Counting Dots: On Learning Numerical Concepts from Visual Data

Andrzej Śluzek[0000−0003−4148−2600]

Institute of Information Technology,
Warsaw University of Life Sciences-SGGW, Poland
andrzej_sluzek@sggw.edu.pl

Abstract. The paper reports selected ideas and preliminary results on NN-based learning simple numerical properties from visual data. Specifically, we use datasets of small images with single-pixel dots (*one* to *ten* dots in each image) to visually estimate integers (ranging from 1 to 10), to distinguish between *even* and *odd* numbers, and to identify some other numerical properties. Small fully-connected NN's and convolutional NN's are used as learning architectures. The obtained results are inconclusive. First, it seems that FCNN architectures are superior to CNN's, which are apparently hardly able to learn the abstractions of numbers. Surprisingly, such conclusions contradict results presented in a recently published paper on similar topics. Secondly, the concepts of *even* and *odd* numbers cannot be learned directly from visual examples, even by FCNN's. We preliminarily hypothesize that basic numerical properties can or cannot be learned by simple NN architectures depending on whether the property divides the considered set of integers into connected subsets (e.g. 1, 2 and larger numbers) or disjointed ones (e.g. even *versus* odd numbers). Nevertheless, the obtained results are still considered preliminary, and they require further theoretical analysis and experimental verification.

Keywords: Numerical Abstractions · Learning · Neural Networks · Visual Data.

1 Introduction and Background

Counting is one of the first abstract concepts developed by children, e.g. [8]. Intuitively, and based on various publications (e.g. [4, 3]) this concept develops from sensory experiences and physical embodiments, with visual inspection playing the vital role.

Machine learning community has been interested in such issues for a long time. Counting and understanding numerical concepts from visual data is an interesting and challenging topics for AI algorithms. Initially, the majority of papers were focused on automatic object counting rather than on 'understanding' the abstractions of numbers, with significant efforts on *counting-by-localizing* sub-tasks, e.g. [5, 7]. This was, obviously, more application-oriented and various sources (e.g. [2, 6]) indicate that true understanding of numbers may not be needed to automatically perform counting tasks.

This paper follows another approach (presented in fewer papers) where visual complexity of training/validation data is reduced, and the focus is on learning the concept of numbers (including abilities to estimate the numbers of objects). In particular, we follow and expand the approaches from [1], where two models are proposed for learning counting (up to 10) in small black images containing a number of white non-overlapping rectangles.

The authors used a standard feed-forward CNN (of very low complexity) and a deep believe network(DBN) model, reporting superior performances of the former, and providing details of the implemented architecture. Compared to that paper, we investigate similar problems. However, other visual datasets and alternative architectures are used. More diversified numerical concepts are considered, and (notably) we obtain results which partially contradict [1].

2 Methodology and Objectives

2.1 Datasets

For the conducted experiments, we generated a simple dataset of 20,000 images. As shown in Fig. 1, image resolution is 10×10 with a number of isolated single-pixel dots over backgrounds of opposite intensities. The images are not binary to better represent the real-world visual conditions (and to minimize effects of overfitting on binary data).

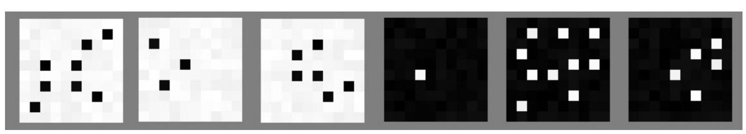


Fig. 1. Exemplary dataset images.

2.2 Tasks

Several learning tasks have been attempted by training selected NN architectures of the dataset. Actually, the dataset is divided into two equal halves, and training/validation is always performed on an arbitrarily selected half (the other one used for extensive testing). The tasks (which are learned separately and independently from the provided visual dataset) are as follows:

a. Learning integers from 1 to 10.
b. Learning even and odd numbers (from the above range).
c. Learning the numbers 1, 2 and larger.

The acquired abilities would be used for estimating the number (or number category) of dots in test images.

2.3 NN Architectures

Following the results reported in [1], we initially tested small CNN's proposed there (with minor changes reflecting differences between the datasets). Eventually, even simpler CNN architectures were used, with much better results.

Nevertheless, because of unsatisfactory performances of CNN's (details in Section 3), fully connected NN's (with correspondingly low complexity) were used as the ultimate choice. For either CNN's and FCNN's, various variants were attempted. Table 1 provides examples of top-performance architectures in each category.

Table 1. Examples of tested architectures (CNN based on [1], our CNN and FCNN).

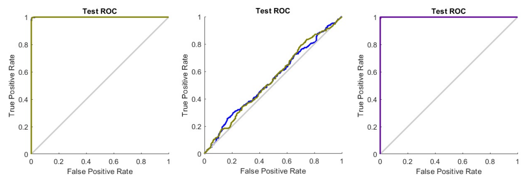
(A) CNN (based on [1])			(B) CNN (our)			(C) FCNN		
Layer	Parameters	Activation	Layer	Parameters	Activation	Layer	Parameters	Activation
input	10×10		input	10×10		input	100	
conv.	3×3, str=1	relu	conv.	2×2, str=2	relu	fc	10	tanh
maxpool	2×2, str=2		maxpool	1×1, str=1		fc	5	tanh
conv.	1×1, str=1	relu	fc	100	tanh	fc	5	tanh
fc	90	tanh	fc	10 (or 2)	softmax	fc	10 (or 2)	softmax
fc	10 (or 2, 3)	softmax	output	10 (or 2, 3)		output	10 (or 2, 3)	
output	10 (or 2)							

3 Experimantal Results

Training of all architectures was performed on 70% of the selected half of the dataset (the remaining 30% used for validation), while the other half of the datatest was used for testing.

Unfortunately, neither CNN architectures similar to Table 1(A) nor CNN architectures similar to Table 1(B) show satisfactory performances for Task(a), i.e. counting from 1 to 10. Accuracy of the former ones is comparable to random choice (10% on both training and testing data). The latter architectures perform better. On training data, the obtained accuracy can exceed 90%, but for validation and testing data we could never get accuracy above 62%. Thus, the experimental results (primarily in a form of ROC curves) are displayed only for FCNN architecture (using the exemplary architecture from Table 1(C)).

As shown in Figs 2(a-d), FCNN almost perfectly handles Task(a) and, in particular, Task(c) (see Section 2.2) but fails to learn the concepts of even and odd numbers (Task(b)). For Task(a), accuracy for test data is 97.34%, while for Task(c) it reaches 99.88%.



(a) (b) (c)

Test confusion matrix (Output class × Target class)

995	6	0	0	0	0	0	0	0	0
5	992	1	0	0	0	0	0	0	0
0	2	990	6	0	0	0	0	0	0
0	0	9	987	6	0	0	0	0	0
0	0	0	7	981	10	0	0	0	0
0	0	0	0	13	985	9	0	0	0
0	0	0	0	0	5	983	12	0	0
0	0	0	0	0	0	8	986	71	0
0	0	0	0	0	0	0	2	836	1
0	0	0	0	0	0	0	0	93	999

(d)

Fig. 2. ROC plots for recognition of numbers: (a) from 1 to 10, (b) *odd* or *even* numbers, (c) numbers 1, 2 or larger. In (d), the confusion matrix (test data) for Task(a) is given.

The results, in spite of very low complexity of applied learning mechanisms, very well correspond to the actual numerical skills of small children (e.g. [8]). Counting *one*, *two*, *many* is the easiest task, counting to 10 comes later (and is more error-prone, with most errors for larger numbers), while understanding the concept of even and odd numbers is seldom achieved before the primary school.

Nevertheless, the obtained results are considered preliminary, and they require further theoretical analysis and experimental verification.

References

1. Creatore, C., Sabathiel, S., Solstad, T.: Learning exact enumeration and approximate estimation in deep neural network models. Cognition **215**, 104815 (2021). https://doi.org/10.1016/j.cognition.2021.104815
2. Fang, M., Zhou, Z., Chen, S., McClelland, J.L.: Can a recurrent neural network learn to count things? Cognitive Science (2018)
3. Fischer, M.H., Shaki, S.: Number concepts: abstract and embodied. Phil. Trans. Royal Society B **373**(1752), 20170125 (2018). https://doi.org/10.1098/rstb.2017.0125
4. Kersey, A.J., Cantlon, J.F.: Primitive concepts of number and the developing human brain. Language Learning and Development **13**(2), 191–214 (2017). https://doi.org/10.1080/15475441.2016.1264878
5. Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems - Vol.1. p. 1324–1332. NIPS'10 (2010)
6. Sabathiel, S., McClelland, J.L., Solstad, T.: Emerging representations for counting in a neural network agent interacting with a multimodal environment. vol. ALIFE 2020: The 2020 Conference on Artificial Life, pp. 736–743 (07 2020). https://doi.org/10.1162/isal_a_00333
7. Walach, E., Wolf, L.: Learning to count with cnn boosting. In: Proceedings of the 14th European Conference on Computer Vision, part II. vol. LNCS 9906, pp. 660–676 (10 2016). https://doi.org/10.1007/978-3-319-46475-6_41
8. Wynn, K.: Children's understanding of counting. Cognition **36**(2), 155–193 (1990). https://doi.org/10.1016/0010-0277(90)90003-3