# Datasets Report

## 1. TVSum Dataset

### Overview

* **Purpose**: Designed to aid the development of algorithms for video summarization, focusing on generating diverse and relevant summaries.
* **Creators**: Introduced by Yale Song, Jordi Vallmitjana, Amanda Stent, and Alejandro Jaimes in the paper "TVSum: Summarizing web videos using titles".
* **Applications**: Useful for benchmarking machine learning models focused on web video summarization.

### Details

* Number of Videos: Contains 50 user-generated videos collected from YouTube.
* Categories: Organized into 10 distinct categories:  
   1. Changing a vehicle tire  
   2. Getting a vehicle unstuck  
   3. Grooming an animal  
   4. Making sandwiches  
   5. Parade  
   6. Flash mob gathering  
   7. Dog show  
   8. Attending a wedding ceremony  
   9. Attempting a bike trick  
   10. Packing for a trip
* Video Lengths: Each video varies between 1 to 6 minutes, with a total runtime of approximately 220 minutes..

### References

* <https://github.com/yalesong/tvsum>
* <https://paperswithcode.com/dataset/tvsum-1#:~:text=Videos%20Using%20Titles-,Title%2Dbased%20Video%20Summarization%20(TVSum)%20dataset%20serves%20as%20a,crowdsourcing%20(20%20per%20video)>

## 2. SumMe Dataset

### Overview

* **Purpose**: Designed to evaluate video summarization techniques by providing user-generated videos with multiple human-created summaries.
* **Creators**: Introduced by D. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool in the paper "Category-specific video summarization".

### Details

* Number of Videos: Contains 25 user-generated videos covering a diverse range of real-world activities.
* Categories and Content: Covers a wide variety of activities, including:  
   - Sports activities  
   - Travel vlogs  
   - Events (e.g., parades, weddings)  
   - Daily activities (e.g., cooking, hiking)  
   - Outdoors and nature scenes  
   - Personal experiences at tourist location
* Video Lengths: Videos typically range from 1 to 6 minutes, with a total runtime of approximately 66 minutes.
* Annotations:  
   - Human-created Summaries: Annotators manually create summaries, selecting key moments and reducing the video length to 15-20% of its original duration.  
   - Variability: Individual preferences lead to high variability in the summaries.  
   - Ground Truth Frames: Provides frame-level ground truth for generating summaries.

### References

* <https://paperswithcode.com/dataset/summe#:~:text=The%20SumMe%20dataset%20is%20a,summaries%20(390%20in%20total)>
* <https://github.com/robi56/video-summarization-resources?tab=readme-ov-file>

## 3. Comparison of TVSum and SumMe

|  |  |  |
| --- | --- | --- |
| Aspect | TVSum | SumMe |
| Purpose | Frame-level importance scores for web videos. | Human-created summaries for diverse activities. |
| Number of Videos | 50 | 25 |
| Categories | 10 predefined categories (e.g., sports, weddings) | Broader range of real-world activities. |
| Video Lengths | 1 to 6 minutes, total runtime ~220 minutes. | 1 to 6 minutes, total runtime ~66 minutes. |
| Annotations | Frame-level scores by 20 annotators. | Summaries created manually by annotators. |
| Summarization Approach | Objective (importance scores). | Subjective (human-created summaries). |