# Helpful or Harmful: Inter-Task Association in Continual Learning

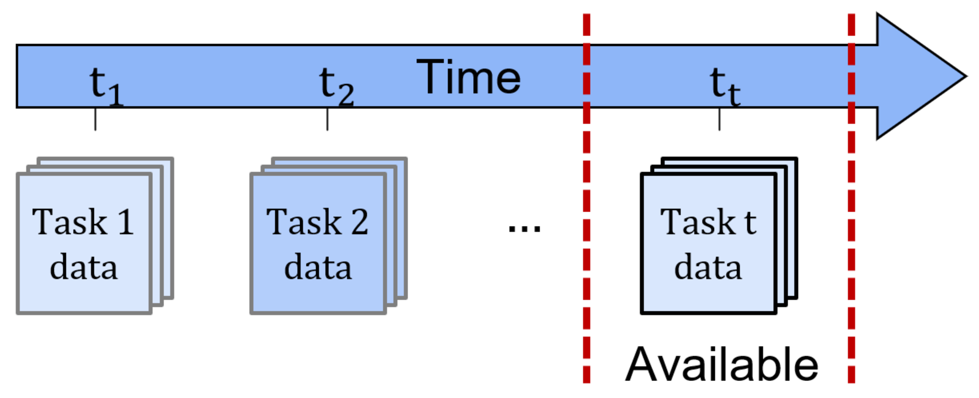
Hyundong Jin and Eunwoo Kim

School of Computer Science and Engineering, Chung-Ang University, South Korea

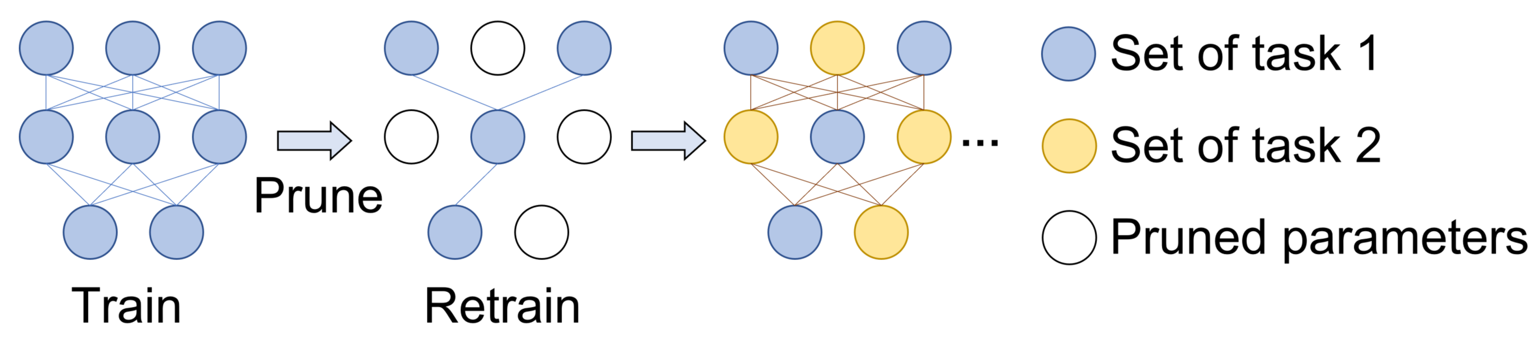
ECCV TEL AVIV 2022

## INTRO & MOTIVATION

- Learning sequentially incoming tasks
- Limit accessibility to the old data
- Forget previously learned tasks when training on new tasks




- Baseline: structural allocation approach [1]




- **Problem: Associate all parameters when inferencing**
- **Goal: Differentiate** helpful and harmful information
- **Approach**: Reformulate the **forgetting problem** into a **task interference problem**

## H²: INTER-TASK ASSOCIATION

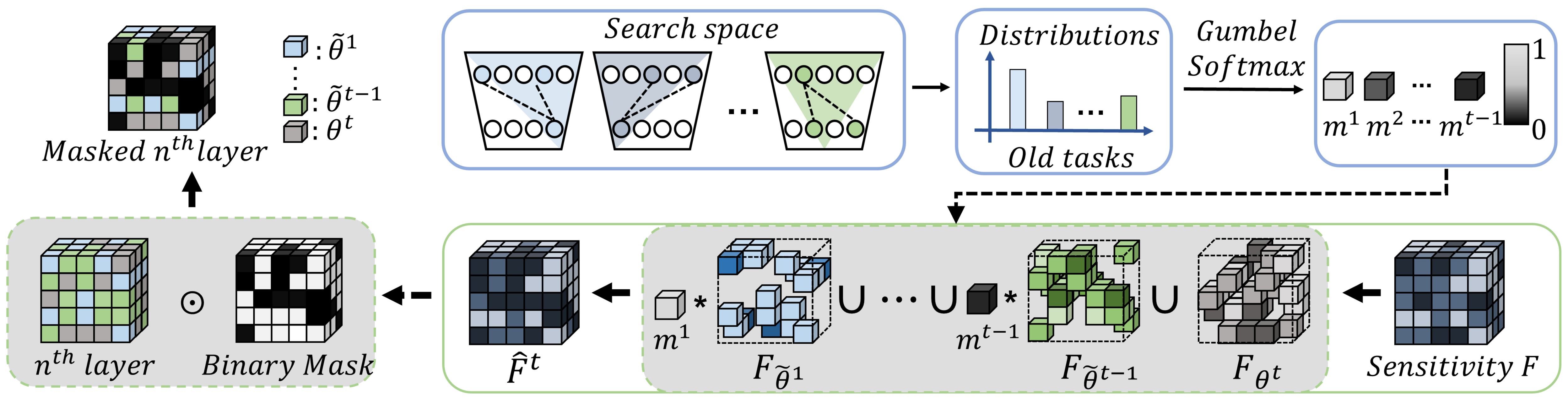
Selects contributable parameters from older tasks in a coarse-to-fine manner

- **Search (Coarse level)**
  - Find an optimal network structure $a^t$
  - Exploit cooperative disjoint set discovered by gradient based search
  - Update the set of parameters $\theta^t$

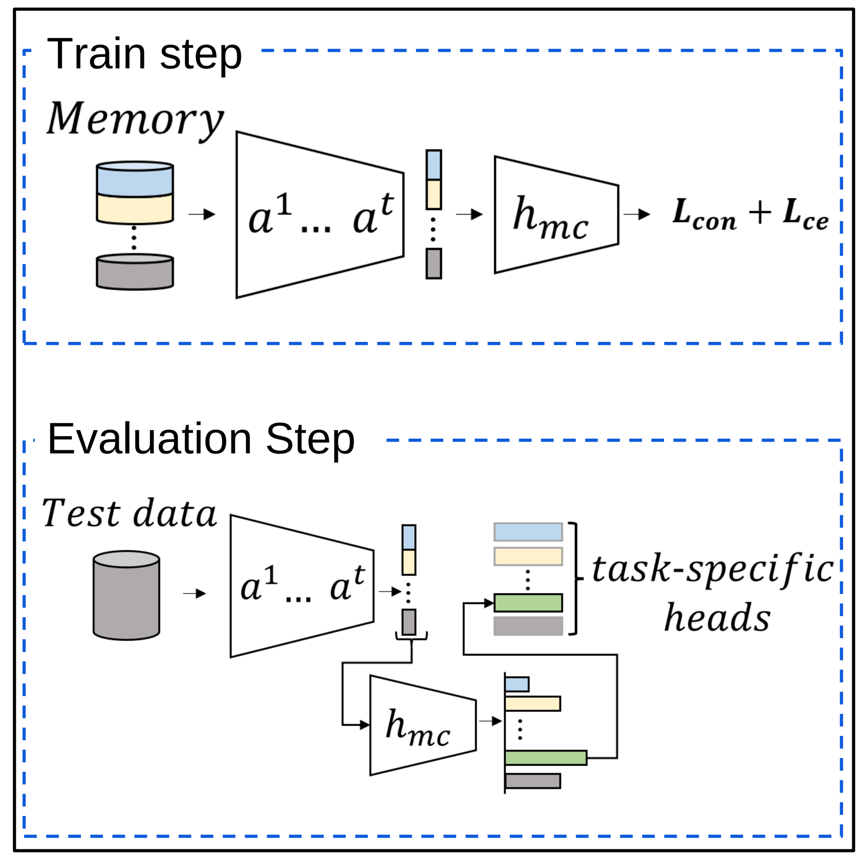
$$\min_{\alpha} \min_{\theta^t} \mathbf{E}_{a^t \sim P_\alpha} [\ \mathcal{L}(a^t, \theta^t)\ ] \qquad a^t = \bigcup_{i=1}^{t-1} \left( m^i \cdot \tilde{\theta}^i \right) \cup \theta^t \qquad m^i = \mathrm{GumbelSoftmax}(\alpha^i | \alpha)$$

- **Sensitivity measure (Finer level)**
  - Further explore useful information from selected old tasks
  - Exploiting both sampled variable and sensitivity

$$F^t(\theta) = \bigcup_{i=1}^{t-1} F_{\tilde{\theta}^i} \cup F_{\theta^t} \qquad a^t = \bigcup_{i=1}^{t-1} \left( \sigma_\phi(m^i \cdot F_{\tilde{\theta}^i}) \odot \tilde{\theta}^i \right) \cup \theta^t$$



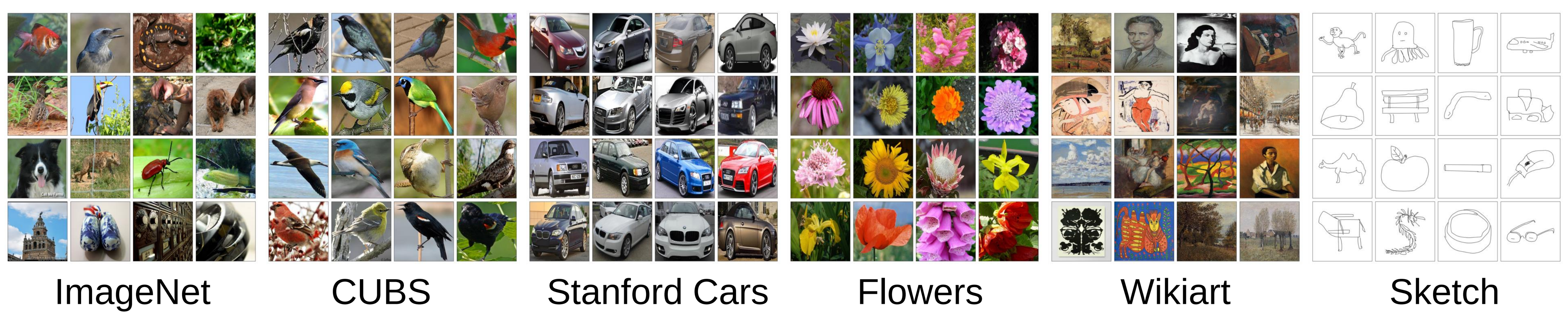

- **Meta Classifier**
  - Require when **task identity is unknown**
  - Predict the task label and allocate the corresponding network
  - Employ contrastive loss for discrimination between tasks



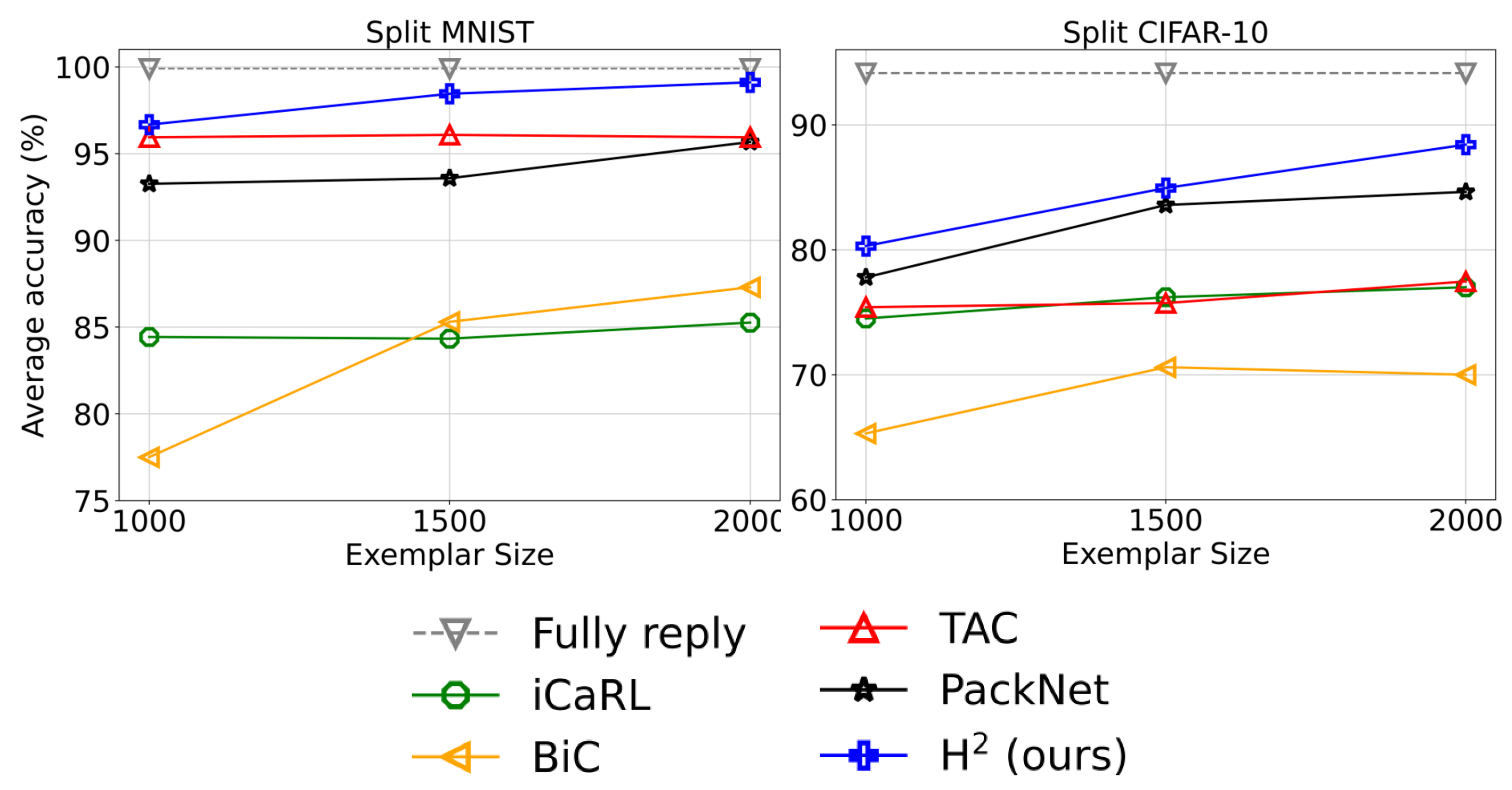

## RESULTS: TASK- AND CLASS- INCREMENTALLEARNING SCEARIOS

- Task-incremental: ImageNet and Find-grained datasets

| Method | ImageNet Acc | ImageNet Param | CUBS Acc | CUBS Param | Stanford Cars Acc | Stanford Cars Param | Flowers Acc | Flowers Param | Wikiart Acc | Wikiart Param | Sketch Acc | Sketch Param | Avg Acc | Avg Param |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EWC-On | 54.5 | 23.4 | 64.1 | 23.4 | 52.1 | 23.4 | 78.7 | 23.4 | 51.4 | 23.4 | 30.7 | 23.4 | 55.3 | 23.4 |
| LwF | 64.5 | 23.4 | 51.7 | 23.4 | 43.9 | 23.4 | 72.9 | 23.4 | 42.7 | 23.4 | 45.5 | 23.4 | 53.6 | 23.4 |
| PackNet | 75.7 | **12.5** | 80.4 | 15.6 | 86.1 | 18.0 | 93.0 | 19.7 | 69.4 | 21.0 | 76.7 | 22.0 | 80.2 | 18.1 |
| Piggyback | **76.1** | 23.4 | 81.5 | 20.5 | 89.6 | 19.7 | 94.7 | 22.3 | 71.3 | 16.3 | **79.9** | 18.0 | 82.2 | 20.0 |
| H² (Ours) | 75.7 | **12.5** | **84.1** | **14.5** | **90.6** | **15.9** | **94.9** | **15.2** | **75.1** | **9.4** | 76.2 | **4.1** | **82.8** | **11.9** |




- Class-incremental: Split MNIST and Split CIFAR-10

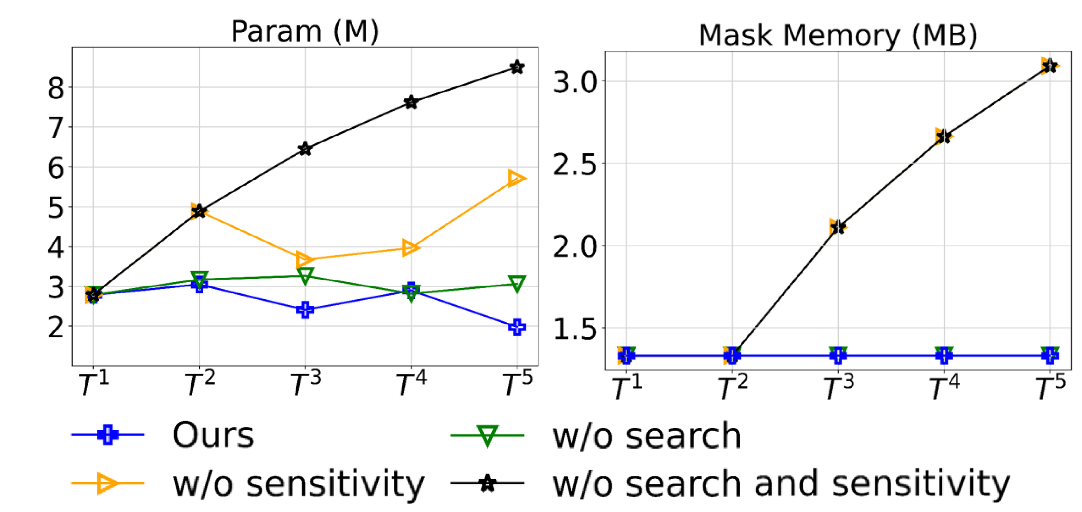



[1] Mallya, A., et al., “Packnet: Adding multiple tasks to a single network by iterative pruning.”, CVPR (2018).

## RESULTS: ABLATIONS

- ResNet-18 (Split CIFAR-10)

| Method | Class-incremental $T^1$ | $T^2$ | $T^3$ | $T^4$ | $T^5$ |
|---|---|---|---|---|---|
| Ours | **98.8** | **91.2** | **87.4** | **86.1** | **80.3** |
| w/o search | **98.8** | 89.5 | 84.9 | 84.5 | 79.6 |
| w/o sensitivity | **98.8** | 90.0 | 83.8 | 82.6 | 78.7 |
| w/o search and sensitivity [27] | **98.8** | 89.3 | 83.5 | 81.1 | 77.4 |
| w/o contrastive loss | **98.8** | 90.7 | 86.7 | 83.2 | 77.1 |
| Random search | **98.8** | 86.3 | 80.8 | 77.9 | 68.7 |




Codes are available at: https://github.com/Jin0316/Helpful-or-Harmful-Inter-Task-Association

Please send any comments or questions to jude0316@cau.ac.kr