Trustworthy AI Autonomy

M4-1 Certification and Digital Twin Generation

Ding Zhao

Assistant Professor

Carnegie Mellon University

Carnegie Mellon University

Safe AI Lab @CMU

Plan for today

- Importance of Certification/Evaluation
- Evaluation and test methods of AI autonomy
  - Naturalistic Field Operational Test (N-FOT) - Test on the public roads
  - Proving ground tests
  - Simulation/digital twins/augmented reality/meta universe
- Concept of Scenarios

FIRSTMILE

Autonomous Vehicle Landscape

Autonomous driving systems

Ike, cruise, tusimple, MOTOR AI, BLUE VISION, MONET, BEAR FLAG ROBOTICS, drive.ai, ENWAY, TTTech, ROADSTAR, auto.ai, PROMOTIVES, MOMENTA, Uber, ARGO, cortica, AIMOTIVE, WAYMO, DATASPEED, Gatik, ZOOX, OXBOTICA, BRAIQ, THE HI-TECH ROBOTIC SYSTEMZ LTD., PILOT, Horizon Robotics, Zong Mu, HOLOMATIC, AIVIS, HEEX, nuTonomy, FIVE AI, Apex.AI, DEEPSCALE, WAYVE, AURORA, POLYSYNC, pony.ai, ClearMotion, RIDE VISION, Brodmann, Valeo, SIGRA, plus.ai, lyft, RENOVO, AURORALABS, AUTONOMOS, divi, algolux, APOSTERA, VAYAVISION, BASELABS, MOBILEYE, TORC ROBOTICS, OTTO, APTIV, CYNGN

Data and simulation

RealSynth, eyesight, thinci, otonomo, Mighty Ai, ottomatika, SILEXICA, tass, SCAPE, foretellix, cognata, SAIPS, HUMANISING AUTONOMY, AAi, rFpro, SMARTMEUP, Applied Intuition, parallel domain, embotech, COGNITIVE PILOT, scale.ai, GEOSPIN, Spring, netradyne, AUTO DRIVE SOLUTIONS, DIBOTICS, MSC Software, LATENT LOGIC

Light detection, ranging and sensing

INNOVIZ, Velodyne LiDAR, Blickfeld, LUMINAR, cepton, QUANERGY, OUSTER, AEYE, PROPHESEE, ADASKY, HESAI, arbe ROBOTICS, Innovusion, TRILUMINA, BWV, visights, ibeo, oryx, TRIEYE, RFISEE, SEEVA, GUARDIAN, NEWSIGHT IMAGING, PHOTONICVISION, TOPOSENS, Brighter AI, blackmore, LeddarTech, XENOMATIX

Autonomous vehicle manufacturers

CLEARPATH, LUCID, EASY MILE, ZMP, Varden Labs, boxbot, ohmio, BYTON, PROTON, TESLA, NIO, 2getthere, SF MOTORS, E/NRIDE, local motors, LeSEE, marble, TeleRetail, XPENG, may mobility, nuro, WELTMEISTER, navya, UNITI, AURO, bossanova, brain corp, OPTIMUS RIDE

Map and location-based services

DEEPMAP, AEVA, ARTISENSE, lvl5, terraloupe, Ushr, atlatec, TELENAV, here, BAREWAYS, aipark, NAVENTIK

Apps and mobility infrastructure

Wunder, autofleet, RIDECELL, AUTONOMIC, Fleetonomy, rideOS, fline, bestmile, ESR LABS, INTEFRA, BlackBerry QNX

Connected car &V2X communication

Autotalks, alcan systems, METAWAVE, HUML, SMART MOBILE LABS, wiGate, ottopia, INRIX

Decoding the Autonomous Driving Landscape July 2019 | Firstmile | www.firstmile.de

Note: All firms shown are either currently or formerly VC / PE-backed

THE FUTURE OF TRANSPORTATION STACK

263 Self-Driving Car Startups

COMET LABS

SERVICES

ROUTE PLANNING
SPATIAL, CITIVIZ, waze, CommuteMaster, X, moovit, Citymapper

PARKING
Waypals, Cloud Parc, Aware, VoicePark, SpotHero, PEARL, seevoder, Parko, FenSens, LUXE, arkys

CAR HAILING + POOLING
TESLOOP, lyft, A, RideCell, 滴滴

OTHER: AFTERMARKET, REPAIR, RENTAL
CARVUE, upshift, WRENCH, PATROL, zipcar, Getaround, URGENTLY, Tantalum, bitkar

SAFETY & SECURITY

PHYSICAL CAR & DRIVER SAFETY +ACCIDENT DETECTION
drivebetter, splitsecnd, nexar, GUARDIAN Optical Technologies, phrame, LyfeLens, SMART AUTOLABS, tourmalinelabs, D, iproc, eLSys, FORESIGHT

EMOTION, FATIGUE & ALCOHOL DETECTION + DISTRACTION AVOIDANCE
CORTEX LABS, eyeris, ALMA care, C6, CARMA, saverone, R, SMART wheel, Radiomize, EASY, carvi

CYBERSECURITY
CYMOTIVE, cognomotiv, VisualThreat, ARGUS, Olympus Sky, Karamba Security, Trillium, Arilou

INTRUSION, TRACKING & RECOVERY
trackntop, nonda, gmt Connect, ROADEYE, caruma

IN-CAR INTELLIGENCE + ASSISTANCE

VEHICLE DIAGNOSTICS & PREDICTIVE MAINTENANCE + SENSOR-BASED VEHICLE SAFETY
TrueMotion, BRAIQ, IQ, metromile, LIGHT METRICS

PASSENGER-FOCUSED SENSORS (INCLUDING USAGE-BASED INSURANCE)
Cartics, Zobie, nauto, SONICLUE, PREDICTIVE SYSTEMS, P, NEBULA, engie

INFOTAINMENT + DISPLAY
link motion, HUDIFY, CARROBOT, PioninCar, swedspot, OverDrive, ROAV, FICONIC SOLUTIONS, RIGHTWARE, PELAGICORE, CLOUDCAR

PERSONAL / VOICE ASSISTANCE
navideck, AUTOMATIC, dash, GRIN, cariq

NAVIGATION ASSISTANCE + PEDESTRIAN ANALYSIS & COMMUNICATIONS
MAXRAY, PERCEPTIVE AUTOMATA, navdy, hudly

AUTONOMY

AUTOMATION SYSTEM
AutoX, drive.ai, POLYSYNC, OXBOTICA, oryx, ARGO AI, AIMOTIVE, PILOT, cruise, nuTonomy, vector.ai, cognivue, 中天安驰 AIDRIVING, BWV, MOMENTA

MAPPING, SIMULATION, & IMAGE RECOGNITION / ANNOTATION
lvl5, minds.ai, NOMOKO, Mapper.ai, CARMERA, TomTom, IMPROBABLE, UDACITY, DEEPMAP, understand.ai, Civil Maps

AUTONOMOUS VEHICLE MAKER + TOOLS
nuro, WAYMO, ZOOX, FARADAY FUTURE

INFRASTRUCTURE + CONNECTED CAR

SENSOR NETWORKING INFRASTRUCTURE (V2V, V2X) - LPWA, CELLULAR, WIFI
DIGIMONDO, U, SAVARI, ZiFiSense, EVVOS, Actility

CONNECTED CAR - DATA, PLATFORM, SOFTWARE
ACTIVESCALER, INRIX, SPARC, mobiliuz, carvoyant, vinli, otonomo, mobile assisted driving, bright box, VEHICLE DATA SCIENCE, mojio

FLEET + TRAFFIC MANAGEMENT
BESTMILE, Cloud Your Car, synovia, Fleetmatics, urban engines, Motionlogic, Travelai, Automile, Zendrive, GLYDEL, locoNav, Immense Simulations, MOTORWAY BUDDY, mentis services, FR, Brisk Synergies

OTA CAR SOFTWARE UPDATION + SMART PHONE ENABLEDTELEMATICS
FIXD, DRIVEBOT, VOYOMOTIVE, ZENE, arynga

INTELLIGENT MANUFACTURING

NEW/ADVANCED MATERIALS
AQUARIUS, NAWA TECHNOLOGIES, SIRRUS, NANOSTEEL

RAPID PROTOTYPING - 3D PRINTING, MODULARIZATION, OPEN SOURCE
DM, Carbon, nanocore, DIVERGENT, LM, OSVehicle, stratasys, VADER

ADVANCED / AUTOMATED ASSEMBLY LINE
Grabit, SYMBIO ROBOTICS, CLEARPATH ROBOTICS, RM ROBOMOTIVE, rethink robotics

MATERIAL CHARACTERIZATION & TESTING
SynTouch, SASTRA ROBOTICS, VIDI, OPTOFIDELITY

ONBOARD SENSORS

LOCATION - GIS, PRECISION POSITIONING, PATH PLANNING
swift NAVIGATION, radiosense, N, EXO TECHNOLOGIES

VISION / CAMERA
DEEP VISION, Carnegie Robotics, machines with vision, ROADSENSE, DEEPSCALE, Chronocam, VAYAVISION

LIDAR
face, INNOVIZ TECHNOLOGIES, PHANTOM INTELLIGENCE, VOYAGE, QUANERGY, AEYE, TRILUMINA, LeddarTech, Velodyne LiDAR, STROBE

RADAR
OCULii, ARTsys360, TOPOSENS, RFISEE, 6TH SENSE, SDS

SPECIALTY VEHICLES

2-WHEELERS
mobike, smarthalo, superpedestrian, Zagster

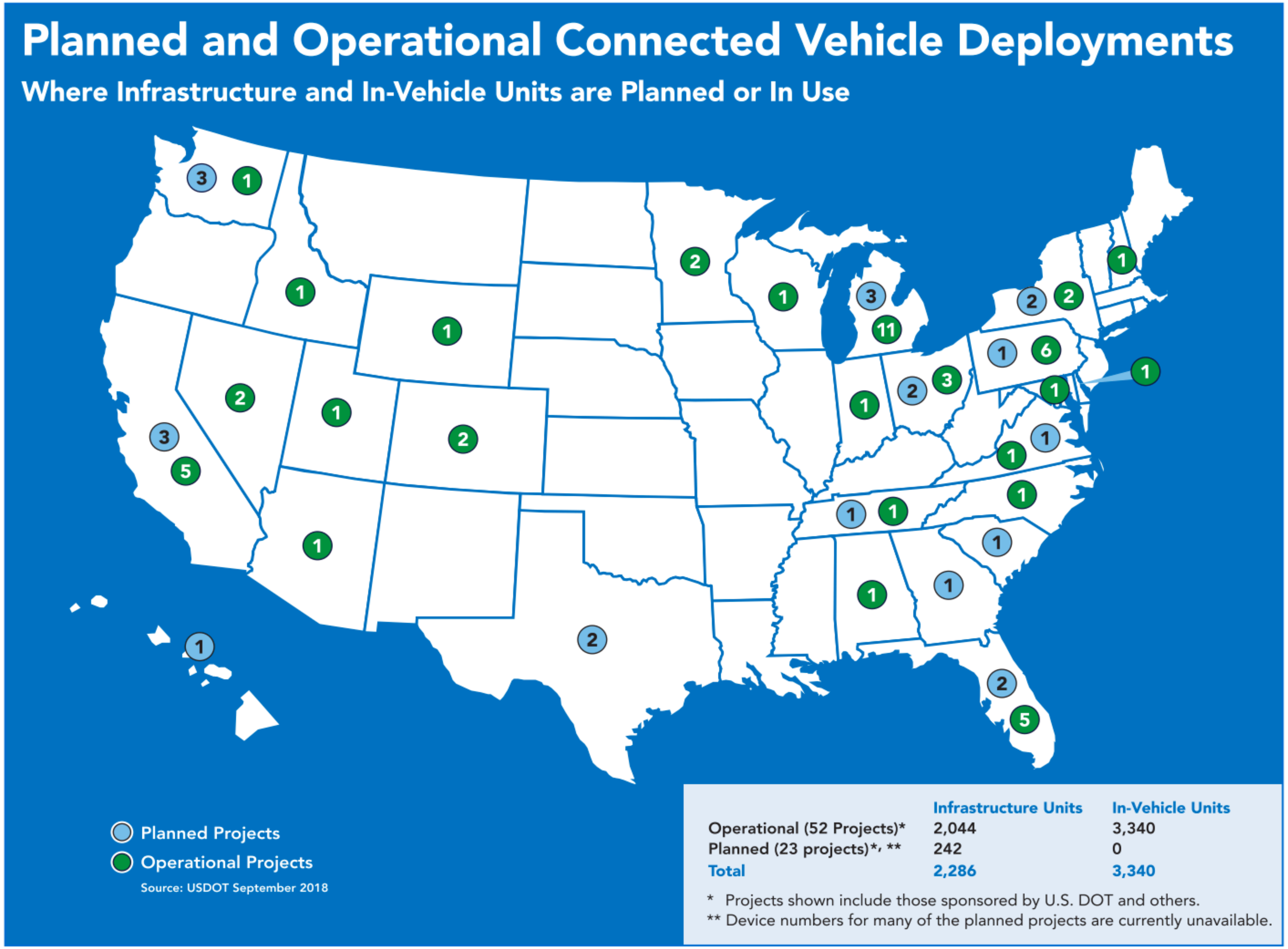
PUBLIC TRANSPORT
nauya, LM, Varden Labs, AURO, EASY MILE

TRUCKS / FREIGHT
JAYBRIDGE ROBOTICS, OTTO MOTORS, PELOTON, TESLA, FLEXPORT, TRUELITE TRACE, STARSKY ROBOTICS

FLIGHT
WRIGHT ELECTRIC, ZUNUM Aero, ZEE, JOBY, Kitty Hawk, LILIUM AVIATION, AEROMOBIL

OTHER: HYPERLOOP, PERSONAL MOBILITY
hyperloop one, ninebot asia, RYNO, WHILL

Regulations



New regulation

DEPARTMENT OF TRANSPORTATION

National Highway Traffic Safety Administration

49 CFR Part 571

Docket No. NHTSA-2021-0003

RIN 2127-AM06

Occupant Protection for Vehicles With

Automated Driving Systems

AGENCY: National Highway Traffic Safety Administration (NHTSA), Department of Transportation.

ACTION: Final rule.

SUMMARY: This final rule amends the occupant protection Federal motor vehicle safety standards (FMVSSs) to account for future vehicles that do not have the traditional manual controls associated with a human driver because they are equipped with Automated Driving Systems (ADS). This final rule makes clear that, despite their innovative designs, vehicles with ADS technology must continue to provide the same high levels of occupant protection that

AUTOS

U.S. clears way for truly driverless vehicles without steering wheels

PUBLISHED FRI, MAR 11 2022•10:31 AM EST | UPDATED FRI, MAR 11 2022•1:04 PM EST

Michael Wayland
@MIKEWAYLAND

SHARE

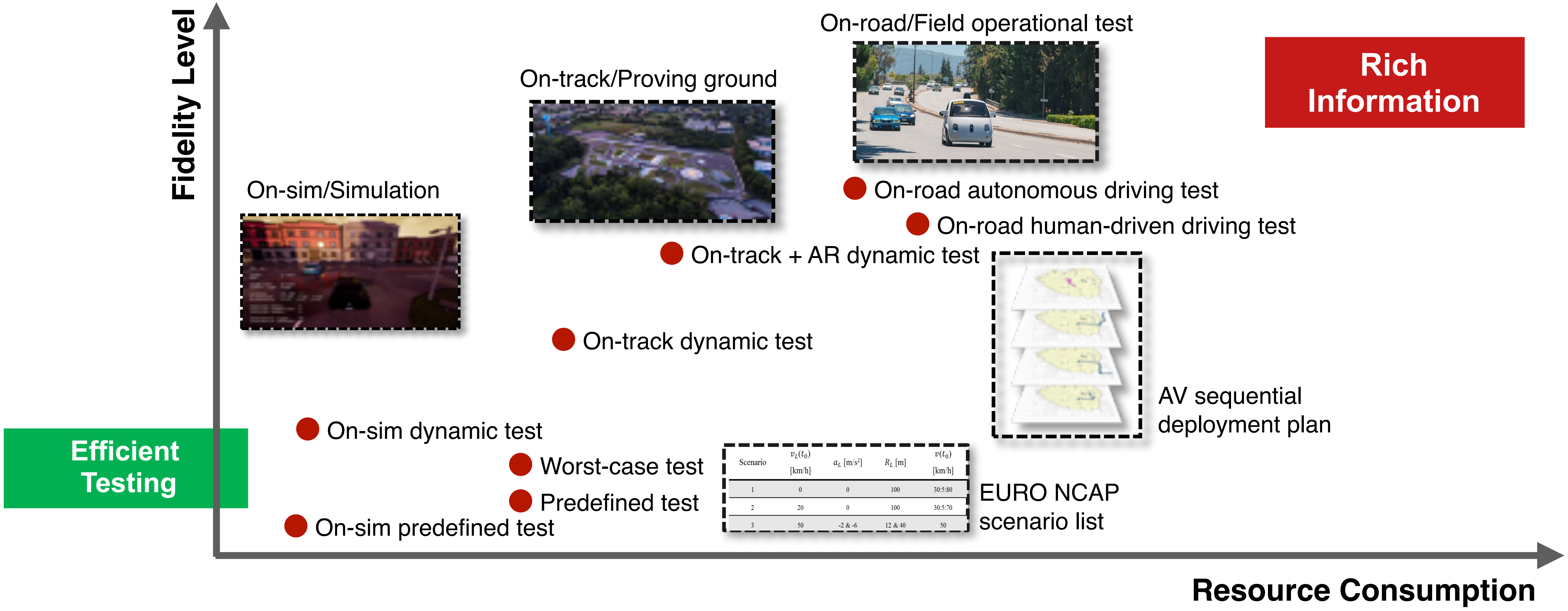
KEY POINTS

- Federal vehicle safety regulators have cleared the way for the production and deployment of truly driverless vehicles that do not include manual controls such steering wheels or pedals.
- The U.S. National Highway Traffic Safety Administration on Thursday issued final rules eliminating the need for highly automated and self-driving vehicles to need such controls.
- The new rule emphasizes such cars “must continue to provide the same high levels of occupant protection as current passenger vehicles.”

TV
Closing
UP NEXT
04:00 pr

https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-03/Final-Rule-Occupant-Protection-Amendment-Automated-Vehicles.pdf

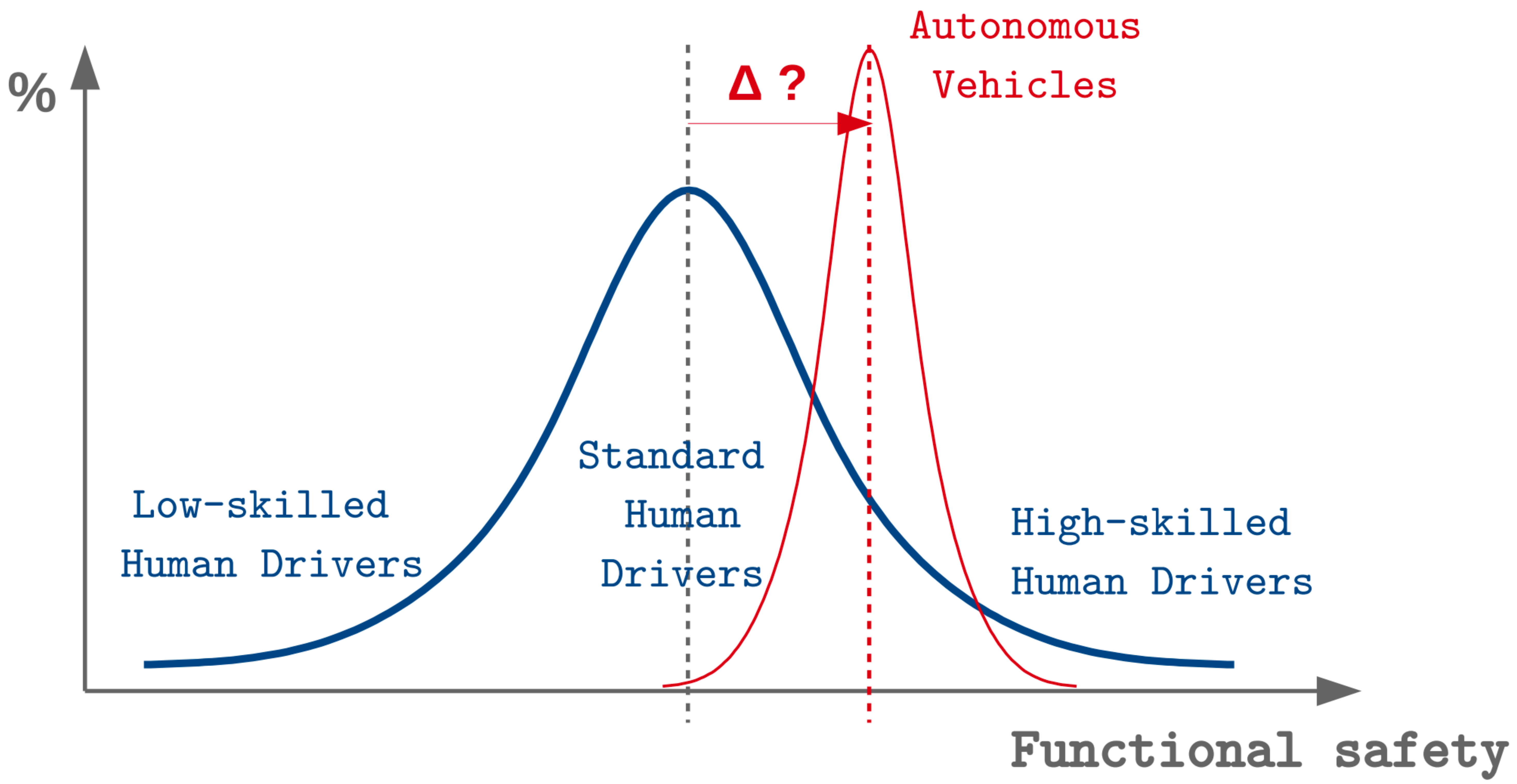
Certification and Evaluation methods



Scenario	$v_L(t_0)$ [km/h]	a_L [m/s^2]	R_L [m]	$v(t_0)$ [km/h]
1	0	0	100	30:5:80
2	20	0	100	30:5:70
3	50	-2 & -6	12 & 40	50

Huang "Synthesis of Different Autonomous Vehicles (AV) Test Approaches ", ITSC, 2018

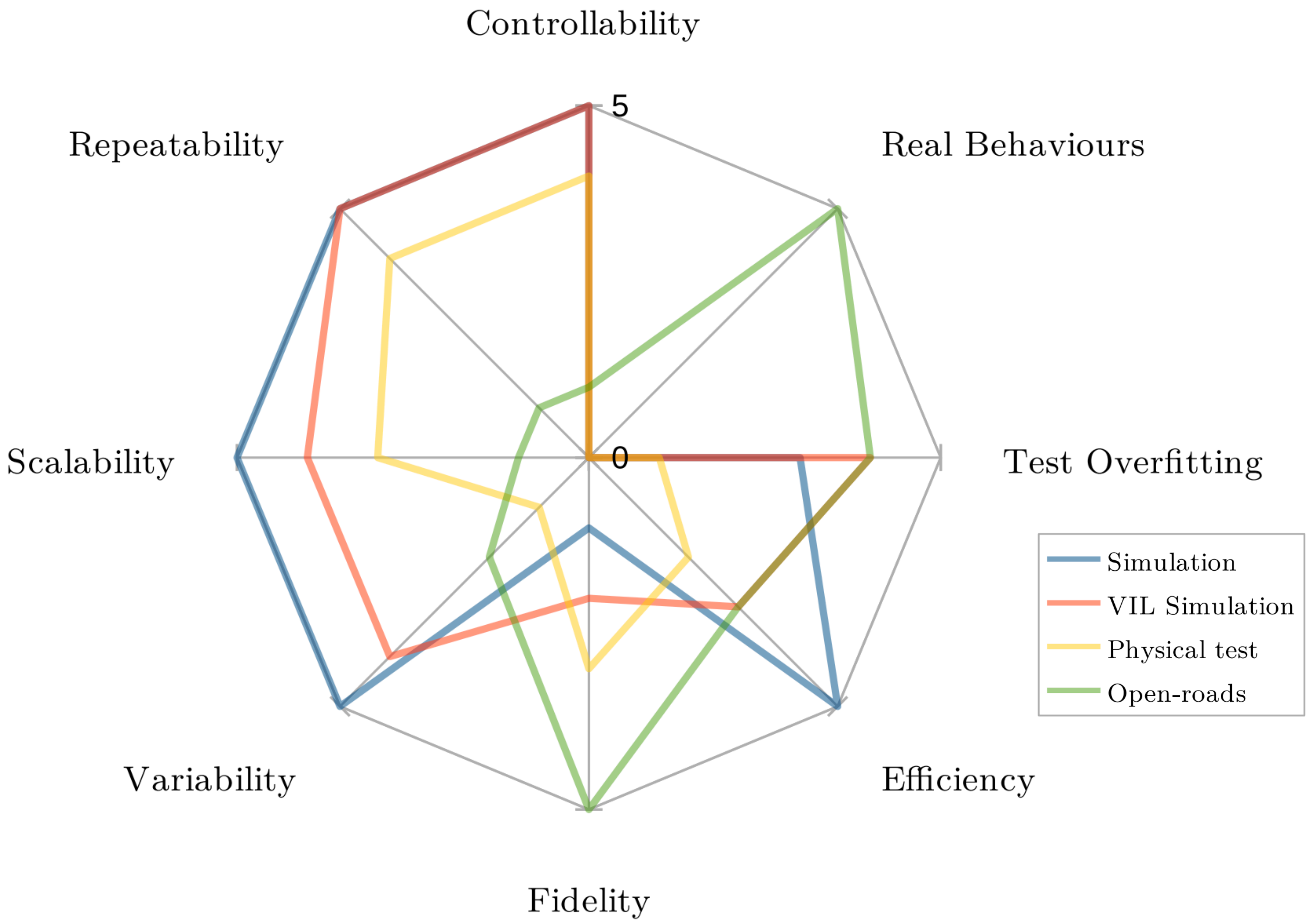
How safe is safe enough for AVs ?

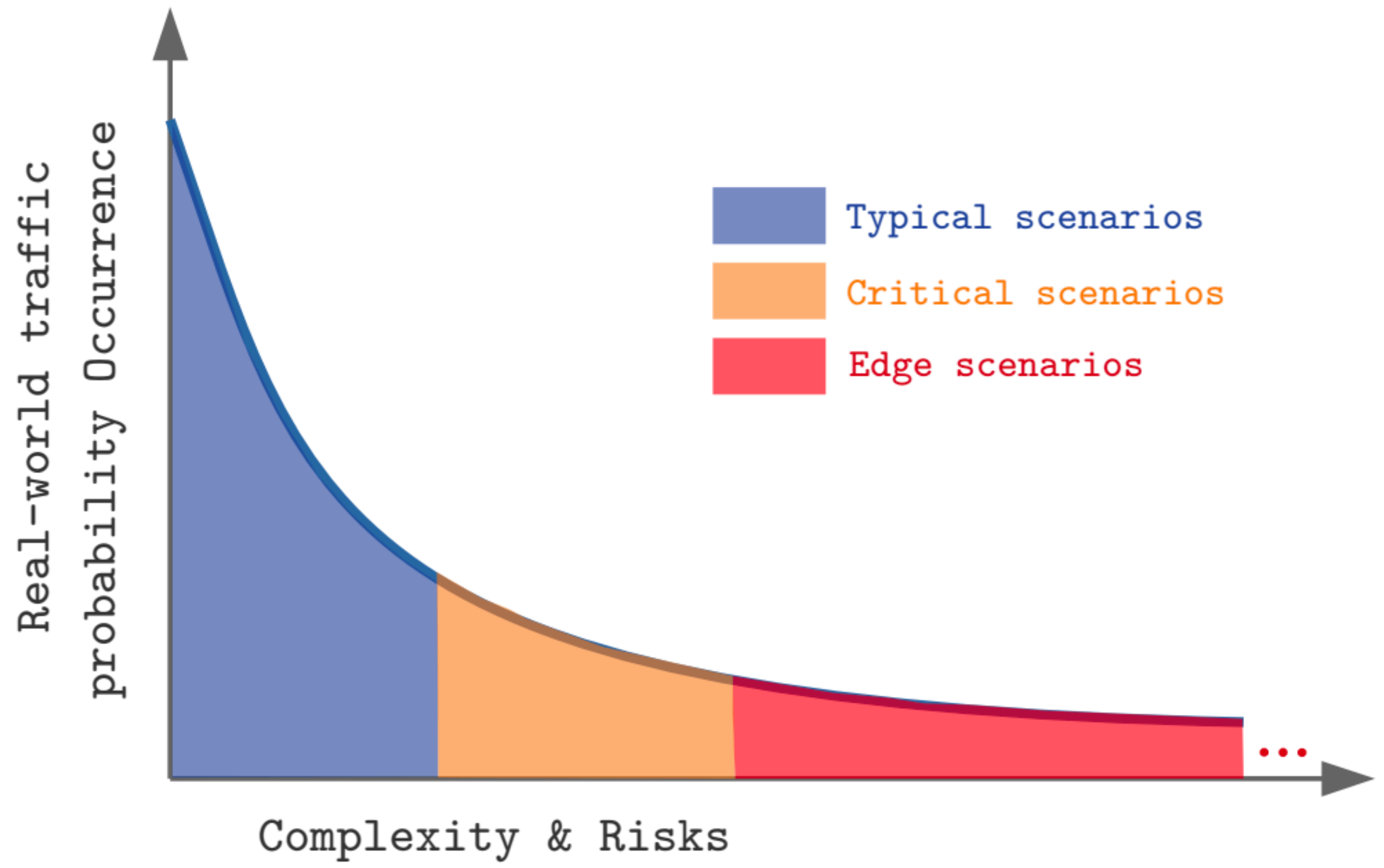


How to measure the safety: Simulation, Vehicle in-the-loop simulation (VIL), physical tests, Open-roads

 https://publications.jrc.ec.europa.eu/repository/handle/JRC127051

How safe is safe enough for AVs ?





Approaches	Typical	Critical	Edge
Simulation	✓	✓	✓
VIL Simulation	✓	✓	✓
Physical track		✓	
Open-roads	✓		

Table 13: Distribution of scenarios by testing approach.

https://publications.jrc.ec.europa.eu/repository/handle/JRC127051

Naturalistic Field Operational Tests (NFOT)

Waymo's self-driving car performing left-turn maneuver

https://storage.googleapis.com/sdc-prod/v1/safety-report/2020-09-waymo-safety-report.pdf

AV Deployment

Source: DMV.org (https://www.dmv.org/articles/top-5-cities-for-self-driving-boom)


San Jose, CA

Ann Arbor, MI


Boston, MA


Pittsburgh, PA


Austin, TX

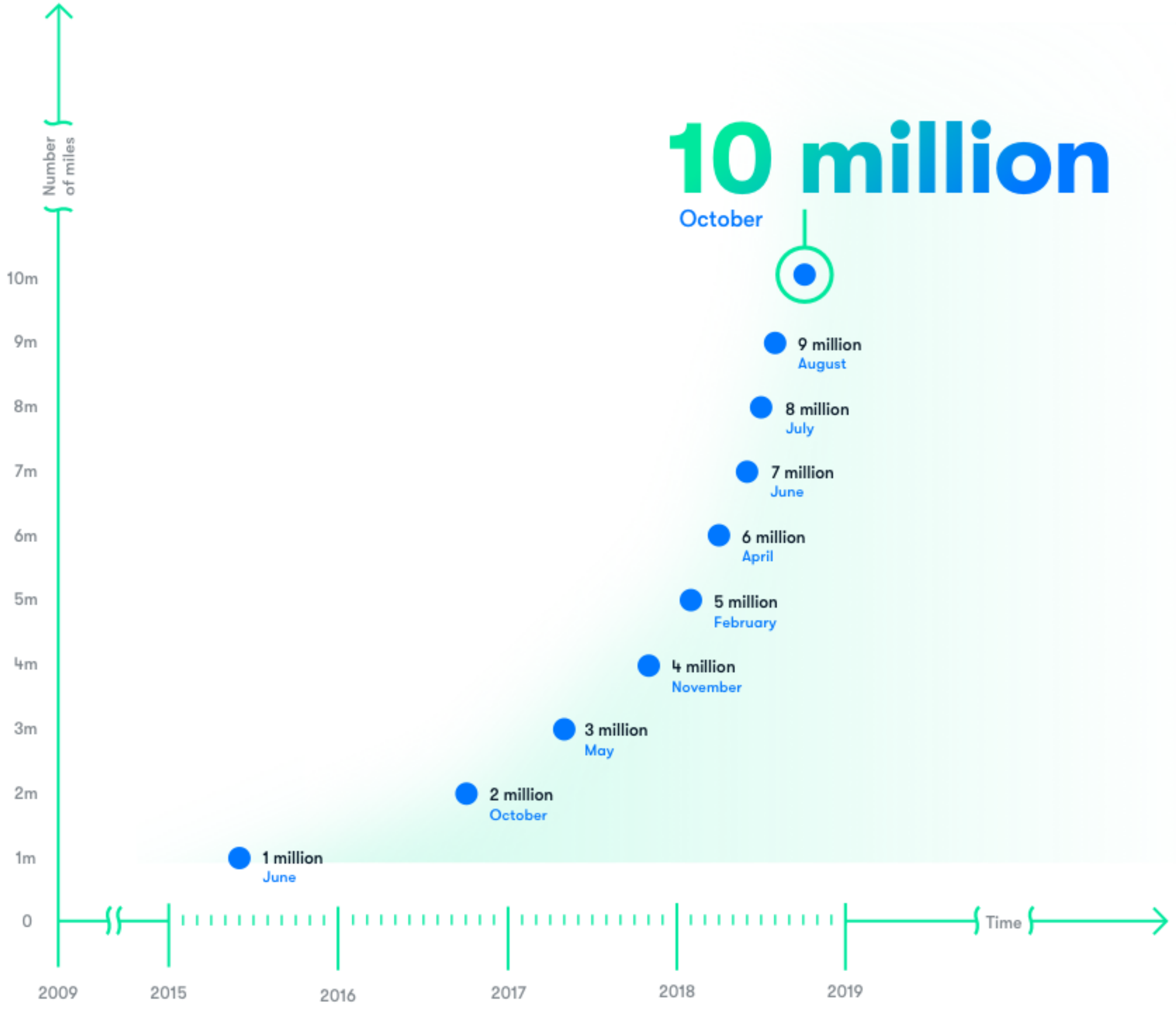
- AV testing in California

> “**Permit Holders**
> As of June 20, 2018, there are 56 Autonomous Vehicle Testing Permit holders.”

Source: DMV.ca.gov (https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing)

Naturalistic Field Operational Tests (NFOT)

[Waymo](#)



20 million miles and counting…

Forbes, January 2020

Driving Datasets for Autonomous Vehicles

Comparison to human baselines

Classification of Waymo-involved collisions (6.1 million AV miles driven)

			Waymo-involved collision-relevant contacts by ISO 26262 severity classification Actual & simulated event counts **(Totals in Bold)**					Human Crash Statistics (Non-Waymo Data)	
Row#	**Event type**	**Manner of Collision** ("Other" = non-Waymo vehicle)	**S0**	**S1 (no airbag deployment)**	**S1 (airbag deployment any vehicle)**	**S2**	**S3**	Collision % Contribution US *	Fatal Collision % Contribution U.S. (Maricopa Cnty, AZ) **
1	**Single Vehicle Events**	**Road Departure, Fixed object, Rollover**	**0**	**0**	**0**	**0**	**0**	20%	27% (21%)
2		**Striking a pedestrian/cyclist**	**0**	**0**	**0**	**0**	**0**	2%	33% (41%)
3		**Struck by pedestrian/cyclist**	**1 (actual) 2 (sim)**	**0**	**0**	**0**	**0**	<0.5%	1% (1%)
4	**Multiple Vehicle Events**	**Reversing**	**1 (actual) 1 (sim)**	**0**	**0**	**0**	**0**	1%	<0.1%
5		Other reversing, Waymo straight	*1 (actual) 1 (sim)*	*0*	*0*	*0*	*0*		
6		Waymo reversing, Other straight	*0*	*0*	*0*	*0*	*0*		
7		**Sideswipe (Same Direction)**	**1 (actual) 8 (sim)**	**1 (sim)**	**0**	**0**	**0**	11%	1% (1%)
8		Other lane change, Waymo straight	*1 (actual) 7 (sim)*	*0*	*0*	*0*	*0*		
9		Waymo lane change, Other straight	*1 (sim)*	*1 (sim)*	*0*	*0*	*0*		
10		**Head-on + Opposite Direction Sideswipe**	**0**	**0**	**1 (sim)**	**0**	**0**	5%	9% (7%)
11		**Rear End**	**11 (actual) 1 (sim)**	**1 (actual) 1 (sim)**	**2 (actual)**	**0**	**0**	34%	5% (5%)
12		Other striking, Waymo struck (stopped)	*8 (actual)*	*0*	*0*	*0*	*0*		
13		Other striking, Waymo struck (slower)	*2 (actual)*	*1 (actual)*	*1 (actual)*	*0*	*0*		
14		Other striking, Waymo struck (decelerating)	*1 (actual)*	*1 (sim)*	*1 (actual)†*	*0*	*0*		
15		Waymo striking, Other struck (stopped)	*0*	*0*	*0*	*0*	*0*		
16		Waymo striking, Other struck (slower)	*0*	*0*	*0*	*0*	*0*		
17		Waymo striking, Other struck (decelerating)	*1 (sim)*	*0*	*0*	*0*	*0*		
18		**Angled**	**4 (sim)**	**6 (sim)**	**1 (actual) 4 (sim)**	**0**	**0**	27%	24% (24%)
19		Same direction - Other turns across Waymo straight travel	*0*	*2 (sim)*	*0*	*0*	*0*		
20		Same direction - Other turns into Waymo straight travel	*3 (sim)*	*0*	*2 (sim)*	*0*	*0*		
21		Opposite direction - Other turns across Waymo straight travel	*0*	*0*	*1 (sim)*	*0*	*0*		
22		Opposite direction - Other turns into Waymo straight travel	*0*	*0*	*1 (sim)*	*0*	*0*		
23		Straight crossing paths	*0*	*1 (sim)*	*1 (actual)*	*0*	*0*		
24		Same direction - Waymo turns across other straight travel	*1 (sim)*	*3 (sim)*	*0*	*0*	*0*		
25		Same direction - Waymo turns into other straight travel	*0*	*0*	*0*	*0*	*0*		
26		Opposite direction - Waymo turns across other straight travel	*0*	*0*	*0*	*0*	*0*		
27		Opposite direction - Waymo turns into other straight travel	*0*	*0*	*0*	*0*	*0*		
28		**Total**	**14 (actual) 16 (sim)**	**1 (actual) 8 (sim)**	**3 (actual) 5 (sim)**	**0**	**0**	100%	100% (100%)

*CRSS 2016-2018, Urban area, ≤ 45 mph roadways
**FARS 2016-2018, Urban area, ≤ 45 mph roadways
† denotes sole collision in driverless operation (without human operator present)

Waymo-involved collision-relevant contacts by ISO 26262 severity classification Actual & simulated event counts **(Totals in Bold)**				
S0	**S1 (no airbag deployment)**	**S1 (airbag deployment any vehicle)**	**S2**	**S3**
14 (actual) 16 (sim)	**1 (actual) 8 (sim)**	**3 (actual) 5 (sim)**	**0**	**0**

A total of 18 dangerous situations observed during data collection and 29 situations during simulation

Schwall, Matthew, et al. "Waymo Public Road Safety Performance Data." arXiv preprint arXiv:2011.00038 (2020).

Proving grounds tests

- Proving grounds provide a physical semi-controllable environment to test AI, e.g. autonomous vehicles (AV)
- Example of AV proving ground facilities:
    - Mcity (UM)
    - The Castle (Waymo)
    - ALMONO (Uber)
    - American Center for Mobility
    - SMART Transportation Research Center (US DOT)
    - Kcity (South Korea)

ALMONO (42 acres)

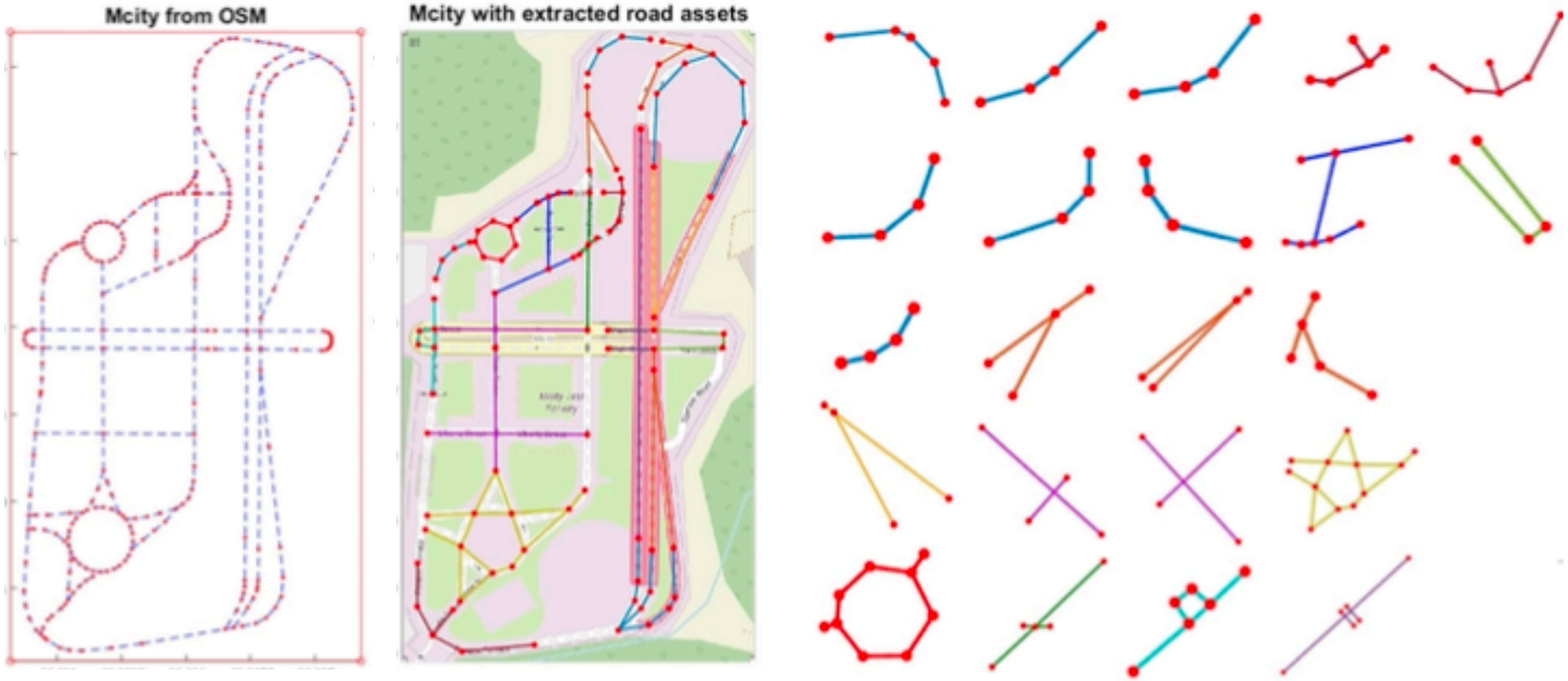
Mcity (32 acres)

Kcity (88 acres)

Rui Chen, Mansur Arief, Weiyang Zhang, and Ding Zhao. "How to Evaluate Proving Grounds for Self-Driving? A Quantitative Approach." *IEEE Transactions on Intelligent Transportation Systems* (2020).

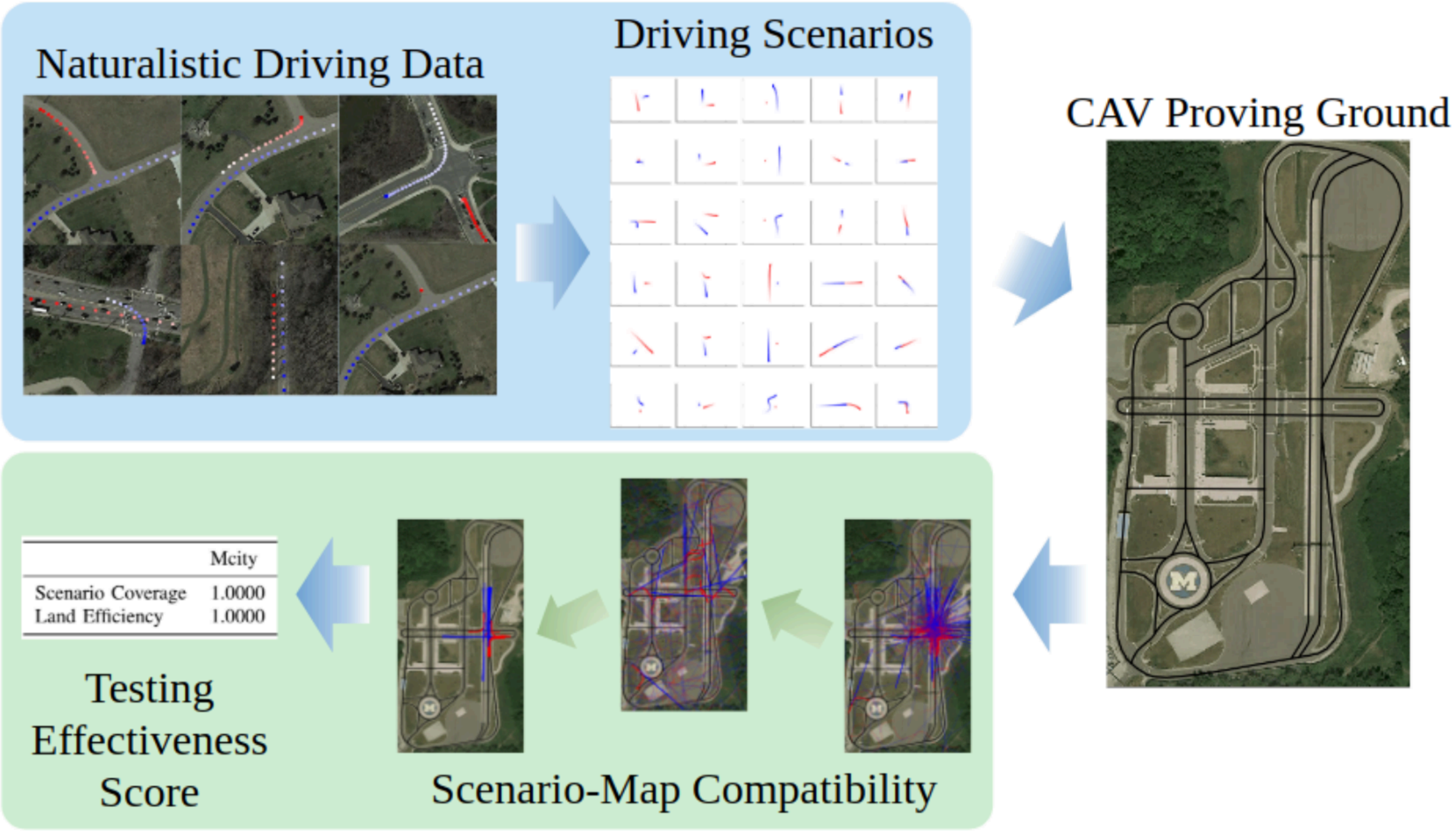
Proving grounds

- Geometric based test scenario generations



Proving grounds

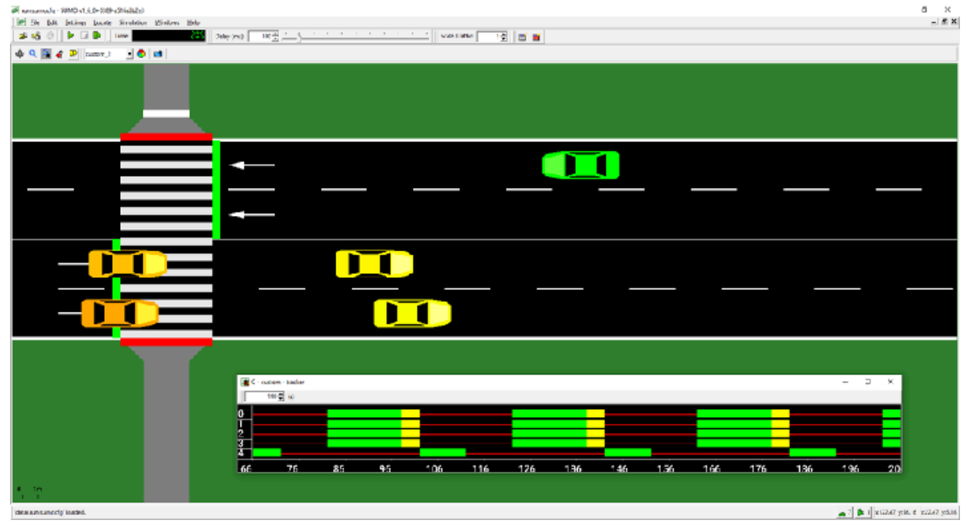
- Data-driven test scenario generations
- Key steps:
  - Extract driving scenarios from driving database
  - Extract proving ground map geometries and assets
  - Optimize the scenario-map compatibility



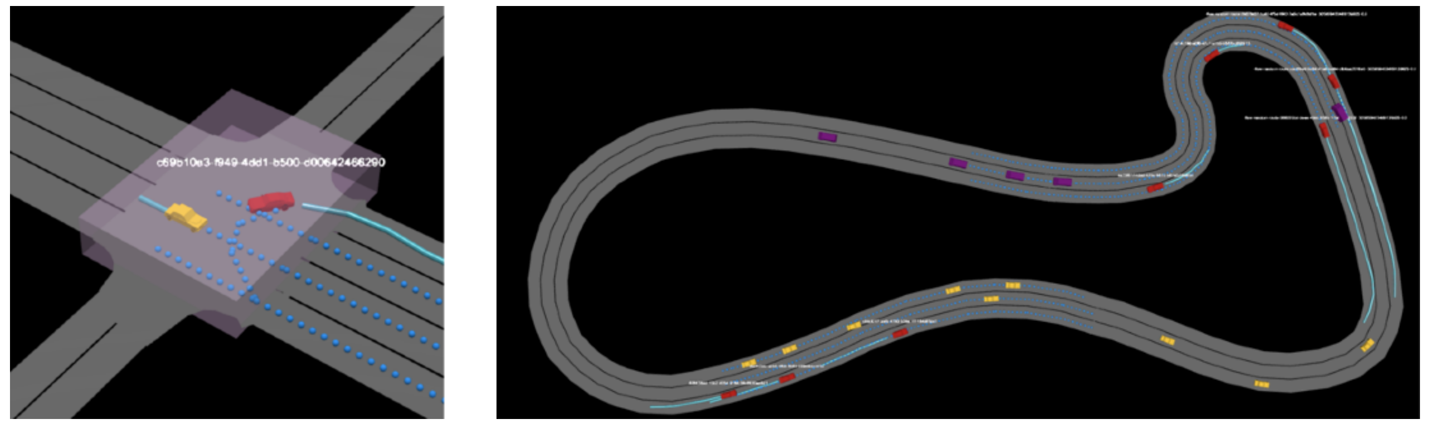
Rui Chen, Mansur Arief, Weiyang Zhang, and Ding Zhao. "How to Evaluate Proving Grounds for Self-Driving? A Quantitative Approach." *IEEE Transactions on Intelligent Transportation Systems* (2020).

Simulations

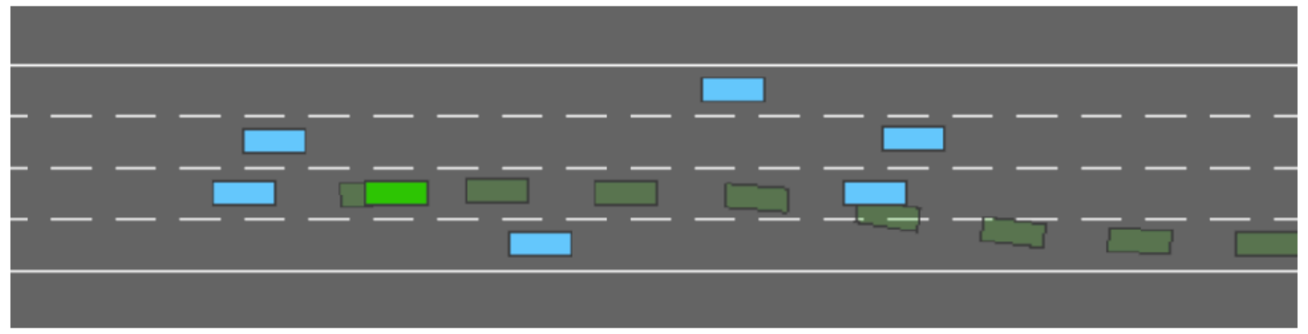
- Simulations allow fast and fully-controlled testing for AI algorithms
- Simulation-based testing is often done at various fidelity level



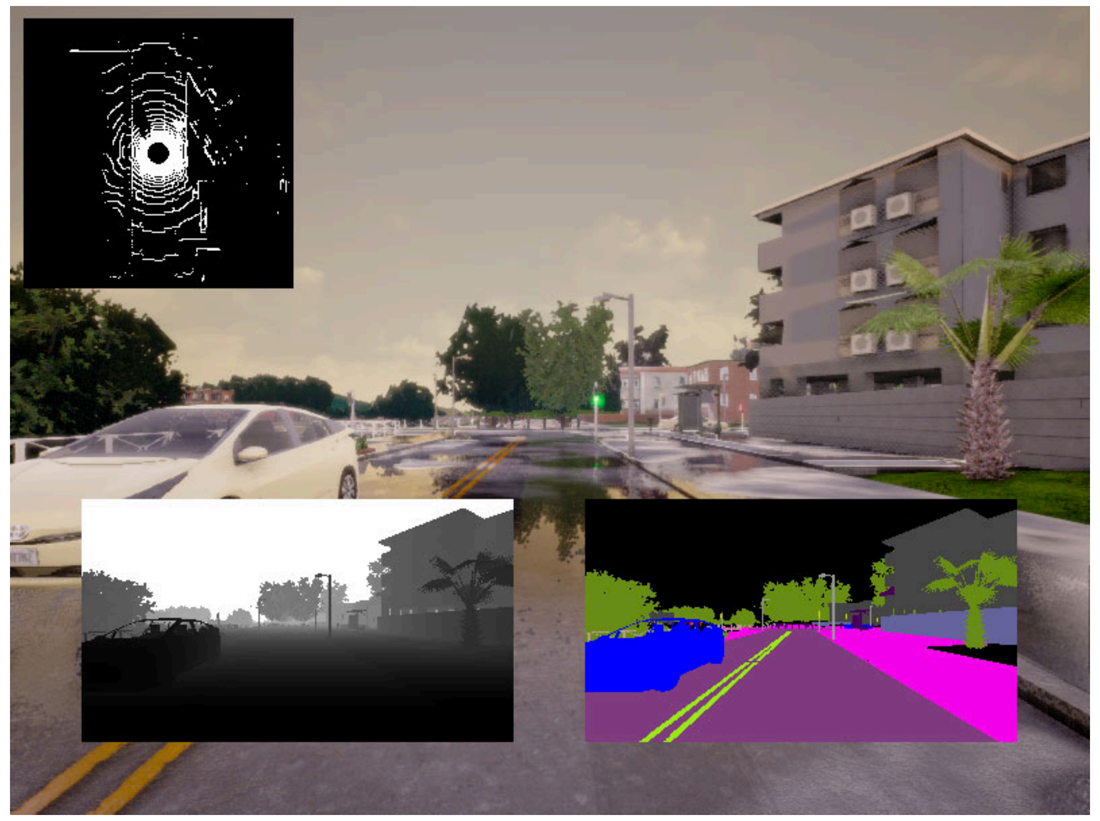
Simulation for Urban Mobility (SUMO)



Scalable Multi-Agent Reinforcement Learning Training School (SMARTS)



Highway Gym Environment (highway-env)



CARLA simulation

WeBot for Automobiles

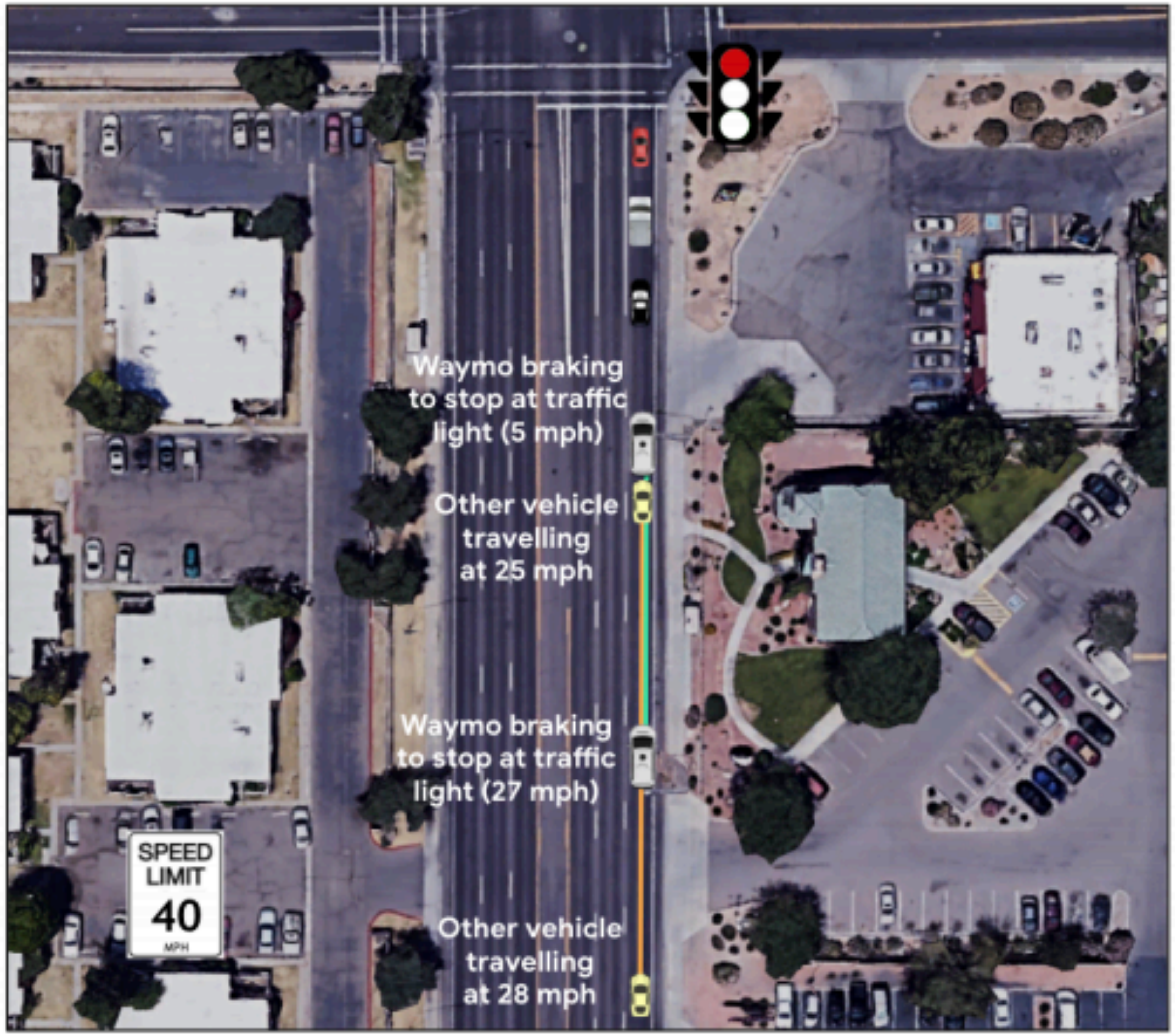
PreScan

Uber ATG simulation platform

Waymo simulated collisions

Head-on collisions

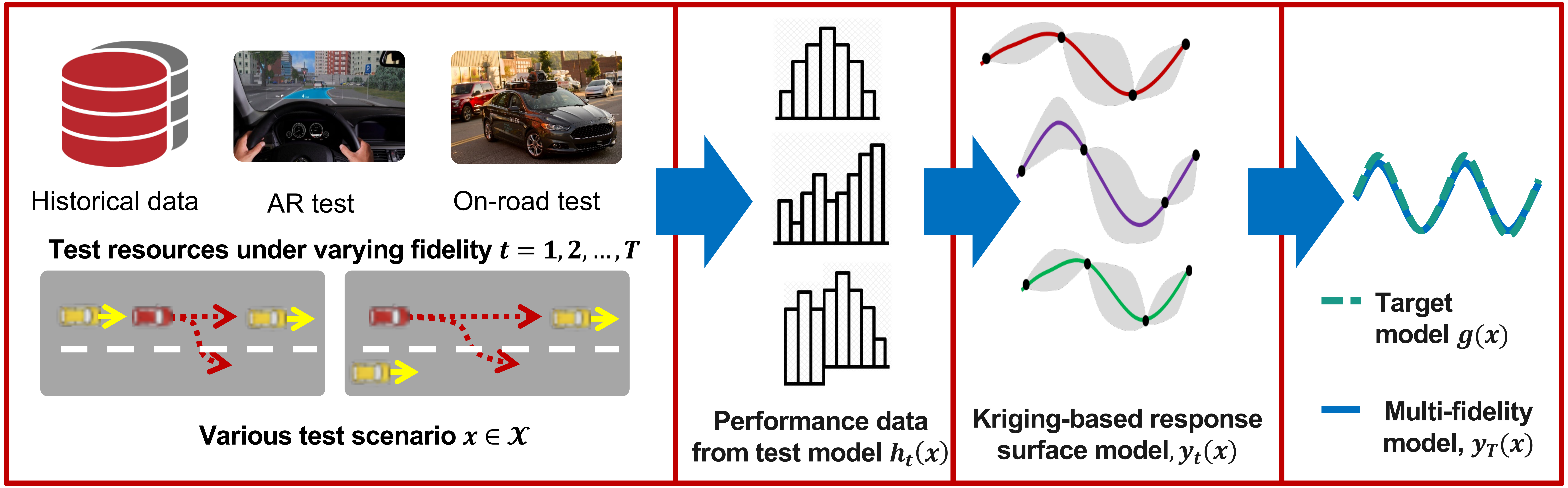
Rear-end collisions



Schwall, Matthew, et al. "Waymo Public Road Safety Performance Data." arXiv preprint arXiv:2011.00038 (2020).

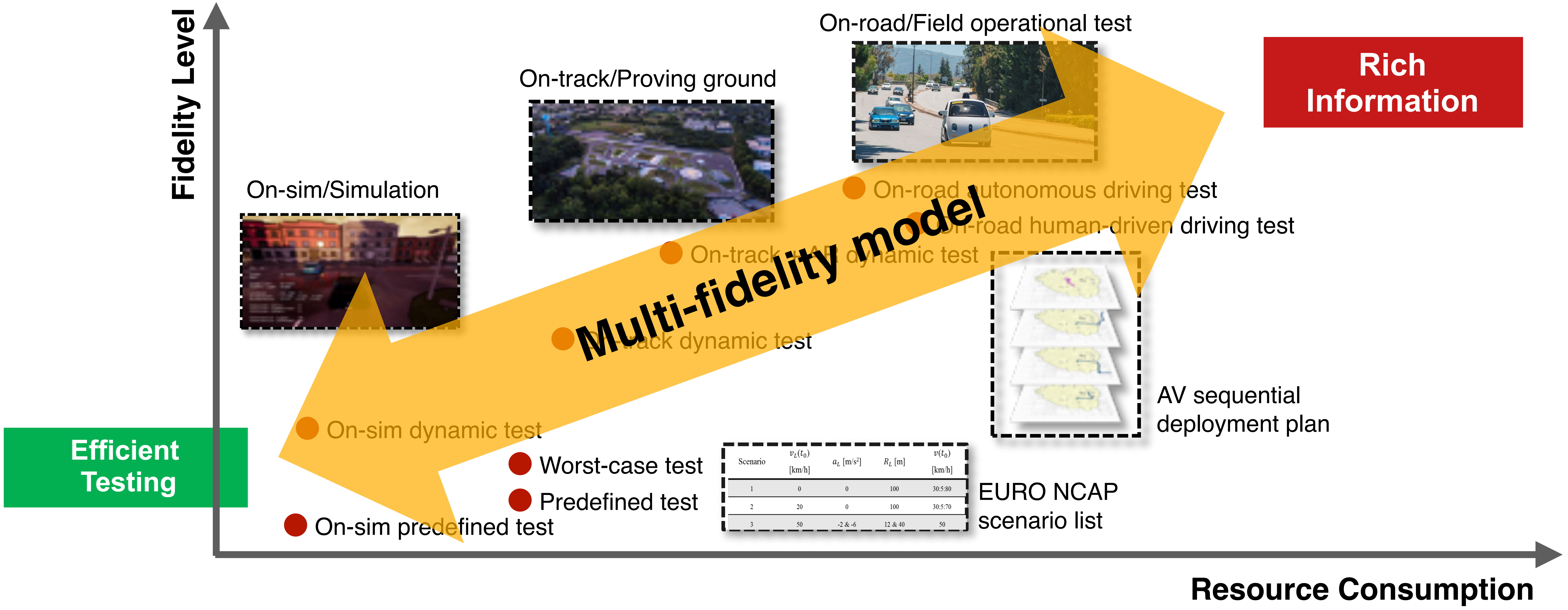
Synthesis tests

- Multi-fidelity models (e.g. Gaussian processes) are promising to synthesize information among various testing modes



Huang "Synthesis of Different Autonomous Vehicles (AV) Test Approaches", ITSC, 2018

Evaluation and test methods for AI autonomy



Scenario	$v_L(t_0)$ [km/h]	a_L [m/s^2]	R_L [m]	$v(t_0)$ [km/h]
1	0	0	100	30:5:80
2	20	0	100	30:5:70
3	50	-2 & -6	12 & 40	50

Huang "Synthesis of Different Autonomous Vehicles (AV) Test Approaches", ITSC, 2018

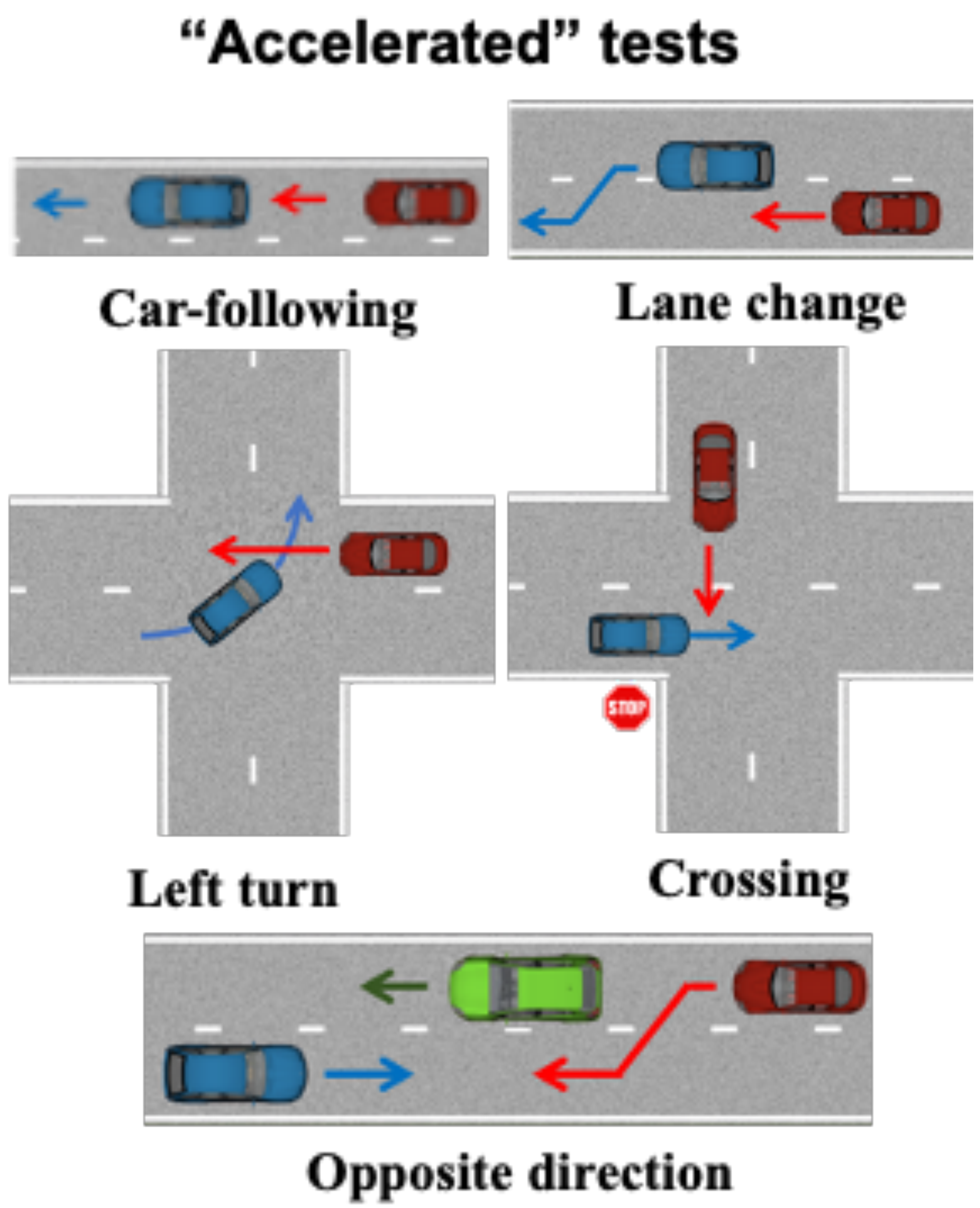
Methods

- Traditional ways to identify scenarios
- Data-based Scenario Generation
- Adversarial Scenario Generation
- Knowledge-based Scenario Generation

Traditional ways to identify scenarios

Table 5. Priority V2V Pre-Crash Scenarios

No	Pre-Crash Scenario	Group	Cost	FYL
1	SCP @ non signal	Junction Crossing	20.4%	20.7%
2	LTAP/OD	LTAP/OD	15.1%	15.3%
3	Rear-end/LVS	Rear-End	14.8%	14.0%
4	Opposite direction/no maneuver	Opposite Direction	14.7%	15.1%
5	Rear-end/LVD	Rear-End	6.1%	5.8%
6	Rear-end/LVM	Rear-End	5.1%	5.1%
7	Changing lanes/same direction	Lane Change	4.2%	4.3%
8	Turning/same direction	Lane Change	3.1%	3.0%
9	Opposite direction/maneuver	Opposite Direction	1.7%	1.8%
10	Drifting/same direction	Lane Change	1.7%	1.8%
Total			86.9%	87.0%



W. G. Najm, S. Toma, J. Brewer, "Depiction of Priority Light-Vehicle Pre-Crash Scenarios for Safety Applications Based on Vehicle-to-Vehicle Communications" (DOT HS 811 732, 2013).

Limitation

- Scenarios manually selected by human may not be able to take the advantage of the big data
- Human and AVs may have different critical scenarios

EURO NCAP AEB

Scenario	$v_L(t_0)$ [km/h]	a_L [m/s^2]	R_L [m]	$v(t_0)$ [km/h]
1	0	0	100	30:5:80
2	20	0	100	30:5:70
3	50	-2 & -6	12 & 40	50

Static

Moving

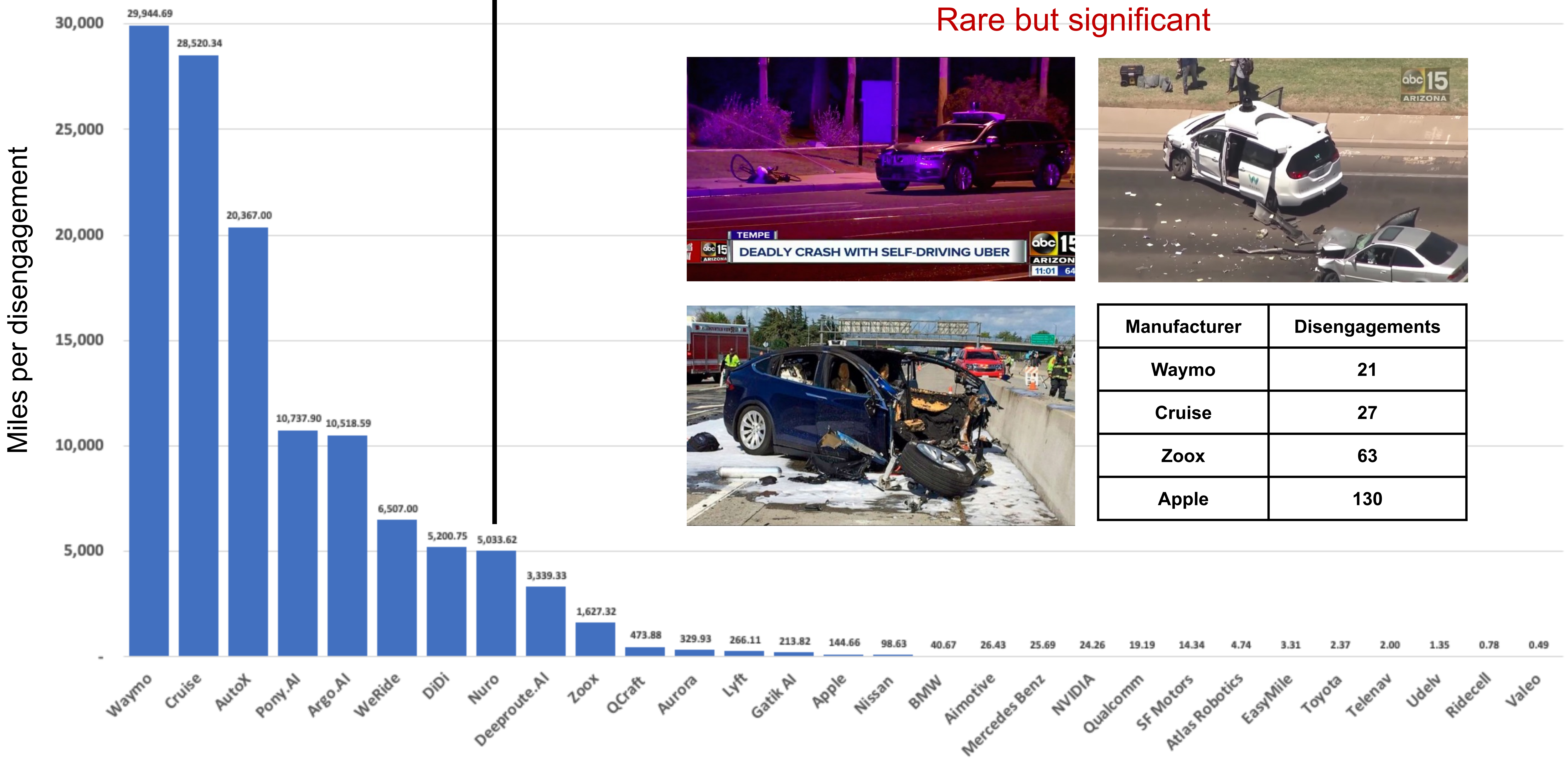
Braking

Realistic safety-critical scenario generation

- Opportunities:
  - Cheaper data access
  - More powerful computational facilities
  - Better machine learning algorithms
- Challenges:
  - Data sparsity /imbalance/rarity
  - Multi-modes
  - Dynamic long-horizon temporal decision making
  - High dimensional sensing input

Why safety-critical scenarios?

Safety?



Rare but significant

Manufacturer	Disengagements
Waymo	21
Cruise	27
Zoox	63
Apple	130

A new stage of evaluating autonomous vehicles: safety-critical scenarios

* *Data source: California Department of Motor Vehicle disengagement report 2020*

How to get safety-critical scenarios?

Collect from real-world road test

- Tremendously rare
- Expensive

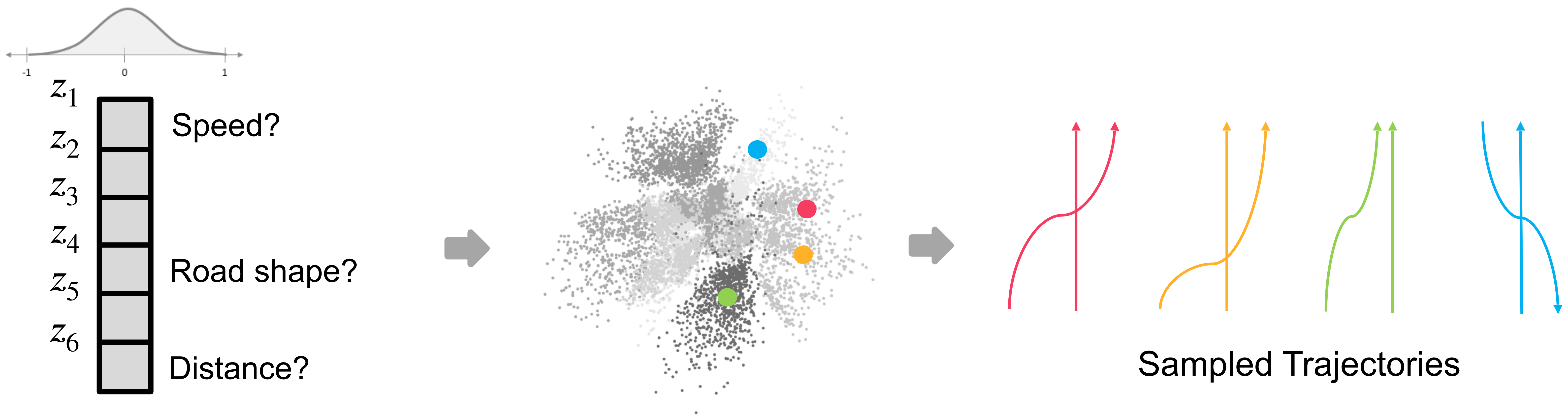
Collect from rule-based simulators

- Lack of diversity
- Not realistic

What's the expectation of a good scenario generator ?

Reality, Adaptability, Controllability, Efficiency, Diversity

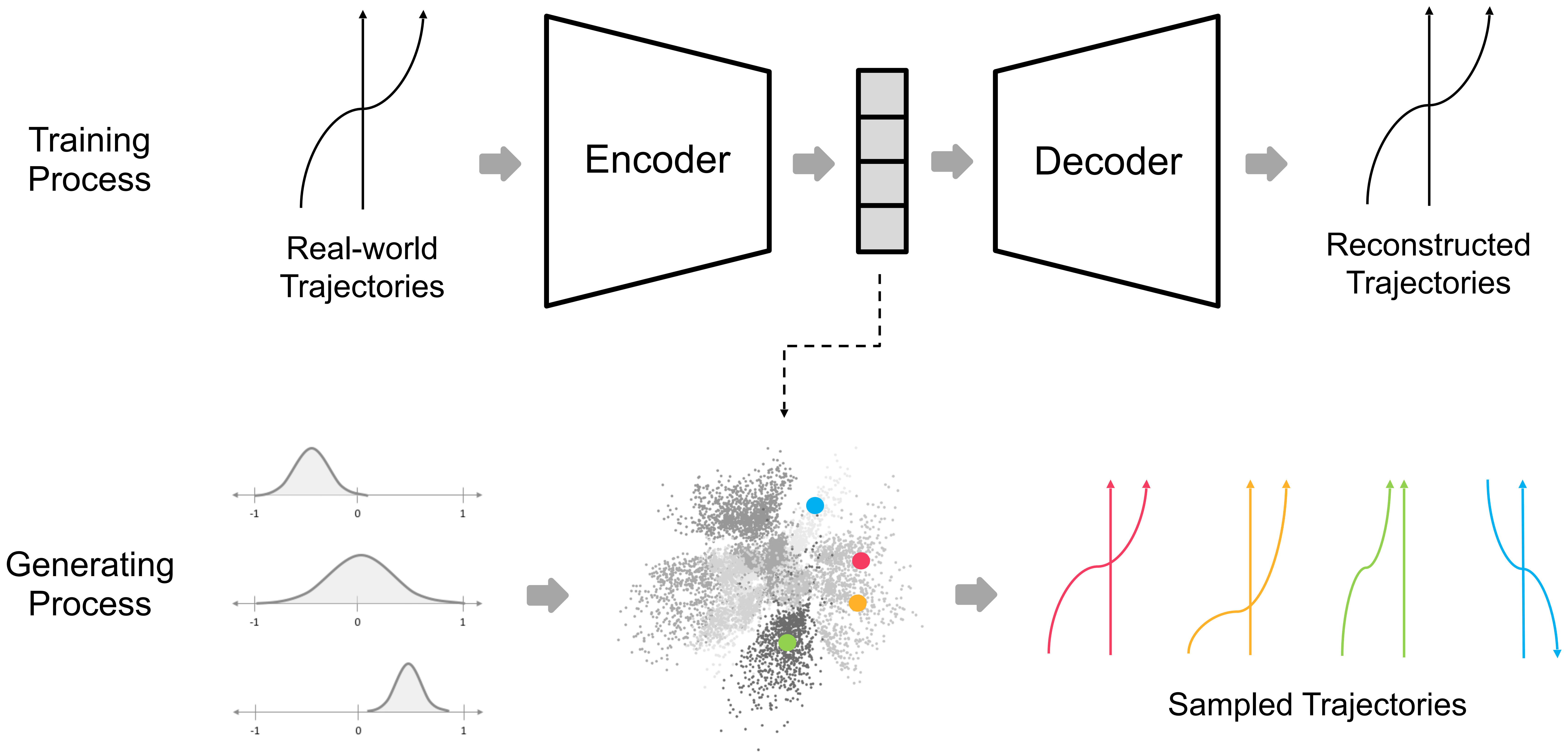
Data-based Scenario Generation



- What does each latent variable mean? How to get the trajectories we want?
- Safety-critical data is still rare in the latent space.

W. Ding, W. Wang, D. Zhao, A new multi-vehicle trajectory generator to simulate vehicle-to-vehicle encounters, ICRA 2019

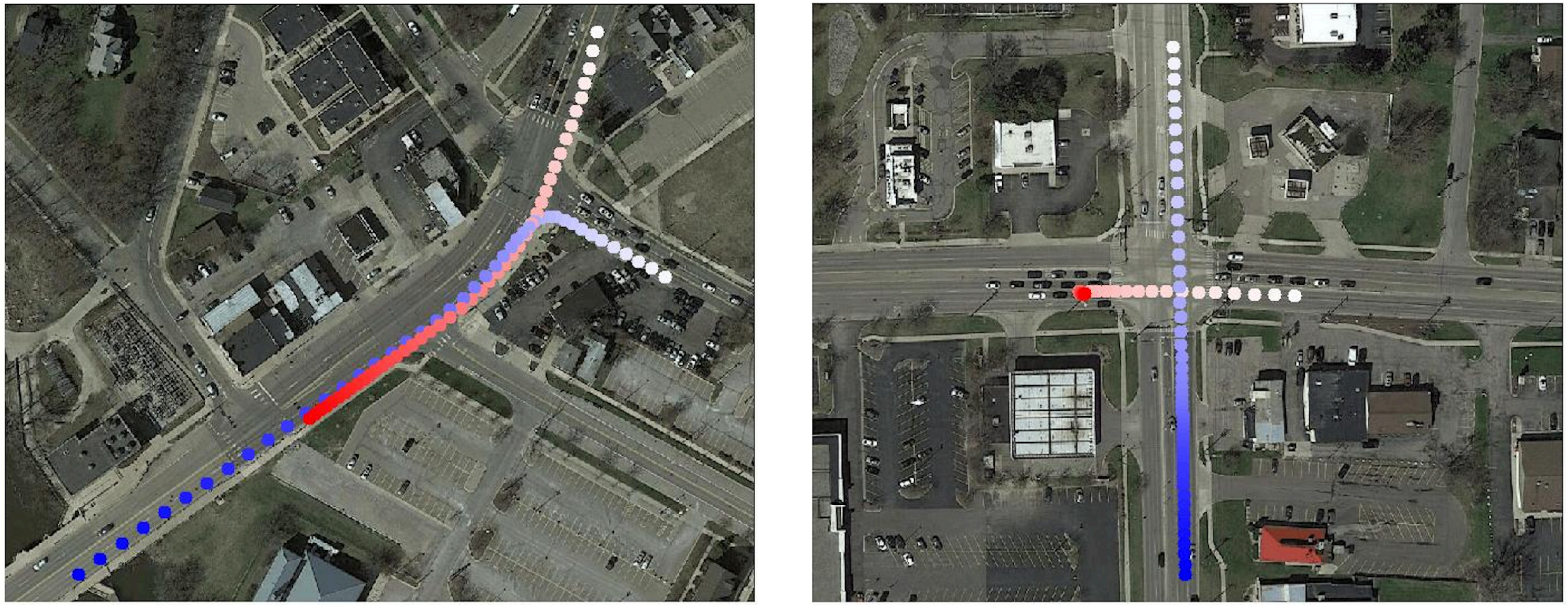
Data-based Scenario Generation



Randomly sample from the learned latent space

W. Ding, W. Wang, D. Zhao, A new multi-vehicle trajectory generator to simulate vehicle-to-vehicle encounters, ICRA 2019

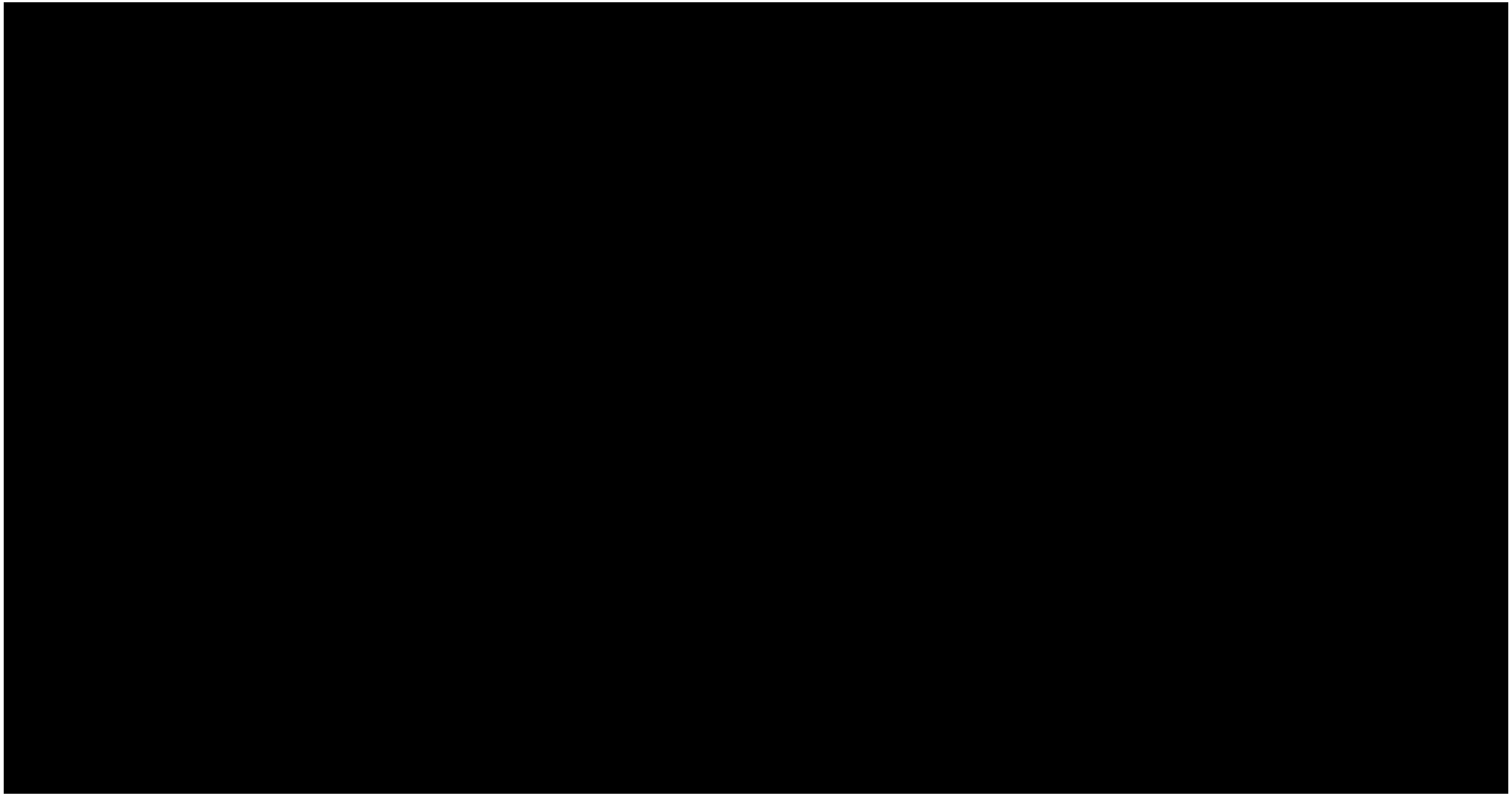
Data-based Scenario Generation



Assume encounter trajectories are formed by two parts: road shape (style), risk level (content)

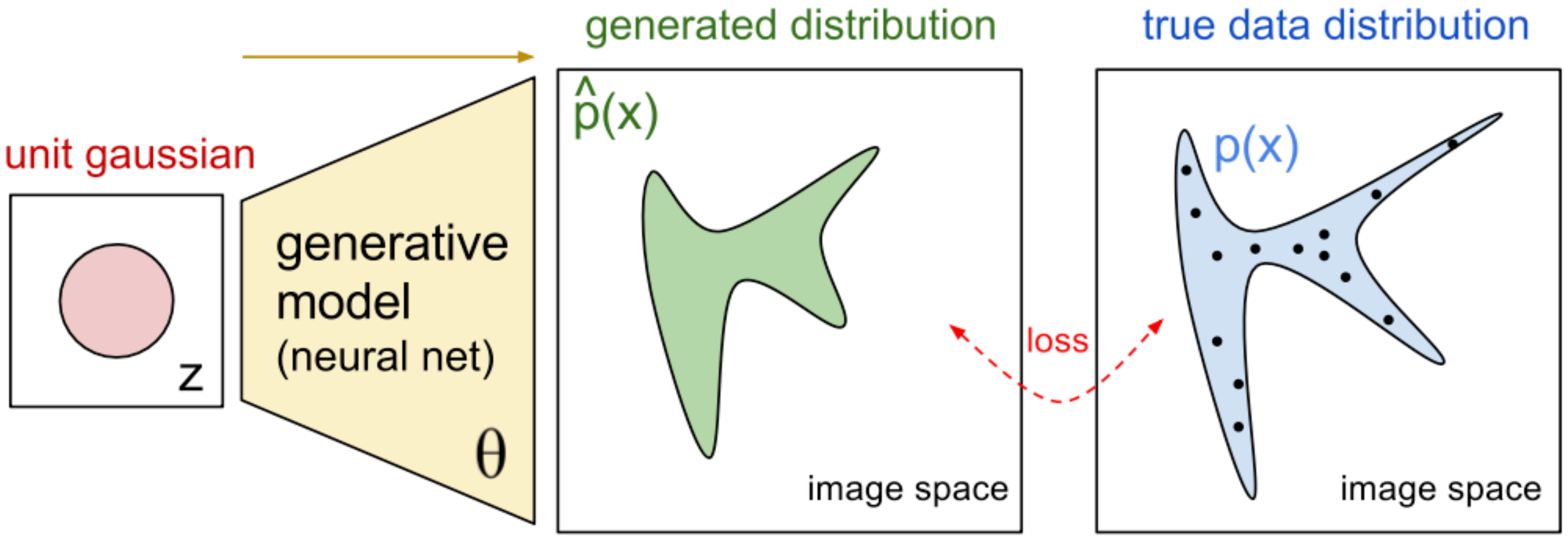
W. Ding, M. Xu, D. Zhao, CMTS: Conditional Multiple Trajectory Synthesizer for Generating Safety-critical Driving Scenarios, ICRA 2020

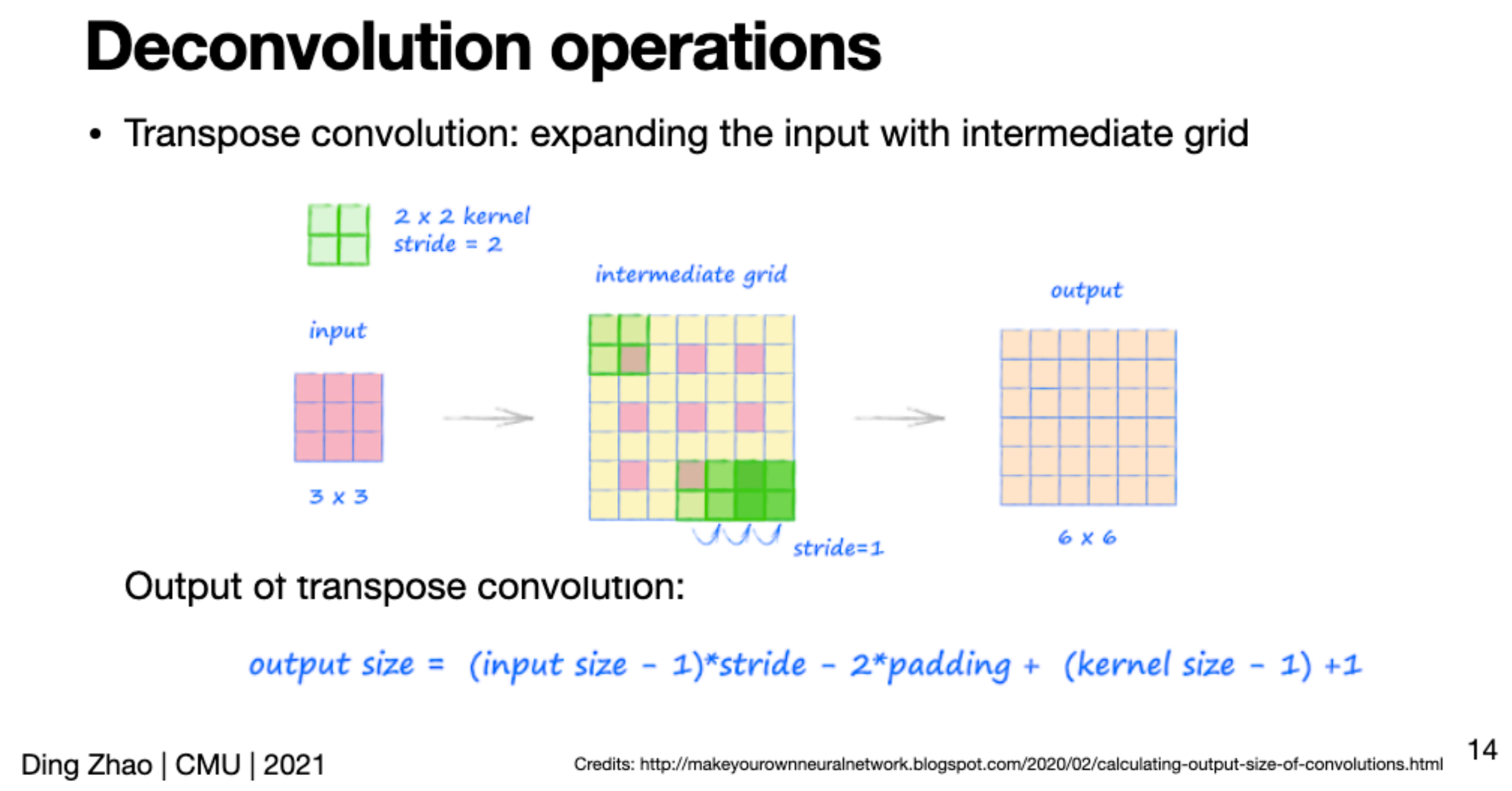
Driving scenario generation



Ding W, Chen B, Li B, et al. Multimodal Safety-Critical Scenarios Generation for Decision-Making Algorithms Evaluation. IEEE Robotics and Automation Letters, 2021

Generative models





Deconvolution operations

- Transpose convolution: expanding the input with intermediate grid

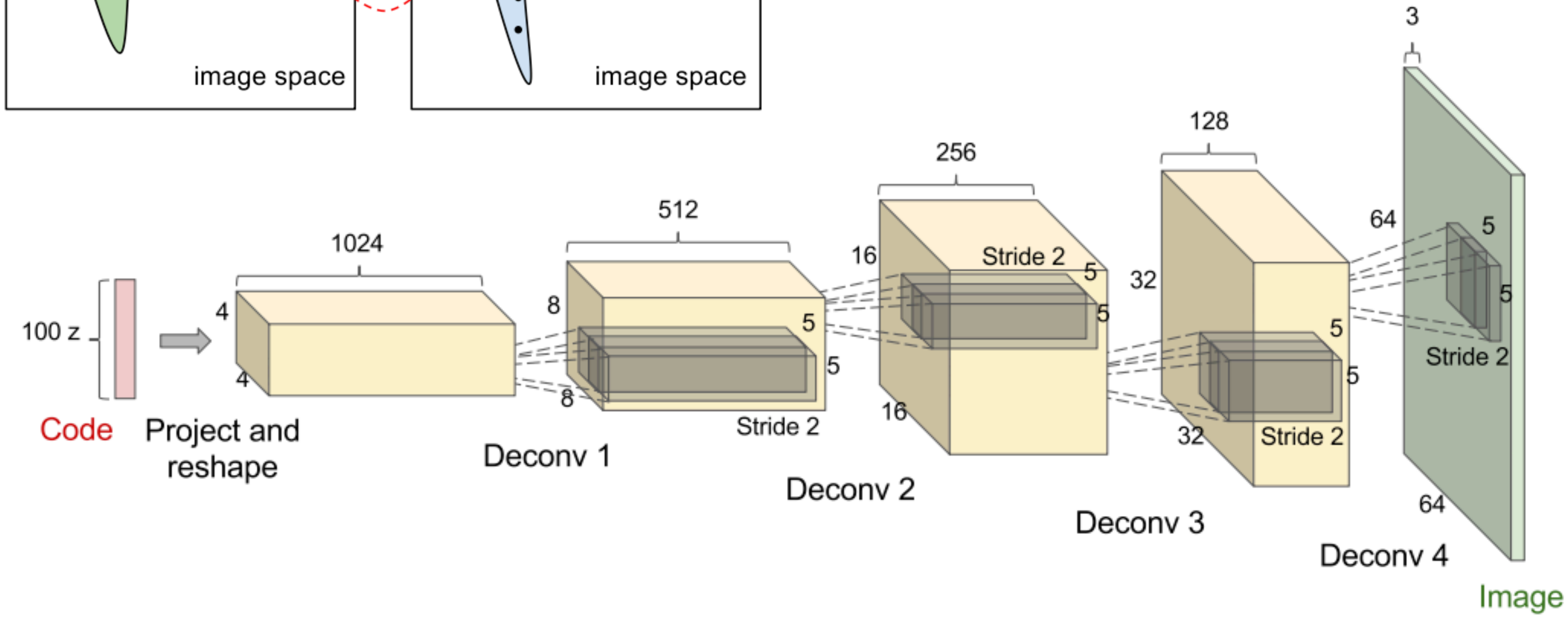
Output of transpose convolution:

output size = (input size - 1)*stride - 2*padding + (kernel size - 1) +1

Ding Zhao | CMU | 2021

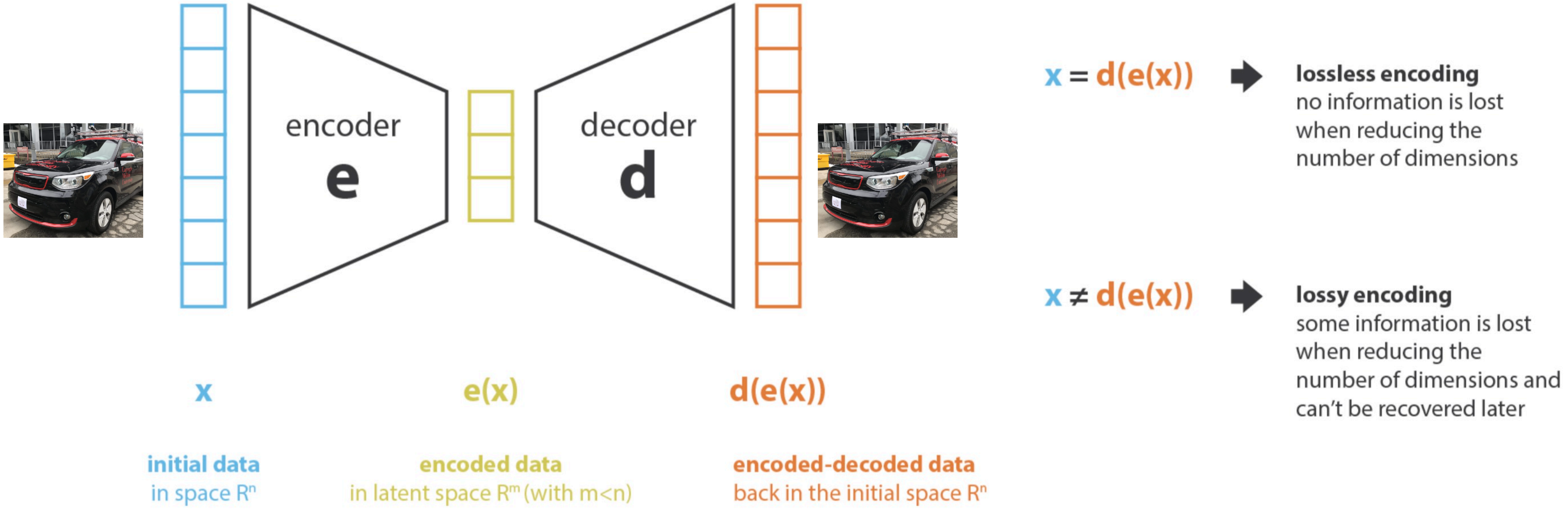
Credits: http://makeyourownneuralnetwork.blogspot.com/2020/02/calculating-output-size-of-convolutions.html

14



https://openai.com/blog/generative-models/

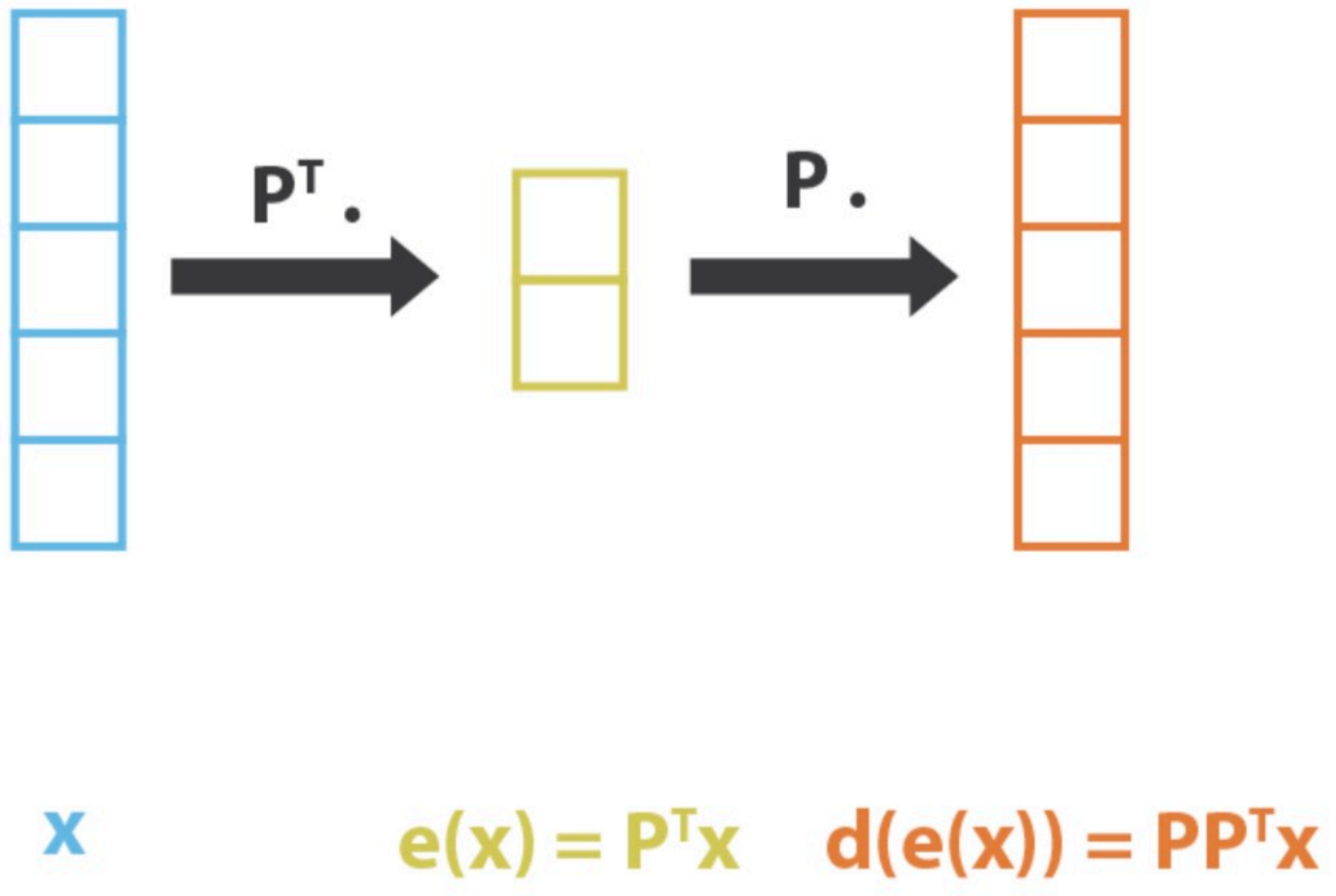
Vanilla autoencoder

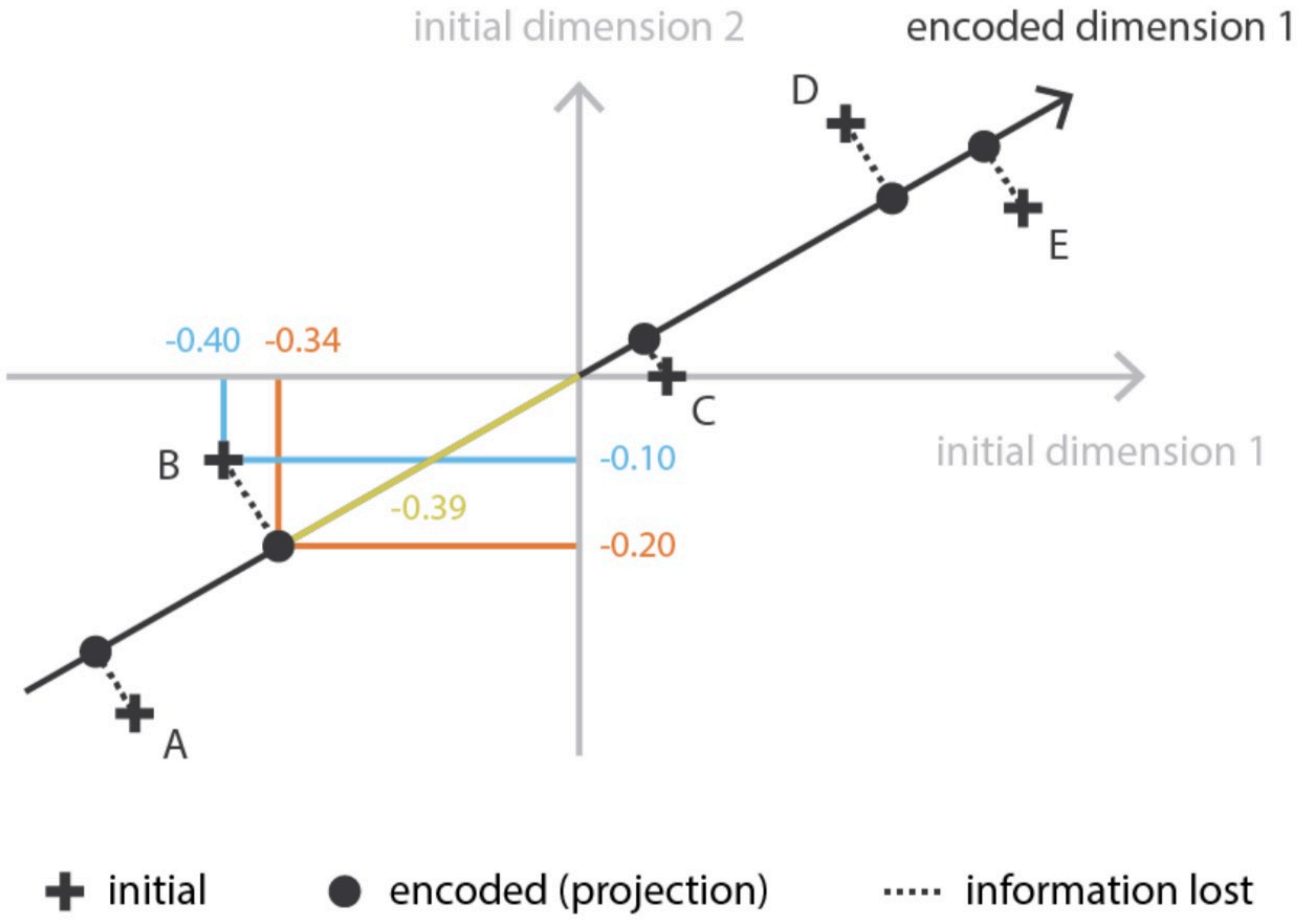


https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Linear autoencoder

- The optimal solution of a linear autoencoder can be obtained with PCA (Principal Component Analysis). The latent space will be by calculating the Singular Value Decomposition (SVD).



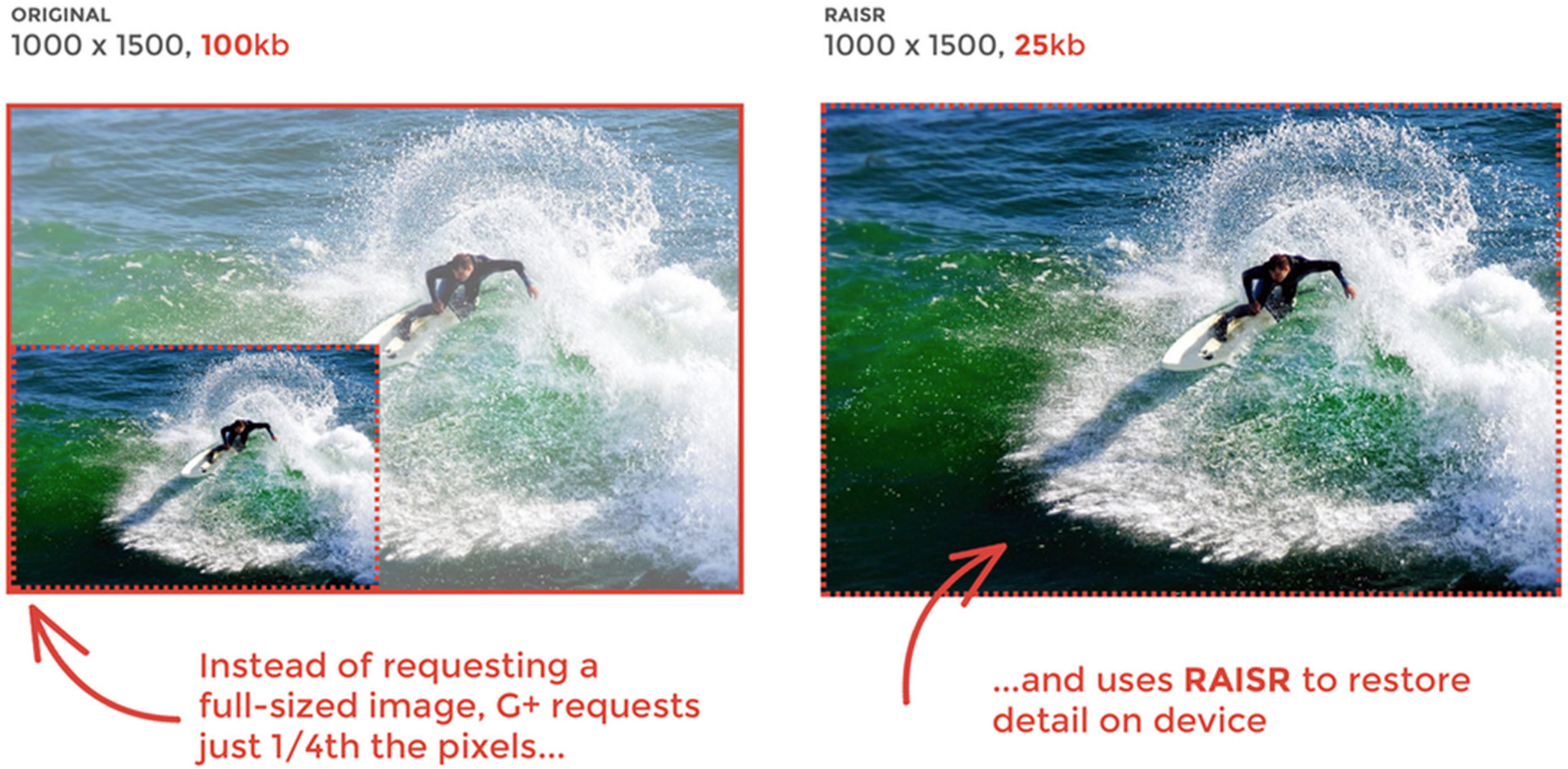


Point	Initial	Encoded	Decoded
A	(-0.50, -0.40)	-0.63	(-0.54, -0.33)
B	(-0.40, -0.10)	-0.39	(-0.34, -0.20)
C	(0.10, 0.00)	0.09	(0.07 0.04)
D	(0.30, 0.30)	0.41	(0.35, 0.21)
E	(0.50, 0.20)	0.53	(0.46, 0.27)

https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

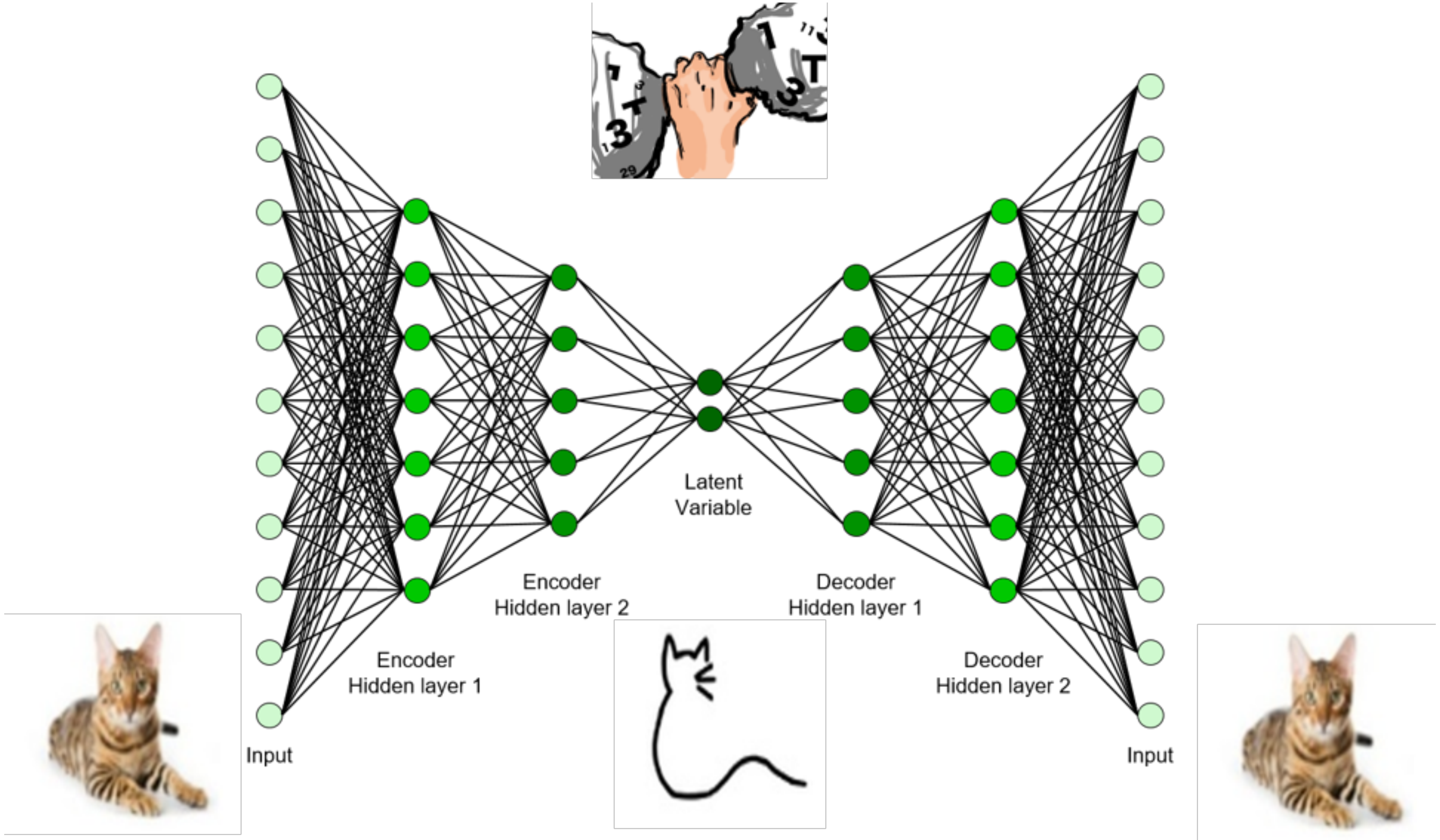
Autoencoder

- Autoencoder can be used as data compression algorithm
  - Google+ sends “latent images” and uses auto encoder to reconstruct images locally



https://www.slrlounge.com/google-raisr-image-resolution-enhancement-straight-out-of-csi/

Autoencoder with neural networks

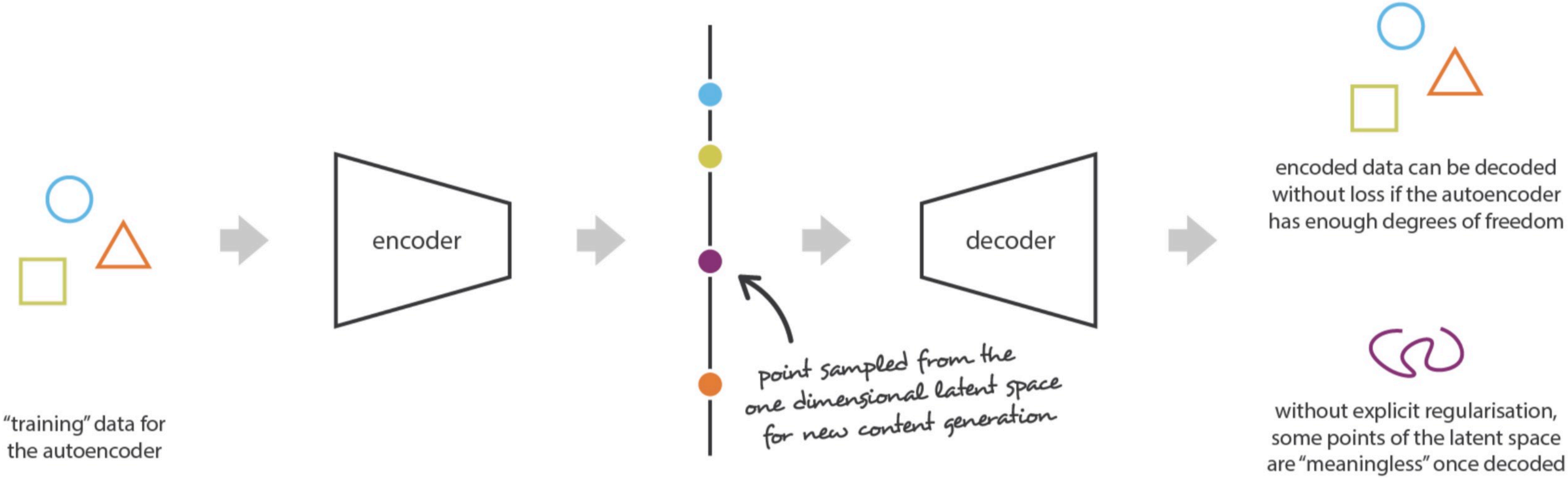


x ϕ z θ $\hat{x}$

https://towardsdatascience.com/extreme-rare-event-classification-using-autoencoders-in-keras-a565b386f098

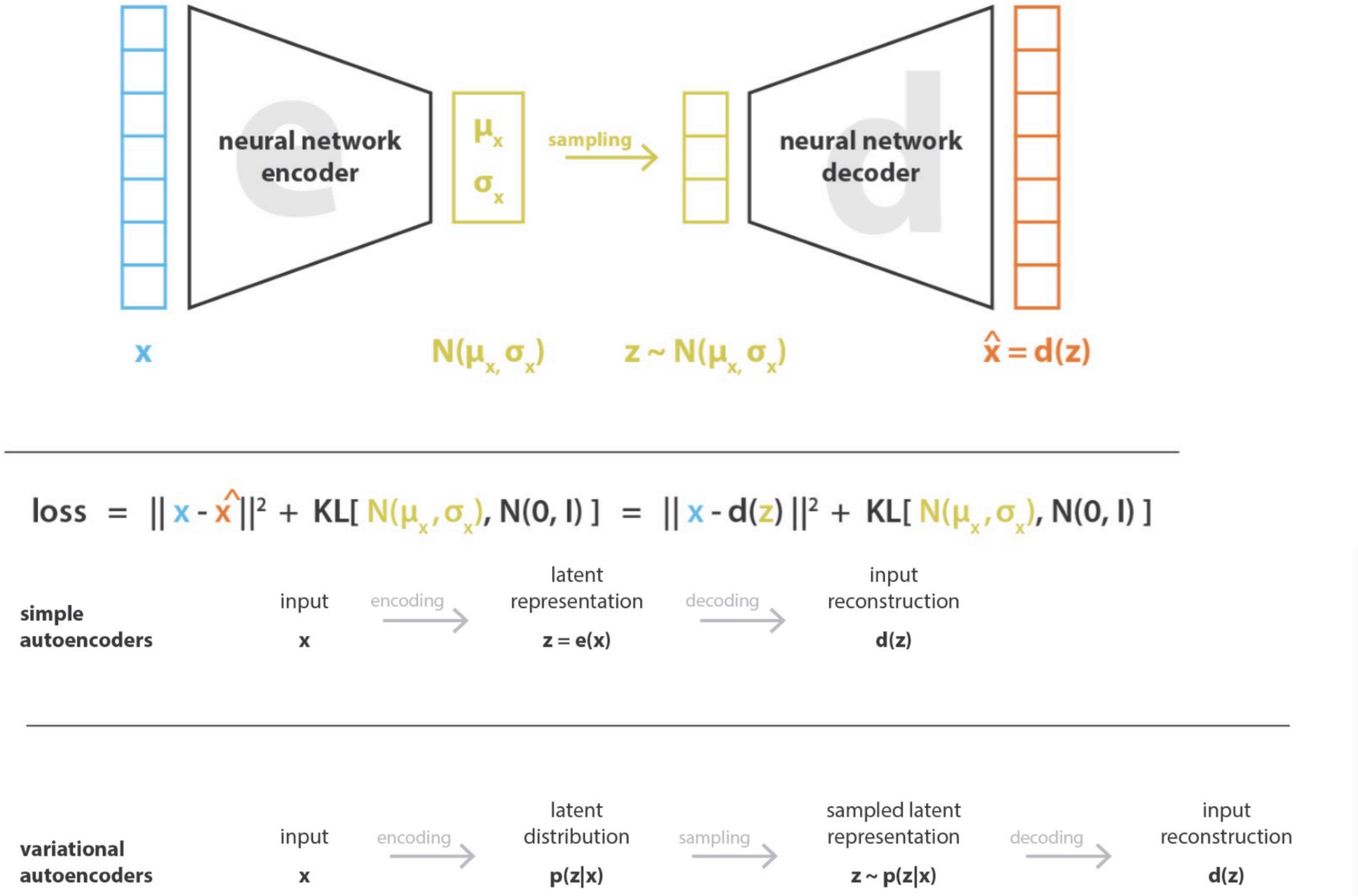
Issues of unregulated autoencoder

- Question: can we use one dimensional number in the latent space?
- Two ideas:
  - 1) add noise to randomize the system; 2) regularize the latent space



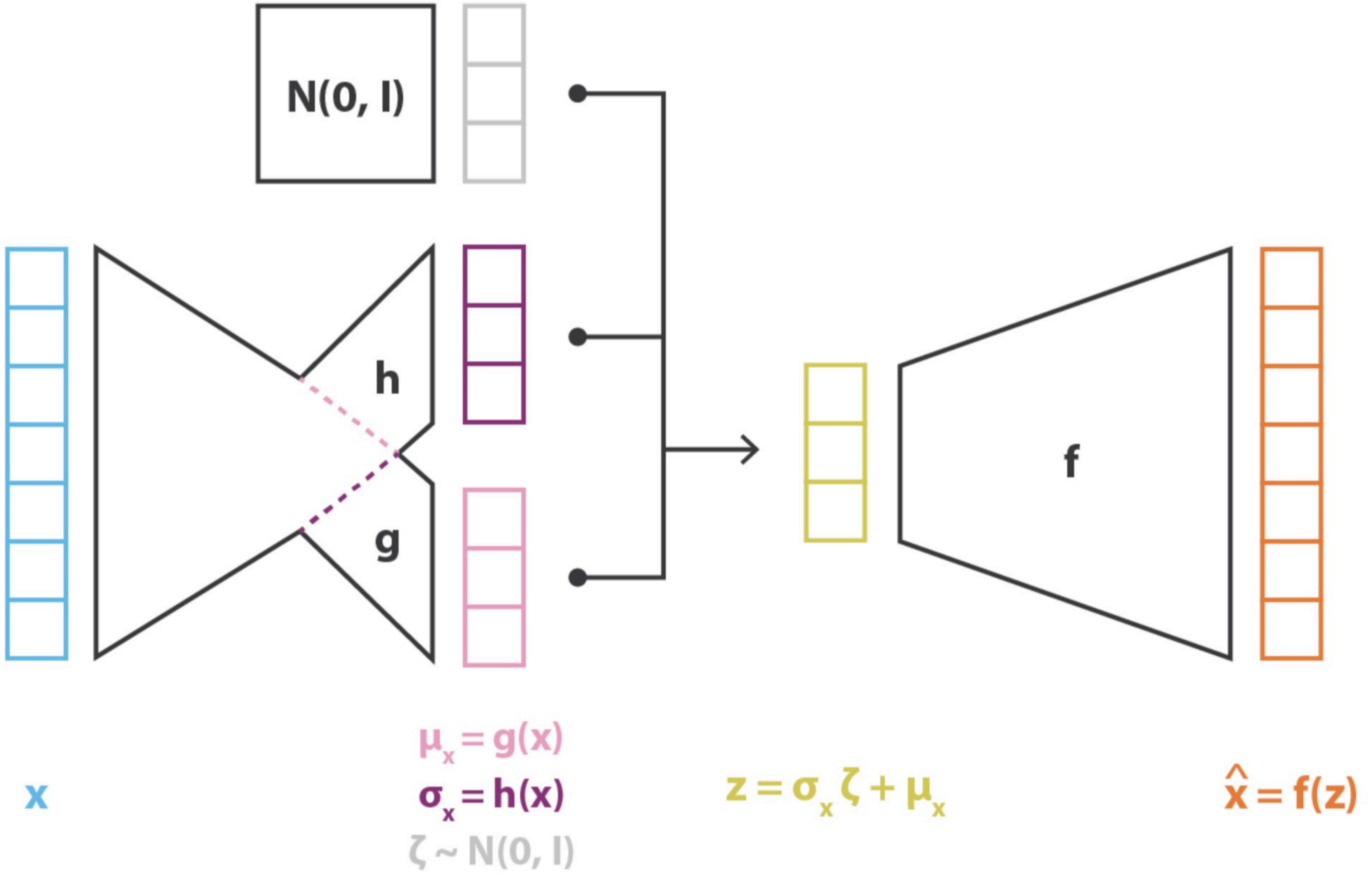
https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Variational Autoencoder



https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Variational Autoencoder

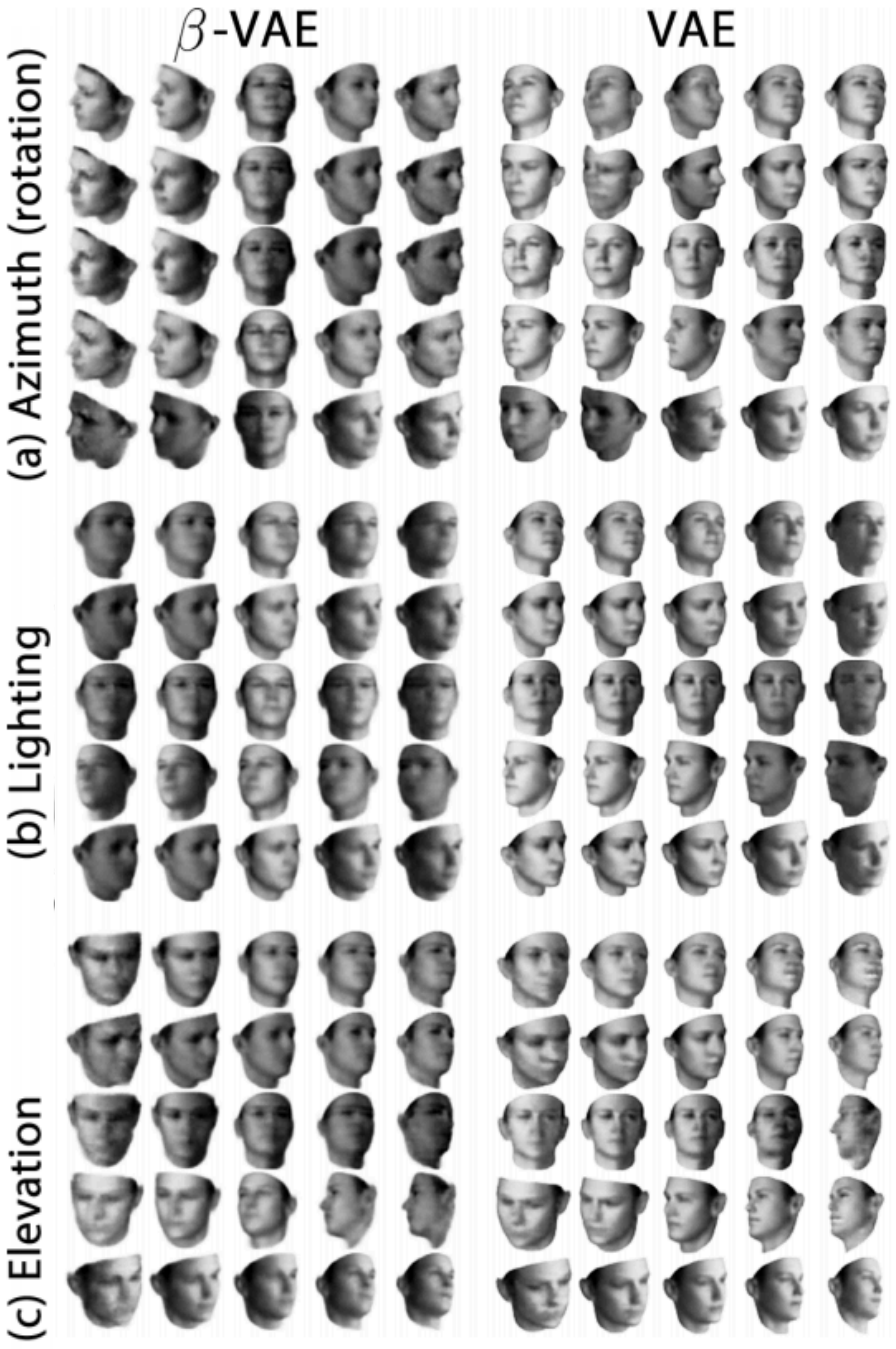


$$\text{loss} = C\,\|x - \hat{x}\|^2 + \mathrm{KL}[\,N(\mu_x, \sigma_x), N(0, I)\,] = C\,\|x - f(z)\|^2 + \mathrm{KL}[\,N(g(x), h(x)), N(0, I)\,]$$

https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

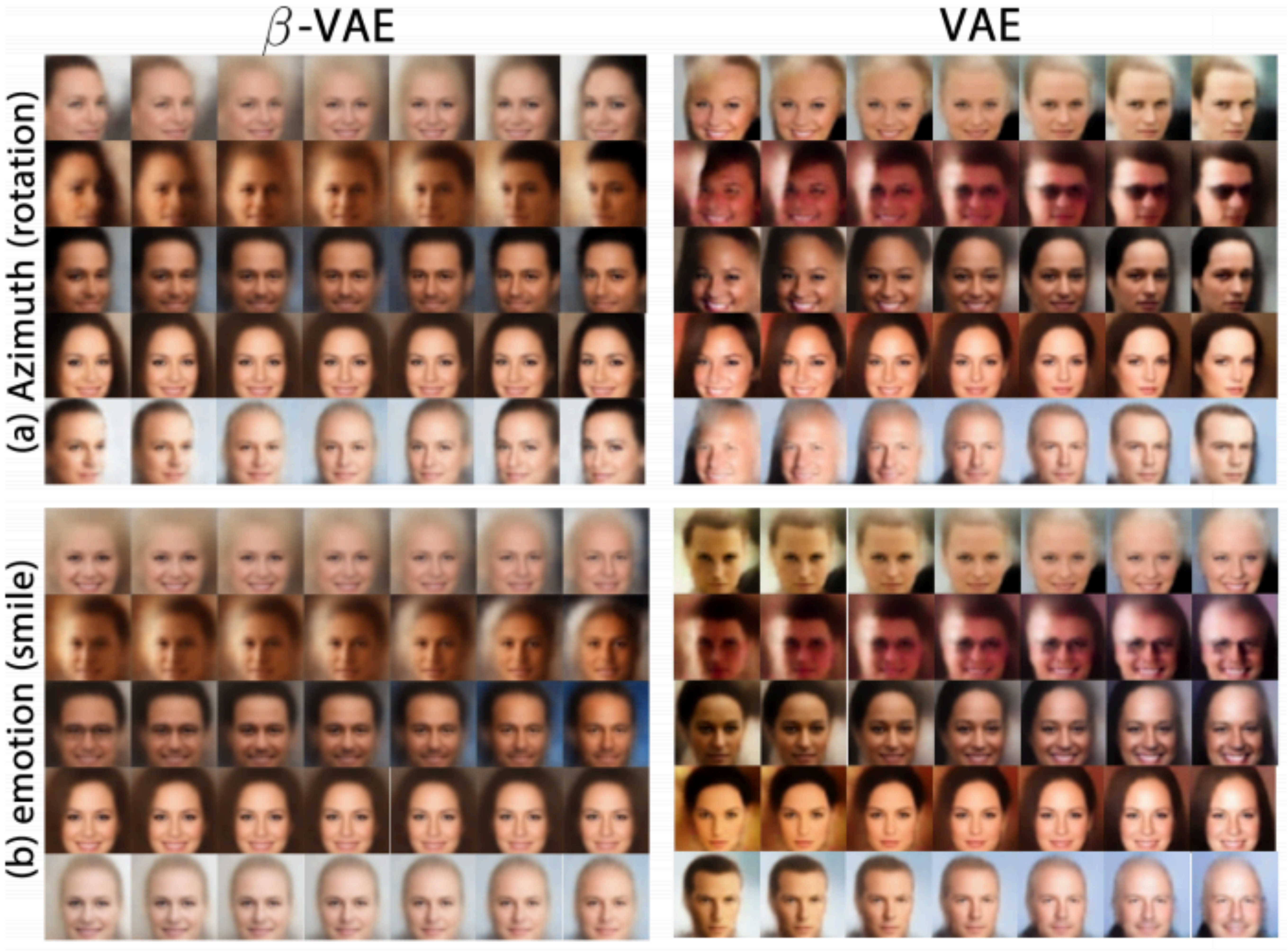
Disentangled VAE (β-VAE)

- Goal: ensuring each dimension of latent vectors learn distinct attributes
- This can be achieved by adding hyperparameter β to the loss function:
$$\mathscr{L}(\theta, \phi, \beta, X, Z) = \mathbb{E}[\log p_\theta(X|Z)] + \boxed{\beta} D_{KL}(q_\phi(Z|X) \| p(Z))$$
- The model then learns to use latent space as efficient as possible



Higgins, Irina, et al. "beta-vae: Learning basic visual concepts with a constrained variational framework." (2016).

Examples of VAE in practice

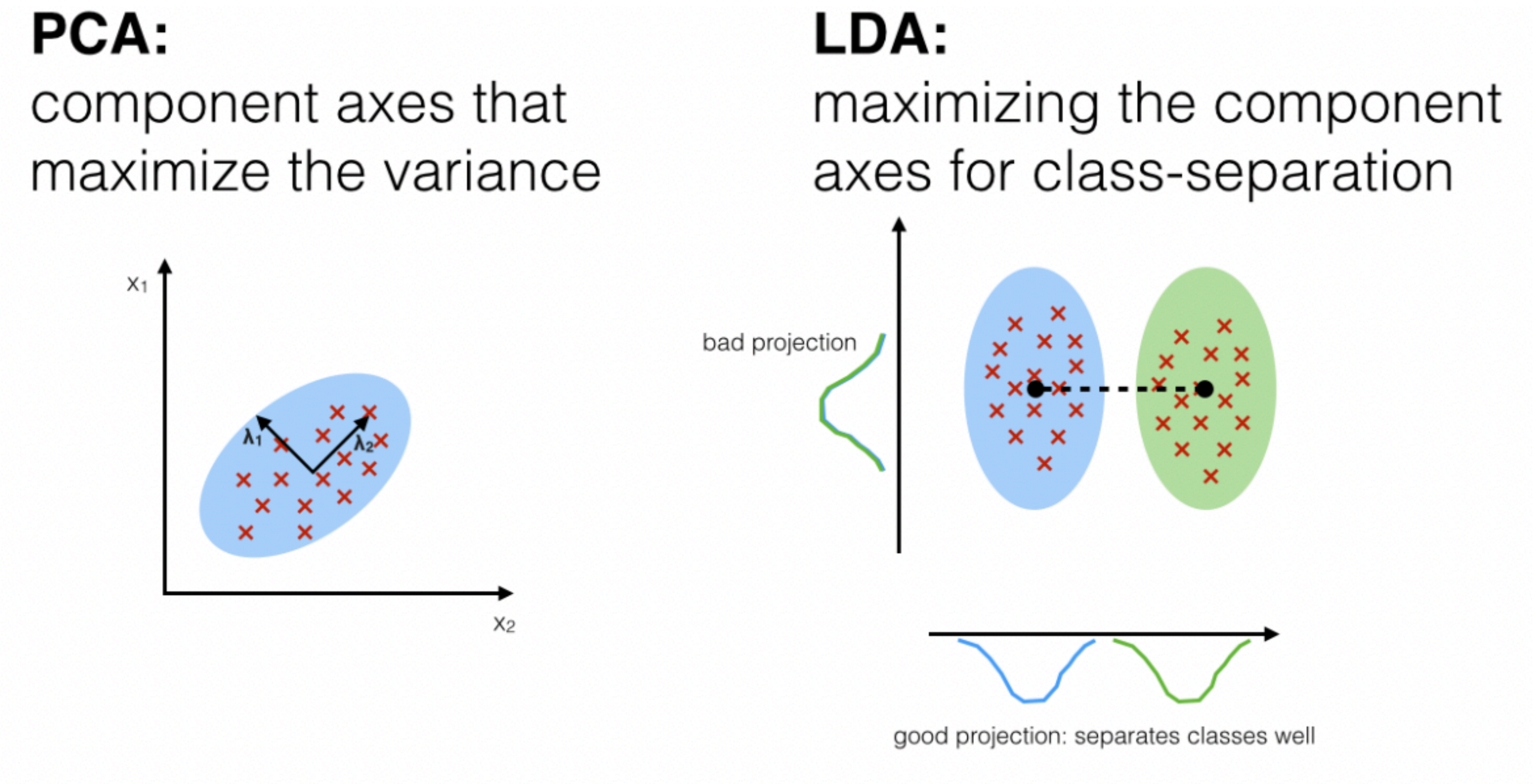


Higgins, Irina, et al. "beta-vae: Learning basic visual concepts with a constrained variational framework." (2016).

Visualize the latent code

- PCA: Principal Component Analysis
    - Linear method, not robust to outliners
- t-SNE: t-distributed stochastic neighbor embedding
    - Nonlinear method, slow, may apply PCA first
- LDA: Linear Discriminant Analysis
    - Maximize the separation between multiple classes
    - Fast, need to know the labels of classes

LDA: Linear Discriminant Analysis



https://towardsdatascience.com/dimensionality-reduction-for-data-visualization-pca-vs-tsne-vs-umap-be4aa7b1cb29

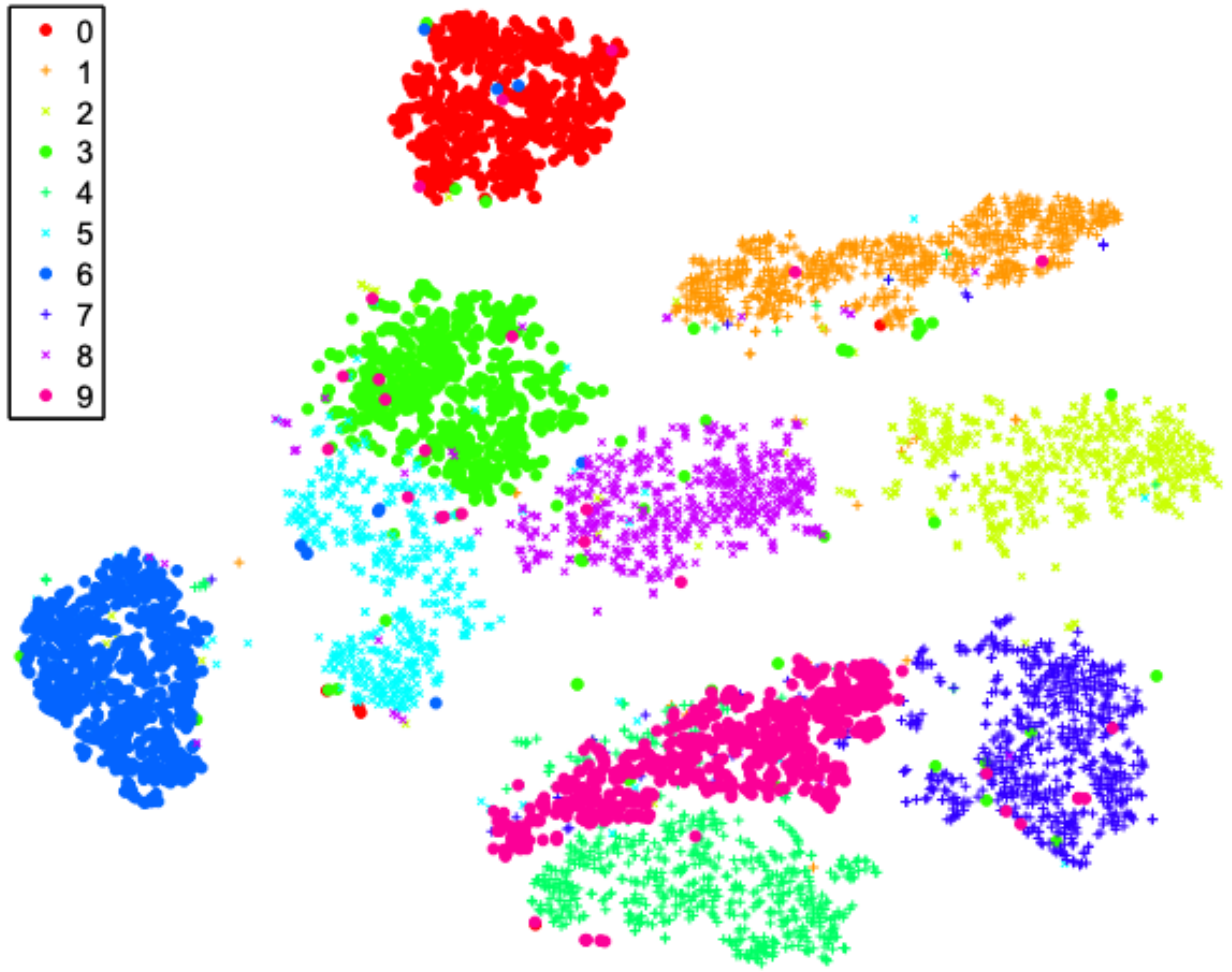
t-distributed stochastic neighbor embedding (t-SNE)

How to visualize the latent space z of VAE?

- Reduce the dimension to 2 or 3
- Unsupervised dimension reduction
- Similar vectors should be close

Toolbox

- sklearn.manifold.TSNE



(a) Visualization by t-SNE.

Visualization of MNIST (digits 0-9) dataset

https://distill.pub/2016/misread-tsne/

https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

Van Der Maaten, L.J.P.; Hinton, G.E. (Nov 2008). "Visualizing Data Using t-SNE". Journal of Machine Learning Research. 9: 2579–2605.

t-distributed stochastic neighbor embedding (t-SNE)

How to visualize the latent space z of VAE?

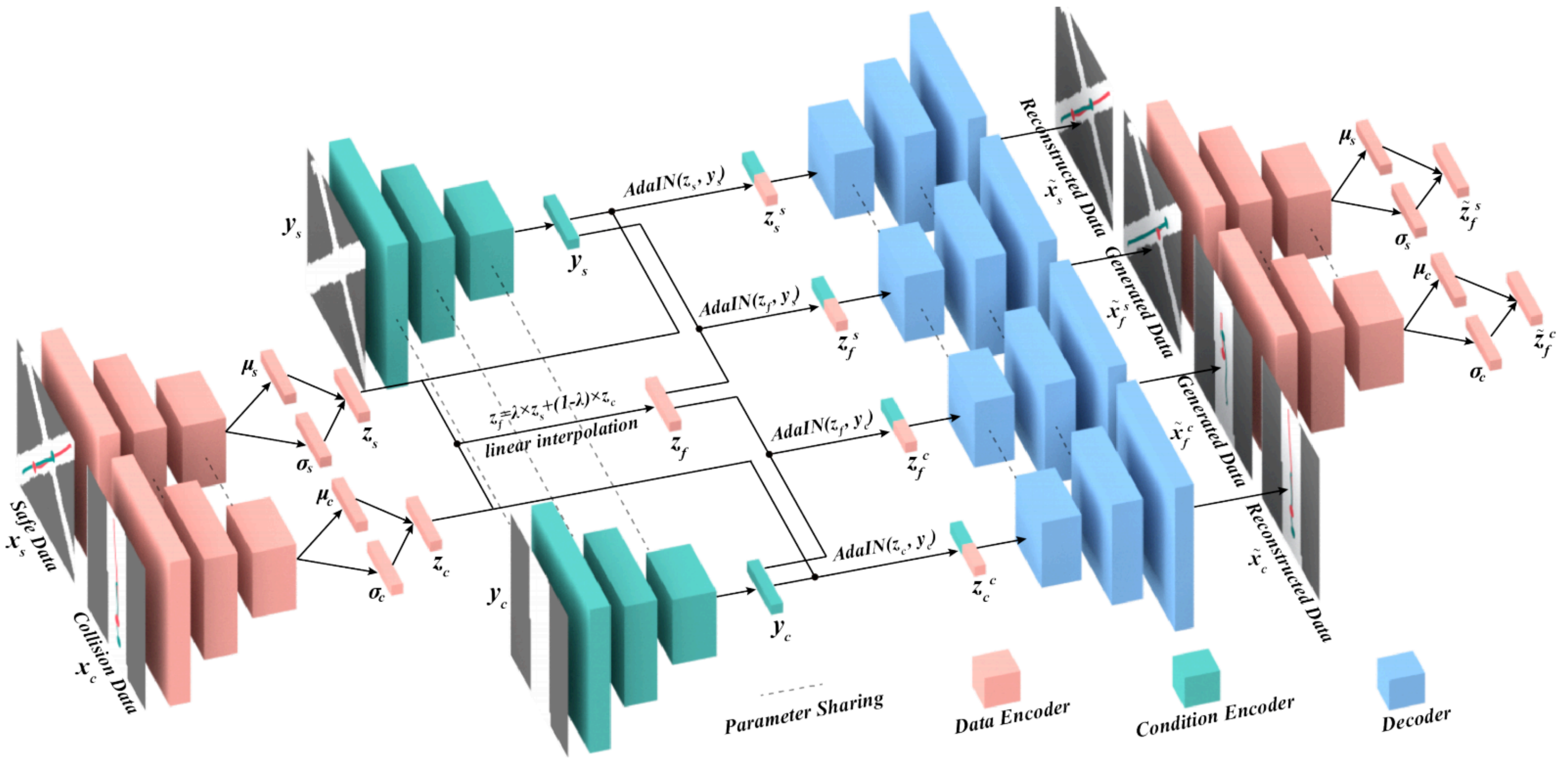
- Reduce the dimension to 2 or 3
- Unsupervised dimension reduction
- Similar vectors should be close

Toolbox

- sklearn.manifold.TSNE

https://www.oreilly.com/people/cyrille-rossant/

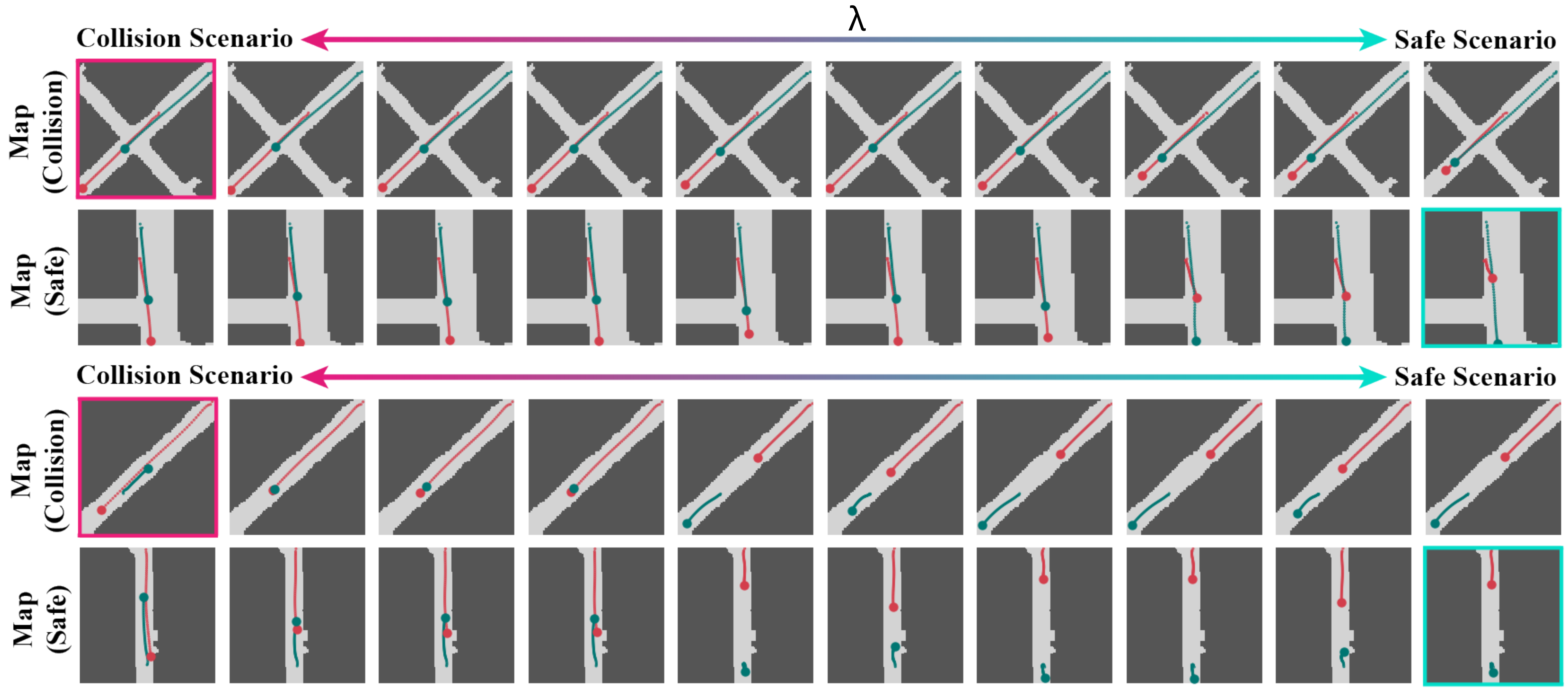
Safety-critical trajectory via VAE



- Use linear interpolation of collision data and safe data to generate safety-critical data
- Use road bird-view image as constraints

Ding W, Xu M, Zhao D. Cmts: A conditional multiple trajectory synthesizer for generating safety-critical driving scenarios[C]//2020 IEEE International Conference on Robotics and Automation (ICRA).

Safety-critical trajectory via VAE



- Different map conditions have different trajectory output
- λ controls the risk value

Ding W, Xu M, Zhao D. Cmts: A conditional multiple trajectory synthesizer for generating safety-critical driving scenarios[C]//2020 IEEE International Conference on Robotics and Automation (ICRA).

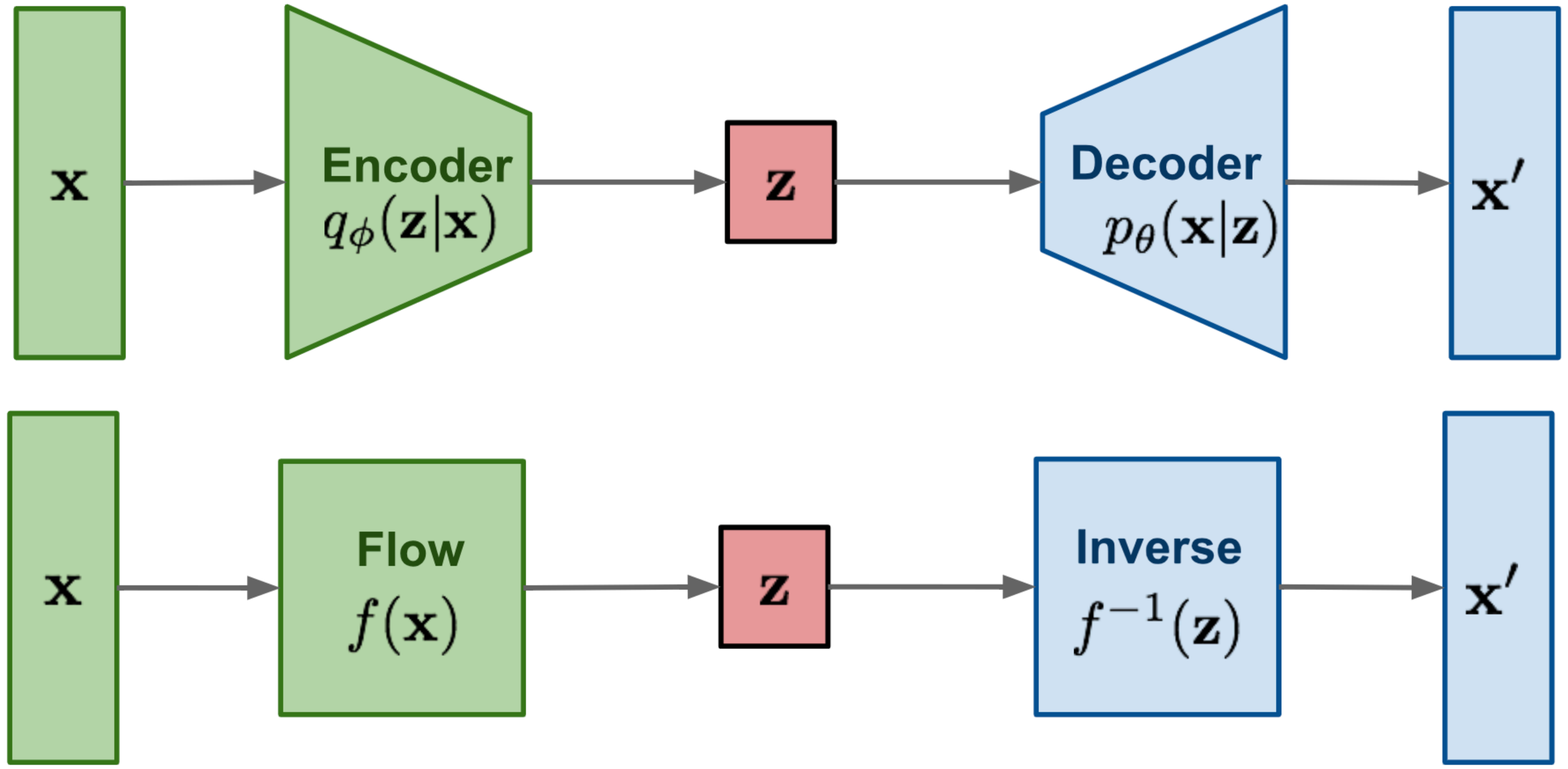
Flow-based generative models

Approximate likelihood

VAE: maximize ELBO.

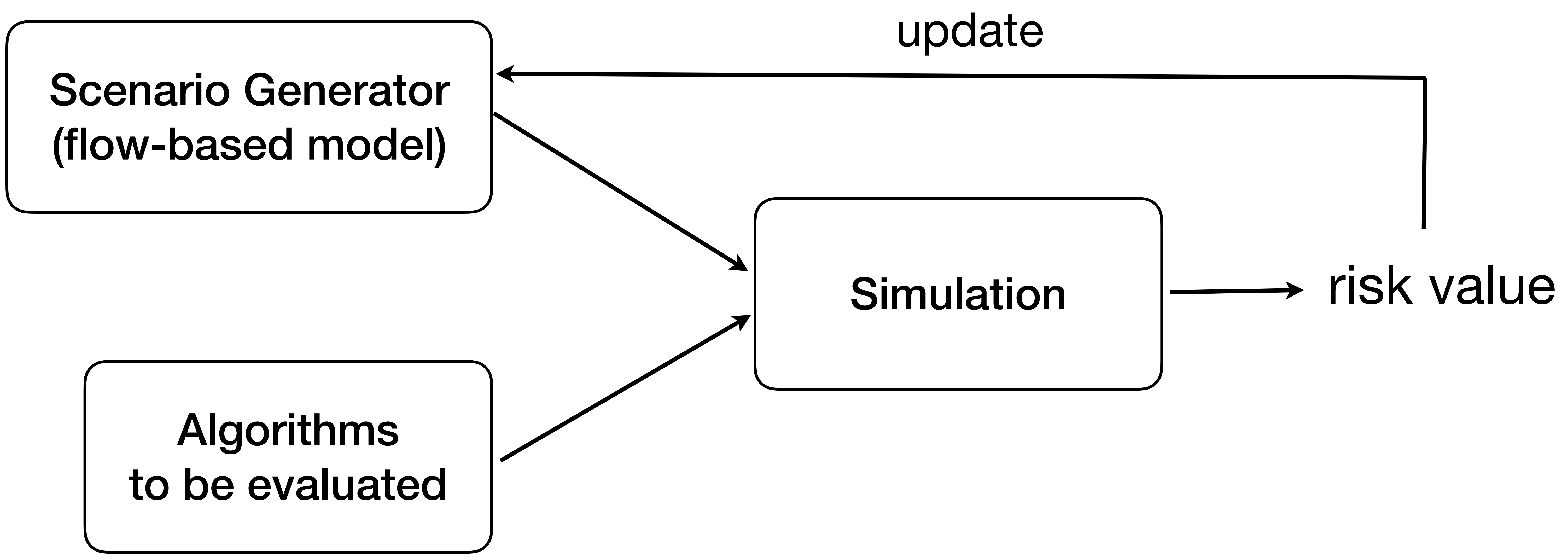
Exact likelihood

Flow-based generative models: minimize the negative log-likelihood



https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.html

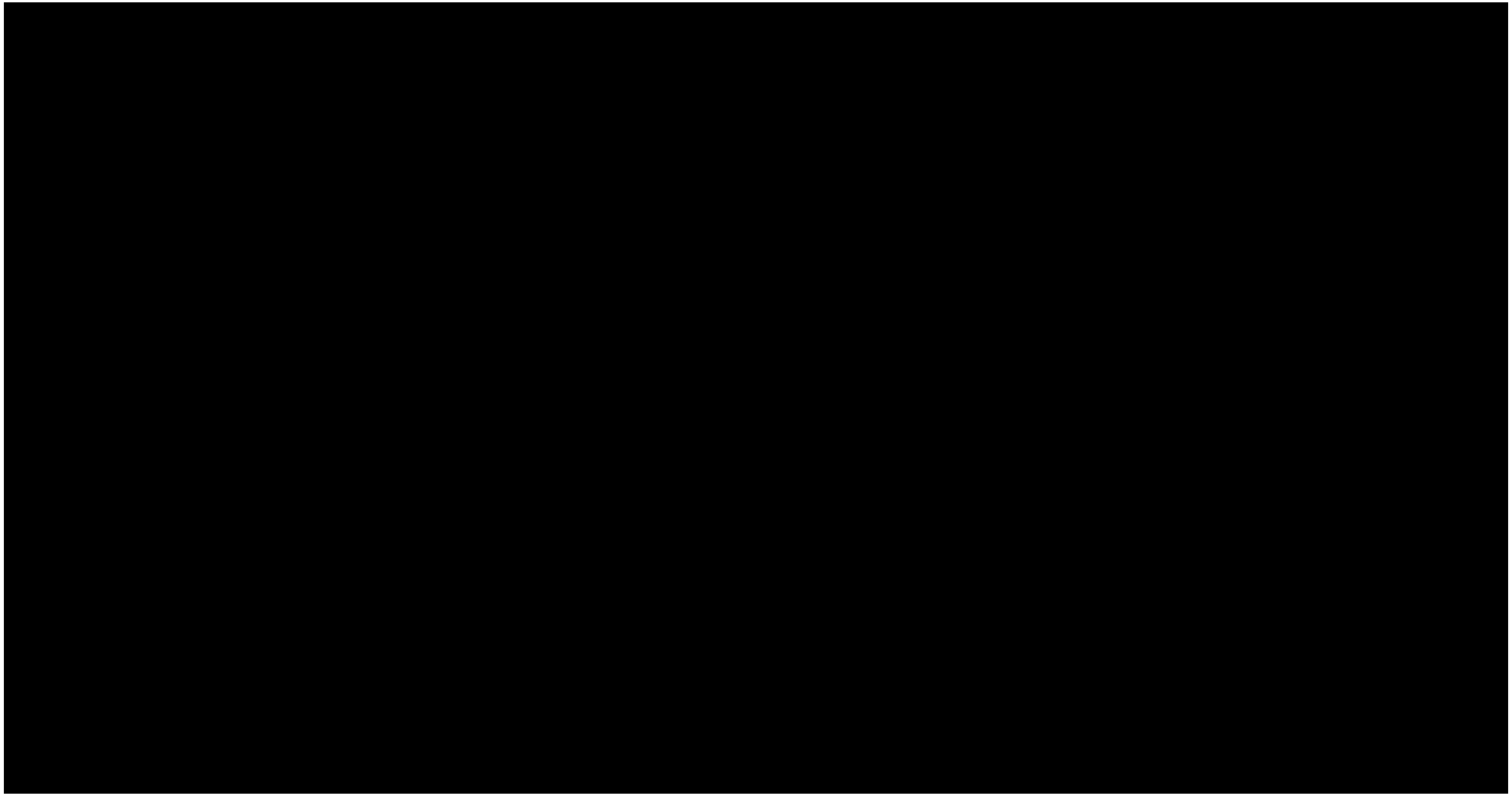
Safety-critical scenarios generation with flow-based model



- An adversarial attack framework
- Use flow-based model to estimate the distribution of risky traffic scenarios

Ding W, Chen B, Li B, et al. Multimodal Safety-Critical Scenarios Generation for Decision-Making Algorithms Evaluation[J]. arXiv preprint arXiv:2009.08311, 2020.

Safety-critical scenarios generation with flow-based model



Ding W, Chen B, Li B, et al. Multimodal Safety-Critical Scenarios Generation for Decision-Making Algorithms Evaluation[J]. arXiv preprint arXiv:2009.08311, 2020.

Data-based Scenario Generation

Summary

- Use the real-world data
- Some kind of controllability

- Poor adaptivity, no interaction with downstream task
- Only use existing data, lack of diversity

Worth Reading

- Beta-VAE
  Higgins, Irina, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. "beta-vae: Learning basic visual concepts with a constrained variational framework." (2016).
  Check open review: https://openreview.net/forum?id=Sy2fzU9gl

- General intro to GAN:
  Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2020. Generative adversarial networks. *Communications of the ACM*, *63*(11), pp.139-144.
  GAN Lab: https://poloclub.github.io/ganlab/