

딥러닝을 이용한 음식 이미지 분류 기술 개발

Food Image Classification using Deep Learning

이가경¹ · 임세연¹ · 양진이² · 유민정² · 김선옥^{2*}

한국항공대학교 소프트웨어학과¹, 한국항공대학교 인공지능학과²

요약

본 연구는 20대와 한국인을 대상으로 한 건강관리 애플리케이션의 음식 이미지 분류 모델을 개선하는 것을 목표로 진행되었다. AI Hub에서 546,194개의 이미지를 수집하여 175개의 음식 클래스를 구성하였으며, ResNet 인공지능 모델을 학습하고 검증하였다. 추가적으로, 실제 촬영한 음식 이미지에 대한 인식 정확도가 상대적으로 낮게 나타나는 원인에 대해 고찰하고, 이를 해결하기 위한 방안으로 모델 성능을 최적화를 위한 다양한 방법을 분석하였다.

■ 중심어 : 딥러닝, ResNet, 음식 이미지 분류

Abstract

This study was conducted with the aim of improving the food image classification model of a health care application targeting Koreans in their twenties. 546,194 images were collected from the Public Data Portal and AI Hub, and 175 food classes were constructed. The ResNet artificial intelligence model was trained and validated. Additionally, we deeply investigated the reasons for the relatively lower recognition accuracy of the actual food images, and we attempted various methods to optimize the model's performance as a solution.

■ Keyword : Deep Learning, ResNet, Food Image Classification

2023년 11월 24일 접수; 2023년 12월 11일 수정본 접수; 2023년 12월 14일 게재 확정.

* 이 출판물은 2021년도 한국항공대학교 교비지원 연구비에 의하여 지원된 연구의 결과임.

이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 “건강관리를 위한 음식 이미지”, “음식 이미지 및 영양 정보 텍스트”를 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에서 다운로드 받으실 수 있습니다.

† 교신저자 (sunok.kim@kau.ac.kr)

I. 서론

2020년 이후로 전 세계적으로 유행한 코로나 팬데믹은 우리 사회의 지대한 영향을 미침과 동시에 건강관리에도 새로운 문제를 제기하였다. 특히 코로나 팬데믹으로 인한 사회적 거리 두기와 외출 제한 등의 영향으로 실내 활동 시간이 증가하면서 부적절한 식습관과 영양 관리는 건강 문제를 악화의 주요 원인으로 작용하였다. 코로나19로 인한 일부 대학생의 식생활 변화 연구[1] 결과에 따르면, 코로나-19로 인해 비대면 수업이 진행에 따라 습관에 미친 부정적인 변화로는 불규칙한 식사, 배달 음식 섭취 증가, 간식 섭취 증가 등이 나타난 것을 알 수 있다. 이러한 식습관의 변화는 불균형한 영양 섭취를 초래할 뿐만 아니라 체중 변화와 함께 다양한 건강 문제의 원인이 될 수 있다. 체중 관리를 위한 균형 있는 영양 섭취를 위해서는 식단 구성에서 잘못된 점을 분석하고 섭취한 음식의 종류와 양을 정확하게 기록하는 것이 중요하다. 그러나 실제로 끼니마다 섭취한 음식을 꾸준히 기록하는 것은 쉽지 않다. 이런 문제를 해결하기 위해, 최근에는 딥러닝 기술을 이용한 음식 자동 분류 및 영양성분 분석에 대한 기술 연구가 진행되고 있다. 본 논문에서는 딥러닝 기반의 이미지 분류 모델에 대해 소개하고, 한국인 다빈도 섭취 외식과 한식 메뉴에 대한 음식 이미지 분류 모델을 제작하여 자동으로 영양 성분 정보를 제공하는 방법에 대해 제시하고자 한다.

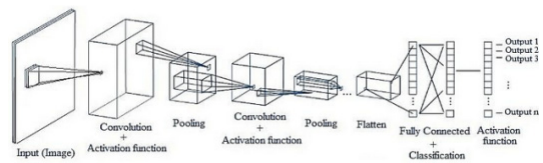
II. 배경 이론

2.1 Convolutional Neural Networks

CNN(Convolutional Neural Networks)는 딥러닝에서 합성곱 신경망을 사용하여 이미지나 영상 데이터를 처리하는 딥러닝 구조이다[3-5,7-9].

DNN(Deep Neural Network)의 경우 2차원 데이터를 1차원으로 변환하며 이미지의 공간적/지역적 정보의 손실이 발생한다는 문제점이 존재한다. 반면 CNN의 경우 초기 입력의 정보를 변형하지 않고 합성곱 신경망을 이용하여 이미지의 부분을 인식하고, 주변 픽셀과의 연관성을 계산한다.

합성곱 신경망의 경우 여러 개의 필터를 사용하며, 각 필터는 입력값을 조각으로 구분한다. 필터의 결과값들에 활성화 함수를 적용하여 하나의 레이어를 완성한다. 풀링 레이어는 풀링 크기에 따라 입력값의 차원을 축소한다. 이와 같은 합성곱 신경망과 풀링 레이어 적용을 반복하며 이미지를 변형하고, 평탄화 작업을 통해 텐서 형태의 값을 벡터 형태로 변환한다. 이후 Softmax를 이용하여 분류 작업을 수행한다.

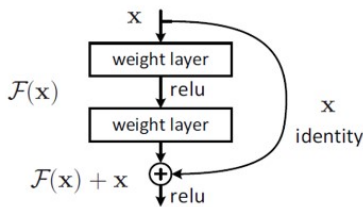


〈그림 1〉 CNN을 이용한 분류 네트워크 예시

2.2 ResNet

ResNet(Residual Network)는 이미지 분류를 위한 딥러닝 아키텍처로, DNN에서 neural network의 깊이가 깊어질수록 발생하는 성능 저하 문제를 해결하기 위해 제안되었다.[2] ResNet은 이름 그대로 잔차(Residual)이라는 개념을 사용하는 데, 이는 기존 네트워크에서의 입력값과 결과값의 차이를 의미한다. 하지만 ResNet은 레이어의 입력값을 레이어 출력값에 더하여 결과값을 입력과 출력의 차이로 지정하며, 해당 값을 0으로 만드는 것을 목표로 학습을 진행한다. 이러한 연산 목표의 변화는 불분명한 목표값을 향해 학습하던 기존 네트워크와 비교하여 입력과 출력이

같은 값이 되도록 하는 목표가 존재한다는 차이점이 존재하기에 학습이 용이해진다는 장점이 있다. 또한 네트워크의 구조 역시 입력값을 출력값에 연결하는 shortcut 추가 외의 큰 변화가 없기에 파라미터의 수와 연산량 급증의 문제 또한 존재하지 않는다. 오히려 shortcut으로 인해 일부 레이어를 생략함으로써 기울기 소실 문제를 해결하였다.



〈그림 2〉 Residual learning 구조

III. 데이터 수집 및 학습 모델 선정

3.1 음식 이미지 데이터셋 구성

본 연구는 AI Hub에서 제공하는 음식 이미지 데이터 2종류를 활용하여 특화된 데이터셋을 구축하였다. 데이터셋 구성 과정에서는 대상 연령대와 국적을 고려하여 한국인 또는 20대가 주로 섭취하는 음식에 집중하였다. 반찬과 같이 여러 음식이 함께 제공되는 경우가 잦은 한식의 특성상, 단일 이미지 인식률을 높이기 위해 반찬 및 음료 등의 클래스를 제거함과 동시에 유사 클래스의 통합을 진행하였다. 이를 통해, ‘건강관리를 위한 음식 이미지’ 데이터셋에서 제공하는 500여 종의 클래스 중 175종을 선별하였고, 대상 집단의 식습관에 부합하는 데이터를 추가하기 위해 ‘음식 이미지 및 영양 정보 텍스트’ 데이터셋에서 20종의 클래스를 추가하여 데이터셋을 구성하였다. 오분류 및 과적합을 방지하기 위해 각 클래스당 이미지 개수를 최소 900개, 최대 5,000개로 조정하여 클래스 간 학습 이미지 개수

의 편차를 줄이도록 수정하였다. 최종 데이터셋은 175개의 클래스로 이루어져 있으며, 학습과 검증을 위해 약 8:2의 비율로 분리하였다.

〈표 1〉 ‘건강관리를 위한 음식 이미지’ 데이터셋 상세 내역

데이터 종류	구축 수량	포함 내용	제공 방식
음식이미지 데이터베이스	500여 종 *60,000장 =30,000,000장	원본 이미지, JSON파일	파일 다운로드

〈표 2〉 ‘음식 이미지 및 영양정보 텍스트’ 데이터셋 상세 내역

데이터 종류	구축 수량	포함 내용	제공 방식
음식이미지 데이터베이스	400종 *2,000장 =800,000장	이미지 포맷, txt, xml	파일 다운로드

〈표 3〉 최종 데이터셋 상세 내역

구분	클래스 개수	포함 내용	제공 방식
학습	175개	546,194	255.8GB
검증		78,397	35.2GB

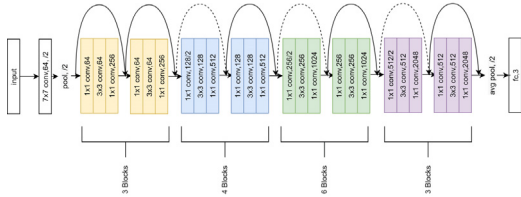
3.2 모델 선택 과정 및 소개

음식 이미지 분류는 다양한 범주와 세밀한 차이를 구분해내야 하는 매우 복잡한 작업으로, 수많은 패턴과 다양한 특징을 학습할 수 있는 깊은 네트워크가 필요하다. 여러 CNN 기반의 모델들이 있지만, Resnet이 이미지 분류 분야에서 많이 사용되고 있다[10-13,18,19]. [14], [15]와 같은 논문이 객체 검출 분야에서 많이 사용되고 있지만, 이미지 분류를 위해 고안된 모델이 아니기 때문에, 레이어 깊이를 늘릴 수 있는 ResNet이 본 논문에서 제안하는 방법에 적합한 구조였다. 따라서 네트워크 깊이에 따른 성능 저하를 방지하고, 빠른 속도와 분류 정확도를 위해 ResNet 구조를 선택하였다.

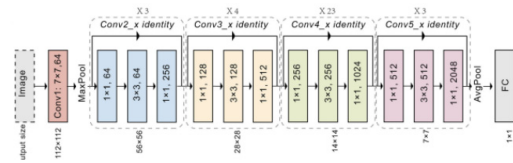
ResNet 구조를 활용한 모델 선택을 위해 50개의 레이어와 101개의 레이어를 가진 두 개의 모델을 선택하여 비교를 진행하였다. ResNet-50의 경우 3개의 레이어로 이루어진 Residual Block 16개를 포함한 구조이며, ResNet-101은 앞선 구조에서 17개의 Residual Block이 추가되어 총 33개의 Block을 포함한다.

레이어 개수 증가에 따른 성능 저하 여부를 확인하기 위해 최종 데이터 셋 구성 이전, 379개의 클래스로 이루어진 음식 이미지 데이터셋을 각

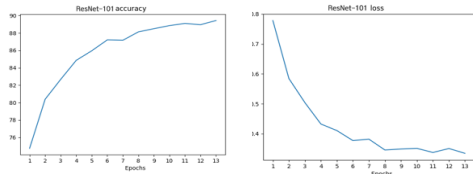
모델에 적용하여 학습을 진행한 후 결과를 확인하였다. ResNet-50, ResNet-101 모두 총 13회 반복 학습을 진행하여 해당 회차의 학습 결과를 분석하였다. ResNet-50의 경우 12회차 학습에서 최대 정확도 89.0017%, 최소 손실값 0.3428을 달성하며 수렴하였고, ResNet-101의 경우 13회차 학습에서 최대정확도 89.4268%, 최소 손실값 0.3345를 달성하였다. 해당 결과에 따라 ResNet-101 구조에 대해 학습 정확도 향상 가능성이 존재한다고 판단하여 최종 데이터셋 구성 이후 ResNet-101 모델을 채택하여 학습을 진행하였다.



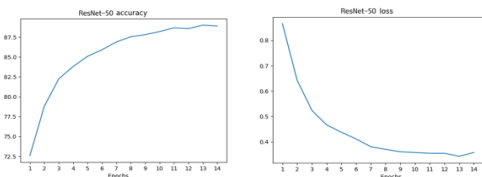
〈그림 3〉 ResNet-50 구조



〈그림 4〉 ResNet-101 구조



〈그림 5〉 성능 확인을 위한 ResNet-50 학습 결과



〈그림 6〉 성능 확인을 위한 ResNet-101 학습 결과

IV. 음식 이미지 분류 모델 학습

4.1 학습 환경 및 초기 설정

이 장에서는 ResNet-101 구조를 이용한 모델 학습에 대한 실험을 진행한다. 실험 환경은 표 4와 같이 진행하였으며, 학습 진행 시 CPU 개입 없이 GPU만을 사용하여 학습 환경을 구성하였다. 모델의 경우 학습을 진행하기 위해 입력 데이터의 크기를 (224,224)로 변환하였으며, 표 5와 같이 인자값을 지정하여 학습을 진행하였다.

모델 학습 최적화를 위한 optimizer로는 SGD

〈표 4〉 모델 학습 환경

장치	환경
운영체제	Ubuntu 20.04
통합 개발 환경	Visual Studio Code, Jupyter Notebook
언어	Python
GPU	NVIDIA RTX 3060

〈표 5〉 인자의 지정값

인자	지정값
학습률	0.01
가중치	0.0001
모멘텀	0.9
배치 크기	16

(Stochastic Gradient Descent) 알고리즘을 선정하였으며, 손실 함수로는 Cross-Entropy Loss를 사용하였다.

4.2 실험 결과 및 해석

이 장에서는 학습된 모델의 검증결과를 통해 모델의 학습 정확도와 손실값을 분석하고 수렴 여부를 확인하며, 해당 모델의 유효성을 분석한다. 그림 7은 ResNet-101구조를 이용한 모델 학습이 진행됨에 따라 점차적으로 성능이 향상되고 있다는 것을 보여준다. 학습이 8회차에서 14회차로 진행되면서 모델의 정확도는 91.94%에서 93.21%로 증가하였으며, 손실값은 0.27에서 0.26으로 감소하였다. 그러나 학습이 11회차에서 12회차로 넘어갈 때, 정확도는 소폭 상승함과 동시에 손실값 또한 약간 증가하는 추세를 보였다. 이는 모델이 일부 데이터에 과적합되고 있을 가능성을 나타내므로 11회 학습 모델이 가장 최적화된 상태라고 판단하여 해당 모델을 최종 학습 모델로 선택하였다.

해당 모델의 유효성을 검증하기 위해 일부 음식 이미지를 이용한 테스트를 진행하였다. 클래스 라벨에 대한 정보를 추출하였고, 이미지 변형과 파라미터 등은 모델 학습과 동일하게 설정하였으며, pytorch[6]의 topk 기능을 이용하여 인식

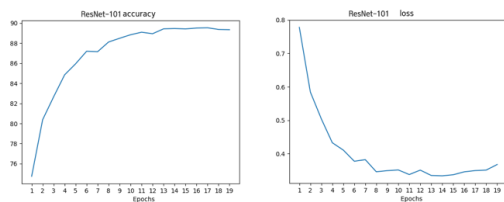
된 이미지에 대한 상위 5개의 예측값과 예측 확률을 확인하였다.

그림 8은 ResNet-101로 학습된 모델의 이미지 테스트 결과를 나타낸다. 단일 음식 이미지를 인식하여 예측값을 나타낸 결과 테스트를 진행한 모든 음식에서 정답 라벨을 가장 높은 확률로 예측했다. 특히 생김새가 유사한 품목이 많은 쌀국수의 경우에도 높은 예측 확률로 분류를 진행한 것을 보아 학습이 성공적으로 진행된 것을 확인할 수 있다.

하지만 모든 이미지에서 예측이 성공한 것은 아니다. 그림 9는 예측이 실패하거나 예측 확률이 낮은 경우를 나타내었다. 훈제 오리를 예측한 경우, 훈제오리뿐만이 아닌 밥, 야채 등이 포함된 사진으로, 상위 5개의 예측값 중에 샐러드와 밥이 포함되어있는 것을 확인할 수 있다. 또한 김밥의 경우 일반적인 김밥과 생김새가 다른 이미지를 인식한 결과 예측 확률이 낮아진 것을 확인

오므라이스		양념치킨		돈가스		탕수육		쌀국수	
									
예측값	예측 확률	예측값	예측 확률	예측값	예측 확률	예측값	예측 확률	예측값	예측 확률
오므라이스	38.3925	양념치킨	25.1109	돈가스	23.0688	탕수육	28.9888	쌀국수	38.1842
양념치킨	11.3514	계정	14.6572	참고김밥	13.2115	오징어냉면	15.0783	우동	24.7890
그라탕	11.2038	후라이드치킨	13.6414	닭다리	12.8368	돈가스	14.4390	국수	24.4687
돈가스	10.2201	푸마리통닭	13.2849	밀밥	11.3257	소금탕수육	13.2631	라면	23.6796
크레페	0.9101	케인	12.0219	가주동	10.9447	야채볶음	13.2500	된장면	22.4968
실행시간	0.0847	실행시간	0.0677	실행시간	0.0717	실행시간	0.1666	실행시간	0.0679

〈그림 8〉 이미지 테스트 성공 사례



〈그림 7〉 최종 모델 학습 결과

〈표 6〉 모델 학습 정확도 및 손실값

학습 회차	10	11	12	13
정확도	91.9408	92.9942	92.9469	92.2167
손실값	0.2675	0.2574	0.2616	0.262

훈제오리		김밥	
			
예측값	예측 확률	예측값	예측 확률
샐러드	11.9813	계장	23.0996
보쌈	10.7260	김밥	14.4595
죽밥	10.0040	전복	14.3378
밥	9.7165	주먹밥	14.3282
카레라이스	9.5699	쇠고기구이	12.7907
실행시간	0.0346	실행시간	0.0319

〈그림 9〉 이미지 테스트 실패 사례

할 수 있다. 이러한 테스트 결과를 통해 단일 음식 이미지에서는 뛰어난 예측 결과를 보여주지만, 다른 음식이 포함되어 있거나 생김새의 편차에 따라 예측확률이 낮아지는 것을 확인할 수 있다. 이러한 한계점은 다양한 음식을 한 상에 올려먹거나 동일 음식 내의 생김새의 편차가 큰 한식의 특성에 따라 치명적 단점이 될 수 있다.

V. 연구의 한계점 및 향후 연구 방향 제시

본 연구는 음식 이미지 분류 모델의 개발을 통해 20대와 한국인을 대상으로 하는 음식 이미지 인식 모델의 정확도를 높이는 것을 목표로 하였다. 그러나, 연구 과정에서 다음과 같은 몇 가지 한계점이 드러났다.

첫째, 테스트 데이터 셋에 대한 정확도는 높게 나타났지만, 실제 촬영한 음식 이미지에 대한 정확도는 상대적으로 떨어졌다. 이는 실제 사용자가 찍는 사진 데이터(조명, 배경, 각도 등)와 학습 데이터의 특성이 상이하다는 점에서 기인하는 것으로 추측된다. 또한 한식의 특성상 음식 촬영 시 메인 음식을 포함하여 다양한 반찬 혹은 곁들여 먹는 음식이 함께 나타날 경우가 많다. 이러한 음식별 객체 탐지 기능의 부재로 인해 타 음식에 대한 정보가 뒤섞이며 정확도가 낮아졌을 가능성이 존재한다.

둘째, 본 연구의 목표는 앱 개발이 아닌 기술 및 성능 개발에 있어서, 실제 활용 단계까지의 업그레이드가 필요하다. 이를 위해서는 고성능의 영양성분 분석을 통한 실제 활용 가능성 향상이 중요하다.

따라서, 향후 연구에서는 다음의 세 가지 방향을 제안한다. 첫째, 실제 사용자가 찍는 사진 데이터의 특성을 반영한 데이터 확보에 집중하여, 모델의 정확도를 더욱 향상시키는 것이다. 둘째, 다중 이미지 인식 기술을 활용하여 객체 탐지 문제를 해결하고, 이를 위한 대량의 학습 데이터

확보에 초점을 맞추는 것이다.

셋째, 영양 성분 분석 데이터와의 결합을 통한 이미지 분석의 세분화가 진행되어야 한다. 이미지 분류와 함께 고성능의 영양 성분 분석을 적용하여 활용 가능성을 향상시켜 연구의 발전을 이룰 수 있다.

이러한 방향성을 따라 한식과 외식을 포함한 글로벌 음식에 대한 데이터를 통해 본 연구의 결과를 더욱 발전시킬 수 있을 것이라 보인다. 발전된 음식 이미지 인식 연구와 영양 성분 분석 연구를 통해 실제 음식 이미지 인식 애플리케이션에서의 높은 활용 가능성을 토대로 전 세계 연구의 건강 증진에 기여할 것을 기대한다.

참 고 문 헌

- [1] 김미현, 연지영, “충청지역 일부 대학생의 코로나-19 이후 식생활 변화, 가정간편식과 배달음식 이용 실태,” *Journal of Nutrition and Health*, vol. 54, no. 4, pp. 383-397, 2021.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. “Imagenet classification with deep convolutional neural networks.” *Advances in neural information processing systems* 25 (2012).
- [4] Sultana, Farhana, Abu Sufian, and Paramartha Dutta. “Advancements in image classification using convolutional neural network.” *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, 2018.
- [5] Rawat, Waseem, and Zenghui Wang. “Deep convolutional neural networks for image classification:

- A comprehensive review.” Neural computation 29.9 (2017): 2352-2449.
- [6] Paszke, Adam, et al. “Pytorch: An imperative style, high-performance deep learning library.” Advances in neural information processing systems 32 (2019).
- [7] Szegedy, Christian, et al. “Going deeper with convolutions.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [8] Simonyan, Karen, and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).
- [9] Huang, Gao, et al. “Densely connected convolutional networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [10] Xie, Saining, et al. “Aggregated residual transformations for deep neural networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [11] Liang, Jiazhi. “Image classification based on RESNET.” Journal of Physics: Conference Series. Vol. 1634. No. 1. IOP Publishing, 2020.
- [12] Ebrahimi, Mohammad Sadegh, and Hossein Karkeh Abadi. “Study of residual networks for image recognition.” Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 2. Springer International Publishing, 2021.
- [13] Zhang, Qi. “A novel ResNet101 model based on dense dilated convolution for image classification.” SN Applied Sciences 4 (2022): 1-13.
- [14] Redmon, Joseph, et al. “You only look once: Unified, real-time object detection.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [15] Girshick, Ross. “Fast r-cnn.” Proceedings of the IEEE international conference on computer vision. 2015.
- [16] Jung, Heechul, et al. “ResNet-based vehicle classification and localization in traffic surveillance systems.” Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.
- [17] Song, Xingguo, Kewei Chen, and Zhongqing Cao. “ResNet-based image classification of railway shelling defect.” 2020 39th Chinese Control Conference (CCC). IEEE, 2020.
- [18] Yahya, Ali Abdullah, et al. “A Novel Image Classification Method Based on Residual Network, Inception, and Proposed Activation Function.” Sensors 23.6 (2023): 2976.

저 자 소 개



이 가 경(Gagyong Lee)

· 2019년 2월~현재: 한국항공대학교 소프트웨어학과 (학사과정)
 <관심분야> 컴퓨터 비전, 인공지능



임 세 연(Seyeon Im)

· 2019년 2월~현재: 한국항공대학교 소프트웨어학과 (학사과정)
 <관심분야> 컴퓨터 비전, 인공지능



양 진 이(Jini Yang)

- 2023년 2월: 한국항공대학교
소프트웨어학과 (공학사)
 - 2023년 3월~현재: 한국항공대
학교 인공지능학과 (석사과정)
- <관심분야> 컴퓨터 비전, 인공
지능



유 민 정(Minjung Yoo)

- 2023년 2월: 한국항공대학교
소프트웨어학과 (공학사)
 - 2023년 3월~현재: 한국항공대
학교 인공지능학과 (석사과정)
- <관심분야> 컴퓨터 비전, 자율
주행



김 선 옥(Sunok Kim)

- 2014년 2월: 연세대학교 전기
전자공학과 (공학사)
- 2019년 2월: 연세대학교 전기
전자공학과 (공학박사)
- 2021년 3월~현재: 한국항공대
학교 인공지능학과 조교수

<관심분야> 준지도 학습, 컴퓨터 비전, 분류, 인공
지능