# 객체 탐지를 이용한 논문 조사

## 빅데이터응용학과 20202801 이강륜

주제: 딥러닝 기반의 무기 소지자 탐지

해당 주제 조사이유: 최근에 빈번한 칼부림 난동 및 모방 심리에 따른 길거리 범죄자 증가

해당 논문에서는 총기를 무기라고 판단하였지만 대한민국은 총기 소지가 불법이므로 칼을 총기로 생각하고 조사하였습니다.

논문에서 총기관련 사고가 전 세계적으로 빈번하게 발생하고, 범죄자들이 총기를 공공장소에 숨기는 문제부터 시작한다. 기존 연구로는 욜로 모델 및 HOG, 딥러님등 다양한 알고리즘, 기술로 모델의 성능을 향상시켰지만 기존 객체 탑지 방법은 개별 객체를 탐지하며, 권총을 든 사람 같이 복합적인 객체는 탐지하기 어렵다는 한계를 말한다. 따라서 한계를 해결하기 위해 복합적인 객체를 탐지하는 알고리즘(APDA, Armed Person Detection Algorithm)를 제안한다

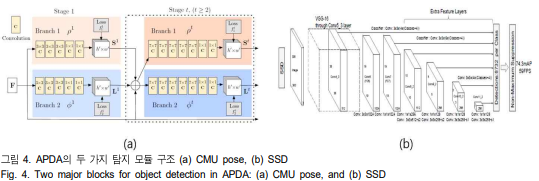
관련 이론으로는 객체 탐지, 무기 소지 탐지, 자세 추정을 소개 한다. 첫번째로 객체 탐지로는 yolo모델은 초당 45프레임으로 빠른 처리 속도를 보여주지만 작은 객체에 대해 낮은 정확도를 모인다. SSD는 순방향 신경망(FeedForward Neural Network, FFNet) 기반의 단일 심층 신경망 을 사용하며, 후보 상자를 탐지하기 위해서 픽셀이나 특징 을 다시 추출하지 않는 최초의 깊은 신경망 기반의 객체 탐지 네트워크로 깊은 쪽의 큰 객체, 얕은 쪽의 작은 객체도 탐지가 가능하다.

PASCAL VOC2007(PASCAL Visual Object Classes Challenge 2007) 에서 73.2% mAP와 초당 7프레임의 성능을 가진 FasterRCNN[11]과 63.4% mAP와 초당 45프레임의 성능을 가진YOLO에 비해 SSD는 300 x 300 입력의 경우 초당 59프레임과 74.3% mAP를 얻었고, 512 x 512 입력의 76.9% mAP 를 달성하여 탐지 정확도 및 속도가 향상됐다. ResNet은기존 CNN의 합성곱 층이 깊어질수록 성능이 저하되는 문제를 해결하기 위한 네트워크다. ResNet 네트워크는 기존의합성곱 층에 층의 입력과 층의 출력을 바로 연결하는 단축연결(Shortcut connection)을 추가하였다. 즉, 참조되지않은 함수를 학습하는 대신 층의 입력을 참조하여 잔차(Residual) 함수를 학습하는 것으로 층을 명시적으로 재구성한다. 그 결과 잔차 네트워크가 최적화하기 쉽고 상당히깊어진 층에서 정확성을 얻을 수 있음을 보였다.

무기 소지 탐지로는 Active Appearance Models(AMMs)을사용하여 칼을 탐지하는데, 이는 Principle Component Analysis(PCA)[13]를 사용하여 칼의 형태를 추정하고, Harris cor- ner-algorithm[14]을 사용하여 가장자리 특징점을 얻어내는과정으로 구성되어있다. CCTV 영상을 활용하여, 칼과 총 각각의 특성에 맞는 탐지 방법을 제안하였다. 우선 칼에 대한 탐지는 MPEC-7 특징 추출 방법을 사용하여특징 지도를 구성하였고, 이를 사용하여 머신러닝의 한종류인 SVM을 통해 칼에 대한 탐지를 진행하였다. 다음으로 총에 대한 탐지는 모폴로지(Mophology)를 이용한 배경탐색과 캐니 에지(Canny Edge) 탐색[16], PCA, 3개의 층으로 구성된 Neural Network, MPEC-7 Visual Descriptor[17] 등을 이용한 칼의 형태 탐색을 통해 진행되었다. [18]은Faster R-CNN의 가장 최신 모델인 VGG-16을 사용하여손에 쥔 총에 대한 탐지를 진행하였다. 학습 및 평가는 김건욱 외 4인: 딥러닝 기반의 무기 소지자 탐지 783 (Geonuk Kim et al.: Armed person detection using Deep Learning) IMFDB[19]를 사용하여 이루어졌으며 탐지의 실험 결과는 SVM, K-Nearest Neighbor(KNN)[20] 모델과의 비교를 통해 성능을 입증하였다.

세번째로 자세추정에는 두 단계로 구성된 하향식(Top-Down) 접근 방식을 사용한다. 첫 번째 단계에서는 Faster R-CNN 을 이용하여 사람들을 포함할 가능성이 있는 경계 상자의 크기와 위치를 예측한다. 두 번째 단계에서는 제안된 각 경 계 상자에 위치하는 사람의 키포인트를 추정한다. 각 인체 특징점에 대해 FCN인 ResNet을 사용하여 밀도가 높은 히 트 맵(Heat map) 오프셋(Offset)을 예측한다. 이후 두 가지 출력을 결합하여 높은 키포인트 예측을 얻는다. 이 알고리 즘의 정확도는 COCO(Common Objects in Context) 테스 트 셋에서 59.8% mAP를 달성하였다. 한편, CMU-Pose는 상향식(Bottom-Up) 접근 방식으로 2-branch로 구성된 다중 단계 CNN을 사용한다. 입력 영상에 대해 하나의 CNN에서 는 각 관절의 위치를 히트 맵 형태로 예측하여 그림 2의 번호로 나타내고, 다른 하나의 CNN에서는 각 신체 부위의 연관성을 표현하는 벡터 맵인 부위 선호도 필드(Part Affinity Fields, PAFs)를 예측한다. 그리고 각각 예측된 정보를결합하여 각 관절을 연관 지어 그림 2와 같이 최종 출력을뽑아낸다. 이 알고리즘의 정확도는 COCO 테스트 셋에서60.5% mAP를 달성했으며, 이전의 작업보다 향상된 정확성을 보인다. 그리고 초당 200프레임의 성능을 가져 실시간자세 추정에 적합하다.

Ⅲ. APDA(Armed Person Detection Algorithm) 인체 특징점을 기반으로, ‘권총 든 사람’을 찾기 위한APDA는 그림 3과 같다. ‘권총 든 사람’을 탐지하는 것이므 로 양 손목의 정보만 사용한다. 영상이 입력되면 CMUpose를 이용하여 그림 4(a)와 같이 입력 영상에 대한 인체 열 지도(Heat map)와 방향 지도(Vector map)를 추정하여 스켈레톤 형태의 사람을 탐지한다. 이는 Bottom-up 방식 을 사용하기 때문에 서로 다른 사람의 특정한 인체 특징 점이 근접할 때 발생하는 오류로부터 복원되는 성능이 뛰 어나다. 그리고 동일 입력 영상에 대해 SSD-MobileNet을 이용하여 권총을 탐지한다. SSD-MobileNet은 그림 4(b) 와 같이 깊이가 다른 복수의 특징 지도를 사용하기 때문 에 작은 인체 주변의 객체 탐지에 적합하고, 분리 가능한 합성곱신경망을 사용하기에 적은 매개변수를 효율적으로 학습한다.



결론에서는 APDA는 두 모듈로 구성되며, 한 모듈은 양 손목의 위치를 찾고, 다른 모듈은 권총의 경계 상자를 찾아내는데, 이러한 모듈들이 병렬적으로 작동하고 정보를 후처리 모듈에서 활용합니다. 권총을 가진 사람을 탐지하기 위해 '손에 쥔 권총' 데이터셋을 활용하여 권총의 탐지 정확도를 향상시키고, 권총과 양 손목 사이의 거리를 최소화하는 사람을 '권총 든 사람'으로 탐지한다.실험 결과를 통해 APDA의 재현율이 기존 방법 대비 46.3% 향상된 89.3%를 보이고, 정밀도는 기존 방법 대비 14.04% 향상된 32.28%를 보여 높은 성능을 갖는다는 결론을 낸다