

FlowLess: Controlling Abstract Image Generation

Supplementary Material

Amir Hertz[✉] and Noah Snavely

Google DeepMind

A Additional Comparisons

A.1 Piece-it-Together

Closely related to our work, Piece-it-Together (PiT) [7] introduced a self supervised approach that fine-tunes a diffusion model to compose input parts. However, each PiT model is trained on a specific non-realistic concept, such as imaginary products or fantasy characters; consequently, it is limited to a narrow range of input parts and output domains. Furthermore, PiT lacks control over the abstraction of the composition and the expressiveness of the output relative to the input. Figure 1 showcases FlowLess’s results on input examples from the PiT paper. As illustrated, our method successfully composes both realistic elements and abstract sketches, whereas PiT fails to generate high-quality results across diverse inputs.

A.2 Abstraction control via closed-source models

To evaluate the visual abstraction capabilities of closed-source multimodal generative models, we queried Nano-Banana Pro (NB) [2] and GPT Image 1.5 (GPTi) [6] using a subset of examples.

Completion. We query NB and GPTi on the completion examples shown in Fig. 7 of the main paper. Both models were provided with images of the input elements and the following instruction:

```
Complete the drawing in Image 1 to create <input-prompt>. (Optional) Incorporate the elements from Images 2-6 into the completion.
```

Figures 2 and 3 demonstrate the performance of NB and GPTi, respectively, on the abstract completion task. As shown, both models failed to generate the desired concept while preserving the original placement of the completion element. While both methods may incorporate input elements into the generation—for instance, the egg-like shape in the first rows—they are generally restricted to cases where the input element is semantically close to the desired prompt. For example, the blue wavy shape is only successfully utilized as decoration representing water or the sky.

Inter-relation controls. We query NB and GPTi on the inter-relation control examples from Fig. 8. in the main paper. Both models were provided with images of the input elements, detail images describing the desired inter-relations, and the following instruction:

Generate an abstract image of <input-prompt> using the materials/curves from images 1-4. Use the circled details in images 5-8 as guides for the relations between these materials. Ignore the circular outlines themselves; focus only on the depicted relationship between the elements when creating the image.

Figures 4 and 5 illustrate the performance of NB and GPTi on abstract composition. While both models produce high-quality images, they struggle to distinguish "detail" inputs from "element" inputs. For instance, the curve-curve results in Figs.4a, 4b, 5a, and 5b do not vary based on the specified inter-relations. Furthermore, the input details in Figs.4c, 4d, 5c, and 5d are often treated as additional elements rather than relational guidance. For example, appearing as circles in the pine tree or used as the tractor wheels.

B Evaluation Setup

Our full evaluation prompts and abstract sets are provided at the end of the appendix, in pages 7 and 8. All generated image results of our method and baselines are provided in the supplementary web page, under "Evaluation Set"

FlowLess Inference. All our results were generated using our fine-tuned Flux.1-dev model with 28 inference steps. We employed the Euler solver, a cosine scheduler with a log-SNR shift of $2 \log(2)$, and classifier-free guidance with a scale of 3.5. The unconditional forward pass differs from the conditional pass only by the text prompt: the abstract input sets and modulation remain identical, while the unconditional prompt is empty.

Baseline implementations. We generated all baseline results using their official implementations¹

The abstract visual elements were provided to each method either as multiple individual images (Flux.2-dev [4], Qwen-Image-Edit [10], and DreamO [5]) or as a single concatenated image (InstantStyle [9], OminiControl [8]).

For inference without a Multimodal Prompt Extension (MMPE) model, we augmented the input prompt as follows: Abstract image of "<input-prompt>" made from the different elements in Image 1, Image 2, ..., Image X., where X represents the number of elements in the input set.

When using an MMPE with Flux.2-dev, we followed the official implementation and expanded our simple evaluation prompt by querying Mistral-Small-3.2-24B-Instruct-2506 [1]. The inputs for this expansion consisted of the abstract set images alongside the following instruction: Write a caption for an abstract image of "<input-prompt>" that is composed of the elements in Image 1, Image 2, ..., Image X. Refer to the items as: Image 1, Image 2, ..., Image X..

¹ Codebases: InstantStyle, OminiControl, DreamO, Qwen-Image-Edit, and Flux.2-dev.

C Additional Results

Different backbone. In addition to our primary Flux.1-dev implementation, we applied our method to Stable diffusion 3.5 Large [3] as a backbone. Results across various abstraction sets, text prompts, and coverage scores are presented in Figures 6, 7, and 8.

Design explorations. Figures 9–27 present further generation results that demonstrate how combining art materials, line strokes, and inter-relations within the abstraction set enables controlled style exploration.

References

1. AI, M.: Mistral-small-3.2-24b-instruct-2506 (2025), <https://huggingface.co/mistralai/Mistral-Small-3.2-24B-Instruct-2506>
2. DeepMind, G.: Introducing nana banana pro (2025), <https://blog.google/innovation-and-ai/products/nano-banana-pro/>
3. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., Podell, D., Dockhorn, T., English, Z., Rombach, R.: Scaling rectified flow transformers for high-resolution image synthesis. In: Proceedings of the 41st International Conference on Machine Learning. ICML’24, JMLR.org (2024)
4. Labs, B.F.: FLUX.2: Frontier Visual Intelligence. <https://bfl.ai/blog/flux-2> (2025)
5. Mou, C., Wu, Y., Wu, W., Guo, Z., Zhang, P., Cheng, Y., Luo, Y., Ding, F., Zhang, S., Li, X., Li, M., Liu, M., Jiang, Y., Wu, S., Zhao, S., Zhang, J., He, Q., Wu, X.: Dreamo: A unified framework for image customization. In: SIGGRAPH Asia 2025 Conference Papers (2025)
6. OpenAI: The new chatgpt images is here (2025), <https://openai.com/index/new-chatgpt-images-is-here/>
7. Richardson, E., Goldberg, K., Alaluf, Y., Cohen-Or, D.: Piece it together: Part-based concepting with ip-priors. arXiv preprint arXiv:2503.10365 (2025)
8. Tan, Z., Liu, S., Yang, X., Xue, Q., Wang, X.: Ominicontrol: Minimal and universal control for diffusion transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14940–14950 (2025)
9. Wang, H., Wang, Q., Bai, X., Qin, Z., Chen, A.: Instantstyle: Free lunch towards style-preserving in text-to-image generation. arXiv preprint arXiv:2404.02733 (2024)
10. Wu, C., Li, J., Zhou, J., Lin, J., et al.: Qwen-image technical report. arXiv preprint arXiv:2508.02324 (2025)

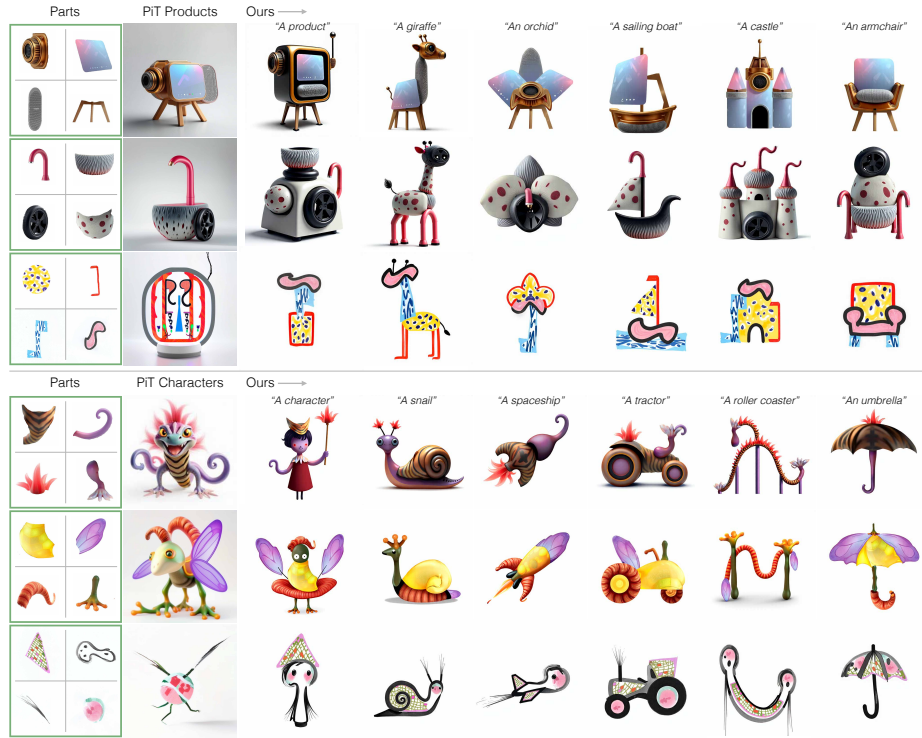


Fig. 1: Comparisons to PiT. Second column: PiT models are specialized to a specific concept, either “Products” or ‘Characters’, and struggles when input abstract elements (3rd and 6th rows). From third column: our model can generate different concepts from both realistic and abstract elements.

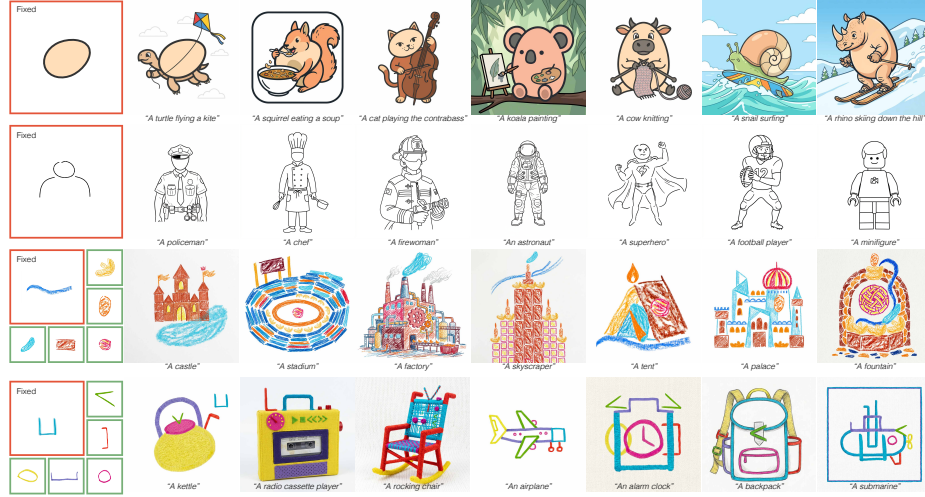


Fig. 2: Image completion by Nano-Banana Pro. Only in rare cases, NB able to complete the drawing without altering the input fixed element.



Fig. 3: Image completion by GPT Image 1.5. In some cases GPTi generates compelling images using the abstraction set, however it fails to preserve the location of input fixed shape.

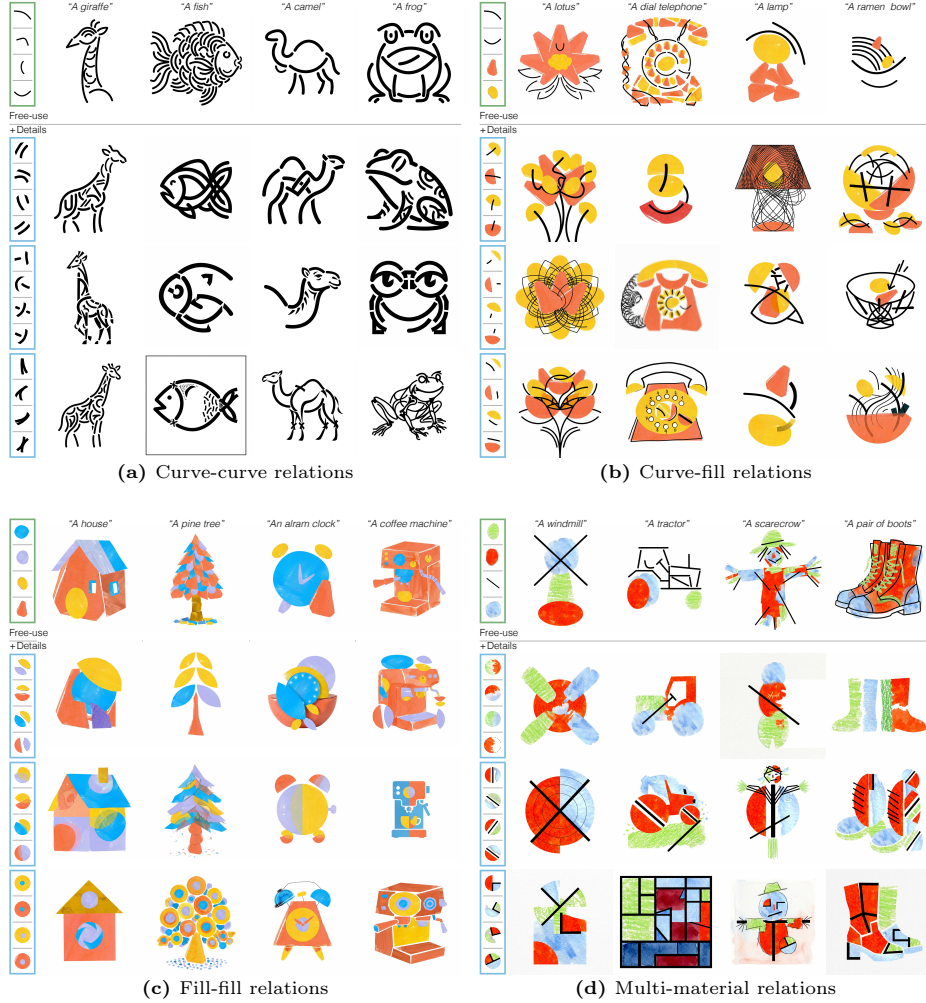


Fig. 4: Control by details with Nano-Banana Pro. Top row in each group shows generation results using the abstraction elements in the green outlines. In bottom rows: generated results using additional visual details inputs (blue outlines).

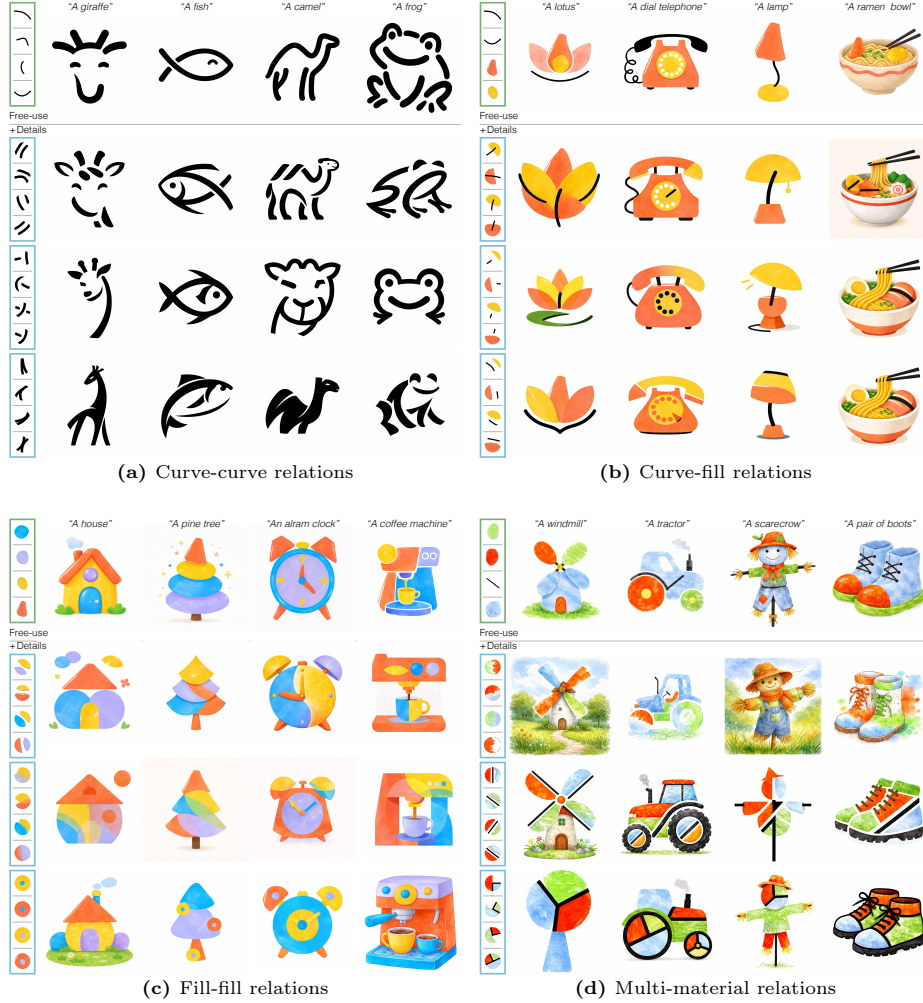


Fig. 5: Control by details with GPT Image 1.5. Top row in each group shows generation results using the abstraction elements in the green outlines. In bottom rows: generated results using additional visual details inputs (blue outlines).

Evaluation abstraction sets:

Watercolors								
Ink + water								
Markers								
Pastels								
Generated (synthetic)								
Pen								
Acrylics								
Vector graphics								
Vector graphics								
Vector graphics								

Evaluation prompts: 001. A giraffe. 002. A turtle flying a kite. 003. The great Sphinx of Giza. 004. A cup of tea and a scone. 005. An elephant rollerskating. 006. A robot. 007. A bird playing the saxophone. 008. An ice cream. 009. A toy train. 010. A daisy flower. 011. A fox hunting. 012. A rose. 013. A slice of pizza. 014. A hot air balloon. 015. A deep-sea submarine. 016. A stack of colorful books. 017. A panda eating bamboo. 018. A medieval castle. 019. A chameleon. 020. A vintage camera. 021. A serene mountain lake. 022. A compass. 023. A steaming bowl of ramen. 024. A field of sunflowers. 025. A squirrel burying an acorn. 026. A powerful lightning storm. 027. A grand lighthouse by the sea. 028. A chocolate cake. 029. A vibrant coral reef teeming with fish. 030. A classical grand piano. 031. A butterfly on a bloom. 032. A sailboat. 033. A basket of fresh berries. 034. A telescope. 035. A wise old owl. 036. A cozy camp. 037. A field of poppies. 038. A bubbling laboratory beaker. 039. A winding river through a forest. 040. A vintage gramophone. 041. A rich espresso with foam. 042. A bioluminescent deep-sea jellyfish. 043. A towering redwood forest. 044. A stack of golden pancakes with syrup. 045. A croissant. 046. A tiny ladybug on a leaf. 047. A glass of sparkling lemonade. 048. A tulip bouquet. 049. A spiral staircase in an old tower. 050. A plate of sushi rolls. 051. A quaint thatched-roof cottage. 052. A curious racoon rummaging. 053. A vintage hot rod car. 054. A blooming cherry blossom tree. 055. A shimmering seashell on a sandy beach. 056. A curious octopus in an undersea cave. 057. A vibrant passion flower. 058. A delicate porcelain teacup. 059. A medieval knight's helmet. 060. A howling wolf. 061. A vintage rotary telephone. 062. A majestic bird of paradise flower. 063. An ancient stone archway. 064. A hovering hummingbird. 065. A swift peregrine falcon diving. 066. A stack of crispy onion rings. 067. A vintage film projector. 068. A parrot on a perch. 069. A forgotten ancient map. 070. A blueberry muffin. 071. A dense bamboo forest. 072. A shimmering dragonfly. 073. A tall oak tree. 074. A flock of migrating geese. 075. A towering sand dune. 076. A collection of antique keys. 077. A blooming cactus flower. 078. A refreshing mojito. 079. A sleek modern skyscraper. 080. A delicious bowl of pasta carbonara. 081. A playful penguin sliding. 082. A graceful swan on a lake. 083. A plate of crispy spring rolls. 084. A dramatic volcanic eruption. 085. An orchid bloom. 086. A shimmering school of fish. 087. A perfectly grilled steak. 088. A lotus flower on water. 089. A winding labyrinthine cave system. 090. A graceful gazelle leaping. 091. A sleeping koala in a eucalyptus tree. 092. A rustic wooden bridge over a stream. 093. A delicious apple pie. 094. A towering cypress tree. 095. A bustling futuristic cityscape. 096. A delicate paper crane. 097. A vintage arcade game cabinet. 098. A hot geyser erupting steam. 099. A futuristic levitating car. 100. A towering saguaro cactus.

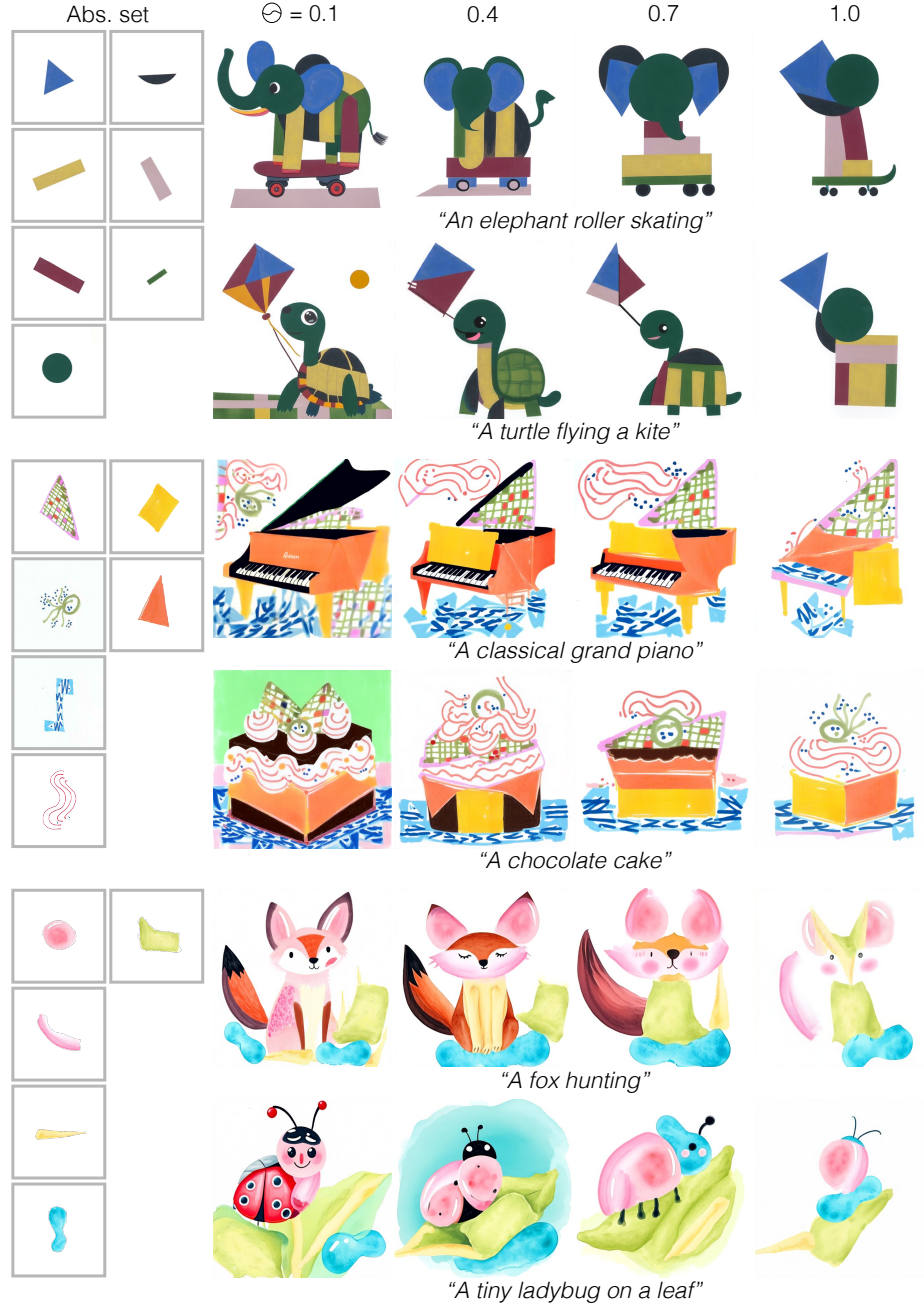


Fig. 6: FlowLess results over Stable Diffusion 3.5 Large. Left to right results were generated using increasing value of coverage score.

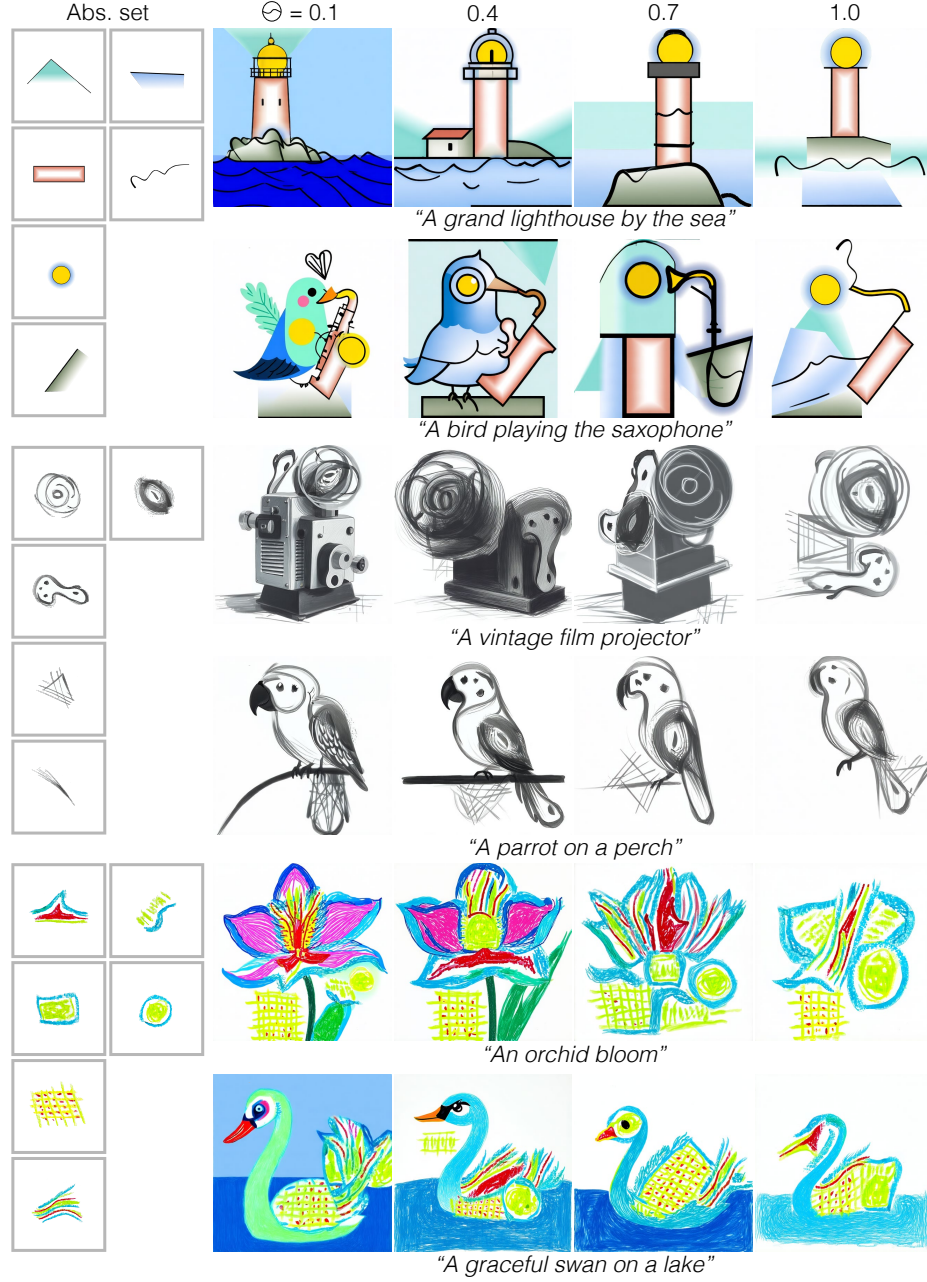


Fig. 7: FlowLess results over Stable Diffusion 3.5 Large. Left to right results were generated using increasing value of coverage score.

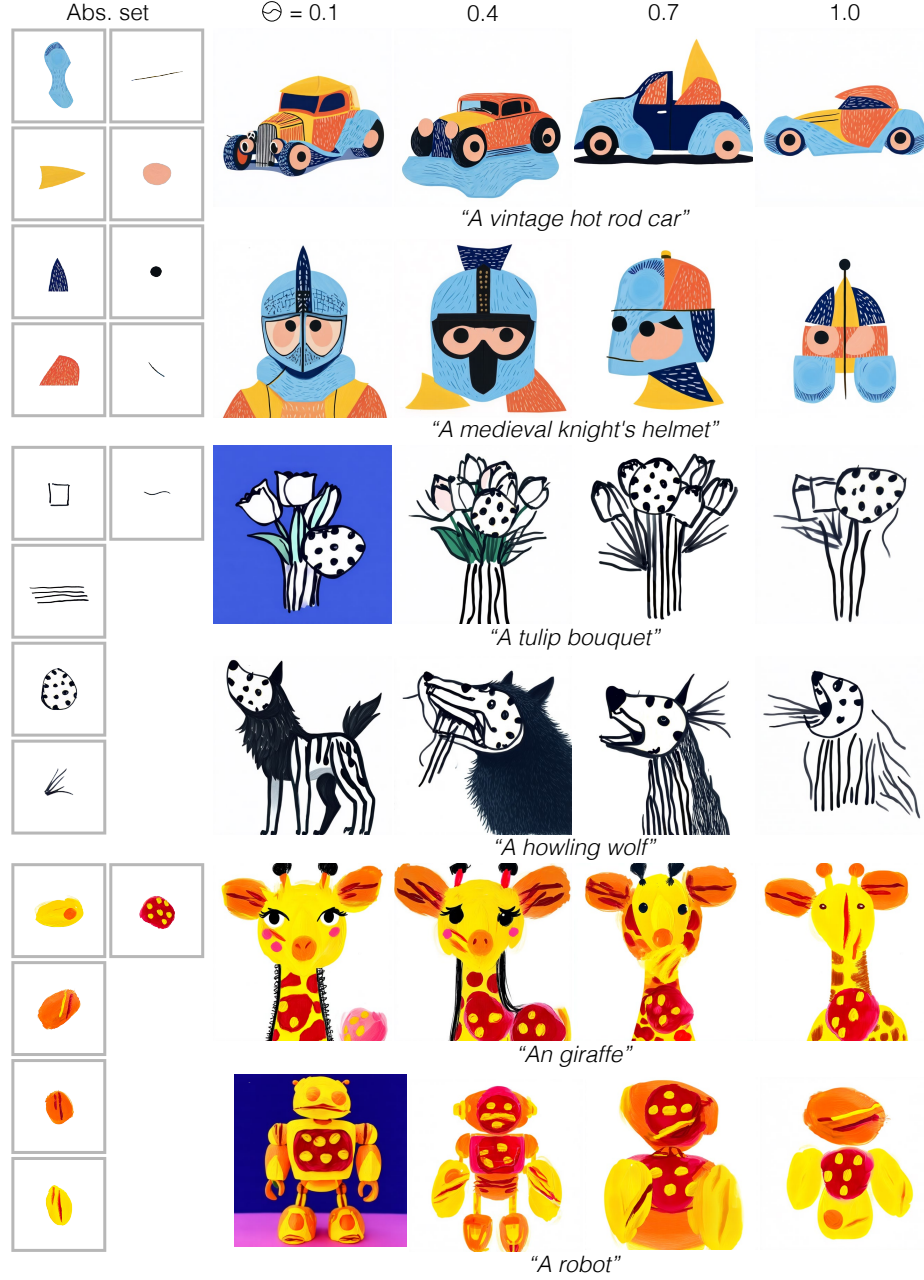


Fig. 8: FlowLess results over Stable Diffusion 3.5 Large. Left to right results were generated using increasing value of coverage score.

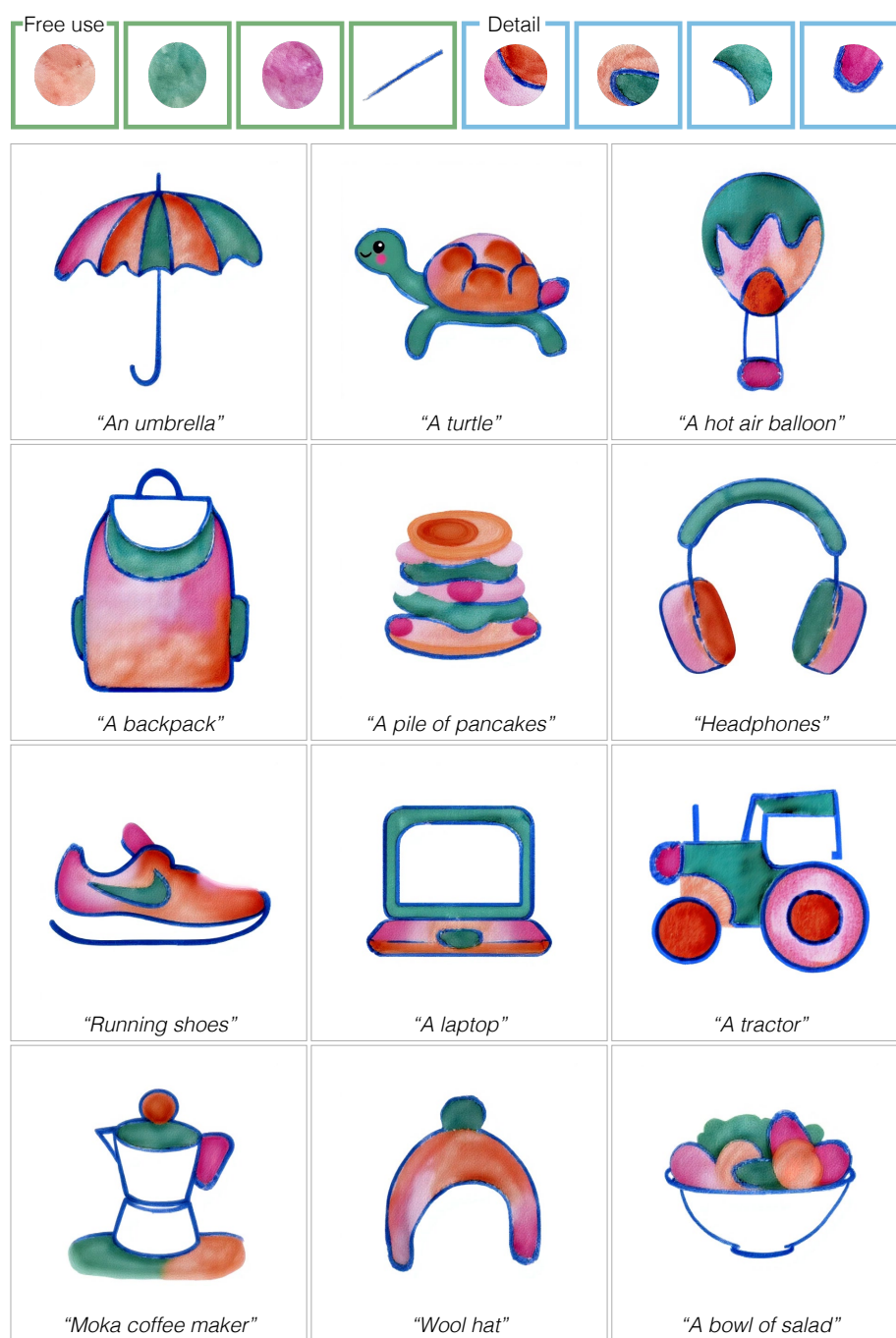


Fig. 9: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

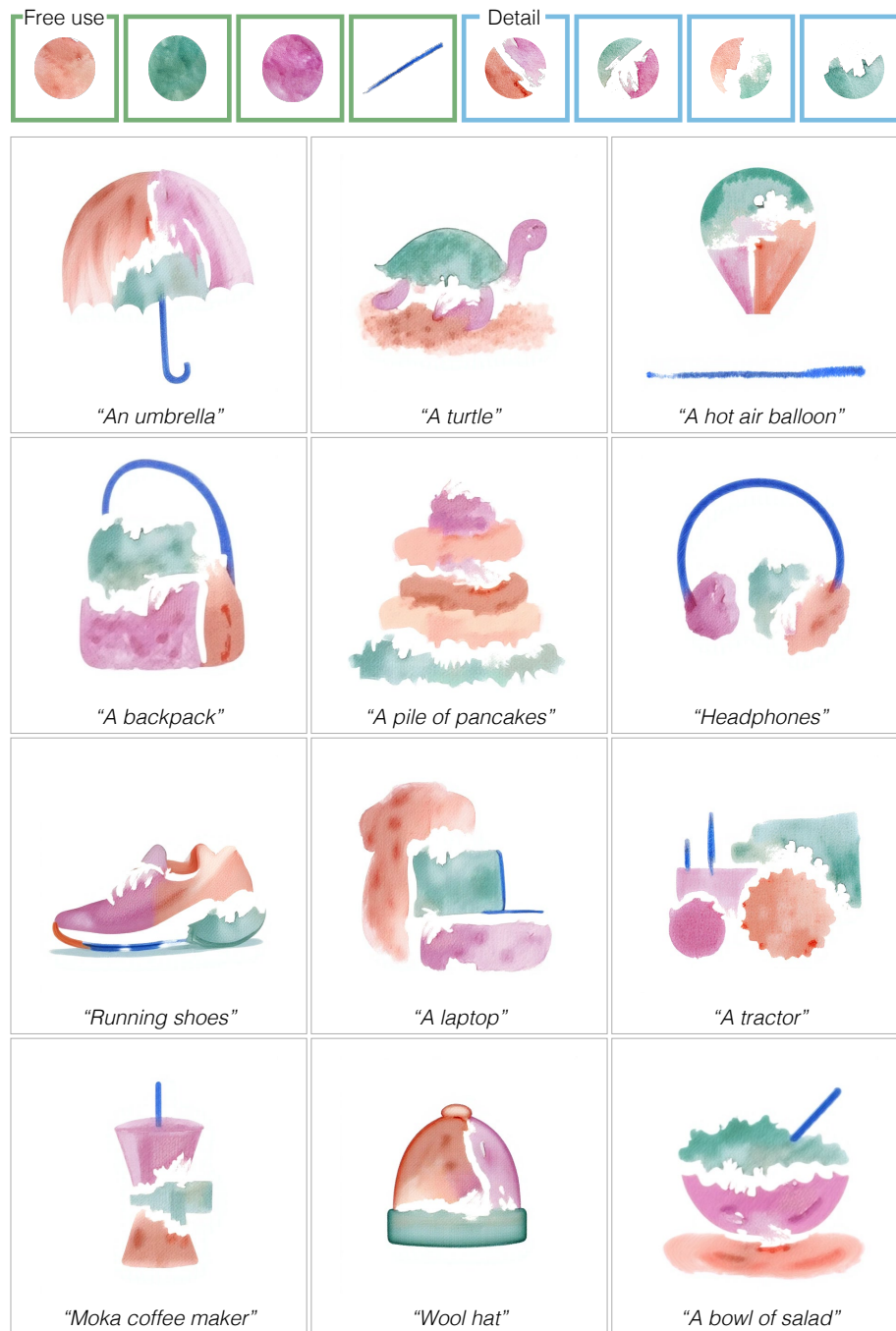


Fig. 10: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

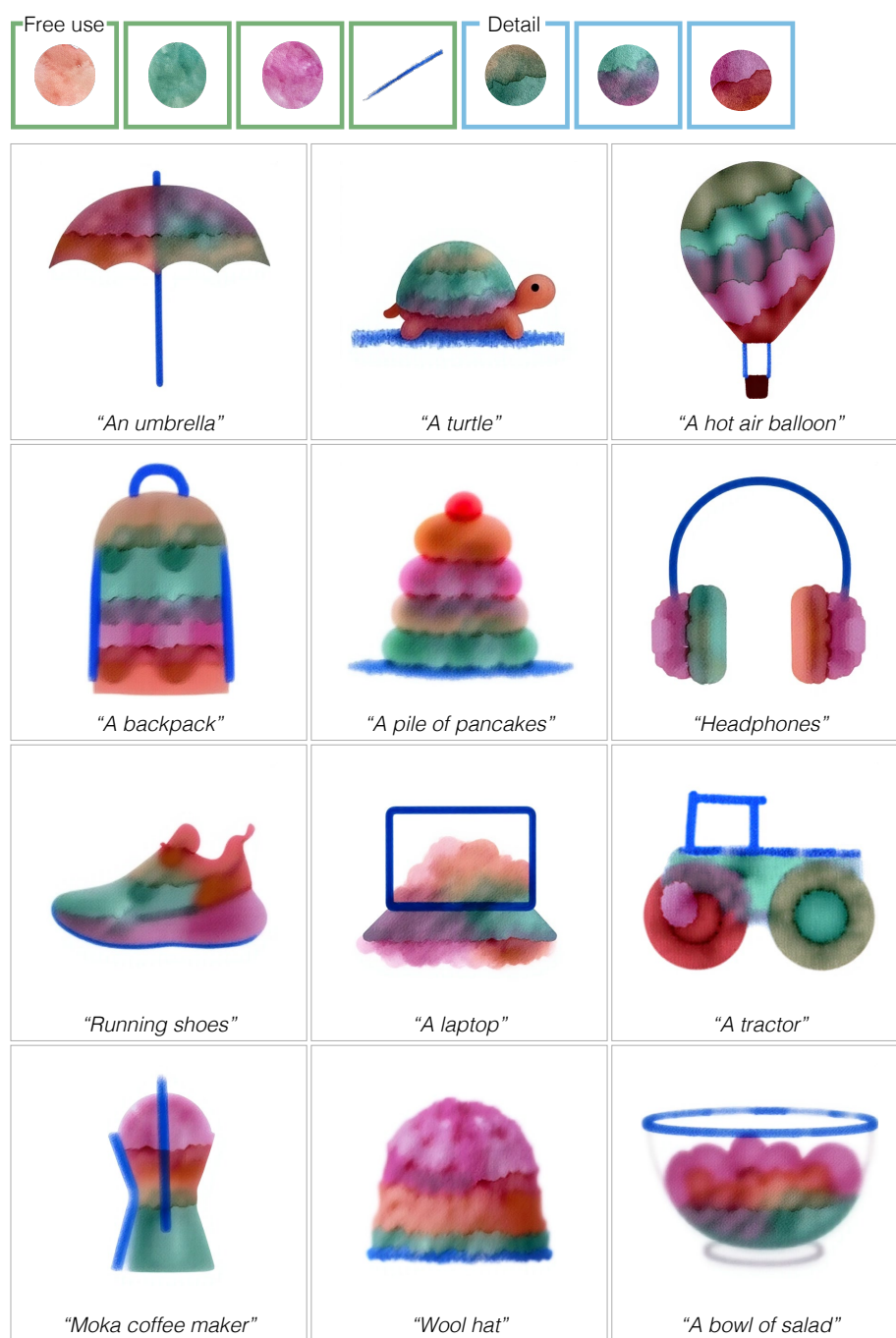


Fig. 11: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).



Fig. 12: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

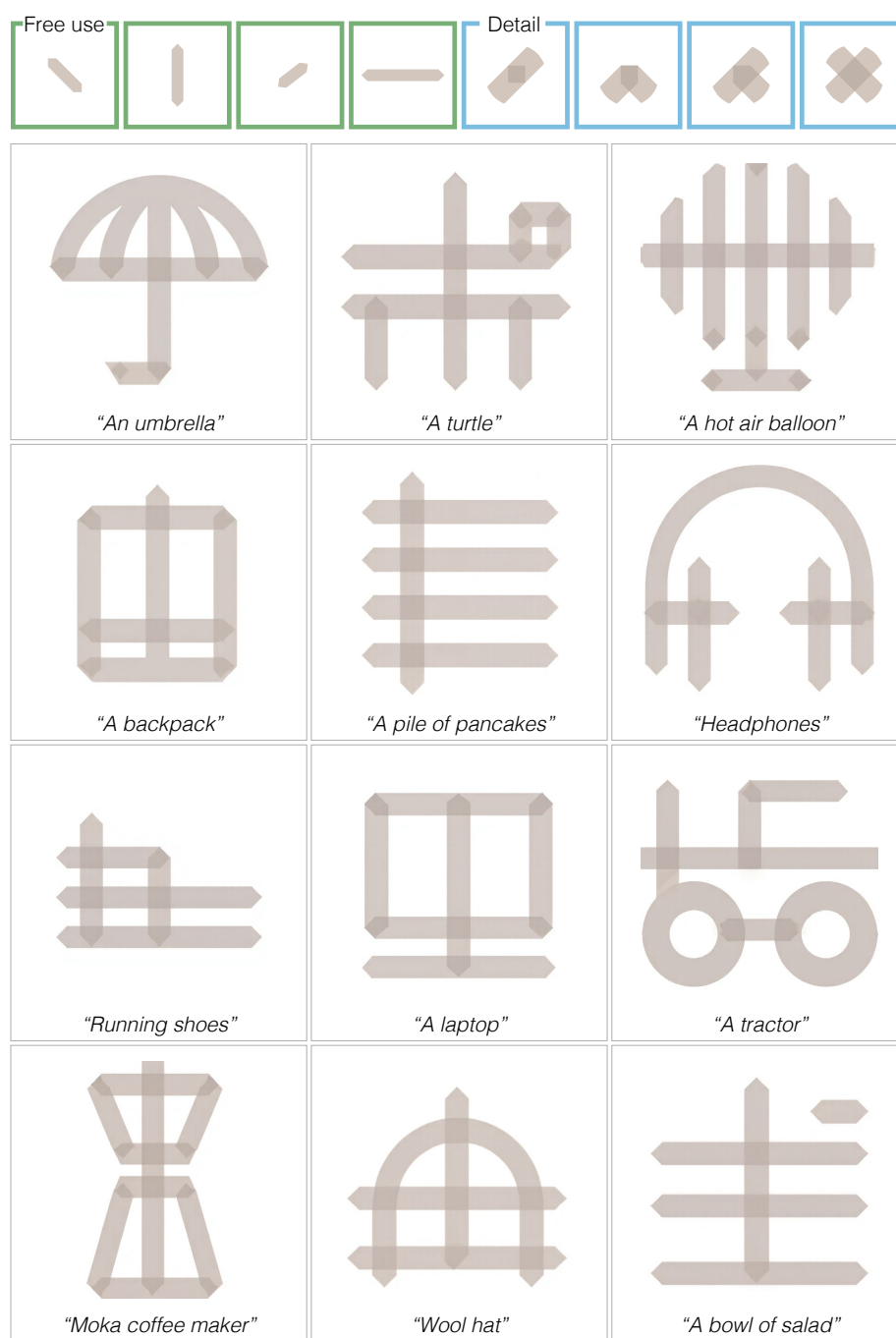


Fig. 13: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

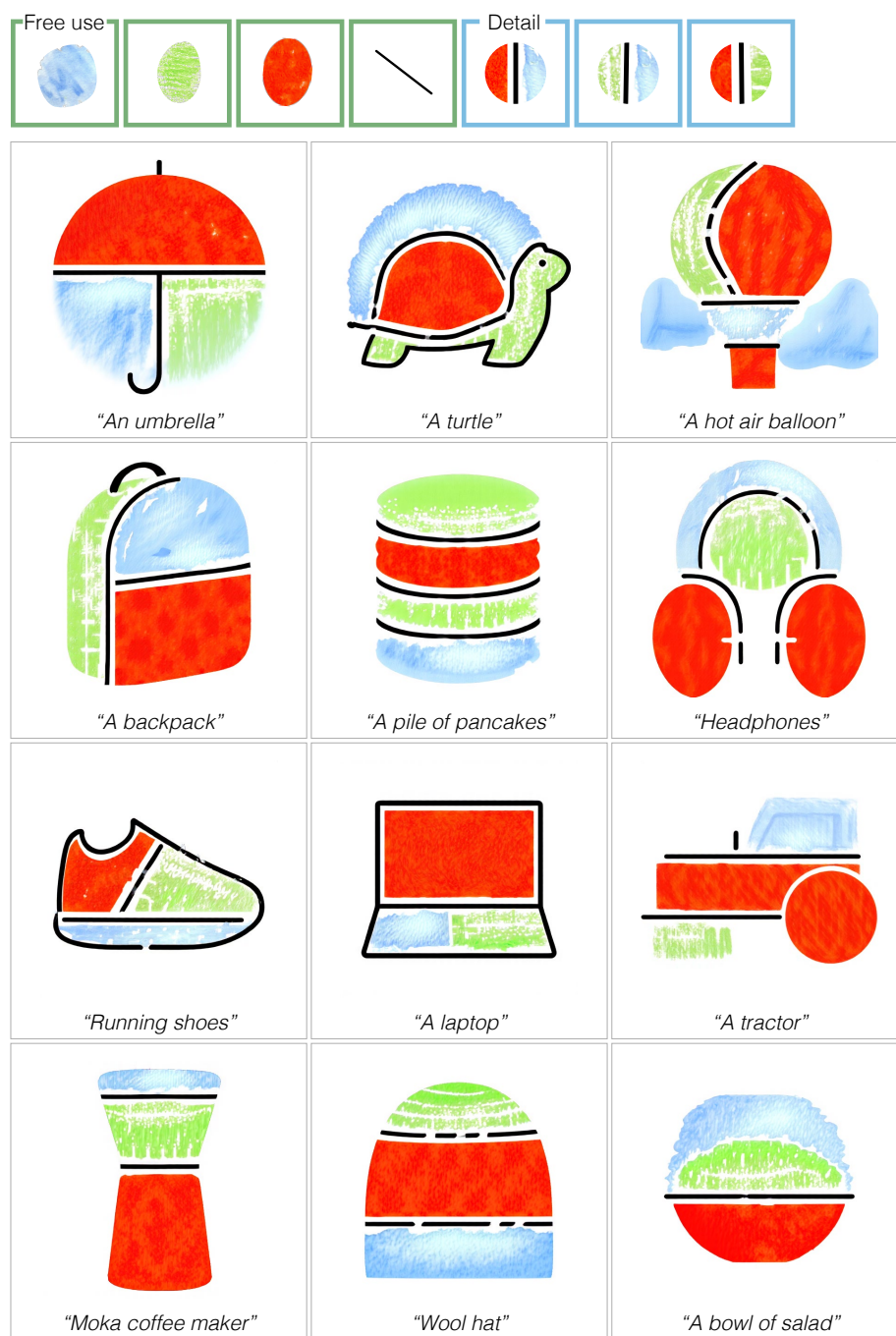


Fig. 14: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

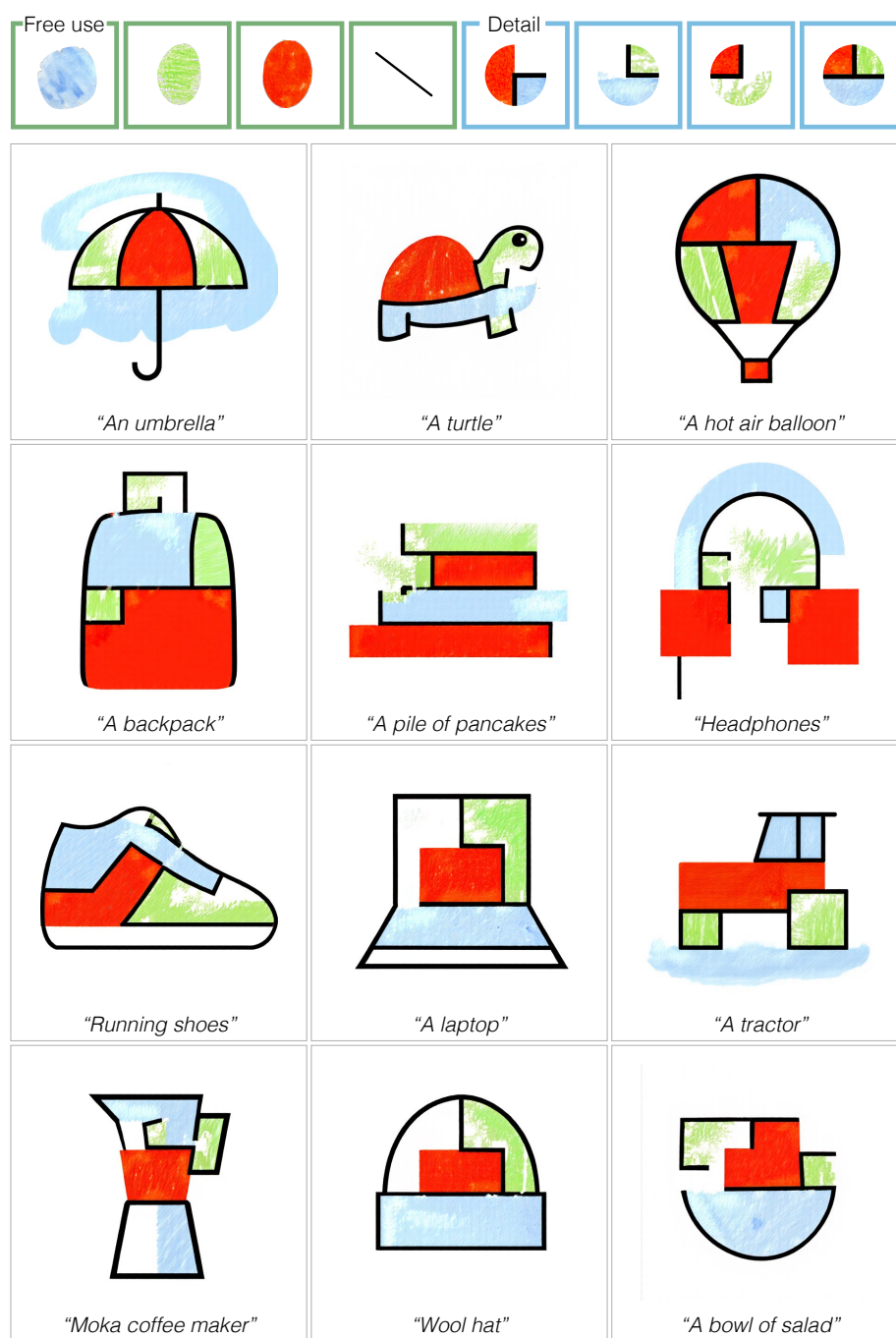


Fig. 15: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

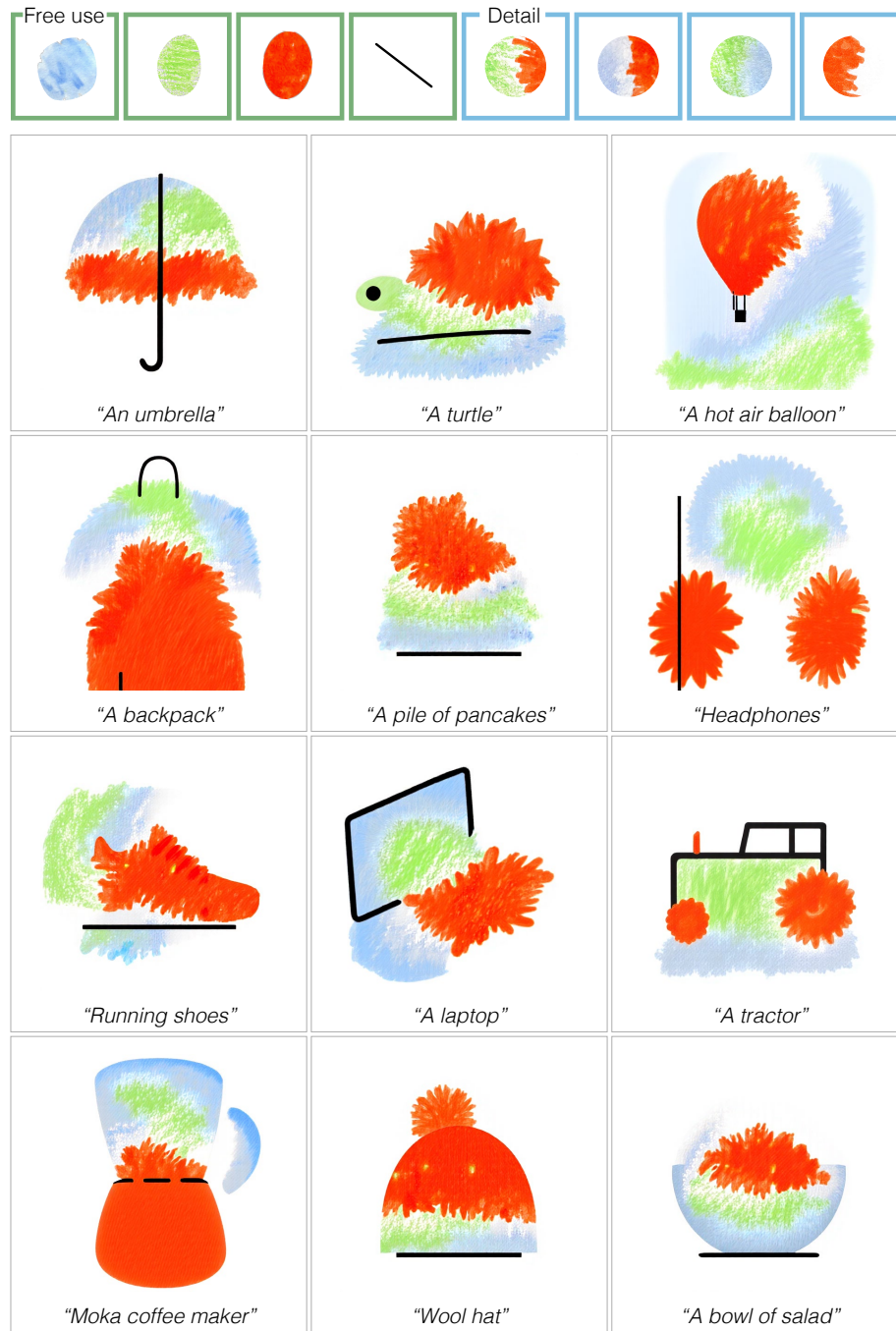


Fig. 16: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

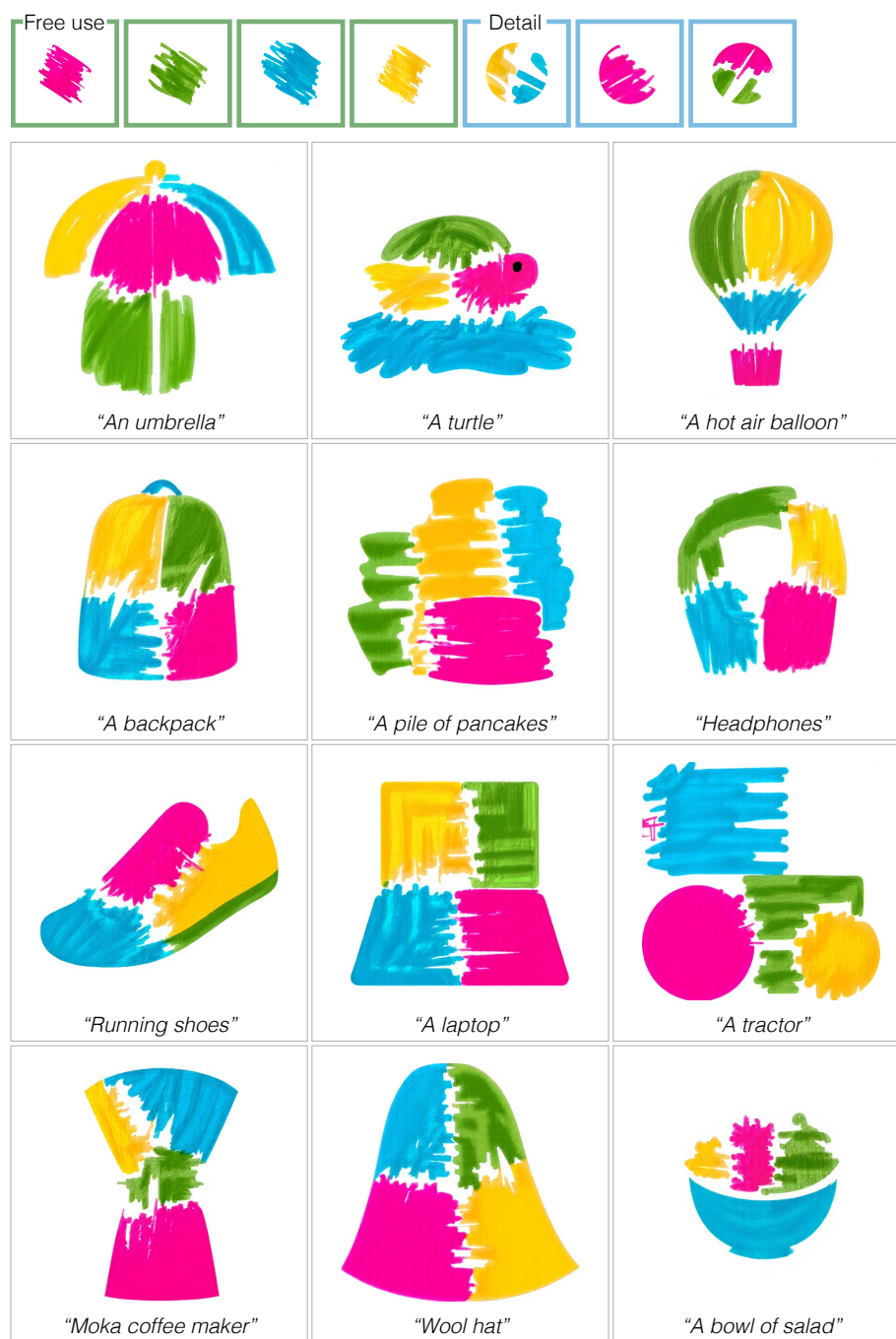


Fig. 17: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

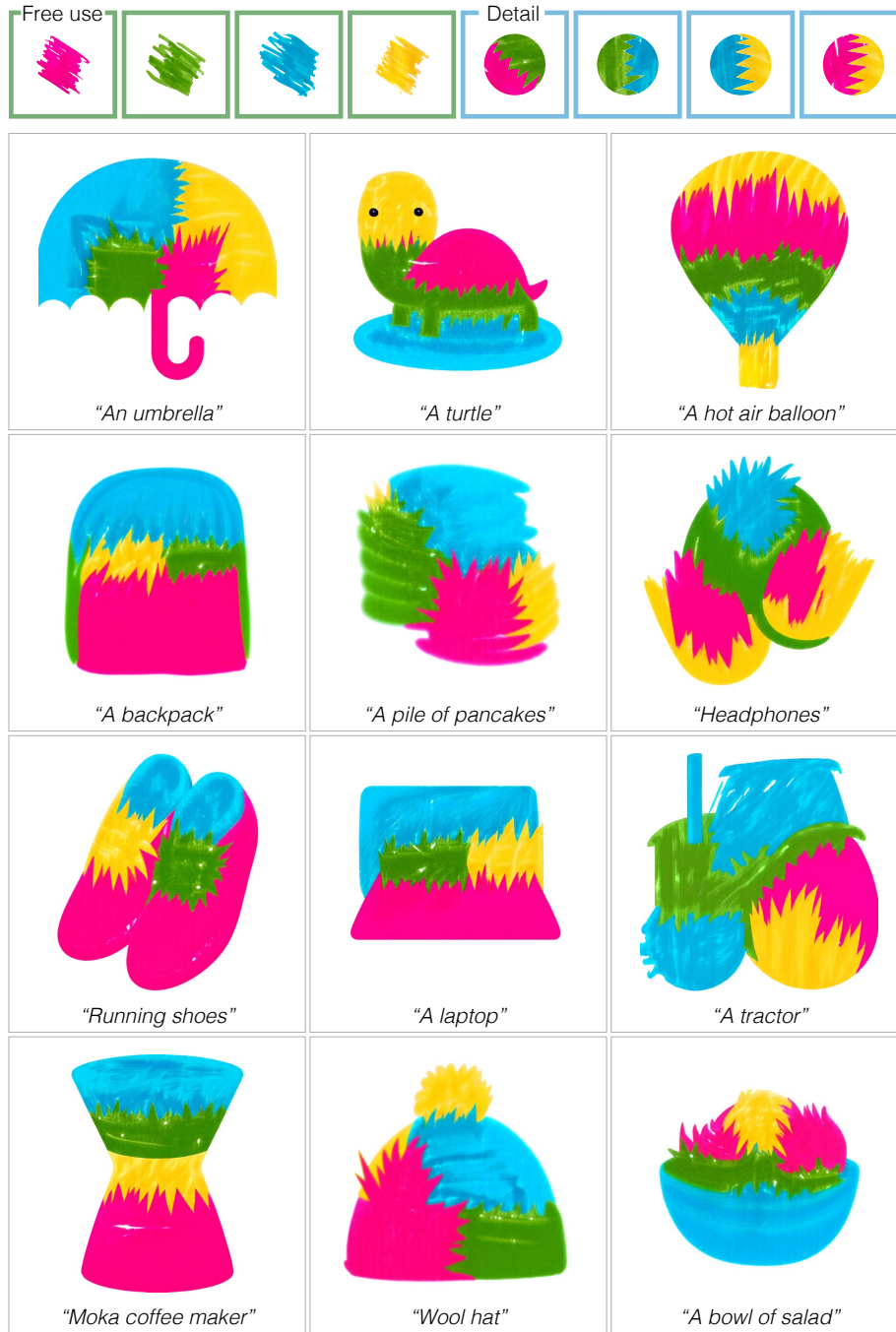


Fig. 18: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).



Fig. 19: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

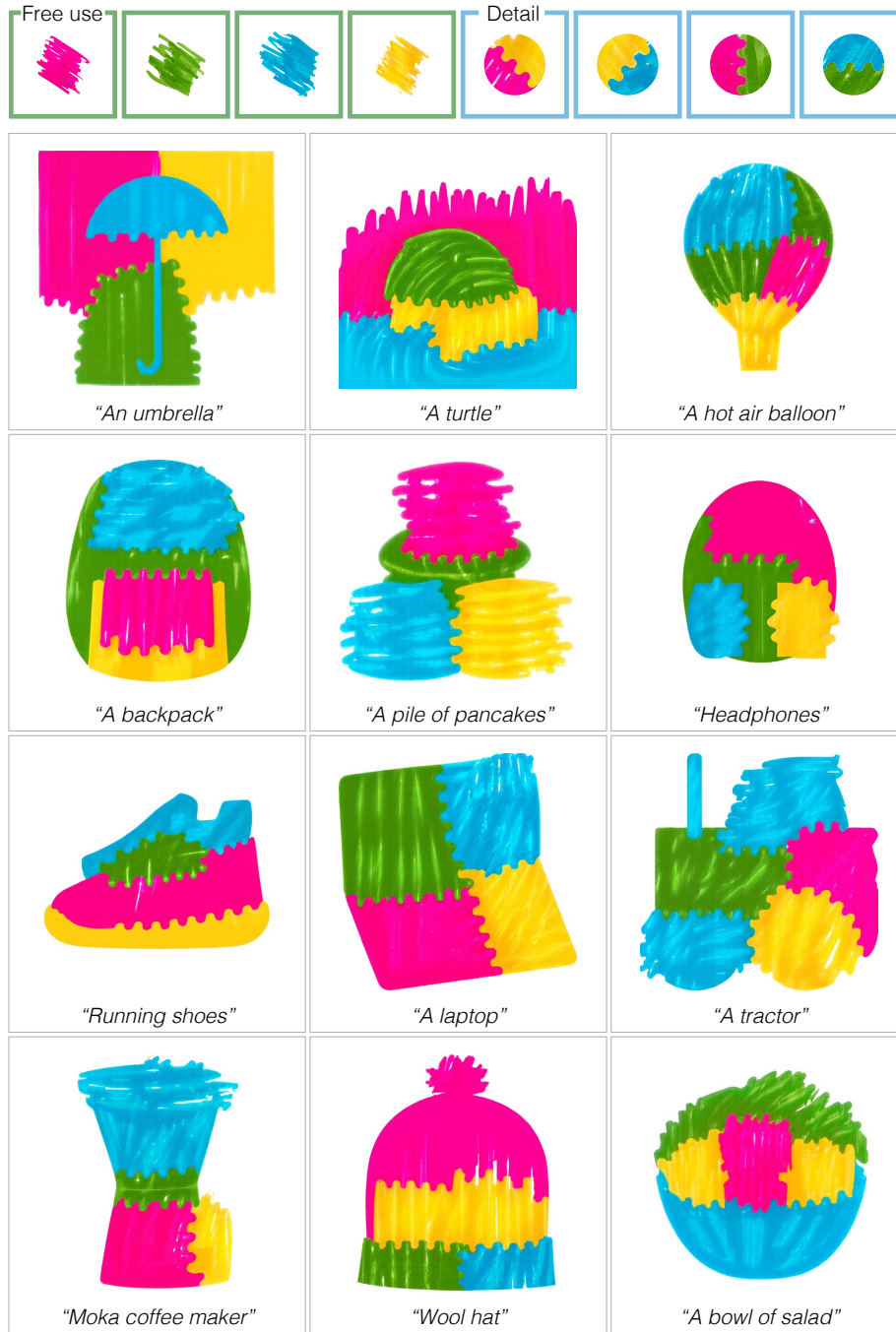


Fig. 20: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).



Fig. 21: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).



Fig. 22: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).



Fig. 23: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

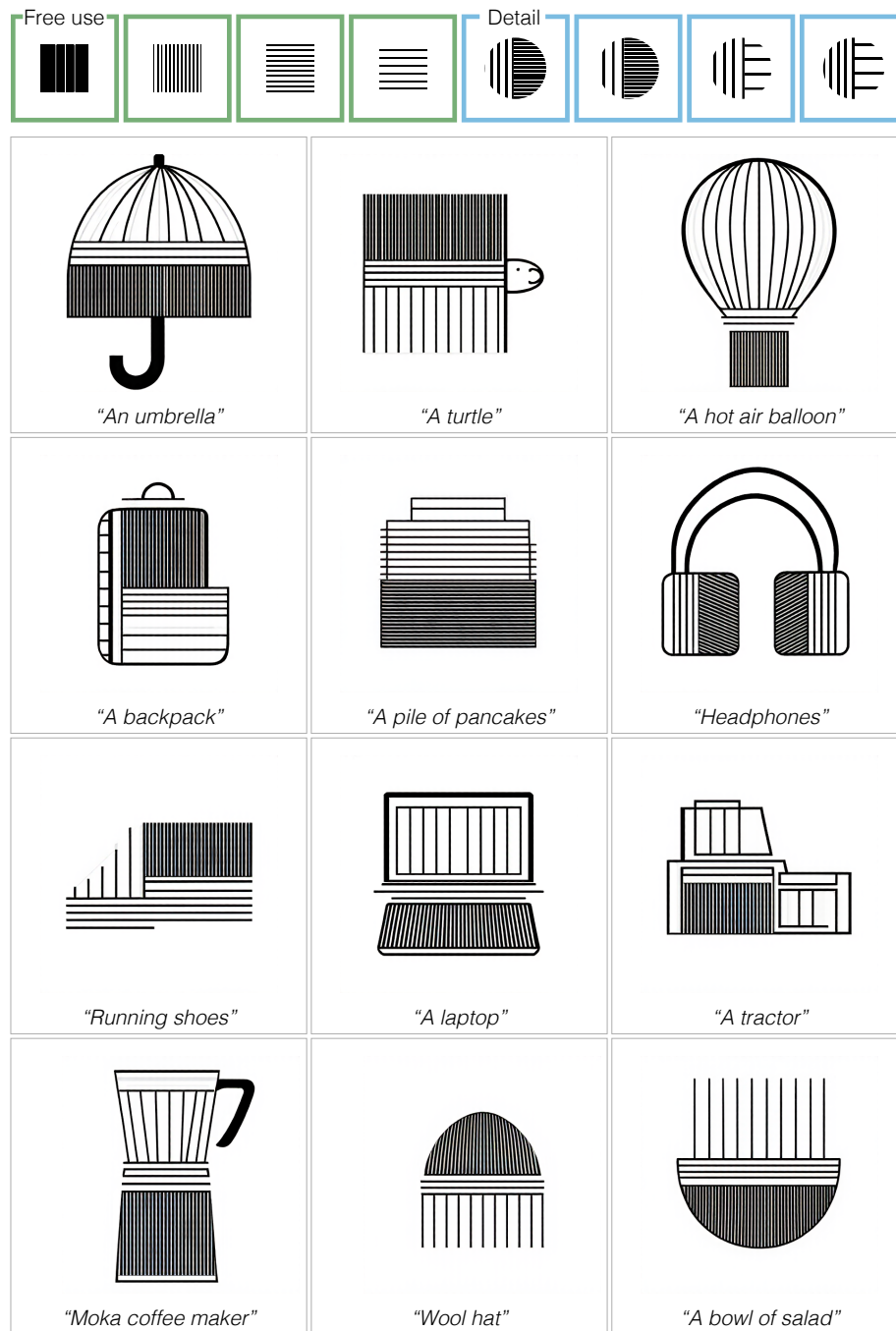


Fig. 24: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

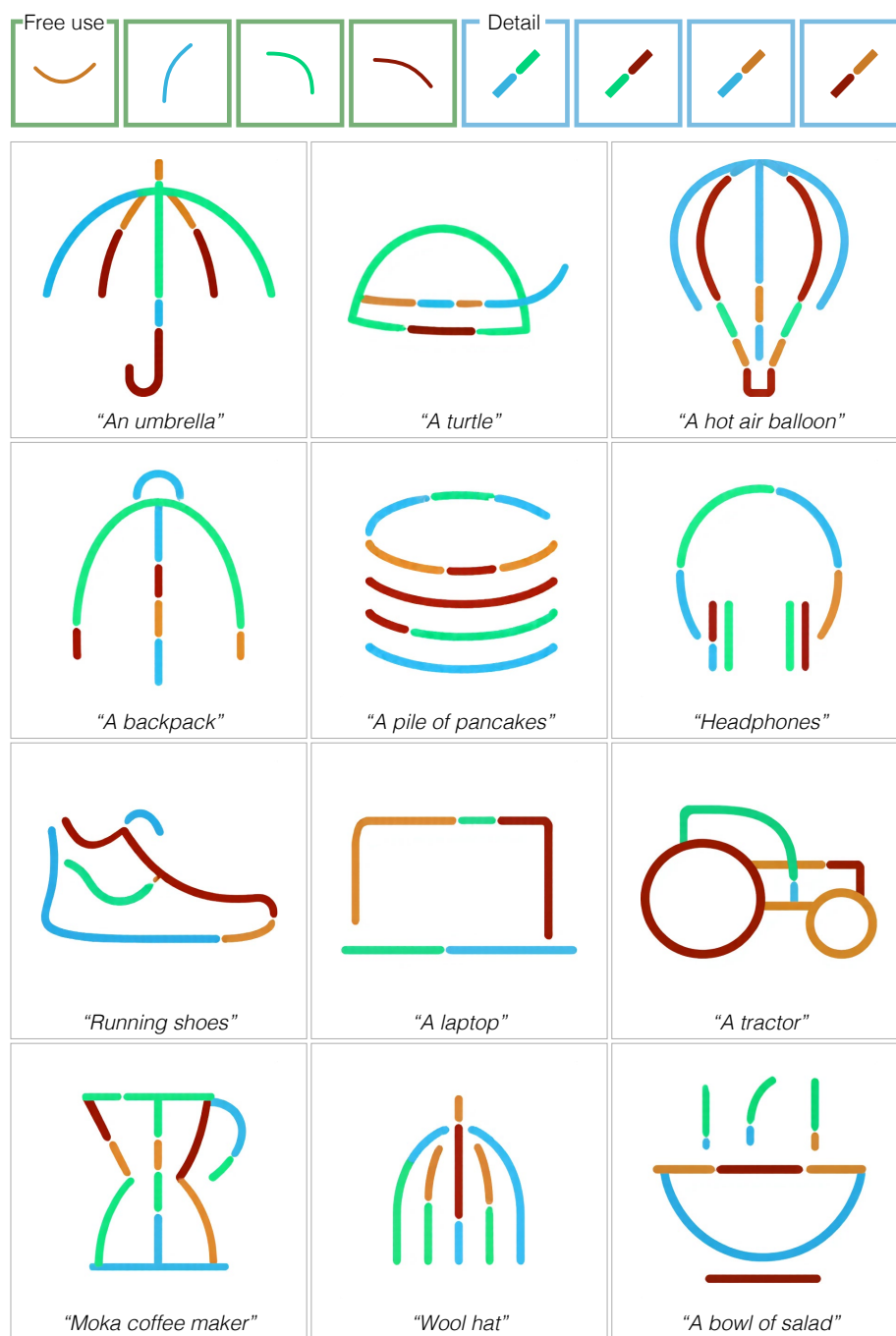


Fig. 25: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

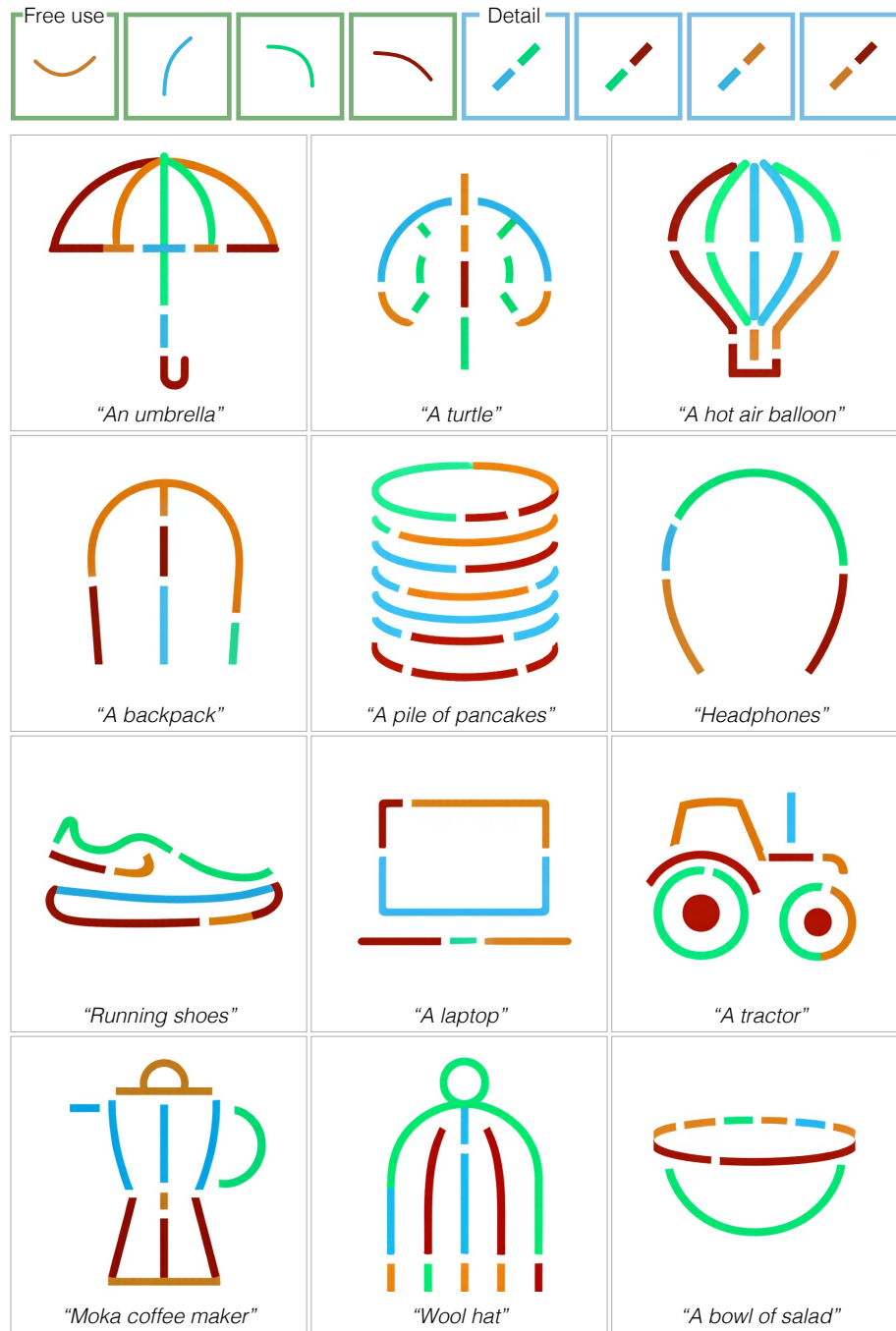


Fig. 26: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).

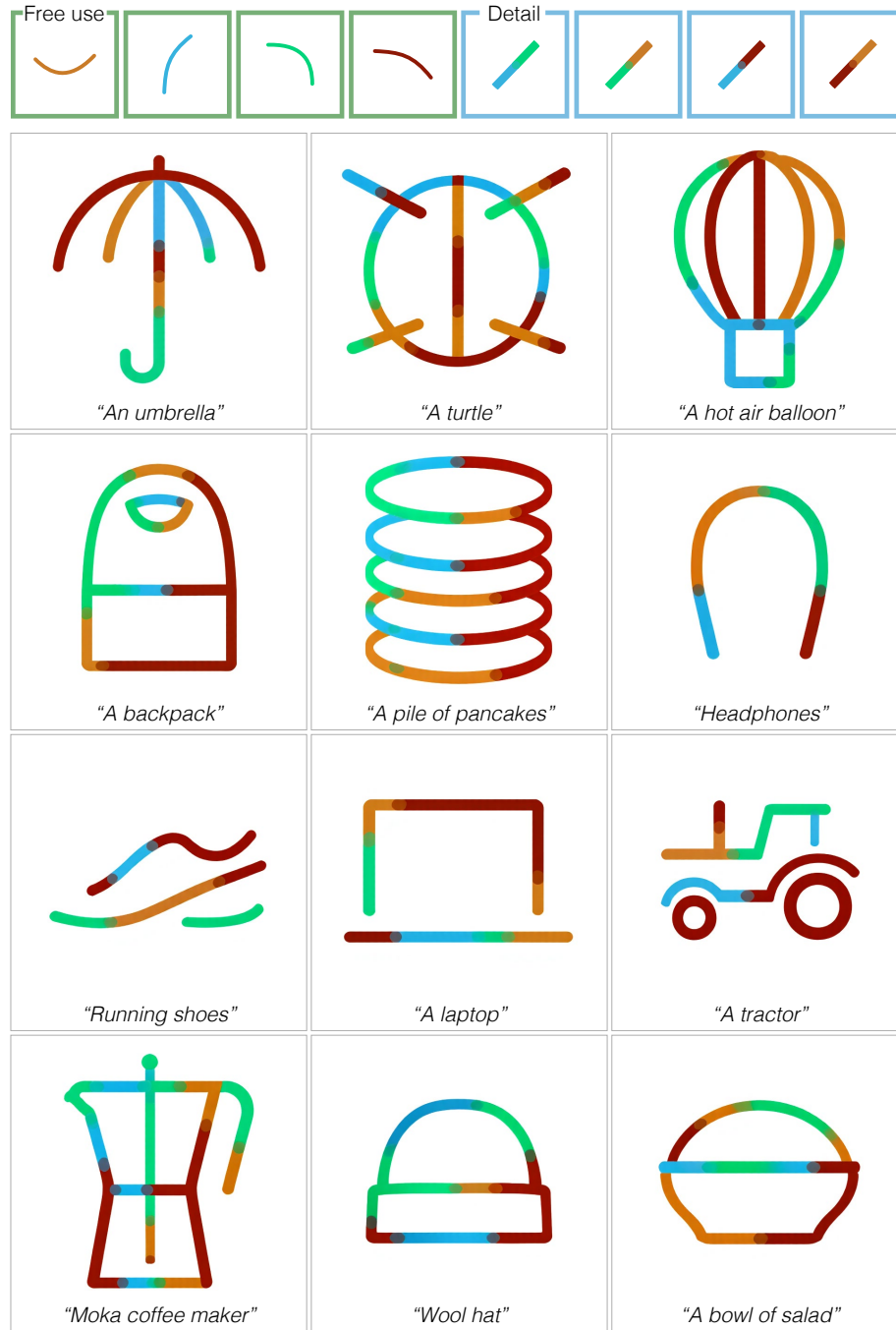


Fig. 27: FlowLess text-to-image generation. Top row shows the input materials (green outline) and details (blue outline).