This folder contains codes used for data pre-processing and splitting before training.

data_processing.py is the main file for data pre-processing.

Python 3.8.5 is used. Packages required for data_processing.py are:

- PyTorch 1.8.0
- pandas 1.1.0
- matplotlib 3.3.1
- numpy 1.19.2
- opencv 4.0.1

To run the code, simply execute data_processing.py through "python data_processing.py" in command line or with other Python IDEs. After execution, new folders containing per-processed and split datasets ready for training will be created in the execution directory. The raw dataset is partitioned into training (70%), validation (10%) and testing (20%) subsets. Seeding (value 0) is used for reproducibility.