Representation Learning for Driving

Alex Kendall @ CVPR, Long Beach, June 2019

WAYVE


112
6VVA091


CITY
SPEED
LIMIT
25
7 Ave
SLOW ZONE
BUSES
ONE WAY
W 41 ST
M7
M20
MTA


20
WX16 FDY

WAYVE

Outline of talk

1. Recipe for success for representation learning
2. Strategy for training data
3. Interpretability & verification

Machine Learning for Autonomous Driving

Some Background

Alex Kendall © Wayve 2019

1989 ALVINN: End-to-End Imitation Learning

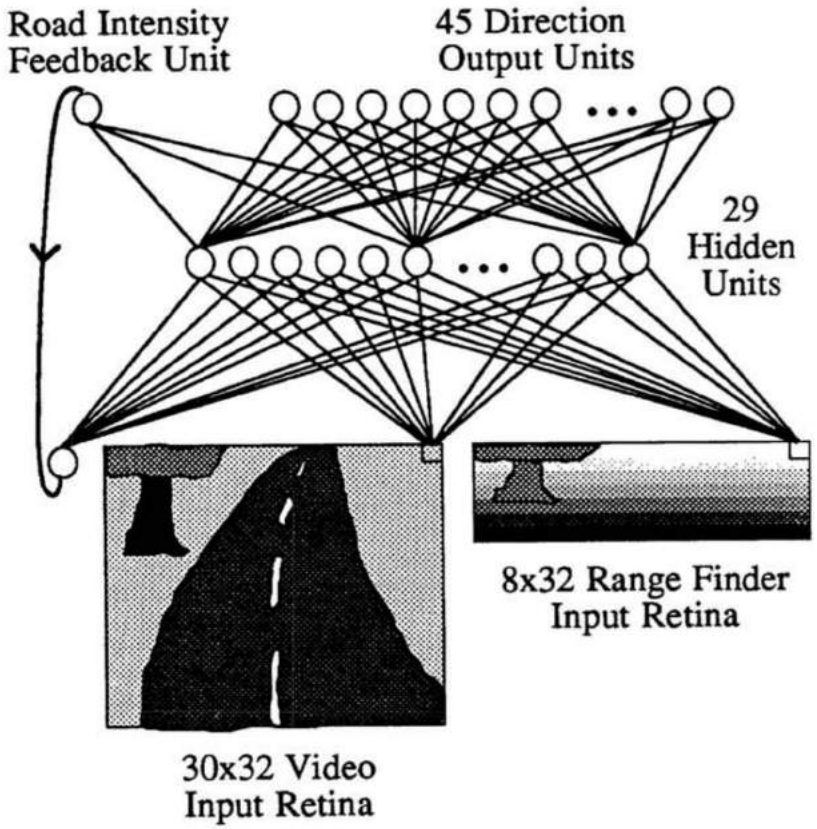


Figure 1: ALVINN Architecture

What's Hidden in the Hidden Layers?

The contents can be easy to find with a geometrical problem, but the hidden layers have yet to give up all their secrets

David S. Touretzky and Dean A. Pomerleau

tions, we fed the network road images taken under a wide variety of viewing angles and lighting conditions. It would be impractical to try to collect thousands of real road images for such a data set. Instead, we developed a synthetic road-image generator that can create as many training examples as we need.

To train the network, 1200 simulated road images are presented 40 times each, while the weights are adjusted using the back-propagation learning algorithm. This takes about 30 minutes on Carnegie Mellon's Warp systolic-array supercomputer. (This machine was designed at Carnegie Mellon and is built by General Electric. It has a peak rate of 100 million floating-point operations per second and can compute weight adjustments for back-propagation networks at a rate of 20 million connections per second.)

Once it is trained, ALVINN can accurately drive the NAVLAB vehicle at about 3½ miles per hour along a path through a wooded area adjoining the Carnegie Mellon campus, under a variety of weather and lighting conditions. This speed is nearly twice as fast as that achieved by non-neural-network algorithms running on the same vehicle. Part of the reason for this is that the forward pass of a back-propagation network can be computed quickly. It takes about 200 milliseconds on the Sun-3/160 workstation installed on the NAVLAB.

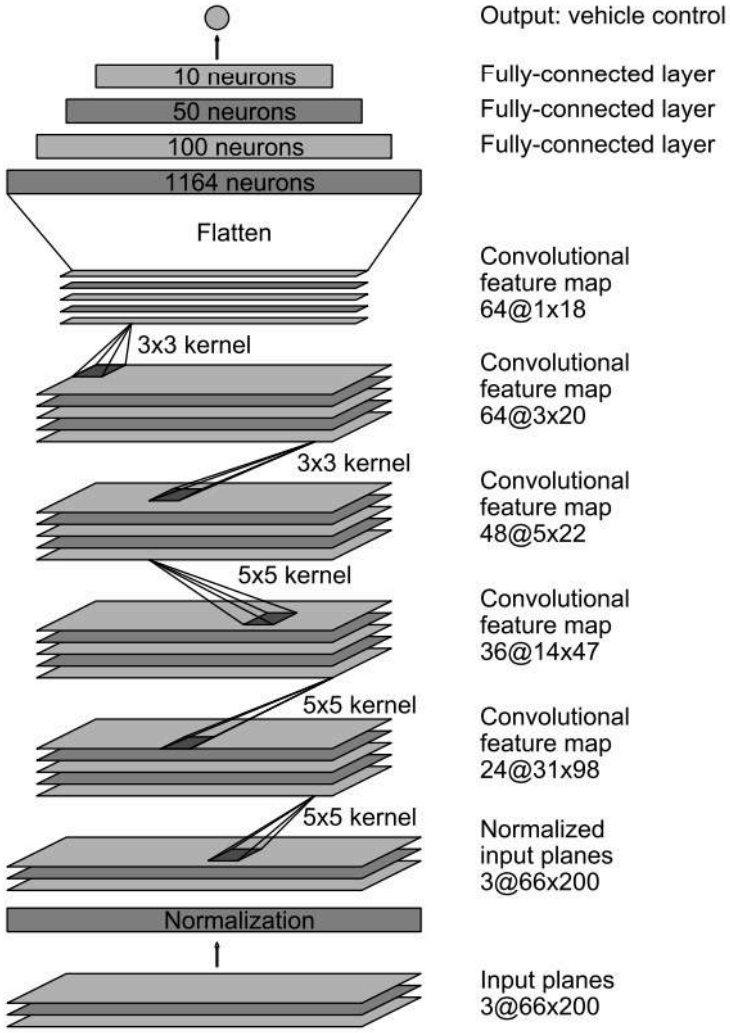
The hidden-layer representations ALVINN develops are interesting. When trained on roads of a fixed width, the network chooses a representation in which hidden units act as detectors for complete roads at various positions and orientations. When trained on roads of variable

continued

Photo 1: *The NAVLAB autonomous navigation test-bed vehicle and the road used for trial runs.*

Pomerleau, Dean A. "ALVINN: An autonomous land vehicle in a neural network." Advances in neural information processing systems. 1989.

2016 NVIDIA: Lane Following on Highways



Bojarski, Mariusz, et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).

Urban driving with end-to-end machine learning

Inputs: camera video and a sat-nav

End-to-end deep learning

Uncertainty propagation from *sensing* to *action*

Outputs: driving commands

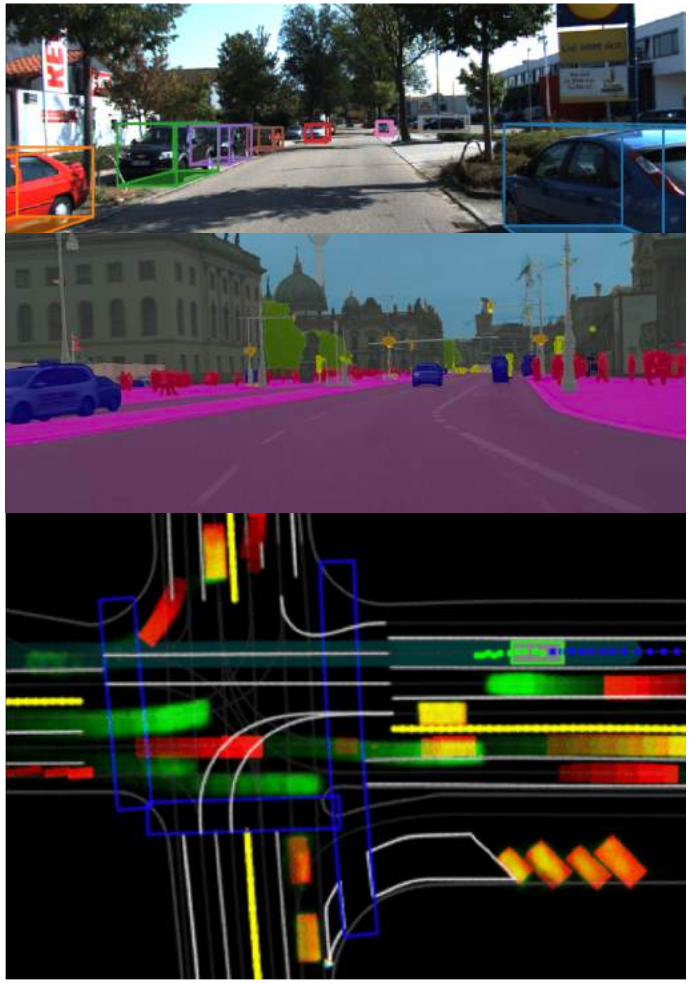
WAYVE

Alex Kendall et al. Learning to Drive in a Day. ICRA, 2019
Wayve, https://wayve.ai/blog/driving-like-human (2019)

A recipe for representing driving

Alex Kendall © Wayve 2019

The Self-Driving State Representation Today



3D Object Detection

Semantic Segmentation

Agent Prediction

Turning indicator detector

HD Map

Driving Affordability Prediction

Traffic sign detection

.....

Autonomous Driving Representation

We can’t enumerate the information we need for every last edge case

Billions of dollars and 10 years of commercial resources can’t do it in a constrained environment like Phoenix, Arizona.

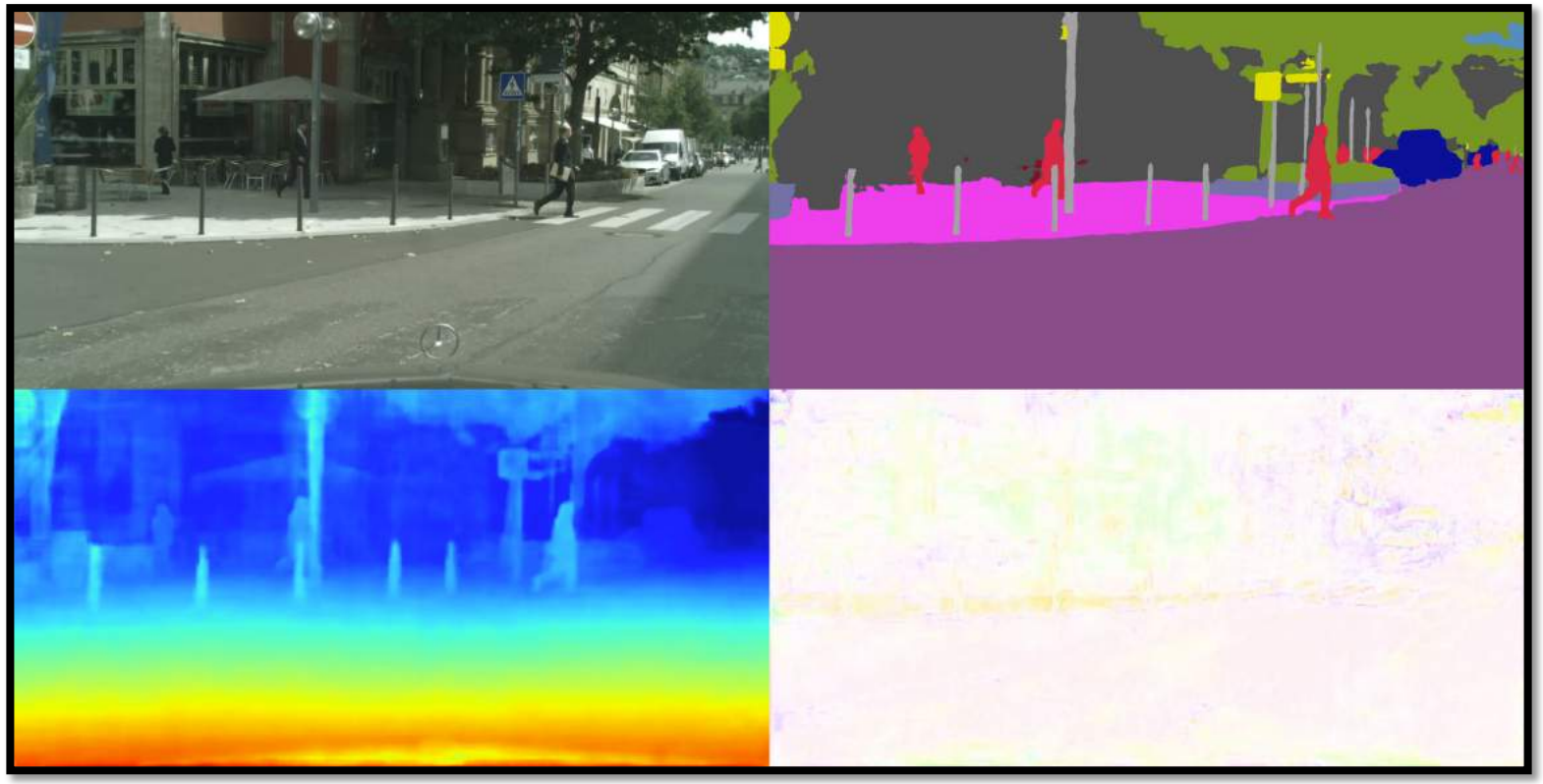
A recipe for a good representation

1. Needs to encode information that we believe is necessary (but not sufficient) for the task
   - For driving, this includes semantics, motion and geometry

2. Should also be optimised w.r.t. the end task
   - Therefore we need an end to end learning signal

3. The decision must be observable in the input data
   - We need the right sensor type and configuration

4. Our representation must have a very good signal to noise ratio
   - We must transform the signal into a compressed, nuisance free & invariant representation

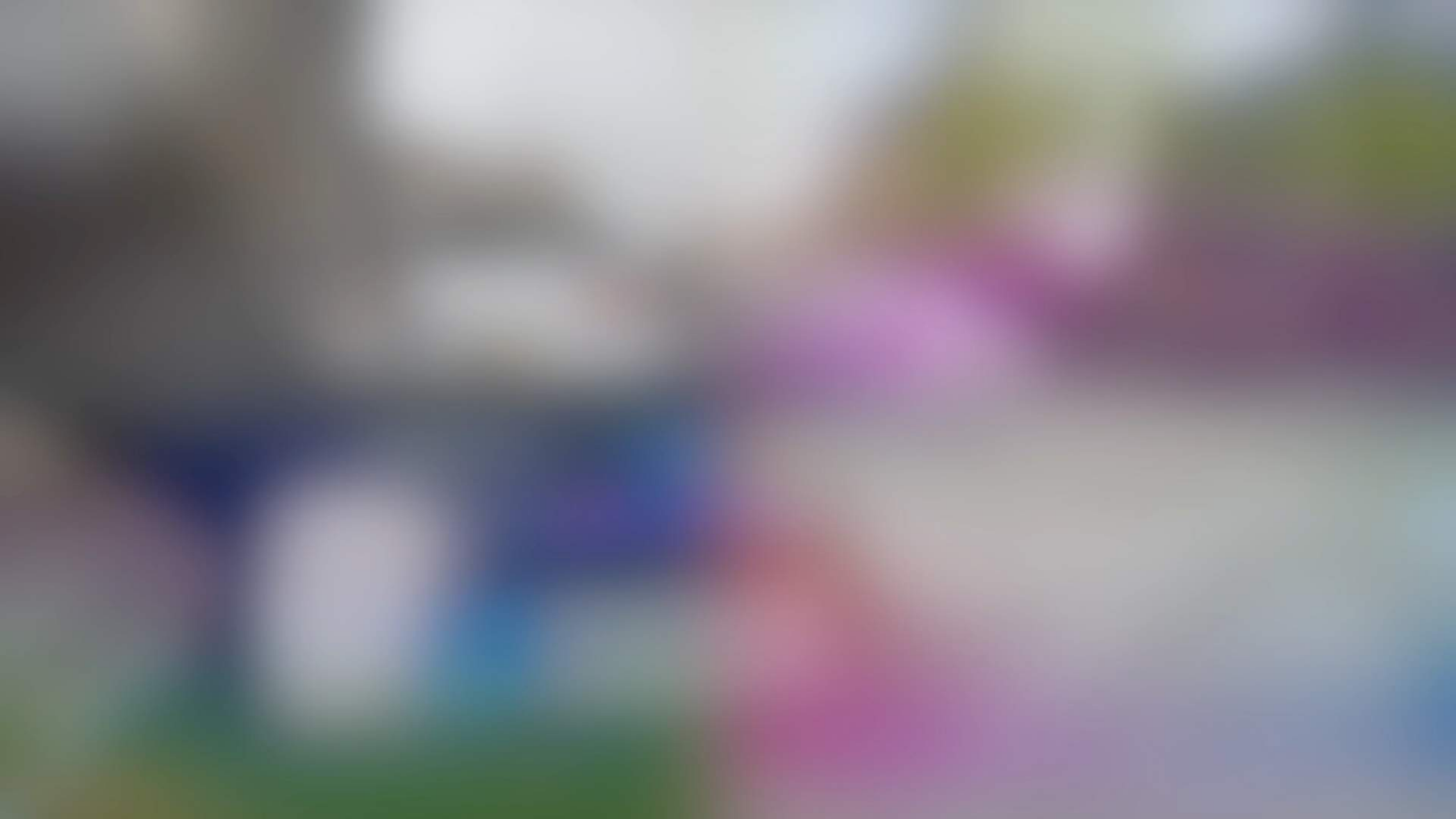


Progression of computer vision from

2015

… to 2018

Badrinarayanan, Kendall, Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. PAMI, 2017.
Kendall, Gal and Cipolla. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. CVPR, 2018.



Training Data

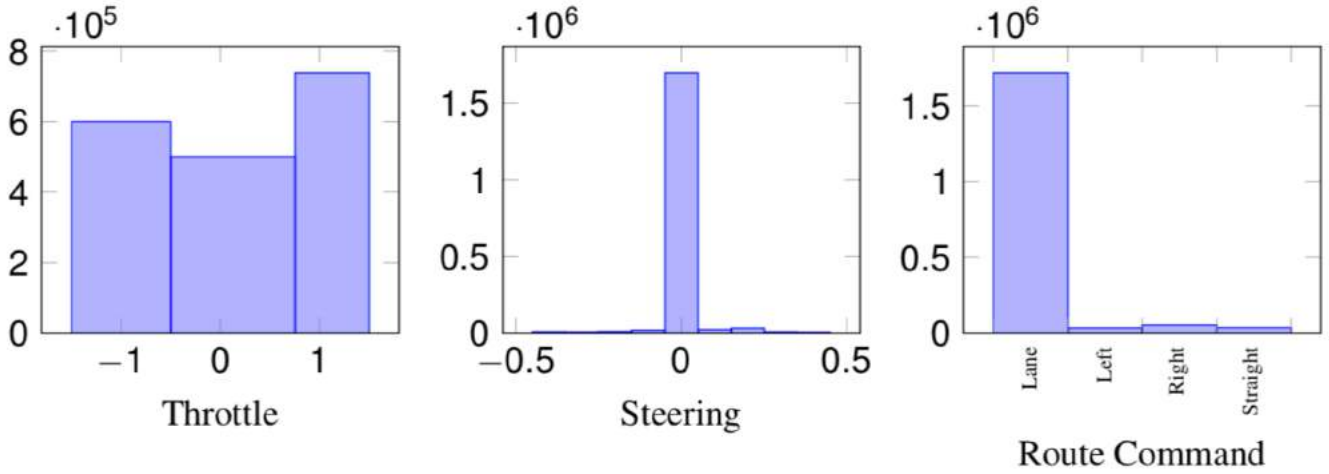
How much and what type do we need?

Alex Kendall © Wayve 2019

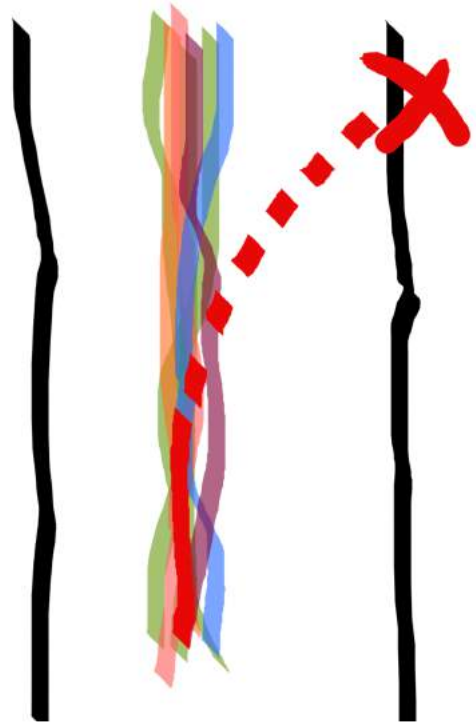
How much data do we need?

- It's not the amount, but the type of data!
- Not all data is created equal
- Important you create a driving curriculum and can seek the right data to improve
  - Off-policy / dash cam data is not good enough!
  - Beneficial to have control over what data is collected
  - Probably need to have on-policy data

Driving data is exceptionally biased

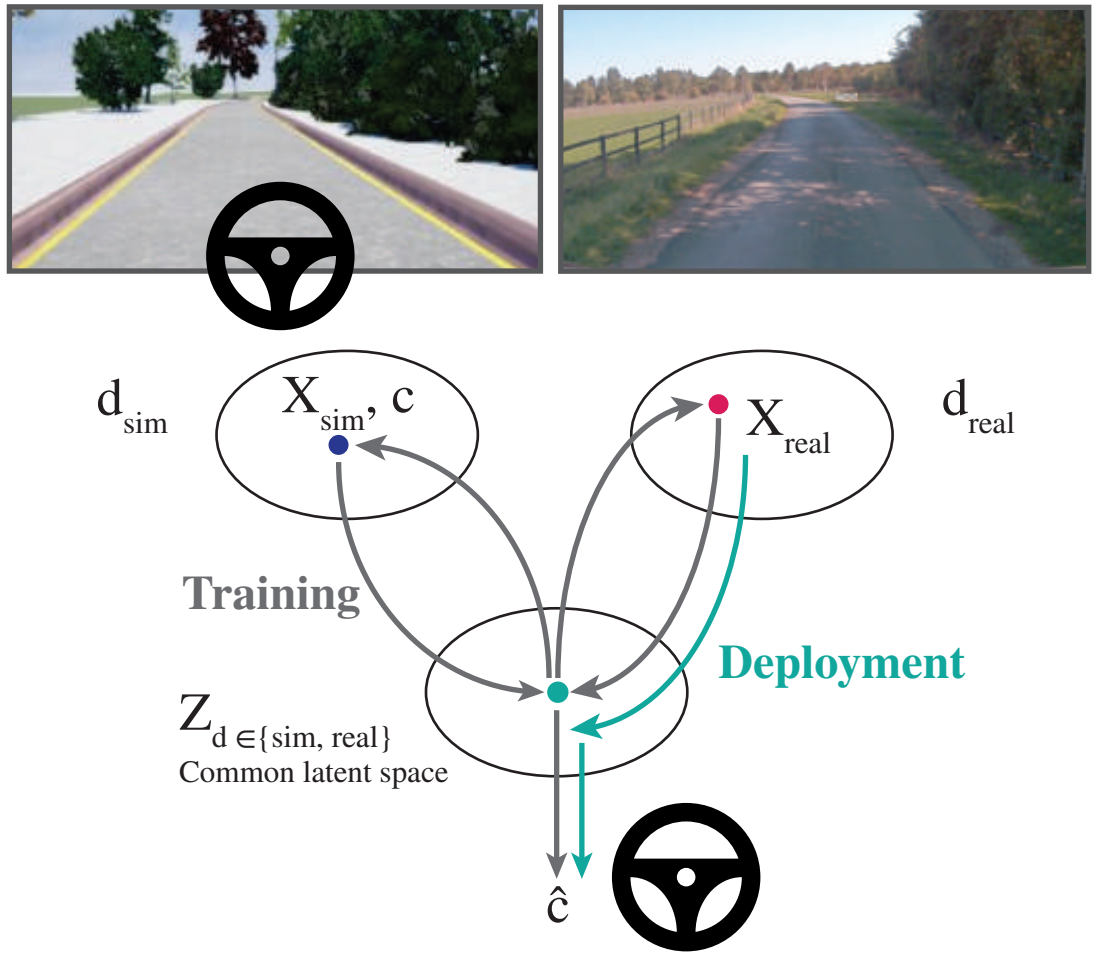


- How much of the state space do we need to explore to learn a good representation?
- If we need training examples densely across all state space, human driving data is not sufficient
- But exploration is dangerous in the real world…



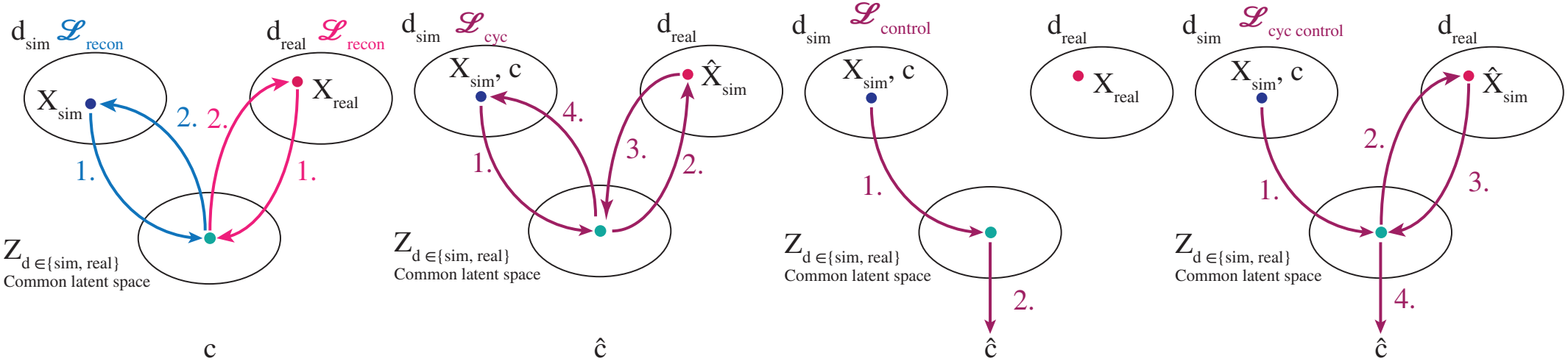
Can we train real-world models in simulated worlds?

- Zero shot sim2real
- Learn to project to a latent space for domain translation and control jointly
- Demonstrate this method can drive 3km+ on public UK roads



Alex Bewley et al. Learning to Drive from Simulation without Real World Labels. ICRA, 2019.

Learning to Drive from Simulation without Real World Labels



(a) Reconstruction loss (b) Cyclic recon. loss (c) Control loss (d) Cyclic control loss

Reconstruction Loss

$$X_d^{recon} = G_d(E_d(X_d))$$

Cyclic Reconstruction Loss

$$X_d^{cyc} = G_d(E_{d'}(G_{d'}(E_d(X_d))))$$

Control Loss

$$\hat{c} = C(E_d(X_d))$$

Cyclic Control Loss

$$\hat{c}^{cyc} = C(E_{d'}(G_{d'}(E_d(X_d))))$$

Not shown: adversarial LSGAN loss, latent reconstruction loss, perceptual loss.

Alex Bewley et al. Learning to Drive from Simulation without Real World Labels. ICRA, 2019.

Comparison to Baseline Methods

	Simulation		Real		
	MAE	Bal-MAE	MAE	Bal-MAE	DPI (metres)
Drive-Straight	0.043	0.087	**0.019**	0.093	$23^{\dagger}$
Simple Transfer	0.055	0.056	0.265	0.272	$9^{\dagger}$
Real-to-Sim Translation	-	-	0.261	0.234	$10^{\dagger}$
Sim-to-Real Translation	-	-	0.059	**0.045**	$28^{\dagger}$
Latent Feature ADA [3]	0.040	0.047	0.032	0.071	$15^{\dagger}$
Ours	**0.017**	**0.018**	0.081	0.087	$>$**3000**

Alex Bewley et al. Learning to Drive from Simulation without Real World Labels. ICRA, 2019.


B&B
Building Contractors
GB

Interpretability & Verification of Deep Learning Representations

Alex Kendall © Wayve 2019

Model-Based Saliency

Suppose $f(\cdot)$ is our driving model and $m(\cdot)$ is our saliency model and $L(\cdot)$ is our loss function for the driving model and the operator $x \cdot m$ degrades the image with noise.

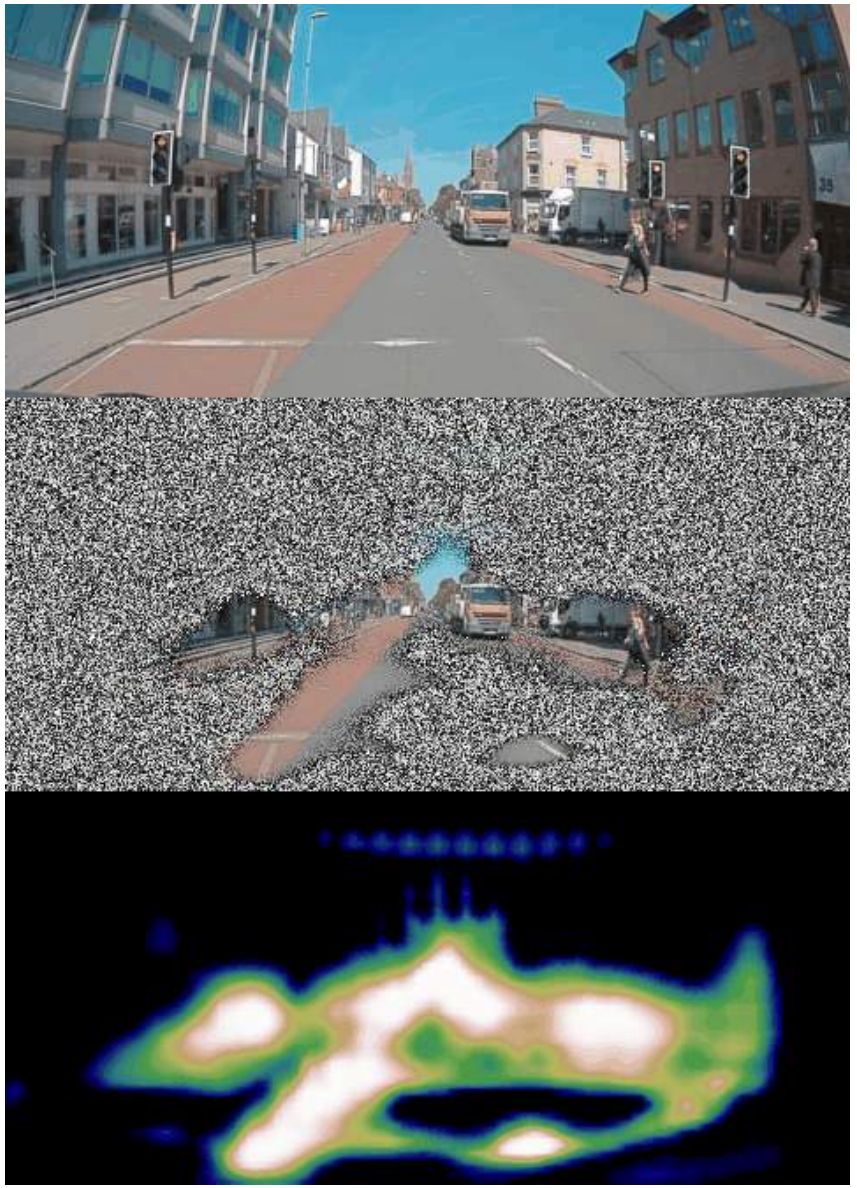
$$L = \lambda_1 |m(x)| + \lambda_2 |\nabla m(x)| + \lambda_3 L_0\left(f\left(x \cdot m(x)\right)\right) + \lambda_4 L_0\left(f\left(x \cdot \left(1 - m(x)\right)\right)\right)^{-\lambda_5}$$

Sparse saliency mask

Smooth saliency mask

Informative saliency mask

Uninformative inverse saliency mask



Dabkowski and Gal. "Real time image saliency for black box classifiers." NeurIPS. 2017.
Fong and Vedaldi. "Interpretable explanations of black boxes by meaningful perturbation." ICCV. 2017.

Model-Based Saliency

Suppose $f(\cdot)$ is our driving model and $m(\cdot)$ is our saliency model and $L(\cdot)$ is our loss function for the driving model and the operator $x \cdot m$ degrades the image with noise.

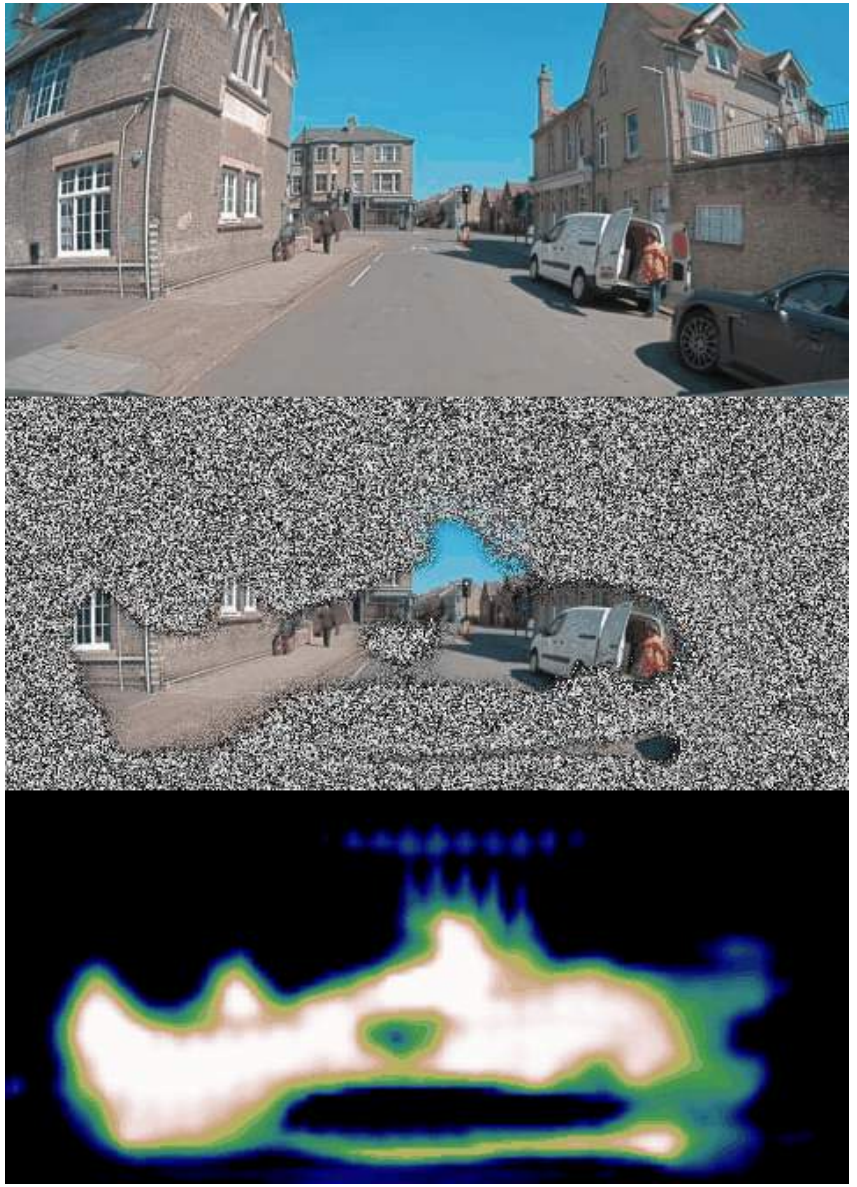
$$L = \lambda_1 |m(x)| + \lambda_2 |\nabla m(x)| + \lambda_3 L_0\left(f\left(x \cdot m(x)\right)\right) + \lambda_4 L_0\left(f\left(x \cdot \left(1 - m(x)\right)\right)\right)^{-\lambda_5}$$

Sparse saliency mask

Smooth saliency mask

Informative saliency mask

Uninformative inverse saliency mask



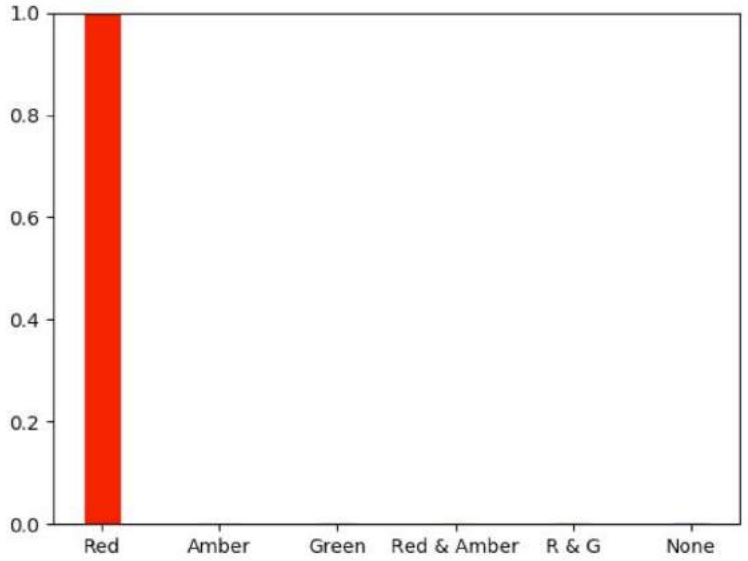
Dabkowski and Gal. "Real time image saliency for black box classifiers." NeurIPS. 2017.
Fong and Vedaldi. "Interpretable explanations of black boxes by meaningful perturbation." ICCV. 2017.

Inspecting the state for traffic light signal

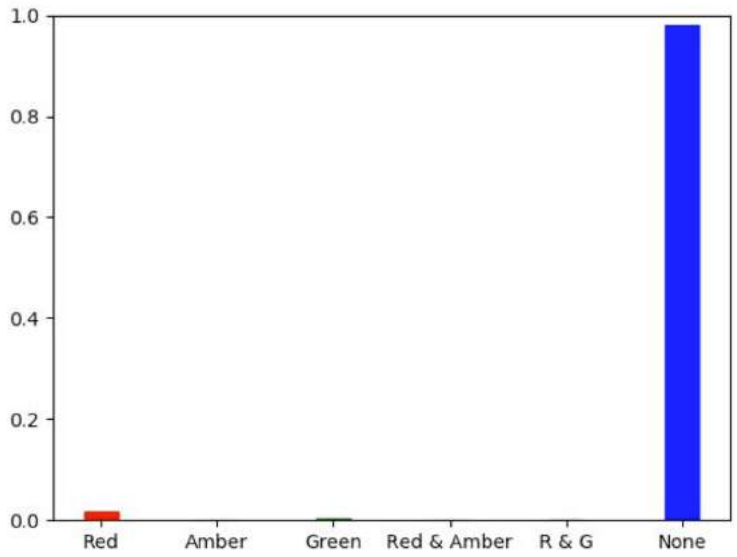
$x_{input} \rightarrow z \rightarrow y_{control}$

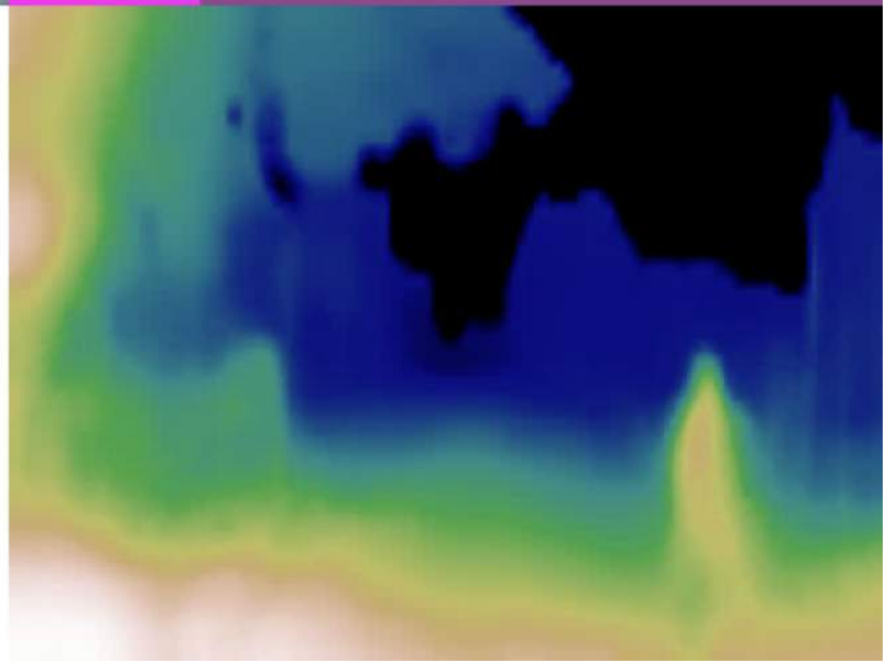
$z \downarrow y_{aux}$

Learn to decode the high dimensional state



Inspecting the state for traffic light signal, semantics and depth





Which metrics do we optimise?

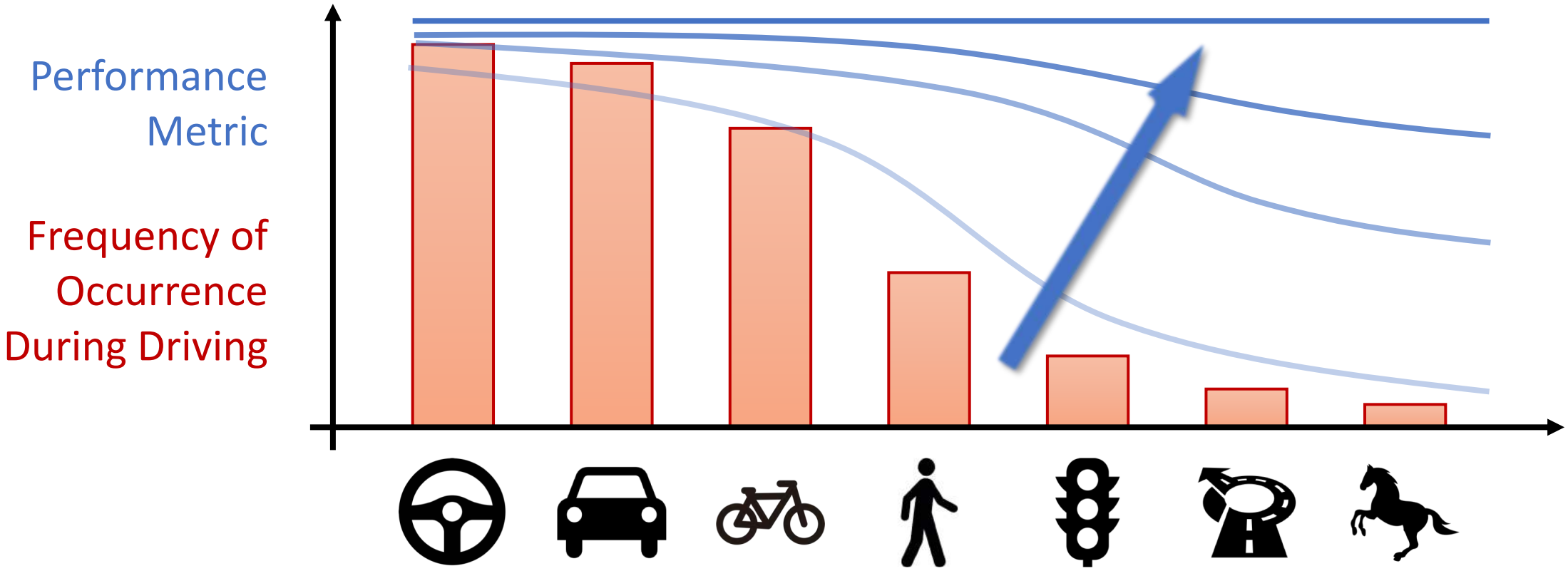
Must move away from component based verification

Improving individual components is no longer a proxy for improving system performance

- It assumes the interface between components is sufficient
- E.g. most KITTI metrics are at 90%+, does improving these metrics increase autonomous driving performance?

Mean Scenario Success

We need to consider complexity of autonomy, not just intervention metrics.



Alex Kendall © Wayve 2019

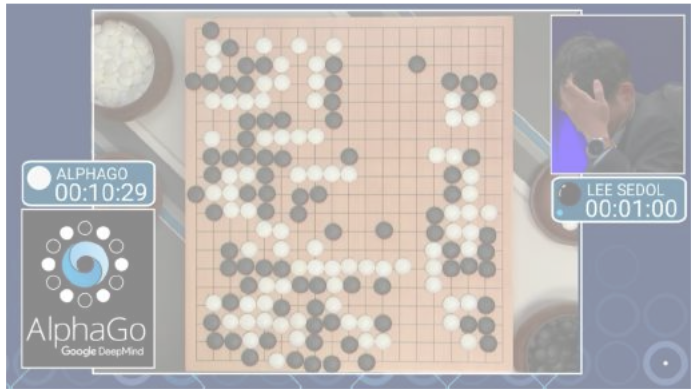
Conclusions

Alex Kendall © Wayve 2019

	Constrained Setting	Open World
Machine Learning	• Low engineering effort to create demo • Brittle representation • No performance guarantees	• Excels with increasing data and scale • Can learn powerful representations which generalise • Validate with statistical evidence
Human Design	• Possible to enumerate all scenarios • Analytical safety guarantees • Limited complexity	• Unachievable to identify all edge-cases • Too complex for safety guarantees • Requires extremely large engineering effort

Alex Kendall © Wayve 2019

Games like Go & DOTA



DeepMind OpenAI

- Incredibly difficult action space: long term strategy, cooperation
- Very basic state space, often discrete, fully observable and noise-free

Autonomous Driving

- Quite easy action space: stop, go, left, right motion primitives
- Super challenging state space: manifold of natural images!

This needs to be solved by the computer vision community!

A complete paradigm shift for AVs

- Low vehicle compute and sensor requirements
- Large training compute and data requirements
- Increased vehicle intelligence
- No reliance on HD-maps
- Ability to leverage simulation for training
- Abundance of open and interesting research questions!

Come work with our team wayve.ai/careers

WAYVE