

## 균열 탐지의 의미론적 분할을 위한 Mean Teacher 학습 구조 최적화

심승보<sup>1\*</sup>

# Mean Teacher Learning Structure Optimization for Semantic Segmentation of Crack Detection

Seungbo Shim<sup>1\*</sup>

**Abstract:** Most infrastructure structures were completed during periods of economic growth. The number of infrastructure structures reaching their lifespan is increasing, and the proportion of old structures is gradually increasing. The functions and performance of these structures at the time of design may deteriorate and may even lead to safety accidents. To prevent this repercussion, accurate inspection and appropriate repair are requisite. To this end, demand is increasing for computer vision and deep learning technology to accurately detect even minute cracks. However, deep learning algorithms require a large number of training data. In particular, label images indicating the location of cracks in the image are required. To secure a large number of those label images, a lot of labor and time are consumed. To reduce these costs as well as increase detection accuracy, this study proposed a learning structure based on mean teacher method. This learning structure was trained on a dataset of 900 labeled image dataset and 3000 unlabeled image dataset. The crack detection network model was evaluated on over 300 labeled image dataset, and the detection accuracy recorded a mean intersection over union of 89.23% and an F1 score of 89.12%. Through this experiment, it was confirmed that detection performance was improved compared to supervised learning. It is expected that this proposed method will be used in the future to reduce the cost required to secure label images.

**Keywords:** Deep learning, Semantic segmentation, Crack detection, Semi-supervised learning, Learning structure optimization

## 1. 서 론

인프라 구조물은 삶의 질을 결정하는 중요한 인자로 볼 수 있다. 특히 도로, 교량, 터널 등과 같은 인프라는 물류의 흐름을 결정하는 중요 요인으로 작용한다(Jeong et al., 2018). 게다가 콘크리트 건축물은 사람의 주거 환경의 안정성을 결정한다. 이러한 인프라 구조물은 대다수 경제 성장 시기에 완공되었고 사회 발전을 견인하는 역할을 수행하였다. 하지만 최근 들어 공용연수가 30년 이상 된 인프라 구조물의 비중이 점진적으로 증가하고 있어 노후 인프라의 사회적인 안전성에 대한 관심도 증가하고 있다(Kim and Kim, 2020).

이러한 실정 속에서 인프라 구조물을 안전하게 유지관리하기 위해서는 주기적인 점검과 적절한 보수가 절실하다. 이 중에서도 객관적인 점검은 현재 구조물의 상태를 정확히 진단하기 위해서 반드시 필요한 행위이다. 구조물 점검 시 작업자가 직접 현장에 방문하여 육안으로 수행하게 될 경우 점검 결과

는 작업자의 상태와 현장의 상황에 따라 상이하게 나타날 수 있다(Adhikari et al., 2014). 이는 부정확한 진단 결과를 야기하고 인프라 구조물의 전반적인 안전성을 저하시키는 요인으로 작용한다. 따라서 이러한 단점을 극복하고 주기적 점검의 정확성과 객관성을 향상시키기 위한 연구가 필요하다.

이를 위해서 최근 들어 고해상도의 카메라와 딥러닝 기술을 융합한 점검 방법들이 활발하게 연구되고 있다(Ye et al., 2019). 분명한 점은 딥러닝 알고리즘의 등장으로 영상 처리의 성능이 현격히 향상되었다. 이는 노후 인프라 점검 분야에서도 마찬가지이다. 그럼에도 불구하고 여전히 극복해야 할 문제가 있다. 본 연구에서는 여러 문제들 가운데, 라벨 데이터를 확보하는 데 필요한 비용과 시간에 대한 문제를 다루고자 한다. 딥러닝은 데이터 학습 기반의 알고리즘으로 반드시 학습 데이터가 필요하다. 마찬가지로 본 연구에서 다루는 균열 탐지에도 학습에 필요한 데이터가 필수적이다. 이들은 균열을 촬영한 영상과 영상 내에서 균열 영역을 표시한 라벨 영상으로 구성된다. 이러한 라벨 영상은 현재 토목 분야 내 관련 전문적 지식을 가진 전문가들에 의해 만들어져야 한다. 그러므로 라벨 데이터를 확보하는 데 많은 고급 노동력이 필요할 수밖에 없다. 본 논문에서는 이 비용을 절감하기 위해 소수의 라벨 영상으로도 균열 탐지 신경망 모델을 학습할 수 있는 방법을 연구하였다.

<sup>1</sup>정회원, 한국건설기술연구원 지반연구본부 수석연구원, 교신저자

\*Corresponding author: seungboshim@kict.re.kr

Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang-Si, Gyeonggi-Do, 10223, Korea  
•본 논문에 대한 토의를 2023년 11월 30일까지 학회로 보내주시면 2023년 12월호에 토론결과를 게재하겠습니다.

## 2. 문헌조사

고해상도 영상을 이용하는 것의 가장 큰 장점은 미세한 균열까지 탐지가 가능하다는 점에 있다. 이 같은 장점을 바탕으로 영상 처리 알고리즘에 대한 연구가 활발히 진행되어 왔다. Abdel-Qader et al.(2003)은 영상으로 촬영한 균열을 탐지하기 위하여 edge 기반 알고리즘을 제안했다. 이 알고리즘은 균열이 콘크리트 면에서 불연속 경계라는 점을 전제로 한다. 그리고 이러한 경계는 edge로써 탐지될 수 있다고 보고 알고리즘을 개발하였다. 다음으로 딥러닝 기반의 신경망 알고리즘은 의미론적 분할(semantic segmentation) 기법을 주로 사용한다. 대표적인 신경망 모델은 fully connected neural network (FCN)가 있다(Long et al., 2015). 이는 균열 영역을 화소 단위로 정확히 추출할 수 있는 방식이다. FCN는 주로 합성곱 연산으로 구성되어 feature map를 생성하고 이후에 up-sampling 과정을 통해 균열 영역을 결정하게 된다. 이와 유사하게 UNet 구조를 가진 신경망 모델을 통해서 균열을 탐지하는 연구 결과들도 다수 소개되었다(Ali et al., 2022). 이 구조는 feature map을 생성하는 암호화 단계와 균열 영역을 결정하는 복호화 단계로 구성되어 있다. 아울러 UNet은 암호화 단계와 복호화 단계를 연결하여 균열 탐지 성능을 향상시키는 점이 특징이다.

균열 탐지 알고리즘을 개발하는 데 있어 신경망 모델의 구조를 개선하는 것뿐만 아니라 학습 구조를 개발하는 방법도 있다. 특히 라벨 데이터가 부족할 때 적용하는 방법으로 준지도 학습이 있다. Li et al.(2020)은 도로 포장 균열을 탐지하기 위해서 적대적 학습을 사용하였다. 이는 라벨 영상의 수가 입력 영상의 수보다 작더라도 탐지 정확도를 높일 수 있도록 판

별자 신경망을 추가한 학습 구조다. 또한, Shim et al.(2022)은 도로 노면 균열을 탐지하기 위해 초해상화 알고리즘과 적대적 학습을 융합하는 학습 구조를 제시하였다. 또한, Shim et al.(2023)은 적대적 학습에 ensemble 기법을 접목시켜 판별자 신경망의 성능을 향상시켰고 결과적으로 균열 탐지 정확도를 높이는 방법을 제안하였다. 그리고 Maeda et al.(2021)은 생성적 적대 신경망을 이용하여 포트홀 영상을 생성하는 알고리즘을 제안하였다. 그들은 이 연구를 통해서 생성된 영상이 탐지 성능에 미치는 영향에 대해 분석하였다. Zhang et al.(2020)은 생성적 적대 신경망 기반의 영상 변환 기법을 통해서 균열을 탐지할 수 있는 방법을 제안하였다.

종합하여 살펴보면, 균열 탐지 정확도의 향상을 목적으로 한 다수의 연구들이 소개되었음을 알 수 있다. 이 중에서도 부족한 학습 데이터의 제약 조건을 극복하기 위해 공통적으로 학습 구조를 발전시키려는 연구들이 진행되고 있다. 특히 균열 영상은 일반적으로 촬영되는 정상 장면이 아니라 비교적 촬영이 어려운 비정상 장면이다. 설상가상으로 이를 학습에 활용하기 위한 라벨 영상도 많은 비용을 통해서만 확보할 수밖에 없는 실정이다. 따라서 높은 정확도를 가진 균열 탐지 알고리즘을 개발하기 위해서는 이 같은 한계를 극복하기 위한 연구가 반드시 필요하겠다.

## 3. Mean teacher 기반 균열 탐지

### 3.1 영상 데이터의 구성

콘크리트 균열 탐지를 위해서 본 논문에서는 라벨 영상 데이터 세트와 비라벨 영상 데이터 세트를 사용하였다(Shim et

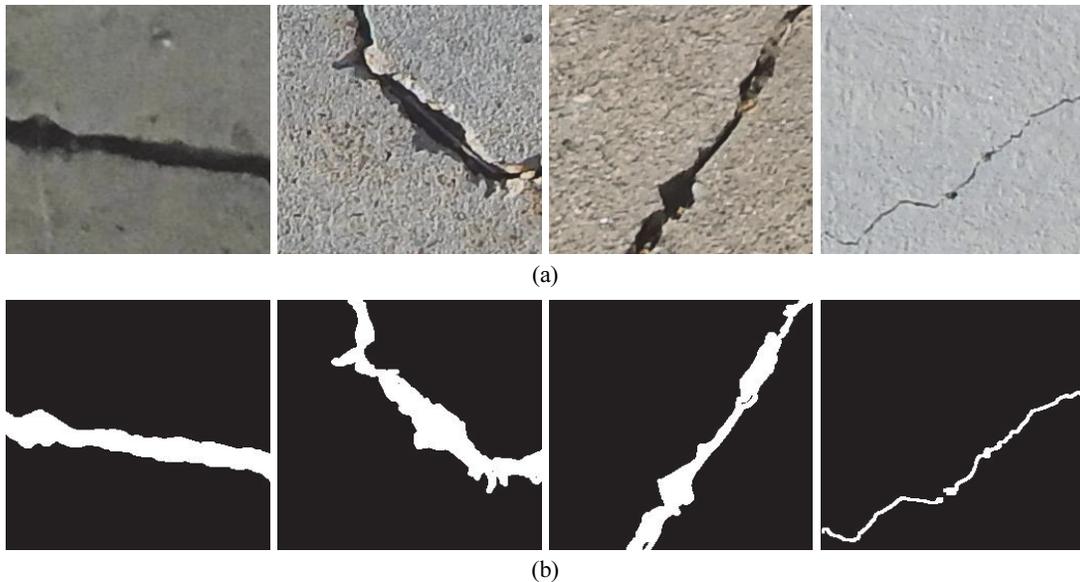


Fig. 1 Dataset: (a) crack image and (b) label image

al., 2020). 라벨 영상 데이터 세트는 Fig. 1과 같이 균열 영상과 라벨 영상이 짝을 이루고 있다. Fig.1(a)은 RGB 색상을 가진 영상이고 Fig.1(b)은 영상 내에서 균열의 영역을 표시한 라벨 영상이다. 비라벨 영상 데이터는 균열 영상만으로 이루어진 데이터다. Fig. 1(a)과 유사하게 다양한 콘크리트 재질에서 발생한 균열을 포함하고 있다.

학습을 위한 구체적인 수량은 라벨 영상 데이터 세트와 비라벨 영상 데이터 세트가 각각 1,196장과 3,000장이다. 모든 영상은 컬러이며 그 크기는  $256 \times 256$  화소이다. 라벨 영상 데이터 세트 중 900장은 훈련용으로 사용하며 나머지는 실험용으로 사용하였다. 비라벨 영상 데이터 세트는 준지도 학습에서 활용될 수 있도록 하였다.

### 3.2 Mean teacher 학습 구조

신경망 모델을 훈련하는 방법은 3가지로 구성된다. 이들은 훈련에 사용되는 라벨의 비율 따라 지도학습, 준지도 학습, 비지도 학습으로 구분된다. 지도학습은 입력 영상과 라벨 영상의 비율이 1대 1인 조건에서 신경망을 훈련하는 학습이다. 비지도학습은 라벨 영상 없이 입력 영상만으로 훈련하는 것이다. 끝으로 라벨 영상의 수가 입력 영상의 수보다 작을 때 준지도 학습을 할 수 있다. 본 논문에서 사용하는 학습 구조는 준지도 학습이다. 이는 균열 영상의 수가 라벨 영상의 수보다 많은 상황에서 적용하는 방법이다. 이 방법의 가장 큰 장점은 라벨 수가 적더라도 탐지 정확도를 향상시킬 수 있다는 점이다. 이를 위해 본 논문에서는 Fig. 2와 같이 mean teacher의 학습 구

조를 사용하였다(Tarvainen and Valpola, 2017). 이 학습 구조는 균열 탐지를 위한 분할 알고리즘 신경망을 사용한다. 이 신경망 모델은 역할에 따라 스승( $Net_T$ ) 신경망과 학생( $Net_S$ ) 신경망으로 구분하여 동일한 모델을 2회 사용한다. 우선 학생 신경망은 균열 탐지를 위한 라벨 영상 데이터 세트에 의해 훈련된다. 균열 영상이 학생 신경망을 통과하여 얻게 되는 예측 결과가 라벨 영상과 유사해지도록 가중치가 갱신된다. 이때 라벨 영상과 예측 영상이 차이가 최소화 되도록 역전파의 방법으로 학생 신경망을 훈련한다. 다음으로는 비라벨 영상 데이터 세트를 이용한 학습이 있다. 이 데이터 세트에서는 균열 영상의 영역을 나타내는 가짜 라벨(pseudo label) 영상이 필요하다. 이를 위해서 스승 신경망의 예측 영상을 활용한다. 결과적으로 스승 신경망의 예측 영상과 유사한 예측 영상을 생산할 수 있도록 학생 신경망은 훈련된다. 이때 역전파의 경로는 학생 신경망으로 향한다. 이처럼 라벨과 비라벨 영상 데이터 세트로 훈련된 학생 신경망의 가중치는 지수이동평균(EMA, Exponential moving average)을 통해 스승 신경망에게 전파된다. 이러한 과정을 반복적으로 거치면서 스승 신경망은 보다 정확하고 신뢰할 수 있는 가짜 라벨 영상을 생산하게 되고 이는 다시 학생 신경망의 성능을 향상시킨다. 이 성능 향상 효과는 또다시 스승 신경망에 전달되어 균열 탐지 정확도가 향상된 신경망 모델로서 완성이 된다.

본 연구에서 사용된 지수이동평균은 Eq.(1)과 같이 정의된다.  $\theta'$  와  $\theta$ 는 각각 스승 신경망과 학생 신경망의 가중치를 가

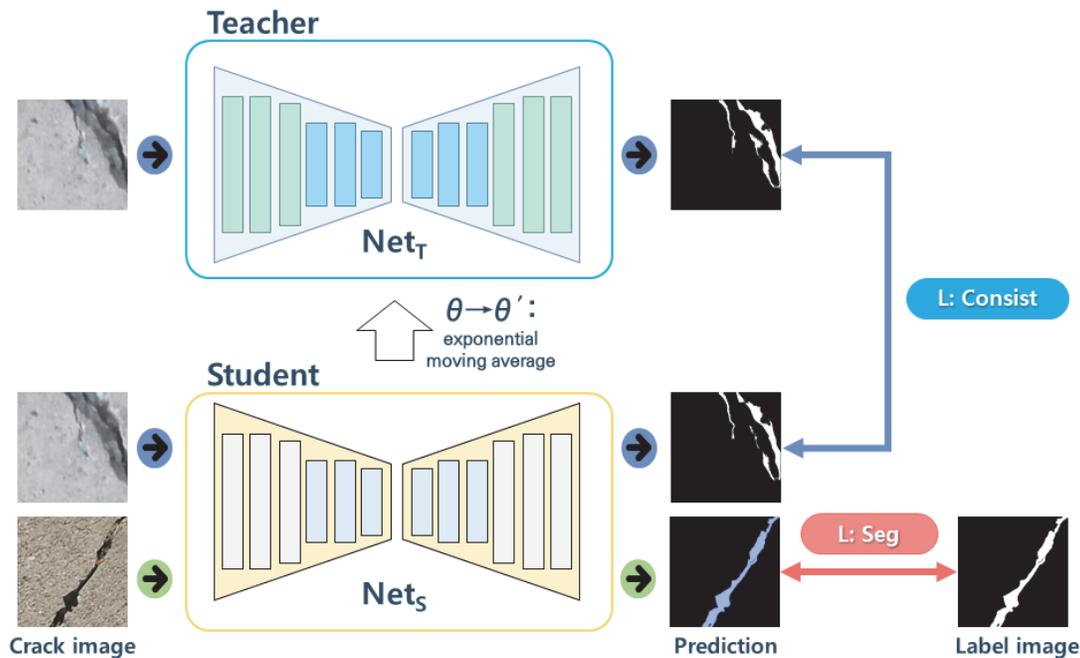


Fig. 2 Learning structure of mean teacher

리킨다. 지수이동평균은 이전 스승 신경망의 가중치를 현재 학생 신경망의 가중치로 갱신하는 과정인데, 갱신하는 속도는  $\alpha$ 에 의해 결정된다.  $\alpha$ 값이 작으면 갱신되는 속도는 빠르고 크면 그만큼 느려진다. 본 연구에서는 이를 0.95로 설정하여 스승 신경망을 훈련하였다. 원래의 mean teacher에서 사용한  $\alpha$ 는 0.99를 사용하였다. 이에 반해 Zheng et al.(2022)에서는 0.90을  $\alpha$ 로 사용하였다. 이러한 점을 고려하여 본 연구에서는 이들의 평균값을 사용하였다.

$$\theta'_t = \alpha\theta'_{t-1} + (1 - \alpha)\theta_t \quad \text{Eq.(1)}$$

Mean teacher 학습 구조에서는 두 가지 손실 함수의 합 ( $L_{Total}$ )이 적용된다. 이는 Eq.(2)과 같이 정의된다. 하나는 라벨 영상 데이터 세트에 사용되는 cross-entropy 손실 함수 ( $L_{Seg}$ )이고 다른 하나는 비라벨 영상 데이터 세트에 사용되는 mean-square-error 손실함수( $L_{Consist}$ )다. 이때  $L_{Consist}$ 의 값이 전체적인 학습에 미치는 영향을 조정하기 balance weight( $\lambda$ )를 사용하였다. 이는 스승 신경망의 예측 영상이 가짜 라벨의 역할을 수행할 때 학생 신경망의 훈련에 미치는 영향 정도를 결정하는 변수다. 이 변수에 따라 균열 탐지의 정확도는 영향을 받는다. 이러한 점을 감안하여 정확한 균열 탐지를 위한 최적의 값을 찾는 연구를 수행하였다.

$$L_{Total} = L_{Seg} + \lambda L_{Consist} \quad \text{Eq.(2)}$$

## 4. 실험 결과 및 논의

### 4.1 평가 지표

균열 탐지의 정확도를 평가하기 위하여 두 가지의 평가지표를 사용하였다. 첫 번째는 평균 중첩 정확도(m-IoU, mean-Intersection over Union)이며, 두 번째는 F1-score (F1)이다. m-IoU는 탐지 영역에 대한 비교로 Eq.(3)과 같이 정의된다. Eq.(3)에 있는  $n_{ij}$ 는 class  $j$ 에 속해 있을 것으로 예측된 class  $i$ 의 모든 화소 수를 나타낸다. 아울러  $n_{cl}$ 는 class의 개수를 가리키고,  $t_i$ 는 class  $i$ 에 속해 있는 모든 화소 수를 지칭한다. F1은 화소 단위의 정확도 비교로 Eq.(4)와 같이 정의된다. Eq.(4)에서 Pr과 Re는 각각 정밀도와 재현율을 가리키고 이들의 조화평균을 평가 지표로 삼았다.

$$\text{m-IoU: } (1/n_{cl}) \sum_i n_{ii} / (t_i + \sum_j n_{ji} - n_{ii}) \quad \text{Eq.(3)}$$

$$\text{F1: } \frac{2 \times \text{Pr} \times \text{Re}}{\text{Pr} + \text{Re}} \quad \text{Eq.(4)}$$

### 4.2 학습 환경

균열 탐지 신경망 모델을 훈련하기 위하여 사용한 하드웨어 사양은 Intel Xeon 6226R 2.9 GHz, 320GB 메모리, NVIDIA Quardro 8000이다. 소프트웨어는 Ubuntu 18.04의 운영 체제 기반의 PyTorch 딥러닝 프레임 워크를 사용하였다. 학습을 위해 사용한 파라미터는 다음과 같다. 전체 에포크의 수는 5,000번이고 배치의 수는 64로 하였다. 최적화 함수는 adam을 사용하였다. 아울러 adam의 파라미터인 learning rate, beta-1, beta-2는 각각 0.001, 0.9, 0.999로 초깃값을 사용하였다. 훈련이 진행되는 가운데 매 10회의 에포크마다 신경망 모델의 가중치를 저장하도록 설정하였다.

### 4.3 성능 평가

균열 탐지를 위한 신경망 모델로 본 논문에서는 ERFNet을 사용하였다(Romera et al., 2017). 이 신경망 모델은 residual module을 활용하여 신경망 구조를 설계한 점이 특징이다. 이 모델을 mean teacher의 학습 구조에 적용하여 균열 탐지의 성능을 관찰하였다. 특히 학습 구조의 손실함수에서 사용되는 balance weight의 변화에 따른 균열 탐지 정확도의 변화를 평가하였다. balance weight의 값은 1.00에서부터 0.01까지 변화하였고 그에 따른 탐지 정확도는 Table 1과 같다. 탐지 정확도는 훈련된 신경망 모델을 실험용 300여 장의 라벨 영상 데이터 세트에 적용하여 평가된 결과다. Balance weight가 값이 1.00에서 점차 작아지면 탐지 정확도는 향상되는 것을 확인할 수 있다. 또한, balance weight가 0.01에 가까워지면서 정확도는 감소되는 것으로 드러났다. 결과적으로 balance weight는 0.05일 때 89.23% m-IoU와 89.12% F1을 기록하며 가장 높은 정확도를 보였다. 이는 비라벨 영상 데이터 세트를 이용한 학습이 미치는 영향에서 기인한 것이다. 신경망 모델에 의해서 생산된 가짜 라벨 영상은 사람에 의해서 만들어진 진짜 라벨 영상보다 신뢰성이 떨어진다고 가정하였다. 그러므로 가짜 라벨 영상을 어느 정도로 신뢰할지를 선택하는 것이 중요하다. 예를 들어, balance weight를 1.00으로 설정하였다면 이는 진짜 라벨 영상과 가짜 라벨 영상을 동일한 수준으로 신뢰하는 것을 의미한다. 이에 반해 balance weight를 0.01을 설정한다면 이는 가짜 라벨 영상을 그만큼 신뢰하지 않는 것으로 볼 수 있다. 가짜 라벨 영상에 대한 신뢰의 정도는 훈련에 사용되

**Table 1** Performance evaluation results [%]

Metrics	Balance weight ( $\lambda$ )					Supervised
	0.01	0.05	0.10	0.50	1.00	
m-IoU	88.94	89.23	89.05	88.53	87.61	88.22
F1	88.79	89.12	88.91	88.29	87.22	87.96

는 데이터와 신경망 모델의 특성에 따라 달라진다. 이러한 점을 고려하여 본 연구에서는 다수의 실험을 통해 가장 높은 탐지 정확도를 낼 수 있는 최적의 balance weight를 선정하였다.

본 연구에서 사용한 mean teacher의 학습 구조는 준지도 학습의 한 종류다. Mean teacher 방식은 하나의 신경망을 복수 사용하는 점이 특징이다. 두 개 이상의 신경망 모델을 사용하는 준지도 학습에서는 서로 다른 신경망 모델을 선정하는 데 많은 경우의 수를 고려해야 한다. 이에 반해 mean teacher 방식은 이러한 경우의 수가 적기 때문에 개발 과정에서 고려해야 할 변수를 줄일 수 있는 장점이 있다. 또한 지도학습과는 달리 균열 영상만을 추가로 사용해도 탐지 정확도를 향상시킬 수 있다. 다시 말해 라벨 영상을 만드는 데 필요한 비용과 시간을 절감하는 효과가 발생한다. 본 논문에서도 이러한 효과를 확인할 수 있다. 지도학습

만으로 신경망 모델을 훈련했을 경우 그 정확도는 88.22% m-IoU와 87.96% F1이다. 이에 반해 mean teacher 학습 구조를 적용했을 때는 지도학습보다 1.01% m-IoU와 1.16%의 F1만큼 탐지 정확도가 향상되었다. 균열 탐지 정확도는 균열 폭의 측정 정확도에 영향을 미친다. 아울러 이는 구조물 상태 진단의 객관성과 신뢰성과도 밀접하게 연관된다. 그러므로 본 논문에서 제안한 소량의 라벨 영상으로 균열 탐지 정확도를 향상시킨 방법은 노후 인프라 구조물의 안전성 향상에도 기여한다.

#### 4.4 탐지 결과 비교 · 분석

학습 구조에 대한 연구의 최종 목적은 제안한 방법을 이용하여 균열 탐지의 정확도를 향상시키는 것에 있다. 이러한 향상은 Fig. 3과 같이 탐지 결과에서 나타난다. Fig. 3의 (a)와 (b)

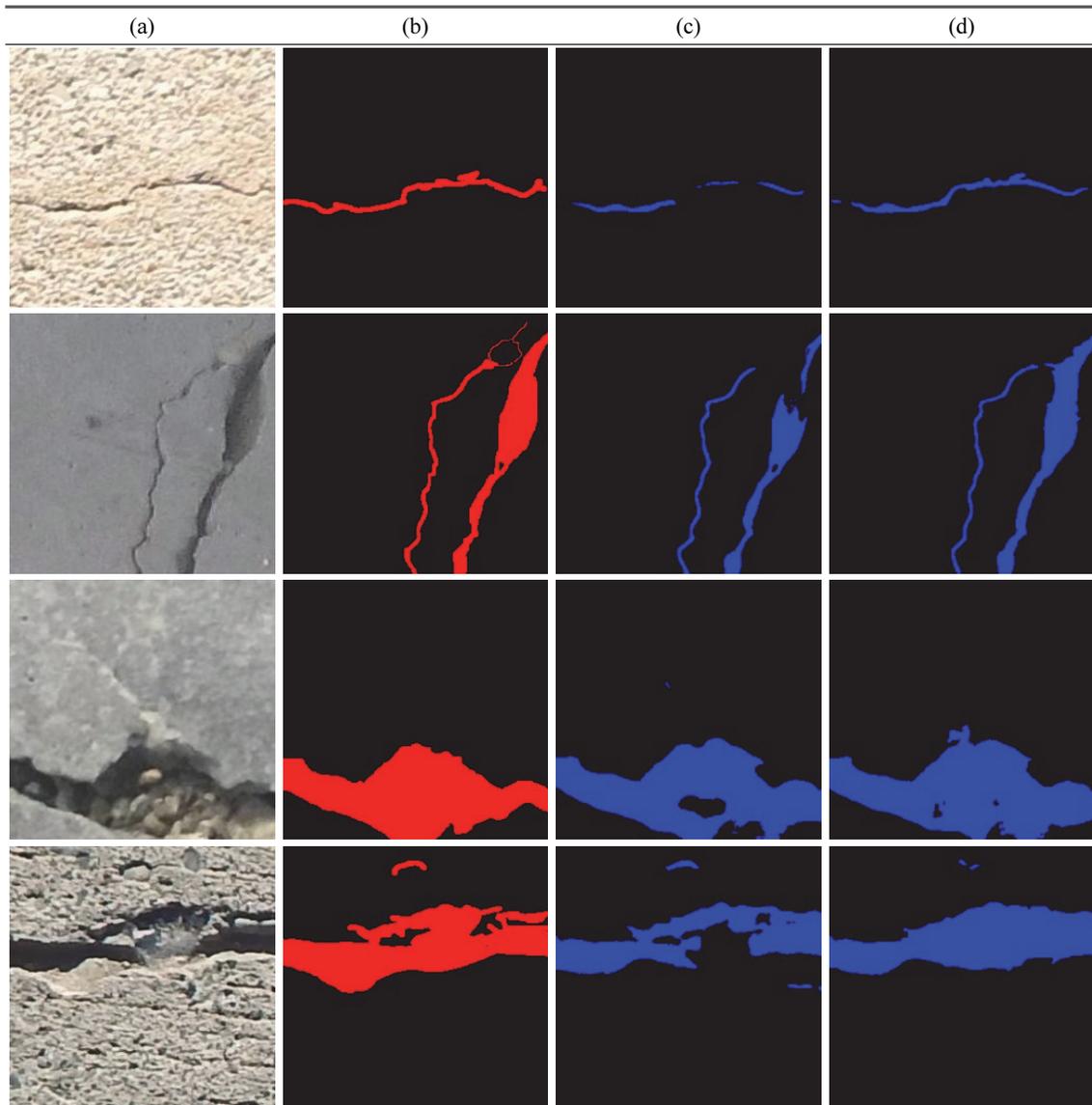


Fig. 3 Result Images: (a) Input images, (b) Label images, (c) Supervised, (d) Mean teacher

는 각각 균열 영상과 라벨 영상이다. 또한, (c)와 (d)는 각각 지도 학습 방법과 mean teacher 방법에 의한 균열 탐지 결과다. 이 때 balance weight는 0.05다. 첫 번째 행의 결과 영상은 미세 균열의 탐지 성능을 보여 준다. Mean teacher 학습 방법을 적용할 때 라벨 영상과 더욱 유사한 예측 결과를 얻을 수 있다. 두 번째부터 네 번째 행의 영상은 넓은 면적의 손상에 대한 탐지 결과다. 지도 학습만을 사용할 경우 손상 면적의 전체를 탐지하지 못하였다. 이에 반해 mean teacher 방식을 적용하면 라벨 영상과 유사한 영역을 예측할 수 있다.

균열 탐지의 신경망 모델을 학습하는 데 영향을 미치는 변수는 다양하게 존재한다. 하지만 mean teacher의 학습 구조에서는 지수이동평균의  $\alpha$ 와 손실 함수의  $\lambda$ 가 주요 변수의 역할을 한다.  $\alpha$ 는 스승과 학생 신경망 간의 관계에서 정의되는 변수이고,  $\lambda$ 는 라벨과 비라벨 데이터 간의 관계에서 활용되는 변수다. 본 연구의 목적은 소량의 라벨 영상으로 탐지 정확도를 향상시키는 방법을 찾는 것에 있다. 이러한 맥락에서  $\lambda$ 에 따른 정확도 변화를 관찰하였다.

Mean teacher 학습을 적용했을 때 정확한 탐지할 가능한 이유는 보다 많은 균열 영상이 학습에 참여했기 때문이다. 비록 라벨이 없는 영상 데이터라 할지라도 다양한 균열의 형상 정보를 가지고 있는 것은 사실이다. 이러한 점을 학습에 이용하여 균열 탐지 성능을 향상시켰다. 물론 훈련 과정에서 균열 정보를 적절히 활용하기 위해서는 비라벨 영상 데이터의 학습 참여율을 결정해야 한다. 본 연구에서는 이를 balance weight라 간주하였고 이 참여율에 따라 탐지 성능의 변화를 관찰하였다. 여러 번의 실험을 통해 최종적으로 탐지 성능을 가장 높일 수 있는 값을 결정하였다. 이러한 연구를 통해 소수의 라벨 영상으로도 균열 탐지의 정확도를 향상시키기 위한 학습 구조를 개발하였다.

준지도 학습의 한 종류인 mean teacher 학습 구조는 라벨과 비라벨 영상 데이터가 활용된다. 비라벨 영상 데이터는 일반적으로 낮은 신뢰성을 갖고 있어 라벨 영상 데이터와 달리 부분적인 훈련 참여가 필요하다. 아울러 이와 같은 참여율은 데이터의 특성에 따라 달라진다. 따라서 본 연구에서는  $\lambda$ 를 사용하여 실험을 수행하였고 균열의 특성에 적합한 비라벨 데이터의 참여율을 제시했다. 이는 균열 탐지 분야에서 mean teacher 학습의 학술적 의미를 갖는다.

본 연구에서는 수천 장 규모의 학습 데이터를 사용하였다. 학습 데이터의 수량은 데이터의 복잡성에 따라 결정된다. 영상이 복잡한 배경과 내용물을 포함하고 있다면 전경을 탐지하기 위해서 다수의 데이터가 필요하다. 본 연구에서 사용한 학습 데이터는 Fig. 1에서 나타난 것과 같이 단일 재질에 불연속 경계선을 포함한 영상들이다. 일반적으로 관찰할 수 있는 콘크리트 표면 누수와 얼룩 그리고 시설물과 지장물 등이 균

열 영상 학습 데이터에는 포함되어 있지 않다. 이러한 점을 살펴보면, 학습 데이터의 복잡성은 상대적으로 낮다고 보는 것이 타당하다. 따라서 본 연구에서 사용한 데이터 세트의 규모로 얻는 연구 결과는 의미를 가질 수 있다.

균열 탐지 알고리즘의 최종적인 목적은 실제 현장에 적용되어 노후 인프라의 객관적인 상태를 평가하는 것이다. 이를 위하여 여러 가지 실용적인 방법이 필요하다. 특히 새로운 환경 조건을 가진 현장에서 신경망 모델이 우수하게 균열을 탐지하기 위해서는 다양한 학습 데이터가 필요하다. 다양한 균열의 폭과 면적을 포함한 영상뿐만 아니라 여러 종류의 콘크리트 재질을 가진 학습 영상이 있어야 한다. 다시 말해 학습 데이터의 다양성을 높이는 방법이 필요하다. 이에 반해 본 연구에서는 이러한 점보다 라벨 영상을 확보하기 위해 필요한 비용을 절감하는 방법에 집중하였다. 다양한 균열이 있더라도 라벨 영상은 반드시 필요하다. 이는 필연적으로 많은 노동력과 시간을 소모한다. 본 연구에서는 이러한 점을 극복하는 데 필요한 방법론을 제안하였다.

## 5. 결 론

콘크리트의 균열을 딥러닝으로 탐지하기 위해서는 학습 데이터가 있어야 한다. 특히 균열 영상에서 균열의 위치를 표시한 라벨 데이터는 필수적이다. 하지만 이러한 라벨 영상을 확보하기 위해서는 많은 비용을 소요될 수밖에 없다. 그러므로 본 연구에서는 비용을 절감하고 동시에 균열 탐지 정확도를 향상시킬 수 있는 mean teacher 학습 구조를 사용하였다. 이 학습 구조는 균열 영상과 동일한 수의 라벨 영상이 필요한 지도학습과 비교하여 성능 평가를 수행하였다. 이 실험을 통해 라벨 영상이 없더라도 균열 영상만으로 탐지 정확도가 향상될 수 있다는 점을 확인하였다. 게다가 비라벨 영상 데이터가 훈련에 미치는 영향을 balance weight를 통해 제어함으로써 정확도 향상을 최대화시킬 수 있는 방법도 제안하였다. 일반적으로 라벨 영상은 확보하는 데 필요한 시간이 영상의 크기에 따라 다르다. 그렇지만 Bang et al.(2019)의 연구에 따르면 아스팔트 도로 균열 영상의 라벨을 만들기 위해서 한 장당 12분의 시간이 소요된다. 또한, 딥러닝의 학습 결과는 라벨 영상의 신뢰도에 따라 크게 영향을 받는다. 그러므로 전문가에 의해서 라벨 영상이 만들어져야 한다. 하지만 전문가의 라벨링 작업은 현실적으로 이루어지기 쉽지 않기 때문에 높은 신뢰성을 가진 다수의 라벨 영상을 확보하는 것은 일정 부분 한계를 수반한다. 본 연구에서 제안하는 방법은 소수의 라벨 영상으로 균열 탐지 신경망을 학습하는 방법으로 부족한 라벨 영상에 대한 일련의 해결 방안이 될 것으로 기대하는 바이다.

## 감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다(No. 2022R1F1A1074663). 지원에 감사합니다.

## References

1. Jeong, Y., Kim, W., Lee, I., and Lee, J. (2018), Bridge inspection practices and bridge management programs in China, Japan, Korea, and US, *Journal of Structural Integrity and Maintenance*, 3(2), 126-135.
2. Kim, H., and Kim, C. (2020), Deep-learning-based classification of point clouds for bridge inspection, *Remote Sensing*, 12(22), 3757.
3. Adhikari, R. S., Moselhi, O., and Bagchi, A. (2014), Image-based retrieval of concrete crack properties for bridge inspection, *Automation in construction*, 39, 180-194.
4. Ye, X. W., Jin, T., Yun, C. B. (2019), A review on deep learning-based structural health monitoring of civil infrastructures, *Smart Structures and Systems*, 24(5), 567-585.
5. Abdel-Qader, I., Abudayyeh, O., and Kelly, M. E. (2003), Analysis of edge-detection techniques for crack identification in bridges, *Journal of Computing in Civil Engineering*, 17(4), 255-263.
6. Long, J., Shelhamer, E., and Darrell, T. (2015), Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, MA, USA, 3431-3440.
7. Ali, R., Chuah, J. H., Talip, M. S. A., Mokhtar, N., and Shoaib, M. A. (2022), Structural crack detection using deep convolutional neural networks, *Automation in Construction*, 133, 103989.
8. Li, G., Wan, J., He, S., Liu, Q., and Ma, B. (2020), Semi-supervised semantic segmentation using adversarial learning for pavement crack detection, *IEEE Access*, 8, 51446-51459.
9. Shim, S., Kim, J., Lee, S. W., and Cho, G. C. (2022), Road damage detection using super-resolution and semi-supervised learning with generative adversarial network, *Automation in Construction*, 135, 104139.
10. Shim, S., Kim, J., Cho, G. C., and Lee, S. W. (2023), Stereo-vision-based 3D concrete crack detection using adversarial learning with balanced ensemble discriminator networks, *Structural Health Monitoring*, 22(2), 1353-1375.
11. Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., and Omata, H. (2021), Generative adversarial network for road damage detection, *Computer-Aided Civil and Infrastructure Engineering*, 36(1), 47-60.
12. Zhang, K., Zhang, Y., and Cheng, H. D. (2020), Self-supervised structure learning for crack detection based on cycle-consistent generative adversarial networks, *Journal of Computing in Civil Engineering*, 34(3), 04020004.
13. Shim, S., Kim, J., Cho, G. C., and Lee, S. W. (2020), Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete structures, *IEEE Access*, 8, 170939-170950.
14. Tarvainen, A., and Valpola, H. (2017), Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.
15. Zheng, M., You, S., Huang, L., Wang, F., Qian, C., and Xu, C. (2022), Simmatch: Semi-supervised learning with similarity matching, *Proceedings of the IEEE conference on computer vision and pattern recognition*, New Orleans, LA, USA, 14471-14481.
16. Romera, E., Alvarez, J. M., Bergasa, L. M., and Arroyo, R. (2017), Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, *IEEE Transactions on Intelligent Transportation Systems*, 19(1), 263-272.
17. Bang, S., Park, S., Kim, H., and Kim, H. (2019), Encoder-decoder network for pixel-level road crack detection in black-box images. *Computer-Aided Civil and Infrastructure Engineering*, 34(8), 713-727.

Received : 09/13/2023

Revised : 10/06/2023

Accepted : 10/10/2023

**요 지** : 인프라 구조물은 대부분 경제 성장기에 완공되었다. 이러한 인프라 구조물은 최근 들어 공용연수가 점차 증가하고 있어 노후 구조물의 비중이 점차 증가하고 있다. 이러한 노후 구조물은 설계 당시의 기능과 성능이 저하될 수 있고 안전사고로까지 이어질 수 있다. 이를 예방하기 위해서는 정확한 점검과 적절한 보수가 필수적이다. 이를 위해서는 우선 미세한 균열까지 정확히 탐지할 수 있도록 컴퓨터 비전과 딥러닝 기술에 수요가 증가하고 있다. 하지만 딥러닝 알고리즘은 다수의 학습 데이터가 있어야 한다. 특히 영상 내 균열의 위치를 표시한 라벨 영상은 필수적이다. 이러한 라벨 영상을 다수 확보하기 위해서는 많은 노동력과 시간이 필요한 실정이다. 이러한 비용을 절감하고 탐지 정확도를 높이기 위해서 본 연구에서는 mean teacher 방식의 학습 구조를 제안하였다. 이 학습 구조는 900장의 라벨 영상 데이터 세트와 3000장의 비라벨 영상 데이터 세트로 훈련되었다. 학습된 균열 탐지 신경망 모델은 300여장의 실험용 데이터 세트를 통해 평가되었고 탐지 정확도는 89.23%의 mean intersection over union과 89.12%의 F1 score를 기록하였다. 이 실험을 통해 지도학습과 비교하여 탐지 성능이 향상된 것을 확인하였다. 향후에 이러한 방법은 라벨 영상을 확보하는데 필요한 비용을 절감하는데 활용될 것으로 기대한다.

**핵심용어** : 딥러닝, 의미론적 분할, 균열 탐지, 준지도 학습, 학습 구조 최적화