[redacted]
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:*

```
#!/bin/bash

# Directory containing the .mp4 files
INPUT_DIR="/home/[redacted]/lindenthal-videos"
OUTPUT_DIR="/home/[redacted]/lindenthal-yolo-preds"
MODEL_PATH="runs/train/exp/weights/best.pt"

# Ensure output directory exists
mkdir -p "$OUTPUT_DIR"

# Iterate over each .mp4 file in the input directory
for video in "$INPUT_DIR"/*.mp4; do
    echo "Processing video: $video"

    # Extract the base name of the video file (without extension)
    base_name=$(basename "$video" .mp4)

    # Set the output directory for the current video
    video_output_dir="$OUTPUT_DIR/$base_name"
    mkdir -p "$video_output_dir"

    # Run YOLOv5 detection on the video file (frame-by-frame)
    # --source is the video, --project is the output directory, --name is the video name
    python3 detect.py --weights "$MODEL_PATH" --source "$video" --project "$video_output_dir" --name "detections"

    echo "Finished processing: $video"
done

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).

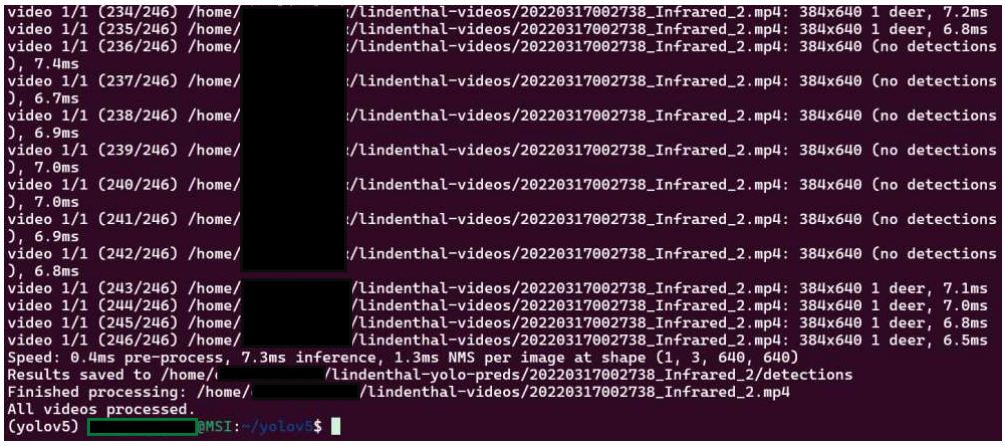
4. *Documentation (non-PhD centered stuff only):*

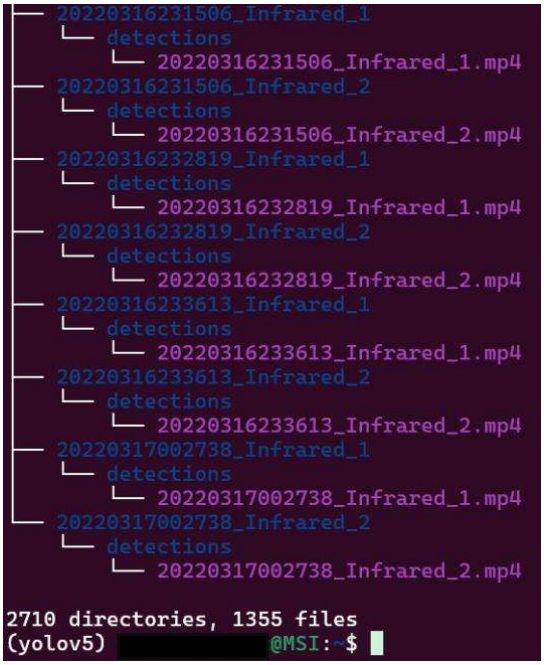
- Fixed bug in video extraction Python script to facilitate the extraction of infrared video data from the .bag files.
- Extracted all video data from all .bag files in the Lindenthal Camera Traps dataset.
- Wrote a bash script (with help from ChatGPT) to run detection on all the extracted video data 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.





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.