Aalborg Universitet

AALBORG UNIVERSITY

Re-Identification of Giant Sunfish using Keypoint Matching

Pedersen, Malte; Haurum, Joakim Bruslund; Moeslund, Thomas B.; Nyegaard, Marianne

Published in:
Proceedings of the Northern Lights Deep Learning Workshop

DOI (link to publication from Publisher):
10.7557/18.6234

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Pedersen, M., Haurum, J. B., Moeslund, T. B., & Nyegaard, M. (2022). Re-Identification of Giant Sunfish using Keypoint Matching. *Proceedings of the Northern Lights Deep Learning Workshop*, *3*. https://doi.org/10.7557/18.6234

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: 11 07, 2026

Re-Identification of Giant Sunfish using Keypoint Matching

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

[1]Visual Analysis and Perception Lab, Aalborg University, Denmark
[2]Ocean Sunfish Research Trust, Auckland, New Zealand

Abstract

We present the first work where re-identification of the Giant Sunfish (*Mola alexandrini*) is automated using computer vision and deep learning. We propose a pipeline that scores an mAP of 60.34% on a full rank of the novel TinyMola dataset which includes 41 IDs and 91 images. The method requires no domain-adaptation or training which makes it especially suited for low-budget or volunteer-based projects, like *Match My Mola*, as part of a human-in-the-loop model.

The pipeline includes segmentation, keypoint detection and description, keypoint matching, and ranking. The choice of feature descriptor has the largest impact on the performance and we show that the deep learning based SuperPoint descriptor greatly outperforms handcrafted descriptors like SIFT and RootSIFT independent of the segmentation level and matching method. Combining SuperPoint and the graph neural network based SuperGlue matching method produces the best results.

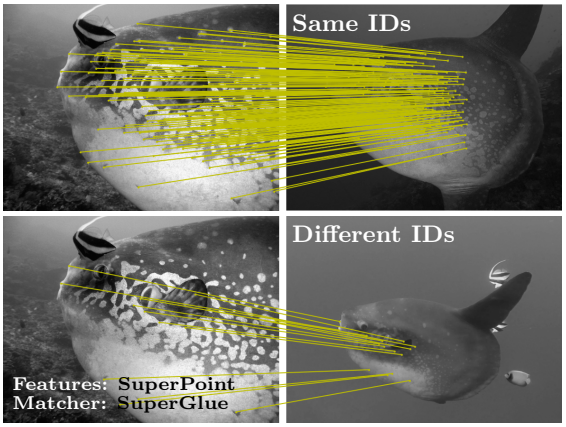

Figure 1: **Re-identification based on the number of keypoint matches**. The images in the top and bottom rows are of the same and different individuals, respectively. The yellow lines connect matched keypoints.

1 Introduction

The world's heaviest bony fish, the elusive 'Giant Sunfish' (*Mola alexandrini*), can reach an impressive weight of 2.3 ton [22]. Globally, they are rarely seen by SCUBA divers, but are frequent seasonal visitors to the Bali area, Indonesia [18]. Here, they seek cleaner fish interaction for removal of skin parasites, and are a highly popular target of the local SCUBA tourism industry [26]. 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 [18].

To investigate this, the citizen science and volunteer based project, Match My Mola [19], collects and curates sunfish images from the Bali area, taken by tourist divers, for photo identification purposes. Images are manually compared pair-wise to assess re-sightings of individuals over time, however, with increasing image numbers match time becomes a significant challenge to this volunteer-based project, and an automated system is critically needed.

Re-identification has been an active research problem within computer vision for decades. However, like in other fields the research has mainly been focused on humans [31] and only few have taken

*Corresponding Author: mape@create.aau.dk
This work has been funded by the Independent Research Fund Denmark under case number 9131-00128B.

https://doi.org/10.7557/18.6234
© The author(s). Licensee Septentrio Academic Publishing, Tromsø, Norway. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

a glance into the aquatic world [2, 24, 15, 5]. In this work we present the first scientific attempt to re-identify sunfish and show that it is possible based on the number of keypoint correspondences as illustrated in Figure 1.

Our contributions include: 1) a re-identification pipeline that requires no domain-adaptation or training, 2) an evaluation of how the segmentation level affects the performance of the system, 3) a comparison between the handcrafted feature descriptors SIFT and RootSIFT and the deep learning based SuperPoint feature descriptor.

2 Related Work

Photographic identification has been used for studying wild marine animals in a non-intrusive manner for decades [14, 29]. It allows researchers to identify the same individual across different years, but requires manual labour to obtain the photographs and match the individuals from the captured footage. Citizen science projects have proven to be an effective and irreplaceable method to gather large amounts of data. But as the databases grow, so does the need for manual labor. Therefore, computer vision has become an essential tool to scale such research.

Image processing and pattern matching techniques have been used to automatically identify individuals of whale-sharks [2], spotted raggedtooth sharks [27], and patterned terrestrial animals [6, 24]. However, recently Siamese networks and the use of triplet loss have become a popular means for handling re-identification problems within marine vision. Wang et al. proposed to use a Siamese network and adversarial training to identify whales by their flukes [28] and Nepovinnykh et al. trained a Siamese network for Ringed Seal re-identification [17]. Deep learning has generally gained a lot of attention during the last decade, which is also seen in the work done by Bouma et al. where they train a ResNet50 model using a triplet loss for identifying dolphins by their fin tail [3]. A ResNet50 was also used by Moskvyak et al. in their work on re-identification of manta rays [15, 16], 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 [23].

A common issue with the aforementioned methods is the requirement for training data and domain adaptation. However, it is demanding to capture images in wild underwater environments and marine image datasets are, therefore, often sparse. This leaves little to no room for the creation of high quality data splits.

3 TinyMola Image Dataset

The dataset used in this work is named 'TinyMola' and it is a subset of the much larger Match My Mola image database, which consists of more than a thousand photo events (PhE). PhEs are 1-3 images of the same individual captured by the same diver during the same dive and the images can be of one or both sides of the fish as illustrated in Figure 2. Manually identifying sunfish between PhEs is both hard and time-consuming and only 29 individuals have been matched and verified by the marine scientists at this point. These individuals form the basis of the TinyMola dataset as no ground truth is available for the remainder of the dataset.

The sunfish have unique markings which are used to identify the individuals. However, the markings on the fish are not identical on the two sides and they cannot be directly compared. Therefore, we frame the re-identification task to be side-specific and provide each side of the fish a unique ID. For each ID there are images from at least two PhEs. However, 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. These 'unpaired' images are named *singles* if they do not match with images from any other PhE. We have a total of 41 IDs shared among 14 left-side, 17 right-side, and 10 single IDs as summarized in Table 1.

Table 1: **TinyMola Dataset composition.**

	Left	Right	Singles	Total
IDs	14	17	10	41
Images	37	44	10	91

(a) Left side of the fish. (b) Right side of the fish.

Figure 2: **Two images from the same photo event**. The images are captured at different distances and angles, and shows both sides of the fish.

(a) Low contrast. (b) Extreme angle.

Figure 3: **Image variations.** Sometimes the objects are only partly visible, the contrast may be low, or the object is captured from an extreme angle.

The quality of the images vary extensively depending on the turbidity of the water, attenuation of light, occlusion, and camera settings [20]. Two examples illustrating some of the variations can be seen in Figure 3. The resolution of the images varies from 0.1 to 16 megapixels (MP) with a mean around 4 MP and an object resolution around 1 MP on average. To standardize the images we resize them to a resolution of 640x480 and convert them to gray-scale.

4 Method

As previously mentioned, re-identification of sunfish is currently conducted manually by marine researchers. The researchers crop the image around the target and look at markings across all overlapping body parts on the two images. If the markings are barely visible the image may be subject to contrast enhancement. The images are then compared pair-wise to images of other sunfish and matches are noted and examined by other matching-experts, to confirm that the images are of the same individual.

We propose a solution inspired by the manual process for ranking the images based on the number of matching keypoints. The pipeline is illustrated in Figure 4 and the modules are described below.

4.1 Region of Interest

We want to investigate whether cropping the image has an effect on the performance of the system. Therefore, we evaluate three levels of segmentation: 1) full image, 2) bounding box, and 3) instance segmentation. An ImageNet [8] pre-trained Mask R-CNN 50 FPN model [10] from Detectron2 [30] is used for segmenting the objects. The model is fine-tuned for 300 epochs with a batch size of 6 and a train split consisting of 100 random images of sunfish from the Match My Mola database, which are not part of the TinyMola dataset. The fine-tuned model achieves an average precision of 87.7% on the segmentation task and the bounding boxes are simply drawn around the segmentations.

4.2 Keypoints

The body of a sunfish is highly rigid, except for the dorsal and pelvic fins. Consequently, the markings on the fish are mainly affected by affine transformations such as rotation and scale. We test and evaluate the performance of two handcrafted feature descriptors (SIFT and RootSIFT) and one state of the art deep learning based feature descriptor (SuperPoint), which are all summarized here.

The Scale Invariant Feature Transform (SIFT) keypoint descriptor was proposed by Lowe [12, 13] and has been among the most popular keypoint descriptors for two decades. Interest points are located in the image by creating a scale-space using difference-of-Gaussians and finding consistent extrema points. A histogram of oriented gradients (HoG) with 36 bins is created from a region around the point and used to assign an orientation to the keypoint. The SIFT descriptor itself is based on a 4x4 matrix of normalized HoG features with 8 bins resulting in a feature vector with 128 values.

Originally, Lowe proposed to match SIFT features by Euclidean distance, however, as noted by Arandjelovic and Zisserman [1] the Hellinger kernel has often been used to compare histograms as it commonly yields superior results compared to Euclidean distance. As the SIFT descriptor is based

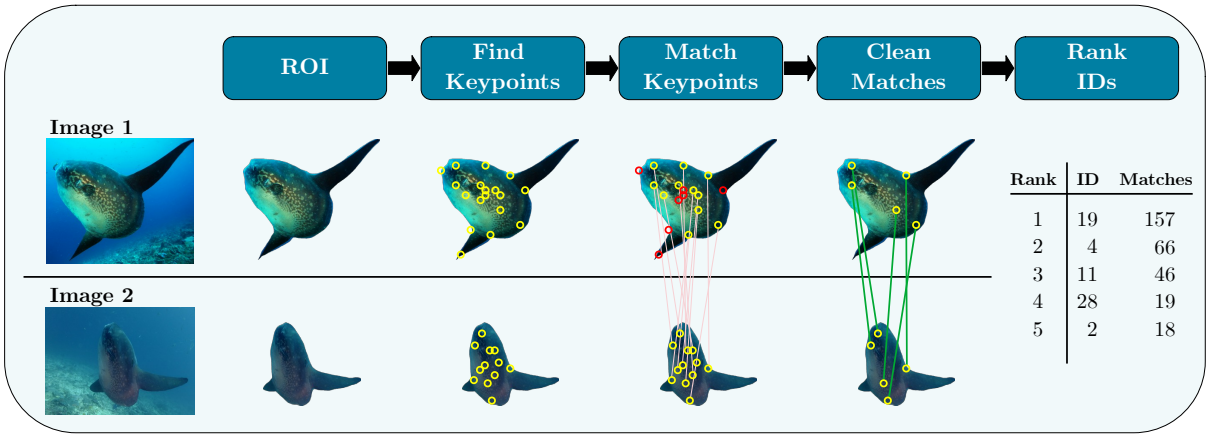

Figure 4: **The proposed pipeline**. In the first module the ROI is either the full image, bounding box, or instance segmentation. Next, keypoints are located (illustrated as yellow circles). During matching some keypoints are not paired (represented by red circles). Ambiguous or weak keypoints are dismissed during cleaning and lastly the images are ranked based on the number of keypoint matches.

on histograms they proposed an enhanced method named RootSIFT, which consists of two additional steps: 1) L1-normalize the SIFT feature vector and 2) square root each element. Consequently, comparing RootSIFT features using Euclidean distance, is equivalent to compare SIFT features using the Hellinger kernel, which often increases performance.

Recently, deep learning has been used to both detect and describe keypoints. SuperPoint [9] is among the state of the art methods that handles both tasks jointly. SuperPoint is a CNN that has been trained on synthetic data of angular shapes, such as triangles, lines, and cubes. Subsequently, the model is finetuned on images from MS-COCO [11] in a self-supervised manner using homographic adaptation, which is the use of random homographies to learn image-to-image transformations that may appear in real world scenarios. The SuperPoint feature vector has a dimensionality of 256.

We evaluate the performance of all three descriptors with default parameter values and a maximum of 1024 keypoints per image. All the keypoint descriptors are calculated on the full image, but is only part of the matching process if they are located within the ROI.

4.3 Keypoint Matching

Finding corresponding keypoints in two images is a matter of determining which pair of features that are most similar (nearest neighbor) determined by a distance function. In the following we will very briefly describe two traditional methods (brute-force and kd-trees) and a graph neural network (SuperGlue) for finding the nearest neighbor.

Depending on the problem, and the dimensionality and nature of the data, keypoint matching has commonly been done using brute-force methods or kd-trees [4]. 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 the TinyMola dataset is small and the task is an offline problem we use brute-force to match the keypoints to get optimal results.

Recently, deep learning has made its entry into keypoint matching and SuperGlue [21], proposed by the team behind SuperPoint, is currently one of the state of the art methods. For each keypoint SuperGlue takes the position and feature descriptor as input and encodes it using a multilayer perceptron. The spatial and visually encoded feature vectors are fed into a graph neural network that utilizes self- and cross-attention to compute matching descriptors. A similarity matrix is computed with added "dustbin" columns and rows to handle non-matched keypoints. Lastly, the Sinkhorn algorithm is used to compute the optimal partial assignment.

The SuperGlue algorithm is designed to be used with SuperPoint which has twice as many elements as SIFT and RootSIFT. Therefore, to make a fair comparison we use brute-force to match the SIFT, RootSIFT, and SuperPoint descriptors. Additionally, we also match SuperPoint features using two pre-trained SuperGlue models: *SGIndoor*, that has been trained on indoor images from ScanNet [7] and *SGOutdoor* that has been trained on a subset of outdoor images from YFCC100M [25].

4.4 Clean Matches

Naively matching the closest keypoints can lead to poor results. For this reason, David G. Lowe introduced the distance ratio test [13] 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.

When using the brute-force method to match the keypoints we clean the matches using the distance ratio test with a threshold of 0.8 as proposed by Lowe [13]. We do not clean the matches proposed by SuperGlue as it dismisses weak and ambiguous candidates through the dustbin and the Sinkhorn assignment scheme.

5 Evaluation

We evaluate the performance of the proposed methods by their mean average precision (mAP) per rank. We view every image of the dataset (except the *single* images) as *probes* and compare each probe against all the other images, which we call the *gallery* images. The *single* images are included in the set of gallery images. There is always at least one gallery image with the same ID as the probe and we name these the *hit* images. We calculate the average precision as

$$\mathrm{AP} = \frac{1}{H}\sum_{n=1}^{k} P_n R_n, \tag{1}$$

where H is the number of hit images, k is the rank, P is the precision, and R is a relevance function. The precision is given by

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{2}$$

where TP is the number of true positives and FP is the number of false positives. The relevance function, R, takes a value of 1 or 0 depending on whether the match is a hit or not. The rank decides the number of matches to take into account and the matches are sorted in a decreasing manner based on the number of keypoint correspondences between the probe and gallery image. The number of hit images is bounded by the rank such that we have $H \leq k$. An example of calculating the AP is given below where $H = 3$ and $k = 5$. The filled and empty circles represent hits and misses, respectively.

Rank	1	2	3	4	5	
Match	●	●	○	●	○	
AP	$\frac{1}{3}(\frac{1}{1}$	$+\ \frac{2}{2}$	$+\ 0$	$+\ \frac{3}{4}$	$+\ 0)$	$= 0.92$

Lastly, the mean AP is given by

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \tag{3}$$

where N is the number of probes.

6 Results

The performance of the system is measured by the mAP which is presented against the rank in Figure 5. There are several interesting aspects that can be seen from the results. One thing to notice is the significant difference in performance between the feature descriptors. For both of the handcrafted descriptors (SIFT and RootSIFT), the mAP at rank 1 is close to zero indicating that the ranking is more or less based on coincidence. On the other hand, the deep learning based SuperPoint descriptor shows promising results both when using brute-force matching and SuperGlue. Decreasing the region of interest seems to have an ambiguous effect; for RootSIFT, SIFT, and SP-SGOutdoor it generally worsen the performance, but for SP-BF and SP-SGIndoor it increases the performance.

Both SIFT and RootSIFT have few hits on the first rank, but the precision increases up to rank 10 and stabilizes. On the other hand, we see that the

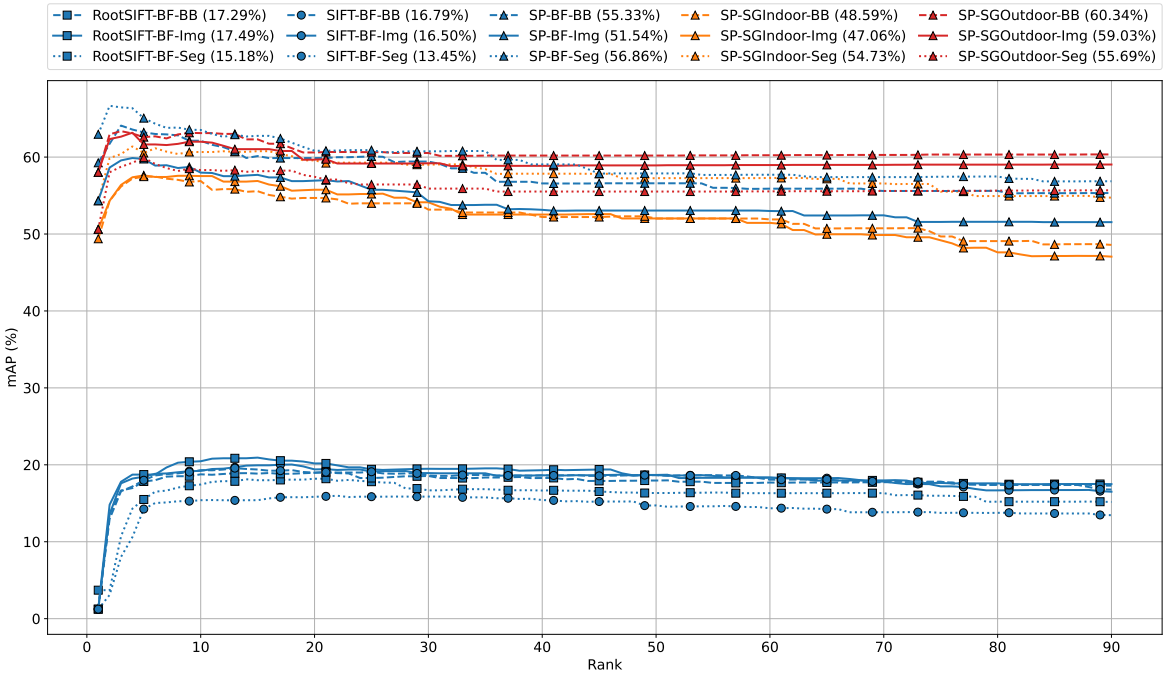

Figure 5: **Evaluation results.** The legends specify *descriptor-matching-segmentation*, e.g., SP-SGIndoor-BB is a combination of the SuperPoint descriptor, SuperGlue indoor matching, and bounding box segmentation. The handcrafted feature descriptors (SIFT and RootSIFT) show weak performance compared to the deep learning based SuperPoint descriptor. The difference between using brute-force matching and the graph-based SuperGlue is less profound and the segmentation level seems to affect the performance ambiguously. The mAP presented in the legends are from the last rank.

SuperPoint based methods all perform well already on rank 1 and their performance increases slightly before dropping and stabilizing, which indicates at least a single hit among the first few ranks.

The SP-BF methods generally perform better than the graph-based SuperGlue model trained on indoor images and the SP-BF-Seg (56.86%) even beats the SP-SGOutdoor-Seg method (55.69%). This indicates that the training data has an effect on the performance of the SuperGlue algorithm. The performance difference between the SGIndoor and SGOutdoor may be due to the outdoor images resembling the underwater domain to a larger degree than indoor images. However, there is most likely a gap between the terrestrial and underwater domains and we suspect that better performance can be achieved by training a SuperGlue model on underwater images. Even so, the SP-SGOutdoor-BB display the strongest performance with an mAP of 60.34% at full rank.

The results indicate that our solution can significantly reduce the search space for the volunteers who are currently manually matching the images in the Match My Mola project. Instead of comparing with every image in the database, the volunteers may only need to look at the top ranked suggestions to find potential strong matches.

7 Conclusion

We propose a pipeline for re-identification of Giant Sunfish (*Mola alexandrini*) that requires no domain adaptation or training. The pipeline is based on publicly available methods for keypoint detection, description, and matching. The evaluation is based on the novel Tiny Mola Dataset, which consists of underwater images of the Giant Sunfish captured in diverse environments.

We found that the largest impact on performance

was based on the choice of descriptor, while the level of segmentation had a low and ambiguous effect. The deep learning based SuperPoint descriptor outperforms the handcrafted keypoint descriptors SIFT and RootSIFT. Good results were obtained with both brute-force matching and the graph neural network based SuperGlue matching. The best performance was achieved using a combination of SuperPoint and SuperGlue with a score of 60.34% mAP on full rank.

None of the methods in the proposed pipeline have been trained or adapted to underwater environments or fish in general. Therefore, the results indicate that the pipeline may also be applicable out-of-the-box in other domains (both terrestrial and underwater). The solution seems especially suited for low-budget or volunteer-based wildlife conservation projects without sufficient data for training supervised machine learning algorithms.

References

[1] R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2911–2918. IEEE, 2012. doi:10.1109/CVPR.2012.6248018.

[2] Z. Arzoumanian, J. Holmberg, and B. Norman. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks rhincodon typus. *Journal of Applied Ecology*, 42(6):999–1011, 2005. doi:10.1111/j.1365-2664.2005.01117.x.

[3] S. Bouma, M. D. Pawley, K. Hupman, and A. Gilman. Individual common dolphin identification via metric embedding learning. In *2018 international conference on image and vision computing New Zealand (IVCNZ)*, pages 1–6. IEEE, 2018. doi:10.1109/IVCNZ.2018.8634778.

[4] S. Brin. Near neighbor search in large metric spaces. In *21th International Conference on Very Large Data Bases (VLDB 1995)*, 1995. URL `http://ilpubs.stanford.edu:8090/113/`.

[5] J. Bruslund Haurum, A. Karpova, M. Pedersen, S. Hein Bengtson, and T. B. Moeslund. Re-identification of zebrafish using metric learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 1–11, 2020. doi:10.1109/WACVW50321.2020.9096922.

[6] J. P. Crall, C. V. Stewart, T. Y. Berger-Wolf, D. I. Rubenstein, and S. R. Sundaresan. Hotspotter—patterned species instance recognition. In *2013 IEEE workshop on applications of computer vision (WACV)*, pages 230–237. IEEE, 2013. doi:10.1109/WACV.2013.6475023.

[7] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. doi:10.1109/CVPR.2017.261.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. doi:10.1109/CVPR.2009.5206848.

[9] D. DeTone, T. Malisiewicz, and A. Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. doi:10.1109/CVPRW.2018.00060.

[10] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. doi:10.1109/TPAMI.2018.2844175.

[11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. doi:10.1007/978-3-319-10602-1_48.

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999. doi:10.1109/ICCV.1999.790410.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. doi:10.1023/B:VISI.0000029664.99615.94.

[14] S. D. McConkey. Photographic identification of the new zealand sea lion: a new technique. *New Zealand Journal of Marine and Freshwater Research*, 1999. doi:10.1080/00288330.1999.9516857.

[15] O. Moskvyak, F. Maire, F. Dayoub, and M. Baktashmotlagh. Learning landmark guided embeddings for animal re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 12–19, 2020. doi:10.1109/WACVW50321.2020.9096932.

[16] O. Moskvyak, F. Maire, F. Dayoub, and M. Baktashmotlagh. Keypoint-aligned embeddings for image retrieval and re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 676–685, January 2021. doi:10.1109/WACV48630.2021.00072.

[17] E. Nepovinnykh, T. Eerola, and H. Kalviainen. 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*, pages 25–34, 2020. doi:10.1109/WACVW50321.2020.9096935.

[18] M. Nyegaard. *There be giants! The importance of taxonomic clarity of the large ocean sunfishes (genus Mola, Family Molidae) for assessing sunfish vulnerability to anthropogenic pressures.* PhD thesis, Murdoch University, 2018. URL `http://researchrepository.murdoch.edu.au/id/eprint/41666`.

[19] Ocean Sunfish Research Trust. Match my mola. `https://oceansunfishresearch.org/matchmymola/`. Accessed: 2021-09-21.

[20] M. Pedersen, N. Madsen, and T. B. Moeslund. No machine learning without data: Critical factors to consider when collecting video data in marine environments. *The Journal of Ocean Technology*, 16(3), 2021. URL `https://www.thejot.net/archive-issues/?id=73`.

[21] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020. doi:10.1109/CVPR42600.2020.00499.

[22] E. Sawai, Y. Yamanoue, M. Nyegaard, and Y. Sakai. Redescription of the bump-head sunfish mola alexandrini (ranzani 1839), senior synonym of mola ramsayi (giglioli 1883), with designation of a neotype for mola mola (linnaeus 1758)(tetraodontiformes: Molidae). *Ichthyological Research*, 65(1):142–160, 2018. doi:10.1007/s10228-017-0603-6.

[23] S. Schneider, G. W. Taylor, and S. C. Kremer. 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*, pages 44–52, 2020. doi:10.1109/WACVW50321.2020.9096925.

[24] C. W. Speed, M. G. Meekan, and C. J. Bradshaw. Spot the match–wildlife photo-identification using information theory. *Frontiers in zoology*, 4(1):1–11, 2007. doi:10.1186/1742-9994-4-2.

[25] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. Yfcc100m: The new data in multimedia research. *Communications of the ACM*, 59 (2):64–73, 2016. doi:10.1145/2812802.

[26] T. Thys, J. P. Ryan, K. C. Weng, M. Erdmann, and J. Tresnati. Tracking a marine ecotourism star: movements of the short ocean sunfish mola ramsayi in nusa penida, bali, indonesia. *Journal of Marine Biology*, 2016. doi:10.1155/2016/8750193.

[27] A. Van Tienhoven, J. Den Hartog, R. Reijns, and V. Peddemors. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark carcharias taurus. *Journal of Applied ecology*, 44(2):273–280, 2007. doi:10.1111/j.1365-2664.2006.01273.x.

[28] W. Wang, R. Solovyev, A. Stempkovsky, D. Telpukhov, and A. Volkov. Method for whale re-identification based on siamese nets and adversarial training. *Optical Memory and Neural Networks*, 29(2):118–132, 2020. doi:10.3103/S1060992X20020058.

[29] B. Wiirsig and T. A. Jefferson. Methods of photo-identification for small cetaceans. *Reports of the International Whaling Commission. Special*, 12:42–43, 1990. URL `https://www.vliz.be/imisdocs/publications/253951.pdf`.

[30] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.

[31] L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016. URL `https://arxiv.org/abs/1610.02984`.