基于预训练模型的鸟鸣识别方法研究

谢捷

南京师范大学

读书明志，文明弘扬；手捧国学，百遍不倦；珠玑格言，与君共勉；
诚信惟善，孝义当先；好书共享，兴味盎然；修身律己，无私奉献。

目录

目录

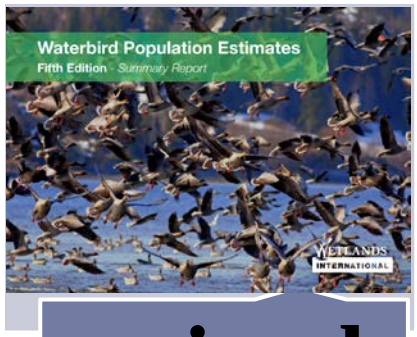
生物多样性损失及监测重要性

Habitat loss

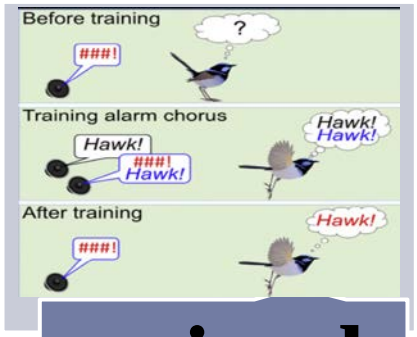
Invasive species

Pollution

Climate change



animal populations



animal behavior

environment

保护 → 监测生物多样性 → 动物监测

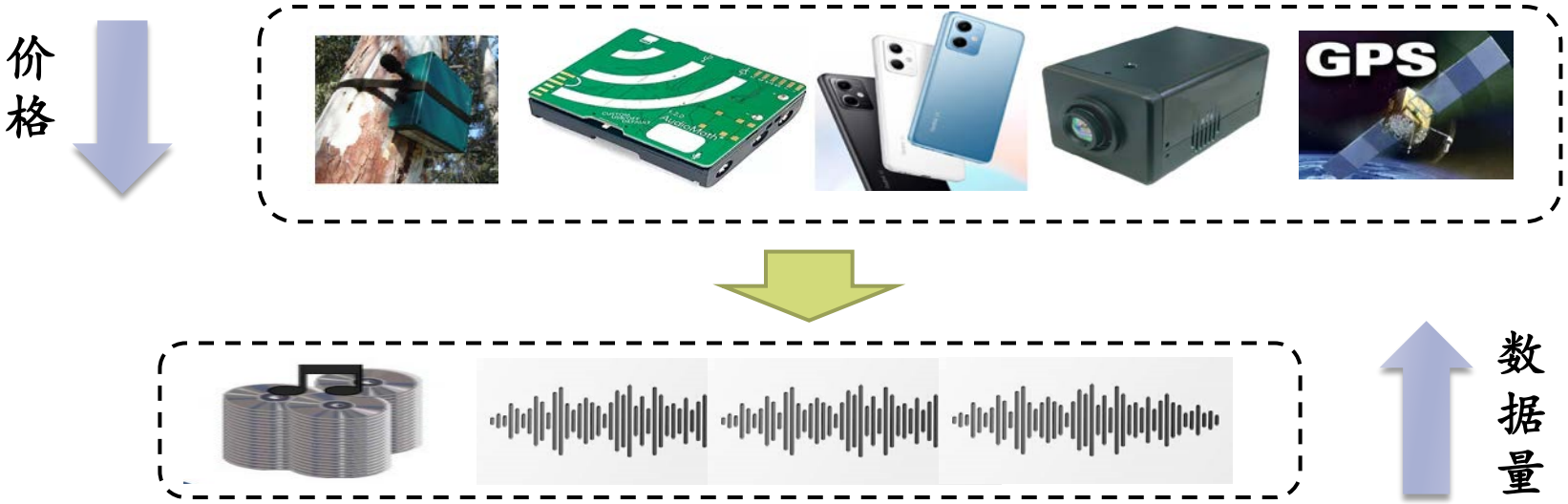
- 人工监测

- 主动监测

Servick, K. (2014). Eavesdropping on ecosystems.

- **被动监测**
  - 红外相机
  - **声传感器**
  - 无人机
  - 其他

被动监测



- 大数据
  - 原始数据多
  - 但，带标签的数据少

大数据

- 自动/半自动数据分析方法

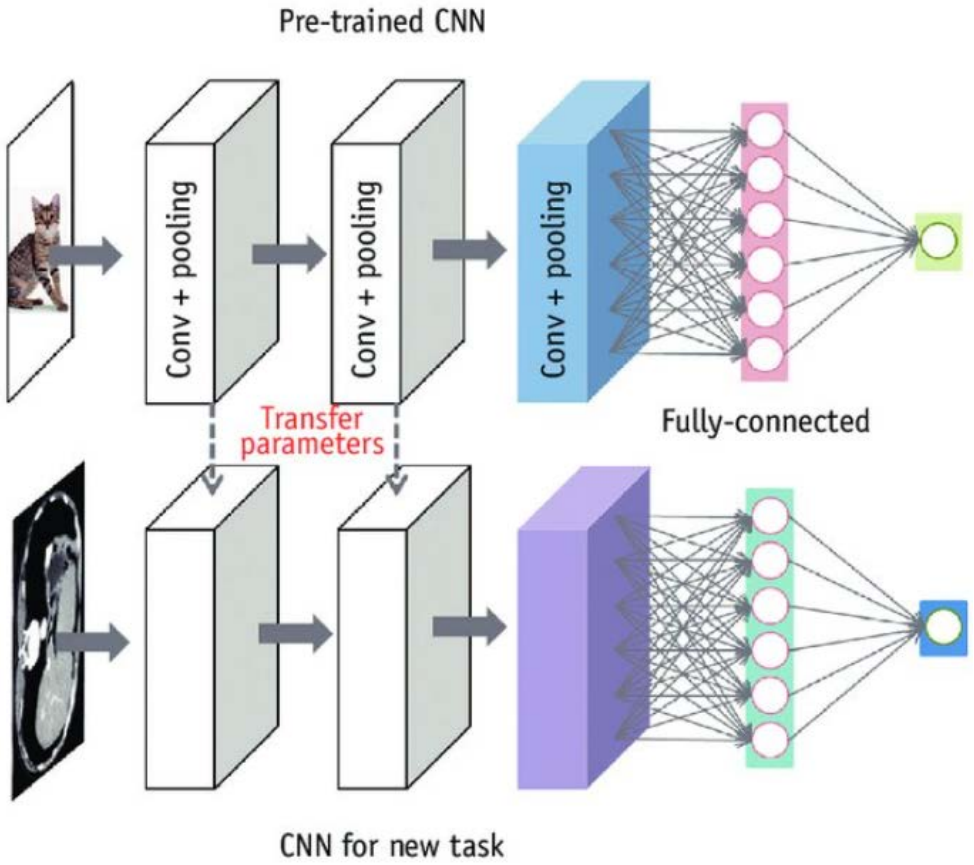
- 深度学习
  - 与采集的原始数据比较，**带标签的数据非常少**
  - 但，在深度学习中，通常需要**大量的训练数据**以便模型能够学习到足够的特征，从而提高泛化性能。

目录

- 研究背景
- **预训练模型**
- 鸟鸣识别
- 总结和展望

预训练模型

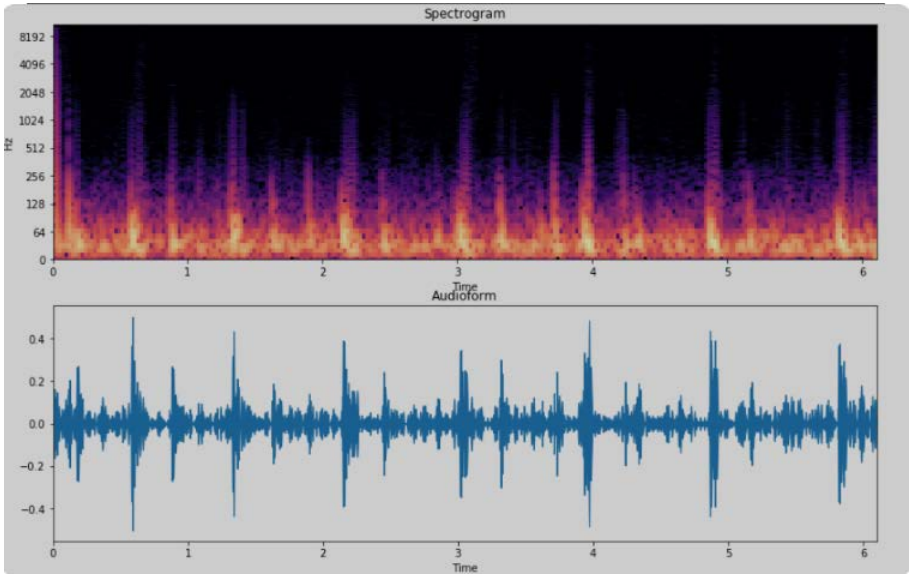
- 预训练模型



Do, S., Song, K. D., & Chung, J. W. (2020). Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning. Korean journal of radiology, 21(1), 33-41.

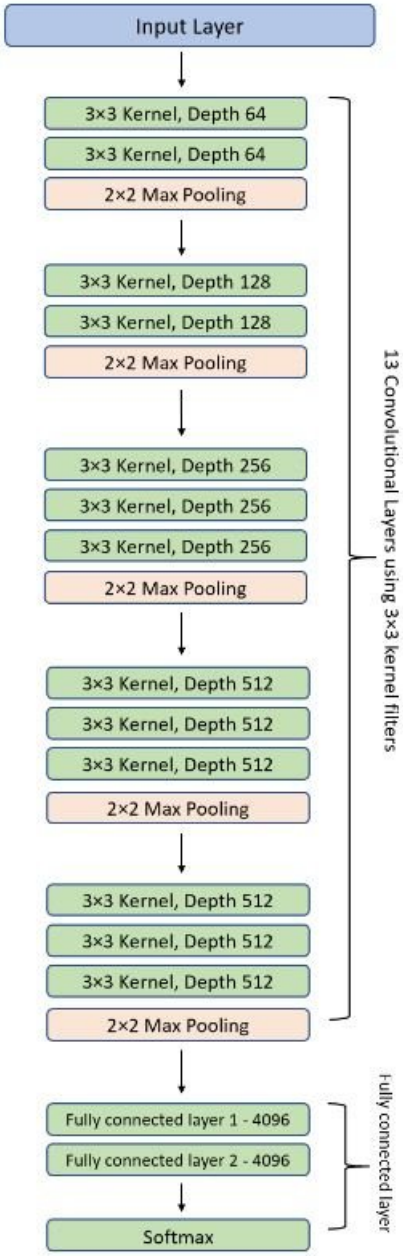
预训练模型

- ImageNet
  - VGG16, VGG19, ResNet50, DenseNet, ...
- AudioNet
  - VGGish, YAMNet, PANNs
- BirdNet
- Transformer-based



模型的训练

- Train from scratch（从头训练）
- Finetune (微调)
- Feature extraction（特征提取）



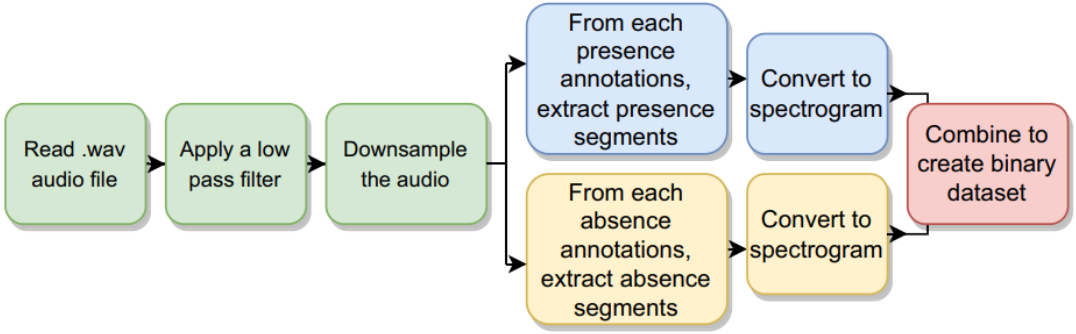
目录

鸟鸣识别1

Ecological Informatics
Volume 70, September 2022, 101688

Passive acoustic monitoring of animal populations with transfer learning

Emmanuel Dufourq [a b c], Carly Batist [d], Ruben Foquet [e], Ian Durbach [f g]



The 12 CNNs compared in this study. The number of trainable network parameters are shown for the case where the feature extractor was fine-tuned (*with FT*) and where the feature extractor was frozen (*without FT*).

Architecture	Study	Parameters with FT	Parameters without FT
DenseNet121	Huang et al. (2017)	6,986,626	32,770
DenseNet169	Huang et al. (2017)	12,537,730	53,250
DenseNet201	Huang et al. (2017)	18,154,370	61,442
InceptionResNetV2	Szegedy et al. (2017)	54,294,626	18,434
InceptionV3	Szegedy et al. (2016)	21,792,930	24,578
MobileNetV2	Howard et al. (2017)	2,275,074	51,202
ResNet101	He et al. (2016a)	42,634,754	81,922
ResNet101V2	He et al. (2016a)	42,610,818	81,922
ResNet152V2	He et al. (2016a)	58,269,826	81,922
ResNet50V2	He et al. (2016a)	23,601,282	81,922
VGG16	Simonyan and Zisserman (2014)	14,731,074	16,386
Xception	Chollet (2017)	20,888,874	81,922

Dufourq, Emmanuel, et al. "Passive acoustic monitoring of animal populations with transfer learning." Ecological Informatics 70 (2022): 101688.

鸟鸣识别1

Table 4
Comparison of the average F1 score across the different network architectures and dataset configurations. The *exponent* approach was used for the spectrogram input. The feature extracted was frozen. The results are averaged across 13 unique executions. The results are ordered (highest to lowed) based on the average of each network architecture across all configurations. The best three performing network architectures on a particular dataset configuration is highlighted in bold.

Method	G 25	G 50	G 100	L 25	L 50	L 100	A 25	A 50	A 100
ResNet101V2	**95.30**	**97.40**	**96.27**	**92.05**	**94.92**	**97.01**	**92.10**	**95.37**	**98.36**
ResNet152V2	95.18	96.92	**96.58**	**91.42**	**95.31**	**96.62**	**93.32**	**94.94**	**98.35**
InceptionResNetV2	94.70	96.75	**96.57**	90.07	**95.35**	95.73	**92.73**	93.95	97.84
ResNet50V2	94.97	97.04	95.13	**91.96**	93.66	**96.36**	90.94	**94.82**	98.12
DenseNet169	94.92	96.95	95.69	89.33	93.78	95.59	91.76	93.32	97.95
DenseNet201	94.84	96.72	95.86	90.08	93.90	95.59	91.02	93.12	98.13
VGG16	**97.26**	**98.09**	94.99	87.74	92.93	95.01	90.26	92.32	**98.74**
DenseNet121	94.58	96.69	95.00	89.90	92.82	95.81	90.26	93.99	98.06
InceptionV3	92.22	95.42	95.40	88.72	93.23	95.45	91.21	93.28	96.93
ResNet101	**96.17**	**97.49**	94.23	90.01	92.21	91.42	91.00	91.02	97.80
Xception	93.88	95.79	95.50	88.15	93.81	94.10	90.74	91.12	97.51
MobileNetV2	94.62	96.65	91.65	88.78	90.85	94.91	83.50	93.40	97.71

- ✓ 重点关注与卷积神经网络 (CNN) 相关的数据稀缺问题。
- ✓ 该方法不太复杂，并且可以很容易地采用。
- ✓ 贡献了四个被动声学数据集，对应 90 小时。
- ✓ 仅采用25个样本，可以获得高达 82% 的 F1 分数。

鸟鸣识别2

Applied Acoustics 166 (2020) 107375

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling

Ming Zhong [a,*], Jack LeBien [b], Marconi Campos-Cerqueira [b], Rahul Dodhia [a], Juan Lavista Ferres [a], Julian P. Velev [c], T. Mitchell Aide [b,d]

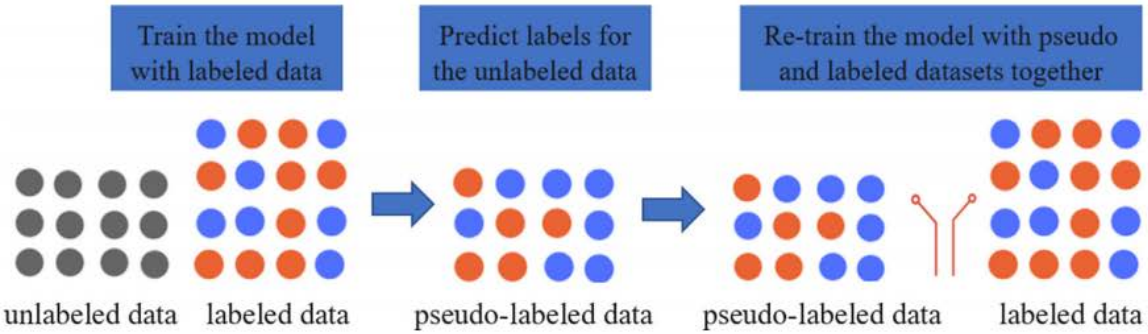


Fig. 2. The diagram of the proposed methodology with pseudo-label generating and model training. The dots with different colors represent observations with different labels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Classification results (sensitivity, specificity and AUC) by each CNN model.

Model ID	Model Description	Sensitivity (%)	Specificity (%)	AUC (%)
1	CNN with VGG16 Architecture	82.1	96.9	97.5
2	Pre-trained ResNet50	84.1	97.7	97.9
3	Pre-trained ResNet50 + Custom Loss Function + Pseudo Labeling	**97.7**	**96.4**	**99.5**

鸟鸣识别3

birds

MDPI

Article

Acoustic Classification of Bird Species Using an Early Fusion of Deep Features

Jie Xie [1,*] and Mingying Zhu [2,3]

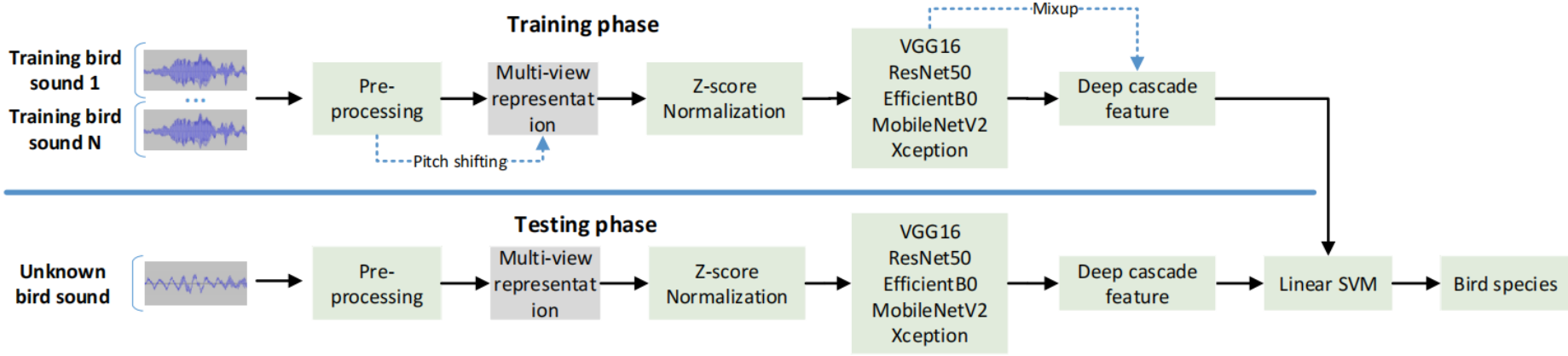


Table 1. Layers to concatenate deep features for five different transfer learning models. For all transfer learning models, four individual layers are selected for extracting features except Xception, where three layers are used for feature extraction.

TL Model	Selected Layers for Generating Concatenated Deep Features			
VGG16	block2_conv2	block3_conv3	block4_conv3	block5_conv3
ResNet50	conv2_block3_out	conv3_block4_out	conv4_block6_out	conv5_block3_out
MobileNetV2	block_13_project_BN	block_14_add	block_15_add	global_average_pooling2d
EfficientNetB0	block4c_add	block5c_add	block6d_add	avg_pool
Xception	block4_sepconv1	block5_sepconv1	block14_sepconv1	—

✓ 基于不同训练层的特征融合

鸟鸣识别3

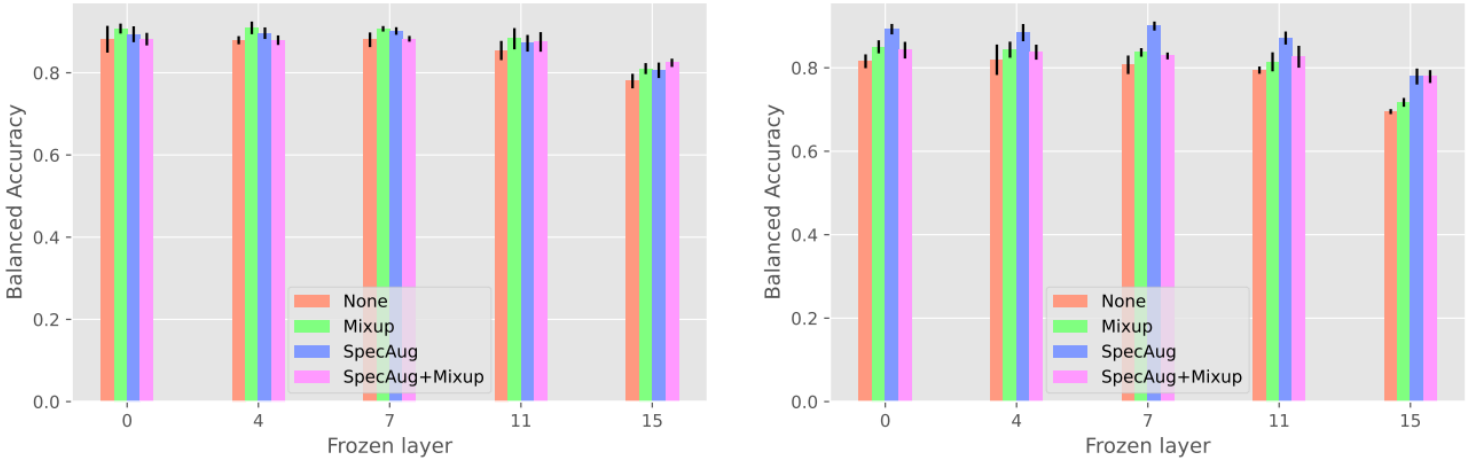


Figure 2. (**a**) Classification performance using deep cascade features with linear SVM, (**b**) End-to-end classification using softmax. For deep cascade features using linear support vector machines (SVM), mixup is slightly better than other methods except for having 15 frozen layers. For end-to-end classification, spectral augmentation (SpecAug) achieves the best performance.

- ✓ 基于深度特征采集Mixup可获得最高的准确率
- ✓ 基于端到端的鸟鸣识别，使用SpecAug可获得最高的准确率

鸟鸣识别3

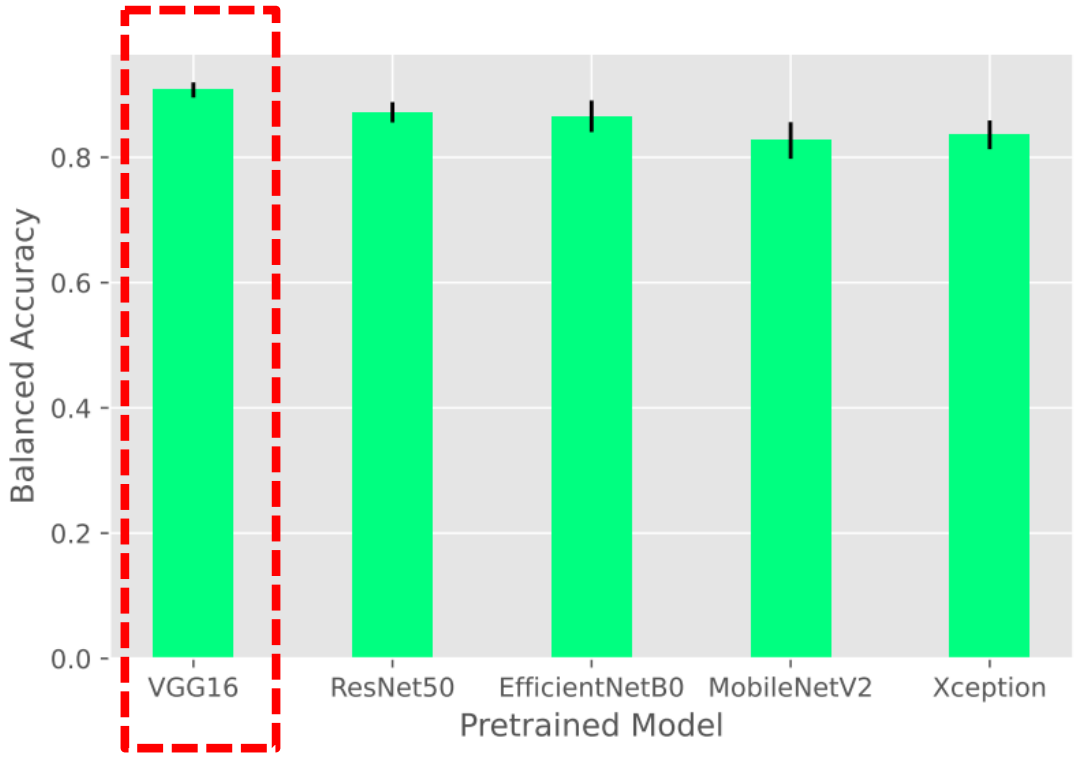
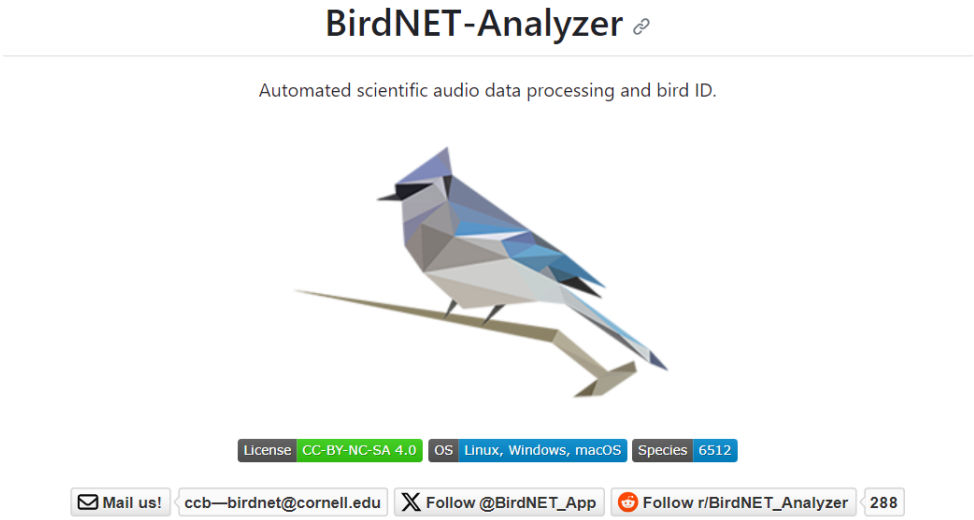


Figure 4. Comparison of different pre-trained models using the same input: repeat-based spectrogram. Here, VGG16 performs the best in terms of balanced accuracy.

✓ 针对CLO43数据集，VGG16获得最高的性能

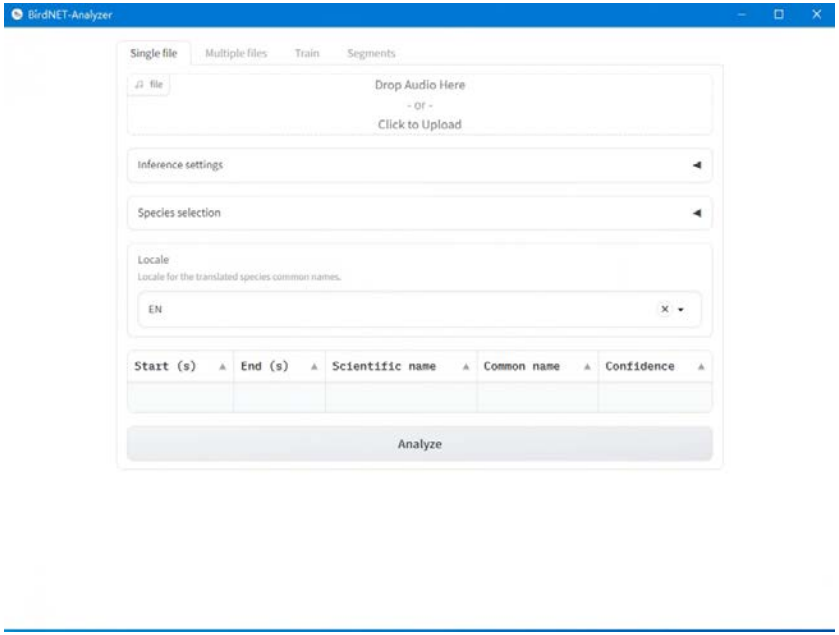
鸟鸣识别4

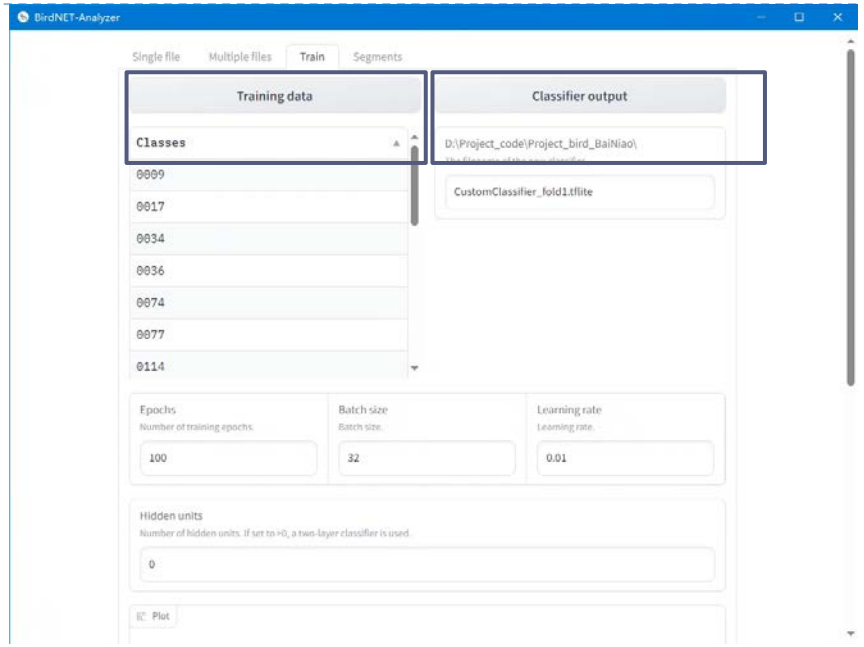
- BirdNet



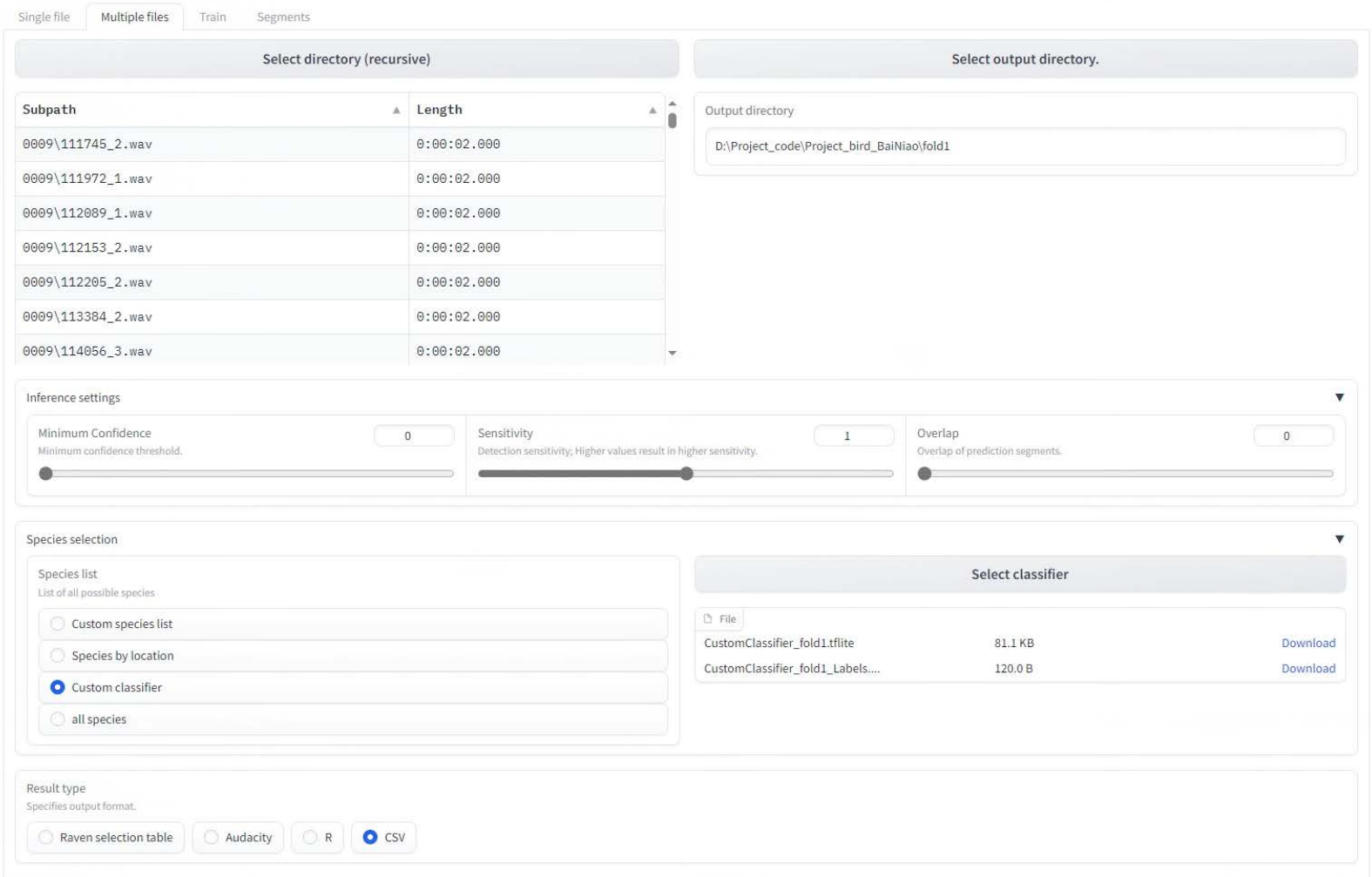
Our Net?

鸟鸣识别4


BirdNET-Analyzer
Single file
Multiple files
Train
Segments
file
Drop Audio Here
- or -
Click to Upload
Inference settings
Species selection
Locale
Locale for the translated species common names.
EN
Start (s)
End (s)
Scientific name
Common name
Confidence
Analyze


BirdNET-Analyzer
Single file
Multiple files
Train
Segments
Training data
Classifier output
Classes
0009
0017
0034
0036
0074
0077
0114
D:\Project_code\Project_bird_BaiNiao\
CustomClassifier_fold1.tflite
Epochs
Number of training epochs.
100
Batch size
Batch size.
32
Learning rate
Learning rate.
0.01
Hidden units
Number of hidden units. If set to >0, a two-layer classifier is used.
0
Plot

鸟鸣识别4


Single file
Multiple files
Train
Segments
Select directory (recursive)
Select output directory.
Subpath
Length
0009\111745_2.wav
0:00:02.000
0009\111972_1.wav
0:00:02.000
0009\112089_1.wav
0:00:02.000
0009\112153_2.wav
0:00:02.000
0009\112205_2.wav
0:00:02.000
0009\113384_2.wav
0:00:02.000
0009\114056_3.wav
0:00:02.000
Output directory
D:\Project_code\Project_bird_BaiNiao\fold1
Inference settings
Minimum Confidence
Minimum confidence threshold.
0
Sensitivity
Detection sensitivity; Higher values result in higher sensitivity.
1
Overlap
Overlap of prediction segments.
0
Species selection
Species list
List of all possible species
Custom species list
Species by location
Custom classifier
all species
Select classifier
File
CustomClassifier_fold1.tflite
81.1 KB
Download
CustomClassifier_fold1_Labels....
120.0 B
Download
Result type
Specifies output format.
Raven selection table
Audacity
R
CSV

目录

- 研究背景
- 预训练模型
- 鸟鸣识别
- **总结和展望**

总结和展望

- 针对特定的鸟鸣识别任务，需要选择合适的预训练模型
- 基于Transformer的预训练模型用于鸟鸣识别
- 构建针对中国鸟类的大模型

谢谢大家！敬请批评指正！

Jie Xie
Nanjing Normal University
Email: xiej8734@gmail.com
Phone: 15906279084