Information Bottleneck Based Data Correction in Continual Learning

Institute of Automation
Chinese Academy of Sciences

Shuai Chen, Mingyi Zhang, Junge Zhang, and Kaiqi Huang

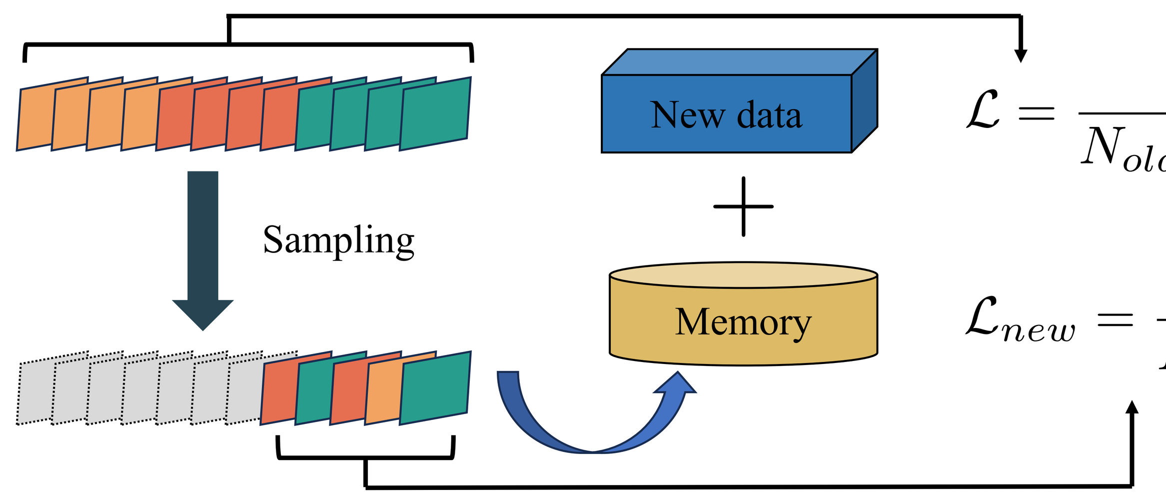
CRISE, Institute of Automation, Chinese Academy of Sciences

School of Artificial Intelligence, University of Chinese Academy of Sciences

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

Motivation:

- The scarcity of old data can create an imbalance between the number of new and old tasks, disrupting the retention of knowledge from the old tasks during the learning process of new tasks.
- During the continual learning process, each time old task data is sampled, some unsampled data is discarded, which introduces bias into the model's estimation of the old task distribution.



$$\mathcal{L} = \frac{1}{N_{old} + N_{new}} \sum_{i=1}^{N} \ell_i(\hat{Y}, Y)$$

$$\not\parallel$$

$$\mathcal{L}_{new} = \frac{1}{N_{buffer} + N_{new}} \sum_{i=1}^{N_n} \ell_i(\hat{Y}, Y)$$

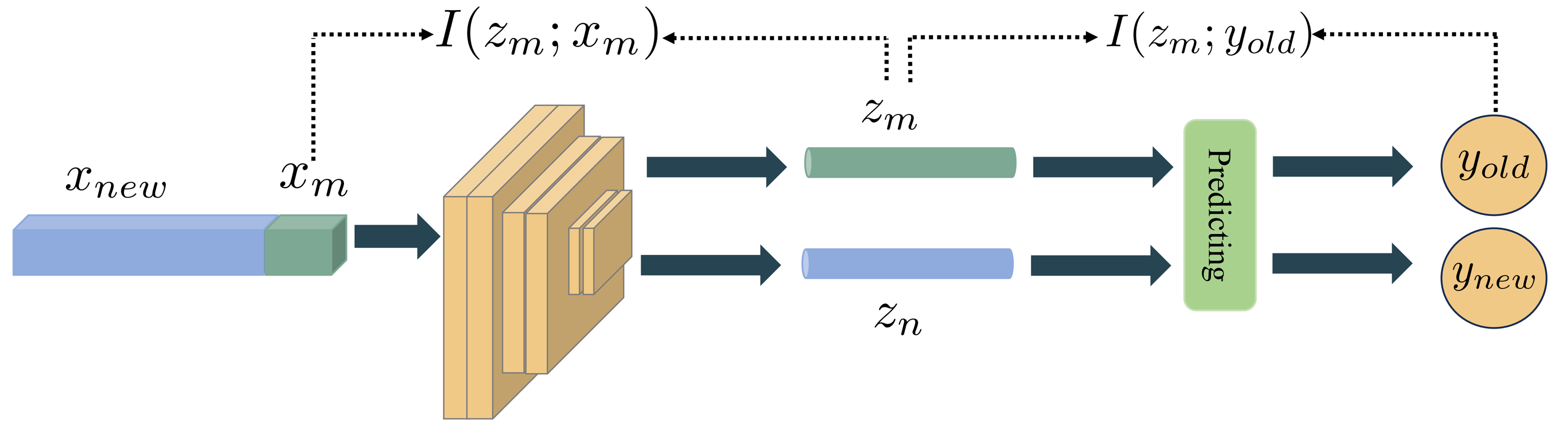
Methods:

- During the update process for new tasks, we introduced a constraint in the learning of buffered data. This constraint encourages the model to learn task-agnostic features associated with old tasks, effectively preventing interference from new tasks.
- After completing the learning of previous tasks, this module models the information of unsampled data, which cannot be directly trained. Our method decouples the features of unsampled data from those of sampled data, enabling the model to leverage the relationship between the two data types to estimate the surrogate influence of unsampled data.

Information Bottleneck Task Against Constraints:

- This module aims to maximize the mutual information between the sampled data features z_m and the buffered dataset x_m to effectively compress information. Furthermore, z_m must accurately predict the labels y_{old} to ensure class distinction, with a weighted parameter α balancing contributions in the objective function.

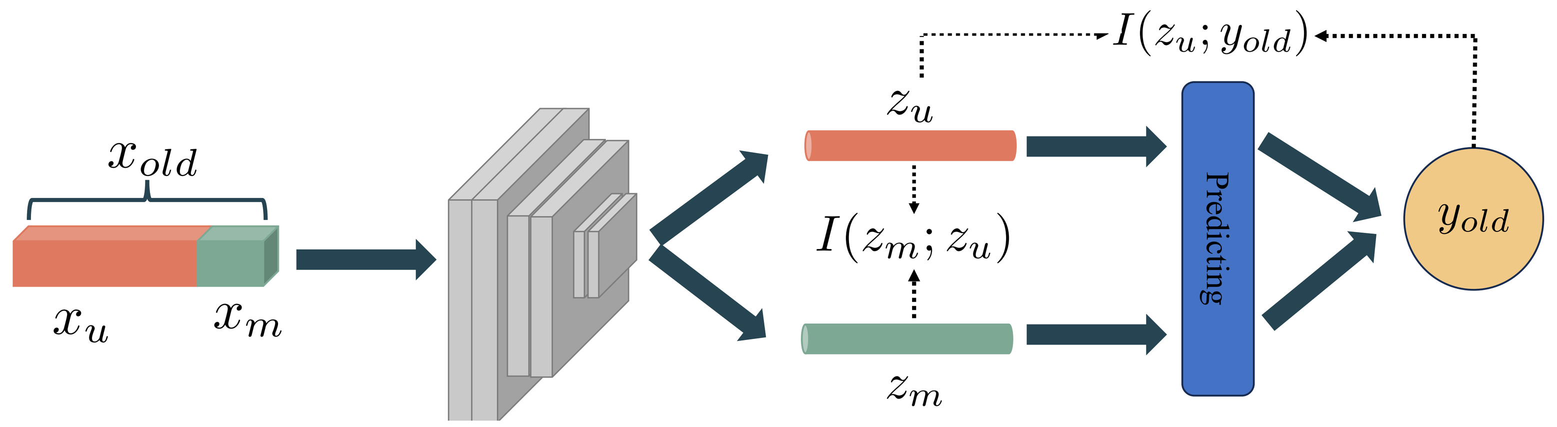
$$\mathcal{L}_{IBTA} := \min \alpha \underbrace{I(z_m; x_m)}_{(1)} - \underbrace{I(z_m; y_{\text{old}})}_{(2)} .$$

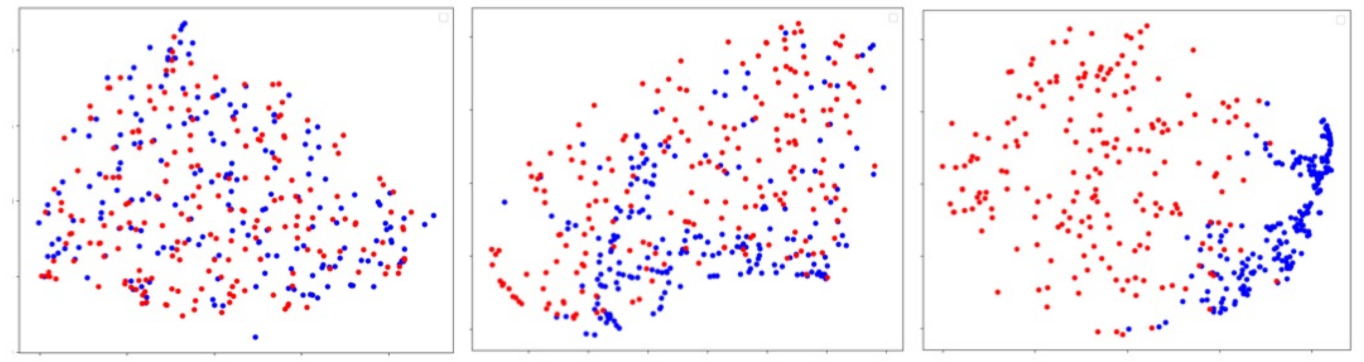


Information Bottleneck Unsampled Data Surrogate:

- This module aims to model information from unsampled data. The method decouples features of unsampled data x_u from those of sampled data x_m after completing the learning of prior tasks. By optimizing this formula, we get the surrogate expression of the unsampled data z_u.

$$\mathcal{L}_{IBDS} := \min I(z_m; z_u).$$





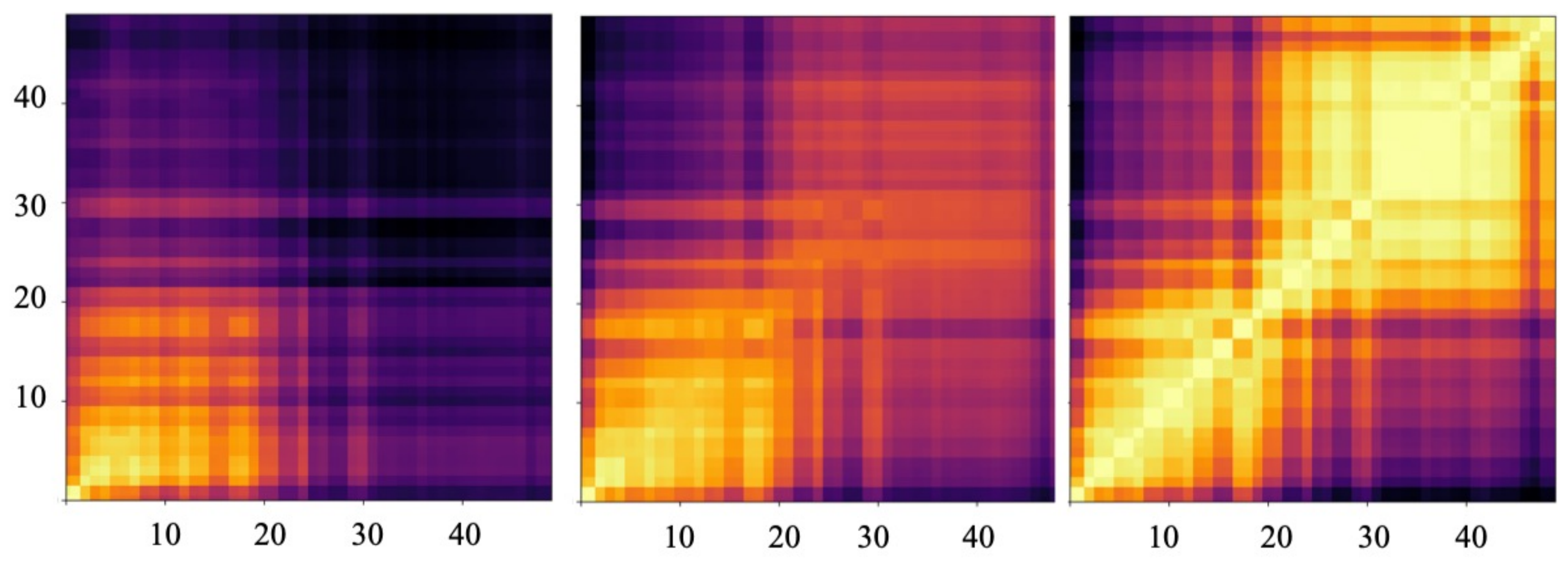
Initialization | Epoch = 20 | Epoch = 50

Results:

- Our method, as a plug-and-play module, can enhance the performance of existing replay-based continual learning approaches..

Method	Split CIFAR-10			Split CIFAR-100			Split ImageNet-100			Split miniImageNet		
Joint	92.38			73.29			80.23			53.55		
FT	19.67			9.29			8.68			9.52		
Buffer size	100	200	500	200	500	2000	200	500	2000	1000	2000	5000
ER	36.39	44.79	57.74	14.35	19.66	36.76	13.63	18.37	34.25	8.37	16.49	24.17
+IBCL	45.43	51.91	63.12	23.98	28.32	42.01	22.72	28.93	43.11	14.79	22.72	27.69
ER-ACE	53.90	63.41	70.53	26.28	36.48	48.41	24.23	37.12	49.55	17.95	22.60	27.92
+IBCL	63.85	70.97	74.82	32.12	40.94	51.89	30.72	41.81	52.62	24.77	27.63	31.02
DER++	57.65	64.88	72.70	25.11	37.13	52.08	26.50	43.65	58.05	18.02	23.44	30.43
+IBCL	66.41	72.52	76.61	33.82	44.21	54.89	32.63	46.57	59.92	27.72	31.41	35.79
X-DER	59.29	65.19	68.10	35.34	44.62	54.44	33.21	46.72	55.23	25.24	26.38	29.91
+IBCL	67.91	73.82	74.14	43.86	48.72	55.91	43.25	49.62	56.59	29.90	30.21	32.66

Method	Split CIFAR-10			Split CIFAR-100			Split miniImageNet		
Buffer size	100	200	500	200	500	2000	1000	2000	5000
ER-ACE	53.90	63.41	70.53	26.28	36.48	48.41	17.95	22.60	27.92
+ sSGD	56.26	64.73	71.45	28.07	39.59	49.70	18.11	22.43	24.12
+ oEwC	52.36	61.09	68.70	24.93	35.06	45.59	19.04	24.32	29.46
+ oLAP	52.76	63.19	70.32	26.42	36.58	47.66	18.34	23.19	28.77
+ OCM	57.18	64.65	70.86	28.18	37.74	49.03	20.32	24.32	28.57
+ LiDER	56.08	65.32	71.75	27.94	38.43	50.32	19.69	24.13	30.00
+ DualHSIC	60.52	68.08	73.78	29.08	38.94	50.55	22.33	25.41	30.12
+ Our	63.85	70.97	74.82	32.12	40.94	51.89	24.77	27.63	31.02
DER++	57.65	64.88	72.70	25.11	37.13	52.08	18.02	23.44	30.43
+ sSGD	55.81	64.44	72.05	24.76	38.48	50.74	16.31	19.29	24.24
+ oEwC	55.78	63.02	71.64	24.51	35.22	51.53	18.87	24.53	31.91
+ oLAP	54.86	62.54	71.38	23.26	34.48	50.80	18.91	25.02	32.78
+ OCM	59.25	65.81	73.53	27.46	38.94	52.25	20.93	24.75	31.16
+ LiDER	58.43	66.02	73.39	27.32	39.25	53.27	21.58	28.33	35.04
+ DualHSIC	64.98	70.28	75.94	31.46	41.86	53.53	24.78	29.37	34.98
+ Our	66.41	72.52	76.61	33.82	44.21	54.89	27.72	31.41	35.79



FT | FT w/Our | ER