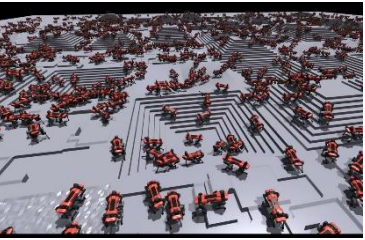
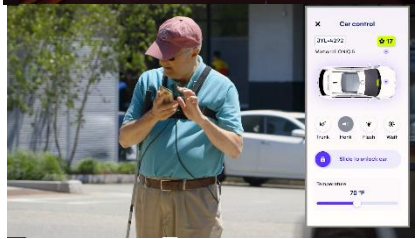
EC500: Robot Learning and Vision for Navigation



Eshed Ohn-Bar

Feb 13, 2023

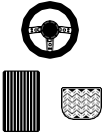


Reading questions on blackboard
Submit with HW1

Big Problem in Robotics (and Natural Agents)

Sensory Input

Action



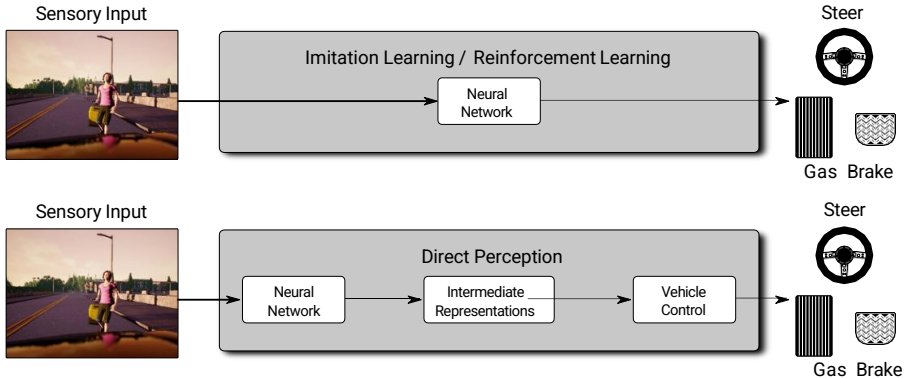
Big Problem in Robotics (and Natural Agents)

Sensory Input

Action

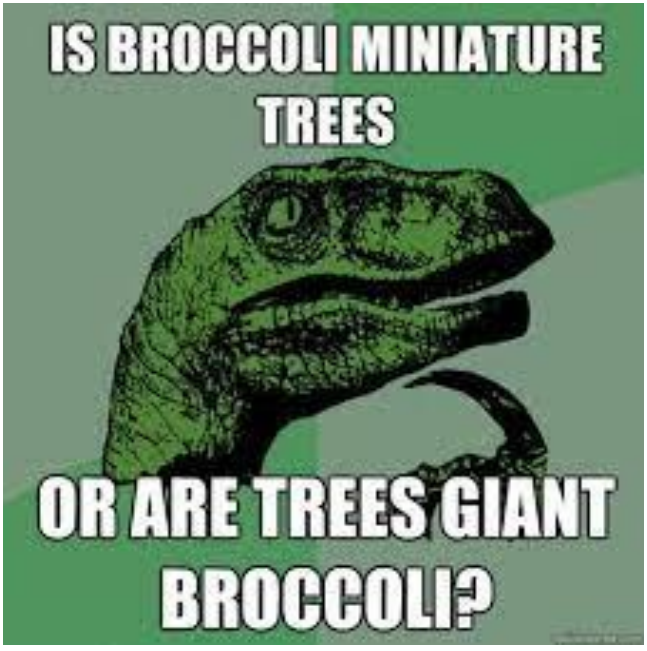
Can we create a representation for joint perception and action?

Direct Perception



Idea of Direct Perception:

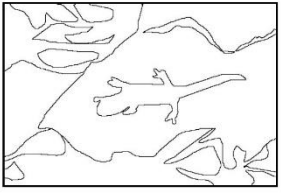
- Learn to predict interpretable low-dimensional intermediate representation
- Exploit classical controllers / finite state machines
- Hybrid model between imitation learning and modular pipelines


IS BROCCOLI MINIATURE TREES
OR ARE TREES GIANT BROCCOLI?

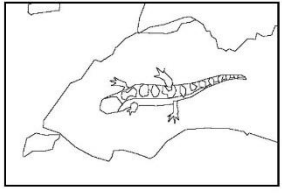
What is an object?

Berkeley Segmentation Dataset

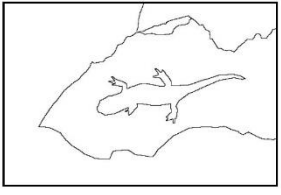
Original Image



Subject 1



Subject 2



Subject 3

“The meaning or value of a thing consists of what it ***affords***.”

JAMES J. GIBSON,
The Ecological Approach to Visual Perception

Affordance – perceived action possibilities, what environments can offer?

Gibson's Affordance (The Ecological Approach to Visual Perception)

The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb *to afford* is found in the dictionary, the noun *affordance* is not. I have made it up. I mean by it something that refers to **both the environment and the animal** in a way that no existing term does. It implies the complementarity of the animal and the environment.

- Action-Specific Perception
- Depends on what the observer ***wants and can*** do
- Action possibilities that are **directly (readily) perceivable** by an actor.

Gibson's Affordance (The Ecological Approach to Visual Perception)

- "... to perceive an affordance is not to classify an object." [4, p. 134].
- "... If you know what can be done with a graspable object, what it can be used for, you can call it whatever you please. ... The theory of affordances rescues us from the philosophical muddle of assuming fixed classes of objects, each defined by its common features and then given a name. ... But this does not mean you cannot learn how to use things and perceive their uses. You do not have to classify and label things in order to perceive what they afford." [4, p. 134].

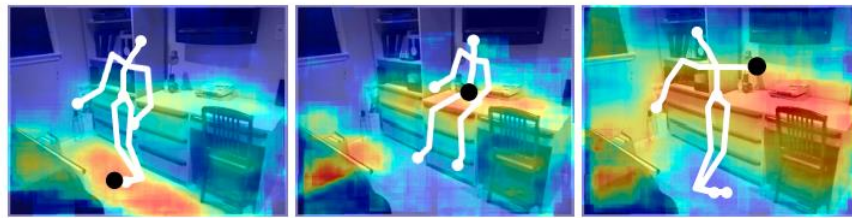
Focus on Functionality/Use

Example Affordance-based Perception

Affordances *are not properties, resources nor features of the environment. Instead they are action-based* ***relations*** *between particular aspects of agents and particular aspects of situations*

Input

Stand Sit Reach



Affordance Labels

In Defense of the Direct Perception of Affordances, Fouhey et al., 2015

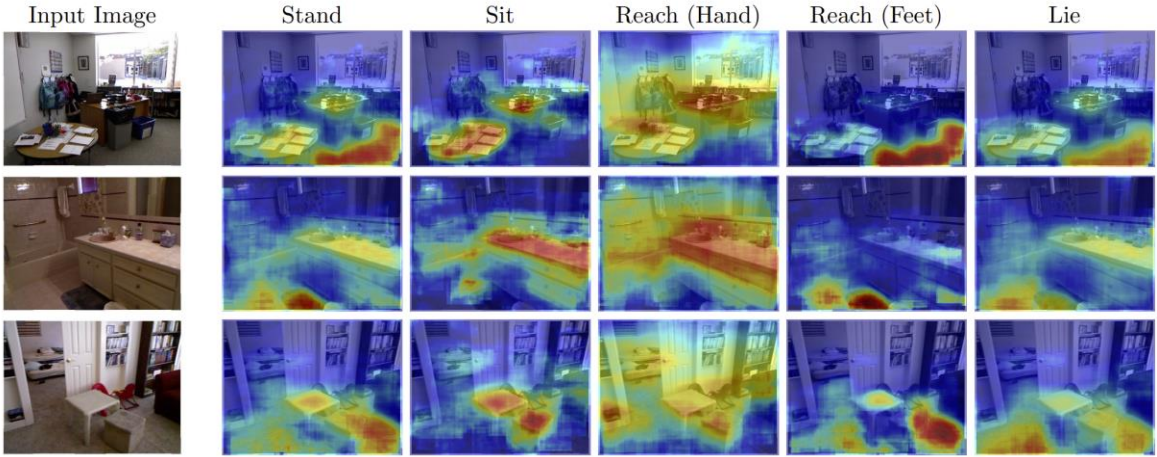


Figure 4: Sample results from our mid-level patch model (red: supports affordance; blue:does not). Our method captures distinctions between the affordances: for instance, one probably cannot touch anything standing on top of the desk in row 1.

In Defense of the Direct Perception of Affordances, Fouhey et al., 2015

Learning pixel-wise affordances for suction and grasping

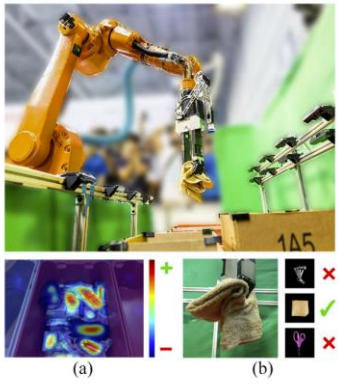


Fig. 1. Our picking system computing pixel-wise affordances for grasping over visual observations of bins full of objects, (a) grasping a towel and holding it up away from clutter, and recognizing it by matching observed images of the towel (b) to an available representative product image. The key contribution is that the entire system works out of the box for novel objects (unseen in training) without the need for any additional data collection or re-training.

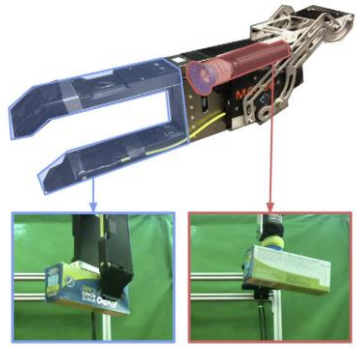
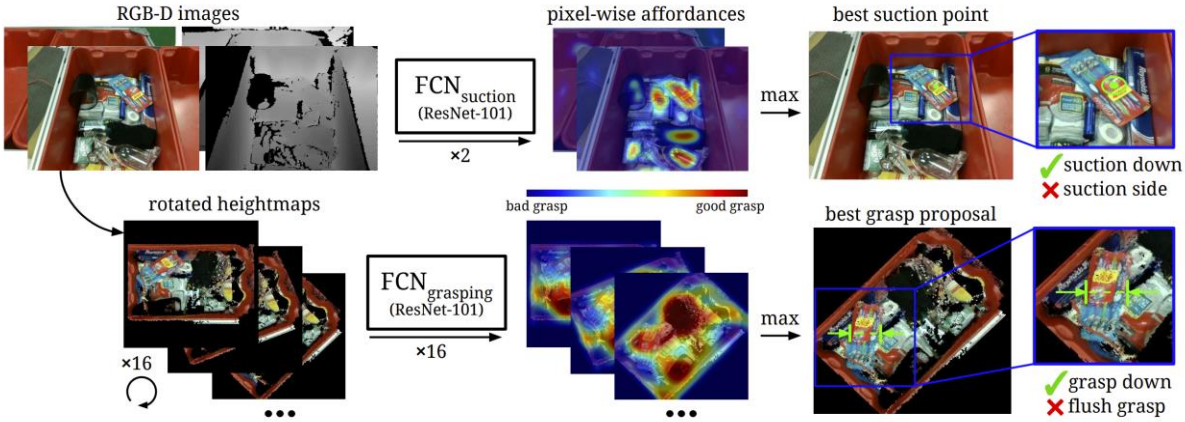


Fig. 3. Multi-functional gripper with a retractable mechanism that enables quick and automatic switching between suction (pink) and grasping (blue).

Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching, IJRR, 2022

Learning pixel-wise affordances for suction and grasping



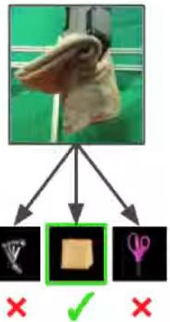
Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching, IJRR, 2022

Robotic Pick-and-Place of Novel Objects in Clutter

with Multi-Affordance Grasping and Cross-Domain Image Matching

Andy Zeng[1], Shuran Song[1], Kuan-Ting Yu[2], Elliott Donlon[2], Francois R. Hogan[2], Maria Bauza[2], Daolin Ma[2], Orion Taylor[2], Melody Liu[2], Eudald Romo[2], Nima Fazeli[2], Ferran Alet[2], Nikhil Chavan Dafle[2], Rachel Holladay[2], Isabella Morona[2], Prem Qu Nair[1], Druck Green[2], Ian Taylor[2], Weber Liu[1], Thomas Funkhouser[1], Alberto Rodriguez[2]

[1]Princeton University [2]Massachusetts Institute of Technology



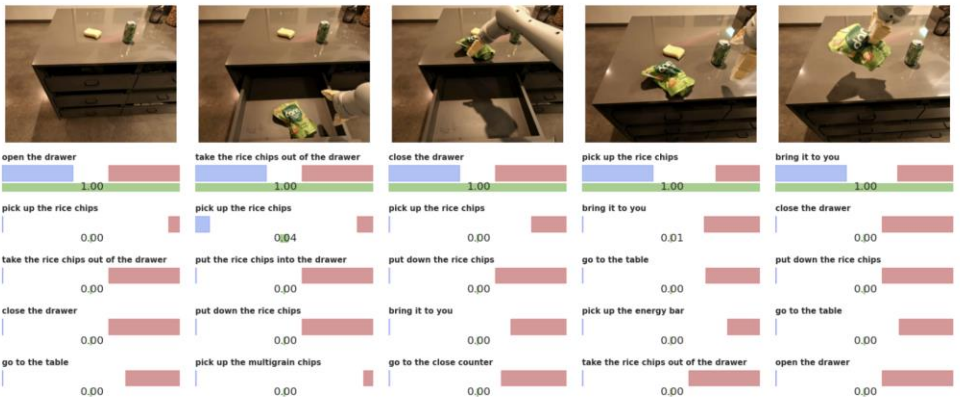
- ✓ Autonomous pick-and-place.
- ✓ Novel objects in clutter with no retraining.
- ✓ 1st place stowing task
  Team MIT-Princeton
  2017 Amazon Robotics Challenge.

© Authors of ICRA 2018 Paper 1502

Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, CoRL, 2022

Human: Bring me the rice chips from the drawer.

Robot: I would
1. Go to the drawer
2. Open the drawer
3. Take the rice chips out of the drawer
4. Close the drawer
5. Pick up the rice chips
6. Bring it to you
7. Put down the rice chips
8. Done



Anticipating Human Activities using Object Affordances for Reactive Robotic Response, RSS, 2009

(a) Robot's RGB-D view. (b) Heatmap of object affordances.

(c) Heatmap of trajectories. (d) Robot opening the door.

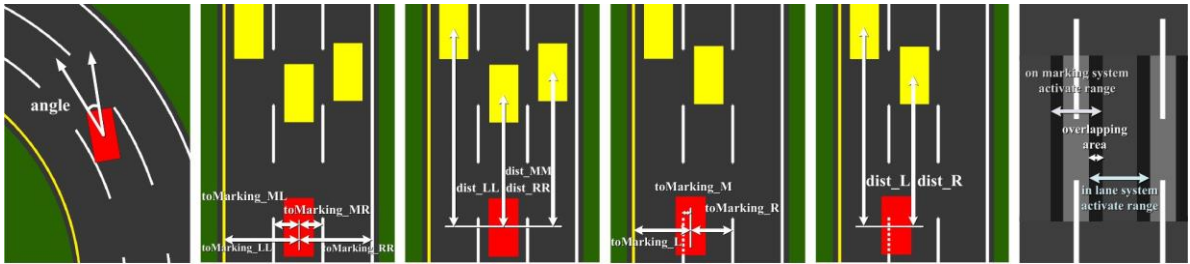
Fig. 1: **Reactive robot response through anticipation:** Robot observes a person holding an object and walking towards a fridge (a). It uses our ATCRF to anticipate the object affordances (b), and trajectories (c). It then performs an anticipatory action of opening the door (d).

Fig. 4: Affordance heatmaps. The first two images show the *reachability* affordance heatmap (red signifies most likely *reachable* locations on the object) and the last two images show the *drinkability* affordance heatmap (red signifies the locations where the object is *drinkable*).

Learning by Observing

Direct Perception for Autonomous Navigation

Direct Perception for Autonomous Driving



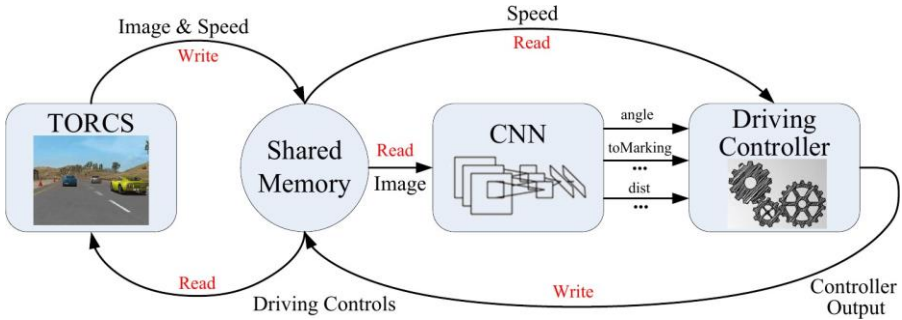
(a) angle (b) in lane: toMarking (c) in lane: dist (d) on mark.: toMarking (e) on marking: dist (f) overlapping area

Affordances:

- Attributes of the environment which limit space of actions [Gibson, 1966]
- Here: 13 affordance indicators
  - 12 of them conditioned on lateral position wrt. road
  - 2 categories: Lane perception and car perception

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

Direct Perception for Autonomous Driving: Overview



- Simulator: TORCS: Open source car racing game simulator
- Network: AlexNet (5 conv layers, 4 fully conn. layers), 13 output neurons
- Training: Affordance indicators trained with L_2 loss (Regression, MSE)

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

Direct Perception for Autonomous Driving: State Machine

always:

1) angle: angle between the car's heading and the tangent of the road

"in lane system", when driving in the lane:

2) toMarking_LL: distance to the left lane marking of the left lane
3) toMarking_ML: distance to the left lane marking of the current lane
4) toMarking_MR: distance to the right lane marking of the current lane
5) toMarking_RR: distance to the right lane marking of the right lane
6) dist_LL: distance to the preceding car in the left lane
7) dist_MM: distance to the preceding car in the current lane
8) dist_RR: distance to the preceding car in the right lane

"on marking system", when driving on the lane marking:

9) toMarking_L: distance to the left lane marking
10) toMarking_M: distance to the central lane marking
11) toMarking_R: distance to the right lane marking
12) dist_L: distance to the preceding car in the left lane
13) dist_R: distance to the preceding car in the right lane

Figure 4: **Complete list of affordance indicators in our direct perception representation.**

while (in autonomous driving mode)
  ConvNet outputs affordance indicators
  check availability of both the left and right lanes
  if (approaching the preceding car in the same lane)
    if (left lane exists **and** available **and** lane changing allowable)
      left lane changing decision made
    else if (right lane exists **and** available **and** lane changing allowable)
      right lane changing decision made
    else
      slow down decision made
  if (normal driving)
    $center\_line$= center line of current lane
  else if (left/right lane changing)
    $center\_line$= center line of objective lane
  compute steering command
  compute $desired\_speed$
  compute acceleration/brake command based on $desired\_speed$

Figure 5: **Controller logic.**

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

Direct Perception for Autonomous Driving: Controller

Steering controller:

$$s = \theta_1(\alpha - d_c/w)$$

- s: steering command θ_1: parameter
- α: relative orientation d_c: distance to centerline w: road width

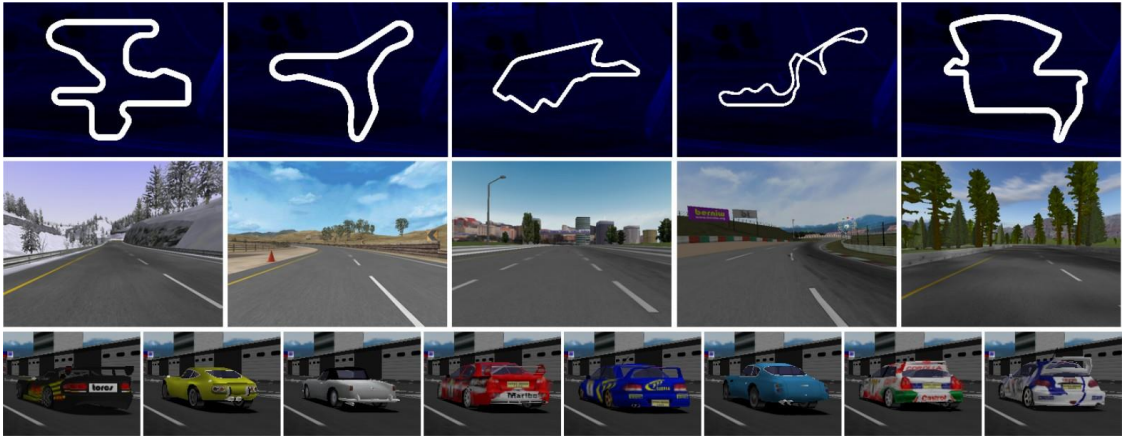
Speed controller:

$$v = v_{max}\,(1 - \exp(-\theta_2 d_p - \theta_3))$$

- v: target velocity v_{max} maximal velocity
- d_p: distance to preceding car $\theta_{2,3}$: parameters

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

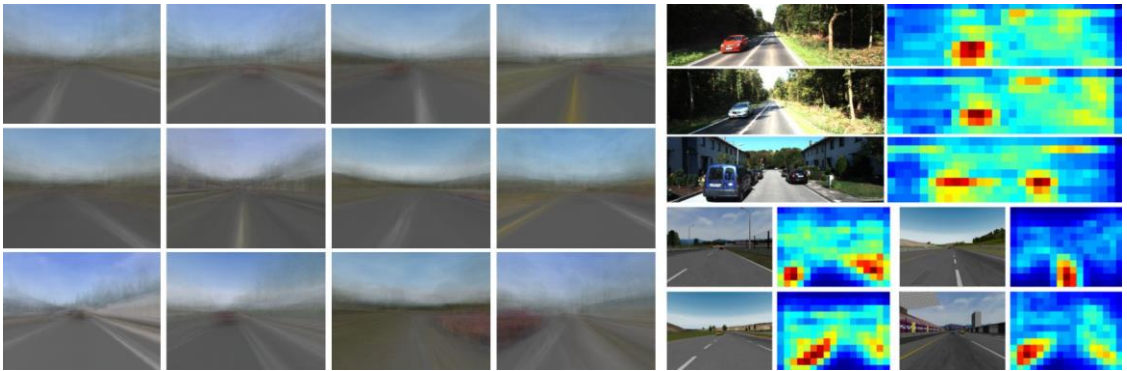
Direct Perception for Autonomous Driving: TORCS Simulator



► TORCS: Open source car racing game http://torcs.sourceforge.net/

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

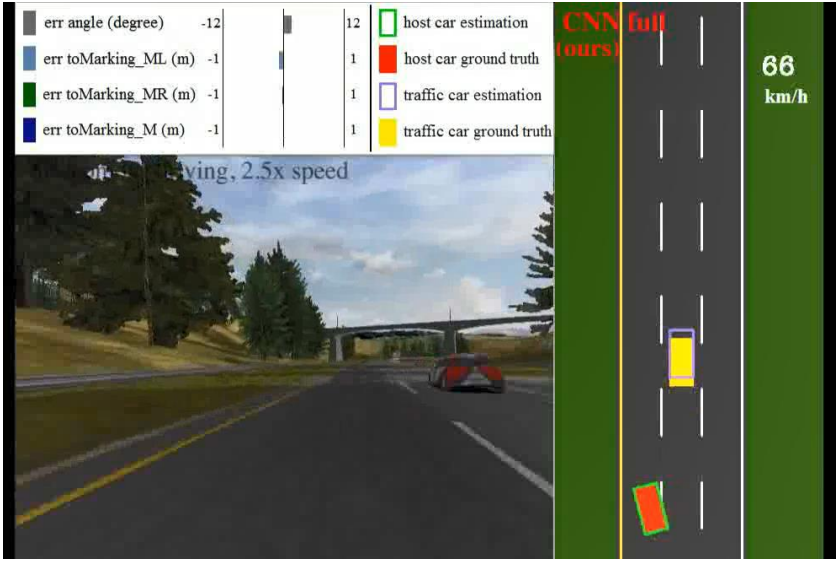
Direct Perception for Autonomous Driving: Visualization



- Left: Averaged top 100 images activating a neuron in first fully connected layer
- Right: Maximal response of 4th conv. layer (note: focus on cars and markings)

Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

Direct Perception for Autonomous Driving: Results



Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.

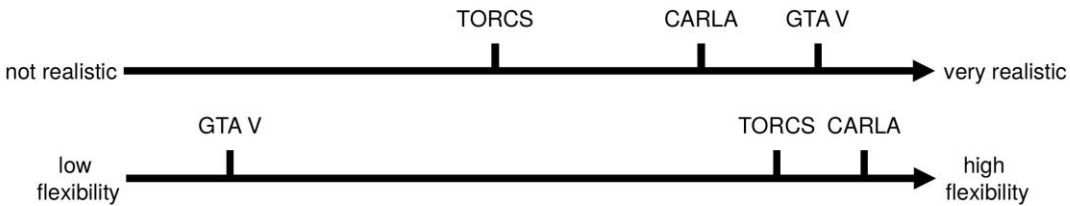
Any Limitations?

Conditional Affordance Learning: Simulators

TORCS

GTA V

CARLA



Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

How can we apply to cities?

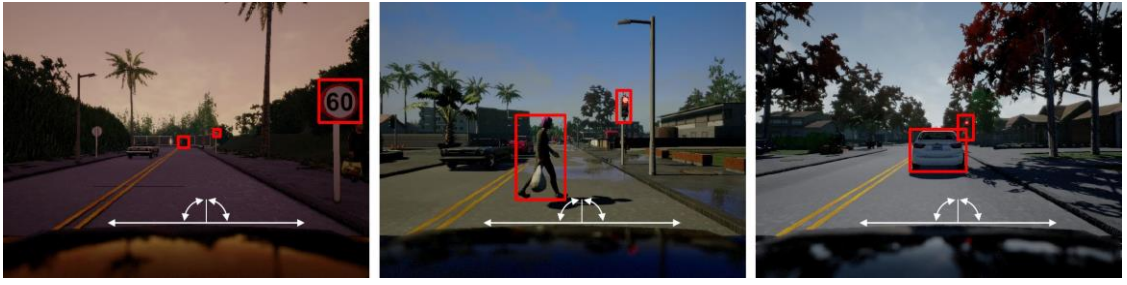
Conditional Affordance Learning

Conditional Affordance Learning: Goals

- Goal: drive from A to B as fast, safely and comfortably as possible
- Infractions:
  - Driving on wrong lane
  - Driving on sidewalk
  - Running a red light
  - Violating speed limit
  - Colliding with vehicles
  - Hitting other objects

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Affordances



Affordances:

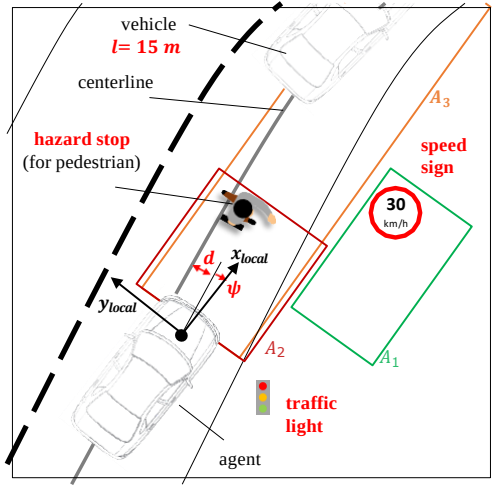
- Distance to centerline
- Relative angle to road
- Distance to lead vehicle
- Speed signs
- Traffic lights
- Hazard stop

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Affordances

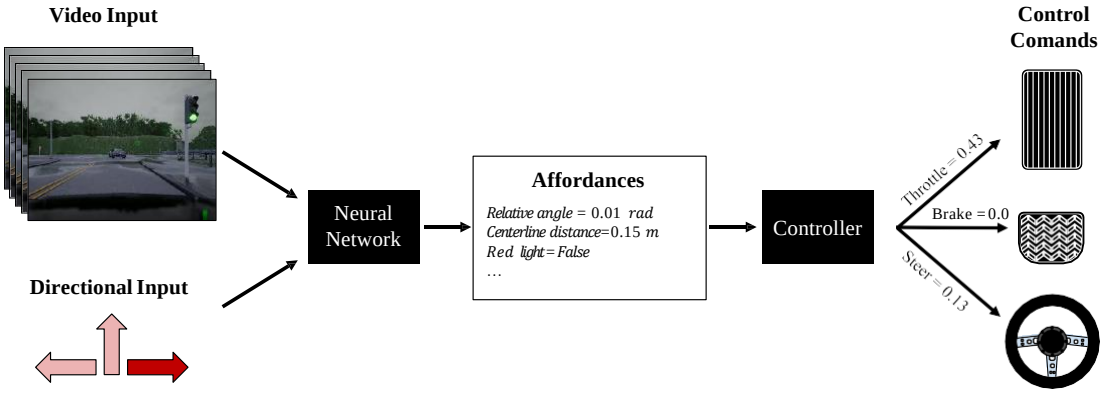
Affordances:

- Distance to centerline
- Relative angle to road
- Distance to lead vehicle
- Speed signs
- Traffic lights
- Hazard stop



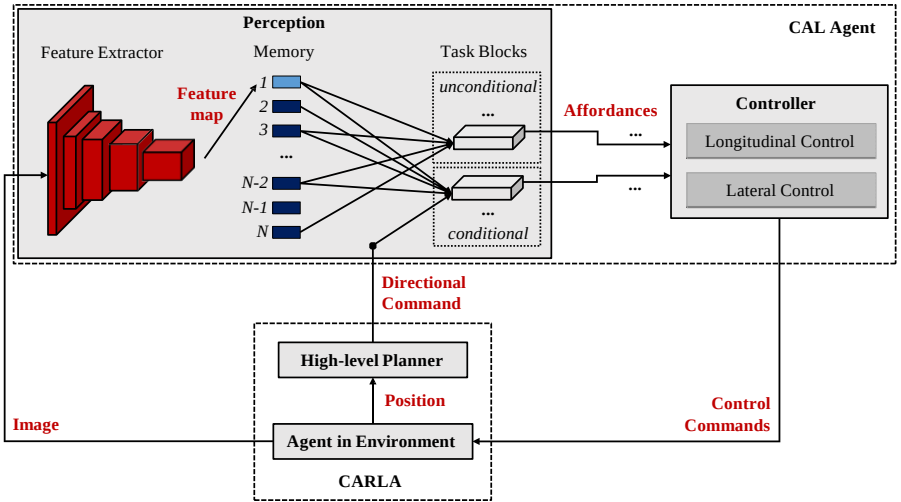
Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning



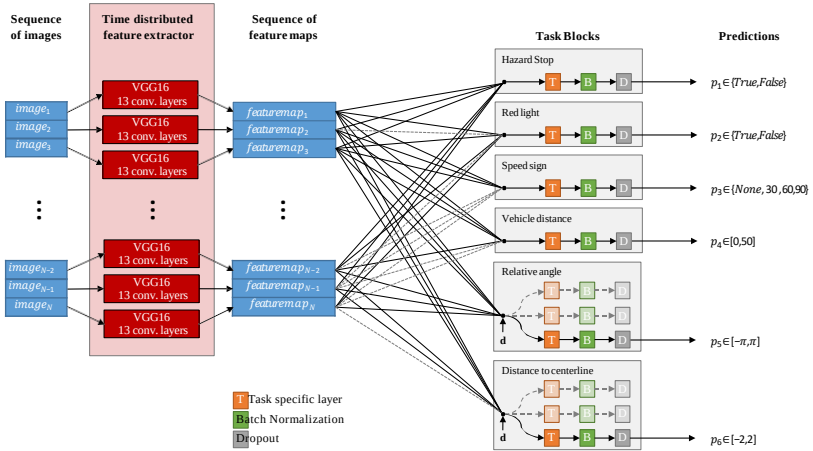
Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: System Overview



Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Perception Stack



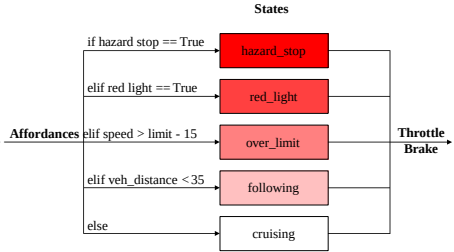
► Feature network: VGG16

► Task network: FC / TCN

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

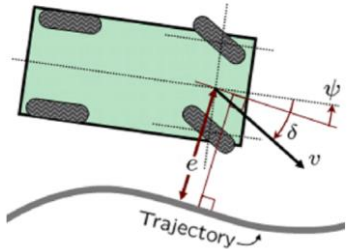
Conditional Affordance Learning: Controller

Longitudinal Control



- Finite-state machine
- PID controller for cruising
- Car following model

Lateral Control



- Stanley controller
- $\delta(t) = \psi(t) + \arctan(\frac{ke(t)}{v(t)})$
- Dampening term (rate) k

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Parameter Estimation

Perception Stack:

- Multi-task learning: single forward pass
- Dataset: random driving using controller operating on GT affordances
  $\Rightarrow$ 240k images with GT affordances
- Loss functions:
  - Discrete affordances: Class-weighted cross-entropy (CWCE)
  - Continuous affordances: Mean average error (MAE)
- Hyperparameter search: temp. window
- Optimized with ADAM (batch size 32)

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

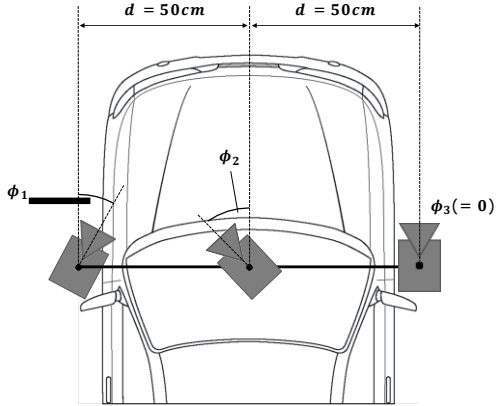
Conditional Affordance Learning: Data Collection

Data Collection:

- Navigation based on true affordances & random inputs
- Image-level annotations

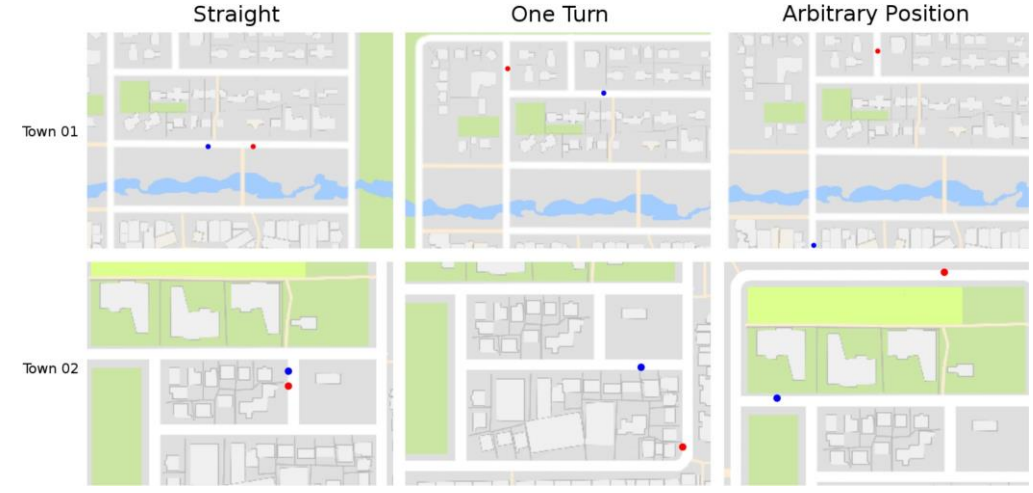
Data Augmentation:

- No image flipping
- Color, contrast, brightness
- Gaussian blur & noise
- Provoke rear-end collisions
- Camera pose randomization



Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Tasks



- Goal-oriented navigation

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Results

	Training conditions				New weather				New town				New town and new weather			
Task	MP	CIL	RL	CAL	MP	CIL	RL	CAL	MP	CIL	RL	CAL	MP	CIL	RL	CAL
Straight	98	95	89	**100**	**100**	98	86	**100**	92	**97**	74	93	50	80	68	**94**
One turn	82	89	34	**97**	95	90	16	**96**	61	59	12	**82**	50	48	20	**72**
Navigation	80	86	14	**92**	**94**	84	2	90	24	40	3	**70**	47	44	6	**68**
Nav. dynamic	77	**83**	7	**83**	**89**	82	2	82	24	38	2	**64**	44	42	4	**64**

Baselines:

- MP = Modular Pipeline [Dosovitskiy et al., CoRL 2017]
- CIL = Conditional Imitation Learning [Codevilla et al., ICRA 2018]
- RL = Reinforcement Learning A3C [Mnih et al., ICML 2016]

Conditional Affordance Learning: Results

Hazard Stop

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Conditional Affordance Learning: Attention

Attention to Hazard Stop

Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Learning to Move with Affordance Maps, ICLR 2020

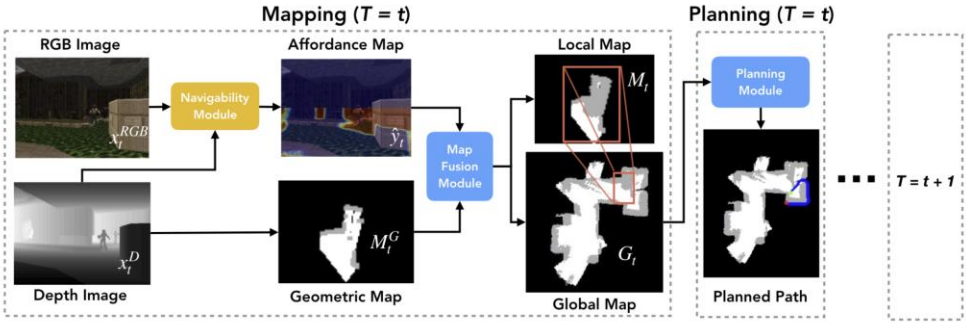


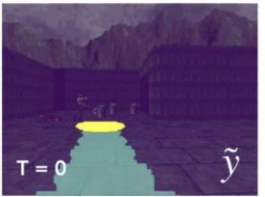
Figure 1: Overview of our proposed architecture for navigation. RGBD inputs x_t are used to predict affordance maps $\hat{y}_t$ and transformed into egocentric navigability maps M_t that incorporate both geometric and semantic information. In the example shown, M_t is labelled as non-navigable in regions near the monster. A running estimate of the current position at each time step is maintained and used to update a global, allocentric map of navigability G_t that enables safe and efficient planning.

https://openreview.net/pdf?id=BJgMFxrYPB

Learning to Move with Affordance Maps, ICLR 2020

Termination Condition

Labeled Navigability



Episode Initialization → Active/Random Trajectory Sampling → Exploration ⇄ Sensor Feedback → Back-Projection

https://openreview.net/pdf?id=BJgMFxrYPB

Results

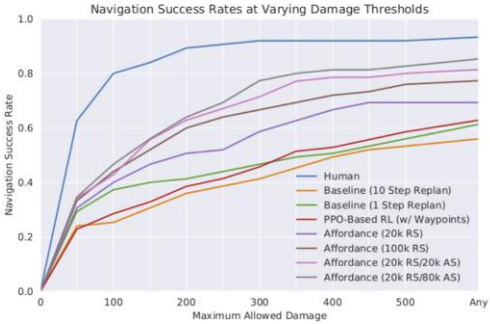


Figure 5: Comparison of navigation performance across all evaluated approaches, plotted as a function of success rate vs. maximum amount of damage permitted per trial (mean results over 5 test runs reported).

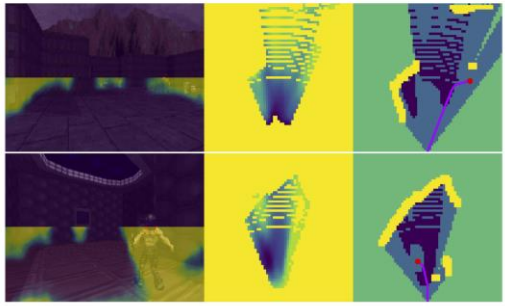


Figure 6: Examples of actively-planned trajectories that maximize label entropy along sampled locations. (**Left**) shows predicted affordances, (**Middle**) shows the projected confidence map, and (**Right**) shows the cost map used to plan the optimal path.

https://openreview.net/pdf?id=BJgMFxrYPB

Learned Visual Navigation for Under-Canopy Agricultural Robots, RSS 2021

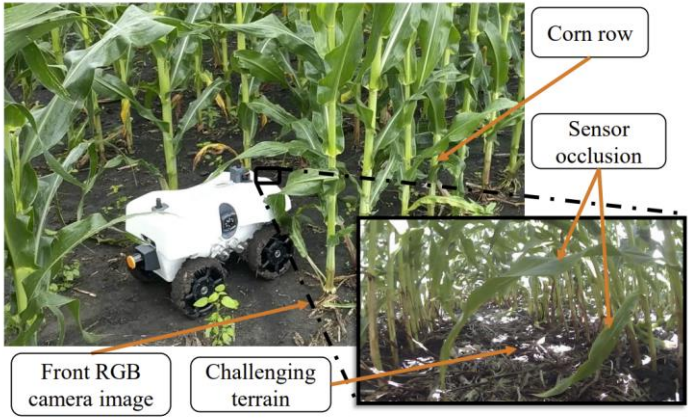


Fig. 1: CropFollow is an autonomous navigation system for under-canopy agriculture robots. It uses RGB images from a front-facing camera to output steering commands to drive the robot in crop rows.

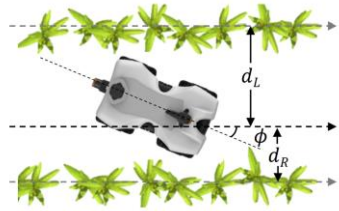


Fig. 3: Our method uses the robot's heading, ϕ and ratio of distance from the left and the right crop row, $d = d_L/(d_L + d_R)$, as the intermediate representation between perception and planning.

Learned Visual Navigation for Under-Canopy Agricultural Robots, RSS 2021

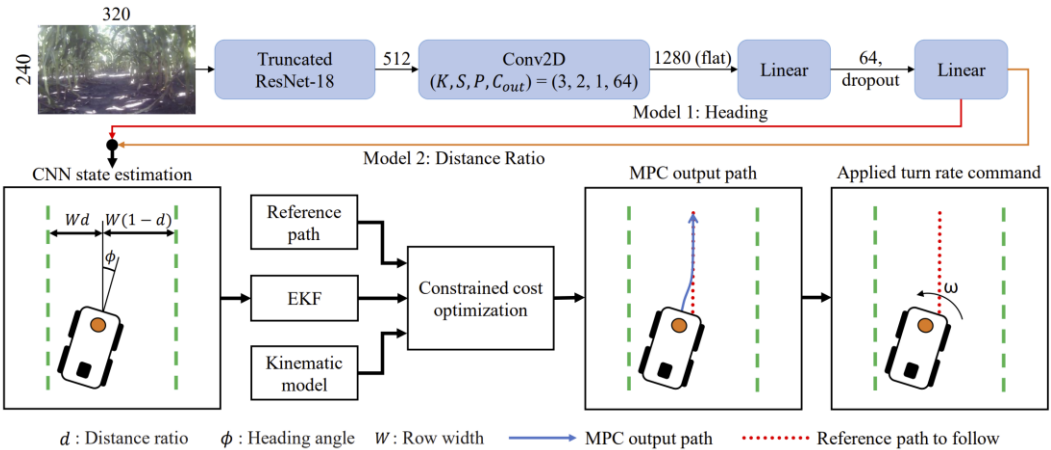
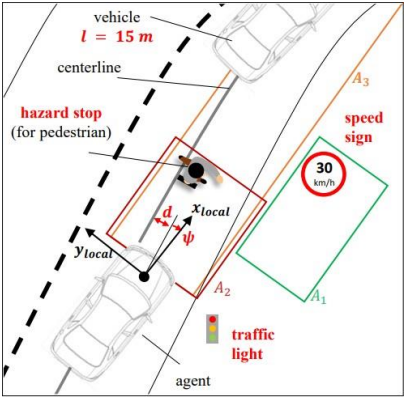


Fig. 2: CropFollow Overview. We use a convolutional network to output robot heading and placement in row. This is used to compute the row center which is used as a reference trajectory. A model predictive controller converts reference trajectories to angular velocity commands.

Conditional Affordance Learning

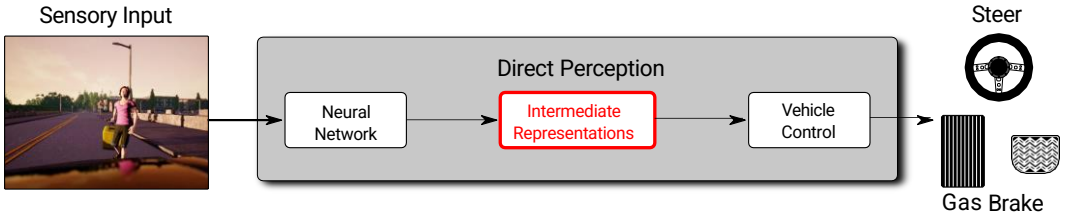
Conditional Affordance Learning:

- Map: Observations → affordances → actions
- Affordances: angle wrt. road, distance to lane boundaries or other cars, etc.
- Decoupling of perception and action
  ⇒ Better generalization?
- Rule-based controller
- Misspecification of affordances



Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

Approaches to Self-Driving



Which intermediate representation?

- Selected/hand-crafted attributes
- Semantic segmentation
- Bounding boxes
- Depth
- Motion/Optical Flow

End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances, CVPR 2020

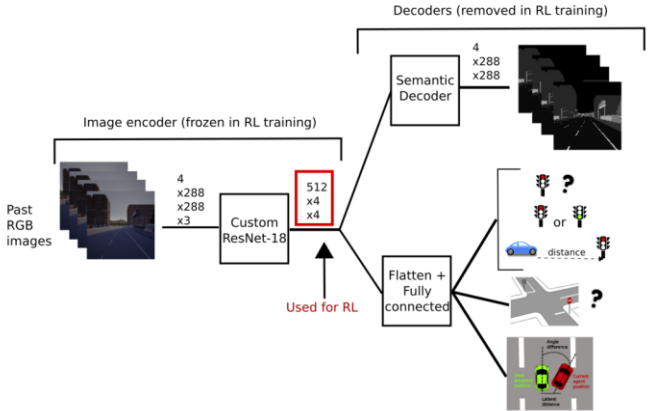


Figure 5. Decoder and losses used to train the encoder: semantic segmentation, traffic light (presence, state, distance), intersection presence, lane position (distance and rotation)

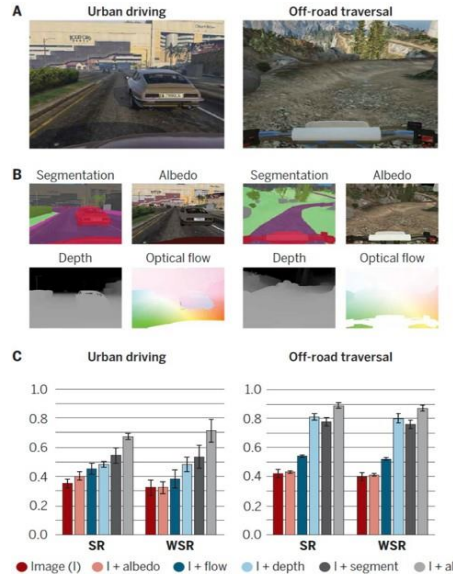
What makes a good abstraction?

- Invariant (hide irrelevant variations from the final policy)
- Universal (applicable to a wide range of scenarios)
- Encode task-relevant knowledge (road is driveable)
- Data and label efficient

Does Computer Vision Matter for Action?

Does Computer Vision Matter for Action?

- ► Analyze various intermediate representations: segmentation, depth, normals, flow…
- ► Intermediate representations improve results
- ► Consistent gains across simulations / tasks
- ► Depth and semantic provide largest gains
- ► Better generalization performance

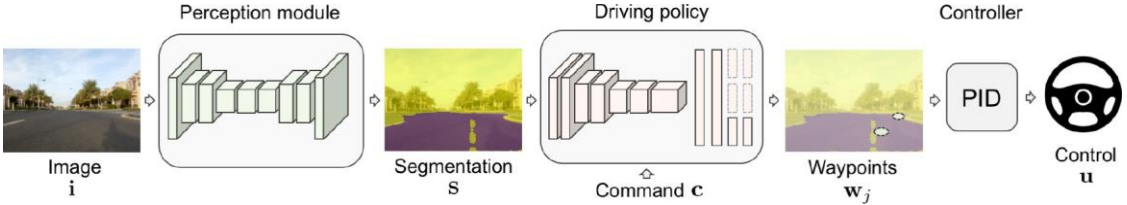


Zhou, Kra˙henbu˙hl and Koltun: Does computer vision matter for action? Science Robotics, 2019.

Does Computer Vision Matter for Action?

Driving Policy Transfer

Driving Policy Transfer: Overview



Problem:

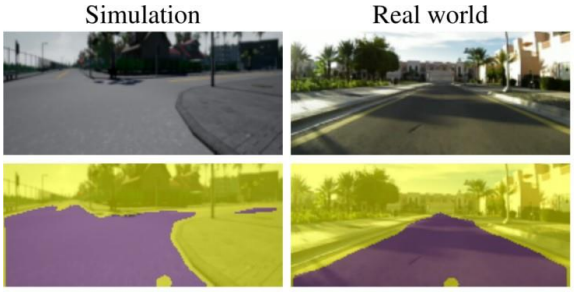
- Driving policies learned in simulation often do not transfer well to the real world

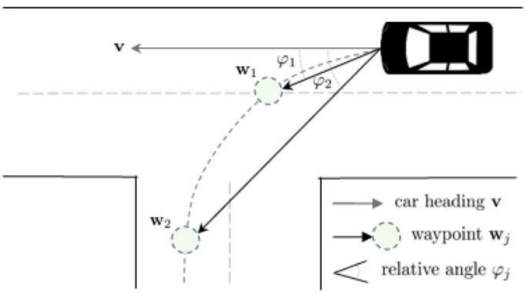
Idea:

- Encapsulate driving policy such that it is not directly exposed to raw perceptual input or low-level control (input: semantic segmentation, output: waypoints)
- Allows for transferring driving policy without retraining or finetuning

Müller, Dosovitskiy, Ghanem and Koltun: Driving Policy Transfer via Modularity and Abstraction. CoRL, 2018.

Driving Policy Transfer: Waypoints



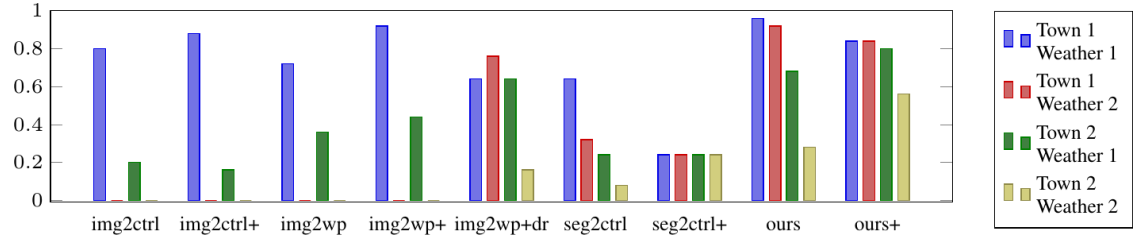


Representation:

- Input: Semantic segmentation (per pixel “road” vs. “non-road”)
  - ERFNet [Romera et al., IV 2017] trained on Cityscapes [Cordts et al., CVPR 2016]
- Output: 2 waypoints (distance to vehicle, relative angle wrt. vehicle heading)
  - One sufficient for steering, second one for braking before turns

Müller, Dosovitskiy, Ghanem and Koltun: Driving Policy Transfer via Modularity and Abstraction. CoRL, 2018.

Driving Policy Transfer: Results



Success Rate over 25 Navigation Trials

- Driving policy: Conditional Imitation Learning (branched)
- Training: Expert agent, random initialization, noise injection
- Control: PID for lateral and longitudinal control
- Results: Full method generalizes best ("+" = with data augmentation)

Müller, Dosovitskiy, Ghanem and Koltun: Driving Policy Transfer via Modularity and Abstraction. CoRL, 2018.

Driving Policy Transfer: Results

Müller, Dosovitskiy, Ghanem and Koltun: Driving Policy Transfer via Modularity and Abstraction. CoRL, 2018.

Further Readings

- Bojarski et al.: VisualBackProp: Efficient Visualization of CNNs for Autonomous Driving. ICRA, 2018.
- Codevilla, Mu¨ller, Lo´pez, Koltun and Dosovitskiy: End-to-End Driving Via Conditional Imitation Learning. ICRA, 2018.
- Chen, Seff, Kornhauser and Xiao: Learning Affordance for Direct Perception in Autonomous Driving. ICCV, 2015.
- Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.
- Muller, Dosovitskiy, Ghanem and Koltun: Driving Policy Transfer via Modularity and Abstraction. CoRL, 2018.
- Zhou et al., Does Computer Vision Matter for Action?, Science Robotics, 2019.