

# Supplementary Material for Chain of Visual Perception: Harnessing Multimodal Large Language Models for Zero-shot Camouflaged Object Detection

This supplementary material contains the following parts:

- Section 1 provides more text prompts we used in this paper.
- Section 2 shows more qualitative results of MMCPF.

We hope this supplementary material can help you get a better understanding of our work.

## 1 MORE TEXT PROMPTS

The text prompt we used in our proposed MMCPF containing:

- (1) Please find a camouflaged object in this image and provide me with its exact location coordinates.
- (2) This image may contain a camouflaged object whose shape, color, texture, pattern and movement closely resemble its surroundings, enabling it to blend in. Can you identify it and provide its precise location coordinates?
- (3) This image may contain a concealed object whose shape, color, texture, pattern and movement closely resemble its surroundings, enabling it to blend in. Can you identify it and provide its precise location coordinates?
- (4) This image may contain a camouflaged object whose shape, color, pattern, movement and texture bear little difference compared to its surroundings, enabling it to blend in. Please provide its precise location coordinates.
- (5) This image may feature a camouflaged object that shares similarities in shape, color, pattern, movement, and texture with its surroundings, enabling it to blend seamlessly. Please provide me with its exact location coordinates.
- (6) This image may feature a concealed object or animal that shares similarities in shape, color, pattern, movement, and texture with its surroundings, enabling it to blend seamlessly. Please provide me with its exact location coordinates.

Note that, in our baseline model, only using vanilla text prompts can result in the failure to output corresponding coordinates of camouflaged objects for certain images. Therefore, for these specific images, we generate four random prompt points to serve as inputs for the VFM. Finally, when we use the text prompts designed in this paper, all images would get their corresponding coordinates of camouflaged objects.

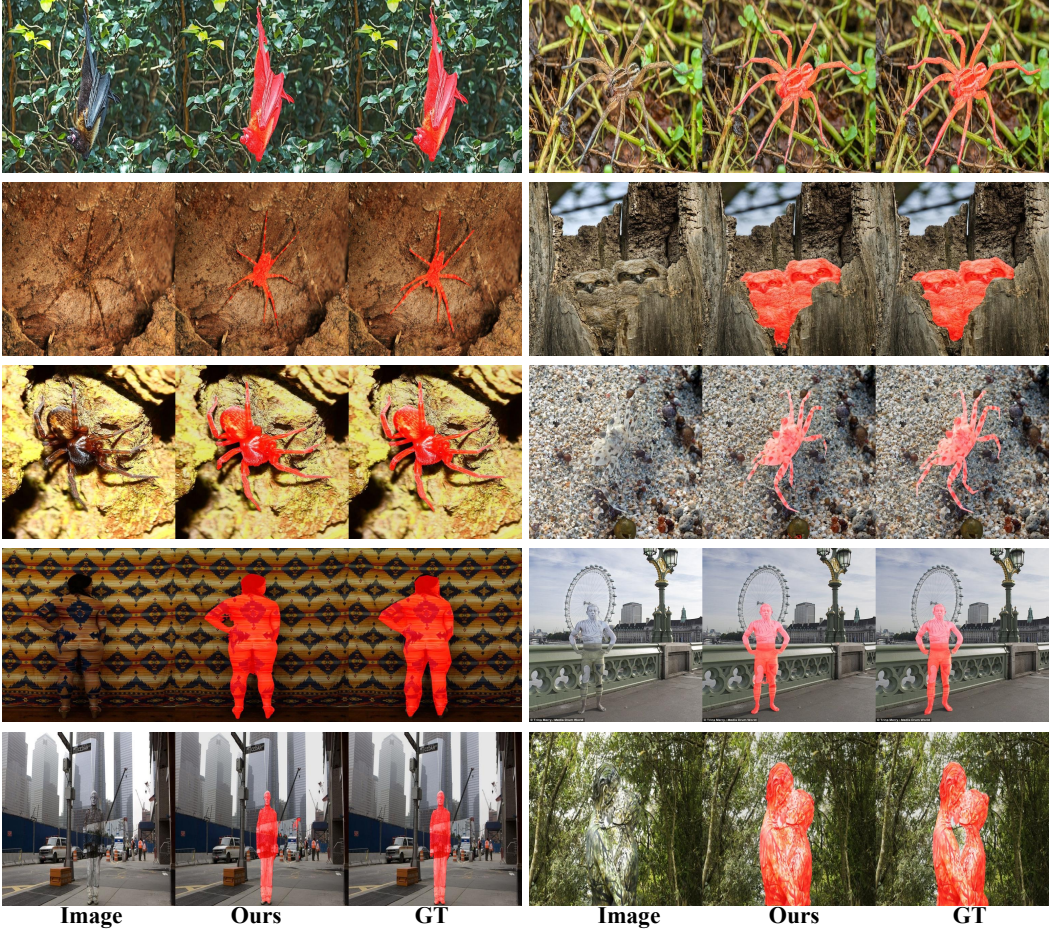


Fig. 1. More qualitative results of the proposed MMCPF framework.

## 2 MORE QUALITATIVE RESULTS

In Fig. 1, we provide more comprehensive qualitative results of our method. The qualitative results clearly demonstrate that our MMCPF is capable of effectively detecting camouflaged objects across different scenes. This versatility further underscores the potential of our framework.