DeepPose:
Human Pose Estimation via Deep Neural Networks

21.5.20
UGRP seminar
곽현우

Abstract

- DeepPose: Human Pose Estimation via Deep Neural Networks
  - 2014년 CVPR에서 발표
  - Deep Learning을 Human Pose Estimation 분야에 최초로 적용한 논문

DeepPose: Human Pose Estimation via Deep Neural Networks

Alexander Toshev Christian Szegedy
Google
1600 Amphitheatre Pkwy
Mountain View, CA 94043
toshev,szegedy@google.com

Figure 1. Besides extreme variability in articulations, many of the joints are barely visible. We can guess the location of the right arm in the left image only because we see the rest of the pose and anticipate the motion or activity of the person. Similarly, the left body half of the person on the right is not visible at all. These are examples of the need for *holistic reasoning*. We believe that DNNs can naturally provide such type of reasoning.

Abstract

We propose a method for human pose estimation based on Deep Neural Networks (DNNs). The pose estimation is formulated as a DNN-based regression problem towards body joints. We present a cascade of such DNN regressors which results in high precision pose estimates. The approach has the advantage of reasoning about pose in a holistic fashion and has a simple but yet powerful formulation which capitalizes on recent advances in Deep Learning. We present a detailed empirical analysis with state-of-art or better performance on four academic benchmarks of diverse real-world images.

mainly by the first challenge, the need to search in the large space of all possible articulated poses. Part-based models lend themselves naturally to model articulations ([16, 8]) and in the recent years a variety of models with efficient inference have been proposed ([6, 19]).

The above efficiency, however, is achieved at the cost of limited expressiveness – the use of local detectors, which reason in many cases about a single part, and most importantly by modeling only a small subset of all interactions between body parts. These limitations, as exemplified in Fig. 1, have been recognized and methods reasoning about pose in a holistic manner have been proposed [15, 21] but with limited success in real-world problems.

In this work we ascribe to this holistic view of human pose estimation. We capitalize on recent developments of deep learning and propose a novel algorithm based on a Deep Neural Network (DNN). DNNs have shown outstanding performance on visual classification tasks [14] and more recently on object localization [23, 9]. However, the question of applying DNNs for precise localization of articulated objects has largely remained unanswered. In this paper we attempt to cast a light on this question and present a simple and yet powerful formulation of *holistic human pose estimation* as a DNN.

We formulate the pose estimation as a joint regression problem and show how to successfully cast it in DNN settings. The location of each body joint is regressed to using as an input the full image and a 7-layered generic convolutional DNN. There are two advantages of this formulation.

DeepPose – Data Input(Data Processing)

- Bounding Box를 이용해 사람이 존재하는 영역 추출
- 행렬 변환으로 image를 고정된 크기로 변환

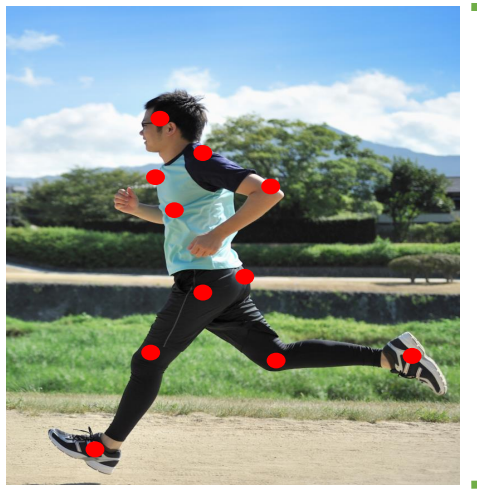
Bounding Box(x,y,w,h)

Normalize Process

$$N(\mathbf{y}_i; b) = \begin{pmatrix} 1/b_w & 0 \\ 0 & 1/b_h \end{pmatrix} (\mathbf{y}_i - b_c)$$

i: 관절 인덱스
w: Bbox 너비
h: Bbox 높이
y_i: i번째 관절에 대한 좌표
b_c: Bbox 중심 좌표

220



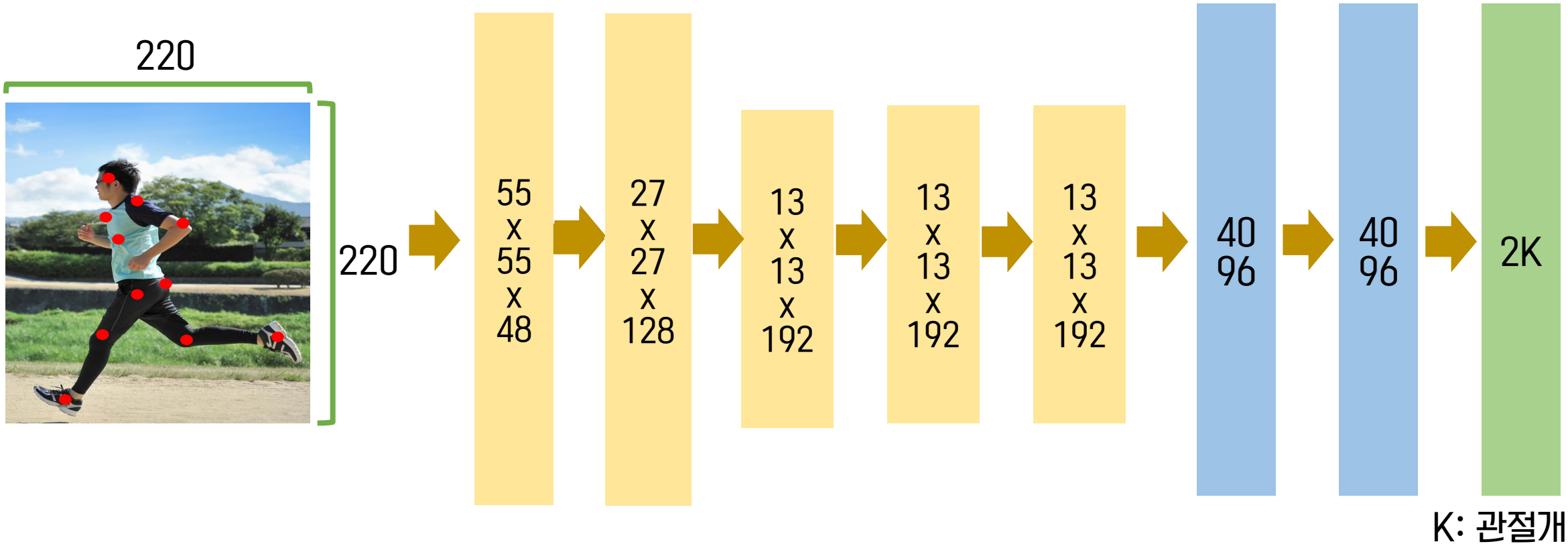
220

DeepPose – Model structure

- 처리된 입력 데이터를 이용해 예측값을 산출하는 모델
- AlexNet 기반으로 Representation 벡터 산출
- 마지막에 Fully Connected Layer를 통해 (2 x 관절 개수)의 예측 벡터 산출

Loss Function

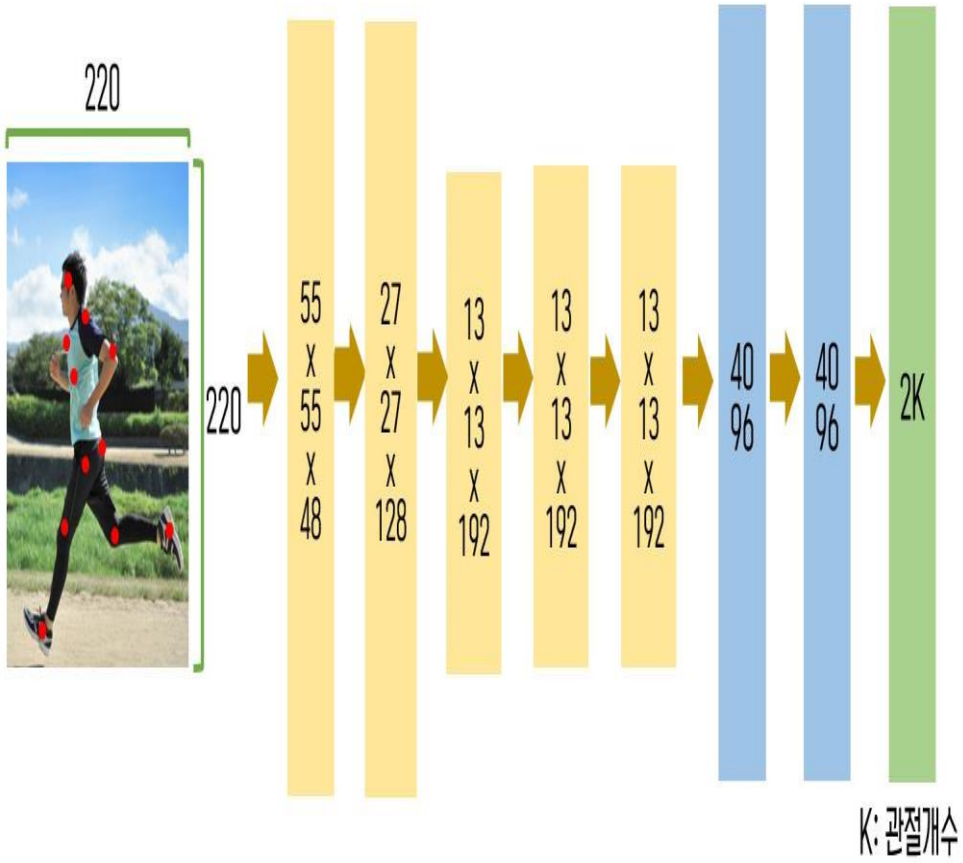
$$MSE = \frac{1}{2k}\sum_{i=1}^{2k}(y_i - \hat{y}_i)^2$$



DeepPose – Training

Stage 1
관절별 대략적인 위치 예측

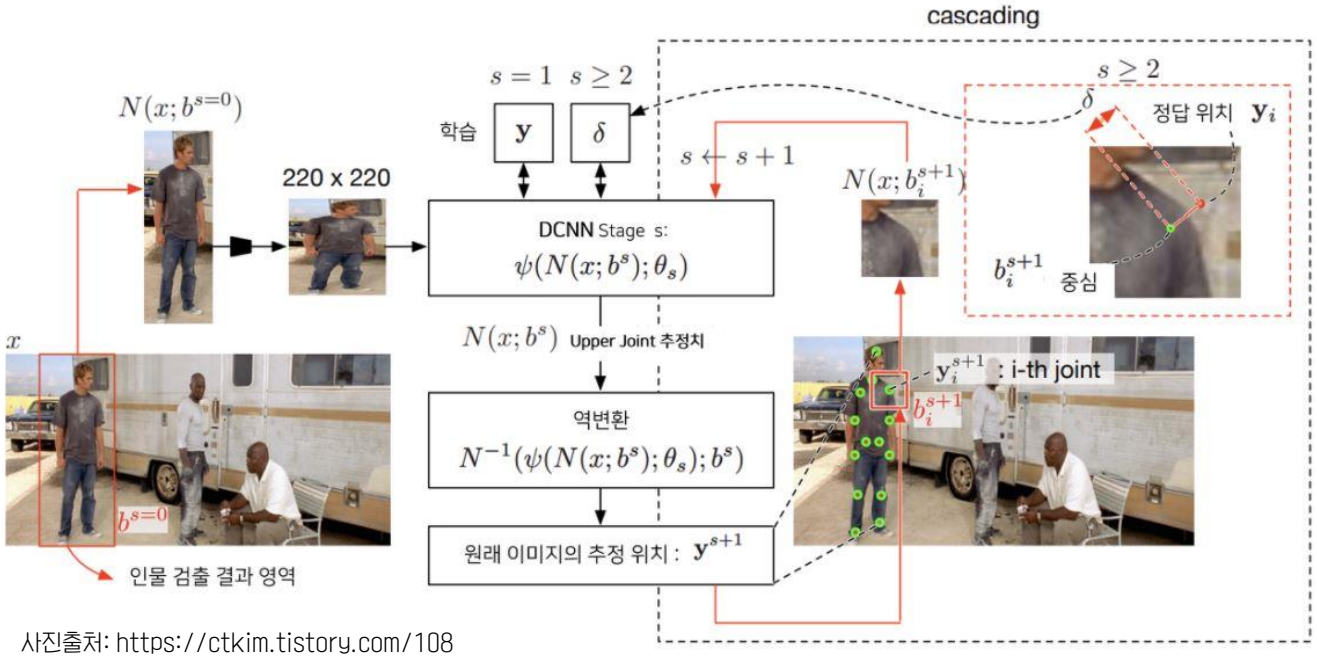
$$\text{Stage 1}: \quad \mathbf{y}^1 \leftarrow N^{-1}(\psi(N(x;b^0);\theta_1);b^0)$$



Stage s(s >= 2)
관절별 개별적위치 예측모델 학습

$$\text{Stage } s: \quad \mathbf{y}_i^s \leftarrow \mathbf{y}_i^{(s-1)} + N^{-1}(\psi_i(N(x;b);\theta_s);b)$$
$$\text{for } b = b_i^{(s-1)}$$
$$b_i^s \leftarrow (\mathbf{y}_i^s, \sigma\text{diam}(\mathbf{y}^s), \sigma\text{diam}(\mathbf{y}^s))$$

실제 이미지 내에서
왼쪽 어깨 예측값과
오른쪽 엉덩이 예측값 거리



사진출처: https://ctkim.tistory.com/108

DeepPose – Evaluate metric

- PDJ(Percent of Detected Joints) 지표

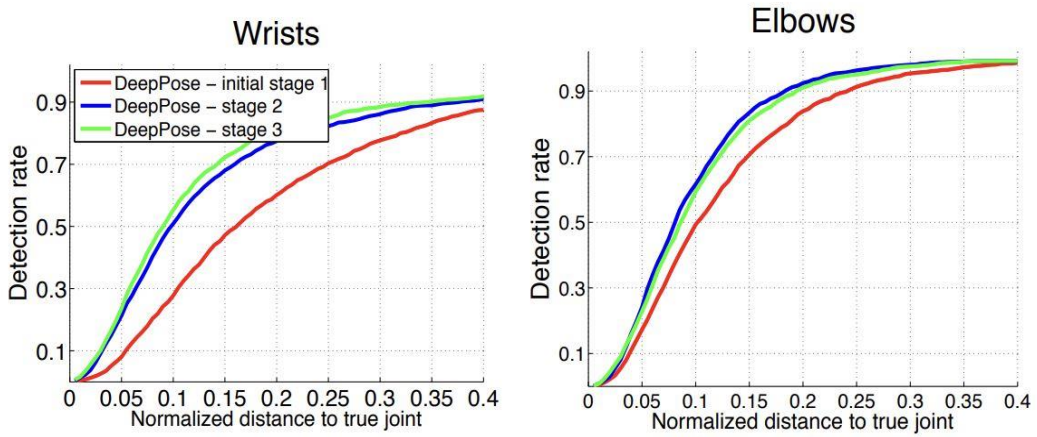
- 몸통 길이 계산(노란색 화살표)
- 특정 임계값(σ) * 몸통 길이가 반지름인 원 생성
- 예측 위치가 원 내부에 있는지 확인
  - 원 내부에 예측 위치 : 1(옳은 결과)
  - 원 외부에 예측 위치 : 0(틀린 결과)
- 예측 결과에 대해 PDJ 계산

*PDJ

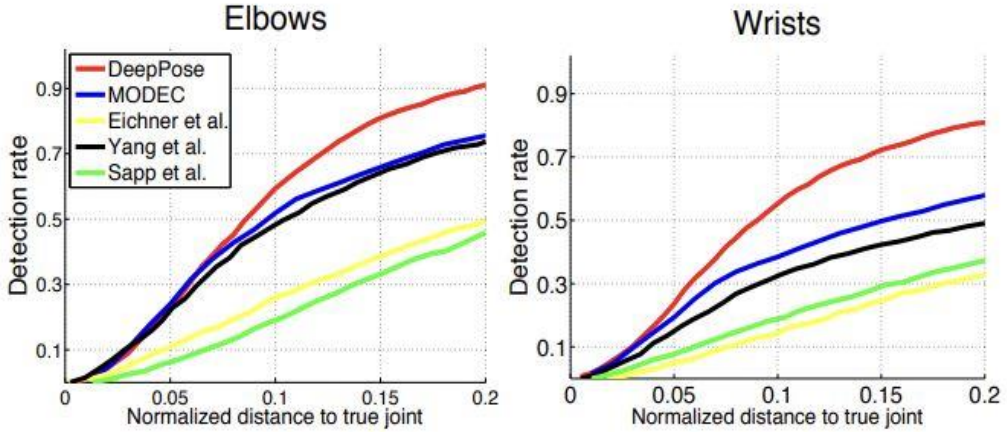
:원 내부에 위치한 예측 관절 위치 수 / 전체 관절 수

DeepPose – Result

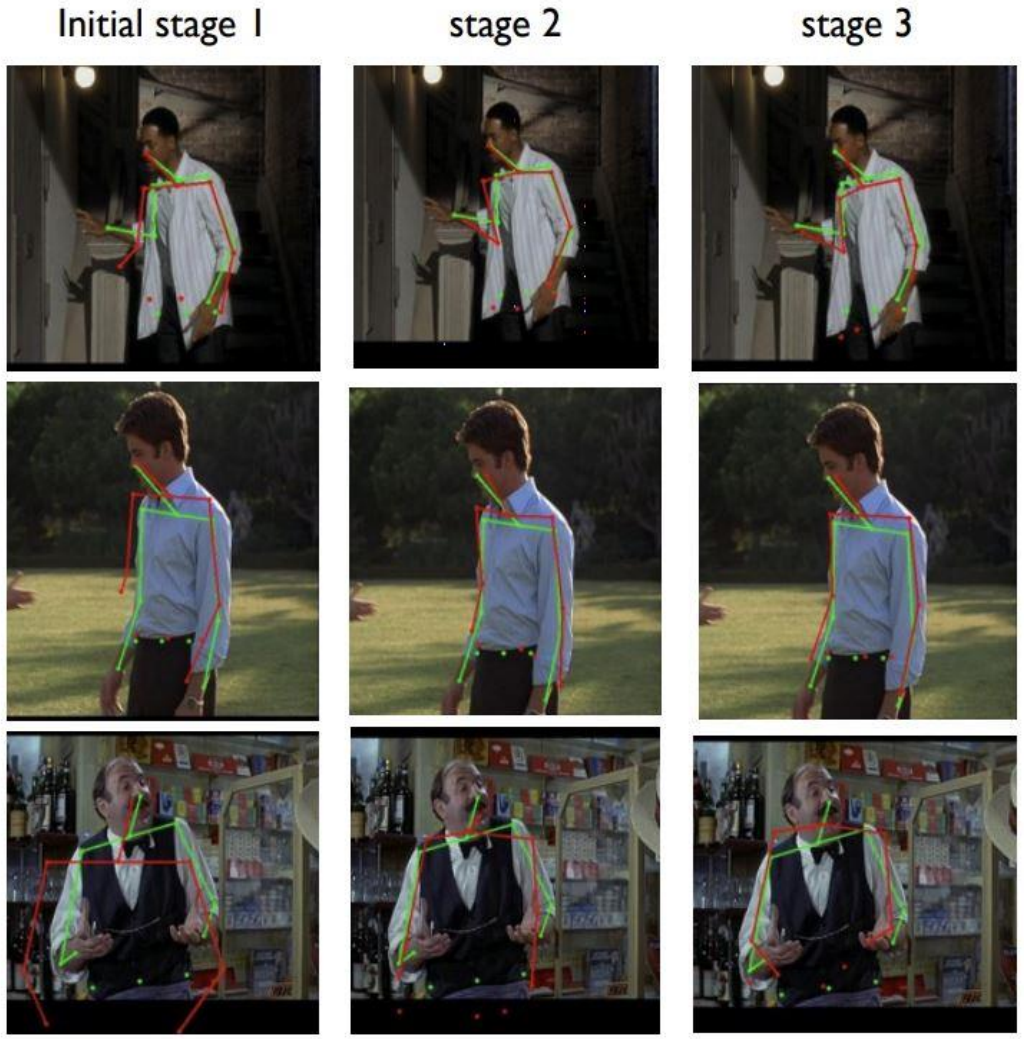
Performance according to stage



Compare with other SOTA



Visualize according to stage



Reference

- DeepPose: Human Pose Estimation via Deep Neural Networks
  https://arxiv.org/abs/1312.4659

- DMQA Open Seminar: Human Pose Estimation
  https://www.youtube.com/watch?v=pK8PeoCLnQc

- Implement Github(not official)
  https://github.com/mitmul/deeppose