Formal Verification of Neural ODE for Safety Evaluation in Autonomous Vehicles

Abdelrahman Sayed Sayed

Univ Gustave Eiffel, COSYS-ESTAS, F-59657 Villeneuve d'Ascq, France
abdelrahman.ibrahim@univ-eiffel.fr

Abstract

Higher autonomy is an increasingly common goal in the design of transportation systems for the cities of the future. Recently, part of this autonomy in both rail and maritime transport has come from the field of artificial intelligence and machine learning, particularly for perception tasks (detection and recognition of rail signals, other vessels, or other elements in the vehicle environment) using neural networks. Although AI-based approaches have gained significant popularity in many application fields due to their good performance, their unpredictability and lack of formal guarantees regarding their desired behavior present a major issue for the deployment of such safety-critical systems in urban areas. The goal of my PhD thesis is to design new formal methods to analyze and ensure the safety of such AI-based perception modules in autonomous vehicles. More specifically, my PhD topic aims to formally evaluate the safety of a recently introduced class of continuous AI models which are neural ODE.

Neural ODE have already been used successfully for image recognition tasks, showing higher performance compared to classical neural networks, but current work in the literature primarily focuses on their training performance, while they have been barely studied in terms of safety and formal guarantees. The main research directions that will be investigated during my PhD include:

- R.D 1: Establishing formal relations between discrete and continuous neural models, and using them to deduce the safety of one model based on the safety verification of the other.
- R.D 2: Analysis of the mathematical properties satisfied by the new continuous models (continuity, monotonicity, contraction, stability, etc.).
- R.D 3: Exploiting these mathematical properties to study and/or enforce the overall behavior of the neural ODE with respect to various features: stability, stabilization, reachability analysis, safety, formal verification.
- Verification and testing of theoretical findings on autonomous underwater vehicles (AUVs) during a 6 month mobility at the Department of Marine Technology, NTNU.

Progress Towards Thesis Objectives and Main Results Obtained

For R.D 1, the strong relation between ResNet with a single residual block and neural ODE was investigated in (Sayed, Meyer, and Ghazel 2025a). This similarity between the two models allowed us to use the properties verified for one model to deduce safety guarantees for the other, i.e., the verification results from one model serve as a proxy for the other. The main contributions obtained so far from this research direction are as follows:

- We derived a rigorous bound on the approximation error between the neural ODE and ResNet models for a given input set.
- We used the derived error bound in conjunction with the reachable set of one model as a proxy to verify the safety properties of the other model, without applying any verification tools to the other model, as illustrated in Figure 1

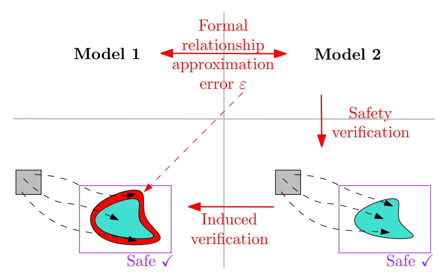

Figure 1: Proposed framework to verify Model 1 based on the outcome of the verification of Model 2

Our proposed error bound tightness was compared with the SOTA, where our error bound was approximately **16 million times** smaller than the SOTA. Thus, the tightness of our proposed approach is very significant.

Another research direction currently being investigated as part of R.D 2 and R.D 3 focuses on the safety verification of neural ODE through mixed monotonicity reachability analysis. The first result from this direction (Sayed, Meyer, and Ghazel 2025b), propose a novel method for lightweight interval-based reachability analysis of neural ODE by adapting continuous-time mixed monotonicity techniques for continuous-time dynamical systems (Meyer, Devonport, and Arcak 2021), our approach efficiently computes interval-based over-approximations, both from the full initial input set and its boundaries, exploiting the homeomorphism property to reduce computational costs. These methods offer lower complexity and faster computation times compared to

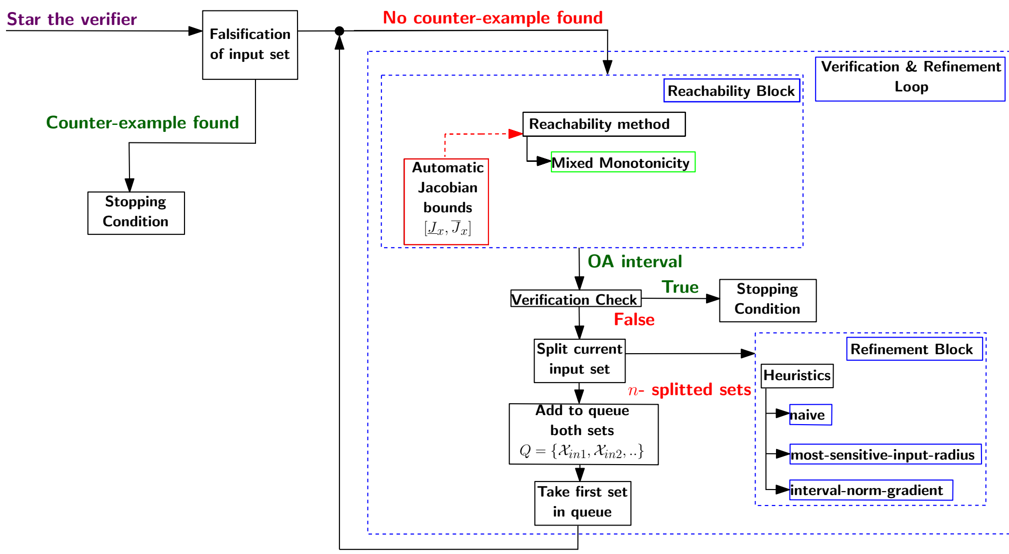

Figure 2: Verifier Architecture

more flexible set representations.

From the numerical results, we demonstrate that our single-step, incremental, and boundary-based reachability approaches yield sound over-approximations, albeit at the cost of tightness. Ultimately, the choice between our interval-based reachability methods and other reachability tools depends on the **trade-off between over-approximation tightness and computational efficiency**, making our approach particularly suitable for high-dimensional and real-time safety verification scenarios. The expected accomplished work by the time of the consortium includes developing a full verifier based on the algorithms and mixed monotonicity reachability method of R.D 2 and R.D 3 for the safety verification of neural ODE, the verifier architecture is illustrated in Figure 2. This verifier will provide the first **iterative refinement approach** to neural ODE verification based on the refinement of the initial input set using the neural ODE reachability analysis methods introduced in (Sayed, Meyer, and Ghazel 2025b) compared to the available tools that only consider the verification of the input set as a whole. Also, in terms of prospective mobility applications, I will be focusing on replacing the current existing AI component for image classification with a neural ODE model, then defining a specification to be verified for this neural ODE model, and finally applying my verifier to ensure a specification/robustness of the neural ODE model are met. An application currently identified that focuses on monitoring underwater zooplankton and some micro-organisms in the water column using AUVs and a particle imager is illustrated in Figure 3, where the verifier will be able to verify the robustness of specific regions of interest (ROI) from collected in situ images.

Future Research Directions

In future work, I plan to explore the scalability of R.D 1 by extending the framework to handle other neural network

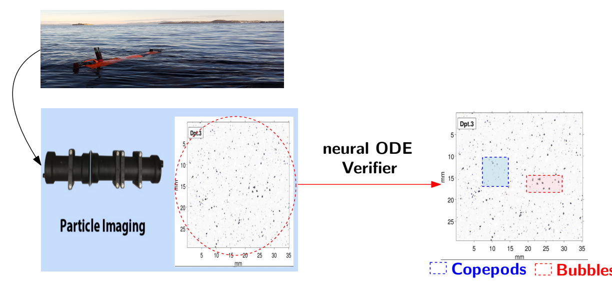

Figure 3: Robustness verification for ROI from collected in situ images

architectures (e.g., RNN and CNN). Additionally, in terms of practical applications, I aim to perform **closed-loop verification** of the entire autonomous loop of the AUV by integrating the neural ODE perception module with the vehicle's control policy. Finally, I intend to achieve **real-time verification** by optimizing the verifier for the embedded hardware of the AUV, facilitating adaptive mission planning.

References

Meyer, P.-J.; Devonport, A.; and Arcak, M. 2021. *Interval reachability analysis: Bounding trajectories of uncertain systems with boxes for control and verification*. Springer Nature.

Sayed, A. S.; Meyer, P.-J.; and Ghazel, M. 2025a. Bridging neural ode and resnet: A formal error bound for safety verification. In *International Symposium on AI Verification*, 97–114. Springer.

Sayed, A. S.; Meyer, P.-J.; and Ghazel, M. 2025b. Mixed Monotonicity Reachability Analysis of Neural ODE: A Trade-Off Between Tightness and Efficiency. *arXiv preprint arXiv:2510.17859*.