

Assignment for EE6222 AY 2022/2023

Action Recognition in the Dark

Video data is one of the most common forms of data widely used in all aspects of our daily life. With the rapid growth of video data (500 hours of videos uploaded to YouTube daily alone), automatic video analysis has become a crucial task in dealing with these vast number of videos. Among various video analysis tasks, **human action recognition (HAR)** is one of the cornerstones, which aims to recognize (classify) a human action automatically. The emergence of various large-scale video datasets, along with the continuous development of deep neural networks have vastly promoted the development of HAR, with increasing application in diverse fields, e.g., security surveillance, autonomous driving, and smart home.

Despite the rapid progress made by current HAR research, most research aims to improve the performance on existing HAR datasets constrained by several factors, one of which concerns the fact that videos in existing datasets are shot under a non-challenging environment, with adequate illumination and contrast. This leads to the observable fragility of the proposed methods, which are not capable to generalize well to adverse environments, including dark environments with low illumination. Take security surveillance as an example: automated HAR models could play a vital role in anomaly detection. However, anomaly actions are more common at nighttime and in dark environments, yet current HAR models are obscured by darkness, and are unable to recognize actions effectively. It is therefore highly desirable to explore methods that are **robust** and could cope with dark environments.

You are assigned a project for **Continuous Assessment (CA)**. The project requires you to propose a method to **perform HAR in videos shot in the dark**. There are many variants of HAR based on the length of video, the number of human actions per video, or even the different modality provided (skeleton or RGB or infrared). For simplicity and clarity, in this CA, your task will be restricted to HAR in a trimmed video with one action per video given RGB frames (shot with commercial cameras). You will be given a set of training data and a set of validation data at the start of this CA. You should train your approach on the training data and validate it leveraging the validation data (or the training data themselves with methods such as k-fold validation). The result obtained during the validation process is deemed the **validation result (VR)**. A set of testing data will be provided after the class on Week 12, and you may examine your method against the testing data by an inference process (no extra training). The result obtained during the testing/inference process is deemed the **testing result (TR)**.

You are required to submit a full report of strictly no less than 3 pages (excluding References) and no more than 9 pages (including References). The report should include the following:

- a) Abstract of the report.
- b) Introduction section which describes the motivation of your proposed method.
- c) Related Work section which briefly reviews related works on HAR/HAR in the dark.
- d) Methodology section which presents your proposed method.
- e) Experiment section which demonstrates the full empirical process.
- f) Conclusion of the report.

Here are some additional directions that you should follow and notes about this CA:

1. The report must be submitted in PDF format, following the provided IEEE Conference template (in LaTeX or Word format). You may use the IEEE Conference Template available on Overleaf (for free). [Link here](#).
2. The report must be done individually. You may discuss with your peers, but NO plagiarism is allowed. The University, School, and the teaching team take plagiarism very seriously. You must submit an originality report, obtained from iThenticate (available on NTULearn). A zero mark will be given to anyone found plagiarizing and a formal report will be handed to the School for further investigation.
3. There will be 10 action classes (drink, jump, pick, pour, push, run, sit, stand, turn, walk) in the provided data. The training, validation, and testing set will all be provided along with files that state the ID of the video, the class ID of the video and the video file name.
4. In the Experiment section of the report, you are required to justify how you construct your method by the validation result (VR). You MUST also include your testing result (TR) to illustrate the actual performance of your method. If there is a gap between your VR and TR, you should discuss the reason behind this gap.
5. To obtain both the VR and TR, you must obtain an output from your method following the output sample file provided to you. A scoring program is also provided to mark the accuracy of your output.
6. You may use any method you deemed necessary for this CA. There are no limitations as to whether deep learning method should be used or what framework/library should be used. However, it should include your original ideas and the report should demonstrate your understanding of your method.
7. It is highly recommended to program your method with **Python** or **MATLAB**. However, other programming languages are also acceptable. You are required to submit a zip file with your code ONLY (no training samples/processing results included).
8. Your score will be based on the **clarity** of the report, the **originality** of your proposed method, and the **reproducibility** of your provided code. Your code may be checked for reproducibility on the testing data to justify you TR. *Note: a high TR does NOT equal to a high CA mark.*
9. The **deadline** for this CA is before class on Week 13, 9 November 2022, at 1829 GMT+8.
10. Submission steps and format:
 - a) Submit your report to the Turnitin portal to obtain your Turnitin report.
 - b) Zip your report, Turnitin report and your code into a **single zip file**, rename the zip file strictly as: **“YourFullName_YourMatriculationNo..zip”** (if your full name as in your matriculation card is Xu Yuecong and your matriculation number is U2000000K, then your submission zip file is named as “Xu Yuecong_U2000000K.zip”)
11. Penalty marks will be applied for late submission or if you do not follow the above submission guideline.

Tips:

- a. For this CA, you are highly encouraged to train and test the method from scratch so that you get to experience how learning methods are designed and examined.
- b. However, if you would like to use deep learning methods BUT do not have the equipment for training (e.g., a computer with GPU), you may consider leveraging the deep learning modules as a tool for feature extraction only. To achieve this without GPU, I will provide you with some commonly used deep video feature extractors (essentially deep video networks) that can be loaded with CPUs. These extractors will be in the form of PyTorch PTH files. An example file showcasing how to use these PTH files are provided for your reference.