Towards using Adversarially Robust Features as alternative features for rendering of Full-Field Foveated Metamers

A. Introduction

Perceptual metamers are visual stimuli that are physically different but appear the same to a human observer. Finding metameric stimuli provide insight into how the human brain stores and processes information, as they suggest a type of perceptual invariance. Here we outline a novel generative model for foveated visual metamers based on adversarially robust deep network feature matching for the distortion effect that mimics visual crowding in the periphery of a human observer. We find that our generative process yields a visually metameric image with respect to the original sample at similar perceptual rates of previous state of the art gradient-descent models tested through image quality assessment models. This provides insight into characteristics of robust high-level embeddings, and can be extended to study how metamerism arises in human perception [1], and also how peripheral computation in humans may be linked to adversarial robustness in machines [3]. Furthermore, this generative modelling framework may aid future work in developing input pipelines to neural networks that could potentially make them adversarially robust without targeted training, and that are also perceptually more similar to the hierarchical and non-uniform processing mechanisms of primate visual cortex.

A.1. Model and Assessment

The novel generative procedure for metamers is outlined in 1. Qualitative results of our metamer synthesis procedure as compared to current state-of-the-art is shown in 2. Psychophysical evaluations for verification of metamerism by humans are underway.

References

[1] Jenelle Feather, Guillaume Leclerc, Aleksander Madry, and Josh H McDermott. Model metamers illuminate divergences between biological and artificial neural networks. *bioRxiv*, 2022. 1

[2] Jeremy Freeman and Eero P Simoncelli. Metamers of the ventral stream. *Nature neuroscience*, 14(9):1195–1201, 2011. 1

[3] Anne Harrington and Arturo Deza. Finding biological plausibility for adversarially robust features via metameric tasks. In *International Conference on Learning Representations*, 2022. 1

[4] Ruth Rosenholtz, Jie Huang, Alvin Raj, Benjamin J Balas, and Livia Ilie. A summary statistic representation in peripheral vision explains visual search. *Journal of vision*, 12(4):14–14, 2012. 1

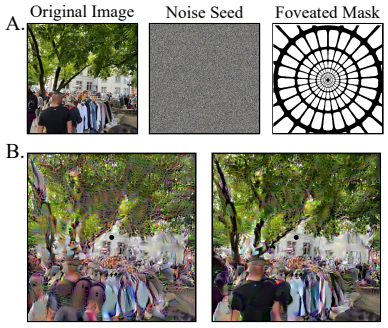

Figure 1. An overview of our Generation Procedure: A) One local synthesis loop consists of iterating through each region in the foveated map from the center outward. For each region, the network is masked such that the effect of the gradient descent towards minimizing the deep network feature loss only alters the noise seed in that region. Each local loop is followed by a small global update to adjust image statistics based on the original image, the result of which is used as the next noise seed for the subsequent local synthesis loop. B) The left shows the noise seed as perturbed after *one* synthesis iteration, and the right after *two* subsequent iterations, notice that these are reminiscent to the metamers of [2, 4] – *without any knowledge of texture statistics*. The images are shown with a black point in the center to emphasize the simulated center point of fixation.

Figure 2. Comparison between foveated metamers synthesized from four different methods, based on four different reference images. These metamers were rendered with a scaling factor held constant of $s = 0.5$. Each model responds different to such stimuli, where the FS model overshoots the texture matching in the image corners, and the NeuroFovea model grossly exaggerates the peripheral distortion for a high scaling factor (it was originally designed to perform will with $s = 0.25$; not shown). *Figure best viewed when zoomed*. Original rendering performed at 512×512.