Supplementary Material: Multi-Scale Semantic Enrichment and Dual Angular Margin Contrast for Few-Shot Class Incremental Learning

Riya Verma
riyavermacse@gmail.com

Sukhendu Das
sdas@iitm.ac.in

Visualization and Perception Lab
Department of Computer Science and Engineering
Indian Institute of Technology Madras
Chennai, India

1 Introduction

In the supplementary material , Section 2 provides detailed results on three benchmark datasets, miniImagnet, CIFAR100, and CUB200, for several previous SOA methods. Section 3 provides visualization, and section 4 presents discussion and hyperparameters analysis.

2 Detailed Results

Tables 1- 3 present detailed results on three benchmark datasets. Our method achieves final accuracy rates of 65.32% for CUB200, 62.78% for miniImagenet, and 61.02% for CIFAR100, outperforming the current state-of-the-art (SOTA) methods by margins of 2.82% (SAVC [15]), 2.04% (MICS [11]), and 2.52% (RCN [19]), respectively. Furthermore, our approach attains average accuracies of 71.43%, 71.29%, and 70.12% across the CUB200, miniImagenet, and CIFAR100 datasets, respectively, surpassing all other methods.

In Table 4, we analyze the False Negative Rate (FNR) and False Positive Rate (FPR) percentages from session 1 as defined in TEEN [18] on the miniImagenet dataset. Notably, a high FPR and a comparatively low FNR are obtained, indicating that new class instances are frequently misidentified as base classes or similar novel classes. Our method stands out by achieving a substantially lower FPR compared to baseline approaches, underscoring the effectiveness of its introduced semantically rich multi-scale feature extraction strategy aided by a contrastive framework.

Figure 1 indicates the Harmonic Mean [9] for different methods under 5-shot incremental scenarios on CUB200 dataset indicated by line graph. We outperform all the benchmarked methods including the latest method OrCo [1]. For 1-shot, we compare with the state-of-the-art, indicated by bar graph. Our multi-scale approach shows its efficiency in the case of

Method	Accuracy in each session (%) ↑									Avg ↑	△last
	0	1	2	3	4	5	6	7	8		
Ft-CNN [16]	61.31	27.22	16.37	6.08	2.54	1.56	1.93	2.60	1.40	13.45	-
iCaRL[a] [13]	61.31	46.32	42.94	37.63	30.49	24.00	20.81	18.80	17.21	33.28	+15.81
EEIL[a] [0]	61.31	46.58	44.00	37.29	33.14	27.12	24.10	21.57	19.58	34.97	+18.18
TOPIC[a] [15]	61.31	50.09	45.17	41.16	37.48	35.52	32.19	29.46	24.42	39.64	+23.02
Rebalancing [0]	61.31	47.80	39.31	31.91	25.68	21.35	18.67	17.24	14.17	30.83	+12.77
SPPR [26]	61.35	63.80	59.53	55.83	52.35	49.60	46.49	43.24	41.92	52.76	+40.52
F2M [4]	72.05	67.47	63.16	59.70	56.71	53.77	51.11	49.21	47.84	57.89	+46.44
CEC [23]	72.00	66.83	62.97	59.43	56.70	53.73	51.19	49.24	47.63	57.75	+46.23
MCNet [8]	72.33	67.70	63.50	60.34	57.59	54.70	52.13	50.41	49.08	58.64	+47.68
MetaFSCIL [0]	72.04	67.94	63.77	60.29	57.58	55.16	52.90	50.79	49.19	58.85	+47.79
MFS3 [20]	73.65	68.91	64.60	61.48	58.68	55.55	53.33	51.69	50.26	59.79	+48.86
FACT [22]	72.56	69.63	66.38	62.77	60.60	57.33	54.34	52.16	50.49	60.70	+49.09
SoftNet [2]	79.77	75.08	70.59	66.93	64.00	61.00	57.81	55.81	54.68	65.07	+53.28
ALICE [1]	80.60	70.60	67.40	64.50	62.50	60.00	57.80	56.80	55.70	63.99	+54.30
LIMIT [5]	72.32	68.47	64.30	60.78	57.95	55.07	52.70	50.72	49.19	59.06	+47.79
WaRP [10]	72.99	68.10	64.31	61.30	58.64	56.08	53.40	51.72	50.65	59.69	+49.25
NC-FSCIL [20]	84.02	76.80	72.00	67.83	66.00	64.04	61.46	59.54	58.31	67.82	+56.91
FCIL [8]	76.34	71.40	67.10	64.08	61.30	58.51	55.72	54.08	52.76	62.37	+51.36
SAVC [13]	81.12	76.14	72.43	68.92	66.48	62.95	59.92	58.39	57.11	67.05	+55.71
RCN* [19]	83.40	78.75	74.94	70.81	67.84	64.89	63.10	60.92	58.53	69.24	+57.13
TEEN [13]	73.53	70.55	66.37	63.23	60.53	57.95	55.24	53.44	52.08	61.41	+50.68
EHS* [0]	71.25	66.65	62.84	59.65	56.90	54.14	51.63	50.05	49.06	58.01	+47.66
MICS [10]	**84.40**	_79.4_	_75.09_	_71.40_	_68.89_	_66.16_	**63.57**	_61.79_	_60.74_	_70.16_	_+59.34_
Ours	_82.97_	**79.96**	**76.93**	**74.07**	**69.39**	**68.98**	_63.48_	**63.04**	**62.78**	**71.29**	**+61.38**

Table 1: Comparison of the performance of different methods on the miniImagenet dataset. SOA results are highlighted in bold, while the second-best outcomes are underlined. a: indicates that results are copied from [23]. *: indicates results directly copied from published literature. △last: Relative improvements of the last session compared to the Ft-CNN [16] model.
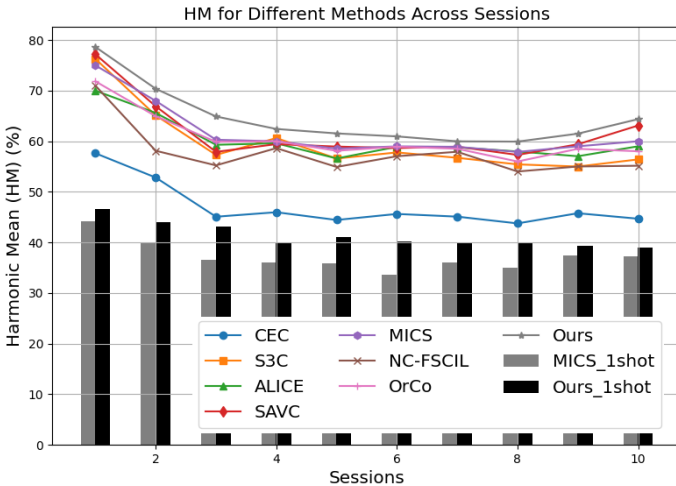

Figure 1: Harmonic Mean for different methods across sessions on CUB200. The line graph depicts the harmonic mean under 5-shot setting, while the bar graph illustrates the harmonic accuracy under 1-shot setting.

Method	Accuracy in each session (%) ↑									Avg ↑	$\triangle_{last}$
	0	1	2	3	4	5	6	7	8		
Ft-CNN [16]	64.10	39.61	15.37	9.80	6.67	3.80	3.70	3.14	2.65	16.54	-
iCaRL[a] [3]	64.10	53.28	41.69	34.73	27.93	25.06	20.41	15.48	13.73	32.93	+11.08
EEIL[a] [1]	64.10	53.11	43.71	35.15	28.96	24.98	21.01	17.26	15.85	33.79	+13.20
TOPIC[a] [16]	64.10	55.88	47.07	45.01	40.11	36.38	33.39	31.55	29.37	42.54	+26.72
Rebalancing [0]	64.10	53.05	43.96	36.97	31.61	26.73	21.23	16.78	13.54	34.22	+10.89
SPPR [26]	63.97	65.86	61.31	57.60	53.39	50.93	48.27	45.36	43.32	54.45	+40.67
F2M [12]	71.45	68.10	64.43	60.87	57.76	55.26	53.53	51.57	49.35	59.15	+46.70
CEC [23]	73.07	68.88	65.26	61.19	58.09	55.57	53.23	51.34	49.14	59.53	+46.49
MCNet [8]	73.30	69.34	65.72	61.70	58.75	56.44	54.59	53.01	50.72	60.40	+48.07
MetaFSCIL [0]	74.50	70.10	66.84	62.77	59.48	56.52	54.36	52.56	49.97	60.79	+47.32
MFS3 [20]	73.42	69.85	66.44	62.81	59.78	56.94	55.04	53.00	51.07	60.93	+48.42
FACT [24]	74.60	72.09	67.56	63.52	61.38	58.36	56.58	54.24	52.10	62.27	+49.45
C-FSCIL	77.47	72.40	67.47	63.25	59.84	56.95	54.42	52.47	50.47	61.64	+47.82
SoftNet [12]	79.88	75.54	71.64	67.47	64.45	61.09	59.07	57.29	55.33	65.75	+52.68
ALICE [13]	79.00	70.50	67.10	63.40	61.20	59.20	58.10	56.30	54.10	63.21	+51.45
LIMIT [25]	73.81	72.09	67.87	63.89	60.70	57.77	55.67	53.52	51.23	61.84	+48.58
WaRP [10]	80.31	75.86	71.87	67.58	64.39	61.34	59.15	57.10	54.74	65.82	+52.09
NC-FSCIL [20]	82.62	78.62	73.34	69.68	66.19	62.85	60.96	59.02	56.11	67.71	+53.46
FCIL [8]	77.12	72.42	68.31	64.47	61.18	58.17	56.06	54.19	52.02	62.66	+49.37
SAVC [13]	78.77	73.31	69.31	64.93	61.70	59.25	57.13	55.19	53.12	63.63	+50.47
RCN* [19]	83.40	78.75	74.94	70.81	67.84	**64.89**	**63.10**	60.92	58.53	69.24	+55.88
TEEN [13]	74.92	72.65	68.74	65.01	62.01	59.29	57.90	54.76	52.64	63.10	+49.99
EHS* [0]	73.98	70.11	66.66	62.75	60.11	57.33	55.59	53.75	51.59	61.31	+48.94
MICS [01]	78.18	73.49	68.97	65.01	62.25	59.34	57.31	55.11	52.94	63.62	+50.29
Ours	**83.44**	**79.99**	**75.57**	**71.42**	**69.54**	64.65	62.95	**62.55**	**61.05**	**70.12**	**+58.40**

Table 2: Comparison of the performance of different methods on the CIFAR100 dataset. SOA results are highlighted in bold, while the second-best outcomes are underlined. a: indicates that results are copied from [23]. *: indicates results directly copied from published literature. $\triangle_{last}$: Relative improvements of the last session compared to the Ft-CNN [16] model.

1-shot as well.

3 Visualizations

Figure 2a compares the confusion matrices generated by the SAVC and our method when applied to the CUB200 dataset. A sharper diagonal in our confusion matrix indicates more correct predictions by our model, and fewer spreads outside the diagonal suggest that our model is less confused and does not often misclassify between different classes compared to SAVC. In figure 4, we visualize the top 3 output probabilities for some input images from miniImagenet dataset for the baseline method, our set of 4 mappers, and the overall result of our method using the Sum-min metric. Our set-based approach performs well on confusing data, acts as an ensemble, and correctly identifies the output classes. The third row of figure 4 displays an inaccurate result for our method; nonetheless, our approach remains superior since the compared method did not rank the true class within the top three scores. Our multi-scale strategy results in fewer errors when classifying novel classes and also addresses the bias towards base classes. Figure 2b we depict the decision boundaries learned from the CIFAR100 dataset using t-SNE [17]. The visualization represents the decision boundaries formed during the base session, where the model is trained on five old classes and five new classes. Our approach evidently reserves space for future classes (indicated by the gray

Method	Accuracy in each session (%) ↑											Avg ↑	$\triangle_{last}$
	0	1	2	3	4	5	6	7	8	9	10		
Ft-CNN [☐]	68.68	43.70	25.05	17.72	18.08	16.95	15.10	10.06	8.93	8.93	8.47	21.97	-
iCaRL[a] [☐]	68.68	52.65	48.61	44.16	36.62	29.52	27.83	26.26	24.01	23.89	21.16	36.6	+12.69
EEIL[a] [☐]	68.68	53.63	47.91	44.20	36.30	27.46	25.93	24.70	23.95	24.23	22.11	36.27	+13.64
TOPIC[a] [☐]	68.68	62.49	54.81	49.99	45.25	41.40	38.35	35.36	32.22	28.31	26.28	43.92	+17.81
Rebalancing [☐]	68.68	57.12	44.21	28.78	26.71	25.66	24.62	21.52	20.12	20.06	19.87	32.49	+11.4
SPPR [☐]	68.68	61.85	57.43	52.68	50.19	46.88	4.65	43.07	40.17	39.63	37.33	49.34	+28.86
F2M [☐]	77.13	73.92	70.27	66.37	64.34	61.69	60.52	59.38	57.15	56.94	55.89	63.96	+47.42
CEC [☐]	75.85	71.94	68.50	63.50	62.43	58.27	57.73	55.81	54.83	53.52	52.28	61.33	+43.81
MCNet [☐]	77.57	73.96	70.47	65.81	66.16	63.81	62.09	61.08	60.41	60.09	59.08	65.57	+50.61
MetaFSCIL [☐]	75.90	72.41	68.78	64.78	62.96	59.99	58.30	56.85	54.78	53.82	52.64	61.93	+44.17
MFS3 [☐]	75.63	72.51	69.65	65.29	63.13	60.38	58.99	57.41	55.55	54.93	53.47	62.45	+45.00
FACT [☐]	75.90	73.23	70.84	66.13	65.56	62.15	61.74	59.83	58.41	57.89	56.94	64.42	+48.47
SoftNet [☐]	78.07	74.58	71.37	67.54	65.37	62.60	61.07	59.37	57.53	57.21	56.75	64.68	+48.28
ALICE [☐]	77.40	72.70	70.60	67.20	65.90	63.40	62.91	60.90	60.50	60.01	60.10	65.75	+51.63
S3C [☐]	80.62	77.55	73.19	68.54	68.05	64.33	63.58	62.07	60.61	59.79	58.95	67.03	+50.48
LIMIT [☐]	75.89	73.55	71.99	68.14	67.42	63.61	62.40	61.35	59.91	58.66	57.41	65.48	+48.94
WaRP [☐]	77.74	74.15	70.82	66.90	65.01	62.64	61.40	59.86	57.95	57.77	57.01	64.66	+48.54
NC-FSCIL [☐]	80.45	75.98	73.20	<u>70.28</u>	68.17	65.16	64.43	63.25	60.66	60.01	59.44	67.28	+ 50.97
FCIL [☐]	78.70	75.12	70.10	66.26	66.51	64.01	62.69	61.00	60.36	59.45	58.48	62.37	+50.01
SAVC [☐]	<u>81.85</u>	<u>77.92</u>	<u>74.95</u>	70.21	<u>69.96</u>	<u>67.02</u>	<u>66.16</u>	<u>65.30</u>	<u>63.84</u>	<u>63.15</u>	62.50	<u>69.35</u>	<u>+54.03</u>
RCN* [☐]	79.86	76.48	73.34	69.72	68.48	65.93	64.58	63.68	62.04	61.48	60.47	67.82	+52.00
TEEN [☐]	77.26	76.13	72.81	68.16	67.77	64.40	63.25	62.29	61.19	60.32	59.31	66.63	+50.84
MICS [☐]	78.77	75.37	72.30	68.72	67.45	65.40	64.72	63.39	61.89	61.89	61.37	67.39	+52.90
Ours	**83.08**	**79.99**	**77.25**	72.35	**71.65**	**69.02**	**68.13**	**67.98**	**65.06**	**65.92**	**65.32**	**71.43**	

Table 3: Comparison of the performance of different methods on the CUB200 dataset. SOA results are highlighted in bold, while the second-best outcomes are underlined. a: indicates that results are copied from [23]. *: indicates results directly copied from published literature. $\triangle_{last}$: Relative improvements of the last session compared to the Ft-CNN [16] model.

Table 4: False Negative Rate/False Positive Rate(%) on miniImageNet dataset for different methods.
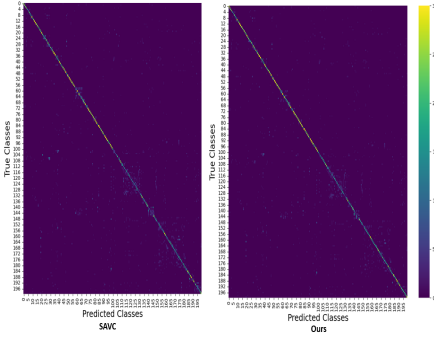
	Session 1		Session 2		Session 3		Session 4		Session 5		Session 6		Session 7		Session 8	
FNR/FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR
CEC [☐]	3.45	68.40	5.60	65.50	7.03	61.93	7.73	58.30	8.47	56.68	9.58	54.50	10.22	52.54	10.93	50.05
TEEN [☐]	8.02	46.40	11.35	38.60	13.12	37.53	15.32	35.20	16.47	32.48	17.38	31.57	18.63	28.03	19.97	26.35
MICS [☐]	11.08	42.79	13.71	30.63	14.00	35.35	10.06	20.08	15.89	27.33	16.04	20.63	21.56	23.45	12.30	16.14
Our	11.24	**35.97**	13.69	**23.47**	14.19	**26.05**	12.94	**15.89**	15.76	**19.58**	17.50	**21.89**	22.78	**17.37**	11.92	**14.85**

areas). The transition to the incremental stage is illustrated, marked by the introduction of 5 new classes (depicted as dots). The preserved space facilitates the incremental learning phase, allowing new classes to integrate into the embedding space without compressing the embedding of existing classes.

4 Discussion and Hyperparameters Analysis

We report the comparison of the number of trainable parameters of several typical methods in Figure 3a. We can infer that the number of parameters used in our model is comparable with most methods, indicating that performance gain is not due to over-parameterization. The training time for CUB200 is shown in Figure 3b ,our method is at par with other related works. Figure 5 illustrates mapper activation on various categories of miniImagenet dataset. This demonstrates that mappers at the low level are frequently engaged similarly to those at the high level. Thus, our approach benefits from a set of features compared to a single feature per image.

Figure 7a compares the (R^2) class separation degree for CE (cross-entropy), SAVC, and our method. We achieved a higher mutual class separation degree, indicating we can separate
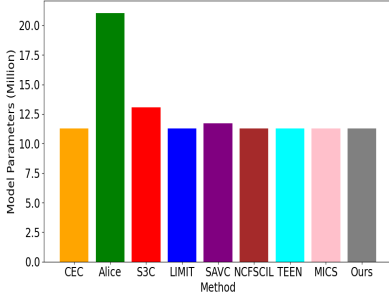
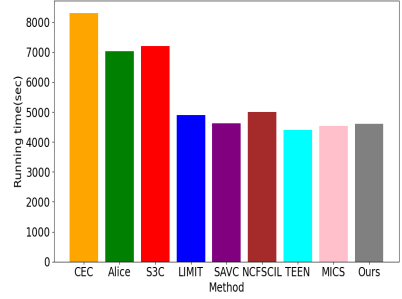(a) Confusion matrices comparison of SAVC and our method on CUB200 dataset.

(b) t-SNE visualization of CIFAR100, Original classes are marked with triangles, new classes with dots, and shaded areas show class decision regions.

Figure 2: Visualizations.

(a) Parameter count in millions.

(b) Training time analysis.

Figure 3: Parameter count and training time analysis on CUB200 for several methods

the novel classes well from the old ones.

For double angular margin contrastive loss used in our method, we analyzed different values of the scale factor s. As shown in figure 6a, s = 30 or 20 gives good results in most sessions. Thus, for all our experiments, we set the s to 30.

Figure 6b analyzes values of positive (m_p) and negative (m_n) margins on CIFAR100 dataset. The plot indicates that it is better to set m_n greater than m_p. Parameter m_n increases the distance threshold for negative pairs. For more diverse datasets like CIFAR100 or miniImagenet, a larger value of m_n helps separate distinct classes well. C manages the numerical range and stability of the cotangent outputs in the loss function. It prevents excessively large gradients by clamping the cotangent function's output, which may cause the learning process to diverge, particularly near its undefined points. Clamping ensures stable, discriminative feature learning, which is crucial for model convergence. Our selection of C=20 is supported by accuracy variations shown in the figure 6c. Temperature $\tau = 20$ gives the optimal performance as depicted in figure 6d.

Figure 7b depicts applying Layerwise Feature Augmentation (LFA) gives better performance over uniform feature augmentation. LFA before mappers performs better as mappers
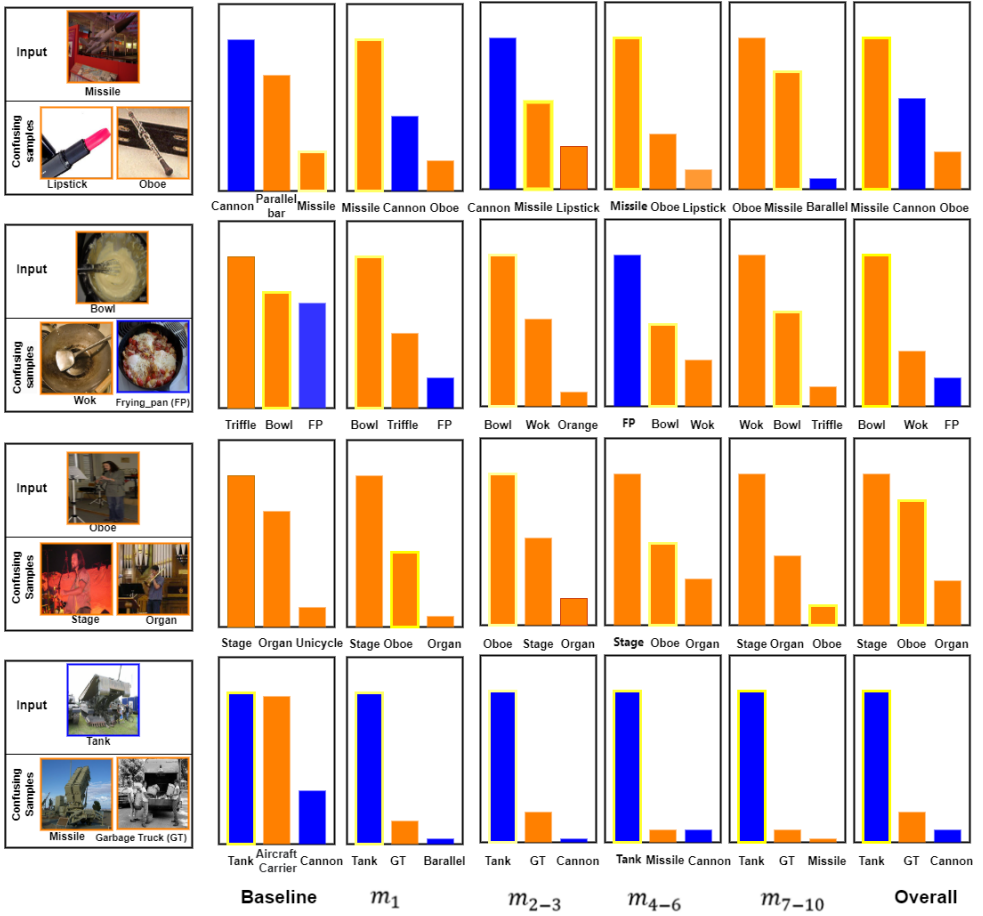
Figure 4: Top 3 prediction probabilities for the baseline [13], mappers (m_1, m_{2-3}, m_{4-6} *and* m_{7-10}) and our cumulative result. Blue and orange colors represent base and incremental classes, respectively. A yellow edge indicates the ground-truth class.
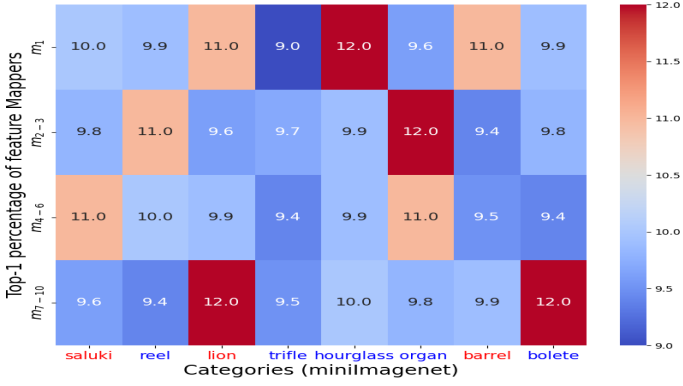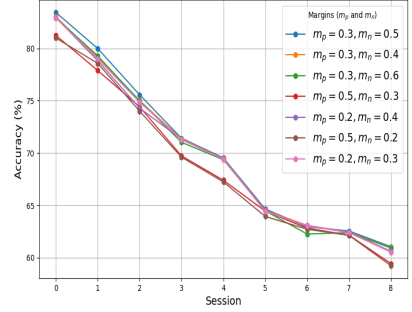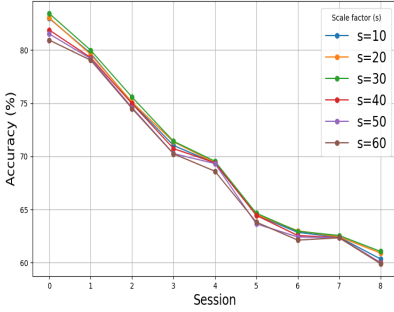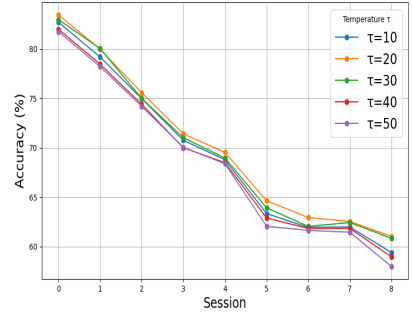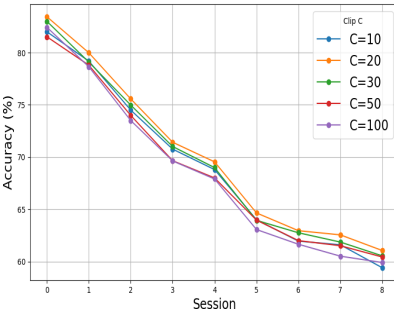
Figure 5: The heatmap represents the frequency (as a percentage) with which each feature mapper (y-axis) is selected when classifying samples from various categories (x-axis) within a subset of the miniImageNet test set. These values are calculated by averaging the selection frequency of each mapper across all 200 samples within their respective categories. Base classes, which are part of the initial training set, are indicated in red, while incremental classes, added during the testing phase, are shown in blue.
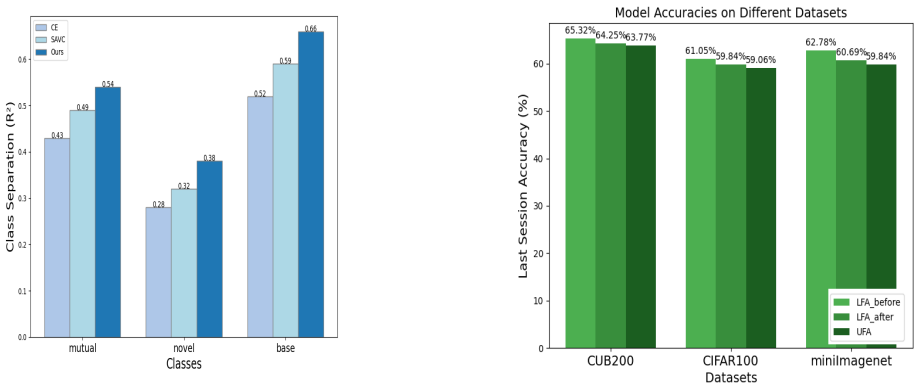

(a) Hyperparameter study for scale factor (s). s = 30 gives the optimal performance

(b) Hyperparameter study for margins. m_p: positive margin, m_n: negative margin.

(c) Hyperparameter study for Clamp C.

(d) Hyperparameter study for Temperature τ.

Figure 6: Hyperparameter Analysis on CIFAR100

further enhance the augmented features learning more generalized representations.

(a) Class separation degree (R^2) defined in SAVC [15]. Higher mutual class separation degree indicates good separation between the novel classes and the old classes.

(b) Analysis of Feature Augmentation. UFA: Uniform Feature Augmentation, LFA_before: LFA applied before mappers, LFA_after:LFA applied after mappers.

Figure 7: Ablation Study on Class Separation Degree and Feature Augmentation.

Table 5 shows the impact of the number of mappers on the average accuracy for the CIFAR100 dataset. The optimal accuracy is achieved with ten mappers. For a fair comparison, we used a ResNet [6] architecture with four blocks. Increasing the number of mappers doesn't necessarily improve performance because it requires reducing the number of filters to maintain comparable parameters, leading to underfitting due to under-parameterization. However, using larger backbones with more mappers could potentially yield better performance.

No. of Mappers	Average Accuracy (%)
4	68.56
8	69.29
10	70.12
12	70.01
14	69.76

Table 5: Impact of the Number of Mappers on Average Accuracy for the CIFAR100 Dataset.

References

[1] Noor Ahmed, Anna Kukleva, and Bernt Schiele. Orco: Towards better generalization via orthogonality and contrast for few-shot class-incremental learning. *arXiv preprint arXiv:2403.18550*, 2024.

[2] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 233–248, 2018.

[3] Zhixiang Chi, Li Gu, Huan Liu, Yang Wang, Yuanhao Yu, and Jin Tang. Metafscil: A meta-learning approach for few-shot class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14166–14175, 2022.

[4] Yao Deng and Xiang Xiang. Expanding hyperspherical space for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1967–1976, 2024.

[5] Ziqi Gu, Chunyan Xu, Jian Yang, and Zhen Cui. Few-shot continual infomax learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19224–19233, 2023.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[7] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 831–839, 2019.

[8] Zhong Ji, Zhishen Hou, Xiyao Liu, Yanwei Pang, and Xuelong Li. Memorizing complementation network for few-shot class-incremental learning. *IEEE Transactions on Image Processing*, 32:937–948, 2023.

[9] Jayateja Kalla and Soma Biswas. S3c: Self-supervised stochastic classifiers for few-shot class-incremental learning. In *European Conference on Computer Vision*, pages 432–448. Springer, 2022.

[10] Do-Yeon Kim, Dong-Jun Han, Jun Seo, and Jaekyun Moon. Warping the space: Weight space rotation for class-incremental few-shot learning. In *The Eleventh International Conference on Learning Representations*, 2022.

[11] Solang Kim, Yuho Jeong, Joon Sung Park, and Sung Whan Yoon. Mics: Midpoint interpolation to learn compact and separated representations for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2236–2245, 2024.

[12] Can Peng, Kun Zhao, Tianren Wang, Meng Li, and Brian C Lovell. Few-shot class-incremental learning from an open-set perspective. In *European Conference on Computer Vision*, pages 382–397. Springer, 2022.

[13] SA Rebuffi, A Kolesnikov, G Sperl, CH Lampert, and icarl. Incremental classifier and representation learning. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5533–5542.

[14] Guangyuan Shi, Jiaxin Chen, Wenlong Zhang, Li-Ming Zhan, and Xiao-Ming Wu. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima. *Advances in Neural Information Processing Systems*, 34:6747–6761, 2021.

[15] Zeyin Song, Yifan Zhao, Yujun Shi, Peixi Peng, Li Yuan, and Yonghong Tian. Learning with fantasy: Semantic-aware virtual contrastive constraint for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24183–24192, 2023.

[16] Xiaoyu Tao, Xiaopeng Hong, Xinyuan Chang, Songlin Dong, Xing Wei, and Yihong Gong. Few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12183–12192, 2020.

[17] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

[18] Qi-Wei Wang, Da-Wei Zhou, Yi-Kai Zhang, De-Chuan Zhan, and Han-Jia Ye. Few-shot class-incremental learning via training-free prototype calibration. *Advances in Neural Information Processing Systems*, 36, 2024.

[19] Ye Wang, Yaxiong Wang, Guoshuai Zhao, and Xueming Qian. Learning to complement: Relation complementation network for few-shot class-incremental learning. *Knowledge-Based Systems*, 282:111130, 2023.

[20] Xinlei Xu, Saisai Niu, Zhe Wang, Wei Guo, Lihong Jing, and Hai Yang. Multi-feature space similarity supplement for few-shot class incremental learning. *Knowledge-Based Systems*, 265:110394, 2023.

[21] Yibo Yang, Haobo Yuan, Xiangtai Li, Zhouchen Lin, Philip Torr, and Dacheng Tao. Neural collapse inspired feature-classifier alignment for few-shot class incremental learning. *arXiv preprint arXiv:2302.03004*, 2023.

[22] Jaehong Yoon, Sultan Madjid, Sung Ju Hwang, Chang-Dong Yoo, et al. On the soft-subnetwork for few-shot class incremental learning. In *International Conference on Learning Representations (ICLR) 2023*. International Conference on Learning Representations, 2023.

[23] Chi Zhang, Nan Song, Guosheng Lin, Yun Zheng, Pan Pan, and Yinghui Xu. Few-shot incremental learning with continually evolved classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12455–12464, 2021.

[24] Da-Wei Zhou, Fu-Yun Wang, Han-Jia Ye, Liang Ma, Shiliang Pu, and De-Chuan Zhan. Forward compatible few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9046–9056, 2022.

[25] Da-Wei Zhou, Han-Jia Ye, Liang Ma, Di Xie, Shiliang Pu, and De-Chuan Zhan. Few-shot class-incremental learning by sampling multi-phase tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[26] Kai Zhu, Yang Cao, Wei Zhai, Jie Cheng, and Zheng-Jun Zha. Self-promoted prototype refinement for few-shot class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6801–6810, 2021.