Commonly Encountered Animal Sounds Recognition in NTHU

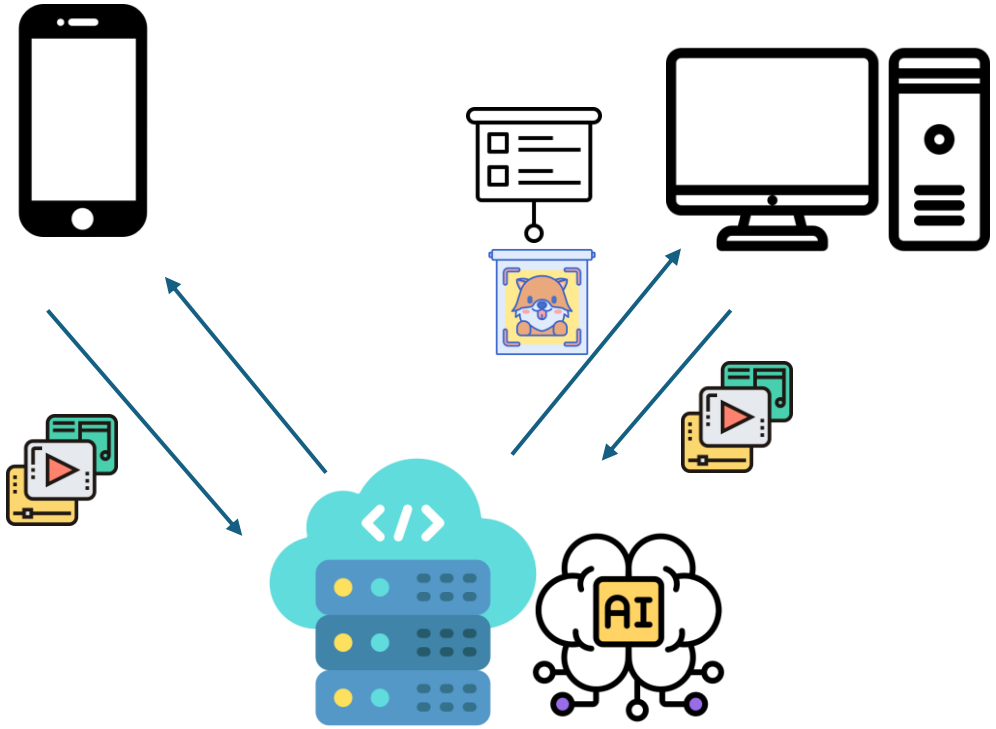
Introduction

- Gathering frequently heard animal sounds through open data platforms

- Our model will be trained using relevant animal sound datasets collected from open data sources

- Through a web platform, users will be able to upload their own audio recordings or files. The system will analyze the input and return relevant information and images of the identified animal

System Flow



Model

- VGGish takes audio waveform as input and produces 128-D embedding representations of the semantic content of the waveform.

- To address the issue of varying audio file lengths, we use data segmentation to split the data into segments every 5 seconds for training and predicting.

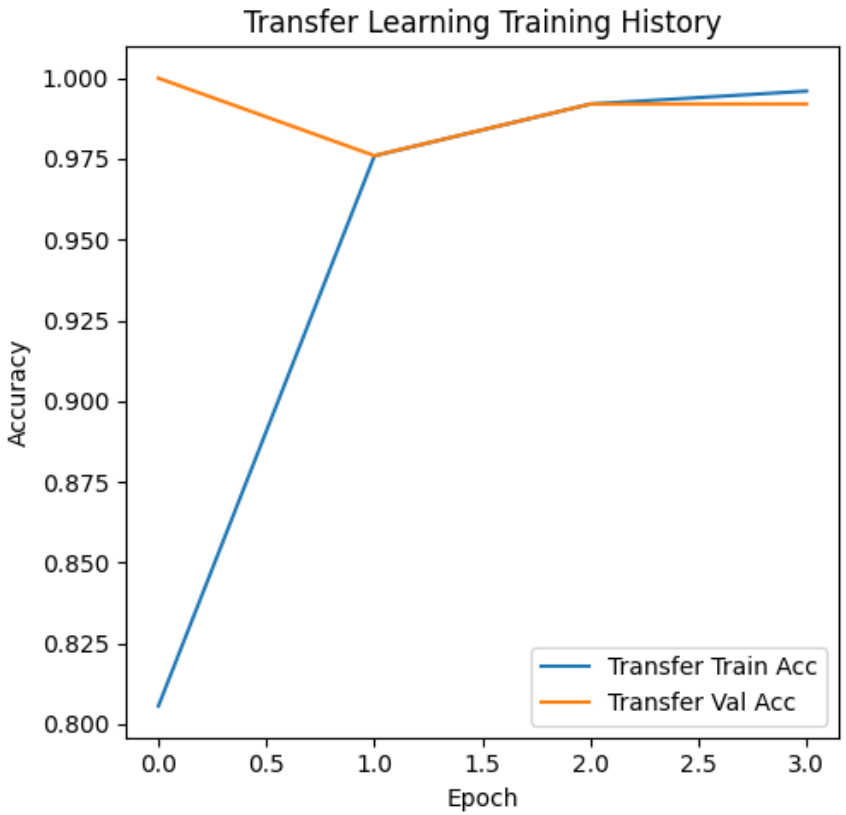
- Furthermore, we can predict animal sound with different format of file

  (.wav, .mp3, .m4a, .aac, .flac, .ogg, .mp4, .avi, .mov, .mkv)

Model

- Dataset:
  One audio file per animal

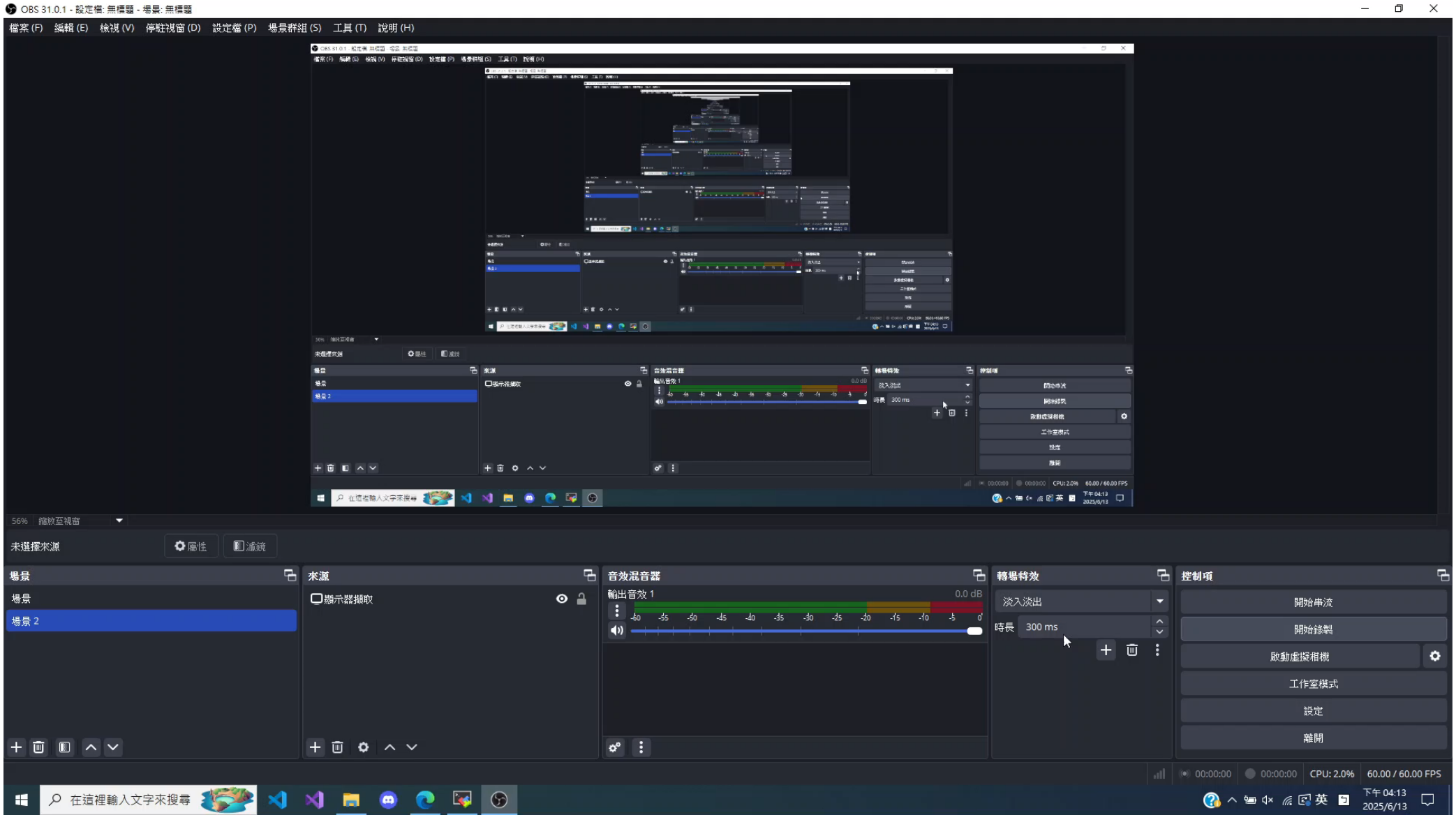
- Training environment:
  Ubuntu 20.04.4 LTS
  Python 3.10
  CPU: AMD Ryzen 9 5900X 12-Core Processor
  GPU: GeForce RTX 3080 Ti



Web

Category	Technologies / Tools Used
Backend Framework	Django (Python)
Frontend Interface	HTML, Bootstrap
Responsive Design	Bootstrap Grid System (adapts to mobile/tablet/desktop)
Database Management	MySQL, Django ORM
File Handling	Django File Upload, 'MEDIA_ROOT' / 'MEDIA_URL' settings
Development Tools	PyCharm, Python virtual environment ('venv')

DEMO


OBS 31.0.1 - 設定檔: 無標題 - 場景: 無標題
檔案 (F) 編輯 (E) 檢視 (V) 停駐視窗 (D) 設定檔 (P) 場景群組 (S) 工具 (T) 說明 (H)
未選擇來源
屬性
濾鏡
場景
場景
場景 2
來源
顯示器擷取
音效混音器
輸出音效 1
0.0 dB
轉場特效
淡入淡出
時長 300 ms
控制項
開始串流
開始錄製
啟動虛擬相機
工作室模式
設定
離開
00:00:00
00:00:00
CPU: 2.0%
60.00 / 60.00 FPS
在這裡輸入文字來搜尋
下午 04:13
2025/6/13