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

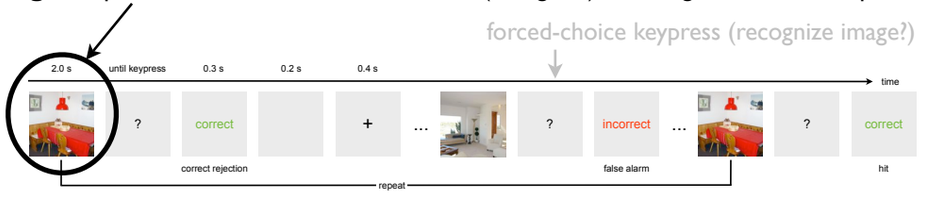
Zoya Bylinskii, Phillip Isola, Antonio Torralba, Aude Oliva
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

IEEE 2015 Conference on Computer Vision and Pattern Recognition

Dataset: http://figrim.mit.edu
Paper: Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., Oliva, A. "Intrinsic and Extrinsic Effects on Image Memorability", Vision Research 2015.

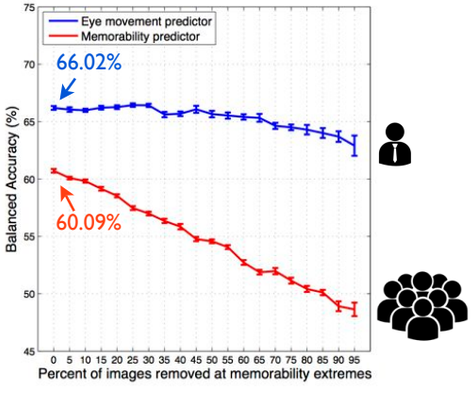
Abstract: We introduce an approach to predict user memory of images: specifically, to infer if a particular user will remember a particular image at a later time.
We measure an unconscious motor signature of memory: eye fixations while a user views images. We train a boosting classifier to differentiate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

goal: predict whether user will remember (recognize) this image later in the sequence



we use eye-movements at encoding to make predictions

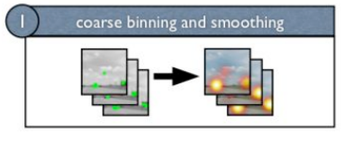
in-lab eyetracking experiments with:
- 630 target natural scene images
- 42 participants (~16.2 per image)



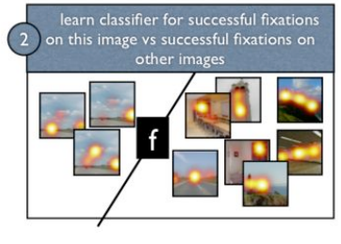
A crowd can be used to make significantly above-chance predictions for an individual user.

When images at the memorability extremes are removed, eye-movements remain robust predictors for an individual user.

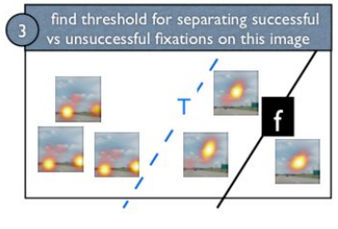
$$\text{balanced accuracy} = \frac{0.5 \times \text{true positives}}{\text{true positives} + \text{false negatives}} + \frac{0.5 \times \text{true negatives}}{\text{true negatives} + \text{false positives}}$$



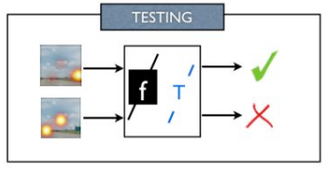
We compute a continuous **fixation map** by coarse binning and smoothing fixation locations (20x20 grid, sigma=2)



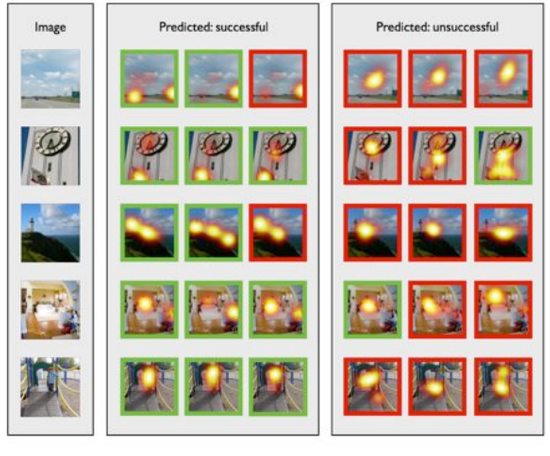
Just the task of classifying whether fixations come from this VS other image achieves 79.7% **balanced accuracy**.



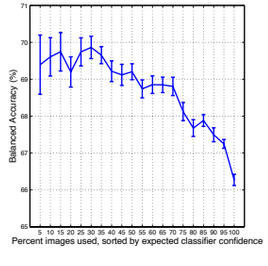
We learn a threshold under each image's **exemplar classifier** for differentiating successful from unsuccessful fixation maps on the image.



A set of fixations that are more likely to have been elicited by the current image as opposed to some other image are **predicted successful** encoding fixations that will lead to a correct recognition of the image.



Not all image are equally predictable
(successful should be distinguishable from unsuccessful fixations)



Applications: an automatic system that monitors the eye movements of a user to determine when to re-present visual material until there is confidence that the user has properly encoded the content.

For more robust performance, can select more predictable images for a given application.