CCTV와 딥러닝을 이용한 응급 상황 인식 시스템

박세준 정범진 이정준
한국산업기술대학교 컴퓨터공학과
bolby@naver.com, last9500@naver.com, jjlee@kpu.ac.kr

Emergency Situation Recognition System Using CCTV and Deep Learning

SeJun Park, Beom-jin Jeong, Jeong-joon Lee
Dept. of Computer Engineering, Korea Polytechnic University

요 약

기존의 CCTV 관리 체계는 사건·사고에 대한 신속한 조치가 불가능하고 정황 파악이나 증거자료 확보 등 사후조치의 성격이 강하다. 본 논문에서는 Mask R-CNN(Regions with CNN)을 이용하여 CCTV가 읽어 들이는 객체가 응급상황인지 판단하는 방법을 제시한다. 사람으로 인식되는 영역을 다층 퍼셉트론(MLP, Multi-Layer Perceptron)으로 학습시켜 해당 대상이 처한 상황을 인지하고 응급상황으로 인식되는 상황이 지속될 경우 관리 모니터를 통해 사용자에게 알림을 준다. 본 연구를 통해 실시간 상호작용적인 CCTV 관리 체계를 구축하여 도움이 필요한 사람의 골든타임을 놓치지 않게 될 것으로 기대한다.

1. 서 론

범죄 예방의 목적으로 CCTV의 필요성이 두드러짐에 따라 지난 5년간 전국에 설치된 CCTV의 수는 두 배 이상으로 늘어났다. 그러나 늘어만 가는 CCTV의 수에 비해 투입되어야 할 관리 인력은 턱없이 모자라고 수많은 화면을 인력만으로 모니터링하기에는 큰 무리가 있다. 그렇기에 기존의 CCTV 관리 시스템은 사건·사고가 발생함에 따라 신속한 대처가 사실상 불가능하고 피해자가 뒤늦게 발견됐을 때에야 비로소 정황파악을 위한 후속조치에 이용되는 경우가 대부분이다.

따라서 CCTV는 받아들이는 영상을 실시간으로 분석하여 현재 보여 지는 상황이 응급상황인지 아닌지 스스로 파악할 수 있는 능력이 요구된다. 응급상황을 인식하는 CCTV는 2000년도 중반부터 연구가 진행되어 왔지만 대부분의 연구가 상황 인식 부분에서 난항을 겪었다.[1] 본 논문에서는 영상처리를 기반으로 이와 같은 상황인식 능력을 갖출 수 있는 방법을 제시한다.

먼저 영상 안의 객체가 응급상황인지 아닌지 판단하기 위해서는 객체 인식이 선행되어야하기에 합성곱 신경망(CNN, Convolutional Neural Network)을 이용하되, 그 중 Mask R-CNN을 이용한다. Mask R-CNN은 기존의 R-CNN처럼 이미지의 객체를 분류하면서 각 픽셀이 해당 객체에 포함 되는지 마스킹해주는 레이어가 추가된다. 이렇게 추출된 각 객체의 실루엣을 이용하여 해당 객체가 취하고 있는 자세를 파악하는 것이 가능하다. 무슨 자세를 취하고 있는지 판단하는 것은 비교적 간단한 데이터에 활용이 가능하고 처리속도가 빠른 다층 퍼셉트론을 이용하도록 한다. 사람으로 인식되는 객체가 쓰러진 상태로 일정 시간 이상 움직임이 없다면 응급상황으로 판단하고 관리자에게 알리며 경찰에 신고하는 기능을 제공한다.

본 연구처럼 객체가 갖는 영역(Segmentation)을 지표로 삼아 학습시킨다면 응급상황 인식에 대한 성능이 충분히 개선되리라 예상한다. 또한 이와 같은 연구가 진행되어 CCTV가 능동적으로 응급상황을 판단 후 관리자에게 알리거나 직접 경찰에 신고해줄 수 있게 된다면 스스로 몸을 움직이지 못하는 응급환자의 골든타임을 확보하여 생명을 구할 수 있을 것이라 기대한다.

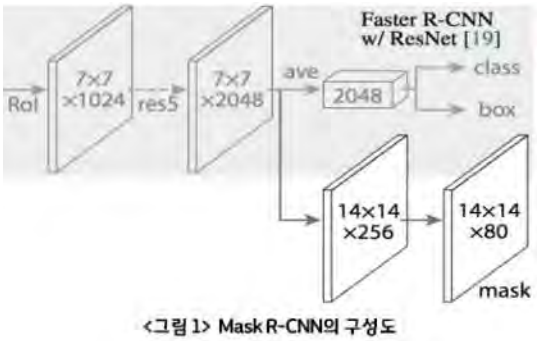
본 논문의 구성은 다음과 같다.

1. 서론
2. 관련 연구
3. 데이터 수집 및 처리
4. 응급상황 인식 구조
5. 결론 및 향후 연구

2. 관련 연구

CCTV로 응급상황을 인식하고자 CCTV 자체에 음성 센서를 부착하여 지능형 음성인식 시스템을 연구한 사례가 있다.[2] 그러나 강도에 의한 피해를 제외하고도 지병을 가져 소리 없이 쓰러지는 환자나 겨울철 길에서 잠든 취객이나 노숙자를 대처하기에는 큰 무리가 있다. 따라서 영상처리를 적극 활용하여 응급 환자에 대해 좀 더 정확한 판단을 내릴 필요가 있다.

2.1 Mask R-CNN

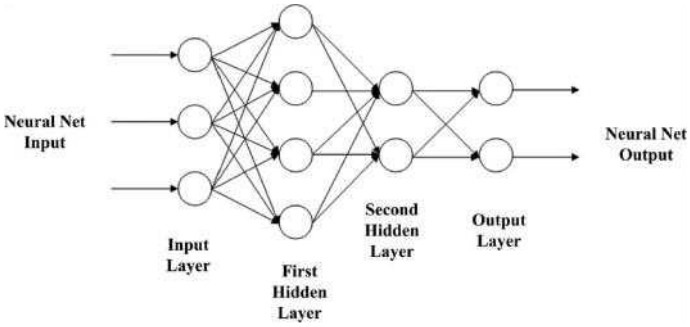


<그림 1> Mask R-CNN의 구성도

Mask R-CNN[3]은 기존 R-CNN에서 속도가 향상된 Faster R-CNN[4]에 하나의 추가적인 가지를 두어 대상 픽셀이 그 객체에 해당하는지 판단하는 binary mask를 출력하는데 사용한다. binary mask는 0과 1로 이루어진 행렬이며 그 픽셀이 객체에 해당되면 1, 아닌 경우는 0을 갖는다. 또한 Mask R-CNN 모델은 COCO 데이터셋을 기반으로 다양한 객체를 인식하므로 사람으로만 이루어진 독자적인 데이터셋을 구축하거나 객체 인식 부분에서 사람만 인식하도록 만들어줄 필요가 있다.

위와 같은 과정으로 얻어진 객체의 binary mask를 기반으로 다층 퍼셉트론을 학습시켜 상황을 인지하게 된다.

2.2 다층 퍼셉트론(Multi-Layer Perceptron)



<그림 2> 다층 퍼셉트론의 구성도

다층 퍼셉트론은 CNN처럼 복잡한 특징을 가진 사물을 분류해내기 힘들지만 간단하고 크기가 작은 데이터셋을 빠르게 분류해낸다. 다층 퍼셉트론을 활용한 대표적인 예로는 MNIST 데이터를 활용한 필기 숫자의 분류를 들 수 있다. Mask R-CNN을 통해 얻은 binary mask를 다층 퍼셉트론으로 학습시켜 어떤 형태의 마스크가 어떠한 상태인지 판단하는 역할을 한다. 객체의 binary mask들은 한 개의 채널로 이루어진 소형의 이미지들이므로 다층 퍼셉트론으로 학습시키기에 매우 적합하다.

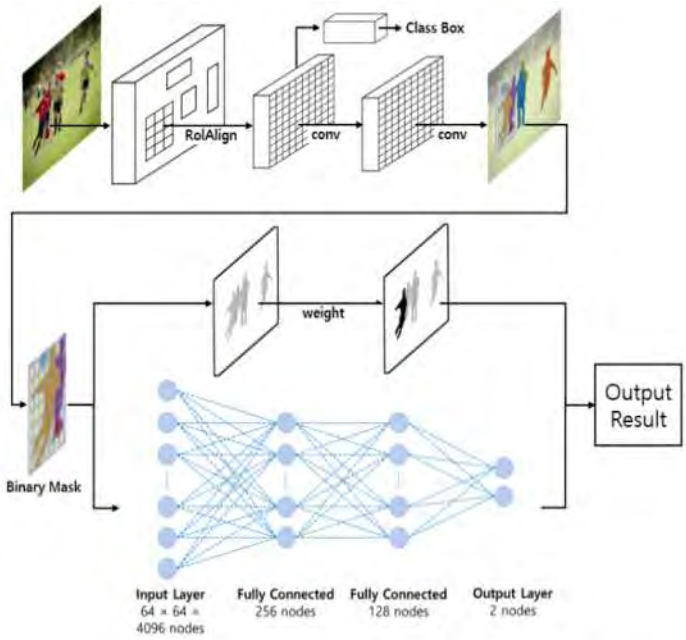
3. 데이터 수집 및 처리

<그림 3> 서있는 사람(좌)과 쓰러진 사람(우)의 형태

CCTV는 설치된 위치나 바라보는 각도에 따라서 읽어 들이는 객체의 형태가 크게 바뀌게 된다. 즉 머리 바로 위에서 수직으로 내려다보는 CCTV의 경우 쓰러진 사람을 서있는 사람으로 인식할 수 있다는 것이다. 이미지 처리를 통해 얻은 객체의 자세에 따라 응급상황 여부를 판단하기 때문에 학습 데이터의 수집은 응급상황을 판단하게 될 CCTV에서 얻는 것이 가장 바람직하다. 그러나 본 연구에서는 최대한 다양한 환경에서 시험해보기 위해 가장 보편적인 세로로 서있는 사람, 그리고 가로로 쓰러진 사람 두 가지를 학습시켰다. 초기 학습을 위한 이미지들은 CCTV가 바라보는 시야각을 생각하여 위에서 사선으로 내려다보는 각도로 다양하게 촬영하였다. 다층 퍼셉트론으로 학습시키기 위해서는 데이터의 정규화가 필요하기 때문에 입력으로 이미지들을 넣으면 사람으로 인식되는 객체들의 binary mask를 추출하여 중앙 정렬된 64x64 크기의 이미지로 정규화 시켜주는 하나의 모듈을 구축한다.

위의 과정을 통하여 <그림3>과 같은 초기 학습 데이터인 서있는 사람의 영역 마스크 200장, 쓰러진 사람의 영역 마스크 100장을 모았다. 쓰러진 자세는 0, 서있는 자세는 1로 라벨링을 해주고 학습 데이터 200장, 테스트 데이터 50장, 검증 데이터 50장으로 다층 퍼셉트론 학습을 진행하였다. 또한 테스트 데이터의 정확도가 일정 수준 이상 나올 때까지 학습시킨 후 우리가 가진 데이터에 과적합(Overfitting)된 것은 아닌지 알아보기 위해 세계 각지의 다양한 라이브 CCTV 영상을 가져와 테스트를 진행하였다. 초기 데이터로 학습 모델이 구축되면 읽어오는 CCTV 영상에서 실시간으로 서있는 사람과 쓰러진 사람을 분류하여 모델에 학습시켜 정확도를 향상시킨다.

4. 응급상황 인식 구조



<그림 4> 응급상황 인식 구조

<그림4>는 CCTV에서 받아온 이미지로부터 어떻게 응급상황을 인식하는지 나타내는 구조도이다. 실시간 영상에 이미지 처리를 적용하기 위해선 고성능의 GPU가 필요하므로 가상서버를 무료로 제공해주는 Google Colaboratory 플랫폼에서 Python과 Tensorflow를 이용하여 구현했다. 입력으로 들어온 CCTV 영상의 프레임은 제일 먼저 Mask R-CNN 모델에 적용된다. Mask R-CNN은 해당 이미지를 합성곱 신경망을 통해 특징 맵(Feature map)들을 추출해낸다. 추출된 특징 맵들은 Regional proposal 알고리즘에 의해 지역화되고 어떤 객체에

해당하는지 판별하게 된다. 이 과정에서 하나의 네트워크가 추가되어 각 픽셀이 판별된 객체에 해당하는지 나타내는 마스크를 반환받는다.

반환받은 객체의 마스크가 무슨 상태인지 판단하는 것은 사전에 학습시킨 다층 퍼셉트론 모델을 통해 진행된다. 사용한 다층 퍼셉트론 모델의 입력층 노드 수는 64x64 크기로 정규화한 이미지를 평활화(Flattening)시켜 받아들이기 때문에 4096개가 존재하고 첫 번째 은닉층은 256개, 두 번째 은닉층은 128개가 존재한다. 또한 과적합을 줄이기 위해 각 은닉층에서 30%의 Dropout을 적용했으며 실시간 영상에 적용되어야하기 때문에 연산속도가 빠른 ReLU 활성화 함수를 사용하도록 한다.

다양한 자세를 학습시킬 수도 있었지만 본 연구에서는 사람이 쓰러졌다는 응급상황과 그 외의 상황만 학습시켜 판단하고자 했기에 손실 함수로는 binary crossentropy, optimizer는 Adam Optimizer를 사용하였다.

다층 퍼셉트론 모델은 입력으로 들어간 객체의 마스크가 쓰러진 상태라면 0, 서 있는 상태라면 1로 판단하게 된다. 그러나 잠깐 넘어진 사람이나 일시적으로 생긴 객체 인식 오판을 응급상황으로 인식하면 안 되기에 응급상황 인식에 하나의 조건을 추가한다. 사람으로 인식된 객체가 쓰러진 상태로 일정 시간 움직임이 없다면 응급상황으로 인식하는 것이다.

움직임이 있는지 판단하기 위해서 객체 영역의 가중치를 담아 놓는 행렬을 추가적으로 만들었다. 이 행렬은 최소 0, 최대 20의 정수를 가지며 사람으로 인식되는 영역에 매 프레임마다 가중치를 더하는 한편 사람으로 인식되지 않는 영역은 가중치를 서서히 감소시킨다. 그리고 다음 받아오는 프레임에서 사람으로 인식되는 영역이 가중치가 최대인 영역과 90% 이상 일치하면 그 사람은 일정 시간 이상 움직이지 않았다고 판단한다.

이와 같은 조건을 종합하여 Mask R-CNN을 통해 사람으로 인식된 객체가 다층 퍼셉트론 모델에 의해 쓰러진 상태로 판단되며, 가중치에 의해 움직이지 않는다고 확인되면 비로소 응급상황으로 인식하게 된다.

<그림 5> 응급상황별 (Normal, Caution, Emergency) 인식 결과

<그림 5>는 본 논문에서 연구한 기술을 실제 CCTV 영상에 적용시킨 것이다. 서 있는 사람은 녹색, 쓰러진 사람은 황색, 쓰러진 상태로 움직이지 못하는 사람은 적색으로 나타내었다. 다층 퍼셉트론 모델에 의해 쓰러졌다고 인식되는 객체는 곧바로 황색으로 표시하고, 그러한 객체들 중 가중치가 높은 객체를 적색으로 표시하도록 만들었다.

5. 결론 및 향후 연구

본 연구에서는 객체인식이 갖는 한계점만으로 응급상황을 인식을 할 수 있는 방법을 제시했다. 응급상황인 사람과 그렇지 않은 사람은 영상 내 외형으로 봤을 때 아주 큰 차이를 갖고 있다. 즉 객체 인식을 잘 활용한다면 상황 인식 기술을 쓰지 않고도 상황 인식을 구현할 수 있다는 것이다.

그 이론에 대한 연구 결과는 매우 만족스러웠다. 다양한 실제 영상에도 적용해 봤을 때 대부분의 경우 응급상황을 확실히 분류해 내는데 성공한 것이다. 응급상황의 오판도 있었지만 CCTV는 위치와 각도에 따라 읽어오는 이미지의 왜곡이 심하기 때문에 기술이 사용될 CCTV에서 학습 데이터를 받아온다면 응급상황 인식률이 향상될 것이라고 예상한다.

그러나 단순 객체의 형태에 따라 응급상황을 판단하기 때문에 자세한 상황을 판단하지 못한다는 한계점을 가지고 있다. 객체의 영역에 따른 상황 인식을 좀 더 세밀하게 하고자 한다면 객체 인식을 할 때 하나의 사람으로 인식하는 것이 아니라 팔, 다리, 머리 등을 분할해서 인식하게 하는 기술이 선행되어야 할 것이다. 또한 다층 퍼셉트론의 은닉층 수를 늘리거나 좀 더 깊이 있는 모델을 사용할 필요가 있다.

참고문헌

[1] S.Y. Lim, J.D. Huh, “Technology Trends of Context Aware Computing Applications”, Electronics and Telecommunications Trends, vol 19, 31-40, 2004
[2] YoungIm Cho, Sungsoon jang, “Implementation of Intelligent Speech Recognition System according to CCTV Emergency Information”, Journal of Korean Institute of Intelligent systems, 415-420, 2009
[3] Kaiming He, Geogia Gkioxari Piotr Dollar, Ross Girshick, “Mask R-CNN”, Facebook AI Research(FAIR), Computer Vision and Pattern Recognition (cs.CV), 2018
[4] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Toward Real-Time Object Detection with Region Proposal, Networks”, Advances in Neural, 2015 Information Processing Systems 28(NIPS 2015), 2016