



中南林业科技大学



2023

复杂环境下的鸟鸣声识别方法研究

汇报人：陈爱斌

人工智能应用研究所

Institute of Artificial Intelligence Application



01

研究背景及课题来源

02

数据集

03

鸟鸣声识别相关工作

04

工作展望

01
Part one

研究背景及课题来源

01 / 研究背景及课题来源



鸟鸣声中蕴含丰富的生物学信息在不同属或种间鸣声特征具有差异性，是鸟类物种识别的重要依据。



鸟类栖息地多为山林或海岛、湿地环境，环境噪声复杂，并受鸟类活动性限制，野外采集数据通常信噪比较低。

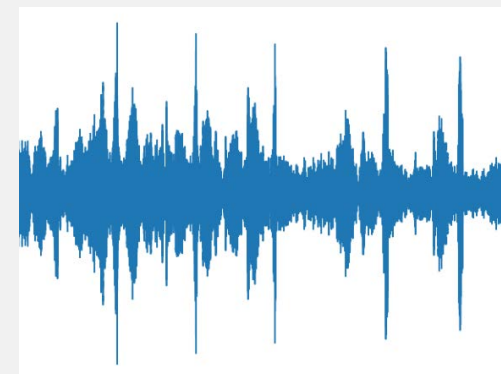


设计一种鲁棒性高、抗噪声能力强的鸟鸣声识别方法对了解鸟类生物多样性具有重要意义。



国家自然科学基金面上项目（62276276，2023-2026）

《融合深度学习和心理声学的南方林区鸟鸣声感知模型研究》



02
Part two

数据集

[xeno-canto](https://www.xeno-canto.org)网站 (https://www.xeno-canto.org)

世界野生鸟类声音网： 是世界上最具有权威性的鸟类声音数据库，目前已经收录了10268个鸟类物种，共609894个鸟类鸣叫的录音文件。

录制环境多样：全球不同的环境下录制，如：森林、草原和湿地等。

- 鸣声种类
 - 鸣叫：短促和简单
 - 鸣唱：时间较长且复杂
- 噪音复杂多样
 - 低频噪音，高频噪音
 - 其他鸟类声音的干扰
 - 声音重叠的干扰

xeno-canto
Sharing wildlife sounds from around the world

Search recordings... Search Advanced Search Tips

About Explore Upload Sounds Forum Mysteries Articles Log in / Register

Brazilian Tepuis

XC303636 0:00 0:46
褐胸蚊鹀 (*Myrmothera simplex*) · call
GABRIEL LEITE

After an **expedition** reached the Serra da Mocidade, a remote mountain range in Roraima, Brazil a few years ago, XC recordist **Gabriel Leite** uploaded a **set of recordings from the area**. It showed many significant records for Tepui species. Just recently another expedition ventured further north and visited the environs of the spectacular Serra do Sol or Wii-tipí and Gabriel uploaded a **set from that Tepui** as well.

Serra do Sol or Wii-tipí, Roraima, Brazil. © Thiago Orsi Laranjeiras.

What is xeno-canto?
xeno-canto is a website dedicated to sharing wildlife sounds from all over the world. Whether you are a research scientist, a birder, or simply curious about a sound that you heard out your kitchen window, we invite you to listen, download, and explore the wildlife sound recordings in the collection.
But xeno-canto is more than just a collection of recordings. It is also a collaborative project. We invite you to share your own recordings, help identify mystery recordings, or share your expertise in the Forums. Welcome!

All groups

Collection Statistics

| | |
|-------------|-----------------|
| 1353348 | IDs |
| 733849 | Recordings |
| 10515 | Species |
| 12215 | Subspecies |
| 9069 | Recordists |
| 11969:55:34 | Recording Time |
| 641815 | Records to GBIF |

More...

Latest New Species

- Ctenodecticus ramburi
- Chorthippus apicalis
- Pseudolus' Saddle Bush-cricket
- Jumping Bush Cricket
- Velarifictorus grylloides

More...

Latest Additions

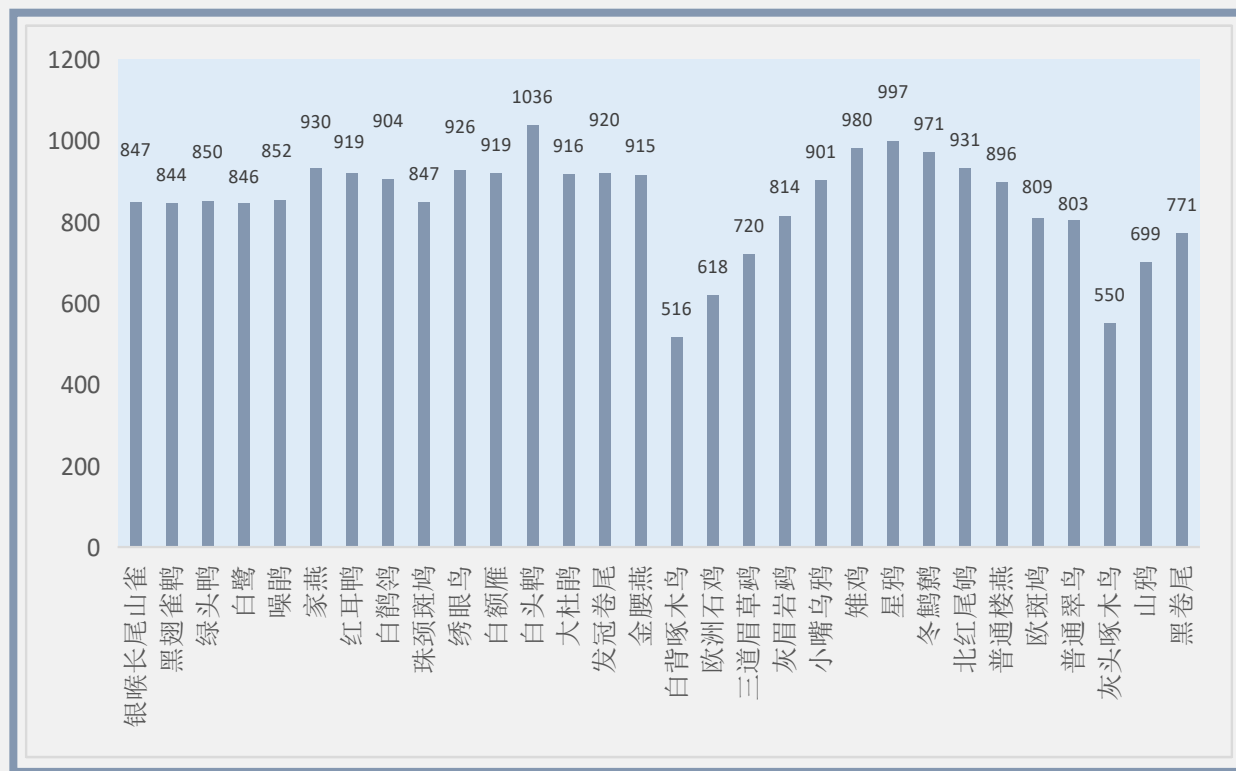
02 /数据集



采样频率为16000HZ，比特率为320000bps



所有的音频片段均由专业人员人工标注。一共30个类别，25454个样本



03

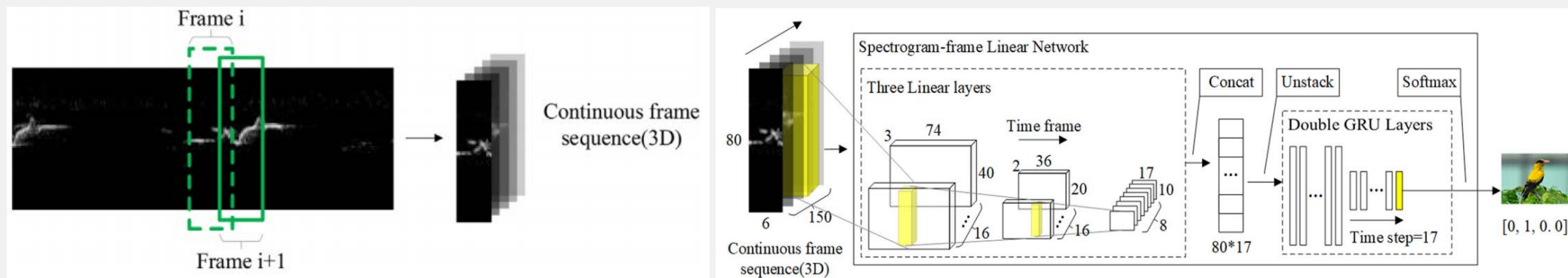
Part three

鸟鸣声识别相关工作

鸟鸣声具有多种频率分布和连续时变特性，CNN无法完全识别鸟类的时频特征。

(1) 提出一种鸟鸣声连续帧序列化方法，用较短的静态帧来作为鸟鸣声的识别单位，多帧组成连续的帧序列，来作为鸟鸣声的动态视觉表示。

(2) 提出一种鸟鸣声的时频帧线性网络，以连续帧序列为输入，强调鸟鸣声的频域分布和时域变化。

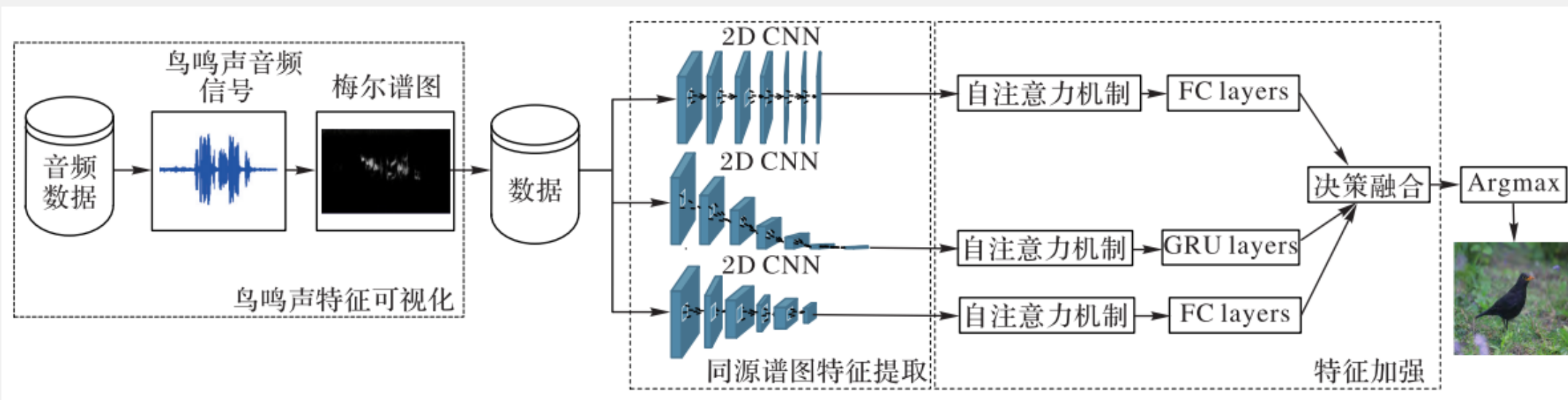


- Zhang X, Chen A, Zhou G. Spectrogram-frame linear network and continuous frame sequence for bird sound classification. **Ecological Informatics** [J]. 2019.

03 / 鸟鸣声识别相关工作

复杂背景下的鸟鸣声具有时域连续性、频域高低性特点。

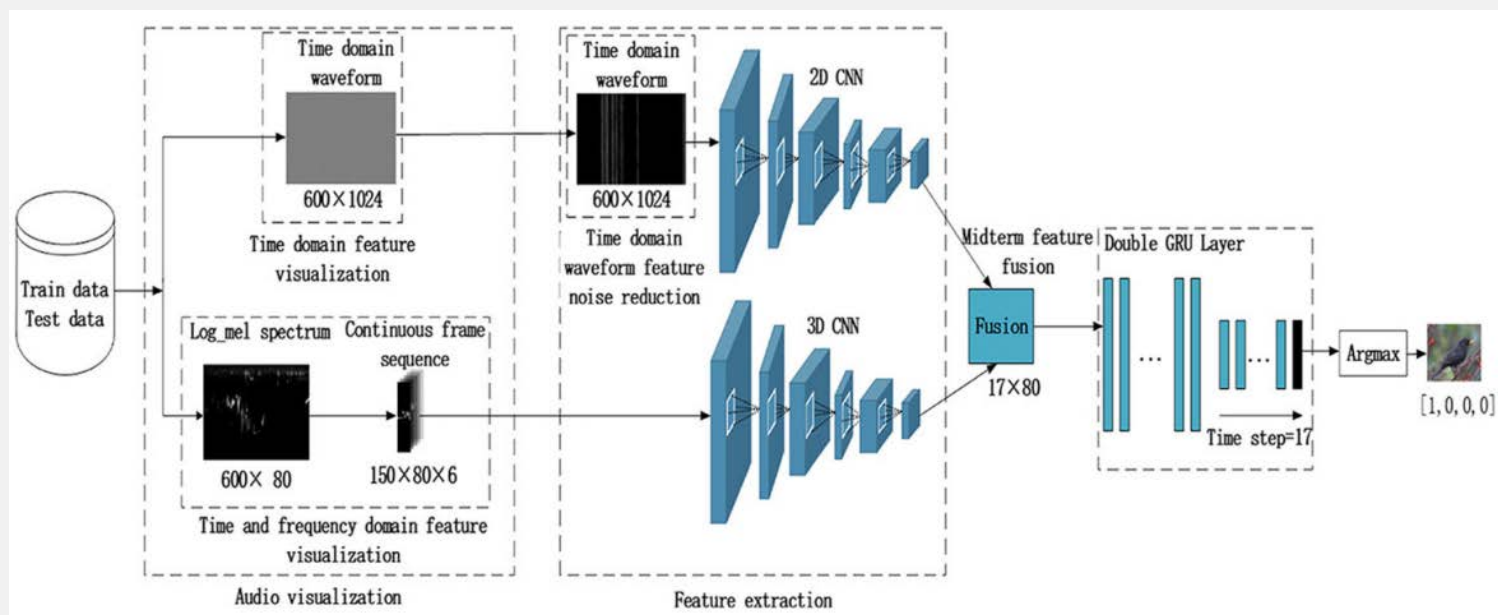
我们将梅尔频谱图采用特定的卷积以及下采样操作在时域和频域上压缩至一维，得到具有连续性以及高低特性的时频域特征。



■ 刘志华,陈文洁,陈爱斌.基于自注意力机制时频谱同源特征融合的鸟鸣声分类[J].计算机应用, 2022, 42(4):1260-1268.DOI:10.11772/j.issn.1001-9081.2021071258.

自然环境采集的鸟鸣声具有时间连续性的本质特征。

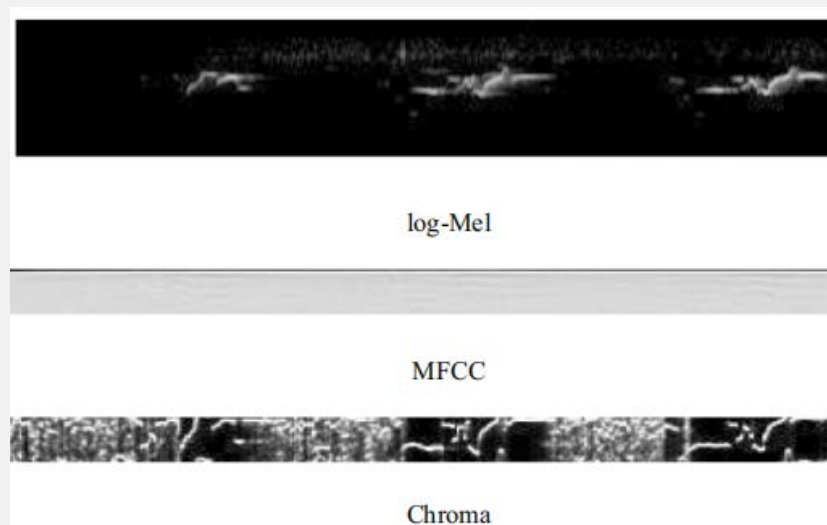
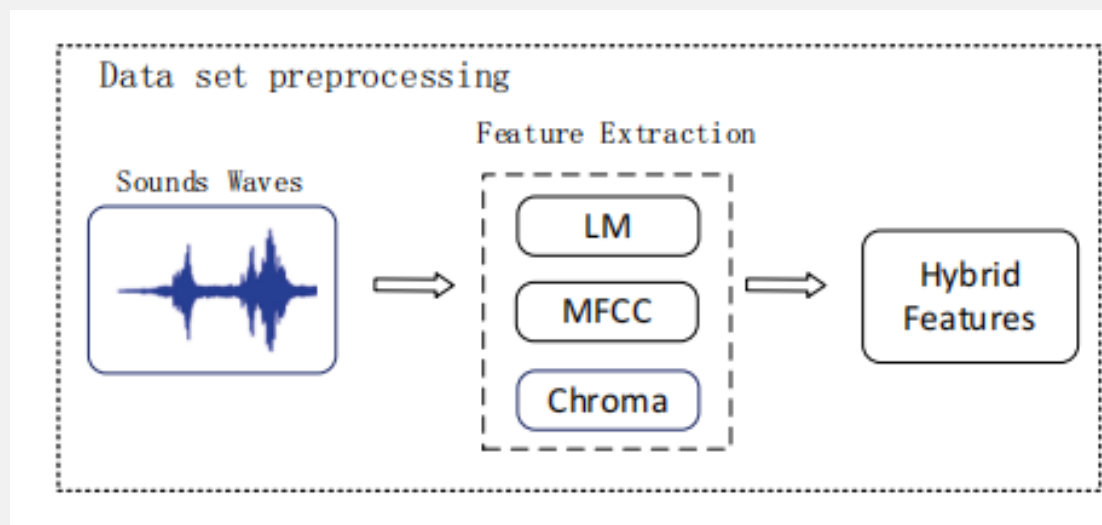
我们分别采用三维卷积神经网络（3DCNN）和二维卷积神经网络（2DCNN）作为log_mel光谱和波形图像的特征提取器。然后将从这两个特征通道中提取的高级特征进行中间阶段融合，并将输出增强特征作为双门控递归单元（d-GRU）网络的输入。



- Liu Z, Chen W, Chen A, et al. Birdsong classification based on multi feature channel fusion[J]. Multimedia Tools and Applications, 2022, 81(11): 15469-15490.

鸟鸣声数据集具有复杂多样音频背景。

我们引入了用于声音和音乐分析的声学特征：色度，并且与常用的鸟声特征：Log-Mel谱图（LM）和Mel频率倒谱（MFCC）拼接融合，丰富单一特征的代表能力。

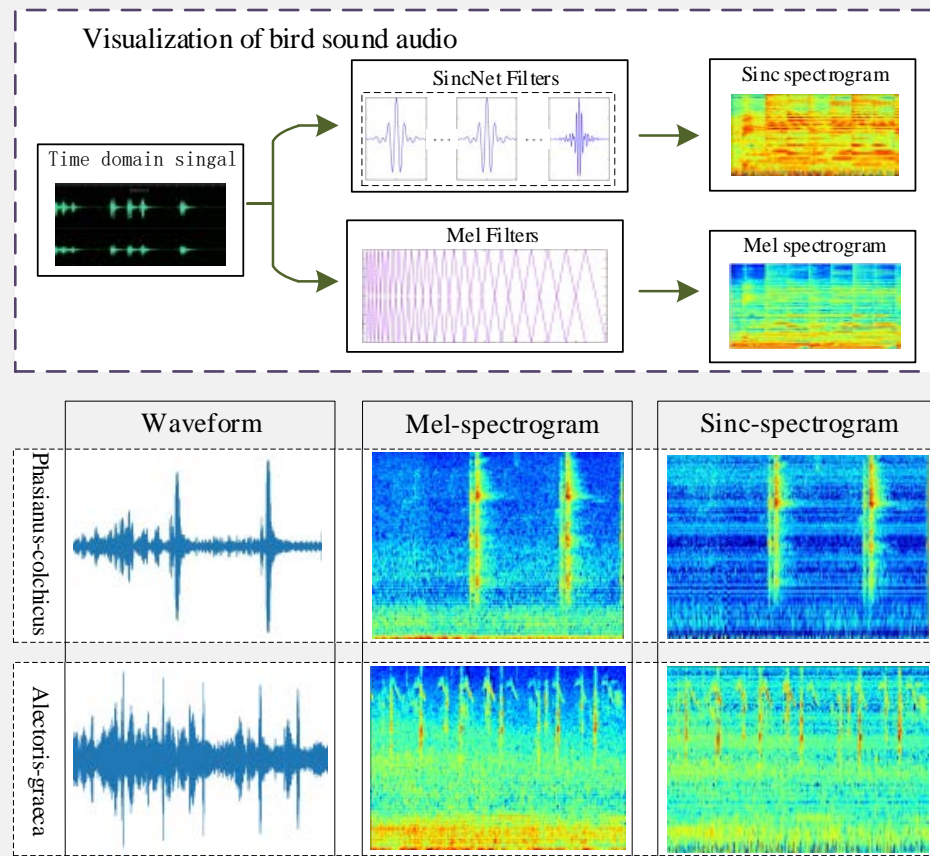


- Yan N , Chen A , Zhou G ,et al.Birdsong classification based on multi-feature fusion[J]. Multimedia Tools and Applications. 2021.DOI:10.1007/s11042-021-11396-9.

03 / 鸟鸣声识别相关工作

鸟鸣声样本通过特征提取生成频谱图进行分类显示出了良好的性能。然而频谱图生成过程中滤波器提取信息会造成信息丢失，限制了鸟鸣声识别的学习能力。

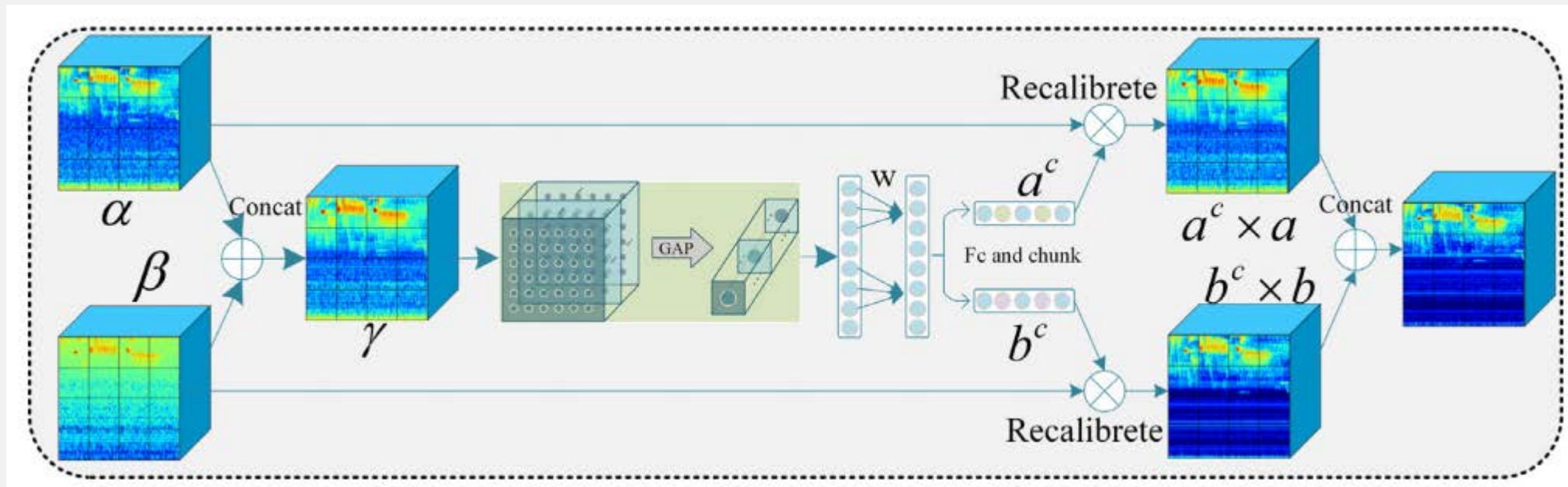
我们对鸟鸣声音频通过Mel滤波器提取具有低频特征优势的Mel-频谱图和Sincnet滤波器提取具有音色特征优势的Sinc频谱图，并对其进行决策前融合，得到鸟鸣声Mel-Sinc频谱图。



- Hu S, Chu Y, Wen Z, et al. Deep learning bird song recognition based on MFF-ScSEnet[J]. Ecological Indicators, 2023, 154: 110844.

由于自然环境的复杂性，基于深度学习的单一的滤波器特征提取音频信息存在信息丢失，高效且快速识别鸟鸣声仍然具有挑战性。

我们基于Mel和Sincnet的特征提取滤波器提取鸟鸣声的低频和音色信息，利用提出的双特征融合模块（FFMS）融合两个特征集间差异性的低频、音色信息。

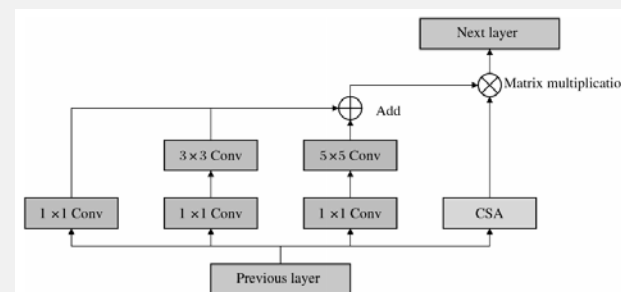
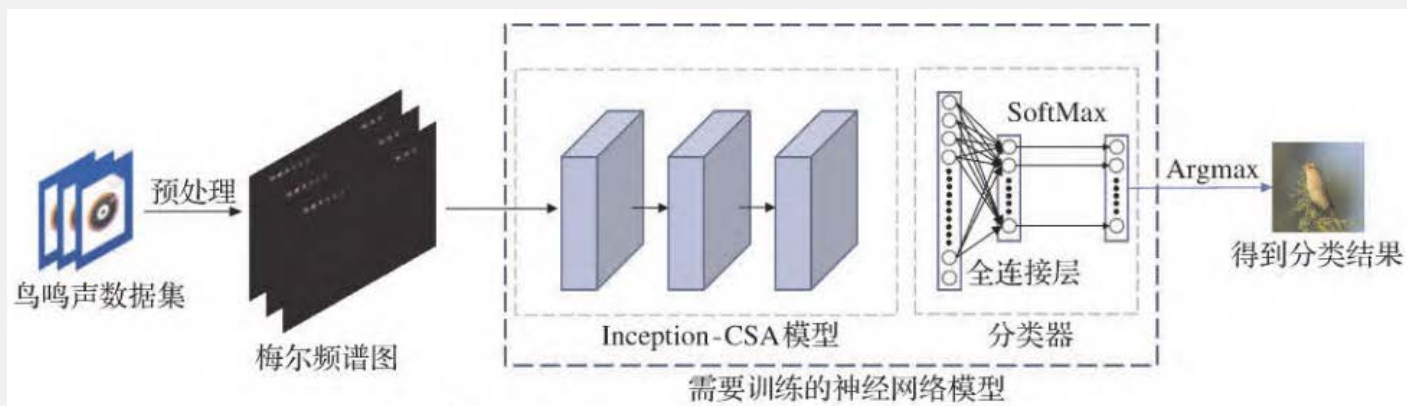


- Hu S, Chu Y, Tang L, et al. A lightweight multi-sensory field-based dual-feature fusion residual network for bird song recognition[J]. Applied Soft Computing, 2023, 146: 110678.

03 / 鸟鸣声识别相关工作

由于音频特征图同时具备时域和频域特征，且卷积神经网络不能捕获远距离特征。

我们利用Inception-CSA模型对鸟鸣声特征图进行特征提取，其中Inception模块提取鸟鸣声特征图中的多尺度局部时频域特征，CSA模块获取鸟鸣声特征图的全局注意力权重，将二者的输出结合得到更强的特征图。

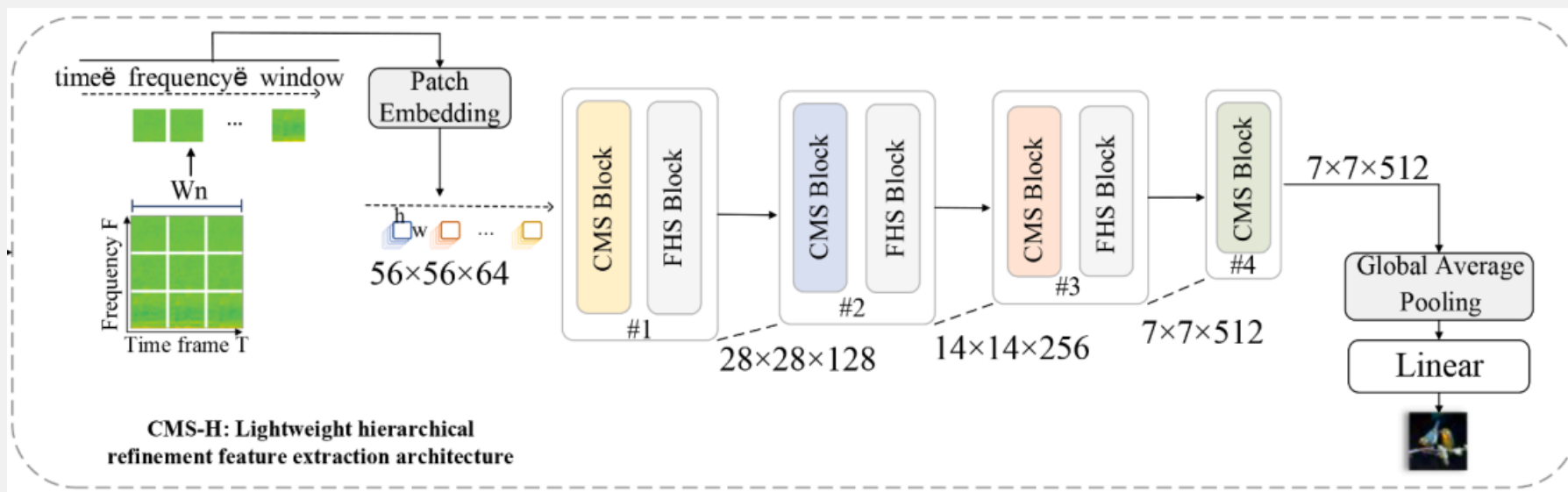


Inception-CSA Block的结构图

- 李怀城, 杨道武, 温治芳, 等. 基于 Inception-CSA 深度学习模型的鸟鸣分类[J]. 华中农业大学学报 (自然科学版), 2023, 42(3): 97-104.

监测区域记录的音频一般具有复杂的背景噪声，鸣声的特征不突出，生物光谱信息不全面。

我们提出了一种结合动态和静态建模的分层鸟叫特征提取架构，以充分挖掘上下文关系，并且解决了Transformer模型中忽略内部结构信息、丢失空间信息的问题。



- Wang Y, Chen A, Li H, et al. A hierarchical birdsong feature extraction architecture combining static and dynamic modeling[J]. Ecological Indicators, 2023, 150: 110258.

04
Part four

工作展望

鸟鸣声识别

基于图像的鸟类识别

鸟类简单情感识别

鸟语识别

宠物声音识别

宠物图像识别

动物个体识别

动物行为识别（金丝猴、黑猩猩）

数据来源？

研究成果和收获

1. Shipeng Hu, Yihang Chu, Lu Tang, Guoxiong Zhou, Aibin Chen, Yurong Sun. A lightweight multi-sensory field-based dual-feature fusion residual network for **bird song recognition**. Applied Soft Computing. 2023
2. Shipeng Hu, Yihang Chu, Zhifang Wen, Guoxiong Zhou, Yurong Sun, Aibin Chen. Deep learning **bird song recognition** based on MFF-ScSEnet. Ecological Indicators. 2023
3. Huaicheng Li, Jizheng Yi, Wenjie Chen, Daowu Yang, Guoxiong Zhou, Weixiong Peng. Environmental **sound classification** based on CAR-Transformer neural network model. Circuits, Systems, and Signal Processing. 2023
4. Yanan Wang, Huaicheng Li, Guoxiong Zhou, Jizheng Yi, Zhiqiang Zhang. A hierarchical **birdsong feature extraction** architecture combining static and dynamic modeling. Ecological Indicators.2023
5. 李怀城, 杨道武, 温治芳, 王亚楠. 基于Inception-CSA深度学习模型的**鸟鸣分类**. 华中农业大学学报. 2023
6. 刘志华, 陈文洁, 陈爱斌. 基于自注意力机制时频谱同源特征融合的**鸟鸣声分类**. 计算机应用. 2022,42(4):1260-1268
7. 杨道武, 陈文洁, 陈爱斌. 基于复值谱图的重参数化结构**声源分离**条件网络. 郑州大学学报(理学版). 2022.54(2):61-66
8. Zihua Liu, Wenjie Chen, Aibin Chen, Guoxiong Zhou, Jizheng Yi. **Birdsong classification** based on multi feature channel fusion. Multimedia Tools and Applications. 2022.02.
9. Yan, N., Chen, A., Zhou, G. et al. **Birdsong classification** based on multi-feature fusion. Multimedia Tools and Applications. 2021.09.
10. 彭宁, 陈爱斌, 周国雄等. 基于正弦注意力表征网络的环境**声音识别**. 应用科学学报. 2021.39(4):641-649

研究成果和收获

11. Ning Peng, Aibin Chen, Guoxiong Zhou. et al. Environment **Sound Classification** Based on Visual Multi-Feature Fusion and GRU-AWS. IEEE ACCESS, 2020(8): 191100-191114
12. Xin Zhang, Aibin Chen, Guoxiong Zhou, Zhiqiang Zhang, Xibei Huang, Xiaohu Qiang. Spectrogram-frame Linear Networks and Continuous Frame Sequence for **Bird Sound Classification**. Ecological Informatics. 2019.
13. 软件著作权：《基于深度学习的鸟类声音分类系统V1.0》（登记号码：2020SR1135506）
14. 软件著作权：《基于深度学习的音频源分离微信小程序》（登记号码：2021SR0500957）
15. 国家发明专利：《基于MFF-ScSNet的狗吠声情绪识别方法》（实质审查阶段）
16. 国家发明专利：《应对低信噪比的鸟鸣声识别分类方法、装置及存储介质》（实质审查阶段）
17. 国家发明专利：《基于表征全局依赖关系的并行注意力机制鸟鸣声识别方法》（实质审查阶段）
18. 国家自然科学基金面上项目：融合深度学习和心理声学的南方林区鸟鸣声感知模型研究（62276276），2023.01-2026.12，53万元



2023

欢迎批评指正！
