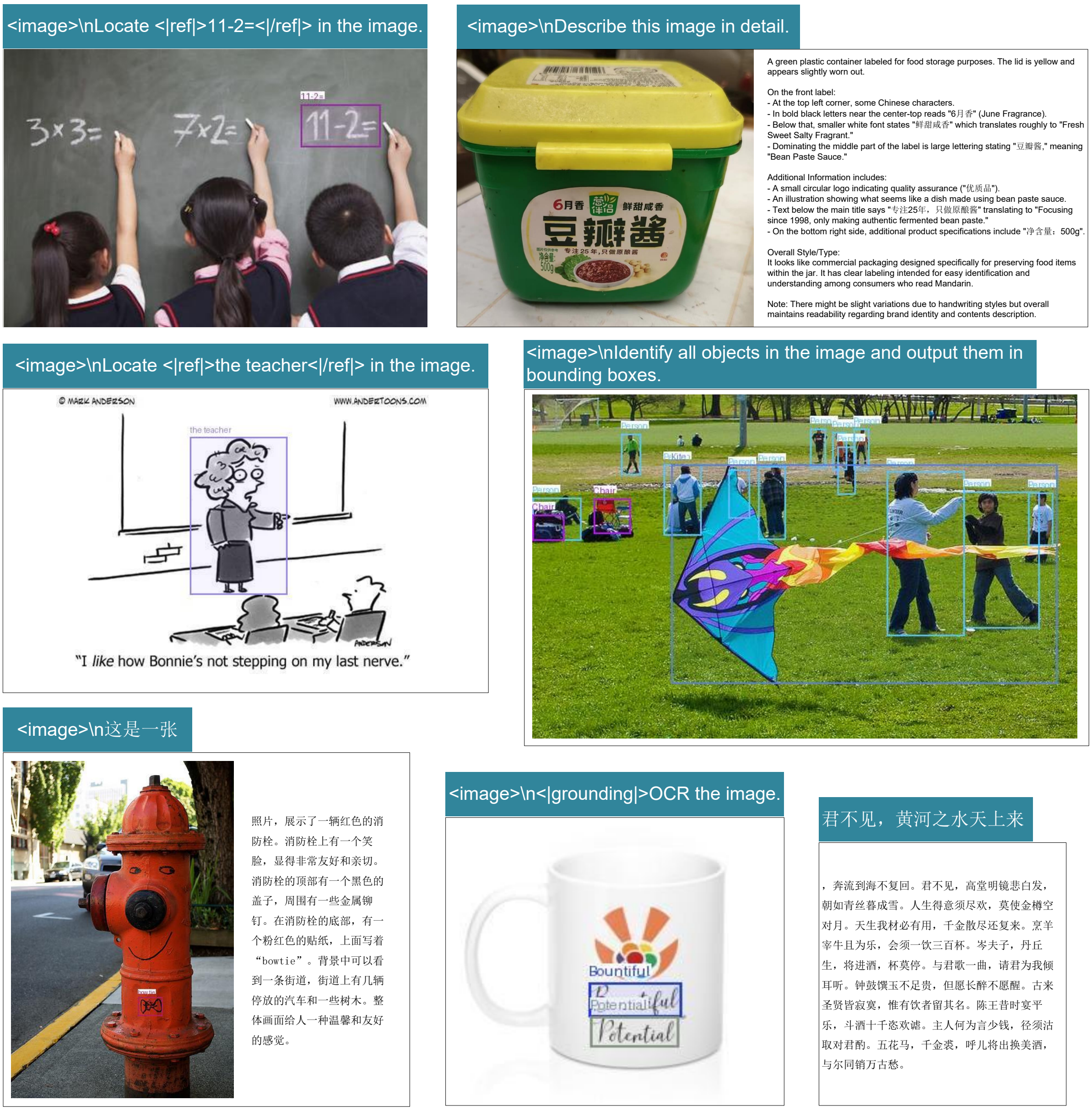

<image>\nLocate <|ref|>11-2=<|/ref|> in the image.

<image>\nDescribe this image in detail.

A green plastic container labeled for food storage purposes. The lid is yellow and appears slightly worn out.

On the front label:
- At the top left corner, some Chinese characters.
- In bold black letters near the center-top reads "6月香" (June Fragrance).
- Below that, smaller white font states "鲜甜咸香" which translates roughly to "Fresh Sweet Salty Fragrant."
- Dominating the middle part of the label is large lettering stating "豆瓣酱," meaning "Bean Paste Sauce."

Additional Information includes:
- A small circular logo indicating quality assurance ("优质品").
- An illustration showing what seems like a dish made using bean paste sauce.
- Text below the main title says "专注25年，只做原酿酱" translating to "Focusing since 1998, only making authentic fermented bean paste."
- On the bottom right side, additional product specifications include "净含量：500g".

Overall Style/Type:
It looks like commercial packaging designed specifically for preserving food items within the jar. It has clear labeling intended for easy identification and understanding among consumers who read Mandarin.

Note: There might be slight variations due to handwriting styles but overall maintains readability regarding brand identity and contents description.

<image>\nLocate <|ref|>the teacher<|/ref|> in the image.

<image>\nIdentify all objects in the image and output them in bounding boxes.

<image>\n这是一张

照片，展示了一辆红色的消防栓。消防栓上有一个笑脸，显得非常友好和亲切。消防栓的顶部有一个黑色的盖子，周围有一些金属铆钉。在消防栓的底部，有一个粉红色的贴纸，上面写着“bowtie”。背景中可以看到一条街道，街道上有几辆停放的汽车和一些树木。整体画面给人一种温馨和友好的感觉。

<image>\n<|grounding|>OCR the image.

君不见，黄河之水天上来

，奔流到海不复回。君不见，高堂明镜悲白发，朝如青丝暮成雪。人生得意须尽欢，莫使金樽空对月。天生我材必有用，千金散尽还复来。烹羊宰牛且为乐，会须一饮三百杯。岑夫子，丹丘生，将进酒，杯莫停。与君歌一曲，请君为我倾耳听。钟鼓馔玉不足贵，但愿长醉不愿醒。古来圣贤皆寂寞，惟有饮者留其名。陈王昔时宴平乐，斗酒十千恣欢谑。主人何为言少钱，径须沽取对君酌。五花马，千金裘，呼儿将出换美酒，与尔同销万古愁。

Figure 12 | We retain DeepSeek-OCR's capabilities in general visual understanding, mainly including image description, object detection, grounding, etc. Meanwhile, due to the inclusion of text-only data, DeepSeek-OCR's language capabilities are also retained. Note that since we do not include SFT (Supervised Fine-Tuning) stage, the model is not a chatbot, and some capabilities need completion prompts to be activated.

5. Discussion

Our work represents an initial exploration into the boundaries of vision-text compression, investigating how many vision tokens are required to decode N text tokens. The preliminary results are encouraging: DeepSeek-OCR achieves near-lossless OCR compression at approximately 10× ratios, while 20× compression still retains 60% accuracy. These findings suggest promising directions for future applications, such as implementing optical processing for dialogue histories beyond k rounds in multi-turn conversations to achieve 10× compression efficiency.