CNN models' sensitivity to numerosity concepts

Neha Upadhyay
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
nupadhyay9@gatech.edu

Sashank Varma
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
varma@gatech.edu

Abstract

The nature of number is a classic question in the philosophy of mathematics. Cognitive scientists have shown that numbers are mentally represented as magnitudes organized as a mental number line (MNL). Here we ask whether CNN models, in learning to classify images, also learn about number and numerosity 'for free'. This was the case. A representative model showed the distance, size, and ratio effects that are the signatures of magnitude representations in humans. An MDS analysis of their latent representations found a close resemblance to the MNL documented in people. These findings challenge the developmental science proposal that numbers are part of the 'core knowledge' that all human infants possess, and instead serve as an existence proof of the learnability of numerical concepts.

1 Introduction

The past 10 years have seen great progress in computer vision models. Recent research is exploring the alignment of these models to behavioral and brain imaging data on human cognition Cichy and Kaiser [2019]. Here, we investigate the sensitivity of CNNs to number.

The nature of number is a classic question in the philosophy of mathematics dating back to Plato's dialogue *Meno*. The classic cognitive science finding is that numbers are represented in the mind as magnitudes akin to the sensory representations of physical quantities Moyer and Landauer [1967]. These magnitude representations are in turn organized as a *mental number line* (MNL; Figure 1).

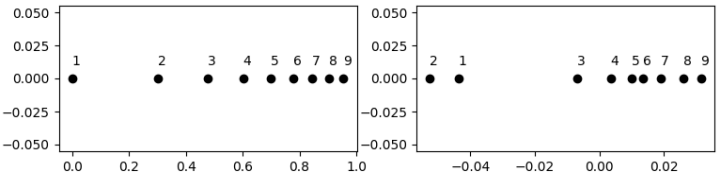

Figure 1: Mental number line representation in (left) human minds and (right) VGG19.

What is the origin of the MNL? A prominent position in developmental science is that magnitude representations of number are part of the 'core knowledge' that all human infants possess Spelke and Kinzler [2007]. An alternative hypothesis is that these representations are learned 'for free' as intelligent systems learn to perceive and navigate their environment. Here, we evaluate this latter hypothesis using CNNs. We propose that these models learn number representations as a 'side effect' of learning to classify images. We evaluate this proposal in experiments that constitute an existence proof of the learnability of numerical concepts without the need to posit core knowledge of number.

37th Conference on Neural Information Processing Systems (NeurIPS 2023) Workshop on MATH-AI.

1.1 Cognitive science evidence that numbers have magnitude representations

That numbers have magnitude representations is supported by multiple experimental findings. We focus on three important findings that have been documented using the number comparison paradigm. In this paradigm, people see a pair of digits (e.g., 1 vs. 3) or numerosities (i.e., sets of objects like 'o' vs. 'o o o'). The *distance effect* is the finding that the time to compare two numbers x and y decreases as the distance $|x - y|$ between them increases Moyer and Landauer [1967]. This is consistent with the following process model: fixate x and y on the MNL and discriminate which one is 'to the right'. The farther apart the two numbers, the easier the discrimination.

The *size effect* is the finding that the time to compare two numbers x and y increases as their average size $(x - y)/2$ increases Parkman [1971]. This suggests that the scaling (i.e., the distance between adjacent numbers) of the MNL is not fixed, as in the conventional number line of mathematics, but rather is psychophysically compressed, as it is for perceptual quantities; see Figure 1. Thus, for example, people are faster to discriminate which of 1 vs. 3 is 'to the right' than 7 vs. 9.

The *ratio effect* combines the distance and size effects: it is the finding that the time to compare two numbers x and y decreases as the ratio of the greater number over the smaller number increases, and this decrease is according to a nonlinear psychophysical function Halberda et al. [2008]. The presence of this effect is considered very strong evidence for magnitude representations of number.

1.2 Numerical sensitivities of computer vision models

Recent research has explored the mathematical capabilities of ML models. Much of this work has focused on the ability of NLP mdels to solve arithmetic, algebra, trigonometric, and calculus problems Welleck et al. [2022] and also to generate proofs in higher-level mathematics (i.e., discrete math, probability, linear algebra, abstract algebra, real analysis, topology) Davies et al. [2021].

Less attention has been paid to the mathematical capabilities of computer vision models. This is perhaps because they are a poor fit for symbolic mathematics. Early research explored pre-CNN models trained on artificially generated numerosity stimuli Stoianov and Zorzi [2012] Zorzi and Testolin [2018] found that such models showed the ratio effect; further analysis of the hidden layers found units tuned to specific numerosities. Subsequent work generalized these findings to networks trained on numerosity images abstracted from naturally occurring images Testolin et al. [2020].

Other researchers have evaluated the number representations of CNN models trained on ImageNet Kim et al. [2021] Nasr et al. [2019]. They have found hidden layer units tuned to specific numerosities. When these representations are used as inputs to a separate network trained to decide whether two images are of the same numeoristy or not, that network shows the distance and size effects.

With respect to the MNL, it has recently been shown that a vision transformer model trained on artificially generated numerosity stimuli learns a latent MNL representation Boccato et al. [2021].

1.3 Research Questions

We investigate whether CNNs learn magnitude representations of number. We present numerosities to the models in a sequence of experiments differing in which incidental visual features are controlled and which are allowed to vary freely (and potentially correlate with number):

1. The items of the two numerosities are circles. The total area of the numerosities is equated to rule out this visual feature as the basis of comparison.
2. Like (1) but the total circumference is equated to rule out this visual feature.
3. Like (2) but the items of the two numerosities are different (e.g., two of the three of circles, squares, triangles) to generalize the findings across shapes.
4. Like (3) but the total area of each numerosity is different to generalize the findings across both shapes and area.
5. Like (4) but the items of each numerosity are random shapes of random area to further generalize the findings.
6. Naturally occurring numerosities found through Google Images that differ on many visual attributes (e.g., shape, size, drawing style, color, etc.) to ensure further generalization.

We evaluate whether the models show the signatures of magnitude representations: the distance, size, and ratio effects. We also explore whether the models learn a latent representation of the MNL.

2 Methods

2.1 Models

We used the pre-trained VGG19 Simonyan and Zisserman [2015] model from the PyTorch model zoo as our primary model. We evaluated the generalizability of our findings by replicating all analyses with the pre-trained Alexnet, Googlenet, Densenet, and Resnet18 models.

2.2 Numerosity stimuli

The stimuli for the six experiments are as described in the six research questions above; see Figure 2 for examples. They spanned numerosities from 1 to 9.

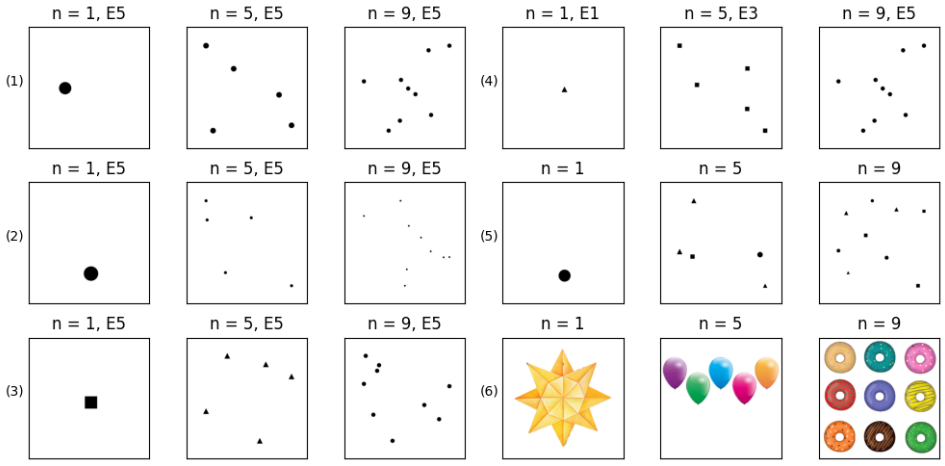

Figure 2: For the 6 experiments, sample stimuli for 3 numerosities.

For the Experiments 1-5, the stimuli were solid black shapes randomly placed on a white background of 720×720 pixels generated using matplotlib. For Experiment 1, the items were circles and the total area of each comparison was equated. Specifically, for each of 5 total areas, we generated 4 stimuli for each of the numerosities 1-9. The total area of a stimulus varied from $A1 = .02\%$ to $A5 = .10\%$ of all pixels in increments of .02%. For Experiment 2, the total circumference of a comparison was equated. The stimuli were generated similarly to Experiment 1 except that the total circumference of a stimulus was varied from $C1 = 100$ to $C5 = 300$ pixels in increments of 50 pixels. Experiment 3 was like Experiment 1 except the items of the two stimuli were different, e.g., one might be circles and the other either squares or triangles. Experiment 4 was like Experiment 3 except the total areas of the two stimuli varied randomly. Experiment 5 was like Experiment 4 except that the items of each stimulus varied randomly both in shape and in area. For Experiment 6, we automatically collected 80-100 images from Google Images for each numerosity 1-9. We stripped the background and manually verified each image. We retained a subset of 40 images per numerosity that clearly showed the target quantity.

2.3 Generating Model predictions

For each experiment, we normalized and resized each stimulus image. For two unequal numerosities each in the range $1 - 9$, there are $(9 \times 8)/2 = 36$ possible comparisons. Denote the numerosities of a comparison as $n1$ and $n2$. We randomly selected stimuli of numerosity $n1$ and $n2$, presented each to the CNN, captured their respective vector representations on the final fully connected layer, and computed the cosine similarity of the two vectors. We repeated this process $M = 20$ times for the first three experiments, $M = 40$ times for Experiment 4, and $M = 60$ times for Experiments 5 and 6, increasing M with increasing noise in the stimulus. Finally, we computed the average cosine similarity when comparing $n1$ and $n2$.

We used these values to estimate the three effects. For the distance effect, we plotted the average cosine similarity (y) at each distance $|n1 - n2|$ (x) and computed the correlation between these two

variables. A distance effect is signalled by a negative correlation close to $r = -1$. We repeated this process for the size effect, with the x variable the size $|n1 - n2|/2$ of the comparison; here, the expectation was for a positive correlation close to $r = 1$. We also repeated this process for the ratio effect, with the x variable the ratio $max(n1, n2)/min(n1, n2)$ of the comparison. Here, we fit a negative exponential function to model the results; the expectation was for an R^2 value close to 1.

3 Results

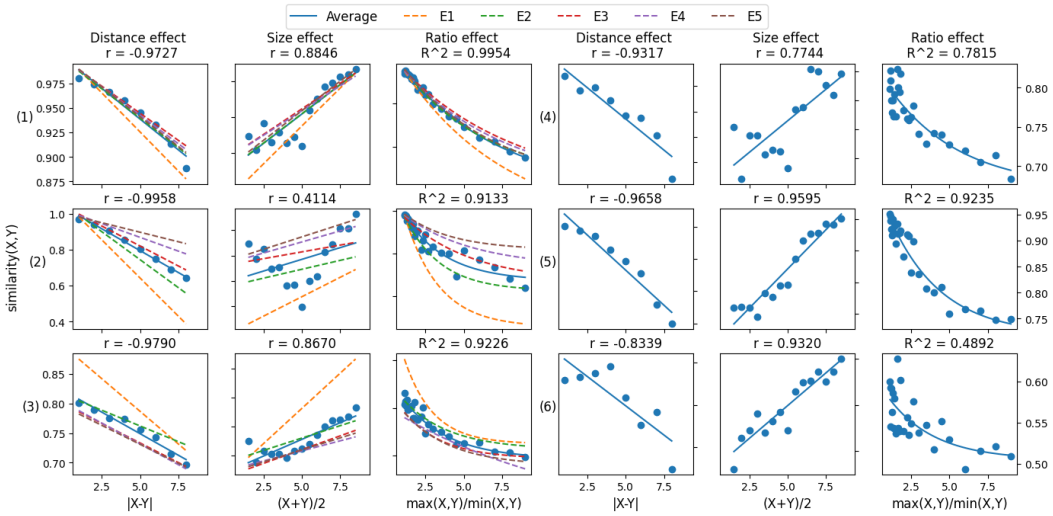

Figure 3: For the 6 experiments, the distance, size, and ratio effects.

Figure 4 presents graphs for all VGG19 experiments showing the distance, size, and ratio effects. The panels correspond to the sample stimuli displayed in Figure 2.

As the graphs show, VGG19 generally displayed the distance, size, and ratio effects across all experiments. Two patterns are worth noting. First, the model's account of the size effect is weaker than its account of the other two effects. This is particularly true of Experiments 2 and 4. That said, it is the distance effect that is the key finding for a MNL, and the ratio effect subsumes both the distance and size effects. Second, the model's distance and ratio effects are weakest for Experiment 6. These stimuli were images returned from a Google Image search, and they naturally varied on many more incidental visual features than the stimuli of the other experiments. The reduced fits may signal the limits of the model's representation of numerosity.

Finally, we estimated the latent MNL representation of VGG19 by organizing the pairwise average cosine similarities of the numerosities $1 - 9$ in Experiment 1 in a matrix, submitting this to MDS, and requesting a 1D solution; see Figure 1. The Stress-I value was 0.008, indicating that a continuum offers a good of these similarities. The representation resembles that of the psychophysically compressed MNL, with the major distortion being the displacement of '2'.

4 Discussion

The current study investigated whether CNN models, in learning to classify images, learn about number and numerosity 'for free'. This was the case: VGG19 showed the distance, size, and ratio effects that are the signature of the magnitude representations in humans. This was true across five experiments using artificially generated stimuli and a sixth using naturally occurring numerosities found via Google Images. In addition, an MDS analysis of the latent representation of the Experiment 1 stimuli found a close resemblance to the psychophysically scaled MNL documented in people.

That CNNs trained on ImageNet learn magnitude representations of number is at odds with the *core knowledge* proposal, which states that the MNL is part of biological endowment of the child Spelke and Kinzler [2007]. This finding is more consistent with the *emergentist* perspective, which

claims that this representation arises from the interplay between neural architecture constraints, domain-general learning mechanisms, and structured environments Zorzi and Testolin [2018].

In ongoing work, we are identifying the earliest layer of CNNs where magnitude representations first manifest as evidenced by distance, size, and ratio effects and by a latent MNL. We are also exploring the numerical sensitivities of state-of the art vision transformer models Boccato et al. [2021].

The current research sets the stage for investigations of the *development* of magnitude representations. The developmental progressions of the distance, size, and ratio effects have been documented in children Halberda et al. [2008] Sekuler and Mierkiewicz [1977]. An interesting question is whether computer vision models show the same progressions in their number representations over training.

References

Tommaso Boccato, Alberto Testolin, and Marco Zorzi. Learning numerosity representations with transformers: Number generation tasks and out-of-distribution generalization. *Entropy*, 23(7):857, 2021.

Radoslaw M Cichy and Daniel Kaiser. Deep neural networks as scientific models. *Trends in cognitive sciences*, 23(4):305–317, 2019.

Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. *Nature*, 600(7887):70–74, 2021.

Justin Halberda, Michèle MM Mazzocco, and Lisa Feigenson. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455(7213):665–668, 2008.

Gwangsu Kim, Jaeson Jang, Seungdae Baek, Min Song, and Se-Bum Paik. Visual number sense in untrained deep neural networks. *Science advances*, 7(1):eabd6127, 2021.

Robert S Moyer and Thomas K Landauer. Time required for judgements of numerical inequality. *Nature*, 215(5109):1519–1520, 1967.

Khaled Nasr, Pooja Viswanathan, and Andreas Nieder. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science advances*, 5(5):eaav7903, 2019.

John M Parkman. Temporal aspects of digit and letter inequality judgments. *Journal of experimental psychology*, 91(2):191, 1971.

Robert Sekuler and Diane Mierkiewicz. Children's judgments of numerical inequality. *Child Development*, pages 630–633, 1977.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental science*, 10(1):89–96, 2007.

Ivilin Stoianov and Marco Zorzi. Emergence of a'visual number sense'in hierarchical generative models. *Nature neuroscience*, 15(2):194–196, 2012.

Alberto Testolin, Will Y Zou, and James L McClelland. Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Developmental science*, 23(5):e12940, 2020.

Sean Welleck, Peter West, Jize Cao, and Yejin Choi. Symbolic brittleness in sequence models: on systematic generalization in symbolic mathematics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8629–8637, 2022.

Marco Zorzi and Alberto Testolin. An emergentist perspective on the origin of number sense. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1740):20170043, 2018.