Machine Number Sense: A Dataset of Visual Arithmetic Problems for Abstract and Relational Reasoning

Wenhe Zhang[1,3] **Chi Zhang**[1,2] **Yixin Zhu**[1,2] **Song-Chun Zhu**[1,2]

[1] UCLA Center for Vision, Cognition, Learning, and Autonomy [2] International Center for AI and Robot Autonomy (CARA) [3] Peking University

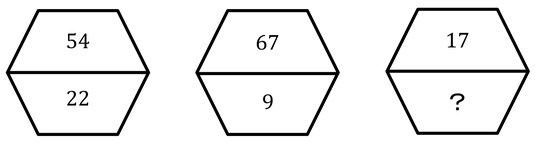
AAAI-20, New York

Motivation

Human Number Sense: the cognitive process of numbers and mathematics

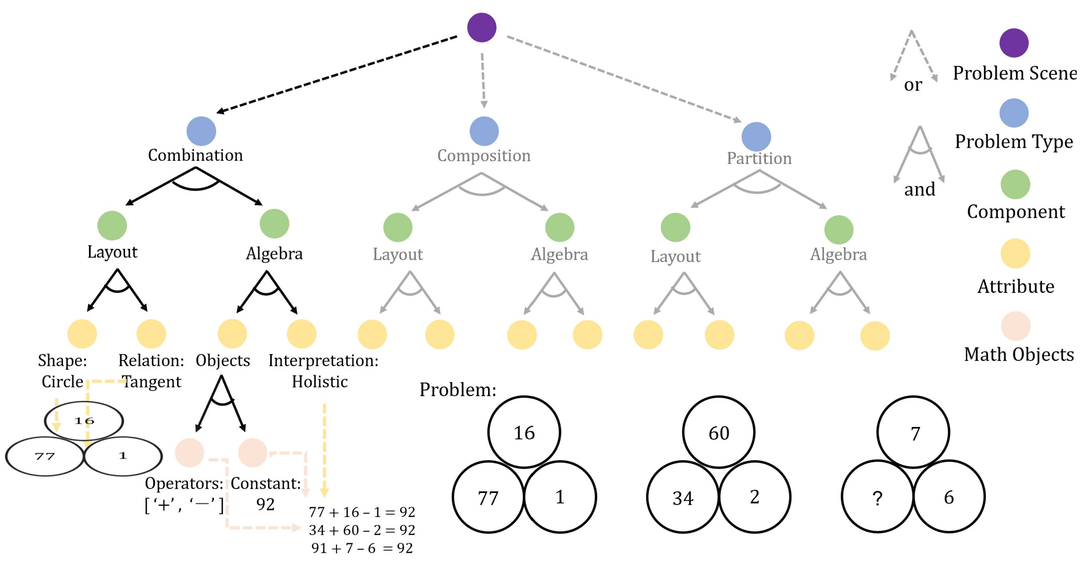
- **Induction of number symbols**:
  - → Abstract understanding of symbols.
  - → Operational relations between numbers.
- **Competence of problem-solving**:
  - → Adaptive representation formulation.
  - → Flexible strategy selection.
- **Vision-based cognitive capacity**:
  - → The understanding of number symbols is based on visual input.
  - → The development and evolution of number sense originate from vision.

Machine Number Sense: a comprehensive test of machine intelligence



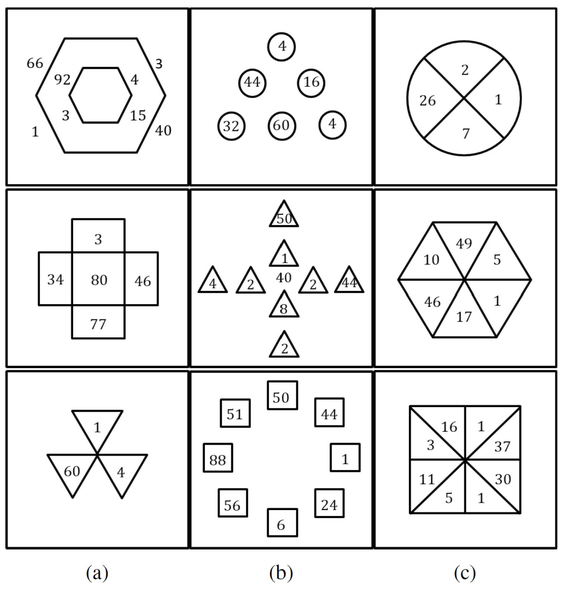
- Combines both crystallized intelligence (knowledge and experience of number processing) and fluid intelligence (adaptive problem-solving in a given situation).
- Can be represented and examined by the proposed dataset—Machine Number Sense (MNS), consisting of various visual arithmetic problems.
- Compared to other mathematical problems in prior work, the problems presented here are unique and difficult in the following aspects:
  - → test machine number sense directly from **pixel input**.
  - → require **flexible hierarchical representations** based on problem context.
  - → focus on **reasoning and understanding**, rather than the traditional tasks (*e.g.*, recognition) in the field of computer vision.
  - → investigate number sense comprehensively from a **cognitive perspective**, instead of the clinical perspective in related human tests.

Dataset Generation



A test is generated by parsing and sampling an And-Or Graph (AOG). Each problem has an internal hierarchical tree structure composed of And-nodes and Or-nodes; an And-node denotes a decomposition of a larger entity in the grammar, and an Or-node denotes an alternative decomposition.

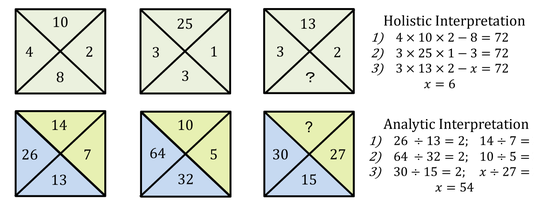
- Problems Types: (a) Combination, (b) Composition, and (c) Partition.

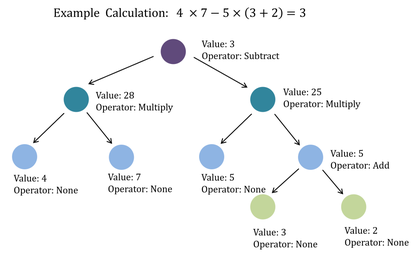


- Each problem contains two important components:
  - → Layout component serves as the **problem context**.
    The attributes vary with different problem types.
  - → Algebra component serves as the **problem content**.
    A crucial attribute is the styles of interpretation — holistic view and analytic view.

Dataset Generation (Cont.)

Numbers are generated by a "Calculator Tree".

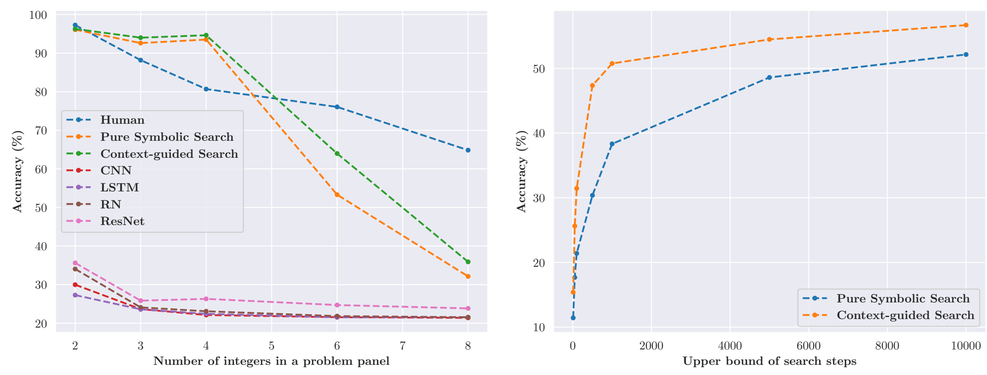




Experiments and Analysis

- We benchmark the proposed MNS dataset using both pre-dominant neural network models and classic search-based algorithms.
  - → Four state-of-the-art neural-network-based CV models for visual problem-solving:
    (i) a front-end **CNN** as feature extractor;
    (ii) a **LSTM** model with a CNN backbone combined with an MLP head;
    (iii) an image classifier based on **ResNet**;
    (iv) a relational network (**RN**).
  - → Two types of the symbolic search-based models:
    (i) **pure symbolic search**; the input is the numbers in each panel;
    (ii) **context-guided search**; the input includes both the numbers and semantic context.
- Additionally, human performance on the MNS dataset has also been collected.

Method	Mean	Combination		Composition		Partition	
		Holistic	Analytic	Holistic	Analytic	Holistic	Analytic
Pure Symbolic Search	52.15%	62.98%	56.83%	22.17%	53.73%	51.29%	71.60%
Context-guided Search	56.70%	64.38%	56.08%	29.81%	61.84%	59.70%	67.59%
CNN	22.71%	25.25%	19.65%	22.53%	20.07%	24.44%	23.25%
LSTM	22.16%	24.57%	21.10%	22.21%	20.12%	23.36%	23.83%
RN	22.96%	27.05%	20.47%	22.93%	20.27%	25.81%	23.64%
ResNet	25.29%	27.90%	24.22%	23.42%	23.73%	26.61%	27.78%
Human	77.58%	66.82%	93.64%	61.36%	78.18%	77.27%	88.18%



- Main Results:
  - → The overall accuracy of neural network models is close to that of pure symbolic search within 100 steps and context-guided search within 50 steps, both of which are relatively small compared to the large problem space.
  - → The performance of search algorithms varies across different types of problem, different styles of interpretation, and different numbers of integers, in strong contrast to the performance consistency of neural network models.
  - → Although pure symbolic search is able to solve some problems, context-guided search has generally better performance, especially on problems with higher complexity.
  - → Compared to the benchmarked computational models, human achieves a significantly higher accuracy in all types of problems without extensive training.
- Possible Reasons:
  - → The representations of number symbols and geometric contexts differ:
    search algorithms: symbolized concepts;
    neural network models: extracted features.
  - → The internal processes of visual information are distinctive:
    search algorithms: process number concepts in a sequential manner;
    neural network models: process visual features in parallel.
  - → The abilities to separate problem content from problem context is also different:
    Search algorithms are advantageous than neural network models, since the number symbols and geometric context information are fed into search algorithms separately.

Conclusions and Discussions

- Compared to simple symbolic search-based models, the poor performance of neural network models suggests its insufficiency in symbolic processing and concept understanding, as well as its difficulty in combining content and context to solve problems flexibly.
- Challenges for future work: how to emerge symbolic concepts directly from pixels with minimal supervisions, how to extract meaningful relations from contextual information, and how to reason and make inductions based on concepts and relations.
- Fusing neural network models' strong capacity of feature extraction in large-scale data processing and search-based algorithms' explicit knowledge structure in fit-for-purpose problem-solving may be an effective method for relational and abstract reasoning.