



특집논문 (Special Paper)
방송공학회논문지 제30권 제4호, 2025년 7월 (JBE Vol.30, No.4, July 2025)
<https://doi.org/10.5909/JBE.2025.30.4.599>
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

루프 기반 지속 학습

안창준^{a)}, 전윤호^{b)}, 황원준^{c)†}

Loop based Continual Learning

Changjun Ahn^{a)}, Yunho Jeon^{b)}, and Wonjun Hwang^{c)†}

요약

최근 지속 학습은 모델이 이전에 학습한 지식을 효과적으로 유지하면서 새로운 작업을 학습할 수 있도록 하는 데 큰 발전을 이루었다. 그러나 대부분의 기존 연구들은 훈련 중 계층 간의 이질성은 고려하지 않는다. 이에 본 논문에서는 반복된 경험과 실패를 통해 점진적으로 최적의 전략을 도출하여 학습하는 방식에서 영감을 기반으로 이를 학습에 적용하여 데이터와 네트워크 계층 간의 이질성을 고려한 학습 루프 기반 지속 학습 방법을 제안한다. 제안된 방법은 같은 작업에 대해 여러 번의 학습 루프를 통해 각 계층의 역할을 파악하고 불필요한 파라미터의 증첩을 방지하고, 중요한 파라미터는 보존하도록 하여 안정적인 학습을 진행한다. 실험 결과를 통해 제안하는 방법이 기존 최신 방법과 비교하여 치명적 망각을 효과적으로 완화하고, 성능 또한 향상시킬 수 있음을 보인다.

Abstract

Continual learning has made significant progress in enabling models to learn new tasks while retaining previously acquired knowledge. However, most existing methods overlook the heterogeneity across network layers during training. In this paper, we propose a training loop-based continual learning method that accounts for both data and layer heterogeneity. Inspired by the way models gradually improve through repeated experiences and failures, our approach identifies the role of each layer through multiple learning loops on the same task, minimizes unnecessary parameter overlap, and preserves important parameters for stable learning. Experimental results demonstrate that our method effectively mitigates catastrophic forgetting and outperforms recent approaches.

Keyword : Continual Learning, Class Incremental Learning, Transfer Learning

a) 아주대학교 인공지능학과(Department of Artificial Intelligence, Ajou University)

b) 국립한밭대학교 인공지능소프트웨어학과(Department of Artificial Intelligence Software, Hanbat National University)

c) 고려대학교 전기전자공학부(School of Electrical Engineering, Korea University)

† Corresponding Author : 황원준(Wonjun Hwang)

E-mail: wjhwang@korea.ac.kr

Tel: +82-2-3290-3235

ORCID: <https://orcid.org/0000-0001-8895-0411>

※ 이 논문은 다음과 같은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(RS-2023-00236245, 약천후/비정형 환경변화에서의 Seamless 자율주행을 위한 인지/판단 AI SW 핵심기술 개발/RS-2025-02283048, 휴먼-AI 동반자 시대를 위한 AGI: 신뢰성·윤리성·적응성을 갖춘 차세대 범용 인공지능 개발)을 받고 또한 연구재단의 중견연구자지원사업(NRF- 2022R1A2C1091402)의 지원으로 수행된 연구임

· Manuscript June 24, 2025; Revised July 9, 2025; Accepted July 10, 2025.

1. 서론

딥러닝 발전^{[18][19]}을 되돌아보면, 모델들은 다양한 분야에서 인간 수준의 성능을 달성할 수 있게 되었다. 기존의 모델들은 대규모 데이터셋에 대해 반복적인 학습을 통해 훈련되는 반면, 데이터가 지속적으로 도착하는 상황에서는 스트리밍 데이터를 학습하는 것이 큰 도전 과제를 제시한다. 전체 데이터셋을 다시 학습하는 것은 계산적으로 비효율적이기 때문이다. 따라서, 지속 학습은 모델이 이전에 습득한 지식을 보존하면서 새로운 클래스를 점진적으로 학습할 수 있는 프레임워크로 주목받고 있다.

지속 학습의 주요 목표는 인공지능 모델이 새로운 정보를 순차적으로 습득하면서도 기존 지식을 잊지 않도록 하는 것이다. 하지만 모델이 새로운 데이터를 학습함에 따라 이전에 학습한 지식이 점차 소실되는 치명적 망각^[1] 현상이 발생한다. 이는 안정성 가소성 딜레마^[2]라 정의한다. 이는 새로운 정보를 통합하는 능력과 기존 지식을 유지하는 능력 사이의 균형을 유지하는 데 어려움이 있다는 것을 의미한다. 이 균형이 무너지면, 지나친 가소성은 치명적인 망각을 야기하고, 반면 지나친 안정성은 새로운 작업에 대한 적응을 방해하게 된다. 따라서 지속 학습 연구의 근본적인 과제는 기존에 학습한 지식을 손상시키지 않으면서 새로운 정보에 효과적으로 적응하고 치명적 망각을 줄이는 것이다.

법^{[3][4][5][6][7]}들이 제안되어 왔다. 특히 피쳐 정보 행렬을 기반으로 한 정규화 방식은 파라미터의 중요도를 추정하고 이를 보존함으로써 치명적 망각을 완화하려는 접근으로 주목받았다. 그러나 기존 연구들은 대부분 피쳐 정보 행렬을 단일 시점에서 정적으로 계산하고, 이를 모든 계층에 동일하게 적용하는 경향이 있다. 이는 각 계층이 수행하는 기능적 역할이 상이하다는 점, 즉 계층 간 이질성을 간과한다는 구조적 한계를 내포한다. 예를 들어 얇은 계층은 일반적인 특징을 학습하는 반면, 깊은 계층은 작업의 구체적인 정보를 처리하는 경향이 있는데, 이를 구분하지 않고 동일한 방식으로 정규화를 적용하면 정보의 전달 효율성과 학습 안정성 모두 저하될 수 있다.

본 논문에서는 이러한 구조적 한계를 극복하고자 학습 루프 기반의 지속 학습 프레임워크를 제안한다. 제안하는 방식은 각 작업별로 모델을 초기화하고 재학습하는 반복 구조를 통해, 학습 과정 전반에 걸쳐 파라미터 중요도의 동적 변화를 추적한다. 이 과정을 통해 각 계층의 피쳐 정보가 어떻게 변화하는지를 누적하여 반영함으로써, 계층 간 이질성을 명확히 식별하고 이에 기반한 선택적 파라미터 보존 전략을 설계할 수 있다. 이를 통해 불필요한 파라미터 중복을 억제하고 중요한 정보의 손실을 줄여, 전체 학습 성능 저하 및 치명적 망각을 효과적으로 완화할 수 있다.

제안하는 방법은 표준 지속 학습 벤치마크에서의 실험을 통해 효율성을 확인하였다. 본 논문의 구성은 다음과 같다. 2장에서는 제안하는 학습 루프 기반 지속 학습에 대해 자세히 설명하며, 3장에서는 실험을 통해 제안하는 방법의 성능을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

II. 제안하는 방법

제안하는 방법은 기존의 지속 학습에서의 훈련 방법과 다르게 학습 작업에 대한 반복 훈련을 통해 치명적 망각을 완화하고 성능을 개선하고자 한다. 본 장에서는 학습 루프 기반 지속 학습의 구현 과정에 대해 설명한다. 이어서 내부에서 각 계층의 중요도 파악 과정에 대해 자세히 설명한다.

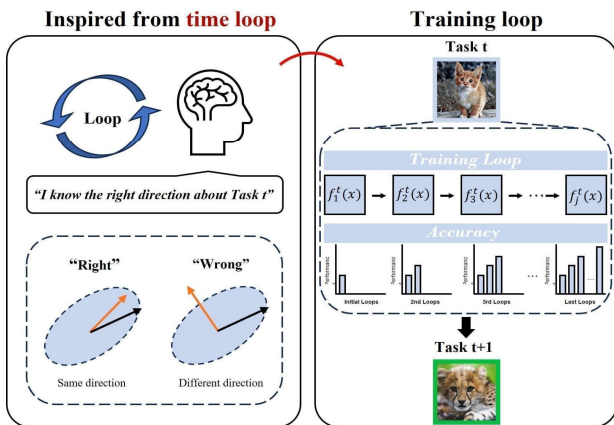


그림 1. 학습 루프 기반 지속 학습의 개요
Fig. 1. Overview of Training Loop based Continual Learning

이러한 문제를 해결하기 위해 다양한 지속 학습 기반 기

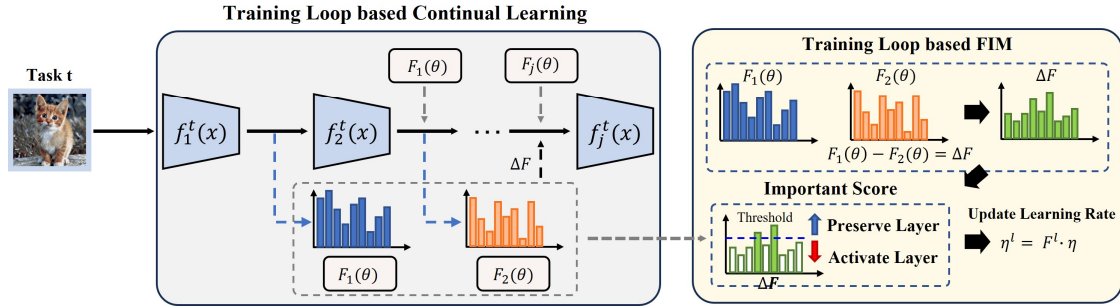


그림 2. 학습 루프 기반 지속 학습의 전체 학습 과정. f_j^t 의 경우 작업 t 에서의 j 번째 학습 루프를 의미. 또한, $F_j(\theta)$ 의 경우 해당 학습 루프의 계층별 피셔 정보 행렬을 뜻함

Fig. 2. The overall training process of Training Loop based Continual Learning

1. 기초 이론

피셔 정보 행렬. 각 계층의 민감도를 추정하려면 로그 우도의 2차 도함수, 즉 헤시안을 계산해야 한다. 민감도는 일반적으로 최적화 지형의 기울기와 곡률을 기반으로 평가되며, 이는 파라미터 업데이트 및 일반화 성능에 중요한 영향을 미친다.

여기서 θ 는 모델의 학습 가능한 파라미터를 의미하며, 손실 함수와 우도 함수는 이 θ 에 의해 정의되고 최적화된다.

로그 우도의 헤시안은 다음과 같이 정의된다.

$$H(\theta) = \nabla^2 \ell(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta^2}. \quad (1)$$

헤시안 행렬은 손실 함수의 2차 곡률 정보를 담고 있으며, 여러 학습 반복에 걸쳐 모델 파라미터의 최적화 경로를 분석하는 데 중요한 역할을 한다. 하지만 헤시안을 직접 계산하는 것은 비용이 매우 높으며, 파라미터 수가 d 일 때 복잡도는 $O(d^2)$ 이다. 이를 해결하기 위해 우리는 피셔 정보 행렬을 사용하여 곡률을 근사화한다. 다음과 같이 정의된다.

$$F(\theta) = \mathbb{E}_{P(x,y)} [\nabla \log P(y|x;\theta) \nabla \log P(y|x;\theta)^T], \quad (2)$$

여기서 $\nabla \log P(y|x;\theta)$ 는 로그 우도의 기울기이다. 피셔

정보 행렬은 1차 기울기만을 사용해 2차 도함수를 근사화하므로 계산 비용이 훨씬 낮다. 결과적으로 데이터 분포에 대한 헤시안의 기대값은 다음과 같이 근사할 수 있다.

$$\mathbb{E} [H(\theta)] \approx F(\theta) \quad (3)$$

대각선 피셔 정보 행렬. 비록 대각선 근사는 파라미터 간의 상호작용을 무시하지만, 각 파라미터가 모델 출력에 미치는 영향을 독립적으로 평가한다. 이로 인해 학습 중 중요한 파라미터를 효과적으로 식별할 수 있다. 또한 계산 복잡도는 $O(d)$ 로 크게 줄어든다. 본 논문에서는 대각선 피셔 정보 행렬을 간단히 F 로 표기한다.

2. 학습 루프 기반 지속 학습

학습 루프 기반 지속 학습은 각 작업 t 에 대해 여러 번의 학습 루프를 통해 학습이 진행된다. 여기서 학습 루프란, 작업 t 에 대해 학습이 반복되는 것을 의미한다. 따라서 학습 루프 내에서 학습 네트워크는 $f_j^t(x)$ 로 표현되며, 여기서 t 는 작업을, j 는 j 번째 학습 루프를 나타낸다. 이를 바탕으로 작업 t 에 대해 학습을 진행하고, j 만큼 학습 루프를 반복한다.

우리는 학습 루프를 통해 보존해야 할 계층과 적극적으로 학습이 필요한 계층을 구분한다. 기존 방법^{[3][4][5][6][7]}과 다르게 작업 t 를 반복 학습하면서 모델의 안정성과 가소성을 점진적으로 향상시킨다. 새로운 작업을 학습하기 전, 모

델은 이전 작업에서 학습된 지식을 바탕으로 초기화된다. 이후에는 동일한 학습 절차를 따라 학습을 수행한다. 학습이 완료되면, 모델은 학습되기 전 상태로 초기화되며 같은 작업에 대해 다시 학습을 반복한다. 이 과정에서, 우리는 이전 학습 루프에서 가장 성능이 좋았던 상태의 계층별 기울기를 저장한다. 이후 학습 루프에서는 현재의 기울기와 이전 학습 루프에서 저장한 기울기 사이의 코사인 유사도를 측정한다. 유사도를 기반으로 이전에 학습된 방향성과 비교하여 학습 정도를 조정함으로써, 각 계층을 동적으로 조정하고 보다 효과적인 적응을 가능하게 한다.

3. 학습 루프 기반 피쳐 정보 행렬

학습 루프를 통해 학습 경로를 정밀하게 추적하고 학습의 효율성을 향상시키기 위해, 벡터 간의 방향 유사도만을 측정하는 코사인 유사도 대신 피쳐 정보 행렬을 활용한다. 코사인 유사도는 파라미터 변화의 절대 크기를 반영하지 못하는 한계가 있기 때문이다. 이를 개선하기 위해, 본 논문에서는 피쳐 정보 행렬을 사용하여 각 파라미터의 중요도를 측정하는 방식을 채택한다. 하지만, 기존의 정적 중요도 추정 방법들은 학습 도중 변화하는 파라미터 중요도를 반영하는 데 한계가 있다.

이를 해결하기 위해 우리는 학습 루프 기반 피쳐 정보 행렬을 제안한다. 이 방법은 각 학습 루프에서 피쳐 정보 행렬을 계산하여 중요도의 상대적 변화를 지속적으로 추적하며, 계층별 중요도 변화에 대해 이해가 가능하다.

구체적으로, 학습 루프가 한 번 끝난 뒤, 모델이 최적 성능을 낼 때의 대각선 피쳐 정보 행렬을 계산한다. 네트워크 내 계층 수가 l , 학습 루프가 j 번째일 때, 이 행렬은 다음과 같이 정의된다.

$$F_j(\theta) = \{F_j^1(\theta), F_j^2(\theta), \dots, F_j^l(\theta)\} \quad (4)$$

이후 학습 루프에서는 이전 학습 루프와의 차이를 반영하여 피쳐 정보 행렬을 업데이트한다. 특히, 피쳐 정보 행렬은 Negative Log-Likelihood (NLL) 손실 \mathcal{L} 은 계층 단위로 계산되며, 다음과 같이 정의된다.

$$\mathcal{L} = -\sum_i \log p(y_i|x_i). \quad (5)$$

여기서 $p(y_i|x_i)$ 는 입력 x_i 에 대해 모델이 예측한 확률이다. $j+1$ 번째 학습 루프에서의 계층별 피쳐 정보 행렬은 $F_{j+1}^l(\theta)$ 로 표기되며, 이는 NLL 손실에 따른 파라미터 중요도 변화를 포착한다.

훈련 전반 과정에서 계층들의 중요도 변화는, 이전 학습 루프의 $F_j^l(\theta)$ 와 현재 학습 루프의 $F_{j+1}^l(\theta)$ 의 차이로 계산된다. 이러한 값들은 계층 내의 각 파라미터별로 정의되므로 다차원 텐서로 구성된다. 그래서 우리는 해석 가능한 비교를 위해, 계층 내 모든 파라미터에 대해 피쳐 정보 행렬값을 평균 내어 하나의 스칼라 대표값으로 변환한다. 층별 훈련 루프 피쳐 정보 행렬의 차이는 다음과 같이 표현된다:

$$\tilde{F} = E[F_j(\theta)] - E[F_{j+1}(\theta)], \quad (6)$$

그러나 \tilde{F} 값이 지나치게 크거나 작아질 수 있어, 이는 학습의 불안정성을 초래할 수 있다. 여기서 \tilde{F} 는 작업 j 와 $j+1$ 에서의 함수 $F_j(\theta)$ 의 기댓값 차이로 정의되며, 이때의 $E[\cdot]$ 는 데이터 분포에 대한 평균값, 즉 확률적 입력에 대해 함수가 가지는 평균적인 출력값을 의미한다. 하지만 이러한 기대값 차이는 데이터의 분포 변화나 분산에 따라 급격하게 변동할 수 있어, 학습 중 불안정성을 유발할 수 있다. 이를 완화하기 위해, 우리는 제곱근을 이용하여 업데이트 크기를 조정하는 변환 방식을 적용한다. 이때 τ 는 이상치의 영향을 제어하고 계층별 차이를 강조하는 지수 하이퍼파라미터이다. 이 변환은 다음과 같이 정의된다:

$$\hat{F} = (\tilde{F})^{\frac{\tau}{2}}, \quad (7)$$

여기서 τ 는 계층별 중요도 차이를 증폭시키기 위해 설정되며, 특정 계층은 거의 변하지 않도록 조정할 수 있게 한다.

식 (7)을 기반으로 우리는 층별 중요도 변화를 반영하고, 이에 따라 학습률을 동적으로 조정한다. 그러나 \hat{F} 는 고정

된 범위가 없기 때문에, Exponential Min-Max Scaler를 적용하여 다음과 같이 정규화한다:

$$\eta^l = F^l \cdot \eta, F^l = \left(\frac{F^l - F_{\min}^l}{F_{\max}^l - F_{\min}^l + \epsilon} \right), \quad (8)$$

여기서 η^l 는 계층별 학습률을 나타내며, 주어진 작업에 대해 민감도와 현재 파라미터 변화를 기반으로 조정된다. 이 과정은 계층 간 학습률 변화가 지나치게 커지는 것을 방지하여, 더 균형 잡히고 안정적인 학습을 가능하게 한다. 또한, ϵ 은 F 가 정의되지 않거나 무한대가 되는 것을 막고 수치적으로 안정적인 계산을 보장한다.

4. 중요도 점수

우리는 학습 루프 기반 피셔 정보 행렬을 활용하여 훈련 중 파라미터 중요도의 변화를 동적으로 반영한다. 피셔 정보 행렬은 학습 중 파라미터의 중요도를 효과적으로 측정할 수 있지만, 특정 작업에서는 변동성이 존재한다. 우리는 학습 루프 기반 피셔 정보 행렬과 함께 중요도 점수^[17]를 도입하여 학습의 안정성을 높이고 중요도 측정을 더욱 정밀하게 한다. 중요도 점수는 다음과 같이 정의된다:

$$\bar{I}(t) = \left(\sum (\theta^t - \theta^{t-1}) \nabla_{\theta} L \right)^2, \quad (9)$$

여기서 t 는 작업을 나타낸다. ∇L 는 이전 훈련 루프에서 해당 파라미터의 손실 함수에 대한 기울기이다. 식 (9)와 동일한 방식으로, 우리는 개별 파라미터 수준이 아닌 계층별 수준에서 중요도를 계산하여 더 구조화된 평가가 가능하다. 식 (7)과 식 (8)을 적용하기 전에, 우리는 중요도 점수를 활용해 중요한 파라미터를 선택한다.

구체적으로, 식 (7)의 출력값들을 중요도 점수에 따라 정렬하고, 특정 임계값 이하의 층들을 보호 대상으로 지정하여 학습의 안정성을 향상시킨다.

III. 실험 결과 및 분석

우리는 제안된 방법의 효과를 실험적으로 증명하기 위해 CIFAR-10, CIFAR-100^[8], Tiny-ImageNet^[9], ImageNet^[10] 네 가지 지속 학습 벤치마크에서 제안 방법의 성능을 실험적으로 검증한다. 제안 방법은 최신의 지속 학습 방법들과 비교되며, 각 구성 요소의 기여도를 평가하기 위한 광범위한 분석도 수행한다.

실험에는 다양한 버퍼 크기와 데이터셋이 사용되며, CIFAR-10은 클래스당 5,000개의 학습 이미지와 1,000개의 테스트 이미지로 구성된다. CIFAR-100은 클래스당 500개의 학습 이미지와 100개의 테스트 이미지로 구성되며, Tiny-ImageNet의 경우는 클래스당 500개의 학습 이미지와 50개의 검증 이미지 그리고 50개의 테스트 이미지로 구성된다. 마지막으로, ImageNet의 경우 클래스당 500개의 학습 이미지와 50개의 테스트 이미지로 구성되고, 해상도는 224x224로 재설정된다.

본 연구에서는 학습 성능을 평가하기 위해 Final Average Accuracy (FAA)와 Final Forgetting (FF)을 사용한다^{[11][12]}. FAA는 최종 모델이 여러 작업에 대해 달성한 모델의 평균 정확도를 측정하며, FF는 망각으로 인해 감소한 정확도를 측정한다. FAA가 높고 FF가 낮을수록 더 나은 성능을 의미한다. 이러한 지표들을 통해 제안 방법의 효과를 검증한다.

모든 실험은 NVIDIA A5000 GPU에서 수행되었으며, 표준 지속 학습 프로토콜을 따른다.

CIFAR-10, CIFAR-100, Tiny-ImageNet의 경우 ResNet-18^[13], ImageNet의 경우는 ResNet-34^[13]를 백본 네트워크를 사용하였다. 버퍼 관리를 위해 클래스 균형이 고려된 reservoir sampling^[14]을 활용하였고, 모델은 SGD 옵티마이저로 학습되며, 훈련 루프의 시작 시점마다 초기화된다.

CIFAR-100의 경우 각 작업당 50 에폭으로 학습된다. Tiny-ImageNet의 경우 각 작업당 100 에폭으로 학습된다. 학습률 감소 스케줄은 데이터셋에 맞게 조정된다. CIFAR-100에서는 각 작업의 에폭 35, 45 시점에서 학습률을 감소시키며, Tiny-ImageNet에서는 에폭 35, 60, 75 시점에서 감소시킨다. 두 경우 모두 지정된 에폭마다 학습률이 10배로 감소된다.

표 1. 다양한 데이터셋 및 버퍼 크기에 대한 기준선과 제안된 방법 간의 비교

Table 1. Comparison of baselines and the proposed method under various datasets and buffer sizes

Method	Class-IL (CIL)			Task-IL (TIL)		
	CIFAR100	CIFAR100	TinyImag	CIFAR100	CIFAR100	TinyImag
Buffer Size	500	2,000	4,000	500	2,000	4,000
iCaRL ^[3]	46.66	52.60	31.47	84.45	86.24	66.06
oEWC ^[4]	8.24	-	-	21.2	-	-
ER ^[5]	20.98	36.06	23.29	73.37	81.09	64.96
ER w/Ours	22.51	37.9	23.92	75.79	84.49	64.98
DER++ ^[6]	36.37	50.72	35.56	75.64	82.43	67.94
DER w/Ours	39.49	54.12	37.99	77.38	84.48	68.13
X-DER ^[7]	47.6	57.83	44.87	85.80	88.78	76.30
X-DER w/Ours	48.65	59.42	45.24	87.65	89.36	77.40

ImageNet은 기존 실험에 포함되어 있지 않기 때문에 실험은 기존 연구^[18]를 참고한다.

제안한 방법은 다양한 기존 모델에 plug and play 방식으로 쉽게 적용될 수 있다는 장점이 있다. 또한, 단순히 에폭 수를 늘리는 기존 방식과 달리, 학습 루프 구조를 활용함으로써 보다 안정적인 학습이 가능하며 치명적 망각을 완화할 수 있다.

FAA 결과는 표 1에 보고되어 있다. FF 결과 또한 같은 테이블에 제시되어 있다. 우리는 네 개의 데이터셋에 대해 각 데이터셋의 클래스 수에 따라 다양한 크기의 메모리 버퍼를 사용하여 실험을 수행한다. 또한, 계산 비용과 효율성을 고려하여 첫 번째 훈련 루프 설정을 가장 합리적인 실험 구성으로 채택하였다.

각 실험은 서로 다른 랜덤 시드를 사용하여 다섯 번 반복되었으며, 평균을 보고한다.

먼저, CIFAR-10에서는 DER++^[6]에 비해 우리의 방법이 버퍼 크기가 200일 때 2.51%의 성능 향상을 보인다. CIFAR-100에서는 버퍼 크기 500에서 3.12% 성능 향상이 관찰되었다. X-DER^[7] 자체는 1.59% 향상을 기록하였다. 이러한 결과는 우리의 방법이 기존 모든 방법보다 일관되게 우수함을 보여준다.

추가적으로, Tiny-ImageNet^[9]에서는 우리의 방법이 2.43%의 성능 향상을 기록했다. 전반적으로, 제안하는 방법은 기존 방법들에 비해 모든 데이터셋에서 일관된 뛰어난 성능을 보여주며, 그 효과성을 입증한다.

표 2에서 나타나듯이, 제안한 방법은 ImageNet 데이터셋에도 기존 방법 ER, DER++ 대비 CIL 및 TIL 모두에서 일

표 2. ImageNet 데이터셋에서의 정량적 성능 지표

Table 2. Performance metrics on the ImageNet datasets

Method	ImageNet	
	CIL	TIL
Buffer Size	20,000	
ER ^[5]	23.11	39.17
w/ Ours	23.37	38.91
DER++ ^[6]	18.82	40.41
w/ Ours	19.01	40.84

관된 성능 향상을 달성하였다. 이는 대규모 복잡한 데이터셋에서도 제안한 접근 방식의 일반성과 효과성을 뒷받침한다. 추가적으로, TIL 설정에서 ER w/ Ours가 ER 대비 낮은 성능을 보이는데, 이는 ER은 단순히 리플레이 기반으로, TIL에서는 task-ID 정보를 명확히 사용하기에 본 논문의 방법론보다 각 분류기에 정답을 직접적으로 감독할 수 있다.

표 3. 학습 루프 기반 학습과 에폭 기반 학습의 성능 지표

Table 3. Performance metrics for loop based learning and epoch based learning

Method	ACC
DER++ ^[6]	37.48
w/ Ours	39.49

표 3을 통해, 기존 DER++에서 학습 에폭을 단순히 두 배 늘린 경우와 DER++ w/ Ours 방식에서 제안하는 학습 루프 기반 학습 방식을 적용한 경우의 성능 차이를 볼 수 있다. 단순히 학습 횟수를 증가시켰을 때 보다, 제안한 학습 루프를 적용했을 때 더욱 뚜렷한 성능 향상이 나타난다. 이는 단순히 반복 학습이 성능 향상의 핵심 요인이 아님을

의미하며, 오히려 클래스 입력에 따라 필요한 파라미터와 불필요한 파라미터 간의 관계를 반영하여 학습을 진행하는 제안 방식이 보다 효과적인 것을 볼 수 있다.

표 4. 학습 루프 증가에 따른 성능 지표

Table 4. Performance as the number of training loops increases

Method	Base	1st	2nd	3rd
DER++ ^[6]	36.37	37.48	38.00	37.64
w/Ours	36.37	39.49	39.88	38.07

표 4를 통해, 기존 DER++와 우리의 제안 방법 간의 성능 차이를 분석한다. 학습 루프만 단독으로 DER++에 적용하는 것은 크게 성능 향상을 시키지 않지만, 제안 방법과 함께 사용할 경우 훨씬 더 큰 성능 향상을 이끌어 낸다. 추가적으로, 우리의 접근 방식은 성능을 효과적으로 개선하지만, 세 번째 학습 루프 이후에는 성능 향상이 점차 감소하는 것을 관찰할 수 있다. 이는 포화 지점에 도달했음을 시사하며, 이후에는 기울기의 발산으로 이어질 수 있다. 이에 따라 계산 비용과 학습 시간을 고려하여 첫 번째 학습 루프를 기본 설정으로 채택하였다. 이는 두 번째 학습 루프보다 효율성과 성능 간 균형이 우수함을 보여준다.

또한, 학습 루프가 증가할 때마다 성능 향상을 보이지만 세 번째 학습 루프와 두 번째 학습 루프를 비교하였을 때, 성능이 소폭 하락한 것을 볼 수 있다. 이는 해당 학습 루프가 반복되면서 더이상 최적화될 수 없는 포화상태로 추론할 수 있다. 그렇기 때문에 더이상의 성능 향상이 없고 오히려 성능이 하락하는 것을 볼 수 있다.

표 5. 기존 중요도 기반 연구의 학습 루프 증가에 따른 정량적 성능 지표 비교

Table 5. Comparison of performance as the number of training loops increases in existing regularization based

Method	Base	1st	2nd	3rd
oEWC ^[4]	8.24	8.27	8.87	8.64
w/Ours	8.24	9.08	9.68	9.16

표 5에서는 학습 루프 기반 피셔 정보 행렬이 중요도 정보와 변화량을 동적으로 반영함으로써 기존 oEWC^[4]와 비교했을 때 단순히 학습 루프를 증가하는 것으로 성능이 소

폭 상승하지만, 본 논문에서 제안하는 학습 루프 기반 피셔 정보 행렬을 적용하여 학습 루프를 반복할 경우, 보다 더 나은 성능 향상을 볼 수 있다.

표 6. Vision Transformer 모델에서 ImageNet 데이터셋의 실험

Table 6. Experiments on the ImageNet dataset using the Vision Transformer

Method	#P	Top-1	
		Avg	Last
DyTox ^[16]	11.01	77.15	69.10
w/Ours	11.01	78.21	70.23

우리는 Vision Transformer^[15]을 이용한 실험을 통해 제안한 방법의 일반화를 검증한다. Vision Transformer^[15] 실험에서는 DyTox^[16]를 기준으로 채택하고, 기존 설정을 유지한 채 학습 루프를 추가하는 실험을 진행하였다. 결과적으로, 모든 설정에서 성능이 개선되었으며, 최종 성능과 평균 성능 모두 높은 수준을 유지하는 것으로 나타났다. 이는 기존에 학습된 정보를 잘 보존하면서 새로운 정보 학습을 촉진하는 것을 알 수 있다. 또한, 본 방법은 추가 모델 생성을 요구하지 않고, plug and play 방식으로 적용 가능하며, 필요한 파라미터 수가 증가하지 않음을 표 6에서 확인할 수 있다.

표 7. 코사인 유사도와 L_2 distance 기반 방식과 피셔 정보 행렬 기반 방식의 성능 비교

Table 7. Compare the performance of methods based on cosine similarity and L_2 distance with methods based on Fisher information matrix

Method	ACC
L_2 distance	28.23
Cosine similarity	38.42
Fisher Information Matrix	39.49

표 7에서는 코사인 유사도 기반 방식과 L_2 distance 기반 방식 간의 성능 차이를 비교하고 있으며, 이를 통해 파라미터 간의 관계를 어떤 방식으로 측정하느냐에 따라 학습 성능이 크게 달라질 수 있음을 보여준다. 실험 결과 L_2 distance를 기준으로 학습한 경우보다 코사인 유사도를 활용한 방식이 더 높은 성능을 나타내며, 이는 단순히 파라미터

간의 거리 자체를 최소화하기보다는, 학습 과정에서의 파라미터 방향성을 유지하는 것이 더 효과적이라는 점을 시사한다. 더 나아가, 피쳐 정보 행렬을 기반으로 한 방식은 이 두 방식보다는 더욱 우수한 성능을 기록하며 가장 효과적인 결과를 보인다. 피쳐 정보 행렬은 각 파라미터가 모델 출력에 미치는 민감도를 반영하는 정보로, 이를 바탕으로 파라미터 공간 내에서 중요한 방향을 강조하여 학습을 유도함으로써, 단순한 거리 기반 접근이나 방향성 유지보다도 더 정교하고 정보적인 학습 경로를 형성할 수 있게 된다. 따라서, 파라미터 간의 거리나 방향만을 고려하는 기존 방식들에 비해, 피쳐 정보 행렬을 활용하여 학습에 있어 중요한 파라미터 방향을 효과적으로 보존하는 것이 성능 향상에 있어 더욱 핵심적인 요소임을 알 수 있다.

표 8. CIFAR-100에서의 Final Forgetting 성능의 정량적 지표
Table 8. Final Forgetting on the CIFAR-100 dataset

Method	CIFAR100	
	CIL	TIL
ER ^[5]	47.62	11.55
w/ Ours	41.18	7.34
DER+ ^[6]	42.58	11.68
w/ Ours	29.80	9.63

Final Forgetting (FF)^{[11][12]}의 경우, 지속 학습에서 모델이 지식을 더 효과적으로 유지하며 더 큰 안정성을 달성하는 것을 의미한다. 제안하는 우리의 방법과 비교했을 때, 더 낮은 FF를 달성하였으며, 우리의 방법이 치명적 망각을 효과적으로 줄일 수 있음을 입증하였다.

IV. 결론

본 논문에서는 치명적 망각 문제를 해결하기 위해 학습 루프 기반 지속 학습 방법을 제안한다. 제안하는 방법은 단일 학습 과정에 대해 여러 번의 학습 루프를 반복하여, 반복적인 iteration을 통해 점진적으로 학습 과정을 개선하는 방식을 따른다. 각 학습 루프마다 계산된 피쳐 정보 행렬을 활용하여, 파라미터 정보의 누적과 중요도의 변화를 동적으로 추적한다. 이는 각 계층의 역할에 대한 보다 정밀한

이해를 가능하게 하며, 계층별로 학습률을 동적으로 조정할 수 있도록 한다. 그 결과, 본 방법은 이전 작업에서 학습한 중요한 파라미터를 효과적으로 보존하면서도 새로운 작업 학습을 위한 유연성을 유지할 수 있게 한다.

본 논문에서 제안한 방식은 정적인 파라미터 중요도 추정 방식의 한계를 넘어, 학습 과정 전반에 걸쳐 중요도를 동적으로 반영함으로써 치명적 망각을 효과적으로 완화하였다. 이러한 구조적 접근은 단순한 성능 개선을 넘어, 실제 지속 학습 기반 시스템 설계에서도 중요하다. 특히 제안 방법은 연속적으로 변화하는 입력 환경에서 적응이 요구되는 다양한 실시간 어플리케이션에 적합하다. 예를 들어, 사용자 행동 변화에 따른 온라인 추천 시스템, 주행 환경 변화에 유연하게 대응해야 하는 자율주행 인식 시스템, 그리고 새로운 제품이나 결합 유형이 지속적으로 발생하는 스마트 제조 및 검사 시스템과 같은 환경에서 효과적으로 적용될 수 있다. 이처럼 본 방법은 제한된 자원 내에서의 연속 학습을 가능하게 하며, 기존 지식의 손실을 최소화하면서도 새로운 정보에 대한 민첩한 적응을 가능하게 함으로써, 지속 학습 기술의 실용적 확장 가능성을 보여준다.

참고 문헌 (References)

- [1] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," arXiv preprint, arXiv:1312.6211, 2013. doi: <https://doi.org/10.48550/arXiv.1312.6211>
- [2] S. T. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*, vol.70, Springer Science & Business Media, 2012. doi: <https://doi.org/10.1007/978-94-009-7758-7>
- [3] S. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental classifier and representation learning," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2001 - 2010, 2017. doi: <https://doi.org/10.48550/arXiv.1611.07725>
- [4] J. Schwarz, W. Czarnecki, J. Luketina, A. Grabska-Barwinska, Y. W. Teh, R. Pascanu, and R. Hadsell, "Progress & compress: A scalable framework for continual learning," *Proceedings of the International Conference on Machine Learning (ICML)*, Stockholm, Sweden, pp. 4528 - 4537, 2018. doi: <https://doi.org/10.48550/arXiv.1805.06370>
- [5] M. Riemer, I. Cases, R. Ajemian, M. Liu, I. Rish, Y. Tu, and G. Tesaro, "Learning to learn without forgetting by maximizing transfer

- and minimizing interference,” arXiv preprint arXiv:1810.11910, 2018, <http://arxiv.org/abs/1810.11910> (accessed Jul. 8, 2025).
doi: <https://doi.org/10.48550/arXiv.1810.11910>
- [6] P. Buzzega, M. Boschini, A. Porrello, D. Abati, and S. Calderara, “Dark experience for general continual learning: A strong, simple baseline,” *Advances in Neural Information Processing Systems*, Vol.33, pp.15920 - 15930, 2020.
doi: <https://doi.org/10.48550/arXiv.1810.11910>
- [7] M. Boschini, L. Bonicelli, P. Buzzega, A. Porrello, and S. Calderara, “Class-incremental continual learning into the extended DER-verse,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, No.5, pp.5497 - 5512, 2022.
doi: <https://doi.org/10.48550/arXiv.2201.00766>
- [8] A. Krizhevsky, Learning multiple layers of features from tiny images, University of Toronto, Technical Report, 2009, Retrieved from <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf> (accessed Jul. 8, 2025).
<https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
- [9] Y. Le and X. Yang, Tiny ImageNet Visual Recognition Challenge, Stanford CS231N Course Project, 2015, <http://tiny-imagenet.herokuapp.com/> (accessed Jul. 8, 2025).
doi: <https://doi.org/10.57702/19nlfifl>
- [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, pp.248 - 255, 2009.
doi: <https://doi.org/10.1109/CVPR.2009.5206848>
- [11] E. Arani, F. Sarfraz, and B. Zonooz, “Learning fast, learning slow: A general continual learning method based on complementary learning system,” arXiv preprint arXiv:2201.12604, 2022, <http://arxiv.org/abs/2201.12604> (accessed Jul. 8, 2025).
doi: <https://doi.org/10.48550/arXiv.2201.12604>
- [12] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. Van De Weijer, “Class-incremental learning: Survey and performance evaluation on image classification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.45, No.5, pp.5513 - 5533, 2022.
doi: <https://doi.org/10.48550/arXiv.2010.15277>
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp.770 - 778, 2016.
doi: <https://doi.org/10.48550/arXiv.1512.03385>
- [14] P. Buzzega, M. Boschini, A. Porrello, and S. Calderara, “Rethinking experience replay: A bag of tricks for continual learning,” *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, pp.2180 - 2187, 2021.
doi: <https://doi.org/10.48550/arXiv.2010.05595>
- [15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16×16 words: Transformers for image recognition at scale,” *International Conference on Learning Representations (ICLR)*, 2021, <https://arxiv.org/abs/2010.11929> (accessed Jul. 8, 2025).
doi: <https://doi.org/10.48550/arXiv.2010.11929>
- [16] A. Douillard, A. Ramé, G. Couairon, and M. Cord, “DyTox: Transformers for continual learning with dynamic token expansion,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp.15534 - 15544, 2022.
doi: <https://doi.org/10.48550/arXiv.2111.11326>
- [17] P. Molchanov, A. Mallya, S. Tyree, I. Frosio, and J. Kautz, “Importance estimation for neural network pruning,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp.11264 - 11272, 2019.
doi: <https://doi.org/10.48550/arXiv.1906.10771>
- [18] H. Koh, M. Seo, J. Bang, H. Song, D. Hong, S. Park, J.-W. Ha, and J. Choi, “Online boundary-free continual learning by scheduled data prior,” *Proceedings of the 11th International Conference on Learning Representations (ICLR)*, Kigali, Rwanda, 2023. <https://openreview.net/forum?id=qco4ekz2Epm>
- [19] J. H. Choi, “The history and development direction of artificial intelligence technology focusing on visual recognition technology,” *Journal of Broadcast Engineering*, vol. 25, no. 1, pp. 14 - 19, Jan. 2020. <https://www.kibme.org/resources/journal/20200206100539343.pdf>
- [20] W. J. Kim, “Recent trends in artificial intelligence-based image quality enhancement technology,” *Journal of Broadcast Engineering*, vol. 25, no. 1, pp. 20 - 27, Jan. 2020. <https://www.kibme.org/resources/journal/20200206100603671.pdf>

저 자 소 개



안 창 준

- 2024년 : 국립한밭대학교 전자제어공학과 학사
- 2024년 ~ 현재 : 아주대학교 인공지능학과 석사과정
- ORCID : <https://orcid.org/0009-0000-4198-1237>
- 주관심분야 : 영상처리, 컴퓨터비전, 딥러닝



전 윤 호

- 2006년 : 서울대학교 수리과학부 이학사, 컴퓨터공학부 공학사
- 2008년 : 서울대학교 전기·컴퓨터공학부 공학석사
- 2008년 ~ 2019년 : 국방과학연구소 연구원
- 2019년 : 한국과학기술원(KAIST) 전기및전자공학부 공학박사
- 2019년 : SK텔레콤 연구원
- 2019년 ~ 2022년 : ㈜모플 CTO
- 2023년 ~ 현재 : 국립한밭대학교 인공지능소프트웨어학과 조교수
- ORCID : <https://orcid.org/0000-0001-8043-480X>
- 주관심분야 : 영상처리, 컴퓨터비전, 시계열 분석, 딥러닝



황 원 준

- 1999년 : 고려대학교 전자공학과 공학사
- 2001년 : 고려대학교 전자공학과 공학석사
- 2016년 : 한국과학기술원(KAIST) 전기및전자공학부 공학박사
- 2001년 ~ 2016년 : 삼성전자 삼성종합기술원 전문연구원
- 2016년 ~ 2020년 : 아주대학교 소프트웨어학과 조교수
- 2020년 ~ 2025년 : 아주대학교 소프트웨어학과 부교수
- 2025년 ~ 현재 : 고려대학교 전기전자공학부 부교수
- ORCID : <https://orcid.org/0000-0001-8895-0411>
- 주관심분야 : 신호처리, 딥러닝, 컴퓨터비전