

## 1. 개발 과제의 개요

어르신들의 건강 및 안전 케어에 대한 관심이 높아지고 있으며, 이를 위한 다양한 서비스가 검토되고 있다. 특히 어르신들의 거주 공간에서 생활 정보를 바탕으로 어르신들의 응급 안전 대응 및 건강 관리를 할 수 있다면 이는 매우 효과적인 서비스가 될 것이다. 본 개발의 목표는 디스플레이형 AI 단말기를 활용하여 65세 이상의 시니어를 대상으로 일상 생활 데이터를 모으고, 앞에서 말한 서비스 중 실시간 행동 분석을 통한 어르신들의 안전 케어 및 위험 상황 판단에 대한 스마트 서비스를 제공하는 시스템을 개발하는 것이다.

## 2. 관련 시장에 대한 분석

### 가. 시장 추이분석

실시간 행동 분석 및 객체 탐지 기술은 여러 분야에 있어 보급되고 있으며, 더 높은 성능 향상을 위한 연구가 계속되서 진행되고 있다. 일례로, 신한은행은 보이스피싱의 타깃이 된 고령 고객의 금융자산을 보호하기 위한 특단의 대책으로 은행권 처음으로 'AI 이상행동 탐지 ATM'을 도입한다고 2022년 3월 7일에 밝혔다. 또한, SKT는 AI 기반 지능형 영상분석 솔루션을 상반기 출시할 예정이며, 자체 AI칩으로 개발을 할 계획을 내비추었다. 해당 기술은 주로 공공 안전, 국방 경계, 산업 안전 등에 활용될 예정이다. 이미 상용화된 기술도 있다. 교통 분야에서는 2020년 '주차유도' 기술을 상용화했다. 영상 내 주차 공간을 인식해 사용자에게 주차 자리 빈 곳을 알려주고, 안정적인 주차를 유도하는 기술이다. 또, 운전자 모습을 영상으로 실시간 촬영해 위험 여부를 알려주는 '안전운전 보조 솔루션'도 선보였다.

앞선 딥러닝 기술들은 어르신들을 위한 서비스 제공에도 이루어지고 있다. 노년층을 위한 '시니어 맞춤형 서비스'가 다양한 산업군에서 쏟아지고 있다. 한글과컴퓨터는 시니어 케어 서비스 사업에 진출하여 '한컴 말랑말랑 행복케어'를 통해 인지훈련 치매 예방 가상현실(VR)과 상호교감이 가능한 AI 로봇 활용 프로그램도 제공할 예정이라고 밝혔다. 또, SKT는 자사의 AI 스피커인 '누구(NUGU)'를 통해 시니어 케어 서비스를 제공하고 있다. 이를 통해 거동이 불편한 어르신들이 음성을 통해 119 호출을 할 수 있는 기능과 말동무 기능 등을 선보였다. 또 다른 이동통신사 KT와 LGU+에서도 시니어 케어 사업에 적극적인 태도를 취하고 있는 것으로 나타났다.

### 나. 마케팅 전략

질병관리청에서 제공한 자료에 따르면 우리나라의 경우, 만 65세 이상인 노인의 신체 손상 원인 중 52.2%가 낙상으로 밝혀졌다. 낙상은 집 등 거주 시설에서 발생하는 경우가 48.0%로 가장 많은 것으로 밝혀졌다. 현재는 노인들의 낙상으로 인한 피해에 즉각적으로 대응하기 어려운 형편이다. 보호자나 도우미들의 도움이 있더라도 시, 공간적 제약 및 물리적 제한이 있어 24시간 종일 관리할 수는 없다.

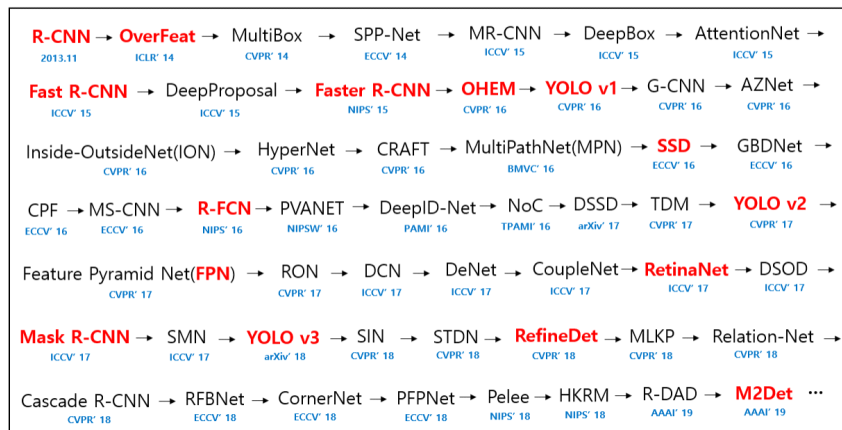
따라서, 본 프로젝트는 기존의 디스플레이형 AI 단말기에 있는 카메라를 통해 실시간으로 어르신들의 위험 상황을 판별하고, 이를 보호자에게 알람을 주는 시스템(앱)을 개발하여 사회문제를 해결하고자 한다. 여러 장비가 필요하지 않은 점과 행동분류모델이 빠르다는 점은 어르신들을 돌보는 입장인 보호자, 도우미, 지자체 쪽에 좋은 선택지가 될 것이고 이쪽을 중점으로 마케팅을 진행할 것이다.

### 3. 관련 기술의 현황

#### 가. State of art

본 프로젝트의 주요 기술인 Object detection에 대한 기술 현황을 구체적으로 서술한다. 각 기술의 자세한 설명은 본문 분량상 생략한다.

#### Object Detection: State of the Art Progress

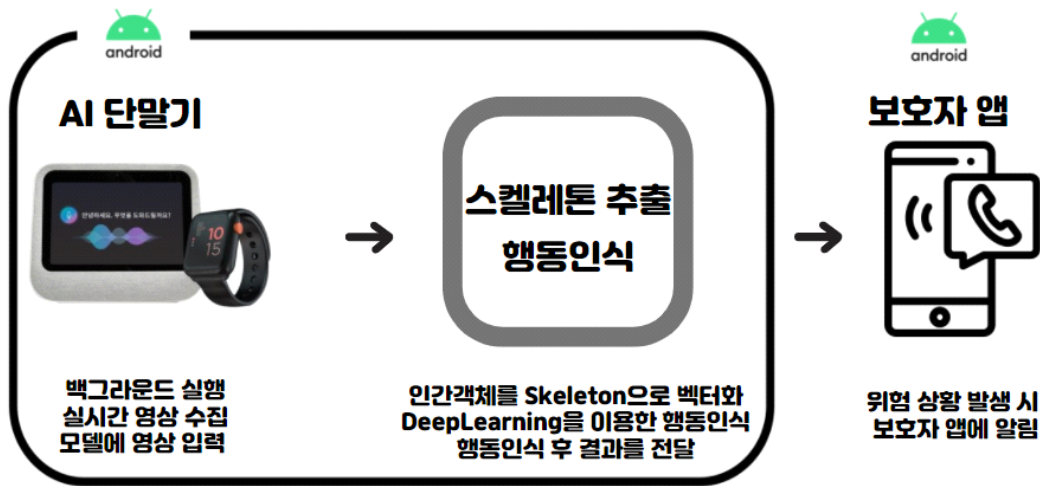


< 그림 1. Object Detection: State of the Art Progress.

source : <https://mlai.iptek.web.id/2019/01/20/object-detection-state-of-the-art-progress/> >

2013년에 최초의 딥러닝 기반 Object detection 모델인 R-CNN이 등장한 후, R-CNN 계열의 모델인 Fast R-CNN, Faster RCNN 모델들이 성과를 보여주었다. 2016년에는 YOLO와 SSD가 등장하면서 새로운 계열의 딥러닝 기반 모델이 등장하였고, 2017년에 RetinaNet이 등장하면서 높은 성능을 보여주기도 했다. 이러한 모델들이 오랜 시간 동안 각광을 받다가 2019년 후반부터 새로운 모델들이 등장하게 되었다. AutoML 기법을 도입한 EfficientNet, EfficientDet 계열이 등장하였고, NLP 분야에서 사용되는 Attention 기반의 transformer 기법을 도입한 DERT 모델 계열이 등장했다. 또한, 처리 속도는 빠르지만 정확도에서의 한계점을 지닌 YOLO 계열의 모델이 version4, 5를 거듭하여 속도뿐만 아니라 높은 정확도를 가지며 다 시금 객체 탐지 분야에서의 존재를 입증했다. 그 이후에도 성능을 발전시킨 Swin Transformer 기반 모델이 등장하였고, 객체 탐지 분야의 성능 평가인 AP 지표가 60을 넘기는 발전을 기록해오고 있다.

## 나. 기술 로드맵



< 그림 2. 기술 로드맵 >

AI 단말기 안에 설치할 Android 앱과 알람을 받을 보호자 Android 앱 2가지로 크게 나누어진다. AI 단말기 속 앱은 내장된 카메라를 실시간으로 제어하여 프레임을 추출하게 된다. 추출된 프레임을 미리 학습한 모델의 입력으로 사용한다. 이때, 모델에는 프레임 속에 존재하는 인간을 스켈레톤으로 벡터화하고, 이 벡터화된 데이터를 딥러닝 모델을 이용하여 어떤 행동을 취하고 있는지 분류하게 된다. 딥러닝 모델이 사전에 정해진 위험상황을 결과로 출력하면, 보호자 앱에 알람을 주는 기능을 통해 위험 상황을 대처하도록 설계했다.

## 다. 특허조사

### 1. 독거노인 케어 시스템 (법정상태 : 거절, 심사진행상태 : 거절결정(재심사))

급작스러운 신체적 문제 발생에 따라 독거노인이 스스로 자신의 상황을 전달할 수 없는 형편인 경우에도 독거노인의 이상징후를 스스로 판단하여 이상 결정 시 이를 관리주체에 통보할 수 있는 독거노인 케어 시스템을 제공하려는 것으로서, 본 발명에 따른 독거노인 케어 시스템은, 노인이 거주하는 세대 내에 설치되는 전기기기 스위치의 온(ON)/오프(OFF) 상태정보를 감시하고, 온/오프 상태정보와 전기기기 스위치의 신원정보를 함께 제공하는 통신부를 갖는 세대별 감시 통신장치이다.

### 2. 리모콘을 이용한 위험 관리 시스템 (공개특허공보)

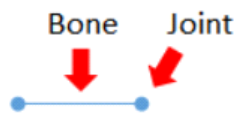
리모콘을 이용한 위험 관리 시스템이 개시된다. 본 발명의 일 실시예에 따른 리모콘을 이용한 위험 관리 시스템은 버튼의 입력정보를 전송하는 TV리모콘 및 상기 TV리모콘과 무선으로 연결되어 있으며, 상기 TV리모콘으로부터 상기 입력정보를 수신하고, 상기 스마트폰의 입력정보를 분석하여 상기 TV리모콘 사용자의 신변에 변화가 발생한 것으로 판단되는 경우 지정된 연락처로 통보하는 스마트폰을 포함한다.

## 4. 과제에서 사용할 세부기술

### 가. 2차원 골격구조 예측 모델

#### 1. OpenPose

OpenPose는 인간 자세 예측(Human Pose Estimation)의 한 분야로 오로지 카메라 한 대를 가지고 사람의 몸, 얼굴, 손가락마디를 정확하게 예측하는 것이다. 세계 최대의 '컴퓨터 비전 및 패턴 인식' 컨퍼런스인 CVPR(IEEE Conference on Computer Vision and Pattern Recognition)에서 2017년에 미국 카네기 멜론대학교에서 개발하여 발표되었다. 딥러닝 합성곱 신경망을 기반으로 영상 및 사진으로만 사람의 몸, 손, 얼굴의 특징점을 추측하는 결과를 선보였다. OpenPose의 발표 전에도 인간의 스켈레톤을 찾는 방법은 연구가 되었지만, 그 전에는 사람을 검출하고, 검출된 사람에 대한 자세를 찾도록 반복 수행하는 Top-Down 방식을 사용하였지만, OpenPose는 Bottom-Up 방식으로 반복처리 없이 수행을 하였다. 여기서 Bottom-Up 방식은 모든 사람의 관절을 추정한 후, 관절 위치를 이어 스켈레톤을 형성하는 것이다.

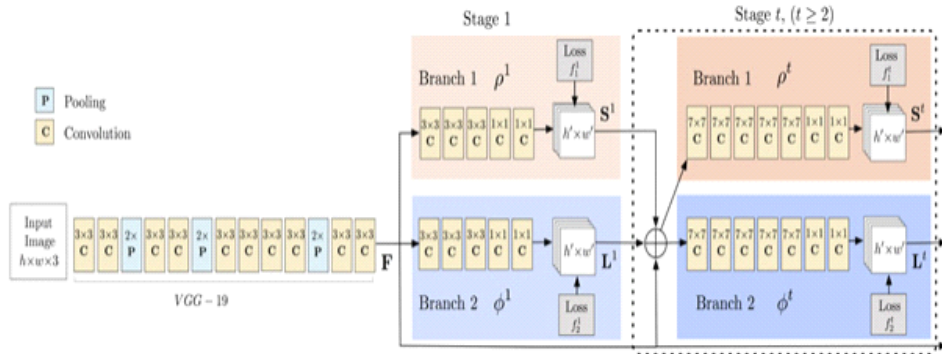


< 그림 3. OpenPose의 스켈레톤 구성 방식 >

OpenPose는 신체의 특징점(관절)을 추론한다. 이후 이 관절들을 이어 주는 방식으로 스켈레톤을 형성하게 된다. 기본 OpenPose에 있는 CNN네트워크는 Body-25, COCO, MPII가 있다. 이 세 가지 구조는 출력하는 관절의 수로 구분지을 수 있다.

BODY-25	COCO	MPII
출력 관절 25개	출력 관절 18개	출력 관절 15개

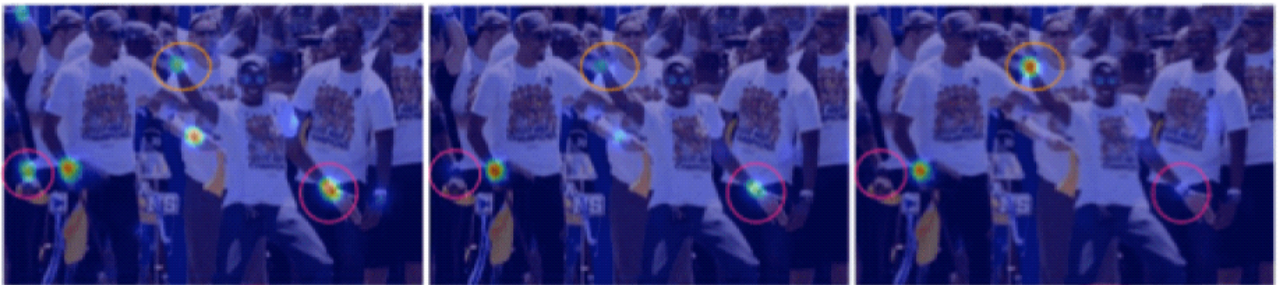
< 표 1. Structure of OpenPose >



< 그림 4. Structure of OpenPose >

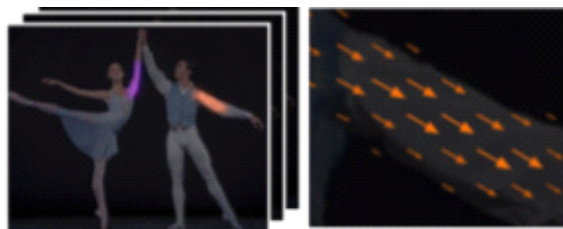
OpenPose의 전반적인 구조를 설명한다. 입력 이미지를 VGG-19 모델을 사용하여 특성맵을 산출한다. 이 특성맵 출력을 2개로 분할하여 Stage에 있는 각각의 Branch에 입력으로 넣는다. 여기서 Branch1은 confidence map을 예측하는 신경망을 의미하고, Branch2는 affinity field를 예측하는 신경망을 의미한다.

confidence map이란 이미지를 보고 사람의 몸에서 어디에 관절이 있을지 찾아내는 역할을 한다. 이 때, 특성맵을 입력해서 Branch1을 거치게 되면 사람의 관절이 있을 것이라 예측되는 곳이 heatmap형식으로 표현되어 나오게 된다. 아래의 사진은 오른쪽 손목 관절 위치를 예측한 결과물로, 오른쪽으로 갈수록 stage를 많이 거처간 결과이다.



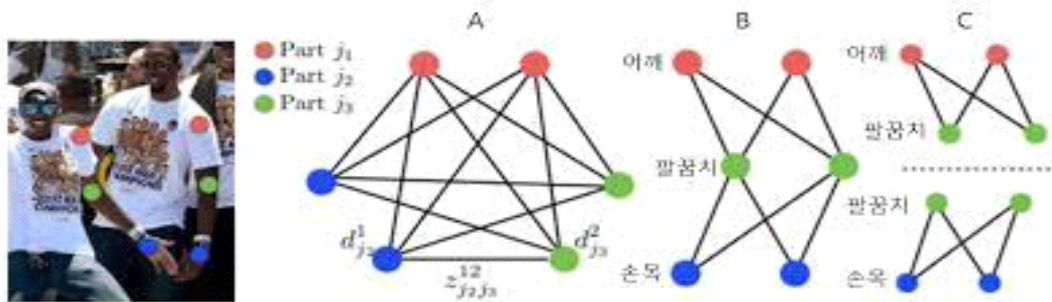
< 그림 5. 오른쪽 손목 관절 예측에 대한 Confience map >

사람의 관절 위치만 찾더라도 이미지 속의 인간의 행동을 예측할 수 없다. 이미지에 사람 한 명만 있으면 사람의 관절관계정보를 이용해서 동작을 검출할 수 있을지도 있지만, 이미지 내에 보이는 사람이 여러명 있고, 그 사람들이 서로 근접, 접촉, 맞물려있다면 관절의 연결관계를 예측하는 것은 쉬운 일이 아니다. 이러한 문제를 해결하기 위해 Affinity Fields라는 2차원 벡터공간을 도입했고 Branch2에서 이를 산출하게된다. 이 벡터공간은 (위치, 방향)을 원소로 하고 있어 2차원 벡터공간이 되는데, 각 관절과 관절 사이에 사람 몸 위에서 방향관계를 표시함으로 현재 관절 다음에 어느 방향에 있는 관절이 동일 인물의 다음 관절이 맞을 지를 예측하는데 도움을 준다.



< 그림 6. Affinity Fields >

결국, OpenPose에서 하고자 하는 것은 keypoint들 간의 정확한 매칭이다. 앞선 구조에서 confidence map을 통해 각 관절 정보가 나온 후, 관절들간의 연결관계를 이분매칭을 한다. 이분매칭은 나 자신이 속한 영역과 그렇지 않은 영역을 나누어 나와 같은 영역의 노드들끼리는 연결하지 않는 방법이다. 이분매칭을 이용하여 다른 영역에 있는 노드들끼리 모두 연결하여 모든 연결 가능성을 고려한다.



< 그림 7. Keypoint Matching >

위 사진에서 빨간색, 초록색, 파란색은 각 관절 영역을 의미하는데, 빨간색 영역 다음에는 초록색이 온다는 것을 알고있기 때문에 이 정보를 이용하여 각 노드들의 연결관계를 단순화 시킨다. 이후 각 노드들이 한 사람의 몸안에서 제대로 연결되었다는 것을 확정할 수 없기 때문에 Affinity Field를 이용하여 각 Keypoint (노드)들 간의 연결관계를 확정한다.

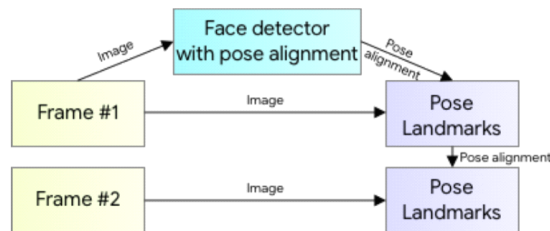
## 2. PoseDetection API

Pose Detection API는 BlazePose를 기반으로 작동하는 API이다.

BlazePose는 모바일에서도 human pose estimation에 대한 real-time inference가 가능한 모델이다. BlazePose는 heatmap과 regression을 모두 이용해 lightweight pose estimation을 한다는 장점이 있다.

기존의 자세 추정 모델은 heatmap을 이용해 각 관절에 대한 heatmap을 생성해 각 관절 좌표에 대한 offset을 수정하는 방식으로 진행되었지만, single person에 대한 실시간 자세 추론을 하기에는 모델이 너무 크다는 단점이 존재했다.

Regression 기반 모델은 컴퓨팅 리소스가 저렴하고 확장성이 크나, 관절이 맞물리는 경우에 예측이 취약하다는 단점이 있지만, BlazePose는 인코더 디코더 신경망을 이용해 모든 관절의 heatmap을 예측한 후, 다른 encoder로 regression을 이용해 각 관절의 좌표를 예측하는 구조를 가지고있으며, 추론당시에 heatmap부분을 제외해 모바일에서도 돌릴 수 있을 정도로 가볍게 만들어진다.



< 그림 8. PoseDetection 동작 방식 >

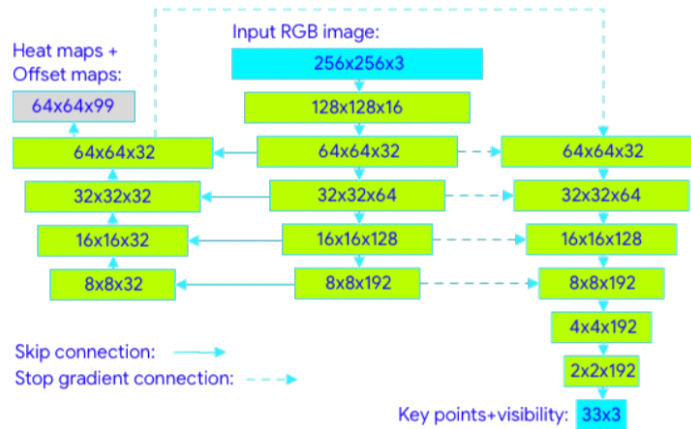
추론 파이프라인은 lightweight body pose detector와 pose tracker 순서로 구성된다. tracker는 keypoint 좌표, 사람 존재 유무, 현재 프레임의 ROI(Region Of Interesting)를 예측한다. Tracker에서 사람이 없다고 판단할 경우 다음 프레임에서 detector를 재실행한다.

최근의 객체 탐지 방법은 대부분 NMS(Non-Maximum Suppression)으로 후처리를 한다. 하지만 NMS는 keypoint가 겹치는 동작에서 오류가 발생한다는 단점이 있다.

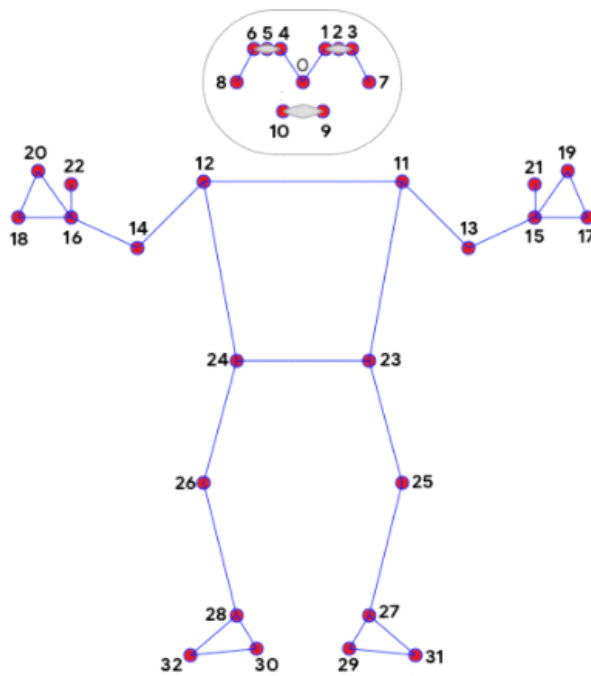
BlazePose에서는 detector의 기준을 명확한 특징이있고, 편차가 적은 얼굴로 정하여, 사람의 골반 중앙 값, 사람을 포함하는 원의 크기, 사람의 기울기 등의 alignment parameter를 예측한다.

BlazePose는 heatmap, offset, regression을 결합한 방식을 이용했다. Heatmap과 offset Loss는 학습 과정에서만 이용하고, 추론을 할 때에는, output Layer를 제거해 경량화하여 regression encoder에서 활용할 수 있도록 했다. 해당 방식은 인코더-디코더 히트맵 기반 신경망 뒤에 regression 신경망이 따라오는 구조를 쌓는 방식으로 적용했다.





< 그림 9. PoseDetection API(BlazePose)의 구조 >

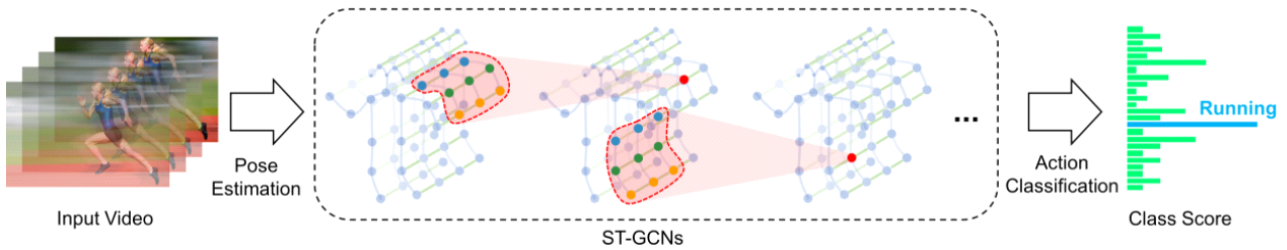


- |                    |                      |
|--------------------|----------------------|
| 0. nose            | 17. left_pinky       |
| 1. left_eye_inner  | 18. right_pinky      |
| 2. left_eye        | 19. left_index       |
| 3. left_eye_outer  | 20. right_index      |
| 4. right_eye_inner | 21. left_thumb       |
| 5. right_eye       | 22. right_thumb      |
| 6. right_eye_outer | 23. left_hip         |
| 7. left_ear        | 24. right_hip        |
| 8. right_ear       | 25. left_knee        |
| 9. mouth_left      | 26. right_knee       |
| 10. mouth_right    | 27. left_ankle       |
| 11. left_shoulder  | 28. right_ankle      |
| 12. right_shoulder | 29. left_heel        |
| 13. left_elbow     | 30. right_heel       |
| 14. right_elbow    | 31. left_foot_index  |
| 15. left_wrist     | 32. right_foot_index |
| 16. right_wrist    |                      |

< 그림 10. Pose Detection API의 출력으로 나오는 33개 관절 정보 >



## 나. ST-GCN



< 그림 11. Structure of ST-GCN >

스켈레톤 데이터를 이용한 이전의 다른 연구들은 한 프레임에서의 관절의 위치 값들을 하나의 feature vector로 나타내고 학습을 시켰다. 그러나, 이러한 방식은 행동인식에서의 관절 간의 관계성이 갖는 중요한 정보를 이용할 수 없고, 학습에 이러한 정보들이 포함되지 않았다. 이러한 문제점을 개선하고, 관절들 사이의 시간적, 공간적 관계성을 찾아내기 위해 ST-GCN 모델이 나타났다. GCN은 그래프 구조의 데이터를 입력으로 받기 때문에 GCN을 사용하기 위해서는 데이터셋을 그래프 모양으로 프로세싱 해야한다. 기존의 GCN을 사용한 행동인식의 다른 연구들은 특정 탐색 규칙 혹은 handcraft part assignment가 필요했다. 해당 모델에서는 그런 과정을 생략하고 성능을 향상 시켰을 뿐만 아니라 서로 다른 데이터셋에서 적용해 일반화 시킬 수 있도록 했다.

해당 모델은 크게 4가지 과정을 거치게 된다.

- (1) Skeleton 추출
- (2) 그래프 구조 생성
- (3) feature 추출
- (4) 분류

위에서 언급한 세부기술 (가)의 OpenPose나 PoseDetection API를 이용하여 Skeleton을 추출하게되면, 각 영상의 프레임마다 Skeleton이 추출되게 된다. 이를 그래프 형태로 프로세싱하는 과정이 (2)에 해당한다. 각각의 관절들은 노드(node)가 되고, 각각의 노드가 이어지는 부분(공간, 시간)을 엣지(edge)로 연결한다. 그 후, 총 9개의 ST-GCN 모듈을 통해 feature를 추출한다. 이 추출된 결과를 토대로 Softmax 함수를 통해 행동을 분류하게 된다.

## 5. 개발과제의 기대효과

### 가. 기술적 기대효과

기존의 연구되었던 실시간 행동 분석 관련 프로젝트에서는 CCTV나 IP 카메라에 연결된 RTSP 서버는 행동인식 서버에 실시간으로 동영상 전송하고, 촬영된 영상은 스마트폰과 같은 사용자의 애플리케이션으로 전송하여 모니터링하는 기법을 사용하고 있다. 하지만, 다수의 CCTV와 연결된 서버는 많은 영상 처리 및 저장과 관련하여 실시간 성능 및 수용력에 한계를 지니며 병목 현상이 이루어 지는 문제가 있다. 이 프로젝트에서 하는 바는 AI 단말기에 동영상 추출, 분석, 알람 기능을 앱 형식으로 구현하면서 별도의 서버 개입이 없어 병목 현상을 방지할 수 있고, 이러한 면에서 많은 사용자들을 수용할 수 있는 수용력을 가진다.

### 나. 경제적 및 사회적 파급효과

AI 단말기가 시니어들의 주거공간에 보급되고 있는 가운데, 또 다른 기기에서 처리하는 것이 아닌, 해당 단말기를 통해 서비스를 제공하므로, 서버가 필요가 없고, 추가적인 단말기 구입 비용이 절약된다. 또한, 현재 국가 주도 및 지자체에서 어르신들의 건강 관리 및 위험 상황들을 파악 및 예방, 대처하려는 노력을 하고 있다. 이에 관련지어, 어르신들의 고독사 문제를 대처하고, 일상생활에서의 이상징후 및 이상행동을 탐지, 경보를 주어 노인분들의 위험상황에 대한 빠른 대처가 가능할 것이고 지자체같이 많은 노인분들을 케어하는 입장에서 좋은 선택이 될 것이다.

## 6. 참고자료

- 노년층 증가로 '시니어 맞춤형 서비스' 출시하는 기업들  
([http://it.chosun.com/site/data/html\\_dir/2020/11/07/2020110701161.html](http://it.chosun.com/site/data/html_dir/2020/11/07/2020110701161.html))
- '초고령사회' 진입 초읽기...이통3사, '시니어 케어' 사업 확대에 적극 나서  
(<https://www.greened.kr/news/articleView.html?idxno=289627>)
- 홀로 죽음 맞는 '고독사' 3배 늘었다...43%는 '65세 이상'  
(<https://www.nocutnews.co.kr/news/5698160>)
- 통화중 ATM 인출 등 이상행동...신한은행, AI로 보이스피싱 예방  
(<https://news.mt.co.kr/mtview.php?no=2022030710540742302>)
- SKT, AI 기반 지능형 영상분석 솔루션 내년 상반기 출시...자체 AI칩으로 기술 격차 자신  
(<http://www.aitimes.com/news/articleView.html?idxno=141159>)
- 지능형 행동인식 기술을 이용한 실시간 동영상 감시 시스템 개발 (장재영 외 4인 ,The Journal of The Institute of Internet, Broadcasting and Communication (IIBC) Vol. 19, No. 2, pp.161-168, Apr. 30, 2019. pISSN 2289-0238, eISSN 2289-0246)
- Sijie Yan, Yuanjun Xiong, Dahua Lin. "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition"
- Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, Matthias Grundmann, Google Research. "BlazePose: On-device Real-time Body Pose tracking"
- Daniele Grattarola, Cesare Alippi. "Graph Neural Networks in TensorFlow and Keras with Spektral"
- Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little. "A simple yet effective baseline for 3d human pose estimation"
- 독거노인 케어 시스템  
(<https://doi.org/10.8080/1020130053844>)
- 리모콘을 이용한 위험 관리 시스템  
(<https://tb.kibo.or.kr/ktms/supplyTe/download.do?rbsIdx=110&idx=1020160018784&tag=2>)