Mimicking the Oracle: An Initial Phase Decorrelation Approach for Class Incremental Learning

Yujun Shi[1] Kuangqi Zhou[1] Jian Liang[3] Zihang Jiang[1]
Jiashi Feng[2] Philip Torr[4] Song Bai[2] Vincent Y. F. Tan[1]
[1]National University of Singapore [2]ByteDance Inc.
[3]Institute of Automation, CAS [4]University of Oxford

1. Background & Introduction

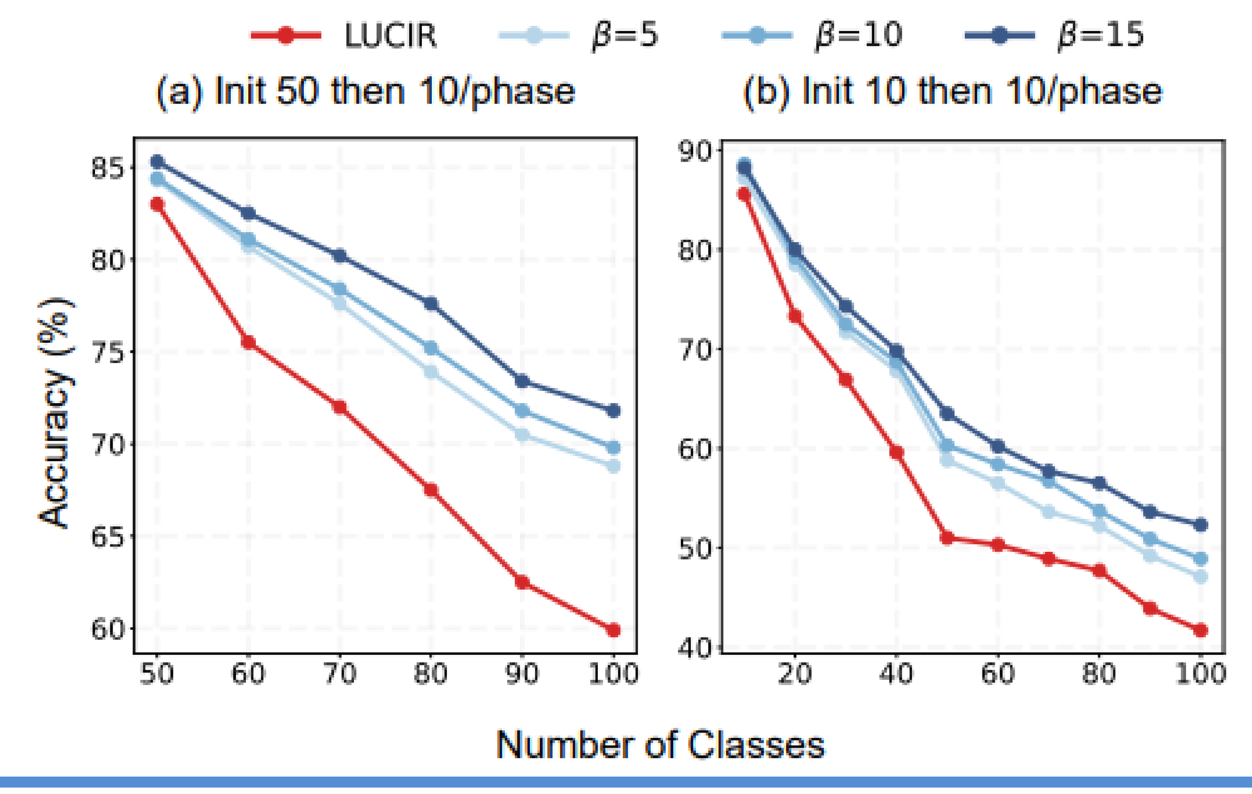
Class Incremental learning (CIL) aims at learning a classifier in a phase-by-phase manner, where only data of a subset of classes are given at each phase. Previous works mainly focus on mitigating forgetting in phases after the initial one. However, **we find that improving CIL at its initial phase is also a promising direction.**

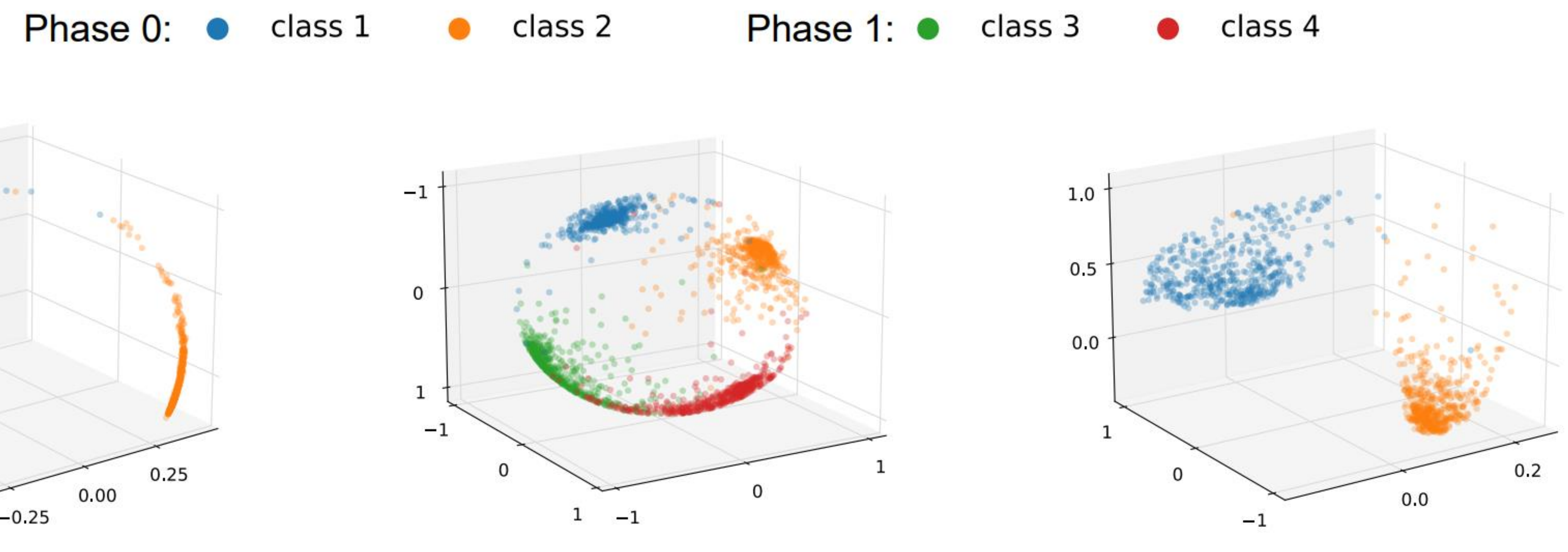
2. An Exploratory Experiment

Firstly, we show that **directly mimicking the oracle model** (i.e., a model jointly trained with all classes) **representations at initial phase** can improve CIL. We use the following objective at initial phase:

$$\min_{\theta} L_{ce}(x, y, \theta) + \beta \left(1 - CosSim\big(f_{\theta}(x), f_{\theta^*}(x)\big)\right),$$

where $f_{\theta}(x)$ and $f_{\theta^*}(x)$ denotes representations output by the initial phase model and the oracle model, respectively. Results are shown below.





1.(a) Naive Training 1.(b) Oracle Model (Joint Training) 1.(c) Applying CwD

3. An observation on representations

Next, we observe class-wise representations produced by an naively trained initial phase model and the oracle model. We find that representations output by the **naively trained model** reside in a **lower dimensional subspace** (see Fig. 1.(a)), while representations output by the **oracle model** can **better use the ambient space** (see Fig. 1.(b)).

4. Class-wise Decorrelation to mimic the oracle

Motivated by our observations, we propose a novel regularization objective at initial phase of CIL, termed **Class-wise Decorrelation (CwD):**

$$L_{CwD} = \frac{1}{Cd^2} \sum_{c=1}^{C} \left\| K^{(c)} \right\|_F^2,$$

where $K^{(c)}$ is the correlation matrix of class c representations. With CwD, representations of each class can better use the ambient space instead of simply residing in a lower dimensional subspace (see Fig. 1.(c)).

5. Experiments

B: initial phase classes. S: classes per phase after the initial one. R: exemplars per class. η: coefficient of CwD objective. Results are averaged over 3 seeds. Without specific description, all results are with ImageNet100 and LUCIR.

Method	CIFAR100 (B=50)			ImageNet100 (B=50)			ImageNet (B=100)	
	S=10	5	2	10	5	2	100	50
LwF [18]	53.59±0.51	48.66±0.58	45.56±0.28	53.62†	47.64†	44.32†	40.86±0.13	27.72±0.12
iCaRL [26]	60.82±0.03	53.74±0.25	47.86±0.41	65.44†	59.88†	52.97†	49.56±0.09	42.61±0.15
BiC [33]	51.58±0.16	48.07±0.02	43.10±0.37	70.07†	64.96†	57.73†	43.23±0.13	38.83±0.12
LUCIR [12]	66.27±0.28	60.80±0.29	52.96±0.25	70.60±0.43	67.76±0.40	62.76±0.22	56.40±0.10	52.75±0.18
+CwD (ours)	67.26±0.16	62.89±0.09	56.81±0.21	71.94±0.11	69.34±0.31	65.10±0.59	57.42±0.11	53.37±0.22
PODNet [8]	66.98±0.13	63.76±0.48	61.00±0.18	75.71±0.37	72.80±0.35	65.57±0.41	57.01±0.12	54.06±0.09
+CwD (ours)	67.44±0.35	64.64±0.38	62.24±0.32	76.91±0.10	74.34±0.02	67.42±0.07	58.18±0.20	56.01±0.14
AANet [19]	69.79±0.21	67.97±0.26	64.92±0.30	71.96±0.12	70.05±0.63	67.28±0.34	51.76*±0.14	46.86*±0.13
+CwD (ours)	70.30±0.37	68.62±0.17	66.17±0.13	72.92±0.29	71.10±0.16	68.18±0.27	52.30*±0.08	47.61*±0.20

S	B	LUCIR	+CwD (ours)	↑
10	10	57.01±0.14	57.90±0.07	+0.89
	20	61.21±0.35	62.49±0.36	+1.28
	30	64.82±0.38	66.54±0.35	+1.72
	40	67.68±0.37	69.70±0.10	+2.02
	50	70.60±0.43	71.94±0.11	+1.33
5	10	50.47±0.31	51.92±0.10	+1.45
	20	56.41±0.37	58.14±0.13	+1.73
	30	61.00±0.09	63.18±0.14	+2.18
	40	63.73±0.23	66.25±0.16	+2.52
	50	67.76±0.40	69.34±0.31	+1.58

S	R	LUCIR	+CwD (ours)	↑
10	40	72.41±0.61	73.29±0.11	+0.88
	30	71.70±0.37	72.63±0.15	+0.93
	20	70.60±0.43	71.94±0.11	+1.34
	10	68.73±0.52	69.77±0.04	+1.04
	5	66.49±0.52	67.63±0.07	+1.14
5	40	70.74±0.49	72.06±0.11	+1.32
	30	68.56±0.42	70.04±0.12	+1.44
	20	67.76±0.40	69.34±0.31	+1.58
	10	64.07±0.38	66.07±0.46	+2.00
	5	60.41±0.77	62.58±0.53	+2.17