What makes an object memorable?

Anonymous ICCV submission

Paper ID ****

Abstract

Recent work by Isola et. al. (2011) has demonstrated that memorability is an intrinsic property of images that is consistent across viewers and can be predicted accurately with current computer vision techniques. Despite progress, a clear understanding of the specific components of an image that drive memorability are still unknown. While previous studies such as Khosla et. al. (2012) have tried to investigate computationally the memorability of image regions within individual images, no behavioral study has systematically explored which memorability of image regions. Here we study which region from an image is memorable or forgettable. Using a large image database, we obtained the memorability scores of the different visual regions present in every image. In our task, participants viewed a series of images, each of which were displayed for 1.4 seconds. After the sequence was complete, participants similarly viewed a series of image regions and were asked to indicate whether each region was seen in the earlier sequence of full images.

1. Introduction

Consider the image and it's corresponding objects in Figure 1. Even though the person on the right is comparable in size to the left person, he is remembered far less by humans (indicated by their memorability scores of 0.18 and 0.64 respectively). People tend to remember the fish in the center and the person on the left, even after 30 minutes have passed (memorability score $= 0.64$). Interestingly, despite vibrant colors and considerable size, the boat is also remembered far less by humans (memorability $= 0.18$).

Just like aesthetics, interestingness, and other metrics of image importance, memorability quantifies something about the utility of a photograph toward our everyday lives. For many practical tasks, memorability is an especially desirable property to maximize. For example, this may be the case when creating educational materials, logos, advertisements, book covers, websites, and much more. Understanding memorability, and being able to automatically predict it, lends itself to a wide variety of applications in each of these areas.rewrite this to draw attention of reviewer to importance of image memorability. Due to this, automatic prediction of intrinsic memorability of images using computer vision and machine learning techniques has received considerable attention in the recent years [5], [6], [4], [2], [7]. While these studies have shed light on what distinguishes the memorability of different images and the intrinsic and extrinsic properties that make those images memorable, the above example raises an interesting question: what exactly about an image is remembered? Despite progress in the computer vision literature on image memorability, a clear understanding of the memorability of the specific components of an image is still unknown. For example, not all objects in an image will be equally remembered by people and as the figure 1 seems to suggest, there exists significant and interesting differences in memorability of objects in an image. Furthermore, the memorability of complex images may be principally driven by the memorability of it's objects. Can specific objects inside images be memorable to

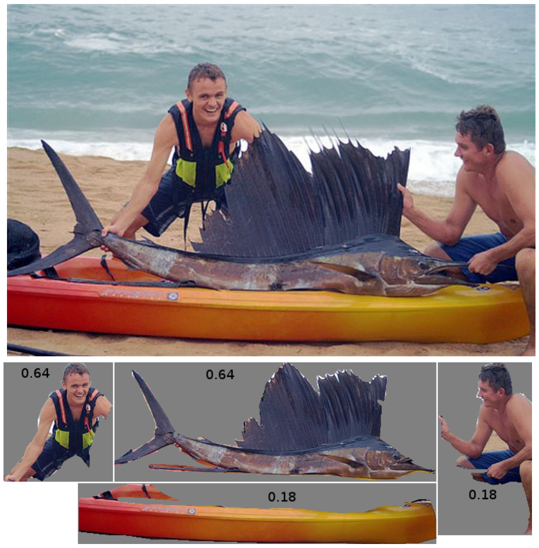

Figure 1: **Memorability of different objects.** Memorability scores of objects for the image in the top row obtained from our pysophysics experiment.

all us and how can we better understand what makes those objects more memorable?

In this paper, we systematically explore the memorability of objects within individual images and shed light on the various factors and properties that drive object memorability by augmenting both the images and object segmentations in the 850 existing images from PASCAL 2010 [3] dataset with memorability scores and class labels. By exploring the connection between object memorability, saliency, and image memorability, our paper makes several important contributions.

Firstly, we show that just like image memorability, object memorability is a property that is shared across subjects and objects remembered by one person are also likely to be remembered by others and vice versa. Secondly, we show that there exists a strong correlation between visual saliency and object memorability and demonstrate insights when can visual saliency directly predict object memorability and when does it fail to do so. While there have been have a few studies that explore the connection between image memorability and visual saliency [2], [11], our work is the first to explore the connection between object memorability and visual saliency. Third, we explore the connection between image memorability and object memorability and show that the most memorable object inside an image can be a strong predictor of image memorability in certain cases. Studying these questions, help not only understand visual saliency, image and object memorability in more detail, but it can also have important contributions to computer vision. For example, understanding which regions and objects in an image are memorable would enable us to modify the memorability of images which can have applications in advertising, user interface design etc. With this in mind, as shown in the section 4, our proposed dataset serves as a benchmark for evaluating object memorability model algorithms and can help usher in future algorithms that try to predict memorability maps.

1.1. Related works

Image Memorability: Describe Isola's first paper n some insights that have been raised on image memorability thus far. Also describe Khosla's comp model but we are the first work to actually describe what humans actually remember and don't

Visual Saliency: Talk about visual attention and models that have been proposed. Also, talk about Pascal-S and how it has helped reduce dataset bias

Saliency and memorability: discuss some results related to saliency and image memorability.

and talk about our work plans on connecting and shedding light on all these phenomena together.

2. Measuring Object Memorability

As a first step towards understanding memorability of objects, we built an image database containing a variety of objects from a diverse range of categories, and measured the probability that every object in each image will be remembered by a large group of subjects after a single viewing. This helps provide ground truth memorability scores for the objects inside the images and allows for a precise analysis of the memorable elements within an image. For this task, we utilized the PASCAL-S dataset [8], a fully segmented dataset built on the validation set of the PASCAL VOC 2010 [3] segmentation challenge. For improved segmentation purposes, we manually cleaned up and refined the segmentations from this dataset. While building the improved ground-truth of full segmentation, we removed all homogenous non-object or background segments such as ground, grass, floor, sky etc, as well as imperceptible object fragments and excessively blurred regions. All remaining object segmentations were tested for memorability. In the end, our final dataset consisted of 850 images and 3414 object segmentations i.e. on average each image consisted of approximately 4 segments for which we gathered the ground truth memorability on.

2.1. Memory Game: Measuring Object Memorability

To measure the memorability of individual objects from our dataset, we created an alternate version of the Visual Memory Game following the basic design in [5], with the exception of a few key differences. We administered the game and collected data through Amazon Mechanical Turk. In our game, participants first viewed a sequence of images one at a time, with a 1.5 second gap in between image presentations. Subjects were asked to remember the contents and objects inside those images as much as they could. To ensure that subjects would not just only look at the salient or center objects, subjects had unlimited time to freely view the images. Once they were done viewing an image, they could press any key to advance to the next image. Following the initial image sequence, participants then viewed a sequence of objects, their task then being to indicate through a key press which of those objects was present in one of the previously shown images. Each object was displayed for 1.5 second, with a 1.5 second gap in between the object sequences. Pairs of corresponding image and object sequences were broken up into 10 blocks. Each block consisted of 80 total stimuli (35 images and 45 objects), and lasted approximately 3 minutes. At the end of each block, the subject could take a short break. Overall, the experiment took approximately took 30 minutes to complete.

Unknown to the subjects, inside each block, each sequence of images was pseudo-random and consisted of 3 'target' images taken from the Pascal-S dataset whose ob-

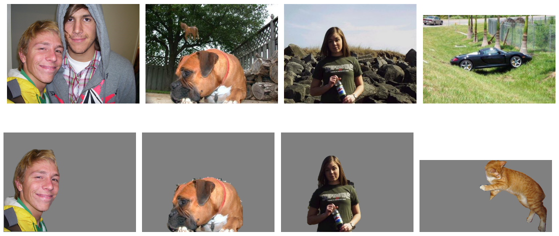

Figure 2: **Example stimuli from the memory game.** Example.

jects the participants were to later identify. The remaining images in the sequence consisted of 16 'filler' images and 16 'familiar' images. The 'filler' images were randomly selected from the DUT-OMRON dataset [12] and the 'familiar' images were randomly sampled from the MSRA dataset proposed in [10]. Similarly, the object sequence was also pseudo-random and consisted of 3 'target' objects (1 object taken randomly from each previously shown target image). The remaining objects in the sequence consisted of 10 'control' objects, 16 'filler' objects, and 16 'familiar' objects. The 'filler' objects were taken randomly from the 80 different object categories in the Microsoft COCO dataset [9] and the 'familiar' objects were the objects taken from the previously displayed 'familiar' images in the image sequence. The filler images and objects helped provide spacing between the target images and target objects, whereas the familiar images and objects allowed us to check if the subjects were paying attention to the task [1], [5]. While the fillers and familiars (both the images and objects) were taken from datasets resembling real world scenes and objects, the 'control' objects were artificial stimuli randomly sampled from the dataset proposed in [1] and helped serve as an additional control to test the attentiveness of the subjects. Repeats on targets i.e. the target images and target objects were spaced $70-79$ stimuli apart, and repeats occurred on the familiars with a spacing of $1-79$ stimuli (i.e. distance between familiar images and their objects). The images and objects appeared only once, and each subject was tested on only one object from each target image. Objects were centered within their parent frame and non-object pixels were set to grey. Participants were required to complete the entire task, which included 10 blocks (overall time approximately 30 minutes), and could not participate in the experiment a second time. After collecting the data, we assigned a 'memorability score' to each target object in our dataset, defined as the percentage of correct detections by subjects. In all our analysis, we removed all subjects whose accuracy on the control objects was below 75% and below 50% on familiar images/objects. A total of 2000 workers from Mechanical Turk ($>95\%$ approval rate in Amazons system) performed the game and on average each object was scored by 20 subjects.

2.2. Is object memorability a shared property across subjects?

Previous work on image memorability has found human consistency to be fairly high. That is, people tend to remember the same images, and exhibit similar performance in doing so. Despite variability due to individual differences and other sources of noise, this level of consistency provides evidence that memorability is an intrinsic property of images that can be predicted. In contrast to full images, this paper focuses primarily on the memorability of individual objects in an image, which may or may not exhibit the same level of human consistency as full images, which often contain complex arrangements of several objects. High consistency in object memorability would indicate that, like full images, objects can potentially be predicted with high accuracy. To assess human consistency in remembering objects, we repeatedly divided our entire subject pool into two equal halves and quantified the degree to which memorability scores for the two sets of subjects were in agreement using Spearmans rank correlation (). We computed the average correlation over 25 of these random split iterations, yielding a final value of 0.76. Such a result confirms that human consistency in remembering particular objects is at least as strong as that of images.

References

[1] T. F. Brady, T. Konkle, G. A. Alvarez, and A. Oliva. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38):14325–14329, 2008. 3

[2] Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and A. Oliva. Intrinsic and extrinsic effects on image memorability. *Vision research*, 2015. 1, 2

[3] M. Everingham and J. Winn. The pascal visual object classes challenge 2010 (voc2010) development kit, 2010. 2

[4] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What makes a photograph memorable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(7):1469–1482, 2014. 1

[5] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 145–152. IEEE, 2011. 1, 2, 3

[6] A. Khosla, J. Xiao, A. Torralba, and A. Oliva. Memorability of image regions. In *Advances in Neural Information Processing Systems*, pages 305–313, 2012. 1

[7] J. Kim, S. Yoon, and V. Pavlovic. Relative spatial features for image memorability. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 761–764. ACM, 2013. 1

[8] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 280–287. IEEE, 2014. 2

[9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014*, pages 740–755. Springer, 2014. 3

[10] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(2):353–367, 2011. 3

[11] M. Mancas and O. Le Meur. Memorability of natural scenes: the role of attention. In *ICIP*, 2013. 2

[12] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3166–3173. IEEE, 2013. 3