# References

[1] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, 2013, v. 32, n. 11, pp. 1231–1237.

[2] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.

[3] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.

[4] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," *arXiv preprint arXiv:1603.00831*, 2016.

[5] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2015, v. 16, n. 6, pp. 3017–3032.

[6] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, 2016, v. 4, n. 1, pp. 15–24.

[7] H. Bradshaw-Martin and C. Easton, "Autonomous or 'driverless' cars and disability: a legal and ethical analysis." *European Journal of Current Legal Issues*, 2014, v. 20, n. 3.

[8] T. Litman, *Autonomous vehicle implementation predictions.* Victoria Transport Policy Institute Victoria, Canada, 2017.

[9] S. Tsugawa, S. Kato, and K. Aoki, "An automated truck platoon for energy saving," in *2011 IEEE/RSJ international conference on intelligent robots and systems.* IEEE, 2011, pp. 4109–4114.

[10] P. Fernandes and U. Nunes, "Platooning with ivc-enabled autonomous vehicles: Strategies to mitigate communication delays, improve safety and traffic flow," *IEEE Transactions on Intelligent Transportation Systems*, 2012, v. 13, n. 1, pp. 91–106.

[11] M. Y. Wang, H. W. Ding, Y. F. Zhao, Z. G. Liang, and L. Z. Zhu, "The application of the exhaustive polling theory in intelligent traffic system," in *MATEC Web of Conferences*, v. 44. EDP Sciences, 2016, p. 02013.

[12] S. A. Fayazi, A. Vahidi, and A. Luckow, "Optimal scheduling of autonomous vehicle arrivals at intelligent intersections via milp," in *2017 American control conference (ACC).* IEEE, 2017, pp. 4920–4925.

[13] I. H. Zohdy, R. K. Kamalanathsharma, and H. Rakha, "Intersection management for autonomous vehicles using icacc," in *2012 15th international IEEE conference on intelligent transportation systems.* IEEE, 2012, pp. 1109–1114.

[14] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transportation Research Part C: Emerging Technologies*, 2016, v. 71, pp. 143–163.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of*

*the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[16] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, 2013, v. 104, n. 2, pp. 154–171.

[17] T. F. Gonzalez, "Handbook of approximation algorithms and metaheuristics," *Handbook of Approximation Algorithms and Metaheuristics*, 2007, pp. 1–1432.

[18] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," *Proceedings of the IEEE International Conference on Computer Vision*, 2005, v. II, pp. 1458–1465.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *Proceedings of the IEEE International Conference on Computer Vision*, 2015, v. 2015 Inter, pp. 1026–1034.

[20] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, v. 8689 LNCS, n. PART 1, pp. 818–833.

[21] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, 2015, v. 2015 Inter, pp. 1440–1448.

[22] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, v. 39, n. 4, pp. 640–651.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, v. 39, n. 6, pp. 1137–1149.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition.* IEEE, 2009, pp. 248–255.

[25] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, pp. 1–14. [Online]. Available: http://arxiv.org/abs/1409.1556

[26] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 2014, pp. 675–678.

[27] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision*, 2017, v. 2017-Octob, pp. 2999–3007.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, v. 2016-Decem, pp. 770–778.

[29] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, 2016, v. 29.

[30] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 761–769.

[31] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, 2017, v. 2017-Octob, pp. 2980–2988.

[32] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, v. 2017-Janua, pp. 5987–5995.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, v. 07-12-June, pp. 1–9.

[34] M. Lin, Q. Chen, and S. Yan, "Network in network," *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014, pp. 1–10.

[35] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[36] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020. [Online]. Available: http://arxiv.org/abs/2004.10934

[37] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, v. 2017-Janua, pp. 6517–6525.

[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, v. 2016-Decem, pp. 779–788.

[39] S. Liu, X. Qi, J. Shi, H. Zhang, and J. Jia, "Multi-scale Patch Aggregation (MPA) for Simultaneous Detection and Segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, v. 2016-Decem, n. 413113, pp. 3141–3149.

[40] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2155–2162.

[41] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

[42] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: An on-line lexical database," *International Journal of Lexicography*, 1990, v. 3, n. 4, pp. 235–244.

[43] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as Points," 2019. [Online]. Available: http://arxiv.org/abs/1904.07850

[44] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, v. 9912 LNCS, pp. 483–499.

[45] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 781–10 790.

[46] G. Ghiasi, T. Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, v. 2019-June, pp. 7029–7038.

[47] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th International Conference on Machine Learning, ICML 2019*, 2019, v. 2019-June, pp. 10 691–10 700.

[48] B. Zoph and Q. V. Le, "Searching for activation functions," *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 2018, pp. 1–13.

[49] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020, n. 2, pp. 12 993–13 000.

[50] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017, pp. 1–16.

[51] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," 2019. [Online]. Available: http://arxiv.org/abs/1910.03151

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, v. 37, n. 9, pp. 1904–1916.

[53] D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function," 2019. [Online]. Available: http://arxiv.org/abs/1908.08681

[54] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8759–8768.

[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017, v. 2017-Decem, n. Nips, pp. 5999–6009.

[56] H. Fausk and D. C. Isaksen, "T-model structures," *Homology, Homotopy and Applications*, 2007, v. 9, n. 1, pp. 399–438.

[57] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, v. 1, n. Mlm, pp. 4171–4186.

[58] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, 2020, v. 21, pp. 1–67.

[59] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Computing Surveys*, 2022, pp. 1–30.

[60] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021. [Online]. Available: http://arxiv.org/abs/2103.14030

[61] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *European conference on computer vision*. Springer, 2016, pp. 354–370.

[62] X. Hu, X. Xu, Y. Xiao, H. Chen, S. He, J. Qin, and P.-A. Heng, "SINet: A scale-insensitive convolutional neural network for fast vehicle detection," *IEEE transactions on intelligent transportation systems*, 2018, v. 20, n. 3, pp. 1010–1019.

[63] Q. Zhao, Y. Wang, T. Sheng, and Z. Tang, "CFENet: Comprehensive feature enhancement module for single-shot object detector," in *Asian conference on computer vision*. Springer, 2018, pp. 325–340.

[64] S. Liu and D. Huang, "RFBNet: Receptive field block net for accurate and fast object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 385–400.

[65] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[66] J. Choi, D. Chun, H. Kim, and H.-J. Lee, "Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 502–511.

[67] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[68] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9627–9636.

[69] A. Casanova, G. Cucurull, M. Drozdzal, A. Romero, and Y. Bengio, "On the iterative refinement of densely connected representation levels for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 978–987.

[70] G. Chen and H. Qin, "Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving," *The Visual Computer*, 2022, v. 38, n. 3, pp. 1051–1063.

[71] A. Petrovskaya, M. Perrollaz, L. Oliveira, L. Spinello, R. Triebel, A. Makris, J. D. Yoder, C. Laugier, U. Nunes, and P. Bessiere, "Awareness of road scene participants for autonomous driving," *Handbook of Intelligent Vehicles*, 2012, v. 2-2, pp. 1384–1432.

[72] E. Amaral, C. Badue, T. Oliveira-Santos, and A. De Souza, "Detecção e rastreamento de veículos em movimento para automóveis robóticos autônomos," *XII simpósio brasileiro de automação inteligente (SBAI)*, 2015, pp. 801–806.

[73] L. Zhang, Q. Li, M. Li, Q. Mao, and A. Nüchter, "Multiple vehicle-like target tracking based on the velodyne lidar?" *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2013, v. 8, n. PART 1, pp. 126–131.

[74] A. Gurghian, T. Koduri, S. V. Bailur, K. J. Carey, and V. N. Murali, "DeepLanes: End-To-End Lane Position Estimation Using Deep Neural Networks," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 38–45.

[75] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Autonomous Robots*, 2009, v. 26, n. 2-3, pp. 123–139.

[76] M. He, E. Takeuchi, Y. Ninomiya, and S. Kato, "Precise and efficient model-based vehicle tracking method using Rao-Blackwellized and scaling series particle filters," in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016, pp. 117–124.

[77] L. Chen, L. Fan, G. Xie, K. Huang, and A. Nüchter, "Moving-object detection from consecutive stereo pairs using slanted plane smoothing," *IEEE Transactions on Intelligent Transportation Systems*, 2017, v. 18, n. 11, pp. 3093–3102.

[78] A. Azim and O. Aycard, "Layer-based supervised classification of moving objects in outdoor dynamic environment using 3D laser scanner," *IEEE Intelligent Vehicles Symposium, Proceedings*, 2014, n. Iv, pp. 1408–1414.

[79] Z. Ge, P. Wang, J. Wang, and Z. Chen, "A 2.5d grids based moving detection methodology for autonomous vehicle," in *2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)*, 2017, pp. 403–407.

[80] M. S. Darms, P. E. Rybski, C. Baker, and C. Urmson, "Obstacle detection and tracking for the urban challenge," *IEEE Transactions on intelligent transportation systems*, 2009, v. 10, n. 3, pp. 475–485.

[81] K. Na, J. Byun, M. Roh, and B. Seo, "Roadplot-datmo: Moving object tracking and track fusion system using multiple sensors," in *2015 international conference on connected vehicles and expo (ICCVE)*. IEEE, 2015, pp. 142–143.

[82] J.-r. Xue, D. Wang, S.-y. Du, D.-x. Cui, Y. Huang, and N.-n. Zheng, "A vision-centered multi-sensor fusing approach to self-localization and obstacle perception for robotic cars," *Frontiers of Information Technology & Electronic Engineering*, 2017, v. 18, n. 1, pp. 122–138.

[83] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, "An Empirical Evaluation of Deep Learning on Highway Driving," 2015, pp. 1–7. [Online]. Available: http://arxiv.org/abs/1504.01716

[84] F. Mutz, V. Cardoso, T. Teixeira, L. F. Jesus, M. A. Golcalves, R. Guidolini, J. Oliveira, C. Badue, and A. F. De Souza, "Following the leader using a tracking system based on pre-trained deep neural networks," *Proceedings of the International Joint Conference on Neural Networks*, 2017, v. 2017-May, pp. 4332–4339.

[85] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, v. 9905 LNCS, pp. 749–765.

[86] H. Wu, Y. Hu, K. Wang, H. Li, L. Nie, and H. Cheng, "Instance-aware representation learning and association for online multi-person tracking," *Pattern*

*Recognition*, 2019, v. 94, pp. 25–34.

[87] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, "Online multi-target tracking with strong and weak detections," in *European Conference on Computer Vision*.  Springer, 2016, pp. 84–99.

[88] M. Yang, Y. Wu, and Y. Jia, "A hybrid data association framework for robust online multi-object tracking," *IEEE Transactions on Image Processing*, 2017, v. 26, n. 12, pp. 5667–5679.

[89] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 300–311.

[90] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4705–4713.

[91] J. Xu, Y. Cao, Z. Zhang, and H. Hu, "Spatial-temporal relation networks for multi-object tracking," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3988–3998.

[92] P. Chu, H. Fan, C. C. Tan, and H. Ling, "Online multi-object tracking with instance-aware tracker and dynamic model refreshment," in *2019 IEEE winter conference on applications of computer vision (WACV)*.  IEEE, 2019, pp. 161–170.

[93] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 941–951.

[94] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE*

*transactions on pattern analysis and machine intelligence*, 2017, v. 40, n. 3, pp. 595–610.

[95] L. Wen, P. Zhu, D. Du, X. Bian, H. Ling, Q. Hu, J. Zheng, T. Peng, X. Wang, Y. Zhang *et al.*, "Visdrone-mot2019: The vision meets drone multiple object tracking challenge results," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[96] F. Guan, Z. Peng, K. Wang, X. Song, and J. Gao, "Multi-step hybrid prediction model of baltic supermax index based on support vector machine," *Neural Network World*, 2016, v. 26, n. 3, p. 219.

[97] B. Pang, Y. Li, Y. Zhang, M. Li, and C. Lu, "Tubetk: Adopting tubes to track multi-object in a one-step training model," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6308–6318.

[98] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, 2021, v. 129, n. 11, pp. 3069–3087.

[99] H. Song, W. Ding, Y. Chen, S. Shen, M. Y. Wang, and Q. Chen, "PiP: Planning-Informed Trajectory Prediction for Autonomous Driving," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, v. 12366 LNCS, pp. 598–614.

[100] J. Pan, H. Sun, K. Xu, Y. Jiang, X. Xiao, J. Hu, and J. Miao, "Lane-attention: Predicting vehicles' moving trajectories by learning their attention over lanes," *IEEE International Conference on Intelligent Robots and Systems*, 2020, pp. 7949–7956.

[101] J. Mercat, T. Gilles, N. El Zoghby, G. Sandou, D. Beauvois, and G. P. Gil, "Multi-Head Attention for Multi-Modal Joint Vehicle Motion Forecasting,"

*Proceedings - IEEE International Conference on Robotics and Automation,* 2020, pp. 9638–9644.

[102] S. Khandelwal, W. Qi, J. Singh, A. Hartnett, and D. Ramanan, "What-If Motion Prediction for Autonomous Driving," 2020. [Online]. Available: http://arxiv.org/abs/2008.10587

[103] C. Luo, L. Sun, D. Dabiri, and A. Yuille, "Probabilistic multi-modal trajectory prediction with lane attention for autonomous vehicles," *IEEE International Conference on Intelligent Robots and Systems,* 2020, pp. 2370–2376.

[104] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "DE-SIRE: Distant future prediction in dynamic scenes with interacting agents," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017,* 2017, v. 2017-Janua, pp. 2165–2174.

[105] S. H. Park, G. Lee, J. Seo, M. Bhat, M. Kang, J. Francis, A. Jadhav, P. P. Liang, and L. P. Morency, "Diverse and Admissible Trajectory Forecasting Through Multimodal Context Understanding," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* 2020, v. 12356 LNCS, pp. 282–298.

[106] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference,* 2014, pp. 1724–1734.

[107] B. Kim, S. H. Park, S. Lee, E. Khoshimjonov, D. Kum, J. Kim, J. S. Kim, and J. W. Choi, "Lapred: Lane-aware prediction of multi-modal future trajectories of dynamic agents," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2021, pp. 14 636–14 645.

[108] S. Narayanan, R. Moslemi, F. Pittaluga, B. Liu, and M. Chandraker, "Divide-and-conquer for lane-aware diverse trajectory prediction," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 794–15 803.

[109] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "HOME: Heatmap Output for future Motion Estimation," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2021, v. 2021-Septe, pp. 500–507.

[110] Y. Liu, J. Zhang, L. Fang, Q. Jiang, and B. Zhou, "Multimodal Motion Prediction with Stacked Transformers," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7573–7582.

[111] D. Lee, "Autonomous Vehicle Implementation Predictions: Implications for Transport Planning," *Transportation Research Board Annual Meeting*, 2020, v. 42, n. 5 June 2020, pp. 1–39.

[112] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "MultiPath: Multiple Probabilistic Anchor Trajectory Hypotheses for Behavior Prediction," 2019, n. CoRL. [Online]. Available: http://arxiv.org/abs/1910.05449

[113] H. Cui, V. Radosavljevic, F. C. Chou, T. H. Lin, T. Nguyen, T. K. Huang, J. Schneider, and N. Djuric, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," *Proceedings - IEEE International Conference on Robotics and Automation*, 2019, v. 2019-May, pp. 2090–2096.

[114] N. N. Sriram, B. Liu, F. Pittaluga, and M. Chandraker, "SMART: Simultaneous Multi-Agent Recurrent Trajectory Prediction," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, v. 12372 LNCS, n. Figure 1, pp. 463–479.

[115] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "CoverNet: Multimodal Behavior Prediction Using Trajectory Sets," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 062–14 071.

[116] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, 2021, v. 9, pp. 16 591–16 603.

[117] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[118] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, and C. Schmid, "VectorNet: Encoding HD maps and agent dynamics from vectorized representation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 522–11 530.

[119] M. Ye, T. Cao, and Q. Chen, "TPCN: Temporal Point Cloud Networks for Motion Forecasting," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 313–11 322.

[120] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, "Learning Lane Graph Representations for Motion Forecasting," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, v. 12347 LNCS, pp. 541–556.

[121] W. Zeng, M. Liang, R. Liao, and R. Urtasun, "LaneRCNN: Distributed Representations for Graph-Centric Motion Forecasting," *IEEE International Conference on Intelligent Robots and Systems*, 2021, pp. 532–539.

[122] J. Ngiam, B. Caine, V. Vasudevan, Z. Zhang, H.-T. L. Chiang, J. Ling, R. Roelofs, A. Bewley, C. Liu, A. Venugopal, D. Weiss, B. Sapp, Z. Chen, and J. Shlens, "Scene Transformer: A unified architecture for predicting

multiple agent trajectories," 2021, n. Figure 1, pp. 1–25. [Online]. Available: http://arxiv.org/abs/2106.08417

[123] G. Song, Y. Liu, and X. Wang, "Revisiting the Sibling Head in Object Detector," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 560–11 569.

[124] L. Fang, Q. Jiang, J. Shi, and B. Zhou, "TPNet: Trajectory Proposal Network for Motion Prediction," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6796–6805.

[125] H. Song, D. Luan, W. Ding, M. Y. Wang, and Q. Chen, "Learning to Predict Vehicle Trajectories with Model-based Planning," 2021, n. CoRL. [Online]. Available: http://arxiv.org/abs/2103.04027

[126] L. Paparusso, S. Melzi, and F. Braghin, "A deep-learning framework to predict the dynamics of a human-driven vehicle based on the road geometry," *arXiv preprint arXiv:2103.03825*, 2021.

[127] S. H. Park, G. Lee, J. Seo, M. Bhat, M. Kang, J. Francis, A. Jadhav, P. P. Liang, and L.-P. Morency, "Diverse and admissible trajectory forecasting through multimodal context understanding," in *European Conference on Computer Vision*. Springer, 2020, pp. 282–298.

[128] H. Cheng, W. Liao, X. Tang, M. Y. Yang, M. Sester, and B. Rosenhahn, "Exploring dynamic context for multi-path trajectory prediction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 795–12 801.

[129] O. Styles, T. Guha, V. Sanchez, and A. Kot, "Multi-camera trajectory forecasting: Pedestrian trajectory prediction in a network of cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 1016–1017.

[130] Y. Liu, Q. Yan, and A. Alahi, "Social nce: Contrastive learning of socially-aware motion representations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 118–15 129.

[131] B. Ivanovic, A. Elhafsi, G. Rosman, A. Gaidon, and M. Pavone, "Mats: An interpretable trajectory forecasting representation for planning and control," *arXiv preprint arXiv:2009.07517*, 2020.

[132] S. H. Park, G. Lee, J. Seo, M. Bhat, M. Kang, J. Francis, A. Jadhav, P. P. Liang, and L.-P. Morency, "Diverse and admissible trajectory forecasting through multimodal context understanding," in *European Conference on Computer Vision*. Springer, 2020, pp. 282–298.

[133] J. Liang, L. Jiang, and A. Hauptmann, "Simaug: Learning robust representations from simulation for trajectory prediction," in *European Conference on Computer Vision*. Springer, 2020, pp. 275–292.

[134] X. Weng, Y. Yuan, and K. Kitani, "Joint 3d tracking and forecasting with graph neural network and diversity sampling," *arXiv preprint arXiv:2003.07847*, 2020, v. 2, n. 6.2, p. 1.

[135] C. Graber and A. Schwing, "Dynamic neural relational inference for forecasting trajectories," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 1018–1019.

[136] A. Monti, A. Bertugli, S. Calderara, and R. Cucchiara, "Dag-net: Double attentive graph neural network for trajectory forecasting," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 2551–2558.

[137] P. Kothari, S. Kreiss, and A. Alahi, "Human trajectory forecasting in crowds: A deep learning perspective," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[138] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, "It is not the journey but the destination: Endpoint conditioned trajectory prediction," in *European conference on computer vision.* Springer, 2020, pp. 759–776.

[139] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer networks for trajectory forecasting," in *2020 25th international conference on pattern recognition (ICPR).* IEEE, 2021, pp. 10 335–10 342.

[140] E. W. Dijkstra *et al.*, "A note on two problems in connexion with graphs," *Numerische mathematik*, 1959, v. 1, n. 1, pp. 269–271.

[141] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE transactions on Systems Science and Cybernetics*, 1968, v. 4, n. 2, pp. 100–107.

[142] S. M. LaValle and J. J. Kuffner Jr, "Randomized kinodynamic planning," *The international journal of robotics research*, 2001, v. 20, n. 5, pp. 378–400.

[143] D. González, J. Pérez, V. Milanés, and F. Nashashibi, "A review of motion planning techniques for automated vehicles," *IEEE Transactions on intelligent transportation systems*, 2015, v. 17, n. 4, pp. 1135–1145.

[144] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE transactions on Robotics and Automation*, 1996, v. 12, n. 4, pp. 566–580.

[145] J. Reeds and L. Shepp, "Optimal paths for a car that goes both forwards and backwards," *Pacific journal of mathematics*, 1990, v. 145, n. 2, pp. 367–393.

[146] J. Funke, P. Theodosis, R. Hindiyeh, G. Stanek, K. Kritatakirana, C. Gerdes, D. Langer, M. Hernandez, B. Müller-Bessler, and B. Huhnke, "Up to the limits: Autonomous audi tts," in *2012 IEEE Intelligent Vehicles Symposium.* IEEE, 2012, pp. 541–547.

[147] W. Xu, J. Wei, J. M. Dolan, H. Zhao, and H. Zha, "A real-time motion planner with trajectory optimization for autonomous vehicles," in *2012 IEEE International Conference on Robotics and Automation.* IEEE, 2012, pp. 2061–2067.

[148] D. González, J. Pérez, R. Lattarulo, V. Milanés, and F. Nashashibi, "Continuous curvature planning with obstacle avoidance capabilities in urban scenarios," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC).* IEEE, 2014, pp. 1430–1435.

[149] R. T. Farouki and T. Sakkalis, "Pythagorean-hodograph space curves," *Advances in Computational Mathematics*, 1994, v. 2, n. 1, pp. 41–66.

[150] R. R. Murphy, *Introduction to AI robotics.* MIT press, 2019.

[151] L. Wang, *Support vector machines: theory and applications.* Springer Science & Business Media, 2005, v. 177.

[152] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.

[153] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, 1997, v. 9, n. 8, pp. 1735–1780.

[154] M. Inoue, T. Yamashita, and T. Nishida, "Robot path planning by lstm network under changing environment," in *Advances in computer communication and computational sciences.* Springer, 2019, pp. 317–329.

[155] J. Hammersley, *Monte Carlo methods.* Springer Science & Business Media, 2013.

[156] R. Coulom, "Efficient selectivity and backup operators in Monte-Carlo tree search," *International conference on computers and games*, 2006, pp. 72–83.

[157] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998, v. 86, n. 11, pp. 2278–2324.

[158] W. Gao, D. Hsu, W. S. Lee, S. Shen, and K. Subramanian, "Intention-net: Integrating planning and deep learning for goal-directed autonomous navigation," 2017, pp. 185–194.

[159] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[160] B. Wymann, E. Espié, C. Guionneau, C. Dimitrakakis, R. Coulom, and A. Sumner, "Torcs, the open racing car simulator," *Software available at http://torcs. sourceforge. net*, 2000, v. 4, n. 6, p. 2.

[161] J. Koutník, G. Cuccu, J. Schmidhuber, and F. Gomez, "Evolving large-scale neural networks for vision-based reinforcement learning," in *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, 2013, pp. 1061–1068.

[162] D. Loiacono, P. L. Lanzi, J. Togelius, E. Onieva, D. A. Pelta, M. V. Butz, T. D. Lönneker, L. Cardamone, D. Perez, Y. Sáez *et al.*, "The 2009 simulated car racing championship," *IEEE Transactions on Computational Intelligence and AI in Games*, 2010, v. 2, n. 2, pp. 131–147.

[163] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[164] S. Wang, D. Jia, and X. Weng, "Deep reinforcement learning for autonomous driving," *arXiv preprint arXiv:1811.11329*, 2018.

[165] C. Ye, H. Ma, X. Zhang, K. Zhang, and S. You, "Survival-oriented reinforcement learning model: An effcient and robust deep reinforcement learning algorithm for autonomous driving problem," in *International Conference on Image and Graphics.* Springer, 2017, pp. 417–429.

[166] J. Zhang and K. Cho, "Query-efficient imitation learning for end-to-end autonomous driving," *arXiv preprint arXiv:1605.06450*, 2016.

[167] X. Pan, Y. You, Z. Wang, and C. Lu, "Virtual to real reinforcement learning for autonomous driving," *arXiv preprint arXiv:1704.03952*, 2017.

[168] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[169] C. Gómez-Huélamo, J. D. Egido, L. M. Bergasa, R. Barea, E. López-Guillén, F. Arango, J. Araluce, and J. López, "Train here, drive there: Simulating real-world use cases with fully-autonomous driving architecture in carla simulator," in *Workshop of Physical Agents.* Springer, 2020, pp. 44–59.

[170] S. Hossain and D.-J. Lee, "Autonomous-driving vehicle learning environments using unity real-time engine and end-to-end cnn approach," *The Journal of Korea Robotics Society*, 2019, v. 14, n. 2, pp. 122–130.

[171] F. Codevilla, M. Müller, A. López, V. Koltun, and A. Dosovitskiy, "End-to-end driving via conditional imitation learning," in *2018 IEEE international conference on robotics and automation (ICRA).* IEEE, 2018, pp. 4693–4700.

[172] X. Liang, T. Wang, L. Yang, and E. Xing, "Cirl: Controllable imitative reinforcement learning for vision-based self-driving," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 584–599.

[173] Y. Jaafra, J. L. Laurent, A. Deruyver, and M. S. Naceur, "Robust reinforcement learning for autonomous driving," 2019. [Online]. Available:

https://openreview.net/forum?id=Bke03G85DN

[174] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. Theodorou, and B. Boots, "Agile autonomous driving using end-to-end deep imitation learning," *arXiv preprint arXiv:1709.07174*, 2017.

[175] J. Li, Y. Chen, X. Zhao, and J. Huang, "An improved dqn path planning algorithm," *The Journal of Supercomputing*, 2022, v. 78, n. 1, pp. 616–639.

[176] M. Bansal, A. Krizhevsky, and A. Ogale, "Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst," *arXiv preprint arXiv:1812.03079*, 2018.

[177] C. Zhou, B. Huang, and P. Fränti, "An advantage actor-critic algorithm for robotic motion planning in dense and dynamic scenarios," *arXiv preprint arXiv:2102.03138*, 2021.

[178] Y. Li, D. Li, W. Zhu, J. Sun, X. Zhang, and S. Li, "Constrained motion planning of 7DOF space manipulator via deep reinforcement learning combined with artificial potential field, Aerospace, v-9, p-163, 2022."

[179] H. Wang, P. Cai, Y. Sun, L. Wang, and M. Liu, "Learning interpretable end-to-end vision-based motion planning for autonomous driving with optical flow distillation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 731–13 737.

[180] M. Jaritz, R. De Charette, M. Toromanoff, E. Perot, and F. Nashashibi, "End-to-end race driving with deep reinforcement learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2070–2075.

[181] J. Zhang and K. Cho, "Query-efficient imitation learning for end-to-end autonomous driving," *arXiv preprint arXiv:1605.06450*, 2016.

[182] R. Hoque, A. Balakrishna, C. Putterman, M. Luo, D. S. Brown, D. Seita, B. Thananjeyan, E. Novoseller, and K. Goldberg, "Lazydagger: Reducing context switching in interactive imitation learning," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 502–509.

[183] X. Chen, I. Vizzo, T. Läbe, J. Behley, and C. Stachniss, "Range image-based lidar localization for autonomous vehicles," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5802–5808.

[184] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou *et al.*, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9710–9719.

[185] 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/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.

[186] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 954–960.

[187] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," *arXiv preprint arXiv:2006.14480*, 2020.

[188] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[189] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[190] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, v. 9905 LNCS, pp. 21–37.

[191] J. Lu, J. Yang, D. Batra, and D. Parikh, "Hierarchical question-image co-attention for visual question answering," *Advances in Neural Information Processing Systems*, 2016, n. c, pp. 289–297.

[192] Y. Zhu, O. Groth, M. Bernstein, and L. Fei-Fei, "Visual7W: Grounded question answering in images," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, v. 2016-Decem, pp. 4995–5004.

[193] P. H. Seo, A. Lehrmann, B. Han, and L. Sigal, "Visual reference resolution using attention memory for visual dialog," *Advances in Neural Information Processing Systems*, 2017, v. 2017-Decem, n. Nips, pp. 3720–3730.

[194] P. H. Seo, Z. Lin, S. Cohen, X. Shen, and B. Han, "Progressive attention networks for visual attribute prediction," *British Machine Vision Conference 2018, BMVC 2018*, 2019, pp. 1–19.

[195] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, v. 11211 LNCS, pp. 3–19.

[196] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4203–4212.

[197] Y. C. Yoon, A. Boragule, Y. M. Song, K. Yoon, and M. Jeon, "Online multi-object tracking with historical appearance matching and scene adaptive detection filtering," *arXiv*, 2018.

[198] A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, "Online multi-target tracking using recurrent neural networks," *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017, pp. 4225–4232.

[199] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics*, 2005, v. 52, n. 1, pp. 7–21.

[200] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 611–625.

[201] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.

[202] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European conference on computer vision*. Springer, 2012, pp. 611–625.

[203] H. Wu, Y. Hu, K. Wang, H. Li, L. Nie, and H. Cheng, "Instance-aware representation learning and association for online multi-person tracking," *Pattern Recognition*, oct 2019, v. 94, pp. 25–34. [Online]. Available: https://doi.org/10.1016%2Fj.patcog.2019.04.018

[204] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, "Online multi-target tracking with strong and weak detections," in *European Conference on Computer Vision*. Springer, 2016, pp. 84–99.

[205] M. Yang, Y. Wu, and Y. Jia, "A hybrid data association framework for robust online multi-object tracking," *IEEE Transactions on Image Processing*, 2017, v. 26, n. 12, pp. 5667–5679.

[206] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 300–311.

[207] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4705–4713.

[208] J. Xu, Y. Cao, Z. Zhang, and H. Hu, "Spatial-temporal relation networks for multi-object tracking," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3988–3998.

[209] P. Chu, H. Fan, C. C. Tan, and H. Ling, "Online multi-object tracking with instance-aware tracker and dynamic model refreshment," in *2019 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2019, pp. 161–170.

[210] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 941–951.

[211] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE transactions on pattern analysis and machine intelligence*, 2017, v. 40, n. 3, pp. 595–610.

[212] L. Chen, H. Ai, C. Shang, Z. Zhuang, and B. Bai, "Online multi-object tracking with convolutional neural networks," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 645–649.

[213] K. Fang, Y. Xiang, X. Li, and S. Savarese, "Recurrent autoregressive networks for online multi-object tracking," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 466–475.

[214] B. Pang, Y. Li, Y. Zhang, M. Li, and C. Lu, "Tubetk: Adopting tubes to track multi-object in a one-step training model," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6308–6318.

[215] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, 2021, v. 129, n. 11, pp. 3069–3087.

[216] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016, pp. 3464–3468.

[217] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[218] X. Wan, J. Wang, Z. Kong, Q. Zhao, and S. Deng, "Multi-object tracking using online metric learning with long short-term memory," in *2018 25th IEEE international conference on image processing (ICIP)*. IEEE, 2018, pp. 788–792.

[219] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, "Towards real-time multi-object tracking," in *European Conference on Computer Vision*. Springer, 2020, pp. 107–122.

[220] W. Feng, Z. Hu, W. Wu, J. Yan, and W. Ouyang, "Multi-object tracking with multiple cues and switcher-aware classification," *arXiv preprint arXiv:1901.06129*, 2019.

[221] Z. Zhou, J. Xing, M. Zhang, and W. Hu, "Online multi-target tracking with tensor-based high-order graph matching," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1809–1814.

[222] F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, and J. Yan, "POI: Multiple object tracking with high performance detection and appearance feature," in *European Conference on Computer Vision*. Springer, 2016, pp. 36–42.

[223] J. Peng, C. Wang, F. Wan, Y. Wu, Y. Wang, Y. Tai, C. Wang, J. Li, F. Huang, and Y. Fu, "Chained-Tracker: Chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking," in *European conference on computer vision*. Springer, 2020, pp. 145–161.

[224] S. Sun, N. Akhtar, H. Song, A. Mian, and M. Shah, "Deep affinity network for multiple object tracking," *IEEE transactions on pattern analysis and machine intelligence*, 2019, v. 43, n. 1, pp. 104–119.

[225] X. Zhou, V. Koltun, and P. Krähenbühl, "Tracking objects as points," in *European Conference on Computer Vision*. Springer, 2020, pp. 474–490.

[226] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, and D. Manocha, "Forecasting Trajectory and Behavior of Road-Agents Using Spectral Clustering in Graph-LSTMs," *IEEE Robotics and Automation Letters*, 2020, v. 5, n. 3, pp. 4882–4890.

[227] M. Wu, H. Ling, N. Bi, S. Gao, Q. Hu, H. Sheng, and J. Yu, "Visual Tracking with Multiview Trajectory Prediction," *IEEE Transactions on Image Processing*, 2020, v. 29, pp. 8355–8367.

[228] M. Huynh and G. Alaghband, "Trajectory Prediction by Coupling Scene-LSTM with Human Movement LSTM," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, v. 11844 LNCS, pp. 244–259.

[229] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[230] D. Lee, Y. Gu, J. Hoang, and M. Marchetti-Bowick, "Joint interaction and trajectory prediction for autonomous driving using graph neural networks," *arXiv*, 2019, n. NeurIPS.

[231] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," *arXiv*, 2019, pp. 1–10.

[232] P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang, and P. Luo, "TransTrack: Multiple Object Tracking with Transformer," 2020. [Online]. Available: http://arxiv.org/abs/2012.15460

[233] J. Sun, Q. Jiang, and C. Lu, "Recursive Social Behavior Graph for Trajectory Prediction," *arXiv*, 2020, pp. 660–669.

[234] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," *Proceedings of the IEEE International Conference on Computer Vision*, 2019, v. 2019-Octob, pp. 2375–2384.

[235] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016.

[236] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep Continuous Fusion for Multi-sensor 3D Object Detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, v. 11220 LNCS, pp. 663–678.

[237] A. Prakash, K. Chitta, and A. Geiger, "Multi-modal fusion transformer for end-to-end autonomous driving," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 7073–7083.

[238] Y. Chen, C. Dong, P. Palanisamy, P. Mudalige, K. Muelling, and J. M. Dolan, "Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2019, v. 2019-June, pp. 1326–1334.

[239] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, 2011, v. 30, n. 2, pp. 129–150.

[240] J. C. Mason and D. C. Handscomb, *Chebyshev polynomials*. Chapman and Hall/CRC, 2002.

[241] X. Li, X. Ying, and M. C. Chuah, "Grip: Graph-based interaction-aware trajectory prediction," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3960–3966.

[242] Y. Li, Y. Chen, N. Wang, and Z. X. Zhang, "Scale-aware trident networks for object detection," *Proceedings of the IEEE International Conference on Computer Vision*, 2019, v. 2019-Octob, pp. 6053–6062.

[243] K. Åhlander, "Einstein summation for multidimensional arrays," *Computers and Mathematics with Applications*, 2002, v. 44, n. 8-9, pp. 1007–1017.

[244] X. Li, X. Ying, and M. C. Chuah, "GRIP++: Enhanced Graph-based Interaction-aware Trajectory Prediction for Autonomous Driving," 2019. [Online]. Available: http://arxiv.org/abs/1907.07792

[245] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "WhereNext: A location predictor on trajectory pattern mining," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 637–645.

[246] S. Carrasco, D. F. Llorca, and M. Á. Sotelo, "SCOUT: Socially-COnsistent and UndersTandable Graph Attention Network for Trajectory Prediction of Vehicles and VRUs," 2021. [Online]. Available: http://arxiv.org/abs/2102.06361

[247] H. He, H. Dai, and N. Wang, "UST: Unifying spatio-temporal context for trajectory prediction in autonomous driving," *IEEE International Conference on Intelligent Robots and Systems*, 2020, pp. 5962–5969.

[248] J. Wang, P. Wang, C. Zhang, K. Su, and J. Li, "F-net: Fusion neural network for vehicle trajectory prediction in autonomous driving," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4095–4099.

[249] G. Kim, D. Kim, Y. Ahn, and K. Huh, "Hybrid Approach for Vehicle Trajectory Prediction Using Weighted Integration of Multiple Models," *IEEE Access*, 2021, pp. 1–1.

[250] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, v. 2019-June, pp. 8475–8484.

[251] S. Julka, V. Sowrirajan, J. Schloetterer, and M. Granitzer, "Conditional Generative Adversarial Networks for Speed Control in Trajectory Simulation," 2021. [Online]. Available: http://arxiv.org/abs/2103.11471

[252] M. Toromanoff, E. Wirbel, and F. Moutarde, "End-to-end model-free reinforcement learning for urban driving using implicit affordances," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7153–7162.

[253] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, "Learning by cheating," in *Conference on Robot Learning*. PMLR, 2020, pp. 66–75.

[254] F. Codevilla, E. Santana, A. M. López, and A. Gaidon, "Exploring the limitations of behavior cloning for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9329–9338.

# List of Publications

**List of papers published/accepted in journals:**

1. Singh, D., Srivastava, R. "Channel spatial attention based single-shot object detector for au- tonomous vehicles," Multimedia Tools and Applications, 2022, v. 81, n. 16, pp. 22 289–22 305.

2. D. Singh and R. Srivastava, "Graph neural network with rnns based trajectory prediction of dynamic agents for autonomous vehicle," Applied Intelligence, 2022, pp. 1–16. 175

3. Singh, D., Srivastava, R. "Multi-scale graph-transformer network for trajectory prediction of the autonomous vehicles," Intelligent Service Robotics, 2022, pp. 1–14.

4. Singh, D., Srivastava, R. "An end to end trained hybrid cnn model for multi-object tracking," Multimedia Tools and Applications, 2022, pp. 1–13.

**List of papers communicated in journals:**

1. D. Singh and R. Srivastava, Semantic Supervision guided image-based Motion Planning of the Autonomous Vehicles in Neural Processing Letters. (SCI-2.908) (Under Review)

**List of papers published in conferences:**

1. D. Singh and R. Srivastava,"Graph convolutional network-based centre-object tracking for the multi-dynamic agents for autonomous vehicles," in 2021 5th International Conference on Information Systems and Computer Networks (IS-CON). IEEE, 2021, pp. 1–8..

2. D. Singh and R. Srivastava, Instance Segmentation Based on Densely Supervised Multi-Scale Object Detector for Autonomous Vehicles in International Conference on Artificial Intelligence: Theory  Applications (Presented)