Visually Grounded Reasoning across Languages and Cultures

UNIVERSITY OF CAMBRIDGE
McGill UNIVERSITY
Fangyu Liu* Emanuele Bugliarello* Edoardo M. Ponti Siva Reddy Nigel Collier Desmond Elliott
Mila
UNIVERSITY OF COPENHAGEN

EMNLP 2021

Existing V&L data

Languages
- Mostly in English or in another IE language

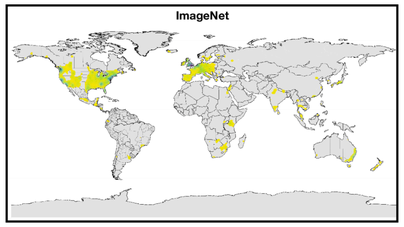
Limits of translation-based dataset creation

A street organ stands in front of a ...

An unusual looking vehicle parked ...

Example from van Miltenburg+ (INLG'17)

Image Sources
- Scraped from the web
- Reflect North American and European cultures

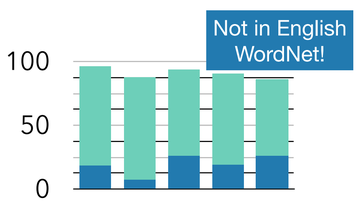


Density map of geographical distribution of images in ImageNet (DeVries+, CVPRW'19)

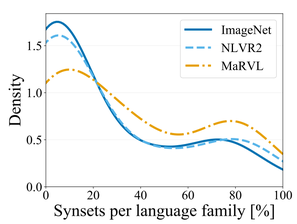
We should not simply translate existing V&L data!

Our Approach

- New protocol driven by **native speakers**
- Universal concepts from Intercontinental Dictionary Series



86-96 concepts per language



MaRVL concepts are in more ***families***

MaRVL: Multicultural Reasoning over Vision and Language

marvl-challenge.github.io

Evaluation data for cross-lingual V&L transfer

Task: Predict if a caption is True/False for 2 images

Languages: Indonesian (id) Swahili (sw) Tamil (ta) Turkish (tr) Mandarin Chinese (zh)

MaRVL-ta ஏறுதழுவல் **(Bull taming)**

இரு படங்களில் ஒன்றில் இரண்டிற்கும் மேற்பட்ட மஞ்சள் சட்டை அணிந்த வீரர்கள் காளையை அடக்கும் பணியில் ஈடுப்பட்டிருப்பதை காணமுடி.

In one of the two photos, more than two yellow-shirted players are seen engaged in bull taming.

Label: True

MaRVL-ta மை **(Vada)**

இரண்டு படங்களிலும் நிறைய மசால் வடைகள் உள்

Both images contain a lot of masala vadas

Label: False

Experiments

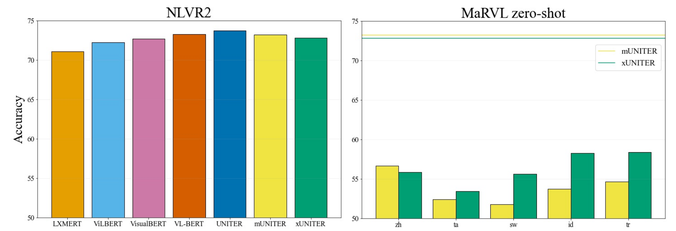
Models
- 5 English V&L BERTs from VOLTA (Bugliarello+, 2021)
- 2 new multilingual models: mUNITER & xUNITER

Fine-tuning

Train on English NLVR2 (Suhr+, 2019)

Test on MaRVL
- Multilingual models: zero-shot, cross-lingual
- English models: translate-test

Results



mUNITER and xUNITER are on par in NLVR2

Zero-shot transfer: -10-20% → chance-level!

Challenges
- Cross-lingual transfer (XLT)
- Out-of-distribution (OOD) generalisation