A COMPUTATIONAL ACQUISITION MODEL FOR MULTI MODAL WORD CATEGORIZATION

[1] The Hebrew University of Jerusalem

[2] University of Melbourne

Uri Berger[12], Gabriel Stanovsky[1], Omri Abend[1], Lea Frermann[2]

Outline: Self-supervised image and text clustering leads to syntagmatic categories and visual zero-shot abilities (object recognition without explicit training)

Motivation

1. Previous studies used **toy scenarios** or **pre-trained** visual encoders
2. How would categories look **without pre-training?**
3. How would they differ from **unimodal** categories?

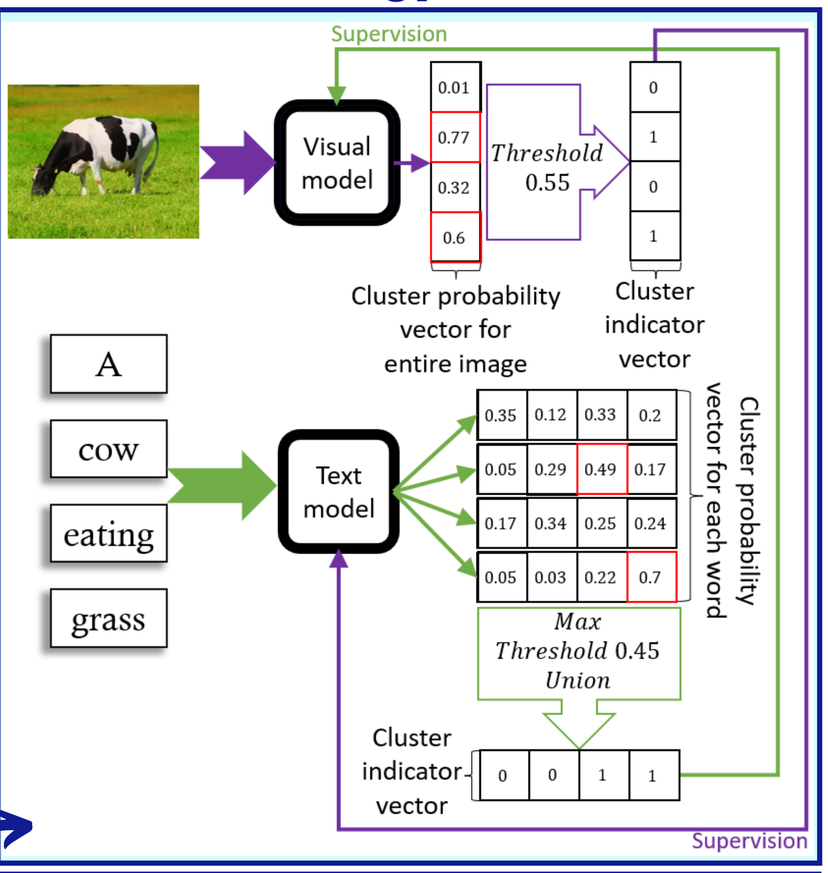
Category Types

Taxonomic:
Boat, Car
Swan, Monkee, Bee
Tree, Flower
Lake

Syntagmatic:
Boat, Swan, Lake
Monkey, Tree
Flower, Bee
Car

- Children shift from syntagmatic to taxononmic
- Visually impaired children prefer taxononmic

Methodology



Visual model (ResNet50 w/o pretraining):

- Text cluster indicator vector **supervises** the visual model
- Visual loss function **compares** visual probability vector to text cluster indicator vector

Text model (co-occurrence count):

- Visual cluster indicator vector **supervises** the text model
- By counting **word-clusters co-occurrence**

Training

Trained the model on **MSCOCO**, a dataset with image-caption pairs (captions created by human annotators):

- **56k images**
- **280k captions** (~5 captions per image)

Results

Categorization:

Model	Taxonomic (F-Score)	Syntagmatic (MAS)
Random	0.15 ± 0.0032	4.23 ± 1.88
Text-only	0.26 ± 0.0098	5.47 ± 0.25
Word2vec	$\mathbf{0.40 \pm 0.0172}$	6.65 ± 0.16
$\text{BERT}_{\text{BASE}}$	0.33 ± 0.011	5.75 ± 0.23
CLIP	0.38 ± 0.0142	7.08 ± 0.41
Ours	0.33 ± 0.0109	$\mathbf{7.45 \pm 0.33}$

Clusters:

Ours 1	skis; axe; sled; parka; sleigh; pants; gloves
Ours 2	sailboat; canoe; swan; raft; boat; yacht; duck; willow; ship; drum
W2V cluster	avocado, walnut, pineapple, grapefruit, coconut, olive, lime, lemon

Object detection:

Conclusion

- Categories are syntagmatic
- Multimodality shifts to syntagmatic
- Model acquired zero-shot abilities (object recognition without explicit training)

Contact

uri.berger2@mail.huji.ac.il

Personal website:

SCAN ME