

Cichlid Computer Vision Project – Weekly Progress

Week ending Friday, December 6, 2024

Summary (please see all detailed attachments below):

Kailey Quesada

1.1 What progress did you make in the last week?

(1) For the Lizard project, I ran B-SOiD on the new dataset. I ran into issues running the new dataset from

a csv, and it randomly worked after the third try.

(2) For the Lizard project, I attempted to run VAME on the new dataset. I ran into issues, most likely

due to the fact that there were a different number of keypoints in the new dataset. Specifically, the number

of keypoints is not the same as the VAME demo. Because of this, I was unable to run VAME on the new

dataset in the given timeframe. Additional troubleshooting must be done.

(3) For the Lizard project, I worked on writing the 16-page lizard report.

(4) For the Lizard project, I created two videos to show the results. One video had just the lizard dataset,

and the other video had both. The lizard dataset video is on YouTube.

(5) For the Lizard and Cichlid projects, I created slides for the final presentation. For the Lizard project, I

created slides and updated the read me and code on the GitHub.

(6) For meetings, I attended the Cichlid CV team meeting, a huddle for the Stroud report, a working meeting

to prepare for the Cichlid CV team presentation, and the end of semester showcase presentation.

1.2 What are you planning on working on next?

- (1) I need to help finish writing the team documentation for the ease of on-boarding handoff.
- (2) I need to finish any other documentation or notes that I have that might be helpful next semester.
- (3) I need to save any important information from PACE before the end of the semester as well.

1.3 Is anything blocking you from getting work done?

None.

Charles Clark

What progress did you make in the last week?

- Fixed bug in video extraction Python script to facilitate the extraction of infrared video data from the .bag files.
- Extracted all video data from all .bag files in the Lindenthal Camera Traps dataset.
- Wrote a bash script (with help from ChatGPT) to run detection on all the extracted video data using my pre-trained YOLOv5s model.
- Ran this custom bash script to generate YOLOv5s predictions on all the extracted videos.
- Continued literature review.
- Attended weekly Cichlid CV team meeting on Tuesday.
- Attended paper writing seminar on Wednesday.
- Attended a follow-up meeting with the Cichlid CV team to work on our final presentation Thursday morning.
- Attended weekly HAAG admin meeting Thursday afternoon.
- Presented with the rest of Cichlid CV at the HAAG Connection event Thursday evening.
- Attended weekly publication seminar Friday afternoon.

What are you planning on working on next?

- Continue working on BioBoost (run detections through SORT to generate annotations).
- Attend weekly HAAG admin meeting.

Is anything blocking you from getting work done?

- Nothing.

Thuan Nguyen

What progress did you make in the last week?

- Presented a slide at the all-HAAG meeting to highlight project updates and contributions.
- Added a section on re-ID training with image crops to the DeepLabCut tutorial. https://github.com/Human-Augment-Analytics/CichlidBowerTracking/blob/master/training_dlc_on_cichlid_data_on_ICE_node.md
- Created a "Meeting Summary" page for the Cichlid CV team project site to log weekly updates and maintain consistent documentation. <https://sites.gatech.edu/cichlid-computer-vision-project/weekly-meetings-demo/>
- Filled out Vikas's onboarding template, focusing on the DeepLabCut experiments and ensuring clarity in project contributions.

What are you planning on working on next?

- With the semester coming to an end, I will find out how my model files and Python scripts currently stored on ICE clusters can be preserved for experiments in the next semester. And I'll continue to find out what the research or experimentation needs are within the Cichlid CV team in order to know what to read or code up for those experiments.

Is anything blocking you from getting work done?

- Not at the moment.

Week 16 Document Submission

Kailey Quesada (Cichlid CV & Lizard CV Team)

December 6, 2024

1 Weekly Project Updates

1.1 What progress did you make in the last week?

- (1) For the Lizard project, I ran B-SOiD on the new dataset. I ran into issues running the new dataset from a csv, and it randomly worked after the third try.
- (2) For the Lizard project, I attempted to run VAME on the new dataset. I ran into issues, most likely due to the fact that there were a different number of keypoints in the new dataset. Specifically, the number of keypoints is not the same as the VAME demo. Because of this, I was unable to run VAME on the new dataset in the given timeframe. Additional troubleshooting must be done.
- (3) For the Lizard project, I worked on writing the 16-page lizard report.
- (4) For the Lizard project, I created two videos to show the results. One video had just the lizard dataset, and the other video had both. The lizard dataset video is on YouTube.
- (5) For the Lizard and Cichlid projects, I created slides for the final presentation. For the Lizard project, I created slides and updated the read me and code on the GitHub.
- (6) For meetings, I attended the Cichlid CV team meeting, a huddle for the Stroud report, a working meeting to prepare for the Cichlid CV team presentation, and the end of semester showcase presentation.

1.2 What are you planning on working on next?

- (1) I need to help finish writing the team documentation for the ease of on-boarding handoff.
- (2) I need to finish any other documentation or notes that I have that might be helpful next semester.
- (3) I need to save any important information from PACE before the end of the semester as well.

1.3 Is anything blocking you from getting work done?

None.

2 Document Submission

2.1 Abstract

Sun, J., et al. “Task Programming: Learning Data Efficient Behavior Representations.” CVPR 2021. https://openaccess.thecvf.com/content/CVPR2021/html/Sun_Task_Programming_Learning_Data_Efficient_Behavior_Representations_CVPR_2021_paper.html.

Specialized knowledge is often needed to accurately label training data for behavior analysis, which can be time-consuming for experts. To address this, the authors propose TREBA, a method that combines self-supervised learning with expert-defined programs to efficiently learn representations of movement data. This approach allows experts to reduce the amount of manual labeling required by creating a few programmed tasks, leading to significant reductions in annotation effort while maintaining classification accuracy. The text discusses how the researchers evaluated the efficiency of their representation for classifying behaviors

in different datasets, such as MARS, CRIM13, and flies. They compared their method, TREBA, against traditional keypoints and expert-designed features to see how much training data was needed to achieve similar performance. The results showed that TREBA significantly improved data efficiency, allowing for accurate classification with much less training data, which means it can learn effectively from fewer examples.

2.2 Scripts and Code Blocks

Code from previous weeks can be found here: <https://github.com/Human-Augment-Analytics/Lizard-Pose-Estimation-and-Evaluation/tree/main>. Below is the run script for VAME that I was trying to get to work this week for the new dataset. This is based directly off of the documentation with changes to make it work, since the documentation is deprecated.

```
1 ...
2
3 # Go Through Working Directory and Add All Files
4 for file_path in working_directory_path.iterdir():
5     if file_path.is_file():
6         if file_path.suffix == '.mp4':
7             videos.append("C:/Users/Username/Path/To/Files/VAME/data/" + file_path.name)
8         elif file_path.suffix == '.csv':
9             poses_estimations.append("C:/Users/Username/Path/To/Files/VAME/data/" +
10                                     file_path.name)
11 # Initialize Project
12 config = vame.init_new_project(
13     project = project,
14     videos = videos,
15     poses_estimations=poses_estimations,
16     working_directory=working_directory_path,
17     videotype='.mp4'
18 )
19
20 # Run an Existing Project
21 # config = 'C:/Users/Username/Path/To/Files/VAME/data/lizard project-Day-2024/config.yaml'
22
23 # Transform DLC CSVs to Numpy Arrays
24 vame.egocentric_alignment(config, pose_ref_index=[0,5])
25
26 # Create Training Dataset for Videos with 6 Keypoints
27 vame.create_trainset(config, pose_ref_index=[0,5])
28
29 # Train the Model
30 vame.train_model(config)
31
32 # Evaluate the Model
33 vame.evaluate_model(config)
34
35 # Pose Segmentation
36 vame.pose_segmentation(config)
37
38 # Create Motif Videos
39 vame.motif_videos(config, videoType='.mp4', parametrization='hmm')
40
41 # Run Community Detection
42 vame.community(config, parametrization='hmm', cut_tree=None, cohort=True)
43
44 # UMAP Visualization
45 fig = vame.visualization(config, label=None, parametrization='hmm')
46
47 # Generative Reconstruction Decoder
48 vame.generative_model(config, mode="centers", parametrization='kmeans')
49
50 # Output Video
51 vame.gif(config, pose_ref_index=[0,4], subtract_background=False, start=None,
52          parametrization='hmm',
```

```
length=100, max_lag=30, label='motif', file_format='.mp4', crop_size=(200,200))
```

Listing 1: Run VAME.

2.3 Documentation

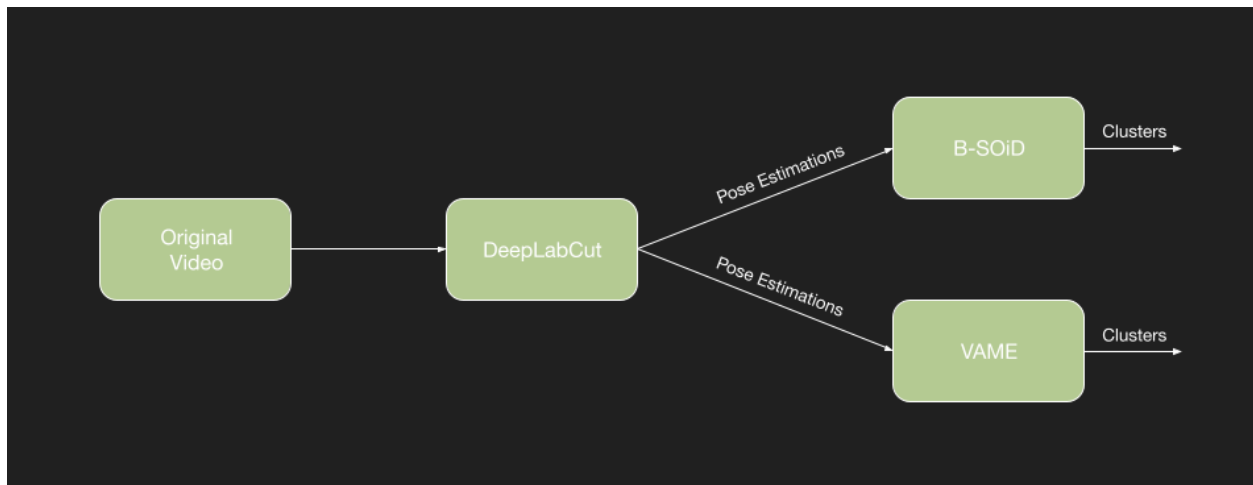
This week, I participated in helping to write a 16-page report for Dr. Stroud. This includes a formal methods section and lots of documentation and results. Because it includes a private dataset, I will not be sharing the link here. Contact Bree to verify.

2.4 Script Validation (Optional)

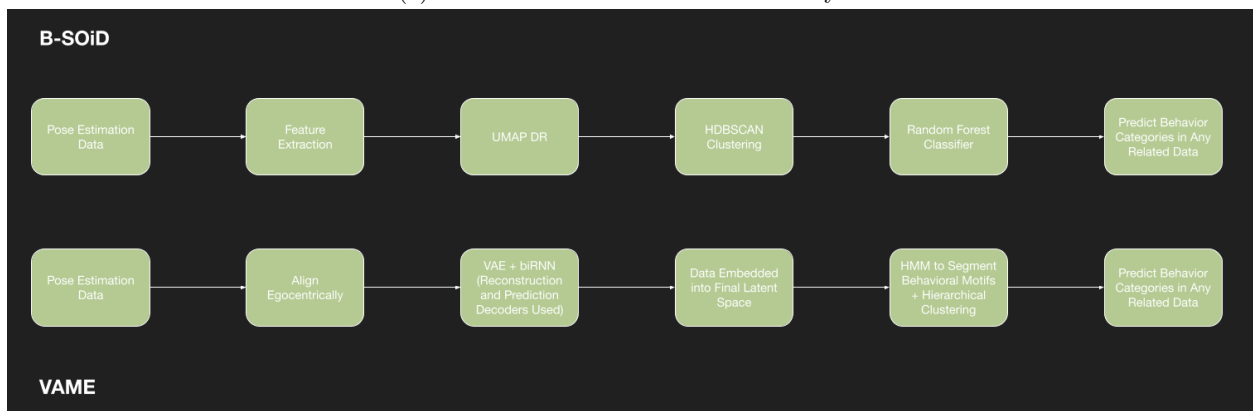
No additional script validation information needed this week.

2.5 Results Visualization

I visualized the workflows of B-SOiD and VAME, as well as the overall Lizard Behavioral Analysis workflow. See them in the figures below.



(a) Overview of Lizard Behavioral Analysis



(b) B-SOiD and VAME Processing Steps

Figure 1: Diagrams for Lizard Project

2.6 Proof of Work

I will need to figure out why specifically VAME would not work on the other dataset. Additionally, I will need to figure out why B-SOiD was randomly failing on the new dataset as well. For VAME, I think it had to do with the number of keypoints, and for B-SOiD I think it had to do with running the GUI from a csv instead of an h5. I will need to figure out workarounds for these problems if I continue on this project.

2.7 Next Week's Proposal

As stated on the first page, this week's goals are the following:

- (1) I need to help finish writing the team documentation for the ease of on-boarding handoff.
- (2) I need to finish any other documentation or notes that I have that might be helpful next semester.
- (3) I need to save any important information from PACE before the end of the semester as well.

Charles R. Clark
CS 6999 – HAAG: Cichlid CV
Fall 2024
December 6, 2024

Week 16 Report

1. *Time-log Response:*

What progress did you make in the last week?

- Fixed bug in video extraction Python script to facilitate the extraction of infrared video data from the .bag files.
- Extracted all video data from all .bag files in the Lindenthal Camera Traps dataset.
- Wrote a bash script (with help from ChatGPT) to run detection on all the extracted video data using my pre-trained YOLOv5s model.
- Ran this custom bash script to generate YOLOv5s predictions on all the extracted videos.
- Continued literature review.
- Attended weekly Cichlid CV team meeting on Tuesday.
- Attended paper writing seminar on Wednesday.
- Attended a follow-up meeting with the Cichlid CV team to work on our final presentation Thursday morning.
- Attended weekly HAAG admin meeting Thursday afternoon.
- Presented with the rest of Cichlid CV at the HAAG Connection event Thursday evening.
- Attended weekly publication seminar Friday afternoon.

What are you planning on working on next?

- Continue working on BioBoost (run detections through SORT to generate annotations).
- Attend weekly HAAG admin meeting.

Is anything blocking you from getting work done?

- Nothing.

2. *Abstracts:*

“An Individual Identity-Driven Framework for Animal Re-Identification”, Wu et al. (2024; arXiv preprint)

- *Abstract:* “Reliable re-identification of individuals within large wildlife populations is crucial for biological studies, ecological research, and wildlife conservation. Classic computer vision techniques offer a promising direction for Animal Re-identification (Animal ReID), but their backbones’ close-set nature limits their applicability and generalizability. Despite the demonstrated effectiveness of visionlanguage models like CLIP in re-identifying persons and vehicles, their application to Animal ReID remains limited due to unique challenges, such as the various visual representations of animals, including variations in poses and forms. To address these limitations, we leverage CLIP’s cross-modal capabilities to introduce a two-stage framework, the Individual Animal IDentity-Driven (IndivAID) framework, specifically designed for Animal ReID. In the

first stage, IndivAID trains a text description generator by extracting individual semantic information from each image, generating both image-specific and individual-specific textual descriptions that fully capture the diverse visual concepts of each individual across animal images. In the second stage, IndivAID refines its learning of visual concepts by dynamically incorporating individual-specific textual descriptions with an integrated attention module to further highlight discriminative features of individuals for Animal ReID. Evaluation against state-of-the-art methods across eight benchmark datasets and a real-world Stoat dataset demonstrates IndivAID's effectiveness and applicability. Code is available at <https://github.com/ywu840/IndivAID>."

- *AI Summary by ChatGPT (4o)*: "This paper introduces IndivAID, a novel framework leveraging the CLIP vision-language model for animal re-identification (ReID). IndivAID addresses challenges unique to Animal ReID, such as pose variation and the lack of descriptive text labels, by generating individual-specific textual descriptions for images. The two-stage training process combines text generation with an attention mechanism to refine visual features for improved ReID performance. Experiments across eight benchmark datasets and a real-world stoat dataset demonstrate significant improvements over state-of-the-art methods."
 - *Key Contributions*:
 - "Two-Stage Framework for Animal ReID:"
 - "Stage 1: IndivAID trains a Text Description Generator to produce textual descriptions that combine static prompts with learnable tokens. These descriptions align images and texts in a shared embedding space using CLIP's encoders."
 - "Stage 2: The framework refines individual-specific textual descriptions through an attention module, optimizing CLIP's image encoder for ReID tasks."
 - "Novel Use of CLIP for ReID: Unlike conventional methods, IndivAID integrates visual features into textual descriptions, addressing the lack of descriptive labels in ReID tasks and improving alignment between image and text modalities."
 - "Enhanced ReID Performance: Evaluation on benchmark datasets and a real-world stoat dataset shows that IndivAID outperforms baselines (CLIP-ZS, CLIP-FT, and CLIP-ReID) in both mAP and Top-1 accuracy, achieving superior generalization across diverse animal species."
 - *Contributions to Knowledge*:
 - "Vision-Language Models in Animal ReID: The paper pioneers the application of CLIP to Animal ReID, introducing a scalable and adaptable approach that overcomes limitations of prior CNN or ViT-based methods."
 - "Individual-Specific Textual Descriptions: The innovative use of learnable tokens to generate personalized text prompts represents a significant advance in aligning image and text embeddings for ReID tasks."
 - "Improved Robustness to Variations: IndivAID handles challenges like pose variations and environmental changes, enhancing its applicability to real-world scenarios."
 - *Future Research Directions*:

- “Scaling to Larger and Multi-Modal Datasets: Expanding IndivAID to incorporate additional modalities (e.g., depth or thermal imaging) and larger datasets could further enhance its generalizability.”
 - “Optimizing Attention Mechanisms: Exploring more efficient attention mechanisms tailored for cross-modal alignment could improve computational efficiency and performance.”
 - “Generalization Across Domains: While focused on animals, IndivAID could be adapted for other domains requiring fine-grained ReID, such as human surveillance or vehicle tracking.”
 - “Semi-Supervised and Few-Shot Learning: Incorporating semi-supervised or few-shot learning strategies might reduce the reliance on large labeled datasets, making IndivAID more practical in data-scarce scenarios.”
- *Link:* <http://arxiv.org/abs/2410.22927>.

3. *Scripts & Code Blocks:*

detect_all.sh

- A bash script which runs a pre-trained YOLOv5s model’s detect.py script on all the videos in a source directory, saving the detections to a target directory.
 - Generated with help from ChatGPT.
- *Code block:*

```

1  #!/bin/bash
2
3  # Directory containing the .mp4 files
4  INPUT_DIR="/home/charlieclark/lindenthal-videos"
5  OUTPUT_DIR="/home/charlieclark/lindenthal-yolo-preds"
6  MODEL_PATH="runs/train/exp/weights/best.pt"
7
8  # Ensure output directory exists
9  mkdir -p "$OUTPUT_DIR"
10
11 # Iterate over each .mp4 file in the input directory
12 for video in "$INPUT_DIR"/*.mp4; do
13     echo "Processing video: $video"
14
15     # Extract the base name of the video file (without extension)
16     base_name=$(basename "$video" .mp4)
17
18     # Set the output directory for the current video
19     video_output_dir="$OUTPUT_DIR/$base_name"
20     mkdir -p "$video_output_dir"
21
22     # Run YOLOv5 detection on the video file (frame-by-frame)
23     # --source is the video, --project is the output directory, --name is the video name
24     python3 detect.py --weights "$MODEL_PATH" --source "$video" --project "$video_output_dir" --name "detections"
25
26     echo "Finished processing: $video"
27 done
28
29 echo "All videos processed."

```

- *Status:* tested and functional.

- *Data*: Requires a source directory with video data, a target directory where detections will be stored, and access to YOLOv5 (preferably a pre-trained model); should be run from the root of the yolov5 Github repo (clone if necessary).

4. *Documentation (non-PhD centered stuff only)*:

- Fixed bug in video extraction Python script to facilitate the extraction of infrared video data from the .bag files.
- Extracted all video data from all .bag files in the Lindenthal Camera Traps dataset.
- Wrote a bash script (with help from ChatGPT) to run detection on all the extracted video data using my pre-trained YOLOv5s model.
- Ran this custom bash script to generate YOLOv5s predictions on all the extracted videos.
- Continued literature review.
- Attended weekly Cichlid CV team meeting on Tuesday.
- Attended paper writing seminar on Wednesday.
- Attended a follow-up meeting with the Cichlid CV team to work on our final presentation Thursday morning.
- Attended weekly HAAG admin meeting Thursday afternoon.
- Presented with the rest of Cichlid CV at the HAAG Connection event Thursday evening.
- Attended weekly publication seminar Friday afternoon.

5. *Script Validation (optional)*: Code written this week is tested, partially functional.

6. *Results Visualization*: Please find here a link to an infrared video that was extracted from a .bag file using the aforementioned Python script: [20201218165020_Infrared_1.mp4](#)

7. *Proof of Work*: For proof of work this week, please find below output from the detect_all.sh script, as well as the output of the “tree” linux command when run from the target directory where all detections were stored.

```
video 1/1 (234/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 7.2ms
video 1/1 (235/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 6.8ms
video 1/1 (236/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 7.4ms
video 1/1 (237/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 6.7ms
video 1/1 (238/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 6.9ms
video 1/1 (239/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 7.0ms
video 1/1 (240/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 7.0ms
video 1/1 (241/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 6.9ms
video 1/1 (242/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 (no detections
), 6.8ms
video 1/1 (243/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 7.1ms
video 1/1 (244/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 7.0ms
video 1/1 (245/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 6.8ms
video 1/1 (246/246) /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4: 384x640 1 deer, 6.5ms
Speed: 0.4ms pre-process, 7.3ms inference, 1.3ms NMS per image at shape (1, 3, 640, 640)
Results saved to /home/charlieclark/lindenthal-yolo-preds/20220317002738_Infrared_2/detections
Finished processing: /home/charlieclark/lindenthal-videos/20220317002738_Infrared_2.mp4
All videos processed.
(yolov5) charlieclark@MSI: ~/yolov5$ █
```

```
├── 20220316231506_Infrared_1
│   └── detections
│       └── 20220316231506_Infrared_1.mp4
├── 20220316231506_Infrared_2
│   └── detections
│       └── 20220316231506_Infrared_2.mp4
├── 20220316232819_Infrared_1
│   └── detections
│       └── 20220316232819_Infrared_1.mp4
├── 20220316232819_Infrared_2
│   └── detections
│       └── 20220316232819_Infrared_2.mp4
├── 20220316233613_Infrared_1
│   └── detections
│       └── 20220316233613_Infrared_1.mp4
├── 20220316233613_Infrared_2
│   └── detections
│       └── 20220316233613_Infrared_2.mp4
├── 20220317002738_Infrared_1
│   └── detections
│       └── 20220317002738_Infrared_1.mp4
└── 20220317002738_Infrared_2
    └── detections
        └── 20220317002738_Infrared_2.mp4

2710 directories, 1355 files
(yolov5) charlieclark@MSI:~$ █
```

8. *Next Week's Proposal (non-PhD centered stuff only):*

- Continue working on BioBoost (run detections through SORT to generate annotations).
- Attend weekly HAAG admin meeting.

9. *Questions:* None.

Week 16 Report

Thuan Nguyen – Cichlid Computer Vision project

Friday, December 6, 2024

Summary

What progress did you make in the last week?

- Presented a slide at the all-HAAG meeting to highlight project updates and contributions.
- Added a section on re-ID training with image crops to the DeepLabCut tutorial. https://github.com/Human-Augment-Analytics/CichlidBowerTracking/blob/master/training_dlc_on_cichlid_data_on_ICE_node.md
- Created a "Meeting Summary" page for the Cichlid CV team project site to log weekly updates and maintain consistent documentation. <https://sites.gatech.edu/cichlid-computer-vision-project/weekly-meetings-demo/>
- Filled out Vikas's onboarding template, focusing on the DeepLabCut experiments and ensuring clarity in project contributions.

What are you planning on working on next?

- With the semester coming to an end, I will find out how my model files and Python scripts currently stored on ICE clusters can be preserved for experiments in the next semester. And I'll continue to find out what the research or experimentation needs are within the Cichlid CV team in order to know what to read or code up for those experiments.

Is anything blocking you from getting work done?

- Not at the moment.

Abstract

Adaptive High-Frequency Transformer for Diverse Wildlife Re-Identification

Chenyue Li, Shuoyi Chen, Mang Ye

<https://arxiv.org/pdf/2410.06977>

+ Introduces the Adaptive High-Frequency Transformer for multi-species wildlife Re-ID, leveraging high-frequency information to capture unique features like fur textures and contours.

+ Implements frequency-domain mixed augmentation and object-aware dynamic selection to enhance discriminative feature learning and reduce noise from natural environments.

- + Demonstrates strong performance in domain generalization, accurately identifying both known and unknown species across diverse datasets.
- + Outperforms state-of-the-art Re-ID methods on terrestrial, aquatic, and aerial species datasets, ensuring adaptability and accuracy in various scenarios.

Work done this week – further details

- + Presented a slide at the all-HAAG meeting to highlight project updates and contributions.
- + Added a section on re-ID training with image crops to the DeepLabCut tutorial.
https://github.com/Human-Augment-Analytics/CichlidBowerTracking/blob/master/training_dlc_on_cichlid_data_on_ICE_node.md
- + Created a "Meeting Summary" page for the Cichlid CV team project site to log weekly updates and maintain consistent documentation. <https://sites.gatech.edu/cichlid-computer-vision-project/weekly-meetings-demo/>
- + Filled out Vikas's onboarding template, focusing on the DeepLabCut experiments and ensuring clarity in project contributions.

Scripts - Documentation - Script Validation - Results Visualization - Proof of Work

For further details, please refer to the details above.

Next week's proposal

- With the semester coming to an end, I will find out how my model files and Python scripts currently stored on ICE clusters can be preserved for experiments in the next semester. And I'll continue to find out what the research or experimentation needs are within the Cichlid CV team in order to know what to read or code up for those experiments.

