ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

ONERA
THE FRENCH AEROSPACE LAB

Certification of AI-based systems: challenges and promises

Certified
AI

Université Fédérale
Toulouse Midi-Pyrénées

Agenda

Introduction

- ANITI institute
- Aeronautical certification

System level analysis

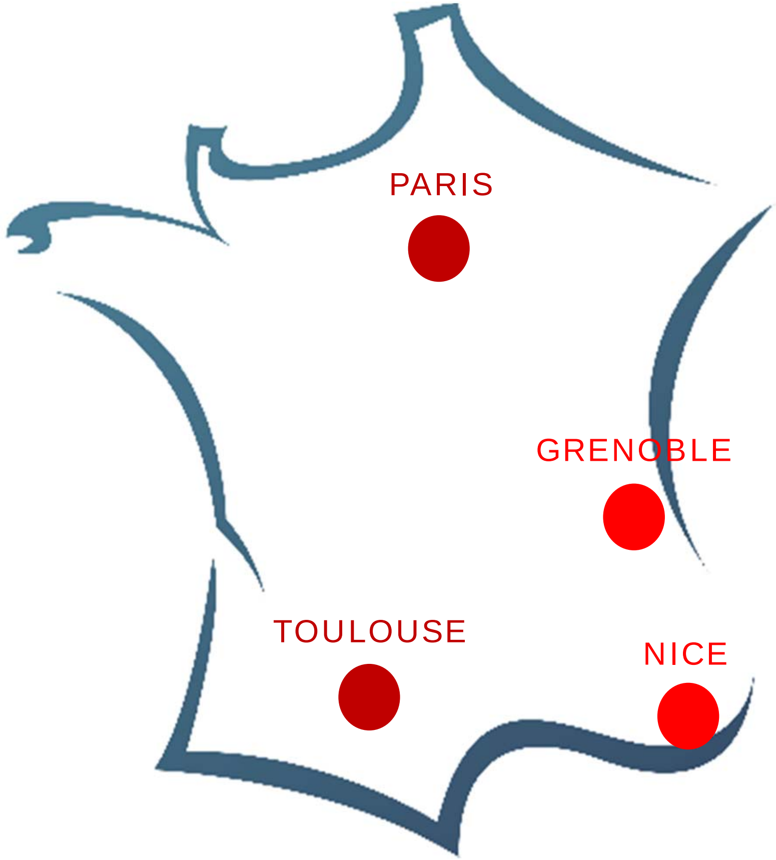
Zoom verification ACAS Xu

Zoom PHYDIAS

Conclusion

3iA: Interdisciplinary Institutes for AI

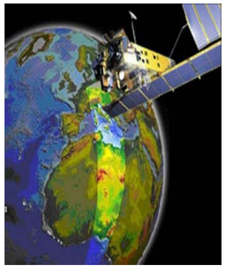
- Networked centers for research, education and economic development, with high international visibility
- 4 institutes
- Kick off: july 2019
- 4-year duration, renewable



ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE
Université Fédérale Toulouse Midi-Pyrénées

ANITI's Ambition

Make possible the sustainable use and development of AI in human critical applicative sectors (transport...) **and in** industry 4.0



Acceptability

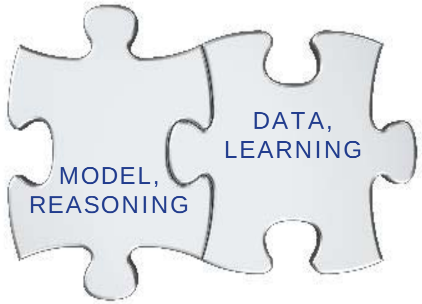
Fairness

Explainability

Robustness

Scalability

Adaptability



Hybrid **AI:** efficient combination of Model-based & Data-based AI

Partners

+50 PARTNERS

More to come !

Context: certification activities

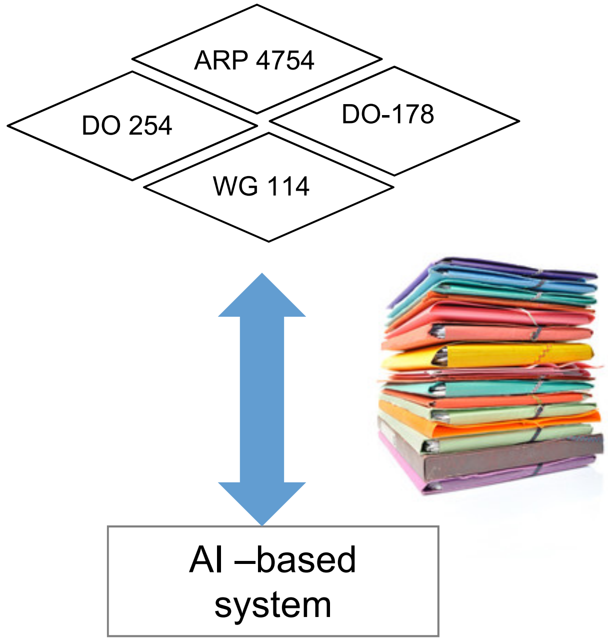
Certification:

- evaluation of an **argumentation**, to convince that a system (i.e., its architecture, its settings, including mitigation means. . .) satisfies **certification objectives** (expressed with AMC standards)

Difficulties :

- Existing standards are inapplicable [BCM+15]
  - o Data oriented specification

[BCM+15] Siddhartha Bhattacharyya, Darren Cofer, David J.Musliner, Joseph Mueller, and Eric Engstrom. Certification considerations for adaptive systems. Technical Report NASA, 2015



Certification: bibliography

- EASA – Concepts of Design Assurance for Neural Networks (CoDANN) – March 2020
- EUROCAE WG 114 / SAE G34 – Artificial Intelligence in Aeronautical Systems SoC (Statement of Concerns) – to be published soon
- AVSI (Aerospace Vehicule Systems Institute) – Machine Learning AFE 87 – June 2020
- White paper ANITI/DEEL/IRT Saint Exupéry: Machine Learning in Certified Systems – to be published soon

Talk: Focus on supervised learning and deep learning only

Agenda

Introduction

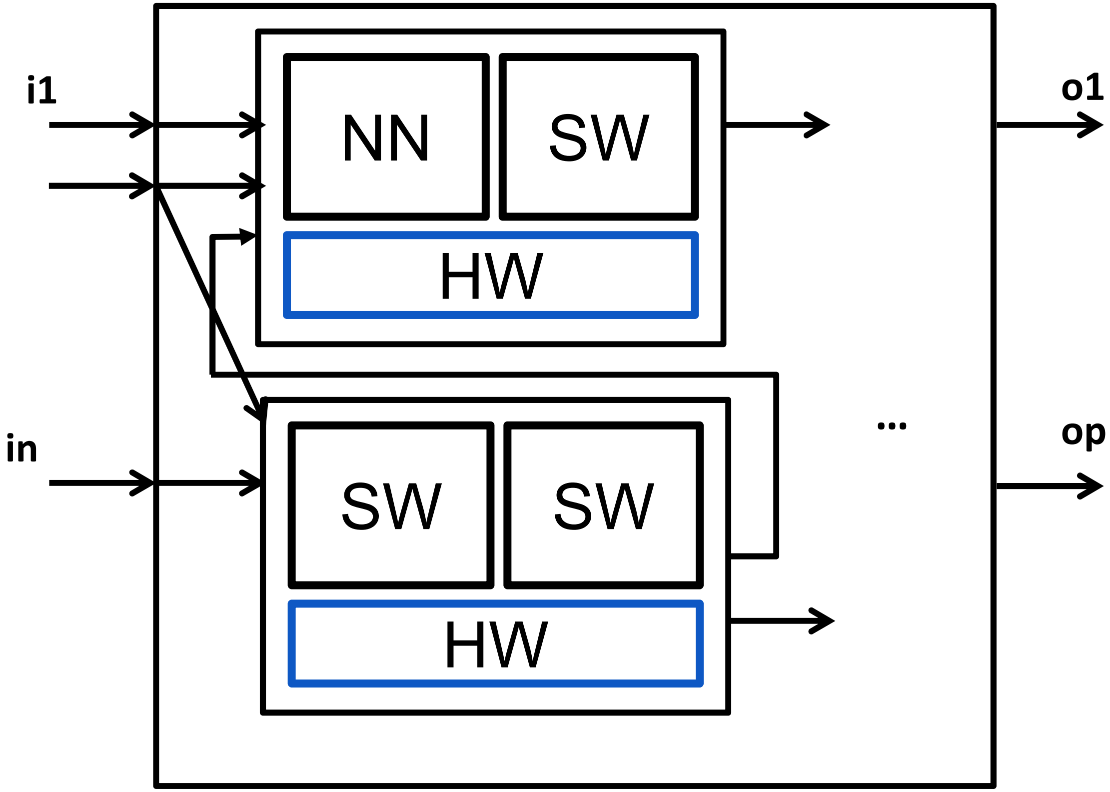
System level analysis

Zoom verification ACAS Xu

Zoom PHYDIAS

Conclusion

AI-based system



Objectives:

- ❑ System loss ≤ 10^{-9}FH
- ❑ Development process, test and verification at software level
- ❑ …

Example: ACAS Xu

GENERAL

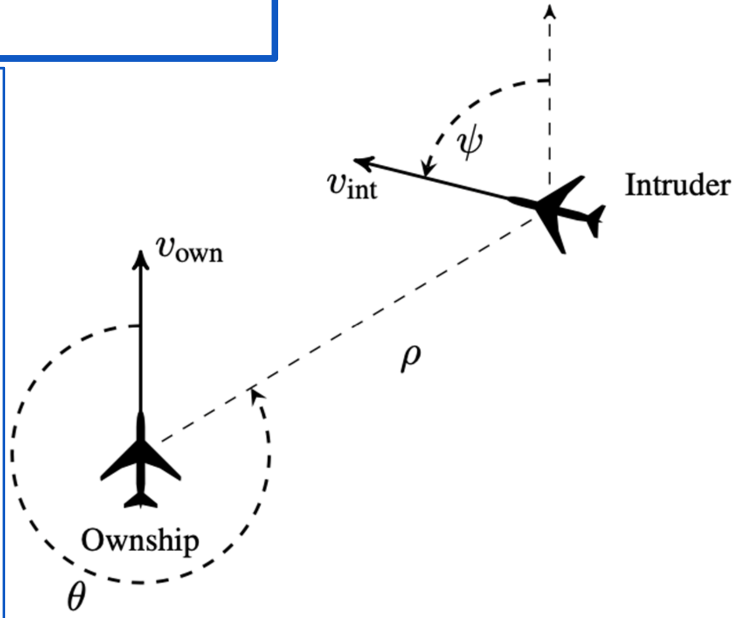
- Avoidance System for vertical and horizontal cooperative and non-cooperative avoidance
- Multi-Intruders

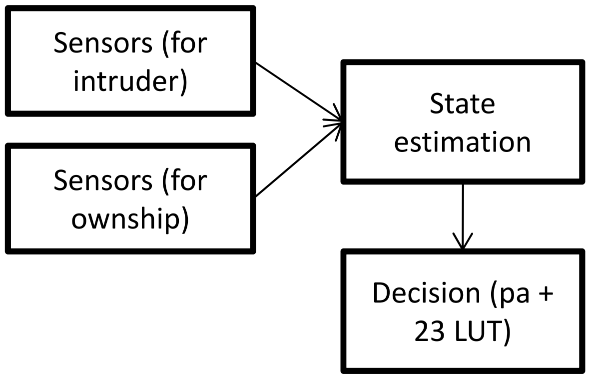
Why within AI consideration?

- On going studies to replace LUT (look-up table) with NN (seminal work Reluplex)

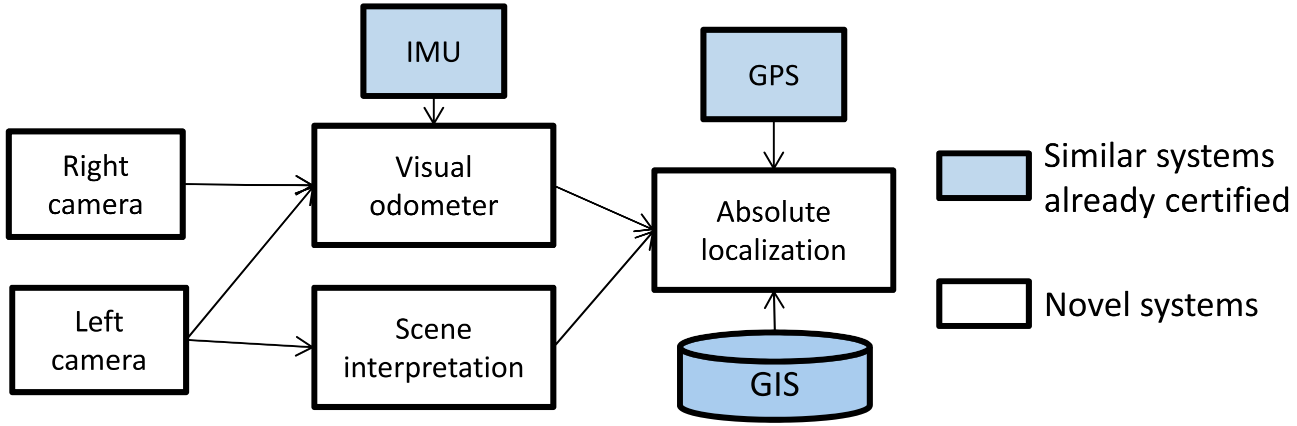
Safety Objective: FC = "the intruder enters the ownship enveloppe“ is Catastrophic

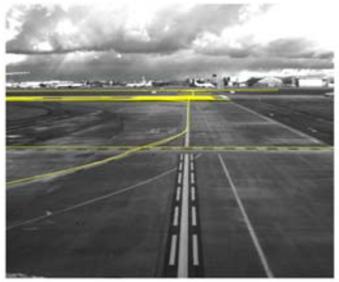
[Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks. Guy Katz, Clark Barrett, David L. Dill, Kyle Julian, Mykel J. Kochenderfer. CAV 2017]





Example: autonomous taxi driving





GENERAL

- Autonomous driving on pre-defined airports

Architecture:

- **Geographic information system (GIS)**: certified data base with airport maps
- **Visual Odometer (VO)**: estimate the trajectory wrt some relative reference
- **Scene Interpretation (SI)**: build a description of the scene
- **Absolute localization (AL)**: estimate the absolute position by fusing information

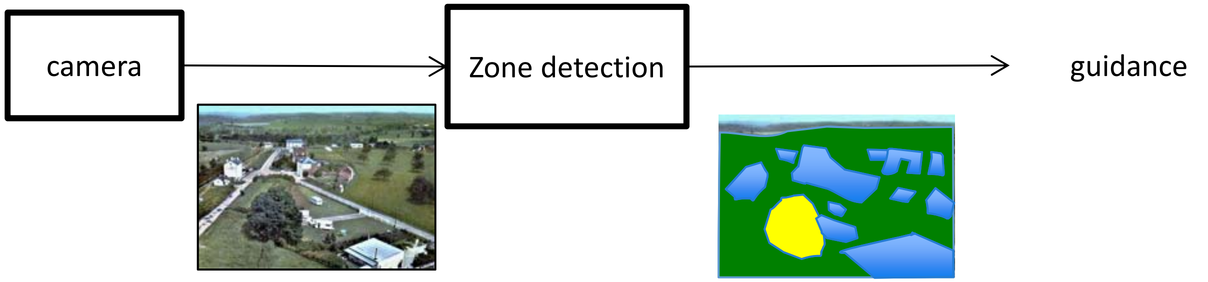
Safety Objective: FC = "the function provides a wrong position without the error being detected“ is Hazardous

Example: UAS emergency landing

GENERAL:

- Autonomous flight on a pre-defined trajectory
- Several back-ups in case of internal failures. Among the back-ups, emergency landing based on vision

Architecture: Mixing scene interpretation algorithms

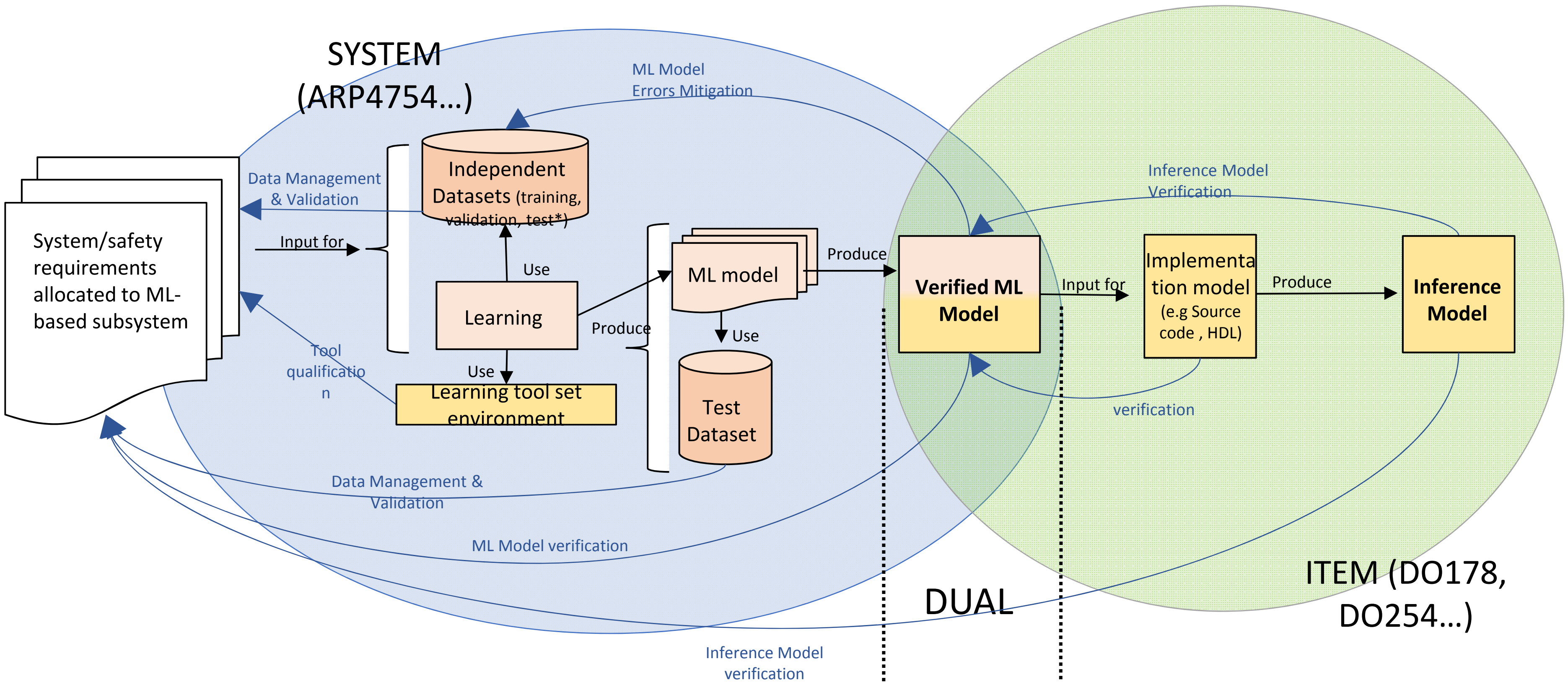


RGB image of the in front ground

Zones identification

Safety Objective: FC = ”deciding to land on a non planar zone, or a zone where a person or a property (car, house, warehouse) stand ” is Hazardous

EUROCAE WG114 current vision



Main examples of use for NN

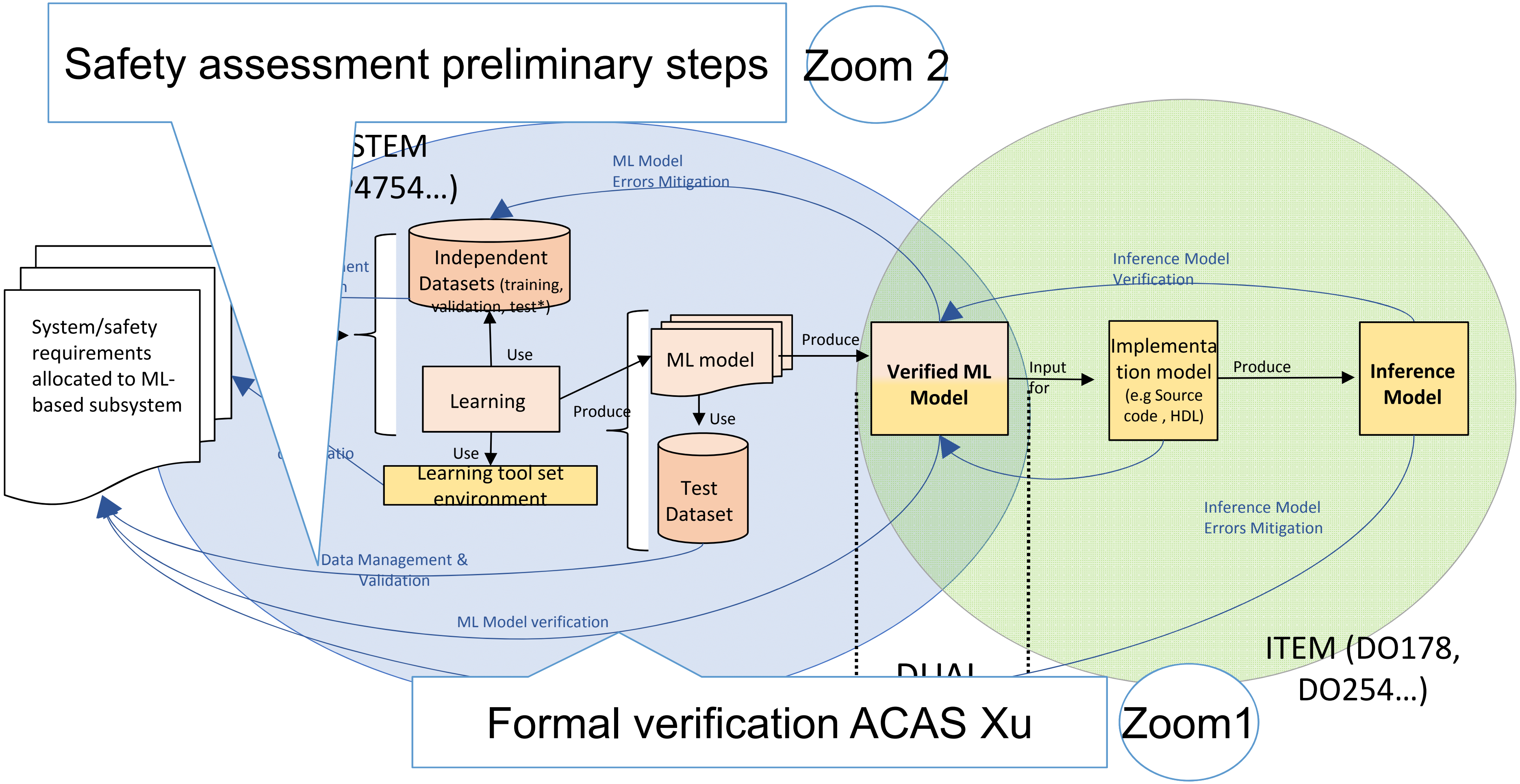
- **Existing certified SW**
  - Ex: certified look up tables replaced by NN
  - Why: increase performance of code (smaller memory footprint)

- **Embedding of design computation code (surrogate model)**
  - E.g.: certified Fortran code that takes 5hours to compute a result
  - Why: increase performance of the aircraft

- **Embedding of fully new system**
  - Ex: obstacle detection with camera
  - Why: increase of autonomy, …

Difficulty / novelty in terms of certification

Technical zooms



Agenda

ANITI
ARTIFICIAL & NATURAL INTELLIGENCE
TOULOUSE INSTITUTE

Introduction

System level analysis

Zoom verification ACAS Xu

- Collaborative work with DEEL partners (Mathieu Damour – Scalian Florence De Grancey – Thales, Christophe Gabreau – Airbus, Adrien Gauffriau – Airbus, Jean-Brice Ginestet – DGA, Alexandre Hervieu – DGA, Ludovic Ponsolle – APSYS)
- Verification tool Arthur Clavière PhD Collins Aerospace (co-supervised with Eric Asselin – Collins Aerospace, Christophe Garion – ISAE Supaéro)

Zoom PHYDIAS

Conclusion

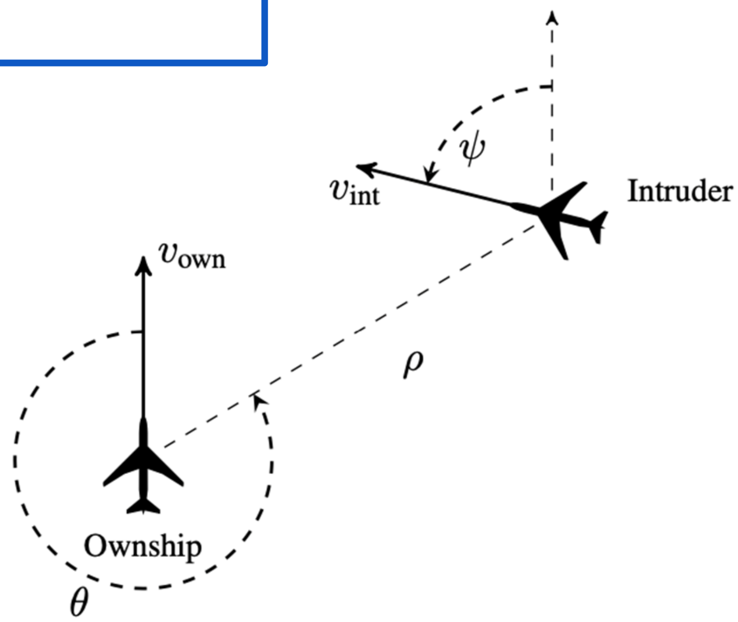
ACAS Xu overview

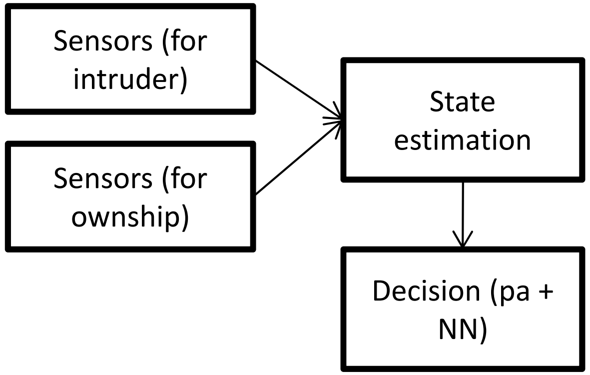
GENERAL

- Avoidance System for vertical and horizontal cooperative and non-cooperative avoidance
- Multi-Intruders
- EUROCAE WG 75.1 / RTCA SC 147

HOW IS IT WORKING

- Model of vehicle with Markov Decision Process
- Dynamic programming to compute Offline cost tables that enable to never have a vehicle in the collision volume
- Validation: large number of simulation and some flight tests



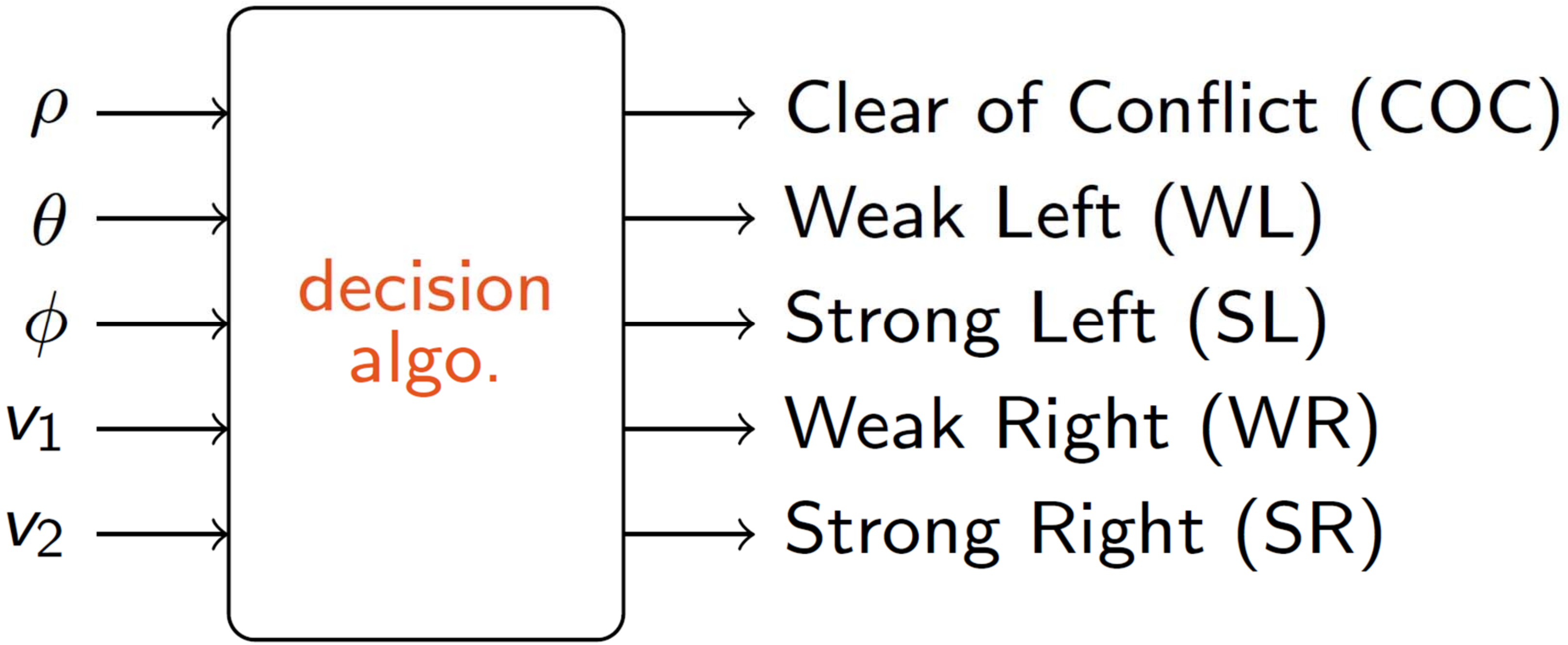


Why neural networks?

Several American universities (Standford, MIT) try to replace the LUT with NN

➔Gain in memory footprint (from 4Go to 3Mb)

➔Good anti collision performance



Certification proposed approach

How to adapt the certification activities of the ACAS Xu

- when replacing the LUT (lookup tables) with NN

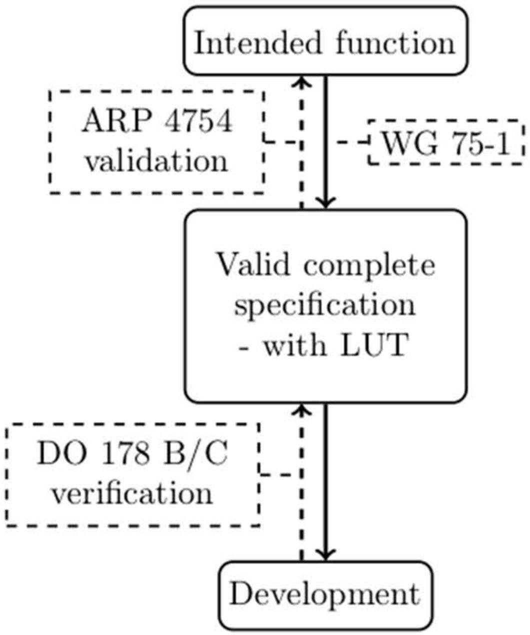


Figure 3: Classic approach

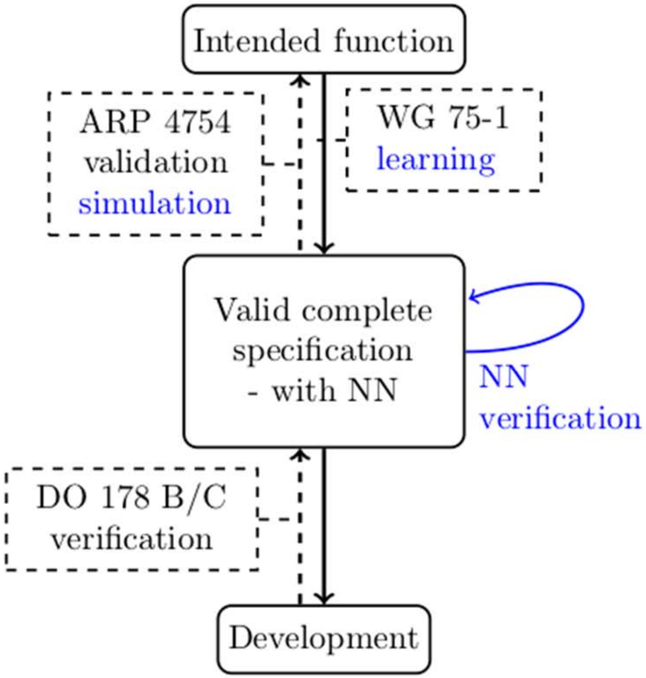
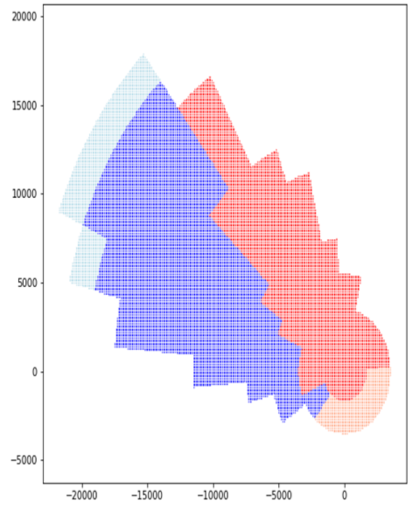
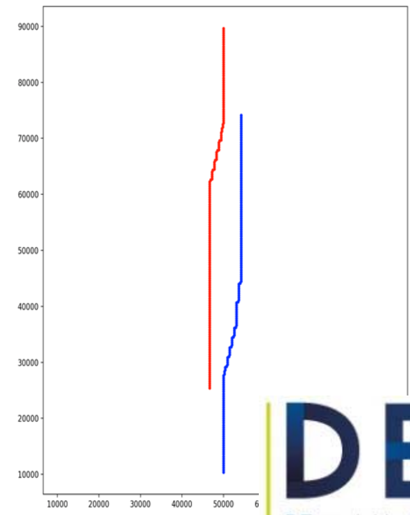


Figure 4: New approach

Developed supporting tools for ACAS-Xu

- **Get Binary tables provided by RTCA**
- **Parsing using documentation and guessing**
- **Enable**
  - Python Notebook to explore configuration
  - ACAS-Xu Simulator





Learning

- **Number:**
  - A la Reluplex: 45 NN (depend on the previous action and vertical)
  - A la Marabou: 1NN
  - 1 per (decision, pa): diverse shape of « function »
- **Structure:**
  - A la Reluplex: 6 layers and 300 neurons per layer
  - Design space exploration to find « optimal » structure
- **Training set:**
  - all LUT in learning data set: to be as close as possible to the LUT
  - Splitting strategies
  - Data augmentation

Verification

- **NN approximate the LUT => not the same exact behaviour**
- **How to formally define an "acceptable behaviour"**
- **Currently: no answer**

- **Literature: 10 properties defined in the Reluplex paper**
  - example property 3: "*If the intruder is directly ahead and is moving towards the ownship, the score for COC will not be minimal.*"
  - Shall hold for all of the 45 NNs except three of them

Input constraints (5D box):

$$(1500 \text{ ft} < \rho < 1800\ ft) \wedge (-0{,}06\ rad < \theta < 0.06\ rad) \wedge (3.10\ rad < \phi < 3.14\ rad) \wedge (980\ ft.s^{-1} < v_1 < 1200\ ft.s^{-1}) \wedge (960\ ft.s^{-1} < v_2 < 1200\ ft.s^{-1})$$

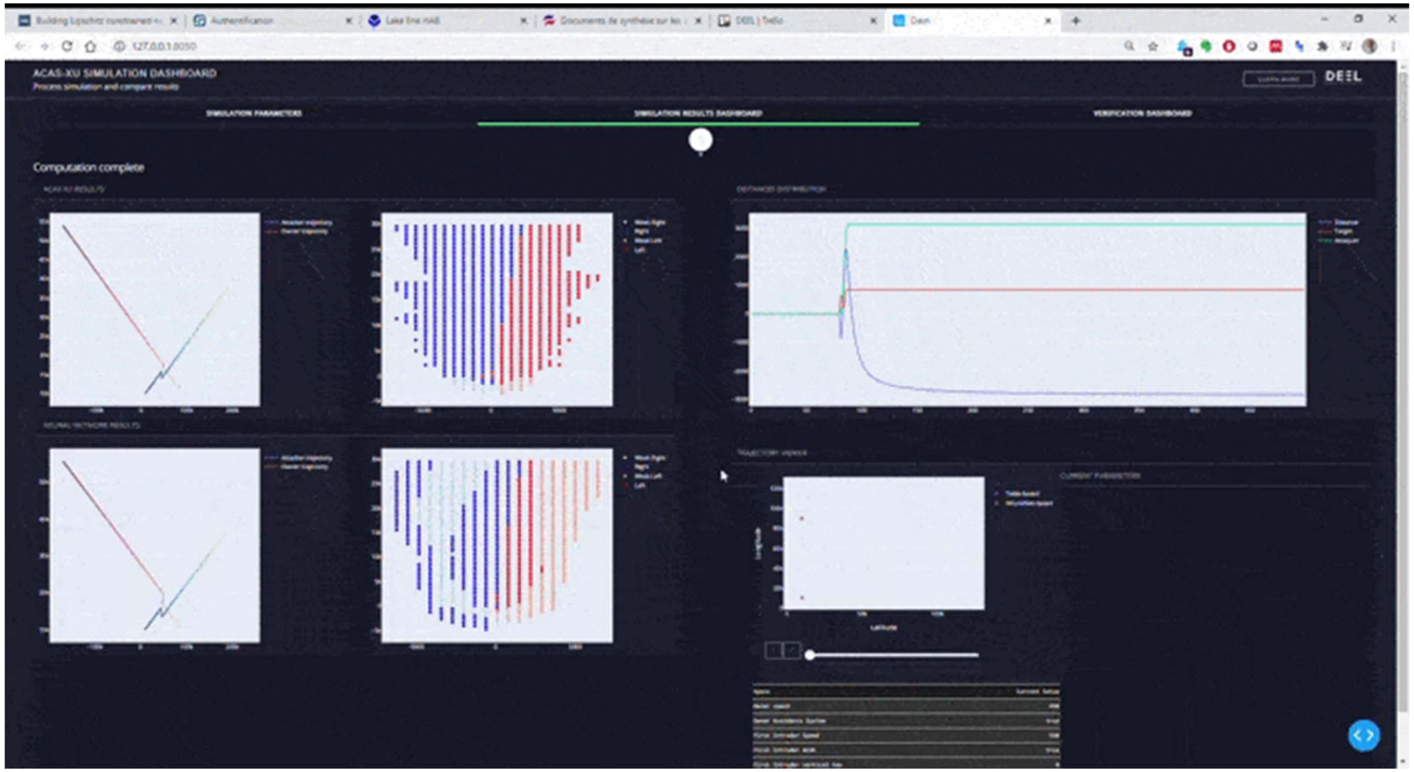
Output constraints (5D Halfspace polytope):

$$(COC > WL) \vee (COC > WR) \vee (COC > SL) \vee (COC > SR)$$

➔ Insufficient from certification perspective. Combination with simulation

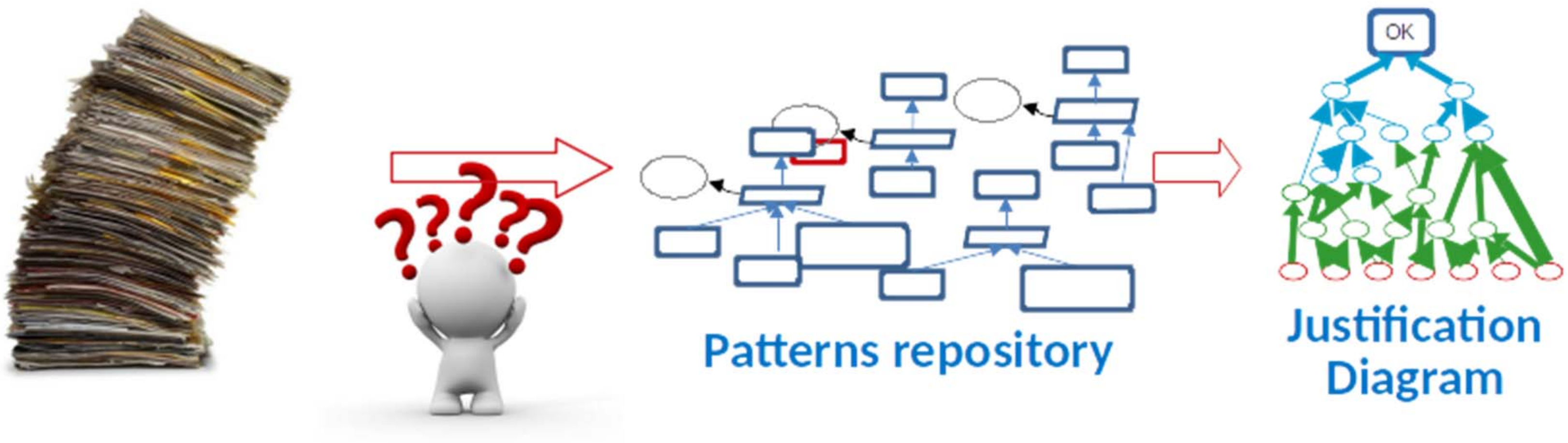
Simulation

- Intensive simulation
- Analysis of several indicators (partial explanation, …)



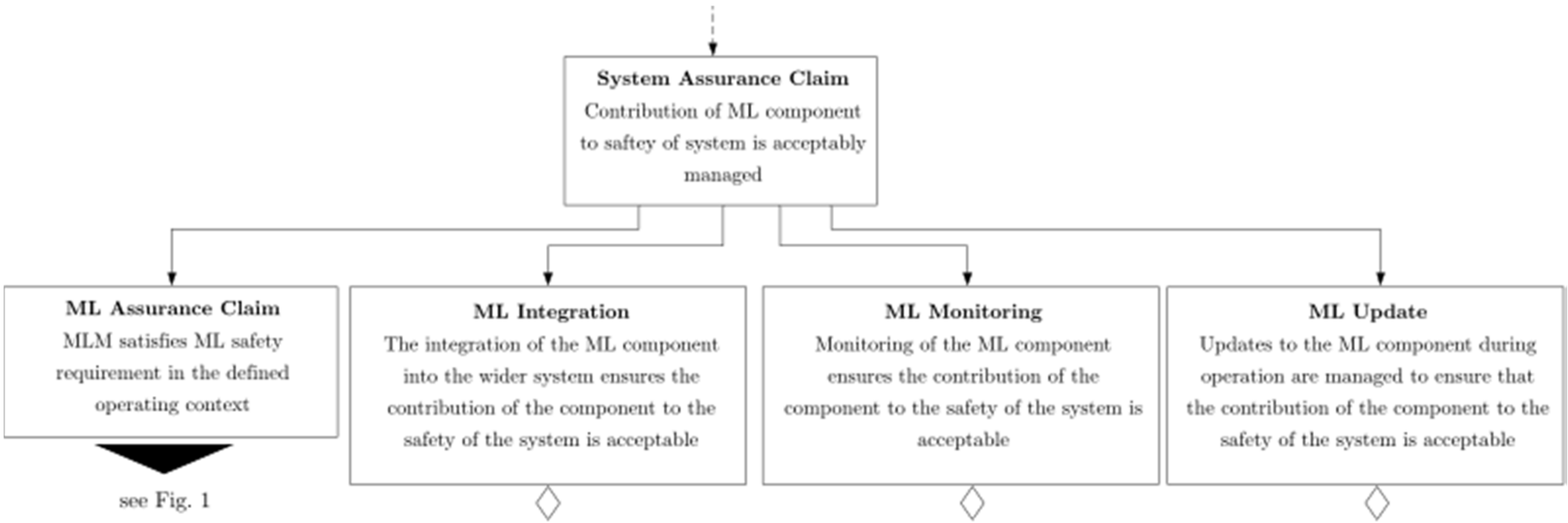
Assurance case reminder

- **structure, organize and share all these V&V items between stakeholders**
- ***an organized argument that a system is acceptable for its intended use with respect to specified concerns (such as safety, security, correctness)***
- **Concretely**
  - list necessary evidence related to the certification
  - structure key evidence (rationale)



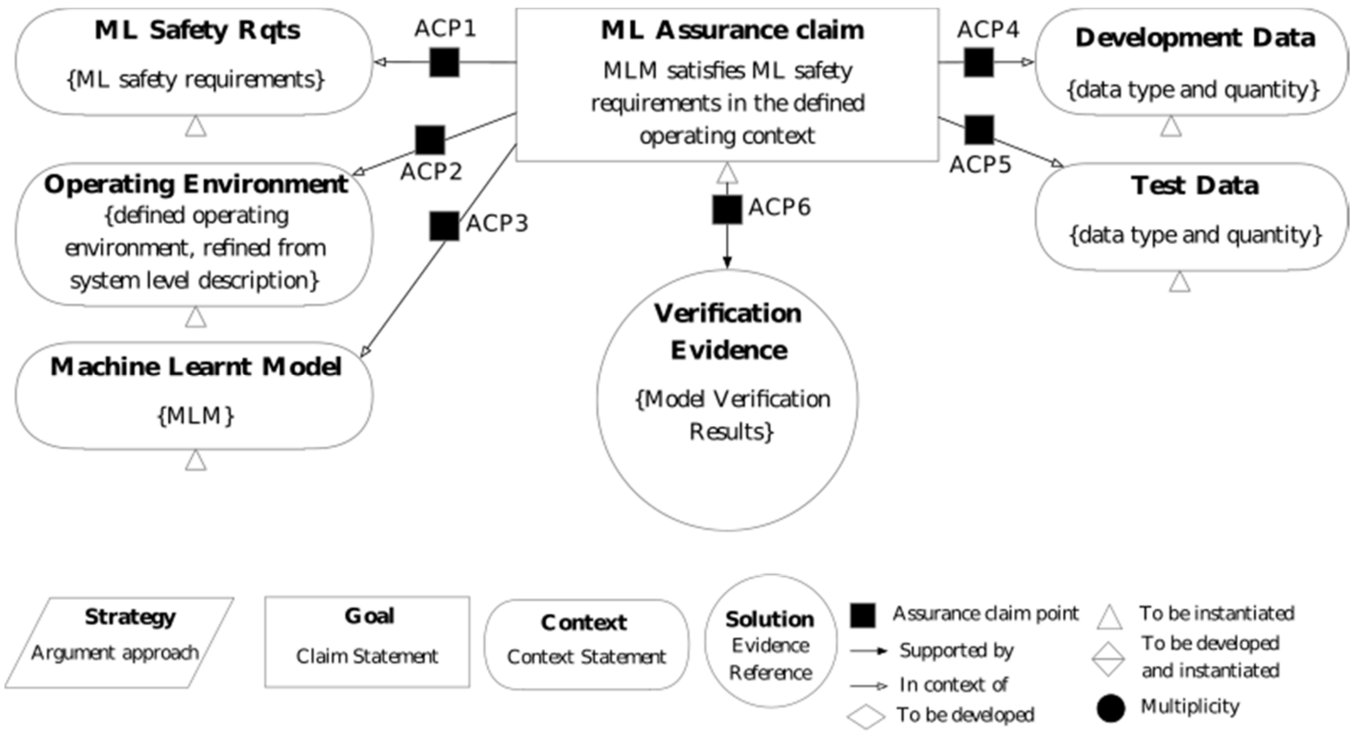
Assurance case for ML

Assurance Argument Patterns and Processes for Machine Learning in Safety-Related Systems. Chiara Picardi, Colin Paterson, Richard Hawkins, Radu Calinescu, Ibrahim Habli. Proceedings of the Workshop on Artificial Intelligence Safety (SafeAI 2020) 2020



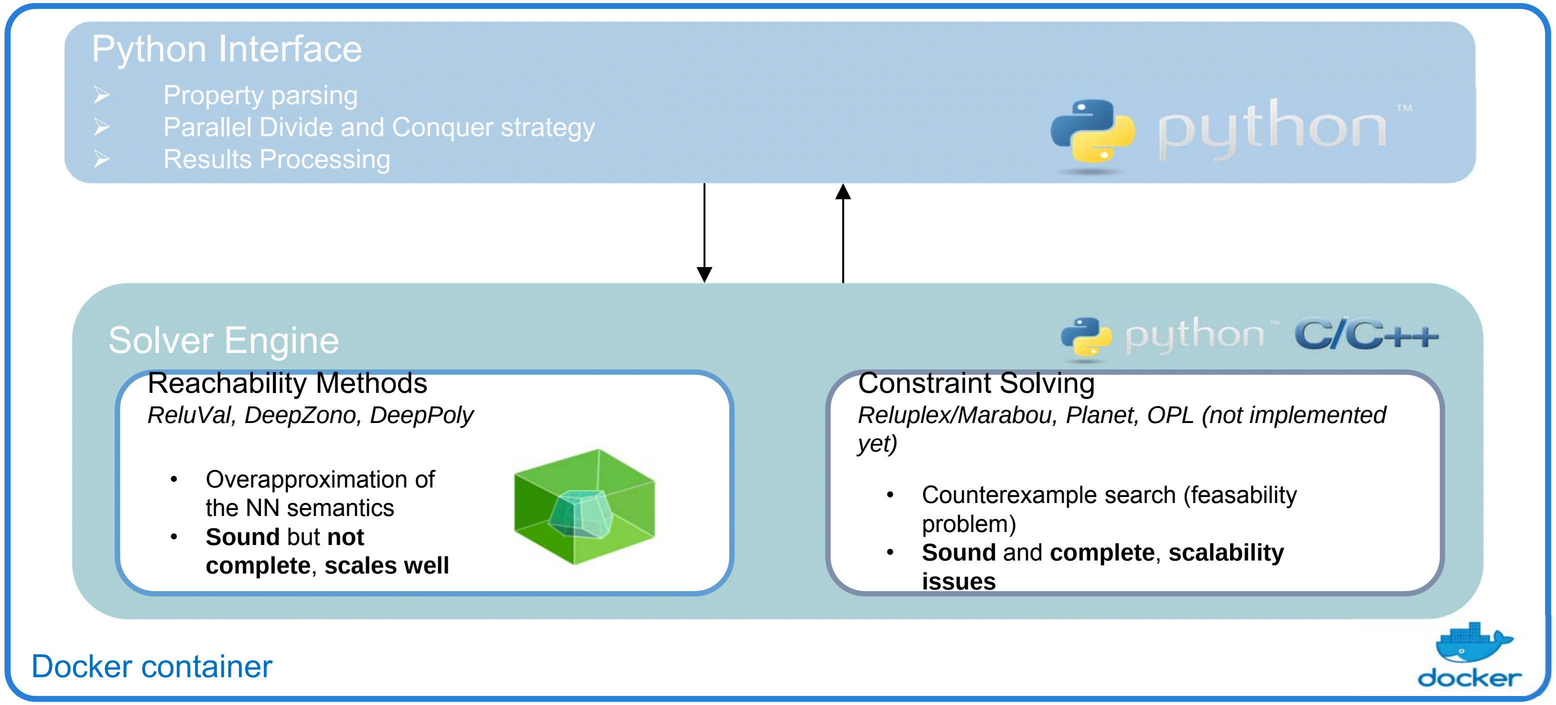
Assurance case for ML

Assurance Argument Patterns and Processes for Machine Learning in Safety-Related Systems. Chiara Picardi, Colin Paterson, Richard Hawkins, Radu Calinescu, Ibrahim Habli. Proceedings of the Workshop on Artificial Intelligence Safety (SafeAI 2020) 2020



A Unified Framework for NN Verification

- **Common julia verification tool developed by C. Liu, T. Arnon, C. Lazarus, C. Barrett, and M. Kochenderfer, "Algorithms for Verifying Neural Networks," 2020 ➔ re-coded from scratch**
- **Proposed approach: Interface to call directly the original tools**



Agenda

Introduction

System level analysis

Zoom verification ACAS Xu

Zoom PHYDIAS

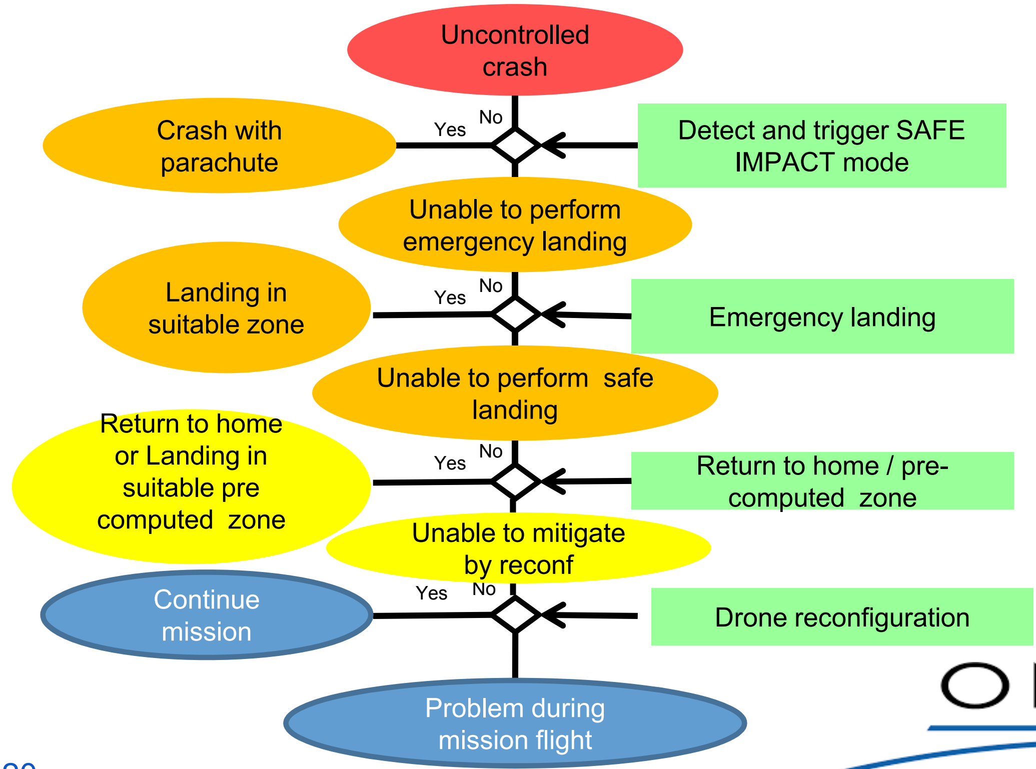
- Collaborative work with Frédéric Boniol, Adrien Chan-Hon-Tong, Kevin Delmas, Alexandre Eudes, Stéphane Herbin, Guy Le Besnerais, Martial Sanfourche

  [Challenges in certification of computer vision based systems for civil aeronautics. Aerospace Lab 2020]

Conclusion

Ground risk management

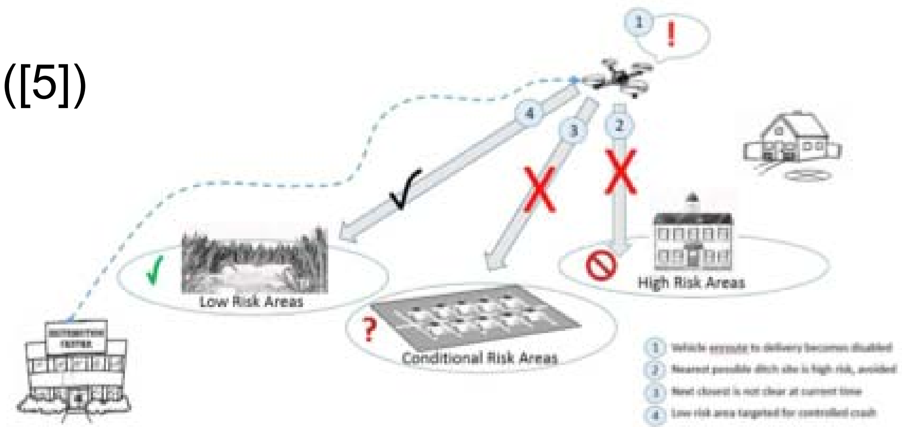
- Nominal flight plan: above sparsely populated zones
- Monitoring of flight plan correct following and health status of the drone
- In case of hazard, pre-defined procedures



State of the art

Numerous work on the topic "Autonomous crash management to a safe and clear site"

- "SafeUAV : Learning to estimate depth and safe landing areas for UAVs from synthetic data". Marcu et al. ECCV 2018. ([1])
- "UAV Emergency Landing Site Selection System using Machine Vision". Faheem et al. Journal of Machine Intelligence. 2015. ([2])
- …
- Safe2Ditch : start-up Nasa ([5])

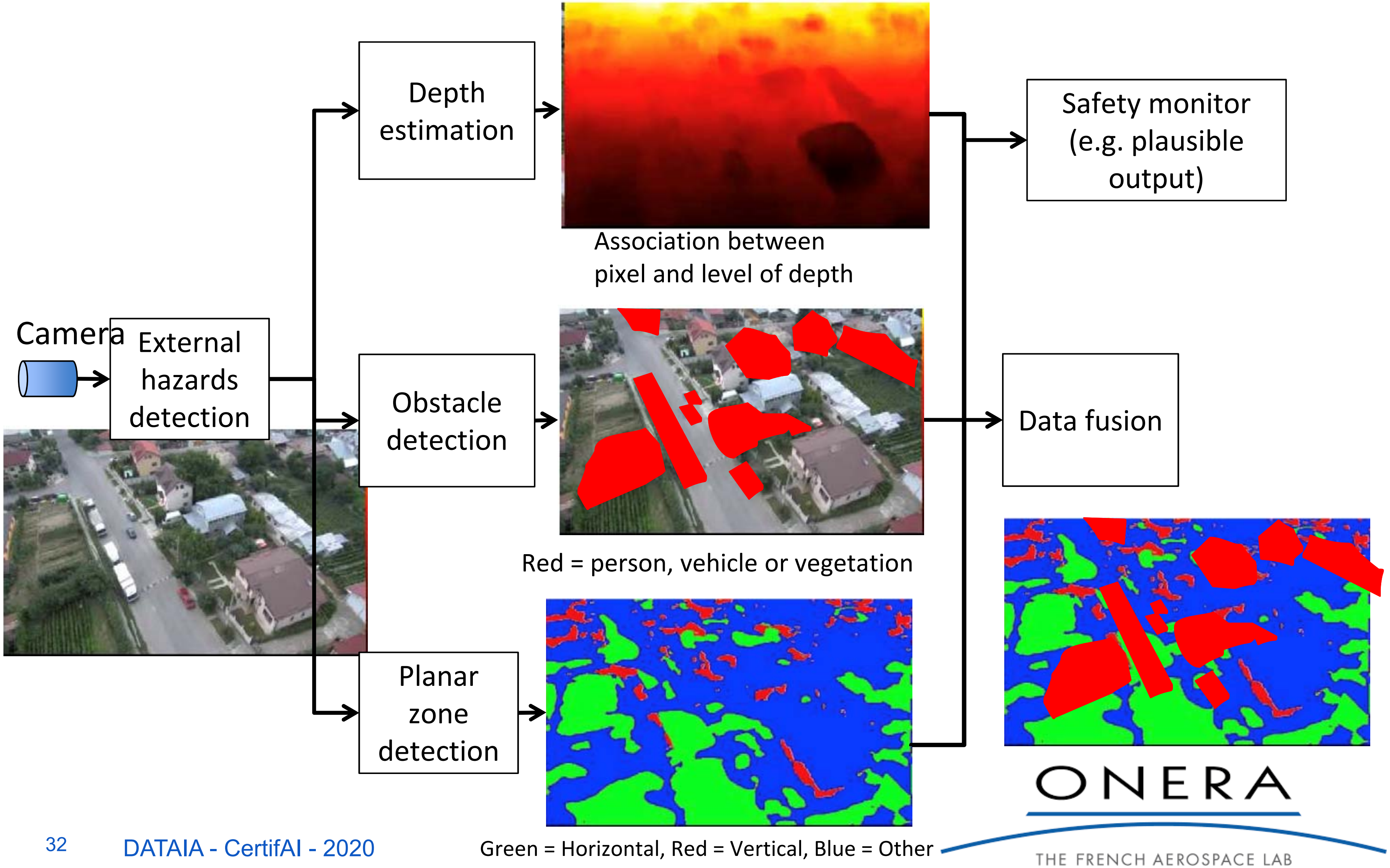


Sample Safe2Ditch Operational Scenario. Image credit: NASA

Safety considerations

- **Case 1 : not considered in the safety argumentation (safe drone)**
  - Emergency landing is an additional barrier, « best effort » (cf literature)
- **Case 2 : unconsidered in the safety argumentation (our case)**
  - FC = ”deciding to land on a non planar zone, or a zone where a person or a property (car, house, warehouse) stand ” is Hazardous
  - What is the detailed architecture?
  - What are the hazards?
  - How to realize the safety assessment?

Detailed architecture – 3 independent chains



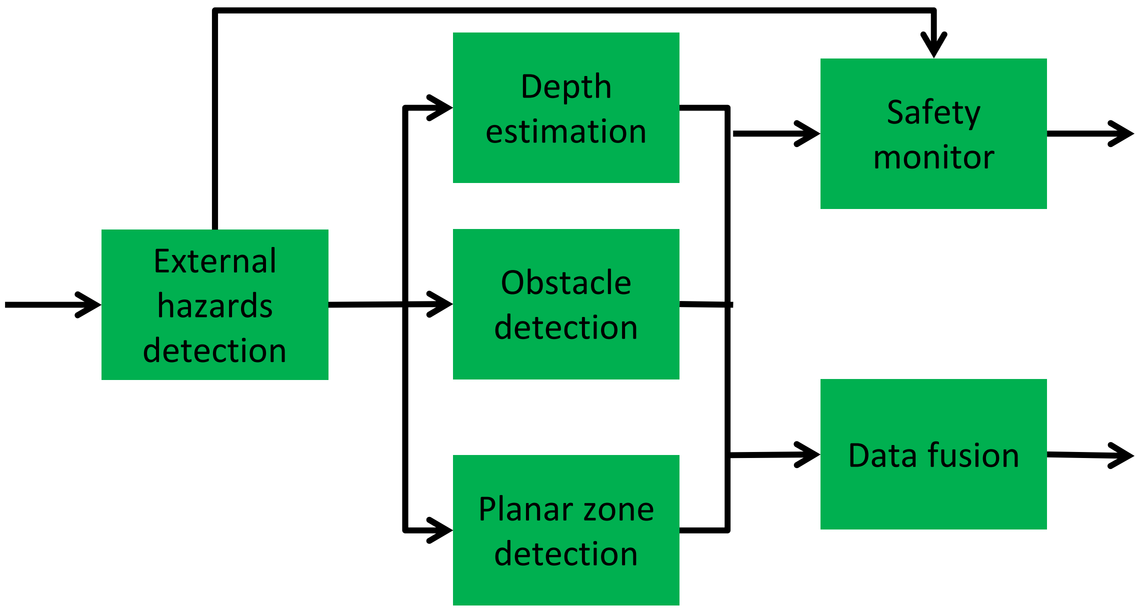
Identification of hazards

- **External events / hazards**
  - Vision hazards [ZMH+17] – CV HAZOP: Illumination (low illumination → low contrast); propagation conditions (e.g. smoke, haze); camera settings (e.g. aperture)…
  - occlusion
  - unreliable contrasted edges between illuminated areas and shadows
  - reflections related to water surface
  - …
- **Algorithm associated hazards** :
  - Incomplete specification: existing data sets for planar ground detection are very small
  - bad generalization
  - lack of robustness
  - …

[ZMH+17] Oliver Zendel, Markus Murschitz, Martin Humenberger, and Wolfgang Herzner. How good is my test data? introducing safety analysis for computer vision. International Journal of Computer Vision, 125(1-3):95–109, 2017.

Safety assessment

- How to associate some failure rate to a failure event that is not a hardware failure?
- How to define the failure propagation?
- How to combine probabilistic behaviour to determine the overall safety?



Agenda

Introduction

System level analysis

Zoom verification ACAS Xu

Zoom PHYDIAS

Conclusion

Conclusion & future work

- **Lot’s of pending work**

- **Finalisation of the ACAS Xu assurance case and associated evidence activities**

- **Safety assessment experiments for the emergency landing**

- **Implementation considerations for neural network**