DELTA: Decoupling Long-Tailed Online Continual Learning

Siddeshwar Raghavan, Jiangpeng He, Fengqing Zhu

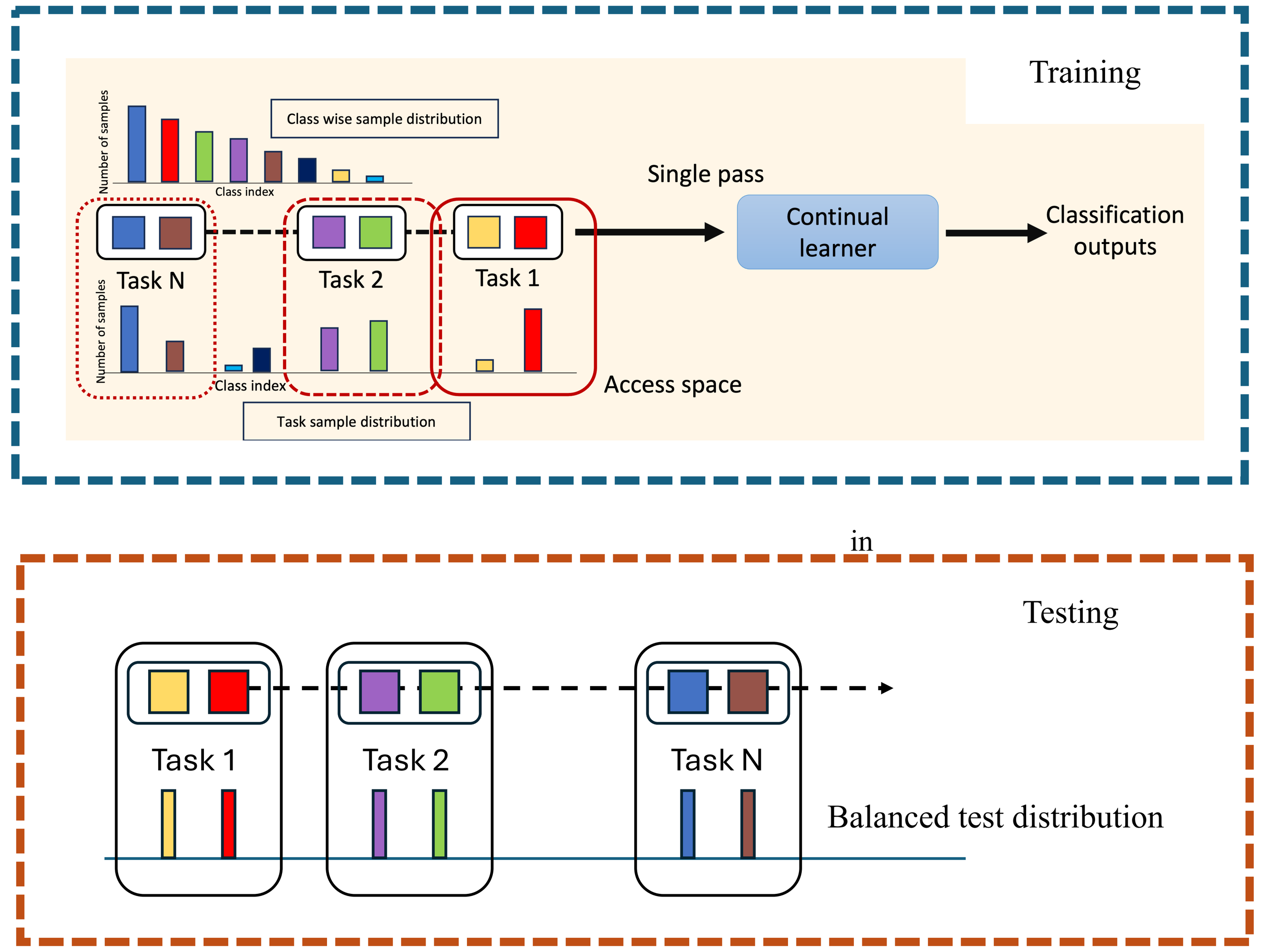
Elmore Family School of Electrical and Computer Engineering, Purdue University

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

CVPR JUNE 17-21, 2024 SEATTLE, WA

Long-Tailed Online Continual Learning



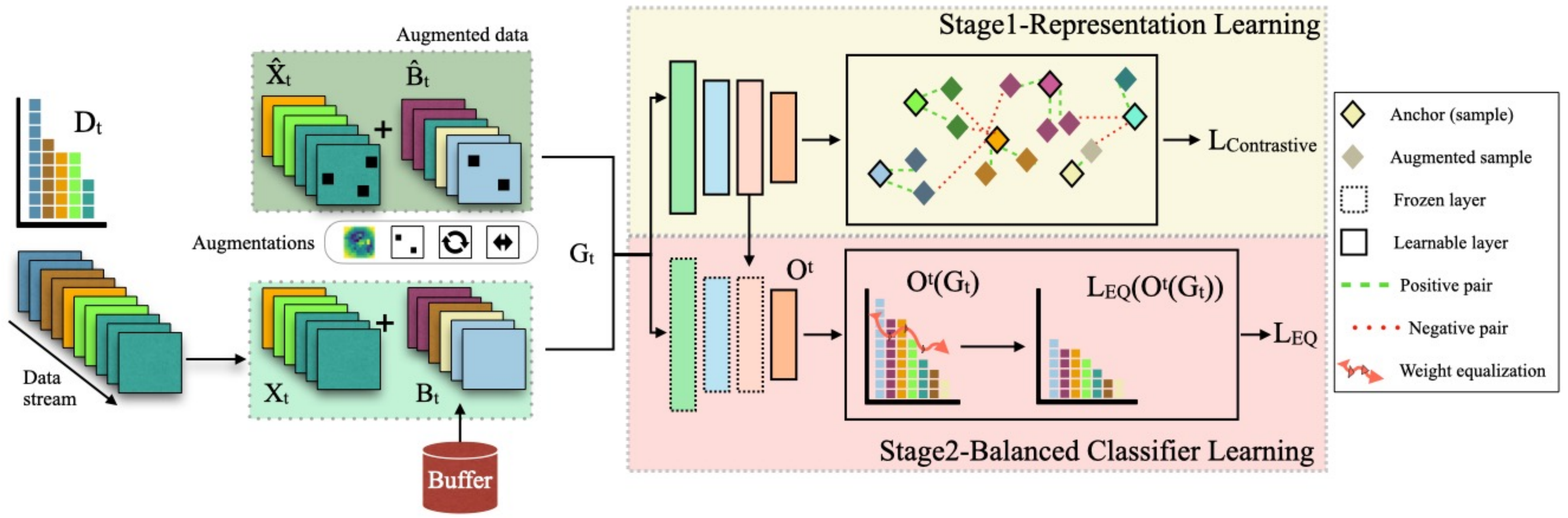
Challenge

- **Imbalanced class representation** - Dealing with real-world scenarios where the data follows long-tailed distributions where each class may not be equally represented.
- **Catastrophic forgetting and model bias** - Limited set of exemplars can be biased towards the frequently appearing classes.
- **Exemplar Management and Data Efficiency**: Utilizing existing continual learning models can introduce distribution challenges and biases, making it difficult to adapt to evolving data distributions while maintaining balanced performance across all classes.

Outcome

- Develop a **two-stage decoupling strategy** that effectively handles real-world distribution scenarios, ensuring freedom from biases and enhancing performance outcomes.
- We propose a **multi-exemplar pairing strategy** to demonstrate the potential performance enhancement in Long-Tail Online Continual Learning

Methodology



Decoupling learning through two stage training.

Stage 1

$$\mathcal{L}_{stage1}(O^t(G_t)) = \mathcal{L}_{contrastive}(O^t(G_t))$$

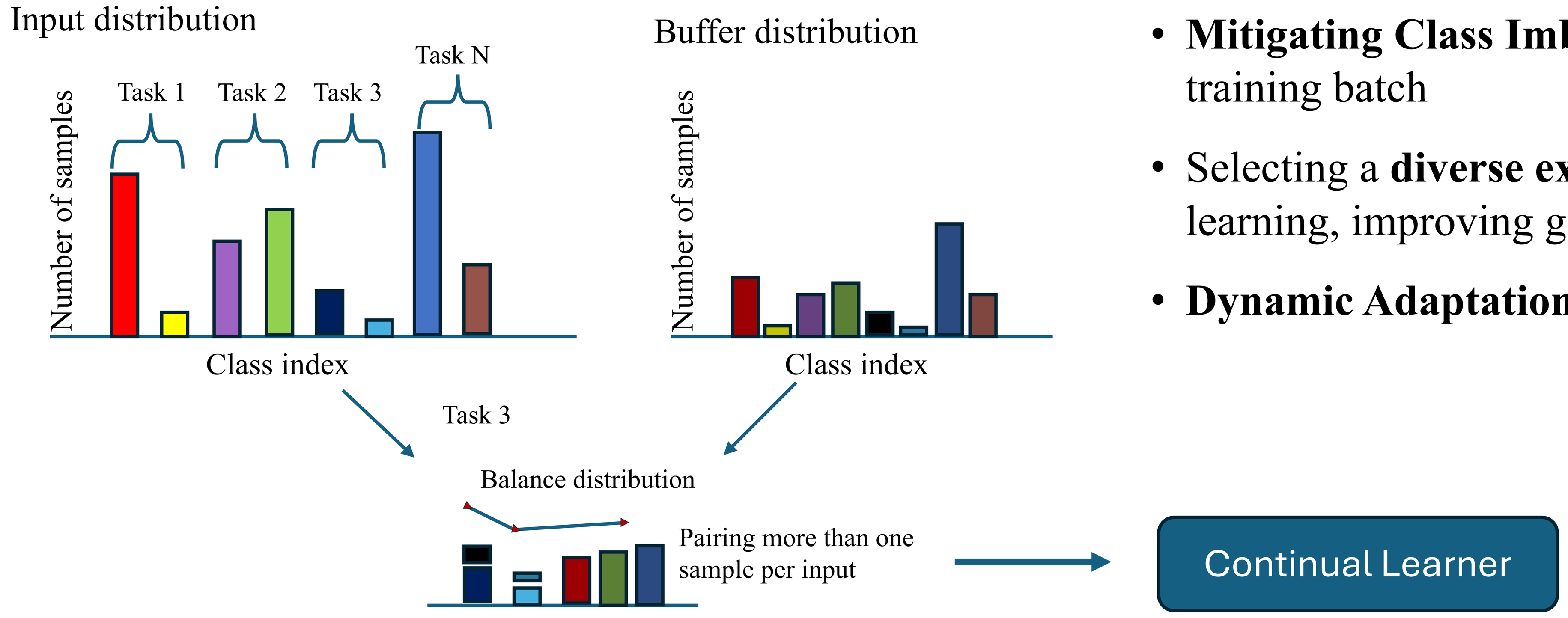
$$\mathcal{L}_{contrastive}(Z_T) = \sum_{j \in T} \frac{-1}{|P(j)|} \sum_{p \in P(j)} \frac{exp(v_j \cdot v_p/\tau)}{\sum_{k \in A(j)} exp(v_j \cdot v_p/\tau)}$$

Stage 2

$$\mathcal{L}_{stage2}(O^t(G_t)) = \mathcal{L}_{EQ}(O^t(G_t))$$

$$\mathcal{L}_{EQ}(O^t(I_x)) = \sum_{i=1}^{k^{1:t}} -I_{y_{(i)}} \sigma([\log[(P(k^t)] + O^t(I_x))])$$

Multi-exemplar pairing



Motivation

- **Mitigating Class Imbalance** and balance the input training batch
- Selecting a **diverse exemplar** set ensures better learning, improving generalization and performance.
- **Dynamic Adaptation** to varying data distribution.

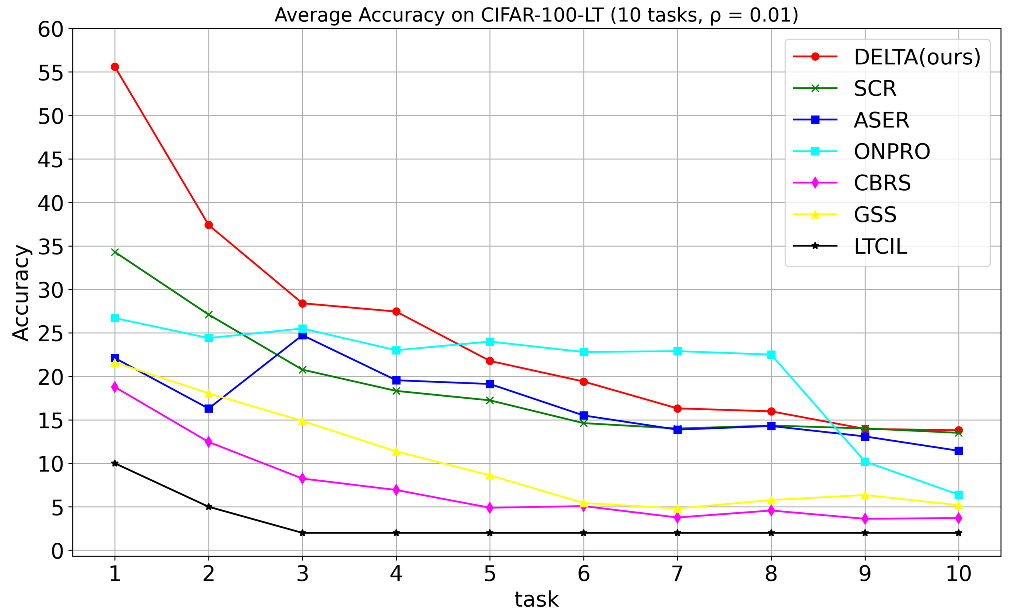
Experiments

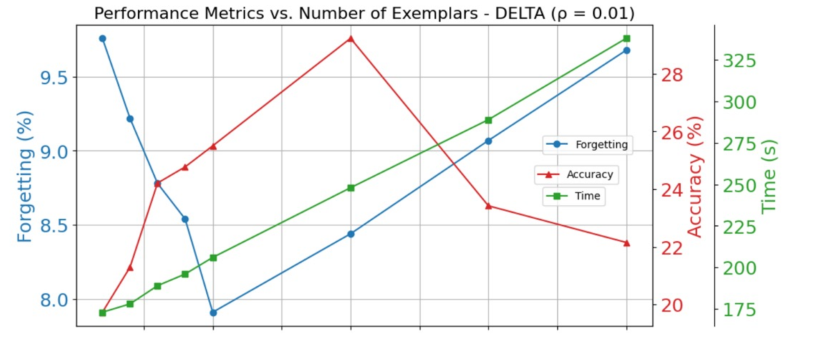
Methods	CIFAR100-LT						VFN-LT					
	20 tasks	20 tasks	20 tasks	10 tasks	10 tasks	10 tasks	15 tasks	15 tasks	15 tasks	7 tasks	7 tasks	7 tasks
	M=0.5K	M=1K	M=2K	M=0.5K	M=1K	M=2K	M=0.5K	M=1K	M=2K	M=0.5K	M=1K	M=2K
OnPRO[ICCV '23]	14.02 ± 0.44	16.28 ± 0.81	18.01 ± 0.22	16.53 ± 0.55	16.92 ± 0.08	18.85 ± 0.32	11.93 ± 0.04	12.77 ± 0.07	13.50 ± 0.05	**8.02 ± 0.60**	9.38 ± 0.21	11.84 ± 0.49
SCR[CVPRW '21]	12.22 ± 0.72	13.48 ± 0.90	15.88 ± 0.79	16.65 ± 0.90	17.02 ± 0.77	17.58 ± 0.66	11.55 ± 0.17	11.82 ± 0.10	12.39 ± 0.73	7.71 ± 0.49	9.19 ± 0.46	9.48 ± 0.47
ASER[AAAI '21]	8.86 ± 0.30	7.86 ± 0.61	8.18 ± 0.31	12.68 ± 0.70	13.76 ± 0.01	15.90 ± 0.91	6.85 ± 0.34	7.61 ± 0.38	7.22 ± 0.36	7.46 ± 1.18	7.52 ± 1.09	6.35 ± 0.19
PRS[ECCV '20]	7.61 ± 0.09	7.54 ± 0.21	7.03 ± 0.13	7.34 ± 0.92	8.95 ± 0.33	9.01 ± 0.39	7.17 ± 0.83	8.72 ± 0.15	8.39 ± 0.19	7.85 ± 0.50	8.66 ± 0.22	9.21 ± 0.30
CBRS[ICML '20]	8.51 ± 0.19	8.66 ± 0.61	8.91 ± 0.33	9.50 ± 0.48	7.22 ± 0.43	7.31 ± 0.08	8.12 ± 0.94	8.35 ± 0.33	8.18 ± 0.44	7.52 ± 0.11	7.64 ± 0.08	7.92 ± 0.34
GSS[NeurIPS '19]	5.16 ± 0.10	5.22 ± 0.22	5.09 ± 0.21	8.97 ± 0.65	10.12 ± 0.02	9.96 ± 0.47	5.86 ± 0.30	6.01 ± 0.91	5.86 ± 0.06	5.92 ± 0.54	4.30 ± 0.22	4.66 ± 0.60
LT-CIL(offline)	3.01 ± 0.77	2.67 ± 0.04	2.43 ± 0.02	1.76 ± 0.11	2.36 ± 0.25	3.76 ± 0.22	1.82 ± 0.45	2.02 ± 0.44	2.38 ± 0.08	3.08 ± 0.71	2.92 ± 0.04	1.99 ± 0.31
DELTA (ours)	**16.53 ± 0.01**	**17.71 ± 0.11**	**19.93 ± 0.07**	**20.25 ± 0.71**	**21.06 ± 0.23**	**22.47 ± 0.51**	**12.5 ± 0.01**	**13.45 ± 0.02**	**13.84 ± 0.01**	8.00 ± 0.39	**10.41 ± 0.52**	**12.84 ± 0.54**

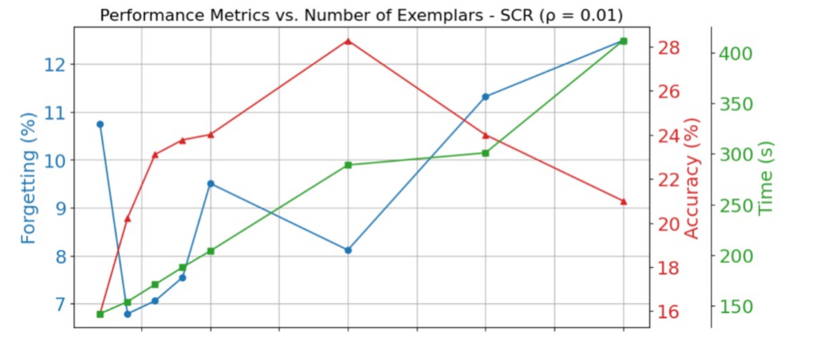
Average Accuracy ± standard deviation (↑) in the long-tailed scenario on Split CIFAR-100-LT, VFN-LT with single exemplar pairing. The highest accuracy is highlighted in red, while the second highest is in blue.

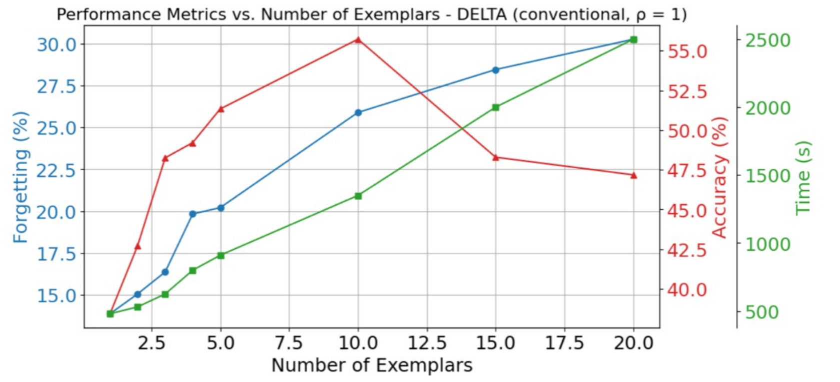
Imbalance ratio (ρ)	SCR	CBRS	DELTA *(ours)*
0.005	3.59 ± 0.64	7.60 ± 0.06	**18.02 ± 0.79**
0.03	4.11 ± 0.42	8.88 ± 0.47	**20.21 ± 0.22**
0.07	8.34 ± 0.04	9.47 ± 0.61	**23.60 ± 0.09**
0.1	6.12 ± 0.35	10.26± 0.38	**24.28 ± 0.60**
1.0 (Conventional)	20.74 ± 0.40	15.79 ± 0.33	**33.23 ± 0.97**

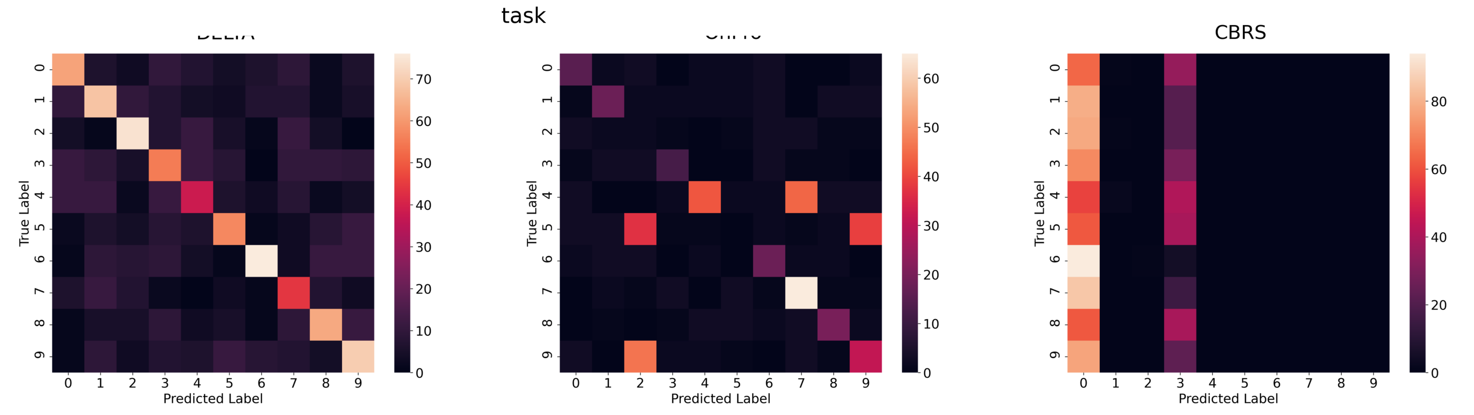
Average Accuracy ± standard deviation (↑) the long-tailed scenario on Split CIFAR-100-LT with varying imbalance factors. Best results in **bold**











Paper

Code

Contact:

Sid Raghavan

raghav12@purdue.edu