Table 1: The List of all Compound Color Names

Catagory	Color Term (Color Code)		
Blended Color:	Orange Red (255, 69, 0)	Orange Yellow (245, 189, 31)	Red Orange (255, 83, 73)
	Red Purple (228, 0, 120)	Yellow Green (154, 205, 50)	Yellow Orange (255, 174, 66)
Modified Color:	Baby Blue(137, 207, 240)	Baby Pink(244, 194, 194)	Dark Brown(101, 67, 33)
	Dark Orange(255, 140, 0)	Dark Purple(48, 25, 52)	Dark Red (139, 0, 0)
	Deep Pink (255, 20, 147)	Hot Pink (255, 105, 180)	Light Blue (173, 216, 230)
	Light Gray (211, 211, 211)	Light Green (144, 238, 144)	Light Orange (254, 216, 177)
	Light Pink (255, 182, 193)	Light Yellow (255, 255, 224)	Middle Blue (126, 212, 230)
	Middle Green (77, 140, 87)	Middle Purple (217, 130, 181)	Middle Red (229, 142, 115)
	Middle Yellow (255, 235, 0)		
Object Color:	Pale Pink(250, 218, 221)	Blood Red (102, 0, 0)	Brick Red (203, 65, 84)
	Canary Yellow (255, 239, 0)	Ghost White (248, 248, 255)	Jungle Green (41, 171, 135)
	Lemon Yellow (255, 244, 79)	Lime Green (50, 205, 50)	Mint Green (152, 255, 152)
	Olive Green (181, 179, 92)	Orchid Pink (242, 189, 205)	Rose Red (194, 30, 86)
	Ruby Red (155, 17, 30)	Salmon Pink (255, 145, 164)	Schoolbus Yellow (255, 216, 0)
	Sea Green (46, 139, 87)	Sky Blue (135, 206, 235)	Turtle Green (138, 154, 91)
	Wood Brown (193, 154, 107)	Yinmn Blue (46, 80, 144)	Zinnwaldite Brown (44, 22, 8)
Signature Color:	Android Green (61, 220, 132)	Barbie Pink (218, 24, 132)	Cambridge Blue (163, 193, 173)
	Caribbean Green (0, 204, 153)	Carolina Blue (86, 160, 211)	Chinese Red (170, 56, 30)
	Duke Blue (0, 0, 156)	French Blue (0, 114, 187)	India Green (19, 136, 8)
	Maya Blue (115, 194, 251)	Navy Blue (0, 0, 128)	Oxford Blue (0, 33, 71)
	Tiffany Blue (10, 186, 181)	Yale Blue (0, 53, 107)	
Abstract Color:	Alice Blue (240, 248, 255)	Bakermiller Pink (255, 145, 175)	Britishracing Green (0, 66, 37)
	Cyber Yellow (255, 211, 0)	Electric Purple (191, 0, 255)	Kelly Green (76, 187, 23)
	St.Patrick's Blue (35, 41, 122)	Shocking Pink (252, 15, 192)	Tickleme Pink (252, 137, 172)

A Complete Compound Color Name Set

Tab. 1 lists all the compound color names collected for this study, along with their corresponding RGB color codes. The color terms are categorized into five types: *blended* (6), *modified* (19), *object-related* (21), *signature* (14), and *abstract* (9), resulting in a total of 69 entries.

B Detailed Prompt Setup for Semantic Disambiguation

The following are the prompts used in Semantic Disambiguation.

Color Term Analysis Prompt:

"Many color expressions in text prompts are compound color terms composed of an intensity modifier (e.g., "rose", "deep", "light") and a basic color term (e.g., "red", "blue"). These modifiers can introduce ambiguity in image generation tasks, as they may be interpreted literally by the model. For instance, in the phrase "rose red", the term "rose" is intended to modify the shade of red, but the model might mistakenly generate a rose flower instead. Given a text prompt, identify the full color term(s), the ambiguous modifier term(s) (if any), and the corresponding basic color term(s)."

Entity Analysis Prompt:

"Given the prompt: prompt and the color term: colorterm identify which token(s) in the prompt each color word is describing. Return only a list of the tokens being described, in the format: [token1, token2, ...]"

Color Code Recommendation:

"Given a scene description prompt, and the color term colorterm mentioned in that scene, return an RGB value for the color term. These RGB values should reflect how the colors would visually appear in the described context, taking into account mood, lighting, environment, and narrative tone. Return only a list of RGB values, in the format: [[r1, g1, b1], [r2, g2, b2], ...]."

"A small girl wearing a ***Barbie*** *pink* dress rocking in a tiny yellow orange chair."

"A woman with a ***orchid*** *pink* and Yale blue shirt on is cleaning the floors."

"A ***lime*** *green* and light blue dog playing in the snow together."

"Man on rocky beach playing with ***baby*** *pink* and cyber yellow dog."

"There is graffiti all over a ***schoolbus*** *yellow* cafe express sign."

Figure 1: Examples of disambiguation generated by the model, which is misled by the compound color terms.

C Limitation of Model-based Metrics

As illustrated in Fig. 1, we employ a **Vision-Language Model (VLM)**, specifically GPT-4o, to assess the quality of generated outputs. The evaluation consists of two main components. First, the VLM is tasked with verifying whether the generated image accurately aligns with the given prompt. Second, the model is asked to detect and report any ambiguous or unclear elements present in the image. This dual evaluation ensures both semantic fidelity to the prompt and clarity in visual representation. In addition, we utilize two metrics: the **CLIP Score**, which measures the global semantic similarity between the image and the prompt, and the **BLIP-VQA** Score, as used in T2I-CompBench, which evaluates the correctness of attribute binding. Finally, to assess the model's ability to handle compositional ambiguity, we measure the **frequency of unintended object appearances** caused by ambiguous color terms in prompts. For instance, in a phrase like "turtle green car," models may incorrectly generate a turtle by misinterpreting the color prefix as a standalone object. We use Grounding DINO to detect whether such prefix terms (*e.g.*, "turtle") are visually present in the generated images. Any detected instance is counted as an ambiguity error if the confidence score exceeds a threshold of 0.85, indicating that the model failed to correctly disambiguate the compositional phrase. *However, in cases involving complex color rendering, we find that model-based metrics often struggle to accurately capture color generation fidelity, occasionally misclassifying generic subjects (e.g., labeling a "person" as "Barbie"), suggesting that such metrics remain insufficient for evaluating fine-grained color understanding.*