Learning to solve symbolic math from visual inputs

Thao Phung
University of Wyoming
phungpthao@gmail.com

Anh Nguyen
Auburn University
anh.ng8@gmail.com

Jeff Clune
Uber AI Labs, University of Wyoming
jeffclune@uwyo.edu

Abstract

Deep neural networks (DNNs) have obtained the state-of-the-art performance on many challenging tasks by learning from only visual inputs such as recognizing and detecting objects from images. The main reason behind DNNs' impressive performance is their ability to automatically extract a hierarchy of abstract features from the inputs that are useful to solve a given problem [2].

Given their impressive pattern recognition capability, can DNNs learn to extract the meanings behind visual symbols? In this research, we explore approaches to train DNNs to solve addition and subtraction problems from images of equations (Fig. 1, right column). This is a challenging task that requires DNNs to learn the underlying values represented by these digit and operator symbols in order to produce the correct answers to the held-out problems (e.g. finding the answer 5 to the equation 2 + 3 = ?).

We show that by training a state-of-the-art DNN on addition images (e.g. 1 + 2) together with their correct labels (here, 3), the DNN predicts correctly 98% of the time on new images *of the equations it had already seen (Fig. 1a & b). However, the DNN performed poorly at 15% accuracy (slightly above random chance) on the equations that it had never seen before (Fig. 1c). This suggests that the DNN only memorizes the visual patterns of the equation images but does not realize the actual values of the numbers to perform addition. Specifically, we teach DNNs to perform addition and subtraction via: (1) curriculum learning i.e. an organized training strategy in which we present a series of gradually increasingly complex problems [1] and (2) indirect learning i.e. training a separate DNN to paint the weights for the main DNN that takes in an equation image as input and outputs the predicted answer [3].*

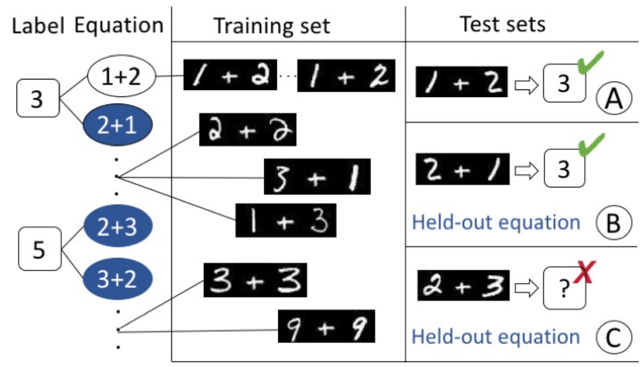

Figure 1: We train a DNN on images of addition problems (e.g. $1 + 2 = ?$; note the training images only contain the left-hand side $1 + 2$ of a full equation) and ask the DNN to output the correct answer. Images are made up of two handwritten digits (taken from the MNIST dataset) and a plus "+" sign. We produce 2000 images for each equation (top row, 3 example images of $1 + 2$), and include in the dataset all combinations of addition problems with two digits (e.g. $0 + 0$, $0 + 1$, ... $9 + 9$).
(A) As expected, the DNN performs well at recognizing new, unseen images of the $1 + 2$ problem that exists in the training set.
(B) Interestingly, the DNN can also correctly solve the $2 + 1 = ?$ problem when the training set only contains images of $1 + 2$ (not its commutative version $2 + 1$). This can be explained as the DNN learns to output "3" whenever it detects the visual patterns of "1", "+", and "2" in the image.
(C) However, when the commutative version is also held-out together with the original problem (e.g. both $2 + 3$ and $3 + 2$), the DNN is not able to correctly predict the answer. To solve this problem, the DNN would have to "understand" the underlying values of digits and symbols. We are exploring different strategies of harnessing curriculum learning [1] and indirect learning [3] to solve this challenging task.

References

[1] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009. 1

[2] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT Press, 2016. 1

[3] D. Ha, A. Dai, and Q. V. Le. Hypernetworks. arxiv preprint. *arXiv preprint arXiv:1609.09106*, 2016. 1