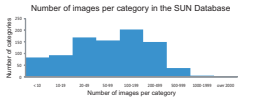
Building a taxonomy of visual scenes: Typicality ratings and hierarchical classification

Krista A. Ehinger[1], Antonio Torralba[2], & Aude Oliva[1]
[1]Department of Brain & Cognitive Sciences, MIT, [2]Computer Science & Artificial Intelligence Laboratory, MIT

Massachusetts Institute of Technology

Scene UNderstanding (SUN) Database

- 122,968 images from 706 semantic categories
- Candidate scene categories were collected from WordNet (all terms corresponding to types of places / environments / landmarks)
- No specific places or places that lack a visual identity (like "territory")

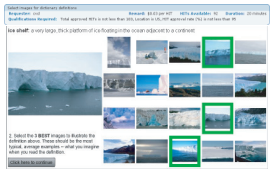


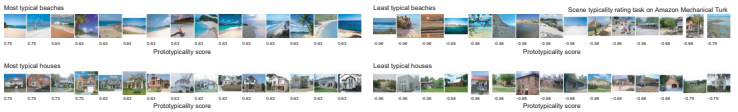
Scene Hierarchy

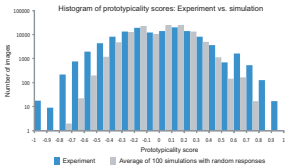
- 20 subjects grouped 8 sets of images (3 indoor, 2 outdoor natural, 3 outdoor man-made), each scene category represented by one typical image
- Hierarchical clustering of grouping distances (how often images were placed in the same group)
- Natural scenes organized by surface terrain / climate, man-made environments organized by function

Rating Scene Typicality

- 675 workers participated in 52,068 trials on Amazon Mechanical Turk
- Workers saw array of images with a category name and definition:
  Task 1. Select the image that matches the definition (4AFC)
  Task 2. Select the 3 best exemplars from a set of 20 images
  Task 3. Select the 3 worst exemplars from the same set of 20 images
- Images were drawn randomly for each trial, with each image appearing 8-10 times across the experiment
- Prototypicality score = ((“best” votes) - 0.9*(“worst” votes)) / appearances



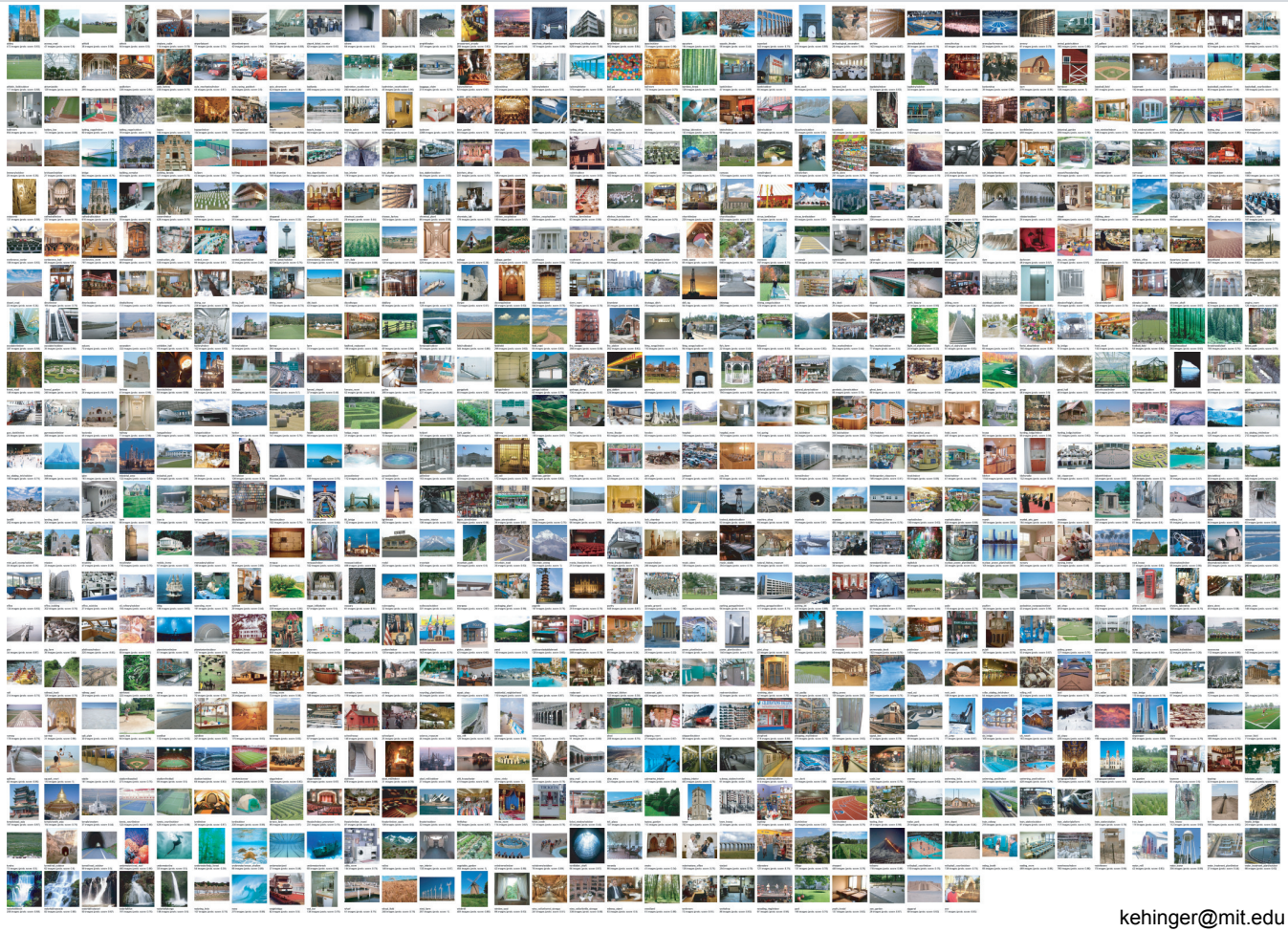




Worker quality and rating consistency

- Average score on 4AFC task was 98%, workers avoided reselecting the same images as both “best” and “worst” on 96% of trials
- Responses to duplicated images were positively correlated (0.53), and “best” / “worst” votes to the same image were negatively correlated (-0.48)
- Extremely high/low prototypicality scores occured much more often than expected by chance: 11% of images scored below the 1st percentile or above the 99th percentile of ratings in a simulated experiment where “workers” responded randomly

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (in press). SUN Database: Large-scale Scene Recognition from Abbey to Zoo. CVPR 2010.



kehinger@mit.edu