Practical Steps:

Collect Data:

- Gather videos based on the task and model requirements.

Preprocess Data:

- Extract frames from videos.
- Write descriptions that capture the sequence of actions. Including key objects, and scene details for each frame.

Train Model:

- Use the collected data to train the chosen model.
- A model will be trained to generate a description for a video taking into account the contextual data presented in the form of frame descriptions.

Evaluate and Generate Captions:

- Test the model on unseen videos.
- Evaluate using metrics like BLEU, METEOR, ROUGE, and CIDEr.
- For the purposes of contrast analysis, the generated description can be compared with the description that was generated without taking into account contextual data(shows the importance of having well annotated dataset)
- Categorization of a video as “dangerous” or “not dangerous” can be performed after getting the final description.

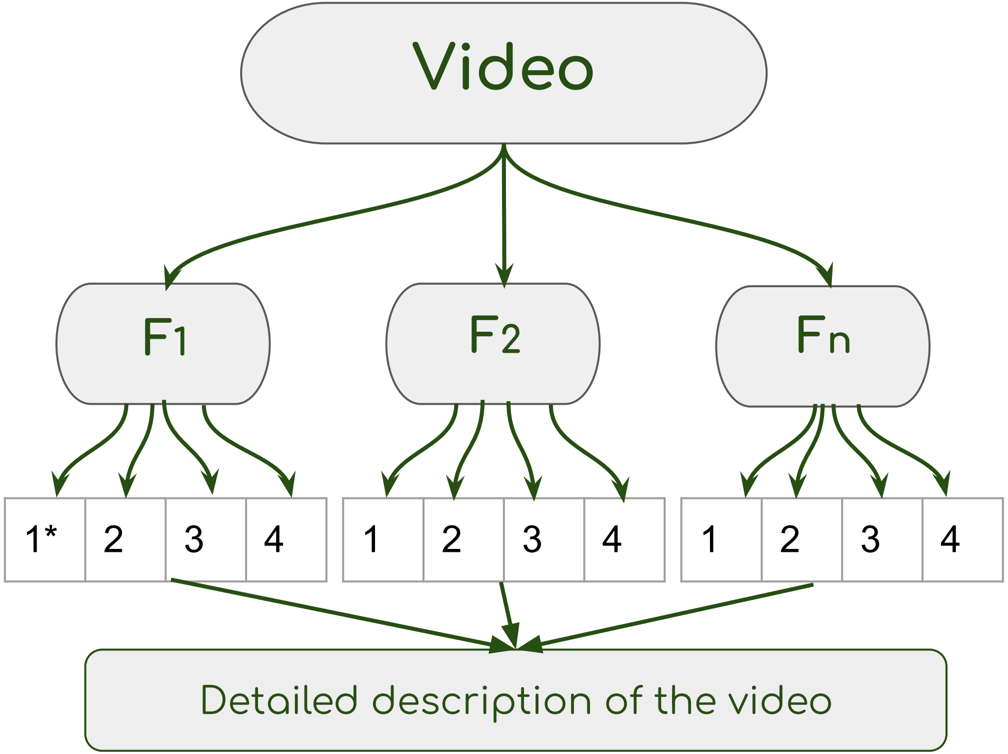
Dataset before preprocessing:

- Videos of normal situations(scenarios)

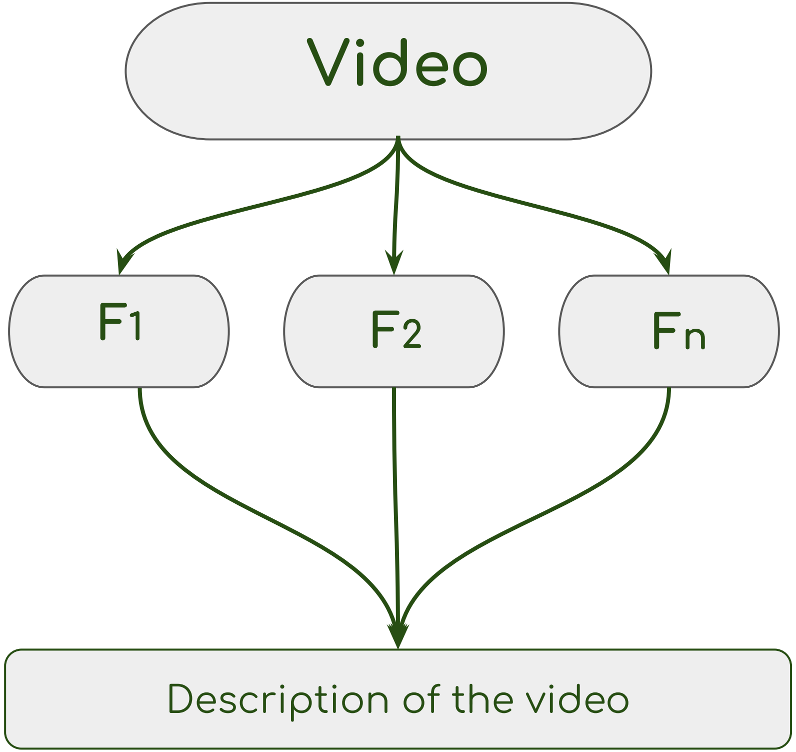
- Videos of abnormal situations(scenarios)

https://paperswithcode.com/dataset/ubnormal

Generation of videos description



generated with *manually predefined frame descriptions



generated without manually predefined frame descriptions