

YOLO 기반 전동 킥보드 사진 인식 시스템 개발

김채현¹, 유사라¹, 윤서영¹, 김가영², 공현정¹, 이진복³, 송성민³, 이기용¹¹숙명여자대학교 컴퓨터학과²숙명여자대학교 통계학과³(주)올룰로

{7chaeny25, rrrr4ra, ysyoun1011, kijh30123, konghj}@sookmyung.ac.kr

{jblee, ssm0219}@olulo.io, kiyonglee@sookmyung.ac.kr

Development of a YOLO-Based Electric Kick Scooter Photo Recognition System

Chaehyeon Kim¹, Sara Yu¹, SeoYoung Yoon¹, Gayoung Kim²,Hyeonjeong Kong¹, Jinbok Lee³, Sungmin Song³, Ki Yong Lee¹¹Department of Computer Science, Sookmyung Women's University²Department of Statistics, Sookmyung Women's University³Olulo

요 약

최근 편리성과 경제성 등의 이유로 개인형 이동장치인 전동 킥보드의 사용이 증가하고 있다. 사용자들은 앱으로 주변의 전동 킥보드 위치를 확인한 뒤, 가까운 기기를 찾아 이용한다. 하지만 전동 킥보드의 위치는 GPS로 표시되기 때문에 10 m 이상의 오차가 날 수 있다. 이를 보완하기 위해 ㈜올룰로의 킥고잉은 사용자가 전동 킥보드 반납 시 촬영한 전동 킥보드 사진을 GPS 위치 정보와 함께 제공한다. 이 사진을 통해 다음 사용자는 더욱 정확히 전동 킥보드를 찾을 수 있다. 하지만 일부 사용자들은 전동 킥보드가 존재하지 않는 사진을 올리기도 하며, 따라서 사용자가 촬영한 사진 중 실제 전동 킥보드가 존재하는 사진들만 제공하는 것은 매우 중요하다. 따라서 본 논문은 사용자가 촬영한 사진 중 실제 전동 킥보드가 존재하는 사진들만 정확히 인식하는 YOLO 기반 시스템을 개발한다. 제안 방법은 (1) 전동 킥보드를 부분별로 탐지하는 기법과 (2) 전동 킥보드를 촬영된 각도에 따라 세분화하여 인식하는 기법을 사용한다. 실제 사용자가 촬영한 사진을 사용한 실험 결과, 제안 방법은 기존 방법에 비해 더욱 정확히 전동 킥보드 사진을 인식하는 것을 확인하였다.

1. 서론

전동 킥보드는 개인형 이동장치의 하나로서, 편리성과 경제성 외에도 교통혼잡, 대기오염, 주차공간 부족과 같은 도시 문제 개선에 적극적으로 기여할 수 있다는 이유로 최근 활발히 사용되고 있다 [1]. 전동 킥보드를 사용하기 위해 사용자들은 앱으로 주변의 전동 킥보드 위치를 확인한 뒤, 가까운 위치로 이동하여 전동 킥보드를 찾아 이용한다. 하지만 앱에 표시된 위치는 GPS를 이용하기 때문에 앱에 표시된 위치와 전동 킥보드의 실제 위치는 10 m 이상 차이가 날 수 있다. 이 경우 사용자는 이용하려는 전동 킥보드를 찾는 데 어려움을 겪을 수 있다.

킥고잉(Kickgoing)은 ㈜올룰로에서 제공하는 국내 최초 전동 킥보드 공유 서비스이다 [2]. 킥고잉은 사용자가 GPS 위치만으로 전동 킥보드를 찾는 어려움을 보완하기 위해 사용자가 전동 킥보드를 반납할 때 반납한 전동 킥보드의 사진을 촬영하여 올리도록 하고 있다. 이후 앱에서 전동 킥보드의 GPS 위치뿐

만 아니라 그의 사진을 함께 제공한다. 따라서 다음 사용자는 이 사진을 사용하여 더욱 정확히 해당 전동 킥보드를 찾을 수 있다. 하지만 일부 사용자는 전동 킥보드를 잘 알아볼 수 없거나 아예 전동 킥보드가 없는 사진을 올리기도 한다. 따라서 사용자가 촬영한 사진 중 실제 전동 킥보드가 존재하는 사진들만 제공하는 것이 매우 중요하다 [3]. 하지만 사용자들은 전동 킥보드를 매우 다양한 각도와 거리에서 촬영하기 때문에 사진에 전동 킥보드가 존재하는지 판단하는 것은 매우 어려운 일이다.

따라서 본 논문은 사용자가 촬영한 사진 중 실제 전동 킥보드가 존재하는 사진들만 정확히 인식하는 시스템을 개발한다. 제안 방법은 (1) 전동 킥보드를 핸들(handle), 바(bar), 발판(bottom) 세 부분으로 나누어 부분별로 탐지하는 기법과 (2) 전동 킥보드를 그의 촬영 각도에 따라 세분화하여 인식하는 기법을 사용한다. 실제 사용자가 촬영한 사진을 이용한 실험 결과, 제안 방법은 기존 방법에 비해 더

욱 정확하게 전동 키포드 사진을 인식하는 것을 확인하였다. 본 논문은 다음과 같이 구성된다. 2장에서는 관련 연구를 간략히 설명하고, 3장에서는 본 논문의 제안 방법을 구체적으로 설명한다. 4장에서는 실험 데이터에 기반한 실험 결과를 제시하고, 마지막 5장에서 결론을 맺는다.

2. 관련 연구

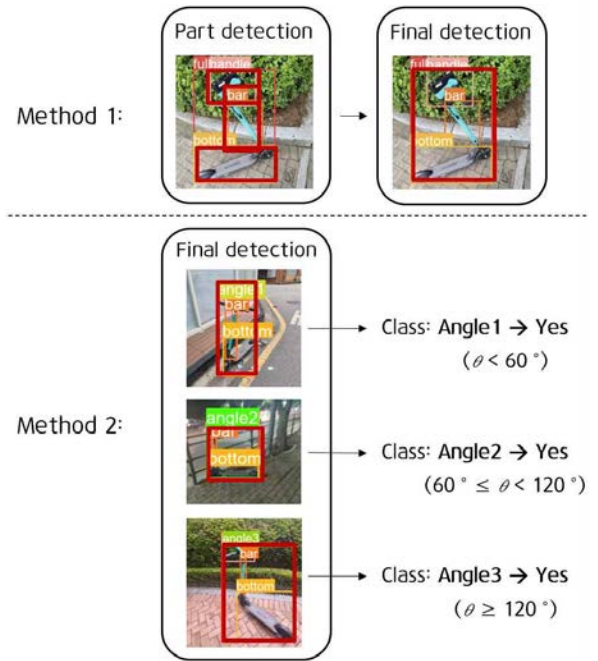
객체 인식(object detection)은 이미지나 비디오상의 객체를 식별하는 기술로, 객체를 식별하고 해당 객체의 위치를 정확하게 찾아주어 이미지 내에서 특정 객체의 존재 여부뿐만 아니라 위치까지 알 수 있게 해준다. 객체 인식 및 객체 위치 검출 기법에는 주로 합성곱 신경망(convolutional neural network, CNN)이 활용되고 있는데, CNN 기술을 활용한 대표적인 기법으로는 R-CNN, YOLO(You Only Look Once) 등이 존재한다 [4].

그중 YOLO는 여러 테두리 상자(bounding box) 생성과 이에 대한 클래스 분류(classification)까지 동시에 수행하는 통합 인식(unified detection) 모델이다. YOLO는 하나의 이미지가 입력되면 이를 그리드 셀(grid cell)들로 분할하고 여러 테두리 상자들을 생성한다. 그다음 각 테두리 상자에 대해 각 객체가 존재할 확률을 예측하여 이를 각 그리드 셀에 부여한다. 마지막으로 각 테두리 상자에 포함된 그리드 셀들이 포함하고 있는 객체에 대한 정보와 IoU(Intersection over Union) 값을 사용하여 테두리 상자 내에 객체가 포함되어 있는지의 여부와 포함되어 있는 객체를 결정한다. 본 논문의 시스템 개발에 사용된 YOLOv5[5]는 YOLO 모델의 5번째 버전으로 신경망의 크기에 따라 Small, Medium, Large, XLarge로 나뉜다. 이들은 순서대로 더 많은 가중치를 가지며 정확도는 높아지지만 속도는 느려지는 단점을 갖는다.

3. 제안 방법

본 장에서는 전동 키포드가 존재하는 사진을 더욱 정확히 인식하기 위해 본 논문에서 개발한 시스템을 설명한다. 본 논문에서 개발한 시스템은 객체 인식에 YOLOv5를 사용하였다. 전동 키포드는 그 모양의 특성상 촬영 각도에 따라 모양이 상당히 달라지기 때문에 단순히 YOLOv5를 사용하는 것만으로는 인식 정확도에 한계가 있다. 따라서 본 논문에서는 다음 2가지 기법을 사용하였다: (1) 전동 키포드를 3부분으로 나누어 부분별로 탐지하는 기법, (2) 전동

키포드를 촬영 각도에 따라 세분화하여 인식하는 기법. (그림 1)은 제안하는 2가지 기법을 나타낸다.



(그림 1) 제안하는 2가지 기법

3.1 전동 키포드 부분별 탐지 기법

전동 키포드는 크게 핸들, 바, 발판 3부분으로 나눌 수 있다. 촬영 각도에 따라 전동 키포드의 전체 모양은 크게 변하지만 각 부분은 비교적 일관된 모양을 유지한다. 따라서 제안 방법은 전동 키포드 사진을 (그림 2)와 같이 핸들, 바, 발판으로 나누어 각각 라벨링을 진행하고, 입력 사진에서 핸들, 바, 발판을 각각 탐지하도록 하였다.



(그림 2) 부분별 라벨링의 예

사진에서 3개 부분이 모두 탐지되면 탐지된 3개의 테두리 상자들 중 가장 좌측 상단의 x, y 좌표와 가장 우측 하단의 x, y 좌표를 모서리로 하는 테두리 상자를 전동 키포드의 전체 테두리 상자로 반환한다. 따라서 사진에서 전동 키포드가 다소 특이한 각도로 나타나더라도 각 부분을 비교적 정확하게 탐지하여 전동 키포드를 검출할 수 있다.

3.2 촬영 각도에 따른 분류 세분화 기법

앞서 언급한 바와 같이 전동 키포드는 보는 각도에 따라 전체적인 모양이 매우 다르다. 가령 전동 키포드가 측면에서 촬영되면 ‘ㄴ’자 모양으로 보여지고, 정면 혹은 후면에서 촬영되었다면 ‘|’자 모양으로 보여지게 된다. 이처럼 전동 키포드는 촬영 각도에 따라 모양이 상당히 다르기 때문에, 이들을 모두 동일한 객체로 라벨링하여 YOLOv5를 훈련시키면 인식 정확도에 한계가 있다. 따라서 제안 방법은 (그림 1)에서와 같이 사진 속 전동 키포드의 바와 발판 사이의 각도 θ 가 60° 미만인 경우 Angle1로, 60° 이상 120° 미만인 경우 Angle2로, 120° 이상인 경우 Angle3으로 나누어 라벨링하고 이 라벨링을 사용하여 YOLOv5를 훈련시킨다. 만약 입력 사진이 (그림 1)에서와 같이 Angle1, Angle2, Angle3 중 하나로 분류되면 해당 사진을 전동 키포드가 존재하는 사진으로 분류하며, 그렇지 않으면 전동 키포드가 존재하지 않는 사진으로 분류한다.

4. 성능 평가

본 장에서는 본 논문에서 개발한 시스템이 전동 키포드가 존재하는 사진을 얼마나 정확히 탐지하는지를 평가하였다. 실험에 사용된 모델은 PyTorch로 구현하였으며, 모델의 성능 척도로는 mAP(mean Average Precision), 정확도(precision), 재현율(recall)을 사용하였다. 실험 데이터로는 ㈜울룰로에서 제공받은 3,102장의 킥고잉 전동 키포드 반납 사진을 사용하였다. 반납 사진에는 전동 키포드가 존재하는 사진과 그렇지 않은 사진이 섞여 있으며, 모델이 전동 키포드가 존재하는 사진을 얼마나 정확히 판별하는지를 평가하였다. 실험에서는 이 사진들을 5:3:2로 나누어 각각 훈련데이터, 검증데이터, 테스트 데이터로 사용하였다.

<표 1> 제안 방법 1의 성능평가 결과

Model	mAP	AP@.75	AP@.85	AP@.95
제안 방법 1	11.861	10.645	9.355	4.194
기존 방법	11.523	10.806	8.710	1.935

<표 1>은 3.1절에서 설명한 첫 번째 제안 방법과 전동 키포드를 부분으로 나누지 않고 전체를 라벨링하여 YOLOv5를 훈련시킨 기존 방법의 성능을 mAP로 평가한 결과이다. 여기서 AP@n은 IOU = n 일 때의 mAP를 나타낸다. <표 1>에서 볼 수 있듯

이 제안 방법 1은 기존 방법에 비해 비교적 좋은 성능을 보여주고 있으며, 특히 IOU 임계값이 증가할수록 더 효과적임을 확인할 수 있다.

<표 2> 제안 방법 2의 성능평가 결과

Model	Precision	Recall	F1-Score
제안 방법 2	91.80%	91.97%	0.92
기존 방법	86.40%	90.58%	0.88

<표 2>는 3.2절에서 설명한 두 번째 제안 방법과 전동 키포드 사진을 촬영 각도에 따라 세분화하지 않고 한 종류로 라벨링한 기존 방법의 성능을 정확도와 재현율로 평가한 결과이다. <표 2>에서 볼 수 있듯이 제안 방법은 기존 방법에 비해 정확도와 재현율을 모두 조금씩 향상시키고 있다.

5. 결론

본 논문에서는 전동 키포드가 존재하는 사진을 더욱 정확히 인식하는 시스템을 개발하였다. 본 논문에서 개발한 시스템은 촬영 각도에 따라 모양이 크게 달라지는 전동 키포드를 인식하기 위해 (1) 전동 키포드를 핸들, 바, 발판으로 나누어 부분별로 탐지하는 기법과 (2) 전동 키포드의 촬영 각도에 따라 세분화하여 인식하는 기법을 사용하였다. 실제 사진을 사용하여 제안 방법의 성능을 평가한 결과, mAP, 정확도, 재현율 측면에서 모두 조금씩 인식 정확도가 향상하는 것을 확인하였다.

Acknowledgement

이 성과는 2022학년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF2022H1D8A303739411)

참고문헌

[1] S. Kim et al, "Study on Shared E-scooter Usage Characteristics and Influencing Factors," The Journal of The Korea Institute of Intelligent Transport Systems, vol. 20, no. 1, pp. 40-53, Feb 2021.
 [2] 킥고잉, <https://kickgoing.io/>.
 [3] S. Gilroy et al, "E-Scooter Rider Detection and Classification in Dense Urban Environment," ArXiv, abs/2205.10184, May 2022.
 [4] Z. Zhao, P. Zheng, S. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov 2019.
 [5] G. Jocher, <https://github.com/ultralytics/yolov5>, 2020.