



외래잡초 분류 : 합성곱 신경망 기반 계층적 구조

유광현*¹, 이재원*², 보호앙트롱*³, 당탄부*⁴, 후이트완녕*⁵, 이주환*⁶, 신도성*⁷, 김진영**

Exotic Weeds Classification : Hierarchical Approach with Convolutional Neural Network

Gwanghyun Yu*¹, Jaewon Lee*², Vo Hoang Trong*³, Dang Thanh Vu*⁴, Huy Toan Nguyen*⁵, JooHwan Lee*⁶, Dosung Shin*⁷, and Jinyoung Kim**

본 연구는 농촌진흥청 연구사업(세부과제번호: PJ01385501)의 지원에 이루어진 것임.

요 약

잡초는 농작물에 막대한 피해를 주는 주요한 개체이다. 잡초를 효과적으로 제거하기 위해서는 정확한 분류를 하고 제초제를 사용하여야 한다. 컴퓨팅 기술의 발전으로, 영상 기반의 기계학습 방법들이 이 분야에서 연구되고 있고, 특히 합성곱 신경망 기반의 모델들이 공공데이터에서 좋은 성능을 보이고 있다. 하지만 실제 응용단계에서는 많은 파라미터 수와 연산량 때문에 GPU와 같은 좋은 하드웨어 조건에서만 잘 작동된다. 본 논문은 계층적 구조 기반의 딥러닝 모델을 제안한다. 실험 결과, 제안된 모델은 적은 파라미터 수로 21개의 외래 잡초 종을 최대 97.2612%의 정확도로 성공적으로 분류하였다. 이를 통해 적은 수의 파라미터를 사용하는 제안된 모델은 네트워크 기반의 분류 서비스에서 적용될 수 있을 것으로 기대된다.

Abstract

Weeds are a major object which is very harmful to crops. To remove the weeds effectively, we have to classify them accurately and use herbicides. As computing technology has developed, image-based machine learning methods have been studied in this field, specially convolutional neural network(CNN) based models have shown good performance in public image dataset. However, CNN with numerous training parameters and high computational amount. Thus, it works under high hardware condition of expensive GPUs in real application. To solve these problems, in this paper, a hierarchical architecture based deep-learning model is proposed. The experimental results show that the proposed model successfully classify 21 species of the exotic weeds. That is, the model achieve 97.2612% accuracy with a small number of parameters. Our proposed model with a few parameters is expected to be applicable to actual application of network based classification services.

Keywords

exotic weeds, weed classification, hierarchical architecture, convolutional neural networks

* 전남대학교 전자컴퓨터공학과

- ORCID¹: <http://orcid.org/0000-0003-0434-5260>
 - ORCID²: <http://orcid.org/0000-0002-3293-4540>
 - ORCID³: <http://orcid.org/0000-0001-6969-3042>
 - ORCID⁴: <http://orcid.org/0000-0003-0743-8251>
 - ORCID⁵: <http://orcid.org/0000-0002-4986-9175>
 - ORCID⁶: <http://orcid.org/0000-0002-1553-9637>
 - ORCID⁷: <http://orcid.org/0000-0003-4924-2062>

** 전남대학교 전자컴퓨터공학과 교수(교신저자)

- ORCID: <http://orcid.org/0000-0002-4986-8980>

· Received: Nov. 01, 2019, Revised: Dec. 18, 2019, Accepted: Dec. 21, 2019

· Corresponding Author: Jinyoung Kim

Dept. of Electronics and Computer Eng, Chonnam National University

Tel.: +82-62-530-1757, Email: beyondi@jnu.ac.kr

I. 서론

잡초는 농작물에 막대한 피해를 주는 개체이다. 수십 년간 잡초를 억제할 수 있도록 엄청난 노력을 했음에도 불구하고, 여전히 농민들은 잡초에 의한 고질적인 피해를 입고 있다. 잡초는 경작지에서 아무데서나 아무 이유 없이 자라고, 농작물과 수분, 영양소, 일조량 등을 경쟁하며, 최종적으로는 생산량에 악영향을 미친다[1]. 수많은 연구 결과들이 농업 생산량 감소와 잡초 증가의 강력한 연관성을 증명하고 있다.[2].

잡초 제어를 위해서는 정확한 잡초 판별이 중요한데, 그 이유는 잡초의 종마다 사용하는 제초제의 성분과 관리하는 방법이 상이하기 때문이다. 이를 위해 잡초 판별 연구는 센서 기반 방법과 영상 판별 분야에서 두드러진 결과를 얻고 있으며 최근에는 딥러닝 기반의 기법이 크게 대두된다.

본 논문은 합리적인 파라미터수와 연산 시간을 줄이면서도 많은 수의 잡초의 종을 분류할 수 있는 딥러닝 모델 구조를 제안한다. 본 논문의 2장에서는 식물 및 잡초를 판별하는 과거의 연구들을 분석하고, 3장에서는 데이터셋과 데이터셋을 활용한 딥러닝 기반의 단일 모델 설명하고, 4장에서는 제안하는 딥러닝 기반의 새로운 구조를 설명하며, 5장에서는 제안하는 모델과 기존의 모델들에 대한 비교 실험 및 결과를 통해, 6장에서 결론을 짓는 것으로 구성된다.

II. 관련 연구

잡초를 제어하기 위해서는 경작지에서 잡초를 정확히 판별하는 것이 중요하다. 잡초의 종마다 사용하는 제초제와 관리 방법이 다르기 때문이다. 판별하려는 대상, 기술, 방법에 따라 다양한 잡초 분류·동정법이 있다.

그 중 가장 고전적으로 사람의 눈과 손을 이용하여 잡초를 분류·동정하는 방법이 있지만[3], 수작업은 지루하고, 잡초에 해당하는 제초제를 사용하려고 할 때, 전문가가 아니면 정확한 동정이 어려우며, 범위가 넓어질수록 인건비가 증가하는 문제가 있다. 기술의 발전으로 센서 또는 영상 기반의 다양한 잡초 판별 방법이 연구되었다.

원격 센서를 이용하여 경작지의 잡초 밀집과 분포를 성공적으로 연구하였고[4], 경작지의 구획을 조밀하게 분석하여 잡초를 분류하는 연구도 진행되었다[5]. 하지만 센서를 이용한 잡초 판별 방법은 개별 잡초의 분류보다는 일정 범위 내의 밀집과 분포를 확인하는데 더 적합한 방법들이다.

영상을 이용한 방법으로는, 8개의 형태학적 특성과 컬러 특성을 하나의 특징 공간으로 만들어 퍼지 논리 기반의 멤버십 함수를 이용하여 정확도를 클래스에 따라 51%에서 95% 달성하였고[6], 10개의 식물 종에 대한 잎의 모양과 질감을 특징으로 제프리 발산법을 사용하여 85%의 정확도를 보였다[7]. 위와 같은 방법들은 식물 분류에 한정되어 있어 잡초 분류에 대한 정확한 결과를 얻을 수 없는 단점이 있다. 또한 활성 모양 모델을 적용하여 이미지로부터 잡초를 추출하고 k개의 신경망을 통해 분류하여 91.3%의 정확도를 보인 방법[8]과 순수 베이지 이론과 가우시안 합성 모델 알고리즘을 이용하여 정확도 85%를 달성한 방법[9]이 있다. 딥러닝 기반의 영상 인식의 성능이 향상되면서, 합성곱 신경망을 이용하여 식물의 특징을 추출하는 방식으로 식물을 분류하고 스마트폰 어플리케이션 시스템을 만든 연구가 진행되었고[10], 이미지 분할 없이 중복되는 패치를 사용하는 입력을 통해 딥러닝 기반의 잡초 분류 모델로 93.8%의 정확도를 보인 연구[11]와 22개종의 식물 10,413장을 ResNet 모델을 이용하여 86.2%의 정확도를 보인 연구[12]가 진행되었다. 위와 같은 모델들은 영상처리 기반부터 딥러닝 기반까지 다양한 알고리즘을 사용하여 잡초 분류를 하였지만, 실제 경작지에 존재하는 잡초 데이터나 많은 종의 잡초를 사용하지 않은 단점이 있다.

위와 같은 문제를 해결하기 위하여 이전 논문에서는 16개의 종의 실제 경작지에서 자라고 있는 외래잡초에 대하여 3가지의 딥러닝 모델을 사용한 분류방법과 앙상블 방법을 제안하여 95.9%의 정확도를 보였다[13]. 하지만 여러 개의 딥러닝 모델을 사용하게 되면 많은 수의 파라미터와 긴 처리 시간이 소요되게 된다. 따라서 본 논문은 전체 파라미터수와 연산 시간을 효과적으로 줄이면서도 많은 수의 외래잡초 종을 분류할 수 있는 딥러닝 모델 구조를 제안한다.

III. 데이터셋 및 딥러닝 모델

3.1 데이터셋

본 논문에서 사용할 잡초 데이터는 외래잡초 21 개종이다. 잡초 영상은 DSLR 카메라, 스마트폰 카메라와 같은 다양한 촬영 장비를 통해 전국 각지에서 촬영되었다. 촬영된 이미지 예는 그림 1과 같다.

촬영된 이미지는 전국 각지의 산, 들판, 도로 및 경작지와 같은 곳에서 촬영된 영상이고 그림 1처럼 대상이 되는 대상 외래잡초 외에 다양한 배경이 포함되어 있다. 원본 영상을 그대로 사용하여 영상처리 기반의 데이터 증강기법을 사용하게 되면, 주 객체 외에 불필요한 배경이 많이 포함되기 때문에, 사람의 눈을 통해 잎, 꽃 및 전초를 정확히 구분하여 잘라내었다. 잘라낸 예시는 그림 2와 같다.



그림 1. 외래잡초 원본 영상
Fig. 1. Original image of exotic weeds



그림 2. 원본 영상에서 잘라낸 이미지
Fig. 2. Cropped image from original image

표 1. 외래잡초 데이터셋

Table 1. Exotic weed dataset

Family	Class	Before augmentation	After augmentation
비름	털비름	247	6,425
	가는털비름	174	4,138
	민털비름	177	5,455
	청비름	168	2,310
	개비름	187	6,612
현삼	큰개불알풀	145	1,883
	개불알풀	117	1,969
명아주	흰명아주	647	2,804
	좁명아주	144	1,692
메꽃	둥근잎유홍초	408	2,300
국화	돼지풀	70	1,219
	단풍잎돼지풀	491	23,819
	미국가막사리	389	20,713
	망초	399	24,512
	큰망초	163	8,836
	털별꽃아재비	590	24,412
	개썩갓	434	23,500
	큰방가지똥	588	17,281
	방가지똥	505	15,448
	큰도꼬마리	504	13,053
	도깨비바늘	248	805

DB를 구축함에 있어 고해상도의 원본 외래잡초 영상에서 외래잡초 각각의 특성을 효과적으로 반영할 수 있도록 잎, 꽃 및 전초를 최대한 배경을 제외하여 직사각형 형태로 잘라내었으며, 종에 따라 최소 4배에서 수십 배의 데이터 증강을 수행하였다. 수작업 이후 구축된 데이터셋은 표 1과 같다.

또한 본 논문에서 제안하는 계층적 구조 기반의 딥러닝 모델을 개발하기 위해 각각의 외래잡초를 과와 클래스(종)로 분류하였다. 분류 기준은 식물분류학에 따르며, 같은 과에 있는 외래잡초 클래스의 특징은 비슷하며, 다른 과에 있는 외래잡초 클래스의 특징은 많이 다르다.

3.2 적용 딥러닝 모델

딥러닝 기반의 이미지 분류 모델로 현재 각광받고 있는 것은 합성곱 신경망(Convolutional neural networks)이다. 합성곱 신경망은 1998년 컨벌루션층(Convolutional layer), 맥스풀링(Max pooling)과 완전연결층(Fully connected layer)을 기반으로 하여

MNIST라는 숫자 인식 데이터셋에서 높은 정확도를 올린 LeNet[14]을 시초로 꾸준히 연구되어 왔다. 2012년 AlexNet[15]이 ImageNet에서 공개한 1000개 종류 분류 경진대회에서 1년 전 대비 혁신적인 결과를 도출하면서, 이미지 분류 분야에서 합성곱 신경망이 좋은 성능을 보인다는 것을 알게 되었다. 2014년 인간의 이미지 분류 테스트 결과인 5% 오차와 거의 근접한 VGGNet[16]과 GoogleNet[17]이 등장하였고 마의 8개 층을 각각 독특한 방법으로 극복하여 각각 19개, 22개의 깊은 망으로 높은 인식률을 보였다. 2015년 ResNet[18]이 152개의 깊은 망을 독자적인 짧은 연결선(Identity short connection) 아이디어를 통해 구현하며 인간의 분류 인식률을 초월하였다. 그 후 매년 오차를 줄이고 있으며, 강화학습(Reinforcement learning)을 접목한 새로운 합성곱 신경망 모델인 NASNet[19]과 학습에 필요한 파라미터수와 처리속도를 개선한 MobileNet[20], 더 적은 파라미터 수로 정확도는 유지하는 SqueezeNet[21] 및 캡슐이라는 다른 방식으로 합성곱 신경망을 재해석한 CapsuleNet[22]이 등장하였다.

본 논문에서는 가장 안정적이고 효율이 좋은 합성곱 신경망 구조라고 일컬어지는 ResNet[18]과 파라미터 수를 획기적으로 줄여 처리 시간을 개선한 MobileNetV2[23]를 중점으로 사용하여 계층적 구조 기반의 합성곱 신경망 모델의 설명 및 실험을 진행한다.

IV. 제안 계층적 구조 기반 CNN

본 논문에서 제안하는 계층적 구조 기반의 합성곱 신경망 모델은 직관적으로 외래잡초와 같이 계층적 구조를 가지고 분류되는 데이터에 대하여 분할 정보 알고리즘 개념을 적용하여 적은 파라미터 수 및 연산 속도를 개선한 효율적인 딥러닝 모델을 구성한 것이다.

4.1 학습 모델에 대한 수식적 분석

이미지 인식 문제는 수학적으로는 조건부 베이저안 확률(Bayesian probability)로 볼 수 있는데, 입력(주어진 외래잡초 데이터 이미지 또는 외래잡초 데이터로부터 어떤 방법을 사용하여 얻어진 특징벡

터)을 X_i , 정답 클래스(Class)를 Y_i 그리고 정답 과(Family)를 F_i 라고 하자. 이 때, $i = 1, \dots, N$ 이며, N 은 학습 데이터 샘플을 의미한다. 모든 정답 클래스 Y_i 그리고 과 F_i 는 온-핫 인코딩(One-hot encoding) 벡터로 구성되며, Y_i 와 F_i 는 $Y_i \in \{0,1\}^M$ 과 $F_i \in \{0,1\}^K$ 이며, M, K 는 각각 정답 클래스와 과의 개수이다. 이와 같은 기본 표기법으로 다음과 같이 세 종류의 학습 모델을 정의할 수 있다.

$$P_w(\hat{Y}|X) \quad (1)$$

$$P_w(\hat{Y}, \hat{F}|X) \quad (2)$$

$$P_w(\hat{Y}|\hat{F}, X) \quad (3)$$

그림 3을 참조하여, 식 (1)의 학습 모델은 클래스를 예측하는 플랫 분류기(Flat classifier)이고, 식 (2)의 학습 모델은 클래스와 과를 결합하여 클래스와 과를 예측하는 글로벌 분류기(Glober classifier)이며, 식 (3)의 학습 모델이 과에 대한 조건부로 클래스를 예측하는 계층적 구조의 로컬 분류기(Hierarchically structured local classifier)이다.

각각의 학습 모델은 자체 학습된 가중치 W 로 예측된다. \hat{Y} 와 \hat{F} 는 입력된 특징 벡터 X 로부터 각각 예측된 정답의 클래스와 과이다. 최대 우도 추정(Maximum likelihood estimation)을 얻기 위해, 예측된 모델 파라미터의 가중치 W 를 고려한다. 그림 3을 참조하여 각 수식에 대한 자세한 설명은 다음과 같다.

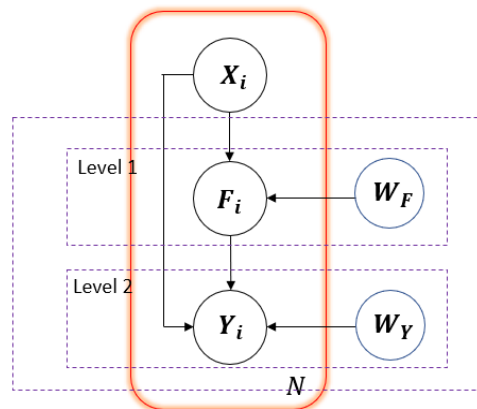


그림 3. 학습 모델에 대한 도식화
Fig. 3. Graphical model of learning framework

식 (2)는 레벨 1과 레벨 2를 결합한 클래스와 과의 예측 모델은 가장 바깥쪽 보라색 점선으로 표시되고, 정답 클래스(Y_i)와 정답 과(F_i)의 모두를 같은 수준으로 둔다. 식 (1)은 레벨 2의 보라색 점선으로 표시된다. 마지막으로 식 (3)은 주황색 박스로 표현되는데, 모든 정답 클래스(Y_i)는 정답 과(F_i)의 조건부로 분포된다.

4.2 목적 함수(Objective function)

학습 모델에 사용되는 가중치를 학습하기 위해 딥러닝 분야에서 잘 알려진 최대 우도 추정 기반의 예측 모델을 사용한다. 다수 클래스 분류 문제에서, 정답을 예측하는 것은 다항 분포를 추정하는 것이다. 클래스 예측 모델은 다음 식과 같다.

$$\hat{Y}_i|X_i \sim M(\mu_1, \dots, \mu_M), P(\hat{Y}_i|X_i) = \prod_{k=1}^M \mu_k^{y_k} \quad (4)$$

식 (4)의 μ_k 는 클래스 k 에 포함되는 X_i 의 확률로, $0 \leq \mu_k = P(\hat{Y}_k=1|X_i) = f_W(X_i) \leq 1$ 로 표현되며, f_W 는 합성곱 신경망 구조로부터 미리 정의된 가중치 함수이다. 모든 학습 데이터의 로그 우도 함수(Log-likelihood function)은 다음과 같다.

$$L = \log \prod_i^N \prod_k^M \mu_{ik}^{y_{ik}} = \sum_i^N \sum_k^M y_{ik} \log(\mu_{ik}) \quad (5)$$

학습 데이터 샘플 (X_i, Y_i)를 사용한다면, 로그 우도 함수를 최대화 하는 것은 크로스-엔트로피 로스 함수(Cross-entropy loss function)를 최소화시키는 것과 같다. 따라서 이것을 다시 표현하면 다음과 같다.

$$CrossEntropy = - \sum_i^N \sum_k^M y_{ik} \log(\mu_{ik}) \quad (6)$$

같은 방법으로 클래스와 과를 결합한 예측 모델과 과에 대한 조건부로 클래스를 예측하는 모델도 다항 분포를 추정한 것처럼 계산할 수 있다. 하지만 위와 같은 모델 각각은 로스(Loss) 함수에 더 추가적인 부분이 필요하다. 글로벌 분류기라고 불리는

클래스와 과를 결합하여 클래스와 과를 예측하는 모델은 한 쌍의 (Y, F)를 동시에 고려해야 한다. 따라서 온-핫 인코딩 벡터는 $M \times K$ 만큼의 요소를 가져야 하고, 이는 $Z = Y \times F^T \in \{0,1\}^{M \times K}$ 를 의미한다. 계층적 구조의 로컬 분류기라고 불리는 모델은 마지막 단계에서 결정되기 전에 이전 단계에서 입력에 대한 정보를 제공받는다. 그러므로 하나의 목적 요소보다 더 많은 것을 포함하게 된다.

4.3 계층적 구조의 합성곱 신경망

계층적 구조의 학습을 설명하는 간단한 내용은 그림 4와 같이 과와 클래스를 단계에 맞게 분리하여 각각의 분류기를 학습하는 것이다.

본 논문에서는 모든 클래스의 학습 데이터를 사용하여 첫 분류기에서 과를 분류하고, 전체 클래스를 과별로 나누어 5개의 과 그룹으로 만들어 각각을 5개의 보조 분류기를 사용하는 것을 확정하였다. 이는 각각의 분류기 모델이 딥러닝 구조에서 각각의 노드로 동작하는 것처럼 보이거나, 결정 트리 구조에서 부모 가지와 자식 가지로 구성되어 학습되는 것과 같다.

계층적 구조의 모델 학습에서 모델 선택을 효율적으로 채택하기 위한 필수 조건은, 계층적 구조에서 모든 모델의 복잡도가 단일 클래스 예측 모델보다 낮아야 한다. 일반적으로 다수 모델의 연산 복잡도는 계층적 구조의 깊이와 수준 및 보조 분류기 개수 등과 같은 지표에 의해 증가한다. 하지만 제안하는 계층적 구조의 분류 모델은 일반적인 단일의 딥러닝 모델과 비교할 때, 단순한 딥러닝 모델을 사용하여 복잡도가 낮아지는 강점이 있다.

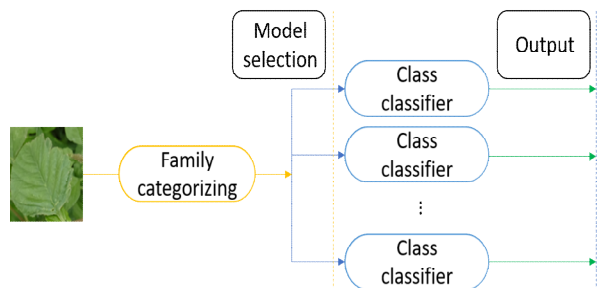


그림 4. 간단한 계층적 구조
Fig. 4. Simplified hierarchical architecture

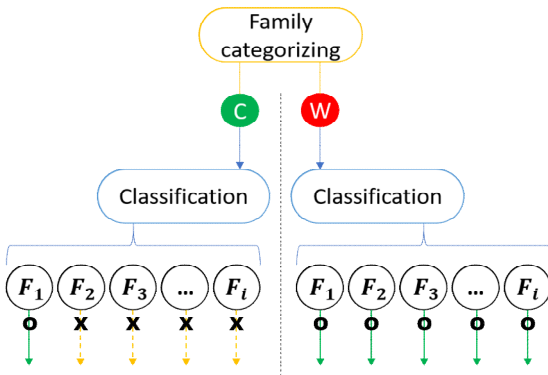


그림 5. 잘못된 과를 분류한 계층적 구조의 보상, C는 정확히 인식, W는 오인식

Fig. 5. Reward of hierarchical architecture from mis-family classification. circled C and W means correct and wrong classification respectively

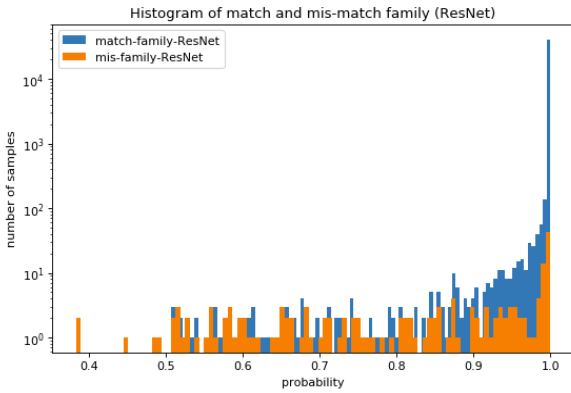


그림 6. ResNet 기반 계층적 구조 모델의 과의 히스토그램(로그 스케일)

Fig. 6. ResNet based hierarchical architecture model of family histogram(log scale)

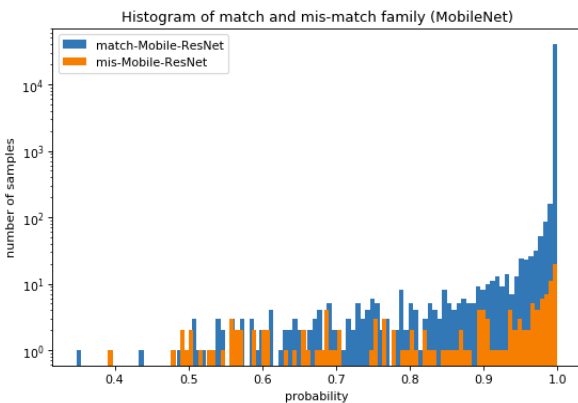


그림 7. MobileNetV2 기반 계층적 구조 모델의 과의 히스토그램(로그 스케일)

Fig. 7. MobileNetV2 based hierarchical architecture model of family histogram(log scale)

본 논문에서는 이와 같은 이점을 활용하여 각각의 딥러닝 모델을 단순한 모델로 사용하여 실험하였다.

하지만 계층적 구조의 모델 학습에 단점이 존재한다. 만약 첫 번째 계층인 과에서 정확한 정답을 예측하지 못하면, 다음 단계 인식기는 의미가 없다. 즉, 정답 과를 정확히 예측하지 못한 입력은 당연히 오인식이 된다.

이를 피하기 위해 그림 5와 같이 오인식 과를 예측했을 때 보상을 줄 수 있는 방법을 제안한다.

$$P(Y=n|X) = \sum_{i=1}^m P(Y=n|F_i, X) P(F_i|X) \quad (9)$$

$$P(Y=n|F_i, X) = \begin{cases} 0 & P_i \text{ is calculated by } i^{\text{th}} \text{ classifier} \\ P_i(Y=n|X) & \text{if class } n \text{ isn't } \in \text{family } F_i \end{cases} \quad (10)$$

식 (10)에서 \$Y\$는 클래스, \$n\$은 클래스 번호, \$X\$는 입력, \$F_i\$는 과의 번호를 의미하며, \$P(Y=n|F_i, X)\$에서 과 분류기가 입력된 이미지를 정확히 정답으로 예측했을 때는 0으로 수렴하여 아무 영향을 미치지 않아 해당되는 클래스 분류기를 사용하게 되고, 반대의 경우에는 모든 클래스 분류기의 정답 확률을 계산하여 가장 높은 확률 값을 가지는 클래스를 찾는다.

테스트에서 과의 계층에서 입력된 영상이 옳게 판별되었는지 아닌지를 확인하기 위하여, 그림 6, 7과 같이, 계층적 구조의 합성곱 신경망 모델을 ResNet 및 MobileNetV2 기반으로 외래잡초 21개 클래스로 실험하여 과의 계층에서 정답을 제대로 예측한 샘플과 정답을 제대로 예측하지 못한 샘플의 히스토그램으로 나타내었다. 과의 계층에 대한 정확한 테스트 결과는 표 2와 같다.

표 2의 결과를 토대로 0.99를 문턱값의 기준으로 두어, 1.0 미만일 경우 \$P_i(Y=n|X)\$로 수렴하게 하여 오인식된 과에 대한 계층에 보상을 준다. 즉, 오인식된 과는 그림 5처럼 W 경우로 진행되어 전체 클래스 분류기의 확률을 모두 계산하고, 그 중 가장 확률이 높은 클래스를 선택하도록 한다.

표 2. 계층적 구조 모델의 과 오인식 비율

Table 2. Hierarchical architecture model based ratio of family mis-classification

Threshold	ResNet[20] based			MobileNetV2[25] based		
	miss	samples	ratio(%)	miss	samples	ratio(%)
0.5	5	170	2.9	5	152	3.3
0.6	27		15.9	20		13.2
0.7	46		27.1	37		24.3
0.8	63		37.1	50		32.9
0.9	87		51.2	70		46.1
0.91	89		52.4	76		50.0
0.92	93		54.7	78		51.3
0.93	96		56.5	90		59.2
0.94	101		59.4	94		61.8
0.95	103		60.6	92		60.5
0.96	109		64.1	98		64.5
0.97	111		65.3	104		68.4
0.98	121		71.2	114		75.0
0.99	123		72.4	123		80.9
1	170		100.0	152		100.0

V. 실험 및 결과

5.1 실험 환경

본 논문에서 제안하는 계층적 구조의 합성곱 신경망 모델 실험을 위해 3장에서 설명한 외래잡초 데이터 21개 클래스를 사용하였으며, 21개 클래스는 다시 5개의 과로 구성된 계층적 데이터이다. 데이터 증강은 일반적으로 딥러닝 모델을 학습할 때 사용하는 방식이 아닌 수작업의 잘라내기를 통해 잡음이 포함된 배경을 최대한 배제하고 관심 객체를 각각의 클래스에 대한 전문가의 식견을 토대로 잘라내었다. 데이터 증강 후 모든 데이터는 무작위로 섞어 6:2:2의 비율로 학습(Train), 검증(Validation) 및 테스트(Test)로 구성하였다.

딥러닝 모델의 학습 및 테스트 구현은 Python 기반 PyTorch와 Keras로 수행하였으며, 학습과 테스트에 사용된 하드웨어 환경은 Intel(R) Core(TM) i9-7900X CPU 3.30GHz의 CPU, 128GByte의 DDR4 RAM, SamSung 970 PRO SSD의 하드디스크 및 NVIDIA TITAN V, 12GByte 내장 RAM으로 구성되었고, 마지막으로 GPU 소프트웨어 환경은 CUDA 10.0 Toolkit과 cuDNN으로 이루어졌다.

5.2 전이학습(Transfer learning)

최근에 딥러닝 신경망을 학습하는데 있어서 초기 값을 랜덤한 가중치와 편이(Bias)를 주기보다는, 이미 학습이 잘 되어 좋은 성과를 내는 모델의 가중치와 바이어스를 가지고 와서 해결하려는 문제에 맞게 재학습시키는 방법이 일반화되고 있다. 이를 전이학습이라고 하는데, 전이학습에서 가장 유명한 데이터셋은 이미지넷 데이터셋[25]이다. 이미지넷 데이터셋은 특히 전체 이미지가 약 999,000개에 이르며 그 중 339,000의 이미지는 식물과 꽃과 관련이 있다. 따라서 전이학습을 채택하게 되면 최적화된 가중치와 바이어스를 학습하는데 드는 시간과 비용을 줄일 수 있다.

PyTorch와 Keras는 전이학습을 위해 미리 학습된 가중치와 편이를 포함한 각각의 유명한 합성곱 신경망 모델들을 제공한다. 본 논문의 실험에서는 VGGNet[16], Inception-ResNet[24], ResNet[18], Mobile Net[20], MobileNetV2[23] 모델을 사용하고, 이미지넷 데이터셋으로 전이학습 된 모델의 일부 층의 가중치와 편이를 고정시켜 새로 학습할 데이터에 영향을 받지 않도록 하는 기법은 사용하지 않는다.

5.3 실험 및 평가

실험의 형평성을 위해서 학습 과정의 모든 모델의 세팅은 다음과 같이 동일하게 구성하였다. 모든 학습 모델은 최대 100번의 학습 에폭(epoch)을 수행하였고, 배치 크기(Batch size)는 32로, 최종 테스트에 사용할 가중치는 학습 중 검증 단계에서 가장 좋은 정확도(Accuracy)를 보인 에폭의 가중치를 선정하였다. 모든 모델은 SGD(Stochastic Gradient Descent) 최적화 방법을 사용하였고, 학습률(Learning rate)은 0.01로 시작하여 일정 학습 기간 동안 10분의 1씩 감소시켰다.

표 3의 구조를 보면, 기준 모델이 되는 ResNet은 하나의 Residual Block이 인접해있는 2개의 블록으로 구성되어 있는데, 이를 자세히 살펴보면 다음과 같이, 컨벌루션 층 → 배치 정규화 층 → 활성화함수 ReLu → 컨벌루션 층 → 배치 정규화 층 으로 되어 있다[18]. 마지막 층은 전통적으로 합성곱 신경망에

서 완전 연결층을 사용하는데, ResNet은 글로벌 평균 풀링(Global average pooling)을 사용하여 파라미터 수를 최소화하고, 마지막 출력 층이 곧바로 소프트맥스(Softmax) 되어 정답 클래스의 확률을 계산할 수 있도록 층의 개수를 정하였다. 본 실험에서, ResNet 기준 모델은 그대로 사용하였고, 계층적 구조의 분류 모델에서는 계층별 클래스의 수에 맞게 구조를 변형하여 사용하였다.

표 2의 실험 결과를 보면, 기준 모델을 그대로 사용하여 21개 클래스를 분류한 결과는 96.2718%를 달성하였지만, 21,298,010개의 높은 수의 파라미터를 사용한다. 본 논문에서 제안하는 계층적 구조의 합성곱 신경망 모델에서는 과만 분류하는 첫 번째 계층에서 99.5364%이고, 첫 번째 계층에서 정답이라고 예측된 입력을 토대로 각각의 과를 분류하는 분류기를 통과하였을 때, 국화과 95.7839%, 명아주과 98.7778%, 비름과 96.1307%, 현삼과 99.0921%와, 메꽃과에 포함된 클래스는 하나라서 과만 분류하여 99.5364%를 달성하였다.

결과적으로 계층적 구조의 합성곱 신경망 모델에서는 평균적으로 95.6122%를 달성하였지만, 전체 파라미터 수는 10,489,950 사용하였다. 이는 ResNet 기준 모델과 비교하였을 때 약 2배 이상의 파라미터 수를 줄였으며, 정확도도 1%밖에 차이 나지 않았

다. 또한 그림 5와 같이 계층적 구조에서 잘못된 과를 예측했을 때를 보정하는 계층적 구조의 합성곱 신경망에서는 실험한 결과 정확도 95.6241%를 달성하여, 기존의 95.6122%보다 성능이 좋아짐을 확인하였다.

계층적 구조의 합성곱 신경망이 망을 구성하는 파라미터 수를 줄이면서도 높은 정확도를 유지할 수 있다는 점을 확인하기 위해 더 단순한 모델을 사용하여 추가 실험을 진행하였다.

표 4는 MobileNetV2를 사용하여 실험을 진행한 구조와 결과이다. MobileNetV2는 ResNet과 마찬가지로 인접해있는 2개의 블록으로 구성되어 있는데, 기본적으로 1x1 컨벌루션 층 → 배치 정규화 층 → 활성화함수 ReLu6 → 컨벌루션 층 → 배치 정규화 층으로 구성되어 있다[23].

마지막 층도 ResNet처럼 글로벌 평균 풀링을 사용하였다. MobileNetV2의 기준 모델이라는 열은 공개된 모델을 그대로 사용한 결과이며, 계층적 구조에서는 계층별 클래스 개수에 따라 구조를 변형하였다. MobileNetV2의 블록들을 추가하고 삭제해보면서 최적의 파라미터 수와 정확도를 가지도록 실험을 반복하였으며 최종 하이퍼 파라미터 모델이 표 3이다.

표 3. ResNet[18] 모델 기반의 실험 결과
Table 3. ResNet[18] model based experimental results

Residual Block	Base model $P(Y X)$	Family model $P(F X)$	국화 class model	명아주 class model	비름 class model	현삼 class model	Total
Architecture							
$\begin{matrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{matrix}$	×3	×2	×2	×3	×2	×2	×12
$\begin{matrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{matrix}$	×4	×2	×2	×4	×2	×2	×14
$\begin{matrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{matrix}$	×6	×2	×2	×0	×2	×0	×10
$\begin{matrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{matrix}$	×3	×0	×0	×0	×0	×0	×0
Average pooling	×1	×1					×1
Output Size	[512,1,1]	[256,4,4]	[256,4,4]	[128,4,4]	[256,4,4]	[128,4,4]	[128,4,4]
Pretrained weights	ResNet34	ResNet18	ResNet18	ResNet18	ResNet18	ResNet18	ResNet18
Total number of parameters							
	21,298,010	2,807,366	2,831,948	1,354,051	2,807,366	689,219	10,489,950
Test accuracy(%)							
	96.2718	99.5364	95.7839	98.7778	96.1307	99.0921	95.6122
Average Accuracy of reward hierarchical architecture(%)							95.6241

표 4. MobileNetV2[25] 모델 기반의 실험 결과
Table 4. MobileNetV2[20] model based experimental results

Inverted Residual Block	Base model $P(Y X)$	Family model $P(F X)$	국화 class model	명아주 class model	비름 class model	현삼 class model	total
Architecture							
$\begin{bmatrix} 1 \times 1, n \\ 3 \times 3, n \end{bmatrix}$	ResNet architecture	×15	×16	×14	×16	×14	×75
Output Size	[512,1,1]	[160,4,4]	[160,4,4]	[160,4,4]	[160,4,4]	[160,4,4]	[160,4,4]
Total Parameters							
	21,298,010	1,084,378	1,368,524	705,475	1,033,158	705,475	4,897,010
Accuracy on test set(%)							
	96.2718	99.5483	97.7796	98.3333	95.9302	99.3515	97.2349
Average Accuracy of reward hierarchical architecture(%)							97.2612

표 5. LateFusion[13] 구조와 계층적 구조 분류 모델의 비교
Table 5. Comparison with LateFusion[13] hierarchical architecture based models

Classification model	LateFusion[13] architecture			Hierarchical architecture			
	Weights combination	Geometric mean	Product Combination	ResNet based		MobileNetV.2 based	
				Normal model	Reward model	Normal model	Reward model
the number of classes	16	16	16	21	21	21	21
the number of parameters	230,387,728			10,489,950		4,897,010	
accuracy(%)	95.892	95.927	95.362	95.6122	95.6241	97.2349	97.2612

실험 결과, 계층적 구조의 분류 모델은 평균 97.2349%의 정확도와 4,897,010의 파라미터 수를 가졌으며, 결과를 보정할 수 있는 계층적 구조의 합성곱 신경망에서는 정확도 97.2612%를 달성하여, 기존의 97.2349%보다 성능이 좋아짐을 확인하였다. 이를 통해 ResNet기반의 기준 모델과 비교하였을 때 약 4배 이상의 파라미터 수 감소를 보였으며, 정확도는 상승한 것을 볼 수 있다.

계층적 구조의 합성곱 신경망이 최대의 파라미터 수를 가지는 딥러닝 앙상블 구조인 LateFusion[13]과 비교를 하였으며, 결과는 표 5와 같다. LateFusion 방법은 직관적으로 전문가 여러 명이 의견을 종합하여 결론을 도출하는 것처럼 다양한 딥러닝 모델의 소프트맥스된 결과를 산술평균, 기하평균 및 Product Combination으로 계산하여 정답을 예측하는 것이다. LateFusion 방법에는 VGG16[16], MobileNet[20], Inception-ResNet[24]이 사용되었으며, 외래잡초 16개 클래스를 분류하였다.

LateFusion이 사용한 전체 파라미터 수는

230,387,728이며, 가장 높은 정확도는 95.892%다. 하지만 ResNet 기반의 계층적 구조의 모델은 LateFusion 모델보다 약 23배를 줄인 10,489,950의 파라미터를 사용하고도 95.6122%의 정확도를 보였으며, 보강모델을 사용하였을 때는 95.6241%로 정확도 상승을 얻었다. 더 적은 수의 파라미터를 사용하는 MobileNetV2 기반의 모델은 각각 약 47배를 줄였으며, 97.2349%의 높은 정확도를 보였고, 보강모델에서는 97.2612%의 가장 높은 정확도를 얻었다.

VI. 결론 및 향후연구

최근에 딥러닝을 이용한 분류 예측 모델은 다양한 분야에서 좋은 성능을 보인다. 하지만 실제 응용 분야에서는 제한된 여건으로 인한 하드웨어 성능 등으로 깊고 복잡하며 많은 수의 파라미터를 가진 좋은 딥러닝 모델들을 사용하기가 어렵다. 이를 위해 적은 수의 파라미터를 사용하면서도 높은 정확도를 달성할 수 있는 딥러닝 모델들이 대두되고

있다.

계층적 구조를 사용하는 합성곱 신경망 모델은 분할-정복 알고리즘을 아이디어로 하여 정답을 정확히 예측하는데 있어 최소의 파라미터를 사용하는 효율적인 딥러닝 구조를 제안한다. 제안된 계층적 구조의 합성곱 신경망은 자체 외래잡초 21개 클래스를 이용하여 최소의 파라미터를 가지면서도 높은 정확도를 얻을 수 있도록 하였다.

실험 결과, 전이학습 기반의 ResNet 단일 모델과 비교하여 파라미터 수를 최소 2배 이상에서 최대 4배 이상까지 줄였으며, LateFusion이라는 최대 파라미터를 사용하여 정확도를 올린 구조와 비교하였을 때, 최대 47배의 파라미터 수 감소를 확인하였다. 또한, 정확도 부분에서 ResNet 기반 모델의 계층적 구조는 95.6122%, MobileNetV2 기반에서는 97.2349%를 달성하였으며, 보강 모델까지 사용하였을 때 MobileNetV2 기반에서 최대 97.2612%로 가장 높았다.

제안된 구조에서 계층적 구조의 합성곱 신경망에 사용되는 기반 모델을 직접 설계하여 각각의 층을 효과적으로 제어할 수 있다면, 더 적은 파라미터 수로 분류 정확도는 현재와 같이 유지하면서 속도는 빨라질 것으로 기대된다.

본 연구를 기반으로 제안된 구조를 종단 간 학습(End-to-end learning)을 통해 하나의 과 분류기와 클래스 분류기를 서로 연결하여 더 좋은 성능을 낼 수 있도록 하며, 학습 데이터의 라벨링 작업을 다양화하여 영역 분할 방법[26] 또는 바운딩 박스(Bounding box)를 활용한 방법[27] 등을 적용하여 객체 분류의 정확도를 높일 수 있는 연구를 진행하도록 하겠다. 이와 같은 결과와 향후연구를 통해 임베디드시스템과 결합된 잡초 분류 분야에서 좋은 성과가 있을 것으로 본다.

References

- [1] T. W. Berge, A. H. Aastveit, and H. Fykse, "Evaluation of an algorithm for automatic detection of broad-leaved weeds in spring cereals", *Precis. Agric.*, Vol. 9, pp. 391-405. Sep. 2008.
- [2] Hamuda, Esmael, Martin Glavin, and Edward Jones, "A survey of image processing techniques for plant extraction and segmentation in the field" *Computers and Electronics in Agriculture*, Vol. 125, pp. 184-199, Jul. 2016
- [3] M. Saber, W. S. Lee, T. F. Burks, J. K. Schueller, C. A. Chase, G. E. MacDonald, and G. A. Salvador, "Performance and Evaluation of Intra-Row Weeder Ultrasonic Plant Detection System and Pinch-Roller Weeding Mechanism for Vegetable Crops", *Proceedings of International Meeting. American Society of Agricultural and Biological Engineers*, pp. 1, 2015
- [4] K. Thorp and L. Tian., "A review on remote sensing of weeds in agriculture", *Precision Agriculture*, Vol. 5, No. 5, pp. 477-508, Oct. 2004.
- [5] S. Christensen, H. T. Sogaard, P. Kudsk, M. Nørremark, I. Lund, E. Nadimi, and R. Jørgensen, "Site-specific weed control technologies", *Weed Research*, Vol. 49, No. 3, pp. 233-241, May 2009.
- [6] J. Hemming and T. Rath, "Computer-vision-based weed identification under field conditions using controlled lighting", *Journal of Agricultural Engineering Research*, Vol. 78, No. 3, pp. 233-243, Mar. 2001.
- [7] T. Beghin, J. S. Cope, P. Remagnino, and S. Barman, "Shape and texture based plant leaf classification", *Advanced Concepts for Intelligent Vision Systems*, Springer, pp. 345-353, Dec. 2010.
- [8] D. M. Tax, M. Van Breukelen, R. P. Duin, and J. Kittler, "Combining multiple classifiers by averaging or by multiplying?", *Pattern recognition*, Vol. 33, No. 9, pp. 1475-1485, Sep. 2000.
- [9] M. Persson and B. Åstrand, "Classification of crops and weeds extracted by active shape models", *Biosystems Engineering*, Vol. 100, No. 4, pp. 484-497, Aug. 2008.
- [10] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification", In

- Computer Vision-ECCV, Springer, pp. 502-516, Oct. 2012.
- [11] S. Haug, A. Michaels, P. Biber, and J. Ostermann, "Plant classification system for crop/weed discrimination without segmentation", Proceedings of IEEE Winter Conference on Applications of Computer Vision, pp. 1142-1149, Mar. 2014.
- [12] M. Dyrmann, H. Karstoft and H. S. Midtby, "Plant species classification using deep convolutional neural network", Biosystems Engineering, Vol. 151, pp. 72-80, Nov. 2016.
- [13] Vo Hoang Trong, Gwang-Hyun Yu, Nazeer Shahid, Seong-Min Hwang, and Jin-Young Kim, "Classification of weeds using Multimodal Convolutional Neural Networks", Proceedings of IEIE, pp. 368-371. Nov. 2018.
- [14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.
- [15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105, Dec. 2012.
- [16] SIMONYAN, Karen and ZISSERMAN, Andrew, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, Sep. 2014.
- [17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, ... and A. Rabinovich, "Going deeper with convolutions", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, Jun. 2015.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, Jun. 2016
- [19] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 8697-8710, Jun. 2018.
- [20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704.04861, Apr. 2017.
- [21] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size", arXiv preprint arXiv:1602.07360, Nov. 2016.
- [22] Sabour Sara, Frosst Nicholas, and Hinton Geoffrey E., "Dynamic routing between capsules", Advances in neural information processing systems, pp. 3856-3866, Dec. 2017.
- [23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, Jun. 2018.
- [24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning", arXiv:1602.07261v2, Aug. 2016.
- [25] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, Jun. 2009.
- [26] He-Jin Yu and Chang-Hwan Son, "Recognizing Apple Leaf Diseases via Segmentation-Aware Deep Convolutional Neural Networks for Smart Farm", Journal of KIIT, Vol. 17, No. 6, pp. 73-83, Jun. 2019.
- [27] Jinso Kim and Jeongho Cho, "YOLO-based Real-Time Object Detection Scheme Combining RGB Image with LiDAR Point Cloud", Journal of KIIT, Vol. 17, No. 8, pp. 95-105, Aug. 2019.

Authors

유 광 현 (Gwanghyun Yu)



2018년 2월 : 전남대학교
전자컴퓨터공학과(공학석사)
2018년 3월 ~ 현재 : 전남대학교
전자컴퓨터공학과(박사과정)
2018년 12월 ~ 현재 : 인섹피디아
회사 창업
관심분야 : 디지털 신호처리, 영상
처리, 머신러닝, 딥러닝

이 재 원 (Jaewon Lee)



2012년 2월: 전남대학교
전자컴퓨터공학과(공학석사)
2017년 8월: 전남대학교
전자컴퓨터공학과(공학박사)
2017년 9월 ~ 현재 : 전남대학교
전자컴퓨터공학과(박사후연구원)
관심분야 : 디지털 영상처리,
컴퓨터비전, 딥러닝

보 호양 트롱 (Vo Hoang Trong)



2017년 9월 : 베트남 호치민
과학대학교 수학, 컴퓨터과학과
(공학사)
2018년 3월 ~ 현재 : 전남대학교
전자컴퓨터공학과(석사과정)
관심분야 : 디지털 신호처리, 영상
처리, 음성 신호처리, 머신러닝

당 탄 부 (Dang Thanh Vu)



2018년 9월 : 베트남 호치민
과학대학교 수학, 컴퓨터과학과
(공학사)
2019년 3월 ~ 현재 : 전남대학교
전자컴퓨터공학과
(석박사통합과정)
관심분야 : 디지털 신호처리, 영상
처리, 음성 신호처리, 머신 러닝

후이 트완 녀 (Huy Toan Nguyen)



2012년 9월 : 베트남 타이넥 기술
대학교 전자공학과(공학사)
2015년 3월 ~ 현재 : 전남대학교
전자컴퓨터공학과
(석박사통합과정)
관심분야 : 컴퓨터 비전, 웨어러블
장치, 시스템 기반의
마이크로프로세서, 머신러닝, 딥러닝

이 주 환 (JooHwan Lee)



2017년 11월 : 더뉴어 회사 창업
2019년 8월 : 전남대학교
지구환경과학부(이학사)
2019년 9월 ~ 현재 : 전남대학교
전자컴퓨터공학과(석사과정)
관심분야 : 디지털 신호처리, 영상
처리, 음성 신호처리, 머신러닝

신 도 성 (Dosung Shin)



1993년 2월 : 동신대학교
정보공학과(공학사)
1998년 2월 : 전남대학교
전자공학과(공학석사)
2017년 2월 : 전남대학교
전자정보공학과(공학박사)

2019년 8월 ~ 현재 : 전남대학교
전자컴퓨터공학과(박사후연구원)
관심분야 : 디지털 신호처리, 영상 처리, 바이오메트릭스

김 진 영 (Jinyoung Kim)



1986년 2월 : 서울대학교
전자공학과(공학사)
1988년 2월 : 서울대학교
전자공학과(공학석사)
1994년 8월 : 서울대학교
전자공학과(공학박사)
1995년 3월 ~ 현재 : 전남대학교

전자컴퓨터공학과 교수
관심 분야 : 디지털 신호처리, 영상 처리, 음성
신호처리, 머신러닝, 딥러닝