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.

Domain	Synsets			
buildings	house$_1$ (364)	bridge$_1$ (297)	shelter$_1$ (169)	restaurant$_1$ (58)
	outbuilding$_1$ (31)	hotel$_1$ (19)	housing$_1$ (17)	place_of_worship$_1$ (12)
people	woman$_1$ (1768)	man$_1$ (1167)	male_child$_1$ (853)	athlete$_1$ (396)
	child$_1$ (333)	creator$_2$ (11)	professional$_1$ (5)	
home	furnishing$_2$ (5355)	vessel$_3$ (525)	kitchen_utensil$_1$ (132)	crockery$_1$ (92)
	cutlery$_2$ (82)	tool$_1$ (72)	lamp$_1$ (34)	
clothing	shirt$_1$ (968)	overgarment$_1$ (786)	dress$_1$ (199)	headdress$_1$ (135)
	neckwear$_1$ (65)	robe$_1$ (27)	glove$_2$ (7)	footwear$_1$ (5)
vehicles	aircraft$_1$ (1208)	train$_1$ (957)	car$_1$ (727)	motorcycle$_1$ (564)
	truck$_1$ (559)	boat$_1$ (499)	ship$_1$ (38)	
food	dish$_2$ (812)	baked_goods$_1$ (770)	foodstuff$_2$ (280)	vegetable$_1$ (48)
	edible_fruit$_1$ (42)	beverage$_1$ (23)		
animals_plants	ungulate$_1$ (2037)	horse$_1$ (833)	feline$_1$ (763)	dog$_1$ (688)
	bird$_1$ (389)	flower$_1$ (44)	rodent$_1$ (27)	insect$_1$ (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	home	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)	dog (654)	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).

2

200
250
201
251
202
252
203
253
204
254
205
255
206
256
207
257
208
258
209
259
210
260
211
261
212
262
213
263
214
264
215
265
216
266
217
267
218
268
219
269
220
270
221
271
222
272
223
273
224
274

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

225
275
226
276
227
277
228
278
229
279
230
280
231
281
232
282
233
283
234
284
235
285
236
286
237
287
238
288
239
289
240
290
241
291
242
292
243
293
244
294
245
295
246
296

247
297
Figure 2: Instructions for AMT annotators for rounds 2 to 4. They were accompanied by the FAQ in Figure 3
248
298
249
299

FAQ:

+ 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" answer---we 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 *pizza* and not *food*?

That's up to you---there 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 *man* or *woman*...

Any name is ok as long as it identifies the right object; *man*, *woman*, *child*, etc. are all valid names for objects. It is important, though, to make sure that you have identified the correct object for which we ask you to give a name. As the instructions say, it 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.).

If you can describe the objects with a singular noun, i.e. they together form some kind of object, such as sandwich or salad, then this is completely fine. It is important though that this "group of objects" is to be named, that is, the bounding box is tightly around 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?

There is no "correct answer", so you are just asked to enter the first name that comes to mind for the marked object. Referring directly to your example, any of the options, *player* OR *snowboarder* OR *man* OR *woman* etc., are good.

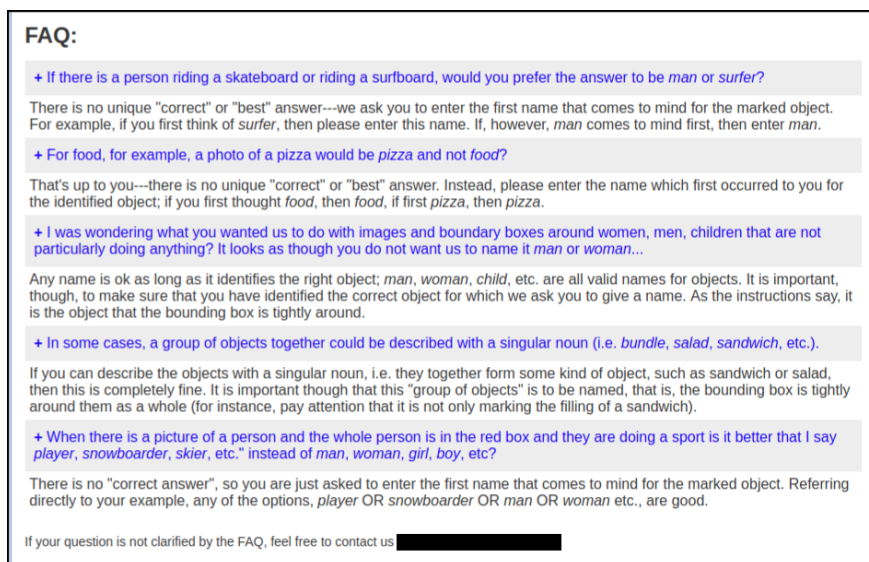If your question is not clarified by the FAQ, feel free to contact us ▮▮▮▮▮▮▮▮

Figure 3: FAQ accompanying the instructions for AMT annotators for rounds 2 to 4.

4