Listen to the Sound of Rare Species!

Data Augmentation for Acoustic Biodiversity Monitoring

AC297r: IACS Capstone Project
Fall 2022

Yang Xiang, Li Sun, Meichen Dong
Yanqi Luo, Ziye Tao

RAINFOREST CONNECTION

IACS

Motivation — Wide Application of Text-to-Image Generation

Tropical ecosystems are particularly characterized by the high number of **rare and inconspicuous species**. Insufficient data samples from them are the main barrier for utilizing advanced data mining techniques to monitor biodiversity.

Therefore, **improving data augmentation methods** in this field is fundamental in order to improve model performance on detecting future rare species and better **protecting biodiversity**.

Research Contributions

Exploration:
Exploratory analysis was carried out on the audio data to better understand the frequency feature and variation of all the rare species across different time and locations.

Audio & Image Augmentation:
Four audio and three image data augmentation methods were compared and evaluated on the model performance.

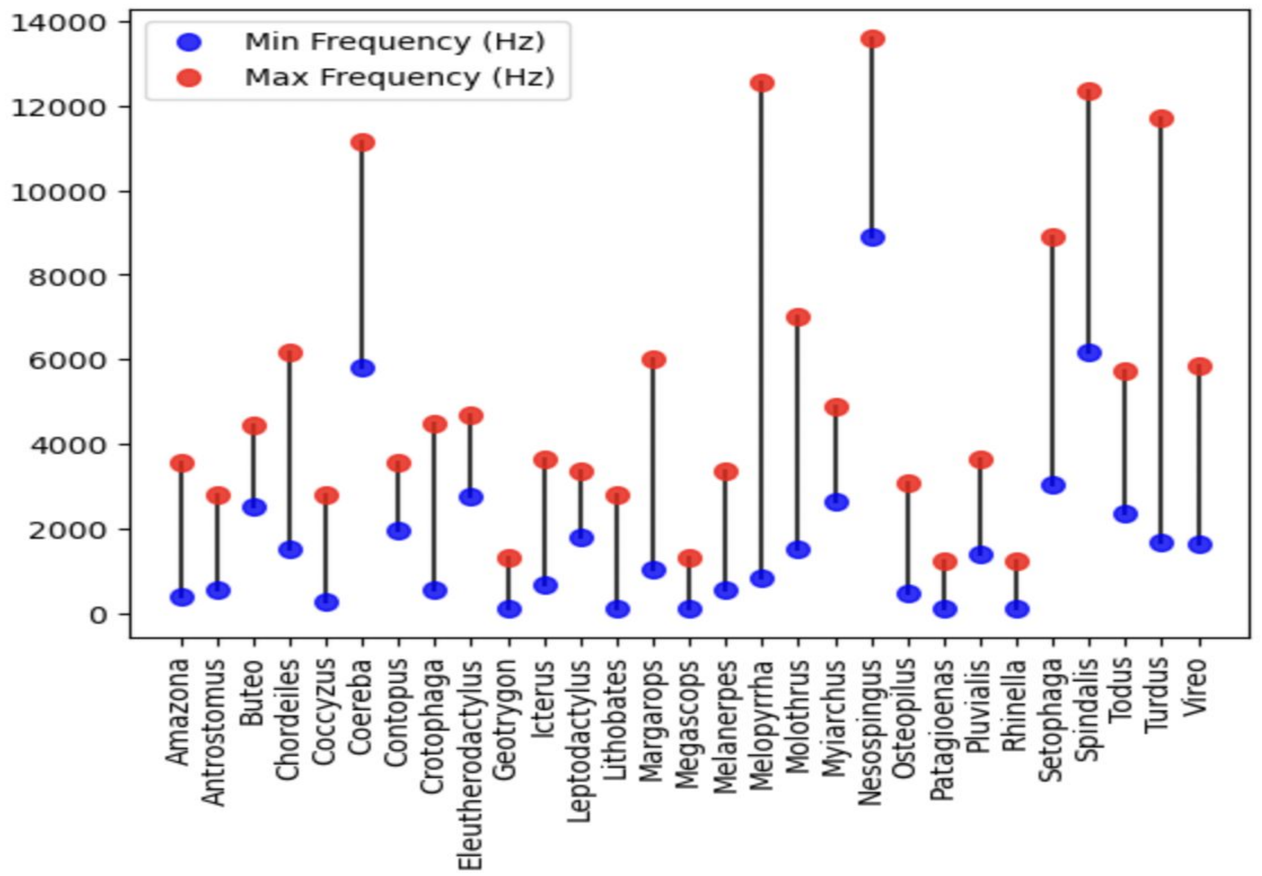
Data Balancing:
Combination of augmentation methods are done and same-level model performance is achieved with much larger size of data which contributes to future detection of species.

Pipeline Construction
Build user-friendly Code pipeline for future usage

Data & Exploratory Data Analysis

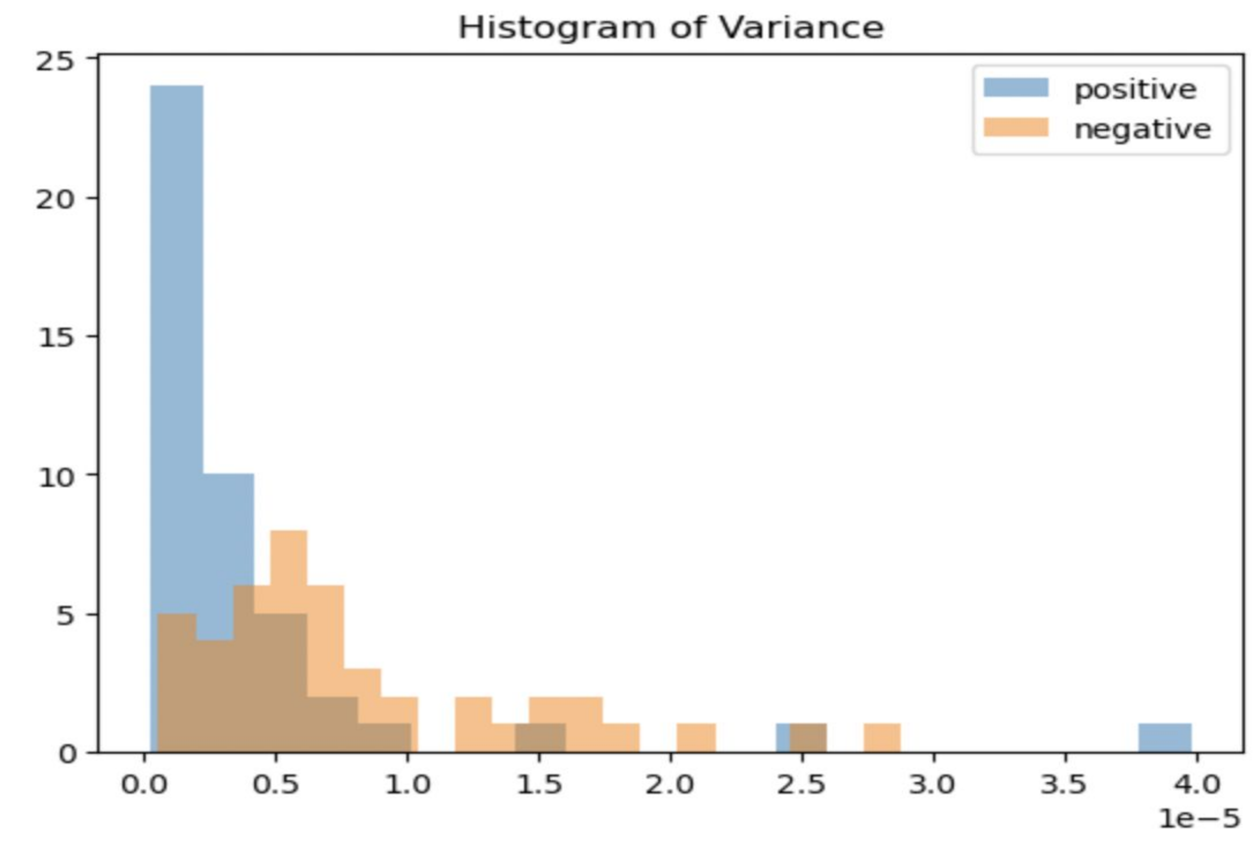
Train: 200 positive & 500 negative samples for 45 species labelled by experts
Test: 2000 1-min-long audio recording

Frequency Range for Each Species

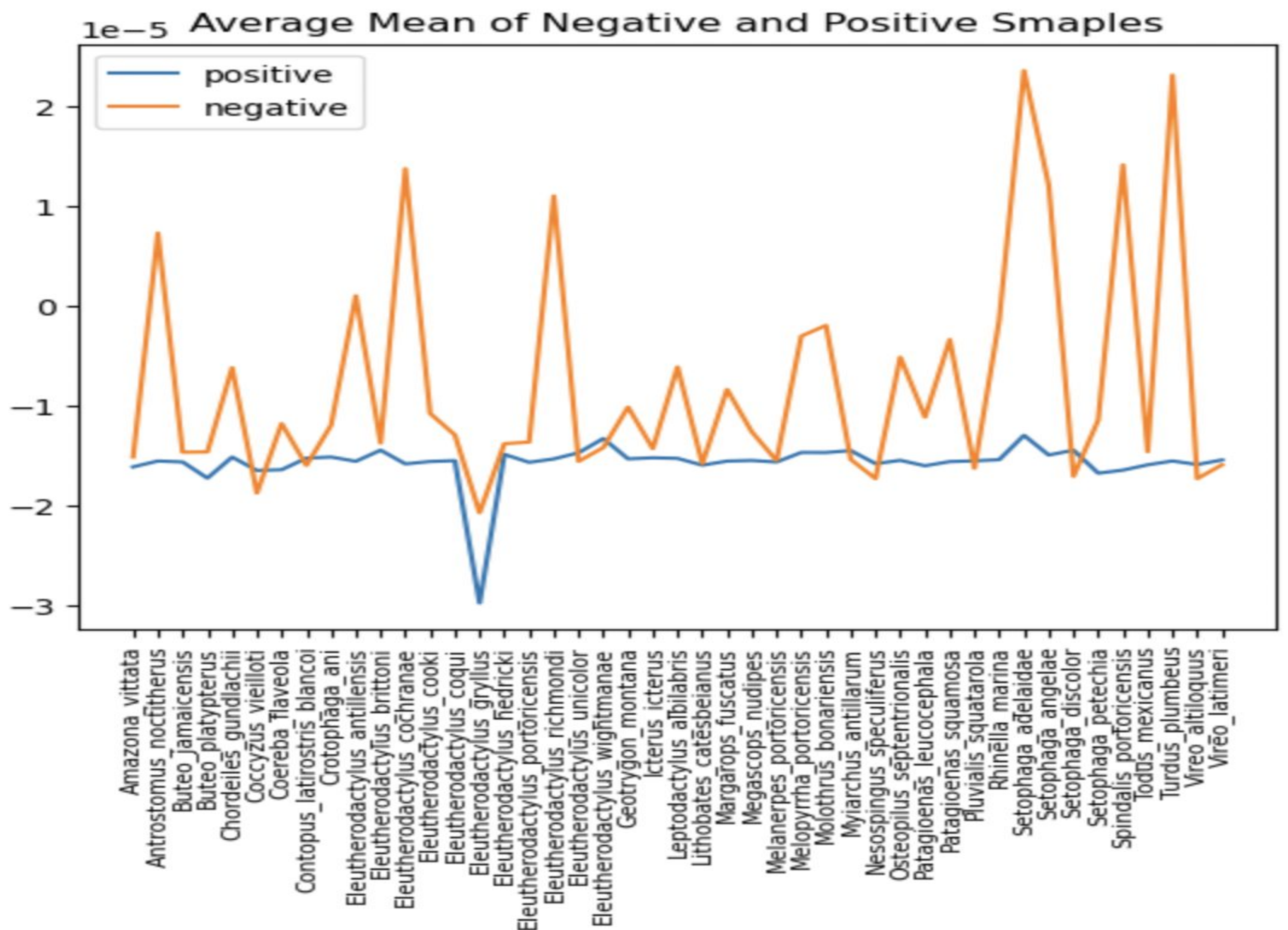


Data Sample Location Distribution

Histogram of Variance of Each Class



Histogram of Mean Values of Each Class

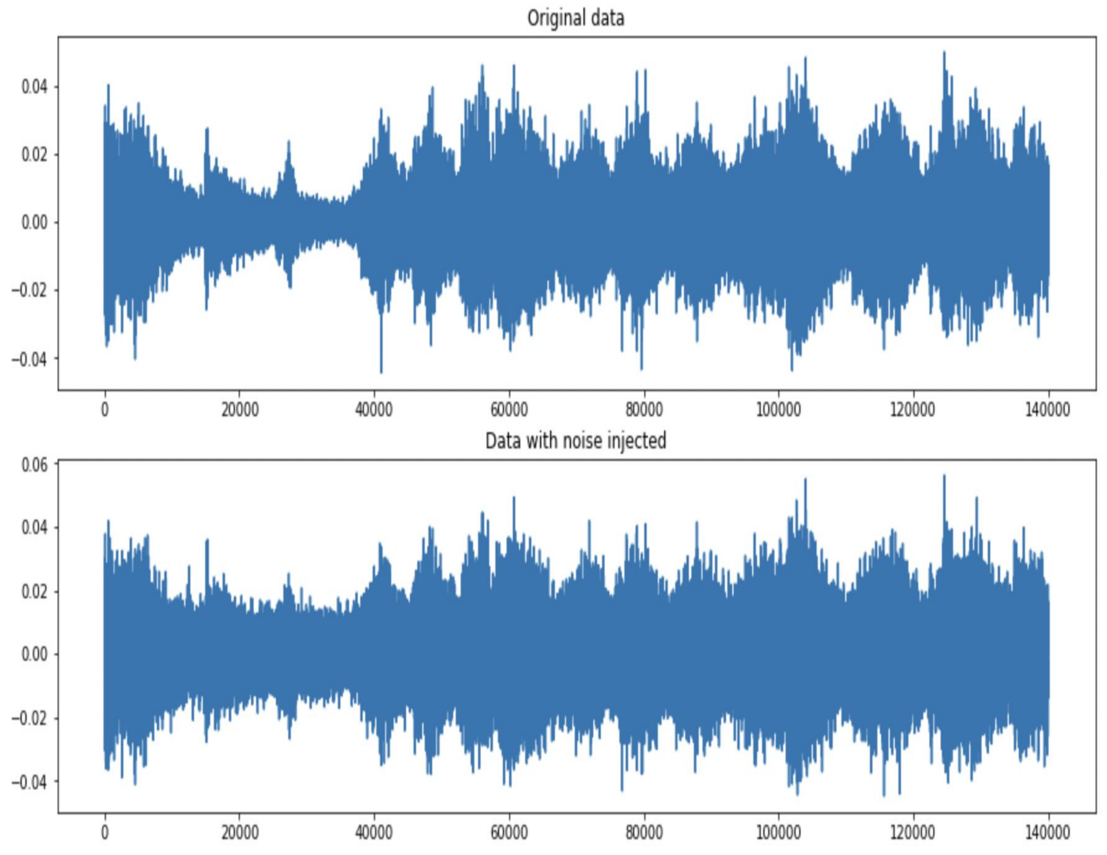


Audio Data Augmentation

Noise Injection

`aug_data = data + noise_factor * noise`

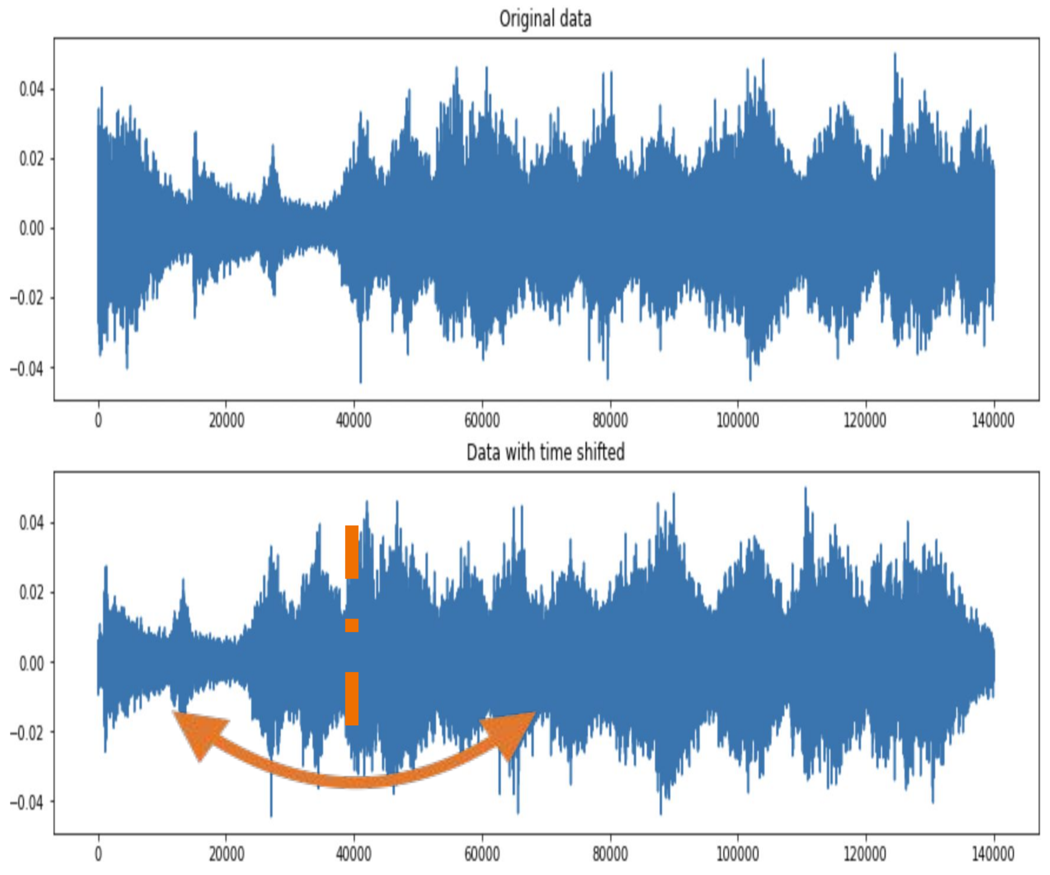
Inject random noise to audio data.



Time Shifting

`np.roll(data, shift)`

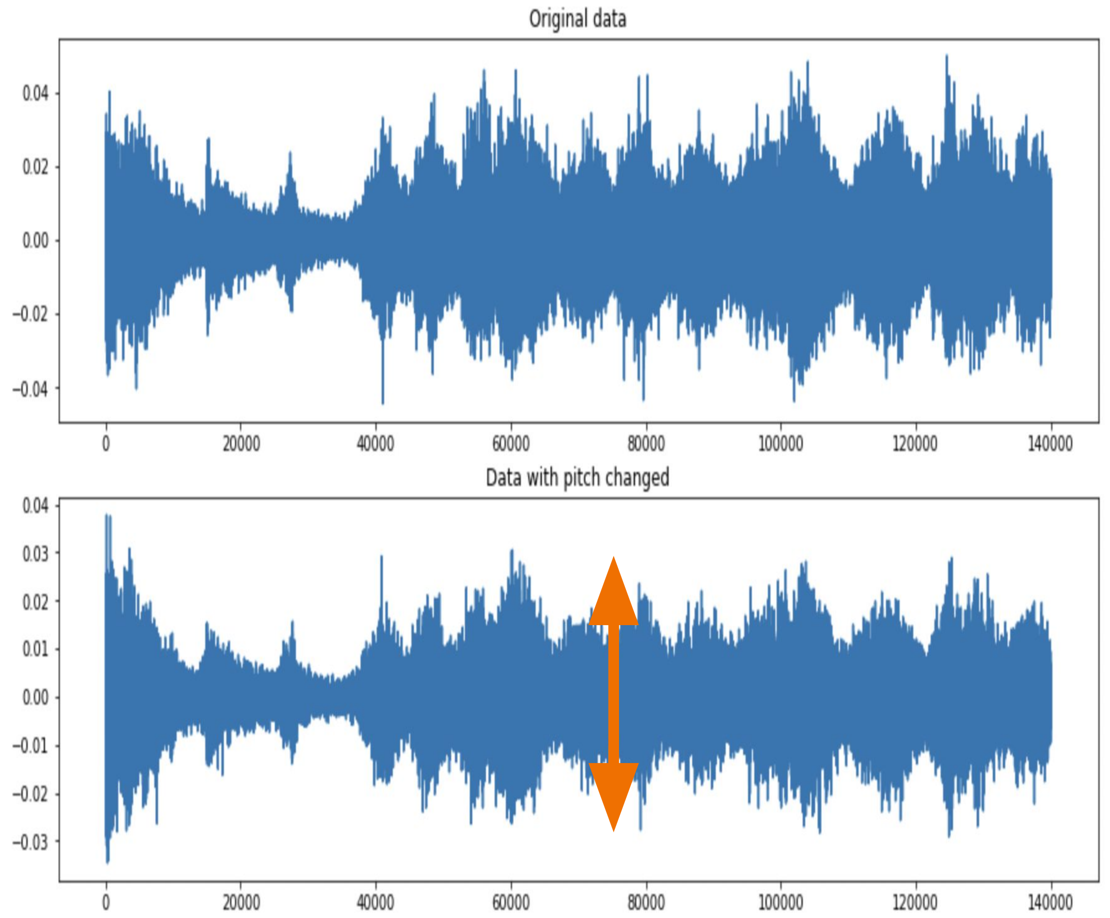
Randomly shift time to left/right to avoid centralization of feature sounds



Pitch Changing

`librosa.effects.pitch_shift(y=data, sr=sample_rate, n\steps=n_step)`

Adjust pitch within reasonable range



Speed Changing

`librosa.effects.time_stretch(data,speed_factor)`

Adjust speed within reasonable range

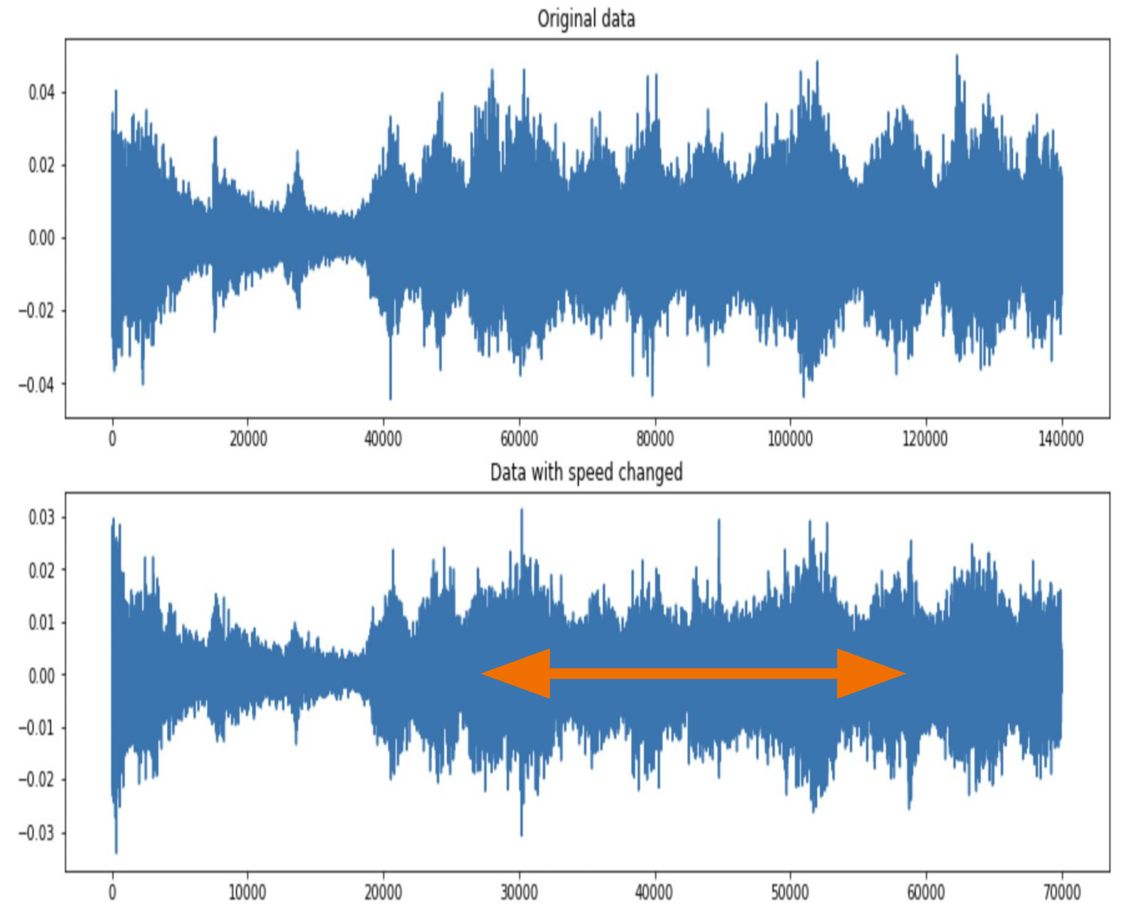
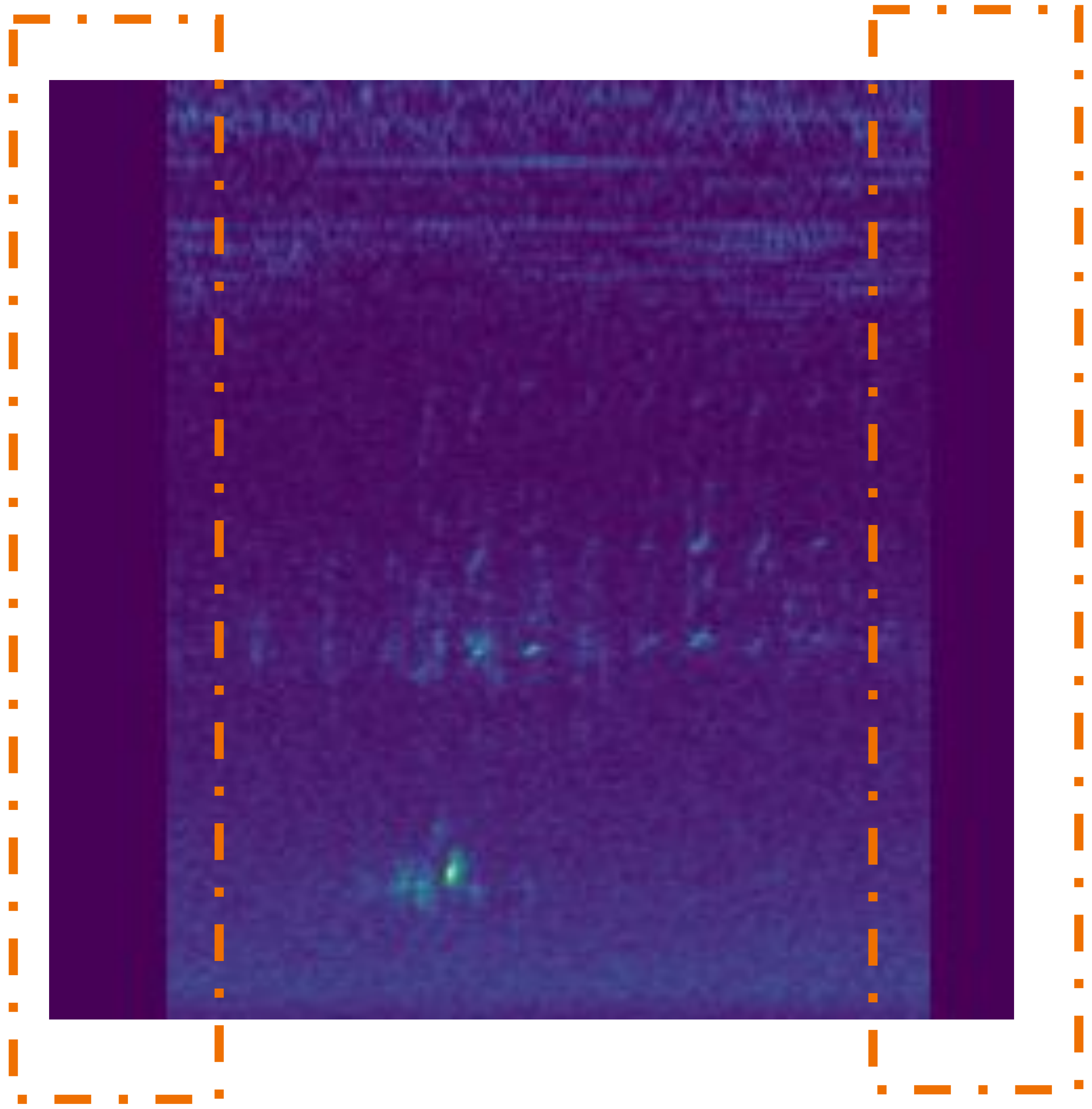


Image Data Augmentation

Loudness Augmenter

nas.LoudnessAug(zone=(zoneL, zoneR),coverage, factor=(0.75, 1.25))

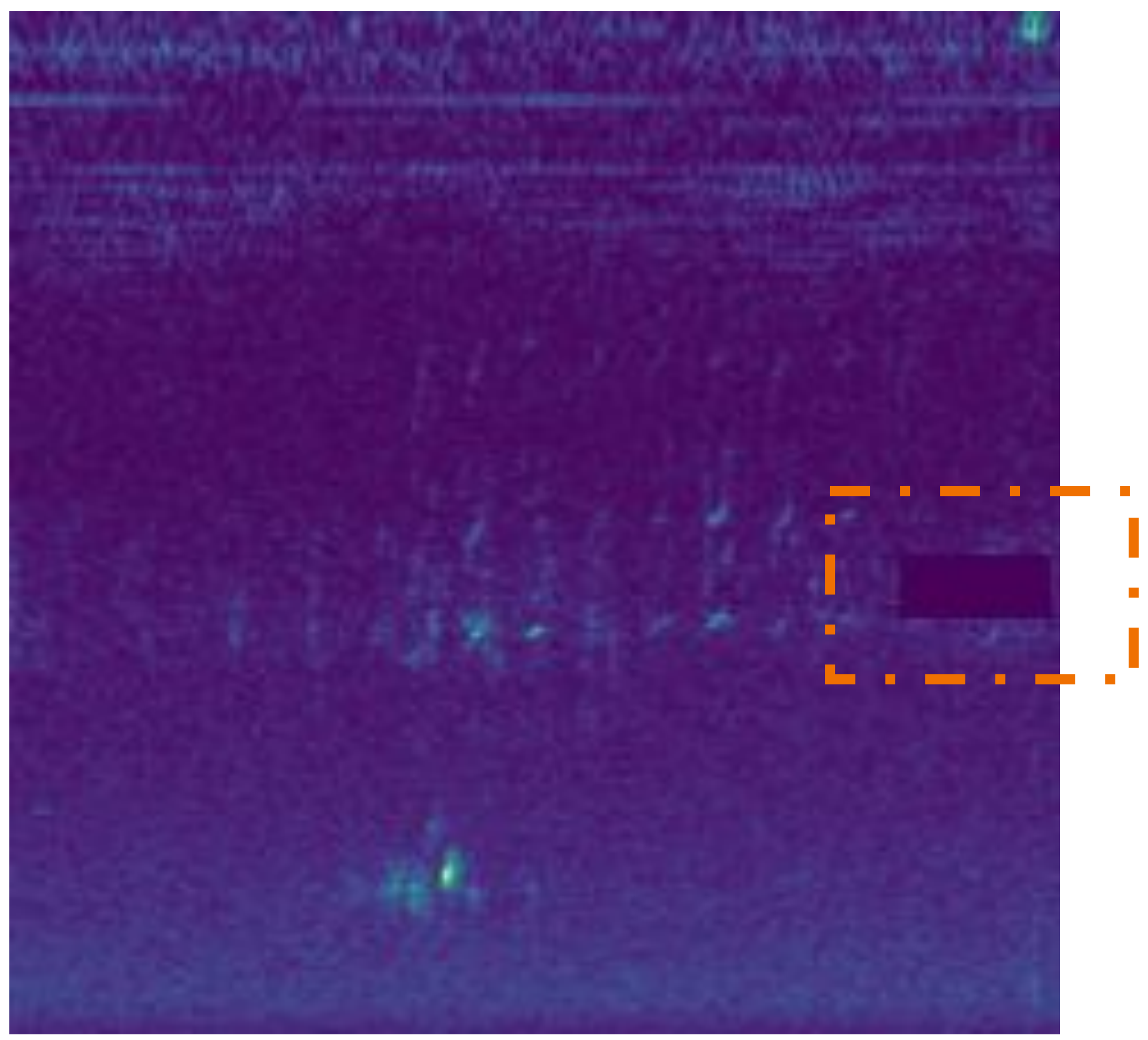
A random spectrogram zone is assigned to apply loudness adjusting operation.



Frequency Masking

nas.FrequencyMaskingAug(zone=(zoneL, zoneR), coverage, factor=(10,20))

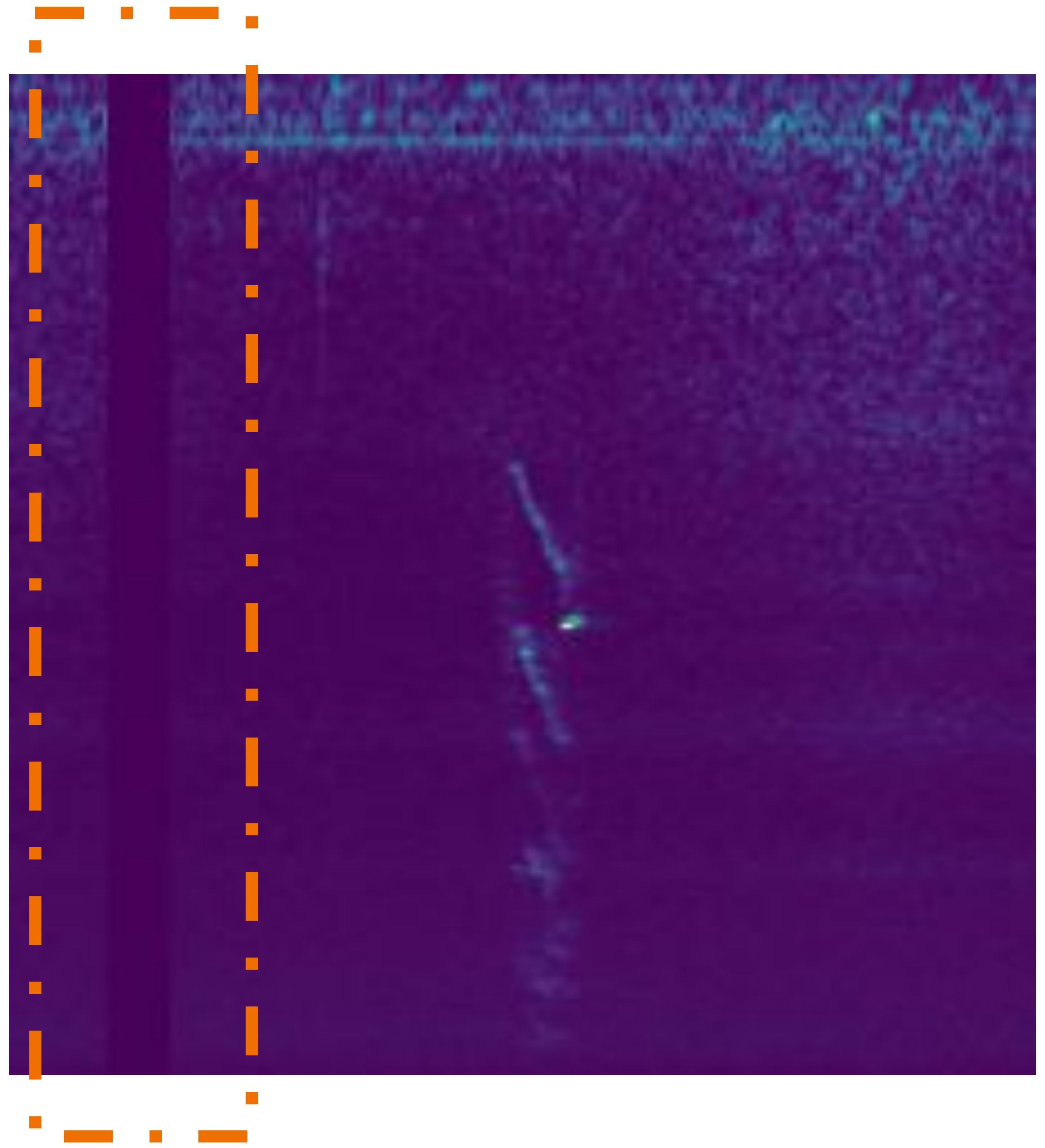
Spectrogram is masked based on frequency by random values.



Time Masking

nas.TimeMaskingAug(zone=(zoneL, zoneR), coverage = coverage)

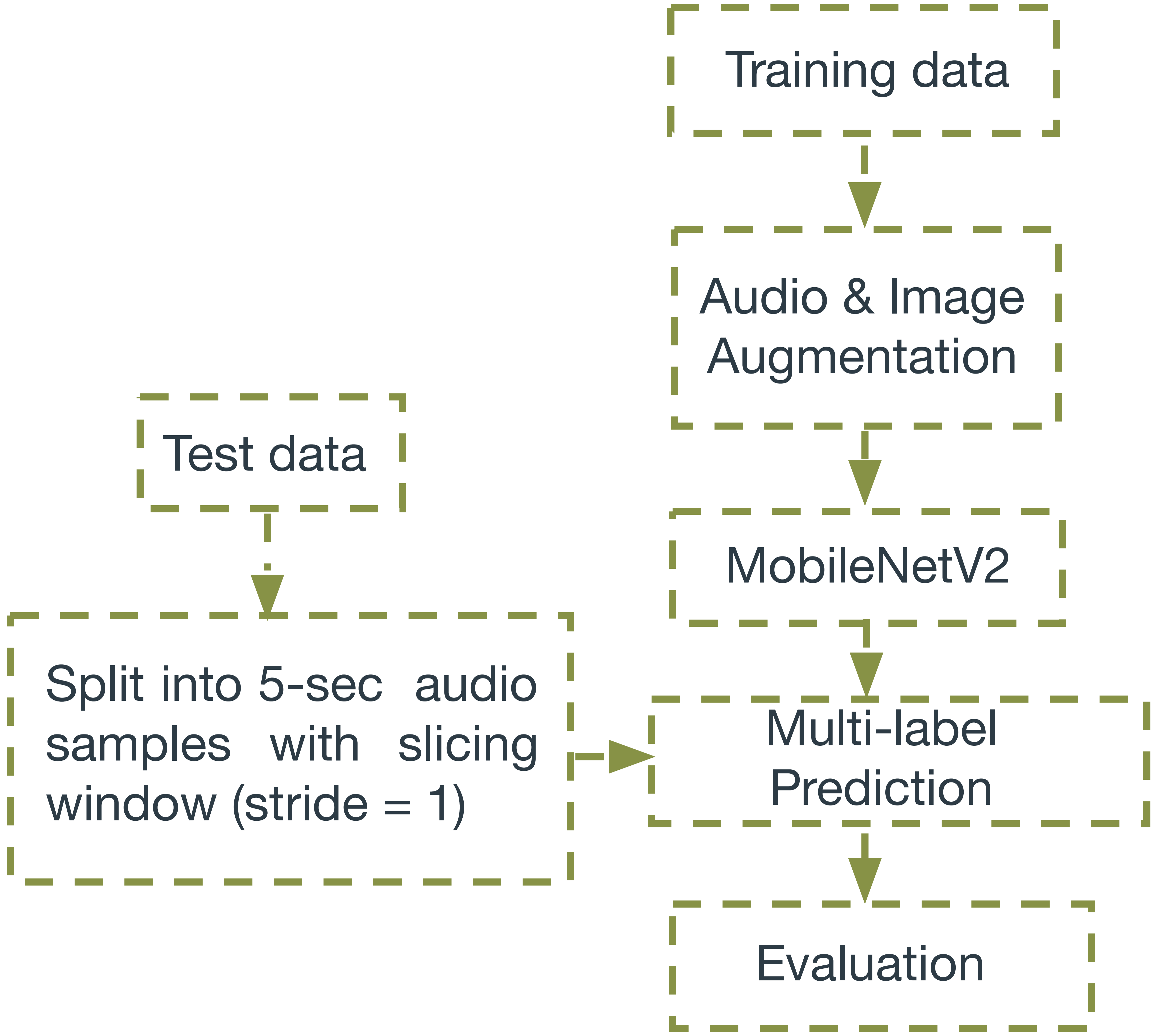
Spectrogram is masked by random values within certain time period



Experiments & Results

Experiment WorkFlow

35 Experiments were done!



Evaluation Results

Validation	Recall	mAP	Precision
Baseline	0.807	0.88	0.89
Noise injection	0.61	0.86	0.879
Shifting time	0.35	0.67	0.85
Changing pitch	0.79	0.9	0.91
Changing speed	0.85	0.84	0.63
All audio augmentation	0.767	0.887	0.894
Loudness	0.745	0.904	0.930
Freq mask	0.492	0.789	0.922
Time mask	0.536	0.840	0.926
All audio & spectrogram augmentation	0.885	0.915	0.886
All audio & spectrogram augmentation w/ mixup	0.360	0.736	0.840

Test	Recall	mAP	Precision
Baseline	0.134	0.27	0.44
Noise injection	0.118	0.235	0.431
Shifting time	0.078	0.158	0.415
Changing pitch	0.39	0.28	0.30
Changing speed	0.187	0.213	0.292
All audio augmentation	0.418	0.312	0.316
Loudness	0.175	0.301	0.419
Freq mask	0.186	0.314	0.417
Time mask	0.071	0.183	0.33
All audio & spectrogram augmentation	0.44	0.28	0.26
All audio & spectrogram augmentation w/ mixup	0.06	0.21	0.38

Discussion & Conclusion

Compared with baseline model:

- ❖ Random combination of audio augmentations achieves best overall performance

  mAP increases 15.5%

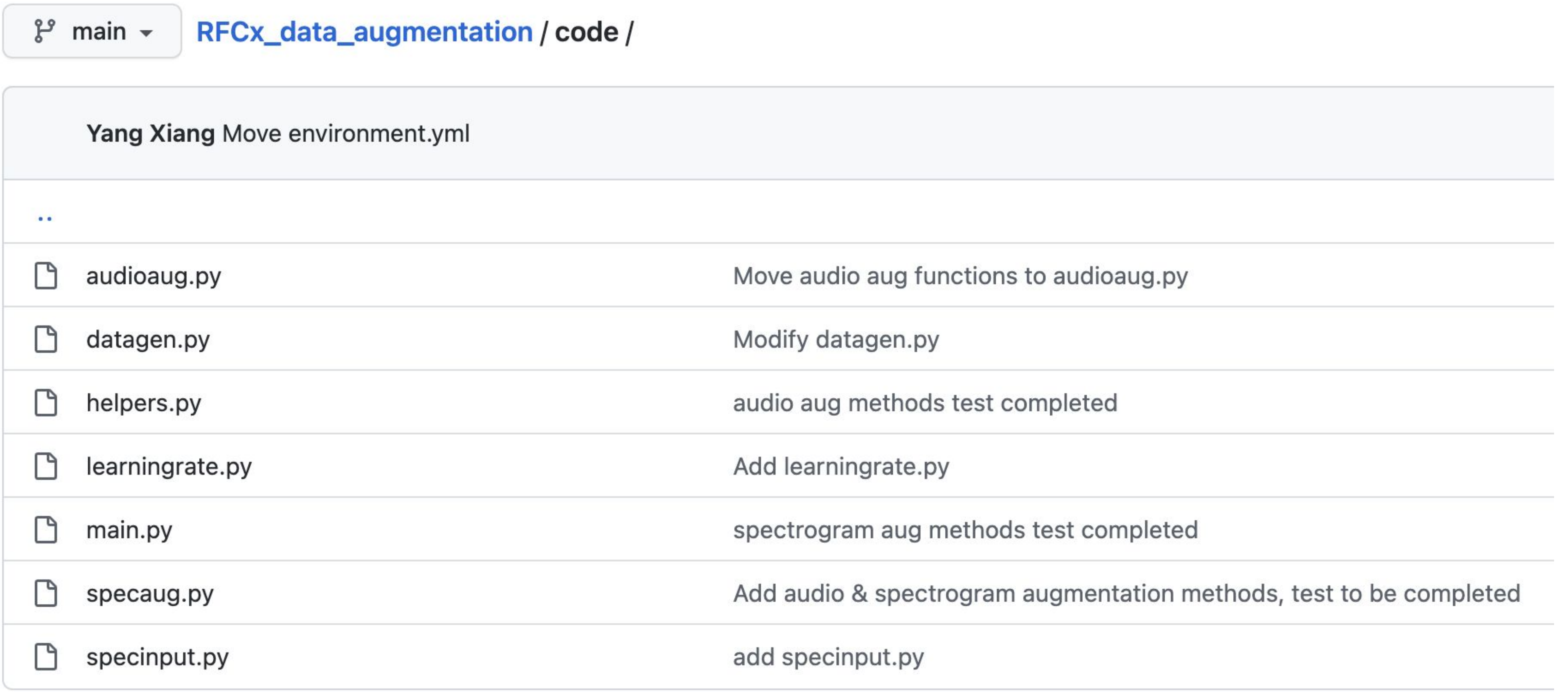
- ❖ Extra spectrogram augmentation doesn't significantly further improve

- ❖ Time-related augmentation methods might be used with caution

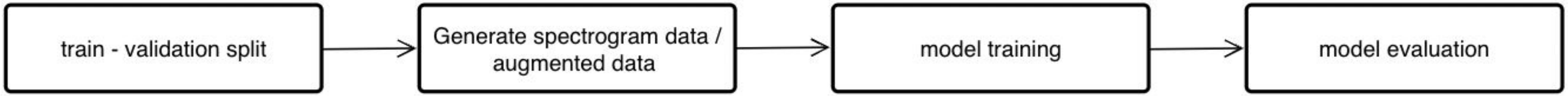
Pipeline Construction Based on Experiments

Pipeline on Github

- Link - https://github.com/YangXiang-Sunny/RFCx_data_augmentation



Full Procedure



Pipeline Construction Based on Experiments

Options

--help / -h
--input / -i: audio / spec
--input_path: input data path
--output_spec_path: output spectrogram data path
--train_val_split / -s: None / 0.8 / …
--aug / -a: noise_injection / shift_time / change_pitch / change_speed / time_mask / freq_mask / loud
--loss / -l: binary_crossentropy / masked_loss
--model_path: model path
--skip_train: evaluation only, skip training process
--test_path: test data path
--output_test_path: path of splitted test data
--slice_test_path: sliced test data path

Command line options

Usage

- Audio input
  python main.py --input audio --input_path ~/data/rfcx-harvard-ds/puerto-rico/train/ --output_spec_path ~/data/spec_data/ --train_val_split 0.8 --loss binary_crossentropy --model_path ~/code/model/model_test/ --test_path ~/data/rfcx-harvard-ds/puerto-rico/test/audio/ --output_test_path ~/data/test_data_preprocessed/ --test_label_path ~/data/rfcx-harvard-ds/puerto-rico/test/test-labels.csv
- Spectrogram input with frequency mask augmentation, skip training
  python main.py --input spec --input_path ~/data/spec_data_freq_mask/ --model_path ~/code/model/model_freq_mask/ --skip_train --test_path ~/data/rfcx-harvard-ds/puerto-rico/test/audio/ --output_test_path ~/data/test_data_preprocessed/ --test_label_path ~/data/rfcx-harvard-ds/puerto-rico/test/test-labels.csv

Sample usages

Reference

Ma.E(2019,June.4) Data Augmentation Audio Medium. https://medium.com/@makcedward/data-augmentation-for-audio-76912b01fdf6.

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecological Informatics, 59, 101113.

Thi-Ly Vu, Zhiping Zeng, Haihua Xu, and Eng-Siong Chng. Audio codec simulation based data augmentation for telephony speech recognition. 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019.

Shengyun Wei, Shun Zou, Feifan Liao, and weimin lang. A comparison on data augmentation methods based on deep learning for audio classification.Journal of Physics: Conference Series, 1453(1):012085,jan 2020.

Lonce Wyse. Audio spectrogram representations for processing with convolutional neural networks. 06, 2017.6

Loris Nanni. Set of texture descriptors for music genre classification. 2014.

Daniel Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin Cubuk, and Quoc Le.Specaugment: A simple data augmentation method for automatic speech recognition. pages 2613–2617,09 2019.

Ilyas Potamitis. Automatic classification of a taxon-rich community recorded in the wild.PLoS ONE,9(5), 2014.

Branko Hilje and T. Mitchell Aide.Calling activity of the common tink frog (diasporus diastema)(eleutherodactylidae) in secondary forests of the caribbean of costa rica.Tropical Conservation Science 5((1):25–37, 2012.

Thomas Lidy and Andreas Rauber. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification.International Society for Music Information Retrieval Conference 2005

Listen to the Sound of Rare Species!

Thank You!

AC297r: IACS Capstone Project
Fall 2022

Yang Xiang, Li Sun, Meichen Dong
Yanqi Luo, Ziye Tao

RAINFOREST CONNECTION

IACS