

---

# Supplementary for Continuous Perception Benchmark

---

Zeyu Wang, Zhenzhen Weng, Serena Yeung-Levy  
Stanford University  
{wangzeyu, zzweng, syyeung}@stanford.edu

## 1 Dataset

### 2 1.1 Motivation

#### 3 For what purpose was the dataset created?

4 The primary aim of the proposed dataset is to assess the video comprehension capabilities of multi-  
5 modal models. Unlike most existing datasets, this one emphasizes evaluating the model’s continuous  
6 perception and processing of visual signals, simulating the way humans handle visual information.

#### 7 Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., 8 company, institution, organization)?

9 The dataset was developed by a group of researchers at Stanford University.

### 10 1.2 Composition

11 **What do the instances that comprise the dataset represent (e.g., documents, photos, people,  
12 countries)?** Are there multiple types of instances (e.g., movies, users, and ratings; people and  
13 interactions between them; nodes and edges)? Please provide a description.

14 Each instance consists of a video, a question, and an answer. The video depicts an office scene  
15 featuring a varying number of objects.

#### 16 How many instances are there in total (of each type, if appropriate)?

17 There are in total 200 video question-answer pairs.

18 **Does the dataset contain all possible instances or is it a sample (not necessarily random)  
19 of instances from a larger set?** If the dataset is a sample, then what is the larger set? Is the  
20 sample representative of the larger set (e.g., geographic coverage)? If so, please describe how  
21 this representativeness was validated/verified. If it is not representative of the larger set, please  
22 describe why not (e.g., to cover a more diverse range of instances, because instances were withheld  
23 or unavailable).

24 The dataset includes all instances we created for the benchmark. Being synthetic, it allows for the  
25 generation of additional instances.

26 **What data does each instance consist of? “Raw” data (e.g., unprocessed text or images) or  
27 features?**In either case, please provide a description.

28 Each instance is composed of a video, a question and an answer for the question.

29 **Is there a label or target associated with each instance?**If so, please provide a description.

30 Yes, there is a groundtruth label for each instance.

31 **Is any information missing from individual instances?**If so, please provide a description, explaining  
32 why this information is missing (e.g., because it was unavailable). This does not include intentionally  
33 removed information, but might include, e.g., redacted text.

34 No.

35 **Are relationships between individual instances made explicit (e.g., users' movie ratings, social  
36 network links)?**If so, please describe how these relationships are made explicit.

37 N/A.

38 **Are there recommended data splits (e.g., training, development/validation, testing)?**If so, please  
39 provide a description of these splits, explaining the rationale behind them.

40 All instances in the dataset are designated for testing.

41 **Are there any errors, sources of noise, or redundancies in the dataset?**If so, please provide a  
42 description.

43 No.

44 **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g.,  
45 websites, tweets, other datasets)?**If it links to or relies on external resources, a) are there guarantees  
46 that they will exist, and remain constant, over time; b) are there official archival versions of the  
47 complete dataset (i.e., including the external resources as they existed at the time the dataset was  
48 created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources  
49 that might apply to a future user? Please provide descriptions of all external resources and any  
50 restrictions associated with them, as well as links or other access points, as appropriate.

51 The dataset is self-contained.

52 **Does the dataset contain data that might be considered confidential (e.g., data that is pro-  
53 tected by legal privilege or by doctor-patient confidentiality, data that includes the content of  
54 individuals non-public communications)?**If so, please provide a description.

55 No.

56 **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening,  
57 or might otherwise cause anxiety?**If so, please describe why.

58 No.

59 **Does the dataset relate to people?**If not, you may skip the remaining questions in this section.

60 No.

61 **Does the dataset identify any subpopulations (e.g., by age, gender)?**If so, please describe how  
62 these subpopulations are identified and provide a description of their respective distributions within  
63 the dataset.

64 N/A.

65 **Is it possible to identify individuals (i.e., one or more natural persons), either directly or  
66 indirectly (i.e., in combination with other data) from the dataset?**If so, please describe how.

67 N/A.

68 **Does the dataset contain data that might be considered sensitive in any way (e.g., data that  
69 reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or  
70 union memberships, or locations; financial or health data; biometric or genetic data; forms  
71 of government identification, such as social security numbers; criminal history)?**If so, please  
72 provide a description.

73 N/A.

74 **1.3 Collection Process**

75 **How was the data associated with each instance acquired?** Was the data directly observable (e.g.,  
76 raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived  
77 from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was  
78 reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If  
79 so, please describe how.

80 It is synthetically generated using the OmniGibson environment.

81 **What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or**  
82 **sensor, manual human curation, software program, software API)?** How were these mechanisms  
83 or procedures validated?

84 It is synthetically generated using the OmniGibson environment by rendering a predefined scene.

85 **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic,**  
86 **probabilistic with specific sampling probabilities)?**

87 N/A.

88 **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and**  
89 **how were they compensated (e.g., how much were crowdworkers paid)?**

90 The researchers listed in the author list involved in the data creation process.

91 **Were any ethical review processes conducted (e.g., by an institutional review board)?** If so, please  
92 provide a description of these review processes, including the outcomes, as well as a link or other  
93 access point to any supporting documentation.

94 No.

95 **Does the dataset relate to people?** If not, you may skip the remaining questions in this section.

96 No.

97 **Did you collect the data from the individuals in question directly, or obtain it via third parties**  
98 **or other sources (e.g., websites)?**

99 N/A.

100 **Were the individuals in question notified about the data collection?** If so, please describe (or show  
101 with screenshots or other information) how notice was provided, and provide a link or other access  
102 point to, or otherwise reproduce, the exact language of the notification itself.

103 N/A.

104 **Did the individuals in question consent to the collection and use of their data?** If so, please  
105 describe (or show with screenshots or other information) how consent was requested and provided,  
106 and provide a link or other access point to, or otherwise reproduce, the exact language to which the  
107 individuals consented.

108 N/A.

109 **If consent was obtained, were the consenting individuals provided with a mechanism to revoke**  
110 **their consent in the future or for certain uses?** If so, please provide a description, as well as a link  
111 or other access point to the mechanism (if appropriate).

112 N/A.

113 **Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data**  
114 **protection impact analysis) been conducted?** If so, please provide a description of this analysis,  
115 including the outcomes, as well as a link or other access point to any supporting documentation.

116 N/A.

117 **1.4 Preprocessing/cleaning/labeling**

118 **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing,**  
119 **tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing**  
120 **of missing values)?**If so, please provide a description. If not, you may skip the remainder of the  
121 questions in this section.

122 No.

123 **Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support**  
124 **unanticipated future uses)?**If so, please provide a link or other access point to the “raw” data.

125 N/A.

126 **Is the software used to preprocess/clean/label the instances available?**If so, please provide a link  
127 or other access point.

128 N/A.

129 **1.5 Uses**

130 **Has the dataset been used for any tasks already?**If so, please provide a description.

131 No.

132 **Is there a repository that links to any or all papers or systems that use the dataset?**If so, please  
133 provide a link or other access point.

134 N/A.

135 **What (other) tasks could the dataset be used for?**

136 In addition to evaluation, the dataset can also be used to train video understanding models.

137 **Is there anything about the composition of the dataset or the way it was collected and prepro-**  
138 **cessed/cleaned/labeled that might impact future uses?**For example, is there anything that a future  
139 user might need to know to avoid uses that could result in unfair treatment of individuals or groups  
140 (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal  
141 risks) If so, please provide a description. Is there anything a future user could do to mitigate these  
142 undesirable harms?

143 No.

144 **Are there tasks for which the dataset should not be used?**If so, please provide a description.

145 N/A.

146 **1.6 Distribution**

147 **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution,**  
148 **organization) on behalf of which the dataset was created?**If so, please provide a description.

149 The dataset is open to public.

150 **How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?**Does the dataset  
151 have a digital object identifier (DOI)?

152 The data will be distributed through Google Drive or other hosting services like Hug-  
153 gingFace. It can be accessed now at <https://drive.google.com/drive/folders/1gvX3JOXd06CMdCSMJGhwoCgWs5wK-nXb?usp=sharing>.

155 **Will the dataset be distributed under a copyright or other intellectual property (IP) license,**  
156 **and/or under applicable terms of use (ToU)?**If so, please describe this license and/or ToU, and

157 provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU,  
158 as well as any fees associated with these restrictions.

159 The dataset will be distributed under the CC BY-SA 4.0 license.

160 **Have any third parties imposed IP-based or other restrictions on the data associated with**  
161 **the instances?**If so, please describe these restrictions, and provide a link or other access point  
162 to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these  
163 restrictions.

164 No.

165 **Do any export controls or other regulatory restrictions apply to the dataset or to individual**  
166 **instances?**If so, please describe these restrictions, and provide a link or other access point to, or  
167 otherwise reproduce, any supporting documentation.

168 No.

## 169 1.7 Maintenance

170 **Who will be supporting/hosting/maintaining the dataset?**

171 The dataset will be maintained by the research team listed in the author list.

172 **How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**

173 The creators can be contacted by the emails mentioned in the author list.

174 **Is there an erratum?**If so, please provide a link or other access point.

175 No.

176 **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete in-**  
177 **stances)?**If so, please describe how often, by whom, and how updates will be communicated to users  
178 (e.g., mailing list, GitHub)?

179 Yes, the dataset will be updated whenever necessary.

180 **If the dataset relates to people, are there applicable limits on the retention of the data associated**  
181 **with the instances (e.g., were individuals in question told that their data would be retained for a**  
182 **fixed period of time and then deleted)?**If so, please describe these limits and explain how they will  
183 be enforced.

184 N/A.

185 **Will older versions of the dataset continue to be supported/hosted/maintained?**If so, please  
186 describe how. If not, please describe how its obsolescence will be communicated to users.

187 Yes.

188 **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for**  
189 **them to do so?**If so, please provide a description. Will these contributions be validated/verified? If  
190 so, please describe how. If not, why not? Is there a process for communicating/distributing these  
191 contributions to other users? If so, please provide a description.

192 Such requests can be communicated to the authors via email.