SAS-GAN: Self-Attention Social GAN on Trajectory Forecasting of Multiple Agents in Autonomous Vehicles

M.H. Nazeri[1], M. Bohlouli[2] and M. Bohlouli ✉[1];

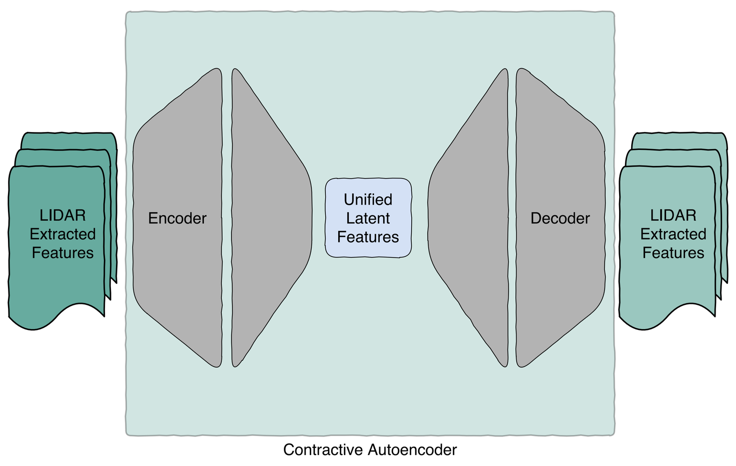
[1]Institute for Advanced Studies in Basic Sciences, Iran; [2]Amir Kabir University of Technology, Iran
Correspondence to: me@bohlouli.com

Poster DOI: xx.xxxx/
github.com/mhnazeri/sasgan

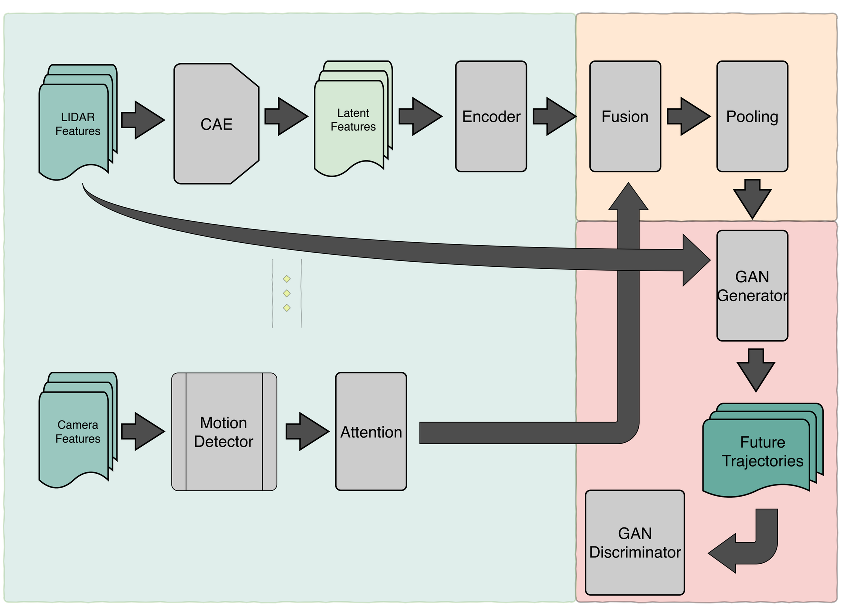
Forecasting the future trajectories of vehicles is a crucial ability for Autonomous Vehicles (AV) to be accepted by society. Having this ability, AVs can understand human behaviors and avoid disastrous situations such as collisions. Our team proposes a novel approach by combining physical attributes of the agents along with their history, which is inspired from the architecture of Social-GAN. To this, it is supposed that the perception system of the AV works as expected. Transferring multi-modal physical features into uniform feature, strengthens accurate reasoning of AVs about future trajectories of various agents in dense traffics. According to the experimental results, proposed approach outperforms state-of-the-art solutions by resulting in x% of forecasting accuracy.

SAS-GAN: Architecture Overview

SAS-GAN uses two different neural networks. The first network, contractive auto-encoder (CAE), is responsible for unifying features and dimensionality reduction of the data (see Fig. 1). CAE trained before and independent of the SAS-GAN. After being trained, it become part of the SAS-GAN. The second part is the main part, SAS-GAN itself (see Fig. 2).



Overview of CAE. It unify and reduce dimensions in two stages.



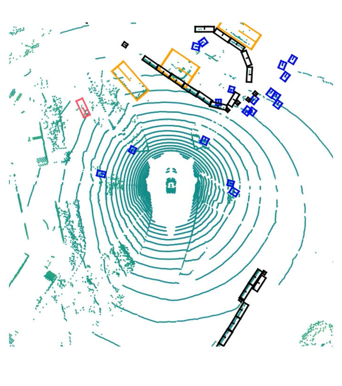
Overview of SAS-GAN architecture. It consists of three parts: feature extractor, social pooling, and generator

SAS-GAN: Trajectory Forecasting

Forecasting future trajectories achieved by understanding the environment, where every agent is located, observe the changes for a while (ex. two seconds) and based on the past observation predict future trajectories.

SAS-GAN: Sensory Inputs

SAS-GAN works on two different types sensory inputs, *LIDAR data* and *camera images*.



Sample LIDAR input to the network. SAS-GAN uses a trained contractive auto-encoder (CAE) to transform raw high-dmensional data to low-dimensional and unified feature vector, that also considers physical attributes of the agent.

Sample camera input to the network. SAS-GAN uses motion detector module to grasp the motion across consecutive frames to reduce the data dimension. Then uses an attention mechanism to extract points of interest in the image which maybe a vehicle far away but coming from the opposite direction.

SAS-GAN: Workflow

After receiving two types of inputs for two seconds, SAS-GAN pool all the information at hand in a latent space where it has control over it. Concat history trajectory with a little bit of noise z to the latent feature vector to preserve the consistency between the past and future. Feed this vector to the LSTM generator of the GAN to produce diverse prediction of the future. The discriminator is responsible for authenticating the acceptability of the generated trajectories.

SAS-GAN: Advantages

The advantages that SAS-GAN has over previous works can be into three parts: *1) works in dense traffics even with 100 agents present at the scene. 2) uses contextual information of the environment. 3) model interactions between different types of agents (ex. the interaction between a bus and a bus is different of a bus and a vehicle).*

SAS-GAN: Future Works

SAS-GAN is not flawless. The biggest imperfection is, it relies on the perception system being competence. Making SAS-GAN to rely on itself for feature extraction from raw sensory inputs can be a huge step towards better planner systems for AVs.

References

[1] Waymo Open dataset: An autonomous driving dataset, 2019.

[2] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human trajectory prediction in crowded spaces. jun 2016.

[3] Claudine Badue, Ranik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius Brito Cardoso, Avelino Forechi, Luan Ferreira Reis Jesus, Rodrigo Ferreira Berriel, Thiago Meireles Paixão, Filipe Wall Mutz, Thiago Oliveira-Santos, and Alberto Ferreira de Souza. Self-driving cars: A survey. *CoRR*, abs/1901.04407, 2019.

[4] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. *arXiv preprint arXiv:1812.03079*, 2018.

[5] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, and Karol Zieba. End to end learning for self-driving cars. *CoRR*, abs/1604.07316, 2016.

[6] Thibault Buhet, Emilie Wirbel, and Xavier Perrotton. Conditional vehicle trajectories prediction in carla urban environment. *ArXiv*, abs/1909.00792, 2019.

[7] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[8] Sergio Casas, Wenjie Luo, and Raquel Urtasun. Intentnet: Learning to predict intention from raw sensor data. In *Conference on Robot Learning*, pages 947–956, 2018.

[9] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8483–8492, 2019.

[10] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[11] Yilun Chen, Chiyu Dong, Praveen Palanisamy, Priyantha Mudalige, Katharina Muelling, and John M Dolan. Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[12] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350. IEEE, 2017.

[13] Nachiket Deo and Mohan M Trivedi. Multi-modal trajectory prediction of surrounding vehicles with maneuver based lstms. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1179–1184. IEEE, 2018.

[14] Nachiket Deo and Mohan Manubhai Trivedi. Convolutional social pooling for vehicle trajectory prediction. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1549–15498, 2018.

[15] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[16] D. González, J. Pérez, V. Milanés, and F. Nashashibi. A review of motion planning techniques for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):1135–1145, April 2016.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[18] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social GAN: Socially acceptable trajectories with generative adversarial networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, jun 2018.

[19] Joey Hong, Benjamin Sapp, and James Philbin. Rules of the road: Predicting driving behavior with a convolutional model of semantic interactions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8454–8462, 2019.

[20] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet. Lyft level 5 av dataset 2019. urlhttps://level5.lyft.com/dataset/, 2019.

[21] Wonhui Kim, Manikandasriram Srinivasan Ramanagopal, Charles Barto, Ming-Yuan Yu, Karl Rosaen, Nick Goumas, Ram Vasudevan, and Matthew Johnson-Roberson. Pedx: Benchmark dataset for metric 3-d pose estimation of pedestrians in complex urban intersections. *IEEE Robotics and Automation Letters*, 4(2):1940–1947, 2019.

[22] Kris M. Kitani, Brian D. Ziebart, James Andrew Bagnell, and Martial Hebert. Activity forecasting. pages 201–214, 2012.

[23] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 137–146. Curran Associates, Inc., 2019.

[24] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B. Choy, Philip H. S. Torr, and Manmohan Chandraker. DESIRE: Distant future prediction in dynamic scenes with interacting agents. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jul 2017.

[25] Wenjie Luo, Bin Yang, and Raquel Urtasun. Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 3569–3577, 2018.

[26] Yuexin Ma, Xinge Zhu, Sibo Zhang, Ruigang Yang, Wenping Wang, and Dinesh Manocha. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents, 2018.

[27] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[28] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. *arXiv preprint arXiv:1903.01568*, 2019.

[29] Sujoy Paul, Jeroen Vanbaar, and Amit Roy-Chowdhury. Learning from trajectories via subgoal discovery. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8409–8419. Curran Associates, Inc., 2019.

[30] Nicholas Rhinehart. *Jointly Forecasting and Controlling Behavior by Learning from High-Dimensional Data*. PhD thesis, Pittsburgh, PA, September 2019.

[31] Nicholas Rhinehart, Kris M. Kitani, and Paul Vernaza. r2p2: A ReparameteRized pushforward policy for diverse, precise generative path forecasting. In *Computer Vision – ECCV 2018*, pages 794–811. Springer International Publishing, 2018.

[32] Nicholas Rhinehart, Rowan McAllister, Kris Kitani, and Sergey Levine. PRECOG: PREdiction Conditioned On Goals in Visual Multi-Agent Settings. *arXiv e-prints*, page arXiv:1905.01296, May 2019.

[33] Nicholas Rhinehart, Rowan McAllister, and Sergey Levine. Deep imitative models for flexible inference, planning, and control. *arXiv preprint arXiv:1810.06544*, 2018.

[34] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. Human motion trajectory prediction: A survey. *arXiv preprint arXiv:1905.06113*, 2019.

[35] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1349–1358, 2019.

[36] Edward Schmerling, Karen Leung, Wolf Vollprecht, and Marco Pavone. Multimodal probabilistic model-based planning for human-robot interaction. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

[37] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7. IEEE, 2018.

[38] Wenyuan Zeng, Wenjie Luo, Simon Suo, Abbas Sadat, Bin Yang, Sergio Casas, and Raquel Urtasun. End-to-end interpretable neural motion planner. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8660–8669, 2019.

[39] Han Zhang, Ian J. Goodfellow, Dimitris N. Metaxas, and Augustus Odena. Self-attention generative adversarial networks. *ArXiv*, abs/1805.08318, 2018.

[40] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. Multi-agent tensor fusion for contextual trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12126–12134, 2019.