ACIL: ACTIVE CLASS INCREMENTAL LEARNING FOR IMAGE CLASSIFICATION

Aditya R. Bhattacharya*, Debanjan Goswami* and Shayok Chakraborty
Department of Computer Science
Florida State University

PROBLEM MOTIVATION

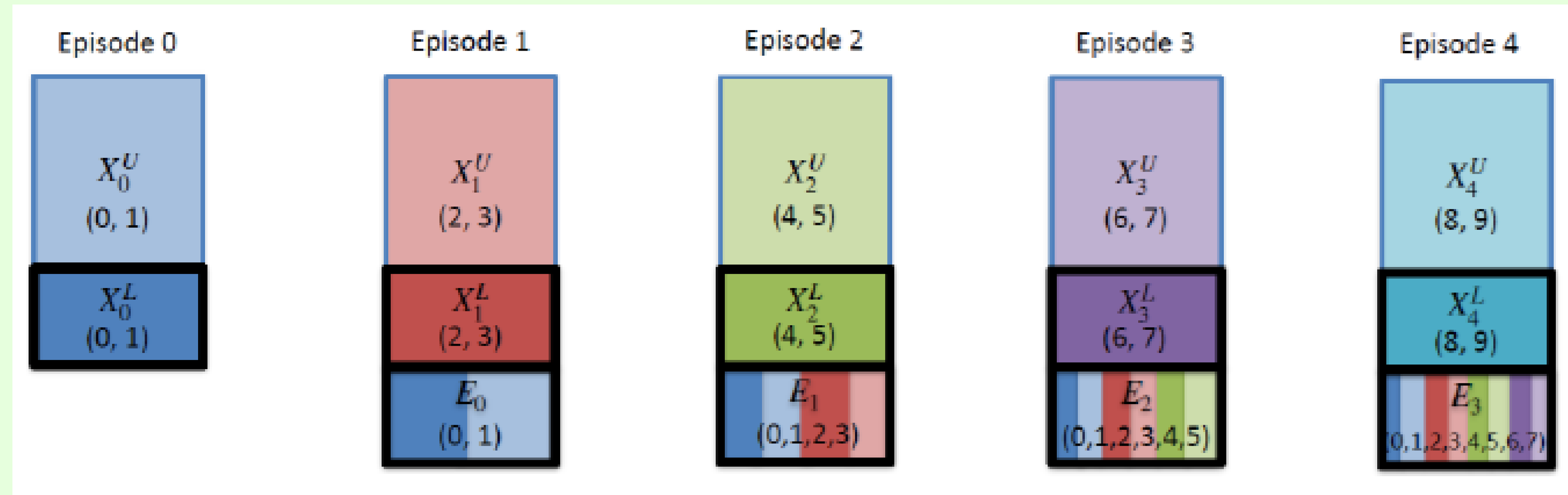
Class Incremental Learning (CIL)

- In many real-world applications, training data arrives in episodes over time
- Deep Neural Networks (DNNs) should learn incrementally from these new classes from their episodic data
- They should also retain knowledge about the formerly learned classes (avoid *catastrophic forgetting*)

Active Learning (AL)

- Automatically identifies salient and exemplar samples from large amounts of unlabeled data
- Reduces human annotation effort
- Model gets trained on the informative samples
- Better generalization capability

Active Class Incremental Learning (ACIL)



- Data arrives sequentially over *n* episodes (*0, 1,… N-1*)
- X^L_n -> labeled data in episode *n*; X^U_n -> unlabeled data in episode *n*; E_{n-1} -> labeled exemplar set from episode *n-1*
- Goal is to select an exemplar set of size *k* from episode *n*, containing a subset of samples from X^L_n, X^U_n and E_{n-1}

PROPOSED FRAMEWORK

Budget Splitting Strategy

- Split the budget to select samples both from the labeled exemplar set and unlabeled set
- These samples will be propagated to the next episode as the exemplar set E_n
- Rationale is to split the budget *k* between X^U_n and E_{n-1} in the same proportion as the number of classes in the two sets:

$$k_{unlabeled} = \left(\frac{|C_{episode}|}{|C_{episode}| + |C_{exemplar}|}\right).k \qquad k_{exemplar} = \left(\frac{|C_{exemplar}|}{|C_{episode}| + |C_{exemplar}|}\right).k$$

- We use the pseudo-labels furnished by the DNN to select samples from each class in X^U_n
- Exemplar set E_n thus contains a good representation of all the classes seen so far

Active Sampling Strategy

- Active sampling is applied class-wise, using a strategy based on uncertainty and diversity
- Partition the samples in *X* into *B* diverse sets using a partition function: *P: X -> {X_1, X_2… X_B}*
- Goal is to minimize the variance of each partition:

$$\sigma^2(X_b) = \frac{1}{2|X_b|^2} \sum_{x_i, x_j \in X_b} ||\mathcal{F}(x_i) - \mathcal{F}(x_j)||^2$$

- Reformulate to minimize the *weighted variance* within each partition:

$$\arg\min_{\mathcal{P}} \sum_{b=1}^{B} \sum_{x \in X_b} \mathcal{I}(x) ||\mathcal{F}(x) - c_b||^2$$

- Each sample is weighted by its prediction uncertainty *I(x)*

Loss Function

- In each episode, DNN is trained on the labeled episodic data X^L_n, and labeled exemplar set E_{n-1}
- Class imbalance may exist in X^L_n *U* E_{n-1} as $|X^L_n| >> |E_{n-1}|$
- Use weighted CE loss to train the DNN:

$$\mathcal{L}_{\mathcal{WCE}}(x_i) = -\sum_{j=1}^{C} \delta(y_i == j) w_j \log p_{ij}$$

- Total loss over all the samples is:

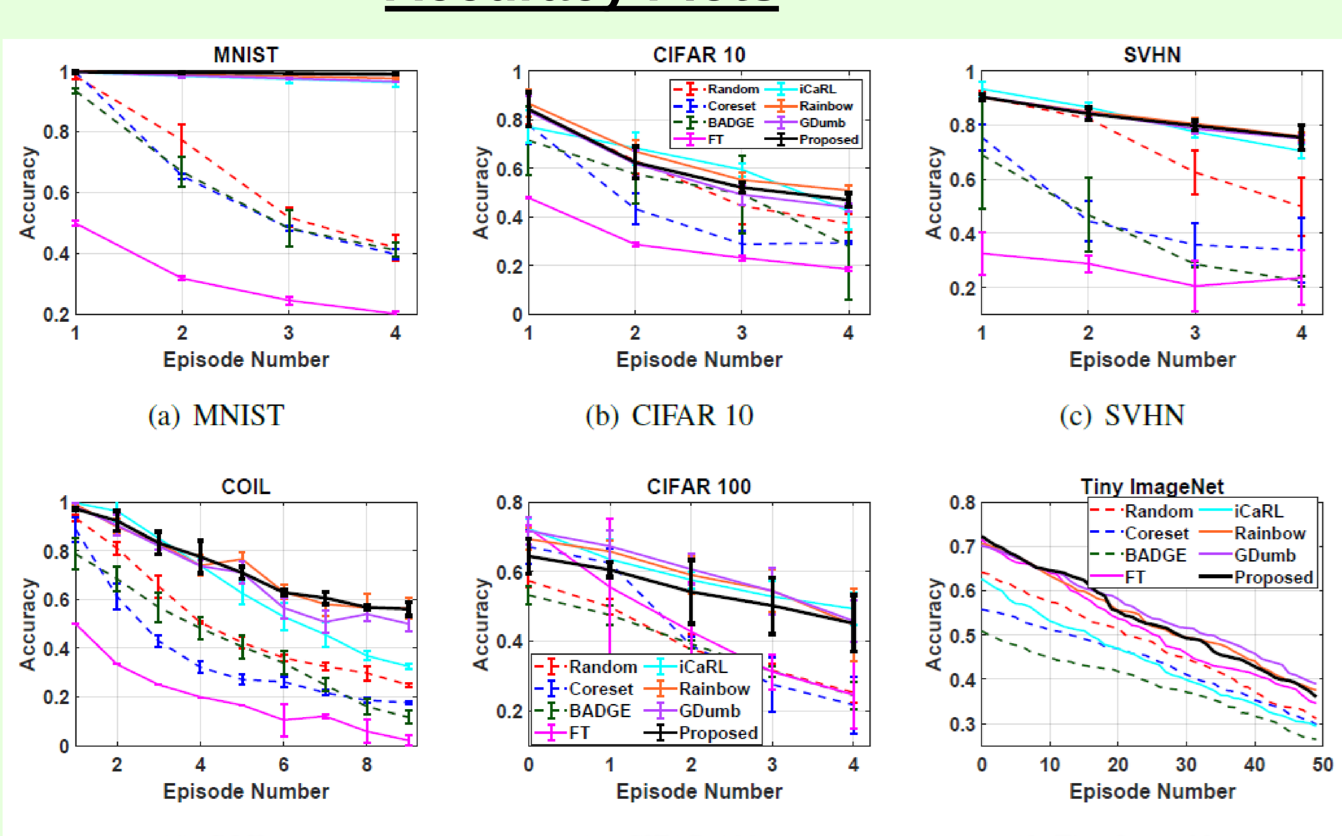
$$\mathcal{L}_{\mathcal{WCE}} = \frac{1}{|L|} \sum_{i=1}^{|L|} \mathcal{L}_{\mathcal{WCE}}(x_i)$$

- Include a distillation loss term between the model predictions in two successive episodes. Final loss function is given by:

$$\mathcal{L} = \mathcal{L}_{\mathcal{WCE}} + \lambda \mathcal{L}_{\mathcal{D}}$$

SAMPLE RESULTS

Accuracy Plots



(a) MNIST (b) CIFAR 10 (c) SVHN (d) COIL (e) CIFAR 100 (f) Tiny ImageNet

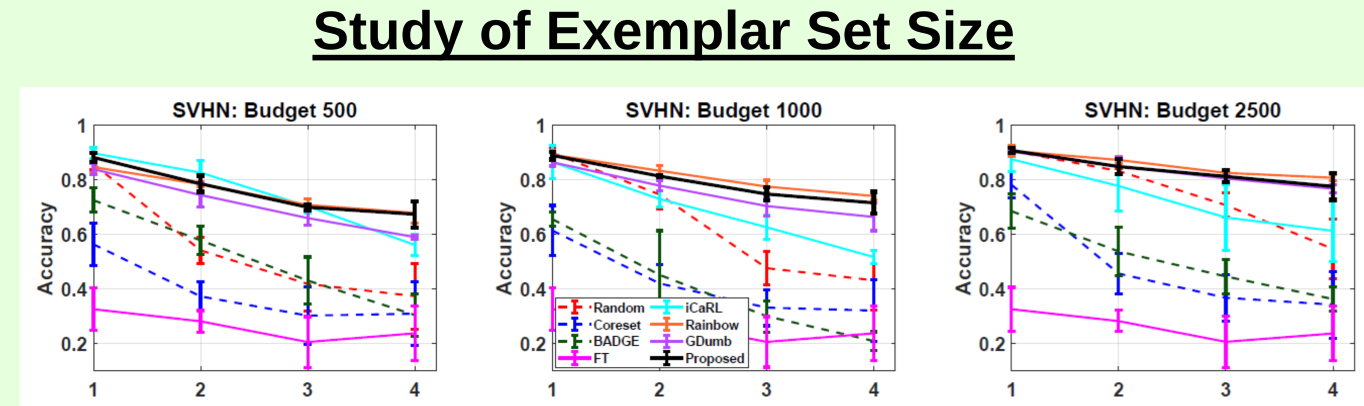
Annotation Effort Analysis

	CIL Baselines	AL Baselines	Proposed
MNIST	$12,000 \pm 478.12$	$3,359.6 \pm 483.43$	$2,899.6 \pm 447.38$
CIFAR 10	$10,000 \pm 0.00$	$2,800 \pm 414.04$	$2,417.2 \pm 344.90$
SVHN	$14,651.40 \pm 5,542.22$	$4,530 \pm 1,181.98$	$3,763.60 \pm 1,455.34$
COIL	570 ± 0.00	204 ± 30.51	153.33 ± 25.38
CIFAR 100	$5,000 \pm 0.00$	$1,450 \pm 152.56$	$1,158 \pm 131.53$
Tiny ImageNet	$1,000 \pm 0.00$	299 ± 9.97	232.07 ± 21.92

Accuracy after Last Episode

	Random	Coreset	BADGE	FT	iCaRL	Rainbow	GDumb	Proposed
MNIST	0.41 ± 0.03	0.39 ± 0.01	0.41 ± 0.02	0.20 ± 0.01	0.96 ± 0.01	$\underline{0.97} \pm 0.01$	0.96 ± 0.01	$\mathbf{0.98} \pm 0.02$
CIFAR 10	0.37 ± 0.03	0.29 ± 0.03	0.27 ± 0.17	0.18 ± 0.04	0.42 ± 0.06	$\mathbf{0.5} \pm 0.02$	0.43 ± 0.01	$\underline{0.47} \pm 0.02$
SVHN	0.49 ± 0.09	0.33 ± 0.1	0.22 ± 0.02	0.23 ± 0.08	$\underline{0.70} \pm 0.02$	$\mathbf{0.75} \pm 0.03$	$\mathbf{0.75} \pm 0.01$	$\mathbf{0.75} \pm 0.04$
COIL	0.24 ± 0.05	0.17 ± 0.02	0.11 ± 0.02	0.02 ± 0.02	0.32 ± 0.01	$\mathbf{0.56} \pm 0.04$	$\underline{0.5} \pm 0.03$	$\mathbf{0.56} \pm 0.02$
CIFAR 100	0.25 ± 0.03	0.22 ± 0.08	0.24 ± 0.04	0.25 ± 0.10	$\mathbf{0.49} \pm 0.05$	0.45 ± 0.10	$\underline{0.46} \pm 0.06$	0.45 ± 0.08
Tiny ImageNet	0.31 ± 0.01	0.30 ± 0.03	0.26 ± 0.04	0.35 ± 0.02	0.29 ± 0.02	$\underline{0.38} \pm 0.05$	$\mathbf{0.39} \pm 0.04$	0.36 ± 0.04

Study of Exemplar Set Size



	CIL Baselines	AL Baselines	Proposed
SVHN: Budget 500	$14,651.40 \pm 5,542.22$	$3,330 \pm 1,068.78$	$3,139.20 \pm 1,166.54$
SVHN: Budget 1,000	$14,651.40 \pm 5,542.22$	$3,730 \pm 1,068.41$	$3,347.20 \pm 1,246.61$
SVHN: Budget 2,500	$14,651.40 \pm 5,542.22$	$4,930 \pm 1,285.86$	$3,972.80 \pm 1,577.21$

* These authors contributed equally to this work