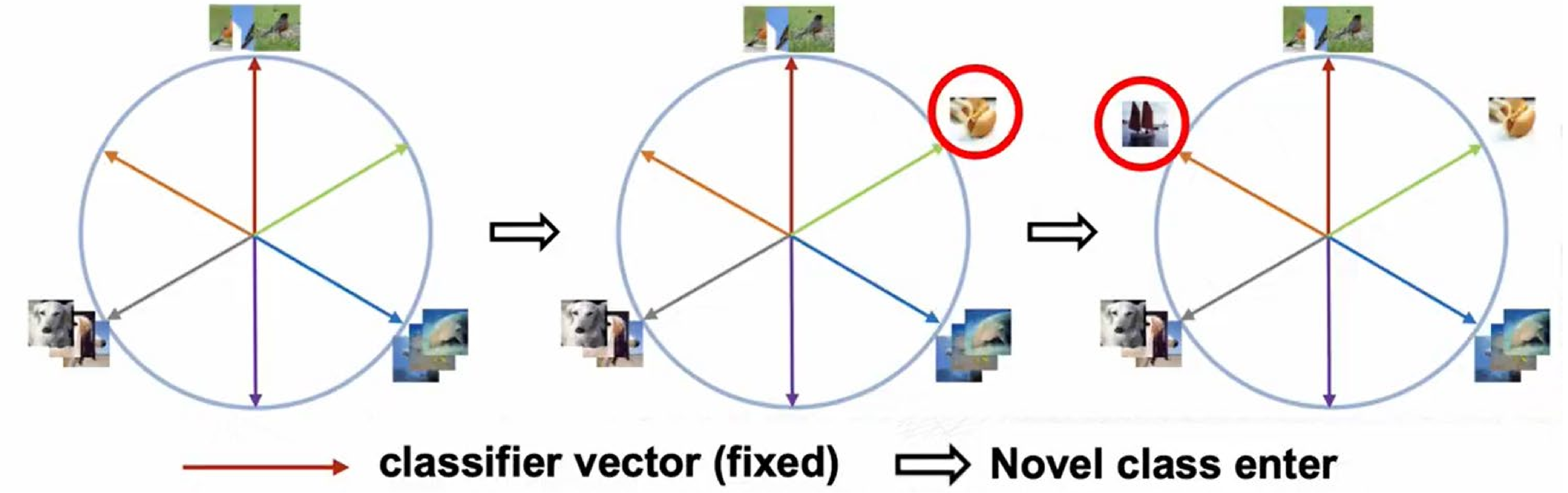
- It was proven that sufficient training on a balanced dataset leadds to Neural Collapse, aphenomenon that classifier vectors and the last layer activations for all classes convergeinto an optimal geometric structure, named the simplex equiangular tight frame (ETF)
- $\mathbf{W}_{\mathrm{ETF}} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K] \in \mathbb{R}^{d \times K}$, which satisfy $\mathbf{w}_i^T \mathbf{w}_j = \begin{cases} 1, & i = j \\ -\frac{1}{K-1}, & i \neq j \end{cases}$
- Ex.



[1] Fang et al., Exploring Deep Neural Networks via Layer-Pedeled Model: MinorityCollapse in Imbalanced Training, PNAS 2021

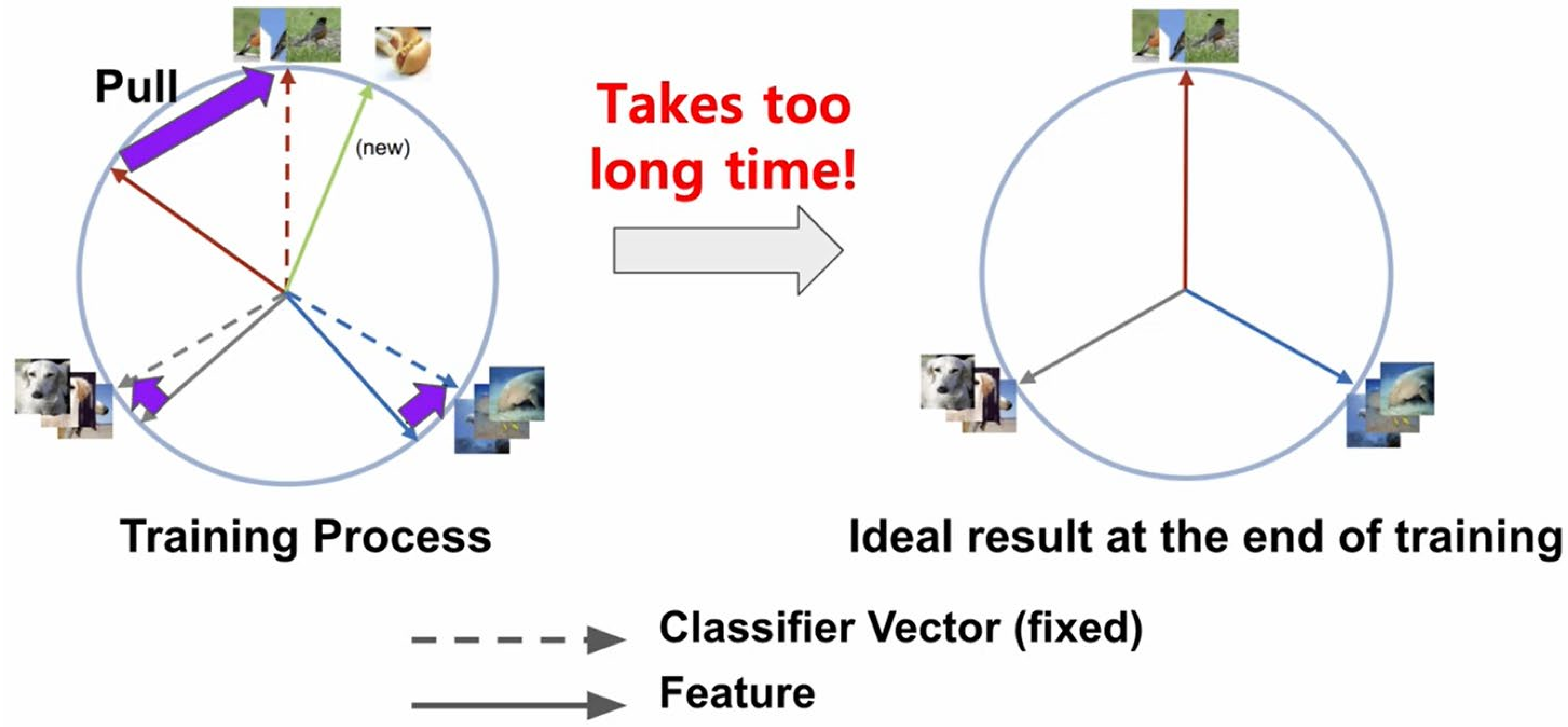
Can we induce neural collapse in CL setups?

Inducing NC in Offline CL Using a Fixed ETF Classifier [1]



[1] Yang et al., Neural Collapse Inspired Feature-Classifier Aliggnment for Few-Shot ClassIncremental Learning, ICLR 2023 Spotlight

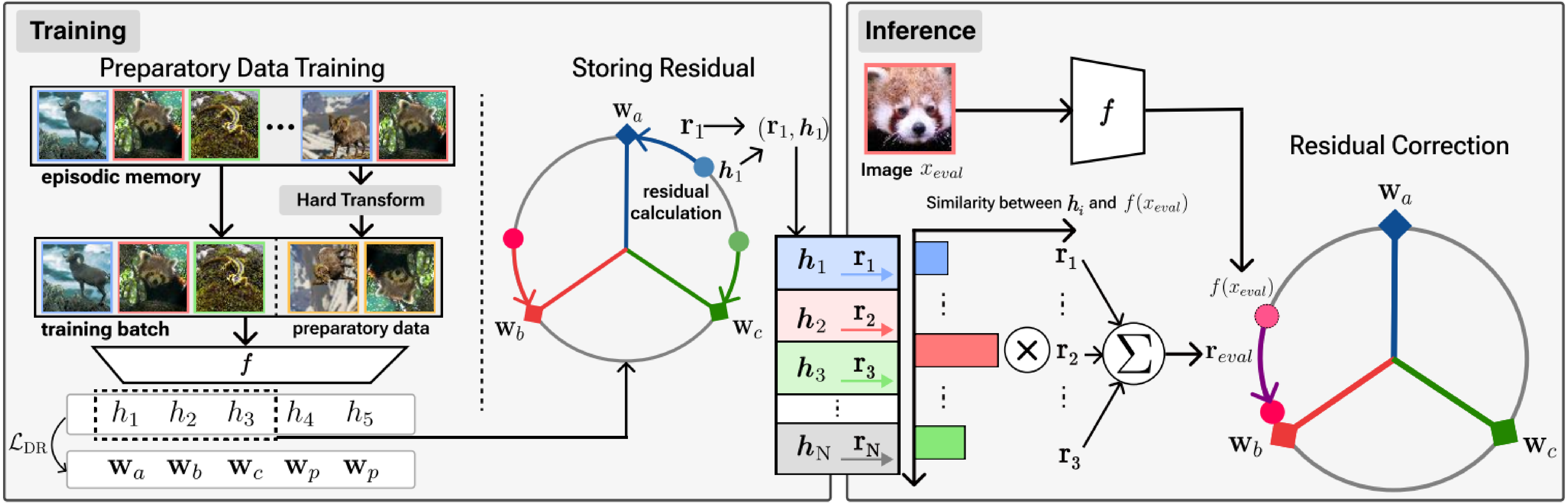
Training: Pulling Features to a Fixed ETF Classifier



Goal

Promote inducing neural collapse in Online CL!

Overview of EARL: Equi-Angular Representation Learning



Training Phase - Preparatory Data Training (PDT)

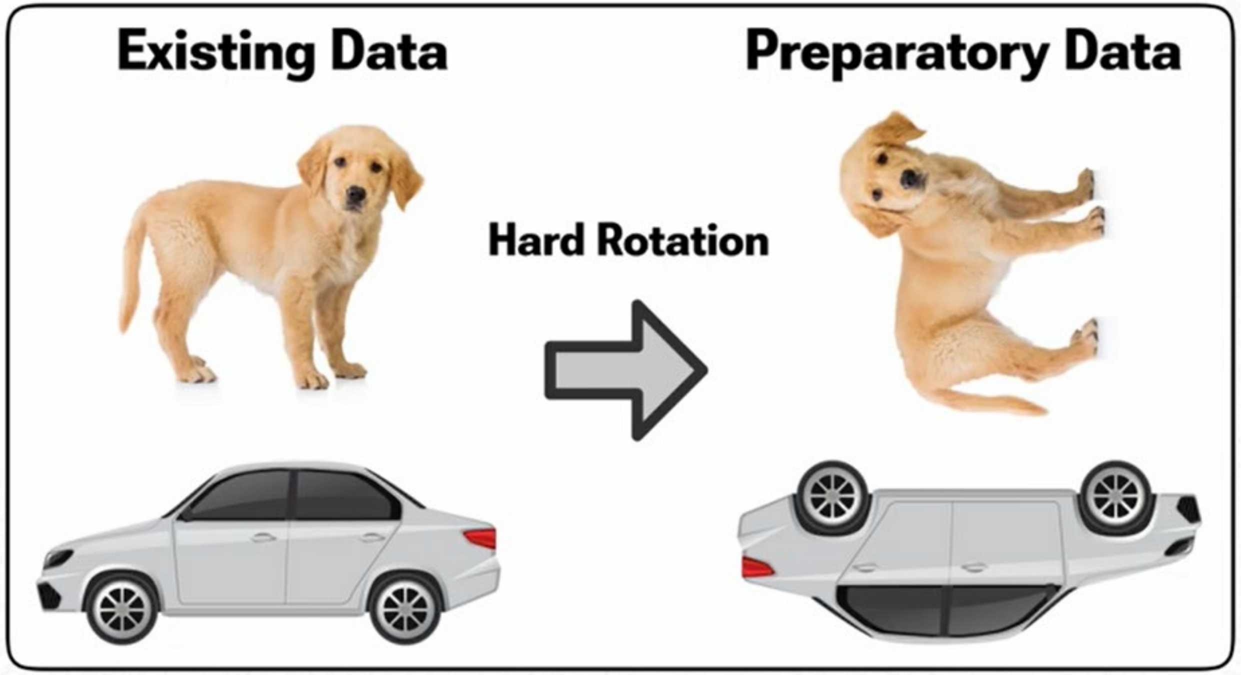
ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

- To mitigate bias, it is ideal to pull data of the class that will beintroduced in the future to another classifier in advance.
  => However, we can't know which class will encounter
- Therefore, we train the model to recognize that data distinguished from existing classes should not be classified to those existing classees.
  => Named **Preparatory Data**
- Pulling features of preparatory data to remaining classifier vecttors
- **Preventing biased predictions when new classes arrive**
  => facilitating rapid convergence to the ETF classifier

Obtaining Preparatory Data Using Negative Transformation

Negative Transformation [1, 2]

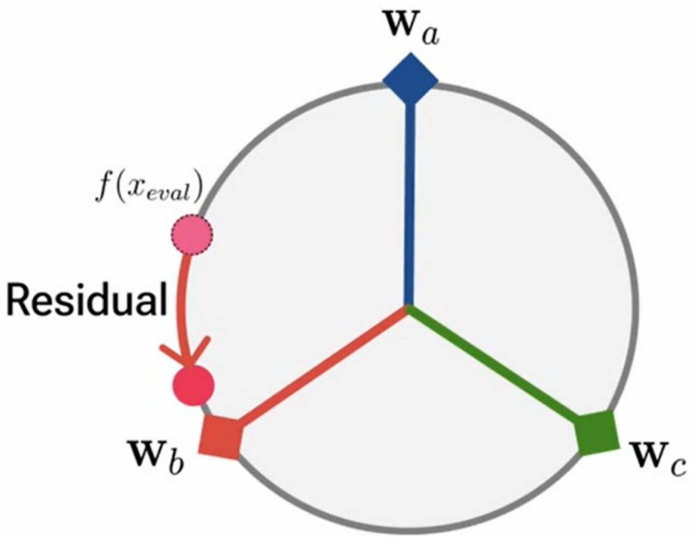
- Transformation to make animage that is distinct from theexisting image
- We use rotation 90, 180, and 270
- ex. dog => upside-down dog



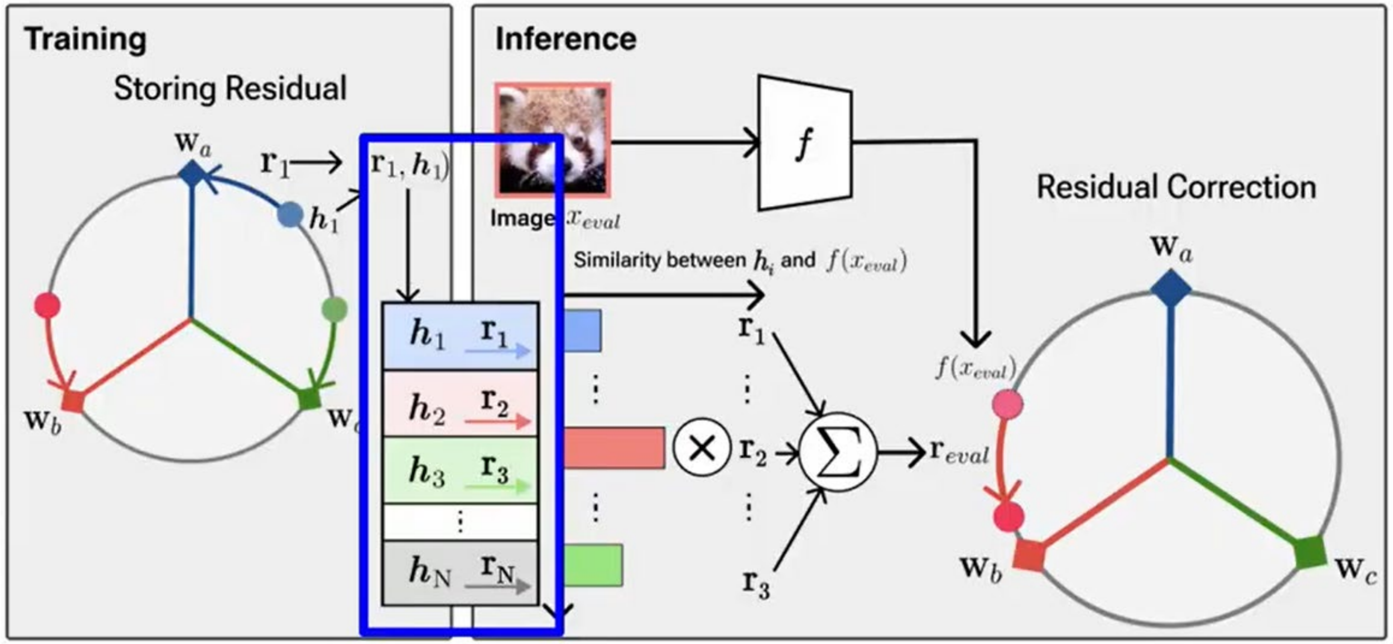
[1] Sinha et al., Negative data augmentation, arXiv 2021
[2] Wang et al., Resmooth: Detecting and utilizing ood samples wheen training with data augmentation. IEEE TNNLS, 2022

Inference Phase - Residual Correction

- Even with preparatory data training, there mayexist discrepancy between newly encounteredclass and corresponding classifiers



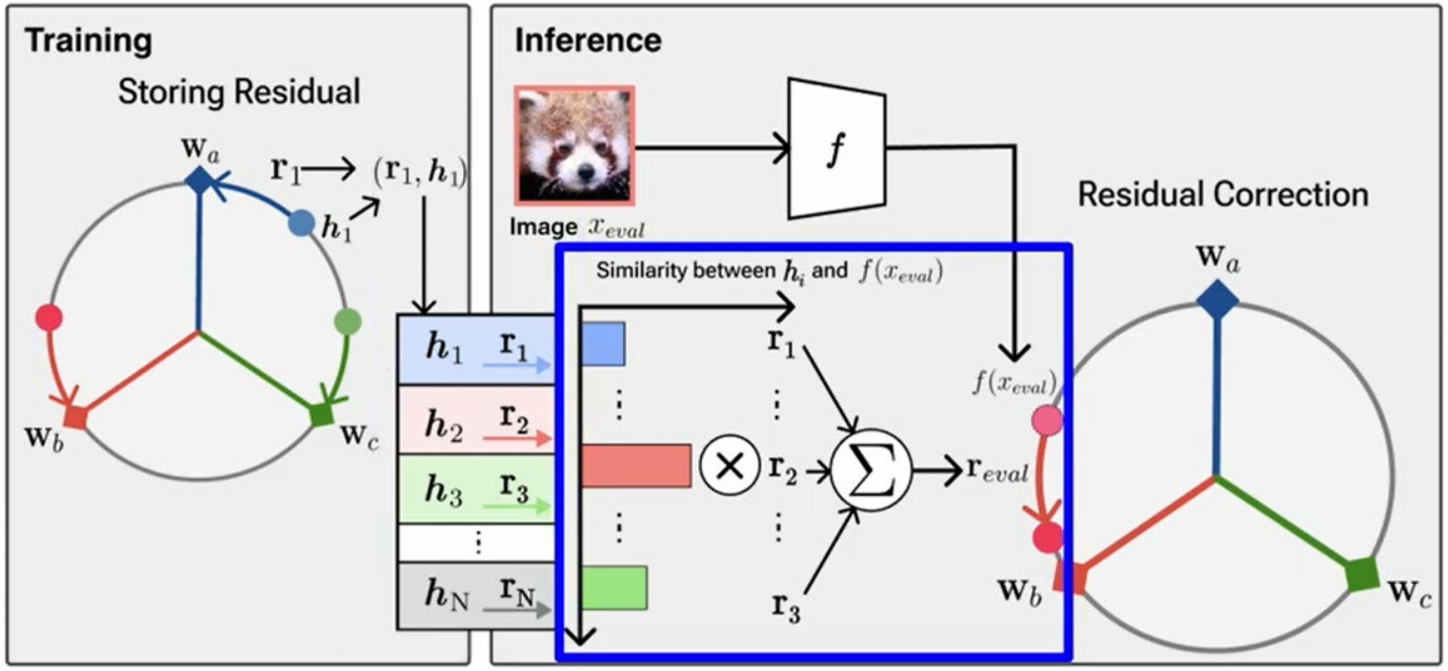
- (feature, residual) pairs arestored in feature-residual Memory during training,



f: model, h: model output, w: ETF classifier vector, each color denotes each classes

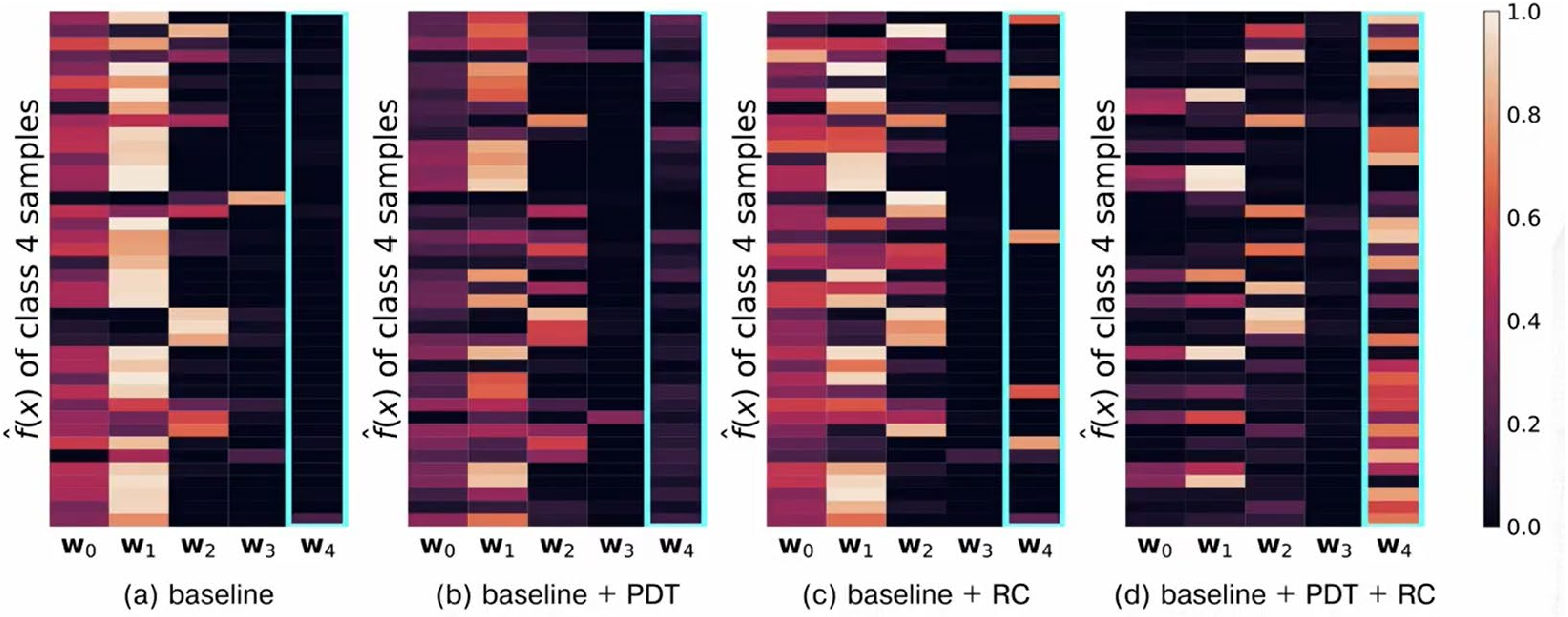
Inference Phase - Residual Correction

- To estimate correct residual, we use similarity btw stored features and inference feature



f: model, h: model output, w: ETF classifier vector, each color denotes each classes

Results: Benefit of the Proposed Components



- PDT and RC increases alignment features and the GT classifier vectors between

Results: Quantitative Analysis

Methods	CIFAR-10				CIFAR-100			
	Disjoint		Gaussian-Scheduled		Disjoint		Gaussian-Scheduled	
	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$
ER	75.94±0.86	63.56±1.32	60.13±0.56	64.81±2.70	52.95±1.25	42.82±0.05	41.12±0.56	42.74±1.09
DER++	74.57±0.89	60.80±1.31	59.88±0.37	64.75±2.29	54.51±1.18	42.86±0.63	43.28±0.57	44.60±1.46
ER-MIR	75.89±1.02	61.93±0.93	60.39±0.48	61.64±3.86	52.93±1.44	42.47±0.13	41.19±0.63	42.93± 1.18
SCR	75.61±0.93	56.52±0.52	60.62±0.43	58.41±2.39	41.84±0.74	36.00±0.83	31.33±0.41	32.11±0.39
EWC	75.25±0.78	60.80±2.20	59.62±0.31	64.24±1.97	52.08±0.83	41.55±0.85	38.22±0.50	42.52±0.58
REMIND	69.55±0.91	53.34±1.01	58.01±0.72	59.27±1.86	40.87±0.76	36.17±1.83	23.40±2.25	28.78±1.71
X-DER	74.34±0.40	62.31±2.28	57.05±3.76	62.89±2.71	52.80±1.61	43.73±0.86	41.94±0.57	44.90±1.04
ODDL	75.03±1.00	61.61±3.55	65.46±0.46	66.19±2.08	40.26±0.50	41.88±4.52	38.82±0.49	41.35±1.08
MEMO	73.21±0.49	62.47±3.38	59.26±0.90	62.01±1.17	40.60±1.11	39.87±0.46	23.41±1.63	32.74±2.11
EARL (Ours)	**79.62±0.62**	**67.58±1.51**	**70.56±0.37**	**71.46±1.00**	**57.12±1.22**	**45.15±0.68**	**48.05±0.49**	**46.59±0.35**

Methods	TinyImageNet				ImageNet-200			
	Disjoint		Gaussian-Scheduled		Disjoint		Gaussian-Scheduled	
	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$	$A_{\text{AUC}} \uparrow$	$A_{last} \uparrow$
ER	37.43±1.05	27.47±0.63	26.37±0.89	25.79±0.44	41.51±0.76	30.87±0.72	32.39±0.36	33.09±0.37
DER++	38.05±1.07	25.41±0.50	31.04±0.67	27.68±0.77	43.20±0.31	34.06±0.50	35.22±0.26	37.88±0.97
ER-MIR	37.81±1.06	26.72±0.86	26.22±0.69	25.11±1.04	38.28±0.38	33.12±0.73	32.17±0.44	33.85±0.93
SCR	34.65±1.08	22.18±0.32	25.86±0.94	22.54±0.59	41.90±0.40	28.92±0.40	33.24±0.32	30.98±0.28
EWC	37.95±0.93	27.50±0.80	25.29±0.81	26.06±0.52	41.84±0.64	31.57±0.80	30.71±0.27	33.33±0.98
REMIND	28.37±0.13	27.68±0.45	10.19±0.60	14.90±1.49	39.25±0.93	31.98±0.84	30.23±0.62	33.98±0.09
X-DER	35.15±2.12	26.67±0.52	29.71±0.86	28.10±0.50	43.21±0.47	33.84±0.98	36.31±0.17	38.31±0.55
MEMO	27.36±0.61	27.57±0.52	10.82±1.23	18.03±1.36	41.55±0.23	34.19±1.47	32.54±0.39	36.11±1.06
EARL (Ours)	**41.77±1.26**	**29.65±0.20**	**35.08±0.70**	**32.49±1.21**	**44.88±0.29**	**34.27±0.55**	**39.14±0.47**	**38.83±0.35**

Thanks