

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Benchmarking mid-level vision with texture-defined 3D objects

#### **Permalink**

<https://escholarship.org/uc/item/7qc5j87f>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

#### **Authors**

Friedman, Yoni  
O'Connell, Thomas  
Siegel, Max  
et al.

#### **Publication Date**

2022

Peer reviewed

# Benchmarking mid-level vision with texture-defined 3D objects

**Yoni Friedman**

MIT, Cambridge, Massachusetts, United States

**Thomas O’Connell**

MIT, Cambridge, Massachusetts, United States

**Max Siegel**

MIT, Cambridge, Massachusetts, United States

**Daniel Bear**

Stanford University, Stanford, California, United States

**Tuan Anh Le**

MIT, Cambridge, Massachusetts, United States

**Bernhard Egger**

Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

**Josh Tenenbaum**

MIT, Cambridge, Massachusetts, United States

## Abstract

We introduce a new benchmark dataset based on classic methods of studying 3D shape perception from texture and motion, inspired by earlier work on Gestalt principles of perceptual organization and the ecological (Gibsonian) approach to perception of structure in moving displays. The dataset consists of parametric 3D shapes (superquadrics) with procedurally generated textures rotating and translating against a similarly textured backdrop. We expect these stimuli to be challenging for current computer vision models, as they depart from the statistics of real-world or realistically rendered stimuli. We test a variety of models’ ability to segment textured stimuli across three training conditions: pre-trained on naturalistic stimuli, pre-trained+fine-tuned on textured stimuli, trained on textured stimuli. While no models generalize to segment textured stimuli without fine-tuning, performance improves with fine-tuning and training on textured stimuli. We will discuss how this benchmark can guide models of scene perception towards more human-like robustness and generality.