Random Path Selection for Incremental Learning

Jathushan Rajasegaran, Munawar Hayat, Salman Khan, Fahad Shahbaz Khan, Ling Shao
Inception Institute of Artificial Intelligence

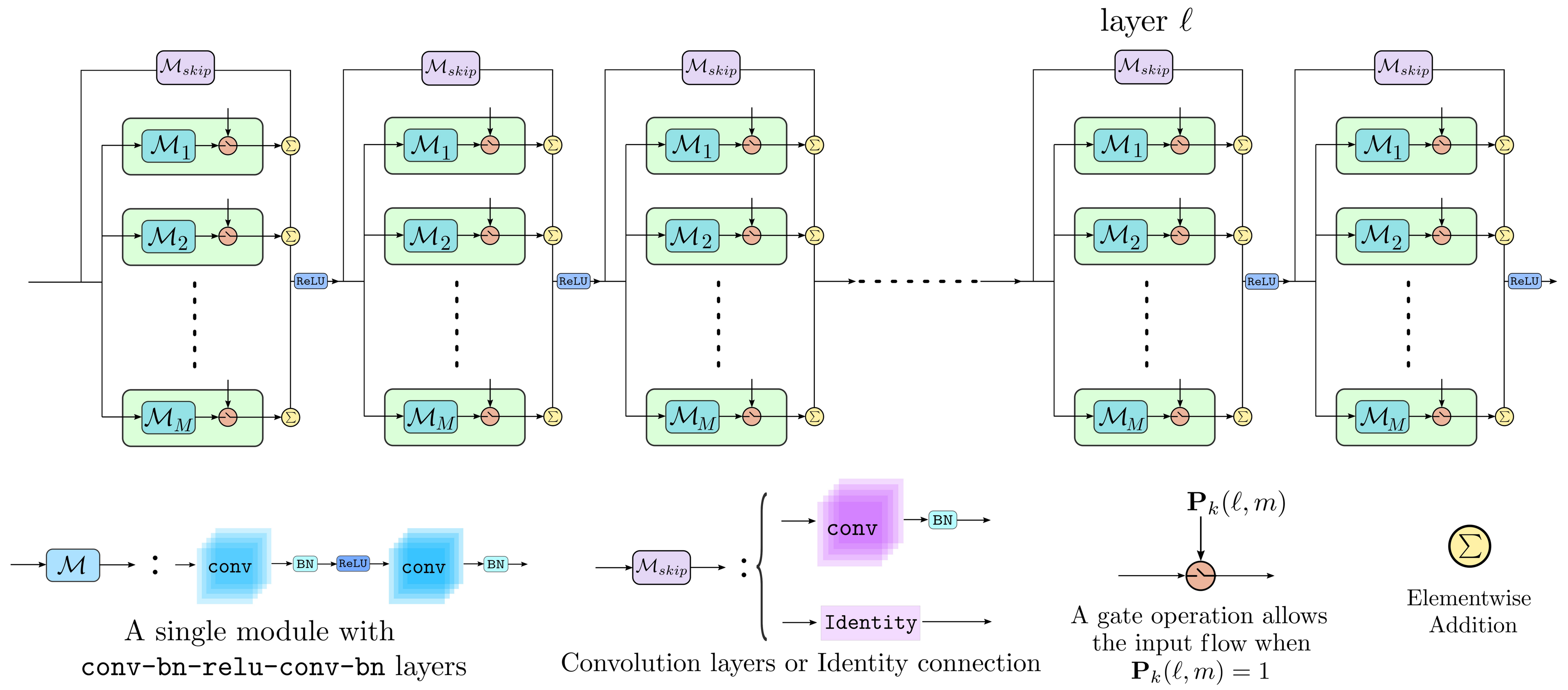
Introduction

- Incremental life-long learning is a main challenge towards the long-standing goal of Artificial General Intelligence.
- Deep neural networks suffer from '*catastrophic forgetting*', when a network is sequentially trained on a series of tasks.
- Existing incremental learning approaches, fall well below the state-of-the-art cumulative models that use all training classes at once.

Problem Setup

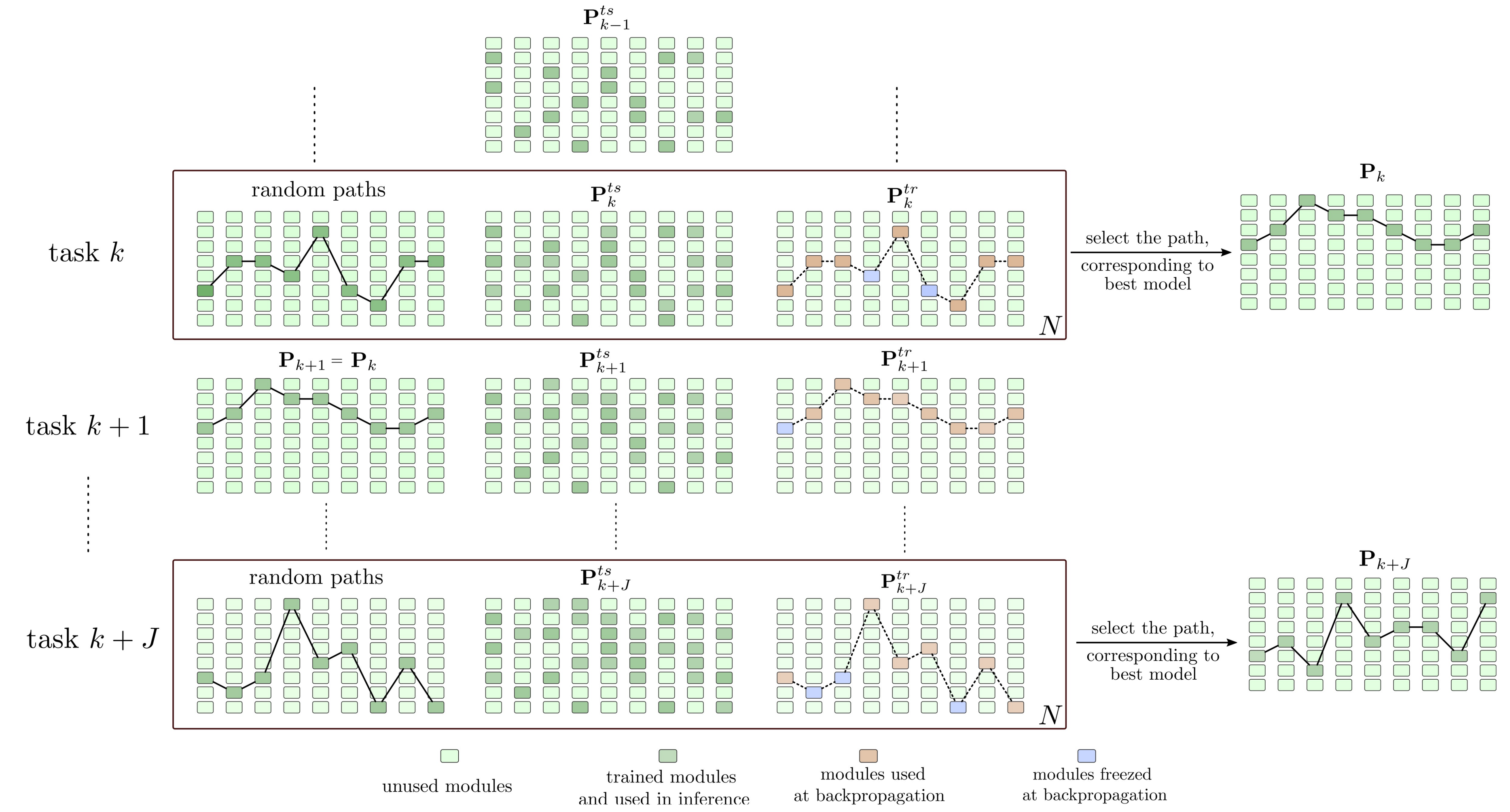
- In real-life settings, learning tasks arrive in a sequence and machine learning models must continually learn to increment already acquired knowledge.
- We consider the recognition problem in an incremental setting where new tasks are sequentially added. Assuming a total of K tasks, each comprising of U classes.
- we restrict exemplar memory budget to random 2000 samples for CIFAR-100 and ImageNet datasets.

RPS Net Architecture



- RPS-Net architecture consists of M residual blocks per layer, with a ever learning skip connection.
- At end of each layer in RPS-Net, all the residual connections and skip connections are combined together using element-wise addition..
- Any single path of RPS-Net can be treated as a single ResNet.

Training and Inference



- For every J tasks, we sample N random paths and train the network separately in parallel.
- The best path is used for next J upcoming tasks.
- When finding the paths at every J tasks, previously trained modules are freezed at training.
- At forward pass, all the previously trained paths are cumulatively used to calculated the loss.
- At backward pass, only the trainable modules are considered, thus its upper bounded by depth of the network.
- Since the number modules are limited, there will be high overlaps down the line, forcing **module reusability** and requires only a small **residual signals** to be learned.

Loss Function

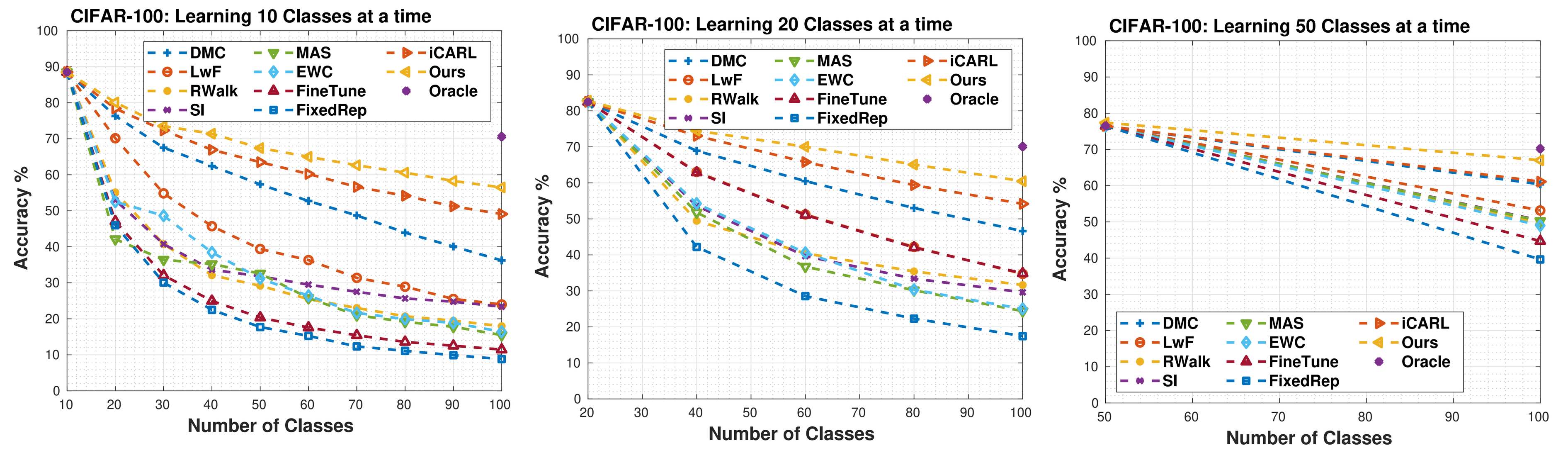
$$\mathcal{L}_{ce} = -\frac{1}{n}\sum_i \mathbf{t}_i[1:k*U]\log(\text{softmax}(\mathbf{q}_i[1:k*U]))$$

$$\mathcal{L}_{dist} = \frac{1}{n}\sum_i \text{KL}\left(\log\left(\sigma\left(\frac{\mathbf{q}_i[1:(k-1)*U]}{t_e}\right)\right), \sigma\left(\frac{\mathbf{q}'_i[1:(k-1)*U]}{t_e}\right)\right)$$

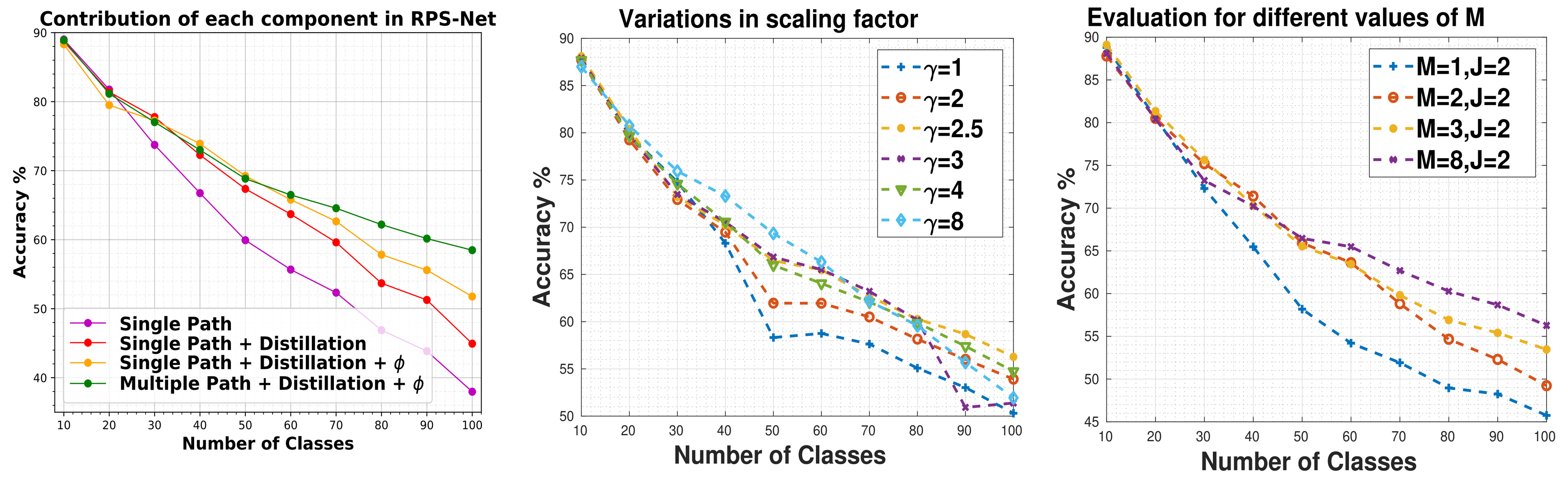
$$\mathcal{L} = \mathcal{L}_{ce} + \phi(k,\gamma)\cdot\mathcal{L}_{dist} \quad \text{Where, } \phi(k,\gamma) = \begin{cases} 1, & \text{if } k \le J \\ (k-J)*\gamma, & \text{otherwise.} \end{cases}$$

Experiments and Results

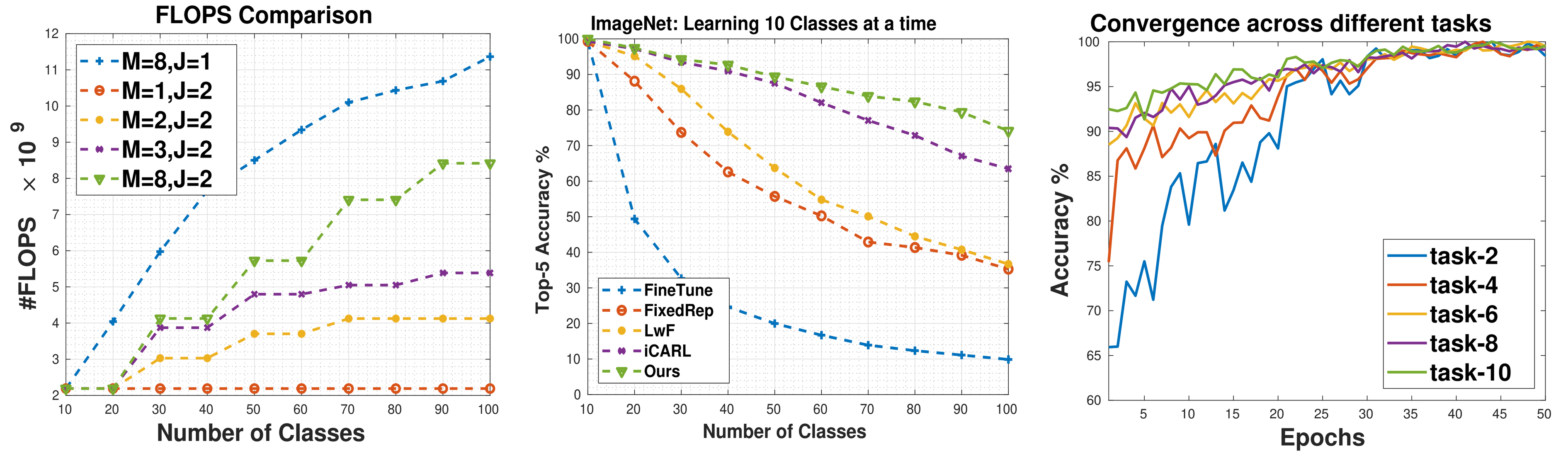
- We use iCIFAR-100 with 10,5 and 2 task. For ImageNet we use first 100 classes with 10 tasks. In all setting we surpass STOA results.



- We study the contributions from each of our contributions extensively.



- RPS-Net complexity increases logarithmically and with forward transfer new tasks converges quickly.



Conclusion

- In this work, we propose a novel network architecture, random path selection strategy and controlled loss function for incremental classifier learning.