
A Turing Test for Artificial Nets devoted to model Human Vision

Jorge Vila-Tomás

Image Processing Lab

Universitat de València

Paterna 46980, València, Spain

jorge.vila-tomas@uv.es

Pablo Hernandez-Cámara

Image Processing Lab

Universitat de València

pablo.hernandez-camara@uv.es

Qiang Li

Image Processing Lab

Universitat de València

TReNDS

Georgia State, Georgia Tech, and Emory

qiang.li@uv.es

Valero Laparra

Image Processing Lab

Universitat de València

valero.laparra@uv.es

Jesús Malo*

Image Processing Lab

Universitat de València

jesus.malo@uv.es

Abstract

In this work² we argue that, despite recent claims about successful modeling of the visual brain using deep nets, the problem is far from being solved, particularly for low-level vision. Open issues include *where should we read from in ANNs to check behavior?*, *what should be the read-out?*, *this ad-hoc read-out is considered part of the brain model or not?*, in order to understand vision-ANNs, *should we use artificial psychophysics or artificial physiology?*, anyhow, *artificial tests should literally match the experiments done with humans?*. These questions suggest a clear need of biologically sensible tests for deep models of the visual brain, and more generally, to understand ANNs devoted to generic vision tasks.

Following our use of low-level facts from *Vision Science* in image processing, we present a low-level dataset compiling the basic spatio-temporal and chromatic facts that describe the adaptive information bottleneck of the retina-V1 pathway, and are not currently available in popular databases such as BrainScore. We propose its use for model evaluation.

As illustration of the proposed methods we check the behavior of three recent models with similar deep architecture: (1) A parametric model tuned via the psychophysical method of Maximum Differentiation [Malo & Simoncelli SPIE 15, Martinez et al. PLOS 18, Martinez et al. Front. Neurosci. 19], (2) A non-parametric model (the *PerceptNet*) tuned to maximize the correlation with humans on subjective image distortions [Hepburn et al. IEEE ICIP 20], and (3) A model with the same encoder as the *PerceptNet*, but tuned for image segmentation [Hernandez-Camara et al. Patt.Recogn.Lett. 23, Hernandez-Camara et al. Neurocomp. 25]. Results on 10 compelling psycho/physio visual facts show that the first (parametric) model is the one with closer behavior to humans in terms of the nonlinear behavior when facing complex spatio-chromatic patterns.

*Corresponding author. Web Site: <https://isp.uv.es/excathedra.html>

²Concept and results first presented at the *AI Evaluation Workshop* at the University of Bristol, June 2022.

1 Introduction

1.1 Prologue

This work reproduces our *talk* (otherwise unpublished in print) at the AI Evaluation Workshop in june 2022 at the AI Dept. of the University of Bristol organized by Prof. Raul Santos of the Eng. Maths Dept. of UoB. That *talk*, that proposed an original methodology (with experimental results) to evaluate deep nets devoted to vision tasks, was the seed of our current work with Prof. Jeff Bowers of the Psychol. Dept. of UoB in the context of the Benjamin Meaker Distinguished Professorship granted to Prof. Jesús Malo in 2024, as a low-level complement to the (high-level) Bowers' proposals in [1, 2]. Journal publication of this 2022 *talk*, is pertinent for a wider audience because this approach based in low-level visual psychophysics is still unusual in the AI and machine learning communities, despite some researchers are independently proposing very similar evaluations quite recently [3, 4]. As shown below, our proposed evaluation program includes facts that go beyond the luminance, color, and contrast masking facts considered in [3, 4]. The work of Rafal Mantiuk's lab shares the same spirit and focus on low-level psychophysics, but his *more quantitative comparison* is in contrast with our proposal, which stress the *qualitative understanding* of the human response curves so that the AI researchers can spot major conceptual errors in deep models in an easy way. Moreover, as explained below, the selected visual stimuli³ (and associated psychophysical facts⁴), allow to intuitively infer modifications in the architectures in order to correct the detected errors.

1.2 Motivation: is that model really human-like?

The motivation for our proposal starts by reviewing the claims about how deep learning models are the ultimate tool to model the visual brain, as recalled in [1]. Claims cited by Bowers et al. include [5, 6, 7, 8, 9, 10]. Other examples in the same vein not cited by Bowers et al. include [11, 12]. Nevertheless, following the skeptical tone of Bowers et al. [1], the crucial comment made by some scientists (for example at the Center of Neural Science of NYU after they carefully listen to the details of your model) is *yes, yes, that is nice, but the brain doesn't work like that, does it?* [13].

Two additional examples of the skepticism in that critical question include *Tomaso Poggio* and *Horace Barlow*. In the 70's David Marr and Tomaso Poggio proposed an interesting taxonomy of the approaches to the vision problem: their famous *separate abstraction levels*, namely, computational, algorithmic and implementation [14, 15]. However, 42 years later, in view of the current tools to optimize models, Poggio himself questioned the separability of these levels [16]. This taxonomy has been inspiring for decades, but now it is under debate [17, 18, 19]. For example, work on color illusions [20] and CSFs in autoencoders [21], and work on subjective distances between images in ANNs [22, 23] stress the relation between the computational and the algorithmic levels, thus questioning previous (purely computational) explanations that disregard architecture [24, 25, 26]. In a similar vein, Horace Barlow, 50 years after his inspiring *Efficient Coding Hypothesis* [27, 28], questioned purely infomax approaches [29], and, for a similar reason, he questioned our preliminary work on the use of Principal Curves to explain color and texture nonlinearities of human vision based on the data, back in 2004 [30]: *that is interesting, but the visual brain may not work like that* [31].

That skepticism (based on the range of empirical behavior explained and the assumptions made to make these explanations) is the core of the spirit in [1], and also the motivation of this work. Therefore, the key ideas of this work are basically two:

- The use of AI techniques (e.g. deep learning) to understand the visual brain may not be as easy as people thought back in 2022, and even now. More explanatory tests are required.
- Our specific proposal here is a low-level Turing test based on 10-points low-level physiological and psychophysical facts (our *Decalogue*) to check if certain artificial model behaves as the (low-level) human visual brain.

1.3 Structure of the paper

Section 2 states that the question *are the models sensible from the point of view of low-level physiology and psychophysics?* remains open from the perspective of modeling and evaluation. In Section 3,

³All made online available here since 2022: <http://isp.uv.es/docs/TuringTestVision.zip>

⁴Original 2022 slides (UoB AI Evaluation Workshop): http://isp.uv.es/docs/talk_AI_Bristol_Malo_et_al_2022.pdf

we propose our contribution: an easy-to-use test (consisting on online available visual stimuli) and associated responses illustrated here for evaluation of deep learning vision models. These stimuli visually illustrate low-level phenomena described by classical *Vision Science*. In Section 4, we illustrate the proposed method through the original evaluation of three recent models: (1) a classically formulated, not end-to-end optimized, model with functional form derived from classical vision science literature, where the specific values of its parameters have been psychophysically measured [32, 33, 34, 35]. (2) A network with a bio-inspired architecture but with free parameters end-to-end optimized to reproduce subjective image quality, the *PerceptNet* [36]. It resembles AlexNet and VGG, but it was specifically designed to accommodate the known aspects of the retina-cortex visual pathway using a constrained version of divisive normalization [37]. And, (3) a model with the same style encoder as the *PerceptNet*, but augmented with a decoder, and both (encoder and decoder) are trained for image segmentation [38, 39], which is also a biologically plausible task. Section 5 discusses what can be generally obtained from the proposed test methods (which might be considered qualitative). Note that even in the engineering case where one does not necessarily need the networks to resemble human vision, one would always want them to have good adaptation properties to achieve good generalization, and potential failures in this regard become clearly evident through the proposed tests. Finally, Section 6 concludes the paper.

2 Open issues in modeling vision

As pointed out in [40] the basic question, as in human vision, is how to deal with deep models which are hardly explainable black-boxes once trained.

2.1 Uncertain computational goal

First, the more general open issue is the discussion on the *computational goal* that eventually explains the organization and behavior of visual systems. Consider architectures/tasks such as the ones presented in Fig. 1. These tasks are related to low-, mid-, and high-level tasks arguably implemented by biological vision. In biology enhancement of the blurry and noisy signal in the retina has been proposed as an explanation of the the LGN, as pursuing this goal may reproduce some of its spatio-chromatic [41, 21] and purely chromatic [20] features. Another example is the compression possibly happening in part at the LGN bottleneck and at the feature selection after V1. Bandwidth limitation, dimensionality reduction and attention focus are sensible goals in this regard [42, 43, 44]. A number of compression algorithms (for images [45, 46, 47, 48, 49, 37] and video [50, 51, 52, 53]) have been based on human vision models. Segmentation is arguably another (mid-level) task that has to be done by biological vision, and biological nonlinearities have been shown to improve segmentation in images [38, 39] and video [52, 53]. Arguably, segmentation is implemented in the *where* channel

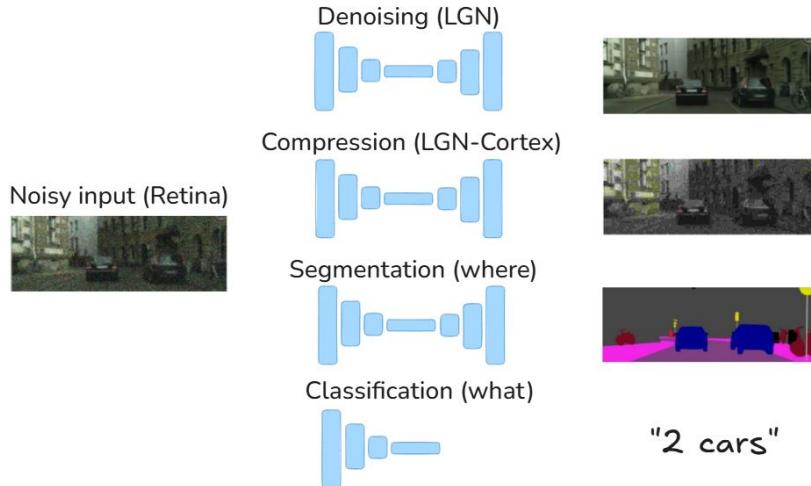


Figure 1: Image Denoising, Image Compression, Image Segmentation and Image Classification architectures with (eventually) biological correlates in the LGN, the V1 and beyond. However, it is not obvious how these tasks may be combined to explain biological vision.

from the lower-level primitives extracted in V1 [54, 55]. Higher-level tasks such as classification are supposed to happen in the *what* channel [56, 57]. And similarly, in standard models such as the one depicted in Fig. 1 biological nonlinearities have been shown to have a significant role in classification [58, 59].

2.2 Uncertain read-out mechanisms

As stated in the introduction, in the age of automatic differentiation where the classical Marr-Poggio levels are not that separated, the *computational goal* is not the only open issue. For instance, in order to check if a (mathematical) model is biologically sensible, where should we read the signals from? The read-out mechanism is also important. Note that the fact that certain layer has the necessary information in order to solve a task (read-out in *any complicated* way, e.g. a highly specialized dense network) is not enough to say that this layer represents the way the visual brain works: the necessary information is already present in the retina (if read in proper way) and, of course, the retina is not a good model for the rest of the visual brain. This problem is illustrated in Fig. 2.

In the case of doing *artificial physiology*, i.e. reading the signals from certain neurons or layer, or *artificial psychophysics*, i.e. trying to make decisions from the responses of the network, for instance to decide if certain stimulus is visible or not, one should propose a *read-out mechanism* to summarize the responses into a decision variable (see Fig. 3). The selection of the *read-out mechanism* is not trivial. In fact, the quality of the read-out information may strongly depend on the complexity of this (arbitrarily selected) mechanism. As a result, one may not be able to tell if the model itself is good, or the good behavior has to be attributed to a clever read-out which is not part of the model. Examples include the use of classifiers at certain location of the network to make a decision on visibility, as in [58, 60], or without classifiers relying on the model output[61]; or the (more classical) use of Euclidean distances between stimuli to tell if they are discriminable [62, 21, 22]. These (arbitrary) decisions definitely affect the characterization of the system, e.g. its frequency response [21, 60]. For example, linear or nonlinear classifiers effectively apply different (non-Euclidean) distance metrics [63] and, hence, they should lead to different decisions.

Another (more particular) discussion is the debate on the summation, which is classical in vision science [64]: for instance, which Minkowski exponent is more physiologically plausible?. Note that using different norms and summation schemes definitely lead to different results [65]. A final (also non obvious) way of assessing stimuli in the network is measuring differences in the statistical properties of the response [66, 67] or measuring information flow along the network [68, 69, 70, 71, 72]. These options require making non-trivial decisions such as which statistical descriptors make sense [73, 67], or how to set the level of noise in the network [68, 69, 70]. In this regard, models can be improved either by changing the architecture and the measures of information [71, 74], or by better estimations of the internal noise [75].

2.3 Uncertain experimental setting

And finally, the third open issue is the way of doing the evaluation: *the experiment implementation matters*. In particular, *should we use artificial physiology or artificial psychophysics?*. Current techniques by the machine learning community to visualize the behavior of the networks [76, 77] are based on classical single cell recordings, such as the very concept of *receptive field* [78, 79, 80], and the identification of sensitive neurons by looking to the stimulus that maximizes the neuron response, which is a common practice in visual neuroscience [81]. However, there are

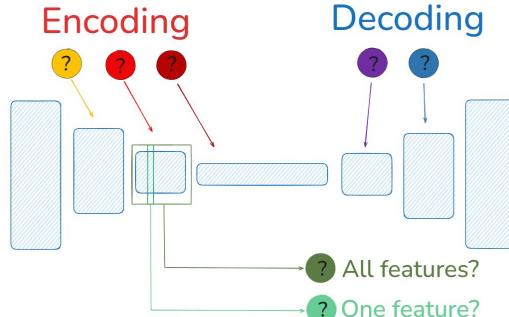


Figure 2: Given a deep model successfully trained for some visual task, the read-out location and read-out mechanism (or decoder) is important to assess its biological plausibility.

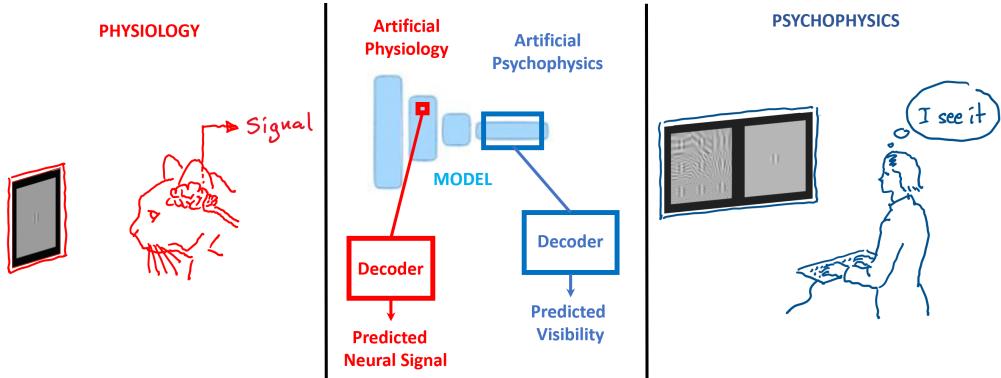


Figure 3: In artificial physiology (left) and in artificial psychophysics (right), the arbitrary decoder to read-out model activations is critical.

more sophisticated techniques such as *reverse correlation* which are used both in physiology [82] and in psychophysics [83], and these are not yet widely used in machine learning. Regarding the experimental setting, *should one go for a literal reproduction of the experiments with humans, or should one try an idealized version of the experiment?*. This open question can be illustrated by the example in Fig. 4 on the spectral sensitivity of a network.

This is a non-trivial question because, for instance, some techniques to assess visual illusions in a model involve the inversion of the inner representation [85, 20], which *does not happen in the human brain* while others, similarly to human psychophysics [86], are based on *matching* the response at the inner representation [87, 88]. As stated above, this has implications as deciding at which layer one should impose the matching (or where to read from).

2.4 Better evaluation techniques are needed

All these non-trivial decisions (despite they all belong to low-level characterizations of the visual system) clearly point out the need of better methodologies for model evaluation in order to assess how close different models may be to the visual brain.

In this context, our proposal here is simple: *just provide the code to generate a set of well selected stimuli that illustrates a number of classical low-level visual psychophysics facts and have them*

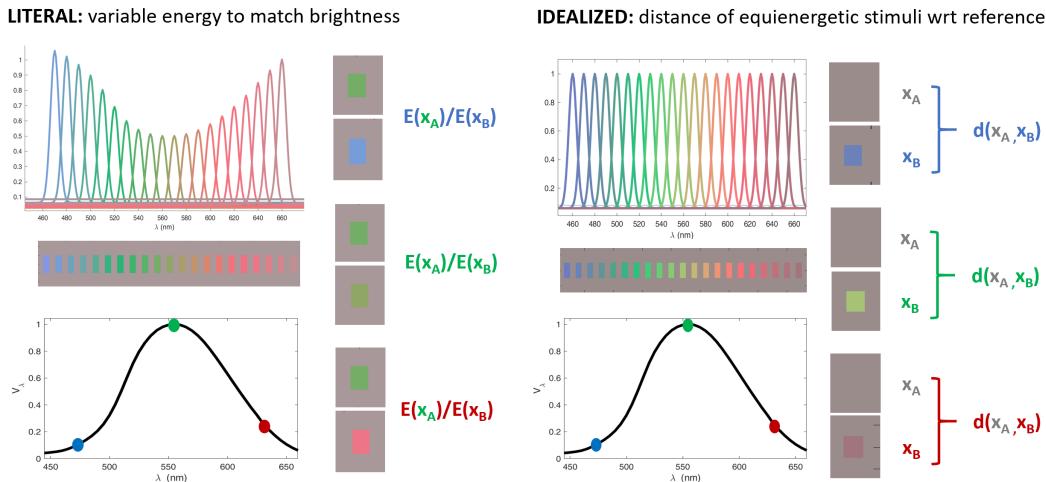


Figure 4: In measuring the spectral sensitivity of certain elements of a network one may try a *literal reproduction* of the human psychophysics (left) or an *idealized* experiment (right). The *literal reproduction* could be done through matching experiments [84]: finding the ratio of energies necessary to match the response to quasi-spectral stimuli of different wavelengths. The *idealized* version of the experiment could be based on measuring the increment of response (distance) due to equal energy quasi-monochromatic stimuli with regard to a common reference.

prepared as inputs to evaluate image-computable models. The first version of such *low-level Turing Test* (back in 2022) included stimuli for 10 well known behaviors (our *Decalogue*). That decalogue is being extended to 20 facts in our (by 2025) on-going collaboration with Prof. Bowers [89].

The selected stimuli (which include color, texture, and motion) are behind the current understanding of early vision as a set of linear-nonlinear layers [90, 91, 33, 34, 92, 35, 93]. Our proposal follows the tradition of previous (too simple) low-level datasets such as the OSA ModelFest initiative [94], but, low-level psychophysics has not been extensively included in the (today popular) BrainScore [95], nor in the high-level criticisms done by Bowers et al. [1, 2].

3 Our Proposal: A low-level vision Turing test for deep-nets

3.1 The Decalogue: facts and foundations

The set of facts and associated stimuli included in our proposal is summarized Table 1. Among the rich literature of low-level visual psychophysics, the selection of those specific facts (and associated stimuli) is founded in two main reasons.

First, they describe the visual information adaptively captured (and discarded) by the front-end of human vision. On the one hand, linear sensitivities describe the spectral, chromatic, and spatio-temporal bandwidth and relative weight given by the system to the frequency components of the input stimuli. This linear description in terms of sensitivity filters is the first order approximation to the visual bottleneck. More interestingly, this bottleneck is adaptive: then in classical models of vision science, extra nonlinear mechanisms are proposed to be in between the linear filters to account for the adaptive responses to the specific eigen-stimuli of the linear filters. The stimuli in the test we compile here were specifically designed to probe those linear and nonlinear mechanisms of human vision. The power and relevance of the selected stimuli for a complete characterization of the low-level bottleneck of image-computable models is suggested by the fact that, for decades, the straightforward use of these facts (with minor or no optimization at all) led to competitive image [45, 46, 47, 48, 49] and video coding algorithms [50, 51, 52, 53] and distortion metrics [96, 97, 62, 98, 99, 65] equipped with color constancy and contrast adaptation [38, 39, 100]. Checking if the response of a network is human-like for those stimuli would imply that the bottleneck of the network would have *statistically* good adaptive behavior [101, 102, 24, 103, 25, 26, 104, 70, 105].

Second, effects elicited by the selected stimuli are visually compelling and hence, the user of the test can check (by the eye) if the model under consideration behaves like humans or not. On the one hand, sensitivity surfaces to simple (isolated) stimuli are standardized and ready for direct quantitative comparison [84, 106, 107, 108, 109, 110, 111, 112]. On the other hand, as illustrated below, nonlinear responses when using stimuli in a context (under adaptation) have specific qualitative behaviors that are easy to see. In this way, that eventual model deviations from human-like behavior are easy to detect.

	Facts	Stimuli	Modality	Response
1	Spectral Sensitivities (achromatic and opponent)	Quasi-spectral	Color	Linear
2	Brightness & Color Response Saturation	Color calibrated	Color	Non-linear
3	Achromatic Contrast Sensitivity (Bandwidth)	Achrom. Gabors/noise	Texture	Linear
4	Chromatic Contrast Sensitivity (Bandwidth)	Chrom. Gabors/noise	Texture	Linear
5	Spatio-Chromatic Receptive Fields	Deltas / noise	Texture	Linear
6	Nonlinear Contrast Response: Saturation	Gabors/noise	Texture	Non-linear
7	Nonlinear Contrast Response: Frequency order	Gabors/noise	Texture	Non-linear
8	Context effects: Energy	Gabors/noise	Texture	Non-linear
9	Context effects: Frequency	Gabors/noise	Texture	Non-linear
10	Context effects: Orientation	Gabors/noise	Texture	Non-linear

Table 1: Facts (and associated stimuli) of our Decalogue that are behind the current understanding of the information bottleneck happening between the retina and the V1 cortex. They include the color, texture and motion processing abilities of human early vision. In the original literature the stimuli were specifically designed to probe the linear or the nonlinear behavior of the system.

3.2 The Decalogue: specific examples

In this section we show four examples of the proposed Decalogue with series of calibrated stimuli (from the colorimetric and the spatial perspectives) that illustrate the nonlinear response of humans to (i) luminance in different backgrounds leading to different perceptions of *brightness*, (ii) deviations in opponent color directions under different induction conditions leading to different perceptions of *hue* and *saturation*, (iii) texture masking due to the energy of the background, and (iv) texture masking due to the similarity between the features of the background and test.

It is important to note that the facts illustrated here (facts 2, 8, 10) are examples of the curves that are not standardized, as opposed to other facts in the proposed Decalogue (facts 1, 3, 4, 6, 7), in which strict quantitative comparisons are possible. The fact that, even in these non-standardized examples, the qualitative behavior is so compelling implies that the associated stimuli are useful to check, rank, and eventually rule out, artificial models.

3.2.1 Luminance and brightness

The first set of stimuli refers to series of luminance-calibrated achromatic samples that illustrate the perception of brightness in backgrounds of different luminance. They illustrate the Weber law [84, 113] and the crispening effect [114], i.e. the achromatic part of fact 2 in Table 1. These effects have been related to the statistics of natural images [115, 25] and with sophisticated models of retinal adaptation [92].

Figure 5 show a series of these stimuli in the (linearly spaced range of luminance $[0.5, 120]$ cd/m^2) on (linearly spaced) backgrounds of luminance in the range $[1, 160]$ cd/m^2 . These stimuli are easily generated in digital levels (i.e. ready to feed conventional artificial models) with the code provided in this work⁵ which makes use of the calibration of the Matlab toolbox Colorlab [116] in a standard computer screen.

Let's describe the perceived brightness of the stimuli in this test.

First, the series of stimuli in the darkest background clearly show the saturation nonlinearity of the brightness vs luminance curve: note that the jumps in perceived brightness for the low-luminance tests are distinctly bigger than the equivalent jumps for the same increments in luminance at the high-luminance end. In the axis of perceived brightness, the above implies that the response (blue curve) has large slope (high sensitivity) at the low-luminance end, and a saturation of such response (lower sensitivity) at the high-luminance end. That makes the *qualitative* saturating blue curve of brightness vs luminance.

Second, when one increases the luminance of the background (e.g. from $1 \text{ cd}/\text{m}^2$ to $40 \text{ cd}/\text{m}^2$), the brightness of the (same) samples is lower than in the previous series, so the *qualitative* brightness response to this second series of stimuli is below the previous one (as depicted by the *qualitative* black curve).

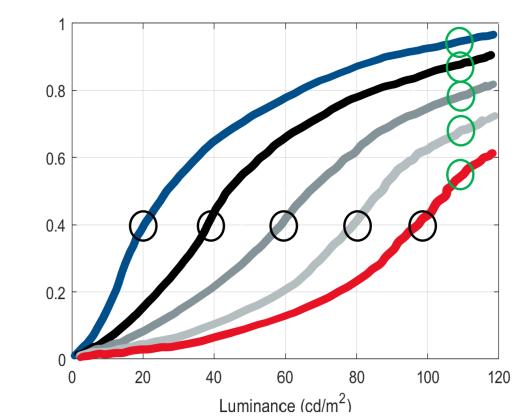
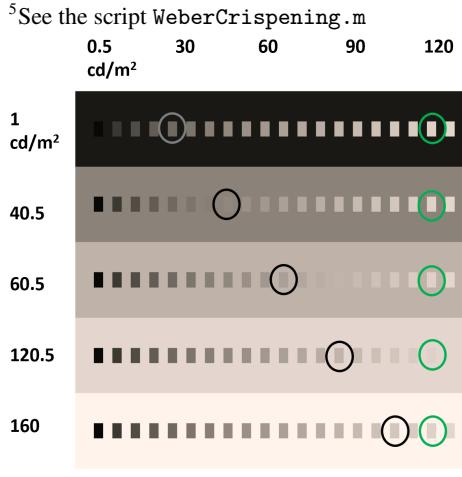


Figure 5: Series of stimuli eliciting nonlinear brightness perception. Luminance calibrated linearly spaced tests in different luminance calibrated backgrounds. This illustrates the Weber law [113] as well as Whittle's crispening effect [114], as summarized in [92].

Third, by looking at the stimuli highlighted in gray in the 1 cd/m^2 and the 40 cd/m^2 backgrounds it is obvious that in the brighter background the stimuli with equivalent brightness are shifted to the right in the scale of luminance, which means that the response in black (for the stimuli in the brighter background) is shifted right-down with regard to the curve in blue (for the stimuli in the darker background). Moreover, this means that the black curve has sigmoidal shape as it should start from zero brightness. Similar visual reasoning implies that this shift progressively increases as one increases the luminance of the background, as *qualitatively* illustrated by the samples highlighted in gray.

Fourth, the (same) stimuli in the brightest background in elicit a brightness response with substantially different shape: the sigmoid has substantially shifted to the right (red curve) and, all in all, one can see a smooth transition of the sigmoidal response curves from the blue curve to the red curve. The crispening effect (increased sensitivity around backgrounds of similar luminance) is illustrated by the shift to the right of the points of maximum slope in the response curves.

Finally, **fifth**, the decreasing brightness of the samples of the same luminance in backgrounds of progressively bigger luminance (as illustrated by the samples highlighted in green) illustrate brightness induction [113].

Of course the *qualitative* visual observations done here, by no means try to substitute the rich *quantitative* literature in which these responses are determined by accurate psychophysics [84, 113]. However, (1) the phenomena are compelling enough so that one can see the qualitative trends of the curves by the eye, and, as seen in the numerical experiments below, (2) these trends (visible in ready to use digital images) are enough to spot divergences with human behavior in certain artificial model or discriminate between models in terms of their similarity to human behavior, which is the ultimate goal of the tests presented here.

3.2.2 Nonlinear response to saturation and color adaptation

Responses to constant deviations from white in the red-green and yellow-blue directions of the Jameson & Hurvich color space [106, 117] with equiluminant stimuli describe the nonlinear perception of hue and saturation as pointed out in [118, 119] in similar opponent spaces, i.e. the chromatic version of fact 2 in Table 1.

Figure 6 shows colorimetrically calibrated stimuli with such deviations (in the range [-20, 20] of the linear RG and YB tristimulus values of the Jameson and Hurvich space) in different backgrounds, which are easy to generate and modify by using the code provided in this work⁶. As in the previous test, let's describe the perceived hue and saturation of the stimuli to infer the qualitative shape of the responses.

First, take the stimuli in gray backgrounds and note that the jumps in perceived hue are bigger around the central (achromatic) stimulus than in the extremes with more saturated stimuli (either red, green, yellow or blue): judge the jumps in saturation close to the achromatic stimulus and at the extremes of the chromatic axes. Similarly to the responses for brightness, these differences imply a sigmoidal response to saturation when the stimuli linearly depart from white in constant steps: see the qualitative responses in gray for both the red-green and the yellow-blue directions. **Second**, these sigmoidal

⁶See the script `Gegenfurther.m` of this work which also uses the Toolbox Colorlab [116] for calibration.

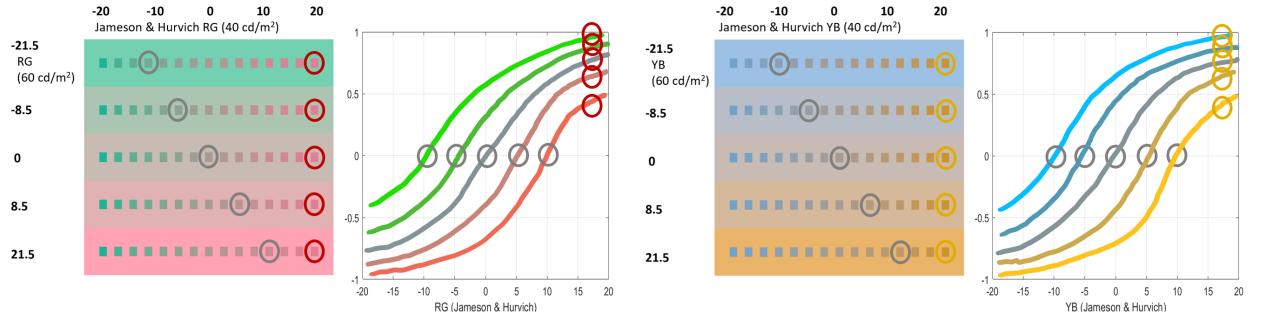


Figure 6: Series of nonlinear perceived saturation (or response of the opponent chromatic channels) versus linearly spaced increments in colorimetrically calibrated color opponent directions in different chromatic contexts. This illustrates the nonlinear effects pointed out in [118, 119].

responses shift to the right or to the left as can be seen from the shift of the stimuli that are perceived as achromatic in the different backgrounds (e.g. see the stimuli highlighted in gray). Note that a stimulus is seen as achromatic when the response of the mechanism tuned to red-green or yellow-blue is zero. See the corresponding shifts in the zero crossings of the sigmoids (also highlighted in gray). Finally, **third**, the shift of the responses is bigger as the saturation of the background is increased.

Again, the goal of this test is not substituting the original accurate psychophysics done on humans [118, 119] to point out these phenomena. On the contrary, they just represent an easy way to get digital images that can be used to test artificial models and check if their responses qualitatively behave like humans.

3.2.3 Texture masking 1 (energy): nonlinear adaptive contrast response

The same kind of qualitative derivation of human-like responses can be applied to the perceived contrast of textured patterns with calibrated frequency content and controlled luminance. The test presented here illustrates the fact that perceived contrast nonlinearly depends with linearly increasing Michelson contrast [120, 121] and this response decreases with (is masked by) the energy of a background of similar texture [122, 123]. This corresponds to fact 8 in Table 1. The stimuli presented in the following example can be reproduced and modified both in frequency orientation, average luminance and contrast with the code provided⁷.

Figure 7 shows Gaussian windowed test noise patches of 4 cycles/degree (cpd) in images subtending 1 degree with average luminance of 50 cd/m^2 and linearly spaced RMSE contrasts (from left to right) in the range $[0, 0.3]$. The different rows show the same tests on different backgrounds of noise of 4 cpd with linearly spaced RMSE contrast in the range $[0, 0.25]$.

Similarly as in the previous cases, let's describe the perceived contrast along the two dimensions of the panel. Test: left to right, and background: top to bottom. Again, the qualitative shape of the responses will be determined by the perceived jumps of contrast of the tests (from left to right) and by their variation as one increases the energy of the background (from top to bottom).

First, for the zero contrast background (first, top row) the jumps in perceived contrast in the low-contrast end (left) are bigger than the jumps in perceived contrast in the high-contrast end (right). See the differences in perceived contrast in the tests highlighted in blue. This implies a saturating contrast response curve (as in the previous examples), i.e. the blue curve.

Second, as the contrast of the background is increased (see stimuli highlighted in orange) the perceived contrast of the test is reduced. This implies that subsequent curves (black and lighter shades of gray) are below the initial blue curve.

⁷See the script `MaskingEnergy.m` which makes extensive use of the Toolbox Vistalab [124].

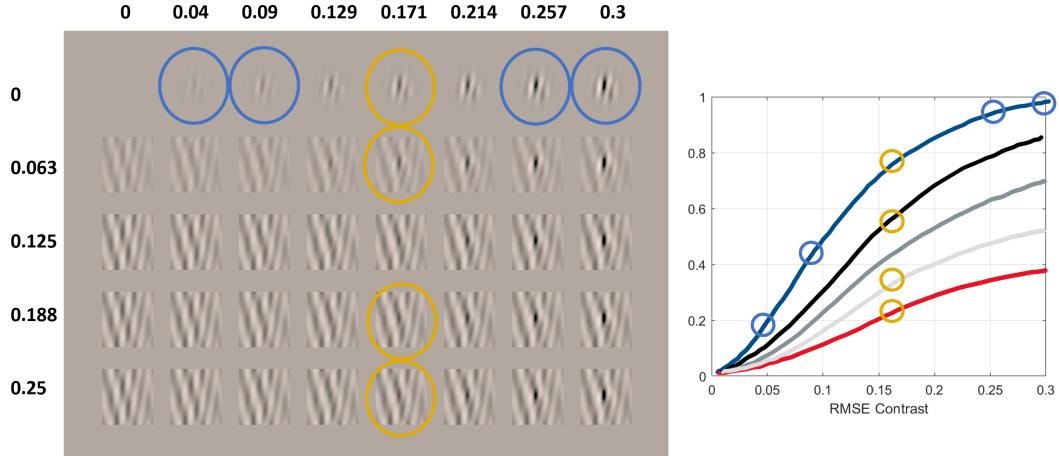


Figure 7: Series of nonlinearly perceived contrast (or response of the mechanisms tuned to certain texture) versus linearly spaced increments in contrast calibrated test of controlled spatial frequency in backgrounds of different (controlled) energy. This illustrates the nonlinear effects pointed out in [109, 120, 121, 110].

Finally, **third**, as in order to perceive the tests with equivalent contrasts in backgrounds of progressively bigger energy the necessary contrast of the test increases, this means that sigmoidal curves shift to the right.

As in the previous examples, the qualitative behavior illustrated by this series of digital images generated by our code should give (in artificial models) corresponding saturating curves with smooth variation from the blue (zero contrast background) condition to the red (high contrast background) condition, and hence the lower response curve.

3.2.4 Texture masking 2 (features): interaction between orientations

Reduction of sensitivity (the so called masking) also happens when certain test is presented on top of a background that shares some feature with the test [125, 122, 123], i.e. facts 9 and 10 in Table 1. The next example, Fig. 8, refers to the specific case of interaction between orientations of test and background. It can be reproduced and modified both in frequency, orientation, contrast and average luminance with the code provided⁸.

Figure 8 shows 6 cpd horizontal Gabor patches with average luminance of 50 cd/m^2 and RMSE contrast increasing linearly from left to right in the range $[0, 0.3]$. These Gabor patches are shown on top of band-pass noise of contrast 0.2, with the same frequency, but different orientation. The numbers in the different rows shows the angular difference between tests and backgrounds. The figure shows some compelling facts that lead to clear qualitative trends in the response curves.

First, the test is better seen (has bigger visibility or perceived contrast) when the background is orthogonal to the test (in the first row). In that row the different jumps in visibility in the low-contrast and high-contrast ends (tests highlighted in cyan) indicate a saturating response as in the previous examples (response curve in blue).

Second, the necessary contrast to detect the test smoothly increases as the difference in orientation between test and background decreases: see that the tests highlighted in blue, black, shades of gray and red, approximately have the same visibility over the different backgrounds with angular differences in the range $[90, 0] \text{ deg}$. The trend is similar for negative angular differences. This implies a smooth variation (decrease) of the response curves in terms of the difference between test and background.

⁸See the script `MaskingOrient.m` which makes extensive use of the Toolbox Vistalab [124].

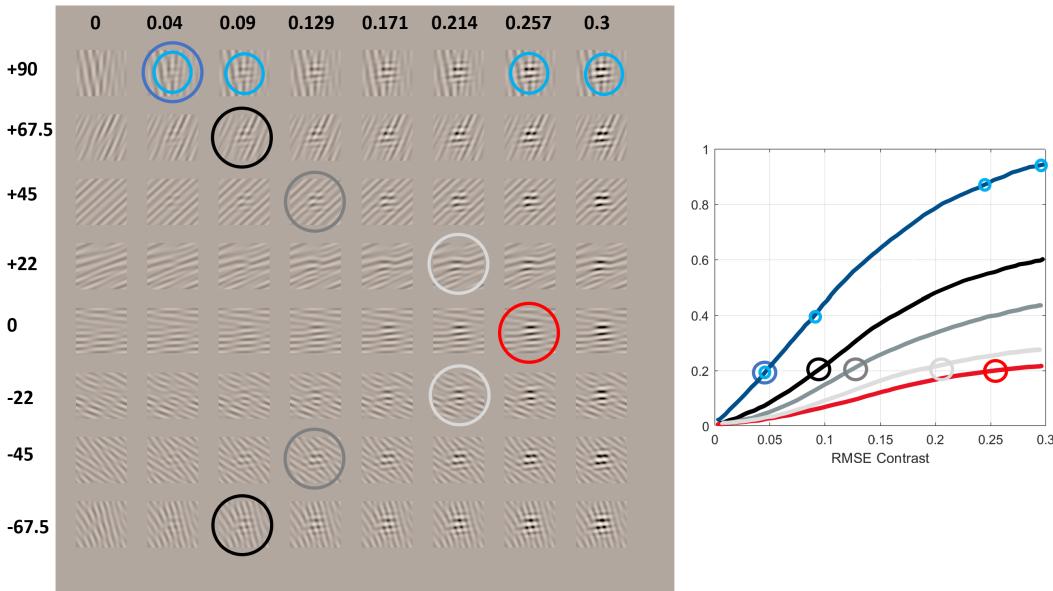


Figure 8: Series of nonlinearly perceived contrast (or response of the mechanisms tuned to certain texture) versus linearly spaced increments in contrast calibrated test of controlled spatial frequency in backgrounds of different orientation. This illustrates the nonlinear effects pointed out in [122, 123].

Third, the biggest masking is obtained when test and background are aligned (the red curve is clearly the lowest response curve). This and the previous fact imply that the general trend is this smooth transition of the nonlinear curves from the blue to the red.

3.3 Proposed methodology

Our proposal is simple: just use the code provided here⁹ to generate the stimuli (digital images well calibrated in luminance, color and spatio-temporal frequency) that illustrate the 10 compelling facts listed in Table 1 that describe the adaptive information bottleneck of low-level human vision. The resulting digital images are organized in series that correspond to progressive stimulation of a vision system in particular ways. The interesting point is that this set of controlled stimulation conditions lead to intuitive responses (as shown above), or even to standardized sensitivity curves or surfaces that are also provided with the code.

Once stimuli are generated they are used to feed any artificial image-computable model. Then, depending on the model, the user decides where to read from the network under consideration and the read-out mechanism to get *visibility* values to generate artificial series of response curves.

In the case of facts 2, 8, 9 and 10 these curves have to be compared with the kind of qualitative curves described above, which given the clarity of the selected stimuli can be drawn by simple visual observation of the stimuli as described above.

In the case of fact 5 (existence of center-surround and Gabor-like receptive fields tuned to achromatic, red-green and yellow-blue patterns [126]), the more straightforward method is checking their presence by reading the response to deltas from single neurons or from the Jacobian of the network at that layer [33, 20, 21]. Other indirect methods could be (1) using reverse correlation feeding the network with controlled noise (also generable using Vistalab [124] following the appropriate literature [83]), or (2) using artificial psychophysics based on adaptation (e.g. the Blakemore and Campbell experiment [127]). However, this very last method to measure fact 5 relies on fulfillment of adaptation facts 6-10, which may not hold in non-human networks.

In the above (non-standardized) cases the general trends of the curves can be qualitatively assessed in detail: general shape of the curves, the blue response and the red curve being the biggest and the lowest respectively, and the transition from one to the other. Note that user of the provided code can change the parameters of the stimuli and infer new curves by applying a similar visual analysis. For the receptive fields they can be analyzed using shape parameters in the spatial or the Fourier domain as classically done in visual neuroscience [80, 82, 128, 129] and the same for the chromatic tuning in standard color spaces [81, 130, 20].

Finally in the case of sensitivity curves or surfaces which are standardized or available in the code (facts 1, 3, 4, 6 and 7) the visibility values obtained from the models can be numerically compared with the provided ground truth.

This qualitative/quantitative methodology is summarized in Fig. 9, and applied in the next experimental section for three illustrative networks.

⁹The Decalogue Toolbox is available here: <http://isp.uv.es/docs/TuringTestVision.zip>

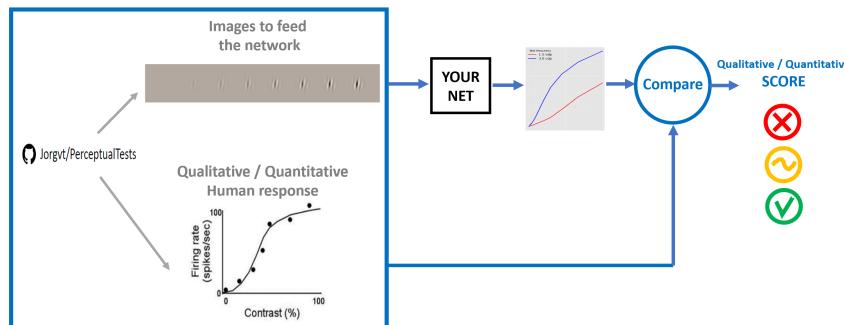


Figure 9: The proposed method: feed the model with series of images, compute responses (using the simplest possible read-out mechanism) and make quantitative comparisons with standard sensitivity surfaces or qualitative comparisons checking the nonlinearity using different adaptation conditions.

4 Experiments: analysis of three illustrative deep-models

4.1 Networks and experimental setting

In our experiments we check the behavior on the proposed *Decalogue* of three recent networks of similar architecture:

1. A parametric vision model, the *BioMultiLayer* network [33], which consists on a cascade of four linear+nonlinear stages that account for (1) color opponency and adaptation, (2) contrast computation, (3) contrast sensitivities and energy masking, and (4) wavelet analysis and cross-masking between textures. The linear parts of all the stages were not optimized but they were directly inspired by classical psychophysical or physiological literature. The nonlinear parts were implemented via Divisive Normalization [131, 49, 65, 35, 132]. The nonlinearities of the 2nd and 3rd stages of the model were tuned via the psychophysical method of Maximum Differentiation in [32]. And the nonlinear parts of the 1st and 4th stages were tuned to reproduce subjective opinions on distortion and contrast masking facts [33, 34]. The statistical properties of the model and its relations with recurrent models were studied in [104] and [35] respectively.
2. A non-parametric model to predict subjective image quality, the *PerceptNet* [36], which starts with a nonlinear front-end at the retina followed by a cascade of three linear+nonlinear stages. The architecture was intended to accommodate similar vision facts that motivated the *BioMultiLayer*. The *PerceptNet* architecture is similar to AlexNet [133] but its nonlinearities were formulated using an end-to-end optimizable Divisive Normalization [134, 37]. Both the linear and the nonlinear parts of *PerceptNet* were end-to-end tuned to maximize the correlation with humans on subjective image distortions [36]. Non-parametric layers of *PerceptNet* are not easy to interpret as pointed out recently [135].
3. An image segmentation model, the *Bio U-Net* [38], with the same style encoder as the non-parametric *PerceptNet* (a cascade of linear + divisive normalization stages), but augmented with a decoder that recovers the original dimension of the input signal and predicts a class per pixel for semantic segmentation. The encoder and the decoder were tuned to optimize segmentation in different databases. The benefits of the biologically-inspired nonlinearities of this model for segmentation have been further studied thereafter [39].

We assumed a visual field of 2 degrees with a sampling frequency of 64 cycles/deg, i.e. we fed the models with 128×128 images. We measured the responses of the models to specific tests through the Euclidean departure between the response to test+background with regard to the response to the isolated background.

4.2 Results

4.2.1 Spectral sensitivities and color responses (properties 1 and 2)

Figure 10-top shows the response of the models to quasi-monochromatic stimuli¹⁰ to get the spectral sensitivity of the neurons (property 1). In order to point out the relevance of the layer where responses are measured from, in the case of the *BioMultiLayer* network, we consider direct read out of the response (with sign) in the first linear layer (subplots A and B) and in the last nonlinear layer (subplots C and D). In this network the first linear layer has achromatic and opponent channels defined by construction so the V_λ [84] (subplot A) and the opponent curves of Jameson & Hurvich [106] (subplot B) are trivially obtained. Interestingly, the spectral sensitivities at the last nonlinear layer are wide-band positive in the first channel of the network and opponent in the other two channels, but their shapes are substantially modified with regard to the human-like behavior at the first layer. These differences justify the qualitative scores given in each case. We can conclude that spectral sensitivity in this model is human-like at the front-end but degrades throughout the network. In other words, as suggested in section 2-b, read-out location matters and certain kind of information should be extracted from a specific place of the model.

¹⁰Spectrally narrow Gaussians (5 nm width) of constant energy centered on different wavelengths along the visible spectrum on top of a low energy flat spectrum, as in Fig. 4. In this way all the stimuli can be faithfully represented in digital values.

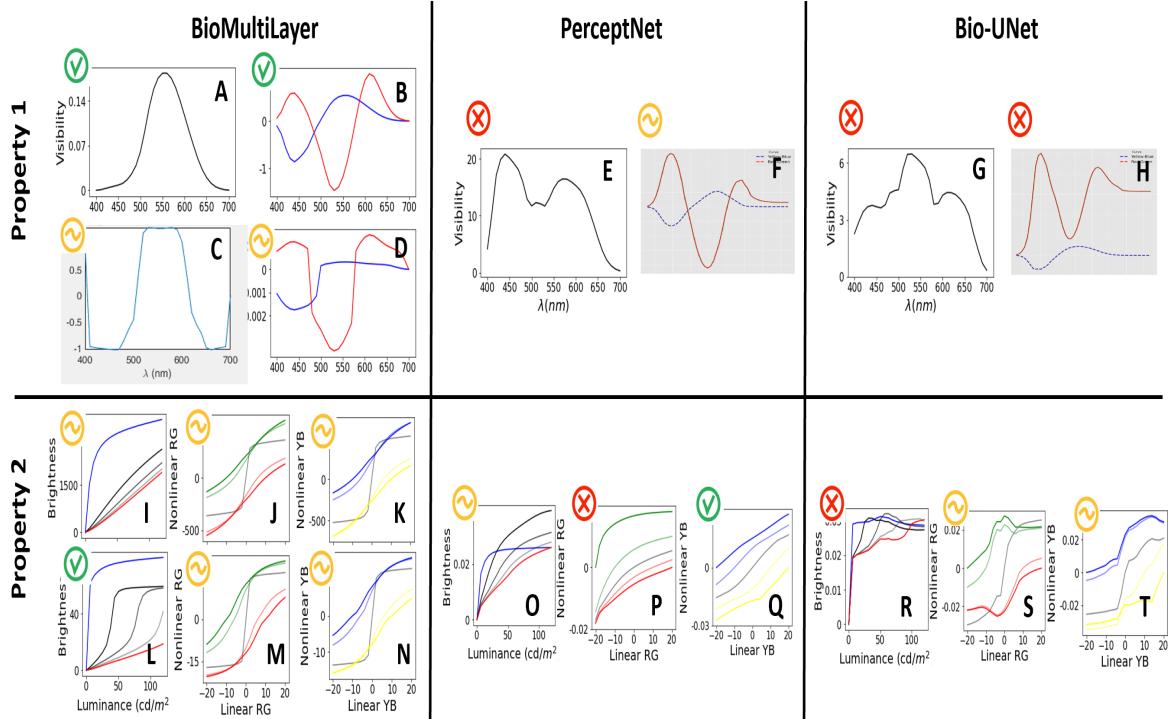


Figure 10: Spectral sensitivities of the considered models (top) and corresponding responses to luminance and linear deviations from white in the cardinal red-green and yellow-blue directions (bottom). On the one hand, knowledge of the standard spectral sensitivity, the CIE V_λ curve [84], or the standard spectral sensitivity of the opponent channels [106] indicates which model is trivially correct in Prop. 1. On the other hand, the stimuli proposed here (Figs. 5 and 6) and the associated human behavior described above indicates the correct trends in the responses of Prop. 2.

The *PerceptNet* has a color space change after the retinal nonlinearity. There is where we measure the sensitivities as the design idea was that achromatic and chromatic channels emerged at that stage. Results show that the first channel displays an all-positive but bimodal response (subplot E) and the other two channels display opponent-like responses (subplot F).

The very same location of the encoder of the segmentation *Bio-U-Net* has very different sensitivities despite it has the same architecture as the *PerceptNet* up to that layer. The sensitivity of the (supposedly) achromatic channel is very noisy and the other two channels display qualitatively opponent oscillations but they are shifted in absolute value (subplots G and H).

On the other hand, Fig. 10-bottom checks property 2 by showing the responses to (i) luminance and to deviations from white in the (ii) red-green and (iii) yellow-blue directions (left, center and right respectively). In the achromatic case, tests in the range $[0.5, 120] \text{ cd/m}^2$ are shown on top of backgrounds of different luminance in the range $[1, 160] \text{ cd/m}^2$. The response curves in different backgrounds are depicted in blue, black, and progressively lighter shades of gray until red, as in Fig. 5. In the chromatic cases, responses are computed with tests on an achromatic background (black curve) and on backgrounds of progressively saturated color (reddish and greenish curves and blueish and yellowish curves as in Fig. 6).

For the *BioMultiLayer* model we have such responses for two different layers: first (I, J, K) and fourth (L, M and N). The achromatic response of the first layer is certainly nonlinear for the darkest background, and the response gets attenuated when the luminance of the background is increased (see the transition from curves in blue to red in subplot I). However, these responses do not reproduce the crispening (sigmoids shifting to high luminance), and responses for high luminance backgrounds are too linear. As a result the achromatic behavior of this layer has been qualified as non-human. The chromatic responses display sigmoidal shape and they shift in the right directions under different backgrounds (subplots J and K). However, the nonlinearities for the chromatic backgrounds are very smooth compared to the sharpness of the nonlinearity for the achromatic background. As a result, the human similarity of chromatic behaviors have been qualified as intermediate. In contrast, the achromatic response of the 4th layer (subplot L) does reproduce the nonlinear behavior and crispening,

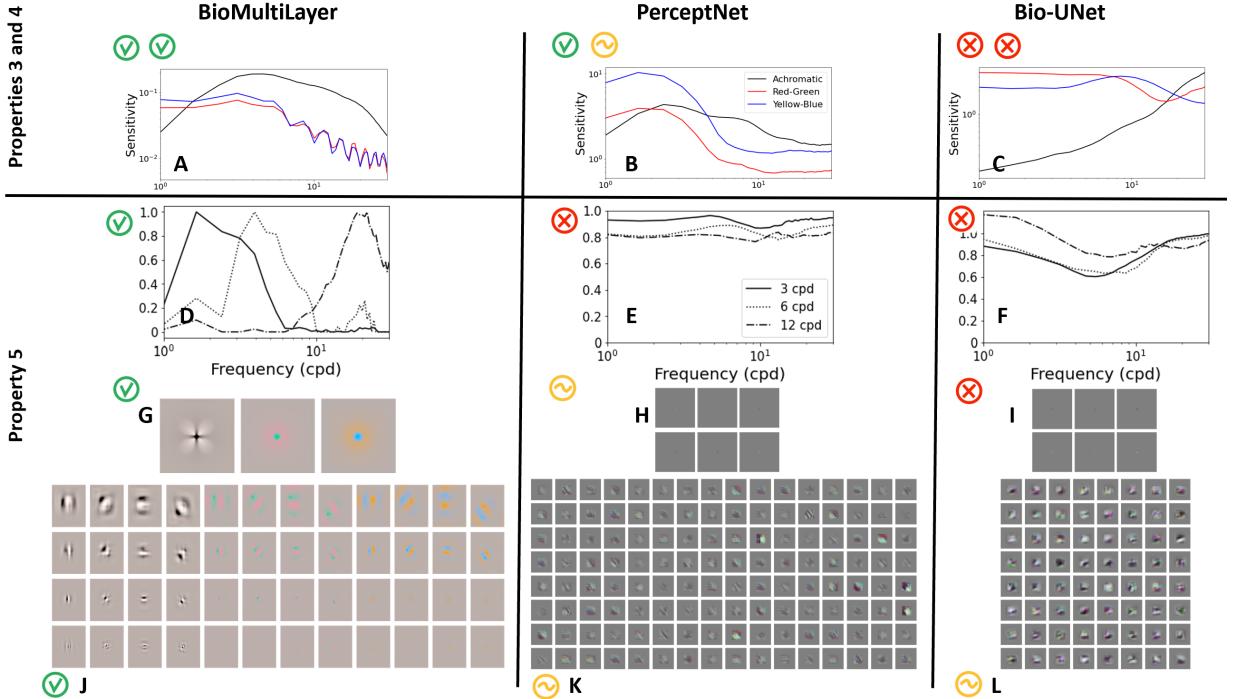


Figure 11: Achromatic and chromatic contrast sensitivities of the considered models (top subplots A, B, C), and different sets of receptive fields computed in different ways at different depths of the models (bottom). See text for details on the psychophysical and the physiological ways to estimate the receptive fields.

so it has been qualified as more human-like than the achromatic response of the 1st layer. Shifts of the chromatic nonlinearities are stronger depending on the background, but the nonlinearities in achromatic backgrounds (black curves in subplots M and N) are still too sharp. Therefore the score remains the same.

The *PerceptNet* model displays nonlinear behavior and crispening in the responses to the achromatic series (subplot O). However, Note how the curves corresponding to light backgrounds exceed the response on dark backgrounds, so human similarity has been qualified as intermediate. The responses to red-green series in *PerceptNet* shift in the right directions on different backgrounds, but they are too linear (and hence wrong) in subplot P. In contrast, the blue-yellow responses (subplot Q) display a rather human behavior.

Finally, the *Bio-U-Net* shows a clearly non-human achromatic response: note the noise and wrong order in the curves with no trace of crispening (subplot R). In contrast, the responses to the chromatic series display the expected sigmoidal shape with the shift in the proper directions for the different chromatic backgrounds (subplots S and T). Noisy and unstable responses is what determined the intermediate score.

4.2.2 Achromatic and chromatic contrast sensitivities and receptive fields (properties 3, 4 and 5)

The top row of Fig. 11 shows the achromatic Contrast Sensitivity Function (property 3, black curve) and the red-green and yellow-blue Contrast Sensitivity Functions (property 4, red and blue curves respectively). These CSFs have been computed from the responses to noise patterns of controlled spatial frequency and the same low contrast ($C_{RMSE} = 0.05$) for every frequency. Patterns were generated in the corresponding color channel of the Jameson & Hurvich color space [106] that isolate luminance, red-green, yellow-blue components. We consider the responses at the last layer of the networks and we plot the Euclidean distance between the responses for each pattern and for a flat image of the same average color.

The CSFs of the *BioMultiLayer* model (subplot A) strongly resemble the human CSFs [107, 108]: the achromatic response is band-pass with peak sensitivity around 4 cpd and high cut-off frequency

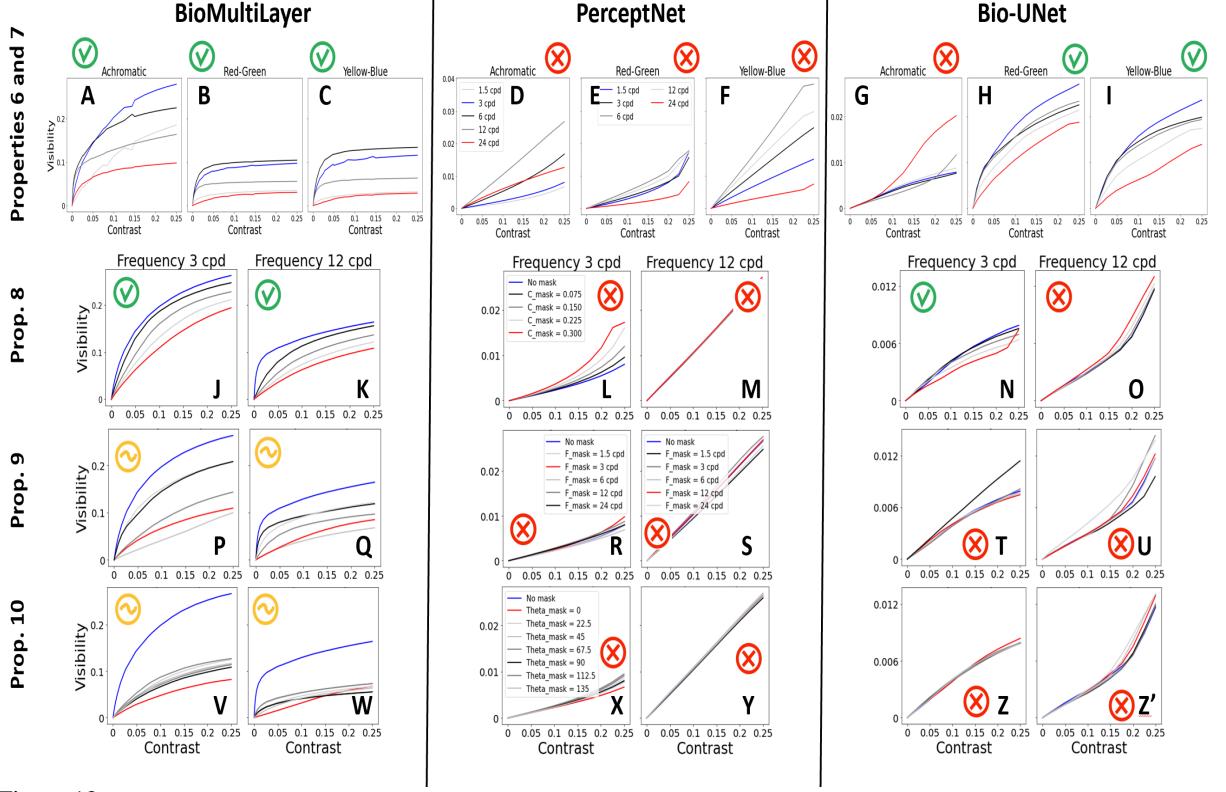


Figure 12: Contrast responses of the considered models in different masking conditions and for achromatic and chromatic textures of different frequencies. The color code (indicated in the subplots corresponding to *Percepnet*, but applicable to the equivalent curves of the other models) has been designed so that human response curves would be in the blue-red order, as in Figs. 7 and 8. In this way it is obvious which model reproduces better the human behavior.

(above 32 cpd), and the chromatic responses are lower and basically low-pass with cut-off frequencies about 15 cpd.

The achromatic CSF of the *PerceptNet* is also band-pass (black curve in subplot B), but the chromatic CSFs are far from human because their shape is also bandpass and the responses to modulations in the YB direction are way bigger than the responses to equivalent achromatic modulations.

Finally, the *Bio-U-Net* (subplot C) displays a strongly non-human behavior: see the non-plausible high-pass behavior of the responses to achromatic gratings, and the bigger responses to chromatic gratings in the mid-frequency range.

The receptive fields of the models (property 5) have been proved in two ways: (1) a *psychophysical* method based on the Blakemore & Campbell experiment [127], which relies on the attenuation of the CSF under adaptation for different frequencies, and (2) a *physiological* method based on recording the response to deltas in the luminance, red-green and yellow-blue channels [33]. In the *BioMultiLayer* model the attenuation of the achromatic CSF when the gratings are shown on top of backgrounds of specific frequencies (subplot D) reveals the existence of narrow-band sensors with bandwidth that increases with frequency, which is consistent with human behavior [127, 136]. This comes from the fact that the linear part of the 4th layer of this model is made of wavelet kernels and their response is nonlinearly attenuated by the activity of neighbor sensors tuned to the same feature through Divisive Normalization.

On the other hand, when checking the shape of the receptive fields using delta functions one gets two biologically plausible results: (a) in the 3rd layer of the *BioMultiLayer* network receptive fields are center-surround patterns in the achromatic, red-green and yellow-blue directions (subplot G), and (b) in the 4th layer one gets local frequency filters with different orientations and scales (subplot J) as happens in biological vision at LGN [137, 126] and V1 [78, 138].

For *PerceptNet* results are quite different: first, the Blakemore and Campbell experiment only shows filters tuned for low frequencies (subplot E) which does not correspond to human behavior. This is not only due to the non-human nature of the CSFs, it also means that any frequency leads to a strong attenuation of the responses to high frequency patterns. This departure from human behavior is also visible when getting the receptive fields from the last layer of the network using deltas: one gets oriented filters but all with the same size and with low-frequency blobs. More over, chromatic information is spread along all the filters (subplot K), as opposed to what happens in the early layers where one does get achromatic, red-green and yellow-blue responses (subplot H).

Finally, in the *Bio-U-Net* model, the Blakemore and Campbell experiment also leads to non-human ratios of the CSFs (subplot F). The receptive fields obtained from deltas in the first layers lead to center-surround blobs despite not in definite chromatic directions (subplot I). In the central layers of the encoder one gets larger receptive fields which have no clear spatial oscillations nor preferred chromatic directions (subplot L).

4.2.3 Contrast saturation, dependence with frequency (properties 6 and 7)

The top row of Fig. 12 shows the visibility response (Euclidean difference of response wrt the response of a uniform gray image) for achromatic patterns of different frequencies and for red-green and yellow-blue patterns of different frequencies all seen in isolation (properties 6 and 7). Line styles for the different frequencies in the achromatic and the chromatic cases is different according to the different order expected for band-pass and low-pass systems. In every case, match with human behavior would be illustrated by having the blue curve at the top and the red curve at the bottom, with a smooth transition from black to light gray, in between.

The *BioMultiLayer* model leads to saturating responses with larger intensities in the achromatic case (left) than in the chromatic cases (subplots A and B-C), as in humans [123, 139]. The achromatic response to mid-frequency (3 cpd, in blue) is clearly bigger than the response to the other frequencies, which is smoothly reduced for higher frequencies (from 6 to 24 cpd) and also attenuated for 1.5cpd. On the other hand, the chromatic responses are basically ordered according to frequency in a low-pass fashion. All these trends are in good agreement with the human behavior.

The achromatic responses of the *PerceptNet*, though band-pass, exhibit a quite linear, non-saturating or even expanding, behavior (subplot D). Moreover, these achromatic responses are not bigger than the response to chromatic patterns, particularly the yellow-blue (subplot F), which is contrary to human perception.

The responses for the chromatic patterns in the *Bio-U-Net* model exhibit human-like saturation and they are in the right frequency order (subplots H and I), but they are larger than the responses for achromatic patterns (subplot G), which is contrary to human perception.

4.2.4 Energy masking and feature masking (properties 8-10)

Each panel of the second row in Fig. 12 shows the responses to a 3 cpd achromatic pattern (left) and a 12 cpd achromatic pattern (right) seen on top of a masking pattern (noise of the same frequency and orientation) with progressively larger RMSE contrast (in the range [0,0.3]) leading to different response curves in different color (from blue to red), thus checking the effect of the energy of the background (property 8). The color code has been selected so that the no-mask case is depicted in blue (less attenuated in humans) and colors from black to light-gray and red are taken for progressively bigger contrasts of the mask.

The responses of the *BioMultiLayer* in Fig. 12 (subplots J,K) progressively attenuate as the energy of the background is increased in line with the reduction in visibility of the test shown in each column of Fig. 7. And this happens both for low and high frequency, with bigger responses for the mid-frequency. Therefore the behavior is qualitatively human. The *PerceptNet* displays a completely non-human behavior: for the 3 cpd tests, progressively larger masks induce enhancement of the expansive (non-saturating) response, and the responses for the high frequency patterns are larger, linear and do not show significant variation with the mask. Finally, the *Bio-U-Net* model does display human-like attenuation of the response to 3 cpd patterns (subplot N). However, the responses to 12 cpd patterns (subplot O) are not human-like because their (large) size and their expansive shape and increase with the energy of the mask.

The panels of the third row of Fig. 12 show the responses for an achromatic test of 3 cpd (left) and 12 cpd (right) seen on top of backgrounds of different frequencies (and 0.2 contrast) compared to the no-mask condition, i.e. it checks the frequency cross-masking (property 9). The color code has been selected so that the no-mask case is depicted in blue (less attenuated in humans) and colors from black to light-gray and red are taken for progressively closer frequencies in mask and test, which lead to increased attenuation of response in humans.

The response of the *BioMultiLayer* model is bigger in the no-mask condition, displays substantial attenuation when the background shares the same frequency of the test (red curves in subplots P and Q) and responses are bigger for 3 cpd than for 12 cpd. In each case the optimal frequency is not the one that leads to the bigger attenuation, but it is close to it. The *PerceptNet* responses are not human because for the low frequency, subplot R, responses are not-saturating regardless of the mask, and the responses for high frequency are larger, linear and the presence of backgrounds leads to larger responses (subplot S). The *Bio-U-Net* does not show human-like trends because in the case that displays a saturating response the presence of a background leads to responses bigger than in the no-mask case (black curve in subplot T). The behavior in subplot U are non-human for the same reasons stated in the subplots N and O.

Finally, the last row of Fig. 12 shows the responses for low- and high-frequency achromatic patterns (left and right, respectively) seen on top of backgrounds of the same frequency but different orientations, i.e. it checks the orientation cross-masking (property 10). Again, the color code has been chosen so that in a human the blue curve would be at top and the red would be at the bottom as in Fig. 8.

For this last example, the *BioMultiLayer* model gets bigger attenuation for the background of the same orientation, particularly for high frequency (see the red curves), and the other orientations lead to responses that are between the no-mask condition (in blue) and the same-orientation background (in red) in subplots V and W. The other models give clearly non-human results because (on top of the arguments used in previous cases) stimulation on backgrounds of the same orientation (red curves) do not lead to the expected attenuation, and bigger attenuation is obtained for backgrounds that are almost orthogonal to the test, which is not what humans experience in Fig. 8.

4.2.5 Summary of results

The qualitative evaluation of the considered models over the proposed tests is summarized in Table 2. From this table there is a clear ranking of the alignment between the models and humans. It is not surprising that the parametric model (the *BioMultiLayer*) has bigger alignment in the linear parts (properties 1 and 5) since sensitivities and center-surround and Gabor receptive fields were parametrically built in that model model.

More interestingly, the band-pass behavior of the sensors emerged from modifications in the CSFs in our simulation of the Blakemore and Campbell experiment. It is also interesting the close reproduction of the band-pass and low-pass behavior and the relative scaling of the CSFs obtained from responses to sinusoids (an original check done here) since they were not built in. This indicates that the (non-trivial) gain of the center-surround cells and the Gabor cells was properly adjusted through the indirect psychophysical experiments done to set its parameters. As a result, the relative order of the (saturated) frequency responses (property 7) is also ok, both for achromatic and chromatic textures. The saturation of the responses to Gabor stimuli in isolation (property 6) is better reproduced in the parametric model than in the Bio-UNet. The difference between them is more evident when one digs in using properties 8-10 because they need proper interaction between texture sensors and this was only easy to do in a parametric model such as the *BioMultiLayer*.

However, note that the reproduction of the interaction between features (both in color, property 2, and in texture, properties 9 and 10) is not properly reproduced not even in the *BioMultiLayer* pointing out that more work is needed in adjusting its parameters as discussed below.

According to the proposed test, the other two models (the -non parametric- *PerceptNet* and the *Bio-U-Net*) are *less human*, in that order of alignment. This also makes sense because the *PerceptNet* was tuned to reproduce low-level human opinion on distortion, while the *Bio-U-Net* was just tuned to reproduce a specific mid-level vision goal such as image segmentation. In the discussion we elaborate more on the combination of goals that may explain the organization of the visual system.

In any case, we see that even with this qualitative application of the proposed test (again, quantitative comparisons could be done with properties 1, 3, 4 and 6, even for moving patterns) a significant ranking is possible, and, as discussed below, the qualitative behaviors, when they are properly understood suggest significant changes in the architectures and training of the models. Quantitative automation of the optimization should be iteratively done by alternating goals of different nature as suggested in [34]: optimize for conventional goals and then fine-tune to reproduce the effects pointed out by the test proposed here (or the other way around).

	Facts	BioMultiLayer	PerceptNet	Bio-UNet
1	Spectral Sensitivities (achromatic and opponent)	✓✓	X~	XX
2	Brightness & Color Response Saturation	✓~	~~	X~
3	Achromatic Contrast Sensitivity (Bandwidth)	✓	✓	X
4	Chromatic Contrast Sensitivity (Bandwidth)	✓	~	X
5	Spatio-Chromatic Receptive Fields	✓✓✓	X~~	XX~
6	Nonlinear Contrast Response: Saturation	✓✓	XX	X✓
7	Nonlinear Contrast Response: Frequency order	✓	X	~
8	Context effects: Energy	✓	X	~
9	Context effects: Frequency	~	X	X
10	Context effects: Orientation	~	X	X

Table 2: Summary of qualitative results is enough to discriminate between the three models.

5 Discussion: What can be learnt from the proposed methodology?

In this section, we discuss the benefits of the proposed *Decalogue* for generic artificial models. Benefits go beyond the evaluation of the human nature of models: even if we don't need that certain model is similar to humans, the behaviors described by the human-like curves elicited by the stimuli in the *Decalogue* imply human-like bottlenecks and adaptation properties that one would like in efficient and robust artificial vision systems. Similarly, we also discuss the benefits of the architectures from classical vision science models that reproduce such behaviors.

5.1 (Non-human) curves suggest changes in the architectures

When measuring the response of conventional networks using the spatially and chromatically calibrated stimuli proposed here one can get human-like behaviors such as the ones shown in Section 3. For instance, shallow autoencoders optimized for image deblurring and denoising display human-like saturation when responding to achromatic and chromatic gratings of controlled spatial frequency: see Fig. 13 (top), reproduced from [21]. In this case, the slope of the response of these autoencoders (their sensitivity) is bigger for achromatic gratings than for red-green and yellow-blue gratings and it reduces with the contrast of the gratings, just as in humans.

This contrast-dependent saturation has been described in vision science with specific input-dependent activation functions such as the Divisive Normalization [132, 103, 65, 91, 33, 34], which has been found equivalent to classical recurrent models of neural interaction in biology [104, 35]. Therefore, this behavior, and their adaptation benefits, can be enforced in conventional networks by imposing this kind of interaction in their architecture. Examples include benefits in autoencoding and compression [49, 37], denoising and enhancement [140, 134], segmentation [38, 39], classification [58, 92, 59], or robustness to adversarial attacks with few layers given the strong nonlinearity due to this biological computation [92]. Moreover, inclusion of these nonlinearities if done parametrically (e.g. by using parametric expressions in the kernel of Divisive Normalization) reduces the training time and increases generalization because of the drastic reduction in the number of parameters of the network [135].

As a result, non-human behaviors that are observed in the models using the proposed stimuli and methodology suggest changes in the parameters of the architecture. See for instance the chromatic examples and the texture examples in Fig. 14 (left and right respectively). The too-sharp behavior of

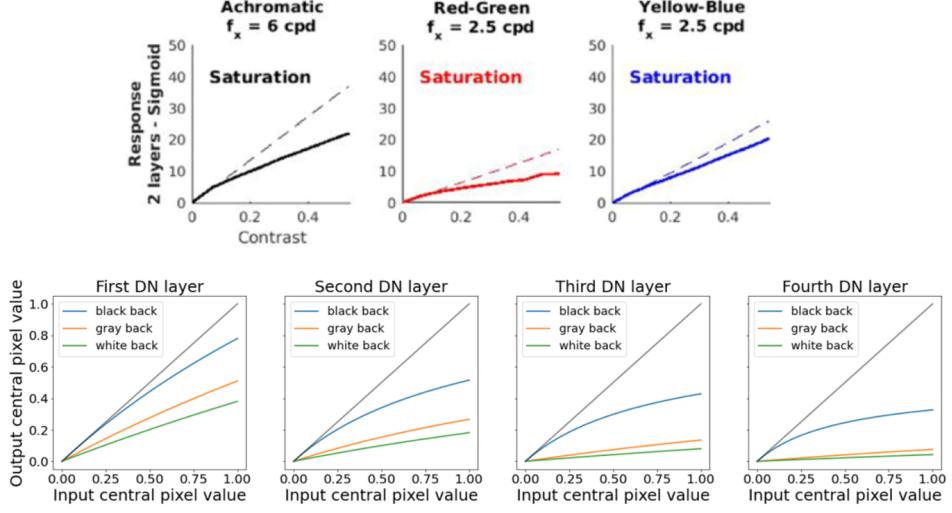


Figure 13: (Top) Human-like saturation behavior in contrast response happening in generic shallow autoencoders [21]. (Bottom) Human-like behavior obtained in image segmentation U-Nets when they are equipped with bio-inspired Divisive Normalization to improve their performance [39].

the response in the achromatic illumination condition could be *corrected* through an illumination dependent term in the denominator of the Divisive Normalization. Similarly, in the texture case, the fact that the no-mask condition (blue curve) is below the low-contrast curve suggests that *facilitation* in the Divisive Normalization is too high. These qualitative errors directly suggest quantitative changes in the architecture (if it is formulated in an explainable way), as pointed out in [34].

5.2 Changes in the optimization goal or training data to get human-like adaptation

The proposed stimuli and the associated human behaviors may also suggest changes in the optimization goal and on the necessary data to train the networks.

For instance, it is known that information maximization arguments lead to the emergence of Gabor-like receptive fields tuned to achromatic and opponent-chromatic directions [141, 130]. However, that sensible goal can be complemented with denoising-deblurring tasks so that center-surround cells and proper contrast sensitivity do emerge [41, 42, 43, 21]. Moreover, if the contrast nonlinearities do not emerge after all these linear stages, or they are not adaptive enough, this may be enforced by the segmentation goal in the encoder, as in [39], see Fig. 13 (bottom). In that case, the behavior in that segmentation network may not be completely human in part by lack of constraints in the Divisive Normalization (free kernels in [38, 39] as opposed to more sensible parametric kernels in [33, 34]), but part of its adaptive behavior may come from the selection of the training data that enforces contrast adaptation. Regarding the poor emergence of plausible receptive fields in the considered non-parametric models (see Table 2), this *error* makes sense in the context of the recently proposed *feature-spreading* problem [135]: if the goal is not demanding enough (as is usual in conventional goals) the features spread along all layers of the net in a way that the weak goal(s) is (are) fulfilled, but the layers remain biologically non-plausible.

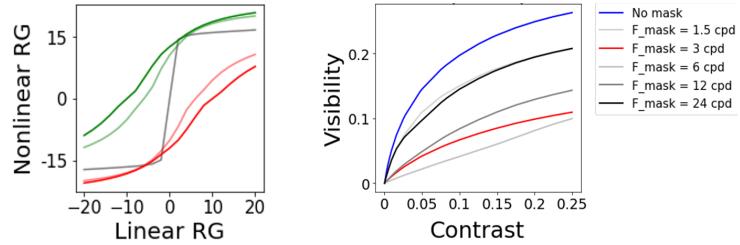


Figure 14: Examples of errors happening in the model [33, 34, 35]. *Left:* too sharp chroma nonlinearity in achromatic context -wrong gray curve- (taken from Fig. 10, subplot M). *Right:* too much / too low masking -light gray curves in wrong order- (taken from Fig. 12, subplot P). As this particular models is explainable, these could be solved by changing the values of the corresponding kernels of divisive normalization.

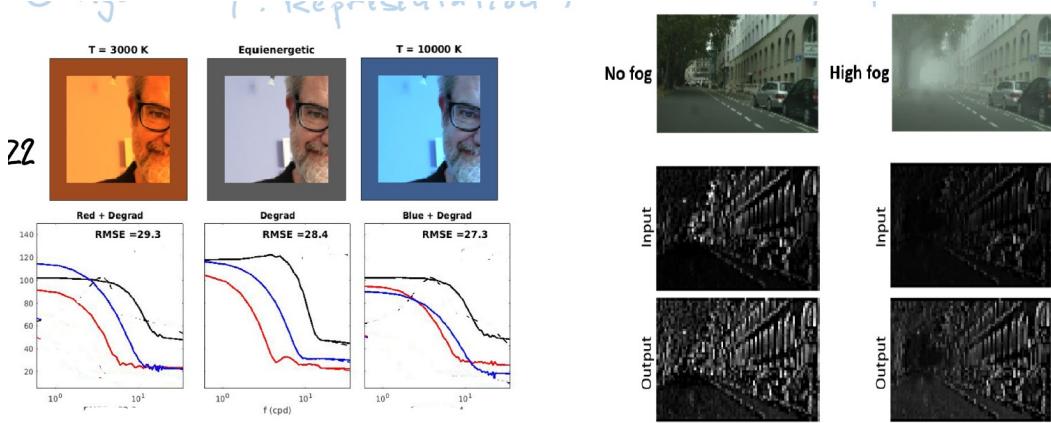


Figure 15: Left: Adaptation in the CSFs in autoencoders obtained from training on the proper (colorimetrically calibrated) environments for color adaptation [21]. Right: Improved contrast perception by using bio-inspired Divisive Normalization (with adaptive contrast responses such as the ones described in our proposal) in the model [38].

Regarding to suggestions on the training data, the behavior of the achromatic and chromatic CSFs proposed here was checked in [21] under data with well controlled illumination. The behavior found in the autoencoder CSFs in those cases resemble Von Kries adaptation, as was anticipated in [130]. Fig. 15 (left) shows that under low-temperature (reddish) illumination the red-tuned channel is relatively attenuated with regard to the blue-tuned channel, and the other way around under a high-temperature (blueish) illumination, as would happen using a Von Kries computation [113] or imposing the shifts in the response curves [118] shown in the Decalogue. Finally, Fig. 15 (right) shows that proper selection of training data (i.e. including images with high fog for segmentation to induce contrast adaptation) leads to the contrast enhancement results, as anticipated by the contrast-dependent nonlinearities shown in Fig. 13 (bottom).

5.3 Human-like curves imply better priors for natural image statistics

Two examples may illustrate how the nonlinear responses to Gabor stimuli shown in textured contexts as presented in the proposed Decalogue capture the statistics of natural images: the described nonlinear behaviors are a robust prior which may benefit whatever network intended to work in vision.

First, in Fig. 16 (top) we show that the energy of neighbor Gabor-like coefficients is correlated (bowtie conditional probabilities of Gabor coefficients in natural images), but the nonlinear responses in textured backgrounds make the resulting coefficients independent [103].

Second, non-Euclidean metrics based on the nonlinear responses to the stimuli presented here (e.g. metrics like those reported in [65, 134, 34, 35]) represent a robust prior of the PDF of natural images as illustrated by the fact shown in Fig. 16 (bottom): in autoencoders with access to very few samples, the use of this kind of perceptual metrics, make the reconstruction of images much more robust than those using (naive) Euclidean metrics because the perceptual metric is already capturing the statistics of natural images although samples are missing [23].

6 Final Remarks

In the first place, we have noted that there are many open problems when we evaluate the human nature of artificial networks: there is a non-trivial relationship between the environment, the task, and the architecture [16, 19, 22]. That complexity implies it is difficult to choose the layer(s) to measure from and the read-out mechanism to check the human nature of the model responses. These problems point out the need of new tests of human alignment that are independent of the training data and goal.

This motivates our proposal: a set of stimuli, *a Decalogue*, based on classical low-level vision science. The stimuli and associated facts (or human responses) describe the adaptive information bottleneck in the retina-V1 pathway. Some of the sensitivity surfaces are standardized [142, 99, 108, 110, 98, 111],

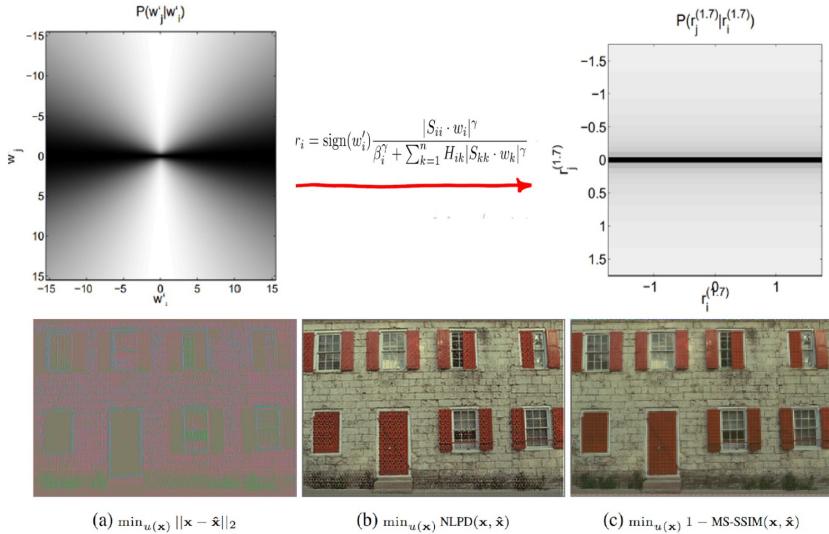


Figure 16: Top: Image PDF factorization from the contrast nonlinearities (div. norm.) illustrated in Figs. 7 and 8, as shown in [103]. Bottom: autoencoders trained with distortion metrics based on the contrast nonlinearities described in our proposal capture natural image statistics despite being trained with few samples [23].

or data is readily available [112] and allow quantitative comparison (namely properties 1, 3, 4, 6), as done in [117, 21], but the ones that involve showing tests in different illuminations or textured backgrounds (namely properties 2, 7-10) or those related to receptive fields (namely Gabor in 5) have only qualitative value.

The qualitative/quantitative nature of the proposed test is not a problem in practice as shown by its application to analyze and rank three illustrative models: (1) a parametric one based on physiology, classical psychophysics and Maximum Differentiation measurements [33, 34, 104, 35], (2) a non-parametric model, the *PerceptNet* [36], that includes trainable Divisive Normalization to reproduce human opinion on subjective image quality, and (3) a U-net with the same encoder as the *PerceptNet* but trained for image segmentation [38, 39]. The proposed test successfully ranks the models according to their different qualitative origin. It is important to stress that the use of this test can be easily extended to quantitative comparisons (as done in [21, 60, 4, 3]), although that is not shown in this presentation. Even the qualitative application of the test works to rank the human alignment and to point out that the two with less alignment have been trained for tasks that are not enough to fully explain the human behavior.

Finally, in the discussion, we have seen that the proposed test can be useful to modify the architecture of the networks, both in their linear and nonlinear parts. The test is useful to question the tasks or restrictions that are used in training (e.g. infoMax, noise, compression bottlenecks, classification, segmentation, etc.). It is also useful to question the data used in the training, either in their generality or balance. Moreover, we discussed how the use of human behaviors represented by the data in the proposed test, gives rise to priors related to the statistics of natural images.

In summary, we argue that the analysis of any kind of network, not only those that are specifically dedicated to modeling the human vision, but any devoted to vision, can benefit, in great measure, from seeing how they respond to the proposed test.

Acknowledgments and Disclosure of Funding

The *invited talk* at the *Artificial Intelligence Evaluation Workshop 2022* was funded by the University of Bristol. The computational work was partially funded by MCIN/AEI/FEDER/UE under Grants PID2020-118071GB-I00 and PID2023-152133NB-I00, by Spanish MIU under Grant FPU21/02256 and by Generalitat Valenciana under Projects GV/2021/074, CIPROM/2021/056, and by the grant BBVA Foundations of Science program: Maths, Stats, Comp. Sci. and AI (VIS4NN). Some computer resources were provided by Artemisa, funded by the EU ERDF through the Instituto de Física Corpuscular, IFIC (CSIC-UV). The audio-draft of this work (*the talk*) was recorded at El Saler beach (Valencia) and then transcription was done at the *Lisboa* restaurant (Valencia): its staff was particularly helpful at writing time during *Fallas 2025*.

References

- [1] Jeffrey S Bowers, Gaurav Malhotra, Marin Dujmović, Milton Llera Montero, Christian Tsvetkov, Valerio Biscione, Guillermo Puebla, Federico Adolfi, John E Hummel, Rachel F Heaton, et al. Deep problems with neural network models of human vision. *Behavioral and Brain Sciences*, 46:e385, 2023.
- [2] V. Biscione et al. MindSet: Vision. a toolbox for testing DNNs on key psychological experiments. *arXiv preprint arXiv:2404.05290*, 2024.
- [3] Yancheng Cai, Fei Yin, Dounia Hammou, and Rafal Mantiuk. Do computer vision foundation models learn the low-level characteristics of the human visual system? *CVPR ArXiV: 2502.20256*, 2025.
- [4] Dounia Hammou, Yancheng Cai, Pavan Madhusudanarao, Christos G. Bampis, and Rafał K. Mantiuk. Do image and video quality metrics model low-level human vision? *ArXiV: 2503.16264*, 2025.
- [5] J. Kubilius et al. Brain-like object recognition with high-performing shallow recurrent anns. *ICLR, Arxiv: 1909.06161*, 2019.
- [6] Johannes Mehrer, Courtney J. Spoerer, Emer C. Jones, Nikolaus Kriegeskorte, and Tim C. Kietzmann. An ecologically motivated image dataset for deep learning yields better models of human vision. *Proc. Nat. Acad. Sci.*, 118(8):e2011417118, 2021.
- [7] C. Zhuang, S. Yan, A. Nayebi, M. Schrimpf, MC. Frank, JJ. DiCarlo, and DLK. Yamins. Unsupervised neural network models of the ventral visual stream. *Proc. Nat. Acad. Sci.*, 118(3):e2014196118, 2021.
- [8] K.R. Storrs, T.C. Kietzmann, A. Walther, J. Mehrer, and N. Kriegeskorte. Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting. *Journal of Cognitive Neuroscience*, 33(10):2044–2064, 2021.
- [9] R. Rajalingham, E.B. Issa, P. Bashivan, K. Kar, K. Schmidt, and J.J. DiCarlo. Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, 38(33):7255–7269, 2018.
- [10] T. Macpherson, A. Churchland, T. Sejnowski, JJ. DiCarlo, Y. Kamitani, H. Takahashi, and T. Hikida. Natural and artificial intelligence: A brief introduction to the interplay between ai and neuroscience research. *Neural Networks*, 144:603–613, 2021.
- [11] SA. Cadena, GH. Denfield, EY. Walker, LA. Gatys, AS. Tolias, M. Bethge, and S. Ecker. Deep convolutional models improve predictions of macaque v1 responses to natural images. *PLoS Comput. Biol.*, 15(4):e1006897, 2019.
- [12] MF. Burg, SA. Cadena, Denfield GH., EY. Walker, AS. Tolias, M. Bethge, and S. Ecker. Learning divisive normalization in primary visual cortex. *PLoS Comput. Biol.*, 16(6):e1009028, 2021.
- [13] L. Paninsky. Personal communication at nyu laboratory for computational vision. 2001.
- [14] D. Marr and T. Poggio. From understanding computation to understanding neural circuitry. *Neurosci. Res. Prog. Bull.*, 15:470–488, 1977.
- [15] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Co., New York, 1978.
- [16] Tomaso Poggio. From marr’s vision to the problem of human intelligence. *MIT-CBMM Memos*, (118), 09/2021 2021.
- [17] Máté Lengyel. Marr’s three levels of analysis are useful as a framework for neuroscience. *The Journal of Physiology*, 602(9):1911–1914, 2024.

- [18] J.W. Pillow. Cross talk opposing view: Marr’s three levels of analysis are not useful as a framework for neuroscience. *The Journal of Physiology*, 602(9):1915–1917, 2024.
- [19] J. Malo and P. Hernández-Cámera. A separate theory-on-top level may be inspiring, but it is neither separate nor enough. *The Journal of Physiology*, 602(9):1918–1918, 2024.
- [20] A. Gomez-Villa, A. Martin, J. Vazquez, M. Bertalmío, and J. Malo. Color illusions also deceive CNNs for low-level vision tasks: Analysis and implications. *Vision Research*, 176:156–174, 2020.
- [21] Qiang Li, Alex Gomez-Villa, Marcelo Bertalmío, and Jesús Malo. Contrast sensitivity functions in autoencoders. *Journal of Vision*, 22(6), 2022.
- [22] Pablo Hernández-Cámera, Jorge Vila-Tomás, Valero Laparra, and Jesús Malo. Dissecting the effectiveness of deep features as metric of perceptual image quality. *Neural Networks*, 185:107189, 2025.
- [23] Alexander Hepburn, Valero Laparra, Raul Santos-Rodriguez, Johannes Ballé, and Jesus Malo. On the relation between statistical learning and perceptual distances. In *International Conference on Learning Representations*, 2022.
- [24] J. Malo and J. Gutiérrez. V1 non-linear properties emerge from local-to-global non-linear ICA. *Network: Computation in Neural Systems*, 17(1):85–102, 2006.
- [25] V. Laparra, S. Jiménez, G. Camps-Valls, and Jesús Malo. Nonlinearities and adaptation of color vision from Sequential Principal Curves Analysis. *Neural Comp.*, 24(10):2751–2788, 2012.
- [26] V. Laparra and J. Malo. Visual aftereffects and sensory nonlinearities from a single statistical framework. *Frontiers in Human Neuroscience*, 9:557, 2015.
- [27] H.B. Barlow. Sensory mechanisms, the reduction of redundancy, and intelligence. *Proc. of the Nat. Phys. Lab. Symposium on the Mechanization of Thought Process*, (10):535–539, 1959.
- [28] H.B. Barlow. Possible principles underlying the transformation of sensory messages. In WA Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA, 1961.
- [29] H.B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253, 2001.
- [30] J. Malo, J. Gutiérrez, and J. Rovira. Perturbation analysis of the changes in V1 receptive fields due to context. *Gordon Research Conference: Sensory Coding and the Natural Environment*, 2004.
- [31] H. Barlow. Personal communication at the GRC sens. coding nat. env. 2004.
- [32] Jesús Malo and Eero P. Simoncelli. Geometrical and statistical properties of vision models obtained via maximum differentiation. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Huib de Ridder, editors, *Human Vision and Electronic Imaging XX*, volume 9394 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 93940L, March 2015.
- [33] M. Martinez, P. Cyriac, T. Batard, M. Bertalmío, and J. Malo. Derivatives and inverse of cascaded linear+nonlinear neural models. *PLOS ONE*, 13(10):1–49, 10 2018.
- [34] M. Martinez, M. Bertalmío, and J. Malo. In praise of artifice reloaded: Caution with natural image databases in modeling vision. *Front. Neurosci. doi: 10.3389/fnins.2019.00008*, 2019.
- [35] J. Malo, JJ. Esteve-Taboada, and M. Bertalmío. Cortical divisive normalization from wilson-cowan neural dynamics. *J. Nonlinear Sci.*, 34(2):35, 2024.
- [36] A. Hepburn, V. Laparra, J. Malo, R. McConville, and R. Santos-Rodriguez. Perceptnet: A human visual system inspired neural network for estimating perceptual distance. In *IEEE ICIP*, pages 121–125, 2020.

- [37] J. Ballé, V. Laparra, and EP. Simoncelli. End-to-end optimized image compression. *ICLR ArXiV:1611.01704*, 2017.
- [38] P. Hernández-Cámara, J. Vila-Tomás, V. Laparra, and J. Malo. Neural networks with divisive normalization for image segmentation. *Patt. Recogn. Lett.*, 173:64–71, 2023.
- [39] P. Hernández-Cámara, J. Vila-Tomás, P. Dauden-Oliver, N. Alabau-Bosque, V. Laparra, and J. Malo. Why divisive normalization works in image segmentation? *Neurocomputing*, 649, 2025.
- [40] Antonio Torralba, Phillip Isola, and William T Freeman. *Foundations of Computer Vision*. MIT Press, 2024.
- [41] J. J. Atick, Z. Li, and A. N. Redlich. Understanding retinal color coding from first principles. *Neural Computation*, 4(4):559–572, 1992.
- [42] Y. Karklin and E. Simoncelli. Efficient coding of natural images with a population of noisy linear-nonlinear neurons. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.
- [43] J. Lindsey, SA. Ocko, S. Ganguli, and S. Deny. The effects of neural resource constraints on early visual representations. *Int. Conf. Learn. Repr. ICLR*, 2019.
- [44] Li Zhaoping. *Understanding Vision: Theory, Models, and Data*. Oxford University Press, 05 2014.
- [45] Gregory K. Wallace. The JPEG still picture compression standard. *Commun. ACM*, 34(4):30–44, April 1991.
- [46] J. Malo, A.M. Pons, and J.M. Artigas. Bit allocation algorithm for codebook design in vector quantization fully based on human visual system non-linearities for suprathreshold contrasts. *Electronics Letters*, 31(15):1229–1231, 1995.
- [47] J. Malo, F. Ferri, J. Albert, J.Soret, and J.M. Artigas. The role of perceptual contrast non-linearities in image transform coding. *Image & Vision Computing*, 18(3):233–246, 2000.
- [48] David Taubman and Michael Marcellin. *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Springer Publishing Company, Incorporated, 2013.
- [49] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli. Non-linear image representation for efficient perceptual coding. *IEEE Transactions on Image Processing*, 15(1):68–80, 2006.
- [50] Didier J. Le Gall. The MPEG video compression algorithm. *Signal Processing: Image Communication*, 4(2):129–140, 1992.
- [51] J. Malo, F. Ferri, J. Gutierrez, and I. Epifanio. Importance of quantizer design compared to optimal multigrid motion estimation in video coding. *Electronics Letters*, 36(9):507–509, 2000.
- [52] J. Malo, J. Gutierrez, I. Epifanio, and F. Ferri. Perceptually weighted optical flow for motion-based segmentation in MPEG-4 paradigm. *Electronics Letters*, 36(20):1693–1694, 2000.
- [53] J.Malo, J.Gutierrez, I.Epifanio, F.Ferri, and J.M.Artigas. Perceptual feed-back in multigrid motion estimation using an improved DCT quantization. *IEEE Transactions on Image Processing*, 10(10):1411–1427, 2001.
- [54] MA. Goodale, AD.; Milner, LS. Jakobson, and DP. Carey. A neurological dissociation between perceiving objects and grasping them. *Nature*, 349(6305):154–156, 1991.
- [55] A. D. Milner and M. A. Goodale. Separate visual pathways for perception and action. *Trends Neurosci.*, 15:20–25, 1992.
- [56] NK. Logothetis and DL. Sheinberg. Visual object recognition. *Ann. Rev. Neurosci.*, 19:577–621, 1996.

- [57] G. Kreiman, C. Koch, and I. Fried. Category specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.*, 3(9):946–953, 2000.
- [58] Ruben Coen-Cagli and Odelia Schwartz. The impact on midlevel vision of statistically optimal divisive normalization in v1. *Journal of Vision*, 13(8):13–13, 07 2013.
- [59] Michelle Miller, SueYeon Chung, and Kenneth D. Miller. Divisive feature normalization improves image recognition performance in alexnet. In *Int. Conf. Learn. Repres. ICLR*, 2022.
- [60] A. Akbarinia, Y. Morgenstern, and K.R. Gegenfurtner. Contrast sensitivity function in deep networks. *Neural Networks*, 164:228–244, 2023.
- [61] P Hernández-Cámara, A Gomez-Villa, JoseManuel Jaén-Lorites, J Vila-Tomás, J Malo, and V Laparra. Contrast sensitivity function of multimodal vision-language models. In *8th Cognitive Computational Neuroscience Conference*, 2025.
- [62] P.C. Teo and D.J. Heeger. Perceptual image distortion. *Proceedings of the SPIE*, 2179:127–141, 1994.
- [63] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.
- [64] Norma V. S. Graham. *Visual pattern analyzers*. Visual pattern analyzers. Oxford University Press, New York, NY, US, 1989. Pages: xvi, 646.
- [65] V. Laparra, J. Muñoz Marí, and J. Malo. Divisive normalization image quality metric revisited. *JOSA A*, 27(4):852–864, 2010.
- [66] Z Wang, A C Bovik, H R Sheikh, and E P Simoncelli. Perceptual image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Processing*, 13(4):600–612, 2004.
- [67] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022.
- [68] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.
- [69] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.
- [70] J. Malo. Spatio-chromatic information available from different neural layers via gaussianization. *The Journal of Mathematical Neuroscience*, 10(18), 2020.
- [71] J. Malo, B. Kheravdar, and Q. Li. Visual information fidelity with better vision models and better mutual information estimates. *Journal of Vision*, 21(9):2351, 2021.
- [72] Qiang Li, Greg Ver Steeg, and Jesus Malo. Functional connectivity via total correlation: Analytical results in visual areas. *Neurocomputing*, 571:127143, 2024.
- [73] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.
- [74] V. Laparra, JE. Johnson, G. Camps-Valls, R. Santos-Rodriguez, and J. Malo. Estimating Information Theoretic Measures via Multidimensional Gaussianization. *IEEE Trans. Patt. Anal. & Mach. Intell.*, 47(02):1293–1308, 2025.
- [75] J. Malo, JJ. Esteve-Taboada, G. Aguilar, M. Maertens, and FA. Wichmann. Estimating the contribution of early and late noise in vision from psychophysical data. *J. Vision*, 25(1):12–12, 2025.

- [76] A. Mahendran and A. Vedaldi. Visualizing deep convolutional neural networks using natural pre-images. *Int. J. Comput. Vis.*, 120:233–255, 2016.
- [77] W. Luo, Y. Li, R. Urtasun, and R. Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.
- [78] David H Hubel, Torsten N Wiesel, et al. Receptive fields of single neurones in the cat’s striate cortex. *J physiol*, 148(3):574–591, 1959.
- [79] David H Hubel and Torsten N Wiesel. Integrative action in the cat’s lateral geniculate body. *The Journal of physiology*, 155(2):385, 1961.
- [80] D.L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.*, 88(1):455–463, 2002.
- [81] C. Tailby, SG. Solomon, NT. Dhruv, and P. Lennie. Habituation reveals fundamental chromatic mechanisms in striate cortex of macaque. *J. Neurosci.*, 28(5):1131–1139, 2008.
- [82] D. Ringach and R. Shapley. Reverse correlation in neurophysiology. *Cognit. Sci.*, 28(2):147–166, 2004. Rendering the Use of Visual Information from Spiking Neurons to Recognition.
- [83] MP. Eckstein and AJ. Ahumada. Classification images: A tool to analyze visual strategies. *J. Vision*, 2(1), 2002.
- [84] G Wyszecki and WS. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, New Jersey, 2000.
- [85] X. Otazu, CA. Parraga, and M. Vanrell. Toward a unified chromatic induction model. *J. Vision*, 10(12):5–5, 10 2010.
- [86] C. Ware and WB. Cowan. Changes in perceived color due to chromatic interactions. *Vision Research*, 22(11):1353–1362, 1982.
- [87] A. Gomez-Villa, A. Martín, J. Vazquez-Corral, M. Bertalmío, and J. Malo. Color illusions also deceive CNNs for low-level vision tasks: Analysis and implications. *Vision Research*, 176:156–174, November 2020.
- [88] A. Gomez-Villa, K. Wang, CA. Parraga, B. Twardowski, J. Malo, J. Vazquez-Corral, and J. van de Weijer. The art of deception: Color visual illusions and diffusion models. *IEEE Comp. Vis. Patt. Recogn. (CVPR)*, 2025.
- [89] J. Malo and J. Bowers. The low-level mindset: compelling low-level visual psychophysics to evaluate image computable vision models. Invited talk, Psychol. Dept. University of Bristol, 2024.
- [90] NC. Rust and JA. Movshon. In praise of artifice. *Nature Neurosci.*, 8(12):1647–1650, 2005.
- [91] HH. Schütt and FA. Wichmann. An image-computable psychophysical spatial vision model. *J. Vision*, 17(12):12–12, 10 2017.
- [92] M. Bertalmío, A. Gomez-Villa, A. Martín, J. Vazquez, D. Kane, and J. Malo. Evidence for the intrinsically nonlinear nature of receptive fields in vision. *Scientific Reports*, 10:16277, 2020.
- [93] M. Bertalmío, A. Durán-Vizcaíno, J. Malo, and FA. Wichmann. Plaid masking explained with input-dependent dendritic nonlinearities. *Sci. Rep.*, 14:24856, 2024.
- [94] T. Carney, SA. Klein, CW. Tyler, AD. Silverstein, B. Beutter, D. Levi, AB. Watson, AJ. Reeves, AM. Norcia, C. Chen, W. Makous, and MP. Eckstein. Development of an image/threshold database for designing and testing human vision models. In *Human Vision and Electronic Imaging IV*, volume 3644, pages 542 – 551. International Society for Optics and Photonics, SPIE, 1999.

- [95] Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv preprint*, 2018.
- [96] S. Daly. Visible differences predictor: An algorithm for the assessment of image fidelity. In A.B. Watson, editor, *Digital Images and Human Vision*, pages 179–206, Massachusetts, 1993. MIT Press.
- [97] A.B. Watson et al. *Digital Images and Human Vision*. MIT Press, Massachusetts, 1993.
- [98] J. Malo, A.M. Pons, and J.M. Artigas. Subjective image fidelity metric based on bit allocation of the human visual system in the dct domain. *Image and Vision Computing*, 15(7):535–548, 1997.
- [99] A. B. Watson and J. Malo. Video quality measures based on the standard spatial observer. In *IEEE Proc. Int. Conf. Im. Proc.*, volume 3, pages III–III, 2002.
- [100] P. Hernández-Cámarra, P. Daudén-Oliver, V. Laparra, and J. Malo. Alignment of color discrimination in humans and image segmentation networks. *Front. Psychol.*, Volume 15 - 2024, 2024.
- [101] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 281:607–609, 1996.
- [102] O. Schwartz and E.P. Simoncelli. Natural signal statistics and sensory gain control. *Nat. Neurosci.*, 4(8):819–825, 2001.
- [103] J. Malo and V. Laparra. Psychophysically tuned divisive normalization approximately factorizes the pdf of natural images. *Neural computation*, 22(12):3179–3206, 2010.
- [104] Alexander Gomez-Villa, Marcelo Bertalmío, and Jesus Malo. Visual information flow in Wilson–Cowan networks. *Journal of Neurophysiology*, 123(6):2249–2268, 2020.
- [105] J. Malo. Information flow in biological networks for color vision. *Entropy*, 24:1442, 2022.
- [106] Leo M. Hurvich and Dorothea Jameson. An opponent-process theory of color vision. *Psychological Review*, 64, Part 1 6:384–404, 1957.
- [107] F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.
- [108] K. T. Mullen. The CSF of human colour vision to red-green and yellow-blue chromatic gratings. *J. Physiol.*, 359:381–400, 1985.
- [109] MA. Georgeson and GD. Sullivan. Contrast constancy: deblurring in human vision by spatial frequency channels. *J. Physiol.*, 252(3):627–656, 1975.
- [110] S. Daly. Application of a noise-adaptive Contrast Sensitivity Function to image data compression. *Optical Engineering*, 29(8):977–987, 1990.
- [111] D. H. Kelly. Motion and vision. ii. stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.*, 69(10):1340–1349, Oct 1979.
- [112] MA. Díez-Ajenjo, P. Capilla, and MJ. Luque. Red-green vs. blue-yellow spatio-temporal contrast sensitivity across the visual field. *J. Mod. Opt.*, 58(19-20):1736–1748, 2011.
- [113] M.D. Fairchild. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2013.
- [114] Paul Whittle. Brightness, discriminability and the “crispening effect”. *Vision Research*, 32(8):1493–1507, 1992.
- [115] S. Laughlin. A simple coding procedure enhances a neuron’s information capacity. *Zeitschrift Für Naturforschung C*, 36(9):910–912, 1981.

- [116] J. Malo and M.J. Luque. ColorLab: A Matlab Toolbox for Color Science and Calibrated Color Image Processing. *Univ. Valencia.* https://isp.uv.es/code/vision_and_color/colorlab/content/, 2002.
- [117] J. Vila-Tomás, P. Hernández-Cámara, and J. Malo. Artificial psychophysics questions classical hue cancellation experiments. *Frontiers in Neuroscience*, 17, 2023.
- [118] J. Krauskopf and K. Gegenfurtner. Color discrimination and adaptation. *Vision Research*, 32(11):2165–2175, 1992.
- [119] Javier Romero, José A. García, Luis Jiménez del Barco, and E. Hita. Evaluation of color-discrimination ellipsoids in two-color spaces. *J. Opt. Soc. Am. A*, 10(5):827–837, May 1993.
- [120] G.E Legge and J.M. Foley. Contrast masking in human vision. *Journal of the Optical Society of America*, 70:1458–1471, 1980.
- [121] G.E Legge. A power law for contrast discrimination. *Vision Research*, 18:68–91, 1981.
- [122] John M. Foley. Human luminance pattern-vision mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A*, 11(6):1710–1719, Jun 1994.
- [123] Andrew B. Watson and Joshua A. Solomon. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14(9):2379–2391, Sep 1997.
- [124] J. Malo and J. Gutierrez. VistaLab: The Matlab toolbox for linear spatio-temporal Vision Models. *Univ. Valencia.* https://isp.uv.es/code/vision_and_color/colorlab/vistalab/, 2002.
- [125] John Ross, Harriet D. Speed, and Fergus William Campbell. Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 246(1315):61–70, 1991.
- [126] Robert Shapley and Michael J. Hawken. Color in the cortex: single- and double-opponent cells. *Vision Research*, 51(7):701–717, 2011. Vision Research 50th Anniversary Issue: Part 1.
- [127] C. Blakemore and F. Campbell. On the existence of neurons selectivity sensitive to the orientation and size of retinal images. *J. Physiol.*, 203:237–260, 1969.
- [128] M. Martinez, LM. Martinez, and J. Malo. Topographic independent component analysis reveals random scrambling of orientation in visual space. *PLoS ONE*, 12(6):e0178345, 2017.
- [129] P. N. Loxley. The two-dimensional gabor function adapted to natural image statistics: A model of simple-cell receptive fields and sparse structure in images. *Neural Computation*, 29(10):2769–2799, 2017.
- [130] M. U. Gutmann, V. Laparra, A. Hyvärinen, and J. Malo. Spatio-chromatic adaptation via higher-order canonical correlation analysis of natural images. *PloS ONE*, 9(2):e86481, 2014.
- [131] M. Carandini and D. Heeger. Summation and division by neurons in visual cortex. *Science*, 264(5163):1333–6, 1994.
- [132] Matteo Carandini and David J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, January 2012. Number: 1 Publisher: Nature Publishing Group.
- [133] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [134] Valero Laparra, Alexander Berardino, Johannes Ballé, and Eero P. Simoncelli. Perceptually optimized image rendering. *J. Opt. Soc. Am. A*, 34(9):1511–1525, Sep 2017.
- [135] Jorge Vila-Tomás, Pablo Hernández-Cámara, Valero Laparra, and Jesús Malo. Parametric perceptnet: A bio-inspired deep-net trained for image quality assessment. *ArXiV*, page 2412.03210, 2025.

- [136] E.P. Simoncelli and E.H. Adelson. *Subband Image Coding*, chapter Subband Transforms, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.
- [137] D. Cai, GC. DeAngelis, and RD. Freeman. Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *J. Neurophysiol.*, 78(2):1045–1061, 1997.
- [138] A.B. Watson. Detection and recognition of simple spatial forms. In O.J. Braddick and A.C. Sleigh, editors, *Physical and Biological Processing of Images*, volume 11 of *Springer Series on Information Sciences*, pages 100–114, Berlin, 1983. Springer Verlag.
- [139] E. Martinez-Uriegas. Color detection and color contrast discrimination thresholds. In *Proc. OSA Meeting*, page 81, 1997.
- [140] J. Gutiérrez, F. Ferri, and J. Malo. Regularization operators for natural images based on nonlinear perception models. *IEEE Tr. Im. Proc.*, 15(1):189–200, 2006.
- [141] A. Hyvärinen, J. Hurri, and PO. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Springer, 2009.
- [142] A.B. Watson and C.V. Ramirez. A Standard Observer for Spatial Vision. *Investig. Opht. and Vis. Sci.*, 41(4):S713, 2000.