

Siamese Network Based Pelage Pattern Matching for Ringed Seal Re-identification

Ekaterina Nepovinnikh

Tuomas Eerola

Heikki Kälviäinen

Computer Vision and Pattern Recognition Laboratory (CVPRL)

Department of Computational and Process Engineering

Lappeenranta-Lahti University of Technology LUT, Lappeenranta, Finland

firstname.lastname@lut.fi

Abstract

In this paper we propose a method to match pelage patterns of the Saimaa ringed seals enabling the re-identification of individuals. First, the pelage pattern is extracted from the seal's fur using a method based on the Sato tubeness filter. After this, the similarities of the pelage pattern patches are computed using a siamese network trained with a triplet loss function and a large dataset of manually selected patches. The similarities are then used to find the best matching patches from the images in the database of known individuals. Furthermore, we employ the proposed pattern matching method to build a full framework for the ringed seal re-identification, consisting of CNN-based animal segmentation, patch correspondence detection, and ranking the images in the database of known seal individuals based on the similarity to the query image. Our experiments on challenging datasets of Saimaa ringed seals show that the proposed method achieves promising identification results, providing a useful tool for the Saimaa ringed seal monitoring.

1. Introduction

Automatic wildlife camera traps and crowd-sourced image material provide novel possibilities to monitor endangered animals species. However, massive image volumes that these methods produce is overwhelming for biologists to go through, which calls for automatic systems to perform the analysis. The main task is to identify the animal individuals in the images to provide the basis for the monitoring, including population size estimation and animal migration tracking.

Saimaa ringed seals (*Pusa hispida saimensis*) are endangered due to various anthropogenic factors, such as random bycatch and climate change. In addition, the risk of their extinction is high due to low genetic diversity and small

population (currently around 400 individuals). The current knowledge about this animal is mainly based on telemetric studies with a relatively small number of individuals. Photo-ID (photo-identification) using camera traps is an approved and effective non-invasive method for studying and monitoring the Saimaa ringed seals [14]. Ringed seals have permanent pelage patterns that are unique to each individual and can be used for the identification (see Fig. 1).

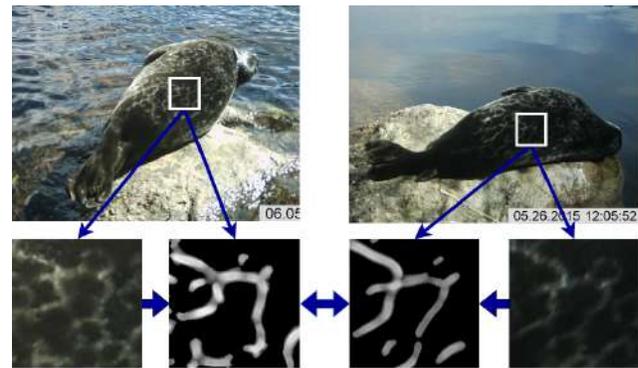


Figure 1. Saimaa ringed seal identification based on pelage pattern patches.

Automatic methods for the re-identification of individual animals have been proposed for various species. However, the re-identification of ringed seals introduces some additional challenges. First, large variation in possible poses is further exacerbated by the deformable nature of those animals. This, in addition to the fact that the pelage pattern is non-uniform and depends on the visible area of the animal, limits the size of the regions that could actually be used for the identification task. Second, the contrast between the ring pattern and the rest of the pelage is low and the appearance of the pattern varies between wet and dry fur. Finally, image quality of automatic camera traps is typically low, which might lead to the loss of details. These challenges make the re-identification considerably more difficult for

ringed seals than, for example, for zebras with clearly visible pattern and limited variation in the pose of the torso.

In this paper, we propose a method for comparison and matching of ringed seal pelage pattern patches. The method starts by extracting the pelage patterns from the images by utilizing the Sato tubeness filter based method [23]. This filter can be used to detect continuous ridges, e.g. tubes, wrinkles, rivers, or, in our case, pelage pattern. It calculates the eigenvectors of the Hessian to compute the similarity of an image region to tubes. This step gets rid of the irrelevant factors, such as illumination, and focuses the attention of the algorithm on the actual pattern itself. This is crucial since most of the data is collected with automatic camera traps and the ringed seals tend to stay in the same region. The same seal is often captured with the same camera, the same background, and the similar illumination and pose. This causes supervised methods to learn superficial features which might appear as good features in the training data, but do not help to re-identify the seals in the real-world application scenarios. To calculate similarities between patterns and to match them to each other, a siamese network trained with a triplet loss function is used.

We build a full framework for the identification of Saimaa ringed seal utilizing the pattern matching. The framework starts with CNN based segmentation of the seal from the background. This allows reliable pattern extraction. Pattern patches are then extracted from the query image and the best matches for each patch are searched from the database of patches from the known seals. The final re-identification is done based on the similarity of the pattern patches. Each patch of the query image is compared to all patches from the database image, and a similarity heatmap is built. Local maxima are used as candidates for this patch projection. A geometrically aware algorithm then selects suitable projection sets for the entire image and ranks the comparative similarity.

In the experimental part of the work, we demonstrate that the proposed pelage pattern matching provides high matching accuracy outperforming siamese network applied for unprocessed patches. Moreover, we show promising identification results, providing a useful semi-automatic re-identification tool for biologists. The system offers N possible candidates and the user manually chooses the corresponding animal individual.

To summarize, this paper makes the following contributions: 1) a pattern extraction algorithm that works on images captured in the wild and reduces the impact of external conditions such as lighting, weather and location, and 2) a novel end-to-end Saimaa ringed seal re-identification method that utilizes a siamese triplet network for comparing distinct image regions and a matching algorithm that checks topological consistency of corresponding points.

2. Related work

2.1. Animal re-identification

Traditional tools for monitoring animals such as tagging requires a physical contact with the animal which causes stress and may change the behavior of the animal. To avoid this, camera-based methods utilizing computer vision algorithms have been developed for animal re-identification. Many of them are species-specific which limits their usability [19, 22, 10].

There have also been research efforts towards creating a unified approach applicable for identification purposes for several animal species. Wildbook [3] is a large-scale project for the study, monitoring and identification of animals with distinguishable marks on the body. Wildbook's computer vision based identification methods are build on the HotSpotter algorithm [7]. This algorithm is not species specific and has been applied to Grevy's and plain zebras, giraffes, leopards, and lionfish. HotSpotter uses viewpoint invariant descriptors and a scoring mechanism that emphasizes the most distinctive keypoints and descriptors. In [26], a species recognition algorithm based on sparse coding spatial pyramid matching (ScSPM) was proposed. It was shown that the proposed object recognition techniques can be successfully used to identify animals on sequences of images captured using camera traps in nature.

Due to the recent progress in deep learning, convolutional neural networks (CNN) have also become popular tools for animal biometrics. For example, in [2], re-identification of the cattle using CNN approach combined with k-NN classifier was proposed. The method achieved the accuracy of over 80% outperforming competing methods. The approach is, however, specific to muzzle patterns of cattle. The muzzle patterns are obtained manually, providing consistent data that simplifies the re-identification.

A typical problem in the wildlife animal re-identification is that it is practically impossible to collect a large dataset with large number of images for all individuals. Often the method needs to be able to identify an individual with only one or few previously collected examples. Moreover, the animal re-identification method should be able to recognize if the query image contains an individual that is not in the database of the known individuals. Recently, siamese neural network based approaches have gained popularity in the task of animal re-identification [13]. These methods provide a tool to classify objects based on only one example image (one-shot learning) and to recognize if it belongs to a class which the network has never seen. For example, in [24], the effectiveness of siamese neural networks for re-identification of Human, Chimpanzee, Humpback Whale, Fruit Fly, and Octopus was demonstrated.

In [20], natural markings of manta rays were used for pose invariant re-identification. The method uses a CNN ap-

proach with the semi-hard triplet mining strategy, the triplet loss function, and an extensive geometric augmentation of the input images. The method achieved 65 % Top-1 accuracy and 97 % Top-10 accuracy. However, it should be noted that the method requires the user input to localize the region of interest.

In [15], a method that combines the CNN baseline with pose-estimation to detect and re-identify Amur tigers was presented. It achieved the Top-5 accuracy of 90%. In [18], a three-module deep CNN architecture in order to learn complementary, non-obvious features as well as obvious ones was used for the Amur tiger identification. The first module learns embeddings from an image as usual, the second module utilizes the same architecture, but receives an image with removed parts that correspond to areas of interest of the first network, and the third module combines their embeddings for the final result. The method achieves the Top-5 accuracy of 91,6%. In [17], Amur tiger images were separated into streams that are utilized by different embedding networks: trunk (body) parts and limb parts. Trunk feature vectors are learned by a network with 8 vertical stripes, while limb feature vectors are learned with multiple-branch network for different limb components. The method demonstrates the best result on the proposed dataset with the Top-5 accuracy of 95,3%.

In [27, 5, 21], the re-identification of the Saimaa ringed seals was considered. In [27], a superpixel based segmentation method and a simple texture feature based ringed seal identification method were presented. In [5], additional pre-processing steps were proposed and two existing species independent individual identification methods were evaluated. However, the identification performance of neither of the methods is good enough for most practical applications. In [21], the re-identification of the Saimaa ringed seals was formulated as a classification problem and was solved using transfer learning. While the performance was high on the used test set, the method is only able to reliably perform the re-identification if there is a large set of examples for each individual. Furthermore, the whole system needs to be retrained if a new seal individual is introduced. Finally, it is unclear if the high accuracy was due to the methods ability to learn the necessary features from the fur pattern, or if it also learned features such as pose, size, or illumination that separated individuals in the used dataset, but do not provide the means to generalize the methods to other datasets.

2.2. Siamese networks

The task of re-identification of people and animals [16, 4, 8] could be formulated in terms of learning a distance metric between individuals. The general strategy is to train a model to discriminate between a collection of same/different pairs. Since the model learns generic embeddings rather than a rigid classifier it is able to better generalize to new classes

that have not been used during the training.

The triplet neural network [12, 25] is an extension of the Siamese Neural Network where three input samples are simultaneously considered in the loss function. The goal is to learn an embedding such that the distance between similar embedded samples is closer than the distance between dissimilar samples.

During the training, three inputs are sampled and run through the same embedding net: anchor x_a , positive x_p , and negative x_n samples. The loss function is calculated using the embedded representations of those samples as follows:

$$L_{triplet}(x_a, x_p, x_n) = \max(0, m + \left\| f(x_a) - f(x_p) \right\|_2^2 - \left\| f(x_a) - f(x_n) \right\|_2^2), \quad (1)$$

where $f(\cdot)$ is the embedding network, and m is some margin. This loss function turns to zero when the positive distance is smaller than the negative by more than a specified margin. If the difference is smaller than the margin, or if the positive distance is larger than the negative, the loss is non-zero.

In [11], the concept of triplet mining is being discussed. Triplet mining refers to a strategy of selecting triplets for training. The issue is that the number of triplets grows cubically with the size of the dataset, and most of those triplets eventually become useless for learning. If one tries to learn the concept of the "same individual" then being shown pictures of individuals with different clothes over and over again does not improve the learning. However, being shown similar-looking but different individuals, or the same individual in different poses, improves the learning dramatically. Hard triplet mining strategy aims to achieve that by only giving the network examples hard negative examples (similar-looking but different) and hard positive examples (differently-looking but similar). The issue with hard triplet mining is that the network is essentially learning on a heavily outlier-biased selection of triplets which can result in poor performance on comparatively "easy" tasks. Semi-hard strategy alleviates that by including "moderate" examples on either negative side, positive side, or both.

3. Saimaa ringed seal re-identification

The proposed re-identification process for the Saimaa ringed seals is shown in Fig. 2. First, the seal is segmented from the background. This step is crucial since most of the images are obtained using static camera traps. Therefore, the same seal is often captured with the same background increasing the risk that the supervised identification algorithm learns to "identify" the background instead of the actual seal if the full image or the bounding box around the seal is used. This may result in a system that is unable to identify the seal in a new environment. After segmentation,

the seal is cropped to bounding box and the pelage pattern is extracted. The region in the pattern image corresponding to the seal segment is then divided into small patches. Finally, the identification is performed by finding the most similar patches in the patch database of the known seals.

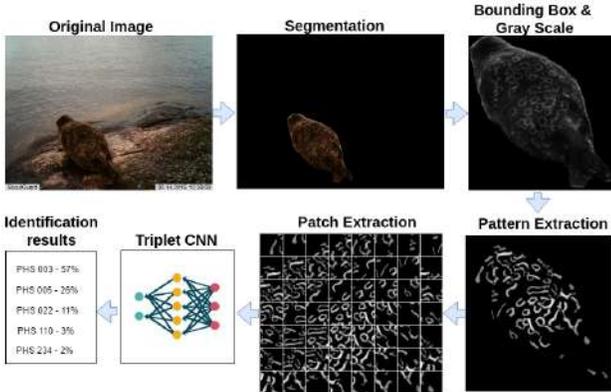


Figure 2. Proposed identification algorithm.

3.1. Segmentation

To segment the seal from the background the Deeplab model [6] is used. DeepLab is a state-of-the-art deep learning model for semantic image segmentation. It contains three main advantages compared to the competing methods. First, It uses atrous convolution which is a powerful tool in dense prediction tasks. Atrous convolution allows to explicitly control the resolution at which feature responses are computed within deep CNNs. Second, atrous spatial pyramid pooling (ASPP) allows to robustly segment objects at multiple scales. ASPP probes an incoming convolutional feature layer with filters at multiple sampling rates and effective fields-of-views, thus capturing objects as well as image context at multiple scales. Third, the localization of object boundaries is improved by combining methods from deep CNNs and probabilistic graphical models. The commonly deployed combination of max-pooling and downsampling in deep CNNs achieves invariance, but has a toll on localization accuracy. The method overcomes this by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF) which is shown both qualitatively and quantitatively to improve localization performance.

To make sure that the pelage pattern is fully covered in the seal segment, two additional postprocessing steps are applied to segmentation maps to close holes and to smooth the borders. In order to close the holes in the pattern (both internal and external) we apply sliding window convex hull on a condition that pattern in a current window is not connected. The image is broken down into many small heavily overlapping square windows. Each is checked for connectivity, and if some part of the current window is not con-

nected then we paint it over with convex hull. This way, the overall smooth structure of the seal outline remains undisturbed, but deep external holes get closed. This allows us to preserve a general concave outline of a seal, while closing potential open holes in the pattern. Smoothing of the image is a simple two-pass Gaussian filter with thresholding to keep the mask binary. Smoothing gets rid of blocky artifacts produced by our hole closing algorithm.

3.2. Pelage pattern extraction

Images of the Saimaa ringed seals are very diverse. They are taken from various distances and viewing angles. The Saimaa ringed seals are characterized by low mobility, but large variability of poses. The fur pattern covers the entire body of the animal and does not have any specific location. In addition, due to the fact that the images were obtained from the camera traps in different weather conditions and with different lighting, there is a significant amount of excess noise.

In order to remove noise, to avoid learning of the superficial characteristics, and to reduce the amount of data needed to train the identification algorithm, the pelage pattern is extracted from the segmented seal images. The pattern extraction algorithm is based largely on the tubeness filter with processing steps to increase fidelity and consists of the following steps:

1. Sato tubeness filter. This filter can be used to detect continuous ridges (tubes, wrinkles, rivers, etc.). It is well-suited for the Saimaa ringed seals pattern extraction since their patterns are mostly continuous ridges that form rings and other shapes.
2. Unsharpening using a mask with a radius of 5 and an intensity of 25. This operation makes the results sharper.
3. Removing segmentation border which heavily influences the Sato filter. This step is necessary since the segmentation border is detected as a "tube", but it does not belong to the pattern.
4. Morphological opening using a disk structuring element with a radius of 3. This operation allows us to remove small artifacts from the grayscale image.
5. Adaptive histogram normalization This operation is performed in order to make the image brighter without losing details.
6. Otsu's thresholding and zeroing out pixels below it. This makes pattern edges well-defined while still keeping pattern smooth.
7. Morphological opening using a disk structuring element with a radius of 3. This operation is also repeated, and this time it removes artifacts left from thresholding.

8. Unsharpening using a mask with a radius of 5 and an intensity of 2. This unsharpening mask is weaker than before. It only needs to slightly sharpen the image after opening in order to keep the pattern well-defined and well-contrasted with black areas.

The result of pattern extraction is a grayscale image with an explicit pattern outline as shown in Fig. 3.

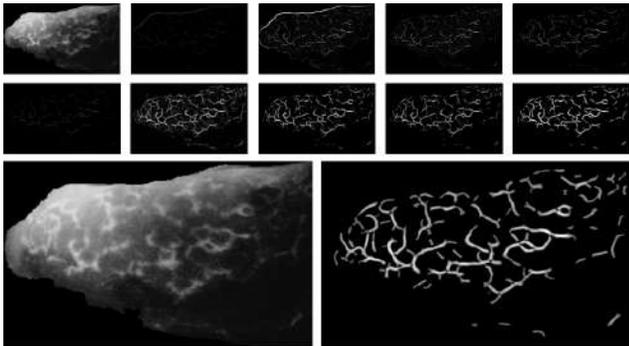


Figure 3. Visualization of pattern extraction result. First row: Steps 1–4 of the algorithm (from the left to the right). Second row: Steps 5–9 of the algorithm (from the left to the right). Third row: the source image (left) and the end result of pattern extraction (right).

3.3. Patch matching

Another noticeable problem with the Saimaa ringed seal re-identification is that the visible part of the pelage pattern varies greatly between different images of the same seal. Therefore, the system should be robust to the seal pose and angle of viewing. To enable this, the pattern image is divided into patches that are then used to find corresponding patches in the known individuals. Dividing the pattern into patches also helps to keep the size of the network used for matching compact. Before the patch extraction, the pattern segment is cropped and scaled to the common size to make scale between images similar relative to the size of the seal. Then overlapping patches with the common size (in our case 160×160 pixels with 50% overlap) are extracted. Finally, the pattern patches with less than 10% non-black pixels are removed.

Triplet Neural Network [12] is used to calculate the similarity between two patches. The network itself consists of 2 halves: a convolutional part and a fully-connected part. We augment the network with rotation-invariance pass. Each image is rotated using a set of predefined angles (we use -30, -20, -10, 0, 10, 20, 30 degrees). Then each rotated version of the same image is passed through the convolutional part of the network, and the results are summed together before being passed to the fully connected part.

Let g be the convolutional part of the network, f is fully connected part, Θ is the set of predefined rotation angles

and x_θ is the image rotated by the angle θ then the result of rotation-invariance pass can be described as such:

$$y(x) = f\left(\sum_{\theta \in \Theta} g(x_\theta)\right) \quad (2)$$

For triplet networks, the training and the evaluation processes differ considerably. During the training, the network receives three samples at a time: anchor, positive, and negative. The anchor is a base image (pattern patch), the positive is a sample of the same class (patch from the same pattern of the same seal) as the anchor, and the negative is a sample of a different class (patch from a different seal) from the anchor. The objective of the network is to encode these samples in a way that the L_2 metric distance between the anchor and the positive is smaller (by a pre-defined margin) than the distance between the anchor and the negative (Fig. 4).

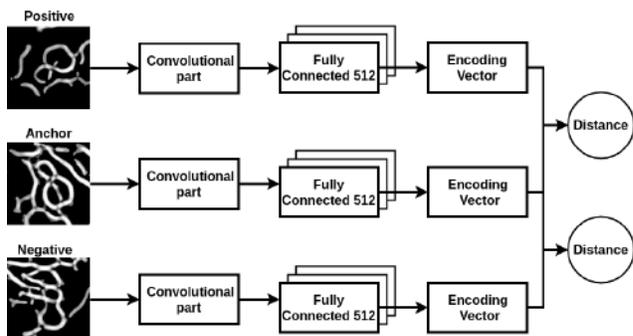


Figure 4. Triplet Network training for patch matching.

When the trained network is applied to the matching task it takes a sample (pattern patch) as the input and produces an encoding vector (feature vector). Our encoding vector consists of 512 dimensions which is the same number as the number of outputs of the CNN. These encodings can be compared using L_2 metric which makes it straightforward and fast to compute distances between them. The patch correspondences can be found by comparing the encodings of patches from the query image to all the patches from the labeled images in the database. Those correspondences provide the basis for the re-identification. It should be noted that the encoding vectors need to be computed only once for each patch enabling efficient computation.

3.4. Individual re-identification

The goal of the individual re-identification algorithm is to predict a seal identifier (unique for each individual animal) given a query image and a gallery of known individuals. The gallery contains a small number of distinct, high quality pattern images of each seal captured from different sides in order for us to be able to reliably perform comparisons with query images. It is used to construct a pattern

patch database which consists of the encoding vectors and the seal identifiers for each patch.

The proposed re-identification algorithm ranks the similarity between two images (the query image and a gallery image) and can be divided into the three main steps: 1) patch-similarity heatmap generation to select candidates of corresponding patches in the gallery image, 2) candidate filtering using topology-preserving projections, and 3) candidate ranking. The patch similarity heatmaps are generated by dividing both images into patches and comparing each query image with all gallery image patches. Local minima of the heatmap (high similarity regions) are found and used as projection candidates.

Candidate filtering is performed by selecting the projection candidates that preserve topological relations between original patches. We use a simple angle-based method to calculate the topological consistency in the candidate projection. The algorithm calculates angles for each three consecutive projection points and compares them to the same angles between patch center points. The total angle difference is the rank (lower is better) of the topological similarity. Finally, the ranking is obtained by calculating the average weight of topologically similar projections and by selecting the one with the lowest average weight. This weight is the total rank of similarity between the query image and the gallery image (see Fig. 5).

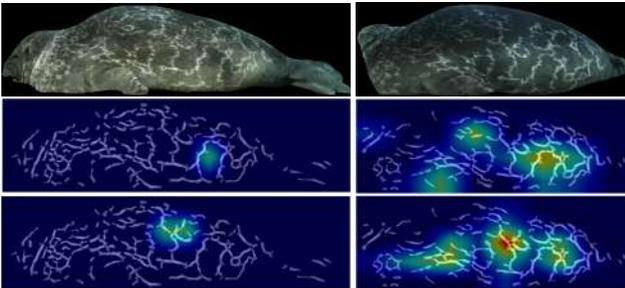


Figure 5. Examples of the region similarity heatmaps: the query image (left) and gallery image (right). Heatmaps for a query image highlight a single region that is being compared to the entire gallery image. Heatmaps for the gallery image show regions which are most similar to the highlighted region from the query image.

4. Experiments

The experiments were performed using a challenging database of the Saimaa ringed seals collected using automatic camera traps (see Fig. 6).

4.1. Segmentation

4.1.1 Data

A large annotated dataset is needed to train the Deeplab model. Annotating the dataset of this size manually is very labor-intensive, so a heuristical, semi-automatic approach



Figure 6. Examples of the seal images.

to generate segmentation ground truth was utilized. First, a pretrained model able to segment similar animals was applied for the dataset of more than 308 846 seal images to obtain a subset of well-segmented seals. The results were checked first automatically by discarding fully black images, the segmentation results with low pixel density, and the segmentation results which took up more than 50% of the image. Then, the rest were filtered manually to remove incorrectly segmented images. In addition, 100 images were manually segmented to complement the dataset with those images that were difficult to segment for the pretrained model. The resulting dataset contained nearly 100 000 well-segmented, high-definition images with binary segmentation masks. The dataset was split as follows: 56 000 images for training, 21 000 images for validation, and 22 800 images for testing.

4.1.2 Results

The Deeplab model was trained using the Tensorflow deep learning framework [1]. A model pretrained on the Pascal VOC dataset [9] was used and transfer learning was applied changing the last network level to separate only two classes: the seal and the background. Examples of the segmentation results are shown in Fig. 7.

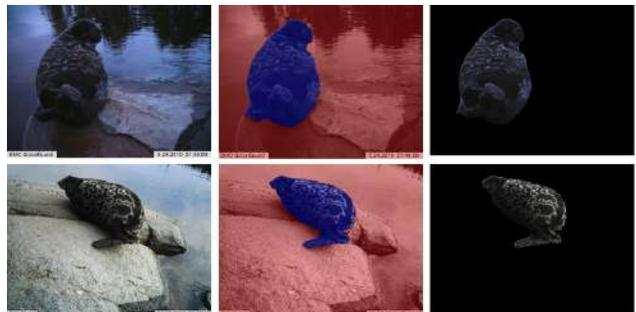


Figure 7. Examples of segmentation results.

Intersection over Union (IoU) between the segmentation result and the ground truth was used as a metric to evaluate the results. The mean IoU over all images in the test set

was 82% without postprocessing the segments. With postprocessing the mean IoU of 91% was achieved. Fig. 8 illustrates the IoU distribution of the segmented images before and after postprocessing. More than 75% of images have IoU more than 90% and 20% have more than 95%.

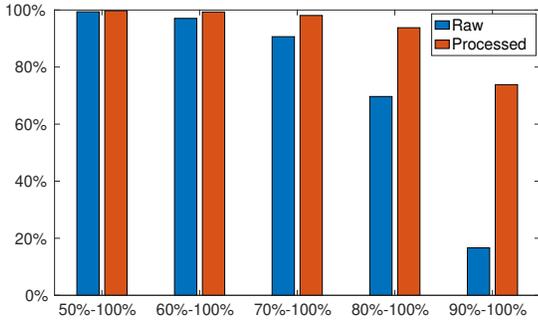


Figure 8. The impact of postprocessing on segmentation quality. The horizontal axis represents the IoU bucket while the vertical axis is the percentage of images that fell into said IoU bucket after segmentation with and without postprocessing.

4.2. Patch matching

4.2.1 Data

In order to train the triplet network for patch matching, a dataset of 3000 different labeled patches belonging to 26 different classes was collected. Each class corresponds to one manually selected location in the pelage pattern of one seal, and each sample from one class was extracted from different images of the same seal. Thus, the class is defined here as the spatially matching corresponding patches in the images of the considered individual seal, not as the class of the individual seal. The dataset was further augmented by random rotations, scaling, and shifts (Fig. 9).

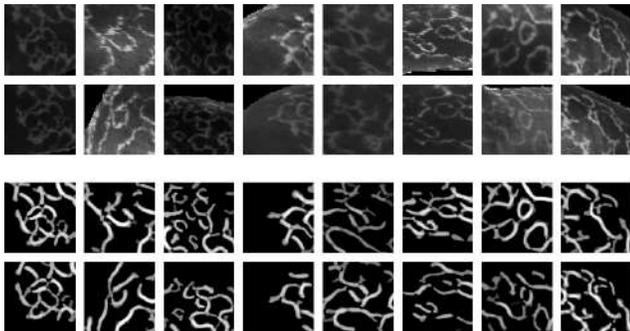


Figure 9. Examples of patches. Original patches (top two rows) and the corresponding pattern patches (bottom two rows).

A different, unrelated set of patches was used to test the method. The dataset of 1500 patches belonging to 28 different classes was collected. None of the classes in the testing

dataset was encountered during the training phase. A patch size of 160 pixels was used.

4.2.2 Results

To demonstrate the usefulness of the proposed pelage pattern method, the triplet network was tested on both the original patches extracted from the segmented images and the processed patches extracted from pattern images. The results are presented in Table 1.

Table 1. Patch matching results.

	Top-1	Top-2	Top-3	Top-4	Top-5
Pattern patches	74.6%	79.2%	81.1%	84.5%	87.0%
Original patches	66.4%	73.0%	76.9%	80.5%	82.5%

The results indicate that the pattern extraction clearly improves the results. This can be explained by the noise and large variation in appearance such as lighting conditions, pattern visibility, and shadows that are unavoidable on non-processed images. This noise prevents the network from generalizing well during the training, and makes it harder to extract relevant features during the evaluation.

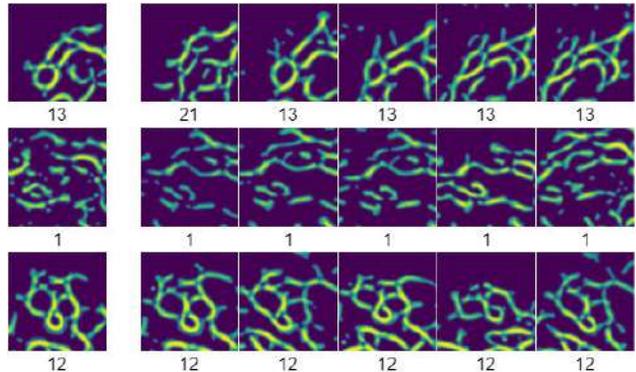


Figure 10. Examples of patch comparison: query patch (left) and best matches in the descending order of distance (right). The first example illustrates the case of robustness to the pattern deformation, the second example shows matching in the case of lost details, and the third shows invariance to small rotations.

Fig. 10 shows examples of patches being compared well despite differences in rotation thanks to the rotation-invariance pass of the network. Patches on the right exhibit different angles of rotation compared to the query patch, and yet they show up in the top-5 comparison.

4.3. Re-identification

4.3.1 Data

The dataset to evaluate the full re-identification framework consisted of 2 000 images of 46 unique seal individuals.

Each image contained one Saimaa ringed seal manually identified by a biologist. The dataset was split into the gallery set (the database of the known seals) and the query set (the test set) as follows: 500 images in the gallery set (on average 10 images per individual, with minimum of 3 and maximum of 12), and the rest (1500 images) in the query set (between 5 and 120 images per individual).

4.3.2 Results

The re-identification framework was tested with several variables. The first is pattern extraction. We performed tests with and without pattern extraction applied in order to quantify its effect on re-identification performance and training quality. The second is network rotation invariance. Our proposed rotation invariance pass was tested against the plain network in order to see how useful it is in the real-world images outside of synthetic tests. Finally, we implemented two different image re-identification strategies based on patch comparison. The first strategy is a simple "patch voting" based on the k nearest neighbors (KNN) classification. In this algorithm, each patch is classified with KNN as belonging to one of individual seals from the gallery set. Then each patch "votes" for the seal it belongs to, and the votes are weighted according to confidence metric. The second strategy is the proposed heatmap-based topologically-aware patch matching. The results are shown in Table 2.

Table 2. Re-identification results with different method variables. OG stands for original images, PAT for pattern extraction, ROT and noROT signify rotation invariance, KNN is KNN-based patch voting and TOP is topologically aware heatmaps

	Top-1	Top-2	Top-3	Top-4	Top-5
OG-noROT-KNN	50.3%	56.2%	60.9%	63.4%	65.6%
PAT-noROT-KNN	58.1%	64.6%	69.3%	74.0%	76.9%
PAT-noROT-TOP	62.7%	68.3%	72.1%	76.9%	83.7%
PAT-ROT-KNN	64.9%	70.5%	75.0%	80.1%	82.5%
PAT-ROT-TOP	67.8%	73.2%	77.2%	81.7%	88.6%

The results demonstrate that the pattern extraction makes a significant difference and increase the performance of the whole re-identification. Moreover, KNN-based patch voting is worse than topologically aware heatmaps at re-identification, especially with non-perfect patch comparison neural network. Finally, rotation invariance gives a small boost in performance and helps to alleviate some of the irregularities in the dataset.

Currently the task of identifying each seal is performed manually by experts which takes considerable time and effort. Because of this, the Top-1 accuracy is not the only important metric for the re-identification system. Limiting the choice to a set of best matches during manual identification is going to speed up the process significantly. High Top-5 accuracy can help experts with identification tasks while

still leaving them with a great degree of manual control. The method can adequately present good potential matches from which an expert can make an accurate conclusion much faster. Fig. 12 shows examples of re-identification results. It is evident that the proposed method is capable of handling complex cases when the pattern is only partially similar or rotated.



Figure 11. Examples of re-identification results: query images (left) and top 4 matches (right). Correct matches (same individual) are highlighted with green and incorrect ones with red.

5. Conclusions

In this paper, we proposed a framework to re-identify the Saimaa ringed seal individuals from camera-trap images for monitoring and conservation purposes. The framework consists of seal segmentation using the state-of-the-art Deeplab model, Sato tubeness filter based pelage pattern extraction method, the siamese network based pattern patch matching for finding patch correspondences, and a re-identification algorithm based on the patch similarities and topology-preserving projections. Our results show that the framework produces promising re-identification results, taking step towards fully automated re-identification system for the Saimaa ringed seals. The method provides a useful tool for conservation biologist by reducing the amount of manual work. The framework is species-agnostic and by replacing the pattern extraction step it can be applied to other animal species with similar pelage or fur patterns.

Acknowledgments

The research was carried out in 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. The authors would like to thank Vincent Biard, Marja Niemi, and Mervi Kunasranta from Department of Environmental and Biological Sciences at University of Eastern Finland for providing the data for the experiments and expert knowledge for identifying the individuals.

References

- [1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghe-

- mawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [2] L. Bergamini, A. Porrello, A. C. Dondona, E. Del Negro, M. Mattioli, N. D’alterio, and S. Calderara. Multi-views embedding for cattle re-identification. In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 184–191. IEEE, 2018.
- [3] T. Y. Berger-Wolf, D. I. Rubenstein, C. V. Stewart, J. A. Holmberg, J. Parham, S. Menon, J. Crall, J. Van Oast, E. Kiciman, and L. Joppa. Wildbook: Crowdsourcing, computer vision, and data science for conservation. *arXiv preprint arXiv:1710.08880*, 2017.
- [4] C.-A. Brust, T. Burghardt, M. Groenenberg, C. Kading, H. S. Kuhl, M. L. Manguette, and J. Denzler. Towards automated visual monitoring of individual gorillas in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2820–2830, 2017.
- [5] T. Chehrsimin, T. Eerola, M. Koivuniemi, M. Auttila, R. Levänen, M. Niemi, M. Kunasranta, and H. Kälviäinen. Automatic individual identification of saimaa ringed seals. *IET Computer Vision*, 12(2):146–152, 2018.
- [6] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv:1802.02611*, 2018.
- [7] J. Crall, C. Stewart, T. Berger-Wolf, D. Rubenstein, and S. Sundaresan. Hotspotter - patterned species instance recognition. *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 230–237, 2013.
- [8] D. Deb, S. Wiper, S. Gong, Y. Shi, C. Tymoszek, A. Fletcher, and A. K. Jain. Face recognition: Primates in the wild. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10. IEEE, 2018.
- [9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>, 2012.
- [10] K. M. Halloran, J. D. Murdoch, and M. S. Becker. Applying computer-aided photo-identification to messy datasets: a case study of thornicroft’s giraffe (giraffa camelopardalis thornicrofti). *African Journal of Ecology*, 53(2):147–155, 2015.
- [11] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.
- [12] E. Hoffer and N. Ailon. Deep metric learning using triplet network. In *International Workshop on Similarity-Based Pattern Recognition*, pages 84–92. Springer, 2015.
- [13] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.
- [14] M. Koivuniemi, M. Kurkilahti, M. Niemi, M. Auttila, and M. Kunasranta. A mark–recapture approach for estimating population size of the endangered ringed seal (phoca hispida saimensis). *PLoS one*, 14(3):e0214269, 2019.
- [15] S. Li, J. Li, W. Lin, and H. Tang. Amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*, 2019.
- [16] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 152–159, 2014.
- [17] C. Liu, R. Zhang, and L. Guo. Part-pose guided amur tiger re-identification. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.
- [18] N. Liu, Q. Zhao, N. Zhang, X. Cheng, and J. Zhu. Pose-guided complementary features learning for amur tiger re-identification. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.
- [19] M. Matthé, M. Sannolo, K. Winiarski, A. Spitzen-van der Sluijs, D. Goedbloed, S. Steinfartz, and U. Stachow. Comparison of photo-matching algorithms commonly used for photographic capture–recapture studies. *Ecology and evolution*, 7(15):5861–5872, 2017.
- [20] O. Moskvayak, F. Maire, A. O. Armstrong, F. Dayoub, and M. Baktashmotlagh. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. *arXiv preprint arXiv:1902.10847*, 2019.
- [21] E. Nepovinskyh, T. Eerola, H. Kälviäinen, and G. Radchenko. Identification of saimaa ringed seal individuals using transfer learning. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 211–222. Springer, 2018.
- [22] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25):E5716–E5725, 2018.
- [23] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical image analysis*, 2(2):143–168, 1998.
- [24] S. Schneider, G. W. Taylor, S. Linquist, and S. C. Kremer. Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. *arXiv preprint arXiv:1902.09324*, 2019.
- [25] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1386–1393, 2014.
- [26] X. Yu, J. Wang, R. Kays, P. Jansen, T. Wang, and T. Huang. Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, 2013(1):52, 2013.
- [27] A. Zhelezniakov, T. Eerola, M. Koivuniemi, M. Auttila, R. Levänen, M. Niemi, M. Kunasranta, and H. Kälviäinen. Segmentation of saimaa ringed seals for identification pur-

poses. In *Proceedings of International Symposium on Visual Computing*, pages 227–236, Las Vegas, USA, 2015.