TKG-DM: Training-free Chroma Key Content Generation Diffusion Model

Supplementary Material

This supplementary material provides additional analysis and results that complement the main submission. Specifically, we delve deeper into the mechanisms and benefits of our proposed Training-Free Chroma Key Content Generation Diffusion Model (TKG-DM). We explore key aspects such as the positive-to-negative ratio in the channel mean shift, the denoising process, the impact of classifier-free guidance, and the effects of negative prompts. All content here is based on the methodology outlined in the main manuscript.

9. Additional Analysis

To further illustrate the capabilities of TKG-DM in generating chroma key content, we present comprehensive visualizations and analyses. These include examinations of the positive-to-negative ratio, detailed observations of the denoising process, an exploration of classifier-free guidance effects, and an assessment of negative prompts.

9.1. Positive-to-Negative Ratio in Channel Mean Shift

To generate images with a specified background color, we introduce the channel mean shift technique as detailed in Section 3.1. This method leverages the relationship between the positive-to-negative ratio of each channel in the initial noise and the resulting color in the generated image. By adjusting this ratio, we aim to optimize the initial noise to produce the desired initial color noise. Fig. 12, 13 and Fig. 14, 15 show results for both single-channel (channel = 2) and multi-channel (channels = 2 and 3) mean shifts for blue and green backgrounds, respectively. These figures demonstrate that as the positive-to-negative ratio increases, the method shifts toward generating a monochromatic background.

For cases without a prompt, both single-channel and multi-channel mean shifts exhibit a similar trend, where the foreground image disappears when the positive-to-negative ratio reaches approximately 7%. In the single-channel scenario, increasing the mean shift results in progressively brighter tones. In contrast, the multi-channel case produces consistent monochromatic images beyond this threshold.

When a text prompt is provided, the foreground object is generated with strong semantic alignment to the prompt on the desired color background. However, if the positive-to-negative ratio is too low, the foreground aligns well with the text, but the background retains unwanted object features. Conversely, if the ratio is too high, the background becomes a solid color, but the semantic alignment of the foreground with the text deteriorates.

These results highlight the trade-off between foreground text alignment and background color uniformity. Our technique optimizes the initial noise distribution by carefully adjusting the positive-to-negative ratio, effectively balancing these factors. This enables the generation of consistent, high-quality chroma key backgrounds across various noise distributions, both single-channel and multi-channel.

9.2. Denoising Process

Denoising Step. To control the size, position, and number of foreground objects, we introduce the init noise selection strategy in Section 3.2. Fig. 19, 20 and 21 show visualizations of each denoising step for a standard SDXL, SDXL with a Green Background Prompt (GBP), and TKG-DM, respectively. As demonstrated in Fig. 20 and 21, using normal initial noise struggles to produce a stable single foreground object or generate a monochromatic background. In contrast, our noise setup enables Stable Diffusion to generate monochromatic background images across various denoising steps. Specifically, with a small denoising step value (denoising step $= 5$), our approach produces coherent semantic structures. These results indicate that our method enables high-precision chroma key image generation even with fewer denoising steps compared to normal initial noise. Further analysis at a denoising step value of 1 shows that, with the introduction of initial color noise, the model produces images with a prominent green element in the background, unlike when using normal initial noise. This demonstrates that our method effectively generates background color elements, even with minimal denoising steps.

Intermediate Step. Next, we examine the intermediate results during the denoising process when the total number of denoising steps is set to 50. For a clear comparison, we conducted this analysis using three configurations: normal SDXL, SDXL with a Green Background Prompt (GBP), and TKG-DM, as shown in Fig. 19, Fig. 20, and Fig. 21, respectively. At an early denoising step (e.g., step = 1), our method exhibits a stronger green component than the normal initial noise, indicating that our initial noise adjustments effectively guide the generation process toward a green chroma key background from the outset.

9.3. Classifier-Free Guidance

Then, we analyze the relationship between TKG-DM and Classifier-Free Guidance (CFG) by adjusting the guidance scale values. Fig. 22, Fig. 23 and Fig. 24 show results for normal SDXL, SDXL with the Green Background Prompt

(GBP), and TKG-DM, respectively. For SDXL with GBP, increasing the guidance scale enhances the prominence of the green background. In contrast, TKG-DM maintains a consistent monochromatic background, unaffected by changes in guidance scale, due to the background control provided by init color noise. This finding demonstrates that in TKG-DM, the background color is independent of the text prompt and unaffected by the guidance scale (CFG). In other words, the background color information relies solely on the init color noise, enabling the generation of monochromatic backgrounds without any dependency on the text prompt.

9.4. Negative Prompt

To further examine TKG-DM's control capabilities, we applied a negative prompt to remove green tones from the foreground. As shown in Fig. 25, we compared results for SDXL with Green Background Prompt (GBP) and TKG-DM. In SDXL with GBP, the background prompt's green setting interferes with the effectiveness of the negative prompt, limiting control over the foreground's green tones. However, in TKG-DM, where init color noise controls the background, the negative prompt selectively influences only the foreground. This demonstrates TKG-DM's superior ability to manage negative prompts for precise color control in the foreground without affecting the background.

Additionally, when generating an elephant, traditional models struggle to adjust size using either prompts or negative prompts. In contrast, as discussed in Section 6, TKG-DM allows users to adjust object size more directly, providing a significant advantage in generating objects at varying scales as needed.

10. Additional Results

10.1. Additional Results: TKG-DM with Green Chroma Key Backgrounds

Fig. 26 and Fig. 27 provide additional examples based on SD1.5 and SDXL, demonstrating TKG-DM's effectiveness in generating high-quality, chroma-keyed foreground images against a green background.

10.2. Additional Results: TKG-DM with Various Chroma Key Backgrounds

TKG-DM allows the user to control the background via the selection channel in channel mean shift (Section 4.1). Fig. 7 provides additional examples of chroma key backgrounds in various colors. This demonstrates our approach's effectiveness in generating diverse chroma key backgrounds.

10.3. Additional Results: TKG-DM with Flow-based Model

To demonstrate the generality of our approach beyond standard Stable Diffusion (SD) variants, we also applied TKG-DM to a flow-based matching model, FLUX [23]. As in our SD-based experiments, we *freeze* the FLUX model parameters and leverage the same Channel Mean Shift technique, highlighting its training-free and wide applicability. Fig. 29 presents example results produced by FLUX with a specified background color using our channel mean shift strategy.

Interestingly, FLUX encodes color information into its latent channels in a way that diverges from SD1.5 or SDXL. Channel 1 often correlates with blue and orange hues, Channel 2 predominantly captures black and white tones, Channel 3 represents pink and green relationships, and Channel 4 governs brightness factors. Although these roles contrast with the channel distributions in SD-based models, our framework remains effective because it only requires identifying which channels to shift for the desired color effect. This flexibility further illustrates that TKG-DM is not tied to any specific network architecture or parameter distribution, and can be seamlessly transferred to other generative pipelines—including flow-based, diffusion-based, or new emerging frameworks—without additional model fine-tuning. Consequently, TKG-DM holds promise for a wide range of downstream tasks where reliable chroma-key content generation is needed, regardless of the underlying generative model's channel semantics.

11. More Results of TKG-DM with application track

Beyond text-to-image tasks, TKG-DM enables chroma key content generation across various applications. The main paper highlights applications involving ControlNet, layout-aware text-to-image, consistency models, and text-to-video models. Additionally, we demonstrate applications of TKG-DM with Multi-Diffusion and text-to-3D generation.

11.1. Application to Multi-Diffusion

Multi-Diffusion [2] enables the generation of different image regions using distinct diffusion processes. By applying TKG-DM to the central portion of the image while keeping other areas under normal noise conditions, we can control which parts of the image have chroma key backgrounds. This flexibility allows precise control over both background and foreground regions, supporting seamless integration into multi-object scene generation and interactive content creation.

11.2. Application to Text-to-3D

Using SV3D [48] with TKG-DM, we address text-to-3D generation. Although a complete text-to-3D pipeline is not yet implemented, TKG-DM's background-free image generation facilitates efficient 3D object extraction without segmentation models, streamlining the creation of clean 3D assets from text prompts. Additionally, TKG-DM offers potential for conditional text-to-3D applications, similar to ControlNet, enabling enhanced control over 3D object generation based on specific textual inputs. This expands the possibility of creating customized 3D models that are aligned with user-defined criteria.

11.3. Additional Results in Application

Fig. 32 shows additional examples, including multi-diffusion and text-to-3D techniques applied without fine-tuning. We also present results where the chroma key processing was applied to foreground generation, allowing efficient and flexible background generation without additional training. These results highlight TKG-DM's versatility across multiple generative tasks.

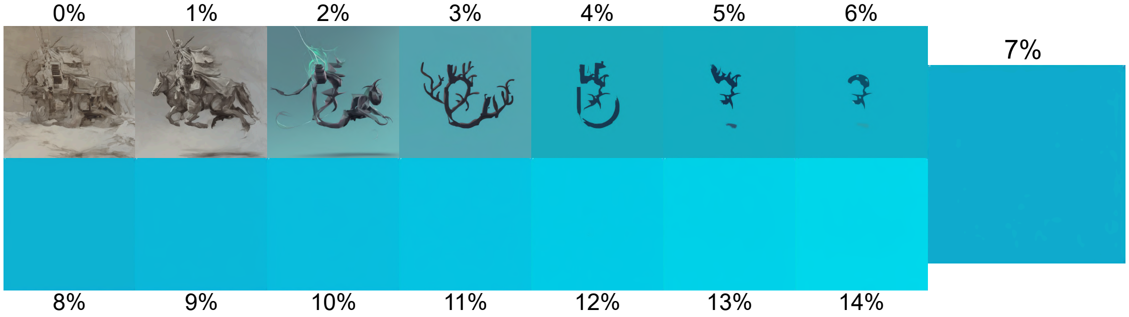

Figure 12. Relationship between the positive-to-negative ratio and single-channel mean shift (channel = 2) without a text prompt. As the ratio increases, the background shifts towards a monochromatic blue.

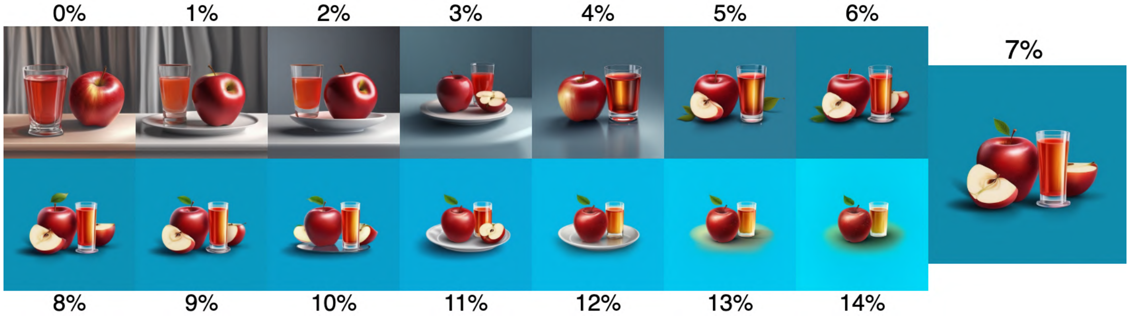

Figure 13. Relationship between the positive-to-negative ratio and single-channel mean shift (channel = 2) with the text prompt "red apple and glass of juice". Higher ratios lead to a monochromatic blue background but may degrade foreground alignment with the prompt.

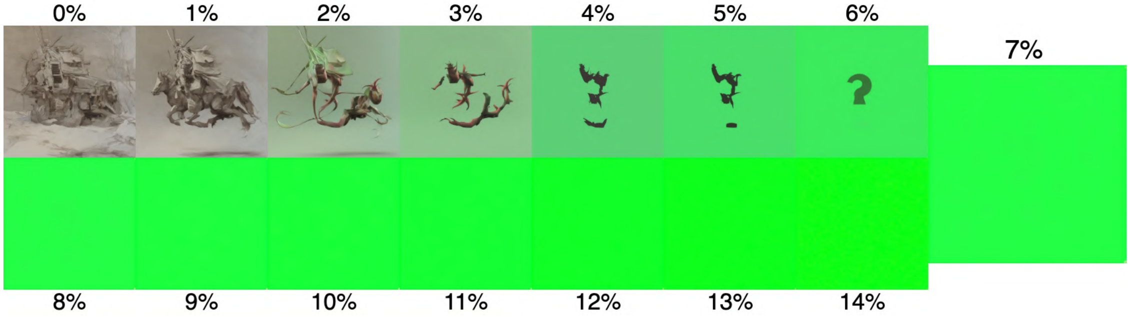

Figure 14. Relationship between the positive-to-negative ratio and multi-channel mean shift (channels = 2 and 3) without a text prompt. Increasing the ratio produces a consistent monochromatic green background.

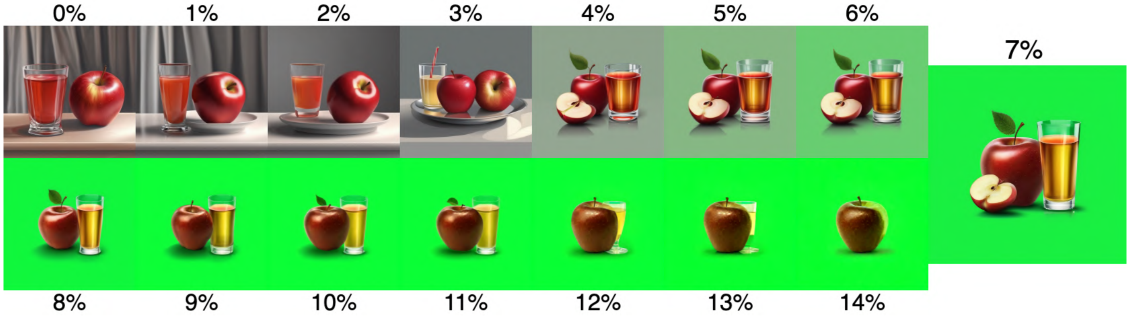

Figure 15. Relationship between the positive-to-negative ratio and multi-channel mean shift (channels = 2 and 3) with the text prompt "red apple and glass of juice". Optimal ratios balance background uniformity and foreground alignment.

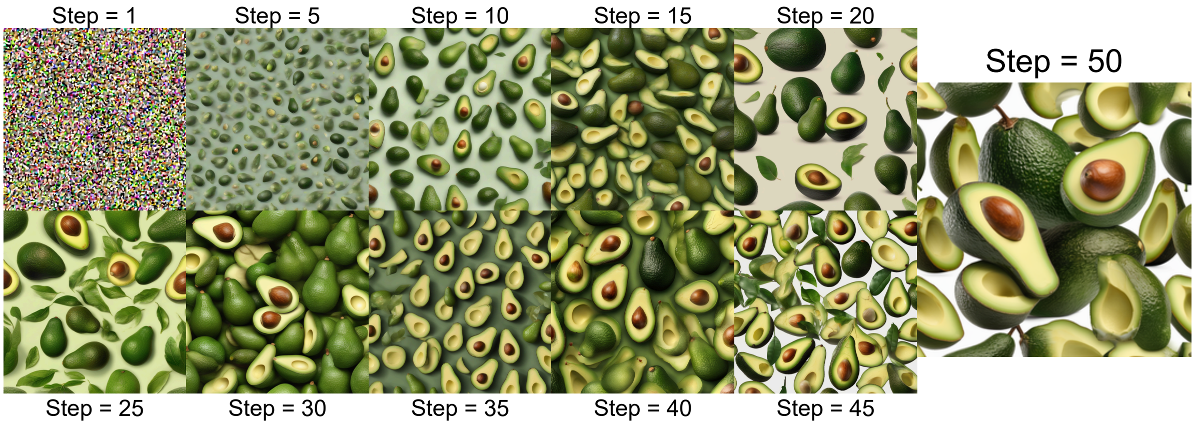

Figure 16. Generated images in SDXL at various denoising steps (1 to 50). The images demonstrate the progression of generated content. The input prompt is “An avocado”.

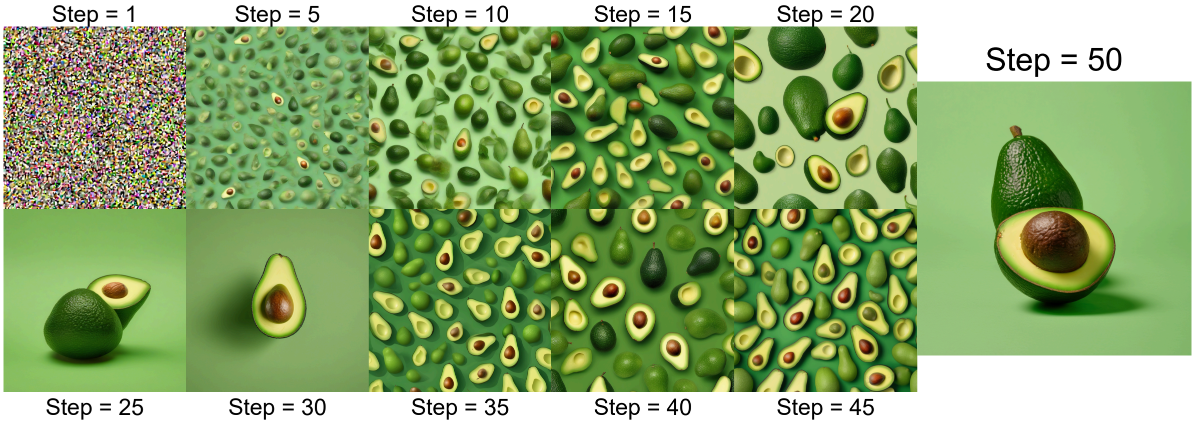

Figure 17. Generated images in SDXL with Green Background Prompt at various denoising steps (1 to 50). The input prompt is “An avocado”. The green background becomes more prominent with more denoising steps but may affect foreground quality.

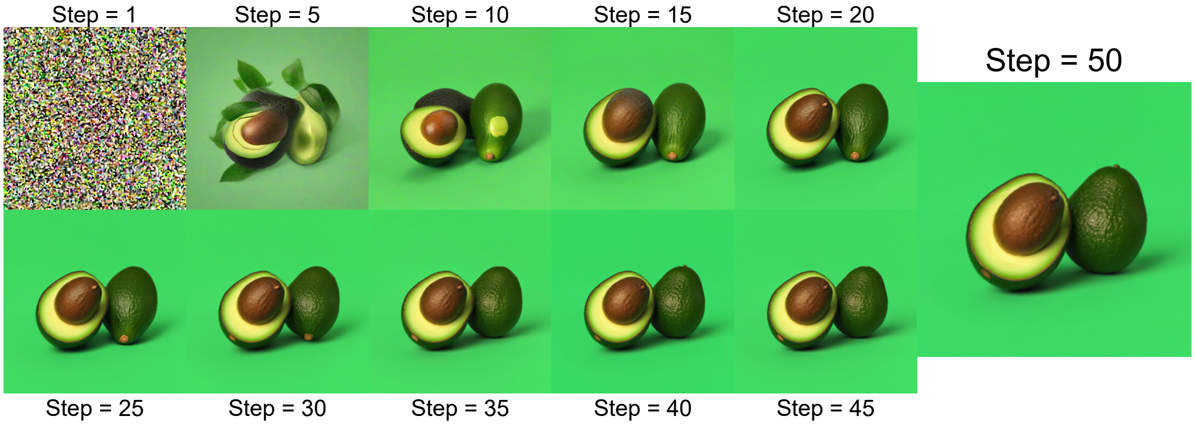

Figure 18. Generated images using TKG-DM at various denoising steps (1 to 50). The input prompt is “An avocado”. Our method maintains a consistent green background across all steps.

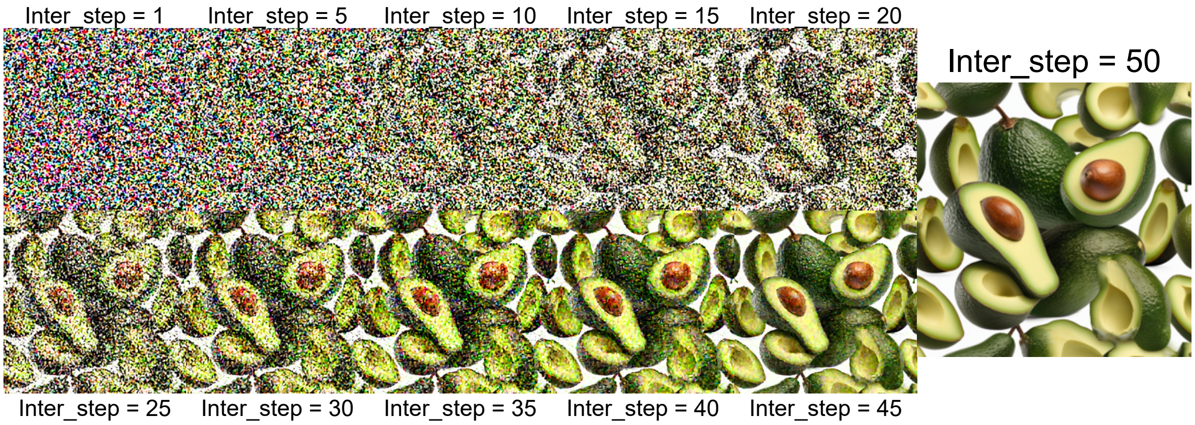

Figure 19. Denoising process progression in SDXL at various steps. The images demonstrate the evolution of generated content from initial noise (Step = 1) to the final output (Step = 50). The input prompt = “An avocado”.

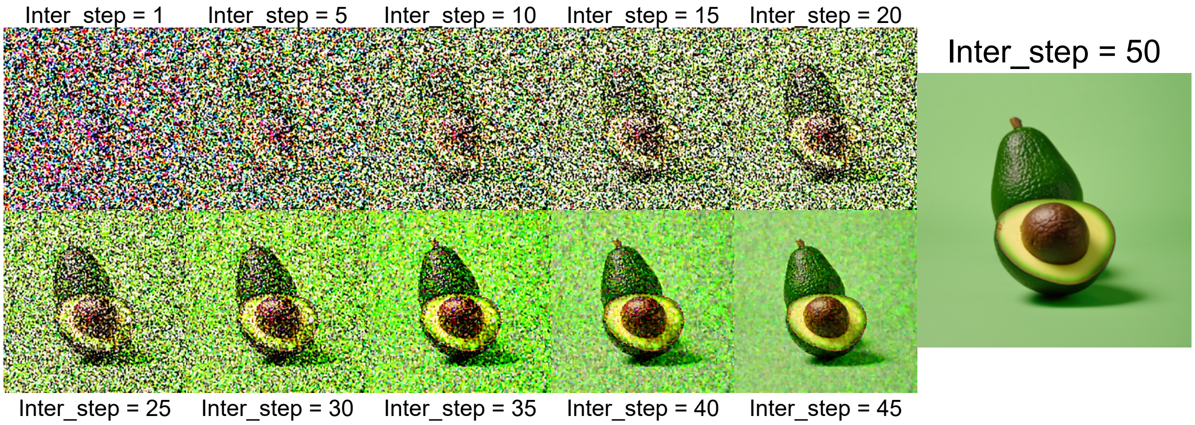

Figure 20. Denoising process progression in SDXL with Green Background Prompt at various steps. The images demonstrate the evolution of generated content from initial noise (Step = 1) to the final output (Step = 50). The input prompt = “An avocado”.

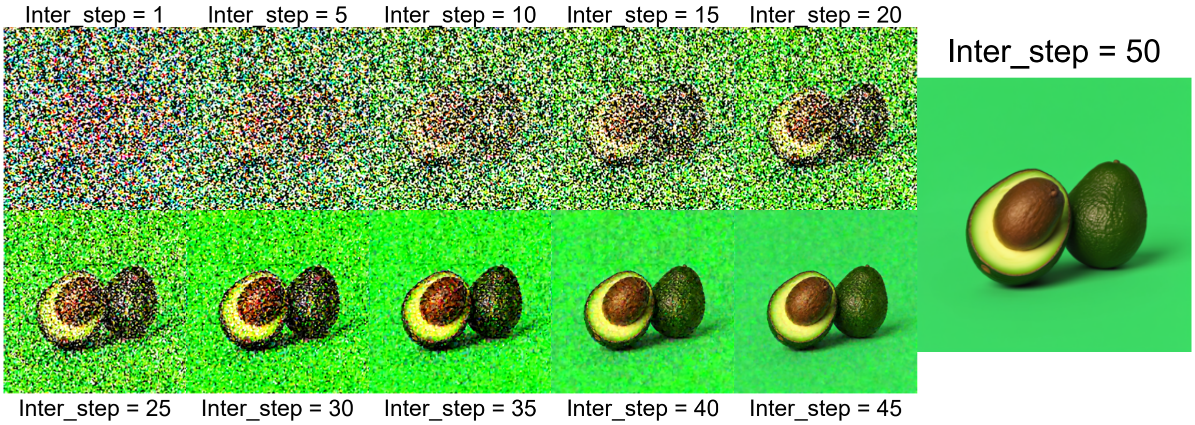

Figure 21. Denoising process progression in our TKG-DM at various steps. The images demonstrate the evolution of generated content from initial noise (Step = 1) to the final output (Step = 50). The input prompt = “An avocado”.

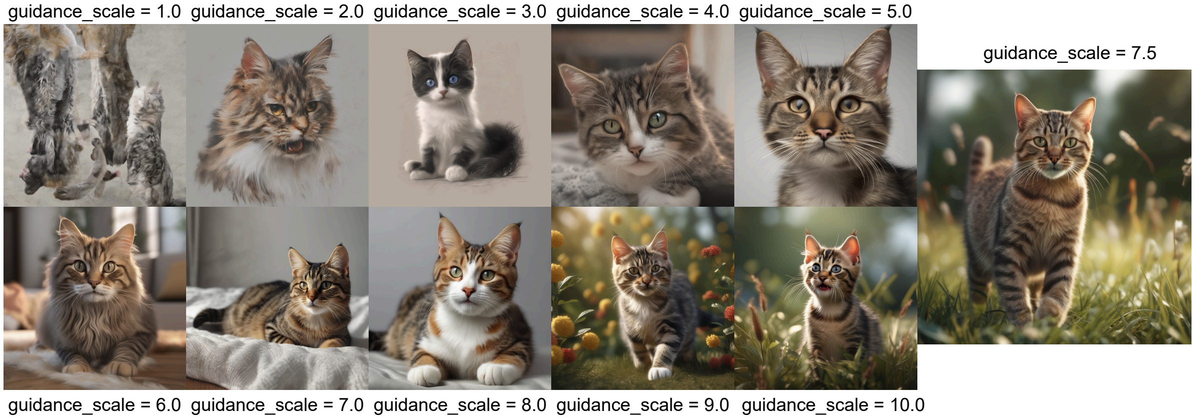

Figure 22. Effect of varying guidance scales in SDXL. The guidance scale ranges from 1.0 to 10.0. The input prompt is “The cat”. Higher guidance scales improve text alignment but may introduce background artifacts.

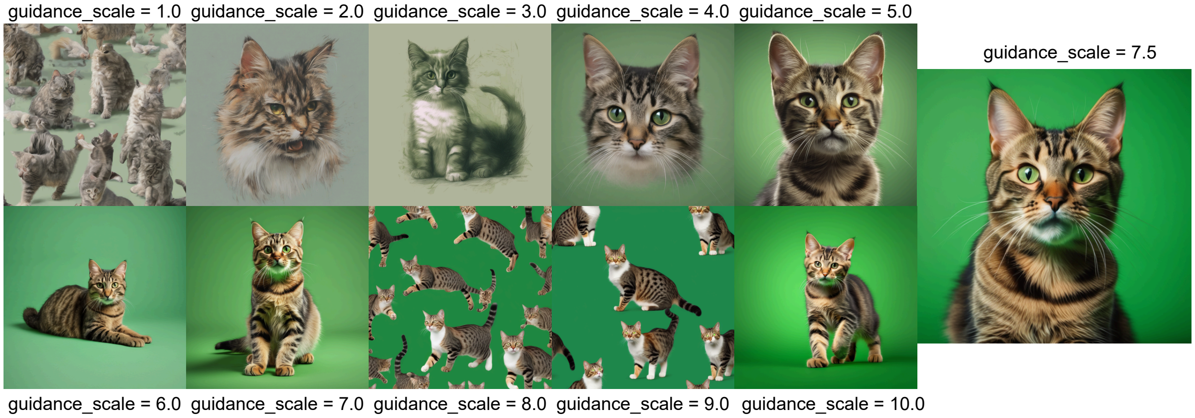

Figure 23. Effect of varying guidance scales in SDXL with Green Background Prompt. The input prompt is “The cat”. Increasing the guidance scale enhances the green background but may affect foreground details.

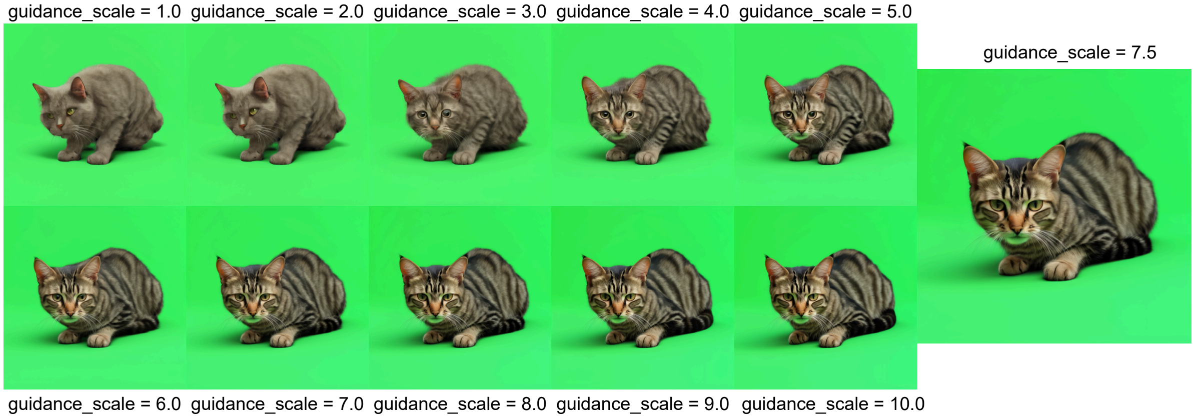

Figure 24. Effect of varying guidance scales in TKG-DM. The input prompt is “The cat”. Our method maintains consistent background and foreground quality regardless of the guidance scale.

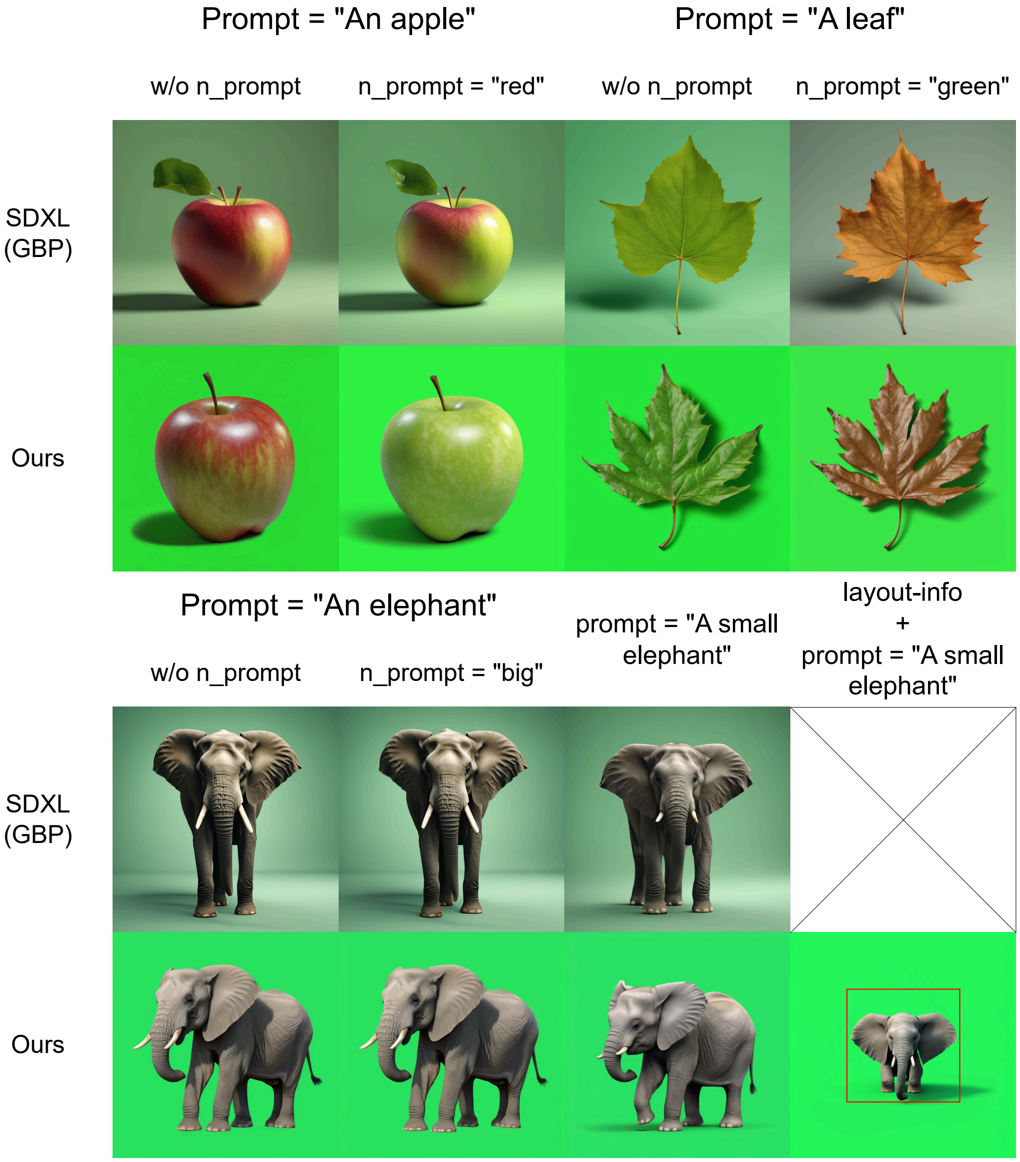

Figure 25. Comparison of results using negative prompts to modify colors and control size in generated images. For color adjustments, the negative prompt removes specific tones, such as green, in both SDXL with Green Background Prompt (GBP) and TKG-DM. In TKG-DM, init color noise enables selective color removal without affecting the background, demonstrating improved control. However, for size control, where negative prompts are generally ineffective, TKG-DM achieves size adjustments through init noise selection, providing enhanced flexibility in generating objects like an elephant at various scales.

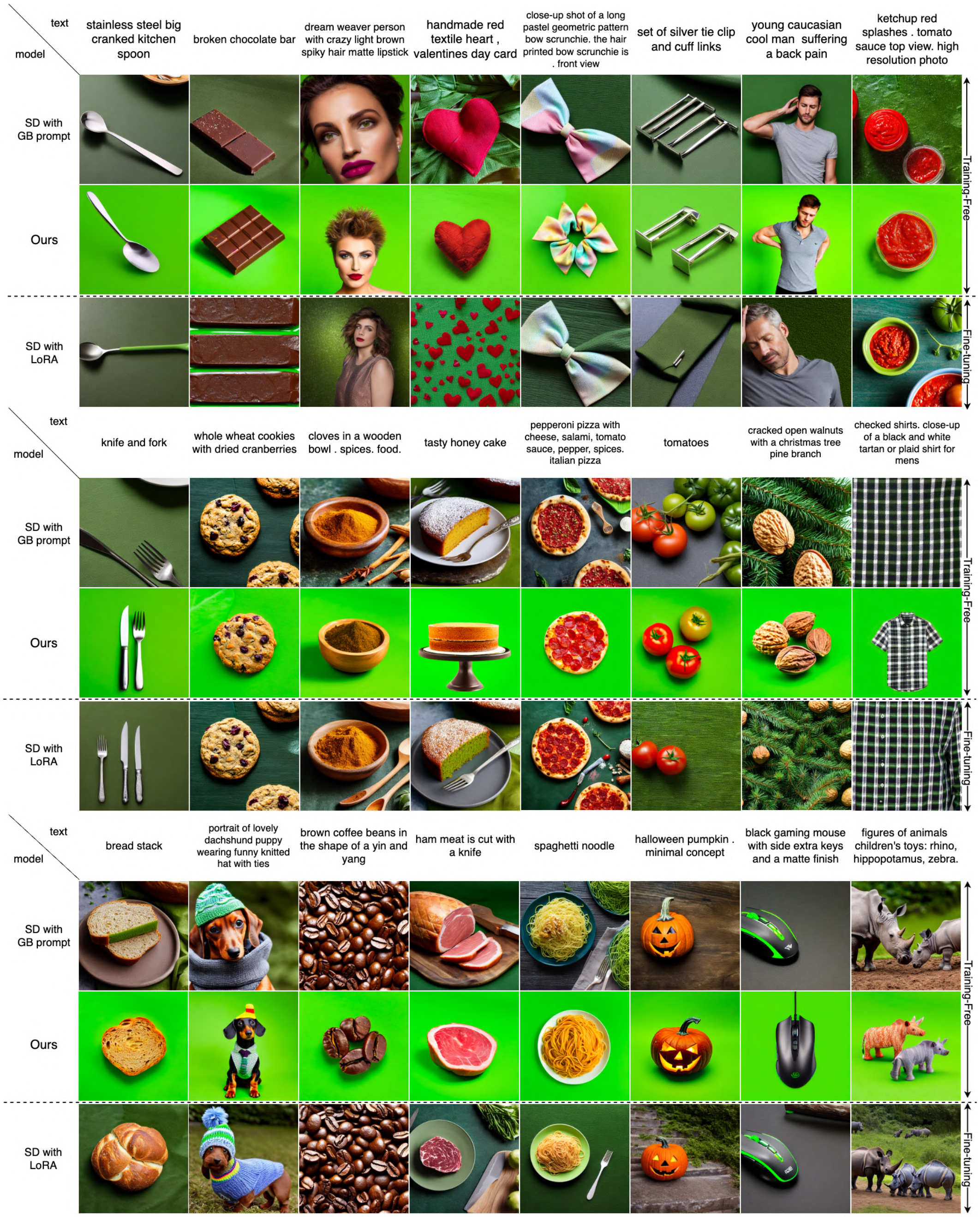
text
model
stainless steel big cranked kitchen spoon
broken chocolate bar
dream weaver person with crazy light brown spiky hair matte lipstick
handmade red textile heart , valentines day card
close-up shot of a long pastel geometric pattern bow scrunchie. the hair printed bow scrunchie is . front view
set of silver tie clip and cuff links
young caucasian cool man suffering a back pain
ketchup red splashes . tomato sauce top view. high resolution photo
SD with GB prompt
Ours
Training-Free
SD with LoRA
Fine-tuning
text
model
knife and fork
whole wheat cookies with dried cranberries
cloves in a wooden bowl . spices. food.
tasty honey cake
pepperoni pizza with cheese, salami, tomato sauce, pepper, spices. italian pizza
tomatoes
cracked open walnuts with a christmas tree pine branch
checked shirts. close-up of a black and white tartan or plaid shirt for mens
SD with GB prompt
Ours
Training-Free
SD with LoRA
Fine-tuning
text
model
bread stack
portrait of lovely dachshund puppy wearing funny knitted hat with ties
brown coffee beans in the shape of a yin and yang
ham meat is cut with a knife
spaghetti noodle
halloween pumpkin . minimal concept
black gaming mouse with side extra keys and a matte finish
figures of animals children's toys: rhino, hippopotamus, zebra.
SD with GB prompt
Ours
Training-Free
SD with LoRA
Fine-tuning

Figure 26. Additional Result in Green Background with SD1.5

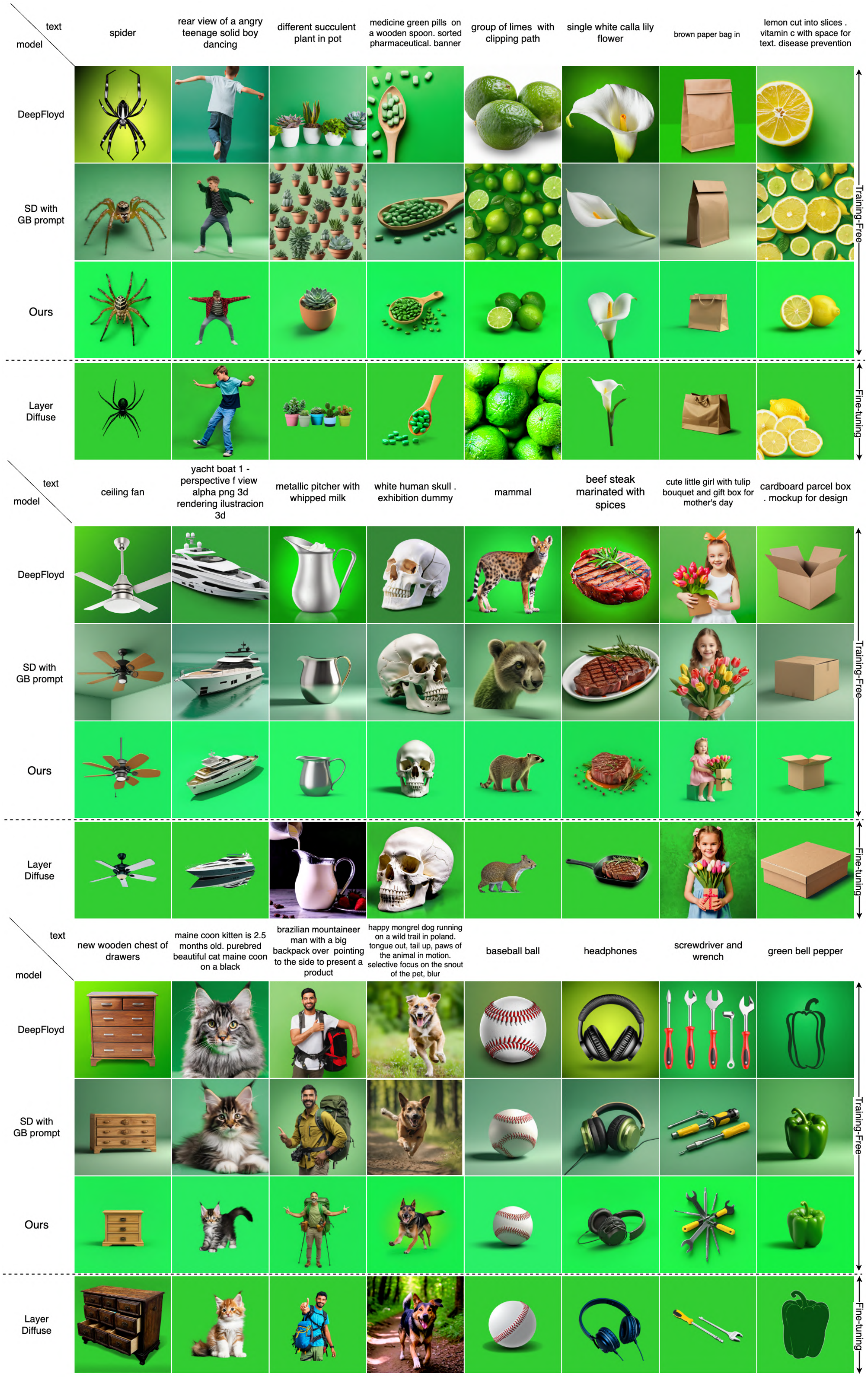
text
model
spider
rear view of a angry teenage solid boy dancing
different succulent plant in pot
medicine green pills on a wooden spoon. sorted pharmaceutical. banner
group of limes with clipping path
single white calla lily flower
brown paper bag in
lemon cut into slices . vitamin c with space for text. disease prevention
DeepFloyd
SD with GB prompt
Ours
Layer Diffuse
Training-Free
Fine-tuning
text
model
ceiling fan
yacht boat 1 - perspective f view alpha png 3d rendering ilustracion 3d
metallic pitcher with whipped milk
white human skull . exhibition dummy
mammal
beef steak marinated with spices
cute little girl with tulip bouquet and gift box for mother's day
cardboard parcel box . mockup for design
DeepFloyd
SD with GB prompt
Ours
Layer Diffuse
Training-Free
Fine-tuning
text
model
new wooden chest of drawers
maine coon kitten is 2.5 months old. purebred beautiful cat maine coon on a black
brazilian mountaineer man with a big backpack over pointing to the side to present a product
happy mongrel dog running on a wild trail in poland. tongue out, tail up, paws of the animal in motion. selective focus on the snout of the pet, blur
baseball ball
headphones
screwdriver and wrench
green bell pepper
DeepFloyd
SD with GB prompt
Ours
Layer Diffuse
Training-Free
Fine-tuning

Figure 27. Additional Result in Green Background with SDXL

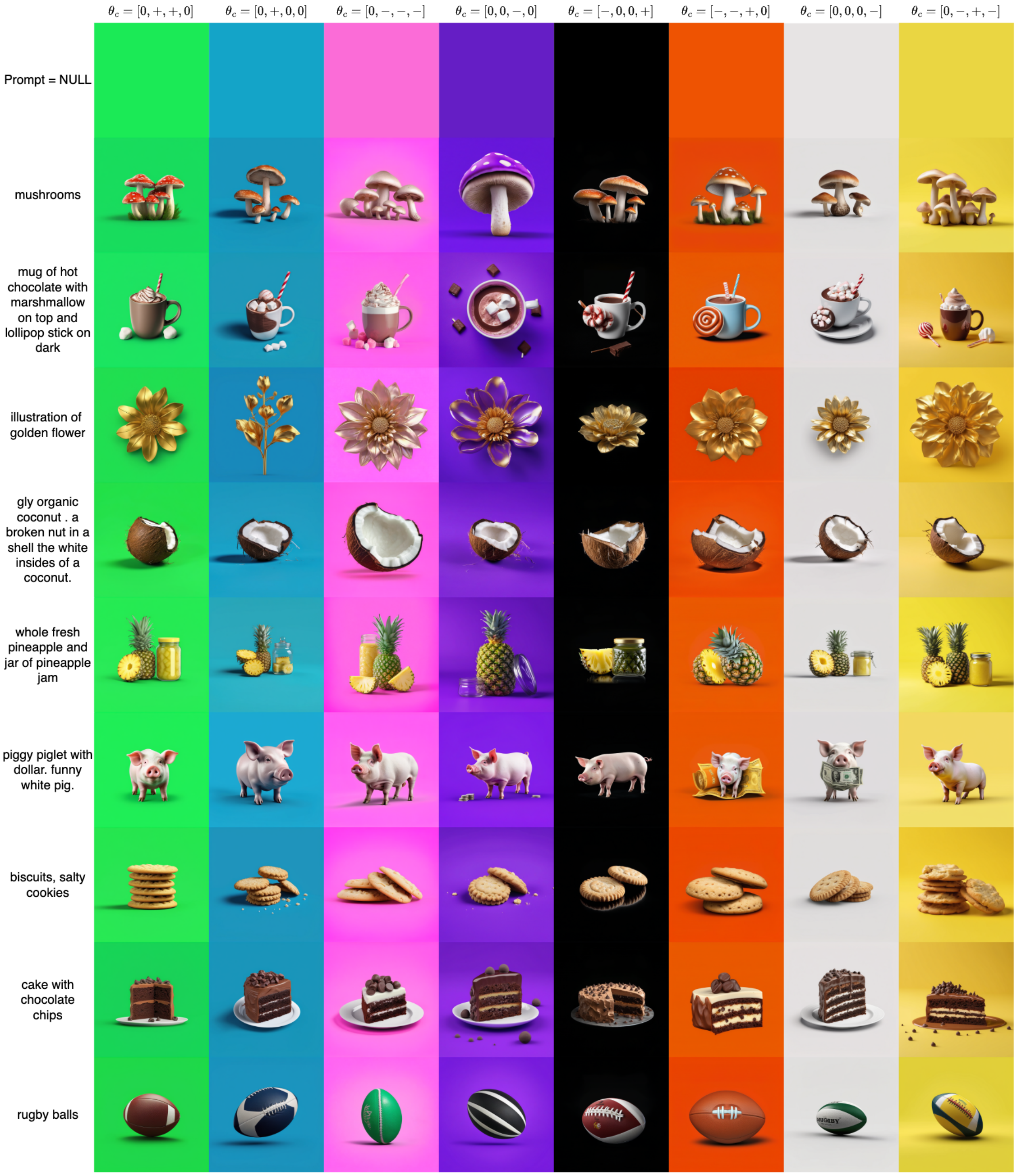
$\theta_c = [0, +, +, 0]$
$\theta_c = [0, +, 0, 0]$
$\theta_c = [0, -, -, -]$
$\theta_c = [0, 0, -, 0]$
$\theta_c = [-, 0, 0, +]$
$\theta_c = [-, -, +, 0]$
$\theta_c = [0, 0, 0, -]$
$\theta_c = [0, -, +, -]$
Prompt = NULL
mushrooms
mug of hot chocolate with marshmallow on top and lollipop stick on dark
illustration of golden flower
gly organic coconut . a broken nut in a shell the white insides of a coconut.
whole fresh pineapple and jar of pineapple jam
piggy piglet with dollar. funny white pig.
biscuits, salty cookies
cake with chocolate chips
rugby balls

Figure 28. Additional Result in various color Background with SDXL

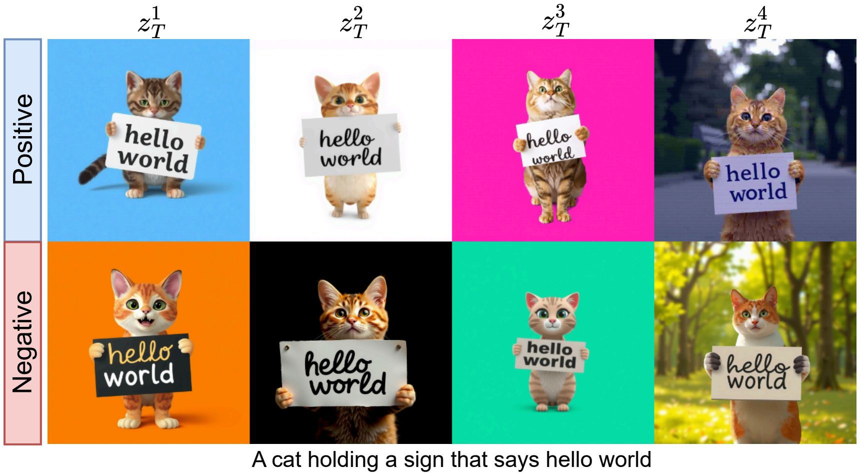

Figure 29. **TKG-DM applied to the FLUX flow-based model.** By adjusting the mean shift in specific channels, TKG-DM generates green and blue backgrounds without fine-tuning the model itself. The differences in each channel's color representation from SD1.5 and SDXL highlight TKG-DM's architecture-agnostic design.

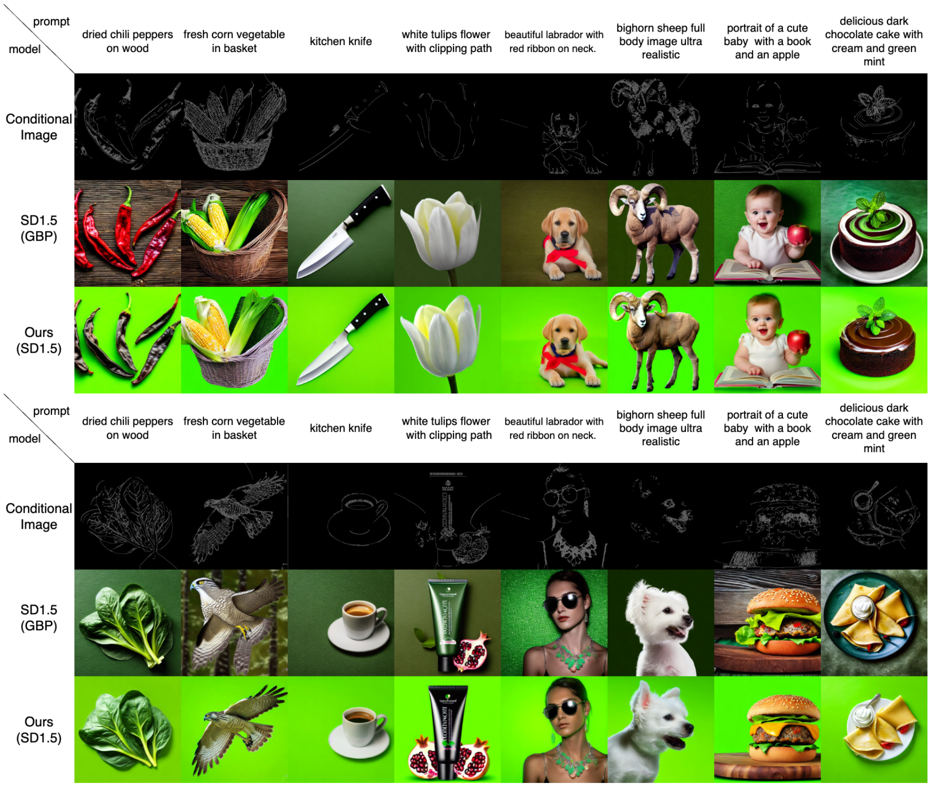
prompt
model
dried chili peppers on wood
fresh corn vegetable in basket
kitchen knife
white tulips flower with clipping path
beautiful labrador with red ribbon on neck.
bighorn sheep full body image ultra realistic
portrait of a cute baby with a book and an apple
delicious dark chocolate cake with cream and green mint
Conditional Image
SD1.5 (GBP)
Ours (SD1.5)
prompt
model
dried chili peppers on wood
fresh corn vegetable in basket
kitchen knife
white tulips flower with clipping path
beautiful labrador with red ribbon on neck.
bighorn sheep full body image ultra realistic
portrait of a cute baby with a book and an apple
delicious dark chocolate cake with cream and green mint
Conditional Image
SD1.5 (GBP)
Ours (SD1.5)

Figure 30. Additional Result in ControlNet with SD1.5

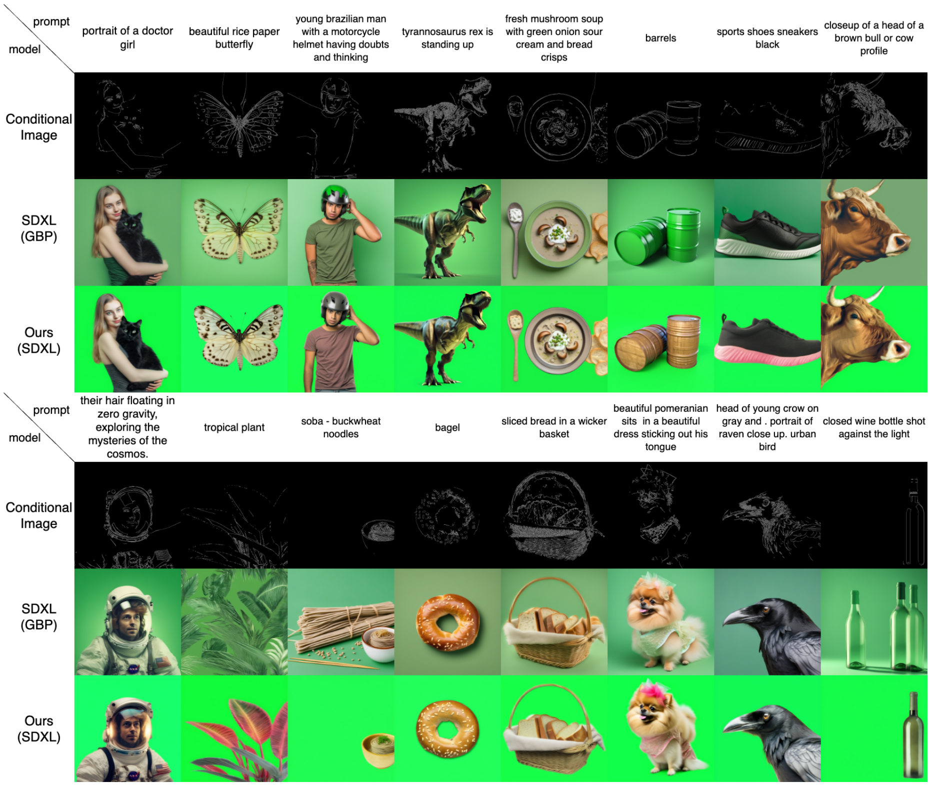
prompt
model
portrait of a doctor girl
beautiful rice paper butterfly
young brazilian man with a motorcycle helmet having doubts and thinking
tyrannosaurus rex is standing up
fresh mushroom soup with green onion sour cream and bread crisps
barrels
sports shoes sneakers black
closeup of a head of a brown bull or cow profile
Conditional Image
SDXL (GBP)
Ours (SDXL)
prompt
model
their hair floating in zero gravity, exploring the mysteries of the cosmos.
tropical plant
soba - buckwheat noodles
bagel
sliced bread in a wicker basket
beautiful pomeranian sits in a beautiful dress sticking out his tongue
head of young crow on gray and . portrait of raven close up. urban bird
closed wine bottle shot against the light
Conditional Image
SDXL (GBP)
Ours (SDXL)

Figure 31. Additional Result in ControlNet with SDXL

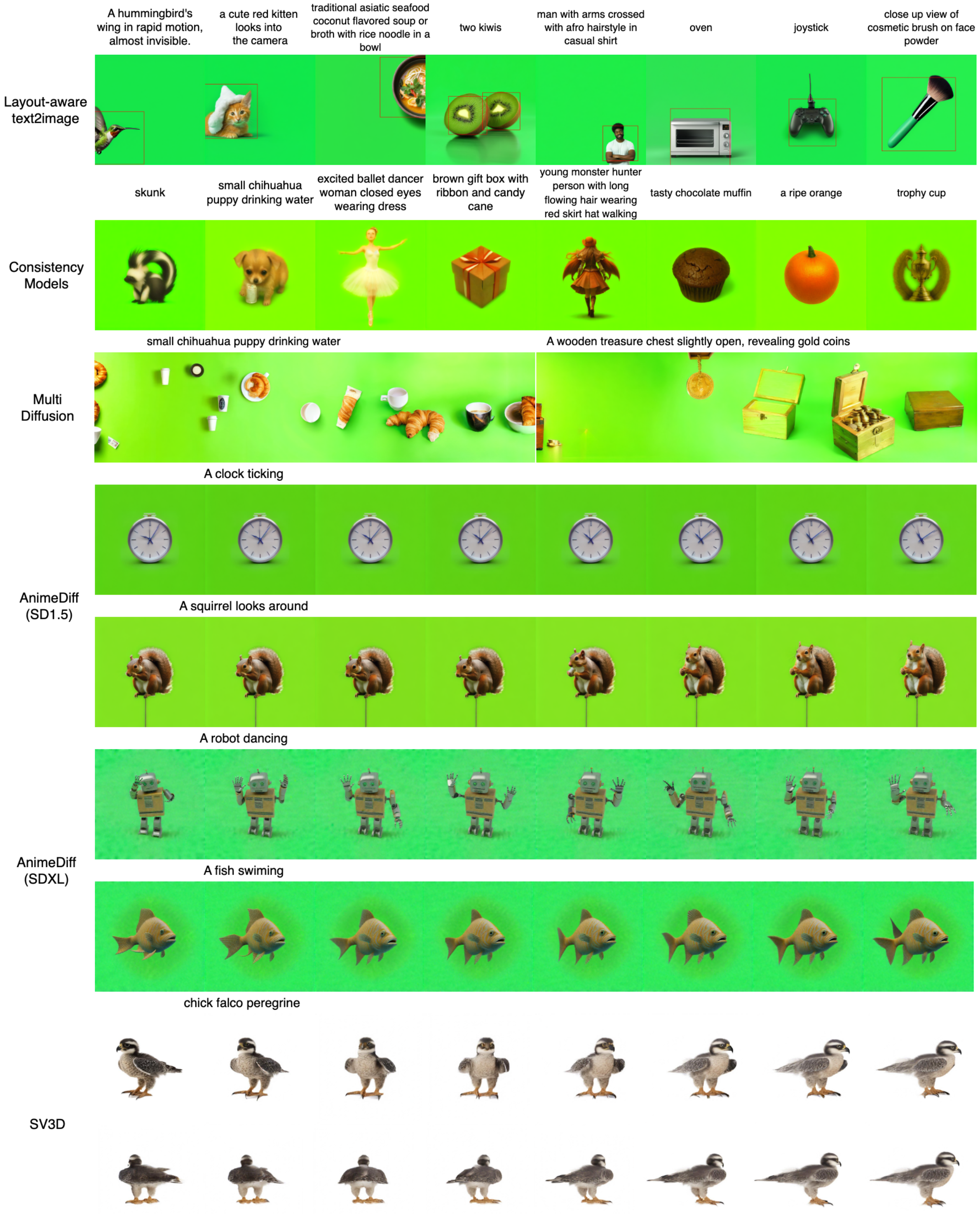
A hummingbird's wing in rapid motion, almost invisible.
a cute red kitten looks into the camera
traditional asiatic seafood coconut flavored soup or broth with rice noodle in a bowl
two kiwis
man with arms crossed with afro hairstyle in casual shirt
oven
joystick
close up view of cosmetic brush on face powder
Layout-aware text2image
skunk
small chihuahua puppy drinking water
excited ballet dancer woman closed eyes wearing dress
brown gift box with ribbon and candy cane
young monster hunter person with long flowing hair wearing red skirt hat walking
tasty chocolate muffin
a ripe orange
trophy cup
Consistency Models
small chihuahua puppy drinking water
A wooden treasure chest slightly open, revealing gold coins
Multi Diffusion
A clock ticking
A squirrel looks around
AnimeDiff (SD1.5)
A robot dancing
A fish swiming
AnimeDiff (SDXL)
chick falco peregrine
SV3D

Figure 32. Additional Result of Application track