ECE/CS 498 DS - Section U/G (R. Iyer)

- **Meeting Time and Place:**
  - **Lecture (Mandatory):** 12:30 PM – 1:50 PM, Mondays and Wednesdays, 1302 Everitt Laboratory
  - **Discussion Session (Optional):** 4:00 PM – 5:00 PM, Fridays, CSL 141
- **Instructor:**
  - **Professor Ravi K. Iyer (rkiyer@illinois.edu)**
  - **Office: 255 Coordinated Science Laboratory (CSL)**
  - **Office Hours: 2:00 PM – 3:00 PM, Mondays, CSL 255**

    (Other times by appointment. Please contact Heidi Leerkamp: leerkamp@illinois.edu)
- **Teaching Assistants:**
  - **James Cyriac (jcyriac2), Vikram Anjur (vanjur2), Shengkun Cui (scui8), Haotian Chen (hc19)**
  - **Office Hours: 4:00 PM – 5:00 PM, Mondays and Wednesdays, CSL 249**

ECE/CS 498 DS - Section U/G (R. Iyer)

- **Piazza:**
  - **All class-related information will be announced via Piazza – Make sure to enroll!**
  - **http://piazza.com/illinois/spring2020/ececs498ds**
- **Class Website:**
  - **https://courses.engr.illinois.edu/ece498dsu/sp2020/**
  - **A detailed class schedule including topics covered and reading for each class is on the class website; we will adjust the schedule as needed**
  - **The classes are being recorded and will be available through class website**
  - **Lecture notes will be posted on the class website weekly**
- **Recommended Text:**
  - **Class Notes/Lecture Slides**
  - Trevor Hastie, Robert Tibshirani, and Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction"
- **Further Reading and Sample Problems:**
  - Daphne Koller and Nir Freidman, "Probabilistic Graphical Models: Principles and Techniques"
  - Ravi Iyer's ECE 313 Class Notes (link)

ECE/CS 498 DS - Section U/G (R. Iyer)

- **Course Components**
  - **Discussion Session:** Optional discussion session will be held Fridays from 4-5 PM.
    - Sessions will either (i) cover further elaboration on that week's lecture material or will (ii) serve as an additional office hour with the TAs
  - **In-Class Activities (ICAs):** Organized sessions during lecture time in which the instructor/TAs help students solve relevant problems.
    - Attendance is mandatory, and ICAs will be preannounced
  - **In-Class Quizzes:** Short 5-minute quizzes administered at the tail-end of random lectures throughout the semester.
    - Attendance is mandatory, and no quiz will occur during week of career fair
  - **Homework (HW)**: Worksheets/assignments related to lecture material that are to be completed within 1 week of release.
  - **Mini-Projects (MPs)**: Programming assignments to be completed in groups of 3 that incorporate end-to-end workflows that solve real-world data science problems.
    - The semester's 3 MPs will range across high impact domains such as autonomous vehicles and health analytics
  - **Final Project (for 4 credit hour students)**: Semester-long data science project to be designed and implemented in groups of 3 by 4-credit hour students.

Grading Policies

Activity	Contribution to Final Grade
Mini-Projects 1, 2, 3	45% (10%, 15%, 20% respectively)
Midterm and Final	35% (Midterm 15%, Final 20%)
Final Project (**4-Credit Students Only**)	30%
Class Participation	10%
Homework	10%

Credit Policy:

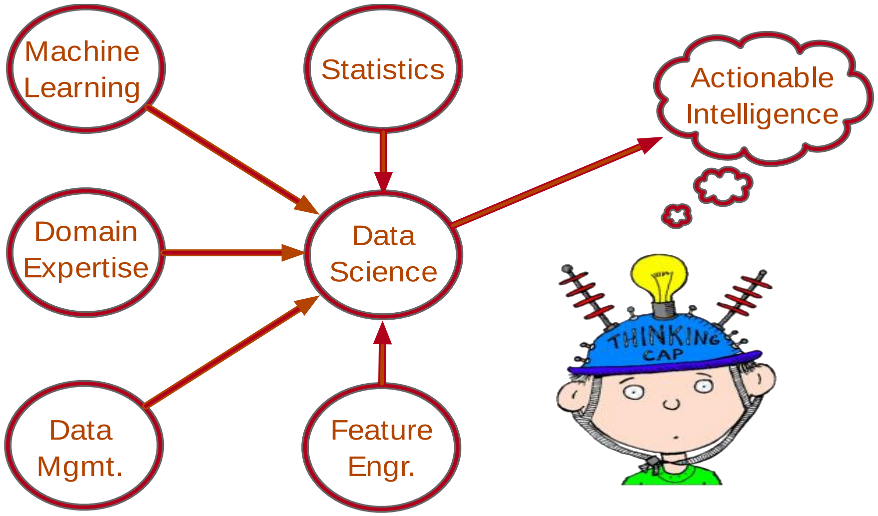
- 4 credit-hour students complete final project in addition to other course material
- Only graduate students are permitted to enroll for 4 credit-hour sections – undergraduates will require instructor approval
- Score distributions for 4 credit-hour students will be normalized to 100%

- **Class Participation:** 10% class participation grade comes from participation in (i) ICAs, (ii) lecture quizzes, (iii) lectures/office hours/discussion sections/Piazza.
  - Instructor/TAs reserve right to track attendance in all of these outlets
- **Late Submissions:** 10% will be taken off each day, prorated (up to 3 days max). 0 credit after that.
- **MP Groups:** **Students will form groups of 3 persons for the projects by the end of week 2**
  - A “Search for Partners” page will be created on Piazza by the end of this week
- **Academic Honesty:** While discussion among groups is encouraged, no sharing of answers/solutions is permitted
  - Groups that submit any identical material will incur penalties for academic dishonesty
- The lowest individual ICA score and in-class quiz score will be dropped

Other Information

- **Prerequisites**
  - Basic probability understanding (e.g. ECE 313/CS 361) and basic Python programming skills are essential.
  - Knowledge of Operating Systems (e.g., ECE 391 or an equivalent course) is beneficial.
  - HW0 to test basic probability skills, HW1 to test Python programming skills
  - **Talk to instructor if you find HW0 to be difficult**
- **Electronics Policy**
  - Students **SHOULD NOT** use electronics (cell phones/laptops/etc.) during lecture
  - Students **SHOULD** bring their laptops to class to use only in case of a quiz
- **Other**
  - **Midterm is scheduled for Wednesday March 11, 2020**
  - **Those requesting DRES accommodations should email instructor/TAs by Friday 1/24**
  - **Project descriptions and due dates will be posted on the course website under "Assignments and Mini-Projects"**
    - Please check "Resources" for tutorials and hints related to projects
  - Over 15 weeks in the spring semester, there will be ~33 hours of mandatory lectures and ~15 hours of optional discussion sections

ECE 498 DS



In this course,

- **Raw datasets to actionable intelligence**
- **Building end-to-end workflows that can work in practice**
- **Work on real-world problems of high societal impact**

Solving problems in real-world involves

- Understanding the problem and the associated domain
- Finding or building relevant datasets

Use of data-science techniques to find solutions by

- Converting raw datasets to usable features
- Choosing a model that best represent the data
- Validation of the models on the field

ECE 498 DS Course Structure

Problem solving and domain-driven analyses

You will be introduced to real-world problems in domains of societal importance: (a) Safety in AVs, (b) Health Analytics, and (c) Intrusion Detection

In-class lectures, quizzes, and activities

You will work on hands-on modeling tutorials with support from instructors

Homework assignments

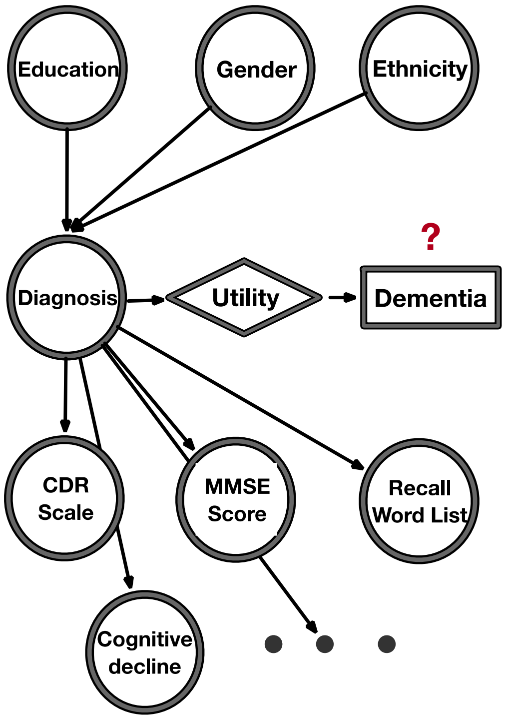
You will be given theoretical questions, and small programming assignments to strengthen your understanding the methods

Outcomes

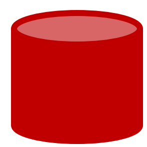
- ***Learn to handle the data***
- ***Learn to decide what model to use and when***
- ***Learn to apply the models to real-world examples***
- ***Learn to derive insights from by combining model solutions with domain knowledge***

ECE 498 DS : Examples from Real-world

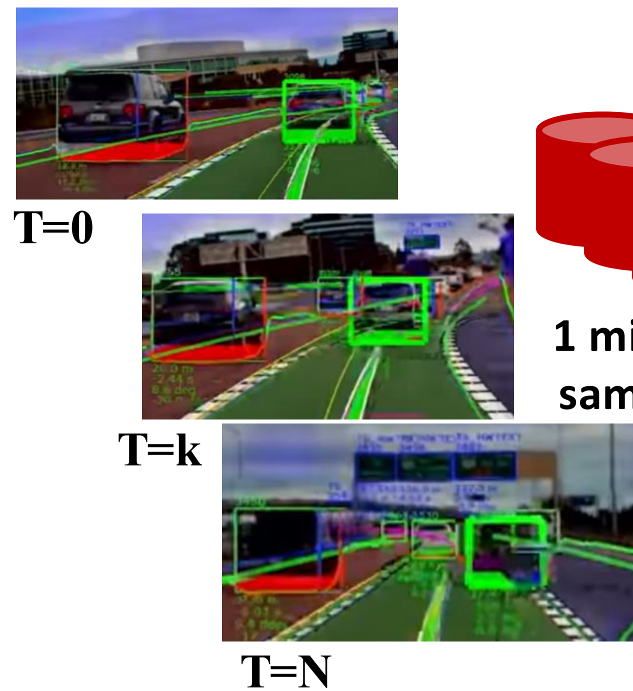
- Uncertainty in measurements, labels etc.
- Structure/Process
- Dataset size: #samples, signal to noise ratio

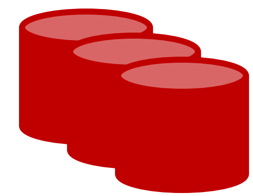


Seixas et al., *A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment*



1000 samples



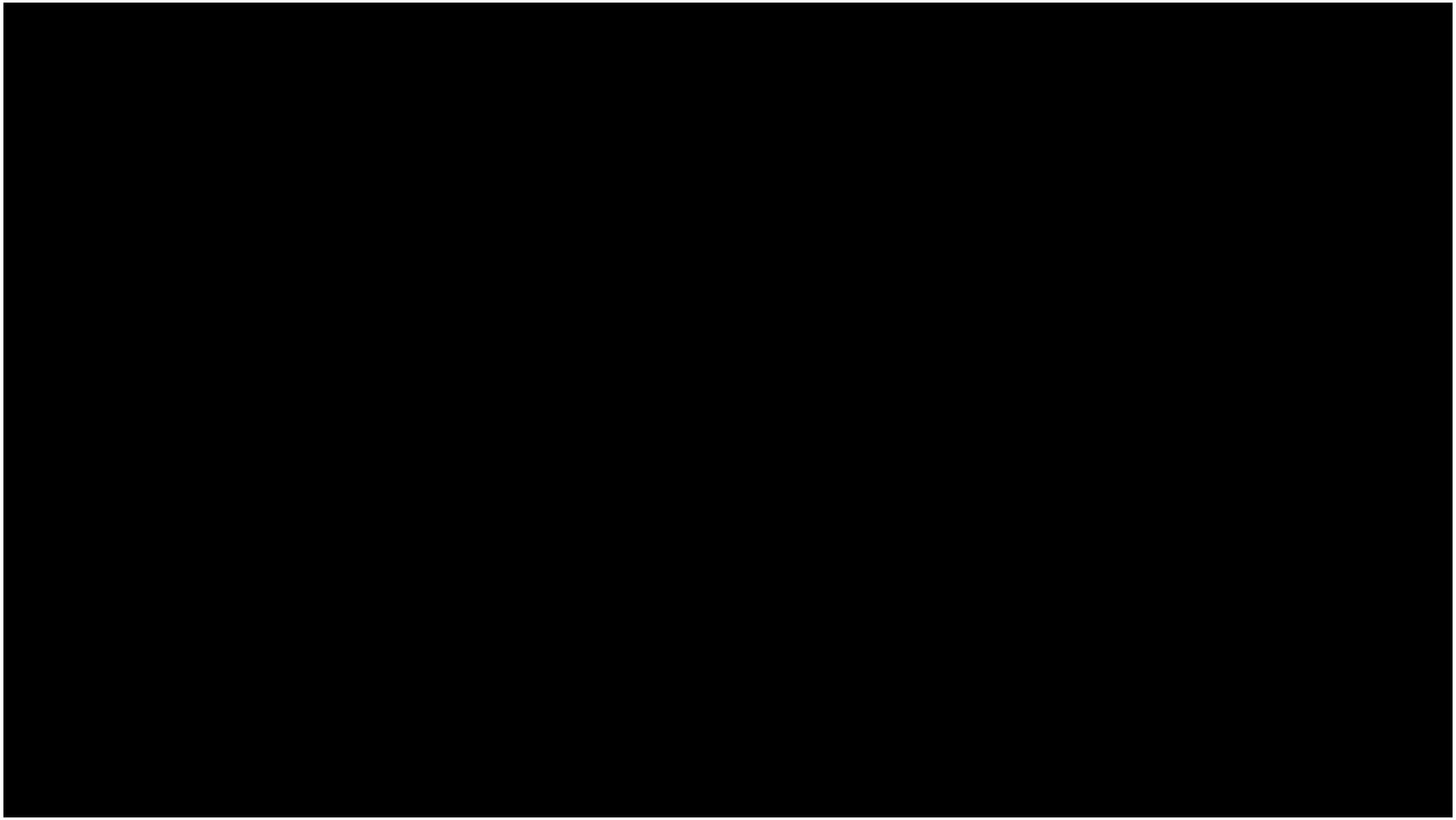


1 million samples

Does the patient suffer from Dementia ?

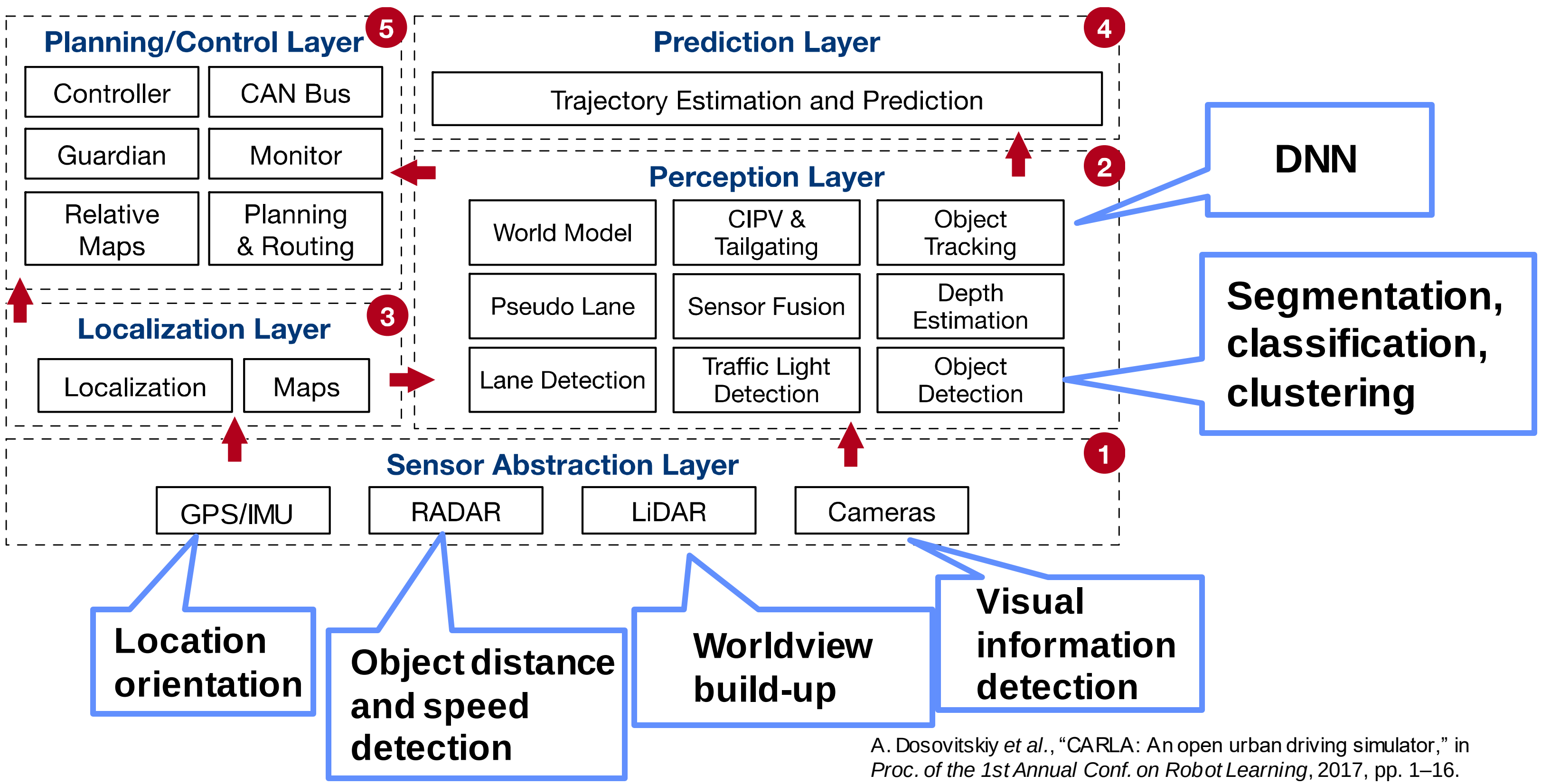
Identify object in front of the AV?

Integrating AI Techniques in a Real System: Autonomous Vehicles (AVs)



Source: https://www.youtube.com/watch?v=B8R148hFxPw

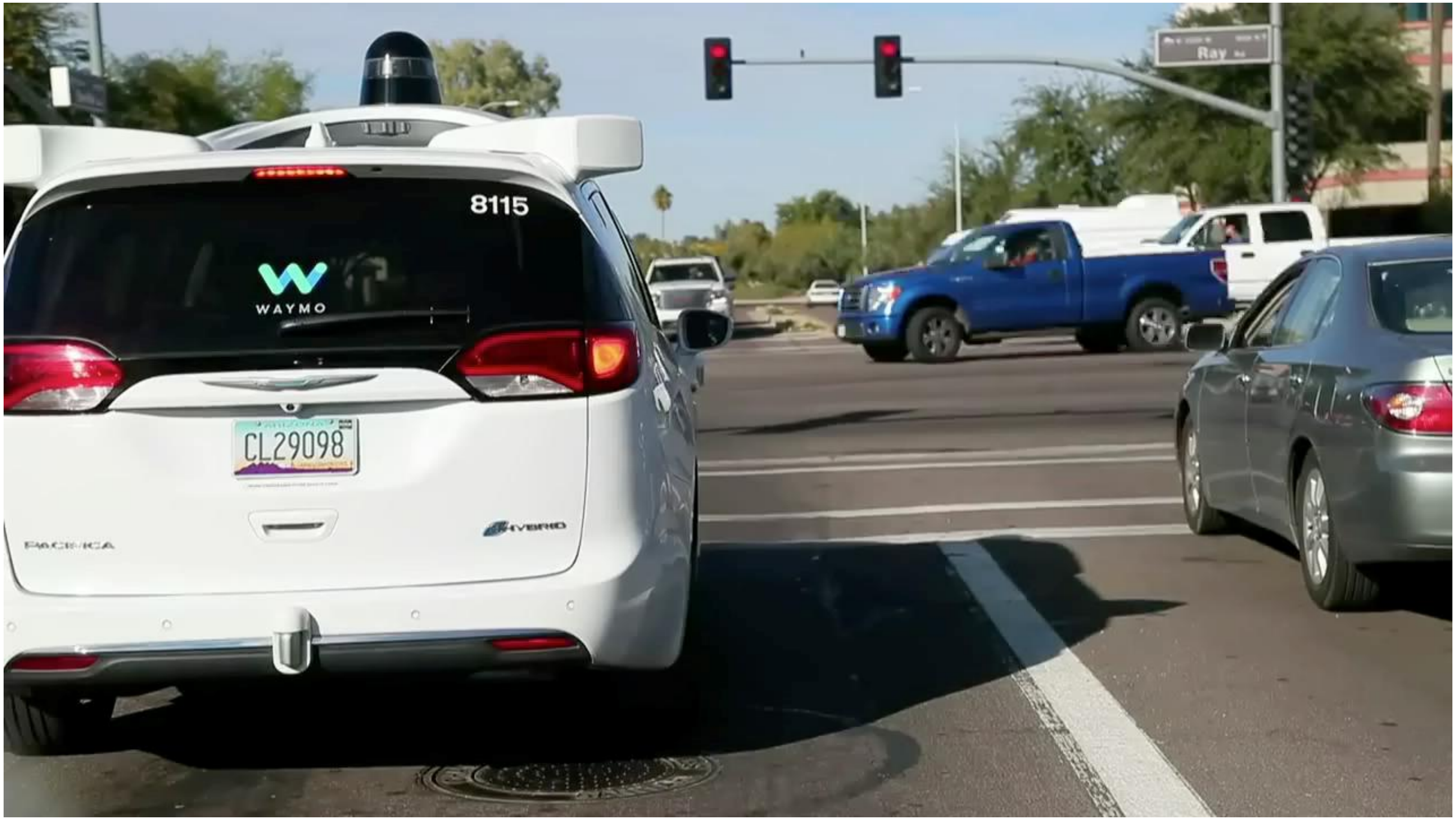
Autonomous Driving System (ADS)



A. Dosovitskiy *et al.*, "CARLA: An open urban driving simulator," in *Proc. of the 1st Annual Conf. on Robot Learning*, 2017, pp. 1–16.

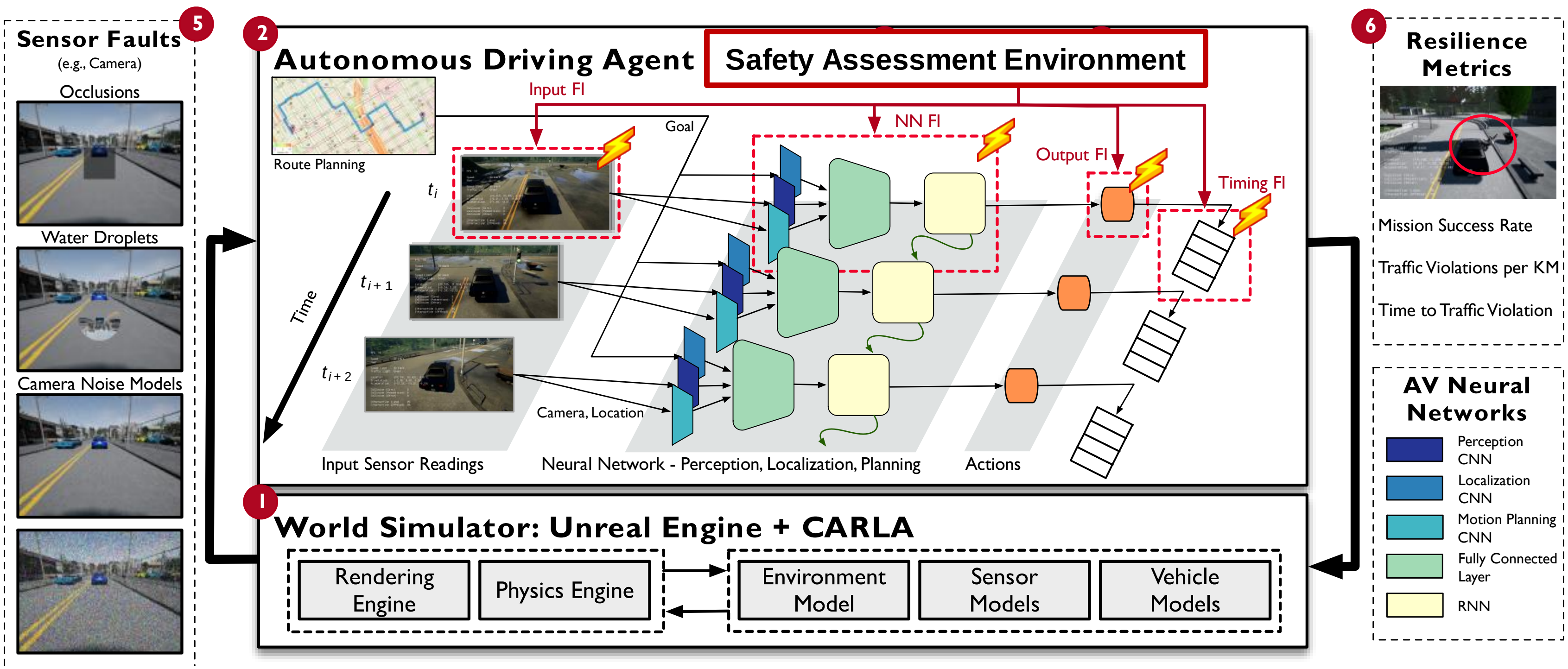
Baidu, "Apollo Open Platform," http://apollo.auto, Accessed: 2018-09-02.

Hands-off the Wheel?



Source: https://www.youtube.com/watch?v=spw176TZ7-8

AV Safety, Reliability, and Dependability Analytics



Intro to Autonomous Vehicles (AVs)

- Autonomous vehicles (AVs) are complex systems that use artificial intelligence (AI) and machine learning (ML) to integrate mechanical, electronic, and computing technologies to make real-time driving decisions.

- AI enables AVs to navigate through complex environments while maintaining a safety envelope that is continuously measured and quantified by onboard sensors (e.g., camera, LiDAR, RADAR).

- Clearly, the safety, resilience, and dependability of AVs are of a significant concern.

Mini Project 1: AV Safety, Reliability, and Dependability Analytics

- Recent media attention on Tesla/Waymo/Uber AVs

- Resilience and Safety characteristics vary across computing kernels and computing systems

- **Methods and tools for Assessing End-to-End Resilience of AV Technology is not available**

TRANSPORTATION UBER RIDE-SHARING

Uber self-driving car saw pedestrian but didn't brake before fatal crash, feds say

The report is more interesting for what it doesn't say than what it does

By Andrew J. Hawkins | @andyjayhawk | May 24, 2018, 11:07am EDT

Safety and Reliability Issues [Banerjee et al., DSN 2018]

- **Data and Machine Learning:** 64% of reports were the result of problems in, or untimely decisions made by, the machine learning system

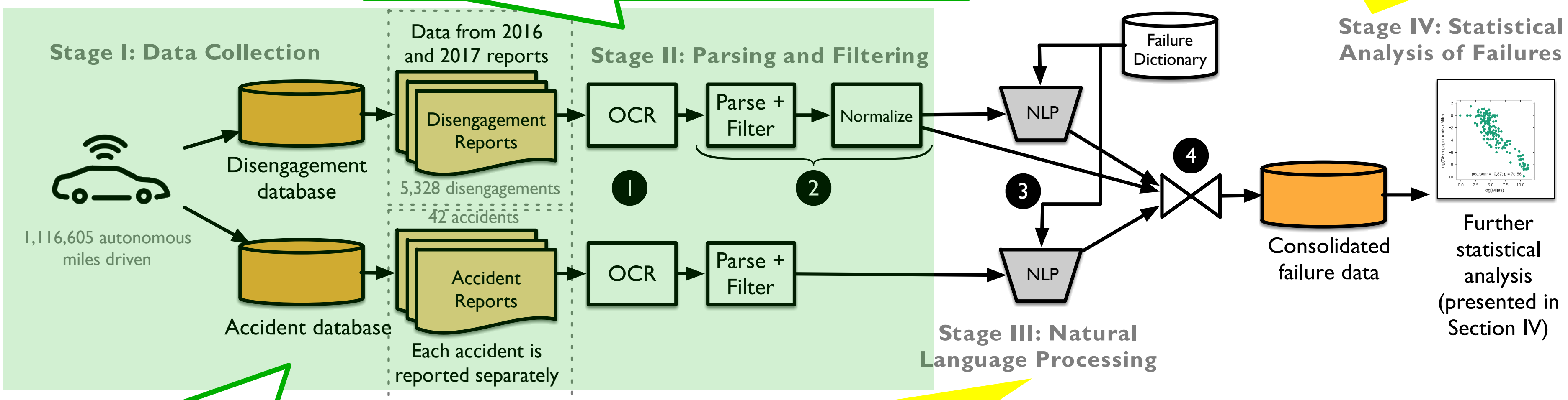
- **Compute system-related**: 30% or more due to failures in computing stack

- **System Design and Integration:** Mismatch between real-world driving behavior, machine learning methods, automotive engineering and compute systems

An End-to-End Workflow for AV Log Data Analysis

- Reports stored as scanned documents.
- Vendor specific parsing & filtering.
- Standardizing data formats across vendors.

- Analyze failure data to quantify
  - Causes
  - Dynamics
  - Impacts



- Vendors are required to collect data as per CA laws.
- CA DMV curates databases of vendor reports.
- No standardized reporting formats.

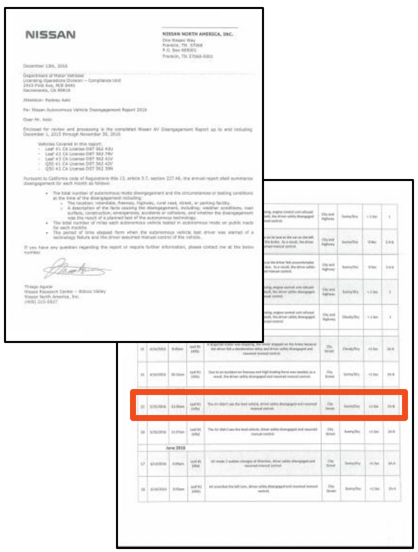
- Localize failures in abstract system model.

Nissan Case Study

❷ **Individual Report**

15	5/25/2016	11:20am	Leaf #1 (Alfa)	The AV didn't see the lead vehicle, driver safely disengaged and resumed manual control.	City Street	Sunny/Dry	<1 Sec	2A-B

❸ **OCR + Parsing + Cleaning**



The AV **didn't see** the lead vehicle…

Analysis shows that AVs do worse than human drivers

Categories: Recognition

❶ **Nissan Disengagement Reports from the CA DMV**

AV Dependability in Adversarial ML

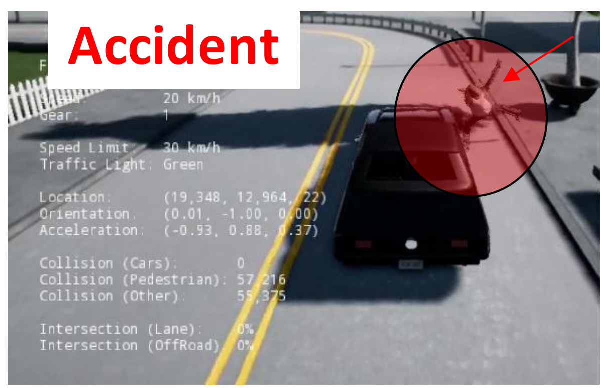
[1]

original fog original rain original translation(40,40) original scale(2.5x)

Weather Conditions

original shear(0.1) original rotation(6 degree) original contrast(1.8) original brightness(50)

Camera and Lighting Conditions



Effect of fault injection

[1] Eykholt, Kevin, et al. "Robust Physical-World Attacks on Deep Learning Visual Classification." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[2] Jha, Saurabh, et al. "AVFI: Fault injection for autonomous vehicles." *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE, 2018.

Example Applications we will come Across

- **AVs**
- **Safety analytics**
- **Genomic computing**
- **Health monitoring**
- **Software design**
- **Heterogenous computing**
- **Security monitoring ……**

Homework 0

- Questions on application of probability in the computing
- Questions cover
  - Basic Probability
  - Conditional Probability – Bayes Formula
  - Probability Distributions Discrete, Continuous
    - Bernoulli's, binomial, Poisson, Exponential Normal …
    - Joint Distribution
  - Variance/Covariance/Mean/Expectation
  - Law of large numbers/Central Limit Theorem
- Be concise and provide all steps
- **Release Date : Monday Jan 27, 2020**
- **Due Date: Monday Feb 3, 2020, 23:59 hrs**
- Students should submit their solutions on Compass (you are encouraged to type your solutions).

Course Outline I

Week 1	Jan 20	Martin Luther King, Jr. Day – No Lecture
	Jan 22	Lecture 1: Course outline and Overview of Mini Projects: a. Autonomous Vehicle (AV) Safety Analytics b. Healthcare analytics c. Security analytics Overview of key data analytics and ML concepts.
Week 2	Jan 27	Lecture 2: Probability Basics Overview, P values, Hypothesis Testing, fitting distributions KS tests KL divergence
	Jan 29	Lecture 3: Python; Jupyter notebook
Week 3	Feb 3	Lecture 4: In-class Activity 1 (Submit in class on Mon Feb 3)
	Feb 5	Lecture 5: Naïve Bayes, Conditional probability, conditional independence
Week 4	Feb 10	Lecture 6: Naïve Bayes, and Bayesian Networks (Continued)
	Feb 12	Lecture 7: In-class Activity 2 on Bayesian Networks. Expectation Maximization. Final project proposals due.
Week 5	Feb 17	Lecture 8: Clustering GMM continued; K-means Clustering
	Feb 19	Lecture 9: Hierarchical Clustering, Regression
Week 6	Feb 24	Lecture 10: Hierarchical Clustering and Linear Regression examples Preparation for Mini-project 2: Introduction to Health-care Domain: Disease Models, Drug Response, Forecasting Disease Progression Mini-project 1 Presentations (out of class)
	Feb 26	Lecture 11: Guest Lecture by a Mayo clinician from Center for Individualized Medicine. Principal Component Analysis (PCA)
Week 7	Mar 2	Lecture 12: In-class Activity 3 on PCA and Clustering
	Mar 4	Lecture 13: Introduction to Probabilistic Graphical Models; revisit Bayesian Networks Naïve Bayes Introduction and Examples
Week 8	Mar 9	Lecture 14: Practice Midterm
	Mar 11	Lecture 15: Midterm

Course Outline II

Week 9	Mar 16	Spring Break – No Lecture
	Mar 18	Spring Break – No Lecture
Week 10	Mar 23	Lecture 16: Midterm Discussion. Markov Models and Hidden Markov Models (HMMs)
	Mar 25	Lecture 17: HMM continued. ICA4
Week 11	Mar 30	Lecture 18: Factor Graphs
	Apr 1	Lecture 19: Intro to Security Domain and Mini-project 3. In-class Lab 3 on Mini-project 3
Week 12	Apr 6	Lecture 20: Factor Graphs Continued + Belief Propagation
	Apr 8	Lecture 21: Belief Propagation continued
Week 13	Apr 13	Lecture 22: In-class Activity 5 on PGMs
	Apr 15	Lecture 23: Supervised Learning (SVM, Neural Nets)
Week 14	Apr 20	Lecture 24: Intro to Deep Learning
	Apr 22	Lecture 25: Decision Trees, Random Forest, Cross Validation
Week 15	Apr 27	Lecture 26: In-class Activity 6 (tentative) on Neural Network + SVM
	Apr 29	Lecture 27: Solved Examples
Week 16	May 4	Lecture 28: Practice Exam for Finals
	May 6	Reading Day
Week17	May 8-15	Final Examination Period. As per the exam calendar