NLVR2: A New Dataset for Visual Reasoning + Natural Language with 107K New Examples

nlvr.ai

TRUE

or

FALSE

(the correct label is true)

One image shows exactly two brown acorns in back-to-back caps on green foliage

FALSE

TRUE

FALSE

A Corpus for Reasoning About Natural Language Grounded in Photographs

Alane Suhr*, Stephanie Zhou*, Ally Zhang, Iris Zhang, Huajun Bai, Yoav Artzi

First two authors contributed equally

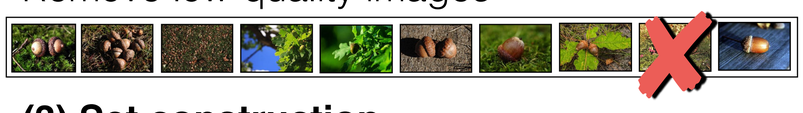
Data Collection

(1) Find images for ImageNet synsets

Use generated search queries

(2) Image pruning

Remove low-quality images

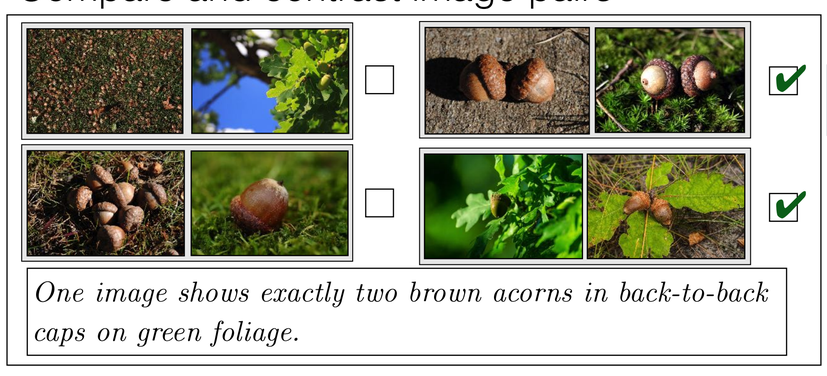


(3) Set construction

Construct sets of interesting images

(4) Sentence writing

Compare and contrast image pairs



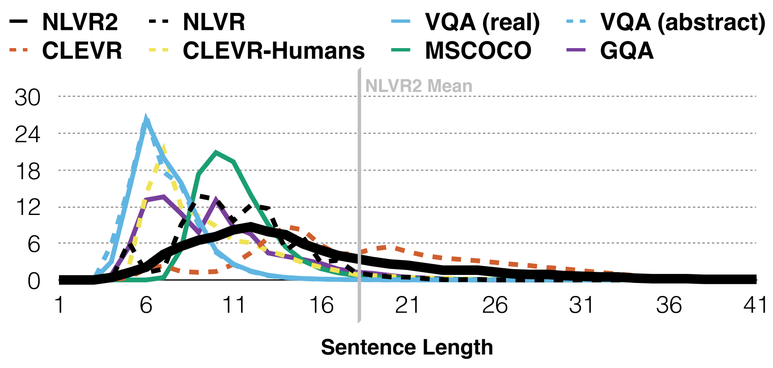
(5) Validation

Label with true or false

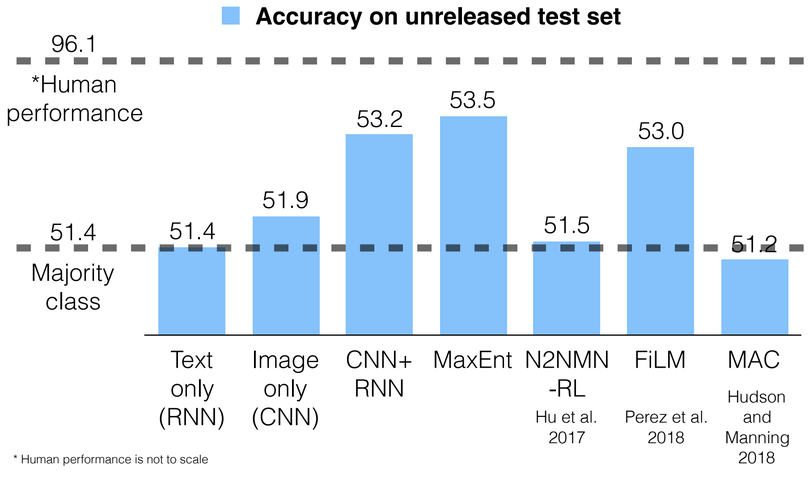
Dataset Statistics

Split	# Sentences	# Examples
Train	23,671	86,373
Dev	2,018	6,982
Public test	1,995	6,967
Private test	1,996	6,970
NLVR² Total	**29,680**	**107,292**
NLVR	**3,962**	**92,244**

- 124 synsets from ImageNet (Russakovsky et al. 2015)
- Agreement: 0.912 α ; 0.889 κ
- Vocabulary: 7,457 word types (NLVR: 252)
- Mean sentence length: 14.8 tokens (NLVR: 11.2)

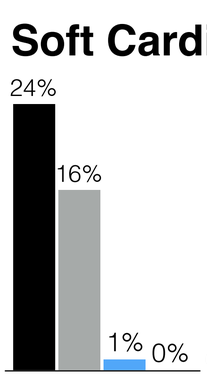


Results



Data Analysis

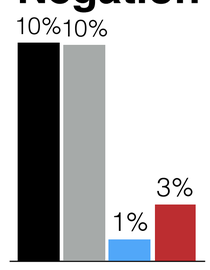
Prevalent linguistic phenomena: hard and soft cardinality, existential and universal quantifier, coordination, coreference, spatial relations, presupposition, preposition attachment ambiguity



Soft Cardinality

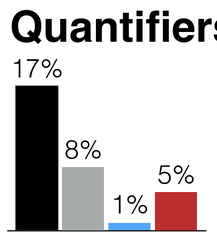
One image contains a single vulture in a standing pose with its head and body facing leftward, and the other image contains a group of <u>at least eight</u> vultures.

Negation



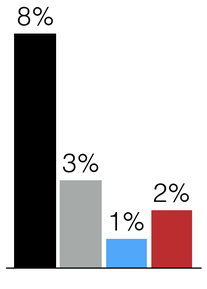
One dog sled team is moving and one is <u>not</u>

Universal Quantifiers



<u>All</u> the chairs have backs.

Comparisons



the left image has 4 balloons of all <u>different</u> colors