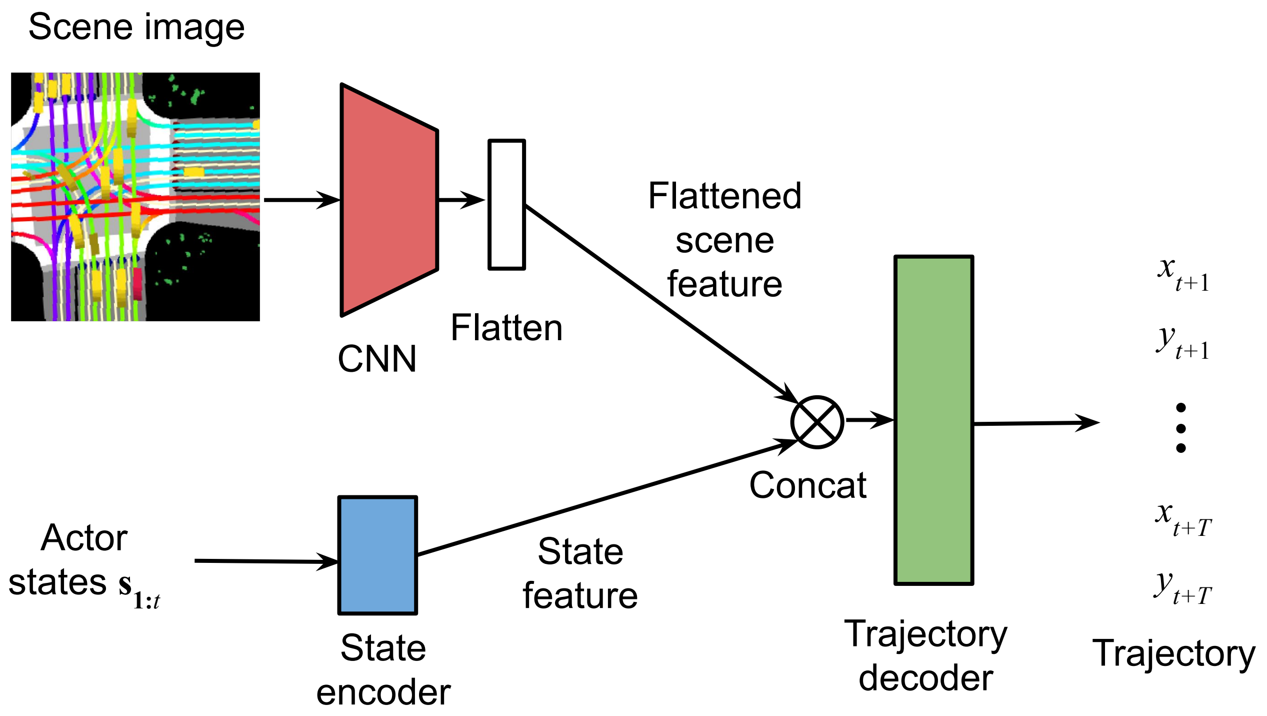
Improving Movement Predictions of Traffic Actors in Bird's-Eye View Models using GANs and Differentiable Trajectory Rasterization

Eason Wang*, Henggang Cui*, Sai Yalamanchi,
Mohana Moorthy, Fang-Chieh Chou, Nemanja Djuric

UBER
ATG

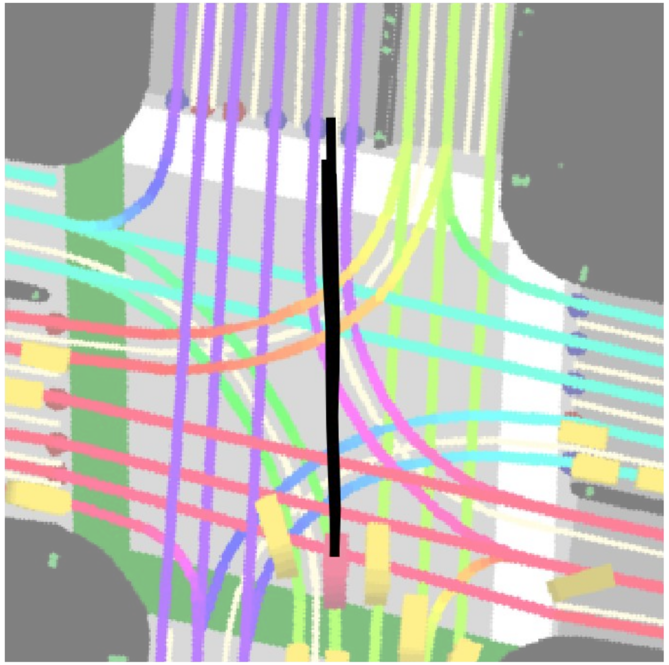
Trajectory prediction

- Given an actor's current and history states and the scene context, the goal is to predict its future trajectory(s)
- Extract scene features with CNNs from the scene image
- Extract actor state features with a state encoder network
- Generate trajectory predictions with a trajectory decoder network



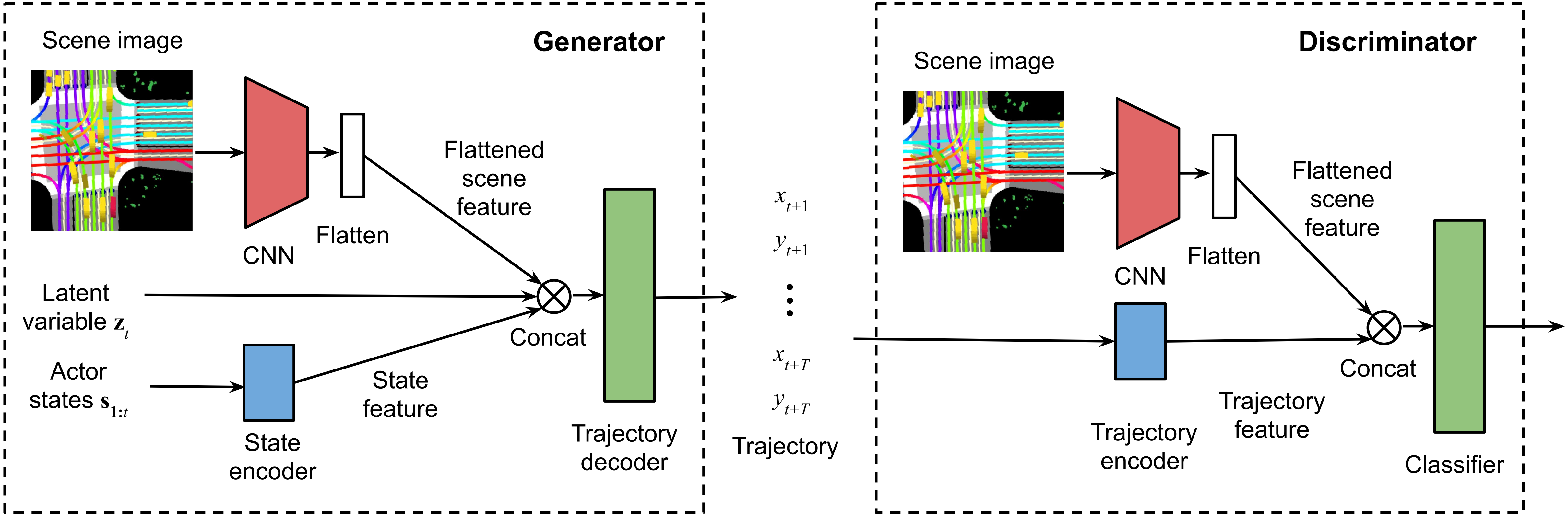
Scene-compliant trajectory prediction

- Predicting the future is not an easy task, but at least we know:
  - An actor is unlikely to drive out of the road
  - An actor is unlikely to straight into the opposing lanes from a left-turn only lane
- As humans, we know this prediction is not correct even without looking at the ground-truth because it’s not *scene-compliant*



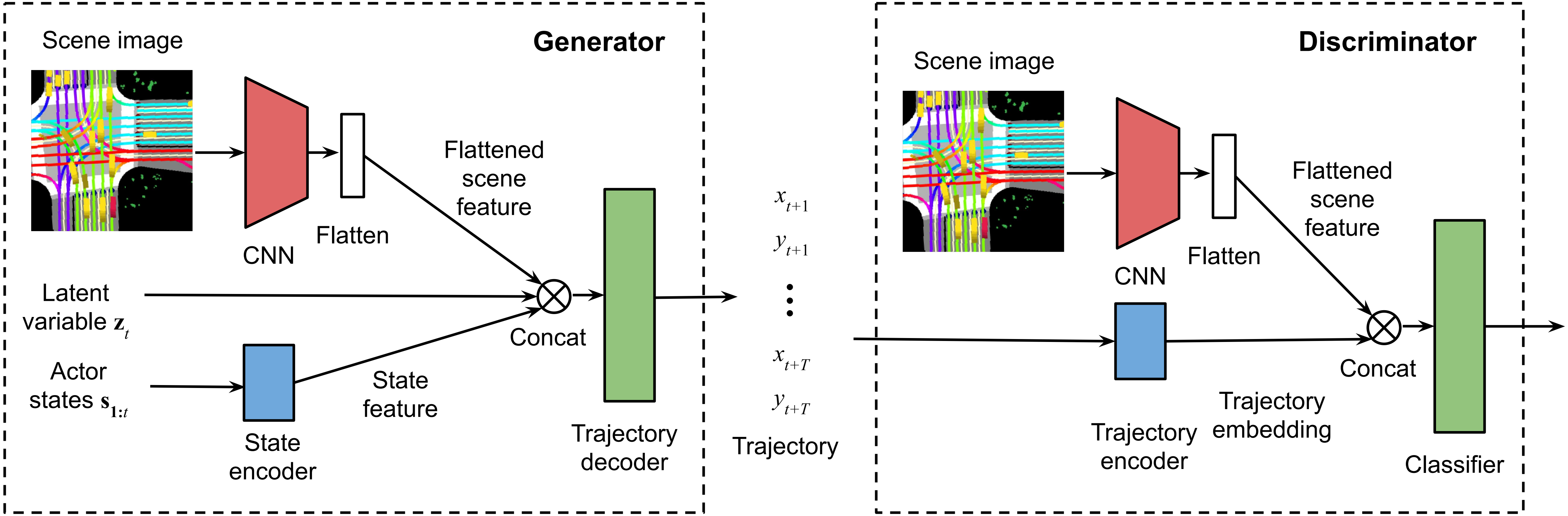
Trajectory prediction with conditional GANs

- Generate trajectory predictions with a conditional GAN model
  - E.g., Social-GAN, Sophie, Social-BiGAT
- A discriminator network is added to discriminate whether a given trajectory is real or fake
- The two networks are trained jointly with some GAN loss



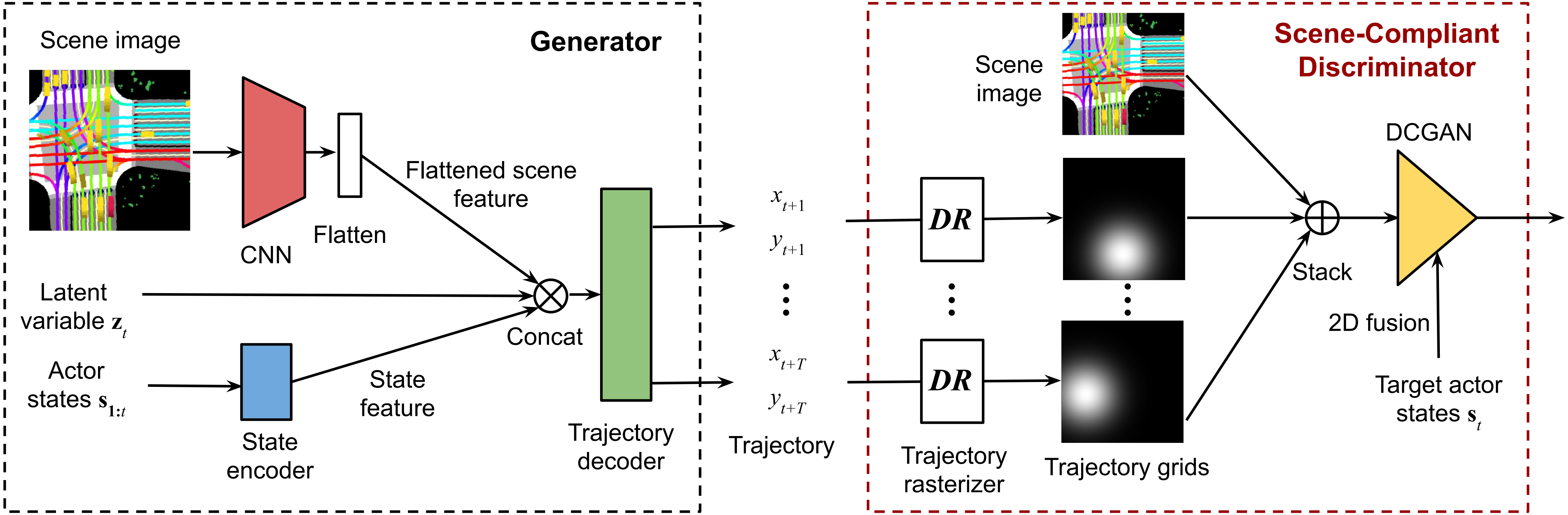
Flaws with the traditional GAN-based approaches

- The scene features are flattened and concatenated with the encoded trajectory embeddings in the discriminator
- It's hard for the discriminator to distinguish scene-compliant and non-scene-compliant trajectories



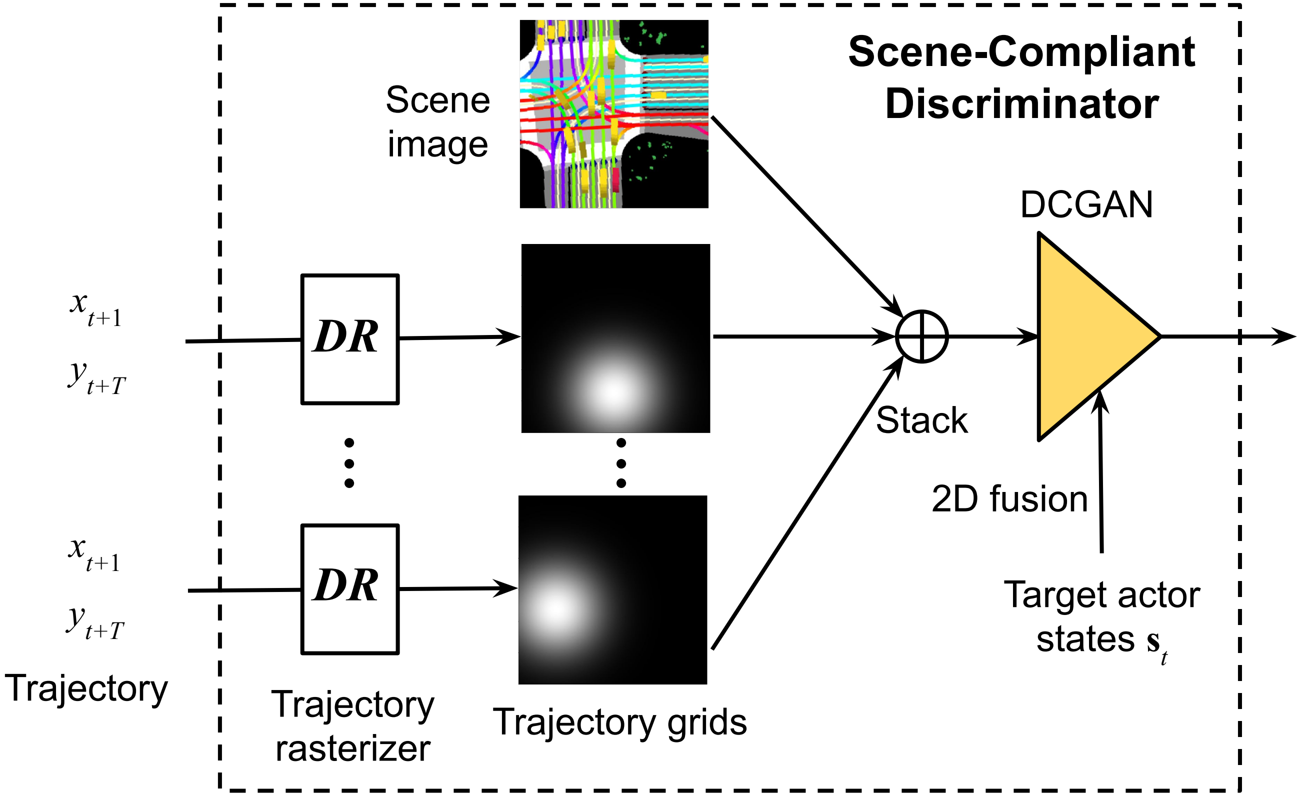
Trajectory prediction with Scene-Compliant GAN

- Scene-Compliant GAN (SC-GAN)
  - The same generator architecture as in the previous works
  - But with a novel *Scene-Compliant Discriminator*



Scene-Compliant Discriminator

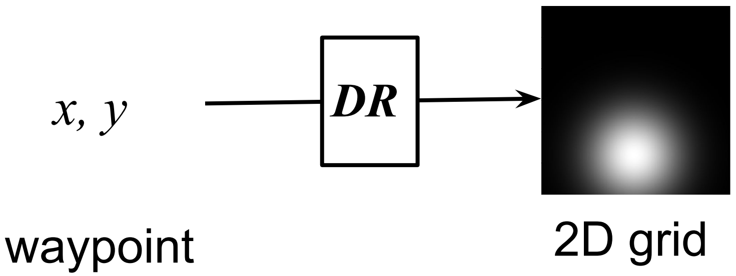
- It rasterizes a trajectory waypoint into a 2D image with a novel *Differentiable Trajectory Rasterizer*
- The trajectory raster images are stacked with the input scene image in the channel dimension
- This reduces the problem to a classic image generation problem
- DCGAN is used as the discriminator architecture which is known to work well for image inputs



UBER ATG

Differentiable Trajectory Rasterizer

- The Differentiable Trajectory Rasterizer differentiably transforms a waypoint into a 2D grid



- For each cell *(i, j)* in the grid, we compute its displacement vector from the waypoint (x, y) as $\boldsymbol{\Delta}^{ij}$
- The value of cell *(i, j)* is set as the density of a 2D Gaussian distribution $N(0, \boldsymbol{\Sigma})$ evaluated at $\boldsymbol{\Delta}^{ij}$

$$\{\mathcal{G}_t\}_{ij} = \mathcal{N}(\Delta_t^{ij}|0, \boldsymbol{\Sigma})$$

- $\boldsymbol{\Sigma}$ = diag(σ^2, σ^2) is a diagonal matrix, and σ controls the probability density of the raster
- The gradients are well-defined, and the direction is aligned with the displacement vector $\boldsymbol{\Delta}^{ij}$

$$\nabla_{[x_t, y_t]}(\{\mathcal{G}_t\}_{ij}) = [\frac{\partial\{\mathcal{G}_t\}_{ij}}{\partial x}, \frac{\partial\{\mathcal{G}_t\}_{ij}}{\partial y}] = -\frac{\{\mathcal{G}_t\}_{ij}}{\sigma^2}\Delta_t^{ij}$$

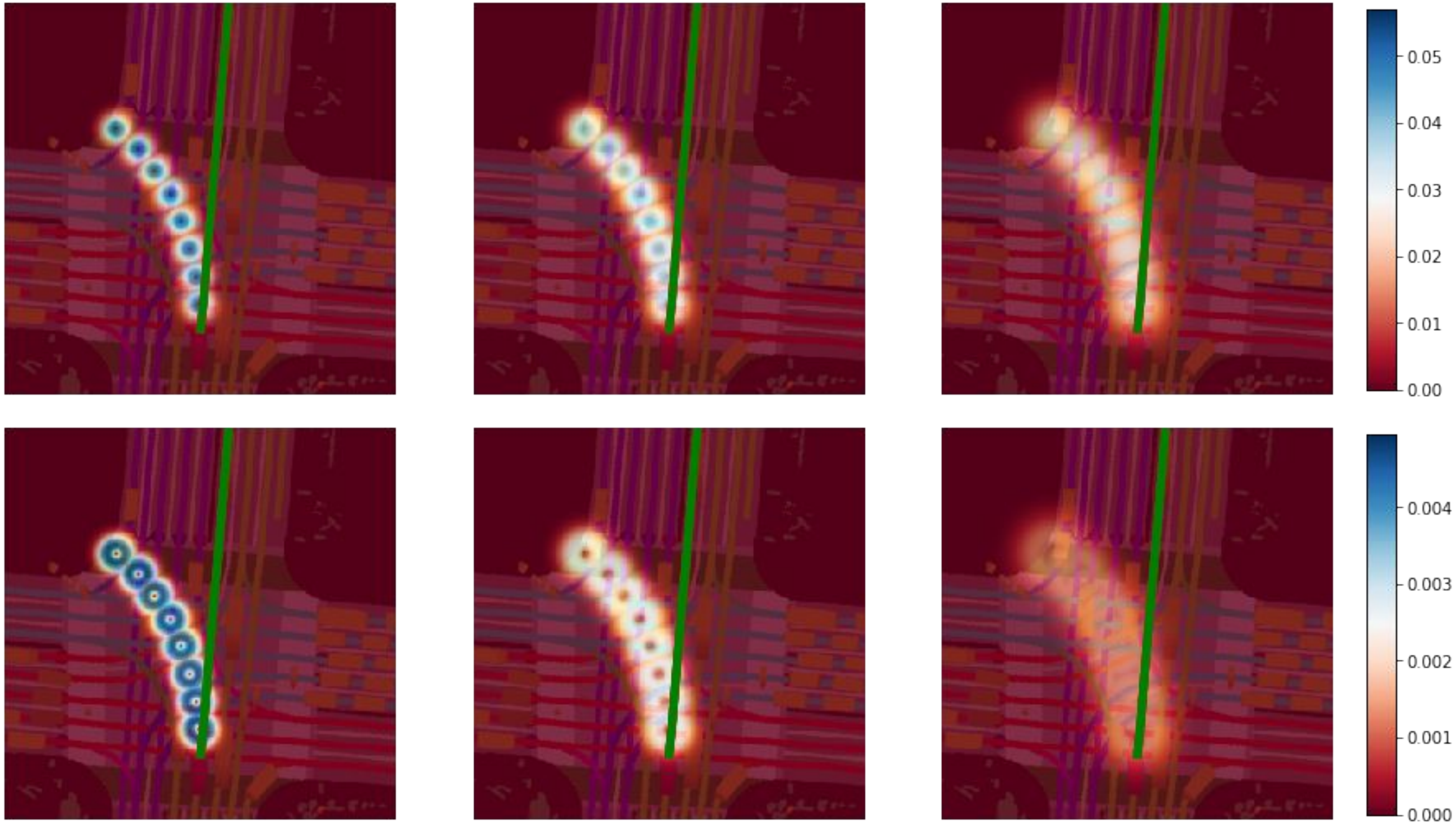
UBER ATG

Differentiable rasterizer

$\sigma = 7$ $\sigma = 10$ $\sigma = 15$

Raster $\{\mathcal{G}_t\}$

Gradients $\nabla_{[x_t, y_t]}(\{\mathcal{G}_t\}_{ij})$



Images are 300x300

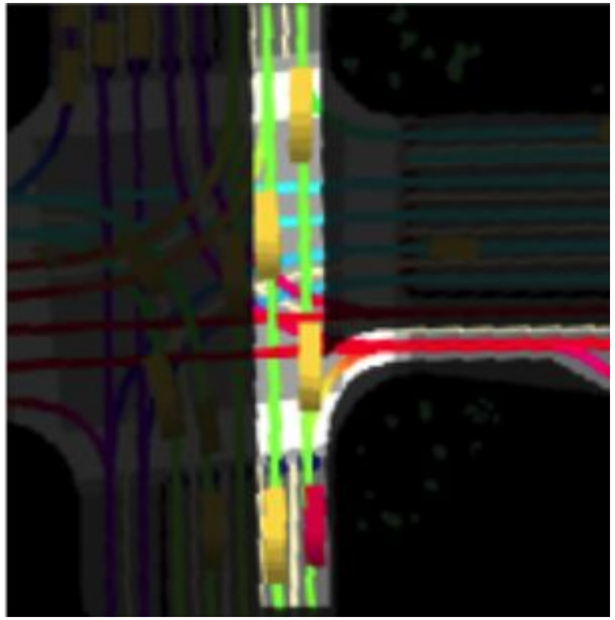
UBER ATG

Model details

- By default, all models are trained with only the GAN loss without the L2 loss
    - Allowing more effective comparisons between the GAN architectures
- We use the Wasserstein GAN loss with gradient penalty as the GAN loss
- Rasters are 300x300 with resolution 0.2 m/pxl
- σ is set to 10

Evaluating scene-compliance

- Identify a *drivable region* for each actor using the path proposal and scoring modules from GBP
- Off-road metrics
  - Off-road distance
    - The distance from the predicted waypoint to the drivable region (if outside)
  - Off-road false-positive %
    - The percentage of predicted waypoints that are outside the drivable region while the corresponding ground-truth is not



drivable region

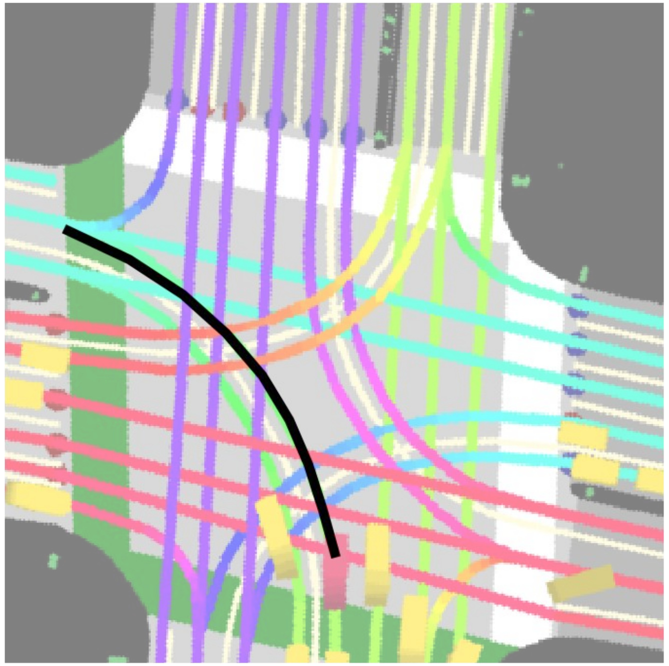
Quantitative results

- Baselines
  - no-scene-GAN (similar to Social-GAN and Sophie)
    - No scene image used in discriminator
  - concat-scene-GAN (similar to Social-BiGAT)
    - Scene image features are flattened and concatenated with trajectory embeddings
- Each model generates multiple trajectory samples
- We measure both the mean and min for L2 and only the mean for off-road metrics
- Scene-Compliant GAN improves off-road distance and off-road false-positives by a large margin

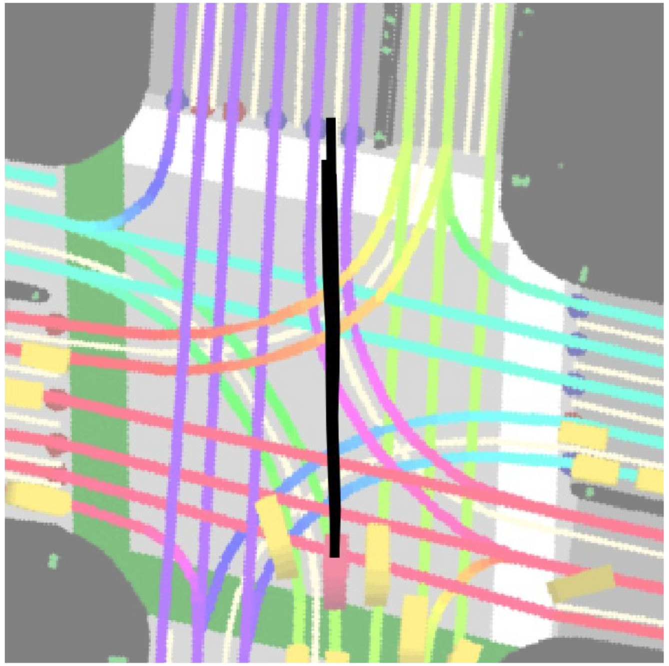
	mean over 3						min over 3		min over 20	
	ℓ_2 [m]		ORD [m]		ORFP [%]		ℓ_2 [m]		ℓ_2 [m]	
Method	**Avg**	**@4s**	**Avg**	**@4s**	**Avg**	**@4s**	**Avg**	**@4s**	**Avg**	**@4s**
no-scene-GAN	4.13	6.57	0.840	1.203	24.50	30.28	3.74	5.87	3.30	5.13
concat-scene-GAN	**2.35**	**5.62**	0.152	0.435	4.40	12.22	1.37	3.13	0.63	1.30
SC-GAN	2.44	5.86	**0.085**	**0.204**	**2.11**	**5.66**	**1.29**	**2.95**	**0.58**	**1.20**

Qualitative results

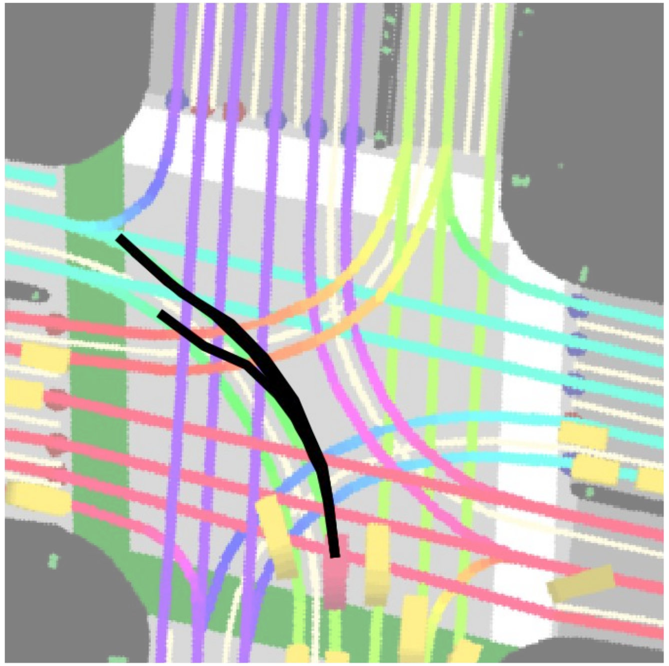
- SC-GAN predicts more scene-compliant trajectories



Ground-truth



concat-scene-GAN

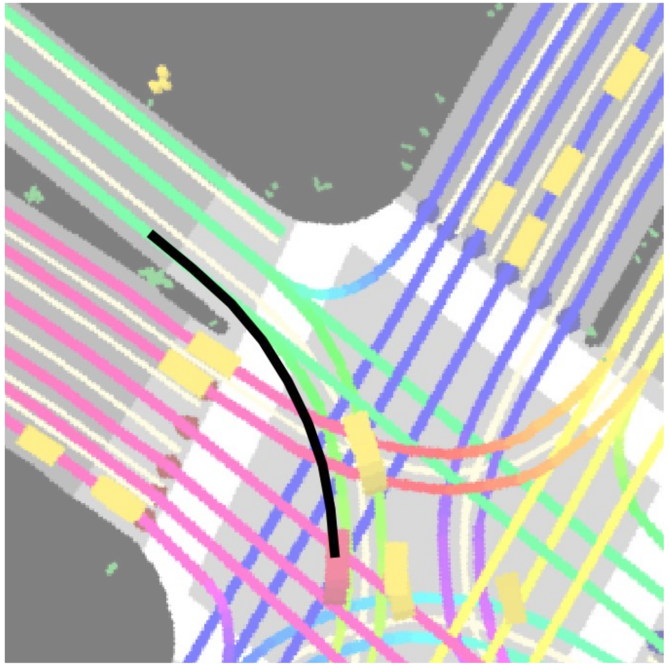


SC-GAN

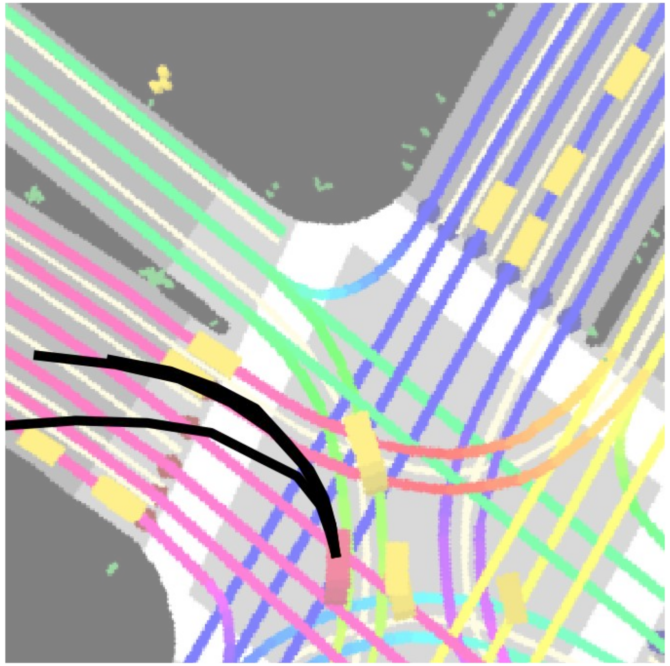
An actor on a left-turn-only lane can only turn left

Qualitative results

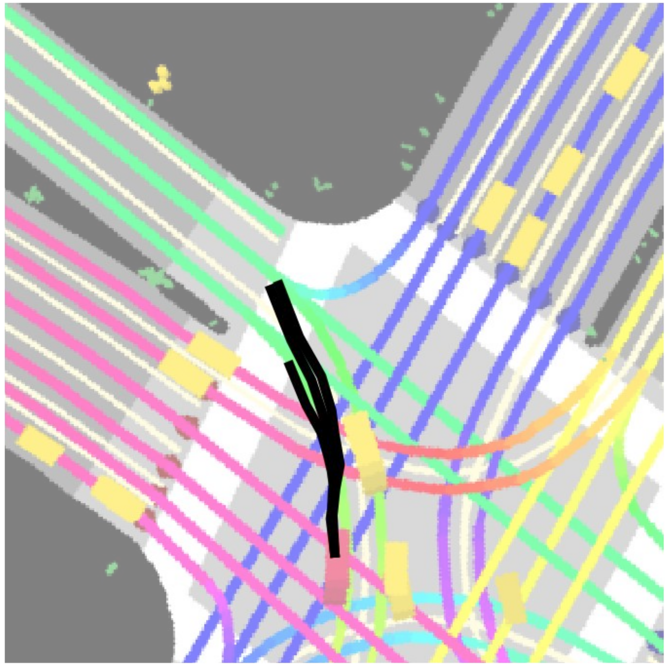
- SC-GAN predicts more scene-compliant trajectories



Ground-truth



concat-scene-GAN



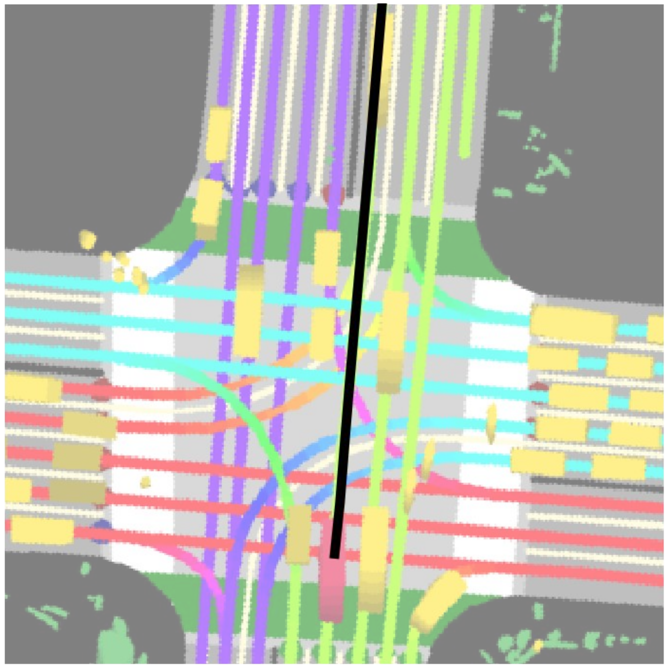
SC-GAN

An actor on a straight-only lane can only go straight

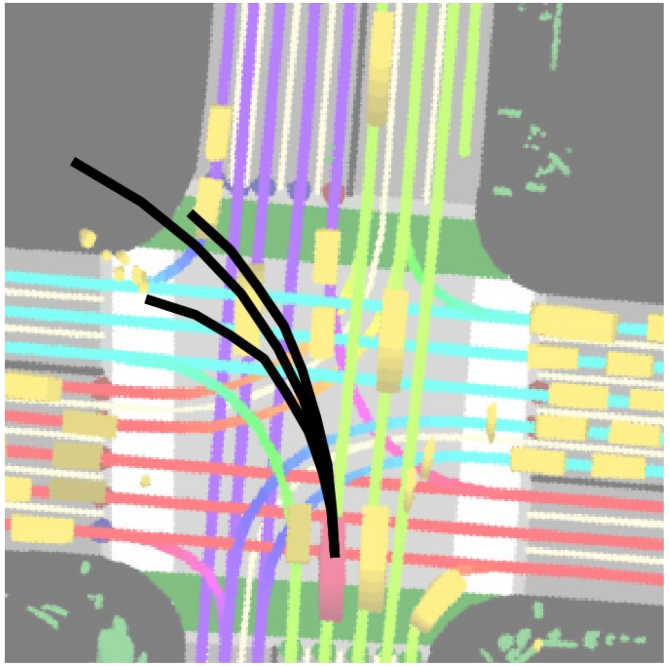
UBER ATG

Qualitative results

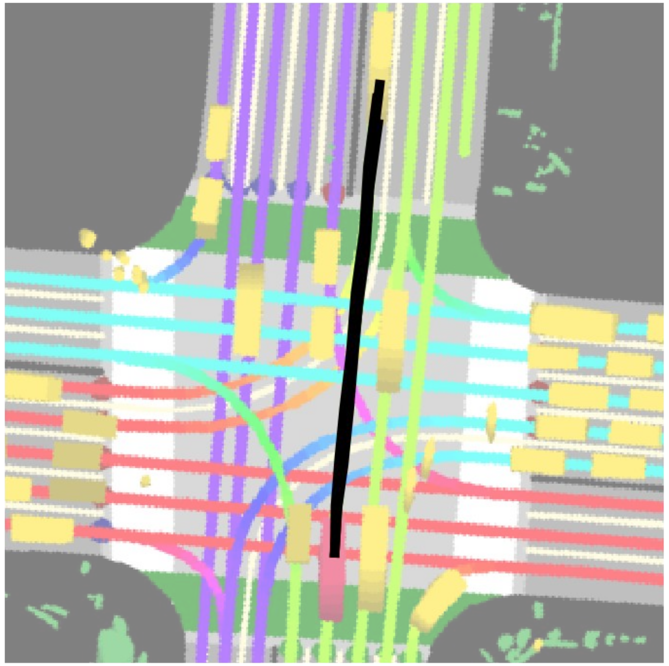
- SC-GAN predicts more scene-compliant trajectories



Ground-truth



concat-scene-GAN

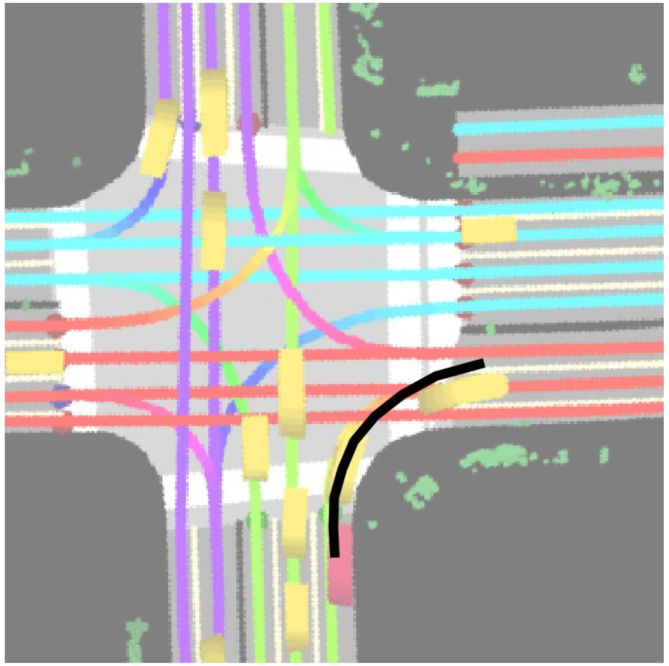


SC-GAN

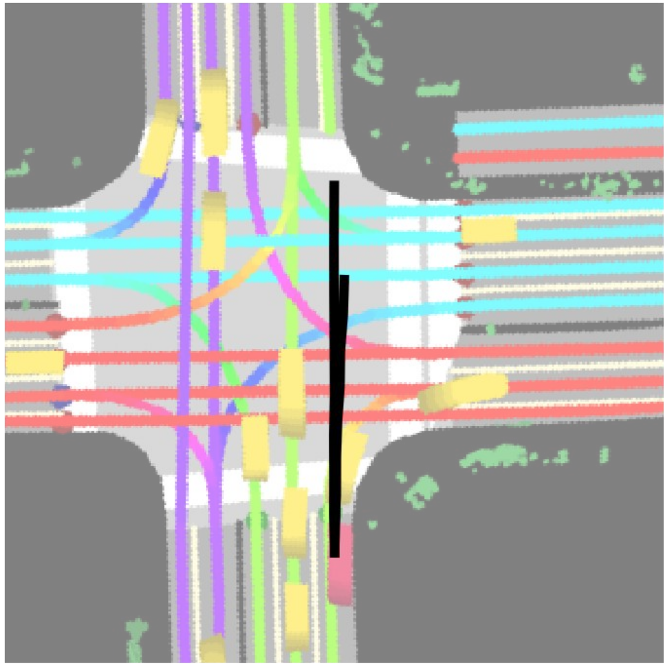
An actor on a right-turn only lane can only turn right

Qualitative results

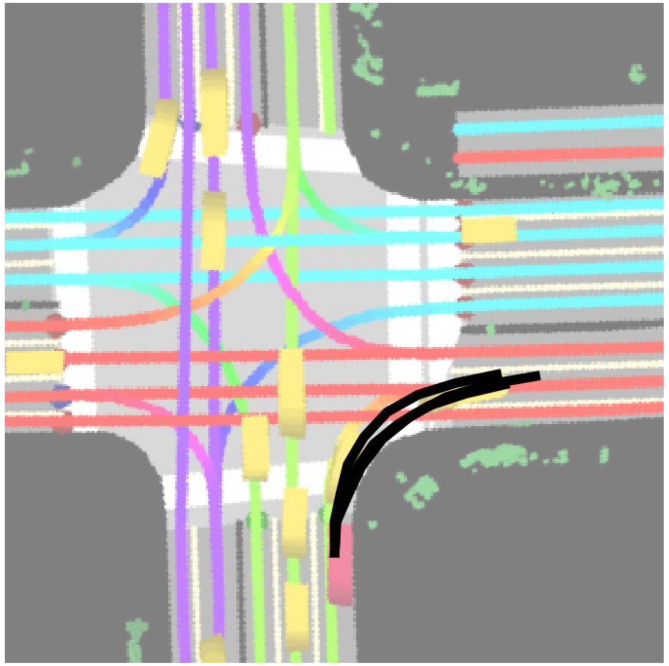
- SC-GAN predicts more scene-compliant trajectories



Ground-truth



concat-scene-GAN



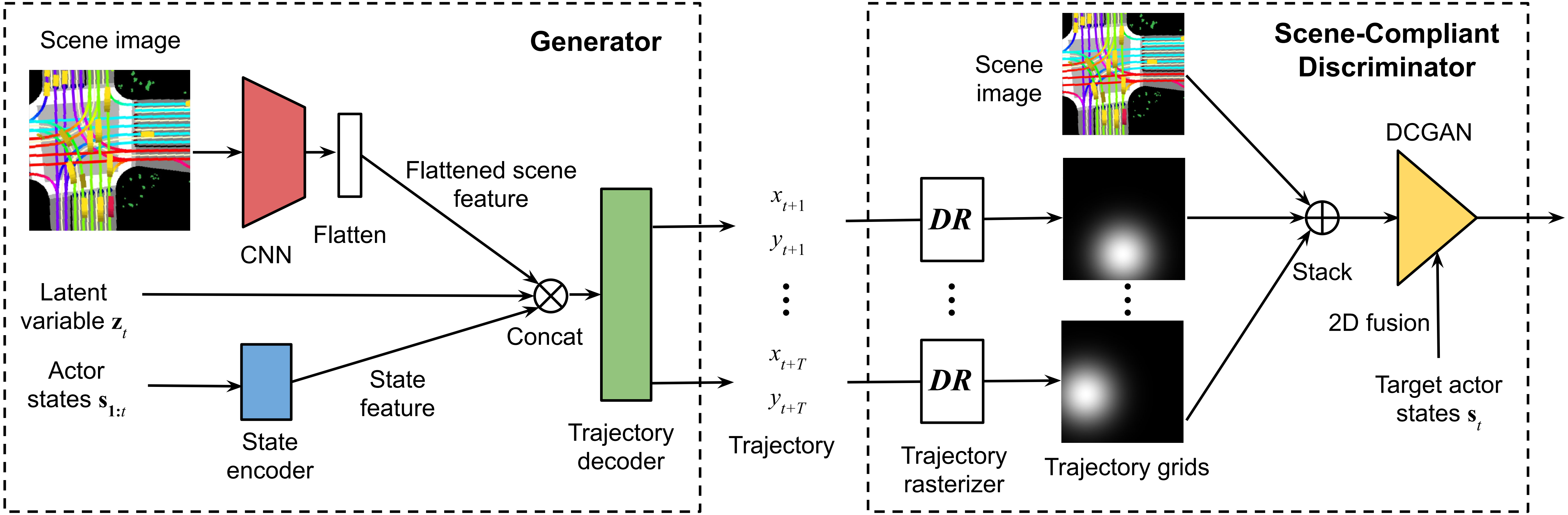
SC-GAN

The trajectories from SC-GAN follow the lanes better

UBER ATG

Conclusions

- We design *SC-GAN* that uses *Differentiable Trajectory Rasterization* (DR) to convert a trajectory into image representation
- SC-GAN is able to predict more scene-compliant trajectories
- DR is a generic component that can be used in other loss functions as well



IEEE Innovation at work

Topics ▾ Courses Events ▾ Podcast Contact Us ▾

How Data Is Being Used to Train Autonomous Vehicles to Navigate Roadways

Uber утверждает, что его AI позволяет беспилотным автомобилям с высокой точностью прогнозировать движение транспорта

16/04/2020 Финансовые новости Blockchain News 0

Uber: Verkehrsprognose mit hoher Genauigkeit

April 17, 2020 in Informationen zu autonomem Fahren

ZME SCIENCE ≡ Uber's AI lets driverless cars accurately predict the movement

Uber's AI lets driverless cars accurately predict the movement of other vehicles

by Alexandra Gerea — April 25, 2020 in Science, Tech, Technology

Uber声称其AI使无人驾驶汽车能够高精度预测交通流量

来源： · 2020-04-29 14:12:42

在本周于预印本服务器Arxiv.org上发表的一篇论文中，优步(Uber)的先进技术集团(ATG)的研究人员提出了一种AI技术，以改善自动驾驶汽车的交通运动预测。它直接适用于Uber自身正在开发的无人驾驶技术，该技术必须能够检测，跟踪和预测周围汽车的轨迹，以便安全地在公共道路上行驶。

众所周知，如果没有能力预测道路上其他驾驶员可能做出的决定，车辆将无法完全自动驾驶。在一个悲剧性的案例中，两年前，Uber自动驾驶原型机在亚利桑那州坦佩市撞死一名行人，部分原因是该车辆未能发现并避开受害者。ATG的研究是新颖的，因为它采用了生成对抗网络(GAN)来进行汽车轨迹的预测，而不是使用不太复杂的体系结构。该研究有望通过将预测的精度提高一个数量级来提高技术水平。 。

AI

Uber claims its AI enables driverless cars to predict traffic movement with high accuracy

KYLE WIGGERS @KYLE_L_WIGGERS APRIL 15, 2020 8:50 AM

Технологии

ИИ позволяет беспилотным автомобилям прогнозировать движение других машин

DEVELOPERS CORNER

HOW UBER OUTPERFORMED EXISTING GANS-BASED BASELINES IN SELF-DRIVING CARS

VĚDA A TECHNIKA HOLIDAY VIRUS TOTAL IT SLUŽBY ⌄ KONTAKT ⌄

English

How Uber Outperformed Existing GANs-Based Baselines In Self-Driving Cars

Auto News / Latest Auto News / Auto Technology

Uber develops AI; enables driverless cars to accurately predict other vehicles motion

The company has developed a Generative Adversarial Networks (GANs) to make car trajectory predictions as opposed to less complex architectures.

ETAuto • Updated: April 26, 2020, 18:19 IST

Nova IA do Uber para carros autônomos prevê trânsito com mais precisão

Rafael Rigues 16/04/2020 15h04

Begin backup slides

Quantitative results

- Compare against public baselines, trained with L2 loss

Method	mean@3		min@3		min@20	
	Avg	@4s	Avg	@4s	Avg	@4s
S-GAN [24]	3.01	7.66	2.36	5.94	1.93	4.77
S-LSTM [44]	2.93	5.17	-	-	-	-
SC-GAN-ℓ_2	**1.75**	**4.17**	**1.03**	**2.26**	**0.54**	**1.01**

Ablation study

-

Method	mean over 3 ℓ_2 [m] Avg	@4s	ORD [m] Avg	@4s	ORFP [%] Avg	@4s	min over 3 ℓ_2 [m] Avg	@4s	min over 20 ℓ_2 [m] Avg	@4s
SC-GAN-1channel	8.68	26.52	**0.040**	**0.033**	**0.98**	**1.23**	8.30	26.28	8.01	25.94
SC-GAN-MNet	3.82	11.18	0.723	3.068	7.21	22.43	2.62	8.15	1.82	6.08
SC-GAN-no-scene	3.79	7.28	0.58	1.16	18.38	33.62	3.52	6.76	2.27	4.61
SC-GAN	**2.44**	**5.86**	0.085	0.204	2.11	5.66	**1.29**	**2.95**	**0.58**	**1.20**