


Pose



Overview

비디오의 인간 포즈 추정은 신체 운동 정량화, 수화 인식 및 전신 제스처 제어와 같은 다양한 응용 분야에서 중요한 역할을 한다. 예를 들어, 그것은 요가, 춤, 피트니스 응용 프로그램의 기초를 형성할 수 있다. 그것은 또한 증강 현실에서 물리적 세계 위에 디지털 콘텐츠와 정보의 오버레이를 가능하게 할 수 있다.

MediaPipe Pose는 ML 키트 포즈 감지 API를 지원하는 BlazePose 연구를 활용하여 RGB 비디오 프레임에서 33개의 3D 랜드마크와 배경 세분화 마스크를 추론하는 하이파이 바디 포즈 추적을 위한 ML 솔루션입니다. 현재의 최첨단 접근 방식은 주로 추론을 위해 강력한 데스크톱 환경에 의존하는 반면, 우리의 방법은 대부분의 최신 휴대폰, 데스크톱/노트북, 파이썬 및 웹에서도 실시간 성능을 달성합니다.

ML Pipeline

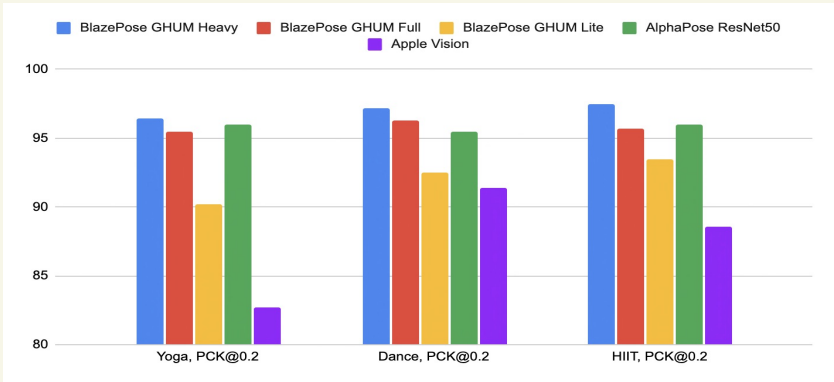
이 솔루션은 MediaPipe Hands 및 MediaPipe Face Mesh 솔루션에서 효과적인 것으로 입증된 2단계 검출기 추적기 ML 파이프라인을 활용합니다. 탐지기를 사용하여 파이프라인은 먼저 프레임 내에서 사람/포즈 관심 지역(ROI)을 찾습니다. 트래커는 이후 ROI로 자른 프레임을 입력으로 사용하여 ROI 내의 포즈 랜드마크와 세분화 마스크를 예측합니다. 비디오 사용 사례의 경우 탐지기는 필요한 경우에만 호출됩니다. 즉, 첫 번째 프레임과 트래커가 더 이상 이전 프레임에서 신체 포즈 존재를 식별할 수 없을 때만 호출됩니다. 다른 프레임의 경우 파이프라인은 이전 프레임의 포즈 랜드마크에서 ROI를 파생합니다.

파이프라인은 포즈 랜드마크 모듈의 포즈 랜드마크 서브그래프를 사용하고 전용 포즈 렌더러 서브그래프를 사용하여 렌더링하는 MediaPipe 그래프로 구현됩니다. 포즈 랜드마크 서브그래프는 내부적으로 포즈 감지 모듈의 포즈 감지 서브그래프를 사용합니다.

참고: 그래프를 시각화하려면 그래프를 복사하여 MediaPipe Visualizer에 붙여넣으세요. 관련 하위 그래프를 시각화하는 방법에 대한 자세한 내용은 시각화 문서를 참조하십시오.

Pose Estimation Quality

다른 잘 수행되는 공개적으로 이용 가능한 솔루션에 대해 모델의 품질을 평가하기 위해, 우리는 요가, 댄스 및 HIIT와 같은 다른 수직을 나타내는 세 가지 검증 데이터 세트를 사용합니다. 각 이미지에는 카메라에서 2-4미터 떨어진 한 사람만 포함되어 있습니다. 다른 솔루션과 일치하기 위해, 우리는 COCO 토폴로지의 17가지 핵심 포인트에 대해서만 평가를 수행합니다.

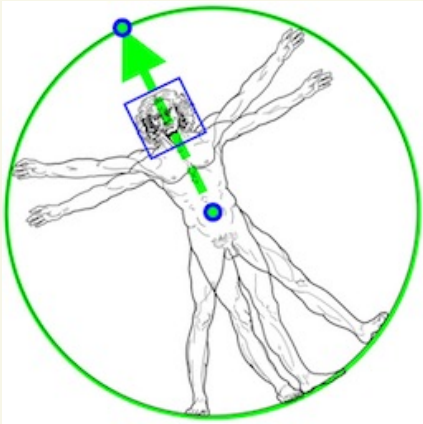


우리는 실시간 인식 사용 사례를 위해 특별히 모델을 설계했기 때문에, 대부분은 대부분의 최신 장치에서 실시간으로 작동합니다.

Models

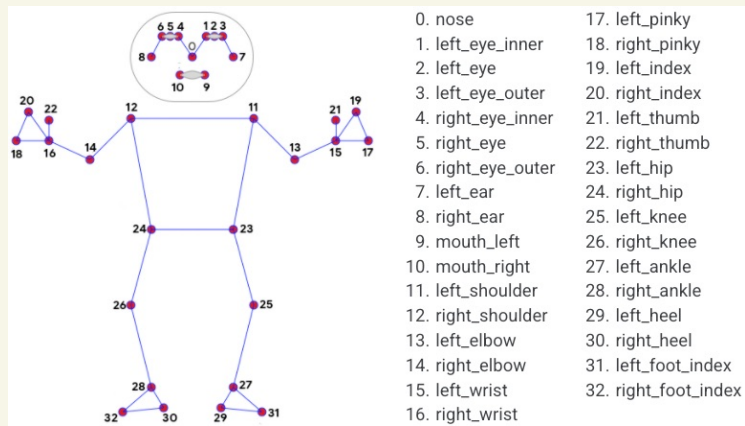
Person/pose Detection Model (BlazePose Detector)

이 탐지기는 MediaPipe Face Detection에서 사람 탐지기의 프록시로 사용되는 자체 경량 BlazeFace 모델에서 영감을 받았습니다. 그것은 인체 중심, 회전 및 스케일을 원으로 확고하게 설명하는 두 개의 추가 가상 키폰트를 명시적으로 예측한다. 레오나르도의 비트루비안 남성에서 영감을 받은 우리는 사람의 엉덩이 중간점, 전체 사람을 제한하는 원의 반경, 어깨와 엉덩이 중간점을 연결하는 선의 경사각을 예측합니다.



Pose Landmark Model (BlazePose GHUM 3D)

MediaPipe Pose의 랜드마크 모델은 33개의 포즈 랜드마크의 위치를 예측합니다(아래 그림 참조).



선택적으로, MediaPipe Pose는 2급 세분화(인간 또는 배경)로 표현되는 전신 세분화 마스크를 예측할 수 있다.

BlazePose Google AI 블로그, 이 논문, 모델 카드 및 출력 섹션에서 자세한 내용을 확인하세요.

Solution APIs

Cross-platform Configuration Options

명명 스타일과 가용성은 플랫폼/언어마다 약간 다를 수 있습니다.

STATIC_IMAGE_MODE

False로 설정하면, 솔루션은 입력 이미지를 비디오 스트림으로 취급합니다. 그것은 첫 번째 이미지에서 가장 눈에 띄는 사람을 감지하려고 노력할 것이며, 성공적인 탐지가 되면 포즈 랜드마크를 더욱 현저화합니다. 후속 이미지에서는 계산과 대기 시간을 줄이기 위해 트랙을 잃을 때까지 다른 탐지를 호출하지 않고 이러한 랜드마크를 추적합니다. True로 설정하면 사람 감지는 모든 입력 이미지를 실행하며, 정적, 관련이 없는 이미지 배치를 처리하는 데 이상적입니다. 기본값은 false입니다.

MODEL_COMPLEXITY

포즈 랜드마크 모델의 복잡성: 0, 1 또는 2. 랜드마크 정확도와 추론 대기 시간은 일반적으로 모델 복잡성에 따라 올라간다. 기본값은 1입니다.

SMOOTH_LANDMARKS

True로 설정하면, 솔루션 필터는 지터를 줄이기 위해 다른 입력 이미지에 랜드마크를 포즈를 취하지만 static_image_mode도 true로 설정하면 무시됩니다. 기본값은 true입니다.

ENABLE_SEGMENTATION

True로 설정하면, 포즈 랜드마크 외에도 솔루션은 세분화 마스크도 생성합니다. 기본값은 false입니다.

SMOOTH_SEGMENTATION

True로 설정하면, 솔루션은 다른 입력 이미지에서 세분화 마스크를 필터링하여 지터를 줄입니다. Enable_segmentation이 false이거나 static_image_mode가 true인 경우 무시됩니다. 기본값은 true입니다.

MIN_DETECTION_CONFIDENCE

탐지가 성공한 것으로 간주될 사람 감지 모델의 최소 신뢰 값([0.0, 1.0]). 기본값은 0.5입니다.

MIN_TRACKING_CONFIDENCE

포즈 랜드마크가 성공적으로 추적되는 것으로 간주될 랜드마크 추적 모델의 최소 신뢰 값([0.0, 1.0]) 또는 다음 입력 이미지에서 사람 감지가 자동으로 호출됩니다. 더 높은 값으로 설정하면 더 높은 대기 시간을 희생시키면서 솔루션의 견고성을 높일 수 있습니다. Static_image_mode가 참이면 무시되며, 사람 감지는 단순히 모든 이미지에서 실행됩니다. 기본값은 0.5입니다.

Output

명명 스타일은 플랫폼/언어마다 약간 다를 수 있습니다.

POSE_LANDMARKS

포즈 랜드마크 목록. 각 랜드마크는 다음과 같이 구성되어 있다:

X와 y: 랜드마크 좌표는 각각 이미지 너비와 높이에 의해 $[0.0, 1.0]$ 으로 정규화되었다.

Z: 엉덩이 중간 지점의 깊이가 원점인 랜드마크 깊이를 나타내며, 값이 작을수록 랜드마크가 카메라에 가까워집니다. Z의 크기는 x와 거의 같은 스케일을 사용한다.

가시성: 이미지에서 랜드마크가 보일 가능성을 나타내는 $[0.0, 1.0]$ 의 값.

POSE_WORLD_LANDMARKS

세계 좌표의 또 다른 포즈 랜드마크 목록. 각 랜드마크는 다음과 같이 구성되어 있다:

X, y 및 z: 엉덩이의 중심에 원점이 있는 미터의 실제 3D 좌표.

가시성: 해당 pose_landmarks에 정의된 것과 동일합니다.

SEGMENTATION_MASK

Enable_segmentation이 true로 설정된 경우에만 예측되는 출력 세분화 마스크. 마스크는 입력 이미지와 너비와 높이가 동일하며, 1.0과 0.0이 각각 "인간"과 "배경" 픽셀의 높은 확실성을 나타내는 $[0.0, 1.0]$ 의 값을 포함합니다. 사용 세부 사항은 아래의 플랫폼별 사용 예제를 참조하십시오.