# Modeling Context Effects on Image Memorability

Zoya Bylinskii Phillip Isola Antonio Torralba Aude Oliva
Computer Science and Artificial Intelligence Lab, MIT, Cambridge USA
{zoya,phillipi,torralba,oliva}@mit.edu

## Abstract

*We use an information-theoretic framework to quantify context differences and image distinctiveness for predicting image memorability. We are able to quantify, using a large natural scene database, the observation that images that are unique or distinct with respect to their image context are better remembered.*

## 1. Introduction

Previous memory studies have suggested that items that stand out from their context are better remembered [4, 6, 5]. Nevertheless, recent work on predicting image memorability [3, 2] has largely ignored the effects of **image context** (the set of images from which the experimental sequence is sampled) on memory performance, instead focusing on the modeling of intrinsic image features alone.

Here, we are able to quantify, using a large natural scene database, the observation that images that are unique or distinct with respect to their image context are better remembered. By systematically varying the image context across experiments, we model the change in context at the feature level. We compute statistics over image features to predict corresponding changes in memorability.

## 2. Experiments

For our experiments, we selected 21 different indoor and outdoor scene categories from the SUN Database [7] for a total of 9K images. From each scene category, 25% of the images were randomly chosen to be **targets** for which we obtained memorability scores[1]. We ran Amazon Mechanical Turk (AMT) studies following the protocol of Isola et al. [3] to collect **memorability scores** (i.e. performance on a recognition memory task) for all 1754 target images. We controlled image context by running two variations of the experiment. In the first (AMT 1), participants saw images from only one scene category at a time. In the second (AMT 2), participants saw images from all scene categories mixed together.

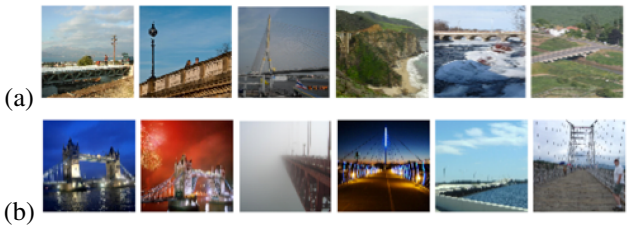


Figure 1. Bridges that are memorable (a) within the bridge category, (b) across all scene categories. For applications where we want images to be memorable, we may want to select those images that are distinct with respect to multiple different contexts.

In all studies, participants were presented with sequences of images for 1 second each, and instructed to press a key when they detected an image repeat. We define a **hit** to be a correct response to an image presented for the second time. A **miss** is when an image was repeated, but not recognized. We measure the memorability of an image as its *hit rate* (**HR**)[2], computed as the ratio of hits to total image presentations (hits+misses). By systematically varying the context for our target images between AMT 1 and AMT 2, we directly measure context effects on image memorability.

## 3. Analysis

We call images **contextually distinct** if they are distinct with respect to their image context. To model context, we estimate a kernel density over the images in a context $C$ as:

$$P_c(f_i) = \frac{1}{\|C\|} \sum_{j \in C} K(f_i - f_j) \quad (1)$$

where $f_i = F(I)$ is a feature vector for image $I$. We used the $fc7$ features from the Places-CNN [8] for the experiments presented here[3], and the Epanechnikov kernel

[1]The FIGRIM dataset, consisting of memorability scores and image features for the 1754 targets tested in 2 different contexts, is available for download at http://figrim.mit.edu

[2]We have also computed the false alarm rate (FAR), and functions of HR and FAR, including d-prime and accuracy. These results can be found in [1]. Here, we only present trends computed with HR, which are representative of the trends computed with the other memorability measures.

[3]Results with other feature spaces can be found in [1].

$K$ with leave-one-out-cross-validation to select the kernel bandwidth. The distinctiveness of an image with respect to a context is then:

$$D(I;C) = -\log P_c(f_i) \quad (2)$$

We also measured the **context entropy** by averaging $D(I;C)$ over all the images in a given image context. This is just the information-theoretic entropy:

$$H(C) = \mathbb{E}_c[D(I;C)] \quad (3)$$

## 4. Results

Denoting $C_1$ as the within-category context of AMT 1 and $C_2$ as the across-category context of AMT 2, we find a Pearson correlation of $r = 0.26$[4] between $D(I;C_1)$ and $\text{HR}(I)$, and $r = 0.24$ between $D(I;C_2)$ and $\text{HR}(I)$ across 1754 target images. Thus, **more contextually distinct images are more memorable**. Moreover, the correlation between $D(I;C_2) - D(I;C_1)$ and $\text{HR}_{C_2}(I)$ - $\text{HR}_{C_1}(I)$ is 0.35. Thus, **the memorability of images can be changed simply by changing which images are presented together**. Consider the images that were memorable with respect to their own category, but became forgettable when combined with other categories. For instance, as in Fig. 1, a bridge that looks like a pasture stands out when combined with other bridges, but drops in memorability when combined with all scene categories (including pastures). In fact, a linear multi-class SVM trained to predict scene categories from $fc7$ features is also more likely to mislabel this bridge instance, and assign it a lower probability of belonging to the bridge class. Across all 1754 target images, the correlation between the probability of the correct scene label and the change in memorability due to context is $r = 0.30$. In other words, the images least likely to belong to their own category experience the greatest drop in memorability when combined with images of other categories. **To make a truly memorable image one must consider the possible contexts in which an image can occur and make the image as distinct as possible with respect to all those contexts.**

Furthermore, categories with many contextually distinct images are more memorable overall. The correlation between $H(C)$ and $\overline{\text{HR}} = \mathbb{E}_c[\text{HR(I)}]$ is $r = 0.53$ (Fig. 2). For instance, the *cockpit* category (low $H(C)$) contains a relatively homogeneous collection and configuration of scene elements: dashboards and buttons. On the other hand, the *amusement park* category (high $H(C)$), consists of a much larger variability of images, containing roller-coasters, concession stands, and rides. **By choosing images that are more distinct from one another, one can increase the total number of images remembered in a single sitting.**

[4] All the correlations reported here are significant with $p < 0.01$.

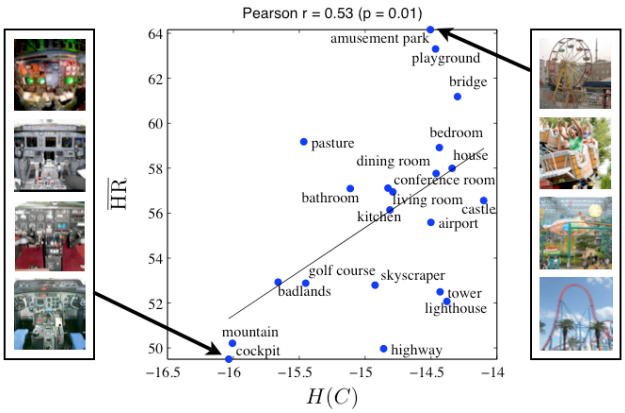


Figure 2. Scene categories with more diverse images are more memorable overall.

## 5. Conclusion

We showed that changes in image context predictably affect the memorability of images, and this can be modeled using an information theoretic framework. This opens up applications for automatic image selection and curation such as creating more memorable photo streams, or making frames stand out in a movie. Memorability predictions that take image context into account are better suited for customizing applications and user interfaces to the individual.

## References


[1] Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and A. Oliva. Intrinsic and extrinsic effects on image memorability. *Vision Research*, 2015. in press. 1

[2] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the Intrinsic Memorability of Images. In *NIPS*, 2011. 1

[3] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1

[4] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects. *J. of Exp. Psych.: General*, 139:558–578, 2010. 1

[5] L. Standing. Learning 10,000 pictures. *The Quart. J. of Exp. Psych.*, 25(2):207–222, 1973. 1

[6] S. Vogt and S. Magnussen. Long-term memory for 400 pictures on a common theme. *Exp. Psych.*, 54(4):298–303, 2007. 1

[7] J. Xiao, J. Hayes, K. Ehinger, A. Oliva, and A. Torralba. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. In *CVPR*, 2010. 1

[8] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition using Places Database. In *NIPS*, 2014. 1