

## 동영상 내 객체 추적을 위한 영상 데이터셋 구축 방법

김지성<sup>1</sup> · 허경용<sup>2</sup> · 장시웅<sup>3\*</sup>

### Building Method an Image Dataset for Tracking Objects in a Video

Ji-Seong Kim<sup>1</sup> · Gyeongyong Heo<sup>2</sup> · Si-Woong Jang<sup>3\*</sup>

<sup>1</sup>Graduate student, Department of Computer Engineering, Dong-Eui University, Busan, 47340 Korea

<sup>2</sup>Associate Professor, Department of Electronic Engineering, Dong-eui University, Busan, 47340 Korea

<sup>3\*</sup>Professor, Department of Computer Engineering, Dong-Eui University, Busan, 47340 Korea

#### 요 약

영상 딥러닝을 위해서는 다량의 영상 데이터셋이 필요한데, 객체의 종류에 따라 영상을 구하고 영상 데이터셋을 구축하는 방법에 많은 차이가 있다. 본 논문에서는 딥러닝을 위한 영상 데이터셋을 구축하는 방법을 제시하고 추적하는 객체에 따라 달라지는 성능을 분석하였다. 제안하는 데이터셋 구축방법을 활용하여 객체를 회전시킨 후 동영상 촬영, 분할하여 커스텀 데이터셋을 구축하고, 성능을 분석한 결과 95% 이상의 객체 검출률을 보였으며, 이동 시 객체의 형상 변화가 적은 경우에 더 높은 성능이 나타났다. 영상 데이터를 구하기 어렵고, 형태의 변화가 적은 객체를 동영상 내에서 추적하기 위한 상황을 위하여는 본 논문에서 제시한 데이터셋 구축방법을 활용하는 것이 효과적인 것으로 판단된다.

#### ABSTRACT

A large amount of image data sets are required for image deep learning, and there are many differences in the method of obtaining images and constructing image data sets depending on the type of object. In this paper, we presented a method of constructing an image data set for deep learning and analyzed the performance that varies depending on the object to be tracked. We took a video by rotating the object, and then created a data set by segmenting the video using the proposed data set construction method. As a result of performance analysis, detection rate was more than 95%, and detection rate of objects with little change in shape was higher performance. It is considered that it is effective to use the data set construction method presented in this paper for a situation in which it is difficult to obtain image data and to track an object with little change in shape within a video.

**키워드** : 인공지능, 딥러닝, 데이터셋, 객체추적

**Keywords** : Artificial intelligence, Deep learning, Datasets, Object tracking

Received 21 October 2021, Revised 22 October 2021, Accepted 18 November 2021

\* Corresponding Author Si-Woong Jang(E-mail:swjang@deu.ac.kr, Tel:+82-51-890-2354)  
Professor, Department of Computer Engineering, Dong-EUI University, Busan 47340, Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.12.1790>

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. 서론

최근 딥러닝 기술을 활용해 영상 내에 객체를 추적하는 기술은 여러 분야에서 많은 연구가 진행되어 왔다 [1-9]. 진행된 연구들은 지하 주차장 CCTV 차량 추적 [1], 양봉가의 피해를 줄이기 위한 말벌 인식[2], 실시간 온라인 수업 태도 행동 인식[3], 알약 종류 인식[4], 음료 정보 시각화를 위한 음료수 인식[5] 등 아주 다양한 분야에서 연구가 진행되고 있다. 이러한 객체 인식 딥러닝의 효과를 높이기 위한 방법 중 하나는 다량의 데이터셋을 확보하는 것이지만 데이터셋을 구축하는 데에는 많은 비용이 소모된다.

따라서, 데이터셋을 구축 또는 증강하는 방법에 대한 연구도 진행되어 왔는데, 목조건물 크랙 감지를 위한 데이터셋 증강 기법[6], 골프 스윙 헤드 궤적 추정 데이터셋 구축 방법[7], 지역 특화 영상 수집 및 데이터셋 구축 [8], 산불 연기 데이터셋 구축[9]과 같이 필요에 맞추어 데이터셋을 구축하는 방법에 대한 연구도 진행 중이다. 본 논문에서는 연구의 필요에 맞추어 다량의 영상 데이터셋을 보다 쉽게 구축하는 방법을 제안하고 추적하고자 하는 객체의 특성에 따른 데이터셋 성능을 분석한다. 논문에서 소개하는 데이터셋은 YOLO(You Only Look Once) 객체 탐지 네트워크에 최적화된 데이터셋이다. 데이터셋으로 구축하고자 하는 객체를 회전판 위에 올린 후 동영상을 촬영, 분할하는 방식을 사용하여 다량의 영상을 생성하고 생성된 영상에서 배경을 제거하는 방식으로 객체의 위치를 찾고 자동 라벨링 하는 방식을 사용하였다. 본 논문의 2장에서는 영상 데이터셋 구축에 관련된 연구들에 대해 살펴보고, 3장에서는 동영상 촬영 및 분할, 자동 라벨링을 통한 데이터셋 구축방법을 기술하고, 4장에서는 구축된 영상 데이터셋을 사용하여 학습하는 방법 및 데이터셋 성능을 분석하고 5장에서는 결론을 맺는다.

## II. 관련 연구

데이터셋을 구축하는 여러 연구들[6-8,10-14]에서는 목조건물, 골프, 자율주행, 산불과 같은 분야에서 필요한 데이터셋 구축 내용을 다루고 있다.

목조건물 크랙 감지를 위한 데이터셋 증강 기법[6]에

서는 목조건물에 크랙이 있는 영상을 이용하여 크랙만을 움직여 데이터셋을 증강하는 기법을 제안하였으며, [7]에서는 2차원 골프 스윙 동영상에서 스윙 자세를 분석하여 클럽 헤드 궤적을 추정하는 방법으로 데이터셋을 구축하는 방법을 제시하였다. 지역 특화 이미지 수집 및 데이터셋 구축[8]에서는 로드뷰 API를 활용하여 자율주행을 위해 해당 지역에 특화된 영상 데이터셋을 구축하는 애플리케이션을 구현하였으며, [10]에서는 국내 비정형 주행 환경을 포함한 자율주행 시뮬레이터에서 발생하는 데이터를 데이터셋으로 구축하여 자율주행 서비스의 안정성을 향상시키는 방법을 제안하고 있다. 위에서 제안하는 데이터셋 구축 방법은 특정 분야에서만 사용 가능하다는 제한점이 있는데, 본 논문에서 제안하는 객체 추적을 위한 영상 데이터셋 구축 방법은 데이터셋 구축에 들어가는 비용을 줄이는 방법을 제시하며 영상 내에서 객체를 추적하는 경우 폭넓은 활용성을 보이고 있다. 또한 객체가 이동하는 특성에 따른 데이터셋 효율을 함께 제시하고 있다.

본 논문에서 제시하는 영상 데이터셋은 YOLOv4 [11-14] 네트워크에 최적화된 데이터셋인데, YOLO 네트워크를 활용한 학습을 위한 데이터셋은 Bounding Box를 사용하여 라벨링을 한다. 라벨링이란 영상 내 객체에 라벨을 달아 분류하는 것이다. YOLO의 Bounding Box는 Class index, box 중심 x좌표, box 중심 y좌표, 너비 비율, 높이 비율 총 5가지 영상 내에 객체 정보를 담고 있다. 가장 보편적인 YOLO 데이터셋 구축 방법은 YOLO\_MARK라는 도구를 사용하여 직접 영상 내에 객체 위치에 Bounding Box를 그려주는 방식을 사용하는데, 이러한 방법은 많은 량의 데이터셋을 구축하기에 어려움이 있기 때문에 본 논문에서 제시하는 데이터셋 구축 방법에는 자동으로 데이터셋 라벨링 하는 방법을 제시한다. YOLO 데이터셋의 내의 객체 Bounding Box 형식은 그림 1과 같다.

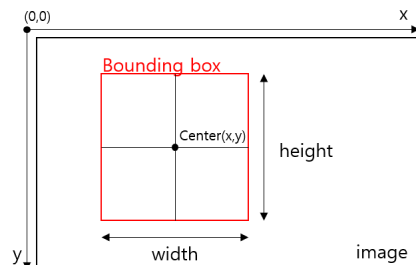


Fig. 1 Bounding box format for object

### III. 영상 데이터셋 구축

3장에서는 YOLO 네트워크에 최적화된 영상 데이터셋을 구축하는 방법을 기술한다. 영상 촬영, 영상분할, 영상 라벨링으로 구성되어 있다.

#### 3.1. 영상 촬영

데이터셋으로 구축할 객체의 동영상을 직접 촬영하여 영상 데이터셋을 생성한다. 촬영 방법은 일정한 속도로 회전하는 회전판 위에 객체를 올려둔 뒤 동영상을 촬영하여 객체의 모든 면을 담고 있는 동영상을 얻는다. 추후 영상처리를 통해 배경의 다양성을 확보하기 위해 배경에는 크로마키 천을 배치하여 동영상을 촬영한다. 영상 촬영 모습은 그림 2와 같다.

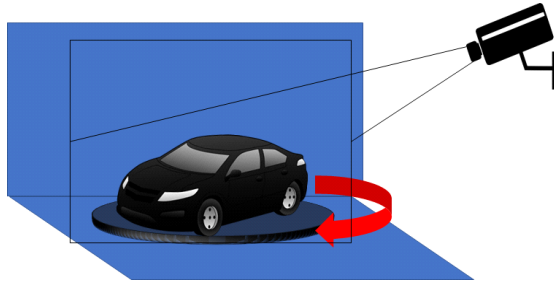


Fig. 2 The structure of the video shoot

#### 3.2. 영상 분할

영상분할 방법은 3.1절에서 촬영한 동영상을 프레임 별로 나눔으로써 객체의 여러 면을 가지는 여러 장의 이미지로 분할한다. 동영상 프레임 분할에서는 객체 동영상을 읽은 뒤 프레임당 회전 각도를 계산하는데 계산식은 수식(1)과 같다.

$$R = \frac{RPM \times 360^\circ}{60 \times FPS} \quad (1)$$

R은 프레임당 회전 각도를 뜻하며 RPM은 분당 회전 수, FPS는 촬영한 영상의 초당 프레임 수이다. RPM과 한 바퀴의 각도인 360°를 곱해 1분간 회전한 각도를 구하고 그 각도를 1분당 프레임 수로 나누어 한 프레임당 회전한 각도를 구할 수 있다. 그다음 분할단위를 계산하는데 여기서 분할단위(U)란 객체의 다면 이미지를 분할하는 각도 단위이다. 분할하기 위한 각도를 한 프레임당 객체의 회전 각도로 나누어 주면 분할단위(U)를 구할 수 있다. 예를 들어 한 프레임당 객체의 회전 각도 R이

2°일 때 객체를 10°단위로 분할하기 위한 분할단위는 5가 된다. 영상 분할 방법의 순서도는 그림 3과 같다.

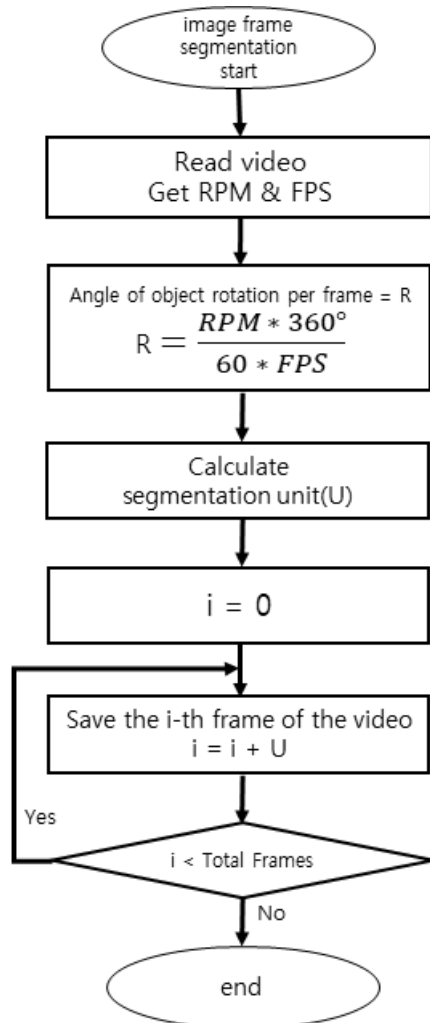


Fig. 3 Image Frame Segmentation department flow chart

#### 3.3. 영상 라벨링

영상 라벨링은 인공지능의 정확한 학습을 위해서 영상 내의 객체에 라벨을 달아 분류하는 것이다. 개인이 영상 데이터셋을 구축하는 데에 있어서 가장 많은 노력이 필요하기 때문에 영상 라벨링 단계를 자동으로 구성하여 데이터셋의 구축 비용을 줄였다.

영상 라벨링은 3.1절에서 촬영된 동영상 데이터를 Gaussian Mixture를 기반으로 이전 프레임과 현재 프레

임 간의 비교를 통해 차분 영상을 얻고, 차분 영상에서 값이 0이 아닌 픽셀이 변화가 있는 곳이기 때문에, 변화가 일어난 픽셀을 이어 객체의 위치로 인식하여 추적하는 방법으로 동영상 내의 객체 위치를 찾는다. 이때 불필요한 배경의 움직임을 제거하여 객체추적의 정확도를 높이기 위해 동영상 내에 객체가 회전하는 위치를 ROI(Region of Interest)로 지정한다. 픽셀값이 0이 아닌 픽셀을 이어서 Bounding Box를 지정하다 보면 객체의 색상에 따라 인식되지 않거나 하나의 객체로 이어지지 않는 경우가 발생하기 때문에 차분 영상의 픽셀값이 0이 아닌 점에 확장 연산을 10회 수행하여 픽셀들을 이어 자동 라벨링의 정확도를 높였다. 확장 연산 적용 전/후 상태는 그림 4와 같다.



Fig. 4 dilatation operation

차분 영상 결과에서 값이 0이 아닌 픽셀을 이어준 다음 Bounding Box를 지정할 수 있다. Bounding Box 지정 결과는 그림 5와 같다.

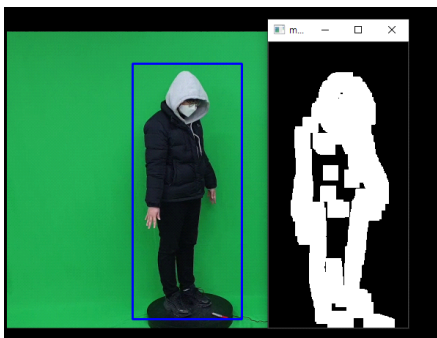


Fig. 5 Specify the bounding box

3.2절에서 계산한 분할단위마다 객체에 라벨링된 Bounding Box 정보를 텍스트 파일로 저장하여 이미지 라벨링을 자동으로 수행할 수 있다. YOLO에서 사용하는 텍스트 파일 내의 데이터 형식은 Class index, box 중심 x 좌표, box 중심 y 좌표, 너비 비율, 높이인데 중심 좌표도 전체 이미지 대비 비율로 표시한다. 지정된 Bounding Box의 좌표, 영상 크기를 YOLO 데이터셋 형식으로 변환해서 텍스트 파일에 저장함으로써 라벨링 된 데이터셋을 구축할 수 있다.

#### IV. 데이터셋 성능 분석

3장에서 기술한 방법대로 사람 데이터셋을 구축하였으며 촬영한 영상의 해상도는 1920x1080 FHD, FPS는 30이며 영상 내의 객체의 RPM은 2 이다. 해당 영상의 프레임당 회전 각도(R)는 0.4°로 10도, 30도, 60도, 90도로 분할하기 위한 분할 단위(U)는 25, 75, 150, 225로 각 분할 단위 마다 분할 시 10도, 30도, 60도, 90도 단위로 분할된다. 각도별로 분할된 데이터셋은 3140, 1048, 524, 348장으로 구성되어 있다. 구축한 데이터셋을 Darknet, YOLOv4를 이용하여 학습 진행하였으며 epoch는 15000으로 지정했다. 학습결과로 동영상 내의 객체의 이동을 추적하는 실험을 진행하였다.

실험 환경은 7~14m 거리에서 촬영한 이동하는 사람 동영상을 프레임단위로 분할하여 688장의 영상으로 분할하였고, 688장으로 구성된 동영상에서 객체를 추적하며 추적을 놓친 영상의 개수를 세어 객체 추적 성능을 측정하였다. 동영상 내에 객체를 탐지한 결과는 그림 6과 같다.

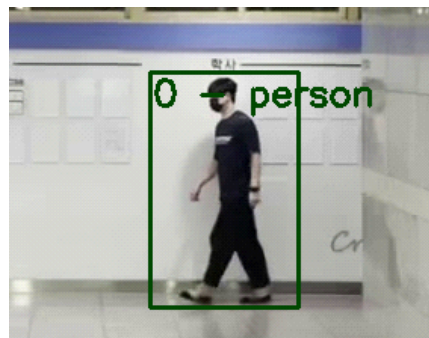


Fig. 6 Results of object detection in the video

학습 반복 횟수에 따라 다른 결과를 보이기 때문에 결과별로 6000, 7000, 9000, 11000, 13000, 15000, best 총 7번씩 테스트를 진행하였으며, best는 학습 진행 과정 중 예상 데이터와 실제 데이터 간의 손실이 가장 적어 최적의 가중치로 판단되는 결과이고, epoch의 숫자는 학습 반복 횟수마다 백업된 가중치이다. 표1은 분할 각도별 학습 횟수별로 객체 추적 테스트 결과 놓친 프레임 수 및 평균, 정확도를 나타내고 있고, 그림 7에서 객체 탐지 결과를 그래프로 나타내고 있다.

Table. 1 Object detection result

| split angle<br>epoch | 10    | 30    | 60    | 90    |
|----------------------|-------|-------|-------|-------|
| 6000                 | 71    | 20    | 62    | 65    |
| 7000                 | 10    | 100   | 37    | 75    |
| 9000                 | 25    | 23    | 123   | 92    |
| 11000                | 5     | 41    | 55    | 136   |
| 13000                | 43    | 25    | 81    | 96    |
| 15000                | 33    | 29    | 132   | 52    |
| best                 | 29    | 47    | 37    | 28    |
| average              | 20.71 | 37.86 | 66.43 | 68.43 |
| detection rate       | 96.99 | 94.50 | 90.34 | 90.05 |

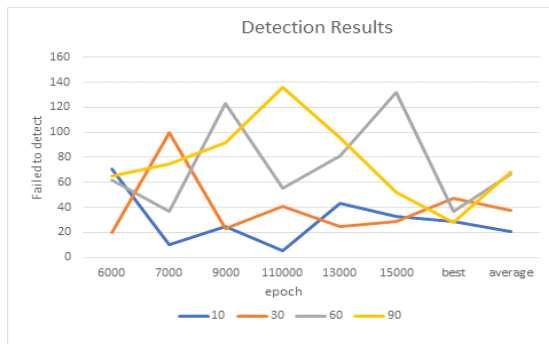


Fig. 7 Object detection result graph

분할 각도가 작을수록 분할한 영상 데이터셋의 영상 개수도 많고 객체의 여러 면을 가지고 있다. 실험 결과 작은 각도로 분할한 데이터셋을 사용할수록 더 높은 검출률을 보였다.

추가적으로 논문에서 제시한 자동 데이터셋 구축 방법과 수동적으로 구축한 데이터셋 간의 성능 분석을 위해 이동하는 사람 데이터셋을 직접 라벨링하여 구축하였다. 이동하는 사람 데이터셋은 5~10미터 거리에서 걸어서 이동하는 사람의 동영상을 촬영 후 프레임별로 분

할한 뒤 직접 라벨링하여 1500장의 데이터셋을 직접 구축하였다. 이동하는 사람의 데이터셋으로 학습한 뒤 앞선 실험에 사용한 이동하는 동영상 내 객체 추적 결과 객체 탐지 실패 수는 0으로 검출률이 100%가 나왔다. 사람의 경우 도보로 이동할 경우팔, 다리 등의 신체 각 부분의 형태가 지속적으로 변화하기 때문에 이동하는 사람의 특징을 많이 포함하고 있는 이동 데이터셋의 결과가 더 좋게 나온 것으로 보인다.

이동 시 형상의 변화 정도에 따른 성능 변화를 분석하기 위해 이동 시 형상의 변화가 적은 객체인 자동차 데이터셋을 구축하여 2차 추적 실험을 진행하였다. 자동차 데이터셋은 촬영 동영상을 매 프레임마다 분할하였으며 데이터셋의 개수는 약 1000장으로 구성되어있다. 실제 자동차를 회전시킬 수 있는 판을 구할 수 없기 때문에 실제 자동차와 동일한 모습을 하고 있는 자동차 모형으로 데이터셋을 구축하였으며 구축 데이터셋은 그림 8과 같다.

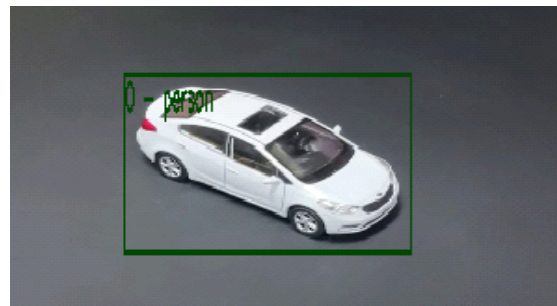


Fig. 8 Car dataset

구축한 자동차 데이터셋으로 이동하는 자동차 동영상 내에 객체 추적 실험을 진행하였으며 실험 결과 모형 자동차와 비슷한 기종의 자동차 검출률이 100%가 나왔다. 이동 시 형상의 변화가 많은 사람보다 높은 인식률을 보였다.

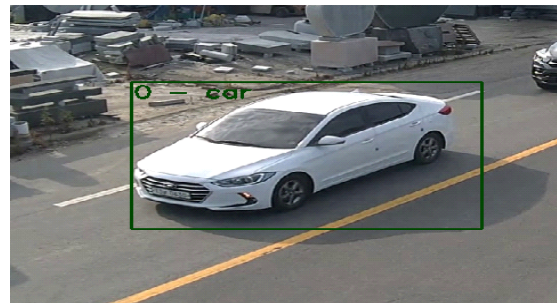


Fig. 9 Car detection results

## V. 결 론

본 논문에서는 YOLO 학습용 데이터셋을 자동으로 구축하는 방법을 제안하고 성능을 분석하였다. 본 논문에서 제시한 데이터셋 구축 방법은 사람이나 자동차 등과 같이 쉽게 데이터셋을 구할 수 있는 경우가 아닌 특정한 객체를 추적해야 하는 경우 개인이 커스텀 데이터셋을 구축하여 사용할 수 있다는 장점이 있다. 동영상 내에 객체를 추적하는 테스트를 진행한 결과 이동 시 형상이 변하는 객체의 경우 회전판을 이용한 데이터셋보다 이동하는 객체로 구축한 데이터셋이 객체의 특징점을 더 많이 담고 있어 더 높은 검출률을 보였고, 자동차는 이동 시 형태 변화가 적기 때문에 회전판을 이용하여 구축한 데이터셋도 우수한 성능을 보였다. 또한 자동차 추적 실험에서 자동차 모형으로 학습을 진행했지만 실제 자동차를 영상 내에서 추적 가능한 점을 고려하면, 고정시킬 수 없는 객체의 경우 실제와 같은 모양의 모형으로 영상을 촬영하여 데이터셋을 구축하여 학습 진행할 수 있다. 실험 결과를 비교 분석하여 객체의 이동 시 형태의 변화 정도와 목표로 하는 정확도 등의 특징을 고려하여 데이터셋을 구축하는 것이 효과적인 것이다. 쉽게 영상 데이터를 구할 수 없는 정형 객체를 동영상 내에서 추적하기 위한 상황에서는 본 논문에서 제시한 회전 원반을 이용한 동영상을 촬영하고, 촬영된 동영상으로부터 데이터셋을 자동으로 생성하는 것이 비용을 줄이고 정확도를 높일 수 있을 것으로 보인다.

### ACKNOWLEDGEMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2021-2020-0-01791) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

## REFERENCES

- [ 1 ] N. H. Thang and J. Kim, "Performance Evaluation of YOLO according to Based Feature Extraction Network in Vespa Detection," *Journal of Korea Multimedia Society*, vol. 24, no. 2, pp. 305-311, Feb. 2021.
- [ 2 ] C. H. Lee and Y. J. Jeong, "Estimation of Moving Direction of Objects for Vehicle Tracking in Underground Parking Lot," *Journal of Apiculture*, vol. 35, no. 3, pp. 149-159, Sept. 2020.
- [ 3 ] J. S. Kim, C. H. Lee, and S. C. Kwon, "Real-time Online Study Attitude Action Recognition And Dataset Construction," *The Conference of The Institute of Electronics and Information Engineers*, Jeju, pp. 1133-1135, 2021.
- [ 4 ] G. Y. Yi, Y. J. Kim, S. T. Kim, H. E. Kim, and K. G. Kim, "AR based Beverage Information Visualization and Sharing System using Deep Learning," *Journal of Korea Multimedia Society*, vol. 22, no. 3, pp. 349-356, Mar. 2019.
- [ 5 ] J. G. Choe, J. H. Lee, D. G. Kang, and S. H. Seo, "Comparison and Verification of Deep Learning Models for Automatic Recognition of Pills," *Journal of Digital Contents Society*, vol. 21, no. 3, pp. 445-452, Mar. 2020.
- [ 6 ] B. J. Kim, I. K. Kim, H. S. Lim, and J. H. Gwak, "Dataset Augmentation Technique for Crack Detection of Wood Building," *Proceedings of the Korean Society of Computer Information Conference*, Jeju, pp. 645-647, 2021.
- [ 7 ] D. H. Lee and J. H. Nang, "A Dataset and Keypoint Detection for Club Head Trajectory Estimation in 2D Golf Swing Video," in *Proceedings of The Korean Institute of Information Scientists and Engineers Conference*, Jeju, pp. 645-647, 2021.
- [ 8 ] D. W. Park, A. R. Goo, and H. J. Park, "Implementation of Region Specific Image Collection And Dataset Building Application Using Roadview API," in *Proceedings of The Korean Institute of Information Scientists and Engineers Conference*, online, pp. 1424-1426, 2020.
- [ 9 ] K. J. Kim, I. S. Jang, and K. T. Lim, "Construction of Wild-fire Smoke Data-set and Comparative Analysis of Detection Method based on Deep Neural Network," in *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, YongPyong, pp. 1172-1173, 2021.
- [ 10 ] S. Y. Lee, Y. J. Kim, Y. B. Shim, H. S. Son, and K. W. Min, "Development of SW technology for recognition, judgement and path control algorithm verification simulation and dataset generation," in *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, YongPyong, pp. 353-354, 2020.
- [ 11 ] T. J. Lee, Y. J. Kim, and H. K. Jung, "Dataset Construction and Model Learning for Manufacturing Worker Safety Management," *Journal of the Korea Institute of Information*

- and *Communication Engineering*, vol. 25, no. 7, pp. 890-895, Jul. 2021.
- [12] J. W. Lee, D. J. Lee, S. J. Jang, D. G. Choi, and J. W. Jang, "Analysis of Deep Learning Model for the Development of an Optimized Vehicle Occupancy Detection System," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 1, pp. 146-151, Jan. 2021.
- [13] G. W. Lee, H. L. Lee, and H. W. Cheong, "Object Detection of AGV in Manufacturing Plants using Deep Learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 1, pp. 36-43, Jan. 2021.
- [14] H. K. Lim, "Overview of Image-based Object Recognition AI technology for Autonomous Vehicles," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 8, pp. 1117-1123, Aug. 2021.



**김지성(Ji-Seong Kim)**

동의대학교 컴퓨터과학 이학사(2020)  
2022년 2월 동의대학교 컴퓨터공학과 공학석사 졸업 예정  
※관심분야: 딥러닝, 영상처리, IoT, 데이터베이스



**허경용(Gyeongyong Heo)**

연세대학교 전자공학과 공학석사 (1996)  
University of Florida 컴퓨터공학과 공학박사 (2009)  
동의대학교 전자공학과 교수 (2012~현재)  
※관심분야: 인공지능, 패턴인식, IoT 시스템



**장시웅(Si-Woong Jang)**

1984년 부산대학교 계산통계학과 이학사  
1993년 부산대학교 전자계산학과 이학석사  
1996년 부산대학교 전자계산학과 이학박사  
1986년 ~ 1993년 대우통신(주) 종합연구소  
2004년 ~ 2005년 University of Texas at Dallas 객원교수  
1996년 ~ 현재 동의대학교 컴퓨터공학과 교수  
※관심분야: IT융합, 차량용 네트워크, 데이터베이스