LaKD: Length-agnostic Knowledge Distillation for Trajectory Prediction with Any Length Observations

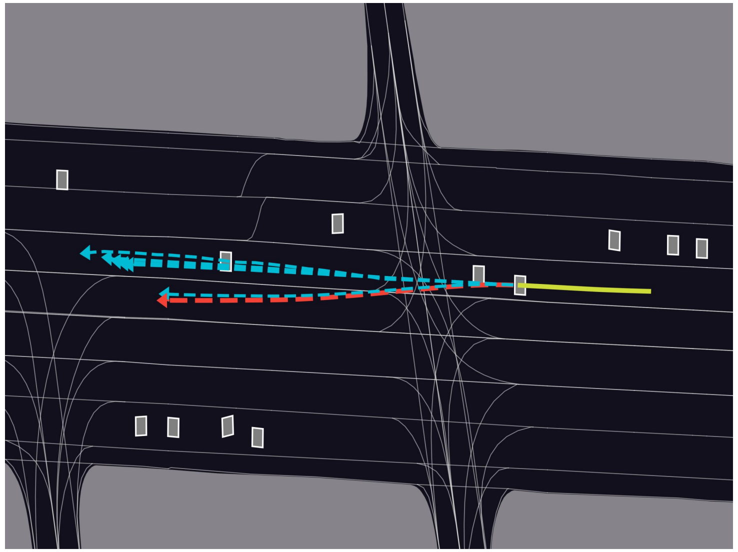
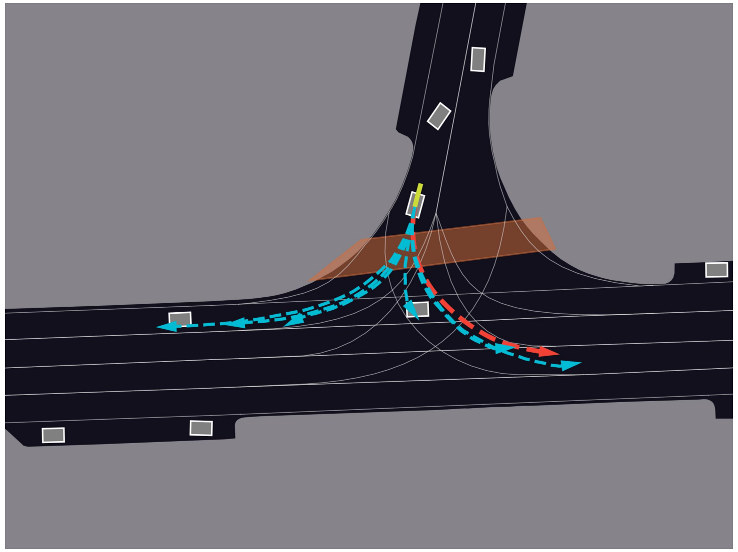
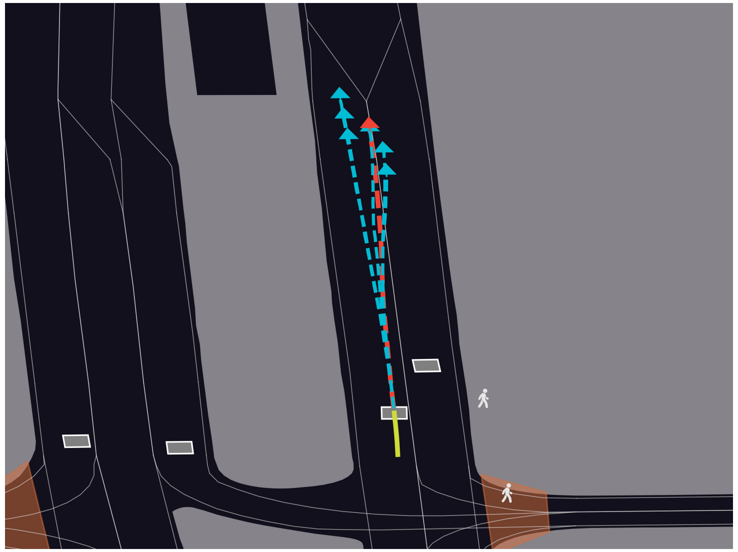
Yuhang Li[1], Changsheng Li[1], Ruilin Lv[1], Rongqing LI[1], Ye Yuan[1], Guoren Wang[1]

[1]Beijing institute of Technology

NEURAL INFORMATION PROCESSING SYSTEMS

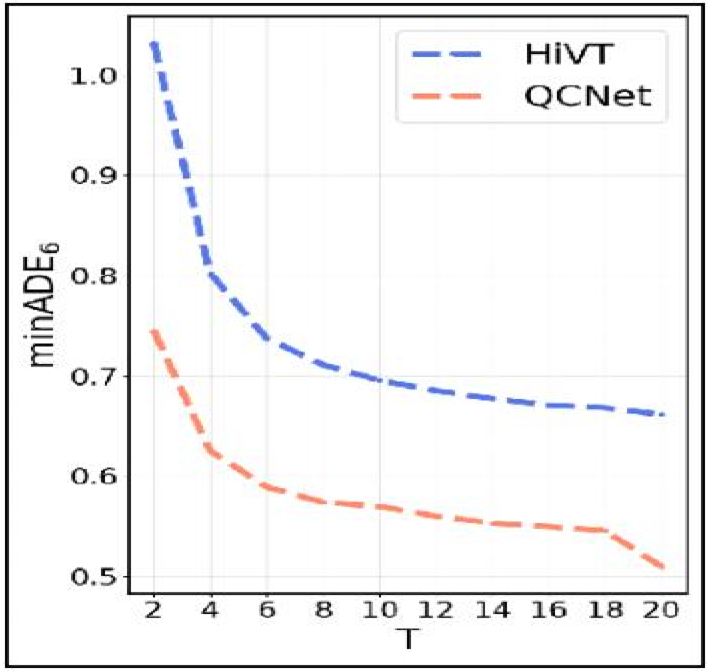
Background & Motivation

- Predicting the future trajectories of dynamic agents in traffic scenarios is a critical task in autonomous driving, enabling autonomous vehicles to make safe decisions .

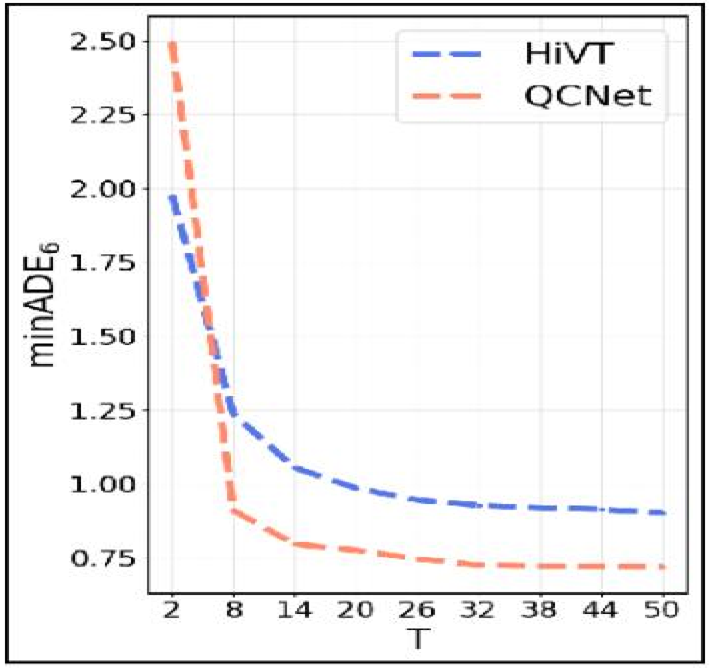




Background & Motivation

- In real-world scenarios, there is often insufficient time to gather an adequate number of observed trajectory points. This poses a new and challenging problem for trajectory prediction, requiring models to make accurate predictions based on observed trajectories of arbitrary lengths.



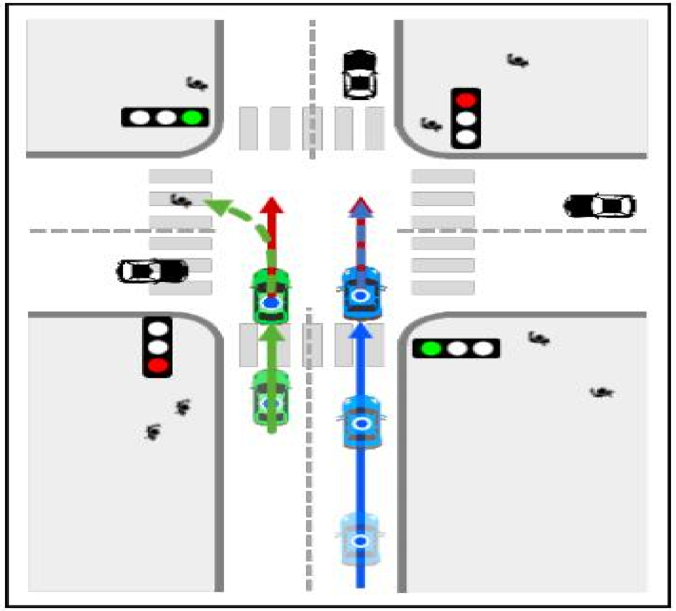
(a) Model performance on Argoverse 1.



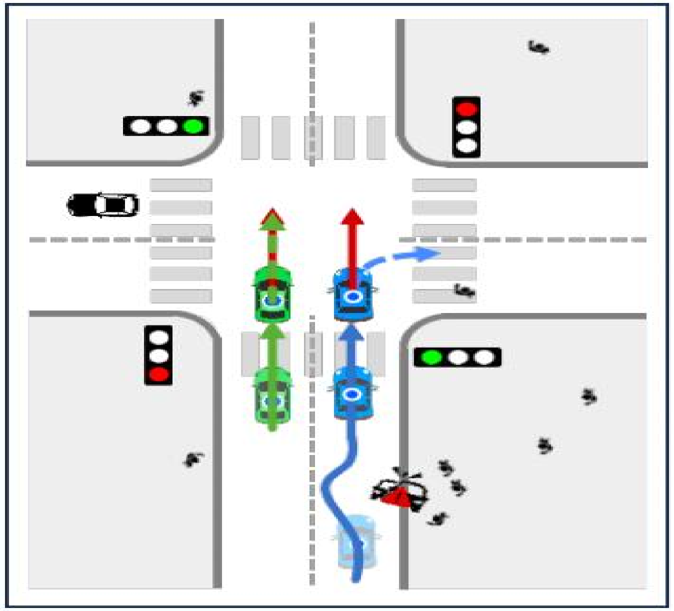
(b) Model performance on Argoverse 2.

Background & Motivation

- We note that longer trajectories often contain more temporal information, which can potentially lead to higher prediction accuracy compared to shorter trajectories. However, as the number of observed trajectory points increases, additional interference might be introduced.

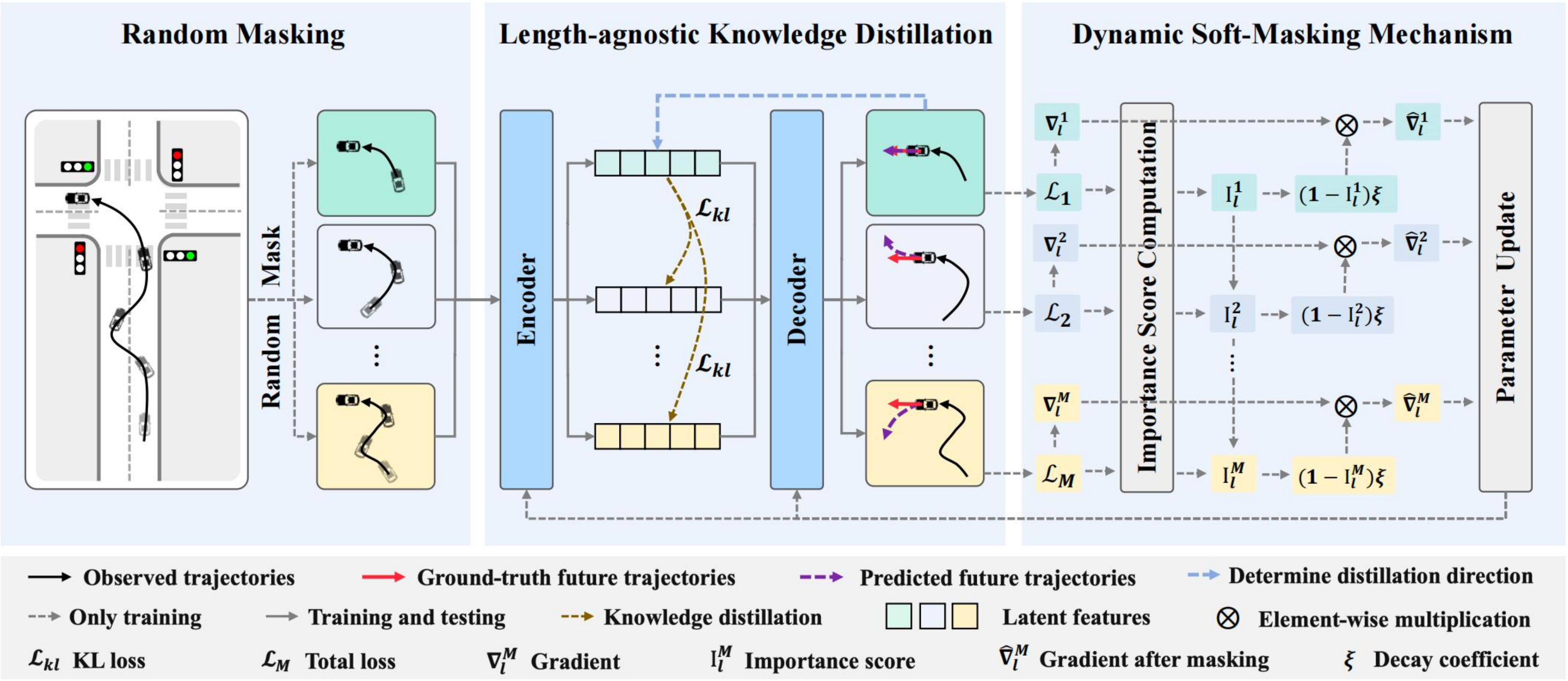


(c) Longer trajectory performs better.



(d) Shorter trajectory performs better.

Method


Random Masking
Length-agnostic Knowledge Distillation
Dynamic Soft-Masking Mechanism
Random Mask
Encoder
Decoder
$\mathcal{L}_{kl}$
$\mathcal{L}_{kl}$
∇_l^1
$\mathcal{L}_1$
∇_l^2
$\mathcal{L}_2$
∇_l^M
$\mathcal{L}_M$
Importance Score Computation
I_l^1
I_l^2
I_l^M
$(1 - I_l^1)\xi$
$(1 - I_l^2)\xi$
$(1 - I_l^M)\xi$
$\hat{\nabla}_l^1$
$\hat{\nabla}_l^2$
$\hat{\nabla}_l^M$
Parameter Update
Observed trajectories
Ground-truth future trajectories
Predicted future trajectories
Determine distillation direction
Only training
Training and testing
Knowledge distillation
Latent features
Element-wise multiplication
$\mathcal{L}_{kl}$ KL loss
$\mathcal{L}_M$ Total loss
∇_l^M Gradient
I_l^M Importance score
$\hat{\nabla}_l^M$ Gradient after masking
ξ Decay coefficient

Experiment

Table 1: Comparisons of different methods on Argoverse 1 and Argoverse 2, evaluated using $\text{min}\overline{\text{ADE}}$, $\text{min}\overline{\text{FDE}}$ and $\overline{\text{MR}}$ metrics. The best results are highlighted in bold.

Dataset	Methods	K=1			K=6		
		$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$	$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$
Argoverse 1	HiVT-Orig	1.4733	3.1834	0.5267	0.7255	1.0740	0.1124
	HiVT-RM	1.4189	3.0599	0.5104	0.7070	1.0447	0.1053
	HiVT-DTO	1.3999	3.0262	0.5056	0.7032	1.0350	0.1039
	HiVT-FLN	1.4011	3.0288	0.5051	0.7026	1.0325	0.1033
	HiVT-LaKD	**1.3317**	**2.8799**	**0.4901**	**0.6807**	**0.9864**	**0.0928**
Argoverse 1	QCNet-Orig	1.1656	2.4021	0.3860	0.5791	0.7399	0.0734
	QCNet-RM	1.0995	2.2550	0.3630	0.5684	0.7115	0.0703
	QCNet-DTO	1.0708	2.2303	0.3563	0.5418	0.6848	0.0671
	QCNet-FLN	1.0631	2.2083	0.3579	0.5411	0.6680	0.0671
	QCNet-LaKD	**0.9982**	**2.0718**	**0.3439**	**0.5240**	**0.6581**	**0.0640**
nuScenes	HiVT-Orig	3.5973	8.3062	0.8518	1.5289	2.8261	0.4377
	HiVT-RM	3.6580	8.4889	0.8647	1.5245	2.8068	0.4716
	HiVT-DTO	3.5860	8.2556	0.8514	1.5105	2.7379	0.4350
	HiVT-FLN	3.5640	8.1928	0.8488	1.5094	2.7489	0.4427
	HiVT-LaKD	**3.4296**	**7.8882**	**0.8369**	**1.4793**	**2.6934**	**0.4329**
nuScenes	QCNet-Orig	4.3134	9.7857	0.8588	1.4719	2.5831	0.4600
	QCNet-RM	4.1723	9.4672	0.8622	1.5255	2.6303	0.4611
	QCNet-DTO	4.1447	9.4552	0.8580	1.4653	2.5798	0.4317
	QCNet-FLN	4.1169	9.3639	0.8562	1.4676	2.5448	0.4344
	QCNet-LaKD	**4.0663**	**9.2524**	**0.8523**	**1.4594**	**2.4901**	**0.4023**
Argoverse 2	HiVT-Orig	2.5502	6.5586	0.7455	1.0561	2.1093	0.3275
	HiVT-RM	2.2848	6.0548	0.7249	0.9457	1.9283	0.2994
	HiVT-DTO	2.2769	6.0548	0.7275	0.9324	1.8946	0.2903
	HiVT-FLN	2.2786	6.0464	0.7240	0.9287	1.8838	0.2891
	HiVT-LaKD	**2.2066**	**5.8769**	**0.7161**	**0.9183**	**1.8686**	**0.2791**
Argoverse 2	QCNet-Orig	2.1006	5.2219	0.6299	0.8339	1.3849	0.1884
	QCNet-RM	1.7452	4.4404	0.5957	0.7508	1.3184	0.1671
	QCNet-DTO	1.7713	4.4900	0.5979	0.7454	1.2924	0.1671
	QCNet-FLN	1.6940	4.2373	0.5808	0.7370	1.2595	0.1596
	QCNet-LaKD	**1.6574**	**4.1505**	**0.5753**	**0.7258**	**1.2420**	**0.1555**

Table 2: Ablation study of our method on the Argoverse 1 dataset.

RM	LaKD	DSM	K=1			K=6		
			$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$	$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$
			1.4733	3.1834	0.5267	0.7255	1.0740	0.1124
✓			1.4189	3.0599	0.5104	0.7070	1.0447	0.1053
✓	✓		1.3619	2.9511	0.5051	0.6851	0.9965	0.0948
✓	✓	✓	**1.3317**	**2.8799**	**0.4901**	**0.6807**	**0.9864**	**0.0927**

Table 3: Analysis of our method with different M on the Argoverse 1 dataset.

M	K=1			K=6		
	$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$	$\text{min}\overline{\text{ADE}}$	$\text{min}\overline{\text{FDE}}$	$\overline{\text{MR}}$
2	1.3457	2.9116	0.4943	0.6808	0.9863	0.0930
3	**1.3317**	**2.8799**	**0.4901**	**0.6807**	**0.9864**	**0.0928**
4	1.3414	2.9017	0.4938	0.6814	0.9867	0.0934
5	1.3563	2.9359	0.5010	0.6851	0.9973	0.0961
6	1.3486	2.9198	0.4977	0.6878	1.0025	0.0943

Experiment

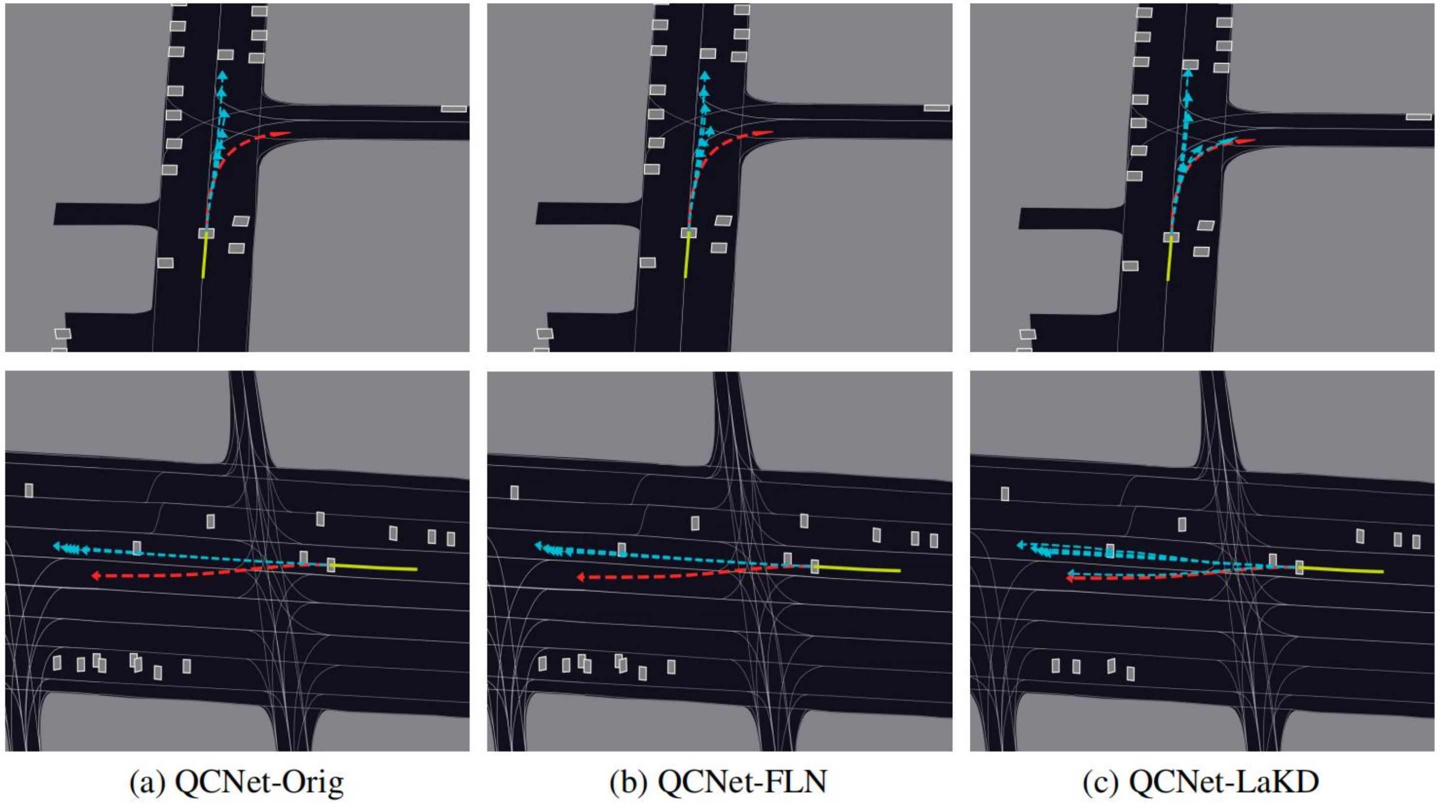


Figure 3: Qualitative results on the Argoverse 2 dataset using (a) QCNet-Orig, (b) QCNet-FLN, and (c) QCNet-LaKD. The observed trajectories, ground-truth trajectories and predicted trajectories are shown in green, red and blue, respectively. Our predicted future trajectories are closer to the ground-truth, compared to other methods.

Thanks for your attention