Supplementary Material for TGIF: A New Dataset and Benchmark on Animated GIF Description

Yuncheng Li	Yale Song	Liangliang Cao	Joel Tetreault
University of Rochester	Yahoo Research	Yahoo Research	Yahoo Research
Larry Goldberg	Alejandro Jaimes	Jiebo Luo	
Yahoo Research	AiCure	University of Rochester	

This supplementary material contains detailed information about the paper, specifically:

- Figure 1: A full version of the instructions shown to the crowd workers iIn reference to Section 3.2. Data Annotation).
- Figure 2: The task page shown to the workers (in reference to Section 3.2. Data Annotation).
- Figure 3: Syntactic validation error messages shown to the workers (In reference to Section 3.2).
- Figure 4: Example animated GIFs with generated descriptions (in reference to Section 5.3 Results and Discussion).
- Figure 5: Side-by-side qualitative comparisons of sentences provided in the LSMDC dataset (professionally annotated descriptive video service (DVS) captions) and sentences annotated by the crowd workers (in reference to Section 2.1. Comparison with LSMDC).

In Figure 4 and Figure 5, we provide one still frame of each animated GIF for easy visualization. However, the still frames are inadequate for showing the dynamic visual content of animated GIFs. We therefore invite the reviewers to the following websites as an alternative to Figure 4 and Figure 5. Each website also contains more examples than we provide in this supplementary material.

- Figure 4: https://goo.gl/xcYjjE
- Figure 5: https://goo.gl/ZGYIYh

Instructions

Task

Below you will see five animated GIFs. Your task is to describe each animated GIF in Drow you wan see in the animate of a front as it to be used to be a set of the second of a one English sentence. You should focus solely on the visual content presented in the animated GIF. Each sentence should be grammatically correct. It should describe the main characters and their actions, but NOT your opinions, guesses or interpretations.

DOs

- Please use only English words. No digits allowed (spell them out, e.g., three).
 Each sentence must contain between 8 and 25 words. Try to be concise.
 Each sentence must contain a vetb.
 If possible, include adjectives that describe colors, size, emotions, or quantity.
 Please pay attention to grammar and spelling.
 Each sentence must express a complete idea, and make sense by itself.
 The sentence should describe the main characters, actions, setting, and ended the sentence should between the main characters. relationship between the objects

DONTS

- The sentence should NOT contain any digits.
 The sentence should NOT mention the name of a movie, film, and character.
 The sentence should NOT mention invisible objects and actions.
 The sentence should NOT make subjective judgments about the GIF.
 The sentence should NOT contain "there is" phrases, such as "there is", "there are", or "there exist"

Remember, please describe only the visual content presented in the animated GIF. Focus on the main characters and their actions.

Examples



Acceptable descriptions

Two lovers holding hands are walking near the beach. A couple holding hands is walking on a beach near an amusement park pier.

Unacceptable descriptions

- · Felicity and Anton being lovebirds in movie Like Crazy. -
- mentioning movie name and the character names
- Be happy, do not worry for nothing, blesses. subjective judgments/comments, slightly ungrammatical
- Couples are walking away from the camera. mentioning invisible objects: camera



Acceptable descriptions

A bald man is putting whipped cream onto his head.
A man wearing a plaid shirt is spraying whip cream on his head

Unacceptable descriptions

Macquark tests priors in live this (IF). – too short (4 words only) and not describing the animated GIF • This (GIF is fumay – too short and not describing the animated GIF Someone is isolaring forward. – too short, and does not describe the main action the man is performing does not describe the main character, nor the main action. • Man and cream – not a complete sentence. • Founty forward musi is worying about something. – does not describe visible actions in the animated GIF.





Figure 2: Task page. Each task page showed 5 animated GIFs and asked the worker to describe each with one sentence. This figure shows one of the five questions in a task page.



(a) A sentence must contain at least 8, but no more than 25 words.

Please describe the animated GIF in one English sentence.

	validation failed		
l	两个人在对视		
	English words only.		

(b) A sentence must contain only English words.

validation failed	validation failed
Two people, one with long hair and the other with mask	A man is glaring, and someone with sunglasses apears
Ungrammatical sentence. A grammatical sentence should have a subject, verb and object. Please refer to the instructions for acceptable sentences. Exa mples of Ungrammatical sentences, A female gymnast on the uneven bars. Someone coming into the house, carrying his skateboard. (Correction: S omeone came into the house, carrying his skateboard.)	Line 1, column 47, Rule ID: MORPOLOGIK_RULE_EN_US Message: Possible spelling mistake found Suggestion: appears; pairs; piers; Spears; spears; spares; amperes; pares; apexes; rapiers; a pears is glaring, and someone with sunglasses apears
(c) A sentence must contain a main verb.	(d) A sentence must be grammatical and free of typos.
Please describe the animated GIF in one English sentence.	
validation failed	
A man is glaring, and someone with sunglasses is is appearing	Please describe the animated GIF in one English sentence.
Line 1, column 47, Rule ID: ENGLISH_WORD_REPEAT_RULE	validation failed

Suggestion: is A man is glaring, and Jordan with sunglass appears ... is glaring, and someone with sunglasses is is appearing Line 1, column 50, Rule ID: PROGRESSIVE_VERBS[1] Message: This verb is normally not used in the progressive form. Try a sim ple form instead. ...glaring, and someone with sunglasses is is appearing

Names such as "Jordan" should not be mentioned

(f) A sentence must contain no named entities, such as an actor/actress, movie names, countries.

(e) A sentence must be grammatical and free of typos.

Figure 3: Syntactic validation. Since the workers provide free-form text, we validate the sentence before submission to reduce syntactic errors. This figure shows example messages shown to the workers in case we detect a syntactic error: error type (red speech balloon) and explanation for the error (caption below text input).



Figure 4: **Example animated GIFs and sentences generated** using the following algorithms: nearest neighbor (N), SMT-FrameNet (S), and LSTM-Finetune (L). The numbers in parentheses show the METEOR score (%) of each generated sentence. More examples with better visualization: https://goo.gl/xcYjjE



Figure 5: Side-by-side qualitative comparisons of sentences provided in the LSMDC dataset (professionally annotated descriptive video service (DVS) captions) and sentences annotated by the crowd workers. We converted movie clips from the LSMDC dataset into the GIF format and used the same pipeline described in our paper (Section 3) to crowdsource descriptions for the clips. Below each movie clip, we show both the crowdsourced description (CW) and the original description found in the LSMDC dataset (MD). We can see that the DVS captions tend to focus on visual content related specifically to the story (e.g., example No.6) or rely on contextual information not actually present within a video clip (e.g., "bear" in example No.11). More examples with better visualization: https://goo.gl/ZGYIYh