

---

# Appendix for Image Textualization: An Automatic Framework for Creating Accurate and Detailed Image Descriptions

---

1 In this appendix, we first provide the detailed prompts for object entity extraction via LLM, which  
2 was adopted in phase 2 for hallucination identification. Then, we illustrate the prompts and in-context  
3 examples for textualized recaptioning. Next, we showcase more qualitative comparisons between  
4 IT-generated and MLLM-generated image descriptions. Finally, we point out the limitation of our  
5 work, which may be addressed in future works.

## 6 **A Detailed Prompt Design for Object Entity Extraction in Phase 2**

7 In table 1, we demonstrate the detailed prompt that guides the LLM to perform entity extraction from  
8 the template description. Specifically, we first indicate the role of the LLM. Then we emphasize the  
9 things to remember during the extraction process: 1) only extract the objects that certainly exists in the  
10 image; 2) avoid extracting the background objects or intangible items; 3) the response should follow  
11 a certain format to facilitate parsing. Next, we provide the LLM with human-annotated in-context  
12 examples to enable better instruction following ability. Lastly, we provide the LLM with the new  
13 template image description to perform entity extraction. In table 2, we showcase the in-context  
14 examples provided to LLM for entity extraction.

## 15 **B Prompt for Textualized Recaptioning in Phase 3**

16 In table 3, we demonstrate the prompt for textualized recaptioning. We first inform the LLM of  
17 its role as a recaptioner. Then, we provide the detailed explanation for the object-level spatial  
18 information, i.e., Relative spatial position, relative depth from lens and relative size of the objects.  
19 Next, we emphasize the following points: 1) avoiding duplication when incorporating new objects  
20 into the description, 2) the photographic characteristics mentioned in the Original Description should  
21 be preserved; 3) the exact values of the spatial information (e.g., bounding boxes) should not be  
22 incorporated into the description. Afterwards, we provide human-annotated in-context examples to  
23 enhance the annotation quality.

24 In table 5, we showcase one of the in-context examples provided to the LLM for textualized re-  
25 captioning. In figure 1, we compare the recaption results without fine-grained object annotation  
26 and in-context examples and observe that object annotation leads to more detailed and accurate  
27 description, while in-context examples effectively prevent the description to contain exact values of  
28 fine-grained information.

**Extract Prompt**

**###TASK DESCRIPTION###**

Your are a helpful entities extractor. Please help me to extract the OBJECTS mentioned in a description about an image.

**###THINGS TO REMEMBER###**

1. Only extract the descriptions of objects that are described with certainty. For example, in the sentence "there's a white car parked, perhaps belonging to one of the hotel guests," the "hotel guests" part is included within "perhaps," indicating uncertainty. Therefore, you only need extract "a white car" that is described with certainty.
2. Avoid extracting abstract or non-specific entities (such as "cozy atmosphere", "excitement", "sky view")!!!
3. Your response should strictly start with "%%%RESPONSE%%%" , following this format: "%%%RESPONSE%%": obj1. obj2. obj3...."

**###IN-CONTEXT EXAMPLES###**

Here are some examples for your reference.

<In-Context Examples>

**###TASK###**

%%%DESCRIPTION%%%: <Description>

%%%RESPONSE%%%:

Table 1: The prompt for extracting entities in the description, <In-Context Examples> is the placeholder for several in-context examples that illustrated in Table 2. <Description> will be replaced by the description to be modified.


	<p><b>w/o annotations + shots</b></p> <p>This image captures a cozy dinner setting ...          Beside each plate ... A wine glass sits on the table alongside a clear glass of water. A knife and a fork are neatly arranged on the table. Additionally, a silver knife rests on one of the plates, emphasizing the meticulous arrangement of the table elements.          The table setting is completed with white candles housed in glass holders, casting a warm and inviting glow over the scene. The presence of a person sitting at the table adds a personal touch, enhancing the atmosphere of intimacy and comfort. The overall composition of the image conveys ...</p>
	<p><b>w/o shots</b></p> <p>This image captures a cozy dinner setting ...          Beside each plate, there are multiple glasses - some filled with water, such as the one positioned at [0.46, 0.48, 0.59, 0.67] with a relative size proportion of 1.99% and another at [0.25, 0.43, 0.39, 0.62] with a relative size proportion of 1.49%, and others with white wine, like the one at [0.27, 0.13, 0.36, 0.34] with a relative size proportion of 0.9%. The table setting is completed with white candles housed in glass holders, casting a warm and inviting glow over the scene, with one candle positioned at [0.0, 0.56, 0.14, 0.74] with a relative size proportion of 2.09%.          The image exudes ... On the table, there are also utensils, including a knife and fork at [0.64, 0.3, 0.75, 0.41] with a relative size proportion of 0.23% and a silver butter knife at [0.43, 0.81, 0.56, 1.0] with a relative size proportion of 0.59%. There is a person sitting at the table at [0.55, 0.0, 1.0, 0.32] with a relative size proportion of 10.78%. The overall composition of the image conveys ...</p>
	<p><b>ours</b></p> <p>This image captures a cozy dinner setting ...          Beside each plate, there are multiple glasses - some filled with water, such as the one positioned near the center-right of the table, and another near the upper-left side, and others with white wine, like the one positioned toward the upper-left corner of the table. The table setting is completed with white candles housed in glass holders, casting a warm and inviting glow over the scene, with one candle positioned toward the lower-left side of image on the table.          The image exudes ... On the table, there are also utensils, including a knife and fork positioned near the upper center-right and a silver butter knife near the bottom-center. There is a person sitting at the table, occupying a large portion of the upper right corner of the image. The overall composition of the image conveys ...</p>

Figure 1: Comparison with results generated without using fine-grained annotation and in-context examples. We

29 **C Qualitative Comparison between IT-generated and MLLM-generate Image**  
 30 **descriptions**

31 We provide more qualitative comparisons between MLLM-generated and IT-generated image descrip-  
 32 tions in table 6, table 7 and table 8. We observe that compared with MLLM-generated descriptions,

In-Context Examples for the Extract Prompt	
<p><b>Example 1:</b></p> <p>%%%DESCRIPTION%%%: This image captures a cozy Christmas scene. At the center of the image is a bed adorned with a <b>red satin bedspread</b> with <b>gold tassels</b> and a <b>gold pillow</b>. The bed is nestled within <b>red curtains</b> that have <b>gold trim</b>, adding to the festive atmosphere. On the bed, there are <b>four teddy bears</b>, two of which are brown and the other two are white. They are arranged in a pile, creating a sense of warmth and comfort. Adding to this feeling is a <b>white cat</b> that is peacefully sleeping next to the teddy bears. The bed is further decorated with <b>red and gold Christmas ornaments</b>, enhancing the holiday spirit. To the right of the bed stands a <b>Christmas tree</b> adorned with <b>gold ornaments</b>, while on the left side of the bed, there are <b>two red candlesticks</b>. Overall, this image exudes a warm and festive atmosphere, perfect for the holiday season.</p> <p>%%%RESPONSE%%%: <b>red satin bedspread. gold tassels. gold pillow. red curtains. gold trim. four teddy bears. white cat. red and gold Christmas ornaments. Christmas tree. gold ornaments. two red candlesticks.</b></p>	
<p><b>Example 2:</b></p> <p>%%%DESCRIPTION%%%: The image depicts a lively scene in an ornately decorated room. At the center of the image stands a <b>man with a long white beard</b>, wearing a <b>tall, cylindrical black hat</b> with a flat top. He is dressed in a <b>white shirt</b> paired with a <b>green tie</b> and carries a <b>black backpack</b>. The room is bustling with <b>people</b>, some of whom are wearing <b>green shirts</b>, adding to the vibrant atmosphere. The room itself is richly adorned, featuring a <b>gold ceiling</b> and <b>red columns</b> that exude an air of grandeur. The man, despite being in the midst of the crowd, stands out due to his unique attire and commanding presence. The <b>people</b> are visible on all sides of him, suggesting that he is at the heart of this gathering.</p> <p>%%%RESPONSE%%%: <b>man with a long white beard. tall cylindrical black hat. white shirt. green tie. black backpack. people wearing green shirts. gold ceiling. red columns.</b></p>	
<p><b>Example 3:</b></p> <p>%%%DESCRIPTION%%%: This image captures a charming scene in a bedroom. The main subject is a <b>black cat</b> standing on its hind legs on a <b>gray nightstand</b>. The cat, full of curiosity, is reaching its paw into a <b>white cup adorned with blue and green designs</b>. The nightstand hosts a few other items: a <b>beige lamp</b>, a <b>white tissue box with blue flowers</b>, and a <b>pink jar of Vaseline</b>. Each item is neatly arranged on the nightstand, creating a harmonious tableau. In the background, you can see a bed covered with a <b>white comforter</b> and adorned with a <b>yellow pillow</b>, adding a pop of color to the scene. The precise location of these objects and their relative positions contribute to the overall composition of the image, creating a snapshot of everyday life with a hint of feline mischief.</p> <p>%%%RESPONSE%%%: <b>black cat. gray nightstand. white cup with blue and green designs. beige lamp. white tissue box with blue flowers. pink jar of Vaseline. white comforter. yellow pillow.</b></p>	

Table 2: In-context examples for object entity extraction.

33 IT-generated ones incorporate more comprehensive visual details, and also demonstrate less hallucination.  
 34

35 **D Limitation**

36 Although we conduct extensive experiments on the quality of the generated image descriptions, we  
 37 did not tune larger multi-modal large language models (e.g., LLaVA 70B) due to limitations on  
 38 computational resources. However, we expect Image Textualization to achieve similar or even better

39 performance on larger models compared with LLaVA-7B model due to their stronger generalization  
40 ability and robustness to forgetting. We will leave this to the future work.

Recaptioning Prompt
<p><b>###TASK DESCRIPTION###</b></p> <p>You are a helpful language assistant. Imagine visualizing an image based on its description. Now, your task is to make the Original Description more detailed. You'll need to use the subsequent provided Objects, along with the corresponding extra information of <i>Relative Spatial Positioning</i>, <i>Relative Distance from the Lens</i>, and <i>Relative Size Proportion in Images (Percentage)</i>, to better assist you in adding these Objects to the original description.</p> <p><b>###EXTRA INFORMATION EXPLANATION###</b></p> <ol style="list-style-type: none"> <li><b>1. <i>Relative Spatial Positioning</i>:</b> It uses a normalized coordinate system where both x (horizontal) and y (vertical) axes range from 0 to 1. The x-coordinate starts at 0 on the image's left edge and increases to 1 towards the right edge. Similarly, the y-coordinate starts at 0 at the top edge and increases to 1 towards the bottom. This system uses four coordinates to define the corners of a rectangle within the image: [x1, y1, x2, y2], representing the top-left and bottom-right corners of the rectangle, respectively. For instance, a positioning of [0.00, 0.00, 0.50, 0.50] means the object's top-left corner is at (0.00, 0.00) and its bottom-right corner is at (0.50, 0.50), placing the object in the upper left quarter of the image. Similarly, [0.50, 0.00, 1.00, 0.50] positions the object in the upper right quarter, with corners at (0.50, 0.00) and (1.00, 0.50). A positioning of [0.00, 0.50, 0.50, 1.00] places the object in the bottom left quarter, with corners at (0.00, 0.50) and (0.50, 1.00), while [0.50, 0.50, 1.00, 1.00] positions it in the bottom right quarter, with corners at (0.50, 0.50) and (1.00, 1.00). Moreover, by comparing these coordinates, you can determine the relative positions of objects. For example, an object with positioning [0.20, 0.20, 0.40, 0.40] is to the left of another with [0.30, 0.30, 0.50, 0.50].</li> <li><b>2. <i>Relative Distance from the Lens</i>:</b> It measures how far objects are from the camera within the image. The closer the value is to 1, the nearer the object is to the camera, placing it in the foreground. Conversely, a value close to 0 indicates that the object is far from the camera, situated in the background. If two objects share the same Object Distance from the Lens value, it suggests they are at the same depth or layer within the image. For instance, if one object has a value of 0.12 and another is at 0.05, these small differences suggest that both objects are likely in the background, even if not exactly at the same depth but relatively close. This helps in understanding the spatial arrangement of elements in the image and how they relate to the viewer's perspective.</li> <li><b>3. <i>Relative Size Proportion in Images (Percentage)</i>:</b> It can tell you the approximate proportion of the object within the image. If the proportion is particularly small or large, it should be emphasized. If the proportion is moderate, it does not need to be highlighted.</li> </ol> <p><b>###THINGS TO REMEMBER###</b></p> <ol style="list-style-type: none"> <li>1. Through the extra information of different Objects, some Objects may represent the same thing. When adding Objects to the Original Description, it is important to avoid duplication.</li> <li>2. The photographic characteristics mentioned in the Original Description should be preserved, such as sizes, locations, camera angles, depths;</li> <li>3. The <i>Relative Spatial Positioning</i>, <i>Relative Distance from the Lens</i>, and <i>Relative Size Proportion in Images (Percentage)</i> of each Object need to be emphasized. However, these details must be expressed without directly using specific values of the extra information, you must convey this information naturally through logical reasoning.</li> </ol> <p><b>###IN-CONTEXT EXAMPLES###</b></p> <p><i>[Chain of thought is placed within a pair of "@@@" (remember only in the Examples will you be provided with a chain of thoughts to help you understand; in the actual task, these will not be given to you.)]</i></p> <p><b>&lt;IN-CONTEXT EXAMPLES&gt;</b></p> <p><b>###TASK###</b></p> <p><i>[**You only need to provide the modified description directly after "%%%Your Modified Description:%%%" that I provided**.]</i></p> <p><b>%% %The Original Description:%% %</b></p> <p><b>&lt;FINE-GRAINED OBJECTS' ANNOTATIONS&gt;</b></p> <p><b>%% %Your Modified Description:%% %</b></p>

Table 3: Recaptioning Prompt for the last phase "Textualized Recaptioning". **<IN-CONTEXT EXAMPLES>** is the placeholder for several in-context examples that illustrated in Table 4, 5. **<FINE-GRAINED OBJECTS' ANNOTATIONS>** will be replaced by the visual detailed textualization created during the second phase.

### In-Context Example of Recaptioning Prompt

**%% % The Original Description: % % %** At the center of the frame is a **black Rolex clock** mounted on a **black pole**. The clock has a **white face** with **black hands**, indicating the time. Behind the clock, there are a **brown tree trunk** with a rough texture and a traffic light. Further back, there's a **white building** with a **red tile roof**, possibly a hotel as indicated by the sign that reads **HOTEL**. In front of the building, there's a **white car** parked, perhaps belonging to one of the hotel guests. Beyond this immediate scene, there's a street with a **white crosswalk**, where a bus is parked.

Hallucinations: a traffic light; a bus

@ @ @

1. In The Original Description, there is a sentence "Behind the clock, there are a brown tree trunk with a rough texture and a traffic light". Since "a traffic light" is a hallucination, it should be removed, then the sentence should be "Behind the clock, there are a brown tree trunk with a rough texture".
2. In The Original Description, "there's a street with a **white crosswalk**, where a bus is parked" should be modified to "there's a street with a **white crosswalk**", since "a bus" is a hallucination that needs to be deleted.

@ @ @

Object1: the clock face is white

Relative Spatial Positioning: [0.23, 0.06, 0.55, 0.31]

Distance from the Lens: 1.0

Relative Size Proportion in Images (Percentage): 5.58

@ @ @

1. The original description already mentions "The clock has a **white face** with **black hands**". Since the object detail is already included, no modification is necessary for this object.

@ @ @

...

Object5: a person walking on the sidewalk

Relative Spatial Positioning: [0.98, 0.57, 1.0, 0.6]

Distance from the Lens: 0.0

Relative Size Proportion in Images (Percentage): 0.05

@ @ @

1. Based on the Object Relative Spatial Positioning [0.98, 0.57, 1.0, 0.6], it can be inferred that "a person walking on the sidewalk" is positioned in the middle of the right side of the image.
2. Because the original description includes the line "Beyond this immediate scene, there's a street with a white crosswalk", we can logically associate "a person walking on the sidewalk" with this statement as a supplemental detail.
3. Furthermore, because the Object Distance from the Lens is 0.0, and the Relative Size Proportion of Objects in Images is 0.05, it indicates that "a person walking on the sidewalk" is positioned very far from the camera and is very small. Therefore, we can modify the original sentence "Beyond this immediate scene, there's a street with a white crosswalk." to "Beyond this immediate scene, there's a street with a white crosswalk. At a distance from the camera, you can see a person walking on the sidewalk, occupying a very small part of the entire image."

@ @ @

**%% % Your Modified Description: % % %** At the center of the frame is a **black Rolex clock** mounted on a **black pole**. The clock has a **white face** with **black hands**, indicating the time. the clock, there's a **brown tree trunk** with a rough texture. Further back, there's a **white building** with a **red tile roof**, possibly a hotel as indicated by the sign that reads **HOTEL**. front of the building, there's a white car parked, perhaps belonging to one of the hotel guests. Closer to the camera, there is a parked silver car near a stop sign. Beyond this immediate scene, there's a street with a white crosswalk. At a distance from the camera, you can see a person walking on the sidewalk, occupying a very small part of the entire image.

Table 4: One of the in-context example of the recaptioning prompt.

### In-Context Example of Recaptioning Prompt

%% % **The Original Description:** %% % In the heart of a winter wonderland, a lone skier, clad in a vibrant orange jacket, carves their way down a pristine, snow-covered slope.

---

Hallucinations:

@ @ @

1. There are no hallucinations.

@ @ @

Object1: a man in a red jacket

Relative Spatial Positioning: [0.8, 0.76, 0.87, 0.93]

Distance from the Lens: 0.96

Relative Size Proportion in Images (Percentage): 0.5

@ @ @

1. In the original description, it mentions "a lone skier, clad in a vibrant orange jacket." Due to the similarity between an orange jacket and a red jacket, and based on "a lone skier," indicating only one person, the detail "a man in a red jacket" can be disregarded.

@ @ @

Object2: a backpack on the skier's back

Relative Spatial Positioning: [0.81, 0.76, 0.83, 0.81]

Distance from the Lens: 0.92

Relative Size Proportion in Images (Percentage): 0.04

@ @ @

1. Because the Object Relative Spatial Positioning and Object Distance from the Lens of this object are very similar to Object1's, "a backpack on the skier's back" most likely refers to "a lone skier, clad in a vibrant orange jacket." Therefore, the detail of "a backpack" should be added, resulting in "a lone skier, clad in a vibrant orange jacket, with a backpack on his back."

@ @ @

---

%% % **Your Modified Description:** %% % In the heart of a winter wonderland, a lone skier with a backpack on his back, clad in a vibrant orange jacket, carves their way down a pristine, snow-covered slope.

Table 5: One of the in-context example of the recaptioning prompt.

## MLLM-generated Description vs IT-Description

### Original Description:

This is a well-organized kitchen with a clean, modern aesthetic. The kitchen features a [white countertop](#) against a [white wall](#), creating a bright and airy atmosphere.

On the countertop, you can see a variety of appliances and items. There's a [sleek coffee machine](#), ready to brew a fresh cup. Next to it is a [paper towel holder](#), standing tall and at the ready for any spills. A [vase](#) adds a touch of elegance to the space, while a [blender](#) suggests the possibility of smoothies or soups being made here. [Various bottles and jars](#) are also present, perhaps containing spices or cooking ingredients.

The objects are mostly colored in shades of white, black, and silver, complementing the modern look of the kitchen. However, there are also pops of color with some yellow and teal objects, adding a bit of cheerfulness to the space.

Above the countertop, shelves hold additional items. The arrangement is neat and everything appears to have its own place.

In the foreground of the image is a [wooden chair](#), perhaps providing a spot to sit while waiting for that coffee to brew or the blender to finish its job. In the background, there's a [window](#) letting in natural light.

Overall, this kitchen is not only functional with its various appliances and ample storage, but also stylish with its color scheme and neat arrangement.



### Modified Description:

This is a well-organized kitchen with a clean, modern aesthetic, featuring a [white countertop](#) against a [white wall](#), creating a bright and airy atmosphere.

On the countertop, an array of appliances and items can be seen. There's a [sleek coffee machine](#) ready to brew a fresh cup of coffee, positioned next to a [sturdy paper towel holder](#). A [vase](#) adds a touch of elegance, while a [blender](#) hints at potential smoothies or soups. [Various bottles and jars](#) are also present, suggesting a collection of spices and cooking ingredients. Next to the countertop, there's a [white refrigerator](#) in the kitchen, occupying a significant portion of the space. It stands out with its pristine white finish. Nearby, a [white microwave oven](#) sits, ready for quick heating or cooking tasks.

Above the countertop, shelves are meticulously arranged with an assortment of items that blend practicality and aesthetic appeal. Here, [bottles of wine](#) are aligned next to [several clear and stemmed wine glasses](#) that catch the light beautifully. A [blue mug](#) and [multiple white coffee mugs](#), additional [clear glasses](#) provide contrast and color to the arrangement. Also, there's a variety of cans including a [blue](#), a [white can](#) and a [can of beer](#), alongside a [silver can](#), contributing to the diversity of the kitchen's storage contents. The kitchen's color palette mostly features shades of white, black, and silver, complementing its modern design. However, pops of color from [the blue pot on the stove](#), and a [small green plant](#) near the cooking area add vibrancy and life to the space.

In the foreground, a [brown wooden table](#) and a [wooden chair with rods](#) offer a cozy spot to sit and savor the culinary delights from this well-appointed kitchen. Nearby, a [brown ceramic cup](#) suggests recent use or a forthcoming coffee break. The background is softly illuminated by natural light streaming through a [window](#), enhancing the overall ambiance and providing a welcoming atmosphere.

Overall, this kitchen not only functions effectively with its various appliances and ample storage but also stands out with its stylish color scheme and orderly arrangement, making it a visually appealing and highly practical space for culinary activities.

Table 6: Visualization of the original description and the modified description.



## MLLM-generated Description vs IT-Description

### Original Description:

The image depicts a well-organized kitchen scene. The primary focus is a **white electric range and oven**. The range features **four burners** and a **digital clock on the back panel**. The oven below has a **window and a handle on its door**. To the right of the range, a **set of knives** is neatly arranged, hanging on the wall. Adjacent to it on the counter, there's a **cookbook** open, perhaps suggesting someone is preparing to cook a meal. On the left side of the range, there's a **black telephone** mounted on the wall, adding a touch of vintage charm to the setting. The backdrop for these items is a **white tile backsplash**, providing a clean and bright atmosphere to the kitchen. The image also contains an OCR text reading "03 22 2010 13 21", but its relevance to the kitchen scene is not clear from the image alone. Overall, this kitchen scene suggests preparation for cooking, with tools and appliances neatly arranged and ready for use.



### Modified Description:

The image depicts a well-organized kitchen scene. The primary focus is a **white electric range and oven**. The range features **four burners** and a **digital clock on the back panel**. The oven below has a **window and a handle on its door**. To the right of the range, a series of knives with black handles are prominently displayed, some hanging on a magnetic strip and others in a cabinet, suggesting a readiness for culinary activities. To the right of the oven, the counter hosts a variety of items: **an open cookbook**, suggesting an ongoing meal preparation, and **several books**, one of which is a **prominent red book**. Nearby, a **box of white paper** is also placed, indicating the kitchen's multipurpose use. On the left side of the range, there is a **spoon**. A **black microwave** is conveniently positioned on the counter and a **jar** finds its place on a shelf, contributing to the organized look. The backdrop features a **white tile backsplash**, enhancing the kitchen's bright and clean atmosphere. The entire setting is punctuated by the presence of utensils and books, each placed thoughtfully to suggest both utility and comfort in the kitchen space. The image also contains an OCR text reading "03 22 2010 13 21", but its relevance to the kitchen scene is not clear from the image alone. Overall, this image portrays a kitchen ready for both cooking and everyday family life, with every item neatly arranged.

Table 7: Visualization of the original description and the modified description.

## MLLM-generated Description vs IT-Description

### Original Description:

The image presents a striking contrast between two distinct bathroom designs.

On the left, we see a bathroom that exudes a cool, calming aura with its **blue upper wall and green lower wall**. The toilet, in harmony with the lower wall, is also green. A **white towel**, neatly hung on the wall, adds a touch of cleanliness and order to the space.

On the right, we are introduced to a bathroom that embraces a warmer, more vibrant color scheme. The wall is painted a lively shade of orange and is adorned with a **tree decal**, adding an element of nature to the indoor space. A **black cabinet** and a **white toilet** form the main fixtures in this bathroom. Above the cabinet, a **black shelf** holds **jars** and a **plant**, further enhancing the natural theme. A **window** opens up to a view of **trees**, connecting the indoors with the outdoors and creating an open, airy atmosphere. In both bathrooms, despite their contrasting designs and color schemes, there is an evident emphasis on cleanliness, orderliness, and harmony with nature. Each design offers its own unique interpretation of comfort and style.



### Modified Description:

The image vividly captures two contrasting bathroom interiors, each with its unique design elements and functional fixtures that reflect distinct aesthetics of comfort and style.

On the left side, the bathroom showcases a soothing atmosphere with its **upper walls painted in a tranquil blue and the lower walls in a refreshing green**. The focal point is a **light green toilet**, perfectly blending with the lower wall, and equipped with a **light green toilet seat**, emphasizing the cool color palette. Nearby, a **towel on a rack** hints at meticulous upkeep and functionality. Additional touches include a **box of tissues** and a **roll of toilet paper**, conveniently placed, ensuring everything is within reach. A **red and white bag** hangs from a hook, injecting a pop of color. A **window** in the bathroom opens to an outdoor view, bridging the interior with the natural world outside. Additionally, a **silver trash can**, subtly positioned near the toilet, further speaks to the thoughtful arrangement of this space.

Transitioning to the right, the bathroom adopts a warmer and more vibrant decor with its lively yellow walls adorned with a **tree decal**, creating an inviting natural ambiance. A **white toilet** anchors this space, complemented by a **dark brown cabinet** mounted on the wall, which hosts a **clear jar with a lid**, **jars with white napkins**, totally **three jars on a shelf**, all contributing to the room's warm and cozy feel. A **small white towel** hangs on a towel rack above a **silver metal bathroom trash can**, combining practicality with style. A **window dressed with white window blinds** offers a view of the outdoors, linking the room with nature and allowing natural light to brighten the space. Both bathrooms, through their deliberate use of colors, fixtures, and decorative elements, emphasize cleanliness, orderliness, and a harmonious integration with nature. The thoughtful placement of everyday items like toiletries on a shelf and a silver trash can ensures that both practicality and aesthetics are beautifully balanced, offering a comforting and stylish environment.

Table 8: Visualization of the original description and the modified description.