Multiple Monkey Pose Estimation Using OpenPose

Salvador BLANCO NEGRETE (1)

Rollyn Labuguen (1) , Jumpei Matsumoto (2) , Yasuhiro Go (3,4) ,

Ken-ichi Inoue (5) , Tomohiro Shibata (1)

(1) Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan

(2) Systems Emotional Science, University of Toyama, Japan

(3) Cognitive Genomics Research Group, ExCELLS National Institutes of Natural Sciences, Japan

(4) Department of System Neuroscience, National Institute for Physiological Sciences Okazaki, Japan

(5) Department of Neuroscience, Primate Research Institute, Kyoto University, Japan

blanco.negrete@gmail.com

Motivation

Animals are widely used for experiments

Why Monkeys ?

Nonhuman primates more closely mirror human physiological and behavioral features.

Manual Annotation vs Computer Based Analysis

Manual :

- Subjective
- Changes by person
- Takes a lot of time

Computer Based :

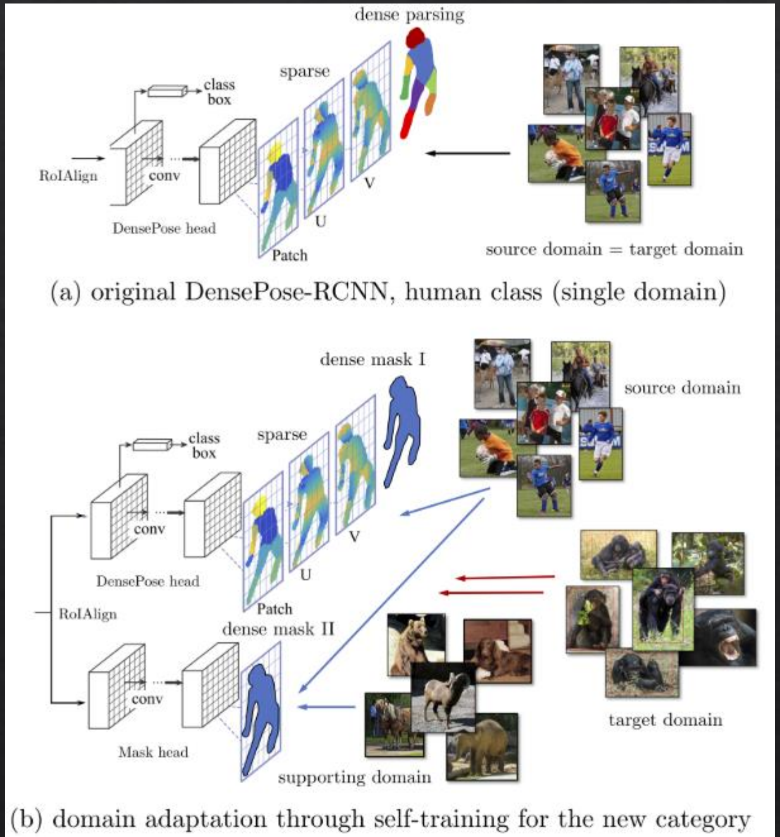
- Quantitative
- Unbiased
- Replicable
- Automated
- Sophisticated data analysis
- Less intrusive

Ethical considerations (Three Rs):

- Replacement: Replace the use of animals
- Reduction: Obtain more information from animals.
- Refinement: Alleviate or minimize potential pain, suffering or distress.

Transferring Dense Pose to Proximal Animal Classes

Sanakoyeu Artsiom, Khalidov Vasil, McCarthy Maureen S., Vedaldi Andrea, Neverova Natalia



(a) original DensePose-RCNN, human class (single domain)

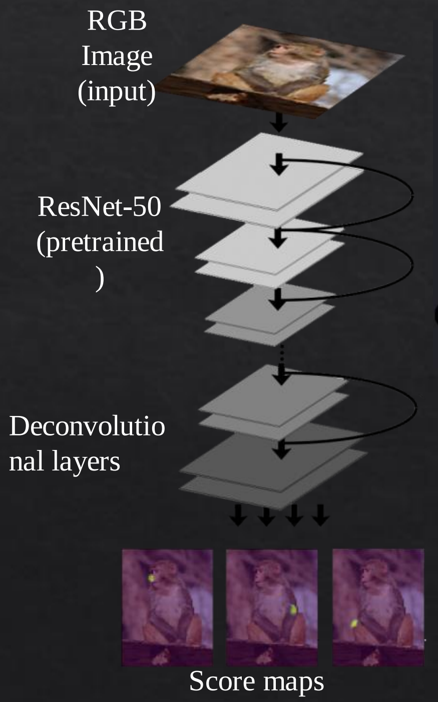
(b) domain adaptation through self-training for the new category

https://arxiv.org/abs/2003.00080

DeepLabCut

Alexander Mathis, Pranav Mamidanna, Taiga Abe, Kevin M. Cury, Venkatesh N. Murthy, Mackenzie W. Mathis, Matthias Bethge (Submitted on 9 Apr 2018)

- Minimal training data ≈ 200



https://arxiv.org/abs/1804.03142
https://github.com/AlexEMG/DeepLabCut

Related Work

OpenPose

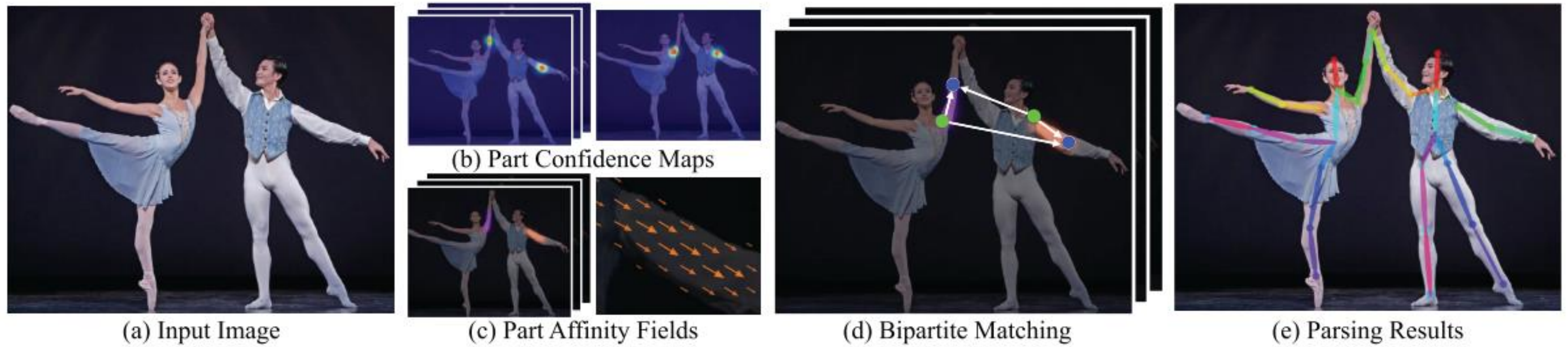
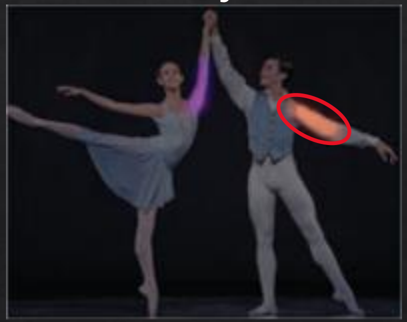
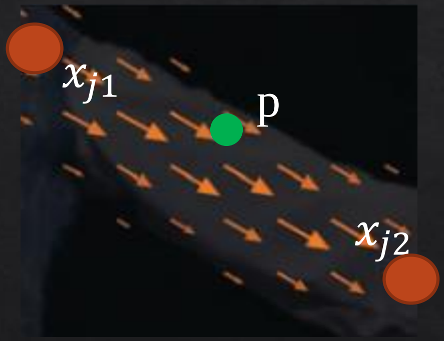


Figure 2. Overall pipeline. Our method takes the entire image as the input for a two-branch CNN to jointly predict confidence maps for body part detection, shown in (b), and part affinity fields for parts association, shown in (c). The parsing step performs a set of bipartite matchings to associate body parts candidates (d). We finally assemble them into full body poses for all people in the image (e).

Part Affinity Fields



$$L_c(p) = \begin{cases} v & \text{if } p \text{ on linb } c \\ 0 & \text{otherwise} \end{cases}$$



$$v = \frac{x_{j2} - x_{j1}}{\left\| x_{j2} - x_{j1} \right\|_2}$$

Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1302–1310, 2017.

Multiple Monkey Pose Estimation

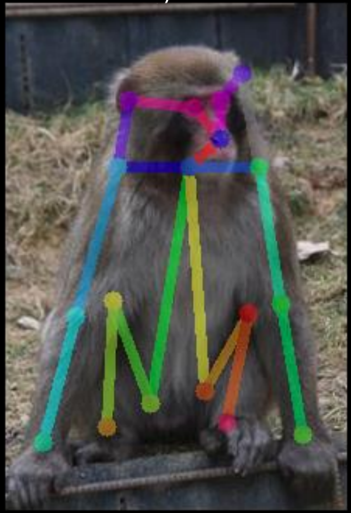
- Model capable of detecting monkey body features and posture
- No restrictions on the number of monkey subjects in the image
- It should work with different monkey behaviors
  - Sniffing, grooming, eating, crawling, etc.
- Robust against colussion and social interactions
- No environment restrictions, useful “In the wild”
  - Not limited to laboratory environments

Method

MacaquePose Dataset

Total number of monkeys	16393
Total number of images	13083

17 Body Features



Nose	
Right ear	Left ear
Right eye	Left eye
Right shoulder	Left shoulder
Right elbow	Left elbow
Right wrist	Left wrist
Right hip	Left hip
Right knee	Left knee
Right ankle	Left ankle

Occlusion Information

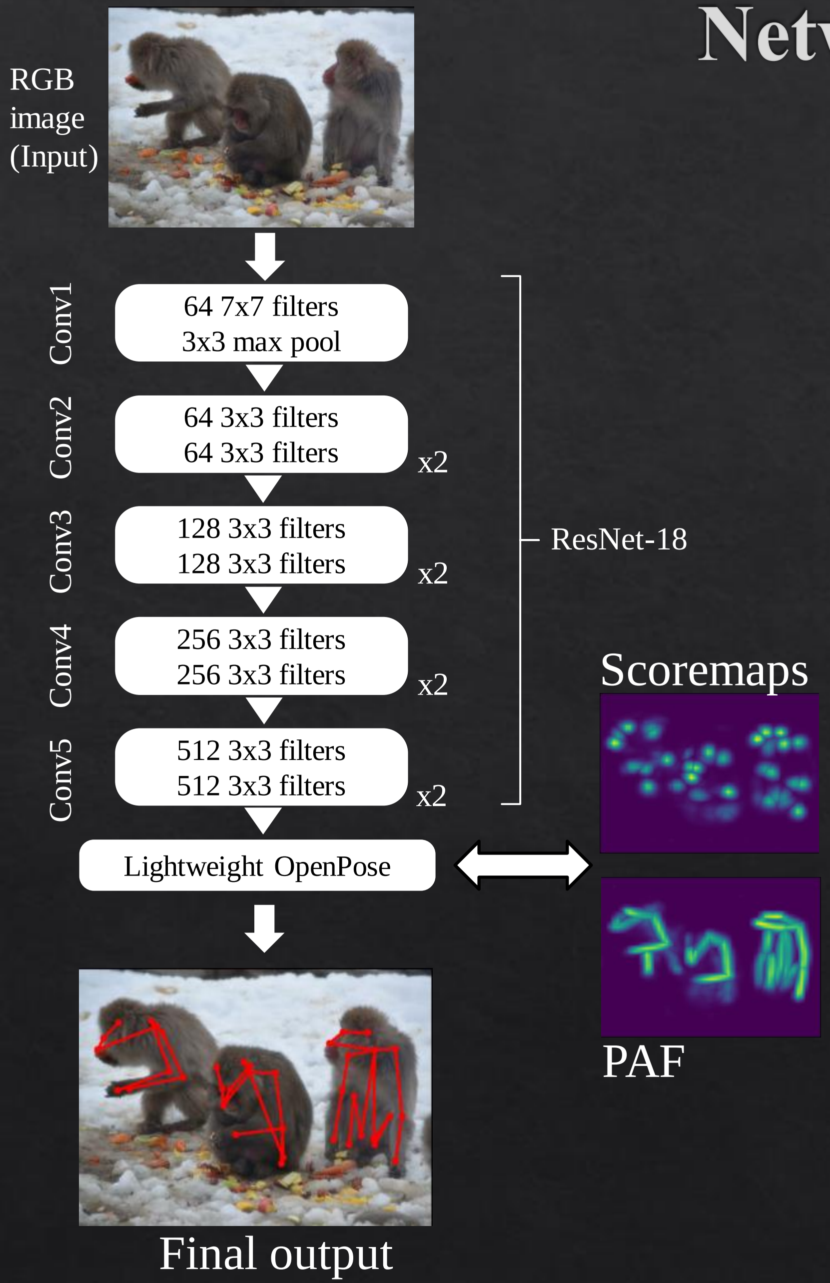
Segmentation masks

https://www.biorxiv.org/content/10.1101/2020.07.30.229989v2

Original OpenPose applied to monkey images

- In a few cases OpenPose is able to detect monkey features.
- The confidence maps show that the network is able to detect some body features, but with low confidence.

Network Architecture

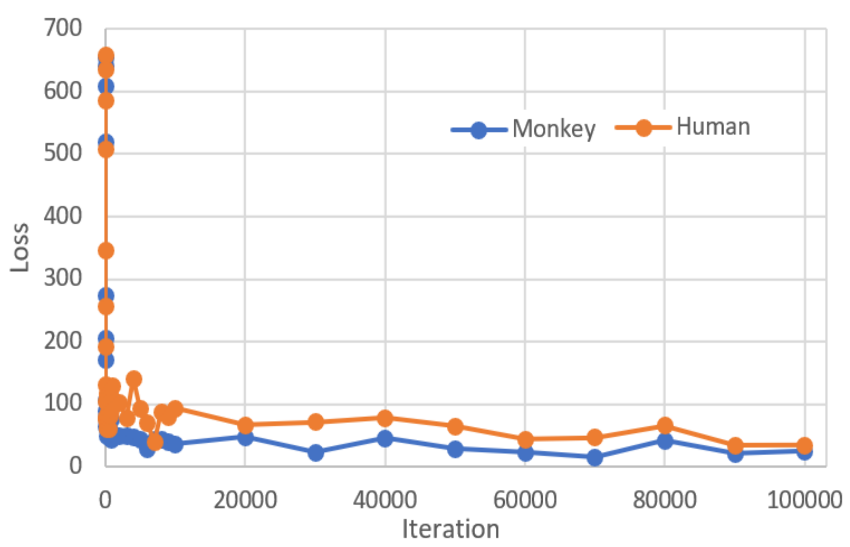


- Images resized to 640 pixels, on height or width.
- ResNet18 allows the network to be lightweight
- Lightweight implementation of OpenPose
- TensorFlow Framework
- Hyperpose an open-source implementation of OpenPose from Tensorlayer
- The network was exported to the ONNX for real-time run

Hyperpose: https://github.com/tensorlayer/hyperpose

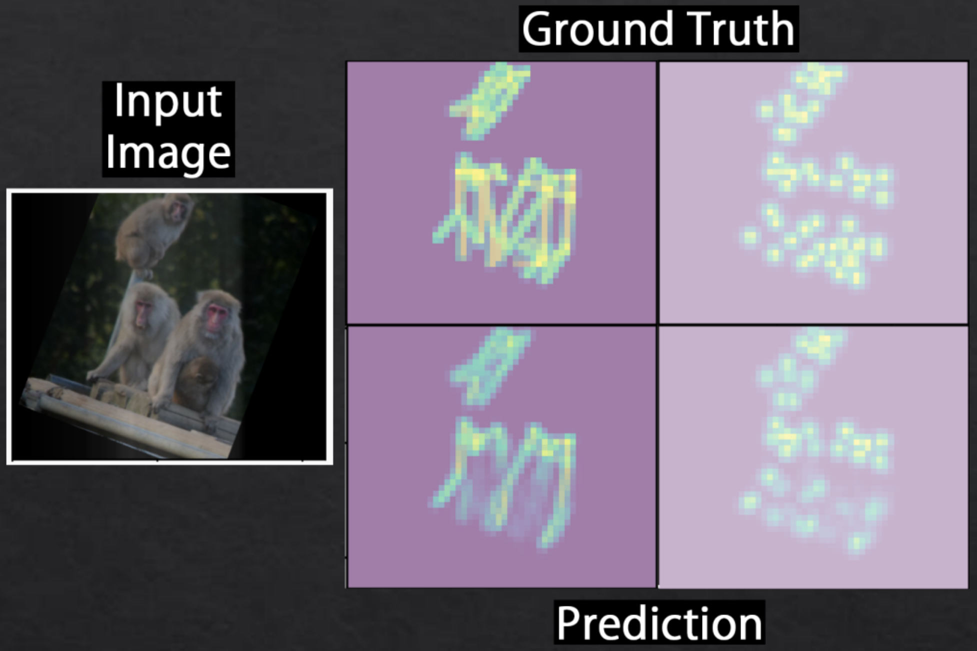
Training

	Train	Val	Total
Images	12265	818	13083
Images %	93.75 %	6.25 %	
Monkeys	15373	1020	16393
Monkeys %	93.77 %	6.22 %	



- The model is trained on a Nvidia GeForce GTX TITAN X
- The model is trained for 100,000 iterations
- The required time to train the network is approximately 24 hours
- Loss significantly decreased during the first 10,000 iterations
- Afterward, the Loss keeps dropping at a slower but steady rate
- The graph also shows a comparison against the same network trained with the human MSCOCO 2017 dataset

Results



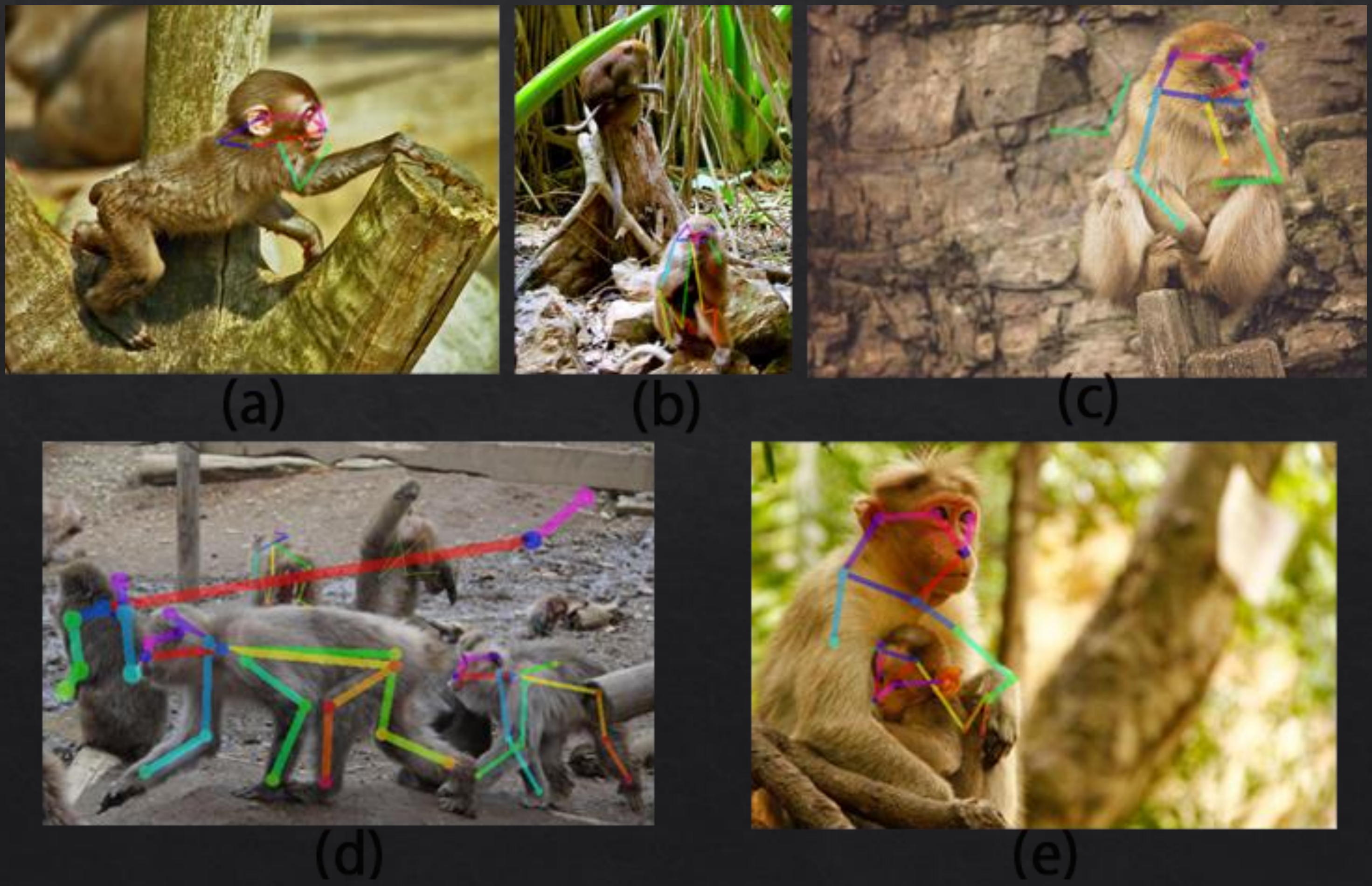
Network	Dataset	Ap^{50}	Ap^{75}	Ap^{M}	Ap^{L}
Lightweight Openpose + ResNet18	Monkey Dataset	80.8	37.3	37.3	42.8
Lightweight Openpose + ResNet18	MSCOCO 2017	38.0	6.9	12.8	15.7
Original OpenPose	MSCOCO 2016	83.4	66.4	55.1	68.1

- Predicted score maps and PAF reassemble the ground-truth
- AP was used as a metric
- Monkey model evaluated on 757 images from the evaluation set
- Human model evaluated on 1831 images from the evaluation set
- Original OpenPose results included in table

Visual Inspection
Success cases

- Unseen images from the dev set
- The sampled images display various activities that include
  - Eating
  - Playing
  - Jumping
  - Crawling
  - Standing
- The backgrounds are rich and varied
  - Natural sceneries
  - Plants and trees
  - City with a sea view

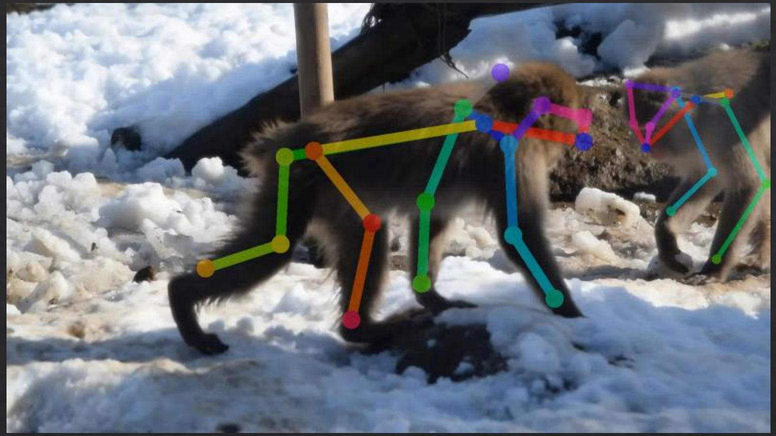
Visual Inspection
Failure cases



- a) Some body parts are not detected
- b) A monkey is not detected
- Body parts detected in the background
  - c) Independent
  - d) Attached to a monkey
- e) Mixed body features

Unsen videos

Sniffing



Mating

Grooming

Gathering

Conclusion

- Trained OpenPose neural network using the MacaquePose monkey dataset.

- The trained model is capable of detecting monkey body features and generating PAF on unseen images.

- The final output is the monkey's posture; there are no restrictions on the number of monkey subjects in the image.
  - It is robust against colussion and social interaction
  - It works on challening backgrounds.

- The model was exported to the ONNX format allowing the model to run in real-time.

- Better results could be achieved
  - Using information from contiguous frames on videos or tracking the subjects.
  - Using a deeper network and transfer learning.

Thank you

References

M. A. Nashaat, H. Oraby, L. B. Peña, S. Dominiak, M. E. Larkum, and R. N. S. Sachdev, " Pixying Behavior: A Versatile Real-Time and Post Hoc Automated Optical Tracking Method for Freely Moving and Head Fixed Animals ," *Eneuro*, vol. 4, no. 1, p. ENEURO.0245-16.2017, 2017.

Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1302–1310, 2017.

T. Nakamura *et al.*, "A markerless 3D computerized motion capture system incorporating a skeleton model for monkeys," *PLoS One*, vol. 11, no. 11, pp. 1–18, 2016.

P. Bala, B. Eisenreich, S. B. M. Yoo, B. Hayden, H. S. Park, and J. Zimmermann, "OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques," *Nat. Commun.*, no. 2020, pp. 1–12, 2020.

R. Labuguen, V. Gaurav, and S. Blanco Negrete, "Monkey Features Location Identification Using Convolutional Neural Networks," *28th Annu. Conf. Japanese Neural Netw. Soc.*, pp. 28–30, 2018.

R. Labuguen, D. K. Bardeloza, S. Blanco Negrete, J. Matsumoto, I. Kenichi, and T. Shibata, "Primate Markerless Pose Estimation and Movement Analysis Using DeepLabCut."

R. Labuguen *et al.*, "MacaquePose: A novel 'in the wild' macaque monkey pose dataset for markerless motion capture," *bioRxiv*, p. 2020.07.30.229989, 2020.

A. Mathis *et al.*, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nat. Neurosci.*, vol. 21, no. 9, pp. 1281–1289, 2018.

A. Sanakoyeu, V. Khalidov, M. S. McCarthy, A. Vedaldi, and N. Neverova, "Transferring Dense Pose to Proximal Animal Classes," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 5232–5241, 2020.

I. F. Rodriguez *et al.*, "Multiple animals tracking in video using part affinity fields," *Work. Vis. Obs. Anal. Vertebr. insect Behav. Proc. 24th Int. Conf. Pattern Recognit. (ICPR), Beijing, China*, pp. 20–24, 2018.

H. Dong *et al.*, "TensorLayer : A Versatile Library for Efficient Deep Learning," no. 1, pp. 1–4.

D. Osokin, "Real-time 2D multi-person pose estimation on CPU: Lightweight OpenPose," *ICPRAM 2019 - Proc. 8th Int. Conf. Pattern Recognit. Appl. Methods*, pp. 744–748, 2019.

T. Y. Lin *et al.*, "Microsoft COCO: Common objects in context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014.