Convolutional Neural Networks Can Be Deceived by Visual Illusions

A. Gomez-Villa*, A. Martín*, J. Vazquez-Corral†, M. Bertalmío*
{alexander.gomez, adrian.martin, marcelo.bertalmio}@upf.edu, j.vazquez@uea.ac.uk

Abstract

Visual illusions teach us that what we see is not always what is represented in the physical world. Their special nature make them a fascinating tool to test and validate any new vision model proposed. In general, current vision models are based on the concatenation of linear and non-linear operations. The similarity of this structure with the operations present in Convolutional Neural Networks (CNNs) has motivated us to study if CNNs trained for low-level visual tasks are deceived by visual illusions. In particular, we show that CNNs trained for image denoising, image deblurring, and computational color constancy are able to replicate the human response to visual illusions, and that the extent of this replication varies with respect to variation in architecture and spatial pattern size. These results suggest that in order to obtain CNNs that better replicate human behaviour, we may need to start aiming for them to better replicate visual illusions.

1. Introduction

Visual illusions are fascinating examples of the complexity of the human visual system, and of the intrinsic difference between perception and reality: while we constantly assume that what we see is a faithful representation of the world around us, visual illusions make clear that what we see is just an internal construct of eyes and brain, because our internal representation and the world itself often do not match.

For instance, Fig. 1 shows a simple color illusion, where three identical cats are seen as having quite different colors depending on their surround. Visual illusions are so striking because, even after we go and check that the three cats have indeed the same triplet RGB value and therefore send the same light to us, we still see them as having different colors.

There are many types of illusions apart from color-based, involving other percepts such as brightness, motion, geometry or grouping, to name a few [1]. For the visual science community the study of visual illusions is key [2, 3], as the mismatches between reality and perception provide insights that can be very useful to develop new vision models of perception or of neural activity [4], and also to validate existing ones. This remains a very challenging open problem, as attested by the variety of vision science models (e.g. perceptual models based on edge-integration, Gestalt-anchoring, spatial-filtering, intrinsic images or purely empirical ones) and the fact that none of them can replicate a wide range of visual illusions; even models that can successfully predict an illusion may fail when a slight modification (like adding noise) is introduced [5].

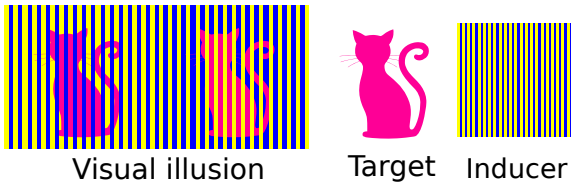

Figure 1. Anatomy of a simple color visual illusion. While the target (cat) is always the same, with the same RGB triplet in the three cases, we perceive it as "pink" when it is isolated, but magenta with one inducing surround and orange with the other.

A very popular approach in vision science is to model neural activity and also perception as a cascade of modules, each consisting of a linear convolution operation followed by a nonlinearity, see [6] and references therein[1]. These are of course the building blocks of convolutional neural networks (CNNs), but while the filters in visual models are designed so that the model best replicates neural or perceptual data, filters in CNNs are learned in a supervised manner in order to perform a specific imaging task, such as classification, recognition or denoising, to name just a few.

The authors find rather stunning that, given the importance of visual illusions for the vision science community, the neural inspiration of CNNs, and that so often the aim of CNNs is replicating human behaviour, there is virtually no work done on linking visual illusions and CNNs. To the

*Department of Information and Communication Technologies (DTIC), Universitat Pompeu Fabra. C. Roc Boronat 138, 08018, Barcelona, Spain
†CMP, University of East Anglia,NR4 7TJ, Norwich, UK

[1]We want to stress that linear+nonlinear cascades are very common but definitely not the only approach to modeling in vision science, given their well-known limitations [7].

best of our knowledge there are only two, very recent, publications in this regard. The first one comes from the vision science field [8], where a CNN trained to predict videos was able to reproduce motion illusions. In the second one, from the perspective of computer vision [9], the authors classify and attempt to generate new visual illusions using generative adversarial networks.

In this paper we report what we consider to be a quite remarkable and surprising finding, namely that CNNs trained on natural image databases for basic low-level vision tasks reproduce the human response to some visual illusion images, i.e. the CNNs are deceived by the visual illusions in the same way that we are deceived by them. Our other main contribution is to study how the ability of these CNNs to replicate visual illusions is affected by common architecture variations and spatial pattern size.

These results have, we believe, important consequences both for visual science and computer vision. For the visual science community, they support the idea that in order to perform low level vision tasks, the human visual system performs operations that as by-product create visual illusions. Moreover, these findings could help vision science in developing a taxonomy of which visual illusions are associated to which visual tasks. For the computer science community, the results build a new bridge between CNNs and the visual system. However, as it is shown on our experiments, this relationship and its possible consequences are constrained by the fact that not all optical illusions are replicated by the CNNs here studied. This can shed light on the limitations of CNNs for mimicking the visual system, and therefore offers an opportunity for the design of new architectures that, by better replicating visual illusions, could behave more like humans do.

2. Methods

2.1. Replication of human visual perception

Let us start by explaining how vision science measures the capacity of a model for replicating visual perception in a particular scenario. Observers first assess their perception of some aspect of the stimulus (e.g. the brightness) in a manner that is quantifiable (e.g. by ranking it on a scale from 0 to 5). Then, subject responses are averaged, and these averages are compared with the output values produced by the model. This last comparison can be either performed qualitatively or quantitatively.

In this paper, we will follow the qualitative paradigm. We say that a CNN replicates a VI if the difference between the values of the input image and the values of the output produced by the CNN agrees qualitatively with human perception. This is, if for instance we perceive a mid-gray level to be a darker gray and the CNN turns a mid-gray input value into an output value that is closer to black.

2.2. Selected visual illusions

There are two antagonistic basic effects in brightness and color visual illusions known as assimilation and contrast. In the case of assimilation effects, the image values change towards those of the neighboring region. Conversely, in the case of contrast effects, the image values move away from those of the neighboring region. We choose two assimilation VIs (Fig. 2a 2b), one contrast VI (Fig. 2d), and two non-defined (not assimilation, neither contrast) VIs (Fig. 2c and 2e). The first row in Figure 3 shows the color version of the same VIs.

The illusions 2a-d present targets that have identical values but that are seen differently depending on their surrounds. The targets are, in the Dungeon illusion ([10], Fig. 2a) the large central squares, in Hong-Shevell ([11], Fig. 2b) the middle rings, in the White illusion ([12], Fig. 2c) the small grey bars, and in the Luminance gradient (Lum.) illusion (combination of [13, 14], Fig. 2d) the circles. The fact that the targets have indeed the same values (0.5 in all cases) can be seen in the second row of Fig. 2, that plots the image values along the segments shown in color over the visual illusions in the top row. The Chevreul illusion [15] presents homogeneous bands of increasing intensity, from left to right, but these bands are perceived to be inhomogeneous, with darker and brighter lines at the borders between adjacent bands. In the color version (first row in Fig. 3), the phenomena are similar: for the Dungeon and the Hong-Shevell cases, the right target must go towards green and the left target towards red; for the White illusion, the left target must go towards yellow and the right target towards red; in the Luminance gradient illusion, the left targets should move towards red and the right targets towards green; finally, the Chevreul illusion should be perceived in the red channel analogously to the grayscale case.

2.3. Low-level vision tasks studied

In this work we consider three key image processing problems that have close correlates in human perception: denoising that relates to our ability to discount noise in images [16], deblurring that relates to our capabilities of avoid perceiving the blur provoked by moving objects [17], and color constancy that relates to the way our perception of colors matches quite well the reflectance properties of objects rather independently of the color of the illuminant.

2.4. CNN architectures

For our core experiment, we chose a very simple architecture for the CNNs: input and output layers of size $128 \times 128 \times 3$ pixels, one hidden layer with eight features maps with a receptive field (kernel size) of five and no stride (stride 1), and sigmoid activation functions. At the end there is a convolutional layer that works as output layer (hence it has three layers for the red, green, and blue channels).

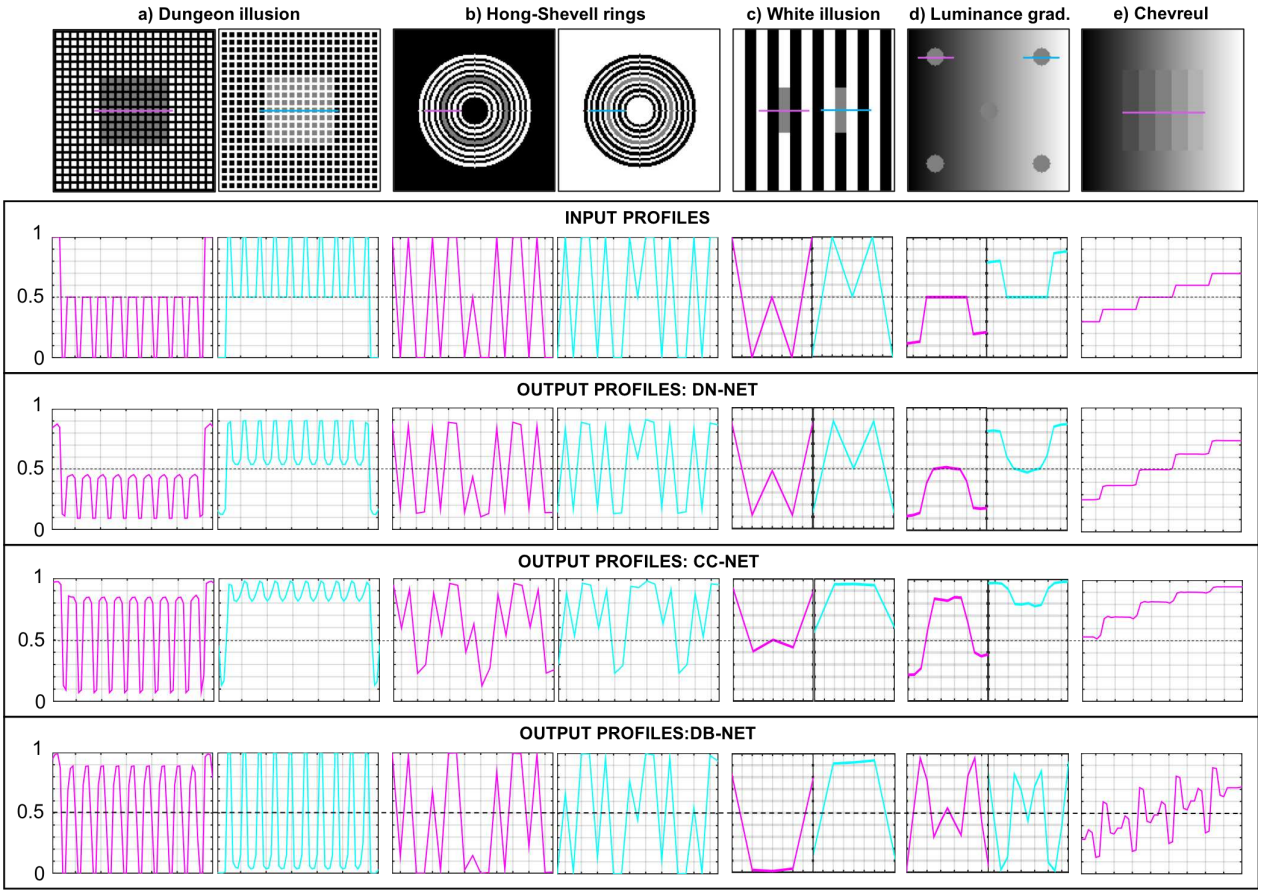

Figure 2. The first row displays the selected grayscale visual illusions as explained in Section 2. The scale of the illusions in the Figure is different from the scale used in the experimenta for displaying purposes. The magenta and cyan lines represent the location in the images of the profiles plotted in the rows 2-4.

Note that no pooling, residual connections or other modifications were added to this architecture. Mean squared error was used as loss function in all the tasks and all the models were implemented[2] using Keras [18]. We name the CNNs based on the task they were trained for. Hence, DN-NET, CC-NET, and DB-NET correspond to denoising, color constancy, and deblurring, respectively

Then, we move to a similar CNN presented by Jain et al. [19], one of the first successful CNNs designed for image denoising. Our implementation of this CNN, that we denote as Jain2009 from now on, has an input/ouput size of 128×128 and is composed of four hidden layers with a kernel size of five and a sigmoid as activation function. This CNN can be considered as a deeper version of the CNNs of the first experiment. We used this CNN to study common variations in the architecture of CNNs and how they affect to the replication of VIs.

Finally, we use a recent CNN for denoising (Zhang2017) [20] to test the response for more complex architectures.

[2]The source code is publicly available at:
https://github.com/alviur/convnets_vs_vi

2.5. Datasets

For denoising we consider the Large Scale Visual Recognition Challenge 2014 CLS-LOC validation dataset [21] (which contains 50k images), and corrupts images with additive Gaussian noise of $\sigma = 25$ after resize them to 128x128. For deblurring we consider the same dataset as before, and blur the images with a Gaussian kernel of $\sigma = 2$. For color constancy we consider the dataset of Cheng *et al.* [22] that provides the color of the illumination for each image. We computed the ground-truth image by applying the inverse of the illuminant color to the original image, and then we performed an end-to-end training between the original image and the ground-truth one. For this problem, we divide each original image into four sub-images in order to increase the pool of available images for the training of the net. By doing this, we end up with a total of 6944 images. In all three cases, the dataset was split in 70% for training, 20% for validation, and 10% for test.

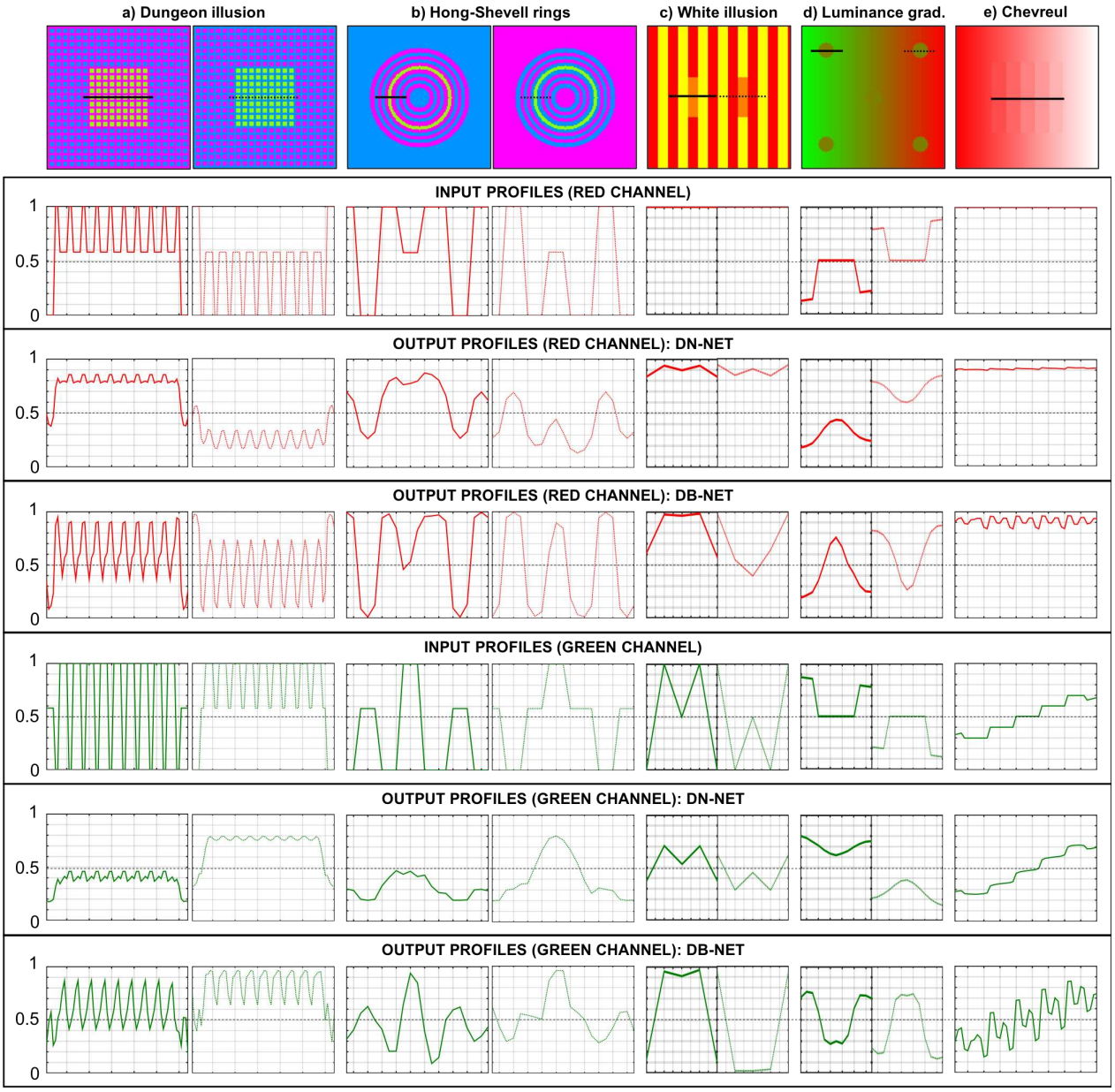

Figure 3. The first row displays the selected color visual illusions as explained in Section 2. The scale of the illusions in the Figure is different from the scale used in the experimenta for displaying purposes. The black continuous and dashed lines represent the location in the images of the profiles plotted in the rows 2-4. Only the profiles from the Red and the Green Channels are displayed.

2.6. Experiments

A base spatial scale for each of the five VIs was fixed in order to evaluate replication. This base scale was 4x4 pixels (px.) target squares for Dungeon (Dun), 1 px. ring width for Hong-Shevell, 4x4 px. target for White, 5 px. diameter target for Luminance gradient (Lum.) and 10 px. step width in Chevreul. A corresponding base size of the receptive field (also referred later as kernel size) was chosen to be 5x5.

In humans there is an observed relationship (see e.g. [1]) between spatial frequency and visual effect. In most cases this relationship states that higher frequencies imply a larger difference between the observed targets. This relationship lead us to study the effect of the spatial scale in the replications of the VIs. For this end, we reasonable assume a relation between the receptive field and the spatial frequency of the patterns. Following this assumption, we increase the illusion's scale by 3, 4 and 6 times the base scale. In the case of the kernel size, we also test receptive field sizes of

Table 1. Summary of experiments performed.

Label	CNN	Visual Illusions	Illusions' scale	Kernel size	Figure
E1	DN-NET, CC-NET, DB-NET	Full set grayscale	Base	5	2
E2	DN-NET, CC-NET, DB-NET	Full set color	Base	5	3
E3a	DN-NET	G/RGB Dun. & Lum.	$\{1, 1.1, 1.2, 1.3\}\times$Base	5	4 a,b
E3b	DN-NET	G/RGB Dun. & Lum.	$\{1, 1.1, 1.2, 1.3\}\times$Base	3, 5, 7, 11, 15	4 c
E4a	Jain2009	G/RGB Dun. & Lum.	Base,	5	5
E4b	Jain2009+Pooling	G/RGB Dun. & Lum.	$\{1, 1.1, 1.2, 1.3\}\times$Base	3, 5, 7, 11, 15	6a
E4c	Jain2009+Dilated Convolutions	G/RGB Dun. & Lum.	$\{1, 1.1, 1.2, 1.3\}\times$Base	3, 5, 7, 11, 15	6b
E4d	Jain2009+Residual Connections	G/RGB Dun. & Lum.	$\{1, 1.1, 1.2, 1.3\}\times$Base	3, 5, 7, 11, 15	6c
E5	Zhang2017	G/RGB Dun. & Lum.	Base	5	7

3x3, 5x5, 7x7, 11x11 and 15x15.

The first two experiments consisted in evaluating the replication of DN-NET, CC-NET and DB-NET when presented with all the selected grayscale (**E1**) and color (**E2**) visual illusions for the base scales and kernel size.

Further experiments in the paper are restricted to the color and grayscale versions of the Dungeon and the Lum. illusions, which are representative of the opposite effects of assimilation and contrast, respectively.

In the third experiment, we studied the replication for DN-NET when the spatial scale and the kernel size are different from the base case. First, with fixed kernel size, we vary the illusion's scale (**E3a**). Secondly, we modify both, the scale of the illusions and the size of the receptive field (**E3b**).

We later moved to a deeper CNN (Jain2009) and we tested its replication for the base scales and kernel size (**E4a**). Then, in this same experiment, we tested all the scales and all the sizes for the illusions and the receptive fields, respectively.

Next, we studied three different variations of the Jain2009 CNN. First, two pooling layers were added to this architecture after the first and second convolutional layers respectively. In order to recover the original scale of the input, after each of the last two hidden convolutional layers an upsampling layer was added. Pooling sizes of 2, 4, and 8 were tested (**E4b**). The second variation was to replace the standard convolutional layers of Jain2009 with convolutional layers with a dilation rate of 2, 4, and 8 (**E4c**). The last test consisted on trying several configurations of Jain2009 with residual connections (**E4d**).

The last experiment studied the replication of a state-of-the-art CNN for denoising, Zhang2017. It was tested for the base scale of the illusions (**E5**).

A summary of all experiments above presented can be found in Table 1.

Finally, we addressed the question if simple image processing algorithms can also explain the VI phenomena. In particular, we study if a classical contrast enhancement (CLAHE [23]) and a classical denoising (Total Variation besed denoising [24]) can reproduce both assimilation and contrast.

3. Results

Due to the outstanding amount of experiments and the difficulty to show its results, in this paper we will focus on some selected cases. Nevertheless, our selected cases represent the main effects and trends found in our proposed experiments. We invite the reader to look at our supplementary material for the full set of experiments.

We present our results as profiles of the output of the CNN. A profile is a 1-dimensional plot of the pixel values of a row from the output image. For each VI, the plotted row is indicated with a color segment over the VI. For instance, in Fig. 2 the input profiles (second row) are the profile of the marked region (as magenta or cyan)in the VI (first row). This qualitative representation of results is common in vision science [2].

The output profiles showed in the figures corresponding to grayscale illusions are the grayscale values obtained using the formula $0.2989\mathrm{R} + 0.5870\mathrm{G} + 0.1140\mathrm{B}$, with R, G and B being the corresponding values in the red, green and blue channels.

In all the results (output profiles) of this paper VI inputs were not contaminated with noise or blurred before feeding them to the CNNs. Additional experiments with contaminated inputs (not shown due to space limitation) resulted in similar replication effects.

3.1. Replication of grayscale VIs (E1)

Fig. 2 shows the results of the experiment E1. We can see that DN-NET is capable of replicating illusions from (a) to (d) (see the row *Output profiles: DN-NET* in Fig. 2). While Dungeon (a) and H.S. (b) are very well replicated, in White (c) and Lum. (d) the effect is less marked. CC-NET replicates illusions from (b) to (d) (see *Output profiles: CC-NET*

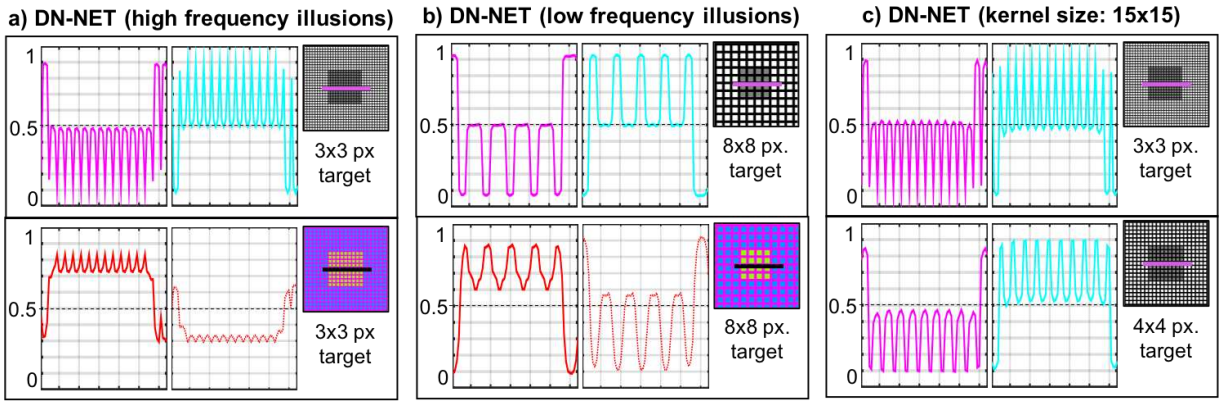

Figure 4. Assimilation results in DN-NET for low and high frequency grayscale and color visual illusions.

in Fig. 2) but produces the opposite effect to that of human perception in (a). Finally, DB-NET replicates illusions from (b) to (e), but presents the same opposite effect as CC-NET in (a). Nevertheless, DB-NET is the only one able to replicate the effect for the Chevreul illusion in grayscale (e).

3.2. Replication of color VIs (E2)

In Fig. 3 we show the results of the experiment E2. DN-NET replicates illusions (a),(b),(c), and (e). For Dungeon (a) and H.S. (b) the right target increases its green value (w.r.t. the input) while the left target increases its value in the red channel. For White (c), the left target gets closer to a yellow color by increasing its green channel value. In the case of Chevreul (e), there is a slight replication in the red channel. Finally, in the case of Lum. (d), DN-NET fails to reproduce the VI.

DB-NET replicates all illusions except for H.S. (b). For illusions Dungeon (a) and White (c) the effect is the same as that observed for DN-NET. For Chevreul (e), the effect is replicated both in the red and the green channels. Finally in Lum. (d), there is a clear increase in the red and the green channels for the left and the right targets respectively, together with a corresponding decrease of the same channels in the opposite target.

3.3. Influence of scale of the VIs (E3a)

DN-NET reduces the replication error when the size of the pattern is increased (increasing the size of the pattern is equivalent to reducing the spatial frequency), therefore emulating the behaviour observed in human perception [1]). However, the reduction of the effect is dependent on the receptive field size and on whether the illusion is in grayscale or color.

The replication effect observed for DN-NET in the Dungeon illusion in grayscale is completely lost when moving to sizes equal or larger than 8 pixels (see the middle column of the first row in Fig. 4). However, the same VI in color still replicates the effect for that size specially in the red channel (middle column of the second row in Fig. 4). The same evolution but in a smaller degree is also observed in the case of Lum.

Furthermore, increasing the spatial frequency leads to an attenuated replication, contrary to the effect produced in human perception. Figure 4a in its first row shows how the assimilation effect in Dungeon almost disappears in grayscale. That is also the case for the contrast effect in Lum. In the case of color, the assimilation effect is still clearly present (Fig. 4a) but not the contrast effect of Lum.

3.4. Effect of receptive field size (E3b)

For the Lum. VI, the use of larger receptive fields lead to an increase of the replication effect. However, for the Dungeon effect when using the largest receptive field size (15×15), moving from a target size of 4 to 3 pixels changes the assimilation into a contrast effect (see Fig. 4c). For the color VIs there were no significant qualitative changes for either illusion.

Although there is a relation between the receptive field and the spatial frequency of the patterns. The nature of this relation is not directly understood from the current experiments. In most of the combinations of pattern's frequency and size of receptive field tested the qualitative results do not change.

3.5. Jain2009, a deeper architecture (E4a)

Figure 5 shows the results of Jain2009. We find replication (although reduced) of both effects in grayscale. Despite being four times deeper than DN-NET, Jain2009 shows qualitatively similar results to the original DN-NET. For the case of color VIs, there is still a replication of the assimilation effect in Dungeon but not of the contrast effect in Lum. (as was the case for DN-NET, see Fig. 3).

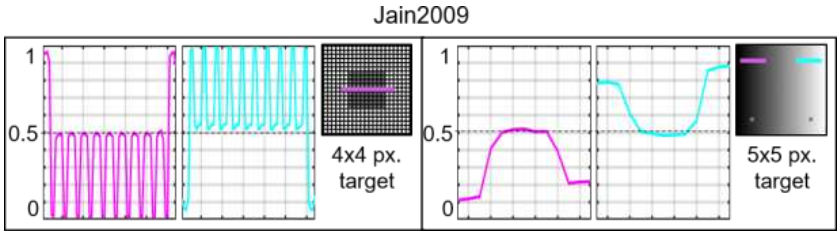

Figure 5. Replication results for Dungeon and Lum. for Jain2009, the architecture based in [19].

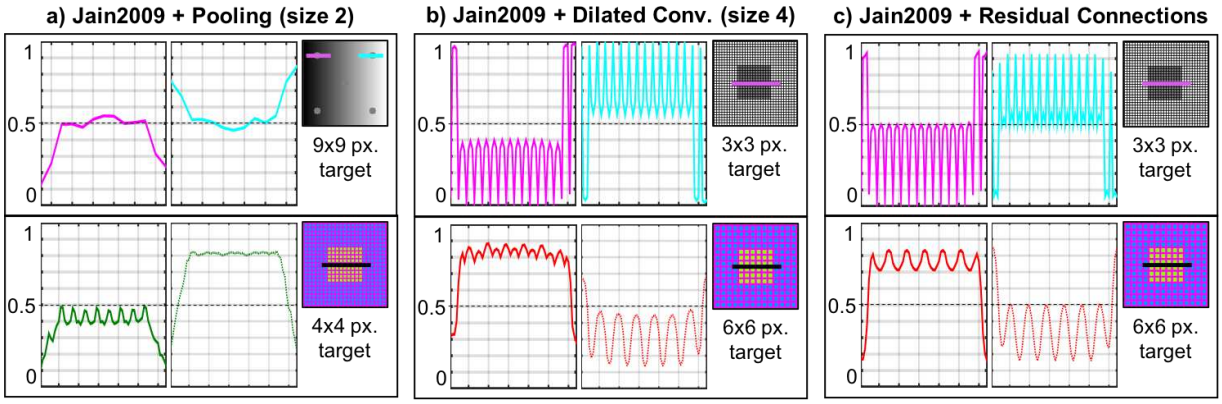

Figure 6. Selected results from Jain2009 when adding Pooling of size two. Assimilation effect is only replicated in color while the contrast effect is replicated in grayscale.

3.6. Adding pooling layers to Jain2009 (E4b)

When pooling layers were used (in this case of size two) the side effect in grayscale images is that higher frequency VIs are destroyed. Also, in the case of Dungeon the replication is lost, in fact, the opposite effect is observed. However, there is still a replication effect in Lum. for bigger target sizes (see Fig. 6a). In the case of color, the same effect of spatial pattern destruction occurs, but the replication effect still remains for the Dungeon VI in the red and green channels. Larger pooling sizes lead to a total spatial destruction of the patterns in the VIs such that further analysis is prevented.

3.7. Adding dilated convolutions to Jain2009 (E4c)

Two main effects are observed when dilated convolutions were added. First, the contrast effect of Lum. is not replicated in grayscale or color for any of the dilation rates. Second, in all the cases the effect in grayscale for Dungeon, when considering targets equal or larger than 4 pixels, is no longer replicated. In fact, it shows a contrast effect instead. However, in the case of color there is still replication for the Dungeon VI even when larger targets are considered (see Fig. 6b second row).

A special case is observed only when using a dilation of size four: Replication does appear in Dungeon in grayscale for the smallest pattern size (shown in the left column of Fig. 6b first row). This is not the case for any other size of the dilation.

3.8. Jain2009 with residual connections (E4d)

Several configurations of Jain2009 with residual connections were tested. They shared the effect of annulling the replication of both Dungeon and Lum. in grayscale. However, an architecture with a single residual connection going from the output of the first convolutional layer to the input of the final output layer was still able to replicate the assimilation effect in the grayscale Dungeon VI for the highest frequency (see the left column in Fig. 6c). In the case of color, for all the different variations of Jain2009 with residual connections, there is a replication for Dungeon even if we increase the pattern size (second row in Fig. 6c) but not for Lum. in none of the cases.

3.9. Replication in recent denoising CNNs (E5)

Figure 7 shows how, to a small degree, Zhang2017 can replicate the effect in both Dungeon and Lum. VIs. This is also the case for the color VI, in where it behaves in the same way as DN-NET and Jain2009.

3.10. VIs and classical image processing methods

It is to be expected that a sharpening filter, or a classic contrast enhancement algorithm like CLAHE [23], may be

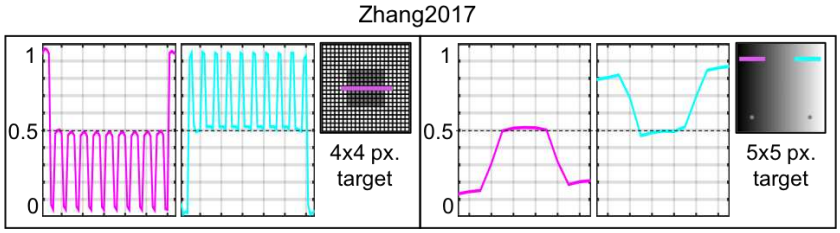

Figure 7. Replication results for Dungeon and Lum. for Zhang2017, the state-of-the-art CNN for image denoising presented in [20].

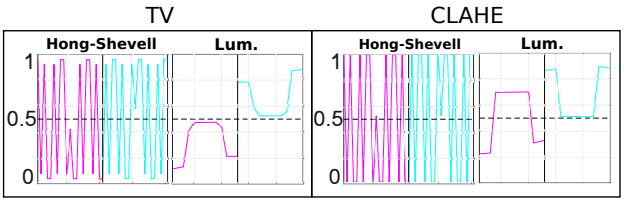

Figure 8. Replication results for classical image processing algorithms. Left: Total-Variation denoising. Right: CLAHE.

able to reproduce VIs that are of the contrast type, i.e. that make the image value of a region to change in the opposite direction to the values of the neighboring pixels. Analogously, a simple average filter, or a classic denoising algorithm like Total Variation based denoising [24], may be able to reproduce VIs that are of the assimilation type, i.e. that make the image value of a region to change towards the values of the neighbors of said region. Nevertheless, these simple filters or classic algorithms do not seem able to reproduce simultaneously both types of illusions, as figure 3.7 shows. TV denoising reproduces Hong-Shevell (assimilation) but not the Luminance illusion (contrast), while CLAHE does the opposite. In contrast, the DB-NET and DN-NET CNNs introduced in this paper were capable of reproducing these two illusions at the same time.

4. Conclusions

In this work we showed that CNNs trained on natural image databases for basic low-level vision tasks reproduce the human response to some visual illusion images, i.e. the CNNs are deceived by the visual illusions in the same way that we are deceived by them. Versions of a single hidden layer CNN trained for denoising, color constancy, and deblurring were tested to replicate five common visual illusions. Deeper architectures and their common modifications (such as pooling layers, dilated convolutions, and residual connections) were explored too in order to evaluate their effect in the replication of visual illusions. It was found that even the simplest single hidden layer with 8 feature maps is already capable of replicating the human response to several grayscale and color illusions. Moreover, changes in the input image or CNN architecture lead to a change in the illusions that the network is able to reproduce.

We argue that the CNNs in this paper reproduce visual illusions as a by-product of the low level vision tasks of denoising, color constancy or deblurring. Albeit clearly different, the biological correlates of all of these tasks aim to improve the efficiency of the representation and the visual processing, so this supports the argument that visual illusions are the price we have to pay in order to optimally use the limited resources of our visual system.

The illusions that the CNNs are able to replicate depend on the task each CNN is solving. It would be interesting, from a vision science perspective, to use this insight to try to associate specific illusions (or families of illusions) with visual processing tasks. Another interesting finding was that CNNs trained with color images can replicate visual illusions in grayscale too: this could maybe give some cues towards answering the question of where precisely in the visual system is the brightness percept derived from color signals, which is still an open one.

Finally, and from a computer vision perspective, if we want CNNs that better replicate human behaviour, we should maybe start aiming for them to better replicate visual illusions. We are currently working along these lines, developing a CNN architecture with the goal of reproducing as many visual illusions as possible, with validations from psychophysical data.

As future work we want to evaluate if CNNs that replicate visual illusions are more resistant to adversarial attacks that do not fool humans. And to generate new visual illusions using for instance generative adversarial networks.

Acknowledgements

This work has received funding from the EU Horizon 2020 programme under grant agreement 761544 (project HDR4EU) and under grant agreement 780470 (project SAUCE) and by the Spanish government and FEDER Fund, grant ref. TIN2015-71537-P (MINECO/FEDER,UE). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] A. G. Shapiro and D. Todorovic, *The Oxford compendium of visual illusions*. Oxford University Press, 2016.

[2] F. A. Kingdom, "Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy," *Vision research*, vol. 51, no. 7, pp. 652–673, 2011.

[3] J. Yeonan-Kim and M. Bertalmío, "Analysis of retinal and cortical components of retinex algorithms," *Journal of Electronic Imaging*, vol. 26, no. 3, p. 031208, 2017.

[4] D. H. Hubel, *Eye, brain, and vision.* Scientific American Library/Scientific American Books, 1995.

[5] T. Betz, R. Shapley, F. A. Wichmann, and M. Maertens, "Noise masking of white's illusion exposes the weakness of current spatial filtering models of lightness perception," *Journal of vision*, vol. 15, no. 14, pp. 1–1, 2015.

[6] M. Martinez-Garcia, P. Cyriac, T. Batard, M. Bertalmo, and J. Malo, "Derivatives and inverse of cascaded linear+nonlinear neural models," *PLOS ONE*, vol. 13, pp. 1–49, 10 2018.

[7] N. V. Graham, "Beyond multiple pattern analyzers modeled as linear filters (as classical v1 simple cells): Useful additions of the last 25 years," *Vision research*, vol. 51, no. 13, pp. 1397–1430, 2011.

[8] E. Watanabe, A. Kitaoka, K. Sakamoto, M. Yasugi, and K. Tanaka, "Illusory motion reproduced by deep neural networks trained for prediction," *Frontiers in psychology*, vol. 9, p. 345, 2018.

[9] R. M. Williams and R. V. Yampolskiy, "Optical illusions images dataset," *arXiv preprint arXiv:1810.00415*, 2018.

[10] P. Bressan, "Explaining lightness illusions," *Perception*, vol. 30, no. 9, pp. 1031–1046, 2001.

[11] S. W. Hong and S. K. Shevell, "Brightness contrast and assimilation from patterned inducing backgrounds," *Vision Research*, vol. 44, no. 1, pp. 35–43, 2004.

[12] M. White, "A new effect of pattern on perceived lightness," *Perception*, vol. 8, no. 4, pp. 413–416, 1979.

[13] E. Bruke, "uber erganzungs und contrasfarben," *Wiener Sitzungsber*, vol. 51, 1865.

[14] E. H. Adelson, "Lightness perception and lightness illusions," *New Cogn. Neurosci*, vol. 339, 2000.

[15] F. Ratliff, *Mach bands: quantitative studies on neural networks*. Holden-Day, San Francisco London Amsterdam, 1965.

[16] W. McIlhagga, "Denoising and contrast constancy," *Vision Research*, vol. 44, no. 23, pp. 2659 – 2666, 2004.

[17] M. Morgan and S. Benton, "Motion-deblurring in human vision," *Nature*, vol. 340, pp. 385–386, 1989.

[18] F. Chollet *et al.*, "Keras." https://keras.io, 2015.

[19] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Advances in Neural Information Processing Systems*, pp. 769–776, 2009.

[20] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[21] Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[22] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *J. Opt. Soc. Am. A*, vol. 31, pp. 1049–1058, May 2014.

[23] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994.

[24] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.