

## 디스플레이형 자판기 사용자 분석을 위한 이중 단계 검출 및 분류 망

서지원<sup>1</sup> · 김미경<sup>2\*</sup>

### 2-Stage Detection and Classification Network for Kiosk User Analysis

Ji-Won Seo<sup>1</sup> · Mi-Kyung Kim<sup>2\*</sup>

<sup>1</sup>Graduate Student, Department of Information Convergence Engineering, Pusan National University, Busan, 46241 Korea

<sup>2\*</sup>Professor, Software Education Center, Pusan National University, Busan, 46241 Korea

#### 요약

시각 정보를 이용한 기계 학습 기술은 주변 상황 인지, 결함 감지, 보안 그리고 사용자 분석과 같이 산업, 서비스 분야에서 활용성이 높아졌다. 그 중 CCTV 영상 분석을 통한 사용자 분석은 시각 정보를 잘 활용하는 실용적인 부분이라고 할 수 있다. 또한 이러한 임베디드 환경에서의 실용성을 높이기 위한 신경 회로망 경량화에 대한 연구가 지속되고 있다. 본 논문에서는 디스플레이형 자판기인 키오스크에서 활용할 수 있는 사람 및 얼굴 검출과 사용자의 나이 및 성별 분류 시스템을 제안한다. 제안하는 모델은 MobileNet, YOLOv2, 생략 연결을 기반으로 설계되었으며, 검출과 분류 망을 개별적으로 학습한 뒤 결합한 2-stage 구조를 띤다. 또한 주의 집중 기법을 사용하여 시스템의 성능을 향상 시키고자 하였다. 제안하는 시스템에 대한 구동과 성능 평가는 소형 그래픽 처리 유닛인 Nvidia Jetson Nano에서 진행하였다.

#### ABSTRACT

Machine learning techniques using visual data have high usability in fields of industry and service such as scene recognition, fault detection, security and user analysis. Among these, user analysis through the videos from CCTV is one of the practical way of using vision data. Also, many studies about lightweight artificial neural network have been published to increase high usability for mobile and embedded environment so far. In this study, we propose the network combining the object detection and classification for mobile graphic processing unit. This network detects pedestrian and face, classifies age and gender from detected face. Proposed network is constructed based on MobileNet, YOLOv2 and skip connection. Both detection and classification models are trained individually and combined as 2-stage structure. Also, attention mechanism is used to improve detection and classification ability. Nvidia Jetson Nano is used to run and evaluate the proposed system.

**키워드**: 기계 학습, 객체 검출, 분류, 모바일넷, 주의 집중 기법

**Keywords**: Machine learning, Object detection, Classification, MobileNet, Attention mechanism

Received 7 April 2022, Revised 19 April 2022, Accepted 20 April 2022

\* Corresponding Author Mi-Kyung Kim(E-mail:ddosun@pusan.ac.kr, Tel:+82-51-510-7052)  
Professor, Software Education Center, Pusan National University, Busan, 46241 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.5.668>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서 론

최근 기계 학습과 심층 학습 기술에 대한 발전이 계속 해서 이루어짐에 따라, 다양한 분야에서 이를 활용하고 있다. 그 중 시각 데이터를 통한 기계 학습은 주변 상황 인지, 결함 감지와 같은 산업 분야와 CCTV를 이용한 사용자 분석, 보안과 같은 서비스 분야에도 많은 영향을 주었다. 이러한 실 사례 중 하나는 무인 주문 시스템을 의미하는 키오스크이다. 키오스크 시스템은 지하철, 식당, 편의 시설 등에 많이 활용되어 단순 자판기의 기능을 넘어서 사용자 분석, 판매 데이터 분석 등의 사물 인터넷과 같은 형태를 가진다.

위와 같은 임베디드 환경에서의 실용성을 고려할 때, 심층 학습 망의 경량화가 중요한 요소 중 하나라고 볼 수 있다. 경량화 방법 중 가지치기(Pruning)는 비교적 중요도가 낮은 신경 회로망의 가중치 연결을 제거하여 매개 변수의 개수를 줄이고, 양자화(Quantization)는 가중치를 표현할 때 사용되는 비트의 수를 의도적으로 감소시켜 신경 회로망이 차지하는 메모리 공간을 줄인다. 가장 직관적인 경량화 방법은 신경 회로망의 구성을 적은 매개 변수로 이루는 방법으로, 사용하는 커널의 크기를 제한하거나 합성곱 연산에 변형을 주어 매개 변수의 개수를 감소시키는 방법 등이 연구되었다.[1]

본 논문에서는 키오스크와 같은 사물 인터넷 플랫폼에서 사용 가능한 사용자 분석을 위한 비전 시스템을 제안한다. 본 시스템은 크게 두 가지의 기능을 수행한다. 첫 번째는 키오스크 주변의 유동 인구의 수를 파악하여 인구 대비 실 사용자의 비율을 계산하는 것이다. 두 번째는 실 사용자의 성별과 나이를 예측하는 것이다. 제안하는 망은 MobileNet을 기반으로 설계되었으며, 검출과 분류 망을 개별적으로 학습한 뒤 결합되는 2-stage의 구조를 띤다. 또한 모든 기능은 소형 폼팩터이자 모바일 그래픽 연산 처리 장치인 Nvidia Jetson Nano에서 실행된다.

## II. 관련 연구

### 2.1. MobileNet

신경 회로망과 더불어 스마트폰과 같은 모바일 기기에 대한 발전 또한 계속되고 있다. 그로 인해 모바일 기

기에서도 실행 가능할 만큼 연산량이 적은 동시에 높은 정확도를 가지는 신경 회로망의 연구가 지속되고 있다.

모바일 환경에서 사용되는 가장 대표적인 신경 회로망은 MobileNet[1]이다. MobileNet은 일반적인 합성곱 연산과 비교하였을 때 연산량이 약 8, 9배가 적은 depthwise seperable convolution을 이용하여, 모바일 기기에서도 사용 가능하도록 설계된 분류 모델이다. MobileNet에서 한 단계 더 진보된 신경 회로망인 MobileNetv2[2]은 더욱 효율적인 계층 구조를 형성하기 위해 선형 병목(Linear bottleneck)과 반전 잔여 블록(Inverted residual block)을 사용한다.

### 2.2. YOLO(You Only Look Once)

객체 검출은 컴퓨터 비전 분야에서 다루는 가장 보편적인 문제 중 하나이다. 특히 심층 학습과 고성능의 그래픽 연산 처리 장치의 도입으로 인해, 실시간성을 확보하는 객체 검출 망의 발전이 다수 이루어졌다.[3, 4]

YOLO[3]가 탄생하기 이전 범용적으로 활용된 객체 검출 시스템은 Faster R-CNN[4]가 있다. Faster R-CNN은 관심 영역(Region of Interest)을 예측하는 별도의 망인 영역 제안 망(Region Proposal Network)을 활용하여 실시간으로 객체를 검출한다. 예측한 각각의 관심 영역에 대한 좌표와 확신 확률을 개별적으로 예측하기 때문에, 예측되는 객체의 수에 따라 수행 시간이 다를 수 있다. 또한 관심 영역을 예측하는 영역 제안 망에 대한 별도의 학습이 필요하기 때문에, 전체 망의 학습 시간이 길어질 수 있다.

YOLO는 이러한 과정을 one-stage로 수행하는 객체 검출 시스템이다. 경계 상자(Bounding box), 확신 확률, 클래스 분류를 동시에 수행하기 때문에, 기존의 객체 검출 시스템에 비해 간단한 구조를 가지고 처리 속도가 빠르다는 장점이 있다.

### 2.3. 주의 집중 기법

주의 집중 기법[5]은 신경 회로망의 성능을 높이는 방법 중 하나이다. 합성곱 신경 회로망(Convolutional Neural Network) 학습 시, 특징 지도(Feature map)가 병목 층을 지날 때 중요한 가중치에 대한 학습을 가속시키는 기능을 한다. 합성곱 신경 회로망은 풀링, 합성곱을 통해 특징 지도의 크기가 줄어들면서 동시에 정보량 또한 감소하는데, 주의 집중 기법을 통해 특징 지도의 중

요하거나 그렇지 않은 부분의 가중치를 조절하면서 문제를 해결한다. 적은 연산량을 통해 성능 향상을 보일 수 있고, 모듈로 생성되어 사용되므로 합성곱 신경 회로망에 쉽게 적용할 수 있도록 설계되었다.

#### 2.4. GIoU(Generalized Intersection over Union)

객체 검출 망 학습 시 경계 상자 예측에 대한 손실 함수로 IoU(Intersection over Union)를 활용하려는 시도가 많아지고 있다. 개별적인 좌표의 값을 직접 예측하거나 평균 절대 오차 혹은 평균 제곱 오차 손실 함수를 사용할 때와 비교하였을 때, 과적합 문제가 발생할 가능성이 적기 때문이다. 그러나 두 경계 상자의 겹침의 정도를 의미하는 IoU의 정의에 따르면, 예측된 경계 상자와 정답 경계 상자가 겹치지 않을 경우 예측된 경계 상자와 정답 경계 상자 간의 거리에 상관없이 IoU 값은 0이 되므로 손실 함수로 사용하기에 무리가 있을 수 있다. GIoU[6]은 이러한 문제를 극복하는 동시에 IoU를 활용하는 방법을 제안하는 손실 함수이다. 두 경계 상자를 둘러싸는 최소의 크기의 가상의 경계 상자를 생성하여, 두 경계 상자의 떨어진 정도를 손실 함수에 적용한다. 이를 통해 겹치지 않은 경계 상자에 대해서도 학습이 가능하게 된다. 수식 1은 GIoU와 이를 이용한 손실 함수에 대한 수식을 나타낸다.  $A$ 와  $B$ 는 IoU 계산을 위한 두 경계 상자,  $C$ 은 두 경계 상자의 영역을 모두 포함하는 최소의 크기의 경계 상자를 의미한다.

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

$$GIoU = IoU - \frac{|C \setminus (A \cup B)|}{C}$$

#### 2.5. Focal 손실

Focal 손실[7]은 교차 엔트로피 손실 사용 시 발생할 수 있는 클래스 불균형 문제를 극복하기 위해 고안된 손실 함수이다. 기존의 교차 엔트로피 손실 함수와 구별되는 점은 예측 확률 값에 대한 승수를 추가로 곱하여, 예측 값의 오차가 클수록 손실에 큰 변화를 준다는 점이다. 이로 인해 예측하기 쉬운 입력에 대한 가중치 변화를 줄이고, 시료 수가 적은 입력에 대한 학습 효과를 높이는 효과를 보이게 된다. 수식 2은 Focal 손실 함수의 수식을 나타낸다.

$$FL(p) = -\alpha(1-p)^\gamma \log(p) \quad (2)$$

$\alpha$ 은 클래스 불균형을 최소화하기 위해 사용되는 하이퍼 파라미터로, 클래스 빈도의 역수를 반영하여 0과 1 사이의 값을 적용한다.  $\gamma$ 은 검출 오류에 대한 손실에 집중하기 위해 사용되는 하이퍼 파라미터로,  $\gamma$  값이 증가할수록 잘 검출된 경우의 손실을 적게 갱신하는 동시에 잘못 검출된 경우의 손실을 극대화하는 효과를 보인다.

### III. 검출 및 분류 모델

본 논문에서 제안하는 망은 보행자 및 얼굴 검출 망, 나이 및 성별 분류 망으로 나뉜다. 그림 1은 전체 망의 구조를 도식화한 것으로, 검출 망을 통해 우선적으로 영역들을 추출한 뒤, 얼굴 영역에 한하여 해당 영역을 분류 망에 입력하여 최종 나이와 성별을 예측한다.

검출 모델에서는 보행자, 얼굴 객체를 검출한다. 보행자 객체는 주변 유동 인구의 수를 파악하기 위해 사용되고, 얼굴 객체는 사용자의 나이 및 성별을 예측하는데 사용된다. 성별은 남성, 여성으로 분류되고, 나이는 15세부터 59세까지에 해당하는 범위로 분류된다.

검출과 분류 망은 모두 MobileNet을 기반으로 설계되었다. 각각의 망은 개별적으로 학습이 이루어지고, 학습 완료 후 하나의 시스템으로 통합된다.

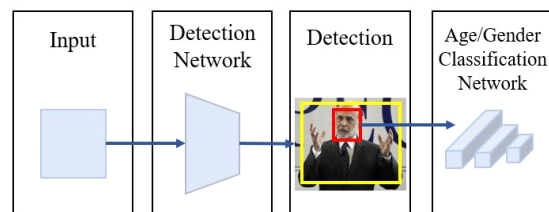


Fig. 1 Proposed model overview

#### 3.1. 사용자 검출 망

검출 알고리즘은 YOLOv2[8]를 기반으로 설계되었다. 기존의 YOLOv2는 Darknet19를 백본으로 사용한다. 이 때 연산에 사용되는 파라미터의 수는 약 5천만 개로, 모바일 환경에서 동작할 경우 속도 저하의 문제가 발생할 수 있다. 따라서 백본 네트워크를 MobileNetv2로 변경하여 파라미터의 수를 약 1/10인 4백만 개로 감소시켰다. 또한 특징 전달을 통한 정확도 향상을 위해,

해당 망에 생략 연결(Skip connection)[9]을 추가하였다. 최종적으로 검출한 객체에 대해 경계 상자의 좌표 값, 확신 확률 그리고 클래스 예측이 이루어진다. 그림 2는 사용자 검출 망을 도식화한 것을 나타낸다.

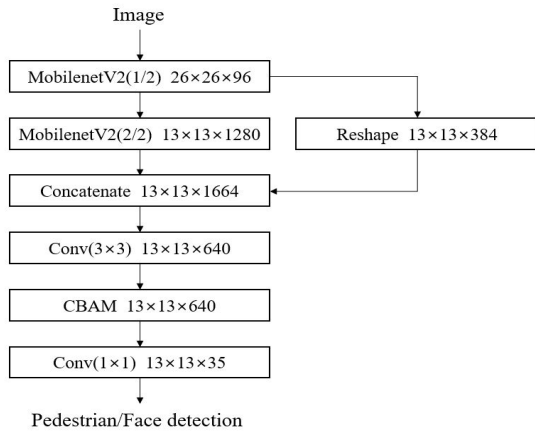


Fig. 2 Structure of detection network

### 3.2. 나이 및 성별 분류 망

나이 및 성별 분류 망은 검출 망에서 추출된 얼굴 영역에 대한 나이와 성별을 예측한다. 검출 망에서와 같은 이유로, 모바일 환경에 적합한 연산량과 처리 속도를 고려하여 백본을 MobileNetV2를 기반으로 설계하였으며, 주요한 특징 추출을 위해 주의 집중 기법인 CBAM[10]을 추가하였다. 헤드는 성별 예측 가지와 나이 예측 가지로 나뉜다. 성별은 남성 혹은 여성, 나이는 15세부터 59세까지 총 45개의 클래스로 분류된다. 그림 3은 나이 및 성별 분류 망의 구조를 도식화한 것을 나타낸다.

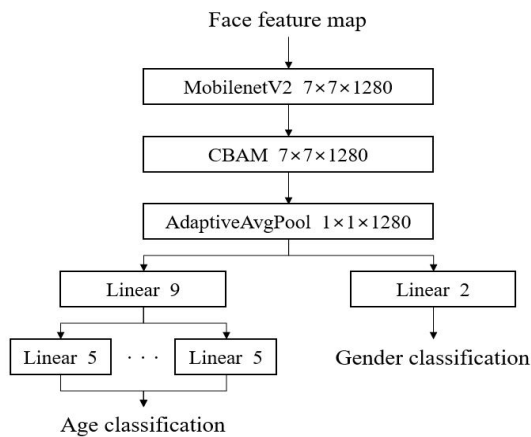


Fig. 3 Structure of age, gender classification network

분류 망의 낮은 복잡도와 학습 데이터 세트의 클래스 불균형 문제를 고려하여, 나이 예측은 총 두 단계로 진행된다. 첫 번째로 45개에 해당하는 나이 클래스를 15~19세, 20~24세, 25~29세, 30~34세, 35~39세, 40~44세, 45~49세, 50~54세 그리고 55~59세와 같이 총 9개의 부모 클래스를 예측한다. 그리고 각각의 부모 클래스를 거쳐 최종 나이 예측이 이루어진다.

### 3.3. 학습 데이터

보행자 및 얼굴 검출 망의 학습에는 MS COCO 2017 Dataset[11]와 Wider Face Dataset[12]를 사용하였다. MS COCO Dataset은 객체 검출, 분할 등 다수의 합성곱 신경망 학습과 성능 평가에 사용되는 데이터 세트이다. 본 논문에서는 MS COCO 2017 Dataset의 90개의 클래스 중에서 사람에 해당하는 클래스만 학습에 반영하였다. Wider Face Dataset은 사람 얼굴 검출을 위한 데이터 세트로, 총 32,203장의 이미지와 약 39만개의 얼굴 객체 정보가 포함되어 있다. 학습 시 38000장의 COCO Dataset의 이미지와 12876장의 Wider Face Dataset의 이미지를 사용하였고, 검증 시 2693장의 COCO Dataset의 이미지와 700장의 Wider Face Dataset의 이미지를 사용하였다. 그림 4, 5는 보행자 및 얼굴 검출 학습에 사용된 데이터 세트의 샘플을 나타낸다.



Fig. 4 Sample images of COCO dataset



Fig. 5 Sample images of Wider Face dataset

나이 및 성별 분류 망의 학습에는 AFAD(Asian Face Age Dataset)[13] 데이터 세트를 사용하였다. AFAD 데이터 세트는 약 16만장의 아시아인의 얼굴 이미지와 나이, 성별 정보를 포함한 데이터 세트이다. 학습 시

104225장의 이미지를, 검증 시 44668장의 이미지를 사용하였다. 그림 6은 나이 및 성별 분류 학습에 사용된 데이터 세트의 샘플을 나타낸다.



Fig. 6 Sample images of AFAD dataset

### 3.4. 손실 함수

손실 함수는 검출 손실과 분류 손실로 구분된다.

검출 손실 함수는 기존의 YOLOv2의 손실 함수를 기반으로 설계되었다. 기존의 YOLO에서 사용하는 손실 함수는 좌표 예측, 확신 확률 예측, 클래스 예측에 대해 모두 합 제공 오차를 사용하였다.

확신 확률 학습 시 예측된 경계 상자와 정답 상자 간의 IoU를 확신 확률에 대한 정답으로 사용한다. 즉 0과 1 사이로 정규화된 값을 예측하게 되는데, 합 제공 오차 손실 함수를 사용할 경우, 기울기가 0이 되어 학습에 지장을 주는 상황이 발생할 수 있다. 마찬가지로 0과 1 사이의 확률을 예측하는 클래스 예측 또한 기존의 손실 함수를 사용할 경우 같은 문제를 야기할 수 있다. 따라서 이와 같은 문제를 해결하기 위해 확신 확률과 클래스 예측에 대해 교차 엔트로피 손실 함수를 사용하였다.

또한 경계 상자 예측에 한하여 GIoU 손실 함수를 사용하였다. GIoU 손실 함수 사용 시 기존의 합 제공 오차 손실을 사용하였을 때보다 학습의 안정성이 증가하는 것을 확인하였다.

분류 손실은 성별 분류 손실과 나이 분류 손실로 구분된다. 각각의 분류 손실은 Focal 손실 함수로 계산되며 더해진다. 수식 3은 각각의 분류 손실에 적용된 Focal 손실 함수의 수식을 나타낸다. 학습 시  $\alpha$ 은 0.25,  $\gamma$ 은 2로 설정하였다.

$$L_{cls} = -\frac{\alpha}{N} \sum_t (1 - \hat{p}_t)^\gamma p_t \log(\hat{p}_t) \quad (3)$$

## IV. 실험 및 결과

실험은 검출 망과 분류 망을 각각 학습한 뒤 병합하는 방식으로 진행하였다.

### 4.1. 실험 환경

제안하는 망은 파이썬(Python)으로 구현되었으며, 심층 학습 라이브러리인 파이토치(PyTorch)를 통해 설계 및 학습되었다. 파이토치는 각각의 망을 모듈로 생성하여 사용하기 때문에 객체 지향적인 설계에 적합한 프레임워크이다. 또한 학습의 가속화를 위해 Nvidia GeForce RTX 3080 그래픽 처리 장치를 사용하였다. 학습에 사용한 최적화 기법으로 적응형 모멘텀 추정 기법(Adam)을 사용하였으며, 초기 학습률을 0.0001로 설정하여 검출 망은 150세대(Epoch), 분류 망은 200세대 만큼 학습하였다.

### 4.2. 평가

제안하는 망의 평가는 Nvidia Jetson Nano 환경에서 검출의 정확도와 분류의 정확도로 구분하여 진행하였다.

객체 검출의 성능에 대한 척도는 mAP를 사용하였다. mAP는 정답이라고 예측한 것 중 실제 정답의 비율인 정밀도와 실제 정답인 것 중 예측한 정답의 비율인 재현율을 이용한 정밀도 재현율 곡선(Precision Recall curve)을 통해 계산된다. 객체 검출의 성능 평가는 COCO 데이터 세트의 테스트 이미지 41000장과 Wider Face 데이터 세트의 테스트 이미지 1400장을 이용하여 진행하였다. 그림 7은 검출 망 학습 시 손실 그래프를 나타낸다.

분류의 성능에 대한 척도는 정확도를 사용하였다. 분류의 성능 평가는 AFAD 데이터 세트의 테스트 이미지

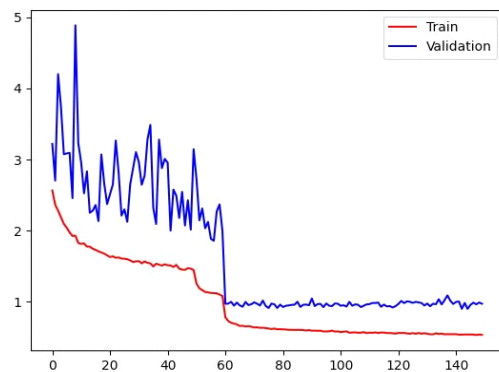


Fig. 7 Loss graph for detection network



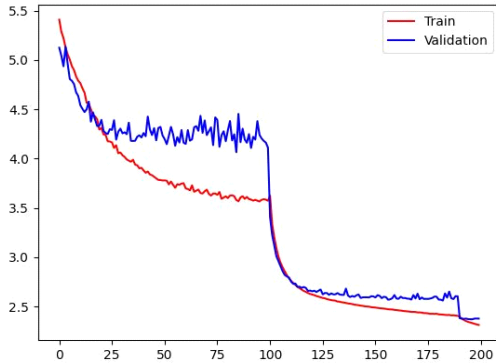


Fig. 8 Loss graph for classification network

5560장을 이용하여 진행하였다. 그림 8은 분류 망 학습 시 손실 그래프를 나타낸다.

표 1은 검출과 분류 망의 성능 지표를 나타낸다.

Table. 1 Evaluation on proposed network

Type	Metric	Value
Detection	mAP	0.354
Classification	Age accuracy(%)	88.47
	Gender accuracy(%)	97.84
Time	FPS	6.02

그림 9는 보행자 검출과 나이 및 성별 분류 결과 이미지를 나타낸다.



Fig. 9 Detection and classification result

## V. 결론

본 논문에서는 Nvidia Jetson Nano 환경에서 사용 가능한 사용자 검출 및 나이 성별 분류 경량 망을 학습하였다. 사전 학습된 MobileNet을 이용한 전이 학습을 통해 검출, 분류 망의 복잡도를 감소하였고, 검출 망 설계 시 MobileNet에 생략 연결을 추가하여 성능을 향상시킬

수 있었다. 나이 분류 망 학습 시 Focal 손실 함수를 사용하여 AFAD 데이터 세트에 포함된 나이 별 데이터의 수의 불균형 문제를 해소하였다. 분류 정확도 성능을 더욱 향상시키기 위해, 학습 시 데이터 수가 부족한 클래스 예측에 대해 가중치를 추가적으로 부여하거나 데이터 균형을 맞추기 위해 시료 수가 많은 클래스의 개수만큼 over-sampling하는 방법을 적용할 수 있을 것으로 보인다.

## ACKNOWLEDGEMENT

This work was supported by a 2-Year Research Grant of Pusan National University.

## REFERENCES

- [ 1 ] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.
- [ 2 ] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City: UT, USA, pp. 4510-4520, 2018.
- [ 3 ] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas: NV, USA, pp. 779-788, 2016.
- [ 4 ] S. Ren, K. He, R. Girchick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems 28*, Montreal, Canada, vol. 1, 2015.
- [ 5 ] B. Zhu, P. Hofstee, J. Lee, and Z. Al-Ars, "An Attention Module for Convolutional Neural Networks," *arXiv preprint arXiv:2108.08205*, 2021.
- [ 6 ] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach: CA, USA, pp. 658-666, 2019.

- [ 7 ] T. -Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2999-3007, 2017.
- [ 8 ] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu: HI, USA, pp. 6517-6525, 2017.
- [ 9 ] 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*, Las Vegas: NV, USA, pp. 770-778, 2016.
- [10] S. Woo, J. Park, J. -Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision*, Munich, Germany, pp. 3-19, 2018.
- [11] T. -Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollar, "Microsoft COCO: Common Objects in Context," *arXiv preprint arXiv:1405.0312*, Feb. 2015.
- [12] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A Face Detection Benchmark," *arXiv preprint arXiv:1511.06523*, 2015.
- [13] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal Regression With Multiple Output CNN for Age Estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas: NV, USA, pp. 4920-4928, 2016.



서지원(Ji-Won Seo)

부산대학교 정보컴퓨터공학부 학사  
부산대학교 정보융합공학과 석사과정 (현재)  
※관심분야: 심층 학습, 컴퓨터 비전, 깊이 추정, 객체 검출



김미경(Mi-Kyung Kim)

부산대학교 전산학과(공학석사)  
부산대학교 멀티미디어 협동과정(공학박사)  
부산대학교 전자공학과 강의전담교수  
부산대학교 소프트웨어교육센터 강의전담교수 (현재)  
※관심분야: 컴퓨터 비전, 물체 추적, 심층 학습