

Data Processing Notebook

Thursday, October 09, 2025 12:30 PM

To Start we have:

- 30fps video (full face)
- 48khz audio
- .mat files outlining ROI
- Txt labels

Step 1: Download dataset

- We'll try download O2M for this exercise

Step 2:

- Convert video to 25fps
- Convert Audio to 16kHz
- Create mouth crops with .mat

Data Augmentation

- Interpolation technique section
 - Can either use 25fps video (crop or full face?) or use original 30fps
 - Probably 30fps because that means less interpolation?
 - Create function that aligns audio in different ways
- Smart Blurring
 - Code to apply this

Prepare manifests and necessary files for testing in neat format