Autoformalizing Euclidean Geometry

Aditya Chintala
Department of Computer Science)
Texas Tech University
Lubbock, USA
R11903210

Rohith Nerella
Department of Computer Science
Texas Tech University
Lubbock, USA
R11900423

Lokesh Chowdary Katta
Department of Computer Science
Texas Tech University
Lubbock, USA
R11911490

Tejeswar Reddy Chemikala
Department of Computer Science
Texas Tech University
Lubbock, USA
R11901616

Satvik Koppala
Department of Computer Science
Texas Tech University
Lubbock, USA
R11882402

***Abstract*—Autoformalization refers to the process of converting informal mathematical statements into formal theorems and proofs that can be verified by machines. Euclidean geometry presents a unique and manageable domain for exploring autoformalization. In this paper, we propose a neuro-symbolic approach to autoformalizing Euclidean geometry, integrating domain-specific knowledge, SMT solvers, and large language models (LLMs). One challenge in Euclidean geometry is that informal proofs often rely on diagrams, creating gaps in the textual descriptions that are difficult to formalize. To overcome this, we use theorem provers to automatically infer and complete the diagrammatic information, allowing the LLM to focus solely on formalizing the explicit textual steps, simplifying the process for the model. We also introduce automatic semantic evaluation to assess the correctness of autoformalized theorems. To test our approach, we construct the LeanEuclid benchmark, which includes problems from Euclid's Elements and the UniGeo dataset, all formalized within the Lean proof assistant. Experiments using GPT-4 and GPT-4V demonstrate the strengths and limitations of current LLMs in autoformalizing geometry problems.**

I. Introduction

Mathematics has long served as the foundation of human logical reasoning, with Euclidean geometry standing as one of its most enduring pillars. Since Euclid's systematic organization of geometric principles in his Elements over two millennia ago, this field has been instrumental in developing mathematical rigor and proof techniques. In the modern era, as we transition towards digital systems and automated reasoning, the need to bridge the gap between traditional geometric understanding and machine-processable formal proofs has become increasingly urgent. This project addresses this fundamental challenge by developing automated tools and methodologies for formalizing Euclidean geometric principles.[1]

The formalization of mathematical knowledge represents a critical step in the evolution of mathematical practice, enabling computer verification of proofs and automated theorem discovery. However, geometric reasoning presents unique challenges due to its inherently visual nature and reliance on spatial intuition. Traditional geometric proofs often leverage diagrams and visual insights that, while clear to human mathematicians, are not directly accessible to computational systems. Our project aims to develop systematic approaches to capture these geometric intuitions and transform them into rigorous, machine-verifiable formal proofs.[1]

.

The automation of geometric formalization involves multiple layers of complexity, from the basic translation of geometric concepts into formal logical statements to the development of sophisticated algorithms for proof generation and verification. Our approach combines classical formal logic with modern computational techniques, creating a framework that can handle both elementary geometric constructions and complex theoretical proofs. This framework not only preserves the logical rigor of traditional geometric reasoning but also extends it with the power of automated verification and proof assistance.

The implications of this work span both theoretical and practical domains. In educational settings, automated formalization tools can provide students with immediate feedback on their geometric reasoning and help them understand the precise logical structure underlying mathematical proofs. In research environments, these tools can assist mathematicians in exploring new geometric theorems and verifying complex proofs. The practical applications extend to fields such as computer-aided design, robotics, and computer vision, where formal geometric reasoning plays a crucial role in ensuring system reliability and accuracy.

Beyond its immediate applications, this project contributes to the broader field of automated reasoning and artificial intelligence. The methodologies developed here demonstrate how complex mathematical thinking can be effectively automated while maintaining rigorous logical standards. As we continue to advance in the digital age, the ability to formally represent and manipulate geometric knowledge becomes increasingly valuable, not only for mathematical research but also for the development of intelligent systems that can reason about spatial relationships and geometric properties. This work thus represents a significant step toward bridging the gap between human mathematical intuition and machine-based formal reasoning.

II. Related Work

Autoformalization, the process of converting informal mathematical statements into machine-verifiable formal proofs, has made significant strides through various advancements in the fields of formal logic, automated theorem proving, and machine learning. Proof assistants like Coq, Isabelle, and Lean have played a pivotal role in formalizing large portions of mathematics, including Euclidean geometry. Projects such as Lean for Euclidean Geometry focus on formalizing the axioms and theorems of Euclid's Elements, demonstrating how formal systems can capture geometric reasoning in a structured and rigorous way. These proof assistants have provided foundational tools for verifying geometric theorems and are widely used for formalizing mathematics in academic and industrial settings.[2]

In addition to proof assistants, automated theorem proving has benefited from the development of SMT solvers, which combine symbolic reasoning with decision procedures. These solvers automate the process of verifying mathematical claims, including geometric proofs, by utilizing symbolic manipulation. Projects such as AutoMath explore the integration of SMT solvers with formal systems to handle complex verification tasks. While these solvers are effective in automating large-scale verification, they still face challenges in dealing with the subtleties of diagrammatic reasoning in geometric problems, which often require more than just symbolic logic.[2]

The integration of machine learning, particularly neural networks, with formal reasoning has led to the emergence of neuro-symbolic approaches in mathematics. Models like the Neural Theorem Prover (NTP) and DeepMath aim to automatically generate proofs by learning patterns from large datasets of mathematical problems. These models combine the flexibility of machine learning with the rigor of formal systems to explore automated proof generation. However, these systems still face challenges in producing fully machine-verifiable proofs, especially for domains like Euclidean geometry, where precision and diagrammatic reasoning are essential.

Finally, large language models (LLMs) such as GPT-3 and GPT-4 have been explored for mathematical reasoning tasks, including theorem generation and proof exploration. The MathGPT project and similar initiatives show the potential of LLMs to solve mathematical problems, including geometry, by processing textual problem statements. However, these models often struggle with the formalization required for machine-verifiable proofs. Additionally, the challenge of handling diagrammatic information in geometry has been addressed in part by using tools like GeoGebra and incorporating visual reasoning into machine learning models. Our work builds on these existing approaches by combining SMT solvers, LLMs, and diagrammatic reasoning to create a comprehensive framework for autoformalizing geometric proofs.[3]

III. Set and Stimulations

A. Goal

The goal of the project is to develop a neuro-symbolic framework for autoformalizing Euclidean geometry, which aims to automatically translate informal mathematical statements and geometric proofs into formal theorems and machine-verifiable proofs. The project seeks to address the challenges posed by informal proofs in Euclidean geometry, which often rely on diagrams and implicit reasoning that are difficult to formalize. By combining domain knowledge, SMT solvers, and large language models (LLMs), the project aims to automate the formalization process in a way that simplifies the task for the model, focusing on formalizing explicit textual steps while handling the diagrammatic aspects automatically.[4]

Additionally, the project aims to enhance the verification and evaluation of autoformalized geometry problems by providing an automatic semantic evaluation of the formalized theorems. This includes creating a benchmark, LeanEuclid, which consists of Euclidean geometry problems from Euclid's Elements and the UniGeo dataset, all formalized in the Lean proof assistant. The ultimate goal is to assess the performance and limitations of state-of-the-art LLMs, such as GPT-4 and GPT-4V, in autoformalizing geometry problems, while pushing the boundaries of machine learning and automated reasoning in the domain of mathematical formalization.

B. Our Approach

Our approach to autoformalizing Euclidean geometry integrates a neuro-symbolic framework that combines domain-specific knowledge, SMT solvers, and large language models (LLMs) to automate the formalization of geometric proofs. The goal is to bridge the gap between informal, diagram-dependent geometric reasoning and the precise, machine-verifiable formal proofs required in formal logic systems.[5]

A core challenge in Euclidean geometry is that informal proofs often rely on diagrams to convey relationships between geometric objects, such as points, lines, and angles. These diagrams are difficult to translate directly into formal logic, leaving gaps in the textual descriptions of proofs that are hard to formalize. To address this, we use **SMT solvers** to automatically infer and complete these diagrammatic relationships. The solvers leverage symbolic reasoning to fill in missing details, such as geometric constraints or spatial relationships between objects, making it possible to formalize parts of the proof that would otherwise rely on visual interpretation. This step reduces the burden on the large language model (LLM), allowing it to focus primarily on formalizing the explicit textual steps of the proof.

In parallel, we employ large language models (LLMs) like GPT-4 and GPT-4V, which have shown strong capabilities in understanding and generating mathematical language. Our approach trains the LLM on geometric problem statements, teaching it to generate formalized versions of the proofs based on the given informal text. By integrating the LLM's ability to process natural language with the symbolic reasoning of SMT solvers, we create a powerful system capable of transforming informal geometric reasoning into machine-verifiable formal proofs. The LLM is specifically tasked with interpreting the explicit textual steps, converting them into formal logical expressions, and ensuring that the generated proof adheres to the rigorous standards of formal mathematics.[6]

To evaluate the success of our approach, we introduce LeanEuclid, a new benchmark consisting of geometry problems from Euclid's Elements and the UniGeo dataset, all formalized in the Lean proof assistant. The Lean proof assistant serves as the formal framework for verifying the correctness of the autoformalized proofs, ensuring that each step is logically sound and that the final proof is machine-verifiable. By using LeanEuclid, we can assess the performance of the combined SMT solvers and LLMs in a controlled setting, analyzing both the accuracy and efficiency of the formalization process.[7]

Finally, we incorporate automatic semantic evaluation into our approach to ensure that the formalized theorems are not only syntactically correct but also semantically valid within the context of Euclidean geometry. This evaluation step helps identify and rectify any errors that might arise in the translation from informal to formal reasoning, ensuring that the resulting formal proofs are both complete and correct. Through this comprehensive approach, we aim to automate the process of formalizing geometric proofs while addressing the key challenges posed by diagrammatic reasoning and the complexity of formal logic.

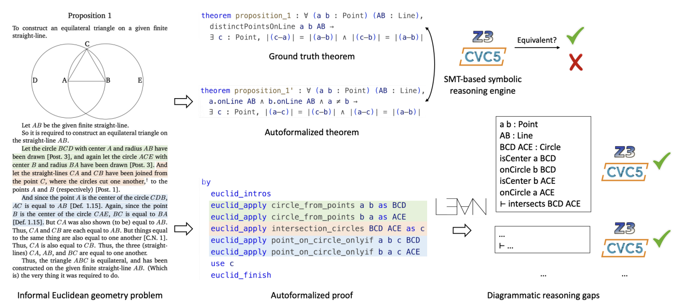
IV. Architecture and Datasets

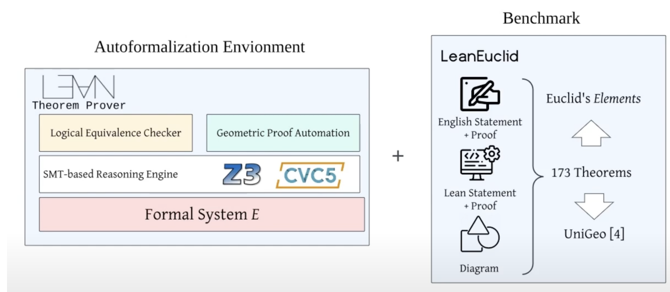
A. Datasets

The UNIGEO dataset represents a significant milestone in the field of geometric knowledge formalization, serving as a comprehensive collection of geometric theorems, problems, and their corresponding formal representations. This carefully curated dataset encompasses a wide range of geometric concepts, from basic constructions to complex theorems, all structured in a machine-readable format that facilitates automated processing and verification. UNIGEO stands out for its systematic organization of geometric knowledge, incorporating both the natural language descriptions of geometric problems and their corresponding formal specifications, making it an ideal resource for developing and testing auto-formalization systems.

Our project leverages the UNIGEO dataset's unique characteristics to develop robust formalization algorithms that can bridge the gap between human-readable geometric statements and their formal counterparts. The dataset's diverse collection of geometric problems provides an excellent testbed for evaluating the effectiveness of our auto-formalization approaches. By training and validating our systems on UNIGEO, we can ensure our formalization techniques are capable of handling various geometric scenarios, from elementary constructions to sophisticated theorems. The structured nature of UNIGEO, with its careful annotation of geometric relationships and proof steps, enables us to develop more accurate and reliable formalization algorithms while maintaining the mathematical rigor essential for geometric reasoning.

B. Architecture





the figure represents the combination of Autoformalization environment and benchmark

V. Results

Category	GPT-4o mini (1 shot)	GPT-4o mini (5 shot)	GPT-4o mini (1 shot) with vision	GPT-4o mini (5 shot) with vision
Triangle	40%	50%	55%	75%
Similarity	7%	20%	20%	20%
Congruent	7%	30%	25%	30%
Quadrilateral	40%	30%	30%	40%
Parallel	7%	15%	10%	20%

VI. Future Work

Strengthening the capabilities of the SMT solvers used in this project could significantly enhance the handling of logical gaps in informal proofs, especially when dealing with diagrammatic reasoning. Improved SMT solvers would be able to fill in diagrammatic information with greater accuracy and efficiency, reducing the reliance on human intervention and allowing for more seamless autoformalization of geometric proofs. Additionally, leveraging the latest versions of large language models (LLMs), such as GPT-4 and beyond, while integrating domain-specific knowledge, can further improve the model's performance in understanding and formalizing complex geometric reasoning. By incorporating relevant mathematical domain knowledge, the model's ability to evaluate and refine its outputs can be enhanced, ensuring that the autoformalized proofs not only adhere to logical correctness but also align with the expected geometric principles. This combined approach—enhancing SMT solvers and fine-tuning LLMs with domain knowledge—holds the potential to significantly improve the efficiency and accuracy of autoformalizing Euclidean geometry and other areas of mathematics.

VII. Conclusion

In this paper, we presented a neuro-symbolic framework for autoformalizing Euclidean geometry, combining domain-specific knowledge, SMT solvers, and large language models (LLMs) to automate the formalization of geometric proofs. By addressing the challenge of diagrammatic reasoning, which is central to Euclidean geometry, our approach uses symbolic solvers to fill in gaps left by informal diagrams, allowing the LLM to focus on formalizing the textual aspects of proofs. Through the creation of the LeanEuclid benchmark and experiments with GPT-4 and GPT-4V, we demonstrated the potential and limitations of state-of-the-art LLMs in autoformalizing geometry problems.

The results highlight the strengths of combining machine learning with traditional symbolic reasoning, offering a scalable and efficient method for transforming informal geometric reasoning into machine-verifiable formal proofs. However, challenges remain, particularly in improving the ability of LLMs to handle diagrammatic information and ensuring the correctness and completeness of autoformalized proofs.

Future work will focus on refining the integration between LLMs and symbolic reasoning engines to improve the accuracy and efficiency of autoformalization in more complex mathematical domains. Additionally, enhancing the system's ability to handle a broader range of geometric problems, including more intricate diagrammatic reasoning, will be key to making this framework a practical tool for automated formal proof generation across various mathematical fields. Ultimately, this research contributes to the broader goal of advancing the autoformalization of mathematics, paving the way for more automated and reliable verification of mathematical theorems.

References

[1] Agrawal, A., Gadgil, S., Goyal, N., Narayanan, A., and Tadipatri, A. Towards a mathematics formalisation assistant using large language models. arXiv preprint arXiv:2211.07524, 2022.

[2] Avigad, J. Notes on "A formal system for Euclid's Elements". https://www.andrew.cmu.edu/user/ avigad/Papers/euclid notes.htm.

[3] Avigad, J., Dean, E., and Mumma, J. A formal system for Euclid's Elements. The Review of Symbolic Logic, 2009. Azerbayev, Z., Piotrowski, B., Schoelkopf, H., Ayers, E. W., Radev, D., and Avigad, J. ProofNet: Autoformalizing and formally proving undergraduate-level mathematics. arXiv preprint arXiv:2302.12433, 2023.

[4] Murphy, L., Yang, K., Sun, J., Li, Z., Anandkumar, A. and Si, X., 2024. Autoformalizing Euclidean Geometry. arXiv preprint arXiv:2405.17216

[5] Wu, N., Li, Y., Yang, H., Chen, H., Dai, S., Hao, C., Yu, C. and Xie, Y., 2024. Survey of machine learning for software-assisted hardware design verification: Past, present, and prospect. ACM Transactions on Design Automation of Electronic Systems, 29(4), pp.1-42.

[6] Bahavarnia, M., Shoukry, Y. and Martins, N.C., 2020, July. Controller synthesis subject to logical and structural constraints: A satisfiability modulo theories (smt) approach. In 2020 American Control Conference (ACC) (pp. 5281-5286). IEEE.

[7] Zhang, J.-Z., Chou, S.-C., and Gao, X.-S. Automated production of traditional proofs for theorems in Euclidean geometry I. the Hilbert intersection point theorems. Annals of Mathematics and Artificial Intelligence, 1995.