

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

# 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 its 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).

NOTE: Big picture way to think about this (rough): A currently massive goal in computer science and artificial intelligence is to model the data driven inference processes that humans make use of for solving vision problems, such as object recognition and scene understanding. A subset of these processes make use of nearly all of the information in the scene, much of which is used implicitly but others are dependent on more explicitly remembered information in



Figure 1: Not all objects are equally remembered. Memorability scores of objects for the image in the top row obtained from our psychophysics experiment.

a scene, that is, the information that the brain has deemed worthy to create special storage for. For example, the recognition of edges in a scene is learned and used automatically, but if we ask a person to report the important information in a scene, only a few high level things are reported. More specifically, humans can only make judgements about elements of an image that they can remember...

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 [15], [19], [14], [7], [20]. 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

108 and interesting differences in memorability of objects in an  
109 image. Furthermore, the memorability of complex images  
110 may be principally driven by the memorability of its objects.  
111 Can specific objects inside images be memorable to  
112 all us and how can we better understand what makes those  
113 objects more memorable?  
114

115 In this paper, we systematically explore the memorability  
116 of objects within individual images and shed light on  
117 the various factors and properties that drive object mem-  
118 orability by augmenting both the images and object seg-  
119 mentations in the 850 existing images from PASCAL 2010  
120 [11] dataset with memorability scores and class labels.  
121 By exploring the connection between object memorability,  
122 saliency, and image memorability, our paper makes several  
123 important contributions.  
124

125 Firstly, we show that just like image memorability, ob-  
126 ject memorability is a property that is shared across sub-  
127 jects and objects remembered by one person are also likely  
128 to be remembered by others and vice versa. Secondly, we  
129 show that there exists a strong correlation between visual  
130 saliency and object memorability and demonstrate insights  
131 when can visual saliency directly predict object memorabil-  
132 ity and when does it fail to do so. While there have been  
133 have a few studies that explore the connection between im-  
134 age memorability and visual saliency [7], [27], our work is  
135 the first to explore the connection between object mem-  
136 orability and visual saliency. Third, we explore the connec-  
137 tion between image memorability and object memorability  
138 and show that the most memorable object inside an image  
139 can be a strong predictor of image memorability in certain  
140 cases. Studying these questions, help not only understand  
141 visual saliency, image and object memorability in more de-  
142 tail, but it can also have important contributions to computer  
143 vision. For example, understanding which regions and ob-  
144 jects in an image are memorable would enable us to modify  
145 the memorability of images which can have applications in  
146 advertising, user interface design etc. With this in mind,  
147 as shown in the section 4, our proposed dataset serves as  
148 a benchmark for evaluating object memorability model al-  
149 gorithms and can help usher in future algorithms that try to  
150 predict memorability maps.  
151

## 1.1. Related works

152 In this section, we briefly discuss existing work related  
153 to visual memory and image memorability. We also review  
154 research related to visual saliency prediction and discuss the  
155 relationship of memorability and visual attention.  
156

157 **Image Memorability:** Describe Isola's first paper n  
158 some insights that have been raised on image memorabil-  
159 ity thus far. Also describe Khosla's comp model but we are  
160 the first work to actually describe what humans actually re-  
161 member and don't  
162

163 **Visual Saliency:** Talk about visual attention and models  
164

165 that have been proposed. Also, talk about Pascal-S and how  
166 it has helped reduce dataset bias  
167

168 **Saliency and memorability:** discuss some results re-  
169 lated to saliency and image memorability.  
170

171 and talk about our work plans on connecting and shed-  
172 ding light on all these phenomena together.  
173

## 2. Measuring Object Memorability

174 As a first step towards understanding memorability of  
175 objects, we built an image database containing a variety of  
176 objects from a diverse range of categories, and measured the  
177 probability that every object in each image will be remem-  
178 bered by a large group of subjects after a single viewing.  
179 This helps provide ground truth memorability scores for the  
180 objects inside the images and allows for a precise analysis  
181 of the memorable elements within an image. For this task,  
182 we utilized the PASCAL-S dataset [23], a fully segmented  
183 dataset built on the validation set of the PASCAL VOC 2010  
184 [11] segmentation challenge. For improved segmentation  
185 purposes, we manually cleaned up and refined the segmen-  
186 tations from this dataset. We removed all homogenous non-  
187 object or background segments such as ground, grass, floor,  
188 sky etc, as well as imperceptible object fragments and ex-  
189 cessively blurred regions. All remaining object segmen-  
190 tations were tested for memorability. In the end, our final  
191 dataset consisted of 850 images and 3412 object segmen-  
192 tations i.e. on average each image consisted of approximately  
193 4 object segments for which we gathered the ground truth  
194 memorability on.  
195

### 2.1. Object Memory Game

196 To measure the memorability of individual objects from  
197 our dataset, we created an alternate version of the Visual  
198 Memory Game through Amazon Mechanical Turk follow-  
199 ing the basic design in [15], with the exception of a few  
200 key differences. In our game, participants first viewed a  
201 sequence of images one at a time, with a 1.5 second gap in be-  
202 tween image presentations. Subjects were asked to remem-  
203 ber the contents and objects inside those images as much as  
204 they could. To ensure that subjects would not just only look  
205 at the salient or center objects, subjects had unlimited time  
206 to freely view the images. Once they were done viewing an  
207 image, they could press any key to advance to the next im-  
208 age. Following the initial image sequence, participants then  
209 viewed a sequence of objects, their task then being to indi-  
210 cate through a key press which of those objects was present  
211 in one of the previously shown images. Each object was  
212 displayed for 1.5 second, with a 1.5 second gap in between  
213 the object sequences. Pairs of corresponding image and ob-  
214 ject sequences were broken up into 10 blocks. Each block  
215 consisted of 80 total stimuli (35 images and 45 objects), and  
216 lasted approximately 3 minutes. At the end of each block,  
217  
218

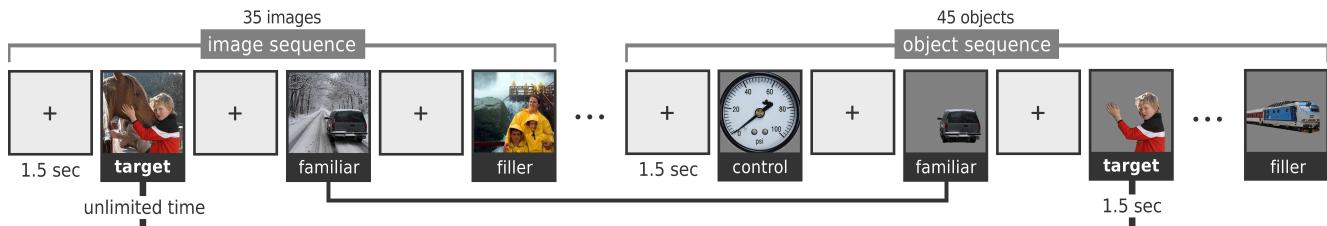


Figure 2: **Main task.** add-in later.

the subject could take a short break. Overall, the experiment took approximately 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 objects 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 [30] and the 'familiar' images were randomly sampled from the MSRA dataset proposed in [25]. 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 [24] and the 'familiar' objects were the objects taken from the previously displayed 'familiar' images in the image sequence. The fillers and familiars helped provide spacing between the target images and target objects, whereas the control objects allowed us to check if the subjects were paying attention to the task [5], [15]. 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 [5] and helped serve as a control to test the attentiveness of the subjects. The target images and the respective target objects were spaced 70 – 79 stimuli apart, and familiar images and their respective objects were spaced 1 – 79 stimuli apart. All 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 70%. In the end, our analysis was performed on a total of 1823 workers from Mechanical

Turk ( $> 95\%$  approval rate in Amazons system). The memorability score of an object corresponded to the number of subjects that correctly detected the repetition of that object. On average, each object was scored by 16 subjects and the average memorability score was 33% ( $SD = 28\%$ ).

## 2.2. Consistency Analysis

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 ( $\rho$ ). We computed the average correlation over 25 of these random split iterations, yielding a final value of 0.76. This high consistency in object memorability indicates that, like full images, object memorability is a shared property across subjects. People tend to remember (and forget) the same objects in images, and exhibit similar performance in doing so. Thus memorability of objects in images can potentially be predicted with high accuracy. In the next section, we study the various factors that possibly drive object memorability in images.

### 3. Understanding Object Memorability

In this section, we aim to better understand object memorability and the factors that make an object more memorable or forgettable to humans. We first investigate the role that simple color features play in determining object memorability.

### 3.1. Can simple features explain memorability?

While simple image features are traditionally poor predictors of memorability in full images [15], and with good reason [21], do they play any role in determining object memorability? We decomposed each image into its hue, saturation, and value components and calculated the mean and standard deviation of each channel. Mean value ( $\rho = 0.1$ ) and variance in value ( $\rho = 0.25$ ) were weakly correlated with object memorability suggesting that brighter and higher contrast objects may be more memorable (Figure 3). On the other hand, essentially no relationship was found between memorability and either hue or saturation (Figure 3). This deviates slightly from the findings in [15] that

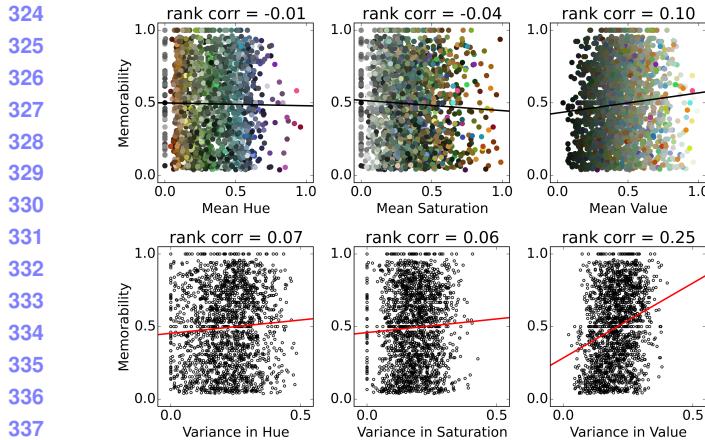


Figure 3: **Correlations between simple color features and object memorability.** Color features were computed from HSV representations of the objects.

showed mean hue to be weakly predictive of image memorability. However, this makes sense since the effect was speculated to be due to the blue and green outdoor landscapes being less memorable than warmly colored human faces and indoor scenes. While our dataset contained plenty of indoor objects and people, outdoor scene-related image regions such as sky and ground were not included as objects. Taken together, these results show that, like image memorability, basic pixel statistics do not play a significant role in determining the memorability of objects in images.

### 3.2. What is the role of saliency in memorability?

Intuitively, the regions within an image that are most salient are likely to have a higher probability of being remembered, since they will draw the attention of viewers and a majority of a viewer's eye fixations will be spent looking at those regions. On the other hand, it is conceivable that some visually appealing regions will not be memorable, especially since aesthetic images are known to be less memorable [15], [14]. When can visual saliency predict object memorability and what are the possible differences between these two phenomena? Quantifying the precise relationship between saliency and memorability will be paramount towards understanding object memorability in greater depth.

To this aim, we utilized the eye fixation dataset made available for the Pascal-S dataset in [23]. With this dataset in hand, we first calculated the number of unique fixation points within the area of each object and computed the correlation between this metric and the object's memorability score (Figure 5 a). We found this correlation to be positive and considerably high ( $\rho = 0.71$ ), suggesting that fixation count and visual saliency may drive object memorability considerably. However, the large concentration of points on the bottom left part of scatter plot in Figure 5 a suggests that part of the reason for this high correlation is that ob-

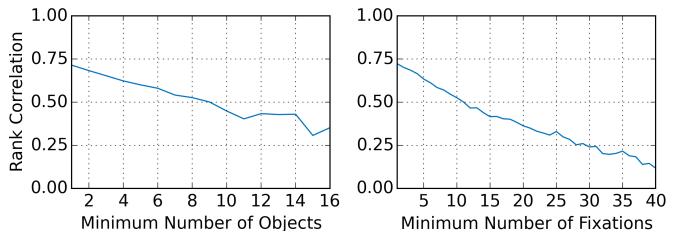


Figure 4: **Correlation between object memorability and number of fixations.** add-in later.

jects that have not been viewed at all have essentially no memorability. Indeed, only objects that have been seen can be remembered. In addition, the points toward the top left appear to decrease in trend. Looking deeper, Figure 4 plots the change in correlation between object memorability and fixations as the minimum number of fixations inside objects increases. The downward monotonic trend indicates that as the number of fixations inside an object increases, the predictive ability diminishes significantly. In addition, Figure 4 plots the correlation between object memorability and number of fixations as a function of total number of objects in an image. Similar to the previous trend, as the number of objects in an image increases, the correlation between saliency i.e. number of fixations and memorability score decreases sharply. This finding is in agreement with the two remaining scatter plots in Figure 5 b (shows that the memorability of an object decreases in the presence of many other objects) and Figure 5 c (shows that number of fixations decreases with the number of objects). This makes intuitive sense since people have more to look at in an image when more objects are present, and so they may look less at any one object, especially if they compete for saliency, and therefore may have a more difficult time remembering those objects.

To sum up, saliency is a surprisingly good index of object memorability in simple contexts where there are few objects in the image, or when an object has few interesting points, but it is a much weaker predictor of object memorability in complex scenes containing multiple objects that have many points of interest (Figure 7).

**Center Bias:** Figure 6 elucidates another example

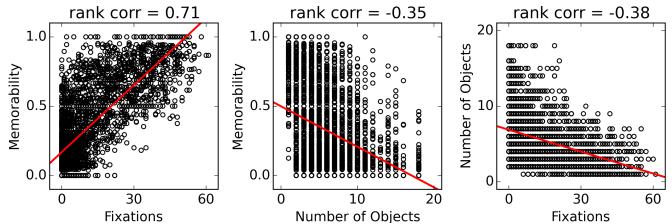
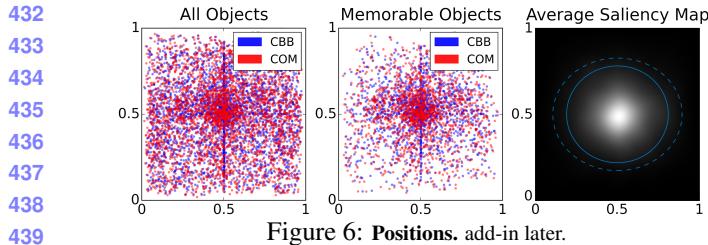
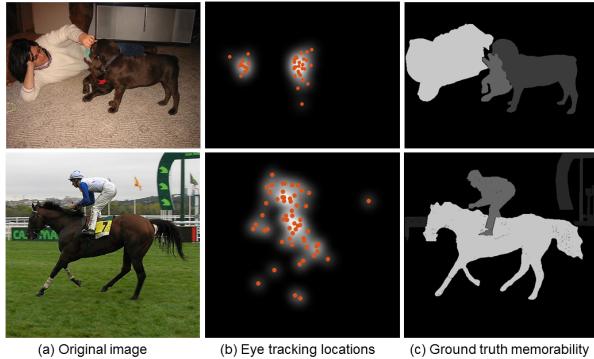


Figure 5: **Correlation between object memorability and number of fixations.** add-in later.

324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431

Figure 6: **Positions.** add-in later.Figure 7: **Saliency Fail cases.** add-in later.

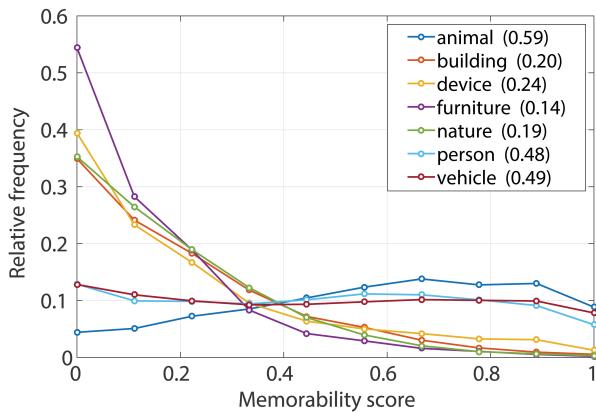
where saliency and memorability diverge. While previous studies related to visual saliency have showed that saliency is heavily influenced by center bias [18], [31], Figure 6 shows that memorability exhibits comparatively less center bias than saliency. This is most apparent when considering the difference in the solid ellipse in the right plot (shows where 95% of fixation positions are located), and the dashed ellipse (shows where the 95% of the above-median memorable objects are located).

### 3.3. How do object categories affect memorability?

In the previous sections, we showed that simple features have little predictive power over object memorability and explored the relationship between visual saliency and object memorability. In this section, we explore how the category of an object influences the probability that the object will be remembered.

#### 3.3.1 Are some classes more memorable than others?

For this analysis, we first assigned three in-house annotators the task of assigning class labels to each object segmentation in our dataset. The annotators were given the original image (for reference) and the object segmentation and asked to assign a single category to the segment out of 7 possible categories: animal, building, device, furniture, nature, person, and vehicle. We choose these high-level categories such that a wide range of object classes could be covered under these categories. For example, device included object segments such as utensils, bottles, televisions, computers etc, nature included segments like trees, mountains, flow-

Figure 9: **Average memorability per object class.** Figure showing some object classes are more memorable than others.

ers, and vehicle contained segments like cars, bikes, buses, airplanes etc.

Figure 9 shows the distribution of the memorability scores for all 7 object classes in our dataset. This visualisation gives a sense of how the memorability changes across different object categories. Animal, person, and vehicle are all highly memorable classes each associated with an average memorability score greater than or close to 0.5. Interestingly, all other object categories have an average memorability score lower than 0.25, indicating that humans do not remember objects from these categories very well. In particular, furniture is the least memorable object class with an average memorability score of only 0.14. This could be possibly due to the fact that most objects from classes like furniture, nature, and building either appear mostly in the background or are occluded which likely decreases their memorability significantly. By contrast, objects from the animal, person, and vehicle classes appear mostly in the foreground, leading to a higher memorability score on average. Interestingly, the topmost memorable objects from building, furniture, and nature tend to have an average memorability score in the range of 0.4 – 0.8, whereas the topmost memorable objects from classes person, animal and vehicle have an average memorability higher than 0.90. This is particularly interesting as these top objects are not occluded and most of them tend to appear in the foreground. While the differences in the memorability of different classes could be driven primarily due to factors like occlusion, size, background/foreground, or photographic bias, the distribution in figure 9 suggests that humans remember some object classes such as person, animal, and vehicle irrespective of external nuisance factors and these object classes are *intrinsically* more memorable than others.

486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539

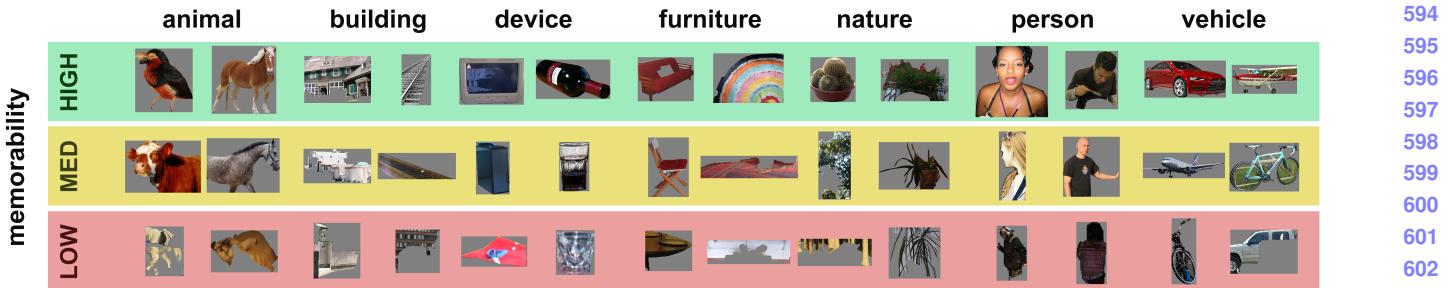


Figure 8: Qualitative results from object categories. add-in later.

### 3.3.2 Why are some objects not memorable in a class?

As demonstrated above, some object classes (i.e. animal, person, vehicle) are more memorable than others. However, not all objects in a class are equally memorable. The examples in Figure 8 show the most memorable, medium memorable, and least memorable objects for each object class. Across classes, non-memorable objects tend to be those that are occluded and obstructed by other objects. What other possible factors could influence the memorability of an object within a class? Among the various possible factors, we explored how object memorability within a particular class is influenced by a) the number of objects in an image and b) the presence of other object classes.

**Number of objects:** We first examined how the memorability of each object class is affected by the number of objects inside an image. Figure 10 shows the change in average memorability of the different object classes with respect to the increase in number of objects in an image. Results indicate that the number of objects present in an image is an important factor in determining memorability. For example, as the number of objects in an image increases, the memorability of animals and vehicles decreases sharply, likely as a result of competition for attention, or decreased spotlight on a single subject of the composition. Interestingly, the memorability of the person class does not change significantly with an increase in number of objects. This suggests that people are not only one of the most memorable object classes, but are also more robust to the presence of clutter in images. This may be because single people in images steal all of the attention of the viewer, but how do they behave in the presence of other people? To answer this, we turn to the question of interclass memorability next.

**Inter-class memorability:** How does the presence of a particular object class influence the memorability of another object class? For all pairwise combinations of object category, we gathered all images that contained at least one object from both categories and computed the change in the average memorability scores for the two object categories. Figure 11 plots these data and visualizes how the memorability of each object class is affected by the pres-

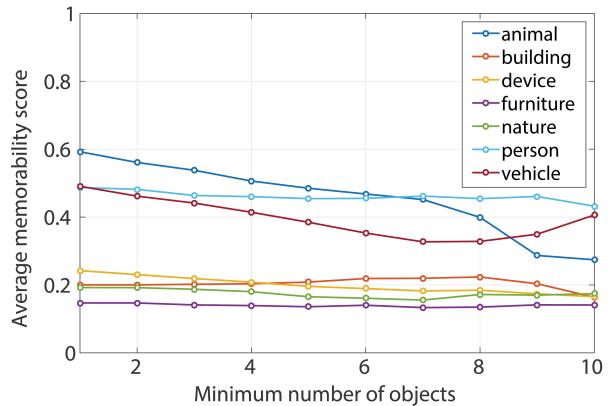


Figure 10: Correlation between object class and number of objects. add-in later.

ence other object classes. The first thing to note is that the values of low-memorability classes (i.e. nature, furniture, device, and building) are not greatly affected by the presence of other object categories. Instead, their memorability tends to remain low across all contexts. The memorability of the animal class remains close to its high average memorability score in presence of most classes, but drops significantly in the presence of other animals, vehicles, and people. The memorability of people also remains close to its average memorability score and tends to be unaffected by the presence of most object categories (including other people). However, the memorability of a person decreases in the presence of vehicles and buildings. This could be due to the fact that people in images containing vehicles or buildings are usually zoomed out and are usually smaller in size (also illustrated in figure 12). The memorability of the vehicle class is strongly affected by the presence of other object categories. In particular, its memorability drops significantly in the presence of another vehicle, people, and animals. Taken together, when an animal, vehicle or a person is present in the same image, the memorability of all three classes usually goes down. However, this pattern of change in memorability varies by class, leading to interesting results. For example, when a vehicle and animal are present in the same image, the animal is generally more

648  
649  
650  
651  
652  
memorable, even though the memorability of both of these  
653 classes drops significantly. When a vehicle or an animal  
654 co-occurs with a person, the person is generally more mem-  
655 orable (also shown in Figure 12).

### 3.4. How are object & image memorability related?

We now know what objects people remember and the factors that influence their memorability, but to what extent does the memorability of individual objects affect the overall memorability of a scene? Moreover, if an image is highly memorable, what can we say about the memorability of the objects inside those images (and vice versa)? To shed light on this question, we conducted a second large-scale experiment on Amazon Mechanical Turk for all images in our dataset to gather their respective image memorability scores. For this experiment, we followed the exact paradigm as the memory game experiment proposed in [15]. A series of images from our dataset and Microsoft COCO dataset [24] (i.e. the 'filler' images) were flashed for 1 second each, and subjects were instructed to press a key whenever they detected a repeat presentation of an image. A total of 350 workers participated in this experiment with each image being viewed 80 times on average. The rank correlation after averaging over 25 random split half trials was found to be 0.70, providing evidence for consistency in the image memorability scores.

Utilizing results from both experiments, we computed the correlation between the scores of the single most memorable object in each image (from Experiment 1) and the overall memorability score of each image (from Experiment 2). We found this correlation to be moderately high ( $\rho = 0.4$ ), suggesting that the most highly memorable object in the image plays a crucial role in determining the overall memorability of an image. To investigate this finding in relation to extreme cases only, we performed the same analysis as above on a subset of the data containing the topmost 100 memorable images and the bottommost 100 memorable

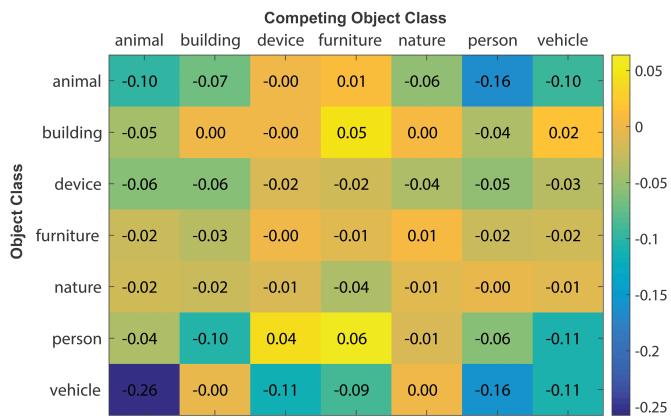


Figure 11: inter-class object memorability relationship. add-in later.



Figure 12: Qual Results. Figure showing how memorability of different classes is effected in presence of other classes. Bottom row is the memorability map

| Animal | Building | Device | Furniture | Nature | Person | Vehicle | All  |
|--------|----------|--------|-----------|--------|--------|---------|------|
| 0.38   | 0.22     | 0.47   | 0.53      | 0.64   | 0.54   | 0.30    | 0.40 |

Table 1: Max object memorability and image memorability. add-in later.

images. The correlation between maximum object memorability and image memorability for this subset of the images increased significantly ( $\rho = 0.62$ ), meaning maximum object memorability serves as a strong indicator of whether an image is *highly* memorable or non-memorable. That is, images that are highly memorable contain at least one highly memorable object, and images with low memorability usually do not contain a single highly memorable object (also shown in Figure 13).

It seems that maximum object memorability is highly explanatory, but does this behavior generalize across object classes? We further computed the correlation between maximum object and image memorability for each individual object class. Results shown in Table 1 show that certain object classes are more strongly correlated than others. For example, images containing animals, buildings, or vehicles as the most memorable objects tend to have varying degree of image memorability (indicated by their lower  $\rho$  values). On the other hand, classes like device, furniture, nature, and person are strongly correlated with image memorability, indicating that if an image's most memorable object belongs to one of these classes, the object memorability score is strongly predictive of the image memorability score. We can imagine scenarios in which this information would be potentially useful. For example, in the case where vision systems are tasked to predict scene memorability, a *single* object and its class can serve as a strong prior in predicting image memorability.

## 4. Predicting Object Memorability

Although the primary goal of our paper is to *understand* what drives memorability of objects in a scene, the current work also makes available the very first dataset containing the ground truth memorability of constituent objects from a highly diverse image set. In this section, we show that our dataset can serve as a benchmark for the purpose of object memorability prediction by making use of the ground truth

756 memorability maps constructed from experimental data.  
 757

758 **Baseline models:** As a first step, we propose a simple  
 759 baseline model that utilizes a conv-net [22], [17] trained on  
 760 the ImageNet database [10]. Since object categories play an  
 761 important role in determining object memorability (??), and  
 762 deep learning models have recently been shown to achieve  
 763 state-of-the-art results in various other recognition tasks, in-  
 764 cluding object recognition and object categorization (cite  
 765 papers here), we believe that this simple model can serve  
 766 as a good initial baseline for object memorability predic-  
 767 tion. We first generated object segments by using MCG, a  
 768 generic object proposal method proposed in [2]. Next, we  
 769 trained an SVR using 6-fold cross-validation on the origi-  
 770 nal segments to map deep features to memorability scores.  
 771 We then used this model to predict the memorability scores  
 772 for the top  $K$  ( $K = 20$ ) segments (obtained via the ranking  
 773 scores provided by the MCG algorithm) for each image. Af-  
 774 ter obtaining the predicted memorability scores, the mem-  
 775 orability maps were generated by averaging the top  $K$  seg-  
 776 ments at the pixel level. Since image features like SIFT [26]  
 777 and HOG [9] have previously been shown to achieve good  
 778 performance in predicting image memorability, we built a  
 779 second baseline model using these features for comparison.  
 780 Training and testing of this model was performed similar to  
 781 the deep-net baseline model.

782 **Evaluation:** To evaluate the accuracy of the predicted  
 783 memorability maps, we computed the rank correlation be-  
 784 tween the mean predicted memorability score inside each of  
 785 the original object segments and their ground truth memo-  
 786 rability scores. From figure 14, we first note that our deep-  
 787 net baseline model performs considerably well ( $\rho = 0.39$ ),  
 788 reaching a level of performance that approaches that of hu-  
 789 man consistency. In contrast, the baseline model trained  
 790 using HOG and SIFT features exhibits much lower over-  
 791 all performance ( $\rho = 0.27$ ). We also included 8 state-  
 792 of-the-art-saliency methods **gb** [12], **aim** [6], **dv** [13], **it**  
 793 [16], **gc** [8], **pc** [28], **sf** [29], and **ft** [1] to our com-  
 794 parison (some of these are the top performing methods ac-  
 795 cording to benchmarks in [4], [3]). Saliency maps gener-  
 796 ated from these methods are likely to have some degree of  
 797 overlap with memorability and are therefore worth compar-  
 798 ing to our baseline, especially given the absence of alterna-  
 799 tive memorability prediction methods. Results from figure  
 800 14) show that the HOG+SIFT baseline is outperformed by  
 801 most saliency methods. Thus, even though models using  
 802 these features have previously demonstrated high predic-  
 803 tive power in predicting image memorability, they may not  
 804 be as well suited for the task of predicting object mem-  
 805 orability. The deep-net baseline model performs better than  
 806 all other saliency methods and only **pc** ( $\rho = 0.38$ ), **sf**  
 807 ( $\rho = 0.37$ ), and **gb** ( $\rho = 0.36$ ) show performance  
 808 comparable to the model. A common factor between these  
 809 saliency methods is that they explicitly add center bias to

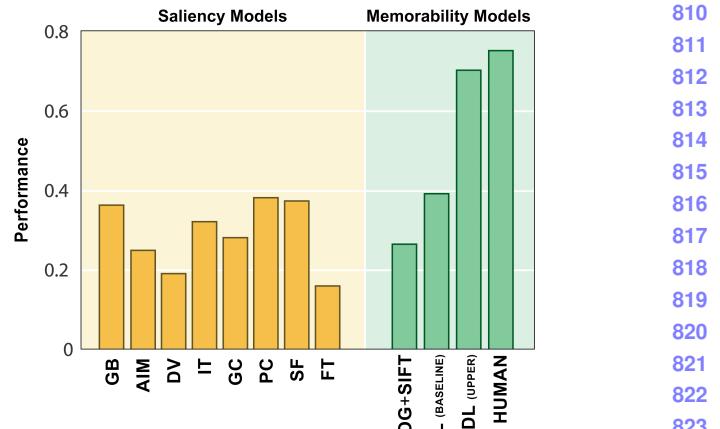


Figure 14: **Main task.** add-in later.

810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841  
 842  
 843  
 844  
 845  
 846  
 847  
 848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863

their implementation, which could be a part of the reason  
 for their favorable performance. Despite this, the deep-net  
 model performs comparably against them and is potentially  
 much better suited for memorability prediction. Thus, we  
 recommend in the future, memorability algorithms com-  
 pare their methods against our deep-net baseline. While  
 the deep-net model performed fairly well, part of the per-  
 formance of such a model is dependent on the quality of  
 the segmentations used. For this reason, we also consider  
 the upper bound of our current predictive power by show-  
 ing the results for our model containing predictions on the  
 original segments (Figure 14). Interestingly, the accuracy  
 of this model is very high and close to human performance  
 $(\rho = 0.7)$ . This demonstrates that the deep-net model has  
 high predictive ability that is suppressed most heavily by  
 constraints of the segmentation task. As the main insight of  
 our evaluation, we have demonstrated that deep fea-  
 tures serve as strong predictors of memorability and selec-  
 tion of higher quality segments can potentially lead to im-  
 proved memorability prediction algorithms.

## References

- [1] R. Achanta, S. Hemami, F. Estrada, and S. Sussstrunk. Frequency-tuned salient region detection. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pages 1597–1604. IEEE, 2009. 8
- [2] P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 328–335. IEEE, 2014. 8
- [3] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In *Computer Vision–ECCV 2012*, pages 414–429. Springer, 2012. 8
- [4] A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *Image Processing, IEEE Transactions on*, 22(1):55–69, 2013. 8

- 864 [5] T. F. Brady, T. Konkle, G. A. Alvarez, and A. Oliva. Visual  
865 long-term memory has a massive storage capacity for object  
866 details. *Proceedings of the National Academy of Sciences*,  
867 105(38):14325–14329, 2008. 3  
868 [6] N. Bruce and J. Tsotsos. Saliency based on information max-  
869 imization. In *Advances in neural information processing systems*,  
870 pages 155–162, 2005. 8  
871 [7] Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and  
872 A. Oliva. Intrinsic and extrinsic effects on image memorability.  
873 *Vision research*, 2015. 1, 2  
874 [8] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and  
875 S.-M. Hu. Global contrast based salient region detection.  
876 In *Computer Vision and Pattern Recognition (CVPR), 2011*  
877 *IEEE Conference on*, pages 409–416. IEEE, 2011. 8  
878 [9] N. Dalal and B. Triggs. Histograms of oriented gradients for  
879 human detection. In *Computer Vision and Pattern Recog-  
880 nition, 2005. CVPR 2005. IEEE Computer Society Conference  
881 on*, volume 1, pages 886–893. IEEE, 2005. 8  
882 [10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-  
883 Fei. Imagenet: A large-scale hierarchical image database.  
884 In *Computer Vision and Pattern Recognition, 2009. CVPR  
885 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 8  
886 [11] M. Everingham and J. Winn. The pascal visual object classes  
887 challenge 2010 (voc2010) development kit, 2010. 2  
888 [12] J. Harel, C. Koch, and P. Perona. Graph-based visual  
889 saliency. In *Advances in neural information processing systems*,  
890 pages 545–552, 2006. 8  
891 [13] X. Hou and L. Zhang. Dynamic visual attention: Searching  
892 for coding length increments. In *Advances in neural infor-  
893 mation processing systems*, pages 681–688, 2009. 8  
894 [14] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What  
895 makes a photograph memorable? *Pattern Analysis and Ma-  
896 chine Intelligence, IEEE Transactions on*, 36(7):1469–1482,  
897 2014. 1, 4  
898 [15] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an  
899 image memorable? In *Computer Vision and Pattern Recog-  
900 nition (CVPR), 2011 IEEE Conference on*, pages 145–152.  
901 IEEE, 2011. 1, 2, 3, 4, 7  
902 [16] L. Itti, C. Koch, and E. Niebur. A model of saliency-based  
903 visual attention for rapid scene analysis. *IEEE Transactions  
904 on pattern analysis and machine intelligence*, 20(11):1254–  
905 1259, 1998. 8  
906 [17] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Gir-  
907 shick, S. Guadarrama, and T. Darrell. Caffe: Convolutional  
908 architecture for fast feature embedding. In *Proceedings of  
909 the ACM International Conference on Multimedia*, pages  
910 675–678. ACM, 2014. 8  
911 [18] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to  
912 predict where humans look. In *Computer Vision, 2009 IEEE  
913 12th international conference on*, pages 2106–2113. IEEE,  
914 2009. 5  
915 [19] A. Khosla, J. Xiao, A. Torralba, and A. Oliva. Memorability  
916 of image regions. In *Advances in Neural Information Pro-  
917 cessing Systems*, pages 305–313, 2012. 1  
918 [20] J. Kim, S. Yoon, and V. Pavlovic. Relative spatial features for  
919 image memorability. In *Proceedings of the 21st ACM inter-  
920 national conference on Multimedia*, pages 761–764. ACM,  
921 2013. 1  
922 [21] T. Konkle, T. F. Brady, G. A. Alvarez, and A. Oliva. Concep-  
923 tual distinctiveness supports detailed visual long-term mem-  
924 ory for real-world objects. *Journal of Experimental Psychol-  
925 ogy: General*, 139(3):558, 2010. 3  
926 [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet  
927 classification with deep convolutional neural networks. In  
928 *Advances in neural information processing systems*, pages  
929 1097–1105, 2012. 8  
930 [23] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The  
931 secrets of salient object segmentation. In *Computer Vision  
932 and Pattern Recognition (CVPR), 2014 IEEE Conference on*,  
933 pages 280–287. IEEE, 2014. 2, 4  
934 [24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ra-  
935 manan, P. Dollár, and C. L. Zitnick. Microsoft coco: Com-  
936 mon objects in context. In *Computer Vision–ECCV 2014*,  
937 pages 740–755. Springer, 2014. 3, 7  
938 [25] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and  
939 H.-Y. Shum. Learning to detect a salient object. *Pattern  
940 Analysis and Machine Intelligence, IEEE Transactions on*,  
941 33(2):353–367, 2011. 3  
942 [26] D. G. Lowe. Distinctive image features from scale-  
943 invariant keypoints. *International journal of computer vi-  
944 sion*, 60(2):91–110, 2004. 8  
945 [27] M. Mancas and O. Le Meur. Memorability of natural scenes:  
946 the role of attention. In *ICIP*, 2013. 2  
947 [28] R. Margolin, A. Tal, and L. Zelnik-Manor. What makes a  
948 patch distinct? In *Computer Vision and Pattern Recog-  
949 nition (CVPR), 2013 IEEE Conference on*, pages 1139–1146.  
950 IEEE, 2013. 8  
951 [29] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung.  
952 Saliency filters: Contrast based filtering for salient region  
953 detection. In *Computer Vision and Pattern Recognition  
954 (CVPR), 2012 IEEE Conference on*, pages 733–740. IEEE,  
955 2012. 8  
956 [30] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang.  
957 Saliency detection via graph-based manifold ranking. In  
958 *Computer Vision and Pattern Recognition (CVPR), 2013*  
959 *IEEE Conference on*, pages 3166–3173. IEEE, 2013. 3  
960 [31] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cot-  
961 trell. Sun: A bayesian framework for saliency using natural  
962 statistics. *Journal of vision*, 8(7):32, 2008. 5  
963 [32] J. Zitnick, M. Uyttendaele, A. Efros, and R. Szeliski. Ef-  
964 fective multi-scale image cloning. In *Computer Vision and  
965 Pattern Recognition (CVPR), 2008 IEEE Conference on*,  
966 pages 1–8. IEEE, 2008. 5  
967 [33] J. Zitnick, M. Uyttendaele, A. Efros, and R. Szeliski. Ef-  
968 fective multi-scale image cloning. *IEEE Transactions on  
969 Pattern Analysis and Machine Intelligence*, 32(8):1422–  
970 1435, 2010. 5  
971

|      |                 |      |      |      |      |      |      |      |      |      |      |      |      |
|------|-----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| 972  |                 |      |      |      |      |      |      |      |      |      |      |      | 1026 |
| 973  |                 |      |      |      |      |      |      |      |      |      |      |      | 1027 |
| 974  |                 |      |      |      |      |      |      |      |      |      |      |      | 1028 |
| 975  |                 |      |      |      |      |      |      |      |      |      |      |      | 1029 |
| 976  |                 |      |      |      |      |      |      |      |      |      |      |      | 1030 |
| 977  |                 |      |      |      |      |      |      |      |      |      |      |      | 1031 |
| 978  |                 |      |      |      |      |      |      |      |      |      |      |      | 1032 |
| 979  |                 |      |      |      |      |      |      |      |      |      |      |      | 1033 |
| 980  |                 |      |      |      |      |      |      |      |      |      |      |      | 1034 |
| 981  |                 |      |      |      |      |      |      |      |      |      |      |      | 1035 |
| 982  |                 |      |      |      |      |      |      |      |      |      |      |      | 1036 |
| 983  |                 |      |      |      |      |      |      |      |      |      |      |      | 1037 |
| 984  |                 |      |      |      |      |      |      |      |      |      |      |      | 1038 |
| 985  |                 |      |      |      |      |      |      |      |      |      |      |      | 1039 |
| 986  |                 |      |      |      |      |      |      |      |      |      |      |      | 1040 |
| 987  |                 |      |      |      |      |      |      |      |      |      |      |      | 1041 |
| 988  |                 |      |      |      |      |      |      |      |      |      |      |      | 1042 |
| 989  |                 |      |      |      |      |      |      |      |      |      |      |      | 1043 |
| 990  |                 |      |      |      |      |      |      |      |      |      |      |      | 1044 |
| 991  |                 |      |      |      |      |      |      |      |      |      |      |      | 1045 |
| 992  |                 |      |      |      |      |      |      |      |      |      |      |      | 1046 |
| 993  |                 |      |      |      |      |      |      |      |      |      |      |      | 1047 |
| 994  |                 | 0.94 | 1.00 | 0.93 | 0.75 | 0.92 | 0.80 | 0.92 | 0.94 | 0.92 | 0.80 | 0.91 | 0.76 |
| 995  | Most memorable  |      |      |      |      |      |      |      |      |      |      |      |      |
| 996  |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 997  |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 998  |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 999  | Least memorable | 0.47 | 0.56 | 0.48 | 0.38 | 0.50 | 0.38 | 0.50 | 0.61 | 0.51 | 0.50 | 0.52 | 0.39 |
| 1000 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1001 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1002 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1003 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1004 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1005 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1006 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1007 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1008 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1009 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1010 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1011 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1012 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1013 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1014 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1015 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1016 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1017 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1018 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1019 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1020 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1021 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1022 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1023 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1024 |                 |      |      |      |      |      |      |      |      |      |      |      |      |
| 1025 |                 |      |      |      |      |      |      |      |      |      |      |      |      |

Figure 13: Qual image-object results. add-in later.