FisHook - 使用 MobileNetV2 进行海洋物种分类的优化方法

Kohav Dey计算机科学与工程系，*PES*大学，*Banashankari*班加罗尔，印度
deykohav@gmail.com

Krishna Bajaj计算机科学与工程系，*PES*大学，*Banashankari*班加罗尔，印度
bajaj.krishna01062002@gmail.com

K SRamalakshmi计算机科学与工程系，*PES*大学，*Banashankari*班加罗尔，印度
rlsri2305@gmail.com

Samuel Thomas电子与通信工程系，*PES*大学，*Banashankari*班加罗尔，印度
samuelthomas1049@gmail.com

Sriram Radhakrishna计算机科学与工程系，*PES*大学，*Banashankari*班加罗尔，印度
sriram.radhakrishna42@gmail.com

arXiv:2304.01524v1 [cs.CV] 4 Apr 2023

摘要—**海洋生态系统对地球的健康至关重要，但气候变化、污染和过度捕捞等人类活动持续威胁着海洋物种。准确分类和监测这些物种，有助于了解其分布、种群动态以及人类活动对它们的影响。然而，由于海洋物种种类繁多，且水下环境复杂，对其进行分类可能颇具挑战。随着计算机性能和基于 GPU 的计算不断进步，深度学习算法如今能够高效地对海洋物种进行分类，从而更便于监测和管理海洋生态系统。本文提出对 MobileNetV2 模型进行优化，并重点说明创建数据集和扩增海洋物种图像的具体指南，使模型达到 99.83% 的平均验证准确率。该迁移学习算法可成功部署于移动应用程序，以便在渔场现场进行分类。**

索引词—**海洋、MobileNetV2、分类、迁移学习、数据集扩增、移动应用程序**

I. 引言

鱼类及其他海鲜已成为全球饮食的重要组成部分，约有 50% 的人口已成为其经常性消费者。此外，根据联合国粮食及农业组织的数据，全球海鲜人均年消费量已超过 20 千克 [1]。面对多变量海洋渔获，需要一种高效的分类算法，用于识别具有食用性限制和特定营养成分的新亚种，同时还需要一个数据库来收录这些物种，供其他鱼类养殖爱好者使用。有效分类这些物种有助于识别较为稀有的近缘物种；结合市场调研，还能促进明智决策，从而确保销售获利并监测海洋物种。定期监测有助于我们了解海洋物种的分布、种群动态以及人类活动的影响，从而制定有效的保护策略。当前使用的传统分类方法则极其耗费人力、容易出错且耗时。



图 1. 本文介绍的模型已部署在一款配备成熟后端的 Android 移动应用程序中。

该提案旨在采用基于图像的方法开发海洋物种分类器，并通过

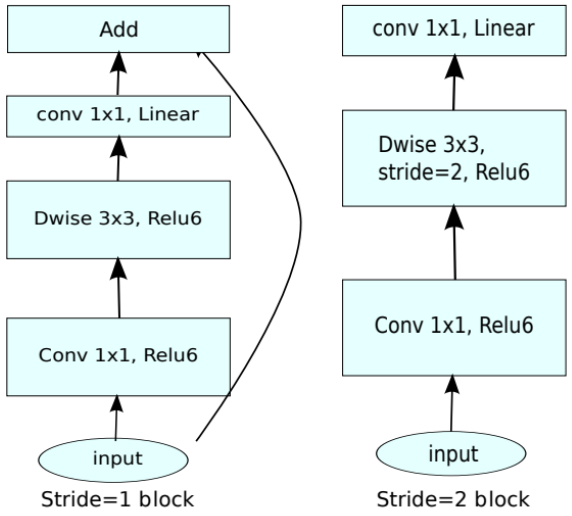


图 2. MobileNetV2 架构。

迁移学习实现。为此，我们将采用 MobileNetV2 [2]架构，该架构已被证明在多种图像分类任务中效果显著。此外，我们还将结合传统的序列神经网络，进一步提高模型的准确率。用于训练和测试模型的数据集包含九种不同的鱼类类别 [3], , 每种类别都有其独特的特征。通过迁移学习，我们将使用以 ImageNet [4] 权重初始化的预训练 MobileNetV2 模型作为特征提取器。这将使模型能够利用 MobileNetV2 架构已有的知识，提高数据集的分类准确率。该海洋物种分类器的主要目标是准确识别并区分数据集中的各种鱼类类别。该分类器可用于多种海洋应用，例如监测和保护工作，以便更好地了解不同鱼类物种的分布和行为。

II. 相关工作

目前，已有多种在不同数据集上训练和测试的机器学习模型投入使用，以达到可接受的分类准确率。Ulucan 等人 [3] 采用基于特征的分类模型，使用支持向量机，在 9 个不同类别上取得了 98.01% 的平均训练准确率和 88.69% 的平均验证准确率。Rossi 等人 [5] 构建了一款配备远程服务器的移动应用程序，用于基于云的机器学习；该应用程序采用人工神经网络分类器，在 7 个不同的鱼类类别上达到 100% 的准确率。Cchabra 等人 [6]构建了一个基于 VGG16 迁移学习的模型，在 8 种不同鱼类上达到 93.8% 的准确率。Liu 等 [7]提出了一种基于 MobileNetV2 的迁移学习模型，在 7 类鱼类上达到 95% 的准确率，鱼类。该优化方法通过提出数据集创建与扩增的指导原则，以及对现有模型进行修改以提高模型准确率，为后续工作提供了参考。

这项文献综述表明，基于 MobileNetV2 的深度学习模型已成功应用于鱼类物种识别与分类任务，并取得了较高的准确率。本综述回顾的研究采用了迁移学习、数据增强、特征提取和混合模型等不同技术，以提升基于 MobileNetV2 的模型性能。本研究论文可在现有文献的基础上，提出一种使用 MobileNetV2 进行鱼类物种识别的新方法，并探索不同技术提升模型准确率的潜力。

III. 概念理论与实现

海洋生物对地球生态系统至关重要，能够提供食物和氧气，并调节气候变化。目前，无数物种持续受到气候变化、污染、过度捕捞及其他人类活动的威胁。准确地分类和监测海洋物种，有助于我们了解其分布、种群动态以及人类活动对它们的影响，从而制定有效的保护策略。由于物种种类繁多，且水下环境复杂，海洋物种分类是一项艰巨的任务。随着计算机性能的提升以及基于 GPU 的计算技术的进步，如今已能够利用深度学习算法处理大量数据，从而准确、高效地分类海洋物种，使海洋生态系统的监测和管理更加容易。本文提出了一种优化方案，旨在提升最准确的图像分类模型在海洋物种分类方面的表现。此前，研究人员已使用 MobileNetV2 模型评估这一问题，并取得了 97.26% 的显著平均分类准确率 [8]。MobileNetV2 是一种神经网络架构，在包括图像分类、目标检测和语义分割在内的多种任务中均达到了最先进的性能。此外，MobileNetV2 的计算效率很高，在保持相近准确率的同时，速度最高可比其前代模型 MobileNetV1 快 3.4 倍。然而，我们发现，通过修改用于训练的数据集类型，可以显著改进该模型。通过明确提出此类数据集的构建指南，我们将平均验证准确率提升至 99.83% ；同时，研究还发现，将同一批海洋物种图像的增强版本加入数据集，并与改进后的模型配合使用，反而会降低模型的有效性。我们采用迁移学习算法，在渔场部署了用于现场分类的方案，并成功将其应用于移动应用程序。

A. 数据集

我们使用由 9 种海洋动物组成的数据集，训练了 Ulucan 等人实现的分类器的一个修改版本 [3] 。其中七个类别——红鲻、金头鲷、竹荚鱼、海鲈、真鲷、黑海西鲱和条纹红鲻——各包含 50 张不同的鱼类图像；其余两个类别——鳟鱼和虾——各包含 30 张不同的图像。该数据集及采集的图像具有足够的多样性：图像从相对于物体表面和自身身体轴线的多个角度拍摄，因而能够呈现对象可能具有的各种形态细节。图像尺寸为 590 x 445 像素。随后，我们将图像输入数据增强算法，对每张图像进行随机角度旋转且不重复，并进行翻转，同时保持图像尺寸一致。数据集的收集和增强完成后，每种海鲜约包含 1,100 张图像。

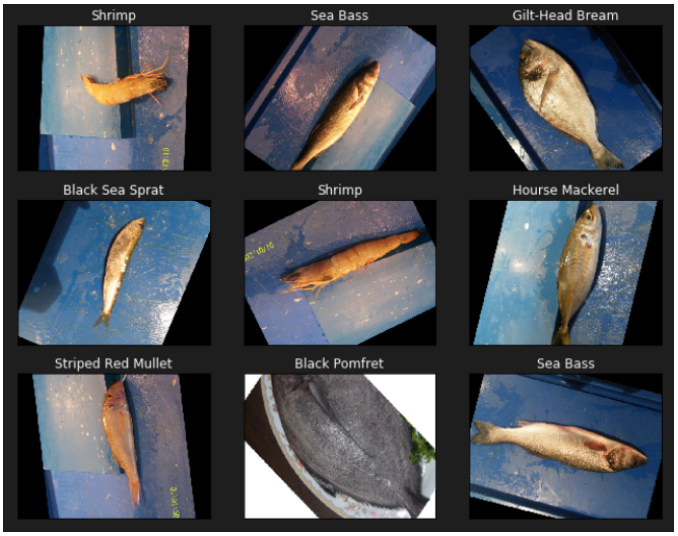


图 3. 数据增强后训练数据集中的示例图像及其标签。

为构建分类器，采用了多种方法。任何需要作为应用程序在边缘设备上运行的基于 TensorFlow 的模型，都需要转换为压缩的 .tflite 模型。在所有模型中，MobileNetV2 的预测结果颇具前景；因此，将 Tensorflow[9] 模型转换为 .tflite [10] 模型格式，以便部署到 Flutter 应用程序中。本文介绍了上述方法的端到端实现，简要概述了发现的各种不足，并探讨了未来的改进方向。

B. 图像预处理

数据集通过 Image- DataGenerator [11] 进行增强，该工具可生成张量图像数据批次，并支持实时数据增强 [12]。训练数据集中的数据增强和预处理通过 0.2 的验证集划分隐式完成。预处理输入 API 会将



图 4. 模型训练所用的原始数据集 [3]。

输入像素缩放到 -1 到 1 之间。不同维度的张量会被重塑为 224X224 像素，作为 MobileNetV2 网络的合适输入，用于特征提取，随后由卷积神经网络进行分类。

C. 迁移学习

迁移学习 [13] 是一种机器学习技术，旨在将预训练模型重新用于相似的深度学习任务，因为训练这类复杂的神经网络需要海量数据、庞大的计算资源和大量时间。由于大多数现实问题缺乏足够的带标签数据点用于训练，这种优化方法能够以相对较少的数据训练深度神经网络，同时缩短训练时间并提升模型性能。

基础网络在基础数据集上针对特定任务进行训练后，其学到的特征会被重新利用或迁移到第二个目标网络中；该目标网络使用不同的目标数据集针对另一项任务进行训练，而该数据集与原始数据集具有相似的高层特征。当模型学到的特征对基础数据集和目标数据集都具有通用性时，迁移学习的效果会更好。这种迁移学习称为归纳迁移。

D. MobileNetV2 - 概述

谷歌推出的 MobileNetV2 [2], 是一种卷积神经网络架构，旨在边缘设备上高效运行。它采用倒置残差结构，其中残差连接位于瓶颈层之间，这是对 MobileNetV1 的改进。MobileNetV1 使用深度可分离卷积，也降低了复杂度开销。中间扩展层采用轻量级深度卷积来筛选特征，为窄层提供非线性。

总体而言，本实现采用 MobileNetV2 架构进行特征提取，以生成特征图，作为传统卷积神经网络全连接 Dense层的输入。该架构包含一个具有 32 个滤波器的初始全卷积层，后接 19 个残差瓶颈层。添加瓶颈层是为了减少网络中的特征图数量，否则特征图数量往往会随层数增加而增多。具体方法是使用 1x1 卷积，

其输出通道少于输入通道，从而降低模型复杂度，使模型更轻量，能够部署在边缘设备（如智能手机）上。

E. 模型架构

本文提出一种基于迁移学习的实现方案，使用MobileNetV2 的全卷积预训练层进行特征提取，并使用带有全连接 Dense层的传统神经网络进行分类。MobileNetV2 在输入层接收大小为 224 x 224x 3 的张量。顶层可训练层的权重被冻结，生成的特征图作为输入送入神经网络的全连接Dense层。网络的隐藏层通过 RELU 激活函数引入非线性，从而可以堆叠多个神经元层，构建深度神经网络。非线性有助于网络学习复杂的数据集，并实现较高的准确率。

$$Relu(x) = max(0, x) \tag{1}$$

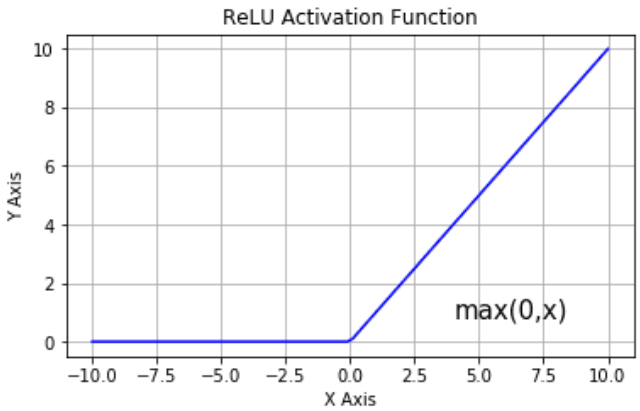


图 5. RELU 激活函数参考图。

图 5 展示了 RELU 函数（公式 1）的图形表示：对于非零输入值，函数输出与输入相同；所有负值均被置为零。

神经网络中的每一层都使用相应的损失函数进行训练，以量化模型在训练阶段的表现。它衡量给定输入的预测输出与实际输出之间的差异，并用于在后续训练迭代中优化模型参数。最小化损失函数可以显著提高模型的准确率，使其更有效地完成预定任务。

分类交叉熵损失 [14]是深度学习中多分类问题常用的损失函数。

$$-\left(y\log(p) + (1-y)\log(1-p)\right) \tag{2}$$

每个观测值对应一个类别标签时，其损失由公式 3 给出。

$$-\sum_{c=1}^{M} y_{o,c}\log(p_{o,c}) \tag{3}$$

如果 M 大于 2（即多分类），则分别计算每个观测值对应的各类别标签的损失，并将结果累加。所得总和即为分类器的分类交叉熵损失（公式 4）。

最终输出层使用Softmax激活，以便更好地解释分类结果，并将NumPy数组中每个类别的预测值缩放至0或1。采用基于MobileNetV2的迁移学习对不同鱼类物种进行分类。由于基础神经网络已从ImageNet数据集中学到了对鱼类分类所需的特征，因此该方法能够取得令人满意的结果。

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad for\ i = 1, 2, \ldots, K \tag{4}$$

公式2从数学上表示了Softmax[15] 函数，该函数会对多分类器中各类别的预测概率进行归一化。

在训练阶段，衡量准确率是评估深度学习模型在训练数据集和验证数据集上性能的重要组成部分。以下是训练期间常用的模型性能评估指标：

1.训练准确率：衡量模型在训练过程中的训练数据集上的准确率。训练准确率以训练数据集中被正确分类的实例所占百分比来表示。

$$Accuracy_{training} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{5}$$

2. 训练损失：衡量训练过程中训练数据集上的预测输出与实际输出之间的误差或差异。训练损失用于优化模型参数，以提升模型性能。

$$L_{train} = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \hat{y}_i) \tag{6}$$

3. 验证准确率：衡量模型在训练过程中的验证数据集上的准确率。验证准确率可用于估计模型在新数据上的表现。

$$Accuracy_{valdation} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i) \tag{7}$$

4. 验证损失：它衡量训练过程中模型在验证数据集上的预测输出与实际输出之间的误差或差异。验证损失用于监控过拟合，即模型在训练数据集上表现良好，但在验证数据集上表现不佳。

$$L_{val} = \frac{1}{m}\sum_{i=1}^{m} L(y_i, \hat{y_i}) \tag{8}$$

在训练过程中，我们的目标是尽量降低训练损失、提高训练准确率，同时避免过拟合。验证准确率和验证损失用于监控模型的性能并防止过拟合。如果验证损失上升而训练损失下降，则表明模型出现了过拟合，需要进行调整。

IV. 部署

将移动应用程序用于数据收集，并结合基于 AWS/GCP 的部署来进行处理和分类，是构建海洋物种分类器的一种很有前景的方法。本节将讨论基于 Flutter 的移动应用程序的设计与实现，以及它与云服务器的集成。

该跨平台移动应用程序使用 Flutter 开发，旨在实现易用、易访问且可定制。用户可以通过该应用程序上传或拍摄海洋物种的图片，随后图片会发送至 Firebase 共享存储，EC2 实例再从中获取图片进行处理和分类 [16]。

AWS 流水线旨在实时高效地处理大量数据。该算法通过 AWS EC2 服务部署；这是一种运行于云端的弹性计算单元，在专用自动化脚本的辅助下，可提供可扩展、可靠且经济高效的模型训练与部署。随着用户不断提交用于分类的数据，模型也会持续利用新数据更新，确保算法长期保持准确且适用。

以这种方式对流程进行流水线化，可以在共享的 Firebase 存储（在此实现中用作存储桶） [17]的交换节点 [18] 处分配训练过程产生的计算负载，由多台远程计算机共同承担。这将大幅提高吞吐量上限，并通过确保可扩展性来支持移动应用程序不断扩大的用户群。

该移动应用程序为用户提供了一种简单便捷的数据收集方式，而云端流水线则确保了对这些数据进行高效处理和分类。将基于 Flutter 的移动应用程序与 AWS 集成，代表了一种强大且创新的海洋物种分类方法。通过实现更准确、高效的数据收集与分类，这种方法有望助力更有效地开展海洋保护工作。

V. 结果

A. 测试规范

TensorFlow 是谷歌开发的一款广受欢迎的开源库，用于机器学习和人工智能。本实验还使用了一个名为 Keras [19] 的辅助软件包，以进一步实现代码抽象。用于训练该

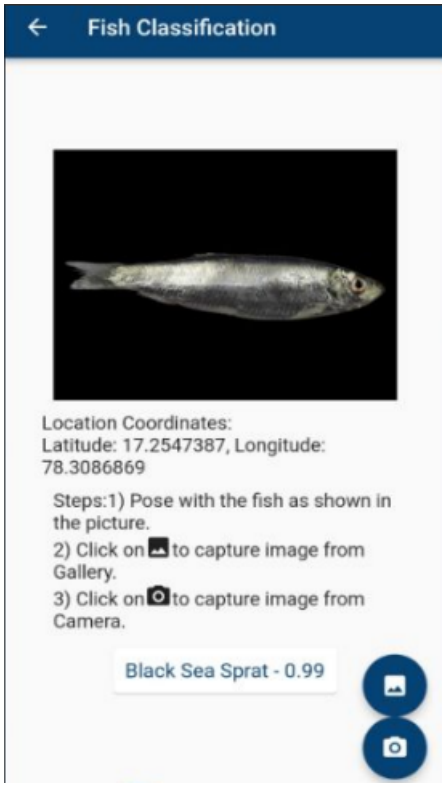


图 6. 模型准确评估的数据集中的图像示例。

模型并在部署前进行训练的实验环境是一台搭载 Ubuntu 20.04 的计算机，配置包括 8 核 16 线程的 AMD Ryzen 9 5900S 处理器、16GB DDR4 内存、1TB 磁盘存储，以及配备 6GB 显存的独立 NVIDIA GeForce RTX 3060 GPU。

B. 训练过程

模型使用分类交叉熵损失函数来找出学习过程中的误差或偏差，并使用Adam优化器通过比较预测结果与损失函数来优化输入权重，该优化器应用于训练过程。

训练准确率、训练损失、验证准确率和验证损失是用于评估模型性能的指标。神经网络以32的批次大小训练了50个训练轮次，并使用提前停止回调；当模型的验证损失增加时，训练过程便会终止。特征提取由MobileNetV2模型的预训练层完成，随后附加了一个由两个全连接Dense层组成的人工神经网络，以执行多类别鱼类分类。模型训练耗时312秒406毫秒（5.223分钟），共进行了3个训练轮次，每轮约耗时104秒468毫秒。

C. 训练与验证结果

供参考，训练准确率、训练损失、测试准确率和测试损失已在表中进行定量分析

1.

结果				
模型	训练准确率	训练 Loss	验证准确率	验证 Loss
InceptionV3 and 自定义未增强基于数据集水下	99.73%	0.0072	91.13%	0.372
MobileNetV1 and 自定义未增强数据集 基于水下	99.29%	0.0199	89.92%	0.5113
MobileNetV2 and 自定义未增强基于数据集水下	99.58%	0.0085	92.89%	0.3165
MobileNetV2 with ANN and 自定义未增强基于数据集外部水体	99.99%	0.0021	99.83%	0.0051

表 I 测试结果的定量分析。

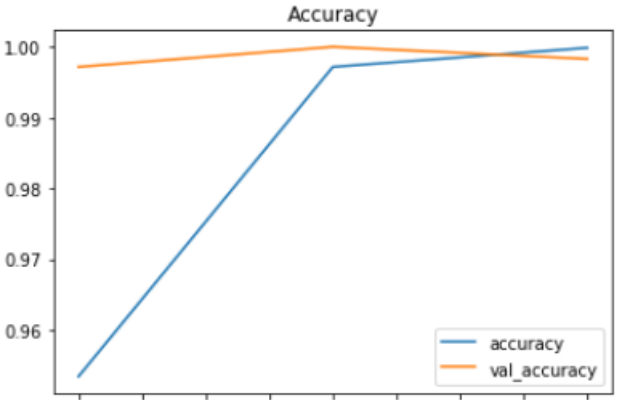


图 7. 训练阶段和验证阶 段的准确率曲线 ase.

从图 7 和图 8 的曲线可以看出，随着网络训练过程中迭代次数的增加，验证准确率和损失也随之上升，并稳定在一定区间内。根据轨迹的特征可以推断，模型使用适当的超参数进行了高效训练，从而获得了可接受的准确率和损失，且过拟合程度极低。

与 Liu 等的结果相比，基于卷积神经网络模型获得的平均准确率为 99.83% [7] ，

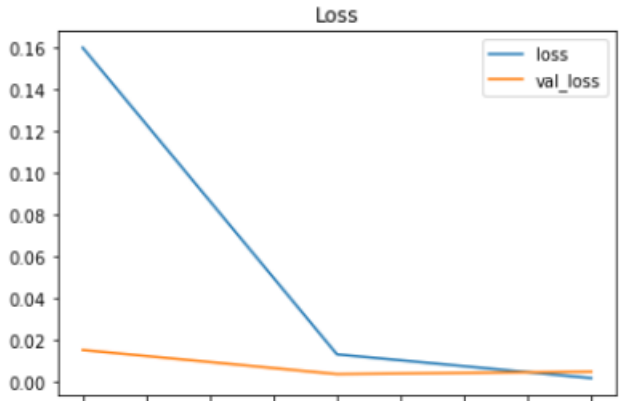


图 8. 训练阶段和验证阶段的损失曲线。

基于 MobileNetV2 架构的模型优于采用支持向量机分类器、并在同一数据集上训练的模型，后者的平均准确率为 98.01%。随后，训练好的模型被转换为一种高效、便携的 FlatBuffers 格式，即 TensorFlow Lite 格式（.tflite），以便部署在边缘设备上。该模型被导入跨平台 Flutter 框架应用中，用于预测一张未见过的图像（图 6）；在此示例中，模型将鱼类识别为黑海西鲱，置信度为 0.99。

值得注意的是，当使用前面提到的增强数据集训练同一模型时，模型的验证准确率远低于 50%。这是因为 Ulucan 等人提供的原始数据集 [3] 本身已隐含地进行了增强，其中包含从多个角度拍摄的同一条鱼的图像，使数据集本身能够有效捕获鱼类整体的独特低级特征 [20] 。因此，进一步增强向数据集中引入了大量噪声，导致网络在整个训练阶段不断调整权重，却未能显著收敛。

D. 部署准确率

尽管该模型在 Ulucan 等人提供的数据集上的训练和测试取得了令人鼓舞的结果， [3], 但需要指出的是，即使遇到从未出现在该数据集中的图像，以及含有额外噪声的图像，模型也能做出准确预测。此类示例见图 9。

从这个演示中可以看出，模型评估的图像中含有大量噪声，包括一名手持金头鲷的人，以及尾部关节和鳃下方区域受到的遮挡。

此时可以说，模型成功做出了一次预测。不过，在使用另外 38 张公开来源的同类图像进行测试后发现，模型以 0.5 或更高的置信度准确分类了其中 35 张，部署准确率达到 92.10%。

目前最合理的假设是，原始数据集中较强的光照

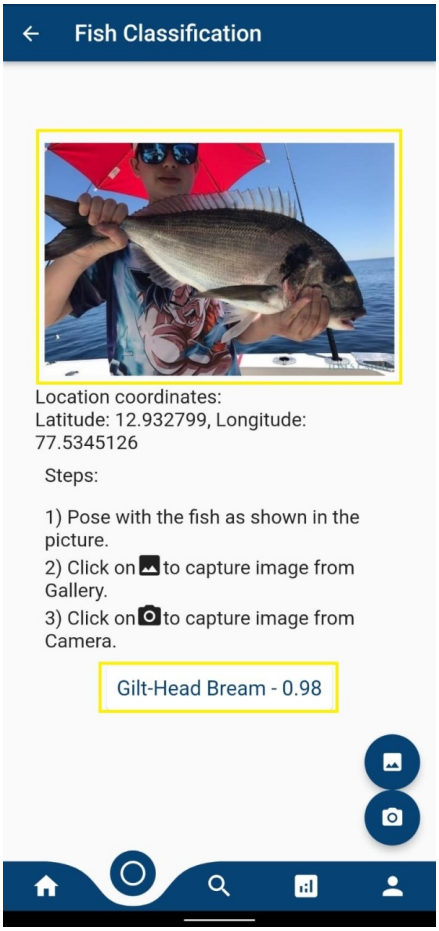


图9. 模型成功评估的一张含有显著噪声的图像示例 [21]。

充分凸显了砧板上的杂散标记，使神经网络能够将其识别为低级特征，因此模型在训练过程中也学会了在预测中考虑显著噪声

VI. 结论与结束语

总之，本研究论文提出的算法展示了出色的准确率，能够根据通过移动应用程序收集的数据对多种海洋物种进行分类。该算法报告的测试准确率为99.83%，结合相较于现状所取得的改进，它有望在海洋保护工作中发挥关键作用。

在 AWS 上的服务器端框架中实现该算法，可根据新数据持续更新和改进模型；这对于海洋生物学这样一个动态变化的领域尤为重要。这项研究有望为持续开展的相关工作作出重要贡献，帮助人们更好地了解和保护海洋生态系统，并为开发更准确、高效的海洋物种分类方法迈出有希望的一步。进一步开发和完善该算法，可能会对海洋保护领域产生深远影响，有助于确保子孙后代能够继续保护海洋物种。

参考文献

[1] F. 联合国粮食及农业组织等，“全球人均鱼类消费量升至每年20千克以上”，2016年。[2] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov和L.-C. Chen，“Mobilenetv2：倒残差与线性瓶颈”，载于《电气电子工程师学会计算机视觉与模式识别会议》（CVPR）论文集，2018年6月。[3] O. Ulucan、D. Karakaya和M. Turkan，“用于鱼类分割与分类的大规模数据集”，载于《2020年智能系统与应用创新会议》（ASYU）。电气电子工程师学会，2020年，第1–5页。[4] S. Kornblith、J. Shlens和Q. V. Le，“更好的ImageNet模型能带来更好的迁移效果吗？”，载于《电气电子工程师学会/计算机视觉基金会计算机视觉与模式识别会议》论文集，2019年，第2661–2671页。[5] F. Rossi、A. Benso、S. Di Carlo、G. Politano、A. Savino和P. L. Acutis，“Fishapp：一种通过图像处理与机器学习技术检测鱼类掺假的移动应用程序”，载于《2016年IEEE自动化、质量与测试、机器人国际会议》（AQTR）。电气电子工程师学会，2016年，第1–6页。[6] H. S. Chhabra、A. K. Srivastava和R. Nijhawan，“一种用于自动鱼类分类的混合深度学习方法”，载于ICETIT 2019论文集。施普林格，2020年，第427–436页。[7] X. Liu、Z. Jia、X. Hou、M. Fu、L. Ma和Q. Sun，“基于嵌入式系统、MobileNet与迁移学习的海洋动物图像实时分类”，载于OCEANS 2019 马赛会议。电气电子工程师学会，2019年，第1–5页。[8] D. Fitrianah、K. M. Suryaningrum、N. T. M. Sagala、V. Ayumi和S. M. Lim，“用于鱼类图像分类的微调MobileNetV2与VGG16算法”，载于《2022年信息学、多媒体、网络与信息系统国际会议》（ICIMCIS）。电气电子工程师学会，2022年，第384–389页。[9] M. Abadi，“TensorFlow：大规模学习函数”，载于《第21届ACM SIGPLAN函数式编程国际会议》论文集，2016年，第1–1页。[10] V. Shah和N. Sajnani，“使用卷积神经网络和TensorFlow Lite进行多类别图像分类”，《工程、科学与管理研究国际期刊》，第3卷，第11期，第65–68页，2020年。[11] P. Arora、V. M. Kapse、S. Sinha和S. Gera，“利用图像数据生成器构建智能系统的肺炎检测分析研究”，载于《2021年第九届可靠性、信息通信技术与优化（趋势与未来方向）国际会议》（ICRITO）。电气电子工程师学会，2021年，第1–4页。[12] T. DeVries和G. W. Taylor，“特征空间中的数据集增强”，arXiv预印本 arXiv:1702.05538，2017年。[13] K. Weiss、T. M. Khoshgoftaar和D. Wang，“迁移学习综述”，《大数据期刊》，第3卷，第1期，第1–40页，2016年。[14] Z. Zhang和M. Sabuncu，“用于带噪声标签的深度神经网络训练的广义交叉熵损失”，《神经信息处理系统进展》，第31卷，2018年。[15] R. A. Dunne和N. A. Campbell，“关于Softmax激活与交叉熵惩罚函数的配对及Softmax激活函数的推导”，载于《第八届澳大利亚神经网络会议论文集》，墨尔本，第181卷。Citeseer，1997年，第185页。[16] G. Juve、E. Deelman、K. Vahi、G. Mehta、B. Berriman、B. P. Berman和P. Maechling，“亚马逊弹性计算云上的科学工作流应用”，载于《2009年第五届IEEE电子科学国际会议研讨会》。电气电子工程师学会，2009年，第59–66页。[17] A. Y. Zomaya等，“并行与分布式计算手册”，图书，1996年。[18] L. Moroney和L. Moroney，“Firebase实时数据库”，《Firebase权威指南：在谷歌移动平台上构建Android应用程序》，第51–71页，2017年。[19] A. Gulli和S. Pal，《使用Keras进行深度学习》。Packt出版有限公司，2017年。[20] A. Khan、A. Chefranov和H. Demirel，“在多层深度卷积神经网络中融合低级特征进行图像场景几何识别”，《神经计算》，第440卷，第111–126页，2021年。[21] T. Komen，“金头鲷垂钓：物种指南、包船与目的地——to m’s catch。”[在线]。可访问：https://www.tomscatch.com/fish-species/gilt-head-bream-35