

Unlabeled Images



Image
Corruption



Good Prompting (step-by-step)

Please provide a **detailed** description of the image, focusing on the following aspects:

(1) **Identify the main subjects (people, animals, objects)** in the image and describe what they are doing

Bad Prompting

(1) Suggest and detail practical items or people that **could logically inhabit** the image's setting.
(2) Augment the scene with details of **potential events or items**.
(3) Incorporate elements that, **though absent**, would seamlessly **fit into the context** of the picture.

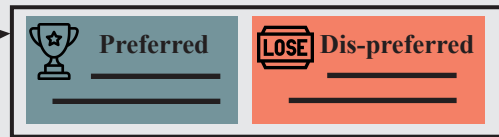


Stage-1
LVLM



Stage-2
LVLM

Stage 1: Image Comprehension Self-Training



DPO alignment fine-tuning w/ regularization



Supervised fine-tuning



Used instruction
fine-tuning data

Stage 2: Description-Infused Fine-Tuning