Aalborg Universitet

Article

Finding Nemo's Giant Cousin: Keypoint Matching for Robust Re-Identification of Giant Sunfish

Malte Pedersen [1,2,*] , Marianne Nyegaard [3] and Thomas B. Moeslund [1,2]

1   Visual Analysis and Perception Lab, Aalborg University, 9000 Aalborg, Denmark
2   Pioneer Centre for AI, 1350 Copenhagen, Denmark
3   Ocean Sunfish Research Trust, Auckland 1010, New Zealand
*   Correspondence: mape@create.aau.dk

Abstract: The Giant Sunfish (*Mola alexandrini*) has unique patterns on its body, which allow for individual identification. By continuously gathering and matching images, it is possible to monitor and track individuals across location and time. However, matching images manually is a tedious and time-consuming task. To automate the process, we propose a pipeline based on finding and matching keypoints between image pairs. We evaluate our pipeline with four different keypoint descriptors, namely ORB, SIFT, RootSIFT, and SuperPoint, and demonstrate that the number of matching keypoints between a pair of images is a strong indicator for the likelihood that they contain the same individual. The best results are obtained with RootSIFT, which achieves an mAP of 75.91% on our test dataset (TinyMola+) without training or fine-tuning any parts of the pipeline. Furthermore, we show that the pipeline generalizes to other domains, such as re-identification of seals and cows. Lastly, we discuss the impracticality of a ranking-based output for real-life tasks and propose an alternative approach by viewing re-identification as a binary classification. We show that the pipeline can be easily modified with minimal fine-tuning to provide a binary output with a precision of 98% and recall of 44% on the TinyMola+ dataset, which basically eliminates the need for time-consuming manual verification on nearly half the dataset.

Keywords: re-identification; keypoint matching; homography; condition number; computer vision; fish

1. Introduction

The world's heaviest bony fish, the elusive 'giant sunfish' (*Mola alexandrini*), can reach an impressive weight of more than two tons [1,2]. Globally, sunfish are rarely seen by divers, but are frequent seasonal visitors to the Bali area, Indonesia [3]. Here, they seek cleaner fish interaction for removal of skin parasites, and are a highly popular target of the local SCUBA tourism industry [4]. Little is known of this seasonal sunfish phenomenon, including if the tourism is reliant on a small, local sunfish population with high site fidelity, or transient sunfish with low re-visitation rates. Understanding this is critical for assessing the potential impacts (and any need for regulation) of diver crowding, which causes disruptions to sunfish–cleaner fish interactions [3].

To investigate this, the citizen science- and volunteer-based project, Match My Mola [5], collects and curates sunfish images from the Bali area, taken by tourist divers, for photo re-identification purposes. Currently, images are manually compared pair-wise, as illustrated in Figure 1, to assess re-sightings of individuals over time. However, with an increasing number of images, match time becomes a significant challenge. Therefore, we previously proposed a pipeline for automated re-identification of giant sunfish based on keypoint matching, which we published as a workshop paper [6]. We found that our solution worked well as part of a human-in-the-loop system where marine researchers were provided the top-n ranked matches, however, the system relied on manual labor to visually match and sort out wrong matches. In this article, we expand upon our previous work, providing more insights and proposing an enhanced pipeline that includes and additional pre-processing

step to increase the performance. Additionally, we discuss the impracticality of a ranked output and propose a modified pipeline, viewing the re-identification problem as a binary classification, which we argue is more suitable for an efficient human-in-the-loop system.

Our contributions include:

1. A computer-vision-based re-identification pipeline that requires no training or fine-tuning;
2. A demonstration of the generalization attributes of the proposed pipeline;
3. A comparison and performance evaluation of the handcrafted and deep-learning-based keypoint descriptors, i.e., ORB, SIFT, RootSIFT, and SuperPoint, with respect to the re-identification task;
4. A discussion on the impracticality of having a ranked output from re-identification systems and a novel solution to make the proposed system more practical.
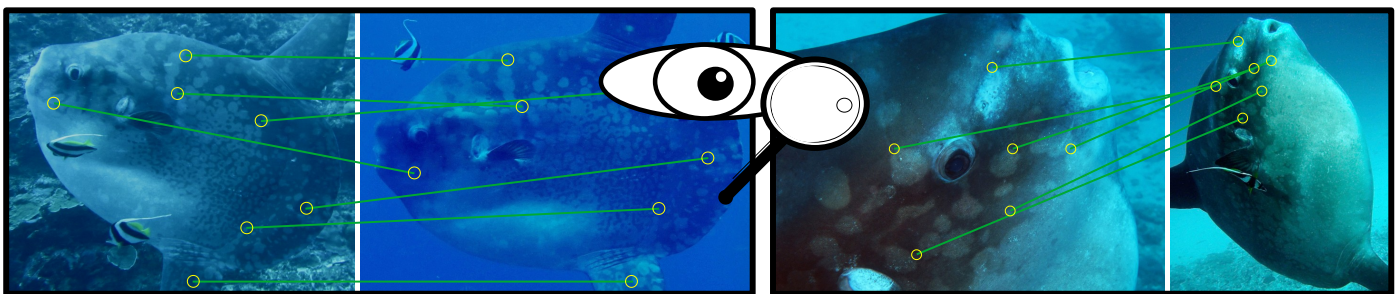

Figure 1. Giant sunfish have unique patterns on their bodies, which can be used for photo identification. Traditionally, marine researchers have matched images by manual visual pattern recognition focused on the body markings, as illustrated in these two examples.

2. Related Work

Photographic identification has been used for studying wild animals for decades [7–10]. It allows researchers to identify the same individual across time and location, but it requires manual labor to obtain the photographs and match the individuals from the captured footage. Citizen science projects and camera traps have proven to be effective and irreplaceable methods to gather large amounts of data. However, as a database grows, so does the need for manual labor for identifying the specimens captured in the images or videos. This has lead to an increase in the use of computer vision systems as an assistive tool within biology and ecology [11–13].

Image processing and pattern matching techniques have been used to identify individuals of many types of animals, including whale-sharks [14,15], spotted raggedtooth sharks [16], fish [17,18], rays [19], seals [20], birds [21], wild terrestrial animals such as zebras, tigers, polar bears, and giraffes [22–26], and farm animals [27,28]. Traditional hand-crafted features, such as SIFT [29], RootSIFT [30], and SURF [31], have been used extensively in animal re-identification [22,27,32–37]. The common pipeline includes a pre-processing step for finding regions of interest in an image concentrated around the target animal. This is often followed by an image enhancement step aimed at distilling unique patterns on the body of the animal. Keypoints are then detected and features (keypoint descriptors) are extracted centered around these points. This is followed by a matching scheme based on minimizing some distance function between the keypoint descriptors in sets of images [38,39]. Lastly, the matches are typically cleaned, e.g., using RANSAC [40], to remove potential outliers, and a final set of matches is used to decide on the IDs of the animals based on some similarity score.

Recently, deep-learning-based methods have become a popular means for handling re-identification problems within marine computer vision. Bergler et al. proposed a multi-stage deep-learning-based framework for detecting and identifying individual killer whales [41]. They trained and used a YOLOv3 model [42] for detecting dorsal fins and a multi-class classification ResNet-34 model [43] for determining the identity of the killer

whales. Wang et al. proposed to use a Siamese network and adversarial training to identify whales by their flukes [44] and Nepovinnykh et al. trained a Siamese network for Saimaa Ringed Seal re-identification [45]. In another work concerning seals, Chelak et al. [20] proposed a new global pooling technique named EDEN and illustrated that the deep features of a modified and fine-tuned ResNet-18 model are suitable for re-identifying Saimaa Ringed Seals. In the work done by Bouma et al., they trained a ResNet-50 model using a triplet loss for identifying dolphins by their flukes [46]. ResNet-50 was also used by Moskvyak et al. in their work on re-identification of manta rays [47,48], where they proposed to embed the feature vectors by body landmark information and use a weighted combination of three losses. On a higher level, Schneider et al. investigated how the performance of CNNs was affected by using either Siamese networks or triplet loss for animal re-identification and found that triplet loss generally outperforms Siamese networks [49].

A common property of all the aforementioned methods is the requirement for training data, parameter fine-tuning, or domain adaptation. However, it is demanding to capture images in wild underwater environments and marine image datasets are, therefore, often sparse. This typically leaves little to no room for high-quality training, testing, and validation splits, as is the case with the giant sunfish dataset that we are evaluating in this work.

3. Match My Mola Re-Identification Dataset

Match My Mola is a citizen science- and volunteer-based project that collects images of sunfish from the Bali area in Indonesia. It is the largest curated collection of sunfish images, that we are aware of, and is a valuable resource for ongoing research in the ecology of the giant sunfish in the Bali area. The images in the database are currently only used for photo identification, but the project aims to expand the use of the images in the future to other areas, such as estimating injury rates from local boats and fishing gear. The photo identification approach allows marine scientists to examine if the same individuals frequent the local reefs several times within and between years, and thereby, better understand if the local tourism industry is reliant on residing or transient individuals. In our previous work on re-identification of sunfish [6], we used a subset of the Match My Mola image database, which we named the 'TinyMola' dataset. It consisted of all the (at that time) manually annotated and verified image pairs of the Match My Mola database which totaled 91 images of 29 individuals. However, the Match My Mola image database contains thousands of photos and researchers and volunteers are continuously matching the images by manual visual inspection. This also means that more individuals have been identified since our previous work was conducted and the annotated part of the Match My Mola database currently contains 224 images of 75 individuals. Therefore, to strengthen our findings, we use an expanded second iteration of the TinyMola dataset that contains the images of the 75 individuals. We name this expanded version of the dataset TinyMola+.

Giant sunfish have unique and intricate whitish body patterns which are well suited to identify individuals [50], as has also been suggested for the close relative *Mola mola* [51]. The contrast of the patterns can vary widely depending both on image quality, environmental factors at the time of photography, but also the physiological state of the patterns themselves. Like many other fish species, giant sunfish are capable of rapid physiological coloration change, whereby low-contrast patterns can become bold and clearly visible in seconds [50]. The patterns themselves, however, are stable during the change, and are also stable over at least 7.2 years [50], and are, therefore, a robust characteristic for photo identification.

The images of the Match My Mola database are grouped in photo events (PhE), which contain 1–3 images per side of the same individual captured by the same diver during the same dive. The markings on the giant sunfish are not identical on the two sides, which is also the case for *Mola mola* [51], and they cannot be directly compared, see Figure 2. Therefore, we frame the re-identification task to be side-specific and provide each side of the fish a unique ID in order to measure the performance of the proposed re-identification pipeline appropriately. For each ID, there are images from at least two PhEs. In Figure 2,

we present images of a giant sunfish named 'Dabra' from two photo events. Notice that this individual has been recorded from both sides, and therefore, has a unique ID for each side.

Figure 2. This giant sunfish has been recorded from both sides in two different photo events (PhE). The patterns on one side cannot be compared to the patterns on the opposite side; therefore, it has a unique ID for each side.

There are cases where two PhEs of the same individual include images from both sides in one of the PhEs but not in the other. With the side-specific composition, TinyMola+ has a total of 224 images containing 83 IDs divided into 41 left-sided and 42 right-sided IDs, with 116 and 108 images, respectively. The images have been gathered from a total of 166 photo events. Most of the images are captured by tourist divers and the quality of the images varies extensively. This is amplified by the turbidity of the water, attenuation of light, occlusion, image compression, and more [52]. Examples illustrating some of the variation of the dataset can be seen in Figure 3. The resolution of the original images varies from 0.1 to 16 megapixels (MP) with a mean around 4 MP and an object resolution around 1 MP on average. However, all the images from the Match My Mola database have been manually cropped around the sunfish by the researchers who curate the database to ease their manual inspections. We use the cropped images in our work to focus on the re-identification task.

Figure 3. Example images from TinyMola+. The quality varies widely between the images and some have high resolutions, clear objects, and distinct patterns, while others have very low resolutions, reduced contrast, and vague patterns.

4. Method

The process of identifying the sunfish on the images is currently conducted manually by marine researchers or trained volunteers. This includes cropping the image around the target and comparing markings across all overlapping body parts on the two images. The images are compared pair-wise and matches are noted and examined by other matching experts, to confirm that the images are of the same individual. Our solution, described below and illustrated in Figure 4, is inspired by the manual process and is an improvement to our previous work on the re-identification of the giant sunfish [6], with an additional pre-processing step.
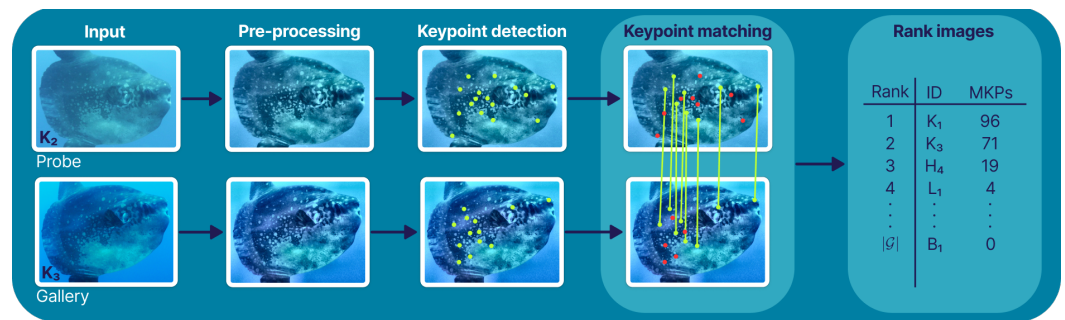
Figure 4. The proposed module-based pipeline with illustrations of an image pair containing the same individual with ID = K (the subscripted number signifies that the images are not the same). In the pre-processing module, the image contrast is enhanced for both images and then the keypoints are detected and matched in the following two modules. Lastly, the gallery images are ranked based on the number of matching keypoints (MKPs), where a higher number indicates a stronger similarity.

4.1. Pre-Processing—Contrast Enhancement

The quality of the images in the TinyMola+ dataset varies to a large degree depending on when and where the images were taken and by whom. This leads to the patterns on the sunfish being less pronounced in some images, e.g., due to the low contrast or low quality of the image. For this reason, we investigated how contrast enhancement may affect the performance of the pipeline. The aim was not to create realistic out-of-the-water images of the sunfish, but rather to enhance the clarity of the patterns to potentially allow for an increased number of distinct keypoints. It should be noted that this module is an addition to our previously proposed method [6].

For enhancing the contrast, we chose the well-proven contrast limited adaptive histogram equalization method [53] (CLAHE). CLAHE enhances the contrast in the image adaptively by processing the image as a set of smaller patches and equalizing the local patch-based histograms (in opposition to a global equalization, which often leads to undesirable results). In addition to the adaptive equalization, CLAHE includes a clipping step to minimize the enhancement of noise. We present a range of varied examples from TinyMola+ processed by CLAHE in Figure 5.

Figure 5. The images in the first row have not been processed. The second row contains the same images after the contrast has been enhanced using the CLAHE algorithm. Notice how the patterns stand out more clearly as the contrast is increased.

4.2. Keypoint Detection

The body of a sunfish is highly rigid, except for the dorsal and anal fins [54]. Consequently, sunfish captured on images at different times and locations are mainly affected by affine transformations, such as rotation and scale, and we can utilize this to determine whether a pair of images contains the same individual. Detecting and describing points affected by such transformations have been studied intensively for decades in fields such as image registration and tracking. Candidate locations are typically known as keypoints

or interest points and they must be characteristic in some manner, e.g., a corner or a high intensity pixel in a low intensity neighborhood. In this work, we evaluated the performance of three handcrafted feature descriptors, i.e., SIFT [29,55], RootSIFT [30], and ORB [56], and the state of the art deep-learning-based feature descriptor SuperPoint [57] with respect to the re-identification of individual sunfish. Each of the aforementioned methods are summarized in this section.

Probably the most widely used hand-crafted keypoint descriptor is the Scale Invariant Feature Transform (SIFT) [29,55], which was proposed two decades ago. SIFT features are based on extrema points that are consistent throughout a difference of Gaussians scale space. When an extrema point has been located, a histogram of oriented gradients (HoG) is created from the region surrounding the pixel. An orientation is assigned to the keypoint based on the normalized HoG features. The SIFT keypoint descriptor itself is based on a 4×4 matrix of normalized HoG features with 8 bins, resulting in a feature vector with 128 values.

Following the publication of SIFT, Arandjelovic and Zissermann noted that matching the features using Euclidean distance, as proposed in the original paper, is not always the best solution [30]. SIFT features are based on histograms and the Hellinger kernel is typically preferred over Euclidean distance when comparing histograms. Therefore, Arandjelovic and Zissermann proposed to L1-normalize the SIFT feature vector and take the square root of each element subsequently, naming this new feature descriptor RootSIFT. Practically, this means that matching RootSIFT features using the Euclidean distance is equivalent to matching SIFT features using the Hellinger kernel, which typically improves the results.

A more recent hand-crafted feature descriptor is the Oriented FAST and Rotated BRIEF [56] (ORB), which, as the name implies, is based on a combination of the FAST keypoint detector [58] and BRIEF keypoint descriptor [59]. FAST and BRIEF were designed to be both fast and accurate, but have several downsides. FAST responds greatly along edges, but has no way of measuring cornerness, which is otherwise typically considered a salient feature. In ORB, this is handled by sorting the keypoints using a Harris corner measure [60] and only picking the top candidates. Furthermore, a scale pyramid is employed and features are produced at each level to handle the lack of scale-invariant keypoints. Lastly, a keypoint orientation is included based on the assumption that the intensity of corners are offset from their centers. This also solves the main problem with the BRIEF descriptor not being able to handle rotations. The orientation of the FAST corner keypoint is used to steer the BRIEF descriptor, which thereby becomes rotation-invariant.

Learned feature descriptors have gained attention with the popularization and accessibility of strong GPUs to efficiently train deep learning models. A seminal work in this field is the self-supervised keypoint detector and descriptor SuperPoint [57], which is a fully convolutional neural network (CNN) with a shared encoder and two decoder heads for keypoint detection and description, respectively. The authors proposed to train a base model with just the decoder head for keypoint detection, named MagicPoint, on purely synthetic data of angular-shaped objects. They then used MagicPoint to create pseudo ground-truth labels on real images. This was done by warping the input images using random homographic transformations, detecting keypoints in the warped images, and aggregating the unwarped set of keypoints into a superset of labels. The homographic adaption of real images allows for jointly self-supervised training of the SuperPoint keypoint detector and descriptor to be invariant to scaling and rotation.

Giant sunfish re-identification presents challenges to all four descriptors. For example, Figure 6 shows two image pairs of the same individual along with the matching keypoints (MKPs). The first example contains two images with relatively low contrast, while the individuals in the second example are rotated in relation to each other. These examples highlight some of the obstacles that can complicate keypoint detection and matching for the respective algorithms.
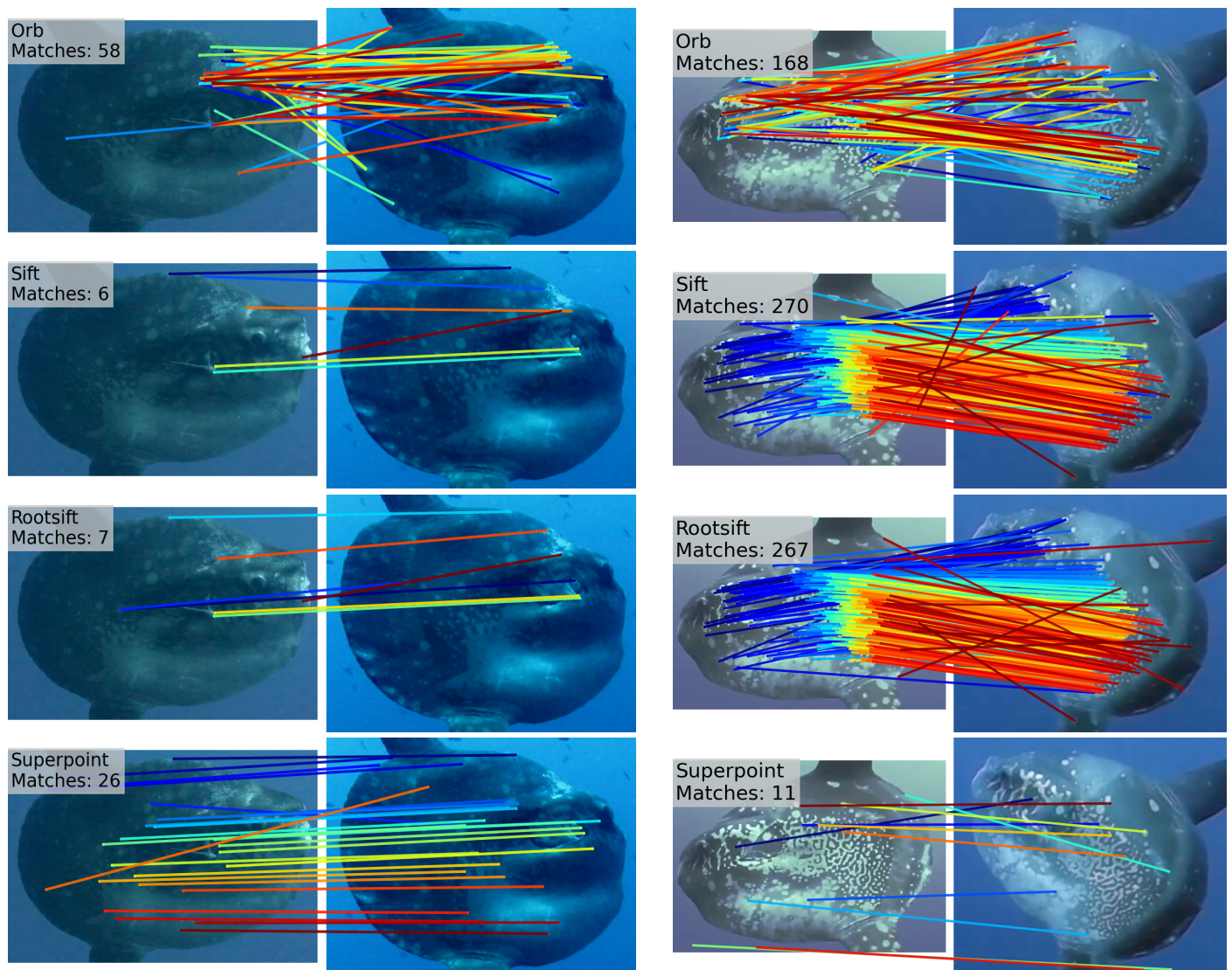
Figure 6. Two examples of image pairs with certain characteristics that can complicate the process of detecting and matching keypoints. The image pairs in the first example (left) have relatively low contrast, while the image pairs in the second example (right) are heavily rotated in relation to each other. In the first example, ORB finds a multitude of matching keypoints, though a large part of them are false positives, SIFT and RootSIFT find few and relatively imprecise keypoints, and lastly, SuperPoint finds a decent amount of mostly precise keypoints. In the second example, ORB finds both correct and incorrect matching keypoints, SIFT and RootSIFT find many correct matches, and SuperPoint finds very few matches. Note that the images have not been contrast-enhanced and the colors of the lines are only for visualization.

4.3. Keypoint Matching

The keypoints are described by feature vectors, and to determine whether keypoints in two images represent the same point on the object, we measured the distance between the vectors. Depending on the problem, dimensionality, and nature of the data, keypoint matching has commonly been performed using brute-force methods or kd-trees [61]. Brute-force methods compare all elements in the two distributions and are guaranteed to find the best match, but the processing time can be high for large distributions. On the other hand, kd-trees do not guarantee to find the best match, but are faster for large distributions. As there are, in our case, no time constraints on the task and the dataset is relatively small, we performed an exhaustive search and matched the keypoints using a brute-force method. SIFT, RootSIFT, and SuperPoint features were matched based on the L^2 distance and the

ORB features were matched based on the Hamming distance due to the binary nature of the features.

Naively matching the closest keypoints can lead to poor results. For this reason, David G. Lowe introduced the distance ratio test [29] as a way to dismiss keypoints that are ambiguous. If the ratio between the distance to the nearest and second nearest neighbor is above a threshold, the keypoint is considered too uncertain and is discarded. The optimal threshold depends on the nature of the data, and if it is too low, too many correct matches may be discarded and vice versa. We used the distance ratio test as the last step of the keypoint matching module to clean our matches.

4.4. Ranking Images

For every image pair, we viewed the number of matching keypoints as a similarity score, where a higher number of MKPs indicates a stronger similarity. We sorted and ranked all images based on the number of MKPs, as illustrated in Figure 4. Note that the example in the figure is hypothetical and only for visualization purposes.

5. Evaluation Protocol

It is not possible to manually determine whether the left and right side of a giant sunfish belong to the same individual, except where photos exist of both sides during the same photo event. Therefore, each side of an individual was assigned different IDs. In cases where a photo event contains images of both sides, but only one of the sides has a match from another photo event in the dataset, the unmatched image was named a *single* and was considered to be noise. Every image, except singles, was considered a *probe* $p \in \mathcal{P}$. Each probe was compared against the *gallery* images, with $g \in \mathcal{G}$ being the set of all images in the dataset except the probe. Note that the singles were included in the gallery and there was always at least one gallery image with the same ID as the probe.

5.1. Performance Metrics

Re-identification systems are typically evaluated based on their ability to rank the gallery images by their similarity to the probe. Two of the most commonly used metrics for evaluating ranking-based re-identification systems are the cumulative matching characteristic (CMC) [62] and the mean average precision (mAP) [63]. The CMC describes the accuracy of the system at a given rank and is often presented as rank-k accuracy. The CMC score is inadequate in cases where the gallery contains multiple images that ID with the probe, as it only refers to the highest ranked true positive gallery sample. Therefore, we also evaluated our system by the mAP, which punishes suboptimal ordering of the ranked gallery images. The CMC score at rank x was computed as follows:

$$\text{CMC}^x = \frac{1}{|\mathcal{P}|} \sum_p^{\mathcal{P}} \begin{cases} 1, & \text{if any of the top-}x \text{ ranked gallery images shares ID with } p \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

We calculated the average precision for probe p at rank x as follows:

$$\text{AP}_p^x = \frac{1}{\mathcal{H}_p} \sum_{n=1}^{x} \text{pr}_n R_n \tag{2}$$

where $\mathcal{H}_p = \min\{|g_p|, x\}$, $|g_p|$ is the total number of gallery images that shares ID with the probe, and R is a relevance function given by:

$$R = \begin{cases} 1, & \text{if the gallery image is a true positive} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Moreover, pr is the precision, calculated as follows:

$$\text{pr} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

where TP is the number of true positives and FP is the number of false positives. Finally, the mAP for rank x was found by:

$$\text{mAP}^x = \frac{1}{|\mathcal{P}|} \sum_p^{\mathcal{P}} \text{AP}_p^x \tag{5}$$

5.2. Pipeline Parameters

An essential aspect of this work was to design a pipeline that is suited for non-technical staff. Therefore, in order to minimize the need for user involvement, we chose default parameters for the methods included in the pipeline. In the pre-processing module, the image contrast was enhanced with CLAHE. We did not fine-tune the CLAHE parameters, but used a patch-size of 8×8 pixels and a clipping limit of 3, which are commonly used settings. We used the default settings for the four keypoint descriptors and we cleaned the matches of SIFT, RootSIFT, and SuperPoint using the distance ratio test with a default threshold of 0.8, as proposed by Lowe in the original SIFT paper [29]. We generally used Python as the programming language and the OpenCV library for implementing the image processing algorithms, such as CLAHE, SIFT, RootSIFT, and ORB. We used the pre-trained SuperPoint model from Magic Leap [57] implemented in PyTorch.

5.3. Testing Data

Beside evaluating our pipeline on the TinyMola+ dataset, we included two additional re-identification datasets with patterned animals for a more thorough assessment of the system, namely, the SealID$_{\text{patches}}$ [20] and OpenCows2020 [28] datasets. As can be seen from the examples in Figure 7, the three datasets vary widely with respect to object appearance, image quality, and contrast, as well as the number of images per individual.

- The TinyMola+ dataset contains 83 individuals and 224 images of varying sizes.
- The SealID$_{\text{patches}}$ test split contains 26 individuals and 836 images of size 256×256.
- The OpenCows2020 test split contains 46 individuals and 496 images of varying sizes.
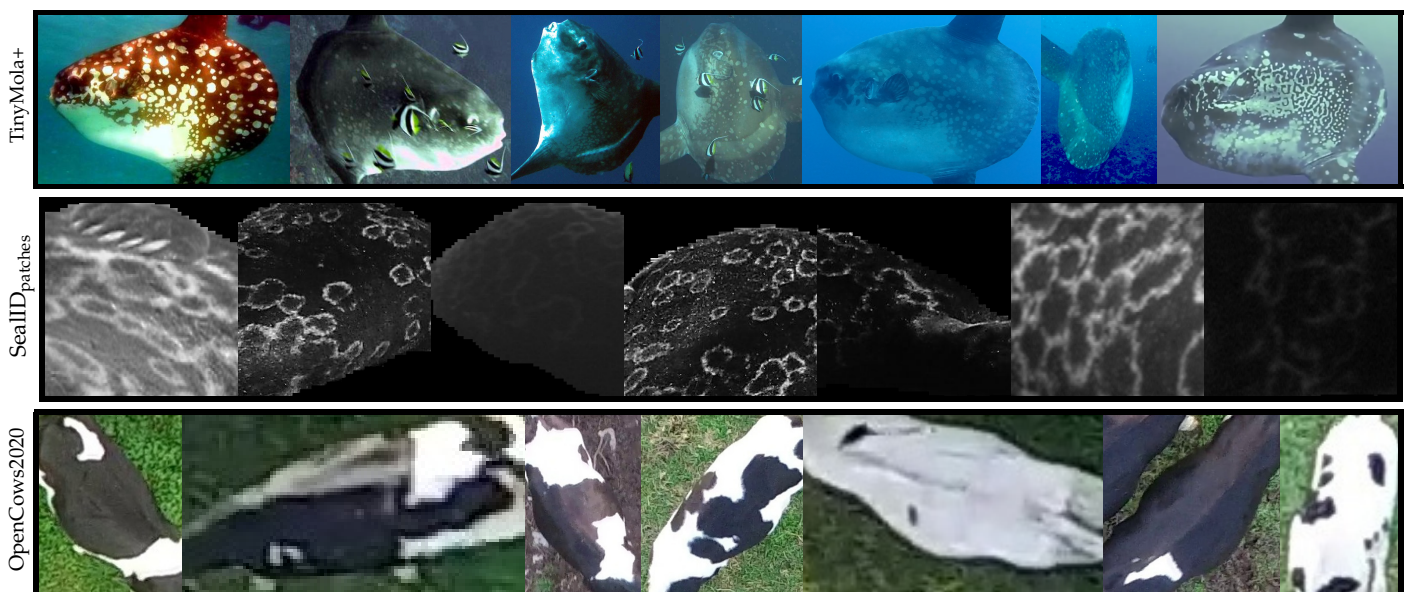

Figure 7. Examples from the TinyMola+, SealID$_{\text{patches}}$ [20], and OpenCows2020 [28] datasets.

During evaluation, all images of the TinyMola+ and OpenCows2020 datasets were resized to a maximum dimension of 640 pixels while keeping the aspect ratio to ensure that the size of the objects were approximately similar between the images. We did not resize the images of the SealID$_{patches}$ dataset, as they had already been resized to 256×256 by the authors of the dataset in order to ensure that the patterns are of approximately the same scale. Note that we exclusively evaluated on the testing splits; we did not use the training splits, as we did not train nor fine-tune our pipeline.

6. Results

Recall that the aim of this work was to develop an automated re-identification pipeline that requires no training data, as it can be extremely difficult and time consuming to capture sufficient data to train robust supervised models for marine tasks due to the harsh underwater environment. Furthermore, underwater environments can vary widely visually, which means that the pipeline should be able to generalize well. Hence, we conducted an evaluation of the efficiency and adaptability of the suggested pipeline equipped with each of the keypoint descriptors, in order to determine the optimal choice among the four candidates.

First, we demonstrate the superiority of the new proposed pipeline compared to the former pipeline [6] on the TinyMola+ dataset. Hereafter, we show that the pipeline generalizes to other re-identification subjects with distinct patterns by evaluating the system on two very different datasets: SealID$_{patches}$ [20] and OpenCows2020 [28]. In Table 1, we present results from our former and new pipeline on the TinyMola+ dataset. We see a tendency indicating that the number of matching keypoints can serve as a strong predictor for determining whether two images contain the same individual. Our new pipeline outperforms the former solution with the SIFT, RootSIFT, and SuperPoint descriptors; however, it performs significantly worse with ORB. The descriptor that obtains the best performance on the TinyMola+ dataset is RootSIFT, which reaches an mAP of 75.91%.

Table 1. Results from the former [6] and current pipeline on the TinyMola+ dataset. We present the CMC score for three ranks (1, 3, and 5). The best results are highlighted in bold. The difference between the former and current solution is highlighted in green if the current solution is better and red otherwise.

Model	TinyMola+			
	CMC^1	CMC^3	CMC^5	mAP
Former$_{SuperPoint}$	69.20	72.32	75.00	60.74
Ours$_{ORB}$	29.02 (-40.18)	36.16 (-36.16)	39.73 (-35.27)	23.97 (-36.77)
Ours$_{SIFT}$	76.79 (+7.59)	82.14 (+9.82)	83.04 (+8.04)	70.20 (+9.46)
Ours$_{RootSIFT}$	**80.36** (+11.16)	**84.38** (+12.06)	**86.16** (+11.16)	**75.91** (+15.17)
Ours$_{SuperPoint}$	72.32 (+ 3.12)	77.23 (+ 4.91)	77.68 (+ 2.68)	63.88 (+ 3.14)

We present an overview of the results from the SealID$_{patches}$ and OpenCows2020 datasets in Table 2. On SealID$_{patches}$, our pipeline outperforms the deep-learning-based and supervised solution proposed by the authors of the dataset [20] with respect to the rank-1 accuracy. This is not the case for the OpenCows2020 dataset, where the authors present a pre-trained ResNet-50 model fine-tuned with a combination of a softmax and reciprocal triplet loss (RTL) [28] that we are not able to match, although we obtained reasonable results. It should be noted that it is unclear exactly how the authors of both datasets calculate their accuracy, but to the best of our knowledge, it is the CMC rank-1 accuracy. Additionally, we evaluated our method on the entire testing split for both datasets and did not consider portions of known/unknown ID's between the testing and training splits as we did not need to train nor fine-tune our pipeline as opposed to the other two solutions that needed annotated training data.

We see that RootSIFT has a marginally higher mAP compared to SuperPoint on the $SealID_{patches}$ dataset, while the opposite is true for the CMC rank-1 score. This indicates that the two descriptors basically perform equally well on the dataset and they are closely followed by SIFT. It is another story when looking at the results for the OpenCows2020 dataset. SIFT and RootSIFT get extremely low CMC rank-1 scores and the mAPs are also substantially lower compared to SuperPoint. This is possibly due to CLAHE not having the desired effect on the OpenCows2020 dataset, which contains a multitude of very small images. A suboptimal configuration of CLAHE may induce an enhancement of noisy elements instead of the actual patterns on the cow (which already have a high contrast due to their naturally black and white colorization).

Table 2. Results from the $SealID_{patches}$ [20] and OpenCows2020 [28] datasets. We present the CMC score for three ranks (1, 3, and 5). The best results are highlighted in bold. Note that our pipeline runs off-the-shelf, while the other solutions are trained specifically for the task at hand.

	SealID$_{patches}$				OpenCows2020			
Model	CMC^1	CMC^3	CMC^5	mAP	CMC^1	CMC^3	CMC^5	mAP
EDEN [20]	86.54	-	-	-	-	-	-	-
ResNet50$_{Softmax-RTL}$ [28]	-	-	-	-	**87.55**	-	-	-
Ours$_{ORB}$	77.87	83.49	86.24	31.70	35.08	43.55	50.60	22.59
Ours$_{SIFT}$	92.82	95.93	96.29	49.95	0.81	77.02	83.67	28.75
Ours$_{RootSIFT}$	93.18	**97.01**	**97.49**	**57.34**	0.81	82.46	**86.69**	31.51
Ours$_{SuperPoint}$	**94.86**	96.89	97.13	56.97	73.79	**83.27**	85.69	**39.52**

We present three examples from the OpenCows2020 and $SealID_{patches}$ datasets in Figures 8 and 9 for each of the descriptors, respectively. All the images in the examples have been processed by CLAHE. We see a tendency that ORB generally finds many, but unreliable, matching keypoints. SIFT and RootSIFT find many true MKPs, but also a portion of false MKPs between images of the same individual. SuperPoint generally finds fewer MKPs compared to the other descriptors, but they seem to be more robust and very few false positives are found.
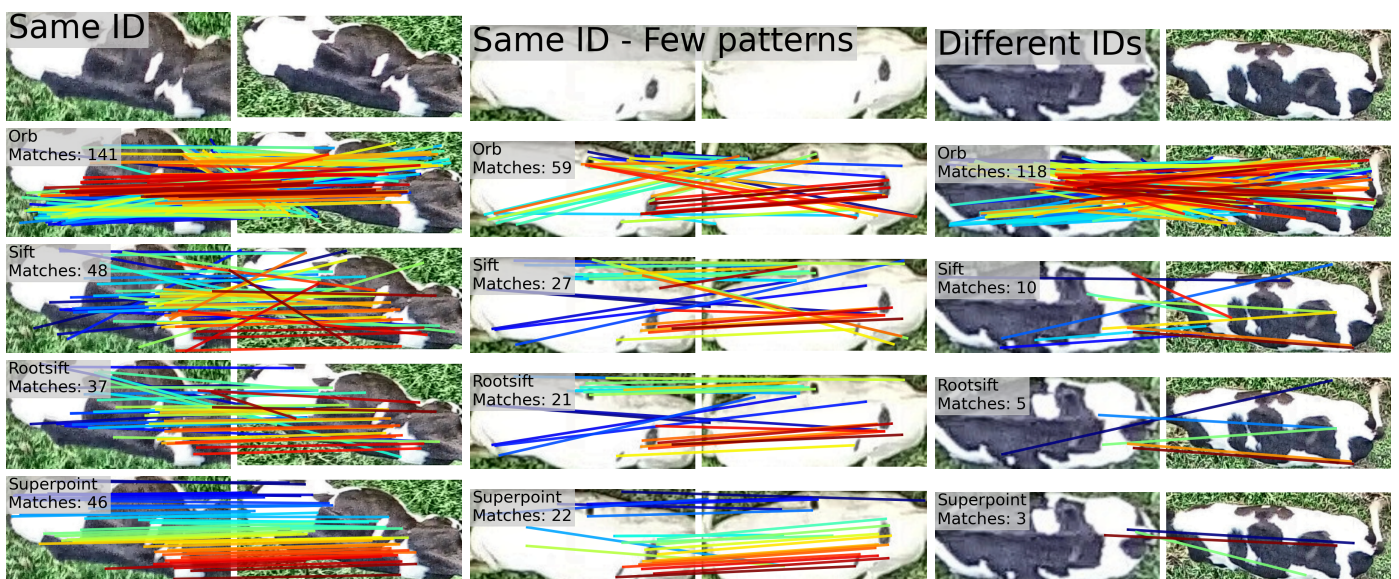

Figure 8. Matching examples from the OpenCows2020 dataset. The images have been processed by CLAHE. The colors of the lines are only for visualization.
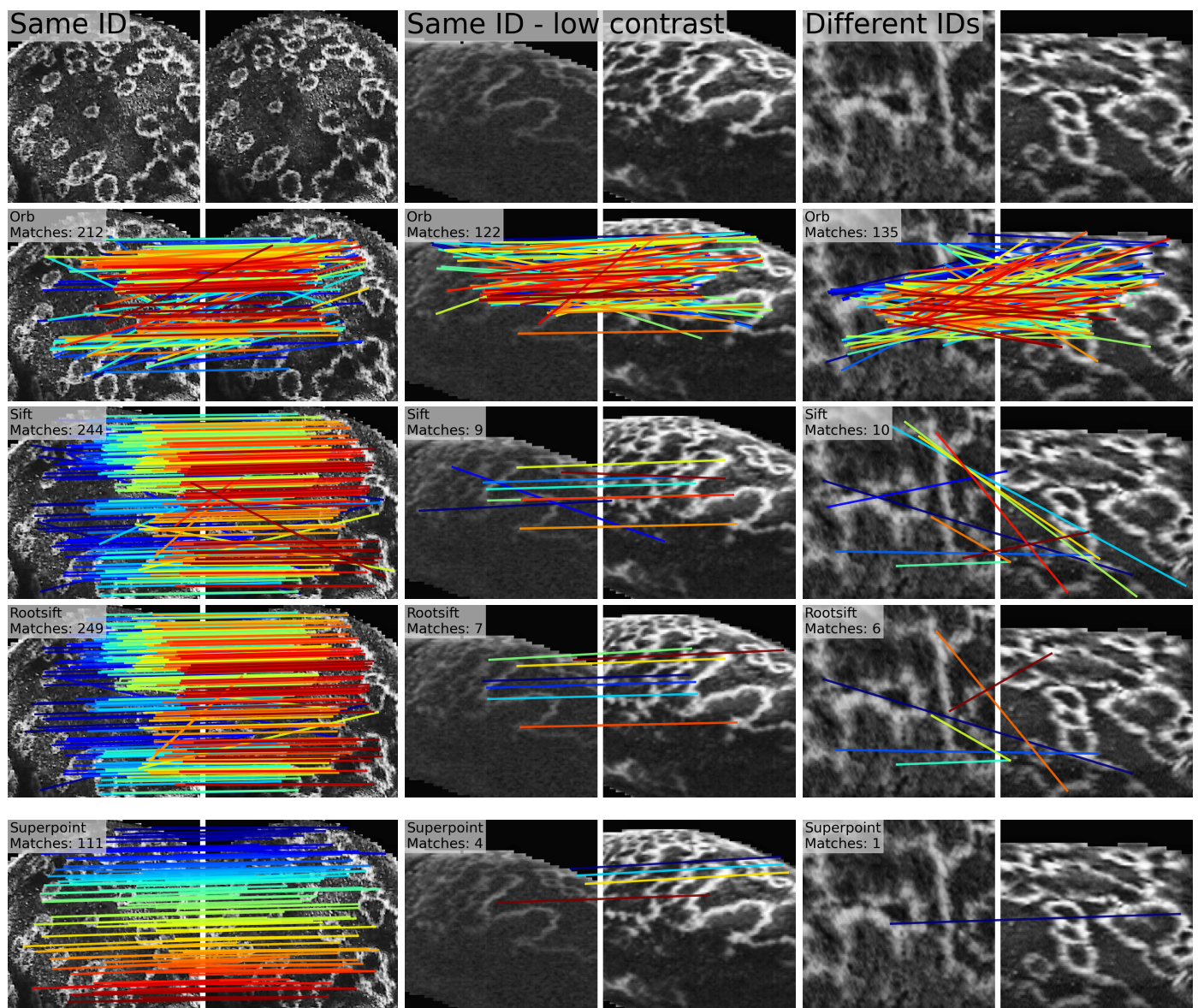
Figure 9. Matching examples from the SealID$_{patches}$ dataset. The images have been processed by CLAHE. The colors of the lines are only for visualization.

7. Discussion

We have seen that our proposed pipeline performs well on the TinyMola+ dataset and also seems to generalize well to other similar tasks. However, conducting meaningful research on wild and elusive animals such as the giant sunfish based on re-identification is challenging due to the time-consuming task of obtaining and analyzing sufficient amounts of data. Therefore, it is not uncommon that volunteers assist in data collection and data curation in environmental and conservation projects. However, this also entails that the personnel on these projects typically have diverse backgrounds and it cannot be expected that they have technical skills to configure or train complicated computer vision systems. Beside proposing a system that works in practice out-of-the-box, an important part of this work is to design a system that requires absolute minimal intervention from the user. A common approach to solve re-identification tasks is by providing the top-n ranked images based on some similarity score, as we did above. However, a ranked output has some negative application-specific attributes:

1. It is not obvious how to decide the optimal rank;
2. It is time consuming to manually verify matches (both positive and negative);

3. It is difficult to evaluate the practicality of the system by standard metrics, such as mAP and CMC.

In real-life applications, there is typically a human in the loop that needs to verify the output of the re-identification system. Often, the user needs to decide on the number of gallery images to look through (the rank). If the rank is too low, the user will miss positive samples, and if the rank is too high, the user will have to look through a multitude of false positives.

In short, a ranking-based output is not very practical in real-life applications. Alternatively, we suggest that the re-identification task can be viewed as a binary classification problem, where a pair of images can either contain the same individual or not. This allows for an arbitrary number of gallery images that share an ID with the probe while liberating the user from deciding on the number of proposals per probe to look through. In the following section, we present a novel binary classification module as an alternative to the ranking module and discuss its strengths and weaknesses.

7.1. Re-Identification as a Binary Classification Problem

Only minor adjustments are required for the pipeline to deliver a binary output. One method is to accept every image pair that has at least a single pair of matching keypoints as a positive sample. However, as we know that all the keypoint descriptors are likely to find noisy MKPs, this will lead to a huge number of disordered false positive identifications, which is even less practical than the ranked output. The key to a robust binary classification module is a very high precision, meaning that very few false positives are accepted.

7.1.1. Thresholding the Minimum Number of Matching Keypoints

A way to minimize the number of false positives is to find a threshold for the minimum number of MKPs needed for an image pair to be considered a positive match. However, such a threshold includes finding a compromise between reducing the number of false positives and increasing the number of false negatives. This compromise can be visualized through a precision–recall curve, where the precision and recall are calculated as:

$$\mathrm{pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{rc} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad (6)$$

where TP are true positives, FP are false positives, and FN are false negatives. A fabricated example with a probe and five gallery images can be seen in Figure 10. In the given example, one image has been correctly matched (a TP) and three others have been wrongly matched (three FPs), illustrated by the green check mark and red crosses, respectively. The right-most image has the same ID as the probe, but has not been matched (an FN). In the given example, $\mathrm{pr} = 0.25$ and $\mathrm{rc} = 0.5$.
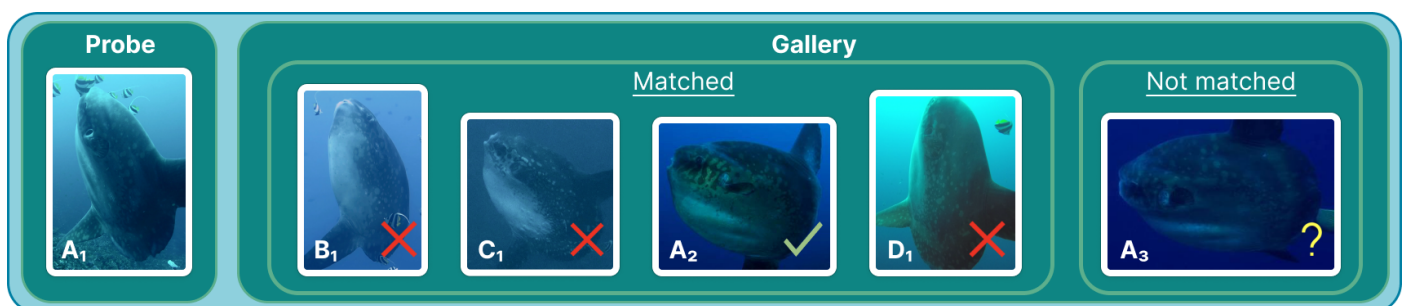

Figure 10. An example with a probe and a gallery containing five images. The ID of each fish is presented in the bottom left corner of the image by a letter. Four of the gallery images have been matched with the probe; however, only the gallery image with the green check mark is correctly matched. The gallery image that has not been matched with the probe shares the ID of the probe. In total, this gives three false positives, one true positive, and one false negative.

The optimal compromise between precision and recall depends on the task at hand. In our case, precision is critical and we want the highest possible precision in order to remove the need for manual labor of verifying the samples. We present two precision–recall plots for the TinyMola+ dataset in Figure 11 based on varying the minimum number of MKPs. Both plots contain curves for each of the four keypoint descriptors.

The left plot contains a traditional precision–recall curve that shows that SuperPoint reaches the best compromise between precision and recall, which is highlighted by the area under the curve (AUC) presented in the legends. The circles refer to the tipping point of the precision recall-curve's 'shoulder', where the precision starts to decrease. The plot on the right is an elaboration of the precision and recall values with respect to the MKPs threshold. Precision and recall are both visualized on the vertical axis, while the horizontal axis indicates the minimum number of MKPs. Note that the circles in the two plots mark the same precision and recall values.
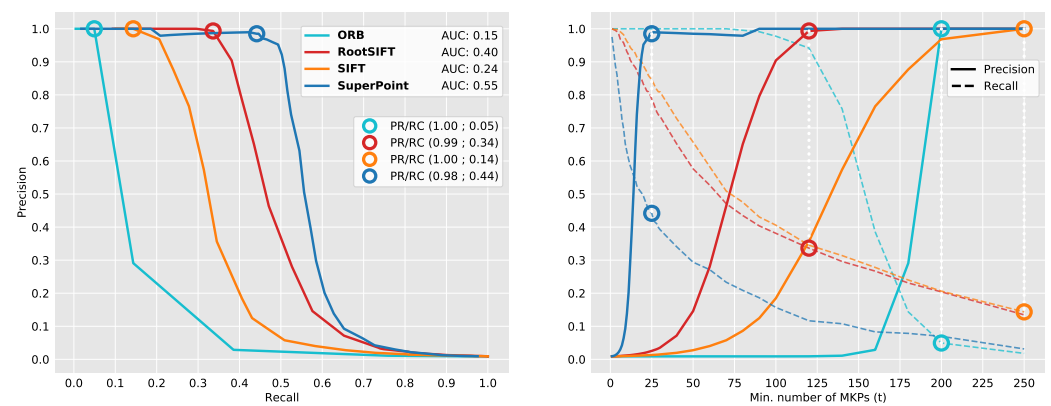

Figure 11. The left plot shows a precision–recall curve based on varying the minimum number of MKPs needed for an image pair to be classified as a positive match. The right plot expands upon the precision–recall curve by showing both precision and recall plotted on the vertical axis and the minimum number of matching keypoints on the horizontal axis. The circles in both plots highlight the precision and recall values at the 'shoulder' of the precision–recall curve.

ORB and SIFT give the worst performance with high precision at the expense of very low recall. RootSIFT outperforms the two other handcrafted descriptors by reaching a recall of 0.34 and a precision of 0.99 at $t = 120$. Lastly, we observe that SuperPoint is able to reach a recall of 0.44 and precision of 0.98 at $t = 25$. This means that 0.44% of the matches of the TinyMola+ dataset are found with a very high precision among the image pairs that shares at least 25 MKPs, and it basically removes the need for user involvement in almost half of the dataset. However, the number of MKPs alone is not the only parameter that we can tune to remove false positives. By analyzing the composition of the matching keypoints, we may be able to allow fewer MKPs, and thereby, a higher recall, while preserving a high precision.

7.1.2. Thresholding the Maximum Condition Number

The body of a sunfish is nearly flat and completely rigid, meaning that it approximates a plane well. We can exploit this by assuming that the matching keypoints between an image pair point have the same spatial positions on two identical planes, which allows us to compute the homography and estimate the change in rotation and distance between the images. In our case, the homography describes the projective transformation between the planes spanned by the bodies of the sunfish, which is naturally constrained. In the odd cases where we have matching keypoints between images of *different* individuals, the projective transformation between the planes will be unconstrained and ambiguous. We can utilize this to minimize the number of false positive matches by discarding image pairs with unlikely projective transformations.

A way to determine the unlikeliness of a projective transformation is by looking at the condition number of the homography matrix. The condition number, κ, indicates to what degree a change in the input affects the output. In the case of a homography matrix, this means that if the matrix is based on a range of correct MKPs, a limited projective transformation is produced, after which a small change to the input will only cause a small change to the output. However, if the matches between the image pairs are wrong, they will point to random spatial positions on the planes and not agree on a common transformation. This leads to a system where even small changes to the input will significantly alter the output. We calculate the condition number as follows:

$$\kappa(H) = \frac{\sigma_{\max}(H)}{\sigma_{\min}(H)} \tag{7}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values, respectively, and H is the homography matrix computed from the matching keypoints. A minimum of four matching keypoints are needed in order to estimate the homography [64], but more MKPs are preferred to minimize the impact of noisy matches. The condition number lies in the interval $[1, +\infty]$, where a lower number implies a stronger candidate for a correct match and a higher number indicates a more complex and unlikely transformation.

The examples presented in Figure 12 are visual illustrations of the correlation between the condition number of the homography matrix and the soundness of the output image. Note that SIFT are used in all the examples; however, similar patterns are observed for the other descriptors. The first example contains an image pair of the same individual captured from different angles. The MKPs are largely correct, and this is illustrated by a relatively simple projective transformation that causes the first image to align nicely with the second image. The following four examples contain different individuals, meaning that all MKPs are wrong. This is illustrated by complicated projective transformations that lead to absurd and unrecognizable projected images.

7.2. Evaluating the Binary Classification

In Figure 13, we present plots with precision–recall curves for each of the keypoint descriptors when thresholding both the minimum number of MKPs and the maximum condition number. Each curve in the plot is based on varying the threshold, t, for the minimum number of MKPs, while the color of the curve indicates the value of the condition number threshold parameter $\mathcal{L}$. The four highlighted curves resemble the curves presented in Figure 11, with no threshold on the condition number. We see that the performance of SIFT, RootSIFT, and ORB can be significantly improved by thresholding both parameters, whereas the gain is negligible for SuperPoint. This is highlighted by the AUC presented in the legends, which is sorted based on the threshold value $\mathcal{L}$.

The best performance, according to the AUC, is obtained when $\mathcal{L} \approx 1 \times 10^5$, while lower AUC scores are seen for both higher and lower threshold values. However, SuperPoint is an exception to the latter as it reaches an optimal and stable performance for $\mathcal{L} \geq 1 \times 10^6$. This indicates that the SuperPoint features are more robust compared to the other descriptors and there is no substantial gain in looking at the condition number, as every image pair that has at least 25 MKPs is practically certain to be a correct match. It is possible to obtain comparable results with SIFT and RootSIFT with respect to the AUC, but it requires the user to fine-tune the minimum number of keypoints as well as the maximum condition number, making them less applicable compared to SuperPoint.
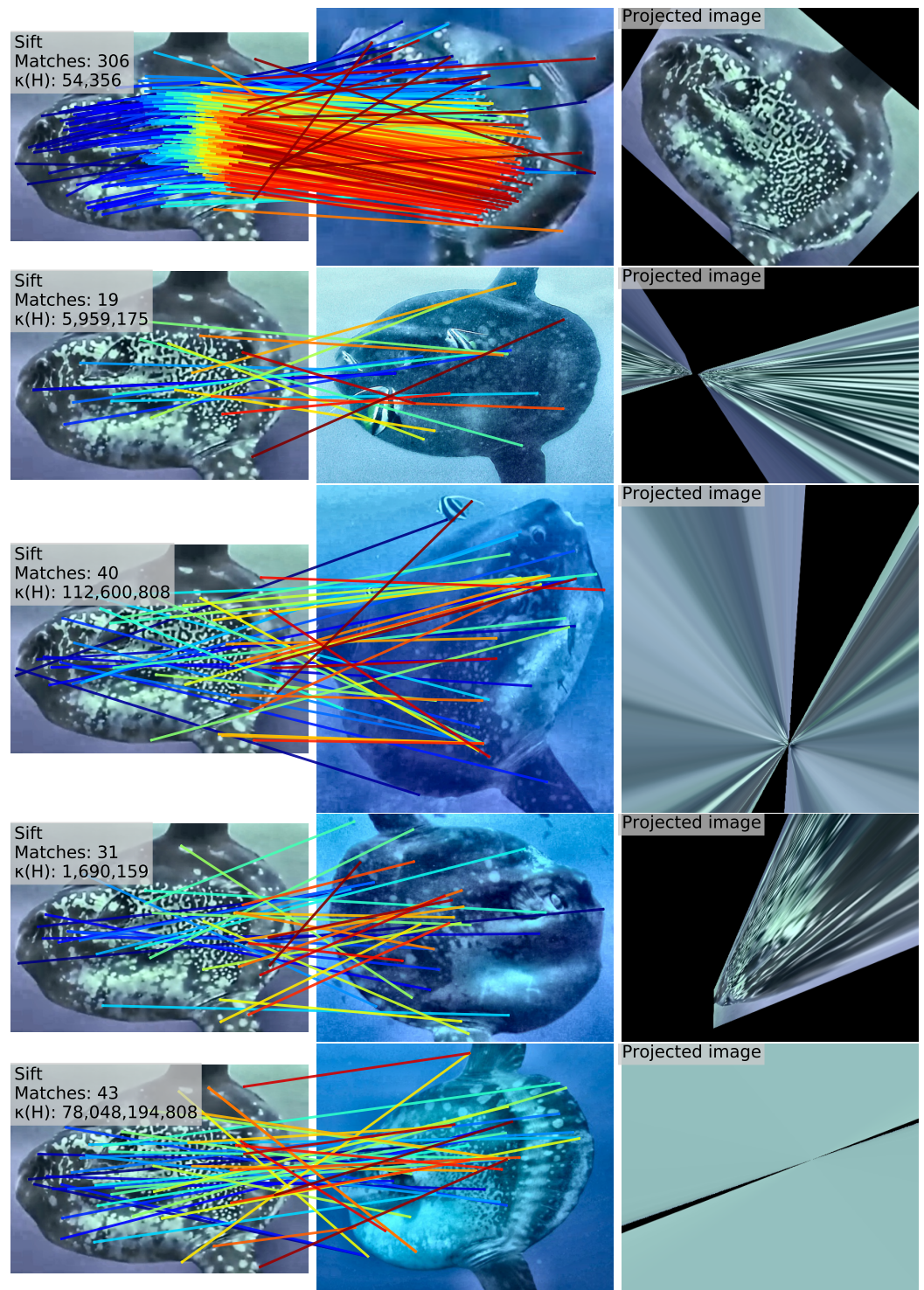
Figure 12. The example in the first row shows an image pair of the same individual and the projected image in the third column is well aligned with the second image, as expected. The following four rows show image pairs that contain different individuals, which means that the matching keypoints are wrong per definition. The consequence is odd homography matrices that lead to absurd projective transformations, as illustrated by the projected images in the third column.
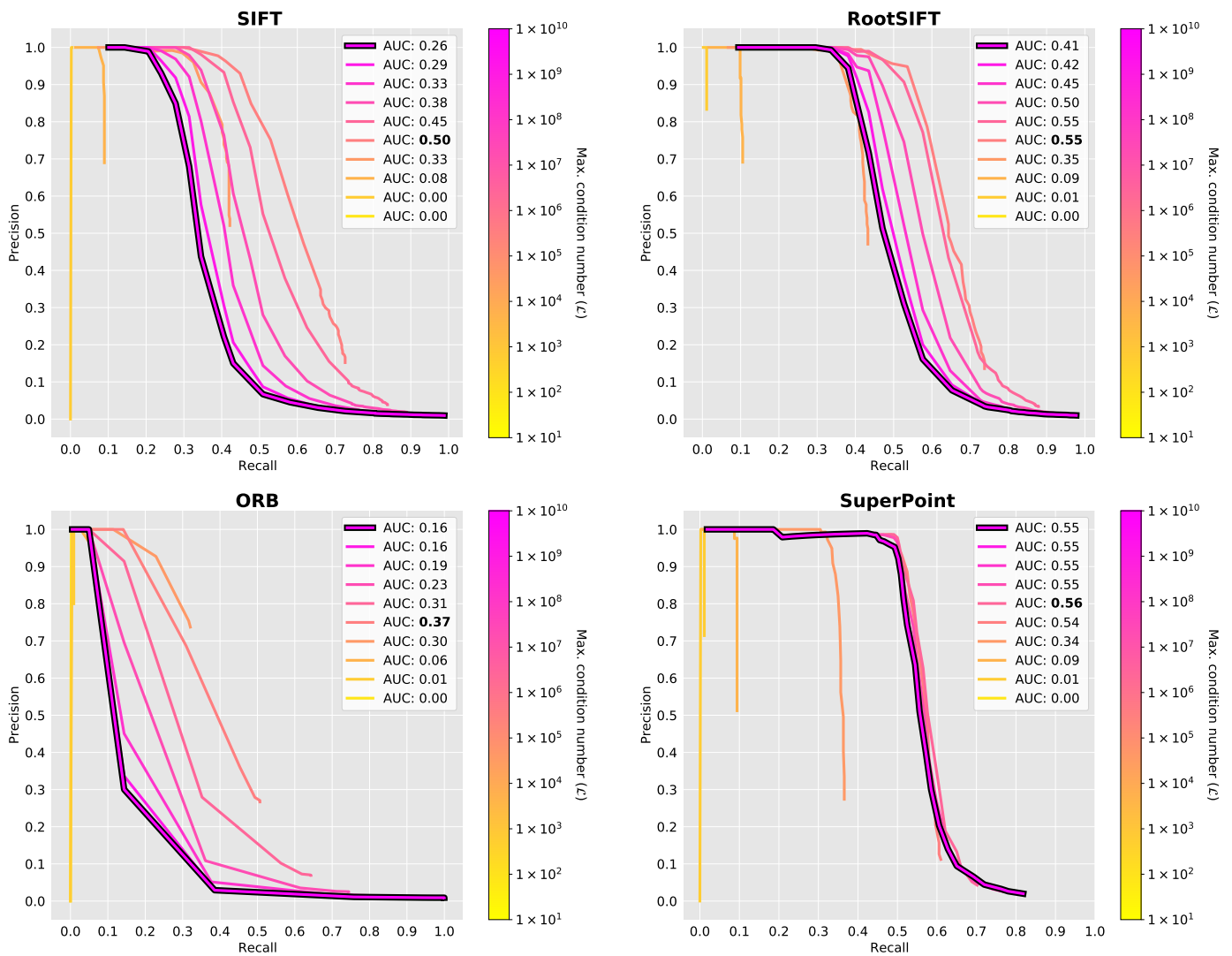
Figure 13. Precision–recall curves for each of the four keypoint descriptors. The curves are based on varying the minimum number of matching keypoints from 1 to 250, while the colors indicate the threshold of the condition number. The legends present the AUC values sorted by the condition number threshold parameter $\mathcal{L}$. Note that thresholding the condition number is nonessential for SuperPoint, as the performance is stable when $\mathcal{L} \geq 1 \times 10^6$.

7.3. Summary

We demonstrate a practical alternative to the typical supervised ranking-based re-identification model with the proposed pipeline equipped with SuperPoint and the novel binary classification module. The main drawback of this approach is the need for a (minimal) fine-tuning of the MKPs threshold parameter. However, the binary output has a range of benefits:

- It allows for an arbitrary number of gallery images that share an ID with the probe (without forcing the user to be concerned about choosing an optimal rank for the proposals);
- It effectively reduces the need for human verification;
- It allows for training supervised models on the automatically labeled data (essentially making them unsupervised);
- The binary and ranked output can be combined by removing the binary classified positive samples and only manually inspecting the top-n proposals of the remainder of the dataset.

Although thresholding the condition number did not result in increased performance for SuperPoint, the use of homography to transform images allows for a practical method of evaluating ranked proposals. Our examples shown in Figure 12 indicate that image pairs depicting different individuals display obscure projective transformations, making it easy to distinguish projected images of matching vs. non-matching individuals. Adding transformed images as an additional source to the original images may enhance the practicality of our pipeline for manual verification of the top-n ranked images, even though its effect cannot be quantified by traditional metrics, such as mAP, CMC, precision, or recall. Future research on the re-identification of giant sunfish, and re-identification in general, should involve real-life assessments that consider the entire human-in-the-loop setup to evaluate the efficacy of different strategies in terms of accuracy, practicality, and efficiency. In particular, the benefits of ranking vs. binary classification should be evaluated.

8. Conclusions

We propose a computer-vision-based pipeline for identifying individual giant sunfish using keypoint matching. We evaluate the pipeline equipped with each of the four keypoint descriptors: ORB, SIFT, RootSIFT, and SuperPoint. The pipeline achieved a mean average precision of 75.91% on the TinyMola+ dataset without any training or fine-tuning. Furthermore, we demonstrate that the pipeline generalizes well to other patterned species, such as seals and cattle, where its performance is comparable to state-of-the-art supervised methods concerning the CMC rank-1 score.

Lastly, we argue that a ranking-based output is not practical for real-life scenarios, as it is challenging for users to determine an optimal rank. Instead, we consider the re-identification task as a binary classification and introduce an alternative output module that identifies image pairs with at least a single pair of matching keypoints as positives. Initially, this approach resulted in a high number of false positives, making it impractical. However, by only accepting image pairs with at least 25 matching keypoints, we demonstrate that giant sunfish can be robustly identified with a precision of 98%, a recall of 44%, and an area under the precision–recall curve of 55%. This approach eliminates the need for human verification of almost half of the TinyMola+ dataset.

Further research is required to thoroughly investigate how automated computer-vision-based re-identification systems can be integrated into practical human-in-the-loop systems. A carefully considered balance between automated and human decision making is required to ensure that such systems are effective and efficient in real-life scenarios and not just on the drawing board.

Author Contributions: Formal analysis, project administration, methodology, software, investigation, visualization, M.P.; data curation, M.P. and M.N.; conceptualization, M.P., M.N. and T.B.M.; funding acquisition, T.B.M. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The images of the TinyMola+ dataset are part of the Match My Mola database which contains images mainly captured by tourist divers (volunteers). Unfortunately, at the time of writing, it has not been possible for the authors to obtain permissions from all volunteers to share the images. Therefore, data sharing is not currently applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gomes-Pereira, J.N.; Pham, C.K.; Miodonski, J.; Santos, M.A.R.; Dionísio, G.; Catarino, D.; Nyegaard, M.; Sawai, E.; Carreira, G.P.; Afonso, P. The heaviest bony fish in the world: A 2744-kg giant sunfish Mola alexandrini (Ranzani, 1839) from the North Atlantic. *J. Fish Biol.* **2022**, *102*, 290–293. [CrossRef] [PubMed]
2. Sawai, E.; Nyegaard, M. A review of giants: Examining the species identities of the world's heaviest extant bony fishes (ocean sunfishes, family Molidae). *J. Fish Biol.* **2022**, *100*, 1345–1364. [CrossRef] [PubMed]

3.   Nyegaard, M. There Be Giants! The Importance of Taxonomic Clarity of the Large Ocean Sunfishes (Genus Mola, Family Molidae) for Assessing Sunfish Vulnerability to Anthropogenic Pressures. Ph.D. Thesis, Murdoch University, Perth, Australia, 2018.

4.   Thys, T.; Ryan, J.P.; Weng, K.C.; Erdmann, M.; Tresnati, J. Tracking a marine ecotourism star: Movements of the short ocean sunfish Mola ramsayi in Nusa Penida, Bali, Indonesia. *J. Mar. Biol.* **2016**, *2016*, 1–6. [CrossRef]

5.   Ocean Sunfish Research Trust. Match My Mola. Available online: https://oceansunfishresearch.org/matchmymola/ (accessed on 21 September 2021).

6.   Pedersen, M.; Haurum, J.B.; Moeslund, T.B.; Nyegaard, M. Re-Identification of Giant Sunfish using Keypoint Matching. In Proceedings of the Northern Lights Deep Learning Workshop 2022, Virtual, 10–12 January 2022; Volume 3. [CrossRef]

7.   Schneider, S.; Taylor, G.W.; Linquist, S.; Kremer, S.C. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol. Evol.* **2019**, *10*, 461–470. [CrossRef]

8.   Hammond, P.S.; Mizroch, S.A.; Donovan, G.P. *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters: Incorporating the Proceedings of the Symposium and Workshop on Individual Recognition and the Estimation of Cetacean Population Parameters*; International Whaling Commission: Impington, UK, 1990.

9.   McConkey, S.D. Photographic identification of the New Zealand sea lion: A new technique. *N. Z. J. Mar. Freshw. Res.* **1999**, *33*, 63–66. [CrossRef]

10.   Würsig, B.; Jefferson, T.A. *Methods of Photo-Identification for Small Cetaceans*; Reports of the International Whaling Commission Special; International Whaling Commission: Impington, UK, 1990; Volume 12, pp. 43–52.

11.   Weinstein, B.G. A computer vision for animal ecology. *J. Anim. Ecol.* **2017**, *87*, 533–545. [CrossRef]

12.   Petrellis, N. Measurement of Fish Morphological Features through Image Processing and Deep Learning Techniques. *Appl. Sci.* **2021**, *11*, 4416. [CrossRef]

13.   Goodwin, M.; Halvorsen, K.T.; Jiao, L.; Knausgård, K.M.; Martin, A.H.; Moyano, M.; Oomen, R.A.; Rasmussen, J.H.; Sørdalen, T.K.; Thorbjørnsen, S.H. Unlocking the potential of deep learning for marine ecology: Overview, applications, and outlook. *ICES J. Mar. Sci.* **2022**, *79*, 319–336. [CrossRef]

14.   Arzoumanian, Z.; Holmberg, J.; Norman, B. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks Rhincodon typus. *J. Appl. Ecol.* **2005**, *42*, 999–1011. [CrossRef]

15.   Holmberg, J.; Norman, B.; Arzoumanian, Z. Estimating population size, structure, and residency time for whale sharks Rhincodon typus through collaborative photo-identification. *Endanger. Species Res.* **2009**, *7*, 39–53. [CrossRef]

16.   Van Tienhoven, A.; Den Hartog, J.; Reijns, R.; Peddemors, V. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark Carcharias taurus. *J. Appl. Ecol.* **2007**, *44*, 273–280. [CrossRef]

17.   Bruslund Haurum, J.; Karpova, A.; Pedersen, M.; Hein Bengtson, S.; Moeslund, T.B. Re-identification of zebrafish using metric learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1–11. [CrossRef]

18.   Olsen, Ø.L.; Sørdalen, T.K.; Goodwin, M.; Malde, K.; Knausgård, K.M.; Halvorsen, K.T. A contrastive learning approach for individual re-identification in a wild fish population. In Proceedings of the Northern Lights Deep Learning Workshop, Tromso, Norway, 10–12 January 2023; Volume 4. [CrossRef]

19.   Gómez-Vargas, N.; Alonso-Fernández, A.; Blanquero, R.; Antelo, L.T. Re-identification of fish individuals of undulate skate via deep learning within a few-shot context. *Ecol. Inform.* **2023**, *75*, 102036. [CrossRef]

20.   Chelak, I.; Nepovinnykh, E.; Eerola, T.; Kälviäinen, H.; Belykh, I. EDEN: Deep Feature Distribution Pooling for Saimaa Ringed Seals Pattern Matching. In *Cyber-Physical Systems and Control II*; Springer International Publishing: Cham, Switzerland, 2023; pp. 141–150. [CrossRef]

21.   Ferreira, A.C.; Silva, L.R.; Renna, F.; Brandl, H.B.; Renoult, J.P.; Farine, D.R.; Covas, R.; Doutrelant, C. Deep learning-based methods for individual recognition in small birds. *Methods Ecol. Evol.* **2020**, *11*, 1072–1085. [CrossRef]

22.   Crall, J.P.; Stewart, C.V.; Berger-Wolf, T.Y.; Rubenstein, D.I.; Sundaresan, S.R. Hotspotter—Patterned species instance recognition. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV) IEEE, Clearwater Beach, FL, USA, 15–17 January 2013; pp. 230–237. [CrossRef]

23.   Shukla, A.; Anderson, C.; Cheema, G.S.; Gao, P.; Onda, S.; Anshumaan, D.; Anand, S.; Farrell, R. A Hybrid Approach to Tiger Re-Identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) IEEE, Seoul, Korea, 27–28 October 2019. [CrossRef]

24.   Speed, C.W.; Meekan, M.G.; Bradshaw, C.J. Spot the match–wildlife photo-identification using information theory. *Front. Zool.* **2007**, *4*, 1–11.

25.   Anderson, C.J.R.; Lobo, N.D.V.; Roth, J.D.; Waterman, J.M. Computer-aided photo-identification system with an application to polar bears based on whisker spot patterns. *J. Mammal.* **2010**, *91*, 1350–1359. [CrossRef]

26.   Parham, J.; Crall, J.; Stewart, C.; Berger-Wolf, T.; Rubenstein, D. Animal population censusing at scale with citizen science and photographic identification. In Proceedings of the AAAI Spring Symposium SS-17-01, Palo Alto, CA, USA, 27–29 March 2017; Association for the Advancement of Artificial Intelligence: Palo Alto, CA, USA, 2017; pp. 37–44.

27.   Andrew, W.; Hannuna, S.; Campbell, N.; Burghardt, T. Automatic individual holstein friesian cattle identification via selective local coat pattern matching in RGB-D imagery. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP) IEEE, Phoenix, AZ, USA, 25–28 September 2016. [CrossRef]

28. Andrew, W.; Gao, J.; Mullan, S.; Campbell, N.; Dowsey, A.W.; Burghardt, T. Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Comput. Electron. Agric.* **2021**, *185*, 106133. [CrossRef]

29. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

30. Arandjelović, R.; Zisserman, A. Three things everyone should know to improve object retrieval. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2911–2918. [CrossRef]

31. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. In *Computer Vision, Proceedings of the ECCV 2006, Graz, Austria, 7–13 May 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417. [CrossRef]

32. Maglietta, R.; Renò, V.; Cipriano, G.; Fanizza, C.; Milella, A.; Stella, E.; Carlucci, R. DolFin: An innovative digital platform for studying Risso's dolphins in the Northern Ionian Sea (North-eastern Central Mediterranean). *Sci. Rep.* **2018**, *8*, 17185. [CrossRef]

33. Zhao, L.; Pedersen, M.; Hardeberg, J.Y.; Dervo, B. Image-Based Recognition of Individual Trouts in the Wild. In Proceedings of the 2019 8th European Workshop on Visual Information Processing (EUVIP) IEEE, Rome, Italy, 28–31 October 2019. [CrossRef]

34. Long, S.L.; Azmi, N.A. Using photographic identification to monitor sea turtle populations at Perhentian Islands Marine Park in Malaysia. *Herpetol. Conserv. Biol.* **2017**, *12*, 350–366.

35. Stoddard, M.C.; Kilner, R.M.; Town, C. Pattern recognition algorithm reveals how birds evolve individual egg pattern signatures. *Nat. Commun.* **2014**, *5*, 4117. [CrossRef]

36. Dunbar, S.G.; Anger, E.C.; Parham, J.R.; Kingen, C.; Wright, M.K.; Hayes, C.T.; Safi, S.; Holmberg, J.; Salinas, L.; Baumbach, D.S. HotSpotter: Using a computer-driven photo-id application to identify sea turtles. *J. Exp. Mar. Biol. Ecol.* **2021**, *535*, 151490. [CrossRef]

37. Bolger, D.T.; Morrison, T.A.; Vance, B.; Lee, D.; Farid, H. A computer-assisted system for photographic mark-recapture analysis. *Methods Ecol. Evol.* **2012**, *3*, 813–822. [CrossRef]

38. Moghimi, A.; Celik, T.; Mohammadzadeh, A.; Kusetogullari, H. Comparison of Keypoint Detectors and Descriptors for Relative Radiometric Normalization of Bitemporal Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4063–4073. [CrossRef]

39. Tareen, S.A.K.; Saleem, Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 3–4 March 2018; pp. 1–10. [CrossRef]

40. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

41. Bergler, C.; Gebhard, A.; Towers, J.R.; Butyrev, L.; Sutton, G.J.; Shaw, T.J.H.; Maier, A.; Nöth, E. FIN-PRINT a fully-automated multi-stage deep-learning-based framework for the individual recognition of killer whales. *Sci. Rep.* **2021**, *11*, 23480. [CrossRef]

42. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]

44. Wang, W.; Solovyev, R.; Stempkovsky, A.; Telpukhov, D.; Volkov, A. Method for Whale Re-identification Based on Siamese Nets and Adversarial Training. *Opt. Mem. Neural Netw.* **2020**, *29*, 118–132. [CrossRef]

45. Nepovinnykh, E.; Eerola, T.; Kalviainen, H. Siamese network based pelage pattern matching for ringed seal re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, Snowmass Village, CO, USA, 2–5 March 2020; pp. 25–34. [CrossRef]

46. Bouma, S.; Pawley, M.D.; Hupman, K.; Gilman, A. Individual common dolphin identification via metric embedding learning. In Proceedings of the 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ) IEEE, Auckland, New Zealand, 19–21 November 2018; pp. 1–6. [CrossRef]

47. Moskvyak, O.; Maire, F.; Dayoub, F.; Baktashmotlagh, M. Learning landmark guided embeddings for animal re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, Snowmass Village, CO, USA, 1–5 March 2020; pp. 12–19. [CrossRef]

48. Moskvyak, O.; Maire, F.; Dayoub, F.; Baktashmotlagh, M. Keypoint-Aligned Embeddings for Image Retrieval and Re-Identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–7 January 2021; pp. 676–685. [CrossRef]

49. Schneider, S.; Taylor, G.W.; Kremer, S.C. Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, Snowmass Village, CO, USA, 1–5 March 2020; pp. 44–52. [CrossRef]

50. Nyegaard, M.; Karmy, J.; McBride, L.; Thys, T.; Welly, M.; Djohani, R. Rapid physiological colouration change is a challenge-but not a hindrace-to successful photo identification of giant sunfish (*Mola alexandrini*, family Molidae). *Front. Mar. Sci.* **2023**, *10*. [CrossRef]

51. Kushimoto, T.; Kakino, A.; Shimomura, N. Possible individual identifications by the body surface marking patterns in the Ocean Sunfish *Mola mola* Sharptail Sunfish *Masturus lanceolatus* (Molidae). *Ichthy Nat. Hist. Fishes Jpn.* **2022**, *19*, 1–7. [CrossRef]

52. Pedersen, M.; Madsen, N.; Moeslund, T.B. No Machine Learning Without Data: Critical Factors to Consider when Collecting Video Data in Marine Environments. *J. Ocean Technol.* **2021**, *16*, 21–30.

53. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [CrossRef]

54. Watanabe, Y.; Sato, K. Functional Dorsoventral Symmetry in Relation to Lift-Based Swimming in the Ocean Sunfish Mola mola. *PLoS ONE* **2008**, *3*, e3446. [CrossRef]

55. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Corfu, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157. [CrossRef]

56. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, IEEE, Barcelona, Spain, 6–13 November 2011. [CrossRef]

57. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 224–236. [CrossRef]

58. Rosten, E.; Porter, R.; Drummond, T. Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef]

59. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary Robust Independent Elementary Features. In *Computer Vision, Proceedings of the ECCV 2010, Crete, Greece, 5–11 September 2010*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 778–792. [CrossRef]

60. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.

61. Brin, S. Near Neighbor Search in Large Metric Spaces. In Proceedings of the 21th International Conference on Very Large Data Bases (VLDB 1995), Zürich, Switzerland, 11–15 September 1995.

62. Moon, H.; Phillips, P.J. Computational and Performance Aspects of PCA-Based Face-Recognition Algorithms. *Perception* **2001**, *30*, 303–321. [CrossRef]

63. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) IEEE, Santiago, Chile, 7–13 December 2015. [CrossRef]

64. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2004. [CrossRef]