Biasing optimization to find more informative model metamers

William F. Broderick
Center for Computational Neuroscience, Flatiron Institute
wbroderick@flatironinstitute.org

Daniel Herrera-Esposito
Princeton University
dherrera1911@gmail.com

Erica Shook
Columbia University
erica.shook@columbia.edu

Eero P. Simoncelli
Center for Computational Neuroscience, Flatiron Institute
New York University
eero.simoncelli@nyu.edu

Abstract

Sensory neuroscientists must carefully choose or design experimental stimuli in order to test hypotheses. Recently, scientists have leveraged computational models and automatic differentiation to synthesize model-optimized stimuli, such as model metamers – stimuli that are physically distinct but that produce identical model responses. However, there exist many different stimuli which satisfy the constraints of the stimulus-generation process, but which may give rise to different scientific interpretations. Here, we propose adding a penalty term to the objective function used to generate the stimuli. This biases the synthesis procedure, allowing researchers to preferentially search for stimuli with certain properties. We demonstrate the use of several penalty functions on a simple LGN-inspired model to increase perceptual diversity among synthesized model metamers. By carefully choosing their penalty functions, researchers can better design stimulus sets to address their scientific question.

Introduction

Sensory neuroscience experiments often require careful design of stimuli to test specific properties of the brain. In recent years, one fruitful approach in visual neuroscience has been the use of synthetic stimuli optimized to match or manipulate some feature of the output of a computational model of the visual sytem. Successful examples of this approach include model metamers (e.g., Freeman and Simoncelli, 2011, Feather et al., 2023), Most Exciting Inputs (e.g., Rose et al., 2021; Walker et al., 2019), and perceptography (e.g., Shahbazi et al., 2024).

However, best practices for the generation of model-optimized stimuli are still developing. Metameric stimuli are defined by invariances of an encoding model, and these invariances are often poorly understood – such that predicting the approximate number of such images and their appearances is impossible a priori. Furthermore, the scientific interpretation of these studies depend upon the generated metamer stimuli, and may change if these stimuli turn out to be biased or influenced by the sampling procedure.

Here, we present one approach to combat this problem in the synthesis of model metamers. We propose adding a secondary penalty to the optimization objective function, which seeks to induce additional properties in the resulting image. We demonstrate that such a penalty can be used to encourage perceptual diversity in the resulting metameric images, showing that the model's invariances are broader than the traditional procedure implies.

Our procedure is implemented in the open source python library plenoptic, which can be found at https://github.com/plenoptic-org/plenoptic/.

Methods

Traditionally, model metamer synthesis initializes optimization with a sample of uniform noise and iteratively updates the pixels so as to minimize the difference between the model output of this optimized image and that of some target image. Here, we propose minimizing a weighted sum of this metamer objective and a generic penalty term.

This procedure can be applied to any model which operates on images. The penalty term in this procedure can be any function that operates on a single image and returns a scalar. We show examples using a simple model inspired by the primate visual Lateral Geniculate Nucleus (LGN) which consists of a single center-surround convolutional filter with a single stage of gain control (the LG model described in Berardino et al., 2017). This model cares only about mid-range frequencies and, because of its gain control, is insensitive to absolute luminance. Thus, its metamers will all differ in their lowest and highest frequencies.

We demonstrate two classes of examples:

- We directly incentivize visual diversity by maximizing the Normalized Laplacian Pyramid Distance (NLPD, Laparra et al., 2016) between two simultaneously-generated metamers.
- We encourage different spectral power densities in two simultaneously-generated metamers. We use the

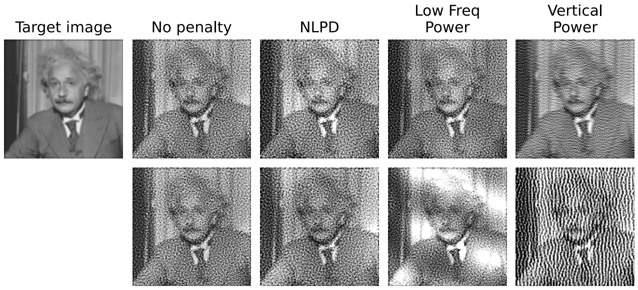

Figure 1: Penalty functions can be used to increase visual diversity of synthesized images. The picture of Albert Einstein in the top left is the target image; the goal of the stimulus synthesis procedure is to produce images whose model output matches that of this image. All other images are model metamers which meet this criterion, synthesized simultaneously in pairs. The first column shows the results of the standard procedure, in which no penalty is used, and the resulting images are indistinguishable. The other three columns shows pairs of images synthesized with an additional penalty, as described in the methods section. The images synthesized so that the NLPD is as large as possible have a low frequency modulation applied, so that Einstein's face is darker in the bottom image and brighter in the top. The bottom image of the pair encouraged to differ as much as possible in the low frequencies is modulated by a high-contrast low frequency diagonal grating. Finally, the pair synthesized to differ as much in the vertical power show a marked difference in orientation, with more vertical power in the bottom image and horizontal in the top.

steerable pyramid (Simoncelli & Freeman, 1995) to decompose each image into orientation and spatial frequency sub-bands and compute the average power in each band. We then use this vector to maximize the difference in low frequencies and (separately) vertical orientation across frequencies.

Results

The metamers for the LGN model can be seen in figure 1. The top-left image of Albert Einstein is the target image: all other images have a model representation that matches that of this natural image. Each column shows a pair of model metamer images which were synthesized together using a different penalty, and all images were initialized with independent patches of white noise. The leftmost pair of images have no penalty and are visually indistinguishable (note that they have a comparable mean-squared error to the other pairs of images, but humans are bad at distinguishing different samples of white noise at high enough resolution). The other pairs are all much more visually distinctive from each other. If one synthesized a set of images without using a penalty, one would likely think that all metameric images for this model are perceptually identical. However, the use of different penalties reveal that this model's metamers include distinguishable images with a variety of appearances. If one wishes to align model and human visual invariances (as in e.g., Broderick et al., 2025), this implies that the model needs to be further improved.

Discussion

We demonstrate a simple modification to the model metamer synthesis procedure which allows greater control over the resulting images. Similar attempts to find more diverse model-optimized stimuli have included initializing with natural images (Broderick et al., 2025) or using equivalent models with alternative training procedures (Feather et al., 2023). While these procedures do result in more diverse stimuli, the penalty function approach presented here demonstrates that even relatively simple and easy-to-reason-about functions can lead to a wider diversity of stimuli.

Through the above examples, we hope to demonstrate how researchers using model-optimized stimuli can find more relevant sets of stimuli for their scientific question through a simple modification of their optimization procedure. The choice of penalty function depends on the specific hypothesis and model, with a wide variety of potential options. Beyond the examples shown here, one could use simple measures of wavelet representational sparsity (Burt & Adelson, 1987) or power spectra to encourage or discourage known properties of natural images, and similar modifications could also be applied to other model-optimized stimulus synthesis procedures.

Disclosure

LLMs were not used in any part of the preparation of this manuscript.

References

Berardino, A., Laparra, V., Ballé, J., & Simoncelli, E. P. (2017, December). Eigen-distortions of hierarchical representations [Presented at: Neural Information Processing Systems 30, Dec 2017, Long Beach, CA]. In I. Guyon, U. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Adv. neural information processing systems (nips*17)* (pp. 1–10, Vol. 30). Curran Associates, Inc.

Broderick, W. F., Rufo, G., Winawer, J., & Simoncelli, E. P. (2025). Foveated metamers of the early visual system. *eLife*. https://doi.org/10.7554/elife.90554.2.

Burt, P. J., & Adelson, E. H. (1987). The laplacian pyramid as a compact image code. In *Readings in computer vision*. Elsevier.

Feather, J., Leclerc, G., Mądry, A., & McDermott, J. H. (2023). Model metamers reveal divergent invariances between biological and artificial neural networks. *Nature Neuroscience*, *26*(11), 2017–2034. https://doi.org/10.1038/s41593-023-01442-0.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889.

Laparra, V., Ballé, J., Berardino, A., & Simoncelli, E. P. (2016). Perceptual image quality assessment using a normalized laplacian pyramid. *Electronic Imaging*, *28*(16), 1–6. https://doi.org/10.2352/issn.2470-1173.2016.16.hvei-103.

Rose, O., Johnson, J., Wang, B., & Ponce, C. R. (2021). Visual prototypes in the ventral stream are attuned to complexity and gaze behavior. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27027-8.

Shahbazi, E., Ma, T., Pernuš, M., Scheirer, W., & Afraz, A. (2024). Perceptography unveils the causal contribution of inferior temporal cortex to visual perception. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-47356-8.

Simoncelli, E. P., & Freeman, W. T. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation. *Proc 2nd IEEE Int'l Conf on Image Proc (ICIP)*, *3*, 444–447. https://doi.org/10.1109/ICIP.1995.537667.

Walker, E. Y., Sinz, F. H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P. G., Ecker, A. S., Reimer, J., Pitkow, X., & Tolias, A. S. (2019). Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience*, *22*(12), 2060–2065. https://doi.org/10.1038/s41593-019-0517-x.