Estimating scene typicality from human ratings and image features

Krista A. Ehinger, Jianxiong Xiao, Antonio Torralba, Aude Oliva
“chair”
• Are there “typical” examples of scene categories?
• Do more typical examples contain more of the visual features relevant for scene categorization?
SUN Database: 899 scene categories, 130,519 images

Xiao, Hays, Ehinger, Oliva, & Torralba (2010)
Scene Understanding (SUN) database

Scenes:
- baseball_field
- warehouse
- theater
- opera_house
- forest

Not scenes:
- equator
- golden_gate_bridge
- outdoors

- forest – evergreen
- forest – deciduous
Rating scene typicality

• 706 categories
• 124,901 images (22 – 2360 images/category)
• 935 people contributed ratings through Amazon’s Mechanical Turk service
living room: a room in a private residence intended for general social and leisure activities

1. Read the word and definition above, then select the image that matches the definition.

Click here to continue
living room: a room in a private residence intended for general social and leisure activities

2. Select the 3 BEST images to illustrate the definition above. These should be the most typical, average examples – what you imagine when you read the definition.

Click here to continue
living room: a room in a private residence intended for general social and leisure activities

3. Select the 3 WORST images to illustrate the definition above. (Note that these are the same images as before, just in a new order.)

[Click here to continue]
Rating scene typicality

Category images

20 images per trial
Rating scene typicality

Category images

20 images per trial

...
Rating scene typicality

Category images

![Category images]

20 images per trial

![20 images per trial]

77,331 trials, each image appeared 12-15 times
Rating scene typicality

Category images

20 images per trial

14 appearances
3 “best”
0 “worst”

12 appearances
1 “best”
1 “worst”

77,331 trials, each image appeared 12-15 times
Typicality score

Typicality = \frac{\text{“best” votes} - k(\text{“worst” votes})}{\text{Number of appearances}}

k = 0.9

Voted a “best” exemplar on 8 out of 12 appearances

Voted a “worst” exemplar on 7 out of 12 appearances

0.67

-0.52
Examples

Most typical beaches

Most typical bedrooms
Examples

Least typical beaches

Least typical bedrooms
Rating scene typicality

Category images

20 images per trial

Simulation: Computer randomly selected images

14 appearances
3 “best”
0 “worst”

12 appearances
1 “best”
1 “worst”
Typicality scores: Experiment vs. simulation

- Score of most typical image in category
- Number of images in category

- Experiment
- Simulation
• Are there “typical” examples of scene categories?
  – Yes, there is a range of typicality in scene categories
• Are there “typical” examples of scene categories?
  – Yes, there is a range of typicality in scene categories

• Do more typical examples contain more of the visual features relevant for scene categorization?
Are more typical images classified more accurately?
Machine-vision classifier

One vs. all SVM classifier trained on 50 random images in each category, tested on 50 images

- Feature histograms (color, edges, texture features)
- GIST descriptor
- SIFT descriptor
- HOG (histograms of oriented gradients)
- SSIM (self-similarity)
- Geometric classes

Felzenszwalb, Girshick, McAllester, & Ramanan (2010); Hoiem, Efros, & Herbert (2007); Lazebnik, Schmid, & Ponce (2006); Oliva & Torralba (2001); Renninger & Malik (2004); Shechtman & Irani (2007)
Classification performance

Humans = 68% correct
Classifier = 38% correct
Chance = 0.25% correct

"All features" classifier
Scene classification vs. scene typicality

Average classifier performance = 38%

Chance = 0.25%
Classification confidence vs. scene typicality

![Bar chart showing the relationship between classification confidence and scene typicality across quartiles. The chart compares correctly classified images (light gray) and incorrectly classified images (dark gray). There is a clear trend indicating that confidence decreases as typicality increases.]
• Are there “typical” examples of scene categories?
  – Yes, there is a range of typicality in scene categories

• Do more typical examples contain more of the visual features relevant for scene categorization?
  – Yes, more typical scenes are classified more accurately
Thank you!

Jianxiong Xiao  Antonio Torralba  Aude Oliva

This work is funded by NSF CAREER award (0546262), NSF grants (0705677 and 1016862), and a NEI grant (EY02484) to A.O; NSF CAREER award (0747120) to A.T; and Google research awards to A.O and A.T.