Elephas:SAM: segmentation performance of Sumatran Elephant in captivity with segment anything model

Fortuno Ery Faqih[1*], Lukman Zaman P.C.S.W[2]
[1,2] Information Technology Department, Faculty of Science and Technology,
Institut Sains dan Teknologi Terpadu Surabaya, East Java, Indonesia
[1*]fortuno.e22@mhs.istts.ac.id, [2]lz@stts.edu

Abstract— **Surabaya Zoo is one of the conservation institutions in Surabaya, which has Sumatran elephants as a collection of endemic Indonesian animals. The Indonesian government protects this animal because of its endangered status. Installing CCTV cameras in elephant captivity helps us to utilize Segment Anything Model (SAM) technology as the initial basis for developing an animal monitoring system in captivity, especially Sumatran Elephant with artificial intelligence (AI). Our investigations differ from past research in that we utilize 60 exclusive images obtained from CCTV footage in an elephant enclosure at Surabaya Zoo over a 30-day period instead of using publicly available datasets. The image set was partitioned into 30 instances taken under Night-time (01:00 WIB) and 30 instances taken under Day-time conditions (15:00 WIB). We perform the evaluation of SAM's predictions scores using the SAM-Point Prompt and SAM-Box Prompt techniques. It was found that, on average, the segmentation prediction scores for 30 Night-time images are higher when the SAM-Point prompt is used (0.941) instead of the SAM-Box prompt (0.939), which is only a 0.002 difference. For a set of 30 vivid images, the SAM-Point Prompt produces a higher average score (0.989) than the SAM-Box Prompt (0.968), indicating a difference of 0.021. The result is that the use of the SAM-Point prompt is more effective than SAM-Box in predicting segmentation scores in Sumatran elephant images at night and during the day.**

Keywords*— *Artificial Intelligence, CCTV, Prediction Score, Segment Anything Model, SAM-Point Prompt, SAM-Box Prompt

I. INTRODUCTION

Surabaya Zoo is an important conservation organization that focuses on breeding the Sumatran Elephant (*Elephas maximus sumatranus*), which is an important species among endangered and protected animals. Sumatran elephants are classified as critically endangered by the International Union for Conservation of Nature (IUCN) [1]. This poses a major challenge, especially in regions such as Bengkulu, Indonesia, where conflicts with local populations have arisen due to increasing elephant populations and the need for large home ranges [2]. The importance of the function of conservation institutions as demonstrated by the Surabaya Zoo is clearly visible in maintaining the survival of these rare animals.

The research was carried out at the Surabaya Zoo, specifically at the Sumatran Elephant House, which had CCTV (closed-circuit television) cameras installed and appropriate research permits were obtained. The task of monitoring and evaluating caged animals using cameras is made difficult due to unpredictable changes in lighting conditions, which directly affect the quality of the images captured [3]. Close monitoring of animals in human care is critical to understanding their physical and psychological health, enabling the study of behavioral changes, and uncovering the root of the problem. Traditional observation technique, which rely on professionals present on site to perform manual evaluations, require a lot of effort and time [4]. The use of artificial intelligence (AI) is critical in increasing the effectiveness of animal monitoring via CCTV cameras, as emphasized by the valuable contributions of [5] [6], and [7].

In April 2023, META introduced the Segment Anything Model (SAM), an innovative method developed for the purpose of segmenting authentic images. SAM, an AI model with the ability to recognize objects in images and videos, is renowned for its versatility and suitability for a wide range of image segmentation applications, without requiring retraining or customization. By training on a dataset of over one billion accurate segmentation masks derived from over 11 million real-life images, SAM has gained the ability to perform "zero-shot" image segmentation. This positions SAM as a powerful tool for image segmentation problems, comparable to the effectiveness of Chat GPT in natural language processing [8].

Although SAM has achieved success in medical image segmentation [9][10][11][12][13][14][15][16][17][18][19][20][21] and other practical uses [22][23][24][25][26], its potential for segmenting captured images of Sumatran elephants has not been studied. This research attempts to meet this need by examining SAM's ability to effectively segment images, especially for Sumatran elephants in captivity.

This research aims to conduct one of the first comprehensive evaluations of SAM's accuracy in immediately segmenting images of captive-bred Sumatran elephants at the Surabaya Zoo, without the need for retraining or fine-tuning. The evaluation involved applying zero-shot SAM to 60 image

datasets, covering a variety of illumination situations such as when night time and day time.

Another important aspect relates to SAM's ability to perform segmentation with minimal input, requiring only a bounding box or a single point as a reference from prediction score [8]. This capability has the potential to reduce human labor during the annotation process. SAM, on the other hand, presents an opportunity to alleviate this time-intensive task.

II. METHOD

A. Overview

The research has been approved by Surabaya Zoo. Figure 4 depicts the all-encompassing structure of our approach. The left panel provides a comprehensive summary of the data collecting procedure for capturing images of Sumatran elephants using closed-circuit television (CCTV) cameras within their enclosure. The right panel illustrates the assessment of SAM for advanced image segmentation in two scenarios: Night-time and Day-time

B. Datasets

The datasets used in this study consists of 60 images of Sumatran elephants in captivity. It includes 30 images taken in night-time situations and another 30 images taken in day-time conditions. Furthermore, a datasets of visual information containing Sumatran elephants was acquired from closed-circuit television (CCTV) recordings captured inside the enclosure. The recordings were made continuously for a duration of 30 days, starting from August 27, 2023, and ending on September 25, 2023. The dataset was meticulously selected by choosing a single image from Night-time settings, particularly at 01:00 AM (WIB), and Day-time conditions, precisely at 3:00 PM (WIB). The datasets are summarized in Fig. 1 (left panel) and listed with more detail in Table 1

C. SAM: brief introduction

The Segment Anything Model (SAM) is built utilizing the largest segmentation datasets available, which includes more than 1 billion accurate segmentation masks obtained from 11 million authorized and privacy-conscious natural images [8]. Upon its debut, SAM's performance was evaluated on 23 datasets that included a wide range of natural images [8]. SAM exhibited exceptional precision in zero-shot applications for these pictures, outperforming other interactive or dataset-specific models without requiring re-training or fine-tuning on novel, unknown datasets and segmentation tasks.

Fig. 1 depicts a SAM architecture comprising an image encoder, specifically a masked auto encoder, which is responsible for extracting embedded pictures. Additionally, there is a prompt encoder that utilizes different sorts of prompts, and a mask decoder that generates masks [8].

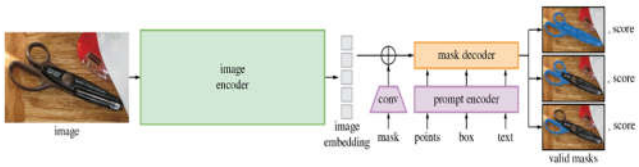

Figure 1. Architecture of SAM

Image Encoder: The core component of this model is a masked auto encoder that employs a vision transformer to achieve exceptional scalability. The ViT-H/16 is a robust vision transformer model specifically designed to process patch sizes of 16x16. The model includes a 14x14 windowed attention mechanism and four global attention blocks that are evenly distributed.

The result produced by the aforementioned encoder is a feature embedding, which is a downscaled replica of the original image, decreased by a factor of 16. The technique of streamlining is crucial for achieving efficient processing while maintaining essential image qualities. This model accepts input images with a size of 1024x1024 pixels and 3 color channels, which is common for high-quality photos. It then transforms this input into a dense representation with a resolution of 64x64 pixels and 256 channels.

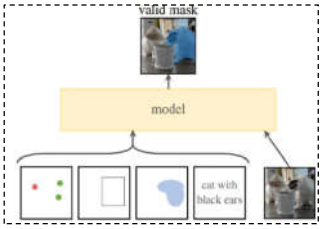

Figure 2. Prompt Encoder

Prompt Encoder: There are two categories of prompts as in Fig. 2: sparse prompts, which consist of points, boxes, and text, and dense, which consist of a mask. As a result of limited time, this study only utilized two types of prompts: points and boxes prompts. A point is denoted by adding the positional encoding of its location to one of the two learnt embedding, which indicates whether it is a foreground or background point. Embedding pairs are used to represent boxes. The positional encoding of the top-left corner is combined with a learnt embedding that represents the concept of the "top-left corner". The same process is applied to the bottom-right corner.

D. Application of SAM to Real-World (Sumatran Elephant in Captivity)

Fig. 4 shows the SAM model with two setting. SAM-Point (single-point prompt setting): We used the center of the body or head of the object (the elephant) as the point prompt. The parameters were configured to the default values, which were identical to those specified in the original SAM article [8]. When three masks are acquired for distinct regions/structures in the image, we choose the mask with the highest score for each mask and employ it for segmentation.

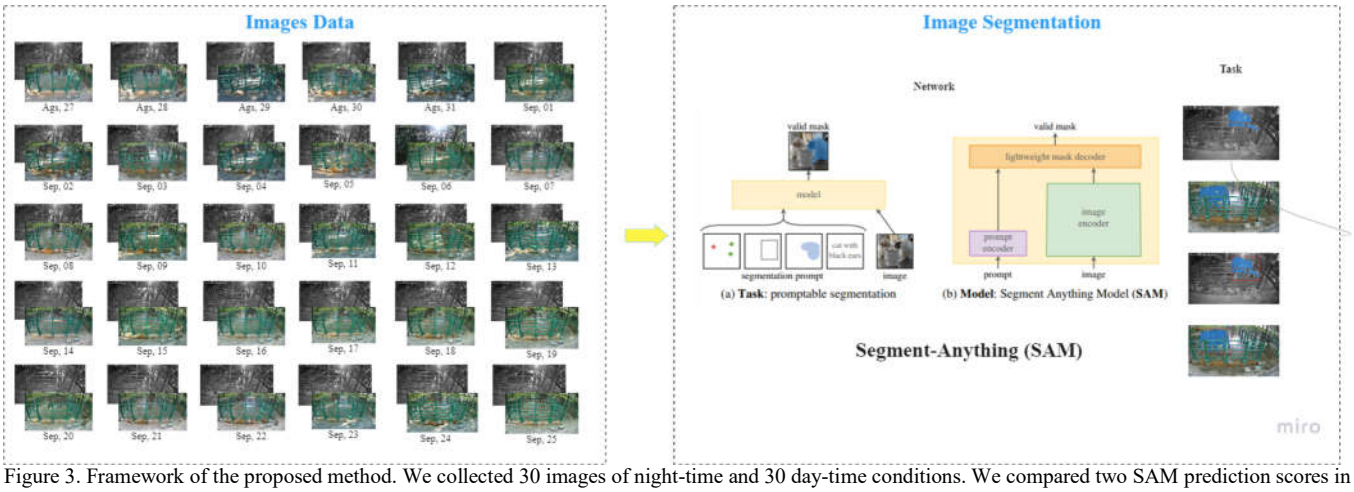

Figure 3. Framework of the proposed method. We collected 30 images of night-time and 30 day-time conditions. We compared two SAM prediction scores in two conditions with two prompts (SAM-Point and SAM-Box), where SAM was trained on more than 1 million natural images.

TABLE I
DATASET CHARACTERISTICS

Datasets	Segmentation Task	Elephant Conditions		Images
		Night-Time	Day-Time	
Augustus 27, 2023	body	standing	standing	2
Augustus 28, 2023	body	standing	standing	2
Augustus 29, 2023	body	standing	standing	2
Augustus 30, 2023	body	lying	standing	2
Augustus 31, 2023	body	standing	standing	2
September 01, 2023	body	lying	standing	2
September 02, 2023	body	lying	standing	2
September 03, 2023	body	standing	standing	2
September 04, 2023	body	standing	standing	2
September 05, 2023	body	lying	standing	2
September 06, 2023	body	standing	standing	2
September 07, 2023	body	lying	standing	2
September 08, 2023	body	lying	standing	2
September 09, 2023	body	standing	standing	2
September 10, 2023	body	standing	standing	2
September 11, 2023	body	lying	standing	2
September 12, 2023	body	standing	standing	2
September 13, 2023	body	standing	standing	2
September 14, 2023	body	lying	standing	2
September 15, 2023	body	standing	standing	2
September 16, 2023	body	standing	standing	2
September 17, 2023	body	lying	standing	2
September 18, 2023	body	lying	standing	2
September 19, 2023	body	lying	standing	2
September 20, 2023	body	standing	standing	2
September 21, 2023	body	lying	standing	2
September 22, 2023	body	lying	standing	2
September 23, 2023	body	lying	standing	2
September 24, 2023	body	lying	standing	2
September 25, 2023	body	lying	standing	2
Total				60

Night time: between 01.00 - 02.00 WIB (Western Indonesian Time)
Day time: between 15.00 - 17.00 WIB (Western Indonesian Time)

SAM-Box (box prompt setting): We calculate the bounding box for SAM according to the position the elephant is in.

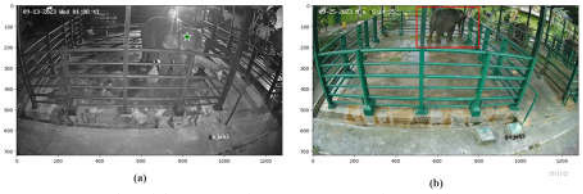

Figure 4. Example with two setting, a: SAM-Point and b: SAM-Box

SAM was explicitly applied to each dataset in both variants. No re-training or fine-tuning was conducted on each elephant image dataset, which is referred to as "zero-shot" application (refer to Fig. 3 right panel, shown by the yellow color). This method is identical to the first SAM test conducted on natural images [8].

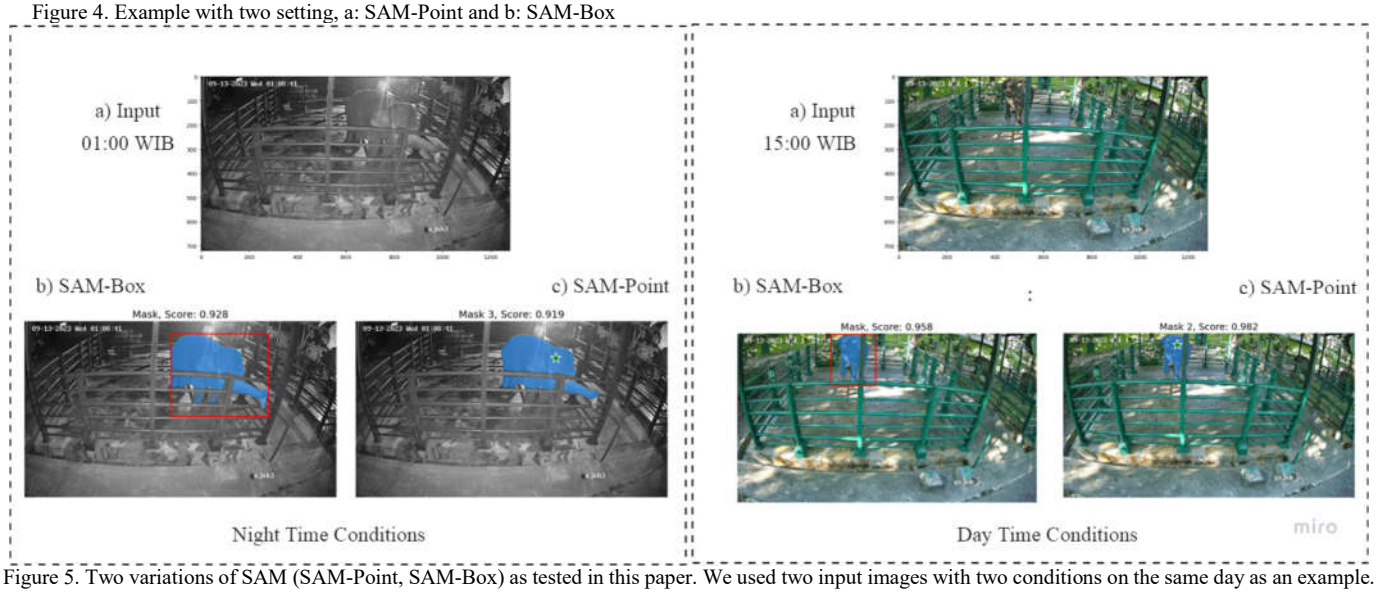

Figure 5. Two variations of SAM (SAM-Point, SAM-Box) as tested in this paper. We used two input images with two conditions on the same day as an example.

III. RESULTS AND DISCUSSION

In this part, we analyze the outcomes of experiments conducted using the Segment Anything Model (SAM) to assess the accuracy of score predictions for images of Sumatran elephants in enclosures under varying illumination conditions. We emphasize the utilization of SAM via Google Colab, its intuitive interface and the absence of a requirement for costly computational hardware

A. Prediction Score Using SAM-Point & SAM-Box

The test focused on assessing the efficacy of SAM in forecasting scores for images taken in Night-time and Day-time conditions. Fig. 5 shows usage with two different prompts: a SAM-point prompt and a SAM-box prompt in two conditions as an example.

Fig. 6 and 7 shows prediction scores that are dependent on both image Day-time ness and prompt type. When using a SAM-point prompt, the segmentation prediction scores for 30 Night-time images had an average of 0.941, whereas the scores for 30 Day-time images had an average of 0.989. Furthermore, the utilization of a SAM-box prompt resulted in distinct results, with an average score of 0.939 for 30 Night-time images and 0.968 for 30 Day-time images (Table 2)

TABLE II
AVERAGE SCORE USING SAM-POINT AND SAM-BOX IN TWO CONDITION

Prompt	Conditions	Average
Single-Point	Night time	0.941
Bounding Box	Night time	0.939
Single-Point	Day time	0.989
Bounding Box	Day time	0.968

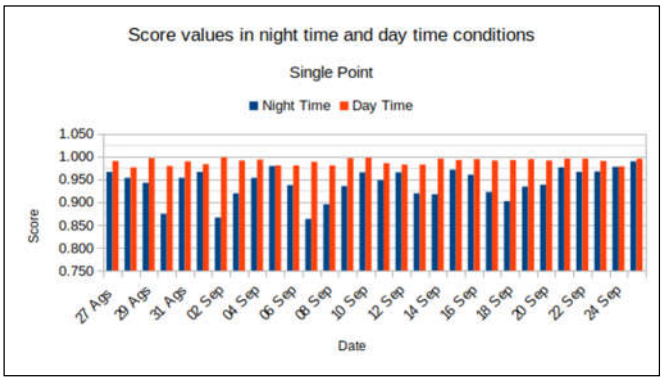

Figure 6. Prediction score SAM in Night-time and Day-time conditions using SAM-Point

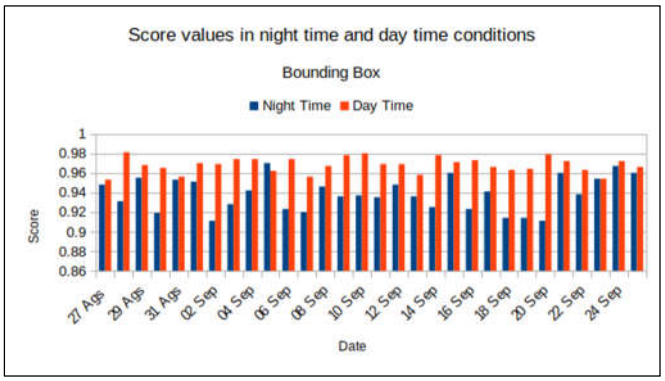

Figure 7. Prediction score SAM in Night-time and Day-time conditions using SAM-Box

B. *Comparing SAM-Point vs SAM-Box*

Significantly, when comparing the two forms of hints, it is clear that the segmentation prediction scores for the 30 Night-time images consistently show higher values when using SAM-point hints. This shows an increase of 0.002 compared to the bounding box approach. Similarly, when analyzing 30 Day-time images, the scores obtained using SAM-point clues were higher than those obtained using bounding box clues, and showed a significant difference of 0.021.

These findings suggest that the choice of prompt type, especially the utilization of SAM-point prompts, can greatly influence segmentation prediction scores. Figure 8 and 9 shows a comparison of score values between SAM-Point and SAM-Box in Night-time and Day-time conditions. Figure 10 and 11 shows several example images after being segmented using SAM-Point and SAM-Box in two conditions

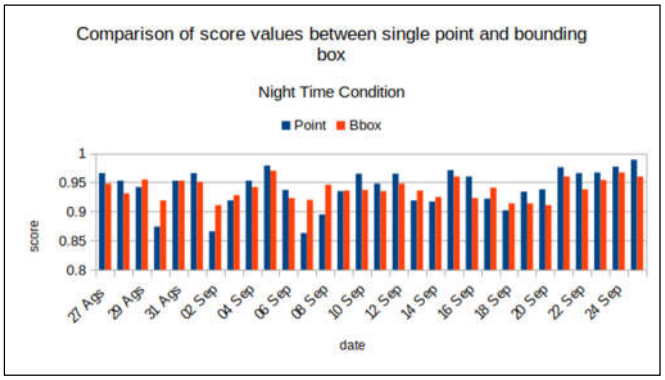

Figure 8. Comparing between SAM points and SAM boxes when Night-time conditions

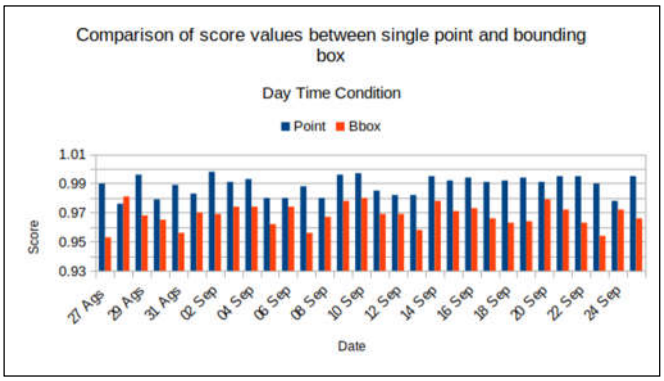

Figure 9. Comparing between SAM points and SAM boxes when Day-time conditions

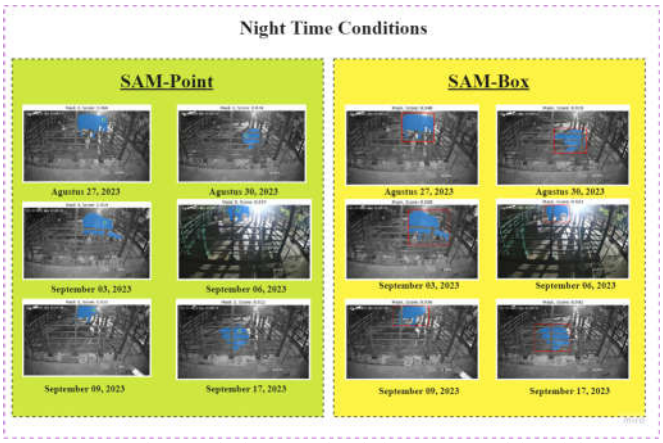

Figure 10. Several example images after being segmented in Night time Conditions

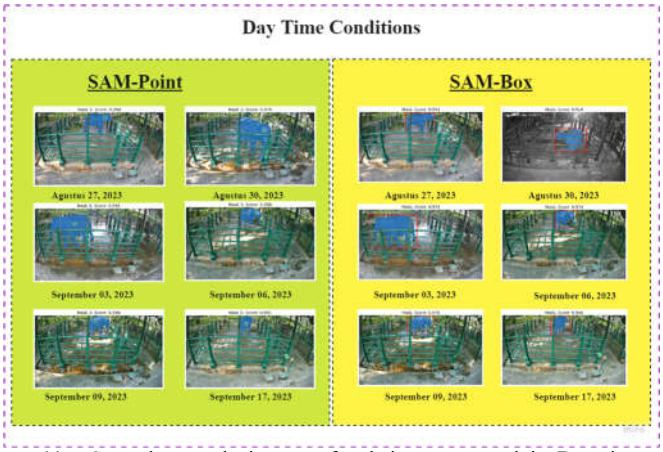

Figure 11. Several example images after being segmented in Day time Conditions

C. *Future Perspective on SAM for Animal Monitoring and Conservation Effort*

SAM has several advantages that make it an attractive choice for machine learning applications for animal monitoring in animal conservation institutions. First, it offers zero-shot generalization to unfamiliar objects and images without requiring additional training [8]. Second, SAM's interactive input process can significantly reduce the time and labor

required for manual image segmentation. The model's ability to generate segmentation masks with minimal input, such as a single point or a bounding box, accelerate the annotation process. This efficiency is crucial for large-scale animal monitoring application, where processing vast amounts of data on time is of utmost importance.

In short, our study focused on demonstrating the potential of SAM adaptability for animal monitoring application in two condition, night-time and day-time. There are some potential research direction and implementation for SAM in animal monitoring applications, which can be condensed as follows:

1. Examining the most effective approaches and techniques for adapting SAM to meet varying datasets with night and day conditions in captivity.
2. Evaluating the potential use of SAM as an alternative to traditional labelling processes and segmentation techniques in a weakly-supervised manner to boost its accuracy and reliability.
3. Investigating potential ways to integrate SAM with prevalent machine learning tools and platforms, such as TensorFlow, to augment the versatility and utility of these systems.

IV. CONCLUSION

Our experiments with images of Sumatran elephants under two conditions, Night-Time and Day-Time provide useful insights into the accuracy of Segment Anything Model (SAM) predictions. The datasets consist of 60 images from our exclusive CCTV collection. We perform the evaluation of SAM's predictions scores using the SAM-Point Prompt and SAM-Box Prompt techniques. It was found that, on average, the segmentation prediction scores for 30 Night-time images are higher when the SAM-Point prompt is used (0.941) instead of the SAM-Box prompt (0.939), which is only a 0.002 difference. For a set of 30 vivid images, the SAM-Point Prompt produces a higher average score (0.989) than the SAM-Box Prompt (0.968), indicating a difference of 0.021.

The result is that the use of the SAM-Point prompt is more effective than SAM-Box in predicting segmentation scores in Sumatran elephant images at night and during the day, no matter what condition the elephant is in, whether lying or standing. This experiment reveals that the prompt capabilities of SAM (point and box), combined with its ability to perform object segmentation with minimal human supervision, can also contribute to a significant reduction in annotation workload.

ACKNOWLEDGMENT

The author would like to thank the management and all related parties at the Surabaya Zoo for their efforts and sharing of data for this research

REFERENCES

[1] A. Gopala et al., "Elephas maximus ssp. sumatranus," IUCN Red List Threat. Species 2011, vol. 8235, no. e.T199856A9129626., pp. 4–29, 2011, [Online]. Available: http://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T199856A9129626.en

[2] A. Wira Putra Wahyudi Arianto, H. Suhartoyo, J. Kehutanan, F. Pertanian, and U. W. Bengkulu Jl supratman, "Mitigasi Konflik Manusia Dengan Gajah Sumatera (Elephas Maximus Sumatranus, Temminck 1847) Di Desa Binaan Lapindo Kabupaten Mukomuko Provinsi Bengkulu," J. Glob. For. Environ. Sci., vol. 3, no. 1, pp. 48–56, 2023

[3] S. Nazir and M. Kaleem, "Advances in image acquisition and processing technologies transforming animal ecological studies," Ecol. Inform., vol. 61, 2021, doi: 10.1016/j.ecoinf.2021.101212

[4] M. Zuerl et al., "Automated Video-Based Analysis Framework for Behavior Monitoring of Individual Animals in Zoos Using Deep Learning—A Study on Polar Bears," Animals, vol. 12, no. 6, 2022, doi: 10.3390/ani12060692

[5] L. Brickson, L. Zhang, F. Vollrath, I. Douglas-Hamilton, and A. J. Titus, "Elephants and algorithms: a review of the current and future role of AI in elephant monitoring," J. R. Soc. Interface, vol. 20, no. 208, 2023, doi: 10.1098/rsif.2023.0367.

[6] J. V. Congdon, M. Hosseini, E. F. Gading, M. Masousi, M. Franke, and S. E. Macdonald, "The Future of Artificial Intelligence in Monitoring Animal Identification, Health, and Behaviour," Animals, vol. 12, no. 13, 2022, doi: 10.3390/ani12131711.

[7] E. Mohammed, T. A. Alsadi, and N. K. El Abbadi, "Scrutiny of Methods for Image Detection and Recognition of Different Species of Animals," Int. J. Recent Technol. Eng., vol. 8, no. 3S3, pp. 151–160, 2019, doi: 10.35940/ijrte.c1046.1183s319.

[8] A. Kirillov et al., "Segment Anything," 2023, [Online]. Available: http://arxiv.org/abs/2304.02643

[9] M. Hu, Y. Li, and X. Yang, "SkinSAM: Empowering Skin Cancer Segmentation with Segment Anything Model," 2023, [Online]. Available: http://arxiv.org/abs/2304.13973

[10] M. Ahmadi et al., "Comparative Analysis of Segment Anything Model and U-Net for Breast Tumor Detection in Ultrasound and Mammography Images," pp. 1–34, 2023, [Online]. Available: http://arxiv.org/abs/2306.12510

[11] T. Chen et al., "SAM-Adapter: Adapting Segment Anything in Underperformed Scenes," pp. 3359–3367, 2023, doi: 10.1109/iccvw60793.2023.00361.

[12] Z. Qiu, Y. Hu, H. Li, and J. Liu, "Learnable Ophthalmology SAM," 2023, [Online]. Available: http://arxiv.org/abs/2304.13425

[13] Y. Li, B. Jing, Z. Li, J. Wang, and Y. Zhang, "nnSAM: Plug-and-play Segment Anything Model Improves nnUNet Performance," vol. 1, 2023, [Online]. Available: http://arxiv.org/abs/2309.16967

[14] S. He et al., "Computer-Vision Benchmark Segment-Anything Model (SAM) in Medical Images: Accuracy in 12 Datasets," no. April, pp. 1–8, 2023, [Online]. Available: http://arxiv.org/abs/2304.09324

[15] S. Roy et al., "SAM.MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model," pp. 1–4, 2023, [Online]. Available: http://arxiv.org/abs/2304.05396

[16] C. Mattjie et al., "Zero-shot performance of the Segment Anything Model (SAM) in 2D medical imaging: A comprehensive evaluation and practical guidelines," 2023, [Online]. Available: http://arxiv.org/abs/2305.00109

[17] S. Pandey, K.-F. Chen, and E. B. Dam, "Comprehensive Multimodal Segmentation in Medical Imaging: Combining YOLOv8 with SAM and HQ-SAM Models," pp. 2584–2590, 2023, doi: 10.1109/iccvw60793.2023.00273.

[18] C. Hu, T. Xia, S. Ju, and X. Li, "When SAM Meets Medical Images: An Investigation of Segment Anything Model (SAM) on Multi-phase Liver Tumor Segmentation," pp. 1–5, 2023, [Online]. Available: http://arxiv.org/abs/2304.08506

[19] M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, and Y. Zhang, "Segment anything model for medical image analysis: An experimental study," Med. Image Anal., vol. 89, 2023, doi: 10.1016/j.media.2023.102918.

[20] A. Ranem, N. Babendererde, M. Fuchs, and A. Mukhopadhyay, "Exploring SAM Ablations for Enhancing Medical Segmentation in Radiology and Pathology," pp. 1–13, 2023, [Online]. Available: http://arxiv.org/abs/2310.00504

[21] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment Anything in Medical Images," pp. 1–9, 2023, [Online]. Available: http://arxiv.org/abs/2304.12306

[22] D. Williams, F. Macfarlane, and A. Britten, "Leaf Only SAM: A Segment Anything Pipeline for Zero-Shot Automated Leaf Segmentation," 2023, [Online]. Available: http://arxiv.org/abs/2305.09418

[23] M. Ahmadi, A. G. Lonbar, A. Sharifi, A. T. Beris, M. Nouri, and A. S. Javidi, "Application of Segment Anything Model for Civil Infrastructure Defect Assessment," pp. 1–31, 2023, [Online]. Available: http://arxiv.org/abs/2304.12600

[24] J. Yang, M. Gao, Z. Li, S. Gao, F. Wang, and F. Zheng, "Track Anything: Segment Anything Meets Videos," 2023, [Online]. Available: http://arxiv.org/abs/2304.11968

[25] S. Mo and Y. Tian, "AV-SAM: Segment Anything Model Meets Audio-Visual Localization and Segmentation," pp. 2–5, 2023, [Online]. Available: http://arxiv.org/abs/2305.01836

[26] S. Ren et al., "Segment anything, from space?," vol. 1, 2023, [Online]. Available: http://arxiv.org/abs/2304.13000