

StoryBench Didemo Annotations

Instructions

bugliarello@ pikinder@ voigtlaender@

Goal: A Dataset with step-by-step stories segmented over time

We have a dataset of videos with short descriptions of actions (sometimes repeated) happening in the video

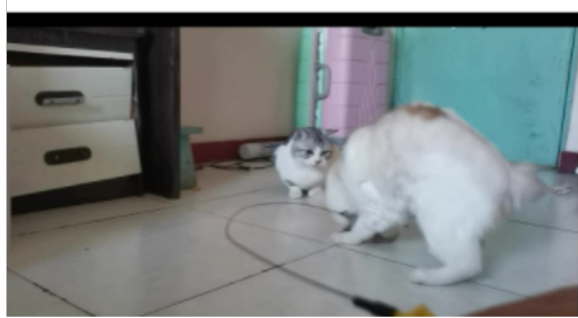
We want to

1. Add a description of the background
2. Specify the number of **important** actors in the video
3. Refine the sentences to create a coherent story
4. Add time stamps for when (the action of) each sentence starts and ends

Note: this task is a variant of the StoryBench task

Your Tasks (1): Add a description of the background

- First, watch the video and understand what is happening.
If necessary, go back and watch it again.
- Then, add a description of the background of the video. For example:



“In the background there is a floor of white tiles, a purple suitcase, and a broken white cupboard.”

Your Tasks (2): Specify the number of *important* actors

- We want to know how many actors appear throughout the video
- By **actor**, we mean entities (e.g. people, animals) who perform actions that are **central to the story** of the video
- Entities that do not play an important role in the video (e.g. a person in the background) are considered un-important.
- That is, can would be able to describe the video without referring to them
- This un-important entities do **not** count towards “number of actors”

Your Tasks: (3) Refine the sentences to create a story of the video

You will get reference descriptions for parts of the video

- These descriptions were collected to identify specific parts of the video
- They may contain
 - **Camera movements** (e.g. “the camera pans to a group of friends”, “the boy runs towards the camera”)
 - **Temporal indicators** (e.g. “first” and “after”)
 - **Spatial indicators** (e.g. “left” and “bottom”)
- Use the references as inspiration, but do not stick exactly to them.
 - You can annotate a different number of sentences, you can merge references together to a sentence, you can write completely new sentences, etc.

You will then need to

- **Describe the actors and their actions** throughout the video with multiple sentences
 - Sentences describe actions that happen one after the other
 - Each sentence should capture a single action
 - If two actions happen in a very short time, put them together into a single short sentence
 - e.g.: If in the video, a kid jumps and immediately after does a flip, the two actions should be one “The kid jumps and flips”
- When you narrate the story, provide important **attributes** to describe them (e.g. “a kid wearing a red shirt”)
 - Do not repeat such information in following sentences when not needed
 - e.g. just use “the kid” if there is only one actor in the video, only one kid in the video, etc.
- Use the original descriptions as references for your step-by-step descriptions
 - You do not need to re-phrase all of them 1:1. They are examples of how parts of the videos have been described before
- Make sure to describe important characteristics, such as **Camera movements, Temporal/Spatial indicators**

Your Tasks: (4) Add action timestamps

- You will then need to record when your sentences start and end
 - You should simply pause the video at start/end time, and click the corresponding “Assign” button to automatically copy the timestamp from the video to the box:

Sentence part	Start time	End time
A boy wearing a yellow blue t - shirt is standing and holding yellow baseball bat.		
	Assign start	Assign end

NB: The end time of a given sentence must be **earlier or equal** to the start time of the next one

Examples



Describe the background:

Number of actors:

Sentence part reference	Sentence part	Start time	End time
kitty using paws to get food out a can	A grey-brown kitty is eating some food out of a can.	0	10.426297
		Assign start	Assign end
can uses paw to move can	It then puts its paw in the can and moves it to the right.	10.426297	17.127522
		Assign start	Assign end
cat puts paw in can for first time.	The cat continues eating from the can.	17.127522	24.417622
		Assign start	Assign end
the cat sticks it's paw in the can.	The cameraman's hand is seen petting the cat's head as it keeps eating from the can.	24.417622	34.59771
		Assign start	Assign end

Examples



Describe the background:

Number of actors:

Sentence part reference	Sentence part	Start time	End time
green jacket comes into view.	A man wearing a green jacket and yellow pants comes into view as a girl wearing a pink sweater and a furry purple character move to the left carrying a sign.	0	3.548506
		Assign start	Assign end
orange furry character walks in front of camera	The man walks in front of camera towards the right dancing, followed by an orange furry character that wears a red jacket and a white mini gown.	3.548506	10.673068
		Assign start	Assign end
a little girl in pink is lifted for a better view.	A little girl in pink is lifted for a better view by a woman on the right, while a woman with fairy wings dances on the street while holding a sign.	10.450475	13.246066
		Assign start	Assign end
woman in the foreground picks up a child in a pink coat.	The camera moves to the left and shows several furry characters dancing.	13.246066	18.530159
		Assign start	Assign end
blond hair woman picks up girl	The camera rotates to the right and shows the orange character and the man in green jacket dancing.	18.530159	23.287233
		Assign start	Assign end
	To the right of the cameraman, a boy wearing a red hat is shown held by a woman in a blue jacket.	23.287233	29.396633
		Assign start	Assign end

Examples

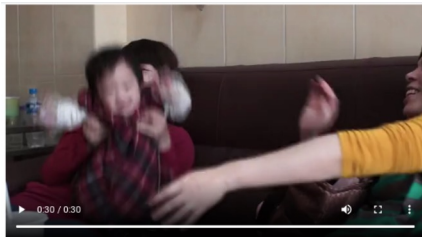


Describe the background:

Number of actors:

Sentence part reference	Sentence part	Start time	End time
yellow sports car pulling up in a driveway.	A yellow sports car pulling up in a driveway from the right.	3.223208	6.171222
		Assign start	Assign end
a yellow car pulls up and parks.	The car comes to a stop.	6.171222	12.662778
		Assign start	Assign end
yellow car comes to a stop.	A lady wearing a white dress walks from the stand towards the car holding an object.	12.662778	16.661221
		Assign start	Assign end
lady goes in	The lady gives the object to the man through the right window.	16.661221	18.929889
		Assign start	Assign end
an exotic yellow car drives off	The lady walks back to the stand.	18.929889	23.340516
		Assign start	Assign end

Examples



Describe the background:

Number of actors:

Sentence part reference

baby is clapping

first time child claps

the young baby claps her hands.

the baby is clapping.

a woman wearing a green and yellow shirt shows her face for the first time

the face of a woman in orange and green is visible for the first time

Sentence part

A woman wearing a red sweater is sitting on the couch and holding a baby who is clapping.

The baby starts jumping held by the woman.

A woman wearing a green and yellow shirt is shown for the first time to the right and moves her arms along with the baby.

The camera focuses on the baby who keeps jumping while the woman on the right claps her hands.

The camera shows again the woman and the baby playing and laughing.

Start time

0

Assign start

4.024307

Assign start

8.500742

Assign start

18.154062

Assign start

22.450037

Assign start

End time

4.024307

Assign end

7.626596

Assign end

18.154062

Assign end

22.450037

Assign end

30.116281

Assign end