CS-588: Autonomous Vehicle System Engineering

D.A. Forsyth, UIUC: Instructor
Rahul Vasanth, UIUC: TA

Class and constraints

- Primary goal:
  - project groups build and execute applications controlling a real AGV
  - all learn some technologies for sensing, planning, mapping and control
    - with emphasis on visual and lidar sensing
- Key issues:
  - keeping safe from COVID
  - not getting hurt by the vehicle
  - you NEED to be registered
- Evaluation:
  - homeworks and final project

Class meetings

- We’ll meet in a “mixed” fashion
  - mostly 2200 Sidney Lu Mech Eng Blg
    - 12h30-13h45
  - some video
  - occasionally at
    - 201 St. Mary’s Road Champaign, IL 61820
    - (High Bay)
    - Google Maps will help!

Topics

- Mainly sensing, some control, some planning
  - for autonomous vehicles
- Vehicles
  - we have access to a Polaris GEM
    - from AutonomousStuff
    - and a simulator for this vehicle
- Structure (!?!)
  - mainly lecture
  - some paper presentations
  - some project practicals

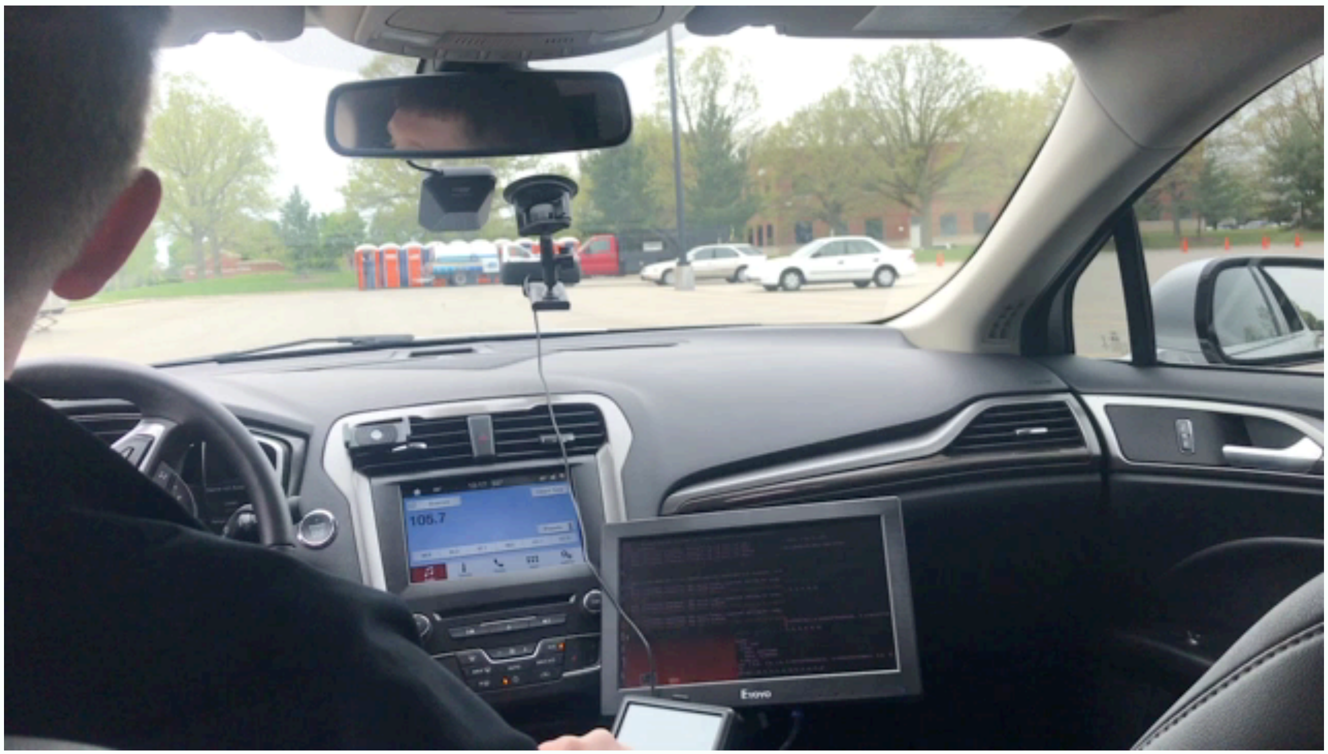
Outline

- Suitable for:
  - CS/ECE Grad students; CS/ECE Undergrads (independent minded)
    - with some experience
      - (AML; Vision; Computational Photog.;etc)
    - independent minded
  - Others
    - won't get much support with minor programming problems
    - limited structure
    - great fun
- Generally, we'll go through a lot of material quite fast
  - and quite superficially

CS598 MAAV - is this for me?

- First homework should help with that!

CS 598 MAAV History



Features

- (We think) only autonomous vehicle courses in country with a physical vehicle
  - very big deal for CS students
    - often have no experience of dealing with physical objects
  - cover
    - detection, sensing, mapping, slam
    - some control
    - some path planning
    - big class projects on real vehicle
- No deaths or injuries to date
  - we're going to be fussy about safety!

Examples

Moving autonomously - MAAV course

- Autonomous car replans its path around a moving obstacle
  - https://www.youtube.com/watch?v=LBv49TwdY2o&feature=youtu.be


CLOSED
DO NOT ENTER

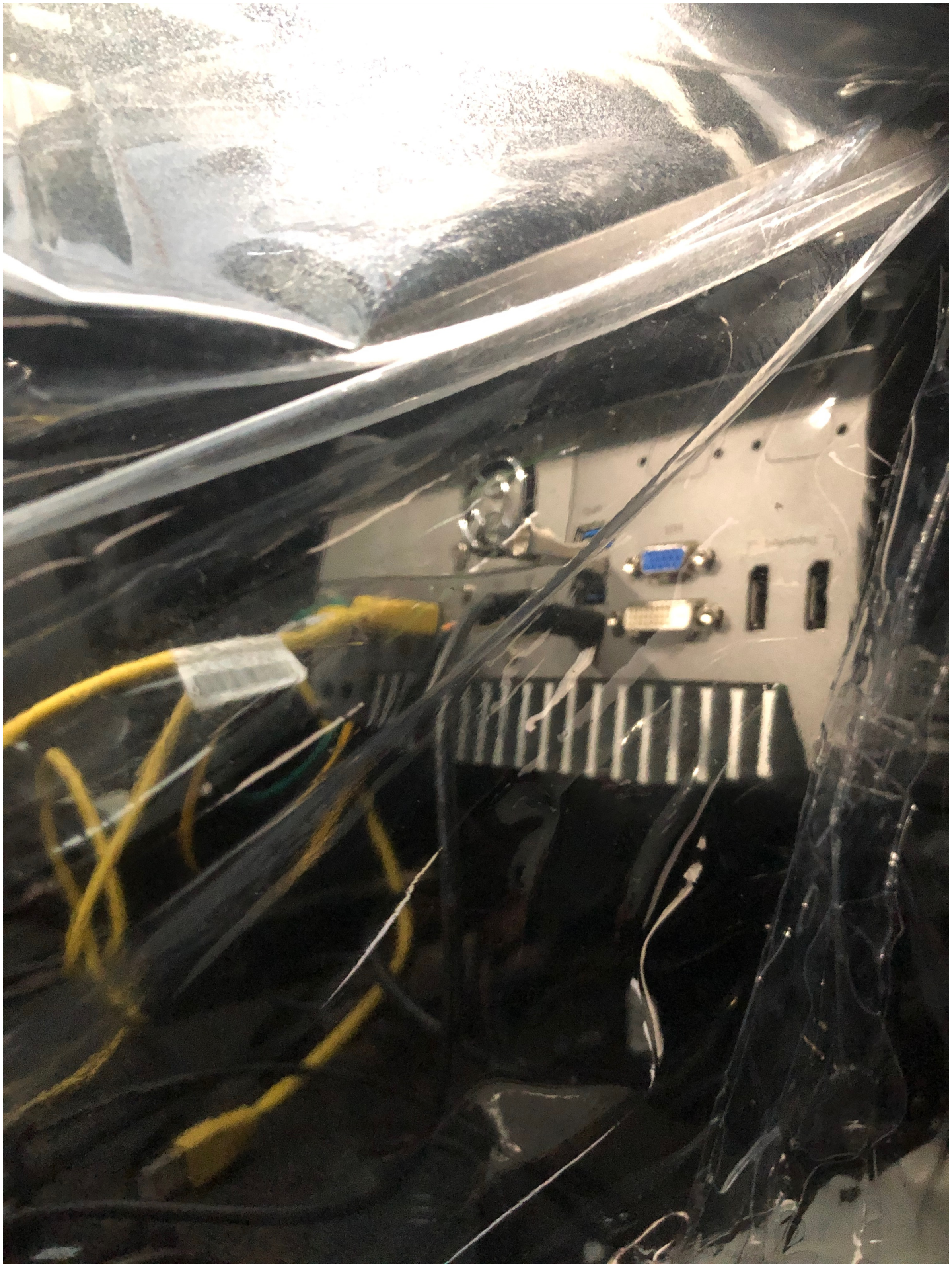
DO YOUR PART
DO NOT SIT HERE
6 FEET APART



STOP


VEHICLE
CABIN
DIVIDER
PROTOTYPE
DO NOT USE
WITH MORE
THAN ONE
PERSON YET
POWERED BY
GEM e2
AUTOMATED RESEARCH DEVELOPMENT PLATFORM

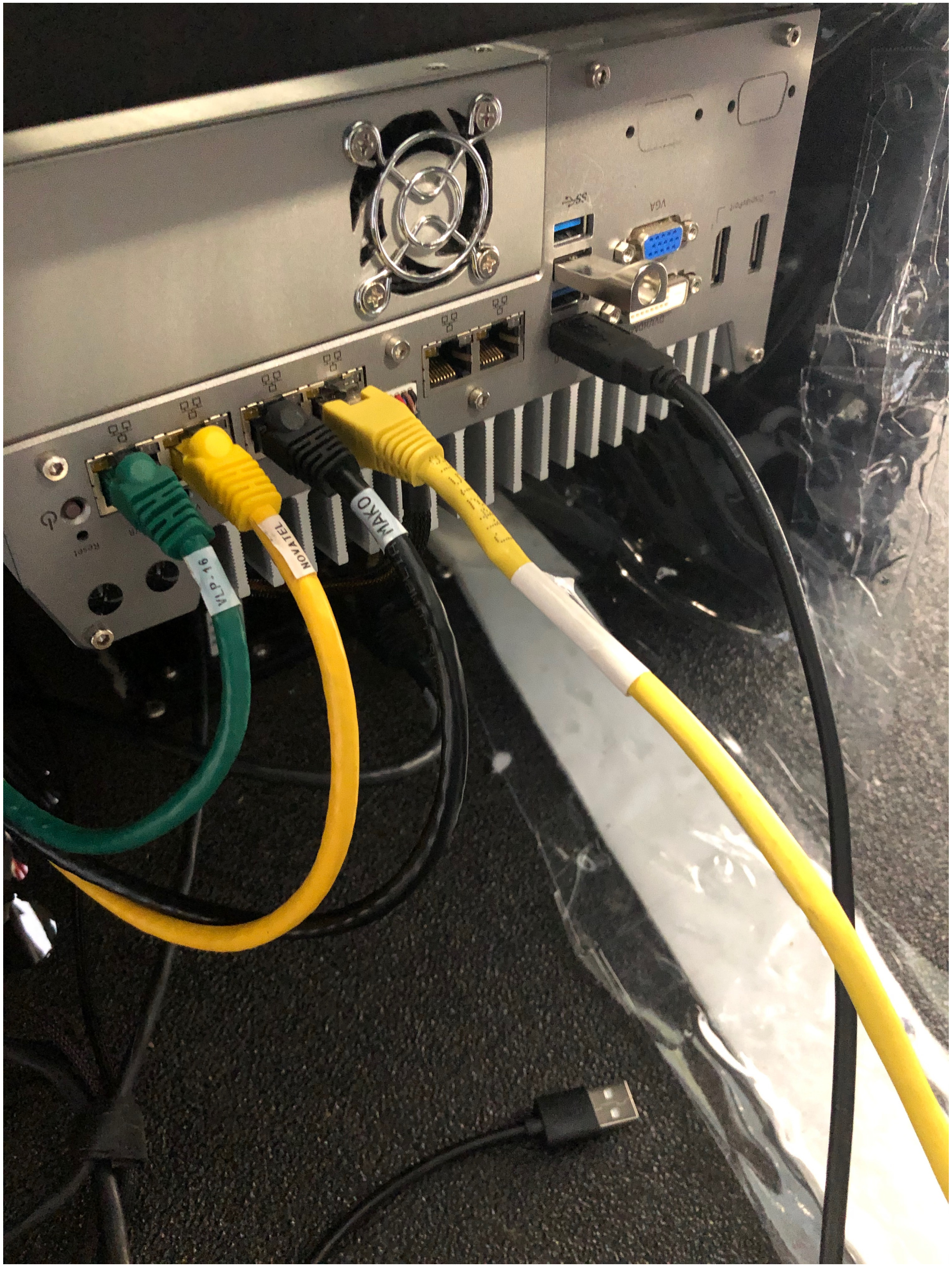

VEHICLE
CABIN
DIVIDER
PROTOTYPE
DO NOT USE
WITH MORE
THAN ONE
PERSON YET.
JOHN




JUL California 2020
7XBM545
AUTONOMOUSTUFF
GEM
POLARIS


58


EMERGENCY
POWERED BY
BY-WIRE
FEATURING PACMod



VGA
VLP-16
NOVATEL
MAKO

Vehicle dangers

- Heavy, but not particularly fast
  - it will hurt you if it hits you
  - ALWAYS a safety driver when moving
- It will hurt a lot if it goes over your foot!
  - this is the most likely accident
  - always use chocks on vehicle
    - either safety driver or chocks should be in

WEAR CLOSED SHOES!

- Do not sit on the back
  - when it breaks, your rear end will land on high output battery connectors
- High center of gravity
  - turning while moving fast is scary
- Infection!

Safety roles

- Safety lookout
  - everyone - we'll train collectively
  - responsible for
    - watching when vehicle moves;
    - ensuring chocks are in;
    - general safety
- Safety driver
  - responsible for ensuring vehicle is safe when moving
    - when chocks are out, safety driver is in
  - some, ideally one per group
  - Rahul/I will train individually


VEHICLE CABIN DIVIDER PROTOTYPE DO NOT USE WITH MORE THAN ONE PERSON YET
POWERED BY
GEM e2
AUTOMATED RESEARCH DEVELOPMENT PLATFORM

The Chocks

Vehicle dangers

Vehicle dangers

Vehicle safety practices

- If the chocks are out, a safety driver MUST be in
- Roles
    - Safety driver
        - (inside vehicle) stop the vehicle from moving into wall, person, etc.
        - MAY NOT DO ANYTHING ELSE
    - Lookout
        - (outside vehicle) watch to ensure others are not in danger, esp feet!
    - Experimenter
        - (inside vehicle) complain at software, etc.

OLD Infection safety practices

- We will break up into project groups
  - I'd like about 5-7
- Only one person per group has access to the vehicle
  - the person could change from day to day, but only one per group at any time
- No more than four people in Highbay at any time
  - one of them is DAF
  - others are in stalls, one per stall
    - there is wifi access
- You can't get in to Highbay without DAF letting you in
  - and you have to show me evidence of recent clean test
- The vehicle has a divider
  - so two people can be in vehicle, but ideally doesn't happen often

This week…

- Friday is a physical visit to High Bay:
  - Safety lookout training
    - I'll need email confirmation from each
      - you should send Rahul email, header: Safety Lookout Confirmation
        - confirming that:
          - you've done the online Lab Safety course
          - you attended safety lookout training in person
            - OR you viewed the safety lookout video.
- You should
  - get vaccinated OR get and keep tested if you want vehicle access
  - follow lectures and written material
  - be trained as a safety lookout
- If you want to be trained as a safety driver
  - Requirements: over 21; US drivers license; safety lookout
  - Contact Rahul with 384 2864 in email header

Next week

- Basic Lecture material on ROS, Neural networks
  - Complete simple exercise by end week 4
  - Safety driver training for some by Rahul

What problems should an AV solve?

- How can I reach my goal safely and efficiently?
  - adjusting goal as required:
  - What action should I use next?
    - How will that action affect goal/safety?
- Possible subsidiary questions
  - Where am I?
  - What is going on around me?
  - What will the likely effects of an action be?
  - What would an expert do? (WWED!)
- How should safety and efficiency trade off?

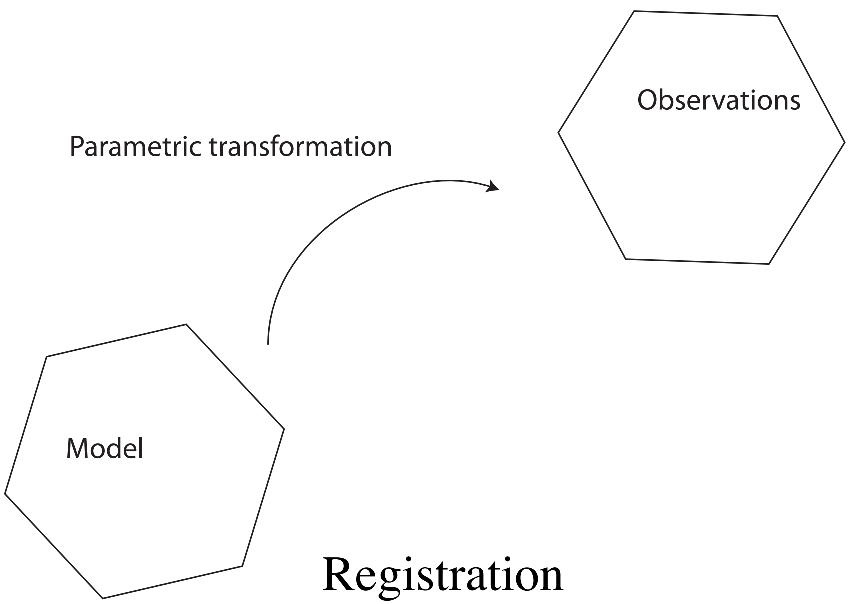
Sensing

- Visual
  - camera/cameras
  - I'll emphasize - cheap sensing, complicated interpretation
- Lidar
- Radar
- Vehicle sensors
  - speed, steering angle, number of wheel revolutions, etc.

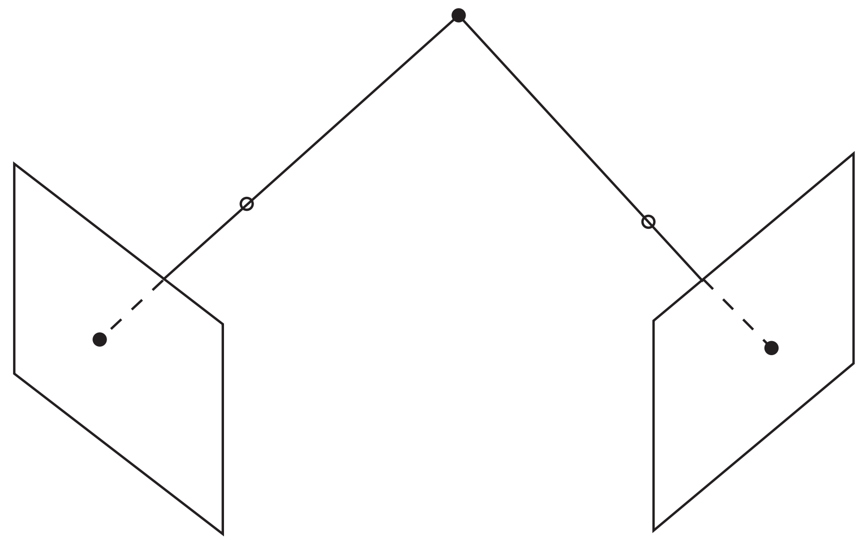
Vision

- Exploded into giant, high-value discipline over last 20 yrs
- Key Tools
  - Geometry
  - Discrimination
  - Regression

Big ideas in vision: Geometry

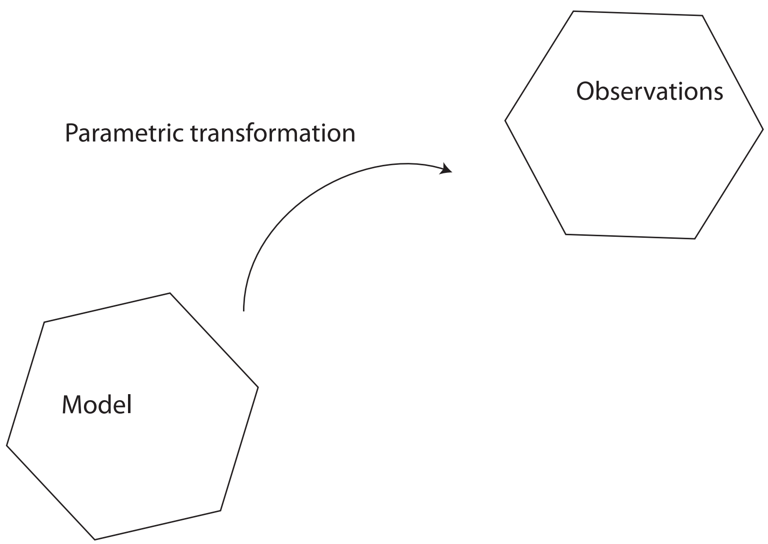


Registration



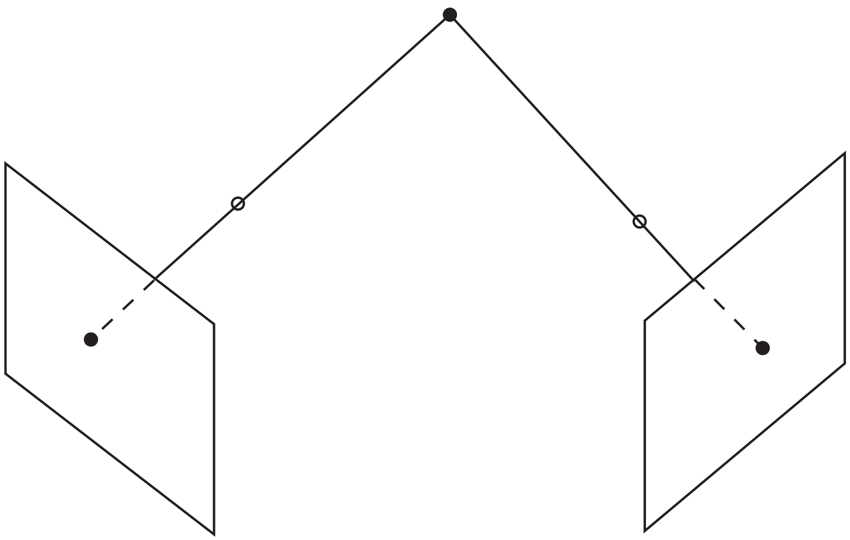
Multiple view geometry

Registration



- Extremely general recipe
  - 2D to 2D; 3D to 3D; 3D to 2D; etc
  - Representations
    - as points
    - as primitives
    - etc
- Fantastically useful

Multiple view geometry

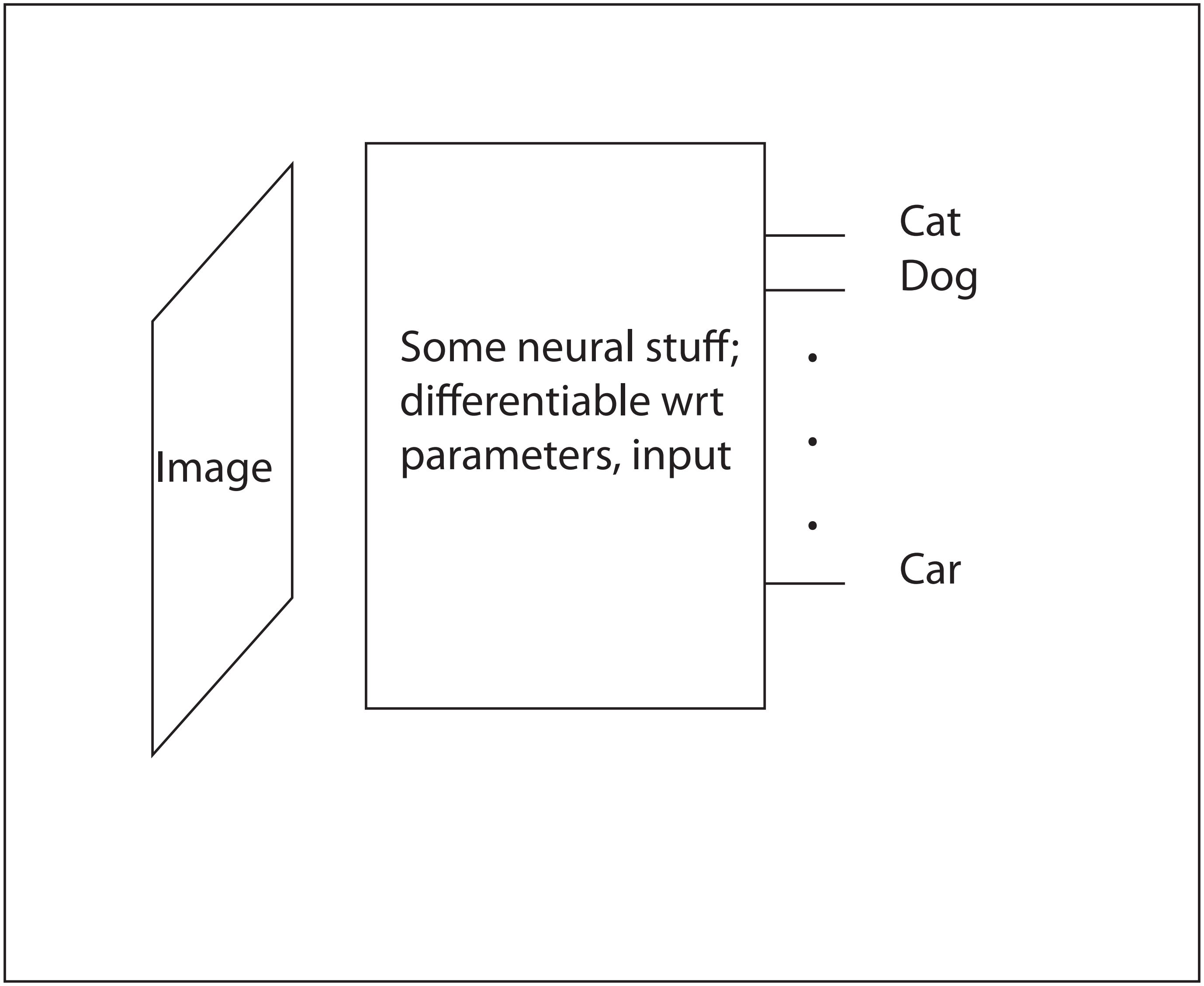


- Q: What can we impute from
  - 2 or more views
- A: Lots
  - Motion of the camera
  - How things are moving in the image
  - 3D structure

Big ideas in vision: Discrimination

- Use some procedure to attach a label to
  - an image; some images; video; range data; lidar data; etc, etc
- “Label” can be very loosely interpreted
- “Procedure” could be
  - learned
  - hand-tuned
  - determined by physics; the problem; etc

Image classification



Key ideas

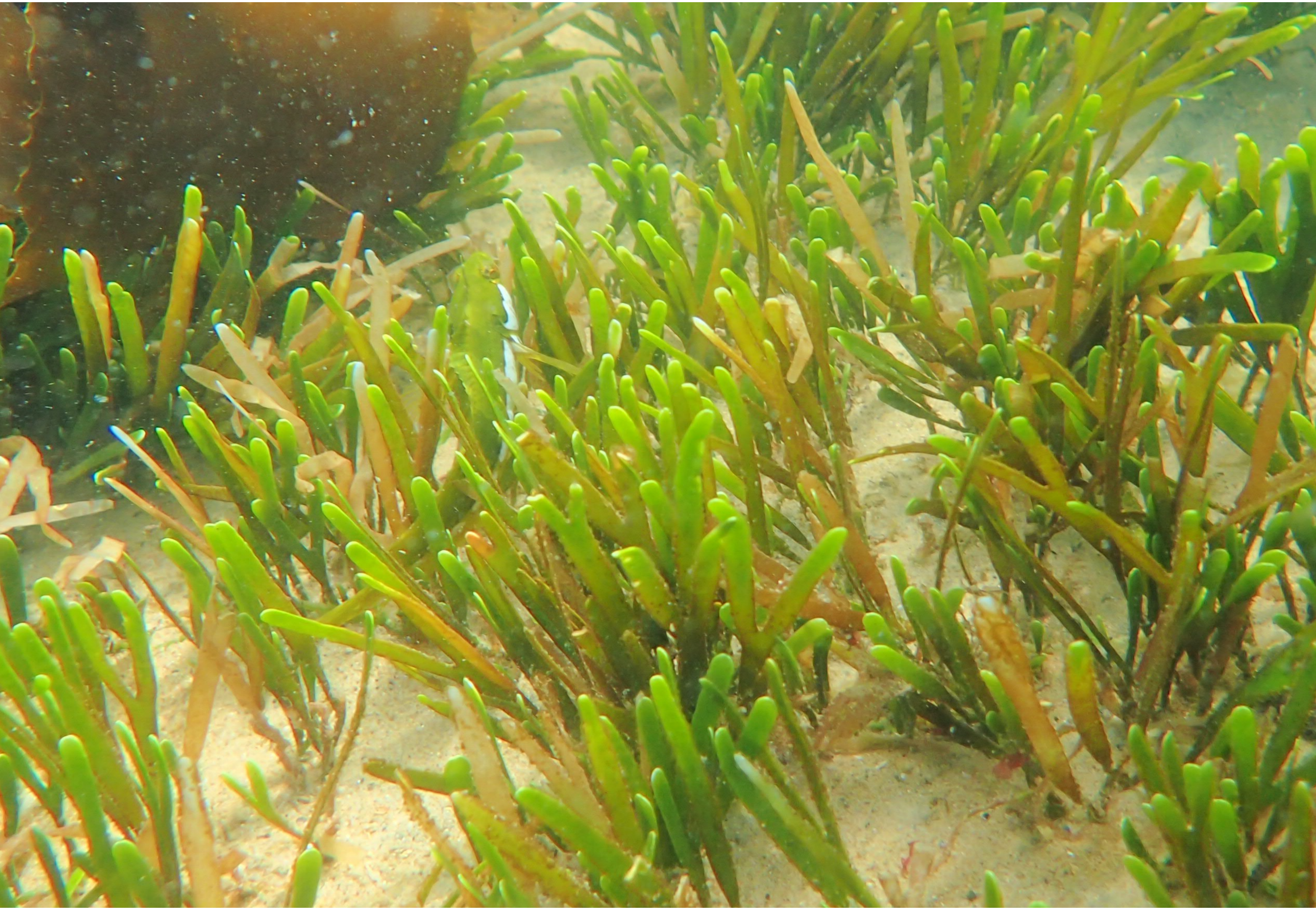
- Goal:
  - Adjust classifier so that it accurately classifies *UNSEEN* data
- Procedure:
  - Adjust so that it
    - classifies training data well
    - generalizes
      - regularization term, either explicit or implicit
- Evaluation:
  - Use held out data to check accuracy on *UNSEEN* data

Classification or detection

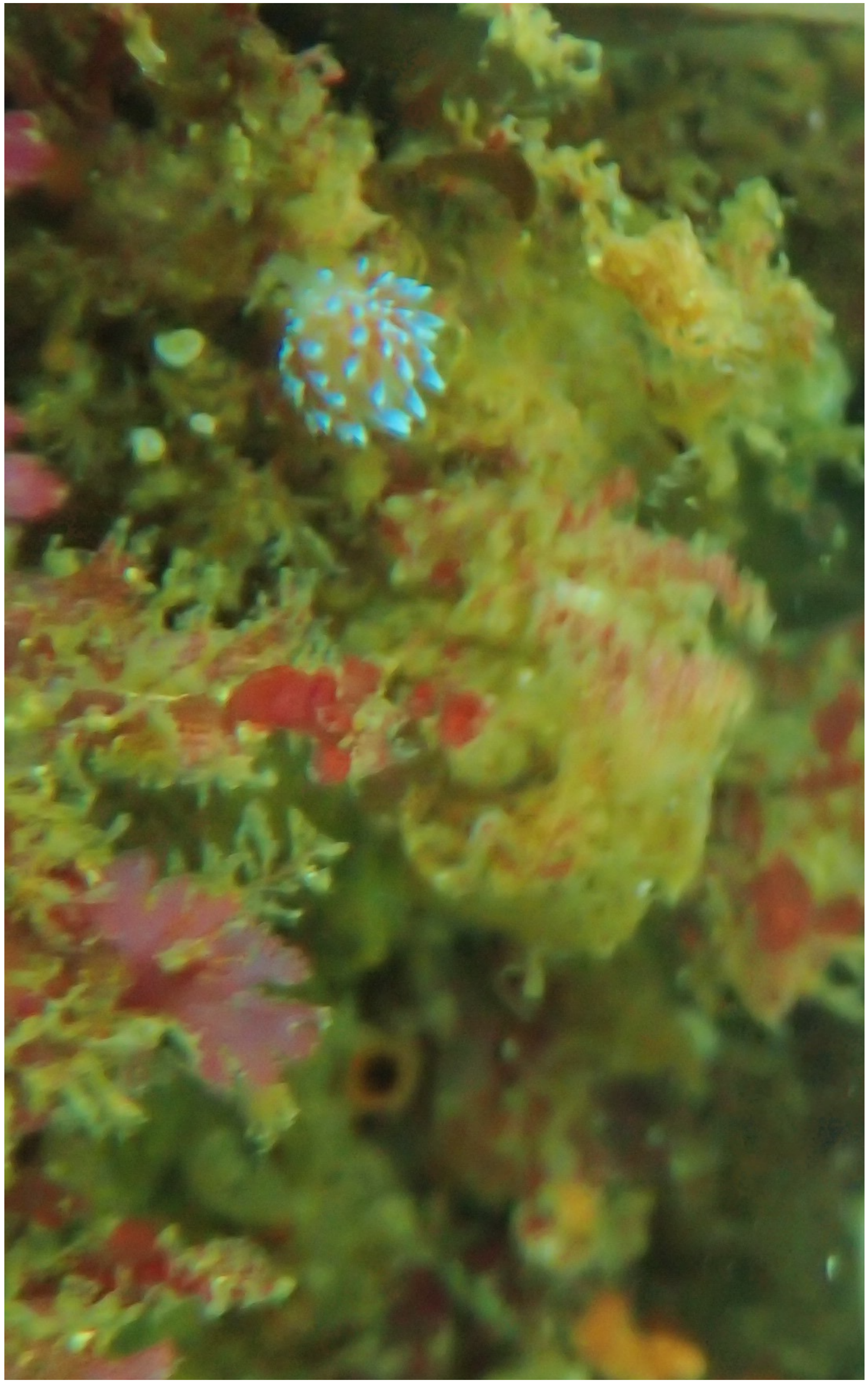
- Classification:
  - there is an X in this image
    - what
- Detection:
  - there is an X HERE in this image
    - what AND where

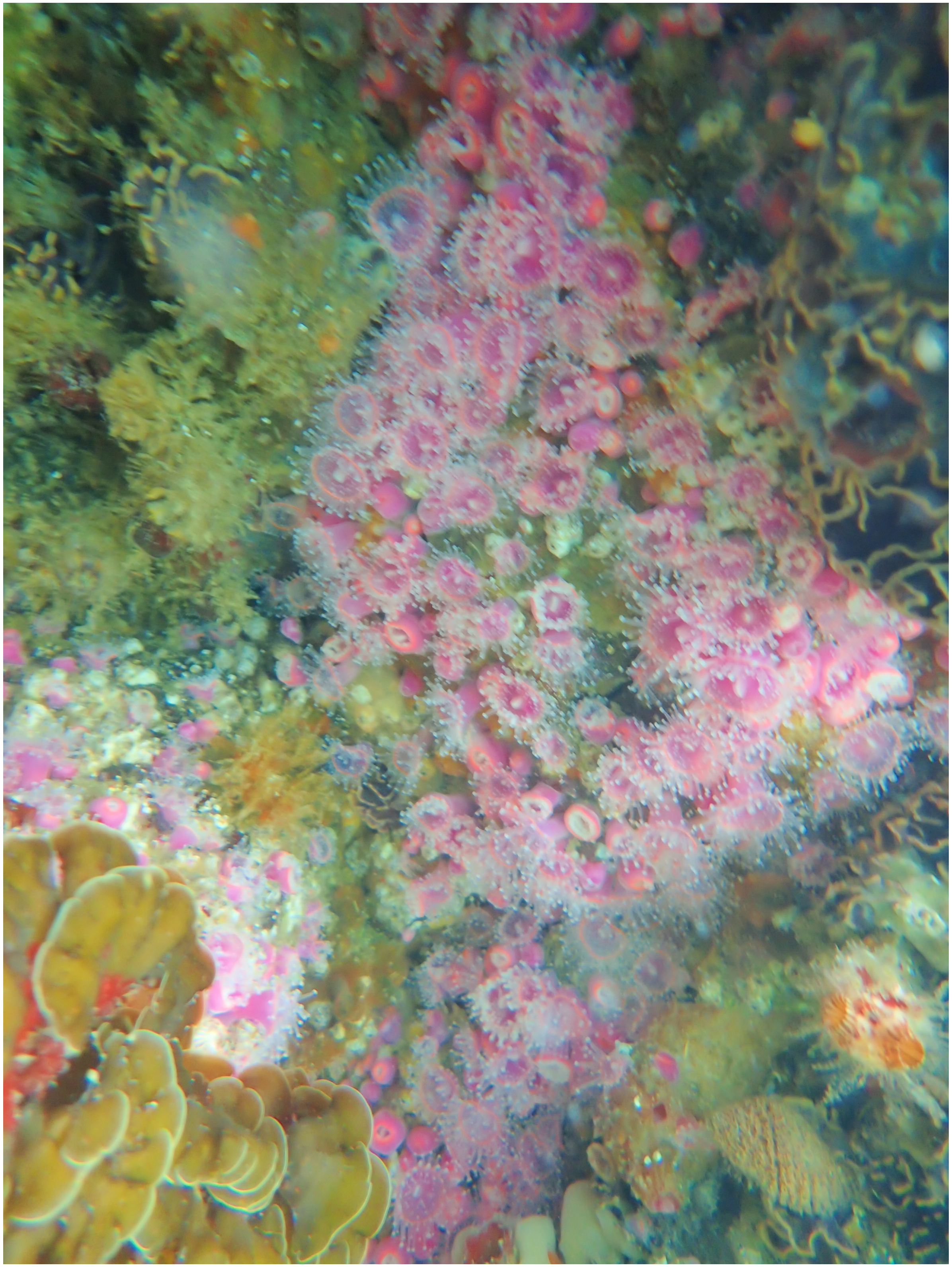
- Key issues
  - how to specify where
  - relationship between what and where
    - efficiency, etc
  - evaluation
    - surprisingly fiddly

It can be hard to find things



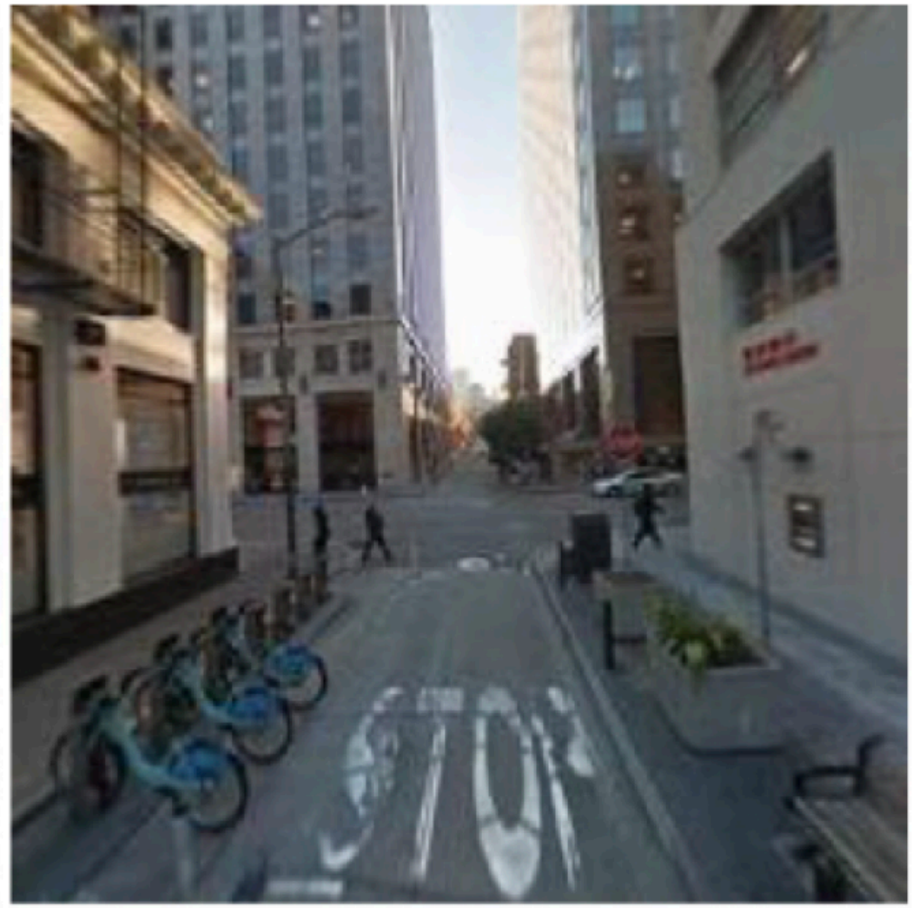
You may not know the right label





Exploiting registration and classification

- Use a classifier to tell:
  - how far to the next intersection?
  - what is it like?
  - is there a bike lane?
  - etc.



Pred = 18.5 m

Road layout maps

- Potential cues
  - streetview
  - openmaps
  -

Partially supervised cues

- Open Street Maps (OSM)

Map data: OpenStreetMap is an open-source mapping project covering over 21 million miles of road. Unlike proprietary maps, the underlying road coordinates and metadata are freely available for download. Accuracy and overlap with Google Maps is very high, though some inevitable noise is present as information is contributed by individual volunteers or automatically extracted from users' GPS trajectories. For example, roads in smaller cities may lack detailed annotations (e.g., the number of lanes may be unmarked). These inconsistencies result in varying-sized subsets of the data being applicable for different attributes.

Seff+Xiao

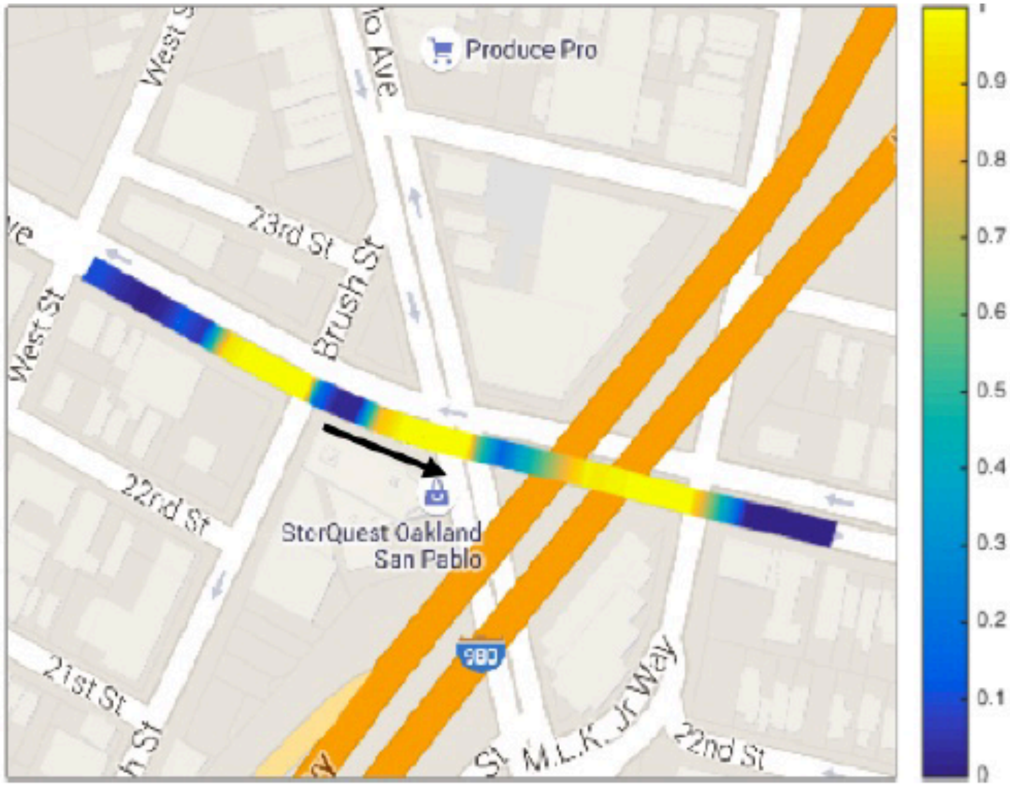


Fig. 3. Intersection detection heatmap. Images are cropped from test set GSV panoramas in the direction of travel indicated by the black arrow. The probabilities of "approaching" an intersection output by the trained ConvNet are overlaid on the road. (The images are from the ground level road, not the bridge.)

Seff+Xiao

Partially supervised cues

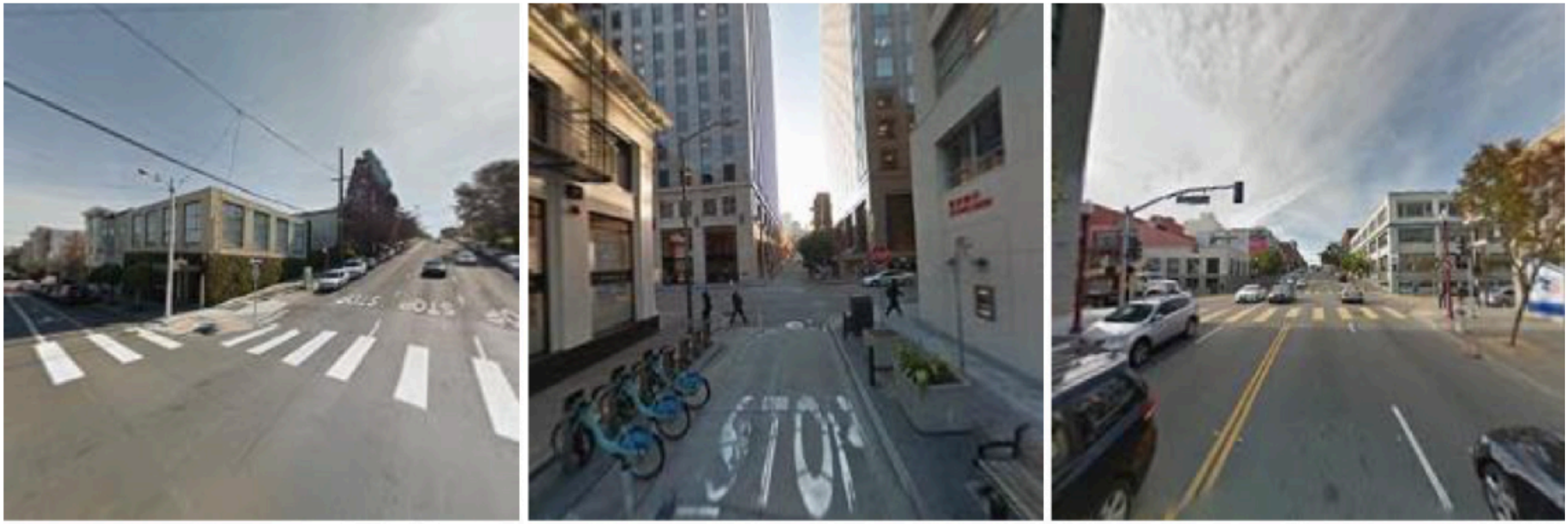
- Google street view

Image collection: Google Street View contains panoramic images of street scenes covering 5 million miles of road across 3,000 cities. Each panorama has a corresponding metadata file storing the panorama's unique "pano_id", geographic location, azimuth orientation, and the pano_ids of adjacent panoramas. Beginning from an initial seed panorama, we collect street view images by running a bread-first search, downloading each image and its associated metadata along the way. Thus far, our dataset contains one million GSV panoramas from the San Francisco Bay Area. GSV panoramas can be downloaded at several different resolutions (marked as "zoom levels"). Finding the higher zoom levels unnecessary for our purposes, we elected to download at a zoom level of 1, where each panorama has a size of 832×416 pixels.

Seff+Xiao

Labelling - I

- Match panoramas to roads
  - panorama center location, orientation is known
  - (essentially) project to plane
  - thresholded nearest neighbor to road center polyline
    - thresholding removes panoramas inside buildings, etc.
  - some noise
    - under bridges, etc.
- Annotations
  - Intersections
  - Drivable heading
  - Heading angle
  - Bike lane
  - Speed limit, wrong way, etc.



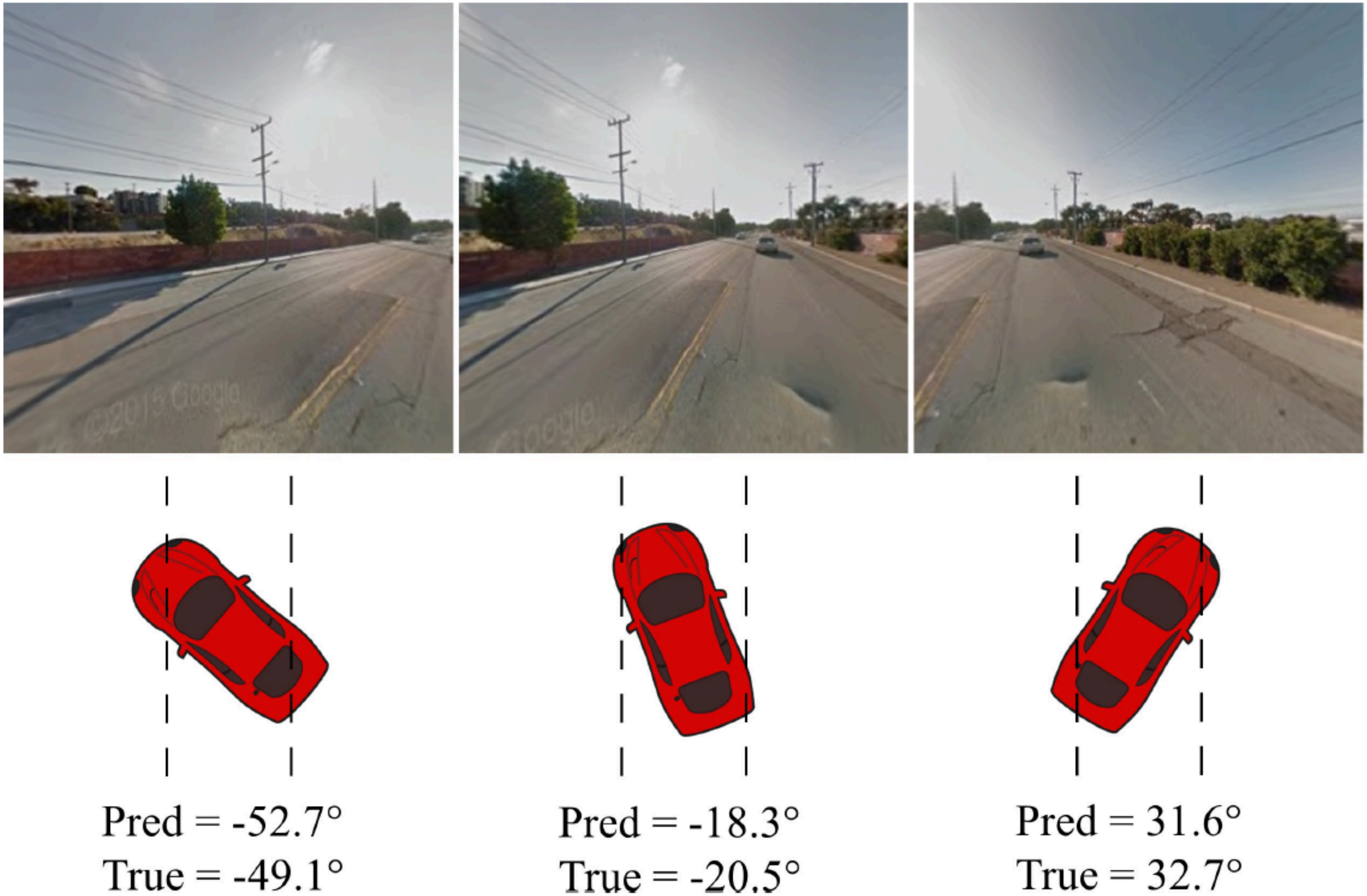
Pred = 0.1 m
True = 1.9 m

Pred = 18.5 m
True = 19.2 m

Pred = 22.9 m
True = 22.4 m

Fig. 4. Distance to intersection estimation. For images within 30 m of true intersections, our model is trained to estimate the distance from the host car to the center of the intersection across a variety of road types.

Seff+Xiao



Seff+Xiao

Fig. 7. Heading angle regression. The network learns to predict the relative angle between the street and host vehicle heading given a single image cropped from a GSV panorama. Below each GSV image, the graphic visualizes the ground truth heading angle.

Pred = 26.1 mph
True = 30 mph

Pred = 30.0 mph
True = 50 mph

Pred = 54.3 mph
True = 50 mph

Fig. 9. Speed limit regression. The network learns to predict speed limits given a GSV image of road scene. The model significantly underestimates the speed limit in the middle example as this type of two-way road with a single lane in each direction would generally not have a speed limit as high as 50 mph.

Seff+Xiao

p(one-way) = 0.207 p(one-way) = 0.226 p(one-way) = 0.848

Fig. 10. One-way vs. two-way road classification. The probability output by the ConvNet of each GSV scene being on a one-way road is shown. From left to right the ground truth labels are two-way, two-way, and one-way. The image on the left is correctly classified as two-way despite the absence of the signature double yellow lines.

Seff+Xiao

p(wrong way) = 0.555 p(wrong way) = 0.042 p(wrong way) = 0.729

Fig. 11. Wrong way detection. The probability output by the ConvNet of each GSV image facing the wrong way on the road is displayed. From left to right the ground truth labels are wrong way, right way, and right way. For two-way roads with no lane markings (left), this is an especially difficult problem as it amounts to estimating the horizontal position of the host car. The problem can also be quite ill-defined if there are no context clues as is the case with the rightmost image.

Seff+Xiao

Pred = 2 | Pred = 2 | Pred = 3
True = 1 | True = 2 | True = 2

Fig. 12. Number of lanes estimation. The predicted and true number of lanes for three roads are displayed along with the corresponding GSV images. For streets without clearly visible lane markings (left), this is especially challenging. Although the ground truth for the rightmost image is two lanes, there is a third lane that merges just ahead.

Seff+Xiao

At this point

- I can tell from an image whether
  - I’m pointing in the right direction
  - going the right way
  - facing an intersection
  - available turns, etc.
  - what and where street signs are
  - …
- Can I build a reliable controller?

Controller strategy

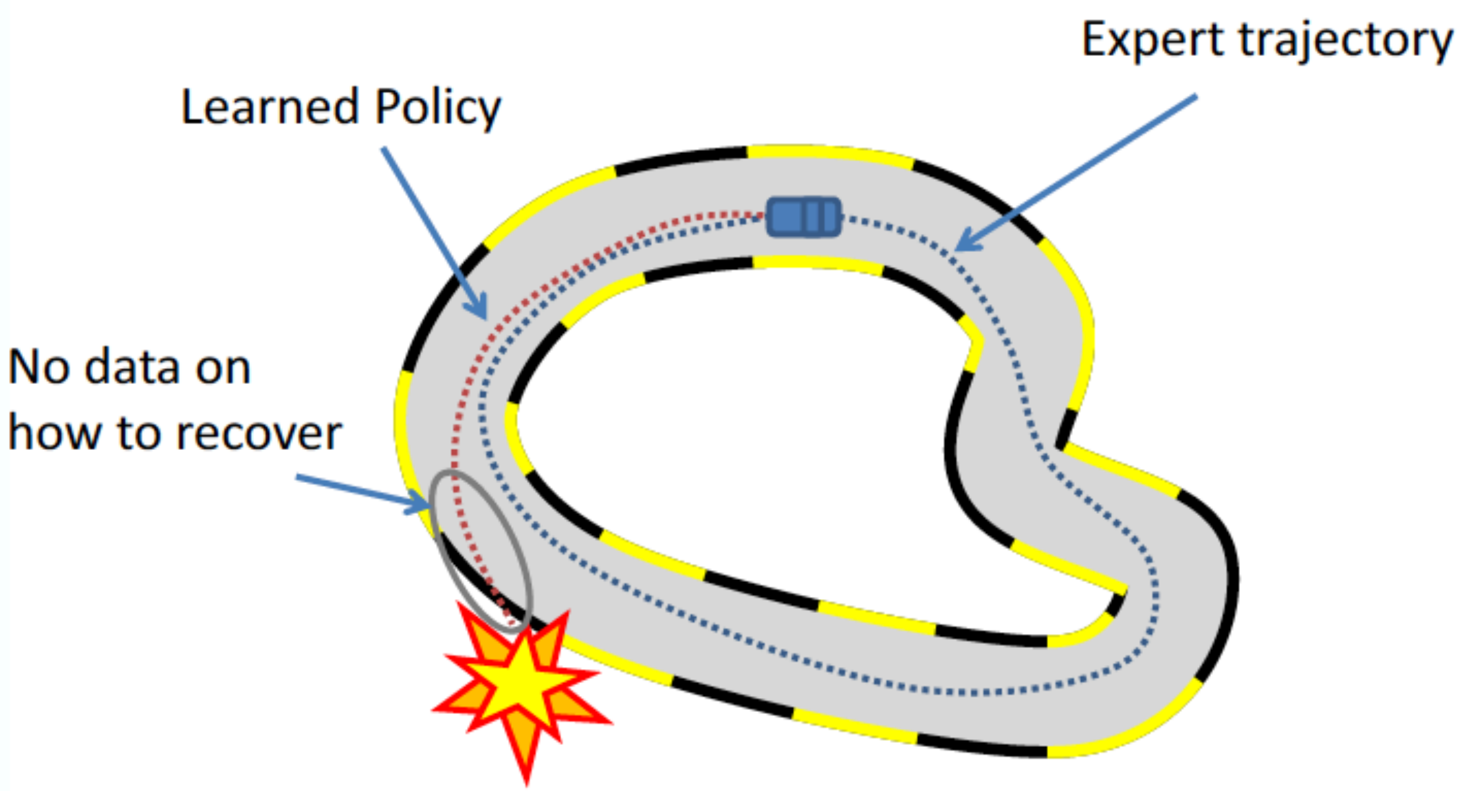
- Watch expert driver
  - at each tick, record
    - location description (speeding signs, dist to intersection, etc.)
    - what expert did
- Train controller for speed, direction
  - accept description
  - produce experts behavior

BIG GOOD QUESTIONS

- Q: WHY IS THIS NOT DRIVING AROUND NOW?
  - A: (pretty obviously) because it doesn’t work
- Q: WHY NOT?
  - A: interesting

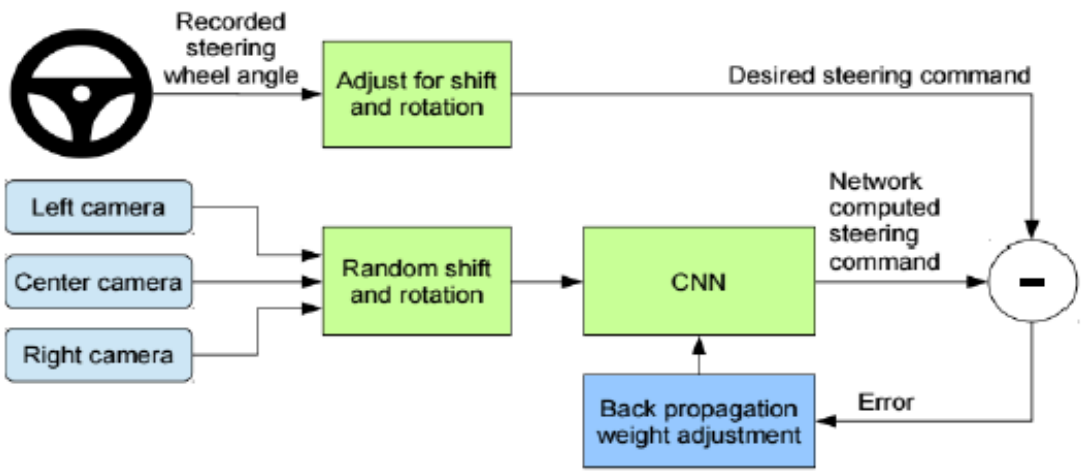
Data Distribution Mismatch!

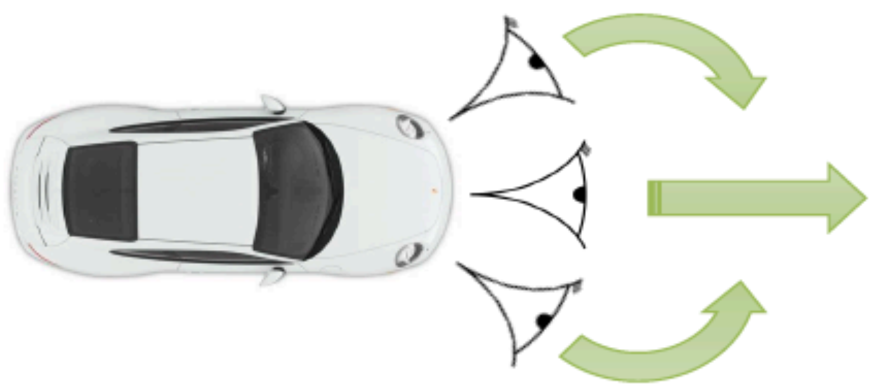
$$p_{\pi^*}(o_t) \neq p_{\pi_\theta}(o_t)$$



Fragkiadaki, ND

Demonstration Augmentation: NVIDIA 2016





Additional, left and right cameras with automatic grant-truth labels to recover from mistakes

"*DAVE-2 was inspired by the pioneering work of Pomerleau [6] who in 1989 built the Autonomous Land Vehicle in a Neural Network (ALVINN) system*. Training with data from only the human driver is not sufficient. The network must learn how to recover from mistakes. …",

End to End Learning for Self-Driving Cars , Bojarski et al. 2016

Fragkiadaki, ND

DAGGER (in simulation)

Dataset AGGregation: bring learner's and expert's trajectory distributions closer by labelling additional data points resulting from applying the current policy

1. train $\pi_\theta(u_t|o_t)$ from human data $\mathcal{D}_{\pi^*} = \{o_1, u_1, ..., o_N, u_N\}$
2. $\boxed{\text{run } \pi_\theta(u_t|o_t) \text{ to get dataset } \mathcal{D}_\pi = \{o_1, ..., o_M\}}$
3. $\boxed{\text{Ask human to label } \mathcal{D}_\pi \text{ with actions } u_t}$
4. Aggregate: $\mathcal{D}_{\pi^*} \leftarrow \mathcal{D}_{\pi^*} \cup \mathcal{D}_\pi$
5. GOTO step 1.

Problems:

- execute an unsafe/partially trained policy
- repeatedly query the expert

Fragkiadaki, ND A Reduction of Imitation Learning and Structured Predictioı

Big ideas in vision: Regression

Why regression?

- It's useful
  - Depth map from a single image
  - Ground map from aerial image
  - Modified/improved image from image
  - etc.

Key ideas

- Goal:
  - Adjust regressor so that it accurately predicts for *UNSEEN* input
- Procedure:
  - Adjust so that it
    - predicts for well for training data
    - generalizes
      - regularization term, either explicit or implicit
- Evaluation:
  - Use held out data to check accuracy on *UNSEEN* data

Key questions

- Generally: produce representations that improve control
  - by interpreting sensor output

- eg
  - Where am I?
    - registration and building simple maps
  - How have I moved?
    - visual odometry
    - (SLAM mentioned, but omitted as requiring filtering)