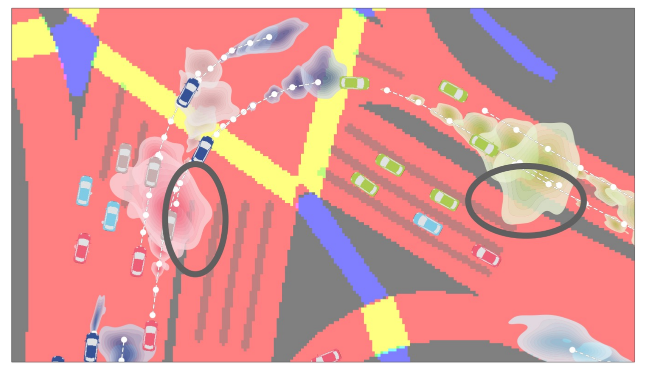
Motion Forecasting with Unlikelihood Training in Continuous Space

Deyao Zhu[1], Mohamed Zahran[1,2], Li Erran Li[3,4], Mohamed Elhoseiny[1]
[1]KAUST, [2]Udacity, [3]AWS AI, Amazon and [4]Columbia University

Limitation of MLE on Motion Forecasting

- Maximum likelihood estimation (MLE) encourages the prediction to be geometrically close to the ground truth.
- However, maintaining the geometrical nearness only is not good enough. Trajectories can be close to the ground truth geometrically but unlikely to happen due to context violation.



- We propose an unlikelihood training [1,2] objective that can explicitly encourage the model to follow the contextual information.

Unlikelihood Training

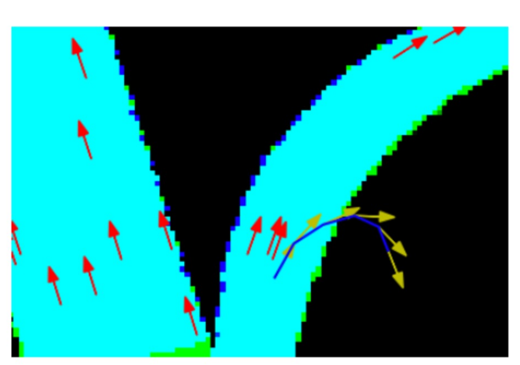
- Design an unlikelihood loss L_{unlike} to minimize the likelihood of context-violated (negative) trajectories $\boldsymbol{Y}_{i,neg}$ besides the original loss L_{orig}.
- Negative trajectories are sampled from the model predicted distribution and selected out by a context checker.
- Our loss can be combined with models that predict a trajectory distribution $p_\theta(\boldsymbol{Y}_i|\boldsymbol{X}_i)$ as output.

$$L_{\text{unlike}} = \mathbb{E}_{\boldsymbol{X}_i,\sim\mathbb{D},\boldsymbol{Y}_{i,neg}\sim p_{\text{neg}}(\boldsymbol{Y}_i|\boldsymbol{X}_i)}[\log p_\theta(\boldsymbol{Y}_{i,neg} \mid \boldsymbol{X}_i)]$$

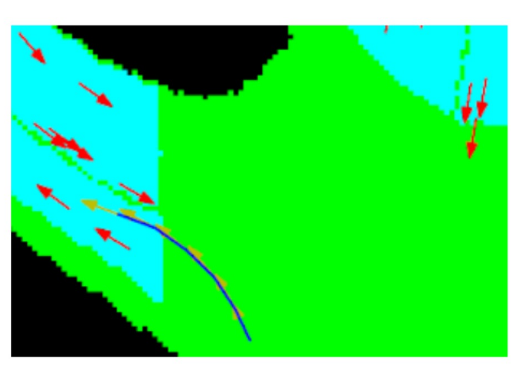
$$L = L_{\text{orig}} + \gamma L_{\text{unlike}}$$

Context Checker

- A map-based checker to judge whether a given trajectory is compliant with context including the drivable region and the driving direction.
- Checker is disable when the ground truth trajectory violates the context to allow context-violated prediction when necessary.



(a) negative case



(b) positive case

Gradient Analysis

- Assume a single-mode prediction case that the future position $y_{gt,t}$ at time step t is modeled by a simple Gaussian distribution $N\ (y_t\ |\hat{\mu}_t\ ,\hat{\sigma}_t\ \boldsymbol{I})$.
- With a single negative position $y_{neg,t}$, the gradients of a simple learning objective with an MLE loss and an unlikelihood loss w.r.t. $\hat{\mu}_t$ and $\hat{\sigma}_t$ are

$$\frac{\partial L_t}{\partial \hat{\boldsymbol{\mu}}_t} = -\frac{1}{\hat{\sigma}_t^2}((\boldsymbol{y}_{gt,t} - \hat{\boldsymbol{\mu}}_t) + (\hat{\boldsymbol{\mu}}_t - \boldsymbol{y}_{neg,t}))$$

$$\frac{\partial L}{\partial \hat{\sigma}_t} = -\frac{1}{\hat{\sigma}_t^3}(||\boldsymbol{y}_{gt,t} - \hat{\boldsymbol{\mu}}_t||^2 - ||\boldsymbol{y}_{neg,t} - \hat{\boldsymbol{\mu}}_t||^2)$$

- The gradients push $\hat{\mu}_t$ towards $y_{gt,t}$ and away from $y_{neg,t}$. When $y_{gt,t}$ is closer to $\hat{\mu}_t$ than $y_{neg,t}$, $\hat{\sigma}_t$ will shrink to exclude the negative region and become a better estimation to the true data distribution.

Experimental Results

- When incorporating our method with Trajectron++ [5] in nuScenes [3] dataset, we avoids 16% context-violated prediction and improves the prediction accuracy performance by more than 8%.

Model	FDE-Full	ADE-Full	Context-Violation-Rate	minFDE10	minADE10
Ground Truth	-	-	5.44%	-	-
Trajectron++	2.74±0.10	1.04±0.05	10.59%±0.54%	1.68±0.05	0.68±0.02
Trajectron++ with L_{unlike}	**2.51±0.06**	**0.95±0.03**	**8.85% ± 0.32%**	**1.52±0.07**	**0.61±0.02**
Relative Improvement	+8%	+9%	+ 16%	+10%	+10%

- When combining our method with Gaussian LaneGCN, a distribution variant of LaneGCN [6], in Argoverse [4] dataset, we avoids 56% of context-violated prediction. ADE-Full and FDE-Full are reduced by 8% and 6%, respectively.

Model	ADE-Full	FDE-Full	Context-Vio.	ADE-1	FDE-1
Ground Truth	-	-	0.85%	-	-
Gaussian LaneGCN	1.81	3.49	9.0%	1.38	3.03
Gaussian LaneGCN with L_{unlike}	**1.67**	**3.29**	**4.0%**	1.38	**3.01**
Relative Improvement	+8%	+6%	+ 56%	0%	+1%

- Experimental results show that our unlikelihood loss help improve the quality of the predicted distribution by making it more accurate and reducing the context-violated prediction.

References

[1] S. Welleck, I. Kulikov, S. Roller, E. Dinan, K. Cho, and J. Weston. Neural text generation with unlikelihood training. arXiv preprint arXiv:1908.04319, 2019.
[2] M. Li, S. Roller, I. Kulikov, S. Welleck, Y. Boureau, K. Cho, and J. Weston. Don't Say That! Making Inconsistent Dialogue Unlikely with Unlikelihood Training. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020
[3] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuScenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[4] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019
[5] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. Trajectron++: Multi-agent generative trajectory forecasting with heterogeneous data for control. In Proceedings of the European Conference on Computer Vision (ECCV), 2020.
[6] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun. Learning lane graph representations for motion forecasting. In Proceedings of the European Conference on Computer Vision (ECCV), 2020.

Ablation Study

- **The Influence of Our Loss Without Map Input:** The effect of L_{unlike} without map input to the model on nuScenes. Results indicate that model can obtain contextual information from L_{unlike} even without map input.

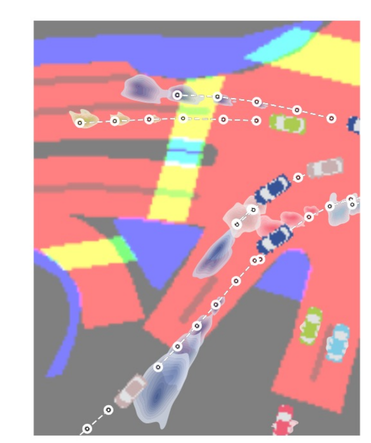
Map Input	L_{unlike}	FDE-Full	ADE-Full
-	-	2.89±0.11	1.08±0.03
-	+	**2.67±0.08**	**1.01±0.02**
+	-	2.74±0.10	1.04±0.05
+	+	**2.51±0.06**	**0.95±0.03**

- **Remove Context-Violated Prediction Directly:** Our method performs better than simply removing the context-violated prediction and still allow context-violation when nessary. This indicates that our unlikelihood loss helps models to understand the context better.

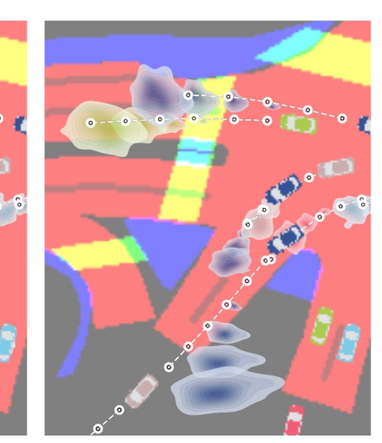
Method	ADE-Full	FDE-Full
Trajectron++	1.04 ± 0.05	2.74 ± 0.10
Trajectron++ with Removing Violation	0.98 ± 0.04	2.55 ± 0.10
Trajectron++ with L_{unlike}	**0.95 ± 0.03**	**2.51 ± 0.06**

Qualitative results

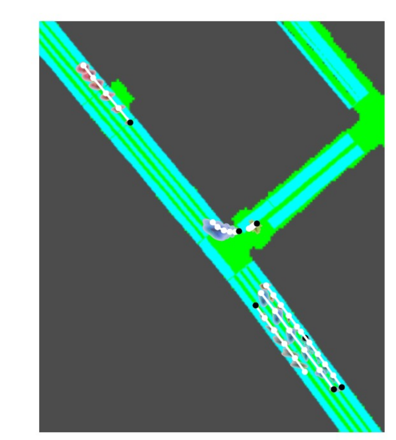
- Qualitative results on nuScenes with Trajectron++ (left) and on Argoverse with Gaussian LaneGCN (right). White points denote the ground truth. Our loss helps reduce the context-violation prediction.



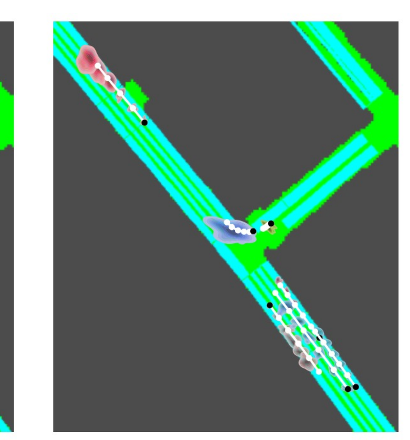
(a) Trajectron++ + Unlikelihood (Ours)



(b) Trajectron++



(a) Gaussian LaneGCN + Unlikelihood (Ours)



(b) Gaussian LaneGCN

Summary

- We propose continuous unlikelihood training for vehicle motion forecasting that encourages models to use contextual information by minimizing the likelihood of context-violated trajectories.
- Our method can be incorporated into state-of-the-art models that predict the future as distributions.
- Experimental results show that our method can improve the quality of the predicted distribution by avoiding maximally 56% context-violated prediction and improving 9% prediction performance.