World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:04, 2022

Trusted Neural Network: Reversibility in Neural Networks for Network Integrity Verification

Authors : Malgorzata Schwab, Ashis Kumer Biswas

Abstract : In this concept paper, we explore the topic of Reversibility in Neural Networks leveraged for Network Integrity Verification and crafted the term ''Trusted Neural Network'' (TNN), paired with the API abstraction around it, to embrace the idea formally. This newly proposed high-level generalizable TNN model builds upon the Invertible Neural Network architecture, trained simultaneously in both forward and reverse directions. This allows for the original system inputs to be compared with the ones reconstructed from the outputs in the reversed flow to assess the integrity of the end-to-end inference flow. The outcome of that assessment is captured as an Integrity Score. Concrete implementation reflecting the needs of specific problem domains can be derived from this general approach and is demonstrated in the experiments. The model aspires to become a useful practice in drafting high-level systems architectures which incorporate AI capabilities.

Keywords : trusted, neural, invertible, API
Conference Title : ICMLA 2022 : International Conference on Machine Learning and Applications
Conference Location : Boston, United States
Conference Dates : April 21-22, 2022