Evaluating Video Datasets for Underwater BioID

Oscar Vanhanen

February 2023

Contents

1 Introduction

1.1 Organisation

This project is based at Plant and Food Research (**PFR**) Lincoln Canterbury. Plant & Food Research is a New Zealand science company delivering research and development designed to grow competitive advantage for clients in the horticulture, wine, cropping, seafood and associated high-value food sectors worldwide.

The seafood production sector is an important sector within NZ providing valuable food items for export and use within NZ. The aquaculture research team at PFR researches the farming of aquatic organisms, such as fish, shellfish, and seaweed. This can include studying these organisms' genetics, breeding, nutrition, and health, as well as investigating ways to improve the efficiency and sustainability of aquaculture systems. They also work on developing new technologies and techniques for aquaculture and evaluating the environmental impacts of different farming methods. The goal of the research is to optimize the production and improve the quality of aquatic products and minimize the environmental impact of aquaculture.

A specific project within this team at PFR is researching how an aquaculture system interacts with a fish's biology. The area of research on aquaculture systems' interaction with a fish's biology is a complex area of active research. A critical component of this research is the ability to measure and re-identify individual fish across multiple points in time from benchtop images and underwater videos. This project aims to develop systems for tracking and understanding those observations as they occur.

1.1.1 Bio-ID

BioID is a biometric identification tool developed by PFR to identify Australasian Snapper. Biometric identification is the process of using a unique combination of biological measurements to identify an individual. BioIDs in fish work on many of the same principles as those in humans (e.g. fingerprinting and facial recognition). The underlying concept is that a unique combination of measurements from an individual could identify who that individual is. In Snapper the BioID is based on the unique spot patterns along their side. BioID gives more precision to measurements of Snapper (body length, weight etc) in noisy environments with many fish observations. Taking measurements in a noisy environment leads to relatively inaccurate results. Associating each measurement with an individual allows for the average measurements per individual and thus increases accuracy significantly.

BioID steps are as follows:

1. Images are captured as fish are transferred between tanks on the benchtop or underwater using a GoPro or similar camera. These images are only within the visible spectrum (wavelengths 380 – 750nm). The image quality is of varying resolution due to the type of camera used (generally 1080p).
2. Features of the fish, fins, body shape, and eyes, are identified by the machine learning models to allow for spot detection.

3. Data is normalised for easier comparison.

4. Various algorithms are used to compare pairs of spot patterns to match individuals.

The nature of the *Frame-Grabber* ML models is based on the Convolutional Neural Network architecture. The models are trained on many Snapper benchtop (clear high quality) images that are manually annotated by personnel knowledgeable about fish morphology. This is a time-consuming and costly procedure.

We are unable to elaborate on the specific nature of the BioID ML model used at Plant & Food Research due to its commercial and research nature.

It is unknown what the best lighting and water quality conditions are required for the best identification of individuals in underwater images. This is what we aim to determine here.

Terminology:

Observation – a single detected fish in a single frame

Tracking event – sequence of linked observations across consecutive frames

Population – a whole collection of one generation of fish bred at one point

Class – a number assigned to a specific section of a generation or collection of multiple generations.

Snapper and fish are used interchangeably onwards.

1.2 Goals

This project aims to develop pipelines to analyse the changes in fish populations tracking observations of fish as they become available, along with tools to query and understand the resulting populations of fish that are formed. The original goals are below ordered by priority:

1. Develop better tools for processing and understanding the underwater data that gets collected.
2. Applying matching model BioID to identify and link individual fish over different time boxes.
3. Building database systems for tracking individual observations and the populations they form.
4. Visualise and analyse the resulting population data.
5. Building web or desktop front-ends for querying and displaying population data.
6. Exploring whether these new datasets allow us to ask novel science questions.

Data science objectives:

1. Extract all the usable frames/observations we can get from numerous underwater videos
2. Time-box the observations
3. Use BioID tools to group observations of the same individual **within a single time box**
4. Use BioID tools to link groups of observations **across different time boxes**
5. Track the location and time of those irregular and infrequent groups of observations, and how they link together
6. Identifying overall population structures

Due to the data quality Objectives changed as this quality determination was uncovered. Moving forward we focused more on the data analysis of the data produced and how this could inform future data collection to improve the ability of BIOID to identify individuals.

1.3 Constraints

The data is the property of PFR and cannot be transferred outside of the organisation due to the commercial interests of PFR customers.

2 Data Overview

2.1 Descriptive statistics

The data repository consists of videos filmed by different cameras in the tanks/sea cages containing Snapper fish at various stages of development.

There were 48 videos determined by the aquaculture team to be suitable for BioID (identifying individual fish) processing and the list were compiled and organised by population class. These videos ranged from 2018 to 2020 and in time 30 sec to 40 min, with differing quality, and frame rate (30 or 60fs). Each video was of a tank that has multiple (hundreds) of one fish species (Snapper) within. Each tank can either contain one generation or multiple generations of fish, but are generally of the same size and maturity.

The variables for each video in this data frame are described in Table 1.

Table 1. Variable descriptions of data frame *videos for BIOID*

Variable name	Description
Species	Species of the fish – SNA: Snapper
year	Year video was captured
session_link	Video file location
video_length	Video length in mins
video_date	Date of capture
PFR_tank	Tank number that the fish class are located in.
tank_volume	Litres of water in the tank
Population_count	Number of fish in the current class
Population_class	Class number (generation number)
Population_age_dph	

There were 2-19 videos for each population class of which there were 6.

A fish population is defined as a group of individuals of the same species or subspecies that are spatially, genetically, or demographically separated from other groups [1].

These videos were processed using a *Frame-Grabber* python-based tool which was incorporated into a *Nextflow* script. Details are described in 3.1. This produced a folder structure for each video as below:

```
Output/
    └── video name/
        └── 1000 - ####/
            ├── frames/
            │   └── jpg files
            ├── individuals/
            │   └── jpg files
            ├── inferences/
            │   └── JSON files
            ├── spots/
            │   └── JSON files + associated jpg files
            └── tracks/
                └── JSON files
```

The *frames* folder contains each frame containing one or more observations of Snapper. Whereas the *individuals* folder contains individual fish observations cropped out from the frames. Inferences are the pixel locations of each individual fish across several frames, forming one continuous tracking event. The *spots* folder contains information on the pixel locations of the spots per individual fish identified.

The descriptive statistics data frame was produced to summarise the information for each video above, with the following variables:

"FileName", "FrameRate", "TotalNumFrames", "NumFramesWithIndividuals", "NumOfIndividuals", "MedianGap", "WaterQuality", "Brightness", "Camera"

2.2 Missing data

The level of missing data could be vast, depending on what in this project is classified as 'missing data'. Fish outside of the frame, therefore not detected by *Frame-Grabber* could be classified as missing data however it is nearly impossible to have an 'eye' on all fish in the tanks. Ideally, all fish within the tank would pass by the camera enabling the detection of each individual eventually in the time of the video. This may not be the case, giving rise to missing fish. There are other possible factors contributing to fish not being detected such as overlapping fish obscuring the fish body thus the spots, water quality, depth of view from camera lenses, or objects could all contribute to preventing the *Frame-Grabber* from identifying fish.

2.3 Outliers

In terms of video length, there were a few videos with a significant video length (nearly an hour) which could introduce bias in the processing time and may contribute to the overrepresentation of this type of water quality and brightness, therefore, introducing bias into the final statistics.

2.4 Imbalance

There is an imbalance in the type of camera used to collect video. The GoPro footage is largely represented over the SnapIT and Phone cameras.

Does not pose an issue as not predicting anything based on this data.

3 Methodology

3.1 Processing videos for frame generation

Frame-Grabber tool is a PFR-developed python-based pipeline that extracts images of individual fish that are suitable for downstream work (e.g., underwater BioID & measurement) from input videos. It uses a tracking module to assign individuals to clusters (tracking events) and uses the measurement feature detection model to identify which images are of high enough quality to measure (e.g., have all fish morphology features present). It is a command line-based application that allows for specific arguments to be passed, most importantly the frame increment which was set to 2 to evaluate every 2^{nd} frame.

Nextflow is a tool that allows for the parallel processing of inputs against an existing pipeline to produce outputs. We used this to run parallel instances of *Frame-Grabber* upon the 48 videos chosen. The command line bash call to *Frame-Grabber* was incorporated into the *Nextflow* script (see Supplementary material). A text file containing all the video paths was passed through, and Slurm was used to manage the task operations on the High-Performance Computing platform. The Slurm Workload Manager, formerly known as Simple Linux Utility for Resource Management (SLURM), or simply Slurm, is a free and open-source job scheduler that is used to manage and allocate resources in high-performance computing environments. Due to the length of some videos a timeout occurred for single tasks on the HPC, to alleviate this problem, the script was changed to split the videos into chunks of 1000 frames. These chunks were then processed iteratively as separate tasks via Slurm, avoiding the time-out issue.

The output was a series of folders per video as described in 2.1.

3.2 Analysing generated frames and observations

Jupyter notebooks with python were used to analyse the resulting frames. Data frame generation and visualisation were separated into two separate notebooks, with the first outputting the CSVs required for the second notebook, allowing it to be re-run without the data frame generation. The first notebook *GenDataframes* looped through all video folders to generate dictionaries which were transformed into a Pandas data frame named *video_df* with the following variables:

"FileName", "FrameRate", "TotalNumFrames", "NumFramesWithIndividuals", "NumOfIndividuals", "MedianGap", "WaterQuality", "Brightness", "Camera"

The variables "Water quality", "Brightness", and "Camera" had to be manually determined and entered into the corresponding dictionaries. Determining these by some code threshold could have been implemented, but this posed challenges around what was considered good/bad quality water and brightness.

A second data frame, *observations_df,* was created to show the number of spots per observation, and enriched with the following variables:

"ObservationPath", "FirstFrameNumber", "IndividualID", "WaterQuality", "Brightness", "Camera", "ObservationAreaPx", "NumSpots"

Table 2. Variable descriptions of data frame *observations_df.*

Variable name	Description
ObservationPath	.json file location
FirstFrameNumber	Year video was captured
IndividualID	First frame number plus individual ID from filename
WaterQuality	Either: clear or murky
Brightness	Either: dim or bright
Camera	Either: gopro, snapit, phone
ObservationAreaPx	Pixel location of observation within the frame
NumSpots	Total number of spot predictions per observation

In the second notebook, *GenPlots* imports the CSV data from the previous notebook, groups the data and visualises the findings.

4 Results

4.1 Video statistics

Analysis of the video datasets found there to be a wide-ranging number of frames per video up to 70,000 frames and 40 mins long, with a skewness towards much shorter 1-15 minute videos (Figure 1).

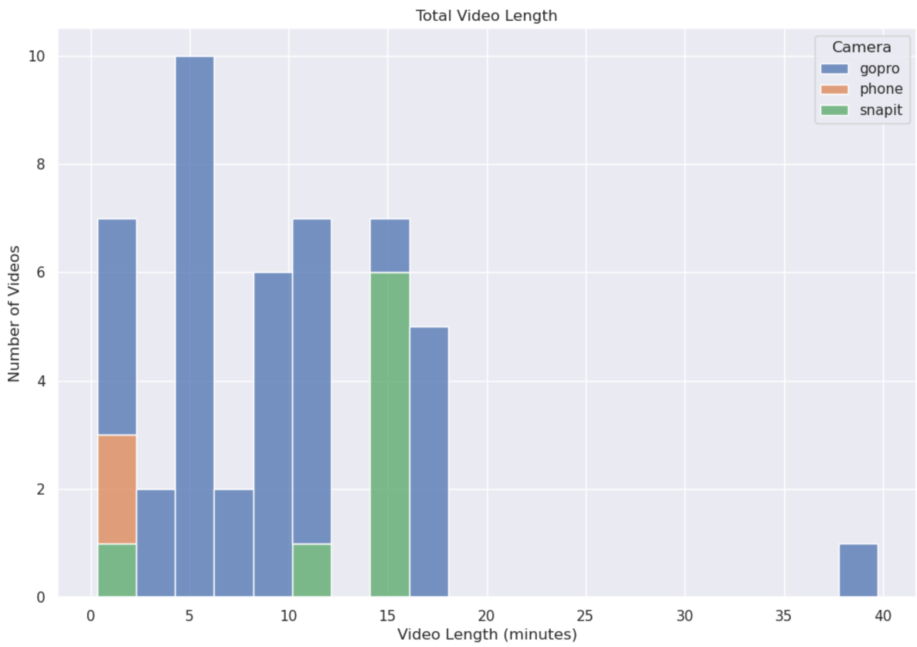

Figure 1. Total number of frames per video of a total of 48 videos and a total of 990,000 frames.

Therefore there was a right skew in the distribution of frames per video as seen in Figure 2. With most videos having 0-20,000 frames. With videos filmed by GoPro again being the major camera type used.

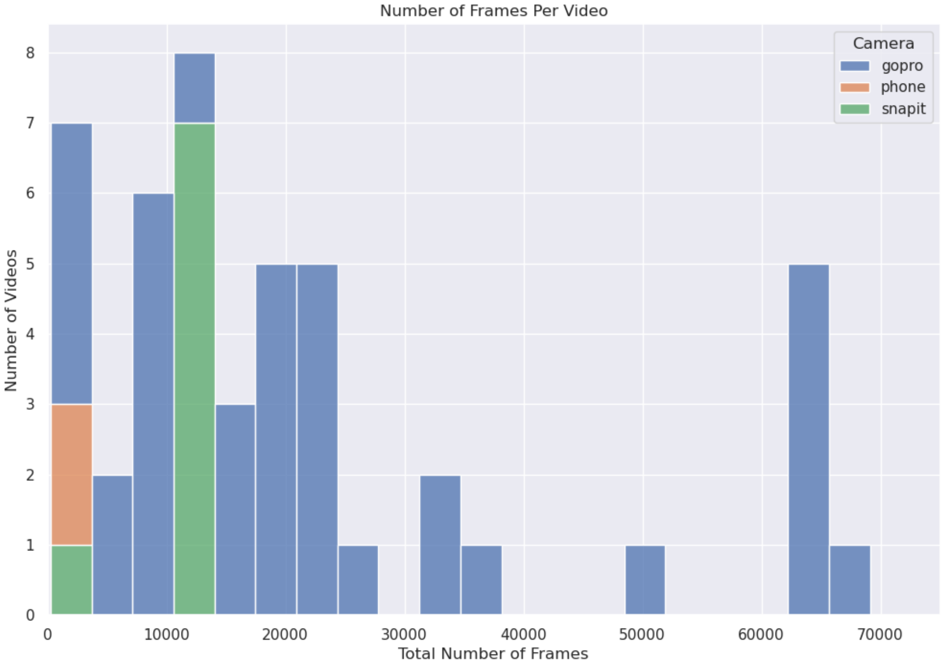

Figure 2. Total number of frames per video of a total of 48 videos and a total of 990,000 frames.

4.2 Observation counts

Within the total number of frames, the *Frame-Grabber* tool found a total of 190,000 observations across 17,000 tracking events. An observation is a single detected fish in a single frame and a tracking event is a sequence of linked observations across consecutive frames. The median number of observations per 1000 frames was calculated to allow comparison between either water quality or camera type (Figure 3). Clearwater quality was associated with a higher IQR (120 – 260) of observations compared to the murky water observations IQR (55-220).

GoPro camera encompassed the full range of observed number of observations per 1000 frames, with an IQR of 100-250 observations (Figure 4). Phone and SnapIT cameras have relatively short observation IQRs. This is due to the low representation of these camera types in the data.

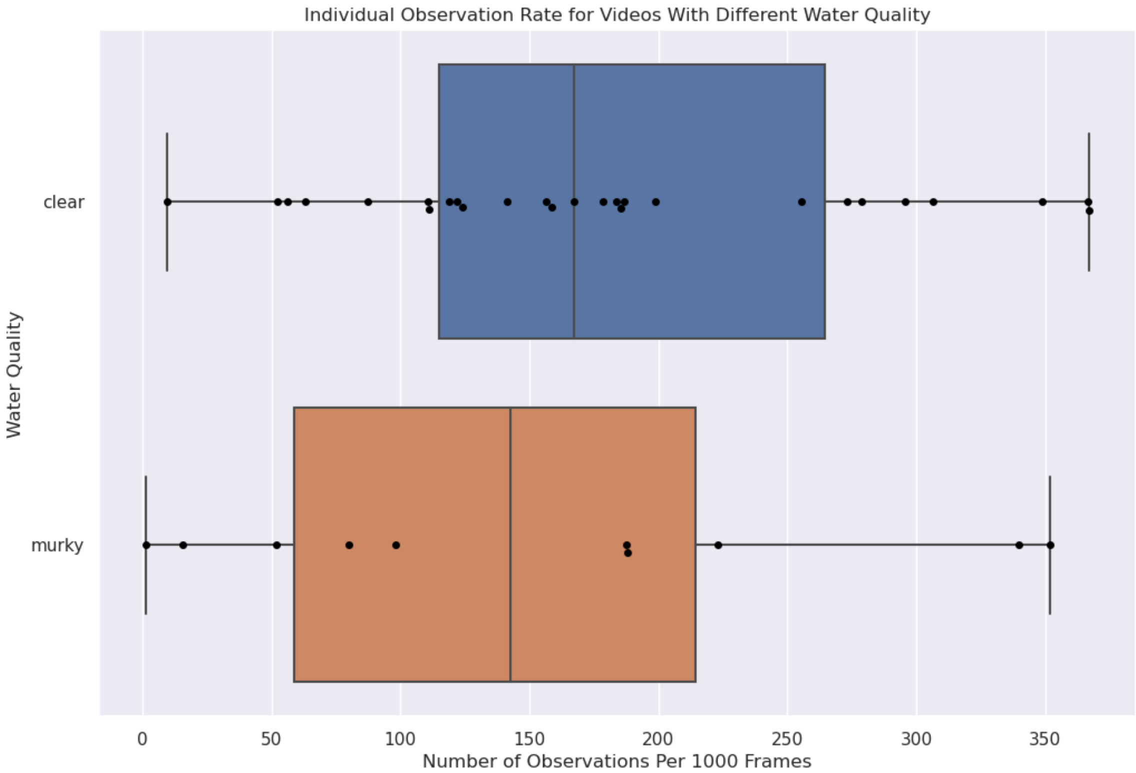

Figure 3. Individual observation rate for videos with different water quality. Only data for the camera GoPro.

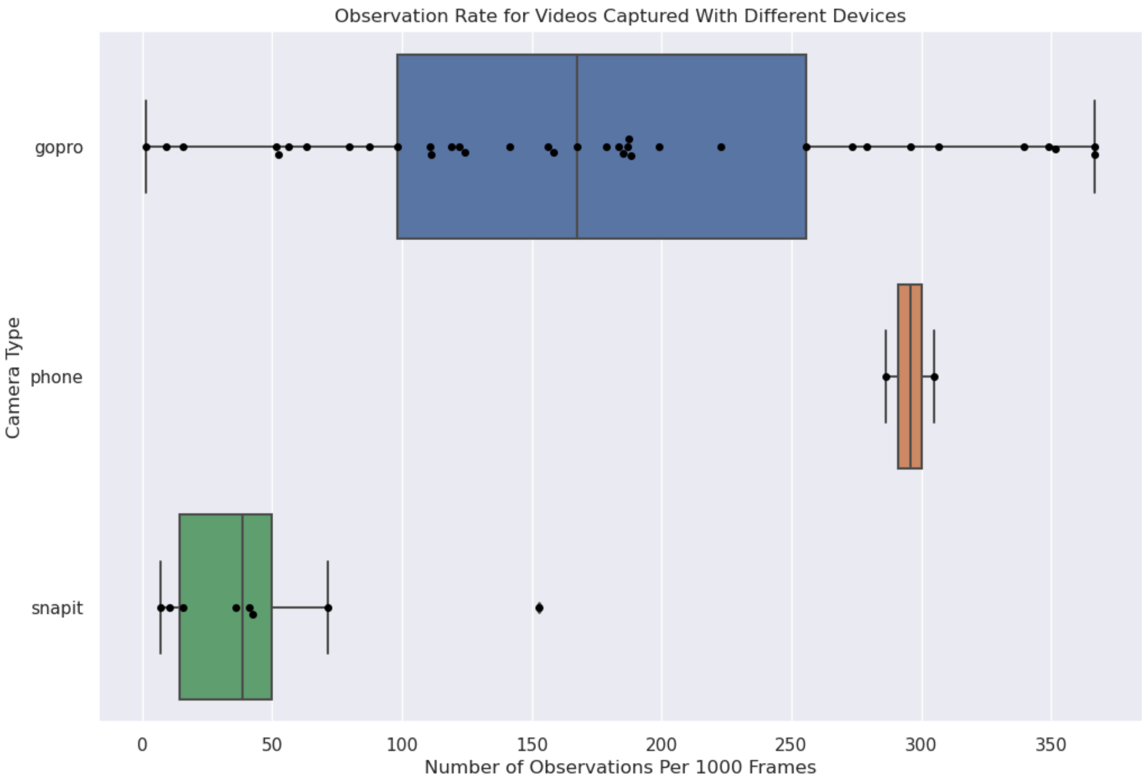

Figure 4. Observation rate for videos captured with different devices. The observation rate is per 1000 frames.

4.2.1 Median gaps between observations

The median number of frames without observations (gaps) between successive observations was calculated for each video in the *GenDataframes* notebook and plotted against water quality. The results are shown in Figure 5, which demonstrates a positive skew for clear water towards a lower number of gaps, but with a wider distribution compared to the gaps in murky water. This runs contrary to the logical expectation that murky water would produce a greater number of (apparently) empty frames. However, this is not the case as can be seen here.

A possible reason for this could be due to the Snapper spot composition. The reflective quality of the spot is not due to pigment, rather it is caused by physical properties (Crystalline structure) in the scales which reflect light possibly of a certain wavelength. In most cases this wavelength of light could be filtered for in murky water, with other confounding wavelengths blocked, giving a much brighter spot. Although this is speculative, an examination of images of murky water quality reveals that the spots are indeed much brighter than when there is bright sunlight shining through the clear water.

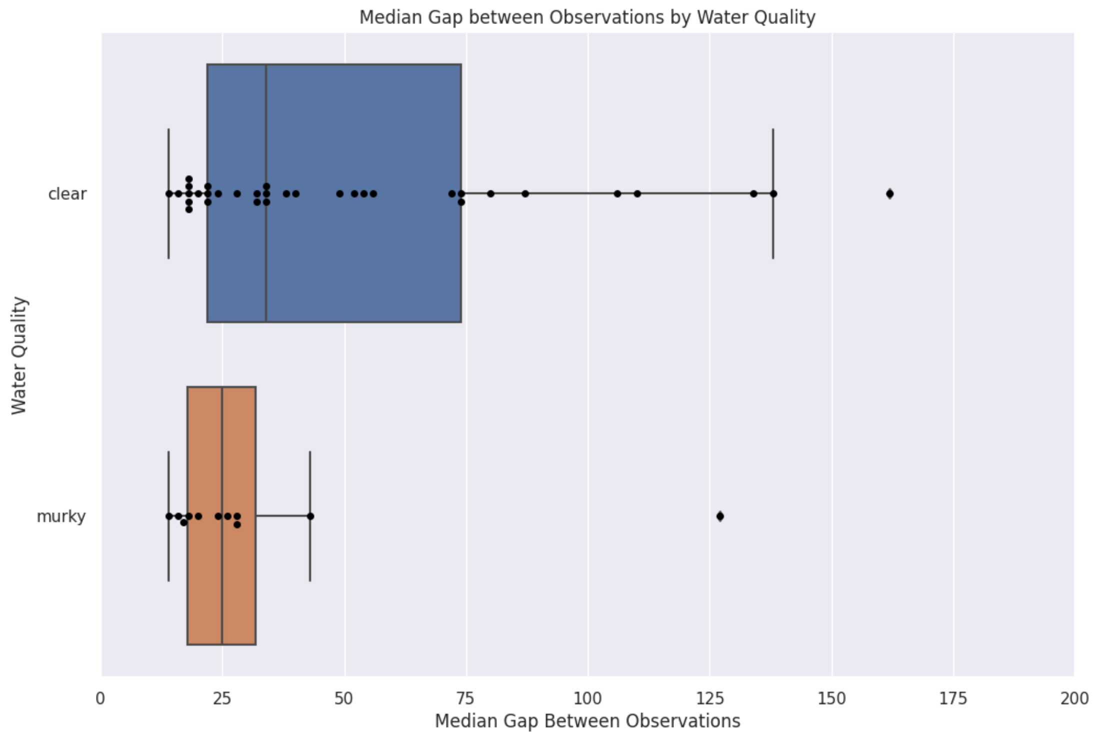

Figure 5. Median Gap between observations by water quality. The median gap refers to the median number of frames without observations of fish.

The median gap between observations varied among different camera types, as shown in Figure 6. The lowest median gaps were observed for phone videos and GoPro cameras, with an interquartile range (IQR) of 20-50 frames. On the other hand, the SnapIT camera had the highest median gap of 75 frames. The GoPro and phone cameras produced the best results with the least number of gaps. Further analysis with more SnapIT videos might help determine the best camera for the application.

Additionally, the low quality of SnapIT frames, with a grainy appearance (as seen in appendix 1), could be affecting the *Frame-Grabber's* ability to accurately distinguish observations from non-observations.

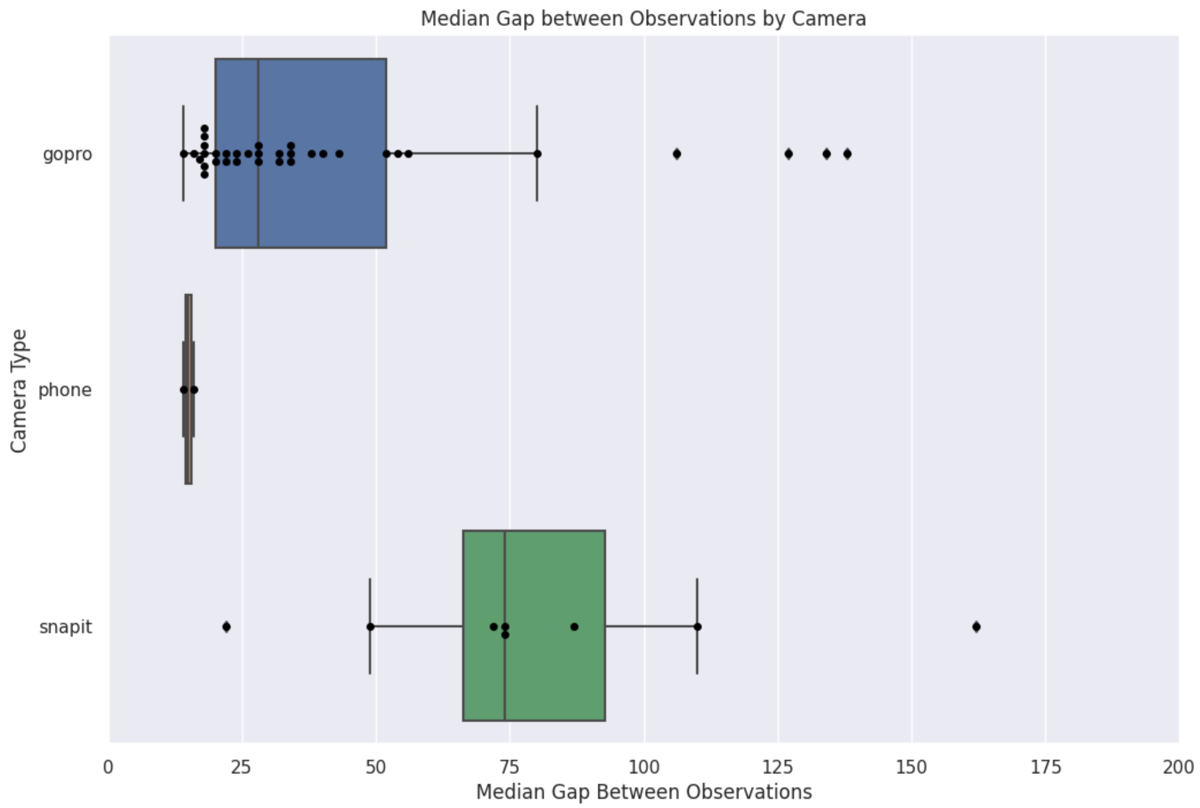

Figure 6. Median Gap between observations by camera type (GoPro, Phone, SnapIT). The median gap refers to the median number of frames without observations of fish.

4.3 Spot counts

Knowing how many spots are detected per observation is important to know for the application of BioID to these observations. BioID requires at least 15, ideally more than 20 spots, for the correct identification of an individual. Water quality is a factor that can impact spot detection, we compared the number of spots detected for each observation against water quality (Figure 7). There were a considerable number of observations with 0-50 spots captured by the GoPro, and so do contain some observations with the required number of spots. All SnapIT videos contained a very low amount of spots and thus no useable observations, but the videos selected were of all of the clear water quality so have the potential for being good observations if there was not high video compression resulting in grainy appearance (Appendix Fig 4).

Looking at the number of usable observations we have can inform us on whether it is ready to apply BioID as we need enough observations to make linkages between individuals. Across the 48 videos that were processed, we concluded there to be 990,000 frames total. Within these frames, there were a total of 190,000 observations across 17,000 tracking events. Of these fish observations, 2300 observations had >15 spots (0.23% of all frames considered) and 1800 observations had >20 spots

(0.18% of all frames considered). The caveat with this is that frames can have multiple observations, i.e spots from other fish in the background can be counted in each observation, therefore some of the observations with >20 spots will be from multiple fish. The number of images with sufficient observations (2300) is enough to try applying BioID, however, the rate compared to the total amount of frames is low and it would be more useful to improve on this before moving forward with underwater BioID.

The number of useable observations is dominated by the GoPro cameras, as to be expected (Figure 8). It is noteworthy to observe that the slope in the range of 15-25 spots per observation is linear on a logarithmic scale. This suggests that a "power law" might be at play, meaning that the relationship between two variables follows a pattern such that one variable is proportional to the power of the other. This observation is interesting because it sheds light on the underlying mechanism driving the number of spots detected per observation.

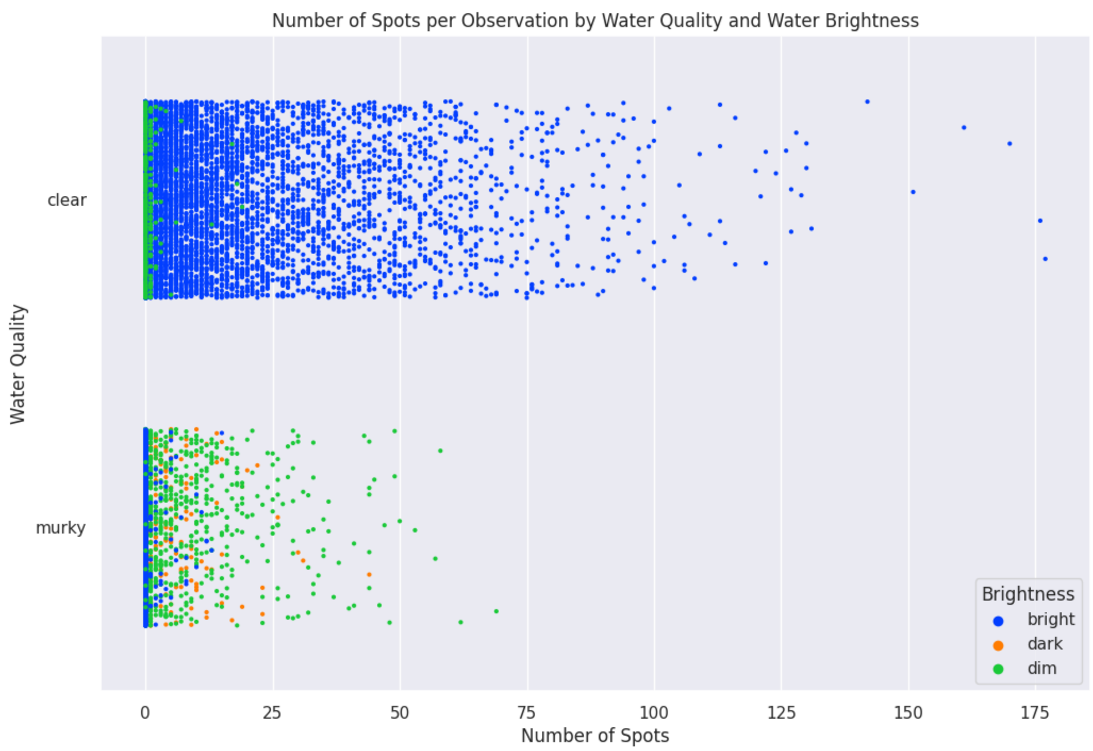

Figure 7. Observations were grouped by video_name and individual_ID, aggregating the number of spots by the maximum.

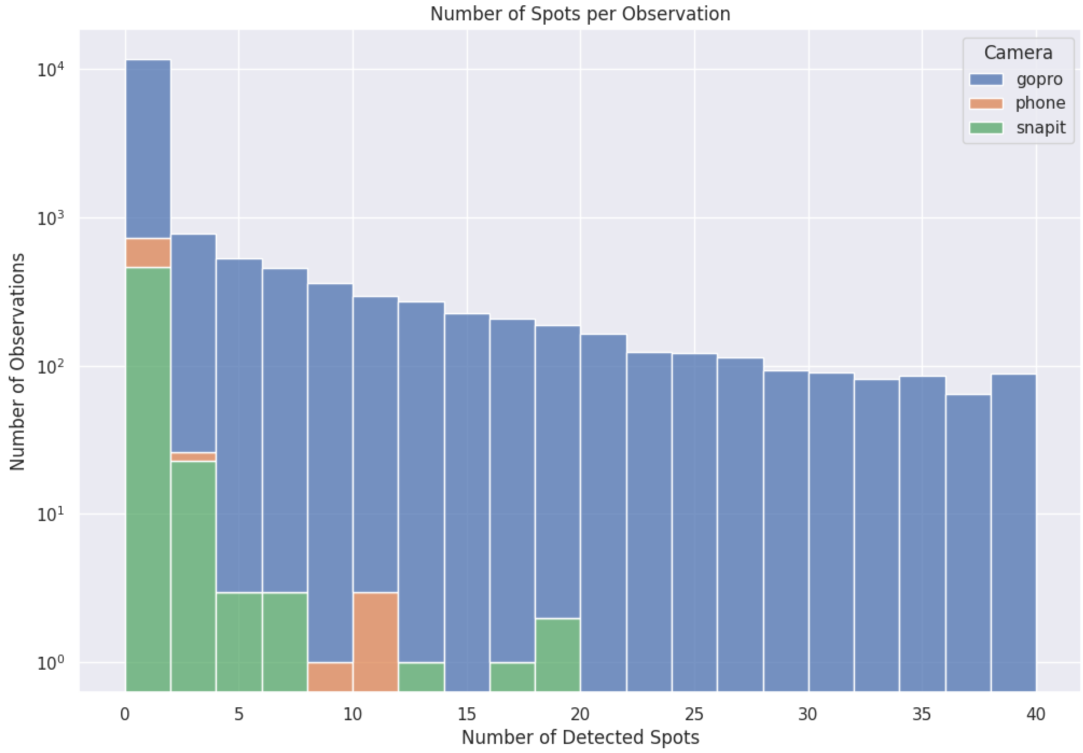

Figure 8. The number of spots per observation of fish. Observations grouped by video_name and individual_ID, aggregated number of spots by maximum.

5 Conclusion & Discussion

One of the initial goals for this project was to develop better tools for processing and understanding the underwater video data that gets collected. The first steps of this goal were to process the available videos to determine the number of observations typically seen per thousand frames, and how many of these observations had the useable number of spots required for BioID identification of individuals. We determined there to be only a few hundred observations of useable quality (15-25 spots), with a success rate of 0.23% out of 990,000 frames processed. Various factors including water quality, video compression and ML model performance contributed to this low number of useable observations. Therefore, it was decided to focus more on improving data collection and observation (fish structure) and spot detection models/pipelines.

In reference to the objectives of this project, some were not accomplished due to the previously described problems. The first goal was successfully achieved as we were able to identify the connections between water quality, camera type, and their impact on the number of observations and spots counted. The data was visualized, enabling the science team to make modifications to the methods and timing of capturing Snapper video data. However, goals 2 to 6 were not achieved due to limitations in the quality of the data set and time constraints. Nevertheless, I did have a brief opportunity to create a web-based application using Python to count small fish from images captured via a phone's web application.

In this section, we will discuss the challenges in processing the videos, comparisons between good/bad images, how to improve data collection, and possible improvements upon the ML models used within the *Frame-Grabber* to extract usable observations.

5.1 Processing

Processing the videos on the high-performance computing system (HPC) was a challenge due to the restrictive timeout settings. Increasing the time at which a particular task in Slurm could take to 36 hours etc was not enough as some tasks would stall. This led to splitting the videos into chunks of 1000 frames, processing each at a time. The time it took to process 48 videos was many days.

Several methods could be used to decrease the time for processing, however, this is not an urgent issue. These include increasing the use of parallel processing techniques (currently about 4 tasks could run at one time) or finding the optimal value of CPUs/GPUs per task etc.

5.2 Comparison of good vs bad images for BioID

Since spot detection was largely dependent on the quality of images it is worthwhile looking into what good and bad images look like. We extracted example observations from each type of water quality and camera type to see what the *Frame-Grabber* was seeing and determining.

In a successful case, the lighting is bright, and the water quality is clear. Leading to bright luminescent spots as can be seen in Appendix Figure 2. Notably, these images are against a dark backdrop which could either be the tank or collection of fish, this may aid in the ML's detection of fish morphology. There are very few videos containing these such clear observations

Lighting and water conditions are integral for spot visibility. As can be seen in Appendix Figure 3, even though the spots are vaguely noticeable to the human eye, the particulate matter and low light are enough to hinder the ML model's ability to identify spots on a correct observation of a fish.

Excessive video compression mainly seen in SnapIT videos leads to the failure of the model to identify spots on correct observations (Appendix Figure 4).

5.3 Data collection

There are multiple ways to enhance the collection of data to improve the detection of fish observations and spots.

Increasing shutter speed: By increasing the shutter speed of the cameras used to collect data, less light can enter the lens, leading to crisper and more detailed images but darker [2]. This could help improve the accuracy of detecting fish observations and spots, as the model will have better-quality images to learn from, with higher detail around the spots.

Less video compression: By reducing the amount of compression applied to the videos, the quality of the data can be improved, increasing the accuracy of the observation and spot detection models.

Camera lens: The number of observations per thousand frames is also a function of the camera lens used. A wider lens may capture more observations in a single frame compared to a narrow lens.

Fish density: The density of fish in the environment being captured can also affect the number of observations per thousand frames. Higher fish densities will generally result in more observations, whereas lower fish densities will result in fewer observations. Using smaller tanks could result in more observations, however, there are considerations such as overlapping of fish and ethical implications of keeping a larger number of fish in smaller spaces.

Remote access: The difficulty of accessing the SnapIT videos, which are far away out in the sea and have only local storage, can make it challenging to minimize video compression. To save enough files on the camera rigs the video needs to be compressed.

5.4 Improving ML models

The observation and spot detection models used in the *Frame-Grabber* step had many observations that were useable in principle but had been cropped incorrectly or spots missed. There are several ways these ML models could be improved.

One way to improve the accuracy of the observation and spot detection models in the *Frame-Grabber* is by collecting more diverse data. This could include a wider range of underwater environments, species, and conditions, which can help the model recognize a greater variety of observations and spots.

Another way to improve the performance of the models is by cleaning and reducing the noise in the training data. This could include removing irrelevant or redundant information, correcting errors, and standardizing the data to make it more consistent and meaningful. However, this is dependent on the model type, whereby in the case of a CNN model it would be better to train in the presence of noise [3].

Feature engineering is the process of creating new features from existing ones to enhance the performance of machine learning models. This can be applied to the models used in the *Frame-Grabber* by extracting new features that better capture the relevant information in the images [4].

The architecture of the machine learning models used in the *Frame-Grabber* could also be improved. This can involve trying different types of models, such as convolutional neural networks (CNNs), Graph Convolutional Networks (GCN) [5] or Capsule Neural Networks [6], to find the best architecture for the data.

Hyperparameters are settings that control the behaviour of machine learning models, such as the number of layers, the learning rate, or the number of neurons in a layer. Hyperparameter tuning is the process of optimizing these settings to find the best combination for a particular task. By tuning the hyperparameters of the observation and spot detection models, the accuracy of the models can be improved.

Annotated data in this context is data that has been labelled with Snapper morphology identifiers. Providing more of this annotated data can help improve the accuracy of the fish morphology and spot detection models by providing more examples for the model to learn from. Annotations are time-consuming and need high expertise to know what features on the fish are required.

Lastly, generating synthetic data for training could help improve the models. Synthetic data is data that is generated artificially, rather than being collected from benchtop or underwater images. Generating synthetic data can help improve the accuracy of the observation and spot detection models by providing additional training examples for the model to learn from. This can be especially useful if it is difficult or expensive to obtain real-world data.

5.5 Future goals

The future goals of this project aim to improve the accuracy and efficiency of evaluating video datasets for underwater BioID by focusing on the following steps:

1. Generating a robust dataset of videos with clearer observations of Snapper. This will be achieved by capturing high-quality videos at specific time points and positions that clearly show the presence of Snapper, which will make it easier to conduct accurate identifications using BioID.
2. Setting up automated pipelines: Automating the processing of new videos and storing observations in databases will help streamline the data collection and analysis process, making it faster and more efficient.
3. Training the BioID ML on a wider range of images: To improve the accuracy of the BioID ML, it will be important to train it on a wider range of images, including murky and dim shots.
4. Making the database searchable from within the finfish facility by Nelson seafood staff. So they will be able to easily access and analyse the data collected. This will help them make informed decisions about their operations and fish management strategies, which will help to ensure the health and sustainability of the fishery populations.

6 References

[1] Wells, Jeffrey. (1995). Populations, Metapopulations, and Species Populations: What Are They and Who Should Care?. *Wildlife Society Bulletin*. 23. 458-462.

[2] T.Holmes. "What Is Shutter Speed for video?" https://wistia.com/learn/production/what-is-shutter-speed (accessed Feb. 4, 2023)

[3] J.Brownlee."Train Neural Networks With Noise to Reduce Overfitting" machinelearningmastery.com https://machinelearningmastery.com/train-neural-networks-with-noise-to-reduce-overfitting/ (accessed Feb. 4, 2023)

[4] J.Brownlee. "Basic Feature Engineering With Time Series Data in Python". https://machinelearningmastery.com/basic-feature-engineering-time-series-data-python/ (accessed Feb. 4, 2023)

[5] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, Volume 1, 2020, Pages 57-81, ISSN 2666-6510.

[6] S.J. Pawan, Jeny Rajan. Capsule networks for image classification: A review. *Neurocomputing*, Volume 509, 2022, Pages 102-120, ISSN 0925-2312.

7 Appendices

7.1 Extra graphs

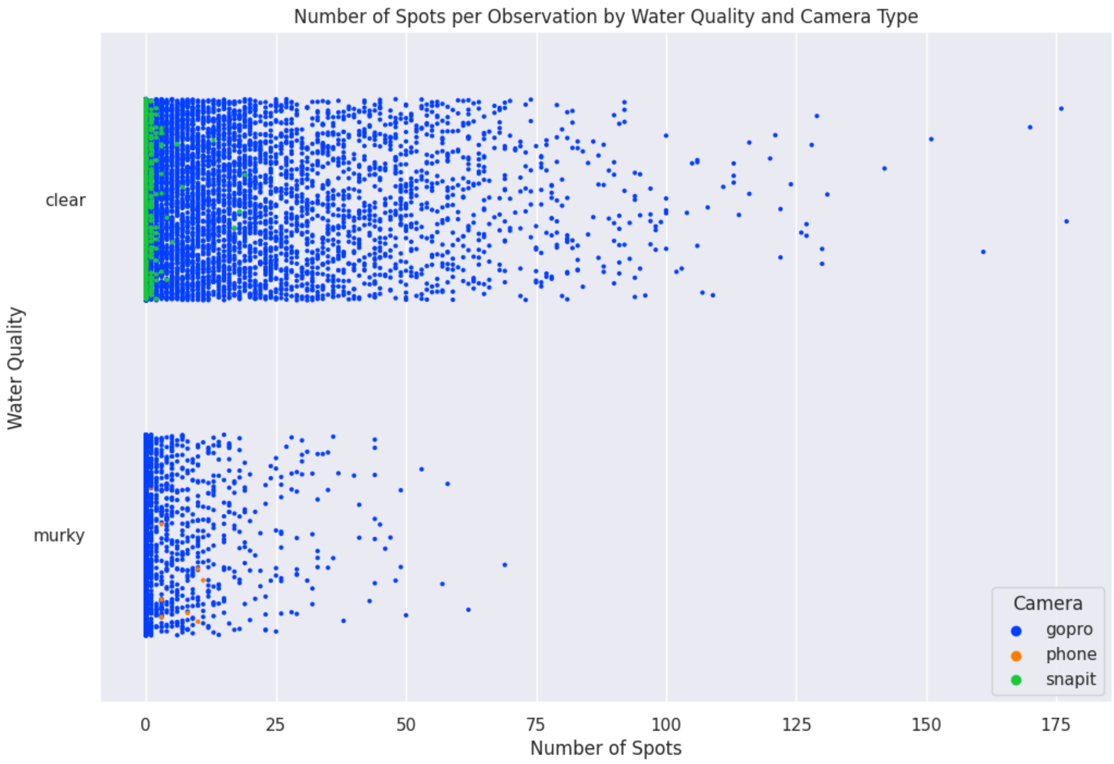

Figure 1. The number of spots per observation by water quality and camera type. Observations grouped by video_name and individual_ID, aggregated number of spots by maximum

7.2 Camera frame examples

Figure 2. Example of a Snapper with highly visible spots. Taken with a GoPro camera under clear water quality and bright conditions.

Figure 3. Example of Snapper with low visible spots. Taken with a GoPro camera under murky water quality and bright conditions.

Figure 4. Example of Snapper with low visible spots. Taken with a SnapIT camera under clear water quality and bright conditions in the open-ocean sea cages. Shows the high compression renders the footage mostly unusable for BioID.

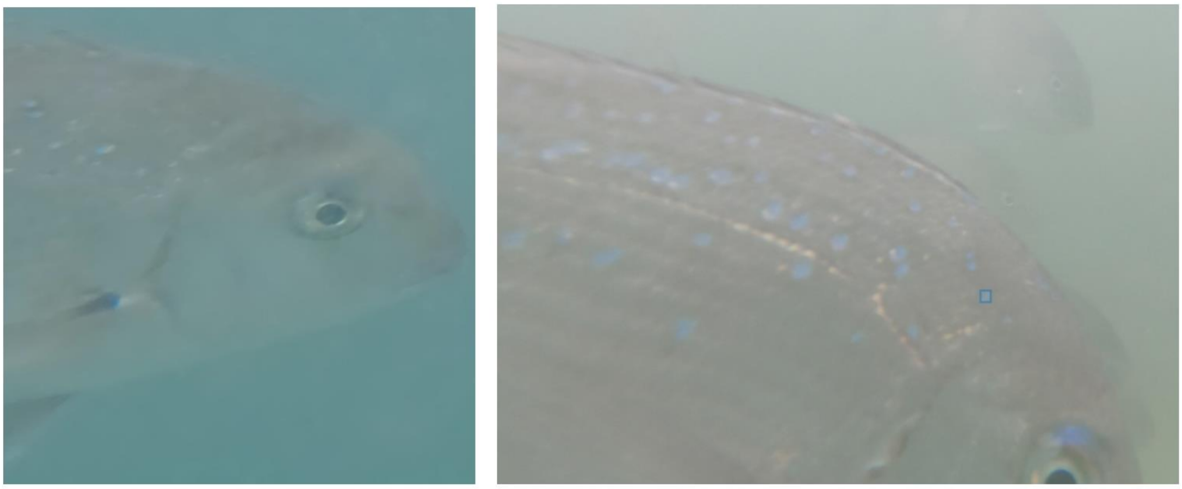

Figure 5. Example of ML model errors. Bounding box detection of fish fails for some observations. Most spots are not detected.

A smart
green
future.
Together.