

基于改进Res2Net模型的淡水鱼类图像分类研究

赵正伟, 朱宏进, 杨根滕, 王金坤

(广西民族大学电子信息学院, 广西 南宁 530006)

✉zzwei@gxun.edu.cn; 1437489585@qq.com; ygt6686@126.com; 1823323212@qq.com



摘要: 针对传统的淡水鱼类图像识别方法速度慢、需要人工提取特征等问题, 提出一种基于改进Res2Net模型的淡水鱼类图像识别算法。提出的改进方案如下: 首先使用CELU激活函数代替ReLU激活函数; 接着将残差块与混合注意力网络相结合; 最后使用三个 3×3 卷积核替代Res2Net模型中第一个卷积层的 7×7 卷积核, 同时在下采样的残差连接中加入平均池化层。实验结果表明, 改进的网络在淡水鱼类图像分类上达到了96.34%的准确率, 比Res2Net的准确率高3.67%, 具有更加优异的性能, 可为淡水鱼类识别提供参考。

关键词: 淡水鱼; 图像识别; 注意力机制

中图分类号: TP391.4 **文献标识码:** A

Research on Classification of Freshwater Fish Images based on Improved Res2Net Model

ZHAO Zhengwei, ZHU Hongjin, YANG Genteng, WANG Jinkun

(College of Electronic Information, Guangxi Minzu University, Nanning 530006, China)

✉zzwei@gxun.edu.cn; 1437489585@qq.com; ygt6686@126.com; 1823323212@qq.com

Abstract: Aiming at the slow speed of traditional freshwater fish image recognition methods and the need to manually extract features, this paper proposes an improved freshwater fish image recognition algorithm based on improved Res2Net model. The proposed improvement scheme is as follows: first, CELU activation function is used instead of ReLU. Then, the residual block is combined with the hybrid attention network. Finally, three 3×3 convolution kernels are used to replace the 7×7 convolution kernels of the first convolutional layer in Res2Net model. At the same time, an average pooling layer is added to the down-sampled residual connection. Experimental results show that the improved network achieves 96.34% accuracy in freshwater fish image classification, which is 3.67% higher than that of Res2Net. It has more excellent performance and provides a reference for freshwater fish recognition.

Keywords: freshwater fish; image recognition; attention mechanism

1 引言(Introduction)

中国是一个淡水鱼类养殖大国^[1], 鱼类的快速识别是实现鱼类养殖自动化的前提, 然而由于不同鱼类外形具有高度的相似性, 这在很大程度上增加了鱼类分类的难度。近年来, 卷积神经网络在计算机视觉领域展现出许多优越性, 诞生了许多优秀的深度学习算法, 例如GAO等于2019年在残差网络ResNet^[2]的基础上提出了一种新的多尺度特征提取方法Res2Net^[3], 通过组合不同数量的卷积核来提取多尺度特征,

达到了更佳的性能。ZHANG等在2021年提出了一种新的注意力机制SANET^[4], 通过分组结合通道注意力机制与空间注意力机制, 提升模型对重要信息的关注度, 可以有效地提高模型的性能。

本研究在分析经典Res2Net模型的基础上, 对模型结构进行改进, 并且引入混合注意力机制提高模型特征学习能力, 通过构建淡水鱼类图像数据集, 实现对10种淡水鱼类图像的精确识别。

2 淡水鱼类图像识别模型构建(Construction of image recognition model for freshwater fish)

2.1 新维度残差网络模型

考虑到构建的淡水鱼类图像数据集图片数量较少, 种类数量繁多, 而且淡水鱼类图像本身具有类间差异小、背景复杂的特点, 同时考虑到算法的时效性, 所选网络模型计算开销不能过大, 因此选用对细粒度图像分类显著的Res2Net-50模型作为淡水鱼类图像识别的分类器。由于对输入特征通道进行了分组处理, 使用不同数量的 3×3 大小的卷积核组合进行分层连接, 使得其在未增加计算开销的情况下, 在粒度更细的级别上表示了图片的特征信息, 大幅度增加了网络对细粒度图像的特征提取能力。同时, 由于使用了残差连接^[2], 解决了深层神经网络由于网络层数过深导致梯度消失或者梯度爆炸而出现的不易收敛的问题, 可以通过增加神经网络的深度或者宽度, 使网络学习到淡水鱼类图像更深层次的语义信息。Res2Net-50模型结构总共由七个部分组成, 包括开始的卷积层与五个类似的残差模块, 残差模块由不同数量的残差单元构成, 最后使用平均池化层与全连接层作为输出端, 其中残差单元结构如图1所示。

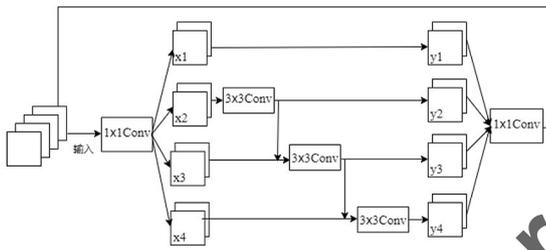


图1 残差单元结构

Fig.1 Residual unit structure

每个残差单元由两个 1×1 卷积层与一组 3×3 卷积核构成, 同时在输入与输出之间添加一个恒等映射的残差连接。每个残差单元中 3×3 卷积层进行多尺度特征提取的计算公式如式(1)所示:

$$y_i = \begin{cases} x_i, & i = 1 \\ K(x_i), & i = 2 \\ K(x_i + y_{i-1}), & i = 3, 4 \end{cases} \quad (1)$$

其中, $x_i \in R^{(B \times W \times H \times C/4)}$, $i = 1, 2, 3, 4$ 表示输入特征图经过Split操作后得到的四个子特征图; B 为输入神经网络的RGB图片数量; $W \times H$ 分别表示特征图的宽和高; C 为输入特征图的通道数量; $K(*)$ 表示使用 3×3 的卷积核进行卷积操作; y_i 为每个子特征图经过多尺度特征提取后对应的输出。因此, Res2Net的残差单元输出由不同粒度大小的感受野组成, 可以很好地提取图像的全局特征以及局部特征, 提高模型的性能。

2.2 改进Res2Net-50模型

本文提出的淡水鱼类图像识别模型是在原始Res2Net-50模型的基础上进行构建与改进的, 包括49个卷积层、2个池化层与1个全连接层。改进后的模型将第一个 7×7 的卷积层使

用3个 3×3 的卷积层进行代替, 同时将原始模型卷积层后的ReLU激活函数代替为CELU激活函数, 最后使用增加了注意力机制与平均池化的改进残差模块代替原始残差模块, 保持模型宽度不变。模型的整体结构如图2所示。

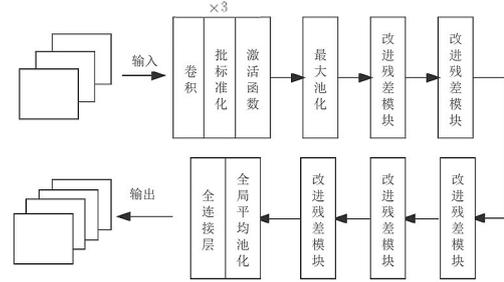


图2 改进的Res2Net-50网络结构

Fig.2 Improved Res2Net-50 network architecture

2.2.1 CELU激活函数

传统Res2net-50模型中使用了ReLU激活函数, 具有运算速度快, 可以有效缓解过拟合的特点, 计算公式如下:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2)$$

由式(2)可以看出, 当神经元的输入为负值时, 计算的输出全部为零。在模型反向传播根据梯度更新参数时, 由于梯度被置零, 因此该神经元参数可能不会进行激活更新, 出现神经元失活现象。为了解决该问题, 本文使用CELU^[5]激活函数代替ReLU激活函数, 其公式如下:

$$f(x) = \max(0, x) + \min(0, a(\exp(x/a) - 1)) \quad (3)$$

由式(3)可以看出, 当输入为正值时, 函数保留了ReLU激活函数的特点; 当输入为负值时, CELU激活函数可以避免梯度被置零, 使参数更新正常进行。实验设置 a 的值为1。

2.2.2 多尺度融合注意力模块

为了加强模型对淡水鱼类图像中重要特征信息的提取效率, 本研究在残差单元中的 1×1 卷积后加入SANET注意力模块, 同时从空间与通道层面加强对鱼类图像关键特征的提取^[6]。SANET具体流程如图3所示。

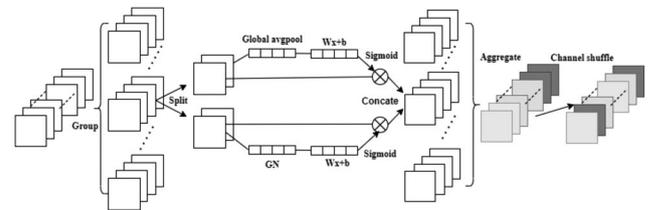


图3 SANET模块流程图

Fig.3 Flowchart of SANET (Shuffle Attention for Deep Convolutional Neural Network) module

首先将特征图 $D \in R^{(B \times C \times H \times W)}$ 在通道维度平均分组为多个子特征图 $D_k \in R^{(B \times C/G \times H \times W)}$, $k = 1, 2, \dots, G$, 接着将每个子特征图平均分成两部分 $D_{k1}, D_{k2} \in R^{(B \times C/2G \times H \times W)}$; 其中一部分执行通道注意力操作, 首先经过全局平均池化, 将特征图的每一个通道压缩成一个实数, 接着通过一个门机制与Sigmoid函数将每

一通道转化为0—1的权重系数，最后将输入特征图逐通道与权重系数相乘得到输出；另一部分执行空间注意力操作，首先经过分组归一化操作^[7]后，同样经过一个门机制与Sigmoid函数得到每个通道的权重系数，逐通道相乘后得到输出，最后将每组的两部分输出拼接后经过Channel shuffle操作得到最终的输出 $D_{out} \in R^{(B \times C \times H \times W)}$ 。上述注意力特征提取过程可以表示为：

$$D_{out1} = \sigma(W_1(\text{globalavgpool}(D_{k1})) + b_1) \cdot D_{k1} \quad (4)$$

$$D_{out2} = \sigma(W_2(\text{GN}(D_{k2})) + b_2) \cdot D_{k2} \quad (5)$$

其中， $W_1, W_2, b_1, b_2 \in R^{(B, C/2G, 1)}$ 表示一个简单的门机制，用于对经过全局平均池化与分组归一化的特征图进行缩放与平移。

2.2.3 卷积核的替换与平均池化的添加

在经典的Res2Net模型结构中，输入图像首先经过 7×7 大小的卷积核进行大块的全局特征提取，但是由于在淡水鱼类图像识别中，各种鱼类的外形高度相似^[8]，需要使用较小的卷积核提取图像更加细微的局部特征来进行区分^[9]，因此将模型第一个卷积层中 7×7 大小的卷积核利用3个 3×3 大小的卷积核进行替换。改进后的模型具有与原始模型相同的感受野大小^[10]，并且通过加深网络，使模型具有更强的非线性拟合能力，同时降低了模型需要训练的参数数量。

同时，传统的Res2Net模型中，在执行下采样操作的残差单元内，残差连接通过使用步长为2、大小为1的卷积来进行特征图的缩小。但是，由于 1×1 大小的卷积操作一般用于调整特征图的通道数，并没有考虑特征图相邻像素之间的联系，会造成部分淡水鱼类图像特征信息的丢失^[11]，因此，本文通过在残差连接中添加步长为2、大小为2的平均池化层来进行特征图的下采样，具体如图4所示。

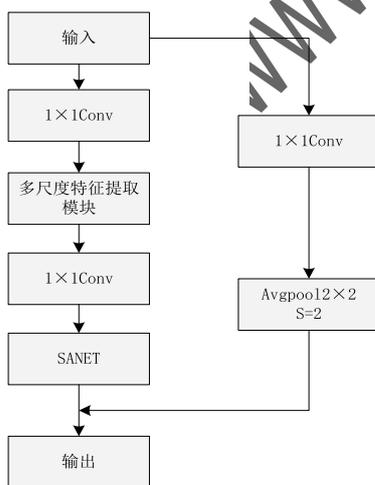


图4 改进残差单元

Fig.4 Improving residual unit

3 实验与结果分析(Experiment and result analysis)

3.1 淡水鱼类图像数据集

本研究选取的淡水鱼类图像数据集共有1,382张图片，包

含10种淡水水域常见的经济鱼类与入侵鱼类，每一类图片数量为60—175张，数据通过手动在互联网上搜集并整理得到，部分图片如图5所示。本实验将淡水鱼类图像数据集的80%划分为训练集，20%作为测试集。



图5 淡水鱼样本数据

Fig.5 Sample data of freshwater fish

由于淡水鱼类图像数据集数据较少，在模型的训练过程中容易出现过拟合的现象，因此在将原始数据集划分为训练集与测试集后，对训练集的图像采用数据增强的方式进行扩增。本实验采用的数据增强方式为垂直水平翻转、顺时针旋转 20° 、 340° 。扩充后淡水鱼类数据集总共包含4,715张图片，其中训练集图像4,442张，测试集图像273张。

3.2 实验设置

首先，将输入图像的尺寸使用Pytorch自带的函数统一缩放为网络所需要输入的标准尺寸 224×224 ，并且以0.5的概率进行图像的随机水平翻转，接着为了加速模型的训练与收敛速度，进行均值方差的归一化。在模型训练过程中使用Adam优化器，学习率设置为固定值0.0001，batch_size设置为32，采用交叉熵损失函数(Cross Entropy Loss)作为模型训练的损失函数，在淡水鱼类图像数据集上训练100个轮次。

3.3 实验环境

本研究使用Pytorch 1.9.1 CUDA深度学习框架搭建淡水鱼类图像识别模型，采用专业的图形处理工作站，使用Windows 10、64位操作系统，GPU为适用专业台式工作站的NVIDIA Quadro p4000 8 GB，CPU为英特尔Xeon(至强) Silver 4210@ 2.00 GHz。

3.4 结果与分析

为了验证本研究提出算法的有效性，将改进的Res2Net-50模型与传统的Res2Net-50模型在经过数据增强的淡水鱼类图像数据集上进行对比分析，保证模型在相同的实验条件下开始训练直至收敛，模型每训练一个轮次就进行一次测试，以便更加直观地查看模型在训练过程中的性能变换情况。

如图6所示为两个模型在淡水鱼类图像数据集上的测试集准确率变换情况。从图6中可知，两个模型在前20个轮次内，

测试集的准确率迅速上升；在模型训练至20个轮次后，模型的准确率曲线变换幅度逐渐减小，整体上表现为上升趋势。在模型逐渐收敛以后，两个模型的准确率曲线存在一个比较明显的高度差，相差约3.67%。

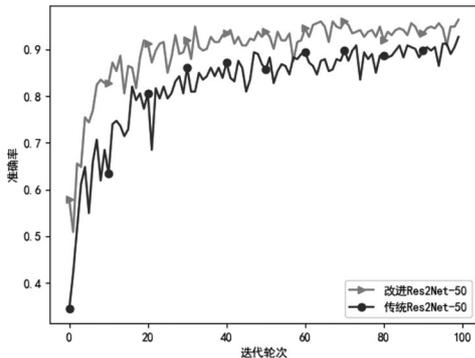


图6 测试集准确率变化

Fig.6 Change in accuracy of test set

为了探究高度差产生的原因，接下来对是否使用数据增强，以及提出的三种改进方法对实验结果的影响进行讨论。

3.5 数据增强对实验结果的影响

利用改进后的Res2Net-50模型分别在经过数据增强与未经过数据增强的淡水鱼类图像数据集上进行实验，结果如表1所示。从表1的结果可以看出，数据集在经过数据增强后，其训练集与测试集的准确率均得到了提升，在训练集上提升了3.8%，在测试集上提升了3.3%。并且经过数据增强后的实验结果较未进行数据增强的实验结果，其测试集与训练集准确率相差更小，这说明在经过垂直翻转、水平旋转一定角度等一系列数据增强操作后，增强了模型的泛化能力与鲁棒性，并且改善了模型在该淡水鱼类图像数据集上存在的过拟合问题。

表1 数据增强对实验结果的影响

Tab.1 The impact of data augmentation on experimental results

数据集	训练集准确率/%	测试集准确率/%
未经过数据增强	91.97	93.04
经过数据增强	95.77	96.34

3.6 不同网络模型对实验结果的影响

为了综合验证本文提出的改进方案对淡水鱼类图像分类的影响，同时为了验证改进Res2Net-50模型分类效果，本文选取了几种常用于图像分类的网络进行多组对比实验。对比实验的软硬件环境完全相同，且训练超参数设置、图像预处理情况完全相同，分类结果如表2所示。从表2的对比结果来看，针对淡水鱼类图像识别，本文提出的改进Res2Net-50模型在测试集上达到了最好的分类效果，准确率为96.34%，相较于AlexNet、SE_Res2Net、DenseNet-121、Xception等模型，准确率分别提高了8.43%、4.76%、6.23%、3.3%，

同时模型训练所需的计算开销并没有明显增加。实验结果表明：通过替换卷积核加深网络深度与添加平均池化层减少特征信息的丢失，更改激活函数减少神经元坏死，添加注意力机制增加网络提取信息的效率与准确性，可以很大程度上提高模型分类准确率，本文提出的改进Res2Net-50淡水鱼类图像分类模型是一种性能优良的模型。

表2 不同卷积神经网络实验结果对比

Tab.2 Experimental result comparison of different convolutional neural networks

模型	测试集准确率/%	模型训练参数	浮点运算数/G
AlexNet	87.91	60.97×10^6	0.7
SE_Res2Net	91.58	26.19×10^6	4.29
DenseNet-121	90.11	6.96×10^6	2.88
Xception	93.04	20.82×10^6	4.59
Res2Net-50	92.67	23.67×10^6	4.29
2.2.1方法	95.24	23.67×10^6	4.28
2.2.1方法+2.2.2方法	95.60	23.67×10^6	4.28
本文改进方法	96.34	23.69×10^6	4.52

4 结论(Conclusion)

本文提出了一种改进Res2Net-50模型的方法，即将原始模型中的第一层 7×7 卷积核使用更小的 3×3 卷积核进行代替，在加深网络深度的同时减少网络参数，并且在下采样部分添加平均池化层减少特征信息的丢失，接着将模型的ReLU激活函数更改为CELU激活函数，提高模型的表达能力，之后加入SA模块，从通道层面与空间层面量化特征图的重要性。将改进的模型应用在淡水鱼类图像数据集上达到了96.34%的准确率，相比于其他模型获得了更好的效果。实验结果表明，本文提出的改进Res2Net-50模型可以显著提高淡水鱼类图像分类准确率，具有更强的鲁棒性。

参考文献(References)

- [1] 万鹏,赵竣威,朱明,等.基于改进ResNet50模型的大宗淡水鱼类识别方法[J].农业工程学报,2021,37(12):159-168.
- [2] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE Computer Society, 2016:770-778.
- [3] GAO S, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture[J/OL]. (2021-1-27) [2021-12-21]. <https://arxiv.org/abs/1904.01169>.
- [4] ZHANG Q L, YANG Y B. Sa-net: Shuffle attention for deep convolutional neural networks[J/OL]. (2021-1-30) [2021-12-

- 21]. <https://arxiv.org/abs/2102.00240>.
- [5] BARRON J T. Continuously differentiable exponential linear units[J/OL]. (2017-4-16) [2021-12-21]. <https://arxiv.org/abs/1704.07483>.
- [6] 徐沁,梁玉莲,王冬越,等.基于SE-Res2Net与多尺度空谱融合注意力机制的高光谱图像分类[J].计算机辅助设计与图形学学报,2021,33(11):1726-1734.
- [7] WU Y, HE K. Group normalization[J/OL]. (2018-6-11) [2020-12-21]. <https://arxiv.org/abs/1803.08494>.
- [8] 冀中,赵可心,张锁平,等.基于空间变换双线性网络的细粒度鱼类图像分类[J].天津大学学报(自然科学与工程技术版),2019,52(05):475-482.
- [9] 薛永杰,巨志勇.基于改进AlexNet的鱼类识别算法[J].电子科技,2021,34(04):12-17.
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. (2015-4-10) [2021-12-21]. <https://arxiv.org/abs/1409.1556>.
- [11] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks[C]// CVPR Organizing Committee. Proceedings of the IEEE Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Computer Society, 2019:558-567.

作者简介:

赵正伟(1980-),男,硕士,高级实验师.研究领域:物联网技术,图像处理.

朱宏进(1998-),男,硕士生.研究领域:深度学习,图像处理.

杨根滕(1998-),男,硕士生.研究领域:数据挖掘,深度学习.

王金坤(1997-),男,硕士生.研究领域:物联网技术,图像处理.

(上接第36页)

- [4] RAVISHANKAR S, BRESLER Y. MR image reconstruction from highly undersampled k-space data by dictionary learning[J]. IEEE Transactions Med Imaging, 2011, 30(5): 1028-1041.
- [5] TREMOULHEAC B, DIKAIOS N, ATKINSON D, et al. Dynamic MR image reconstruction-separation from undersampled k-t-space via low rank plus sparse prior[J]. IEEE Transactions Med Imaging, 2014, 33(8):1689-1701. ◆
- [6] PRUESSMANN K P, WEIGER M, SCHEIDEGGER M B, et al. Sense: Sensitivity encoding for fast MRI[J]. Magnetic Resonance in Medicine, 1999, 42(5):952-962.
- [7] JUNG H, SUNG K, NAYAK K S, et al. K-T Focuss: A general compressed sensing framework for high resolution dynamic MRI[J]. Magnetic Resonance in Medicine, 2009, 61(1): 103-116.
- [8] LINGALA S G, HU Y, DIBELLA E, et al. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR[J]. IEEE Transactions on Medical Imaging, 2011, 30(5): 1042-1054.
- [9] AWATE S P, DIBELLA E V R. Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity[C]// Alejandro F, Andrés S, Raimund O, et al. 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). Barcelona, Spain: IEEE, 2012:318-321.
- [10] CABALLERO J, PRICE A N, RUECKERT D, et al. Dictionary learning and time sparsity for dynamic MR data reconstruction[J]. IEEE Transactions on Medical Imaging, 2014, 33(4):979-994.
- [11] CHANDRASEKARAN V, SANGHAVI S, PARRILO P A, et al. Rank-sparsity incoherence for matrix decomposition[J]. SIAM Journal on Optimization, 2011, 21(2):572-596.
- [12] KAFALI S G, SHIH S F, RUAN D, et al. Adaptive locally low rank and sparsity constrained reconstruction for accelerated dynamic MRI[C]// MATHEWS J, JONG C Y. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa, USA: IEEE, 2020:930-934.
- [13] 马杰,王晓云,张志伟,等.一种基于全变分正则化低秩稀疏分解的动态MRI重建方法[J].光电子·激光,2016,27(1):87-96.
- [14] LUSTIG M, PAULY J M. Spirit: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space[J]. Magnetic Resonance in Medicine, 2010, 64(2):457-471.
- [15] 史加荣,郑秀云,周水生.矩阵补全算法研究进展[J].计算机科学,2014,41(4):13-20.

作者简介:

杨青海(1995-),男,硕士生.研究领域:图像处理.

杨敏(1969-),男,博士,副教授.研究领域:图像处理与模式识别.