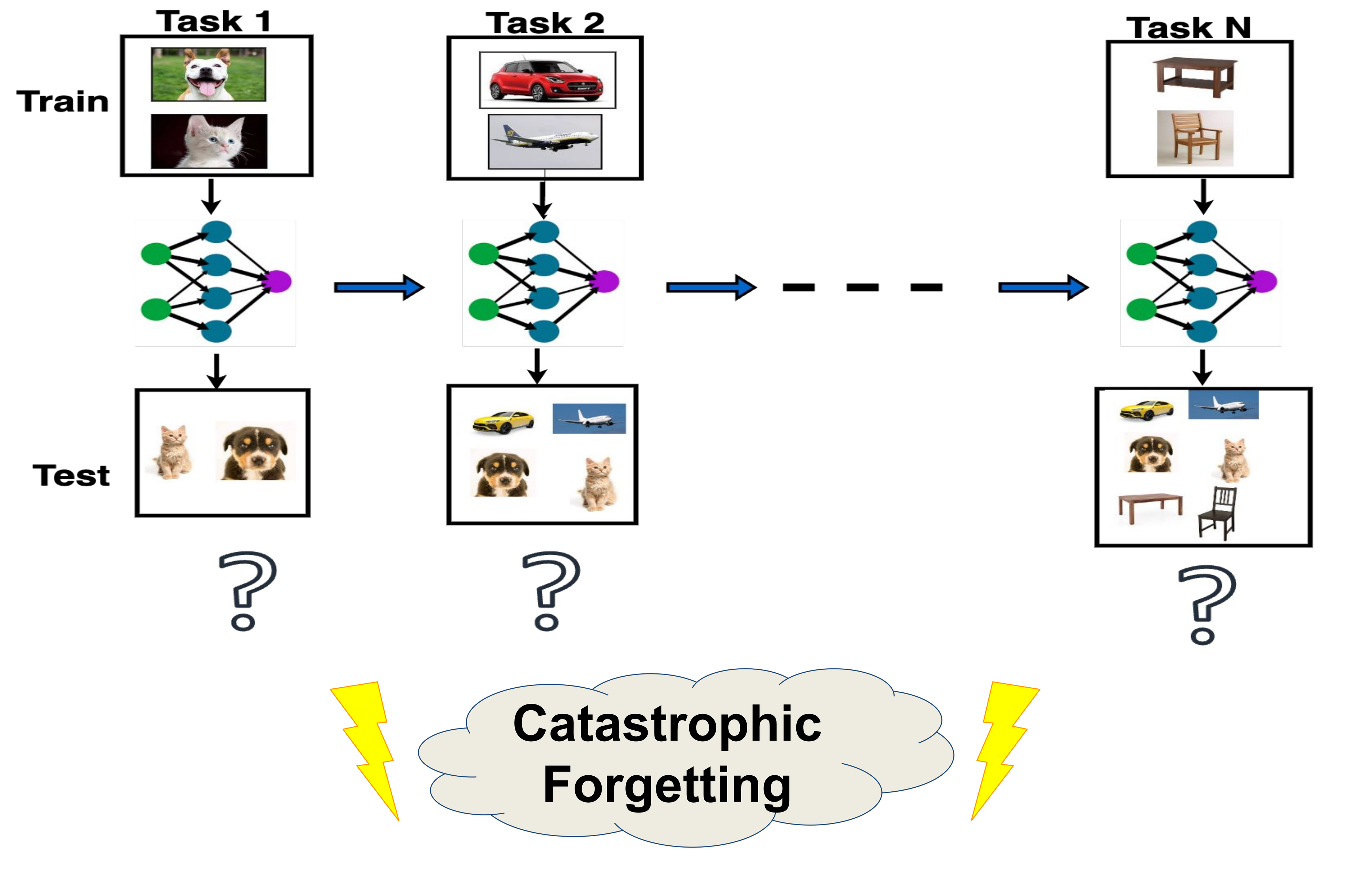
Convolutional Prompting meets Language Models for Continual Learning

Anurag Roy[1], Riddhiman Moulick[1], Vinay Verma[2], Saptarshi Ghosh[1], Abir Das[1]

[1]IIT Kharagpur, [2]IML Amazon India

CVPR SEATTLE, WA JUNE 17-21, 2024

Motivation



- **Class-Incremental Continual Learning (CIL):** Training on new classes of images, while continuing to perform well on all classes encountered till now.
- **Motivation:** Prompt-tuning based approaches have proven to be effective for CIL by using small learnable vectors on top of pre-trained vision transformers to adapt to emerging tasks
- **Challenge: Catastrophic Forgetting** Retraining on new data causes the model to forget previously learned features.

Our Contributions

- **ConvPrompt** for class-incremental continual learning using vision transformers, with *convolution-based dynamically generated learnable vectors* (prompts) being used to overcome catastrophic forgetting
- Dynamically decide the number of prompts to be added for each task, using *text-based attribute similarity* where the attributed for the classes in each task are generated using LLMs

Method

Training:

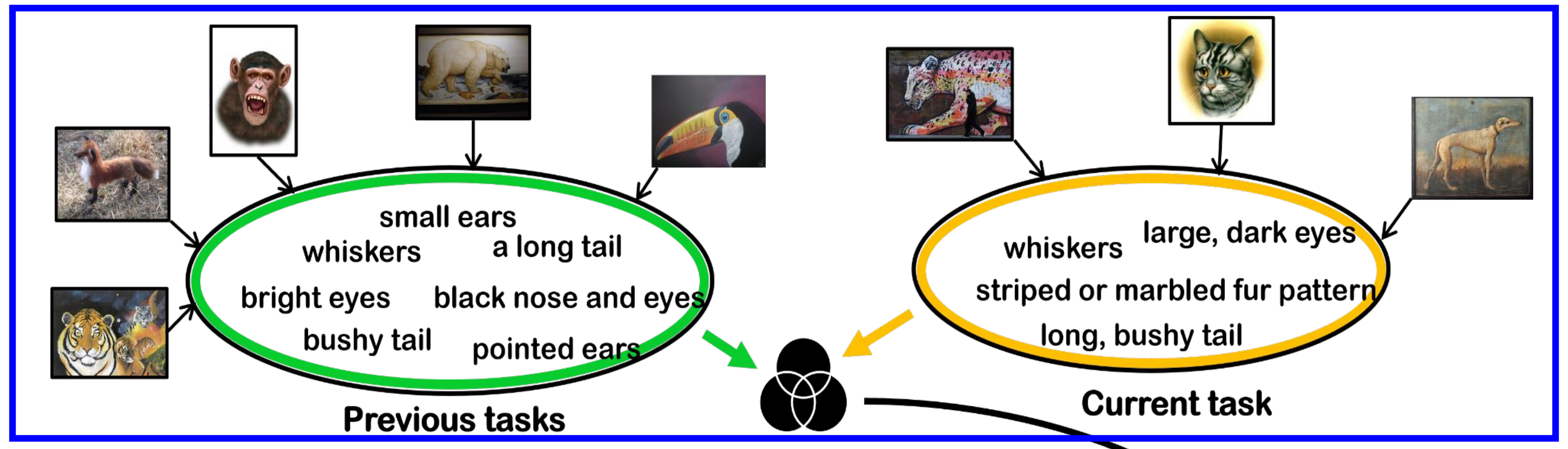
- For every task, a set of convolutional kernels (prompt generators) are learnt which act on the [*cls*] token to produce the necessary prompt components for the image

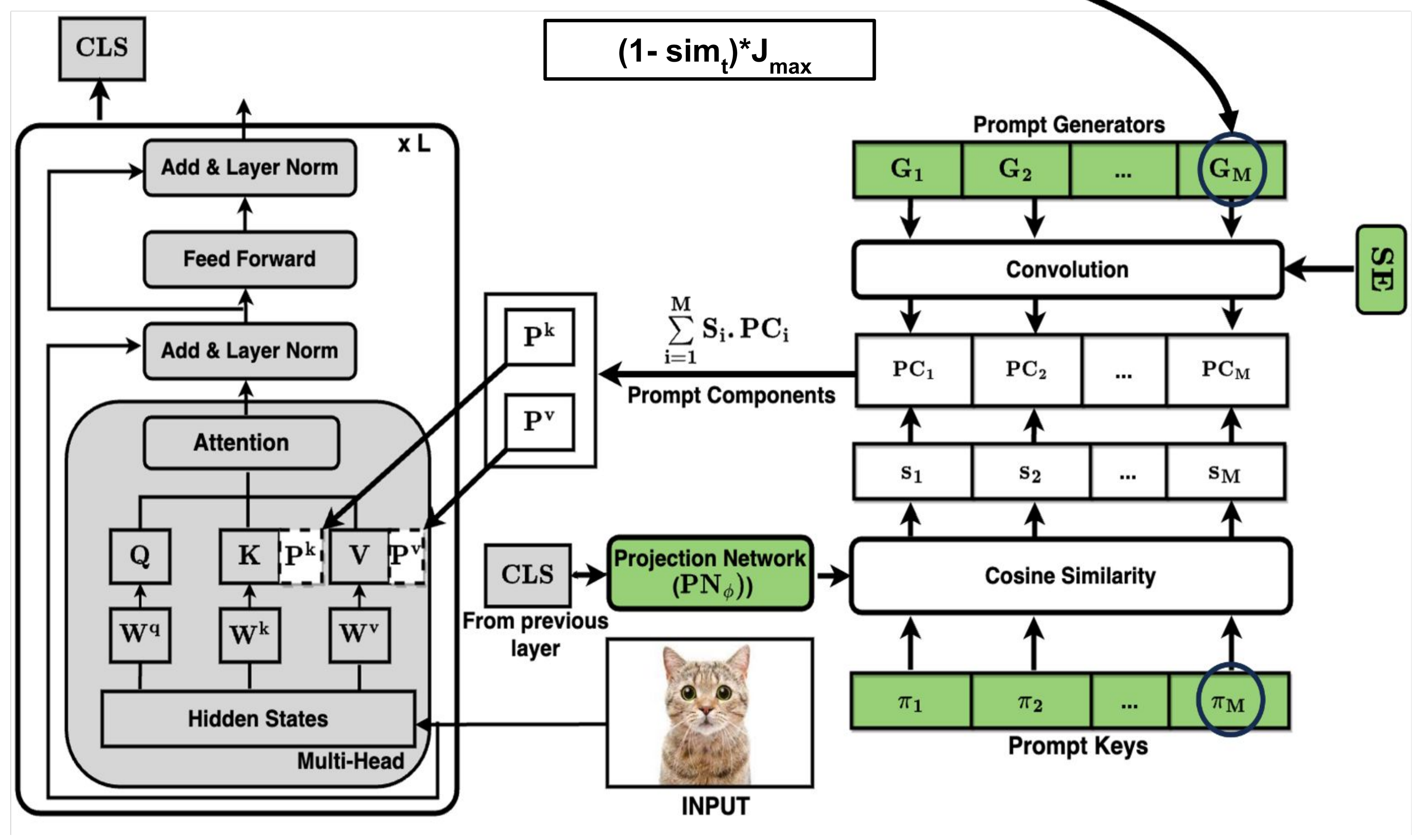
- A weighted combination of the prompt components generated using Convolution over the shared embedding gives the final prompt

-The number of prompt generators required is decided beforehand based on attribute similarity between the current and previous tasks

Inference:

- The entire pool of learned prompt generators is used to generate the prompt components and a weighted-combination of these gives the final-prompt during test-time

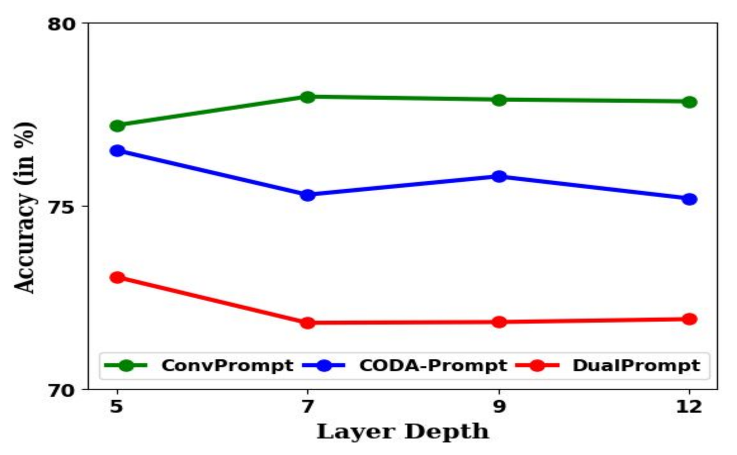


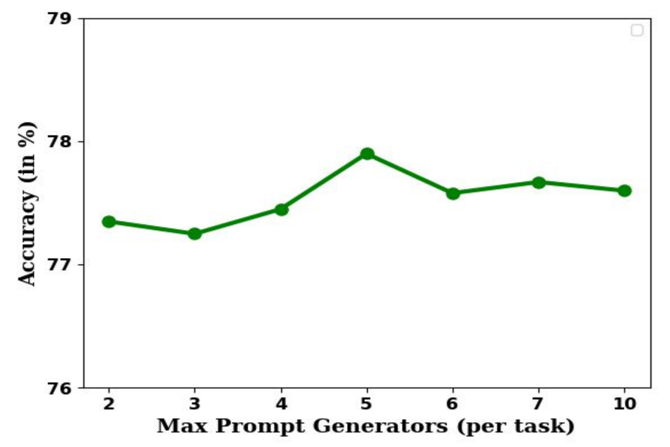


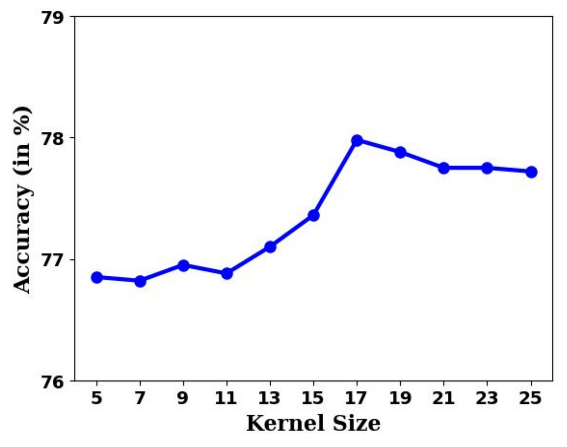
Experiments

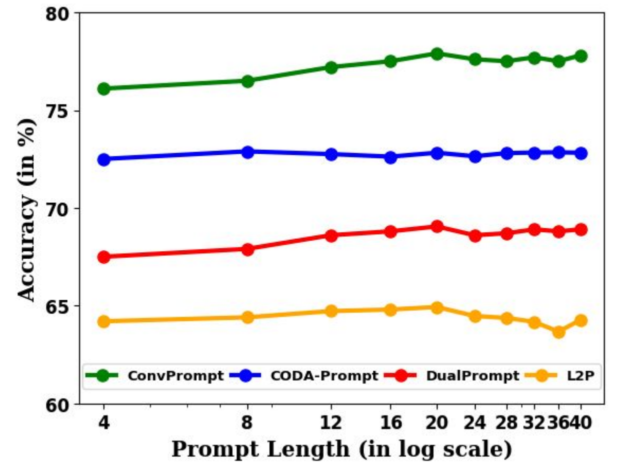
Tasks	Split CIFAR-100		Split CUB-200		$N_{param}(\downarrow)$
Method	$A_T(\uparrow)$	$F_T(\downarrow)$	$A_T(\uparrow)$	$F_T(\downarrow)$	Train/Total
Joint-FT (upper bound)	93.22 ± 0.16	–	88.00 ± 0.15	–	100/ 100
Seq-FT	8.6 ± 0.43	42.67 ± 0.13	23.87 ± 0.54	62.52 ± 0.57	100/ 100
ER (buffer size 5000)	82.30 ± 0.42	16.30 ± 0.24	60.73 ± 0.23	8.71 ± 0.65	100/ 100
LwF [27]	64.56 ± 1.23	25.27 ± 1.32	48.73 ± 1.46	25.18 ± 0.31	100/ 100
L2P [54]	82.76 ± 1.17	7.86 ± 0.39	62.21 ± 1.92	7.12 ± 0.33	**0.7/ 100.7**
L2P + LGCL [23]	84.33 ± 0.06	5.83 ± 0.23	–	–	**0.7/ 100.7**
DualPrompt [53]	85.07 ± 0.49	5.57 ± 0.20	66.00 ± 0.57	$\mathbf{4.4 \pm 0.31}$	1.3/ 101.3
DualPrompt + LGCL [23]	87.23 ± 0.21	5.10 ± 0.15	–	–	1.3/ 101.3
CODA-Prompt [46]	87.00 ± 0.38	4.78 ± 0.24	74.40 ± 0.74	6.40 ± 0.34	4.6/ 104.6
ConvPrompt	$\mathbf{88.87 \pm 0.33}$	$\mathbf{4.75 \pm 0.15}$	$\mathbf{80.2 \pm 0.52}$	5.6 ± 0.38	2.0/ 102.0

Method	MACs
L2P [54]	35.85B
DualPrompt [53]	33.72B
CODA-Prompt [46]	33.72B
ConvPrompt	17.98B









Acknowledgement

We would like to thank Google Research India and ACM India for travel support.

Project Page