

ANOMALYLM: BRIDGING GENERATIVE KNOWLEDGE AND DISCRIMINATIVE RETRIEVAL FOR TEXT-BASED PERSON ANOMALY SEARCH

Anonymous authors

Paper under double-blind review

A APPENDIX

A.1 APPENDIX OUTLINE

This supplementary material includes 2 aspects:

1. Visualization:
 - Cloze generation
 - Cloze completion
2. Statement: Usage of LLMs

A.2 VISUALIZATION

A.2.1 CLOZE GENERATION

We present additional results for the cloze generation step in Figure 1. In these examples, meaningful verbs and color descriptions are substituted with the placeholders **<VERB>** and **<COLOR>**.

	Text Query	Cloze
(a)	A person is <u>standing</u> on a rocky ledge by a body of water, wearing <u>blue</u> shorts. He is about to <u>slide</u> into the water. He is <u>holding</u> onto the tree trunk with one hand and supporting the ground with the other. The surrounding area is lush with greenery and trees.	A person is <VERB> on a rocky ledge by a body of water, wearing <COLOR> shorts. He is about to <VERB> into the water. He is <VERB> onto the tree trunk with one hand and <VERB> the ground with the other. The surrounding area is lush with greenery and trees.
(b)	The image shows a child in a <u>green</u> shirt and <u>pink</u> skirt <u>playing</u> on a dirt ground. The little girl is <u>swinging</u> and <u>bumping</u> into another bent child, dressed in <u>colorful</u> clothing. There is a <u>red</u> chair and some mats in the background.	The image shows a child in a <COLOR> shirt and <COLOR> skirt <VERB> on a dirt ground. The little girl is <VERB> and <VERB> into another bent child, dressed in <COLOR> clothing. There is a <COLOR> chair and some mats in the background.
(c)	A man with a beard and sunglasses is wearing a <u>red</u> shirt and <u>beige</u> shorts. He is <u>falling</u> off the <u>pink</u> scooter because the pull rod of the scooter <u>broke</u> . The background features a <u>white</u> building with a window and a <u>yellow</u> curb.	A man with a beard and sunglasses is wearing a <COLOR> shirt and <COLOR> shorts. He is <VERB> off the <COLOR> scooter because the pull rod of the scooter <VERB> . The background features a <COLOR> building with a window and a <COLOR> curb.

Figure 1: (a-c) *Visualizations of cloze generation*: (left) Original text query, (right) Cloze. Verb and color words are replaced with placeholders **<VERB>** and **<COLOR>**, respectively.

A.2.2 CLOZE COMPLETION

We provide additional results for the cloze completion step in Figure 2. Figure 2 compares results of the cloze completion step using paired anomalous images (left) versus unpaired normal images

(right). The results show that crucial visual cues from paired anomalous images guide the LMM toward semantically aligned completions, whereas the absence of such cues in unpaired normal images leads to dissimilar outputs.

A.3 USAGE OF LLMs

In the preparation of this manuscript, the authors utilized a Large Language Model (LLM) as a general-purpose writing assistance tool. The use of the LLM was strictly limited to improving the language and readability of the text. Specifically, GPT is employed for tasks such as correcting grammatical errors, rephrasing sentences for better clarity and flow, and ensuring stylistic consistency.

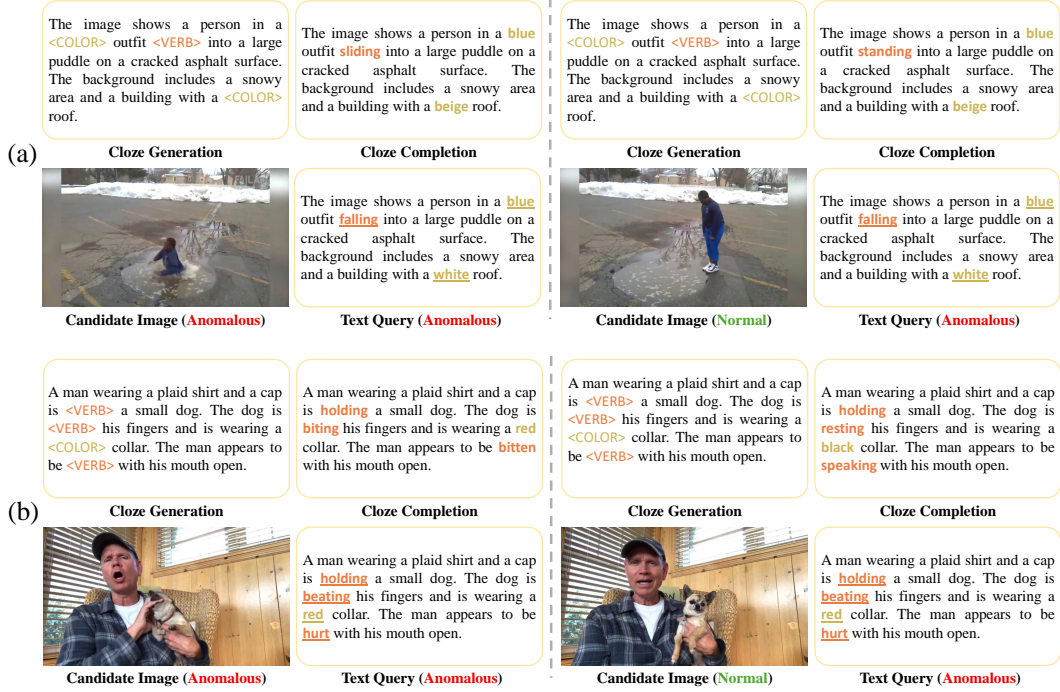


Figure 2: (a,b) Visualizations of cloze completion: (left) Cloze completion for anomalous-action candidate, (right) Cloze completion for normal-action candidate.