Bird Species identification using Neural Networks (BirdCLEF 2024)

Amzi Jeffs
Junichi Koganemaru
Salil Singh
Ashwin Tarikere

https://www.kaggle.com/competitions/birdclef-2024/overview

BirdCLEF 2024

- Kaggle competition hosted by the Cornell Lab of Ornithology
- Participants are asked to train models reporting probabilities of whether each of 182 given species are present in a hidden test set of ~1100 audio clips
- Each test clip contains potentially multiple bird calls
- The models are scored on a custom macro-averaged ROC-AUC metric

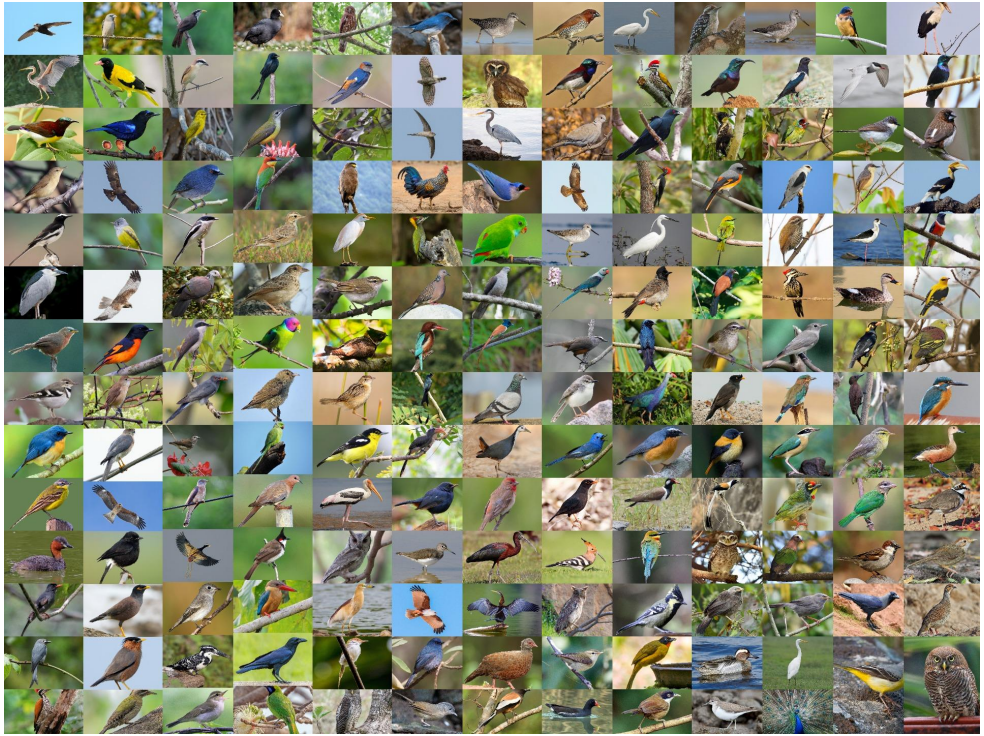
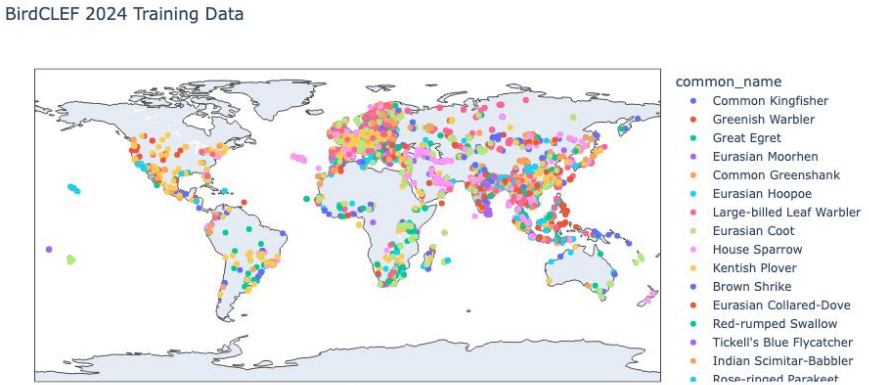
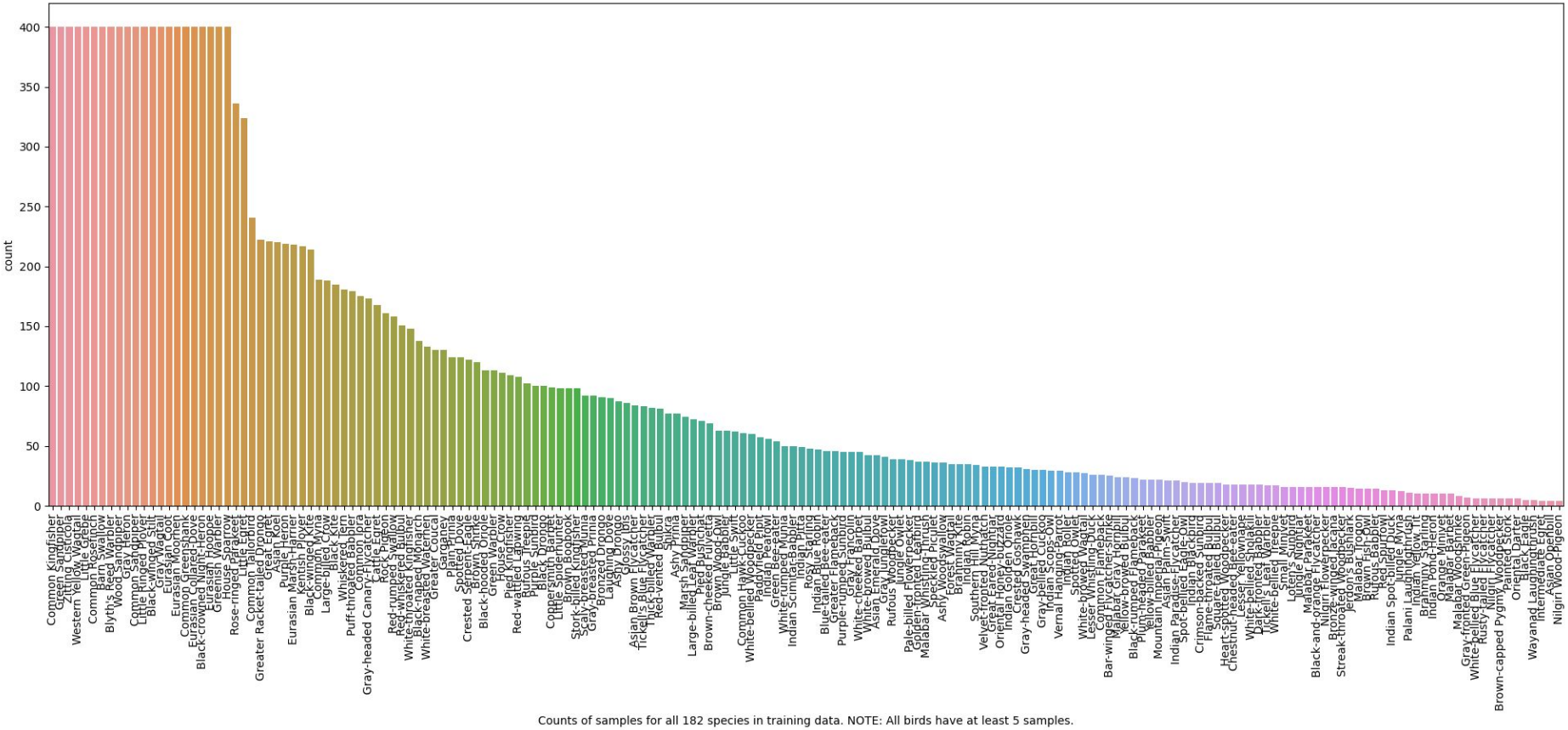


Image source: https://www.kaggle.com/competitions/birdclef-2024/discussion/491083

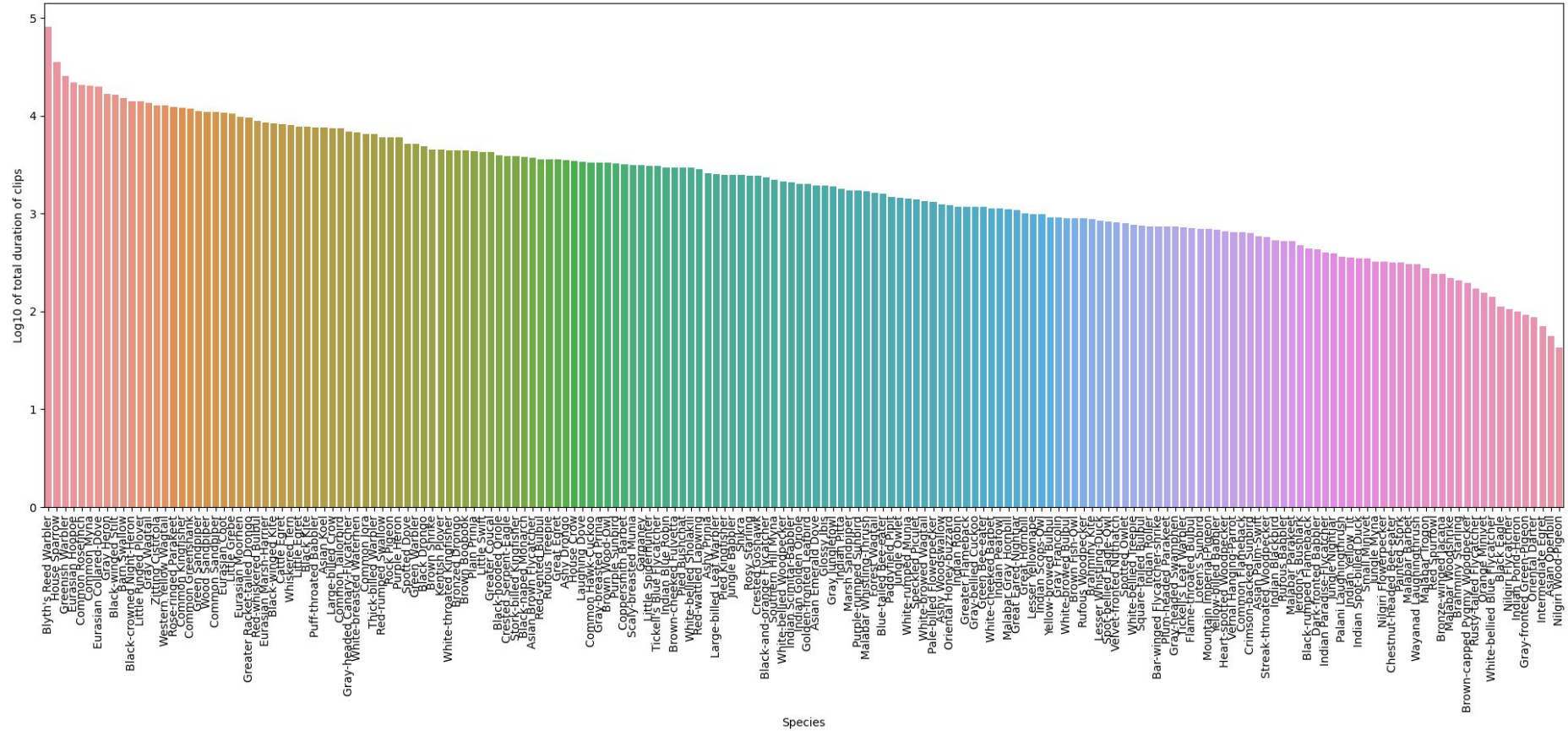
Dataset

- The organizers provided a dataset (23.43GB) containing training audio clips (7.81GB), metadata with information on the clips, and unlabeled soundscapes (15.61GB)
- For now we have treated this as a supervised problem by only using only the training audio clips and metadata
- True testing data is hidden, so we held out 20% of given data as a holdout set to validate performance

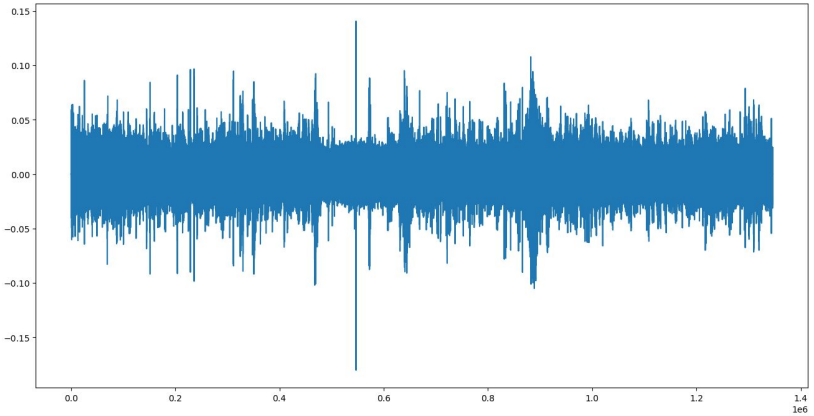


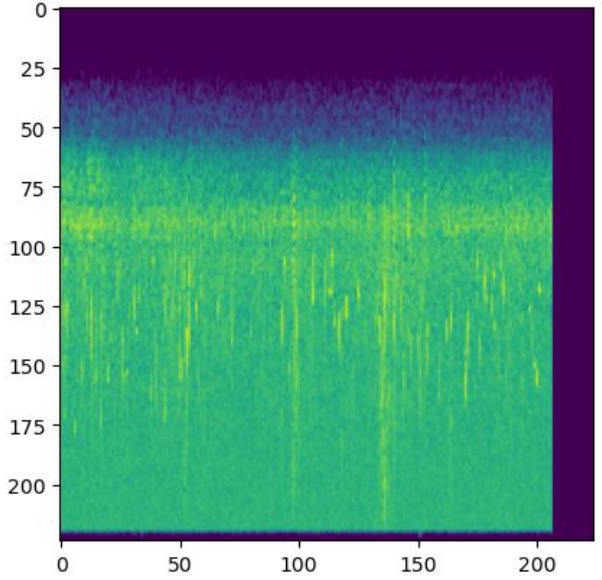


Exploratory data analysis: 19567 total clips

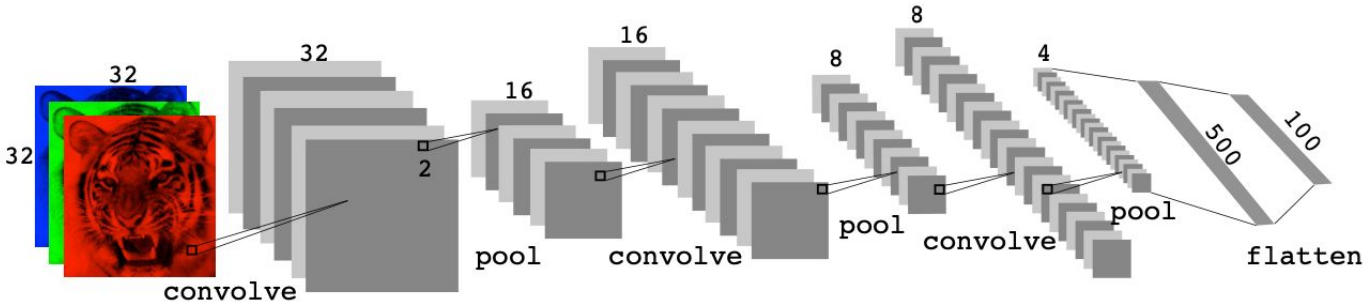


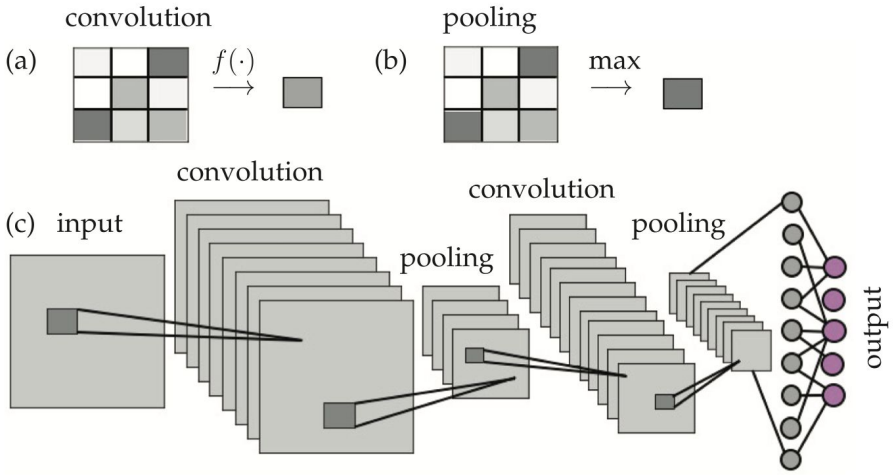
Exploratory data analysis: duration of audio clips in log plot





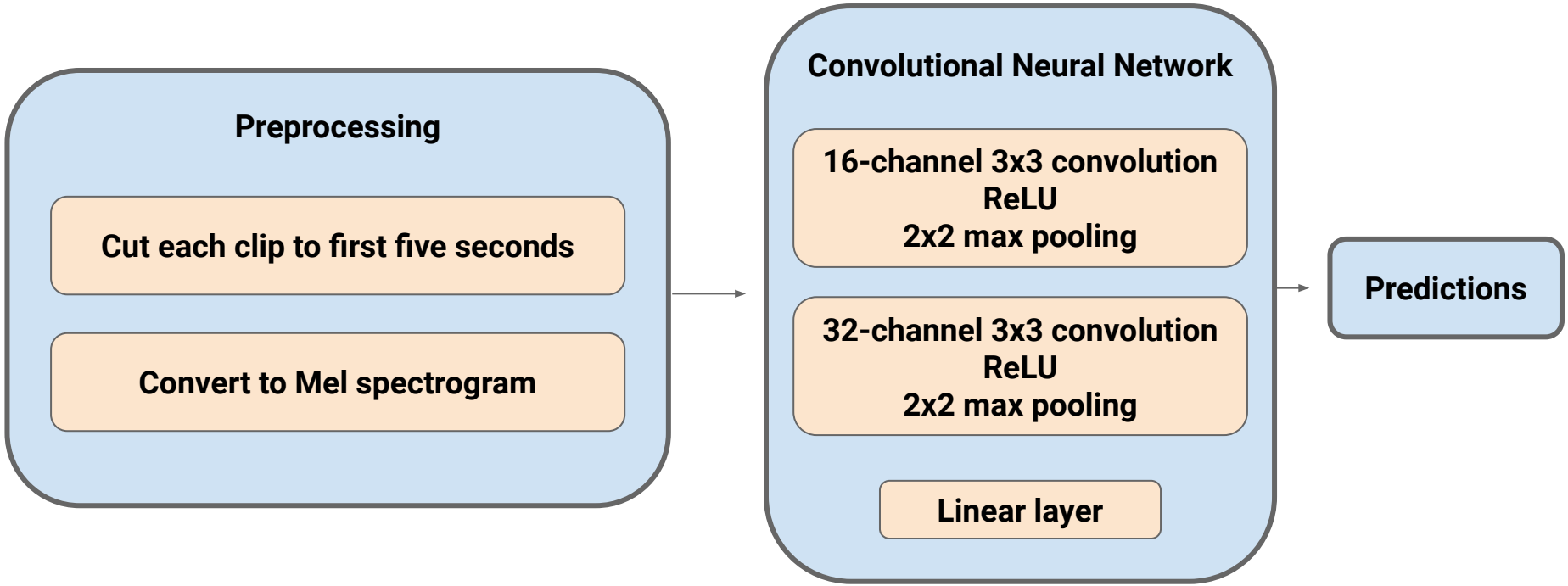
Main approach: audio to mel spectrogram as input into convolutional neural network (CNN)



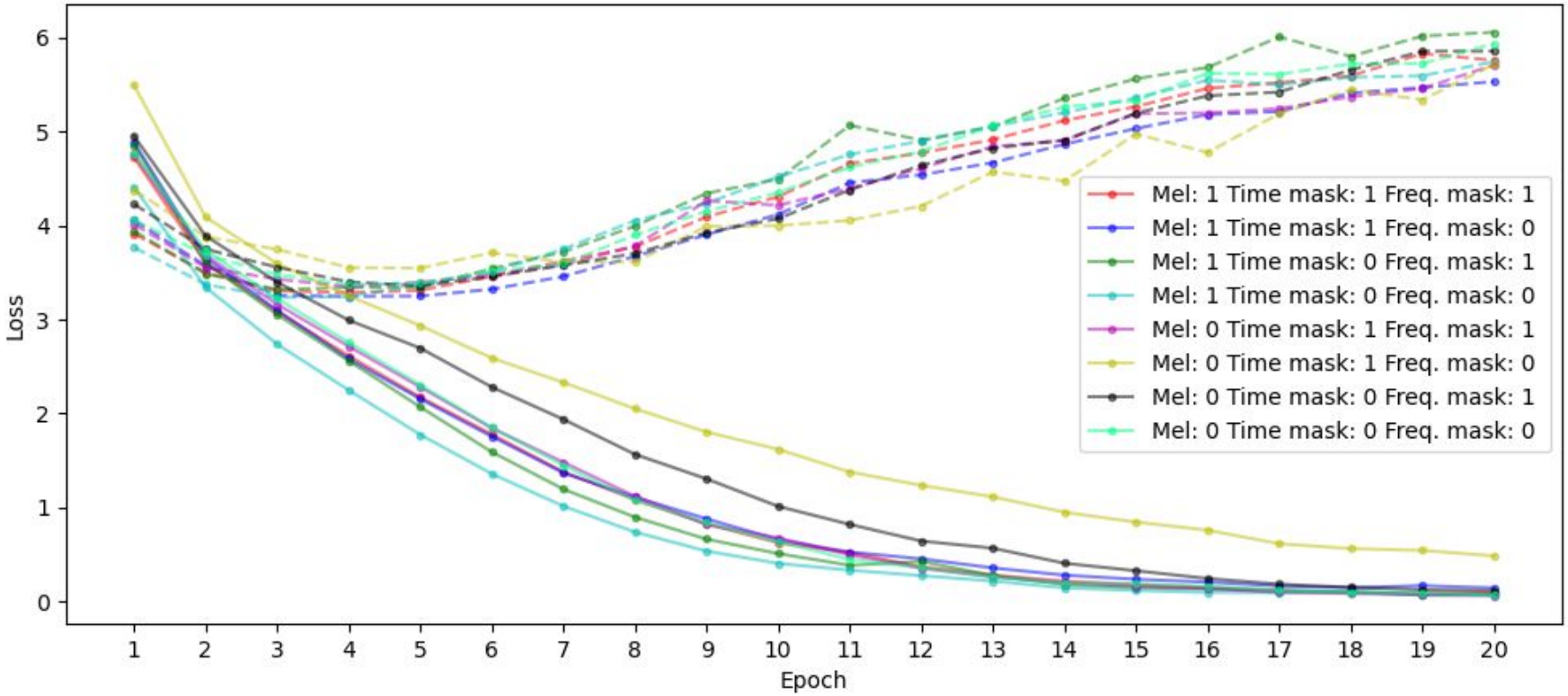


Neural Network Architecture: typical CNN architecture

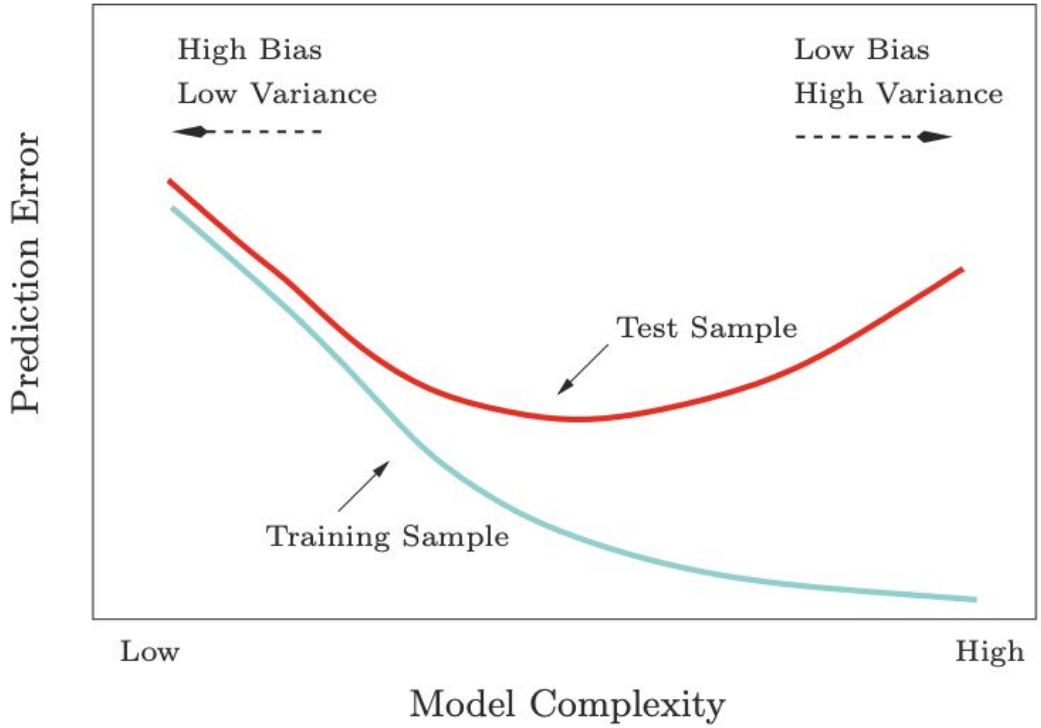
Image sources: Data-Driven Science and Engineering by Brunton and Kutz, An Introduction to Statistical Learning by James et. al.



Our simplest baseline, pared down from an example by Valerio Velardo.



2-Layer baseline training and validation losses, various preprocessing approaches



Bias-Variance trade off, underfitting vs overfitting

Key challenges

- Our models have a **tendency to overfit**, so while the model was able to learn from the existing training data, they did not generalize well to the hidden test data
- **Runtime** is 4~6 hours on an external cluster with a 128 core GPU
- The competition has a strict inference limit **2 hours on CPU only,** so the use of larger models (including ensemble models) is heavily penalized

Our approaches

Data augmentation:

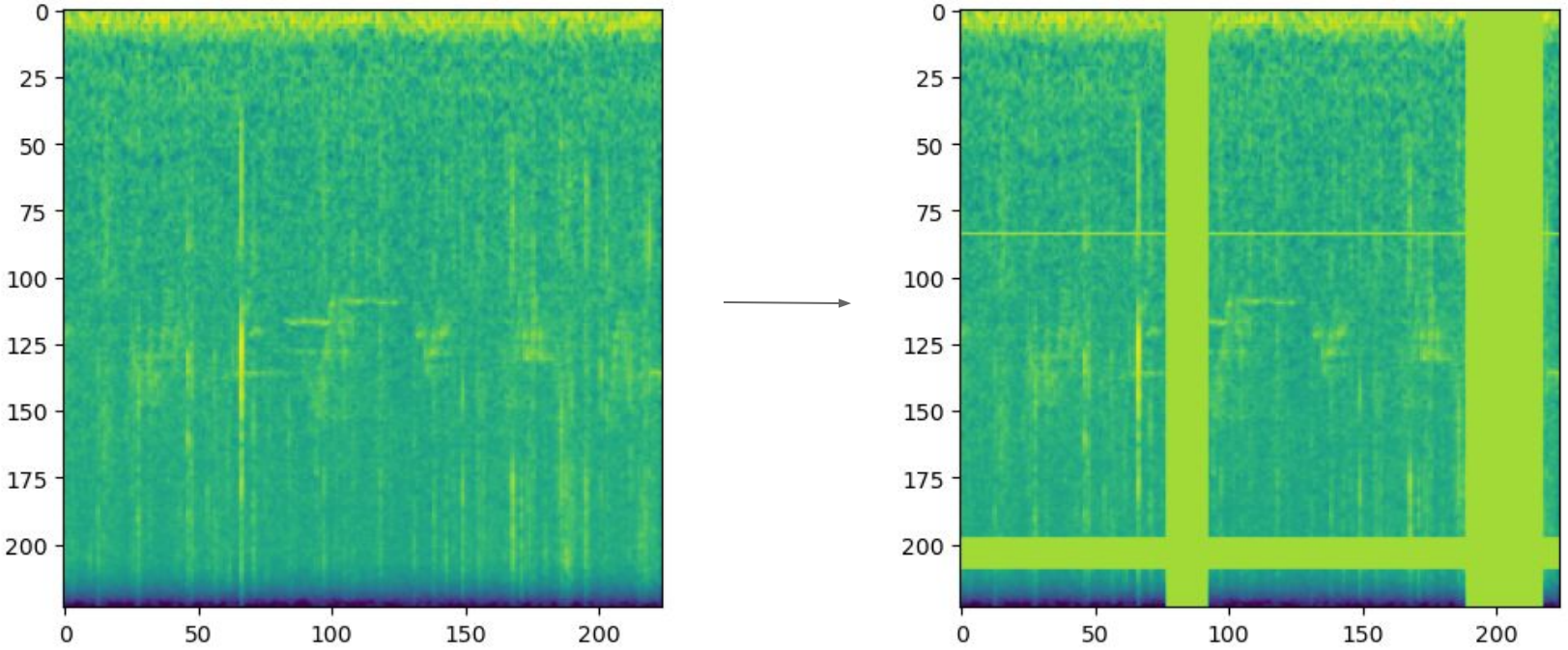
- Frequency and time masking
- Increase/decrease contrast (power)

Careful training

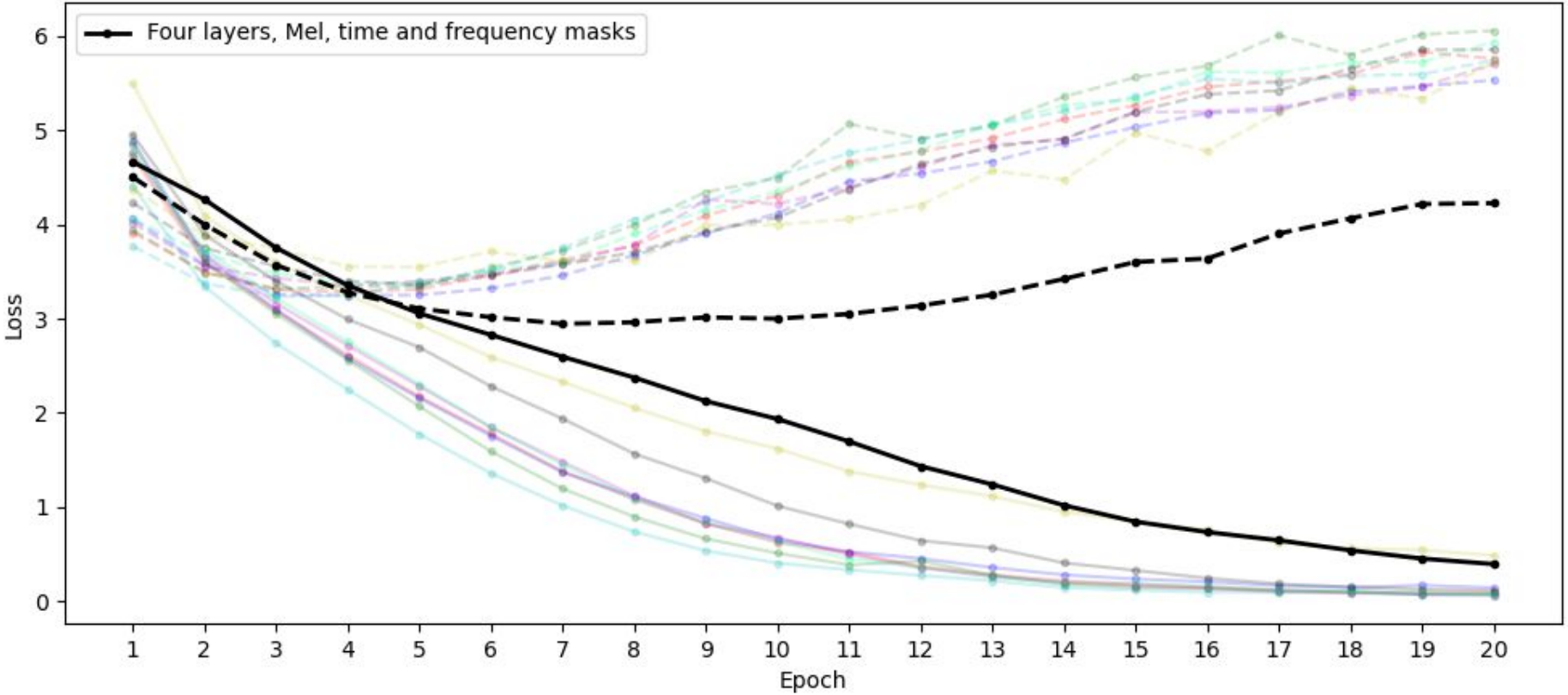
- K-fold cross-validation
- Weighted sampling of classes
- Learning rate scheduler
- Early stopping

Better architecture

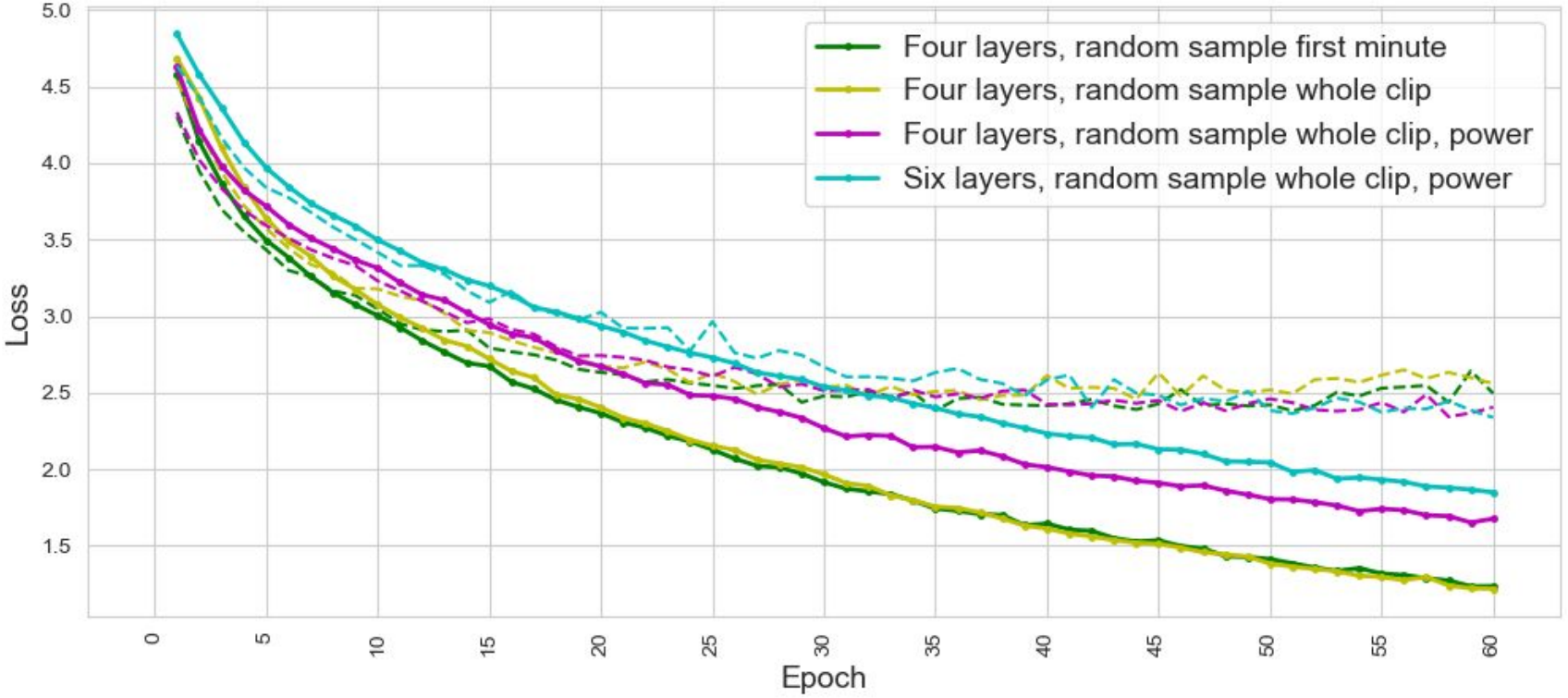
- Dropout layers
- Richer architectures



Preprocessing: frequency and time masking



Better architecture gave some improvements…



…more data and preprocessing gave far more!

Conclusion and future work

- **More Data**: Using additional data from past years or Xeno-canto database.
- **Better Data**: Devising a novel call-no-call detection function to improve sample quality of 5s clips and use techniques like gradient boosting decision trees to eliminate false positives.
- **Realistic Data**: Noising, denoising, altering context – via distortions or resizing – of signal to resemble the test data.

Conclusion and future work

- **Task-tailored loss functions**
- **Architecture tweaks**: Temporal context; more OOTB pretrained model weights.
- **Ensemble methods** with knowledge distillation. **Inference speed up** using ONNX or OpenVINO
- **Few shot techniques** and synthetic data generation.
- **Penalization** to reduce output probabilities in proportion to training data.
- **One week left on Kaggle!** Current leaderboard score is 0.61, which is in the middle of the pack. Aiming for 0.65+.

References

- Holger Klinck, Maggie, Sohier Dane, Stefan Kahl, Tom Denton, Vijay Ramesh. (2024). BirdCLEF 2024. Kaggle. https://kaggle.com/competitions/birdclef-2024
- Valerio Velardo. PyTorch for audio tutorial series. 2021. https://github.com/musikalkemist/pytorchforaudio/tree/main
- Brunton, Steven L., and J. Nathan Kutz. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge: Cambridge University Press, 2019.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York, Springer, 2009.
- All the Birds We Cannot See. https://www.kaggle.com/competitions/birdclef-2024/discussion/491083
- Why there is a gap between CV and LB. https://www.kaggle.com/competitions/birdclef-2024/discussion/498404