

StoryBench VidLN Annotations

Instructions

bugliarello@ pikinder@ voigtlaender@

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

- We have a dataset of videos with long descriptions of the story of the video
 - Descriptions are “per-actor,” describing the story from the viewpoint of an entity (e.g. “kid”)
- The description for each video is one long sentence
 - e.g. “A kid wearing grey pants is running on a white floor while putting a black-white-brown printed cloth on the head then collides with toy vehicles and falls on the floor.”

We want to

1. Split the long sentence into shorter sentences, each describing actions that happen one after the other
2. Add time stamps for when (the action of) each sentence starts and ends

Your Tasks: (1) Fix short sentences

- You will get
 - The original long sentence, and
 - An *automatic split* of the long sentence into shorter sentences
- You will then need to
 - Make sure that each short sentence is faithful to the original long sentence.
 - We want to keep the **same words** (in principle, all words) as the original long one, unless it is not grammatically correct (e.g. due to the splitting)
 - e.g.: it is OK to keep the verb “flies” instead of the original “flying” in the short sentence
 - When part of a caption is ambiguous or has major issues, you can refine it by adding or editing words
 - Fix errors in the short sentences compared to the original long sentence
 - If the same subject phrase is repeated across sentences “A brown cat is walking” “A brown cat falls down.”
→ Use “**The...**” from the 2nd sentence on: “**The** brown cat falls down.”
 - Add missing words whenever required (we noticed automatic splits drop words sometimes)
 - If two actions happen in a very short time, put them together into a single short sentence
 - e.g.: If in the video, a boy jumps and immediately after does a flip, the two actions should be merged into one (as in the original, long sentence)
 - “The boy jumps.” “He flips.” → “The boy jumps and flips”
 - Fix actor naming when possible
 - Another man / man two → a man (+ descriptive attributes)

Your Tasks: (2) Add action timestamps

- You will also get
 - The original video
- You will then need to
 - Record the time when a short sentence starts and ends
 - 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 action must be earlier or equal to the start time of the next action

Example Interface



Combined sentence: A boy wearing a yellow blue t-shirt is standing and holding yellow baseball bat and then hits the ball and then looks in front and makes faces.

Sentence part reference	Sentence part	Start time	End time
A boy wearing a yellow blue t - shirt is standing and holding yellow baseball bat.	A boy wearing a yellow blue t - shirt is standing and holding yellow baseball bat.		
		Assign start	Assign end
The boy hits the ball.	The boy hits the ball.		
		Assign start	Assign end
The boy looks in front and makes faces.	The boy looks in front and makes faces.		
		Assign start	Assign end
		Assign start	Assign end
		Assign start	Assign end

☐ Multiple stories
☐ Unimportant actor

- There are 4 empty extra rows that can be used if the sentences need to be split further. In most cases, you can ignore those rows.

Example: Fixing of Sentence Split (1)

- Combined sentence:
 - Man one wearing brown trousers is sitting on a black sofa on a brown surface playing with a controller and then moving while raising his hands and legs and looking at man two.
- Automatically split sentences:
 - 1) Man one wearing brown trousers is sitting on a black sofa on a brown surface. Man one is playing with a controller.
 - 2) Man one is moving while raising his hands and legs.
 - 3) Man one is looking at man two.
- Fixed split sentences:
 - 1) Man one wearing brown trousers is sitting on a black sofa on a brown surface. ~~Man one is playing with a controller.~~
 - Keep original words whenever possible
 - 2) Man one is moving while raising his hands and legs.
 - 3) Man one is looking at man two.

Example: Fixing of Sentence Split (2)

- Combined sentence:
 - An ostrich is standing on the right side, looking at the piece of food held by the man, and then suddenly grabs the cup of food and starts eating at once.
- Split sentences:
 - 1) An ostrich is standing on the right side. The ostrich is looking at the piece of food held by the man. The ostrich grabs the cup of food.
 - 2) The ostrich starts eating at once.
- Fixed split sentences:
 - 1) An ostrich is standing on the right side. ~~The ostrich is,~~ looking at the piece of food held by the man. ~~The ostrich grabs the cup of food.~~
 - Keep original words whenever possible
 - 2) The ostrich **suddenly** grabs the cup of food **and** ~~The ostrich~~ starts eating at once.
 - The last sub-sentence in 1) happens with 2) in the video
 - Keep original words whenever possible

Example Answer 1

Combined sentence: A kid wearing grey pants is running on a white floor while putting a black-white-brown printed cloth on the head then collides with toy vehicles and falls on the floor.

Sentence part reference	Sentence part	Start time	End time
A kid wearing grey pants is running on a white floor. The kid puts a black - white - brown printed cloth on the head.	A kid wearing grey pants is running on a white floor. The kid puts a black - white - brown printed cloth on the head.	0	3
The kid collides with toy vehicles and falls on the floor.	The kid collides with toy vehicles and falls on the floor.	3	7

- Sometimes the automatic split is good and doesn't need to be edited

Example Answer 2

Combined sentence: A boy in a grey t-shirt is playing with the football by swirling his legs then gut imbalance and falls down on a stone's surface and got hurt his elbow.

Sentence part reference

A boy in a grey t - shirt is playing with the football by swirling his legs.

The boy in a grey t - shirt got imbalance and falls down on a stone 's surface.

The boy in a grey t - shirt got hurt his elbow.

Sentence part

A boy in a grey t-shirt is playing with the football by swirling his legs.

The boy in a grey t-shirt got imbalanced and falls down on a stone's surface.

The boy in a grey t-shirt hurt his elbow.

- Sometimes the automatic split is good and doesn't need to be edited

Example Answer 3

Combined sentence: A white dog is walking down the stairs while rubbing its lower paws.

Sentence part reference

A white dog is walking down the stairs.

The dog is rubbing its lower paws.

Sentence part

A white dog is walking down the stairs while rubbing its lower paws.

- Both actions happen at the same time, so we merged it back to the same sentence

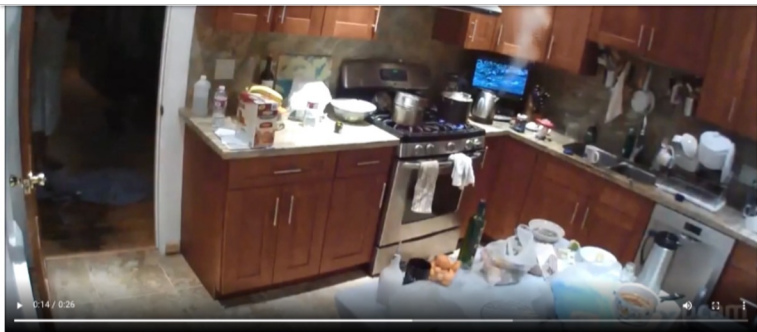
Multiple Stories Checkbox

- Sometimes the video shown on the top of the user interface actually consists of multiple shorter stories (clips) that are arranged together, one after the other.
For example, these screenshots are taken from the same video:



- If this is the case
 - Please make sure that all the start and end times are **within** to the corresponding video clip
 - **Always** fix actor naming: Another man / man two → a man (+ descriptive attributes)
 - ~~Another~~ man → A man wearing blue shorts
 - Click the “multiple stories” checkbox and Add the story number:
 - If the description applies to an actor from the **first story** ⇒ write **1**, and so on.
 - Add **Story start time**: when the video clip corresponding to the described story starts
 - Add **Story end time**: when the video clip corresponding to the described story ends
- NB: This doesn't apply if only the last few moments of the video are from another clip

Example Answer Multiple-Stories (same video, story 1)



Combined sentence: A man wearing a grey robe catches a fire on his left hand, screams and tries to extinguish by patting the robe while walking, moves out of the gate, takes off the robe, throws it on the ground and stomps it with his leg.

Sentence part reference

A man wearing a grey robe catches a fire on his left hand.

The man screams and tries to extinguish the fire by patting the robe.

The man walks out of the gate. The man takes off the robe. The man throws the robe on the ground. The man stomps the robe with his leg.

Sentence part

A man wearing a grey robe catches a fire on his left hand.

The man screams and tries to extinguish the fire by patting the robe while walking.

The man takes off the robe, throws it on the ground and stomps it with his leg.

Start time	End time
0.63534	2.662898
Assign start	Assign end
2.662898	8.555148
Assign start	Assign end
8.555148	14.553431
Assign start	Assign end

Multiple stories ☒

Story number 1

Story start time 0

Story end time 14.553431

Assign start time of story

Assign end time of story

Example Answer Multiple-Stories (same video, story 2)



Combined sentence: Another man wearing a black t-shirt is holding an object and runs, falls on a yellow plastic sheet to slide but fails and stops there on the spot.

Sentence part reference

Another man wearing a black t-shirt is holding an object and runs.

The man falls on a yellow plastic sheet to slide.

The man falls and stops there on the spot.

Sentence part

A man wearing a black t-shirt is holding an object and runs.

The man falls on a yellow plastic sheet to slide but fails and stops there on the spot.

Start time

14.935268

Assign start

End time

15.768668

Assign end

Assign start

Assign end

Assign start

Assign end

Multiple stories ☒

Story number

2

Story start time

14.935268

Assign start time of story

Story end time

22.438031

Assign end time of story

Unimportant actor ☐

NOTE: This second story (#2) starts right after the end of the first story (see slide before) and it ends at around 22s.

NOTE: In the first sentence: Another man → A man
The second and third sentences have been merged.

Unimportant Actor Checkbox

- Sometimes, the long sentence describes an entity which does not play an important role in the video (e.g. a person in the background)
- If this is the case
 - Please check the “unimportant actor” checkbox
 - When the checkbox is checked, you can skip all other annotations for this example (no need to correct the sentences or give the time stamps)
 - If an actor consists of multiple entities (e.g. a group of people) and they perform different actions, please check the “unimportant actor” checkbox
 - Example: “A group of people is standing and a few are riding bikes.”
- Example: “A brown puppy with only his mouth visible is looking at the men.”