https://doi.org/10.2352/CIC.2024.32.1.17

Color Terms and Stable Diffusion

Nathan Moroney, Numantic Solutions

Abstract

Stable diffusion is a generative algorithm for creating images from text prompts. This paper explores prompts with color terms and proposes a process to generate, visualize and assess these results. Automated prompts are used to generate and render a `color term`*, an* `object` *and a* `context`*. The results are then evaluated using two dashboard views of the underlying images. First is a sampling based on a collection of frequently used color terms. Second is a sampling by object prompts, such as apples and boxes. This paper considers the following questions: how effectively are the colored objects generated? how do the colors generated by stable diffusion compare to human color naming? How might color terms be useful in visualizing properties and features of generative algorithms? The dashboard view of color terms suggests that less frequently used color terms may be generated less consistently. In addition, even the most common color terms can fail to be correctly generated. Likewise, objects with more frequent color associations, such as apples or pumpkins, will result in less accurate color generation.*

Introduction

Stable diffusion is one algorithm for generating images from text prompts.[1] Other algorithms[2] exist for this type of generative processing but this paper focuses on stable diffusion given the availability of an open implementation of the trained model. Specifically `stable-diffusion.cpp`[3] with version 1.4 weights are used and run on a laptop. Stable diffusion is a type of denoising autocoder that learns to synthesize images using forward and reverse pass diffusion denoising.[4] This diffusion is done in a latent space[5] and uses cross-attention as a general-purpose conditioning of signals, such as text. The diffusion process was trained on pairs of images and captions taken from LAION-5B.[6] The CLIP ViT-L/14 text embedding module, was trained the on the proprietary "WebImageText" (WIT) dataset of images and captions.[7]

At an overly simplistic level, stable diffusion transforms a text string to matrix of colors or red, green and blue values. How does stable diffusion perform when color terms are also used as the input? This question has been considered in the context of quantifying model alignment, but with a focus on the basic color terms.[8] The accuracy of the colors generated is one form of attribute binding[9] and contributes to the image quality of the resulting renderings. But what might be learned about stable diffusion by focusing on color terms?

Human color naming is well researched and includes linguistic debates about relativist versus universalist theories of language.[10,11] More broadly, color categorization has also been studied in infants[12] and corvids[13]. In the color naming literature and for this paper the basic terms correspond to the following eleven colors : red, green, blue, yellow, brown, pink, purple, black, gray and white. Non-basic color terms are additional one word names, such as cyan, or multi-word names such as sky blue. In addition, a range of color naming datasets have been collected and analyzed in different contexts.[14-20]

One simple observation is that fuchsia is surprisingly difficult for English speakers to spell.[21] Figure 1 demonstrates that stable diffusion will generate color objects corresponding to both fuschia [*sic*] and fuchsia. It is likely that these and many other color terms are present in the LAION-5B and WIT datasets but at the time of publication, both were unavailable for analysis. This paper proposes the use of structured color prompts and interactive modal dashboards to better understand how color terms like fuchsia (and it's misspellings) are rendered by stable diffusion.

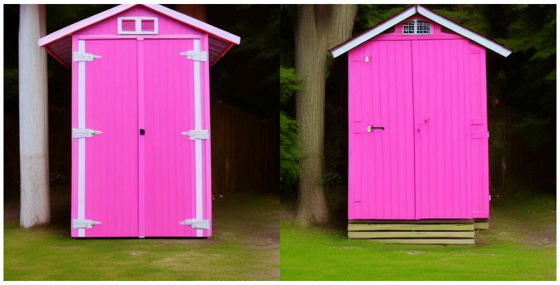
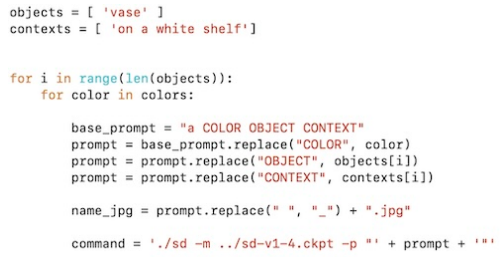


Figure. 1. *Stable diffusion results for "a fuschia door on a narrow shed" (left) and "a fuchsia door on a narrow shed" (right)*

Color Prompts

This paper uses color prompts with the following overall structure : a `color term`, an `object` and a `context`. A color term is a color name, such as 'green' or 'fuchsia'. The object is the subject of the rendering, such as a 'door' or a 'bowl'. Finally, the context is the setting or environment, such as 'on a narrow shed' or 'on a white shelf'. The example python code shown in Fig. 2 provides an example implementation.



```
objects = [ 'vase' ]
contexts = [ 'on a white shelf']


for i in range(len(objects)):
    for color in colors:

        base_prompt = "a COLOR OBJECT CONTEXT"
        prompt = base_prompt.replace("COLOR", color)
        prompt = prompt.replace("OBJECT", objects[i])
        prompt = prompt.replace("CONTEXT", contexts[i])

        name_jpg = prompt.replace(" ", "_") + ".jpg"

        command = './sd -m ../sd-v1-4.ckpt -p "' + prompt + '"'
```

Figure 2. *Python code demonstrating a color prompt consisting of a* `color term`*, an* `object` *and a* `context`*.*

Figure 3 below shows the stable diffusion results for the basic color terms (top 4 rows) and non-basic color terms (bottom 4 rows) and prompts with the object of a 'bowl' and a context of 'on a kitchen counter'. The top of this figure shows bowls for red, green, blue, yellow, brown, orange, pink, purple, black, gray, grey and white. The bottom of this figure shows bowls for beige, celadon, cyan, dark blue, fuchsia, light green, lilac, lime green, tan, taupe, teal and violet. The top images are recognizable based on the basic color terms used in the prompt. For the bottom images, the celadon, cyan, fuchsia and taupe bowls are an identical white bowl. This is hypothesized to indicate a potential gaps in the training data. This infilling of colors can be contrasted with cognitive color and attention effects.[22]

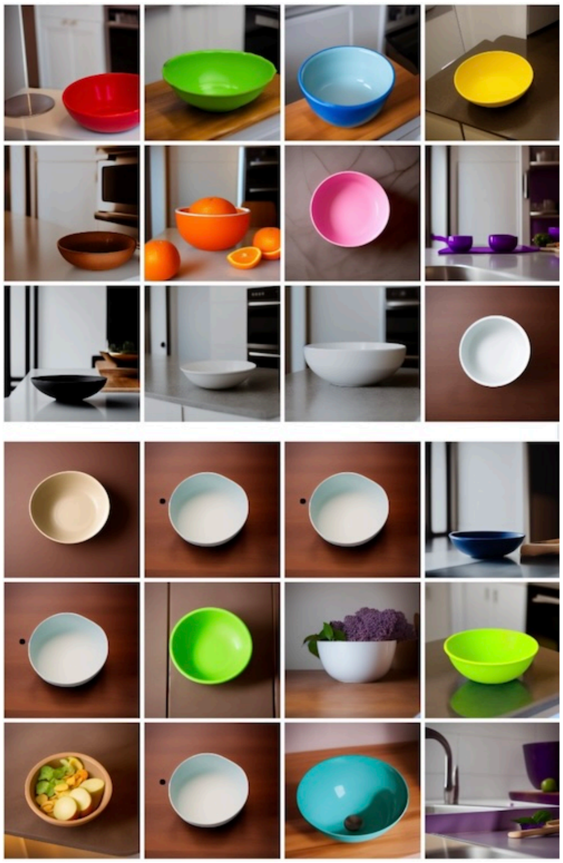


Figure 3. *Basic (top 12) and non-basic (bottom 12) color terms used with a color prompt with an `object` of 'bowl' and a `context` of 'on a kitchen counter'. See the text for a complete list of the color terms.*

The C++ stable diffusion implementation includes a number parameters. One of these parameters is the number of iterations applied to the diffusion process. The default value is 20 and this value (and all of the other defaults) are used for all examples in this paper. With respect to the number of iterations, its is possible to use a color prompt to visualize the time course of the resulting colors.

Figure 4 shows the results for a color prompt in which the `color term` is 'red', the `object` is a 'table cloth' and the `context` is 'on a round table' (or 'a red table cloth on a round table'). The specific number iterations used was : 1, 2, 3, 4, 7, 9, 11, 13, 17, 18, 19, and 20. Note that the first and last rows are increments of 1 while the middle row uses larger increments. This sampling was chosen from the qualitative observation that the middle iterations in the sequences were largely variations on unstable elements in the composition, in this case the place settings. From the color analysis perspective the red of the table cloth was largely consistent across the number iterations. This and additional experiments suggests that analysis could be accelerated based on fewer iterations.

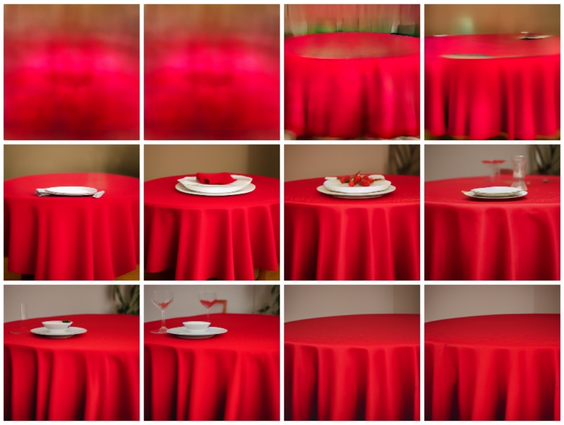


Figure. 4. *Red table cloths on round tables with steps of 1, 2, 3, 4, 7, 9, 11, 13, 17, 18, 19, 20*

The stable diffusion implementation used for this paper also provides a parameter for a seed value for a pseudo-random generator. This means that an identical prompt can yield different generated images based on the seed value. Figure 5 shows the results for the prompt 'a blue bowl on a kitchen counter'. Qualitatively, these bowls appear a similar blue but it is unclear how much color variation might result from changing the seed values.

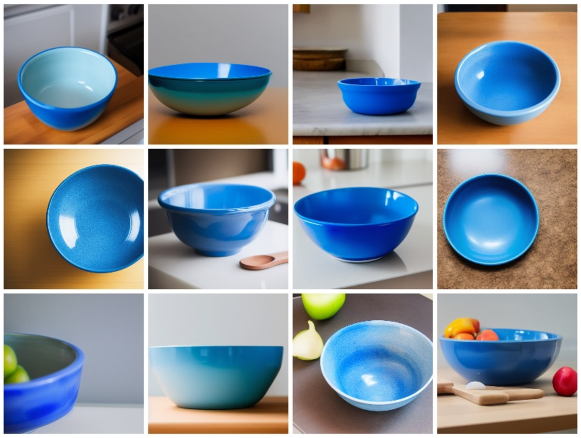


Figure 5. *Results for the color prompt "a blue bowl on a kitchen counter" with different random seed values.*

Assessment Dashboards

As an initial assessment of the results for color terms and stable diffusion, color prompts were processed in batches of hundreds. These images were then integrated into two Streamlit[23] dashboards with a selection slider. This slider allowed rapid evaluation of either color terms or objects. The images were display as a 4 by 4 array of thumbnails. Figure 6 shows the results for the color terms 'lime green' (top) and 'periwinkle' (bottom).

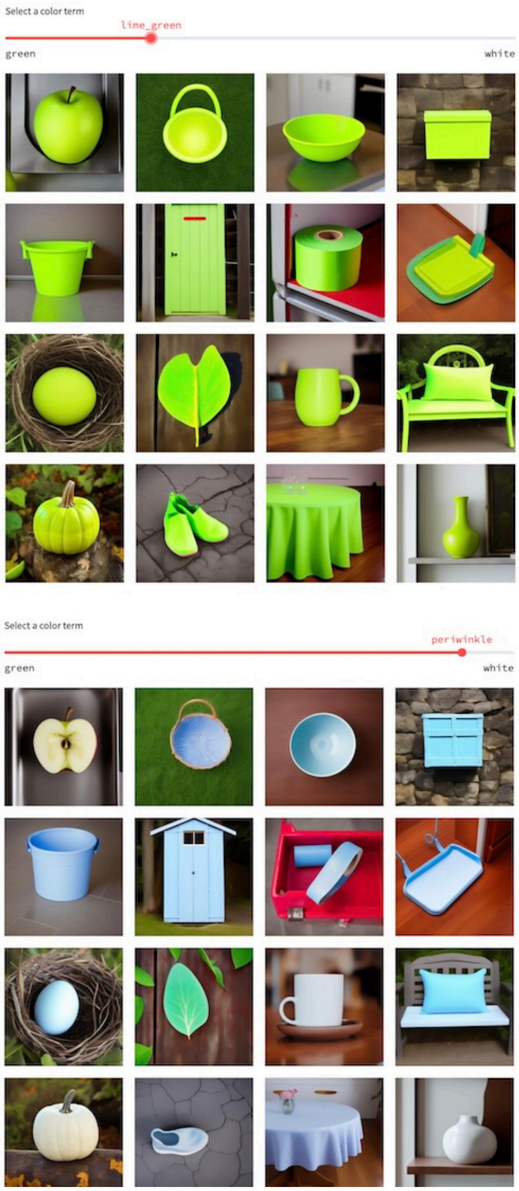


Figure 6. *Color term dashboard for 'lime green' (top) and 'periwinkle' (bottom).*

The results in Figure 6 allow a quick evaluation of lime green as a color term that is relatively consistent across a range of objects and contexts. In comparison, the periwinkle results show white objects or cases in which the generated colors are missing. The selection slider at the top, lists the color from more frequent on the left to less frequent on the right. This allows an efficient comparison of color terms. In addition the objects and contexts are in fixed locations (vases on shelves are always in the lower right corner).

The second dashboard, is an objects dashboard in which a fix set of color terms is used with a given object and context. Figure 7 shows the results for boxes (top) and apples (bottom). The context for the boxes is 'on a stone wall' and for apples is 'on a mental tray'.

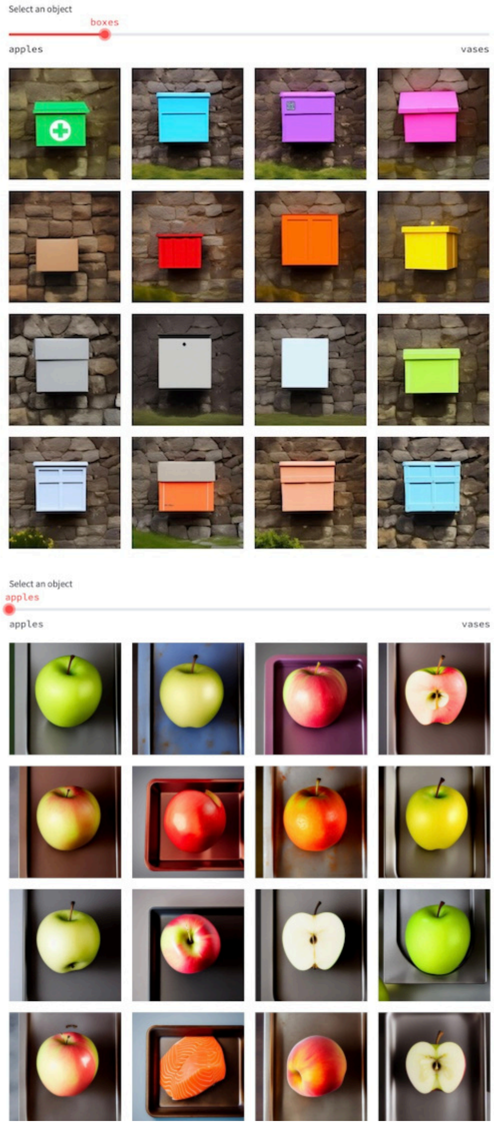


Figure 7. *Object dashboard for box (top) and apple (bottom).*

The color terms for figure 7 are from top-left to bottom-right : green, blue, purple, pink, brown, red, orange, yellow, gray, black, white, lime green, lavender, salmon, peach and periwinkle.

Qualitatively the colors generated for the boxes is consistent with human color naming. However the results for apples are less satisfactory. The object dashboard is helpful for exploring which objects are more accurately generated (doors or eggs) and those which are less accurately generated (pumpkins and leaves). It also demonstrates that ambiguous color terms like salmon can be generated as both the fish (bottom Fig. 7) and as a solid color (top Fig. 7).

Discussion

The stable diffusion model has hundreds of millions parameters and as was noted previously a range of additional execution parameters. It is also based on a number of large, unavailable training datasets. This makes it challenging to formulate and test specifics assertions about the model. This paper has used automated batch prompts with a simple `color term` + `object` + `context` format to sample the resulting rendered images. Use of color and object dash boards has yields some possible model characteristics which could be the subject of future study:

- Even basic color terms are not always consistently generated, for example black rendered as other than black occurred 1 out 3 times in one test run
- For a subset of objects, such as table cloths and shoes, the color term error rates are near zero indicating qualitatively correct rendering for the top 50 most frequent color terms.
- In contrast, objects with strong pre-existing color associations, such as apples or pumpkins have error rates of over 50 percent.
- Objects with relatively simplistic geometries, such as boxes and eggs, may be promising for larger area sample color patches for testing purposes
- Incorrect or missing coloration of objects tends to white or the typical color for that object
- Errors can occur when a color term is applied to nearby object, such as a saucer under a mug, instead of the prompt object
- Failures can also occur for color terms which are also objects, such as a violet basket be rendered as a tan basket filled with flowers

Figure 8 shows a bar plot of manually identified errors (as a fraction) for the top 50 most frequent color terms. A larger value indicates more errors in the rendered color with respect to the color specified in the prompt. Note that these values do not reflect other attributes of the result. For example, all dust pans generated had a relatively close color but often the geometry was unlike that of a typical dust pan.

With respect to specific color terms, Fig. 9 shows a CIELAB a* vs b* bubble chart of color terms. The coordinates of the points are the centroids for the prompts including the object ‘egg’ and the context ‘in a birds nest’. The radius is the ΔE^*_{ab} color difference with respect to the human nominal color centroid. The larger the bubble the larger the color difference, such as for the blues and cyans. This suggest that even for color prompts which render color for a wide range of color terms, there will be variations in the accuracy of these colors.

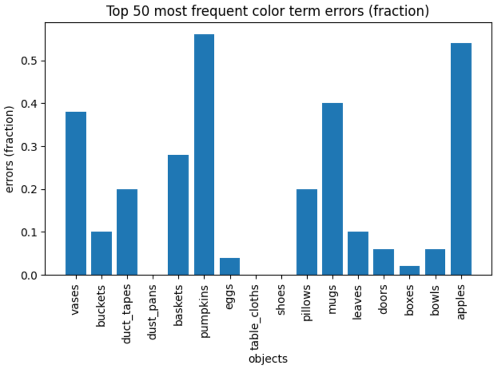


Figure 8. *Errors (as a fraction) for top 50 most frequent color terms for a collection of different objects.*

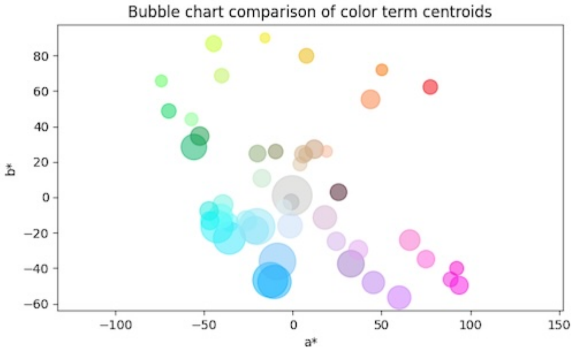


Figure 9. *Color centroids for the object prompt ‘egg’ and color difference with respect to nominal human color centroids as the radii.*

The results presented in this paper are for version 1.4 of stable diffusion. Future work should consider similar testing of additional generative models, including the newest versions of each. However, these results provide an initial reference point for assessing the color quality of one version of a highly cited generative algorithm. Work is also ongoing on further automating these results, such as computing color difference statistics between rendered object color centroids and human color naming centroids. In addition expanding the color term prompt to more complex appearance, such as glossy, matte or metallic is another promising research direction.

Conclusions

This paper has described a process for generating and visualizing stable diffusion generated color terms.[24] Automated prompts are created using the structure : `color term` + `object` + `context`. For example ‘a pink roll of duct tape on a red toolbox’ or ‘a beige dust pan on a hallway floor’. The resulting image collections can be efficiently assessed using color term and object dashboards. This has yielded an initial batch of hypotheses listed in the discussion section. These dashboards also provide an informative tool for presentation during an interactive session. These dashboards show that different objects will yield different errors in color terms, with table cloths having fewer errors than vases. Likewise for color prompts with a full range of color terms, the accuracy will vary. Some colors such as blues and cyans will have larger errors than reds and oranges when compared to the corresponding human nominal centroids.

References and Notes

[1] https://github.com/CompVis/stable-diffusion

[2] Zhang, Chenshuang, et al. "Text-to-image diffusion models in generative ai: A survey." arXiv preprint arXiv:2303.07909 (2023).

[3] stable-diffusion.cpp : `https://github.com/leejet/stable-diffusion.cpp` used with the weights : **sd-v1-4.ckpt**

[4] Lee, Seongmin, et al. "Diffusion explainer: Visual explanation for text-to-image stable diffusion." arXiv preprint arXiv:2305.03509 (2023).

[5] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10 684–10 695.

[6] https://laion.ai/blog/laion-5b/

[7] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.

[8] Grimal, Paul, et al. "TIAM-A Metric for Evaluating Alignment in Text-to-Image Generation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024.

[9] Hartwig, Sebastian, et al. "Evaluating Text to Image Synthesis: Survey and Taxonomy of Image Quality Metrics." arXiv preprint arXiv:2403.11821 (2024).

[10] Berlin, Brent, and Paul Kay. Basic color terms: Their universality and evolution. Univ of California Press, (1991).

[11] Saunders, Barbara. "Revisiting basic color terms." Journal of the Royal Anthropological Institute 6.1, pp. 81-99 (2000).

[12] Skelton, Alice E., John Maule, and Anna Franklin. "Infant color perception: Insight into perceptual development." Child development perspectives 16.2 (2022): 90-95.

[13] Apostel, Aylin, et al. "Corvids optimize working memory by categorizing continuous stimuli." Communications Biology 6.1 (2023): 1122.

[14] Moroney, Nathan. "Unconstrained web-based color naming experiment." Color imaging VIII: Processing, hardcopy, and applications. Vol. 5008. SPIE, 2003.

[15] Mojsilovic, Aleksandra. "A computational model for color naming and describing color composition of images." IEEE Transactions on Image processing 14.5 (2005): 690-699.

[16] Moroney, Nathan, Pere Obrador, and Giordano Beretta. "Lexical image processing." Color and Imaging Conference. Vol. 2008. No. 1. Society for Imaging Science and Technology, 2008.

[17] Mylonas, Dimitris, Lindsay MacDonald, and Sophie Wuerger. "Towards an online color naming model." Color and imaging conference. Vol. 18. Society of Imaging Science and Technology, 2010.

[18] Lindner, Albrecht, et al. "A large-scale multi-lingual color thesaurus." Color and Imaging Conference. Vol. 20. Society of Imaging Science and Technology, 2012.

[19] Heer, Jeffrey, and Maureen Stone. "Color naming models for color selection, image editing and palette design." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012.

[20] Mylonas, Dimitris, and Lindsay MacDonald. "Augmenting basic colour terms in English." Color Research & Application 41.1 (2016): 32-42.

[21] Parraman, Carinna, and Alessandro Rizzi. "CREATE: building a multi-disciplinary project in Europe." COLOUR CODED (2010): 6, pages 290-295, “The many misspellings of fuchsia” by N. Moroney

[22] Schwitzgebel, Eric. "Why did we think we dreamed in black and white?." Studies in History and Philosophy of Science Part A 33.4 pp. 649-660 (2002)

[23] https://streamlit.io/

[24] https://github.com/NMoroney/color_terms_and_stable_diffusion has code, prompts, images and other content relating to this paper.