

Describe the image in detail.



(a)

**mPLUG-Owl3:** The image features a train traveling down the tracks. ... There is a car visible in the background, likely waiting for the train to pass. The scene also includes a few people scattered around the area, possibly waiting for the train or observing it as it goes by.

**+Ours:** The image features a train traveling down the tracks. ... **There is no car visible in the background;** only the train and railway infrastructure are present. **The scene does not include any people scattered around the area.**

Describe the image in detail.



(b)

**LLaVA-1.5-13B:** The image captures a lively scene at a park, where *a group of people* is playing a game of frisbee ... In the background, there are several cars parked along the park, and *a truck is also visible* ...

**+Ours:** The image captures a lively scene at a park, where **two people** is playing a game of frisbee ... In the background, there are several cars parked along the park ...

Describe the image in detail.



(c)

**LLaVA-1.5-13B:** The image features a street scene with a wooden bench situated. ... There are two people in the scene, one standing closer to the left side of the image and the other person standing near the right side. ... and a car is visible in the background.

**+Ours:** The image features a street scene with a wooden bench situated. **The storefront in the background has signage that reads "LA TECA DE LAVIA" and there is an awning above it. A red trash bin labeled "FRIGO" is placed near the entrance of the store, which also displays various posters and signs outside its windows.** ... **There are no people in the scene. ... and there is no car visible in the background.**