

논문 2023-18-35

멀티 모달 지도 대조 학습을 이용한 농작물 병해 진단 예측 방법 (Multimodal Supervised Contrastive Learning for Crop Disease Diagnosis)

이 현 석, 여 도 엽, 함 규 성, 오 강 한*
(Hyunseok Lee, Doyeob Yeo, Gyu-Sung Ham, Kanghan Oh)

Abstract : With the wide spread of smart farms and the advancements in IoT technology, it is easy to obtain additional data in addition to crop images. Consequently, deep learning-based crop disease diagnosis research utilizing multimodal data has become important. This study proposes a crop disease diagnosis method using multimodal supervised contrastive learning by expanding upon the multimodal self-supervised learning. RandAugment method was used to augment crop image and time series of environment data. These augmented data passed through encoder and projection head for each modality, yielding low-dimensional features. Subsequently, the proposed multimodal supervised contrastive loss helped features from the same class get closer while pushing apart those from different classes. Following this, the pretrained model was fine-tuned for crop disease diagnosis. The visualization of t-SNE result and comparative assessments of crop disease diagnosis performance substantiate that the proposed method has superior performance than multimodal self-supervised learning.

Keywords : Crop Disease Diagnosis, Multimodal Data, Supervised Contrastive Learning, Deep Learning

1. 연구배경

농작물 질병은 생산량 감소와 품질 저하를 초래하는 중요한 문제이다. 따라서 농작물 질병을 조기에 감지하여 치료 또는 예방하는 것은 농가의 수익을 보호하는 중요한 역할을 한다 [1].

농작물 질병 진단 방법은 크게 시각적 증상 관찰, 진단 키트 활용, 유전자 분석으로 분류된다. 특히, 시각적 증상 관찰은 비용이 저렴하며 즉시 결과를 확인할 수 있기 때문에 많이 사용되는 방법이다. 과거에는 사람이 직접 눈으로 확인하며 진단하였으나, 작물에 따라 질병의 종류가 다르며 그 진행 양상도 다르기 때문에 그 한계가 있다. 따라서 최근에는, 컴퓨터 비전과 인공지능을 이용한 자동화 병해 진단기술이 활발히 연구되고 있다 [2-5].

최근 IoT 기술과 스마트팜의 확산으로 환경데이터의 수집이 용이해지고 있다. 농장에서는 자동화, 작물모니터링 등의 목적으로 농작물 환경데이터를 수집하기 시작했다. 온도, 습도, 이산화탄소 농도 등 수집된 다양한 환경데이터는 농작물 영상 데이터와 함께 농작물 질병 예측에 유용하게 활용 될 수 있다.

Alves, L. 등은 온도, 습도, 누적열량 등의 데이터를 측정 후, Random Forest를 이용하여 올리브와 포도의 질병 유무

를 예측하였다 [6]. Sannakki, S. 등은 온도, 압력, 습도, 풍속 등을 측정하여 포도 노균병, 흰가루병, 그리고 탄저병을 진단하였다 [7].

대부분의 기존 연구에서는 영상 또는 시계열 데이터를 단일 모달리티의 데이터만을 사용하여 농작물 병해 또는 작물 병해 예측을 연구해왔다. 하지만, 서로 다른 센서로 측정된 두 가지 이상의 데이터를 함께 사용하여 농작물 병해를 예측하는 연구는 아직 부족한 상황이다.

딥러닝 기반의 농작물 영상 분석을 위하여 전이학습 (Transfer Learning)이 유용하게 활용될 수 있다. 하지만, 시계열 형태의 환경데이터를 분석하기 위해서는 기존의 전이 학습 방식으로는 한계가 있다. 시계열을 다루는 RNN (Recurrent Neural Network)와 LSTM (Long-Short Term Memory)를 이용한 모델은 구성이 정형화되어 있지 않으며 ImageNet [8]과 같은 사전 학습된 모델이 존재하지 않기 때문에 전이학습이 어려운 상황이다.

이러한 문제를 해결하기 위해 대조 학습 (Contrastive Learning)을 활용할 수 있다. 대조 학습은 사용자가 정의한 Pretext Task와 이를 통해 학습된 모델을 원하는 Task에 적용하는 Downstream Task로 구성된다. 잘 정의된 Pretext Task를 수행함으로써 Downstream Task 수행에 도움이 될 수 있는 사전학습 모델을 얻을 수 있다.

본 연구에서는 농작물 영상 데이터와 시계열 환경 데이터를 동시에 활용하는 멀티 모달 지도 대조 학습 (Multimodal Supervised Contrastive Learning)을 제안하고 이를 이용한 농작물 병해 진단 방법을 제안한다. 이를 위하여 기존의 대조 손실 함수를 멀티 모달 데이터에 적용할 수 있도록 수정하였다. 또한, t-SNE를 이용한 특징 (Feature) 시각화와 농

*Corresponding Author (khokh888@wku.ac.kr)

Received: Sep. 20, 2023, Revised: Oct. 25, 2023, Accepted: Nov. 7, 2023.

H. S. Lee: Daegu Gyeongbuk Medical Innovation Foundation (Researcher)

D. Y. Yeo: Korea Atomic Energy Research Institute (Senior Researcher)

G. S. Ham: Jeonbuk National University (Post Doctor)

K. H. Oh: Wonkwang University (Assistant Professor)

※ 이 성과는 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2021R1G1A1093546).

작물 병해 진단 예측 성능을 멀티 모달 자기 지도 대조 학습과 비교하여 그 우수함을 보이고자 한다.

II. 관련연구

1. 멀티 모달 학습

멀티 모달 데이터란 시각, 청각 그리고 온도와 같이 서로 다른 센서로부터 획득한 서로 다른 개별적인 특성을 가진 데이터를 의미한다 [9]. 딥러닝 모델이 이러한 멀티 모달 데이터를 학습하는 것을 멀티 모달 학습이라 한다. 일반적으로 한 종류의 데이터를 사용했을 때에 비하여 멀티 모달 데이터를 사용했을 때 성능이 더 향상된다고 알려져 있다. 최근 단일 데이터를 사용하여 해결하였던 문제를 멀티 모달 데이터를 이용하여 성능을 개선한 연구가 많이 연구되고 있다 [10-12]. 하지만 농작물 질병 진단을 위해서는 주로 영상 또는 환경데이터 하나만을 사용하는 연구가 주로 이루어졌으며, 멀티 모달 데이터가 주어졌을 때 어떻게 활용할 수 있는지에 대한 연구는 거의 이루어지지 않고 있다.

2. 대조 학습

대조 학습의 목적은 데이터로부터 좋은 특징을 추출할 수 있는 표현 능력 (Representation)을 학습하는 것을 목적으로 한다. 비지도 학습 중 하나인 자기 지도 대조 학습 (Self-Supervised Contrastive Learning)은 라벨링 없이 영상데이터만으로 좋은 표현능력을 얻을 수 있었다 [13]. 자기 지도 대조 학습은 불균형한 데이터 셋에 대해서도 강인하며 [14], 영상과 정형데이터를 함께 사용하는 멀티 모달 자기 대조 학습으로 확장될 수 있음을 보였다 [15, 16].

하지만 자기 지도 대조 학습은 배치 크기와 데이터 증강 방법에 민감하며 같은 레이블을 가진 데이터에 대해서도 유사성이 낮아지게 학습하는 모순점이 있다. 이를 개선하기 위하여 지도 대조 학습 (Supervised Contrastive Learning)이 제안되었으며 자기 지도 대조 학습보다 더 좋은 성능을 보여주었다 [17]. 하지만, 아직 멀티 모달 데이터에 대한 지도 대조 학습 연구는 아직 알려져 있지 않다.

본 논문에서는 멀티 모달 데이터에 대한 지도 대조 학습 방법과 멀티 모달 지도 대조 손실 함수를 제안하고 농작물 병해 진단 예측에 적용하고자 한다.

III. 농작물 병해 진단

본 연구에서는 국내 인공지능 경진대회 데이터 셋 '농업 환경 변화에 따른 작물 병해 진단 AI 경진대회'를 이용하였다 [18]. 대회의 목표는 주어진 작물 이미지와 시계열 형태의 환경데이터를 이용하여 작물의 종류, 병해 그리고 병해의 정도를 예측하는 것이다. LG AI Research에서 데이터를 수집하였으며, 데이콘(주)에서 데이터를 다운 받을 수 있다.

제공되는 데이터는 작물의 병해 이미지와 환경데이터가

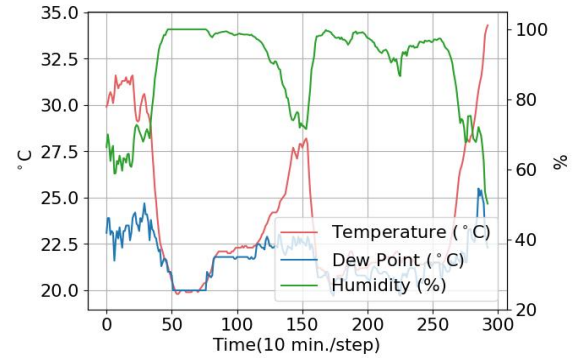


그림 1. 탄저병 증기에 걸린 포도의 환경데이터
Fig. 1. The environment data of grape with anthracnose-intermediate



그림 2. 탄저병 증기에 걸린 포도 사진
Fig. 2. The image of grape with anthracnose-intermediate

한 쌍을 이루며 이에 해당하는 농작물 종류, 병해의 이름, 그리고 병해의 정도가 라벨로서 제공된다. 작물 병해 이미지에는 병해를 입은 작물의 잎, 채소 또는 과일이 영상 중앙에 위치하도록 촬영되었다. 환경 데이터는 작물 병해 이미지 촬영 전 48 시간 동안 '온도', '습도', 그리고 '이슬점'의 환경 정보가 10분 주기로 데이터가 기록되어 있다.

그림 1과 2에 탄저병에 감염되어 증기 상태의 포도에 대한 사진과 환경데이터 일부 (온도 평균, 이슬점 평균, 그리고 습도 평균)를 도시하였다.

다음의 표 1에서 농작물 데이터를 농작물의 종류, 병해 그리고 병해의 정도에 따라 25개의 클래스로 나누었다. 본 연구에서는 대회의 목적과 동일하게 주어진 농작물 사진과 환경데이터를 이용하여 총 25개의 클래스를 예측하는 것을 목표로 한다.

IV. 멀티 모달 대조 학습

1. 대조 학습

이번 장에서는 멀티 모달 지도 대조 학습 설명을 위하여 대조 학습을 간략히 소개한다. 대조 학습은 사용자가 정의한 Pretext Task와 사전 학습 모델을 본래의 목적에 맞게 미세조정 (Fine-tuning)하는 일을 Downstream Task라 한

표 1. 농작물 병해 분류에 사용된 클래스 종류와 데이터 개수
Table 1. Class types and the number samples used for crop disease diagnosis

Crop	Disease	Severity	The number of samples
Strawberry	Normal		810
Tomato	Normal		143
Tomato	Powdery Mildew	Intermediate	189
Paprika	Normal		1177
Paprika	Powdery Mildew	Early	154
Paprika	Powdery Mildew	Intermediate	111
Paprika	Powdery Mildew	Terminally	42
Paprika	Calcium Deficiency	Early	166
Paprika	Deficiencies in Macronutrient (N)	Early	142
Paprika	Deficiencies in Macronutrient (P)	Early	156
Paprika	Deficiencies in Macronutrient (K)	Early	153
Cucumber	Normal		917
Chili	Normal		69
Chili	Anthracoese	Intermediate	99
Chili	Deficiencies in Macronutrient (N)	Early	148
Chili	Deficiencies in Macronutrient (P)	Early	159
Chili	Deficiencies in Macronutrient (K)	Early	157
Grape	Normal		828
Grape	Anthracoese	Early	40
Grape	Anthracoese	Intermediate	12
Grape	Downey Mildew	Early	13
Grape	Downey Mildew	Intermediate	29
Grape	Sunscald	Early	18
Grape	Sunscald	Terminally	14
Grape	Corky Core	Early	21
Total			5,767

다. Pretext Task에서는 동일한 영상을 데이터 증강을 통하여 서로 다른 두 영상을 생성한다.

자기 지도 학습에서는 같은 영상으로부터 증강된 두 영상이 Positive Pair가 되며 다른 영상들은 모두 Negative Pair가 된다. 각 영상은 Encoder f 와 Projection Head P 를 통과하여 저차원 벡터 z 로 사상된다. 본 연구에서는 z 는 데이터에 대한 특징이라 부르기로 한다. 그 후, Positive Pair의 특징은 특징공간에서 유사도가 가까워지도록 (Pull) 학습을 하고, Negative Pair는 유사도가 멀어지도록 (Push) 학습한다. 유사도의 계산은 Distance 함수를 사용할 수 있지만, 본 연구에서는 Cosine Similarity를 사용하였다.

지도 대조 학습에서는 Positive Pair 구성을 확장하여 같은 라벨을 가진 모든 영상을 포함하며, 다른 라벨을 가진 영상은 Negative Pair로 구성한다. 그 외 학습방법은 자기지도 학습과 유사하다.

2. 멀티 모달 지도 대조 학습

이번 장에서는 제안하는 멀티 모달 지도 대조 학습을 설명하고자 한다. 멀티 모달 지도 대조 학습을 설명하는 대표도는 그림 3과 같다. 멀티 모달 지도 대조 학습은 Positive Pair를 더 확장하여 같은 라벨을 가진 영상데이터와 함께 같은 라벨을 가진 환경데이터를 포함시킨다. 마찬가지로 Negative Pair는 서로 다른 라벨을 가진 영상데이터와 함께 환경데이터로 구성한다.

대조 학습을 위한 데이터 증강을 위하여 지도 대조 학습 [17]에서 사용한 방법과 같은 RandAugment [19]를 사용한다. 영상을 위한 데이터와 시계열 데이터를 위한 RandAugment 방법은 각각 독립적으로 적용되나 구현의 편의를 위하여 하이퍼파라미터는 동일한 값을 적용하였다. RandAugment에 적용된 데이터 증강방법의 종류는 'V 실험' 장에서 설명한다.

증강된 영상데이터와 시계열 데이터는 각 모달리티별로 구성된 f 와 P 를 통과하여 각 모달리티에 대한 특징 z 를 생성한다. 이 과정은 수식 (1)로 표현될 수 있다. 이 때, 아래 첨자 m 은 모달리티 (Image 또는 Sequence)를 의미하며 i 또는 s 가 될 수 있다.

$$z_m = P_m(f_m(x_m)), \quad m \in (i, s). \quad (1)$$

본 연구에서는 영상데이터를 처리하기 위한 f_i 로서 널리 사용되는 ImageNet을 사전 학습한 Resnet50 [20]을 사용하였고, 시계열 데이터를 처리하기 위한 f_s 로서 가장 간단한 형태의 LSTM 기반의 모델을 구성하였다. LSTM 기반 모델은 9개의 채널을 가진 시계열 입력 데이터를 입력받는다. Hidden_dim=200, num_layers=1을 사용하였다. Projection Head P 는 2개의 Dense Layer, Batch Normalization Layer, 그리고 ReLU 활성화 함수로 구성된 Fully Connected Layers이다. Projection Head를 통과한 특징 z 는 128차원이다.

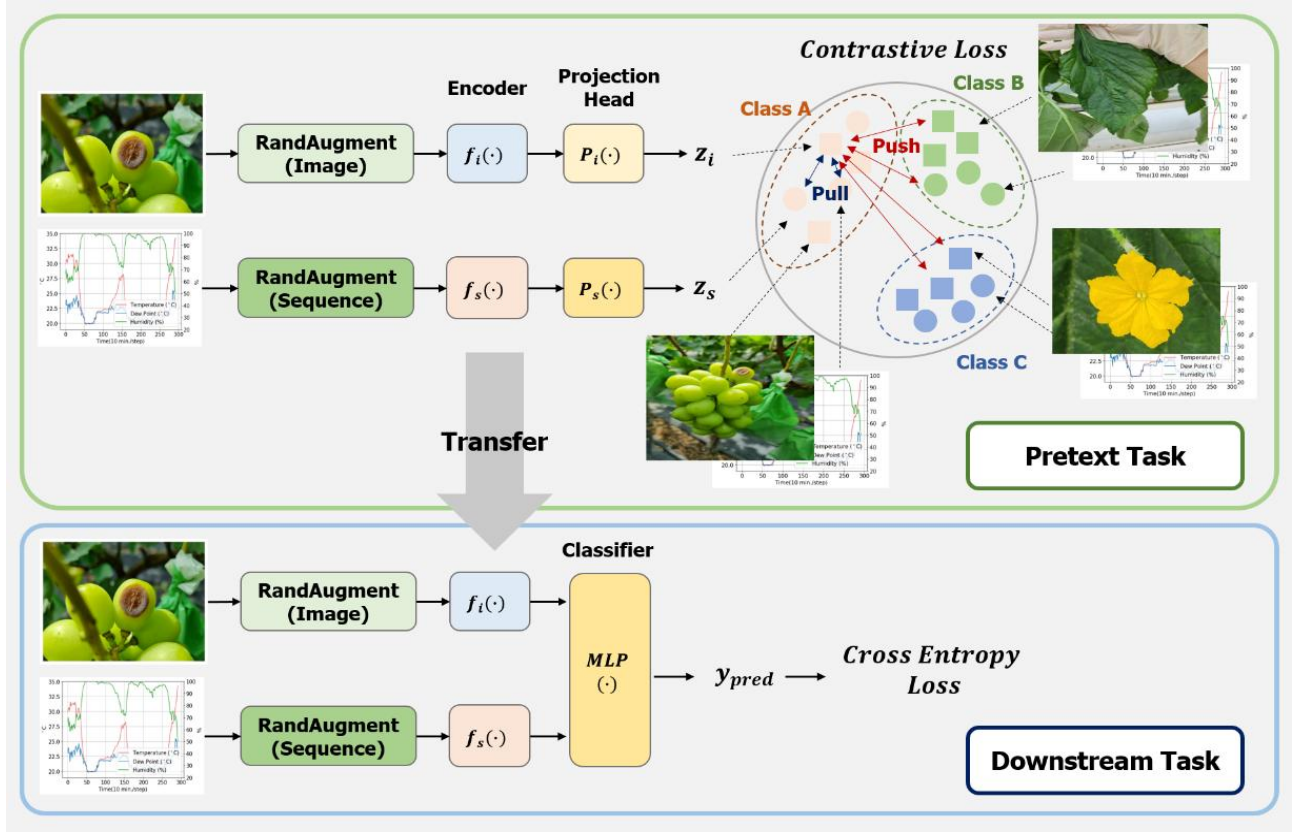


그림 3. 멀티 모달 지도 대조 학습 기반의 농작물 병해 진단 모델
Fig. 3. Multimodal supervised contrastive model for crop disease diagnosis

3. 멀티 모달 지도 대조 손실 함수

멀티 모달 지도 대조 손실 함수를 소개하기 위하여 몇 가지 기호를 정의한다. 먼저 우리는 대조 학습이 Stochastic Gradient Descent 방법을 사용하며 각 미니 배치의 크기는 n 이라고 가정한다. 주어진 배치에 대하여 데이터 셋의 인덱스 집합은 $I = \{1, \dots, n\}$ 이다. k 번째 샘플데이터는 영상데이터 x_i^k 와 시계열 환경데이터 x_s^k 가 한 쌍으로 존재하며, 각 데이터에 대한 라벨은 y^k 이다.

영상 데이터 x_i^k 에 대하여 멀티 모달 자기 지도 대조 학습에서는 다른 모달리티의 데이터 x_s^k 를 Positive Pair로 구성하였다. 멀티 모달 지도 대조 학습에서는 Positive Pair를 확장하여, $y^p = y^k$ 를 만족하는 모든 p 번째 샘플데이터 데이터, 즉 x_i^p 와 x_s^p 를 Positive Pair를 포함한다. 우리는 Positive Pair를 만족하는 인덱스 집합 $P(k) = \{k \in I : y^p = y^k\}$ 라 정의한다. 그 외 모든 데이터 셋을 Negative Pair로 구성한다. 위 내용을 바탕으로 영상데이터에 대한 지도 대조 손실 함수 $L(i, s)$ 는 기존의 지도 대조 손실 함수 [17]를 확장하여 수식 (2)와 같이 정의할 수 있다. 이 때, τ 는 온도상수로서 하이퍼파라미터이다.

마찬가지로, 환경데이터 x_s^k 의 Positive Pair는 x_i^k 와 $y^p = y^k$ 를 만족하는 모든 p 번째 샘플데이터 데이터 x_i^p 와 x_s^p 로 구성된다. 환경데이터에 대한 지도 대조 손실함수 $L(s, i)$

는 수식 (2)에서 i 와 s 위치를 바꿈으로 쉽게 구할 수 있다.

전체 손실 함수는 $L(i, s)$ 와 $L(s, i)$ 의 합으로 구성되며 스케일링 하이퍼파라미터를 λ 를 이용하여 다음의 수식 (3)으로 정의한다.

$$L(i, s) = \frac{\exp\left(\frac{\cos(z_i^k \cdot z_s^k)}{\tau}\right) + \sum_{\substack{l \in P(k) \\ l \neq k}} \exp\left(\frac{\cos(z_i^k \cdot z_l^l)}{\tau}\right) + \exp\left(\frac{\cos(z_i^k \cdot z_s^l)}{\tau}\right)}{\exp\left(\frac{\cos(z_i^k \cdot z_s^k)}{\tau}\right) + \sum_{\substack{a \in I_k \\ a \neq k}} \exp\left(\frac{\cos(z_i^k \cdot z_a^a)}{\tau}\right) + \exp\left(\frac{\cos(z_i^k \cdot z_s^a)}{\tau}\right)},$$

$$L_{\text{contrastive}} = \lambda L(i, s) + (1 - \lambda) L(s, i). \quad (3)$$

V. 실험

1. 데이터 셋

실험을 위하여 2장에서 소개한 '농업 환경 변화에 따른 작물 병해 진단 AI 경진대회' 데이터 셋을 이용하였다. 경진대회에서 제공하는 데이터 셋은 레이블을 모두 제공하는

훈련용 데이터와 레이블을 제공하지 않는 테스트 데이터로 구성된다. 본 연구에서는 레이블을 제공하는 훈련용 데이터만을 이용하였다. 데이터의 개수는 총 5767개이며, 본 연구에서는 학습데이터:검증데이터:테스트데이터 비율을 5:3:2로 나누어 사용하였다. 클래스에 따른 데이터가 불균형하기 데이터 셋을 나눌 때 그 비율이 유지되도록 나누었다. 또한, Weighted Sampler를 사용하여 부족한 클래스에 대한 데이터가 더 많이 학습되도록 설정하였다. 검증데이터는 오버피팅을 막기 위한 중간 저장용으로 사용하였고 성능평가는 테스트 데이터에 대하여 이루어졌다. 농작물 병해 진단 예측 성능지표로 AI 경진대회에서 사용한 성능지표와 동일한 Macro F1 Score를 사용하였다. Macro F1 Score는 각 클래스에 대한 F1 Score의 산술평균으로 정의되며, 다중 클래스 분류에서 많이 사용되는 지표 중 하나이다. Macro F1 Score는 식 (4)와 같이 표현될 수 있으며, 여기서 n 은 전체 클래스의 수이다.

$$Macro\ F1\ Score = \frac{1}{n} \sum_{i=1}^n F1\ Score_i. \quad (4)$$

2. 모델 학습

이번 장에서는 Pretext Task와 Downstream Task에서 딥러닝 모델이 학습한 환경에 대하여 서술하고자 한다. Pretext Task와 Downstream Task에서 증강된 데이터를 생성하기 위하여 우리는 사전연구 [16]에서 사용된 RandAugment를 사용하였다. RandAugment는 간단한 Grid Search만으로 지도 대조 학습에서 좋은 성능을 보장할 수 있는 방법이다. 영상과 시계열 데이터 증강에 사용된 방법은 표 2 와 같다. 시계열 데이터의 증강 방법에 대한 자세한 설명은 사전연구 [21]를 참고 바란다.

증강된 영상데이터는 256×256으로 크기가 축소되어 f_i 로 입력된다. 증강된 시계열 환경데이터는 데이터 앞부분의 320개 신호만을 추출하여 f_s 로 입력된다.

Pretext Task를 수행하기 위하여 Optimizer는 Adam, Learning Rate는 0.001, 그리고 Learning Rate Scheduler는

표 2. RandAugment에 사용된 증강 방법
Table 2. Augmentation types for RandAugment

Types of augmentation method for image		
• Identity	• ShearX	• ShearY
• TranslateX	• TranslateY	• Rotate
• AutoContrast	• Invert	• Equalize
• Solarize	• Posterize	• Color
• Brightness	• Sharpness	
Types of augmentation method for time series		
• Identity	• Flip	• Jitter
• WindowSlice	• Scalling	• TimeWarp
• MagitudeWarp		

Cosine Annealing Warm-Restarts를 사용하였다. 모든 실험에서 하이퍼파라미터 τ 는 0.07, 그리고 $\lambda=0.5$ 를 사용하였다. Pretext Task에서 Batch Size의 크기가 클수록 성능이 좋고 알려져 있다 [13]. 하지만, 본 연구에서는 실험을 통하여 너무 큰 Batch Size보다는 128개에서 가장 좋은 성능을 확인하였다. Epoch는 2000번 학습하였다.

Downstream Task는 Learning rate 0.0001, Batch size를 32개, 그리고 Epoch을 200번으로 설정하여 학습하였다. 미세조정단계에서는 정확도 수렴이 빠르기 때문에 낮은 학습률과 적은 학습 횟수를 선택하였다. 손실함수는 Cross-Entropy를 사용하였다. 또한 Encoder의 표현 능력만을 평가하기 위하여 Encoder의 역전과 학습은 멈추고 Linear Classifier만을 미세조정하였다. 이는 사전연구 [13]의 Linear Evaluation Protocol을 반영한 것이다.

학습사양은 AMD EPYC 7742 64-Core Processor, DDR 2TB, 그리고 NVIDIA A100-SXM4 환경에서 수행되었으며, Pytorch 버전 1.13.0으로 구현되었다. Downstream Task의 경우, 매 Epoch에 검증데이터에 대한 평가를 수행 후 가장 좋은 F1-Score를 가지는 모델을 최종 모델로 선정하였다.

3. t-SNE를 이용한 특징 시각화

멀티 모달 지도 대조 학습이 표현 능력을 잘 학습했는지를 직관적으로 이해하기 위하여, 우리는 t-SNE (t-distributed stochastic neighbor embedding)을 이용하여 Embedding 공간을 시각화 하였다. 테스트 데이터 셋을 입력 받고 각 모달리티에 대한 f 를 통과한 후 계산된 특징 z 를 t-SNE를 이용하여 2차원 공간에 시각화하였다. 우리는 제안하는 멀티 모달 지도 대조 학습에 대한 표현 학습 결과를 멀티 모달 자기 지도 대조 학습 [15, 16]과 비교를 하였다. t-SNE를 이용한 특징 시각화 결과, 같은 클래스에 속한 특징들은 가까워지고 다른 클래스에 속한 특징들은 멀어져야 좋은 표현 능력을 학습했다고 할 수 있다.

그림 4를 통하여, $f_i^{SupCon} > f_i^{SimCLR} > f_s^{SupCon} > f_s^{SimCLR}$ 순으로 표현 능력이 좋음을 확인할 수 있다. 위첨자는 멀티 모달 지도 대조 학습 (Multimodal SupCon) 또는 멀티 모달 자기 지도 학습 (Multimodal SimCLR)을 의미한다. 아래첨자는 데이터의 모달리티 (Image 또는 Sequence)를 의미한다.

영상 데이터에 대한 Encoder는 f_i^{SupCon} 와 f_i^{SimCLR} 모두 표현 능력을 잘 학습하였음을 확인할 수 있다. 반면, 환경데이터에 대한 표현 능력은 영상데이터에 비하여 표현 능력이 부족함을 확인할 수 있다. f_s^{SupCon} 은 같은 클래스에 속한 특징들이 군집화를 이루었지만, f_s^{SimCLR} 는 학습이 거의 되지 않은 것처럼 보인다.

우리는 위 결과로 부터 영상 데이터와 시계열 데이터에서 모두 멀티 모달 지도 대조 학습 결과가 멀티 모달 자기 지도 학습보다 더 좋은 표현 학습 능력을 가졌음을 확인할 수 있다.

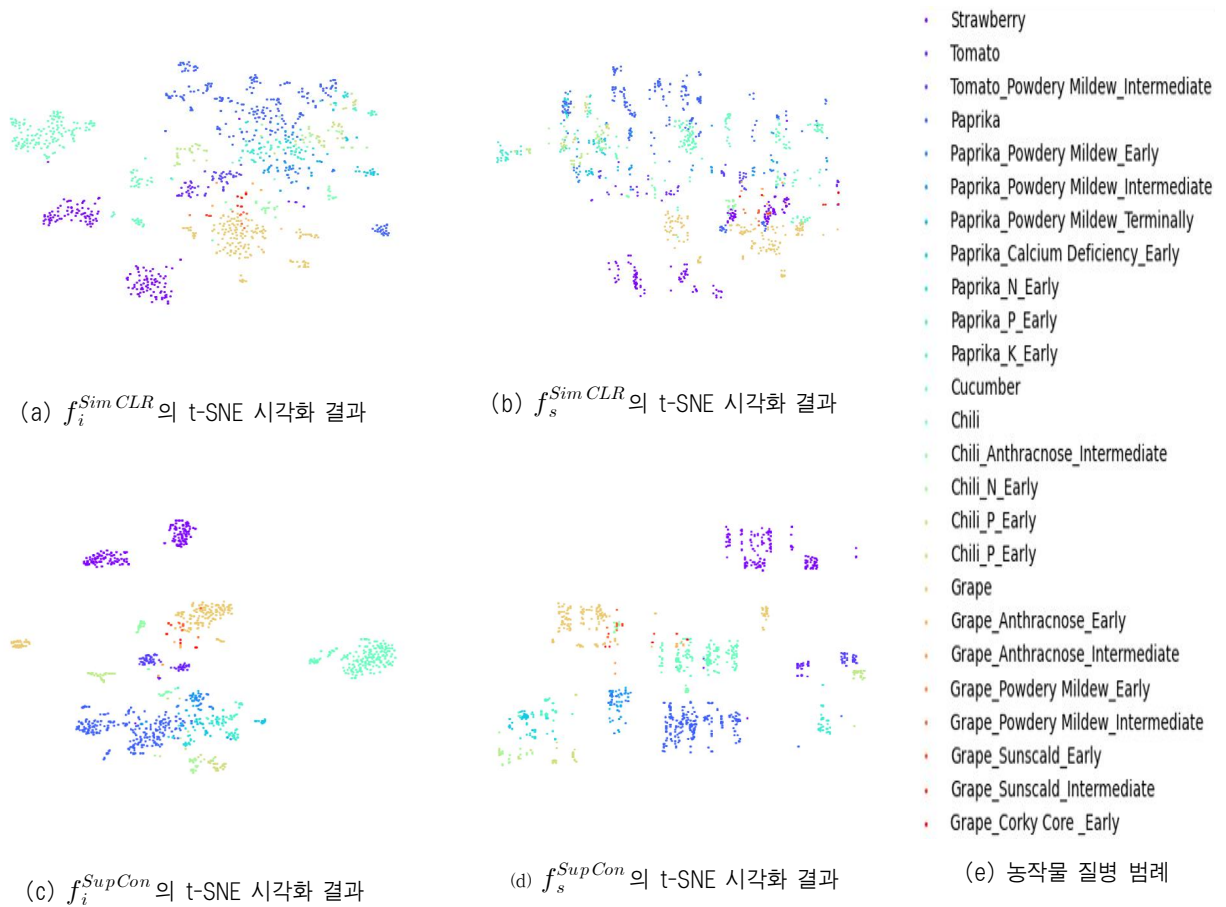


그림 4. 테스트 데이터 셋에 대한 t-SNE 시각화 결과
 Fig. 4. The visualization of t-SNE result on test dataset

표 3. 농작물 병해 진단 성능 비교표
 Table 3. Performance comparison for crop disease diagnosis

Method	Precision	Recall	F1 Score
w/o Pretext Task	0.909	0.862	0.878
Multimodal SimCLR [15, 16]	0.890	0.915	0.891
Multimodal SupCon	0.922	0.915	0.916

4. 농작물 병해 진단 성능

사전 학습된 모델의 정량적 평가를 위하여 농작물 병해 진단 예측 문제에 대해 미세조정을 수행하였다. 멀티 모달 지도 대조 학습의 성능 개선 효과를 확인하기 위하여 Pretext Task가 없는 일반적인 학습, 멀티 모달 자기 지도 대조 학습을 같은 방법으로 학습하였고 표 3에 표기하였다.

성능 지표는 Macro - Precision, Recall, F1-Score가 사용되었으며 모든 지표에서 멀티 모달 지도 대조 학습이 가장 우수함을 확인할 수 있다.

VI. 결론

최근 IoT 기술의 발전과 스마트팜의 확산으로 작물 관련 데이터 획득이 영상 데이터뿐만 아니라 온도, 이산화탄소 농도, 습도 등 다양한 시계열 환경 데이터를 포함하는 멀티 모달 데이터를 이용한 딥러닝 기반의 작물 병해 진단 연구가 점차적으로 중요해지고 있다. 본 연구에서는 기존의 멀티 모달 자기 지도 학습 개념을 확장하여 멀티 모달 지도 학습을 활용한 농작물 병해 진단 방법을 제안하였다.

먼저, 영상데이터와 시계열 환경데이터의 증강데이터를 RandAugment를 통해 생성한다. 증강된 데이터는 각 모달리티 별로 Encoder와 Projection Head를 통과하여 저차원 특징으로 사상된다. 그 후, 제안된 멀티 모달 지도 손실함수를 통해 같은 클래스에 속한 특징들은 서로 가까워지고, 다른 클래스에 속한 특징들은 서로 멀어지도록 학습된다. 이러한 과정을 거쳐 사전 학습된 모델을 농작물 병해 진단 과제에 미세조정하여 최종 모델을 생성한다.

제안된 방법이 멀티 모달 자기 지도 학습보다 성능이 우수함을 t-SNE를 이용한 시각화와 농작물 병해 진단 성능

비교를 통해 증명하였다.

본 연구의 후속 연구로 멀티 모달 데이터를 처리하기 위한 Encoder간의 성능 차이를 줄이는 방법을 연구하고자 한다. 'V 실험' 장 t-SNE의 시각화 결과에서 확인한 바와 같이 f_i 의 표현 능력이 f_s 보다 좋음을 확인할 수 있다. Pretext Task 학습 중, f_i 의 역전파를 중지하고 f_s 만을 학습하는 방법이 해결책이 될 수 있을 것이다.

또한, 농작물 질병과 질병의 정도는 계층구조를 이루고 있지만, 본 연구에서는 단순히 서로 다른 클래스로 분류하는 지도 대조 학습을 수행하였다. 저차원으로 사상된 표현 공간에서 '시설포도-탄저병-중기'의 특징과 '시설포도-탄저병-말기' 특징의 거리가 어느 정도 유지가 되어야 하지만, '고추-탄저병-중기'와의 거리보다는 거리가 짧아야 할 것이다. 특징공간에서 클래스의 계층 구조를 반영할 수 있는 대조 학습 방법을 후속연구로 수행하고자 한다.

References

- [1] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, Vol. 145, pp. 311-318, 2018.
- [2] J. A. Pandian, V. D. Kumar, O. Geman, M. Hnatiuc, M. Arif, K. Kanchanadevi, "Plant Disease Detection Using Deep Convolutional Neural Network," *Applied Sciences* Vol. 12, No. 14, pp. 6982, 2022.
- [3] H. Yu, J. Liu, C. Chen, A. A. Heidari, Q. Zhang, H. Chen, M. Mafarja, H. Turabieh, "Corn Leaf Diseases Diagnosis Based on K-means Clustering and Deep Learning," *IEEE Access*, Vol. 9, pp. 143824-143835, 2021.
- [4] H. S. Yoon, S. B. Jeong, "Performance Comparison of Base CNN Models in Transfer Learning for Crop Diseases Classification," *Journal of the Society of Korea Industrial and Systems Engineering*, Vol. 44, No. 3, pp. 33-38, 2021.
- [5] J. Li, Y. Qiao, S. Liu, J. Zhang, Z. Yang, M. Wang, "An Improved YOLOv5-based Vegetable Disease Detection Method," *Computers and Electronics in Agriculture*, Vol. 202, 107345, 2022.
- [6] L. Alves, R. R. Silva, J. Bernardino, "System to Predict Diseases in Vineyards and Olive Groves using Data Mining and Geolocation," *ICSOFIT*. 2018.
- [7] S. Sannakki, V. S. Rajpurohit, F. Sumira, H. Venkatesh, "A Neural Network Approach for Disease Forecasting in Grapes Using Weather Parameters," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, pp. 1-5, 2013.
- [8] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, "Imagenet: A Large-scale Hierarchical Image Database," 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248-255, 2009.
- [9] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A. Y. Ng, "Multimodal Deep Learning," *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 689-696, 2011.
- [10] J. Venugopalan, L. Tong, H. R. Hassanzadeh, M. D. Wang, "Multimodal Deep Learning Models for Early Detection of Alzheimer's Disease Stage," *Scientific reports*, Vol. 11, No. 1, pp. 3254, 2021.
- [11] W. Liu, W. L. Zheng, B. L. Lu, "Emotion Recognition Using Multimodal Deep Learning," *Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16 - 21, 2016, Proceedings, Part II 23*. Springer International Publishing, pp. 521-529, 2016.
- [12] T. G. Kim, B. J. Kang, M. Rho, S. Sezer, E. G. Im, "A Multimodal Deep Learning Method for Android Malware Detection Using Various Features," *IEEE Transactions on Information Forensics and Security*, Vol. 14, No. 3, pp. 773-788, 2018.
- [13] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," *International Conference on Machine Learning*. PMLR, pp. 1597-1607, 2020.
- [14] J. Zhang, J. Zou, Z. Su, J. Tang, Y. Kang, H. Xu, Z. Liu, S. Fan, "A Class-aware Supervised Contrastive Learning Framework for Imbalanced Fault Diagnosis," *Knowledge-Based Systems*, Vol. 252, 109437, 2022.
- [15] A. Taleb, M. Kirchler, R. Monti, C. Lippert, "Contig: Self-supervised Multimodal Contrastive Learning for Medical Imaging with Genetics," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 20908-20921, 2022.
- [16] P. Hager, M. J. Menten, D. Rueckert, "Best of Both Worlds: Multimodal Contrastive Learning with Tabular and Imaging Data," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23924-23935, 2023.
- [17] P. Khosla, P. Teterwak, C. Wang, "Supervised Contrastive Learning," *Advances in Neural Information Processing Systems* 33, pp. 18661-18673, 2020.
- [18] DACON AI Challenge Website. accessed: 2023-05-12. <https://dacon.io/competitions/official/235870/overview/description>
- [19] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, "RandAugment: Practical Automated Data Augmentation with a Reduced Search Space," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 702-703, 2020.
- [20] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [21] B. K. Iwana, S. Uchida, "An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks," *Plos one*, Vol. 16, No. 7, e0254841, 2021.

Hyunseok Lee (이 현 석)

2014 Department of Mechanical Engineering from Pusan National University (B.S.)
2016 Department of Mechanical Engineering from KAIST (M.S.)

Career:

2017~2019 Senior Research Engineer, LG Electronics
2019~2020 Senior Research Engineer, Korea Institute of Medical Microrobotics
2020~Researcher, Daegu Gyeongbuk Medical Innovation Foundation
Field of Interests: Contrastive Learning, Multimodal Deep Learning
Email: hs.lee@kmedihub.re.kr

Doyeob Yeo (여 도 엽)

2009 Department of Mathematical Sciences from KAIST (B.S.)
2012 Department of Mathematical Sciences from KAIST (M.S.)
2017 Department of Mathematical Sciences from KAIST (Ph.D.)

Career:

2017~2020 Postdoctoral Researcher, Electronics and Telecommunications Research Institute
2020~2021 Senior Research Engineer, Electronics and Telecommunications Research Institute
2021~Senior Research Engineer, Korea Atomic Energy Research Institute
Field of Interests: Generative Deep Learning & Image Processing
Email: yeody@kaeri.re.kr

Gyu-Sung Ham (함 규 성)

2018 Department of Computer Engineering from Wonkwang University (B.S.)
2020 Department of Computer Engineering from Wonkwang University (M.S.)
2023 Department of Computer Engineering from Wonkwang University (Ph.D.)

Career:

2023~Postdoctoral Researcher, Jeonbuk National University
Field of Interests: Deep Learning, Big data System, Distributed System, System Authentication.
Email: hamsung1231@gmail.com

Kanghan Oh (오 강 한)

2010 Department of Computer Engineering from Honam University (B.S.)
2013 Department of Electronics & Computer Engineering from Chonnam National University (M.S.)
2017 Department of Electronics & Computer Engineering from Chonnam National University (Ph.D.)

2021~Department of Computer and Software Engineering, Wonkwang University (Prof.)

Career:

2017~2020 Postdoctoral Researcher, Jeonbuk National University
2020~2021 Senior Researcher, Daegu Gyeongbuk Medical Innovation Foundation
Field of Interests: Computer Vision, Deep Learning
Email: khoh888@wku.ac.kr