International Journal of Computer Vision
https://doi.org/10.1007/s11263-024-02071-1

Species-Agnostic Patterned Animal Re-identification by Aggregating Deep Local Features

Ekaterina Nepovinnykh[1] · Ilia Chelak[2] · Tuomas Eerola[1] · Veikka Immonen[1] · Heikki Kälviäinen[1] · Maksim Kholiavchenko[3] · Charles V. Stewart[3]

Received: 2 September 2023 / Accepted: 27 March 2024

Abstract
Access to large image volumes through camera traps and crowdsourcing provides novel possibilities for animal monitoring and conservation. It calls for automatic methods for analysis, in particular, when re-identifying individual animals from the images. Most existing re-identification methods rely on either hand-crafted local features or end-to-end learning of fur pattern similarity. The former does not need labeled training data, while the latter, although very data-hungry typically outperforms the former when enough training data is available. We propose a novel re-identification pipeline that combines the strengths of both approaches by utilizing modern learnable local features and feature aggregation. This creates representative pattern feature embeddings that provide high re-identification accuracy while allowing us to apply the method to small datasets by using pre-trained feature descriptors. We report a comprehensive comparison of different modern local features and demonstrate the advantages of the proposed pipeline on two very different species.

Keywords Computer vision · Image processing · Animal biometrics · Re-identification · Ringed seals · Convolutional neural networks

1 Introduction

Animal biometrics, especially image-based individual re-identification, has recently gained extensive attention due to both its importance for ecology and conservation and the availability of large volumes of wildlife image data gathered via automatic game cameras and participatory science projects. The benefits of automated re-identification methods are evident as they allow valuable data for conservation efforts to be obtained, for example, accurate population size estimates and novel information about animal migration and behavior patterns (McCoy et al., 2018; Araujo et al., 2020). Compared to traditional methods such as tagging, which may cause stress and change the behavior of the animal, image-based re-identification offers a non-invasive technique for monitoring of endangered species (Norouzzadeh et al., 2018).

A fundamental challenge for animal identification is the problem of small labeled datasets. This arises in several variations. Firstly, there is an overall lack of images labeled with known individual ids. Generating ground truth animal ids for algorithm training requires a combination of (a) expertise, (b) good heuristics about appearance and location, (c) extensive searching, and (d) effective software tools (Kulits et al., 2021), making the generation of ground truth expensive, time-consuming, and focused on only the most charismatic species. Secondly, there is generally a long-tailed distribution in the number of sightings per individual animal, with many individuals seen just once or a few times, and fewer individuals seen frequently (see Fig. 10). This problem arises in part because of the just-enumerated difficulties in generating ground truth labels, and in part due to the inherent difficulty of obtaining the original data: some individuals are rarely in locations where images are acquired. Thirdly,

Communicated by Helge Rhodin.

✉ Ekaterina Nepovinnykh
ekaterina.nepovinnykh@lut.fi

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

[2] Department of Computer Science, Faculty of Science, University of Helsinki, P.O. Box 4, 00100 Helsinki, Finland

[3] Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Published online: 30 April 2024

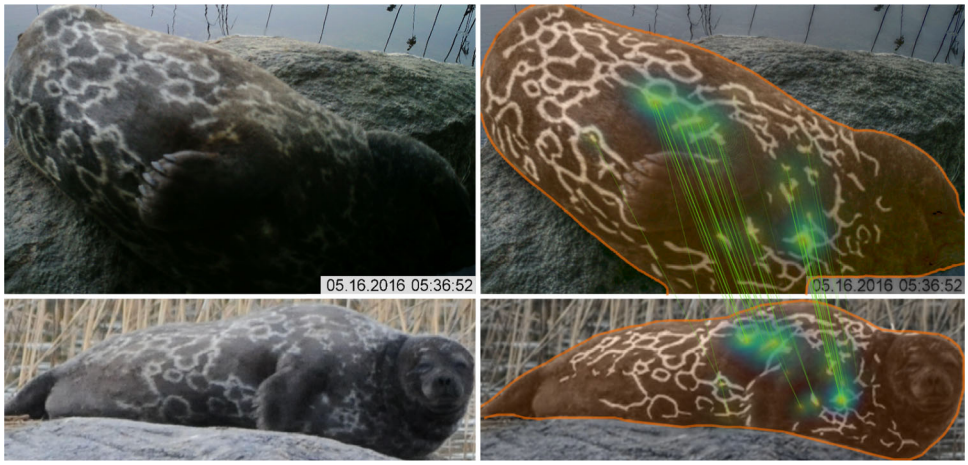

Fig. 1 Visualization of the proposed re-identification method called Aggregated Local Features for Re-Identification (ALFRE-ID). The input pictures are on the left and the results are on the right. The animal is segmented (orange outline), and the matching regions of the fur pattern are highlighted and connected with lines. The intensity of the highlights corresponds to the similarity of the matched regions

animal id is generally an open set identification problem: except in special circumstances (Christin, 2015), it is rare that an entire population is represented by the photos in the database. Hence, any set of images added to the database may show new individuals. Effectively addressing these concerns will significantly broaden the utility of animal identification.

A variety of methods for image-based identification exist that utilize distinct characteristics in fur, feather, and skin patterns (Crall et al., 2013; Berger-Wolf et al., 2015; Moskvyak et al., 2021a; Li et al., 2020) or adapt techniques developed for human face re-identification (Deb et al., 2018; Crouse et al., 2017; Agarwal et al., 2019). Traditional methods require the least prior information, and therefore in practice are still being used extensively (Berger-Wolf et al., 2017), but they are significantly limited in how they exploit any available training data. Methods that learn without identity labels require manually selecting the features—such as ear, fin and fluke contours Weideman et al. (2020)—and are limited by both the need for manual generation of feature training data and the ability to select these features in the first place. Finally, deep learning methods, which offer the most power and flexibility are data-hungry and therefore greatly challenged by the limited-data scenario that can occur for animal re-identification.

In this paper, we propose a pipeline that combines the best of these approaches. This is obtained by utilizing deep CNN-based local features and feature aggregation. We call the pipeline Aggregated Local Features for Re-Identification (ALFRE-ID). By aggregating learnable local features, it is possible to obtain representative pattern feature embeddings that provide high re-identification accuracy similar to deep metric learning-based methods. At the same time, the possibility of using pretrained local feature descriptors allows us to apply the method to small datasets much more accurately than end-to-end deep learning methods. The proposed pipeline is inspired by content-based image retrieval (CBIR) methods and builds on earlier work (Nepovinnykh et al., 2020) where Siamese networks were utilized to learn a similarity metric for local patches of pelage patterns. We further develop this approach by utilizing affine invariant local CNN features and aggregating them into a fixed-size embedding vector describing global features. The full re-identification pipeline consists of tone mapping, animal segmentation, feature extraction, computation of aggregated pattern feature embeddings, selection of potential matches by finding the most similar embeddings in the database of known individuals, and geometric verification and final match ranking by analyzing the spatial consistency of the pattern similarities (see Fig. 1). The pipeline follows a modular approach where individual techniques such as local feature extractors can be changed to address differences between animal species.

Our contributions are summarized as follows:

1. We propose a CBIR-motivated pipeline for individual animal identification called ALFRE-ID that includes interchangeable learned local features, feature aggregation, and feature embeddings to address the limitations of current methods, especially on small labeled datasets.
2. We experimentally demonstrate the advantages and tradeoffs of our pipeline in comparison to widely-used traditional methods based on non-learned, hand-crafted features (Hotspotter) and end-to-end deep learning methods.
3. We evaluate the pipeline's performance on two very different and challenging animal species showing trade-offs between various component options, demonstrating that a flexible pipeline of components is crucial for performance on small training datasets.

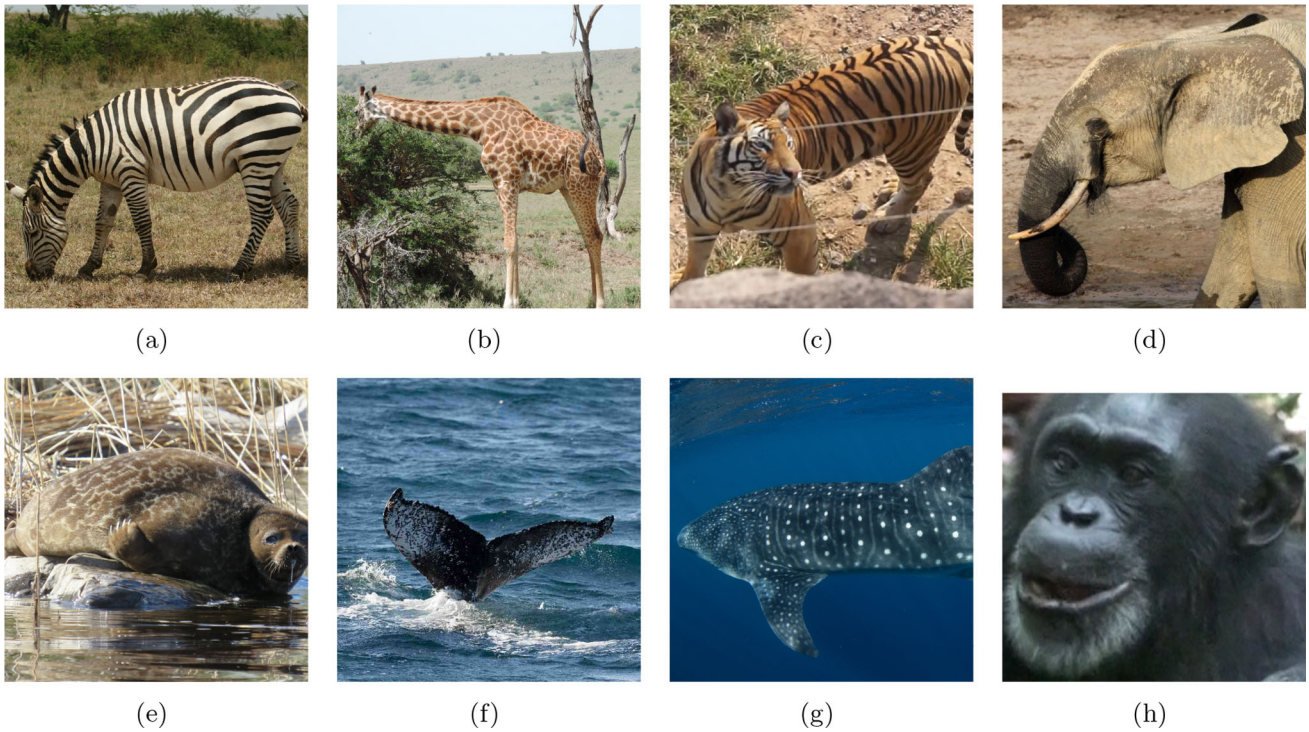

Fig. 2 Example images of the main identifiable features from publicly available re-identification data sets: **a** Plains zebra (*Equus quagga*) (Parham et al., 2017): stripe fur pattern; **b** Masai giraffe (*Giraffa tippelskirchi*) (Parham et al., 2017): spot fur pattern; **c** Amur tiger (*Panthera tigris*) (Li et al., 2020): stripe fur pattern; **d** African elephant (*Loxodonta africana*) (Korschens & Denzler, 2019): head shape; **e** Saimaa Ringed seal (*Pusa hispida saimensis*) (Nepovinnykh et al., 2022c): ringed fur pattern; **f** Humpback whale (*Megaptera novaeangliae*) (Cheeseman et al., 2017): fluke shape; **g** Whale shark (*Rhincodon typus*) (Holmberg et al., 2009): skin spot pattern; **h** Chimpanzee (*Pan troglodytes*) (Freytag et al., 2016): face

2 Related Work

2.1 Animal Re-identification

Animal re-identification is a broad term referring to the process of identifying an individual animal based on its features. The features are based on biological traits, and they can be captured in a number of ways, for example, acoustically (Hartwig, 2005; Pruchova et al., 2017) or visually in the form of images (Vidal et al., 2021) or videos (Zuerl et al., 2023). Currently, image-based methods are the most widely utilized approach due to the relative ease of data acquisition and manual analysis (Schneider et al., 2019).

Various animal species can be re-identified by different types of visually unique biological traits such as fur pattern, face, or fin shape. Examples of such traits are presented in Fig. 2. Re-identification methods can be divided into three categories: (1) traditional, non-learning methods that depend on hand-crafted local features, (2) methods that learn feature descriptions by manually selecting the biological traits, and (3) end-to-end deep learning methods. The first category consists of methods that extract and match hand-crafted local features such as SIFT (Lowe, 1999) between images and perform the re-identification typically by quantifying the similarity of the matching regions or the geometric consistency of the matched point pairs. For example, HotSpotter (Crall et al., 2013) is a SIFT-based re-identification algorithm that uses viewpoint invariant descriptors and a scoring mechanism that emphasizes the most distinctive key points, called "hot spots," on an animal pattern. A similar approach was proposed in Pedersen et al. (2023) where multiple local feature descriptors including SIFT, SURF, and SuperPoint were compared on giant sunfish re-identification. Lalonde et al. (2022) proposed to use transformer-based local features (Sun et al., 2021), instead of traditional hand-crafted features and a simple point correspondence confidence based matching criteria for blue whale re-identification. Algorithms in this category are species-agnostic and can be applied to wide-variety biological traits. The HotSpotter algorithm has been successfully used for re-identification of zebras (*Equus quagga*) (Crall et al., 2013) and giraffes (*Giraffa tippelskirchi*) (Parham et al., 2017), jaguars (*Panthera onca*) (Crall et al., 2013), ocelots (*Leopardus pardalis*) (Nipko et al., 2020), and leopards (*Panthera pardus*) (Suessle et al., 2023).

The second category of methods utilizes species-specific traits such as ear [e.g., Asian elephant (De Silva et al.,

2022)], fin [great white sharks (Hughes & Burghardt, 2017)], and fluke contours [e.g., humpback whale (Weideman et al., 2017, 2020)]. Both traditional feature-engineering based approaches and deep learning methods have been proposed to compute the feature (e.g., shape) representation for the selected traits. Examples of efficient algorithms for deep learning edge-based re-identification include CurvRank (Weideman et al., 2017), finFindR (Thompson et al., 2019, 2022), OC/WDTW (Bogucki et al., 2019) and the ArcFace-based method by Cheeseman et al. (2022). These methods have been applied to marine mammals such as bottlenose dolphins (*Tursiops truncatus*) (Tyson Moore et al., 2022; Thompson et al., 2019, 2022; Patton et al., 2023), humpback whales (*Megaptera novaeangliae*) Webber et al. (2023); Patton et al. (2023), right whales (*Eubalaena glacialis*) (Khan et al., 2022; Patton et al., 2023), and they use the unique shape of tail or fins to identify the animals. Similar deep learning methods have been also used to learn feature descriptors for cattle muzzle (Kumar et al., 2018) and primate faces (Deb et al., 2018; Brust et al., 2017). Since the methods in this category operate by quantifying the specific visual traits, distinguishing the individuals of the species of interest, they can be often trained without identity labels. However, this also makes the methods species-specific which limits their wider usability.

The third category consists of methods that utilize deep learning techniques such as convolutional neural networks (CNNs) to learn the feature embeddings or re-identification in an end-to-end manner without the need to manually select the biological traits to be used for the re-identification. These methods can be divided into classification and metric-based approaches (Vidal et al., 2021). The classification-based approaches (see e.g. de Silva et al. 2022) assume that the database of individuals is known and fixed, allowing the final algorithm to only identify individuals from that database. The metric-based methods (see e.g. Schneider et al. 2022), on the other hand, aim to learn a similarity metric between the input images. The re-identification is then performed by clustering or matching based on the similarity, which means that metric-based approaches are not limited by the initial database and can be applied to new individuals without retraining. Metric-based methods are generally preferred since obtaining the full dataset containing all individuals is practically impossible for any wildlife application. However, it should be noted that it is possible to extend the classification-based methods to tackle the open-set problem. For example, Kim et al. (2022) proposed to use a CNN-based classifier with the OpenMax layer to address the missing individuals in the training set.

Most recent methods for animal re-identification utilize deep learning, particularly CNNs (Schneider et al., 2019, 2020). CNNs have been successfully applied for re-identification of Amur tigers (Panthera tigris) (Li et al., 2020; Liu et al., 2019a, b), zebras (*Equus quagga*) and giraffes (*Giraffa tippelskirchi*) (Badreldeen Bdawy, 2021), undulate skate (*Raja undulata*) (Gómez-Vargas et al., 2023), and bumblebees (*Bombus terrestris*) (Borlinghaus et al., 2023). In order to improve re-identification accuracy, pose estimation and key point alignment have been proposed (Yeleshetty et al., 2020; Yu et al., 2021; Moskvyak et al., 2021b).

PIE (Moskvyak et al., 2021a) is a deep learning-based method for matching of individuals which is invariant to the pose. The method receives shape embedding and pose embedding separately and normalizes the shape to match the individual regardless of the specific pose. PIE was originally developed for manta rays (Moskvyak et al., 2021a), but it has been also used for humpback whale flukes, orcas, and right whales. Apart from CNNs, also vision transformers have been proposed for animal re-identification (Zheng et al., 2022). While end-to-end deep learning methods have been shown to produce state-of-the-art performance when the amount of training data is large, their data-hungry nature limits their applicability on species for which large-scale databases are not available.

A number of methods for the re-identification specific to Saimaa ringed seals—one of our target species—have been proposed (Zhelezniakov et al., 2015; Chehrsimin et al., 2018; Nepovinnykh et al., 2018, 2020; Chelak et al., 2021; Nepovinnykh et al., 2022a, b, 2023; Immonen et al., 2023). Saimaa ringed seals are especially challenging species for re-identification due to several issues: (i) a large variation in possible poses, exacerbated by the deformable nature of the seals, (ii) non-uniform pelage patterns, limiting the size of the regions that can be used for the re-identification task, (iii) low contrast between the ring pattern and the rest of the pelage, and (iv) extreme dataset bias since the collected dataset contains disproportionally more images of some selected individuals and the variety of the backgrounds is extremely small due to the limited number of camera trap locations. These challenges have been addressed by proposing various approaches to preprocess the images and to encode the pattern features (Zhelezniakov et al., 2015; Chelak et al., 2021; Nepovinnykh et al., 2020, 2022a). The most successful methods employ a pattern extraction step (Nepovinnykh et al., 2020, 2022a) to construct a binary representation of the pelage pattern and metric learning-based pattern encoding.

Individual whale sharks can be identified based on the spot pattern on their skin. Arzoumanian et al. (2005) applied a blob detection to find the individual spots, and used pattern-matching algorithm (Groth, 1986) originally developed for astronomical images (star patterns) to compare the patterns. Kholiavchenko (2022) utilized a U-Net-based model for spot detection and a metric learning-based approach generated pattern embeddings for the re-identification of individuals.

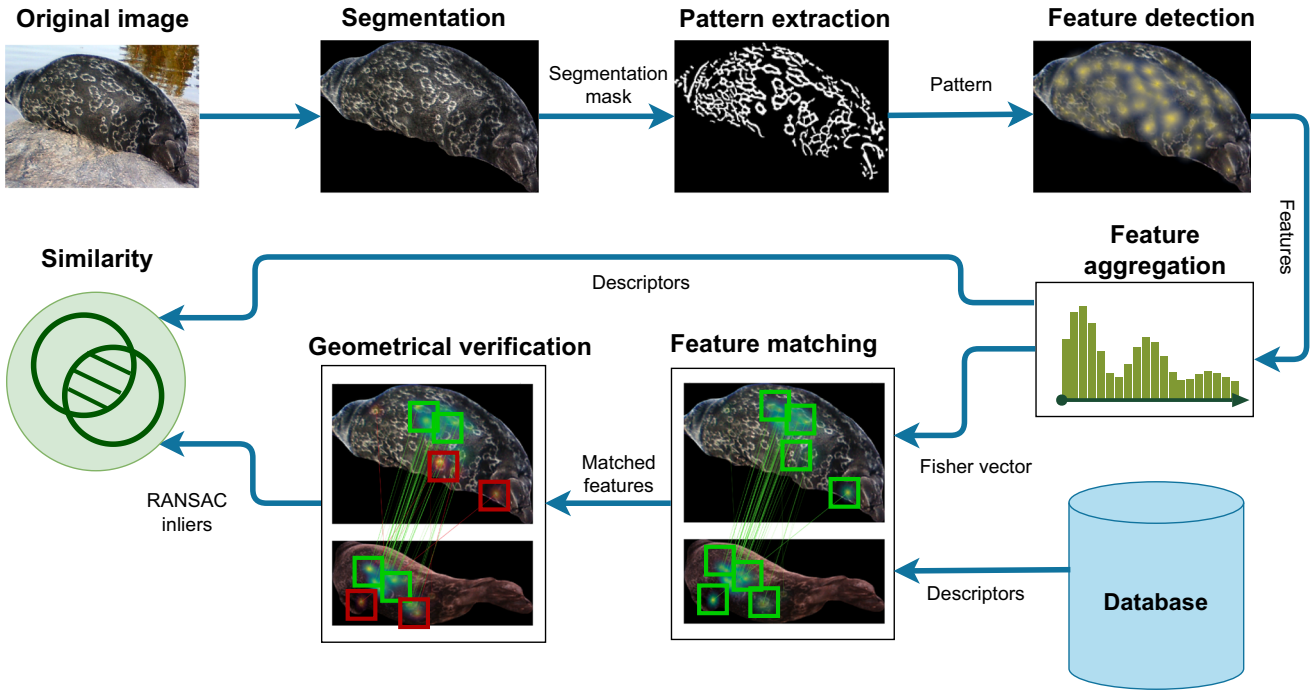

Fig. 3 ALFRE-ID re-identification pipeline

2.2 Content Based Image Retrieval

The task of visual animal re-identification can be formulated as a task of finding the most similar image from the database to the given query image. This formulation matches the definition of content-based image retrieval (CBIR) (Smeulders et al., 2000) and motivates the study of the suitability of CBIR methods for animal re-identification. CBIR methods have already been applied to the task of animal re-identication (Nepovinnykh et al., 2022a).

CBIR methods usually consist of two main steps: feature extraction and feature aggregation. The feature extraction problem can be solved using standard hand-crafted features, such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004; Arandjelović & Zisserman, 2012), or extraction by convolutional neural networks [see, e.g., (Mishchuk et al., 2017))]. Then, feature aggregation creates a descriptor for each image that can be used to find the most similar image from the database. Traditional methods such as Bag of Words (BOW) (Sivic, 2003), Vector of Locally Aggregated Descriptors (VLAD) (Jégou et al., 2010) and the Fisher Vector (Perronnin & Dance, 2007; Perronnin et al., 2010; Hutchison et al., 2010) do the aggregation using a specially constructed codebook. The codebook is usually created by an unsupervised clustering algorithm. For example, k-means (MacQueen et al., 1967) is used for VLAD, and a Gaussian Mixture Model (GMM) (McLachlan & Basford, 1988) is used for the Fisher Vector. Finally, fixed-size descriptors are created for each image based on the vocabulary and extracted features. The distance between these descriptors is inversely proportional to the visual similarity.

Due to the availability of data and the convenience of end-to-end approaches, deep learning-based methods for CBIR are becoming increasingly popular such as NetVLAD (Arandjelovic et al., 2016) where a generalized VLAD layer is used to aggregate CNN-extracted features.

Also, visual localization (Sarlin et al., 2019) shares similarities with CBIR and the animal re-identification. In visual localization, the task is to find a location in an environment that corresponds to a given image. While the formulation of CBIR is more closely related to the animal re-identification, similar steps are utilized also in visual localization including pose estimation, feature aggregation, database search, and geometrical verification.

3 Pipeline

The proposed ALFRE-ID pipeline is inspired by CBIR techniques and consists of seven steps (see Fig. 3): (1) image prepossessing, (2) instance segmentation, (3) pelage pattern extraction, (4) feature extraction, (5) feature aggregation, (6) individual re-identification, and (7) geometric verification.

Fig. 4 Examples of the image processing of camera trap images. The images on the left are the originals. The right column demonstrates the result of the tone-mapping

Some of these steps involve choices of different methods depending on the species.

3.1 Image Preprocessing

Depending on illumination conditions, variation in the contrast of the images can be rather high. This could lead to a loss of detail in the region of interest, i.e., the animal and its fur pattern. In order to rectify this issue, we employ the tone-mapping approach to equalize the contrast in dark and bright image regions. The algorithm proposed by Mantiuk et al. (2006) is used due to its ability to produce realistic tone-mapped images without introducing visual artifacts. This method considers contrast on multiple spatial frequencies while using gradient methods with some additional extensions to ensure that the global brightness levels are not reversed and low-frequency details are properly reconstructed. Examples of images before and after prepossessing are presented in Fig. 4.

3.2 Instance Segmentation

The instance segmentation step is important in the common scenario where datasets are collected using static camera traps. This together with the fact that individual animals tend to use the same sites or areas inter-annually causes one individual to be very often captured with the same camera (the same background). This increases the risk that the supervised re-identification algorithm learns to identify the background instead of the actual animal if the full image or the bounding box around it is used. Consequently, this algorithmic behavior may lead to such a situation where the method is unable to identify the animal in a new environment.

The model selection depends on the species. Animals captured in groups require instance segmentation such as Mask R-CNN (He et al., 2017) while for solitary animals, the segmentation can be solved with simpler semantic segmentation models. For various common animal species, pretrained models are already available (Bello et al., 2021; Chen & Belbachir, 2023; Dai & Liu, 1966). If this is not the case, transfer learning can be utilized. Recent natural language processing based promptable segmentation methods such as the Segment Anything Model (SAM) (Kirillov et al., 2023) provide flexible segmentation models for new target species via zero-shot transfer. Details on how the instance segmentation was implemented for the target species are given in Sect. 4.1.

3.3 Pelage Pattern Extraction

The main identifying feature of many species is their fur, feather, or skin pattern. Often the pattern is both permanent and unique to each individual, and therefore, quantifying the pattern can form the basis of individual re-identification. Depending on the species it can be beneficial to focus the attention on the pattern and discard irrelevant information causing database bias—such as illumination and other visual factors—by extracting the pattern from the images (Nepovinnykh et al., 2022a). The pattern extraction can be formulated as an image binarization problem and solved using encoder-decoder networks. The result of the pattern extraction step is a binary image containing only the pattern.

Due to the differences in fur patterns between species, the pattern extraction can be unnecessary or require a custom model. Detailed descriptions of the pattern extraction step for the target species are provided in Sect. 4.1. Since the animal is first segmented and the pattern colors often follow a bimodal distribution (dark pattern on light background or vice versa), reasonably good pattern extraction accuracy can be obtained with segmentation models pretrained on other species if necessary training data is not available for the target species. For example, Immonen et al. (2023) applied successfully a pattern extraction model trained on Saimaa ringed seals to whale sharks.

3.4 Feature Extraction

Local feature extraction and description have shown to be efficient tools for animal re-identification (Berger-Wolf et al., 2017; Nepovinnykh et al., 2022a). However, traditional hand-crafted local features such as SIFT are significantly limited in how they exploit any available training data. Modern learning-based local features, on the the hand, leverage the benefits of deep learning and CNNs to obtain representative feature descriptors, making them an attractive alternative for animal re-identification.

Wild animals can be found in a variety of poses resulting in distorted and warped patterns on images. While the pattern as a whole is transformed in a non-linear way, it can be argued that small local regions experience close to affine transformations, making an affine invariant feature extractor suitable for the task. Modern CNN-based local feature extraction approaches allow learning affine invariant feature descriptors using general-purpose datasets. This makes the feature extraction step flexible in a sense that different feature detector and descriptor combinations can be used for different species without the need for additional training.

HesAffNet (Mishkin et al., 2018) is a modification of the classical Hessian Affine Region detector (Mikolajczyk & Schmid, 2002, 2004) where the shape estimation step is done by the AffNet CNN. The detector is based on the Harris cornerness measure (Harris & Stephens, 1988) which uses a second moments matrix to find regions of interest by estimating the most prominent gradient directions. This method is combined with the multiscale approach from (Lindeberg, 1998) which uses Laplacian of Gaussian to find extrema in the scale space. The same concept can be further extended to all affine transformations, not just the scale. However, the degree of freedom is much higher for affine transformations, which complicates the process and requires a special shape adaptation algorithm. The original Hessian Affine detector used Baumberg iteration (Baumberg, 2000), which is replaced by an AffNet CNN in HesAffNet.

AffNet and HardNet are closely related, sharing the same architecture and using similar training procedures. During the training of HardNet (Mishchuk et al., 2017), batches of matching patch pairs are chosen, each containing an anchor a_i and positive match p_i. Each patch is encoded by the network, and a matrix of pair-wise distances between all anchors and positive matches is computed. For each pair, the 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 $d(\cdot, \cdot)$ is the distance function, $p_{j\,\min}$ 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, the main difference being that the derivative for the negative term in the loss is set to 0. This loss is called hard negative constant and helps avoid situations where positive samples cannot be moved closer together because of a negative sample lying between them in the metric space. The training procedure for AffNet is also more complicated since it is learning affine shapes and not just a distance metric. Therefore, spatial transformers are used to transform input patches according to the predicted shape, which are then fed into a descriptor network, e.g., HardNet, and only then is the loss calculated and backpropagated through both networks. The example of HesAffNet application to a preprocessed image is visualized in Fig. 5.

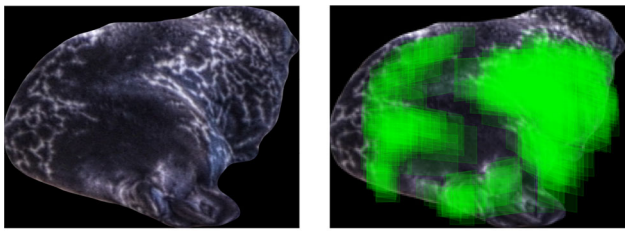

Fig. 5 Visualisation of Hessian Affine patch extraction: **a** segmented image; **b** HesAffNet-based patch extraction. Note that while original images are used for visualization purposes, the features are extracted from pattern images. Extracted regions are highlighted in green (Color figure online)

Going a step further, the Hessian detector can be replaced by a learned CNN-based method. Key.Net (Barroso-Laguna & Mikolajczyk, 2022) uses a combination of manually hardcoded and learned filters along with a multi-scale pyramid. This feature detector can be used in conjunction with AffNet and HardNet for a full feature detection and encoding pipeline.

Alternatively, DISK (Tyszkiewicz et al., 2020) provides an end-to-end framework for both feature extraction and encoding. DISK uses the U-net network as the backbone and utilizes reinforcement learning in order to train it. The network is trained to obtain a high number of correct matches by using a policy gradient method, keeping training and inference very close to each other. The network outputs dense descriptors and keypoints heatmap, which together could be combined to obtain discriminative sparse keypoints.

3.5 Feature Aggregation

Features are aggregated using Fisher Vector (Perronnin & Dance, 2007; Perronnin et al., 2010; Hutchison et al., 2010). First, Principal Component Analysis (PCA) is applied to the feature embeddings to decorrelate the features and reduce the dimensionality. This is important for Fisher Vectors, which are known to produce large descriptors. The images in the database of known individuals are used to learn principal components. Next, a visual vocabulary (codebook) is constructed by applying a Gaussian Mixture Model (GMM) to the features from the database. Then, 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.

3.5.1 Fisher Vector

Let $X = \{x_t, t = 1, \ldots, T\}$ be a sample of T observations and u_λ be a probability density function modeling the distribution of the data, where λ is a vector of its parameters. The score is defined as the gradient of the log-likelihood of the data on the model:

$$G_\lambda^X = \nabla_\lambda \log u_\lambda(X). \tag{2}$$

This score function can be used to define the Fisher Information Matrix (FIM) (Amari & Nagaoka, 2000):

$$F_\lambda = E_{x\sim u_\lambda}[G_\lambda^X G_\lambda^{X'}], \tag{3}$$

which acts as a local metric for a parametric family of distributions. This metric can also be used to measure the similarity between 2 samples using the Fisher Kernel (FK) (Jaakkola & Haussler, 1999):

$$\begin{aligned} K_{FK}(X, Y) &= G_\lambda^{X'} F_\lambda^{-1} G_\lambda^Y \\ &= G_\lambda^{X'} L_\lambda{}' L_\lambda G_\lambda^Y \\ &= \mathscr{G}_\lambda^{X'} \mathscr{G}_\lambda^Y, \end{aligned} \tag{4}$$

where $L'_\lambda L_\lambda$ is the Cholesky decomposition of F_λ^{-1}, G_λ^X and G_λ^Y are the Fisher Vectors of samples X and Y respectively. By using Fisher Vectors, it is possible to calculate the kernel as a simple dot product, which can be efficiently utilized by linear classifiers. When constructing a Fisher Vector for an image, a set of local features is assumed to be independent, meaning that the final descriptor can be constructed as a sum of Fisher Vectors for each local feature, i.e.,

$$G_\lambda^X = \sum_{t=1}^{T} L_\lambda \nabla_\lambda \log u_\lambda(X). \tag{5}$$

Usually, a Gaussian Mixture Model (GMM) is used as u_λ, since it can be used to approximate any continuous distribution with arbitrary precision (Titterington et al., 1985). The gradients of the GMM parameters are concatenated into a vector of size $2DK$ where D is the dimensionality of samples and K is the number of components in GMM. It has been shown (Hutchison et al., 2010) that $L2$ and power normalization generally improve the performance of the method. Therefore, it is common to apply power and $L2$ normalization to the Fisher Vector to get the final descriptor.

3.6 Individual Re-identification

Re-identification is done by calculating the cosine distance from the query image descriptor to each image descriptor in the database of known individuals as

$$d_L = 1 - \frac{\Phi_q \cdot \Phi_{db}}{||\Phi_q||_2 ||\Phi_{db}||_2}, \tag{6}$$

where Φ_q is the Fisher vector for query image and Φ_{db} is the Fisher vector for a database image. This distance quantifies the dissimilarity of the aggregated local pattern appearances between the images. The individuals in the database are ranked based on the distances, the first-ranked being the most likely match.

3.7 Geometric Verification

Aggregated local pattern appearance does not take into account the global spatial structure of the pattern. To further incorporate this information to the pattern matching, the geometric consistency of the local similarities is analyzed. This is done using a similar method as the spatial reranking step of the HotSpotter algorithm (Crall et al., 2013) and the object retrieval method proposed in Philbin et al. (2007). Local interest points extracted from each image are matched to find the feature correspondences between query and database images. The matching is done by computing cosine distances between the embeddings of individual feature pairs.

The image coordinates of feature correspondences are then normalized to have the zero mean and the maximum distance of 1 to the origin. Outliers (and inliers) are detected by estimating the parameters of a homography between the query image and database image using RANSAC. The assumption is that if the patterns do not match, the inconsistency in the global arrangements of feature correspondences causes a low number of inliers. Therefore, the number of inliers, n, is a good metric for geometric similarity of patterns. It should be noted that due to the large pose variation of animals, it is recommended to have a high inlier threshold to ensure successful outlier detection in the case of matching patterns.

The final re-identification of the animal individual in the query image is performed by searching the most similar pattern from the database of known individuals. To compute the dissimilarity (distance) a novel combination of the dissimilarity of aggregated local pattern appearance and geometric dissimilarity of patterns is used. We use the following reranking rule:

$$d_C = (d_L)^n, \tag{7}$$

where d_L is the cosine distance between Fisher vectors (aggregated local pattern appearance) and n is the number of inliers. The geometric consistency, defined as a number of inliers n, has an exponential influence on the cosine distance

($d_L \leq 1$). If the number of individuals in the database is large, re-identification can be made more efficient by using the aggregated Fisher vector for quick database searches and using the geometric similarity only as a reranking or verification step.

4 Experiments and Results

Our experiments are focused on two key issues: (i) the impact of modern pre-trained feature extraction algorithms on content-based-retrieval approaches to individual animal identification and (ii) the impact of training data size on the relative efficacy of local feature-based methods as opposed to end-to-end deep learning based methods.

4.1 Data

We consider two very different patterned animals: Saimaa ringed seals and whale sharks. Saimaa ringed seal patterns consist of local arrangements of ring-like shapes. The regions enabling the re-identification often constitute a rather small portion of the whole pattern. This together with the fact that the contrast between the pattern and the rest of the body is low and the appearance of the pattern varies, makes this a challenging dataset. Whale shark patterns, on the other hand, consist of small spots with similar appearance and the main trait allowing the re-identification is the geometric arrangement of the spots. Small differences between individuals and a large variation in image quality due to underwater imaging further complicate the re-identification task.

4.1.1 Saimaa Ringed Seals

The re-identification dataset consists of 57 individual seals with a total of 2080 images. The dataset is divided into two subsets: the database subset (430 images) and the query subset (1650 images). 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 of the same individuals as in the database. It should be noted that the high-quality images were prioritized when constructing the database and, therefore, images in the query subset often have lower quality. Examples of images from both subsets are presented in Fig. 6. The dataset has been made publicly available. For further description of the dataset, see Nepovinnykh et al. (2022c).

Images were segmented using Mask R-CNN (He et al., 2017). A segmentation model trained for Ladoga ringed seals from Nepovinnykh et al. (2022b) was utilized. 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 captured 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 Nepovinnykh et al. (2022b). After the segmentation masks were obtained, morphological opening and closing operations were applied to close the holes and smooth the borders by using morphological closing and opening. Examples of segmentation results are presented in Fig. 7.

Fig. 6 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)

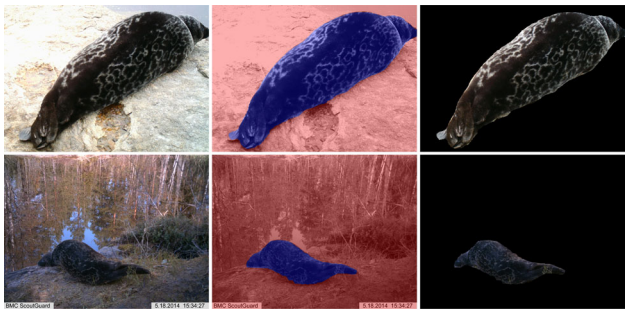

Fig. 7 Examples of the segmentation masks. The images on the left are the originals. The mask is highlighted in blue and the background is highlighted in red on the middle images. The last column shows the result of the segmentation (Color figure online)

Saimaa ringed seal pattern was extracted using the U-net encoder-decoder architecture (Ronneberger et al., 2015). The pattern image was further post-processed to remove small noise by using unsharp masking and morphological opening. Finally, all images were resized in such a way that the mean width of the pattern lines was the same for all images, bringing them into the same scale. The line width for each image can be approximated as a ratio of the number of all white pixels to the number of all pixels in the morphological skeleton. This operation is necessary since the images were obtained from a variety of sources and have a large variation in image resolution. Example results for pattern extraction are shown in Fig. 8. For a more detailed explanation of the seal pattern extraction step, as well as the comparison to other methods, see Zavialkin (2020).

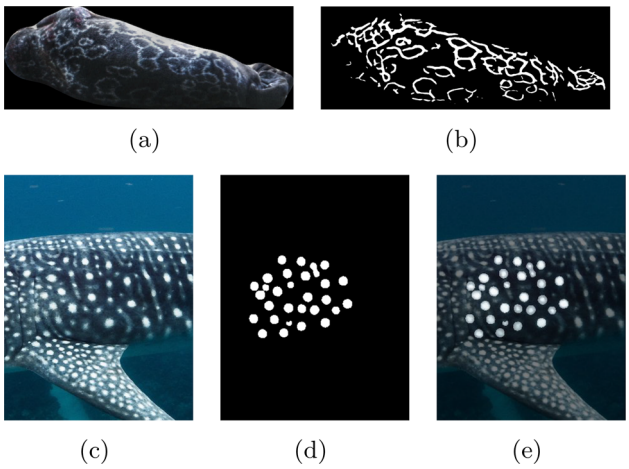

Fig. 8 Visualization of the pattern extraction step for the Saimaa ringed seals: (**a**) and (**b**), and for the whale sharks: (**c**)–(**e**)

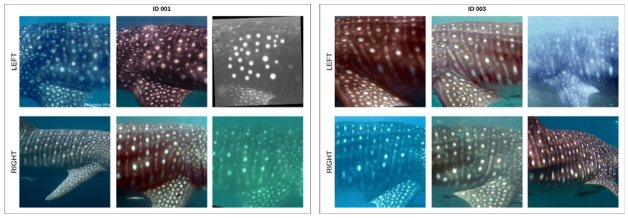

Fig. 9 Sample images from the whale shark dataset

4.1.2 Whale Sharks

To study the re-identification, the whale shark identification dataset provided by Wild Me (Holmberg et al., 2009; Blount et al., 2022) has been used. The database of whale sharks was curated using the semi-automatic Modified Groth algorithm (Arzoumanian et al., 2005; Holmberg et al., 2009) to suggest matches, that were verified whale shark experts. Each image in the dataset is accompanied by a bounding box delineating the torso of the whale shark's body, an individual identification tag, and the viewpoint of the animal (right or left). Therefore, examples of whale shark images cropped according to bounding boxes are presented in Fig. 9. The dataset is divided into training and test subsets for training neural network-based methods. The training subset comprises a total of 5409 annotated sightings, specifically pertaining to 235 distinct whale shark viewpoints (unique combinations of an individual and a viewpoint). The test subset consists of 1543 sightings belonging to 412 unique viewpoints. No individuals present in the training set are included in the test set. The image distribution of images for the training subset can be seen in Fig. 10. Since the query/database split is not provided in this dataset, a leave-one-out strategy is used to assess the re-identification accuracy. That is, each image is compared to all other images.

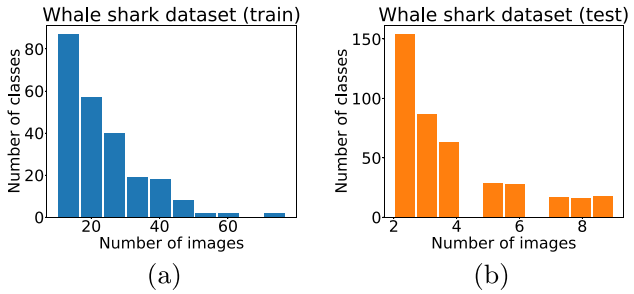

Fig. 10 Image distribution across the whale shark dataset, displaying the number of classes (individual + viewpoint) and the corresponding number of images along the x-axis: **a** training subset; **b** test subset. For example, in the test subset, there are around 150 classes with less than 2 images per class

Whale sharks exhibit a pattern characterized by an array of spots, which adorn their massive bodies. These spots, varying in size and spacing, create a unique mosaic-like arrangement that serves as a natural identifier for individual whale sharks. To accurately extract the pattern, we adopt a specialized approach that centers on segmenting these white spots. The segmentation process involves the neural network to perform image classification at the pixel level, where it precisely identifies and delineates each individual white spot on the whale shark's body. In our work, we adopt the U-net architecture (Ronneberger et al., 2015) with the SEResNet34 (Hu et al., 2018) backbone that has been successfully applied for similar problems such as blood vessel segmentation from medical images (?). The U-net architecture consists of encoder and decoder parts. The encoder part hierarchically encodes input images into a latent representation, effectively capturing essential features. A decoder part employs up-sampling layers to expand the latent representation to match the original image dimensions. Encoder layers pass over information to the corresponding decoder layers with the help of special connections. This helps to transfer the classification context to the localization part. The neural network's objective is to recognize the presence of white spots and outline their boundaries, down to the finest details. The resulting outcome is a set of binary masks, where every non-zero pixel value corresponds to the location of a white spot on the whale shark's body.

4.2 Comparison of Feature Descriptors

In order to select the most suitable feature descriptor for each dataset, re-identification for each dataset has been performed using three different deep local features: (1) HesAffNet (Mishkin et al., 2018) for feature detection and HardNet (Mishchuk et al., 2017) for feature description, (2) Key.Net (Barroso-Laguna & Mikolajczyk, 2022) + AffNet (Mishkin et al., 2018) for feature detection and HardNet (Mishchuk et al., 2017) for description (which

Table 1 Experiments with different descriptors on the SealID dataset

Method	GV	Original		Pattern	
		Top-1	Top-5	Top-1	Top-5
HessAffNet +HardNet	N	52.55	66.06	77.03	85.15
	Y	58.61	69.82	**81.64**	**87.33**
Key.Net+HardNet	N	37.21	54.00	60.55	73.58
	Y	42.12	58.18	68.42	77.09
DISK	N	33.88	49.39	45.70	59.58
	Y	38.79	54.97	50.55	62.79

GV column indicates whether geometrical verification was used (Y) or not (N). "Original" and "Pattern" indicate whether pattern extraction was skipped or not respectively
The best results are highlighted in bold

will be referred to as Key.Net + HardNet for the sake of brevity), and (3) DISK (Tyszkiewicz et al., 2020) feature detectors/descriptors.

4.2.1 Saimaa Ringed Seals

Results for the SealID dataset are presented in Table 1. It is clear that the choice of the feature extractor and descriptor greatly affects the final accuracy. The difference between the DISK and HessAffNet + HardNet is around 30%, with the DISK showing the worst results and HessAffNet + HardNet the best. The results also indicate that the pattern extraction step is integral to the re-identification on the SealID dataset, increasing the accuracy by about 20%.

4.2.2 Whale Sharks

For the whale shark dataset, only a training subset was used to create codebooks for PCA and GMM. It should be noted that since we are using pre-trained local feature detectors and descriptors, no method training is needed. Codebook generation does not require identity labels which makes it possible to test two realistic scenarios: (1) the codebook is generated for the same set of images the re-identification is applied (*fine-tuned codebook*) and (2) the codebook is generated and tested with a different set of images (*pre-generated codebook*). The first corresponds to a scenario where re-identification is applied to a fixed set of images collected earlier. In this case, the re-identification process starts with the generation of a codebook and proceeds to re-identify the animal in each image. In the second scenario, the codebook is generated beforehand (offline) and the re-identification happens online while new images are collected. It is good to notice that since the subset used to generate the codebook (training set) and subset used to test the re-identification accuracy do not contain the same individuals, this is even more challenging than a typical scenario, where, at least, some individuals have been captured earlier and can be used for generating the codebook.

Both fine-tuned and pre-generated codebooks were tested using a leave-one-out strategy. The results for the dataset without the pattern extraction step are presented in Table 2. DISK approach achieves the highest re-identification accuracy in contrast to the SealID dataset, where it performed the worst. The results for the pre-generated codebooks are consistently worse than for the fine-tuned codebook, which is the expected consequence of the fact of how the codebooks were created. Results for different feature extractors and descriptors on the whale shark dataset with pattern extraction step are presented in Table 3. Moreover, DISK applied to the original images outperforms all other feature extractors and descriptors both with and without the pattern extraction step. With the addition of the pattern extraction step, both Key.Net + HardNet and DISK perform comparably well, achieving higher accuracy than HessAffNet + HardNet, with DISK producing slightly higher accuracy scores when using pre-generated codebook than Key.Net + HardNet. Surprisingly, while pattern extraction significantly increases accuracy for the HessAffNet + HardNet and Key.Net + HardNet approaches, the DISK method produces better results using original images.

4.3 Comparison to PIE

PIE (Moskvyak et al., 2021a) is an end-to-end deep learning method for re-identification. The main problem with PIE and similar methods for re-identification is their need for a large amount of the labeled training data which is often not available for wildlife applications. Acquiring and labeling large datasets of animal individuals is a difficult and tedious task requiring expertise, time and effort. With that in mind, one of the main advantages of the proposed ALFRE-ID pipeline is that the core of the algorithm does not require training on the target dataset as the feature extractors and descriptors are pretrained and the codebook generation does require labeled data. In order to simulate a real-world scenario where fully labeled data is scarce, we compared the ALFRE-ID pipeline to PIE with different sizes of training set: 100%, 50%, and 25% of the original training sets. For ALFRE-ID, the training set was only used to generate the PCA and GMM codebooks. The reduction is done on a per-individual basis, i.e. for each individual only 50% of available images from the full train set are used for the training/codebook generation. 100% of the training set corresponds to the standard split.

In order to compare PIE to ALFRE-ID on the SealID dataset, a special train-test split has been used. The whole dataset, i.e., the union of the query and database subsets, has been divided in the following manner: if the individual contains more than 6 samples, it is assigned to the training set and otherwise to the test set. Therefore, the training and test sets contain a different set of individuals similar to the whale shark data set. The final scores are presented for the

Table 2 Experiments with different descriptors on the whale shark dataset without pattern extraction

Method	GV	Fine-tuned codebook		Pregenerated codebook	
		Top-1	Top-5	Top-1	Top-5
HessAffNet + HardNet	N	58.90	73.06	43.68	57.61
	Y	69.97	80.01	54.37	64.67
Key.Net + HardNet	N	36.43	52.81	24.69	35.83
	Y	49.95	60.97	35.85	42.46
DISK	N	72.58	84.12	52.81	66.49
	Y	**83.00**	**88.92**	**67.40**	**74.91**

The best results are highlighted in bold

Table 3 Experiments with different descriptors on the whale shark dataset with pattern extraction

Method	GV	Fine-tuned codebook		Pregenerated codebook	
		Top-1	Top-5	Top-1	Top-5
HessAffNet + HardNet	N	74.04	85.13	43.09	57.03
	Y	73.00	84.10	43.23	55.61
Key.Net + HardNet	N	81.12	89.81	51.78	64.22
	Y	**81.17**	89.65	51.32	63.97
DISK	N	80.03	88.89	56.38	70.57
	Y	81.16	**90.20**	**59.49**	**72.19**

The best results are highlighted in bold

leave-one-out re-identification on the test set. The results are presented in Table 4.

As expected the size of the set used to generate the codebook does not have a large influence on the re-identification accuracy of the ALFRE-ID method. The difference in accuracy between the full and a half training set for different datasets is between 1% and 7%. Further reducing the size of the training set to 25% of its original size does not have a negative effect on the accuracy. Contrary to the ALFRE-ID, the accuracy of PIE drops significantly when the size of the training set is reduced. The accuracy on the whale shark dataset drops from 86% to 51% when the size of the training set is reduced to a quarter of its original size. When using pregenerated codebook for ALFRE-ID, PIE shows higher accuracy only if 50% or more of the available images are used for training. In SealID, a similar large drop can be observed. Results on fine-tuned codebook are again considerably better than on pregenerated codebook. However, it should be noted that the accuracies on fine-tuned codebook are not fully comparable with those on PIE as the test set is different.

4.3.1 Comparison to Hotspotter

Hotspotter (Crall et al., 2013) is another popular species-agnostic re-identification algorithm that uses local features (SIFT) for the re-identification. The comparison between ALFRE-ID and HotSpotter for both datasets is presented in Table 5. ALFRE-ID outperforms Hotspotter on both datasets, with a lead of about 20%. Moreover, only small differences between Top-1 and Top-5 scores for Hotspotter can be observed, while the increase in accuracy for the ALFRE-ID method is clear. That means that ALFRE-ID would provide more benefit in the semi-automatic re-identification scenario where the set of best matches is provided for an expert for the final verification. The results indicate that the modern CNN-based local features together with feature aggregation significantly increase the re-identification accuracy compared to traditional local feature-based methods.

5 Conclusion

In this paper, a novel pipeline for patterned animal re-identification called Aggregated Local Features for Re-Identification (ALFRE-ID) was proposed. The pipeline utilizes modern deep learning-based local features and feature aggregation inspired by content-based image retrieval techniques. The full re-identification pipeline consists of image enhancement, animal instance segmentation, optional fur pattern extraction, feature extraction, feature aggregation, individual re-identification by database search, and geometric verification steps. The pipeline follows a modular approach where individual techniques can be changed to address differences between animal species. The main benefit of the proposed approach is that by utilizing pretrained local feature descriptors no labeled training data is needed to deploy the re-identification model to new species. At the same time, powerful feature representations are obtained via feature aggregation enabling comparable re-identification accuracy to deep learning-based end-to-end models that

Table 4 Experiments with different sizes of the training set

Dataset	%	No	ALFRE-ID		PIE
			Fine-tuned codebook	Pregenerated codebook	
SealID	100	1988	83–88	76–83	47–73
SealID	50	974	76–82	76–83	45–71
SealID	25	465	76–81	78–87	34–58
Whale shark	100	5155	83–89	67–75	86–93
Whale shark	50	2479	83–89	66–74	73–84
Whale shark	25	1107	82–89	66–74	51–68

HesAffNet + HardNet is used for SealID and DISK is used for whale sharks. The % column specifies the percent of the training set used to create codebooks. The results are presented in pairs as *a*–*b*, where *a* is top-1 accuracy and *b* is top-5 accuracy

Table 5 Comparison with Hotspotter. HessAffNet + HardNet feature descriptor is used for the SealID dataset when testing ALFRE-ID. The DISK feature descriptor without pattern extraction is used for the whale shark dataset

Method	SealID		Whale shark			
			Fine-tuned codebook		Pregenerated codebook	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ALFRE-ID	**81.64**	**87.33**	**83.00**	**88.92**	**67.40**	**74.91**
Hotspotter	69.58	76.24	63.30	63.58	49.51	51.07

The best results are highlighted in bold

require significantly larger amount of training data. This makes it possible to apply the pipeline to the new animal species for which large-scale labeled databases are not available. We evaluated the method against other state-of-the-art data-driven and hand-crafted animal re-identification methods on two challenging datasets of Saimaa ringed seals and whale sharks. Our method clearly outperformed the competing methods under limited training data scenarios. As future work, we plan to apply and test our method on more animal species.

Acknowledgements 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.

Author Contributions T. Eerola and H. Kälviäinen were responsible for the supervision of the research, and project administration; E. Nepovinnykh, I. Chelak, T. Eerola, V. Immonen, H. Kälviäinen, and C. Stewart participated designing methodology; E. Nepovinnykh, I. Chelak, V. Immonen, M. Kholiavchenko implemented the algorithm; E. Nepovinnykh, I. Chelak, T. Eerola, H. Kälviäinen, M. Kholiavchenko, and C. Stewart 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)).

Data Availability The SealID dataset is publicly available at https://doi.org/10.23729/0f4a3296-3b10-40c8-9ad3-0cf00a5a4a53 The whale shark dataset is not publicly available.

Code Availability The codes for the described experiments are available at https://github.com/kwadraterry/Norppa.

Declarations

Conflict of interest We declare no conflict of interest.

Consent for Publication All authors consent that the publisher has the author's permission to publish research findings. All authors guarantee that the research findings have not been previously published.

References

Agarwal, M., Sinha, S., Singh, M., et al. (2019) Triplet transform learning for automated primate face recognition. In *International conference on image processing (ICIP)*. https://doi.org/10.1109/ICIP.2019.8803501

Amari, S., & Nagaoka, H. (2000). *Methods of Information Geometry*. American Mathematical Society.

Arandjelović, R., & Zisserman, A. (2012). Three things everyone should know to improve object retrieval. In *Conference on computer vision and pattern recognition (CVPR)*. https://doi.org/10.1109/CVPR.2012.6248018

Arandjelovic, R., Gronat, P., Torii, A., et al. (2016). NetVLAD: CNN architecture for weakly supervised place recognition. In *Confer-*

ence on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/CVPR.2016.572

Araujo, G., Ismail, A., McCann, C., et al. (2020). Getting the most out of citizen science for endangered species such as Whale Shark. *Journal of Fish Biology, 96*, 864–867. https://doi.org/10.1111/jfb.14254

Arzoumanian, Z., Holmberg, J., & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of Whale sharks Rhincodon typus. *Journal of Applied Ecology, 42*(6), 999–1011.

Badreldeen Bdawy Mohamed, O. (2021). Metric learning based pattern matching for species agnostic animal re-identification. Master's thesis, Lappeenranta-Lahti University of Technology LUT, Finland

Barroso-Laguna, A., & Mikolajczyk, K. (2022). Key.net: Keypoint detection by handcrafted and learned CNN filters revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*, 698–711. https://doi.org/10.1109/iccv.2019.00593

Baumberg, A. (2000). Reliable feature matching across widely separated views. In *Conference on computer vision and pattern recognition (CVPR)*. https://doi.org/10.1109/CVPR.2000.855899

Bello, R. W., Mohamed, A. S. A., & Talib, A. Z. (2021). Contour extraction of individual cattle from an image using enhanced mask R-CNN instance segmentation method. *IEEE Access, 9*, 56984–57000. https://doi.org/10.1109/ACCESS.2021.3072636

Berger-Wolf T, Rubenstein D, Stewart C, et al (2015) Ibeis: Image-based ecological information system: From pixels to science and conservation. In: Bloomberg Data for Good Exchange Conference

Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V., et al. (2017). Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv preprint arXiv:1710.08880

Blount, D., Gero, S., Van Oast, J., et al. (2022). Flukebook: An open-source AI platform for cetacean photo identification. *Mammalian Biology, 102*, 1005–102. https://doi.org/10.1007/s42991-021-00221-3

Bogucki, R., Cygan, M., Khan, C. B., et al. (2019). Applying deep learning to right whale photo identification. *Conservation Biology, 33*, 676–684. https://doi.org/10.1111/cobi.13226

Borlinghaus, P., Tausch, F., & Rettenberger, L. (2023). A purely visual re-id approach for bumblebees (Bombus terrestris). *Smart Agricultural Technology, 3*, 100135.

Brust, C. A., Burghardt, T., Groenenberg, M., et al. (2017) Towards automated visual monitoring of individual gorillas in the wild. In *International conference on computer vision workshop (ICCVW)*. https://doi.org/10.1109/iccvw.2017.333

Cheeseman, T., Johnson, T., & Muldavin, N. (2017) Happywhale: Globalizing marine mammal photo identification via a citizen science web platform. Paper SC/67A/PH/02 presented to the Scientific Committee of the Report to the International Whaling Commission.

Cheeseman, T., Southerland, K., Park, J., et al. (2022). Advanced image recognition: A fully automated, high-accuracy photo-identification matching system for humpback whales. *Mammalian Biology, 102*(3), 915–929.

Chehrsimin, T., Eerola, T., Koivuniemi, M., et al. (2018). Automatic individual identification of Saimaa ringed seals. *IET Computer Vision, 12*, 146–152. https://doi.org/10.1049/iet-cvi.2017.0082

Chelak, I., Nepovinnykh, E., Eerola, T., et al. (2021). EDEN: Deep feature distribution pooling for saimaa ringed seals pattern matching. arXiv preprint arXiv:2105.13979

Chen, I. H., & Belbachir, N. (2023). Using mask R-CNN for underwater fish instance segmentation as novel objects: A proof of concept. In *Proceedings of the Northern lights deep learning workshop* (Vol. 4). https://doi.org/10.7557/18.6791

Crall, J., Stewart, C., Berger-Wolf, T., et al. (2013). Hotspotter—patterned species instance recognition. In *Winter conference on applications of computer vision (WACV)*. https://doi.org/10.1109/2013.6475023

Crouse, D., Jacobs, R., Richardson, Z., et al. (2017). Lemurfaceid: A face recognition system to facilitate individual identification of lemurs. *BMC Zoology, 2*, 1–14. https://doi.org/10.1186/s40850-016-0011-9

Dai, Y., Liu, Y., & Zhang, S. (2021). Mask R-CNN-based cat class recognition and segmentation. *Journal of Physics: Conference Series, 1966*(1), 012010. https://doi.org/10.1088/1742-6596/1966/1/012010

De Silva, M., Kumarasinghe, P., De Zoysa, K., et al. (2022). Reidentifying asian elephants from ear images using a cascade of convolutional neural networks and explaining with gradcam. *SN Computer Science, 3*(3), 192.

de Silva, E. M., Kumarasinghe, P., Indrajith, K. K., et al. (2022). Feasibility of using convolutional neural networks for individual-identification of wild asian elephants. *Mammalian Biology, 102*(3), 931–941.

Deb, D., Wiper, S., Gong, S., et al. (2018). Face recognition: Primates in the wild. In *International conference on biometrics theory, applications and systems (BTAS)*. https://doi.org/10.1109/btas.2018.8698538

Freytag, A., Rodner, E., Simon, M., et al. (2016). Chimpanzee faces in the wild: Log-Euclidean CNNs for predicting identities and attributes of primates. In *German conference on pattern recognition (GCPR)*. https://doi.org/10.1007/978-3-319-45886-1_5

Gómez-Vargas, N., Alonso-Fernández, A., Blanquero, R., et al. (2023). Re-identification of fish individuals of undulate skate via deep learning within a few-shot context. *Ecological Informatics, 75*, 102036.

Groth, E. J. (1986). A pattern-matching algorithm for two-dimensional coordinate lists. *Astronomical Journal, 91*, 1244–1248.

Harris, C. G., & Stephens, M. J. (1988). A combined corner and edge detector. In *Alvey vision conference*. https://doi.org/10.5244/c.2.23

Hartwig, S. (2005). Individual acoustic identification as a non-invasive conservation tool: An approach to the conservation of the African wild dog Lycaon pictus (Temminck, 1820). *Bioacoustics The International Journal of Animal Sound and its Recording, 15*, 35–50. https://doi.org/10.1080/09524622.2005.9753537

He, K., Gkioxari, G., Dollár, P., et al. (2017). Mask R-CNN. In *International conference on computer vision (ICCV)*. https://doi.org/10.1109/iccv.2017.322

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

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Hughes, B., & Burghardt, T. (2017). Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision, 122*, 542–557.

Hutchison, D., Kanade, T., & Kittler, J., et al. (2010). Improving the fisher kernel for large-scale image classification. In *European conference on computer vision (ECCV)*. https://doi.org/10.1007/978-3-642-15561-1_11

Immonen, V., Nepovinnykh, E., Eerola, T., et al. (2023). Combining feature aggregation and geometric similarity for re-identification of patterned animals. arXiv preprint arXiv:2308.06335

Jaakkola, T., & Haussler, D. (1999). Exploiting generative models in discriminative classifiers. In *Conference on neural information processing systems (NeurIPS)*.

Jégou, H., Douze, M., Schmid, C., et al. (2010). Aggregating local descriptors into a compact image representation. In *Conference on*

computer vision and pattern recognition (CVPR). https://doi.org/10.1109/CVPR.2010.5540039

Khan, C., Blount, D., Parham, J., et al. (2022). Artificial intelligence for right whale photo identification: From data science competition to worldwide collaboration. *Mammalian Biology, 102*(3), 1025–1042.

Khan, C. B. & Shashank, W. K. (2015). Right whale recognition. https://kaggle.com/competitions/noaa-right-whale-recognition

Kholiavchenko M (2022) Comprehensive deep learning pipeline for whale shark recognition. Master's thesis, Rensselaer Polytechnic Institute (RPI), USA

Kim, J., Woo, S., Park, B., et al. (2022). Temporal flow mask attention for open-set long-tailed recognition of wild animals in camera-trap images. In *2022 IEEE international conference on image processing (ICIP)* (pp. 2152–2156). IEEE.

Kirillov, A., Mintun, E., Ravi, N., et al. (2023). Segment anything. arXiv preprint arXiv:2304.02643

Korschens, M., & Denzler, J. (2019). ELPephants: A fine-grained dataset for elephant re-identification. In *International conference on computer vision workshop (ICCVW)*. https://doi.org/10.1109/iccvw.2019.00035

Kulits, P., Wall, J., Bedetti, A., et al. (2021). Elephantbook: A semi-automated human-in-the-loop system for elephant re-identification. In *ACM SIGCAS conference on computing and sustainable societies* (pp. 88–98).

Kumar, S., Pandey, A., Sai Ram Satwik, K., et al. (2018). Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement, 116*, 1–17. https://doi.org/10.1016/j.measurement.2017.10.064

Lalonde, M., Landry, D., & Sears, R. (2022). Automated blue whale photo-identification using local feature matching. In *International conference on pattern recognition* (pp. 460–473). Springer.

Li, S., Li, J., Tang, H., et al. (2020). ATRW: A benchmark for amur tiger re-identification in the wild. In *ACM international conference on multimedia*. https://doi.org/10.1145/3394171.3413569

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision, 30*, 77–116. https://doi.org/10.1023/A:1008045108935

Liu, C., Zhang, R., & Guo, L. (2019a). Part-pose guided amur tiger re-identification. In *International conference on computer vision workshop (ICCVW)*. https://doi.org/10.1109/ICCVW.2019.00042

Liu, N., Zhao, Q., Zhang, N., et al. (2019b). Pose-guided complementary features learning for amur tiger re-identification. In *International conference on computer vision workshop (ICCVW)*. https://doi.org/10.1109/ICCVW.2019.00038

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *International conference on computer vision (ICCV)*. https://doi.org/10.5555/850924.851523

MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. In *Berkeley symposium on mathematical statistics and probability*

Mantiuk, R., Myszkowski, K., & Seidel, H. P. (2006). A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception, 3*, 286–308. https://doi.org/10.1145/1166087.1166095

McCoy, E., Burce, R., David, D., et al. (2018). 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. https://doi.org/10.3389/fmars.2018.00271

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. M. Dekker.

Mikolajczyk, K., & Schmid, C. (2002). An affine invariant interest point detector. In *European conference on computer vision (ECCV)*. https://doi.org/10.1007/3-540-47969-4_9

Mikolajczyk, K., & Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision, 60*, 63–86. https://doi.org/10.1023/B:VISI.0000027790.02288.f2

Mishchuk, A., Mishkin, D., Radenovic, F., et al. (2017) Working hard to know your neighbor's margins: Local descriptor learning loss. In *Conference on neural information processing systems (NeurIPS)*

Mishkin, D., Radenović, F., & Matas, J. (2018). Repeatability is not enough: Learning affine regions via discriminability. In *European conference on computer vision (ECCV)*. https://doi.org/10.1007/978-3-030-01240-3_18

Moskvyak, O., Maire, F., Dayoub, F., et al. (2021a). 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)*. https://doi.org/10.1109/DICTA52665.2021.9647359

Moskvyak, O., Maire, F., Dayoub, F., et al. (2021b). Keypoint-aligned embeddings for image retrieval and re-identification. In *Winter conference on applications of computer vision (WACV)*. https://doi.org/10.1109/48630.2021.00072

Nepovinnykh, E., Eerola, T., Kälviäinen, H., et al. (2018). Identification of Saimaa ringed seal individuals using transfer learning. In *International conference on advanced concepts for intelligent vision systems (ACIVS)*. 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 *Winter conference on applications of computer vision workshops (WACVW)*. https://doi.org/10.1109/wacvw50321.2020.9096935

Nepovinnykh, E., Chelak, I., Eerola, T., et al. (2022a). NORPPA: Novel ringed seal re-identification by pelage pattern aggregation. arXiv preprint arXiv:2206.02498

Nepovinnykh, E., Chelak, I., Lushpanov, A., et al. (2022b). Matching individual Ladoga ringed seals across short-term image sequences. *Mammalian Biology* 1–16. https://doi.org/10.1007/s42991-022-00229-3

Nepovinnykh, E., Eerola, T., Biard, V., et al. (2022c). SealID: Saimaa ringed seal re-identification database. arXiv preprint arXiv:2206.02260

Nepovinnykh, E., Vilkman, A., Eerola, T., et al. (2023). Re-identification of saimaa ringed seals from image sequences. In *Scandinavian conference on image analysis* (pp. 111–125).

Nipko, R., Holcombe, B., & Kelly, M. (2020). Identifying Individual Jaguars and Ocelots via pattern-recognition software: Comparing HotSpotter and wild-ID. *Wildlife Society Bulletin, 44*, 424–433. https://doi.org/10.1002/wsb.1086

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., et al. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences, 115*, 5716–5725. https://doi.org/10.1073/pnas.1719367115

Parham, J. R., Crall, J., Stewart, C., et al. (2017). Animal population censusing at scale with citizen science and photographic identification. In *AAAI spring symposium series*

Patton, P. T., Cheeseman, T., Abe, K., et al. (2023). A deep learning approach to photo-identification demonstrates high performance on two dozen cetacean species. *Methods in Ecology and Evolution, 14*(10), 2611–2625.

Pedersen, M., Nyegaard, M., & Moeslund, T. B. (2023). Finding nemo's giant cousin: Keypoint matching for robust re-identification of giant sunfish. *Journal of Marine Science and Engineering, 11*(5), 889.

Perronnin, F., & Dance, C. (2007). Fisher kernels on visual vocabularies for image categorization. In *Conference on computer vision and*

pattern recognition (CVPR). https://doi.org/10.1109/CVPR.2007.383266

Perronnin, F., Liu, Y., Sánchez, J., et al. (2010). Large-scale image retrieval with compressed Fisher vectors. In *Conference on computer vision and pattern recognition (CVPR)*. https://doi.org/10.1109/CVPR.2010.5540009

Philbin, J., Chum, O., Isard, M., et al. (2007). Object retrieval with large vocabularies and fast spatial matching. In *IEEE Conference on computer vision and pattern recognition* (pp. 1–8).

Pruchova, A., Jaška, P., & Linhart, P. (2017). Cues to individual identity in songs of songbirds: Testing general song characteristics in Chiffchaffs Phylloscopus collybita. *Journal of Ornithology, 158*, 911–924. https://doi.org/10.1007/s10336-017-1455-6

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer assisted intervention (MICCAI)*. https://doi.org/10.1007/978-3-319-24574-4_28

Sarlin, P. E., Cadena, C., Siegwart, R., et al. (2019). From coarse to fine: Robust hierarchical localization at large scale. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12716–12725).

Schneider, S., Taylor, G. W., Linquist, S., et al. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution, 10*, 461–470. https://doi.org/10.1111/2041-210x.13133

Schneider, S., Taylor, G., & Kremer, S. (2020). Similarity learning networks for animal individual re-identification—beyond the capabilities of a human observer. In *Winter applications of computer vision workshops (WACVW)*. https://doi.org/10.1109/WACVW50321.2020.9096925

Schneider, S., Taylor, G. W., & Kremer, S. C. (2022). Similarity learning networks for animal individual re-identification: An ecological perspective. *Mammalian Biology, 102*(3), 899–914.

Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *International conference on computer vision (ICCV)*. https://doi.org/10.1109/ICCV.2003.1238663

Smeulders, A., Worring, M., Santini, S., et al. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1349–1380. https://doi.org/10.1109/34.895972

Suessle, V., Arandjelovic, M., Kalan, A. K., et al. (2023). Automatic individual identification of patterned solitary species based on unlabeled video data. arXiv preprint arXiv:2304.09657

Sun, J., Shen, Z., Wang, Y., et al. (2021) LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8922–8931).

Thompson, J., Zero, V., Schwacke, L., et al. (2019). finFindR: Computer-assisted Recognition and Identification of Bottlenose Dolphin Photos in R. bioRxiv, p. 825661. https://doi.org/10.1101/825661

Thompson, J. W., Zero, V. H., Schwacke, L. H., et al. (2022). finFindR: Automated recognition and identification of marine mammal dorsal fins using residual convolutional neural networks. *Marine Mammal Science, 38*(1), 139–150.

Titterington, D. M., Afm, S., Smith, A. F., et al. (1985). *Statistical analysis of finite mixture distributions*. Wiley

Tyson Moore, R. B., Urian, K. W., Allen, J. B., et al. (2022). Rise of the machines: Best practices and experimental evaluation of computer-assisted dorsal fin image matching systems for bottlenose dolphins. *Frontiers in Marine Science, 9*, 849813.

Tyszkiewicz, M., Fua, P., & Trulls, E. (2020). Disk: Learning local features with policy gradient. *Advances in Neural Information Processing Systems, 33*, 14254–14265.

Vidal, M., Wolf, N., Rosenberg, B., et al. (2021). Perspectives on individual animal identification from biology and computer vision. *Integrative and Comparative Biology, 61*, 900–916. https://doi.org/10.1093/icb/icab107

Webber, T., Lewis, T., Talma, S., et al. (2023). Cetaceans of the Saya de Malha bank region, Indian Ocean: A candidate important marine mammal area. *Regional Studies in Marine Science, 66*, 103164. https://doi.org/10.1016/j.rsma.2023.103164

Weideman, H., Stewart, C., Parham, J., et al. (2020). Extracting identifying contours for african elephants and humpback whales using a learned appearance model. In *IEEE/CVF winter conference on applications of computer vision* (pp. 1276–1285).

Weideman, H. J., Jablons, Z. M., & Holmberg, J., et al. (2017). Integral curvature representation and matching algorithms for identification of dolphins and whales. In *International conference on computer vision workshop (ICCVW)*. https://doi.org/10.1109/iccvw.2017.334

Yeleshetty, D., Spreeuwers, L., & Li, Y. (2020). 3D face recognition for cows. In *International conference of the biometrics special interest group (BIOSIG)*

Yu, H., Xu, Y., Zhang, J., et al. (2021). AP-10k: A benchmark for animal pose estimation in the wild. In *Conference on neural information processing systems (NeurIPS) datasets and benchmarks track*

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

Zhelezniakov, A., Eerola, T., Koivuniemi, M., et al. (2015). Segmentation of Saimaa ringed seals for identification purposes. In *International symposium on visual computing (ISVC)*. https://doi.org/10.1007/978-3-319-27863-6_21

Zheng, Z., Zhao, Y., Li, A., et al. (2022). Wild terrestrial animal re-identification based on an improved locally aware transformer with a cross-attention mechanism. *Animals, 12*(24), 3503.

Zuerl, M., Dirauf, R., Koeferl, F., et al. (2023). PolarBearVidID: A video-based re-identification benchmark dataset for polar bears. *Animals, 13*, 801. https://doi.org/10.3390/ani13050801