Supplemental Material for submission "Do Objects in Real-World Images Have a Canonical Name?"

Anonymous EMNLP-IJCNLP submission

1 Dataset: Supplementary Information

Figures 1 and 2 contain the instructions for our subjects in the first round and rounds 2-4, respectively. After the first round, based on the opt-out annotation, we kept images that met all the following conditions (thresholds were estimated via manual inspection):

- they were not marked as occluded by any subject;
- "Bounding box is unclear" was marked at most twice:
- at most 17% of elicited names were in plural form (to remove cases where the bounding box contains several objects);
- the most frequent elicited name is of the same domain as the VisualGenome name.

This yielded 25,596 images, discarding 5,497. Table 1 gives an overview of the collection synsets, grouped into the 7 domains, and Table 2 shows the 7 domains together with the top 10 VG names.

150

151

152

153154155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178 179

180

181

182

183

184 185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

103
104
105
106
107
108
109
110
111
112
113
114
115
116

100

101

102

117118119

120

121

128

129

130

143144145

142

149

Domain Synsets buildings house₁ (364) bridge₁ (297) $shelter_1$ (169) restaurant₁ (58) outbuilding₁ (31) $hotel_1$ (19) housing₁ (17) place_of_worship₁ (12) people woman₁ (1768) man₁ (1167) male_child₁ (853) athlete₁ (396)child₁ (333) creator₂ (11) professional₁ (5) furnishing₂ (5355) home vessel₃ (525) kitchen_utensil₁ (132) crockery₁ (92) lamp₁ (34) cutlery₂ (82) $tool_1$ (72) clothing shirt₁ (968) overgarment₁ (786) dress₁ (199) headdress₁ (135) glove₂ (7) neckwear₁ (65) robe₁ (27) $footwear_1$ (5) vehicles $train_1$ (957) aircraft₁ (1208) car₁ (727) motorcycle₁ (564) boat₁ (499) $truck_1$ (559) ship₁ (38) food $foodstuff_2$ (280) dish₂ (812) baked_goods₁ (770) vegetable₁ (48) edible_fruit₁ (42) $beverage_1$ (23) animals_plants ungulate₁ (2037) horse₁ (833) feline₁ (763) dog₁ (688) bird₁ (389) flower₁ (44) rodent₁ (27) insect₁ (12) $fish_1$ (11)

Table 1: Overview of our dataset: Synset nodes for each domain (subscript indicates synset number; number of instances in parentheses).

vehicles	food	animals_plants	shome	buildings	people	clothing
train (954)	pizza (518)	giraffe (915)	bed (888)	house (340)	boy (853)	shirt (904)
car (642)	cake (261)	horse (822)	bench (714)	bridge (274)	man (806)	jacket (451)
plane (485)	bread (186)	cat (754)	table (687)	dugout (91)	woman (766)	coat (267)
airplane (479)	sandwich (153)		desk (672)	tent (53)	girl (650)	dress (190)
motorcycle (466) bun (143)	zebra (461)	counter (516)	restaurant (33)	lady (342)	hat (77)
truck (465)	cheese (110)	cow (324)	couch (366)	overpass (23)	guy (330)	t-shirt (62)
boat (450)	donut (78)	bird (295)	chair (365)	grill (22)	child (230)	tie (51)
jet (106)	salad (70)	sheep (216)	carpet (307)	garage (18)	batter (110)	blazer (43)
aircraft (85)	sauce (68)	bull (48)	bowl (219)	hotel (16)	kid (85)	hood (26)
van (76)	apple (33)	flower (40)	curtain (182)	castle (14)	skateboarder (80)	cap (20)

Table 2: Overview of our ManyNames dataset: Top-10 VG names for each domain (number of instances in parentheses).

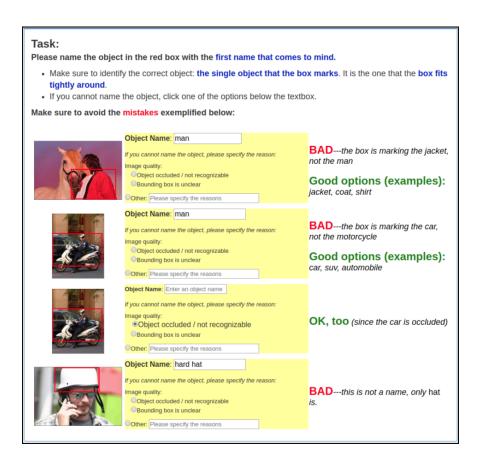


Figure 1: Instructions for AMT annotators for the first round.

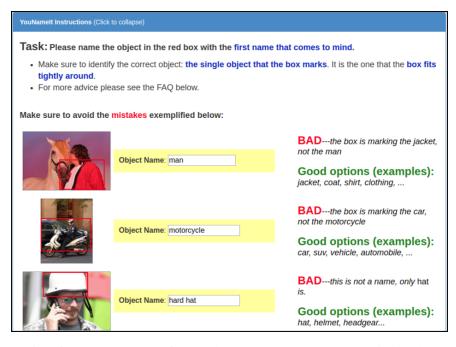


Figure 2: Instructions for AMT annotators for rounds 2 to 4. They were accompanied by the FAQ in Figure 3

Confidential Review Copy. DO NOT DISTRIBUTE.

_	
F	AQ:
	+ If there is a person riding a skateboard or riding a surfboard, would you prefer the answer to be man or surfer?
	There is no unique "correct" or "best" answerwe ask you to enter the first name that comes to mind for the marked object. For example, if you first think of surfer, then please enter this name. If, however, man comes to mind first, then enter man.
	+ For food, for example, a photo of a pizza would be <i>pizza</i> and not <i>food</i> ?
	That's up to youthere is no unique "correct" or "best" answer. Instead, please enter the name which first occurred to you for the identified object; if you first thought food, then food, if first pizza, then pizza.
	+ I was wondering what you wanted us to do with images and boundary boxes around women, men, children that are not particularly doing anything? It looks as though you do not want us to name it <i>man</i> or <i>woman</i>
t	only name is ok as long as it identifies the right object; man, woman, child, etc. are all valid names for objects. It is important, nough, to make sure that you have identified the correct object for which we ask you to give a name. As the instructions say, is the object that the bounding box is tightly around.
	+ In some cases, a group of objects together could be described with a singular noun (i.e. bundle, salad, sandwich, etc.).
t	you can describe the objects with a singular noun, i.e. they together form some kind of object, such as sandwich or salad, nen this is completely fine. It is important though that this "group of objects" is to be named, that is, the bounding box is tightly round them as a whole (for instance, pay attention that it is not only marking the filling of a sandwich).
	+ When there is a picture of a person and the whole person is in the red box and they are doing a sport is it better that I say player, snowboarder, skier, etc." instead of man, woman, girl, boy, etc?
	here is no "correct answer", so you are just asked to enter the first name that comes to mind for the marked object. Referring irectly to your example, any of the options, player OR snowboarder OR man OR woman etc., are good.