Perception Challenge: Autonomous Vehicles

Leilani H. Gilpin
with Adam Amos-Binks and Dustin Dannenhauer
Autonomous Vehicles are Prone to Failure

Uber Example in my PhD Work
Uber Example in my PhD Work

Lack of Data and Challenges for AVs

• Existing Challenges
  • Targeted as optimizing a mission or trajectory and not safety.
  • Data is hand-curated
• Failure data is not available
  • Unethical to get it (cannot just drive into bad situations).
  • Want the data to be realistic (usually difficult in simulation).
• Develop a set of challenges and stress tests that generate new errors.
Existing Challenges and Benchmarks
Not Focused on Out of Domain Errors
Other Challenges Not Anticipatory

Not Focused on Error Detection
Autonomous Vehicle Limitations

• Complexity
  • Complex system build out of sensors, opaque software, and machinery.
  • It’s difficult to trace back what happened.

• Opaqueness
  • Proprietary mechanisms
  • Computer vision systems that are too opaque and dense to understand.
Current Approaches for Robust AVs

1. Error and failure analysis is **post mortem and reactive** instead of anticipatory.

2. Explanations are a **post mortem** tool.

3. **Lack of Redundancy**: Unlike aircrafts (that purposely has components that are repeated), autonomous vehicles that rely entirely on a single system for perception (e.g., Tesla camera system) and it is prone to failure and error.
Approach: Content Generation

Anticipatory Thinking Layer for Error Detection

DALL-E Generates “A chair in the shape of an avocado”

Synthetic images produced by StyleGAN, a GAN created by Nvidia researchers.
Need for Context

“Realistic” Adversarial examples
Approach: How it Works
Use Adversarial Images in Dev Testing

- Solution: Use a cognitive architecture that helps to anticipate and understand these failure cases.

- Assess autonomous vehicles for their risk management capabilities before being deployed and provide incident level risk management explanations in human readable form.
Isolated error detection

Integrated error detection

Explanatory Error Detection

Content generation
Larger Approach

Cognitive model of Navigation
- General navigation model
- Identify navigation landmarks
- Extend work on explanations

Landmark similarity metric
- General image embedding
- Identify similar landmarks
- Reuse existing datasets

Natural Adversarial Generation
- Generate 'chimera' landmarks
- Stochastic output
- Extend VAE/GANs (e.g., DALL-E)

Perception System Stress Test
- Dev-Training
- Dev-Generated
- Deploy
- All Errors
Impact
Anticipatory Thinking Layer for Error Detection

• Goal - Develop methods that a priori can explain an autonomous vehicle’s ability to manage the risks stemming from errors in perceiving their environment.

• One possible solution is to explain why the autonomous behavior is safe (or risky, trustworthy, etc.) or not.

• Impact - Consumer confidence and safety features, appropriate legal and regulatory oversight.