

---

**Algorithm 1** Pseudo Code for Text-Correlated Attention Heatmap Generator

---

```
1:  $E_t \leftarrow model.e1$  ▷ Define  $E_t$  as model embedding  
2:  $image\_tensor \leftarrow INPUTIMAGE TENSOR$  ▷ Input image tensor  
3:  $img\_mean \leftarrow model.get\_vision\_tower().image\_processor.image\_mean$   
4:  $img\_std \leftarrow model.get\_vision\_tower().image\_processor.image\_std$   
5:  $image\_tensor \leftarrow SQUEEZEFIRSTDIM(image\_tensor)$  ▷ Remove batch dimension  
6:  $img\_show \leftarrow PERMUTETOHWC(image\_tensor)$  ▷ To Height x Width x Channel  
7:  $img\_show \leftarrow CONVERTTONUMPYANDREVERSECHANNELS(img\_show)$   
8:  $img\_show \leftarrow UNNORMALIZE(img\_show, img\_std, img\_mean) \times 255$   
9:  $img\_score \leftarrow NORMALIZE(E_t)$  ▷ Normalize scores to [0, 1]  
10:  $img\_score\_avg \leftarrow MEAN(img\_score, 1)$  ▷ Compute average scores  
11:  $img\_score\_reshaped \leftarrow RESHAPE TO 24 \times 24(img\_score\_avg)$   
12:  $threshold \leftarrow FIND50TH HIGHEST VALUE(img\_score\_reshaped)$   
13:  $img\_score\_reshaped \leftarrow ZEROBELOW THRESHOLD(img\_score\_reshaped, threshold)$   
14:  $img\_score\_upscaled \leftarrow UPSCALE TO MATCH IMAGE(img\_score\_reshaped, image\_tensor.shape[1], image\_tensor.shape[2])$ 
```

---



The image features a white vase filled with **yellow flowers**, placed on a table or countertop. This vase is accompanied by two smaller vases, each containing yellow flowers. The arrangement creates a visually appealing and vibrant display of yellow flowers. The table or countertop appears to be a **white surface**, which complements the color of the flowers.



The image features a young girl **standing outside**, holding a blue umbrella to protect herself from the rain. She is wearing a pink jacket and a scarf, adding a touch of color to the scene. The **umbrella** is open, covering the girl and providing shelter from the elements. The girl appears to be looking up, possibly observing something above her.



Figure 1: Example of Knowledge Ability Comparative Experiments



The image features a **yellow sports** car parked on the side of a street, likely in a parking lot or a designated parking area near a business district. **Several people** can be seen walking around the area. The street is lined with **various vehicles** parked along the side. The presence of multiple people and vehicles suggests that this is a busy urban area with a mix of commercial and residential spaces.

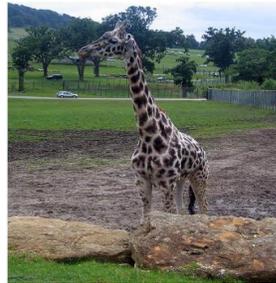


In the image, a woman is feeding a **giraffe** from her hand, while the giraffe leans its head to reach her. The woman is wearing **a hat** and is standing close to the giraffe. The scene appears to be taking place in a zoo setting, as the giraffe is in an enclosure, surrounded.

Figure 2: Example of Knowledge Ability Comparative Experiments



The image depicts **a group of cows walking down a street** in a residential area. There are **three cows** in total. The cows appear to be **moving in the same direction**, possibly heading towards a **fence** or a **gate**. The street is lined with houses on both sides, creating a cozy and residential atmosphere. The cows are walking on the sidewalk, which is located between the houses and the street.



The image features a **giraffe** standing **in a dirt field**, surrounded by a fence. **The giraffe is positioned near the center of the scene**, with its long neck and legs prominently visible. In the background, there are several cars parked, likely belonging to visitors to the zoo.

Figure 3: Example of Knowledge Ability Comparative Experiments



The image captures a thrilling moment of **a snowboarder in the air**, performing a **jump on a snow-covered slope**. The snowboarder is **in the middle of the jump**, showcasing their skill and athleticism. The snowboard is clearly visible beneath the snowboarder, as they soar through the air. The scene is set against a backdrop of snow-covered ground, creating a picturesque winter landscape. The snowboarder's impressive jump is the focal point of the image, highlighting their talent and the excitement of the sport.



The image features a **baby sitting** in a **high chair**, looking at a colorful birthday cakes placed in front of her on the tray of the high chair. The baby appears to be curious and excited about the cakes, possibly waiting for someone to cut and serve them.



Figure 4: Example of Knowledge Ability Comparative Experiments



The image features **a man standing in a grassy field**, flying a **kit** in the air with a string attached. In the background, **there are trees**, creating a pleasant outdoor setting for kite-fighting enthusiasts to engage their skills on windless days like **the man in jean short sleeveless shirts** is doing in his park-like surroundings on a clear day with blue skies and green grassy field.



The image features **a man walking on the beach**, holding a **surfboard**. The man is wearing a **wetsuit**, which suggests that he is likely preparing to go surfing. He is walking along the shimmering ocean, enjoying his time by its beautiful waters.



Figure 5: Example of Knowledge Ability Comparative Experiments

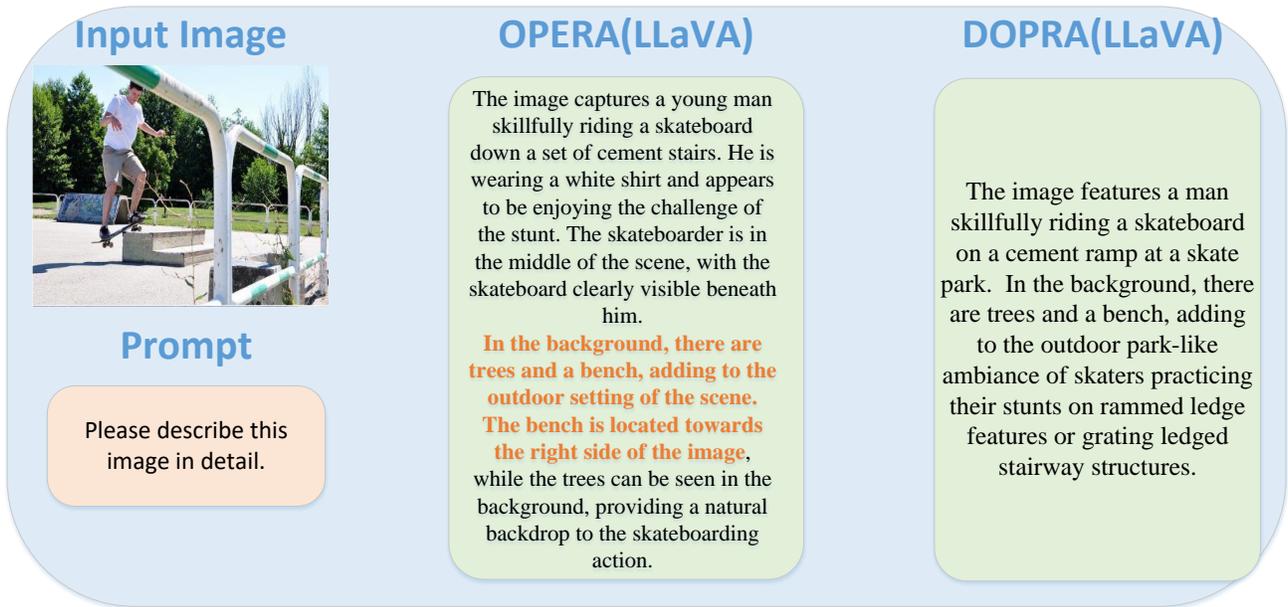


Figure 6: Examples of OPERA vs. DOPRA comparisons

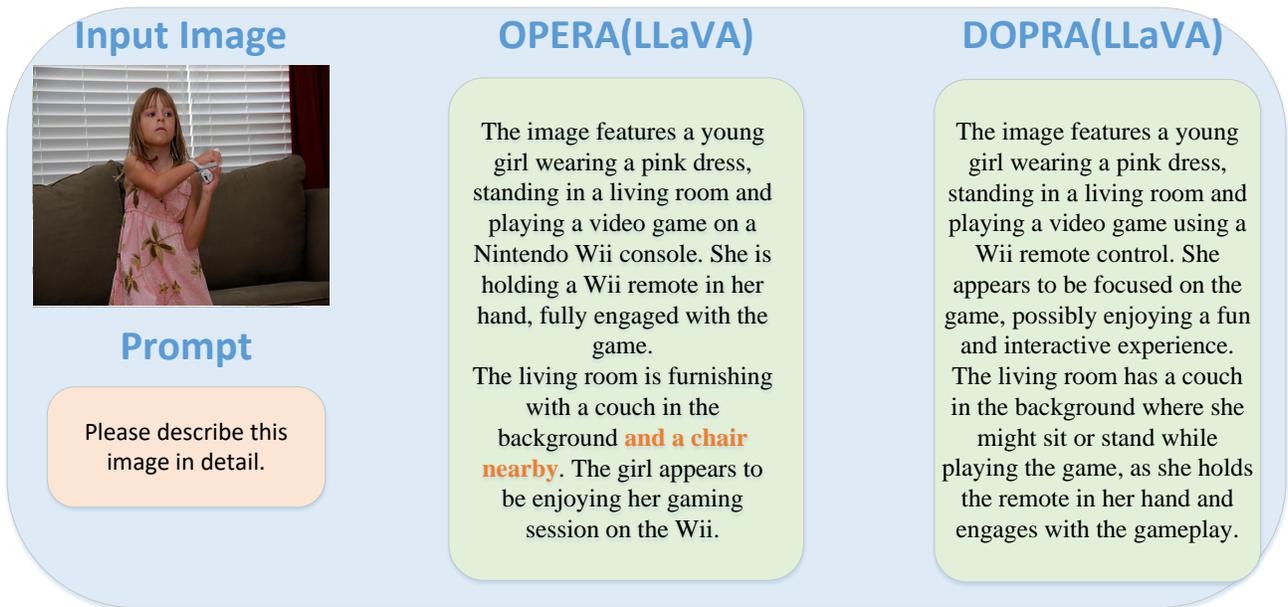


Figure 7: Examples of OPERA vs. DOPRA comparisons

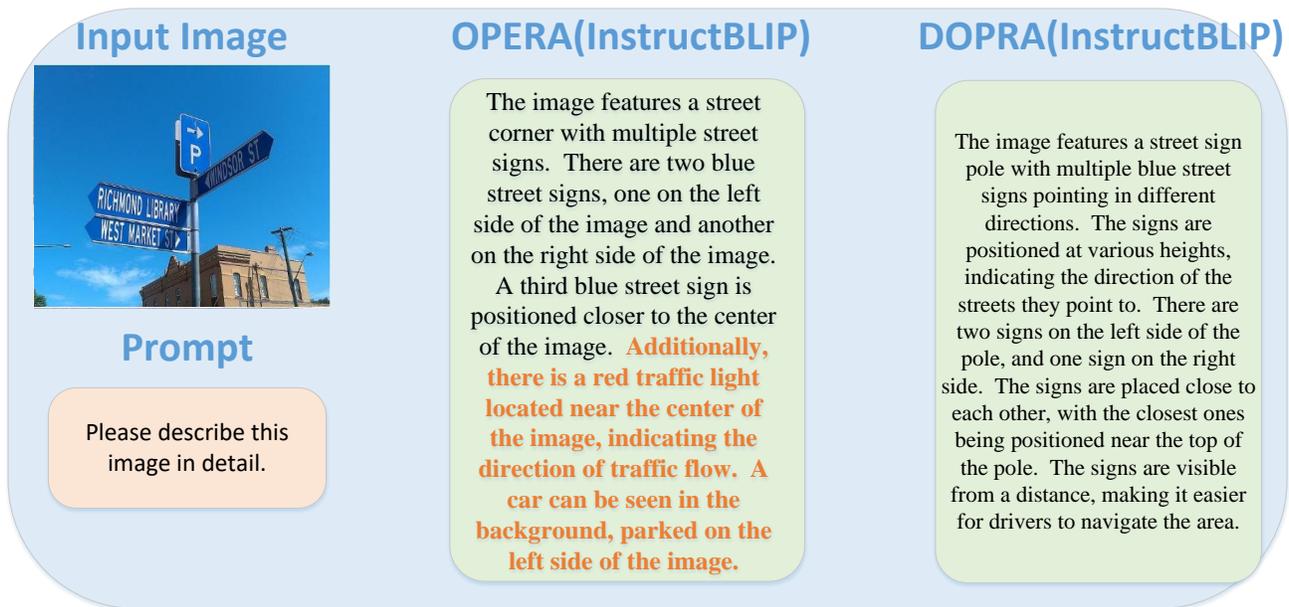


Figure 8: Examples of OPERA vs. DOPRA comparisons

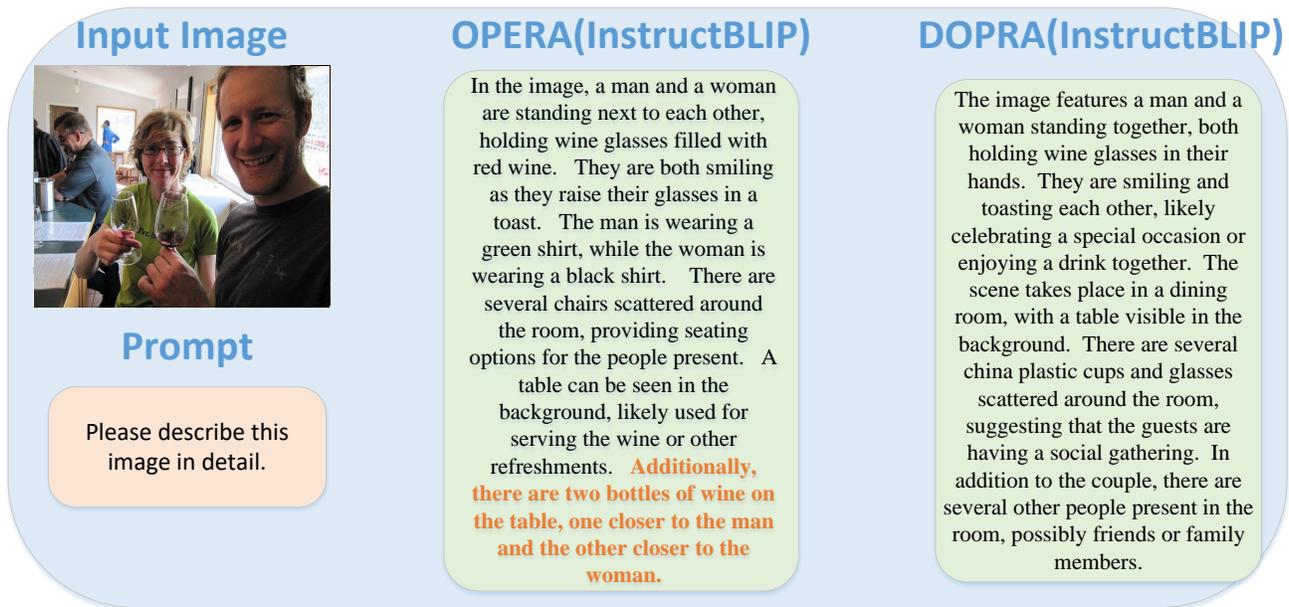


Figure 9: Examples of OPERA vs. DOPRA comparisons

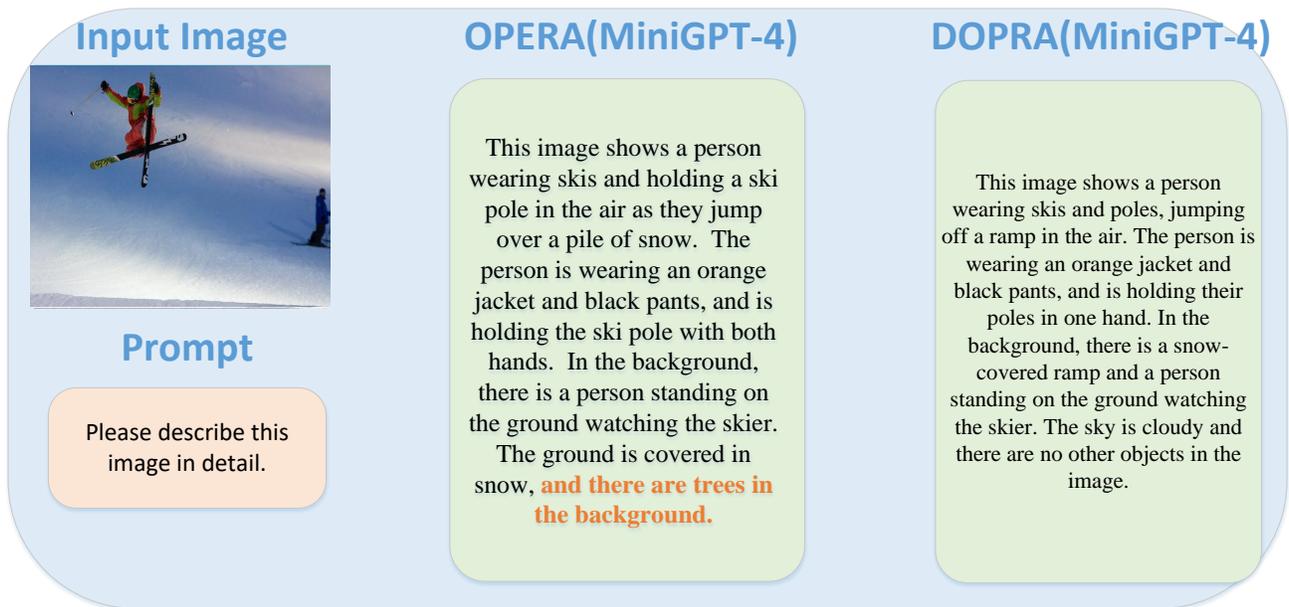


Figure 10: Examples of OPERA vs. DOPRA comparisons

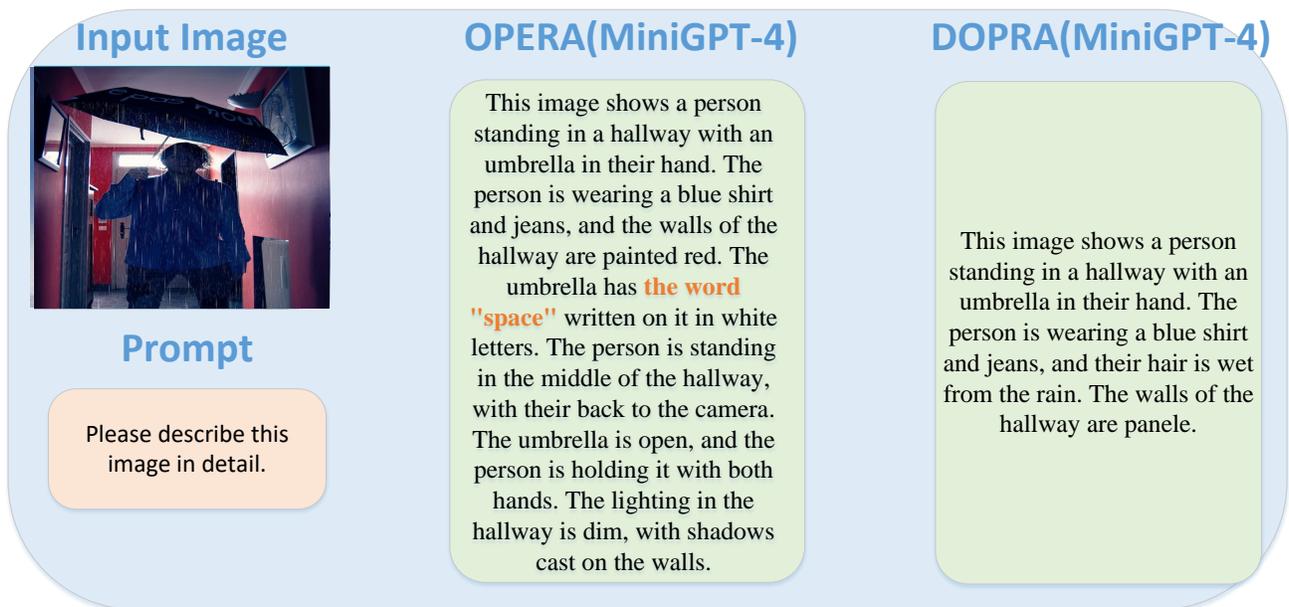


Figure 11: Examples of OPERA vs. DOPRA comparisons