S3C: Self-Supervised Stochastic Classifiers for FSCIL

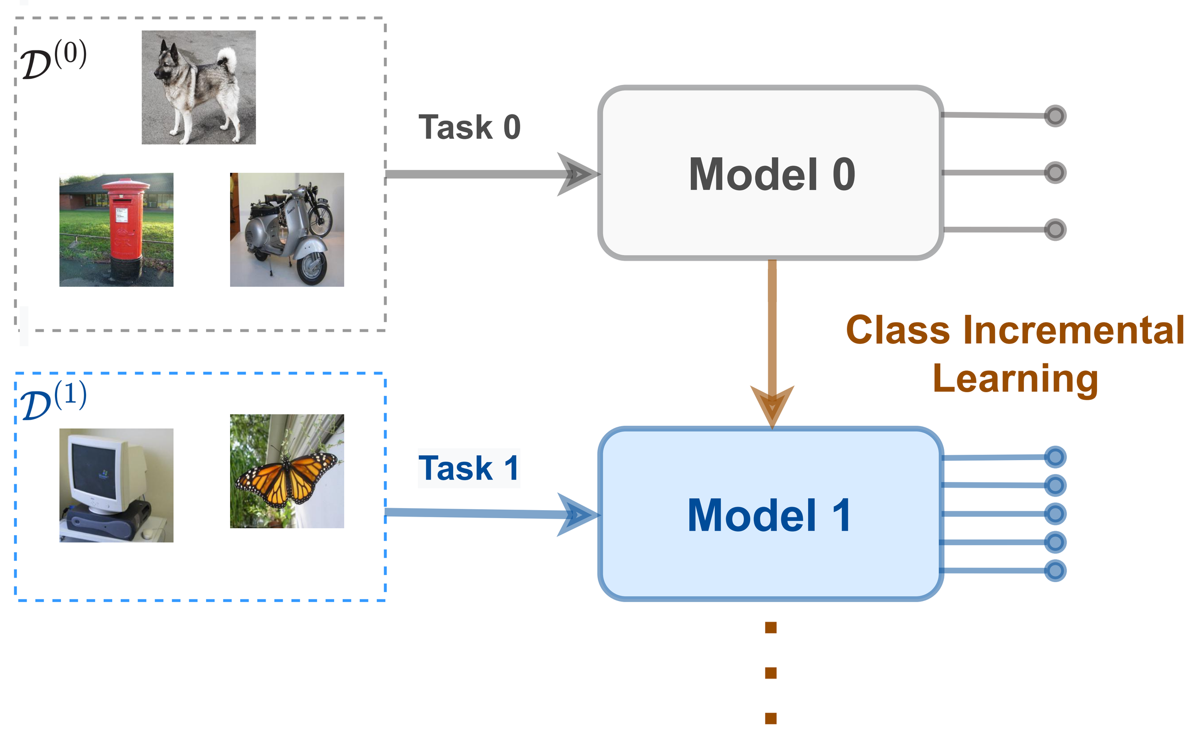
Jayateja Kalla & Soma Biswas

Department of Electrical Engineering, Indian Institute of Science, Bengaluru, India.

TEL AVIV 2022

Class-Incremental Learning

- Ability to learn from continuously evolving data is important for many real-world applications.
- In class-incremental learning, the model tries to update its knowledge continuously, when the set of new classes are available.



Why we need class-incremental learning?

- Data collection
- Data unavailability
- Computational expensive

Major Challenge: Catastrophic forgetting on old classes

Few-Shot Class-Incremental Learning

- The training data at each incremental task contains few labeled samples per class.
- Eliminates the need for collecting large amount of annotated data at each incremental step.

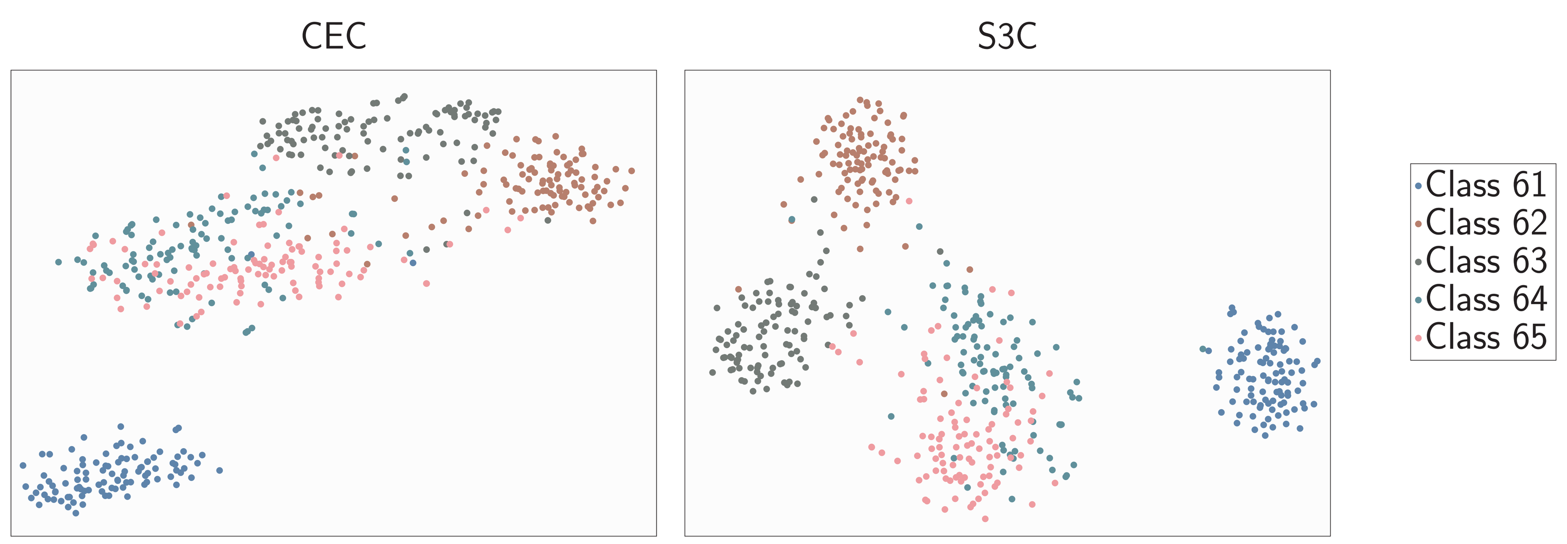
Major Challenges:

- **Over-fitting** on new classes
- **Catastrophic forgetting** on old classes

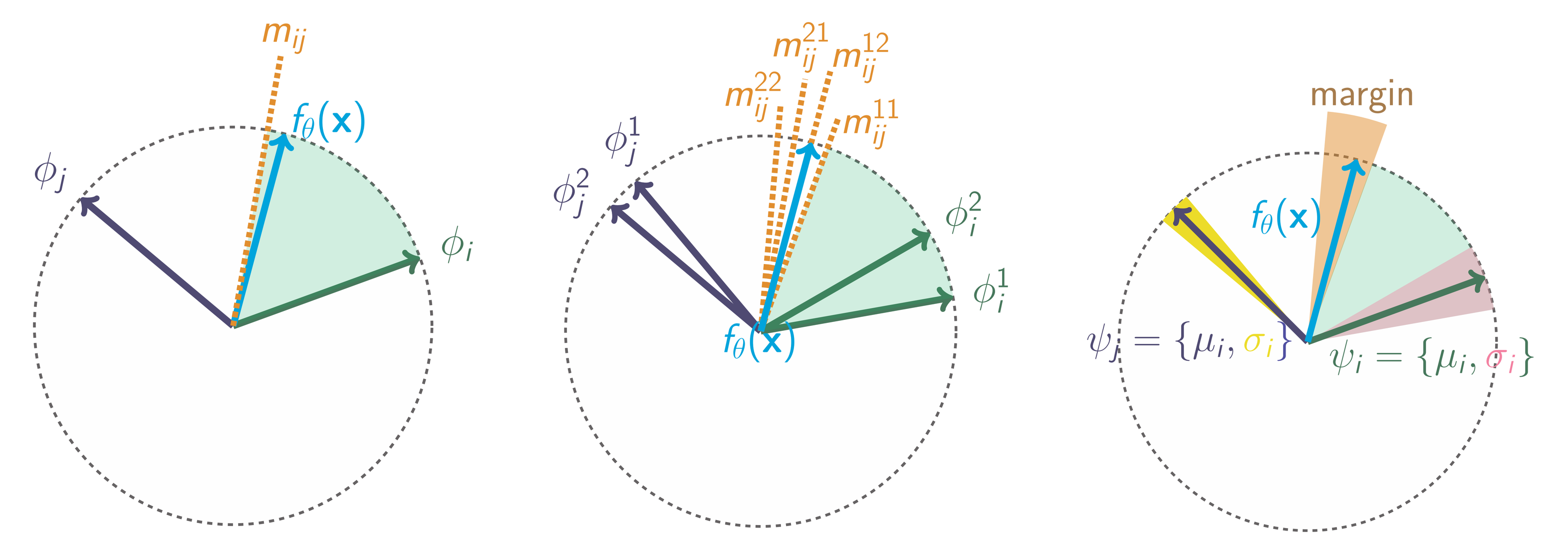
Contributions

- Proposed a novel framework, termed **S3C (Self-Supervised Stochastic Classifiers)** to address the FSCIL task.
- Studied the effectiveness of stochastic classifiers for FSCIL to mitigate over-fitting on new classes.
- Self-supervision along with stochastic classifiers can be used to effectively learn new classes, without hindering the performance on the old classes.
- Extensive experiments on three benchmark datasets, namely CIFAR100, CUB200 and miniImageNet show the effectiveness of proposed approach.

Effectiveness of Proposed S3C on new task classes

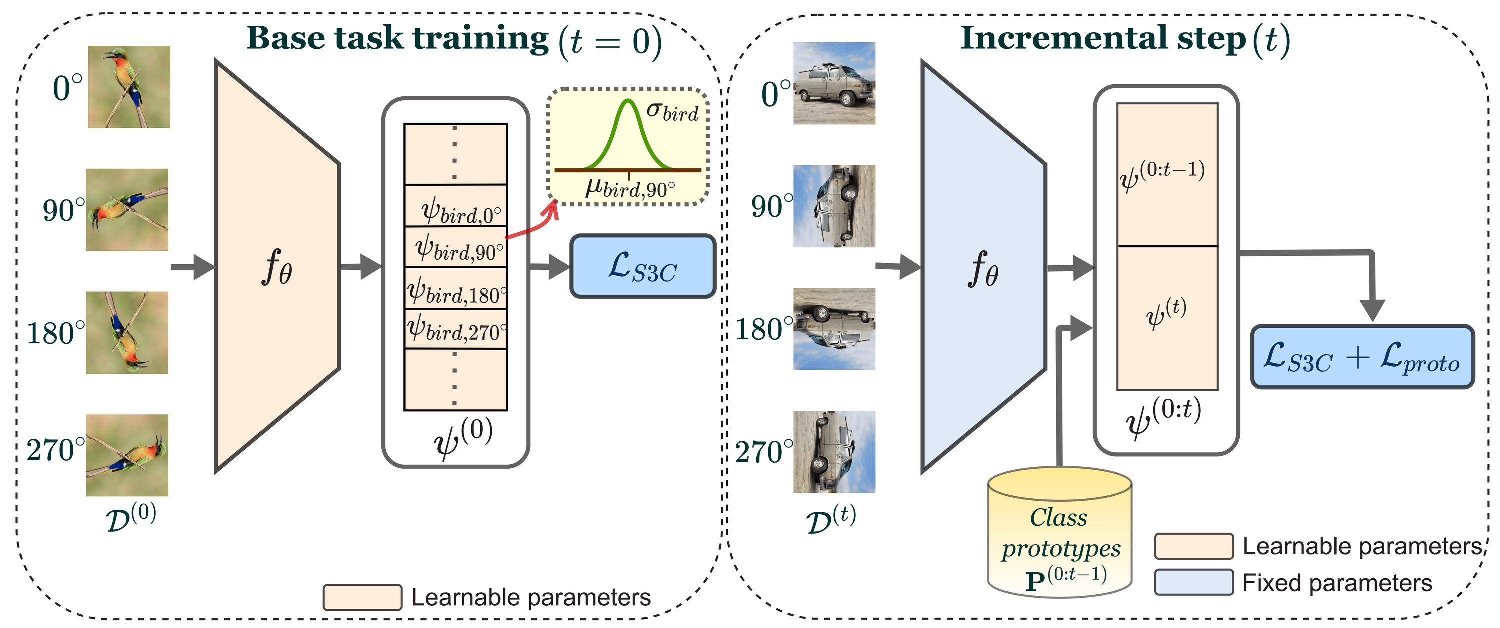


Stochastic Classifiers



a). Standard Classifier (ϕ) b). 2-Classifiers/Class (ϕ) c). Stochastic Classifier (ψ)

Proposed S3C framework



$$\mathcal{L}_{S3C}(\mathbf{x}, y;\ \theta, \psi^{(0:t)}) = -\frac{1}{M}\sum_{r=1}^{M}\log(\rho_{yr}^{(s)}(\widetilde{\mathbf{x}}_r;\ \theta, \psi^{(0:t)}))$$

$$\mathcal{L}_{base} = \mathcal{L}_{S3C}(\mathbf{x}, y;\ \theta, \psi^{(0)})$$

$$\mathcal{L}_{inc}^{(t)} = \lambda_1 \cdot \mathcal{L}_{proto}(q, \check{y}, \psi^{(0:t)}) + \lambda_2 \cdot \mathcal{L}_{S3C}(\mathbf{x}, y;\ \theta, \psi^{(0:t)})$$

Testing

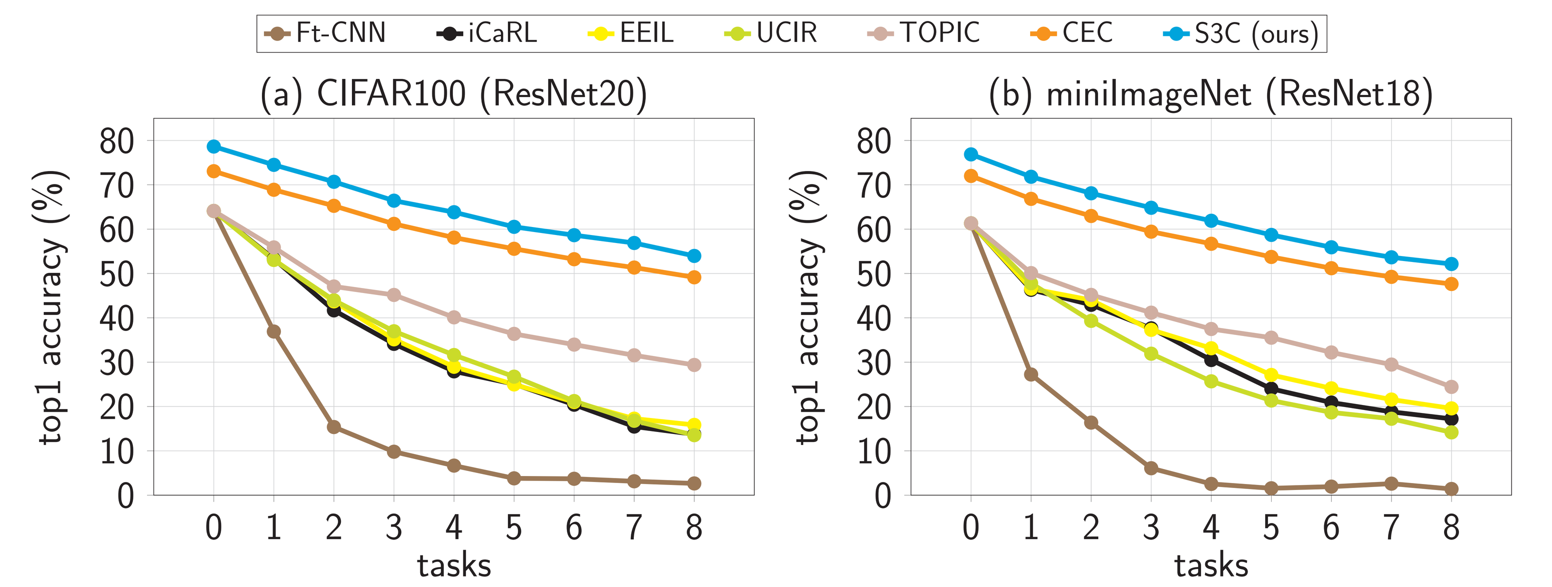
The aggregate score for the i^{th} class in task s for a given test image $\mathbf{x}$ is computed as $z_i^{(s)} = \frac{1}{M}\sum_{r=1}^{M}\eta\ \langle\ \overline{\mu_{ir}^{(s)}}, \overline{f_\theta(\widetilde{\mathbf{x}}_r)}\rangle$.

Then the final prediction $\hat{i}, \hat{s} = arg \max_{i,s} P_{agg}(i, s/\mathbf{x})$. where $P_{agg}(i, s/\mathbf{x}, \theta, \psi^{(0:t)}) = \frac{\exp(z_i^{(s)})}{\sum_{j=0}^{t}\sum_{k=1}^{|C^{(j)}|}\exp(z_k^{(j)})}$

Ablation Study on CIFAR100

self-supervision	classifier	After task 0 base task accuracy	After task $\mathcal{T}$ top1 accuracy	After task $\mathcal{T}$ harmonic mean
✗	linear	74.70	48.98	26.76
✓	linear	76.14	53.55	41.80
✓	stochastic	78.03	53.96	45.22

Experimental Results on CIFAR100 and miniImageNet



Dataset	Method	Harmonic Mean (%) ↑							
		1	2	3	4	5	6	7	8
CIFAR100	CEC	41.57	38.75	32.36	31.53	32.55	32.40	32.25	31.27
	S3C (Ours)	**61.60**	**54.57**	**48.94**	**47.60**	**47.00**	**46.75**	**45.96**	**45.22**
miniImageNet	CEC	31.68	30.86	29.52	29.01	26.75	24.46	26.14	26.24
	S3C (Ours)	**35.30**	**38.18**	**40.62**	**38.86**	**35.02**	**34.49**	**36.06**	**36.20**

Experimental Results on CUB200

Method	Accuracy in each session (%) ↑											PD ↓	Our relative improvement
	0	1	2	3	4	5	6	7	8	9	10		
Ft-CNN	68.68	43.7	25.05	17.72	18.08	16.95	15.1	10.6	8.93	8.93	8.47	60.21	+**39.83**
iCaRL	68.68	52.65	48.61	44.16	36.62	29.52	27.83	26.26	24.01	23.89	21.16	47.52	+**26.69**
EEIL	68.68	53.63	47.91	44.2	36.3	27.46	25.93	24.7	23.95	24.13	22.11	46.57	+**25.74**
UCIR	68.68	57.12	44.21	28.78	26.71	25.66	24.62	21.52	20.12	20.06	19.87	48.81	+**27.98**
TOPIC	68.68	62.79	54.81	49.99	45.25	41.4	38.35	35.36	32.22	28.31	26.28	42.40	+**21.97**
CEC	75.85	71.94	68.50	63.50	62.43	58.27	57.73	55.81	54.83	53.52	52.28	23.57	+**2.74**
S3C (Ours)	**80.62**	**77.55**	**73.19**	**68.54**	**68.05**	**64.33**	**63.58**	**62.07**	**60.61**	**59.79**	**58.95**	**20.83**	

Method	Harmonic Mean (%) ↑									
	1	2	3	4	5	6	7	8	9	10
CEC	57.63	52.83	45.08	45.97	44.44	45.63	45.10	43.76	45.77	44.69
S3C (Ours)	**76.29**	**65.12**	**57.30**	**60.63**	**56.59**	**57.79**	**56.73**	**55.43**	**55.48**	**56.41**

References

1. Tao, X., Hong, X., Chang, X., Dong, S., Wei, X., Gong, Y.: Few-shot class-incremental learning. In: CVPR. pp. 12183–12192 (2020)
2. Zhang, C., Song, N., Lin, G., Zheng, Y., Pan, P., Xu, Y.: Few-shot incremental learning with continually evolved classifiers. In: CVPR. pp. 12455–12464 (2021)
3. Lee, H., Hwang, S.J., Shin, J.: Self-supervised label augmentation via input transformations. In: ICML. pp. 5714–5724. PMLR (2020)

Acknowledgments and Code

This work is partly supported through a research grant from SERB, Department of Science and Technology, Govt. of India and Google Research, India.

GitHub Code

{jayatejak|somabiswas}@iisc.ac.in

code:https://github.com/JAYATEJAK/S3C