**Common bugs and frequently asked questions (FAQ)**

1. What are the things I should keep in mind when recording videos using a drone?

Drone flight design depends on the environment, the study objective and the sensitivity of the species to noise form the drone. In cases when individuals are scattered across a large area, it is ideal to fly the drone at relatively high heights to record most animals. However, if the animals are very close to each other (like in case of dolphins), it is recommended to fly low if individual tracks are required. In any case, the drone height should be kept fixed, and drone movement to a minimum throughout the duration of the flight; camera angles can be changed slightly if flying with a nadir view is not possible. Note that it is crucial to have start times for both the drone flight and when the video was recorded at the highest resolution possible to ensure a precise synchronisation of flight and video parameters. Details on the exact flight start time is generally available in the fight log name while video timestamps are recorded in the subtitle file of the video (which may need to be extracted using the *ffmpeg* library if the subtitles are embedded in the video file). However, it is a good practice to manually record both values using a dedicated watch in case one or both are missing in different drone models.

1. I have used a different drone than the one this package was written for (DJI Mini 3 Pro). How do I modify the modules to analyse data from my own drone?

While it is possible to change the scripts for each module to better align with the data formatting of your own drone, it would be ideal to modify the format of the data recorded from your drone to match the exemplar dataset (e.g., the flight log file) instead. The process of detecting and tracking animals in the video, either through our machine learning models or manually is indifferent to drone models as long as some drone-specific information is available. Video parameters, specifically its resolution and fps information need to be added in the script. The framework for converting pixel coordinates to geo-coordinates is fixed as well. Column names in the datasets of different drones and their formatting may vary, but they can be changed to match the column names of the example files for a smoother experience. Overall, the data required is

* 1. The field of view of the drone camera when recording a video, which can be estimated using simple experiments described in the manuscript (sec 2.1).
  2. The flight parameters at a microsecond level recorded in the flight log. Specifically, data is needed on drone fly time, speed, GPS coordinates, height and orientation, and camera gimbal pitch. These may be recorded with different names in flight logs of different models.
  3. The flight start time (FST; time when the drone motors were switched on) needs to be recorded at the highest precision possible. This is because often drone flight logs may show different timestamps that local time. Drone fly time can be added to FST to get precise timestamps for drone flight.
  4. Timestamps on the video need to be extracted so that detections in the video can be synchronised with flight parameters to geo-reference their positions. If timestamps are not available, video time can be added to the start time of video recording to obtain timestamps. Start time must be noted manually when recording starts in case it is not recorded in the subtitles.
  5. The difference in altitude between where the drone launched from and where the animals are, must be recorded. Our geo-referencing algorithm uses information on the height of the drone form the surface (when the animals are) to estimate their tracks. If the drone if launched from a different height (say a hilltop), and the animals being tracked are on a different plane (e.g., sea surface), the height of the hilltop is essential to add to the drone’s flight height.

1. How do I obtain flight log and SRT files?

Flight log files are stored in the drone controller after every flight and can be copied using a USB cable. In case of DJI drones, these files need to be decoded and converted to their respective csv files using the PhantomHelp website. Similarly, SRT files are saved along with the video in the drone’s memory and have precise timestamps at the resolution of milliseconds. In some models, they may be embedded in the video file and thus need to be extracted. One way to do this is to use the ffmpeg software in command prompt, executed with the line *ffmpeg -i “path to video.mp4” “output path for srt”*.

1. Can I combine data from multiple drones for geo-referencing?

Our algorithm can be applied to any drone system and trajectories estimated from videos from one drone can be supplemented with trajectories from videos of another drone. Nevertheless, both videos will need to be analysed separately. Error in geo-referencing, and therefore the accuracy of tracks may vary across drone models, depending on their sensors and image processing.

1. How can I reduce geo-referencing error?

Planning flights to have minimal movement will ensure robust estimation of trajectories. Geo-referencing error can vary depending on the accuracy of onboard sensors and how precisely data from the log and the video are synchronised. In addition, since videos are processed differently by each drone’s software, they may have different levels of distortion, which can impact geo-referencing accuracy. Although we have not been able to test our algorithm using different drones, we believe these errors may be minimized in case of high-end models.

1. How do I choose the different parameters in each module for analysing videos?

The choice of parameters is dependent on the videos recorded. The main parameters for automated detection of animals include:

* 1. grabsize: Our detection model first identifies regions where animals may potentially be and then draws a bounding box (a square box) around each region to extract the image within the bounding box for model inference. The size of the bounding box should be such that it covers the full body on an individual. If individuals are of different sizes, the box should be such it is able to cover the largest of individuals. Bounding box size will need to be approximated from video frames and grabsize is the half of this value. E.g., if an animal is completely covered in a box of 150x150 pixels, grabsize will be 75. Note that when generating training data by clicking on animals, grabsize will need to be modified for the new frame size. E.g., if the frame is resized by a factor of 2.5, the new grabsize would be 30. All training data will need to be resized to a fixed dimension (150 x 150 in this case). The value of grabsize and the rescaling factor needs to be constant across scripts.
  2. Threshold range for grey-scaled images: This should be a range within which the pixels occupied by animals are distinctly visible. Rathore et al. (2023) provide an option in their MOTHE interface to dynamically identify these thresholds for different videos. These may also be kept uninformed with a min and max of 0 and 255 respectively.
  3. Threshold range for pixel area occupied by individuals: Every individual occupies some area within the bounding box. These represent the min and max possible value of this area. This range also needs to be approximated from video frames.

When geo-referencing, IDs that are not detected for at least 2 seconds are removed. This parameter can be changed depending on individual requirement. The pix\_to\_pos function also provides information on whether the drone or its camera were moving at each timeframe. This assumes that detections are filtered at a resolution of 1 second. If this resolution is decreased (to say of 5 seconds), the code to detect drone movement will need to be modified accordingly. A key argument of the geo-referencing function is the offset (‘off’) parameter. It captures the lag between the drone video and when its flight parameters are recorded. For example, a camera movement may be detected at time 11:12:12.500 in the video but it may be recorded later in the flight log (say at 11:12:14.500; after correcting for flight start time). The lag here is 2 seconds and the ‘off’ parameter is defined as the lag\*logging frequency, where logging frequency is how frequently the drone logs flight information. In this case, if the logging frequency is 200ms, then ‘off’ = 2\*5 = 10.

1. My drone constantly moved and changed altitude when recording videos. How will this impact automated tracking?

Although our geo-referencing algorithm can obtain trajectories from a moving drone, minimizing drone movement in any direction during recording is highly recommended. This is because fast movements can impact accurate matching of flight parameters from the drone’s log file and detection parameters from the video. Slow movement on the xy plane will reduce errors; the final csv containing information on the objects lat/lon also contains columns about the movement of the drone and its gimbal. Rows where such movement is high (recorded as a ‘1’) may also be eliminated before analysis.

In case of changes in drone height, tracking animals becomes more challenging. The automated detection algorithm assumes that the size of the animal/object being tracked does not change much throughout the video. As the drone drops down, different parts of the same animal may be identified as different individuals or the animal may not be identified at all. Similarly as the drone climbs up, it may become difficult to separate individuals from one another. Nevertheless, if data from such flights are to be extracted, we recommend using manual tracking instead.

1. Can I change the temporal resolution at which the detections are geo-referenced?

When using automated detections, all frames of a video are processed, and only 1 frame per second is geo-referenced. This is set keeping in mind the practical of fine-scale tracking vs the time it takes to geo-reference. Typically, a finer resolution is only needed to assess synchrony in movement or similar behaviours. For such analysis, we recommend using pixel-level trajectories instead of the geo-referenced ones.

1. There are too many/too few detections in my video.

Detection quality depends on the model which in turn depends on the training data and ultimately on the heterogeneity of the study area. It is never possible to build a model that works perfectly in all scenarios. Therefore, manual processing of detections, either by removing obvious false positives, or adding the ones missed out is often a necessity. If the model performs too poorly, there is always an option to manually track animals.

1. The detection model is running well but does not detect any animal.

This usually happens when the area and/or grayscale threshold range are outside desired values. You can debug the script to check for keypoint identities and locations on the frame and adjust the thresholds accordingly. When in doubt, use uninformed grayscale thresholds (0,255) and lower the min area threshold.

1. Error: cannot resize image of dimensions xx to yy.

Ensure that the values of grabsize and window\_rescale\_factor are the same as when generating training data of building the model. If the size of animals in the video to be analysed are larger than what the model was trained for, grabsize must be an exact multiple of the original grabsize such that it can be resized to the original size for model inference.