2020 IMEKO TC-19 International Workshop on
Metrology for the Sea
Naples, Italy, October 5-7, 2020

NNPool in SPIR pipeline for Risso's dolphins identification

R. Maglietta[1], V. Renò[1], R. Caccioppoli[2], S. Bellomo[3], FC. Santacesaria[3], G. Cipriano[4], E. Stella[1], K. Hartman[6], C. Fanizza[3], G. Dimauro[2], R. Carlucci[4]

[1]*Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, via Amendola 122, Bari, Italy*
[2]*Department of Computer Science, University of Bari, Via Orabona 4 - 70125 Bari, Italy*
[3]*Jonian Dolphin Conservation, Viale Virgilio 102 - 74121 Taranto, Italy*
[4]*Department of Biology, University of Bari, Via Orabona 4 - 70125 Bari, Italy*
[5]*CoNISMa, Piazzale Flaminio 9 - 00196 Rome, Italy*
[6]*Nova Atlantis Foundation, Risso's Dolphin Research Centre, Rua Dr. A. F. Pimentel 11, 9930-309, Santa Cruz das Ribeiras, Pico, Azores, Portugal*

***Abstract* – Photo-identification is designed to recognize single individuals of a species by exploiting unique and distinctive marks identifiable in one or multiple photographs. The Risso's dolphin's distinctive marks and scars on the dorsal fin proved to be very useful to photo-identify individuals. The aim of this paper is to improve the performances of the photo-identification studies on this species through machine learning techniques. This paper proposes the introduction of *NNpool*, based on Convolutional Neural Networks, in the SPIR pipeline to identify Risso's dolphins in a new photograph.**

I. INTRODUCTION

The Risso's dolphin *Grampus griseus* is one of the least-known cetacean species on a global scale, with Mediterranean subpopulation ranked as Data Deficient by the IUCN Red List [1]. To bridge the gap of understanding this species, a key component is executed through photo-identification (photo-ID) studies, based on the recognition of single individuals thanks to the specific markers on their dorsal fins obtained in photographs. It is their appearance that makes the Risso's dolphins particularly suitable for this kind of research. Generally, adult Risso's dolphins present extensive white scarring on their bodies, (Fig. 1) presumably caused by infraspecific interactions [2], can appear as scratches, stains, or circular marks, and in some animals can cover most of their body surface. As a result, the unique markings on their dorsal fin can be successfully analyzed to identify single individuals.

The state of the art for the photo-identification of Risso's dolphins is the algorithm SPIR (Smart Photo-identification of Risso's dolphin) [3], [4], where Scale Invariant Feature Transform (SIFT) [5] is applied to detect the key-points over the scars on the dorsal fin of Risso's dolphins. The purpose of these photo-ID tool is to match the most similar dolphins, in terms of SIFT features, in a catalogue of *known* and labelled Risso's dolphins. The characteristic of SPIR is that it is a best-matching algorithm, assigning the most similar dolphin in the catalogue to the query image (i.e. the best-matching dolphin) and taking into account the reliability of its identification. In fact, SPIR predicts the identity of the dolphin in the query image as that of the dolphin in the catalogue with the highest number of equally-oriented matched SIFT features with the query image. If this number is less than 4, or if two or more dolphins in the catalogue have the same highest number the system will return a warning to the user. In the pipeline, this step reminds pay attention on the case of *unknown* dolphins, i.e. those individuals that are not part of the reference catalogue, but it does not solve the problem. If we were to have photographs of new fins, for example acquired during a new survey, in order to bring about a more powerful photo-identification, the *unknown* class should be considered among the possible identities for the query dolphin. Machine learning algorithms provide us with strategies able to classify examples never seen before, taking into consideration all desirable classes, and in particular the *unknown* one [6].

Fig. 1. Risso's dolphin Grampus griseus.

The main novelty of this paper is the application of NNpool strategy, recently proposed in [7], in the SPIR pipeline, to manage the *unknown* dolphins class in the Risso's dolphin photo-identification. In particular NNPool consists in a pool of multiple Convolutional Neural Networks, simultaneously queried to photo-identify single individuals, whose outputs are opportunely merged. The DolFin [1] catalogue [3], [7] which collects Risso's dolphins data and photos acquired in the period 2013-2018 in the Northern Ionian Sea (Central-eastern Mediterranean Sea), was used to carry out experiments. To validate experimental results a data set, collecting Risso's dolphins images from Azores, was used. Results show that the combination of this two algorithms improves the performance of the automated Risso's dolphin photo-identification and it is suitable for large data set studies.

II. MATERIALS AND METHODS

A. Survey area and data collection

Sighting data of *G. griseus* were collected from July 2013 to August 2018 during vessel-based surveys conducted on board of a 12-m catamaran, investigating an area of about 960 km^2, in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean Sea). Date, daytime, sea weather conditions, geographic coordinates, group size (number of specimens), and depth (m) were recorded. In addition, a collection of images for photo-ID were taken using a Nikon D3300 camera with Nikon AF-P Nikkor 70-300 mm, f4,5-6,3G ED lens. 93 different Risso's dolphins were photo-identified using the SPIR algorithm [3], [4] and their photographs are freely accessible using the DolFin 1 platform. The NNPool training set, D_{NN}, consists of 582 images of 24 different dolphins among the 93 collected in DolFin. Since the left and right side of the dolphin's fin are perceived and analyzed independently, as if they belong to separate individuals, there are 28 different models.

To validate the capability of NNPool to recognize and discard *unknown* dolphins and compair it to SPIR on its own, a new data set, D_v, was built. It contains 500 images of Risso's dolphin fins obtained from Azores, specifically off Pico island covering approximately 540 km^2 during May-September 2019. Risso's dolphins were first located from a land based look out (38.4078 N and 28.1880 W) using 25x80 binoculars (Steiner observer) [8], and encountered during ocean based surveys, using a 5.8 meters long zodiac, equipped with a 50 HP outboard engine. This data are unseen and completely new for both, NNPool and SPIR.

B. NNPool

NNPool is a methodology, based on Deep Learning [9], a type of machine learning (such as [10]–[13]) in which a model learns how to perform classification tasks directly from images. Through the various levels, then, manages to interpret concepts with a high level of complexity simply by composing concepts of lower complexity, easier to learn. In an image, for example, the first level of a neural network learns the presence or absence of edges and the orientation of the image. The second layer typically identifies the arrangement of edges, regardless of small variations in them. Finally, the third layer can assemble recurring structures into larger combinations that correspond to parts of familiar objects, and subsequent layers will detect objects as combinations of these parts.

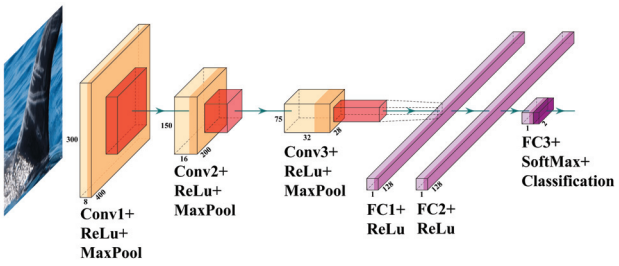

Fig. 2. A graphical representation of the NN build for for each dolphin.

In detail, NNPool uses one of the most popular algorithms for Deep Learning, that is the Convolutional Neural Network (CNN) [14]. A CNN can have dozens or hundreds of layers, each able to learn to detect different features of an image. Like other neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between. These layers perform operations that alter the data with the intent of learning features specific to the data. The most common layers are:

- Convolutional (Conv) which puts the input images through a set of convolutional filters, each activating a certain feature from the images.
- Rectified linear unit (ReLu) which allows faster and more effective training by mapping negative values to zero and maintaining positive values, thus adding non-linearity.
- Pooling (MaxPool) which simplifies the output by performing nonlinear downsampling, reducing the number of parameters that the network needs to learn.
- Fully Connected (FC) which multiplies the input by a weight tensor and adds a bias vector. All neurons are connected together and typically these kind of layers are used to actually classify the extracted features from previous layers.
- SoftMax which allows to transform a set of values into probabilities associated to the classes, [15].

[1] http://dolfin.ba.issia.cnr.it/

- Classification which computes the cross entropy loss for multi-class classification problems with mutually exclusive classes.

By mixing these types of layers, we obtain the neural network architecture proposed in this paper. In order to recognize the fin image, we use the following combination of layers for three times: Conv-ReLu-MaxPool, each time doubling the learned filters number on the convolutional layers from 8 to 16 to 32 (Fig. 2).

The image shows an architecture divided into four blocks, in detail:

- Frist block: a convolution layer that preserves the input size, with eight 3x3 size filters, followed by a ReLu type layer used to insert non-linearity. As a last step there is a 3x3 size max pooling layer and stride 2, to reduce the output size of this block.

- Second block: a convolution layer that preserves the input size, with sixteen 3x3 size filters, followed by a ReLu type layer, followed by a 3x3 size max pooling layer and stride 2, to reduce the output size of this block.

- Third block: a convolution layer that preserves the input size, with thirty-two 3x3 size filters, followed by a ReLu type layer, followed by a 3x3 size max pooling layer and stride 2, to reduce the output size of this block.

- Fourth block: a pair of fully connected layers, each with ReLu activation type, with output size 128, and finally another fully connected layer with output size 2, followed this time by a classification layer with softmax activation function.

A CNN model is then trained for each of the 28 models in the D_{NN} data set with a one-vs-all technique and a Crossing Validation strategy with n_{CV} cycles, empirically set to 10.

Class imbalance is managed with a downsampling strategy. Given a dolphin d_i among the 28 dolphins in D_{NN}, let n_i be the number of images available for d_i. Then, the *unknown* class is composed of m_i images, where $m_i = 27 \times \kappa^*$ with $\kappa^* = min\{\kappa \in \mathbb{N} \,|\, 27 \times \kappa \geq n_i\}$, i.e. the first multiple of 27 greater than n_i. This way, κ images for each of the remaining 27 individuals are taken into account. The training set is composed of $(n_i + m_i)$ photos. Subsequently, an oversampling technique based on image augmentation is used to increase the number of images in the training set. Two geometric transformations have been applied to every image: random rotation of ± 45 degrees (over both y and x axes) and translation of ± 20 pixels (over both axes, too).

i	CNN_i	p_i	CNN_i	p_i
1	FRANGETTA_R	48%	TI_L	90%
2	PREZZEMOLO_L	43%	BLACK_L	50%
3	PINNA_L	41%	PINNA_L	33%
4	TI_L	40%	ZANTE_L	30%
5	BLACK_L	36%	DALMATA_L	27%
6	ZANTE_L	32%	DUBBIO_L	25%
7	CARL_R	30%	CUPIDO_L	12%
8	DELTA_R	28%	PERONI_R	12%
9	DUBBIO_L	24%	TRIS_L	10%
10	ALT_R	20%	FRANGETTA_R	8%
11	PERONI_R	19%	PREZZEMOLO_L	7%
12	CUPIDO_L	18%	PREZZEMOLO_R	6%
13	TRIS_L	10%	SVIRGOLO_L	5%
14	ELE_R	10%	CARL_R	3%
15	PREZZEMOLO_R	6%	DELTA_R	3%
16	SVIRGOLO_L	5%	SVIRGOLO_R	1%
17	DALMATA_L	2%	VITO_R	1%
18	ZANTE_R	2%	ERARD_R	0,8%
19	VITO_R	1%	EMME_R	0,5%
20	JHONATAN_L	1%	ALT_R	0,2%
21	SMILE_R	1%	CUPIDO_R	0,2%
22	ERARD_R	0,5%	JHONATAN_L	0,2%
23	MENO_R	0,1%	MENO_R	0,1%
24	SVIRGOLO_R	0%	ELE_R	0%
25	EMME_R	0%	SMILE_R	0%
26	HUGO_L	0%	JAX_L	0%
27	JAX_L	0%	HUGO_L	0%
28	CUPIDO_R	0%	ZANTE_R	0%
	MODEL WITH p_i>51%	0	MODEL WITH p_i>51%	1
	RESULT LABEL	***UNKNOWN***	RESULT LABEL	***KNOWN***

Fig. 3. Example of NNPool output. The NNPool output is a vector P containing p_i values with i = 1, 2, ...,28, where p_i is the probability of the fin in the photo to belong to the dolphin reported to the CNN_i column. On the left (right) a result of the classification of unknown (known) dolphin.

So, NNPool consists of the mixing of the CNN_i networks, with i=$1, 2, ..., 28$, where CNN_i is made of n_{CV} trained models. NNPool is composed by the $(28 \times n_{CV})$ models.

Every CNN_i was built with the following parameters, empirically set:

- Solver Name: *stochastic gradient descent with momentum*, with momentum set to 0.9;

- Inital Learning Rate: set to 0.00001;

- Mini Batch Size: $\frac{1}{4}$ of the Training Set size;

- Max Epochs: 60, shuffling the data at every epoch.

Basically, the pipeline input is a cropped images and, to that regard, an algorithm, devoted to the fin cropping of dolphins photos, has been recently presented in literature [16]. So, every time we want to predict the label of a new photo, it will be first automatically cropped and resized to the dimension required by the input layer of all the CNNs, which is 300x400 pixels. Successively, the photo can be used as NNPool input, giving a P vector as output (as shown in figure 3). If there is only one $p_i \in P > 51\%$, the new photo will be labelled as *known*, otherwise it will be labelled as *unknown*.

III. EXPERIMENTS AND RESULTS

All data are analyzed using Matlab (MathWorks, Natick, MA). The performance of CNN photo-identification, illustrated in [7], was good, with almost all accuracies values above 80%. Moreover it has been proved proved that CNN performance is influenced by the training set size and the images quality. In fact when few images are available for the specimen, CNN sensitivity decreases with images quality, and when the number of images for the specimen increases, good sensitivity values are achieved even if fair or poor quality images are used. In general, CNNs performance is very good both in case of many low quality images as well as in case of few high quality images. Hence, quality of images surely impacts on the algorithm performances and therefore should be taken into account. Several methods to evaluate image quality have been discussed in literature [17], [18]. For NNPool the Perception based Image Quality Evaluator (PIQE scores)[19] was used to evaluate the image quality and was computed for all the images used to train each M_i classifiers. These are no-reference image quality scores, with values in the range $[0, 100]$, inversely correlated to the perceptual quality of an image. A low PIQE score value indicates high perceptual quality and high score value indicates low perceptual quality.

Here we present results of the automated photo-identification of Risso's dolphins captured in photos collected in the datadset D_v, using the combination of NNpool and SPIR. Figure 4 shows the number of False Positive (FP) and True Negative (TN) of NNPool+SPIR and SPIR photo-identification of unknown dolphins. FP represents *unknown* specimens, errouneusly associated to *known* models, TN represents *unknown* specimens, correctly discarded. Among the 500 examples, NNPool+SPIR correctly identifies 392 photos as belonging to *unknown* individuals (i.e TN) and it wrongly classifies 108 examples as *known* (i.e. FP). On the other hand, SPIR correctly identifies 238 images as *unknown* individuals (i.e TN) and wrongly associates to *known* models 262 photo (i.e. FP). Therefore, using NNPool + SPIR the number of False Positives recognized in D_v decreases compared to SPIR, which instead shows comparable FP and TN values. Finally, Table 1 shows a figure of merit ϵ that reflects the ability of recognizing unknown dolphins of the two methodologies on D_v dataset. The ϵ has been calculated as percentage of error made in identification by algorithms, i.e. the percentage of false positives on the total number of photos analyzed. The formula is:

$$\epsilon = \frac{FalsePositive(FP)}{TotalImages} \quad (1)$$

Results highlighted that the combination of NNPool within SPIR pipeline is essential as it can reduce the photo-identification error of unknown individulas.

Table 1. The error, ϵ, in SPIR pipeline with and without NNPool.

	SPIR	NNPool+SPIR
ϵ	0.52	0.22

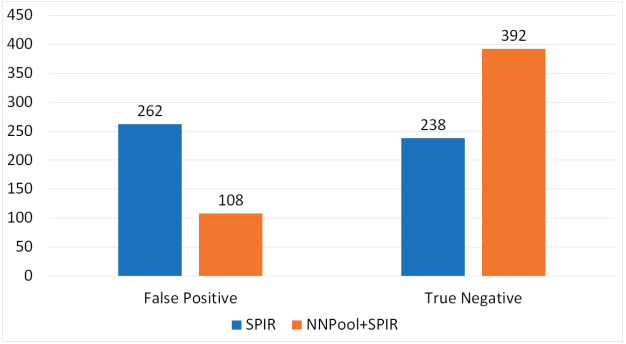

Fig. 4. In orange (blue) the values of FP and TN metrics result from NNPool (SPIR) in photo-identification on D_v.

IV. CONCLUSION AND FUTURE WORKS

Overall, the performance of NNPool+SPIR pipeline appears quite promising, in addition to the fact that this methodology automatically processes large amounts of data with no interaction by the user. When paired with a photo-ID algorithm, such as SPIR [3], the ability of NNPool to identify *unknown* Risso's dolphins, namely those dolphins never encountered during previous surveys, can open new frontiers to photo-identification studies. Combining the advantages and disadvantages of both methodologies to achieve better photo-identification. In fact this double-step processing mechanism is definitely the best approach with as first step a NNPool run, in order to identify the *unknown* dolphins. Subsequently, the *unknown* dolphins should be set apart through a manual photo-identification procedure, while the remaining *known* data set will be analyzed using SPIR algorithm, in order to automatically photo-identify the individuals portrayed in the photographs. This combination of algorithms will provide higher photo-identification performances on novel and larger datasets.

V. REFERENCES

[1] S. Gaspari and A. Natoli, "*Grampus griseus* (Mediterranean subpopulation). The IUCN Red List of Threatened Species," *IUCN*, vol. e.T16378423A16378453, 2012.

[2] K. L. Hartman, "Risso's dolphin: *Grampus griseus*," in *Encyclopedia of Marine Mammals*, Elsevier, 2018, pp. 824–827. DOI: 10.1016/b978-0-12-804327-1.00219-3.

[3] R. Maglietta, V. Reno, G. Cipriano, C. Fanizza, A. Milella, E. Stella, and C. R, "DolFin: an in-

novative digital platform for studying Risso's dolphins in the Northern Ionian Sea (North-eastern Central Mediterranean)," *Scientific Reports*, vol. 8, p. 17 185, 2018.

[4] V. Reno, G. Dimauro, G. Labate, E. Stella, C. Fanizza, G. Cipriano, R. Carlucci, and R. Maglietta, "A SIFT-based software system for the photo-identification of the Risso's dolphin," *Ecological Informatics*, vol. 50, pp. 95–101, 2019.

[5] P. M. Panchal, S. R. Panchal, and S. K. Shah, "A comparison of SIFT and SURF," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 2, pp. 323–327, 2013.

[6] R. Maglietta, A. Bruno, V. Reno, G. Dimauro, E. Stella, C. Fanizza, S. Bellomo, G. Cipriano, A. Tursi, and R. Carlucci, "The promise of machine learning in the Risso's dolphin *Grampus griseus* photo-identification," in *IEEE International Workshop on Metrology for the Sea*, 2018, DOI: 10.1109/MetroSea.2018.8657839.

[7] R. Maglietta, V. Reno, R. Caccioppoli, E. Seller, S. Bellomo, F. C. Santacesaria, R. Colella, G. Cipriano, E. Stella, K. Hartman, C. Fanizza, G. Dimauro, and R. Carlucci, "Convolutional neural networks for risso's dolphins identification," *IEEE Access*, pp. 1–1, 2020. DOI: 10.1109/access.2020.2990427.

[8] K. Hartman, F. Visser, and A. Hendriks, "Social structure of Risso's dolphins (Grampus griseus) at the Azores: a stratified community based on highly associated social units," *Canadian Journal of Zoology*, vol. 86, pp. 294–306, 2008.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. DOI: 10.1038/nature14539.

[10] R. Maglietta, N. Amoroso, M. Boccardi, S. Bruno, A. Chincarini, G. Frisoni, P. Inglese, A. Redolfi, S. Tangaro, A. Tateo, R. Bellotti, and ADNI, "Automated hippocampal segmentation in 3D MRI using random undersampling with boosting algorithm," *Pattern Analysis and Applications*, vol. 19, p. 579, 2016.

[11] R. Maglietta, N. Amoroso, S. Bruno, A. Chincarini, G. Frisoni, P. Inglese, S. Tangaro, A. Tateo, and R. Bellotti, "Random Forest classification for hippocampal segmentation in 3D MR images," in *12th International Conference on Machine Learning and Applications*, 2013, pp. 264–267.

[12] V. Vapnik, "An overview of statistical learning theory," *IEEE Trans on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[15] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Neurocomputing*, F. F. Soulié and J. Hérault, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 227–236.

[16] V. Renò, G. Losapio, F. Forenza, T. Politi, E. Stella, C. Fanizza, K. Hartman, R. Carlucci, G. Dimauro, and R. Maglietta, "Combined color semantics and deep learning for the automatic detection of dolphin dorsal fins," *Electronics*, vol. 9, no. 5, p. 758, May 2020. DOI: 10.3390/electronics9050758.

[17] G. Dimauro, "A new image quality metric based on human visual system," in *Proceedings of IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS)*, 2012, pp. 69–73.

[18] G. Dimauro, N. Altomare, and M. Scalera, "PQMET: A digital image quality metric based on human visual system," in *Proceedings of 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2014, pp. 1–6.

[19] N. Venkatanath, D. Praneeth, B. Chandrasekhar, S. Channappayya, and S. Medasani, "Blind image quality evaluation using perception based features," in *Proceedings of the 21st National Conference on Communications (NCC) Piscataway, NJ: IEEE*, 2015.