

DOI:10.16136/j.joel.2022.11.0075

多尺度混合注意力胶囊网络的海洋鱼类识别

许学斌^{1,2*}, 刘 燊 莲^{1,2}, 路 龙 宾^{1,2}, 刘 晨 光^{1,2}

(1. 西安邮电大学 计算机学院, 陕西 西安 710121; 2. 西安邮电大学 陕西省网络数据分析与智能处理重点实验室, 陕西 西安 710121)

摘要: 针对胶囊网络 (capsule network, CapsNet) 特征提取结构单一和数据处理中参数量过大的问题, 提出多尺度混合注意力胶囊网络模型。首先, 在网络初始端添加不同尺度的卷积核来多角度提取特征, 并引入混合注意力机制, 通过聚焦更具分辨性的特征区域来降低复杂背景干扰。其次, 采用局部剪枝算法优化动态路由, 减少参数量, 缩短模型训练时间。最后, 在海洋鱼类数据集 F4K (Fish4Knowledge) 上验证, 结果表明, 与传统残差网络 (residual network50, ResNet-50)、双线性网络 (bilinear convolutional neural network, B-CNN)、分层精简双线性注意力网络 (spatial transformation network and hierarchical compact bilinear pooling, STN-H-CBP) 以及 CapsNet 模型相比, 该算法识别精度为 98.65%, 比 ResNet-50 模型提升了 5.92%; 训练时间为 2.2 h, 相比于 CapsNet 缩短了近 40 min, 验证了该算法的可行性。

关键词: 胶囊网络 (CapsNet); 图像识别; 动态路由算法; 注意力机制; 多卷积核

中图分类号: TP391 **文献标识码:** A **文章编号:** 1005-0086(2022)11-1158-07

Identification of marine fish using multi-scale mixed attention capsule network

XU Xuebin^{1,2*}, LIU Shenlian^{1,2}, LU Longbin^{1,2}, LIU Chenguang^{1,2}

(1. School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an, Shaanxi 710121, China; 2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts & Telecommunications, Xi'an, Shaanxi 710121, China)

Abstract: A multi-scale hybrid attention capsule network (CapsNet) model is proposed to solve the problem of insufficient feature extraction due to single feature extraction structure in CapsNet and excessive amount of parameters in data processing. First, convolution kernels of different scales are added at the initial end of the network to extract features at multiple angles, and channel attention (CA) mechanism and spatial attention (SA) mechanism are introduced to reduce complex background interference by focusing on features of more resolved regions. Second, a local pruning algorithm is adopted to optimize the dynamic routing algorithm, which reduces calculation parameters and training time. Finally, validation on open marine fish data set F4K (Fish 4 Knowledge) shows that the model recognition accuracy in this paper is 98.65% compared with traditional residual network50 (ResNet-50), bilinear convolutional neural network (B-CNN), spatial transformation network and hierarchical compact bilinear pooling (STN-H-CBP) and CapsNet models, 5.92% higher than ResNet-50 model; The training time is 2.2 h, which is nearly 40 min shorter than that of CapsNet, which verifies the feasibility of the proposed algorithm.

Key words: capsule network (CapsNet); image recognition; dynamic routing algorithm; attention mechanism; multi-convolution kernel

* E-mail: ccp9999@126.com

收稿日期: 2022-03-21 修订日期: 2022-05-28

基金项目: 国家自然科学基金面上项目 (61673316)、陕西省重点研发计划项目 (2017GY-071, 2018GY-135)、陕西省教育厅项目 (16JK1697)、陕西省技术创新引导计划项目 (2017XT-005)、咸阳市科技计划项目 (2017K01-25-3) 和西安邮电大学研究生创新基金 (CXJJLY202004) 资助项目

1 引言

海洋鱼类资源丰富,是海洋生命的主要构成部分。随着对海洋资源的开采,海洋鱼类生存面临巨大挑战。因此,引入科学技术对海洋鱼类进行识别和监控,对保护海洋鱼类多样性具有重要意义。

传统海洋鱼类识别主要是依赖人工进行特征分析归类^[1],容易受外界因素影响,需专业知识储备,且调参过程复杂,往往只能提取浅层特征,分类精度不高,泛化性差^[2]。随着深度学习的发展,越来越多相关技术应用到鱼类图像识别中^[3]。文献[4]使用卷积神经网络(convolutional neural network, CNN)对鱼类进行识别,与传统识别方式相比,其准确率大幅度提升。文献[5]提出基于 CNN 来构建鱼类识别模型,提高了模型的鲁棒性。但由于 CNN 的池化操作,只重视特征存在与否,并不关注特征姿态,容易丢失矢量信息。文献[6]强调了这两点:一是 CNN 对经过仿射变换的物体鲁棒性差;二是 CNN 不考虑识别目标的空间关系。为了克服 CNN 的缺陷,在 2017 年, SABOUR 等^[6]提出了胶囊网络(capsule network, CapsNet)。区别于 CNN, CapsNet 中摒弃了池化层,其输入输出都是具有方向的神经元,可通过各个特征之间的关系来预测整个物体^[7]。但 CapsNet 在计算和内存消耗方面的代价较大,特征提取结构较浅,仅在简单数据集上识别效果优于其他深度学习网络,对于复杂数据的识别效率并不高^[8]。

针对以上问题,本文提出多阶段混合注意力 CapsNet 模型。将海洋鱼类图像输入模型中,在特征提取层引入大小为 1×1 、 3×3 、 5×5 、 7×7 的卷积核,融合通道注意力(channel attention, CA)和空间注意力(spatial attention, SA),更全面地提取有用信息。最后,当胶囊间耦合系数低于固定阈值时,采取局部剪枝策略,优化动态路由算法,在其反向传播时不对相应的权重进行更新,以达到降低参数的目的。

2 CapsNet

2.1 网络结构

CapsNet 主要是由胶囊组成的,可根据局部信息来预测全局信息,保留了特征的空间关系^[9]。它的网络结构如图 1 所示。

在网络模型中,第 1 层采用普通的卷积层来提取基础特征;第 2 层为主胶囊层,将上层输出结果用 8 组卷积核卷积,得到大小为 $6 \times 6 \times 32$ 的低层胶囊。

第 3 层为数字胶囊层,通过动态路由计算,预测低层胶囊与高层胶囊的相似度,最终输出大小为 10×16 的向量,每个向量的值代表存在概率,方向为特征的姿态^[10]。

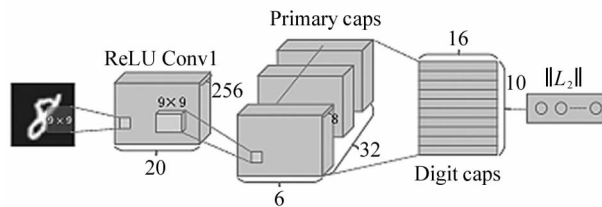


图 1 CapsNet 基本结构图

Fig. 1 Basic structure of capsule network

2.2 CapsNet 工作原理

整个胶囊运作过程是将输入的第 L 层胶囊 i 经过动态路由算法,最终预测输出第 $L+1$ 层胶囊 j ^[11]。由于胶囊输入输出是具有方向的神经元,又称之为向量神经元。其基本结构如图 2 所示。

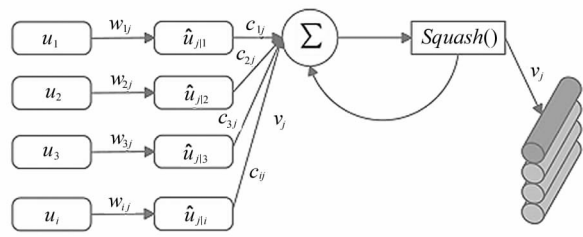


图 2 向量神经元基本结构图

Fig. 2 Basic structure of vector neuron

在动态路由迭代更新中,会产生耦合系数 c_{ij} 。当 c_{ij} 越大时,表明低层特征与高层特征的表达越相似。其更新过程如下:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \quad (1)$$

$$b_{ij} = b_{ij} + u_{j|i} \cdot v_j, \quad (2)$$

式中, b_{ij} 为胶囊 i 连接到胶囊 j 的先验概率,初值设为 0。通过预测向量 $u_{j|i}$ 和输出向量 v_j 的内积来更新 b_{ij} ,进而更新 c_{ij} 。当两个向量相似时,内积为正, b_{ij} 增大, c_{ij} 也随之增大,表明物体存在的概率越大。若两个向量不相似时,则反之。

首先计算胶囊 i 到胶囊 j 的预测向量 $u_{j|i}$,再将 $u_{j|i}$ 与之对应的 c_{ij} 相乘求和作为对下一层的输入:

$$u_{j|i} = w_{ij} \cdot u_i, \quad (3)$$

$$s_j = \sum_i c_{ij} \cdot u_{j|i}, \quad (4)$$

式中, w_{ij} 为变换矩阵, u_i 为低层胶囊。用 squashing 函数挤压 s_j , 最终得到输出向量 v_j 。挤压不会改变

向量的方向,只是确保向量的模长压缩至 0—1 之间。squashing 函数表达式如下:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|}, \quad (5)$$

式中,若 s_j 趋向于无穷大时, v_j 模长接近于 1,表示物体存在概率越大。当 s_j 趋向于 0 时, v_j 模长接近于 0,存在概率越小。同时,采用 Margin 损失函数来减少误差。损失函数的计算式如下:

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2, \quad (6)$$

式中, T_k 表示 k 类存在与否,若 $T_k=1$,表示存在, $T_k=0$,表示不存在。 m^+ 表示惩罚假阳性,取值为 0.9, m^- 为惩罚假阴性,取值为 0.1。 λ 表示超参数,取值

为 0.5。

3 文中算法

基于原模型的优化可分为两部分。其一,在普通卷积层后加入多卷积核提取基本特征,增加图像的感受野。由于海洋图像背景较为复杂,加入混合注意力机制,根据目标任务分配不同通道,通过增加或抑制不同通道的特征权重来聚焦有用信息,减少复杂背景对识别效果的干扰。其二,优化动态路由算法,将低于固定阈值的耦合系数采取剪枝策略,在反向传播时不更新相应的权重矩阵,减少计算量,降低网络计算负载。基本结构如图 3 所示。

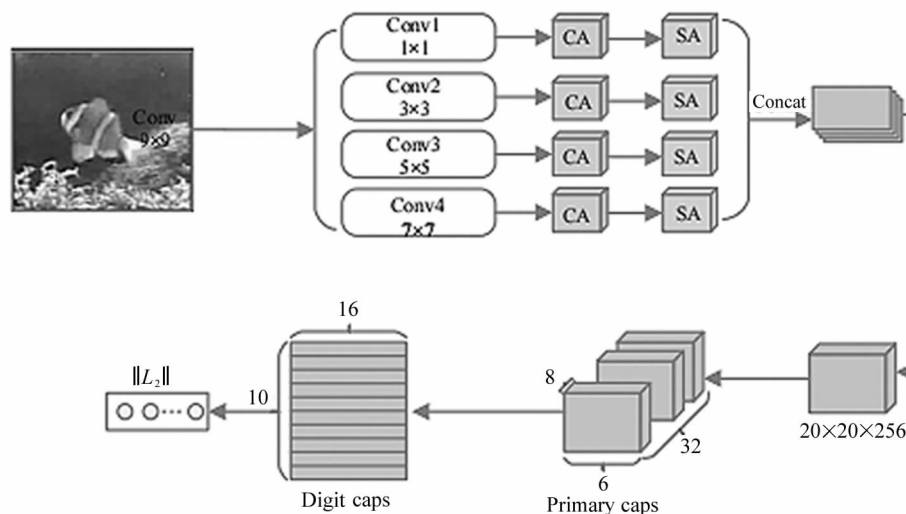


图 3 优化 CapsNet 的鱼类识别算法结构

Fig. 3 Structure of fish identification algorithm of optimized CapsNet

3.1 网络结构

传统 CapsNet 缺乏对低层特征的关注,导致特征提取不充分,识别复杂数据时效果不佳。针对此问题,在网络初始端引入多尺度卷积核,大小分别为 1×1 、 3×3 、 5×5 、 7×7 。4 个卷积核采取的是并行方式来进行特征提取,扩大了特征感受野,并行有利于减少训练时间。

由于海洋环境复杂,对识别造成一定的干扰。因此,在每层卷积核后引入 CA 机制,根据不同的特征来建立不同的通道,按照识别目的增加或抑制每个特征通道的权重。之后再添加 SA 机制,通过聚焦任务相关特征,减少复杂背景对识别精度的影响。

CA 机制通过不同特征来建立相应的通道,根据目标任务去增加或抑制不同通道的权重^[12]。其中,将平均池化和最大池化操作后的特征图丢进共享网

络,最后生成 CA 特征图 M_c 。计算式如下:

$$\begin{aligned} M_c(F) &= \delta(MLP(AvgPool(F)) + \\ &MLP(MaxPool(F))) = \\ &\delta(W_1(W_0(F_{avg}^c)) + \\ &W_1(W_0(F_{max}^c))), \end{aligned} \quad (7)$$

式中, δ 为激活函数, W_1 和 W_0 为 MLP 的权重, F_{avg}^c 为平均池化特征, F_{max}^c 为最大池化特征。其结构如图 4 所示。

SA 机制只关注任务相关区域,寻找网络中最重要的信息进行变换来获取权重^[13]。将经过最大池化 F_{max}^s 和平均池化 F_{avg}^s 操作得到的特征图融合,结合普通的卷积处理,使用 Sigmoid 函数激活。计算式如下:

$$\begin{aligned} M_s(F) &= \\ &\delta(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \end{aligned}$$

$$\delta(f^{7 \times 7}([\bar{F}_{\text{avg}}^s; F_{\text{max}}^s])), \quad (8)$$

式中, f 为不同通道的特征加权。最终得到二维空间注意力特征图 M_s , 并通过编码进行特征强调或者抑制。结构如图 5 所示。

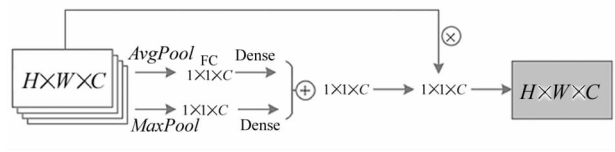


图 4 CA 机制

Fig. 4 Channel attention mechanism

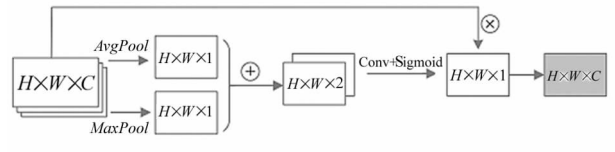


图 5 SA 机制

Fig. 5 Spatial attention mechanism

3.2 优化算法

CapsNet 中动态路由计算过程需要更新大量的参数,非常消耗计算资源且训练时间较长。针对此问题,提出局部剪枝动态路由算法。当子胶囊预测与父胶囊越接近时,耦合系数越大;反之,耦合系数越小。当耦合系数 c 小于固定阈值时,则采取剪枝策略,将其对应的权重参数 w_{ij} 置为 0,反向传播时停止对该参数进行更新,减少计算的参数量与时间。剪枝的示意图 6 如下所示。

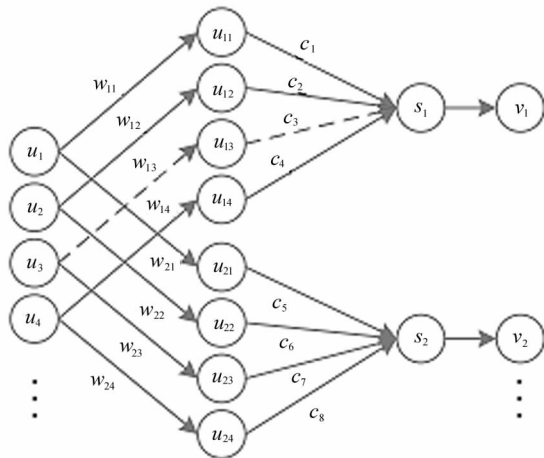


图 6 剪枝示意图

Fig. 6 Pruning diagram

改进后的动态路由算法步骤如下:

算法 1 优化后的动态路由算法

1) 程序路由

2) 初始化耦合系数对数

3) for r to:

4) $c[c < \text{threshold}] = 0$

5) 对在 l 和 $l+1$ 层之间的胶囊 i 和胶囊 j

$$c_{ij} \leftarrow \text{softmax}(b_{ij})$$

$$b_{ij} \leftarrow b_{ij}^{(x,y)} + \hat{u}_{j|i}^{(x,y)} v_j^{(p,q)}$$

6) 对于在 $L+1$ 层上的胶囊 j

$$s_j^{(p,q)} \leftarrow \sum_t c_{ij}^{(x,y)} \hat{u}_{j|i}^{(x,y)}, j \in \{1 : T^{L+1}\}$$

$$v_j^{(p,q)} \leftarrow \text{squash}(s_j^{(p,q)})$$

$$w_{ij} \leftarrow \text{BackPropagation}(w_{ij})$$

4 实验过程与结果

4.1 数据预处理

数据集 F4K(Fish4Knowledge)是从海底观测平台实时录像中拍摄截取下来的海洋鱼类数据集。它包含 23 个种类,共 27 370 张图像。由于海底环境恶劣,拍摄时受到其他生物、光线强度和位置等因素的影响,对识别造成一定的干扰。因此,先给图像做预处理,减少复杂背景对实验的影响。为了更接近真实世界,通过空间变换、中心裁剪出 100×100 的区域、提升亮度等操作来进行图像增强,实现了图像数据的平稳输入,再将数据做归一化处理,避免产生欠拟合或过度拟合。图 7 是预处理之后的部分图像。

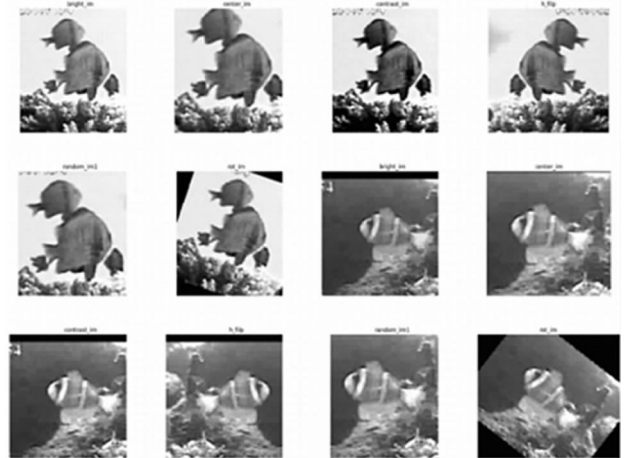


图 7 处理后的部分图像

Fig. 7 Processed partial image

4.2 实验参数的选取

本实验将 F4K 数据集集中的 6/7 作为训练集,1/7 为测试集,采用先粗调再细调的方式对新增加的网

络进行训练。微调后的模型部分参数设置如下: batch_size=8,迭代次数为30次。

学习率的选择决定着网络模型能否找到全局最小值,进而找到全局最优参数。由表1可知,当学习率为0.001时,改进的模型比原模型的测试精度高出5.39%。当学习率为0.01和0.0001时,改进的模型验证精度都高于原模型,但还没达到全局最优,获得的结果均不理想。造成这种结果的原因,一是学习率过大时,模型往往会忽略数据中的某些特征直接跳到下一阶段,导致网络模型发散,无法收敛;二是学习率过小时,网络模型收敛速度慢,且在训练中过程中,极易把局部最优点当成全局最优解进行收敛,致使识别精度下降。可见,采取学习率为0.001时,模型效果最佳。

表1 CapsNet与本文模型精度比较
Tab.1 Comparison of accuracy between CapsNet and proposed model

Model number	Model name	Learning rate	Training accuracy/%	Test accuracy/%
1	CapsNet	0.01	89.54	82.35
2	CapsNet	0.001	97.63	92.73
3	CapsNet	0.0001	88.67	81.43
4	Our	0.01	98.43	97.65
5	Our	0.001	99.68	98.12
6	Our	0.0001	93.32	91.22

剪枝阈值的选择会直接对模型产生影响,尤其对训练时间的影响。若阈值过大,会将低层特征和高层特征表达相似时的部分参数裁剪,导致识别精度下降。若阈值过小,则反之。在本实验中,剪枝阈值选取固定值。根据表2可知,随着阈值的增大,识别精度大体上变化较小,模型训练时间随着参数剪枝越多而缩短。将阈值为0.02时与阈值为0.03、0.04、0.05相比,训练时间变化差别不大,但识别精度并没有达到最优效果;将阈值为0.02与阈值0.01相比,识别精度相差无几,但训练时间缩短了将近20 min。因此,该实验的局部剪枝阈值为0.02。

表2 局部剪枝阈值的选取
Tab.2 Selection of local pruning threshold

Local pruning threshold	Improve the test accuracy of the model/%	Training time/h
0	98.36	2.5
0.01	98.64	2.4
0.02	98.65	2.2
0.03	98.18	2.1
0.04	98.21	2.1
0.05	98.19	2.0

4.3 实验结果分析

4.3.1 消融实验

通过消融实验来验证该算法对海洋鱼类图像的识别性能。结果如表3所示,a-CapsNet代表在原模型中融合多卷积核,b-CapsNet代表在CapsNet网络添加CA机制,c-CapsNet为在CapsNet网络添加SA机制,d-CapsNet是多层卷积核、CA和SA相互融合到CapsNet模型中。实验表明,4种不同的融合模型识别精度均比原模型高。但CA机制更注重给不同特征加权重,容易忽略特征空间位置关系;SA机制注重目标相关区域,而容易忽略其他特征。这导致b-CapsNet和c-CapsNet识别精度有所提升,但效果并不是最佳的。因此,将卷积核和注意力机制融合到原CapsNet中,在海鱼数据集F4K上进行验证,精确度达到98.61%。

表3 消融实验结果对比
Tab.3 Comparison of ablation results

Data set	Model	Recognition accuracy/%
F4K	CapsNet	92.73
	a-CapsNet	94.83
	b-CapsNet	97.47
	c-CapsNet	97.29
	d-CapsNet	98.61

在优化CapsNet特征提取器的基础上,在动态路由算法过程中采取局部剪枝策略来减少计算资源消耗。结果如表4所示,在优化动态路由算法之后,识别精度相对于CapsNet提高了5.92%,与只优化特征提取模型d-CapsNet相比无明显提高,但训练时间缩短了近30 min,有效减少了训练时间。

表4 改进动态路由算法
Tab.4 Improving dynamic routing algorithms

Model	Recognition accuracy/%	Training time/h
CapsNet	92.73	2.6
d-CapsNet	98.61	2.5
Algorithm in this paper	98.65	2.2

4.3.2 实验结果

表5为残差网络(residual network50, ResNet-50)、双线性网络(bilinear convolutional neural network, B-CNN)、分层精简双线性注意力网络(spatial transformation network and hierarchical compact bilinear pooling, STN-H-CBP)以及CapsNet模型与本

文提出的 CapsNet 优化模型的实验结果对比。其中传统 CNN 均使用大量卷积层和池化层。可看出本文算法较上述模型至少提高 0.61%，验证了本模型的有效性。

表 5 实验结果对比

Tab. 5 Comparison of experimental results

Data set	Model	Recognition accuracy/%
F4K	ResNet-50 ^[14]	93.68
	B-CNN ^[15]	97.33
	STN-H-CBP ^[16]	98.04
	CapsNet	92.73
	Algorithm in this paper	98.65

图 8 是优化模型的实验结果可视化曲线。子图(a)为损失函数收敛曲线,子图(b)为测试精度曲线。其中,实线代表训练精度,虚线代表测试精度。由图

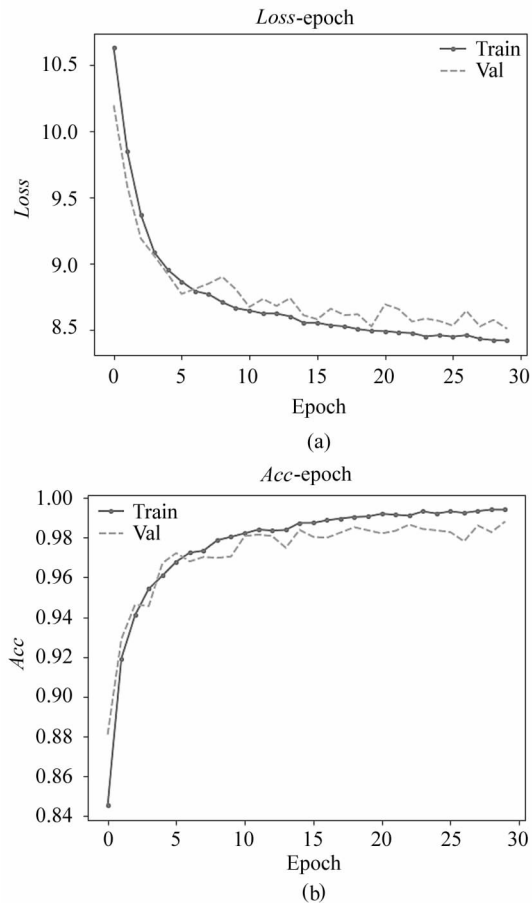


图 8 实验结果可视化:(a) 损失函数收敛曲线;
(b) 测试精度曲线

Fig. 8 Visualization of experimental results:
(a) Loss function convergence curves;
(b) Test accuracy curves

8 可见,从 0—10 轮之间损失函数快速收敛,测试精度快速上升。到 epoch 为 25 时,损失函数和验证精度趋于平稳,验证了算法的可行性。

5 结 论

本文对 CapsNet 的特征提取结构和动态路由算法进行了一定的改进,有效地解决了特征提取结构不充分的问题,减少了运算过程中的参数量,降低了计算代价。在特征提取层中,添加多卷积核和混合注意力机制,多角度提取特征,同时有效地抑制无用信息,提高模型特征提取能力。在计算效率方面,采取剪枝策略来优化动态路由算法,降低网络运行时间。此外,CapsNet 较好保留特征的空间关系,解决了 CNN 的缺陷。通过实验可知,本模型在海鱼识别中优于其他传统的识别模型。

参考文献:

[1] VAN TREECK R, VAN WICHELEN J, WOLTER C. Fish species sensitivity classification for environmental impact assessment, conservation and restoration planning [J]. *Science of the Total Environment*, 2020, 70 (8): 135173.

[2] ZANG X, YIN T, HOU Z, et al. Deep learning for automated detection and identification of migrating American eel *Anguilla rostrata* from imaging sonar data [J]. *Remote Sensing*, 2021, 13(14): 2671.

[3] SEBASTIEN V, DAVID M, MARC C, et al. A deep learning method for accurate and fast identification of coral reef fishes in underwater images [J]. *Ecological Informatics*, 2018, 48: 238-244.

[4] HRIDAYAMI P, PUTRAI K G D, WIBAWAK S. Fish species recognition using VGG16 deep convolutional neural network [J]. *Journal of Computing Science and Engineering*, 2019, 13(3): 124-130.

[5] ZHANG J, ZENG G, QIN R. Fish recognition method for submarine observation video based on deep learning [J]. *Journal of Computer Applications*, 2019, 39(2): 376-381.

[6] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules [C]// *Advances in Neural Information Processing Systems*, December 4-9, 2017, Long Beach, USA. Red Hook: Curran Associates, 2017: 3856-3866.

[7] XIANG C, ZHANG L, TANG Y, et al. MS-CapsNet: a novel multi-scale capsule network [J]. *IEEE Signal Processing Letters*, 2018, 25(12): 1850-1854.

[8] HUANG R, LI J, LI W, et al. Deep ensemble capsule network for intelligent compound fault diagnosis using multi-

- sensory data[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(5): 2304-2314.
- [9] CHEN J, LIU Z. Mask dynamic routing to combined model of deep capsule network and U-Net[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2653-2664.
- [10] LI X, KONG X, ZHANG J, et al. A study on fault diagnosis of bearing pitting under different speed condition based on an improved inception capsule network[J]. Measurement, 2021, 181: 109656.
- [11] CHEN M S, WU R R, ZHANG Z G, et al. Traffic sign classification model based on improved capsnet [J]. Computer Application Research, 2020, 37(S2): 367-368+371.
陈名松, 吴冉冉, 张泽功, 等. 基于改进 CapsNet 的交通标志分类模型[J]. 计算机应用研究, 2020, 37(S2): 367-368+371.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [13] ZHAO L, YANG F, BU L, et al. Driver behavior detection via adaptive spatial attention mechanism[J]. Advanced Engineering Informatics, 2021, 48: 101280.
- [14] JIA Y X, FAN S C, YI X M. Research on fish recognition based on salience enhancement and transfer learning[J]. Fishery Modernization, 2020, 47(1): 9.
贾宇霞, 樊帅昌, 易晓梅. 基于显著性增强和迁移学习的鱼类识别研究[J]. 渔业现代化, 2020, 47(1): 9.
- [15] LUO J H, WU J X. Overview of fine-grained image classification based on depth convolution feature[J]. Journal of Automation, 2017, 43(8): 13.
罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, 43(8): 13.
- [16] DONG S J, LIU W, CAI W W, et al. Fish recognition based on hierarchical compact bilinear attention network [J]. Computer Engineering and Application, 58(5): 186-192.
董绍江, 刘伟, 蔡巍巍, 等. 基于分层精简双线性注意力网络的鱼类识别[J]. 计算机工程与应用, 58(5): 186-192.

作者简介:

许学斌 (1974—), 男, 副研究员, 博士/博士后, 主要研究方向人工智能、生物特征识别。