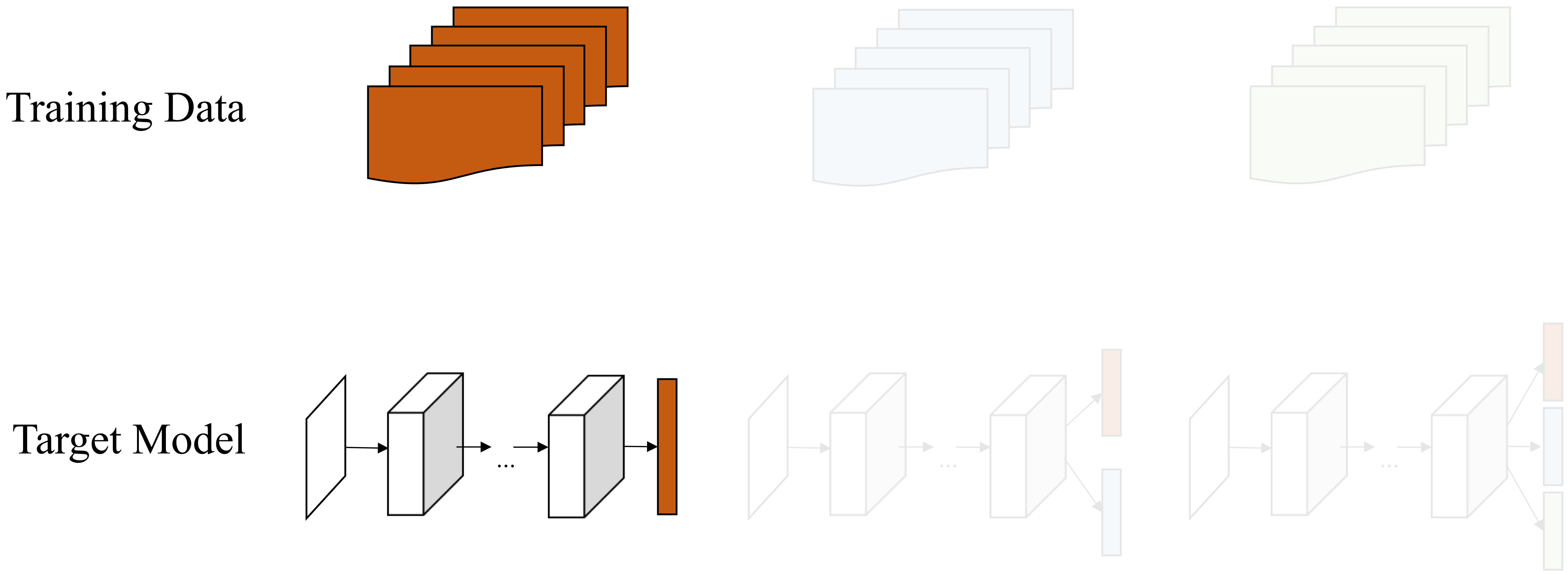
Incremental Classification: First Step into Lifelong Learning

PAN Xinyu
MMLab, Department of IE

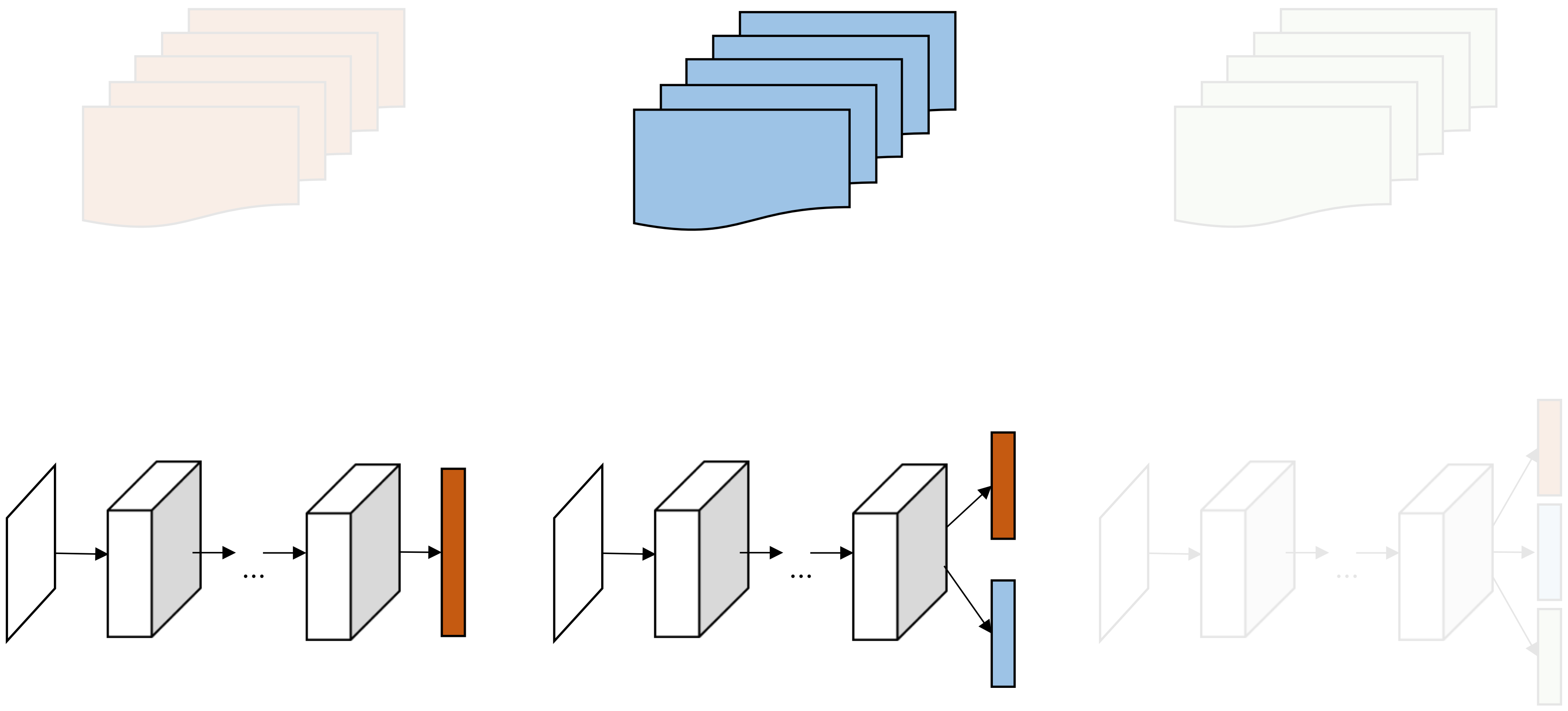
Multi-task Incremental Classification: Setup



Multi-task Incremental Classification: Setup

Training Data

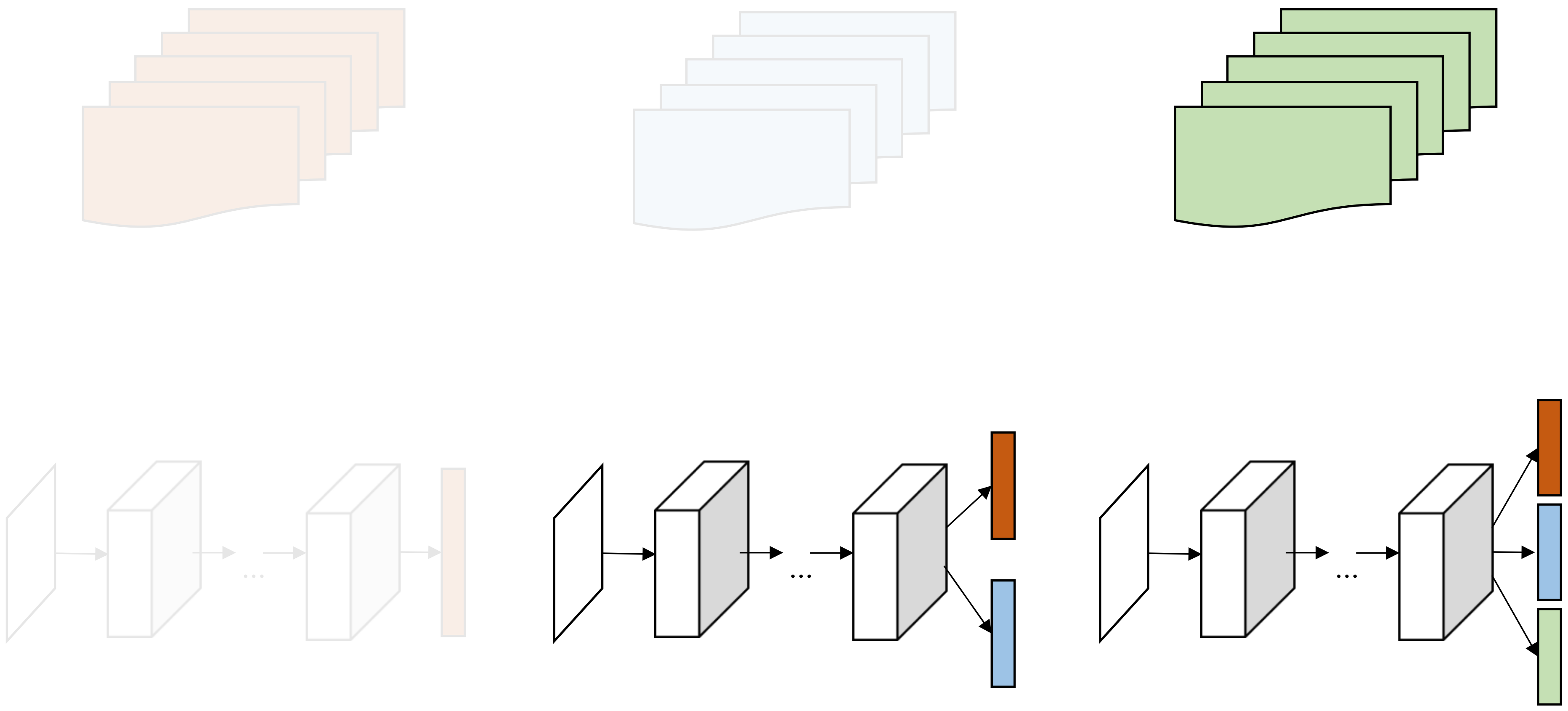
Target Model



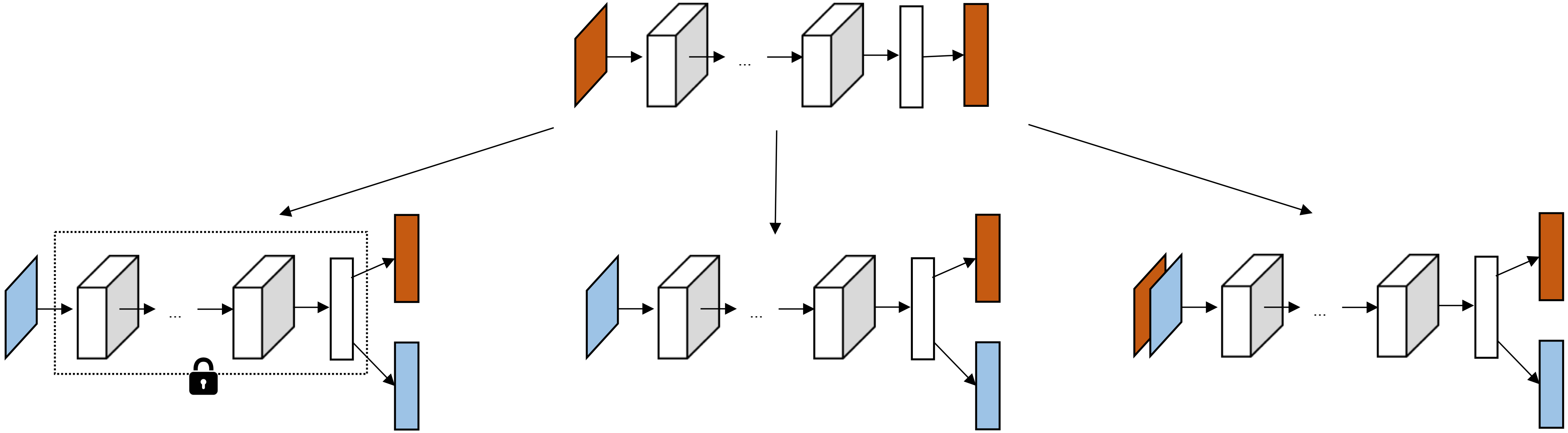
Multi-task Incremental Classification: Setup

Training Data

Target Model



Multi-task Incremental Classification: Baseline



Feature Extraction	*Finetuning*	*Re-training*
Sub-optimal for the new task	Catastrophic forgetting	Time consuming

Potential Application Scenarios

- Limited storage budget that can not keep all sequential data.
- The collected data will expire due to privacy issues.
- Efficient deployment of the model for incremental data.
- …

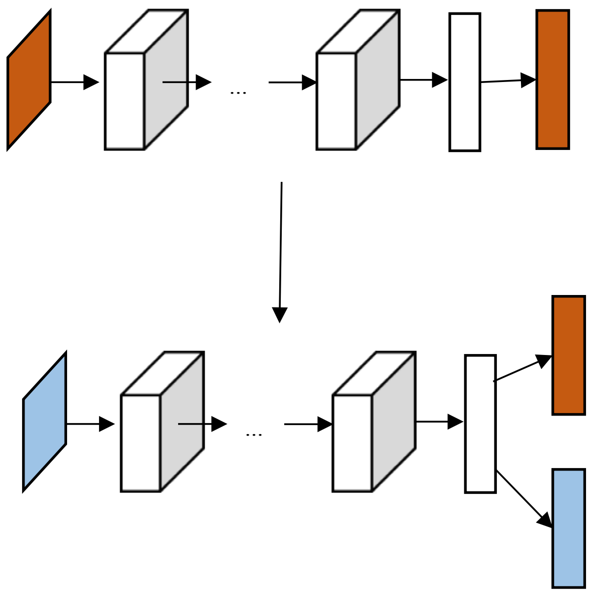
Lifelong Learning via Progressive Distillation and Retrospection

Saihui Hou[1*] Xinyu Pan[2*] Chen Change Loy[3] Zilei Wang[1] Dahua Lin[2]

[1] University of Science and Technology of China [2] The Chinese University of Hong Kong
[3] Nanyang Technological University [* indicates joint first authorship]

(Accepted in ECCV 2018)

Handle Catastrophic Forgetting

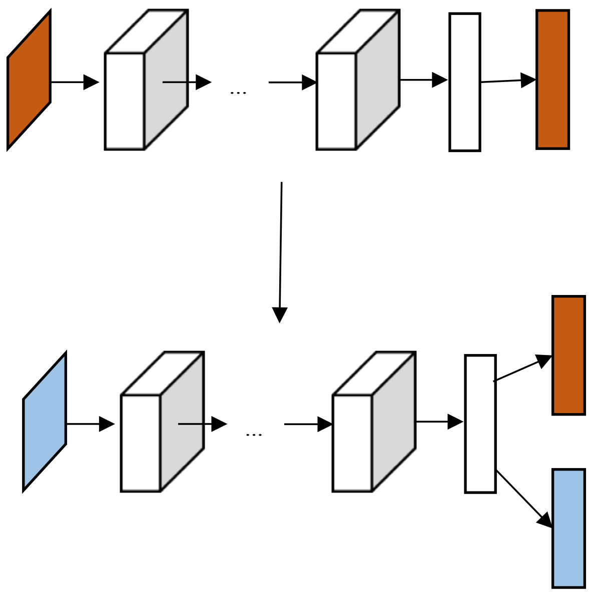


Finetuning

Catastrophic forgetting

How to prevent performance drop in the old task?

Handle Catastrophic Forgetting



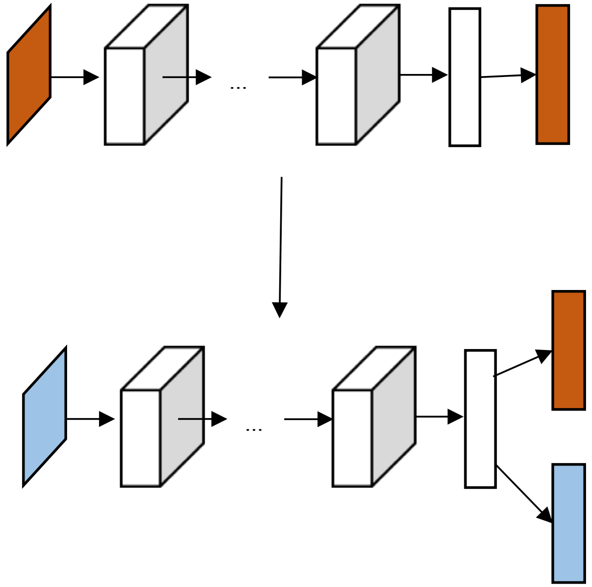
Finetuning

Catastrophic forgetting

How to prevent performance drop in the old task during training?

We need an *indicator*.

Handle Catastrophic Forgetting



Finetuning

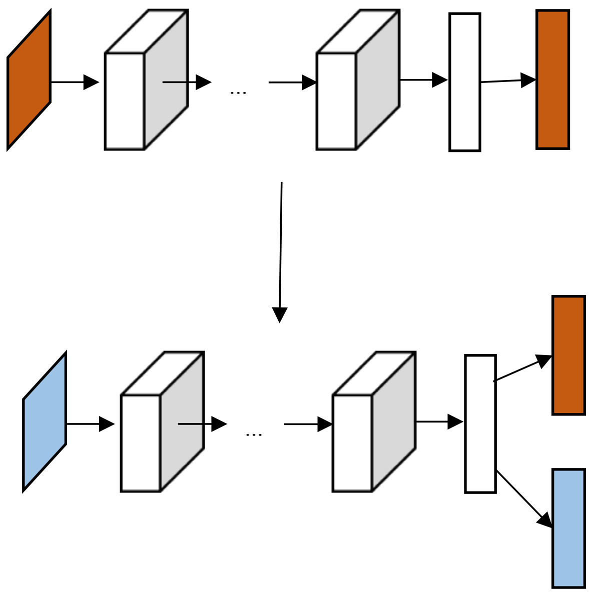
Catastrophic forgetting

How to prevent performance drop in the old task during training?

We need an *indicator*.

How to construct an indicator if we do not reserve any of old data?

Handle Catastrophic Forgetting



Finetuning

Catastrophic forgetting

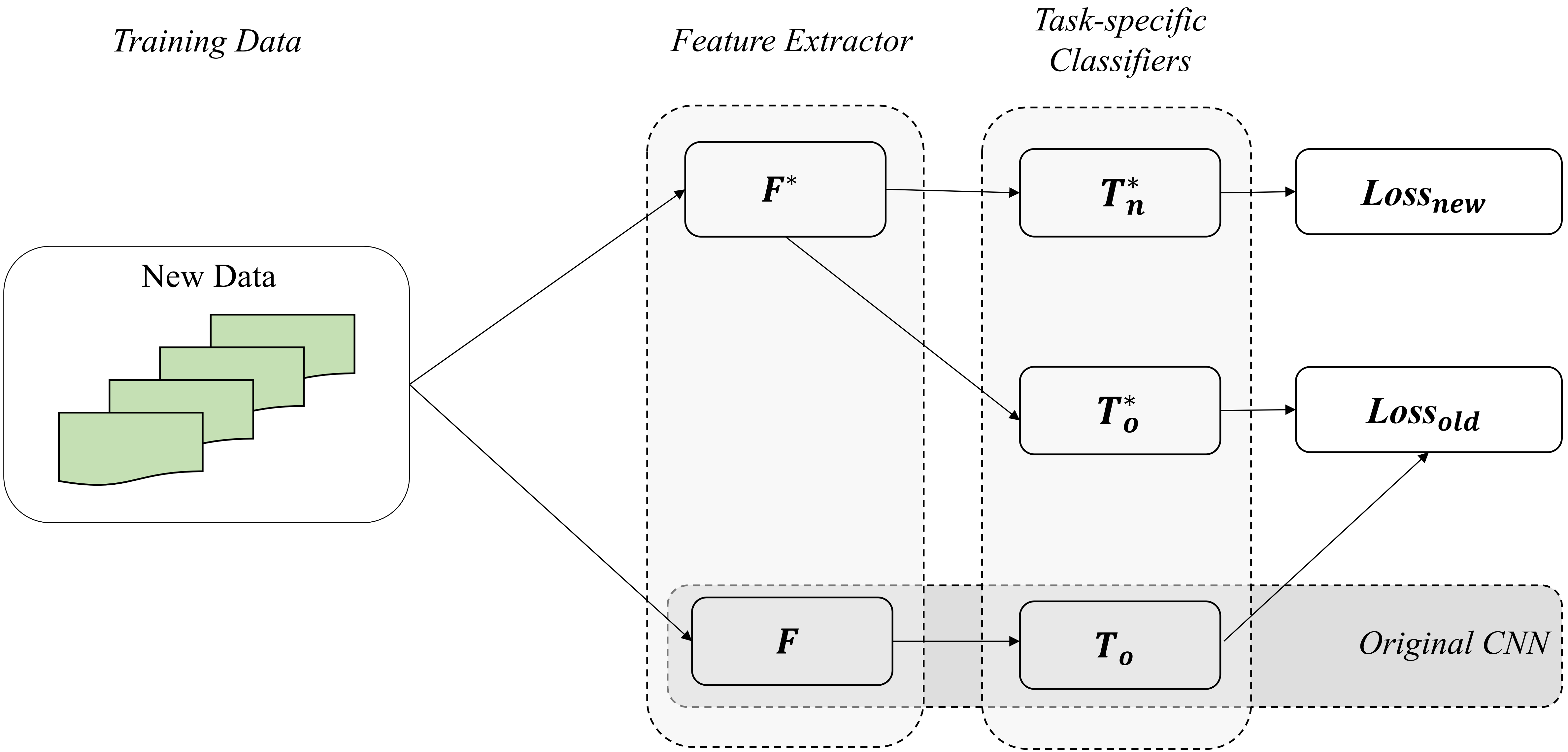
How to prevent performance drop in the old task during training?

We need an *indicator*.

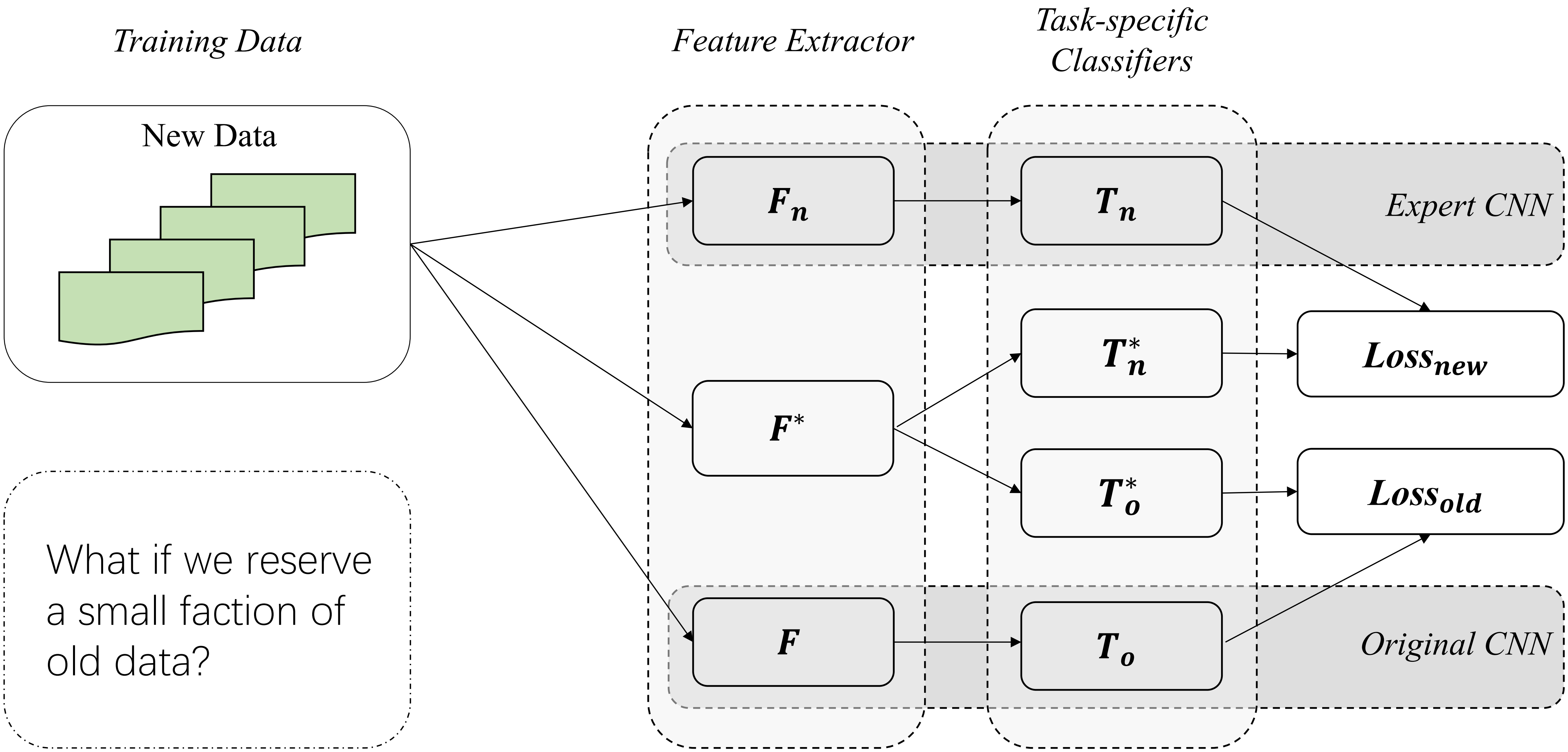
How to construct an indicator if we do not reserve any of old data?

Take new data as *fake* old data.

Learning without Forgetting (Accepted in ECCV 2016)

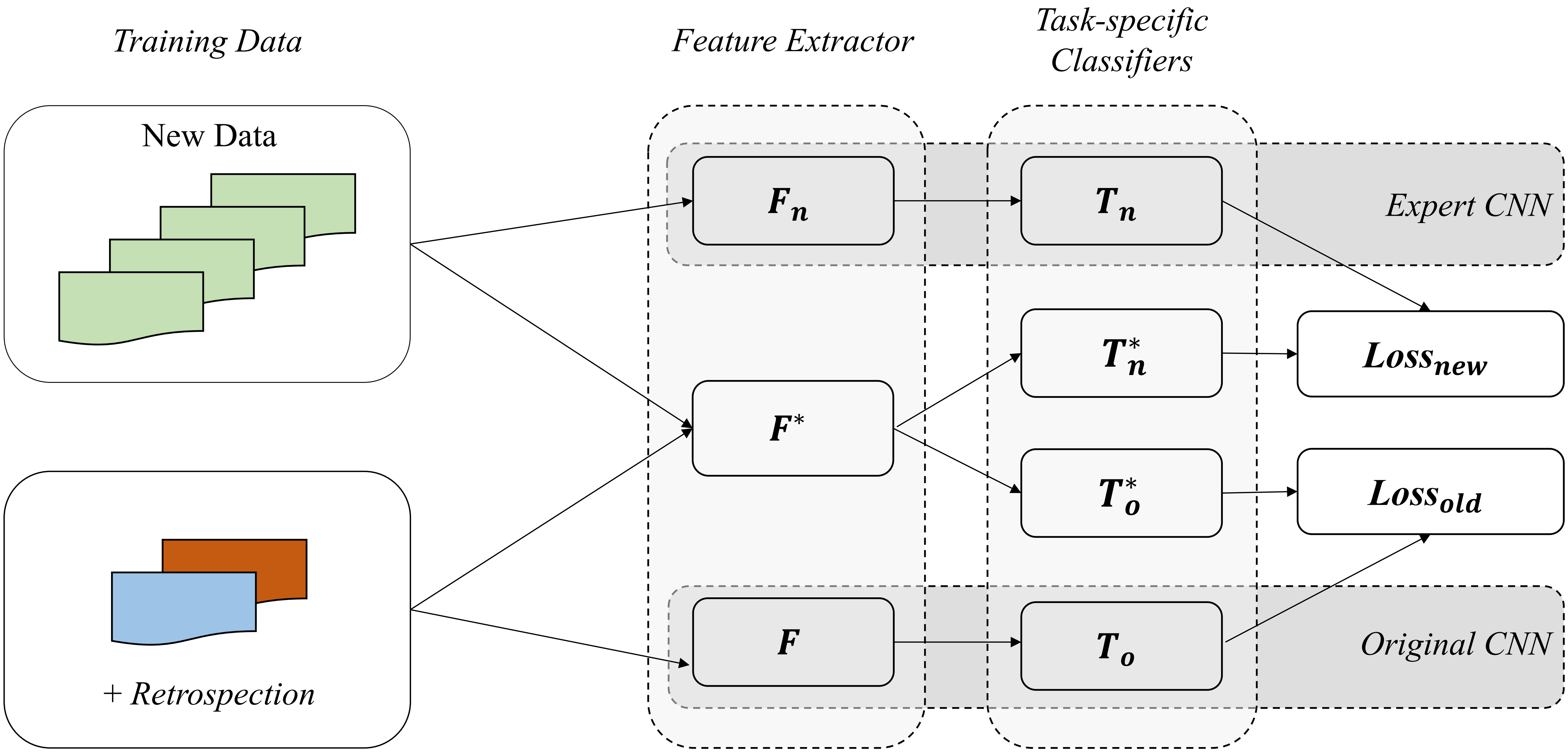


Adaptation by Distillation

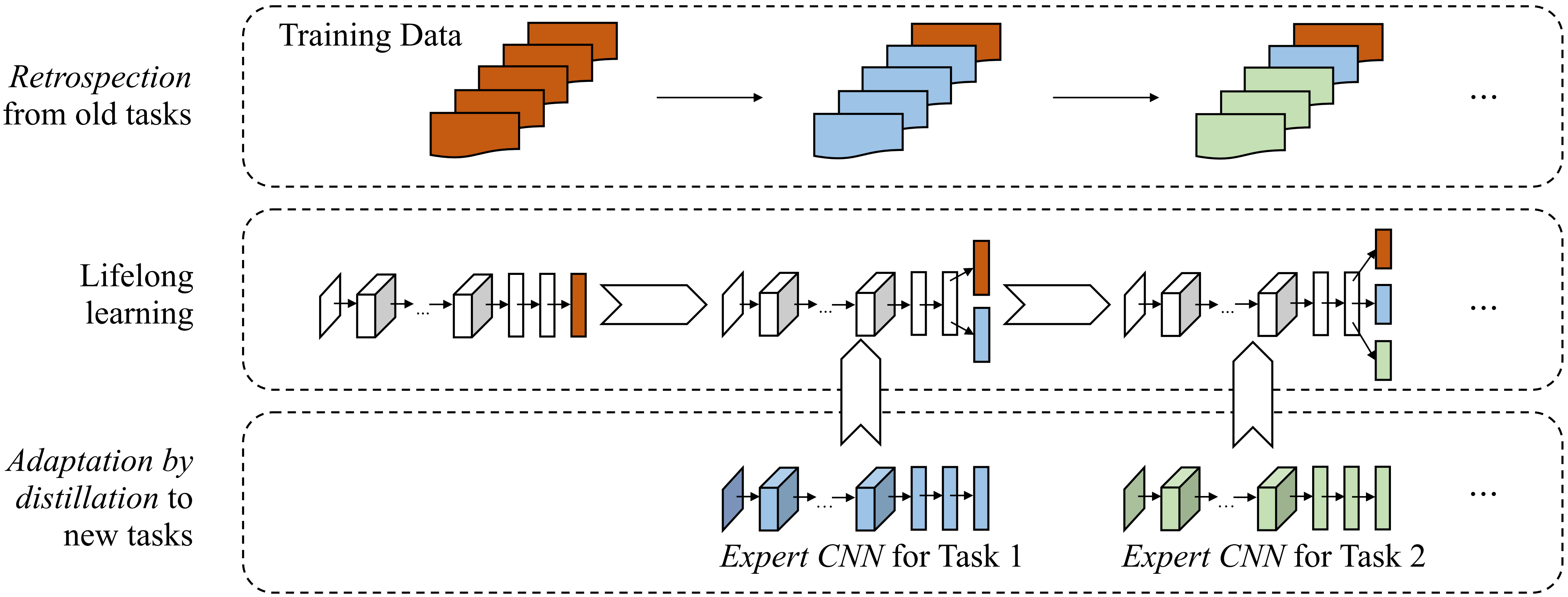


What if we reserve a small faction of old data?

Adaptation by Distillation + Retrospection



Overview of Distillation and Retrospection

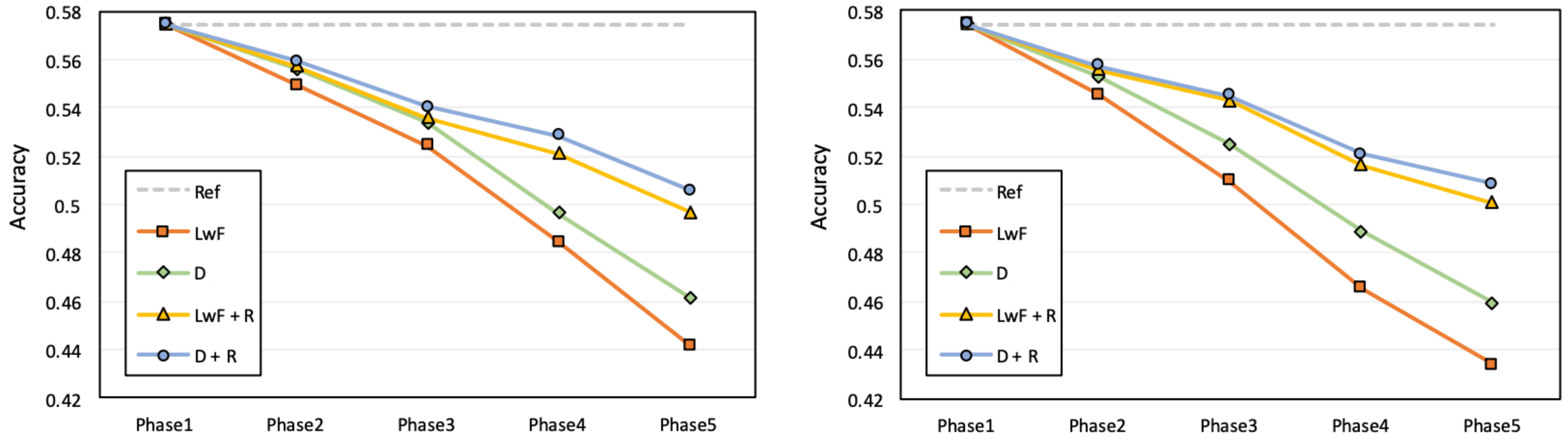


Dataset

Table 1. The statistics of the datasets used in this work.

Task	Datasets	#Category	#Training	#Test
ImageNet	ILSVRC-2012 [19]	1000	1,281,167	50,000
Birds	CUB-200-2011 [26]	200	5994	5794
Flowers	Oxford Flowers [18]	102	2040	6149
Scenes	MIT Scenes [19]	67	5360	1340
Aircrafts	FGVC-Aircrafts [16]	100	6667	3333

Some Results



(a) Imagenet→Scenes→Birds→Flowers→Aircrafts. (b) Imagenet→Birds→Flowers→Aircrafts→Scenes.

Fig. 3. Accuracy degradation on ImageNet in five-task scenario. D for *Distillation*, and R for *Retrospection*.

Ablation Study on #Reserved Samples

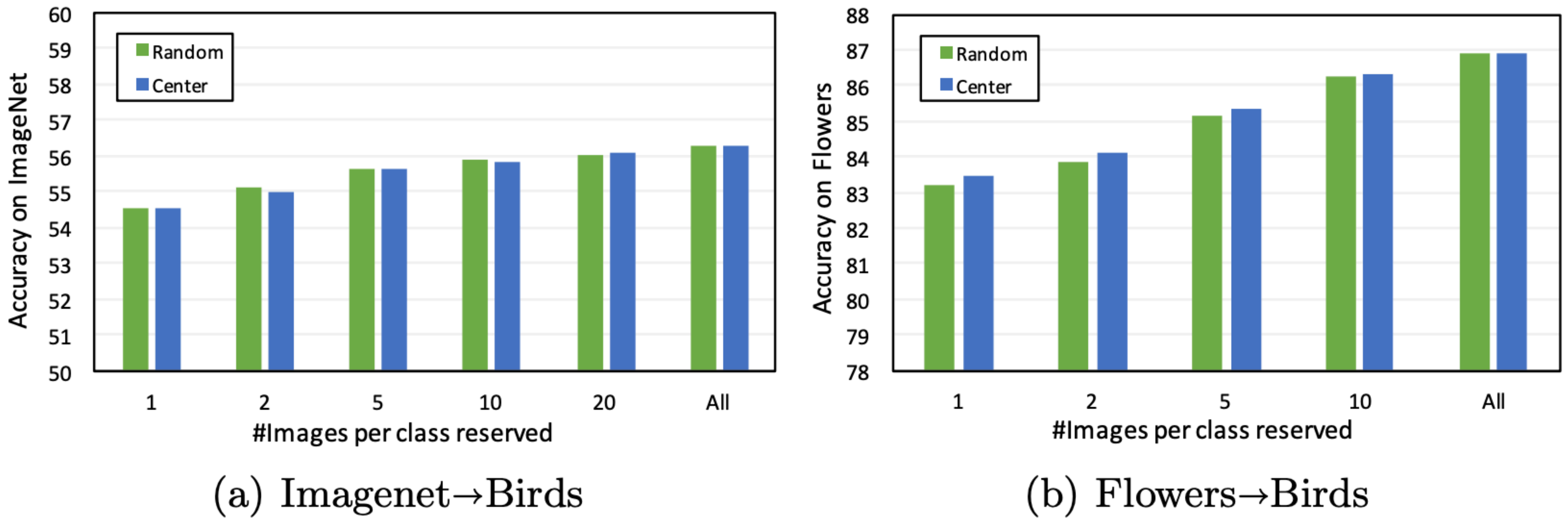


Fig. 4. Ablation study on *Retrospection* strategy. **Random** for random selection, and **Center** for selecting images close to the class center. The accuracy on the old task increases with the increasing number of images reserved for each class. Choosing images close to the class center is not significantly superior to random selection.

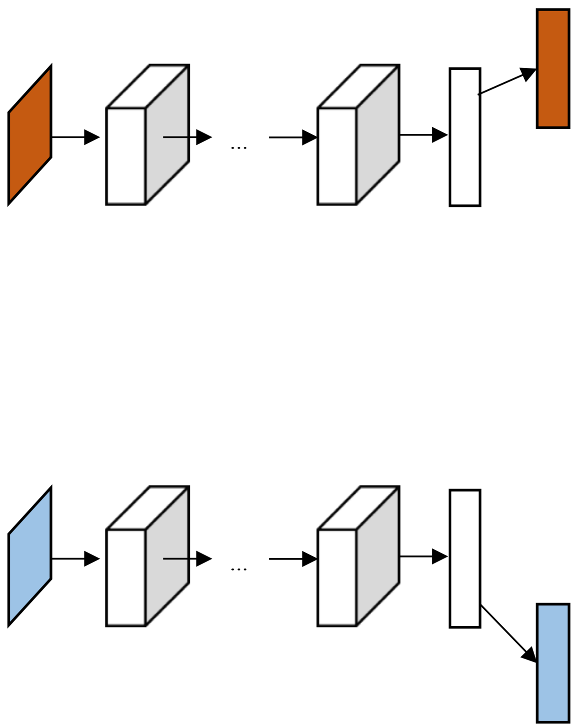
Learning a Unified Classifier Incrementally via Rebalancing

Saihui Hou[1*] Xinyu Pan[2*] Chen Change Loy[3] Zilei Wang[1] Dahua Lin[2]

[1] University of Science and Technology of China [2] The Chinese University of Hong Kong
[3] Nanyang Technological University [* indicates joint first authorship]

(To appear in CVPR 2019)

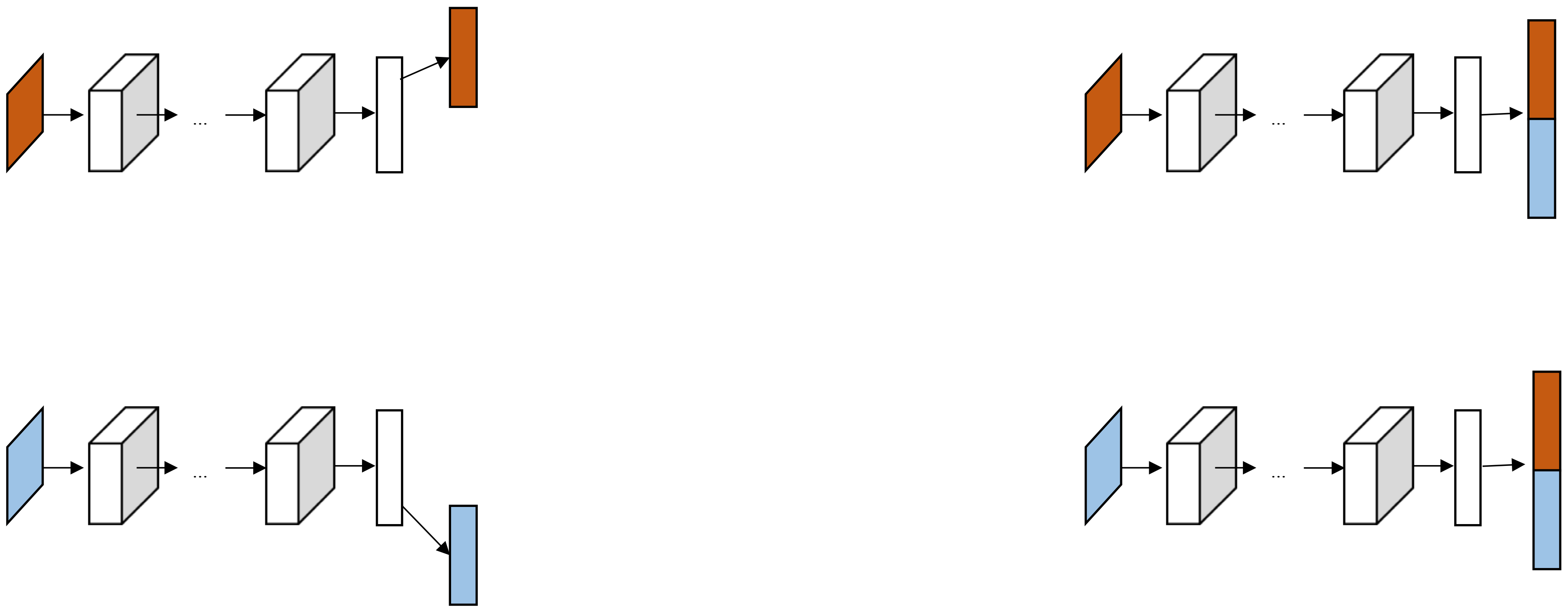
From Multi-task to Multi-class



Multi-task Setting

There is an oracle to tell which classifier should be used at inference time.

From Multi-task to Multi-class



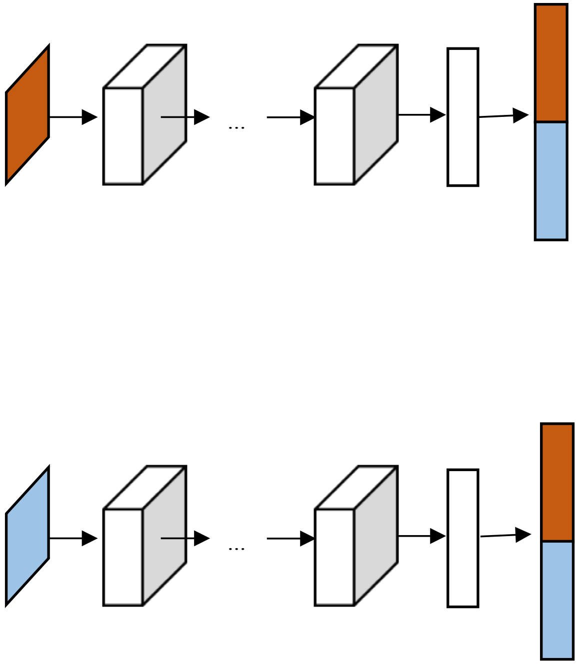
Multi-task Setting

Multi-class Setting

From Multi-task to Multi-class

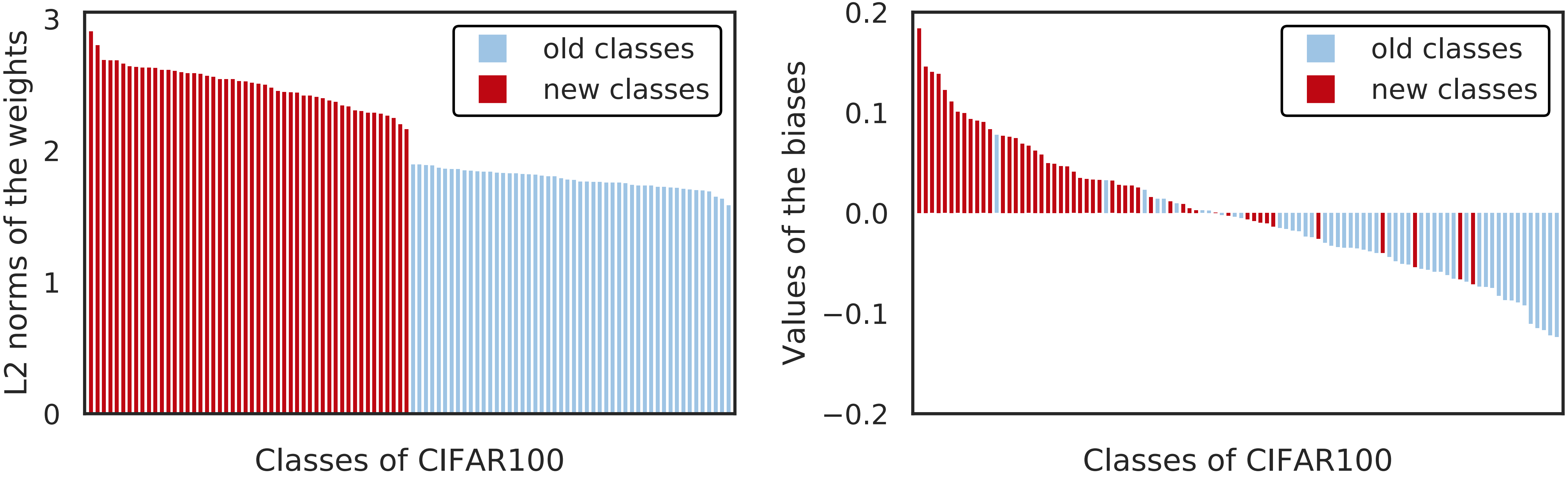
There is no oracle here. But can we simply adapt distillation and retrospection to this setup?

Multi-task Setting

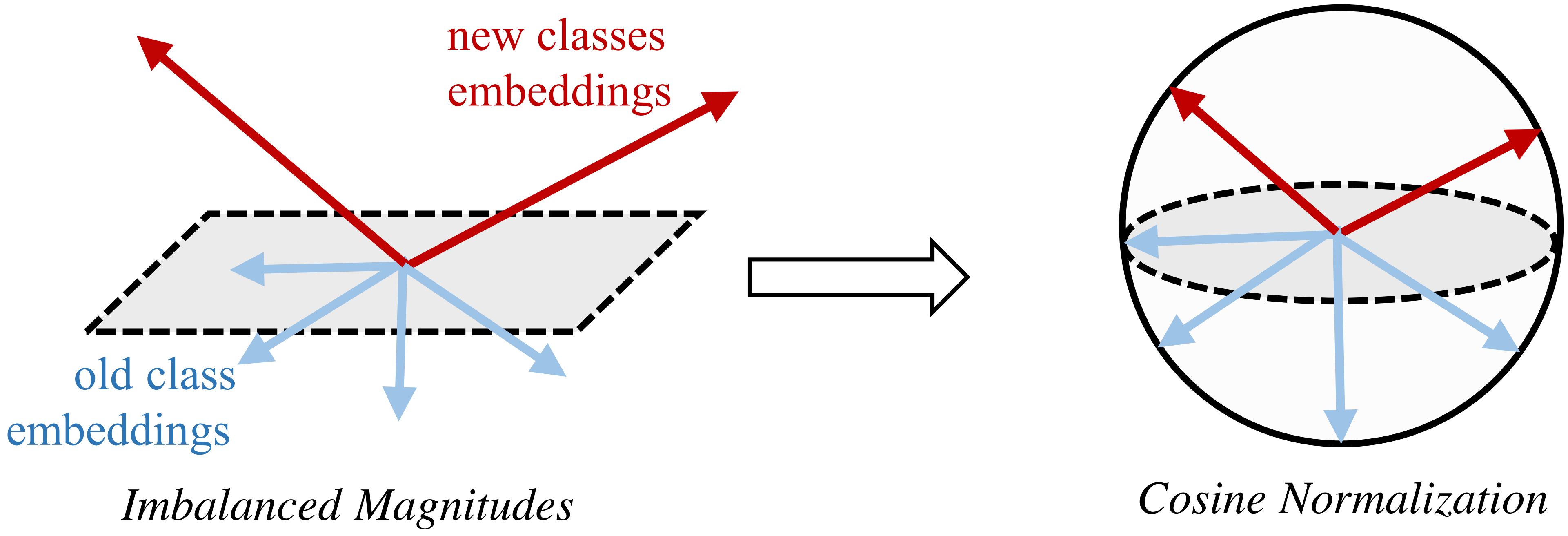


Multi-class Setting

A Toy Example to Visualize Imbalance

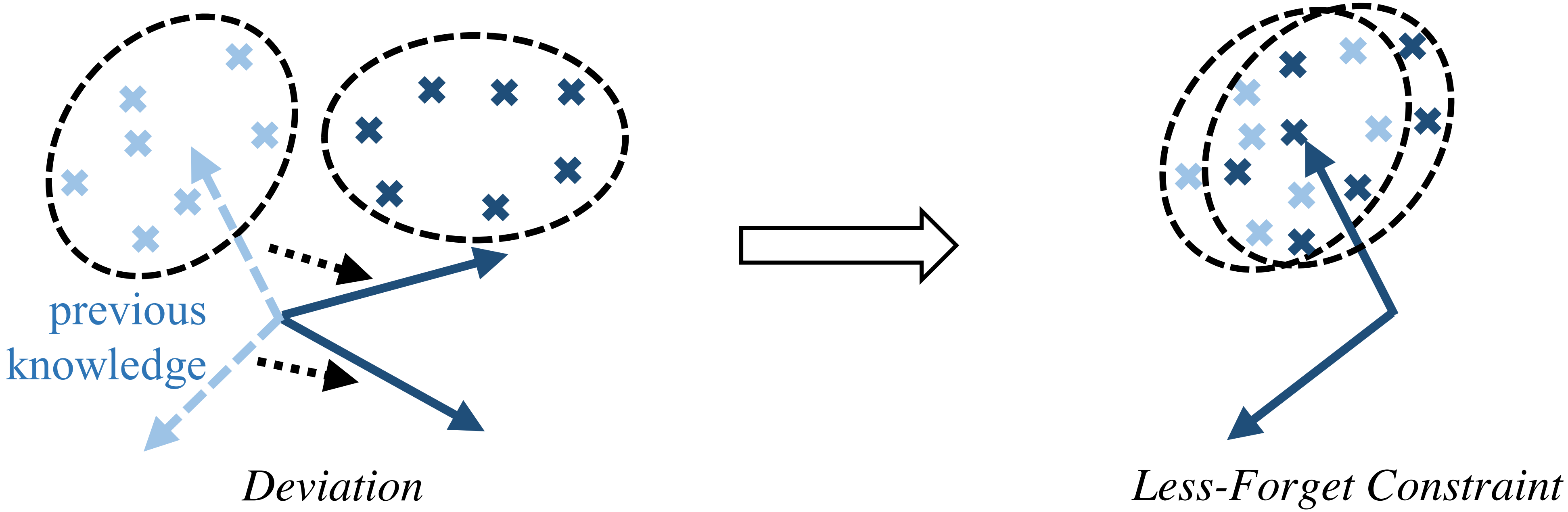


Handle the Imbalance

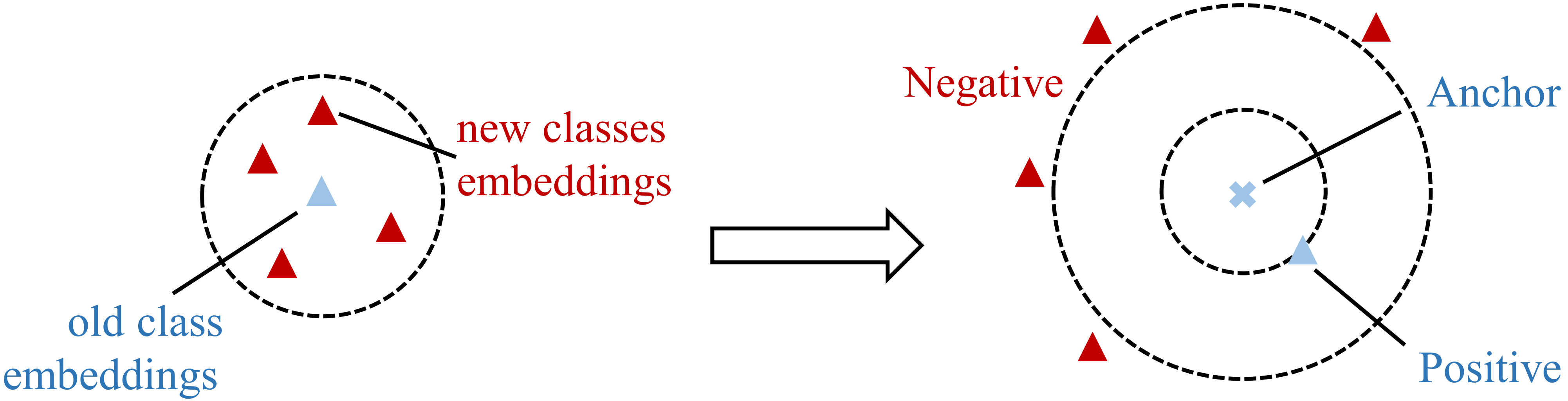


(We will use *embedding* and the weights of last fully-connected layer alternatively in the following.)

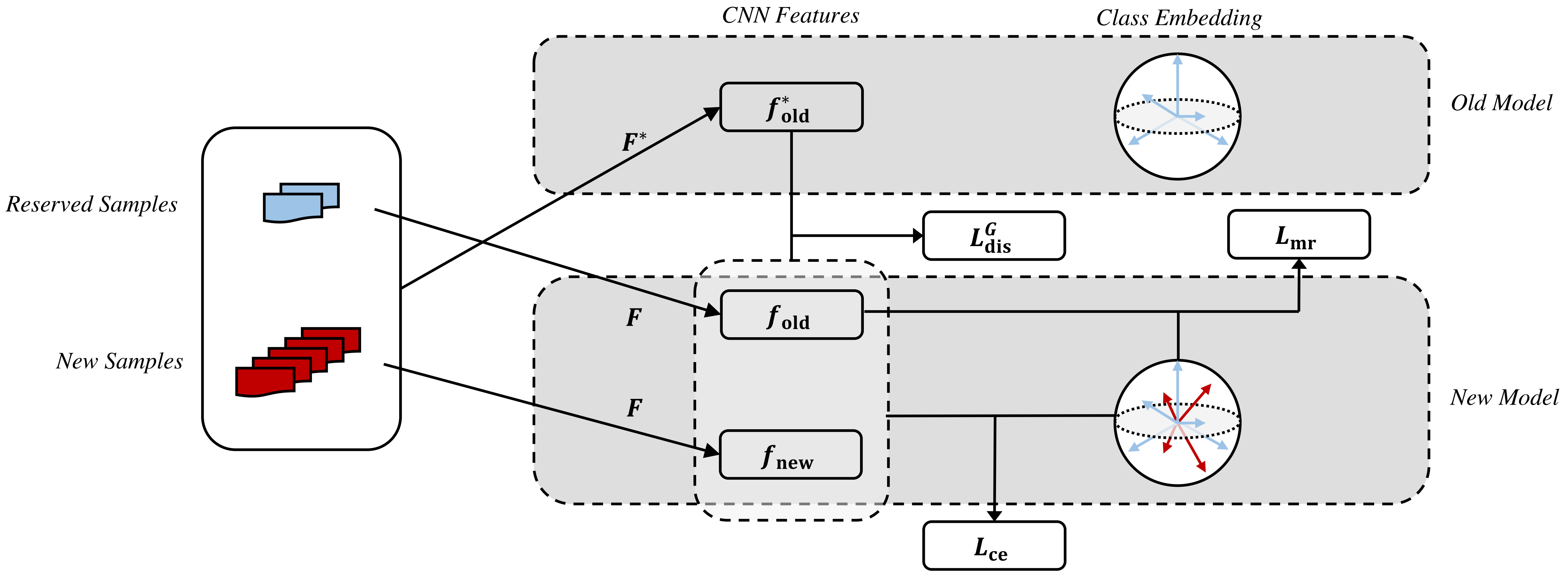
Handle the Imbalance



Handle the Imbalance

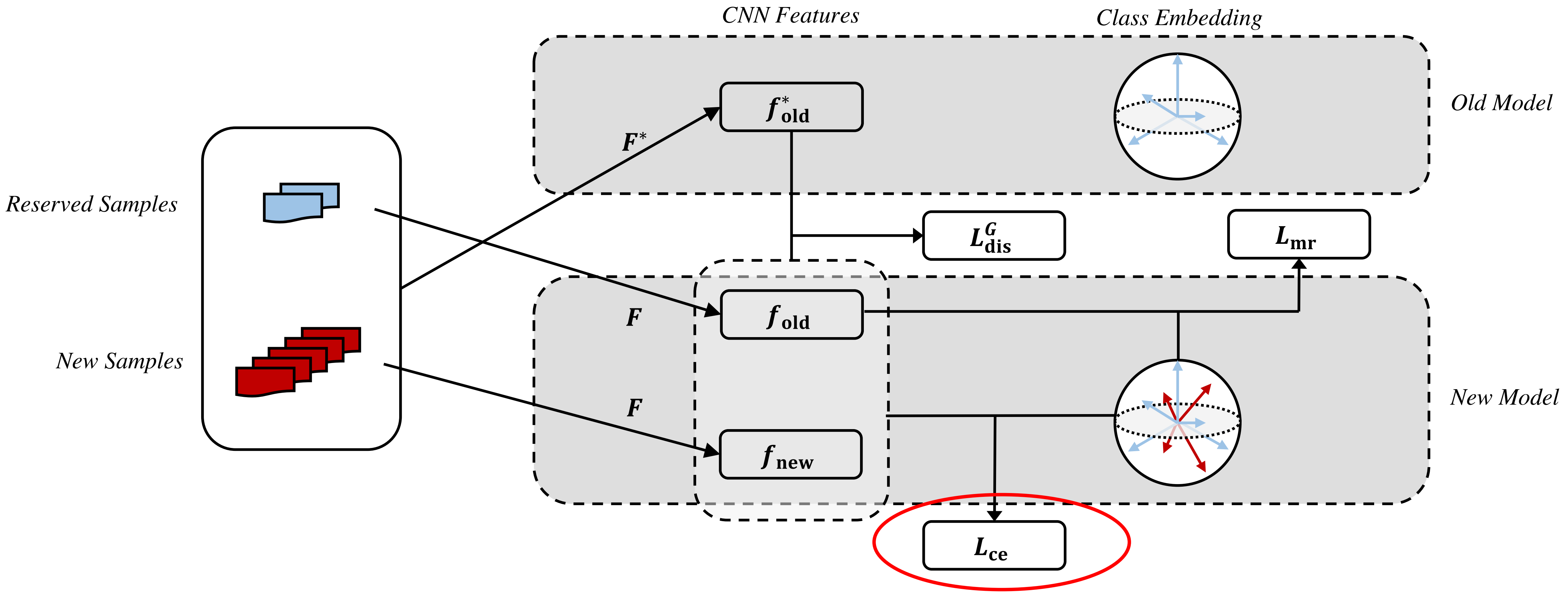


Overview



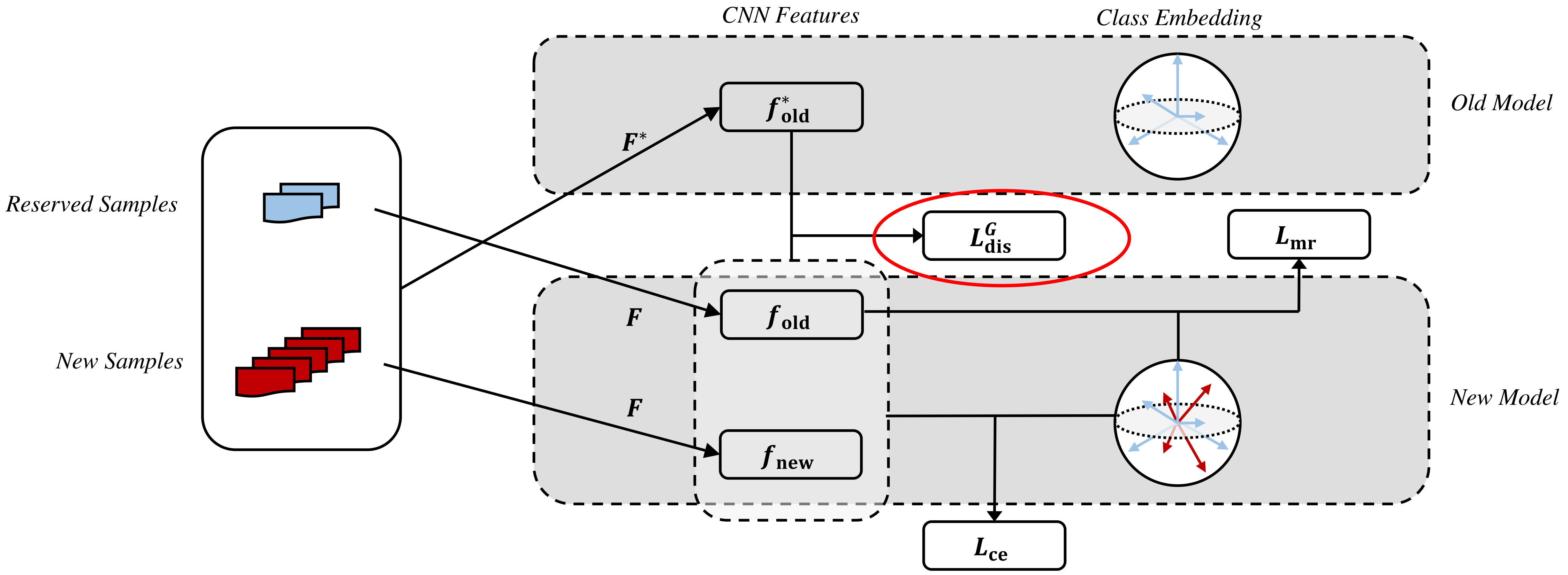
CNN Features
Class Embedding
Old Model
New Model
Reserved Samples
New Samples
F^*
F
F
f^*_{old}
f_{old}
f_{new}
L^G_{dis}
L_{mr}
L_{ce}

Overview



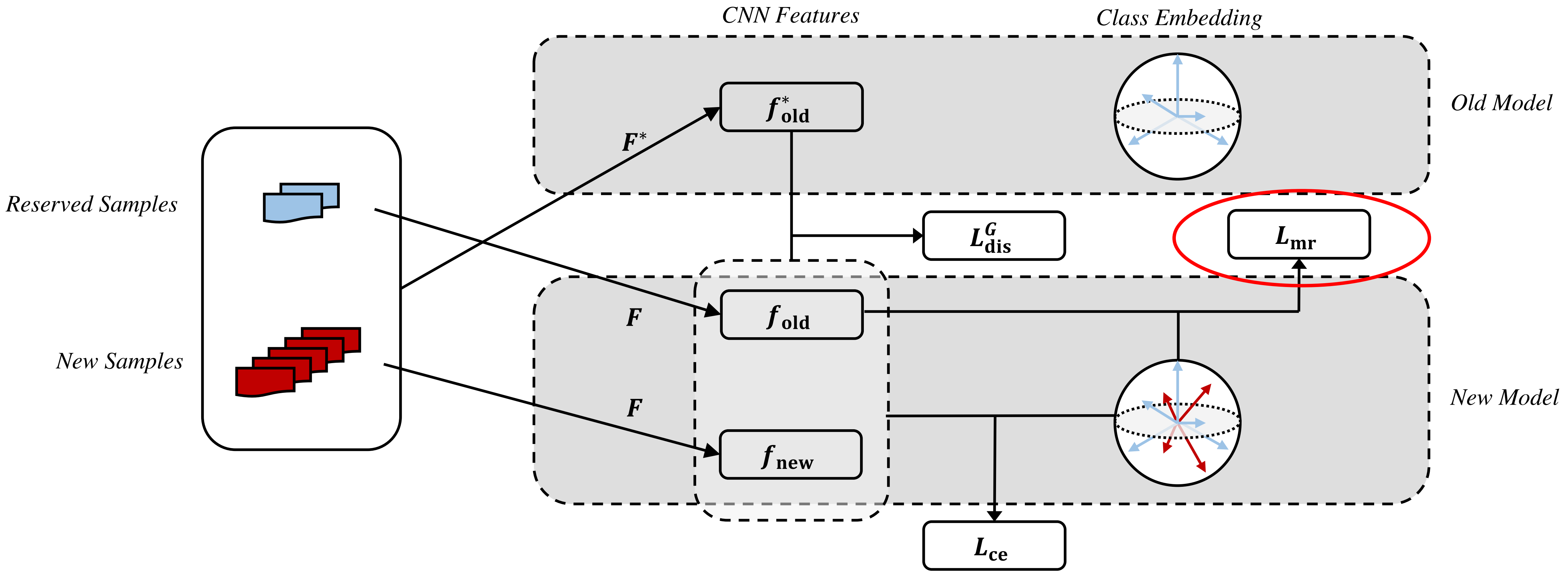
$$L_{\mathrm{ce}}(x) = -\sum_{i=1}^{|\mathcal{C}|} y_i \log\left(p_i\right),$$

Overview



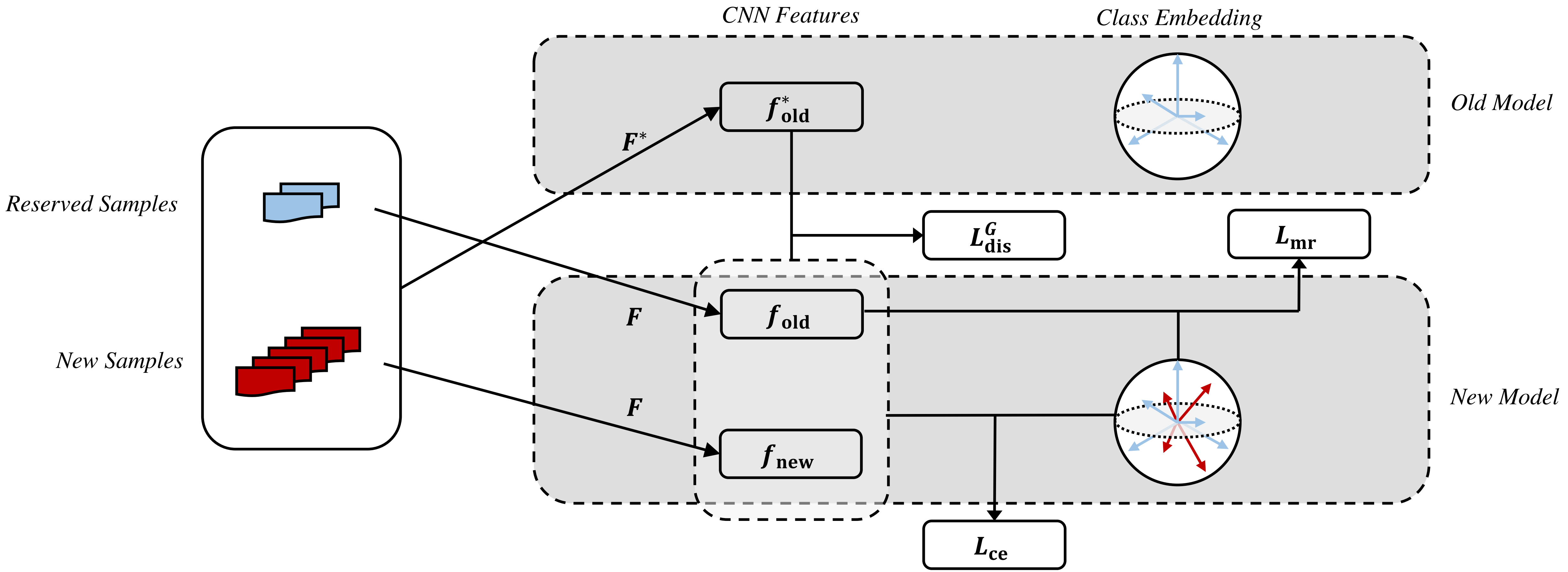
$$L^{\mathrm{G}}_{\mathrm{dis}}(x) = 1 - \langle \bar{f}^*(x), \bar{f}(x) \rangle,$$

Overview



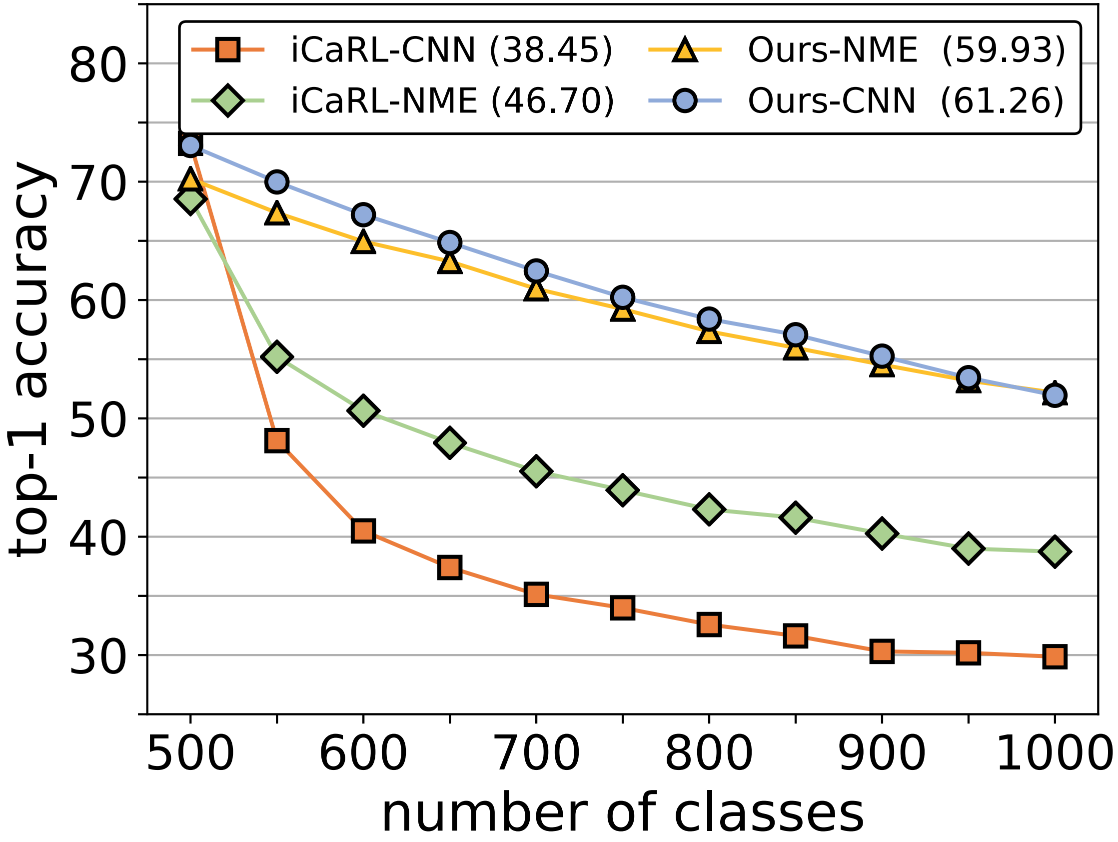
$$L_{\text{mr}}(x) = \sum_{k=1}^{K} \max(m - \langle \bar{\theta}(x), \bar{f}(x) \rangle + \langle \bar{\theta}^k, \bar{f}(x) \rangle,\ 0),$$

Overview

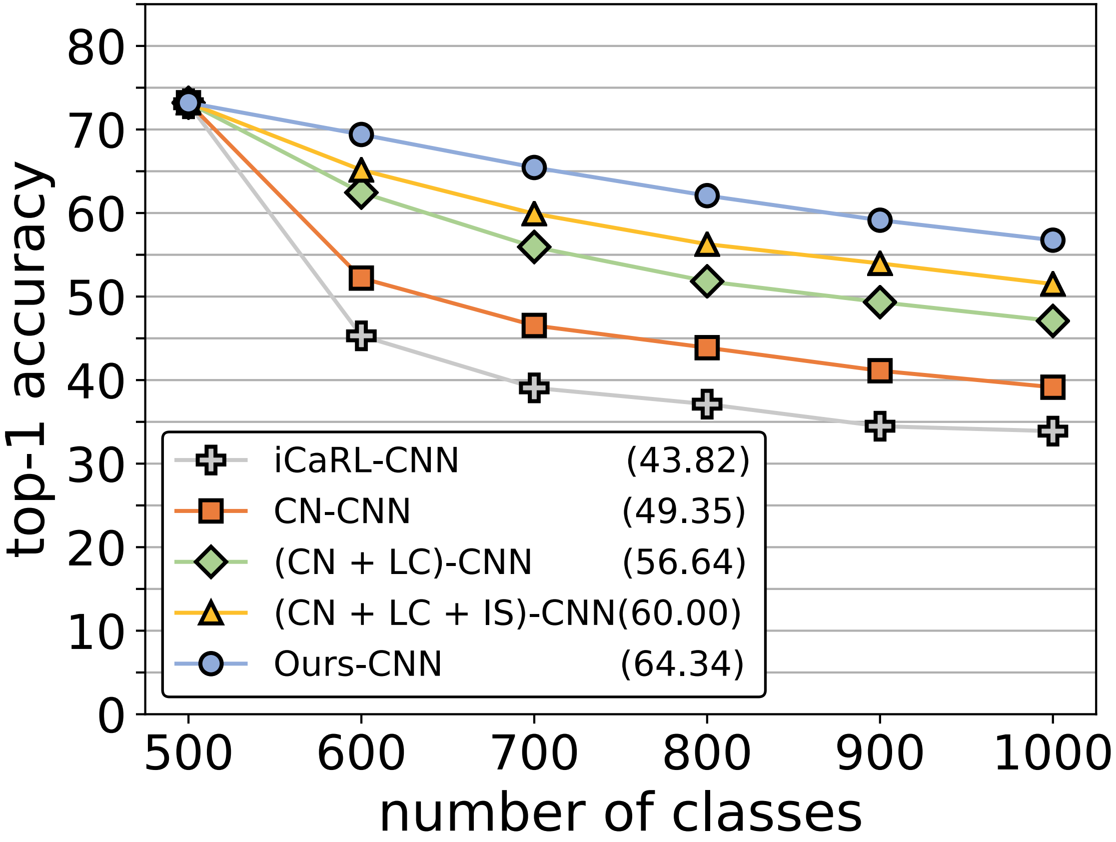


$$L = \frac{1}{|\mathcal{N}|}\sum_{x\in\mathcal{N}}(L_{\text{ce}}(x) + \lambda L^{\text{G}}_{\text{dis}}(x)) + \frac{1}{|\mathcal{N}_{\text{o}}|}\sum_{x\in\mathcal{N}_{\text{o}}} L_{\text{mr}}(x),$$

Some Results



10 phases



5-phase ablation study

Thank you!