Journal of the Korean Institute of Industrial Engineers
Vol. 50, No. 1, pp. 11-22, February 2024.
ISSN 1225-0988 | EISSN 2234-6457
https://doi.org/10.7232/JKIIE.2024.50.1.011


# 심층 강화학습을 활용한 이미지 분류

고병은 · 김성범†

고려대학교 산업경영공학과

# Deep Reinforcement Learning for Image Classification

Byeongeun Ko · Seoung Bum Kim

Department of Industrial and Management Engineering, Korea University

We present a deep reinforcement learning-based approach that leverages data augmentation techniques to improve the accuracy of image classification tasks. We define the essential components of deep reinforcement learning and create an environment that enables the model to learn more autonomously when additional learning is required. Our experiments with various benchmark datasets(CIFAR-10, SVHN and WM-811K) demonstrate that the proposed deep reinforcement learning-based approach outperforms the conventional convolutional neural networks in terms of classification accuracy, highlighting its potential to effectively address classification problems using deep reinforcement learning.

***Keywords:*** Data Augmentation, Image Classification, Reinforcement Learning, Sequential-Decision Making Process

## 1. 서 론

강화학습은 기계학습의 한 종류로 어떠한 환경(environment)에서 에이전트(agent)가 행동(action)을 취하고, 잘된 행동인지 잘못된 행동인지를 판단하여 보상(reward)함으로써 학습을 수행하는 분야이다(Shin *et al.*, 2019). 에이전트는 단순히 한 번의 행동으로 끝나지 않고, 순차적인 행동을 취한 후 누적 보상(accumulated reward)을 최대화 하는 방향으로 수많은 시행착오를 거치며 학습이 진행된다. 이러한 강화학습은 강력한 장점을 가지고 있는데 기존의 전통적인 딥러닝 방법처럼 인간의 지식을 이용하여 모델을 학습시키는 지도학습(supervised learning)이 아니라 강화학습의 에이전트가 환경과 상호 작용하며 새로운 행동을 탐험하고, 최적 행동을 찾아가는 알고리즘으로써 스스로 학습 데이터를 쌓고 학습을 할 수 있기 때문이다(AlMahamid *et al.*, 2021). 이러한 강화학습의 기본적인 학습 흐름을 <Figure 1>에서 도식화 하였다. 하지만 이와 같은 고전적인 강화학습은 단순한 상태 공간과 행동 공간을 가지는 문제에만 적용할 수 있으므로 복잡한 현실 세계의 문제들을 해결하는데 한계가 있다(Jang *et al.*, 2019). 따라서 이러한 한계를 인공 신경망 구조를 도입하여 실제와 같은 환경에서도 동작할 수 있는 심층 강화학습(deep reinforcement learning)이 제안되었다. 심층 강화학습은 최근 딥러닝 분야에서 가장 주목받는 분야 중 하나이며 순차적인 의사결정이 필요한 자율주행, 무인 로봇 그리고 게임 등과 같은 분야에서 두각을 나타내며

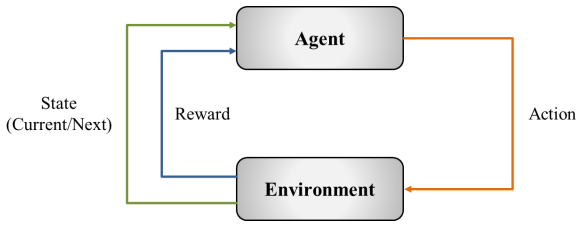



**Figure 1.** Basic Framework of Reinforcement Learning. The agent, upon taking an action receives the corresponding reward and the subsequent state from the environment.

본 논문은 한국연구재단 BK21 FOUR 사업의 지원을 받아 연구되었음.
† 연락저자 : 김성범 교수, 02841, 서울특별시 성북구 안암로 145 고려대학교 산업경영공학부, Tel : 02-3290-3397, Fax : 02-929-5888, E-mail : sbkim1@korea.ac.kr
2023년 11월 14일 접수; 2024년 1월 23일 게재 확정.

활발하게 연구되고 있다(AlMahamid *et al.*, 2021). 심층 강화학습은 DeepMind의 알파고(AlphaGo)와 같이 특정 문제를 해결하기 위한 방법론으로 사용되기도 하고(Silver *et al.*, 2017), OpenAI의 ChatGPT 서비스처럼 미세조정 단계에서 보조적인 역할로 사용되기도 한다(Zhao *et al.*, 2023).

이처럼 최근 널리 쓰이고 있는 심층 강화학습 방법론을 전통적으로 강화학습이 사용되지 않았던 이미지 분류 문제에 적용하려는 연구가 이루어지고 있다(Qiao *et al.*, 2018; Lin *et al.*, 2020; Stember *et al.*, 2022). 이미지 분류 문제는 딥러닝 분야에서 가장 기초적인 분야이며 실제 현장에서 유용하게 적용되고 있다. 예를 들어 제조현장에서 불량을 자동으로 탐지하기 위한 분류, 헬스케어 분야에서 질병의 판독을 위한 의료 이미지 분류 기술 등이 활발하게 사용되고 있다. 그러나 심층 강화학습 방법론을 활용하여 이미지 분류 문제를 해결하는 연구는 그리 많지 않은 상황이다. 그 이유는 에이전트가 상호작용 할 수 있는 환경 없이 주어진 데이터셋만 사용하여서는 강화학습이 가지는 장점을 제대로 활용하기 쉽지 않기 때문이다. 따라서 본 논문에서는 심층 강화학습 방법론을 사용하여 동일한 구조의 네트워크를 사용하였을 때 전통적인 학습 방법보다 높은 성능을 달성하는 것을 목표로 한다. 이를 위하여 강화학습 방법론을 적용할 수 있도록 강화학습의 필수 요소인 상태, 행동 그리고 보상을 적절하게 정의하고 기존 관련 연구의 한계점을 개선한 순차적인 의사결정 프로세스와 이미지 증강 기법을 도입하여 보다 효과적인 방법론을 제안하고자 한다. 본 연구의 주요 기여점은 아래와 같다.

- 심층 강화학습 방법론을 이미지 분류 문제에 적용하기 위해 강화학습에 필수적인 상태, 행동 그리고 보상 요소를 정의하였다.
- 3가지의 오픈 데이터셋(CIFAR-10, SVHN, WM-811K)을 활용한 실험을 통해 가장 높은 성능을 달성하여 이미지 분류 문제에 심층 강화학습 방법론의 적용 가능성을 확인하였다.
- 제안 방법론은 순차적인 의사결정 프로세스 및 이미지 증강 기법의 종류/세기 등을 자유롭게 변경할 수 있는 구조를 가지므로 데이터셋 및 도메인에 따라 다양하게 변형 적용이 가능하다.

본 논문의 구성은 다음과 같다. 제2장에서 배경 이론을 설명하였다. 제3장에서는 제안 방법론에 대하여 구체적으로 설명하였다. 제4장에서는 실험에 사용된 데이터와 실험 설계에 대하여 기술하였다. 제5장에서는 실험 결과를 정리하였고, 끝으로 제6장에서는 본 연구의 결론 및 향후 연구 방향을 서술하였다.

## 2. 배경 이론

### 2.1 강화학습과 Deep Q-Network

기존 강화학습은 벨만(Bellman) 방정식에 근간을 두고 있으며 식 (1)과 같이 표현할 수 있다. 이때 $s$는 상태, $a$는 행동을 의미하며 $s$상태에서 행동 $a$를 취하였을 때 도달하게 되는 상태를 $s'$, 즉시 얻게 되는 보상을 $r(s,a)$라 한다. $Q(s,a)$는 상태 $s$에서 행동 $a$를 취하였을 때 최종적으로 받게 되는 총 보상의 기대값(expected value)을 나타내며, 이를 Q-value라 한다. 마지막으로 $\gamma$는 할인율(discount factor)이며, 현재와 미래 가치에 대한 중요도를 조절하는 파라미터이다.

$$Q(s,a) = E[r(s,a) + \gamma \max_a Q(s',a)] \tag{1}$$

다시 정리하자면 현재 상태 $s$에서 행동 $a$를 취할 때 받을 수 있는 모든 보상의 총 합 $Q(s,a)$는 현재 행동을 취해서 즉시 받을 수 있는 보상 $r(s,a)$와 미래에 받을 것으로 예상되는 미래 보상의 최대값 $\max_a Q(s',a)$의 합으로 계산할 수 있다. 이 값을 최대화 하도록 최적의 행동을 선택하는 것이 에이전트의 목표이다. 여러 가정하에 식 (1)을 활용하여 실제 Q-value 값을 계산할 수 있는데 재귀적인 성질을 활용하여 초기에는 부정확한 값으로 추측하더라도 최종적으로는 정답 값에 수렴하게 된다. 이와 같은 학습 방식을 Q-러닝(Q-learning) 알고리즘이라고 하며 모든 가능한 상태-행동 조합을 테이블 형태로 저장하고 값을 갱신하는 방법으로, 상태와 행동이 유한한 경우에만 적용 가능하다는 한계를 가지고 있다. 따라서 이러한 고전적인 강화학습의 한계를 극복하기 위해 인공 신경망을 활용한 심층 강화학습 방법론이 제안되었다. 심층 강화학습은 인공 신경망을 적용하여 상태와 행동이 무한한 경우에도 사용할 수 있게 되었으며, 이로써 학습을 효율적으로 수행할 수 있고 매우 복잡한 상태-행동 공간을 가지는 문제도 다룰 수 있게 되었다(Jang *et al.*, 2019). 그 중에서도 Mnih *et al.*(2013)의 연구 Deep Q-Network(DQN)는 무한한 상태 공간과 유한한 행동 공간을 다룰 수 있어 가장 널리 사용되는 대표적인 심층 강화학습 알고리즘이다. DQN 방법론의 인공 신경망은 비용함수를 사용하여 학습이 진행되는데, 식 (1)에서 최종적으로 수렴하여 좌변과 우변이 같아지게 되는 것이 목적이므로 식 (1)의 좌변과 우변의 차이로 정의되는 비용함수를 최소화 하는 방향으로 모델을 학습할 수 있다. 이를 수식으로 표현하면 식 (2)와 같다.

$$Cost = E\left[\left\{r(s,a) + \gamma \max_a Q(s',a;\theta) - Q(s,a;\theta)\right\}^2\right] \tag{2}$$

이때 $\theta$는 네트워크의 가중치(weight) 뜻하며 $Q(s,a;\theta)$는 상태 $s$에서 행동 $a$를 취할 때 네트워크에 의하여 출력되는 Q-value 값을 뜻한다. 이와 같은 DQN은 이해하기 쉽고 직관적이라는 장점을 가지고 있기 때문에 DQN을 기반으로 다양한 방법론들이 연구되고 있다(AlMahamid *et al.*, 2021). 본 논문에서는 기본적인 DQN 구조에 재현 메모리(replay buffer)와 목표 네트워크(target network)를 추가한 모델을 사용하였다. 재현 메모리는 강화학습의 에이전트가 과거의 경험을 기억하고 재사용해서 학습하게 함으로써 높은 시간 상관관계(temporal

correlation)를 가지는 문제와 과거의 경험을 빠르게 잊는 문제를 극복하기 위해 사용된다. 목표 네트워크는 식 (2)에서 현재 네트워크가 가고자 하는 방향인 $\max_a Q(s', a; \theta)$ 값이 지나치게 변동하여 학습 안정성이 떨어지는 것을 방지하기 위해 고안되었다. 기존 네트워크와 동일한 구조를 가지지만 가중치를 고정하고 일정 주기마다 최신으로 업데이트 하는 방법을 통해 목표 값의 변동을 줄여 학습 안정성을 높인다. 재현 메모리와 목표 네트워크가 추가된 최종 비용 함수는 식 (3)과 같고 학습 흐름을 <Figure 2>와 같이 표현하였다. $\hat{Q}$는 목표 네트워크의 Q-value이며 $\hat{\theta}$은 목표 네트워크의 가중치이다.

$$Cost = E\left[\left\{r(s, a) + \gamma \max_a \hat{Q}(s', a; \hat{\theta}) - Q(s, a; \theta)\right\}^2\right] \quad (3)$$

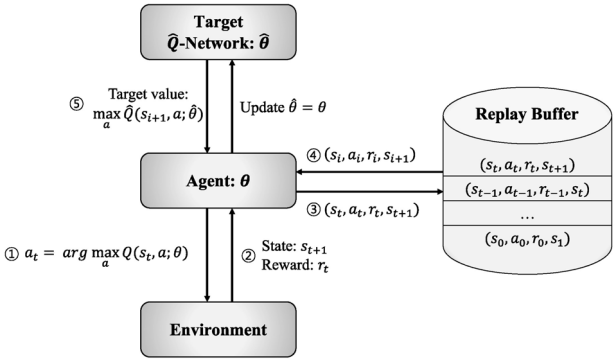



**Figure 2.** Learning Process of Deep Q-Network. The replay buffer stores the experiences (state, action, reward and next state). The agent uses this stored data to learn and periodically updates the parameters of the target network.

### 2.2 심층 강화학습을 활용한 이미지 분류

심층 강화학습은 최근 높은 활용성을 보이고 있으며 심층 강화학습을 활용하여 이미지 분류 문제를 해결하려는 연구 또한 이루어지고 있다. 이러한 심층 강화학습은 이미지에서 보다 본질적인 특징을 추출하고 기존의 전통적인 방법으로는 쉽게 해결할 수 없는 문제를 해결하는 것을 목적으로 한다. Qiao *et al.*(2018)은 손 글씨를 이미지로 표현한 MNIST(modified national institute of standards and technology) 데이터셋의 분류 문제를 다루었다. 오토인코더(autoencoder) 구조를 활용하여 이미지로부터 특징을 추출하고 추출된 특징을 에이전트의 상태로 정의하여 Q-러닝 알고리즘으로 해당 이미지가 어느 클래스에 속하는지를 출력하게 된다. 이후 클래스를 맞추면 1의 보상을 받고, 틀릴 경우 -0.5의 보상을 받아 모델의 학습을 유도한다. 이러한 방식은 오토인코더를 통하여 노이즈에 강건한 특징을 추출하고 추출된 특징을 강화학습 방법론을 통하여 학습함으로써 효과적인 학습이 진행된다. 저자는 MNIST 데이터셋에 심층 강화학습을 최초로 적용하여 높은 성능과 효율성을 가진다고 주장한다. 하지만 해당 연구에는 강화학습의 핵심 요소인 순차적인 의사결정 과정이 없다. 모델은 입력 받은 특징을 바탕으로 어떤 클래스에 속하는지 출력하고 출력한 결과는 보상으로만 받으며 더 이상의 프로세스는 진행되지 않는다. 이는 끝없이 환경과 상호작용하며 모델이 선택한 행동이 어떤 변화를 일으켰는지를 관측하지 않는다는 점에서 한계가 존재한다. 다음으로 Lin *et al.*(2020)은 심층 강화학습을 활용하여 클래스간 데이터가 불균형한 경우 적합한 방법론인 DQNimb를 제안하였다. 강화학습의 필수 요소인 상태, 행동 그리고 보상은 앞선 연구(Qiao *et al.*, 2018)와 동일하게 설정하였다. 저자는 보상의 크기를 조절함으로써 클래스 불균형을 간단하게 해결할 수 있다고 주장했다. 주류 클래스를 맞췄을 때의 보상은 적게, 비주류 클래스를 맞췄을 때는 보상을 크게 설정하여 모델이 비주류 클래스 데이터에 보다 집중할 수 있도록 설계하였다. 해당 연구는 단순히 보상의 크기만 조절함으로써 인위적인 이미지의 증감 없이 클래스 불균형을 간단하게 해결하였다는데 의의가 있으며 이를 <Figure 3>으로 도식화 하였다.

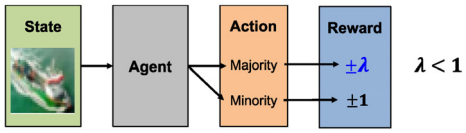



**Figure 3.** Overall Structure of DQNimb(Lin *et al.*, 2020). The model achieves highest performance when lambda is set equal to the imbalanced ratio (i.e., $\rho = \frac{|D_{\text{minority}}|}{|D_{majority}|}$) where $D_{dass}$ is the set of samples each class.

하지만 클래스를 주류 및 비주류 클래스로 구분하여 단순 이진분류 문제로 변환하여 실험을 수행한 점은 한계점으로 클래스가 3개 이상인 경우 성능이 하락하는 문제를 가진다. 또한 앞선 관련 연구와 동일하게 순차적인 의사결정 프로세스가 없다는 한계점을 가진다. 마지막으로 Stember *et al.*(2022) 연구는 한번 이미지가 입력되면 여러 단계에 걸쳐 순차적인 의사결정 과정을 가지는 차별점을 가진다. 하지만 클래스를 맞추면 초록색 음영을 추가하고, 틀릴 경우 빨간색 음영을 추가하여 다음 단계를 시작하는 매우 단순한 구조를 가지고 있다. 이러한 구조로 학습을 지속할 경우 주어진 데이터에 정답이 표현되기 때문에 정답 유출의 문제가 발생할 수 있고 동일한 이미지를 반복해서 학습하므로 과적합에 우려가 있다. 결과적으로 이러한 관련 연구들의 한계점을 본 논문에서 극복하고자 하였으며 순차적인 의사결정 프로세스와 이미지 증강 기법의 도입으로 이를 해결하고자 하였다. 자세한 사항은 3장에서 서술하였다.

## 3. 제안 방법론

본 장에서는 심층 강화학습 알고리즘을 적용한 제안 방법론의 전체적인 구조를 설명하고, 제안 방법론의 핵심 요소인 순차

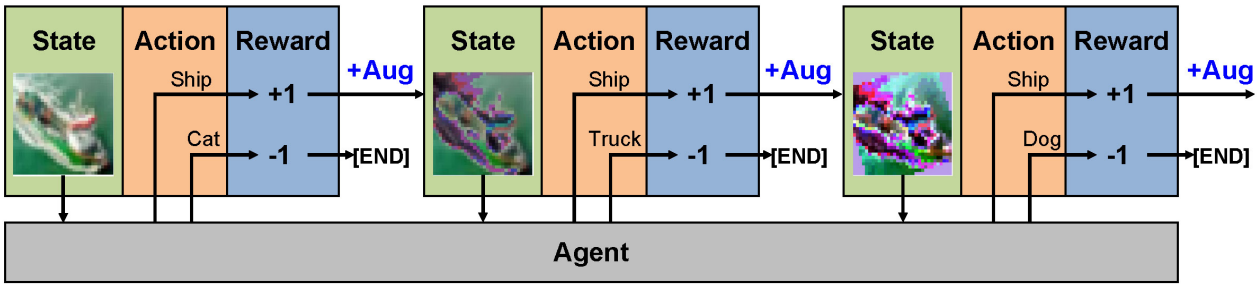



**Figure 4.** Overview of the Proposed Method. The image from the dataset is defined as a state, each class as an action, and the reward is defined as a numeric value according to the classification results. Successful classifications yield positive rewards, prompting the re-augmentation and re-input of the image into the agent. However, if the agent provides an incorrect answer, the episode concludes. The classification process is iterated up to 10 times within a single episode.

적인 의사결정 프로세스와 이미지 증강 기법에 대하여 구체적으로 설명한다.

### 3.1 제안 방법론 전체 구조

강화학습 방법론은 에이전트와 환경이 서로 상호작용하며 학습이 진행되므로 강화학습의 필수 요소인 상태, 행동 그리고 보상을 적절하게 정의하는 것이 필수적이다. 본 연구에서는 일반적인 강화학습과는 다르게 실시간으로 상호작용 할 수 있는 환경 없이 주어진 데이터셋을 활용하여야 하므로 필수 요소를 정의하는데 보다 어려움이 따른다. 따라서 기본적인 구조는 관련 연구에서 착안하여 상태를 데이터셋의 이미지, 행동을 클래스, 그리고 보상을 정답 유무에 따른 수치 값으로 정의하였다. 이는 매우 간단하고 직관적인 구조를 가진다. 하지만 관련 연구(Qiao *et al.*, 2018; Lin *et al.*, 2020; Stember *et al.*, 2022)들은 에이전트가 선택한 행동과 그 행동으로부터 얻게 되는 새로운 정보를 고려하지 못한다. 따라서 모델이 행동을 취하는 것은 모두 독립적으로 수행되고 결과적으로 기존 전통적인 방법들과 큰 차이를 보이지 못한다. 본 연구에서는 에이전트가 행동을 취하고 그 행동의 결과에 따라서 학습의 방향이 변화하는 것을 의도하였다. 제안 방법론은 에이전트가 행동을 선택하고 분류에 성공할 경우 주어진 이미지를 증강하여 다시 에이전트에게 입력함으로써 단 한 번의 분류 후 학습이 종료되는 것을 방지하였다. 에이전트는 다시 입력된 증강된 이미지로 분류를 수행하며 정답을 맞출 경우 또다시 증강된 이미지를 입력 받고, 정답을 맞추지 못할 경우 해당 에피소드(episode)가 종료된다. 이러한 제안 방법론의 구조는 이미지 증강의 기회가 최대가 되도록 설계하였으며 <Figure 4>에 도식화 하였다. 이는 사람이 학습하는 방식과도 유사하게 구성되었다. 즉, 쉬운 문제를 맞추면 더 어려운 문제를 제공 받아 학습을 지속하고, 만약 문제를 틀릴 경우 다시 쉬운 문제를 접하게 하는 방식으로 학습이 진행된다. 이때 에피소드는 하나의 이미지로부터 파생된 학습 과정을 뜻하며 하나의 에피소드에서 계속하여 정답을 맞출 경우 최대 10회까지 반복하여 점차 어려운 이미지를 학습 할 수 있도록 설계하였다. 그리고 분류 과정에서 발생하는 성공/실패 경험은 재현 메모리에 저장되어 학습에 활용된다. 이러한 순차적인 의사결정 프로세스를 3.2절에서 보다 상세히 설명하였다. 또한 점진적으로 어려운 이미지를 만드는 방법은 이미지 증강 기법을 중첩 적용하여 활용하였고 보다 상세한 설명은 3.3절에 서술하였다.

### 3.2 순차적인 의사결정 프로세스

본 연구에서는 에이전트가 클래스 분류에 성공할 경우 에피소드가 종료되지 않고 증강된 이미지가 에이전트로 재입력되어 최대 10회까지 진행하고, 분류에 실패할 경우 즉시 에피소드를 종료하고 다음 이미지로 동일 프로세스를 진행하도록 구성하였다. 기본적인 강화학습은 초기에는 다양한 행동을 수행하며 탐색하는 과정을 거치다가 점차 학습이 진행될수록 높은 보상을 얻을 수 있는 행동에 집중하게 된다. 따라서 이와 유사한 환경을 만들고자 하였고 정확히 분류하여 보상을 얻은 경우 에피소드를 계속 진행하고, 분류에 실패할 경우 에피소드를 즉시 종료하여 재현 메모리에 정답을 맞춘 사례가 더욱 많이 저장될 수 있도록 하였다. 결과적으로 재현 메모리에는 다양하게 증강된 이미지가 저장되므로 이를 통해 학습한 에이전트는 이미지에서 본질적인 특징을 추출할 수 있게 된다. 각 이미지 별로 앞서 서술한 학습 진행 상황을 <Figure 5>에 도식화 하였다.

이러한 순차적인 의사결정 과정은 매우 다양하게 설계가 가능하다는 장점이 있다. 예를 들어 에이전트가 분류에 성공할 경우 에피소드를 즉시 종료하고, 분류에 실패하였을 경우 이미지를 약간 변형하여 다시 재입력하는 과정을 반복하여 실패 사례를 재현 메모리에 많이 쌓을 수도 있다. 뿐만 아니라 학습의 진행 정도에 맞추어 이미지 증강의 세기 또는 횟수를 조절할 수도 있다. 이처럼 자유로운 구성이 가능하므로 풀고자 하는 도메인과 데이터셋에 맞게 최적화할 수 있다는 장점이 있다. 본 연구에서는 다양한 환경을 구성하여 실험하였고 가장 성능이 높은 환경을 제안 방법론으로 제시하였다.

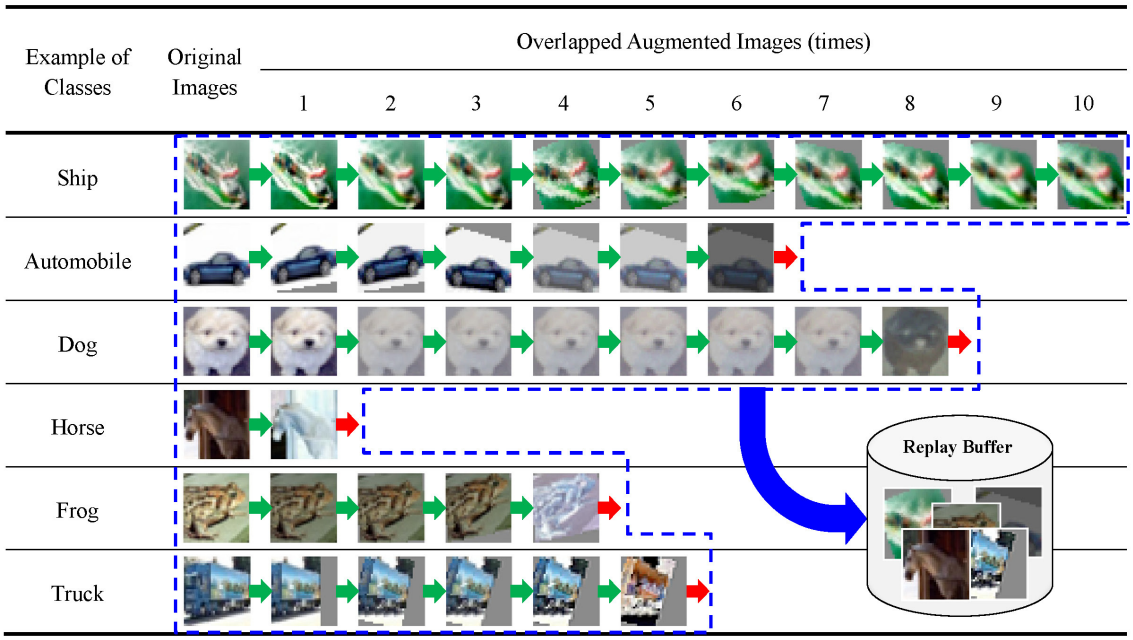



**Figure 5.** Continuous Decision-making Process Using CIFAR-10 Dataset. A green arrow indicates successful classification, while a red arrow indicates the agent's failure to classify. Upon successful classification, the image undergoes augmentation, facilitating continuous learning. However, if classification fails, the episode concludes, and the next episode begins. Data generated through this process is stored in the replay buffer and utilized for subsequent learning.

### 3.3 이미지 증강 기법

순차적인 의사결정 프로세스를 거치며 에이전트는 보다 어려운 이미지를 입력 받게 된다. 이때 어려운 이미지는 증강기법이 여러 번 중첩된 이미지로 정의하였다. 에이전트가 이미지 분류에 성공하여 주어진 이미지를 증강할 때 사전 정의한 증강 기법 중 무작위로 선택하였으며 클래스 보전을 위하여 원본을 크게 훼손하지 않는 범위에서 증강의 세기를 무작위 적용하였다. 이미지 증강 기법이 여러 번 중첩 적용되며 보다 분류가 어려운 이미지가 생성되는 과정을 <Figure 6>에 도식화 하였다. <Figure 6>에서 이미지 밑에 각 픽셀이 가지는 값을 히스토그램으로 표시하였고, 원본 이미지로부터 점차 증강되어 우측으로 이동할수록 픽셀 값의 분포가 정보를 잃어가는

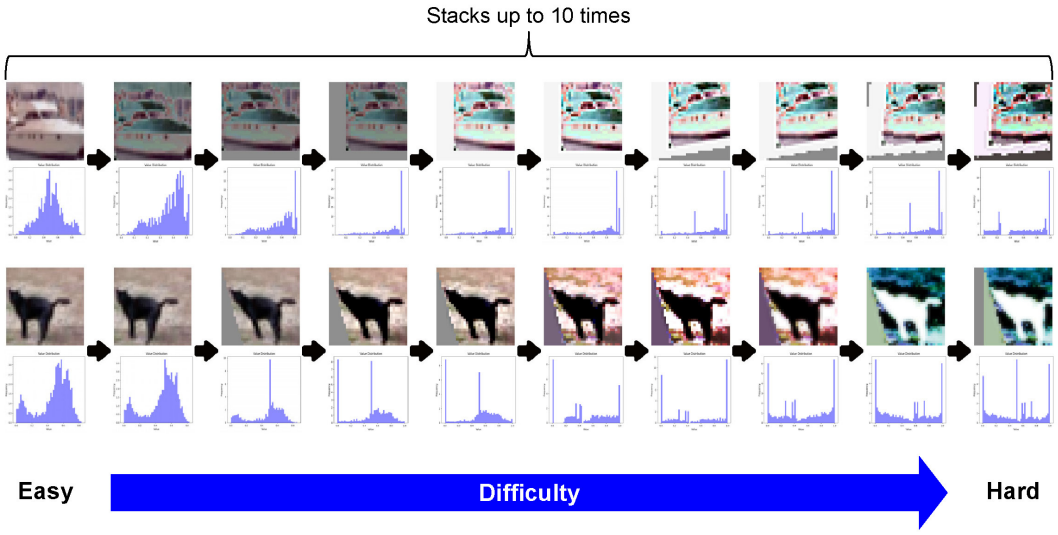



**Figure 6.** Example of the Process of Creating Difficult Images Using the CIFAR-10 Dataset. The pixel values of each image are presented as a histogram. As the image undergoes incremental augmentation from its original state, the corresponding histogram depicting pixel values progressively loses information.

과정을 확인할 수 있다. 따라서 증강 기법이 중첩 적용되면 원본 이미지와 분포가 달라지며 정보를 잃기 때문에 보다 어려운 이미지라고 할 수 있다. 결과적으로 원본부터 10번 중첩 증강된 이미지까지 모델이 경험할 수 있게 되며 이는 과적합을 방지하고 각 클래스의 주요한 특징을 추출할 수 있으므로 더욱 강건한 모델이 생성되게 된다.

본 논문에서 사용한 이미지 증강 기법과 세기는 Cubuk *et al.*(2019)의 연구를 참고하였다. 해당 연구는 최적의 이미지 분류 성능을 달성하기 위하여 데이터셋을 어떻게 증강하여야 하는지를 강화학습 방법론으로 탐색하는 연구이다. 해당 연구에서 사용한 기법 외에도 다양한 이미지 증강 기법이 연구되고 있지만(Xu *et al.*, 2023) 단일 이미지만 사용하여 간단하고 빠르게 적용할 수 있는 기법들이므로 본 연구에 적합하다고 판단하였다. 그리고 본 논문에서 WM-811K 데이터셋에 사용한 증강 기법은 다른 데이터셋과는 다소 다르게 적용하였다. 이는 해당 데이터셋이 RGB 채널별로 원-핫 인코딩(one-hot encoding)되어 있으므로 특정 증강 기법(invert, solarize, equalize, autocontrast)을 적용할 시 이미지의 변화가 없거나 클래스 변경의 여지가 있기 때문이다. 따라서 해당 기법은 삭제하고 증강 기법의 다양성을 위하여 3개를(zoom, flip, mirror) 추가하여 사용하였다. Cubuk *et al.*(2019)에서 제시한 이미지 증강 기법과 본 연구에서 사용한 증강 기법을 <Table 1>과 <Table 2>와 같이 표시하였다. <Table 1>, <Table 2>에서 대괄호는 해당 범위 안에서 값을 무작위로 가질 수 있음을 의미한다.

**Table 1.** Image Augmentation Techniques Applied in the Cubuk *et al.*(2019). The listed operations are derived from the python imaging library(PIL) and two additional techniques(cutout, sample pairing) were added. A blank entry in "Range of Magnitudes" column is a techniques that does not require a magnitude value.

| Operation Name | Range of Magnitudes |
|---|---|
| ShearX(Y) | [-0.3, 0.3] |
| TranslateX(Y) | [-150, 150] |
| Rotate | [-30, 30] |
| Solarize | [0, 256] |
| Posterize | [4, 8] |
| Contrast | [0.1, 1.9] |
| Color | [0.1, 1.9] |
| Brightness | [0.1, 1.9] |
| Sharpness | [0.1, 1.9] |
| Cutout | [0, 60] |
| Sample Pairing | [0, 0.4] |
| AutoContrast | |
| Invert | |
| Equalize | |

**Table 2.** Image Augmentation Techniques used in this Study. Variations in augmentation techniques were applied to each dataset because of potential differences in image characteristics or class variations.

| CIFAR-10 | | SVHN | | WM-811K | |
|---|---|---|---|---|---|
| Operation Name | Range of Magnitudes | Operation Name | Range of Magnitudes | Operation Name | Range of Magnitudes |
| ShearX(Y) | [-0.3, 0.3] | ShearX(Y) | [-0.3, 0.3] | ShearX(Y) | [-0.3, 0.3] |
| TranslateX(Y) | [-150, 150] | TranslateX(Y) | [-150, 150] | TranslateX(Y) | [-150, 150] |
| Rotate | [-30, 30] | Rotate | [-30, 30] | Rotate | [-30, 30] |
| Solarize | [0, 256] | Solarize | [0, 256] | Posterize | [4, 8] |
| Posterize | [4, 8] | Posterize | [4, 8] | Contrast | [0.1, 1.9] |
| Contrast | [0.1, 1.9] | Contrast | [0.1, 1.9] | Color | [0.1, 1.9] |
| Color | [0.1, 1.9] | Color | [0.1, 1.9] | Brightness | [0.1, 1.9] |
| Brightness | [0.1, 1.9] | Brightness | [0.1, 1.9] | Sharpness | [0.1, 1.9] |
| Sharpness | [0.1, 1.9] | Sharpness | [0.1, 1.9] | Zoom | [-3, 3] |
| AutoContrast | | AutoContrast | | Flip | |
| Invert | | Invert | | Mirror | |
| Equalize | | Equalize | | | |

## 4. 실험 설계

### 4.1 데이터 소개 및 전처리

본 연구에서는 이미지 분류에 널리 사용되고 있는 CIFAR-10 (canadian institute for advanced research), SVHN(street view house numbers)과 Wu *et al.*(2014)가 공개한 WM-811K 데이터셋을 사용하였다. 전통적인 학습 방법과의 성능 비교가 본 논문의 목적이므로 이미지 분류 분야에서 가장 널리 사용되어지는 데이터셋인 CIFAR-10과 SVHN을 선정하였다. 또한 산업 현장에서 제안 방법론의 적용 가능성을 확인하기 위하여 공개되어 있는 WM-811K 데이터셋을 추가로 활용하여 실험을 진행하였다. 데이터셋 별 클래스와 샘플 수, 데이터 분리 비율 등의 간략한 개요는 <Table 3>과 같다.

첫 번째 데이터셋인 CIFAR-10은 비행기, 자동차, 새, 고양이, 사슴, 강아지, 개구리, 말, 배, 트럭의 10가지 클래스로 구성되어 있으며, 픽셀당 0~255까지의 범위를 가지는 숫자가 32×32 크기의 2차원 컬러 이미지로 구성되어 있다. 총 60,000개의 이미지로 구성되어 있으며 이 중에서 50,000장은 학습용 데이터, 10,000장은 테스트용 데이터로 구분되어 있다. 본 실험에서는 검증을 위해서 50,000장의 학습데이터를 다시 재분할하여 검증용 데이터셋으로 구분하였다. 최종적으로 학습, 검증 그리고 테스트 데이터셋의 비율은 4:1:1이 되도록 설정하였다. CIFAR-10 데이터셋의 샘플 이미지는 <Figure 7>과 같다.

두 번째 데이터셋인 SVHN은 실제 거리에서 주택 주소를 촬영한 이미지 데이터셋으로 한 장의 이미지 안에 여러 개의 숫자로 이루어져 있으며 다양한 크기의 이미지로 구성되어 있다. 본 실험에서는 편의를 위하여 32×32 크기로 제공하는 Format 2를 사용하였다. Format 2는 인식 대상 숫자가 가운데에 위치하도록 전처리가 이루어져 있으며 <Figure 8>에서 샘플을 확인할 수 있다. 하나의 이미지에 표시되는 여러 숫자 중 가운데에 위치한 숫자로 레이블링 되어있으며 0부터 9까지 총 10개의 클래스를 가진다. 데이터는 총 99,289장으로 이루어져

**Table 3.** Brief Overview of Each Dataset

| Dataset | Number of Classes | Number of Samples (Data Separation Ratio) | Channel | Image Size |
|---|---|---|---|---|
| CIFAR-10 | 10 | 60,000 (Train : Valid : Test = 4 : 1 : 1) | RGB | 32x32 |
| SVHN | 10 | 99,289 (Train : Valid : Test ≈ 9 : 1 : 3.5) | RGB | 32x32 (Format2) |
| WM-811K | 8 | 25,519 (Train : Valid : Test ≈ 9 : 1 : 2.5) | RGB (One-hot-encoded) | 64x64 (Resized) |

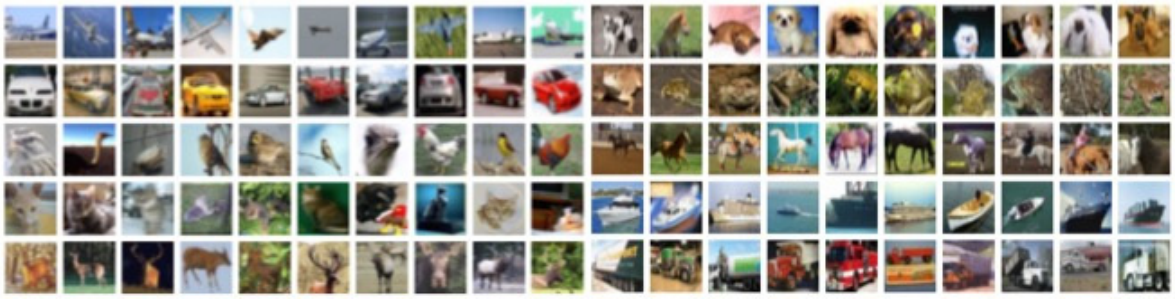


**Figure 7.** Example of CIFAR-10 Dataset. There are 10 classes and completely mutually exclusive classes: Airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck

**Figure 8.** Example of SVHN Dataset in which the Numbers are Located in the Center of the Image. Classification is based on identifying the digit located in the middle among those represented in the image

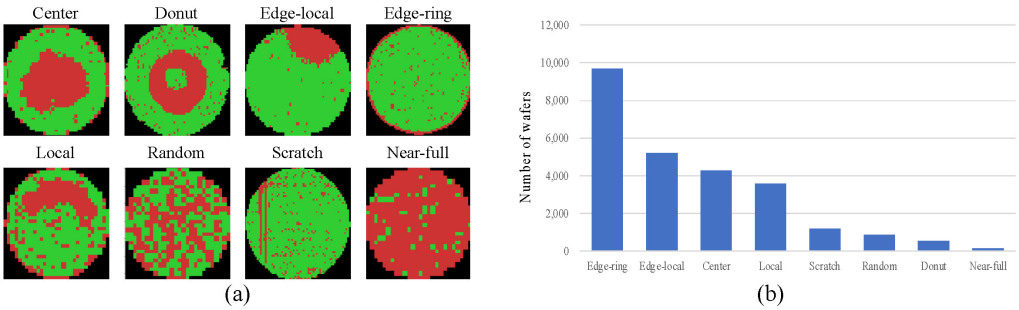



**Figure 9.** Example of WM-811K Image Dataset. (a) Defect patterns in wafer images. The red color is a defect chip and the green color is a normal chip. (b) A bar chart illustrating the class imbalance, with Edge-ring having 65 times more data then Near-full defect

있으며 학습, 검증 그리고 테스트를 위하여 약 9:1:3.5의 비율로 분리하여 실험을 진행하였다.

마지막으로 사용한 데이터셋인 WM-811K은 반도체 공정 중 완성된 웨이퍼 내 반도체칩이 정상적으로 작동하는지 전기적인 검사를 수행하고 검사 결과를 웨이퍼 빈 맵(wafer bin map)으로 표시한 이미지 데이터셋이다. 웨이퍼 빈 맵에는 불량 여부에 따른 검사 결과가 원-핫 인코딩의 형태로 저장되어 있으며 다양한 형태의 불량 패턴이 나타난다. 이러한 불량 패턴은 엔지니어가 해당 불량의 원인을 분석하고 판단하는데 중요한 역할을 수행하기 때문에 웨이퍼 빈 맵의 패턴 분류 문제는 매우 중요하다. WM-811K 데이터셋 또한 다양한 크기의 이미지로 구성되어 있으므로 실험을 위하여 64×64 크기로 이미지 사이즈를 조정하였고, 총 811,457장의 이미지 중 명확하게 불량 패턴 레이블이 있는 25,519장을 선별하여 사용하였다. 사용된 이미지는 불량 패턴의 형태에 따라 센터(center), 도넛(donut) 등 총 8개의 클래스로 구성되어 있으며 클래스 예시를 <Figure 9(a)>에서 표현하였다. 또한 불량 패턴별로 최대 65배의 차이가 있는 심각한 클래스 불균형 문제를 가지고 있으므로 실제 현장에서의 적용 가능성을 확인하는데 보다 적합하다. 클래스 별 불균형 정보는 <Figure 9(b)>에서 바 차트로 표현하였다. 마지막으로 앞선 데이터셋과 동일하게 검증을 위하여 데이터셋을 분리하여 사용하였으며 학습, 검증 그리고 테스트 데이터셋의 비율은 약 9:1:2.5로 구성하였다.

### 4.2 합성곱 신경망

본 연구에서 사용한 네트워크는 합성곱 신경망(convolution neural networks, CNN) 기반으로 구성하였다. 총 4개의 컨볼루션 레이어(convolution layer)를 사용하였으며, 사용된 커널(kernel) 사이즈는 3×3을 사용하였고, 각각 64에서 512까지의 채널을 갖는다. 또한 클래스별 출력을 제외한 2개의 완전 연결층(fully-connected layer)을 사용하였다. 활성화 함수(activation function)는 ReLU(rectified linear unit)를 사용하였으며 과적합 방지를 위하여 드롭아웃(dropout, dropout rate $p=0.2$) 기법을 사용하였다. 네트워크의 전체적인 구조는 <Figure 10>과 같이 표현하였다. 파란색 음영은 컨볼루션 레이어 부분이며, 초록색 음영은 완전 연결층을 표시하였다. 본 연구에서는 동일한 구조의 네트워크를 사용하였을 때 전통적인 학습 방법과 심층 강화학습 방법론의 성능을 비교하는 것이 목적이므로 비교적 깊지 않은 네트워크 구조를 사용하였다.

### 4.3 평가지표

학습 방법론 별 성능 비교를 위하여 정확도(accuracy)와 매크로 평균(macro average) F1-score를 평가 지표로 사용하였다. 정확도는 클래스 분류 문제에서 널리 사용되는 지표이기 때문에 CIFAR-10과 SVHN 데이터셋을 활용한 결과를 비교할 때 적합하다. 하지만 WM-811K와 같이 클래스별 불균형이 심한

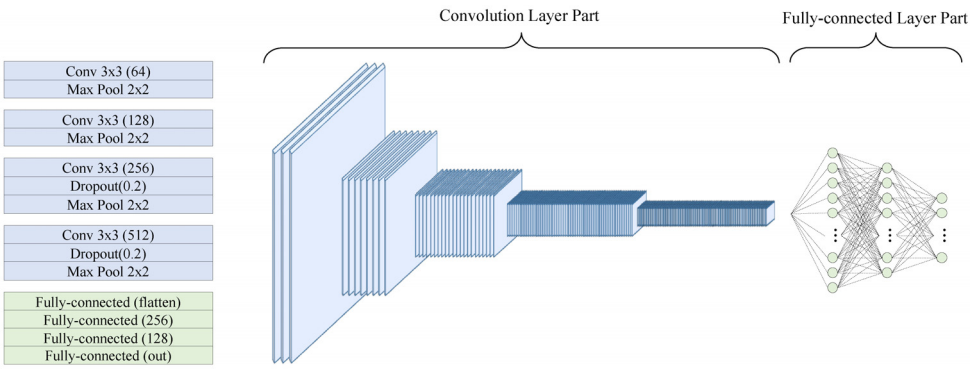



**Figure 10.** Overview of CNN Configuration

경우 단순 정확도만으로는 모델의 성능을 비교하기에는 어려움이 따른다. 특히 산업현장에서는 불량 또는 이상 데이터의 개수가 극히 드물기 때문에 데이터의 수를 고려한 평가 지표가 필수적으로 요구 된다. 매크로 평균 F1-score는 클래스별 F1-score를 산술평균한 값으로써 데이터의 개수가 적은 클래스의 성능이 데이터 개수가 많은 클래스에 가려져 성능 평가가 제대로 되지 않는 단점을 보완한 평가 지표이다. 클래스별 F1-score는 식 (4)와 같으며 이를 이용하여 매크로 평균 F1-score를 식 (5)와 같이 표현할 수 있다. 따라서 본 연구에서는 매크로 평균 F1-score를 평가지표로 추가하여 보다 공정하게 모델의 성능을 비교 하고자 하였다.

$$F1\ score_{class} = (2 \times (precision_{class} \times recall_{class}) \\ / (precision_{class} + recall_{class}) \quad (4)$$

$$Macro\ averaged\ F1\ score = (\sum_{class} F1score) \\ / (number\ of\ classes) \quad (5)$$

### 4.4 실험 조건

공통적으로 모델의 손실 함수는 크로스 엔트로피(cross entropy)를 사용하였고 최적화 알고리즘은 Adam(adaptive moment estimation)을 사용하였다. 그리고 비교군인 전통적인 학습방법의 최적 성능 도출을 위하여 학습률(learning rate)을 변경하며 실험을 진행하였다. 배치 사이즈(batch size)는 32로 고정하였으며 실험 진행 중 더 이상의 성능 개선이 없다면 조기 종료(early stop)하도록 설정하였다. 다음으로 심층 강화학습은 앞서 방법론과 동일한 배치 사이즈를 사용하였으며, 에피소드 외에도 추가 하이퍼파라미터(hyperparameter)를 다양하게 변경하며 실험을 수행하였다. 모든 실험은 무작위로 seed를 변경하며 10회 반복 후 평균값을 사용하였으며 전통적인 방법론의 하이퍼파라미터 설정은 <Table 4>, 제안 방법론의 세부적인 하이퍼파라미터 설정은 <Table 5>와 같다. 각 항목의 대괄호 내의 범위에서 값을 변경하며 실험을 수행하였고, 괄호가 없는 하이퍼파라미터는 해당 값으로 고정하여 실험을 수행하였다.

**Table 4.** List of Hyperparameters for Traditional Learning Method. The experiment was conducted by changing the values in the corresponding section in parentheses.

| Hyperparameters | Values |
|---|---|
| Epoch | 100 |
| Learning Rate | [0.0001, 0.001] |
| Batch Size | 32 |
| Patience(Early Stopping) | [3, 10] |
| Delta(Early Stopping) | 0 |

**Table 5.** List of Hyperparameters for the Proposed Method

| Hyperparameters | Values |
|---|---|
| Episodes | [1,000,000, 6,000,000] |
| Learning Rate | [0.0001, 0.001] |
| Batch Size | 32 |
| Reward | [±1, ±10] |
| Gamma | [0.80, 0.98] |
| Start Value (e-greedy) | [0.1, 0.8] |
| End Value (e-greedy) | [0.01, 0.8] |
| Decrement (e-greedy) | 0.001 |
| Decay (e-greedy) | 1000 |
| Replay Buffer Size | [50,000, 200,000] |
| Alpha (Prioritized Replay) | [0, 1] |
| Beta (Prioritized Replay) | [0, 1] |
| Beta Increment (Prioritized Replay) | $[2.5e^{-7}, 1e^{-5}]$ |
| Target Network Update Frequency | 500 |

## 5. 실험결과

### 5.1 비교 실험 결과

전통적인 학습방법과 심층 강화학습 방법론의 성능 비교 결과는 아래 <Table 6>과 같다. 성능이 가장 높은 항목을 진하게,

**Table 6.** Comparison of Performance between Traditional Learning Methods and the Proposed Method in Terms of Accuracy and F1-score Across CIFAR-10, SVHN, and WM-811K

| Methods \ Datasets | | | CIFAR-10 | | SVHN | | WM-811K | |
|---|---|---|---|---|---|---|---|---|
| | | | Accuracy | F1-score | Accuracy | F1-score | Accuracy | F1-score |
| Traditional Learning Method | Vanilla | | 0.7707 (0.0076) | 0.7710 (0.7710) | 0.9275 (0.0017) | 0.9207 (0.0018) | 0.9323 (0.0048) | 0.8976 (0.0119) |
| | Augmented Image | 100% | 0.7896 (0.0065) | 0.7895 (0.0065) | 0.9379 (0.0028) | 0.9320 (0.0033) | 0.9304 (0.0071) | 0.8979 (0.0107) |
| | | 1,000% | 0.8314 (0.0026) | 0.8316 (0.0023) | 0.9587 (0.0010) | 0.9552 (0.0012) | 0.9434 (0.0031) | 0.9203 (0.0080) |
| Deep Reinforcement Learning | DQNimb (Lin *et al.*, 2020) | | 0.7367 (0.0066) | 0.7361 (0.0063) | 0.9085 (0.0039) | 0.9008 (0.0039) | 0.9152 (0.0076) | 0.8801 (0.0121) |
| | Proposed Method | | **0.8806 (0.0024)** | **0.8798 (0.0026)** | **0.9608 (0.0015)** | **0.9575 (0.0018)** | **0.9463 (0.0028)** | **0.9255 (0.0066)** |

표준편차를 괄호로 표시하였다. 모든 데이터셋에서 제안 방법론의 성능이 가장 높았으며 클래스별 샘플 수에 따라 보상의 차등을 둔 DQNimb(Lin *et al.*, 2020) 방법론을 능가하는 결과를 확인할 수 있었다. 이는 제안하는 방법론의 순차적인 의사결정 구조와 이미지 증강 기법이 효과적으로 적용되었음을 보여준다. 제안 방법론은 이미지 증강 기법을 사용하였기 때문에 공정한 비교를 위하여 전통적인 학습 방법에서도 이미지를 증강하여 실험을 수행하였다. 학습에 주어진 이미지를 모두 무작위로 증강하여 원본 데이터와 합쳐 총 2배수의 데이터로 실험을 수행하였고, 그 결과를 <Table 6>의 두 번째(augmented image-100%) 행에서 확인할 수 있다. 그리고 주어진 데이터셋을 증강 기법을 최대 10번까지 중첩 적용하며 총 10배수의 데이터를 새로 생성하였고 기존 데이터와 합쳐 실험을 수행한 결과를 <Table 6>의 세 번째(augmented image-1,000%) 행에 수록하였다. 이처럼 기존 방법은 학습전에 증강된 이미지를 모두 사전 생성하여 제공하여야 한다는 단점이 존재한다. 이미지를 몇 장 증강하여 기존 데이터셋에 추가시킬 것인지를 결정하는 것은 매우 어렵고 시간이 많이 소요되는 작업이기 때문이다. 반면에 제안 방법론은 모델의 학습정도에 따라서 동적으로 증강된 이미지의 양을 모델이 스스로 달리하여 학습할 수 있다는 장점을 가진다.

### 5.2 구성 요소 별 성능 기여도 평가

본 연구에서 제안하는 순차적인 의사결정 구조와 중첩 이미지 증강 기법의 기여 정도를 확인하기 위하여 구성 요소별 성능 기여도 평가를 수행하였고 그 결과는 <Table 7>과 같다. <Table 7>에서 핵심 구성 요소를 전혀 사용하지 않은 첫 번째 행은 관련 연구인 DQNimb(Lin *et al.*, 2020)와 유사하나 클래스 별 샘플 수에 따라 보상의 크기를 조절하지 않았다는 점에서 <Table 6>의 네 번째 행의 결과와는 다소 차이가 있다. 주요 구성 요소를 하나씩 추가하며 실험을 수행한 결과 단순히 순차적인 의사결정 구조를 적용한 경우 성능 향상의 폭이 크지 않은 것을 확인할 수 있다. 이미지 증강 기법 또한 중첩없이 적용할 경우 성능 향상의 폭이 크지 않거나, 데이터셋에 따라 오히려 하락하는 결과를 보여주는 것을 확인할 수 있다. 따라서 순차적인 의사결정 과정과 중첩 이미지 증강 기법이 동시에 적용될 때 성능 향상의 폭이 크게 나타나 본 연구에서 제안하는 핵심 요소가 상호 보완적이며 효과적임을 보여준다.

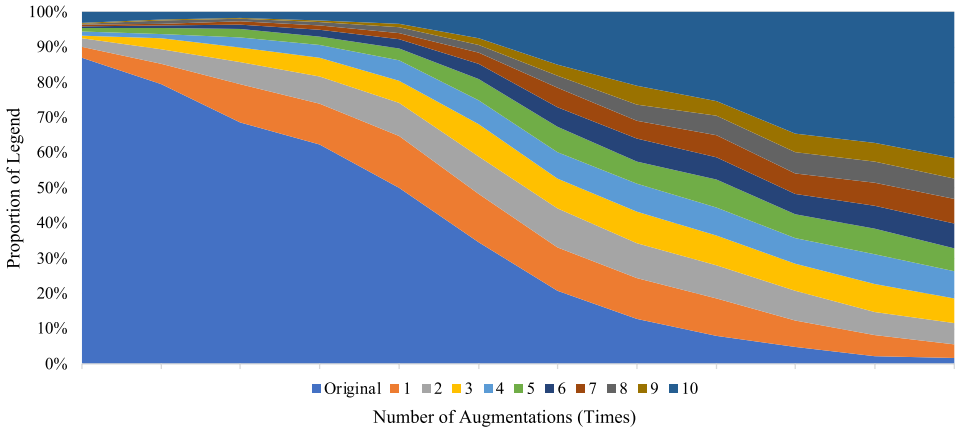
### 5.3 실험 결과 해석

제안 방법론은 순차적인 의사결정 구조와 중첩된 이미지 증강 기법 도입을 통하여 학습 정도에 따라 학습의 난이도를 조절하는 커리큘럼 러닝(curriculum learning) 방식으로 동작하도록 설계되었다. 따라서 학습 초기에는 원본 이미지 또는 1~2회 증강된 이미지, 즉 쉬운 이미지를 주로 학습하나 점차 모델이 고도화 됨에 따라 여러 번 중첩된 어려운 이미지도 분류할 수 있게 된다. 이는 실험 결과를 토대로 확인할 수 있으며 <Figure 11>에 도식화 하였다. <Figure 11>의 (a), (b), (c)는 각각 CIFAR-10, SVHN 그리고 WM-811K 데이터셋의 학습 과정을 2차원 영역형으로 표시한 차트이다. 각 차트의 가로축은 학습의 진행 정도를 뜻하며 좌측이 학습 초기, 우측이 학습 말기를 의미한다. 차트의 범례는 에피소드 마다 모델이 몇 번 증강된 이미지에서 분류에 실패하였는지를 뜻하고 이를 세로축에 비율로 표시하였다. 예를 들어 모델이 원본 이미지와 1회 증강된 이미지 분류에 성공하고, 2회 증강된 이미지 분류에 실패할 경우 <Figure 11>에서 회색 범례(2 times)에 속하게 된다. <Figure 11> (a)에서 학습 초기에는 약 85%의 에피소드에서 원본 이미지(original image)를 제대로 분류하지 못하였으나 학습 말기에는 거의 대부분의 에피소드에서 원본 이미지는 정답으로 분류하는 것을 확인할 수 있다. 그리고 학습이 진행됨에 따라 여러 번 중첩된 이미지의 분류 정답율이 높아지는 것을 확인 할 수 있으며 학습 종료 시점에는 절반이상의 에피소드에서 최대 10번 중첩된 이미지까지 모델이 경험하게 된다. 그리고 <Figure 11> (a)보다 <Figure 11> (b)와 (c)에서 10번 중첩된 이미지의 학습 비율이 빠르게 증가하는 것을 확인할 수 있는데 이러한 경향은 데이터셋의 분류 난이도에 따라 형태가 달라지는 것으

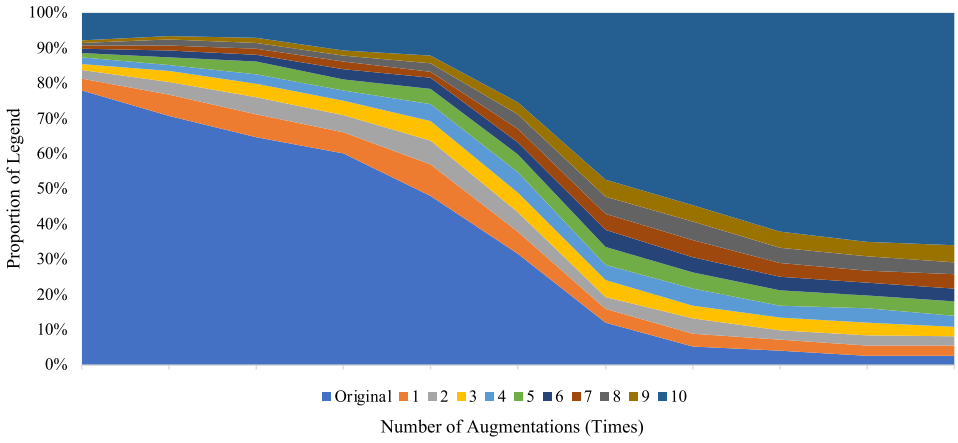
**Table 7.** Impact of Key Components of the Proposed Method for CIFAR-10, SVHN, and WM-811K Datasets

| Methods | Key Components | | | CIFAR-10 | | SVHN | | WM-811K | |
|---|---|---|---|---|---|---|---|---|---|
| | Sequential Structure | Augmentation | Overlapped Augmentation | Accuracy | F1-score | Accuracy | F1-score | Accuracy | F1-score |
| Deep Reinforcement Learning | X | X | X | 0.7367 (0.0066) | 0.7361 (0.0063) | 0.9060 (0.0050) | 0.8963 (0.0058) | 0.9244 (0.0051) | 0.8892 (0.0055) |
| | O | X | X | 0.7969 (0.0043) | 0.7965 (0.0043) | 0.9081 (0.0038) | 0.8992 (0.0039) | 0.9177 (0.0032) | 0.8805 (0.0073) |
| | O | O | X | 0.8356 (0.0051) | 0.8352 (0.0049) | 0.9175 (0.0037) | 0.9098 (0.0040) | 0.9132 (0.0052) | 0.8727 (0.0131) |
| | O | O | O | **0.8806 (0.0024)** | **0.8798 (0.0026)** | **0.9608 (0.0015)** | **0.9575 (0.0018)** | **0.9463 (0.0028)** | **0.9255 (0.0066)** |

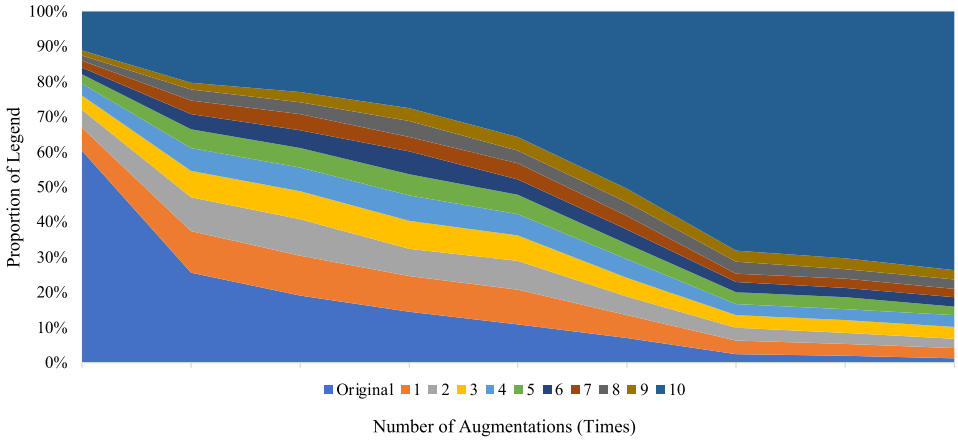
로 해석할 수 있다. 결과적으로 학습이 진행되면서 모델이 접하는 증강된 이미지의 비율 변화를 시각화 하여, 제안하는 심층 강화학습 방법론과 환경이 잘 적용되는지 확인 가능하다. 뿐만 아니라 학습이 진행되는 동안 모델의 학습 정도를 모니터링하며 이미지 증강의 세기와 종류를 도메인 및 데이터셋에 최적화하여 조절할 수 있다는 장점이 있다.




(a) CIFAR-10




(b) SVHN




(c) WM-811K

**Figure 11.** Diagram of the Learning Progress for (a) CIFAR-10, (b) SVHN, and (c) WM-811K. The horizontal axis indicates the degree of learning progress. The vertical axis represents the ratio within the legend. The legend Indicates the extent of overlap observed within a given episode before classification failed.

## 6. 결론

본 논문에서 심층 강화학습을 이미지 분류 문제에 적용하는 방법론을 제안하였다. 제안 방법론은 기존의 전통적인 학습 방법과 관련 연구(Lin *et al.*, 2020) 보다 높은 성능을 달성하였고 특히 2가지의 핵심 요소를 제안하였다. 첫 번째로 순차적인 의사결정 구조를 설계하였고 이는 모델이 학습을 진행함에 따라 자동적으로 이미지 증강의 기회를 조절하는 것과 같은 효과가 있다. 따라서 이미지 증강의 기회가 최대가 되도록 하여 이미지에서 핵심적인 특징을 추출하는데 도움이 되었다. 두번째로 중첩된 이미지 증강 기법을 도입하였다. 이미지 증강 기법을 여러 번 중첩하여 적용하면 데이터 소실로 인하여 분류가 어려운 이미지가 생성되게 되고 모델이 기존보다 어려운 이미지를 접할 수 있게 된다. 따라서 주어진 데이터셋 보다 다양한 분포의 이미지를 학습할 수 있게 되고 이는 과적합을 방지하고 강건한 모델을 생성할 수 있게 된다. 결과적으로 심층 강화학습 방법론을 사용할 경우 이미지 분류 문제에서 성능 향상이 가능함을 보였다. 뿐만 아니라 이미지 분류 벤치마크로 주로 활용되는 CIFAR-10과 SVHN 외에 실제 산업 현장에서 수집된 WM-811K 데이터셋에서도 성능 향상을 보여 이 분야의 가능성을 확인하였다. 하지만 근본적으로 모델과 상호작용 할 수 있는 환경이 없으므로 주어진 데이터셋 이상의 새로운 정보는 얻기 힘들다는 한계가 존재한다. 따라서 모델과 사용자와의 상호 작용 또는 생성형 모델을 활용하여 상태 간의 전이를 새롭게 제시할 수 있다면 더욱 효과적인 학습이 진행될 수 있을 것이다. 이처럼 이미지 분류 문제를 위하여 강화학습의 필수 요소를 정의하고 구조를 설계하는 것은 매우 자유로우며 다양한 연구가 이루어질 수 있다. 본 연구의 후속 연구로 학습 진행 과정을 모니터링하여 이미지 증강의 종류와 세기를 동적으로 변경하는 연구를 수행하고자 한다. 이를 통해 무작위로 증강 기법을 선택하는 것보다 효과적인 학습을 진행하고, 이미지 증강 기법 별 분류 정답률의 변화를 파악할 수 있을 것으로 기대된다. 마지막으로 증강 기법 별 효과 측정을 통해 이미지 분류 성능 향상을 달성하기 위한 데이터셋 별 최적 이미지 증강 기법 탐색도 가능하리라 예상된다.

## 참고문헌

AlMahamid, F. and Grolinger, K. (2021), Reinforcement Learning Algorithms: An Overview and Classification, In *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, IEEE, 1-7.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. (2019), Autoaugment: Learning Augmentation Strategies from Data, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 113-123.

Jang, S. Y., Yoon, H. J., Park, N. S., Yun, J. K., and Son, Y. S. (2019), Research Trends on Deep Reinforcement Learning, *Electronics and Telecommunications Research Institute*, **34**(4), 1-14.

Lin, E., Chen, Q., and Qi, X. (2020), Deep reinforcement learning for imbalanced classification, *Applied Intelligence*, **50**, 2488-2502.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013), Playing Atari with Deep Reinforcement Learning, arXiv preprint arXiv:1312.5602.

Qiao, J., Wang, G., Li, W., and Chen, M. (2018), An adaptive deep Q-learning strategy for handwritten digit recognition, *Neural Networks*, **107**, 61-71.

Shin, S. J., Cho, C. L., Jeon, H. S., Yoon, S. H., and Kim, T. Y. (2019), A Survey on Deep Reinforcement Learning Libraries, *Electronics and Telecommunications Research Institute*, **34**(6), 87-99.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... and Hassabis, D. (2017), Mastering the game of go without human knowledge, *Nature*, **550**(7676), 354-359.

Stember, J. and Shalu, H. (2022), Deep Reinforcement Learning Classification of Brain Tumors on MRI. In *Innovation in Medicine and Healthcare: Proceedings of 10th KES-InMed 2022*, Singapore: Springer Nature Singapore, 119-128.

Wu, M. J., Jang, J. S. R., and Chen, J. L. (2014), Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-scale Data Sets, *IEEE Transactions on Semiconductor Manufacturing*, **28**(1), 1-12.

Xu, M., Yoon, S., Fuentes, A., and Park, D. S. (2023), A Comprehensive Survey of Image Augmentation Techniques for Deep Learning, *Pattern Recognition*, 109347.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... and Wen, J. R. (2023), A Survey of Large Language Models, arXiv preprint arXiv:2303.18223.

## 저자소개

**고병은 :** 연세대학교 수학과에서 2015년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구 분야는 인공지능, 머신러닝을 제조 현장에 응용하는 연구를 수행하고 있다.

**김성범 :** 고려대학교 산업경영공학부 교수로 2009년부터 재직하고 있으며, 인공지능공학연구소 소장, 기업산학연협력센터 센터장, 한국데이터마이닝학회 회장을 역임했다. 미국 University of Texas at Arlington 산업공학과에서 교수를 역임하였으며, 한양대학교 산업공학과에서 학사학위를 미국 Georgia Institute of Technology에서 산업시스템공학 석사 및 박사학위를 취득하였다. 인공지능, 머신러닝, 최적화 방법론을 개발하고 이를 다양한 공학, 자연과학, 사회과학 분야에 응용하는 연구를 수행하고 있다.