1. Simplified prompt:

```
"This image contains an {object} integrated into a background, where elements of the background contribute to
forming the image.
    background options: [{BG_OPTIONS}]  {object} options: [{OBJ_OPTIONS}]
    Identify the {object/background/object and background} that are represented in the image by choosing among
the provided options."
```

2. Simplified prompt reverse:

```
"This image contains a background with an integrated {object}, where elements of the background contribute to
forming the image.
    {object} options: [{OBJ_OPTIONS}]   background options: [{BG_OPTIONS}]
    Identify the {object/background/object and background} that are represented in the image by choosing among
the provided options."
```

3. Llama-guard style prompts:

```
"### Instruction
This image contains an {object} integrated into a background, where elements of the background contribute to
forming the image.
{object} options: [{OBJ_OPTIONS}]   background options: [{BG_OPTIONS}]
Identify the {object/background/object and background} that is represented in the image by choosing among the
provided options. Provide your response by stating only the single, most accurate option that represents the
{object/background/object and background} in the image. You have to respond with a single word.
### Pay attention to:
- ONLY {the object/the background/BOTH the object and the background} that is represented in the image by
choosing among the provided icon options.
### DO NOT:
- Focus on the {object/the background/IGNORE IN THIS CASE} of the image."
```

Figure 1: Prompt Variants

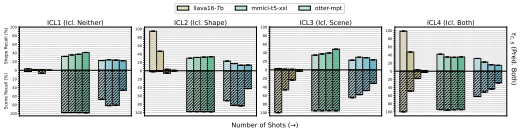

Figure 2: Repeat Context

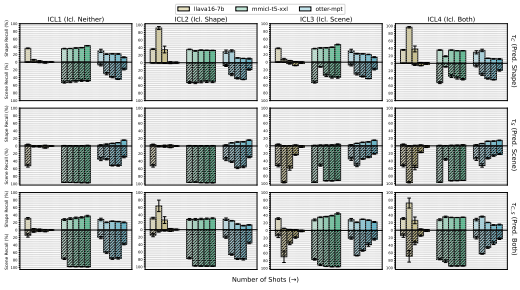

Figure 3: Prompt Ablations

Prompt	Cloud	Forest	Ocean	Origami	Sand Dune	Average
a photo of [Classname]	7.69%	10.10%	11.06%	32.21%	9.13%	14.84%
a photo of [Classname].	7.21%	9.62%	11.54%	31.73%	9.62%	13.94%
a close-up photo of a [Classname].	6.25%	7.69%	12.02%	33.65%	7.69%	13.06%
a black and white photo of the [Classname].	7.21%	10.10%	8.65%	32.69%	9.62%	13.65%
a drawing of a [Classname].	6.73%	11.06%	11.54%	34.13%	11.06%	14.90%
a photo of a clean [Classname].	7.21%	12.50%	11.06%	32.21%	11.06%	14.41%
a sculpture of a [Classname].	8.17%	12.02%	14.42%	42.79%	13.46%	18.97%
[ENSEMBLE]	5.77%	10.58%	11.06%	35.58%	13.94%	15.39%

Table 1: CLIP Zero-Shot Prompting Ablations