Gradient Reweighting: Towards Imbalanced Class-Incremental Learning

Jiangpeng He

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

VIPER

Video and Image Processing Laboratory

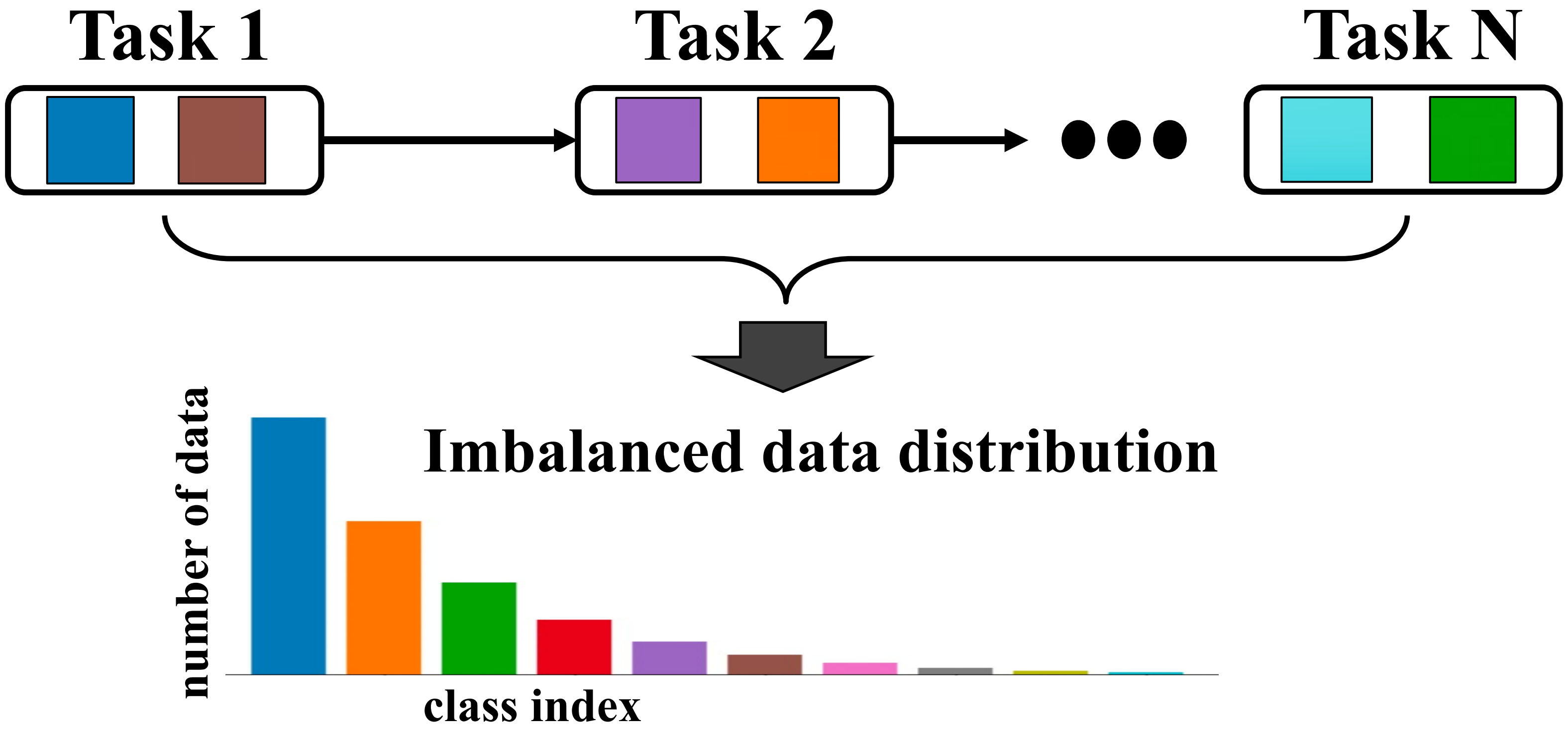
CVPR

JUNE 17-21, 2024

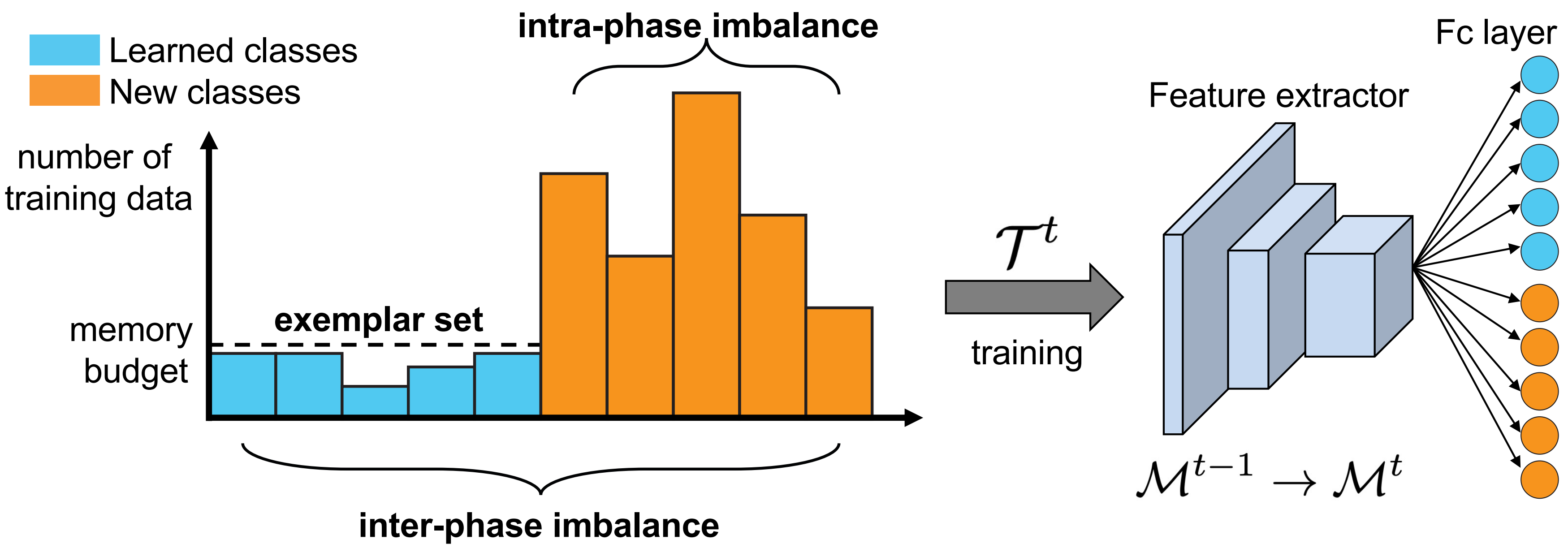
SEATTLE, WA

Imbalanced Class-Incremental Learning

- **Learning classes incrementally from non-uniform data distribution**



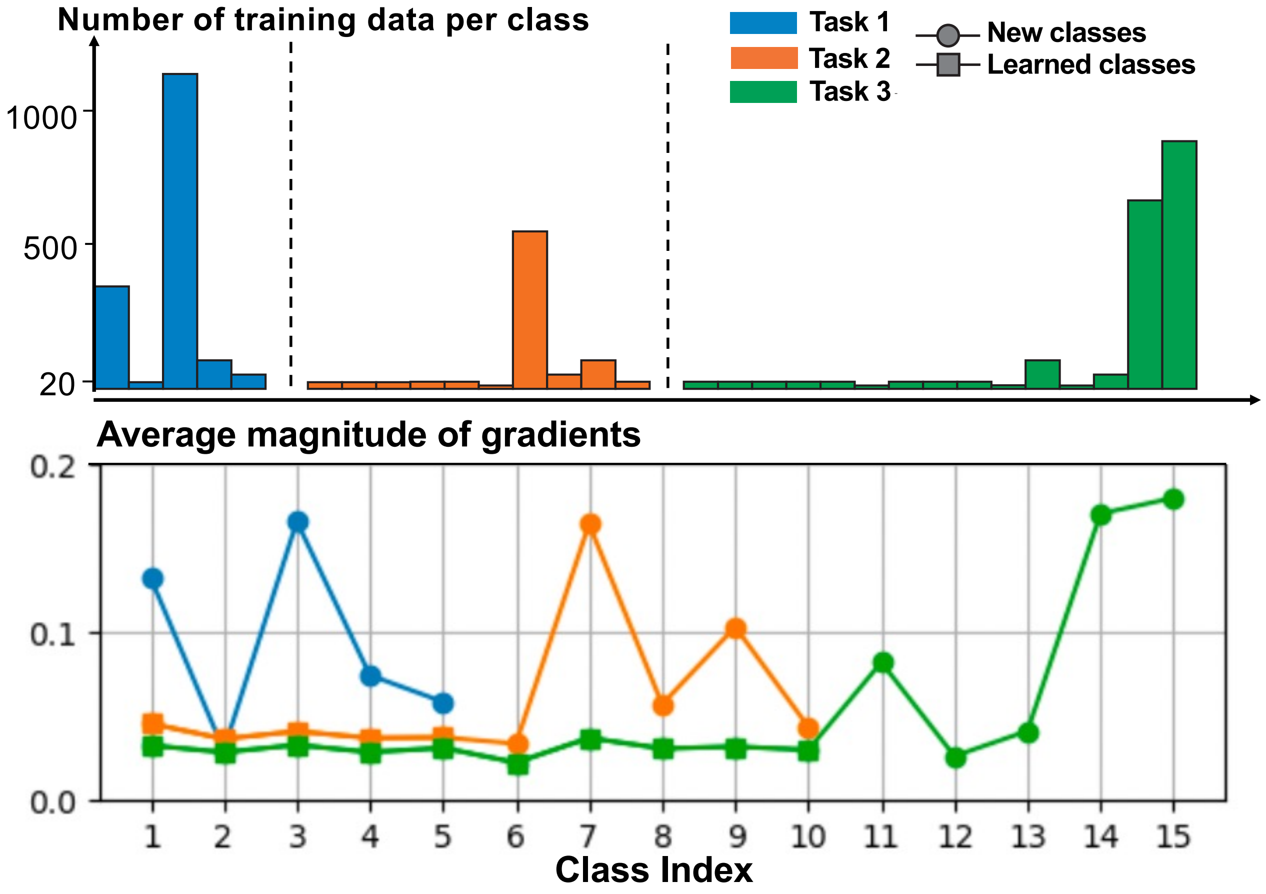
Challenge - Dual Imbalance



- ***<u>Inter-phase imbalance</u>*: disparities between stored exemplars of old tasks and new class data**
- ***<u>Intra-phase imbalance</u>*: severe class imbalances within each individual task**

Motivation

- **The magnitude of weight update significantly biased towards instance-rich classes**



Weight Update

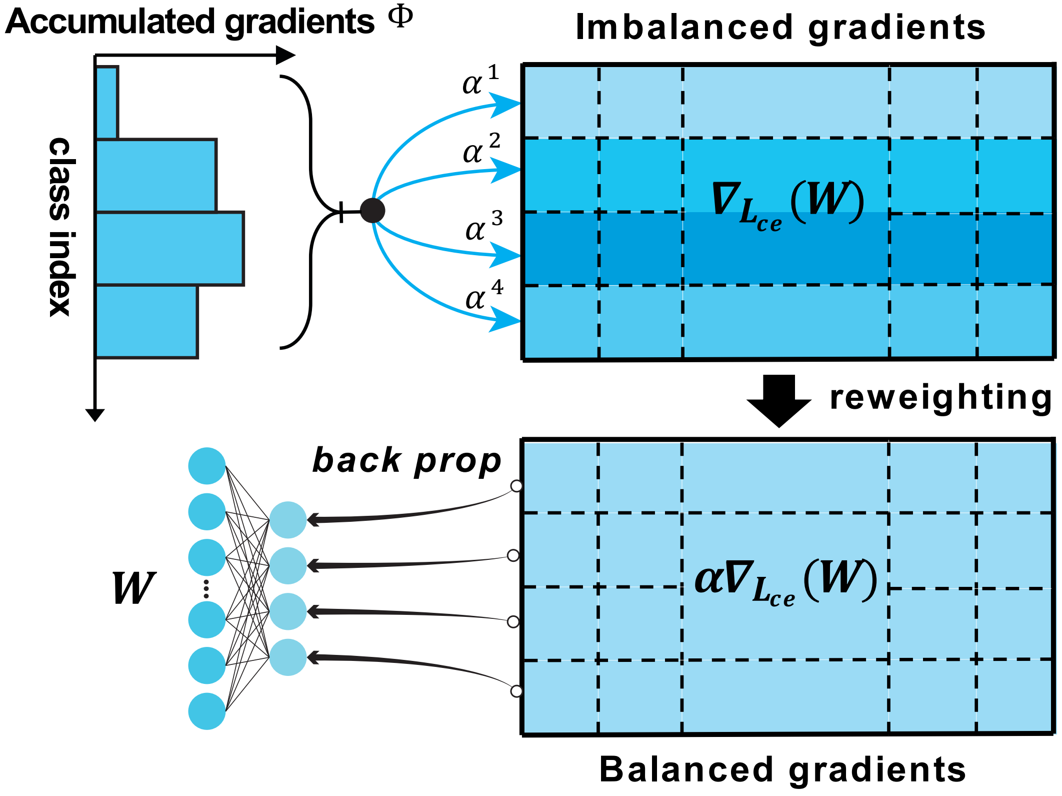
$$W_{i+1} = W_i - \boxed{\eta \nabla_{\mathcal{L}}(W_i)}$$

$$\eta ||\nabla_{\mathcal{L}}(W_i)||$$

magnitude

Intra-Phase Gradient Reweighting

- **Leveraging the per-class balance ratio during training process**



Class balance ratio

$$W_{i+1}^{j} = W_{i}^{j} - \eta \boxed{\alpha^{j}} \nabla_{\mathcal{L}}(W_{i}^{j})$$

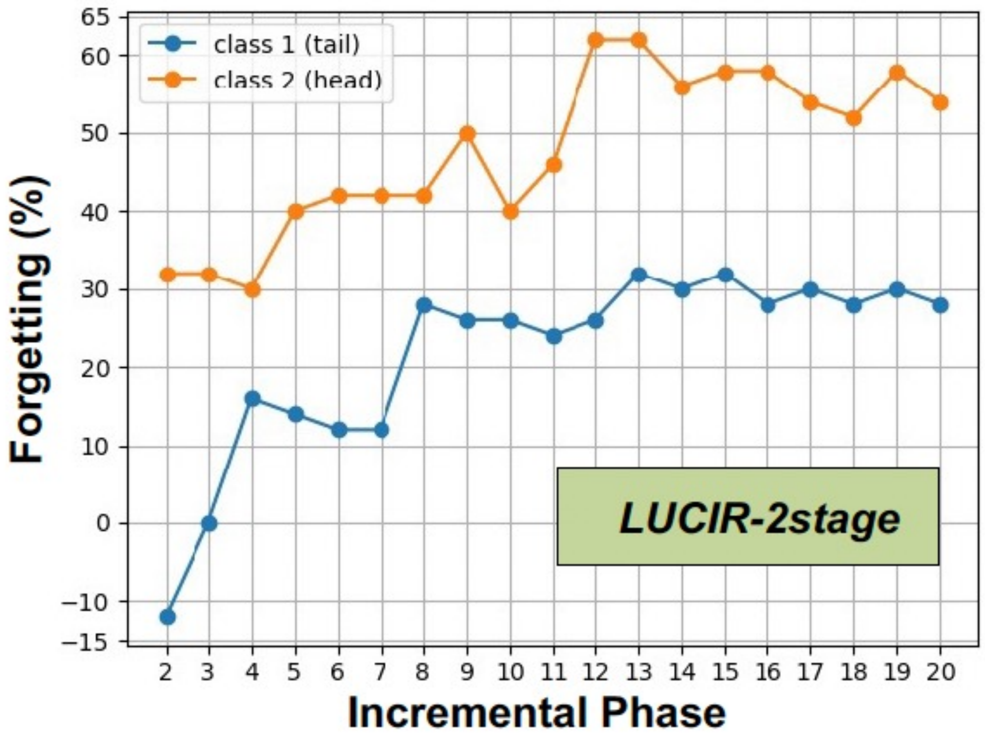
Accumulated Gradients

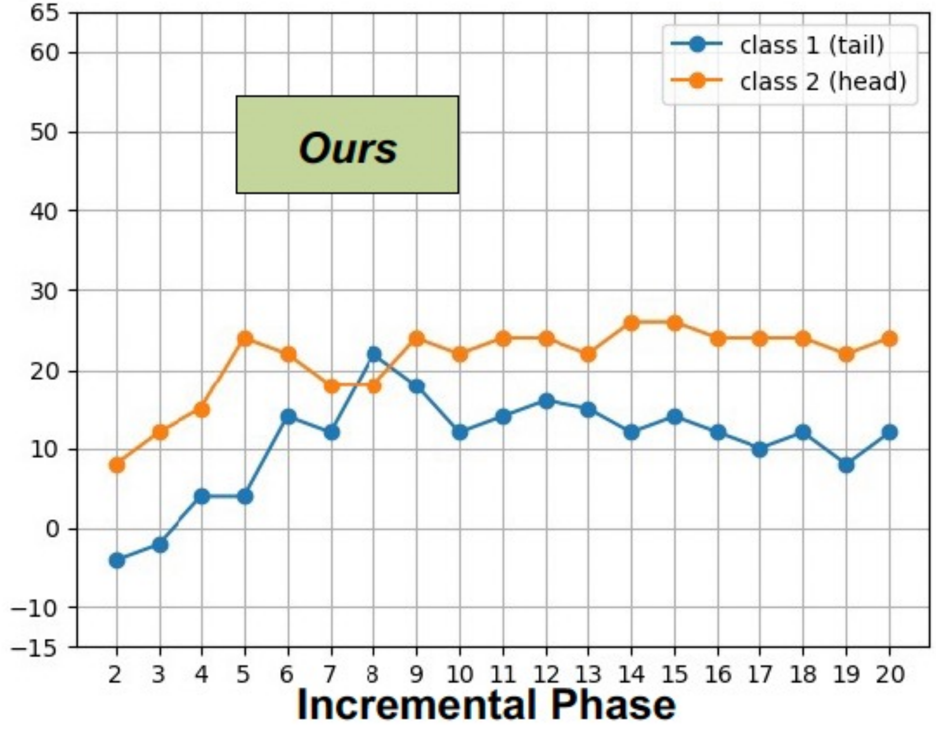
$$\alpha_i^j = \frac{\min\limits_{m \in \mathcal{Y}^1} \Phi_i^m}{\Phi_i^j}, \quad \Phi_i^j = \sum_{n=1}^{i} ||\nabla_{\mathcal{L}_{ce}}(W_n^j)||$$

Imbalanced Catastrophic Forgetting

- **Head classes suffers more forgetting than tail classes as most the training data becomes unavailable in subsequent incremental phases**

ImageNetSubset-LT (*LFS*, $N = 20, n_\varepsilon = 20$)





Distribution-Aware Knowledge Distillation

- **The DAKD loss prioritize preserving knowledge for classes with more training data lost during incremental learning**
- **Decouple the original distillation loss into a weighted sum of two parts**
  - $\boldsymbol{\sigma} = [\mathbf{0}, \mathbf{1}]$ **and** $\boldsymbol{\sigma} = \mathbf{1}$ **indicates balanced data lost and the DAKD will performs the same as regular knowledge distillation loss**

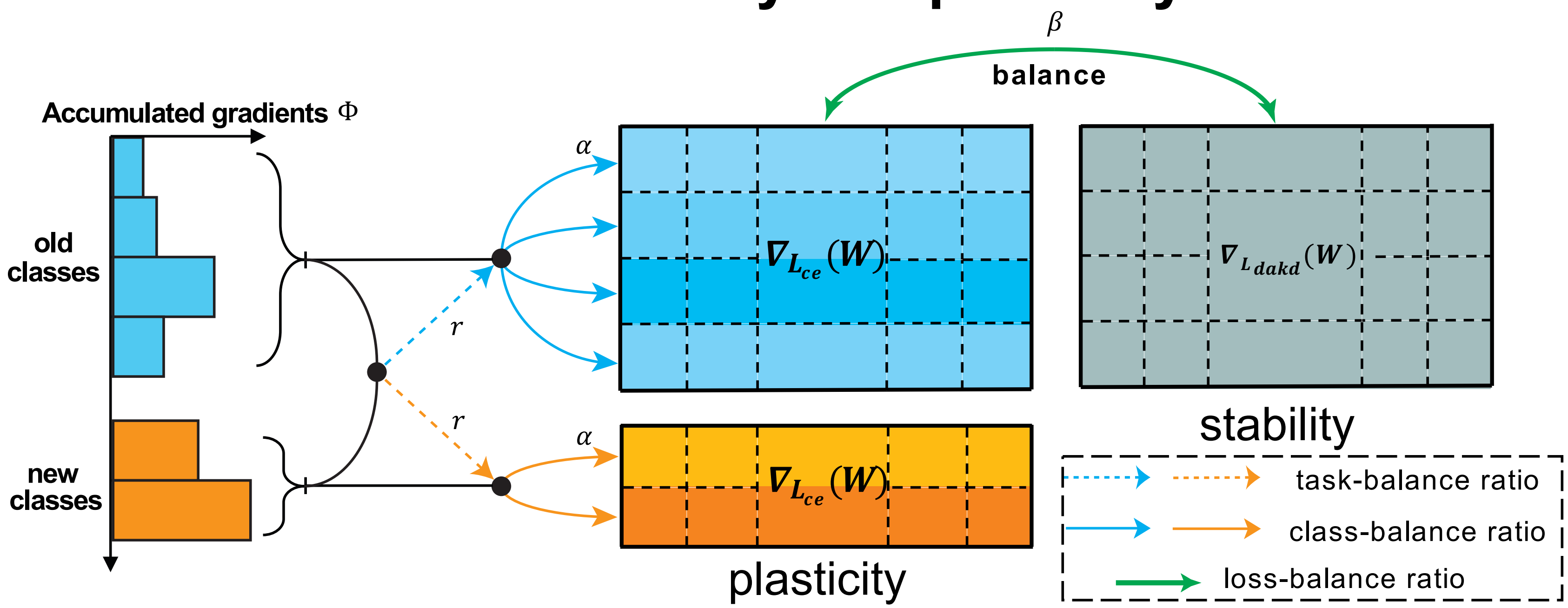
Measure the imbalance

$$\mathcal{L}_{dakd}(z, \hat{z}|\mathbf{s}) = \boxed{\sigma}\mathcal{L}_{kd}(z, \hat{z}) + (1 - \sigma)\mathcal{L}_{kd}^{imb}(\tilde{z}, \hat{z})$$

$$\tilde{z}_j = \frac{s_j}{\sum_m^{|\mathcal{Y}^{1:t-1}|} s_m} z_j + (1 - \frac{s_j}{\sum_m^{|\mathcal{Y}^{1:t-1}|} s_m})\hat{z}_j$$

Inter-Phase Decoupled Gradient Reweighting

- **Reweight the gradient for new tasks and learned tasks separately**
- **Tune the balance between stability and plasticity**

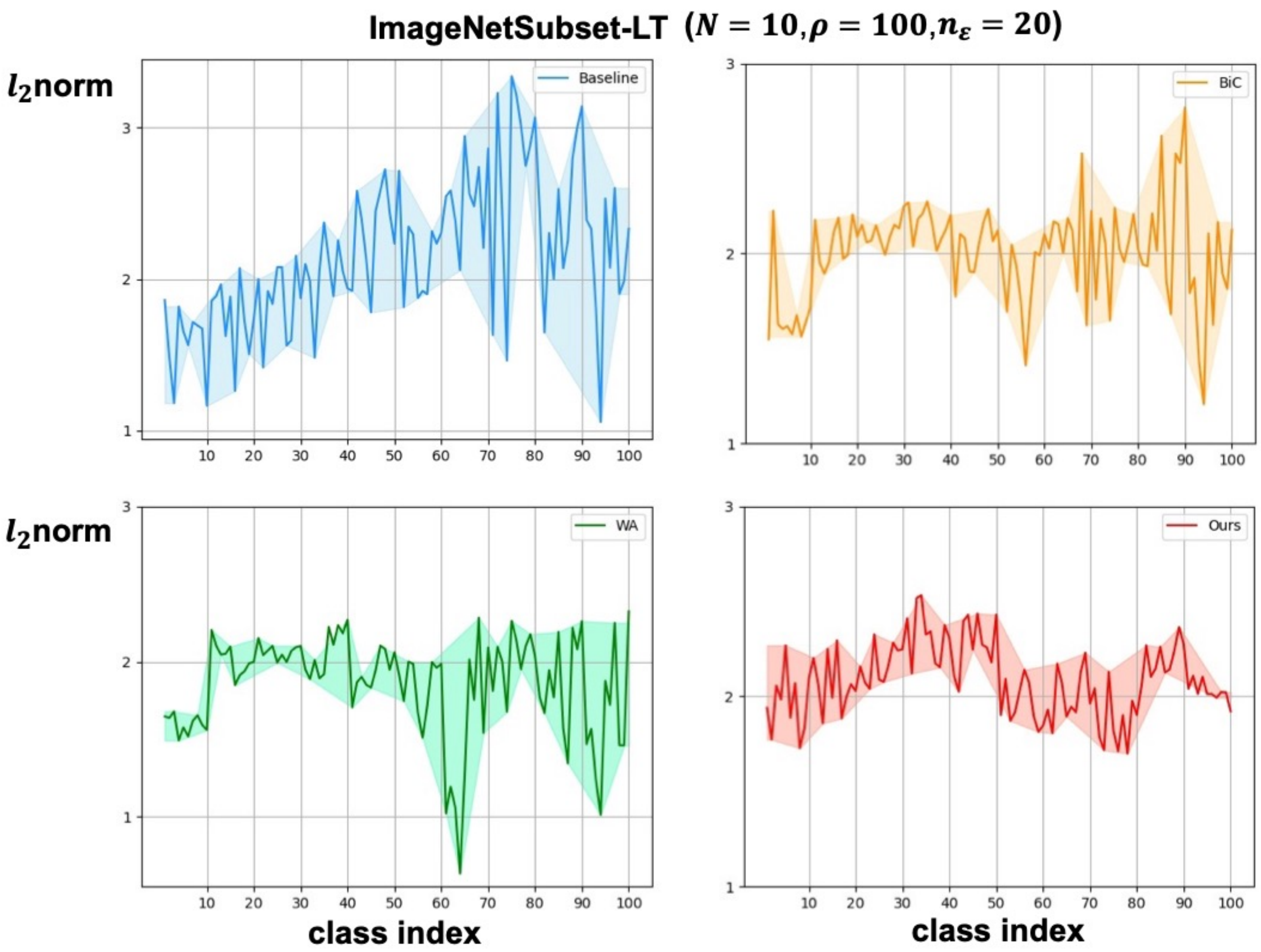


$$W^j_{i+1} = W^j_i - \overbrace{\underbrace{\eta(\alpha^j_i r^j_i \nabla_{\mathcal{L}_{ce}}(W^j_i)}_{j\in\mathcal{Y}^t} + \beta_i \nabla_{\mathcal{L}_{dakd}}(W^j_i))}^{j\in\mathcal{Y}^{1:t-1}}$$

Experiments

Datasets	CIFAR100-LT				ImageNetSubset-LT				Food101-LT			
Evaluation protocol	*LFS*		*LFH*		*LFS*		*LFH*		*LFS*		*LFH*	
Total tasks N	10	20	5	10	10	20	5	10	10	20	5	10
iCaRL [37]	21.83	24.28	28.68	28.33	33.75	29.71	41.82	40.21	18.13	12.50	21.83	21.31
IL2M [5]	31.37	29.99	34.90	33.42	31.70	25.20	40.75	39.08	16.11	16.27	23.93	22.48
BiC [46]	28.89	20.10	25.68	25.95	33.31	30.86	33.18	29.23	16.94	16.81	22.80	20.75
WA [53]	27.63	23.48	32.07	26.85	32.58	29.03	32.62	28.10	16.58	15.99	18.45	19.45
SSIL [1]	26.07	26.15	30.72	29.21	30.38	25.99	38.97	35.18	16.86	15.65	21.65	19.03
FOSTER [44]	30.43	29.96	37.25	37.91	34.38	29.75	46.51	43.88	24.27	20.45	32.39	31.46
MAFDRC [11]	32.67	31.95	37.94	38.51	**40.01**	34.48	48.23	44.12	26.93	19.21	34.22	30.91
EEIL-2stage [10, 26]	**33.64**	**32.25**	36.40	34.91	36.84	30.39	43.62	41.49	19.75	20.02	22.65	22.83
LUCIR-2stage [21, 26]	31.09	31.03	38.47	37.86	39.87	**34.79**	**48.97**	47.39	**27.65**	**24.68**	**36.05**	35.06
PODNet-2stage [14, 26]	30.41	30.37	38.38	38.45	35.47	31.71	48.02	**47.74**	23.78	21.13	35.42	**35.22**
FOSTER-2stage [26, 44]	31.27	30.68	**40.26**	**39.43**	36.47	33.95	48.89	46.93	25.82	22.28	35.69	33.48
Ours	**35.66**	**34.35**	**40.18**	**39.11**	**45.12**	**40.79**	**50.57**	**49.13**	**29.05**	**26.42**	**36.84**	**36.19**

Weight Bias Correction Effects



The L2 norm of learned weight vectors

Source Code

https://github.com/JiangpengHe/imbalanced_cil

Contact

he416@purdue.edu

Gradient Reweighting: Towards Imbalanced Class-Incremental Learning

Jiangpeng He

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

VIPER

Video and Image Processing Laboratory

CVPR

JUNE 17-21, 2024

SEATTLE, WA