Novelty Assessment Report

Paper: Generating metamers of human scene understanding
PDF URL: https://openreview.net/pdf?id=cSDXx8V6K9
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Human vision combines low-resolution “gist” information from the visual periphery with sparse but high-resolution information from fixated locations to construct a coherent understanding of a visual scene. In this paper, we introduce MetamerGen, a tool for generating scenes that are aligned with latent human scene representations. MetamerGen is a latent diffusion model that combines peripherally obtained scene gist information with information obtained from scene-viewing fixations to generate image metamers for what humans understand after viewing a scene. Generating images from both high and low resolution (i.e. “foveated”) inputs constitutes a novel image-to-image synthesis problem, which we tackle by introducing a dual-stream representation of the foveated scenes consisting of DINOv2 tokens that fuse detailed features from fixated areas with peripherally degraded features capturing scene context. To evaluate the perceptual alignment of MetamerGen generated images to latent human scene representations, we conducted a same-different behavioral experiment where participants were asked for a “same” or “different” response between the generated and the original image. With that, we identify scene generations that are indeed metamers for the latent scene representations formed by the viewers. MetamerGen is a powerful tool for understanding scene understanding. Our proof-of-concept analyses uncovered specific features at multiple levels of visual processing that contributed to human judgments. While it can generate metamers even conditioned on random fixations, we find that high-level semantic alignment most strongly predicts metamerism when the generated scenes are conditioned on viewers’ own fixated regions.

Core Task Landscape

This paper addresses: **Generating Image Metamers from Foveated and Peripheral Visual Information**

A total of **35 papers** were analyzed and organized into a taxonomy with **11 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Computational Models of Peripheral Vision and Metamer Generation**
- **Rendering Applications and Display Systems**
- **Perceptual Validation and Psychophysical Studies**
- **Biological and Neural Foundations**

Complete Taxonomy Tree

- Generating Image Metamers from Foveated and Peripheral Visual Information Survey Taxonomy
- Computational Models of Peripheral Vision and Metamer Generation
  - Early Visual System Pooling Models ★ (10 papers)
  - [0] Generating metamers of human scene understanding (Anon et al., 2026) View paper
  - [1] Foveated metamers of the early visual system (William F. Broderick, 2023) View paper
  - [11] Author response: Foveated metamers of the early visual system (William F. Broderick, 2025) View paper
  - [14] VSS 2023: Foveated metamers of the early visual system (Broderick, 2023) View paper
  - [21] eLife assessment: Foveated metamers of the early visual system (Behrens, 2023) View paper
  - [24] Reviewer #2 (Public Review): Foveated metamers of the early visual system (Lange, 2023) View paper
  - [25] Reviewer #1 (Public Review): Foveated metamers of the early visual system (William F. Broderick, 2023) View paper
  - [29] Effects of Foveation on Early Visual Representations (Broderick, 2022) View paper
  - [30] VSS 2020: Estimating scaling of retinal and cortical pooling using metamers (Broderick, 2023) View paper
  - [35] Metamers of the Early Visual System (Kong, 2014) View paper
  - Ventral Stream and Higher-Level Feature Models (2 papers)
  - [2] Metamers of the ventral stream (Simoncelli Eero, 2011) View paper
  - [15] Metamers of the ventral stream revisited (Thomas S. A. Wallis, 2015) View paper
  - Texture Statistics and Windowed Feature Methods (2 papers)
  - [18] Accelerated Texforms: Alternative Methods for Generating Unrecognizable Object Images with Preserved Mid-Level Features (Arturo Deza, 2019) View paper
  - [27] A Foveated Model Of Visual Discrimination Based On Windowed Texture Statistics (Simoncelli, 2024) View paper
  - Neural Network and Deep Learning Approaches (4 papers)
  - [6] Towards metamerism via foveated style transfer (Arturo Deza, 2017) View paper
  - [12] Neural Metameric Enhancement for Foveated Rendering (Jiannan Ye, 2024) View paper
  - [13] Gaze-Centric Metamer Computation Based on Peripheral Encoding (Zhenhao Ma, 2023) View paper
  - [28] Real-time peripheral vision metamer computing method (Z.-D. Ma, 2023) View paper
- Rendering Applications and Display Systems
  - Foveated Rendering for Virtual and Augmented Reality (4 papers)

- [3] Beyond blur: Real-time ventral metamers for foveated rendering (DR Walton, 2021) View paper
- [20] Generation of images for stereoscopic displays using selected perceptual features of human visual system (Wernikowski, 2022) View paper
- [33] CUDA-Optimized real-time rendering of a Foveated Visual System (Malkin, 2020) View paper
- [34] Beyond blur (David R. Walton, 2021) View paper
- Holographic Display Systems (3 papers)
- [4] Metameric varifocal holograms (David R. Walton, 2022) View paper
- [8] Perceptually guided computer-generated holography (Kaan Aksit, 2022) View paper
- [23] Metameric Varifocal Holography (KavaklÄ±, 2022) View paper
- Image Warping and Inpainting Applications (2 papers)
- [7] Metameric inpainting for image warping (Rafael Kuffner dos Anjos, 2022) View paper
- [32] Beyond Flicker, Beyond Blur: View-coherent Metameric Light Fields for Foveated Display (Prithvi Kohli, 2022) View paper

• Perceptual Validation and Psychophysical Studies
- Metamer Discriminability and Detection Tasks (1 papers)
- [22] Testing models of peripheral encoding using metamerism in an oddity paradigm (Thomas S. A. Wallis, 2016) View paper
- Visual Search and Peripheral Perception Studies (3 papers)
- [16] Quantifying peripheral and foveal perceived differences in natural image patches to predict visual search performance. (Anna E. Hughes, 2016) View paper
- [17] Peripheral Representations: from Perception to Visual Search (Arturo Deza, 2018) View paper
- [31] Quantitative measures of crowding susceptibility in peripheral vision for large datasets (Shumikhin, 2020) View paper

• Biological and Neural Foundations
- Non-Cone Photoreceptor Contributions (3 papers)
- [5] Form vision from melanopsin in humans (A. Allen, 2019) View paper
- [19] Domain of metamers exciting intrinsically photosensitive retinal ganglion cells (ipRGCs) and rods (FranÃ§oise ViÃ©not, 2012) View paper
- [26] Daylights with high melanopsin stimulation appear reddish in fovea and greenish in periphery (Hirokuni Higashi, 2023) View paper
- Adversarial Robustness and Neural Network Alignment (2 papers)
- [9] Doppelgangers and Adversarial Vulnerability (Kamberov, 2025) View paper
- [10] Finding biological plausibility for adversarially robust features via metameric tasks (Harrington, 2021) View paper

Narrative

Core task: generating image metamers from foveated and peripheral visual information. The field centers on creating images that appear perceptually identical to originals despite differing physically, exploiting the human visual system's spatially varying sensitivity. The taxonomy reveals four main branches. Computational Models of Peripheral Vision and Metamer Generation focuses on algorithmic approaches to pooling and texture synthesis that mimic early visual processing, with works like Ventral Stream Metamers[2] and Foveated Metamers[1] establishing foundational pooling models. Rendering Applications and Display Systems translates these models into practical technologies for virtual reality and holographic displays, as seen in Metameric Varifocal Holograms[4] and Real-time Ventral Metamers[3]. Perceptual Validation and Psychophysical Studies empirically tests whether generated metamers truly match human perception through controlled experiments. Biological and Neural Foundations grounds the computational work in neuroscience, examining how retinal and cortical mechanisms constrain metamer generation.

A particularly active line explores trade-offs between computational efficiency and perceptual fidelity, with Real-time Ventral Metamers[3] and Gaze-Centric Metamer Computation[13] pushing toward interactive applications while maintaining perceptual accuracy. Another thread investigates the biological plausibility of pooling models, questioning whether texture statistics alone suffice or whether deeper neural constraints matter, as in Biological Plausibility Metamers[10]. Generating Scene Metamers[0] sits within the Early Visual System Pooling Models cluster, closely aligned with foundational works like Foveated Metamers[1] and Early Visual Metamers[35] that emphasize texture-statistic pooling. Compared to Foveated Metamers Response[11] and eLife Foveated Metamers[21], which focus on refining perceptual validation, Generating Scene Metamers[0] appears more oriented toward extending computational synthesis methods to complex natural scenes, bridging classical pooling theory with modern scene generation challenges.

Related Works in Same Category

The following **9 sibling papers** share the same taxonomy leaf node with the original paper:

1. Foveated metamers of the early visual system

Authors: William F. Broderick, Gizem Rufo, Jonathan Winawer, Eero P. Simoncelli, J. Winawer | **Year/Venue**: 2023 | **URL**: View paper

Abstract

The ability of humans to discriminate and identify spatial patterns varies across the visual field, and is generally worse in the periphery than in the fovea. This decline in performance is revealed in many kinds of tasks, from detection to recognition. A parsimonious hypothesis is that the representation of any visual feature is blurred (spatially averaged) by an amount that differs for each feature, but that in all cases increases with eccentricity. Here, we examine models for two such feature...

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, using eccentricity-scaled pooling windows to generate metamers that simulate peripheral vision limitations. The candidate paper focuses on low-level features (luminance and spectral energy) averaged in pooling regions to model early visual processing, while the original paper uses a latent diffusion model (MetamerGen) conditioned on DINOv2 features from actual human fixations to generate scene-level metamers that capture higher-level scene understanding beyond simple feature pooling. The key difference is that the candidate employs traditional pooling-based synthesis of basic visual statistics, whereas the original leverages deep generative models to create metamers reflecting semantic scene representations formed through naturalistic viewing behavior.

---

2. Author response: Foveated metamers of the early visual system

Authors: William F. Broderick, Gizem Rufo, Jonathan Winawer, Eero P. Simoncelli | **Year/Venue**: 2025 | **URL**: View paper

Abstract

N/A

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, focusing on generating image metamers through computational models of peripheral vision. The original paper (Generating metamers of human scene understanding) uses a latent diffusion model

conditioned on DINOv2 features from both foveated fixations and blurred peripheral information to generate scene-level metamers, validated through same-different behavioral experiments. The candidate paper (Foveated metamers of the early visual system) appears to focus on lower-level pooling models that average basic visual features like luminance and spectral energy in eccentricity-scaled windows, representing a more traditional approach to early visual system modeling rather than the deep learning-based scene understanding approach of the original.

---

3. VSS 2023: Foveated metamers of the early visual system

Authors: William Ferguson Broderick | **Year/Venue**: 2023 | **URL**: View paper

Abstract

N/A

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, focusing on generating metamers through computational models of peripheral vision. While the original paper (MetamerGen) develops a latent diffusion model that combines peripheral gist information with high-resolution fixation data to generate scene metamers and validates them through behavioral experiments, the candidate paper appears to be a conference presentation (VSS 2023) that likely focuses on foveated metamers using traditional pooling-based approaches in early visual processing. The key difference is that MetamerGen employs modern deep learning techniques (DINOv2, Stable Diffusion) for scene-level understanding with dynamic fixations, whereas the candidate likely uses classical pooling models for simpler visual stimuli in controlled peripheral viewing conditions.

---

4. eLife assessment: Foveated metamers of the early visual system

Authors: Timothy E Behrens | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Human ability to discriminate and identify visual attributes varies across the visual field, and is generally worse in the periphery than in the fovea. This decline in performance is revealed in many kinds of tasks, from detection to recognition. A parsimonious hypothesis is that the representation of any visual feature is blurred (spatially averaged) by an amount that differs for each feature, but that in all cases increases with eccentricity. Here, we examine models for two such features: loca...

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, using eccentricity-scaled pooling windows to generate metamers that simulate peripheral vision limitations. The original paper (MetamerGen) generates scene metamers by combining peripheral gist with high-resolution fixation information using a latent diffusion model conditioned on DINOv2 features, while the candidate paper focuses on simpler pooling models that average low-level features (luminance and spectral energy) in fixed eccentricity-scaled windows to study basic discrimination thresholds. The key difference is that MetamerGen addresses dynamic scene understanding with multiple fixations and semantic content generation, whereas the candidate examines static pooling of elementary visual features across large visual fields.

---

5. Reviewer #2 (Public Review): Foveated metamers of the early visual system

Authors: Floris P. de Lange | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Human ability to discriminate and identify visual attributes varies across the visual field, and is generally worse in the periphery than in the fovea. This decline in performance is revealed in many kinds of tasks, from detection to recognition. A parsimonious hypothesis is that the representation of any visual feature is blurred (spatially averaged) by an amount that differs for each feature, but that in all cases increases with eccentricity. Here, we examine models for two such features: loca...

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, which focuses on generating metamers through eccentricity-scaled pooling of low-level features. The candidate paper examines luminance and spectral energy pooling models with large stimuli to determine critical scaling parameters for peripheral vision, while the original paper (MetamerGen) uses a latent diffusion model conditioned on DINOv2 features from both foveal fixations and peripheral information to generate scene metamers that align with human scene understanding. The key difference is that the candidate uses simple averaging in pooling windows for low-level features, whereas the original employs deep learning to generate complex scene metamers from combined foveal and peripheral representations.

---

6. Reviewer #1 (Public Review): Foveated metamers of the early visual system

Authors: William F. Broderick, Gizem Rufo, Jonathan Winawer, Eero P. Simoncelli | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Human ability to discriminate and identify visual attributes varies across the visual field, and is generally worse in the periphery than in the fovea. This decline in performance is revealed in many kinds of tasks, from detection to recognition. A parsimonious hypothesis is that the representation of any visual feature is blurred (spatially averaged) by an amount that differs for each feature, but that in all cases increases with eccentricity. Here, we examine models for two such features: loca...

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, which focuses on generating metamers through eccentricity-scaled pooling of low-level features. The candidate paper examines luminance and spectral energy pooling models with psychophysical experiments to determine critical scaling parameters for peripheral vision, while the original paper (MetamerGen) uses a latent diffusion model conditioned on DINOv2 features from both foveal fixations and peripheral information to generate scene metamers that align with human scene understanding. The key difference is that the candidate employs traditional pooling-based synthesis of simple features, whereas the original leverages deep generative models to create complex scene metamers from naturalistic viewing behavior.

---

7. Effects of Foveation on Early Visual Representations

Authors: WF Broderick | **Year/Venue**: 2022 | **URL**: View paper

Abstract

â¦ However, it is not true that peripheral vision is just a blurrier version of foveal vision, and â¦ 59 3 Foveated metamers of the early visual system 69 3.1 Abstract . . . . . â¦

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, focusing on generating metamers through eccentricity-dependent visual processing. While the original paper (MetamerGen) uses a latent diffusion model conditioned on DINOv2 features from actual human fixations and peripheral blur to generate scene metamers for understanding high-level scene perception, the candidate paper appears to focus on the fundamental effects of foveation on early visual representations and metamer generation at lower levels

of the visual hierarchy. The key difference is that MetamerGen targets scene understanding through behavioral validation with a same-different paradigm, whereas the candidate examines how foveation affects early visual system representations more directly.

8. VSS 2020: Estimating scaling of retinal and cortical pooling using metamers

Authors: William Ferguson Broderick | **Year/Venue**: 2023 | **URL**: View paper

Abstract

Materials related to the virtual VSS 2020 poster "Estimating scaling of retinal and cortical pooling using metamers"

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, focusing on generating metamers through eccentricity-scaled pooling mechanisms that simulate peripheral vision. The original paper (Generating metamers of human scene understanding) generates scene-level metamers by combining peripheral gist with high-resolution fixation information using a latent diffusion model conditioned on DINOv2 features, while the candidate paper (VSS 2020: Estimating scaling of retinal and cortical pooling using metamers) focuses on estimating the scaling parameters of retinal and cortical pooling windows themselves, using metamers as a psychophysical tool to characterize early visual processing rather than generating complex scene understanding.

9. Metamers of the Early Visual System

Authors: W Kong | **Year/Venue**: 2014 | **URL**: View paper

Abstract

â¦ The visual information processing of human starts at retina. â¦ to generate metamers for this model of early visual system and â¦ Shortening viewing distance ends up testing more peripheral â¦

Relationship Analysis

Both papers belong to the Early Visual System Pooling Models category, focusing on generating metamers through computational models of peripheral vision. The original paper (MetamerGen) generates scene metamers by combining foveated fixation information with peripheral gist using a latent diffusion model conditioned on DINOv2 features, validated through behavioral experiments testing human scene understanding. The candidate paper focuses on metamers of the early visual system using simpler pooling models that average low-level features (luminance, spectral energy) in eccentricity-scaled windows, representing a more foundational approach to modeling early visual processing without the high-level semantic integration or behavioral validation present in the original work.

Contributions Analysis

Overall novelty summary. MetamerGen introduces a latent diffusion model that generates scene metamers by combining peripheral gist information with high-resolution fixation data. The paper resides in the Early Visual System Pooling Models leaf, which contains ten papers focused on eccentricity-scaled feature averaging to simulate early visual processing. This is the most populated leaf in the Computational Models branch, indicating a crowded research direction where foundational pooling approaches have been extensively explored. The work extends classical pooling theory to complex natural scenes using modern generative architectures.

The taxonomy reveals neighboring leaves addressing ventral stream modeling (two papers on mid-to-high level features), texture statistics methods (two papers on windowed feature distributions), and neural network approaches (four papers on deep learning-based metamer generation). MetamerGen bridges early visual pooling with deep learning methods, sitting at the boundary between hand-crafted feature models and learned representations. The dual-stream DINOv2 conditioning mechanism connects to the Neural Network and Deep Learning Approaches leaf, which explores encoder-decoder and diffusion architectures, though those works typically do not emphasize foveated input structures or scene-level gist integration.

Among eleven candidates examined, the core MetamerGen contribution shows one refutable candidate from five examined, suggesting some overlap with prior diffusion-based metamer generation. The dual-stream DINOv2 conditioning examined two candidates with no clear refutation, indicating potential novelty in the specific architectural fusion of peripheral and foveated features. The behavioral paradigm contribution examined four candidates without refutation, though this may reflect the limited search scope rather than definitive novelty. The analysis covers top-K semantic matches and does not constitute an exhaustive literature review.

Based on the limited search scope, MetamerGen appears to offer architectural innovations in fusing foveated and peripheral representations within diffusion models, though the core idea of generating metamers via learned features has precedent. The work's positioning in a crowded leaf suggests incremental refinement rather than paradigm shift, but the dual-stream conditioning and scene-level synthesis may represent meaningful extensions. A broader literature search would be needed to assess whether similar foveated diffusion architectures exist outside the examined candidates.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: MetamerGen: a latent diffusion model for generating scene metamers

Description: The authors propose MetamerGen, a latent diffusion model that combines peripheral gist information with fixation-based foveal information to generate image metamers aligned with human scene understanding. The model uses a dual-stream representation of foveated scenes with DINOv2 tokens to fuse detailed features from fixated areas with peripherally degraded features.

This contribution was assessed against **5 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Uncertainty Quantification in HSI Reconstruction using Physics-Aware Diffusion Priors and Optics-Encoded Measurements

URL: View paper

Brief Assessment

HSI Diffusion Priors[40] focuses on hyperspectral image reconstruction from compressed measurements using physics-aware diffusion priors for uncertainty quantification, not on generating scene metamers aligned with human visual representations or scene understanding.

2. Modeling human scene understanding fixation-by-fixation using generative models

URL: View paper

Brief Assessment

Fixation Scene Understanding[39] appears to focus on fixation-by-fixation scene understanding using generative models and scene metamers, but the provided candidate context is too limited (only fragments) to determine whether it presents the same dual-stream DINOv2-based architecture combining peripheral gist with foveal fixation information that characterizes MetamerGen.

3. Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding

URL: View paper

Brief Assessment

Vision Decoding Diffusion[41] focuses on decoding visual stimuli from fMRI brain recordings to reconstruct images, not on generating scene metamers aligned with human scene understanding from fixation and peripheral vision data.

---

4. Unraveling Metameric Dilemma for Spectral Reconstruction: A High-Fidelity Approach via Semi-Supervised Learning

URL: View paper

Brief Assessment

Spectral Reconstruction Metameric[42] addresses spectral reconstruction from RGB images in hyperspectral imaging, not scene understanding or human visual perception. The candidate focuses on resolving metameric ambiguities in spectral signals for material identification, while the original paper generates scene metamers aligned with human scene representations using fixation-based conditioning.

---

5. Seen2Scene

URL: View paper

Prior Art Analysis

Seen2Scene[36] demonstrates that a latent diffusion model combining peripheral and foveal information for scene generation was developed prior to the original paper's submission. Both papers use dinov2 features to extract foveal information from fixated regions and peripheral/gist information from the broader scene, then integrate these through adapter-based frameworks into stable diffusion's cross-attention mechanism. The architectural approach of dual-stream conditioning (foveal + peripheral) and the use of dinov2 tokens for representing fixated vs. non-fixated regions are nearly identical between the two works.

Evidence

Evidence 1 - **Rationale**: Both papers describe latent diffusion models that combine fixation-based foveal information with peripheral scene information to generate scenes aligned with human understanding. - **Original**: metamergen is a latent diffusion model that combines peripherally obtained scene gist information with information obtained from scene-viewing fixations to generate image metamers for what humans understand after viewing a scene - **Candidate**: seen2scene, a framework for modeling human scene understanding by controlling the inputs used to generate a visual hypothesis of the scene. seen2scene uses a self-supervised encoder to extract features from fixated scene regions, which guide a pre-trained text-to-image latent diffusion model through...

Evidence 2 - **Rationale**: Both papers use dinov2's dual representation structure to model foveal (patch tokens) and peripheral (global tokens) information, demonstrating the same technical approach to representing fixated vs. non-fixated scene regions. - **Original**: we introduce a dual-stream representation of the foveated scenes consisting of dinov2 tokens that fuse detailed features from fixated areas with peripherally degraded features capturing scene context - **Candidate**: dinov2 provides multiple types of embeddings: patch tokens that capture local spatial information in a grid covering the input image and global tokens (cls and register tokens) that capture broader contextual features (darcet et al., 2024). this multi-scale structure aligns with human vision; patch...

Evidence 3 - **Rationale**: Both papers cite the same IP-adapter work and describe the same approach of integrating dinov2 embeddings into stable diffusion's cross-attention mechanism, indicating identical architectural foundations. - **Original**: similar to ip-adapters (ye et al., 2023), which integrate clip image embeddings into stable diffusion, we learn how to incorporate dinov2 patch embeddings into the cross-attention mechanism of the text-to-image stable diffusion model - **Candidate**: seen2scene builds on stable diffusion by replacing its text conditioning with dinov2 visual embeddings through the unet's cross-attention mechanism (h. y e et al., 2023; z. y e et al., 2025)

Contribution 2: Dual-stream DINOv2-based conditioning mechanism for foveated image synthesis

Description: The authors develop a novel conditioning approach that uses separate DINOv2 feature streams for foveal (fixated) and peripheral (blurred) visual information. These streams are integrated through adapter-based cross-attention mechanisms into a pretrained Stable Diffusion model, enabling generation from variable-resolution inputs.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Extensive ProGAN: a robust model for human face frontalization

URL: View paper

Brief Assessment

ProGAN Face Frontalization[43] focuses on face frontalization using progressive GANs with a two-pathway architecture for facial regions, not on dual-stream conditioning mechanisms that fuse foveal and peripheral visual features for general image synthesis using DINOv2 and diffusion models.

---

2. A Gated Peripheral-Foveal Convolutional Neural Network for Unified Image Aesthetic Prediction

URL: View paper

Brief Assessment

Peripheral-Foveal Aesthetic Prediction[44] focuses on aesthetic assessment using peripheral and foveal vision concepts with CNNs, not on conditioning mechanisms for diffusion-based image synthesis or generation tasks.

Contribution 3: Real-time behavioral paradigm for identifying scene metamers

Description: The authors introduce a gaze-contingent experimental paradigm where participants view scenes for variable fixations, followed by a brief presentation of either the original or a MetamerGen-generated image. This paradigm enables identification of which generated scenes are true metamers for human scene understanding through same-different judgments.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Spatial structure, phase, and

URL: View paper

Brief Assessment

Spatial Structure Phase[38] focuses on detecting contrast-defined targets in natural images using same-different judgments, but does not address gaze-contingent paradigms or scene metamers generated from fixation patterns. The candidate examines phase alignment effects on target detection, not metamer identification from viewing behavior.

---

2. Are summary statistics enough? Evidence for the importance of shape in guiding visual search

URL: View paper

Brief Assessment

Summary Statistics Search[37] uses a gaze-contingent paradigm for visual search tasks with statistically-matched object replacements, not for evaluating scene metamers or scene understanding representations as in the original paper.

---

3. Metameric varifocal holograms

URL: View paper

Brief Assessment

Metameric Varifocal Holograms[4] focuses on computer-generated holography for display systems using gaze-contingent rendering, not on behavioral paradigms for evaluating scene understanding metamers through same-different judgments as in the original paper.

---

4. Seen2Scene

URL: View paper

Brief Assessment

Seen2Scene[36] does not describe a behavioral paradigm for identifying metamers through same-different judgments. The candidate focuses on computational evaluation using clip and dreamsim metrics, not human behavioral experiments to identify perceptual metamers.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Generating metamers of human scene understanding View paper
- [1] Foveated metamers of the early visual system View paper
- [2] Metamers of the ventral stream View paper
- [3] Beyond blur: Real-time ventral metamers for foveated rendering View paper
- [4] Metameric varifocal holograms View paper
- [5] Form vision from melanopsin in humans View paper
- [6] Towards metamerism via foveated style transfer View paper
- [7] Metameric inpainting for image warping View paper
- [8] Perceptually guided computer-generated holography View paper
- [9] Doppelgangers and Adversarial Vulnerability View paper
- [10] Finding biological plausibility for adversarially robust features via metameric tasks View paper
- [11] Author response: Foveated metamers of the early visual system View paper
- [12] Neural Metameric Enhancement for Foveated Rendering View paper
- [13] Gaze-Centric Metamer Computation Based on Peripheral Encoding View paper
- [14] VSS 2023: Foveated metamers of the early visual system View paper
- [15] Metamers of the ventral stream revisited View paper
- [16] Quantifying peripheral and foveal perceived differences in natural image patches to predict visual search performance. View paper
- [17] Peripheral Representations: from Perception to Visual Search View paper
- [18] Accelerated Texforms: Alternative Methods for Generating Unrecognizable Object Images with Preserved Mid-Level Features View paper
- [19] Domain of metamers exciting intrinsically photosensitive retinal ganglion cells (ipRGCs) and rods View paper
- [20] Generation of images for stereoscopic displays using selected perceptual features of human visual system View paper
- [21] eLife assessment: Foveated metamers of the early visual system View paper
- [22] Testing models of peripheral encoding using metamerism in an oddity paradigm View paper
- [23] Metameric Varifocal Holography View paper
- [24] Reviewer #2 (Public Review): Foveated metamers of the early visual system View paper
- [25] Reviewer #1 (Public Review): Foveated metamers of the early visual system View paper
- [26] Daylights with high melanopsin stimulation appear reddish in fovea and greenish in periphery View paper
- [27] A Foveated Model Of Visual Discrimination Based On Windowed Texture Statistics View paper
- [28] Real-time peripheral vision metamer computing method View paper
- [29] Effects of Foveation on Early Visual Representations View paper
- [30] VSS 2020: Estimating scaling of retinal and cortical pooling using metamers View paper
- [31] Quantitative measures of crowding susceptibility in peripheral vision for large datasets View paper
- [32] Beyond Flicker, Beyond Blur: View-coherent Metameric Light Fields for Foveated Display View paper
- [33] CUDA-Optimized real-time rendering of a Foveated Visual System View paper
- [34] Beyond blur View paper
- [35] Metamers of the Early Visual System View paper
- [36] Seen2Scene View paper
- [37] Are summary statistics enough? Evidence for the importance of shape in guiding visual search View paper
- [38] Spatial structure, phase, and View paper
- [39] Modeling human scene understanding fixation-by-fixation using generative models View paper
- [40] Uncertainty Quantification in HSI Reconstruction using Physics-Aware Diffusion Priors and Optics-Encoded Measurements View paper
- [41] Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding View paper
- [42] Unraveling Metameric Dilemma for Spectral Reconstruction: A High-Fidelity Approach via Semi-Supervised Learning View paper
- [43] Extensive ProGAN: a robust model for human face frontalization View paper
- [44] A Gated Peripheral-Foveal Convolutional Neural Network for Unified Image Aesthetic Prediction View paper