

딥러닝을 이용한 실신 감시 시스템 연구

김태경*, 조현종*,**

강원대학교 IT대학 전자공학과*, 강원대학교 BIT의료융합학**

A Research on the Fainting Monitoring System using Deep Learning

Tae-Kyeong Kim*, Hyun-Chong Cho*,**

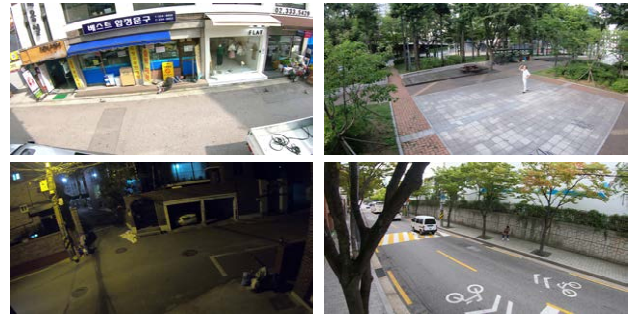
Department of Electronics Engineering, Kangwon National University*

Interdisciplinary Graduate Program in BIT Medical Convergence, Kangwon National University**

Abstract -사회적 기대수명이 증가하면서 현재의 고령사회는 초고령사회를 앞두고 있다. 또한, 성인의 3%가 겪을 수 있는 실신은 높은 비율로 노인, 여성에게 발생한다. 실신과 같은 위험한 상황임에도 불구하고 주변 사람 혹은 CCTV를 통해 감시하는 요원이 없다면 신고가 늦어질 수 있다. 본 논문에서는 YOLOv4 알고리즘을 통하여 CCTV로 받아들인 영상으로부터 객체의 실신을 감지하는 실신 감시 시스템을 구현하였다. 실신 감시 시스템은 AI Hub의 이상행동 CCTV 영상에서 실신(Swoon) 데이터 세트를 통해 구축되었다. 개발한 시스템의 목적은 일정 시간 이상 실신으로 감지되면 관련 관공서에 알려 신속한 개입을 통해 객체의 사고를 미리 방지하고자 한다.

1. 서 론

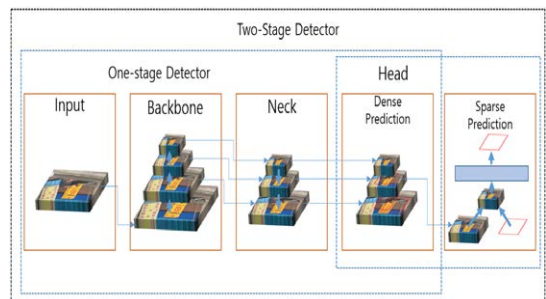
경제가 발전하여 영양과 위생상태가 좋아지고 보건, 의료가 발전하여 사회적 기대수명이 증가하였다. 2020년 기준 65세 이상 고령 인구가 현재 우리나라 전체 인구의 15.7%인 812만 5,000명이며 2025년에는 20.3%에 이르러 초고령사회로 진입할 것으로 전망된다[1]. 서울대학교병원 의학 정보에 따르면 인구의 20%에서 40%까지 일생에 한 번 실신을 경험하며 노인과 여성에게 높은 비율로 발생한다[2]. 실신으로 일어나는 사고를 방지하려면 주변 사람의 도움이나 CCTV로 지켜보는 요원의 신고가 필요하지만 그렇지 못하여 발견이 늦어진다면 큰 사고로 이어질 수 있다. 그래서 실신을 비롯한 여러 사건을 방지하기 위하여 CCTV의 설치량을 아래 그림 1과 같이 늘리고 있지만 늘어나는 CCTV를 사람이 모두 관찰할 수 없기에 본 연구에서 딥러닝을 통하여 CCTV 영상에서 실신을 감지할 수 있는 실신 감시 시스템을 제안한다[3].



〈그림 2〉 데이터 예시

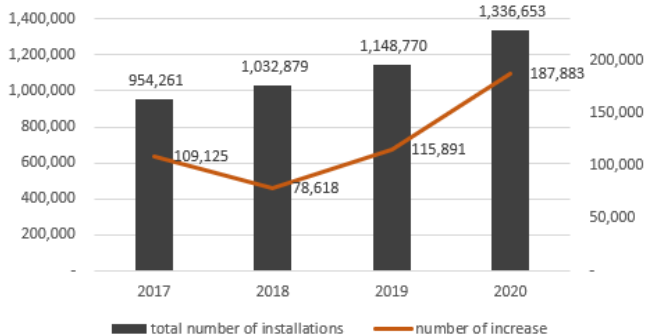
2.2 실신 감시 시스템 네트워크

본 논문에서는 Pytorch로 구현된 YOLOv4(You Only Look Once) 네트워크를 사용하여 Object detection을 통한 실시간 실신 감시 시스템을 구현하였다. YOLOv4는 빠르고 정확한 객체 검출을 위한 합성곱 신경망 네트워크로 본 모델은 한 개의 GPU에서 학습할 수 있도록 목표하였다. 아래의 그림 3과 같이 YOLOv4의 Head 부분은 YOLOv3의 아이디어를 기반으로 되어있다. 이는 Dense prediction과 Sparse prediction으로 나뉘는데 1-Stage에서는 Dense prediction을 사용하고 2-Stage에서는 Sparse prediction을 사용한다. Dense prediction은 클래스 예측과 박스 위치 교정이 통합되어 있고 Sparse prediction은 분리되어 있다. Backbone은 연산 능력이 낮은 환경에서 감당하기 힘든 연산량을 완화할 수 있는 모델인 CSPNet(Cross-Stage Partial network)기법을 Darknet53 Backbone에 적용한 CSPDarkNet53이 적용되었으며 Neck에서는 특징 맵을 동일한 $d \times d$ 블록으로 나눈 뒤 spatial pyramid를 만든 뒤 bag-of-words를 사용해 특징을 추출하는 SPM(Spatial Pyramid Matching)을 CNN에 통합하고 bag-of-word 대신 max-pooling 연산을 적용한 SPP(Spatial Pyramid Pooling)와 YOLOv3에서의 FPNet(Feature Pyramid Network) 대신 PANet(Path Aggregation Network)을 적용하였다. 그리고 본 모델은 전처리, 학습방법과 관련된 알고리즘인 BoF(Bag of Freebies)와 아키텍처, 후처리 관련된 알고리즘 BoS(Bag of Specials)를 적용하였다. BoF에는 CutOut, Mixup, Cutmix 등의 Data Augmentation과 IoU(Intersection over Union), GIoU(Generalized IoU), CIoU(Complete IoU) 등의 Loss Function이 적용되었고 BoS에서는 SPP, Mish, BN(Batch Normalization) 등이 적용되었다. 또한, Mosaic 알고리즘이 적용되었는데 이는 그림 4처럼 네 장의 이미지를 하나의 이미지로 합쳐 1장의 이미지로 4장의 이미지를 학습하는 효과를 볼 수 있다[5]. 실신 감시 시스템에선 CIoU, SPP, Mish, BN, greedy nms, leaky ReLU(Rectified Linear Unit) 등이 적용되었다. 실신 감시 시스템의 개발환경은 Window 10, NVIDIA TITAN XP, CUDA 10.2, Python 3.6, Torch 1.6, Tensorboard 1.14, OpenCV 4.1을 사용하였다.



〈그림 3〉 실신 객체 감지 네트워크 구조

CCTV installation status in public institutions

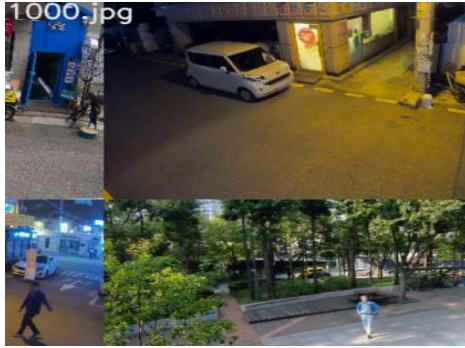


〈그림 1〉 공공기관 CCTV 총 설치 수 및 증가 수

2. 본 론

2.1 실신 데이터

본 논문에서 학습에 사용된 데이터는 한국지능정보사회진흥원이 운영하는 AI Hub에서 제공하는 '이상행동 CCTV 영상'에서 '실신(Swoon)'으로 분류된 자료를 사용하였다. 해당 데이터는 실제 상황이 아니라 연출로 제작된 영상이며 총 84시간의 5분가량으로 된 영상 912개이다. 그림 2와 같이 실데이터를 제외한 실외 CCTV 영상 11가지의 여러 상황, 장소를 바뀌거나 촬영한 영상만 사용하였다. 이때 다양한 상황에서 시스템이 활용되었을 때의 성능을 확인하기 위해 약 20개의 장소 중 9개의 장소는 학습에 사용되지 않고 검증 및 테스트에서만 사용하였다. CCTV 영상을 프레임 단위로 잘라 총 30,374장 중 18,429장을 학습데이터, 5,905장을 검증 데이터, 6,040장은 테스트 데이터로 사용하였다[4]. 학습데이터 18,429장 중 3,347장이 실신한 사람이 반드시 1명 이상이 있는 데이터이다.



〈그림 4〉 Mosaic 알고리즘 적용 예시

2.3 연구 결과

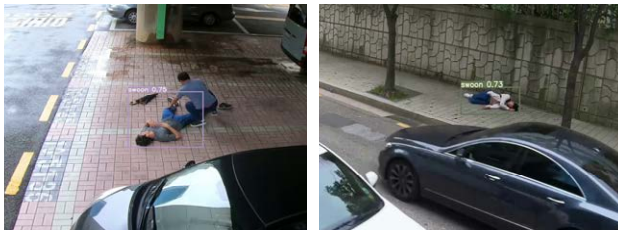
학습에 사용되지 않은 6,040장의 테스트 데이터를 사용하여 IoU Threshold를 0.5로 두고 학습 결과를 도출하였을 때 결과는 아래 표 1과 같다. Epoch 220번의 학습정밀도 Precision의 값은 0.84이며 재현율 Recall 값은 0.504이며 F1-Score는 0.63이다. F1-Score이란 Precision 값과 Recall 값의 조화 평균으로 구할 수 있다.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

그 외 지표로 mAP(mean Average Precision)를 측정하였다. AP(Average Precision)는 PR 곡선(precision-recall curve)에서 그래프 아래쪽의 면적으로 계산된다. mAP란 각 클래스당 AP를 구한 다음 평균을 낸 수치이다. mAP_0.5는 IoU Threshold를 0.5로 하여 구한 값으로 0.71로 나타났다.

〈표 1〉 실신 감시 시스템 감지 결과

Precision	0.838
Recall	0.504
F1-Score	0.6294
mAP_0.5	0.714



〈그림 5〉 실신 검출 결과



〈그림 6〉 실신 검출 오류

실신 감시 시스템을 이용하여 검출한 결과 그림 5를 보았을 때 정확히 실신을 검출된 부분도 있었지만 그림 6과 같이 시간대가 밤이거나 혹은 어느 물건 위에 실신한 경우 그림자를 실신으로 잡거나 실신을 제대로 감지하지 못하는 경우가 발생하였다. 실신 위치의 환경 변화 따라 제대로 검출하지 못하는 경우가 있었다.

3. 결 론

본 논문에서는 CCTV와 같은 영상에서 실신한 사람을 감지하여 사고를 방지하는 실신 감시 시스템을 구현하였다. 딥러닝 모델로는 Pytorch를 이용하여 구현한 YOLOv4를 사용하였으며 데이터는 AI Hub에서 제공하는 이상행동 CCTV 영상 중 실신 부분의 외부 CCTV 영상만 사용하였다. 실신 감시 시스템의 Precision 값과 Recall 값은 0.84, 0.50이고, 이를 통해 계산된 F1-Score는 0.63이고 mAP_0.5는 0.71의 성능을 확인할 수 있었다. 실제 실시간 CCTV를 연동하여 더 다양한 장소, 시간 등 CCTV 영상들을 데이터로 추가하여 학습하고 더 나아가 폭행, 침입, 강도, 납치 등 다양한 상황에서 활용된다면 주변 행인이 없거나 CCTV를 감시하고 있는 요원이 없어 신고가 늦어져 사고로 이어지는 것을 미리 방지할 수 있을 것이다. 향후 연구로 배경이 제거된 실내의 데이터를 추가하거나 색감 모델과 같은 이미지 전처리를 통하여 데이터를 증대한다면 더 정확한 결과를 얻을 수 있을 것이다.

감사의 글

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2017R1E1A1A03070297). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2018-0-01433) supervised by the IITP (Institute for Information & communications Technology Promotion).

[참 고 문 헌]

- [1] 통계청. (2020). 『2020 고령자통계』
- [2] 질병관리청 국가건강정보포털-실신. (2020). Retrived from <https://health.kdca.go.kr/healthinfo/biz/health/gnrlzHealthInfo/gnrlzHealthInfo/gnrlzHealthInfoView.do>
- [3] e-나라지표. (2021). 『공공기관 CCTV 설치 및 운영』
- [4] AI Open Innovation Hub, “이상행동 CCTV 영상”. <https://aihub.or.kr/aida/139>
- [5] Bochkovskiy, A., Wang C., Mark Liao, H. (2020). “YOLOv4: Optimal Speed and Accuracy of Object Detection”. arXiv:2004.10934v1 [cs.CV] 23 Apr 2020