

AlPhaPose

2018250033 유규빈
2018250056 함상진

목차

01 Introduction

02 Related Work

03 Whole-Body Multi Person Pose Estimation

PART 1.

Introduction

1

빠르고 정확한 위치 파악을 위한 대칭 적분 키포인트 회귀(SIKR)

2

인간 중복 탐지를 제거하기 위한 매개 변수 포즈 비최대 억제(P-NMS)

3

포즈 추정 및 추적을 동시에 하기 위한 포즈 인식 아이덴티티 임베딩

PGPG(Part-Guided Proposal Generator)와 다중 도메인 지식 증류 방법 사용

자세 추정 데이터셋으로 테스트 결과
속도와 정확도 부분에서 최신 기술들보다 개선



먼저 경계 상자를 감지한 다음 각 상자 내의 포즈를
추정하는 하향식(Top-Down) 프레임워크를 따름



1

객체 검출을 실패할 경우 포즈 추정기가 인간 포즈를 추정할 수 없음

2

그래서 정확성이 높은 객체 검출을 사용하지만 두 단계의 추론을 느리게 함



객체 탐지 누락 문제를 완화하기 위해 탐지 신뢰도와 NMS 임계값을 낮춰
후속 포즈 추정을 위한 더 많은 후보를 제공함

AlphaPose에서 다단계 동시 파이프라인을 설계해 실시간으로 실행 가능

Whole-Body Estimation

001 >> 양자화 오류

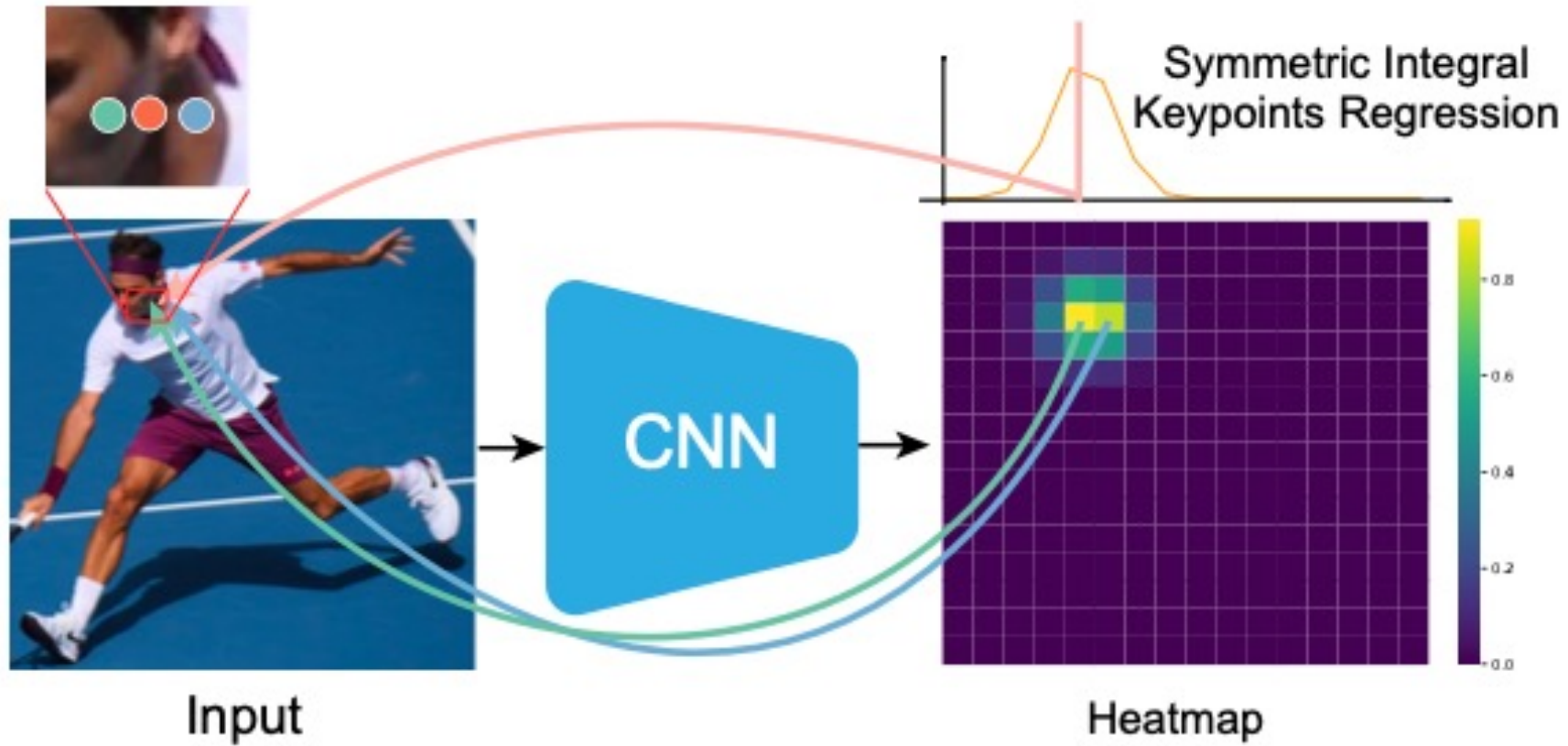
하향식, 상향식 프레임워크에서 키포인트에 대해 현재 가장 많이 사용되는 표현은 히트맵
하지만 계산 리소스의 한계로 인해 히트맵 크기는 일반적으로 입력 이미지의 1/4임
하지만 신체, 얼굴 및 손의 키포인트를 동시에 1/4로 국소화하는 경우 히트맵이 이산적으로 표현할 때
히트맵의 인접 그리드 모두 정확한 위치를 놓치는 양자화 오류가 일어날 수 있음

002 >> 기존 오류 해결 방안

하위 네트워크를 추가하거나 ROI-Align 을 선택하여 특성 맵을 확대함
그러나 특히 다중 인원 검출할 때 두 가지 방법 모두 계산이 어려움

003 >> 오류 해결 방안 제시

양자화 오류를 제거하면서 히트맵 표현과 동등한 정확도를 가질 수 있는 회귀 방법인
대칭 적분 키포인트 회귀 방법을 제안



입력 영상이 들어오면 히트맵이 이산적인 표현으로 변환

부족한 훈련 데이터셋

1

일반 자세 추정과는 달리 데이터셋이 하나뿐인 전신 자세 추정

2


기존 데이터셋에 없는 새로운 관절 요소가 추가로 추정이 필요

3

기존 데이터셋에 주석을 달아 Halpe 라는 이름의 데이터셋 생성


신체 부위들의 데이터셋 훈련 데이터들을 통합하는
다중 도메인 지식 증류법을 사용

훈련 샘플을 증가시키기 위한
Part-guided human proposal generator (PGPG) 사용



하향식에서 전신 포즈 추정과 사람 식별을 동시에 하는
포즈 인식 아이덴티티 임베딩을 도입

포즈 추정기에 사람 재식별 브랜치가 부착되어 있으며
포즈 추정 및 사람 식별을 동시에 수행



PART 2.

Related Work

2.1 Multi Person Pose Estimation

001 >> **OpenPose**

Part Affinity Fields (PAFs)를 도입하여 개인과 신체 부위 간 연관 점수를 인코딩
이를 이용하여 연관 매칭 문제를 이분 매칭 하위 문제 집합으로 분해하여 해결한 방법

002 >> **Associative embedding**

감지된 신체 부위에 대해 속한 개인을 나타내는 식별 태그를 학습해 객체를 구별한 방법

003 >> **DeeperCut**

ResNet에 기반해 DeepCut보다 더 나은 신체 부위 검출과 최적화 향상 전략을 사용

2.1 Multi Person Pose Estimation

001 >> **마스크 R-CNN**

ROI Align 후 기존 경계 상자 인식과 동시에 포즈 추정을 해 엔드 투 엔드 훈련을 가능하게 함
으로써 Faster R-CNN을 확장한 방법

002 >> **PandaNet**

앵커(Anchor) 기반 방법을 제안하여, 다중 인물 3D 포즈 추정을 한 번에 처리하고 높은 효율성을 달성한 방법

003 >> **Sparsely-Labeled Videos**

이미지에서 비디오로 확장하여, 일부 데이터만 레이블링된 비디오에서 포즈 워핑(Pose warping)을 학습하는 방법을 제안한 방법

2.1 Multi Person Pose Estimation

시간이 많이 걸리는 2단계를 동시에 처리하고
빠른 추론을 가능하게 하는 다단계 파이프라인 개발

2.2 Whole-Body Keypoint Localization

001 >> **OpenPose**

OpenPose는 계단식 방법을 개발
먼저 PAF를 사용하여 신체 키포인트를 감지한 다음
얼굴 경계와 손 키포인트를 추정하기 위해 두 개의 별도 네트워크를 채택
이러한 설계는 시간을 비효율적으로 만들고 추가 계산 리소스를 소비

002 >> **Single-network whole-body pose estimation**

전체 신체 키포인트를 추정하기 위한 단일 딥러닝 모델 제안
그러나 출력 해상도가 제한되어 얼굴 및 손과 같은 미세한 부분에서 성능이 저하

003 >> **Whole-body human pose estimation in the wild**

ROIAlign을 사용하여 특징 맵에서 손과 얼굴 영역을 자르고
크기를 조정한 후 키포인트를 예측하는 ZoomNet을 제안

2.2 Whole-Body Keypoint Localization

soft-argmax 표현이 전신 자세 추정에 더 적합하다는
주장을 제시하고 더 높은 정확도를 제공하는
개선된 soft-argmax 버전을 개발

2.3 Integral Keypoints Localization

비대칭 기울기 문제

크기 종속적 키포인트 점수

2.4 Multi Person Pose Tracking

다인 자세 추적은 영상의 다인 자세 추정에서 확장되어
시간이 지남에 따라
각 예측 키폰트에 대응하는 아이덴티티를 제공

2.4 Multi Person Pose Tracking

001 >> 비 온라인 방식

bottom-up 포즈 추정 방법에 의해 감지된 키포인트를 사용하여 시간 및 공간 그래프를 구성
그러나 시간 및 공간 그래프의 전제조건은 그래프 컷 최적화가 온라인 방식으로 실행되는 것을 방지하므로
시간이 많이 걸리고 메모리가 비효율적

002 >> 온라인 방식

단일 프레임을 입력한 다음 아이덴티티 매칭을 위해, designed pose flow, GCN, optical flow, transformation를 사용
온라인 이미지 스트림이 불안정하거나 인간이 빠르게 이동할 때 만족하지 못할 수 있는 포즈의 공간 연속성에만 의존

003 >> re-ID feature

기존 연구들의 문제를 해결하기 위해 인간 re-ID feature을 채택
잠재적인 배경 노이즈를 피하기 위해 pose-guided re-ID feature 추출을 설계
또한 상자, 포즈 및 re-ID features을 동시에 활용할 수 있는 multi-stage 정보 병합 방법을 설계

PART 3.

Whole-Body Multi Person Pose Estimation

3.1 Symmetric Integral Keypoints Regression (SIKR)

1

비 대칭 기울기 문제 해결(Asymmetric gradient problem)

2

관절 크기에 따른 키포인트 점수 문제 해결
(Size-dependent Keypoint Scoring Problem)

3.1.1 Asymmetric gradient problem

001 >> soft-argmax $\hat{\mu} = \sum x \cdot p_x,$

적분 회귀라고도 하며 미분 가능

따라서 heatmap 기반 접근 방식을 회귀 기반 접근 방식으로 전환하고 end-to-end 훈련을 가능하게 함

여기서 x 는 각 픽셀의 좌표, p_x 는 정규화 후 heatmap에서의 픽셀 가능성

002 >> $\frac{\partial \mathcal{L}_{reg}}{\partial p_x} = x \cdot \text{sgn}(\hat{\mu} - \mu).$ $\mathcal{L}_{reg} = \|\mu - \hat{\mu}\|_1.$

각 픽셀의 기울기 값은 위와 같이 공식화 됨

훈련 중에 손실 함수를 적용하여 예측된 관절 $\hat{\mu}$ 와 실측 위치 μ 사이의 ℓ_1 norm 최소화.

003 >> 비대칭성

그레디언트 진폭(amplitude)은 비대칭

그레디언트의 절대값은 실측 값에 대한 상대적 위치 대신 픽셀의 절대 위치(즉, x)에 의해 결정

따라서 동일한 거리 오차가 있을 경우 키포인트가 다른 위치에 있을 때 기울기가 다르게 나타남.

이러한 비대칭성은 CNN 네트워크의 변환 불변성을 깨서 성능 저하로 이어짐

3.1.1 Asymmetric gradient problem

004 >> Amplitude Symmetric Gradient

학습 효율성 향상을 위해 역방향(backward) 전파에서 진폭 대칭 기울기(ASG) 함수를 제안
이는 실제 기울기에 근사한 값

005 >> $$\delta_{ASG} = A_{grad} \cdot \text{sgn}(x - \hat{\mu}) \cdot \text{sgn}(\hat{\mu} - \mu),$$

여기서 A_{grad} 는 기울기의 진폭
학습 과정에서 이 대칭 기울기 분포는 heat map의 이점을 더 잘 활용
보다 직접적인 방식으로 실측 위치를 근사화

3.1.2 Size – dependent Keypoints Scoring problem

001 >> 기존 정규화(소프트 맥스)의 문제점

soft-Argmax를 수행하기 전에 예측된 heatmap의 원소 합을 $\sum x = 1$ 로 정규화
다인의 경우 관절 위치뿐만 아니라 자세 NMS 및 mAP 계산을 위한 관절 신뢰도가 필요하기 때문에 소프트 맥스 연산을 사용
기존 방법에서는 열지도의 최대값이 관절 신뢰도로 간주됐는데, 이는 크기에 따라 다르고 정확하지 않았음

002 >> 1단계 정규화의 문제점

소프트 맥스와 같은 1단계 정규화를 채택하면 열지도의 최대값은 분포의 규모에 반비례하며,
이는 신체 관절의 예상 크기에 크게 의존함.
따라서 큰 관절(예: 왼쪽 엉덩이)은 작은 관절(예: 코)보다 작은 신뢰값을 생성하여 예측된 신뢰값의 신뢰성을 손상

003 >> Two-step Heatmap Normalization

신뢰값 예측과 적분 회귀를 분리하기 위해 2단계 heatmap 정규화 방법을 제안

3.1.2 Size – dependent Keypoints Scoring problem

004 >> $c_x = \text{sigmoid}(z_x),$

첫 번째 단계에서 요소별 정규화를 수행하여 confidence heatmap(신뢰도 히트맵) C 를 생성
 z_x 는 위치 x 의 정규화되지 않은 로짓(logit) 값,
 c_x 는 위치 x 의 confidence heatmap 값.

005 >> $conf = \max(\mathbf{C}).$

관절 신뢰도는 heatmap의 최대값으로 표시 가능
정규화의 첫 단계에 원소 별 시그모이드 연산을 하고 C 의 합을 1로 강제하지 않기 때문에
 C 의 최대값은 관절의 크기에 영향을 받지 않음
이러한 방식으로 예측된 joint 신뢰도는 예측된 위치에만 관련이 있음

006 >> $p_x = \frac{c_x}{\sum \mathbf{C}}.$

두 번째 단계에서 확률 히트맵(probability heatmap) P 을 생성하기 위한 전역 정규화 수행
확률 heatmap P 의 원소 합은 1이며, 이는 예측된 관절 위치 μ^* 이 heatmap 영역 내에 있음을 보장하고 훈련 과정을 안정화
첫 번째 단계를 통해 관절 신뢰도를 얻고 두 번째 단계에서 생성된 열 지도에서 관절 위치를 얻음

3.2 Multi-Domain Knowledge Distillation

001 >> 네트워크 성능

네트워크의 성능은 추가적인 훈련 데이터로부터 더 많은 이점을 얻을 수 있음

300Wface, FreiHand 및 InterHand의 세 가지 추가 데이터 세트 채택

이러한 데이터 세트를 결합하여 네트워크는 일상 이미지에 대한 얼굴 및 손 키포인트를 정확하게 예측 가능

002 >> 훈련 배치 구성

주석이 달린 데이터 세트에서 1/3 샘플링, coco 전체 본체에서 1/3 샘플링

나머지는 300Wface와 FreiHand에서 동일하게 샘플링

각 샘플에 대해 데이터 세트별 증식 적용

003 >> pose-guided proposal generator의 확장 필요성

도메인별 데이터 세트는 정확한 intermediate supervision을 제공할 수 있지만, 데이터 distribution은 실제 이미지와 상당히 다름

이 문제를 해결하기 위해 pose-guided proposal generator(포즈 유도 제안 생성기)를 전신 시나리오로 확장하고 통합된 방식으로 데이터 증식을 수행

3.3 Part-Guided Proposal Generator

001 >> 데이터 증식의 필요성

2단계 포즈 추정의 경우, 인간 검출기에 의해 생성된 인간 `proposal`은 일반적으로 실측 인간상자와 다른 데이터 분포를 생성
한편, 얼굴과 손의 공간 분포는 일상 전신 이미지와 데이터 세트의 부분 전용 이미지 사이에서도 다름
훈련 중에 적절한 데이터 증식이 없으면 포즈 추정기는 탐지된 인간에 대한 테스트 단계에서 제대로 작동하지 않을 수 있음

002 >> part-guided proposal generator

인간 검출기의 출력과 유사한 분포를 가진 훈련 샘플을 생성하기 위해 `part-guided proposal generator`를 제안
많은 경계 상자가 있는 다양한 신체 부위에 대해 `proposal` 생성기는 인간 검출기의 출력 분포와 일치하는 새 상자를 생산

3.3 Part-Guided Proposal Generator

005 >> 오프셋 분포 모델링

각파트에대한 실측 경계상자가이미있으므로,
이문제를 검출된 경계상자와 대응되는 실측 경계 상자 사이의 상대적인 오프셋 분포를 모델링하는 문제로 단순화.
이 상대적인 오프셋 분포는 각 부위마다 다르게 변할 수 있음

$$P(\delta x_{min}, \delta x_{max}, \delta y_{min}, \delta y_{max} | p)$$

$$\begin{aligned} \delta x_{min} &= \frac{x_{min}^{detect} - x_{min}^{gt}}{x_{max}^{gt} - x_{min}^{gt}}, \\ \delta x_{max} &= \frac{x_{max}^{detect} - x_{max}^{gt}}{x_{max}^{gt} - x_{min}^{gt}}, \end{aligned}$$

$$\delta y_{min}, \delta y_{max}, p$$

$\delta x_{min}/\delta x_{max}$ 는 인간 검출기에 의해 생성된 경계 상자의 왼쪽/
오른쪽 좌표와 실측 경계 상자의 좌표 사이의 정규화된 오프셋

실측 부위 타입 (ground truth part type)

분포를 모델링할 수 있다면, 우리는 인간 검출기에 의해 생성된 인간 제안과 유사한 많은 훈련 샘플을 생성 가능

3.3 Part-Guided Proposal Generator

007 >> 분포 모델링을 위한 Human detection 생성

Halpe-FullBody 데이터 세트에 대한 인간 감지를 생성

데이터 세트의 각 인스턴스에 대해 얼굴, 몸 및 손의 주석 분리

각 분리된 부위에 대해, 타이트하게 둘러싸인 경계 상자와 전체 사람의 감지된 경계 상자 사이의 오프셋을 계산

수평 및 수직 방향의 상자 분산은 일반적으로 독립적이기 때문에 원래 분포를 아래의 식으로 모델링하여 단순화

$$P_x(\delta x_{min}, \delta x_{max} | p), P_y(\delta y_{min}, \delta y_{max} | p).$$

008 >> Halpe-FullBody 데이터 처리

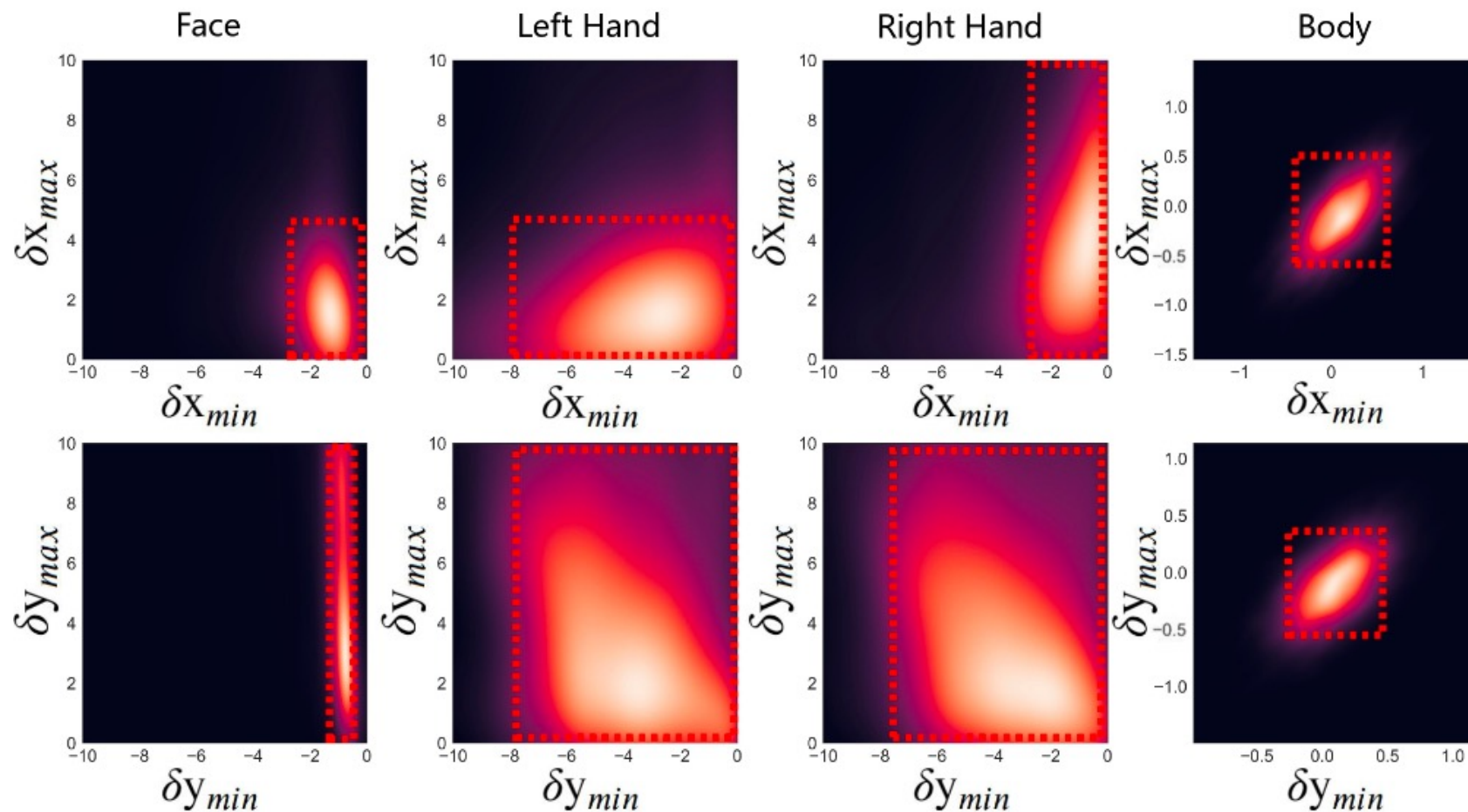
Halpe-FullBody에서 모든 인스턴스를 처리한 후

오프셋은 주파수 분포를 형성

데이터를 가우스 혼합 분포에 맞추어, 신체 부위별로 가우스 혼합 분포 파라미터가 다름

다음 그림에서 분포와 대응되는 부위를 시각화

3.3 Part-Guided Proposal Generator



3.3 Part-Guided Proposal Generator

009 >> 증식된 훈련 proposal 생성

자세 추정기의 훈련 단계에서, 신체부위 p 에 속하는 훈련 샘플의 경우,

$P_x(\delta x_{min}, \delta x_{max} | p)$ 및 $P_y(\delta y_{min}, \delta y_{max} | p)$ 에 따라 dense sampling하여 추가적인 오프셋을 생성하여 실측 바운딩 박스에 대한 증식된 훈련 proposals을 생성할 수 있음

실제로, 대략적인 균일 분포(그림 3의 점선이 있는 빨간색 상자)에서의 샘플링도 유사한 성능을 낼 수 있다는 것을 발견함

3.4 Parametric Pose NMS

001 >> 하향식 접근법의 단점: early commitment

인간 감지기가 사람을 감지하지 못하면 포즈 추정기가 이를 복구할 힘이 없음
대부분의 하향식 기반 방법은 중복 포즈를 피하기 위해 검출 신뢰도를 높은 값으로 설정하기 때문에 이 문제 발생

002 >> early commitment 해결 방법

위와 반대로 높은 탐지 리콜을 보장하기 위해 탐지 신뢰도를 낮은 값(실험에서 0.1)으로 설정
이 경우 인간 탐지기는 불가피하게 일부 사람에 대한 중복 탐지를 생성하므로 중복 포즈 추정이 발생
따라서 중복을 제거하기 위해 Non-Maximum Suppression (NMS)가 필요
기존의 방법들은 효율적이지 않고, 정확하지도 않아 본 논문에서는 parametric pose NMS 방법을 제안

3.4 Parametric Pose NMS

003 >> 포즈 P_i 에 대한 관절

m개의관절이있는포즈 P_i 는 $\{< k_i^1, c_i^1 >, \dots, < k_i^m, c_i^m >\}$ 로 표시
여기서 k_i^j, c_i^j 는각각관절의j번째위치와신뢰점수

004 >> NMS scheme

NMS포즈를다음과같이다시검토함
먼저가장신뢰도높은포즈를기준으로선택하고,그에가까운일부포즈는제거기준을적용하여제거.
중복포즈가제거되고고유한포즈만보고될때까지나머지포즈set에서이프로세스를반복

3.4 Parametric Pose NMS

005 >> Elimination Criterion

서로 너무가깝고 유사한 포즈를 제거하기 위해 포즈 유사성을 정의
포즈 유사성을 측정하기 위해 포즈 거리 메트릭 $d(P_i, P_j | \Lambda)$ 를 정의하고 임계값 η 를 제거 기준으로 정의
 Λ 은 $d(\cdot)$ 의 매개변수 집합

006 >> $$f(P_i, P_j | \Lambda, \eta) = \mathbb{1}[d(P_i, P_j | \Lambda, \lambda) \leq \eta]$$

위의식은 제거 기준
 $d(\cdot)$ 이 η 보다 작다면, $f(\cdot)$ 의 출력은 1
이는 기준 포즈 P_j 와의 중복성 때문에 P_i 가 제거 되어야 함을 의미

3.4 Parametric Pose NMS

007 >> Pose Distance

거리 함수 $d_{pose}(P_i, P_j)$

우리는 P_i 박스가 B_i 라고 가정하여 다음 소프트 매칭 함수를 아래와 같이 정의

$$K_{Sim}(P_i, P_j | \sigma_1) = \begin{cases} \sum_n \tanh \frac{c_i^n}{\sigma_1} \cdot \tanh \frac{c_j^n}{\sigma_1}, & \text{if } k_j^n \text{ is within } \mathcal{B}(k_i^n) \\ 0 & \text{otherwise} \end{cases}$$

여기서 $\mathcal{B}(k_i^n)$ 는 k_i^n 의 상자 중심이고, $\mathcal{B}(k_i^n)$ 의 각 차원은 원래 상자 B_i 의 1/10
 \tanh 연산은 낮은 신뢰도 점수를 가진 포즈를 걸러냄
해당하는 두 관절이 모두 높은 신뢰 점수를 가질 경우 출력은 1에 가까움
이 거리는 자세 간 일치하는 관절의 수를 부드럽게 계산

3.4 Parametric Pose NMS

008 >>
$$H_{Sim}(P_i, P_j | \sigma_2) = \sum_n \exp\left[-\frac{(k_i^n - k_j^n)^2}{\sigma_2}\right]$$

신체 부위들의 공간적 거리 또한 고려하여 위의 식으로 작성됨

009 >>
$$d(P_i, P_j | \Lambda) = K_{Sim}(P_i, P_j | \sigma_1) + \lambda H_{Sim}(P_i, P_j | \sigma_2)$$

앞서본 두 개의 식을 합쳐서 최종 거리 함수를 작성

여기서 λ 는 두 거리와 $\lambda = \{\sigma_1, \sigma_2, \lambda\}$ 의 가중치를 균형 있게 조정할 값임

이전 포즈 NMS은 포즈 거리 파라미터 및 임계값을 수동으로 설정

하지만, 우리의 매개 변수는 데이터 중심 방식으로 결정 가능

010 >> 최적화

감지된 중복 포즈가 주어지면 제거 기준 $f(p_i, p_j | \lambda, \eta)$ 의 4개 매개 변수는 검증 세트에 대한 최대 mAP를 달성하도록 최적화됨
4D 공간에서의 철저한 검색은 다루기 어렵기 때문에,
다른 두 개의 매개 변수를 반복적인 방식으로 수정하여 한 번에 두 개의 매개 변수를 최적화
수렴이 완료되면 파라미터가 고정되고 테스트 단계에서 사용