

Describe the image in detail



Vision  
Language  
Model

→ I

The image depicts a busy city street filled with cars, traffic lights, and **pedestrians**. The street is lined with **houses** on the **right** side of the **road**. A **yellow traffic light**... Overall, the image depicts a **rainy** day...

+

**HalLocalizer**

pedestrian	0.9	obj
green traffic light	0.5	attr
houses-right-road	0.8	rel
rainy	0.9	scene



Attend to the **attribute** of traffic light



Attend to the **scene** details



Focus on tokens with high hallucination probability

