Visually Grounded Reasoning across Languages and Cultures

Fangyu Liu* Emanuele Bugliarello* Edoardo M. Ponti Siva Reddy Nigel Collier Desmond Elliott

EMNLP 2021

Existing V&L data

Languages
- Mostly in English or in another IE language

Limits of translation-based dataset creation

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

Existing V&L data

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)

Our Approach

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

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

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

Results

mUNITER and xUNITER are on par in NLVR2

MaRVL concepts are in more families

86-96 concepts per language

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

Translate MaRVL-zh into en -15%

Translate NLVR2 into zh

-8%

0% - chance-level!

Evaluate 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)

Results

mUNITER and xUNITER are on par in NLVR2

MaRVL concepts are in more families

86-96 concepts per language

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

Translate MaRVL-zh into en -15%

Translate NLVR2 into zh

-8%

0% - chance-level!

Evaluate 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)

Results

mUNITER and xUNITER are on par in NLVR2

MaRVL concepts are in more families

86-96 concepts per language

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

Translate MaRVL-zh into en -15%

Translate NLVR2 into zh

-8%

0% - chance-level!

Evaluate 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)

Results

mUNITER and xUNITER are on par in NLVR2

MaRVL concepts are in more families

86-96 concepts per language

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

Translate MaRVL-zh into en -15%

Translate NLVR2 into zh

-8%

0% - chance-level!