

000
001
002
003
004
005
006
007
008
009
010
011012
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
053

What makes an object memorable?

Anonymous ICCV submission

Paper ID ****

054
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

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 [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

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 In this paper, we systematically explore the memorability
115 of objects within individual images and shed light
116 on the various factors and properties that drive object
117 memorability by augmenting both the images and object
118 segmentations in the 850 existing images from PASCAL
119 2010 [3] dataset with memorability scores and class labels.
120 By exploring the connection between object memorability,
121 saliency, and image memorability, our paper makes several
122 important contributions.

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

149 1.1. Related works

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

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

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

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

162 2. Measuring Object Memorability

163 As a first step towards understanding memorability of
164 objects, we built an image database containing a variety of
165 objects from a diverse range of categories, and measured the
166 probability that every object in each image will be remembered
167 by a large group of subjects after a single viewing.
168 This helps provide ground truth memorability scores for the
169 objects inside the images and allows for a precise analysis
170 of the memorable elements within an image. For this task,
171 we utilized the PASCAL-S dataset [8], a fully segmented
172 dataset built on the validation set of the PASCAL VOC
173 2010 [3] segmentation challenge. For improved segmentation
174 purposes, we manually cleaned up and refined the segmentations
175 from this dataset. While building the improved
176 ground-truth of full segmentation, we removed all homogeneous
177 non-object or background segments such as ground,
178 grass, floor, sky etc, as well as imperceptible object
179 fragments and excessively blurred regions. All remaining object
180 segmentations were tested for memorability. In the end, our
181 final dataset consisted of 850 images and 3414 object
182 segmentations i.e. on average each image consisted of approx-
183 imately 4 segments for which we gathered the ground truth
184 memorability on.

185 2.1. Memory Game: Measuring Object Memorability

186 To measure the memorability of individual objects from
187 our dataset, we created an alternate version of the Visual
188 Memory Game following the basic design in [5], with the
189 exception of a few key differences. We administered the
190 game and collected data through Amazon Mechanical Turk.
191 In our game, participants first viewed a sequence of images
192 one at a time, with a 1.5 second gap in between image pre-
193 sentations. Subjects were asked to remember the contents
194 and objects inside those images as much as they could. To
195 ensure that subjects would not just only look at the salient
196 or center objects, subjects had unlimited time to freely view
197 the images. Once they were done viewing an image, they
198 could press any key to advance to the next image. Following
199 the initial image sequence, participants then viewed a se-
200 quence of objects, their task then being to indicate through
201 a key press which of those objects was present in one of
202 the previously shown images. Each object was displayed
203 for 1.5 second, with a 1.5 second gap in between the ob-
204 ject sequences. Pairs of corresponding image and object
205 sequences were broken up into 10 blocks. Each block con-
206 sisted of 80 total stimuli (35 images and 45 objects), and
207 lasted approximately 3 minutes. At the end of each block,
208 the subject could take a short break. Overall, the experiment
209 took approximately took 30 minutes to complete.



Figure 2: **Example stimuli from the memory game.** Example filler, familiar, and control images and objects.

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 [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 (ρ). 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.

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 if simple color and geometric factors like simple object features like color etc can predict if an object would be memorable or not.

3.1. Can simple color features explain object memorability?

While simple image features are traditionally poor predictors of memorability in full images [Oliva, CVPR 2011], and with good reason [cite [11] from Oliva, CVPR 2011]], it is important to verify that this finding generalizes to individual objects. To do this, we first examined a number of simple color statistics. We decomposed each image into its hue, saturation, and value components and calculated the mean and standard deviation of each channel. Essentially no relationship existed between memorability and either mean hue ($\rho = 0$) or mean saturation ($\rho = 0.02$). This deviates somewhat from the findings related to images that show hue to be weakly predictive of memorability. However, this makes sense since the effect has been speculated to be due to the blue and green outdoor landscapes being less memorable than warmly colored human faces and indoor

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

324
325
326
327
328
329
330
331
332
333
334
335

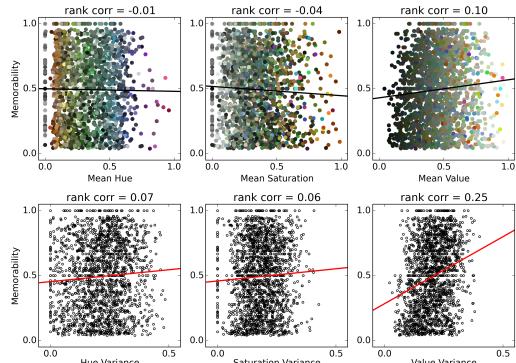


Figure 3: **Correlations between simple color features and object memorability.** Color features were computed from HSV representations of the objects.

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. This may explain why the effect was not present in our dataset. Looking at the variance of each channel, we see that variance in hue ($\rho = 0.16$) and saturation ($\rho = 0.15$) are weakly correlated with memorability, while value has a medium correlation ($\rho = 0.38$). This suggests that variation in the color hue and color purity of an object contribute somewhat to object memorability. This may be because people are memorable regardless of the color of their clothing, or simply that color inhomogeneity draws ones attention. The finding that variation in value contributes considerably to object memorability indicates that high contrast objects are more memorable. Although the correlation seems high compared to the performance of past simple features in past research, the scatterplot in FIGURE X suggests that it may not produce reliable predictions. Next, we computed image size, calculated by the number of pixels that make up the object normalized by the total number of pixels in the parent image. Not surprisingly, this metric correlated strongly with object memorability ($\rho = 0.53$). This makes sense given that large objects, especially that take up the majority of the parent image frame, are more likely to be seen and identified.

3.2. What is the relationship between visual saliency and object memorability?

Note: add in small intro about past work

To examine the relationship between visual saliency and object memorability, we used two indexes of saliency to describe each object. The first was calculated as the number of unique fixation points within the area of each object. The correlation between object memorability and number of fixations within those objects was quite high ($\rho = 0.71$). However, this correlation decreases significantly as the number of total fixations within the object increases. In figure X,

Figure 3 consists of three scatter plots arranged horizontally. The first plot, titled 'All Objects', shows a dense cloud of points with a color legend: blue for CBB and red for COM. The second plot, titled 'Memorable Objects', shows a similar distribution of points. The third plot, titled 'Fixation Positions', shows a distribution of points with a grayscale intensity gradient, indicating higher density in the center. All three plots have axes ranging from 0.0 to 1.0.

Figure 4: **Positions.** add-in later.

we show the correlation between object memorability and number of fixations for subsets of the objects that have at least some minimum number of fixations. On the leftmost side of the plot, the correlation is at its highest since the set of objects with a fixation count greater than or equal to zero includes all objects and so the correlation is identical to FIGURE FIX-CORR. However, as the minimum number of fixations increases, we look only at objects that contain an increasingly higher number of distinct fixation points, and the correlation decreases. By the time we reach a minimum fixation count of around 35, the correlation falls to around 0.2. This may mean that the predictive value of saliency decreases when there are many areas of interest within a single object. (NOTE: We may want to show examples of images with lots of fixation points that were not well predicted. We could just plot the fixation locations as points on the image) The reason for this decline may be that...[make arguments here] Toward the right end of the plot, few objects contain large numbers of fixations, and so the correlations become sporadic. The red line marks the point for which the correlations cease to be statistically significant, just at the end of the main trend line. Alternatively, FIGURE Z shows a similar relationship in which the correlation between memorability and fixation count is plotted as a function of minimum number of objects present in the parent image of the object. In a similar pattern, we see that as the minimum number of objects increases, the correlation between fixation count and object memorability decreases. This follows from the fact that fixation count and number of parent image objects are negatively correlated ($\rho = -0.38$), meaning that if there are more objects in an image, there are less fixation points in any one object, which may be a consequence of needing to distribute attention across the objects in the scene. In both scenarios, the ability to predict object memorability sharply decreases. ===== To examine the relationship between visual saliency and object memorability, we used two indexes of saliency to describe each object. The first was calculated as the number of unique fixation points within the area of each object. The correlation between object memorability and number of fixations within those objects was quite high ($\rho = 0.71$). However, this correlation decreases significantly as the number of total fixations within the object increases. In figure X, we show the correlation between object memorability and number of fixations for subsets of

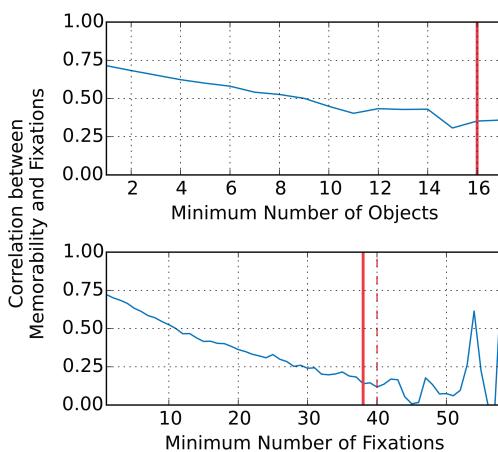


Figure 5: Correlation between object memorability and number of fixations. add-in later.

the objects that have at least some minimum number of fixations. On the leftmost side of the plot, the correlation is at it's highest since the set of objects with a fixation count greater than or equal to zero includes all objects and so the correlation is identical to FIGURE FIX-CORR. However, as the minimum number of fixations increases, we look only at objects that contain an increasingly higher number of distinct fixation points, and the correlation decreases. By the time we reach a minimum fixation count of around 35, the correlation falls to around 0.2. This may mean that the predictive value of saliency decreases when there are many areas of interest within a single object. (NOTE: We may want to show examples of images with lots of fixation points that were not well predicted. We could just plot the fixation locations as points on the image) The reason for this decline may be that...[make arguments here] Alternatively, FIGURE Z shows a similar relationship in which the correlation between memorability and fixation count is plotted as a function of minimum number of objects present in the parent image of the object. In a similar pattern, we see that as the minimum number of objects increases, the correlation between fixation count and object memorability decreases. This follows from the fact that fixation count and number of parent image objects are negatively correlated ($\rho = -0.38$), meaning that if there are more objects in an image, there are less fixation points in any one object, which may a consequence of needing to distribute attention across the objects in the scene. In both scenarios, the ability to predict object memorability sharply decreases.

The second metric we used... (put map method here?)

3.3. What is the relationship between object memorability and object categories?

In the previous sections, we showed that simple features have little predictive power over object memorability and

there exists a weak relationship between visual saliency and object memorability. In this section, we show that object memorability is heavily influenced by it's category and investigate the possible reasons behind the same.

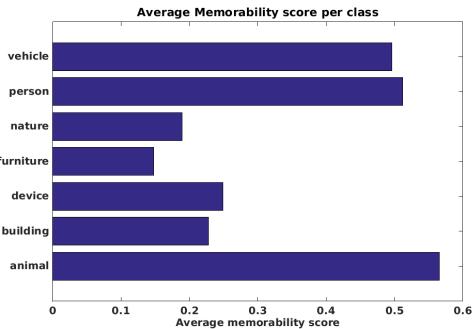
3.3.1 Are some object classes more memorable than others?

Our dataset contains 3414 object segmentations from the Pascal-S dataset. 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 1 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 like utensils, bottle, tv, computer etc, nature included segments like tree, mountain, flowers, and vehicle covered segments like car, bike, bus, aeroplane etc. **add a small table here which shows stats of the object classes n their frequency in the dataset.**

Figure 6 shows the average memorability score for the all 7 different object classes in our dataset. This visualisation gives a sense of how the memorability of different object classes differs. Animal, person, and vehicle are all highly memorable classes with each of them having an average memorability score greater than 0.5. Interestingly, all the 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 has an average memorability score of only 0.18 and is the least memorable object class. 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 could decrease their memorability significantly. On the contrary, objects from animal, person, and vehicle classes appear mostly in the foreground leading to a higher memorability score on average. Table 1 shows the average memorability of the top 20 most memorable objects for each object class. This analysis is particularly interesting as these top objects are not occluded and most of them tend to appear in the foreground. The top 20 most memorable objects from person, animal and vehicle have memorability higher than 0.90 whereas the average memorability of the top 20 objects from building, furniture, and nature is lesser than 0.70. While the differences in the memorability of different classes could be driven primarily due to factors like occlusion, size, background/foreground, the results in table 1 suggest that memorability could be an intrinsic property of an object class and some object classes like person, animal, vehicle are in general intrinsically more

540 541 542 543 544 545 546 547 548	Object Class	Memorability of top 20 objects
549	Animal	0.94
550	Building	0.65
551	Device	0.86
552	Furniture	0.63
553	Nature	0.67
554	Person	0.93
555	Vehicle	0.96

550 Table 1: Average memorability of top 20 objects for each
551 class



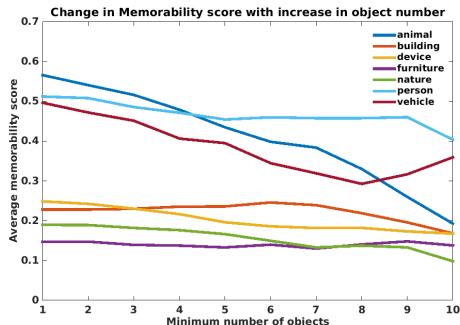
564 Figure 6: **Average memorability per object class.** Figure showing
565 some object classes are more memorable than others.

566
567
568
memorable than classes like furniture, nature etc.

570 3.3.2 Why are some objects not memorable within a 571 class?

573 As demonstrated above, some object classes like animal,
574 person, vehicle are more than others either due to their
575 intrinsic property or external factors like occlusion, fore-
576 ground/background etc. However, not all objects in a class
577 are equally memorable. The examples in figure 7 show the
578 most memorable and least memorable objects for each ob-
579 ject class. A common property between these objects is that
580 the least memorable objects within each object class tend
581 to be occluded and obstructed by other objects. What other
582 possible factors could influence the memorability of an ob-
583 ject within a class?

584 **Relationship with number of objects in an image:** We
585 first tested how the memorability of each object class is af-
586 fected by the number of objects inside an image. We took in
587 all images that had at least 'n' objects (where $1 \leq n \leq$
588 10) inside them and computed the average memorability of
589 each object class. Figure 8 shows the change in average
590 memorability of the different object classes with respect to
591 the increase in number of objects in an image. This figure
592 shows that the number of objects present in an image
593 is an important factor that results in the variation in the



594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

Figure 8: **Correlation between object class and number of objects.**
add-in later.

memorability of objects within a class. As the number of objects in an image increase, the memorability of the object classes decreases. In particular, the memorability of objects belonging to highly memorable classes like animal, person, and vehicle decreases significantly with the increase in number of objects in an image. Intuitively, this could mean that with an increase in number of objects, humans cannot concentrate on a single object in an image and have a difficulty remembering objects in those images. Thus, if a person, animal or vehicle is present in a highly cluttered image, humans can not remember those objects very well.

Relationship of inter-class memorability: While the previous analysis sheds light on how the memorability of an object class is effected by factors such as occlusion, and number of objects in an image, how does the presence of a particular object class effect the memorability of another object class? To investigate this, for each possible pair of object categories we took all images that contained at least one object from both the categories and computed the average memorability scores for the two object categories. This way we were able to capture the relationship between pairs of object categories and how their presence effects their respective memorability scores. Figure 9 plots these results and for each object class, shows how it's memorability is effected in the presence of other object classes. The red mark on each line shows the over all average memorability of that class. Firstly, the memorability of low memorable classes i.e. nature, furniture, device, and building is not effected a lot in the presence of other object categories and their memorability tends to be low throughout. The memorability of person class also remains close to it's average memorability score and tends to be unaffected by the presence of most object categories. However, the memorability of person does get effected in the presence of buildings as the average memorability of person class drops below 0.4 for images containing both persons and buildings. This could be due to the fact that there exists images wherein the building is quite large and the persons in those images are usually smaller or less significant in contrast to the build-

648
649
650
651
652
653
654
655
656

Figure 7: Qual Results. Figure showing top and bottom objects for each class.

702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

ing (also illustrated in figure 10). The memorability of the animal class remains close to it's average memorability in presence of most classes, however it drops significantly to below 0.4 in presence of the person class (also shown in figure 10). The memorability of vehicle class is strongly affected by the presence of other object categories. In particular, it's memorability drops significantly in presence of person and animal classes also shown in figure 10.

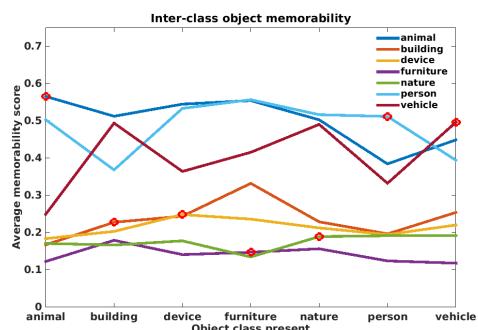


Figure 9: inter-class object memorability relationship. add-in later.

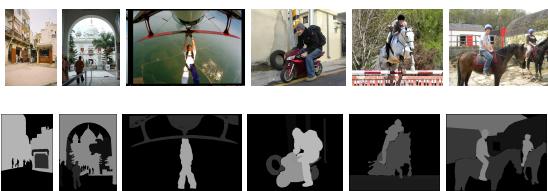


Figure 10: Qual Results. Figure showing how memorability of different classes is effected in presence of other classes. Bottom row is the memorability map

Taken together, the results from this section show that the memorability an object within a particular class is greatly effected by occlusion, number of objects present in an image and presence of other object classes in the image.

3.4. What is the relationship between object and image memorability?

[Note: only the skeleton of this section is added, this still needs significant work]

While in the previous sections, we have investigated the various factors that could effect object memorability, an important question remains. Now that we know what objects people remember and the factor that influence their

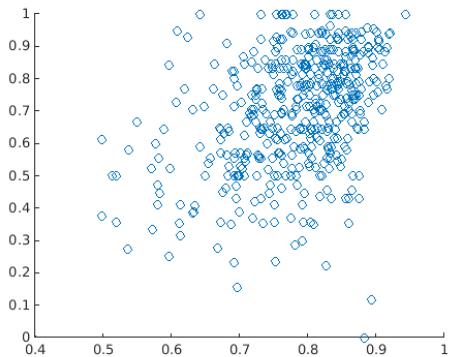
memorability, how does the memorability of objects effect the overall image memorability? To shed light on the relationship between image memorability and object memorability, we conducted another large-scale experiment on Amazon Mechanical turk for the images in our dataset to gather their respective memorability scores. For this experiment, we followed the exact paradigm as the memory game experiment proposed in [5]. [add details about the fillers, add details about the experiment including number of workers, number of times each image was viewed by workers and explain the consistency analysis]

After collecting the image memorability scores, we conducted several analysis to answer - if an image is highly memorable, what happens to memorability of objects inside the image (and vice versa)? For each image in the dataset, we firstly computed the maximum object memorability within the image. Figure 11 shows the high correlation of maximum object memorability with image memorability ($\rho = 0.43$). Although there is a high correlation between maximum object and image memorability, it still doesn't predict image memorability completely. However, table 2 also demonstrates that maximum object memorability is a very strong predictor of whether an image is highly memorable or low memorability. If we only take the top-most 20 and bottommost 20 images from the dataset, then their average max object memorability is highly correlated with the image memorability scores. Similar trend is observed for 50 most and 100 most top and bottom memorable images. This is also shown via the figure XX [show the qualitative examples here to show memorable images contain memorable objects and non-memorable images do not]. This shows although simply the max object memorability cannot predict whether an image is medium memorable, max object memorability serves as a strong features in predicting if an image is highly memorable or not. Images that are lowly memorable, is possibly due to the fact that they do not contain a single highly memorable object [refer to qual figure again].

4. Predicting Object Memorability

here the benchmarking and baseline code comes up

756	Top and bottom images	max object memorability correlation
757	20	0.8
758	50	0.6
759	100	0.5

761
762 Table 2: correlation for top and bottommost images
763764
765
766
767
768
769
770
771
772
773
774
775 Figure 11: Correlation between max object memorability and image
776 memorability. add-in later.
777778

References

- 782 [1] T. F. Brady, T. Konkle, G. A. Alvarez, and A. Oliva. Visual
783 long-term memory has a massive storage capacity for object
784 details. *Proceedings of the National Academy of Sciences*,
105(38):14325–14329, 2008. 3
- 785 [2] Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and
786 A. Oliva. Intrinsic and extrinsic effects on image memorability.
787 *Vision research*, 2015. 1, 2
- 788 [3] M. Everingham and J. Winn. The pascal visual object classes
789 challenge 2010 (voc2010) development kit, 2010. 2
- 790 [4] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What
791 makes a photograph memorable? *Pattern Analysis and Ma-*
792 *chine Intelligence, IEEE Transactions on*, 36(7):1469–1482,
793 2014. 1
- 794 [5] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an
795 image memorable? In *Computer Vision and Pattern Recog-*
796 *nition (CVPR), 2011 IEEE Conference on*, pages 145–152.
797 IEEE, 2011. 1, 2, 3, 7
- 798 [6] A. Khosla, J. Xiao, A. Torralba, and A. Oliva. Memorability
799 of image regions. In *Advances in Neural Information Pro-*
800 *cessing Systems*, pages 305–313, 2012. 1
- 801 [7] J. Kim, S. Yoon, and V. Pavlovic. Relative spatial features for
802 image memorability. In *Proceedings of the 21st ACM inter-*
803 *national conference on Multimedia*, pages 761–764. ACM,
804 2013. 1
- 805 [8] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The
806 secrets of salient object segmentation. In *Computer Vision*
807 *and Pattern Recognition (CVPR), 2014 IEEE Conference on*,
808 pages 280–287. IEEE, 2014. 2
- 809 [9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ra-
manan, P. Dollár, and C. L. Zitnick. Microsoft coco: Com-

- 810 mon objects in context. In *Computer Vision–ECCV 2014*,
811 pages 740–755. Springer, 2014. 3
- 812 [10] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and
813 H.-Y. Shum. Learning to detect a salient object. *Pattern*
814 *Analysis and Machine Intelligence, IEEE Transactions on*,
33(2):353–367, 2011. 3
- 815 [11] M. Mancas and O. Le Meur. Memorability of natural scenes:
816 the role of attention. In *ICIP*, 2013. 2
- 817 [12] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang.
818 Saliency detection via graph-based manifold ranking. In
819 *Computer Vision and Pattern Recognition (CVPR), 2013*
820 *IEEE Conference on*, pages 3166–3173. IEEE, 2013. 3
- 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