

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 2022 work^{2 3} we argued that, despite recent claims about successful modeling of the visual brain using artificial nets, the problem is far from being solved (even for low-level vision). First we introduce open issues such as *where should we read from in ANNs in order to reproduce human behavior?*, *what should be the read-out mechanism?*, *this ad-hoc read-out is considered part of the brain model or not?*, in order to understand the behavior of ANNs, *should we use artificial psychophysics or artificial physiology?*, in the case of ANNs, *artificial experiments should literally match the experiments done with humans?*. This means that there is a clear need of rigorous procedures for experimental tests for ANNs devoted to model the visual brain, and more generally, to understand ANNs devoted to generic vision tasks.

Following our experience in using low-level facts from *Quantitative Visual Neuroscience* in image processing and computer vision, in this work we presented the idea of developing a low-level dataset compiling the basic spatio-temporal and chromatic facts that are known to happen in the retina-V1 pathway, and they are not currently available in existing databases such as BrainScore.

In our illustrative results we checked the behavior of three recently proposed models with similar 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 called *PerceptNet* tuned to maximize the correlation with human opinion about subjective distance between images [Hepburn et al. IEEE ICIP 19], and **(3)** A model with the same encoder as PerceptNet, but tuned for image segmentation (later published as Hernandez-Camara et al. Patt.Recogn.Lett. 23). Results on 10 compelling psycho/physio visual facts show that the first model (the parametric one) is the one with closer behavior to the human observers in terms of receptive fields (obviously because they were parametrically imposed), but more interestingly, on the nonlinear behavior when facing complex spatio-chromatic patterns of a range of luminances and contrasts.

*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.

³Original presentation: http://isp.uv.es/docs/talk_AI_Bristol_Malo_et_al_2022.pdf

1 Introduction

1.1 Deep Neural Networks are the Best Models of the Human Visual System

Deep neural networks (DNNs) currently offer some of the most advanced models for understanding visual information processing in the primate brain [Jacob et al. Nat. Commun. 21, Nikolaus. Annu. Rev. Vis. Sci. 15, Li et al. J. Vis. 22]. Recent research has highlighted the potential for deep learning to inform and refine our theories of brain function, with many findings suggesting that the principles behind DNNs are closely aligned with the way the brain processes visual information [Richards et al. Nat. Neurosci. 19]. The functional similarities and differences between DNNs and the visual brain have been widely explored, making this a dynamic and rapidly growing area of research within both neuroscience and computer vision.

1.2 Using DNNs to Understand the Visual Brain May Not Be That Easy

DNN, while powerful, are often poorly explainable, presenting challenges when applying them to understanding the visual brain. Issues such as denoising, which functions similarly to the Lateral Geniculate Nucleus (LGN), and compression, which mimics the information preprocessing from the LGN to the visual cortex. But tasks like segmentation (which are difficult to explain from a visual science perspective), classification (which are also challenging to interpret), and other networks further complicate the process. Is there a unified deep network that can explain all visual functions? These complexities make it difficult to directly map neural network operations onto brain functions in a meaningful way.

Another important consideration is whether to focus on reproducing artificial physiology or artificial psychophysics. This distinction involves balancing a literal reproduction of brain processes with an idealized, abstract model of human behavior. For example, should we focus on mimicking response summation and read-out in a way that aligns with human vision, or should we prioritize matching the results of psychophysical experiments? These questions reflect a broader challenge of matching DNN models to human behavior in a way that is both faithful and effective, going beyond the scope of high-level metrics like BrainScore (<https://www.brain-score.org/>).

Furthermore, when applying DNNs to model the visual brain, one key challenge is determining where to read features from within the network and which layers to focus on for encoding or decoding information. Should we use just one feature, or should we consider all features across different layers? This decision is akin to the physiology-psychophysics problem,

These questions ultimately highlight the significant issues related to model plausibility. Understanding how closely DNN models can replicate the complexity of human vision requires addressing many nontrivial challenges. To evaluate whether our models are truly capturing human-like visual processing, we might need to consider a more sophisticated "Turing Test" for visual models, ensuring that the behavior of the artificial networks aligns not just with low-level brain activity but also with high-level cognitive functions.

1.3 Our Proposal: An Easy-To-Use Turing-Test from Vision Science

To test the capabilities of DNNs in mimicking human visual functions, we propose an approach inspired by key findings from vision science. By digesting the literature and selecting compelling facts related to stimuli and their corresponding responses, we aim to evaluate how well DNNs align with known physiological and psychophysical principles.

In our proposal, we suggest two approaches for testing DNNs: one linear and one nonlinear, as shown in the table⁴ in Fig.1. For color processing, we first test DNNs from a linear perspective, focusing on spectral sensitivities, and then from a nonlinear perspective to examine adaptation, nonlinear responses, and color illusions. Similarly, for texture processing, we begin by testing the linear aspects, such as receptive fields and contrast sensitivities, and then explore nonlinear responses, including masking effects and contrast illusions. This dual approach provides a comprehensive framework for evaluating how DNNs replicate the complexity of visual processing observed in the human brain.

⁴The graphical quality of tables, figures and diagrams in this preprint will be updated in the journal version.

	FACTS	MODALITY	LINEAR
1	Spectrum \rightarrow Opponent spectral sensitivity	COLOR	L
2	Brightness & Color Response Saturation	COLOR	NL
3	(Linear) Spatio-Chromatic Receptive Fields	TEXTURE	L
4	Achromatic Contrast Sensitivity (Bandwidth)	TEXTURE	L
5	Chromatic Contrast Sensitivity (Bandwidth)	TEXTURE	L
6	Nonlinear Contrast response: Saturation	TEXTURE	NL
7	: Frequency order	TEXTURE	NL
8	Context effects: Energy	TEXTURE	NL
9	: Frequency	TEXTURE	NL
10	: Orientation	TEXTURE	NL

Figure 1: Turing-test checklist. This table presents a checklist of tests for evaluating DNNs, examining both linear and nonlinear methods to assess how well they mimic human visual processing.

2 Methods

Three biological neural networks are evaluated here: BioMultiLayer [Martinez et al. PLOS 18], Perceptnet [Hepburn et al. IEEE ICIP 19], and Bio U-Net [Hernandez et al. Patt. Recogn. lett. 23] models.

BioMultiLayer: A cascade of isomorphic L+NL modules based on canonical Divisive Normalization [Carandini et al. Nat. Rev. Neurosci. 11]. Which optimized for modeling human low-level visual pathway. The model processes the spatial distribution of spectral irradiance at the retina. It consists of several layers that combine linear and nonlinear transformations to simulate human visual processing. In the first layer, the input is processed by three LMS spectral sensitivities and a linear recombination, producing three tristimulus values representing luminance and opponent chromatic channels. These values undergo adaptive saturation and a Weber-like nonlinearity to enhance low luminance regions. The second layer calculates the deviation of brightness from the local average and normalizes it nonlinearly to compute local contrast. In the third layer, local contrast responses are filtered by center-surround receptive fields and normalized again, increasing responses in low-contrast regions. Finally, the fourth layer applies a wavelet transform and normalizes responses based on surround activity, again boosting responses in low-input regions.

Perceptnet: Each stage of the PerceptNet model corresponds to a process in the human visual system. It begins with gamma correction, followed by conversion to an opponent color space, and then a Von Kries transform. Next, the model applies center-surround filters and LGN normalization, which mimics the early processing stages of visual information. The network then moves to orientation-sensitive and multiscale processing in V1, and concludes with divisive normalization in V1, reflecting higher-level visual processing.

Bio U-Net: A deep network, such as U-Nets, incorporates canonical computations like divisive normalization [Carandini et al. Nat. Rev. Neurosci. 11], which accounts for adaptation in biological neurons, and achieves strong performance on segmentation tasks.

We evaluated three models using achromatic and chromatic brightness and masking perception. In human vision, for brightness perception, the threshold increases, and the response becomes saturated. For fixed luminance, brightness decreases with the background, while for fixed brightness, luminance increases with the background, as shown in Fig.2A. For masking perception, as shown on the left side of Fig.2B, from top to bottom, the pattern with increasing contrast C is overlaid on a noisy

background (mask) of similar frequency, with an increase in contrast C_m . Meanwhile, on the right side of Fig.2B, a mask with a more similar orientation θ has a more pronounced effect.

Building on these functional results from vision science, we can use DNNs to reproduce these responses and compare their performance to that of the human visual cortex. This approach helps us gain a deeper understanding of visual processing and guides the design of artificial models that more closely mimic human vision.

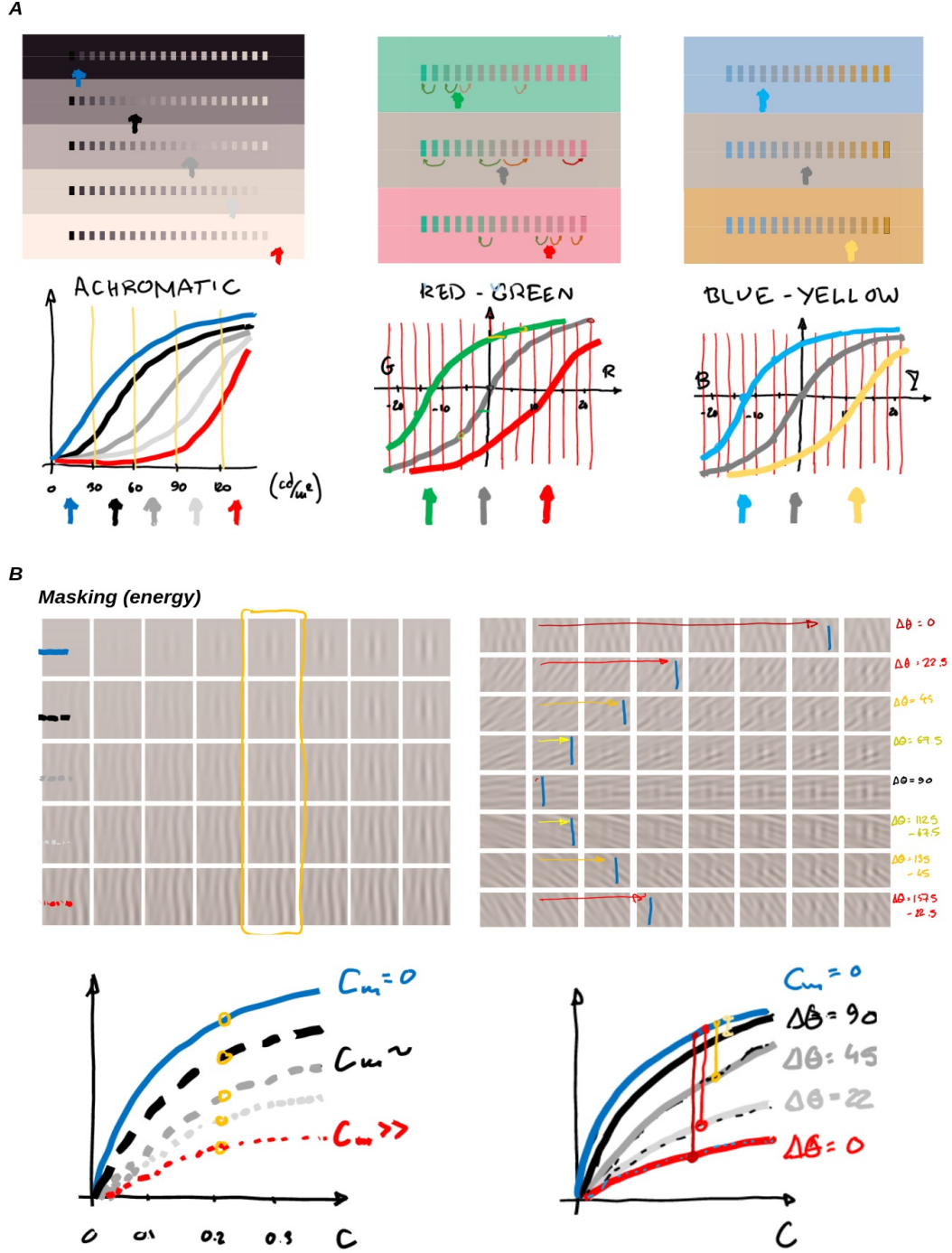


Figure 2: Human brightness and masking perception: Panel A presents achromatic (white-black) and chromatic (red-green, yellow-blue) brightness perception, while Panel B illustrates human visual perception for masking, including both energy and orientation masking.

3 Results

The vision "Turing-Test" was conducted on three biologically-oriented models, with their performance shown in Fig.3. Overall, the BioMultiLayer model outperformed the others in each test, followed by Perceptnet, which performed better than Bio U-Net. BioMultiLayer's superior performance can be attributed to its design, which closely mimics human visual information processing at every stage, reflecting the functional of the human visual system. Perceptnet also incorporates key elements of visual processing, such as opponent color space, center-surround filters, and divisive normalization, which contributed to its strong performance in the test. In contrast, Bio U-Net only includes divisive normalization, which explains why its performance was less competitive in our Turing test.

	BioMultiLayer PLoS 18 Front. Neurosci. 19	Perceptnet IEEE ICIP 20	Bio U-Net arXiv 22
FACTS			
1 Spectrum → Opponent spectral sensitivity	✓	~	✗
2 Brightness & Color Response Saturation	~	~	~
3 (Linear) Spatio-Chromatic Receptive Fields	✓ ✓	~ ✗	✗ ✗
4 Achromatic Contrast Sensitivity (Bandwidth)	✓	✓	✗
5 Chromatic Contrast Sensitivity (Bandwidth)	✓	✗	✗
6 Nonlinear Contrast response: Saturation	✓	✓	✓
7 Frequency order	✓	✗	✓
8 Context effects: Energy	✓	~	✓
9 Frequency	~	✗	✗
10 Orientation	~	✗	✗

Figure 3: Turing test evaluation checklist for BioMultiLayer, Perceptnet, and Bio U-Net models. This table provides a comprehensive set of tests to evaluate the performance of these three deep neural networks, examining both linear and nonlinear methods to assess how effectively they replicate human visual processing.

Key insights from the Vision "Turing-Test" that we can learn are: Understanding how architectural and functional adjustments, such as incorporating linear and nonlinear components like divisive normalization, affect performance is crucial. These changes help bridge the gap between artificial models and biological vision systems. Additionally, modifying tasks and representations used during training can significantly improve our ability to explain and interpret behavioral outcomes. Careful attention to training methods and the choice of datasets is essential to ensure models exhibit reliable and sensible behavior, as the data used strongly influences how well a model generalizes and reflects true human vision. Incorporating robust priors from biological vision systems is another important step, as these priors help guide the model's responses and enhance its ability to handle variations in real-world visual input. In summary, through the vision "Turing-Test", we can deepen our understanding of vision systems and create artificial models that more closely mimic human visual processing.

4 Conclusions

When evaluating the human-like qualities of artificial networks, it is crucial to recognize the nontrivial interplay between data, task, and architecture. This complexity calls for the development of new testing methods that are independent of the training process itself. To address this, we propose a Vision-Science-based low-level checklist that allows for both qualitative and quantitative evaluation of artificial networks in terms of their alignment with human perception. This checklist has proven effective in discriminating between three bio-inspired models, offering a tangible method to assess how closely an artificial network's behavior resembles that of the human visual system.

The checklist not only serves as a tool for comparison but also opens up avenues for further refinement of artificial network architectures. It can help guide modifications to both linear and nonlinear

structures, improving their ability to mimic human processing. Additionally, the checklist encourages a critical reassessment of the tasks and constraints applied during network training. By questioning frameworks like infomax, noise management, and the bottleneck effect, it becomes possible to rethink how tasks are formulated to ensure more human-like results.

Furthermore, the checklist emphasizes the importance of training methodologies. It highlights the need for better, more representative data sets to avoid biases and ensure that the network's performance is not limited by skewed inputs. In doing so, it suggests that refining the priors used in training can significantly enhance the alignment of artificial networks with human-like behavior. Ultimately, this approach offers a more comprehensive path forward, enabling improvements in both the architecture and training of artificial networks to bring them closer to human-like performance.

Acknowledgments and Disclosure of Funding

Authors thank Prof. Raul Santos for his invitation to participate in the Workshop on Evaluating Artificial Intelligence (Univ. of Bristol 2022). This work was partially funded by MICIIN projects obtained 2020 and 2021 by J. Malo and V. Laparra.

The compilation of this conference paper for arxiv was first written by J. Malo at Pinedo Beach in "El Velero" restaurant at Valencia, Spain at jan. 2025. The food is so so, but it is a great place to write. J. Malo thanks the staff of El Velero for their attitude and their electric plug. The first version was later organized by Q. Li towards the journal version that is on the way.

References

- [Malo & Simoncelli SPIE 15] Geometrical and statistical properties of vision models obtained via maximum differentiation. Published in Proc SPIE Conf on Human Vision and Electronic Imaging (HVEI XX), vol.9394 Feb 2015.
- [Martinez et al. PLOS 18] M Martinez-Garcia, P Cyriac, T Batard, M Bertalmío, J Malo (2018). Derivatives and inverse of cascaded linear+nonlinear neural models. PLoS ONE 13(10): e0201326.
- [Martinez et al. Front. Neurosci. 19] M Martinez, M Bertalmío J Malo (2019). In Praise of Artifice Reloaded: Caution With Natural Image Databases in Modeling Vision. Front. Neurosci., 18 February, Volume 13.
- [Hepburn et al. IEEE ICIP 19] A Hepburn, V Laparra, J Malo, R McConville, R, Santos (2019). PerceptNet: A Human Visual System Inspired Neural Network for Estimating Perceptual Distance. 2020 IEEE International Conference on Image Processing.
- [Hernandez et al. Patt. Recogn. lett. 23] P Hernández-Cámara, J Vila-Tomás, V Laparra, J Malo (2023). Neural networks with divisive normalization for image segmentation. Pattern Recognition Letters 173, 64-71.
- [Jacob et al. Nat. Commun. 21] G Jacob, R. T. Pramod, H Katti, S. P. Arun (2021). Qualitative similarities and differences in visual object representations between brains and deep networks. Nature Communications. 12.
- [Nikolaus. Annu. Rev. Vis. Sci. 15] K Nikolaus (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annual Review of Vision Science. 1. 417-446.
- [Richards et al. Nat. Neurosci. 19] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. d. Berker, S. Ganguli, C. J. Gillon, D. Hafner, A. Kepecs, N. Kriegeskorte, P. Latham, G. W. Lindsay, K. D. Miller, R. Naud, C. C. Pack, P. Poirazi, P. Roelfsema, J. Sacramento, A. Saxe, B. Scellier, A. C. Schapiro, W. Senn, G. Wayne, D. Yamins, F. Zenke, J. Zylberberg, D. Therien, K. P. Kording (2019). A deep learning framework for neuroscience. Nature Neuroscience. 22. 1761-1770.
- [Li et al. J. Vis. 22] Q Li, A Gomez-Villa, M Bertalmío, J Malo (2022). Contrast sensitivity functions in autoencoders. Journal of Vision. 22. 8.
- [Carandini et al. Nat. Rev. Neurosci. 11] M. Carandini, D. J. Heeger (2011). Normalization as a canonical neural computation. Nature reviews. Neuroscience. 13. 51-62.