

Openpose 정리

- **Openpose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields**

- the first real-time multiperson system to jointly detect human body, foot, hand, and facial keypoints (in total 135 keypoints) on single images(단일 이미지)

- **Abstract**

1. Realtime system & multi-persion
2. a nonparametric representation => Part Affinity Fields (PAFs)
 - (PAFs = associate body parts with individuals in the image)
3. bottom-up system
 - high accuracy and realtime performance, regardless of the number of people in the image a runtime comparison to Mask R-CNN and Alpha-Pose

- **Imporvement notice (2017 to 2019)**

1. PAF와 body part location을 모두 refine하는 것보다 PAF만 refine하는 것이 더 성능이 높음
 - runtime performance: 200% ⬆, accuracy: 7% ⬆
2. Body keypoint + foot keypoint 를 동시에 예측하는 모델을 개발
 - 독립적으로 예측하는 것보다 inference time을 줄이고 accuracy는 유지
3. vehicle keypoint estimation에도 적용 가능
 - 일반성 입증
4. open-source로 배포 (body, foot, hand, facial keypoints)

PAF 정리 (Abstract – 1, 2)

- PAF = 네트워크의 첫번째 스테이지 예측 대상 & 이분그래프의 가중치의 재료
- PAF의 효과
 - 혼잡한 이미지에서 정확도를 향상
 - 모든 관절을 연결할 때, PAF connections가 중복됨. (귀-어깨, 어깨-손목) 특히, 이러한 중복성은 혼잡한 이미지에서 정확도를 향상시킴.
 - 가능한 쌍을 **PAF 기준으로 정렬**했기 때문에, 다른 사람과 **연결할 가능성이 줄어듦**.
 - 만약 연결이 이미 다른 사람에게 할당된 두 신체 부위를 연결하려고 한다면, 알고리즘은 이것이 더 높은 신뢰도의 PAF 연결과 모순된다는 것을 인식하고, 현재 연결은 무시된다.
 - 저자의 이전 논문, multi-person parsing algorithm에서는 root component 부터 시작
- PAF 핵심 기여 (PAF 개념 다시 정리)
 - bottom-up 방식의 약점은 전역 컨텍스트 정보를 잘 활용하지 못한다는 것에 있음
 - -> 이 문제를 해결하기 위해 PAF를 도입(제안)
 - PAF: 이미지 도메인에서 팔다리의 위치와 방향을 인코딩하는 역할을 하는 2D 벡터
 - 두 개의 관절이 연결되어 있는지 여부와, 해당 방향 정보를 알 수 있음

PAF – 네트워크 성능 관련

- PAF와 body part location을 모두 refine하는 것보다 PAF만 refine하는 것이 runtime performance, accuracy가 올라감
 - 이전 버전: PAF와 CM 네트워크는 병렬적으로 학습함 -> 현재 버전: PAF 먼저 학습 후, CM 학습 (직렬 구조)
 - PAF의 정확도를 최대화하는 것은 중요한 반면, CM의 정확도는 그다지 중요하지 않음을 실험적으로 증명
- ☞ 그 결과 네트워크 구조가 변경됨 (최종 네트워크 구조는 5PAF + 1CM을 채택함)
- Results on the COCO Keypoints Challenge
 - PAF는 반복해서 refine하면 성능이 상승 (1CM)
 - CM을 늘리는 것은 성능에 많은 영향을 끼치지 못함
 - PAF를 먼저 refine해야 성능이 상승 (3CM - 3PAF)
- Results on the MPII Multi-Person Dataset
 - PAF vs One/Two midpoint : PAF 도입한 경우가 성능이 높음 { 정확도:)2.9% higher than one-midpoint) > One-midpoint(2.3% higher than two-midpoint)}
 - 스테이지 별 precision 변화: monotonically 단조롭게 증가함

Method	AP	AP ⁵⁰	AP ⁷⁵	AP ^M	AP ^L	Stages
5 PAF - 1 CM	65.3	85.2	71.3	62.2	70.7	6
4 PAF - 2 CM	65.2	85.3	71.4	62.3	70.1	6
3 PAF - 3 CM	65.0	85.1	71.2	62.4	69.4	6
4 PAF - 1 CM	64.8	85.3	70.9	61.9	69.6	5
3 PAF - 1 CM	64.6	84.8	70.6	61.8	69.5	4
3 CM - 3 PAF	61.0	83.9	65.7	58.5	65.3	6

TABLE 5: Self-comparison experiments on the COCO validation set. CM refers to confidence map, while the numbers express the number of estimation stages for PAF and CM. Stages refers to the number of PAF and CM stages. Reducing the number of stages increases the runtime performance.

COCO

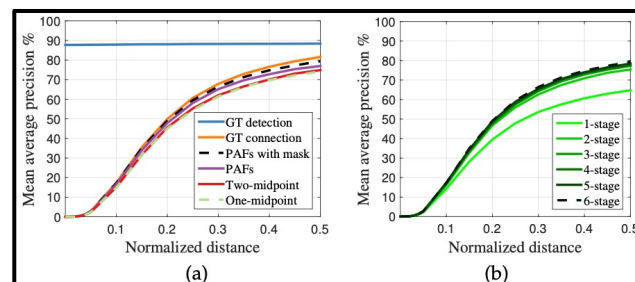
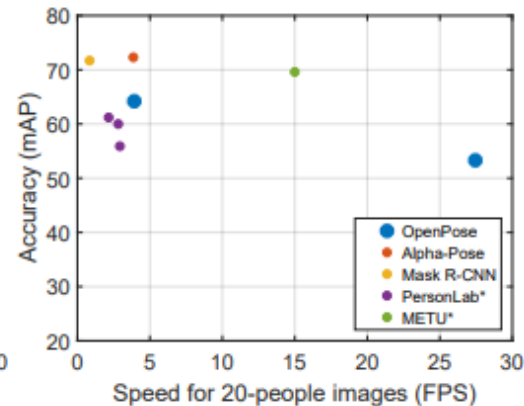
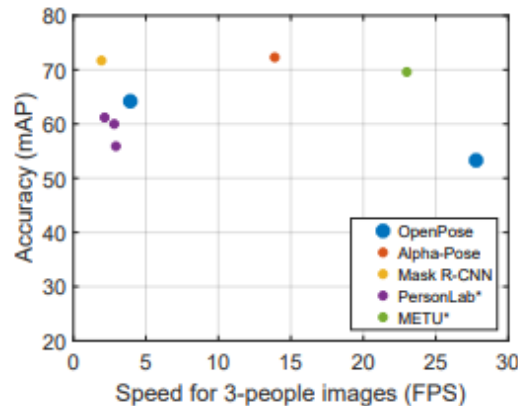
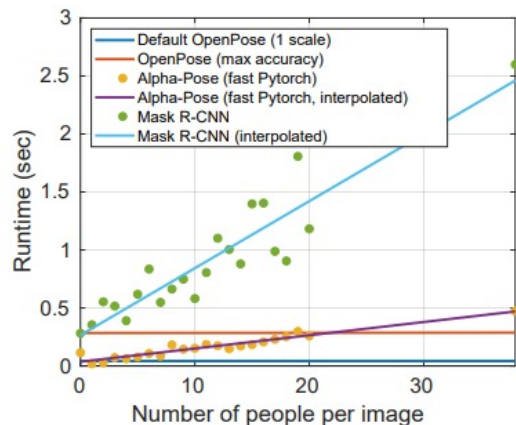


Fig. 11: mAP curves over different PCKh thresholds on MPII validation set. (a) mAP curves of self-comparison experiments. (b) mAP curves of PAFs across stages.

MPII

실시간이 가능한 이유

- 사람 수에 상관 없이 realtime으로 high accuracy (Realtime, Multi-Person 환경에서도 우수한 성능(다른 모델 보다 유리))
 - Top-down 방식: 사람의 수에 비례해서 inference time 선형적으로 증가
 - 단점 1) detection 실패 시 예측 못 함 (robust) 단점 2) inference time 은 사람 수의 비례 (양의 상관관계)
 - Bottom-up 방식: inference time invariant
- Trade-off between Speed and Accuracy
 - 사실 MaskR-CNN, Alpha-Pose 등 다른 방법들에 비해 Openpose의 Accuracy가 엄청나게 좋지 않은
 - 하지만, 정확도가 높은 위의 모델들은 이미지에 20명 이상의 사람이 있는 경우, 너무 느려짐 (효율성 Bad)



Openpose System

- **System**

- 다양한 플랫폼에서 실행 가능: Ubuntu, Windows, Mac OS X 및 임베디드 시스템(예: Nvidia Tegra TX2) 등
- 다양한 하드웨어를 지원: CUDA GPU, OpenCL GPU, CPU 전용 장치 등
- 활용분야: many vision and robotics topics (person reidentification, GAN-based video retargeting of human faces and bodies, Human-Computer Interaction, 3D pose estimation, and 3D human mesh model generation).

- **사용자 지정 tunig**

- (이미지, 비디오, 웹캠 및 IP 카메라) 스트리밍 사이에서 입력 선택, 결과를 표시 여부, 디스크에 저장 여부, 각 검출기 (몸, 발, 얼굴, 손)의 활성화 여부, 픽셀 좌표 정규화를 활성화 여부, 사용할 GPU 수를 제어, 더 빠른 처리를 위한 프레임을 건너뛰기 여부

- **** 성능이 잘 안 나오는 경우**

- 일반적이지 않은 자세나 위아래가 뒤집힐 경우
- Body occlusion
- 여러 사람이 겹칠 경우