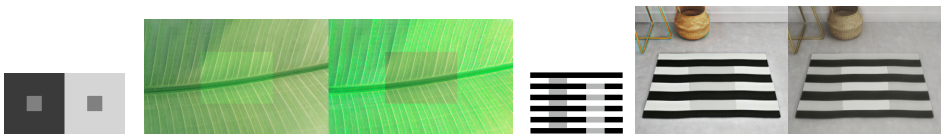
A Statistical Story of Visual Illusions —
A unified approach for explaining visual illusions & generating them

Elad Hirsch and Ayellet Tal

Technion – Israel Institute of Technology
{eladhirsch@campus,ayellet@ee}.technion.ac.il



(a) Simultaneous-Contrast Illusion (b) White's Illusion

Fig. 1. Examples of visual illusions. (a) Look at the central rectangles. Do they have the same intensity/color? It seems they do not, but they actually do. (b) The same effect happens for the gray long rectangles—they are perceived as having different intensities, although their intensities are equal. This paper provides a general, data-driven explanation to these (and other) illusions. It also presents a model that generates illusions by modifying given natural images in accordance with the above explanation (the right illusions in (a) & (b)). Please watch the illusions on a computer screen.

Abstract. This paper explores the *wholly empirical paradigm* of visual illusions, which was introduced two decades ago in Neuro-Science. This data-driven approach attempts to explain visual illusions by the likelihood of patches in real-world images. Neither the data, nor the tools, existed at the time to extensively support this paradigm. In the era of big data and deep learning, at last it becomes possible. This paper introduces a tool that computes the likelihood of patches, given a large dataset to learn from. Given this tool, we present an approach that manages to support the paradigm and explain visual illusions in a unified manner. Furthermore, we show how to generate (or enhance) visual illusions in natural images, by applying the same principles (and tool) reversely.

1 Introduction

The physical state of the world differs from our subjective measure of its properties, such as lightness, color, and size, as demonstrated in Fig. 1. This gap leads to visual illusions, which is a fascinating phenomenon that plays a major role in the study of vision and cognition [5]. This, in turn, has made human-made

illusions very popular; they are used for education, in business, and as a form of (op-)art and of entertainment (e.g. "Illusion of the Year Contest") [30,38,39].

Illusions come in a wide variety. In this paper we focus on those that are caused by the lack of sufficient information to reconstruct the real scene. Reality, as projected on the retina, can represent an infinite number of scenes, thus reconstruction is an ill-posed inverse problem. For instance, the color of a pixel in an image may correspond to an infinite number of combinations of environmental illumination, surface reflectance and atmospheric transmittance.

A variety of explanations have been proposed for visual illusions, mostly in Psychology and Neuro-biology [17,24,20,4,12]. In this paper, we follow the *wholly empirical paradigm* proposed by [25]. Briefly, our interpretation of the scene is highly dependant on our lifetime experience. Thus, the statistics of the projections of natural scenes and the likelihood of patches in them determine our perception. Visual illusions are caused when the ill-posed inverse problem is resolved statistically and this interpretation contradicts the original stimuli.

For instance, in Fig. 1(a) two identical gray boxes are surrounded by different backgrounds. They are perceived differently: The one on the bright background looks darker than the other. Fig. 1(b) demonstrated the opposite effect, as identical gray blocks look lighter when surrounded by an overall brighter background. How would one explain it? It turns out that natural patch statistics can explain both illusions and many more.

This *wholly empirical paradigm* was supported by a set of experiments with limited data. The era of deep learning and big data opens new opportunities, providing the means to support the theory based on solid grounds. This is a major goal of this paper—give a general and unified explanation to a variety of seemingly-unrelated visual illusions. We note that this goal is inline with the classical motivation of computer vision algorithms—mimicking cognitional mechanisms, some of which might be affected from this reality-perception gap.

Towards this end, we introduce a statistical tool that estimates the likelihood of image patches to occur, based on the learned statistics of natural scene patches. We will show how this tool assists us to explain three different types of well-known visual illusions.

An important property of our proposed tool is being reversible. This enables a controlled statistical manipulation of image patches, i.e. making a patch more (or less) likely with respect to a natural dataset. Thus, it allows us not only to explain visual illusions, but also to create ones, as illustrated in Fig. 1. Unlike the synthetic illusions that can be found in textbooks, our generated illusions appear as natural images, as after all, these are the illusions of our everyday life.

We are not the first in computer vision to be fascinated by visual illusions. In 2007, Corney et al [8] proposed to use a shallow neural network and predict surface reflectance in synthetic images. This work managed to show that this network was deceived by several lightness illusions similarly to humans. A decade later, Gomez-Villa et al [14] trained deep convolutional neural networks (CNNs) on datasets of natural images. The networks were trained to perform low-level vision tasks of denoising, deblurring and color constancy. However, it

was demonstrated that each network (one trained for denoising, one for deblurring and one for color constancy) is not deceived by all the illusions. One may view these works as implicit support to the connection between the statistics of natural images and visual illusions. We will show that a single empirical method can explain all of these illusions.

In [22,37] a video frame prediction network was proposed, which could predict illusory motion. In [40] a GAN was trained to generate illusions out of a dataset of visual illusions. It was claimed that this approach is unable to fool human vision. In [15] synthetic visual illusions were generated, by adding an illusion discriminator that quantifies the perceptual difference between two target regions [14], to a GAN that generates backgrounds for the targets. The choice of the pre-trained illusion discriminator and the balance of the losses of the discriminators lead to different kinds of results, thus lacking generality.

Our method enables us to apply the same unified mechanism found in synthetic illusions on natural images, in order to demonstrate the effects in natural contexts. This complements the consistent explanation to these illusions. Thus, this paper makes three key contributions:

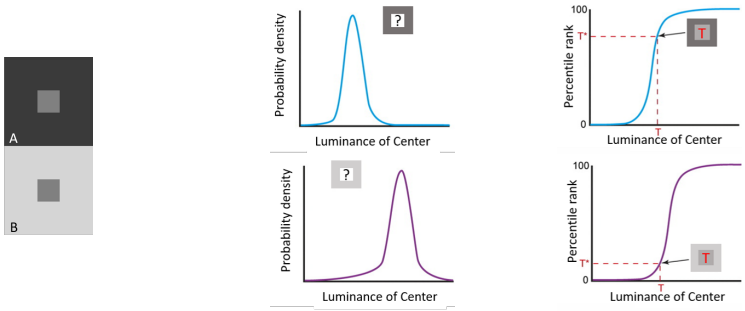
1. It introduces a novel statistical tool for estimating the likelihood of image patches (Section 3). This should be the basic building block of the empirical paradigm of visual illusions, however such a tool did not exist until now.
2. It supports the empirical paradigm, using a large dataset of natural images, and demonstrates it for three different illusions (Section 4).
3. It proposes a general method to automatically generate "natural" visual illusions, by manipulating the likelihood of image patches (Section 5).

2 Background: Wholly Empirical Paradigm of Vision

Various paradigms have been proposed for explaining visual illusions [17,24,20,4,12,27]. In this paper we focus on the *wholly empirical paradigm,* presented by Purves et al. [27,26,29], which is powerful in its ability to explain a whole range of illusions. This section briefly introduces it.

This theory is motivated by the observation that a 2D projection of a 3D world makes the inverse optical problem highly ill-posed. For instance, the measured luminance of a surface corresponds to an infinite number of combinations of environmental illumination, surface reflectance and the atmospheric transmittance. Similarly, the projection of a line on the retina at a certain length and orientation may correspond to an infinite number of possible 3D lines in the world, in different lengths, distances and orientations.

Hence, visual perception does not aim to recover real world properties (e.g., color, length and orientation) explicitly, as it is impossible. Instead, its role is to promote useful behavior out of the 2D retinal image. In other words, vision generates useful perception for successful behaviors without recovering real-world properties. This requires a continuous learning of the frequency of occurrence



(a) simultaneous-contrast illusion (b) probability of center (c) percentile rank of center

Fig. 2. Empirical explanation of contrast illusion [29]. (a) The same gray area in the center is perceived lighter in the dark surrounding than in the bright surrounding. (b) The probability of the luminance value of the center when surrounded by dark background (top) and by bright background (bottom): The peak is at the value of the surrounding. (c) The percentile rank ($\propto$ cumulative probability) of each probability: For a given center luminance value T, the percentile rank is higher under dark background than under bright one. Therefore, the same gray area T appears lighter when surrounded by dark background.

of image patterns and their impact on survival. Biological vision, for this purpose, relies on recurring scale-invariant patterns within images to rank perceptual properties. Studying the statistics of natural scenes reveals environmental behaviors that the unconscious observer deduct over experience. This theory is supported by psycho-physical studies that find a remarkable connection between perceptual properties and the likelihood of measured physical properties.

This theory also provides explanations for a large variety of visual illusions [26,25,16,42,41,33,28]. These are explained through the likelihood of the occurrence of relevant patches. For instance, the basic simultaneous-contrast visual illusion in Fig. 2(a) is composed of two equal gray areas with different surroundings. The area surrounded by the darker surrounding looks lighter. Why is that? We examine the probability of the intensity of the center area of a patch in natural images, given a surrounding (Fig. 2(b)). As expected, the maximum probability will be attained when the center has the same intensity of that of the surrounding (i.e., creating a smooth patch), and decreases as it gets farther away. This implies that for a given gray level T, the *percentile rank* in T under dark surrounding is larger than that under bright surrounding (Fig. 2(c)); the *percentile rank* of a score is defined as the percentage of scores that are smaller or equal to it. This rank corresponds to the perceived intensity. The same concept holds for other properties, such as color, size, orientation and more.

To support this theory, statistics of patches in the real world must be provided. In the era pre-big data, this was not easy to do. Thus, the empirical analysis was evaluated on relatively small datasets (up to 4000 images) [36]. Furthermore, the probability of each property (lightness, size, etc.) was estimated

by applying uniform sampling of patches or templates of patches and estimating the likelihood function of the property according to its relative occurrence in the sampled patches; for instance, how many times the center of a patch had value x when its surrounding was uniformly y. This exhaustive uniform sampling has a couple of limitations. First, it requires sampling thousands of image patches that approximately fit each template (e.g. a uniform background of value y). Second, it relies on a marginal distribution of patches that fit the template and therefore neglects more complex relations during the estimation process.

Section 4 provides empirical support for the *wholly empirical* paradigm, by utilizing the current rich natural datasets (of over 1.5M images of natural scenes) and state-of-the-art computer vision tools. Our model naturally overcomes the drawbacks of uniform sampling. It therefore opens new opportunities to explain and demonstrate the phenomenon of visual illusions in an empirical manner.

3 Measuring Patch Likelihood

Recall that according to the empirical approach, visual illusions depend on the frequency of recurring patterns in projections of natural scenes, which determines their likelihood. How shall the likelihood of patches be measured? While patch likelihood is not measured explicitly in computer vision tasks, implicitly, it has a tremendous importance in applications of image restoration [9,44,32]. We seek after a general explicit method, which is not application-dependent. This method should also overcome the shortcoming of the uniform sampling, discussed in Section 2. Furthermore, we require that this method would have generative capabilities, in order for it not only assist to explain visual illusions, but also to generate ones.

Section 3.1 proposes a patch likelihood estimation model that can efficiently and accurately learn a high-dimensional distribution of a large dataset of natural scene patches. This raises an interesting question of how to evaluate the behavior of the proposed model. In the patch case, visual assessment of sampled patches is irrelevant and a quantitative ground truth does not exist. In Section 3.2 we introduce two measures, one is quantitative and the other is qualitative.

3.1 Framework

Since we aim at explicitly estimating the likelihood of properties (e.g., intensity, saturation etc.), as well as modifying these properties, we turn to likelihood-based generative models. These models can be classified into three main categories: (1) *Autoregressive models* [35,34] (2) *Variational Autoencoders (VAEs)* [6,18] and (3) *Flow-based models* [19,10,11]. We focus on the flow-based model, for three reasons: First, it optimizes the exact log-likelihood of the data. Second, the model learns to fit a probabilistic latent variable model to represent the input data, which is important for the specific generative property we seek after. Third, the model is reversible.

In particular, we base our framework on *Glow* [19], a recent flow-based architecture. Hereafter, we briefly introduce the theory behind this model, adapted to our patch case (as [19] handles full images). Let x be a patch, sampled from an unknown distribution of natural scene patches $p^*(x)$. Let D be a dataset of samples $\{x_i|i = 1, ..., N\}$ taken from the same distribution. We look for a model $p_\theta(x)$ that will minimize

$$\mathcal{L}(D) = \frac{1}{N}\sum_{i=1}^{N} -\log(p_\theta(x_i)). \tag{1}$$

The generative process is defined by the latent variable $z \sim p_\theta(z)$, where $p_\theta(z)$ is a simple multivariate Gaussian distribution $N(0, I)$. The transformation of the latent variable to the input space is done by an invertible function $x = g_\theta(z)$ s.t.

$$z = g_\theta^{-1}(x) = f_\theta(x). \tag{2}$$

The function $f_\theta(x)$ is a composition of K transformations, termed a *flow*. We denote the output of each inner transformation f_i with h_i. Then, the probability density function of the model, given a sample x is:

$$\log(p_\theta(x)) = \log(p_\theta(z)) + \sum_{i=1}^{K} \log(|\det(dh_i/dh_{i-1}|). \tag{3}$$

For the families of functions used in flow-based architectures, this term it efficient to compute, and both the forward path and the backward path are feasible.

Implementation. Fig. 3 illustrates our model. It is based on the architecture of Glow, with a single *flow* and $K = 32$ composed transformations. The input consists of image patches of size 16×16, a size that manages to capture textured structures and to allow stable training. The network was trained on random patches, sampled from *Places* [43], which is a large scene dataset.

3.2 Evaluation

There is no ground truth for the likelihood of patches. Hereafter, we propose two measures that may be used to evaluate the performance of our model.

Quantitative evaluation—Center of patch test. Many of the experiments of [26] were based on direct uniform sampling of many natural scene image patches and calculating the probability of some hand-crafted features, e.g. calculating the probability of the color of the center area of a patch, given its surroundings (Fig. 2). This approach requires sampling of many patch templates, for instance different colors of the center area or the surrounding.

We propose a similar, but much simpler approach: We generate the target patches with different values of the center and the surrounding. These are injected to our pre-trained network, which provides us with a likelihood score for each patch. This approach can be used to explain many visual illusions, with

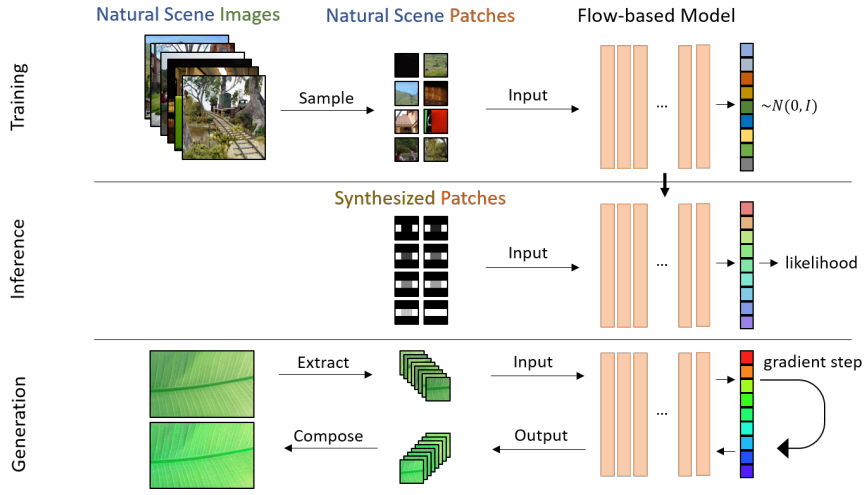


Fig. 3. Framework. During training, the likelihood of natural scene patches is learned. During inference (Section 4), this pre-trained model is fed with synthetic patches that represent common visual illusions, and estimates their likelihood. To generate visual illusions given an image (Section 5), the image patches are first extracted and their corresponding latent variables are manipulated with a gradient step. They are then fed back into the reverse path, in order to reconstruct a new image.

slight adjustments for each illusion. Section 4 will demonstrate that our results are amazingly similar to the empirical experiments of [26] for the simultaneous-contrast illusion they studied.

Qualitative evaluation—Min-Max patch test. Sorting patches of an image by their network's scores may also provide a sanity check for the network's behavior. We would expect smooth patches to be more likely than textured ones, and among the textured patches we would expect a reasonable ranking—one that expresses the learned statistics. Since it is impossible to determine whether a ranking is reasonable with respect to a large dataset of images, we propose to determine it by training on patches of a single image. This would allow visual evaluation of the results. Furthermore, a comparison of the ranking of the internal statistics (relative to single image) to that of the external statistics (relative to the entire dataset) could help in the evaluation of our tool.

Fig. 4 demonstrates the results on two images. For each image, we present 8 random patches from the internal most/least likely 100 patches and 8 patches from the external most/least likely 100 patches (trained on Places). The most likely internal and external patches are very similar—they are both very common in the source image and are relatively smooth. There is, however, a clear difference between the internal and the external least likely patches. The internal least likely patches are indeed very unique in the source images (the white windows and the red-on-orange pattern). This confirms our sanity check! We may

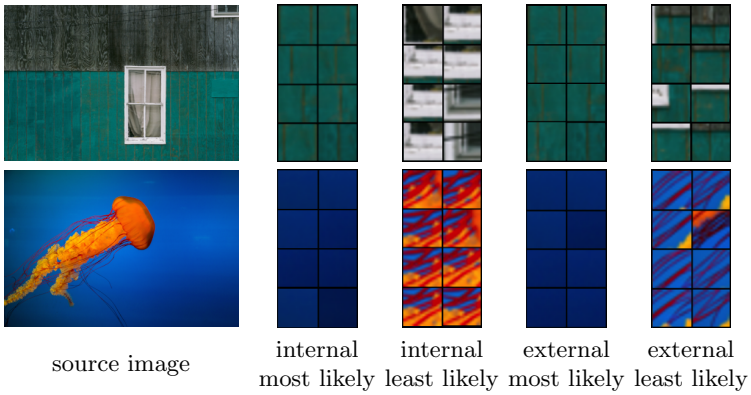


Fig. 4. Min-Max patches. Samples from 100 most-likely and 100 least-likely patches of images, w.r.t. the internal image statistics (trained on a single image) & the external statistics (trained on Places). While the most-likely patches are smooth in both cases, the least-likely patches differ. For instance, white patches (top) are very likely in the world, but are unlikely in the image. Images taken from [1].

now believe our tool, according to which the external least likely patches are the green-brown-white and red strings on blue background, respectively. Our results indicate that these patches are unique in the world, even though they appear much more in the source image.

4 Explaining Visual Illusions

This section attempts to support the empirical paradigm regarding deceived perception of common visual illusions, equipped with our deep learning-based probabilistic tool. The underlying assumption is that the statistics of features in large datasets of natural scenes (Places [43] in our case) is similar to the feature statistics in the "dataset" of retinal images. We introduce a general technique for explaining visual illusions, given information regarding patch statistics. Furthermore, we demonstrate the statistical reasoning on three common visual illusions, focusing on intensity and color illusions. For each illusion we ask what the likelihood is of the illusory pattern to appear in a projection of a natural scene. We note that a statistical explanation of the first illusion was given by [26], whereas statistical explanations of the other two were not given in the literature.

Method. Our method proceeds as follows: Given an illusion, we define an illusion-dependent template & target. For instance, in the case of Fig. 1(a), the template is the rectangular surrounding and the target is the inner rectangle. Then, we generate 256 instances of the template with the same surrounding (context) but with different values (intensity, saturation, hue etc.) of the target area. This yields 256 patches, which differ from one another only in the target area. Our goal is to evaluate the likelihood of these patches, as it expresses the

probability of the target values given a specific context. This is done by providing the pre-trained network from Section 3 with these patches as input. The system returns the likelihood of this pattern.

Recall that perception reacts according to the *percentile rank* (Section 2). Let A, B be two different backgrounds of the same target area, having value T. If the percentile rank of value T in background A is higher than the percentile rank of value T in background B, this means that statistically we expect the target area to have a higher value (e.g. lighter, in the case of intensity) in A than in B. Therefore, the perceived value in the target will be higher in A than in B. In terms of the likelihood function, this means that the peak of the likelihood value in A is attained in a lower value than in B, as discussed in Section 2.

Illusions. Hereafter we demonstrate the results of our method on three illusions.

1. Simultaneous lightness/color contrast illusion [13,23]. In this illusion, two identical patches are placed in the center of different backgrounds. While the color of these central patches is the same, it appears darker when surrounded by a brighter background than by a darker background (Fig. 5(a)).

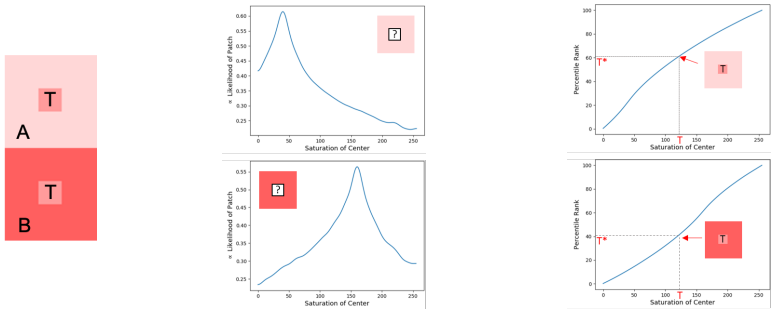
Our experiment is performed both on the hue, saturation and value (in HSV color space) and on the intensity (in gray-scale). The template is a uniform background and a uniform center area (target). For each property (e.g. hue), we set a value in the range $[0, 255]$ and generate 256 patches in which the background has this value; the value of the center of the patch increases from 0 to 255.

Fig. 5(b) shows the likelihood graphs for (a), as outputted by our network, for saturation. The likelihood is maximal when the center and the surrounding have the same saturation and drops as they get farther from each other. As shown in Fig. 5(c), the same saturation T of the center areas would have a higher percentile rank when surrounded by less saturated (lighter) background than by more saturated background; hence, it would be perceived as darker. The same analysis applies to the hue & value properties of the HSV color space.

2. White's illusion. In this illusion, black and white horizontal bars are interrupted by identical target gray blocks (Fig. 6(a)). The gray blocks appear darker when they interrupt the white bars and lighter when they interrupt the black [3]. Interestingly, the illusory effects of the edges of the target gray blocks in White's illusion and in the simultaneous-contrast illusion, are reversed. Here, when the target block has more of dark edges it looks darker and when it has more of bright edges it looks lighter.

We aim to show that when a rectangular gray target patch is surrounded by black bars from below and from above and by white on the sides, it appears darker than in the inverse (B&W) case. Therefore, in the first template, the pixels of the top & bottom thirds are black and the middle bar is split horizontally, such that left and the right quarters are white. The target area interrupts the middle bar and its value increases from 0 to 255, leading to 256 patches. In the second template, the roles of the black and the white are reversed.

Fig. 6(b) presents our results–the likelihood graphs of the target value, given its surrounding. These graphs support the findings: When the gray block interrupts the white bar, it is more likely to be light, thus it has a low percentile



(a) Contrast illusion (b) Likelihood of background (c) Percentile rank of background

Fig. 5. Experiment—contrast illusion (a) The backgrounds have the same hue & value and different saturation. The center, of color T, is perceived more saturated on the top than on the bottom, though it is the same. (b) shows the likelihood of the saturation value of the center when surrounded by the top/bottom background; the peaks are when the value T is the same as the background. (c) Since the percentile rank of T on top is higher than on the bottom, it will be perceived as more saturated.

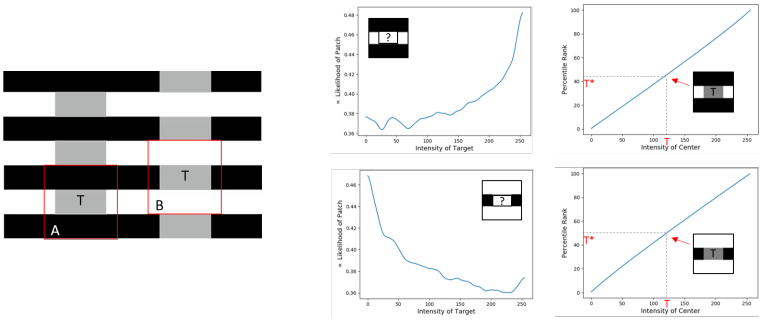
rank, and vise versa (Fig. 6(c)). As before, this low percentile rank explains the dark appearance when interrupting the white bar, while the high rank explains the light appearance when interrupting the black bar.

3. Hermann grid. Fig. 7(a), which consists of uniformly-spaced vertical and horizontal white bars on a black background, illustrates this illusion. Stare at an intersection; this intersection appears white when it is in the center of gaze, but gray blobs appear in the peripheral intersections [31]. This illusion relates not only to the statistics of patches, but also to the receptive field, which is smaller in the center of gaze than in the periphery.

Therefore, to explain this illusion we must also emulate the receptive field. This is done by considering a low-scale image as effectively corresponding to a large receptive field, and a high-scale image as corresponding to a smaller receptive field. We therefore feed our network first with patches of a high-scale grid image (512×512) and then with patches of a low-scale image (256×256), but with the same patch size.

The difference in the likelihood maps of the two cases can be observed in Fig. 7(b)-(c), as heat-maps for each patch. In high-scale, the white intersections are highly likely (brown). This is not surprising, as it represents the small receptive field, which captures the center of the intersections as smooth (& likely) white patches. However, in low scale, the white intersections become unlikely (yellow). This is so because white crosses on black backgrounds are indeed unlikely in real life.

To explain why the peripheral intersections look darker, we employ our method. The template in this case is a white cross on a black background, which looks like the intersection area in low scale. The target area is the center of the cross and its value increases from 0 to 255, leading to 256 different patches. Fig. 7(d) shows the results: the likelihood of the value of the center of the cross



(a) Illusion & Examined Pattern (b) Likelihood graphs (c) Percentile rank

Fig. 6. Experiment—White's illusion (a) The targets T have the same value, but they are perceived differently, depending on their surrounding patterns A, B (in red). (b) The likelihood of the target grayscale intensity, when interrupting the white(/black) bar. (c) These graphs explain why, when the target area interrupts the white bar, it has a lower percentile rank, and hence it looks darker, and vise versa. The same statistical paradigm explains both White's illusion and the simultaneous-contrast illusion, although they cause an opposite contrast effect.

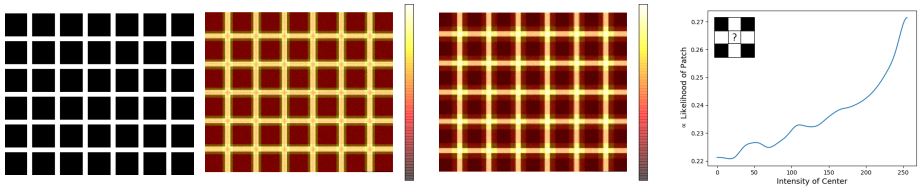
increases as the gray-scale value approaches white. In the periphery, where these intersections are not pure white [2], they would have a lower percentile rank and therefore would look darker, as before.

5 Generating Visual Illusions

Generating visual illusions is a grand challenge [15]. This section introduces a novel method for doing so. Furthermore, we aim at generating illusions in the context of natural images, by enhancing illusory effects in a given image. This requirement adds a couple new difficulties, in comparison to generating synthetic illusions. First, it is impossible to choose a uniform target area. Second, due to the amount of details in a natural image, the target area should be large in order to be noticeable. However, if the target is large, since the neighborhood of its inner parts do not change, the illusory effects might be reduced (since illusory effects depend on the surrounding, including adjacent inner parts).

The key idea of our approach is based on the principle of the empirical paradigm: We can generate illusory effects by controlling the likelihood of image patches. In particular, given an image, we could generate context (surrounding) that is slightly more likely or slightly less likely, as described hereafter.

Method. Given an image and target areas, the algorithm first extracts all the image's overlapping patches, except for those of the target. Second, these patches are fed-forward into our pre-trained network (Section 3), resulting with a latent variable z (Eq. 2) and a likelihood score for each patch. Third, a gradient step is performed on the latent variable, such that the associated patch's likelihood would slightly increase or slightly decrease. A manipulated patch is generated



(a) Illusion (b) NLL, high scale (c) NLL, low scale (d) Likelihood of intersection

Fig. 7. Experiment—Hermann grid. Stare at one specific white intersection (a). It appears white when it is in the center of the gaze, but illusory gray blobs appear when not at the center. In the patch *negative log-likelihood (NLL)* map of the grid, (b) the white intersection is likely (brown) in high scale, & (c) unlikely (yellow) in low scale. (d) presents the likelihood graph of the gray-scale value of the center of an intersection in low scale. Due to the low resolution in the peripheral vision, the value of the center has a lower percentile rank and would therefore be perceived as darker.

by injecting the above manipulated latent variable into the reverse pass of the network, which generates its corresponding manipulated patch. Finally, the manipulated patches compose an image that is similar to the input image, except that each patch (excluding the target) has a new likelihood.

The core of this method is the controlled likelihood manipulation of image patches. This operation is feasible due to (1) the reversibility of our network and (2) the form of the latent variable. We modify z, while our goal is to modify the likelihood of its corresponding patch x. This will indeed happen, since as observed by [10], regions with high density, in patch (input) space, shall also have a large log-determinant value and a large value of $p_\theta(z)$ (Eq. 3).

We use our prior knowledge regarding the distribution of the latent variable to manipulate input patches according to their likelihood. Let x be a patch and $z = f_\theta(x)$ be its latent representation. Manipulating the latent variable z to z' and back-projecting it with the reversible flow to the patch space result in a patch $x' = f_\theta^{-1}(z')$. Applying manipulation operation Ψ yields:

$$x' = f_\theta^{-1}(z') = f_\theta^{-1}\Big(\Psi\big(f_\theta(x)\big)\Big). \tag{4}$$

As the distribution of z is Gaussian (Section 3), Ψ is implemented as a simple gradient step:

$$z' = \Psi(z) = z + \eta \cdot \Big[- z \cdot e^{-\frac{z^2}{2}} \Big]. \tag{5}$$

The step size η determines the amount of likelihood manipulation. In our experiments, $\eta_1 = 0.6$ to increase the likelihood and $\eta_2 = -0.8$ to decrease it.

Results. Fig. 8 demonstrates our results. The manipulation is not limited to a single property of the image, such as hue, saturation, etc. Instead, changing the patches based on their likelihood, result in different changes in various regions, for different properties of the image, depending only on the input itself.

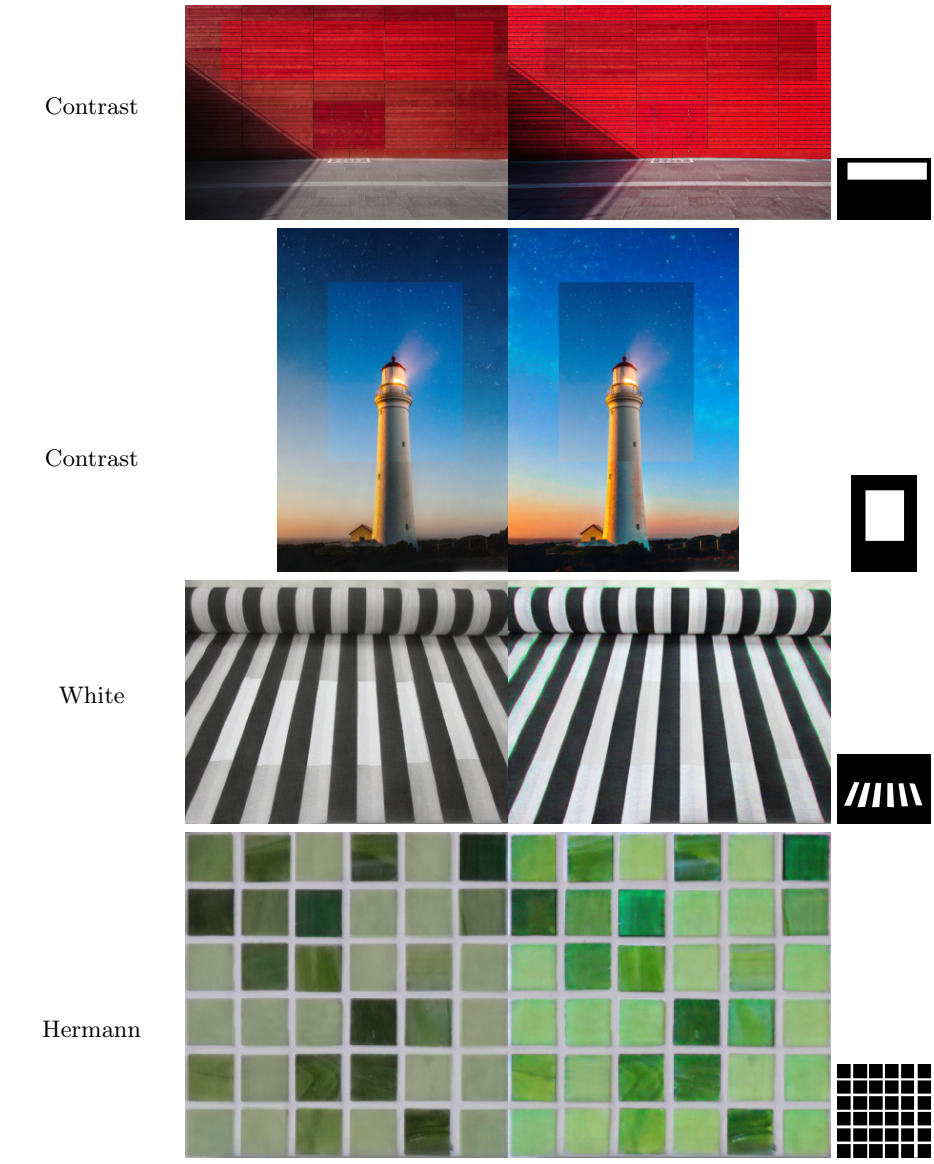


Fig. 8. Generated visual illusions. Each input image was manipulated patch-wise in order to change the surrounding of the masked area (white areas in (c)). Simultaneous contrast-like & White-like illusions: While the masked regions have the same color in the left and in the right images, they are not perceived as so. Hermann's-like illusion: The illusory gray blobs in the white intersections are enhanced in the left image and are reduced in the right. More examples are given in the supplemental material. Please watch this figure on a computer screen.

(a) some examples from [15] (b) some examples of ours

Fig. 9. Comparison of simultaneous-contrast illusions. (a) In [15] the background of identical target areas (rectangles or rings) is automatically generated, after training a GAN with an additional illusion discriminator on DTD [7] or CIFAR10 [21] datasets. (b) Our approach does not generate backgrounds from scratch. Instead, it manipulates the context (backgrounds) of any target area in a given natural image, creating two manipulated versions of the input, in which the target areas are identical.

Fig. 8(top two) shows simultaneous-contrast illusions. Two identical target areas (the white area in the mask) are perceived as having different colors, thanks to their different backgrounds. The left background was generated by manipulating the source image with η_1, and the right with η_2. Fig. 9 compares our result to that of [15], where illusions of this type were synthesized by adding a pre-trained block that acts as an illusion discriminator in a GAN framework.

Fig. 8(middle) illustrates a White-like illusion, which was generated by our system. Again, the left background was generated by manipulating the source image with η_1 and the right image by manipulation with η_2. In the result, the targets that interrupt the fabric's darker stripes (left) look lighter, although surrounded from right and left by brighter stripes than in the right image.

Fig. 8(bottom) demonstrates a Hermann grid-like illusion, as generated by our system. In these two natural grids the white lines are the same, but the colored squares were manipulated as before (left with η_1, right with η_2). The gray illusory blobs are enhanced in the left image, where the blocks were manipulated to be more likely, and reduced in the right.

6 Conclusions

The empirical paradigm of vision argues that human vision does not aim to better represent reality, but to statistically resolve an ill-posed inverse problem, even if it contradicts reality. In this paper we support this paradigm by proposing a unified method, which is able to explain a variety of visual illusions, by analyzing the statistics of image patches in big data. Furthermore, the paper shows that reversing the process, by changing the likelihood of patches in an image, manages to enhance visual effects for the same analyzed illusions. Both the support of the paradigm and the generation of illusions are possible thanks to a novel tool that measures the likelihood of image patches and has generative properties.

In the future, we intend to automatically choose the best target areas for the generation process; currently, the process depends on manual selection. Furthermore, more illusions could be studied using the proposed tool, both color-based and geometric (e.g. mis-perceiving size or direction).

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017)
2. Alonso, J., Chen, Y.: Receptive field. Scholarpedia **4**(1), 5393 (2009)
3. Anderson, B.L.: Perceptual organization and white's illusion. Perception **32**(3), 269–284 (2003)
4. Barlow, H.B., et al.: Possible principles underlying the transformation of sensory messages. Sensory communication **1**, 217–234 (1961)
5. Carbon, C.C.: Understanding human perception by human-made illusions. Frontiers in Human Neuroscience **8**, 566 (2014)
6. Chen, X., Kingma, D.P., Salimans, T., Duan, Y., Dhariwal, P., Schulman, J., Sutskever, I., Abbeel, P.: Variational lossy autoencoder. In: 5th International Conference on Learning Representations, ICLR (2017)
7. Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., , Vedaldi, A.: Describing textures in the wild. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2014)
8. Corney, D., Lotto, R.B.: What are lightness illusions and why do we see them? PLoS computational biology **3**(9), e180 (2007)
9. Deledalle, C.A., Denis, L., Tupin, F.: Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. IEEE Transactions on Image Processing **18**(12), 2661–2672 (2009)
10. Dinh, L., Krueger, D., Bengio, Y.: NICE: non-linear independent components estimation. In: 3rd International Conference on Learning Representations, ICLR (2015)
11. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real NVP. In: 5th International Conference on Learning Representations, ICLR (2017)
12. Gibson, J.J.: The ecological approach to visual perception: classic edition. Psychology Press (2014)
13. Gilchrist, A.: A gestalt account of lightness illusions. Perception **43**(9), 881–895 (2014)
14. Gomez-Villa, A., Martin, A., Vazquez-Corral, J., Bertalmio, M.: Convolutional neural networks can be deceived by visual illusions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
15. Gomez-Villa, A., Martín, A., Vazquez-Corral, J., Malo, J., Bertalmío, M.: Synthesizing visual illusions using generative adversarial networks. arXiv preprint arXiv:1911.09599 (2019)
16. Howe, C.Q., Purves, D.: Perceiving geometry: Geometrical illusions explained by natural scene statistics. Springer Science & Business Media (2005)
17. Hubel, D.H., Wiesel, T.N.: Brain and visual perception: the story of a 25-year collaboration. Oxford University Press (2004)
18. Kingma, D.P., Welling, M.: An introduction to variational autoencoders. Foundations and Trends in Machine Learning **12**(4), 307–392 (2019)
19. Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In: Advances in Neural Information Processing Systems. pp. 10215–10224 (2018)
20. Kording, K.P.: Bayesian statistics: relevant for the brain? Current opinion in neurobiology **25**, 130–133 (2014)

21. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images (2009)
22. Lotter, W., Kreiman, G., Cox, D.: A neural network trained to predict future video frames mimics critical properties of biological neuronal responses and perception. arXiv preprint arXiv:1805.10734 (2018)
23. Lotto, R.B., Purves, D.: An empirical explanation of color contrast. Proceedings of the National Academy of Sciences **97**(23), 12834–12839 (2000)
24. Marr, D.: Vision: A computational investigation into the human representation and processing of visual information **2**(4.2) (1982)
25. Purves, D., Lotto, R.B.: Why we see what we do: An empirical theory of vision. Sinauer Associates (2003)
26. Purves, D., Lotto, R.B.: Why we see what we do redux: A wholly empirical theory of vision. Sinauer Associates (2011)
27. Purves, D., Morgenstern, Y., Wojtach, W.: Perception and reality: Why a wholly empirical paradigm is needed to understand vision. Frontiers in Systems Neuroscience **9** (11 2015)
28. Purves, D., Shimpi, A., Lotto, R.B.: An empirical explanation of the cornsweet effect. Journal of Neuroscience **19**(19), 8542–8551 (1999)
29. Purves, D., Wojtach, W.T., Lotto, R.B.: Understanding vision in wholly empirical terms. Proceedings of the National Academy of Sciences **108**(Supplement 3), 15588–15595 (2011)
30. Rabin, M.: The nobel memorial prize for daniel kahneman. The Scandinavian Journal of Economics **105**(2), 157–180 (2003)
31. Schiller, P.H., Carvey, C.E.: The hermann grid illusion revisited. Perception **34**(11), 1375–1397 (2005)
32. Sulam, J., Elad, M.: Expected patch log likelihood with a sparse prior. In: International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition. pp. 99–111. Springer (2015)
33. Sung, K., Wojtach, W.T., Purves, D.: An empirical explanation of aperture effects. Proceedings of the National Academy of Sciences **106**(1), 298–303 (2009)
34. Van Den Oord, A., Kalchbrenner, N., Espeholt, L., kavukcuoglu, k., Vinyals, O., Graves, A.: Conditional image generation with pixelcnn decoders. In: Advances in Neural Information Processing Systems 29, pp. 4790–4798 (2016)
35. Van Den Oord, A., Kalchbrenner, N., Kavukcuoglu, K.: Pixel recurrent neural networks. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. p. 17471756 (2016)
36. Van Hateren, J.H., van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society of London. Series B: Biological Sciences **265**(1394), 359–366 (1998)
37. Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M., Tanaka, K.: Illusory motion reproduced by deep neural networks trained for prediction. Frontiers in Psychology **9**, 345 (2018)
38. Webpage: Best illusion of the year contest, `http://illusionoftheyear.com/`
39. Webpage: Prof. akiyoshi kitaoka's illusion pages, `http://www.ritsumei.ac.jp/akitaoka/index-e.html`
40. Williams, R.M., Yampolskiy, R.V.: Optical illusions images dataset. CoRR **abs/1810.00415** (2018)
41. Wojtach, W.T., Sung, K., Purves, D.: An empirical explanation of the speed-distance effect. PLoS One **4**(8) (2009)

42. Wojtach, W.T., Sung, K., Truong, S., Purves, D.: An empirical explanation of the flash-lag effect. Proceedings of the National Academy of Sciences **105**(42), 16338–16343 (2008)
43. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
44. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: International Conference on Computer Vision. pp. 479–486 (2011)