

 README.md

SparrowVis Code

This readme provide details for what each part of the pipeline does, to allow you to adapt and customize each part of it!

Quick reminder, here is the pipeline and the general workflow:

1. Open source software called Deep Meerkat (Weinstein, 2018) to identify movement frames
2. Process output and group frames into “events”
3. Extract short video clips of the events
4. Summarize processed videos into a biological meaningful “visit rate”

Figures

Running the Pipeline

1. Running Deep Meerkat

Deep Meerkat is an open source software. Here is the original github repository. To run deep meerkat, you have to run

```
python ../DeepMeerkat/DeepMeerkat/Meerkat.py --input [Video] --path_to_model ../DeepMeerkat/DeepMeerkat/model/
--output [Output]
```

With **[Video]** being the path to the input video and **[Output]** being the path to the output destination. Deep Meerkat outputs the .jpg images of the movement frames, **annotation.csv** that records all the movement frames and info, and lastly **parameters.csv** which saves important information regarding the video.

Note that Deep Meerkat is quite slow, with approximately a 1:1 run time. This means that a 1 hour video will likely take around an hour to run.

2. Defining Events

Next, the R script ProcessFrameInfo.R reads the **annotation.csv** output from Deep Meerkat and groups it into events. An event is defined as any cluster of movement frames that is at least 2 frames long and less than 40 frames apart. To run the script, use:

```
Rscript ProcessFrameInfo.R [VidName]
```

Where **[VidName]** is the name of the video without the .mp4 extension. The script will output a new dataframe called **Short.csv** within each of the output folders.

3. Extracting Clips

Next, the python script ExtractFrames.py takes the events and timestamps from **Short.csv** from before, and writes separate 7 second video files for each event. To run the script, run:

```
python ExtractFrames.py [VidName]
```

Where **[VidName]** is the name of the video without the .mp4 extension as above. Note that this script pulls the full video file from the RawVideos/ directory, so make sure the original video file is there.

4. Extracting Visit Rates

Lastly, once all events are defined, the script OrganizeOutput.R takes all the files located under **Output/** and tally up the number of visit events detected for each video. To run it, type:

```
Rscript OrganizeOutput.R
```

This will create a csv file called "AutoVisitRate.csv" which stores counts the visits detected for each video and the effective observation time (see Nakagawa et al., 2007)

Running on the HPC cluster

If you are more experienced with computational techniques and have access to the higher performance computing cluster in Imperial (or similar), I provided additional scripts that I used to run the pipeline on it. This would save more time because you can basically run the pipeline in parallel for multiple videos.

To run Deep Meerkat, I used the script `RunMeerkat_bash.sh`, which checks if the script runs successfully and moves the video file to the output only if it is successful. This was due to a bug that prevents the code from running on the HPC cluster, and as of Nov 2021 (when I am writing this), this issue has still not been solved.

To run the script, I used the `-J` argument to allow multiple videos to be ran at once. Sample for running 100 videos at once:

```
qsub RunMeerkat_bash.sh -J 1-100
```

If you are trying to run the pipeline on a computing cluster and is having trouble, do not hesitate to contact me!! Good Luck!

References

- Nakagawa, S., Gillespie, D.O.S., Hatchwell, B.J., Burke, T., 2007. Predictable males and unpredictable females: sex difference in repeatability of parental care in a wild bird population. *J. Evol. Biol.* 20, 1674–1681.
- Weinstein, B.G., 2018. Scene-specific convolutional neural networks for video-based biodiversity detection. *Methods Ecol. Evol.* 9, 1435–1441.

Author and Affiliations

Alex Chan Hoi Hang
the.alex.chan2@gmail.com/ hhc4317@ic.ac.uk
MRes Computational Methods in Ecology and Evolution
Department of Life Sciences
Imperial College Silwood Park
UK. SL5 7PY