

More Details, Please: Improving Autoformalization with More Detailed Proofs



Guillem Tarrach, Albert Q. Jiang,
Daniel Raggi, Wenda Li, Mateja Jamnik

gg457@cam.ac.uk

UNIVERSITY OF
CAMBRIDGE
Department of Computer
Science and Technology

1. Problem statement

- Autoformalization is the process of automatically translating informal (natural language) mathematical proofs into formal (logic-based) proofs and verifying them.
- Informal proofs omit many details that are necessary for formal proofs. Autoformalizers must fill in the gaps, which presents a significant challenge for current approaches to autoformalization.
- To address this, we use large language models to add details to steps in informal proofs that are difficult to verify with automated provers.

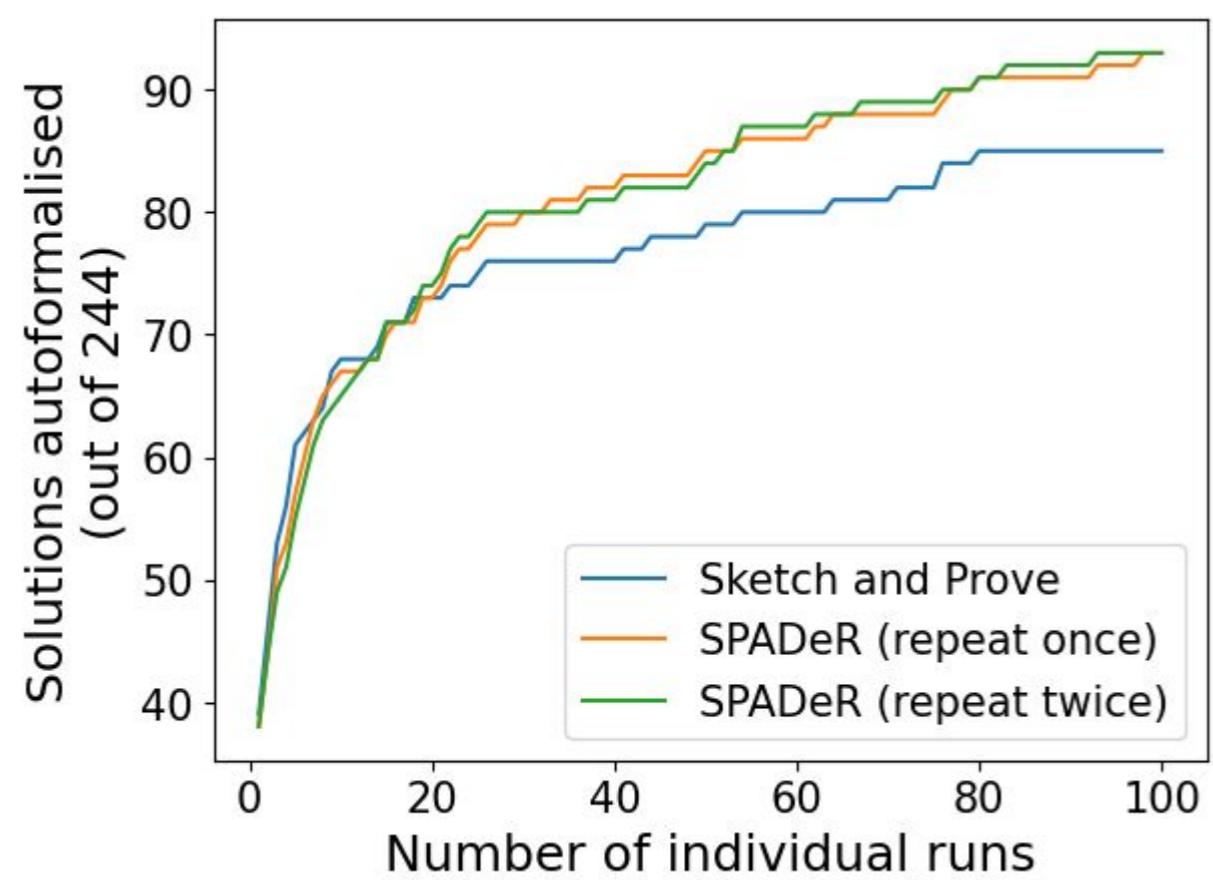
2. Sketch, Prove, Add Details & Repeat (SPADeR)

We propose an easy-to-implement method that enhances existing approaches to autoformalization:

- Generate a formal **sketch** of the informal proof. The sketch follows the informal proof, and may be incomplete (contain open conjectures).
- Prove** as many of the gaps in the formal sketch as possible with an off-the-shelf automated prover. If all the gaps are filled, we have successfully autoformalized the proof.
- Add details** to the steps in the informal proof that could not be proved formally using few-shot prompting with a large language model. The model is prompted to generate a new informal proof by replacing the problematic steps with more detailed versions, while leaving the rest of the proof unchanged.
- Repeat** the process for the new, more detailed informal proof for a specified number of iterations, or until a formal proof is found.

3. Results

- Adding detail with GPT-4o improves the success rate on the miniF2F dataset from 34.8% to 38.1%.



- Multiple detailing passes do not improve performance on the miniF2F problems.

4. Main Takeaways

- LLMs can understand and explain implicit details in informal proofs.
- Adding details to difficult steps in informal proofs improves the performance of autoformalizers.

