

1

Hierarchy-guided Neural Networks for Species Classification

Mohannad Elhamod,^{1,*} Kelly M. Diamond,² A. Murat Maga,^{2,3} Yasin Bakis,⁴ Henry L. Bart Jr.,⁴ Paula Mabee,⁵ Wasila Dahdul,⁶ Jeremy Leipzig,⁷ Jane Greenberg,⁷ Brian Avants⁸ and Anuj Karpatne¹

¹Virginia Tech, Blacksburg, 24060, VA, USA, ²Seattle Children's Research Institute, Seattle, 98199, WA, USA, ³University of Washington, Seattle, 98115, WA, USA, ⁴Tulane University, New Orleans, 70118, LA, USA, ⁵National Ecological Observatory Network, Battelle, Boulder, 80304, CO, USA, ⁶University of California, Irvine, Irvine, 92623, CA, USA, ⁷Metadata Research Center, Drexel University, Philadelphia, 1910, PA, USA and ⁸University of Virginia, Charlottesville, 22904, VA, USA

*Corresponding author. elhamod@vt.edu

2

3 Abstract

- 4 1. Species classification is an important task that is the foundation of industrial, commercial, ecological, and scientific applications
- 5 involving the study of species distributions, dynamics, and evolution.
- 6 2. While conventional approaches for this task use off-the-shelf machine learning (ML) methods such as existing Convolutional
- 7 Neural Network (ConvNet) architectures, there is an opportunity to inform the ConvNet architecture using our knowledge of
- 8 biological hierarchies among taxonomic classes.
- 9 3. In this work, we propose a new approach for species classification termed Hierarchy-Guided Neural Network (**HGNN**), which
- 10 infuses hierarchical taxonomic information into the neural network's training to guide the structure and relationships among
- 11 the extracted features. We perform extensive experiments on an illustrative use-case of classifying fish species to demonstrate
- 12 that **HGNN** outperforms conventional ConvNet models in terms of classification accuracy, especially under scarce training data
- 13 conditions.
- 14 4. We also observe that **HGNN** shows better resilience to adversarial occlusions, when some of the most informative patch regions
- 15 of the image are intentionally blocked and their effect on classification accuracy is studied.

16 **Introduction**

17 Depicting the branching pattern of taxa, phylogeny represents a hypothesis of evolutionary relationships based on shared similarities
18 derived from common ancestry (Hennig, 1966). From conservation to zoology, phylogenetic relationships are critical for interpreting
19 study results and implications in the biological sciences. One area, however, where this hierarchical information has yet to be
20 fully incorporated is that of machine learning and image classification. Deep neural networks have found immense success in image
21 classification problems with state-of-the-art ConvNet models (e.g., GoogleNet (Szegedy et al., 2015), AlexNet (Krizhevsky et al.,
22 2012), and VGGNet (Simonyan and Zisserman, 2014)) reaching unprecedented performance on large-scale benchmark datasets such
23 as ImageNet (Deng et al., 2009) and CIFAR (Krizhevsky, 2009). By design, deep neural networks function similarly to phylogenetic
24 analyses by extracting a hierarchy of simpler to more complex forms of abstraction in hidden layers—simpler features at lower
25 depths (e.g., edges and texture) are non-linearly composed to form complex features at higher depths (e.g., eyes and fins). This has
26 motivated several recent architectural innovations in deep learning such as ResNet (He et al., 2016), ResNeXt (Xie et al., 2017),
27 and DenseNet (Huang et al., 2017), that have enabled the learning of deep and complex hierarchy of hidden features. However, the
28 innate hierarchy extracted by neural networks from data is not necessarily tied to known evolutionary relationships in real-world
29 applications. In this work, we explore the question: *Is it possible to make use of known phylogenetic classes to inform the learning*
30 *of features, and can it lead to better generalization and robustness?*

31 Image classification in real-world biological problems such as species classification is fraught with several challenges that limit
32 the usefulness of state-of-the-art deep learning methods trained on benchmark datasets. First, real-world images of specimens suffer
33 from various data quality issues such as damaged specimens and occlusions of key morphological features (Fox and Hartman, 2019),
34 which can crucially impact classification performance. Figure 1 shows some relevant examples. Second, real-world datasets for
35 classification are limited in their scale in comparison to benchmark datasets, with limited representative power in terms of number
36 of species (Rathi et al., 2018; Ogunlana et al., 2015; Costa et al., 2013; Larsen et al., 2009; Lee et al., 2008; Allken et al., 2019; Rauf
37 et al., 2019; Ding et al., 2017), or number of images per species (Rodrigues et al., 2010; Lee et al., 2003). This is especially true for
38 rare species (Villon et al., 2021). Third, the hierarchy of features extracted by conventional deep learning frameworks, while useful
39 for prediction, do not conform to known biological hierarchies and hence do not directly translate to advancing scientific knowledge,
40 which is often a more important goal than improving predictive performance for a scientist (Karpatne et al., 2017). While these
41 challenges are applicable to species classification problems involving a variety of taxa, in this study we focus on the problem of
42 classifying the species of a fish specimen given a 2D image. We selected fishes for our study because they are a highly diverse,
43 well-studied, and an ancient group of animals that comprise almost half of all vertebrate species (Helfman et al., 2009). Further, the
44 phylogenetic relationships of fishes are well-studied (Betancur-R et al., 2017; Hughes et al., 2018), and the taxonomic classification
45 of fishes is generally aligned with phylogeny.

46 Early work on automated fish classification used basic computer vision and image processing techniques to extract shape features
47 such as landmarks and measurements and used tools such as decision trees, discriminant function analysis, and support vector
48 machines to classify species based on these features (Lee et al., 2003, 2008; Larsen et al., 2009; Ogunlana et al., 2015). Others have
49 applied scale-invariant feature transform (SIFT) and principal component analysis (PCA), and then used nearest neighbor search
50 for classification (Rodrigues et al., 2010). Only recently has the use of raw image features in its intrinsic high-dimensionality become



Fig. 1: Fish images from museum collections, demonstrating the challenges of curating fish image datasets.

51 more feasible, likely because of advances in computational capabilities. For example, (Hasija et al., 2017) employed graph-embedding
52 discriminant analysis, which reduces the image set matching problem to a point-to-point classification problem.

53 Advances in computing power have also enabled researchers to use more flexible and powerful classification methods such as
54 ConvNets, especially designed to work with high-dimensional images. The basic idea of a ConvNet is to learn convolutional kernels
55 (or filters) of a fixed size at every layer, that are applied to the input image to generate multiple channels of image outputs for the
56 next layer, followed by a final block of a max-pooling layer and a softmaxed fully connected layer to return class labels (Goodfellow
57 et al., 2016). The number of feature maps is referred to as the width of the ConvNet, while the number of layers is termed as its
58 depth. To further boost ConvNet's performance, image preprocessing techniques can be used. For example, (Rathi et al., 2018)
59 pre-processed the fish images by means of Gaussian blurring, erosion and dilation and Otsu thresholding (Otsu, 1979).

60 More recently, researchers have taken advantage of state-of-the-art architectures available in the field of deep learning for biological
61 classification. For example, in a work by (Rauf et al., 2019), the technique of *transfer learning* was explored for fish classification,
62 where neural network models *pre-trained* over large and diverse benchmark datasets were used as building blocks and then *fine-*
63 *tuned* on the fish images. Transfer learning eliminates much of the arduous task of hyper-parameter tuning otherwise required in the
64 field of deep learning, and allows researchers to build on top of well-tested benchmark neural network models. It also saves model
65 development time and boosts classification performance, especially when the available task-specific training sets are small (Yosinski
66 et al., 2014). This technique has already been successfully applied in other prior works on fish classification (Siddiqui et al., 2018;
67 Allken et al., 2019) and fish detection (Salman et al., 2019).

68 Extensions of ConvNets have also been used for several tasks such as fish detection, counting, and classification. For example,
69 (Salman et al., 2019) have used R-CNNs (Girshick et al., 2014) along with background subtraction and optical flow features to detect
70 fish in underwater videos. Similarly, (Jalal et al., 2020) attack the problems of fish detection and classification using a YOLO deep
71 neural network (Redmon et al., 2016) combined with a mixture of Gaussians model and optical flow features. In a different approach,
72 (Villon et al., 2020) post-process the prediction of a deep learning model with confidence thresholding to obtain a misclassification
73 risk estimation, which is particularly useful for identifying rare species. Finally, (Villon et al., 2021) have proposed using few-
74 shot learning (Wang et al., 2020) to achieve better results on rare species. This, however, is at the expense of less robustness at
75 distinguishing species that look too similar.

76 Our current method aims for a generic method that incorporates hierarchy to improve neural network models. Here we use
77 taxonomic relationships from fish classification to serve as an example training dataset. Specifically, we present a novel deep learning

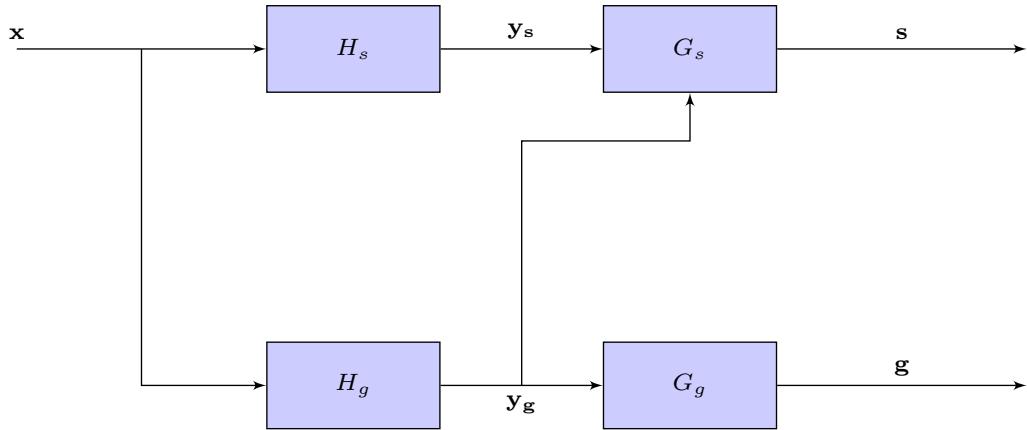


Fig. 2: Schematic diagram of **HGNN**. The top ResNet predicts the species (s) of the input fish image (x), while the bottom ResNet predicts the genus (g). To leverage the relationship between genus and species classes for guiding the hidden features of our neural network, we harness the genus features learned at an intermediate depth (y_g) of the genus ResNet and aggregate them with the species features learned at the y_s level of the species ResNet. The combination of both species and genus features are then used to make species class predictions. This architecture is described in details in the Materials and Methods section.

78 architecture termed Hierarchy-Guided Neural Network (**HGNN**) that incorporates known hierarchy among classes (available as a
 79 two-level taxonomy: genus and species) to guide the learning of features at the hidden layers of the neural network. This work builds
 80 on a history of multi-label and hierarchical classification techniques using pre-built taxonomies (Silla and Freitas, 2011; Zhang and
 81 Zhou, 2013). Our proposed architecture shown in Figure 2 consists of two sub-modules (top and bottom rows) of ResNet models
 82 operating in parallel. We use the ResNet architecture in our work because it is currently among the most widely-used and best-
 83 performing ConvNet models for benchmark computer vision problems, including fish identification (Khan et al., 2020; Jalal et al.,
 84 2020; Villon et al., 2020; Ditría et al., 2020a), although our proposed idea of **HGNN** is generic and can work with any deep learning
 85 architecture. In Figure 2, the top row ResNet predicts the species class s of the input fish image x , while the bottom row predicts
 86 the genus class g . These ResNets learn a hierarchy of features (from simple to complex) at their hidden layers useful for the tasks
 87 of species and genus classification, respectively. While both these sub-modules can be viewed as learning separate features, we know
 88 that the genus features learned in the bottom ResNet represents features at a higher level of abstraction that are directly useful for
 89 the task of species classification. Building upon this knowledge in our proposed **HGNN** framework, we harness the genus features
 90 learned at an intermediate depth H_g of the genus sub-module, and aggregate them with the species features learned at the H_s depth
 91 of the species sub-modules. The combination of both species and genus features is then used for the task of species prediction.

92 While using taxonomic information for automated fish classification is not novel (Kutlu et al., 2017), to our knowledge, the
 93 only body of work that has researched it before in the context of deep learning is by (dos Santos and Gonçalves, 2019). However,
 94 our proposed method is distinguished in two ways. First, while they have used the family and order information, we use the genus
 95 information. We argue that incorporating the genus yields more information gain as it involves more discriminative features than the
 96 order and family. Second, their model only uses the taxonomic information in the last fully-connected layer, while our philosophy
 97 is to use it at a convolutional level of the network as that allows for capturing localized visual features that are taxonomically
 98 plausible.

99 We demonstrate the effectiveness of our proposed **HGNN** model in learning meaningful, diverse, and robust features at the
 100 hidden layers of the neural network leading to better generalization performance in the target application of fish species classification,

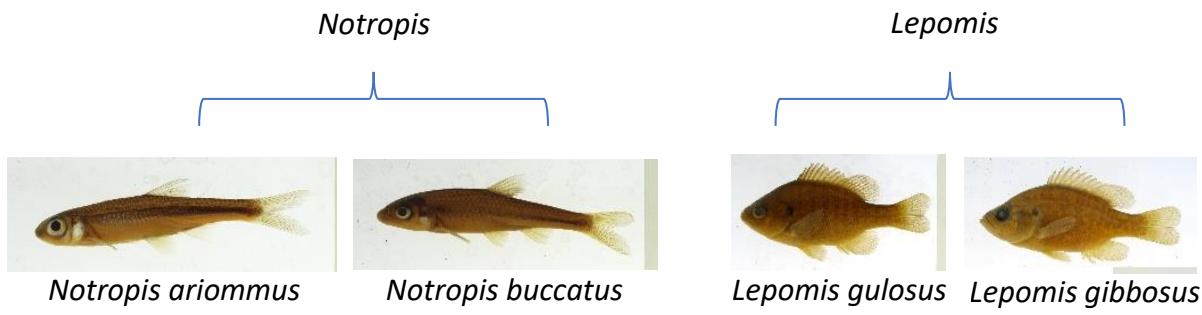


Fig. 3: Species that belong to the same genus exhibit features that are similar because of common ancestry.

101 even in the paucity of training data. We also empirically test the robustness of our model to synthetically generated image occlusions,
102 where salient regions of the input images were intentionally occluded to adversely affect classification performance. We observe that
103 by anchoring our learned features to the biologically known hierarchy among genus and species classes, our model is much more
104 robust to occlusions as compared to a data-only ‘black-box’ model that only uses image data and predicts the species with no genus
105 information (i.e. using only the top ResNet in Figure 2).

106 Materials and Methods

107 HGNN framework

108 We first present our proposed **HGNN** architecture that incorporates hierarchy among genus and species classes in neural network
109 construction. We consider the problem of predicting the target species s given input image x using a composition of neural network
110 layers. We are also given the genus level class g for every input x .

111 We make two observations to motivate our proposed **HGNN** framework. First, we assume that the hierarchical taxonomy of
112 genus and species classes captures a notion of derived similarity in terms of the discriminatory input features of every class. This is
113 true, as illustrated in Figure 3, in the context of fish classification because species classes that belong to the same genus are more
114 closely related phylogenetically than species classified in different genera. In the case of the species and genera analyzed here, with
115 only a few exceptions, this is the case (Supporting Information, Table ??). As a result, species that map to the same genus g should
116 generally share similar features at the internal representation of the neural network (e.g., filters learned at the convolutional layers).

117 This observations seems to align with some earlier work (dos Santos and Gonçalves, 2019). Second, while the mapping from s to
118 g is one-to-one, the inverse mapping from g to s is not unique. Hence, along with the shared features learned for every g , we also
119 need to learn unique features for every s to differentiate between species belonging to the same genus.

120 Building upon these two observations, we consider the following architectural composition of our neural network as shown in
121 Figure 2. First, we use a functional block of layers H_g to extract hidden features at some intermediate depth of the neural network
122 that are useful for predicting g as well as s . These hidden features are passed to another functional block G_g that predicts g . The
123 complete chain of function compositions from x to g can be represented as $F_g(x)$, where $F_g = G_g \circ H_g$ and \circ represents the function
124 composition operator. Second, we learn another functional block H_s that extracts hidden features unique to every species. Finally,

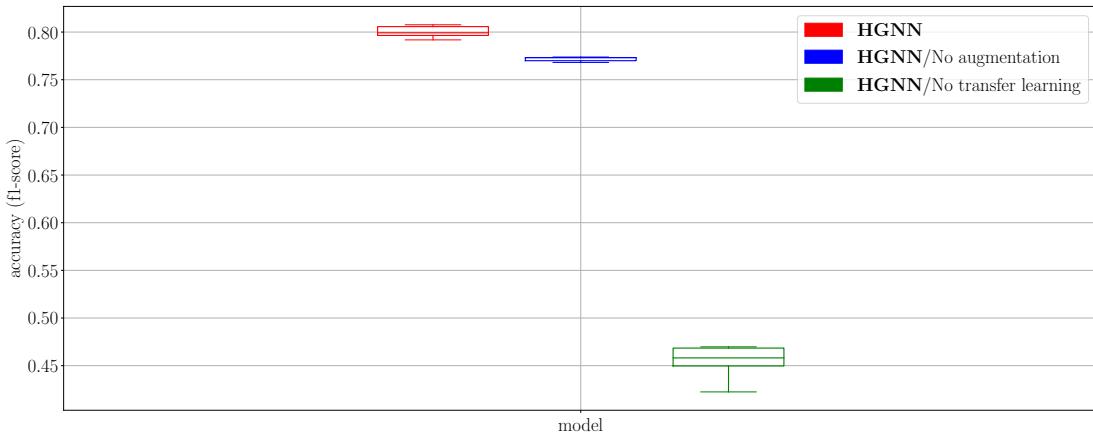


Fig. 4: Comparison among different models, showing the impact of data augmentation and transfer learning on the classification performance of **HGNN** models.

125 the features from H_s and H_g are combined using matrix addition and fed to another functional block of layers, G_s that predicts
 126 the target species s . The composition of functions mapping \mathbf{x} to s can thus be given by $f(\mathbf{x})$, where $F = G_s \circ (H_g + H_s)$.
 127 To train the functional blocks in the complete **HGNN** architecture, we consider minimizing the following objective function:

$$\min_{H_s, H_g, G_s, G_g} \lambda_s L_s(s, t_s) + \lambda_g L_g(g, t_g) \quad (1)$$

128 where L_s and L_g are loss (or error) functions defined on the space of species labels and genus labels, respectively, on the training set.
 129 Specifically, these loss functions act as a measure of difference between the correct classification (t_s and t_g), and the prediction (s
 130 and g) on the training samples, respectively. We used the cross-entropy function as our preferred choice of loss function. Further, λ_s
 131 and λ_g are trade-off hyper-parameters balancing the relative importance of L_s and L_g , respectively; their values are automatically
 132 assigned using the adaptive smoothing algorithm proposed in (Murugesan et al., 2016). Both the softmaxed outputs of our neural
 133 network model, s and g , are probability vectors whose entries range from 0 to 1 proportional to the model's credence about each
 134 species and genus class, respectively.

135 As mentioned in the Introduction, our model is composed of two identical ResNets. The first ResNet comprises of H_g and G_g ,
 136 while H_s and G_s constitute the other. In our experiments, we found that the best point to extract the intermediate genus features
 137 (i.e. the point between H_g and G_g) is right before the final max-pooling layer. The same point in the other ResNet is used to
 138 combine the genus and species features. Instead of initializing our neural network parameters (or weights) with arbitrary values, we
 139 used pre-trained weights of ResNet trained on the ImageNet benchmark dataset as a good starting solution for our target problem
 140 of fish classification. Then, by optimizing the loss function in equation (1) on the fish training dataset of interest, we fine-tuned the
 141 parameters of the entire network to be more specialized for our target task. This technique, which is called transfer learning (Tan
 142 et al., 2018a), is widely adopted in the field of deep learning particularly in applications of computer vision, and has proven its
 143 effectiveness in scenarios with data paucity. In our preliminary experiments, as shown in Figure 4, we have found using this mode
 144 of transfer learning to increase the model's average performance by about 35%.

145 Evaluation

146 **Data Collection and Pre-processing**

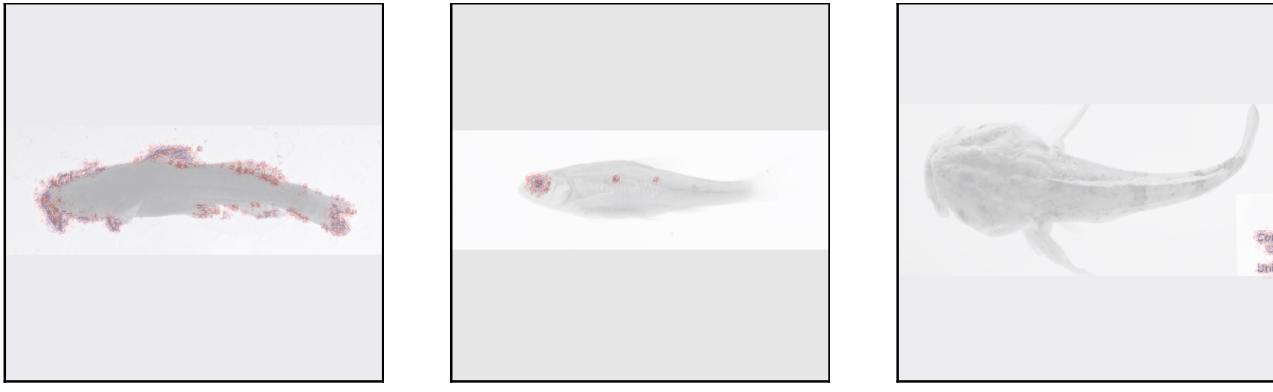
147 Our dataset comprises of images contributed by five museums that participated in the Great Lakes Invasives Network Project (GLIN).
148 More information about this project can be found in the Data Availability section. This dataset, as is typical for biological species
149 images, is highly imbalanced; some species have only a few images, while others have thousands. To alleviate this problem, and
150 for computational feasibility, we created a number of subsets of the dataset for the purpose of training and evaluation. Specifically,
151 we created two subsets that differ in terms of classification complexity (or difficulty). The first subset is called **Easy** and comes
152 from a single museum (Illinois Natural History Survey). Therefore, its images are homogeneous in terms of lighting and camera
153 conditions. The second is called **Hard** and its images are aggregated from across all museums, making it a larger, more diverse, and
154 more complex dataset. Comparing results from these two datasets helps illustrate the effects of dataset complexity on classification
155 performance. We further created two subsets of the **Easy** dataset by capping the number of images per species in the **Easy** dataset
156 to 50 or 100. These different dataset sizes help illustrate how training data paucity impacts the model's classification performance.
157 Henceforth, the suffix of the dataset will refer to the number of images per species. For example, **Easy/100** has 100 images per
158 species. Table 1 gives a statistical summary of each dataset considered in this study. More details can be found in the Supporting
159 Information document, Tables ??, ??, and ??

160 The acquired fish images typically contained a ruler, specimen label(s), and species tags along with the fish specimen. To retain
161 only the fish region in the images, we trained a 2D Unet model (Goodfellow et al., 2016) using a small portion of our data in the
162 ANTsRNet software (Tustison et al., 2018). We manually segmented the background, fish, scale bar, and field notes on 550 images
163 using 3D Slicer (Kikinis et al., 2014). We used weights from the trained model to automatically mask and crop the fish specimen
164 portion of the remainder images. With the exception of rare cases where the fish overlapped the scale bar and/or the field notes,
165 which were discarded, this pipeline resulted in successful generation of RGB fish-only images at the original resolution. The pipeline
166 was implemented in R using ANTsR (Avants, 2019) and ANTsRNet.

167 Once the cropped fish images were obtained, we performed data augmentation by randomly applying standard image
168 transformations used in deep learning for computer vision, including translations of up to 0.25 of the image dimension, flips
169 with a probability of 30%, rotations of up to 60°, and Gaussian random intensity variations using PCA with $\sigma = 0.1$ of the color
170 channel value (Krizhevsky et al., 2012). Data augmentation is critical when using ConvNets for image processing and is a common
171 practice for fish classification (Villon et al., 2020, 2021), especially when the available data is limited (Shorten and Khoshgoftaar,
172 2019). By training the model on variations of the same image, the model is deterred from learning nuanced patterns in the images
173 that can lead to spurious performance, such as the intensity of the background, and encouraged to be robust under variable input
174 conditions. Our preliminary results, as shown in Figure 4, indicate that data augmentation boosted the model's accuracy by about
175 2.8%.

176 **Evaluation Setup for Comparing Classification Performance**

177 In the process of training black-box neural network architectures, it is common to observe higher generalization errors when the
178 amount of training data is small. However, in **HGNN**, we show that by including a biological knowledge-guided loss term (see
179 Equation 1) in the learning objective of neural networks, we can achieve reasonably good generalization performance even in
180 situations where training data are scarce. This is in alignment with the observations made in a previous work by (Jia et al., 2019).



(a) Barbels and fins

(b) Eye

(c) Label paper

Fig. 5: Saliency maps of different fish images obtained for **Blackbox-NN**. Pixels in red denote image regions with high saliency scores, indicating higher importance of those regions for fish classification as perceived by the model.

181 To test for this hypothesis in the context of fish classification, we compared the classification performance of our proposed model
182 to a baseline black-box neural network architecture (termed **Blackbox-NN**) comprising of a ResNet of the same size and shape
183 as that of one of the ResNets of our proposed model. Specifically, we compared the performance of **HGNN** and **Blackbox-NN**
184 on each of the three data subsets mentioned in Table 1. For each of these subsets, we used 64% of the data for training, 16% for
185 validation, and the remainder for testing. To measure classification performance, we used the f1-score of the correct species class
186 (Tan et al., 2018b). Throughout this paper, we used box plots to show the model’s performance over five random runs of neural
187 network training. To obtain the best-performing neural network models, we performed an explorative Naïve-Bayes approach for
188 hyper-parameter search and fine-tuning. Then, we picked those parameters that performed best on the validation set.

189 **Tools for Deep Learning Visualization and Assessing Robustness to Adversarial Occlusions**

190 Saliency maps (Simonyan et al., 2014) are heatmaps of the gradients of a neural network model’s output with respect to its input.
191 In other words, a saliency map shows how strongly do changes in pixel values of a certain region of the image cause a change in the
192 species’ probability, highlighting the areas of the image that are most decisive for the classification problem. While other tools, such
193 as GradCAM (Selvaraju et al., 2017), have been used for the same purpose (dos Santos and Gonçalves, 2019), we found saliency
194 maps to be more powerful and capable of detecting the most subtle visual features. Figure 5 shows some examples of saliency maps
195 obtained for **Blackbox-NN**. The code we used for generating these saliency maps is inspired by FlashTorch (Ogura and Jain, 2020),
196 an implementation tool based on Guided Back-propagation (Springenberg et al., 2015). As we can see in Figure 5, the baseline
197 model is quite sensitive to different features of the input fish image for different species, including barbels and fins in Figure 5a
198 and the eye in Figure 5b. Saliency maps are also a good debugging tool as they can reveal cases where the model is “cheating” or
199 looking at irrelevant features of the image that are not biologically meaningful for the purpose of fish classification. An example of
200 such a case is presented in Figure 5c, where the model is incorrectly picking up pixels around the note on the label paper in the
201 image as regions with high saliency scores. In this way, saliency maps can be used for “interpreting” the learned features of neural
202 network models.

203 Along with offering interpretability, saliency maps can also be used for investigating the resiliency (or robustness) of neural
204 networks to adversarial occlusions. For example, by occluding regions (or patches) in the input image with high saliency scores, a

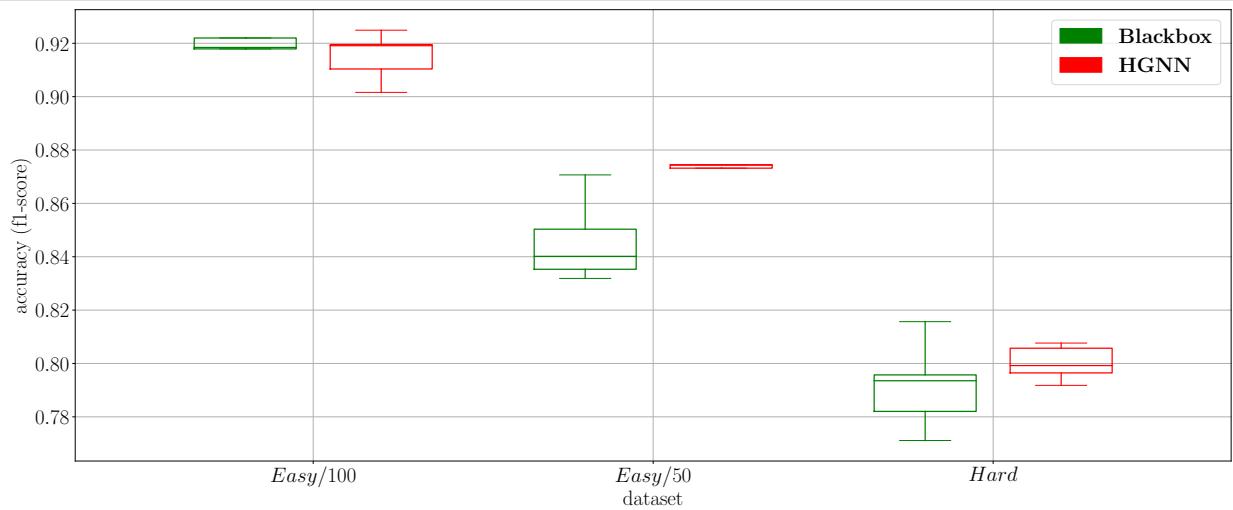


Fig. 6: Classification performance across different subsets of the GLIN dataset for **HGNN** and **Blackbox-NN**. By definition, as the boxes for the two models do not overlap for **Easy/50** and **Hard**, it means there is at least 95% confidence (McGill et al., 1978) that the median accuracy of **HGNN** is higher than **Blackbox-NN**

205 neural network model's reliability at making correct predictions can be stress-tested even when it is starved off information from
 206 salient image regions. To measure the robustness of a model at every round of adversarial occlusions, we calculated the average
 207 probability of the correct class predicted by the model on an input image \mathbf{x} , averaged over all test images as $\mathbb{E}_{\mathbf{x}}(P_{t_s}(\mathbf{x}))$. The
 208 higher this metric, the less confused the model is about the input. Further, by measuring drops in this metric as a consequence of
 209 adversarial occlusions, we can evaluate if a model is too sensitive to selective regions of the input image (with the highest saliency
 210 score contributions), which when obstructed can confuse a model into making incorrect predictions. We make use of this metric to
 211 assess the robustness of **Blackbox-NN** and **HGNN** in our experiments.

212 Results

213 Effect of Dataset Complexity and Training Size

214 Figure 6 shows a comparison between **HGNN** and **Blackbox-NN** on three subsets of the GLIN dataset: **Easy/100**, **Easy/50**, and
 215 **Hard**. Two observations can be made from this figure. First, as datasets become more complex (e.g., the **Hard** dataset) and/or
 216 subject to less training data (e.g., the **Easy/50**), the performance of the model deteriorates. Second, and more importantly, the
 217 impact of our method is more pronounced exactly when data is scarce and the dataset is complex. As Figure 6 shows, while the
 218 median performance of **HGNN** is almost equal to that of **Blackbox-NN** for **Easy/100**, which is the easiest of the datasets, the
 219 former clearly outperforms the latter on both **Easy/50** and **Hard**. This highlights our model's power and ability to compensate for
 220 the relative lack of data with respect to dataset complexity by incorporating biological knowledge.

221 Effect of Adversarial Occlusion

222 To demonstrate **HGNN**'s resiliency to adversarial occlusions, we iteratively cover regions (or patches) in an image with the highest
 223 saliency scores and report the probability of the correct class predicted by the model over the occluded image. Figure 7 shows an
 224 example of this process on an illustrative fish specimen from the **Easy/50** dataset. From left to right, the figure shows a progression

225 from an image with no occlusion towards applying more patches of adversarial occlusions (seen as green square patches) on the
226 same image. Below each image is the model's predicted probabilities over the 5 most probable species sorted in descending order,
227 for both **HGNN** (top row) and **Blackbox-NN** (bottom row). We make a number of observations here. First, all of the saliency
228 maps highlight the features of importance for classifying this fish, namely the eye, nostrils, and the dorsal fin. However, notice
229 that the saliency maps for **HGNN** are slightly different from that of **Blackbox-NN**, demonstrating that the two models are not
230 looking at the image in the exact same way (i.e., they have distinct saliency maps). This difference is important for making a fair
231 comparison between the two models. Second, even when there is no occlusion, while **Blackbox-NN** makes the correct prediction, its
232 probability of the correct species class is significantly lower than that of **HGNN**'s. This demonstrates **HGNN**'s ability to extract
233 more useful and generalizable features from images for fish classification. Third, after applying two patches of occlusions (in the
234 middle column), we notice that even though both models get the species right, **Blackbox-NN**'s second guess is not within the
235 correct genus. Finally, and most importantly, after applying four patches of occlusions (in rightmost column), we notice that while
236 both models start predicting the wrong class, **HGNN** is still within the correct genus, while **Blackbox-NN** is not. It follows that
237 the **Blackbox-NN** model is not learning phylogenetic features that could be used in other tasks, such as trait segmentation. To
238 drive this point home, we automate this process for the entire dataset and compute the average predicted probability of the correct
239 class across all images, as a function of the number of adversarial occlusions applied to the images. Table 2 reports the results for
240 each number of patches ranging from 0 (no occlusion) to 4. We can see that **HGNN** shows higher average probability of the correct
241 class across all number of patches in comparison with **Blackbox-NN**. This demonstrates **HGNN**'s ability to generalize and handle
242 image imperfections better, especially when the most informative (or salient) regions of the image are occluded.

243 **Discussion**

244 In this paper, we have shown that embedding the hierarchical taxonomy of the genus and species classes in the design and learning of
245 neural networks leads to solutions with better generalization, superior accuracy, and better resiliency to adversarial occlusions. Most
246 of the deep learning methods currently in the literature perform tasks without learning biologically-relevant features. Our proposed
247 method leverages a particularly important aspect of species classification—the hierarchical arrangement of taxon names—which
248 improves model interpretability and biological-validity. The aim of our method is to provide biologists not only with the correct
249 classification, but also with a plausible one when it fails.

250 An ultimate goal of this research is to augment biological information on the connections among phenotype, genotype and
251 environment into deep learning, so that an understanding of genealogical relationships among species is discovered by our neural
252 networks. While we have not fully investigated these relationships here, a future direction of our project is to explore how the
253 anatomical features of species learned by our models relate to the environments the species were collected from and how closely
254 related the species are. This would increase understanding of how the environment and genealogy shape the phenotypes of species.
255 Moreover, we plan to investigate how such learned features aid us in other relevant tasks, such as segmenting the phenotypic traits
256 of species. Finally, we also plan to exploit other forms of hierarchical information such as phylogenetic tree-based distances among
257 species to better understand how this informs biologically-informed neural network feature learning.

258 Recent advances in image computation are enabling automated methods of extracting phenotypic data from specimen images.
259 We hope that our present framework for leveraging biological information in training machine learning models will have a direct

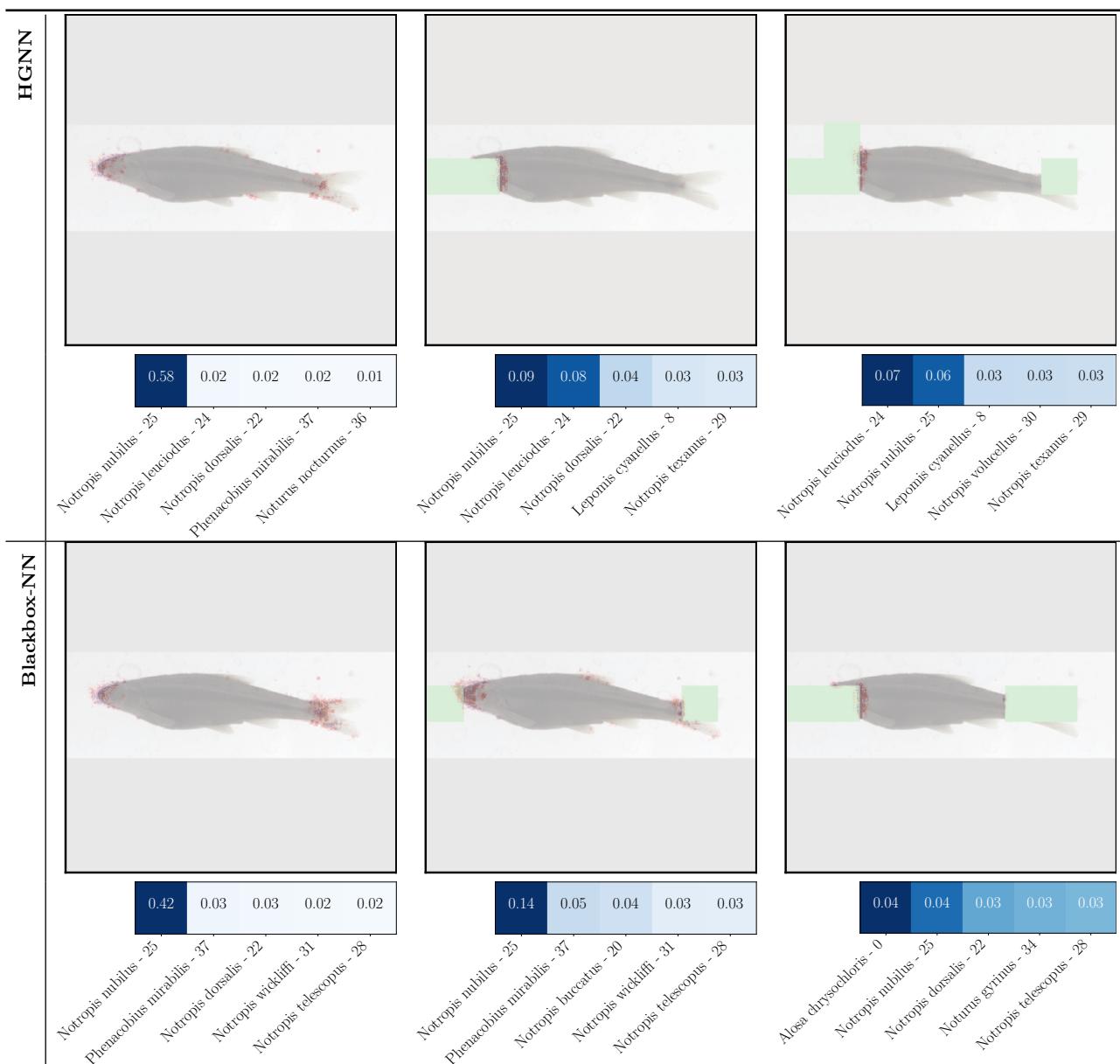


Fig. 7: Saliency maps showing the effect of adversarial occlusions (shown as green square patches) on the predicted probabilities of the species class produced by **HGN** (top row) and **Blackbox-NN** (bottom row) on an example fish image. The left-most column corresponds to the case with no occlusion, while the number of occlusions increase as we go from left column to the middle column (2 patches) to the right-most column (4 patches).

260 impact on several biologically relevant computer vision tasks, including species detection (Li et al., 2016), tracking and counting
 261 (Spampinato et al., 2008), segmentation (Chuang et al., 2013; Yao et al., 2013), and classification (Ding et al., 2017; Rathi et al.,
 262 2018; Sarigul, 2017). This automation effort is essential as manual annotation is laborious and requires expertise (Villon et al.,
 263 2020), especially with the large amount of data that has become recently available (Ditria et al., 2020a). Moreover, it has been
 264 shown that automation can be more accurate than human annotation (Ditria et al., 2020b).

265 In this paper, we have focused on teleost fishes as a model system for species classification due to their high diversity and
 266 importance economically and scientifically. Fishes are the targets of recreation (Arlinghaus and Cooke, 2009), aquaculture and
 267 fisheries (Lynch et al., 2016), and conservation (Arthington et al., 2016). Fishes make up more than half of all vertebrates and they

268 play critical roles in Earth ecosystems (Near et al., 2012; Villon et al., 2020). However, our framework of **HGNN** is quite generic
269 and can be potentially applied to incorporate hierarchical knowledge into machine learning models for a broad variety of other
270 biological problems involving phenotypic trait discovery and understanding in other taxonomic groups.

271 **Acknowledgment**

272 This work is supported by the National Science Foundation through Harnessing the Data Revolution Ideas Lab program awards
273 1940322 (Bart); 1940233 (Greenberg); 2022042 (Mabee); 1940247 (Karpatne) and 1939505 (Maga), and also a startup allocation
274 from the XSEDE (TG-DEB200005).

275 **Data Availability**

276 For this work, and in an effort to create a diverse and statistically substantial dataset, we aggregated more than 60,000 images of
277 fish specimens from five ichthyological research collections (Field Museum of Natural History <http://www.tubri.org/HDR/FMNH/>,
278 Illinois Natural History Survey <http://www.tubri.org/HDR/INHS/>, J. F. Bell Museum of Natural History (<http://www.tubri.org/HDR/JFBM/>), Ohio State University Museum of Biological Diversity (<http://www.tubri.org/HDR/OSUM/>), and the University of
279 Wisconsin-Madison Zoological Museum (<http://www.tubri.org/HDR/UWZM/>)) that participated in the Great Lakes Invasives Network
280 Project (GLIN)¹. The GLIN project is digitizing 1.73 million historical biological specimens representing 2,550 species, including
281 fishes, clams, snails, mussels, algae and plants that are potentially invasive to the Great Lakes Region of the U.S. The computer
282 code used for running our experiments is found at (Elhamod, 2021).
283

284 **Conflict of Interest**

285 The authors have no conflict of interest to declare.

286 **Authors' contribution**

287 M.E designed the methodology from machine learning perspective and conducted and analysed the experiments. K.D, A.M.M and
288 B.A pre-processed the data and helped design the experiments from phylogeny perspective. Y.B collected and labelled the data.
289 H.B, P.M and W.D critiqued the methodology and provided suggestions to improve it. They also helped in selecting the set of
290 species to work on and the general direction of research to explore. J.L and J.G helped setting up the pipeline for managing image
291 metadata and creating workflows for model deployment. A.K provided overall supervision across all tasks conducted in this work.
292 All authors contributed to writing the manuscript and gave final approval for publication.

293 **References**

294 Vaneeda Allken, Nils Olav Handegard, Shale Rosen, Tiffanie Schreyeck, Thomas Mahiout, and Ketil Malde. Fish species identification using
295 a convolutional neural network trained on synthetic data. *ICES Journal of Marine Science*, 76(1):342–349, jan 2019. ISSN 1054-3139.
296 doi: 10.1093/icesjms/fsy147.

¹ <http://greatlakesinvasives.org/>

297 Robert Arlinghaus and Steven J. Cooke. Recreational Fisheries: Socioeconomic Importance, Conservation Issues and Management Challenges.

298 *Recreational Hunting, Conservation and Rural Livelihoods: Science and Practice*, pages 39–58, 2009. doi: 10.1002/9781444303179.ch3.

299 Angela H. Arthington, Nicholas K. Dulvy, William Gladstone, and Ian J. Winfield. Fish conservation in freshwater and marine realms:

300 status, threats and management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26(5):838–857, 2016. ISSN 10990755. doi:

301 10.1002/aqc.2712.

302 Brian B Avants. *ANTsR: ANTs in R: Quantification Tools for Biomedical Images*, 2019. R package version 0.5.4.2.

303 Ricardo Betancur-R, Edward O Wiley, Gloria Arratia, Arturo Acero, Nicolas Bailly, Masaki Miya, Guillaume Lecointre, and Guillermo Orti.

304 Phylogenetic classification of bony fishes. *BMC evolutionary biology*, 17(1):162, 2017.

305 M. Chuang, J. Hwang, and C. S. Rose. Aggregated segmentation of fish from conveyor belt videos. In *2013 IEEE International Conference*

306 *on Acoustics, Speech and Signal Processing*, pages 1807–1811, 2013.

307 C. Costa, F. Antonucci, C. Boglione, P. Menesatti, M. Vandepitte, and B. Chatain. Automated sorting for size, sex and skeletal anomalies

308 of cultured seabass using external shape analysis. *Aquacultural Engineering*, 52:58–64, jan 2013. ISSN 01448609. doi: 10.1016/j.aquaeng.

309 2012.09.001.

310 J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

311 G. Ding, Y. Song, J. Guo, C. Feng, G. Li, B. He, and T. Yan. Fish recognition using convolutional neural network. In *OCEANS 2017 -*

312 *Anchorage*, pages 1–4, 2017.

313 Ellen Ditría, Michael Sievers, Sebastian Lopez-Marcano, Eric L. Jinks, and Rod M. Connolly. Deep learning for automated analysis

314 of fish abundance: the benefits of training across multiple habitats. *bioRxiv*, 2020a. doi: 10.1101/2020.05.19.105056. URL <https://www.biorxiv.org/content/early/2020/05/22/2020.05.19.105056>.

315 Ellen M. Ditría, Sebastian Lopez-Marcano, Michael Sievers, Eric L. Jinks, Christopher J. Brown, and Rod M. Connolly. Automating the

316 analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Frontiers in Marine Science*, 7:429,

317 2020b. ISSN 2296-7745. doi: 10.3389/fmars.2020.00429. URL <https://www.frontiersin.org/article/10.3389/fmars.2020.00429>.

318 Anderson Aparecido dos Santos and Wesley Nunes Gonçalves. Improving pantanal fish species recognition through taxonomic ranks in

319 convolutional neural networks. *Ecological Informatics*, 53:100977, 2019. ISSN 1574-9541. doi: <https://doi.org/10.1016/j.ecoinf.2019.100977>. URL <https://www.sciencedirect.com/science/article/pii/S1574954119301001>.

320 Elhamod. Hierarchy-guided Neural Networks for Species Classification, October 2021. URL <https://doi.org/10.5281/zenodo.5562809>.

321 David Glynne Fox and Thomas PV Hartman. Photographing fluid-preserved specimens. In *Biobanking*, pages 149–153. Springer, 2019.

322 Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic

323 segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014. doi: 10.1109/CVPR.2014.

324 81.

325 Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

326 Snigdhaa Hasija, Manas Jyoti Buragohain, and S. Indu. Fish species classification using graph embedding discriminant analysis. In

327 *Proceedings - 2017 International Conference on Machine Vision and Information Technology, CMVIT 2017*, pages 81–86. Institute of

328 Electrical and Electronics Engineers Inc., mar 2017. ISBN 9781509049936. doi: 10.1109/CMVIT.2017.23.

329 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on*

330 *Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

331 Gene Helfman, Bruce B Collette, Douglas E Facey, and Brian W Bowen. *The diversity of fishes: biology, evolution, and ecology*. John

332 Wiley & Sons, 2009.

333 Willi Hennig. *Phylogenetic systematics*. 1966.

336 Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of*
337 *the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

338 Lily C Hughes, Guillermo Ortí, Yu Huang, Ying Sun, Carole C Baldwin, Andrew W Thompson, Dahiana Arcila, Ricardo Betancur-R,
339 Chenhong Li, Leandro Becker, et al. Comprehensive phylogeny of ray-finned fishes (actinopterygii) based on transcriptomic and genomic
340 data. *Proceedings of the National Academy of Sciences*, 115(24):6249–6254, 2018.

341 Ahsan Jalal, Ahmad Salman, Ajmal Mian, Mark Shortis, and Faisal Shafait. Fish detection and species classification in underwater
342 environments using deep learning with temporal information. *Ecological Informatics*, 57:101088, 2020. ISSN 1574-9541. doi: <https://doi.org/10.1016/j.ecoinf.2020.101088>. URL <https://www.sciencedirect.com/science/article/pii/S1574954120300388>.

343 Xiaowei Jia, Jared Willard, Anuj Karpatne, Jordan Read, Jacob Zwart, Michael Steinbach, and Vipin Kumar. *Physics Guided RNNs for*
344 *Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles*, pages 558–566. 05 2019. ISBN 978-1-61197-567-3.
345 doi: 10.1137/1.9781611975673.63.

346 Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza
347 Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on*
348 *Knowledge and Data Engineering*, 29(10):2318–2331, 2017.

349 Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional
350 neural networks. *Artificial Intelligence Review*, 53(8):5455–5516, Apr 2020. ISSN 1573-7462. doi: 10.1007/s10462-020-09825-6. URL
351 <http://dx.doi.org/10.1007/s10462-020-09825-6>.

352 Ron Kikinis, Steve D. Pieper, and Kirby G. Vosburgh. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical
353 Support. In *Intraoperative Imaging and Image-Guided Therapy*, pages 277–289. Springer, New York, NY, 2014. ISBN 978-1-4614-7656-6
354 978-1-4614-7657-3. doi: 10.1007/978-1-4614-7657-3_19. URL https://link.springer.com/chapter/10.1007/978-1-4614-7657-3_19.

355 Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

356 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural
357 networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural*
358 *Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

359 Yakup Kutlu, Bilal Iscimen, and Cemal Turan. Multi-stage fish classification system using morphometry. *Fresenius Environmental Bulletin*,
360 26:1910–1916, 03 2017.

361 Rasmus Larsen, Hildur Olafsdottir, and Bjarne Kjær Ersbøll. Shape and texture based classification of fish species. In *Lecture Notes in*
362 *Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5575
363 LNCS, pages 745–749. Springer, Berlin, Heidelberg, 2009. ISBN 3642022294. doi: 10.1007/978-3-642-02230-2_76.

364 D. J. Lee, Sharon Redd, Robert Schoenberger, Xiaoqian Xu, and Pengcheng Zhan. An Automated Fish Species Classification and Migration
365 Monitoring System. In *IECON Proceedings (Industrial Electronics Conference)*, volume 2, pages 1080–1085, 2003. doi: 10.1109/IECON.
366 2003.1280195.

367 Dah Jye Lee, James K. Archibald, Robert B. Schoenberger, Aaron W. Dennis, and Dennis K. Shiozawa. Contour matching for fish species
368 recognition and migration monitoring. *Studies in Computational Intelligence*, 122:183–207, 2008. ISSN 1860949X. doi: 10.1007/
369 978-3-540-78534-7_8.

370 Xiu Li, Min Shang, Hongwei Qin, and Liansheng Chen. Fast accurate fish detection and recognition of underwater images with Fast R-CNN.
371 In *OCEANS 2015 - MTS/IEEE Washington*. Institute of Electrical and Electronics Engineers Inc., feb 2016. ISBN 9780933957435. doi:
372 10.23919/oceans.2015.7404464.

375 Abigail J. Lynch, Steven J. Cooke, Andrew M. Deines, Shannon D. Bower, David B. Bunnell, Ian G. Cowx, Vivian M. Nguyen, Joel Nohner,
376 Kaviphone Phouthavong, Betsy Riley, Mark W. Rogers, William W. Taylor, Whitney Woelmer, So Jung Youn, and T. Douglas Beard.
377 The social, economic, and environmental importance of inland fish and fisheries. *Environmental Reviews*, 24(2):115–121, 2016. ISSN
378 11818700. doi: 10.1139/er-2015-0064.

379 Robert McGill, John W Tukey, and Wayne A Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

380 Keerthiram Murugesan, Hanxiao Liu, Jaime Carbonell, and Yiming Yang. Adaptive smoothed online multi-task learning.
381 In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing
382 Systems*, volume 29, pages 4296–4304. Curran Associates, Inc., 2016. URL <https://proceedings.neurips.cc/paper/2016/file/a869ccbcb9568808b8497e28275c7c8-Paper.pdf>.

383 Thomas J. Near, Ron I. Eytan, Alex Dornburg, Kristen L. Kuhn, Jon A. Moore, Matthew P. Davis, Peter C. Wainwright, Matt Friedman,
384 and W. Leo Smith. Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences
385 of the United States of America*, 109(34):13698–13703, 2012. ISSN 00278424. doi: 10.1073/pnas.1206625109.

386 S O Ogunlana, O Olabode, S A A Oluwadare, and G B Iwasokun. Fish Classification Using Support Vector Machine. Technical Report 2,
387 2015. URL www.ajocict.net.

388 Misa Ogura and Ravi Jain. Flashtorch. <http://doi.org/10.5281/zenodo.3596650>, 2020.

389 Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):
390 62–66, 1979.

391 Dhruv Rathi, Sushant Jain, and S. Indu. Underwater Fish Species Classification using Convolutional Neural Network and Deep Learning.
392 In *2017 9th International Conference on Advances in Pattern Recognition, ICAPR 2017*, pages 344–349. Institute of Electrical and
393 Electronics Engineers Inc., dec 2018. ISBN 9781538622414. doi: 10.1109/ICAPR.2017.8593044.

394 Hafiz Tayyab Rauf, Muhammad Ikram Lali, Saliha Zahoor, Syed Zakir Shah, Abd Rehman, and Syed Ahmad Chan Bukhari. Visual features
395 based automated identification of fish species using deep convolutional neural networks. *Computers and Electronics in Agriculture*, 11
396 2019. doi: 10.1016/j.compag.2019.105075.

397 Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE
398 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. doi: 10.1109/CVPR.2016.91.

399 Marco T.A. Rodrigues, Flávio L.C. Pádua, Rogério M. Gomes, and Gabriela E. Soares. Automatic fish species classification based on robust
400 feature extraction techniques and artificial immune systems. In *Proceedings 2010 IEEE 5th International Conference on Bio-Inspired
401 Computing: Theories and Applications, BIC-TA 2010*, pages 1518–1525, 2010. ISBN 9781424464388. doi: 10.1109/BICTA.2010.5645273.

402 Ahmad Salman, Shoaib Ahmad Siddiqui, Faisal Shafait, Ajmal Mian, Mark R Shortis, Khawar Khurshid, Adrian Ulges, and Ulrich
403 Schwanecke. Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. *ICES
404 Journal of Marine Science*, 77(4):1295–1307, 02 2019. ISSN 1054-3139. doi: 10.1093/icesjms/fsz025. URL <https://doi.org/10.1093/icesjms/fsz025>.

405 Mehmet Sarigul. Comparison of different deep structures for fish classification. *International Journal of Computer Theory and Engineering*,
406 9, 10 2017. doi: 10.7763/IJCTE.2017.V9.1167.

407 Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual
408 explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*,
409 pages 618–626, 2017. doi: 10.1109/ICCV.2017.74.

410 Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

411 Shoaib Ahmed Siddiqui, Ahmad Salman, Muhammad Imran Malik, Faisal Shafait, Ajmal Mian, Mark R Shortis, and Euan S Harvey.
412 Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited

415 labelled data. *ICES Journal of Marine Science*, 75(1):374–389, jan 2018. ISSN 1054-3139. doi: 10.1093/icesjms/fsx109.

416 Carlos N Silla and Alex A Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge*
417 *Discovery*, 22(1-2):31–72, 2011.

418 Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*
419 *arXiv:1409.1556*, 2014.

420 Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and
421 saliency maps. *CoRR*, abs/1312.6034, 2014.

422 C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. Fisher. Detecting, tracking and counting fish in low quality unconstrained
423 underwater videos. In *Proc. 3rd Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 514–519, 2008.
424 ISBN 978-989-8111-21-0.

425 Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin A. Riedmiller. Striving for simplicity: The all convolutional net.
426 In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA,*
427 *USA, May 7-9, 2015, Workshop Track Proceedings*, 2015. URL <http://arxiv.org/abs/1412.6806>.

428 C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with
429 convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

430 Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International*
431 *conference on artificial neural networks*, pages 270–279. Springer, 2018a.

432 P.N. Tan, M. Steinbach, A. Karpatne, and V. Kumar. *Introduction to data mining (2nd edition)*. Pearson Addison Wesley Boston, 2018b.

433 Nicholas J. Tustison, Zixuan Lin, Xue Feng, Nicholas Cullen, Jaime F. Mata, Lucia Flors, James C. Gee, Talissa A. Altes, John P. Mugler III,
434 and Kun Qing. Convolutional neural networks with template-based data augmentation for functional lung image quantification. *Academic*
435 *Radiology*, 2018. URL <https://www.ncbi.nlm.nih.gov/pubmed/30195415>.

436 Sébastien Villon, David Mouillot, Marc Chaumont, Gérard Subsol, Thomas Claverie, and Sébastien Villéger. A new method to control
437 error rates in automated species identification with deep learning algorithms. *Scientific Reports*, 10:10972, 07 2020. doi: 10.1038/
438 s41598-020-67573-7.

439 Sébastien Villon, Corina Iovan, Morgan Mangeas, Thomas Claverie, David Mouillot, Sébastien Villéger, and Laurent Vigliola. Automatic
440 underwater fish species classification with limited data using few-shot learning. *Ecological Informatics*, 63:101320, 2021. ISSN 1574-9541.
441 doi: <https://doi.org/10.1016/j.ecoinf.2021.101320>. URL <https://www.sciencedirect.com/science/article/pii/S1574954121001114>.

442 Yaqing Wang, Quanming Yao, James Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning, 2020.

443 Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In
444 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

445 Hong Yao, Qingling Duan, Daoliang Li, and Jianping Wang. An improved K-means clustering algorithm for fish image segmentation.
446 *Mathematical and Computer Modelling*, 58(3-4):790–798, 2013. ISSN 08957177. doi: 10.1016/j.mcm.2012.12.025.

447 Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Proceedings of*
448 *the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, page 3320–3328, Cambridge, MA,
449 USA, 2014. MIT Press.

450 Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*,
451 26(8):1819–1837, 2013.

Table 1. Statistics of the subsets of the GLIN dataset used in this study for training and evaluation.

Dataset	# of images	# of species	# of genera	# of images per species
GLIN (All)	63758	575	187	1 to 7935
Hard	4882	102	26	30 to 50
Easy/100	3762	38	11	63 to 100
Easy/50	1900	38	11	50

Table 2. Average probability of the correct species class predicted by **Blackbox-NN** and **HGNN** over **Easy/50**, as a function of the number of adversarial occlusions applied to every image. From left to right, we start with non-occluded images and progressively add more patches of occlusions.

Model	Number of Occlusion Patches				
	0	1	2	3	4
Blackbox-NN	0.473	0.355	0.291	0.232	0.187
HGNN	0.482	0.369	0.307	0.256	0.215