Appendix

A. Additional Analysis

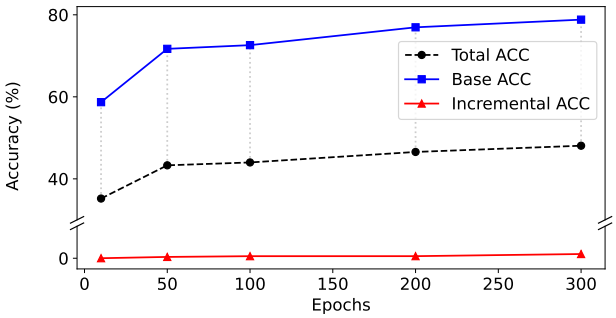


Figure A1. Average accuracies of base and incremental classes in standard joint training on the CIFAR-100 test set by the number of base session training epochs. All results are reported from the last incremental session.

Performance Bias in Standard Joint Training. Figure A1 illustrates the average accuracies of base and incremental classes in the FSCIL setting for standard joint training. We observe that standard joint training achieves near-zero accuracy on incremental classes, and its overall performance is heavily biased toward base classes. Such gap between base and incremental accuracies grows linearly as the number of training epochs in the base session increases.

Table A1. Average false positive (FP) rate and false negative (FN) rate of standard joint training on CIFAR-100 test set under both the CIL and FSCIL settings.

Method	FP rate	FN rate
Standard Joint Training (CIL)	0.254	0.255
Standard Joint Training (FSCIL)	0.488	0.521

Inter-Task Class Separation (ICS) in Joint Training. To highlight the ICS problem in standard joint training under the FSCIL setting, we quantitatively evaluate the quality of decision boundary formation using false positive (FP) and false negative (FN) rates. As shown in Table A1, the joint training model exhibits higher FP and FN rates in the FSCIL setting than in the CIL setting on average. This shows that the ICS problem is more severe in standard joint training under the FSCIL setting compared to the CIL setting.

Exploring Imbalanced Learning in FSCIL. Figure A2 compares the base and incremental class accuracies across the top-5 trials for each method. The imbalance-aware joint training approach achieves markedly higher incremental accuracy, whereas most other approaches—including the standard joint training—show near-zero accuracy on incremental classes, indicating severe overfitting to base classes.

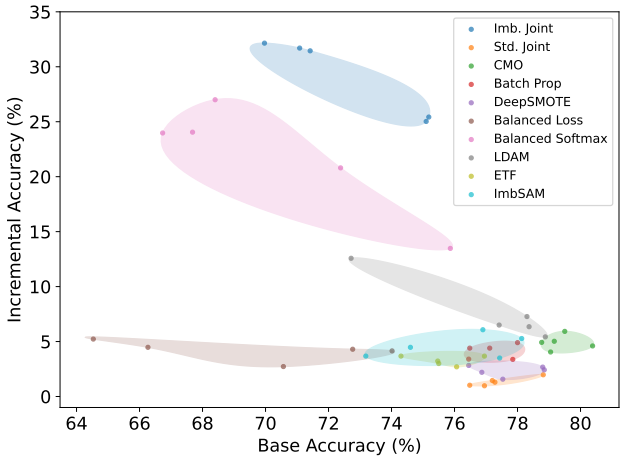


Figure A2. Base and incremental class performance comparison of imbalance-aware joint training (*Imb. Joint*), standard joint training (*Std. Joint*), and 8 imbalance learning techniques on CIFAR-100 test set in the last session of a 5-way 5-shot FSCIL setting. The figure shows the base and incremental class accuracies of the top-5 trials for each method after 30 iterations of hyperparameter random search. The best trials are determined according to the *aAcc*. *Imb. Joint* consists of CMO, Balanced Softmax, and ImbSAM.

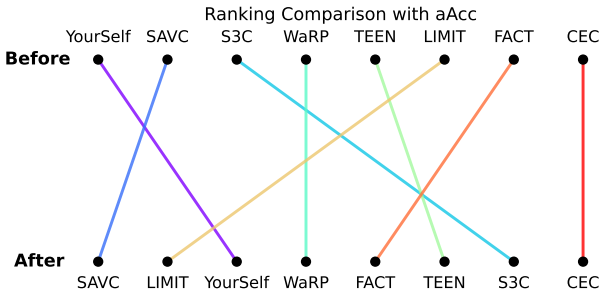
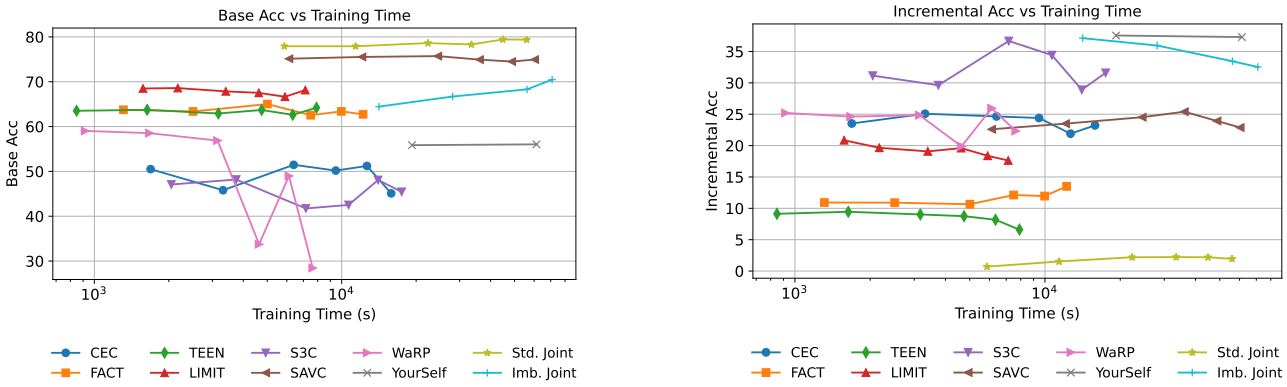


Figure A3. Comparison of FSCIL method performance rankings on CIFAR-100 test set before and after applying our standardized evaluation protocol. Rankings are presented in descending order from left to right.

Performance Rankings Before and After Protocol Standardization. Figure A3 illustrates the ranking shifts of existing FSCIL methods in terms of *aAcc* before and after applying our standardized evaluation protocol. The "**Before**" rankings are based on the values reported in prior studies, whereas the "**After**" rankings reflect the results reproduced under our proposed protocol. All methods except for CEC and WaRP show changes in ranking. Particularly, S3C and LIMIT exhibit the largest shifts, each moving by four positions. These ranking shifts highlight inconsistencies in previous FSCIL experiments and evaluation, underscoring the need for a standardized protocol in future studies.



(a) Base Classes

(b) Incremental Classes

Figure A4. Performance comparison of base and incremental classes across different training times on CIFAR-100 test set. Training times are presented on a log scale. *Std. Joint* and *Imb. Joint* denote standard joint training and imbalance-aware joint training, respectively.

Base and Incremental Performance by Training Time. Figure A4 shows the *aAcc* of base and incremental classes with respect to training time across different methods. For base classes, standard joint training and SAVC achieve the highest performance, benefiting from their longer training durations. In contrast, YourSelf and imbalance-aware joint training mark the best results for incremental classes. Overall, imbalance-aware joint training maintains strong performance on both the base and incremental classes. Notably, TEEN and LIMIT exhibit a trend in which extended training improves base class performance but degrades incremental class performance, likely due to overfitting to base classes as the number of training epochs increases.