Contents

List of Figures

List of Tables

1

Introduction

1.1 Cyber-Physical Systems

Introduce the concept of safety critical, CPS. What makes a system safety critical?

1.2 Artificial Neural Networks

Introduce AI, more specifically ANNs. What they can do?

1.3 Synchronous Programming

What are synchronous languages, why are they used, where are they used?

1.4 Contribution

Creating methods of implementing ANNs in safety critical systems using synchronous semantics.

1.5 Thesis Structure

...?

2
Background

2.1 Artificial Neural Networks

2.1.1 Machine Learning

What is machine learning. Where is it used. Some examples.

2.1.2 Efficacy of Machine Learning

How does machine learning (ANNs) improve systems these days?

2.1.3 Structure of an Artificial Neural Network

How does an ANN run, how does it train, what does it look like.

2.1.4 Safety of Artificial Neural Networks

Safety drawbacks of ANNs: Unpredictable Hard to verify Slow Etc

2.2 Synchronous Languages

2.2.1 Synchronous Semantics

Briefly explain synchronous semantics: what is synchronous programming and how does it work?

2.2.2 Safety of Synchronous Languages

What guarantees does synchronous programming provide? Determinism Causality Easy to formalise Etc

2.2.3 Timing Correctness

Time predictability.

2.2.4 Functional Correctness

Runtime Enforcement

2.3 Time Predictable Cyber-Physical Systems

2.3.1 Worst Case Execution Time

Explain what WCET is in safety critical systems. Why is it important for deadlines to be met and how does this relate to WCET?

2.3.2 Worst Case Reaction Time

WCRT vs WCET. Throughput vs Time.

2.3.3 Timing Techniques

What timing techniques exist for synchronous and asynchronous systems? Which of these are relevant to our work?

2.3.4 Patmos Time Predictable Processor

Brief description of Patmos and why we use it.

2.4 Verification of Artificial Neural Networks

This section looks into ANN verification methods and why they are good/bad.

2.4.1 The Problem of Artificial Neural Network Verification

ANN verification is an issue. Why?

2.4.2 Existing Techniques for Verifying ANNs

Rundown of existing techniques of ANN verification.

2.4.3 Limitations of Artificial Neural Network Verification

Describe where verification is limited and why. Where can it do better?

2.5 Run-time Enforcement

2.5.1 Run-time Assurance

What is RA and why is it useful? Formally defined.

2.5.2 Run-time Monitoring

How does this relate and improve on RA?

2.5.3 Run-time Enforcement

Benefits? Synchronous vs asynchronous systems.

2.5.4 Safety (Timed) Automata

These are used in our run-time enforcement techniques. Discuss these and how they work.

2.6 Summary

Summary of above.

3

Synchronous Neural Networks

3.1 Synchronous Neural Networks

3.1.1 Fundamentals of Artificial Neural Networks

Explain what is needed in an ANN library.

3.1.2 Architecture of Synchronous Neural Networks

Describe how SNNs run and how they work as an ANN library.

3.1.3 Formalalisation of Synchronous Neural Networks

This subsection provides a formal definition for feed forward SNNs

3.1.4 Implementation of Synchronous Neural Networks

Describe how to use the SNN library. Describe both the C and Esterel aspects of SNNs.

3.2 Meta Neural Networks

This section describes how MNNs relate to SNNs

3.2.1 Architecture of Meta Neural Networks

Explain how MNNs are created and how they function

3.2.2 Formalisation of Meta Neural Networks

Provide the formal definition for MNNs

3.2.3 Implementation of Meta Neural Networks

Explain how MNNs are implemented and why they are useful. Maybe give some basic examples.

3.3 Timing Analysis of Synchronous Neural Networks

3.3.1 Patmos

Brief description of the Patmos used in timing analysis.

3.3.2 Time Predictable Software for Patmos

What is required of software to be time analysable on Patmos?

3.3.3 Timing of Meta Neural Networks

Describe how the timing of MNNs was accomplished.

3.4 A Case Study for Time Predictable Synchronous Neural Networks

3.4.1 An Energy Storage System for an Electric Vehicle Charging Station

Cite Kalpesh's paper and describe the system we are using as our case study. Why it is unsafe and how we can make it better.

3.4.2 Energy Storage System and Machine Learning

Describe the role of machine learning for the ESS and how the system can be improved. Describe the training methods used to train the ANNs, and elaborate on how the ANNs work for this system.

3.4.3 Synchronous Neural Network in an ESS

Methodology behind implementing the C and Esterel code for the trained ANN.

3.4.4 Timing Analysis of the ESS

Explain how timing was important for this system and how it can improve the safety of this system. WCET and WCRT.

3.5 ESS Case Study Results

Results of the ESS case study.

3.6 Synchronous Neural Network Timing Benchmark Results

Results of the benchmarks

3.6.1 RABBIT Game

3.6.2 Adder

3.6.3 AI-BRO

3.6.4 XOR

3.7 Summary

Summary of everything previously mentioned in this chapter

3.8 Discussion

Discussion of the results of the case study and benchmarks. Discuss methodology and results.

4

Runtime Enforcement of Synchronous Neural Networks

4.1 Runtime Enforcement of Safety Critical Systems

Introduce the concept of runtime enforcement for safety critical systems.

4.1.1 Safety Automata

Explain how timed (safety) automata are used as policies for runtime enforcement. Why these are good.

4.1.2 Synchronous Runtime Enforcement

How does RTE works with synchronous systems and what are the benefits.

4.2 Runtime Enforcement of Synchronous Neural Networks

Section showing how we are using RTE with SNNs.

4.2.1 Safe Synchronous Neural Networks

Introduce the concept of combining SNNs and RTE to create a safe SNN

4.2.2 Semantics of Safe Synchronous Neural Networks

SSNN formal semantics in detail?

4.2.3 Safe Meta Neural Networks

How does the use of RTE expand to MNNs?

4.2.4 Semantics of Safe Meta Neural Networks

SMNN formal semantics in detail?

4.2.5 Timing of Safe Synchronous Neural Networks

Briefly discuss how these can be timed, etc.

4.3 A Case Study for Safe Synchronous Neural Networks

This is the created AV system with runtime enforcement.

4.3.1 Autonomous Vehicles

What are AVs, where are they used, why are they dangerous?

4.3.2 Safety Autonomous Vehicle Systems

How is this system unsafe? What did it do to make it unsafe.

4.3.3 Runtime Enforcement of Autonomous Vehicles Systems

Where can RTE be introduced to AV systems. What are its limitations? Where does our work take RTE of this system.

4.3.4 Architecture of an Autonomous Vehicle Simulation

We created a simulation of an AV system. What was done to ensure they were as similar as possible.

4.3.5 Safety Policy for an Autonomous Vehicle Simulation

The enforced safety policy for the above system.

4.4 Autonomous Vehicle Case Study Results

Results of the AV case study.

4.5 Safe Synchronous Neural Network Benchmark Results

Results of the above-mentioned benchmarks

4.5.1 RABBIT Game

4.5.2 Energy Storage System

4.6 Summary

Summary of everything previously mentioned in this chapter

4.7 Discussion

Discussion of the results of the case study and benchmarks. Discuss methodology and results.

5

Runtime Verification of Synchronous Neural Networks

5.1 Verification of Deep Artificial Neural Networks

Deep Artificial Neural Networks (ANNs) [7], ANNs with large, complex inputs and a large, complex layer structure and multiple layers, are hard to verify due to their complex nature [4]. Lots of work has been put into the verification of ANNs, but the results yielded from this work have many limitations and are generally time consuming.

Previously, this thesis introduced the concept of using runtime enforcement to dynamically "verify" an ANN as it runs. While this works very well for simple, feed forward ANNs, this does not extend to more complicated Convolutional Neural Networks (CNNs).

5.1.1 Runtime Enforcement of Deep Artificial Neural Networks

Runtime Enforcement, as a solution to the verification of ANNs and introduced in the previous chapter, only works when the inputs to the ANN are known. This works very well with feed forward ANNs, as their inputs are, generally, defined and known. Take the AI-BRO ANN discussed in Section **??**; this ANN has a vector of defined inputs representing the objects "seen" by the vehicle. Since dangerous situations are defined by both inputs and outputs, the outputs of this ANN can be enforced because both the inputs and outputs are known at all times.

Complex ANN inputs (such as images for CNNs) are unknown, and the purpose of the ANNs is to classify said inputs. Since the inputs cannot be defined as something recognisable, the decision made by the ANN cannot be classified as dangerous because a dangerous situation is defined by known inputs and outputs. Take a CNN that classifies pedestrians; we are using the CNN because the software cannot recognise whether the input image shows a pedestrian or not. If the pedestrian is misclassified as a tree, the system has no way to know this based off of the inputs to the CNN alone. Thus, behaviour of ANNs with complex inputs (most commonly classification ANNs) cannot be enforced using only the inputs to the ANN.

The solution proposed is to use runtime enforcement as a means to safely combine the outputs from various sensors in the system and use these to enforce the outputs of the ANN.

5.1.2 Input Perturbation for Deep Artificial Neural Networks

This chapter also tackles another problematic aspect of using CNNs: input perturbations. A group showed that very slight modifications to the input image of a Convolutional Neural Network (CNN), such as the discolouration of a few pixels, could cause the image to be misclassified [4]. A CNN can train to high accuracy on the training set, but simple perturbations to the CNN's input can lead to drastically reduced accuracy once the CNN

has been deployed. This thesis shows two methods of increasing the prediction accuracy of an CNN affected by input perturbations: the first is the above-mentioned sensor fusion with runtime enforcement and the other is using a Meta Neural Network (MNN) ensemble of different CNNs, each classifying a different aspect of the input image and congregating to make a decision as a whole.

5.1.3 Sensor Fusion and Runtime Verification for Deep Artificial Neural Networks

Sensor fusion is a highly researched topic for increasing the accuracy of object detection CNNs [10]. Sensor fusion is the combination of two or more different sensor types to increase the overall sensor detection accuracy of the system. Where Autonomous Vehicles (AVs) are concerned safety is of utmost importance, making this the biggest area of research for sensor fusion using CNNs.

The proposed approach to safe deep ANNs combines runtime enforcement and sensor fusion to safely increase the detection accuracy of CNNs in AVs. The sensors in question are a 360◦ Light Detection and Ranging (LiDAR) sensor and three front-facing cameras. The sensor outputs are combined synchronously using a synchronous runtime enforcer, however the runtime enforcer does not enforce the outputs of the detection system. Rather, the runtime enforcer verifies the integrity of the detected outputs, and in the case of a possible misclassification the system is put into a safe mode where control of the vehicle is returned to the driver. This type of runtime enforcement has been termed as “runtime verification”. Using a Safe Synchronous Neural Network (SSNN) as the CNN in the AV system, the system is kept synchronous, time predictable and easy to formalise.

Additionally, to reduce the impact of perturbations to the CNN’s inputs, the SSNN used in the system is actually a MNN composed of three MNN ensembles, each with three synchronous CNNs. This MNN aims to greatly increase the prediction accuracy by splitting the prediction process amongst multiple, different CNNs such that the weakness of each CNN is addressed by the other CNNs.

5.2 A Object Detection Case Study for Perturbed Inputs

Autonomous Vehicles (AVs) are quickly growing, companies such as Uber and Tesla the biggest contributors to this field.

5.2.1 Object Misclassification in Autonomous Vehicle Systems

Describe what sort of object are misclassified, some examples of issues with misclassifications (Uber, etc) and why these occur.

5.2.2 Simulation of an Autonomous Vehicle Object Detection System

We simulated an AV object detection system. How close is it to the original and why we simulated what we did.

5.3 Simulating a Traffic Sign Recognition AV System

The system designed for this case study was made to reflect a Autonomous Vehicle (AV) and its object detection mechanisms. The system used multiple techniques to tackle the inherent issues of the AV system, i.e. weakness to perturbed inputs and misclassification detection. The system's sensors included an overhead, 360° Light Detection and Ranging (LiDAR) apparatus, and a single, frontal facing camera. A solitary camera was sufficient to prove the efficacy of this solution, however it is to be noted that AV systems generally use multiple cameras, facing different directions, so that the controller can make properly informed decisions. The system used can be seen in Figure 5.3.

5.3.1 Architecture of a Traffic Sign Recognition AV System

Utilising synchronous semantics [1], a Meta Neural Network (MNN), containing three other MNN ensembles, was created. A MNN is structure containing at least one Synchronous Neural Network (SNN) [6] synchronously combined with any number of other functional blocks [6]. Each of the three MNN ensembles were used to classify shape, colour and object type of each object being output by the camera. Each ensemble synchronously combined the outputs of three different convolutional SNNs, providing increased prediction accuracy for shape, colour and object type. Each SNN was implemented in the synchronous language Esterel [2]. Synchronous programming guarantees code that is deterministic, causal and bounded. Additionally, the calculation of Worst Case Execution Time (WCET) and Worst Case Reaction Time (WCRT) is enabled by using Esterel with C. These ensembles ran in synchronous concurrency with each other, forming a single, larger MNN. The outputs of each ensemble were then combined into a batch of outputs forming the *predicted output*.

5.3.2 Sensor Fusion for a Traffic Sign Recognition AV System

Two sensors were used for the purpose of sensor fusion; LiDAR and cameras. The LiDAR for this system was accurate 93% of the time [9], to closely simulate a real LiDAR system. The simulated camera outputs consisted of test images from both the Visual Object Classes (VOC) 2012 [3] and German Traffic Sign Recognition Benchmark (GTSRB) [8] datasets, in a combination of people, vehicles and various traffic signs. The LiDAR and camera outputs were handled by different parts of the controller. The camera outputs were fed into a MNN (see Figure 5.1) where they were classified by shape, colour and object type. Sensor fusion happened after classification occurred and was done using an enforcer during runtime.

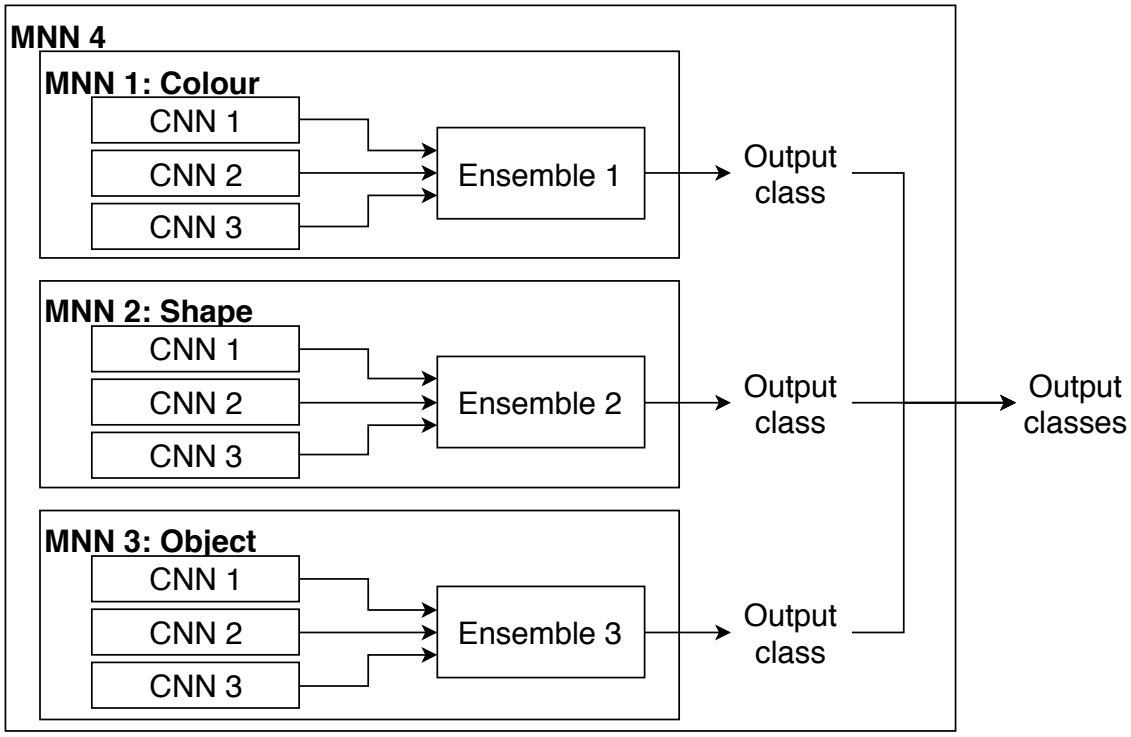


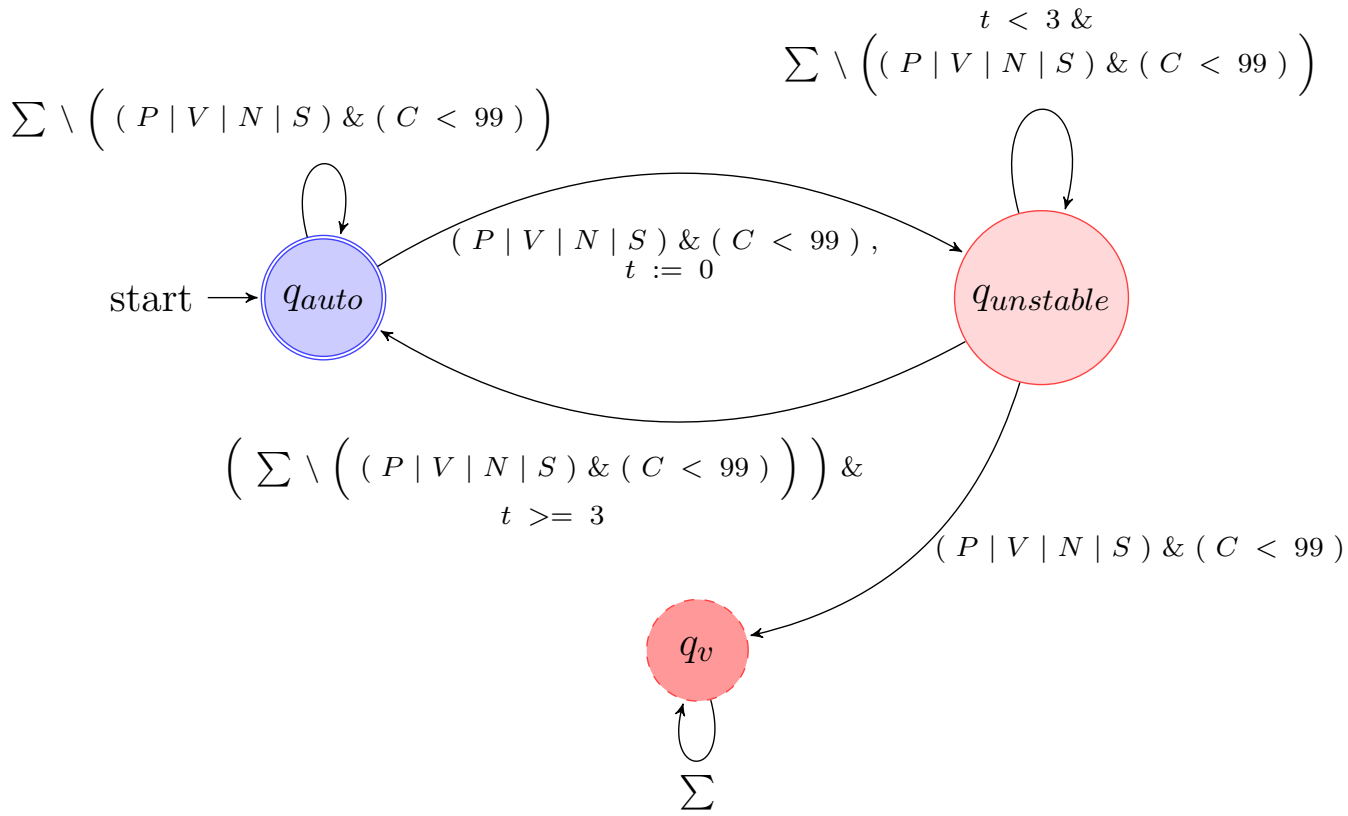
Figure 5.1: Block diagram showing the Meta Neural Network (MNN) ensemble

5.3.3 Input Perturbation Safety Policy

A safety (timed) automata **??** was designed to verify predictions made by the MNN and ensure utmost safety at all times. The automata started in a safe state, where control of the vehicle was autonomously handled by the system controller. If a misclassification was detected, the enforced policy entered an unstable state, still under autonomous control. Once enough time passed without further misclassifications, the vehicle entered the safe state again. However, if another misclassification was detected while unstable, the enforced policy entered a violation state and forced control of the AV to the driver. The vehicle would not enter autonomous mode again until the system was restarted. A diagram of the enforced policy's safety automaton is shown in Figure 5.2.

5.3.4 Runtime Enforcer for a Traffic Sign Recognition AV System

The system controller was encapsulated by a run-time enforcer [5] that used sensor fusion to check for misclassifications made by the MNN. This type of run-time enforcement, where neither the inputs nor outputs of the sensors or controller are enforced, has been termed as *run-time verification*. *Run-time verification* refers to the verification of system parameters during run-time, while ensuring that the system is aware of any failed guards in the enforced policy.



- P: Misclassification of a person.
- V: Misclassification of a vehicle.
- N: Classification of an object when there is nothing.
- S: Misclassification of a traffic sign.
- C: Confidence rating of the SNN classification.
- t: Timer for the unstable state.

Figure 5.2: Enforcer policy for the Autonomous Vehicle (AV) prediction system

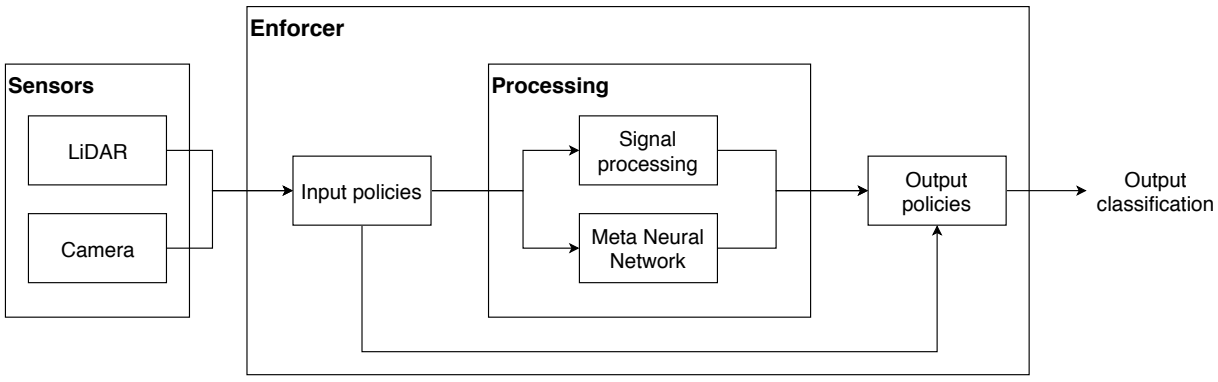


Figure 5.3: Block diagram showing the acfAV system with enforcer

5.4 Results of the Runtime Verified AV System

This research provides a solution for two aspects of Autonomous Vehicle (AV) systems: predicting accurately with perturbations to the system's inputs and safely dealing with misclassifications by the system. The issue of input perturbations was addressed using a Meta Neural Network (MNN) of different convolutional Synchronous Neural Networks (SNNs), each SNN working in tandem to predict more accurately. Misclassification by the system's controller was addressed by implementing sensor fusion between cameras and LiDAR. This was done using a run-time enforcer that enforced a safety automaton.

To test the MNN's ability to deal with perturbations, the input images (taken from the VOC 2010 and GTSRB datasets) were perturbed by randomly replacing approximately 7% of the image pixels with randomly coloured pixels. Table 5.1 shows that without perturbations, the accuracy of the MNN was increased from 87.07% to 93.7% when using an enforced policy. Table 5.2 shows that the accuracy of the MNN hugely decreases when perturbations are present, from 87.07% to 41.6%. However, with sensor fusion and a relative safety policy, the accuracy is increased to 90.48%.

Test case	Person	Vehicle	Nothing of relevance	Street sign	Total
Number of misclassifications	1.23	3.82	3.27	4.61	12.93
Caught misclassifications	1.00	3.64	3.20	3.75	11.59
False negatives	0.79	0.90	0.37	2.90	4.96
Missed misclassifications	1.02	1.09	0.44	3.75	6.30

Table 5.1: Table showing results of the AV prediction SNN using the original images

Test case	Person	Vehicle	Nothing of relevance	Street sign	Total
Number of misclassifications	3.59	9.20	6.58	39.03	58.40
Caught misclassifications	2.73	8.00	6.37	35.87	52.98
False negatives	1.11	0.58	0.12	2.29	4.10
Missed misclassifications	1.97	1.78	0.32	5.45	9.52

Table 5.2: Table showing results of the AV prediction SNN using perturbed images

5.5 Summary

Summary of everything previously mentioned in this chapter

5.6 Discussion

Discussion of the results of the case study and benchmarks. Discuss methodology and results.

6
Conclusions

This is the conclusion

6.1 Future Work

A

Appendix

References

[1] A. Benveniste, P. Caspi, S. A. Edwards, N. Halbwachs, P. Le Guernic, and R. De Simone, "The synchronous languages 12 years later," *Proceedings of the IEEE*, vol. 91, no. 1, pp. 64–83, 2003.

[2] G. Berry and G. Gonthier, "The ESTEREL synchronous programming language: Design, semantics, implementation," *Sci. Comput. Program.*, vol. 19, no. 2, pp. 87–152, Nov. 1992.

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[4] T. Gehr, M. Mirman, D. Drachsler-Cohen, P. Tsankov, S. Chaudhuri, and M. T. Vechev, "Ai2: Safety and robustness certification of neural networks with abstract interpretation," *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 3–18, 2018.

[5] S. Pinisetty, P. Roop, S. Smyth, N. Allen, S. Tripakis, and R. von Hanxleden, "Runtime enforcement of cyber-physical systems," vol. 16, pp. 1–25, 09 2017.

[6] P. S. Roop, H. Pearce, and K. Monadjem, "Synchronous neural networks for cyber-physical systems," in *MEMOCODE '18 Proceedings of the 16th ACM-IEEE International Conference on Formal Methods and Models for System Design*, 2018.

[7] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[8] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, no. 0, pp. –, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608012000457

[9] P. Wei, L. Cagle, T. Reza, J. Ball, and J. Gafford, "LiDAR and Camera Detection Fusion in a Real Time Industrial Multi-Sensor Collision Avoidance System," *ArXiv e-prints*, Jul. 2018.

[10] T. Wu, C. Tsai, and J. Guo, "Lidar/camera sensor fusion technology for pedestrian detection," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec 2017, pp. 1675–1678.