

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], [7], [4], [2], [8]. 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], [13], 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 [10], 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. We removed all homogenous
176 non-object or background segments such as ground, grass,
177 floor, sky etc, as well as imperceptible object fragments and
178 excessively blurred regions. All remaining object segmentations
179 were tested for memorability. In the end, our final
180 dataset consisted of 850 images and 3412 object segmentations
181 i.e. on average each image consisted of approximately
182 4 object segments for which we gathered the ground truth
183 memorability on.

184 2.1. Object Memory Game

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

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

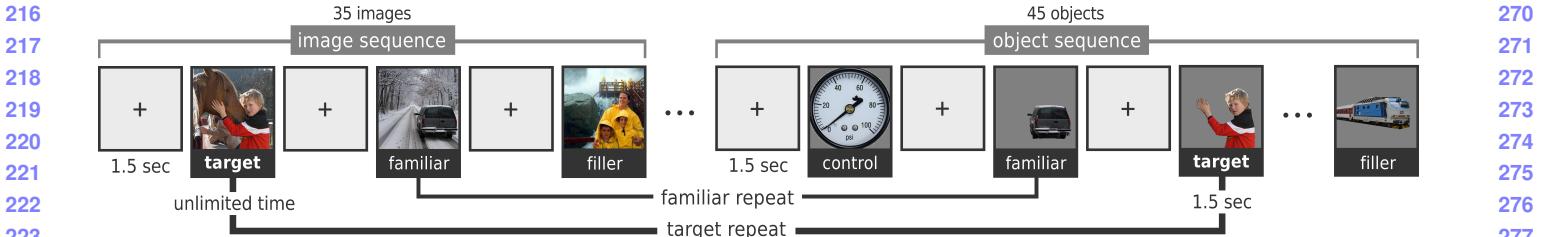


Figure 2: Main task. add-in later.

jects the participants were to later identify. The remaining images in the sequence consisted of 16 'filler' images and 16 'familiar' images. The 'filler' images were randomly selected from the DUT-OMRON dataset [14] and the 'familiar' images were randomly sampled from the MSRA dataset proposed in [12]. 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 [11] 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 [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 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). 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 (ρ). 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 [5], and with good reason [9], 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 [5] that 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

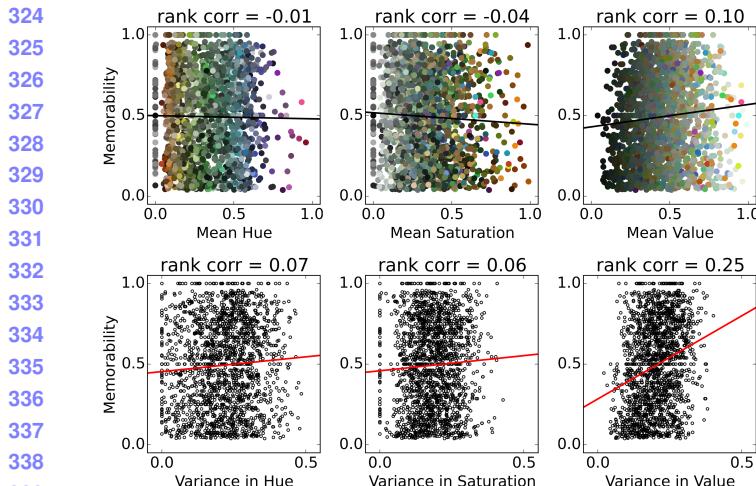


Figure 3: **Correlations between simple color features and object memorability.** Color features were computed from HSV representations of the objects.

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 viewers 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 [5], [4]. 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 [10]. 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 objects 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 objects 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

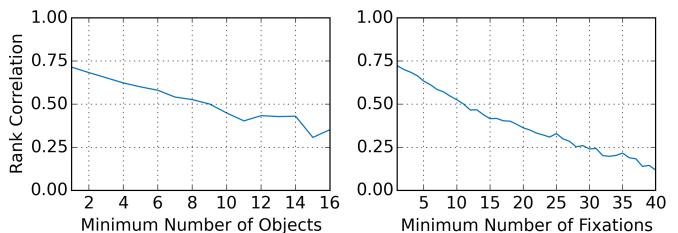


Figure 4: **Correlation between object memorability and number of fixations.** add-in later.

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.

Center Bias: Figure 6 elucidates another example where saliency and memorability diverge. While previous studies related to visual saliency have showed that saliency is heavily influenced by center bias [6], [15], Figure 6 shows that memorability exhibits comparatively less center bias

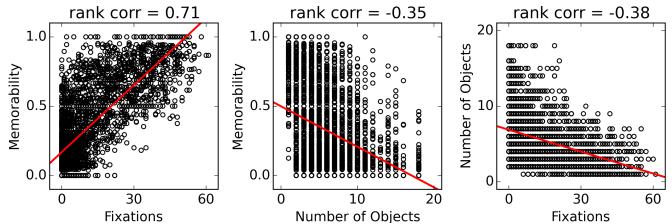
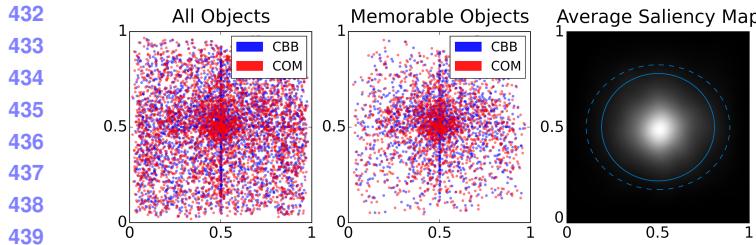


Figure 5: **Correlation between object memorability and number of fixations.** add-in later.

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.

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).

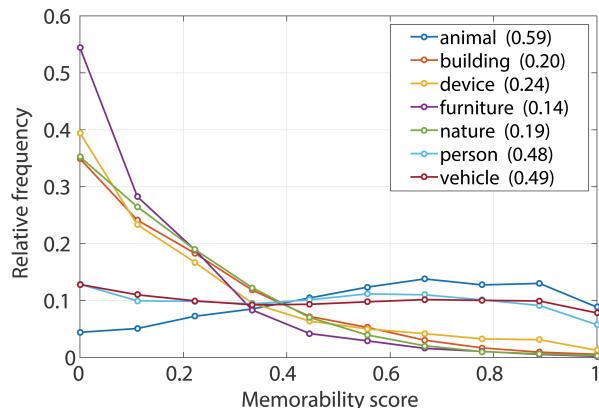
3.3. How do object categories effect memorability?

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

Figure 7 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 memo-

Figure 7: **Average memorability per object class.** Figure showing some object classes are more memorable than others.

rability 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 7 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.

3.3.2 Why are some objects not memorable within a class?

As demonstrated above, some object classes like animal, person, vehicle are more memorable than others. However, not all objects in a class are equally memorable. The examples in figure ?? show the most memorable and least memorable objects for each object class. A common property between these objects is that the least memorable objects within each object class tends to be occluded and obstructed by other objects. What other possible factors could influence the memorability of an object within a class? Among various possible factors, we explored how object memorability within a particular class is influenced by a) number of objects and b) presence of another object class.

Number of objects: We first tested how the memorability of each object class is affected by the number of objects inside an image. We took in all images that had at least 'n' objects (where $1 \leq n \leq 10$) inside them and com-

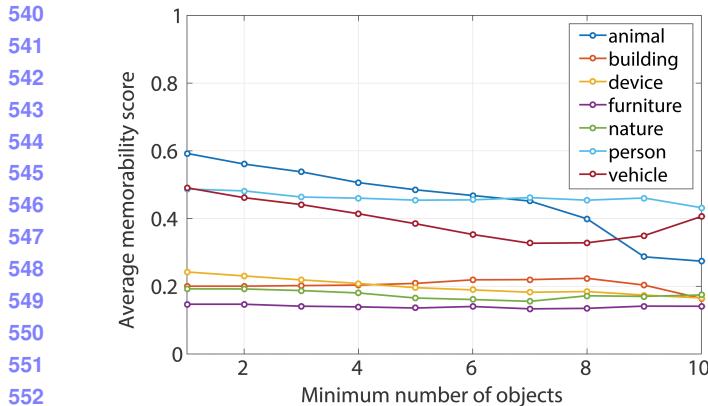


Figure 8: Correlation between object class and number of objects. add-in later.

puted the average memorability of each object class. Figure 8 shows the change in average memorability of the different object classes with respect to the increase in number of objects in an image. Results from this figure show that the number of objects present in an image is an important factor in determining memorability especially for classes like animal and vehicle. As the number of objects in an image increase, the memorability of animal and vehicle classes decreases sharply. This could mean that with an increase in number of objects, viewers cannot concentrate on a single object in an image and have a difficulty remembering objects in those images. Interestingly, the memorability of the person class does not change significantly with an increase in number of objects. This suggests that humans tend to remember people even in highly cluttered images.

Inter-class memorability: How does the presence of a particular object class influence the memorability of another object class? For each possible pair of object categories, we took all images that contained at least one object from both the categories and computed the change in the average memorability scores for the two object categories. Figure 9 captures the relationship between different pairs of object categories and shows how the memorability of each object class is affected in the presence of other object classes. Firstly, the memorability of low memorable classes i.e. nature, furniture, device, and building is not effected greatly in the presence of other object categories and their memorability tends to be low throughout. The memorability of the animal class remains close to its average memorability in presence of most classes, but it drops significantly in the presence of vehicle and person class. The memorability of person class also remains close to its average memorability score and tends to be unaffected by the presence of most object categories. However, the memorability of person does go lower in the presence of vehicle and building classes. This could be due to the fact that persons in images containing vehicles or buildings are usually zoomed

out and are usually smaller in size (also illustrated 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. Taken together, when an animal, vehicle or a person are present in the same image, the memorability for all these object classes usually goes down. However, this change in memorability is different for all these classes which leads to interesting results. When a vehicle or an animal is present alongside a person, the person is generally more memorable. When a vehicle and an animal is present in the same image, even though the memorability of both these classes drops significantly, the animal is generally more memorable than the vehicle. We also show examples of such inter-class memorability effects in Figure 10.

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 re-

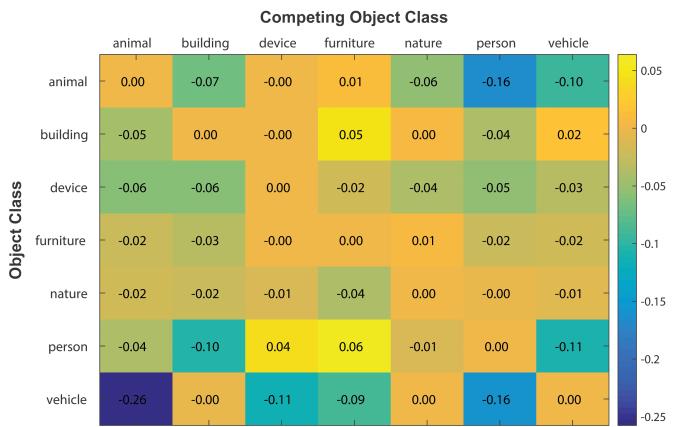


Figure 9: inter-class object memorability relationship. add-in later.

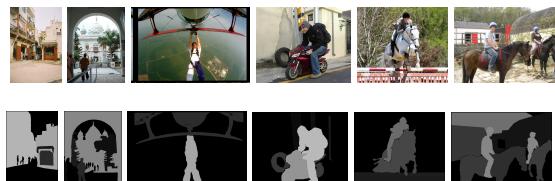


Figure 10: Qual Results. Figure showing how memorability of different classes is effected in presence of other classes. Bottom row is the memorability map

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

648	Animal	Building	Device	Furniture	Nature	Person	Vehicle	Everingham and J. Winn. The pascal visual object classes	702
649	0.38	0.22	0.47	0.53	0.64	0.54	0.30	challenge2010 (voc2010) development kit, 2010. 2	703

650									704
-----	--	--	--	--	--	--	--	--	-----

651									705
-----	--	--	--	--	--	--	--	--	-----

652									706
-----	--	--	--	--	--	--	--	--	-----

653									707
-----	--	--	--	--	--	--	--	--	-----

654									708
-----	--	--	--	--	--	--	--	--	-----

655									709
-----	--	--	--	--	--	--	--	--	-----

656									710
-----	--	--	--	--	--	--	--	--	-----

657									711
-----	--	--	--	--	--	--	--	--	-----

658									712
-----	--	--	--	--	--	--	--	--	-----

659									713
-----	--	--	--	--	--	--	--	--	-----

660									714
-----	--	--	--	--	--	--	--	--	-----

661									715
-----	--	--	--	--	--	--	--	--	-----

662									716
-----	--	--	--	--	--	--	--	--	-----

663									717
-----	--	--	--	--	--	--	--	--	-----

664									718
-----	--	--	--	--	--	--	--	--	-----

665									719
-----	--	--	--	--	--	--	--	--	-----

666									720
-----	--	--	--	--	--	--	--	--	-----

667									721
-----	--	--	--	--	--	--	--	--	-----

668									722
-----	--	--	--	--	--	--	--	--	-----

669									723
-----	--	--	--	--	--	--	--	--	-----

670									724
-----	--	--	--	--	--	--	--	--	-----

671									725
-----	--	--	--	--	--	--	--	--	-----

672									726
-----	--	--	--	--	--	--	--	--	-----

673									727
-----	--	--	--	--	--	--	--	--	-----

674									728
-----	--	--	--	--	--	--	--	--	-----

675									729
-----	--	--	--	--	--	--	--	--	-----

676									730
-----	--	--	--	--	--	--	--	--	-----

677									731
-----	--	--	--	--	--	--	--	--	-----

678									732
-----	--	--	--	--	--	--	--	--	-----

679									733
-----	--	--	--	--	--	--	--	--	-----

680									734
-----	--	--	--	--	--	--	--	--	-----

681									735
-----	--	--	--	--	--	--	--	--	-----

682									736
-----	--	--	--	--	--	--	--	--	-----

683									737
-----	--	--	--	--	--	--	--	--	-----

684									738
-----	--	--	--	--	--	--	--	--	-----

685									739
-----	--	--	--	--	--	--	--	--	-----

686									740
-----	--	--	--	--	--	--	--	--	-----

687									741
-----	--	--	--	--	--	--	--	--	-----

688									742
-----	--	--	--	--	--	--	--	--	-----

689									743
-----	--	--	--	--	--	--	--	--	-----

690									744
-----	--	--	--	--	--	--	--	--	-----

691									745
-----	--	--	--	--	--	--	--	--	-----

692									746
-----	--	--	--	--	--	--	--	--	-----

693									747
-----	--	--	--	--	--	--	--	--	-----

694									748
-----	--	--	--	--	--	--	--	--	-----

695									749
-----	--	--	--	--	--	--	--	--	-----

696									750
-----	--	--	--	--	--	--	--	--	-----

697									751
-----	--	--	--	--	--	--	--	--	-----

698									752
-----	--	--	--	--	--	--	--	--	-----

699									753
-----	--	--	--	--	--	--	--	--	-----

700									754
-----	--	--	--	--	--	--	--	--	-----

701									755
-----	--	--	--	--	--	--	--	--	-----