(a) Qualitative Case Study (Image Generation)

+think

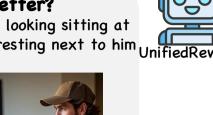
(Ours)

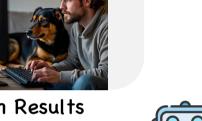
Which of the following images is better?

Text Caption: [A man wearing a brown cap looking sitting at his computer with a black and brown dog resting next to him UnifiedReward on the couch.]

Image 1:

Image 2:





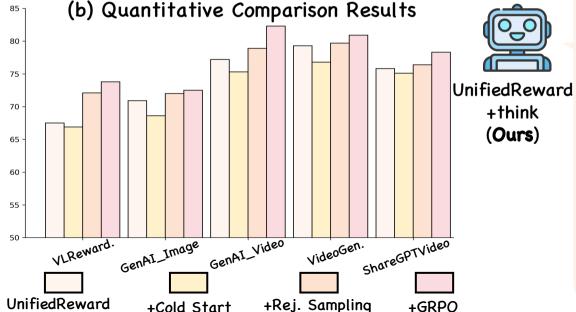


Image 1 is better



<think>

1. Semantic consistency: - Image 1 (7/10): The man is wearing a cap and is sitting with a dog, but the dog is not black and brown as described.

The setting is cozy, but the computer is not clearly visible. - Image 2 (9/10): The man is wearing a brown cap and is clearly using a computer. The dog is black and brown, matching

the description. The setting is consistent with the caption. 2. Aesthetics:

- Image 1 (8/10): The image has a warm, inviting color palette and good composition, with a cozy atmosphere.

- Image 2 (9/10): The image is well-composed with good lighting and focus on the man and the dog, creating a realistic and appealing scene.

3. Authenticity:

- Image 1 (6/10): The image has a more artistic,

illustrative style, which reduces realism. - Image 2 (9/10): The image appears realistic and detailed, with attention to the textures of the cap, dog fur, and computer.

Total score: Image 1: 7+8+6=21 Image 2: 9+9+9=27

</think>

<answer>Image 2 is better</answer>

