SEALNET: Facial recognition software for ecological studies of harbor seals

Zachary Birenbaum[1], Hieu Do[1], Lauren Horstmeyer[1], Hailey Orff[1], Krista Ingram[1], and Ahmet Ay[1]

[1]Colgate University

January 6, 2022

Abstract

Methods for long-term monitoring of coastal species such as harbor seals, are often costly, time-consuming, and highly invasive, underscoring the need for improved techniques for data collection and analysis. Here, we propose the use of automated facial recognition technology for identification of individual seals and demonstrate its utility in ecological and population studies. We created a software package, SealNet, that automates photo identification of seals, using a graphical user interface (GUI) software to identify, align and chip seal faces from photographs and a deep convolutional neural network (CNN) suitable for small datasets (e.g., 100 seals with five photos per seal). We piloted the SealNet technology with a population of harbor seals located within Casco Bay on the coast of Maine, USA. Across two-years of sampling, 2019 and 2020, at seven haul-out sites in Middle Bay, we processed 1529 images representing 408 individual seals and achieved 88% (93%) rank-1 accuracy in closed set (open set) seal identification. We identified four seals that were photographed in both years at neighboring haul-out sites, suggesting that some harbor seals exhibit site fidelity within local bays across years, and that there may be evidence of spatial connectivity among haul-out sites. Using capture-mark-recapture (CMR) calculations, we obtained a rough preliminary population estimate of 4386 seals in the Middle Bay area. SealNet software outperformed a similar face recognition method developed for primates, PrimNet, in identifying seals following training on our seal dataset. The ease and wealth of image data that can be processed using SealNet software contributes a vital tool for ecological and behavioral studies of marine mammals in the emerging field of conservation technology.

SEALNET: Facial recognition software for ecological studies of harbor seals

Zach Birenbaum[1,*], Hieu Do[1,2,*], Lauren Horstmyer[3], Hailey Orff[3], Krista Ingram[3,#], Ahmet Ay[2,,3,#]

Running title: *A Complete Pipeline for Facial Recognition of Harbor Seals*

[1] Colgate University, Department of Computer Science, 13 Oak Dr., Hamilton, NY, USA

[2] Colgate University, Department of Mathematics, 13 Oak Dr., Hamilton, NY, USA

[3] Colgate University, Department of Biology, 13 Oak Dr., Hamilton, NY, USA

**Equal Contribution*

#*Corresponding authors*

Email:

Krista Ingram: kingram@colgate.edu

Ahmet Ay: aay@colgate.edu

ABSTRACT

Methods for long-term monitoring of coastal species such as harbor seals, are often costly, time-consuming, and highly invasive, underscoring the need for improved techniques for data collection and analysis. Here, we propose the use of automated facial recognition technology for identification of individual seals and demonstrate its utility in ecological and population studies. We created a software package, SealNet, that automates photo identification of seals, using a graphical user interface (GUI) software to identify, align and chip seal faces from photographs and a deep convolutional neural network (CNN) suitable for small datasets (e.g., 100 seals with five photos per seal). We piloted the SealNet technology with a population of harbor seals located within Casco Bay on the coast of Maine, USA. Across two-years of sampling, 2019 and 2020, at seven haul-out sites in Middle Bay, we processed 1529 images representing 408 individual seals and achieved 88% (93%) rank-1 accuracy in closed set (open set) seal identification. We identified four seals that were photographed in both years at neighboring haul-out sites, suggesting that some harbor seals exhibit site fidelity within local bays across years, and that there may be evidence of spatial connectivity among haul-out sites. Using capture-mark-recapture (CMR) calculations, we obtained a rough preliminary population estimate of 4386 seals in the Middle Bay area. SealNet software outperformed a similar face recognition method developed for primates, PrimNet, in identifying seals following training on our seal dataset. The ease and wealth of image data that can be processed using SealNet software contributes a vital tool for ecological and behavioral studies of marine mammals in the emerging field of conservation technology.

INTRODUCTION

Marine ecosystems are dynamic, and conservation of key species that inhabit these ecosystems requires long-term monitoring of populations across a range of temporal and geographic scales Methods for long-term monitoring of coastal species, including harbor seals, are often invasive, costly, and time-consuming (Cunningham 2009), underscoring the need for new techniques for systematic data collection and analysis. The automation of these processes can be an effective technique for monitoring population dynamics, as automation increases reproducibility while decreasing cost and labor (.

Harbor seals are an ideal species for long-term monitoring as they are highly mobile animals that inhabit a large geographic range and are ecologically and economically important as top predators (Aarts et al., 2019). In addition, harbor seals can be observed non-invasively as they congregate at "haul-out" sites—essential areas where seals come out of the water to rest on rocky islets, allowing them to thermoregulate and avoid predation—which make them easily visible to researchers from afar (. As top predators, seal populations effect ecosystem dynamics, with healthy populations likely decreasing competition among species such as flounder, sole and dab, and, in turn, influencing the balance of both ecologically and economically critical fish populations . Increases in seal populations along the Atlantic coast have also increased the numbers of sharks that inhabit coastal waters, potentially affecting tourism revenue in addition to local ecosystems .

Harbor seals are important indicators of ecosystem health because they are susceptible to climate change and, given their extensive overlap with human activities both in and out of the water, are particularly vulnerable to increased anthropogenic activity (Allen et al., 1984). Over the last century, the Atlantic coast populations of harbor seals in northeastern North America have a history of heavy exploitation. Following the Marine Mammal Protection Act of 1972, populations of harbor seals off the Northeast coast of the U.S, successfully rebounded to healthy population numbers, but the steep decline in abundance prior to any legislation is evidence of the potential vulnerability of the population to acute or chronic ecological challenges.

As key regulators and indicators of ecosystem health, monitoring harbor seal population levels and movement patterns is essential. Tagging methods have been widely used in the past, however these GPS-monitoring devices are expensive, ranging from $1000 to $3000 for one device (*GPS and VHF Tracking Collars Used for Wildlife Monitoring* , 2017). In addition, the attachment of external devices may interfere with behaviors such as swimming speed, oxygen consumption, and metabolic rate, potentially corrupting the data collected or harming or disturbing the individual . Aerial and visual observation methods limit interference with seal behavior, but both techniques are time consuming and expensive . Photo based identification techniques also have the advantage of being non-invasive, but manual interpretation of photographs is time-intensive and often limited to small-scale projects. For seals, manual matching based on fur colors is difficult due to changing coat colors as seals mature and during annual molting. However, some promising progress has been made using analysis of pelage markings, i.e. spots on the seal's coat that can be reliably used as diagnostic tools (Cunningham, 2009).

Here, we propose the use of automated facial recognition technology as a system for identification of seal individuals for ecological and population studies. We used deep learning methods and convolutional neural networks to develop SealNet, a redesign of the PrimNet software (Deb et al., 2018) developed for primates. CNN-based facial recognition software achieves identification accuracies of 93.8% with lemurs , 92.5% with chimpanzees , and 97.27% with pandas . Another software, BearID, recently achieved close to 100% face chipping accuracy (number of faces recognized in an unprocessed photo) despite an overall pipeline identification accuracy of 82.4% SealNet contributes a new software package to automate the process of seal identification for use by researchers in the field.

In this paper, we outline the creation of a graphical user interface (GUI), that allows the user to automatically identify, align and chip seal faces to facilitate the processing of raw data. Then, using a deep convolutional neural network (CNN) suitable for small datasets (e.g., 100 seals with five photos per seal), we developed a seal face recognition software. We trained and tested this software on a wild population of Atlantic harbor seals in Casco Bay, Maine, U.S.A. We compare the performance of SealNet with its predecessor PrimNet and show that SealNet outperforms this software in the prediction of harbor seal identities. In a time of rapid ecosystem changes, SealNet represents a new tool non-invasive tracking of seals for use in ecological and behavioral studies.

METHODS

Photographic Data Collection

In the summers of 2019 and 2020, we captured 2267 photos across seven haul-out sites around Casco Bay, Maine, U.S.A. (Seal Rock, Wilson Cove, Brandt Ledges, Mitchell Fields, Branning Ledge, Whaleboat, and Bustin's Ledge; see **Figure 1**). During a single visit to each site, we took photos for 30 minutes to one hour from a 22-foot Eastern motorboat equipped with a 90-horsepower engine, an open deck, and a low-profile console. All site visits occurred in the summer (moulting season) of each year, with exact dates dependent on weather and tides, as some of the sites are inaccessible at high tide. We used a Nikon COOLPIX P1000 digital camera with a 125x optical zoom, 16MP backside illuminated CMOS sensor, built-in NIKKOR lens, 35mm equivalent focal length range of 24–3000mm, and a 250x Dynamic Fine digital zoom which gives the user an equivalent focal length of 6000mm. The camera is also equipped with Nikon's Dual Detect Optical Vibration Reduction, which provides 5 stops of optical image stabilization to ensure clear telephoto shots. We photographed at a minimum distance of 54.9 m (60 yards) from haul-out sites with the engine in low throttle or off to create minimal disturbance to the seals. We took multiple photographs of each individual seal as the boat drifted past the site. For this preliminary study, we processed a total of 415 images in 2019 and 1114 images in 2020 across the seven locations in Casco Bay (**Table 1**).

Development of SealNet

Data Cleaning:

We manually processed the total number of photos in the database (>5000 images) to remove blurry photos, shots of sky or water, and duplicates. We cropped each photo in the condensed dataset (n=2267) to focus on the seal faces to minimize the amount of time the software takes to identify faces. Once photos have been cropped, they are ready to be viewed in the graphical user interface and analyzed with the face detection software (steps outlined in **Figure 2**).

Face Detection:

We created the graphical user interface (GUI) in C++ by modifying*imglab* tools for image annotation We trained the interface to detect seal faces, allowing for automated detection of all seal faces in each photo. In addition, the GUI allows for the option to manually select seal faces by drawing boxes around valid faces in the application. A valid seal face is determined based on the quality and clarity of the image, as well as the angle of the seal face to the camera. Invalid faces are those that are too blurry, not facing the camera, or are partially obstructed. Invalid faces are ignored by the software, as are regions of the image not marked as faces. Variations in illuminations, lighting, and other conditions can introduce noise to the data and impede analysis. We next converted the photos to grayscale to help the model learn based on physical features and color patterns rather than the colors, which also serves to reduce overfitting during training. After all photos were aligned and cropped, we manually grouped photos of the same seals into folders by individual. To train our face detector, we selected all seal faces from the 516 photos taken at Brandt Ledges on January 29$^{\text{th}}$, 2020.

Our imglab based face detection software is a CNN network which uses Max-Margin Object Detection (loss function. The first three layers of the network downsample the input images by 8 and output a feature map of 32 channels. This feature map will go through 4 more convolutional layers with batch normalization and Rectified Linear Unit (ReLU) as nonlinearity. The final output will only have 1 channel; a large value will indicate that the network has found an object at that location and vice versa.

Using the full 2020 dataset, we measured the accuracy of the model using 5-fold stratified cross-validation. Each strata (i.e, each location and date) was split into 5 sections. For each fold, 4 of the 5 sections were chosen as a training set while the remaining section was used as a validation set. For each fold, the training set contained ˜413 photos from all 5 locations, and the validation set contained ˜103 photos from the same 5 locations. The accuracy of the face detector is measured by two metrics: precision and recall (**Figure 3**).

Face Alignment and Chipping:

Face alignment is critical for the accuracy of our face recognition software. As a result, prior to chipping the individual seal faces, we aligned them using the manually tagged eye locations in each photo by performing in-plane rotation to align the eyes along the x-axis. Once the eyes are manually located in each photo, the GUI automatically aligns and chips the faces to the desired size (e.g., 112x112 pixels). We followed an approach similar to that used by the developers of LemurFaceID to align faces: Given $(l_x,\ l_y)$ and$(r_x,\ r_y)$ to be the center of the left and right eyes respectively, one can calculate the rotation matrix M to be used in an affine transformation of the image. Let$x = \frac{l_x\ +\ r_x}{2}andy = \ \frac{l_y\ +\ r_y}{2}$ and$\theta = atan(\frac{r_{\text{y}}\ -\ l_y}{r_x - l_x})$, so$(x,\ y)$ will be the location of the midpoint between the centers of the two eyes and θ be the rotational angle. Then M will be calculated as:

$$M = \begin{bmatrix} cos\theta & -sin\theta & x(1 - cos\theta) + ysin\theta \\ sin\theta & cos\theta & y(1 - cos\theta) - xsin\theta \end{bmatrix}$$

Inter-pupil Distance (IPD) is the distance between the center of the two eyes, or$IPD = \sqrt{(r_x - l_x)^2 + (r_{\text{y}}\ -\ l_y)^2}$. We scaled each image automatically so that each eye would be$0.5 \times \text{IPD}$ away from the closest side edge and$0.6 \times \text{IPD}$ away from the top edge of the cropped face image. We chose these values by sampling 30 seal images and determining the optimum face to background ratio for facial recognition.

Thus, at the end of this step, each face image was rotated and resized to 112x112 pixels in preparation for facial recognition. The name of each of these images will contain information about its original image and the location within the original image from which it was chipped.

SealNet Architecture and Training:

The CNN-based face recognition classifier is the main component of our software package. We train this classifier with photos that have been aligned, chipped, and normalized. We trained the CNN for 100 epochs, using mini-batch gradient descent and a batch size of 16. We started with a learning rate of 0.01 and used

ADADELTA optimizer for our gradient descent. Each input image underwent four convolutional blocks and a final bottleneck layer to output an embedded vector of length 512 that contained learned features of the input image (**Figure 4**). Each convolutional block contained a convolutional layer with a kernel size of 3x3 and a stride of 1 followed by a max-pooling layer with a kernel size of 3x3 and a stride of 2. These blocks may contain an additional Squeeze and Excitation (SE) block that performs feature recalibration. The addition of SE blocks to our CNN helped the model better learn the interdependencies between channels, highlighting informative features while disregarding unimportant features . As proposed by PrimNet, our convolutional block also employs group convolutions followed by channel shuffling to make the network sparser to reduce vulnerability to overfitting .

SealNet was trained using a GeForce RTX 2080 TI graphics card that took about 4 minutes and 33 seconds on average to finish training each fold. We trained the model on an average of 485 images of the same resolution for each fold of the closed-set and on 533 images with dimensions (112, 112) for the open-set.

Validation of SealNet

In this biometric system, the probe set refers to the collection of biometric identities to be recognized, while the gallery set refers to identities that have been previously enrolled into the system. The gallery set acts as a database from which each probe identity will be searched. We measured the accuracy of SealNet with two standard recognition tasks: closed-set and open-set identification. Both closed-set and open-set refer to 1:N matching scenarios where each identity in the probe set will be searched against multiple identities in the gallery. The SealNet face recognition software produces a similarity score for each probe-gallery pair and the result will be sorted in descending order so that the identity with the highest score will be the most likely matched candidate. In closed-set identification, it is guaranteed that the identity in the probe is present in the gallery; whereas in open-set identification, it is uncertain whether that is the case.

We validated SealNet's face recognition capabilities using k-fold cross-validation with k = 5. That is, we divide the training data into 5 sections/folds, and at some point, use each section as a test set. We filtered the data to only include seals with more photos than the number of folds.

For closed-set identification, in each fold, we summarized the accuracy of our system using a Cumulative Match Characteristic (CMC) curve which plots the True Positive Identification Rate (TPIR) against the ranking of seals. TPIR measures the probability of observing a correct match within each rank. A correct match between probe p and an identity$g$ in the gallery has rank k if the similarity score between p and g is the k th largest score .

For open-set identification, prior to splitting the dataset into 5 folds, we randomly select half of the seals with enough photos to be eligible for training, and put them, along with all seals lacking sufficient data to be included in training, into each of the testing sets as new seals. This method of training exclusions provided the best balance between the quantities of open-set testing photos and training photos. Whenever a probe-gallery pair's similarity score exceeded our acceptance threshold, we "accepted" that individual; i.e., marked it as having been seen before (i.e the probe has a match in the gallery), while any probe with a similarity score tha was less than the threshold value was rejected as an imposter.

Using the information on whether the probe was truly an imposter or not, all probes were categorized as follows: True Positives (TP) scored above the threshold and correct match was predicted within top "Rank" similarity scores. False Positives (FP) scored above the threshold but had no true match in gallery. False Negatives (FN) contained a match in gallery but had a top similarity score below the threshold, or the correct prediction for gallery member was not within the top "Rank" similarity scores. True Negatives (TN) had no match in the gallery and top predicted match had a similarity score below the threshold. Accuracy is measured as the ratio between the sum of TP and TN over all queries, which is equivalent to $\frac{TP+TN}{TP+TN+FP+FN}$. This formula is identical for open and closed set, but since the closed set inherently has no 'True Negatives' or 'False Positives', the closed set accuracy computation can be simplified to $\frac{\text{TP}}{TP+FN}$.

Comparison of SealNet with PrimNet

To see how well our software performed compared to a previously developed facial recognition software, PrimNet, we trained and tested it and SealNet models using the same data and parameters. To further ensure fairness, we tested the Rank-1 F1-Score results for each model at all threshold values in 0.01 increments and present the values for the run with the highest score. The Rank-5 scores presented use the same threshold as the best performing rank one, with loosened constraints for being classified as a True Positive. We used F1-Scores as there were only 74 in-set seal photos as opposed to the 571 photos with no corresponding seal in the gallery. Because F1-Score provides a better measure of propensity for incorrect classifications than accuracy it is more applicable to imbalanced datasets like ours. Baseline accuracy is the accuracy score of the model assuming all probes were rejected. TPR, or true positive rate, is the most intuitive measure of model performance, and shows the proportion of correctly classified probes at a given threshold.

Training of SealNet with New Data

We also measured how SealNet performs as we increase the size of our seal database (gallery set), for instance by adding seals from new locations or new dates. Each time new data was added, we evaluated the model's closed-set accuracy by running 5-fold cross validation and calculating its average true identification rate.

Probing of the dataset with new individuals

After adding new data into their respective folders of either probe or gallery images, we ran the RecognitionGUI software, allowing for predictions of individuals. The RecognitionGUI software provides scores for each individual in the gallery, assigning them a number based on their facial biometric similarities to the individual in the probe. Similarity scores are provided for the top five ranked matches; top matches are confirmed by visual check in the output graphic.

To add new individuals to the gallery database, a 'match' probe photo can be automatically merged to its matching file. If the new individual does not match previous individuals in the dataset, a novel ID is created, and the seal photos and ID are added to the existing database. To speed up this process, we created a quick Python program, Tkinter GUI .

RESULTS

Automatic Face Detection

We found that SealNet's face detector has precision (the percentage of predictions that are seal face) of 85.43% and a recall (the percentage of total seal faces that are correctly predicted) of 86.94% after being trained on a dataset of 516 photos from one haul-out site on a single day that contained 1,178 valid seal faces. **Figure 3** shows the accuracy of our model across different threshold levels for predicting a seal face. As the value of threshold decreases, the precision decreases to 0 while the recall approaches to 1. On the other hand, if threshold increases, the precision increases to 1 but the recall will decrease to 0. We chose threshold 0 for our face detector because it gives the best precision-recall trade-off.

We detected 49 false positives, that is, faces detected by SealNet that were not faces. Most were caused by vegetation or other parts of the seal that had face-like shapes (**Supplementary Figure 1**). SealNet missed on average 43 faces, mostly ones that were angled away from the camera (false negatives, **Supplementary Figure 2**). We identified a total 408 unique seals, with an average of 2.9 photos per). Among these, 74 seals appeared in at least 5 photos.

Accuracy in Seal Identification

Our closed set data contained 74 seals that had at least 5 photos (607 photos in total). For each fold, the testing set contains one-fifth of the number of photos of each of the 74 seals and the training set will contain the remaining photos of those seals. We trained and tested both PrimNet and SealNet on the same data for each fold. Our average rank-1 accuracy was 88% and our average rank-5 accuracy was 96% across 5 -folds (**Figure 5**). PrimNet yielded 70% rank-1 accuracy and 91% rank-5 accuracy on the same dataset.

Our open set data also included 74 seals with at least 5 photos and 571 photos from seals with fewer than

5 photos. Both PrimNet and SealNet models were trained and tested utilizing the same splits of data and equivalent parameters for number of epochs and batches per epoch to ensure fairness. SealNet outperformed PrimNet with an accuracy that is 9% better at correctly classifying Rank 5 and Rank 1 matches for probes, respectively (**Table 2**). F1 scores, a measure of model performance for unbalanced datasets, showed a similar result with SealNet performing 37% and 39% better than PrimNet.

SealNet's Performance on a Growing Dataset

We performed expansion of our data five times, each time adding a new folder to the probe, so our database increases from 194 to 406 unique seals. We calculated the average accuracy of both rank-1 and rank-5 training runs as shown in **Table 3** . Our data suggests that our model performs consistently (at the same accuracy level) as the size of our dataset increased.

Semiautomatic Expansion of the Dataset

Our tkinter GUI allows for direct comparison of our existing seal database with new probed individuals, supported by similarity scores of the individuals in the gallery to the current probe (**Figure 4**). These similarity scores are based on facial biometrics of each individual, and they are determined after each image within a seal individual folder is compared to each image found in a gallery seal individual folder. On average, the similarity scores for individuals that are a match with the probe individual are approximately 0.65 with a SD of 0.1 (range 0.55-0.85). Similarity scores with individuals without a match range in similarity scores from 0.2-0.5.

Ecological Results

SealNet identified four individual seals that were photographed in both 2019 and 2020: 015_Armani, 198_Petal, 211_Clove, and 393_Cystine. All four seals were originally photographed on Brandt Ledges in 2019 and were re-photographed on Mitchell Fields (198_Petal and 211_Clove), Whaleboat (393_Cystine), or Branning Ledges (015_Armani) during the 2020 season. Our mark-recapture estimate based on our total number of seals photographed in both 2019 and 2020, with four individuals "recaptured," reveals an approximate initial population size estimate of 4,386 individuals in the Middle Bay region of Casco Bay.

DISCUSSION

Here we present the utility of a new software package, SealNet, a complete, automated pipeline to non-invasively identify individual seals in photographic images. We describe a novel face detector GUI trained to detect harbor seal faces and the development of a new neural network to recognize individual seal faces. Following validation of the technique, we use SealNet in a preliminary study to explore site-fidelity of harbor seals in the Casco Bay, Maine region of the northwestern Atlantic coast. Our initial validation analyses confirm the efficiency and accuracy of our facial recognition technology in the photo identification of an economically important coastal marine mammal.

Our trained face detector had a precision of 85%, and a recall of 87% despite having very little restrictions on the position of the seals within the photo with the only limitation that both eyes of the seals are visible. This feature enables the successful use of SealNet software to conduct studies on animals in the wild without having the animals looking directly at the camera. The precision and recall of our detector could be increased by restricting the possible angle and pose of the seal, but this would limit the number of photos that meet such requirements in field studies.

In a direct performance comparison of the recognition task, SealNet performs better than PrimNet on average at all ranks with up to 18% improved classification accuracy at rank-1 for closed-set and 9% for open-set identification with SealNet. It is also important to note that our model performs consistently well as our database increases in size. The consistent performance of our model demonstrates that SealNet generalizes well (i.e., overfitting is not an issue).

For our recognition software, we have achieved a remarkable accuracy in both close-set (rank-1: 88% and rank-5: 96%) and open-set (rank-1 and rank-5: 93%), but there is still room for improvement. FaceNet which was trained on more than 3 million images of almost 10,000 unique human individuals, achieved an accuracy of almost 100%. Other biometric models for chimpanzees and lemurs also yield an accuracy around 92%–94%. Therefore, with a larger dataset with more photos per seal individual, it is possible that we can further improve our accuracy.

Preliminary ecological results

Using the SealNet facial recognition software package and a small, initial dataset sampled across two years (2019 and 2020) in Casco Bay, we identified four individuals in the datasets from both years, indicating a small degree of local site fidelity across years during the months of June and July. All four seals were found on the haul-out site, Brandt Ledges, in 2019. In 2020, the four seals were photographed again within 1-3 nautical miles of Brandt Ledges: one was photographed on Branning Ledges, two were photographed at the Mitchell Field site, and one was photographed on the Whaleboat Island site. This result supports previous results suggesting site fidelity among harbor seals off the coast of NE Scotland . It is also interesting to note that two of the individuals found in the dataset from both years, Clove and Petal, were found together initially on Brandt Ledges on one day in 2019, and then found together again in 2020 at the Mitchell Field site. These results suggest that SealNet software may be useful in future long-term studies of social relationships in harbor seals. Previous studies have examined competitive relationships among harbor seals , however further research is needed to examine other questions related to social behavior, including social fidelity, persistence of family groups, and other social dynamics.

Our preliminary ecological results suggest some site-fidelity of harbor seals in Middle Bay as well as site-fidelity to neighboring haul-out sites within the bay. A more extensive ecological study is underway to determine the degree of site fidelity and spatial connectivity of haul-out sites in this region. In addition, our results provide an estimated population size of approximately four thousand seals utilizing Middle Bay, although more extensive photographic data will help refine this population estimate. Current estimates of harbor seal abundance are outdated, suggesting a population of 38,014 individuals in the whole of Maine in 2001 (followed by an aerial survey done in 2012 determining a population of 75,834 individuals (). Accurate local and regional population estimates are imperative to understanding dynamics of seal abundance in relationship to anthropomorphic and climate changes to coastal marine environments as well as the impact of an increasing great white shark population.

The use of facial recognition software to identify individuals in wild populations is a relatively new area of research being primarily utilized in studies of land mammals such as lemurs and brown bears (Crouse et al., 2017). Our research extends the use of such methods to marine mammal species. Facial biometrics are not the only measure that can be used for automated identification of seals. For example, a recent, groundbreaking study utilized pelage markings found on the seals coat to identify grey seal individuals near Wales (Langley et al., 2020). Given that coat patterns may change across seasons during molting or over time in harbor seals, facial biometrics may offer and additional and/or more reliable method of identification. Thus, developing facial recognition techniques for harbor seals allows for a rapid, non-invasive means for detailed study of an economically and ecologically important species. Importantly, researchers can customize the software and the GUI to suit their own needs at each step of data collection—training the face detector for additional species, modifying the alignment procedure, or preprocessing images for face recognition.

Limitations

Although SealNet produced promising results, there are still limitations that need to be addressed. First, our SealNet software still requires some manual work during the data collection process—after running the automatic face detector, researchers are still required to manually locate the eyes, nose and mouth in order for the program to automatically align and chip the seal faces. Thus, one possible improvement that we can implement in the future is to add a landmark detector to be used in conjunction with the face detector. Secondly, to generate training data, researchers must manually group multiple face chips belonging to the same individual. Not only is this process laborious, it may be also error-prone. A more sustainable approach would be to implement a classifier; however, researchers would still be required to manually check if the classification is accurate.

Although SealNet does well in closed-set classification, open-set verification performance could be dramatically improved by reducing the similarity scores between such seals. This success could be achieved in two ways. First, changes in preprocessing methodology could yield greater differentiation between very similar seals by removing lighting and other environmental effects without affecting color pattern information the way grayscaling does. Second, changes to our model architecture could improve such performance. However, the inherent complexity in any attempt to leverage specificity, while simultaneously avoiding overfitting, presents a difficult balance which all recognition models struggle to strike. Thus, the best approach to this problem would be to maximize the quantity and quality of information available to the model through preprocessing improvements prior to making changes to the CNN architecture itself.

Conclusion

The use of SealNet facial recognition software to identify individual harbor seals has multiple future applications to aid in decision-making for conservation efforts, including assessments of seal abundance, evaluation of site fidelity within and across coastal regions, determination of trends in migration patterns, and the exploration of patterns in social behavior among harbor seals at haul-out sites. The ease and wealth of data that can be collected with non-invasive photography, coupled with the predictive ability of the SealNet to identify individuals, provides researchers with a robust toolkit that has the potential to transform ecological studies of wild populations of harbor seals.

DATA AVAILABILITY STATEMENT

SealNet is an open-source application available on GitHub at https://github.com/zbirenbaum/SealFaceRecognition. The code and models are both also archived at (DRYAD Accession number will be added following manuscript acceptance). Owing to file sizes, raw images will only be available upon request to the authors.

COMPETING INTERESTS STATEMENT

The author(s) declare no competing interests.

AUTHOR CONTRIBUTIONS

KKI and AA conceptualized the study; AA, ZB & HD completed the face recognition methodology and LH, HD and KKI completed the ecological methodology; data curation and visualization was completed by LH, HD, and ZB; Writing-original draft was completed by AA, ZB, HD, LH and KKI; Writing-review & editing was completed by KKI.

ACKNOWLEDGMENTS

We would like to acknowledge the assistance of Rebecca Gowen, Daniel Jaris and Nick Knight in the collection of seal images in Casco Bay.

REFERENCES

Year	**Date**	**Location**	**Total # Seals (# Chips)**	**Unique # Seals (# IDs)**
2019	7/16	Brandt Ledges	50	11
	7/20	Brandt Ledges	56	17
	7/24	Seal Rock	45	8
	7/27	Wilson Cove	19	4
	7/30	Bustin's Ledge	15	3
2020	7/1	Whaleboat	39	8
	7/10	Branning Ledges	820	197
	7/28	Whaleboat	33	8
	7/29	Branning Ledges	254	65

Year	Date	Location	Total # Seals (# Chips)	Unique # Seals (# IDs)
	7/31	Mitchell Fields	434	127

Table 1. Dataset summary. The number of chips is the number of faces recognized by the facial recognition software. The number of IDs represents the number of individuals identified after grouping the chips according to individual.

		TPR	FPR	FNR	TNR	Baseline Accuracy	Accuracy	Precision	F1-Score
SealNet	**R1**	0.6486	0.0434	0.0355	0.9566	0.8695	0.9286	0.6000	0.6234
	R5	0.6757	0.0434	0.0329	0.9566	0.8695	0.9310	0.6098	0.6410
PrimNet	**R1**	0.2703	0.1030	0.0754	0.8970	0.8153	0.8399	0.2083	0.2353
	R5	0.3243	0.1030	0.0702	0.8970	0.8153	0.8448	0.2400	0.2759
Difference	**R1**	0.3784	-0.0596	-0.0399	0.0596	0.0542	0.0887	0.3917	0.3881
	R5	0.3514	-0.0596	-0.0373	0.0596	0.0542	0.0862	0.3698	0.3652

Table 2. Comparison of open-set performance between SealNet and PrimNet for key metrics of model evaluation. In open-set evaluation, any probe with a similarity score for its best match in the gallery less than the value of the threshold was rejected as an 'imposter'. True Positives scored above the threshold and correct match was predicted within top "Rank" similarity scores (TPR). False Positives scored above the threshold but had no true match in gallery (FPR). False Negatives contained a match in gallery but had a top similarity score below the threshold, or the correct prediction for gallery member was not within the top "Rank" similarity scores (FNR). True Negatives had no match in the gallery and top predicted match had a similarity score below the threshold (TNR). Baseline accuracy is the accuracy score of the model assuming all probes were rejected. F1-Score provides a better measure of propensity for incorrect classifications than accuracy, suited to unbalanced datasets.

# Seals	Rank 1 Accuracy
0.85	0.9498
0.8708	0.9664
0.8652	0.9684
0.8556	0.9555
0.8714	0.9582

Table 3. Iterative training accuracy for closed-set data. The average rank-1 and rank-5 accuracy levels for each iterative training run following the probing of individuals from each date during 2020; accuracies are relatively robust to numbers of individuals.

Hosted file

`Seals_Fig 1.docx` available at https://authorea.com/users/454247/articles/551919-sealnet-facial-recognition-software-for-ecological-studies-of-harbor-seals

Hosted file

`Seals_Fig 2.docx` available at https://authorea.com/users/454247/articles/551919-sealnet-facial-recognition-software-for-ecological-studies-of-harbor-seals

Hosted file

Seals_Fig 3.docx available at https://authorea.com/users/454247/articles/551919-sealnet-facial-recognition-software-for-ecological-studies-of-harbor-seals

Hosted file

Seals_Fig 4.docx available at https://authorea.com/users/454247/articles/551919-sealnet-facial-recognition-software-for-ecological-studies-of-harbor-seals

Hosted file

Seals_Fig 5.docx available at https://authorea.com/users/454247/articles/551919-sealnet-facial-recognition-software-for-ecological-studies-of-harbor-seals