Fine-Grained Fish Species Image Classification

Ricardo J.M. Veiga
rjveiga@ualg.pt
João M.F. Rodrigues
jrodrig@ualg.pt

LARSyS & ISE, Universidade do Algarve,
8005-139 Faro, Portugal

Abstract

Fine-grained fish species classification is especially important for ecological studies and fisheries management, as it helps with ecosystem evaluation, environmental monitoring, and biodiversity conservation. Manual fish species identification, which was historically the primary method, is difficult and time-consuming, while deep learning techniques offered the possibility of automation and improvement in both efficiency and accuracy. This work investigates the usage of the Swin Transformer in conjunction with a novel plug-in module for fine-grained fish identification (classification), resulting in promising results. While the present method does not yet outperform the state-of-the-art directly, it obtains 96.20% accuracy on Croatian fish dataset, and 100% accuracy when the top three predictions are considered.

1 Introduction

Image classification and fine-grained image classification are computer vision tasks that are related but separate. The former is concerned with broader image labeling, whereas the fine-grained is focused with distinguishing extremely similar objects or species within the categories.

Fine-grained taxonomy of fish species is critical for ecological studies and fisheries management. It aids in the assessment of aquatic ecosystem health, the tracking of environmental changes, and the preservation of biodiversity in ecology. It ensures sustainable practices, prevents overfishing, and ensures compliance with legislation in fisheries management. Manual fish species identification, is not a trivial issue, as can be seen in Fig. 1, however, is the primary method for fish identification, it is time-consuming, error-prone, and difficult due to the diversity of species and minute variances that necessitate specialist expertise.

In this paper, we focused on the use of deep learning (DN), Swin Transformer, to automate and improve fish species classification, increasing efficiency and accuracy. With this purpose, we present and adaptation a previous model used to (fine-grained) birds classification to classify fish.

2 Literature Review

For this study it was used the Croatian fish dataset (CFD), introduced by Jäger et al. [2]. It is a small-scale fine-grained fish classification dataset accompanied by a baseline results obtained using Convolutional Neural Networks (CNN) and a linear Support Vector Machine (SVM). The authors work [2] emphasized the need for specialized datasets in this domain, setting the stage for subsequent research.

Even so, in the literature there are several models presented for fish classification, namely, Qiu et al. [6] improved the transfer learning with Bilinear-CNNs (BCNN). Their approach significantly improved classification accuracy, highlighting the importance of leveraging pre-existing knowledge. Later, the same authors [7] extended their work exploring the use of Squeeze-and-Excitation (SE) networks, emphasizing the role of feature engineering in enhancing the discriminative power of models for fine-grained fish classification. Okafor et al. [4] explored the importance of data augmentation methods, investigating the images rotation and color constancy effects across intra-class variations and limited data.

Figure 1: Samples of CFD [2], illustrating the diversity of fish species and the challenges of fine-grained classification.

Semi-supervised learning techniques, demonstrating the effectiveness of incorporating unlabeled data to enhance classification performance, through the use of modified Deep Convolutional Generative Adversarial Networks (DCGAN), was used by [11]. A Genetic Programming (GP) approach to discern between low-quality images in order to effectively extract meaningful features, applying restoration operators to the most affect areas of the images was presented in [10]. Pang et al. [5] applied Knowledge Distillation (KD) through Feature Similarity Alignment (FSA) and used a teacher-student model, lowering the noise in the raw data with KL-Divergence, achieving the highest accuracy on the small-scale Croatian Dataset. Finally, more recent models such as [9] conjoined a refined CNN with SVM and additional SE blocks into a Hybrid DeepCNN, further improving the accuracy on the small-scale dataset through post-training after initializing with a different fish dataset. In [8] the authors improved the quality of the raw images prior training using Underwater Image Enhanced Generative Adversarial Network (UIEGAN), requiring only smaller VGG and RESNET models while achieving higher accuracy on the small-scale datasets.

All the previous methods relied on a CNN architecture approach, progressively increasing the accuracy spanning from data augmentation, transfer learning, feature engineering, semi-supervised learning, or by combining different techniques.

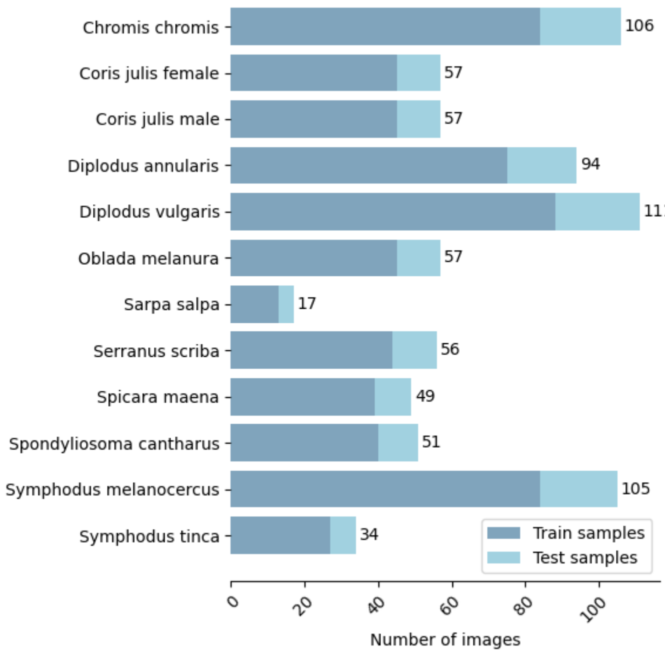

Figure 2: Fishes' species sampling distribution of the CFD [2].

3 Methodology

While CNNs have traditional dominated image-related tasks, recent advances with Transformer [3], which were originally build for sequential data, show promise in fine-grained categorization. Therefore, in this study we explored the use of the Swin Transformer, through a novel plug-in module presented by Chou et al. [1], that used to classify birds. These authors used two fine-grained bird identification datasets, namely CUB200-2011 and NABirds (see details in [1]). Here, we explore the same principles but for the FGIC challenge in the small-scale Croatian fish dataset.

The CFD comprises on the distribution of 794 variable low-quality images across 12 fish species, with some being very similar, as can be observed in Fig. 1. This distribution is unbalanced, as can be verified in Fig. 2, further increasing the FGIC challenge. It is possible to observe in the same figure the 0.8:0.2 training and test samples split.

The plugin-in module [1] (PIM) used to train the model from scratch for this study focused on exploring the main difference between coarse-grained classification versus fine-grained classification, with the former involving dissimilar categories, and the latter dealing with highly similar categories. This task often requires expert labeling, or the use of automatically discriminative regions identification, which often involves multi-stage architectures. This module produces pixel-level feature maps and integrates filtered features to improve fine-grained visual classification.

Although multiple CNNs can be used as backbone with PIM, we chose the Swin Transformer [3]. The results obtained can be reproduced using the seed 42. We trained for 20 epochs with Stochastic Gradient Descent as the optimization algorithm, automatic mixed precision, the raw data resized to 384×384 pixels, with a batch size of 16.

Figure 3: Samples of the output heatmaps obtained with PIM.

4 Results & Discussion

As described in section 2, different data augmentation approaches, CNN architecture variations and multi-modal models improved the accuracy on the small-scale Croatian fish dataset, as can be compared in Tab. 1.

The work presented is an initial approach, although the result obtained (96.20%) in this study does not yet surpass the current state-of-the-art achieved by [5] (99.22%), when considering only the combiner top 1 results. If we extend the results range to the combiner top 3, we achieved 100.0% through multiple experiments with different seeds.

When used on this dataset, our method, which was adapted from [1], produced intriguing findings. Notably, in Fig. 3 in the lower left and center images, the heatmaps show slight variations in the pertinent traits that discriminate between female and male *Coris Julis*.

Author(s)	Model/Architecture	Accuracy
Jäger et al. [2] (baseline)	*CNN + SVM*	*66.78%*
Qiu et al. [6]	B-CNN	81.30%
Okafor et al. [4]	Fine-tuned CNN	82.18%
Zhao et al. [11]	DCGAN	82.90%
Qiu et al. [7]	B-CNN+refined SE Blocks	83.92%
Yan et al. [10]	Genetic Programming	87.90%
Veluswami et al.[9]	CNN+Refined SE+SVM	94.99%
Sudhakara et al. [8]	UIEGAN+DCGAN+VGG	95.64%
Our[1]	**Swin Transformer**	**96.20%**
Pang et al. [5]	FSA+KL-Divergence	99.22%

Table 1: Comparison of accuracy results.

5 Conclusion & Future Work

This study introduced the Transformers to the field of fine-grained fish classification, with the baseline obtaining results surpassing almost the current state-of-the-art. Nevertheless, these results were achieved without performing data augmentation to the raw data, allowing space to improve in future work.

With Pang et al. [5] achieving the highest accuracy performing data regeneration on the variable low-quality small-scale, the next step in our research will focus on automatically adapting the input raw data to further express the relevant fine features, as well as, concentrate on data augmentation and data regeneration.

Acknowledgments

This work was financed by the Portuguese Foundation for Science and Technology (FCT), FCT - PhD grant 2022.11602.BD and project LARSyS - FCT Project UIDB/50009/2020.

References

[1] Po-Yung Chou, Cheng-Hung Lin, and Wen-Chung Kao. A novel plug-in module for fine-grained visual classification. *arXiv preprint arXiv:2202.03822*, 2022.

[2] Jonas Jäger, Marcel Simon, Joachim Denzler, Viviane Wolff, Klaus Fricke-Neuderth, and Claudia Kruschel. Croatian fish dataset: Fine-grained classification of fish species in their natural habitat. *Swansea: Bmvc*, 2, 2015.

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[4] Emmanuel Okafor, Lambert Schomaker, and Marco A Wiering. An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals. *Journal of Information and Telecommunication*, 2(4):465–491, 2018.

[5] Jian Pang, Weifeng Liu, Baodi Liu, Dapeng Tao, Kai Zhang, and Xiaoping Lu. Interference distillation for underwater fish recognition. In *Asian Conference on Pattern Recognition*, pages 62–74. Springer, 2021.

[6] Chenchen Qiu, Jinna Cui, Shaoyong Zhang, Chao Wang, Zhaorui Gu, Haiyong Zheng, and Bing Zheng. Transfer learning for small-scale fish image classification. In *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–5. IEEE, 2018.

[7] Chenchen Qiu, Shaoyong Zhang, Chao Wang, Zhibin Yu, Haiyong Zheng, and Bing Zheng. Improving transfer learning and squeeze-and-excitation networks for small-scale fine-grained fish image classification. *IEEE Access*, 6:78503–78512, 2018.

[8] M Sudhakara, M Janaki Meena, K Reddy Madhavi, P Anjaiah, et al. Fish classification using deep learning on small scale and low-quality images. *International Journal of Intelligent Systems and Applications in Engineering*, 10(1s):279–288, 2022.

[9] Jansi Rani Sella Veluswami and Panneerselvam Nivetha. Multi-species fish identification using hybrid deepcnn with refined squeeze and excitation architecture. *Aquatic Sciences and Engineering*, 37 (4):220–228, 2022.

[10] Zichu Yan, Ying Bi, Bing Xue, and Mengjie Zhang. Automatically extracting features using genetic programming for low-quality fish image classification. In *2021 IEEE Congress on Evolutionary Computation (CEC)*, pages 2015–2022. IEEE, 2021.

[11] Jian Zhao, Yihao Li, Fengdeng Zhang, Songming Zhu, Ying Liu, Huanda Lu, and Zhangying Ye. Semi-supervised learning-based live fish identification in aquaculture using modified deep convolutional generative adversarial networks. *Transactions of the ASABE*, 61(2): 699–710, 2018.

[1] adaptation of the model proposed by Chou et al. [1] for bird classification.