Training-Free Uncertainty-guided Logit Adjustment for Few-Shot Class-Incremental Learning

Supplementary Material

Within the supplement, we provide further implementation details in Section A, followed by additional analysis of the proposed method B. Additional experimental results are presented in Section C.

A. Further Implementation Details

Evaluation Metrics. In FSCIL, the commonly used class-wise average accuracy (*aAcc*) tends to be biased toward base classes. The average accuracy at session i is defined as

$$aAcc_i = \frac{\frac{|\mathcal{Y}_{\text{base}}|}{|\mathcal{Y}_{\text{novel}}|} A_i^1 + \sum_{j=2}^{i} A_i^j}{\frac{|\mathcal{Y}_{\text{base}}|}{|\mathcal{Y}_{\text{novel}}|} + (i-1)} \quad (6)$$

, where $|\mathcal{Y}_{\text{base}}|$ denotes the number of base classes and $|\mathcal{Y}_{\text{base}}|$ denotes the number of incremental classes. Since the number of base classes $|\mathcal{Y}_{\text{base}}|$ is much larger than that of incremental classes $|\mathcal{Y}_{\text{novel}}|$, the overall performance is dominated by base-class accuracy. For instance, the ratio $\frac{|\mathcal{Y}_{\text{base}}|}{|\mathcal{Y}_{\text{novel}}|}$ equals 12 for CIFAR-100 and *mini*ImageNet, and 10 for CUB-200, indicating that *aAcc* is heavily influenced by the base subset. To better capture the balance between base and incremental performance, recent works [2, 23, 43] have adopted harmonic mean (hAcc), which penalizes large disparities between the two accuracies and provides a more balanced evaluation. More recently, Tang et al. [34] introduced a generalized accuracy (gAcc) as a global variant of *aAcc*, offering a more comprehensive view of the overall performance trend across sessions. In addition to accuracy-based metrics, Wang et al. [38] proposed measuring the false positive rate (FPR) by defining base classes as positive and incremental classes as negative. This metric quantifies the proportion of incremental samples misclassified as base classes, thus serving as an indicator of prediction bias. Following these practices, we report *aAcc*, hAcc, gAcc, and FPR to comprehensively evaluate both overall performance and bias tendencies in our experiments.

Details of Baseline Models. We evaluate our method using four representative incremental-frozen approaches widely adopted in FSCIL. These include a conventional cross-entropy–based prototype classifier, the forward-compatible training strategy FACT [46], the contrastive-learning–based SAVC [33], and the geometry-aware NC-FSCIL [40]. A brief summary of each baseline is provided below.

- **CE**: The model is first trained with standard cross-entropy loss during the base session. In subsequent incremental sessions, the feature extractor remains frozen and only the classifier parameters are updated for new classes.
- **FACT** [46]: FACT expands the base training data by generating virtual novel classes using manifold mixup, thereby encouraging the feature extractor to reserve separable regions for unseen categories.
- **NC-FSCIL** [40] introduces pre-assigned classifier weights (i.e. prototypes) based on a simplex equiangular tight frame (ETF), and fine-tunes a projection layer to align the output features with their corresponding prototypes.
- **SAVC** [33]: SAVC employs contrastive learning in the base session to obtain compact and discriminative representations. For the incremental phase, multiple prototypes estimated from the few-shot samples are aggregated to form the classifier parameters for each new class.

CE, FACT, and SAVC adopt a ResNet-18 backbone for the *mini*ImageNet and CUB200 datasets, and a ResNet-20 backbone for CIFAR100. NC-FSCIL uses a ResNet-12 backbone on *mini*ImageNet and CIFAR-100, and a ResNet-18 backbone on CUB200.

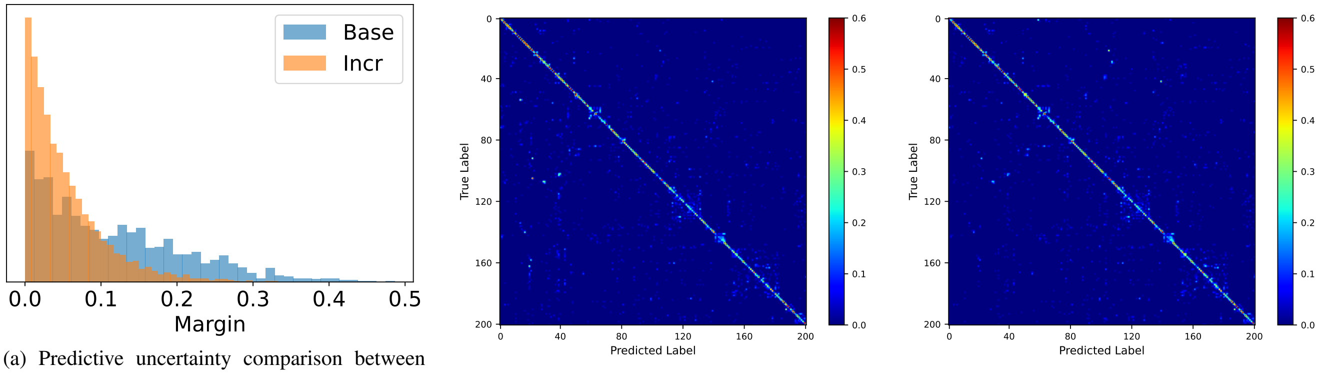
Training Details. Following previous studies, for the CUB200 dataset, we initialize all networks with ImageNet-1K pretrained weights for the backbone encoder. All experiments are conducted on a single NVIDIA RTX 3090 GPU with 24GB of memory. Since our method is a *plug-and-play* approach, we adhere to the hyperparameter settings and original architecture settings used in each of the four baseline models described above.

Table A. Session-wise performance under the extreme setting.

ULA	Session 1	Session 8	Session 15	aAcc
w/o	74.85	56.61	51.11	60.78
w/	**75.32**	**57.14**	**51.48**	**61.33**

B. Additional Analysis of Uncertainty-Guided Logit Adjustment

Effectiveness in Scenarios with Diverse Error Distributions. We further analyze the behavior of ULA in scenarios where the number of incremental classes significantly exceeds the number of base classes. To this end, we construct an extreme FSCIL setting on CUB200 by restricting the base session to 50 classes and conducting 15 incremental sessions (10-way, 5-shot each) with a frozen encoder



(a) Predictive uncertainty comparison between base and incremental samples.

(b) Confusion matrix before applying ULA.

(c) Confusion matrix after applying ULA.

Figure A. In an extreme FSCIL setting where the number of incremental classes exceeds that of base classes, a large portion of incremental samples are misclassified as base classes without ULA. Applying ULA effectively mitigates this base-biased error.

trained only on the base classes. Incremental samples are randomly selected from the training set with a fixed seed to ensure reproducibility and no overlap with base data.

Even in this setting, predictive uncertainty still exhibits a group-wise disparity between base and incremental samples, as shown in Fig. Aa. Although inter-incremental confusions increase as the number of incremental classes grows, a substantial portion of incremental samples are still misclassified as base classes (Fig. Ab). This indicates that base-biased prediction remains a dominant error mode under the incremental-frozen protocol. After applying ULA, these errors are effectively mitigated (Fig. Ac), leading to consistent improvements across sessions, as summarized in Table A. As a result, ULA remains actively triggered and effective even when incremental classes substantially outnumber base classes.

Robustness to Hyperparameter Selection. We additionally analyze the sensitivity of ULA to hyperparameter choices. ULA introduces three tunable parameters: the scaling factor, margin threshold, and entropy threshold, along with the temperature parameter applied to the final logits(an existing hyperparameter in the baseline). The first three parameters are constrained to bounded ranges in [0,1], defining a relatively small search space.

To evaluate robustness, we randomly sample parameter configurations within the predefined ranges and measure performance across trials. As shown in Fig. B, ULA consistently improves performance across a wide range of values. This indicates that ULA does not rely on precise hyperparameter tuning to remain effective.

C. Additional Results

In this section, we report the harmonic accuracy of CIFAR100 in Table C, and we also report the base and incremental average accuracies, together with the overall *aAcc* and *gAcc* values for all sessions on three datasets, as summarized in Tables D to F. We further evaluate ULA on an additional FSCIL baseline (ADBS) [20], and observe consistent performance gains across datasets, as shown in Table B.

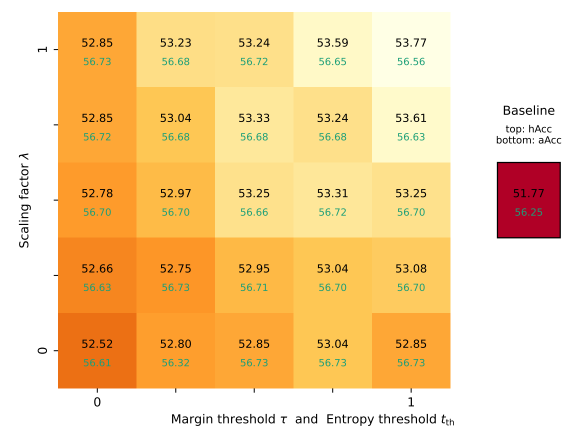


Figure B. Performance stability of ULA under randomly sampled hyperparameter configurations.

Table B. Performance of ULA applied to the ADBS baseline.

Method	Dataset	aAcc	aHM
ADBS [20]	CIFAR100	62.30	4.95
+ ULA		**63.44**	**17.93**
ADBS [20]	CUB200	67.75	57.25
+ ULA		**68.09**	**59.01**
ADBS [20]	miniImageNet	65.91	27.14
+ ULA		**66.02**	**30.84**

Method	Harmonic accuracy in each session (%)								aHM	*aAcc* (%)	*gAcc* (%)
	1	2	3	4	5	6	7	8		Last (8)	Last (8)
C-FSCIL [11]	31.47	28.32	26.67	23.11	24.84	24.28	23.04	23.48	25.65	49.70	–
CEC [41]	41.97	39.18	34.25	33.25	33.84	33.47	32.90	32.15	35.12	49.20	–
LIMIT [47]	39.97	36.98	33.25	32.63	32.27	32.65	31.96	31.39	33.88	50.80	–
BIDist [44]	47.86	42.68	40.18	37.01	35.41	33.80	33.25	30.93	37.64	59.43	–
CE	43.42	38.64	34.44	32.46	31.90	33.42	31.83	31.95	34.75	51.29	40.44
+ULA (Ours)	47.49	42.10	37.59	35.58	35.00	36.42	34.32	34.50	37.87	51.53	41.40
FACT [46]	46.43	38.65	34.36	32.46	33.47	34.44	33.54	33.34	35.83	51.67	41.11
+ULA (Ours)	47.71	40.13	36.54	34.78	35.45	36.14	35.18	35.03	37.62	51.78	41.72
NCFSCIL [40]	59.22	50.85	44.50	43.98	43.68	44.26	41.27	41.64	46.17	54.66	45.74
+ULA (Ours)	60.79	53.10	45.73	44.69	44.27	44.99	41.81	42.45	47.22	54.52	46.00
SAVC [33]	43.41	34.18	29.10	28.68	30.23	30.12	30.26	29.42	31.92	51.39	39.75
+ULA (Ours)	52.65	42.90	37.94	36.64	37.68	37.34	36.59	36.16	39.73	51.45	41.92

Table C. Harmonic mean accuracy in each session, average harmonic mean accuracy across all sessions (aHM), and global accuracy (*aAcc*, *gAcc*) on CIFAR100.

Method	Class Group	Session-wise Accuracy (%)									Means	*aAcc*	*gAcc*
		0	1	2	3	4	5	6	7	8			
CE	Base	76.70	75.95	75.25	74.54	73.94	73.21	72.73	72.48	71.78	73.73	62.23	53.00
	Incremental	-	30.40	25.99	22.40	20.80	20.40	21.69	20.40	20.55	22.82		
+ULA (Ours)	Base	76.70	75.70	74.81	73.96	73.28	72.45	71.85	71.58	70.66	73.03 (-0.70)	62.39	53.79
	Incremental	-	34.60	29.30	25.20	23.49	23.08	24.40	22.57	22.82	25.68 (+2.86)		
FACT [47]	Base	77.08	76.13	75.31	74.51	73.91	72.98	72.36	72.13	71.65	73.62	62.42	53.46
	Incremental	-	33.40	26.00	22.33	20.80	21.72	22.60	21.85	21.72	23.80		
+ULA (Ours)	Base	77.08	75.86	75.01	74.00	73.35	72.33	71.63	71.38	70.78	73.04 (-0.58)	62.43	53.88
	Incremental	-	34.80	27.40	24.26	22.80	23.48	24.16	23.34	23.27	25.43 (+1.63)		
NC-FSCIL [40]	Base	82.33	79.48	78.53	77.17	75.42	73.25	72.67	73.03	71.52	75.13	66.18	58.68
	Incremental	-	47.20	37.60	31.27	31.05	31.12	31.83	28.77	29.37	33.52		
+ULA (Ours)	Base	82.33	79.03	77.85	76.60	74.88	72.65	71.95	72.37	70.63	74.49 (-0.64)	66.01	58.84
	Incremental	-	49.40	40.30	32.60	31.85	31.84	32.73	29.40	30.35	34.80 (+1.28)		
SAVC [33]	Base	78.27	77.21	76.66	76.03	75.61	74.61	74.16	73.95	73.38	75.20	62.81	52.95
	Incremental	-	30.20	22.00	18.00	17.70	18.96	18.90	19.02	18.40	20.39		
+ULA (Ours)	Base	78.27	76.30	75.30	74.36	73.56	72.01	71.28	70.93	69.45	72.89 (-2.31)	62.89	54.70
	Incremental	-	40.20	30.00	25.46	24.40	25.52	25.30	24.65	24.45	27.49 (+7.10)		

Table D. Base and Incremental accuracy shown per session for CIFAR100.

Method	Class Group	Session-wise Accuracy (%)											Means	*aAcc*	*gAcc*
		0	1	2	3	4	5	6	7	8	9	10			
CE	Base	78.90	78.26	78.03	77.83	77.58	77.33	76.99	76.68	76.48	76.41	76.28	77.18	67.30	62.83
	Incremental	-	48.18	47.04	41.91	44.68	43.51	44.55	45.03	44.06	45.74	45.90	45.06		
+ULA (Ours)	Base	78.90	77.77	77.29	77.12	76.64	76.25	75.63	75.12	74.85	74.55	74.42	75.96 (-1.22)	67.93	64.27
	Incremental	-	59.42	53.84	46.44	49.08	47.43	48.66	49.03	47.48	48.95	48.95	49.92 (+4.86)		
FACT [46]	Base	75.84	74.97	74.77	74.67	74.29	74.15	73.67	73.54	73.22	72.88	72.78	73.89	63.84	59.25
	Incremental	-	53.46	47.60	38.72	40.62	38.53	39.96	40.37	38.80	40.52	40.18	41.87		
+ULA (Ours)	Base	75.84	74.66	74.36	74.29	73.74	73.57	72.94	72.66	72.56	72.12	71.91	73.28 (-0.61)	64.31	60.30
	Incremental	-	61.04	52.98	42.31	43.98	41.42	42.54	42.57	40.86	42.56	42.04	45.23 (+3.36)		
NC-FSCIL [40]	Base	80.80	76.82	78.32	78.53	77.23	76.75	77.09	76.78	76.68	75.94	76.15	77.02	67.64	62.05
	Incremental	-	66.31	47.17	42.25	46.05	43.37	45.31	45.48	42.49	43.57	44.06	46.60		
+ULA (Ours)	Base	80.80	75.80	77.34	77.90	76.33	75.49	75.52	75.31	74.97	74.16	74.65	75.74 (-1.28)	67.70	62.86
	Incremental	-	71.68	50.71	45.14	48.80	45.69	48.06	48.31	44.41	46.05	46.42	49.52 (+2.92)		
SAVC [33]	Base	78.58	78.08	77.68	77.61	77.28	77.08	76.70	76.33	76.06	75.92	75.69	76.84	67.13	62.60
	Incremental	-	48.96	48.16	43.19	44.81	42.08	43.85	44.76	43.70	45.35	45.91	45.07		
+ULA (Ours)	Base	78.58	77.65	77.15	77.05	76.37	75.96	75.42	74.84	74.67	74.47	74.16	75.77 (-1.07)	67.46	63.61
	Incremental	-	56.07	52.23	46.46	48.17	45.52	47.00	47.84	46.41	47.83	48.28	48.58 (+3.51)		

Table E. Base and Incremental accuracy shown per session for CUB200.

Method	Class Group	Session-wise Accuracy (%)									Means	*aAcc*	*gAcc*
		0	1	2	3	4	5	6	7	8			
CE	Base	74.02	73.36	72.85	72.40	72.01	71.65	71.31	71.00	70.54	71.89	60.33	51.14
	Incremental	-	22.00	21.10	20.73	21.10	20.08	18.63	19.51	20.57	20.46		
+ULA (Ours)	Base	74.02	72.78	71.68	70.66	69.93	69.05	68.36	67.63	66.93	69.62 (-2.27)	60.78	53.50
	Incremental	-	33.00	30.50	29.80	29.40	28.20	26.43	27.02	27.45	28.97 (+8.51)		
FACT [46]	Base	75.58	74.93	74.53	73.96	73.71	73.53	73.28	73.08	72.95	73.74	60.06	49.28
	Incremental	-	16.80	15.00	14.06	12.80	12.76	11.80	12.40	12.92	13.56		
+ULA (Ours)	Base	75.58	74.71	74.10	73.33	73.08	72.80	72.45	72.01	71.83	73.03 (-0.71)	60.46	50.47
	Incremental	-	21.00	18.80	18.73	16.95	16.28	14.86	15.80	16.25	17.33 (+3.77)		
NC-FSCIL [40]	Base	83.88	78.88	77.32	74.18	75.50	75.90	76.83	76.80	77.52	76.61	68.24	61.45
	Incremental	-	48.20	45.30	43.40	42.45	37.76	32.83	32.63	29.73	39.03		
+ULA (Ours)	Base	83.88	77.35	75.47	72.80	72.62	73.57	75.03	75.00	75.85	74.71 (-1.90)	67.70	61.83
	Incremental	-	54.40	48.70	46.60	45.25	40.32	35.27	35.54	33.32	42.42 (+3.39)		
SAVC [33]	Base	80.15	79.25	78.73	78.46	78.05	77.76	77.45	77.10	76.96	77.97	65.21	55.17
	Incremental	-	27.40	23.40	23.20	22.90	20.08	18.83	19.57	20.65	22.00		
+ULA (Ours)	Base	80.15	78.05	76.95	76.01	75.28	74.31	73.80	72.95	72.75	75.01 (-2.96)	65.76	58.27
	Incremental	-	41.80	38.40	36.13	34.60	30.68	28.23	28.48	29.00	33.41 (+11.41)		

Table F. Base and Incremental accuracy shown per session for mini-ImageNet.