Matching individual Ladoga ringed seals across short-term image sequences

Nepovinnykh Ekaterina, Chelak Ilia, Lushpanov Andrei, Eerola Tuomas, Kälviäinen Heikki, Chirkova Olga

DOI: 10.1007/s42991-022-00229-3

Mammalian Biology
https://doi.org/10.1007/s42991-022-00229-3

METHODOLOGY

Matching individual Ladoga ringed seals across short-term image sequences

Ekaterina Nepovinnykh[1] · Ilia Chelak[2] · Andrei Lushpanov[1,3] · Tuomas Eerola[1] · Heikki Kälviäinen[1] · Olga Chirkova[4]

Received: 17 March 2021 / Accepted: 7 January 2022

Abstract
Automated wildlife reidentification has attracted increasing attention in recent years as it provides a non-invasive tool to identify and to track individual wild animals over time. In this paper, the first steps are taken towards the automatic photo-identification of the Ladoga ringed seals (*Pusa hispida ladogensis*). A method is proposed that takes a sequence of images, each containing multiple individuals as the input, and produces cropped images of seals grouped based on one certain individual per group. The method starts by detecting each seal from the images and proceeds to matching the individual seals between the images. It is shown that high grouping accuracy can be obtained with a general-purpose image retrieval method on an image sequence taken from the same location within a relatively short period of time. Each resulting group contains multiple images of one individual with slightly different variations, for example, in pose and illumination. Utilizing these images simultaneously provides more information for the individual re-identification compared to the traditional approach, i.e., which utilizes just one image at a time. It is further demonstrated that a convolutional neural network based method can be used to extract the unique pelage patterns of the seals despite the low contrast. Finally, a method is proposed and experiments with the novel Ladoga ringed seals data are carried out to provide a proof-of-concept for the individual re-identification.

Keywords Animal re-identification · Convolutional neural networks · Instance segmentation · Ladoga ringed seal · Photo-identification

Handling editors: Daniel I. Rubenstein and Stephen C.Y. Chan.

This article is a contribution to the special issue on "Individual Identification and Photographic Techniques in Mammalian Ecological and Behavioural Research – Part 1: Methods and Concepts" — Editors: Leszek Karczmarski, Stephen C.Y. Chan, Daniel I. Rubenstein, Scott Y.S. Chui and Elissa Z. Cameron.

✉ Ekaterina Nepovinnykh
ekaterina.nepovinnykh@lut.fi

Ilia Chelak
chelak.ilrost@gmail.com

Andrei Lushpanov
andrew9691@yandex.ru

Tuomas Eerola
tuomas.eerola@lut.fi

Heikki Kälviäinen
heikki.kalviainen@lut.fi

Olga Chirkova
chirkovaolga.spb@gmail.com

[1] Computer Vision and Pattern Recognition Laboratory, Department of Computational Engineering, School of Engineering Science, Lappeenranta-Lahti University of Technology LUT, P.O.Box 20, 53851 Lappeenranta, Finland

[2] Department of Artificial Intelligence Institute of Computer Science and Technology, Peter the Great St. Petersburg Polytechnic University, Polytechnicheskaya, 29, Saint Petersburg, Russian Federation 195251

[3] Department of Computer Science and Computational Experiment, Southern Federal University, Rostov-on-Don, Russian Federation 344006

[4] Interregional Charitable Public Organization "Biologists for Nature Conservation" (BFNC), 24 line 3-7, Saint Petersburg, Russian Federation 199106

Introduction

Ladoga ringed seals (*Pusa hispida ladogensis*) are a vulnerable subspecies of the ringed seals only found in Lake Ladoga, Russian Federation (Fig. A1). According to recent

Published online: 08 April 2022

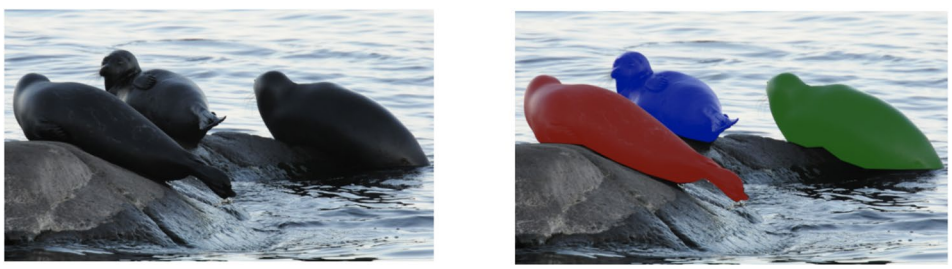
studies, around 5500–8000 seals inhabit the lake (Trukhanova et al. 2013; Trukhanova 2013). The landlocked population faces various threats associated with fishing bycatch, the industrialization of the area, and climate change motivating the monitoring of the population. Despite their close phylogenetic proximity, the Ladoga ringed seals are considerably less studied than its sister population, called the Saimaa ringed seals (*Pusa hispida saimensis*) found in Lake Saimaa, Finland (Kunnasranta et al. 2021). However, recently, the first efforts to employ wildlife photo-identification techniques to study the Ladoga ringed seals have been initiated.

Automated wildlife photo-identification has gained prominent attention as a potential tool to monitor animal populations in a non-invasive manner. The basic idea is to utilize computer vision techniques to automatically analyze large volumes of image data, to identify the individual animals in the images, and in this way, produce useful information on population processes and attributes such as survival, dispersal, site fidelity, reproduction, health, as well as size and density. Individual identification can be based on distinctive permanent characteristics visible in images, such as fur, feather, or skin patterns, scars, or shape. The Ladoga ringed seals have a pelage pattern that is unique to each seal, enabling the identification of the individuals over their whole lifetime.

This study is based on the earlier works on the automatic photo-identification of the Saimaa ringed seals (*Pusa hispida saimensis*) (Zhelezniakov et al. 2015; Chehrsimin et al. 2018; Nepovinnykh et al. 2018, 2020; Chelak et al. 2021). Despite being the sister populations, there are two major differences that make the photo-identification of the Ladoga ringed seals more challenging: (1) the pelage pattern of the Ladoga ringed seals has low contrast which makes it hard to extract the necessary features for identification, and (2) the Ladoga ringed seals are more social, and therefore, images often contain large number of individuals. Images of the Saimaa ringed seals typically contain only one animal so the detection (segmentation) step can be formulated as a binary classification task for pixels (the seal and the background) (Zhelezniakov et al. 2015; Nepovinnykh et al. 2020). The Ladoga ringed seals, on the other hand, require a method that is able to detect and delineate each instance of a seal in the image separately (see Fig. 1).

In this paper, the above challenges are tackled by proposing a method to process and analyze sequences of Ladoga ringed seal images. First, the seal instance segmentation method, Mask R-CNN, of He et al. (2017) is utilized to detect and segment each seal in an image. After each instance has been cropped, it is matched sequentially with instances contained in other images. As a result, a set of image groups is obtained, each corresponding to one individual and containing multiple images with varying pose, illumination, and quality (see Fig. 2). These groups can then be used for the re-identification of the seal individual and, matching the individual with images in a database of the known individuals. Utilizing multiple images of the individual has the potential to improve the accuracy of the re-identification as compared to traditional methods that utilize only one image at a time. The re-identification algorithm can aggregate more information about the pattern. For example, if the seal turns around new parts of the pattern might become visible, thus improving the chances of finding a match to images in the database. In this way, it is possible to extend expand the database with previously unseen parts of the seal's coat pattern.

The image sequences considered in this study consist of sequential images obtained using game cameras (Scout Guard, UVision, and Atl Acorn models) or sets of images from the same group of seals captured using DSLR or other handheld cameras (the model Pentax K5 Vivitar 400mm f5.6) (Gromov et al. 2021). This means that the images in one sequence are obtained from the same location within a short period of time leading to relatively small variations in the appearance of the seals on the consecutive images. This makes it possible to use a general-purpose image retrieval method for individual matching (grouping). A convolutional neural network (CNN) based method with the Generalized-Mean (GeM) pooling layer (Radenović et al. 2016, 2019) is proposed for the task. Moreover, a CNN based method for pelage pattern extraction using the well-known U-Net encoder–decoder architecture (Ronneberger et al. 2015) is



(a) (b)

Fig. 1 Instance segmentation: **a** Original image; **b** Segmented image

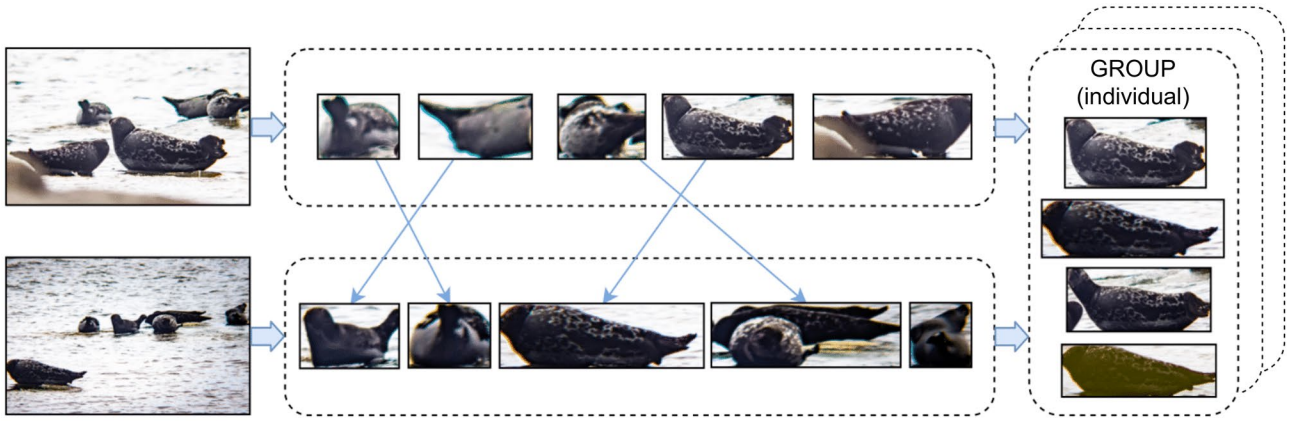


Fig. 2 Ladoga ringed seal individual grouping

employed. Finally, the re-identification part is solved using CNN-extracted pattern features, that are aggregated into Fisher vectors (Perronnin and Dance 2007; Perronnin et al. 2010a, b) that generate an image descriptor. As a result, a full framework for automated photo-identification of the Ladoga ringed seals is obtained.

Related work

Animal detection and instance segmentation

The first step of a typical automatic re-identification pipeline begins with animal detection. Detection might be done in several different ways. A general classification might be used to determine whether an object is present in an image. Localization might be used to return the spatial location of an object, for example, with a bounding box. Semantic segmentation is used to classify each pixel in the image individually. Instance segmentation similarly classifies each pixel. However, it is also able to separate individual instances of each class, which is especially important for the re-identification task.

Currently, methods based on Convolutional Neural Network (CNN) have become the standard for the detection tasks (Liu et al. 2020). The methods can be roughly divided into one-stage and two-stage frameworks.

Two-stage frameworks, such as R-CNN (Girshick et al. 2014), Fast R-CNN (Girshick 2015), Mask R-CNN (He et al. 2017) first generate region proposals, and then apply a classifier to those regions. In the case of Mask R-CNN (He et al. 2017), a fully connected network is used for the instance segmentation. An extension of this idea to a larger number of stages (cascades) should produce state-of-the-art results in instance segmentation (Chen et al. 2019).

One-stage frameworks, such as YOLO (Redmon et al. 2016), SSD (Liu et al. 2016) and CornerNet (Law and Deng 2020) use a single end-to-end network to perform object detection. Even though two-stage frameworks have higher accuracy (Liu et al. 2020), one-stage frameworks are simpler and easier to train than two-stage frameworks, making them more suitable for mobile devices and real-time applications.

While the general-purpose detection methods described above might be used for animal detection as well, sometimes a more specialized approach might be necessary. Many early animal detection methods were based on face and head detection (Burghardt and Calić 2006; Zhang et al. 2011). Such methods are typically highly sensitive to the pose of the depicted animal which limits their applicability. Today, CNNs are widely applied to animal detection (Parham et al. 2018; Verma and Gupta 2018; Kellenberger et al. 2019). Zhelezniakov et al. (2015) justified the use of segmentation for animal re-identification for the case when the data are obtained using static game cameras since capturing a common background in each image might bias the machine learning training process of the re-identification algorithms. For images containing a single animal, semantic segmentation is enough. Nepovinnykh et al. (2020) proposed such method for the Saimaa ringed seal re-identification where a CNN-based DeepLab model (Chen et al. 2018) was utilized for the seal segmentation.

Automatic wildlife re-identification

The main task of wildlife re-identification when the query image contains only one animal, is to find the corresponding individual from a gallery set of the known individuals. In practice, this can be accomplished by determining whether the animals in two images (the query image and the gallery image) are the same individual. This can be done based on

characteristics unique to each individual such as fur patterns or tail shapes.

The WildBook (Berger-Wolf et al. 2017) project aims to help with conservation efforts using crowd-sources data and computer vision models. The Hotspotter (Crall et al. 2013) algorithm is included in the IBEIS (Image-Based Ecological Information System) (Berger-Wolf et al. 2015) which is a part of the WildBook project. Hotspotter is a species-agnostic re-identification algorithm for patterned species. It has been successfully applied to the re-identification of zebras (*Equus quagga*) and giraffes (*Giraffa tippelskirchi*) (Parham et al. 2017), whale sharks (*Rhincodon typus*) (Holmberg et al. 2009), hawksbill turtles (*Eretmochelys imbricata*) (Dunbar et al. 2021), leopards (*Prionailurus bengalensis euptilurus*) (Park et al. 2019), and burying beetles (*Nicrophorus*) (Quinby et al. 2021). The algorithm is based on the affine invariant keypoints (hot-spots) with RootSIFT (Lowe 2004; Arandjelović and Zisserman 2012) descriptors which are used to match (re-identify) a query image to the database images.

Recent advances in deep learning have popularized the use of CNNs also for animal re-identification (Bouma et al. 2018; Deb et al. 2018; Schneider et al. 2019; Moskvyak et al. 2020). Li et al. (2020) sought new solutions for the photo-identification of the Siberian tiger *(Panthera tigris tigris)* focusing on Amur Tiger Re-identification in the Wild Challenge. Various CNN architectures were proposed for solving the re-identification task on the dataset following the lead of others (Liu et al. 2019a, b; Cheng et al. 2020).

Image retrieval

Content-based image retrieval (CBIR) is a computer vision problem with the goal of understanding how to search and retrieve query images from a database based only on the visual content of the image (Smeulders et al. 2000). This task is similar to the animal re-identification task where the matching image is searched for from the database of the known individuals. The traditional image retrieval methods, such as Bag of Words (BOW) (Sivic and Zisserman 2003), Vector of Locally Aggregated Descriptors (VLAD) (Jégou et al. 2010) and Fisher vector (Perronnin and Dance 2007; Perronnin et al. 2010a, b), consist of three steps: extraction of the features, creation of the codebook, and image encoding.

The first step, feature extraction, can be done using traditional hand-crafted features such as Scale Invariant Feature Transform (SIFT) (Lowe 2004; Arandjelović and Zisserman 2012), even though CNN are also suitable (Mishchuk et al. 2018). The codebook is then created using the descriptors from the database, usually by applying a clustering algorithm with the number of clusters corresponding to a number of visual words. Based on the codebook, image features can be transformed to fixed-length vectors by encoding the relationship of the feature to the clusters. The vectors are then used to measure similarity between the images. The main difference among image retrieval methods is how they create a codebook and how they convert them to fixed-sized vectors for image representation. Finally, the similarity between the images is measured using distances between fixed-length vector representations for both the image in question and images in the database which are then ranked.

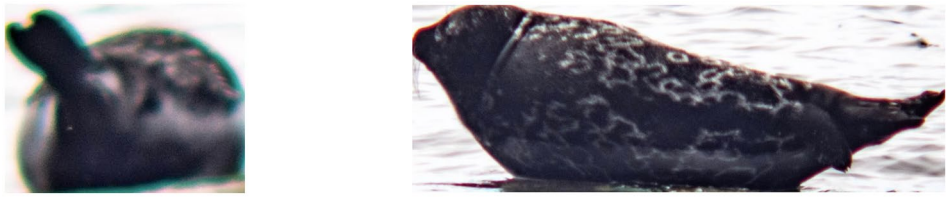
Due to the success of CNNs in different computer vision tasks, many CNN-based algorithms have been developed and applied to image retrieval tasks. The usual approach is to use a CNN to extract features, then apply specialized layers to construct a final encoding vector. For example, NetVLAD (Arandjelović et al. 2016) is a CNN inspired by the classical VLAD (Jégou et al. 2010) algorithm which uses a generalized VLAD layer to aggregate CNN-extracted features. The layer encodes cluster residuals in the same manner as the original VLAD, with the main modification being the change from hard to soft assignment to make it differentiable. This is necessary for the network to be trainable with gradient descent. Tolias et al. (2016) performed max-pooling over overlapping image regions to generate the final descriptor. The use of regions allows encoding spatial information, which is lost when pooling all features globally. Radenović et al. (2019) proposed to generalize global mean pooling to increase the influence of relevant features as follows:

$$\mathbf{f}^{(g)} = \left[f_1^{(g)} \dots f_k^{(g)} \dots f_K^{(g)}\right]^\top, \quad f_k^{(g)} = \left(\frac{1}{|\mathcal{X}_k|}\sum_{x\in\mathcal{X}_k} x^{p_k}\right)^{\frac{1}{p_k}}, \quad (1)$$

where $\mathcal{X}_k$ is the kth channel of the input and $f^{(g)}$ is the resulting pooled vector. The parameter p is responsible for how the features are selected. It can be treated as a network parameter and can be included in the learning process. Thus, by increasing the parameter p, it is possible to increase the impact of strong (relevant) features on the result.

The proposed method

The main problem in the re-identification of the Ladoga ringed seals is the fact that the most reliable way to identify a Ladoga seal is by analyzing its pelage pattern (Gromov et al. 2021) which often has low contrast. Due to poor image quality, seal pose, various obstructions, or lack of illumination, the pattern might be impossible to reliably segment or even miss from the image. Example images with and without recognizable patterns are presented in Fig. 3. However, by utilizing the information about the time and the place of taken pictures, it is possible to group individual seals within the series of images taken from one site within a relatively



(a) (b)

Fig. 3 Example images of the same individual. Both images were taken from the same site within a relatively small time window. Notice how only the second image contains an identifiable pattern

narrow time window. This is possible to do because seals are generally not a very mobile and rarely move far from the place initially sighted. This suggests that the background and the general visual similarity serve as good indicators of whether two seals are the same individual. This is why we propose a separate individual grouping step that could be used before the final re-identification.

The proposed framework for the Ladoga ringed seal re-identification is visualized in Fig. 4. Given a set of images obtained from the same group of individuals (usually images obtained on the same site within a given time window, usually a day), the seals in each image are first detected and cropped using an instance segmentation algorithm. The cropped images are then matched with others to obtain grouped sets of images each containing one uniquely identified individual. The fur pattern is then extracted from the cropped images. Segmentation masks obtained earlier are utilized to remove the background that could negatively affect the accuracy of pattern extraction. Images, where the pattern could not be extracted, are removed from further processing. Finally, all satisfactory pattern images in the group are used to identify the individual. This paper focuses mainly on the detection (instance segmentation) and grouping steps. However, pattern extraction and the final re-identification steps are also considered and discussed.

Seal instance segmentation

For the seal instance segmentation, Mask R-CNN (He et al. 2017) is used. Each seal image is cropped based on the bounding box coordinates and the segmentation masks are saved for later use.

For the backbone of the Mask R-CNN, two variations of the ResNet (He et al. 2016) architecture are used with 50 and 101 layers, respectively. Furthermore, three different modifications of the ResNet backbone are considered. Since the original ResNet was intended to be used mainly for classification, those modifications are necessary for applying the network to the segmentation task. The main difference is how they deal with different scales of objects, which is an essential part of the problem since the scale of seals on photos varies greatly. Three backbone variants are employed, of which the first two are taken from the original publication (He et al. 2017). In the first one, ResNet is combined with Feature Pyramid Network (FPN) (Lin et al. 2017) that uses lateral connections, thus generating a feature pyramid from a single-scale input within the network. In the second one, the original Faster R-CNN with ResNet features from the final layer of the fourth stage (C4) is used. In the third one, the features are extracted from the fifth stage with the dilated convolution (DC5) (Li et al. 2017).

Individual grouping

Individual grouping is performed for a sequence of images from the same location within a short period of time. Therefore, images can be expected to contain mostly the same group of individuals. Moreover, it can be assumed that consecutive images will contain relatively small variations in the seal pose and illumination. Thus, the same seal individual should exhibit a similar appearance in the different images of the sequence. This makes it possible to utilize general-purpose image retrieval approaches to find matching seal individuals among the images using visual similarity. However, when images are collected by photographers using handheld cameras, variation in time gaps, view angle and zoom can be large, rendering the tracking methods described above for camera traps less suitable for this particular task. For example, a photographer might randomly decide to zoom in on a particular seal or subgroup of seals, leaving other seals out of frame, then switching his focus to another group, and so on.

The cropped seal images (instances) obtained from the instance segmentation step each contain a single individual. The instances are cropped using their bounding boxes, meaning that at least a small piece of background information is present, which is important for extreme cases when the viewpoint or seal pose changes. Those instances are used as input for the individual grouping. First, the ResNet-101 network with GeM pooling (Radenović et al. 2016, 2019) is applied to calculate descriptor vectors for each instance. The network was pretrained on the general image dataset (Radenović et al. 2016) for the retrieval tasks. The goal was to utilize general features that are inherent to the natural images to group the individuals by a general visual likeness. Next, the similarities between the instances cropped from different images are measured using the distance between the descriptors. The individual grouping is based on the similarities and is performed using the following algorithm:

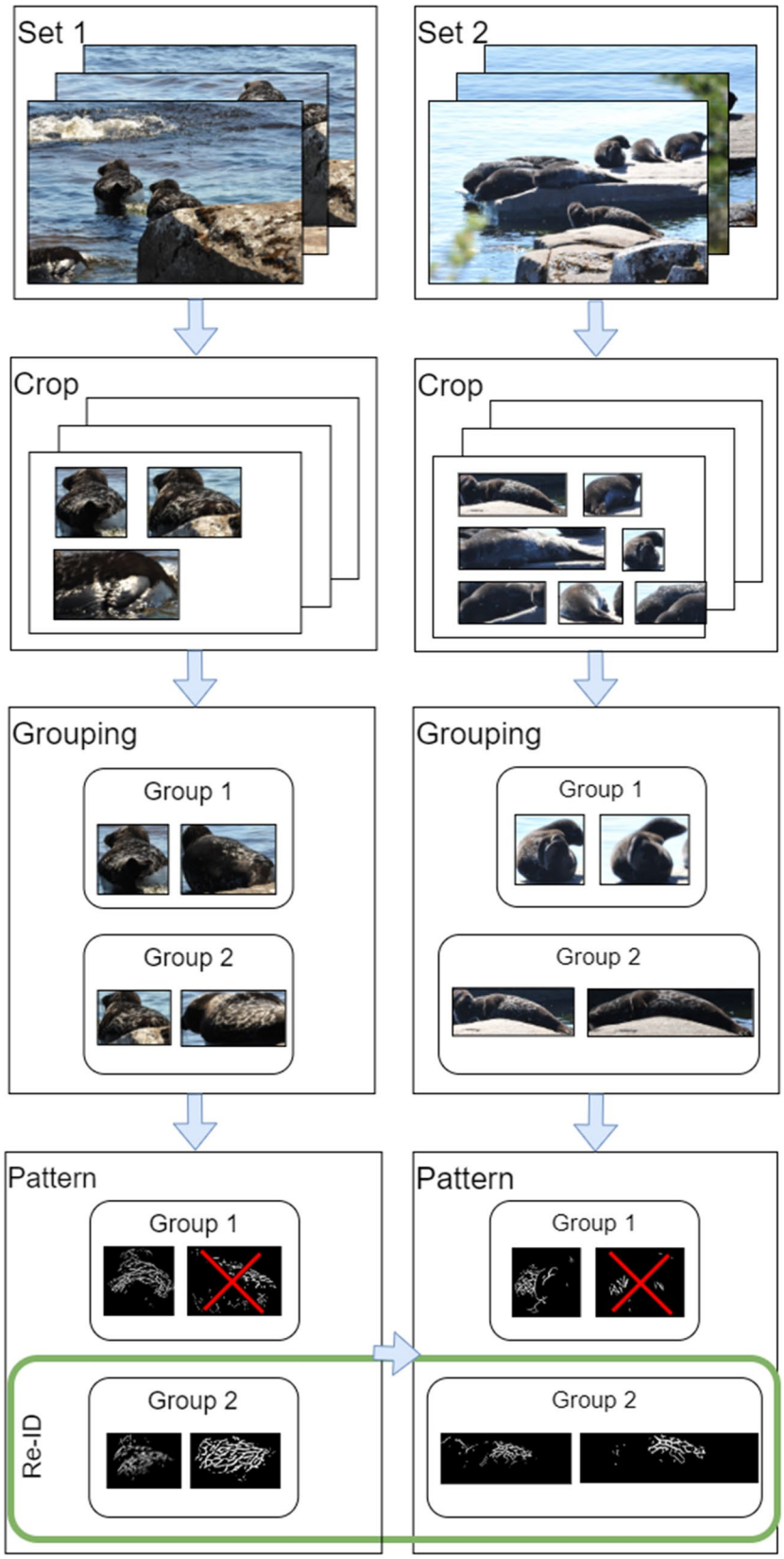


Fig. 4 Schematic of the Ladoga ringed seal re-identification pipeline. Images without a recognizable pattern are crossed out

1. Find an image with the highest number of seal instances (N). Initialize N groups using those cropped seals.
2. Choose the next image in chronological order. For each seal instance from that image, calculate distances to all previously grouped seals and aggregate them to get the mean distance to each group, resulting in a set of individual-group distances.
3. For an individual-group pair with the minimum distance out of all remaining pairs, assign that individual to that group and remove that individual and group from further consideration.
4. Repeat Step 3 until there are no more unassigned individuals.
5. Repeat Steps 2–4 until all seal individuals are grouped.

As a result, a set of groups each containing cropped images of one individual is obtained.

Pattern extraction

The characteristic allowing the re-identification of the Ladoga ringed seal individuals is the pelage pattern which consists of gray rings. Therefore, for an automatic method to succeed in the re-identification task, the method must be able to extract this pattern from an image. This is not an easy task due to the low contrast between the patterns. In this work, the CNN-based pelage pattern extraction method originally developed for the Saimaa ringed seals (Zavialkin 2020) is used. The same network that was pretrained on the Saimaa ringed seals patterns is used since the patterns of the two sister species are extremely similar in appearance. To increase the accuracy of the pattern extraction, the background is first removed using the segmentation mask obtained in the instance segmentation step. The pattern extraction method utilizes the well-known UNet encoder–decoder architecture (Ronneberger et al. 2015) that is used to transform the input image to a binary mask that corresponds to the ring pattern. Morphological opening and closing are used to remove small unconnected components and close gaps in the pattern respectively. The method is visualized in Fig. 5.

The varying quality of the image data, the low resolution of the cropped seal images, and low contrast limit the success rate of the pattern extraction. However, given multiple images of the individual in the considered group, the pattern can be often successfully extracted, at least, for some of them allowing the re-identification.

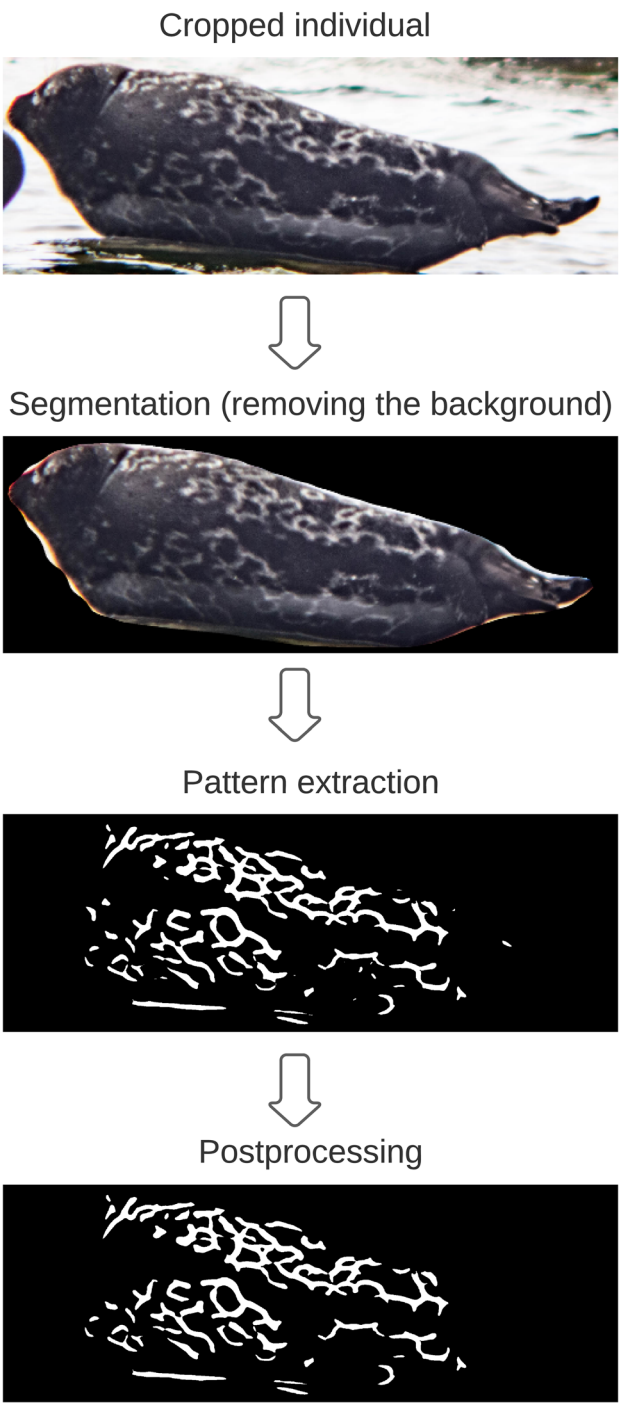


Fig. 5 Pattern extraction pipeline

Re-identification

For the final re-identification step, a modified version of the algorithm developed for the Saimaa ringed seals (Nepovinnykh et al. 2020; Chelak et al. 2021) is used. The re-identification method consists of the following steps: (1) cutting the pattern into small patches, (2) computing patch descriptors using CNN, and (3) re-identification based on Fisher vectors created by aggregating patch descriptors from an image and by comparing the query Fisher vector descriptor with the ones from the database. The patches are filtered out depending on the proportion of non-black pixels, i.e. patches with less than 10% (taken from Nepovinnykh et al. 2020) of

white pixels are considered unusable due to not containing enough pattern to be recognizable. Images with all patches filtered out are considered unrecognizable and filtered out.

Instead of using the standard triplet loss (Hoffer and Ailon 2015) that was used in Nepovinnykh et al. (2020) for embedding patches, the SphereFace loss (Liu et al. 2017) is used for the patch embedding step of the Ladoga ringed seals. Both losses can be used to solve the metric learning problem. However, one of the advantages of the SphereFace loss is that it omits the triplet mining step. This step is required for the triplet loss, meaning that during the training, triplets of samples should be chosen according to a predefined strategy, often only using the hardest samples. Choosing a proper strategy and then choosing appropriate samples during the training is usually quite challenging. SphereFace bypasses this step by formulating metric learning as a closed set classification for the duration of training. Moreover, SphereFace implies an additional constraint for the feature vectors. That is, all the vectors should lie on the hypersphere of some predefined size. Thus, the loss is able to achieve better separability of individuals using angle distances between the feature vectors without the need for complicated triplet mining.

Due to the lack of an annotated dataset for the Ladoga ringed seal pattern images, the network is trained on artificial pattern patches generated by the Adversarial generator-encoder (AGE) network (Ulyanov et al. 2018). The AGE network is a generative adversarial network (GAN) trained on the dataset of the Saimaa ringed seal fur patterns from (Nepovinnykh et al. 2020). The training dataset for GAN contained a total of 1320 pattern images in the train dataset and 660 images in the validation dataset. Training hyperparameters were taken from the original paper (Ulyanov et al. 2018) without any modification. The fur patterns of the Saimaa and Ladoga ringed seals are similar enough for the network to learn representative features of the patterns of both species. The AGE network is a generative autoencoder. A decoder part was used on noise to create a dataset of 100 classes with a total of 10000 artificial patches that were used to train the SphereFace network.

The SphereFace network utilizes ResNet-18 (He et al. 2016) as a backbone which is modified by implementing the second order attention (Ng et al. 2020) after the 3rd and 4th ResNet blocks. In addition, all the original ResNet pooling layers are replaced with SoftPool (Stergiou et al. 2021). Finally, GeM (generalized mean pooling) (Radenović et al. 2019) is utilized as a global pooling of a feature map produced by the final convolutional layer. The global pooling is then followed by a fully-connected layer with the output size of 512 and an L^2 normalization layer. Such architecture of final layers is chosen after the original GeM paper (Radenović et al. 2019) to provide features with rotation and translation invariance.

Principal Component Analysis (PCA) is applied to all patch embeddings (descriptors produced by the SphereFace network) to reduce the dimensionality and decorrelate the patch descriptors. Multiple images of the same individual in a group and multiple patches per image results in a large amount descriptors for each seal. These need to be combined to perform the re-identification that computes similarity between query seal and a known individual.

Fisher Vector (Perronnin and Dance 2007; Arandjelović and Zisserman 2012) is used to create a descriptor for the full seal by aggregating descriptors of patches from the image or the image group. Using all the images in the group adds extra information for the matching process, especially when different parts of the pattern are visible on different images from the same group. The codebook for Fisher vectors is created by applying the Gaussian Mixture Model (GMM) method to the database patches. Fisher vectors themselves are constructed from feature gradients with respect to the GMM parameters. Finally, cosine distances between Fisher vectors are used to rank database images based on their similarity to the query. The method is visualized in Fig. 6.

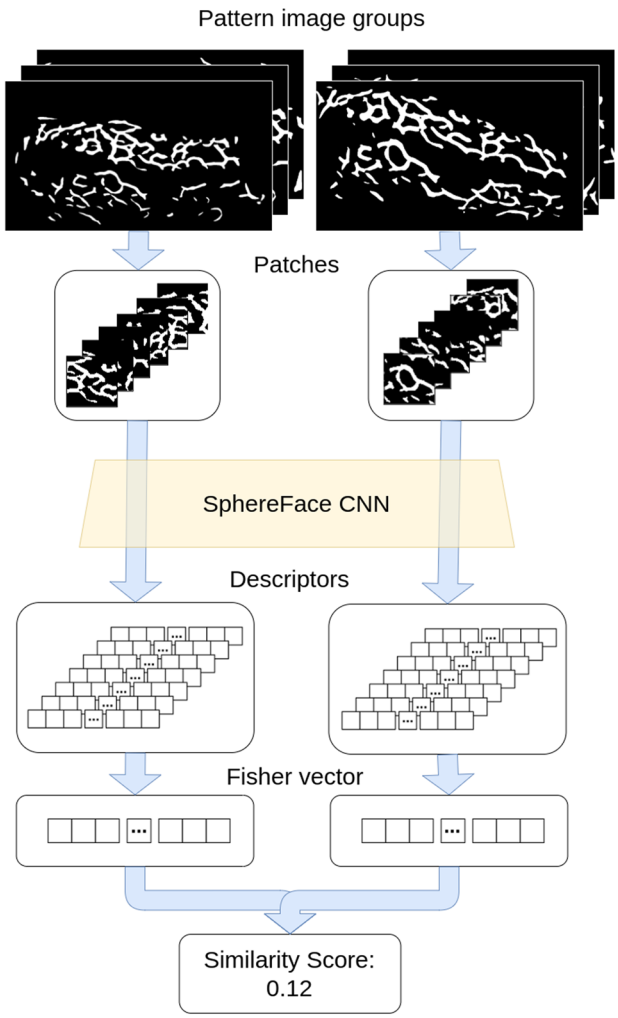


Fig. 6 Re-identification pipeline schematic

Experiments

Data

The image data were collected using two methods: (1) game cameras capturing images within a fixed time interval, and (2) DSLR or other handheld cameras, with multiple consecutive images of the same group (Gromov et al. 2021). Example images are presented in Fig. 7.

The images were collected during July of 2019 and June of 2020. The exact dates and the distribution of images are presented in Fig. 8. There is a maximum of 11 seals present in the same image. Images with smaller numbers of seals are more frequent. The exact distribution of images in relation to the number of seals per image is shown in Fig. 9. All images were collected in sets to ensure that an image set contains images collected during one session. Image sets with few images are most frequent. Some image sets contain upwards of 69 images. The exact distribution of the number of images in relation to a number of image sets is presented in Fig. 10.

Three datasets were prepared to evaluate the different steps of the proposed framework. For the instance segmentation step, 150 images were selected varying in the type, quality, number of seals, and weather conditions to provide a representative set for both model training and testing. The number of individuals in images varied from 1 to 19. The contour of each seal was manually annotated for all images as shown in Fig. 11. The dataset was divided into

Fig. 7 Examples of dataset images

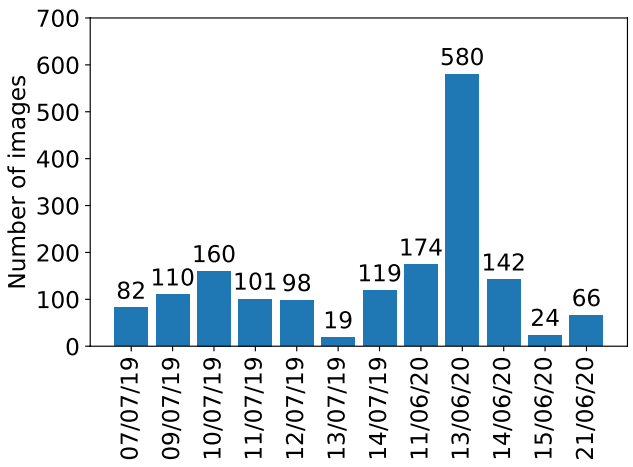


Fig. 8 Distribution of images in relation to dates

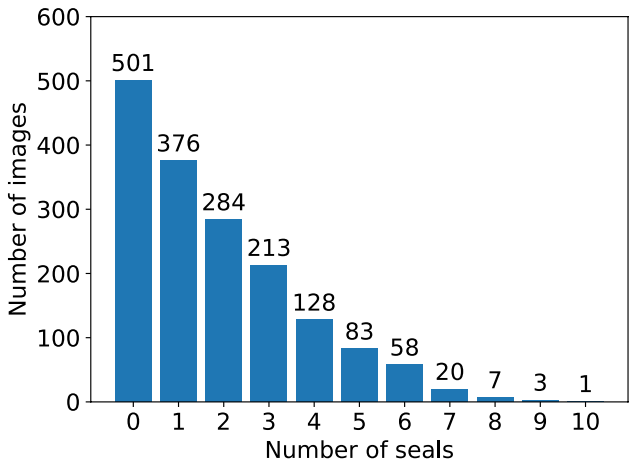


Fig. 9 Distribution of images in relation to a number of seals per image

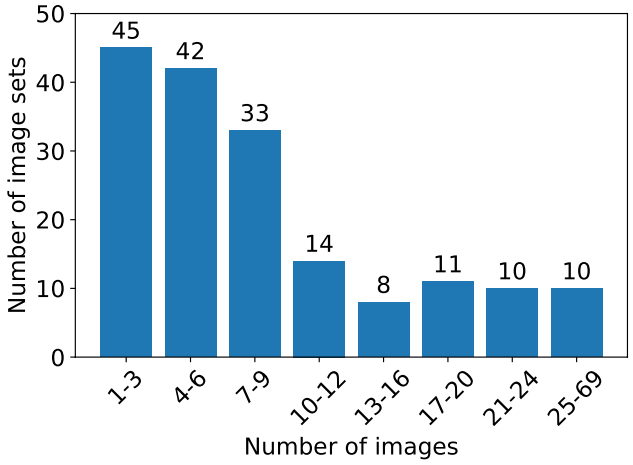


Fig. 10 Distribution of images in relation to image sets

Fig. 11 An example of manual seal annotation

a training set containing 100 images and a test set containing 50 images. The training procedure and the dataset are described in detail in (Lushpanov 2020).

For evaluating the individual grouping step, image sequences obtained using handheld cameras were used. Each sequence was obtained during a short period of time lasting only a couple of hours, and contains images from the same group of the seals with variations on the individuals that are visible in the images. The dataset contains 60 image sequences with 3 to 42 images per sequence. The total amount of images is 689. The number of seals in an image varies from 1 to 14. To evaluate the re-identification steps, multiple image sequences had to contain the same seal individuals. In total, 21 seal individuals have images in multiple sequences. It should be noted that a large-scale Ladoga ringed seal photo-identification database with expert annotated seal IDs does not exist yet.

For the re-identification step, a small dataset of the known individuals is created from the previously described images. This dataset contains 50 individuals, with 81 images of segmented seals in the database and a total of 299 images in the query. The query contains 37 groups that are used for experiments with re-identification with grouping.

Results

Instance segmentation

Six models with different backbone architectures were compared to find the best one for the Ladoga ringed seal detection and segmentation. The training dataset contains 100 annotated seal images. All pre-trained models were then fine-tuned on the seal training dataset. The learning rate was fixed as 0.00025 and the detection threshold of 0.8 was used. The models were evaluated using the mean Average Precision (mAP) and the F_1 score. Both F_1-score and mAP use Precision and Recall metrics. Precision and Recall are calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives. Then, F_1-score is the harmonic mean of Precision and Recall as follows:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

and mAP is computed by varying the threshold for the IoU detection and calculated as the area under the curve of function $y(x) = \text{precision}(\text{recall})$.

The results are presented in Table 1. Figure 12 shows example results with each model. The ResNet architecture with 101 layers combined with FPN was found to provide the best accuracy and was selected.

Individual grouping

For the initial grouping, the image descriptors were computed using the pre-trained ResNet101 models from (Radenović et al. 2016, 2019). Three models were compared. All models use whitening, which is a method for the decorrelation of data and centering of data such that it has a unit variance. The models differ by the final feature extraction method, whitening, and the dataset used for training. The first two models use generalized mean pooling and differ only in whitening and were pretrained on a large retrieval dataset RetrievalSfM120k (Radenović et al. 2016).

Table 1 Accuracy of architectures based on mAPs and F_1 scores with the testing threshold equal to 0.8

Architecture	mAP		F_1 score	
	Segmentation	Bounding box	Segmentation	Bounding box
ResNet-50-FPN	66.9%	61.5%	0.259	0.233
ResNet-50-C4	68.5%	59.9%	0.270	0.239
ResNet-50-DC5	68.0%	64.0%	0.273	0.261
ResNet-101-FPN	**71.1%**	**66.4%**	**0.290**	**0.276**
ResNet-101-C4	66.8%	52.5%	0.263	0.212
ResNet-101-DC5	69.9%	62.9%	0.280	0.261

Best results are indicated in bold

The first model uses the fully-connected layer for the whitening. Linear discriminant projections proposed by (Mikolajczyk and Matas 2007) is used for whitening in the second model. The third model uses maximum activation pooling (MAC) instead of GeM and was pretrained on a standard ImageNet (Krizhevsky et al. 2017) dataset. The Rand index (Rand 1971) was used as an evaluation metric. If all possible pairs of elements are considered and the grouping task is formulated as an attempt to classify them as "same class" or "different class", the Rand index corresponds to the accuracy of such classification. Table 2 shows that the best results were obtained with GeM for pooling and the fully connected layer for the whitening.

Pattern extraction

Examples of the pattern extraction results are shown in Fig. 13. Due to the lack of ground truth annotations, it was not possible to compute exact pattern extraction accuracy. However, based on visual analysis the proposed method was able to extract the satisfactory pattern from 42% of images. The pattern extraction step was further used to filter out cropped images where the pattern was not visible. The filtering step can be thought of as a classification problem with two classes: "pattern is suitable for re-identification" and "pattern is not suitable for re-identification or absent". The ground truth was created by visual assessment. The classification accuracy of 85.6% was achieved for the filtering step. Out of the image groups where there is a pattern visible to the human eye, the proposed method was able to successfully extract the pattern from at least one image for 93.3% of the groups.

Re-identification

To train the SphereFace network used for identification, the AMSGrad (Reddi et al. 2019) version of the AdamW (Loshchilov and Hutter 2019) optimizer was used. The batch size was set to 32, the initial learning rate was 10^{-5}, and the weight decay was 10^{-3}. The network was trained for 5 epochs with the learning rate being cut to 1.5×10^{-6} after 3 epochs.

First, an experiment to determine the optimal dimensionality for PCA and the number of clusters for the codebook creation was conducted. The values that produced the best accuracy were chosen and used in all subsequent experiments. The resulting values are 512 clusters and 128 dimensions after PCA.

Re-identification experiments were done with and without the grouping step. An example of found matches for the re-identification without grouping is presented in Fig. 14. A comparison of the results for the method applied to the pattern images versus the original images was performed as well. The results are presented in Table 3. For each query, possible matches from the database are ordered by their similarity to the query in descending order. Top-n refers to the percent of images for which a correct match is found in the n closest matches from the database. For example, a Top-5 score of 50% would mean that for half of the queries at least one correct match has been found in the closest 5 matches from the database. It should be noted, however, that for the no grouping version with pattern extraction, images where the pattern is not visible or recognizable are counted as wrong matches since they cannot be matched with that method. The results indicate that both the pattern extraction and the grouping steps significantly improve the re-identification accuracy.

Table 2 Rand index of individual grouping

Architecture	Training data on	Accuracy
ResNet-101+GeM+FC	**RetrievalSfM120k** (Radenović et al. 2016)	**95.4%**
ResNet-101+GeM+WHITENING	RetrievalSfM120k (Radenović et al. 2016)	94.5%
ResNet-101+MAC	ImageNet (Krizhevsky et al. 2017)	92.0%

The best model is indicated in bold

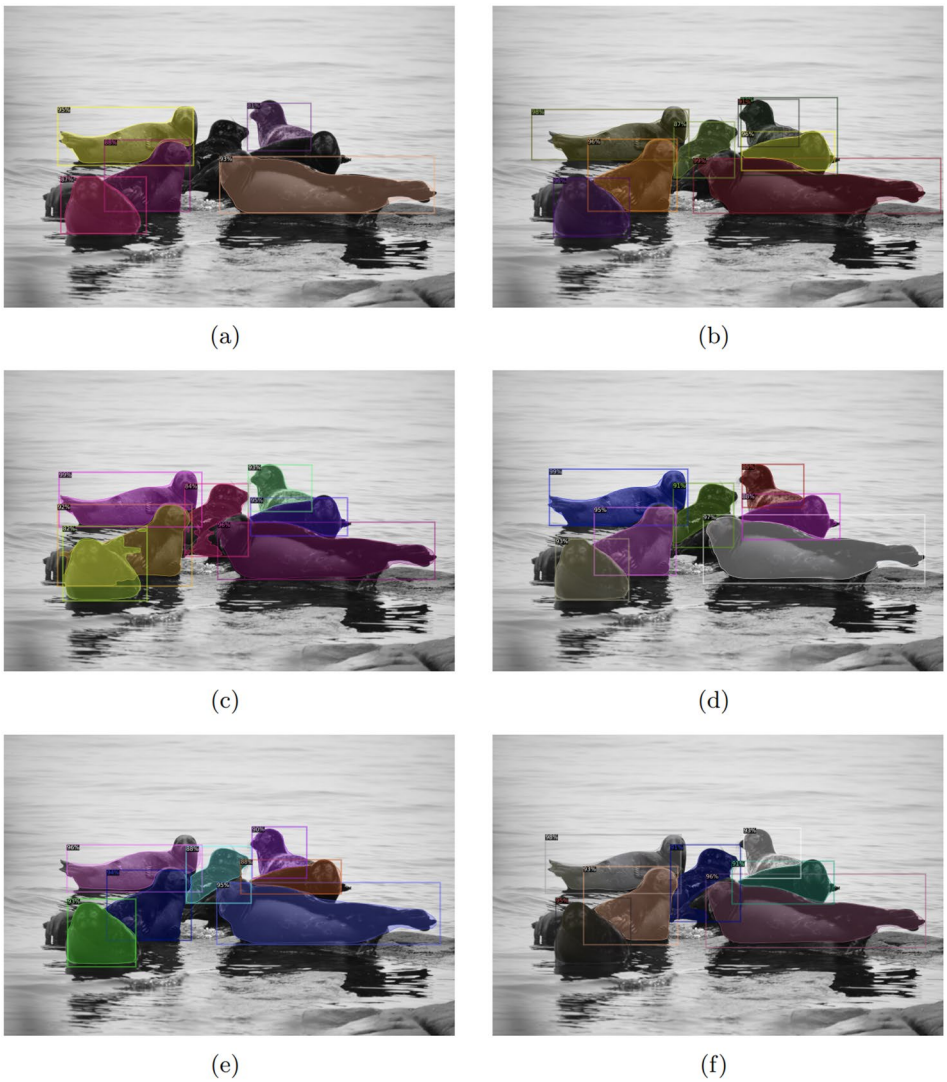


Fig. 12 Examples of instance segmentation results: **a** ResNet-50-FPN; **b** ResNet-50-C4; **c** ResNet-50-DC5; **d** ResNet-101-FPN; **e** ResNet-101-C4; **f** ResNet-101-DC5

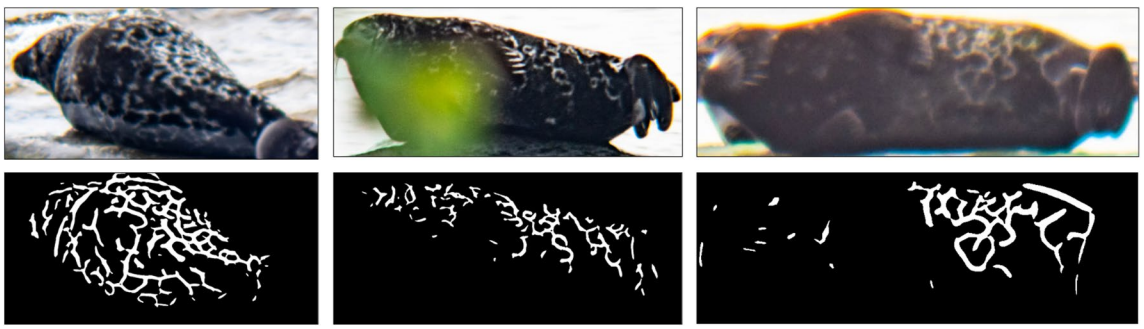


Fig. 13 Pattern extraction example results: original images (top row); extracted patterns (bottom row)

Table 3 Re-identification accuracy for different variants of the algorithm

Grouping	Pattern extraction	Top-1	Top-2	Top-3	Top-4	Top-5
No	No	53.18%	60.87%	64.88%	67.89%	68.23%
	Yes	60.74%	66.11%	70.47%	75.17%	77.18%
Yes	No	83.95%	89.97%	91.97%	93.98%	94.65%
	Yes	**93.62%**	**94.30%**	**94.30%**	**96.64%**	**96.64%**

Best results are indicated in bold

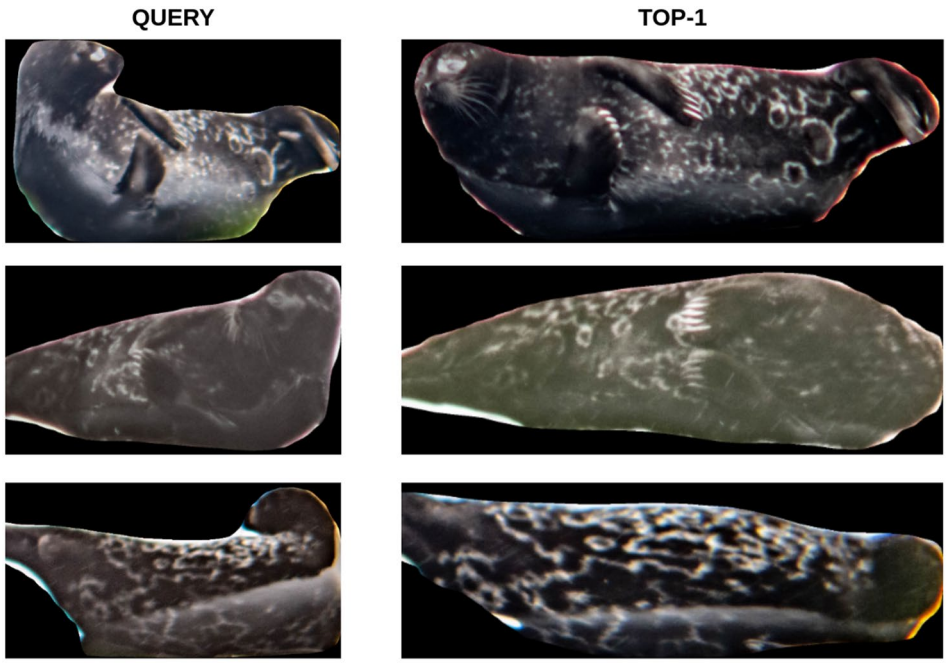


Fig. 14 Example of correct re-identification results: The query images (in the left), the corresponding closest matches from the database (in the right)

Conclusions

A pipeline for the processing of image data and re-identification of the Ladoga ringed seals has been successfully developed and deployed. It consists of four steps: seal instance segmentation, individual grouping, pattern extraction, and re-identification. Mask R-CNN was selected for the instance segmentation and demonstrated good accuracy on the Ladoga ringed seal images. Various backbone architectures for Mask R-CNN were compared and a combination of ResNet-101 with Feature Pyramid Network produced the best segmentation accuracy. An image retrieval method is used to group the detected seals based on the visual similarity from the image sequences obtained from the same location within a short time period, resulting in groups each containing cropped images of one individual. These image groups then could be used to re-identify the individual by searching for the match from a database of the known individuals. Having multiple images of the same individual as a query greatly increased the re-identification accuracy compared to the traditional methods that utilize only one image at a time. For pattern extraction, the CNN-based method utilizing the UNet encoder–decoder architecture was able to extract the patterns from the Ladoga ringed seal images despite being trained on the Saimaa ringed seal data. Finally, a modification of a pattern matching originally developed for the Saimaa ringed seals using Fisher vectors computed from the SphereFace embeddings of the pattern image patches was used for re-identification. This step utilizes previously computed grouping information for the creation of descriptors for each group rather than an individual image. This approach greatly improved the re-identification accuracy as compared to a standard image-to-image matching-based re-identification.

Appendix

Fig. A1 Ladoga ringed seals (*Pusa hispida ladogensis*) in Lake Ladoga, Russian Federation

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42991-022-00229-3.

Acknowledgements The authors would like to thank Gromov V., Shakhnazarova V., Dmitriev P., Fedeneva Y., Bakunovich P., Kouprianov A., Trukhanova I. and the Interregional Charitable Public Organisation Biologists for Nature Conservation (BFNC) from the Russian Federation for providing the Ladoga ringed seal imagery.

Author contributions TE and HK were responsible for the supervision of the research, designing methodology, and project administration; AL developed the segmentation algorithm, EN and IC developed a grouping algorithm, EN and IC implemented the re-identification algorithm. OC collected and processed the data; EN, AL, TE, and HK prepared the original draft of the manuscript. All the authors gave the final approval for publication.

Funding Open Access funding provided by LUT University (previously Lappeenranta University of Technology (LUT)). The research is a part of the CoExist project (Project ID: KS1549) funded by the European Union, the Russian Federation, and the Republic of Finland via The South-East Finland–Russia CBC 2014-2020 Programme. We would also like to thank the Raija ja Ossi Tuuliaisen Säätiö foundation for additional financial support.

Availability of data and material Not available.

Code availability Not available.

Declarations

Conflict of interest We declare no competing interests.

Ethics approval Fieldwork to collect imagery of Ladoga ringed seal was carried out in accordance with the Valaam Archipelago Nature Park regulations and by the written approval of the Ministry of the Natural Resources and Environment of the Republic of Karelia dated 13.01.2020.

References

Arandjelović R, Zisserman A (2012) Three things everyone should know to improve object retrieval. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2911–2918. https://doi.org/10.1109/CVPR.2012.6248018

Arandjelović R, Gronat P, Torii A, Pajdla T, Sivic J (2016) NetVLAD: CNN architecture for weakly supervised place recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 5297–5307. https://doi.org/10.1109/CVPR.2016.572

Berger-Wolf T, Rubenstein D, Stewart C, Holmberg J, Parham J, Crall J (2015) Ibeis: Image-based ecological information system: From pixels to science and conservation. In: Bloomberg data for good exchange conference, vol 2

Berger-Wolf T, Rubenstein D, Stewart C, Holmberg J, Parham J, Menon S, Crall J, Van Oast J, Kiciman E, Joppa L (2017) Wildbook: crowdsourcing, computer vision, and data science for conservation. arxiv:1710.08880

Bouma S, Pawley M, Hupman K, Gilman A (2018) Individual common dolphin identification via metric embedding learning. In: International conference on image and vision computing New Zealand (IVCNZ), pp 1–6. https://doi.org/10.1109/IVCNZ.2018.8634778

Burghardt T, Calić J (2006) Analysing animal behaviour in wildlife videos using face detection and tracking. IEE Proc Vis Image Signal Process 153:305. https://doi.org/10.1049/ip-vis:20050052

Chehrsimin T, Eerola T, Koivuniemi M, Auttila M, Levänen R, Niemi M, Kunnasranta M, Kälviäinen H (2018) Automatic individual identification of Saimaa ringed seals. IET Comput Vis 12:146–152. https://doi.org/10.1049/iet-cvi.2017.0082

Chelak I, Nepovinnykh E, Eerola T, Kälviäinen H, Belykh I (2021) EDEN: Deep Feature Distribution Pooling for Saimaa Ringed seals pattern matching. arxiv:2105.13979

Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. Eur. Conf. Comput. Vis. (ECCV) 11211:833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Chen K, Ouyang W, Loy CC, Lin D, Pang J, Wang J, Xiong Y, Li X, Sun S, Feng W, Liu Z, Shi J (2019) Hybrid task cascade for instance segmentation. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 4969–4978. https://doi.org/10.1109/CVPR.2019.00511

Cheng X, Zhu J, Zhang N, Wang Q, Zhao Q (2020) Detection features as attention (Defat): a keypoint-free approach to amur tiger re-identification. In: IEEE international conference on image processing (ICIP), pp 2231–2235. https://doi.org/10.1109/ICIP40778.2020.9190667

Crall J, Stewart C, Berger-Wolf T, Rubenstein D, Sundaresan S (2013) HotSpotter—patterned species instance recognition. In: IEEE workshop on applications of computer vision (WACV), pp 230–237. https://doi.org/10.1109/WACV.2013.6475023

Deb D, Wiper S, Gong S, Shi Y, Tymoszek C, Fletcher A, Jain AK (2018) Face recognition: primates in the wild. In: IEEE 9th international conference on biometrics theory, applications and systems (BTAS), pp 1–10. https://doi.org/10.1109/BTAS.2018.8698538

Dunbar S, Anger E, Parham J, Kingen C, Wright M, Hayes C, Safi S, Holmberg J, Salinas L, Baumbach D (2021) HotSpotter: using a computer-driven photo-id application to identify sea turtles. J Exp Mar Biol Ecol 535:151490. https://doi.org/10.1016/j.jembe.2020.151490

Girshick R (2015) Fast R-CNN. In: IEEE international conference on computer vision (ICCV), pp 1440–1448

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 580–587. https://doi.org/10.1109/CVPR.2014.81

Gromov VV, Shakhnazarova VY, Chirkova OA, Trukhanova IS (2021) Development of a database for photo-identification of the Ladoga ringed seal *Pusa hispida ladogensis*. In: Proceedings of the conference "Marine mammals of the holarctic" **(in press)**

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/CVPR.2016.90

He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. In: IEEE international conference on computer vision (ICCV), pp 2980–2988.https://doi.org/10.1109/ICCV.2017.322

Hoffer E, Ailon N (2015) Deep metric learning using triplet network. In: Similarity-based pattern recognition, pp 84–92. https://doi.org/10.1007/978-3-319-24261-3_7

Holmberg J, Norman B, Arzoumanian Z (2009) Estimating population size, structure, and residency time for whale sharks *Rhincodon typus* through collaborative photo-identification. Endanger Species Res 7:39–53. https://doi.org/10.3354/esr00186

Jégou H, Douze M, Schmid C, Pérez P (2010) Aggregating local descriptors into a compact image representation. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 3304–3311. https://doi.org/10.1109/CVPR.2010.5540039

Kellenberger B, Marcos D, Lobry S, Tuia D (2019) Half a percent of labels is enough: efficient animal detection in UAV imagery using deep CNNs and active learning. IEEE Trans Geosci Remote Sens 57:9524–9533. https://doi.org/10.1109/TGRS.2019.2927393

Krizhevsky A, Sutskever I, Hinton G (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60:84–90. https://doi.org/10.1145/3065386

Kunnasranta M, Niemi M, Auttila M, Valtonen M, Kammonen J, Nyman T (2021) Sealed in a lake–biology and conservation of the endangered Saimaa ringed seal: a review. Biol Cons 253:108908. https://doi.org/10.1016/j.biocon.2020.108908

Law H, Deng J (2020) CornerNet: detecting objects as paired keypoints. Int J Comput Vision 128:642–656. https://doi.org/10.1007/s11263-019-01204-1

Li Y, Qi H, Dai J, Ji X, Wei Y (2017) Fully convolutional instanceaware semantic segmentation. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 4438–4446. https://doi.org/10.1109/CVPR.2017.472

Li S, Li J, Tang H, Qian R, Lin W (2020) ATRW: a benchmark for Amur tiger re-identification in the wild. In: International conference on multimedia (ACM), pp 2590–2598. https://doi.org/10.1145/3394171.3413569

Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 936–944. https://doi.org/10.1109/CVPR.2017.106

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg A (2016) SSD: single shot multibox detector. In: European conference on computer vision (ECCV), pp 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Liu W, Wen Y, Yu Z, Li M, Raj B, Song L (2017) SphereFace: deep hypersphere embedding for face recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 6738–6746. https://doi.org/10.1109/CVPR.2017.713

Liu C, Zhang R, Guo L (2019a) Part-pose guided Amur tiger re-identification. In: IEEE international conference on computer vision workshop (ICCVW), pp 315–322. https://doi.org/10.1109/ICCVW.2019.00042

Liu N, Zhao Q, Zhang N, Cheng X, Zhu J (2019b) Pose-guided complementary features learning for Amur tiger re-identification. In: IEEE international conference on computer vision workshop (ICCVW), pp 286–293. https://doi.org/10.1109/ICCVW.2019.00038

Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M (2020) Deep learning for generic object detection: a survey. Int J Comput Vision 128:261–318. https://doi.org/10.1007/s11263-019-01247-4

Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. arxiv:1711.05101

Lowe D (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60:91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lushpanov A (2020) Instance segmentation of Ladoga ringed seals. Master's thesis, Lappeenranta-Lahti University of Technology LUT, Finland

Mikolajczyk K, Matas J (2007) Improving descriptors for fast tree matching by optimal linear projection. In: IEEE international conference on computer vision (ICCV), pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408871

Mishchuk A, Mishkin D, Radenović F, Matas J (2018) Working hard to know your neighbor's margins: local descriptor learning loss. arxiv:1705.10872

Moskvyak O, Maire F, Dayoub F, Baktashmotlagh M (2020) Learning landmark guided embeddings for animal re-identification. In: IEEE winter applications of computer vision workshops (WACVW), pp 12–19. https://doi.org/10.1109/WACVW50321.2020.9096932

Nepovinnykh E, Eerola T, Kälviäinen H, Radchenko G (2018) Identification of saimaa ringed seal individuals using transfer learning. In: Blanc-Talon J, Helbert D, Philips W, Popescu D, Scheunders P (eds) Advanced concepts for intelligent vision systems. Springer, Cham, pp 211–222. https://doi.org/10.1007/978-3-030-01449-0_18

Nepovinnykh E, Eerola T, Kälviäinen H (2020) Siamese network based pelage pattern matching for ringed seal re-identification. In: IEEE winter applications of computer vision workshops (WACVW), pp 25–34. https://doi.org/10.1109/WACVW50321.2020.9096935

Ng T, Balntas V, Tian Y, Mikolajczyk K (2020) SOLAR: second-order loss and attention for image retrieval. In: European conference on computer vision (ECCV), pp 253–270. https://doi.org/10.1007/978-3-030-58595-2_16

Parham J, Crall J, Stewart C, Berger-Wolf T, Rubenstein D (2017) Animal population censusing at scale with citizen science and photographic identification. In: AAAI spring symposium series

Parham J, Stewart C, Crall J, Rubenstein D, Holmberg J, Berger-Wolf T (2018) An animal detection pipeline for identification. In: IEEE winter conference on applications of computer vision (WACV), pp 1075–1083. https://doi.org/10.1109/WACV.2018.00123

Park H, Lim A, Choi T-Y, Baek S-Y, Song E-G, Park Y (2019) Where to spot: individual identification of leopard cats (*Prionailurus*

bengalensis euptilurus) in South Korea. J Ecol Environ 43:39. https://doi.org/10.1186/s41610-019-0138-z

Perronnin F, Dance C (2007) Fisher kernels on visual vocabularies for image categorization. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8. https://doi.org/10.1109/CVPR.2007.383266

Perronnin F, Liu Y, Sánchez J, Poirier H (2010a) Large-scale image retrieval with compressed Fisher vectors. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 3384–3391. https://doi.org/10.1109/CVPR.2010.5540009

Perronnin F, Sánchez J, Mensink T (2010b) Improving the Fisher Kernel for large-scale image classification. In: European conference on computer vision (ECCV), pp 143–156. https://doi.org/10.1007/978-3-642-15561-1_11

Quinby B, Creighton C, Flaherty E (2021) Estimating population abundance of burying beetles using photo-identification and mark recapture methods. Environ Entomol 50:238–246. https://doi.org/10.1093/ee/nvaa139

Radenović F, Tolias G, Chum O (2016) CNN image retrieval learns from BoW: unsupervised fine-tuning with hard examples. In: European conference on computer vision (ECCV), pp 3–20. https://doi.org/10.1007/978-3-319-46448-0_1

Radenović F, Tolias G, Chum O (2019) Fine-tuning CNN image retrieval with no human annotation. IEEE Trans Pattern Anal Mach Intell 41:1655–1668. https://doi.org/10.1109/TPAMI.2018.2846566

Rand W (1971) Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66:846–850. https://doi.org/10.1080/01621459.1971.10482356

Reddi S, Kale S, Kumar S (2019) On the convergence of Adam and beyond. arxiv:1904.09237

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 779–788. https://doi.org/10.1109/CVPR.2016.91

Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention (MICCAI), pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Schneider S, Taylor G, Linquist S, Kremer S (2019) Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol Evol 10:461–470. https://doi.org/10.1111/2041-210X.13133

Sivic J, Zisserman A (2003) Video Google: a text retrieval approach to object matching in videos. In: IEEE international conference on computer vision (ICCV), vol 2, pp 1470–1477. https://doi.org/10.1109/ICCV.2003.1238663

Smeulders A, Worring M, Santini S, Gupta A, Jain R (2000) Content based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22:1349–1380. https://doi.org/10.1109/34.895972

Stergiou A, Poppe R, Kalliatakis G (2021) Refining activation downsampling with SoftPool. In: IEEE international conference on computer vision (ICCV), pp 10357–10366

Tolias G, Sicre R, Jégou H (2016) Particular object retrieval with integral max-pooling of CNN activations. arxiv:1511.05879

Trukhanova I (2013) The ladoga ringed seal (*Pusa hispida ladogensis*) under changing climatic conditions. Russ J Theriol 12:41–48

Trukhanova I, Guratie E, Sagitov R (2013) Distribution of Hauled-Out Ladoga Ringed Seals (*Pusa hispida ladogensis*) in Spring 2012. Arctic, vol 66, pp 417–428

Ulyanov D, Vedaldi A, Lempitsky V (2018) It takes (only) two: adversarial generator-encoder networks. In: The AAAI conference on artificial intelligence, vol 32

Verma G, Gupta P (2018) Wild animal detection using deep convolutional neural network. In: International conference on computer vision & image processing (cvip), pp 327–338. https://doi.org/10.1007/978-981-10-7898-9_27

Zavialkin D (2020) CNN-based ringed seal pelage pattern extraction. Master's thesis, Lappeenranta-Lahti University of Technology LUT, Finland

Zhang W, Sun J, Tang X (2011) From tiger to panda: animal head detection. IEEE Trans Image Process 20:1696–1708. https://doi.org/10.1109/TIP.2010.2099126

Zhelezniakov A, Eerola T, Koivuniemi M, Auttila M, Levänen R, Niemi M, Kunnasranta M, Kälviäinen H (2015) Segmentation of saimaa ringed seals for identification purposes. In: Advances in visual computing, pp 227–236. https://doi.org/10.1007/978-3-319-27863-6_21