Motivation and Overview

- We explore the role of generative AI in visual programming domains such as Hour of Code Maze Challenge by Code.org and Karel.
- The available set of tasks on existing platforms is very limited, posing a major hurdle for novice students in mastering the missing concepts.
- We develop a novel technique to synthesize tasks for a given specification.
- Our technique can enable AI systems to provide personalized feedback to students as new simpler tasks and worked examples for scaffolding.

Our Technique NEURTaskSyn

- Purely Neural Techniques
  - The generative process is highly brittle and the output task could be incorrect w.r.t. the input specification.
- Purely Symbolic Techniques
  - The generative process is time-inefficient and not suitable for applications that require real-time synthesis.
- Neuro-guided Symbolic Engine
  - We develop NEURTaskSyn that can synthesize high-quality tasks while being robust and efficient.

Illustrative Example

(a) 1st and solution c^\\text{neural}
(b) 1st and solution c
(c) Specifications φ
(d) 1st and c

Figure 1: Illustration of an AI system providing a new simpler task to a student

Problem Setup

Visual Programming Task T := (T_δ, T_{code})

- T_δ denotes the visual puzzle
- T_{code} denotes additional constraints on a solution code

Solution Code C of a Task T

- C successfully solves T_δ
- C respects T_{code}

Task Synthesis Specification ψ := (ψ_δ, ψ_{code}, ψ_δ)

- ψ_δ is a partially initialized visual puzzle of the task to be synthesized
- ψ_{code} and ψ_δ capture the constraints that should be followed by solution codes of the task to be synthesized

Synthesis Objective for t^\text{neural} given φ^\text{neural}

- O1: Validity. t^\text{neural} respects φ^\text{neural} and there exists one solution C for t^\text{neural}
- O2: Concepts. t^\text{neural} conceptually captures specification φ^\text{neural}, i.e., its solution codes respect the concepts and complexity of φ_{\text{code}}
- O3: Trace. The quality of execution trace of solution codes on t^\text{neural}
- O4: Overall. All the above objectives (O1, O2, O3) are satisfied
- O5: Human. The quality of t^\text{neural} from a human expert’s point of view

Experimental Results

<table>
<thead>
<tr>
<th>Technique</th>
<th>O1: Validity</th>
<th>O2: Concepts</th>
<th>O3: Trace</th>
<th>O4: Overall</th>
<th>O5: Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEURTaskSyn</td>
<td>1.00</td>
<td>0.83</td>
<td>0.90</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>BaseTaskSyn</td>
<td>0.97</td>
<td>0.37</td>
<td>0.33</td>
<td>0.33</td>
<td>0.20</td>
</tr>
<tr>
<td>GPT4TaskSyn-converse</td>
<td>0.97</td>
<td>0.57</td>
<td>0.60</td>
<td>0.43</td>
<td>0.27</td>
</tr>
<tr>
<td>GPT4TaskSyn-laws</td>
<td>0.80</td>
<td>0.37</td>
<td>0.57</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td>EXPERT</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Conclusions and Future Work

- We propose NEURTaskSyn, a novel neuro-symbolic technique for synthesizing visual programming tasks.
- Our results show that GPT-4 faces challenges in code execution, symbolic operations, and visual planning.
- Visual programming domains can serve as benchmarks for assessing the capabilities of generative models.
- It would be interesting to fine-tune LLMs to improve their capabilities for visual programming domains.

NeurIPS’23 Workshop on Generative AI for Education (GAIED)