Exploration of Data Driven Solutions to Non-Human Primate Pose Estimation

Daniel Barton   Daniel Chang   Ruoxuan Wei   Spencer Holgate

Abstract

Non-human primate (NHP) pose estimation is important in many fields such as behavioral studies. In order to create an improved NHP pose estimation model, we first utilized DeepLabCut, an animal pose estimation toolkit, and the OpenMonkeyChallenge dataset, to create a state-of-the-art baseline. We then identified four limitations of the baseline: multiple primate occlusion, general occlusion, physiological differences between species, and neglection of human-transfer learning. We propose and implemented solutions for the latter three, including artificial occlusion data augmentation, splitting models based on species meta-data, and utilization of the COCO data-set for human-transfer learning. Using the PASCAL VOC dataset to implement artificial occlusion data augmentation, we were ultimately able to create an augmented model that had a mean per joint position error of 0.068, a 23.6% decrease relative to the baseline.

1. Introduction

Pose estimation is a computer vision technique used to infer the pose of people or objects in image or video. While human pose estimation is extensively researched, Non-human primate (NHP) research lags significantly. This is important, as NHP research is important in many fields, such as biomedicine, neuroscience, psychology, anthropology, epidemiology, and ecology. Moreover, automated tracking of NHPs can benefit conservation efforts.[1]

In this study, we will analyze the performance of a state-of-the-art pose estimation baseline model on the OpenMonkeyDataset, and propose novel solutions that overcome the limitations of this baseline.

1.1. OpenMonkeyChallenge

OpenMonkeyChallenge [30] is a computer vision benchmark challenge for Non-Human Primate (NHP) pose tracking, consisting of 112,360 images of 26 species of primates (6 New World monkeys, 14 Old World monkeys, and 6 apes), with annotated 17 landmarks on each, including nose, left eye, right eye, head, neck, left shoulder, left elbow, left wrist, right shoulder, right elbow, right wrist, hip, left knee, left ankle, right knee, right ankle and tail.

The dataset is split into training (60%), validation (20%), and test (20%) sets. Bounding boxes and species metadata are included in all three sets, while landmark visibility and locations are not given in the test set. The goal of this challenge is to predict the landmark locations in the 22,306 test images. Prediction performance is evaluated using three standardized pose estimation evaluation metrics (MPJPE, PCK, AP) as defined in Yao et al[30].

2. State-of-the-art Pose Estimation

Human pose estimation is a foundational task for computer vision. There are two main subclasses of pose estimation techniques: top-down models and bottom-up models. Top-down models first utilize a subject (ex., primate) detector to create a bounding box around the subject, and then identify the landmarks of the subject within the box. On the other hand, bottom-up models detect all instances of landmarks globally throughout the image, and combine the results to construct a skeleton and final estimations of landmarks [18]

There are many existing top down pose estimation models such as DeepLabCut [24], Convolutional Pose Machine [29], Hourglass [25], and HRNet [26], and bottom up pose estimation models such as HigherHRNet [7]. Due to the abundance of human pose data, these models are able to generalize effectively on human poses. However, pose estimation in other domains, such as with NHPs, lags far behind the current state-of-the-art for human pose estimation. Existing NHP datasets such as OpenMonkeyPose[17] and Macaque-Pose[19] are limited in their diversity of NHP environments, species, and camera angles. As we will discuss later, primates are a diverse group, with varying lifestyles and evolutionary paths, which in turn results in a diverse range of physiological features and thus diverse poses.

The purpose of this study is to utilize these existing pose estimation models and the OpenMonkeyChallenge dataset as a starting point to improve the performance of state-of-the-art NHP pose estimation. As we will discuss in the next section, DeepLabCut was selected as our baseline model.

3. Baseline: DeepLabCut

We selected DeepLabCut (DLC) as our baseline model. DLC uses a ResNet50 backbone, pretrained on the ImageNet dataset.

Using mostly default parameters (batch size = 1, iterations = 400,000), we obtained a baseline score (Table 1).

Metric	Baseline
$MPJPE$	0.089
PCK@0.2	0.871
PCK@0.5	0.911
mAP	0.732

Table 1: Baseline Score

4. Baseline Limitations and Proposed Solutions

After analyzing the results of the baseline model, DeepLabCut, on the OpenMonkeyDataset, we have identified four limitations: multiple primate occlusion, general occlusion, negligence of species metadata, and negligence of utilizing human pose estimation data for transfer learning.

4.1. Multiple Primate Occlusion

Top down pose estimation models struggle in the presence of multiple NHPs. Top down models rely on a bounding box that detects the subject. However, when the primate of interest is occluded by a secondary primate, top down models will struggle to distinguish between the primate of interest and the occlusion.

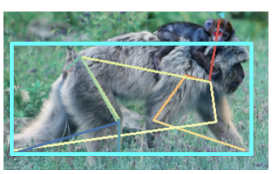

Figure 1: Multiple primate occlusion

It may be intractable to completely isolate the primate of interest by optimizing (tightening) the bounding boxes without removing the primate of interest (Fig. 1).

Moreover, primate occlusion is a special case of occlusion in general. For these reasons, we will defer the issue of primate occlusion to the handling of general occlusion. However, it is important to note that we do not address this issue directly.

4.2. Dealing With General Occlusion

In many instances, objects such as tree branches occlude important landmarks of the primate of interest, disrupting the baseline model (Fig. 2). This is an issue, as nearly all primates are arboreal animals. This issue of general occlusion has also been shown to be a difficult problem to solve in other areas of computer vision such as facial recognition[10]. However, there have been some promising solutions within human pose estimation that deal with self-occlusion and other-occlusion [12] [14].

Figure 2: General Occlusion (Tree Branch)

One of the ways occlusion is handled is through temporal solutions[2][8][13]. These temporal solutions require repeated frames. However, these types of solutions cannot be applied to our challenge.

Another way occlusion can be handled is through artificial occlusion data augmentation. Artificial occlusion data augmentation has been found to be successful in multiple areas, facial recognition[21], and 3D pose estimation[27].

Liu and Sárándi *et. al* employed artificial occlusion in two different ways, the first being strictly black colored shapes – and the second was to add random objects (e.g., a television) to the image to occlude the object of interest. Liu was able to demonstrate that the accuracy in artificial occlusion trends similarly to real occlusion training [21].

Sárándi et. al demonstrate a significant decrease in MPJPE when employing artificial occlusion data augmentation, even after augmenting only 10 or 20 percent of the training images.

4.3. Primate Physiological Differences

It is obvious that primates have a diverse range of body proportions. Some primates look more like rodents, while others, like Chimpanzees, look absolutely ripped. For this reason, an optimal model should utilize the species metadata.

A naive method to implement this optimization is to partition each dataset (train/val/test) into 26 disjoint sets (one for each species), and train a separate model for each species. However, this method may unnecessarily limit the dataset sizes of each species model. Certain species have very similar evolutionary paths and thus similar pose patterns.

A UMAP dimensionality reduction [30] of the dataset reveals that certain primate species cluster together.

For this reason, it would be fruitful to experiment with clustering algorithms to reap the benefits of creating a model that fits stronger to each species while minimizing the negative effects of reducing dataset sizes.

4.4. Human Pose Estimation Transfer Learning

In order to provide a more robust and diverse data-set, we can utilize transfer learning[16]. Pose data from the human domain can supplement our limited data, as humans are primates. Human pose transfer learning has proved to be useful in the pose estimation of ancient vase paintings[22], anime characters[6] horses[23][5] and 3D human pose estimation [9].

When adding data from another domain to the training data-set, We can use domain adaptation to bridge the gap between the human and NHP domains. A domain adaptation via style-transfer method was used by Madhu, Prathmesh *et. al* to improve pose estimation in ancient vase paintings[22]. A synthetic data-set was produced from the COCO [20] human pose data-set using adaptive instance normalization (AdaIN) [15] [22]. In order to bridge the domain gap between ancient vases and human poses, the algorithm combined semantic information from the human pose data and texture information from the ancient vase input.

Another domain adaptation has also been implemented between human and horse poses [5]. This implementation used a joint supervised scheme for cross domain joint position learning.

5. Methodologies

5.1. Artificial Occlusion

Two types of artificial occlusion data augmentation were performed. (1) Black Box Occlusion, and (2) Regular Object Occlusion. For both types, the same random subsampling of the original training images of primates was performed. To observe the influence of sub-sampling rates, we used both 20% and 50% sub-sampling rates.

Figure 3: Example of Black Box Occlusion

For black box occlusion, for each sub-sampled training image, a black box occlusion image 20% of the size of the training image was generated and superimposed on the training image of interest, centered on a randomly selected key-point location (Fig. 3).

For regular object occlusion, we utilized the PASCAL 2012 data-set to obtain regular objects for occluding the training images. More information on this data-set can be found in section 6.1.

The regular objects were resized using openCV's INTER_AREA interpolation and superimposed in the same way as the black boxes using the same training sub-sample (Fig. 4)

Figure 4: Example of Regular Object Occlusion

5.1.1 Limitations

By having a global arbitrary scaling for the occlusion images, the occlusion was not scale invariant to the scale of the target primate.

Another issue is that we include the background of their environment rather than extracting the object itself, introducing undesirable noise.

5.2. Species Informed Models

We employed two methods of dataset partitioning by species, which we will describe in the following two sections.

The first (naive) approach simply created 26 subsets of the dataset (one for each species) and trained a model for each subset.

A limitation of this approach is that the distribution of species counts in the training set is heavily right skewed (Fig. 5). For example, the training set had 29,424 Japanese Macaque images, but only 63 Tufted Capuchin images. As we will discuss later, this may have a negative effect on the model performance.

The second, "improved" approach involved clustering the species by pose similarities, partitioning the dataset into fewer (and subsequently larger sized) subsets, and training a model for each subset.

Before performing sample clustering, we performed a dimensionality reduction on the dataset using a similar

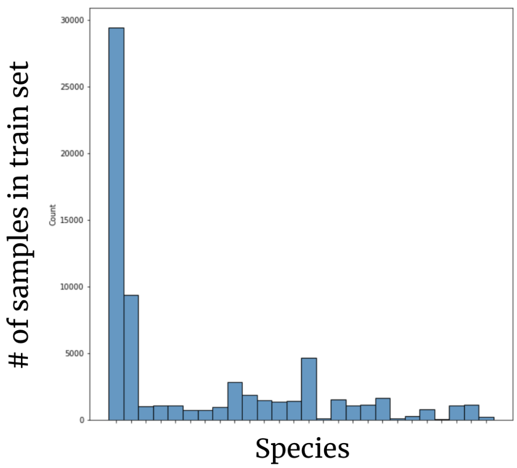

Figure 5: Distribution of species counts in training set

methodology as that of Yao *et. al*. The purpose of the dimensionality reduction was to improve the performance of the clustering algorithm, as manifold learning is a powerful technique that can extract non-linear relationships from pose data [28] [4] [3] that clustering algorithms using non hand-crafted distance metrics are unable to learn.

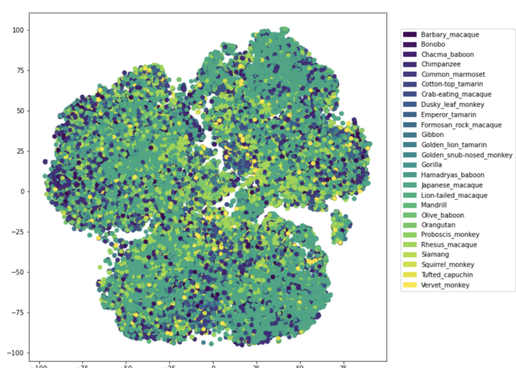

Figure 6: t-SNE Dimensionality Reduction

First, in order to allow coherent and meaningful clustering, we normalized the training set pose labels using the technique described in Yao *et. al*. Specifically, we subtracted landmark coordinates by the hip joint position, scaled relative coordinates using the given bounding box annotations, and rotated the coordinates such that the head landmark was at zero degrees.

After pose normalization, we performed a dimensionality reduction (Fig. 6) on the dataset using t-distributed stochastic neighbor embedding (t-SNE).

Afterwards, we applied k-means clustering on the sample embedding. Finally, after obtaining sample clusters, each species was assigned to the sample cluster that had the highest proportion of the species, effectively creating k species clusters. Using these species clusters, we partitioned the dataset (one subset per species cluster), and trained the DLC model on each subset.

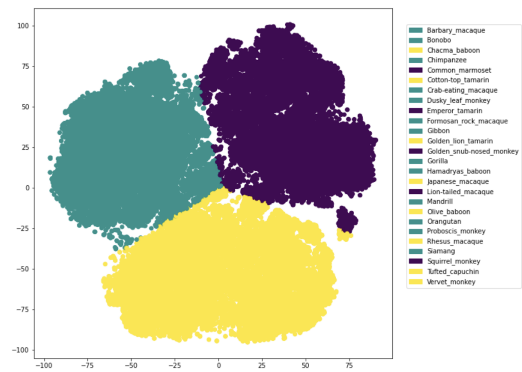

Figure 7: Species K-Means Assignment

5.3. Human Pose Transfer Learning

As we mentioned earlier, humans and NHPs have similar anatomical structures, so by first training a pose estimation model on a human dataset, we can transfer the weights/features learned for use in pose estimation of NHPs.

We utilized the COCO human keypoint dataset [20], which contains 262,465 images.

5.3.1 Preprocessing Details

Many images in the COCO dataset were either close up images or were poorly annotated and thus lacked keypoints. To address this, we filtered the dataset such that all samples contained at least 11 annotated keypoints, reducing our dataset to 98,862 samples.

We then further preprocessed the COCO dataset to be more similar to our primate dataset by merging hip landmarks, dropping ear landmarks, and creating a neck landmark that was the averaged coordinate value of the nose and shoulder locations.

The model was first trained on the preprocessed COCO dataset for 500k iterations. The model weights were then frozen and the model was subsequently trained on the open monkey pose data for 500k iterations to produce our final model (Fig. 8).

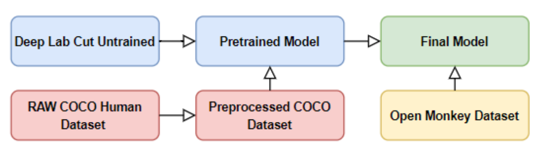

Figure 8: Human Pose Transfer Learning Method

5.3.2 Limitations

Human pose transfer learning (HTL) is helpful when dataset sizes are limited. However, issues may arise due to inherent differences in environments or pose patterns. Future works can mitigate this issue by implementing domain adaptation techniques discussed earlier.

The human data also lacked several important landmarks for NHP pose estimation such as the tail, which resulted in certain issues downstream, which we will discuss later.

6. Results and Discussion

The results of all the augmented models (Fig. 9) indicate that our best performing model with respect to MPJPE was the regular occlusion trained model and the best model with respect to mAP was the naive species informed model.

Metric	Baseline	Black occl (20%)	Reg. occl (20%)	Species informed	Species informed (clustered)	Human transfer learning
MPJPE	0.089	0.0689	0.0682	0.0684	0.0749	0.079
PCK@0.2	0.871	0.9182	0.9197	0.919	0.907	0.898
PCK@0.5	0.911	0.9825	0.9832	0.984	0.980	0.9768
mAP	0.732	0.835	0.834	0.836	0.810	0.790

(best, mediocre, worst)

Figure 9: Scores

6.1. Artificial Occlusion

Using artificial occlusion, we were able to demonstrate a 23% increase in MPJPE performance with artificial occlusion at a 20% subsampling rate for both black box occlusion and regular object occlusion. This follows suit with the results given by Sárándi et. al [27].

The sub-sampling amount had negligible affect on the performance of the model (Fig. 10).

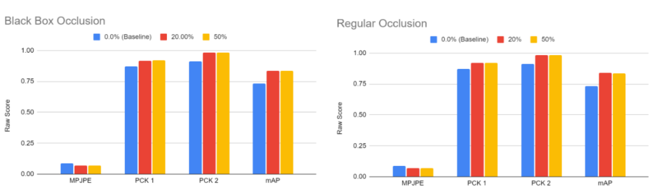

Figure 10: Sub Sampling Performance Comparison

What we have been able to identify here is a significant increase in performance by utilizing artificial occlusion data augmentation. We also have identified that the difference in performance is negligible between black boxes and regular objects (Fig. 9). This allows for the removal of needing an external database that provides occlusion images.

6.2. PASCAL Data-set

The data-set used for regular objects was the PASCAL VOC 2012 development kit data-set[11]. The PASCAL VOC 2012 dataset is best described by the table from their documentation, but is included for ease of reference:

Table 1: Statistics of the main image sets. Object statistics list only the 'non-difficult' objects used in the evaluation.

	train		val		trainval		test	
	img	obj	img	obj	img	obj	img	obj
Aeroplane	327	432	343	433	670	865	–	–
Bicycle	268	353	284	358	552	711	–	–
Bird	395	560	370	559	765	1119	–	–
Boat	260	426	248	424	508	850	–	–
Bottle	365	629	341	630	706	1259	–	–
Bus	213	292	208	301	421	593	–	–
Car	590	1013	571	1004	1161	2017	–	–
Cat	539	605	541	612	1080	1217	–	–
Chair	566	1178	553	1176	1119	2354	–	–
Cow	151	290	152	298	303	588	–	–
Diningtable	269	304	269	305	538	609	–	–
Dog	632	756	654	759	1286	1515	–	–
Horse	237	350	245	360	482	710	–	–
Motorbike	265	357	261	356	526	713	–	–
Person	1994	4194	2093	4372	4087	8566	–	–
Pottedplant	269	484	258	489	527	973	–	–
Sheep	171	400	154	413	325	813	–	–
Sofa	257	281	250	285	507	566	–	–
Train	273	313	271	315	544	628	–	–
Tvmonitor	290	392	285	392	575	784	–	–
Total	5717	13609	5823	13841	11540	27450	–	–

Figure 11: Description of PASCAL VOC 2012 Objects [11]

The object types used from this set (Fig. 11) were: Bird, Cat, Cow, Dog, Person, PottedPlant, Sheep. These object types were used since they intuitively would be the most natural to be in an environment with primates.

Data-set Limitations The data-set included several object types that would not be expected to be in the environment that a a primate would be in. Even among the most intuitively natural object classes within the data-set (e.g. Birds), the types of birds may not be natural to the habitat of these primates. This may be detrimental since the images that the model is to predict upon would instead have different object classes that are more frequent in a primate's environment, such as tree branches. It may be better to have sourced a data-set with more common object types that could be used as artificial occlusion, such as trees, vines, rocks, etc.

6.3. Species Informed Models

The naive species informed model outperformed the baseline method across the board and performs competitively with the artificial occlusion models (Fig. 9).

However, due to the right skewed species distribution in the training set (Fig. 5), certain species (Tufted Capuchin) are underrepresented and subsequently, models trained on those subsets underperformed severely (Fig. 12).

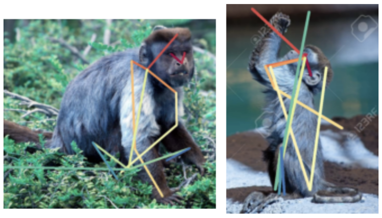

Figure 12: Species Informed: Tufted Capuchin Miserable Results

One limitation of our analysis is that we evaluated our results on the hidden test set so we are unable to inspect the performances for each of the 26 models. In future works, further analysis on a validation set may be fruitful.

The clustered species informed model also outperformed the baseline method. However, it underperformed relative to all other augmented models except for the HTL model. This may be attributed to lack of hyperparameter optimization and choice of clustering algorithms, which was a result of time constraints: we were only able to test the model with k=3 clusters.

However, even with our limited testing, we identified potential benefits using this method. With 3 clusters, we noticed that the turquoise cluster (Fig. 7) contained all of the Apes, which shows how the t-SNE + k-means process is identifying pose patterns effectively, as Apes have very distinct body proportions compared to monkeys.

6.4. Human Transfer Learning

As expected, the HTL model beat the baseline model across the board (Fig. 9, right column). However, it underperformed relative to the other augmented models. As we mentioned, this may be attributed to issues such as lack of tail keypoints in the COCO dataset.

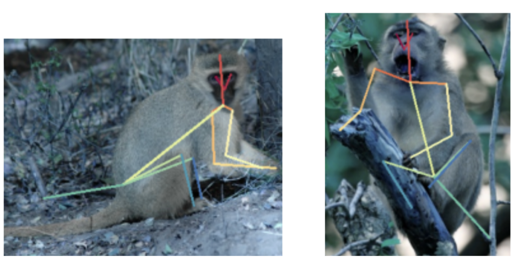

Figure 13: HTL: failed tail predictions

Besides validating this limitation qualitatively (Fig. 13), we can utilize DeepLabCut's likelihood value for each keypoint prediction, which is taken from the DLC model's spatial density function output.

Average likelihood values for the tail landmark, which was absent in COCO, was significantly lower than those for nose and eyes, which were present (Table 2).

	Tail	Nose	Eye
Likelihood	0.3744	0.916	0.912

Table 2: DLC Likelihood Scores

7. Conclusion

In this analysis, we identified the following limitations of our state-of-the-art baseline model: multiple monkey occlusion, occlusion in general, and the neglection of using species metadata and human data.

Here is a comparison of MPJPE performance increase (i.e., MPJPE decrease) relative to the baseline for all our solutions (Fig. 14).

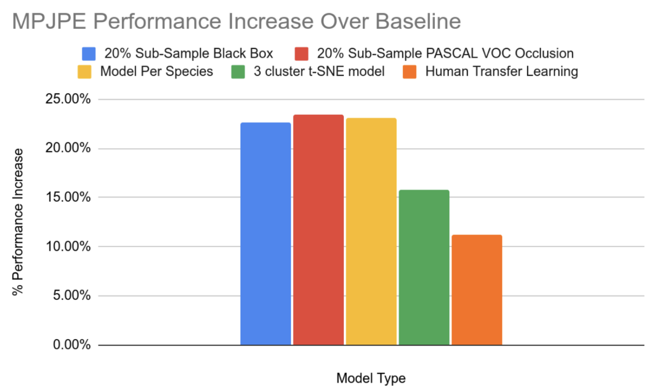

Figure 14: MPJPE Performance Increase

We addressed occlusion issues by exploring two artificial occlusion augmentation methods: black box and regular image artificial occlusion. We noticed that both of these methods had similar results, where the most prominent performance increase was in MPJPE by 23%.

We leveraged species metadata by exploring two methods of partitioning the dataset by species. Our naive method yielded a similar performance as the occlusion methods. While our clustered method outperformed the baseline, it did not yield an improvement over the naive method, possibly due to the lack of hyperparameter optimization.

Lastly, we utilized the COCO human keypoint dataset for transfer learning. Again, this model outperformed the baseline, but underperformed relative to all other methods, which could be attributed to the lack of important landmarks, such as tail and neck.

8. Code and Data Availability

Github Repository

References

[1] Pose estimation guide. 1

[2] Federico Angelini, Zeyu Fu, Yang Long, Ling Shao, and Syed Mohsen Naqvi. 2d pose-based real-time human action recognition with occlusion-handling. *IEEE Transactions on Multimedia*, 22(6):1433–1446, 2020. 2

[3] Vineeth Nallure Balasubramanian, Jieping Ye, and Sethuraman Panchanathan. Biased manifold embedding: A framework for person-independent head pose estimation. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007. 4

[4] Mai Bui, Sergey Zakharov, Shadi Albarqouni, Slobodan Ilic, and Nassir Navab. When regression meets manifold learning for object recognition and pose estimation. *CoRR*, abs/1805.06400, 2018. 4

[5] Jinkun Cao, Hongyang Tang, Hao-Shu Fang, Xiaoyong Shen, Cewu Lu, and Yu-Wing Tai. Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3

[6] Shuhong Chen and Matthias Zwicker. Transfer learning for pose estimation of illustrated characters. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 793–802, January 2022. 3

[7] Bowen Cheng, Bin Xiao, Jingdong Wang, Honghui Shi, Thomas S. Huang, and Lei Zhang. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[8] Yu Cheng, Bo Yang, Bo Wang, Yan Wending, and Robby Tan. Occlusion-aware networks for 3d human pose estimation in video. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 723–732, 2019. 2

[9] Carl Doersch and Andrew Zisserman. Sim2real transfer learning for 3d human pose estimation: motion to the rescue. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 3

[10] Hazım Kemal Ekenel and Rainer Stiefelhagen. Why is facial occlusion a challenging problem? In Massimo Tistarelli and Mark S. Nixon, editors, *Advances in Biometrics*, pages 299–308, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. 2

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 5

[12] Lianrui Fu, Junge Zhang, and Kaiqi Huang. Beyond tree structure models: A new occlusion aware graphical model for human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. 2

[13] Mehwish Ghafoor and Arif Mahmood. Quantification of occlusion handling capability of 3d human pose estimation framework. *IEEE Transactions on Multimedia*, pages 1–1, 2022. 2

[14] Yanlei Gu, Huiyang Zhang, and Shunsuke Kamijo. Multi-person pose estimation using an orientation and occlusion aware deep learning network. *Sensors*, 20(6), 2020. 2

[15] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 3

[16] Mahbub Hussain, Jordan J. Bird, and Diego R. Faria. A study on cnn transfer learning for image classification. In Ahmad Lotfi, Hamid Bouchachia, Alexander Gegov, Caroline Langensiepen, and Martin McGinnity, editors, *Advances in Computational Intelligence Systems*, pages 191–202, Cham, 2019. Springer International Publishing. 3

[17] Bala PC;Eisenreich BR;Yoo SBM;Hayden BY;Park HS;Zimmermann J;. Automated markerless pose estimation in freely moving macaques with openmonkeystudio. 1

[18] Sheng Jin. Towards multi-person pose tracking: Bottom-up and top-down methods. 1

[19] Rollyn Labuguen, Jumpei Matsumoto, Salvador Blanco Negrete, Hiroshi Nishimaru, Hisao Nishijo, Masahiko Takada, Yasuhiro Go, Ken-ichi Inoue, and Tomohiro Shibata. Macaquepose: A novel "in the wild" macaque monkey pose dataset for markerless motion capture, Jan 1AD. 1

[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing. 3, 4

[21] Hao Liu, Huiping Duan, Hongyu Cui, and Yunjie Yin. Face recognition using training data with artificial occlusions. In *2016 Visual Communications and Image Processing (VCIP)*, pages 1–4, 2016. 2

[22] Prathmesh Madhu, Angel Villar-Corrales, Ronak Kosti, Torsten Bendschus, Corinna Reinhardt, Peter Bell, Andreas Maier, and Vincent Christlein. Enhancing human pose estimation in ancient vase paintings via perceptually-grounded style transfer learning, 2020. 3

[23] Alexander Mathis, Thomas Biasi, Steffen Schneider, Mert Yuksekgonul, Byron Rogers, Matthias Bethge, and Mackenzie W. Mathis. Pretraining boosts out-of-domain robustness for pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1859–1868, January 2021. 3

[24] Tanmay Nath, Alexander Mathis, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie Weygandt Mathis. Using deeplabcut for 3d markerless pose estimation across species and behaviors, Jun 2019. 1

[25] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. *CoRR*, abs/1603.06937, 2016. 1

[26] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer*

Vision – ECCV 2016, pages 483–499, Cham, 2016. Springer International Publishing. 1

[27] István Sárándi, Timm Linder, Kai Oliver Arras, and Bastian Leibe. Synthetic occlusion augmentation with volumetric heatmaps for the 2018 ECCV posetrack challenge on 3d human pose estimation. *CoRR*, abs/1809.04987, 2018. 2, 5

[28] Chao Wang, Yuanhao Guo, and Xubo Song. Head pose estimation via manifold learning. In Paul Bracken, editor, *Manifolds*, chapter 6. IntechOpen, Rijeka, 2017. 4

[29] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1

[30] Yuan Yao, Abhiraj Mohan, Eliza Bliss-Moreau, Kristine Coleman, Sienna M. Freeman, Christopher J. Machado, Jessica Raper, Jan Zimmermann, Benjamin Y. Hayden, and Hyun Soo Park. Openmonkeychallenge: Dataset and benchmark challenges for pose tracking of non-human primates. *bioRxiv*, 2021. 1, 2

Project Information

Daniel Barton

- Data processing for artificial occlusion
- Sourcing PASCAL data-set

Daniel Chang

- Data processing for species informed models
- Species clustering methods
- Model training

Ruoxuan Wei

- Collecting human transfer learning data
- Data processing for human pose transfer learning

Spencer Holgate

- Collecting human transfer learning data
- Data processing for human pose transfer learning

CodaLab Team Name

chan1975