
Intermediate Human Pose Estimation

2021. 08. 27

Data Mining and **Quality Analytics** Lab

김상민

발표자 소개

❖ 김상민(Sangmin Kim)

- 현재 석사과정 재학 중 (지도교수: 김성범)
- Data Mining & Quality Analytics Lab
- E-mail: sanmiz@korea.ac.kr
- 연구분야
 - ✓ Graph Neural Network
 - ✓ Human Pose Estimation



Contents

1. Overview

Human Pose Estimation

2. Single Person Pose Estimation

Heatmap-based model

3. Multi Person Pose Estimation

Bottom-up approach model

4. Conclusion

Human Pose Estimation

Overview

Introduction

❖ Pose(자세)

- 몸가짐이나 일정한 태도를 취하고 있는 모습
- 행동의 기본 단위
- 사람을 이해하는 중요한 정보



<https://www.news1.kr/articles/?4385246>

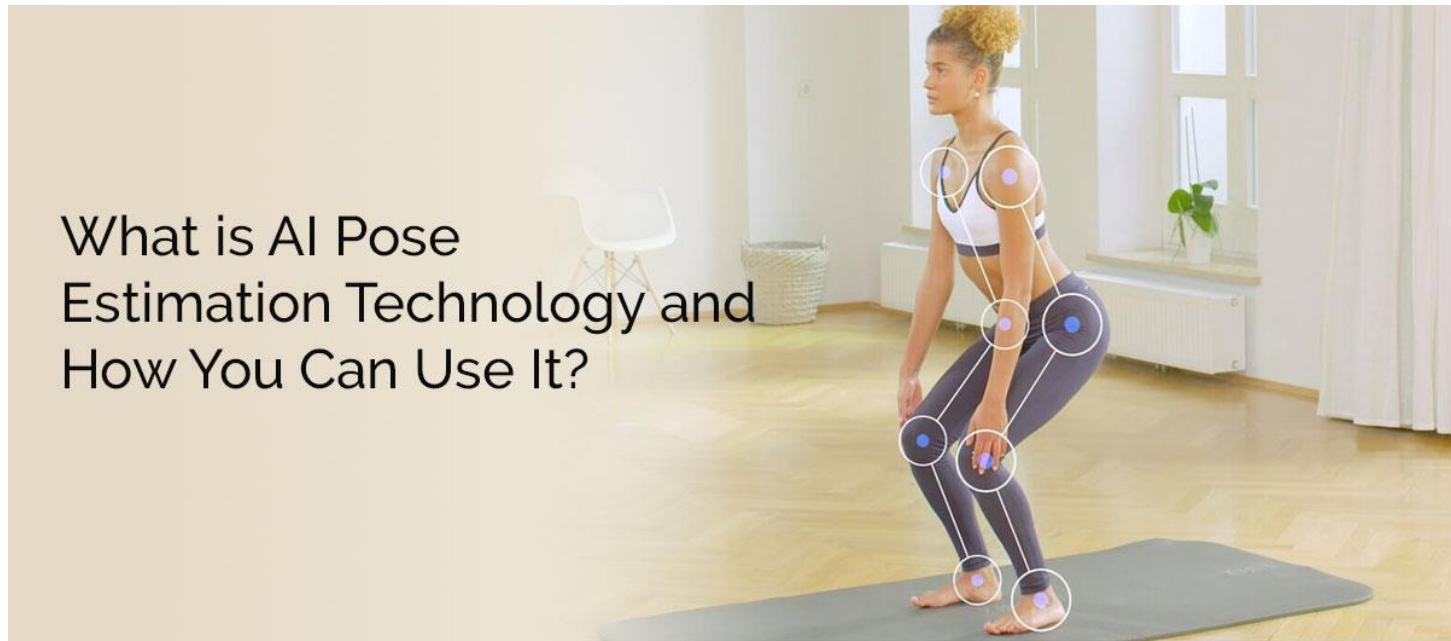
<https://www.youtube.com/watch?v=GBpnsFfLt2Q>

Overview

Introduction

❖ Human Pose Estimation

- **Pose Estimation** 분야는 딥러닝이 부상하기 전부터 **Computer Vision** 분야에서 연구되어 오던 분야
- 딥러닝의 출현과 함께 날개를 달고, 연구 및 상업적으로도 활용의 범위를 넓혀가고 있음
- 핵심은 인체 부분 중 머리, 몸체, 팔, 다리와 같은 관절의 위치를 올바르게 추정하는 것



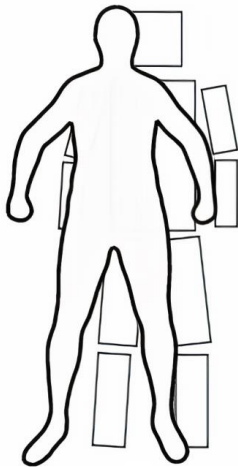
<https://www.quytech.com/blog/human-pose-estimation-technology/>

Overview

Human body models

❖ 연구 동향 – 과거

- **Contour-based model**
- Skeleton-based model
- Volume-based model



Contour-based model



Skeleton-based model



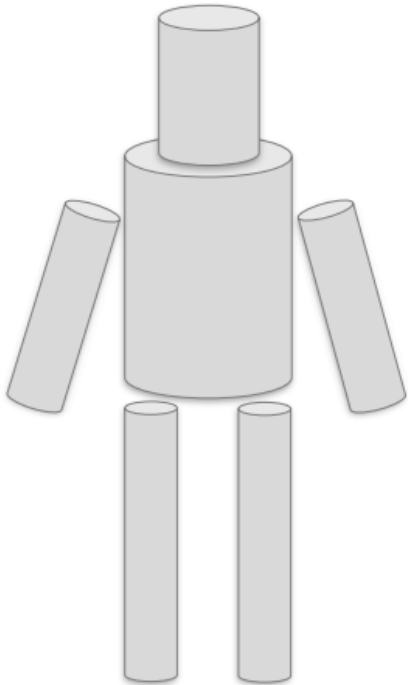
Volume-based model

Overview

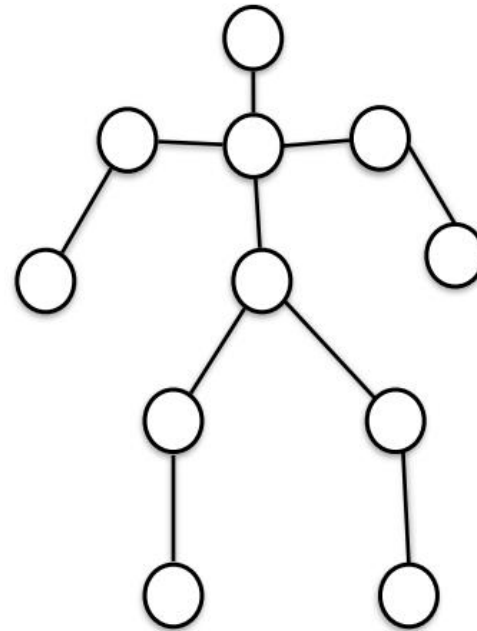
Human body models

❖ 연구 동향 - 과거

- 1970년대, Stanford와 SRI에서 '모든 객체는 단순한 기하학적 형태로 표현 가능하다' 라는 개념을 제안



Generalized Cylinder
(Brooks & Binford, 1979)



Pictorial Structure
(Fischler & Elschlager, 1973)

Overview

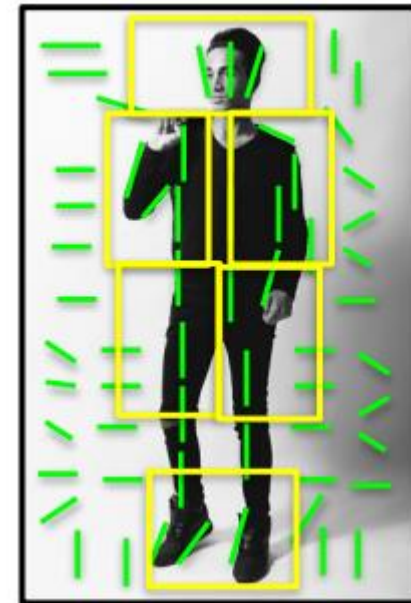
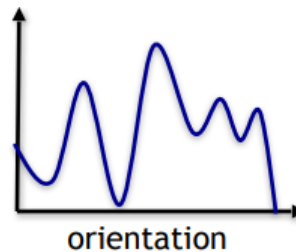
Human body models

❖ 연구 동향 - 과거

- 2000년대, 인터넷과 디지털 카메라 등 기술의 발전과 더불어 사람 모델링 연구의 발전



Histogram of Gradients
(Dalal & Triggs, 2005)



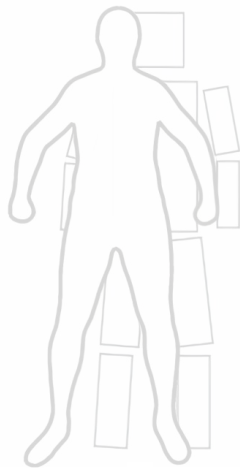
Deformable Part Model
(Felzenswalb, McAllester, Ramanan, 2009)

Overview

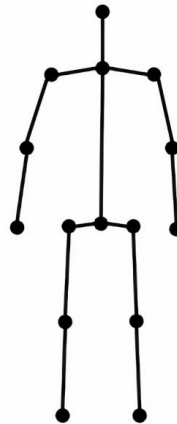
Human body models

❖ 연구 동향 – 현재

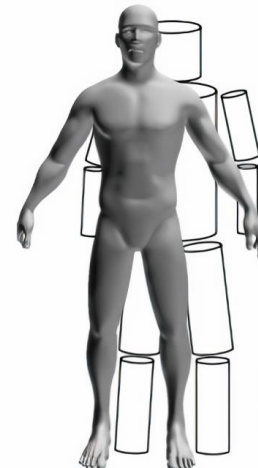
- Contour-based model
- **Skeleton-based model**
- **Volume-based model**



Contour-based model



Skeleton-based model



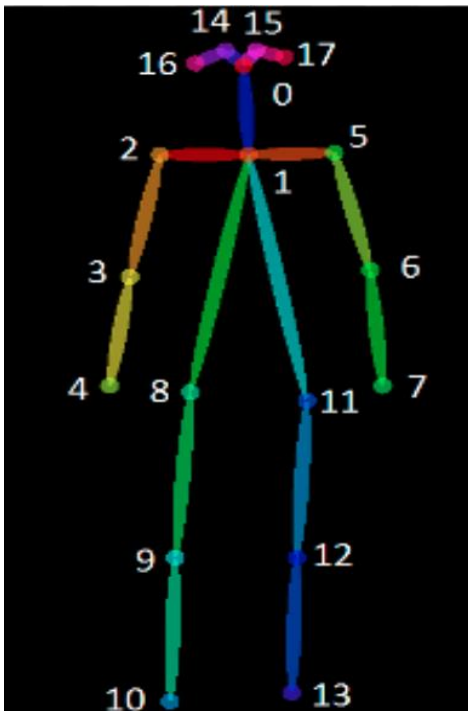
Volume-based model

Overview

Human body models

❖ Skeleton-based model

- 신체 골격 구조를 구성하는 **관절(joint, key point)**로 이루어진 모델
- 2D pose에서는 (x, y) , 3D pose에서는 (x, y, z) 좌표가 사용됨

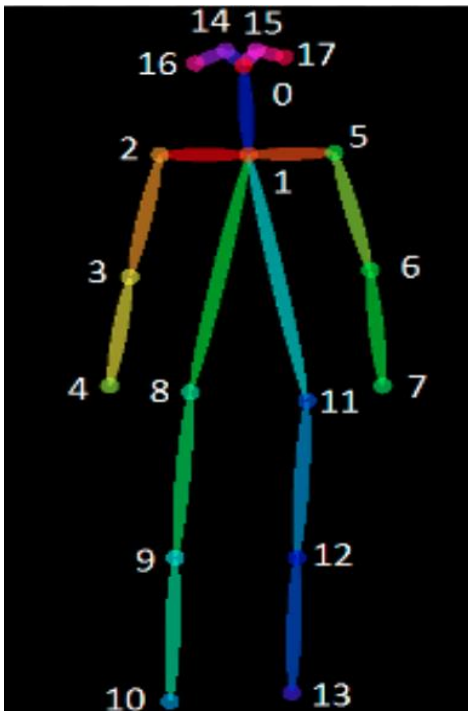


Overview

Human body models

❖ Skeleton-based model

- Part(joint, key point): 관절
- Limb(part pair, part connection): 두 관절 사이 연결
 - ✓ 단, 코- 왼쪽 눈과 같이 관절의 연결이라고 보기 어려운 pair도 존재

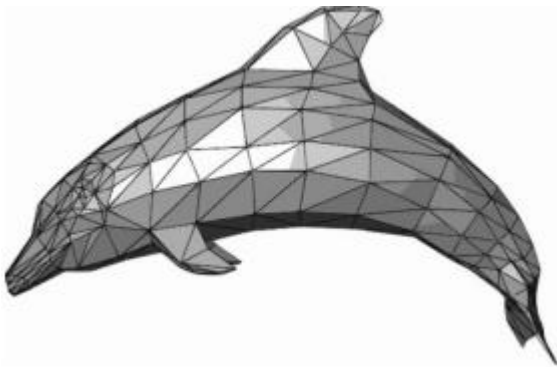


Overview

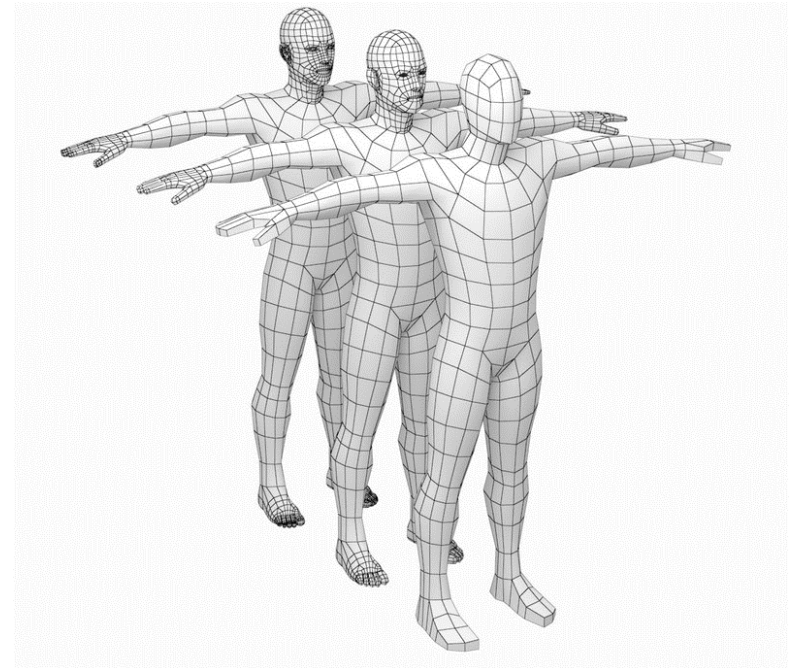
Human body models

❖ Volume-based model

- 3D body shape 및 3D pose 추정에 일반적으로 활용
- 3D mesh 데이터를 활용하여 모델링
 - ✓ Mesh 데이터는 다각형 면(삼각형 혹은 사각형의 면)으로 구성



Polygonal mesh data



3D mesh modeling

https://en.wikipedia.org/wiki/Polygon_mesh

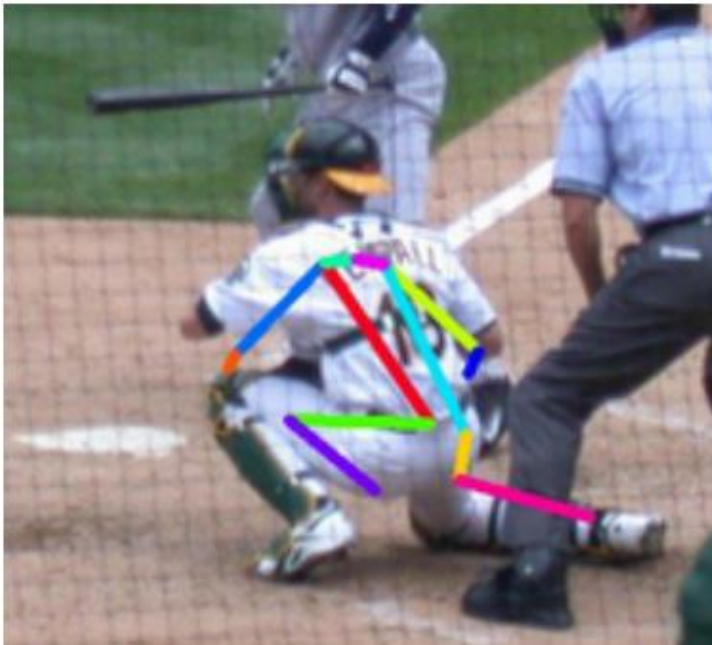
<https://www.turbosquid.com/3d-models/mesh-male-body-t-pose-3d-model-1268686>

Overview

Challenges

❖ Occlusion

- **폐색 현상:** 사람이 사람에 의해 가려지거나, 다른 객체에 의해 가려지는 현상
- 이로 인해 각 관절에 대한 낮은 확률 값(Confidence)을 얻게 되고, 올바른 추정이 어려움



<https://eehoeskrap.tistory.com/466>

Overview

Challenges

❖ 이 외 문제점

- 인접한 사람들 사이에서의 관절 위치 중복
- 다양한 카메라 각도 및 빛의 세기, 의상에 따른 신체 외형 변화에 따른 어려움



Golda, T., Kalb, T., Schumann, A., & Beyerer, J. (2019, September). Human pose estimation for real-world crowded scenarios. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-8). IEEE.

Overview

Applications

❖ 활용 분야

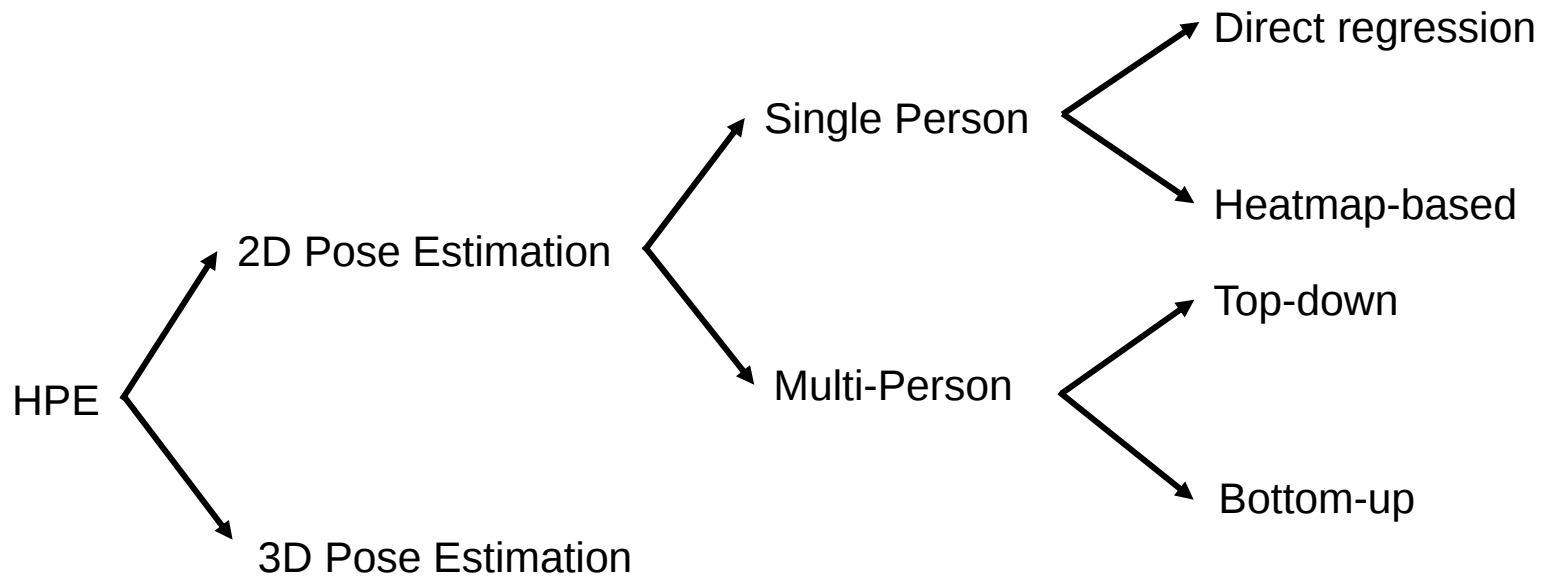
- 자세 교정, 행동 인식, 이상 행동 감지, 안전 예방 시스템, 증강·가상 현실, CG 제작 등에 활용



Overview

Hierarchy of HPE

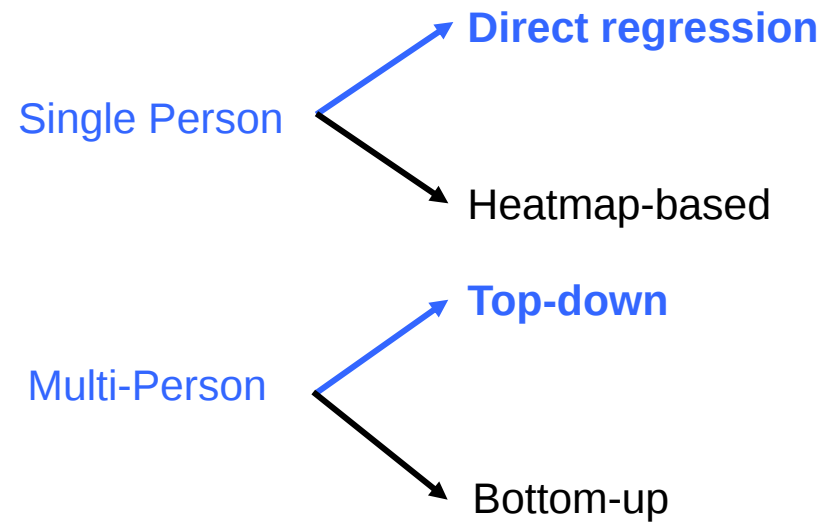
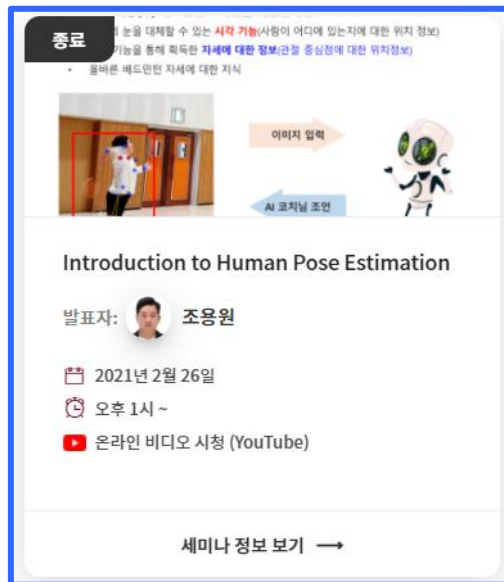
❖ Human Pose Estimation 계층도



Overview

Hierarchy of HPE

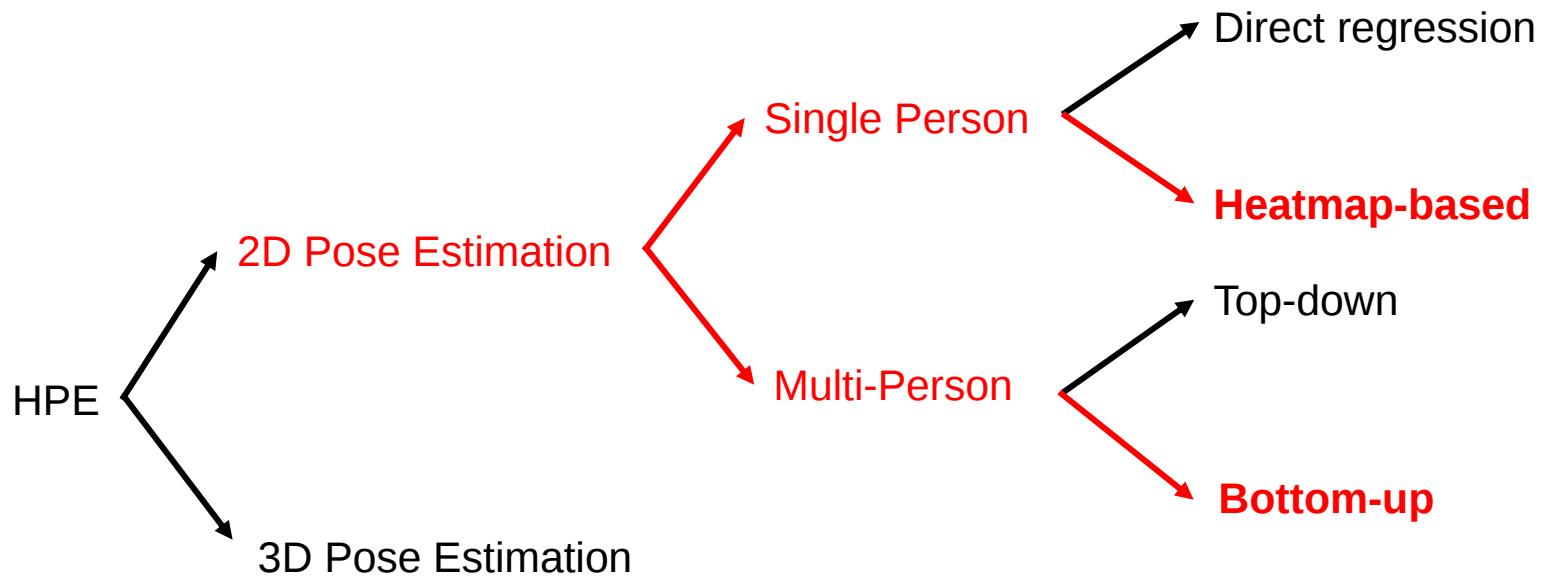
❖ Introduction to Human Pose Estimation (2021.02.26)



Overview

Hierarchy of HPE

❖ Intermediate Human Pose Estimation (2021.08.27)

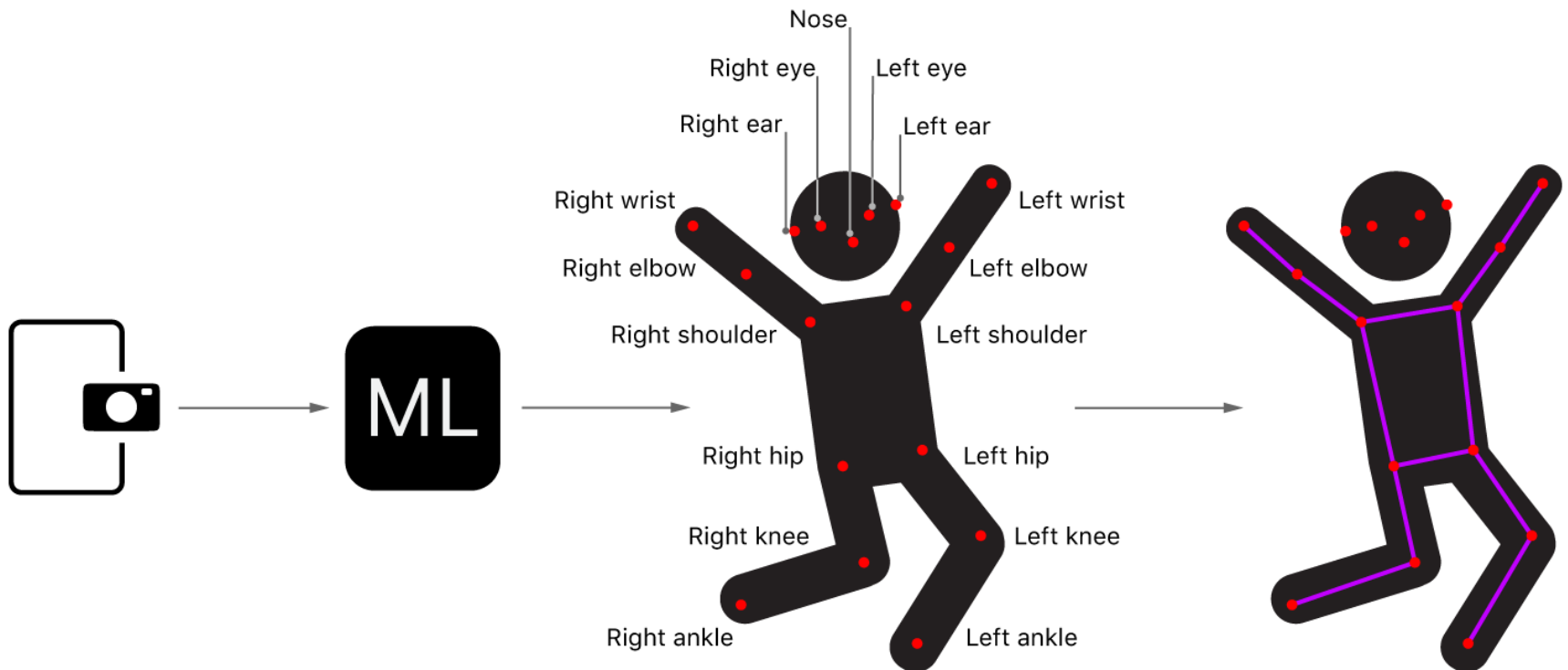


Overview

Input data and target(label)

❖ 2D Pose Estimation 입력 및 출력 데이터

- 입력 데이터는 일반적으로 **RGB 이미지**
- 출력 데이터는 관절 17가지에 해당하는 **(x, y) 좌표**



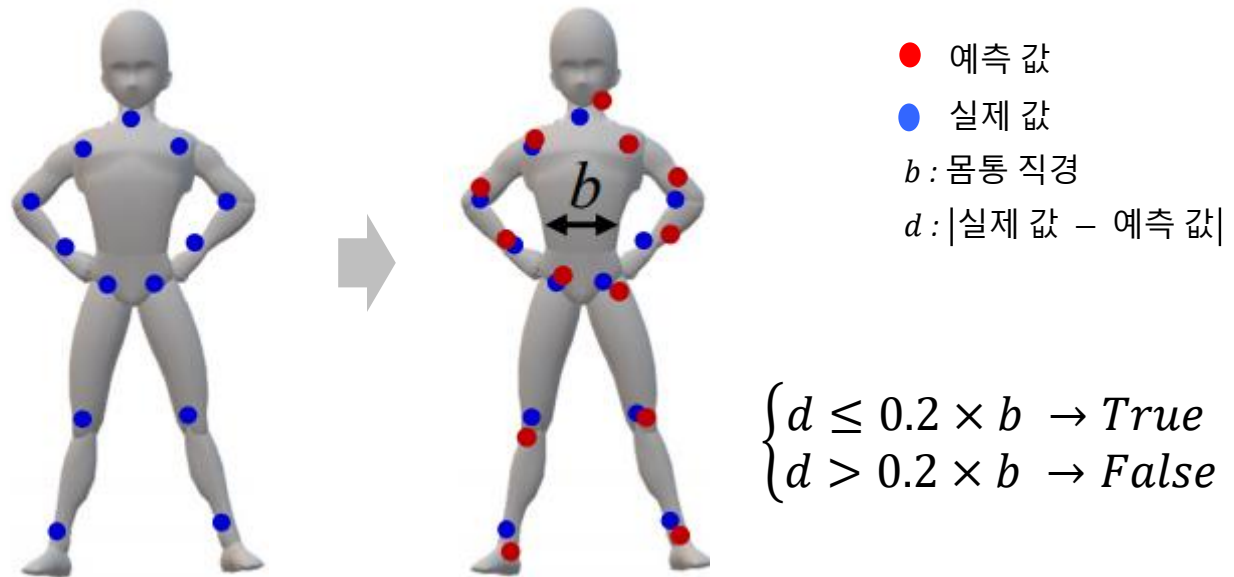
https://developer.apple.com/documentation/coreml/detecting_human_body_poses_in_an_image

Overview

Evaluation metrics

❖ 2D Single Person Pose Estimation 평가지표

- Percentage of Correct Key points(PCK)
 - ✓ 관절 추정 좌표와 정답 좌표의 거리가 임계값보다 작으면 정답으로 분류
 - ✓ PCKh @ 0.2: 임계값을 0.2 x 몸통 직경으로 반영한 PCK
 - ✓ PCKh @ 0.5: 임계값을 0.5 x 머리 길이로 반영한 PCK



$$\begin{cases} d \leq 0.2 \times b \rightarrow True \\ d > 0.2 \times b \rightarrow False \end{cases}$$

$$PCK = \frac{True \text{ 관절의 수}}{\text{총 관절의 수}} \times 100$$

<http://koreascience.or.kr/article/JAKO202021853968943.pdf>

Overview

Evaluation metrics

❖ 2D Multi Person Pose Estimation 평가지표

- Average Precision(AP)
 - ✓ 여러 인물 자세 추정 시 사용되는 평가 지표로, MS COCO 데이터셋과 AI Challenger 데이터셋에서 사용
 - ✓ 10가지 다른 임계값에 따른 Object Key point Similarity를 구하여 평균 낸 값
- Object Key point Similarity(OKS) – **Introduction to Human Pose Estimation 세미나 참조**
 - ✓ 주석된 관절 추정 좌표와 정답 좌표의 유사성 측정 방법
 - ✓ 인물의 추정 자세와 정답 자세가 완전히 일치하면 값이 1이 됨

$$OKS = \frac{\sum_i [\exp(\frac{-d_i^2}{2s^2k_i^2})\delta(v_i > 0)]}{\sum_i [\delta(v_i > 0)]}$$

d_i : 정답 관절과 검출된 관절 사이의 Euclidean 거리

v_i : 관절 점의 주석 여부

k_i : 관절 점 종류에 따른 설정된 상수

s : 해당 객체 세그먼트 영역의 제곱근

❖ Human Pose Estimation

- 인체 부분 중 머리, 몸체, 팔, 다리와 같은 관절의 위치를 올바르게 추정하는 것
- 신체 모델링에 대한 연구는 과거 1970년대부터 이어져왔으나, 카메라 및 인터넷의 발전과 더불어 성장이 가속됨
- 여전히, 극복해야할 문제들이 많으며, 특히 occlusion(폐색 현상)이 대표적인 어려움으로 꼽힘
- HPE 분야는 크게 2D 및 3D 자체 추정으로 나뉘며, 2D 자세 추정은 이미지 내 사람의 수에 따라 구분, 이는 세부적 접근 방식에 따라 구분됨
- 일반적으로 RGB 이미지를 입력 변수로 하여 모델을 통해 관절 좌표(x,y)를 추출

Single Person Pose Estimation

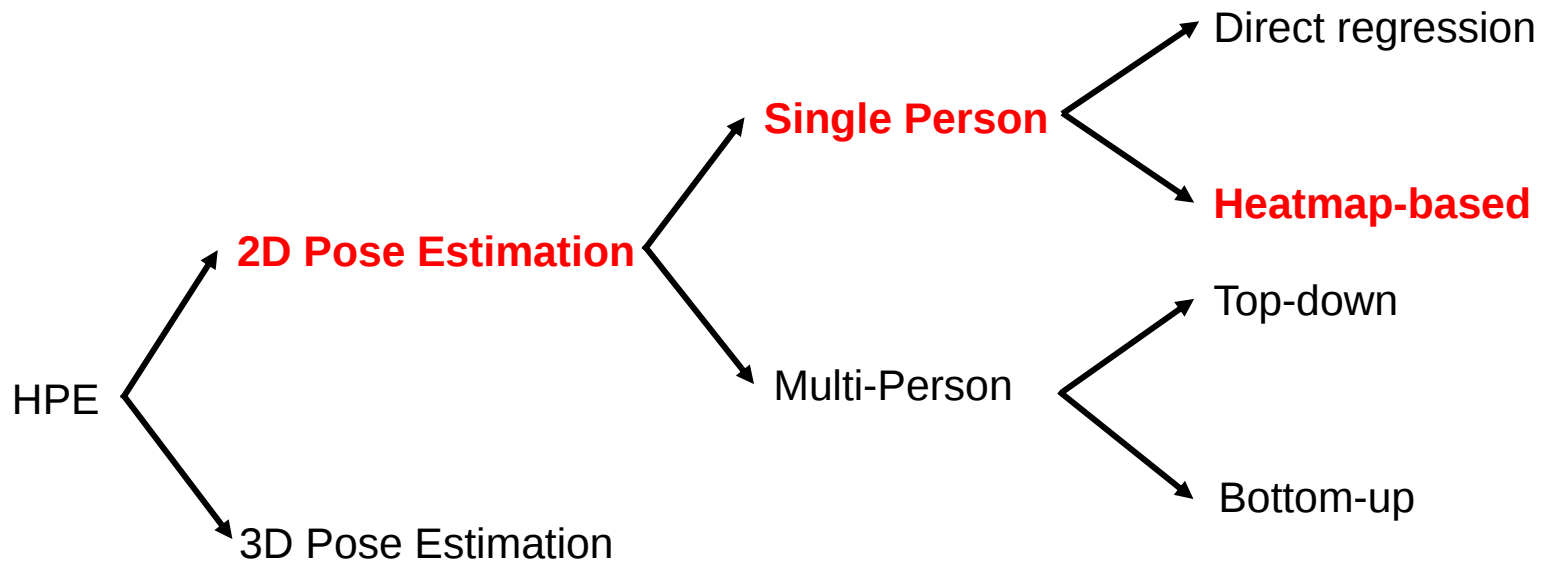
Heatmap-based method

Single Person Pose Estimation

Background

❖ Approach

- Hierarchy of HPE



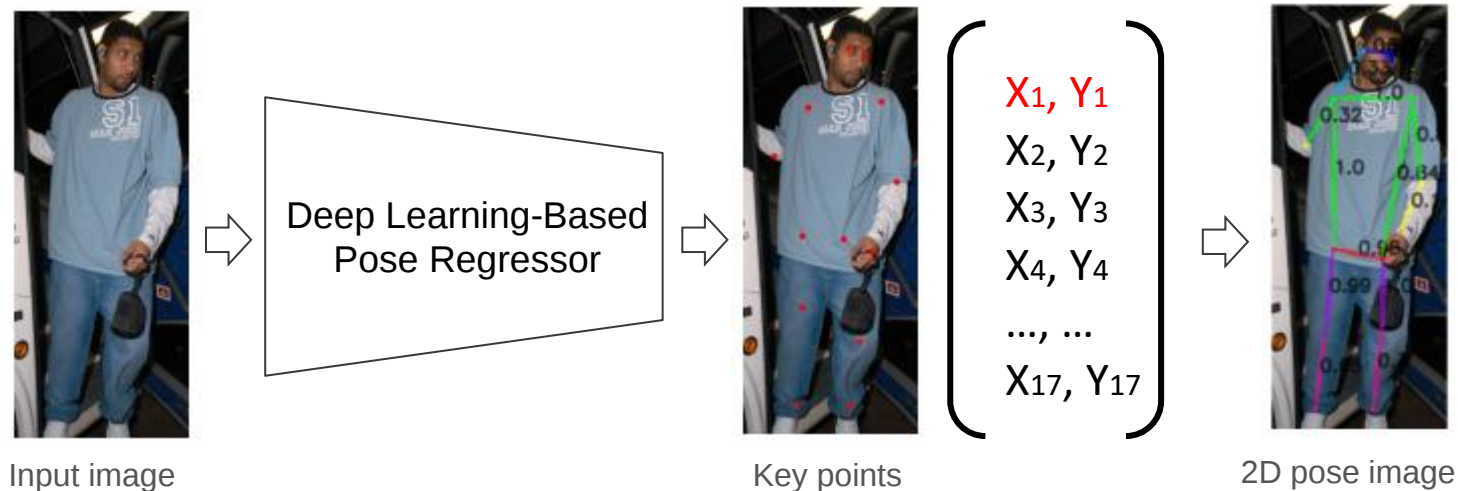
Single Person Pose Estimation

Background

❖ Approach

- **Direct regression**

- ✓ 관절 별 좌표를 딥러닝 기반 예측 모델을 통해 추정
- ✓ 딥러닝 방법론을 HPE 분야에 최초로 적용한 DeepPose(2014, CVPR) 논문이 대표적



Direct regression method

Single Person Pose Estimation

Background

❖ Approach

- **Direct regression**
 - ✓ 관절 별 좌표를 딥러닝 기반 예측 모델을 통해 추정
 - ✓ 딥러닝 방법론을 HPE 분야에 최초로 적용한 DeepPose(2014, CVPR) 논문이 대표적

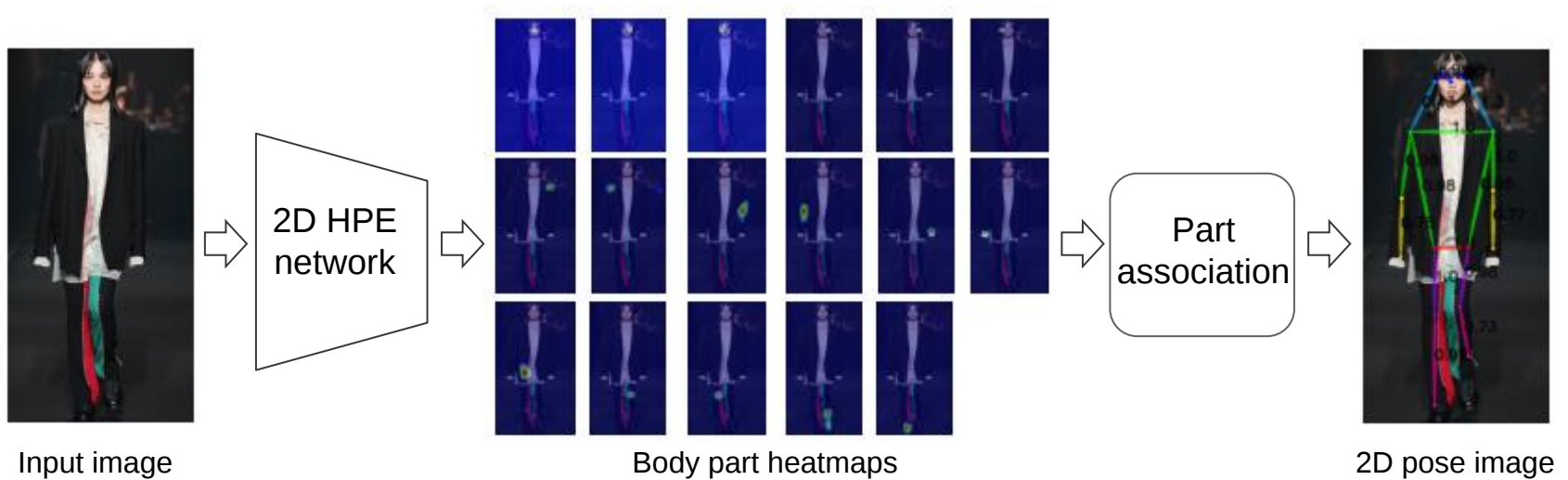


Single Person Pose Estimation

Background

❖ Approach

- **Heatmap-based**
 - ✓ 관절 별 좌표 대신 각 관절이 나타날 확률(heatmap)을 추정
 - ✓ 2D HPE에서 이후 논문들의 주요 방법론으로 자리매김



Heatmap-based method

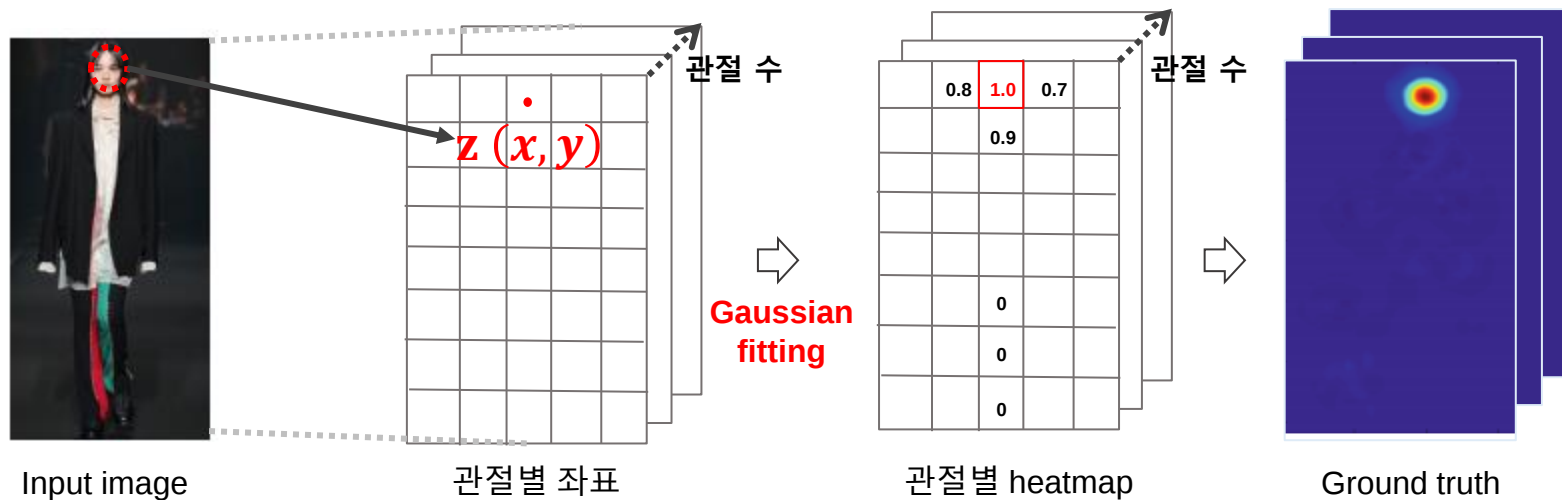
Single Person Pose Estimation

Background

❖ Approach

- Heatmap-based

- ✓ 관절 별 좌표 대신 각 관절이 나타날 확률(heatmap)을 추정
- ✓ Gaussian fitting 을 통해 Ground truth heatmap 생성



Heatmap regression method

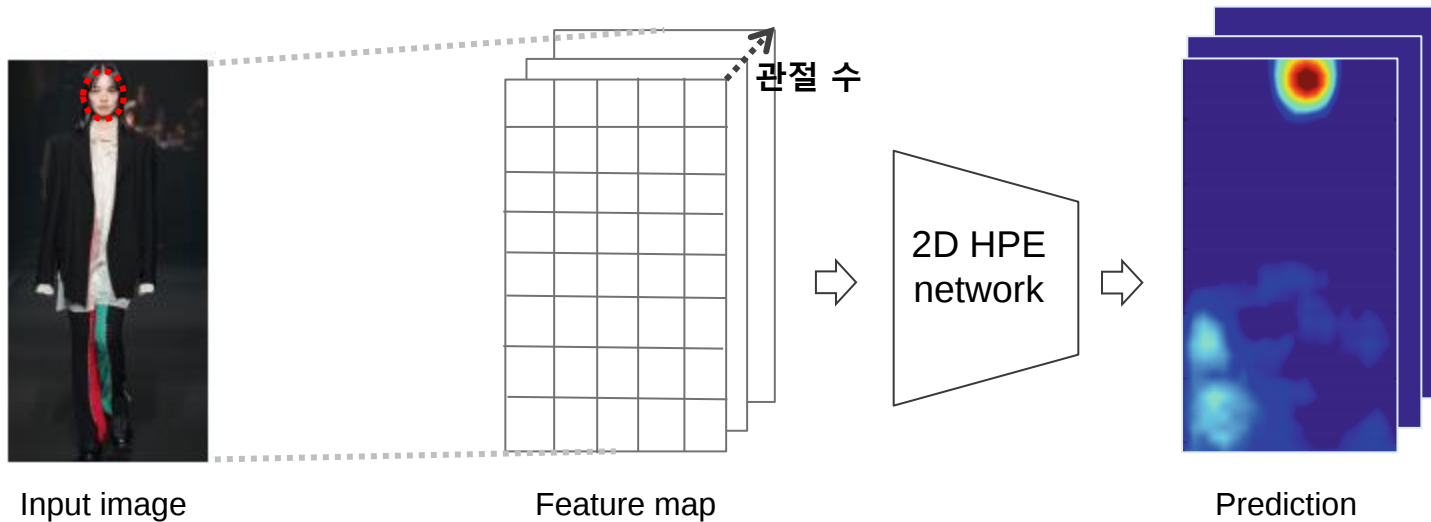
Single Person Pose Estimation

Background

❖ Approach

- Heatmap-based

- ✓ 관절 별 좌표 대신 각 관절이 나타날 확률(heatmap)을 추정
- ✓ CNN을 통해 feature map에서 관절이 있을 법한 부분이 높은 확률을 갖도록 학습



Heatmap regression method

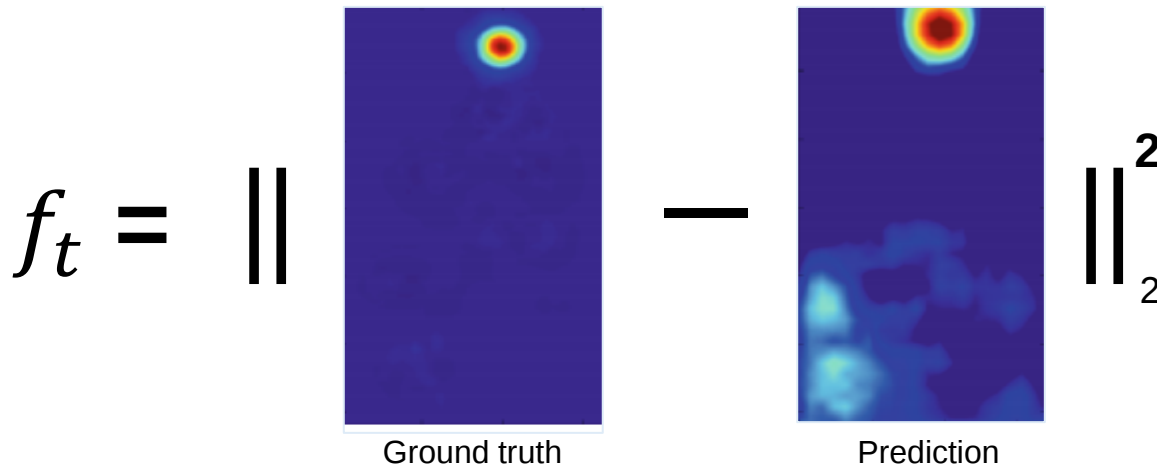
Single Person Pose Estimation

Background

❖ Approach

- Heatmap-based

- ✓ 관절 별 좌표 대신 각 관절이 나타날 확률(heatmap representation)을 추정
- ✓ 손실 함수는 Euclidean distance로 학습



Single Person Pose Estimation

Introduction

❖ Convolutional Pose Machines

- 2016년 IEEE conference on Computer Vision and Pattern Recognition에서 발표
- 저자들은 CMU(Carnegie Mellon Univ.)의 Robotics Institute 소속
- Pose Machines: Articulated Pose Estimation via Inference Machines(2014, ECCV) 후속작
- 2021년 08월 27일 기준으로 2198회 인용

Convolutional Pose Machines

Shih-En Wei
shihenw@cmu.edu

Varun Ramakrishna
vramakri@cs.cmu.edu

Takeo Kanade
Takeo.Kanade@cs.cmu.edu

Yaser Sheikh
yaser@cs.cmu.edu

The Robotics Institute
Carnegie Mellon University

Abstract

Pose Machines provide a sequential prediction framework for learning rich implicit spatial models. In this work we show a systematic design for how convolutional networks can be incorporated into the pose machine framework for learning image features and image-dependent spatial models for the task of pose estimation. The contribution of this paper is to implicitly model long-range dependencies between variables in structured prediction tasks such as articulated pose estimation. We achieve this by designing a sequential architecture composed of convolutional networks that directly operate on belief maps from previous stages, producing increasingly refined estimates for part locations, without the need for explicit graphical model-style inference. Our approach addresses the characteristic difficulty of vanishing gradients during training by providing a natural learning objective function that enforces intermediate supervision, thereby replenishing back-propagated gradients and conditioning the learning procedure. We demonstrate state-of-the-art performance and outperform competing methods on standard benchmarks including the MPII

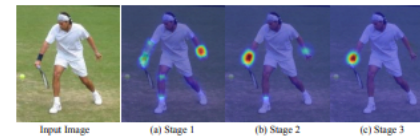


Figure 1: A Convolutional Pose Machine consists of a sequence of predictors trained to make dense predictions at each image location. Here we show the increasingly refined estimates for the location of the *right elbow* in each stage of the sequence. (a) Predicting from local evidence often causes confusion. (b) Multi-part context helps resolve ambiguity. (c) Additional iterations help converge to a certain solution.

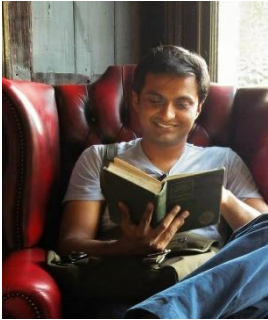
of each part. At each stage in a CPM, image features and the belief maps produced by the previous stage are used as input. The belief maps provide the subsequent stage an expressive non-parametric encoding of the spatial uncertainty of location for each part, allowing the CPM to learn rich image-dependent spatial models of the relationships between parts. Instead of explicitly parsing such belief maps either using graphical models [28, 38, 39] or specialized part processing steps [39, 40], we learn directly

Single Person Pose Estimation

Background

❖ 연구 동향

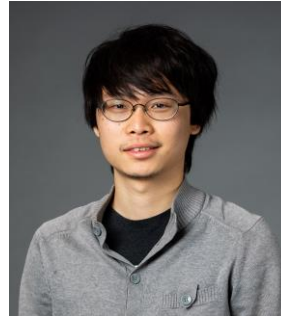
- CMU Robotics Institute에서 Human Pose Estimation 연구 연대기



Varun Ramakrishna

Pose Machines

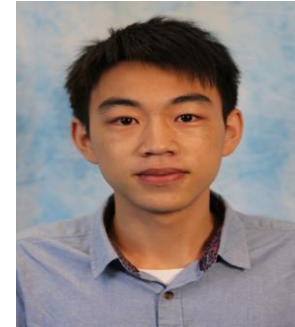
2014



Shih-En Wei

Convolutional Pose Machines

2016



Zhe Cao

OpenPose

2017, 2019

Single Person Pose Estimation

Pose Machines

❖ Pose Machines

- Concept



Single Person Pose Estimation

Pose Machines

❖ Pose Machines

- Concept



Single Person Pose Estimation

Pose Machines

❖ Pose Machines

- 지역적인 이미지(local image)로 학습 시 관절에 따라 정확도 차이가 큼
- 따라서, **Context** 정보와 **신체 구조 정보**를 활용하여 관절 위치 추정



Single Person Pose Estimation

Convolutional Pose Machines

❖ Convolutional Pose Machines

- Convolution 연산을 거친 Receptive field를 통해 **신체 구조 정보**를 활용
- 여러 stage를 통해 나온 heatmap의 **context** 정보를 활용



Single Person Pose Estimation

Convolutional Pose Machines

❖ Convolutional Pose Machines

- Convolution 연산을 거친 Receptive field를 통해 신체 구조 정보를 활용
- Stage를 반복하여 추출한 heatmap의 **context** 정보를 활용

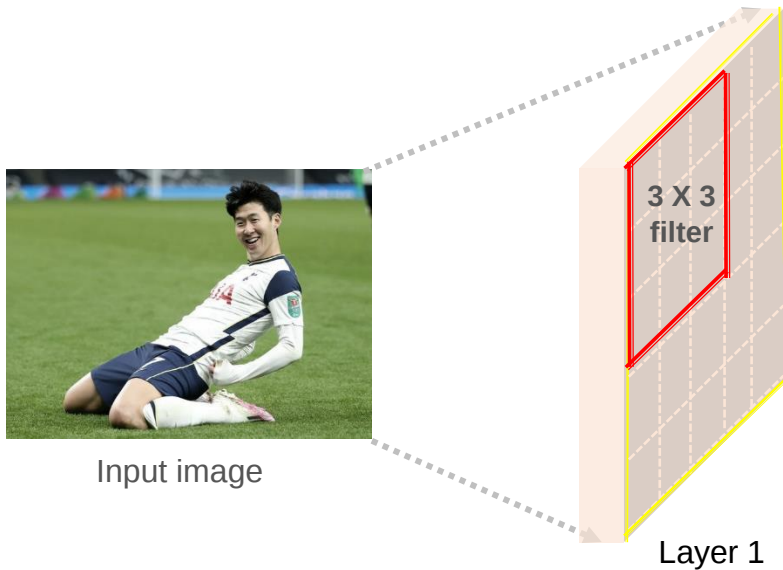


Single Person Pose Estimation

Background

❖ Receptive field

- 각 단계의 입력 이미지에 대해 하나의 필터가 커버할 수 있는 이미지 영역의 일부
- Convolution 레이어를 거치면서 더 많은 영역을 수용

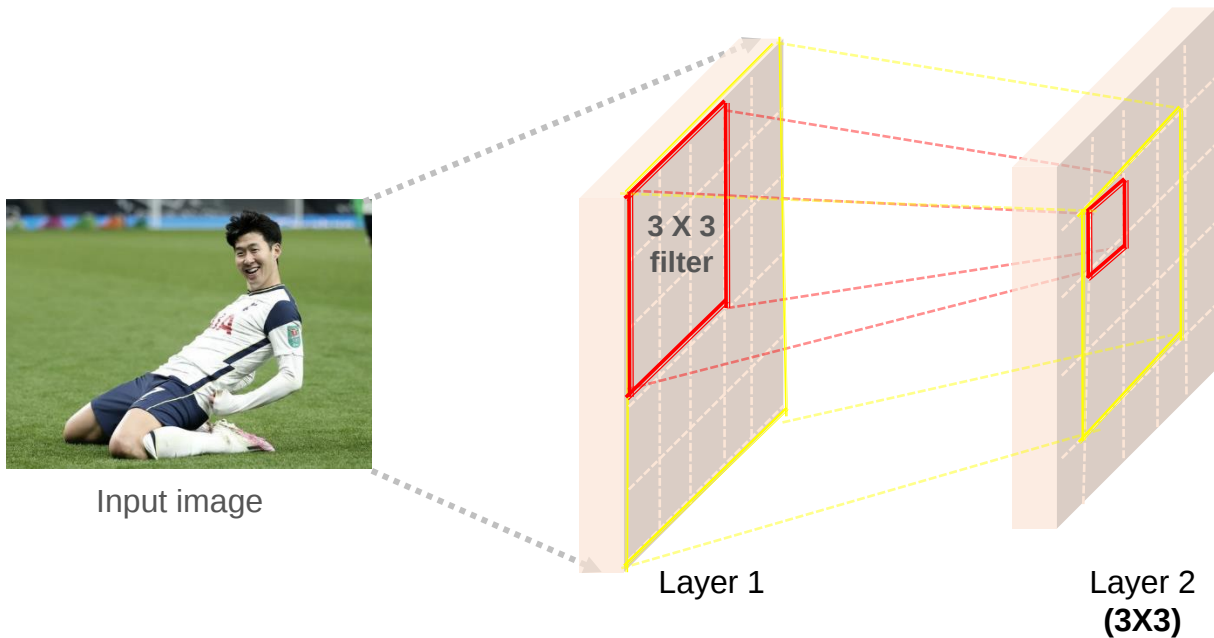


Single Person Pose Estimation

Background

❖ Receptive field

- 각 단계의 입력 이미지에 대해 하나의 필터가 커버할 수 있는 이미지 영역의 일부
- Convolution 레이어를 거치면서 더 많은 영역을 수용

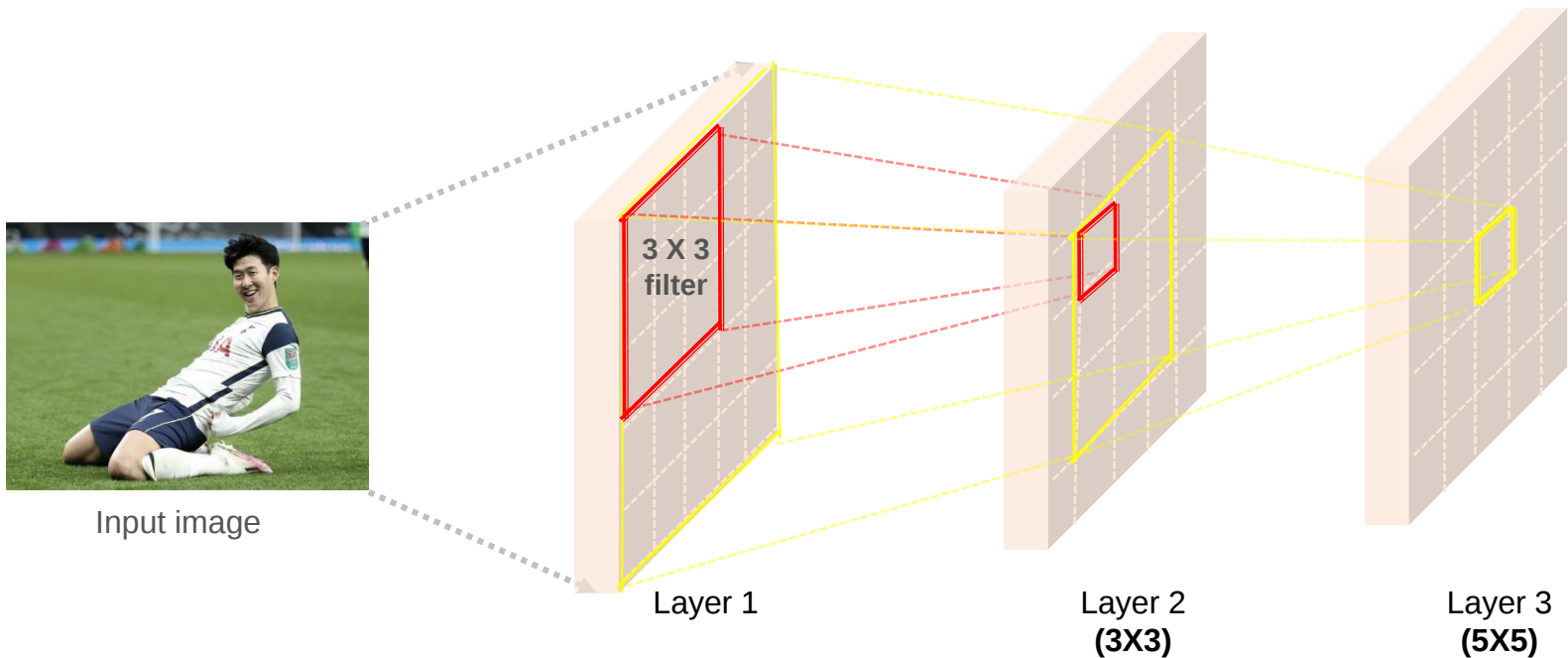


Single Person Pose Estimation

Background

❖ Receptive field

- 각 단계의 입력 이미지에 대해 하나의 필터가 커버할 수 있는 이미지 영역의 일부
- Convolution 레이어를 거치면서 더 많은 영역을 수용

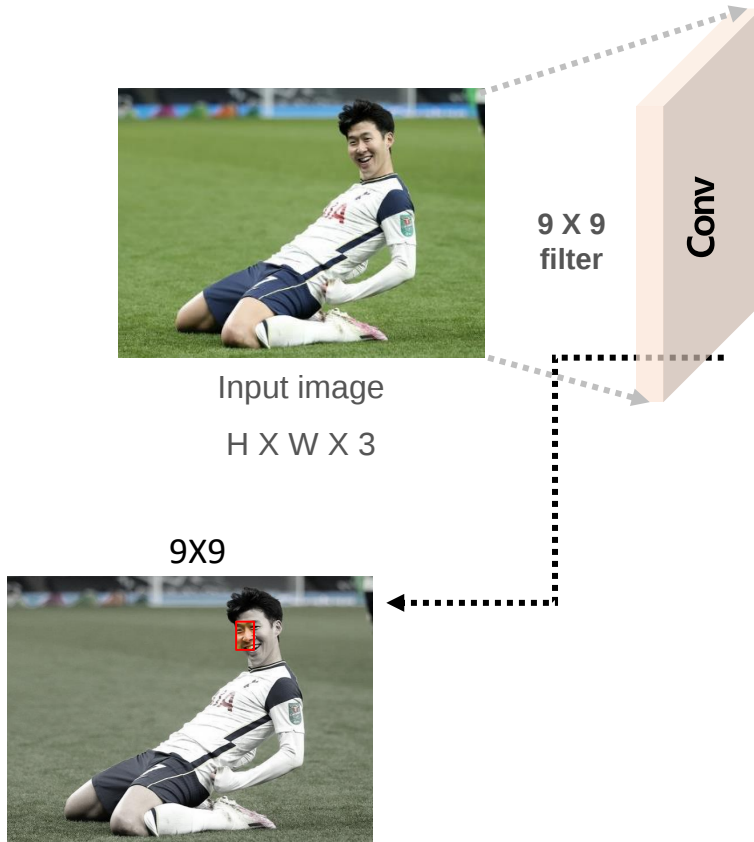


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 **receptive field**는 하위 레이어에서는 **local**한 영역, 상위 레이어로 가면서 **global**한 영역을 해석

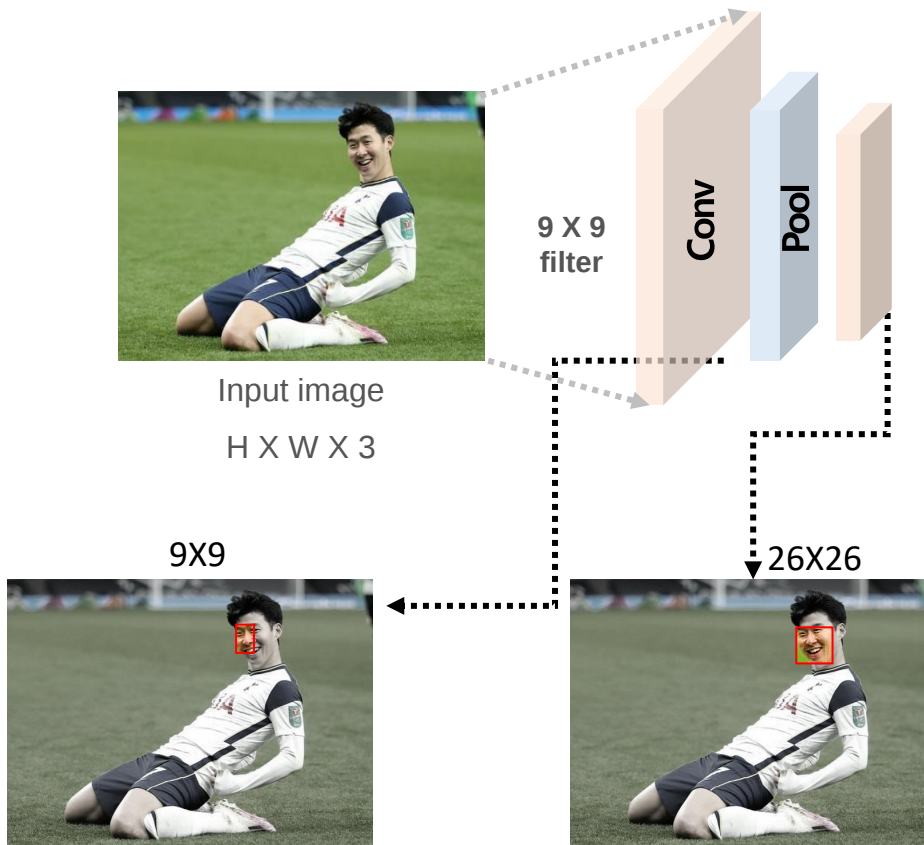


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 **receptive field**는 하위 레이어에서는 **local**한 영역, 상위 레이어로 가면서 **global**한 영역을 해석

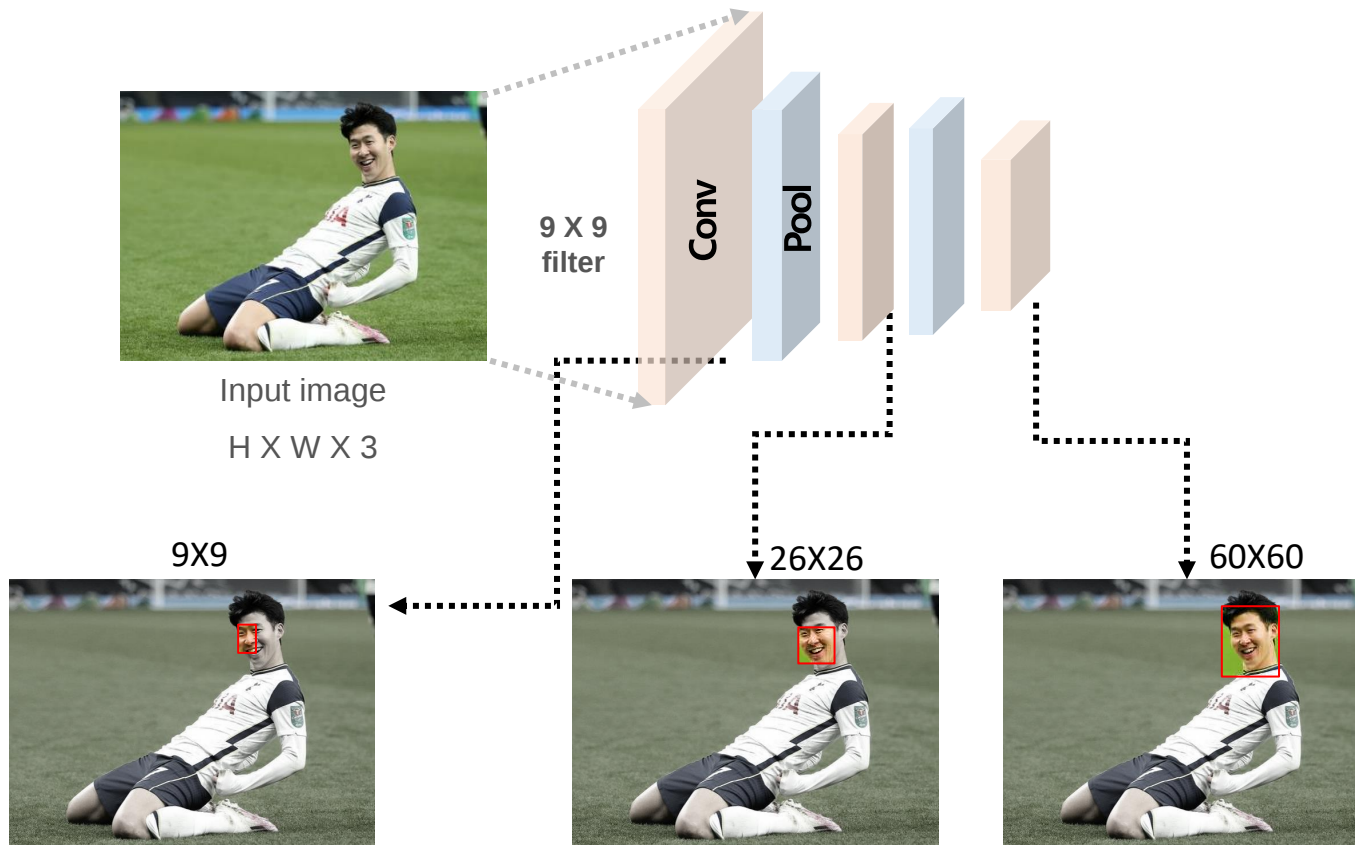


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 **receptive field**는 하위 레이어에서는 **local**한 영역, 상위 레이어로 가면서 **global**한 영역을 해석

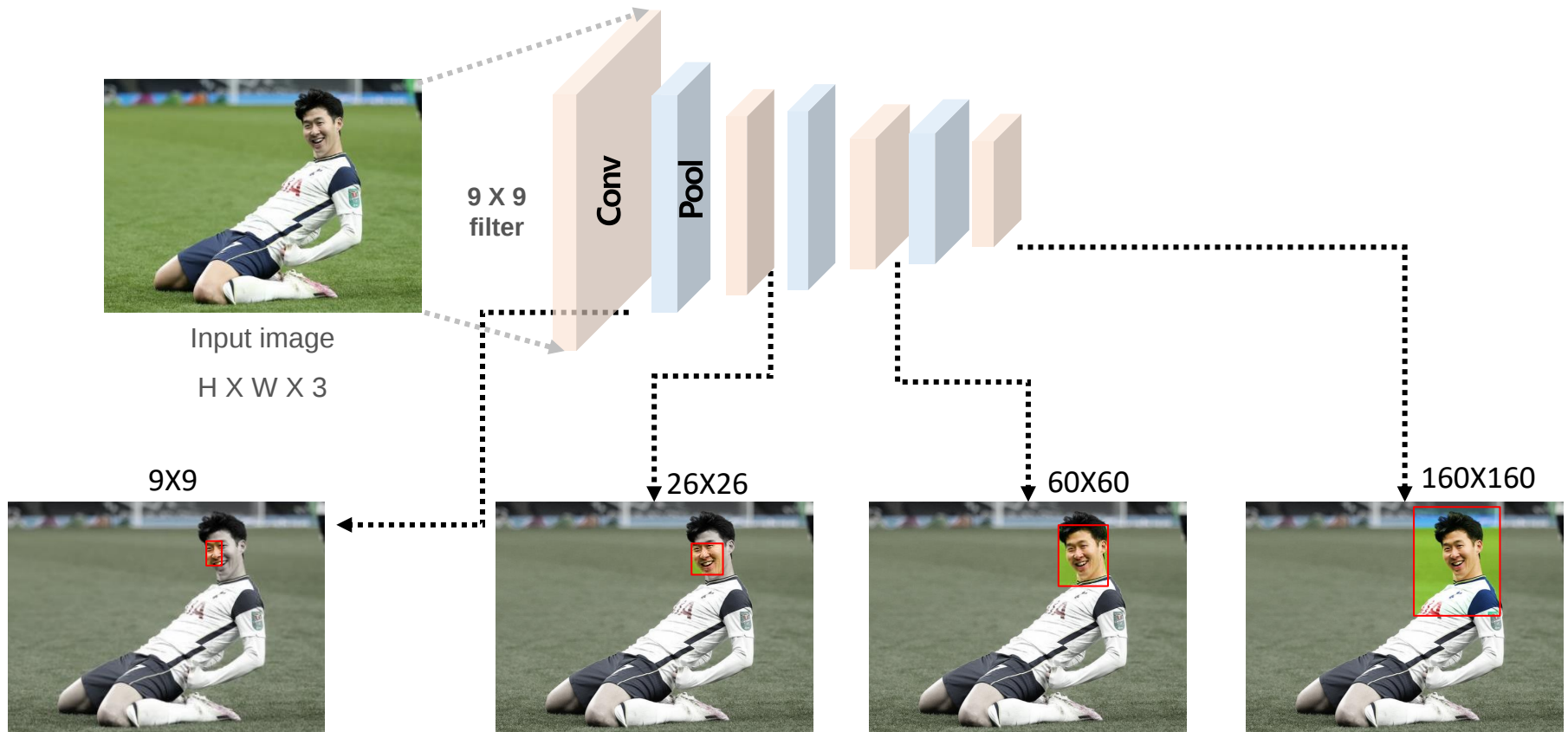


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 **receptive field**는 하위 레이어에서는 **local**한 영역, 상위 레이어로 가면서 **global**한 영역을 해석



Single Person Pose Estimation

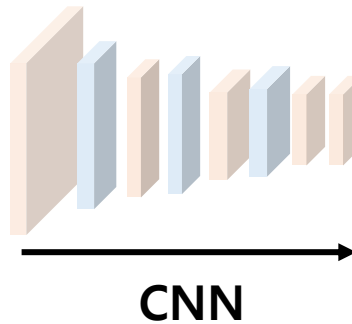
Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 레이어를 지난 receptive field에서 다른 부위의 관절의 정보도 포함될 때, 타겟 관절을 찾기 용이함



9X9



160X160

Single Person Pose Estimation

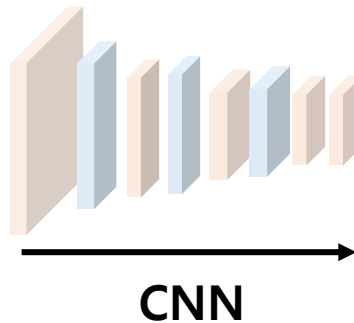
Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 레이어를 지난 receptive field에서 다른 부위의 관절의 정보도 포함될 때, 타겟 관절을 찾기 용이함



9X9



160X160

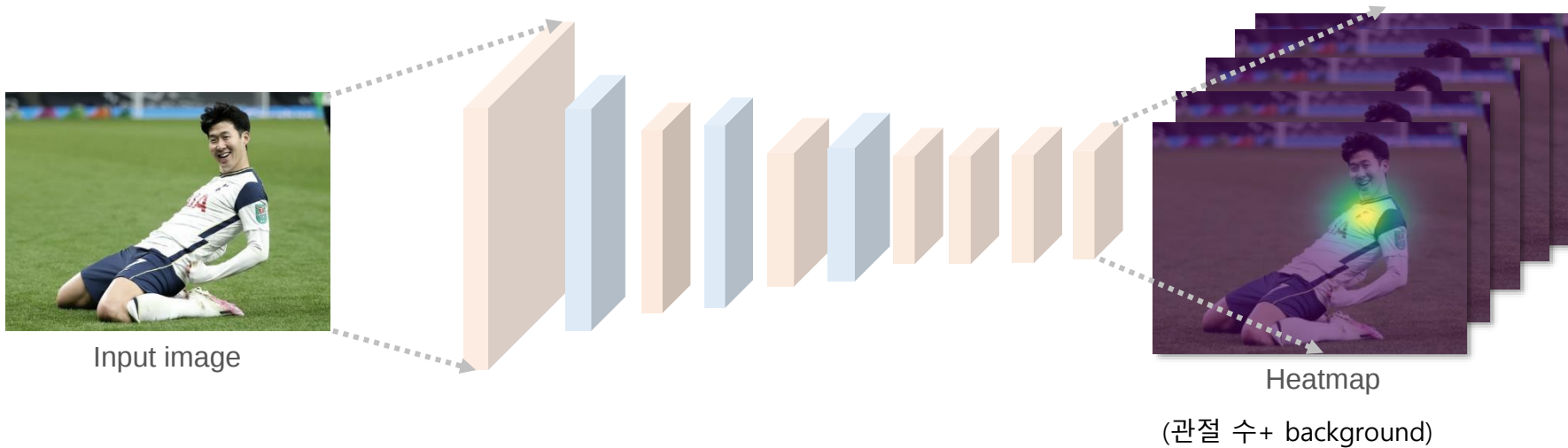
신체 구조를 활용하여 관절 위치 인식

Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 나온 heatmap을 관절 수만큼 예측

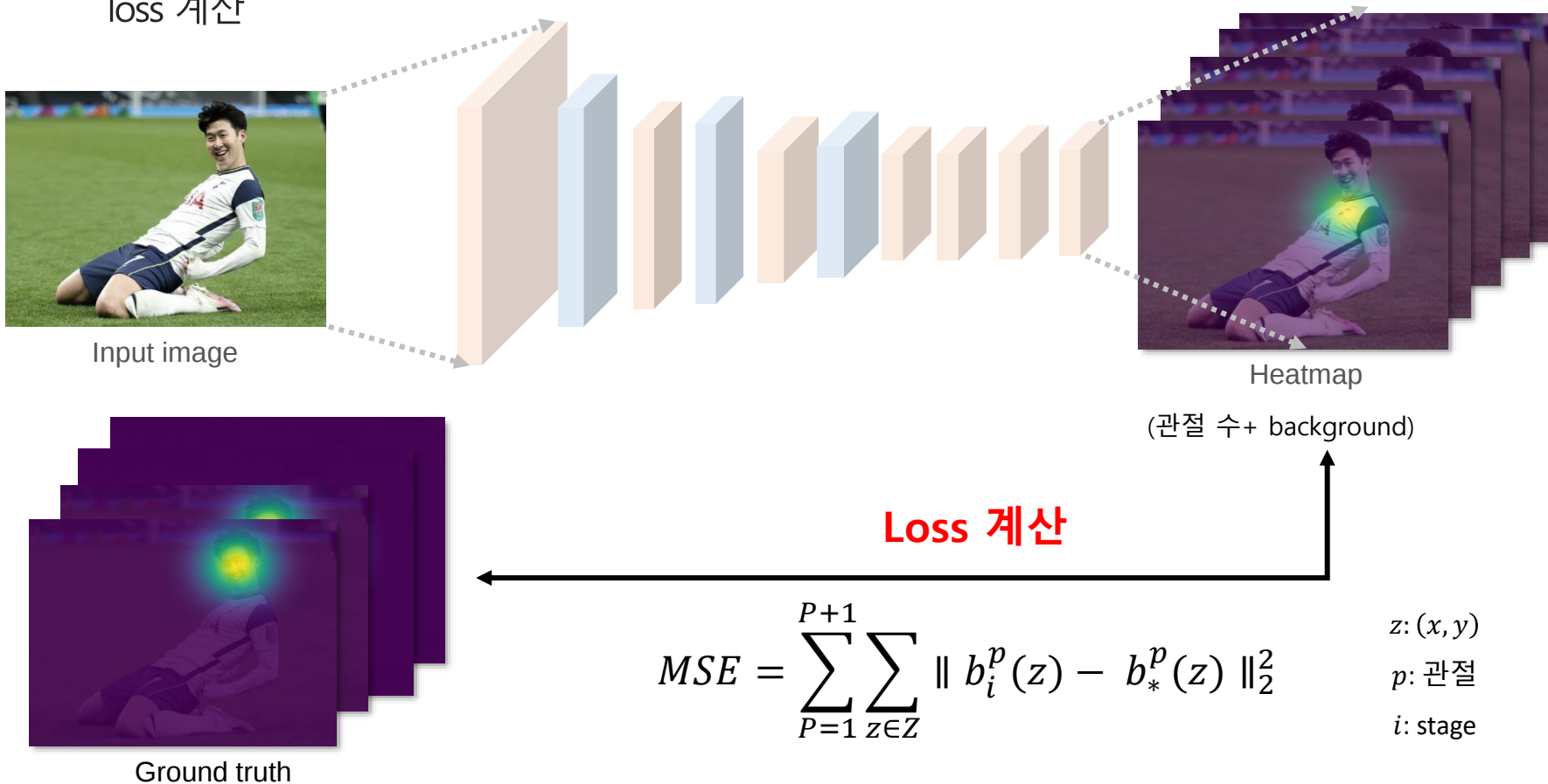


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage 1

- Convolution 연산을 통해 나온 heatmap을 관절 수만큼 예측하여, heatmap ground truth 과 loss 계산

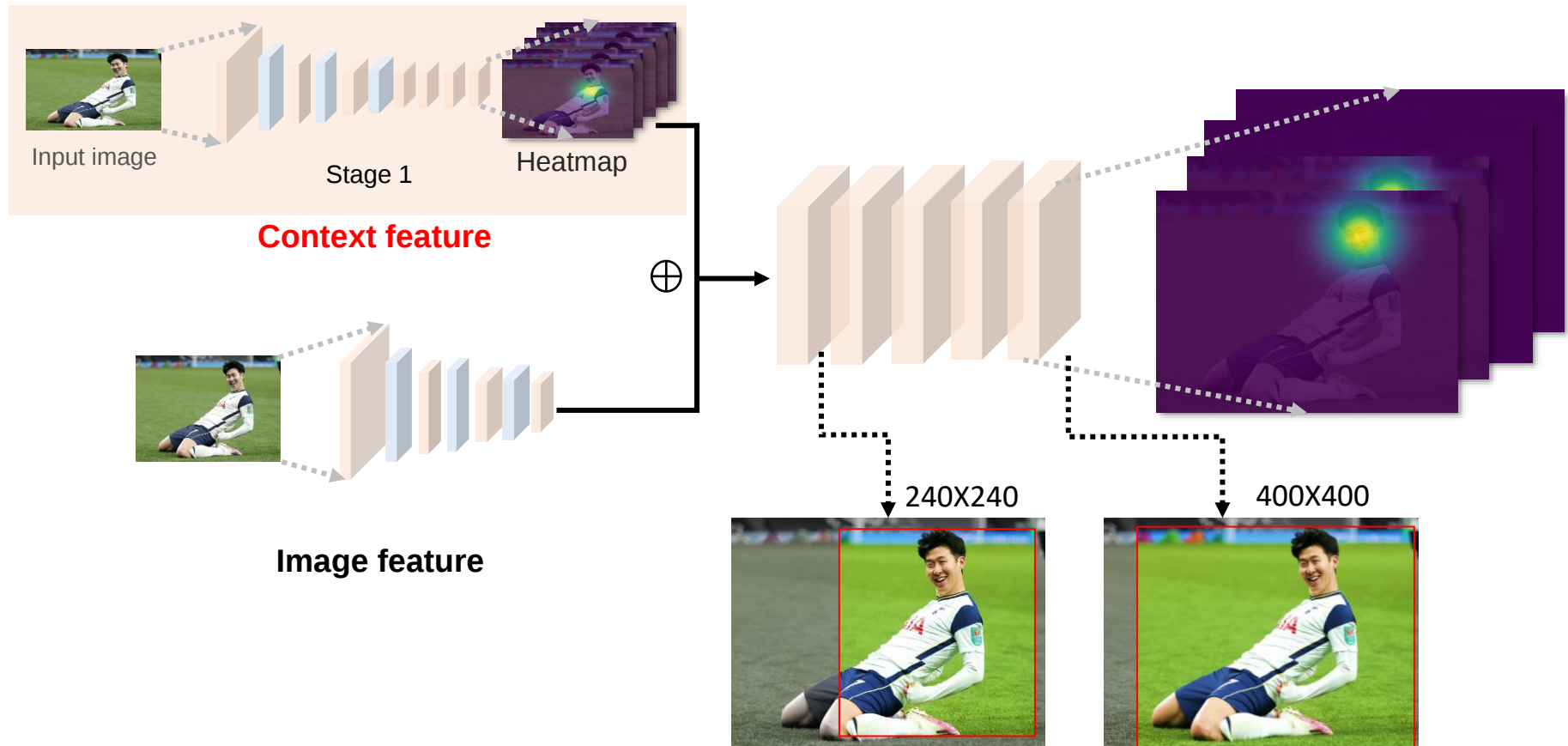


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage ≥ 2

- Stage 2에서의 입력변수는 stage 1에서의 heatmap을 변환한 **context feature**와 입력 이미지를 재차 convolution 연산을 거친 **이미지 feature**

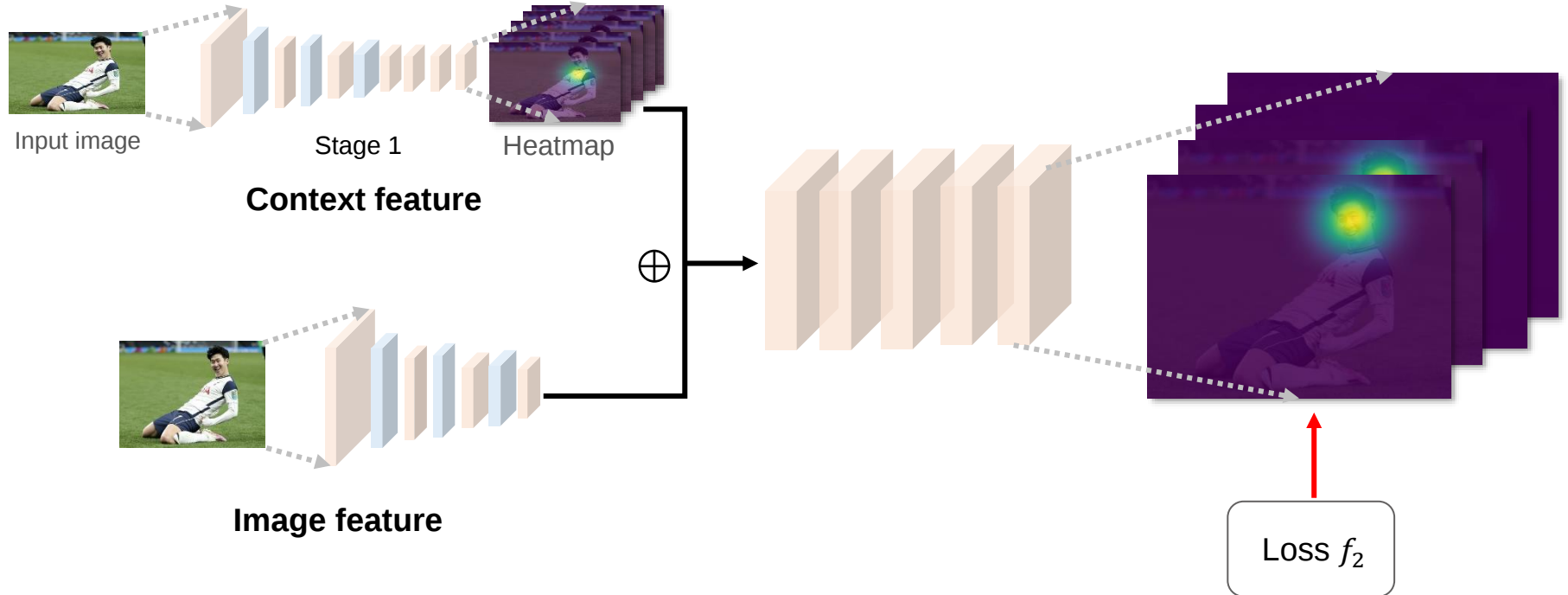


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage ≥ 2

- Stage 1에서와 마찬가지로 예측한 heatmap과 정답 heatmap과의 차이를 통해 손실 값 계산

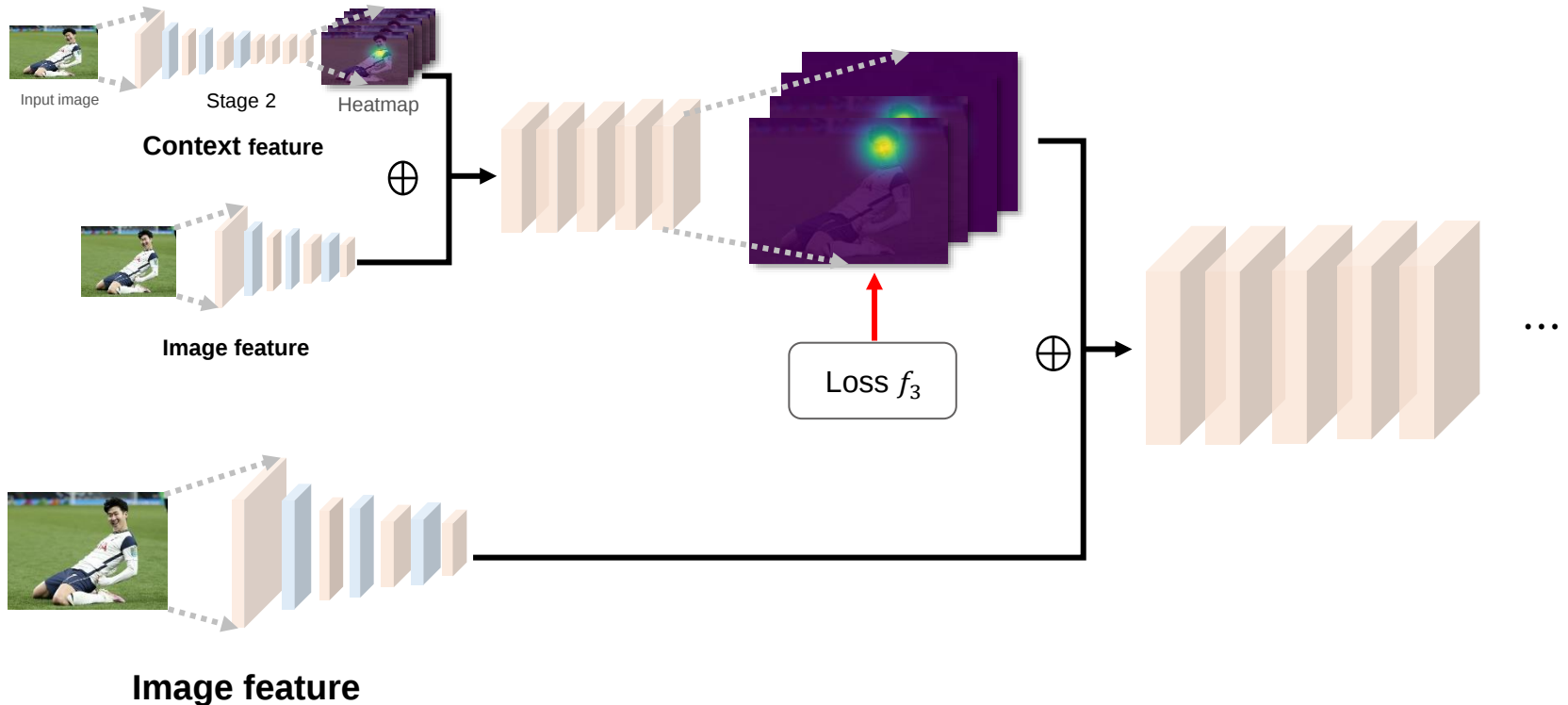


Single Person Pose Estimation

Convolutional Pose Machines

❖ Architecture – Stage ≥ 2

- Stage 6번째까지 같은 네트워크 구조를 반복하여 연산 진행
- Stage 6번째까지의 총 손실 합을 줄이는 방향으로 학습 진행

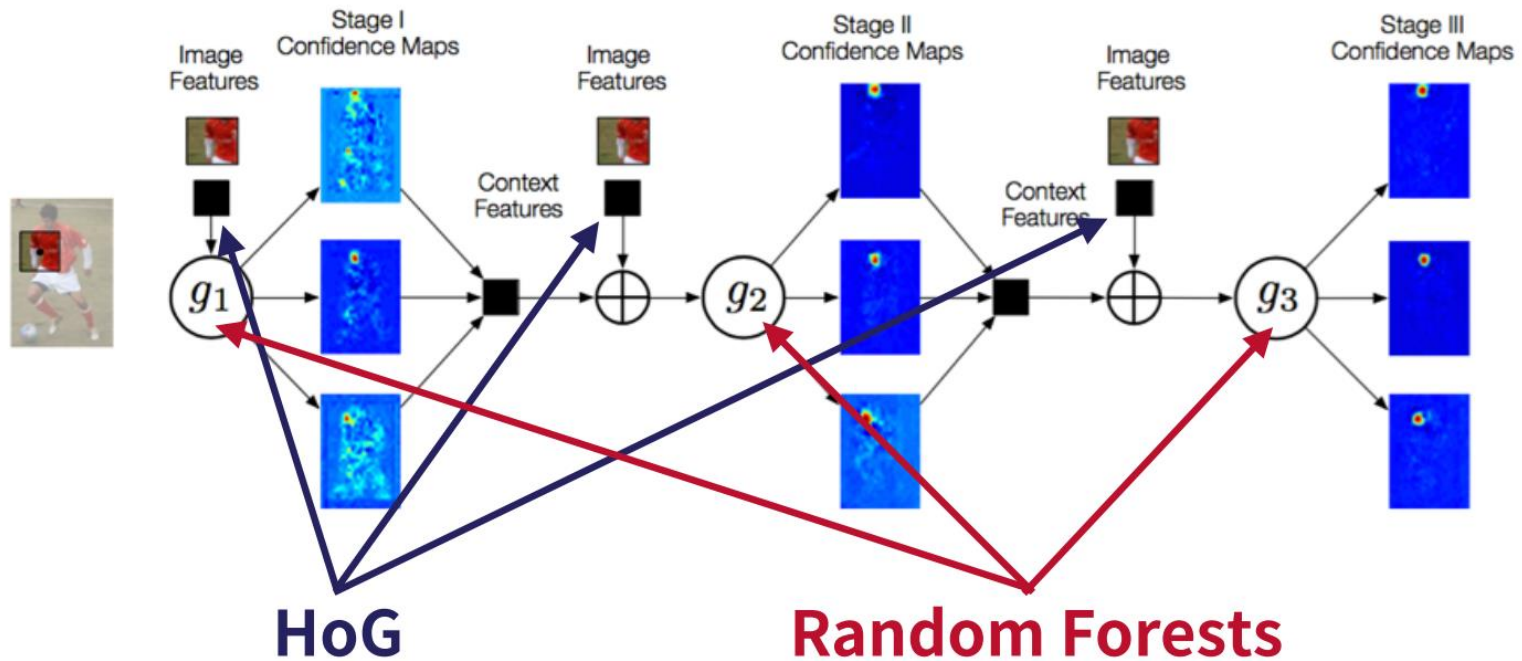


Single Person Pose Estimation

Pose Machines

❖ Comparison

- 기존 Pose Machines에서의 이미지 feature 는 histogram of gradient 방법으로 생성
- Heatmap 생성 이후, 복잡한 연산 과정을(fixed hand-crafted) 통해 context feature 생성

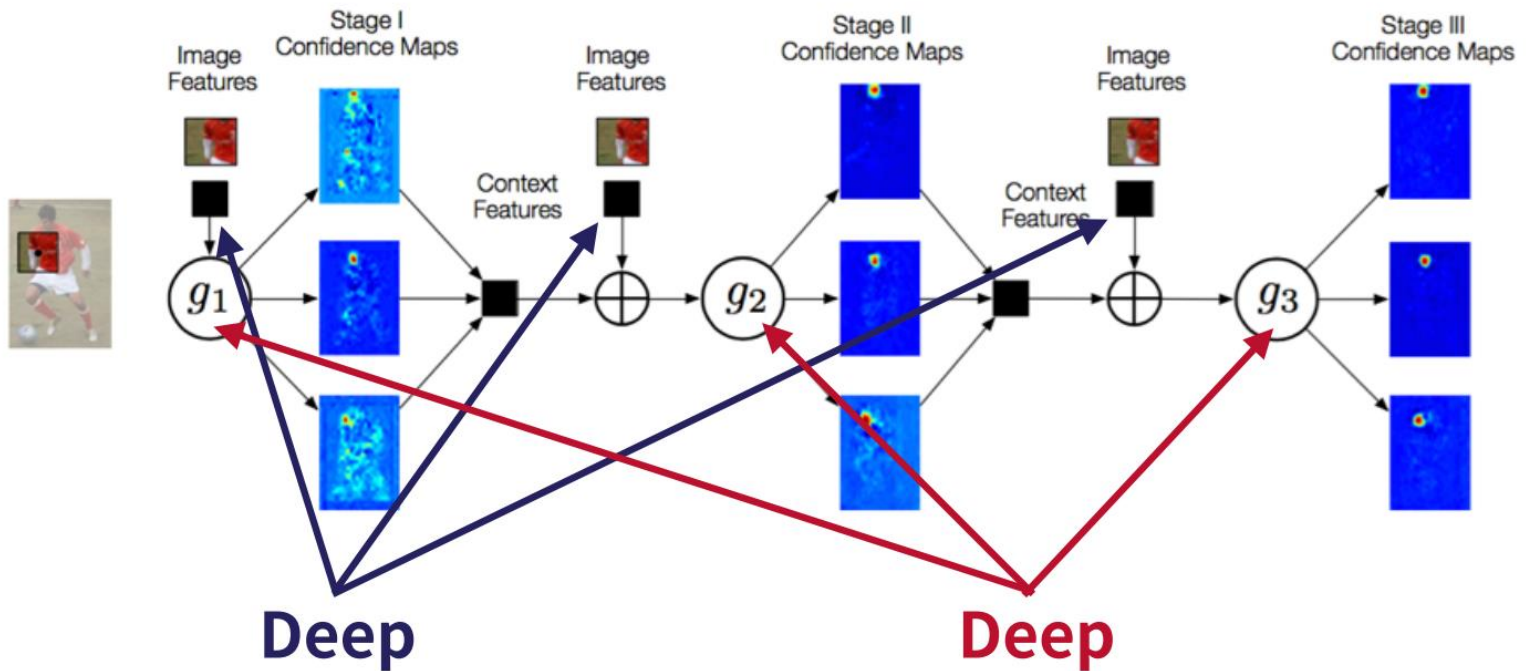


Single Person Pose Estimation

Convolutional Pose Machines

❖ Comparison

- 기존 Pose Machines에서의 이미지 feature 및 context feature를 CNN을 활용하여 생성



Single Person Pose Estimation

Convolutional Pose Machines

❖ Limitation

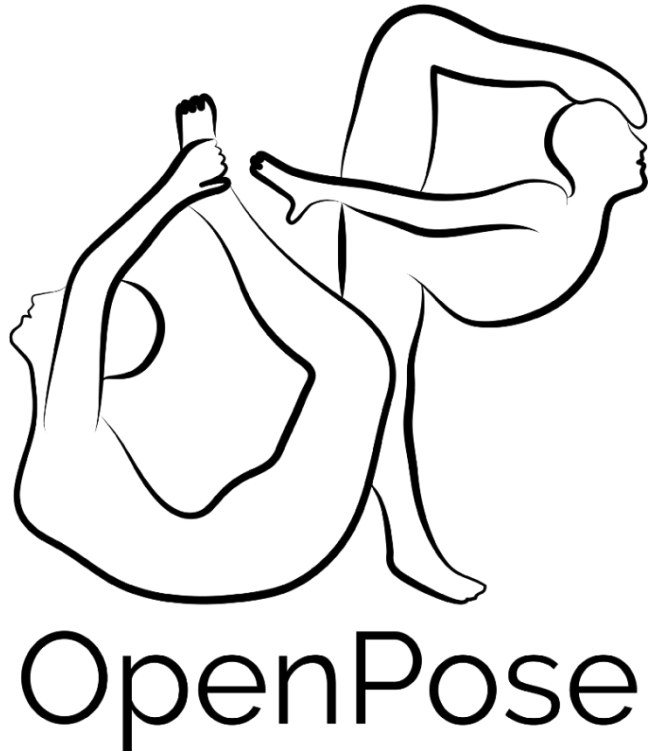


Single Person Pose Estimation

Convolutional Pose Machines

❖ Limitation

- CPM의 한계점을 보완 및 개선하여 OpenPose 논문 등장



Multi Person Pose Estimation

Bottom-up approach

Multi Person Pose Estimation

Introduction

❖ OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

- 2017년 IEEE conference on Computer Vision and Pattern Recognition에서 발표
- 2019년 IEEE Transactions on Pattern Analysis and Machine Intelligence에 업데이트 버전 게재
- 저자들은 CMU(Carnegie Mellon Univ.)의 Robotics Institute 소속
- 2021년 08월 27일 기준으로 각각 3914회, 1685회 인용
- MPPE에서 open-source library로 최초 공개하며, 해당 분야에서 대중적인 모델로 자리잡음

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

Zhe Cao, *Student Member, IEEE*, Gines Hidalgo, *Student Member, IEEE*,
Tomas Simon, Shih-En Wei, and Yaser Sheikh

Abstract—Realtime multi-person 2D pose estimation is a key component in enabling machines to have an understanding of people in images and videos. In this work, we present a realtime approach to detect the 2D pose of multiple people in an image. The proposed method uses a nonparametric representation, which we refer to as Part Affinity Fields (PAFs), to learn to associate body parts with individuals in the image. This bottom-up system achieves high accuracy and realtime performance, regardless of the number of people in the image. In previous work, PAFs and body part location estimation were refined simultaneously across training stages. We demonstrate that a PAF-only refinement rather than both PAF and body part location refinement results in a substantial increase in both runtime performance and accuracy. We also present the first combined body and foot keypoint detector, based on an internal annotated foot dataset that we have publicly released. We show that the combined detector not only reduces the inference time compared to running them sequentially, but also maintains the accuracy of each component individually. This work has culminated in the release of OpenPose, the first open-source realtime system for multi-person 2D pose detection, including body, foot, hand, and facial keypoints.

Index Terms—2D human pose estimation, 2D foot keypoint estimation, real-time, multiple person, part affinity fields.

1 INTRODUCTION

IN this paper, we consider a core component in obtaining a detailed understanding of people in images and videos: human 2D pose estimation—or the problem of localizing anatomical keypoints or “parts”. Human estimation has largely focused on finding body parts of *individuals*. Inferring the pose of multiple people in images presents a unique set of challenges. First, each image may contain an unknown number of people that can appear at any position or scale. Second, interactions between people induce complex spatial



Multi Person Pose Estimation

Approach

❖ Top-Down 방식

- 사람을 먼저 감지를 한 다음 각 사람의 자세를 추정
- 대표적 방법으로 Mask-RCNN(2017, CVPR) 존재
- 문제점: 사람을 인식하지 못하면 측정이 어려우며, 사람 수가 많아지면 계산이 복잡함



사람 탐지



자세 추정

Top-down approach

Multi Person Pose Estimation

Approach

❖ Bottom-up 방식

- ✓ 이미지 내 관절을 먼저 감지하여 서로 연결하여 모든 사람의 자세를 추정
- ✓ 선행 연구로 Deepcut(2016, CVPR) 가 존재하나, 개별 이미지를 처리하는데 몇 초에서 이미지에 따라 수백 초까지 걸리는 문제점이 존재
- ✓ 따라서 본 논문은 이러한 속도 문제를 대폭 개선했으며, 단일 이미지를 넘어서 비디오 영상 처리에서도 좋은 성능을 보임



Bottom-up approach

Multi Person Pose Estimation

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

❖ Concept

- Convolutional Pose Machines(CPM)을 통해 이미지 내 사람의 관절 위치 추정은 가능
- 그러나, 동일한 인물 내 관절끼리 연결하는 것은 어려운 문제



Multi Person Pose Estimation

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

❖ Concept

- Convolutional Pose Machines(CPM)을 통해 이미지 내 사람의 관절 위치 추정은 가능
- 그러나, 동일한 인물 내 관절끼리 연결하는 것은 어려운 문제
 - ✓ 관절 사이 중간 지점(midpoint)을 생성하여 해결하려 했으나, 이미지가 복잡한 경우, 연결할 수 있는 여러 경우가 발생(False association)



False association

Multi Person Pose Estimation

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

❖ Concept

- **Part Affinity Fields(PAFs)**
 - ✓ Affinity fields란 **2차원 벡터 공간을 의미하며, 위치 정보와 방향 정보로 구성**
- PAFs를 잘 추정한다면, **동일 인물 내 관절 간 연결을 올바르게 표현 가능**
- 두 관절 사이의 연결(limb)을 2차원 벡터로 인코딩할 수 있는 필터를 학습



Multi Person Pose Estimation

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

❖ Architecture (2019 TPAMI 버전)



Input Image



Part Affinity Fields



Part Confidence Maps



Bipartite Matching



Parsing Results

Multi Person Pose Estimation

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

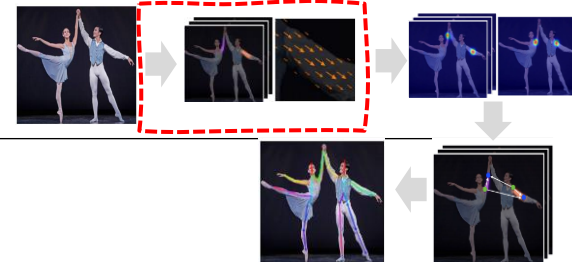
❖ Architecture (2019 TPAMI 버전)

- 입력 이미지를 VGG-19 모델을 통해 Feature map을 추출
 - ✓ 2017년 초기 버전 대비 VGG-19 모델을 통해 커널 사이즈를 줄이면서 연산량 감소



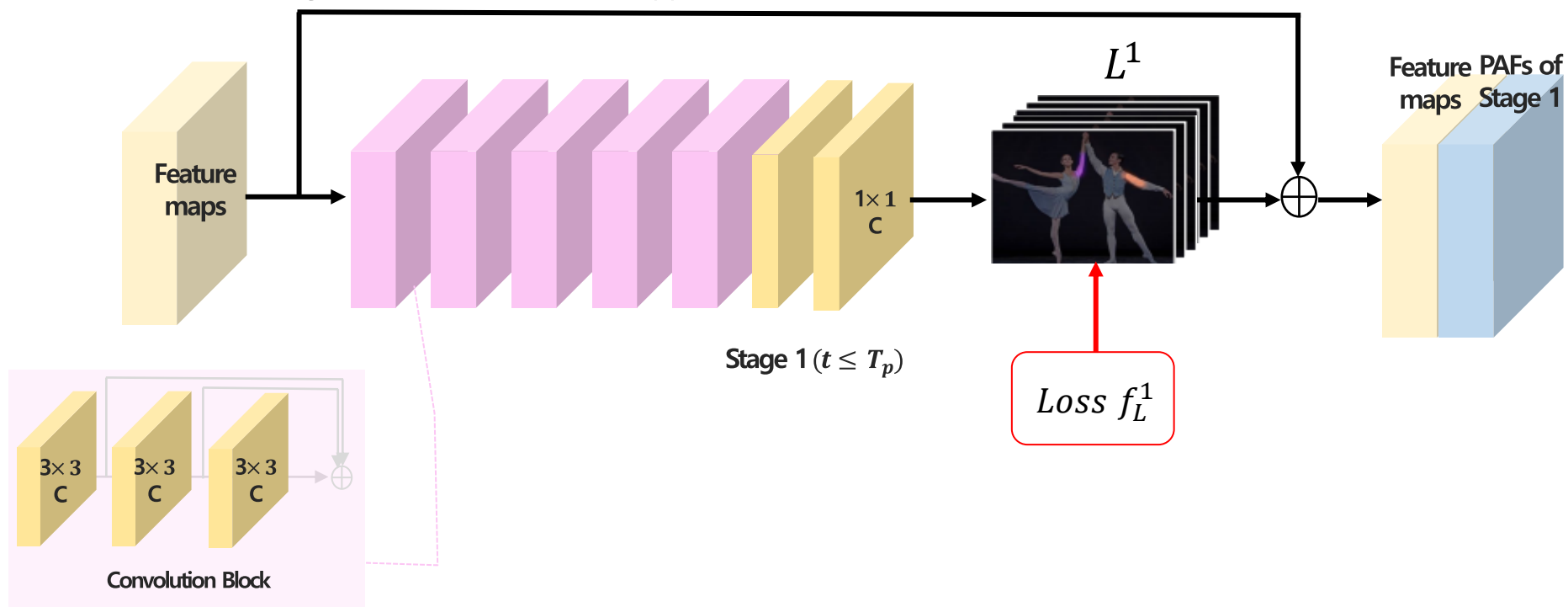
Multi Person Pose Estimation

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



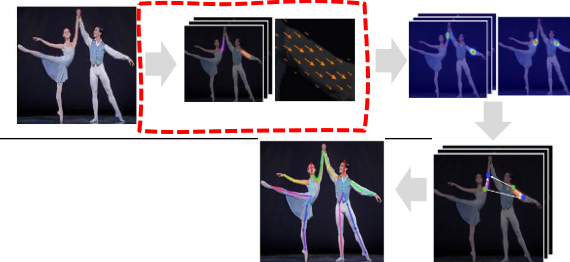
❖ Architecture – Part Affinity Fields

- VGG-19를 통해 추출한 feature maps을 입력 변수로 Part Affinity Fields(PAFs)부터 추정
- 5개의 convolution block과 2개의 1x1 필터를 사용하는 convolution layer를 거쳐 PAFs를 생성
- 생성된 PAFs와 PAFs ground truth 값과 손실 값 계산, 이 때 손실 함수는 MSE 를 사용
 - ✓ PAFs ground truth 생성 과정은 appendix 참조



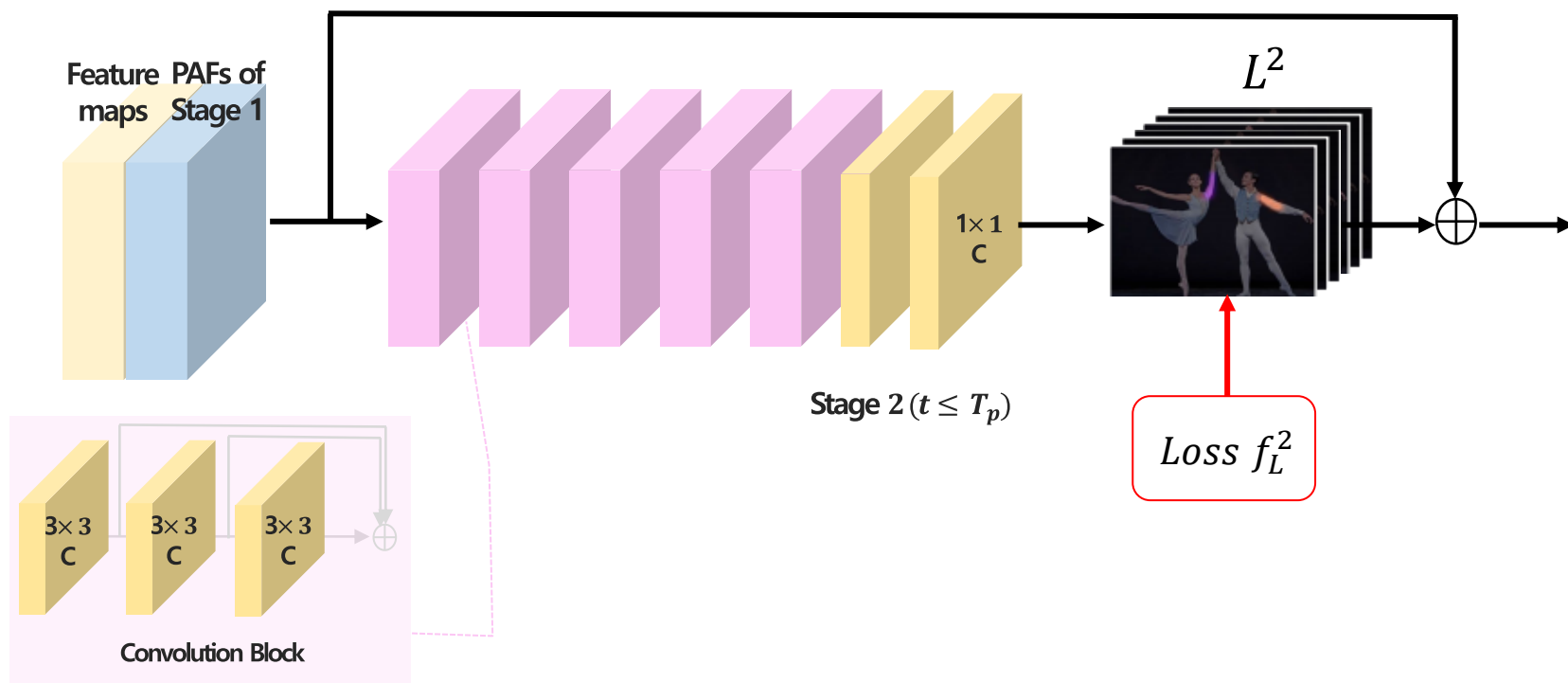
Multi Person Pose Estimation

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



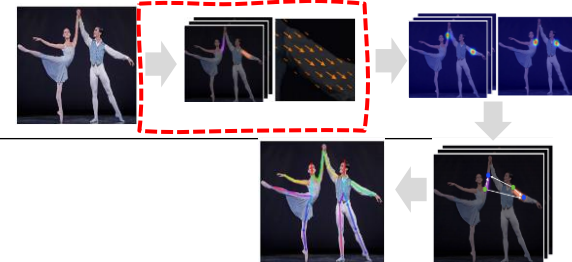
❖ Architecture – Part Affinity Fields

- 두번째 stage부터 입력 변수를 feature maps과 이전 단계의 stage에서 나온 PAFs를 concat하여 사용



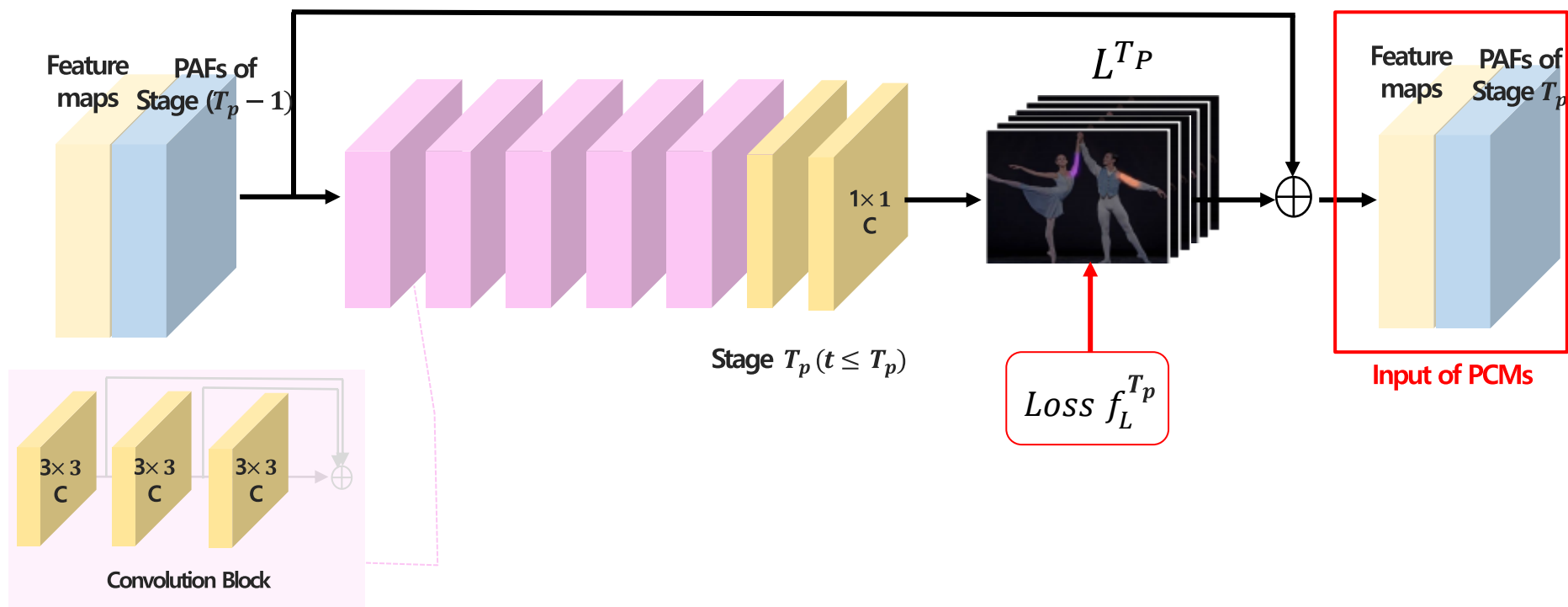
Multi Person Pose Estimation

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



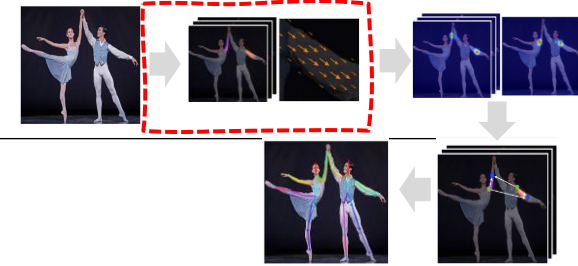
❖ Architecture – Part Affinity Fields

- 최종 stage를 거쳐 나온 PAFs(L^{T_P})는 Part Confidence Maps(PCMs) 네트워크의 입력 변수로 활용



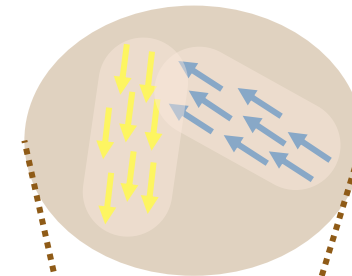
Multi Person Pose Estimation

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



❖ Architecture – Part Affinity Fields

- Stage를 지남에 따라 생성된 팔뚝(right forearm) PAFs
- Stage를 거듭할수록 정밀한 결과를 볼 수 있음



Stage 1

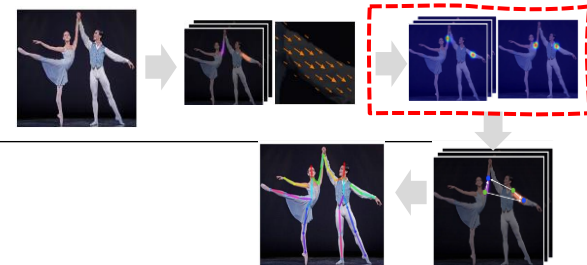


Stage 2



Stage 3

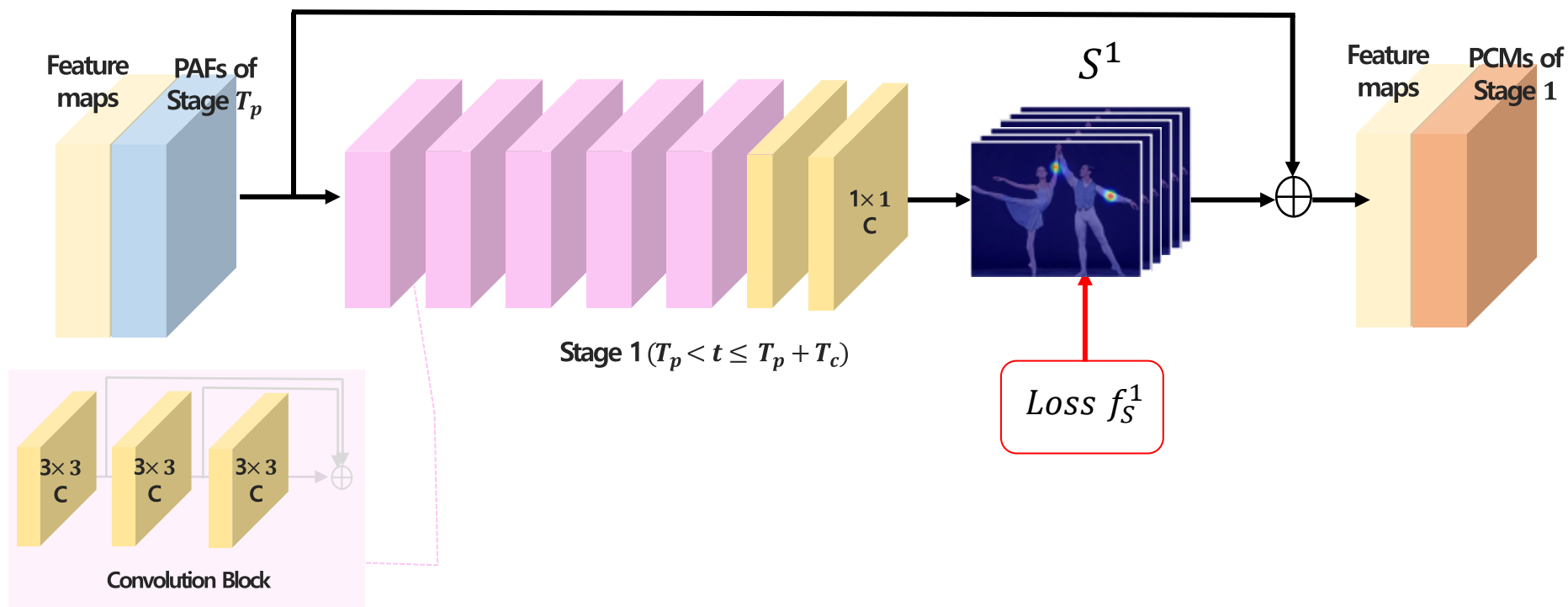
Multi Person Pose Estimation



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

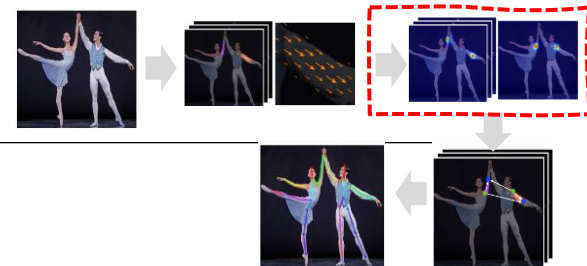
❖ Architecture – Part Confidence Maps

- PAFs와 동일한 네트워크 과정을 통해 $PCMs(S^t)$ 를 추출
- Convolutional Pose Machines의 heatmap과 PCM은 동일하며, 같은 손실 함수를 사용



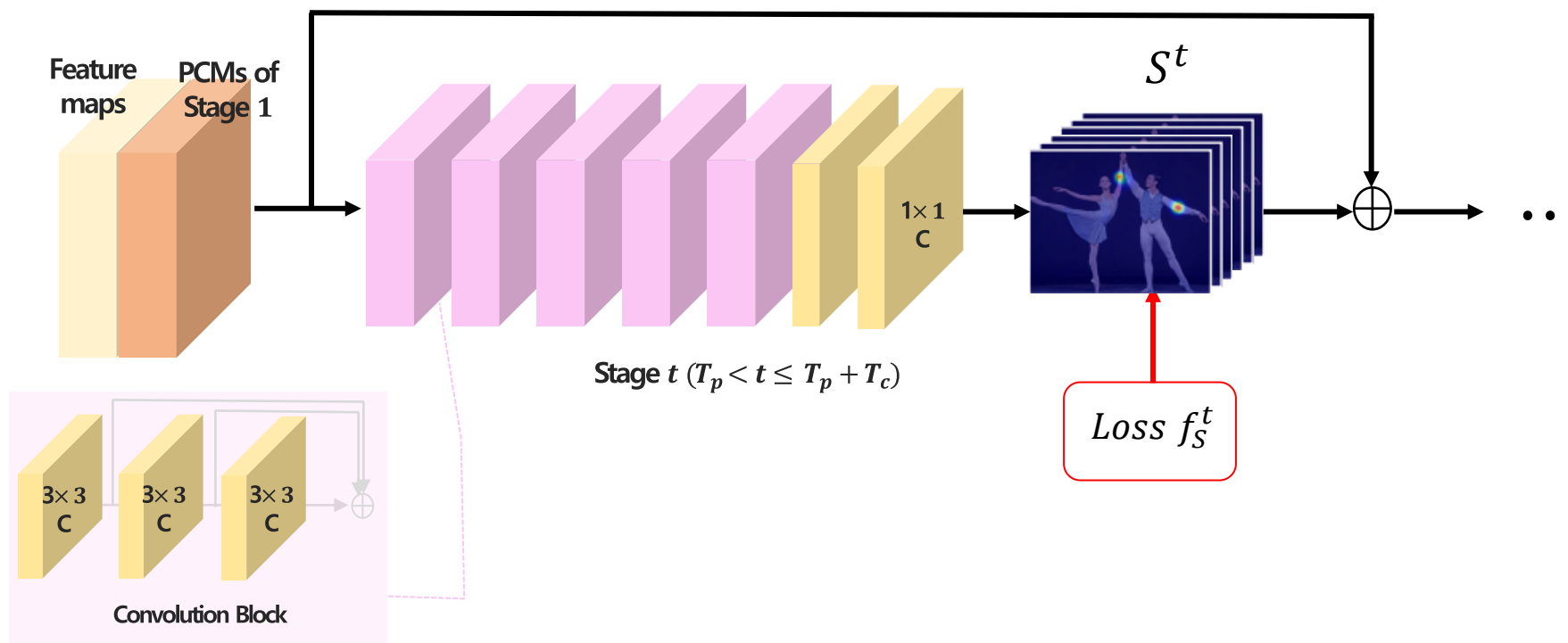
Multi Person Pose Estimation

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



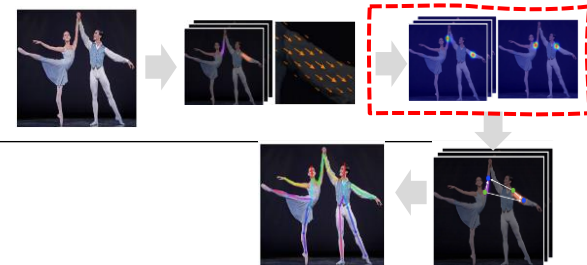
❖ Architecture – Part Confidence Maps

- PAFs와 동일한 네트워크 과정을 통해 $PCMs(S^t)$ 를 추출
- Convolutional Pose Machines의 heatmap과 PCM은 동일하며, 같은 손실 함수를 사용



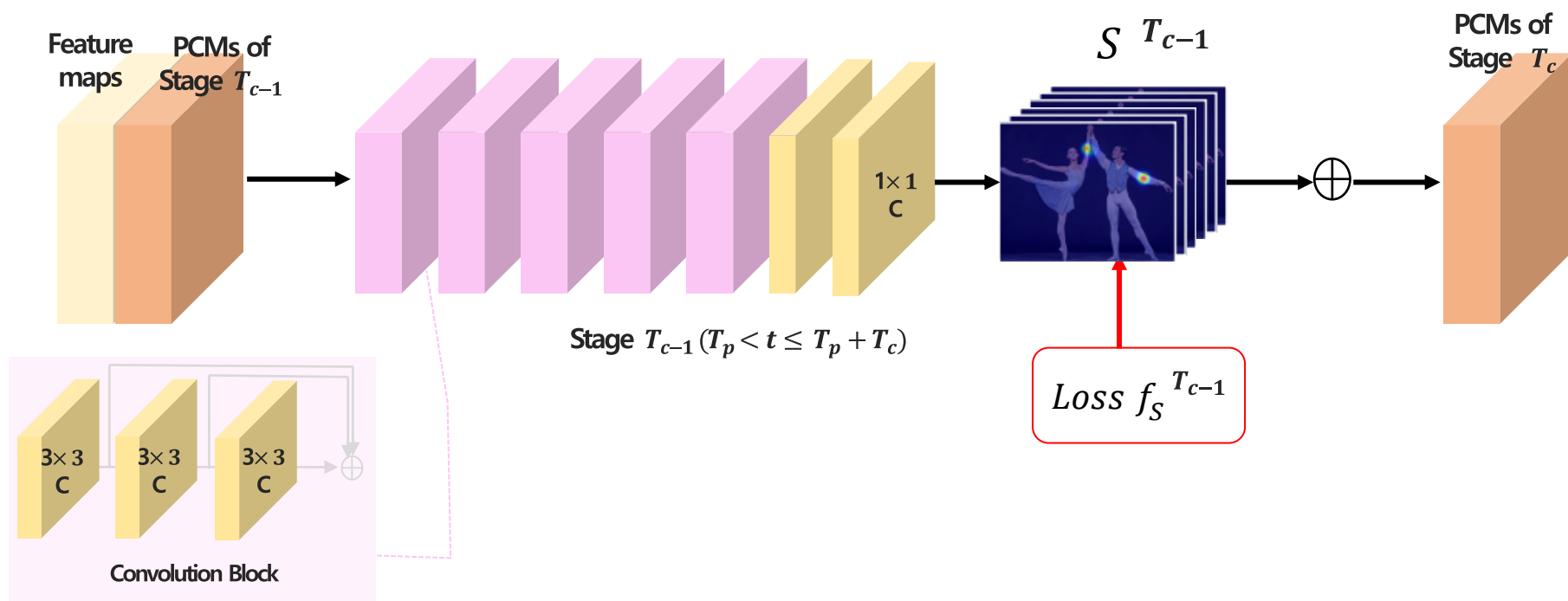
Multi Person Pose Estimation

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



❖ Architecture – Part Confidence Maps

- PAFs와 동일한 네트워크 과정을 통해 $PCMs(S^t)$ 를 추출
- Convolutional Pose Machines의 heatmap과 PCM은 동일하며, 같은 손실 함수를 사용
- 최종 $PCMs(S^{T_c})$ 를 구하여 관절들의 위치를 추정



Multi Person Pose Estimation

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

❖ Architecture (2019 TPAMI 버전)



Input Image



Part Affinity Fields



Part Confidence Maps



Parsing Results



Bipartite Matching

Appendix 참조

Multi Person Pose Estimation

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

❖ Results

- Bottom-up 접근 방법의 대표적 baseline인 Deepcut에 비해 높은 정확도와 빠른 연산 처리 성능을 보여줌
- Top-down 접근 방법의 대표적 baseline인 Mask-R-CNN에 대비 부족한 성능을 보여줌

Method	<i>mAP</i>	<i>s/image</i>
DeepCut	59.5	485
OpenPose	75.6	0.005

MPII dataset

Method	<i>AP</i>	<i>AP^L</i>
Mask R-CNN	69.2	76.3
OpenPose	64.2	68.2

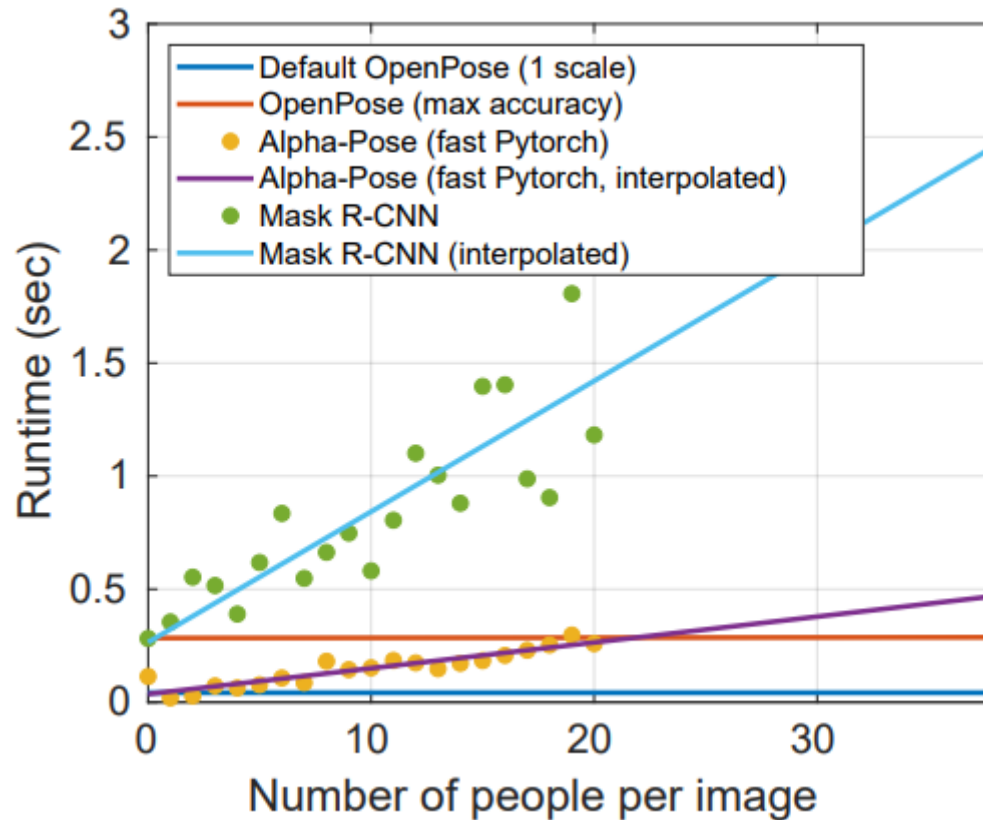
COCO dataset

Multi Person Pose Estimation

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

❖ Results

- 그러나 이미지 내 사람의 수의 증가에 따른 연산 속도 측면에서 OpenPose가 압도적인 결과를 보여줌



Multi Person Pose Estimation

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

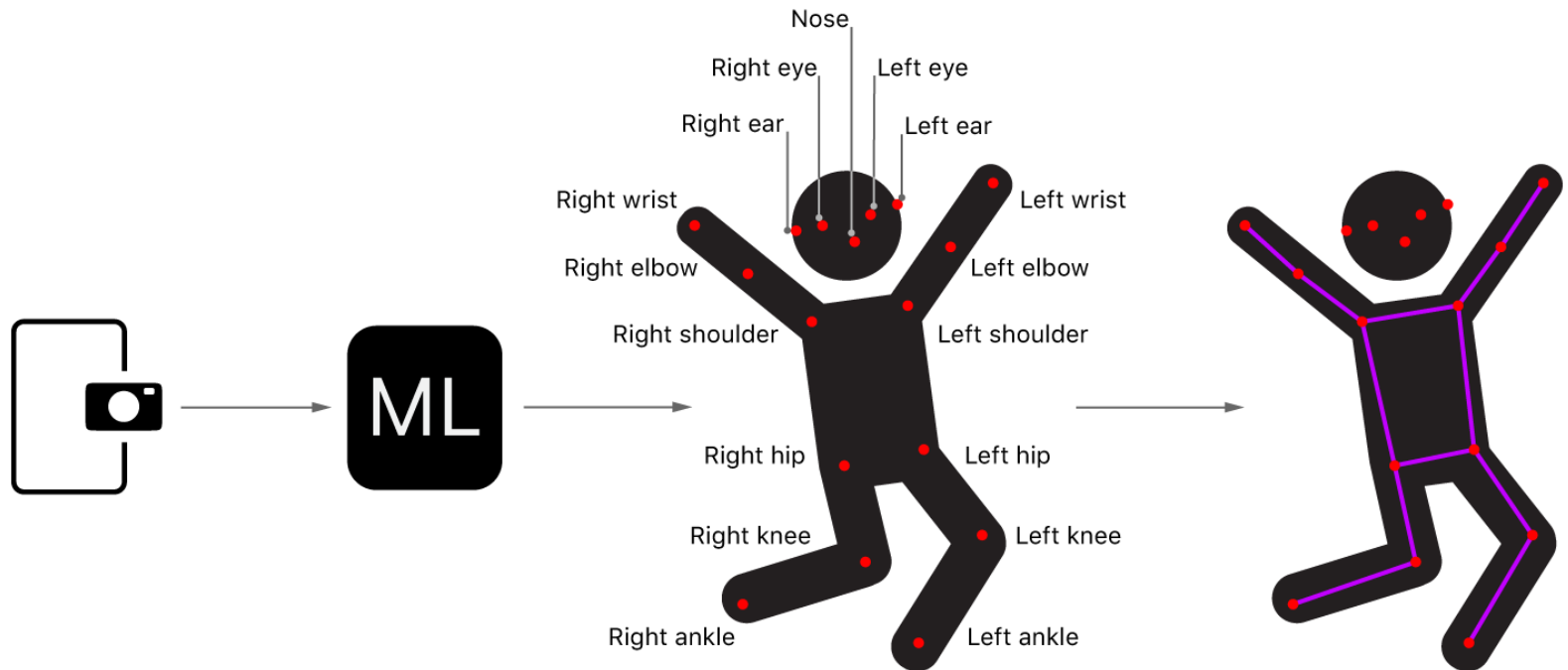
❖ Results



Conclusion

❖ 결론

- Human pose estimation: 인간의 관절별 좌표를 예측하는 문제
 - ✓ 입력 데이터는 일반적으로 RGB 이미지
 - ✓ 출력 데이터는 인간에 해당하는 관절별 좌표
- 해당 문제는 Single person estimation과 Multi-person estimation으로 구분

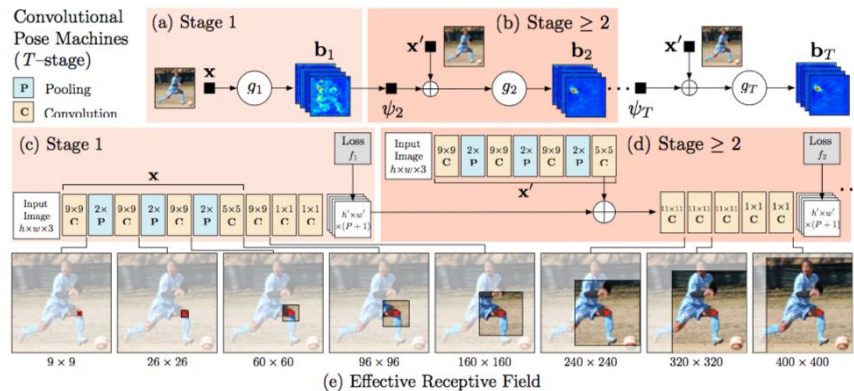


Conclusion

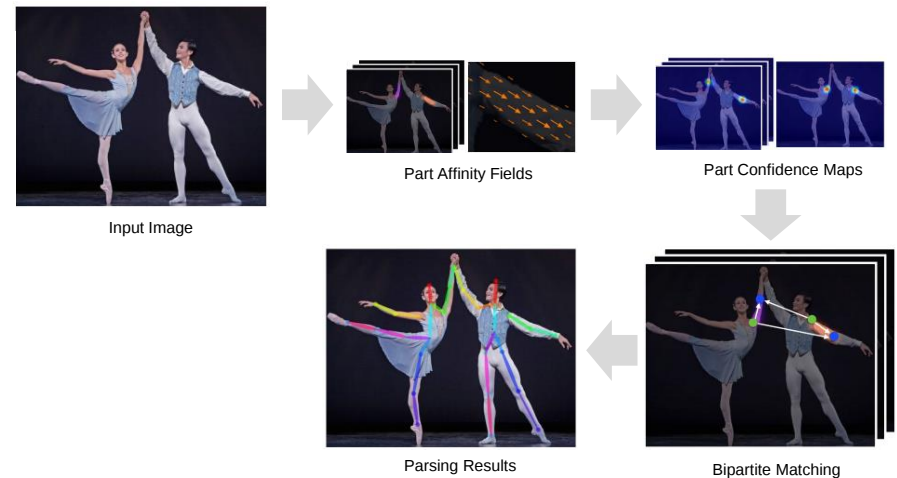
❖ 결론

- Human pose estimation: 인간의 관절별 좌표를 예측하는 문제
 - ✓ 입력 데이터는 일반적으로 RGB 이미지
 - ✓ 출력 데이터는 인간에 해당하는 관절별 좌표
- 해당 문제는 Single person estimation과 Multi-person estimation으로 구분
 - ✓ Single person estimation 대표 알고리즘: Convolutional Pose Machines(2016)
 - ✓ Multi-person estimation 대표 알고리즘: OpenPose (2017, 2019)

Single person estimation



Multi-person estimation



Reference

1. Dang, Q., Yin, J., Wang, B., & Zheng, W. (2019). Deep learning based 2d human pose estimation: A survey. *Tsinghua Science and Technology*, 24(6), 663-676.
2. Chen, Y., Tian, Y., & He, M. (2020). Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*, 192, 102897.
3. Ramakrishna, V., Munoz, D., Hebert, M., Bagnell, J. A., & Sheikh, Y. (2014, September). Pose machines: Articulated pose estimation via inference machines. In *European Conference on Computer Vision* (pp. 33-47). Springer, Cham.
4. Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 4724-4732).
5. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), 172-186.

감사합니다

Appendix

Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향·위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결



PAFs



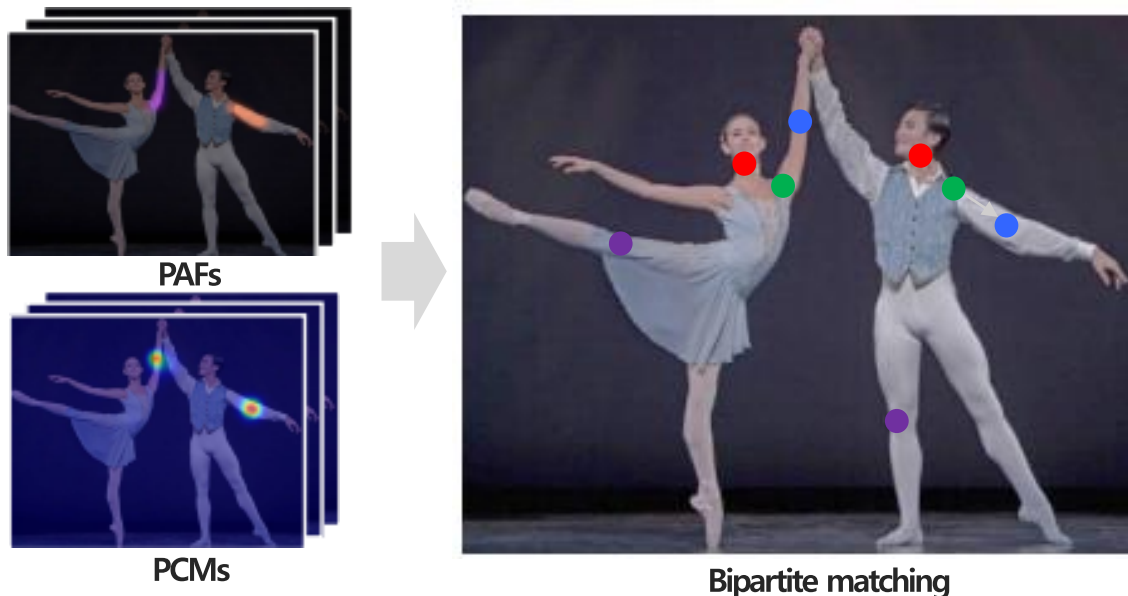
PCMs

Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향·위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결

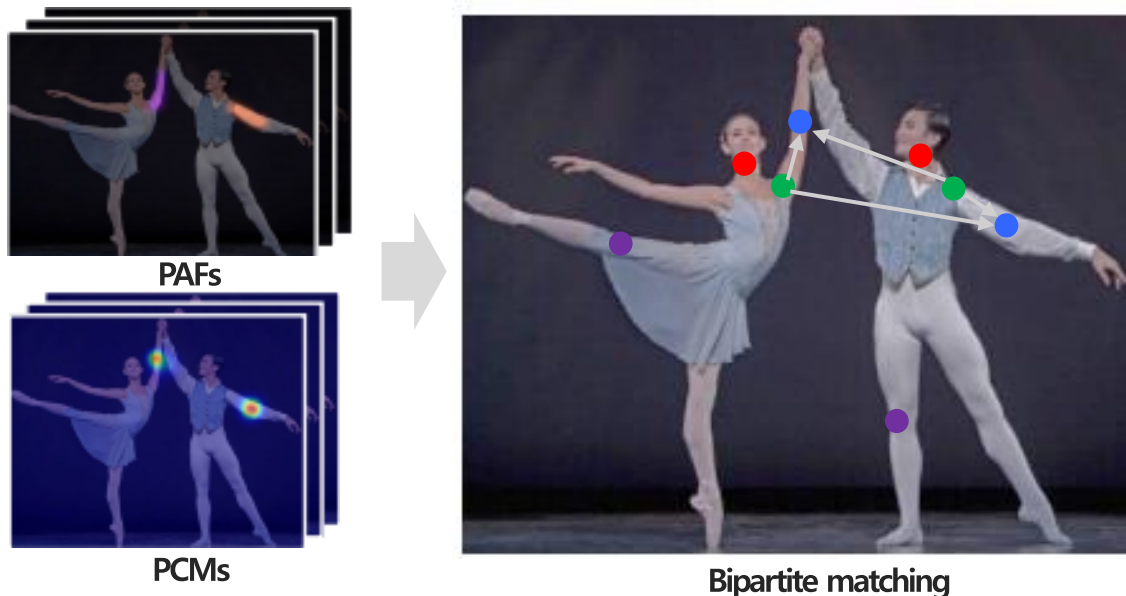


Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향 · 위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결

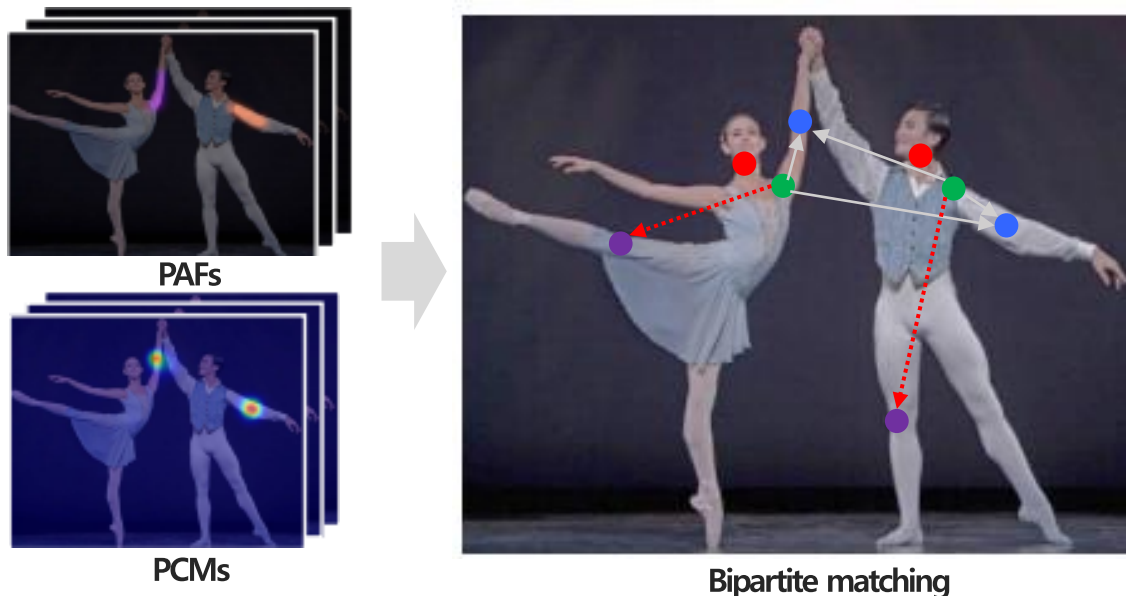


Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향 · 위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결

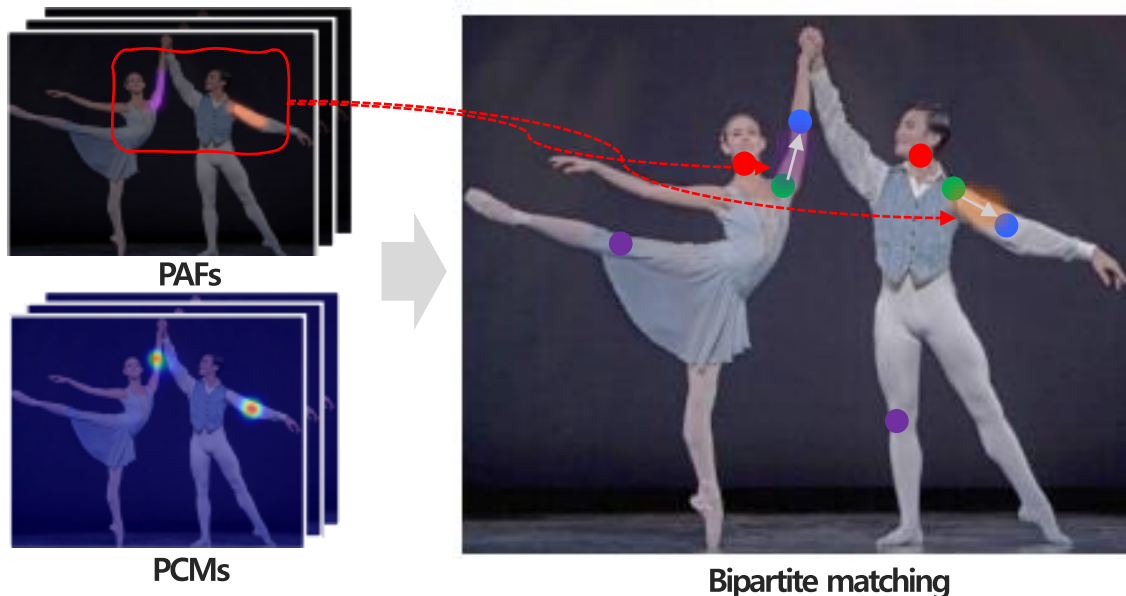


Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향·위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결

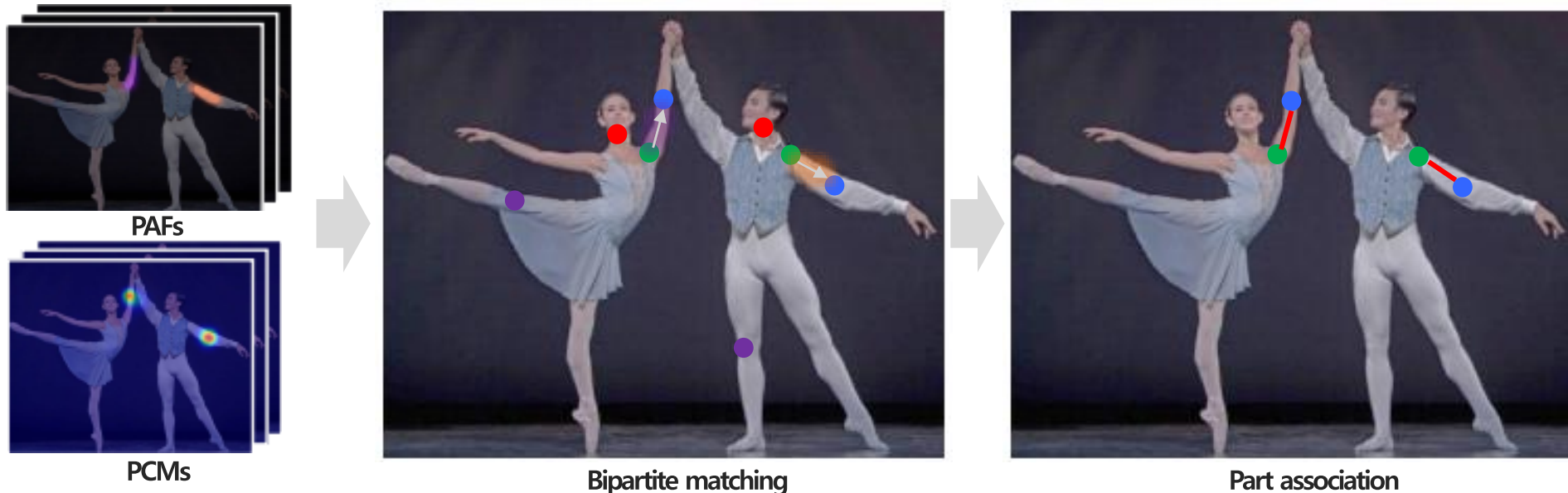


Appendix

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

❖ Architecture -Bipartite Matching(이분 매칭)

- Multi-stage CNN을 통해 관절 위치(PCMs)와 관절 사이의 방향·위치 정보(PAFs)를 생성함
 - ✓ PCMs 과 PAFs 를 통해 관절 candidate를 추정
- 동일 인물 내 관절 간의 연결을 위해 Weighted Bipartite matching 기법을 사용하여 연산량 감소
 - ✓ 관절을 node, 관절 간의 연결을 edge로 설정하여 이웃한 관절 두개씩 매칭
 - ✓ 인체이기에 연결 가능한 관절 쌍은 정해져 있음 ex) 왼쪽 어깨와 오른쪽 무릎 간 연결은 불가능
 - ✓ PAFs 스코어를 산출하여 가장 높은 값을 가지는 관절 쌍을 선택하여 연결



Appendix

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

❖ Architecture - Merging

- Bipartite matching을 통해 구한 관절 쌍이 어떤 사람의 관절인지 결정하는 문제
 - ✓ 사람 한 명당 하나의 관절 쌍만 가진다는 가정
- 구한 모든 관절의 연결 수 만큼 사람이 존재한다고 가정하여 실제 이미지 내 존재하는 사람의 수 만큼 줄여나가는 과정
 - ✓ N개의 connection이면, N명의 사람이 존재한다고 가정
 - ✓ Connection 두 개가 공유하는 관절이 존재하면, 이를 한 사람의 관절이라고 판단
 - ✓ 상기 과정을 반복하여 실제 존재하는 사람의 수만큼 줄여 나감



관절 연결 수 = 사람 수 = 4

Appendix

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

❖ Architecture - Merging

- Bipartite matching을 통해 구한 관절 쌍이 어떤 사람의 관절인지 결정하는 문제
 - ✓ 사람 한 명당 하나의 관절 쌍만 가진다는 가정
- 구한 모든 관절의 연결 수 만큼 사람이 존재한다고 가정하여 실제 이미지 내 존재하는 사람의 수 만큼 줄여나가는 과정
 - ✓ N개의 connection이면, N명의 사람이 존재한다고 가정
 - ✓ Connection 두 개가 공유하는 관절이 존재하면, 이를 한 사람의 관절이라고 판단
 - ✓ 상기 과정을 반복하여 실제 존재하는 사람의 수만큼 줄여 나감



관절 연결 수 = 사람 수 = 4



관절 연결 수 - 공유하는 관절의 수 = 2

Appendix

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

❖ Architecture - Merging

- Bipartite matching을 통해 구한 관절 쌍이 어떤 사람의 관절인지 결정하는 문제
 - ✓ 사람 한 명당 하나의 관절 쌍만 가진다는 가정
- 구한 모든 관절의 연결 수 만큼 사람이 존재한다고 가정하여 실제 이미지 내 존재하는 사람의 수 만큼 줄여나가는 과정
 - ✓ N개의 connection이면, N명의 사람이 존재한다고 가정
 - ✓ Connection 두 개가 공유하는 관절이 존재하면, 이를 한 사람의 관절이라고 판단
 - ✓ 상기 과정을 반복하여 실제 존재하는 사람의 수만큼 줄여 나감



관절 연결 수 = 사람 수 = 4



관절 연결 수 - 공유하는 관절의 수 = 2



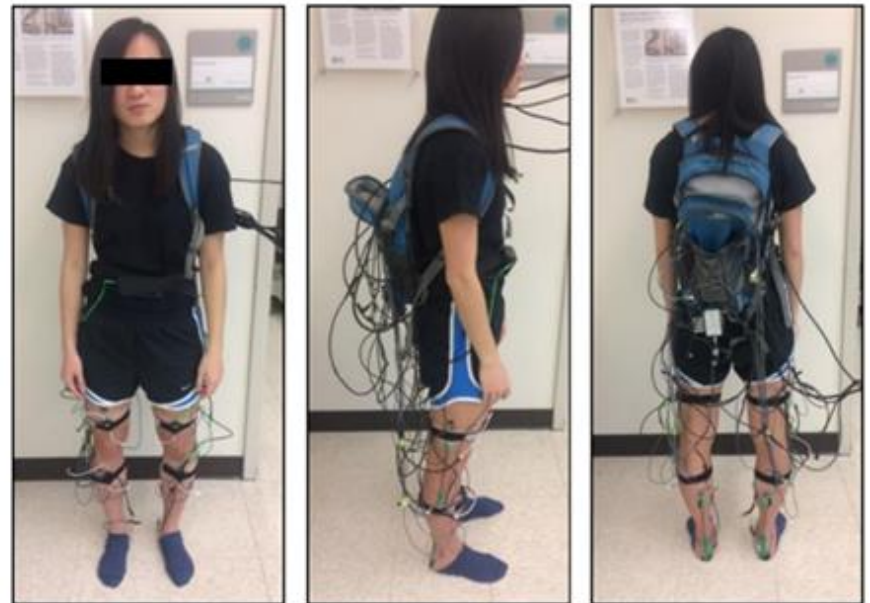
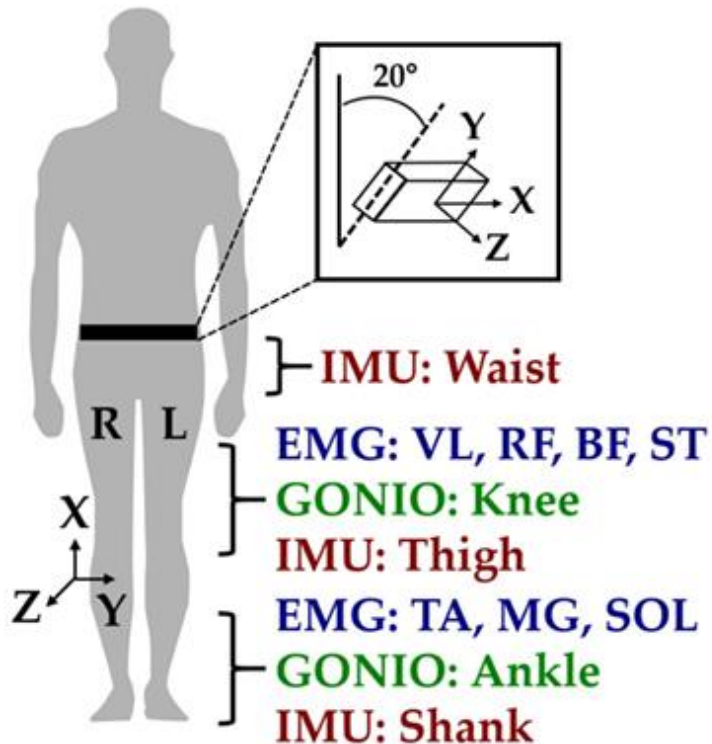
Parsing Results

Appendix

Human Action Recognition

❖ Human Action Recognition(HAR)

- 비디오 영상 내 사람 탐지, 사람 자세 추정, 시계열 데이터를 통한 분석 등 많은 부분을 다룸
- Action classification은 사전에 정해진 행동 범주 중 하나의 행동으로 분류
- Action detection은 행동의 시작과 끝 지점을 탐지, 공간 상의 위치, 행동 분류 등을 수행



센서 데이터를 통한 행동 분류 실험

Appendix

Human Action Recognition

❖ Human Action Recognition(HAR)

- 비디오 영상 내 사람 탐지, 사람 자세 추정, 시계열 데이터를 통한 분석 등 많은 부분을 다룸
- Action classification은 사전에 정해진 행동 범주 중 하나의 행동으로 분류
- Action detection은 행동의 시작과 끝 지점을 탐지, 공간 상의 위치, 행동 분류 등을 수행

