NORPPA: NOvel Ringed seal re-identification by Pelage Pattern Aggregation

Ekaterina Nepovinnykh, Tuomas Eerola, Heikki Kälviäinen
Computer Vision and Pattern Recognition Laboratory (CVPRL)
Department of Computational Engineering
School of Engineering Sciences
Lappeenranta-Lahti University of Technology LUT, Lappeenranta, Finland
firstname.lastname@lut.fi

Ilia Chelak
Department of Computer Science, Faculty of Science,
University of Helsinki, Helsinki, Finland
firstname.lastname@helsinki.fi

Abstract

*We propose a method for Saimaa ringed seal (*Pusa hispida saimensis*) re-identification. Access to large image volumes through camera trapping and crowdsourcing provides novel possibilities for animal conservation and monitoring and calls for automatic methods for analysis, in particular, when re-identifying individual animals from the images. The proposed method NOvel Ringed seal re-identification by Pelage Pattern Aggregation (NORPPA) utilizes the permanent and unique pelage pattern of Saimaa ringed seals and content-based image retrieval techniques. First, the query image is preprocessed, and each seal instance is segmented. Next, the seal's pelage pattern is extracted using a U-net encoder-decoder based method. Then, CNN-based affine invariant features are embedded and aggregated into Fisher Vectors. Finally, the cosine distance between the Fisher Vectors is used to find the best match from a database of known individuals. We perform extensive experiments of various modifications of the method on challenging Saimaa ringed seals re-identification dataset. The proposed method is shown to produce the best re-identification accuracy on our dataset in comparisons with alternative approaches.*

1. Introduction

Image-based individual re-identification of animals has recently gained significant attention due to the availability of large volumes of wildlife image data from automatic game cameras and citizen science projects. Automated re-identification methods have clear advantages over traditional methods, such as tagging, as they offer a non-invasive approach to monitor endangered species without causing stress or behavior changes [37]. The benefits of this technique are demonstrated by the valuable data it provides for conservation efforts, including accurate population size estimates and novel information on animal migration and behavior patterns [3, 26].

In this study, we focus on the Saimaa ringed seal (*Pusa hispida saimensis*), an endangered species endemic to Lake Saimaa in Finland, with a population currently numbering no more than 500 individuals. The conservation of this species necessitates the regular evaluation of population size, migration patterns, and behavior, as exemplified in [19–21]. To achieve this, the Photo ID method is employed, which entails the re-identification of each distinctive individual. The re-identification of Saimaa ringed seals is made feasible by their unique and enduring ring pattern that encompasses their entire body. Presently, re-identification is carried out manually, but it can be optimized through the utilization of computer vision-based methods.

A variety of methods for animal re-identification exist that utilize distinct characteristics in fur, feather and skin patterns [7, 13, 18, 22, 23, 31], and methods originally developed for human re-identification have been successfully applied to animals [1, 12, 14, 42]. Visual animal re-identification can be formulated as a task of finding a match for the given query image from a database of known individuals, which is equivalent to a content-based image retrieval (CBIR) problem [44] where an image is searched from a database based on the image content. However, despite the clear similarity between CBIR and re-identification tasks, utilizing utilization of CBIR approaches for animal re-identification has remained largely unstudied.

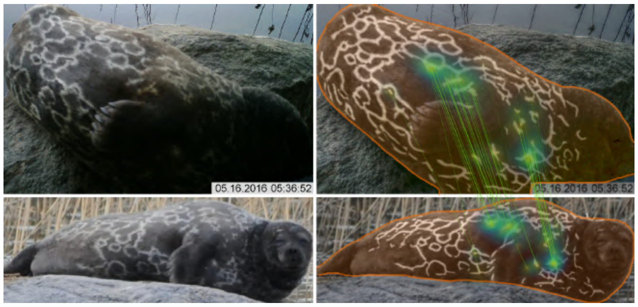


Figure 1. Visualisation of the proposed re-identification method. Input pictures are on the left and the results are on the right. The seal is segmented (orange outline), and matching regions of the pelage pattern are highlighted and connected with lines. The intensity of the highlights corresponds to the similarity of the matched regions.

Multiple methods have been proposed for ringed seal re-identification [10, 11, 32, 34, 35, 46]. Re-identifying individual Saimaa ringed seals from images is particularly challenging due to several factors: limited viewing angles, narrow range of poses and locations, biased image data collected using static game cameras, low sociality and high site fidelity of the seals, domain shift and database bias due to heterogeneous images captured by different cameras, and challenges posed by the seals' deformable nature, non-uniform pelage patterns, low contrast between ring pattern and fur, and large variation in possible poses. Compared to other animals such as zebras, which have clearly visible and distinct patterns with limited pose variation, re-identifying Saimaa ringed seals is considerably more challenging, which limits the accuracy of the existing methods.

In this paper, we address the above challenges by proposing the NOvel Ringed seal re-identification by Pelage Pattern Aggregation (NORPPA) method for automatic Saimaa ringed seal re-identification (Fig. 1). The proposed work is the first application of CBIR methods to the animal individual re-identification task to the best of the authors' knowledge. We further develop this approach by proposing an improved pattern feature embedding, which is done by utilizing affine invariant local CNN features and aggregating them into a fixed size embedding vector describing global features. The advantage of the system is that it does not have to be reconfigured or retrained in case new individuals are added to the database. In the experimental part of the work, we show that the proposed method outperforms previously developed re-identification methods for Saimaa ringed seals as well as HotSpotter [13]. In addition, different variations of the method are comprehensively evaluated to find the best pattern feature embeddings for the task. The code is available

The main contribution of this paper can be summarized as follows: (i) a novel Saimaa ringed seal re-identification method (NORPPA) inspired by content-based image retrieval methods, (ii) a novel combination of local affine-covariant region learning and CNN-based descriptors and feature aggregation to obtain a single fixed size pattern embedding vector with high discrimination power, and (iii) extensive evaluation of the method and its modifications on a challenging Saimaa ringed seal dataset. While the method was developed for Saimaa ringed seals, it is also possible to apply it to other patterned species as shown in [4].

2. Method

The proposed NORPPA method is illustrated in Fig. 2. It consists of four steps: 1) image prepossessing 2) feature extraction, 3) feature aggregation and 4) individual re-identification based on the aggregated vectors.

2.1. Image preprocessing

2.1.1 Tone-mapping

The images can have high contrast variation depending on the illumination conditions. This can cause a loss of detail in the region of interest, namely, the seal and its pelage pattern. To address this issue, we employ the tone-mapping approach to equalize the contrast in dark and bright image regions. The tone-mapping algorithm proposed in [25] is used due to its ability to produce realistic tone-mapped images without visual artifacts. The algorithm adjusts the contrast at different spatial frequencies using gradient methods with some extensions. These extensions prevent the reversal of global brightness levels and the loss of low-frequency details. Examples of images before and after prepossessing are presented in Fig. 3.

2.1.2 Seal instance segmentation

Seal instance segmentation is crucial because most of the images come from static camera traps. This, along with the fact that seal individuals tend to use same sites or areas inter-annually, makes it likely that one seal individual is captured repeatedly by the same camera. This may cause the supervised re-identification algorithm to learn to identify the background of the image instead of the actual seal if the full image or the bounding box is used. As a result, the algorithm may fail to identify the seal in a new environment.

Instance segmentation is performed using Mask R-CNN [16]. A segmentation model trained for Ladoga ringed seals from [32] is utilised. This is possible due to the two species being visually almost indistinguishable. Ladoga ringed seals are more numerous than Saimaa ringed seals and they are often captures in large groups which makes it easier to collect and annotate large training data for the segmentation. For more details about the instance segmentation model and training procedure see [32].

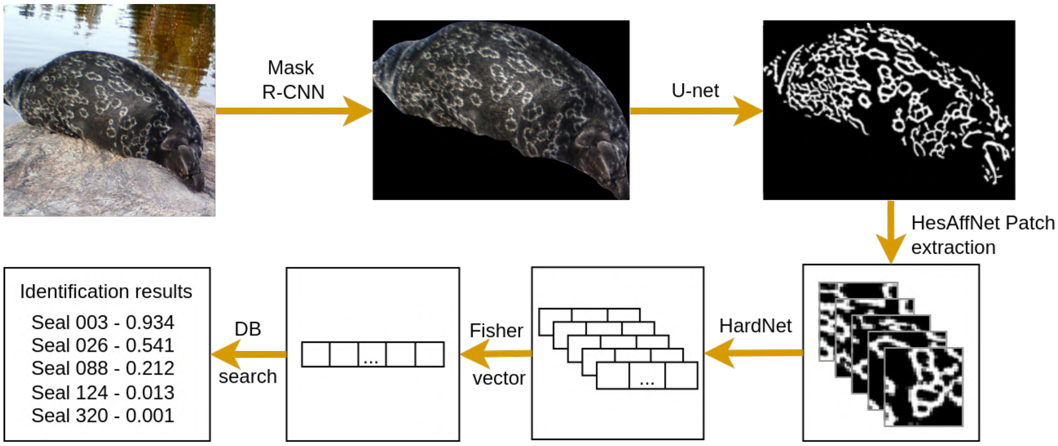


Figure 2. NORPPA re-identification pipeline.

Figure 3. Examples of the image processing of camera trap images. Images on the left are the originals. The right column demonstrates the result of the tone-mapping.

After the segmentation masks are obtained, additional morphological operations are applied to close the holes and smooth the borders by using morphological closing and opening. The examples of segmentation results are presented in Fig. 4.

2.2. Feature extraction

2.2.1 Pelage pattern extraction

The main distinguishing feature of a seal is its pelage pattern, which is both permanent and unique to each seal allowing the identification of individuals over their whole lifetime. The pelage pattern forms the basis for the proposed re-identification method. In order to focus the attention on the pattern and discard irrelevant information causing database bias such as illumination and other visual factors

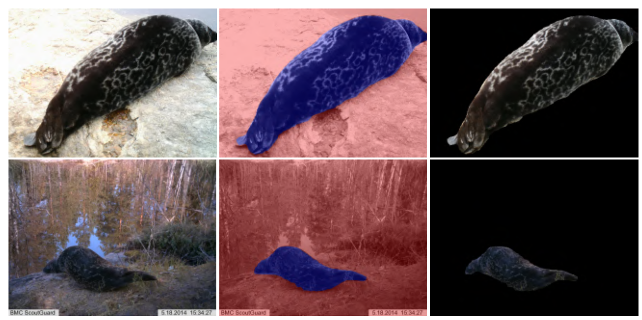


Figure 4. Examples of the segmentation masks. The images on the left are the originals. The mask is highlighted in blue and the background is highlighted in red on the middle images. The last column shows the result of the segmentation.

(e.g., wet fur looks different from the dry fur), the pattern is segmented. This is done using a CNN based method utilizing the U-net encoder-decoder architecture [40] that has been successfully applied to segmentation of other similar thin structures (e.g. veins in medical image data). The output of the method is a binarized image of the pelage pattern(see Fig. 5). The network has been trained and tested on the manually annotated dataset consisting of 520 images. The pattern image is further post-processed to remove small noise by using unsharp masking and morphological opening. All images are then resized in such way that the mean width of the pattern lines is the same for all images, bringing them into the same scale. This is necessary because the images are obtained from various sources and the image resolution has a large variation. For a more detailed explanation of the pattern extraction step, further training details, as well as the comparison to other segmentation methods, see [45].

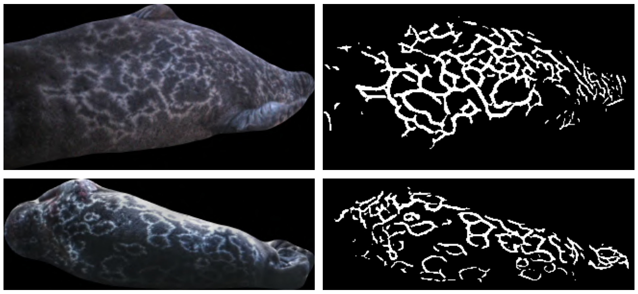


Figure 5. Example of pattern extraction output.

2.2.2 Local feature extraction

Seals can be found in a variety of poses. The deformable nature of seals body results in distorted and warped patterns on images. The pattern undergoes a non-linear transformation as a whole, but small local regions typically have similar affine transformations. Therefore, an affine invariant feature extractor is suitable for the task. For this purpose a combination of HesAffNet [30] detector and HardNet [29] descriptor is used.

The combination of a Hessian-Affine detector [28] with RootSIFT [2] used to be considered a gold standard for local feature extraction and description. However, with the increasing size of available datasets and rapidly developing field of deep learning, CNN-based methods are now able to outperform previous handcrafted features. The combination of HesAffNet [30] and HardNet [29] is able to provide state-of-the-art results in image retrieval tasks, which makes those methods particularly useful for animal re-identification as well.

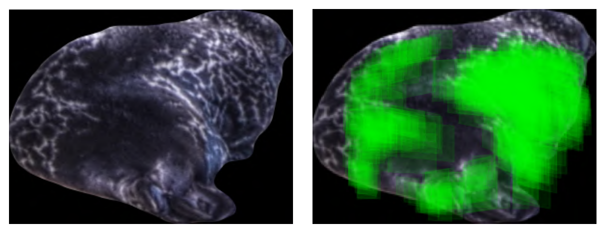
HesAffNet modifies the Hessian Affine Region detector [27, 28] by using the AffNet CNN for shape estimation. The detector finds regions of interest based on the Harris cornerness measure [15], which uses a second moments matrix to estimate the dominant gradient directions. It also uses the multiscale approach from [24], which finds extrema in the scale space by using Laplacian of Gaussian. This concept can extend to all affine transformations, not just scale. However, affine transformations have more degrees of freedom, which make the process more complex and need a special shape adaptation algorithm. The original Hessian Affine detector used Baumberg iteration [6], which HesAffNet replaces with an AffNet CNN.

AffNet and HardNet have a similar architecture and training procedure. HardNet is trained on batches of matching patch pairs, each with an anchor a_i and a positive match p_i. Each pair correspond to a different location, so there are no other matches except for the ones in each pair. The network encodes each patch, and computes a matrix of pairwise distances between all anchors and positive matches. For each pair, a closest non-matching descriptor from the batch is chosen, and a final hard negative margin loss is computed as

$$L = \frac{1}{n}\sum_{i=1}^{n} \max(0, 1 + d(a_i, p_i)) - \min(d(a_i, p_{j\,\min}), d(a_{j\,\min}, p_i)), \quad (1)$$

where $p_{j\,\min}$ is the closest non-matching positive to a_i, and $a_{j\,\min}$ is the closest non-matching anchor to p_i.

AffNet utilizes a slightly different training procedure, with the main difference being that it sets the derivative for the negative term in the loss to 0. This loss, called hard negative-constant, helps avoid the situations where a negative sample in the metric space prevents positive samples from moving closer together. The training procedure for AffNet is also more complex, since it learns affine shapes and not just a distance metric. It uses spatial transformers to transform input patches according to the predicted shape, and then feeds them into a descriptor network, such as HardNet. Only after that, it calculates and backpropagates the loss through both networks. Fig. 6 visualizes regions extracted using HesAffNet.



(a) (b)

Figure 6. Visualisation of Hessian Affine patch extraction: ((a)) segmented image; ((b)) HesAffNet-based patch extraction. Extracted regions are highlighted in green.

2.3. Feature aggregation

HardNet and HesAffNet produce a set of local region embeddings for each image. To obtain a single embedding for the whole image, the features are aggregated using Fisher Vector [17, 38, 39]. First, Principal Component Analysis (PCA) is applied to the resulting the feature embeddings to decorrelate the features and reduce the dimensionality. This is an important since Fisher Vectors are known to produce large descriptors. Principal components are learned using the images in the database of known individuals. Next, a visual vocabulary is constructed by applying Gaussian Mixture Model (GMM) to the features from the database. Fisher Vectors are created for each image by computing the partial derivatives of the log-likelihood function with respect to the GMM parameters and concatenating them. Kernel PCA [43] is applied to further reduce the dimensionality of the resulting image descriptors which reduces the storage requirements for the database and speeds up the database search for the re-identification.

2.4. Individual re-identification

Re-identification involves finding the individual with the lowest cosine distance from the Fisher vector of the query image. It is a common practice to use Euclidean distance or dot product [41], and since Fisher vectors are l_2-normalized, cosine distance and dot-product are essentially the same metrics. Heatmaps are generated by calculating the distance between all query and database patches. The inliers of the final homography are highlighted with ellipses that are directly proportional to their similarity. To handle previously unseen individuals, a threshold similarity value can be set. The algorithm considers a seal a new individual if the highest similarity score is below the threshold. Experts can verify the addition of new individuals to the database in a semi-supervised manner. No reconfiguration of the algorithm is required for the addition of new individuals.

To address novel (previously unseen) individuals, a threshold value for the similarity can be set. During re-identification, for a given query image, the top-k images with the smallest cosine distances between their Fisher vectors and the query's Fisher vector are returned. If the smallest distance is larger than a threshold, then the match is rejected, meaning the query contains a new individual. The addition of new individuals to the database does not require reconfiguration of any part of the algorithm. However, to keep the database of known individual clean, it is better to conduct this step in semi-supervised manner, where an expert verifies that the individual is not found in the database.

3. Experiments and results

3.1. Data

For the evaluation of the proposed method, publicly available SealID dataset have been used [33]. The dataset consists of 57 individual seals with a total of 2080 images (see Fig. 7). The dataset is divided into two subsets: database and query. The database subset contains a minimal number of high-quality unique images that are enough to cover the full body pattern of each seal. The query subset contains the remaining images and contains the same individuals as in the database.

To train and evaluate the patch embedding (feature extraction) and matching (finding the corresponding patch in other images) a separate dataset containing in total 4599 pattern image patches was utilized [11]. The training subset contains 3016 images and 16 classes. The testing subset contains 1583 images and 26 classes that are different from the training classes in the training set. Each class corresponds to one manually selected location in the pelage pattern of one individual seal (see Fig. 8). The images that were used to construct the dataset of pattern image patches are not included in the database and query subsets of the re-identification dataset.

Figure 7. Examples from the database and query datasets. Every row contains images of an individual seal. For every image from the query dataset (left) there is a corresponding subset of images from the database (right).

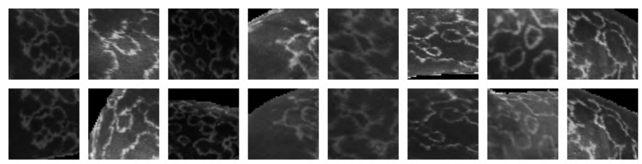


Figure 8. Examples of pattern image patches. The patches in the second row match the patches in the first row.

The accuracy is calculated as the ratio of correctly identified instances to the total number of instances from the query subset.

3.2. Feature extraction

The feature extraction step contains two differences compared to the previous version of the Saimaa ringed seal re-identification algorithm [34]. The first difference is that the region of interest detection approach uses the affine invariant regions (HesAffNet) instead of dense patches. The second difference is a switch to HardNet network to compute patch embedding. HardNet was compared to Triplet Network from [34] and ArcFace Network from [11]. To assess the necessity of each of these changes both modifications were evaluated separately. Hyperparameters for all versions of the algorithm were chosen using the Tree Parzen Estimator [8] algorithm. The results of the experiments are presented in Table 1.

As can be seen, both HesAffNet for region of interest detection and HardNet for patch embedding computation improve the accuracy noticeably. This finding leads to the conclusion that the dense patches approach cannot handle more general cases, whereas fine invariant features provide much needed robustness to various imaging conditions.

In order to evaluate the effect of the pelage pattern ex-

Table 1. Re-identification accuracy for different variants of the algorithm.

Patch extraction	Patch embedding	Top-1, %	Top-3, %	Top-5, %
Dense	Triplet [34]	52.06	60.36	65.70
	ArcFace [11]	39.94	50.06	56.67
	HardNet	52.18	61.70	67.27
HessAffNet	Triplet [34]	60.42	69.27	73.52
	ArcFace [11]	47.03	55.58	60.55
	HardNet	**77.64**	**82.97**	**85.27**

Table 2. Comparison of re-identification results by the NORPPA method on the SealID dataset with and without the pattern extraction step.

Input data	Top-1, %	Top-5, %	Top-10, %	Top-20, %
Original images	55.03	68.48	76.36	84.73
Pattern images	77.64	85.27	89.09	92.18

traction on the algorithm's accuracy, an ablation study has been performed. The results with and without the pattern extraction step are presented in Table 2. It is clear that the pelage feature extraction significantly increases the accuracy of the algorithm.

3.3. Patch embedding network

Training and fine-tuning of HardNet on different datasets were conducted in order to further improve the method.

The original HardNet was trained on the union of HPatches [5] and Brown [9] datasets. Typically, fine-tuning a machine learning model on domain-specific training data improves the method performance in a new domain. To test this on Saimaa ringed seal re-identification, we fine-tuned the HardNet model on patches of pelage pattern images. Fine-tuned models were compared to the pretrained model, a model trained from scratch on the pattern patches, and a model trained on the union of all datasets.

The results are presented in Table 3. For the training, all hyperparameters and random seeds were taken from the original implementation of HardNet [29].

While fine-tuning on the patches dataset improved the accuracy of the patch matching, the overall accuracy of the full-image matching dropped significantly. One possible reason is that the patches dataset was created using patches of the same scale, while the patches extracted by HesAffNet during the full re-identification algorithm vary in scale, leading to a different level of detail.

Training on the union of all datasets showed no considerable improvements. This result can be explained by the size of the pelage pattern patches dataset in comparison to

Table 3. Comparison of results for HardNet trained and fine-tuned on various datasets. We report mean with standard deviation.

Training	Patches TOP-1, %	Full TOP-1, %	Full TOP-5, %
Pattern patches	86.5	59.9	71.4
Brown +HPatches	93.02	**77.2**	**85.1**
Brown +HPatches +Pattern patches	**93.76**	70.7	80.5

the combined sizes of the Brown and HPatches datasets. In other words, since HardNet utilizes triplet sampling during the training stage, the probability of an image from the pelage pattern dataset appearing in the triplet is extremely small.

3.4. Qualitative evaluation

Visual examples of the re-identification results for the proposed NORPPA method are presented in Fig. 9. For the final version we use HardNet trained on Brown and HPatches datasets. Upon inspecting the results with highlighted areas, it is evident that the proposed method learns to perform the matching between query and database images based on the characteristics of the pelage pattern. Furthermore, it can be seen that the method is able to find the corresponding regions in the patterns in very challenging cases (Fig. 10).

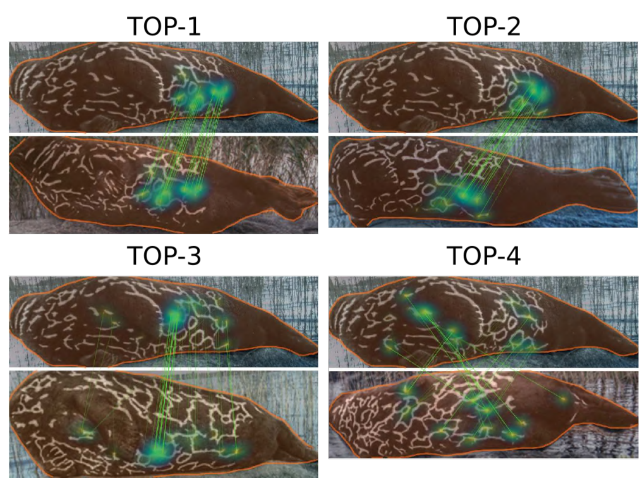
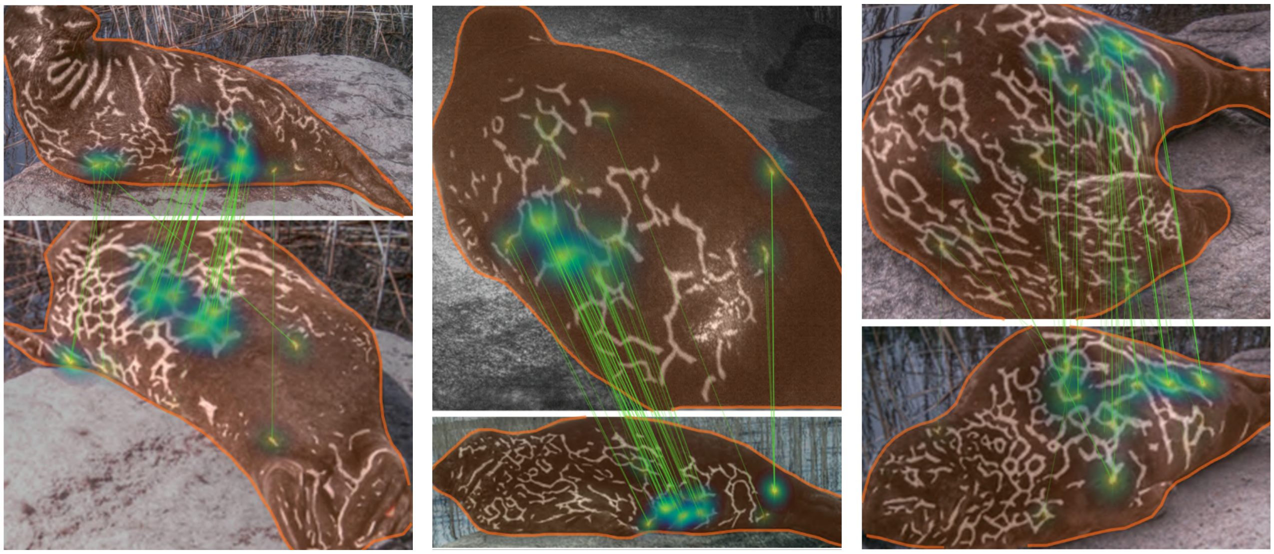


Figure 9. TOP-4 examples for the NORPPA method. For the given query image, the four best matches in decreasing order of similarity. Matched hotspots are highlighted in green. TOP-1–TOP-3 matches are correct. TOP-4 is incorrect.



(a) (b) (c)

Figure 10. Examples of some challenging cases. Top images are matched to the bottom images. The seal segmentation is shown in orange. The matching regions are highlighted and connected with green lines, the intensity corresponds to the similarity of a matched pair. The algorithm is able to match patterns even when the pose and point of view change significantly.

Table 4. Comparison of re-identification results between HotSpotter, NORPPA and previous iterations of the algorithm: SaimaaReID [34] and LadogaReID [32].

Method	TOP-1	TOP-3	TOP-5
SaimaaReID [34]	35.23%	44.61%	60.39%
LadogaReID [32]	39.94%	50.06%	56.67%
HotSpotter [13]	61.87%	63.63%	64.42%
NORPPA (ours)	**77.64%**	**82.97%**	**85.27%**

3.5. Quantitative evaluation

SaimaaReID [34], LadogaReID [32] without grouping step and NORPPA seal re-identification methods have been compared to HotSpotter [13], which is another method developed for patterned animal re-identification. HotSpotter is species-agnostic, and as such can be applied to Saimaa ringed seals as well. The results of NORPPA and HotSpotter for the Saimaa ringed seal dataset are presented in Table 4. It can be seen that the proposed method clearly outperforms HotSpotter based on TOP-1 accuracy. The difference is even more clear on TOP-5 accuracy, implying that even when NORPPA fails to correctly re-identify the seal, it is often able to provide a high rank for the correct match in the database. This is especially useful when the method is applied in a semi-supervised manner where the algorithm provides a set of possible matches for the expert to verify.

By considering a larger number of top matches, it is possible to further increase the chances of finding a correct individual. The plot of the top-k accuracy relative to the k value is presented in Fig. 11. The relationship for the NORPPA, SaimaaReID and LadogaReID methods is logarithmic in nature with fast growth for small k values, which slows down significantly with higher values. HotSpotter, on the other hand, exhibits almost no improvement after TOP-2 accuracy, with the difference between TOP-1 and TOP-5 accuracy being only about 2%, while the difference for NORPPA is almost 10%. The improvement in accuracy is a desirable property for a semi-automatic approach, offering a considerable accuracy improvement in exchange for a relatively small increase in the manual work required (as compared to a fully manual approach). Depending on the final application and available data, the relationship between the top-k accuracy and k can be used to determine the optimal number of matches to be returned by the algorithm.

4. Discussion

The proposed method has demonstrated promising results and can be utilized for the automatic re-identification of ringed seals. To mitigate the extreme dataset bias and to focus attention on the pelage pattern for re-identification, a series of steps were performed, including tonemapping of the original image, segmentation of seals from the background, and segmentation of the pelage pattern from the seal image. Furthermore, to address the issue of variability in the viewpoints and poses of the seals, local pattern patches were extracted from the images and pattern features were aggregated over the images. It should be noted

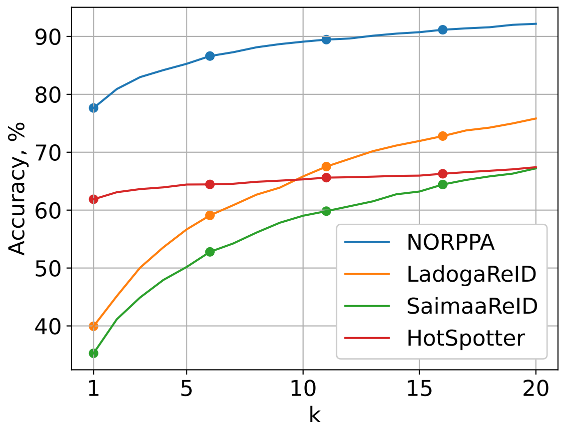


Figure 11. Plot of top-k re-identification accuracy for the proposed NORPPA method relative to k.

that the accuracy of re-identification is influenced by various factors, such as image quality, the distance between the camera and the seal, weather conditions, illumination, and the seal's pose.

The obtained TOP-1 accuracy of 77.6% motivates to utilize the proposed method in a semi-automatic manner, thereby noticeably reducing the manual labor. Expert verification is useful to ensure the accuracy and reliability of the results. Additionally, the incorporation of new individuals into the database does not require the reconfiguration of any of the system's components.

The proposed framework was evaluated on the challenging SealID dataset, which comprises high and poor-quality images captured using game and hand-held cameras. A significant advantage of the NORPPA method is that it does not require a large dataset for training, and it operates on the white-box principle, whereby the outcome of each step can be assessed and improved. For example, the pattern extraction step may not always eliminate grass occlusions, leading to incorrect matches with other images with the same distortion. The re-identification and ranking steps can be enhanced by incorporating geometric verification.

While the method was developed for Saimaa ringed seals, the future plans include its evaluation on other patterned animal species. The method is relatively easily transferable and has only 2 steps that require supervised training: segmentation and pattern extraction. While those steps benefit the accuracy of the method, they could be omitted, whereas the feature extraction and re-identification steps require unsupervised training only.

One additional benefit of the proposed method is that it allows features to be aggregated over multiple images. This opens interesting possibilities for further research as sequences of game camera images can be utilized to create a single descriptor for a larger portion of a pelage pattern by filling in the gaps created by obstructions and viewpoints. This has been demonstrated in [36].

5. Conclusion

A novel method for Saimaa ringed seal re-identification called NOvel Ringed seal re-identification by Pelage Pattern Aggregation (NORPPA) was proposed in this paper. The method utilizes pelage pattern extraction and feature aggregation inspired by content-based image retrieval techniques. The re-identification pipeline consists of image enhancement, seal instance segmentation by Mask R-CNN, U-net based pelage pattern extraction, pattern feature extraction, feature aggregation, and individual re-identification by database search.

Improved pattern feature embeddings were proposed by employing affine-invariant region of interest detection, CNN based feature descriptors, and Fisher Vector feature aggregation to obtain fixed size embedding vectors with high discriminative power. The proposed method was applied to a novel and challenging Saimaa ringed seal dataset and showed superior performance compared to HotSpotter and earlier versions of the Saimaa ringed seal re-identification method by the authors.

Acknowledgements

The authors would like to thank Raija ja Ossi Tuuliaisen Säätiö Foundation, the project CoExist (Project ID: KS1549) for funding the research. In addition, authors would like to thank Vincent Biard, Piia Mutka, Marja Niemi, and Mervi Kunnasranta from the Department of Environmental and Biological Sciences at the University of Eastern Finland (UEF) for providing the data of Saimaa ringed seals and their expert knowledge of identifying each individual.

References

[1] Mohit Agarwal, Sanchit Sinha, Maneet Singh, Shruti Nagpal, Richa Singh, and Mayank Vatsa. Triplet transform learning for automated primate face recognition. In *International Conference on Image Processing (ICIP)*, 2019. 1

[2] Relja Arandjelović and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 4

[3] Gonzalo Araujo, Abdul Ismail, Cat McCann, David McCann, Christine Legaspi, Sally Snow, Jessica Labaja, B. Manjaji-Matsumoto, and Alessandro Ponzo. Getting the most out of citizen science for endangered species such as Whale Shark. *Journal of Fish Biology*, 96:864–867, 2020. 1

[4] Ola Badreldeen Bdawy Mohamed. Metric learning based pattern matching for species agnostic animal re-

identification. Master's thesis, Lappeenranta-Lahti University of Technology LUT, Finland, 2021. 2

[5] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 6

[6] A. Baumberg. Reliable feature matching across widely separated views. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000. 4

[7] TY Berger-Wolf, DI Rubenstein, CV Stewart, J Holmberg, J Parham, and J Crall. Ibeis: Image-based ecological information system: From pixels to science and conservation. In *Bloomberg Data for Good Exchange Conference*, 2015. 1

[8] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for Hyper-Parameter Optimization. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2011. 5

[9] Matthew Brown and David G. Lowe. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision*, 74:59–73, 2007. 6

[10] Tina Chehrsimin, Tuomas Eerola, Meeri Koivuniemi, Miina Auttila, Riikka Levänen, Marja Niemi, Mervi Kunnasranta, and Heikki Kälviäinen. Automatic individual identification of Saimaa ringed seals. *IET Computer Vision*, 12:146–152, 2018. 2

[11] Ilia Chelak, Ekaterina Nepovinnykh, Tuomas Eerola, Heikki Kälviäinen, and Igor Belykh. Eden: Deep feature distribution pooling for saimaa ringed seals pattern matching. *Cyber-Physical Systems and Control II*, pages 141–150, 2023. 2, 5, 6

[12] Melanie Clapham, Ed Miller, Mary Nguyen, and Russell C Van Horn. Multispecies facial detection for individual identification of wildlife: a case study across ursids. *Mammalian Biology*, 102(3):921–933, 2022. 1

[13] J.P. Crall, C.V. Stewart, T.Y. Berger-Wolf, D.I. Rubenstein, and S.R. Sundaresan. Hotspotter - patterned species instance recognition. In *Winter Conference on Applications of Computer Vision (WACV)*, 2013. 1, 2, 7

[14] Debayan Deb, Susan Wiper, Sixue Gong, Yichun Shi, Cori Tymoszek, Alison Fletcher, and Anil K Jain. Face recognition: Primates in the wild. In *International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018. 1

[15] Christopher G. Harris and M. J. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, 1988. 4

[16] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *International Conference on Computer Vision (ICCV)*, 2017. 2

[17] David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the Fisher Kernel for Large-Scale Image Classification. In *European Conference on Computer Vision (ECCV)*, 2010. 4

[18] Bingliang Jiao, Lingqiao Liu, Liying Gao, Ruiqi Wu, Guosheng Lin, Peng Wang, and Yanning Zhang. Toward re-identifying any animal. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 1

[19] Meeri Koivuniemi, Miina Auttila, Marja Niemi, Riikka Levänen, and Mervi Kunnasranta. Photo-ID as a tool for studying and monitoring the endangered Saimaa ringed seal. *Endangered Species Research*, 30:29–36, 2016. 1

[20] Meeri Koivuniemi, Mika Kurkilahti, Marja Niemi, Miina Auttila, and Mervi Kunnasranta. A mark–recapture approach for estimating population size of the endangered ringed seal (Phoca hispida saimensis). *PLOS ONE*, 14:214–269, 2019. 1

[21] Mervi Kunnasranta, Marja Niemi, Miina Auttila, Mia Valtonen, Juhana Kammonen, and Tommi Nyman. Sealed in a lake – Biology and conservation of the endangered Saimaa ringed seal: A review. *Biological Conservation*, 253:108908, 2021. 1

[22] Jessy Lauer, Mu Zhou, Shaokai Ye, William Menegas, Steffen Schneider, Tanmay Nath, Mohammed Mostafizur Rahman, Valentina Di Santo, Daniel Soberanes, Guoping Feng, et al. Multi-animal pose estimation, identification and tracking with deeplabcut. *Nature Methods*, 19(4):496–504, 2022. 1

[23] Shuyuan Li, Jianguo Li, Hanlin Tang, Rui Qian, and Weiyao Lin. ATRW: A Benchmark for Amur Tiger Re-identification in the Wild. In *ACM International Conference on Multimedia*, 2020. 1

[24] Tony Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30:77–116, 1998. 4

[25] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception*, 3:286–308, 2006. 2

[26] Emer McCoy, Raul Burce, David David, Elson Aca, Jennifer Hardy, Jessica Labaja, Sally Snow, Alessandro Ponzo, and Gonzalo Araujo. Long-Term Photo-Identification Reveals the Population Dynamics and Strong Site Fidelity of Adult Whale Sharks to the Coastal Waters of Donsol, Philippines. *Frontiers in Marine Science*, 5:271, 2018. 1

[27] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *European Conference on Computer Vision (ECCV)*, 2002. 4

[28] Krystian Mikolajczyk and Cordelia Schmid. Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60:63–86, 2004. 4

[29] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2017. 4, 6

[30] Dmytro Mishkin, Filip Radenović, and Jiři Matas. Repeatability Is Not Enough: Learning Affine Regions via Discriminability. In *European Conference on Computer Vision (ECCV)*, 2018. 4

[31] Olga Moskvyak, Frederic Maire, Feras Dayoub, Asia O. Armstrong, and Mahsa Baktashmotlagh. Robust re-

identification of manta rays from natural markings by learning pose invariant embeddings. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2021. 1

[32] Ekaterina Nepovinnykh, Ilia Chelak, Andrei Lushpanov, Tuomas Eerola, Heikki Kälviäinen, and Olga Chirkova. Matching individual Ladoga ringed seals across short-term image sequences. *Mammalian Biology*, pages 1–16, 2022. 2, 7

[33] Ekaterina Nepovinnykh, Tuomas Eerola, Vincent Biard, Piia Mutka, Marja Niemi, Mervi Kunnasranta, and Heikki Kälviäinen. Sealid: Saimaa ringed seal re-identification dataset. *Sensors*, 22:7602, 2022. 5

[34] Ekaterina Nepovinnykh, Tuomas Eerola, and Heikki Kälviäinen. Siamese Network Based Pelage Pattern Matching for Ringed Seal Re-identification. In *Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2020. 2, 5, 6, 7

[35] Ekaterina Nepovinnykh, Tuomas Eerola, Heikki Kälviäinen, and Gleb Radchenko. Identification of Saimaa Ringed Seal Individuals Using Transfer Learning. In *International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, 2018. 2

[36] Ekaterina Nepovinnykh, Antti Vilkman, Tuomas Eerola, and Heikki Kälviäinen. Re-identification of saimaa ringed seals from image sequences. In *Scandinavian Conference on Image Analysis (SCIA)*, 2023. 8

[37] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115:5716–5725, 2018. 1

[38] Florent Perronnin and Christopher Dance. Fisher Kernels on Visual Vocabularies for Image Categorization. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 4

[39] Florent Perronnin, Yan Liu, Jorge Sánchez, and Hervé Poirier. Large-scale image retrieval with compressed Fisher vectors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 4

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015. 3

[41] Jorge Sánchez, Florent Perronnin, Thomas Mensink, and Jakob Verbeek. Image Classification with the Fisher Vector: Theory and Practice. *International Journal of Computer Vision*, 105:222–245, 2013. 5

[42] Stefan Schneider, Graham W Taylor, and Stefan C Kremer. Similarity learning networks for animal individual re-identification: an ecological perspective. *Mammalian Biology*, pages 1–16, 2022. 1

[43] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10:1299–1319, 1998. 4

[44] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000. 1

[45] Denis Zavialkin. CNN-based ringed seal pelage pattern extraction. Master's thesis, Lappeenranta-Lahti University of Technology LUT, Finland, 2020. 3

[46] Artem Zhelezniakov, Tuomas Eerola, Meeri Koivuniemi, Miina Auttila, Rikka Levänen, Marja Niemi, Mervi Kunnasranta, and Heikki Kälviäinen. Segmentation of Saimaa ringed seals for identification purposes. In *International Symposium on Visual Computing (ISVC)*, 2015. 2