



“Look, some green circles!” Learning to quantify from images

Ionut-Teodor Sorodoc, Angeliki Lazaridou, Gemma Boleda,
Aurélie Herbelot, Sandro Pezzelle and Raffaella Bernardi

{firstname.lastname@unitn.it}
CIMEC (University of Trento)

Motivation

1. Current models in both language and vision are very effective in representing the meaning of content words, but are less focused on function words like **natural language quantifiers** (*no, some, all*)
2. In grounded contexts children can provide quantification estimates before learn to count via Approximate Number System (**ANS**)

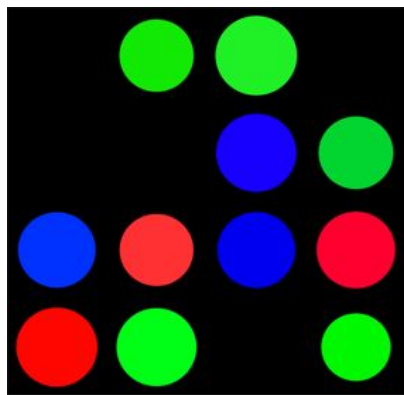
Hypothesis

We investigate whether a neural network model can learn the meaning of quantifiers **no**, **some**, and **all** from utterances grounded in vision

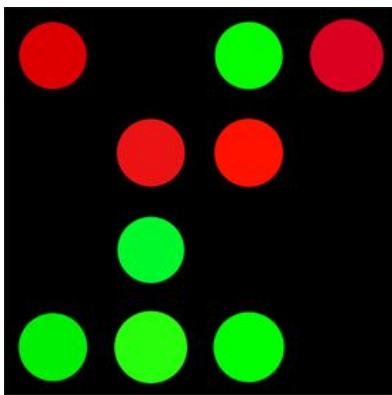
Consistently with human ANS, we hypothesize that quantifiers can be learnt from visual scenes and that **counting** is neither sufficient nor necessary for the acquisition of quantifiers

Task

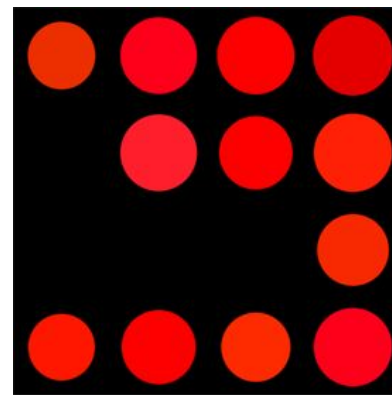
Given a set of objects (circles) with different properties (colors), the model learns to apply the correct quantifier to the scenario



some circles are green

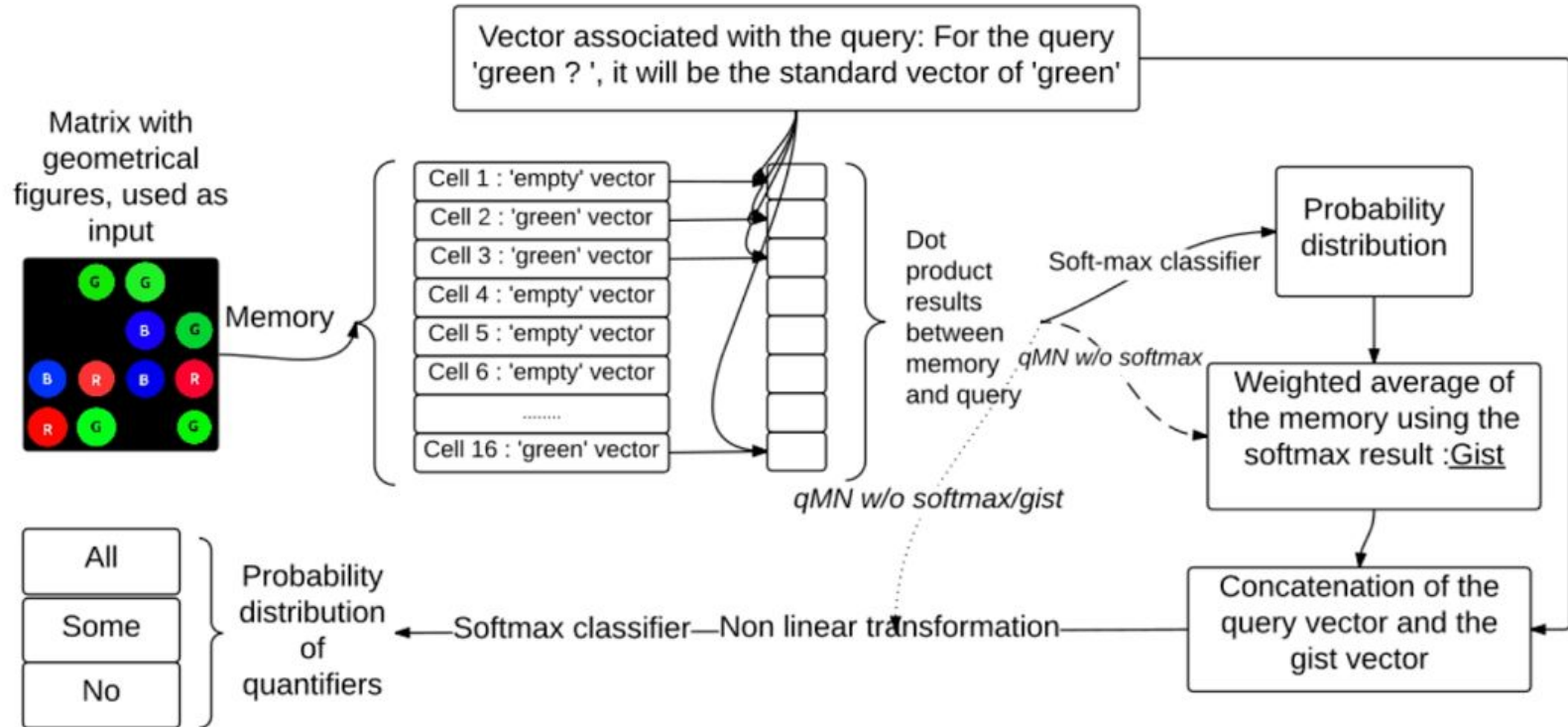


no circles are blue



all circles are red

Quantifier Memory Network Model (qMN)



Other Models

1. **Counting model:**

Model implemented to simulate serial counting of the cells

2. **Recurrent Neural Network:**

State-of-the art model to be tested against our qMN model

Experimental Setup

We test each model over 3 experimental setups:

1. **familiar**: 5K datapoints randomly split in train/val/test set
2. **unseen quantities**: no overlap train/test wrt number of objects in the image
3. **unseen colors**: train with 10 colors and test with 5 unseen colors

Results

Models	familiar	unseen quantities	unseen colors
RNN	65.7	62.0	49.7
Counting	86.5	78.4	32.8
qMN	88.8	97.0	54.9
-softmax	85.9	66.6	54.4
-softmax/gist	51.4	51.8	44.4

Conclusions

Counting is neither necessary nor sufficient to quantify over images

Current investigation is aimed at:

- extending the work to other quantifiers (*few*, *most*)
- experimenting with real images

Thank you for your attention!