Deep High-Resolution Representation Learning for Human Pose Estimation

COMP8240 Project - Final Presentation - Group G

David Dela Cruz
Rigel Ng
Srikar Parimi
Shreyas Kumar Singh

Macquarie University

November 2, 2021

Outline

Overview of the Research

- Deep High-Resolution Representation Learning for Human Pose Estimation
- Human Pose Estimation Problem
  - High-Resolution Deep Neural Network (HRNet)
  - Top-Down: Human Detection - Single Person Keypoint Detection
    - ★ Input: Person Bounding Boxes, Images
    - ★ Output: Keypoint Coordinates
  - Human Keypoint Estimation
    - ★ Nose, Eyes, Ears, Shoulders, Elbows, Wrists, Hips, Knees & Ankles

Replication of Original Research

Testing Environment

- Environment: Google Colab
- Requirements:
  - Python
  - COCOAPI
  - Pytorch
- Models: HRNet-W32, HRNet-W48
- Evaluation Script Inputs:
  - YAML Config File: Person Bounding Box JSON File, Ground-Truth JSON File, Image Dataset
  - Pre-trained Model
- Evaluation Metrics: Average Precision (AP), Average Recall (AR), Object Keypoint Similarity (OKS)

Replication of Original Research

Original Results (COCO 2017 Validation Set)

Model	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
HRNet-W32	75.8	90.6	82.7	71.9	82.8	81.0
HRNet-W48	76.3	90.8	82.9	72.3	83.4	81.2

Replicated Results (COCO 2017 Validation Set)

Model	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
HRNet-W32	75.8	90.6	82.5	72.0	82.7	80.9
HRNet-W48	76.3	90.8	82.9	72.3	83.4	81.2

Issues Encountered

- GPU Limitation
  - Update python test scripts accordingly
  - Longer execution time for Keypoint estimation
- Evaluation Script
  - Naming convention for Dataset Images
- Absence of the Person Detection Model
  - Top-Down Approach: Model requires Person Bounding Boxes
  - As a solution, the model used the Ground-Truth Person Bounding Boxes

Creation of New Dataset

Image Sources

- Obtained images from various datasets/sources
  - Stanford 40 Actions - The Stanford 40 Action Dataset contains images of humans performing 40 actions
  - COCO 2014 Test Dataset

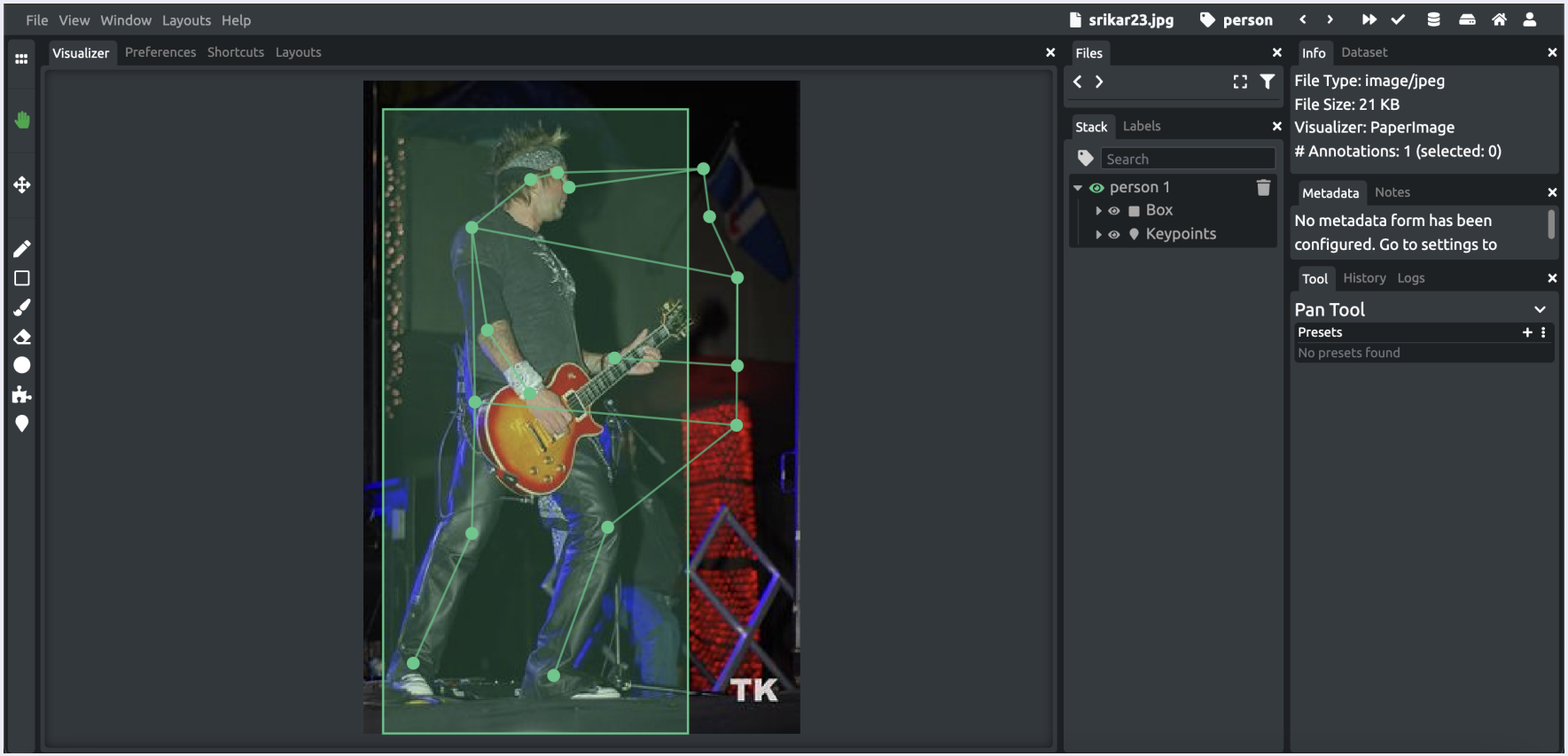
Datasets Created

- Full General Dataset - 106 Images w/ 158 Person & Keypoint Annotations
- Orientation Based Subsets
  - Front-Facing
  - Back-Facing
  - Side-Facing
  - Multiple People

Creation of New Dataset

DataTorch.io

- Export Person Bounding Box & Keypoint Annotations in a JSON file in the similar format of the Original COCO Dataset.
- We have created a python script to re-format and clean the DataTorch JSON file

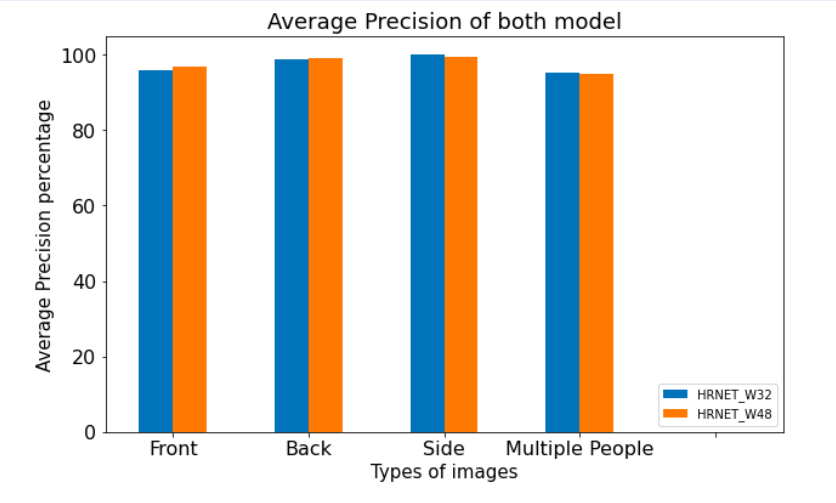


New Dataset Results

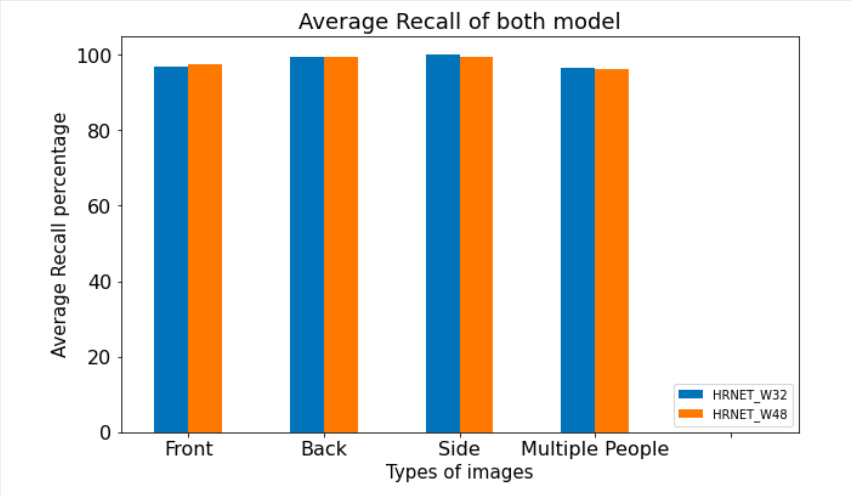
New Dataset (Full Dataset) - Results

Model	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
HRNet-W32	94.9	96.7	96.7	97.5	95.0	95.9
HRNet-W48	95.0	96.8	96.8	97.5	95.1	96.0

New Dataset (Orientation Based Dataset) - Results



Average Precision



Average Recall

Conclusion and Summary

- Reproducibility
  - Pros - Overall documentation is good - able to completely reproduce what we set out to do
  - Cons - Creating new dataset without crowdsourcing / lots of money from microsoft is extremely time-consuming.
- High accuracy
  - Model is incredibly accurate if bounding boxes are accurate.
  - Annotation should be on every subject if possible.