Primate pose estimation with OpenMonkeyChallenge

Chahyon Ku, Gustav Baumgart, and Max Scheder-Bieschin
Dec 14, 2022
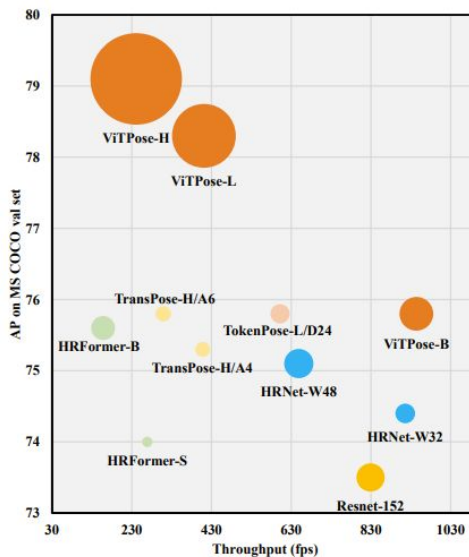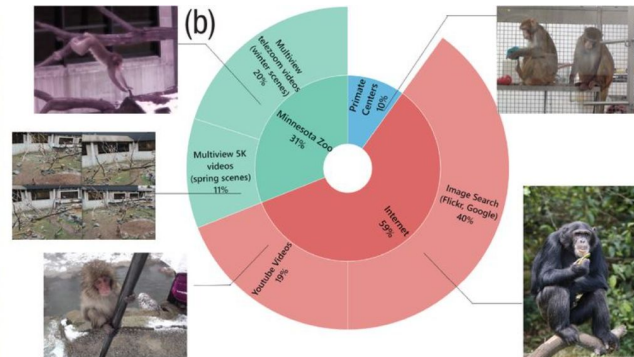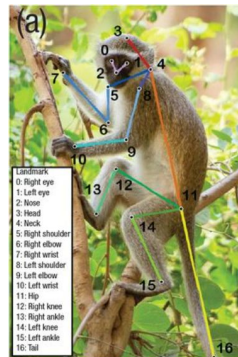
UNIVERSITY OF MINNESOTA
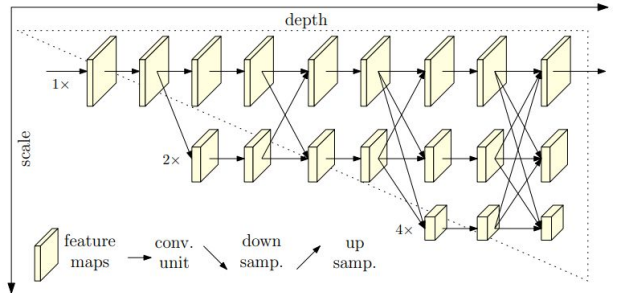Driven to Discover℠

Dataset & Project Goals

- **111,529** total samples
  - Up to **17** landmarks per sample
  - Primarily internet aggregate
- **26** species **3** families

- HRNet as Baseline
- ViTPose
  - Augmentation
  - Context

HRNet
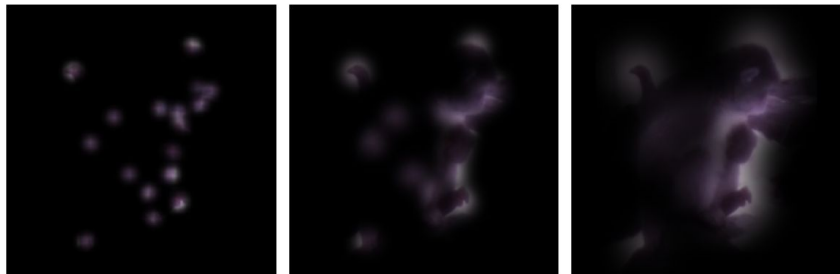
Timm pretrained hrnet_w{18, 32, 48, 64} for 20 epochs

	MPJPE	PCK0.2	mAP	#Param
HRNet_w18	0.060	0.939	0.822	11M
HRNet_w32	0.058	0.944	0.830	31M
HRNet_w48	**0.058**	**0.945**	**0.830**	**67M**
HRNet_w64	0.058	0.944	0.830	118M

MPJPE ▲	PCK@0.2 ▲	PCK@0.5 ▲	mAP ▲
0.047 (15)	0.964 (1)	0.996 (1)	0.925 (1)
0.053 (14)	0.957 (2)	0.994 (2)	0.901 (2)
0.068 (13)	0.920 (4)	0.983 (5)	0.838 (3)
0.075 (11)	0.918 (5)	0.987 (4)	0.789 (4)
0.071 (12)	0.939 (3)	0.991 (3)	0.786 (5)
0.095 (10)	0.842 (8)	0.980 (6)	0.668 (6)
0.101 (9)	0.866 (7)	0.976 (8)	0.657 (7)
0.105 (8)	0.872 (6)	0.978 (7)	0.590 (8)
0.199 (7)	0.711 (9)	0.887 (10)	0.449 (9)
0.228 (4)	0.676 (10)	0.895 (9)	0.411 (10)
0.219 (5)	0.665 (11)	0.871 (12)	0.380 (11)
0.213 (6)	0.596 (12)	0.872 (11)	0.260 (12)
1.286 (1)	0.000 (15)	0.025 (15)	0.000 (13)
1.001 (2)	0.010 (14)	0.157 (14)	0.000 (13)
0.725 (3)	0.014 (13)	0.199 (13)	0.000 (13)

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

ViTPose

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.
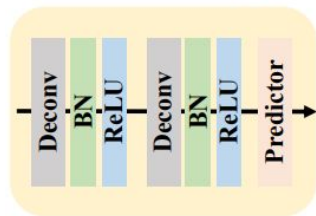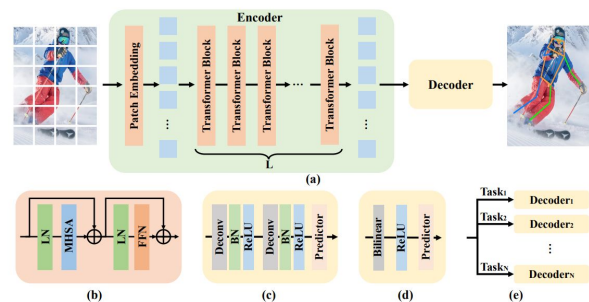
Timm pretrained vit_base_patch16_224 for 5 epochs

Std of Gaussian Heatmap (4, 8, 16)

Number of UpConv Layers (2, 3, 4)

(c)

	MPJPE	PCK0.2	AP	#Param	Time
s8u2	0.055	0.957	0.833	105M	1h 15m
s4u2	0.058	0.949	0.819	105M	1h 15m
s16u2	0.058	0.958	0.813	105M	1h 15m
s8u3	0.051	0.959	0.852	115M	2h 15m
s8u4	**0.051**	**0.959**	**0.855**	**124M**	**4h 45m**
hrnet	0.058	0.945	0.830	67M	-

s: std of heatmap, u: number of upconv layers

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

RandAugment

- Color, Spatial Transformation
- Randomly apply N transformations of magnitude M
  - N = 3, M = 0, 2, 4 for experiments
- 10 epochs
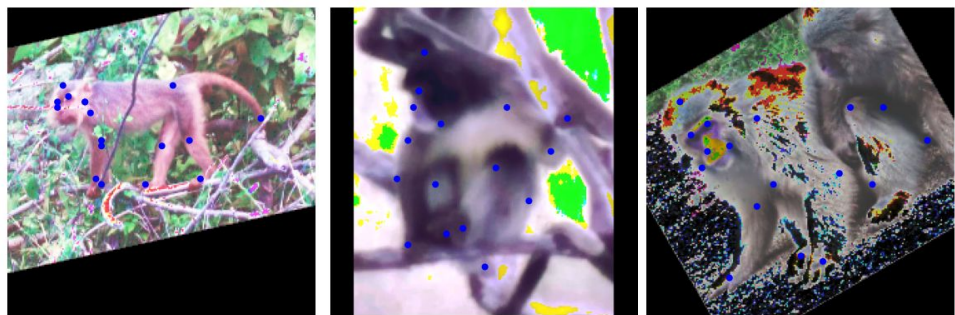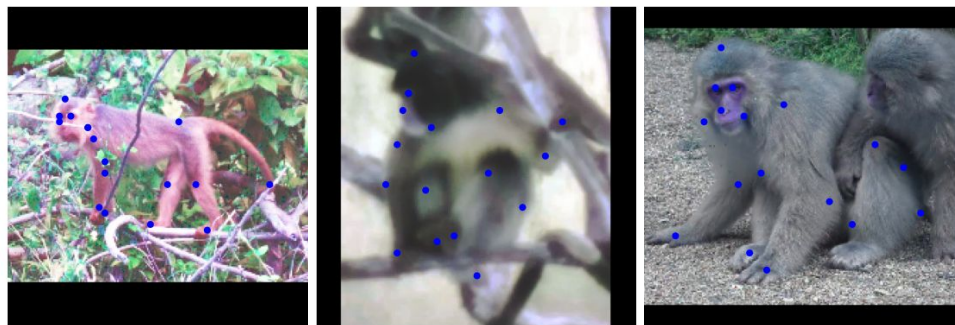
	MPJPE	PCK0.2	AP	#Param
s8**m0**u2	0.055	0.958	0.833	105M
s8**m2**u2	**0.052**	**0.963**	**0.844**	105M
s8**m4**u2	0.052	0.962	0.843	105M
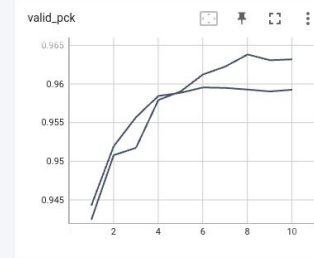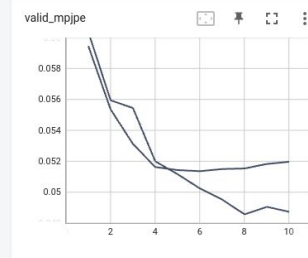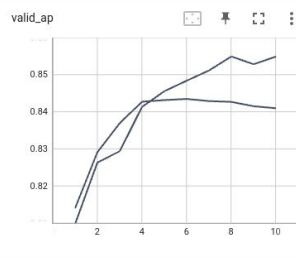
S: std of heatmap, U: number of upconv layers
M: magnitude of randaugment

RandAugment: Practical automated data augmentation with a reduced search space (2019)

Species Context Prefix Token

Vision Transformer (ViT)

x 26 (1 / specie)

An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (2020)
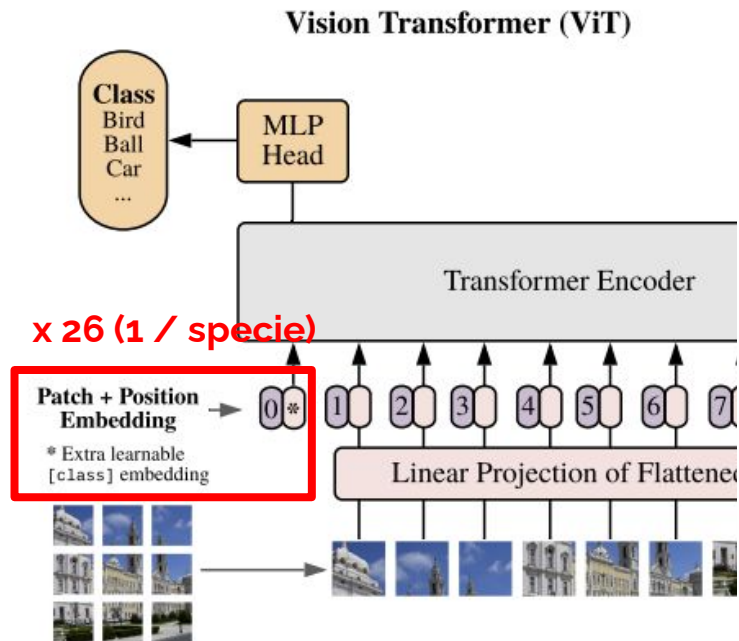
- Instead of having 1 [class] embed, have 26 for each species
- 10 epochs

	MPJPE	PCK0.2	mAP	#Param
s8m0u2	0.055	0.958	0.833	105M
s8m0u2**c**	0.054	0.959	0.833	105M
s8m2u2	0.052	0.963	0.844	105M
s8m2u2**c**	**0.050**	**0.964**	**0.850**	105M
s8m4u2	0.052	0.962	0.843	105M
s8m4u2**c**	0.052	0.963	0.844	105M

Final Results

20 Epochs

	MPJPE	PCK0.2	mAP	#Param	Time
s8m2u3	0.047	0.967	0.867	115M	8h
s8m2u3c	0.047	0.968	0.870	115M	8h
s8m4u3c	0.047	0.968	0.869	115M	8h
s8m2u4c	0.046	0.967	0.874	124M	17h
s8m4u4c				124M	17h

MPJPE ▲	PCK@0.2 ▲	PCK@0.5 ▲	mAP ▲
0.047 (15)	0.964 (1)	0.996 (1)	0.925 (1)
0.053 (14)	0.957 (2)	0.994 (2)	0.901 (2)
0.068 (13)	0.920 (4)	0.983 (5)	0.838 (3)
0.075 (11)	0.918 (5)	0.987 (4)	0.789 (4)
0.071 (12)	0.939 (3)	0.991 (3)	0.786 (5)
0.095 (10)	0.842 (8)	0.980 (6)	0.668 (6)
0.101 (9)	0.866 (7)	0.976 (8)	0.657 (7)
0.105 (8)	0.872 (6)	0.978 (7)	0.590 (8)
0.199 (7)	0.711 (9)	0.887 (10)	0.449 (9)
0.228 (4)	0.676 (10)	0.895 (9)	0.411 (10)
0.219 (5)	0.665 (11)	0.871 (12)	0.380 (11)
0.213 (6)	0.596 (12)	0.872 (11)	0.260 (12)
1.286 (1)	0.000 (15)	0.025 (15)	0.000 (13)
1.001 (2)	0.010 (14)	0.157 (14)	0.000 (13)
0.725 (3)	0.014 (13)	0.199 (13)	0.000 (13)

Final Results

20 Epochs

	MPJPE	PCK0.2	mAP	#Param	Time
base-u3	0.0468	0.9677	0.8699	115M	8h
base-u4	0.0463	0.9671	0.8737	124M	17h
large-u3	0.0458	0.9690	0.8741	355M	20h
large-u4	**0.0449**	**0.9690**	**0.8793**	**371M**	**40h**

MPJPE ▲	PCK@0.2 ▲	PCK@0.5 ▲	mAP ▲
0.047 (15)	0.964 (1)	0.996 (1)	0.925 (1)
0.053 (14)	0.957 (2)	0.994 (2)	0.901 (2)
0.068 (13)	0.920 (4)	0.983 (5)	0.838 (3)
0.075 (11)	0.918 (5)	0.987 (4)	0.789 (4)
0.071 (12)	0.939 (3)	0.991 (3)	0.786 (5)
0.095 (10)	0.842 (8)	0.980 (6)	0.668 (6)
0.101 (9)	0.866 (7)	0.976 (8)	0.657 (7)
0.105 (8)	0.872 (6)	0.978 (7)	0.590 (8)
0.199 (7)	0.711 (9)	0.887 (10)	0.449 (9)
0.228 (4)	0.676 (10)	0.895 (9)	0.411 (10)
0.219 (5)	0.665 (11)	0.871 (12)	0.380 (11)
0.213 (6)	0.596 (12)	0.872 (11)	0.260 (12)
1.286 (1)	0.000 (15)	0.025 (15)	0.000 (13)
1.001 (2)	0.010 (14)	0.157 (14)	0.000 (13)
0.725 (3)	0.014 (13)	0.199 (13)	0.000 (13)