Neuro-Symbolic Multi Digit Arithmetic

Ruan Mackridge 23846879

March 2023

Abstract

This paper shows the results of an evaluation of Neuro-Symbolic AI models on different arithmetic operations on MNIST-images. The evaluation compares an end-to-end Neuro-Symbolic model to a segmented Neuro-Symbolic model and a baseline convolutional neural network. The results of the evaluation show that the end-to-end Neuro-Symbolic model outperforms the other two models. This suggests that Neuro-Symbolic AI models are a promising approach to solving problems that require both symbolic reasoning and classification.

1 Introduction

In recent years, the field of artificial intelligence has had a surge of advancements and popularity. This has led to an AI revolution across many industries, including healthcare, finance, and manufacturing. It is predicted that AI technologies could increase global GDP by $15.7 trillion by 2030 [10]. However, despite these advancements, there are still many challenges that need to be addressed in order to achieve human-level intelligence. This task proves difficult because humans possess superior pattern processing [8] and reasoning capabilities [3]. A challenge in the development of AI is the integration of symbolic reasoning with neural networks, which has been a long-standing goal in the field of artificial intelligence.

The two primary paradigms to creating artificial intelligence are symbolic artificial intelligence (classical AI) and connectionist artificial intelligence (neural networks)[6]. Symbolic AI is a rule-based approach that that uses high-level reasoning to solve problems. This approach is most effective when given problems that have clear rules and logic to follow, but performs poorly when given problems that require pattern recognition or that have noisy data. Connectionist AI is a neural network-based approach that uses pattern recognition to solve problems. This approach lets the system learn the rules and logic from the data rather than having it explicitly programmed. This approach is highly effective when given problems that require pattern recognition or that have unstructured data. The integration of symbolic reasoning with neural networks offers the potential to combine the strengths of both paradigms, providing enhanced interpretability, explainability, and generalization capabilities in AI models [6].

AI has been shown to have some failings stemming from its inability to reason. Leading to undesired behaviors like mentioned in the paper by Chanda et al. [2].

This paper evaluates the performance of Neuro-Symbolic AI models on the task of MNIST-digit addition, subtraction, less-than, greater-than and multiplication. Demonstrating the potential of Neuro-Symbolic AI in solving real-world problems that require a combination of symbolic reasoning and neural network-based image processing techniques. To perform the evaluation an end-to-end Neuro-Symbolic AI model is compared with two baseline models. The first baseline model is a standard convolutional neural network (CNN) and the second baseline model is a segmented Neuro-Symbolic AI model that consists of a CNN which classifies the images, and then a probabilistic logic operation which performs an operation on the two images.

This paper finds that the end-to-end Neuro-Symbolic AI model outperforms the other two models consistently. Suggesting that combining symbolic reasoning with neural networks is a promising approach to solving problems. And could increase the performance of AI models in real-world applications.

The paper is organized as follows, Section 2 gives a brief overview of the concepts that are used in this paper. It also gives a brief overview of the related work that has been done in this field. Section 3 describes the primary approach evaluated in this paper. Section 4 describes the empirical process that was used to evaluate the primary approach. Section 5 presents the results of the evaluation of the model. Section 6 concludes the paper and Section 7 discusses future work.

2 Background

This section will give a brief overview of the concepts that are used in this paper. It also gives a brief overview of the related work that has been done in this field. Section 2.1 gives an overview of Symbolic AI. Section 2.2 gives an overview of Convolutional Neural Networks. Section 2.3 gives an overview of Neuro-Symbolic AI. Section 2.4 gives an overview of the related work that has been done in this field.

2.1 Symbolic AI

Symbolic artificial intelligence is a collection of AI approaches that are rule-based and contain high-level reasoning [9]. Symbolic AI is designed to imitate the way humans reason. This was seen as one of the main ways forward for artificial intelligence during the late 20th century [9]. This approach to AI uses logic and symbols to make decisions similar to how humans would. It is most effective when given problems where there are clear rules and logic to follow. Symbolic AI is not effective when the input that it receives is noisy and difficult to classify into rule sets. Regrettably for Symbolic AI, this tends to be the situation when dealing with real-world data, which often contains noise and defies easy categorization.

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a class of artificial neural networks that are well adapted to solve computer vision tasks. CNN are composed of convolutional layers, pooling layers, and fully connected layers. In the case of image classification, the convolutional layers and pooling layers extract features from the given images after which the fully connected layers map the features to an output. The output usually represents the prediction of the labels for the images [11].

2.3 Neuro-Symbolic AI

Neuro-Symbolic AI is a combination of Symbolic AI and neural networks. It is simple to see how Symbolic AI and neural networks can complement each other. In the specific problem presented in this paper, Symbolic AI is ineffective when given unstructured data and the CNN's excel in classifying this type of data, like images. Whereas CNN is ineffective when given problems that require high-level reasoning and Symbolic AI is effective in these situations. This led to the dawn of Neuro-Symbolic AI. An approach that takes the classification of a CNN and feeds it into a symbolic AI program. This approach allows the Symbolic AI to be able to work with non-perfect data and it makes the AI more comprehensible [5]. An example of this combined approach is DeepProbLog, proposed by Manhaeve et al.[7].

2.4 Related work

The concept of Neuro-Symbolic AI, which combines neural networks with logical and probabilistic reasoning, has gained significant attention in recent years. De Raedt et al. [4] proposed a Neuro-Symbolic framework that integrates symbolic logic and probabilistic reasoning with neural networks, offering a holistic approach to artificial intelligence. The authors argued that this integration allows for the benefits of both symbolic and connectionist approaches, providing improved interpretability, explainability, and generalization capabilities in AI models. The Neuro-Symbolic framework proposed by De Raedt et al. has been widely cited in the literature and has spurred further research in the field of Neuro-Symbolic AI, paving the way for novel approaches and applications that leverage the synergy between neural, logical, and probabilistic reasoning in artificial intelligence.

DeepProbLog, proposed by Manhaeve et al. [7], is a notable example of a Neuro-Symbolic AI framework that combines neural networks with probabilistic logic programming. DeepProbLog extends the traditional Probabilistic Logic Programming (PLP) paradigm by incorporating neural networks as differentiable modules, allowing for joint learning of symbolic and sub-symbolic representations. The framework enables the integration of declarative logical reasoning with deep learning techniques, offering the potential to combine the strengths of symbolic reasoning, probabilistic modeling, and neural networks. DeepProbLog has been applied to various tasks, such as probabilistic program induction, learning from incomplete data, and relational reasoning. The introduction of DeepProbLog has opened up new avenues for research in Neuro-Symbolic AI, providing a powerful and flexible framework for developing AI models that integrate neural networks with probabilistic logic programming for enhanced interpretability, explainability, and learning capabilities.

The work by Augustine et al. [1] on Visual Sudoku Puzzle Classification presents a unique application of Neuro-Symbolic AI in the context of solving Sudoku puzzles using a collective neuro-symbolic approach. The authors propose a suite of collective neuro-symbolic tasks that involve integrating symbolic reasoning with neural networks for solving Sudoku puzzles, a popular logic-based game. The approach involves a combination of rule-based symbolic reasoning with neural networks for image processing and pattern recognition. The authors demonstrate the effectiveness of their Neuro-Symbolic AI approach in accurately classifying Sudoku puzzles as solvable or unsolvable, showcasing the potential of Neuro-Symbolic AI in solving complex real-world problems that require a combination of symbolic reasoning and neural network-based image processing techniques. This work expands the application of Neuro-Symbolic AI into the domain of visual puzzle solving and highlights the versatility of this approach in diverse problem domains beyond traditional image classification and sequence prediction tasks.

3 Neural and Neural-Symbolic Approaches to MNIST-digit operations

This section discusses the approach that is evaluated in this paper and the two baseline models that are used for comparison. The primary approach is an end-to-end Neuro-Symbolic AI model and is discussed in Section 3.1. The first baseline model is a standard CNN and is discussed in Section 3.2. The second baseline model is a segmented Neuro-Symbolic AI model and is discussed in Section 3.3.

3.1 End-to-End Neuro-Symbolic AI model

The primary approach that is evaluated in this paper is an End-to-End Neuro-Symbolic AI model (E-ENS). This model is based on the DeepProbLog model proposed by Manhaeve et al. [7]. This approach consists of a CNN which has two convolutional layers, two max-pooling layers, three linear layers, and a softmax layer. The forward function of this network has been modified to incorporate the probabilistic logic operation. Explicitly defining the operation that the model is trying to learn. It does this by creating a matrix that has a value of 1 if the operation on the two images equals the target label and a value of 0 if the operation on the two images does not equal the target label. This matrix is then multiplied by the output of the softmax layer. This ensures that the model starts to learn the operation.

This approach is expected to perform better than the other two models as it can learn how the operation works and then use this to aid in performing better. It utilizes the strengths of both the CNN and the probabilistic logic operation. The CNN is used to classify the individual images but by implementing the logic operation in the forward function of the CNN it can also learn the operation. This allows the model to learn the operation and use this to correctly predict the target label.

3.2 CNN

The first baseline model that is used for comparison is a standard CNN. This model is a standard CNN with two convolutional layers, two max-pooling layers, three linear layers, and a softmax layer. The hyper-parameters of the CNN are chosen to be the same as the ones used by Manhaeve et al. [7] in their Neuro-symbolic model, this was done to ensure a fair comparison between the models. The final linear layer of the CNN outputs 19 values, one for each possible outcome of an operation on the two images, instead of 10 values for each possible label of the individual images.

This approach attempts to solve the problem by learning the operation implicitly. The model is not given any information about the operation that it is trying to learn. It is only given the images and the target label.

This model is not expected to perform as well as the other two models as it is not given any information about the operation that it is trying to learn. It is only given the images and the target label, and in cases where the same target label can be achieved by multiple combinations of the individual images, the model is expected to struggle to learn the operation.

3.3 Segmented Neuro-Symbolic AI model

The second baseline model that is used for comparison is a segmented Neuro-Symbolic AI model (SNS). This model consists of a CNN which is trained to classify the individual images, and a probabilistic logic operation which performs the chosen operation on the two images. The probabilistic logic operation receives probability estimates from the CNN and uses these to calculate the probability of the two images equalling the target label.

The CNN in this model has three convolutional layers and one average pooling layer. The network in this model is different as it only needs to train on the individual images and not the target label. This is because this model only needs to classify the images. After which the probabilistic logic operation will perform the operation on the two estimates for the images.

The probabilistic logic operation takes in two tensors of probabilities of size 10 and outputs a tensor of size equal to the solution space. The input tensors represent the probability of an image equalling digits 0 to 9. The output tensor represents the probability of the two images equalling the solution to the chosen problem. The operation is defined as follows:

$$P_z = \vee_{i=0}^{9} \vee_{j=0}^{9} \left(P_{x_i} \wedge P_{y_j}\right), \text{ where } z = i \oplus j \tag{1}$$

The probability of the two images equalling the target label is equal to the probability of the first image equalling digit i and the second image equalling digit j for all possible combinations of i and j, where $z = i \oplus j$. And $\oplus$ is the operator for the chosen problem.

This model is expected to perform better than the CNN and have similar performance to the end-to-end Neuro-Symbolic AI model. This is because it is given the same information as the end-to-end Neuro-Symbolic AI model. The difference is that the end-to-end Neuro-Symbolic AI model learns the operation during training, whereas the segmented Neuro-Symbolic AI model uses a predefined operation. As both the end-to-end Neuro-Symbolic AI model and the segmented Neuro-Symbolic AI model are both explicitly given the operation, they are expected to perform similarly.

4 Empirical Process

This section outlines the practical methodology employed for assessing various models. Section 4.1 provides an overview of the dataset used in this evaluation, while Section 4.2 expands on the specific problems used to evaluate the models. The training process of the models is described in Section 4.3, and the subsequent evaluation procedures are explained in Section 4.4.

4.1 MNIST-dataset

The experimental evaluation will use the MNIST dataset, which consists of gray-scale images of handwritten digits from zero to nine. This dataset is chosen, because it has a finite set of labels that can be assigned. The dataset will be divided into training, validation, and test sets with a ratio of 80:10:10 respectively.

The data is preprocessed by pairing up the images and setting a target label for each pair. The target label is set to match the expected output for the problem being used.

4.2 Problems

There are five different problems used to evaluate the models in this paper. All the problems take in two images and then perform an operation on the two images. The operators used are: $+$, $-$, $<$, $>$ and $\times$.

The addition and subtraction problems both have 19 possible outcomes. The less-than and greater-than problems have two possible outcomes, a boolean describing the statement: $Image1 < Image2$. Where 1 represents true and 0 represents false. Then finally the multiplication problem has 37 different outcomes.

4.3 Model Training

Five experiments will be conducted. Evaluating the three models on each of the different problems described above. In all cases the dataset will be modified such that the target labels will be the result of the operation performed on the labels of the individual images respectively.

Before training commences the seed for the model and problem combination is set to 7. Each model is trained on the training set for 15 epochs and evaluated on the validation set. They are then tested on the test set and the accuracy is recorded in a csv file. This process is repeated 20 times for each model and problem.

The segmented Neuro-Symbolic model is pre trained to classify the individual images until it achieves an accuracy of 95%. It is then evaluated on the operation performed on the images, after performing the probabilistic logic operations.

4.4 Evaluation

The average accuracy of the models are calculated over the 20 tests for each model and problem. The accuracy of each model is determined by calculating the percentage of correct predictions on the test set.

The accuracy of the models are also compared using pair wise Mann-Whitney U tests. To determine if the difference in accuracy between the models is statistically significant.

5 Results

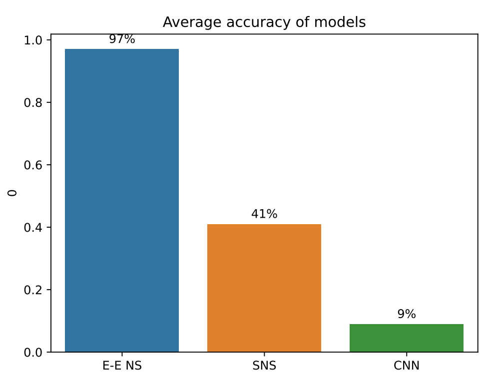
This section presents the results of the evaluation of the models. The accuracy of the models on the different problems are presented in Section 5.1. The results of the pairwise Mann-Whitney U tests are presented in Section 5.2.

5.1 Accuracy of models

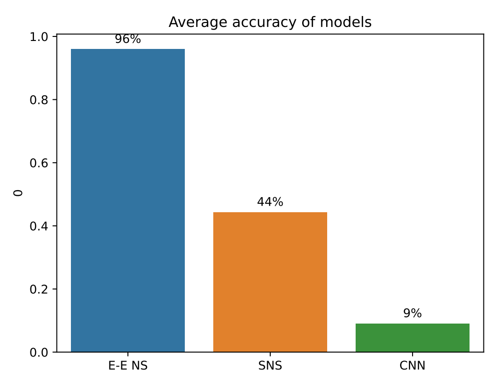
The average accuracy of the models on the five different problems are displayed below.

The average accuracy of the three models over the 20 addition tests are shown in Figure 1a. The graph in Figure 1b shows the average accuracy of the three models over the 20 subtraction tests.

Figure 1



(a) Average accuracy of the models over 20 tests for addition.



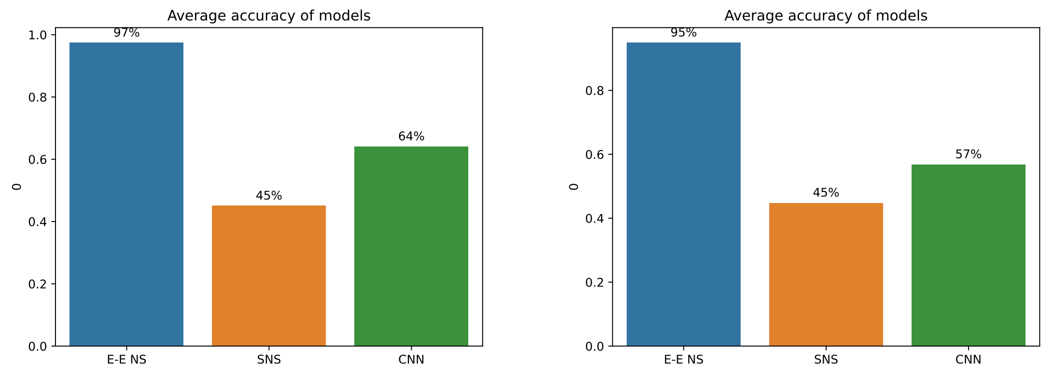
(b) Average accuracy of the models over 20 tests for subtraction.

The average accuracy for the greater-than and less-than operators are shown in Figure 2a and Figure 2b respectively.

Figure 3 shows the average accuracy of the models on the multiplication problem.

From these results it is evident that the E-E NS model has better performance than the other two models. It greatly outperformed them by having an average accuracy of 95% and greater in all the problems. In the majority of the problems the SNS model outperformed the

Figure 2



(a) Average accuracy of the three models over 20 tests for the greater-than operator.

(b) Average accuracy of the three models over 20 tests for the less-than operator.

CNN model. However, in the greater-than and less-than problems the CNN outperformed the SNS model. This is most likely due to the fact that these problems are binary classification problems, where only two outcomes are available.

The SNS and CNN models performed better on the multiplication problem than on the addition and subtraction problems. This is an unexpected result as the multiplication problem had 37 different outcomes compared to the 19 of the addition and subtraction problems. However, the E-E NS model performed similarly in this problem to the other problems.

5.2 Mann-Whitney U tests

Each model was compared to the other two models using pair wise Mann-Whitney U tests on each problem. This was done to determine if the difference in accuracy between the models is statistically significant. A "tournament" for each problem was held and if a model's accuracy was shown to be significantly different to the other models it is granted a "win" and the other model a "loss". In the case that there is not a statistically significant difference both models are granted a "draw".

Each Mann-Whitney U test was performed with a significance level of 0.05. The models are then ranked according to the best performance in each problem. The result are shown in Table 1 below.

In the table the E-E NS model refers to the end-to-end Neuro-Symbolic AI model, the SNS model refers to the segmented Neuro-Symbolic AI model, and the CNN model refers to the baseline CNN model. This table shows that the E-E NS model outperformed the other two models in all the problems. The SNS model outperfomed the CNN model on the addition, subtraction and multiplication problems, with the CNN performing better than the SNS model in the greater-than and less-than problems.

6 Conclusion

I have compared and evaluated three models on the addition, subtraction, multiplication, less-than and greater-than operators on MNIST images. A baseline CNN model, a segmented

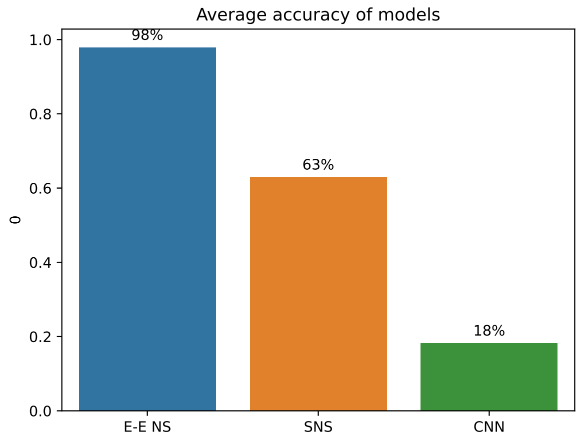


Figure 3: Average accuracy of the three models over 20 tests for multiplication.

Neuro-Symbolic AI model, and an end-to-end Neuro-Symbolic AI model.

The results of the evaluation of the models suggests that the E-E NS model performs better than the other two models. It suggests that the E-E NS model is a better performing model when working with problems that require both symbolic reasoning and classification, like the addition and subtraction of MNIST images.

This is a promising result in the field of Neuro-Symbolic AI. Suggesting that there are possible real world applications where the intertwining of symbolic reasoning and classification can be used to solve problems.

7 Future work

More complex problems that consist of multiple operations can be used to evaluate the performance of Neuro-Symbolic AI models. This will allow for a more in-depth evaluation of the performance of Neuro-Symbolic AI models.

The performance of Neuro-Symbolic AI models can also be evaluated on different metrics, for example taking into account the difference in training time and computational resources needed.

8 Bibliography

[1] Eriq Augustine et al. “Visual Sudoku Puzzle Classification: A Suite of Collective Neuro-Symbolic Tasks”. In: 16th Internatonal Workshop on Neural-Symbolic Learning and Reasoning. NeSy 2022 (Cumberland Lodge). Ed. by Artur d’Avila Garcez and Ernesto Jiménez-Ruiz. 2nd International Joint Conference on Learning & Reasoning (IJCLR 2022) 3212. UK, Windsor: CEUR-WS, 2022.

[2] Sasanka Sekhar Chanda and Debarag Narayan Banerjee. “Omission and commission errors underlying AI failures”. en. In: *AI Soc.* (Nov. 2022), pp. 1–24.

[3] Roberto Colom et al. “Human intelligence and brain networks”. en. In: *Dialogues Clin. Neurosci.* 12.4 (2010), pp. 489–501.

Table 1: Ranking of the different methods for each problem.

Model	Problem	Wins	Losses	Draws	Difference	Rank
E-E NS	Addition	2	0	0	2	1
	Greater than	2	0	0	2	1
	Less than	2	0	0	2	1
	Multiplication	2	0	0	2	1
	Subtraction	2	0	0	2	1
SNS	Addition	1	1	0	0	2
	Greater than	0	2	0	-2	3
	Less than	0	2	0	-2	3
	Multiplication	1	1	0	0	2
	Subtraction	1	1	0	0	2
CNN	Addition	0	2	0	-2	3
	Greater than	1	1	0	0	2
	Less than	1	1	0	0	2
	Multiplication	0	2	0	-2	3
	Subtraction	0	2	0	-2	3

[4] KULeuven DTAI Group. *ProbLog*. Version 2.2.4. Oct. 17, 2022. URL: `https://dtai.cs.kuleuven.be/problog/`.

[5] Artur S. d'Avila Garcez et al. "Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning". In: *CoRR* abs/1905.06088 (2019). arXiv: `1905.06088`. URL: `http://arxiv.org/abs/1905.06088`.

[6] Pascal Hitzler et al. "Neuro-symbolic approaches in artificial intelligence". In: *National Science Review* 9.6 (Mar. 2022). nwac035. ISSN: 2095-5138. DOI: `10.1093/nsr/nwac035`. eprint: `https://academic.oup.com/nsr/article-pdf/9/6/nwac035/43952953/nwac035.pdf`. URL: `https://doi.org/10.1093/nsr/nwac035`.

[7] Robin Manhaeve et al. *DeepProbLog: Neural Probabilistic Logic Programming*. 2018. DOI: `10.48550/ARXIV.1805.10872`. URL: `https://arxiv.org/abs/1805.10872`.

[8] Mark P Mattson. "Superior pattern processing is the essence of the evolved human brain". en. In: *Front. Neurosci.* 8 (Aug. 2014), p. 265.

[9] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Vol. 82. Jan. 2003, pp. 42–48.

[10] Darrell M. West and John R. Allen. "How artificial intelligence is transforming the world". In: *Brookings* (2018). URL: `https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world/`.

[11] Rikiya Yamashita et al. "Convolutional neural networks: an overview and application in radiology". In: *Insights into Imaging* 9.4 (Aug. 2018), pp. 611–629.