Our dataset consists of indoor home scenarios:
The total duration of the dataset is 349.25 minutes and the average length is 100 videos within 20~30 sec (25 from Wyze cam + 75 from Internet)

Our proposed solution is an end-to-end machine learning model with an Encoder-Decoder type architecture.
The encoder is a 3D CNN which extracts multi-scale frame features from the pre-trained feature extractor.
The decoder has three parallel heads to generate captions, predict temporal boundaries for events in a video and count the number of events in the input.
The generated caption is refined using the YOLO Object Detection Model for obtaining the correct object class in the video.

The decoder has three parallel heads to generate captions, predict temporal boundaries for events in a video and count the number of events in the input.

A dataset designed to evaluate the performance of the video caption generated on about 50 indoor videos and 10 outdoor videos using the metrics mentioned below.

**EVALUATION DATASET:**
- A dataset designed to evaluate the performance of the video caption generated on about 50 indoor videos and 10 outdoor videos using the metrics mentioned below.

**COMPARE THE MODEL:**
- Pre-Trained
- Fine-Tuned

**METRIC NAME** | **METRIC NAME** | **METRIC NAME** | **METRIC NAME** | **METRIC NAME**
--- | --- | --- | --- | ---
BLEU_1 | BLEU_2 | BLEU_3 | BLEU_4 | METEOR
--- | --- | --- | --- | ---
0.538 | 0.557 | 0.222 | 0.132 | 0.244 | 0.484 | 0.445
0.776 | 0.698 | 0.626 | 0.570 | 0.466 | 0.756 | 2.521
0.887 | 0.875 | 0.850 | 0.825 | 0.607 | 0.920 | 2.956

**QUALITATIVE RESULTS:**
- **INDOOR**
- **OUTDOOR**

**SOFTWARE SYSTEM:**
- Using Streamlit with Ngrok and Colab to build the website.
- Features:
  - Gallery: User can test some sample videos.
  - Upload indoor/outdoor video: User can upload indoor/outdoor videos and get the quantitative results through their mobile and computer.

**REFERENCES:**

**FUTURE WORK:**
- Enhancements of the PDVC Model Architecture.
- Creating an enhanced UI capable of user authentication and authorization.
- Integration of the system with the Wyze Hardware for seamless integration as a single system.
- Generating better captions for anomaly, vehicle based and outdoor scenarios.