Estimating the Size of Wild Chimpanzees from Images

J. Alscher, T. Bala, J. Brandinger, C. Demuynck, & K. Godhwani

[1]*Department of Data Science, Tufts University*
419 Boston Ave, Medford, MA 02155, United States

Abstract

The purpose of this project is to develop an automated process to estimate the size of wild chimpanzees from images. We use a combination of models and techniques to measure the shoulder-to-rump distance from images. The models developed in this project provide researchers with a non-invasive method for monitoring the growth of chimpanzees which will aid in conservation research.

Key words. size estimation, computer vision, instance segmentation, animal pose estimation

1. INTRODUCTION

Understanding the precise size of chimpanzees in the wild holds significant importance within the realm of biology and biological anthropology. It offers insight into the various aspects of their behavior and ecology and also could offer reasons for conservation. Accurate size estimates can contribute greatly to studies on locomotion, habitat use, and social/competitive interactions within and between chimpanzee communities. Meanwhile, tracking the body size of chimpanzees all across Africa can help direct conservation funds to communities with chimpanzees suffering from stunted growth or any unhealthy or unnatural abnormalities. However, measuring them in the wild can prove to be a difficult task, as researchers are requested to not interact with the chimpanzees and cause any scientific or ethical interference. Also, proximity between humans and chimpanzees poses an increased risk of passing disease (Wildlife Chimpanzees, n.d.).

Estimating the size of chimpanzees from images allows researchers to gather crucial data without resorting to invasive measurement techniques, which minimizes the disturbances to these animals' natural habitats. This approach would help facilitate longitudinal studies that enable researchers and conservationists to track changes in population dynamics and individual growth over time. Since chimpanzees are the closest living relatives of the human race, gaining insight into their behavior can reveal a lot about ourselves. Therefore, accurately being able to predict the size of chimpanzees would contribute greatly to research and conservation efforts.

Our customer, Dr. Zarin Machanda, who teaches here at Tufts, is currently working on research that would require the annotation of several hundred images to extract accurate measurements of the East African subspecies of chimpanzees (Pan troglodytes schweinfurthii) in the community on her field site in Kanyawara, Kibale National Park, Uganda. Her field assistants use a Canon EOS 40D as their camera to take pictures whenever convenient while gathering focal data on the chimpanzees. The camera used consists of a digital SLR with a laser box mounted on it. The laser box performs parallel laser photometry: a single laser enters a beam splitter and two parallel lasers exit and hit the bodies of the chimpanzees at a set distance apart. This parallel laser technique employed has been modified from a previous technique that was successfully used on red colobus monkeys (Rothman et al., 2008). The exact distance varies based on the laser system being used but is documented such that the ratio between the actual distance and the pixel distance between the laser pointers can be used to find the actual distance between the shoulder and rump on the bodies of the chimpanzees. Earlier, the team was using green laser pointers to project onto the bodies of the chimpanzees. However, more recently, Dr. Machanda's team has shifted to using red laser pointers, which the chimpanzees seem to be less startled by. Dr. Machanda's team has a combination of green and red laser pointer pictures — although they are clearly

separated into different datasets — and incoming images from the field site will only contain red laser pointers.

In order to use this ratio, the images have been quality checked for focus and angle such that the body of the chimpanzee is perpendicular to the camera with the shoulder and rump visible and the chimpanzee is in a quadrupedal position, which means that it is on all fours. This measurement allows us to estimate the spinal length of the chimpanzees, which has been found to be a fair determinant of the body size of chimpanzees since it is a measure of skeletal length. Among adults, this measurement should be fixed through the rest of their lifespan, while for infants, the change in this measurement reflects skeletal growth, which is a key aspect of chimpanzee development and health.

The problem that is currently faced is that Dr. Machanda's team must manually measure each of the chimpanzee images to find the size estimate, which can be tedious. Currently, her team spends approximately 2 minutes per image to calculate by hand the size of the chimpanzee pictured. For hundreds of images, this can total to an incredibly large amount of time, which her team can spend in other ways, even perhaps in gathering more data from the field site. If this process can be automated by a pipeline that can process the images of the chimpanzees and output the shoulder-to-rump measurements in inches, this would save the team a great deal of time. Dr. Karen Panetta, Director of the Simulation Laboratory at Tufts University, took up this project for her lab. Under the guidance of her Ph.D. student, Obafemi Jinadu, we have worked on this project by using AI such as mass segmentation and pose estimation to find the body size of the chimpanzees.

2. RELATED WORK

With the advancement of computer vision techniques like instance segmentation and pose estimation, researchers are increasingly applying these techniques to unlock new discoveries in areas like animal neuroscience and animal husbandry. Recent research involving deep learning has identified some advances and challenges in animal pose estimation and has discussed how neuroscience laboratories can leverage these practical, fast, and accurate tools for better quantification of animal behavior (Mathis & Mathis, 2020). Recent advances in computer vision address the challenging manual, on-body measurement methods that are required to obtain livestock phenotypic data (Ma et al., 2024). To combat this, the authors present various computer vision techniques, such as 3D reconstruction technology and live animal weight estimation, to estimate the body size and weight of animals. Similarly, a combination of computer vision techniques and regression machine learning have been applied to weigh live sheep in farms (Sant'Ana et al., 2021). These non-contact tools reduce the stress on the animals, limit economic losses, and are rapid and efficient. We employ such a tool, Meta's Segment Anything Model (SAM), to conduct instance segmentation on our images (Kirillov et al., 2023).

3. METHODOLOGY

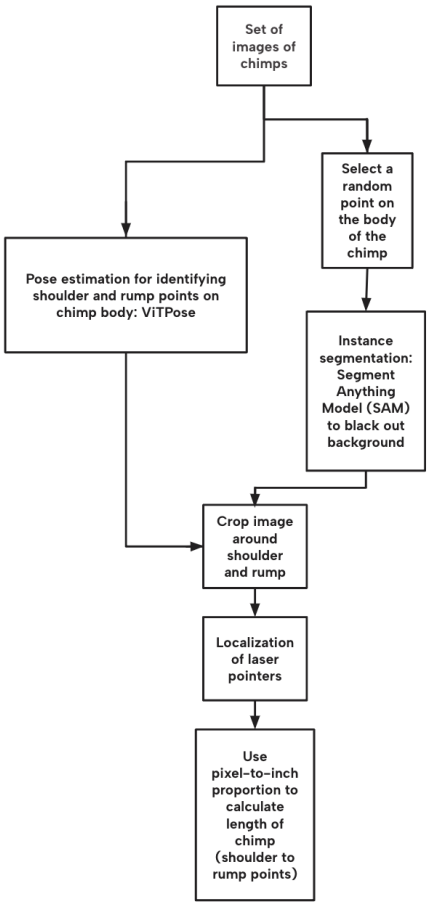

Fig. 1: Diagram Showing Pipeline Flow

3.1. Animal Pose Estimation

Pose estimation is a computer vision task that involves estimating the pose of a person from an image or a video by estimating the locations of key body joints (Xu et al., 2022). A pose estimation model, for example, takes a camera image as input and outputs estimates about where key body joints are (e.g., the elbow is located at (x, y)). These joints can be connected to draw a stick skeleton of the pose.

Pose estimation is challenging due to variations of occlusion, truncation, scales, and human appearances. To combat these, deep learning-based methods have been developed that typically use convolutional neural networks. Recently, vision transformers

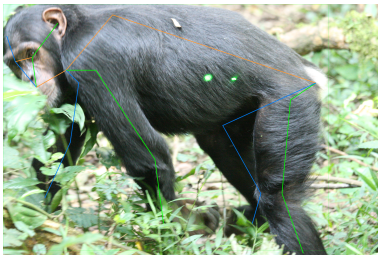

Fig. 2: ViTPose on Chimpanzee

have shown great potential in many vision tasks. As a result, different vision transformer structures have been deployed for pose estimation. One such example is ViTPose. Unlike other models, ViTPose leverages plain vision transformers.

As described by its authors, "ViTPose employs plain and non-hierarchical vision transformers as backbones to extract feature maps for the given person instances, where the backbones are pre-trained with masked image modeling pretext tasks, to provide a good initialization. Then, a lightweight decoder processes the extracted features by upsampling the feature maps and regressing the heatmaps w.r.t. the keypoints, which is composed of two deconvolution layers and one prediction layer" (Xu et al., 2022).

Additionally, ViTPose follows the common top-down setting for human pose estimation. In this top-down approach, a detector is first used to detect person instances and ViTPose is then employed to estimate the body joints and other keypoints within the detected bounding boxes (Xu et al., 2022). Despite no elaborate designs in the model, ViTPose obtains state-of-the-art performance on the MS COCO Keypoint dataset, which is a large-scale object detection, segmentation, and captioning dataset with 80 object categories (objects easily picked up or handled, such as animals, vehicles, and household items) and 91 stuff categories (background or environmental items such as sky, water, and roads) (Yang et al., 2022). In fact, a version of the ViTPose model with a very big vision transformer model as the backbone (ViTAE-G) has the best mean average precision (AP) on the MS COCO Keypoints. Beyond this strong performance, ViTPose presents other advantages: simplicity in model structure, scalability in model size, flexibility in training paradigm, and knowledge transferability. Due to these advantages and its superior performance, we chose to employ a version of the ViTPose model to perform pose estimation on the chimpanzees in our dataset. Specifically, we chose to use the ViTPose+-H model as it is the model in the open-source ViTPose library with the best performance on the APT-36K dataset (Xu et al., 2022). The APT-36K dataset contains 36,000 still frames of 30 animal species. High-quality keypoint annotations (manually labeled) are provided for all of the animal instances (Yang et al., 2022). Because this dataset was created as a benchmark for animal pose estimation, we determined that performance on APT-36k is most relevant for our task of chimpanzee pose estimation. Details for the ViTPose+-H model are shown below (Xu et al., 2022).

Model	Dataset	Resolution	AP
ViTPose+-H	COCO+AIC+MPII+AP10K+APT36K+WholeBody	256x192	82.3

Fig. 3: Details for ViTPose+-H Model

3.2. Instance Segmentation

A key component of our pipeline required separating the chimpanzee from its background. To achieve this, we use instance segmentation, a deep learning algorithm that outlines and separates individual objects within an image, providing a detailed mask of the desired object. For our pipeline, we implemented a model called Segment Anything Model (SAM), developed by Meta AI. SAM was trained on 11 million images and 1.1 billion segmentation masks. SAM's design allows it to adapt to new image distributions and tasks without prior knowledge, a feature known as zero-shot transfer. The model employs a powerful image encoder, a prompt encoder, and a lightweight mask decoder. This unique architecture enables flexible prompting and real-time mask computation. Our choice of this model was primarily based on its ability to detect a wide range of objects and in our cases animals without the need of additional training(Kirillov et al., 2023).

Fig. 4: Segmentation Ouptut

3.3. Blob Detection

In order to detect laser points on the chimpanzees, we utilized a blob detection algorithm from scikit-image, specifically, Laplacian of Gaussian. This method combines Gaussian smoothing with the Laplacian operation to enhance the detection of blob-like structures in an image. First, the image is smoothed with a Guassian filter to reduce noise and details in the image that are smaller than the blobs you are interested

in detecting. Next, the smoothed image is processed with a Laplacian filter. The Laplacian is a second-order derivative filter that highlights regions of rapid intensity change, which are typically the edges of blobs. This process is performed at multiple scales to detect blobs of different sizes (van der Walt et al., 2014).

To enhance this operation and mitigate mislocalization we applied a mask on the chimpanzee to isolate its torso. This specific mask was generated by overlaying two masks, the output from SAM along with a geometric mask to remove the chimpanzees' head and rear.

4. PIPELINE SETUP

Our system expects images of chimpanzees with the following three characteristics:

1. Two green/red laser pointers on the chimpanzee's body are a known distance apart. This distance is guaranteed in our image set (as explained in the Background).
2. The photographed chimpanzee is perpendicular to the camera (i.e., facing a direction perpendicular to the lasers).
3. The chimpanzee is photographed on all fours.

Fig. 5: Ideal Photo

Provided a set of images satisfying these criteria, our system follows the below steps to compute a shoulder-to-rump measurement for each chimpanzee:

1. Our system first prompts the user to position an input point to be placed anywhere on the body of the photographed chimpanzee in focus.
2. Meta's Segment Anything Model (SAM) leverages the provided input point to identify and separate the chimpanzee in the image from the rest of the background of foliage. The segmentation mask from SAM is used to convert this background to black.
3. ViTPose (pose estimation algorithm) is run on the original image of the chimpanzee to predict the locations of key joints, including the chimpanzee's shoulder and rump.
4. ViTPose's prediction points are used to compute the distance between the chimpanzee's shoulder and rump in pixels.
5. The predicted shoulder and rump points are used to crop the modified image (from step 2). The crop is determined by calculating boundaries perpendicular to the line that passes through the shoulder and rump points identified by ViTPose.
6. This cropped image is parsed to identify and localize the red/green laser pointers on the chimpanzee's body. The distance between the laser pointers in pixels is now computed and saved.
7. The laser pointer distances (pixels and centimeters) are leveraged to convert the shoulder-to-rump measurement from pixels to centimeters:

$$\frac{\text{laser dist. (pixels)(known)}}{\text{laser dist. (cm)(known)}} = \frac{\text{shoulder-rump dist. (pixels)(known)}}{\text{shoulder-rump dist. (cm) (unknown)}}$$

8. To calculate our desired unknown metric, we solve for the shoulder-to-rump distance (in cm) in the above equality to get:

$$\text{shoulder-rump dist (cm)} = \frac{\text{shoulder-rump dist. (pixels)} \times \text{laser dist. (cm)}}{\text{laser dist. (pixels)}}$$

5. APPROACH

The dataset provided by our client consists of 347 quality-checked images, satisfying the criteria described earlier. We were provided with the ground truth measurements (based on manual calculations on the images by Professor Machanda's team) for 202 of them, giving us a reference to analyze the performance of our model.

First working with this subset of 202 images, we passed them through our pipeline as specified in the section above. For best performance, we had to set it up in an environment that used a GPU runtime type. To do so, we set up a GitHub repo that we cloned and ran our image data through on Tufts' High-Performance Cluster.

After running the 202 images through our pipeline, we found that our model had consistent difficulty in certain cases. As a result, we decided to establish a stricter set of criteria for quality-checking our dataset. This allowed us to narrow down to a set of images, which we refer to as a "perfect subset" for comparison. To analyze the performance of our model without external factors, we excluded images as follows:

- Images containing more than one chimpanzee. VitPose struggled to handle these cases accurately.
- Images in which one or both of the laser pointers were difficult to identify even with the human eye (ie, dim laser pointers).

Fig. 6: Example of Unusable Picture

- Images in which one or both of the laser pointers are blurred or obstructed by a glare.
- Images with other red spots (e.g., ears) besides the two laser pointers.

As we assessed these failure cases, we additionally noticed that the laser point detection code had problems with images featuring other spots of red. Specifically, there was a pattern of mistaking the red ears and behinds of the chimpanzees for laser pointers. To manage these cases, we added an additional step in our pipeline, step 5 as seen above. The purpose of this step was to crop the images around the torso for a more focused view.

We then ran the remaining 145 images through our pipeline to produce the desired measurements of the chimpanzees to provide our client with.

6. RESULTS

We applied our pipeline and produced results for three cases: 1) the dataset of 202 images, 2) the "perfect subset" of 60 images, and 3) the total dataset of 347 images. Note that we had ground truth measurements in cases 1 and 2 only. The results produced in case 3 are our client deliverable.

Full Labeled Dataset of 202 images:

Via our model, we calculated the body size measurement associated with each image. Having access to the ground truth measurements, we calculated the percent error on each prediction, by dividing the difference between the hand-calculated and pipeline-calculated shoulder-to-rump distances by the hand-calculated distances. Referencing the magnitude of our measurements, we identified that a threshold of 10% error was reasonable. From our list of ground truth shoulder-to-rump measurements, we saw that the average body size of a chimpanzee is about 51.73 inches, which means that thresholding at a level of 10% error would result in a maximum of a 5-inch deviation from the true body size of the chimpanzee. More importantly, we were able to confirm later that there was a fairly normally distributed percent error centered at around 0% error so the 10% error on average does not bias towards underestimation or overestimation of body size. We have been trying to get our customer to confirm this as an acceptable threshold of error.

On the dataset of 202 images, 35 images failed at some point in the pipeline. These failures were disregarded in our accuracy analysis. Of the remaining 167 images, 114 images had predictions with an absolute percent error of less than 10%, giving our system an accuracy of 68.3%. The median absolute percent error was 5.8%, the minimum was 0.2%, and the maximum was 470.8%. The mean absolute percent error of 24.27% was driven up by several outliers, and these failure cases will be discussed later in this section. The distribution of absolute percent errors can be seen in Figure 7 for the following:

1. **Final Error:** predicted shoulder-rump distance (in) vs. actual shoulder-rump distance (in)
2. **Laser Error:** predicted laser distance (pixels) vs. actual laser distance (pixels)
3. **Body Error:** predicted shoulder-rump distance (pixels) vs. actual shoulder-rump distance (pixels).

	Measurement	Mean	Median	Std Dev	Min	Max
0	Final Error	27.289820	5.8	64.336236	0.2	470.8
1	Laser Error	21.666467	1.0	61.328892	0.0	424.6
2	Body Error	6.129940	4.3	7.900993	0.1	67.9

Fig. 7: Summary of Absolute Percent Errors

The distributions pictured in Figure 8 and Figure 9 display the spread of non-absolute percent errors among the predictions on the fully labeled dataset.

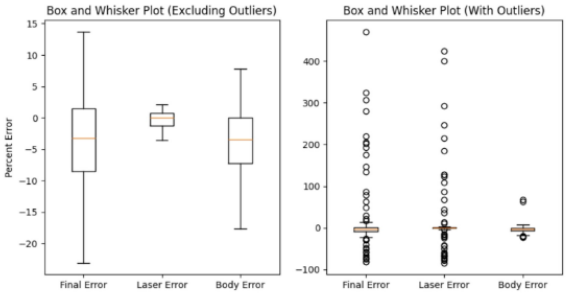

Fig. 8: Scatterplots of Non-Absolute Percent Errors

6.1. FAILURE CASES

In narrowing down from the full dataset of labeled images (size 202) to the perfect subset (size 60), we

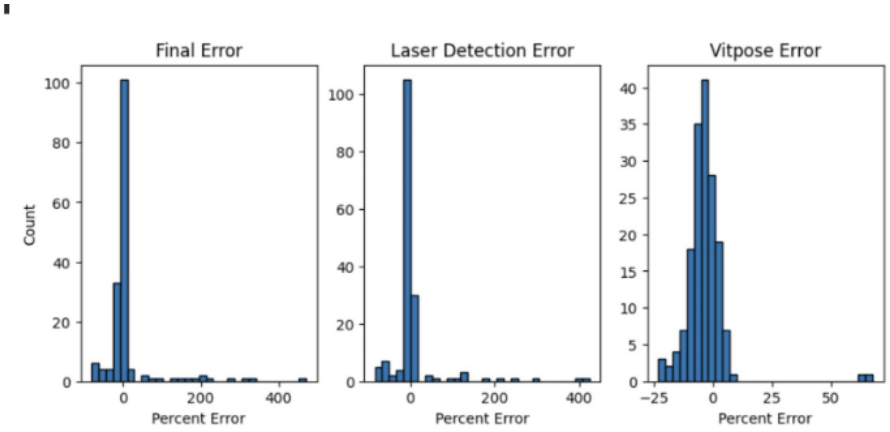

Fig. 9: Histograms of Non-Absolute Percent Errors

identified certain failure cases that resulted in bad performance for either our laser detection code or VitPose.

Laser Detection Failure cases:

- **Glare:** In certain images in our dataset, there is a bright glare that obscures the red laser pointers. The glare makes the laser pointers difficult to see even with the naked eye and oftentimes prevents our detection algorithm from identifying the laser pointers.

  In this first example, the laser fails to recognize the laser pointers due to the glare and incorrectly identifies other spots on the chimpanzee's body as laser pointers.

Fig. 10: Bright Glare Example

  In this second example, the laser detection algorithm fails to recognize the red laser pointers and returns no identified points at all.

Fig. 11: Failure to Recognize Laser Pointers

- **Blurred Laser Pointers:** In an ideal image, the red laser pointers are circular dots on the chimpanzee's body. However, at times, the laser pointers can appear blurred and thus smeared / "spread out" on the chimpanzee's body. When the red of a laser pointer is smeared across a large enough length on the chimp's body, our laser detection may incorrectly identify two locations on the single blurred laser pointer or fail to identify the center of the blurred laser pointer. This results in our system determining a poor estimate for the pixel length between the laser pointers. To solve this, we attempted to set a constant minimal length between the laser pointers. If the algorithm finds two laser pointers in an image with a distance less than our minimal length, then it is forced to continue until it identifies two laser pointers with a distance above the minimal length. This solution was successful in certain cases but did not generalize well to all images due to the dramatic amount of variation in how zoomed-in each image is (i.e., in some images the distance between the laser pointers is $\sim$30 pixels whereas the circumference of a single laser dot in others is $\sim$30 pixels).

  The below example shows an image in which the laser detection code identifies two points on one laser dot due to the laser points being blurred/smeared on the chimpanzee's body.

Fig. 12: Smeared Laser Pointers

  In the below example, the chimpanzee does not identify two points on one laser dot, but it misplaces one of the dots due to the dot being blurred/spread out on the chimpanzee's body.

Fig. 13: Blurred Laser Pointers

- **Red Coloration that Resembles Laser Pointers:** Some of the photoed chimpanzees have red coloration on places like their ears, fingers, and rear end that the laser detection code can incorrectly identify as being laser pointers. Examples of these are shown below.

  In this first example, a laser point is incorrectly identified on the red of the chimpanzee's behind.

Fig. 14: Incorrect Laser Pointers Identification: Behind

In the below two examples, a laser point is incorrectly identified on the red of the chimpanzee's ear (the blue points are the detected laser points).

Fig. 15: Incorrect Laser Pointers Identification: Ear

Fig. 16: Incorrect Laser Identification: Ear

In this second example, a laser point is incorrectly identified on the red of another chimpanzee's finger.

Fig. 17: Incorrect Laser Identification: Finger

- **Dim Laser Pointers:** In certain images, the laser pointers are not identified by our laser detection algorithm because they are dim and lack red color. In the below examples, the laser detection algorithm failed to detect both laser pointers because at least one of them was too dim in color:

Fig. 18: Failure to Recognize One Laser Point

Pose Estimation Detection Failure Cases:

- **Multiple Chimpanzees in the Image:** ViTPose performs poorly on images with multiple chimpanzees. Provided an image with multiple chimpanzees, we found that ViTPose may fail to draw the poses of all of the pictured chimpanzees. In these cases, if the chimpanzee to be measured (i.e., the one with laser pointers) does not have a stick skeleton, then the chimpanzee cannot be measured. Additionally, for images with a child chimpanzee on top of a mother chimpanzee, VitPose occasionally considers both chimpanzees as one animal and draws a deformed skeleton.

  Below is an example of VitPose not drawing the pose of the chimpanzee we are interested in measuring due to the image containing multiple chimpanzees:

Fig. 19: Multiple Chimpanzee Error

Below is an example of VitPose treating a mother and child chimpanzee as one animal and thus incorrectly identifying the shoulder and rump of the chimpanzee to be measured.

Fig. 20: Two Chimpanzees as One

- **Foliage Obscuring Parts of the Chimpanzee:** VitPose performed poorly for images where foliage covered parts of the chimpanzee and obscured the view of the full chimpanzee. See the below examples of this: In this first example, the rump of the chimpanzee is partially obscured by the tree branch. Due to this, ViTPose seems to compensate by making a rump prediction far closer to the shoulder than the rump is.

Fig. 21: Branch Covering Rump

Branches and leaves obscure parts of the chimpanzee, including parts of the chimpanzee's arm, back legs, and rump. VitPose, in turn, incorrectly marks the shoulder point too far down the chimpanzee's right arm and the rump point too far down the chimpanzee's back leg.

Fig. 22: Branch Covering Arm, Leg, and Rump

"The Perfect Subset"

From identifying these patterns in the failure cases, we created a "perfect subset" of 60 images from the set of 202 images with ground truth values. This perfect subset consisted of images where the laser pointers were easily visible to the human eye, the entire torso of the chimpanzee was visible in the frame of the image, and there was only one chimpanzee pictured. From here, we repeated the process and analysis that was used for the dataset of 202 images. The "perfect subset" of 60 images returned 53/60 images with less than 10% absolute prediction error, for an accuracy of about 88%. The median absolute percent error was 4.15%. There were still failure cases within this subset although the maximum percent error decreased and there were fewer outliers than the full 202 image dataset. Figure 23 provides the statistics on running this perfect subset of images through our pipeline on the absolute percent error. We see that the mean is consistently (but only slightly) higher than the median, indicating an influence of a couple of outliers.

	Measurement	Mean	Median	Std Dev	Min	Max
0	Final Error	5.683333	4.15	6.035373	0.3	39.7
1	Laser Error	2.371667	0.85	8.894045	0.0	66.4
2	Body Error	4.473333	3.75	3.525047	0.1	16.8

Fig. 23: Distribution of Absolute Percent Errors

Drilling down into this observation, we create a distribution of prediction errors as pictured in Figure 26. We see that there are a couple of outliers with large magnitude percent error coming from the laser detection. We also see a case where both ViTPose and the laser pointer localization do a fairly accurate job at predicting, as seen in Figure 24, but they both overestimate the size of the chimpanzee: the laser pointers are marked closer than they are while ViTPose chooses a shoulder point closer to the front, rather than closer to the ear of the chimpanzee. This compounds the error that we see on the image, resulting in an error of 21.7%.

Fig. 24: Incorrect Laser Detection

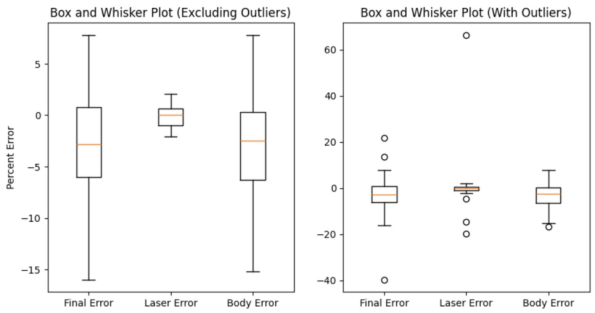

Fig. 25: Box and Whisker Plots

Nevertheless, excluding these outliers, from the boxplot in Figure 25, we see that the range of the prediction is smaller for the laser point localization

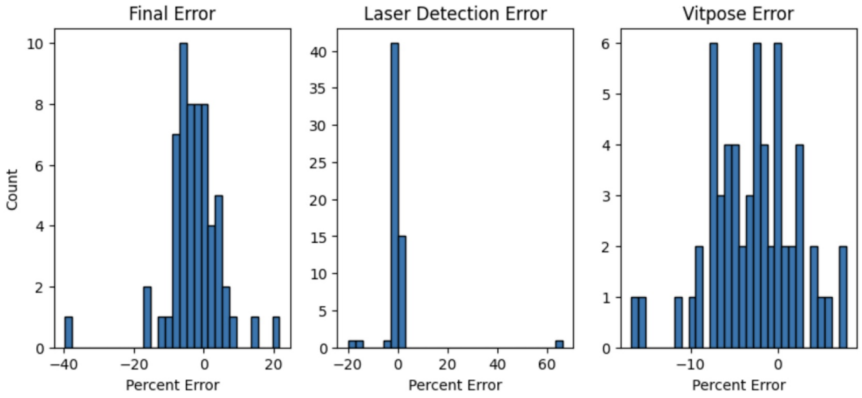

Fig. 26: Distribution of Prediction Percent Errors

as compared to the ViTPose prediction. This makes sense because wherever we see the laser point error, it is usually when there is a detection of a red area near the rump or the face of the chimpanzee pictured (which is far away from the center of the body of the chimpanzee, where the laser pointers tend to be), whereas the shoulder and rump points on the chimpanzee's body can be more ambiguously determined.

Finally, we ran the entire dataset, provided by our client, through our pipeline. We exported our results for the 145 images that we weren't provided ground truth measurements for into a spreadsheet with columns "Image name" and "Predicted shoulder-to-rump measurement." These will be delivered to our client as Professor Machanda's team does not have measurements for these images. The distribution in Figure 27 shows the spread of chimpanzee sizes, with estimated spine length averaging 58.66 inches. The standard deviation of our predictions is 44.67 inches. While we do not have information about the age, sex, or names of the chimpanzees in these 145 images, our client and their team of chimpanzee experts will be able to use these predictions with further context.

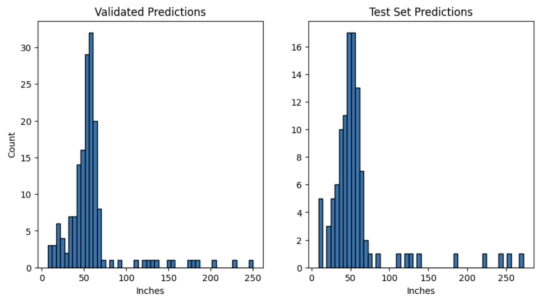

Fig. 27: Distribution of Predictions

7. DISCUSSION

Although non-contact tools such as SAM for identifying animal images have their benefits, they can struggle with variations in lighting conditions and the posture of livestock. Ji et al. (2023) explore in detail these conditions under which modern computer vision technologies, particularly SAM, do not perform well. They show that SAM performs especially poorly in concealed scenes involving camouflaged animals (e.g., a camouflaged owl in a tree), industrial defects, and medical lesions. Further, Yang et al. (2023) address the scarcity of available animal behavior data and the high cost of labeling a large amount of such data. To address these issues, the authors propose a novel approach that leverages instance segmentation-based transfer learning to not be heavily reliant on labeled data, and they integrate the system with SAM to significantly improve efficiency and accuracy. The resulting system supports multiple animal tracking and behavior analyses, and the authors show the exceptional, human-level performance of the method through a series of experiments, demonstrating that it is a valuable asset to animal behavior researchers.

8. CONCLUSION

The ability to measure the size of chimpanzees in the wild is important in understanding aspects of their behavior and ecology. This can aid biologic research and conservation efforts, yet obtaining such measurements is difficult when maintaining non-invasive methods. The purpose of our research was to develop a computer vision technique for measuring the skeletal growth of wild chimpanzees. In collaboration with Dr. Zarin Machanda's research team, we created and tested a pipeline for obtaining the shoulder to rump measurement of known chimpanzees.

To summarize our pipeline: we began with perpendicular images of chimpanzees in a quadrupedal position. Each image contained two laser points on

the chimpanzee of interest's body and gave us a reference of distance. We used Meta's Segment Anything Model to separate the chimpanzees from their background. The segmentation mask was then used to convert the background to black. From here, we parsed the modified image to localize the laser pointers. We then used ViTPose (a pose estimation algorithm) on the original image to predict the locations of key joints. Extracting the identified chimpanzee's shoulder and rump gave us a relevant distance to measure. The laser pointer distances (pixels and centimeters) were leveraged to convert the shoulder-to-rump measurement from pixels to centimeters as desired.

Throughout this process we had success in using SAM and ViTPose to manipulate our dataset of images. Having manually calculated shoulder-to-rump measurements of each chimpanzee, we were able to compare our calculated results and found reasonable results once further quality checking the dataset. As discussed in the 'approach' section, our research led us to develop new criteria for quality-checking images which will be communicated to our client. Another draw of our approach is its automated nature and time/computation benefit. Rather than relying on time consuming manual measuring, we were able to measure 347 images in 47 minutes, a process which our client estimates would have previously taken 694 minutes (about 11.5 hours).

In conclusion, we contribute a non-invasive method for measuring the shoulder to rump size of chimpanzees in the wild. From images taken at a distance, we were able to calculate a relevant measurement that can benefit wildlife researchers and wild animal populations alike.

9. FUTURE WORK

Based on the discussion of the results, the following would be logical avenues to explore if we were to expand on this project:

1. To tackle laser point localization error, we can spend more time analyzing the data to identify patterns in where the laser pointers are placed on the body of the chimpanzees. Perhaps if they are always placed closer to the middle of the torso we can further restrict the mask within which our code searches for the laser pointers.

2. Apply a lower and upper bound on body length measurements on a chimpanzee to filter out outliers produced by our pipeline that should be checked with manual measurements. If we are provided with further metadata on each image such as the age of the chimpanzee we can also work with our client to figure out size ranges that are customized for each chimpanzee that would flag measurements out of the range and allow our client's team to manually check on these examples.

3. When experimenting with a different implementation of ViTPose from a different repository that used the yolov8 model, it was able to successfully perform pose estimation on a segmented image. We should further explore this method as an alternative to our current usage or any other methods we can use in our current implementation of ViTPose on a segmented image with only one chimpanzee. This will eliminate the error of ViTPose choosing shoulder and rump points on the chimpanzee that does not have the laser points on its body.

4. Discuss adding new quality-checking criteria to check for clear laser pointers and the complete presence of chimpanzees within the frames of the images to use our pipeline.

5. We provided a system to compute a shoulder-to-rump measurement for each chimpanzee to assess skeletal growth. A possible next step is creating a system to measure change in mass. One way of doing so could involve using instance segmentation and pose estimation to draw a mask on the chimpanzee's torso. Speaking with our professor, this measurement would be an accurate way of assessing a chimpanzee's change in mass.

REFERENCES

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). *Segment Anything* (arXiv:2304.02643). arXiv. https://doi.org/10.48550/arXiv.2304.02643

Ma, W., Qi, X., Sun, Y., Gao, R., Ding, L., Wang, R., Peng, C., Zhang, J., Wu, J., Xu, Z., Li, M., Zhao, H., Huang, S., & Li, Q. (2024). *Computer Vision-Based Measurement Techniques for Livestock Body Dimension and Weight: A Review. Agriculture*, 14(2), Article 2. https://doi.org/10.3390/agriculture14020306

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. Current Opinion in Neurobiology, 60, 1–11. https://doi.org/10.1016/j.conb.2019.10.008

Rothman, J. M., Chapman, C. A., Twinomugisha, D., Wasserman, M. D., Lambert, J. E., & Goldberg, T. L. (2008). Measuring physical traits of primates remotely: The use of parallel lasers. American Journal of Primatology, 70(12), 1191–1195. https://doi.org/10.1002/ajp.20611

Sant'Ana, D. A., Pache, M. C. B., Martins, J., Soares, W. P., de Melo, S. L. N., Garcia, V., de Moares Weber, V. A., da Silva Heimbach, N., Mateus, R. G., & Pistori, H. (2021). Weighing live sheep using computer vision techniques and regression machine learning. Machine Learning with Applications, 5, 100076. https://doi.org/10.1016/j.mlwa.2021.100076

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... Contributors, T. S.-I. (2014). scikit-image: image processing in Python. PeerJ, 2, e453. doi:10.7717/peerj.453

Wildlife Chimpanzees. (n.d.). Wildlife Conservation Society. Retrieved March 30, 2024, from https://www.wcs.org/our-work/species/chimpanzees

Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. arXiv [Cs.CV]. Retrieved from http://arxiv.org/abs/2204.12484

Yang, Yuxiang, et al. "APT–36K: A Large-scale Benchmark for Animal Pose Estimation and Tracking." arXiv, Cornell University Library, 13 Oct. 2022, arxiv.org/abs/2206.05683.