

基于多层次的海洋生物分类^①

赵东¹, 程远志^{1,2}

¹(青岛科技大学 信息科学技术学院, 青岛 266061)

²(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

通信作者: 程远志, E-mail: yzcheng2007@163.com



摘要: 本文提出了一种多层次海洋生物分类方法。海洋生物种类繁多,且同门类生物具有较强的类间相似性,而不同门类生物具有较大的差异。我们利用物种间的相似性,帮助网络学习生物先验知识,设计出了一种多层次分类方法。设计了C-MBConv模块,并结合多层次分类方法改进了EfficientNetV2网络架构,改进后的网络架构称为CM-EfficientNetV2。我们的实验表明CM-EfficientNetV2比原网络EfficientNetV2有着更高的准确率,在南麂列岛潮间带海洋生物数据集上准确率提高了1.5%,在CIFAR-100上准确率提高了2%。

关键词: 分类; 多层次; 卷积; 海洋生物; 图像识别; 深度学习

引用格式: 赵东,程远志.基于多层次的海洋生物分类.计算机系统应用,2024,33(4):226–234. <http://www.c-s-a.org.cn/1003-3254/9452.html>

Multi-hierarchical Classification for Marine Organisms

ZHAO Dong¹, CHENG Yuan-Zhi^{1,2}

¹(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

²(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: This study proposes a multi-hierarchical classification method for marine organisms. Marine organisms are diverse, and organisms of the same phylum have strong inter-class similarity, while organisms of various phyla have large differences. Meanwhile, a multi-hierarchical classification method is designed by utilizing the similarity among species to help the network learn biological prior knowledge. Additionally, this study designs a C-MBConv module and improves the EfficientNetV2 network architecture by combining the multi-hierarchical classification method, and the improved network architecture is called CM-EfficientNetV2. The experiments show that CM-EfficientNetV2 has higher accuracy than the original network EfficientNetV2, with an accuracy improvement of 1.5% on the inter-tidal marine biology dataset of the Nanji Islands and 2% on CIFAR-100.

Key words: classification; multi-hierarchical; convolution; marine organism; image recognition; deep learning

随着海洋开发的不断推进,海洋生物分类任务的重要性逐渐凸显。由于海洋环境开阔、背景复杂、亮度低等原因,导致在海洋生物栖息地中的分类任务是一项挑战。人工区分鱼类需要大量的检查工作,工作繁琐且乏味,还需要具有过硬专业知识的海洋生物学家。利用技术自动化海洋生物分类和识别将有助于海洋生物科学进一步发展,自动化系统能有效协助海洋生物

学家对各种海洋生物进行分类。

很少有能够实时监测海洋生物,将海中捕获的图像进行分类的应用。这一类系统能够用于海洋监测活动,如评价种群数量、分类当地常见海洋生物种类和各种鱼类洄游^[1]。准确识别海洋生物种类对研究人员、海洋科学家和生物学家非常有帮助,并且可以辅助确定海洋中的生物量水平和地质变化^[2]。

① 收稿时间: 2023-10-08; 修改时间: 2023-11-09; 采用时间: 2023-11-15; csa 在线出版时间: 2024-01-30

CNKI 网络首发时间: 2024-02-01

现阶段海洋生物数量由于环境因素以及人为因素影响造成了很大变化,典型的因素例如全球变暖、过度开采海洋自然资源等原因^[3].这些因素进一步推动了建立一个标准的、经济有效的、值得信赖的方法来监测整个海洋栖息地的海洋生物^[4].随着深度学习的发展,图像分类早已成为火热的研究领域,卷积神经网络并不需要显式的特征提取方法.不少优秀的卷积神经网络在分类任务中表现十分出色,例如 EfficientNetV2^[5].

在海洋生物分类任务中的大多数技术都与一般分类保持一致,都是通过一次分类完成任务.但是海洋生物种类繁多,且识别难度大,为此引入生物学先验知识设计了一种多层次分类方法.为了使用于一般分类任务的神经网络更能贴合多层次分类任务,并且提高识别的精确度,因此对网络进行了优化,设计了一个新的用于多层次分类任务的模块,称为 C-MBConv 块,以及一个新的用于网络二阶分类的网络组件,精细分类模块.结合所提出的模块与组件,生成了用于本多层次分类任务的网络架构 C-EfficientNetV2.

多层次分类任务与传统的分类任务有很大的不同,在传统分类任务中,只需要进行一次分类即可完成任务,但是多层次分类需要进行多次分类.这就导致多层次分类存在很大的风险,如果第 1 次分类出现错误,会导致后续分类基于前一次的错误结果继续分类,显然会导致了整个预测结果的错误.也可以说,多层次分类在进行分类时会具有一定风险,且该风险会传递至下次分类.为了尽可能地降低错误的分类给后续决策任务带来的风险问题,设计了一种将风险最小化的模式.结果对比图如图 1 所示,从中可以看出,我们的方法明显优于其他方法.

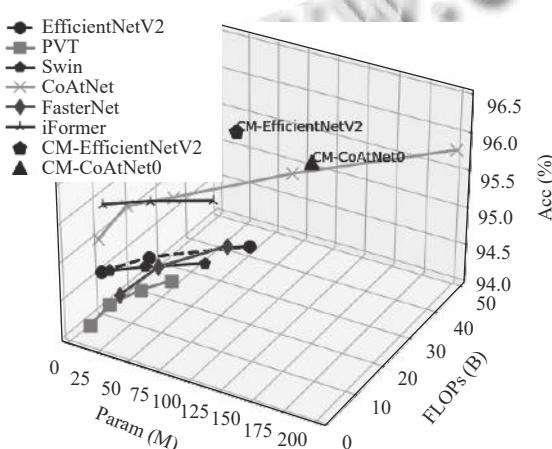


图 1 在南麂列岛潮间带海洋生物数据集上评估精度

1 相关介绍

1.1 多层次分类

在深度学习中,视觉多阶分类任务问题备受关注.目前,深度学习在视觉识别任务中表现出卓越能力,但在处理视觉相似的细粒度对象类时,现有方案往往存在问题^[6,7].这些方案通常采用端到端方式学习代表性特征和 N 路分类器,但忽略了对象类之间的视觉相似性,可能导致难以识别特定对象类^[8,9].

为了解决视觉识别中学习对象类间相似性困难的问题,研究人员提出了利用任务间关系的方法.这包括使用树形结构^[10,11]和深度多任务学习技术^[12,13].然而,大多数现有的深度多任务学习技术假设所有任务都是同等相关的,忽略了任务间关系的显著差异^[10].因此,为了更有效地区分视觉相似的对象类,需要开发新的算法来处理这种巨大差异,实现网络与视觉层次结构的自适应学习.

在现实世界的知识系统中,分层结构被广泛应用于组织数据^[14].已有研究尝试利用各种方法来学习层次结构,包括树结构、层次聚类等.这些方法在不同领域如文本分类^[15]、视觉识别^[16]、肺病分类^[17]、基因功能预测^[18]和植物物种识别^[19]等都得到了应用.方法可以将复杂任务划分为一组相对较小且容易处理的子任务,分层方法比一般分类技术在性能上有更好的表现^[20].

然而,尽管现有研究中提出了一些深度学习和层次分类的方法,仍然面临一些挑战,例如如何解决层间错误传播问题,以及如何使预训练的树结构能够适应深度网络随时间的改进等.总体而言,多层次分类问题和分层视觉识别方法是深度学习中一个备受关注的领域,有待进一步的研究和探索.我们所研究的方法充分利用了类别间的层次关系,更深入学习对象类之间的视觉相似性关系,在海洋生物分类任务中取得了更好的性能和效果.

1.2 神经网络架构

计算机视觉领域的主流架构是卷积神经网络(CNN),其中部分研究利用低秩近似的原理,如 Mobile-Nets^[21,22]、ShuffleNets^[23,24]、GhostNet^[25]、EfficientNet^[26]、Xception^[27]专注于提高参数效率,使得较少的参数能够实现更好的精度,利用深度可分离卷积等技术来减少参数和 FLOPs,但特征映射之间的相关性和冗余性尚

未得到充分利用。

Transformer 在视觉分类任务中也有出色表现, 如 PVT^[28]、Swin^[29]、ViT^[30], 与 CNN 相比, 其全局运算要更胜一筹。还有许多研究将卷积和注意力机制相结合, 如 CoAtNet^[31]。总体来说, CNN 及其变体仍然是计算机视觉主流网络, 并且不断有新的架构和改进被提出, 推动着整个领域的发展。我们的网络模型改进了 EfficientNetV2, 使其更能适应此类任务, 并取得更高的精度。

2 方法介绍

2.1 层次划分

为了解决海洋生物分类任务中, 同门类别包含较强的类间相似性, 异门类别差异大的问题。设计了一种将海洋生物类别进行分层次的方法, 通过逐层次分类, 达到精准分类的目的。此处的分层次, 指的是将海洋生物种类中具有较强类间相似性的种类组成一组, 如鱿鱼和乌贼, 差异性较大的类别将存在于不同的组当中, 如鱿鱼和龙虾。所分的各个组将被称为一阶类别, 组成一阶类别的各个小类将被称为二阶类别。

具体来讲, 将在本海洋分类任务中, 引入生物学先验知识。将同一门类的海洋生物归为一阶类别, 不同门类的海洋生物为不同的一阶类别。如鱿鱼和乌贼为同一阶类别, 鱿鱼和龙虾为不同一阶类别。所有生物都属于不同的二阶精细类别, 例如鱿鱼、乌贼、龙虾。利用生物学先验知识划分好不同层次后, 再使用聚类指标来保证每个一阶类别的样本平衡性、以及类别平衡。

类别平衡指的是各个一阶类别之间的二阶类别数

量需保证平衡, 类别不平衡会导致严重的数据不平衡问题, 从而影响到最终分类器的性能, 为了避免类别失衡, 定义参数 B_{class} 来评估类别平衡性, 其定义如式(2)所示。另外, 不仅需要考虑到保证类别平衡, 还需要保证样本平衡, 如果每个一阶类别的样本数量差异很大, 则会严重影响分类器的训练以及分类性能, 定义参数 B_{sample} 用于评估样本平衡性, 其定义如式(3)所示。将 B_{class} 与 B_{sample} 结合生成平衡参数 $B(c)$ 同时保证样本平衡以及类别平衡。 $B(c)$ 的定义如式(1):

$$B(c) = \frac{B_{\text{class}} \cdot B_{\text{sample}}}{c} \quad (1)$$

$B(c)$ 的值越小则样本的平衡性就越好, 即生成的一阶类别合理。

$$B_{\text{class}} = \sum_{i=1}^C \left[1 + \left(\frac{b_i - b_a}{b_a} \right)^2 \right] \quad (2)$$

其中, C 代表的是一阶类别的数量, b_a 代表被评估的一阶类别中二阶类别的平均数量, b_i 代表第 i 个一阶类中的二阶类别总数量。 B_{class} 代表了一阶类别的类平衡, 此参数越大, 代表越不平衡。

$$B_{\text{sample}} = \sum_{i=1}^C \left[1 + \left(\frac{s_i - s_a}{s_a} \right)^2 \right] \quad (3)$$

其中, s_i 表示为一个一阶类别当中的样本数, s_a 表示一阶类别的平均数量。为了同时保证样本平衡以及类别平衡, 所以需要将其结合生成平衡参数 $B(c)$ 。

如图 2 所示为利用生物学先验知识划分好的多层次海洋生物数据集。

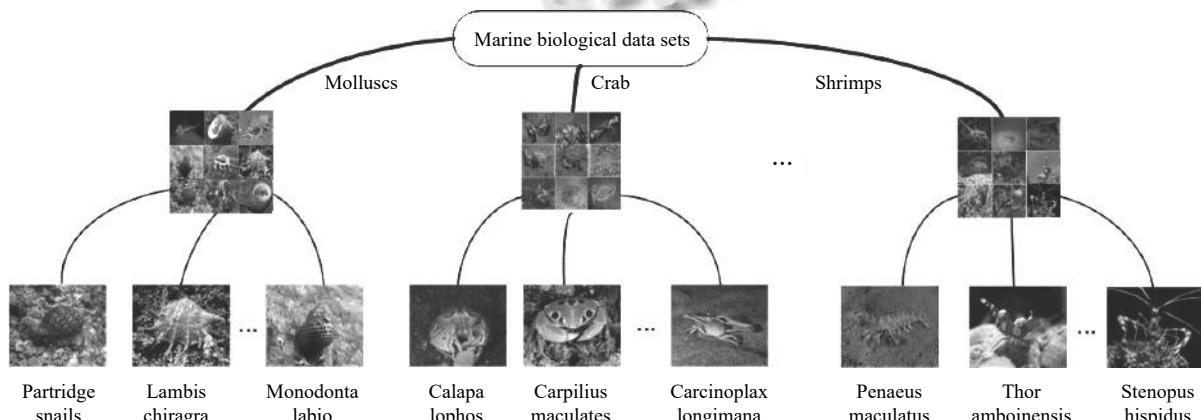


图 2 层次划分

2.2 适应多层次网络

为匹配主干网络所提取特征, 并保证提取模式不

改变, 设计出一个新的模块。参考原主干网络中不同阶段的模块组合, 充分结合 MBConv 块和 Fused-MBconv

块的特点,提出一种新的C-MBConv (combine MBConv) 块.在进行维度扩张后,模型学习通道间依赖关系和空间编码的能力大大增强,经过SE,允许学习通道与通道之间更深层次的相关性,深度可分离卷积再将不同通道的特征图重新调整和组合,分别提取通道特征以

及空间特征,增强模型的表达能力.卷积最后再同时提取空间及通道特征,进一步增强模型的泛化能力.C-MBConv 将更能适应多层次分类任务,使其在多层次分类中获得更好的鲁棒性.

所设计的 C-MBConv 的详细构成如图 3 所示.

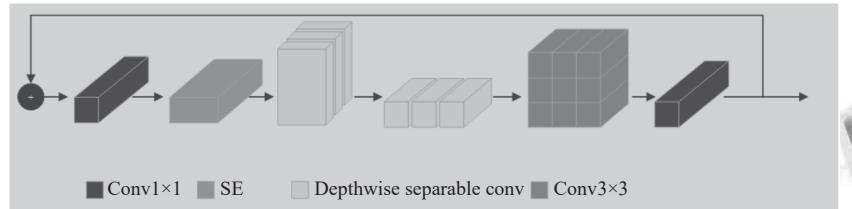


图 3 C-MBConv

基于所提出的 C-MBConv 块,对使用 NAS 框架的 EfficientNetV2 网络架构进行调整,重新调整后的网络命名为 C-EfficientNetV2. C-EfficientNetV2 将主要使用 C-MBConv 块,首先由一个常规卷积来进行通道扩充,随后是由不同数量的 C-MBConv 组成的多个阶段,最后一个阶段为全连接层. C-EfficientNetV2 网络的组成如图 4 所示.

为了使新设计出的 C-EfficientNetV2 更能适应本多层次任务,额外设计了一个新的组件,称其为精细分类模块,使 C-EfficientNetV2 更适合进行多层次分类.额外设计的精细分类模块主要任务为结合主干网络中的特征图,并利用初步决策生成的一阶类别作为辅助,生成用于进行二阶分类的精细特征图.使二阶类别分类精度进一步提高.精细分类模块的主要由 C-MBConv 构成.

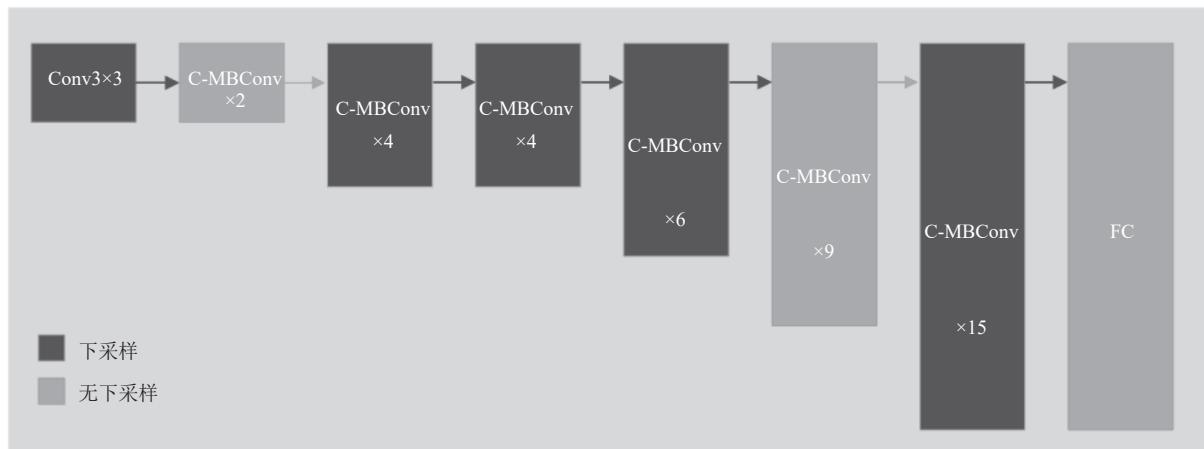


图 4 C-EfficientNetV2

2.3 基于风险最小化的多层次分类

由于多层次分类需要进行多次分类,导致多层次分类存在很大的风险,如果第 1 次分类出现错误,会导致后续分类基于前一次的错误结果继续分类,显然会导致了整个预测结果的错误.例如一个样本的二阶类别为紫海胆,它的实际一阶类别为棘皮动物门,但是被错误地判断为了节肢动物门,那么后续的分类任务会将决策其为截肢动物门中某个二阶类别.也可以说,多

层次分类在进行分类时会具有一定风险,且该风险会传递至下次分类.

为了尽可能地降低错误的分类给后续决策任务带来的风险问题,设计了一种将风险最小化的模式.为所设计的 C-EfficientNetV2 网络增加了对层次分类的置信度估计.利用贝叶斯方法,来对新网络的输出重新建模,提供对预测结果的置信度估计,而不再是一个单个值.综上所述,给出风险最小化决策的定义,如式(4):

$$\hat{y}_{\min} = \arg \min \left(\sum_{i=1}^C L(\hat{y}, y) \cdot P(\hat{y}|x) \right) \quad (4)$$

其中, \hat{y} 为网络对所有一阶类别可能的决策, $P(\hat{y}|x)$ 为对所有可能决策的贝叶斯后验概率, 以及 $L(\hat{y}, y)$ 为对应的损失值. 利用设计的网络模型做出一阶分类决策, 再通

过贝叶斯推理对模型决策结果做出延伸, 实现对决策的最小化风险控制.

完整的多层次分类流程如图 5 所示, 主要为 3 部分组成, 左侧为 C-EfficientNetV2 网络架构, 中部结合风险最小化产生一阶决策, 右侧为精细决策产生二阶类别.

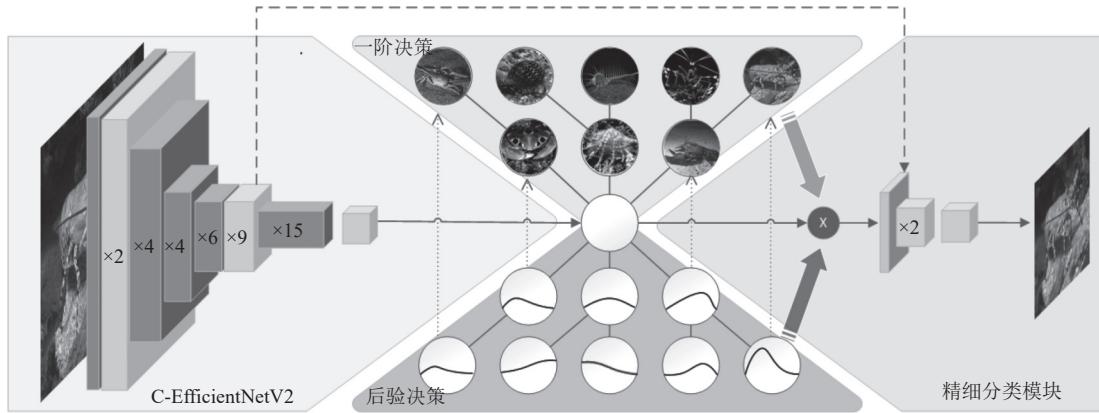


图 5 多层次分类

3 实验结果与分析

本节主要介绍了本方法的消融实验部分以及对比实验部分.

数据集: 实验中主要使用了两个数据集, 第 1 个数据集为南麂列岛潮间带海洋生物数据集, 该数据集由中国科学院海洋所提供, 包含 342 个海洋生物种类(二阶类别), 可以归于 9 个大类(一阶类别)中, 总共约 34.2M 张海洋生物图片, 数据集架构如图 2 所示, 按照 4:1 的比例划分训练集和测试集. 为了验证本方法的可扩展性, 第 2 个数据集使用了公共数据集 CIFAR-100^[32], 该数据集包含 100 个类别, 总计共 60M 张图片, CIFAR-100

的 100 个类别又归属于 20 个超类, 符合本实验的多阶类别设定, 按本实验划分如图 6 所示. 要注意的是, 对于 CIFAR-100 的测试, 我们对所有模型都没有使用其他数据集的预训练, 而是直接训练.

实验设置: 我们的实验将图像调整大小为 256×256, 并中心裁剪为 224×224 进行训练, 使用了一些数据增强方法, 包括 Mixup^[33]、CutMix^[34]、label smoothing^[35] 和 RandAugment^[36], 训练的 batch_size 设置为 128, 训练 512 个周期, 并由 AdamW^[37] 优化, 损失函数设置为 SoftTargetCrossEntropy, 初始学习率为 1×10^{-3} 并且使用余弦调度^[38], 权重衰减设置为 0.05.

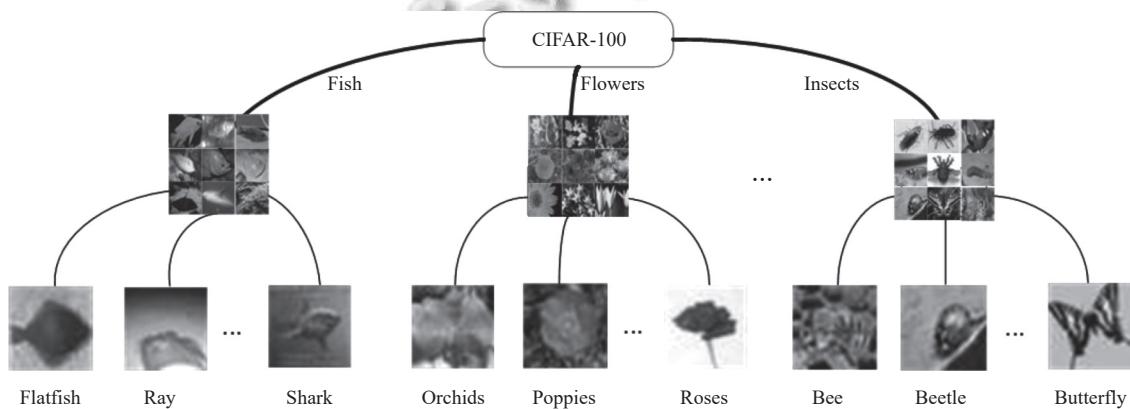


图 6 CIFAR-100 多层次划分

实验环境: 本实验在 Ubuntu 22.04 LTS 系统下进行, 使用 PyTorch 1.13.1 深度学习框架, Python 版本为

3.7.16, CUDA12.0, CPU 型号为 i7-13790F, GPU 使用 NVIDIA GeForce RTX 4090, 128 GB 内存.

3.1 消融实验

本部分将依次对方法所涉及的超参数, 以及所设计的模块对方法的影响进行探究.

3.1.1 超参数

对于超参数 E , 一阶类别权重, 将增强属于相同一阶类别的权重, 且降低不同一阶类别权重, 实验环境保证都是在相同复杂度的 C-EfficientNetV2 模型下进行, 并使用南麂列岛潮间带海洋生物数据集评估, 表 1 展示了超参数 E 值设置的大小对性能的影响.

表 1 一阶类别权重 E 对模型的影响

E	Top-1 Acc (%)	E	Top-1 Acc (%)
0	70.52	0.6	86.36
0.1	79.28	0.7	85.76
0.2	80.95	0.8	85.72
0.3	82.74	0.9	86.10
0.4	85.69	1.0	85.69
0.5	85.61		

要注意的是, 当 E 为 1 的时候, 表示实验变为只使用原特征图, 与只使用 C-EfficientNetV2 无异. 当 E 为 0 的时候, 代表完全忽视异类对特征的影响, 而只注意相同类别之间的关系. 通过表 1 的结果显示, 当 E 为 0 的时候, 准确率达到最低的 70.52%, 当过于忽视不同一阶类别对特征影响的时候, 准确率会变得很差, 这代表网络需要学习到异类之间的信息, 而不能只学习到相同类别之间的相似度关系. E 设置为 0.6 的时候, 准确率达到了最高的 86.36, 而 E 为 1.0 时准确率为 85.69%, 这代表了可以通过适当提高相似一阶类别权重, 使得网络模型能更好学习到类别之间的相似度信息.

对于超参数 α , 风险最小化权重, 将决定 \hat{y}_{min} 对一阶决策特征图所产生的影响. 使用相同的实验环境与数据集, E 设置为 0.6, 对超参数 α 进行估计, 表 2 展示了超参数 α 值设置的大小对模型的影响.

表 2 α 大小对模型的影响

α	Top-1 Acc (%)
0	95.95
0.3	95.79
0.5	96.18
0.7	96.02
1	95.62

当 α 为 0 时代表风险最小化权重完全主导决策, 与只使用贝叶斯层无异. 当 α 为 1 时, 代表只使用原特征图, 即用 C-EfficientNetV2, 而不用风险最小化方法. 观察表 2 可知, 不论是只是用原特征图还是只用贝叶

斯层, 准确率都未能达到最高. 当 α 的值为 0.5 时, 精度达到最高, 这说明了当原网络与贝叶斯层相辅相成时, 可以令模型具有更好的性能, 进一步说明了风险最小化策略的有效性.

对于超参数 β , 一阶损失权重. 由于我们的方法由网络主体生成一阶类别, 由精细分类模块生成二阶类别, 这就包含了两处需要计算损失地方, 因此就需要一个超参数来权衡一阶损失与二阶损失. 采用与上述实验相同的实验环境, 对超参数 β 进行估计, 表 3 展示了超参数 β 的大小对模型精度的影响.

表 3 β 大小对模型的影响

β	Top-1 Acc (%)	β	Top-1 Acc (%)
0	86.36	0.6	84.67
0.1	86.42	0.7	84.05
0.2	86.65	0.8	84.66
0.3	85.76	0.9	83.44
0.4	85.2	1.0	83.35
0.5	85.05		

当 β 设置为 0 时, 代表只使用二阶损失, 则模型不考虑一阶损失. 当 β 设为 1 时, 则相反. 观察表 3 的结果可知, 单独使用二阶损失或一阶损失准确率都没有最优, 且当二阶损失权重比例偏大的时候, 准确率会倾向于更高. 当 β 设置为 0.2 的时候, 准确率达到了最高, 结果表明, 一阶损失需在本任务中占有一定比例时, 可以使模型表现得更出色, 进一步说明了本方法利用一阶损失提升训练效果的有效性.

3.1.2 C-MBConv

所提出的 C-MBConv 在本文第 3.2 节中做了介绍, 为了验证本模块的有效性, 我们将主要对 EfficientNetV2 利用 C-MBConv 来进行改动, 随后与原网络对比性能. 额外地, 由于 CoAtNet 中也包含了 MBConv, 对 CoAtNet 也进行同样的变动, 对比网络性能, 注意, 本部分未使用精细分类模块和多层次分类部分, 只考虑 C-MBConv. 表 4 展示了 C-MBConv 对网络性能的影响.

表 4 C-MBConv 对模型性能的影响

模型_模块	#Param (M)	Top-1 Acc (%)
EfficientNetV2_s_222221	107	95.10
EfficientNetV2_s_222211	35	95.14
EfficientNetV2_s_222111	23	94.51
EfficientNetV2_s_221111	21	94.65
EfficientNetV2_s	20	94.87
EfficientNetV2_m	53	95.14
CoAtNet0_C	31	95.57
CoAtNet0	17	95.27
CoAtNet1	32	95.56

对于 EfficientNetV2_s 的改动主要在其中间的 6 个阶段中, 数字 2 表示改动, 1 表示不改动, 而对于 CoAtNet0 则直接替换了原网络的 MBConv。观察结果表明, 改动后的网络都比原网络精确度都有了提升, 对比同参数量的变体依然能有更好表现。因此可知 C-MBConv 可以使得网络性能提升, 综合模型大小与精确度, 我们在后续实验中选择出 EfficientNetV2_s_222211 变体作为最优模型, 后续将直接称其为 C-EfficientNetV2_s。

3.1.3 精细分类模块

设计的精细分类模块在本文第 3.2 节已做过说明, 本节将探讨此模块对本方法的有效性。同样主要针对 EfficientNetV2 进行实验, 为 EfficientNetV2_s 增加精细分类模块, 对比其与原网络的性能。[表 5](#) 列出了精细分类模块对模型性能的影响。

[表 5](#) 精细分类模块对模型性能的影响

模型	精细分类块 N(个)	#Param (M)	Top-1 Acc (%)
EfficientNetV2_s	0	20	94.87
	1	56	95.27
	2	65	95.34
	3	74	95.50
	5	91	95.46
	8	117	95.33
	15	178	95.40
EfficientNetV2_l	0	118	95.36

N 为精细分类模块的个数, 当 N 为 0 时, 代表使用的是原架构。观察[表 5](#)中的结果可以发现, 增加精细分类模块可以使得网络得准确率逐步获得提高, 当 N 值为 3 时到达峰值, 准确率达到了 95.50%。因此结果表明, 精细分类模块确实可以提升网络的性能。综合模型大小与精确度, 最终我们选择了 N 设置为 3 的模型, 作为后续实验最优方法, 结合入 C-EfficientNetV2_s 中。

3.1.4 多层次分类

在本节中, 将测试多层次分类方法与原网络的性能。具体实验设置是为 C-EfficientNetV2 增加精细分类模块, 且融入多阶分类先验知识, 最后结合风险最小化, 将此模型称为 CM-EfficientNetV2。同样也对 CoAtNet0 设置了多层次分类方法, 暂且称为 CM-CoAtNet0。[表 6](#)、[表 7](#)列出了多层次分类方法对模型性能的影响。

观察[表 6](#)和[表 7](#)中结果可知, 应用了本方法的模型都获得了提升, 且对比相同参数量原网络架构时依然能取得出色性能, 其中 CM-EfficientNetV2 在海洋数据集上的准确率可达到 96.47%, 超过原架构 1.6%。在

CIFAR-100 数据集上不使用其他数据集预训练模型的基础上, 准确率达到 85.89%, 超过原架构 1.76%。

[表 6](#) 在海洋数据集上多层次分类方法对模型性能的影响

模型	#Param (M)	FLOPs (B)	Top-1 Acc (%)
CM-EfficientNetV2	90	19.5	96.47
EfficientNetV2_s	20	2.92	94.87
EfficientNetV2_m	53	5.5	95.14
EfficientNetV2_l	118	12.4	95.36
CM-CoAtNet0	134	26.3	96.07
CoAtNet0	17	3.4	95.27
CoAtNet1	32	6.9	95.66

[表 7](#) 在 CIFAR-100 数据集上多层次分类方法对模型性能的影响

模型	#Param (M)	FLOPs (B)	Top-1 Acc (%)
CM-EfficientNetV2	88	19.5	85.89
EfficientNetV2_s	20	2.92	84.13
EfficientNetV2_m	53	5.5	84.87
EfficientNetV2_l	118	12.4	84.96
CM-CoAtNet0	134	26.3	83.89
CoAtNet0	17	3.4	82.54
CoAtNet1	32	6.9	83.69
CoAtNet2	54	12.9	83.85

3.2 对比试验

在本节中, 我们将对比我们的方法与现阶段先进的技术。对比网络包括 PVT、Swin、CoAtNet、FasterNet^[39]、iFormer^[40]。各个模型的对比结果如[表 8](#)、[表 9](#)所示。整体对比结果如[图 1](#)所示。

[表 8](#) 海洋数据集模型对比结果

模型	#Param (M)	FLOPs (B)	Top-1 Acc (%)
CM-EfficientNetV2	90	19.5	96.47
EfficientNetV2_s	20	2.92	94.87
EfficientNetV2_m	53	5.5	95.14
EfficientNetV2_L	118	12.4	95.36
pvt_t	12	1.9	94.15
pvt_s	24	3.7	94.43
pvt_m	43	6.5	94.62
pvt_L	61	9.5	94.73
Swin_t	27	3.0	94.92
Swin_s	49	5.8	94.98
Swin_b	87	10.2	95.06
CM-CoAtNet0	134	26.3	96.07
CoAtNet0	17	3.4	95.27
CoAtNet1	32	6.9	95.66
CoAtNet2	55	12.9	95.68
CoAtNet3	116	27.7	95.79
CoAtNet4	202	48.6	95.82
FasterNet_s	30	4.6	94.55
FasterNet_m	52	8.8	94.89
FasterNet_l	92	15.5	95.13
iFormer_s	19	4.5	95.70
iFormer_b	47	8.8	95.73
iFormer_L	86	13.3	95.77

表9 CIFAR-100 数据集模型对比结果

模型	#Param (M)	FLOPs (G)	Top-1 Acc (%)
CM-EfficientNetV2	88	19.5	85.89
EfficientNetV2_s	20	2.92	84.13
EfficientNetV2_m	53	5.5	84.87
EfficientNetV2_L	118	12.4	84.96
pvt_t	12	1.9	76.04
pvt_s	24	3.7	77.03
pvt_m	43	6.5	77.67
pvt_L	60	9.5	78.01
Swin_t	27	3.0	79.76
Swin_s	48	5.8	80.13
CM-CoAtNet0	134	26.3	83.89
CoAtNet0	17	3.4	82.54
CoAtNet1	32	6.9	83.69
CoAtNet2	54	12.9	83.85
CoAtNet3	116	27.7	84.16
CoAtNet4	202	48.6	84.91
FasterNet_s	30	4.6	81.78
FasterNet_m	52	8.8	82.42
iFormer_s	19	4.5	82.67
iFormer_b	47	8.8	82.89
iFormer_L	86	13.3	83.02

观察表8和表9结果可知,我们的多层次分类方法在南麂列岛潮间带海洋生物数据集以及CIFAR-100数据集上都获得了优异的成绩。由此可知,我们所使用的多层次分类方法是一种有效提高模型精度的方法,并且可以扩展至其他数据集。

4 结论与展望

本文提出了一种海洋生物多层次分类的方法CM-EfficientNetV2,本方法利用生物先验生成多阶类别,由风险最小化方法降低异类类别的错误可能。其网络架构主要基于C-MBConv模块构成,可以更细致地学习通道特征与空间特征的相关性。为适应多层次分类任务延伸出精细分类模块,充分结合粗细粒度特征与风险最小化方法,生成二阶精细类别。提出的方法在CIFAR-100数据集和南麂列岛潮间带海洋生物数据集上不仅超越了EfficientNetV2且都表现出了优异性能。实验结果表明了本网络架构具有很强的可扩展性,期望未来可以扩展至其他领域。

参考文献

- Knausgård KM, Wiklund A, Sørdalen TK, et al. Temperate fish detection and classification: A deep learning based approach. *Applied Intelligence*, 2022, 52(6): 6988–7001. [doi: 10.1007/s10489-020-02154-9]
- Shafait F, Mian A, Shortis M, et al. Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science, 2016, 73(10): 2737–2746. [doi: 10.1093/icesjms/fsw106]
- Villon S, Chaumont M, Subsol G, et al. Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and HOG+SVM methods. *Proceedings of the 17th International Conference on Advanced Concepts for Intelligent Vision Systems*. Lecce: Springer, 2016. 160–171. [doi: 10.1007/978-3-319-48680-2_15]
- Siddiqui SA, Salman A, Malik MI, et al. Automatic fish species classification in underwater videos: Exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*, 2018, 75(1): 374–389. [doi: 10.1093/icesjms/fsx109]
- Tan MX, Le QV. EfficientNetV2: Smaller models and faster training. *Proceedings of the 38th International Conference on Machine Learning*. ICML, 2021. 10096–10106.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe: NIPS, 2012. 1106–1114.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Proceedings of the 3rd International Conference on Learning Representations*. San Diego: ICLR, 2015.
- Sfar AR, Boujemaa N, Geman D. Vantage feature frames for fine-grained categorization. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland: IEEE, 2013. 835–842.
- Deng J, Krause J, Fei-Fei L. Fine-grained crowdsourcing for fine-grained recognition. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland: IEEE, 2013. 580–587.
- Fan JP, Zhao TY, Kuang ZZ, et al. HD-MTL: Hierarchical deep multi-task learning for large-scale visual recognition. *IEEE Transactions on Image Processing*, 2017, 26(4): 1923–1938. [doi: 10.1109/TIP.2017.2667405]
- Kontschieder P, Fiterau M, Criminisi A, et al. Deep neural decision forests. *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1467–1475.
- Fan JP, Zhou N, Peng JY, et al. Hierarchical learning of tree classifiers for large-scale plant species identification. *IEEE Transactions on Image Processing*, 2015, 24(11): 4172–4184. [doi: 10.1109/TIP.2015.2457337]
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 3431–3440.
- Silla CN, Freitas AA. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 2011, 22(1–2): 31–72. [doi: 10.1007/s10618-010-0175-9]

- 15 Esuli A, Fagni T, Sebastiani F. Boosting multi-label hierarchical text categorization. *Information Retrieval*, 2008, 11(4): 287–313. [doi: [10.1007/s10791-008-9047-y](https://doi.org/10.1007/s10791-008-9047-y)]
- 16 Griffin G, Perona P. Learning and using taxonomies for fast visual categorization. *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage: IEEE, 2008. 1–8.
- 17 Chang YJ, Kim N, Lee Y, et al. Fast and efficient lung disease classification using hierarchical one-against-all support vector machine and cost-sensitive feature selection. *Computers in Biology and Medicine*, 2012, 42(12): 1157–1164. [doi: [10.1016/j.combiomed.2012.10.001](https://doi.org/10.1016/j.combiomed.2012.10.001)]
- 18 Wang HX, Shen XT, Pan W. Large margin hierarchical classification with mutually exclusive class membership. *The Journal of Machine Learning Research*, 2011, 12: 2721–2748.
- 19 Gopal S, Yang YM. Hierarchical Bayesian inference and recursive regularization for large-scale classification. *ACM Transactions on Knowledge Discovery From Data*, 2015, 9(3): 18.
- 20 Deng J, Satheesh S, Berg AC, et al. Fast and balanced: Efficient label tree learning for large scale object recognition. *Proceedings of the 25th Annual Conference on Neural Information Processing Systems*. Granada: NIPS, 2011. 567–575.
- 21 Howard AG, Zhu ML, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv:1704.04861*, 2017.
- 22 Sandler M, Howard AG, Zhu ML, et al. MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4510–4520.
- 23 Zhang XY, Zhou XY, Lin MX, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 6848–6856.
- 24 Ma NN, Zhang XY, Zheng HT, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 122–138.
- 25 Han K, Wang YH, Tian Q, et al. GhostNet: More features from cheap operations. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 1577–1586.
- 26 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*. Long Beach: ICML, 2019. 6105–6114.
- 27 Chollet F. Xception: Deep learning with Depthwise separable convolutions. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 1800–1807.
- 28 Wang WH, Xie EZ, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE, 2021. 548–558.
- 29 Liu Z, Lin YT, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE, 2021. 9992–10002.
- 30 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the 9th International Conference on Learning Representations*. ICLR, 2021.
- 31 Dai ZH, Liu HX, Le QV, et al. CoAtNet: Marrying convolution and attention for all data sizes. *Proceedings of the 34th Advances in Neural Information Processing Systems*. NeurIPS, 2021. 3965–3977.
- 32 Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Technical Report, Toronto: University of Toronto, 2009.
- 33 Zhang HY, Cissé M, Dauphin YN, et al. Mixup: Beyond empirical risk minimization. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: ICLR, 2018.
- 34 Yun S, Han D, Chun S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 6022–6031.
- 35 Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 2818–2826.
- 36 Cubuk ED, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle: IEEE, 2020. 3008–3017.
- 37 Loshchilov I, Hutter F. Decoupled weight decay regularization. *Proceedings of the 7th International Conference on Learning Representations*. New Orleans: ICLR, 2019.
- 38 Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. *Proceedings of the 5th International Conference on Learning Representations*. Toulon: ICLR, 2017.
- 39 Chen J, Kao S, He H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023. 12021–12031.
- 40 Si CY, Yu WH, Zhou P, et al. Inception transformer. *Proceedings of the 36th NeurIPS 2022*. New Orleans: NeurIPS, 2022.

(校对责编: 孙君艳)