Is CLIP Fooled by Optical Illusions?

Jerry Ngo
MIT CSAIL*
ngop@mit.edu

Swami Sankaranarayanan
MIT CSAIL
swamiviv@mit.edu

Phillip Isola
MIT CSAIL
phillipi@mit.edu

Abstract

Recent large machine learning models have achieved impressive performance on perception tasks such as classification or object detection, especially on unseen data. While it is unclear if they model the human cognitive process, they provide a compelling framework for case study and analysis. In this work, we test one such hypothesis: To what extent do large vision language models mimic the human cognitive system? We attempt to answer this question by focusing our attention on the ability of such models to perceive optical illusions. We analyze the CLIP model as a visual system, presenting stimuli in the form of image and text prompts and observing how the model's classification score changes under different illusory strengths. Our results show that CLIP is fooled by different types of illusions relating to lightness and geometry.

1 Introduction

Foundation models (Bommasani et al., 2021) are a class of large over-parameterized models that accomplish a wide variety of tasks relating to human perception. These include object classification and detection in computer vision, text-to-image synthesis, language translation and generation in natural language processing, etc. In this work, we analyze one such large vision-language model, CLIP, which has been shown to possess properties similar to human cognition such as multimodal neurons that encode concepts across vision and language, understanding emotion composition (Goh et al., 2021), exhibiting Stroop effect (Radford et al., 2021) and sound symbolism (Near, 2022) among others. We consider CLIP as a visual system and probe it with a variety of stimuli, record its response and compare those responses with how humans interpret the same stimuli. We focus our experiments on optical illusions which have been known to pose a challenge to human cognition.

Optical illusions are a phenomenon where the perceived visual world is different from reality. Studying illusions can lead to interesting observations as it can reveal how our brain processes visual information (Adelson, 1999). The existence of optical illusions has long intrigued scientists across diverse fields (Buckle et al., 2013). The advent of artificial neural networks (ANN), which have been modeled after the structure of the human brain, provide a good platform for testing the effects of optical illusions and thereby exploring the similarities between human and machine vision.

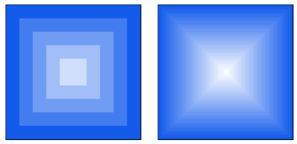
There are several previous works in machine learning exploring optical illusions from the perspective of low-level vision – training a simple CNN on natural images exhibits Gestalt Closure (Kim et al., 2019), VGG and ResNet models pre-trained on ImageNet can detect the Scintillating grid illusion (Sun & Dekel, 2021) etc. However, all previous works only use the visual modality, i.e. images, to measure the effect. In this work, we explore a new way to probe the ability of machine learning models to capture illusions by using multimodal stimuli i.e. image and text.

*The author's original affiliation is Beloit College. This work was done when visiting CSAIL.

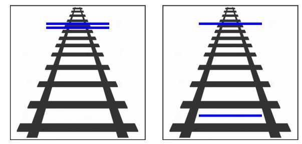
Preprint. Under review at SVRHM.

2 Setup

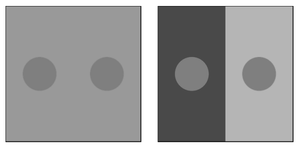
2.1 Included stimuli



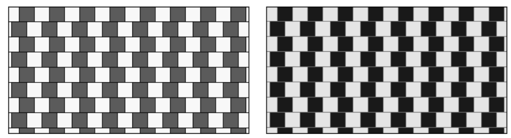
(a) Vasarely illusion



(b) Ponzo illusion



(c) Simultaneous contrast illusion



(d) Café wall illusion

Figure 1: This figure illustrates all the illusions used in this paper. The left image of every illusion is of a weak illusion strength variation, while the right image is one of strong illusion strength.

The Vasarely illusion comprises of superimposed squares with ordered levels of lightness. As the number of squares increases, an 'X' like shape appears along the diagonal of the image. This 'X' does not exist.

The Ponzo illusion depicts a rail/ladder from the lower end's point of view. On the rail, there are two horizontally parallel lines of the same length. However, the line in the lower end seems to be shorter.

The simultaneous contrast effect is the phenomenon where the surrounding luminance affects how the inner luminance is perceived. In this work, we focus on a simple version of this illusion. In this variation, the illusion is a square image divided into two areas with different luminance levels. Each area has a circle at the center, and both circles have the same luminance level. The circles in the darker area appear to be lighter and vice versa.

The café wall illusion consists of a checkboard-like pattern. Even though the board is made up of parallel lines of black and white tiles, those lines appear to be sloped.

3 Experiments and results

In this section, we discuss the different experimental setups used to assess how CLIP perceives illusions. First, we describe how to use CLIP as a zero-shot illusion classifier. We also introduce a calibration method to debias CLIP predictions. We then show that CLIP can be fooled by optical illusions by classifying illusion images into manually-designed prompts. Finally, we perform a case study where the text prompts are collected from human observers. This experiment is done to assess the effect of any bias in the prompt-designing process and ensure that our illusion analysis can generalize to multiple human observers.

3.1 Illusion classification using CLIP

Given an image I and a prompt T, we represent the CLIP similarity as $\text{CLIP}(I, T)$. For each stimulus, we create a zero-shot classifier using the CLIP model to match the input image to a prompt that describes the illusory signal (illusion prompt) or a prompt that ignores the illusion (non-illusion prompt). If CLIP scores the illusory prompt higher, then it implies that the input image is classified as an illusion by CLIP. Our method is illustrated in figure 2.

3.2 Calibrating CLIP comparison scores

Given an image and a text prompt, CLIP provides a score indicating the similarity between the semantic content of the image and text prompt. Using this raw score to classify an image into one of

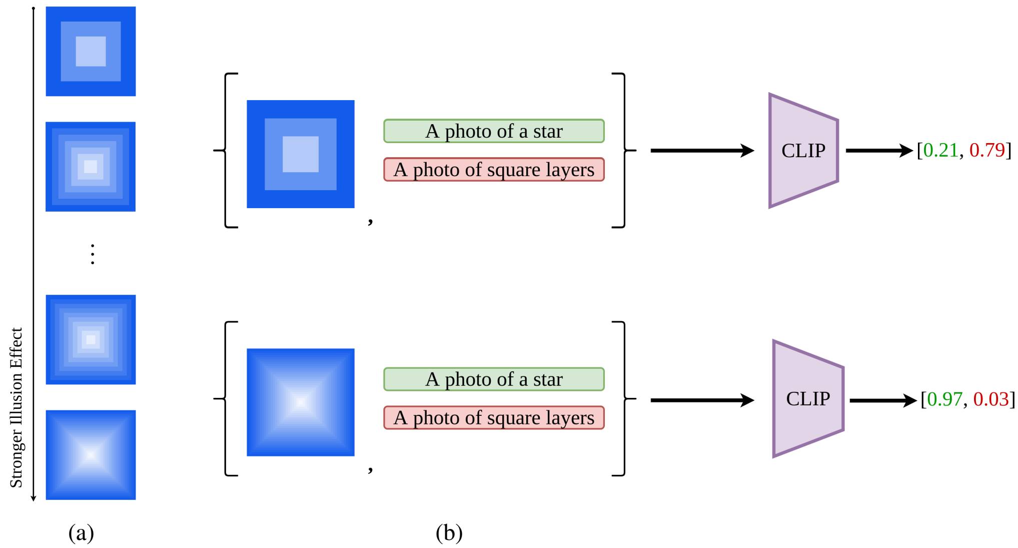


Figure 2: (a) We show the spectrum of illusory signals from weak to strong illusion effects. (b) To assess how CLIP perceives an illusion, we prepare a pair of prompts that includes an illusion prompt that acknowledges the illusion and a non-illusion prompt that ignores the illusion. We then classify images of different illusion strengths to one of the two prompts and compare their output comparison scores.

many prompts can be an effective tool when the concepts represented in the prompts have a clear semantic meaning, such as objects or animals. However, in our case, the prompts describe more abstract concepts, and utilizing raw scores can result in a biased prediction. For example, it could be unclear if the CLIP score is high for an illusion image and prompt pair due to CLIP being fooled by the illusion or if there is a random bias unaccounted for in the text prompt.

To remove this bias, we follow the method described in (Zhao et al., 2021) which discusses how large language model (LLM) results could provide unstable results depending on the prompt format, in-context sample choice, and permutation due to factors like training data distribution. They propose to obtain a content-free prediction to calibrate LLM's output, which helps reduce variance and improve performance. We apply a similar approach to calibrate the CLIP scores.

Let I be an image and T be a text prompt that describes the image. We define $\tilde{s} = \text{softmax}(\text{CLIP}(I, T))$ as the uncalibrated similarity score. Now let $\{I_c^{(i)}\}_{i=1}^N$ be a set of N random natural images that will be used for calibration. We define $\bar{s} = \frac{1}{N}\sum_{i=1}^N \text{softmax}(\text{CLIP}(I_c^{(i)}, T))$ as the average CLIP score for each prompt. Finally, we get a calibrated score $s = \text{softmax}(\tilde{s}/\bar{s})$. For our experiments, $\{I_c^{(i)}\}_{i=1}^N$ is a set of 128 images randomly drawn from the ImageNet dataset's training set. Unless specified, all the comparison scores in this paper are calibrated.

3.3 Prompt engineering experiment

In this experiment, we aim to design a set of prompts that fool CLIP for each illusion. We design this set such that each prompt clearly describes the illusory component in the input image. We also try to reduce the bias between the illusion and non-illusion prompts by keeping them at similar lengths and having an equivalent amount of contextual information. For each illusion, we use the strong illusion stimulus from figure 1 and iteratively modify existing prompts to maximize the illusion prompt similarity score with the image.

Table 1 shows the pair of prompts we designed for each illusion. In all these pairs, the calibrated illusion prompt score is higher than the non-illusion prompt, which signifies CLIP can be successfully fooled in all cases. With non-calibrated CLIP, the contrast and café wall illusion still fool CLIP. This suggests that CLIP is more robust to some illusory stimuli (image/text prompt combinations) than others.

Table 1: Prompt Engineering Results: We select a pair of prompts for each illusion. In every pair, the first prompt, with an asterisk at the end, is the illusion prompt, and the second is the non-illusion prompt.

Illusion	Prompt	Uncalibrated Score	Calibrated Score
Vasarely illusion	A photo of a star*	0.22	**0.58**
	A photo of square layers	**0.78**	0.42
Ponzo illusion	A photo of a train rail where two horizontal blue dashes of different sizes lay on*	0.48	**0.60**
	A photo of a train rail where two horizontal blue dashes of the same sizes lay on	**0.52**	0.40
Contrast illusion	A photo of two circles of two different shades of gray*	**0.66**	**0.75**
	A photo of two circles of the same shades of gray	0.34	0.25
Café wall illusion	A board of horizontally tilted lines*	**0.94**	**0.94**
	A board of horizontally parallel lines	0.06	0.06

3.4 Human case study

In this experiment, we want to measure the agreement between human descriptions of illusory signals and compare it to CLIP's assessment of the same stimuli. We designed a human survey to collect a diverse set of prompts for each illusion. We attach screenshots of the survey and the instructions in the appendix. We sent the survey with detailed instructions to four people, who each provided three illusion prompts and three non-illusion prompts per illusion. Through the case study, we collected 12 illusion prompts and 12 non-illusion prompts for each illusion (except for the café wall illusion with 11 prompts instead). From those prompts, we construct the set of all ordered pairs (illusion prompt, non-illusion prompt), which results in 144 pairs of prompts per illusion (121 pairs for Café wall illusion). We then follow the metric described below to determine whether an illusion fools CLIP.

3.4.1 Metric

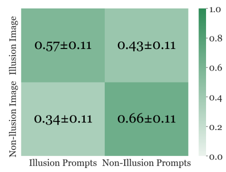
Given an illusory stimulus X, and a pair (p, q) of an illusion prompt and non-illusion prompt respectively, the CLIP score for illusion and non-illusion image-text pair can be given as:

$$S^i_{ill}, S^i_{non-ill} = \text{softmax}(\text{CLIP}(X, [p_i, q_i])) \tag{1}$$

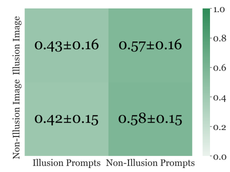
As we have multiple pairs of prompts for each image, we can calculate the mean and variance of the illusion or non-illusion prompts' comparison scores:

$$\mu = \frac{1}{N}\sum_i^N S^i \qquad \sigma^2 = \frac{1}{N}\sum_i^N (S^i - \mu)^2 \tag{2}$$

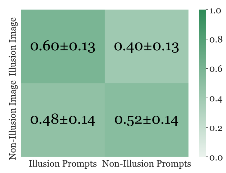
Applying (1) and (2), we calculate comparison scores' mean and variance between paired combinations of strong/weak-illusion images (from figure 1) and illusion/non-illusion prompts. Given an illusion image input, if an illusion fools CLIP, we would expect the resulting mean to be higher for an illusion prompt compared to a non-illusion prompt.



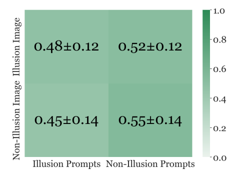
(a) Vasarely illusion



(b) Ponzo illusion



(c) Simultaneous contrast illusion



(d) Café wall illusion

Figure 3: Mean and variance of every illusion between strong-weak images and illusion and illusion/non-illusion prompt. The greater mean makes the grid's green darker. If an illusion fools CLIP, we would expect to see a darker color along the diagonal.

3.4.2 Results & discussion

Figure 3 shows the results of the described statistics for each illusion. Looking at the mean, we could tell that the Vasarely Illusion, Simultaneous contrast illusion, and Café wall illusion potentially fooled CLIP. We then perform two t-tests per illusion: one between illusion prompts and non-illusion prompts comparison score of illusion image, one between illusion prompts and non-illusion prompts comparison score of non-illusion image. The results show that the Vasarely illusion has significantly different means in both tests ($p < 0.05$). The simultaneous contrast illusion has significantly different means in the first test ($p < 0.05$). However, the café wall illusion has similar means in both tests.

From the above discussion based on figure 3, our results indicate that similar to humans, the CLIP model can indeed be fooled by optical illusions – Vasarely and simultaneous contrast illusions. However, we note that this result depends to an extent on how good the text prompt is in describing the illusory component in the image. As an example, let us consider an illusion prompt from the Vasarely illusion (fools CLIP) and Ponzo illusion (does not fool CLIP):

Sample Vasarely prompt : *A white, X-shaped light with blue background that is formed by superimposing squares*
Sample Ponzo prompt : *two lines with the upper one longer*

More descriptive prompts that give importance to the illusion concept in text are rated higher by CLIP and the difference in such prompt pairs (illusion vs non-illusion) is higher. More examples of prompt pairs obtained from our human study can be found in the Appendix.

4 Conclusion

Large foundation models such as CLIP not only are adept at solving perception-related problems but also have been shown to possess a great ability to generalize from few examples. In this work, we have considered CLIP as a visual system, providing us with a unique opportunity to probe the system with illusion stimuli, that successfully fool the human visual system. We find that while we can engineer text prompts to fool CLIP for any given illusion, the results from our human case study suggest that the descriptive nature of the prompt plays an important role in how CLIP is affected by the illusion. Of the 4 illusions we studied, we find that we can fool CLIP reliably with 2 of them, which is an intriguing effect that merits future research into what makes large models like CLIP susceptible to such inputs.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] Code will be released in the future. We have included survey responses in Appendix, full data will be released publicly in the future.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [Yes]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

References

Edward H. Adelson. Lightness perception and lightness illusions. 1999.

Michael Bach. Café wall illusion, 2022. URL https://michaelbach.de/ot/ang-cafewall/.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Christopher E. Buckle, Viyan Udawatta, and Christopher M. Straus. Now you see it, now you don't: Visual illusions in radiology. *RadioGraphics*, 33(7):2087–2102, 2013. doi: 10.1148/rg.337125204. URL https://doi.org/10.1148/rg.337125204. PMID: 24224600.

Gabriel Goh, Nick Cammarata †, Chelsea Voss †, Shan Carter, Michael Petrov, Ludwig Schubert, Alec Radford, and Chris Olah. Multimodal neurons in artificial neural networks. *Distill*, 2021. doi: 10.23915/distill.00030. https://distill.pub/2021/multimodal-neurons.

Been Kim, Emily Reif, Martin Wattenberg, Samy Bengio, and Michael C. Mozer. Neural networks trained on natural scenes exhibit gestalt closure. 3 2019. URL http://arxiv.org/abs/1903.01069.

Dominique Makowski, Zen J. Lau, Tam Pham, W. Paul Boyce, and S.H. Annabel Chen. A parametric framework to generate visual illusions using python. *Perception*, 50(11):950–965, 2021. doi: 10.1177/03010066211057347. URL https://doi.org/10.1177/03010066211057347. PMID: 34841973.

Asara Near. The bouba/kiki effect and sound symbolism in clip, 2022. URL https://nearcyan.com/the-bouba-kiki-effect-and-sound-symbolism-in-clip.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. 2021. URL https://github.com/OpenAI/CLIP.

Eric D. Sun and Ron Dekel. Imagenet-trained deep neural networks exhibit illusion-like response to the scintillating grid. *Journal of Vision*, 21:1–13, 10 2021. ISSN 15347362. doi: 10.1167/jov.21.11.15.

Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. Calibrate before use: Improving few-shot performance of language models. In *International Conference on Machine Learning*, pp. 12697–12706. PMLR, 2021.

A Appendix

A.1 Stimuli generation

A.1.1 Vasarely illusion

We generated illusion photos with different amounts of squares. For lightness, we used the HSL color model to modify the color of each imposed square. We used the 50 lightness (L) values for each color: from 50 to 100. If an image contains n square, the lightness values used in that image is the set $\{\lfloor 50 + \frac{50a}{n} \rfloor \mid 0 \leq a < n, a \in N\}$. Accordingly, we had the amounts of imposed squares going from 1 to 50. All images are in the 1:1 ratio with a size of 336 to 400 pixels. All the images are resized to 336 pixels by a pre-processing function before being passed into CLIP.

A.1.2 Ponzo illusion

We generated the Ponzo illusion by using a 336x336 image of a rail. Both horizontal lines have the same length of 226 pixels. The upper line is fixed to 45 pixels from the top. The lower line's location varies from 275 pixels to 30 pixels from the top, and every location jump is 10 pixels apart.

A.1.3 Simultaneous contrast illusion

As the difference between the surroundings and the inner circle increase, the illusion effect also increases. Thus, we could test how CLIP interprets the effect by varying the brightness of both canvases. The canvases are created using the package of Makowski et al. (2021). Using the HSL color system, we fixed the two circles with the color code of [160, 0, 50]. The left area uses the same color, with the L values varying from 51 to 100. Whereas the right area has L values varying from 49 to 0. Combining two areas together, we got a total of 50 photos of different pairs of brightness.

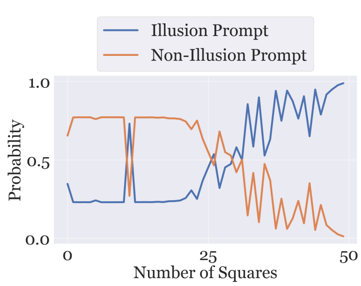
A.1.4 Café wall illusion

Two images were obtained from (Bach, 2022). They were both on the "b/w" and "1/2 checks" settings. The strong illusion image was taken at the default "b/w" colors, while the non-illusion image's colors were adjusted so that the mortar line color is darker than the two tiles' colors.

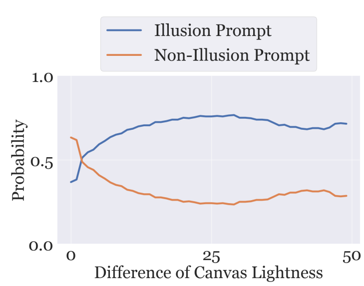
A.2 Illusion strength experiment

We only test out one image of each illusion in the previous experiments. However, as an illusion varies its illusion strength by doing small modifications, we want to observe whether CLIP tracks the illusion strength trend. In this experiment, we use pairs of prompts from the previous experiment and then classify a batch of images into one of two prompts. From the human case study experiment, we decided only to assess the Vasarely illusion and the simultaneous contrast illusion, which were proved to fool CLIP. For each illusion, we start with the lowest illusion-strength image and repeatedly apply the same modification to increase the illusion strength. For the Vasarely illusion, we increase the number of squares, while for the simultaneous contrast illusion, we increase the contrast between two canvases' color.

Figure 4 shows the trend of CLIP response to the illusionary change using two prompts' calibrated scores. In the Vasarely illusion and simultaneous contrast illusion, CLIP thinks the non-illusion prompt is more associated with the first few weak-illusion-effect images. Overall, in all illusions, we observe an upward trend for the illusion prompt and vice versa for the non-illusion prompt. This signifies that CLIP is able to detect the illusionary change and predict the trend in the correct direction. However, future studies on how humans associate textual descriptions with illusion change under the same modification are needed to determine the match of other attributes like the rate of change.



(a) Vasarely illusion



(b) Contrast illusion

Figure 4: For each illusion, we used the same pair of prompts that were engineered in section 3.3.

A.3 Survey

We used Google Forms to conduct the case study. The survey begins with one sample illusion, the checkerboard illusion, with one picture of the illusion, and sample non-illusion and illusion prompts. This is followed by four sections collecting the prompts for four illusions used in the paper. Each section includes a general description of what the illusion effect is, a section asking for three illusion prompts with a strong illusion strength photo, and a section asking for three non-illusion prompts with a weak illusion strength photo.

A.3.1 Survey form

Illusion Annotation

* Required

Survey Description

In this survey, you will annotate a set of illusions. Your response will be used as a label for the classification task of machine learning models (CLIP).

You should describe what you see in the photo in detail with information like objects, color, shape and etc.

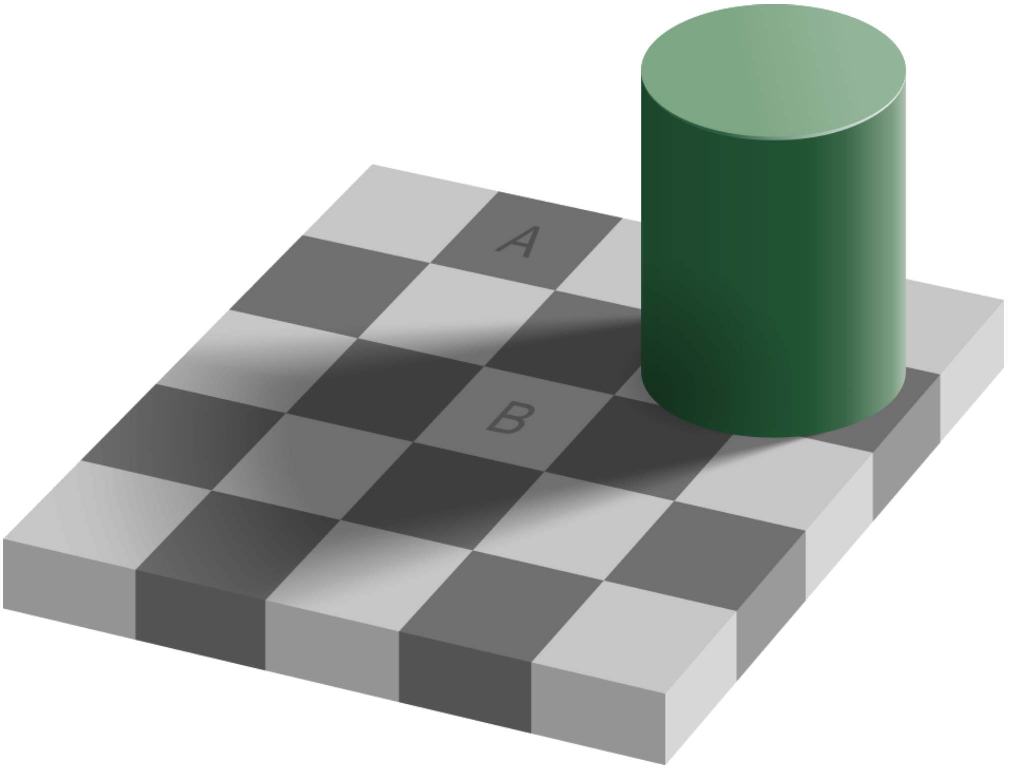
1. Name

Sample Illusion Description	In this section, we will describe a sample illusion as an example.

Checker Board Illusion

The image below depicts a checkerboard with two squares, A and B.

Squares A and B have the same brightness. However, **square A looks darker, while square looks lighter.** We would focus on this illusion property when we design the prompt.



Sample Illusion Prompts

Illusion prompts should acknowledge the existence of the illusion effect/property. Sample illusion prompts for checkerboard illusion:

- A photo of letter B on white background and letter A on gray background
- A photo of a white square containing letter B and a gray square containing letter A
- A photo of a checkerboard with a dark box containing letter A and a light box containing letter B

Sample Non-Illusion Prompts

Non-illusion prompts should ignore the illusion effect. Sample non-illusion prompts for checkerboard illusion:

- A photo of letter A on gray background and letter B on gray background
- A photo of a square containing letter B and a square containing letter A, both squares have the same brightness

Vasarely Illusion Description:

Vasarely illusion comprises of superimposed squares with ordered levels of brightness. As the number of squares increase, **an invisible X/star-like shape appears gradually**.

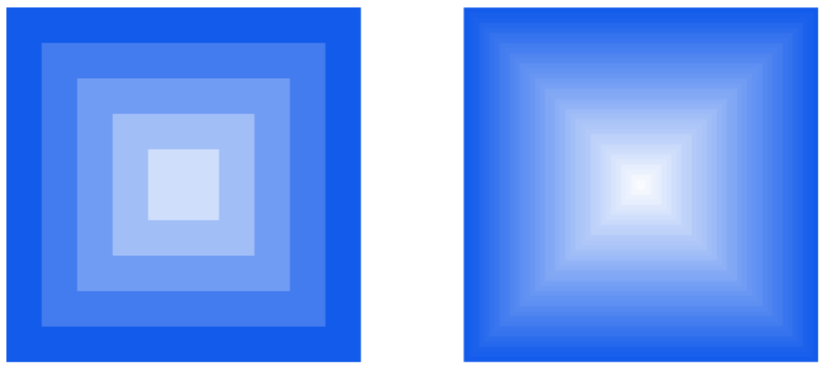
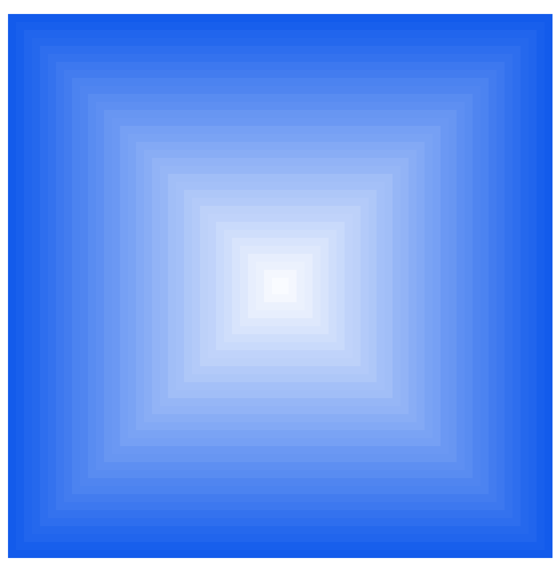


Image for Illusion Prompt Description



2. Description #1 *

3. Description #2 *

4. Description #3 *

5. Description #4

6. Description #5

Image for Non-Illusion Prompts Description

7. Description #1 *

8. Description #2 *

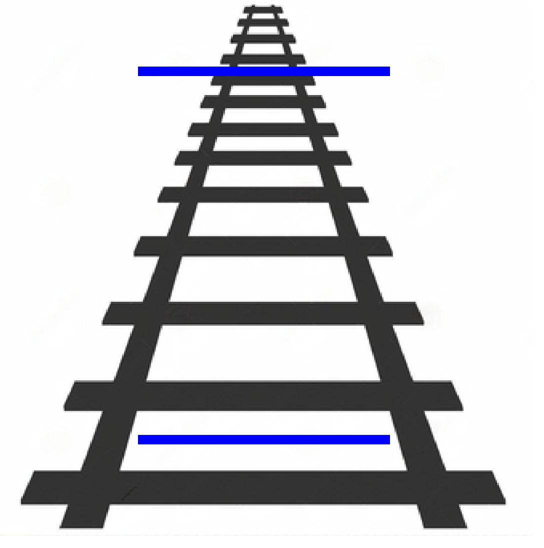
9. Description #3 *

10. Description #4

11. Description #3

Ponzo Illusion Description:	Ponzo illusion depicts a rail track/ladder where two parallel lines of the same length are laid out on the track one above the other. However, as the vertical distance between the lines increases, **the lower line looks shorter than the upper line**.

Image for Illusion Prompts Description:



12. Description #1 *

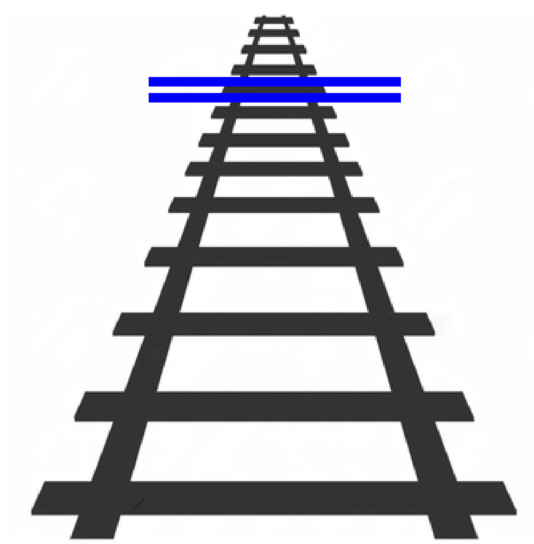
13. Description #2 *

14. Description #3

15. Description #4

16. Description #5

Image for Non-Illusion Prompts Description:



17. Description #1 *

18. Description #2 *

19. Description #3 *

20. Description #4

21. Description #5

Simultaneous Contrast Illusion:

This illusion consists of a square image separated in the middle into two equally sized areas with different brightness levels. Each area has a circle at the center, and both circles have an identical brightness. However, **the circles in the darker area appear to be lighter and vice versa**.

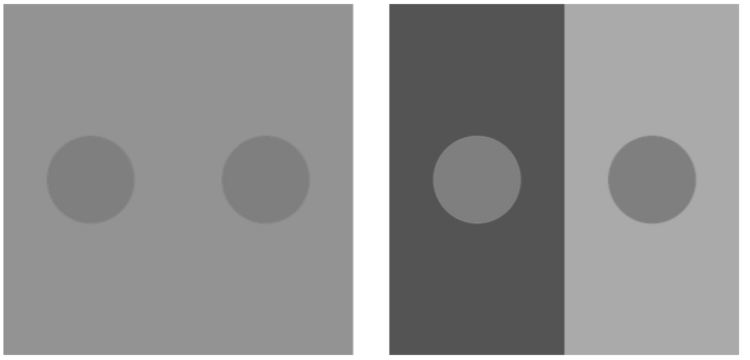
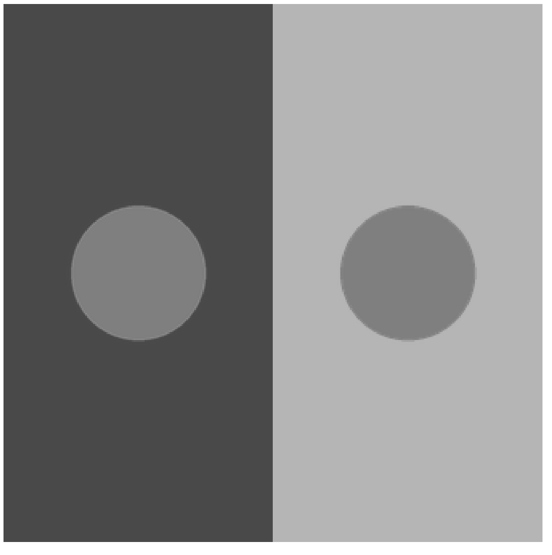


Image for Illusion Prompts Description:



22. Description #1 *

23. Description #2 *

24. Description #3 *

25. Description #4

26. Description #5

Image for Non-Illusion Prompts Description:

27. Description #1 *

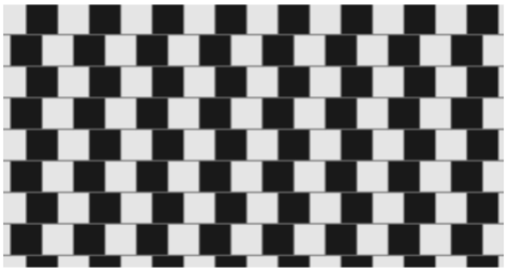
28. Description #2 *

29. Description #3 *

30. Description #4

31. Description #5

Café Wall Illusion:	The café wall illusion is a geometrical-optical illusion in which the parallel straight dividing lines between staggered rows with alternating dark and light "bricks" **appear to be sloped, not parallel as they really are.**



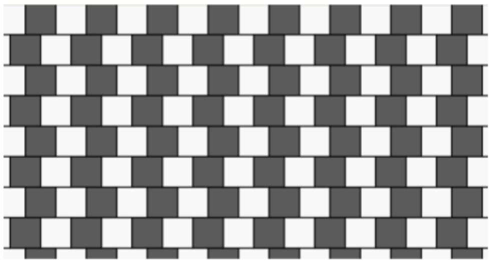
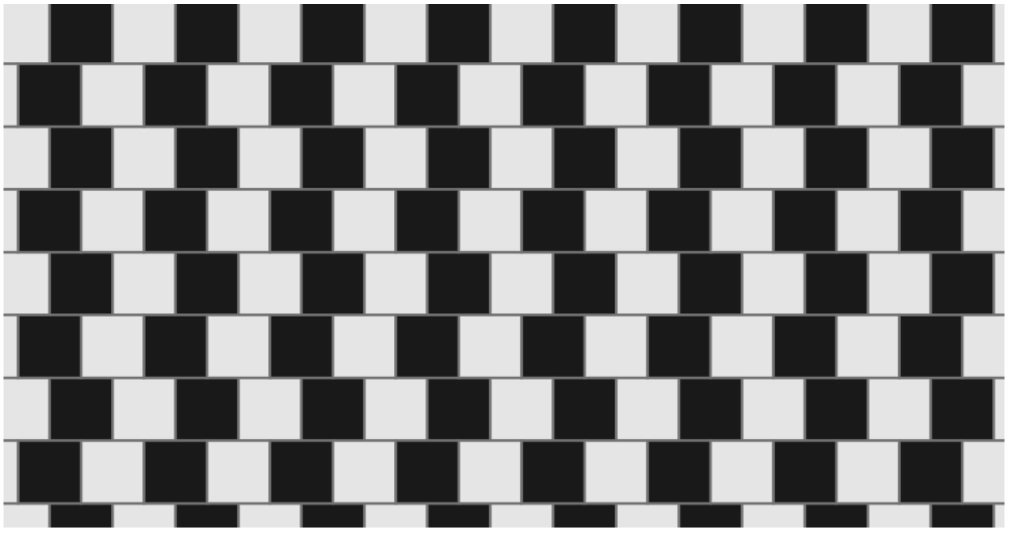


Image for Illusion Prompts Description:



32. Description #1 *

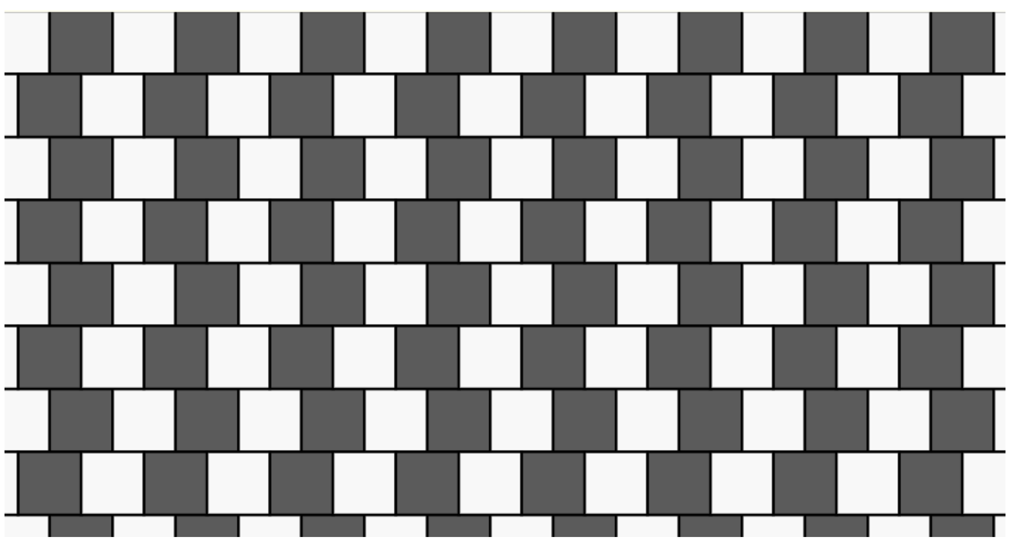
33. Description #2 *

34. Description #3 *

35. Description #4

36. Description #5

Image for Non-Illusion Prompts Description:



37. Description #1 *

38. Description #2 *

39. Description #3 *

40. Description #4

41. Description #5

A.3.2 Survey responses

Vasarely illusion:

Illusion prompts:

- White X shape
- a set of boxes stacking on each other, with the diagonal being brighter
- Many superimposed squares with decreased sizes and increased levels of brightness.
- blue color gradient
- X shape on blue background
- a nested squared making an X shape
- A white, X-shaped light with blue background that is formed by superimposing squares.
- star with a faded center
- X shaped centered in the center of the image
- a blue box containing a diagonal X shape
- A cross-shaped light with blue background that's formed by many squares.
- stacked squares with gradient colors

Non-illusion prompts:

- Stacked squares
- nested boxes with top layers having lighter color
- A few superimposed squares with decreased sizes and increased levels of brightness.
- squares colored shades of blue stacked on one another
- "Concentric" squares
- a set of layered boxes ranging from deep blue to light blue
- 5 blue squares. The smaller the square, the brighter the colors.
- dark blue border with fading center
- Squares of different sizes and brightness
- five square boxes stacking together
- Stack 5 blue squares together. The top one should have smaller size and brighter color.
- 5 progressively smaller and faded blue squares stacked

Ponzo illusion:

Illusion prompts:

- Top line is longer than the bottom line
- the upper blue line is longer than the lower blue line
- A track with two blue lines where the lower one looks shorter than the top one while they actually have the same length.
- top line is longer than lower line
- Top line looks longer because of the perspective
- two lines with the upper one longer
- A track with two blue lines where the upper one looks longer than the lower one while they actually have the same length.
- two parallel lines of varying lengths on a railway track
- Bottom line looks shorter and thiner
- two lines lie on a rail road with the upper one being longer

- A track with two blue lines which actually have the same length but the top one looks longer than the lower one.
- two parallel lines of varying lengths on a ladder

Non-illusion prompts:

- Two lines close to each other
- two lines lying on the rail road
- A track with two blue lines that are close to each other.
- two parallel blue lines
- Two lines of equal shape
- two equivalent blue lines parallel to each other
- Two blue lines that are close to each other and sit on top of a track.
- two parallel blue lines of same length on a ladder
- Two horizontal blue lines of same length
- two blue lines sitting together with equal length
- A track with two blue lines that have small distance.
- two parallel blue lines of same length on a railway track

Simultaneous Contrast Illusion:

Illusion prompts:

- Two circles on differently lit backgrounds
- the right circle is darker than the left one
- Place two circles that have equal brightness in one dark region and the other light region.
- light grey circle on dark grey background on the left and dark grey circle on light grey background on the left
- Circle over darker background looks brighter
- two circles in the image with the left one to be brighter
- Split the image into a bright region and dark region. Place the same circle inside each region.
- color contrast background and circle
- Circle over bright background looks darker
- the left circle is in grey color while the right one in dark grey
- Draw left half of the image with a dark background and place a gray circle inside it. Draw the other half of the image with a light gray background and place the same circle inside it.
- right side dark grey circle on light grey background on left side light grey circle on dark grey background

Non-illusion prompts:

- Two circles
- two identical circles in both shape and color
- Pace two circles inside a gray background.
- two circles on a colored background
- Identical circles side by side
- two dark grey circles lying in a light grey canvas
- Draw two circles.
- light grey background with two dark grey circles
- Two dark circles in a square

- two circles with the same dark grey color
- Draw one circle on the left of the image and draw the other one on the right.
- two dark circles on a light background

Café wall illusion:

Illusion prompts:

- Non-aligned Checkerboard pattern
- sloped brick lines parallel to each other
- Generate a set of black-white grids that appear not be arranged regularly. However, they are actually arranged regularly.
- sloped checkered bricks
- Checkerboard with non-parallel lines
- lines sloped toward two opposite directions are interleaved
- Draw a set of black-or-white grids that appear to be sloped, not parallel as they really are.
- skewed asymmetric checkered pattern
- Curved horizontal lines
- No blocks of the same color should sit with each other. Draw these blocks as if they are sloped but they are actually not.
- non-parallel checkered pattern

Illusion prompts:

- Checkerboard pattern that is not aligned vertically
- parallel and staggered rows
- Draw a set of black-and-white blocks.
- faded parallel checkered pattern
- Parallel horizontal lines
- Draw a set of black-and-white squares.
- parallel checkered pattern with faded black and white
- Horizontal stripes
- staggered rows that are parallel to each other
- Fill the image with black-and-white squares.
- dark grey and white checkered parallel pattern