

Affective Computing 2025 Exercise 1:

Micro-Gesture Annotation

General Information

Data collection is the first and one of the most important steps in scientific research. How to collect, clean, and annotate data for scientific purposes is the first step.

In this assignment 1, we use one example to demonstrate the whole process of dataset preparation, using [our research work about micro-gesture](#) published on CVPR 2021 (the most well-known AI conference in the field). You can briefly read the paper to better understand the background story, but it's completely fine to move to the next without reading.

In this assignment, you will gain hands-on experience in **micro-gesture annotation** from video data.

Micro-gestures are subtle, short-duration movements that often carry critical information about human intentions, emotions, or decision-making.

Multimodal AI systems are increasingly used for video understanding, but their performance relies on the large-scale training with human-manually labeled ground truths (i.e., we need to label the data to teach AI). Especially for fine-grained micro-gesture detection, the performance of current AI systems is still **poor, unstable, and inconsistent**. Therefore, **manual annotation remains essential** for creating reliable datasets and for benchmarking the progress of automated methods.

This assignment is designed to provide you with both practical skills in manual labeling with our own developed annotation tool and a deeper understanding of the strengths and limitations of current AI approaches.

The overall goals of this assignment are: 1. Learn how to annotate data for scientific purposes from raw samples. 2. Learn how to develop various tools for data annotation. 3. Learn how to conduct a data annotation quality check

Total points: 10 points (exercise grading) + 5 bonus points (optional, will be added directly to final grading)

What to Return

At the end of this assignment, you must submit the following items:

1. **Per-video annotations** in .txt format (one file per video).

2. **Merged submission file** submission.csv (produced with the provided merging script).
3. **Kaggle competition submission & score screenshot**
4. **Analysis summary** (optional, for the bonus task).

Tasks

Task 1 (2 points)

Familiarize yourself with the annotation platform. Refer to **Usage of annotation platform.pdf**. It's important to have a well-developed annotation tool for data labeling and analysis, and we will introduce our self-developed annotation tool as an example to conduct the data annotation. You need to:

- Use the platform at: [Annotation Platform Link](#) . (You can also find the **annotation_review_new.html** in the downloaded exercise1 folder.)
- Explore its functions for video playback, segment selection, and saving labels.
- Screenshots and usage instructions will be found in **Usage of annotation platform.pdf**.

Task 2 (3 points)

Annotate the assigned videos.

- Each student is assigned a group of videos (see **Data grouping information.xlsx** for your allocation).
- Download the corresponding videos from [Data Annotation Link](#)
- Annotate **micro-gestures** in each video using the platform.
- Export annotations as .txt files.
- **Important:** Do **not** change the filenames of the .txt files.

Task 3 (1 point)

Merge your annotations into one submission file.

- Refer to the provided helper script to merge all .txt files into a single CSV file:

S3_merge_txt_to_csv.py

- The final submission.csv must contain the following columns:

ID, video_id, class, start_time, end_time

- ID must be a unique row index (0, 1, 2, ...).

Task 4 (4 points + 2 bonus points)

Participate in the Kaggle competition. The Kaggle platform is one of the biggest platforms for the AI and data scientist community, with various datasets and competitions available, feel free to explore the platform and see some [previous competitions](#) hosted by us.

We opened a new track on Kaggle to simulate a micro-gesture data annotation competition. You need to submit your annotation, and there will be a cross-validation on your annotated results. The purpose of this task is to show you that human annotation can be random, and there should be some quality control methods to reduce human error.

- Register for a Kaggle account (if you don't already have one).
- Join the competition:

[Kaggle Competition Link](#)

- Submit your submission.csv file, and the server will automatically check your annotation against our labeled ground truths.
- Check your **score and ranking** on the leaderboard.
- Make sure your annotation score gets at least 60 to get the full 4 points.
- You may resubmit multiple times before the deadline to improve your performance, and the top 3 annotation results will get 2 bonus points.

Task 5 (Open-ended, bonus 3 points)

Compare manual annotation with AI-generated annotations.

- Try using an existing **open-source model** for gesture or action segmentation. Possible candidates include:
 - **OpenGVLab InternVideo / InternVideo2** (video understanding foundation models).
 - **VideoMAE** (masked autoencoder for video).
 - **ActionFormer / SlowFast** (gesture/action recognition baselines).
 - **GPT-4V / LLaVA / Video-LLaMA** (multimodal LLMs with video input, if available).
- Apply the model to your assigned videos to produce predicted gesture annotations.
- Compare AI-generated annotations with your manual labels.
- Write a short summary analyzing:
 - Where the model performs well.
 - Typical errors (missed gestures, false positives, temporal misalignment).
 - Advantages and limitations of automated annotation vs. manual labeling.