

(a) Qualitative comparison on FLMS

(b) Qualitative comparison on CaptureGuide-Bench (PSG)

User: Recommend a 16:9 composition.

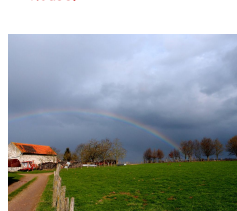
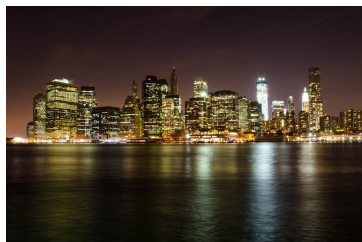
Capture-MLLM: ... the overall visual weight is too high...it is recommended to **adjust the framing range** so that the horizon line is placed in a more appropriate position. ...

User: I want to take a photo of **Osaka Castle**. Could you recommend a composition?

Capture-MLLM: ...It is recommended to use a **centered composition**...

User: I want to take a photo of this **rainbow** and the **orange-roofed house**.

Capture-MLLM: ...the visual weight is clearly shifted downward...it is recommended to **use the rule of thirds**, and place the house and rainbow...



(c) Demonstration of Capture-MLLM's interpretable and interactive aesthetic cropping