Recurrence over feature maps improves **object enumeration** ability of CNNs.

Teaching CNNs to Count: A 2D Recurrent Approach for Artificial Number Awareness

Marcus A Werren, Anna S Bosman *(University of Pretoria, South Africa)*

1 Numerosity

- The **neurocognitive function** that provides us with **number awareness**.
- **Rapid** and **accurate** enumeration of a small set of objects (without counting) to a **high confidence** level.

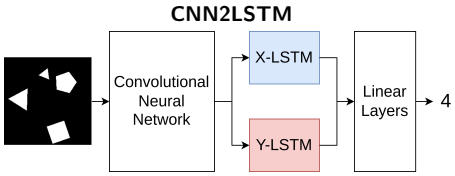
2 Proposed method

*Can **recurrent connections** establish generalised representations for **numerosity perception**?*

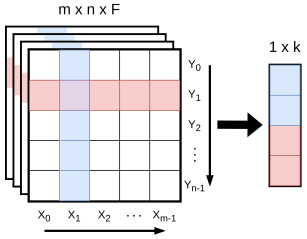
1. Process images into feature maps using a convolutional neural net (CNN).
2. Pass the feature maps to the **recurrent** component to create a numerosity representation.
3. Use resulting representation to predict numerosity for the visual stimuli.

3 Implementation

Recurrent deep learning architecture: output in the form of classification or regression.



Recurrent process: 2-dimensional long short-term memory **(2D-LSTM)** accepts CNN features as time series input.



m, n, F: feature map width, feature map height, number of feature maps
k: latent space dimension

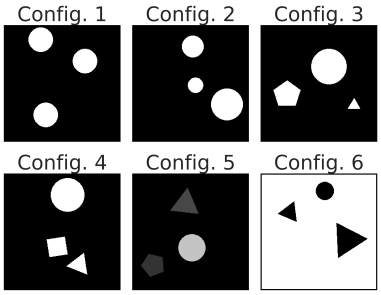
Resulting X-LSTM and Y-LSTM outputs are concatenated for the linear layer's input.

4 Results

Experiments were conducted on synthetic (Sec. 4.1) data and real-world (Sec. 4.2) data.

4.1 Synthetic Data

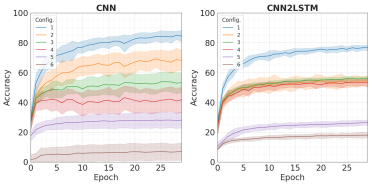
Different levels of **object complexity**:



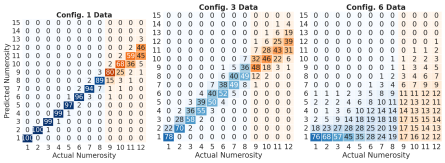
4.1.1 Config. 1: Generalisation

Each model was trained with config. 1 data only, and tested on all configurations.

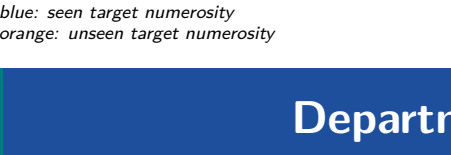
Test accuracy for each data configuration:



CNN - prediction accuracy



CNN2LSTM - prediction accuracy

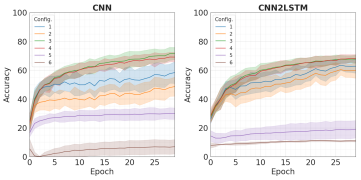


blue: seen target numerosity
orange: unseen target numerosity

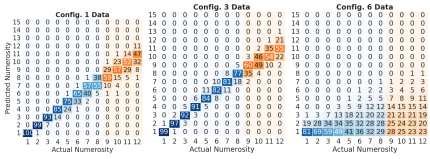
4.1.2 Config. 3: Generalisation

Each model was trained with config. 3 data only, and tested on all configurations.

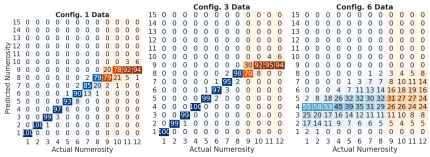
Test accuracy for each data configuration



CNN - prediction accuracy



CNN2LSTM - prediction accuracy

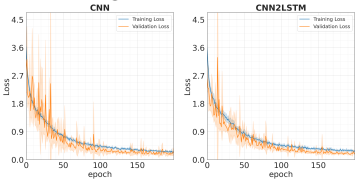


blue: seen target numerosity
orange: unseen target numerosity

4.2 Real-World Data

ResNet-18 was used as the CNN backbone. The models were trained with a custom **human-face** data set extracted from publicly available image data sets.

Training and Validation Loss



CNN - ScoreCAM Activation Maps

CNN2LSTM - ScoreCAM Activation Maps

CNN2LSTM produces more defined activation areas while reducing noise in earlier blocks.

5 Conclusion

When the maximum target numerosity is known, the CNN2LSTM models **improve test accuracy** and are **less sensitive to hyperparameter settings** than CNNs. Furthermore, recurrent connections **improve shape generalisation**.

Department of Computer Science

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and Information Technology
Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

Contact the author →

SCAN ME