Harvard John A. Paulson
School of Engineering
and Applied Sciences

Silence to Sound

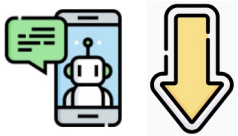
Generate Visually Aligned Sound for Videos

Bailey Bai, Danni Lai, Tiantong Li, Yujan Ting, Yong Zhang, Hanlin Zhu
Oct 24th, Midterm Presentation

Problem Statement

Motivation

- Video has no sound
- Noise in soundtrack



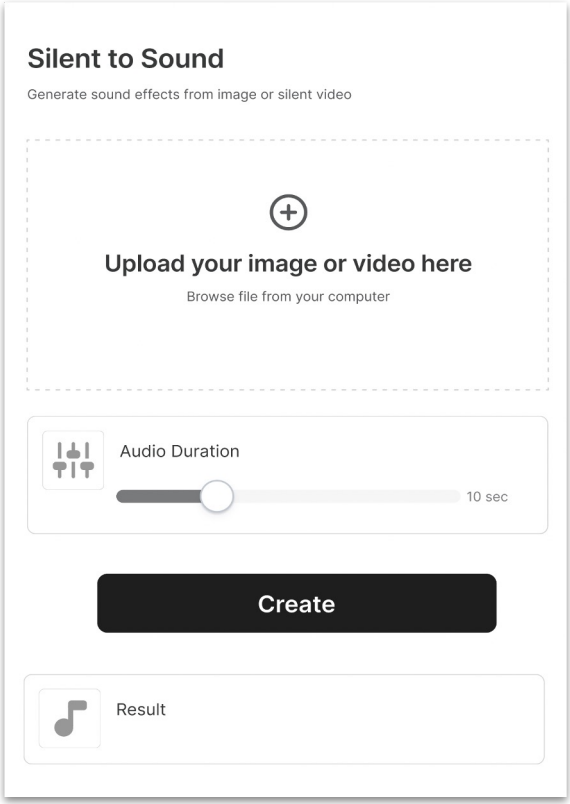
Our Solution: API that generates sound

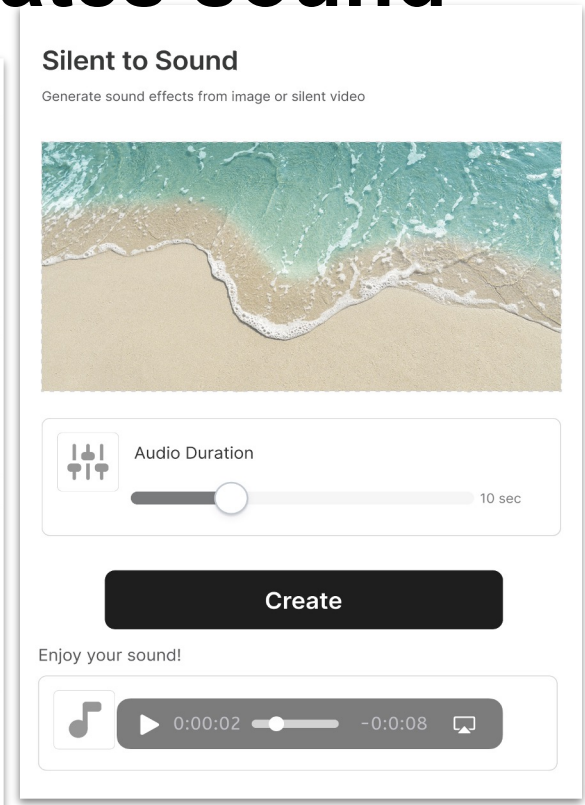
Motivation

- Video has no sound
- Noise in soundtrack

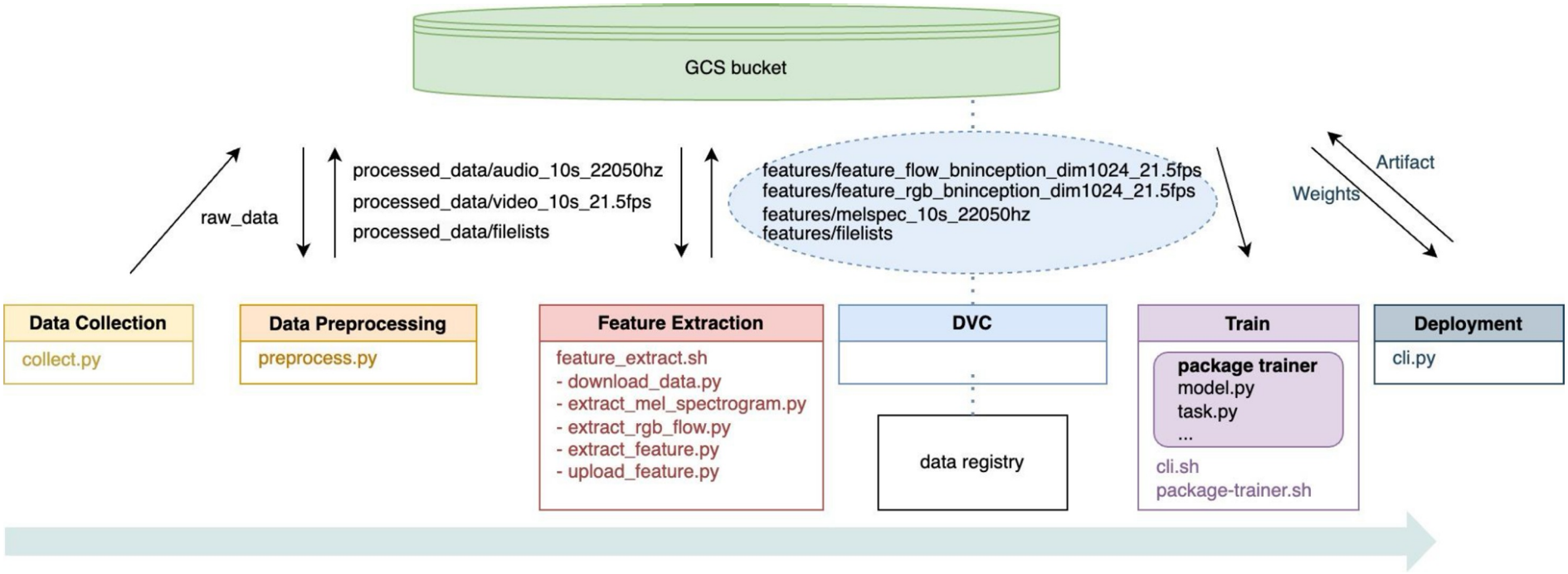
Potential Usage

- Video sound restoring
- Video de-noising
- Viewer experience enhancement
- Accessibility

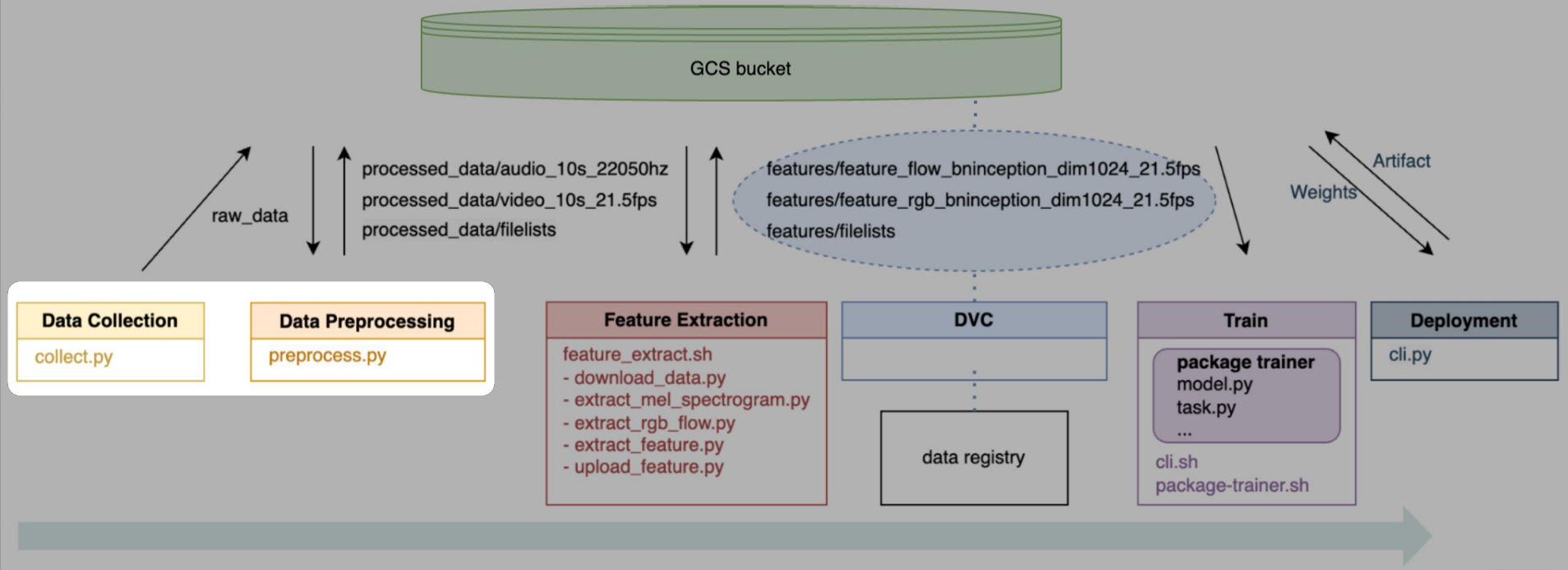




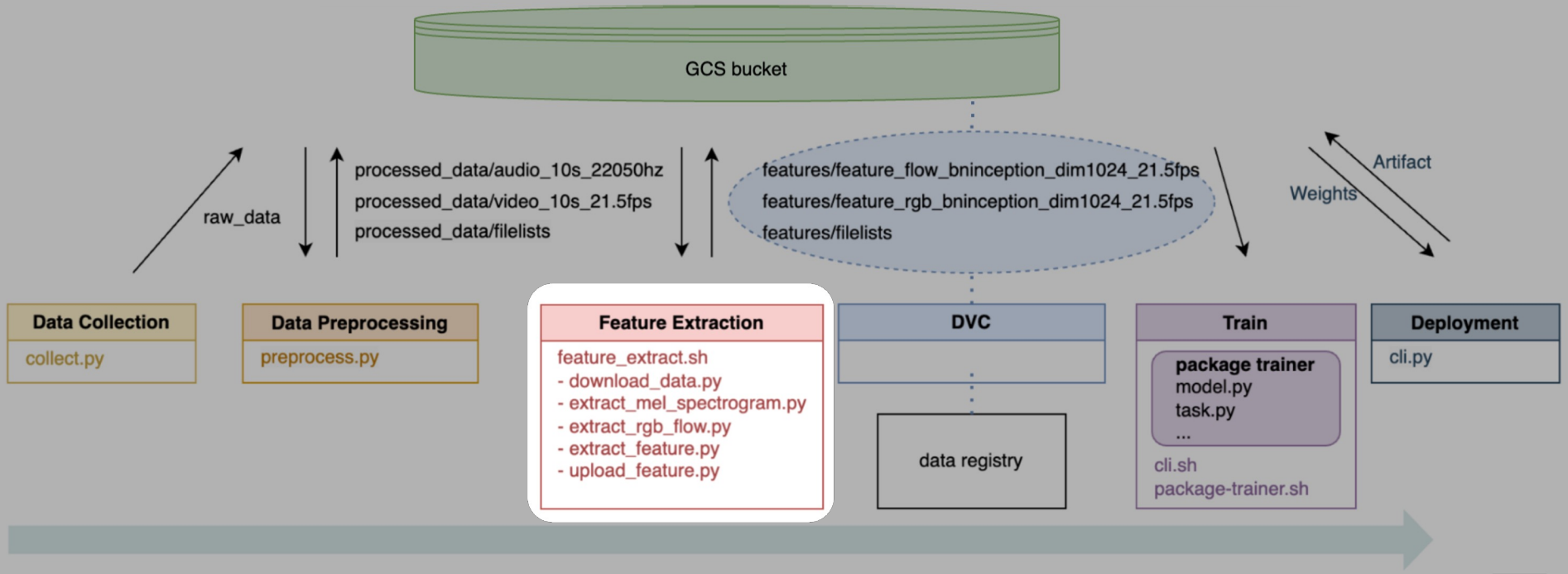
Pipeline Overview



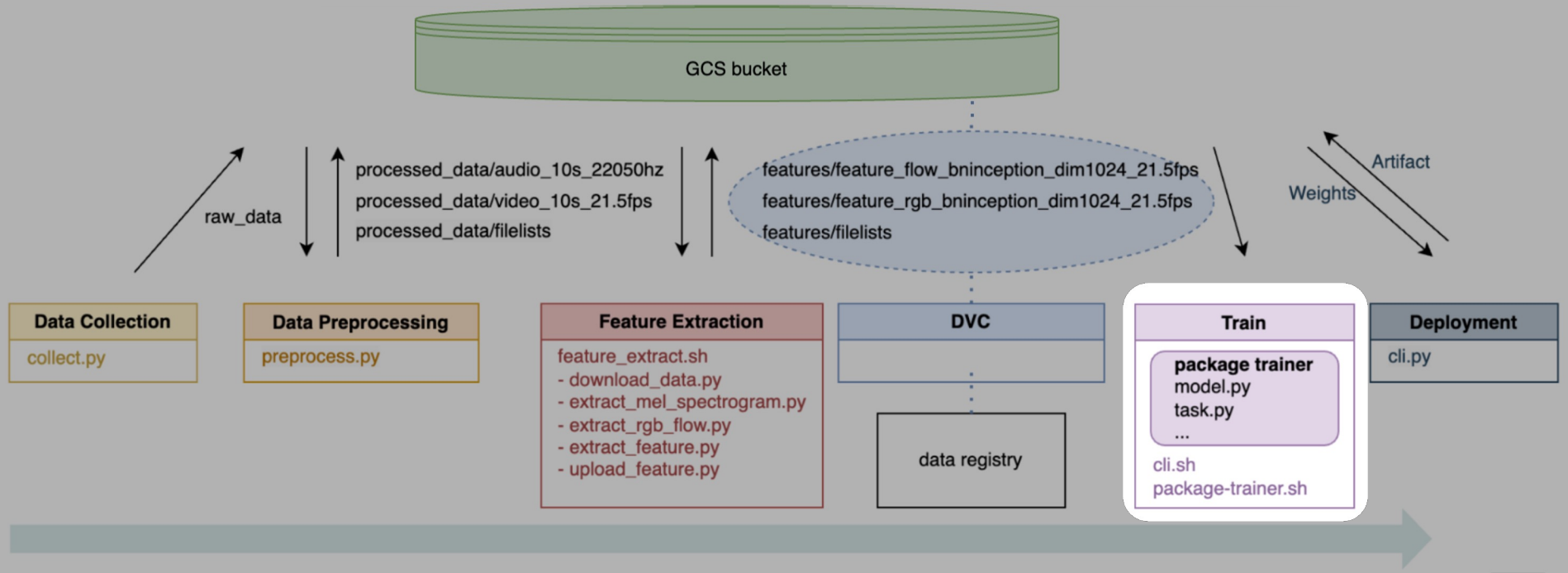
Pipeline Overview



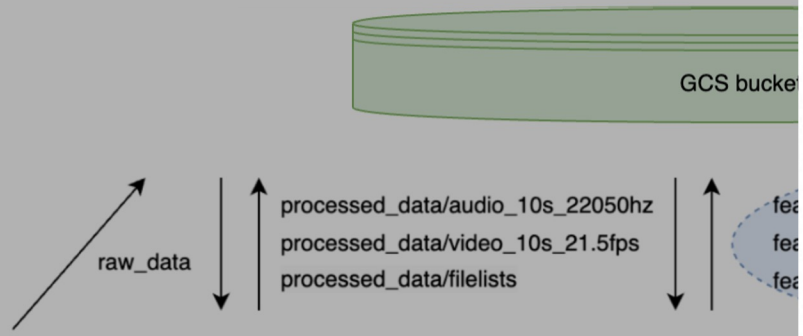
Pipeline Overview



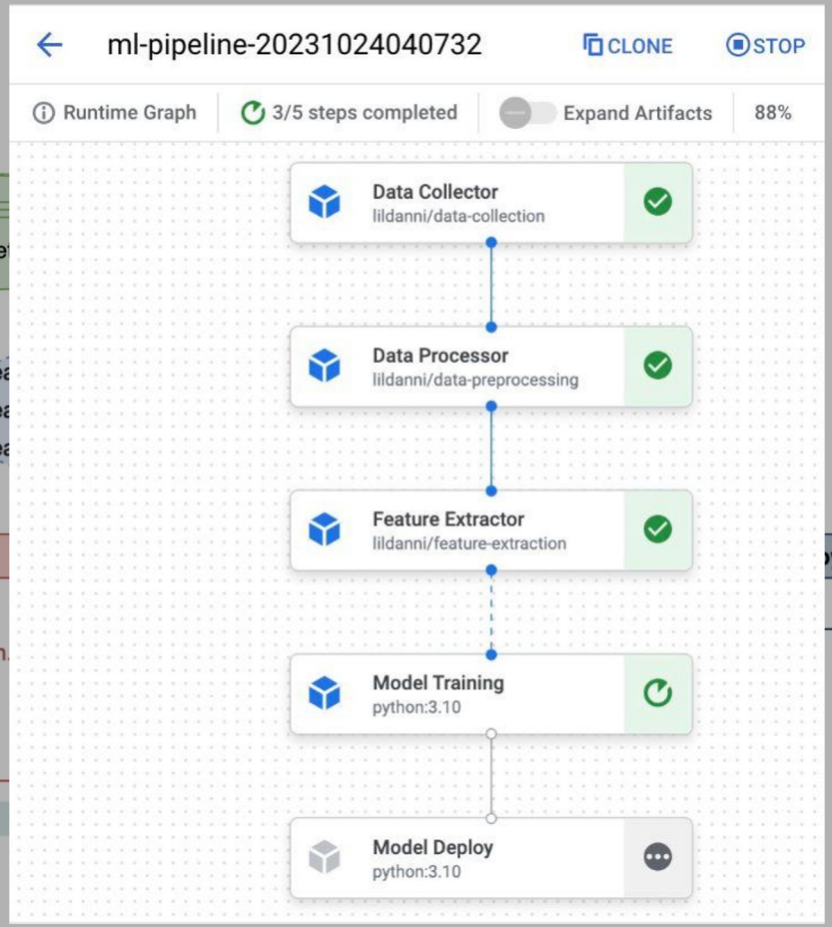
Pipeline Overview



Pipeline Overview



Data Collection	Data Preprocessing	Feature Extraction
collect.py	preprocess.py	feature_extract.sh - download_data.py - extract_mel_spectrogram. - extract_rgb_flow.py - extract_feature.py - upload_feature.py



Model Training

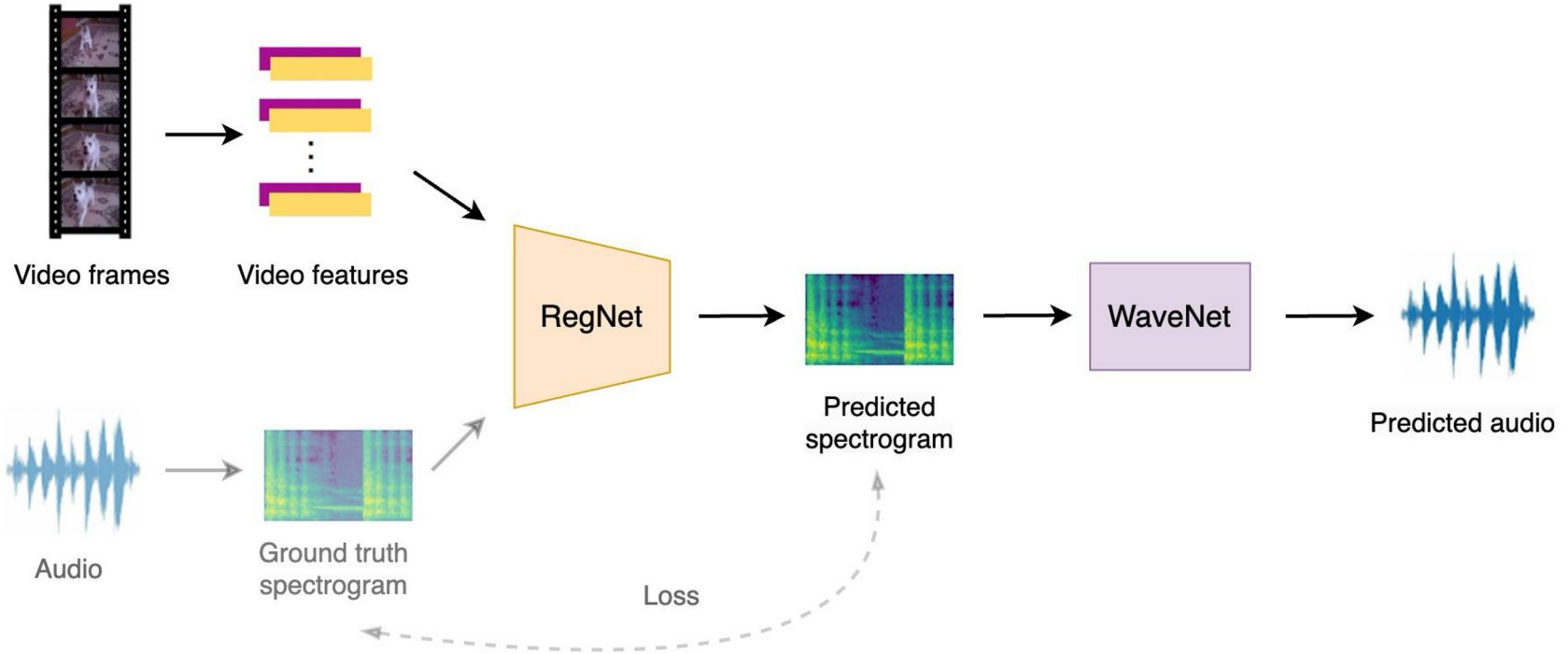


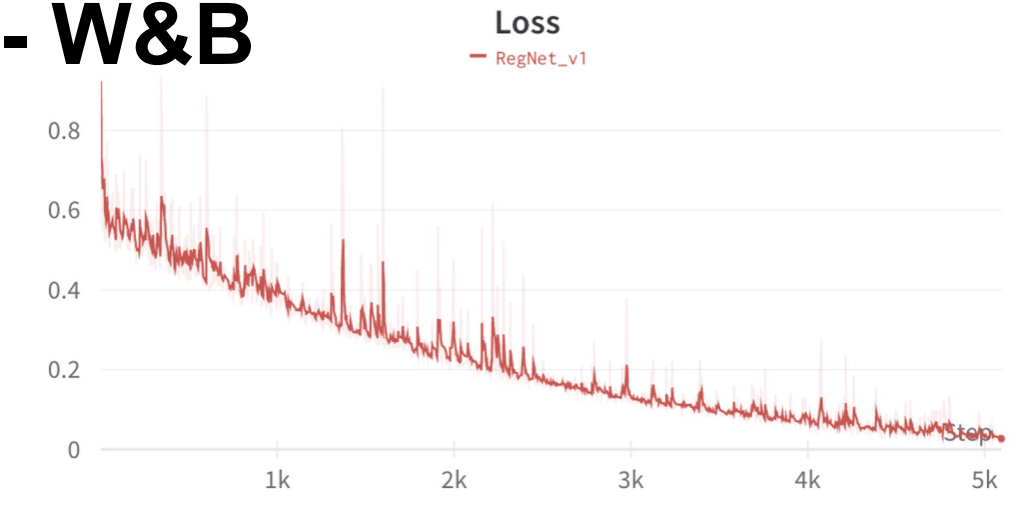
Image symbols from Chen et al. (2020). Generating Visually Aligned Sound From Videos. IEEE Trans. on Image Processing. [https://doi.org/10.1109/tip.2020.3009820]

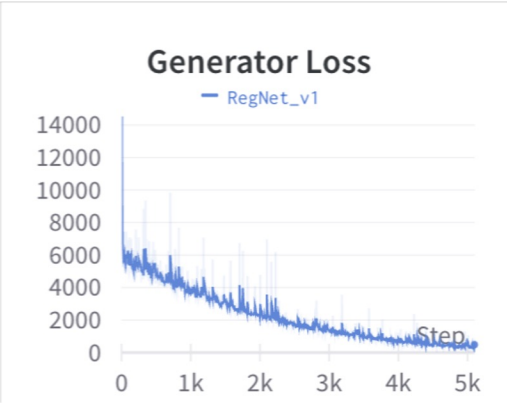
Training History - W&B

- 30 Epochs
- lr = 2e-4
- batch_size = 4

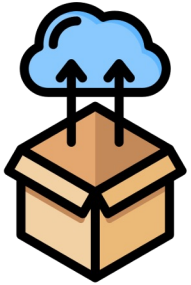
Regularization

- weight_decay = 1e-6
- grad_clip_thresh = 1.0

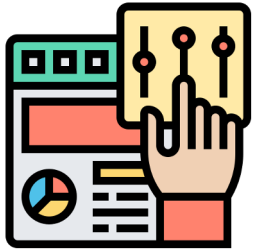




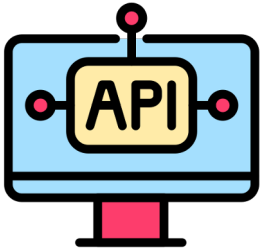
Next Steps



Finalize Model Deployment



Design User Interface



APIs & Frontend-Backend Integration

Thank you!