第 39 卷 第 2 期
2020 年 3 月

应用声学
Journal of Applied Acoustics

Vol. 39, No. 2
March, 2020

◇ 研究报告 ◇

多特征融合的鸟类物种识别方法*

谢将剑[1,2†] 杨 俊[1] 邢照亮[3] 张 卓[3] 陈 新[3]

(1 北京林业大学工学院 北京 100083)

(2 林业装备与自动化国家林业和草原局重点实验室 北京 100083)

(3 先进输电技术国家重点实验室 (全球能源互联网研究院有限公司) 北京 102211)

摘要: 深度学习输入特征的选择直接影响其分类性能,为了进一步提高基于深度学习的鸟类物种识别模型的分类性能,该文提出一种多特征融合识别方法。该方法首先通过短时傅里叶变换、梅尔倒谱变换和线性调频小波变换分别计算得到鸣声信号的3种语图样本集,然后分别利用3种语图样本集训练3个基于VGG16迁移的单一特征模型,将3个模型的输出进行自适应加权求和实现融合,并修正了加权交叉熵函数以克服样本不平衡的问题,最后对语图进行分类实现鸟类物种的识别。以ICML4B鸣声库的35种鸟类为研究对象,对比了4种模型的平均识别准确率(MAP),结果表明特征融合模型较单一特征模型的MAP最大提高了0.307;选择输入语图的持续时间分别为100 ms、300 ms以及500 ms,对比不同持续时间下4种模型的测试MAP值,结果表明持续时间为300 ms时4种模型的MAP值均为最高;对比了不同信噪比下4种模型的识别效果,多特征融合模型的识别准确率随着信噪比的下降降低最少。说明在选择合适的语图持续时间后,该文提出的特征融合模型能得到更高的识别准确率,具有一定的抗噪能力,且训练参数少,更适合于少样本鸟类的识别。

关键词: 鸟类物种识别;深度卷积神经网络;多特征融合

中图法分类号: TP181 **文献标识码:** A **文章编号:** 1000-310X(2020)02-0199-08

DOI: 10.11684/j.issn.1000-310X.2020.02.005

Bird species recognition method based on multi-feature fusion

XIE Jiangjian[1,2] YANG Jun[1] XING Zhaoliang[3] ZHANG Zhuo[3] CHEN Xin[3]

(1 *School of Technology, Beijing Forestry University, Beijing* 100083, *China*)

(2 *Key Lab of National Forestry and Grassland Administration for Forestry Equipment and Automation, Beijing* 100083, *China*)

(3 *State Key Laboratory of Advanced Transmission Technology, Global Energy Interconnection Research Institute Co. Ltd., Beijing* 102211, *China*)

Abstract: The choice of input feature directly affects the classification performance of the deep learning, a multi-feature fusion recognition method was proposed to improve the classification performance of the bird species recognition model. In this method, firstly three kinds of spectrogram samples of vocalization signals were calculated through short time Fourier transform, Mel-frequency cepstrum transform and Chirplet transform respectively, then three single feature models which based on VGG16 transfer learning were trained using these three kinds of spectrogram samples accordingly, modified weighted cross entropy function was used to fix the problem of imbalanced data set, the outputs of three models were fused to classify the spectrograms and realize

2019-06-16 收稿; 2019-11-28 定稿

*国家自然科学基金资助项目 (31670553), 国家电网公司科技项目 (SGGR0000WLJS1801082), 国家重点研发项目 (2017YFC1403503), 中央高校基本科研业务费专项 (2016ZCQ08)

作者简介: 谢将剑 (1988–), 男, 江西鹰潭人, 博士, 副教授, 研究方向: 深度学习在林业生态环境检测中的应用。

†通信作者 E-mail: shyneforce@bjfu.edu.cn

the recognition of bird species. Taken the 35 kinds of bird in ICML4B database for study subject, the MAPs were compared, results show that the mean average precision (MAP) of feature fusion model is highest increased by 0.307 contrast to the single feature model; Three spectrogram durations, 100 ms, 300 ms and 500 ms were chosen to compare the test MAP of four models, the results reveal that the 300 ms duration is the best; the precision of 4 models with different SNR were compared, the precision reduction of feature fusion model as the SNR decreased is the least. The proposed model can achieve better performance with suitable duration, have anti-noise ability in some degree, and the trainable parameters are less, which is more suitable for birds with little samples.

Keywords: Bird species recognition; Deep convolutional neural networks; Multi-feature fusion

0 引言

鸟类鸣声具有一定的稳定性和明显的物种鉴别特征，是鸟类物种识别的主要方式之一[1]。利用鸟鸣声语图中鸣声区域的图像特征可以区分鸟类物种[2−3]，进而应用于鸟类物种的调查与监测，具有高效率、非损伤、低干扰、大范围等优势，应用前景巨大[4]。基于鸣声的鸟类物种识别的关键在于提取合适的鸟鸣声差异特征、选择高性能的分类器对差异特征进行分类。深度学习具有较强的自动学习特征和进行分类的能力[5−6]，在基于鸣声的鸟类物种识别方法中得到了广泛的研究。Chakraborty 等[7] 利用深度神经网络 (Deep neural network, DNN) 采用音频信号的梅尔滤波能量系数作为输入，实现了基于鸣叫的鸟类识别，最佳识别准确率达到 98.48%。Piczak[8] 以音频信号的梅尔频域谱图为输入，对比了 3 种不同结构的深度卷积神经网络 (Deep convolutional neural network, DCNN) 的识别效果，结果表明输入谱图的大小、网络的层数以及网络结构都会对识别效果产生影响。Martinsson[9] 利用 18 层的深度残差神经网络 (Deep residual neural network, DRNN) 对 BirdCLEF 竞赛的鸣声样本进行识别，平均识别精度达到 53.8%，相比官方提供的深度卷积网络识别方法低 2%。谢将剑等[10] 研究了 3 种不同语图作为输入时，利用 VGG16 模型进行鸟类物种识别的性能，结果表明基于线性调频小波变换生成的语图作为输入，相比其他语图作输入时，识别准确率和效率都得到了改善。以上研究表明，选择合适的语图作为输入、DCNN 作为识别模型，可以得到较好的识别性能。

通过多特征融合可以提高模型的分类性能[11]，本文提出一种基于深度卷积神经网络的多特征融合模型，该模型利用短时傅里叶变换 (Short-time Fourier transform, STFT)、梅尔倒谱变换 (Mel frequency cepstrum transform, MFCT) 和线性调频小波变换 (Chirplet transform, CT) 分别计算得到鸣声信号的语图，基于 3 种语图样本集分别训练 3 个基于 VGG16 迁移的单一特征模型，并利用自适应线性加权对 3 种特征进行融合，最终基于融合特征实现鸟类物种的识别。以鸣声库 ICML4B 的鸟鸣声为研究对象，通过对比实验验证了本文提出模型的优越性。

1 鸟鸣声语图样本集

1.1 鸟鸣声信息

本文采用的鸣声库是法国国立自然博物馆提供的 ICML4B 鸣声库。该鸣声库共包含 35 种鸟类鸣声，具体的物种信息如表 1 所示。每种鸟类各包含 1 个音频信号。每个鸣声音频信号均为持续时间 30 s、采样频率 44.1 kHz、16 bit 输出、WAV 格式的数字信号，信噪比在 20 ∼ 60 dB 之间[12]。

1.2 鸣声信号的处理

鸣声信号的预处理主要包括预加重、分割、分帧以及加窗。预加重用于补偿鸟鸣声在传播时高频成分的衰减，采用一阶高通滤波器来实现，预加重系数取 0.95。

ICML4B 鸣声库中鸟鸣声信号的信噪比较高，采用能量阈值法对鸣声进行分割。分割前需要对信号进行无重叠的分帧，为了最大限度保留鸣声信号，帧长选择为 50 ms。计算每帧信号的能量后，将能量大于最大能量 60% 的帧认为是鸣声区域，予以保留，去除其他非鸣声区域，实现鸣声信号的分割，使各有效鸣声段连续，同时可以降低背景噪声的影响。

鸣声信号是一种典型非平稳随机信号，在对分割后的信号进行时频变换前，需要对信号分帧。同

时对每帧信号加窗，以避免分帧后信号两端可能造成的不连续性。本文选择帧长为 50 ms，重叠 30%，窗函数为汉明窗。

表 1　ICML4B 鸣声信号的信息

Table 1　Detail information of ICML4B bird vocalization signal

物种	语图数
北长尾山雀 Aegithalos caudatus	35
云雀 Alauda arvensis	54
黑额黑雁 Anthus trivialis	39
加拿大黑雁 Branta canadensis	30
欧金翅雀 Carduelis chloris	44
短趾旋木雀 Certhia brachydactyla	16
斑尾林鸽 Columba palumbus	39
小嘴乌鸦 Corvus corone	24
杜鹃 Cuculus canorus	26
大斑啄木鸟 Dendrocopos major	34
黄鹀 Emberiza citrinella	21
欧亚鸲 Erithacus rubecula	29
苍头燕雀 Fringilla coelebs	24
松鸦 Garrulus glandarius	13
夜莺 Luscinia megarhynchos	27
白鹡鸰 Motacilla alba	40
金黄鹂 Oriolus oriolus	17
青山雀 Parus caeruleus	42
大山雀 Parus major	23
沼泽山雀 Parus palustris	37
蓝孔雀 Pavo cristatus	41
环颈雉 Phasianus colchicus	29
红尾鸲 Phoenicurus phoenicurus	37
棕柳莺 Phylloscopus collybita	37
绿啄木鸟 Picus viridis	25
林岩鹨 Prunella modularis	29
普通鸸 Sitta europaea	35
灰斑鸠 Streptopelia decaocto	48
灰林鸮 Strix aluco	19
紫翅椋鸟 Sturnus vulgaris	25
黑顶林莺 Sylvia atricapilla	32
鹪鹩 Troglodytes troglodytes	39
乌鸫 Turdus merula	34
欧歌鸫 Turdus philomelos	41
槲鸫 Turdus viscivorus	34

1.3　鸣声语图的计算

语图中的鸣声区域可以看成是图片中的特殊“物体”，通过识别鸣声区域的特征，可以实现鸟鸣声的分类[3]。为了得到不同的语图特征，选择了 3 种时频变换方法计算语图：

(1) 短时傅里叶变换

STFT 是最常用的一种时频分析方法，它通过时间窗内的一段信号来表示某一时刻的信号特征。计算得到每一帧的时频矩阵，便可以画出对应的语图。

(2) 梅尔频域倒谱变换

人耳听到的声音高低与声音的频率并不成线性正比关系，通常采用 Mel 频率尺度来模拟人耳的听觉特性[13]。鸣声信号经过快速傅里叶变换之后，通过一系列三角形 Mel 频率滤波器组，然后对所有滤波器输出进行对数运算，再进一步做离散余弦变换便可得到梅尔倒谱系数 (Mel frequency cepstrum coefficient, MFCC)。本文计算得到 32 维梅尔倒谱系数后，去掉表征平均值的第 0 维，选择余下的 31 维系数转换成梅尔谱图。

(3) 线性调频小波变换

CT 是一种线性时频表示，可以看成是短时傅里叶变换和小波变换的综合，在表征短时平稳信号时具有明显优势[14]。对每一帧信号进行线性调频小波变换，利用快速 Chirplet 分解算法计算得到小波系数[14]，然后利用小波系数生成语图。

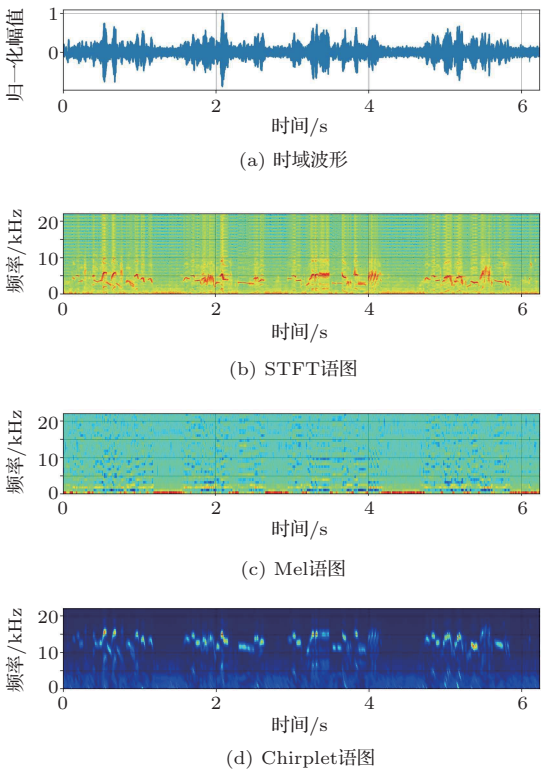

图 1　棕柳莺的鸣声语图

Fig. 1　Spectrograms of Phylloscopus collybita

对鸣声进行预处理后，分别采用以上3种时频变换方法生成语图，图1为棕柳莺的鸣声信号及计算得到的语图，从上到下依次为鸣声信号的时域波形、STFT语图、Mel语图以及Chirplet语图。

1.4 语图样本集的建立

利用语图特征进行分类时，选择音节特征作为输入比选择鸣唱特征的分类效果更好[15]。因此，本文将分割后的鸣声信号继续分帧，得到持续时间为500 ms的鸣声信号，计算其对应的语图，保存成大小为224×224的彩色图像，作为模型的输入，以通过语图图像特征的差异，实现鸟类物种的识别。最终计算得到35种鸟类的鸣声语图数量如表1所示。利用3 种不同的时频变换对鸣声信号进行计算，便可得到3个不同的鸣声语图样本集。

2 多特征融合的鸟类物种识别模型

2.1 基于VGG16的特征迁移模型

DCNN利用多层卷积层和池化层的组合自主学习图像特征，配合全连接层对特征进行分类，进而实现图像的识别。DCNN可以通过局部连接、权值共享及池化操作等有效地降低网络的复杂度，减少训练参数的数目[5,16]。VGG16是一种典型的DCNN，由于其在ImageNet图片分类中的优异性能，在图像识别领域得到了广泛的应用[17−19]。DCNN的模型的参数随着深度的增大而增加，训练过程需要输入更多的已标注样本。如果缺乏足够的已标注样本，训练时容易导致过拟合，无法得到有效的识别模型。

基于迁移学习的思想，利用预训练好的模型作为特征提取器，冻结特征提取模型的参数，训练时不再参与更新，只更新用于分类器的参数，可以大大减小对已标注样本的需求量[20−21]。本文将鸟鸣声的识别等效成对鸣声语图的识别，基于VGG16模型将图像识别问题迁移到基于鸣声语图的鸟类物种识别中。选择ImageNet预训练好的VGG16模型参数作为特征提取模型参数的初始值，通过训练对模型参数进行微调，可以提高训练效率，同时有利于样本数据量小的情况下的模型训练。

2.2 多特征融合模型

由不同的时频变换方法计算得到的不同语图，可以表征鸟鸣声的不同特征。采用不同的语图样本集作为输入时，模型的识别效果不同[15]。对同一对象的不同特征进行融合，得到的特征更全面，有利于提高分类的效果。多特征的融合方法直接影响融合特征的表达能力，目前，常用的多特征融合方法有直接叠加、串行或者并行连接以及加权求和等[11]。前两类方法不能体现不同特征的差异性，同时还扩大了特征的维数，增大了计算量。因此，采用加权求和的方式，引入权重的概念，表征不同的特征在识别过程中的贡献度。

为了充分利用3种语图特征，进一步提升识别性能，首先构建3个不同的基于VGG16的特征迁移模型，分别提取3种语图的特征。然后，将提取的3种特征进行自适应线性加权，一方面实现基于特征的融合，另一方面保持特征维度，可以不增加模型参数。融合后的特征F如式(1)所示：

$$F=\sum_{n=1}^{3}\omega^n y^n, \tag{1}$$

其中，ω^n和y^n分别表示特征n相应的权值和特征向量。不同的权值可以表征不同的特征在识别过程中的贡献度，且满足条件$\sum_{n=1}^{3}\omega^n=1$。该权值参与训练和更新过程，通过迭代自动获得最优的权值。

最后，将融合后的特征输出到2个全连接层和1个Softmax输出层组成的分类器中，基于多特征融合模型的鸟类物种识别流程如图2所示。

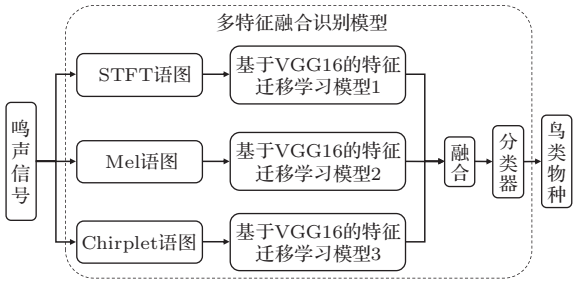

图2 基于多特征融合模型的鸟类物种识别流程

Fig. 2 Procedure of bird species recognition based on multi-feature fusion

在训练时，先利用3种不同的语图分别作为样本集，训练出3个基于VGG16的特征迁移模型；在多特征融合模型中，这3个基于VGG16的特征迁移模型的参数不再参与更新，以克服由于模型增大后带来的参数数量增大进而对样本数量需求增大的缺陷。

3 实验结果与分析

3.1 模型训练设置

实验在Ubuntu16.04 64位系统下，基于深度学习框架Tensorflow1.4.1完成，采用的硬件平台为E5-2620CPU (6×2.1 GHz，32 GB内存)和GTX1080ti GPUs (11 GB内存)的工作站。

实验时，3个鸣声语图样本集均按照8 : 1 : 1的比例被随机划分成训练集、验证集和测试集，用于本文提出的识别模型的训练、验证以及测试，具体的实验流程如图3所示。

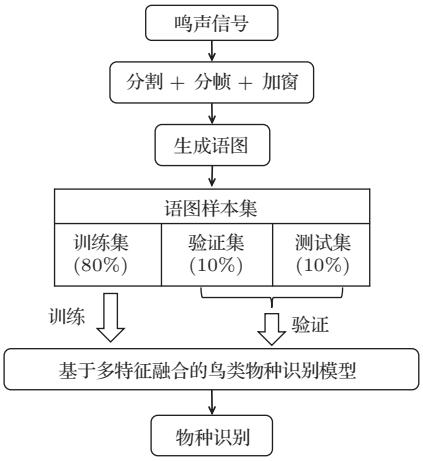

图3　识别模型训练流程图

Fig. 3　Train flow of recognition model

在训练过程中，为了加快数据的处理速度，将数据集分成多个分区(Batch)，适当增大分区大小(Batchsize)可以提高训练的效率。综合考虑到实验用的电脑内存有限，选择分区大小为32。模型训练的参数如表2所示。

表2　训练参数

Table 2　Train parameters

参数类型	值或方法
初始化	正态分布的随机初始化
优化算法	Adam
分区大小	32
学习率	0.001
指数衰减	0.8/100步
损失函数	加权交叉熵函数

从表1可以看出不同鸟类的语图数量相差较大，属于不均衡样本集，不利于DCNN模型的训练。为了克服样本不平衡的问题，引入加权交叉熵损失函数作为模型的损失函数，该方法通过提高少样本类别在损失函数中的权重，进而解决不平衡数据的问题。对于多类别分类时，改进后每个batch中第j ($j=1,2,\cdots,32$)个样本属于第$i(i=1,2,\cdots,35)$类时的交叉熵损失函数值如式(2)所示：

$$\mathrm{WCE}_{ij}=-\omega_i y_i \lg \widehat{y}_i-(1-y_i)\lg(1-\widehat{y}_i), \quad (2)$$

式(2)中，ω_i是类别i的权值；y_i为该样本是否属于类别i的实际标签，属于则为1，不属于则为0；$\hat{y}_i$为该样本预测为类别i的概率。类别i的权值ω_i可由式(3)计算得到：

$$\omega_i=\frac{1-\beta_i}{\beta_i}, \quad (3)$$

式(3)中，β_i是所有训练样本集中属于类别i的样本数占总样本集大小的比例。

进一步得到代价函数为

$$\mathrm{Cost}=\frac{1}{32}\sum_{j=1}^{32}\beta_i\cdot\mathrm{WCE}_{ij}. \quad (4)$$

3.2 单一特征模型和融合模型性能对比

选择持续时间为500 ms的语图作为语图样本集，首先分别训练3个基于VGG16的特征迁移模型，然后将3个模型的特征提取部分进行冻结，通过全连接层组合形成融合模型，再训练融合模型的分类器部分的参数，得到最终的融合模型。

通常利用平均识别准确率(Mean average precision, MAP)来评价识别模型的好坏，本文提出模型的MAP计算公式如式(5)所示：

$$\mathrm{MAP}=\frac{\sum_{q=1}^{35}\mathrm{Ave}P(q)}{35}, \quad (5)$$

式(5)中，q为鸟类物种的编号，Ave$P(q)$为对应物种的识别正确率。

图4为不同模型在验证集上的MAP随着迭代次数增加的变化，Ch代表Chirplet语图特征模型，Mel代表梅尔语图特征模型，Spe代表STFT语图特征模型，Fuse代表融合模型。

从图4中可以看出，融合模型在76次迭代中达到最大MAP值，而其他模型要到250次以后才趋于最大MAP值。融合模型相比单一特征模型达到最大MAP值的时间要更短，说明融合模型的训练效

率要更高。而且对比单一特征模型，融合模型的最大MAP值也最大。

图5为4种模型的平均识别准确率对比，图中的MAP值均为5次运算的MAP平均值。

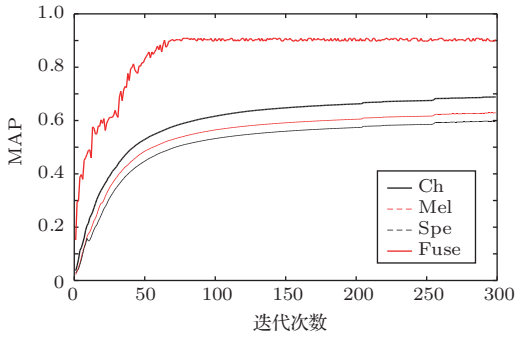

图4　不同模型的验证MAP随着迭代次数的变化

Fig. 4　Variation of validation MAPs with epochs increasing

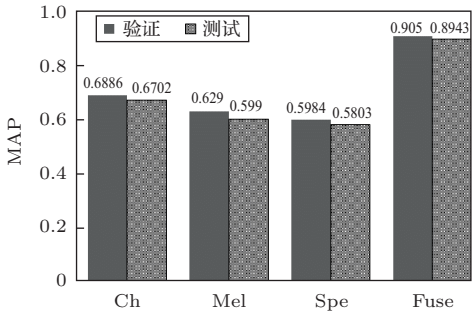

图5　不同模型的验证MAP和测试MAP对比

Fig. 5　Comparison of validation MAP and test MAP with different model

从图5可以看出：(1) 单一特征模型中，Chirplet语图作为输入时的MAP时最大，STFT语图的MAP最小，和文献[14]相吻合；(2) 通过多特征融合后，融合模型的MAP较单一特征模型提升较大，相比STFT语图的提升了30%左右。

综上所述，将不同特征进行融合，再利用分类器来进行分类的方法可以大大提高模型的识别能力，说明本文提出的多特征融合模型是可行的。同时本文提出的融合模型中，待训练的参数只包含分类器的参数，参数数量相对于VGG16模型大大减少，可以降低对样本数量的需求。

3.3　语图不同持续时间的性能对比

输入DCNN的图像大小是固定的，选择不同的语图持续时间，会改变语图中鸟鸣声区域的特征，进而影响识别模型的性能。为了研究语图持续时间对识别性能的影响，选择持续时间为100 ms、300 ms和500 ms，分别计算得到3个语图样本集。按照本文提出的建模方法得到不同持续时间下的单一特征模型和融合模型。

图6为不同持续时间时，不同模型的测试MAP对比。

从图6可以看出，语图持续时间不同时，对于同一个模型，持续时间为300 ms的MAP值最大，100 ms的最小，而且4种模型的变化规律一致。进一步对35种鸟类鸟鸣声的音节持续时间进行了统计分析，得到了不同持续时间的音节数量分布如图7所示，71.7%的音节持续时间在100 ~ 300 ms之间。

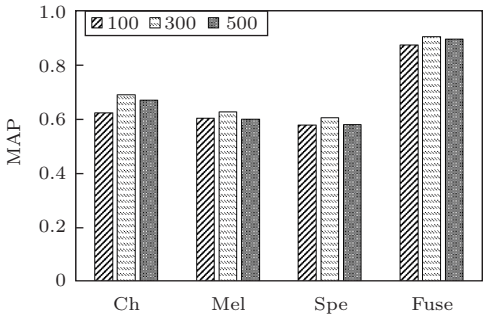

图6　不同持续时间语图下不同模型的测试MAP对比

Fig. 6　Comparison of test MAP with different model and duration

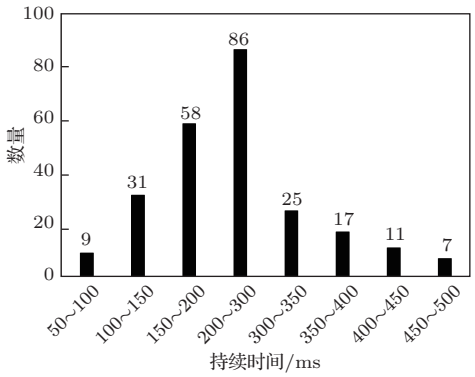

图7　不同持续时间的音节数量分布

Fig. 7　Distribution of syllable quantity with different duration

依据统计结果分析可得，持续时间为300 ms时效果最佳的原因在于：本文提出的方法是基于语图中鸟鸣声区域的图像特征实现语图的分类，达到识别鸟类物种的目的。因此图像当中鸟鸣声区域的完

整性对于准确识别鸟类物种影响较大。鸟鸣声的音节持续时间各不相同，当音节的持续时间小于语图持续时间时，在语图中能够完整显示音节的特征，可以提高识别的准确率。而在完整性保证的基础上持续时间更长时，得到的语图数量减小，等效于训练样本数量下降，导致训练的效果下降。

综上所述，不同语图持续时间会影响模型的识别性能。在数据量足够大时，可以尽量选择较长的持续时间，使每幅语图中的鸣声区域保持完整。如果数据量有限，则需要根据鸟鸣声音节持续时间的分布，选择合适的持续时间。

3.4　不同信噪比时模型的性能对比

为了对比不同信噪比下不同模型的识别性能，重新整理语图持续时间为300 ms时在测试集上的实验结果。按照信噪比大小重新整理成3个子集，为了保证每个集样本数均衡，分为强噪声集(信噪比在20 ∼ 35 dB)、中噪声集(信噪比在35 ∼ 45 dB)以及低噪声集(信噪比在45 ∼ 60 dB)。计算得到不同子集的识别准确率如图8所示。

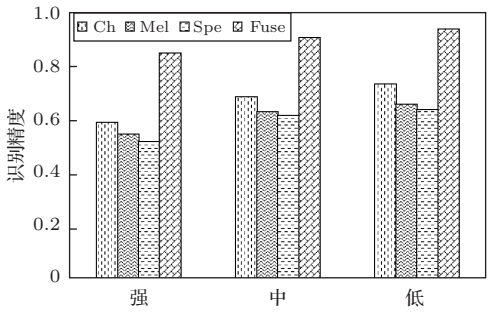

图8　不同信噪比下的模型的识别精度

Fig. 8　Precision under different SNR

从图8中可以看出，随着信噪比的升高，4种模型的识别精度都在下降。进一步计算得到强噪声集和低噪声集上识别精度的相对误差：Ch、Mel、Spe和Fuse的分别为19.64%、17.15%、18.81%和9.73%，3个单一特征模型的下降的数值更大，说明多特征融合模型的抗噪能力较其他3个模型更强。

4　结论

为了进一步提高识别的准确率，本文提出一种基于Chirplet语图、Mel语图以及STFT语图3 种语图特征融合的鸟类物种识别方法。该方法首先建立3个不同语图输入时的基于VGG16特征迁移的单一特征模型，然后将其进行加权求和融合得到特征融合模型。以ICML4B鸣声库的35种鸟类为研究对象，对比了持续时间为500 ms的语图作为输入时4种模型的MAP值，特征融合模型较前3个模型在MAP值和训练效率上均有较大的提升，验证了本文提出的多特征融合模型的可行性及优势；为了研究语图持续时间的影响，选择持续时间分别为100 ms、300 ms以及500 ms的语图作为输入，对比不同模型的MAP值，结果表明持续时间300 ms的MAP值最高。对比了不同模型识别不同信噪比鸣声的识别效果，结果表明多特征融合模型抗噪声的能力最强。因此，根据鸟鸣声的音节持续时间分布，选择合适的语图持续时间，利用本文提出的多特征融合可以提高鸟类物种识别的准确率。而且该融合模型的训练参数少，适合于样本数量小的鸣声数据集的分类和识别，这对于有些珍稀鸟类的识别具有较高的应用价值。

参　考　文　献

[1] Mielke A, Zuberbühler K. A method for automated individual, species and call type recognition in free-ranging animals[J]. Animal Behaviour, 2013, 86(2): 475–482.

[2] Ruiz-Muñoz J F, Castellanos-Dominguez G, Orozco-Alzate M. Enhancing the dissimilarity-based classification of birdsong recordings[J]. Ecological Informatics, 2016, 33: 75–84.

[3] Priyadarshani N, Marsland S, Castro I. Automated birdsong recognition in complex acoustic environments: a review[J]. Journal of Avian Biology, 2018, 49(5): 1–27.

[4] 马克平. 生物多样性监测依赖于地面人工观测与先进技术手段的有机结合 [J]. 生物多样性, 2016, 24(11): 1201–1202.
Ma Keping. Biodiversity monitoring relies on the integration of human observation and automatic collection of data with advanced equipment and facilities[J]. Biodiversity Science, 2016, 24(11): 1201–1202.

[5] 杨洋, 汪毓铎. 基于改进卷积神经网络算法的语音识别 [J]. 应用声学, 2018, 37(6): 940–946.
Yang Yang, Wang Yuduo. Speech recognition based on improved convolutional neural network algorithm[J]. Journal of Applied Acoustics, 2018, 37(6): 940–946.

[6] 李云红, 梁思程, 贾凯莉, 等. 一种改进的DNN-HMM 的语音识别方法 [J]. 应用声学, 2019, 38(3): 371–377.
Li Yunhong, Liang Sicheng, Jia Kaili, et al. An improved speech recognition method based on DNN-HMM model[J]. Journal of Applied Acoustics, 2019, 38(3): 371–377.

[7] Chakraborty D, Mukker P, Rajan P, et al. Bird call identification using dynamic kernel based support vector ma-

chines and deep neural networks[C]. Proceedings of the IEEE International Conference on Machine Learning and Applications, IEEE, 2017: 280–285.

[8] Piczak K. Recognizing bird species in audio recordings using deep convolutional neural networks[C]. CEUR Workshop Proceedings, 2012, 43(9): 87–90.

[9] Martinsson J. Bird species identification using convolutional neural networks[D]. Gothenburg: Chalmers University of Technology, 2017.

[10] 谢将剑, 李文彬, 张军国, 等. 基于 Chirplet 语图特征和深度学习的鸟类物种识别方法 [J]. 北京林业大学学报, 2018, 40(3): 122–127.
Xie Jiangjian, Li Wenbin, Zhang Junguo, et al. Bird species recognition method based on Chirplet spectrogram feature and deep learning[J]. Journal of Beijing Forestry University, 2018, 40(3): 122–127.

[11] 随婷婷, 王晓峰. 一种基于 CLMF 的深度卷积神经网络模型 [J]. 自动化学报, 2016, 42(6): 875–882.
Sui Tingting, Wang Xiaofeng. Convolutional neural networks with candidate location and multi-feature fusion[J]. Acta Automatica Sinica, 2016, 42(6): 875–882.

[12] Goeau H, Glotin H, Vellinga W P, et al. LifeCLEF bird identification task 2014[R]. CLEF2014: 585–597.

[13] Wang J C, Wang J F, Weng Y S. Chip design of MFCC extraction for speech recognition[J]. Integration-The VLSI Journal, 2002, 32(1/2): 111–131.

[14] Glotin H, Ricard J, Balestriero R. Fast chirplet transform to enhance CNN machine listening-validation on animal calls and speech[J]. arXiv: 1611.08749, 2017.

[15] Potamitis I, Ntalampiras S, Jahn O, et al. Automatic bird sound detection in long real-field recordings: applications and tools[J]. Applied Acoustics, 2014, 80(4): 1–9.

[16] Marban A, Srinivasan V, Samek W, et al. Estimating position & velocity in 3D space from monocular video sequences using a deep neural network[C]. Proceedings of the IEEE International Conference on Computer Vision Workshop, IEEE, 2018: 1460–1469.

[17] 张建华, 孔繁涛, 吴建寨, 等. 基于改进 VGG 卷积神经网络的棉花病害识别模型 [J]. 中国农业大学学报, 2018, 23(11): 161–171.
Zhang Jianhua, Kong Fantao, Wu Jianzhai, et al. Cotton disease identification model based on improved VGG convolution neural network[J]. Journal of China Agricultural University, 2018, 23(11): 161–171.

[18] 刘文定, 李安琪, 张军国, 等. 基于 ROI-CNN 的赛罕乌拉国家级自然保护区陆生野生动物自动识别 [J]. 北京林业大学学报, 2018, 40(8): 123–131.
Liu Wending, Li Anqi, Zhang Junguo, et al. Automatic identification method for terrestrial wildlife in Saihanwula National Nature Reserve in Inner Mongolia of northern China based on ROI-CNN[J]. Journal of Beijing Forestry University, 2018, 40(8): 123–131.

[19] Yang T, Long X, Sangaiah A K, et al. Deep detection network for real-life traffic sign in vehicular networks[J]. Computer Networks, 2018, 136(8): 95–104.

[20] 付晓峰, 吴俊, 牛力. 小数据样本深度迁移网络自发表情分类 [J]. 中国图象图形学报, 2019, 24(5): 753–761.
Fu Xiaofeng, Wu Jun, Niu Li. Classification of small spontaneous expression database based on deep transfer learning network[J]. Journal of Image and Graphics, 2019, 24(5): 753–761.

[21] 胡满满, 陈旭, 孙毓忠, 等. 基于动态采样和迁移学习的疾病预测模型 [J]. 计算机学报, 2019, 42(10): 2339–2354.
Hu Manman, Chen Xu, Sun Yuzhong, et al. A disease prediction model based on dynamic sampling and transfer learning[J]. Chinese Journal of Computers, 2019, 42(10): 2339–2354.