

Project Note: VoxML Standardization

Manolito Octaviano Jr.
National University, Philippines
mvocavianojr@national-u.edu.ph

Abstract

This paper presents a short project notes and recommendations on the annotation shared task. The paper is divided into 4 parts: overview of the shared task; notes on annotation tool; points on annotation manual; and deeper investigation that can be done on the next cycle of the shared task.

1 Shared Task

The goal of the shared task is to crowd-source image annotations from different groups of participants. This is part to develop the VoxML (Pustejovsky and Krishnaswamy, 2016) into an ISO Standard for the representation of visual information. The use of crowdsourcing approach to gather annotations ensures a large number of contributors for fast and efficient development of voxicon (i.e. library of object defined within VoxML).

In the shared task, participants are instructed to annotate as many images as possible using the provided tool¹ that must be installed. A guideline how the images should be annotated is also provided.

2 Annotation Tool

The annotation tool utilized in the shared task is shown in Figure 1. The tool can easily be used by anyone even a non-computing background user. It has a clean design and well-structured organization to provide a smooth user experience. Below is the list of recommendations that can be adopted to improve the user's experience for utilizing the tool:

Technical Guideline: Non-computing users might face technical problems when installing/setting the tool on local device. A technical guideline with step-by-step procedure and images on how the tool should be installed

must be included. The guideline must show what the user should expect to see after each step is accomplished.

Skip Button: There are some images that are hard to analyze and takes time to annotate. In order to maximize the good quality of annotation from the user, a skip button should be added. Skipped image can be rolled again to the user at least after 10 new images.

Delete Button (add object): Accidentally pressing the “add object” in the tool is inevitable. Furthermore, the tool does not allow the user to submit the annotation if the additional object is blank. Thus, delete button is necessary to avoid this case.

VoxML Annotation Survey

Provide a brief caption of the image.

Caption

What activity (if any) is the focus of the scene?

Focus Activity

Identify the objects/entities in the scene (including people and animals).
For each entity, add activities associated with that object.
Check the "Shown in Scene" box if the listed activity is depicted in the image.
Otherwise, use the next field to indicate what would need to change for that activity to be possible.

Object

Associated Activity

Shown in Scene

To Activity: "Object" must

ADD ACTIVITY

ADD OBJECT

Identify the major spatial and configurational relations between entities in the scene.
See the provided guideline for sample relations.

Spatial Relation

ADD WORD

SUBMIT & GO TO NEXT

Figure 1: The interface of the tool.

¹<https://github.com/csu-signal/VoxML-Track-Annotation-2021>

3 Annotation Manual

The annotation manual provides a detailed explanation on how the annotators will annotate (see section 1 and 2 of the manual) the different images using the tool. The example annotation (see section 3 of the manual²) provided to the users is very helpful as reference for annotating the images. Furthermore, it gives an overview of what the shared task is all about (see section 4 of the manual). Below is the list of recommendations that can be added in the manual:

Research Ethics: A short discussion of research ethics considered in the study must be added. It is important to discuss with the annotators the ethics associated to research, especially on the photographs of people (Wiles et al., 2008).

Relations: The manual should provide a complete list of word relations that can be used by the annotators. This is to ensure that all annotators will use the same set of vocabulary.

Depicted Action³: A template on how the participants will annotate a depicted action must be provided. It will ensure that annotators will have the same way of writing the action. This will answer the question: “Are the annotators will also use the format *To <activity>, <object> must* and checking the ‘Shown In Scene’ box?”

4 Deeper Investigation

Below is the list of recommendations that can be considered on the next cycle of image annotation.

Basic Information: The tool should also record some basic information of the annotators such as age and race since these factors could affect the quality of annotation produced.

Image Domain: The images provided are random that came from a different domain. Thus, each image consists of different objects and activities. This makes the annotation hard since most images are unrelated to each other. In order to solve this, it will be much easier for the participants to annotate images came from a specific domain. For example, images that will be rolled are all sports-related, school-related, or food-related.

²<https://github.com/csu-signal/VoxML-Track-Annotation-2021/blob/main/ISA-17-guideline.pdf>

³The guide on how to annotate the depicted action is not as clear on how to annotate the activity that is not depicted by the image.

Emotion: Emotion depicted in the image, if any, can also be added in annotation (Poretski et al., 2019). Specifically, the six basic emotion (Ekman, 1992) or neutral can be adopted⁴ in annotation.

⁴In the tool, this can be implemented using radio button for the ease of use.

References

- Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- Lev Poretzki, Joel Lanir, and Ofer Arazy. 2019. Feel the image: The role of emotions in the image-seeking process. *Human-Computer Interaction*, 34(3):240–277.
- James Pustejovsky and Nikhil Krishnaswamy. 2016. Voxml: A visualization modeling language. *arXiv preprint arXiv:1610.01508*.
- Rose Wiles, Jon Prosser, Anna Bagnoli, Andrew Clark, Katherine Davies, Sally Holland, and Emma Renold. 2008. Visual ethics: Ethical issues in visual research.