Annotation guidelines:
Identification of target objects in images

[This is an English translation of the annotation guidelines, which were originally prepared in the local language.]

Overview

The data to be annotated are descriptions for objects in photographs generated by different systems.
The target objects for the descriptions are each marked with a bounding box. The aim is to annotate whether the respective target object has been **correctly identified** by the system, i.e. whether the description contains a **valid description of the type of the target object**.
In most cases, type designations are realized as nouns. Example: If the bounding box in the image marks a dog, an expression such as "the dog on the right" would contain *dog* as a valid type description, as the target object is of type *dog*. Possible alternatives could be e.g. *animal* or *pet*, as the depicted dog also falls into these categories. "The cat" would have to be annotated as false. Descriptions such as "it is on the right", "upper left", "the grey one" or "the thing" do not contain type designations for the target object and would have to be annotated accordingly. **Only the type designations are relevant for the annotation**. It can be ignored whether the descriptions of other properties (e.g. color, position in the image, ...) are true. The sentence "the dog on the left" should therefore be annotated as correct with regard to the type label as long as the marked object is a dog, even if it is on the right side of the image. Similarly, "the red suitcase" should also be annotated as correct if the marked object was a suitcase, but was actually blue.
The outputs of several systems are to be annotated. For each system there is an HTML file and a corresponding CSV file, each with the same file name (apart from the file extension). The HTML file contains photographs with marked target objects, an ID and a generated description for each object. The CSV file contains the same IDs and descriptions in the same order, as well as an empty column in which the annotation is to be entered.

Categories

Each item should be assigned **exactly one of the following four categories**:

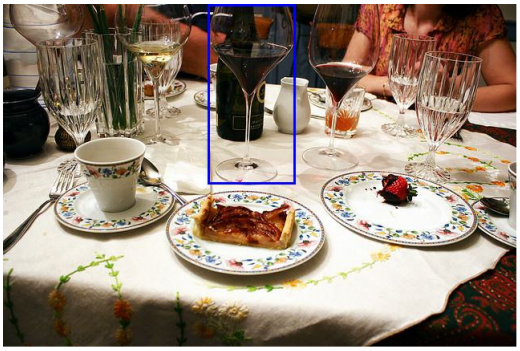
- **A** *(Adequate)*: The description contains a **valid type identifier** for the target.
- **O** *(Omission)*: The description contains **no type identifier** for the target (e.g. due to pronominalization, general nouns such as *thing*, or the restriction to non-type properties such as *blue*; *large* or the position in the image, e.g. *bottom left corner*; *on the right*; *top left*).
- **M** *(Misaligned)*: The description contains a **type identifier for a different object** than the intended target, which is also (partially) covered by the bounding box.
- **F** *(False)*: The description contains a **false type identifier** that does not apply to the target (and does not refer to any other object in the bounding box).

Examples for the individual categories are appended to the end of this file. For complete examples, please refer to the files *demo_file.csv* and *demo_file.html*.

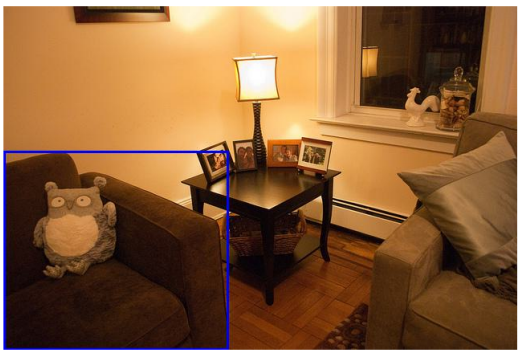
Caution: As previously stated, **only the type descriptions** must be taken into account for the categorization. It is irrelevant whether other attributes or aspects are described truthfully (e.g. whether *the left dog* is actually on the left in the image, or whether *the blue suitcase* is actually blue). It should also be ignored whether a description could also refer to other objects in the image (such as *the dog* if several dogs are shown).
In borderline cases, the most intuitively appropriate category can be used. In rare cases, it may be debatable what constitutes the object designation in the sentence. Here, you should generally choose designations of (a) concrete entities that (b) do not represent attributes or specifications for other object designations. For "bowl of apples", "bowl" would therefore be the relevant term (since "bowl" denotes a concrete thing and "of apples" a specification); for "bunch of apples", on the other hand, it would be "apples" (since "bunch" does not denote a concrete thing).

Examples

Category A



index: 9
ann_id: 667576
expression: middle wine glass



index: 69
ann_id: 117421
expression: left couch

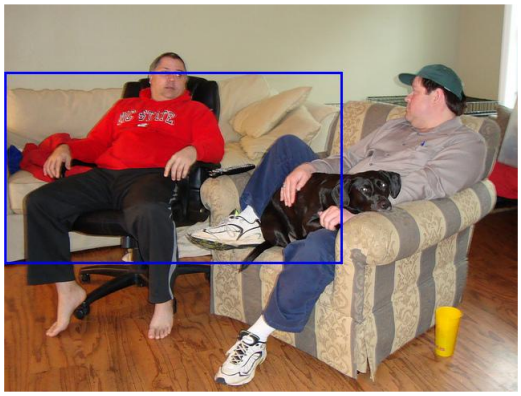
Category O

index: 15
ann_id: 322419
expression: bottom left

index: 24
ann_id: 65175
expression: right side of screen

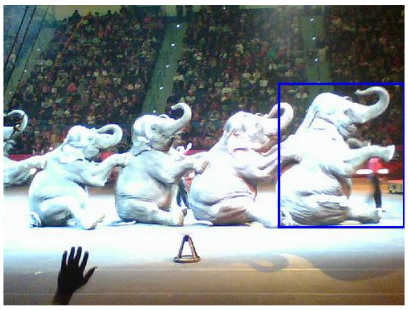
Category M

index: 5
ann_id: 97835
expression: chair in middle

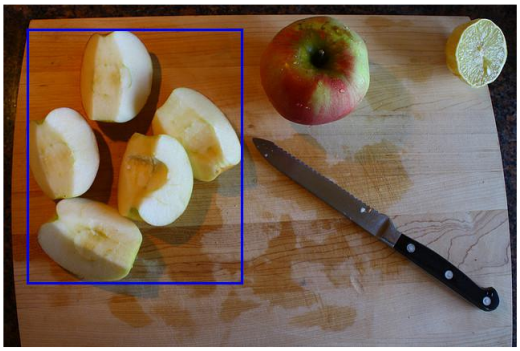


index: 82
ann_id: 111364
expression: red shirt

Category F



index: 10
ann_id: 580317
expression: white dog



index: 30
ann_id: 1048809
expression: banana slices on left

index: 96
ann_id: 1118003
expression: white book