How you look at a picture determines if you will remember it

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 introduce an approach to predict user memory of images: specifically, to infer if a particular user will remember a particular image at a later time. We measure an unconscious motor signature of memory: eye fixations while a user views images. We train a boosting classifier to differentiate eye movements that lead to a successful memory of an image from those that lead the image to be forgotten.

1. Introduction

Previous image memorability studies [3, 2] have shown that people are highly consistent in which images they remember and forget, using this insight to estimate the memorability of an image, independent of the observer. Here, we demonstrate that it is possible to make better predictions about a particular individual on an individual trial by leveraging the individual's eye movements to determine if an image will be later remembered.

2. Eyetracking experiments

We used a similar set-up to [3] to run our memorability studies. We collected the memorability scores and eye movements of a total of 40 participants (averaging 14.1 per image) on 630 target images from the FIGRIM dataset[1]. Participants saw a sequence of images, presented for 2 seconds each, and were asked to respond (by keypress) anytime they recognized an image recurring in the sequence. Eyetracking was performed using an EyeLink1000, with image stimuli presented at a resolution of 1000×1000 pixels. More details can be found in [1].

3. Eye movement predictor

Given a set of fixations on an image, we want to predict: will the viewer remember this image at a later point in time? Our key assumption is that if a viewer's fixations differ from the fixations expected on an image, the viewer may not have encoded (the relevant parts of) the image correctly. We label a set of fixations that have a high probability of being elicited by an image as successful encoding fixations for the image, and predict that they will lead to a correct recognition of the image later. Otherwise, we predict that the image will be forgotten (see Fig. 1).

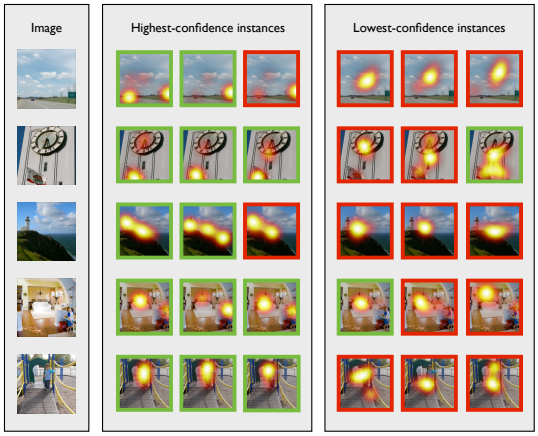
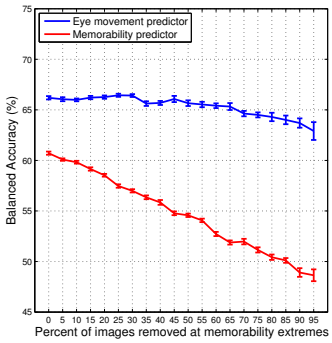


Figure 1. Individual viewers' fixation maps overlaid on top of the images viewed. For each of these 5 example images, we include the 3 highest and 3 lowest confidence instances under the image's classifier (trained to differentiate fixations on this image from fixations on other images). Fixations that later led to a correct recognition of the image are outlined in green, and those where the image was unsuccessfully remembered are in red. This depicts some of the successes and failure modes of our model.

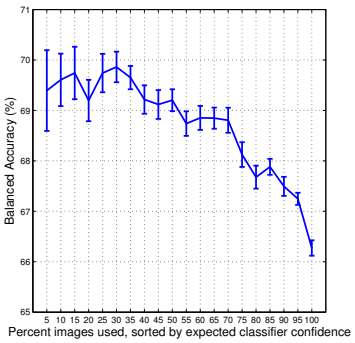
Fixations were coarsely binned (20×20 grid) and smoothed (Gaussian with $\sigma = 2$) into a **fixation map**. For each image, we trained an boosting classifier $G_i = g(I)$ to differentiate fixation maps on image I (positive examples) from fixation maps on all other images (negative examples). Next, we evaluated both successful and unsuccessful fixations on image I under the classifier G_i to obtain confidence values for each set of fixations. We learned a threshold on the confidence values from the training data that best differentiates successful from unsuccessful fixations on the train-

[1]The FIGRIM dataset, consisting of memorability scores for 1754 images across 21 scene categories of the SUN database [4], is available at http://figrim.mit.edu

(a)



(b)



(c)

Figure 2. (b) When we prune images at the memorability extremes, memorability scores fall to chance as a predictor of per-trial memory performance while our eye movement predictor remains robust. (c) Our classifier makes more accurate predictions when it has higher expected confidence. (a) Images sorted by expected confidence (from least to most). Overlaid on top of each image is the average fixation map computed over all successful encoding fixations.

ing data.

At test time, for a held-out set of participants, we evaluated a participant's encoding fixations under the classifier G_i to obtain a confidence value. We thresholded this confidence value with the threshold chosen during training to produce the final prediction: whether the participant's fixations are successful or unsuccessful.

4. Eye movements predict image memories

As a baseline we used an image's memorability score[1] as a predictor to make trial-by-trial predictions for whether a particular individual will remember a particular image. This predictor, which does not take into account individual variability, achieves a balanced accuracy of 60.09% ($SD : 1.55\%$), significantly above chance (50%).

Our eye movement predictor uses a viewer's eye movements to predict whether an image will be remembered. Over 15 different splits of participant data, this classifier obtained a balanced accuracy of 66.02% ($SD : 0.83\%$).

To further highlight the power of the individual trial predictor, we sorted images by their memorability scores. As we progressively removed images at the extremes (Fig.2b), memorability scores fell to chance as expected. Meanwhile, our eye movement predictor remained robust.

5. Not all images are equally predictable

An image with all of the important content in the center might not require the viewers to move their eyes very much and this makes prediction particularly difficult because successful and unsuccessful fixations may not be that different. Thus, we can separate images into those on which confident predictions can be made from those on which prediction will be difficult. For an image I, we compute the expected confidence of classifier G_i as the average confidence value over its positive training examples. Sorting images by this expected confidence measure (Fig.2a), we obtain the results in Fig.2c. Our predictor's accuracy reaches almost 70% on the images for which it is most confidence.

Thus, it is possible to select images that our classifier is expected to do well on. This becomes an important feature for applications where we have a choice over the images that can be used, and need to have a system that can robustly predict from eye fixations, whether an image will be later remembered.

6. Future Applications

Imagine an automatic system that monitors the eye movements of a student on a set of lecture slides and uses this information to determine whether or not the student is "paying attention". If not, the system may either alert the student to increase attentiveness at this point in time, or else the system may continue to re-present the material again until it has acquired some confidence that the student has properly encoded the content.

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, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What Makes a Photograph Memorable? *PAMI*, 36(7):1469–1482, 2014. 1

[3] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *CVPR*, 2011. 1

[4] 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