Capstone 2 - Milestone Report I
Burns, Echelle
Feb 2020

Identifying Humans in Drone Footage from Local Beaches

1. *Introduction*

Throughout the years, researchers have used a variety of different technologies to address the movement patterns and behavior of particular shark species. In more recent years, however, shark researchers have begun to use drones and helicopters to capture footage of shark species to answer questions like: *How do these animals interact with one another when they are not being actively tagged or followed?* Using drone data itself presents a lot of issues when it comes to data processing, such as having a suitable reference item to accurately predict the sizes of objects in the drone's field of view and figuring out the orientation of the drone in order to georeference the resulting footage. However, although drone footage is relatively new to the field of marine biology, data scientists have been analyzing video footage and images for quite a while.

Shark attacks from a variety of species, including White Sharks (*Carcharodon carcharias*), Tiger Sharks (*Galeocerdo cuvier*), and Bull Sharks (*Carcharhinus leucas*) are heightened in the media and instill terror in much of the human population. Although shark researchers are aware that shark attacks do not happen as frequently as the media portrays, no research projects have addressed how frequently sharks and humans actually interact with each other, and how many times those interactions result in a shark bite. A graduate student at California State University, Long Beach is using video footage from drones and helicopters in order to answer this question, by identifying not only when White Sharks and humans are in the water at the same time, but also by categorizing how the sharks respond when approached by their human counterparts. However, gathering drone footage for such a project inevitably produces a lot of raw footage (several 10+ minute videos per day) that likely contains neither sharks nor humans. Going through each frame of a video one-by-one may take up a significant amount of one's time, leaving little time for true data analysis techniques. Therefore, the goal of this project is to take stills from drone video footage and train a deep neural network to identify how many humans are present within each frame. Although this work will be catered to a specific researcher and a specific problem, the methods from this project may also be adapted for researchers who are attempting to answer similar research questions or have similar streams of data.

2. *Materials and Methods*

Data Sources

A total of 2,465 drone images (3840x2160 pixels) were provided by the California State University, Long Beach Shark Lab. These images were taken at local beaches along the southern California coast and include features such as: humans that are participating in a variety of different activities (e.g. walking on the beach, wading in the shoreline, swimming, paddleboarding, kayaking, or surfing), White Sharks swimming near the ocean's surface, or other forms of marine life (e.g. dolphins, stingrays, kelp). The images were not yet labeled when they were received. Example images can be found in: SpringBoard-Capstone2/data.

Image Labeling and Splitting

Images were labeled by creating an interactive python script that would display images on the screen and allow the user to place dots on top of locations within the image where humans were present (SpringBoard-Capstone2/Data_Processing/Labeling_Images). For ease of labeling, the images were resized to 960x540 pixels. This size allowed the labeler to see the entire image at one time, and still allowed for easy identification of human subjects. There was no differentiation between different forms of human activity for this project, but future work should try to implement different labeling methods for different spots (e.g. walking on the beach, wading in the shoreline, swimming, paddleboarding, kayaking, or surfing).

The raw version of each image was then saved at 960x540 pixels in true color, grayscale, and HSV color formats (SpringBoard-Capstone2/Data_Processing/ DataWrangling_PhotoContrasts, Figures 1-3). Labeled images were saved as arrays with values of 0 (no human present) and 255 (location of human; Figure 4). Both raw images and labeled images were then split into 25 smaller images (192x108 pixels) in order to run the model more efficiently (SpringBoard-Capstone2/Data_Processing/ DataWrangling_ImageSlicing, Figure 5).

Data Wrangling

Since data were images, no extreme data wrangling procedures were implemented.

Exploratory Data Analysis

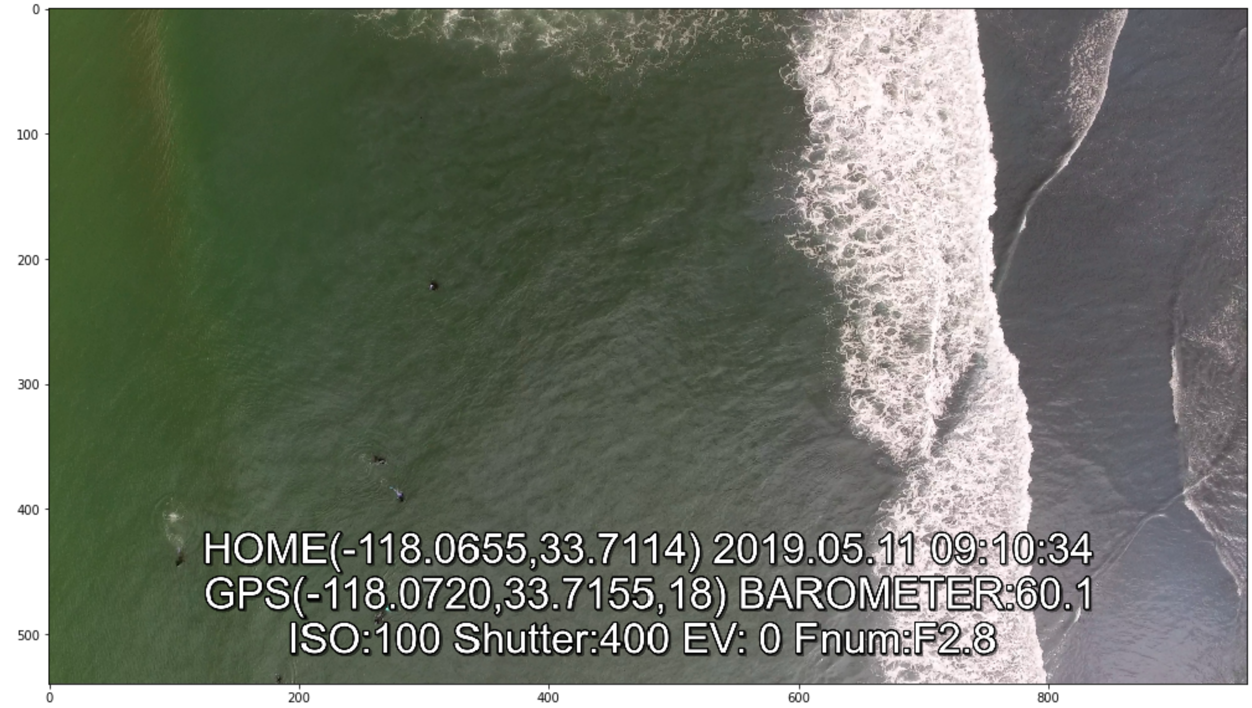
Of the 2,465 drone images that were provided, only 521 included at least one human subject. Therefore, only images that originally had a human in the frame were used for model training and validation. Once split into 25 smaller images, the dataset was comprised of 39,075 images (some with 0s present) to use for the model. Exploratory data analyses were run on the full size images (960x540 pixels) and split images (192x108 pixels) on this set of data, primarily to determine the underlying distribution of the number of people in each image (SpringBoard-Capstone2/Exploratory_Data_Analysis/Descriptive_Statistics).

3. *Results*

Exploratory Data Analysis

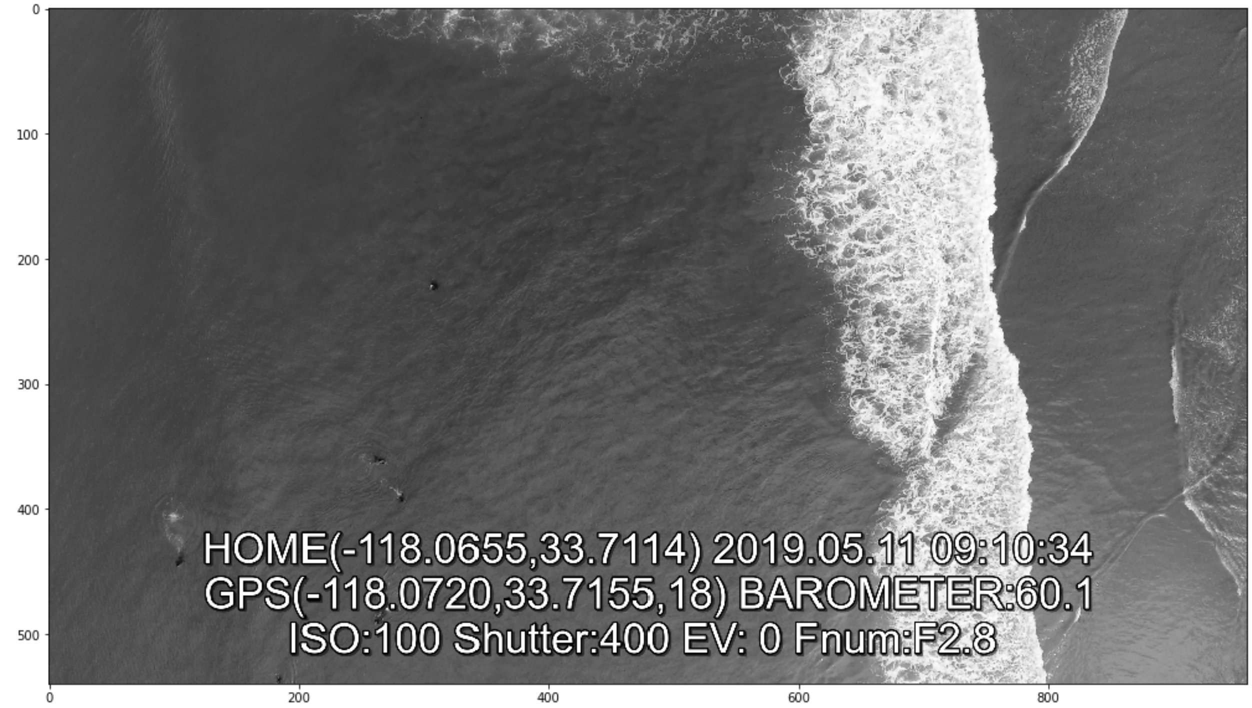
Most of the full size images had one person, and the frequency of images with more than one person decreased as the number of people in the image increased (Figure 6). Most images that were included in this dataset were comprised of images with 7 humans or less. Of the split images, however, over 90% had 0 people in the frame and approximately 7% of the images had only one person in the frame (Figure 7).

Figures



HOME(-118.0655,33.7114) 2019.05.11 09:10:34
GPS(-118.0720,33.7155,18) BAROMETER:60.1
ISO:100 Shutter:400 EV: 0 Fnum:F2.8
0
100
200
300
400
500
0
200
400
600
800

Figure 1. True color image of drone footage that includes human subjects (swimmers on the left of the figure.



HOME(-118.0655,33.7114) 2019.05.11 09:10:34
GPS(-118.0720,33.7155,18) BAROMETER:60.1
ISO:100 Shutter:400 EV: 0 Fnum:F2.8
0
100
200
300
400
500
0
200
400
600
800

Figure 2. Grayscale image of drone footage that includes human subjects (swimmers on the left of the figure). Content is the same as Figure 1.

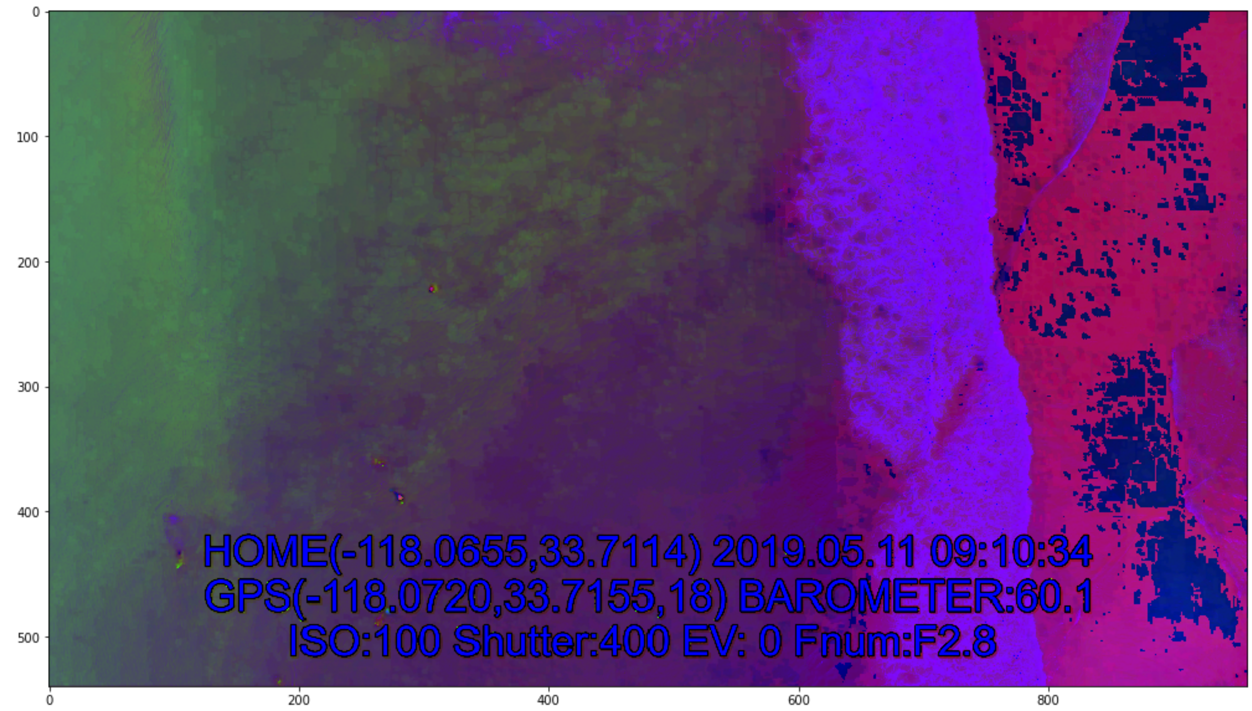

0
100
200
300
400
500
0
200
400
600
800
HOME(-118.0655,33.7114) 2019.05.11 09:10:34
GPS(-118.0720,33.7155,18) BAROMETER:60.1
ISO:100 Shutter:400 EV: 0 Fnum:F2.8

Figure 3. HSV image of drone footage that includes human subjects (swimmers on the left of the figure). Content is the same as Figure 1.

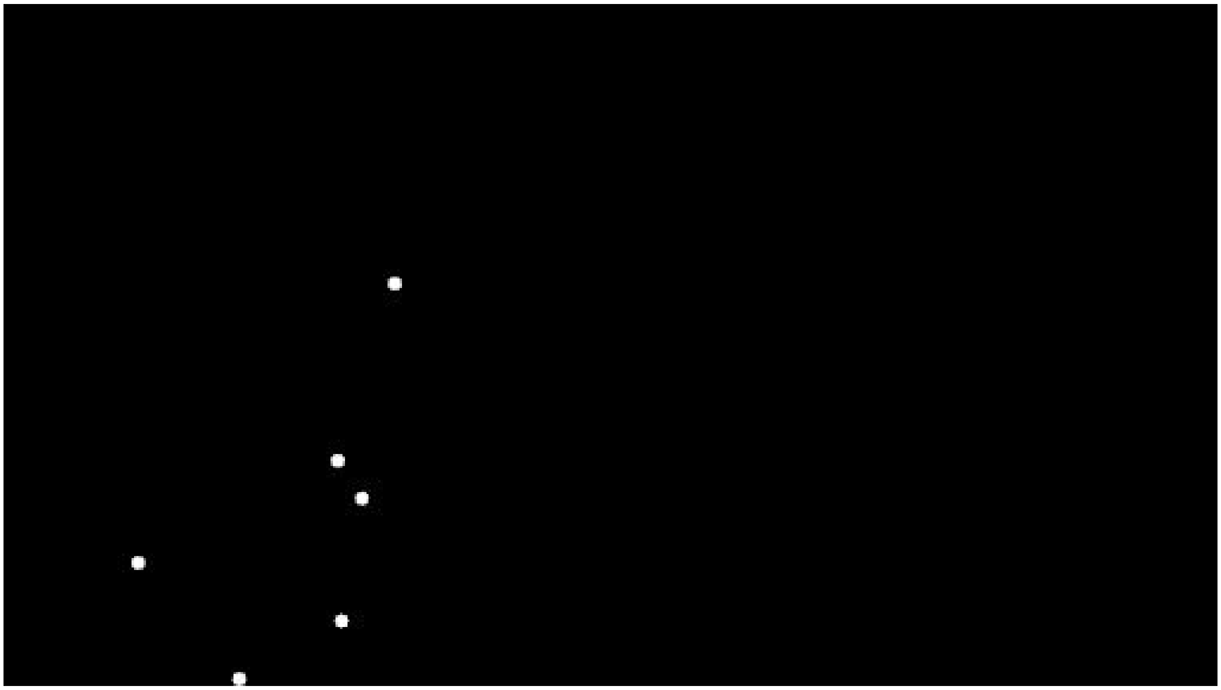


Figure 4. Labeled image of drone footage that includes human subjects, where all black (0) indicates no humans and white (255) indicates the location of a human. Content is the same as Figure 1.

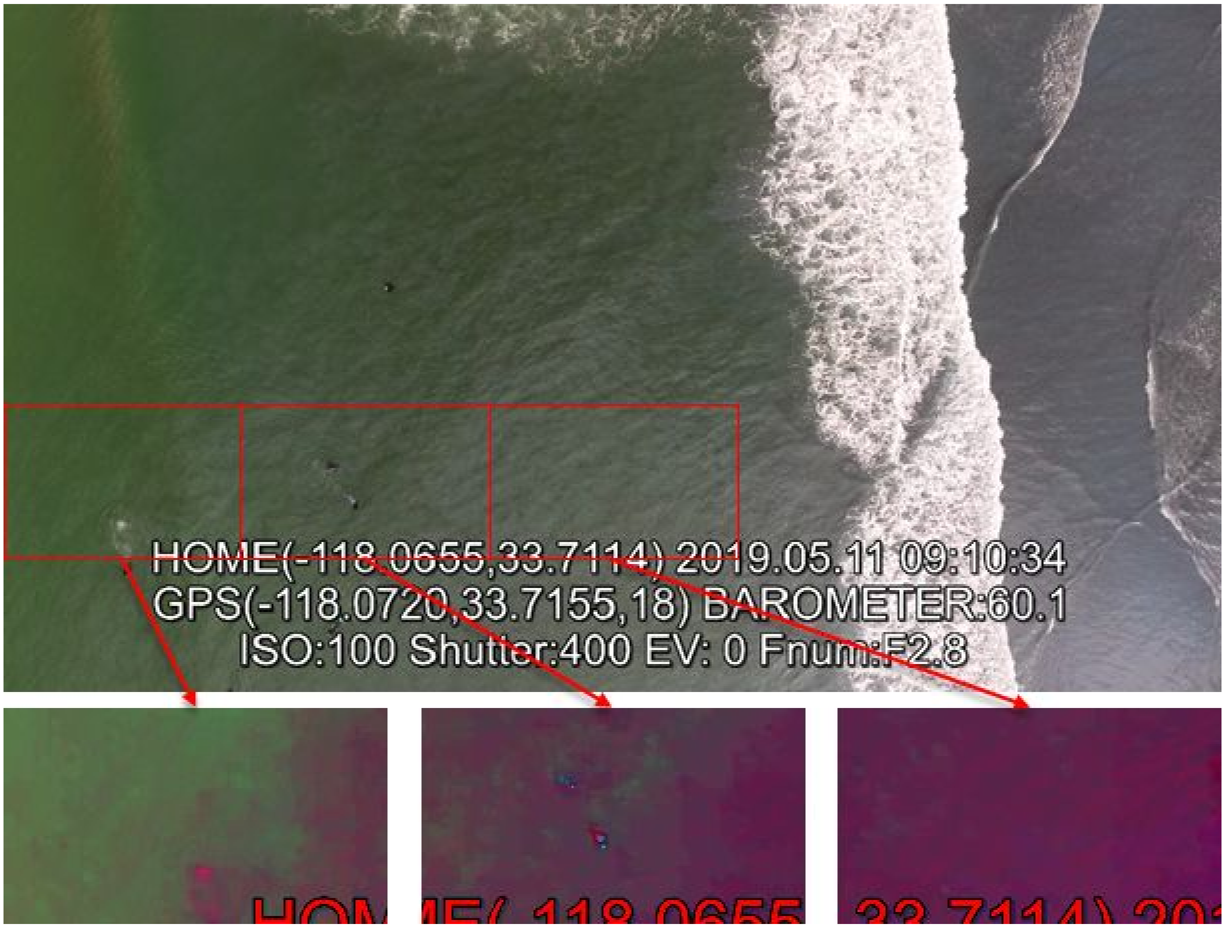

HOME(-118.0655,33.7114) 2019.05.11 09:10:34
GPS(-118.0720,33.7155,18) BAROMETER:60.1
ISO:100 Shutter:400 EV: 0 Fnum:F2.8

Figure 5. Three images from a raw image (960x540 pixels) that was split into 25 smaller images (192x108 pixels) and converted to HSV. Content is the same as Figure 1.

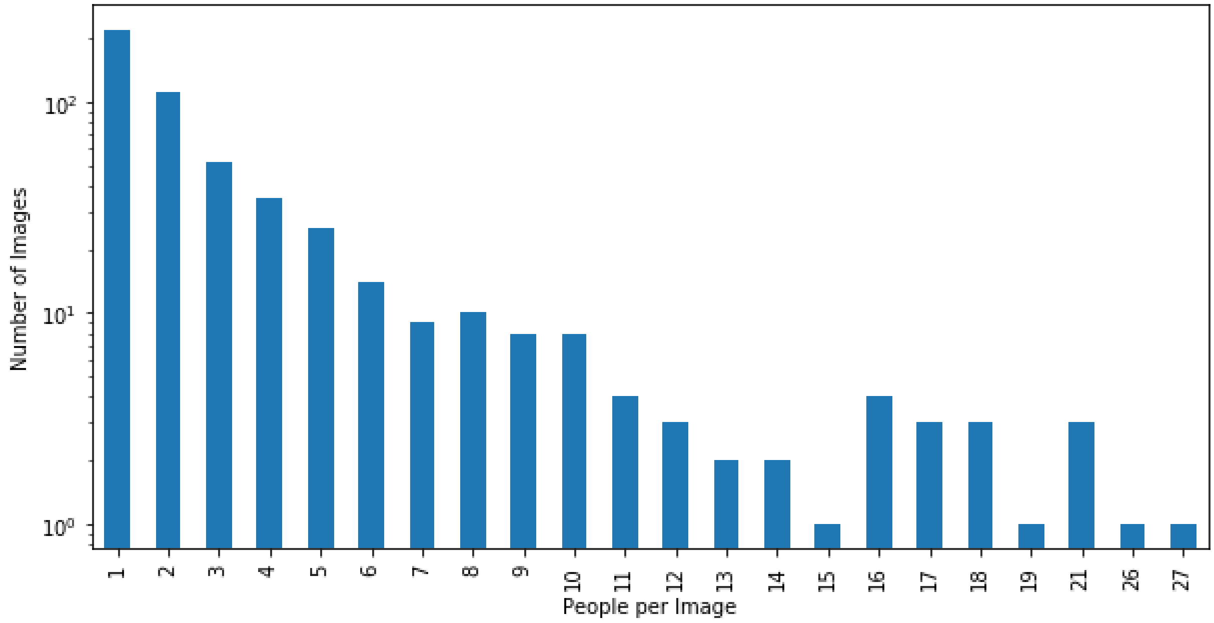


Figure 6. The frequency of raw images (980x540 pixels) that had varying numbers of human subjects. Data were taken only from frames that included at least one person.

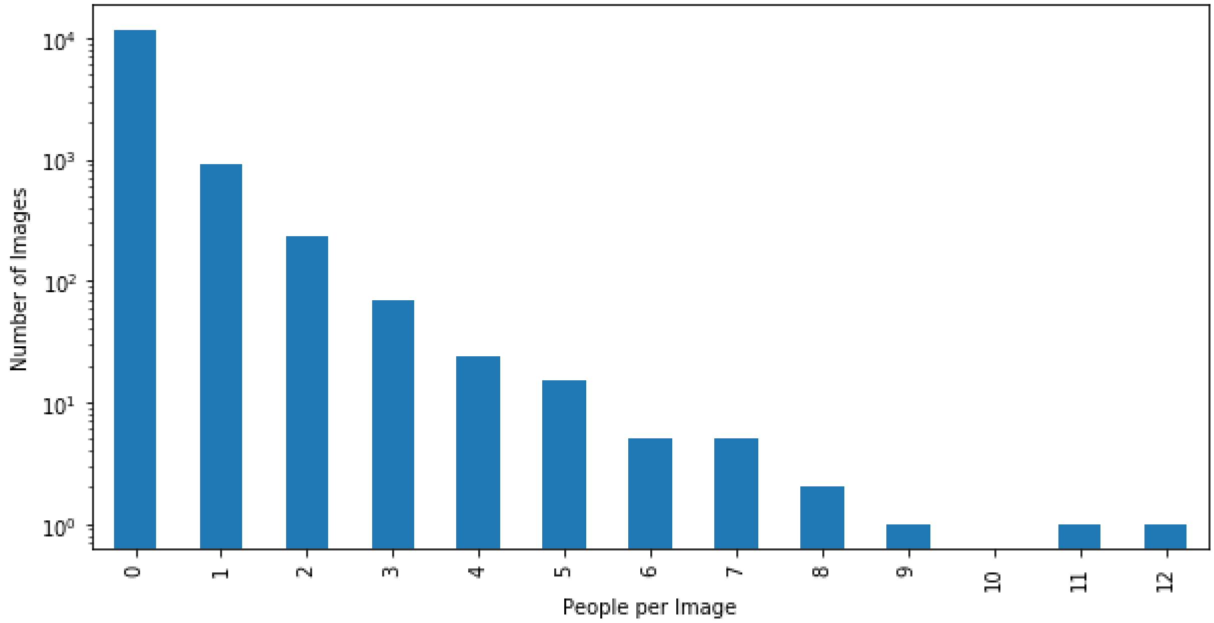


Figure 7. The frequency of images that were split (25 images of 192x108 pixels per raw image) that had varying numbers of human subjects. Raw images included at least one person.