

Sequential Storytelling Image Dataset (SSID)

Research Paper Title

Sequential Vision to Language as Story: A Storytelling Dataset and Benchmarking

Dataset Distribution:

Partition	distribution	# Image	# Album	# Sentence	# Story
Training Images	90%	15,625	3,125	62,500	12,500
Validation Images	5%	870	174	3,480	696
Testing Images	5%	870	174	3,480	696
Total	100%	17,365	3,473	69,460	13,892

Dataset Organization:

The images obtained from videos on YouTube were saved in a folder with a unique ID for each image. For instance, images 1 through 5 provide the collection of the first set of images, while images 6 through 10 present the following set of images in a similar fashion. A JSON file organized all the images while including the ground truth corresponding to each image. The structure of the JSON file is as follows:

```
{
  "annotations":
  [[{"storylet_id": Int,
    "storytext": str,
    "youtube_image_id": int,
    "album_id": int,
    "story_id": int,
    "image_order": int }]]
}
```

where *storylet_id* is the unique story identifier, and *storytext* is a string value consisting of the story's first sentence. *youtube_image_id*, *album_id*, and *story_id* are integer values containing the ID for each complete story. Finally, *image_order* represents the image's position within the set of images.