

# PyCinemetricsV2: Computational Cinema Studing Tool

[115305288@qq.com](mailto:115305288@qq.com) <https://movie.yingshinet.com>

## Install

1. <https://github.com/CBD-Lab/pyCinemetricsV2>
2. After cloning, the only missing folder is `models`. After downloading the `models` folder from the cloud drive, just place it in the directory, and it will run.
3. Files shared through the cloud drive:

Model files and test\_videos:

[Baidu Netdisk]

(<https://pan.baidu.com/s/1GMIOYvglmvSolclowuM0A?pwd=1234>)

[Google Drive]

(<https://drive.google.com/drive/folders/1ho48Bx6KF-fZewnwBpoHmdl5F1XCY6Xm?usp=sharing>)

4. Model and functionality correspondences:

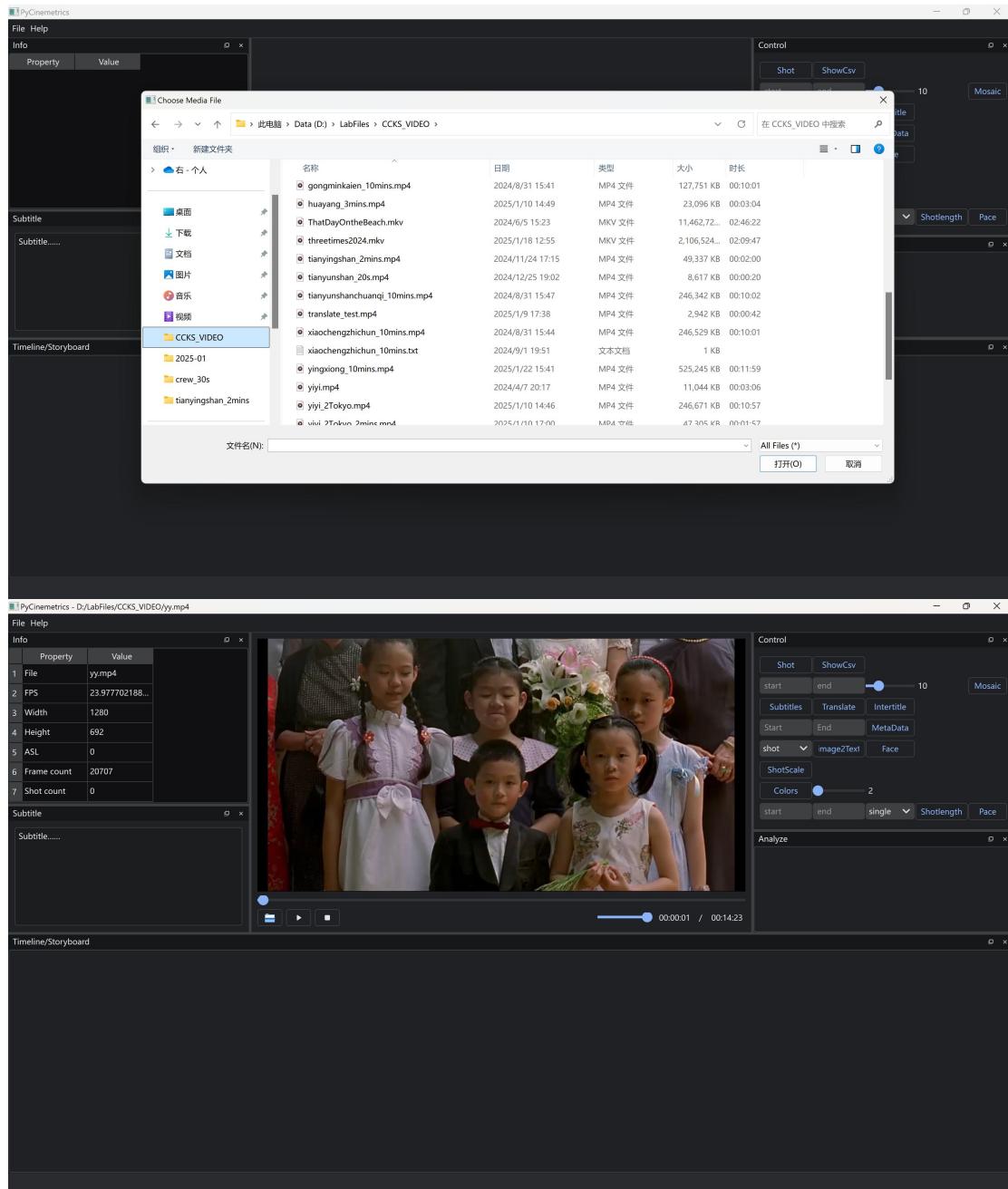
- Video boundary detection: transnetv2
- Face recognition: buffalo\_l
- Speech-to-text subtitles: faster-whisper-base
- Subtitle detection: paddleocr
- Object detection: git-base
- Translation: opus-mt-en-zh
- Shot type recognition: pose net

5. Potential issues:

- If `pip install insightface` fails, please install it manually. [Link to GitHub] (<https://github.com/Gourieff/Assets/tree/main/Insightface>)
- Some models are outdated, so the latest numpy cannot be used. Version 1.26.0 is compatible.

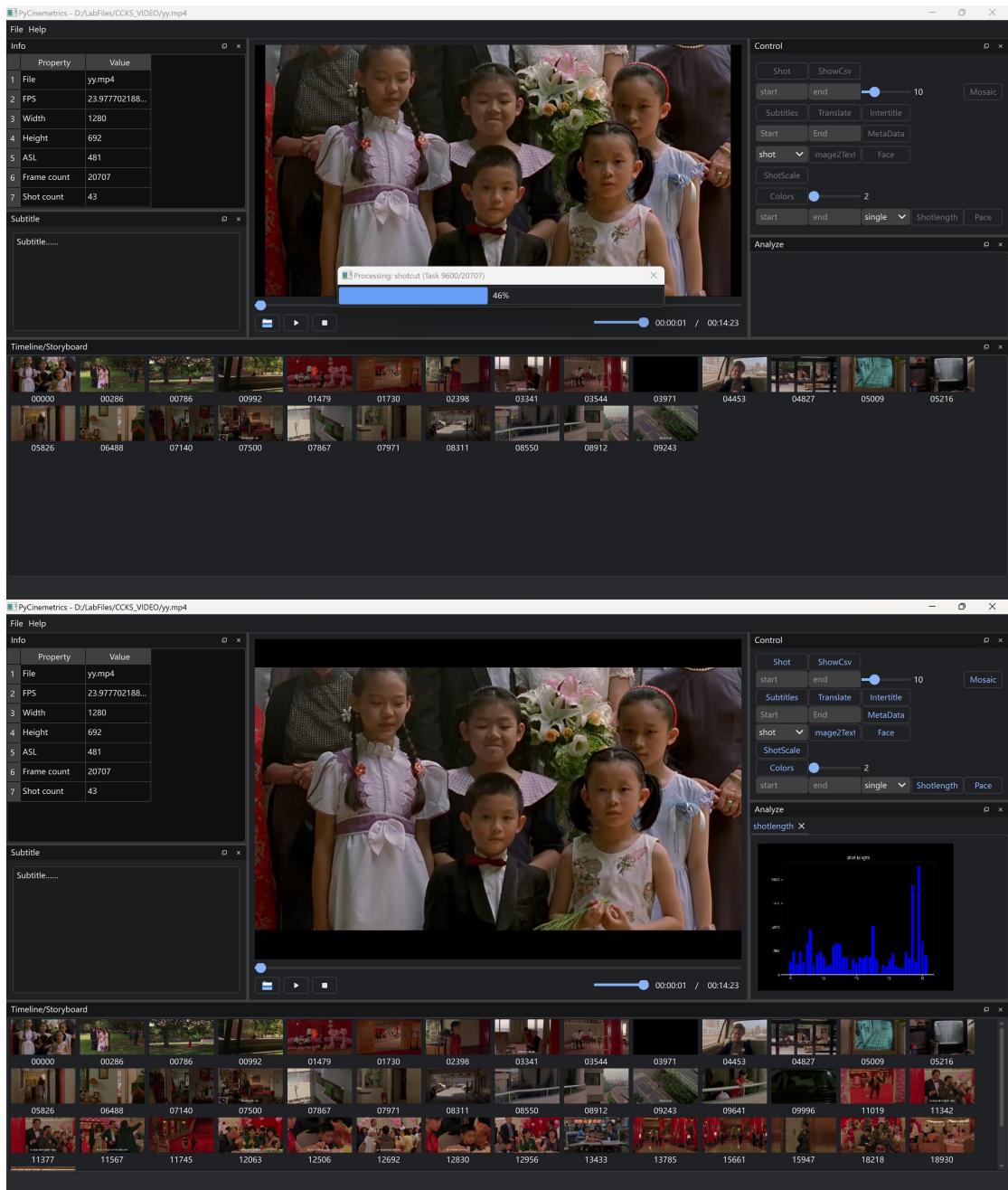
## 1- Open video file

Select a video file from the local disk, and the video will be played. Basic information will be listed in the Info area.



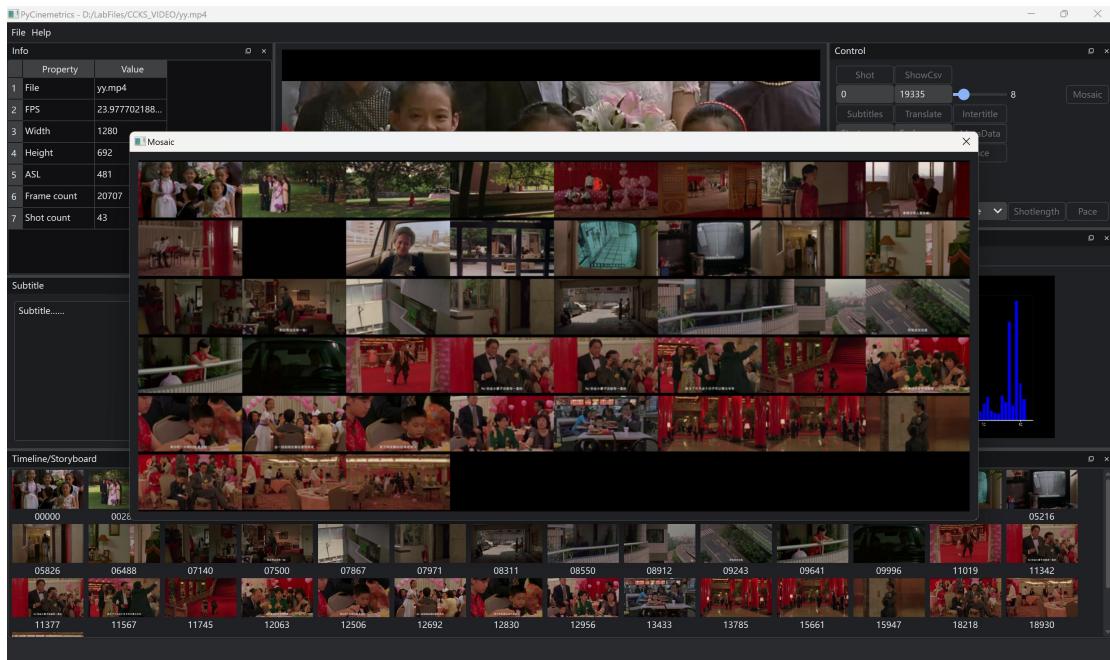
## 2- Shot Boundary Detection (Transnetv2)

In the **control** panel, click the **[shot]** button to trigger the shot boundary detection. First, a folder with the same name as the video file will be created in the '**'img'**' folder to store the output files. Then, the TransnetV2 model will be called to perform shot boundary detection on the video. The first frame of each shot will be displayed in the **Timeline** area and saved in the '**'frame'**' folder inside the output folder. Additionally, a shot length graph will be output in the **Analyze** area. **(Please make sure to execute the shot first, as many subsequent modules depend on the shot frames generated by this function.)**



### 3- Mosaic

Enter the start frame and end frame in the '**start**' and '**end**' boxes to adjust the number of images to be stitched per row. Click the [**Mosaic**] button, and it will automatically search for the boundary frames in the [**start**, **end**] range, then stitch them into a single image and display it.



## 4- Subtitle Detection (Faster-whisper)

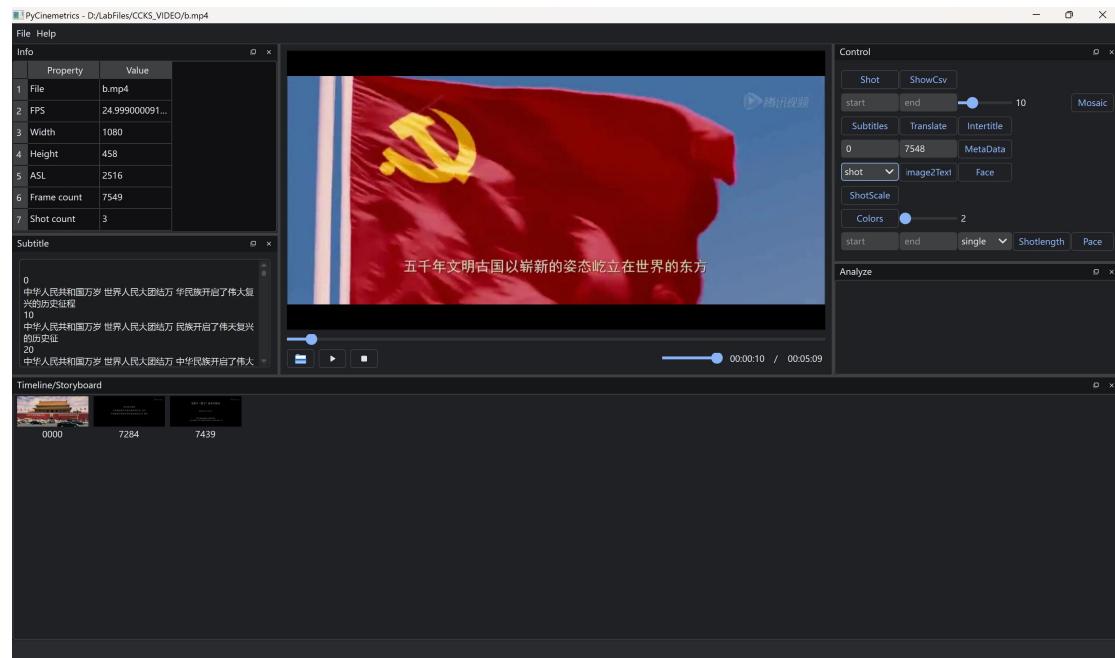
In the **control** panel, click the [**Subtitles**] button to run the subtitle recognition. When this function is executed, it will first extract the audio file, then divide the audio file into segments of 10 minutes each for recognition. Once the recognition is complete, the subtitles will be displayed in the **Subtitle** panel, and **CSV** and **SRT** files will also be generated. The three columns in the CSV file represent the start time, end time, and content of each subtitle segment. Next to the [**Subtitles**] button, there is also a [**Translate**] button that provides English to Chinese translation.



## 5- Metadata (PaddleOcr)

Enter the **start and end frame**, with default values of 0 and 9000 frames (if the total number of frames in the video is less than 9000, the default will be set to the total frames). Click the [**Metadata**] button, and the PaddleOCR algorithm will be invoked to extract subtitles from each shot frame (with a interval of 10 frames). The metadata will be stored in a **CSV** file, **PNG** files, and

## SRT files.



## 6- Translate (Opus-mt-en-zh)

In the **control** panel, click the [Translate] button to use the opus-mt-en-zh model for Chinese-English translation. The translation results will be stored in a **CSV** file and **SRT** file, and the translated subtitles will be displayed in the **Subtitle** area.





## 7- Intertitle (PaddleOcr)

In the **control** panel, click the [**Intertitle**] button to invoke the PaddleOCR algorithm to extract intertitles from each shot frame. The intertitle data will be stored in a **CSV** file, **PNG** files, and **SRT** files.



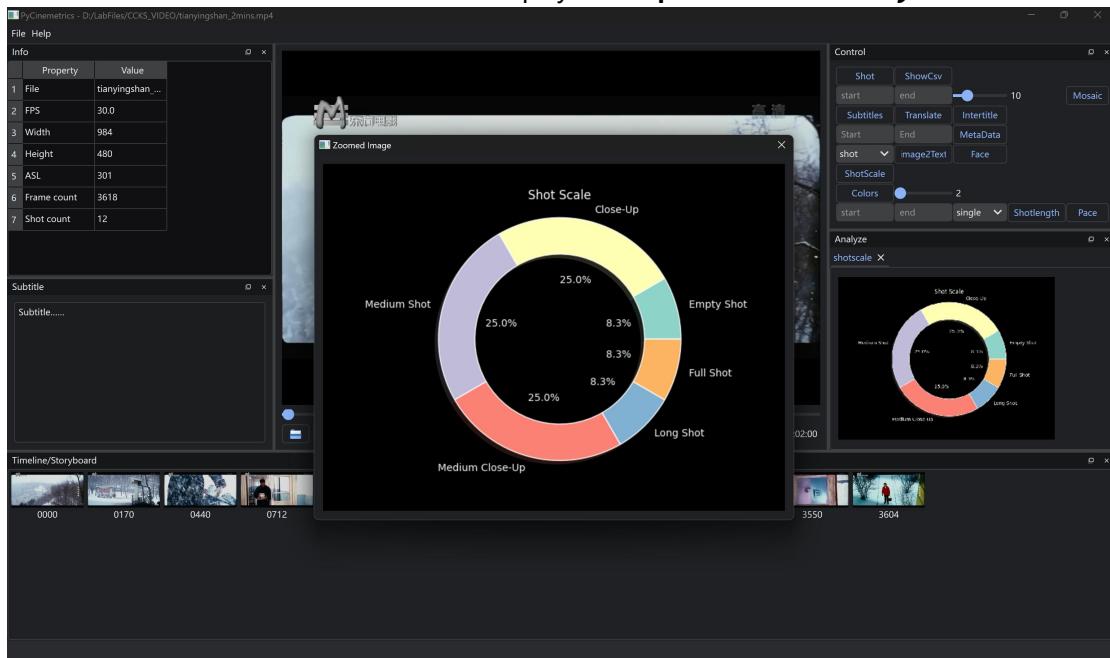
## 8- Image2Text (git-base)

In the **control** panel, click the [**Image2Text**] button to use the Git-base model for image description. The images for description can be selected from the dropdown menu above. If 'shot' is selected, it will detect the previously generated shot frames. If 'keyFrame' is selected, it will extract the video keyframes and perform the detection. The results will then generate a **word cloud** and be displayed in the **Analyze** area.



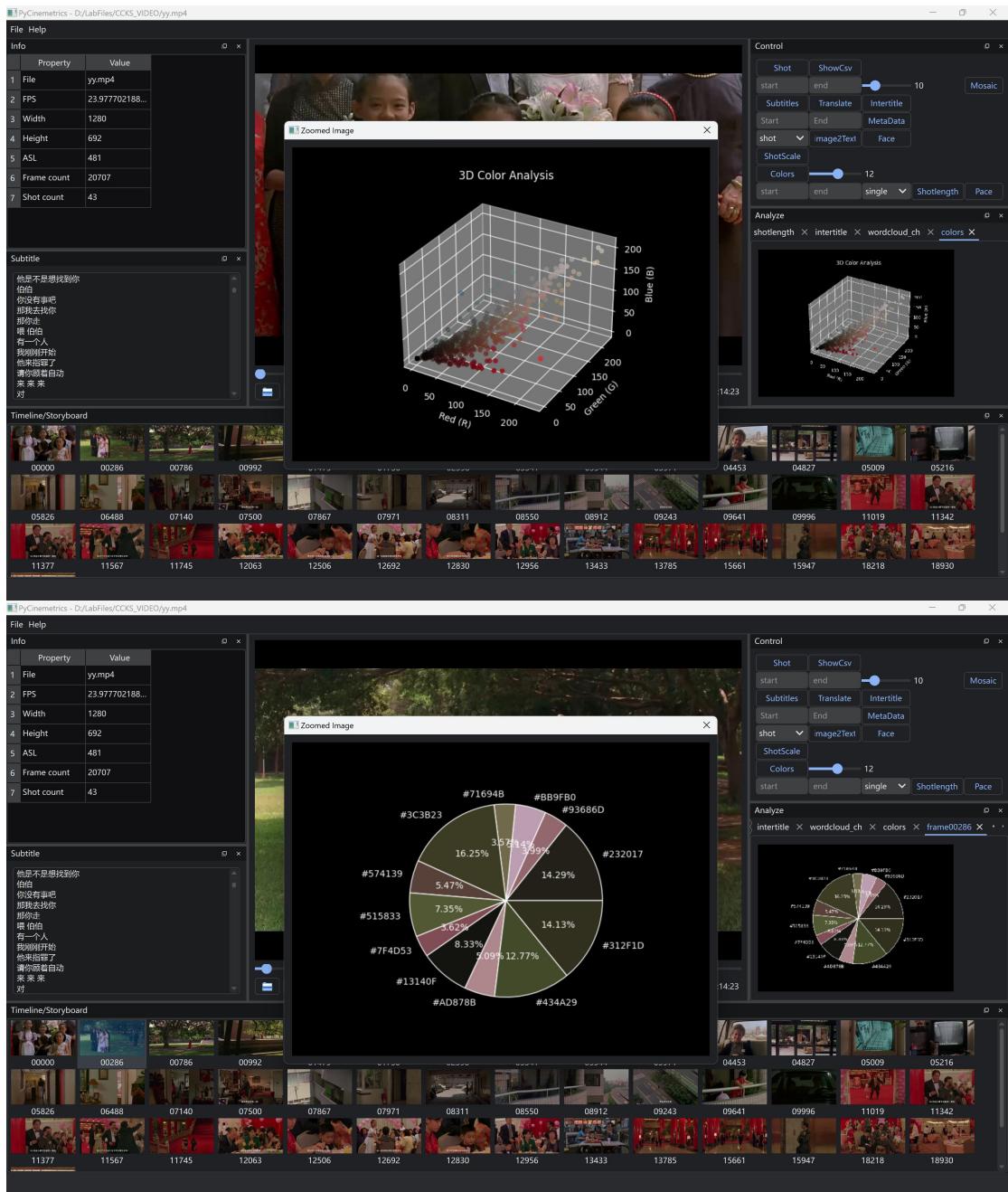
## 9- ShotScale (Pose net)

In the **control** panel, click the [**ShotScale**] button to use the OpenPose model to detect the human skeleton map in the shot frames. Based on the skeleton map, the scene type will be classified. The classification results will be displayed as a pie chart in the **Analyze** area.



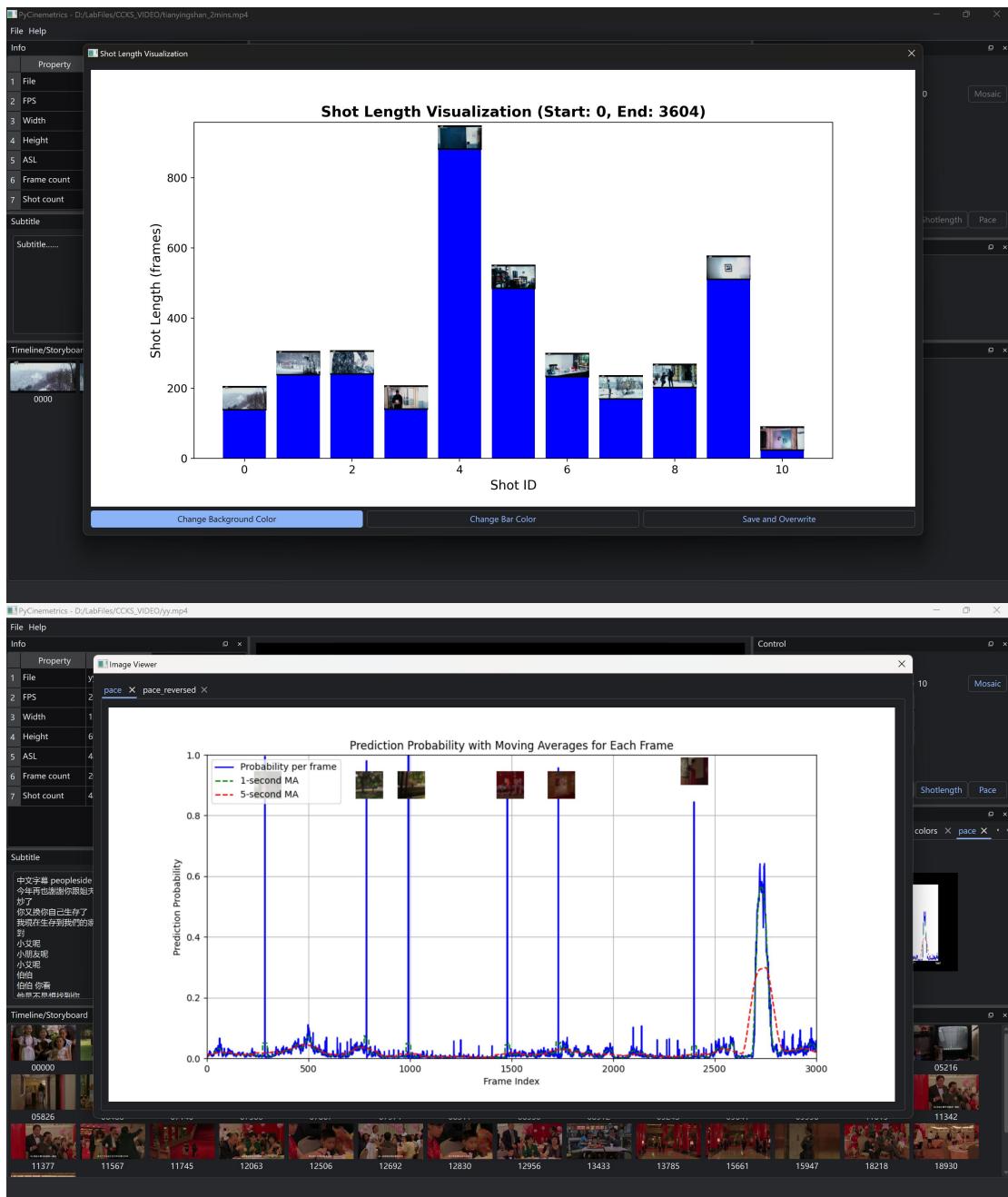
## 10-Colors

In the **control** panel, click the [**Colors**] button to invoke the K-means clustering algorithm to extract colors from each shot frame (defaulting to 2 colors). The number of K-means clusters can be adjusted next to the button. You can also **double-click** on a shot frame in the Timeline area to get the color wheel chart for that frame.



## 11-Shotlength image modification and Pace

Enter the start frame and end frame to define the analysis interval. Click the [**ShotLength**] button to generate and modify the shot length image, including changes to the bar color and background color. Click the [**Pace**] button to generate and visualize the frame-to-frame similarity.



## 12-Face (Buffalo-L)

In the **control** panel, click the "**Face**" button to open the facial recognition panel. The panel is divided into three sections:

**Top Bar:** Click on "**Face Recognition**" to start the function. The feature has two adjustable parameters: first, you can choose whether to use storyboard frames or keyframes for recognition. The second parameter is the **recognition result granularity**, which determines how strictly facial categories are retained. The higher the value, the fewer face categories are recognized. Increasing this value helps to reduce the recognition of background actors' faces.

**Middle Bar:** This section displays the saved facial images. All images are cropped into circular shapes. Clicking on an image will enlarge it for a closer view. To **change the name** of a face, simply edit the text below the image. The "**Delete**" button allows you to remove the selected face image.

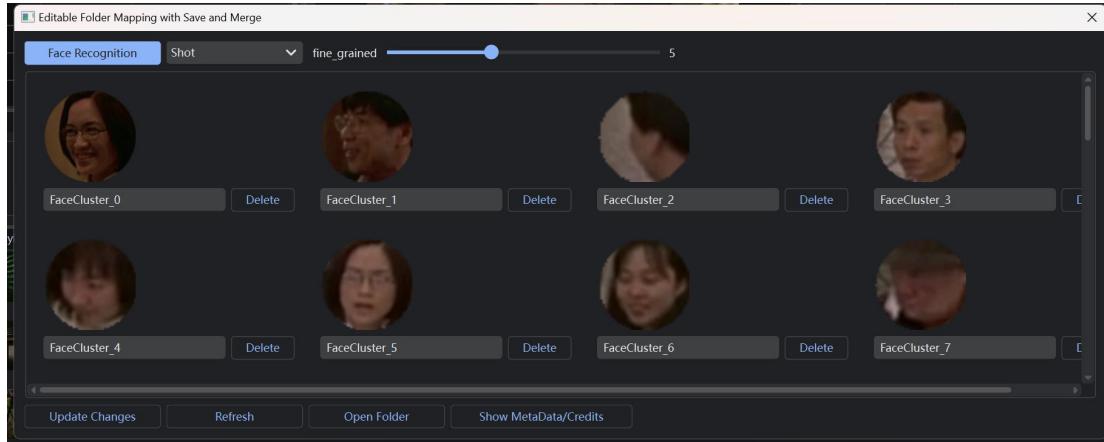
**Bottom Bar:** This section contains four function buttons:

**Save Changes:** After modifying a face's name, click this button to save all changes.

**Refresh:** Reload the images from the folder, undoing any unsaved modifications.

**Open Folder:** Open the facial images folder.

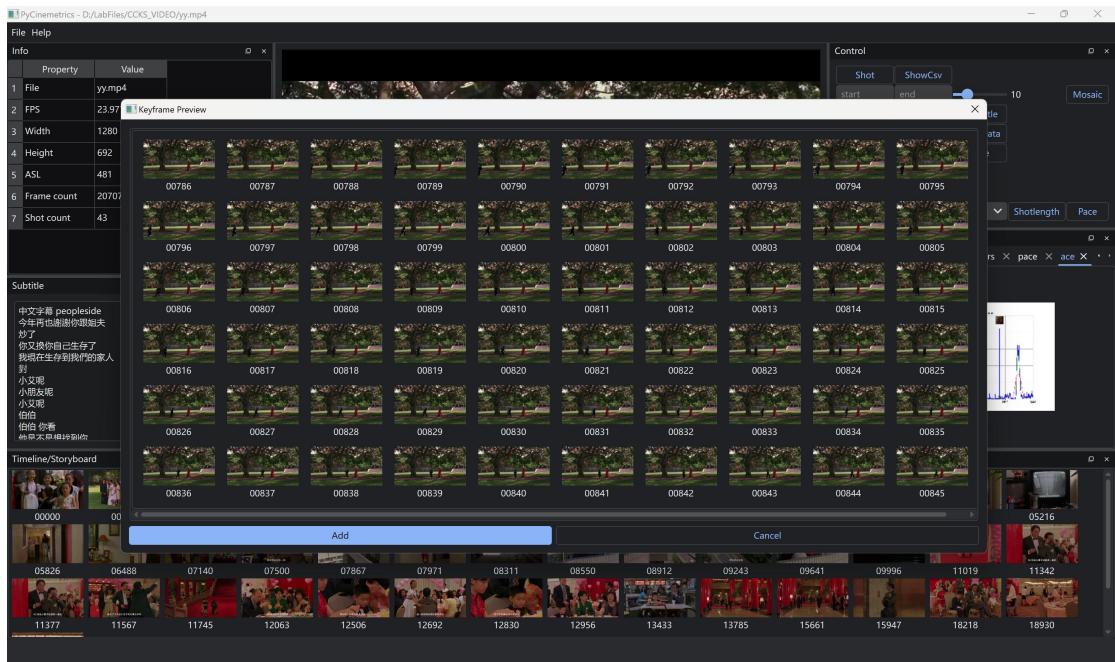
**Show Cast List:** Click this button will display the cast in the **subtitle** panel if have.



## 13-Adding and removing shot frames in the Timeline

The shot frames generated by the model may not always be perfect, so you can **add or remove** shot frames that have already been generated. **Right-click** on a video frame to choose options like deleting a frame, adding a **shot frame**, or adding a **keyframe**. When adding frames, a selection box will pop up to choose the frames to be added. The difference is whether all the frames within the shot are displayed or only the keyframes within the shot. You can select multiple frames at once.





## 14-Viewing CSV files and output files

In the **control** panel, click the [**ShowCsv**] button to view the generated **CSV** files. All output files are saved in the '**img/[video name]**' folder."

