Tampere University

A Practical Overview of Safety Concerns and Mitigation Methods for Visual Deep Learning Algorithms

Saeed Bakhshi Germi, Esa Rahtu

Tampere University

Objectives

- Categorizing faults and underlying causes in a visual deep learning algorithm
- Providing a practical and complete list of safety concerns in each stage
- Listing potential state-of-the-art mitigation methods to deal with the concerns
- Discussing the effectiveness of the mitigation methods

Safety Concerns and Mitigation Methods

- Incomplete Dataset
- Inadequate Dataset
- Insufficient / Noise Dataset
- Ill-Matched Architecture
- Unfitting Metrics
- Incompatible Benchmarks
- Black-Box Behavior
- Defective Hardware
- Hostile Environmental

Conclusion

- State-of-the-art mitigation methods have vital drawbacks:
  - Accuracy-vs-Safety trade-off, High resource/time demand, Requirement of expert knowledge, etc.
- A demand to move away from traditional broad-spectrum standardization
- Utilizing safety concern lists to create a safety case argument for the application