



HAL
open science

Automatic underwater fish species classification with limited data using few-shot learning

Sebastien Villon, Corina Iovan, Morgan Mangeas, Thomas Claverie, David Mouillot, Sébastien Villéger, Laurent Vigliola

► To cite this version:

Sebastien Villon, Corina Iovan, Morgan Mangeas, Thomas Claverie, David Mouillot, et al.. Automatic underwater fish species classification with limited data using few-shot learning. *Ecological Informatics*, 2021, 63, pp.101320. 10.1016/j.ecoinf.2021.101320 . hal-03415715

HAL Id: hal-03415715

<https://hal.umontpellier.fr/hal-03415715>

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Automatic underwater fish species
2 classification with limited data using
3 few-shot learning

4
5 Sébastien Villon^{a,*}, Corina Iovan^a, Morgan Mangeas^a, Thomas Claverie^{b,c}, David
6 Mouillot^{c,d}, Sébastien Villéger^c, Laurent Vigliola^a

7
8 ^a ENTROPIE, IRD, University of New-Caledonia, University of La Reunion, CNRS,
9 Ifremer, Labex Corail, Noumea, New-Caledonia, France

10 ^b CUFR Mayotte, France

11 ^c MARBEC, University of Montpellier, CNRS, IRD, Ifremer, Montpellier,
12 France

13 ^d Institut Universitaire de France, Paris, France

14
15 * Corresponding author: sebastien.villon@ird.fr

16

17

18

19

20

21 Abstract:

22 Underwater cameras are widely used to monitor marine biodiversity, and
23 the trend is increasing due to the availability of cheap action cameras.
24 The main bottleneck of video methods now resides in the manual
25 processing of images, a time-consuming task requiring trained experts.
26 Recently, several solutions based on Deep Learning (DL) have been
27 proposed to automatically process underwater videos. The main limitation
28 of such algorithms is that they require thousands of annotated images in
29 order to learn to discriminate classes (here species). This limitation
30 implies two issues: 1) the annotation of hundreds of common species
31 requires a lot of efforts 2) many species are too rare to gather enough
32 data to train a classic DL algorithm. Here, we propose to explore how
33 few-shot learning (FSL), an emerging research field, could overcome DL
34 limitations. Few-shot learning is based on the principle of training a Deep
35 Learning algorithm on “how to learn a new classification problem with
36 only few images”. In our case-study, we assess the robustness of FSL to
37 discriminate 20 coral reef fish species with a range of training databases
38 from 1 image per class to 30 images per class, and compare FSL to a
39 classic DL approach with thousands of images per class. We found that
40 FSL outperform classic DL approach in situations where annotated images
41 are limited, yet still providing good classification accuracy.

42

43 Keywords: few-shot learning, Deep learning, video, marine biodiversity

44

45 Introduction

46

47 The world's ecosystems have entered an era of anthropogenic
48 defaunation where human activities have triggered global decline in
49 animal abundance, species range contraction and a new wave of species
50 extinction [1]. This global change is threatening ecosystem services
51 worldwide hence the stability of our food systems, economies, and health.
52 Defaunation is more advanced in terrestrial and freshwater ecosystems
53 than in the marine environment where it started centuries later. However,
54 the pace of defaunation is accelerating in oceans mostly due to the
55 advent of industrial fishing since a century ago [2]. Given this context of
56 global changes rapidly affecting fish communities, it is imperative to
57 monitor fish biodiversity over time, on a large scale and using non-
58 destructive methods.

59 Fish biodiversity surveys in the marine environment are typically
60 performed by divers. Although dive visual censuses provide a great deal
61 of information on some shallow habitats, there are many limitations. First,
62 divers are limited by depth and can hardly perform long dives to count
63 fish below 30 m, ignoring mesophotic habitats and deeper ecosystems.
64 Second, divers are limited by time and generally focus their 2-4 dives per
65 day in the most speciose hard-substrate habitats, and ignore less rich and
66 often immense adjacent soft-bottom habitats. Third, dive surveys provide
67 data at a slow rate so that the compilation of global fish biodiversity
68 database takes decades of efforts by multiple teams of highly skilled
69 taxonomic divers (e.g. [3], [4]). This is a major restriction to the
70 necessary temporal monitoring of global marine ecosystems, although a
71 few time series exist in some countries¹ [5].

¹ AIMS, Long-Term Monitoring Program: Visual Census Fish Data (Great Barrier Reef)
<https://apps.aims.gov.au/metadata/view/5be0b340-4ade-11dc-8f56-00008a07204e>

72 Underwater videos (UV) are increasingly used [6] to overcome the
73 limitations of diver-based surveys to quickly collect large amounts of
74 data. For instance, more than 15,000 video stations were deployed in 58
75 countries in just three years for the first global assessment of the
76 conservation status of reef sharks [7]. Furthermore, underwater video
77 surveys can be performed in [many](#) habitats, with some example in
78 shallow reefs [8], sandy lagoons [9], deep sea [10], and even in the
79 pelagic ecosystem [11]. Deploying underwater video stations does not
80 require expert taxonomists and is now quite inexpensive with the
81 improvement of cheap action cameras since a few years. The bottleneck
82 to analyse this data now resides in the manual processing of the videos.
83 Indeed, manually extracting fish biodiversity and abundance data from
84 raw videos requires unsustainable workload by highly trained taxonomic
85 experts. [Although this annotation work can be improved through citizen
86 science \[12\]–\[14\], such time-consuming and expensive task cannot
87 match the increasing size of datasets, up to 20,000 hours of videos for
88 global surveys \[7\] and the necessary monitoring of global oceans over
89 time.](#)

90

91 As the demand for automatic methods to analyze underwater videos is
92 [rising](#), the latest generation of deep learning algorithms (DL), and in
93 particular convolutional neural networks (CNNs) [are](#) increasingly used for
94 species identification [14]–[17] and fish detection [18]–[21]. However,
95 these algorithms require a large dataset of annotated images (thumbnails
96 hereafter) in order to train a robust model, able to provide satisfying
97 results. Therefore, this method still requires collecting an important image
98 dataset manually annotated by experts. This is especially problematic in
99 highly diverse faunas such as coral reef fish that encompass nearly 6500
100 species worldwide [22]. Furthermore, a universal pattern in species
101 distribution, including fish communities, is that both rare and common
102 species are found in every community, with the fraction of rare species

103 more important in rich ecosystems, such as coral reefs [23], [24]. It is
104 therefore almost impossible to gather enough thumbnails of rare species
105 to efficiently train a deep neural network in a “classic” way, which
106 requires thousands of images per species [25]–[27]

107

108

109 There are two ways to tackle this problem of lack of data. The first one
110 consists of directly addressing the data itself, through data augmentation
111 [28]–[30]. The second option is to change the classification algorithm.
112 Few-shots learning (FSL) algorithms [31], [32] are designed to compute a
113 classification task (query, noted Q) with only a few thumbnails to train
114 (Support Sets, noted SS), and it has been increasingly studied since 2017
115 [33]. Few-shots learning methods are divided into three main
116 approaches. Metric-based methods are embedding both queries (Q) and
117 support sets (SS), before assigning to the query a class, according to
118 distances computed between Q and SS ([34]–[36]). The second approach
119 consists of 1) training a model on a large database, and 2) adapt this
120 model to a new task with few examples, while not forgetting the concepts
121 learned previously [37], [38]. Finally, optimization-based methods are
122 designed to adapt quickly to new tasks, hence able to learn a
123 classification task with few examples [33], [39], [40]. Optimization-based
124 algorithms showed promising results in deep learning few-shot
125 classification [33], [41], [42]. Such methods propose to pre-train (or
126 “meta-train”) a model with existing databases (e.g. MiniImageNet [43] ,
127 Ominglot [44]) on different tasks so it can adapt easily to a new one. For
128 object identification, a task is defined by the classes the model has to
129 discriminate. Once this model, called “meta-model” has been trained, it
130 can then be tuned to operate on a new task with a very limited dataset,
131 usually only 1-5 thumbnails per class.

132

133 In this study we propose to compare the efficiency of optimization-based
134 few-shot learning and standard large dataset deep-learning methods to
135 identify coral reef fish species on images. More specifically, we aim to
136 determine how well a classic deep learning architecture trained with
137 thousands of images [and the benefit of data augmentation](#) (hereafter DL)
138 and FSL algorithms perform in situations where training thumbnail
139 dataset is large or limited. To achieve this, we first trained a classic DL
140 architecture built for image classification [45] on a large dataset of
141 69,169 thumbnails, and on a more limited dataset of 6,320 thumbnails for
142 20 coral reef fish species. Then, we trained a few-shots, optimization-
143 based learning algorithm [39] on the exact same training datasets while
144 varying the number of shots from 1 to 30. Finally, we compared the
145 capacity of DL and FSL models to correctly identify species on an
146 independent thumbnail dataset, and modelled the asymptotic relationship
147 between classification accuracy and the number of thumbnails in the
148 training datasets for both classic DL and FSL algorithms

149

150

151 Material and methods

152

153 Thumbnail datasets

154 We used three fish thumbnail datasets (T_0 , T_1 , and T_2) extracted from
155 175 underwater videos recorded on reefs around Mayotte Island (Western
156 Indian Ocean) using GoPro Hero 3+ and GoPro hero 4+ cameras with a
157 resolution of 1920x1080 pixels. A thumbnail is defined as an image
158 containing a single labelled fish belonging to one of the 20 most common
159 fish species in the videos, and representing a broad range of sizes, colors,
160 body orientations, and background (Supp. Fig. 1, Supp. Fig. 2).

161 T_0 is composed of 69,169 thumbnails extracted from 130 videos, with a
162 range of 1,134 to 7,345 thumbnails per species (Table 1). T_1 is composed
163 of 6,320 thumbnails extracted from 20 videos with 40-1,436 images per
164 species whereas T_2 is composed of 13,232 thumbnails extracted from 25
165 videos with 55-3,896 images per species. Thumbnails size originally
166 ranged from 55x55 pixels to 500x450 pixels, but were resized to 84x84
167 pixels before being processed through FS and DL algorithms.

168 The datasets T_1 and T_2 correspond to two real scenarii where videos were
169 recorded during two trips in the field of a week each.

170 The three thumbnails datasets are fully independent, as they were
171 extracted from videos recorded at different sites, with different conditions
172 (weather, lighting, depth, time of the day, seascape) and on different
173 days.

174

175 To train our DL architecture, we applied data augmentation to T_0 and T_1 .
176 For each natural thumbnails in T_0 and T_1 , we created 9 thumbnails
177 through contrast augmentation or diminution, and horizontal flip. We then
178 obtained augmented datasets composed of 691,690 (AT_0) and 63,200

179 (AT1) images respectively [Supp. Table 1](#). Further details on thumbnail
 180 datasets and data augmentation are given in [46].

181

182 Table 1: Number of natural thumbnails extracted from the videos to build
 183 our three datasets

Family	Species	Training dataset T_0	Training dataset T_1	Test dataset T_2
Acanthuridae	<i>Acanthurus leucosternon</i>	3,259	235	491
Acanthuridae	<i>Acanthurus lineatus</i>	1,008	114	864
Acanthuridae	<i>Naso brevirostris</i>	1,134	539	1932
Acanthuridae	<i>Naso elegans</i>	7,345	1,435	3,896
Acanthuridae	<i>Zebrasoma scopas</i>	4,970	48	579
Chaetodontidae	<i>Chaetodon auriga</i>	2,134	737	502
Chaetodontidae	<i>Chaetodon guttatissimus</i>	1,182	221	68
Chaetodontidae	<i>Chaetodon trifascialis</i>	5,234	41	630
Chaetodontidae	<i>Chaetodon trifasciatus</i>	4,421	71	82
Labridae	<i>Gomphosus caeruleus</i>	3,131	57	173
Labridae	<i>Halichoeres</i>	3,192	40	287

	<i>hortulanus</i>			
Labridae	<i>Thalassoma hardwicke</i>	4,951	181	275
Lethrinidae	<i>Monotaxis grandoculis</i>	3,893	797	1,422
Monacanthidae	<i>Oxymonacanthus longirostris</i>	2,553	54	55
Pomacentridae	<i>Abudefduf vaigiensis</i>	5,124	376	216
Pomacentridae	<i>Amblyglyphidodon indicus</i>	1,188	636	1,310
Pomacentridae	<i>Chromis opercularis</i>	1,525	81	93
Pomacentridae	<i>Chromis ternatensis</i>	3,640	300	156
Pomacentridae	<i>Pomacentrus sulfureus</i>	5,409	270	142
Zanclidae	<i>Zanclus cornutus</i>	3,876	86	59
TOTAL		69,169	6,320	13,232

184

185

186

187

188 Experimental design

189 To compare classic deep-learning and few-shot algorithms in situations of
 190 large or small thumbnail datasets, [we led five experiments using datasets](#)
 191 [T0, T1, T2, AT0 and AT1](#) described in Supp. Table 1 :

- 192 1) We trained a classic DL algorithms architecture with our biggest
193 dataset *A70* as a baseline for the DL accuracy;
- 194 2) We trained the same DL architecture with the same hyper-
195 parameters (e.g. model architecture and training process) but on a
196 much more limited dataset (*A71*). Hyper-parameters are the
197 parameters defining the architecture (number of layers, number
198 and size of convolutions, connections between layers) and the
199 training process of a Deep Model (learning rate, neurone activation,
200 back-propagation computation).;
- 201 3) We trained the same DL architecture with limited datasets obtained
202 by subsampling T0 to 250 and 500 images per class (here after
203 "species" when we are referring to our experiments), corresponding
204 to 2500 and 5000 thumbnails in *A70*;
- 205 4) We pre-trained a FSL architecture on the 64 training classes of
206 MiniImageNet (Supp. Fig. 3) and used T0 to build support sets (*SS*)
207 with 1, 5, 15 and 30 thumbnails for each fish species;
- 208 5) We pre-trained the same FSL architecture on MiniImageNet and
209 used the more limited T1 dataset to build support sets with 1, 5, 15
210 and 30 images per species.

211

212 We used ResNet 100 [45] as our classic deep-learning algorithm. Resnet
213 is a convolutional neural network (CNN), a DL architecture which is able
214 to both extract features from images and classify these images thanks to
215 those features [47]. In order for a CNN to build an image classification
216 model, the architecture is fed a large dataset, composed of pairs of labels
217 and images. Using this dataset, the algorithms change their inner
218 parameters in order to minimize the classification error, through a
219 process called back-propagation. The ResNet architecture achieved the
220 best results on ImageNet Large Scale Visual Recognition Competition
221 (ILSVRC [43]) in 2015, considered the most challenging image

222 classification competition. It is still one of the best classification
223 algorithms, while being easy to use and implement.

224

225 For the few-shot implementation, we used the Reptile algorithm [39].
226 Few-shot learning algorithms are specific DL algorithms, whose goal is to
227 be able to fit a model with very few training images. The Reptile algorithm
228 is based on the well-known MAML architecture [33], and more precisely
229 on the first-order version of MAML [48]. The Reptile algorithm is based on
230 the division of the training dataset into a number of tasks T_i , a task being
231 a learning problem. Through repetitively changing the task during the first
232 training phase (known as meta-training), this algorithm produces a quick
233 learner, i.e. a learner than can quickly adapt to a new task with a small
234 number of examples.

235 Here, the few-shot algorithms were tested on a classic *n-ways k-shots*
236 procedure, n being the number of classes per support set, and k the
237 number of images per class in the support set. [For instance, a 5-ways 1-](#)
238 [shots consists of training 5 classes with supports sets composed of 1](#)
239 [image per classes \(e.g. species\).](#) We set $n=5$ [34], [36], [40], [42], [49]
240 and allowed k to vary between 1 shot and 30 shots for both experiments
241 4 and 5. We did not use data augmentation for FSL experiments for
242 several reasons. First, the goal of FSL is to adapt quickly with a very
243 limited number of images. Second, to have similar settings for method
244 comparison. There were no data-augmentation in the original paper, so
245 we reproduced that. It also allowed us to compare our results with those
246 obtained on benchmarks. Third, the reason behind the use of raw data
247 instead of augmented data in few-shot learning paper is that with very
248 few training samples and few conditions, the risk of overfitting by using
249 the same image modified multiple times is far greater than in classic
250 approaches with important datasets with many conditions.

251

252

253 Model comparison

254 All the DL and FSL models were tested on the independent *T2* dataset.

255 First, we compared the results of experiments 1 and 2 in order to
256 estimate the decrease in performance of a classic ResNet DL architecture
257 when trained on a large dataset *AT0* (i.e. between 11,340 and 73,450
258 images per species after data augmentation, with an average of 3458
259 natural thumbnails per species) or trained on a more limited dataset *AT1*
260 (i.e. between 400 and 14,360 images per species after data
261 augmentation, with an average of 315 natural thumbnails per species).

262 Second, we compared the results of experiments 1 and 4 in order to
263 evaluate if the ResNet architecture outperforms the Reptile architecture in
264 a real-case situation where thumbnail dataset is large (*T0 and AT0*).

265 Finally, we compared the results of experiments 2 and 5 to determine
266 whether and to which extent a Reptile model performs better than a
267 ResNet model in a real-case situation where thumbnail dataset is limited
268 (*T1 and AT1*).

269

270

271 In order to better evaluate the performance of ResNet and Reptile
272 algorithms, we also modelled the relationship between model accuracy
273 and the number of thumbnails used to train the models. To achieve this,
274 we fitted the following asymptotic function to the results of experiments
275 1, 3 and 4 (obtained through training DL and FSL architectures on
276 datasets of various size obtained from *AT0* and *T0*):

277

278 $Accuracy = Accuracy_{\infty} \cdot (1 - \exp(-R \cdot N_{image}))$ (eq.1)

279 where $Accuracy_{\infty}$ is the asymptotic model accuracy when the number of
280 thumbnails N_{image} is infinite, and R is the rate at which the asymptote is
281 reached.

282 Equation 1 was fitted by non-linear mixed-effect modelling (NLME [50])
283 using species as a random effect .This method is widely used for fitting
284 asymptotic processes. It allows estimating and comparing asymptotic
285 accuracies of both FSL and DL algorithms, and the number of image to
286 reach these asymptotic accuracies. The number of images required to
287 reach the asymptotic accuracy was calculated as the number of images
288 corresponding to an accuracy of 0.99 times the asymptotic value,
289 meaning the asymptote was reached within 1%.

290

291

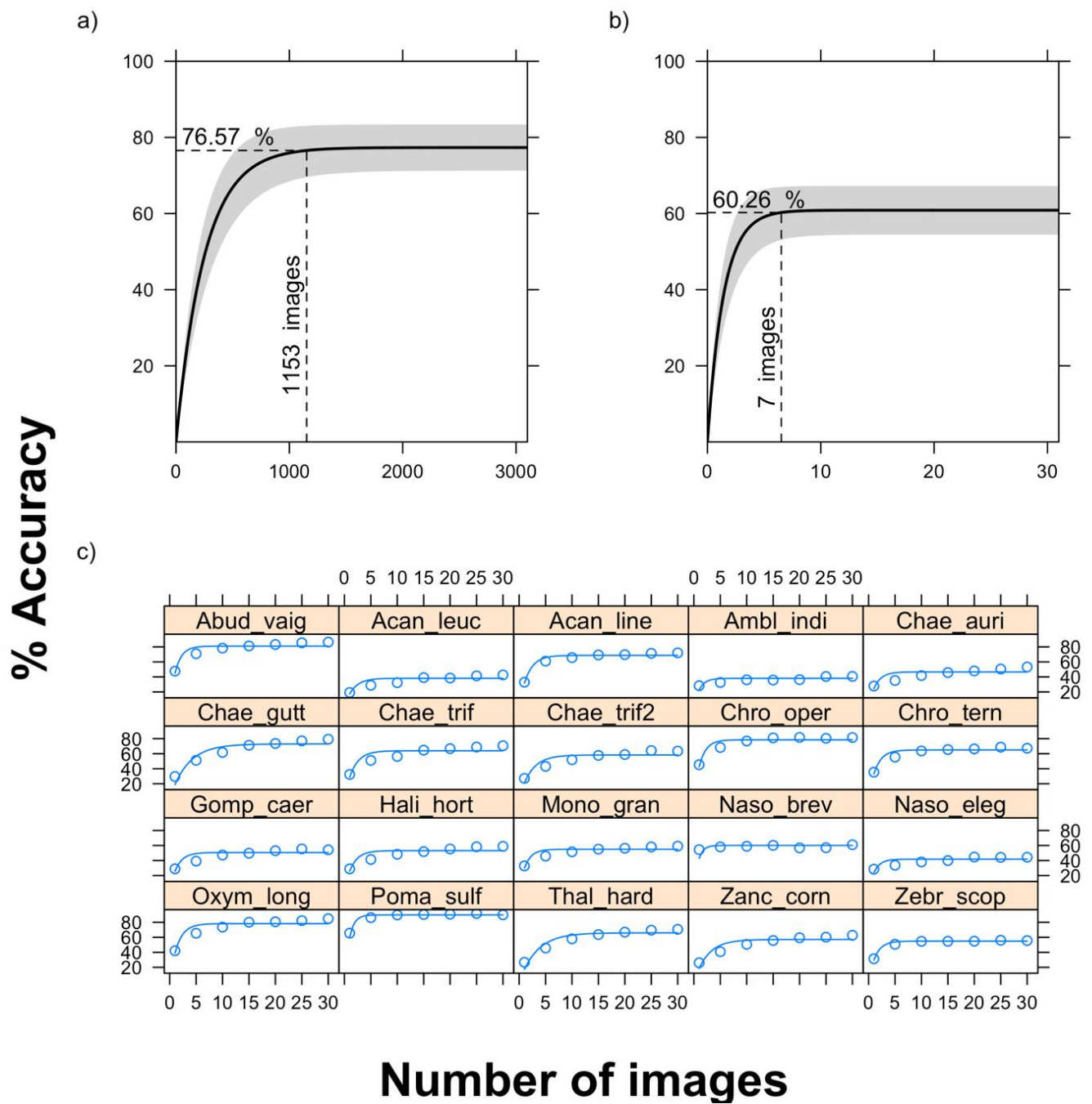
292 Results

293 The deep ResNet model trained on the large *AT0* dataset (3458 natural
294 thumbnails in average per species) during the first experiment obtained a
295 mean accuracy (i.e. percentage of correct classification) of 78.00%
296 (standard deviation (SD) of 15.16%) on *T2* test-dataset (Table 2). With
297 this model, accuracy varied among species between 54.14% (*Naso*
298 *brevirostris*) and 99.07% (*Abudefduf vagiensis*). The same ResNet DL
299 model trained on smaller *AT1* (315 natural thumbnails in average per
300 species) during the second experiment showed highly degraded
301 performance with a mean accuracy of only 42.21% (SD=24.95%). Among
302 species variation ranged with this model from only 3.49% (*Chaetodon*
303 *trifascialis*) to 85.86% (*Chaetodon auriga*).

304 The few-shot Reptile architecture trained on limited *T1* dataset during our
305 fifth experiment obtained a mean accuracy of 32.04% for the 1-shot
306 learning (SD= 12.70%) and 51.77% mean accuracy for the 30-shots
307 learning (SD= 18.96%) (Table 2,). In this scenario of limited *T1* training
308 dataset, the few-shot Reptile algorithm nearly equalled the ResNet DL
309 model with only 5 shots (41.47% accuracy for 5-shots learning on *T1* vs
310 42.21% for DL on *AT1*), and performed better beyond 10 shots (45.92%

311 of accuracy on *T2* with 10-shots learning). A pairwise proportion test
312 showed a p-value <0.0001 , assessing that FSL was significantly better
313 than DL in this scenario beyond 10 shots (Supp. Table 3) accuracy of
314 Reptile models had a standard deviation from 12.70% with one-shot
315 learning, to 18.96% with 30-shots learning, indicating important variation
316 in accuracy among species. However, this standard deviation was smaller
317 than that of the ResNet algorithm trained on the same *AT1* limited
318 dataset (24.95%).

319 Figure 1: Relationship between the number of natural thumbnails per species for
 320 training and the accuracy of deep-learning and few-shot learning models. Non-
 321 linear mixed effects asymptotic model fit for (a) DL architecture at the fixed-
 322 effect level, and for FSL architecture at (b) the fixed-effect level and (c) the
 323 random-effect level. Grey areas represent 95% CI in fixed-effects estimates.
 324 Dotted lines represent the NLME estimate of the number of images per species
 325 required to reach the 99% asymptote value. We obtained similar magnitude with
 326 the 95% asymptote value, reached with 750 images for DL and 5 with FSL.



328

329

330

331 The same few-shot Reptile architecture trained on subsets of T0 during
332 the fourth experiment obtained even better results than when trained on
333 T1, with a mean accuracy on T2 of 34.57% for 1-shot, 50.23% for 5-
334 shots, and up to 64.92% for 30-shots (Table 2).

335 Mixed-effects modelling (NLME) of T0 and AT0 experimental data showed
336 a clear pattern of asymptotic increase of accuracy with the number of
337 natural thumbnails for both Resnet and Reptile architectures (Figure 1).

338 NLME models included significant species random effect for both DL and
339 FSL (Log-likelihood tests, $P < 0.0001$).

340 The fixed-effect asymptotic value of accuracy was higher for ResNet
341 model ($Accuracy_{\infty} = 77.34\%$, 95% CI: 71.26-83.41%) than for Reptile
342 model ($Accuracy_{\infty} = 60.87\%$, 95% CI: 54.48-67.26%), illustrating higher
343 classification power of ResNet over Reptile when large numbers of
344 thumbnails are available. However, the slope of the asymptotic model
345 was two-orders of magnitude higher for Reptile (0.707, 95% CI: 0.559-
346 0.854) than for ResNet architecture (0.0040, 95% CI: 0.0032-0.0048),
347 illustrating the high capacity of Reptile FSL algorithm to learn from only a
348 few images. NLME modelling further showed that average asymptotic
349 accuracy was reached with only 7 natural thumbnails per species for
350 Reptile architecture, compared to 1153 natural thumbnails per species for
351 ResNet, confirming the strong power of Reptile method in situation of
352 limited thumbnail training dataset. However, model random effects
353 showed that some variation existed among species. For Reptile
354 architecture, asymptotic accuracy values ranged from 38.09%
355 (*Amblyglyphidodon indicus*) to 89.78% (*Pomacentrus sulfureus*), and was
356 reached with 4 to 16 training images per species. For DL architecture,
357 species asymptotes varied from 62.72% (*Monotaxis grandoculis*) to

358 96.81% (*Abudefduf vaigiensis*), and could be reached with 786-1776
359 thumbnails per species (*Supp. Table 4*).

360

361 *Table 2: Accuracy of our ResNet deep-learning (DL) and Reptile few-shots learning (FSL) models trained on T0 or*
 362 *T1 thumbnails datasets for different number of shots. Accuracy is the % correct classification of models on T2 test*
 363 *dataset. DL models were trained from T0 and T1 after data augmentation (AT0 and AT1).*

364

Image per species (on average)	DL		FSL					
	T0	T1	T1			T0		
	3458	315	1 shot	5 shots	30 shots	1 shot	5 shots	30 shots
<i>Abudefduf vaigiensis</i>	99.07	69.91	16.08	11.39	11.38	47.67	70.9	86.35
<i>Acanthurus leucosternon</i>	86.15	44.67	25.51	30.71	38.74	19.23	28.8	42.66
<i>Acanthurus lineatus</i>	59.72	20.37	39.86	56.04	72.50	32.93	61.01	72.02
<i>Amblyglyphidodon indicus</i>	58.78	60.78	25.75	26.74	32.86	28.26	32.55	40.64
<i>Chaetodon auriga</i>	87.05	85.86	18.16	25.68	36.56	27.8	35.18	53.20
<i>Chaetodon guttatissimus</i>	85.50	44.12	33.58	44.21	58.26	29.61	51.18	79.29
<i>Chaetodon trifascialis</i>	90.00	3.49	29.02	25.48	28.44	27.14	43.17	63.51
<i>Chaetodon trifasciatus</i>	87.80	28.05	38.73	50.63	66.72	32.41	51.07	70.63

<i>Chromis opercularis</i>	61.29	9.68	44.01	61.81	62.94	45.34	68.28	81.50
<i>Chromis ternatensis</i>	59.61	55.77	18.91	24.94	35.07	35.4	55.44	67.22
<i>Gomphosus caeruleus</i>	75.72	20.81	26.01	38.99	58.74	28.96	39.16	54.22
<i>Halichoeres hortulanus</i>	82.93	17.07	31.82	44.81	57.01	28.94	41.35	58.87
<i>Monotaxis grandoculis</i>	57.10	53.37	32.03	41.52	50.64	32.8	45.85	59.13
<i>Naso brevirostris</i>	54.14	68.60	47.06	54.26	64.66	54.47	58.08	61.00
<i>Naso elegans</i>	93.24	79.43	34.54	43.11	52.17	28.47	33.71	44.36
<i>Oxymonacanthus longirostris</i>	96.43	14.54	39.29	53.48		42.15	65.44	
					66.26			84.86
<i>Pomacentrus sulfureus</i>	90.14	61.97	70.90	88.18	93.93	65.53	86.21	90.00
<i>Thalassoma hardwicke</i>	90.90	51.64	25.64	44.4	67.96	26.72	45.7	70.60
<i>Zanclus cornutus</i>	81.36	40.68	18.33	31.28	44.44	26.08	40.82	62.72
<i>Zebrasoma scopas</i>	63.04	13.30	25.56	31.81	36.05	31.42	50.69	55.70
MEAN	78.00	42.21	32.04	41.47	51.77	34.57	50.23	64.92
SD	15.16	24.95	12.70	16.93	18.96	11.14	14.75	14.55

365

366

367 Discussion

368 Our experiments demonstrated that few-shot learning methods based on Reptile
369 architecture can be effectively used to drastically reduce the number of annotated
370 images for underwater fish identification. Accuracy levels obtained with few-shot
371 learning algorithm trained with only five training images are close to those of a
372 standard Deep Learning architecture such as ResNet trained with 400-14350 images
373 per species. Further, FSL architecture trained with 10 images outperformed a
374 ResNet 100 architecture trained with at least 400 images per species. This is a very
375 promising result in situations where many species need to be identified from models
376 trained with a few images, a typical characteristic in marine biodiversity
377 applications.

378 However, the important standard deviation among the different trained species
379 (18.96 SD on 30-shots) showed that few-shot algorithms may not be robust enough
380 to discriminate among similar species showing only subtle differences. Nevertheless,
381 in our 2nd experiment, our ResNet model achieved an accuracy under 40% for all
382 the species with fewer training images than 1140 (after data augmentation, i.e. 114
383 natural images), and only 7 species were identified with an accuracy greater than
384 45%. These species were represented with a range of 2700-14350 images during
385 the training phase. [We also show better results with the model trained on T0 than
386 the model trained on T1. As expected, increasing the number of images per shot
387 rely on better performances as well as increasing the per species images variability.
388 However, in real conditions, few-shot learning is to be used in a context where very
389 few images per classes are at disposal. Therefore, the dataset T1 corresponded
390 more to a real use case scenario.](#)

391 Thus, there is a trade-off to make between accuracy and robustness on one hand,
392 and the cost of video annotation by experts on the other.

393

394 Modeling the accuracy of neural networks using NLME allowed to understand the
395 number of images per species required for the Few-shot and Deep architecture to
396 reach 99% of their maximum potential accuracy. In our case study, there was a
397 150-fold factor between the average number of images required for a Deep
398 Learning architecture (1153 images) and for a Few-shot architecture (7) to reach
399 asymptotic accuracy. However, it is important to note that these numbers could
400 vary according to the number and complexity of classes fed to the deep classifier.

401 In this work we used a Reptile FSL architecture. As the field of few-shot learning is
402 quickly improving, new methods are proposed at a fast rate. While Reptile obtained
403 a mean accuracy of 61.98% on the MiniImageNet dataset (the most used
404 benchmark for few-shot learning methods) through a 5-shots learning, [51] recently
405 achieved 80.51% of accuracy on the same dataset. Although further studies are
406 required, we can reasonably assume that the improvements of FSL algorithms will
407 further expand the possible use of few-shot learning for real-life use cases.

408
409 Applied to marine and coral reef ecology, such methods requiring few examples to
410 fit a model on an identification task could be used for studies on species rarely seen
411 on screen. A key characteristic of highly diverse ecosystems is that they are
412 composed of few very common species and a large proportion of less-common and
413 rare species. Hence, the important effort required to build databases with a
414 sufficient number of images of all these rare species is the main bottleneck
415 preventing the use of Deep Learning on a large number of species. The
416 improvement of few-shot learning algorithms offers promises to build efficient
417 identification models to automatically process images and videos to localise and
418 identify rare fish species. [Such models could then be paired with more classic deep
419 architectures, more efficient to identify abundant species with the leverage of
420 important datasets.](#)

421

422 Acknowledgements:

423 This study was funded by the French National Research Agency project ANR
424 18-CE02-0016 SEAMOUNTS.

425

426

427

428

430 References

- 431 [1] R. Dirzo, H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen,
432 "Defaunation in the Anthropocene," *Science (80-.)*, vol. 345, no. 6195, pp.
433 401–406, 2014.
- 434 [2] H. S. Young, D. J. Mccauley, M. Galetti, and R. Dirzo, "Patterns , Causes , and
435 Consequences of Anthropocene Defaunation," *Annu. Rev. Ecol. Evol. Syst.*, no.
436 August, pp. 333–358, 2016.
- 437 [3] J. E. Cinner *et al.*, "Meeting fisheries, ecosystem function, and biodiversity
438 goals in a human-dominated world," *Science (80-.)*, vol. 311, no. April, pp.
439 307–311, 2020.
- 440 [4] R. D. Stuart-smith *et al.*, "Integrating abundance and functional traits reveals
441 new global hotspots of fish diversity," *Nature*, vol. 501, no. 7468, pp. 539–
442 542, 2013.
- 443 [5] A. Heenan *et al.*, "Long-term monitoring of coral reef fish assemblages in the
444 Western central pacific," *Sci. Data*, vol. 4, pp. 1–12, 2017.
- 445 [6] S. K. Whitmarsh, P. G. Fairweather, and C. Huveneers, "What is Big BRUVver
446 up to ? Methods and uses of baited underwater video," *Rev. Fish Biol. Fish.*,
447 vol. 27, no. 1, pp. 53–73, 2017.
- 448 [7] M. Aaron MacNeil *et al.*, "Global status and conservation potential of reef
449 sharks," *Nature*, no. July 2019, 2020.
- 450 [8] J. Juhel, L. Vigliola, L. Wantiez, T. B. Letessie, J. J. Meeuwig, and D. Mouillot,
451 "Isolation and no-entry marine reserves mitigate anthropogenic impacts on
452 grey reef shark behavior," *Sci. Rep.*, vol. 9, no. November 2018, pp. 1–11,
453 2019.
- 454 [9] M. Cappo, G. De, and P. Speare, "Inter-reef vertebrate communities of the
455 Great Barrier Reef Marine Park determined by baited remote underwater video
456 stations," *Mar. Ecol. Prog. Ser.*, vol. 350, pp. 209–221, 2007.

- 457 [10] V. Zintzen, M. J. Anderson, C. D. Roberts, E. S. Harvey, and L. Andrew,
458 "Effects of latitude and depth on the beta diversity of New Zealand fish
459 communities," *Sci. Rep.*, vol. 7, no. July, pp. 1–10, 2017.
- 460 [11] Tom B Letessier *et al.*, "Remote reefs and seamounts are the last refuges for
461 marine predators across the Indo- Pacific," *PLoS Biol.*, vol. 17, pp. 1–20,
462 2019.
- 463 [12] C. J. Torney *et al.*, "A comparison of deep learning and citizen science
464 techniques for counting wildlife in aerial survey images," *Methods Ecol. Evol.*,
465 vol. 10, no. October 2018, pp. 779–787, 2019.
- 466 [13] E. C. McClure *et al.*, "Artificial Intelligence Meets Citizen Science to
467 Supercharge Ecological Monitoring," *Patterns*, vol. 1, no. 7, p. 100109, 2020.
- 468 [14] M. Willi *et al.*, "Identifying animal species in camera trap images using deep
469 learning and citizen science," *Methods Ecol. Evol.*, vol. 10, no. 1, pp. 80–91,
470 2019.
- 471 [15] Z. Miao *et al.*, "Insights and approaches using deep learning to classify
472 wildlife," *Sci. Rep.*, no. May, pp. 1–9, 2019.
- 473 [16] M. Lasseck, "Audio-based Bird Species Identification with Deep Convolutional
474 Neural Networks Audio-based Bird Species Identification with Deep
475 Convolutional Neural Networks," no. January, 2020.
- 476 [17] Y. Shiu *et al.*, "Deep neural networks for automated detection of marine
477 mammal species," pp. 1–12, 2020.
- 478 [18] D. Rathi, S. Jain, and S. Indu, "Underwater Fish Species Classification using
479 Convolutional Neural Network and Deep Learning. (arXiv:1805.10106v1
480 [cs.CV])," no. June, 2018.
- 481 [19] A. Salman *et al.*, "OCEANOGRAPHY : METHODS Fish species classification in
482 unconstrained underwater environments based on deep learning," pp. 570–
483 585, 2016.
- 484 [20] S. Villon *et al.*, "A Deep Learning algorithm for accurate and fast identification

- 485 of coral reef fishes in underwater videos," *PeerJ Prepr.*, vol. 6, p. e26818v1,
486 2018.
- 487 [21] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "DeepFish: Accurate underwater
488 live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, pp.
489 49–58, 2016.
- 490 [22] P. Chabanet, S. R. Floeter, A. Friedlander, J. Mcpherson, and R. E. Myers,
491 "Global Biogeography of Reef Fishes : A Hierarchical Quantitative Delineation
492 of Regions," *PLoS One*, vol. 8, no. 12, 2013.
- 493 [23] A. P. Hercos, M. Sobansky, H. L. Queiroz, A. E. Magurran, and A. Andre, "Local
494 and regional rarity in a diverse tropical fish assemblage," *Biol. Sci.*, vol. 280,
495 pp. 81–101, 2013.
- 496 [24] G. E. Jones, M. J. Caley, and P. L. Munday, "Rarity in Coral Reef Fish
497 Communities," *Coral reef fishes Dyn. Divers. a complex Ecosyst.*, pp. 88–101,
498 2002.
- 499 [25] L. Liu, T. Zhou, G. Long, J. Jiang, and C. Zhang, "Many-Class Few-Shot
500 Learning on Multi-Granularity Class Hierarchy," *IEEE Trans. Knowl. Data Eng.*,
501 pp. 1–14, 2020.
- 502 [26] A. Li, T. Luo, Z. Lu, T. Xiang, and L. Wang, "Large-Scale Few-Shot Learning :
503 Knowledge Transfer With Class Hierarchy," in *Proceedings of the IEEE
504 Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7212–
505 7220.
- 506 [27] P. Zhuang, Y. Wang, and Y. Qiao, "WildFish : A Large Benchmark for Fish
507 Recognition in the Wild," in *Proceedings of the 26th ACM international
508 conference on Multimedia*, 2018, vol. 2, pp. 1301–1309.
- 509 [28] J. Wang and L. Perez, "The Effectiveness of Data Augmentation in Image
510 Classification using Deep Learning," *arXiv Prepr. arXiv1712.04621.*, 2017.
- 511 [29] D. A. Van Dyk and X. Meng, "The Art of Data Augmentation," *J. Comput.
512 Graph. Stat.*, vol. 8600, no. 2001, pp. 1–50, 2012.

- 513 [30] S. C. Wong, M. D. McDonnell, G. Adam, and S. Victor, "Understanding data
514 augmentation for classification : when to warp?," in *2016 international
515 conference on digital image computing: techniques and applications (DICTA)*,
516 2016, pp. 1–6.
- 517 [31] L. Fei-fei, R. Fergus, S. Member, and P. Perona, "One-Shot Learning of Object
518 Categories," *EEE Trans. pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–
519 611, 2006.
- 520 [32] M. Fink, "Object Classification from a Single Example Utilizing Class Relevance
521 Metrics," in *Advances in neural information processing systems*, 2005, pp.
522 449–456.
- 523 [33] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast
524 Adaptation of Deep Networks," *arXiv Prepr. arXiv1703.03400*, 2017.
- 525 [34] F. Sung, Y. Yang, and L. Zhang, "Learning to Compare : Relation Network for
526 Few-Shot Learning Queen Mary University of London," in *Proceedings of the
527 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,
528 2018, pp. 1199–1208.
- 529 [35] Liu Yanbin *et al.*, "Learning to propagate labels: Transductive propagation
530 network for few-shot learning," in *arXiv preprint arXiv:1805.10002*, 2019, pp.
531 1–14.
- 532 [36] J. Victor, Garcia Bruna, "FEW-SHOT LEARNING WITH GRAPH NEURAL
533 NETWORKS," in *arXiv preprint arXiv:1711.04043, 2017.*, 2018, pp. 1–13.
- 534 [37] S. Gidaris, P. Paristech, N. Komodakis, and P. Paristech, "Dynamic Few-Shot
535 Visual Learning without Forgetting," in *Proceedings of the IEEE Conference on
536 Computer Vision and Pattern Recognitio*, 2018, pp. 4367–4375.
- 537 [38] B. Hariharan, R. Girshick, and F. Ai, "Low-shot Visual Recognition by Shrinking
538 and Hallucinating Features," in *Proceedings of the IEEE International
539 Conference on Computer Vision*, 2017, pp. 3018–3027.
- 540 [39] A. Nichol and J. Schulman, "Reptile : a Scalable Metalearning Algorithm," *arXiv
541 Prepr. arXiv1803.02999, 2018*, pp. 1–11, 2018.

- 542 [40] Q. Sun and Y. L. T. Chua, "Meta-Transfer Learning for Few-Shot Learning,"
543 *Conf. Comput. Vis. Pattern Recognit.*, pp. 403–412, 2018.
- 544 [41] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few
545 Examples: A Survey on Few-Shot Learning arXiv : 1904 . 05046v2 [cs . LG]
546 13 May 2019," 2019.
- 547 [42] M. A. Jamal and H. Cloud, "Task Agnostic Meta-Learning for Few-Shot
548 Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and
549 Pattern Recognition (CVPR)*, 2019.
- 550 [43] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge,"
551 *Int. J. Comput. Vis.*, pp. 211–252, 2015.
- 552 [44] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "The Omniglot challenge :
553 a 3-year progress report," *COBEHA*, vol. 29, pp. 97–104, 2019.
- 554 [45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image
555 Recognition," in *Proceedings of the IEEE conference on computer vision and
556 pattern recognition*, 2016, pp. 770–778.
- 557 [46] S. Villon, D. Mouillot, M. Chaumont, and G. Subsol, "A new method to control
558 error rates in automated species identification with deep learning algorithms,"
559 *Sci. Rep.*, vol. 10, pp. 1–13, 2020.
- 560 [47] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, pp. 436–444,
561 2015.
- 562 [48] A. Nichol, J. Achiam, and J. Schulman, "On First-Order Meta-Learning
563 Algorithms," *arXiv*, pp. 1–15, 2018.
- 564 [49] Y. Wang, Q. Yao, and L. M. Ni, "Generalizing from a Few Examples : A Survey
565 on Few-shot Generalizing from a Few Examples : A Survey on Few-shot," *ACM
566 Comput. Surv.*, vol. 53, no. June, 2020.
- 567 [50] J. Pinheiro and D. Bates, *Mixed-effects models in S and S-PLUS*. 2006.
- 568 [51] H. Li, D. Eigen, S. Dodge, M. Zeiler, and X. Wang, "Finding Task-Relevant
569 Features for Few-Shot Learning by Category Traversal," in *Proceedings of the*













570 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),*
571 2019, vol. 1.









572

573

574

575

			
<i>Abudefduf vaigiensis</i>	<i>Acanthurus leucosternon</i>	<i>Acanthurus lineatus</i>	<i>Amblyglyphidodon indicus</i>
			
<i>Chaetodon auriga</i>	<i>Chaetodon guttatissimus</i>	<i>Chaetodon trifascialis</i>	<i>Chaetodon trifasciatus</i>
			
<i>Chromis opercularis</i>	<i>Chromis ternatensis</i>	<i>Gomphosus caeruleus</i>	<i>Halichoeres hortulanus</i>

			
<i>Monotaxis grandoculis</i>	<i>Naso brevirostris</i>	<i>Naso elegans</i>	<i>Oxymonacanthus longirostris</i>
			
<i>Pomacentrus sulfureus</i>	<i>Thalassoma hardwicke</i>	<i>Zanclus cornutus</i>	<i>Zebrasoma scopas</i>

578 *Supp. Fig. 1 : The 20 reef fish species considered in this study*

579

580

581

582

583

584



585 *Supp. Fig. 2: Diversity of individuals of the same species and of their*
586 *environments.*

587



588 *Supp. Fig. 3: Examples of classes' images in MiniImageNet*

589 *Supp. Table 1: Dataset usage during our experiments*

Dataset name	Building	Number of annotations	Usage
T0	Human Annotation	69,169	Building of supports sets of 1,5,15 and 30 images for our fourth experiment
T1	Human Annotation	6,320	Building of supports sets of 1, 5, 15 and 30 images for our fifth experiment
T2	Human Annotation	13,232	testing dataset
AT0	Data augmentation applied on T0	691,690	DL training for our first experiment
AT1	Data augmentation applied on T1	63,200	DL training for our second experiment

590

591

592

593 Supp. Table 2: Mean accuracy obtained with FSL models trained with
 594 1,5,10,15,20,25 and 30 images per species. All the images used for the
 595 supports set are from T1.

Species	Number of thumbnails in the Support set						
	1	5	10	15	20	25	30
<i>Abudefduf vaigiensis</i>	16.08	11.39	13.18	12.59	14.02	12.98	11.38
<i>Acanthurus leucosternon</i>	25.51	30.71	34.90	36.68	37.16	36.95	38.74
<i>Acanthurus lineatus</i>	39.86	56.04	63.52	66.21	70.02	70.93	72.50

<i>Amblyglyphidodon</i>							
<i>indicus</i>	25.75	26.74	28.06	28.80	30.24	32.20	32.86
<i>Chaetodo auriga</i>	18.16	25.68	30.67	32.63	33.78	35.42	36.56
<i>Chaetodon</i>							
<i>guttatissimus</i>	33.58	44.21	47.40	49.26	54.12	55.91	58.26
<i>Chaetodon trifascialis</i>	29.02	25.48	27.44	28.02	28.18	29.65	28.44
<i>Chaetodon</i>							
<i>trifasciatus</i>	38.73	50.63	53.47	60.17	63.19	64.38	66.72
<i>Chromis opercularis</i>	44.01	61.81	61.85	63.85	64.67	61.98	62.94
<i>Chromis ternatensis</i>	18.91	24.94	26.27	31.21	33.87	33.49	35.07
<i>Gomphosus</i>							
<i>caeruleus</i>	26.01	38.99	46.84	52.88	53.96	58.69	58.74
<i>Halichoeres</i>							
<i>hortulanus</i>	31.82	44.81	50.73	53.52	54.11	55.85	57.01
<i>Monotaxis</i>							
<i>grandoculis</i>	32.03	41.52	45.68	47.59	48.34	49.19	50.64
<i>Naso brevirostris</i>	47.06	54.26	61.01	59.46	59.30	62.40	64.66
<i>Naso elegans</i>	34.54	43.11	48.38	50.19	51.31	50.55	52.17
<i>Oxymonacanthus</i>							
<i>longirostris</i>	39.29	53.48	59.71	58.84	62.70	64.37	66.26
<i>Pomacentrus</i>							
<i>sulfureus</i>	70.90	88.18	90.92	93.08	92.67	94.13	93.93
<i>Thalassoma</i>							
<i>hardwicke</i>	25.64	44.40	57.80	62.03	64.97	67.86	67.96
<i>Zanclus cornutus</i>	18.33	31.28	37.27	41.71	41.70	42.95	44.44

Zebrasoma scopas 25.56 31.81 33.32 34.62 37.02 37.22 36.05

Mean 32.04 41.47 45.92 48.17 49.77 50.86 51.77

SD 12.70 16.93 17.69 18.00 18.10 18.52 18.96

596

597

598 Supp. Table 3: Probability values of the pairwise proportional test used to
 599 assess the significance of the difference between accuracies obtained
 600 through few-shot models with 5, 10, 15, 20, 25 and 30 shots and deep
 601 learning model trained on T1.

	DL	FS1	FS5	FS10	FS15	FS20	FS25
FS1	<2e-16						
FS5	0.28	<2e-16					
FS10	1.20E-08	<2e-16	3.80E-12				
FS15	<2e-16	<2e-16	<2e-16	0.0016			
FS20	<2e-16	<2e-16	<2e-16	4.10E-09	0.0379		
FS25	<2e-16	<2e-16	<2e-16	1.30E-14	8.90E-05	0.2362	
FS30	<2e-16	<2e-16	<2e-16	<2e-16	3.90E-08	0.0059	0.28

602
 603 *Supp. Table 4: Value of the asymptotic accuracy predicted by the NLME*
 604 *models, and number of natural images required for both Deep Learning*
 605 *architecture and Few-shot Learning architecture to reach 99% of this*
 606 *asymptote.*

Species	Deep Learning		Few-Shot Learning	
	Number of images required to reach 99% of the asymptote.	Accuracy asymptote value	Number of images required to reach 99% of the asymptote.	Accuracy asymptote value
<i>Naso brevirostris</i>	1,506.47	67.43	5.04	38.10
<i>Monotaxis grandoculis</i>	1,769.71	62.72	7.59	38.14
<i>Amblyglyphidodon indicus</i>	1,766.55	62.89	5.34	41.76
<i>Chromis ternatensis</i>	1,492.65	67.75	6.16	46.55
<i>Acanthurus</i>	1,425.04	69.30	6.42	50.59

<i>lineatus</i>				
<i>Zebrasoma scopas</i>	1,776.17	62.74	5.92	54.94
<i>Gomphosus</i>				
<i>caeruleus</i>	1,341.80	71.51	6.90	53.15
<i>Chromis</i>				
<i>opercularis</i>	1,112.28	78.67	5.85	54.95
<i>Zanclus cornutus</i>	1,069.22	80.57	3.66	60.08
<i>Halichoeres</i>				
<i>hortulanus</i>	1,250.29	74.18	11.60	57.14
<i>Acanthurus</i>				
<i>leucosternon</i>	1,184.97	76.31	10.21	58.28
<i>Chaetodon</i>				
<i>guttatissimus</i>	868.64	90.93	6.46	64.95
<i>Chaetodon auriga</i>	1,053.06	81.31	14.97	65.94
<i>Chaetodon</i>				
<i>trifasciatus</i>	1,076.99	80.34	7.70	63.88
<i>Chaetodon</i>				
<i>trifascialis</i>	1,087.93	79.94	7.42	68.66
<i>Pomacentrus</i>				
<i>sulfureus</i>	960.99	85.63	15.41	72.98
<i>Thalassoma</i>				
<i>hardwicke</i>	994.27	84.00	5.86	78.42
<i>Naso elegans</i>	1,074.08	80.48	6.66	78.13
<i>Oxymonacanthus</i>				
<i>longirostris</i>	833.75	93.29	5.68	80.97
<i>Abudefduf</i>				
<i>vaigiensis</i>	785.95	96.82	4.12	89.78
Mean	1221.54	77.34	7.45	60.87

607

608

