DMQA Open Seminar

Representation-based Continual Learning

Seokho Moon

School of Industrial and Management Engineering, Korea University

danny232@korea.ac.kr

June 9, 2023

❖ 문석호 (Seokho Moon)

- 고려대학교 산업경영공학부 졸업(2019.08)
- 고려대학교 산업경영공학과 대학원 재학 중
- Data Mining & Quality Analytics Lab (김성범 교수님)
- 석박사통합과정 (2019.09 ~)

❖ 관심 연구 분야

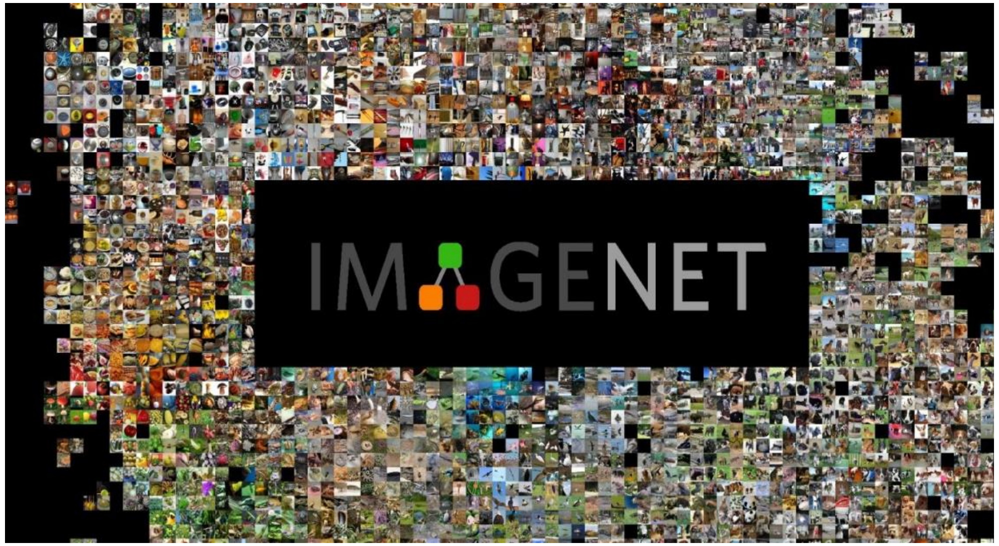
- Continual/Incremental Learning
- Self/Semi-supervised learning
- Anomaly Detection

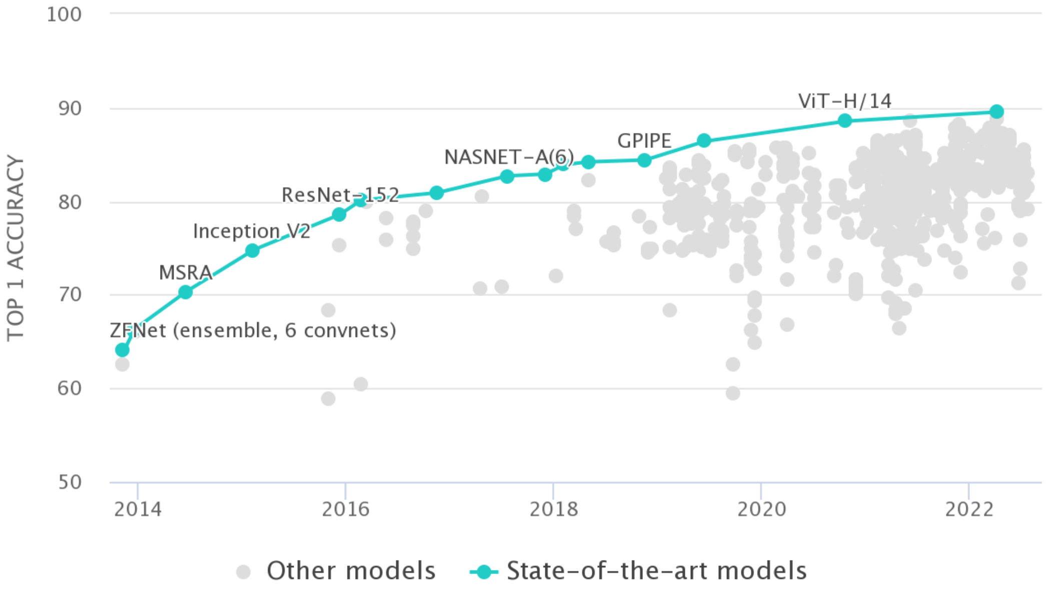
❖ E-mail

- danny232@korea.ac.kr

Introduction

- 현재의 machine learning 및 deep learning은 특정 task에서 매우 우수한 성능을 보여주고 있음
- 예를 들어 image classification의 경우, 인간의 분류 성능을 이미 넘어섰으며 지속적으로 성능이 증가하는 상황





https://paperswithcode.com/sota/image-classification-on-imagenet

Introduction

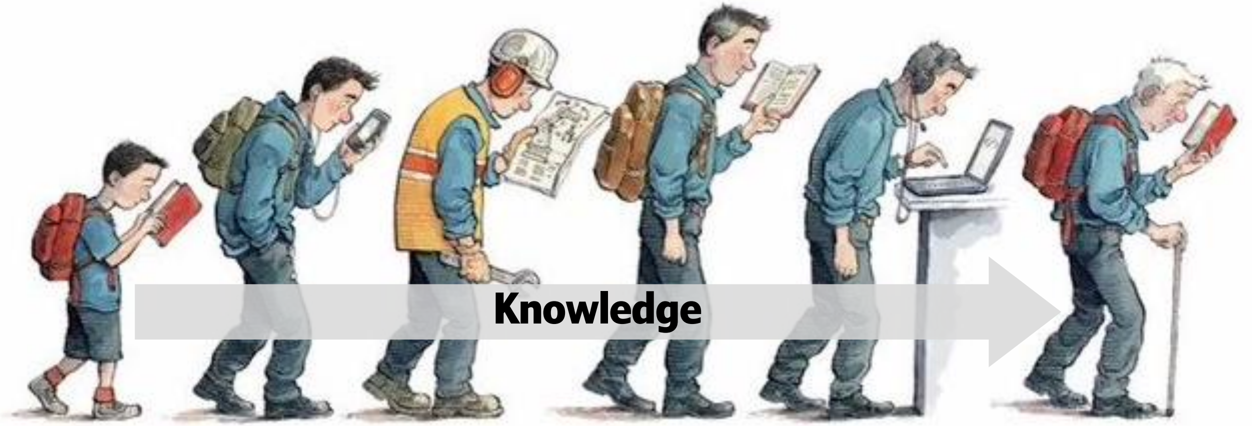
- 그러나 앞선 예시의 ImageNet에서 학습하지 않은 데이터에 대한 inference 성능이 저하되는 등의 문제가 있으며, 인간과 유사한 Artificial Intelligence (AI) 으로 보기에는 **여러가지 한계점**이 존재

1. 지속적으로 학습하지 않으며, knowledge가 **축적/이동되지 않음**
2. 문제 상황이 **close-world**를 가정하고 있어서, 데이터가 새롭게 추가되는 실제 현실에 대한 대응이 어려움
3. 한번 학습이 완료된 모델을 test task에 적용한 후에, 해당 모델이 **새롭게 학습되지 않음**

이러한 문제를 해결하기 위해

continual learning 방식을 제안

Continual Learning of Humans



https://www.economist.com/special-report/2017/01/12/lifelong-learning-is-becoming-an-economic-imperative

Liu, Bing, and Sahisnu Mazumder. "Lifelong and continual learning dialogue systems: learning during conversation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 17. 2021.

Introduction

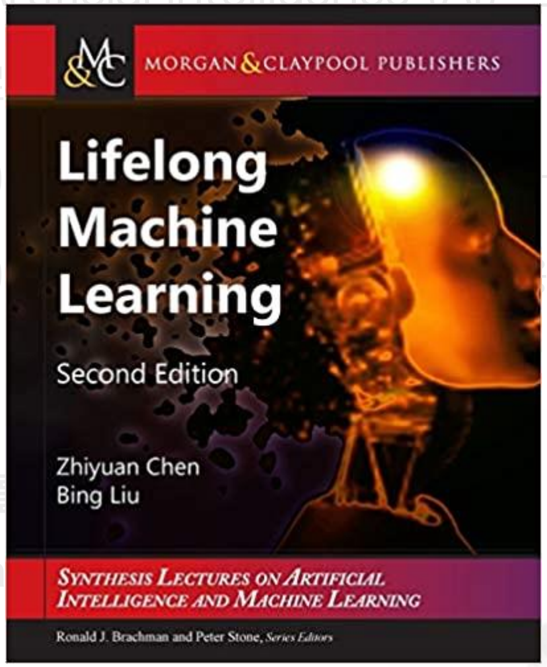
- 그러나 앞선 예시의 ImageNet에서 학습하지 않은 데이터에 대한 inference 성능이 저하되는 등의 문제가 있으며, 인간과 유사한 Artificial Intelligence (AI) 으로 보기에는 여러가지 한계점이 존재

1. 지속적으로 … 축적/이동되지 …
2. 문제 상황이 … , 데이터가 새롭 … 이 어려움
3. 한번 학습이 … 한 후에, 해당 도 …

Contin … ans

이러한 문제를 …

continual lea …



MORGAN & CLAYPOOL PUBLISHERS

Lifelong Machine Learning

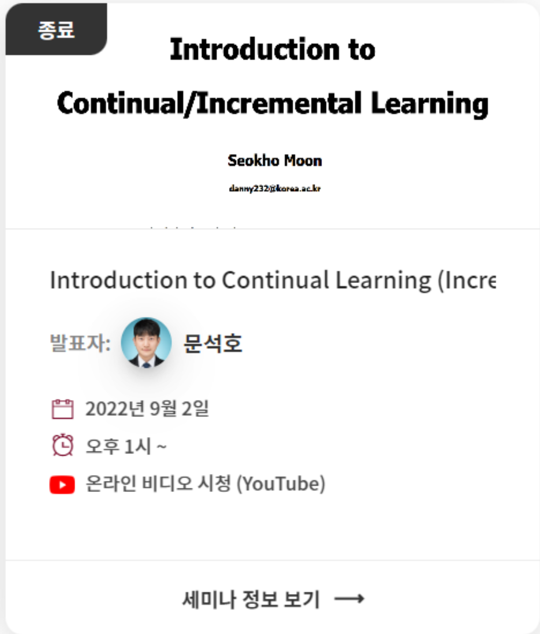
Second Edition

Zhiyuan Chen
Bing Liu

SYNTHESIS LECTURES ON ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Ronald J. Brachman and Peter Stone, *Series Editors*

https://www.cs.uic.edu/~liub/Part-1-continual-learning-slides.pdf
https://www.cs.uic.edu/~liub/lifelong-machine-learning.html



종료

Introduction to
Continual/Incremental Learning

Seokho Moon
danny232@korea.ac.kr

Introduction to Continual Learning (Incre

발표자: 문석호

2022년 9월 2일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

http://dmqm.korea.ac.kr/activity/seminar/378

Knowledge

Continual Learning 에 대한 기본 자료.

https://www.economist.com/special-report/2017/01/12/lifelong-learning-is-becoming-an-economic-imperative

Introduction

- Continual Learning은 접근 방식에 따라 3~5개 정도로 나뉠 수 있음
- 논문들에서 여러 접근 방식을 결합하여 아이디어 제시 가능 (e.g., experience replay + knowledge distillation)

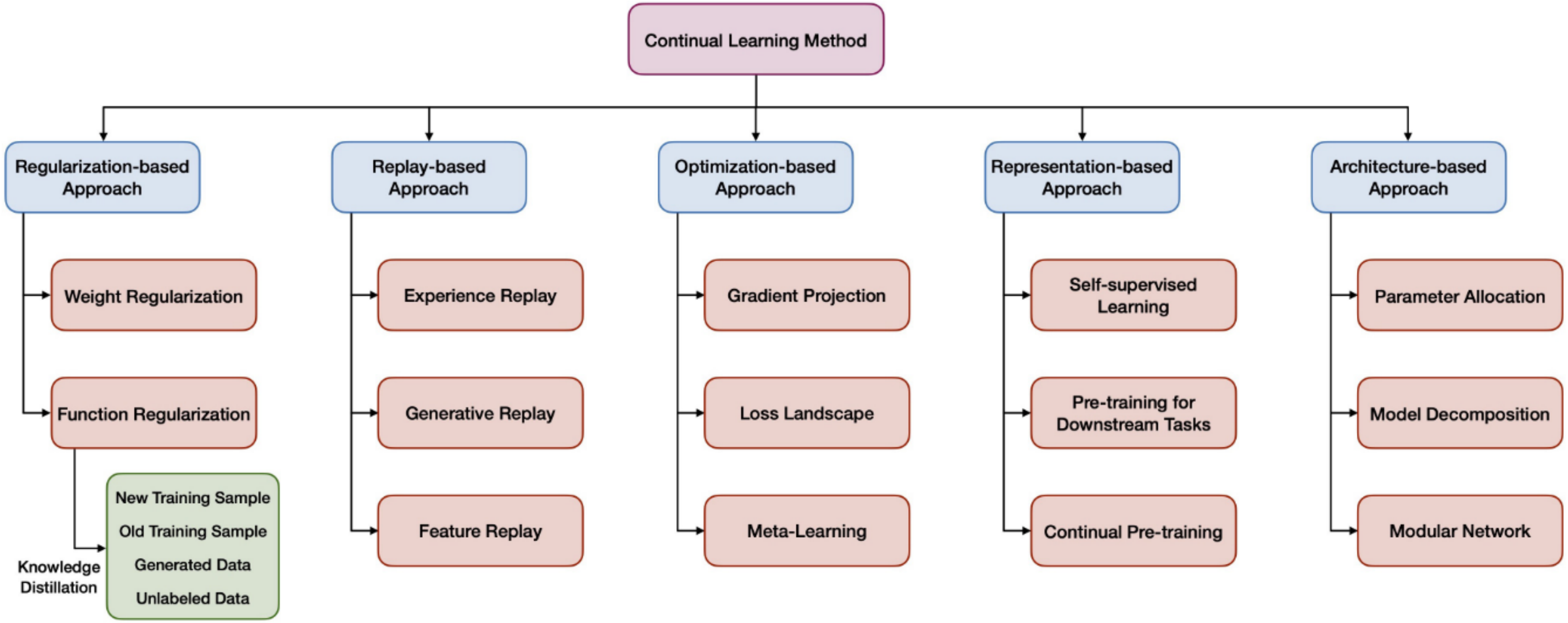
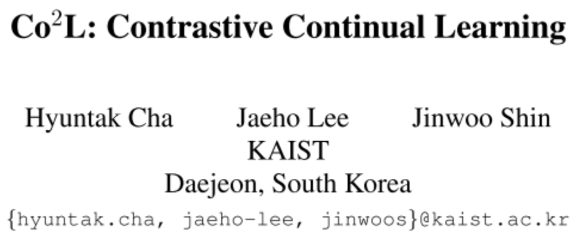


Fig. 3. A state-of-the-art and elaborated taxonomy of representative continual learning methods. We have summarized 5 main categories (blue blocks), each of which is further divided into several sub-directions (red blocks).

Wang, Liyuan, et al. "A comprehensive survey of continual learning: Theory, method and application." arXiv preprint arXiv:2302.00487 (2023).

Paper Review

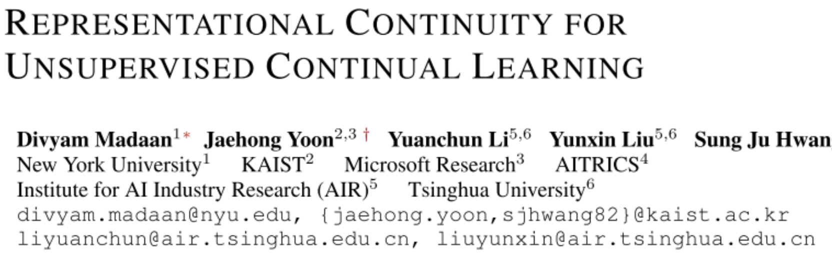
Co^{2}L: Contrastive Continual Learning (ICCV 2021)



Co^{2}L: Contrastive Continual Learning

Hyuntak Cha Jaeho Lee Jinwoo Shin
KAIST
Daejeon, South Korea
{hyuntak.cha, jaeho-lee, jinwoos}@kaist.ac.kr

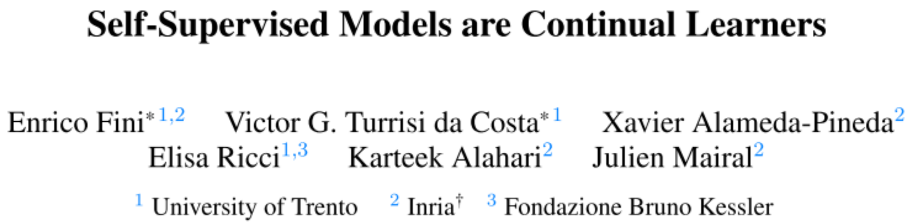
Representational Continuity for Unsupervised Continual Learning (ICLR 2022)



REPRESENTATIONAL CONTINUITY FOR UNSUPERVISED CONTINUAL LEARNING

Divyam Madaan[1*] **Jaehong Yoon**[2,3 †] **Yuanchun Li**[5,6] **Yunxin Liu**[5,6] **Sung Ju Hwang**[2,4]
New York University[1] KAIST[2] Microsoft Research[3] AITRICS[4]
Institute for AI Industry Research (AIR)[5] Tsinghua University[6]
divyam.madaan@nyu.edu, {jaehong.yoon,sjhwang82}@kaist.ac.kr
liyuanchun@air.tsinghua.edu.cn, liuyunxin@air.tsinghua.edu.cn

Self-Supervised Models are Continual Learners (CVPR 2022)



Self-Supervised Models are Continual Learners

Enrico Fini[*1,2] Victor G. Turrisi da Costa[*1] Xavier Alameda-Pineda[2]
Elisa Ricci[1,3] Karteek Alahari[2] Julien Mairal[2]
[1] University of Trento [2] Inria[†] [3] Fondazione Bruno Kessler

Paper Review: Co^{2}L (ICCV 2021)

- Introduction

- 현실 세계에서는 데이터가 순차적으로 수집되는 경우가 많은데, 이 경우에 하나의 모델을 지속적으로 업데이트하면 기존 정보를 잃는 **catastrophic forgetting** 현상이 발생함.
- 이러한 문제를 해결하기 위해 replay-based, regularization-based 등의 접근 방식으로 연구가 진행되었음.

What type of knowledge is likely to be useful for future tasks (and thus not get forgotten),

and how can we learn and preserve such knowledge?

- 대부분의 선행 연구는 catastrophic forgetting을 해결하기 위해 기존 지식(knowledge)를 보존하는데 집중하였으나, 저자들은 **transferable representations**을 학습하는 것이 더 중요하다고 주장함.

  → 다시 말해, 미래 task에도 필요한 knowledge를 찾아내고 전달하는 것이 continual learning에서 중요.

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co^{2}L (ICCV 2021)

- Introduction

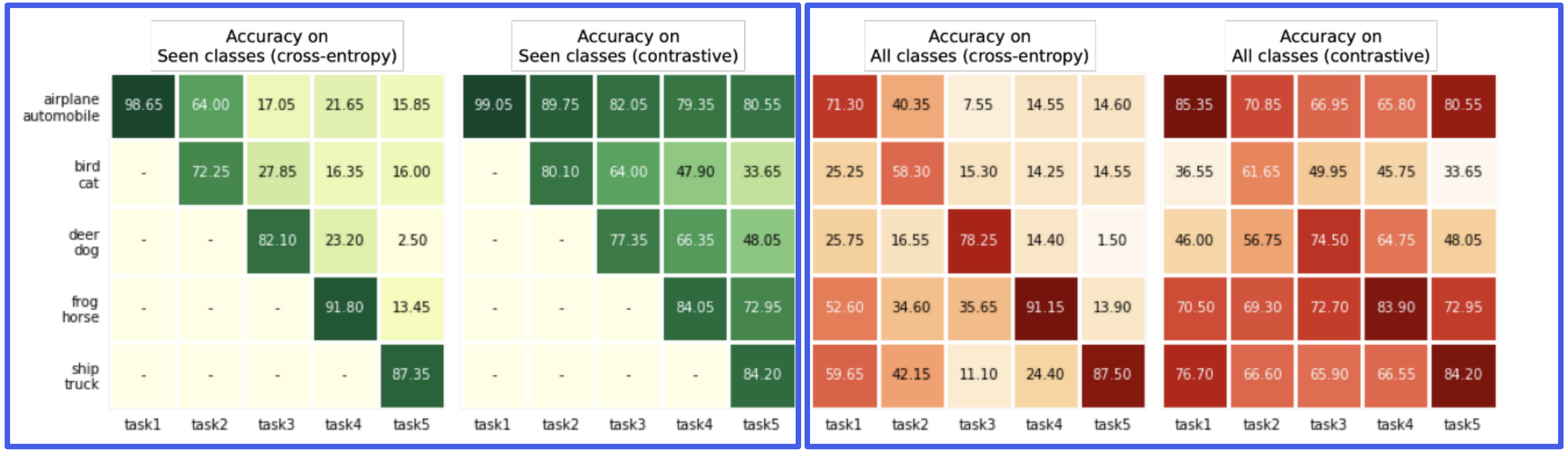
- **Transferable representations**을 잘 학습하기 위한 방법으로 저자들은 contrastive learning을 선택. (SupCon)

① Contrastive learning이 transferable representations을 잘 추출할 것이라는 가설.

② Cross-entropy 방식과 contrastive learning 방식을 forgetting 관점과 useful for future task 관점으로 비교하여 증명.

기존 정보를 forgetting 하는지에 대해 비교

Unseen(future) objects에 대해 유용한 정보를 학습하는지 비교



Accuracy on Seen classes (cross-entropy)

	task1	task2	task3	task4	task5
airplane automobile	98.65	64.00	17.05	21.65	15.85
bird cat	-	72.25	27.85	16.35	16.00
deer dog	-	-	82.10	23.20	2.50
frog horse	-	-	-	91.80	13.45
ship truck	-	-	-	-	87.35

Accuracy on Seen classes (contrastive)

	task1	task2	task3	task4	task5
airplane automobile	99.05	89.75	82.05	79.35	80.55
bird cat	-	80.10	64.00	47.90	33.65
deer dog	-	-	77.35	66.35	48.05
frog horse	-	-	-	84.05	72.95
ship truck	-	-	-	-	84.20

Accuracy on All classes (cross-entropy)

	task1	task2	task3	task4	task5
airplane automobile	71.30	40.35	7.55	14.55	14.60
bird cat	25.25	58.30	15.30	14.25	14.55
deer dog	25.75	16.55	78.25	14.40	1.50
frog horse	52.60	34.60	35.65	91.15	13.90
ship truck	59.65	42.15	11.10	24.40	87.50

Accuracy on All classes (contrastive)

	task1	task2	task3	task4	task5
airplane automobile	85.35	70.85	66.95	65.80	80.55
bird cat	36.55	61.65	49.95	45.75	33.65
deer dog	46.00	56.75	74.50	64.75	48.05
frog horse	70.50	69.30	72.70	83.90	72.95
ship truck	76.70	66.60	65.90	66.55	84.20

Figure 3. Observation on two learning schemes, cross-entropy loss training and contrastive representation learning on Seq-CIFAR-10 without any design used for the continual learning settings. As new task arrives, each model is trained only with current task samples with model weights without re-initialization. After each task training ends, a new linear classifier is trained on the fixed current representation with samples observed so far (denoted by "seen objects") or all samples including ones from future tasks (denoted by "all objects"). The pair of left figures shows contrastively trained representations suffer less from forgetting than the ones trained with cross entropy loss. The right pair shows contrastively learned representation is much more useful to perform unseen objects classification tasks.

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

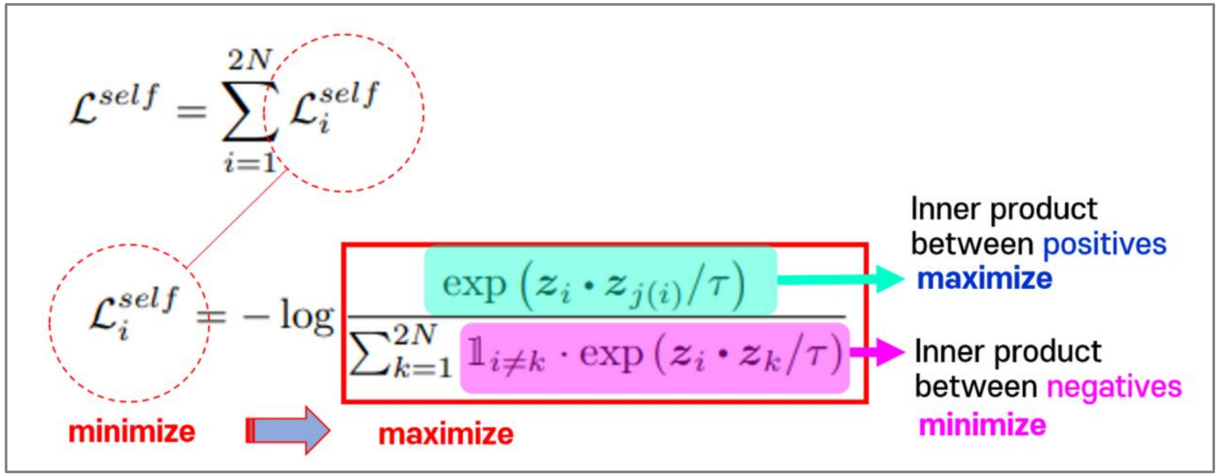
Paper Review: Co^{2}L (ICCV 2021)

- Preliminaries: Contrastive Learning

- 본 논문에서는 SupCon[1](Supervised Contrastive learning, NeurIPS 2020) 방식을 사용함.

- 참고자료[2]

  Self-supervised contrastive learning vs SupCon 비교 설명.

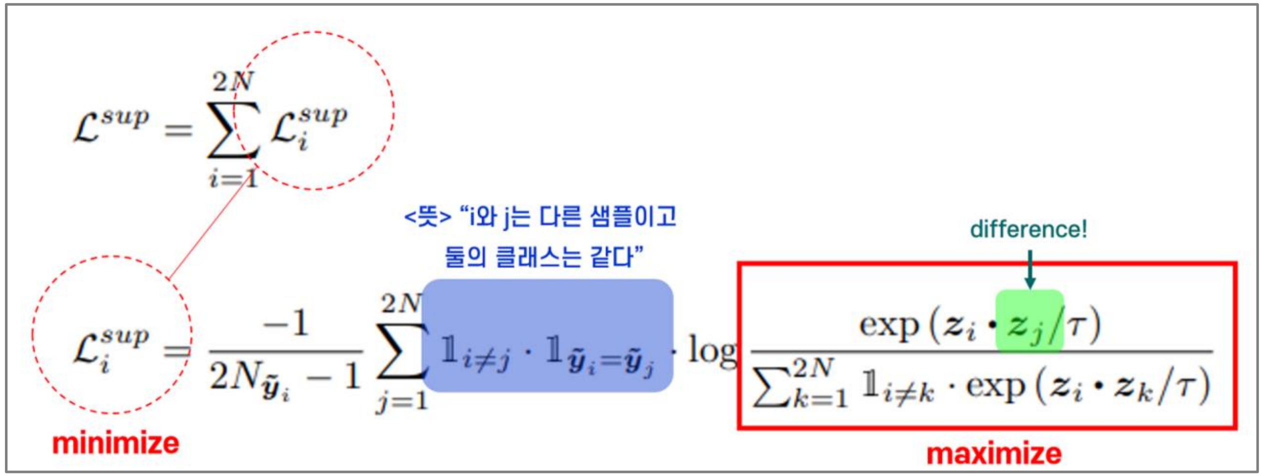


$$\mathcal{L}^{self} = \sum_{i=1}^{2N} \mathcal{L}_i^{self}$$

$$\mathcal{L}_i^{self} = -\log \frac{\exp\left(\boldsymbol{z}_i \bullet \boldsymbol{z}_{j(i)} / \tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp\left(\boldsymbol{z}_i \bullet \boldsymbol{z}_k / \tau\right)}$$

Inner product between positives maximize
Inner product between negatives minimize
minimize
maximize

[Self-supervised contrastive learning]



$$\mathcal{L}^{sup} = \sum_{i=1}^{2N} \mathcal{L}_i^{sup}$$

$$\mathcal{L}_i^{sup} = \frac{-1}{2N_{\tilde{\boldsymbol{y}}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{y}}_j} \cdot \log \frac{\exp\left(\boldsymbol{z}_i \bullet \boldsymbol{z}_j / \tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp\left(\boldsymbol{z}_i \bullet \boldsymbol{z}_k / \tau\right)}$$

[Supervised Contrastive learning]

[1] Khosla, Prannay, et al. "Supervised contrastive learning." *Advances in neural information processing systems* 33 (2020): 18661-18673.

[2] https://blog.naver.com/heianjung/222616494664

Paper Review: Co²L (ICCV 2021)

- Preliminaries: Contrastive Learning

- 본 논문에서는 SupCon[1](Supervised Contrastive learning, NeurIPS 2020) 방식을 사용함.

- 앞서 설명한 내용을 해당 논문에서 정리한 수식으로 보면 아래와 같음.

$$\mathcal{L}^{\text{sup}} = \sum_{i=1}^{2N} \frac{-1}{|\mathfrak{p}_i|} \sum_{j \in \mathfrak{p}_i} \log \left(\frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_j / \tau)}{\sum_{k \neq i} \exp(\mathbf{z}_i \cdot \mathbf{z}_k / \tau)} \right)$$

$$\mathfrak{p}_i = \left\{ j \in \{1, \ldots, 2N\} \;\middle|\; j \neq i,\ y_j = y_i \right\}$$

The index set of positive samples

[1] Khosla, Prannay, et al. "Supervised contrastive learning." *Advances in neural information processing systems* 33 (2020): 18661-18673.

Paper Review: Co²L (ICCV 2021)

- Co²L: Contrastive Continual Learning

- 제안 방법론은 아래처럼 두 가지의 loss의 합으로 이뤄져 있음. 각 부분에 대해 뒤에서 자세히 설명.

  ① 현재 들어온 data(task)에 대한 학습 부분 (sup asym; Asymmetric Supervised Contrastive Loss)

  ② 기존 지식을 보존하는 부분 (IRD; Instance-wise Relation Distillation)

$$\mathcal{L} = \underbrace{\mathcal{L}^{\text{sup}}_{\text{asym}}}_{\text{(1) learning}} + \underbrace{\lambda \cdot \mathcal{L}^{\text{IRD}}}_{\text{(2) preserving}}$$

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co²L (ICCV 2021)

- Co²L: (1) Asymmetric Supervised Contrastive Loss

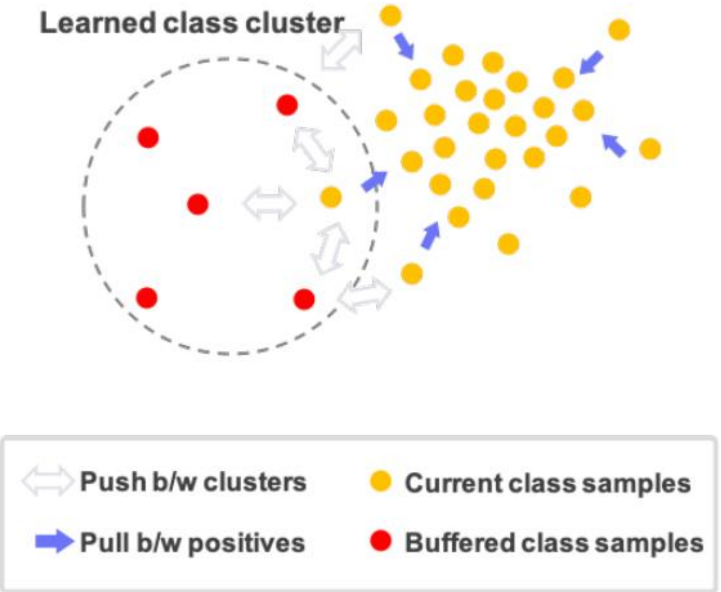
- 참고 사항: 제안 방법론은 memory buffer를 사용해서 **일부 past task data에 대해 보관**하는 rehearsal-based 방식 사용.

  → 따라서 학습 시 각 batch에는 current task samples와 past task samples가 섞여서 존재.

- 본 논문은 기존 SupCon을 asymmetric 한 형태로 수정해서 loss를 제안.
  ① Current task samples → anchors로 설정
  ② Past task samples → negative samples로 설정

$$\mathcal{L}^{\text{sup}}_{\text{asym}} = \sum_{i \in S} \frac{-1}{|\mathfrak{p}_i|} \sum_{j \in \mathfrak{p}_i} \log \left(\frac{\exp\left(\mathbf{z}_i \cdot \mathbf{z}_j / \tau\right)}{\sum_{k \neq i} \exp\left(\mathbf{z}_i \cdot \mathbf{z}_k / \tau\right)} \right)$$

$S \subset \{1, \ldots, 2N\}$ ← the set of indices of current task samples in the batch



(a) Asymmetric SupCon Loss

해당 방법은 기존 SupCon과 비교하여, past task samples에 대한 overfitting을 방지하여 성능 향상

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co²L (ICCV 2021)

- Co²L: (2) Instance-wise Relation Distillation (IRD)

- 해당 아이디어는 SEED[1] (ICLR 2021)에서 영감을 얻어 제안.
  - ✓ 요약: Self-supervised learning의 지속적인 발전에 비해, 소형 기기 등에서 사용가능한 알고리즘 및 성능의 개선이 부족하였음. 따라서, teacher model은 MoCo-V2로 학습시키고 knowledge distillation을 통해 student model(EfficientNet, MoblieNet 등)을 학습시켜 성능을 개선함.

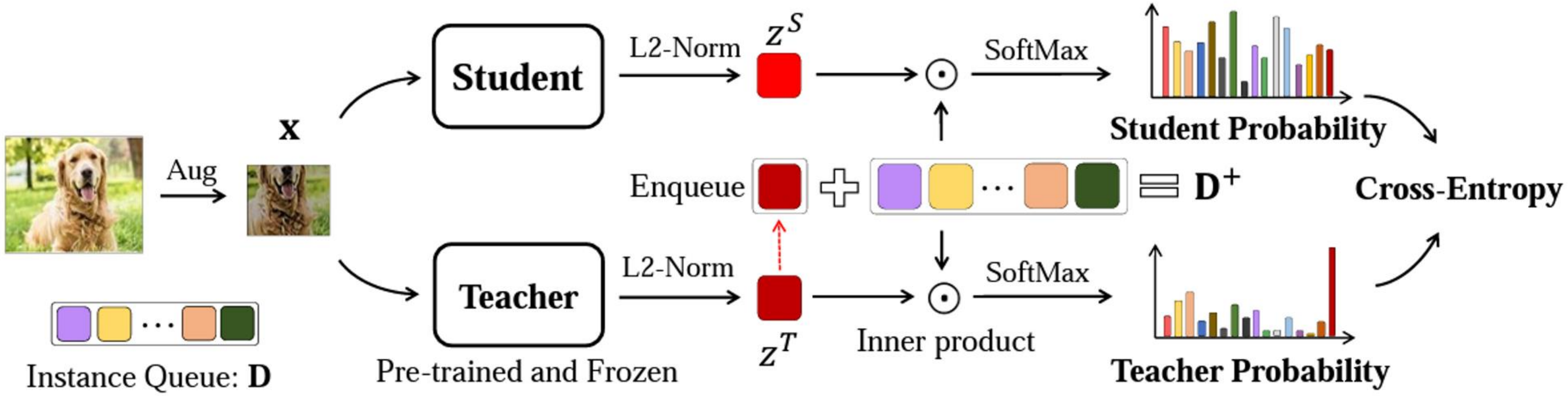
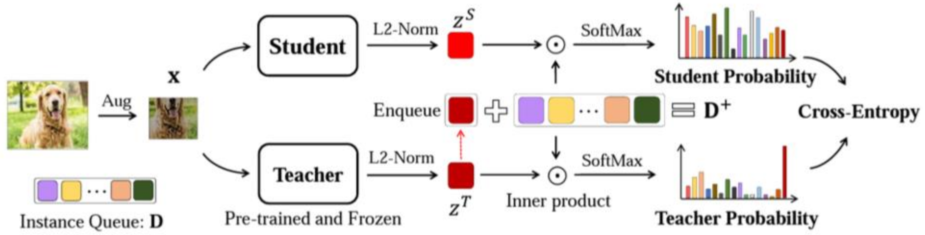


Figure 2: Illustration of our self-supervised distillation pipeline. The teacher encoder is pre-trained by *SSL* and kept frozen during the distillation. The student encoder is trained by minimizing the cross entropy of probabilities from teacher & student for an augmented view of an image, computed over a dynamically maintained queue.

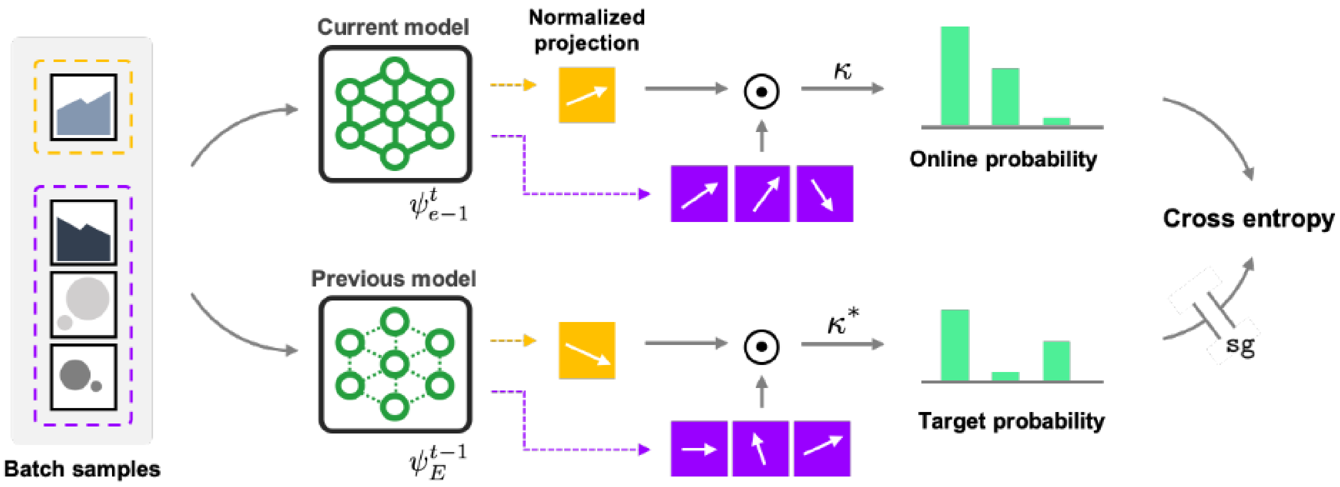
[1] Fang, Zhiyuan, et al. "SEED: Self-supervised Distillation For Visual Representation." *International Conference on Learning Representations*. 2021.

Paper Review: Co²L (ICCV 2021)



- Co²L: (2) Instance-wise Relation Distillation (IRD)

- Current model과 previous model에 batch(2N) 내의 데이터를 통과시킨 후 normalized projection 값을 사용하여 instance-wise similarity를 계산.
- Current representation과 past representation의 **discrepancy**를 IRD loss로 하여, <past → current> 방향으로 knowledge transfer가 가능하도록 유도.



(b) Instance-wise Relation Distillation Loss

$$\mathbf{p}\left(\tilde{\mathbf{x}}_i;\psi,\kappa\right)=\left[p_{i,1},\ldots,p_{i,i-1},p_{i,i+1},\ldots,p_{i,2N}\right]$$

[instance-wise similarity vector]

$$p_{i,j}=\frac{\exp\left(\mathbf{z}_i\cdot\mathbf{z}_j/\kappa\right)}{\sum_{k\neq i}^{2N}\exp\left(\mathbf{z}_i\cdot\mathbf{z}_k/\kappa\right)}$$

[instance-wise similarity]

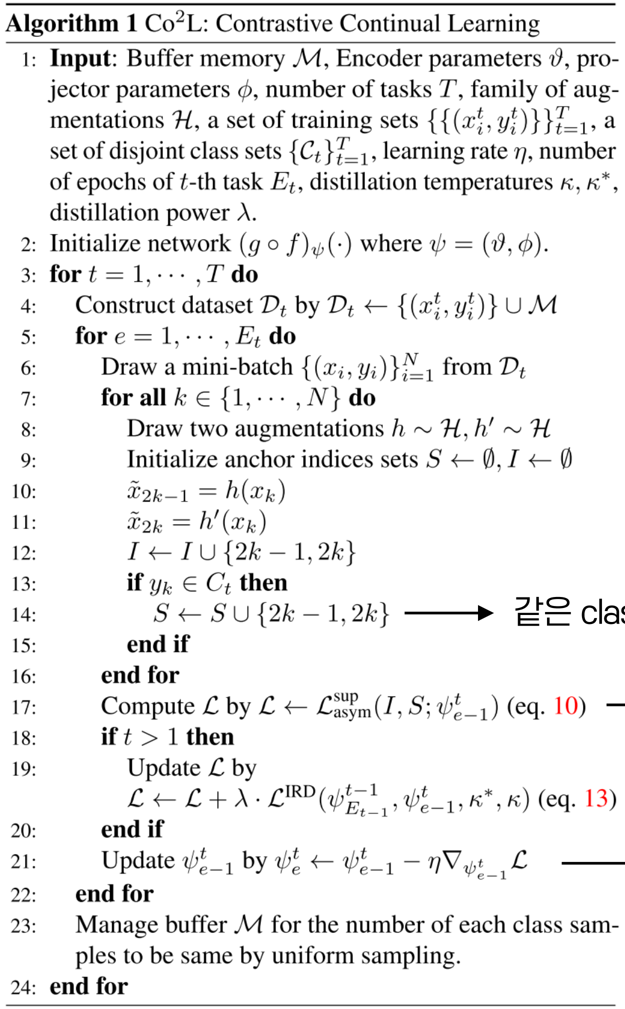
$$\mathcal{L}^{\mathrm{IRD}}=\sum_{i=1}^{2N}-\mathbf{p}\left(\tilde{\mathbf{x}}_i;\psi^{\mathrm{past}},\kappa^*\right)\cdot\log\mathbf{p}\left(\tilde{\mathbf{x}}_i;\psi,\kappa\right)$$

[Instance-wise relation loss]

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co²L (ICCV 2021)

- Co²L: Algorithm/Framework



```
Algorithm 1 Co²L: Contrastive Continual Learning
Input: Buffer memory $\mathcal{M}$, Encoder parameters $\vartheta$, projector parameters $\phi$, number of tasks $T$, family of augmentations $\mathcal{H}$, a set of training sets $\{\{(x_i^t, y_i^t)\}\}_{t=1}^T$, a set of disjoint class sets $\{\mathcal{C}_t\}_{t=1}^T$, learning rate $\eta$, number of epochs of $t$-th task $E_t$, distillation temperatures $\kappa, \kappa^*$, distillation power $\lambda$.
Initialize network $(g \circ f)_\psi(\cdot)$ where $\psi = (\vartheta, \phi)$.
for $t = 1, \cdots, T$ do
    Construct dataset $\mathcal{D}_t$ by $\mathcal{D}_t \leftarrow \{(x_i^t, y_i^t)\} \cup \mathcal{M}$
    for $e = 1, \cdots, E_t$ do
        Draw a mini-batch $\{(x_i, y_i)\}_{i=1}^N$ from $\mathcal{D}_t$
        for all $k \in \{1, \cdots, N\}$ do
            Draw two augmentations $h \sim \mathcal{H}, h' \sim \mathcal{H}$
            Initialize anchor indices sets $S \leftarrow \emptyset, I \leftarrow \emptyset$
            $\tilde{x}_{2k-1} = h(x_k)$
            $\tilde{x}_{2k} = h'(x_k)$
            $I \leftarrow I \cup \{2k-1, 2k\}$
            if $y_k \in \mathcal{C}_t$ then
                $S \leftarrow S \cup \{2k-1, 2k\}$
            end if
        end for
        Compute $\mathcal{L}$ by $\mathcal{L} \leftarrow \mathcal{L}^{\text{sup}}_{\text{asym}}(I, S; \psi^t_{e-1})$ (eq. 10)
        if $t > 1$ then
            Update $\mathcal{L}$ by
            $\mathcal{L} \leftarrow \mathcal{L} + \lambda \cdot \mathcal{L}^{\text{IRD}}(\psi^{t-1}_{E_{t-1}}, \psi^t_{e-1}, \kappa^*, \kappa)$ (eq. 13)
        end if
        Update $\psi^t_{e-1}$ by $\psi^t_e \leftarrow \psi^t_{e-1} - \eta \nabla_{\psi^t_{e-1}} \mathcal{L}$
    end for
    Manage buffer $\mathcal{M}$ for the number of each class samples to be same by uniform sampling.
end for
```

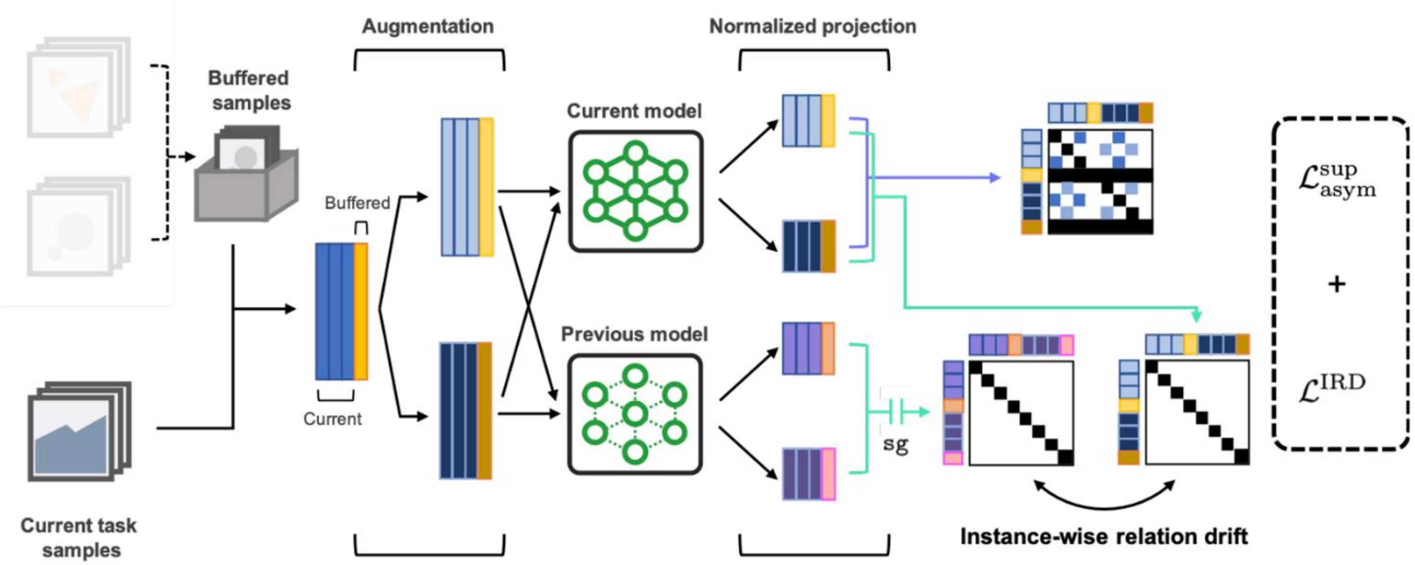
→ 같은 class 데이터의 augmentation를 S에 저장.

→ Epoch 마다 Loss 계산

→ 전 task(t-1)의 최종 모델($\psi^{t-1}_{E_{t-1}}$, $encoder + projection$)과 현재 task(t)의 전 epoch의 모델(ψ^t_{e-1})과 IRD loss 계산

→ Loss 에 따라 모델 업데이트

→ Memory buffer가 각 class의 비율이 동일하게 유지하도록 조정.

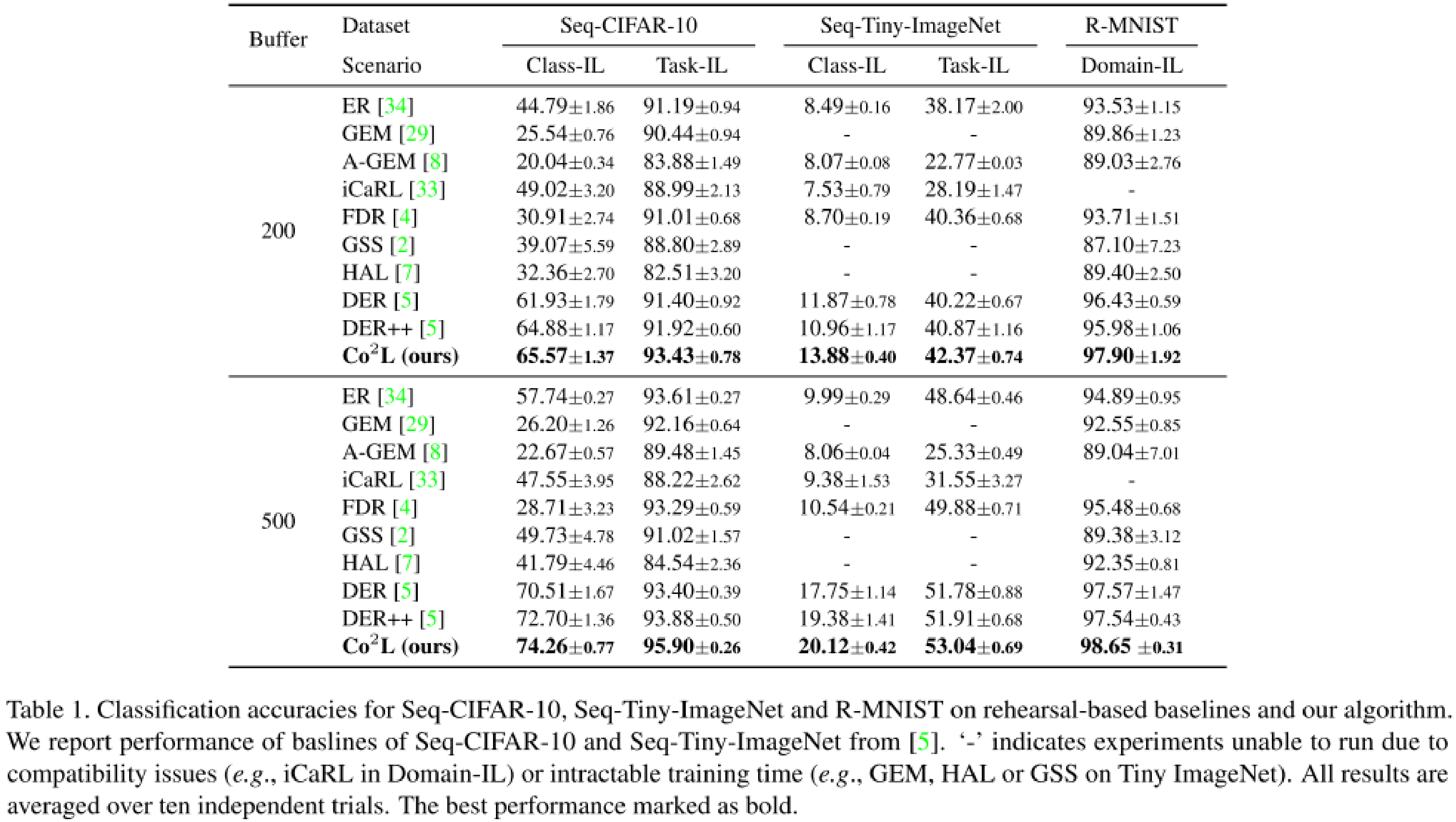


Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co²L (ICCV 2021)

- Experiment

- Rehearsal-based continual learning 방법들과 성능 비교.
- 학습된 모델을 통해 나온 representation을 사용하여 classifier를 학습시켜 classification 진행한 성능.



Buffer	Dataset	Seq-CIFAR-10		Seq-Tiny-ImageNet		R-MNIST
	Scenario	Class-IL	Task-IL	Class-IL	Task-IL	Domain-IL
200	ER [34]	44.79±1.86	91.19±0.94	8.49±0.16	38.17±2.00	93.53±1.15
	GEM [29]	25.54±0.76	90.44±0.94	-	-	89.86±1.23
	A-GEM [8]	20.04±0.34	83.88±1.49	8.07±0.08	22.77±0.03	89.03±2.76
	iCaRL [33]	49.02±3.20	88.99±2.13	7.53±0.79	28.19±1.47	-
	FDR [4]	30.91±2.74	91.01±0.68	8.70±0.19	40.36±0.68	93.71±1.51
	GSS [2]	39.07±5.59	88.80±2.89	-	-	87.10±7.23
	HAL [7]	32.36±2.70	82.51±3.20	-	-	89.40±2.50
	DER [5]	61.93±1.79	91.40±0.92	11.87±0.78	40.22±0.67	96.43±0.59
	DER++ [5]	64.88±1.17	91.92±0.60	10.96±1.17	40.87±1.16	95.98±1.06
	Co²L (ours)	**65.57±1.37**	**93.43±0.78**	**13.88±0.40**	**42.37±0.74**	**97.90±1.92**
500	ER [34]	57.74±0.27	93.61±0.27	9.99±0.29	48.64±0.46	94.89±0.95
	GEM [29]	26.20±1.26	92.16±0.64	-	-	92.55±0.85
	A-GEM [8]	22.67±0.57	89.48±1.45	8.06±0.04	25.33±0.49	89.04±7.01
	iCaRL [33]	47.55±3.95	88.22±2.62	9.38±1.53	31.55±3.27	-
	FDR [4]	28.71±3.23	93.29±0.59	10.54±0.21	49.88±0.71	95.48±0.68
	GSS [2]	49.73±4.78	91.02±1.57	-	-	89.38±3.12
	HAL [7]	41.79±4.46	84.54±2.36	-	-	92.35±0.81
	DER [5]	70.51±1.67	93.40±0.39	17.75±1.14	51.78±0.88	97.57±1.47
	DER++ [5]	72.70±1.36	93.88±0.50	19.38±1.41	51.91±0.68	97.54±0.43
	Co²L (ours)	**74.26±0.77**	**95.90±0.26**	**20.12±0.42**	**53.04±0.69**	**98.65 ±0.31**

Table 1. Classification accuracies for Seq-CIFAR-10, Seq-Tiny-ImageNet and R-MNIST on rehearsal-based baselines and our algorithm. We report performance of baslines of Seq-CIFAR-10 and Seq-Tiny-ImageNet from [5]. '-' indicates experiments unable to run due to compatibility issues (*e.g.*, iCaRL in Domain-IL) or intractable training time (*e.g.*, GEM, HAL or GSS on Tiny ImageNet). All results are averaged over ten independent trials. The best performance marked as bold.

[Class-IL, Task-IL]

Seq-CIFAR-10: 5 tasks

Seq-Tiny-ImageNet: 10 tasks

[Domain-IL]

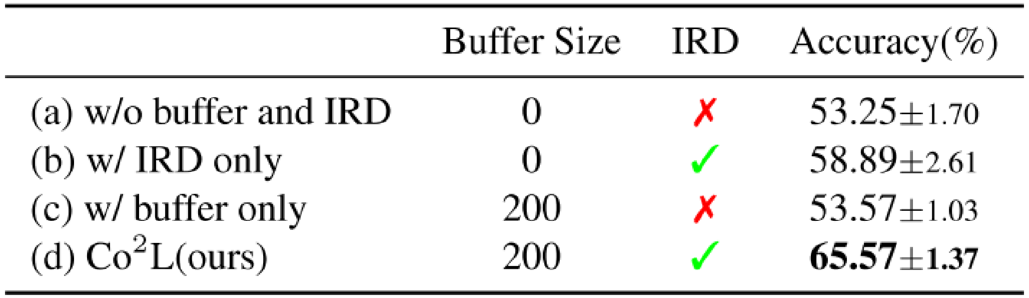
R-MNIST: 20 tasks

Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: Co^{2}L (ICCV 2021)

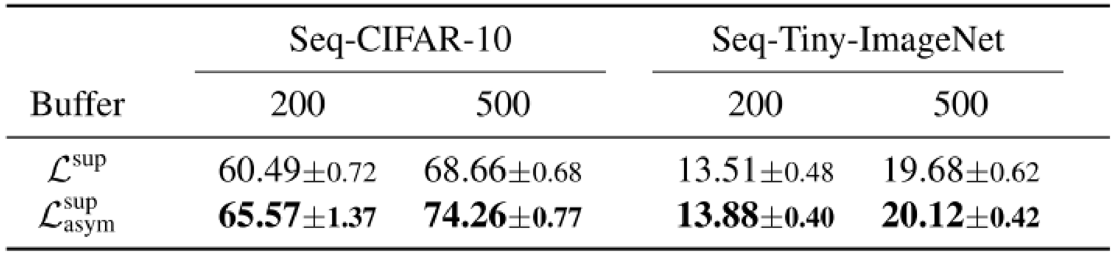
- Experiment (Ablation studies)

- IRD loss 방식의 우수성과 asymmetric SupCon loss의 우수성을 ablation studies를 통해 입증.



	Buffer Size	IRD	Accuracy(%)
(a) w/o buffer and IRD	0	✗	53.25±1.70
(b) w/ IRD only	0	✓	58.89±2.61
(c) w/ buffer only	200	✗	53.57±1.03
(d) Co^{2}L(ours)	200	✓	**65.57±1.37**

Table 2. Ablation study of Instance-wise Relation Distillation (IRD). We train our model on Seq-CIFAR-10 dataset under class-IL scenario (identical to the setup in Section 5.2) with ablated Co^{2}L. IRD brings significant gain with or without replay buffer. All results are averaged over ten independent trials.



	Seq-CIFAR-10		Seq-Tiny-ImageNet	
Buffer	200	500	200	500
$\mathcal{L}^{\text{sup}}$	60.49±0.72	68.66±0.68	13.51±0.48	19.68±0.62
$\mathcal{L}^{\text{sup}}_{\text{asym}}$	**65.57±1.37**	**74.26±0.77**	**13.88±0.40**	**20.12±0.42**

Table 3. The effectiveness of asymmetric SupCon loss ($\mathcal{L}^{\text{sup}}_{\text{asym}}$) versus the original SupCon loss ($\mathcal{L}^{\text{sup}}$), combining with the IRD loss. All results are averaged over ten independent trials.

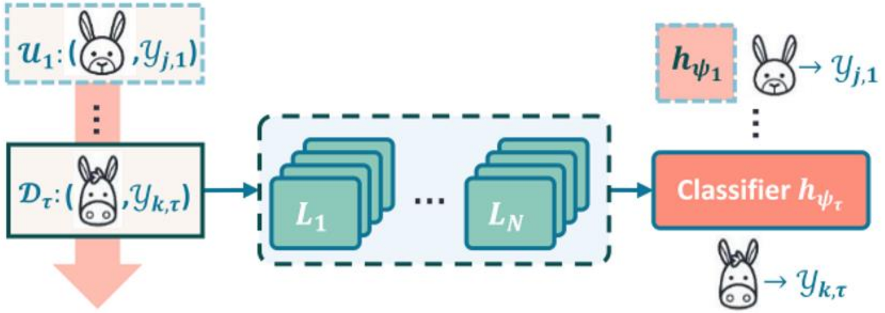
Cha, Hyuntak, Jaeho Lee, and Jinwoo Shin. "Co2l: Contrastive continual learning." *Proceedings of the IEEE/CVF International conference on computer vision*. 2021.

Paper Review: LUMP (ICLR 2022)

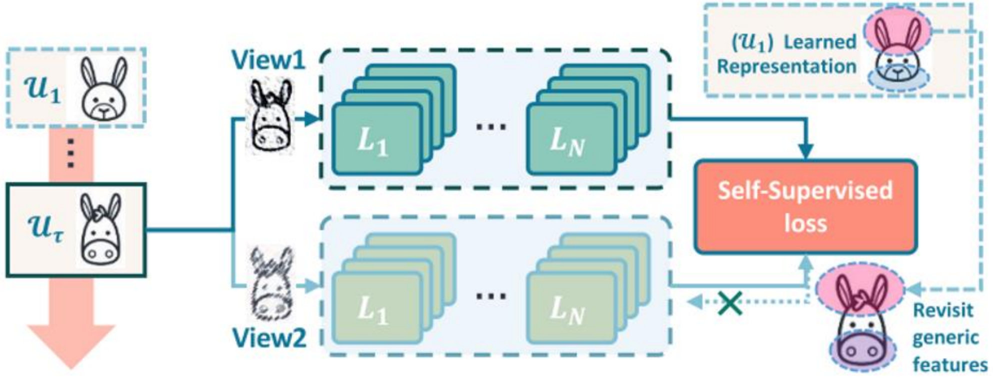
- Introduction

- Continual learning에 관한 대부분 연구들은 supervised learning paradigm에 집중하고 있는 상황.
- 하지만, 현실 세계에는 high-quality label을 얻는 것이 어려운 경우가 많으므로 unsupervised 환경에 대해 연구 필요.

- 따라서, 저자들은 **unsupervised continual learning (UCL)** 문제를 해결해야 한다고 주장.
- 또한, 해당 분야에 관한 연구는 거의 없으며 기존 연구(CURL[1], NeurIPS 2019)는 고해상도 데이터에 사용이 어렵고 모델 구조도 MLP를 사용하여 확장성이 낮다는 단점이 존재.



SUPERVISED CONTINUAL LEARNING (SCL)



UNSUPERVISED CONTINUAL LEARNING (UCL)

Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." *International Conference on Learning Representations*. 2022.

[1] Rao, Dushyant, et al. "Continual unsupervised representation learning." *Advances in Neural Information Processing Systems* 32 (2019).

Paper Review: LUMP (ICLR 2022)

- Related Work

- **Continual learning.**
  - ✓ Regularization approaches
  - ✓ Architectural approaches
  - ✓ Rehearsal approaches

- **Representation learning.**
  - ✓ Contrastive learning 기반으로 발전했으며 SimCLR (ICML 2020), MoCo (CVPR 2020) 등이 존재.
  - ✓ 큰 batch size와 negative sample pair이 필요하다는 기존 방법의 단점을 극복한 SimSiam (CVPR 2021) 등장.
  - ✓ Cross-correlation을 사용한 Barlow Twins (ICML 2021)가 새롭게 제안되었음.

Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." *International Conference on Learning Representations*. 2022.

Paper Review: LUMP (ICLR 2022)

- Unsupervised Continual Learning

- 기존 CL 방법론을 간단히 변형하여 UCL 상황에서 실험할 수 있도록 세팅.
  - ✓ Regularization-based method: SI (Synaptic Intelligence) ICML 2017
  - ✓ Architectural-based method: PNN (Progressive Neural Network) Arxiv 2016
  - ✓ Rehearsal-based method: DER (Dark Experience Replay) NeurIPS 2020
- DER은 distillation loss가 softmax 형식이므로 UCL 상황으로 변환하는 과정 소개.

$$\mathcal{L}^{\text{DER}}_{\text{SCL}} = \mathcal{L}^{\text{FINETUNE}}_{\text{SCL}} + \alpha \cdot \mathbb{E}_{(x,p)\sim\mathcal{M}} \boxed{\left[\|\text{softmax}(p) - \text{softmax}(h_\psi(x_{i,\tau}))\|_2^2 \right]}$$

$$\mathcal{L}^{\text{DER}}_{\text{UCL}} = \mathcal{L}^{\text{FINETUNE}}_{\text{UCL}} + \alpha \cdot \mathbb{E}_{(x)\sim\mathcal{M}} \boxed{\left[\|f_{\Theta_\tau}(x) - f_\Theta(x_{i,\tau})\|_2^2 \right]}$$

Softmax를 구할 수 없기 때문에
feature 간의 거리로 변경.

$$\mathcal{L}^{\text{FINETUNE}}_{\text{UCL}} = \frac{1}{2}D(p^1_{i,\tau}, \texttt{stopgrad}(z^2_{i,\tau})) + \frac{1}{2}D(p^2_{i,\tau}, \texttt{stopgrad}(z^1_{i,\tau})),$$

$$\text{where } D(p^1_{i,\tau}, z^2_{i,\tau}) = -\frac{p^1_{i,\tau}}{\|p^2_{i,\tau}\|_2} \cdot \frac{z^2_{i,\tau}}{\|z^2_{i,\tau}\|_2},$$

[SimSiam loss]

$$\mathcal{L}^{\text{FINETUNE}}_{\text{UCL}} = \sum_i (1 - \mathcal{C}_{ii})^2 + \lambda \cdot \sum_i \sum_{j \neq i} {\mathcal{C}_{ij}}^2,$$

$$\text{where } \mathcal{C}_{ij} = \frac{\sum_{\mathcal{B}} z^1_{\mathcal{B},i} z^2_{\mathcal{B},j}}{\sqrt{\sum_{\mathcal{B}} (z^1_{\mathcal{B},i})^2}\sqrt{\sum_{\mathcal{B}} (z^2_{\mathcal{B},j})^2}}$$

[Barlow Twins loss]

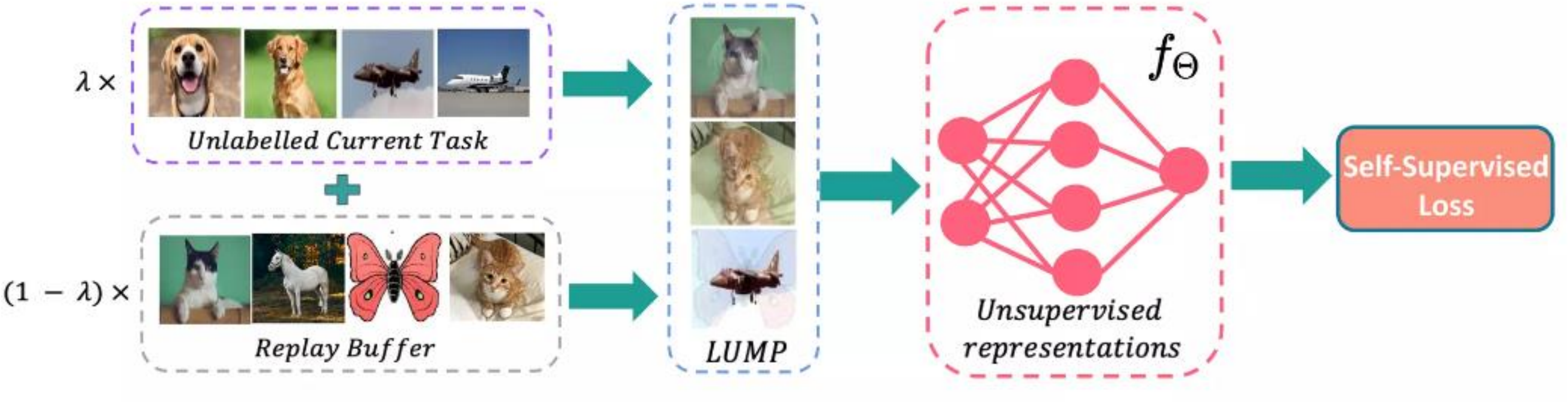
Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." *International conference on machine learning*. PMLR, 2017.
Rusu, Andrei A., et al. "Progressive neural networks." *arXiv preprint arXiv*:1606.04671 (2016).
Buzzega, Pietro, et al. "Dark experience for general continual learning: a strong, simple baseline." *Advances in neural information processing systems* 33 (2020): 15920-15930.

Paper Review: LUMP (ICLR 2022)

- Lifelong Unsupervised Mixup (LUMP)

- 본 논문을 통해 제안된 LUMP은, mixup을 통해 만들어진 데이터를 가지고 model을 학습시키는 방법.
- Past data와 current data의 mixup을 통해 만들었으므로, 유용한 attributes에 revisiting해서 robust한 representation을 얻을 수 있다고 주장.

- 저자들이 작성한 슬라이드 그림 참고.



Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." *International Conference on Learning Representations*. 2022.
https://www.slideshare.net/MLAI2/representational-continuity-for-unsupervised-continual-learning

Paper Review: LUMP (ICLR 2022)

- Experiment

- 학습된 representation을 동일하게 KNN classifier에 학습시켜 각 실험세팅의 성능을 확인.

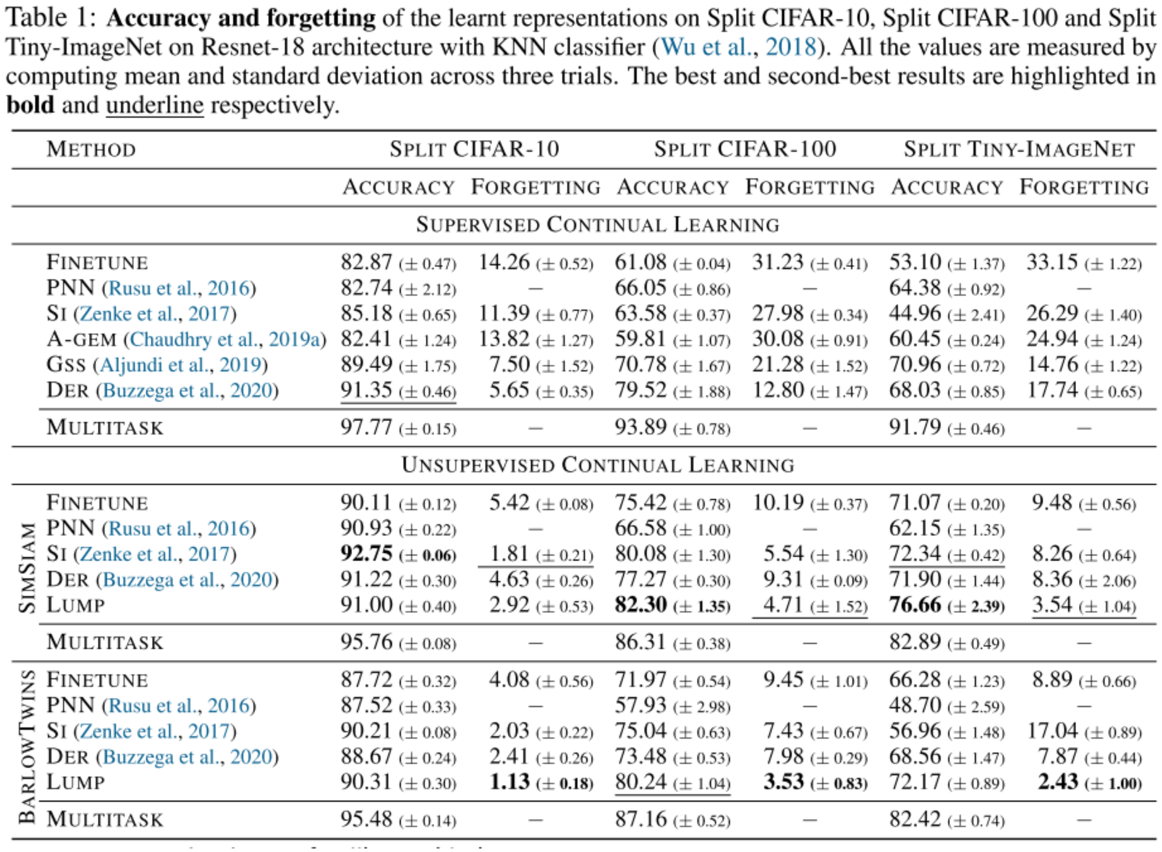


Table 1: **Accuracy and forgetting** of the learnt representations on Split CIFAR-10, Split CIFAR-100 and Split Tiny-ImageNet on Resnet-18 architecture with KNN classifier (Wu et al., 2018). All the values are measured by computing mean and standard deviation across three trials. The best and second-best results are highlighted in **bold** and underline respectively.

	Method	Split CIFAR-10		Split CIFAR-100		Split Tiny-ImageNet	
		Accuracy	Forgetting	Accuracy	Forgetting	Accuracy	Forgetting
	Supervised Continual Learning						
	Finetune	82.87 (± 0.47)	14.26 (± 0.52)	61.08 (± 0.04)	31.23 (± 0.41)	53.10 (± 1.37)	33.15 (± 1.22)
	PNN (Rusu et al., 2016)	82.74 (± 2.12)	–	66.05 (± 0.86)	–	64.38 (± 0.92)	–
	SI (Zenke et al., 2017)	85.18 (± 0.65)	11.39 (± 0.77)	63.58 (± 0.37)	27.98 (± 0.34)	44.96 (± 2.41)	26.29 (± 1.40)
	A-GEM (Chaudhry et al., 2019a)	82.41 (± 1.24)	13.82 (± 1.27)	59.81 (± 1.07)	30.08 (± 0.91)	60.45 (± 0.24)	24.94 (± 1.24)
	GSS (Aljundi et al., 2019)	89.49 (± 1.75)	7.50 (± 1.52)	70.78 (± 1.67)	21.28 (± 1.52)	70.96 (± 0.72)	14.76 (± 1.22)
	DER (Buzzega et al., 2020)	91.35 (± 0.46)	5.65 (± 0.35)	79.52 (± 1.88)	12.80 (± 1.47)	68.03 (± 0.85)	17.74 (± 0.65)
	Multitask	97.77 (± 0.15)	–	93.89 (± 0.78)	–	91.79 (± 0.46)	–
	Unsupervised Continual Learning						
SimSiam	Finetune	90.11 (± 0.12)	5.42 (± 0.08)	75.42 (± 0.78)	10.19 (± 0.37)	71.07 (± 0.20)	9.48 (± 0.56)
	PNN (Rusu et al., 2016)	90.93 (± 0.22)	–	66.58 (± 1.00)	–	62.15 (± 1.35)	–
	SI (Zenke et al., 2017)	**92.75 (± 0.06)**	1.81 (± 0.21)	80.08 (± 1.30)	5.54 (± 1.30)	72.34 (± 0.42)	8.26 (± 0.64)
	DER (Buzzega et al., 2020)	91.22 (± 0.30)	4.63 (± 0.26)	77.27 (± 0.30)	9.31 (± 0.09)	71.90 (± 1.44)	8.36 (± 2.06)
	LUMP	91.00 (± 0.40)	2.92 (± 0.53)	**82.30 (± 1.35)**	4.71 (± 1.52)	**76.66 (± 2.39)**	3.54 (± 1.04)
	Multitask	95.76 (± 0.08)	–	86.31 (± 0.38)	–	82.89 (± 0.49)	–
BarlowTwins	Finetune	87.72 (± 0.32)	4.08 (± 0.56)	71.97 (± 0.54)	9.45 (± 1.01)	66.28 (± 1.23)	8.89 (± 0.66)
	PNN (Rusu et al., 2016)	87.52 (± 0.33)	–	57.93 (± 2.98)	–	48.70 (± 2.59)	–
	SI (Zenke et al., 2017)	90.21 (± 0.08)	2.03 (± 0.22)	75.04 (± 0.63)	7.43 (± 0.67)	56.96 (± 1.48)	17.04 (± 0.89)
	DER (Buzzega et al., 2020)	88.67 (± 0.24)	2.41 (± 0.26)	73.48 (± 0.53)	7.98 (± 0.29)	68.56 (± 1.47)	7.87 (± 0.44)
	LUMP	90.31 (± 0.30)	**1.13 (± 0.18)**	80.24 (± 1.04)	**3.53 (± 0.83)**	72.17 (± 0.89)	**2.43 (± 1.00)**
	Multitask	95.48 (± 0.14)	–	87.16 (± 0.52)	–	82.42 (± 0.74)	–

1. Supervised CL 보다 unsupervised CL이 대부분의 실험에서 성능이 우수함을 확인 (continual learning에는 label 정보보다 데이터의 본질적인 정보가 더 중요)

2. 제안한 LUMP 방법론은 Split CIFAR-10을 제외하고, Split CIFAR-100과 Split Tiny-ImageNet에서 가장 좋은 정확도 성능을 보여줌.

3. 또한, forgetting 부분에서 특히 LUMP의 성능이 우수함을 확인. 이는 mixup을 통해 과거의 중요 attributes에 revisiting 할 수 있었기 때문으로 보임.

Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." *International Conference on Learning Representations*. 2022.

Paper Review: LUMP (ICLR 2022)

- Experiment

- CIFAR-100을 20tasks를 순차적으로 학습 완료한 후에 모델을 가지고 apple(t_0)과 raccoon(t_{13}) 의 feature map을 시각화.
- SCL은 UCL에 비해 class 정보가 적고, 일관된 패턴이 잘 보이지 않음.

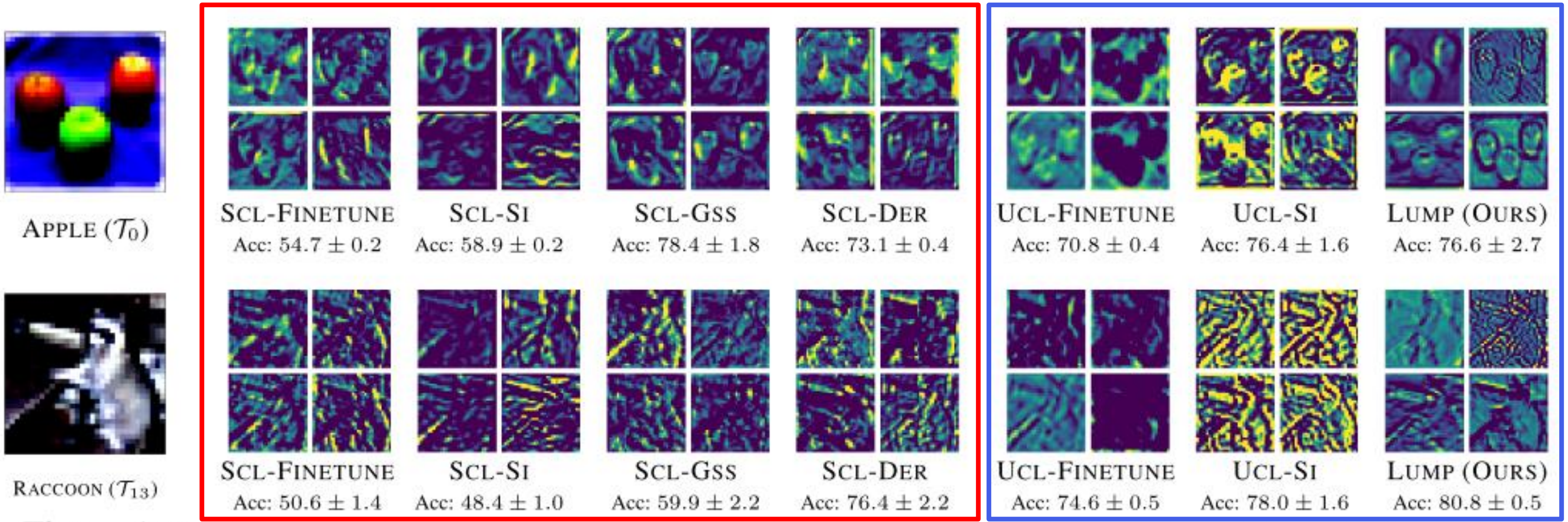


Figure 4: **Visualization of feature maps** for the second block representations learnt by SCL and UCL strategies (with Simsiam) for ResNet-18 architecture after the completion of CL for Split CIFAR-100 dataset ($n = 20$).

Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." *International Conference on Learning Representations*. 2022.

Paper Review: CaSSLe (CVPR 2022)

- Introduction

- Self-supervised learning(SSL)은 지난 몇 년간 매우 인기있는 방법론 중에 하나.
- 앞선 LUMP 논문과 마찬가지로, continual learning 상황 속에서 self-supervised learning을 사용한 경우가 적음을 강조.

- Related Work

- <u>Self-supervised learning.</u>
  - ➢ SimCLR, MoCo, SimSiam, SwAV, DeepCluster, DINO, BarlowTwins, VICReg 등
- <u>Continual learning.</u>
  - ➢ 대부분의 CL method들은 supervision(label)을 사용하는 방식으로 발전.
  - ➢ Unsupervised continual learning 관한 연구는 많지 않으며, 최근 연구 두 개에 대해서도 단점이 존재.
    - ① Continual Contrastive Learning[1] (ICME 2022): Co^{2}L을 unsupervised setting으로 확장했으나, contrastive learning에 국한되어 다른 SSL 방식으로 확장이 어려움.
    - ② LUMP (ICLR 2022): experiments에 사용된 데이터가 작으며, class-IL 상황만을 실험했고 SSL 방법도 두 가지(SimSiam, Barlow Twins)만을 사용했다는 한계 존재.

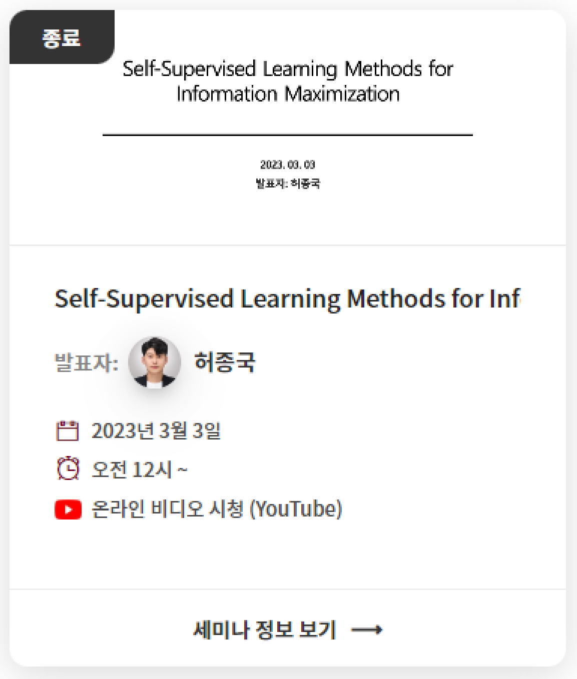
[1] Lin, Zhiwei, Yongtao Wang, and Hongxiang Lin. "Continual Contrastive Learning for Image Classification." *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022.

Paper Review: CaSSLe (CVPR 2022)

- Introduction

- Self-supervised learning(SSL)은 지난 몇 년간 매우 인기있는 방법론 중에 하나.
- 앞선 LUMP 논문과 마찬가지로, continual learning 상황 속에서 self-supervised learning을 사용한 경우가 적음을 강조.

- Related Work



종료

Self-Supervised Learning Methods for Information Maximization

2023. 03. 03
발표자: 허종국

Self-Supervised Learning Methods for Inf

발표자: 허종국

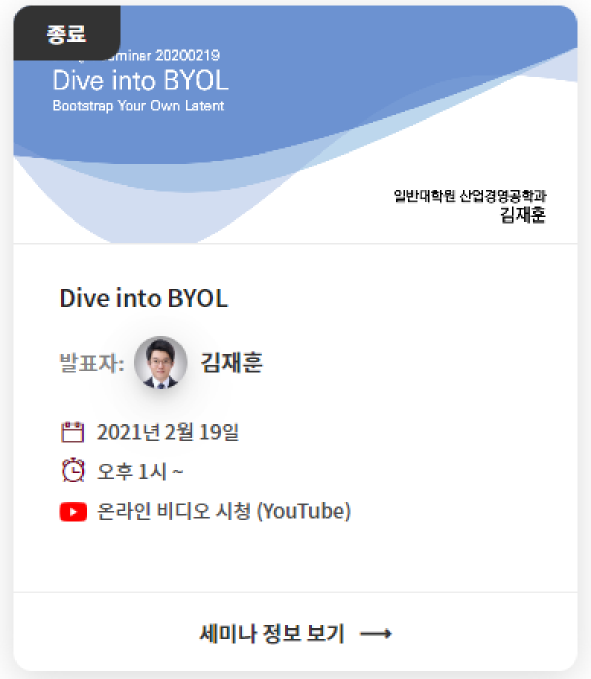
2023년 3월 3일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

http://dmqm.korea.ac.kr/activity/seminar/399



종료

minar 20200219

Dive into BYOL

Bootstrap Your Own Latent

일반대학원 산업경영공학과
김재훈

Dive into BYOL

발표자: 김재훈

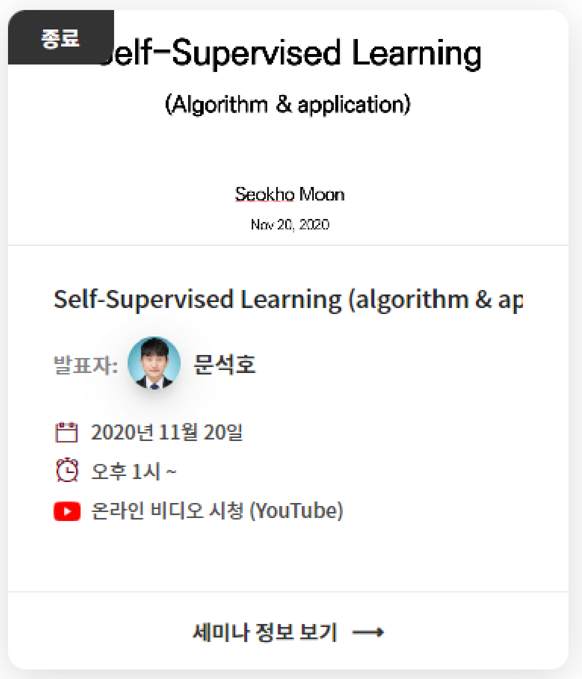
2021년 2월 19일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

http://dmqm.korea.ac.kr/activity/seminar/310



종료

elf-Supervised Learning

(Algorithm & application)

Seokho Moon

Nov 20, 2020

Self-Supervised Learning (algorithm & ap

발표자: 문석호

2020년 11월 20일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

http://dmqm.korea.ac.kr/activity/seminar/302

Paper Review: CaSSLe (CVPR 2022)

- Continual Self-Supervised Learning (CSSL)

- 여러 continual learning 방식에서 사용되는 개념 설명. (Co^{2}L, LUMP 논문 recap.)

$$\mathcal{L} = \underbrace{\mathcal{L}^{\text{sup}}_{\text{asym}}}_{\text{(1) learning}} + \underbrace{\lambda \cdot \mathcal{L}^{\text{IRD}}}_{\text{(2) preserving}}$$

$$\mathcal{L}^{\text{DER}}_{\text{UCL}} = \mathcal{L}^{\text{FINETUNE}}_{\text{UCL}} + \alpha \cdot \mathbb{E}_{(x)\sim\mathcal{M}}\left[\, \|f_{\Theta_\tau}(x) - f_{\Theta}(x_{i,\tau})\|_2^2 \,\right]$$

Current task에 맞게
학습시키는 loss

Past/current task 사이의
knowledge를 전달하는 방식

Fini, Enrico, et al. "Self-supervised models are continual learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Paper Review: CaSSLe (CVPR 2022)

- Continual Self-Supervised Learning (CSSL)

- z: $after\ current\ network$, $\bar{z}$: $after\ frozen\ network$

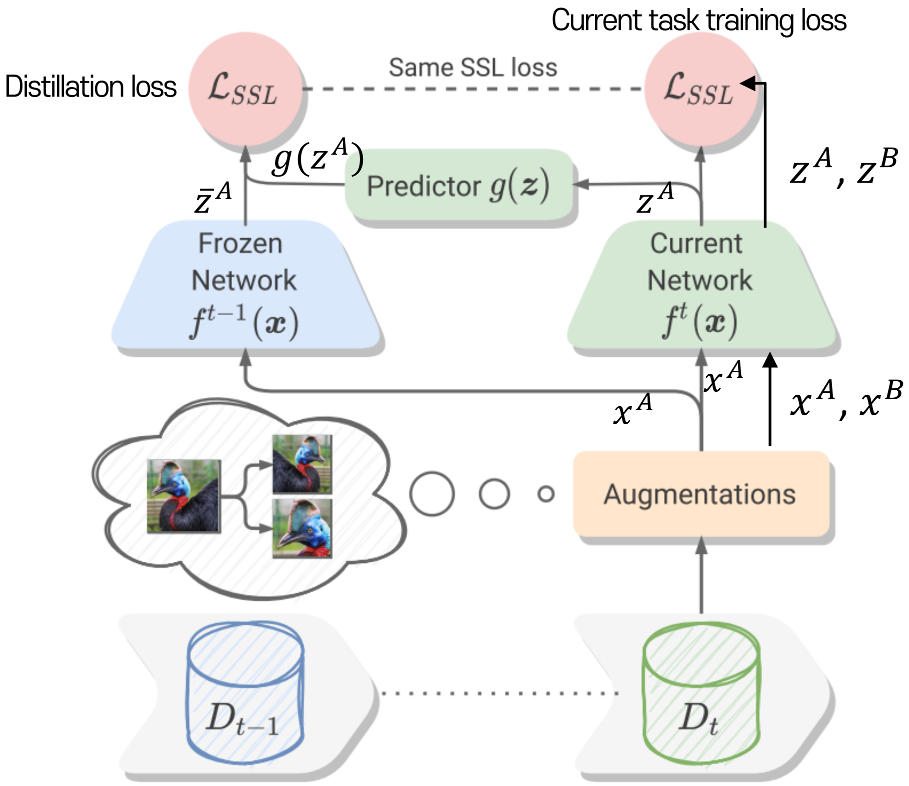


Figure 2. Overview of the CaSSLe framework.

Current task에 맞게 학습시키는 loss (SSL 방식에 의존)

Past/current task 사이의 knowledge를 전달하는 방식 (distillation loss 미정 상태)

$$\mathcal{L} = \boxed{\mathcal{L}_{SSL}(\boldsymbol{z}^A, \boldsymbol{z}^B)} + \boxed{\mathcal{L}_D(\boldsymbol{z}^A, \bar{\boldsymbol{z}}^A)}$$

∗ A, B는 하나의 image에서 각각 augmentation 된 데이터

$$= \mathcal{L}_{SSL}(\boldsymbol{z}^A, \boldsymbol{z}^B) + \boxed{\mathcal{L}_{SSL}(g(\boldsymbol{z}^A), \bar{\boldsymbol{z}}^A)}$$

Network를 학습하는 SSL loss를 distillation loss로 사용하자는 것이 논문 아이디어.

Fini, Enrico, et al. "Self-supervised models are continual learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Paper Review: CaSSLe (CVPR 2022)

- Continual Self-Supervised Learning (CSSL)

- 왜 supervised CL에서 주로 쓰는 distillation loss를 쓰면 안되는지에 대한 저자들의 주장.
    - ✓ 기존 Supervised CL 방식에서의 distillation loss처럼 기존 지식을 보존하려는 형태로 학습한다면, SSL 방법 특성상 모델이 current task에 대한 지식 습득이 방해될 수 있음.
    - ✓ Distillation methods(e.g., POD, Less-Forget)에서 사용한 regularization(l2-norm)과 SSL 방식(e.g., Barlow Twins)의 regularization이 맞지 않으면 성능 저하가 있을 수 있음.

- 저자들이 Predictor g를 사용하는 이유.
    - ✓ 바로 $z, \bar{z}$ 값이 유사해지도록 학습하면 current model이 제대로 학습되지 않을 수 있음.

- 저자들이 SSL loss를 본 논문의 distillation loss로 사용하는 이유.
    - ✓ g(z)가 augmentation에 대해 invariant 하길 원하므로

Fini, Enrico, et al. "Self-supervised models are continual learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Paper Review: CaSSLe (CVPR 2022)

- Experiment

- 실제 다양한 실험환경(class-IL, data-IL, domain-IL) 및 데이터셋(CIFAR100, ImageNet100, DomainNet)에서 제안한 아이디어의 타당성 입증. (논문 참조)

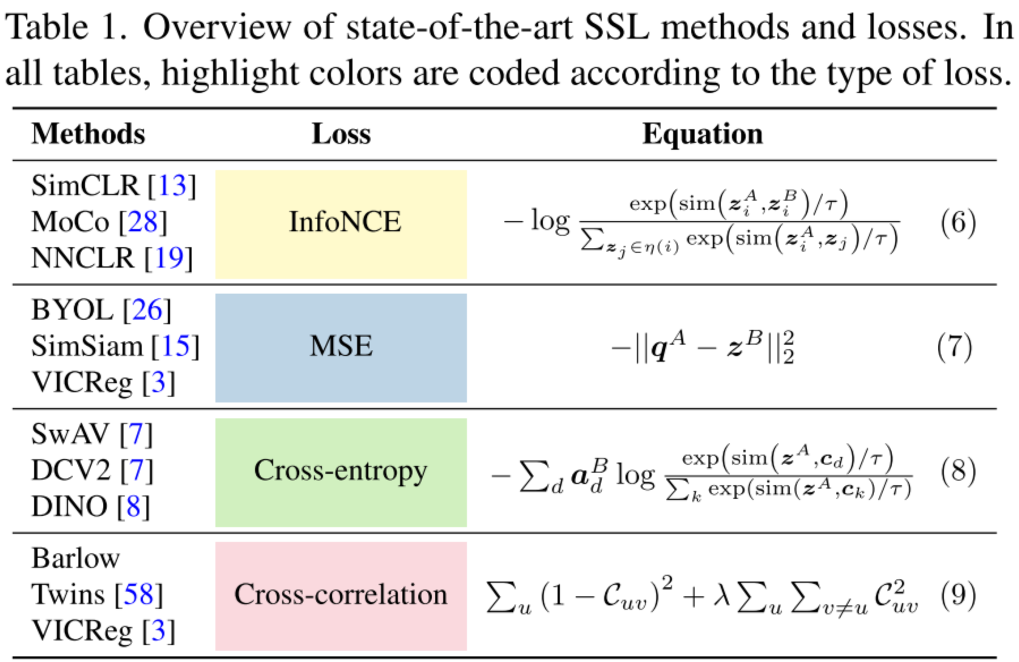


Table 1. Overview of state-of-the-art SSL methods and losses. In all tables, highlight colors are coded according to the type of loss.

Methods	Loss	Equation
SimCLR [13] MoCo [28] NNCLR [19]	InfoNCE	$-\log \frac{\exp\left(\text{sim}\left(\boldsymbol{z}_i^A, \boldsymbol{z}_i^B\right)/\tau\right)}{\sum_{\boldsymbol{z}_j \in \eta(i)} \exp\left(\text{sim}\left(\boldsymbol{z}_i^A, \boldsymbol{z}_j\right)/\tau\right)}$ (6)
BYOL [26] SimSiam [15] VICReg [3]	MSE	$-\lvert\lvert \boldsymbol{q}^A - \boldsymbol{z}^B \rvert\rvert_2^2$ (7)
SwAV [7] DCV2 [7] DINO [8]	Cross-entropy	$-\sum_d \boldsymbol{a}_d^B \log \frac{\exp\left(\text{sim}\left(\boldsymbol{z}^A, \boldsymbol{c}_d\right)/\tau\right)}{\sum_k \exp(\text{sim}(\boldsymbol{z}^A, \boldsymbol{c}_k)/\tau)}$ (8)
Barlow Twins [58] VICReg [3]	Cross-correlation	$\sum_u (1 - \mathcal{C}_{uv})^2 + \lambda \sum_u \sum_{v \neq u} \mathcal{C}_{uv}^2$ (9)

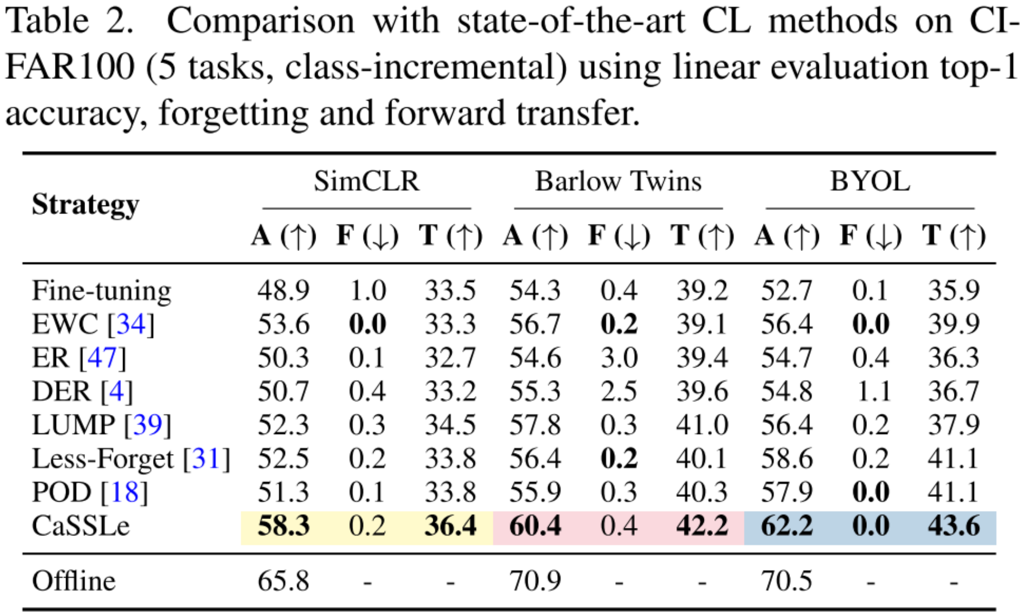


Table 2. Comparison with state-of-the-art CL methods on CIFAR100 (5 tasks, class-incremental) using linear evaluation top-1 accuracy, forgetting and forward transfer.

Strategy	SimCLR			Barlow Twins			BYOL		
	A (↑)	F (↓)	T (↑)	A (↑)	F (↓)	T (↑)	A (↑)	F (↓)	T (↑)
Fine-tuning	48.9	1.0	33.5	54.3	0.4	39.2	52.7	0.1	35.9
EWC [34]	53.6	**0.0**	33.3	56.7	**0.2**	39.1	56.4	**0.0**	39.9
ER [47]	50.3	0.1	32.7	54.6	3.0	39.4	54.7	0.4	36.3
DER [4]	50.7	0.4	33.2	55.3	2.5	39.6	54.8	1.1	36.7
LUMP [39]	52.3	0.3	34.5	57.8	0.3	41.0	56.4	0.2	37.9
Less-Forget [31]	52.5	0.2	33.8	56.4	**0.2**	40.1	58.6	0.2	41.1
POD [18]	51.3	0.1	33.8	55.9	0.3	40.3	57.9	**0.0**	41.1
CaSSLe	**58.3**	0.2	**36.4**	**60.4**	0.4	**42.2**	**62.2**	**0.0**	**43.6**
Offline	65.8	-	-	70.9	-	-	70.5	-	-

Fini, Enrico, et al. "Self-supervised models are continual learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Paper Review: CaSSLe (CVPR 2022)

- Experiment

- 실제 다양한 실험환경(class-IL, data-IL, domain-IL) 및 데이터셋(CIFAR100, ImageNet100, DomainNet)에서 제안한 아이디어의 타당성 입증. (논문 참조)

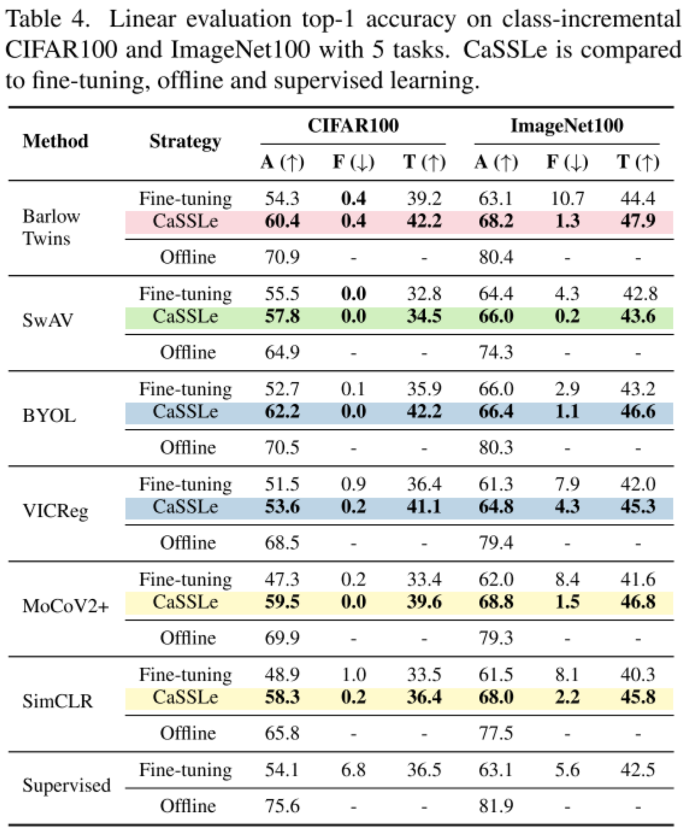


Table 4. Linear evaluation top-1 accuracy on class-incremental CIFAR100 and ImageNet100 with 5 tasks. CaSSLe is compared to fine-tuning, offline and supervised learning.

Method	Strategy	CIFAR100			ImageNet100		
		A (↑)	F (↓)	T (↑)	A (↑)	F (↓)	T (↑)
Barlow Twins	Fine-tuning	54.3	**0.4**	39.2	63.1	10.7	44.4
	CaSSLe	**60.4**	**0.4**	**42.2**	**68.2**	**1.3**	**47.9**
	Offline	70.9	-	-	80.4	-	-
SwAV	Fine-tuning	55.5	**0.0**	32.8	64.4	4.3	42.8
	CaSSLe	**57.8**	**0.0**	**34.5**	**66.0**	**0.2**	**43.6**
	Offline	64.9	-	-	74.3	-	-
BYOL	Fine-tuning	52.7	0.1	35.9	66.0	2.9	43.2
	CaSSLe	**62.2**	**0.0**	**42.2**	**66.4**	**1.1**	**46.6**
	Offline	70.5	-	-	80.3	-	-
VICReg	Fine-tuning	51.5	0.9	36.4	61.3	7.9	42.0
	CaSSLe	**53.6**	**0.2**	**41.1**	**64.8**	**4.3**	**45.3**
	Offline	68.5	-	-	79.4	-	-
MoCoV2+	Fine-tuning	47.3	0.2	33.4	62.0	8.4	41.6
	CaSSLe	**59.5**	**0.0**	**39.6**	**68.8**	**1.5**	**46.8**
	Offline	69.9	-	-	79.3	-	-
SimCLR	Fine-tuning	48.9	1.0	33.5	61.5	8.1	40.3
	CaSSLe	**58.3**	**0.2**	**36.4**	**68.0**	**2.2**	**45.8**
	Offline	65.8	-	-	77.5	-	-
Supervised	Fine-tuning	54.1	6.8	36.5	63.1	5.6	42.5
	Offline	75.6	-	-	81.9	-	-

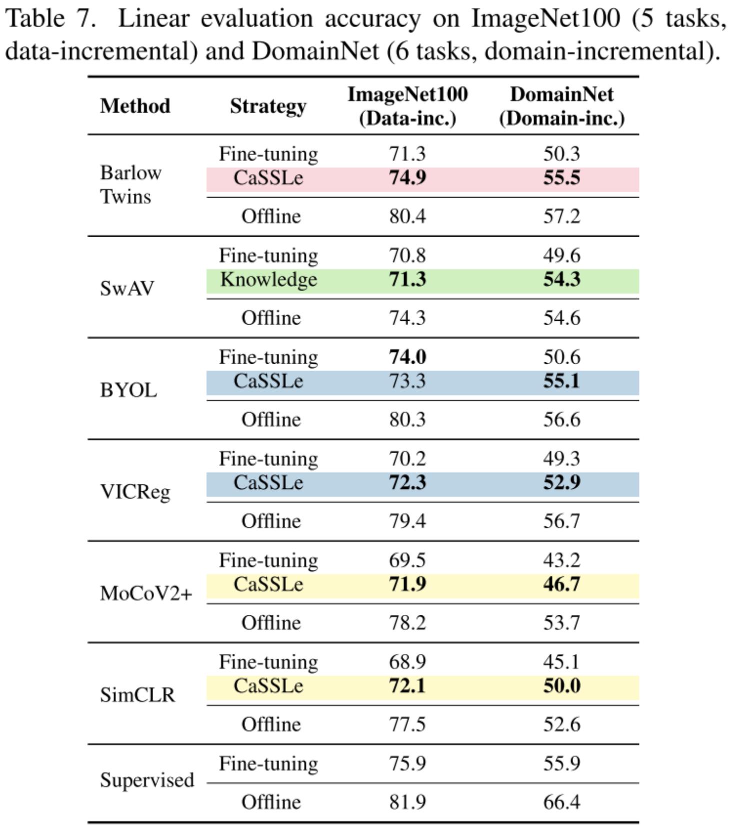


Table 7. Linear evaluation accuracy on ImageNet100 (5 tasks, data-incremental) and DomainNet (6 tasks, domain-incremental).

Method	Strategy	ImageNet100 (Data-inc.)	DomainNet (Domain-inc.)
Barlow Twins	Fine-tuning	71.3	50.3
	CaSSLe	**74.9**	**55.5**
	Offline	80.4	57.2
SwAV	Fine-tuning	70.8	49.6
	Knowledge	**71.3**	**54.3**
	Offline	74.3	54.6
BYOL	Fine-tuning	**74.0**	50.6
	CaSSLe	73.3	**55.1**
	Offline	80.3	56.6
VICReg	Fine-tuning	70.2	49.3
	CaSSLe	**72.3**	**52.9**
	Offline	79.4	56.7
MoCoV2+	Fine-tuning	69.5	43.2
	CaSSLe	**71.9**	**46.7**
	Offline	78.2	53.7
SimCLR	Fine-tuning	68.9	45.1
	CaSSLe	**72.1**	**50.0**
	Offline	77.5	52.6
Supervised	Fine-tuning	75.9	55.9
	Offline	81.9	66.4

Fini, Enrico, et al. "Self-supervised models are continual learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Conclusion

- Continual learning은 현실 세계에 좀 더 부합하는 개념이기 때문에, 최근 많은 연구들이 진행 중.

- 이번 세미나에서는 representation-based 접근 방법론에 대해 살펴봄.

- Representation 기반 연구들은 대부분 self-supervised learning에서 나온 아이디어를 많이 차용하였고, label이 없는 상황에서 어떤 식으로 knowledge preserve/transfer 할 지에 대한 아이디어가 주를 이룸.

- 최근 pre-trained model과 prompt learning을 사용하여 SOTA를 달성하고 있는 연구들도 추후 소개 예정.
  → L2P (CVPR 2022), S-Prompts (NeurIPS 2022), DualPrompt (ECCV 2022), CODA-Prompt (CVPR 2023)

Thank you

Seokho Moon

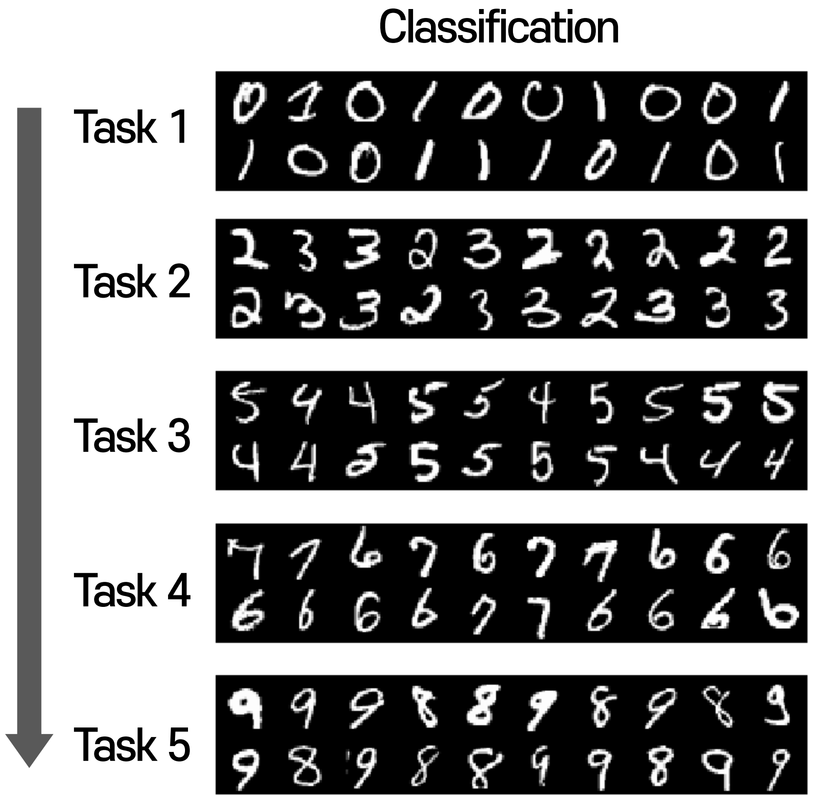
danny232@korea.ac.kr

School of Industrial and Management Engineering, Korea University

Appendix

Example of Tasks

- MNIST 데이터셋을 예시로 사용하여 2 class classification을 하는 5개의 task를 도식화
- 각 task는 1→2→3→4→5 의 순서를 가지고 **순차적**으로 학습이 진행되며, network는 1개로 구성



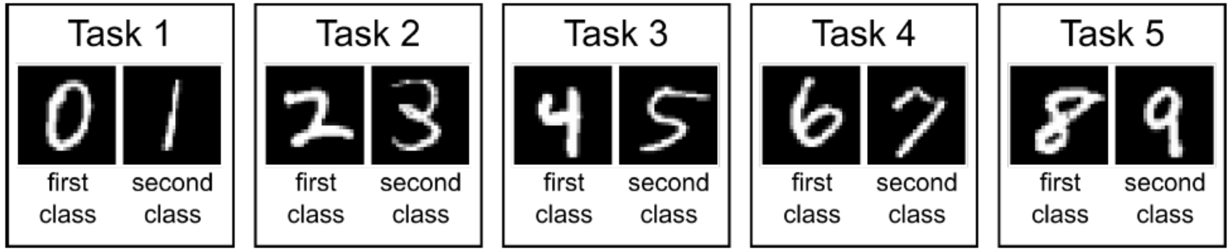


Figure 1: Schematic of split MNIST task protocol.

LeCun, Yann. "The MNIST database of handwritten digits."http://yann. lecun. com/exdb/mnist/(1998).
Van de Ven, G. M., & Tolias, A. S. “Three scenarios for continual learning.” (2018). Neural Information Processing Systems Workshops, 2018.

Three continual learning scenarios

- Continual learning은 크게 3가지의 scenario를 가지며, 각각은 task incremental learning(**Task-IL**), domain incremental learning(**Domain-IL**), and class incremental learning(**Class-IL**) 으로 되어 있음

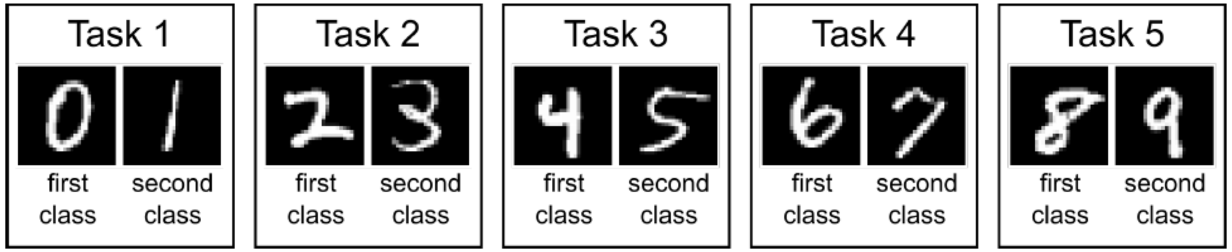
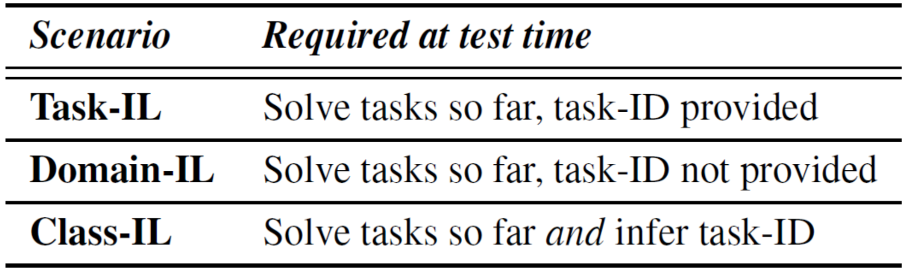


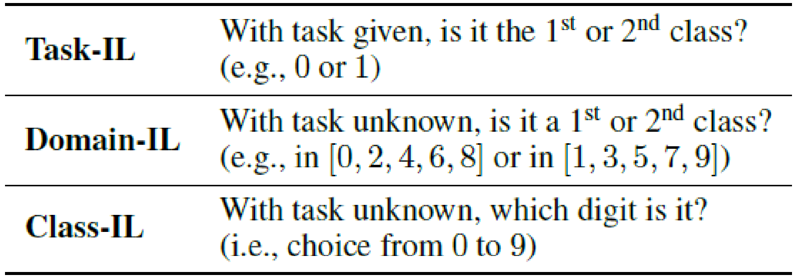
Figure 1: Schematic of split MNIST task protocol.

Table 1: Overview of the three continual learning scenarios.



Scenario	*Required at test time*
Task-IL	Solve tasks so far, task-ID provided
Domain-IL	Solve tasks so far, task-ID not provided
Class-IL	Solve tasks so far *and* infer task-ID

Table 2: Split MNIST according to each scenario.



Task-IL	With task given, is it the 1st or 2nd class? (e.g., 0 or 1)
Domain-IL	With task unknown, is it a 1st or 2nd class? (e.g., in [0, 2, 4, 6, 8] or in [1, 3, 5, 7, 9])
Class-IL	With task unknown, which digit is it? (i.e., choice from 0 to 9)

LeCun, Yann. "The MNIST database of handwritten digits."http://yann. lecun. com/exdb/mnist/(1998).
Van de Ven, G. M., & Tolias, A. S. “Three scenarios for continual learning.” (2018). Neural Information Processing Systems Workshops, 2018.

Three continual learning scenarios

- Task 5까지 순차적으로 학습이 완료된 1개의 network를 사용하는 상황 (test 시 목표에 따라 구분)

1. **Task-IL :** 특정 task(e.g., task 4)를 선정하고 학습된 network를 통해 해당 task의 성능(e.g., accuracy)을 확인
2. **Domain-IL :** 각 Task에 사용된 데이터의 domain이 다른 상황에서, 학습된 network의 task 성능을 확인(task는 동일)
3. **Class-IL :** 모든 task에 사용된 class를 전부 통합하여, 기존에 학습된 network를 통해 class 분류 성능 확인

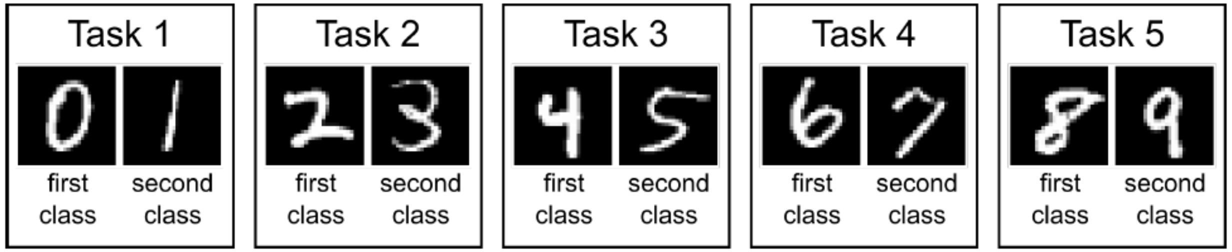
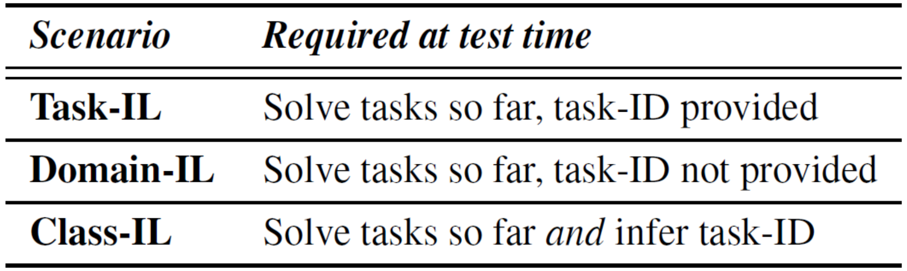


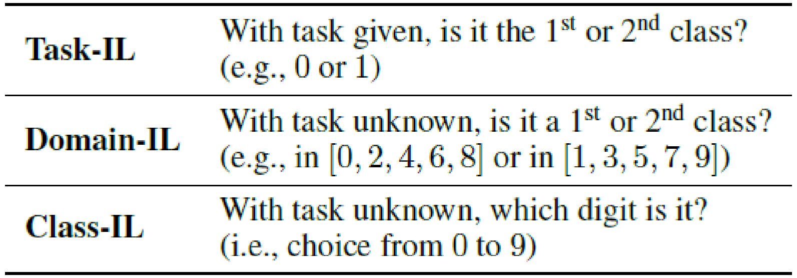
Figure 1: Schematic of split MNIST task protocol.

Table 1: Overview of the three continual learning scenarios.



Scenario	*Required at test time*
Task-IL	Solve tasks so far, task-ID provided
Domain-IL	Solve tasks so far, task-ID not provided
Class-IL	Solve tasks so far *and* infer task-ID

Table 2: Split MNIST according to each scenario.



Task-IL	With task given, is it the 1st or 2nd class? (e.g., 0 or 1)
Domain-IL	With task unknown, is it a 1st or 2nd class? (e.g., in $[0, 2, 4, 6, 8]$ or in $[1, 3, 5, 7, 9]$)
Class-IL	With task unknown, which digit is it? (i.e., choice from 0 to 9)

LeCun, Yann. "The MNIST database of handwritten digits."http://yann. lecun. com/exdb/mnist/(1998).
Van de Ven, G. M., & Tolias, A. S. “Three scenarios for continual learning.” (2018). Neural Information Processing Systems Workshops, 2018.

Three continual learning scenarios

- Task 5까지 순차적으로 학습이 완료된 1개의 network를 사용하는 상황 (test 시 목표에 따라 구분)

1. **Task-IL :** 특정 task(e.g., task 4)를 선정하고 학습된 network를 통해 해당 task의 성능(e.g., accuracy)을 확인
2. **Domain-IL :** 각 Task에 사용된 데이터의 domain이 다른 상황에서, 학습된 network의 task 성능을 확인(task는 동일)
3. **Class-IL :** 모든 task에 사용된 class를 전부 통합하여, 기존에 학습된 network를 통해 class 분류 성능 확인

Domain-IL example

Domain 1 **Task**

One task is to classify car reviews as positive or negative and another task is to classify camera reviews as positive or negative. Car and camera are two domains.

Domain 2 **Task**

Class-IL example

Domain : animal **Task 1**

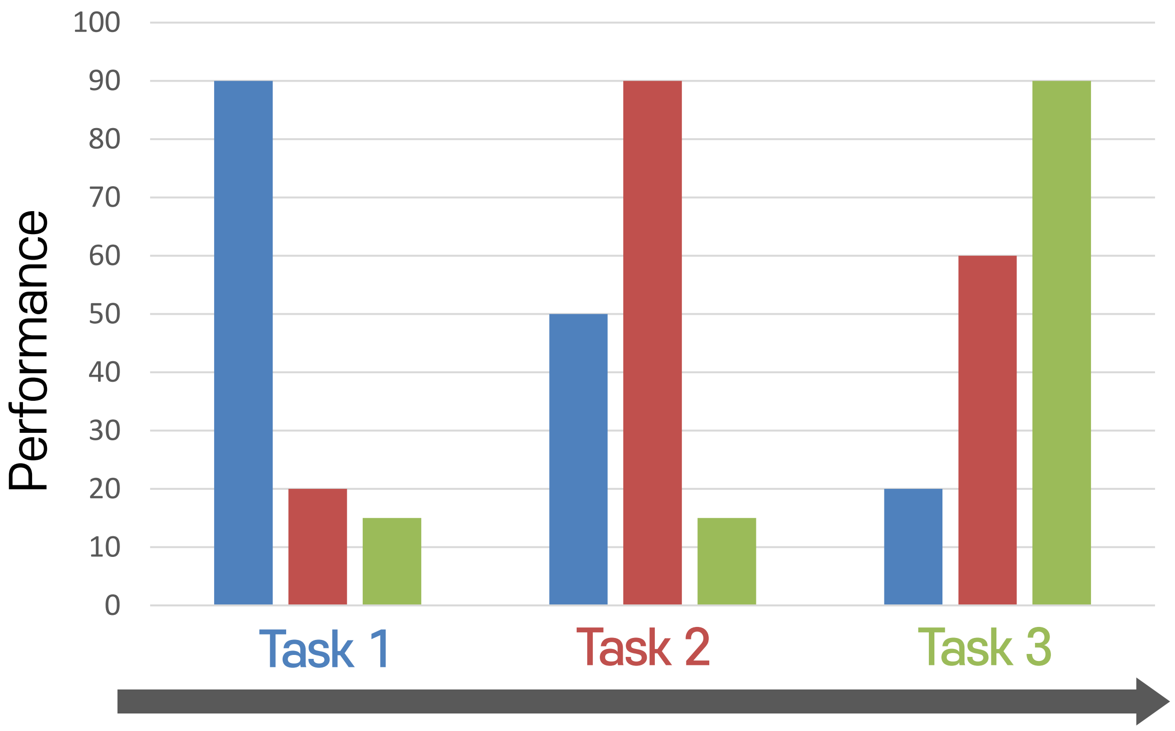
Today we learn to recognize pig and chicken, and tomorrow, we also learn to recognize sheep.

Task 2

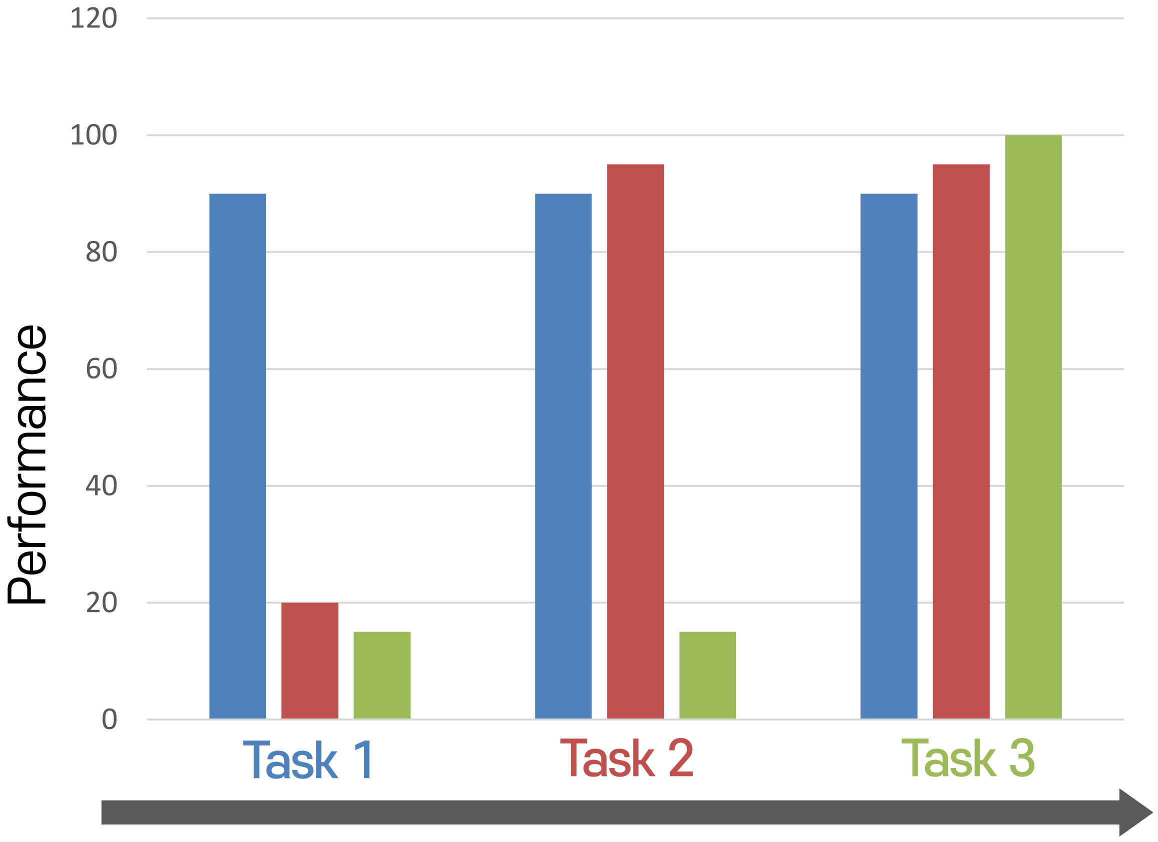
Van de Ven, G. M., & Tolias, A. S. “Three scenarios for continual learning.” (2018). Neural Information Processing Systems Workshops, 2018.

Task-IL scenario

- Task-IL scenario에서 catastrophic forgetting과 (knowledge) forward transfer 상황 설명

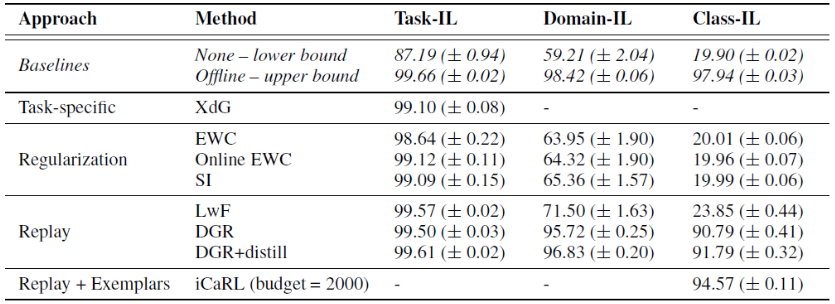


Scenario : Catastrophic forgetting



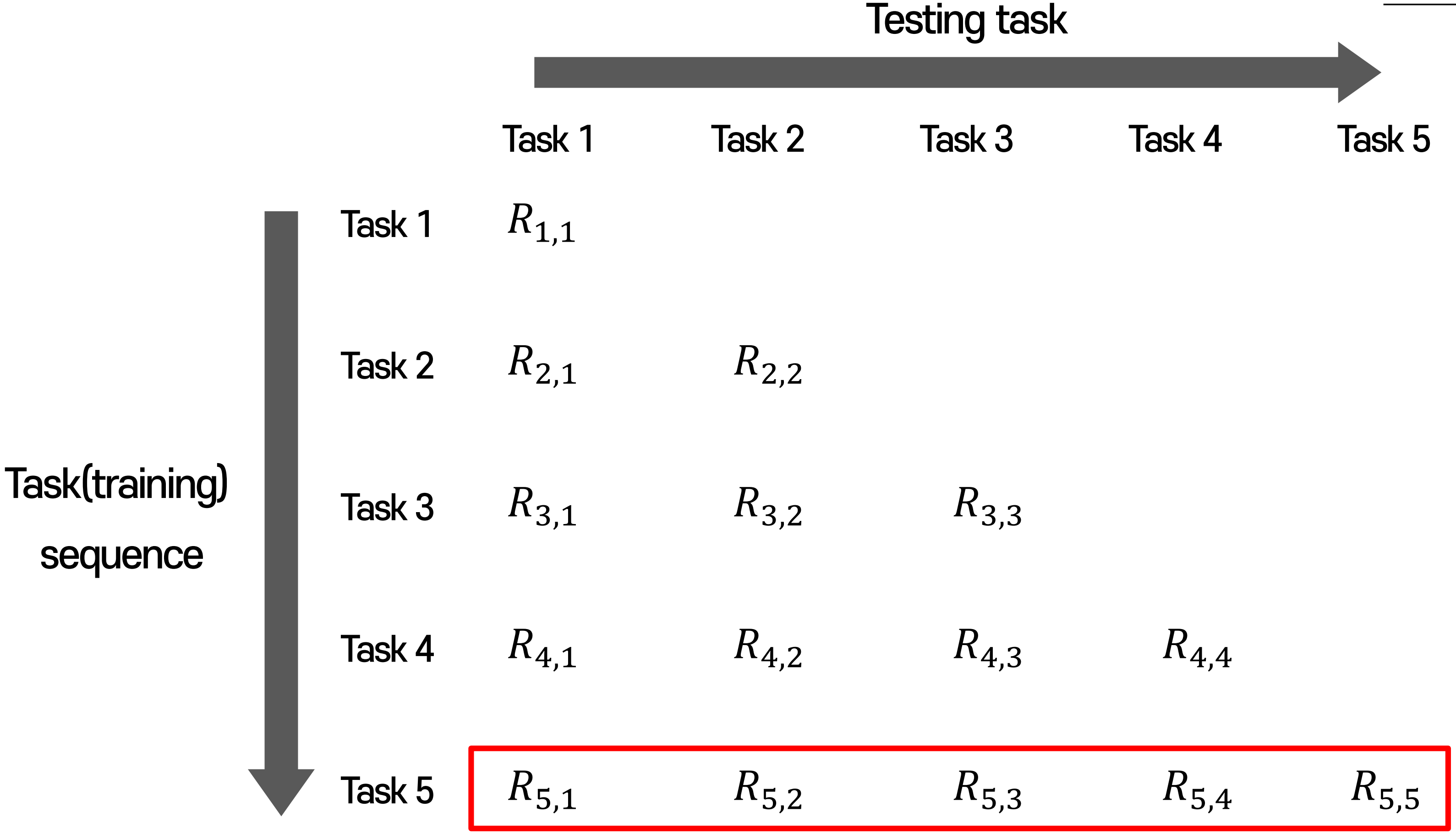
Scenario : Forward transfer

Evaluation metrics



Approach	Method	Task-IL	Domain-IL	Class-IL
Baselines	*None – lower bound* *Offline – upper bound*	*87.19 (± 0.94)* *99.66 (± 0.02)*	*59.21 (± 2.04)* *98.42 (± 0.06)*	*19.90 (± 0.02)* *97.94 (± 0.03)*
Task-specific	XdG	99.10 (± 0.08)	-	-
Regularization	EWC Online EWC SI	98.64 (± 0.22) 99.12 (± 0.11) 99.09 (± 0.15)	63.95 (± 1.90) 64.32 (± 1.90) 65.36 (± 1.57)	20.01 (± 0.06) 19.96 (± 0.07) 19.99 (± 0.06)
Replay	LwF DGR DGR+distill	99.57 (± 0.02) 99.50 (± 0.03) 99.61 (± 0.02)	71.50 (± 1.63) 95.72 (± 0.25) 96.83 (± 0.20)	23.85 (± 0.44) 90.79 (± 0.41) 91.79 (± 0.32)
Replay + Exemplars	iCaRL (budget = 2000)	-	-	94.57 (± 0.11)

- $R_{m,n}$: m번째 task까지 학습한 모델의 n번째 task의 성능



Testing task

Task(training) sequence

	Task 1	Task 2	Task 3	Task 4	Task 5
Task 1	$R_{1,1}$				
Task 2	$R_{2,1}$	$R_{2,2}$			
Task 3	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$		
Task 4	$R_{4,1}$	$R_{4,2}$	$R_{4,3}$	$R_{4,4}$	
Task 5	$R_{5,1}$	$R_{5,2}$	$R_{5,3}$	$R_{5,4}$	$R_{5,5}$

Task 5까지 학습한 모델 성능

= 각 testing task 성능의 평균

https://www.cs.uic.edu/~liub/Part-1-continual-learning-slides.pdf