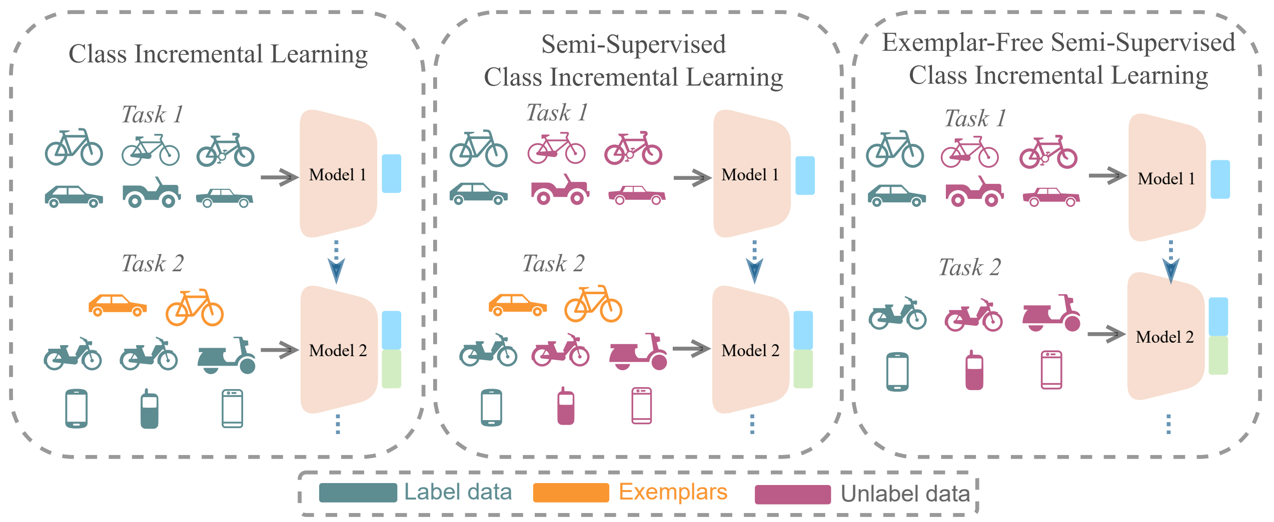
TACLE: TAsk and CLass-awarE Exemplar-free Semi-supervised Class Incremental Learning

Jayateja Kalla* Rohit Kumar* Soma Biswas

Image Analysis and Computer Vision Lab
Department of Electrical Engineering
Indian Institute of Science, Bangalore, INDIA.

WACV 2025
TUCSON, ARIZONA • FEB 28 - MAR 4

Class Incremental Learning Scenarios



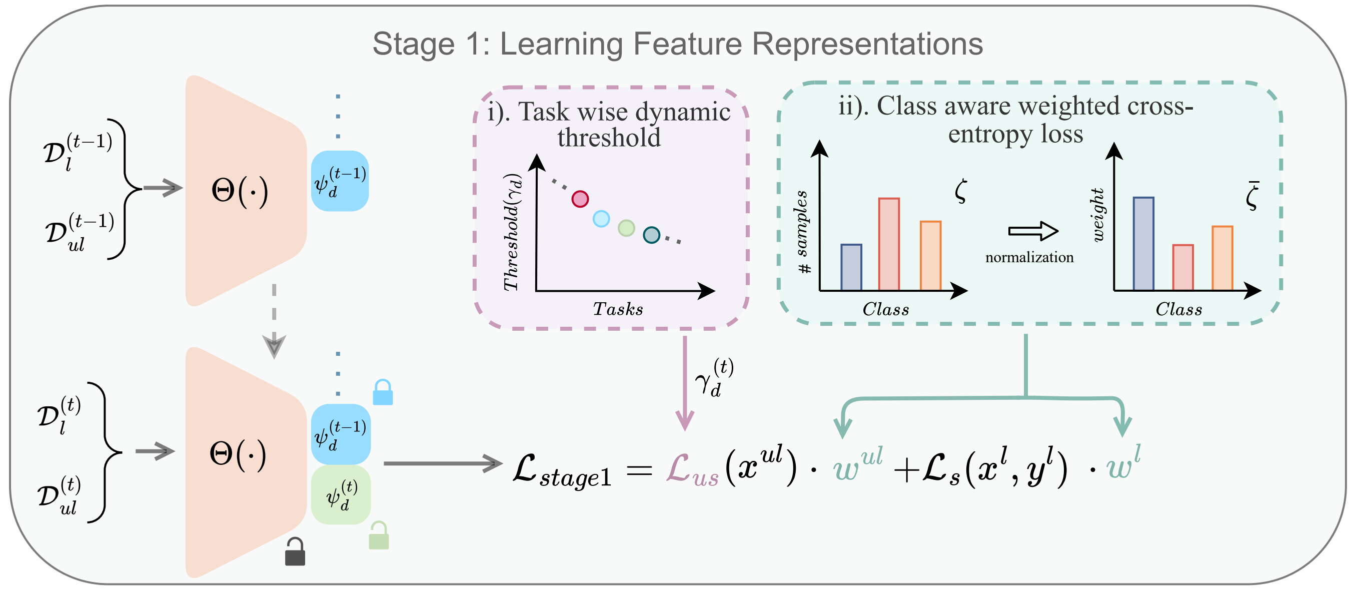
- The ability to learn from continuously evolving data is important for many real-world applications.
- In class incremental learning, the model updates its knowledge continuously when a new set of classes becomes available.
- In Semi-supervised CIL, the model updates using both labeled and unlabeled data at each task.

Why we need class incremental Learning?

- Data Collection of new classes.
- Data Unavailability of old classes.
- Computationally expensive to train model from scratch.

Stage1 : Learning Features Using both Labeled and Unlabeled Data

Task-wise adaptive threshold facilitates effective utilization of unlabeled data, while class-aware weighted loss improves performance on under-represented classes.



(i). Supervised Cross Entropy Loss: $\mathcal{L}_s(\mathbf{x}_i^l, y_i^l) = \mathcal{H}(p_i^l, y_i^l)$

(ii). Task-wise threshold loss for UnSupervised data: $\mathcal{L}_{us}(\mathbf{x}_i^{ul}) = \mathbb{I}(\max(p_i^{ul}) > \gamma_a^{(t)}) \cdot \mathcal{H}(\hat{p}_i^{ul}, \arg\max(p_i^{ul}))$

(iii). Inverse Sigmoid Function: $\gamma_a^{(t)} = \dfrac{\alpha}{1+e^{\alpha t}} + \beta$

(iv). Class aware weighted loss for stage1: $\mathcal{L}_{stage1} = \mathcal{L}_s(\mathbf{x}_i^l, y_i^l) \cdot w_i^l + \mathcal{L}_{us}(\mathbf{x}_i^{ul}) \cdot w_i^{ul}$

Experimental Results

Method	Model	CIFAR 100			CIFAR 10		
		0.8%	5%	25%	0.8%	5%	25%
Fine Tuning	RN18*	1.8 ± 0.2	5.0 ± 0.3	7.8 ± 0.1	13.6 ± 2.9	18.2 ± 0.4	19.2 ± 2.2
oEWC		1.4 ± 0.1	4.7 ± 0.1	7.8 ± 0.4	13.7 ± 1.2	17.6 ± 1.2	19.1 ± 0.8
ER (500)	RN18*	8.2 ± 0.1	13.7 ± 0.6	17.1 ± 0.7	36.3 ± 1.1	51.9 ± 4.5	60.9 ± 5.7
iCaRL (500)		3.6 ± 0.1	11.3 ± 0.3	27.6 ± 0.4	24.7 ± 2.3	35.8 ± 3.2	51.4 ± 8.4
FOSTER (500)		4.7 ± 0.6	14.1 ± 0.6	21.7 ± 0.7	43.3 ± 0.7	51.9 ± 1.3	57.1 ± 2.0
X-DER (500)		8.9 ± 0.3	18.3 ± 0.5	23.9 ± 0.7	33.4 ± 1.2	48.2 ± 1.7	58.9 ± 1.5
PseudoER (500)	RN18*	8.7 ± 0.4	11.4 ± 0.5	18.3 ± 0.2	50.5 ± 0.1	56.5 ± 0.6	57.0 ± 0.6
CCIC (500)		11.5 ± 0.7	19.5 ± 0.2	20.3 ± 0.3	54.0 ± 0.2	63.3 ± 1.9	63.9 ± 2.6
PAWS (500)		16.1 ± 0.4	21.2 ± 0.4	19.2 ± 0.4	51.8 ± 1.6	64.6 ± 0.6	65.9 ± 0.3
CSL (500)		23.6 ± 0.3	26.2 ± 0.5	29.3 ± 0.3	64.5 ± 0.7	69.6 ± 0.5	70.0 ± 0.4
NNCSL (500)		27.4 ± 0.5	31.4 ± 0.4	35.3 ± 0.3	73.2 ± 0.1	77.2 ± 0.2	77.3 ± 0.1
PseudoER (5120)	RN18*	15.1 ± 0.2	24.9 ± 0.5	30.1 ± 0.7	55.4 ± 0.5	70.0 ± 0.3	71.5 ± 0.2
CICC (5120)		12.0 ± 0.3	29.5 ± 0.4	44.3 ± 0.1	55.2 ± 1.4	74.3 ± 1.7	84.7 ± 0.9
ORDisCo (12500)		-	-	-	41.7 ± 1.2	59.9 ± 1.4	67.6 ± 1.8
CSL (5120)		23.7 ± 0.5	41.8 ± 0.4	50.3 ± 0.8	64.3 ± 0.7	73.1 ± 0.3	73.9 ± 0.1
NNCSL (5120)		27.5 ± 0.7	46.0 ± 0.2	56.4 ± 0.5	73.7 ± 0.4	79.3 ± 0.3	81.0 ± 0.2
SLCA (0)	ViTs†	66.43 ± 0.04	81.86 ± 0.02	86.95 ± 0.01	93.55 ± 0.03	94.45 ± 0.01	**96.19** ± 0.01
SLCA+FT (0)		71.67 ± 0.09	83.96 ± 0.06	86.91 ± 0.02	94.07 ± 0.07	95.35 ± 0.05	96.08 ± 0.02
TACLE (ours) (0)		**79.51** ± 0.08	**85.58** ± 0.05	**87.24** ± 0.02	**94.59** ± 0.08	**95.49** ± 0.05	96.02 ± 0.01
SLCA (0)	ViTs‡	63.67 ± 0.03	91.38 ± 0.02	93.69 ± 0.01	91.64 ± 0.02	97.79 ± 0.01	98.56 ± 0.01
SLCA+FT (0)		88.23 ± 0.04	93.30 ± 0.03	94.08 ±0.01	98.45 ± 0.03	98.26 ± 0.02	**98.89** ± 0.02
TACLE (ours) (0)		**92.35** ± 0.06	**93.59** ± 0.04	**94.10** ± 0.02	**98.61** ± 0.03	**98.44** ± 0.03	98.86 ± 0.02

Method	Model	ImageNet100-Subset		
		1%	5%	25%
Fine-tuning	ResNet18*	1.5 ± 0.2	2.7 ± 0.1	4.1 ± 0.2
ER (5120)		12.2 ± 0.8	26.3 ± 0.7	38.8 ± 1.0
FOSTER (5120)		14.8 ± 1.1	32.8 ± 0.7	42.1 ± 1.5
X-DER (5120)		10.8 ± 1.1	27.4 ± 1.6	45.3 ± 1.0
CCIC (5120)		13.5 ± 1.2	19.5 ± 0.7	25.9 ± 0.9
CSL (5120)	ResNet18*	26.8 ± 0.4	47.9 ± 0.2	56.3 ± 0.5
NNCSL (5120)		29.7 ± 0.4	51.3 ± 0.1	65.6 ± 0.3
SLCA (0)	ViTs†	78.30 ± 0.04	79.29 ± 0.02	82.39 ± 0.01
SLCA+Fixed Threshold (0)		79.72 ± 0.08	82.21 ± 0.05	**83.08** ± 0.02
TACLE (ours) (0)		**80.82** ± 0.09	**82.42** ± 0.04	83.01 ± 0.02

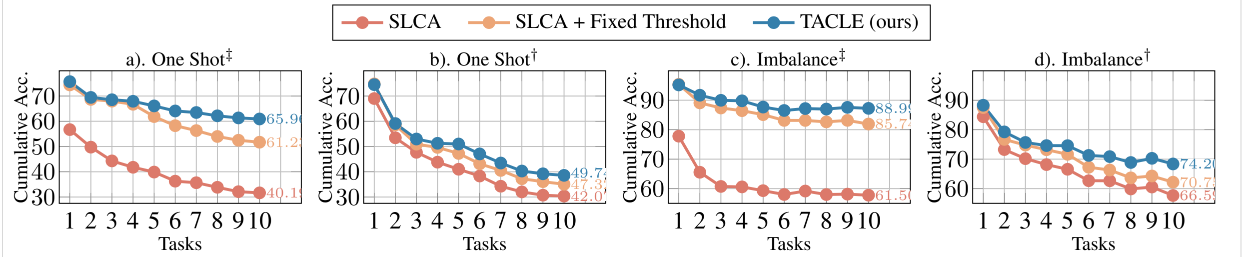
Method	Data		Components			Pre-trained	
	Labelled	Unlabeled	C1	C2	C3	ImageNet	MoCo v3
SLCA	✓	✗	✗	✗	✗	63.37	66.43
SLCA + Fixed Threshold	✓	✓	✗	✗	✗	88.23	71.67
TACLE	✓	✓	✓	✗	✗	89.10	75.29
	✓	✓	✓	✓	✗	91.32	77.19
	✓	✓	✓	✓	✓	**92.35**	**79.51**

c1: Task aware threshold

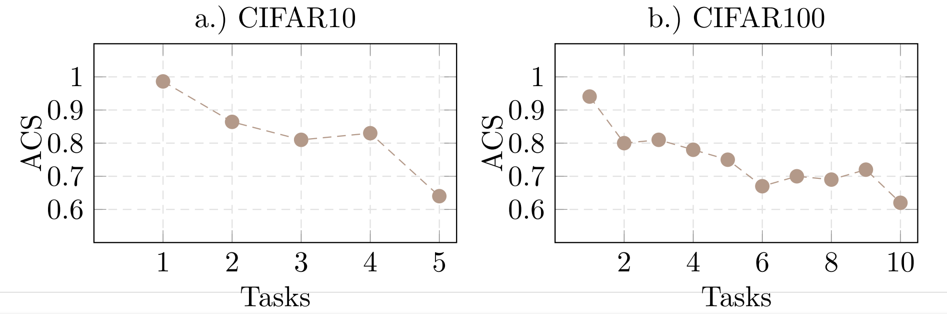
c2: Class aware weightage

c3: Unable data for stage2

Analysis on one-shot SS-CIL and imbalance SS-CIL problem settings

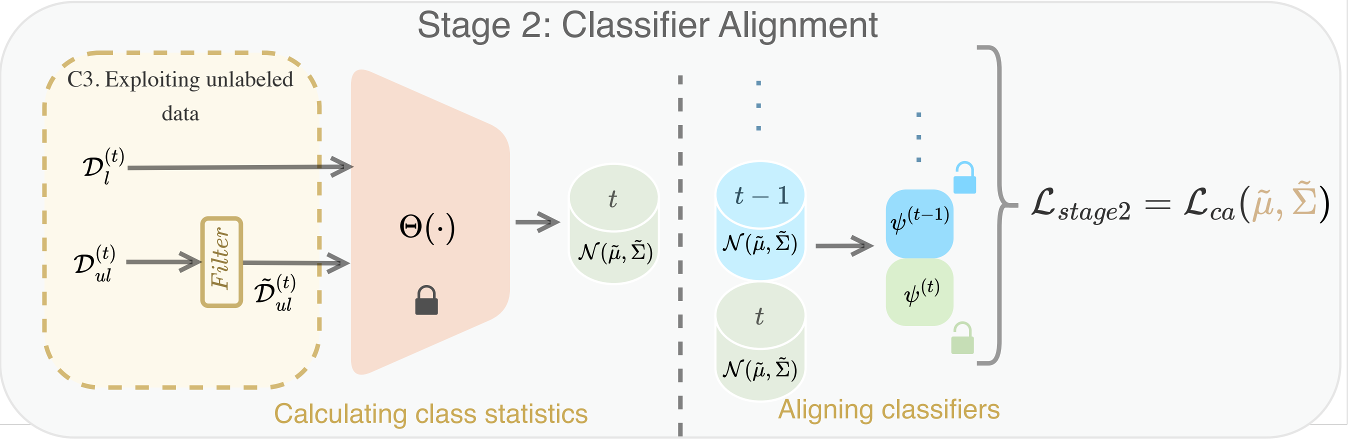


Average Confidence Score (ACS) for Unlabeled Data Across Tasks



Stage2 : Classifier Alignment Using both Labeled and Unlabeled Data

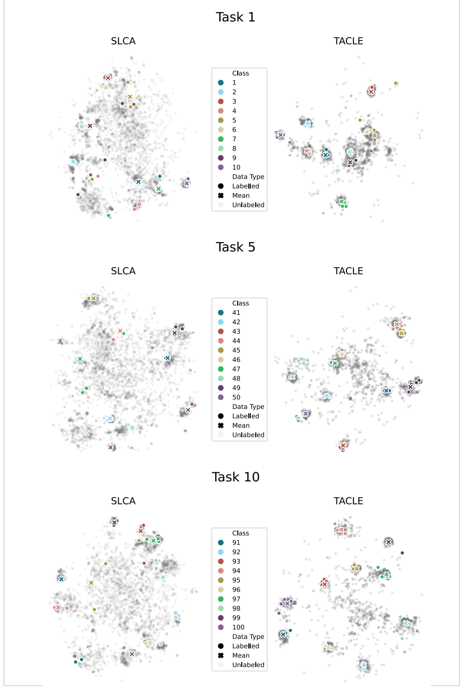
Leveraging unlabeled data for better class statistics estimation, which further enhances classifier calibration.



Algorithm 1: TACLE for semi-supervised class incremental learning

Input: $\{\Theta, \psi\} \leftarrow$ Model; $\{\mathcal{D}^{(1)}, \mathcal{D}^{(2)}, \dots, \mathcal{D}^{(\mathcal{T})}\} \leftarrow$ Data stream;
$E_{s1} \leftarrow$ No. of epochs for stage 1; $E_{s2} \leftarrow$ No. of epochs for stage 2;
for $t \leftarrow 1$ *to* $\mathcal{T}$ **do**
  $\mathcal{D}_l^{(t)} = \{\mathbf{x}_i^l, y_i^l\}_{i=1}^{N_l^{(t)}}$; $\mathcal{D}_{ul}^{(t)} = \{\mathbf{x}_i^{ul}\}_{i=1}^{N_{ul}^{(t)}}$;
  $\zeta \leftarrow$ Uniform distribution across all classes
  // #Stage 1: Feature Representation Learning //
  for $e_{s1} \leftarrow 1$ *to* E_{s1} **do**
    $\mathcal{B}_l$ = SampleMiniBatch($\mathcal{D}_l^{(t)}$); $\mathcal{B}_{ul}$ = SampleMiniBatch($\mathcal{D}_{ul}^{(t)}$);
    $\hat{\mathcal{B}}_{ul}$ = ImageAugmentations($\mathcal{B}_{ul}$);
    $\mathcal{O}_l, \mathcal{O}_{ul}, \hat{\mathcal{O}}_{ul} = \Theta(\psi^{(t)}(\mathcal{B}_l, \mathcal{B}_{ul}, \hat{\mathcal{B}}_{ul}))$;
    $w^l \leftarrow$ Assigning class-aware weights for labeled data $\mathcal{B}_l$ using $\bar{\zeta}$;
    $w^{ul} \leftarrow$ Assigning class-aware weights for unlabeled data $\mathcal{B}_{ul}$ using $\bar{\zeta}$;
    $\mathcal{L}_{stage1} \leftarrow \mathcal{L}_s(\mathcal{B}_l) \cdot w^l + \mathcal{L}_{us}(\hat{\mathcal{B}}_{ul}) \cdot w^{ul}$; // Total loss for stage1
    $\zeta \leftarrow$ Update the histogram distribution using $\mathcal{D}_{ul}^{(t)}, \gamma_a^{(t)}$;
    $\bar{\zeta} \leftarrow (2 - \zeta)$; // Normalization
    $\{\Theta, \psi^{(t)}\} \leftarrow$ Update model parameters using $\mathcal{L}_{stage1}$;
  //#Stage 2: Classifier Alignment//
  $\tilde{\mathcal{D}}^{(t)} \leftarrow$ Expanded labelled data set using $\mathcal{D}_l^{(t)}, \mathcal{D}_{ul}^{(t)}, \gamma_a^{(t)}$;
  $\tilde{\mu}_k^{(t)}, \tilde{\Sigma}_k^{(t)} \leftarrow$ Estimate mean and variance using $\tilde{\mathcal{D}}^{(t)}$; // where $k \in 1, 2, .., |\mathcal{C}^{(t)}|$
  for $e_{s2} \leftarrow 1$ *to* E_{s2} **do**
    $\mathcal{L}_{stage2} \leftarrow \mathcal{L}_{ca}(\tilde{\mu}_k^{(1:t)}, \tilde{\Sigma}_k^{(1:t)})$; // Alignment loss for classifiers
    $\psi^{(1:t)} \leftarrow$ Update classifier parameters using $\mathcal{L}_{stage2}$;

t-SNE Visualizations - CIFAR100 Dataset



References:

1. Zhiqi Kang, Enrico Fini, Moin Nabi, Elisa Ricci, and Karteek Alahari. A soft nearest-neighbor framework for continual semi-supervised learning. In ICCV, pages 11868–11877, 2023.
2. Zhang, Gengwei, et al. "Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. In ICCV, pages 19148-19158, 2023.

{ jayatejak, krohit, somabiswas }@iisc.ac.in

Code: https://github.com/rokmr/TACLE