Primate pose estimation with OpenMonkeyChallenge

Chahyon Ku, Gustav Baumgart, Max Scheder-Bieschin, Josh Spitzer-Resnick

Oct 31, 2022

University of Minnesota
Driven to Discover℠

Overview

- Motivation
- Applications
- Related work
- Dataset
- Methodology
  - Baseline method
  - Proposed method
- Current status
- Next steps

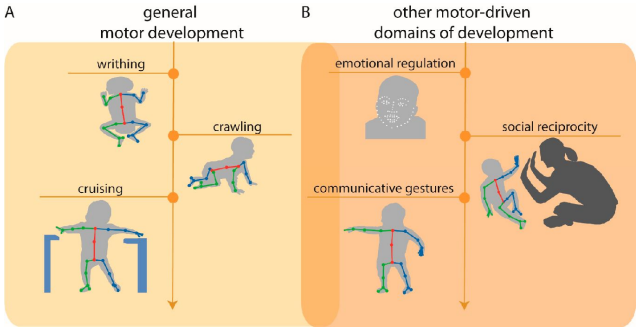
Motivation: 2D pose estimation

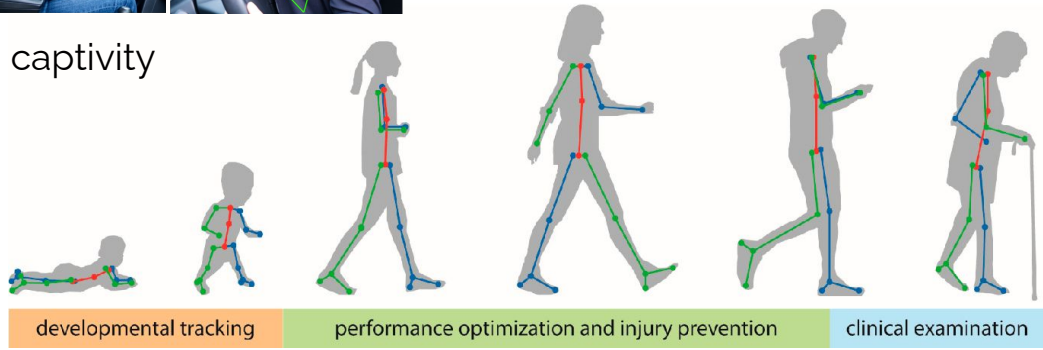
- detect ***keypoint (joint) pixel location*** of animate objects
- help us understand ***movement***
- has been studied extensively for humans, less so for other species
- issues in ***robustness*** of previous models
  - limited scope of previous dataset collection
  - environment, species, # of landmarks
  - expensive to annotate
- many applications...

- this dataset: 2D, single pose

Applications

- humans
  - **healthcare** - assisted living
  - assisted driver systems
  - **virtual reality**, video games
  - sports
- primates
  - monitoring health in the wild + in captivity
  - understanding **social behaviors**



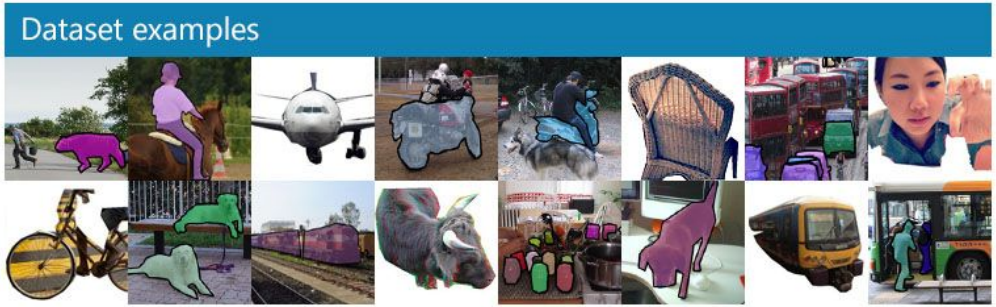


A survey on Pose Estimation using Deep Convolutional Neural Networks (2021)
Applications of Pose Estimation in Human Health and Performance across the Lifespan (2021)
POSEidon: Face-from-Depth for Driver Pose Estimation (2017)
Human detection and Pose Estimation with Deep Learning for Sport Analysis (2018)

Related work

- 2D **Human** Pose Estimation
  - Pioneered by Google
  - Datasets: **MPII** and **COCO**
  - CPMs, HR-Net, RSN, ViTPose

Microsoft COCO: Common Objects in Context (2014)



2D Human Pose Estimation: New Benchmark and State of the Art Analysis (2014)

Related work

- 2D **Non-human** Pose Estimation
  - More limited than 2D human pose estimation
  - Sometimes limited to **macaque** monkeys

Some top search results for “pose estimation in monkeys”

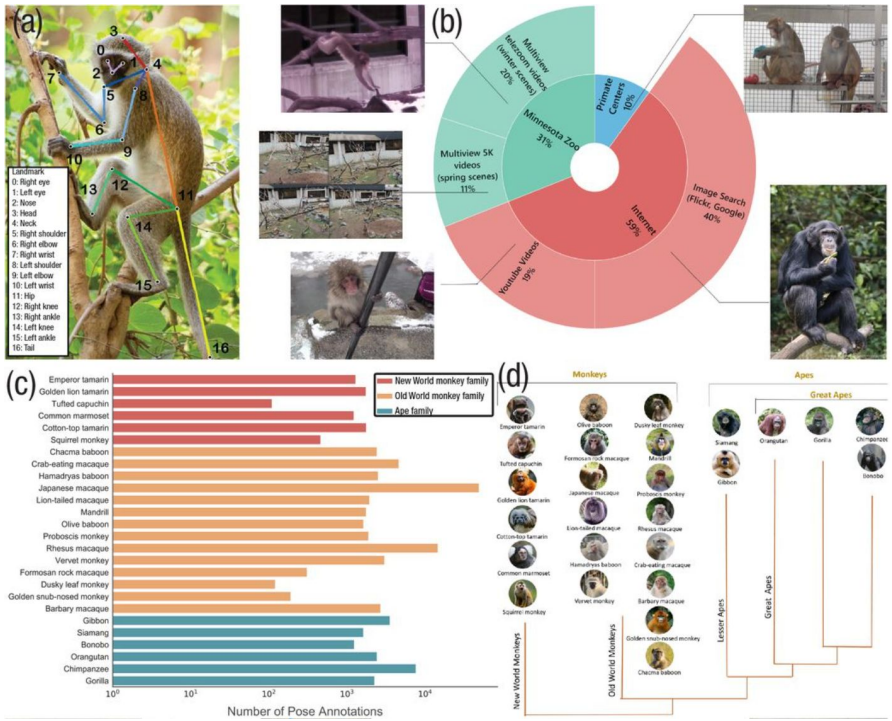
An Attention-Refined Light-Weight High-Resolution Network for Macaque **Monkey Pose Estimation**
S Liu, Q Fan, S Liu, S Li, C Zhao - Information, 2022 - mdpi.com
… of markerless macaque **pose estimation**. This dataset … **monkey pose estimation** based on MacaquePose. The experimental results show that the **pose estimation** accuracy for **monkeys** …
☆ Save 99 Cite All 2 versions »

[HTML] Automated markerless **pose estimation** in freely moving macaques with OpenMonkeyStudio
PC Bala, BR Eisenreich, SBM Yoo, BY Hayden… - Nature …, 2020 - nature.com
… **monkey pose estimation** in mind. Relative to other more readily trackable species, **monkeys** … colored fur covering), have much richer **pose** repertoires, and have much richer positional …
☆ Save 99 Cite Cited by 77 Related articles All 10 versions Web of Science: 43 »

Openmonkeystudio: Automated markerless **pose estimation** in freely moving macaques
PC Bala, BR Eisenreich, SBM Yoo, BY Hayden… - BioRxiv, 2020 - biorxiv.org
… **monkey pose estimation** in mind. Relative to other more readily trackable species, **monkeys** … colored fur covering), have much richer **pose** repertoires, and have much richer positional …
☆ Save 99 Cite Cited by 26 Related articles All 4 versions »

Dataset

- **111,529** total samples
  - Up to **17** landmarks per sample
  - Primarily internet aggregate
- Multiple sources and resolutions
- **26** species **3** families
- Semi-automated annotation
- Hidden test validation annotations



Baseline: HRNet

Deep **High-Resolution** Representation Learning for Human Pose Estimation (2019)

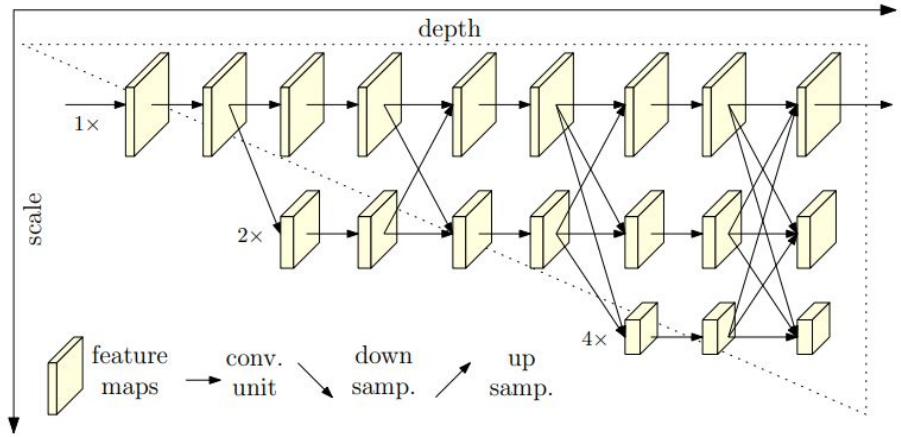


Figure 1. Illustrating the architecture of the proposed HRNet. It consists of parallel high-to-low resolution subnetworks with repeated information exchange across multi-resolution subnetworks (multi-scale fusion). The horizontal and vertical directions correspond to the depth of the network and the scale of the feature maps, respectively.

Proposed idea: ViTPose

ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation (2022)

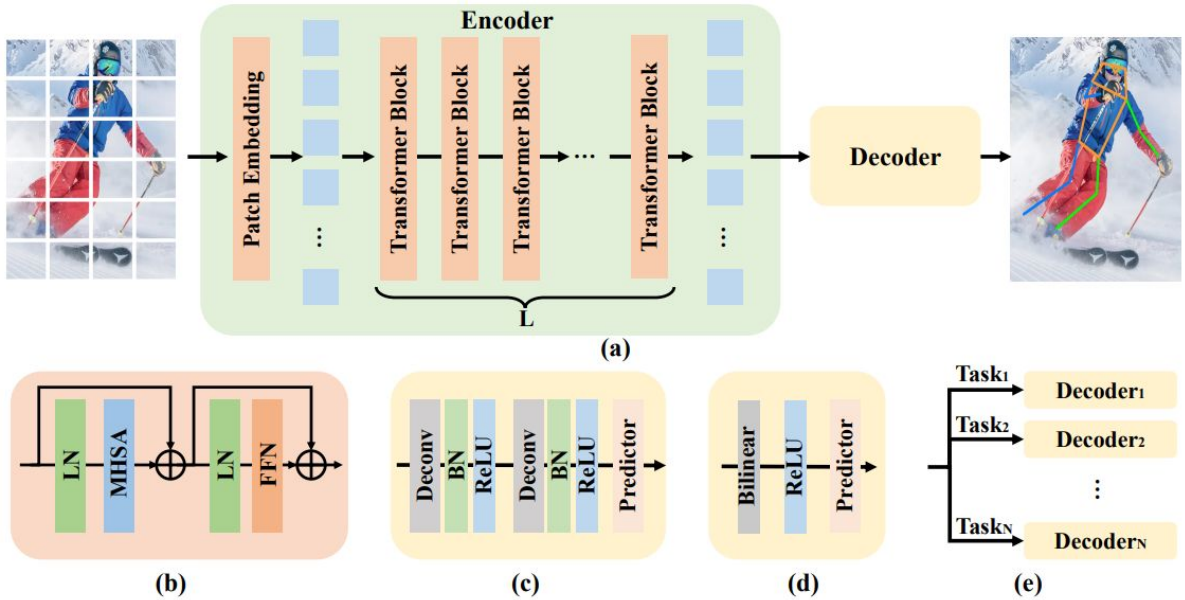


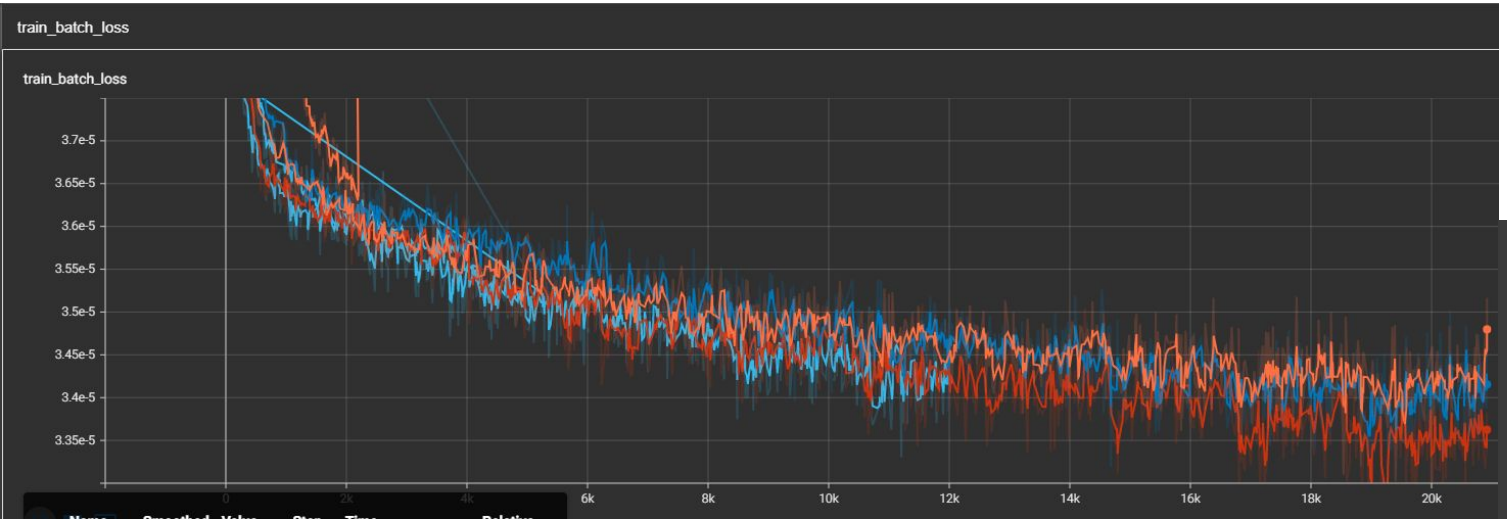
Figure 2: (a) The framework of ViTPose. (b) The transformer block. (c) The classic decoder. (d) The simple decoder. (e) The decoders for multiple datasets.

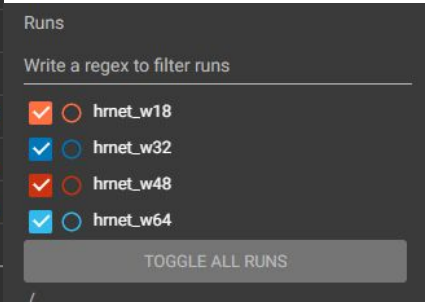
Status: implementation

Dataset (png -> h5) and data loading

Environment setup on **MSI** for larger scale training

Fine-tuning pre-trained **HRNet** for 10 epochs

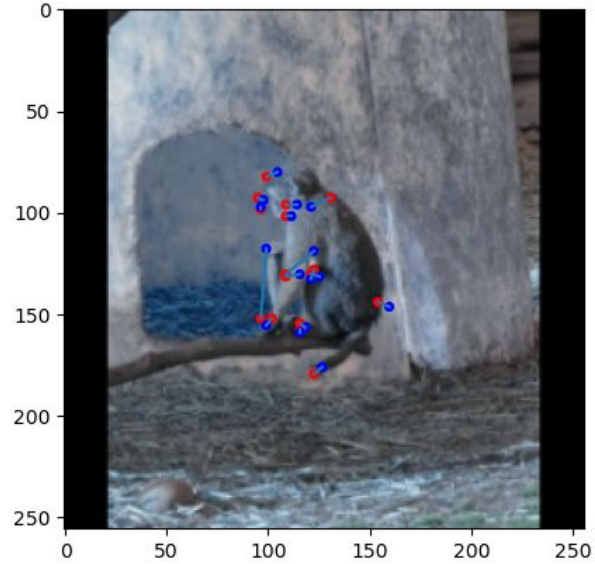


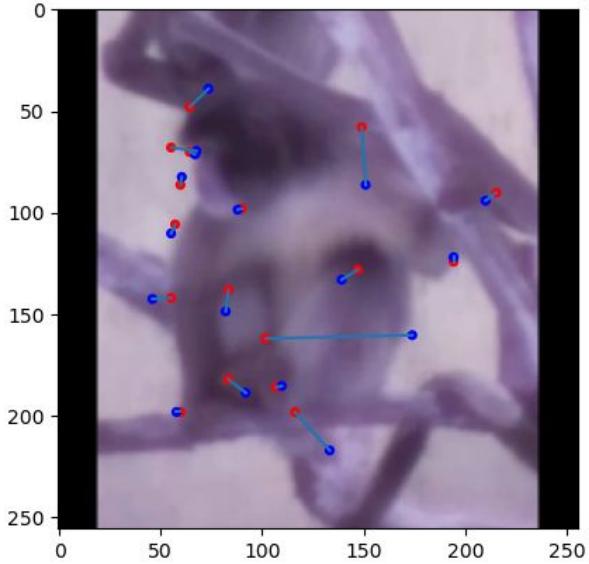


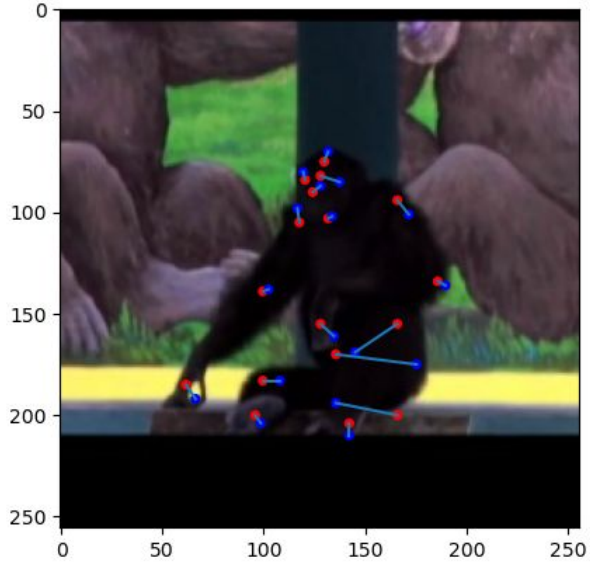
Status: qualitative results

Ground truth (blue) vs. **prediction** (red)

Line indicates corresponding joints







Status: quantitative results

Mean Per Joint Position Error (MPJPE) on DEV set: **0.0643**

Probability of Correct Keypoint@0.2 (PCK) on DEV set: **0.929**

-> If TEST metrics are similar, puts us in 3rd position

$$\mathrm{PCK@}\epsilon = \frac{1}{17J}\sum_{j=1}^{J}\sum_{i=1}^{17}\delta\left(\frac{\|\widehat{\mathbf{x}}_{ij} - \mathbf{x}_{ij}\|}{W} < \epsilon\right)$$

$$\mathrm{MPJPE}_i = \frac{1}{J}\sum_{j=1}^{J}\frac{\|\widehat{\mathbf{x}}_{ij} - \mathbf{x}_{ij}\|}{W}$$

MPJPE ▲	PCK@0.2 ▲
1.286 (1)	0.000 (15)
1.001 (2)	0.010 (14)
0.725 (3)	0.014 (13)
0.228 (4)	0.676 (10)
0.219 (5)	0.665 (11)
0.213 (6)	0.596 (12)
0.199 (7)	0.711 (9)
0.105 (8)	0.872 (6)
0.101 (9)	0.866 (7)
0.095 (10)	0.842 (8)
0.075 (11)	0.918 (5)
0.071 (12)	0.939 (3)
0.068 (13)	0.920 (4)
0.053 (14)	0.957 (2)
0.047 (15)	0.964 (1)

Next steps & work distribution

Implement and test ViTPose

Do larger scale training with >10 epochs

Champion programming: Ku and Gustav

Assist and lead write-ups: Max and Josh