

REFERENCES

- [1] M. Y. Arafat, M. M. Alam, and S. Moh, “Vision-Based Navigation Techniques for Unmanned Aerial Vehicles: Review and Challenges,” *Drones*, vol. 7, no. 2, 2023.
- [2] A. Elamin, A. El-rabbany, and S. Jacob, “Event-Based Visual / Inertial Odometry for UAV Indoor Navigation,” 2025.
- [3] V. Niculescu, L. Lamberti, F. Conti, L. Benini, and D. Palossi, “Improving Autonomous Nano-Drones Performance via Automated End-to-End Optimization and Deployment of DNNs,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, no. 4, pp. 548–562, 2021.
- [4] L. Lamberti, “ENABLING MULTI-TASKING AI-BASED PERCEPTION ON AUTONOMOUS NANO-UAVS,” Ph.D. dissertation, 2024. [Online]. Available: https://amsdottorato.unibo.it/id/eprint/11643/3/lamberti_lorenzo_tesi.pdf
- [5] M. F. Bonner and R. A. Epstein, “Coding of navigational affordances in the human visual system,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 18, pp. 4793–4798, 2017.
- [6] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, “Grad-CAM: Why did you say that?” pp. 1–4, 2017. [Online]. Available: <http://arxiv.org/abs/1611.07450>
- [7] A. Garcia, S. S. Mittal, E. Kiewra, and K. Ghose, “A Convolutional Neural Network Feature Detection Approach to Autonomous Quadrotor Indoor Navigation,” *IEEE International Conference on Intelligent Robots and Systems*, pp. 74–81, 2019.
- [8] Y. Lu, Z. Xue, G. S. Xia, and L. Zhang, “A survey on vision-based UAV navigation,” *Geo-Spatial Information Science*, vol. 21, no. 1, pp. 21–32, 2018. [Online]. Available: <http://doi.org/10.1080/10095020.2017.1420509>
- [9] M. Y. Arafat and S. Moh, “Bio-inspired approaches for energy-efficient localization and clustering in uav networks for monitoring wildfires in remote areas,” *IEEE Access*, vol. 9, no. civil, pp. 18 649–18 669, 2021.
- [10] C. Wu, B. Ju, Y. Wu, X. Lin, N. Xiong, G. Xu, H. Li, and X. Liang, “UAV autonomous target search based on deep reinforcement learning in complex disaster scene,” *IEEE Access*, vol. 7, pp. 117 227–117 245, 2019.

- [11] T. C. Knowles, A. G. Summerton, J. G. Whiting, and M. J. Pearson, “Ring Attractors as the Basis of a Biomimetic Navigation System,” *Biomimetics*, vol. 8, no. 5, pp. 1–16, 2023.
- [12] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, “Vision based MAV navigation in unknown and unstructured environments,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 21–28, 2010.
- [13] K. Mohta, M. Watterson, Y. Mulgaonkar, S. Liu, C. Qu, A. Makineni, K. Saulnier, K. Sun, A. Zhu, J. Delmerico, K. Karydis, N. Atanasov, G. Loianno, D. Scaramuzza, K. Daniilidis, C. J. Taylor, and V. Kumar, “Fast, autonomous flight in GPS-denied and cluttered environments,” *Journal of Field Robotics*, vol. 35, no. 1, pp. 101–120, 2018.
- [14] I. Sa, M. Kamel, M. Burri, M. Bloesch, R. Khanna, M. Popovic, J. Nieto, and R. Siegwart, “Build Your Own Visual-Inertial Drone,” *IEEE Robotics & Automation Magazine*, no. March 2018, pp. 89–103, 2017.
- [15] G. Loianno, C. Brunner, G. McGrath, and V. Kumar, “Estimation, Control, and Planning for Aggressive Flight with a Small Quadrotor with a Single Camera and IMU,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 404–411, 2017.
- [16] P. Foehn, E. Kaufmann, A. Romero, R. Penicka, S. Sun, L. Bauersfeld, T. Laengle, G. Cioffi, Y. Song, A. Loquercio, and D. Scaramuzza, “Agilicious: Open-source and open-hardware agile quadrotor for vision-based flight,” *Science Robotics*, vol. 7, no. 67, 2022.
- [17] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, “A 64mW DNN-based Visual Navigation Engine for Autonomous Nano-Drones,” 5 2018. [Online]. Available: <http://arxiv.org/abs/1805.01831><http://dx.doi.org/10.1109/JIOT.2019.2917066>
- [18] D. Palossi, N. Zimmerman, A. Burrello, F. Conti, H. Muller, L. M. Gambardella, L. Benini, A. Giusti, and J. Guzzi, “Fully Onboard AI-Powered Human-Drone Pose Estimation on Ultralow-Power Autonomous Flying Nano-UAVs,” *IEEE Internet of Things Journal*, vol. 9, no. 3, pp. 1913–1929, 2022.
- [19] A. Loquercio, A. I. Maqueda, C. R. Del-Blanco, and D. Scaramuzza, “DroNet: Learning to Fly by Driving,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088–1095, 2018.
- [20] L. Lamberti, L. Bellone, L. Macan, E. Natalizio, F. Conti, D. Palossi, and L. Benini, “Distilling Tiny and Ultrafast Deep Neural Networks for Autonomous

Navigation on Nano-UAVs,” *IEEE Internet of Things Journal*, vol. 11, no. 20, pp. 33 269–33 281, 2024.

- [21] D. Palossi, A. Loquercio, F. Conti, E. Flamand, D. Scaramuzza, and L. Benini, “A 64-mW DNN-Based Visual Navigation Engine for Autonomous Nano-Drones,” *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8357–8371, 2019.
- [22] S. N. Fry and R. Wehner, “Honey bees store landmarks in an egocentric frame of reference,” *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology*, vol. 187, no. 12, pp. 1009–1016, 2002.
- [23] M. Kutbi, H. Li, Y. Chang, B. Sun, X. Li, C. Cai, N. Agadakos, G. Hua, and P. Mordohai, “Egocentric Computer Vision for Hands-Free Robotic Wheelchair Navigation,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 107, no. 1, 2023.
- [24] B. Pan, A. W. Harley, C. K. Liu, and L. J. Guibas, “LookOut: Real-World Humanoid Egocentric Navigation,” 2025. [Online]. Available: <http://arxiv.org/abs/2508.14466>
- [25] A. Prakash, R. Tu, M. Chang, and S. Gupta, “3D Hand Pose Estimation in Everyday Egocentric Images,” pp. 183–202, 2025.
- [26] C. Plizzari, G. Goletto, A. Furnari, S. Bansal, F. Ragusa, G. M. Farinella, D. Damen, and T. Tommasi, “An Outlook into the Future of Egocentric Vision,” *International Journal of Computer Vision*, vol. 132, no. 11, pp. 4880–4936, 2024. [Online]. Available: <https://doi.org/10.1007/s11263-024-02095-7>
- [27] Z. Qiu, Z. Liu, W. Niu, T. Bhattacharjee, and S. Kalantari, “EgoCogNav: Cognition-aware Human Egocentric Navigation,” 2025. [Online]. Available: <http://arxiv.org/abs/2511.17581>
- [28] W. Wang, C. K. Liu, and M. Kennedy, “Egocentric Scene-aware Human Trajectory Prediction,” 2024. [Online]. Available: <http://arxiv.org/abs/2403.19026>
- [29] J. Krantz, A. Gokaslan, D. Batra, S. Lee, and O. Maksymets, “Waypoint Models for Instruction-guided Navigation in Continuous Environments,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 15 142–15 151, 2021.
- [30] P. Anderson, Q. Wu, D. Teney, J. Bruce, M. Johnson, N. Sunderhauf, I. Reid, S. Gould, and A. Van Den Hengel, “Vision-and-Language Navigation: Interpreting Visually-Grounded Navigation Instructions in Real Environments,” *Proceed-*

ings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3674–3683, 2018.

- [31] S. Liu, H. Zhang, Y. Qi, P. Wang, Y. Zhang, and Q. Wu, “AerialVLN: Vision-and-Language Navigation for UAVs,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 15 338–15 348, 2023.
- [32] C. Chen, S. Majumder, Z. Al-Halah, R. Gao, S. K. Ramakrishnan, and K. Grauman, “Learning To Set Waypoints for Audio-Visual Navigation,” *ICLR 2021 - 9th International Conference on Learning Representations*, pp. 1–16, 2021.
- [33] B. R. Mantha and B. Garcia de Soto, “Designing a reliable fiducial marker network for autonomous indoor robot navigation,” *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, no. Isarc, pp. 74–81, 2019.
- [34] A. Giusti, J. Guzzi, D. C. Ciresan, F. L. He, J. P. Rodriguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. D. Caro, D. Scaramuzza, and L. M. Gambardella, “A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots,” *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 661–667, 2016.
- [35] A. Garcia, E. Mattison, and K. Ghose, “High-speed vision-based autonomous indoor navigation of a quadcopter,” *2015 International Conference on Unmanned Aircraft Systems, ICUAS 2015*, pp. 338–347, 2015.
- [36] A. Garcia, S. S. Mittal, E. Kiewra, and K. Ghose, “A convolutional neural network vision system approach to indoor autonomous quadrotor navigation,” *2019 International Conference on Unmanned Aircraft Systems, ICUAS 2019*, no. May, pp. 1344–1352, 2019.
- [37] R. P. Padhy, S. Verma, S. Ahmad, S. K. Choudhury, and P. K. Sa, “Deep Neural Network for Autonomous UAV Navigation in Indoor Corridor Environments,” *Procedia Computer Science*, vol. 133, pp. 643–650, 2018. [Online]. Available: <https://doi.org/10.1016/j.procs.2018.07.099>
- [38] S. S. Mansouri, P. Karvelis, C. Kanellakis, A. Koval, and G. Nikolakopoulos, “Visual Subterranean Junction Recognition for MAVs based on Convolutional Neural Networks,” *IECON Proceedings (Industrial Electronics Conference)*, vol. 2019-Octob, pp. 192–197, 2019.
- [39] E. Dack, L. Brigato, M. McMurray, M. Fontanellaz, T. Frauenfelder, H. Hoppe, A. Exadaktylos, T. Geiser, M. Funke-Chambour, A. Christel, L. Ebner, and

- S. Mougiakakou, “An Empirical Analysis for Zero-Shot Multi-Label Classification on COVID-19 CT Scans and Uncurated Reports,” *Proceedings - 2023 IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2023*, pp. 2626–2635, 2023.
- [40] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.
- [41] Z. C. Lipton, “The mythos of model interpretability,” *Communications of the ACM*, vol. 61, no. 10, pp. 35–43, 2018.
- [42] L. V. Haar, T. Elvira, and O. Ochoa, “An analysis of explainability methods for convolutional neural networks,” *Engineering Applications of Artificial Intelligence*, vol. 117, no. October 2022, p. 105606, 2023. [Online]. Available: <https://doi.org/10.1016/j.engappai.2022.105606>
- [43] D. Palossi, F. Conti, and L. Benini, “An open source and open hardware deep learning-powered visual navigation engine for autonomous nano-UAVs,” *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*, pp. 604–611, 2019.
- [44] E. Flamand, D. Rossi, F. Conti, I. Loi, A. Pullini, F. Rotenberg, and L. Benini, “GAP-8: A RISC-V SoC for AI at the Edge of the IoT,” *Proceedings of the International Conference on Application-Specific Systems, Architectures and Processors*, vol. 2018-July, pp. 1–4, 2018.
- [45] K. He, “Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification,” 2014.
- [46] E. Spera, A. Furnari, S. Battiato, and G. M. Farinella, “Egocart: A benchmark dataset for large-scale indoor image-based localization in retail stores,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1253–1267, 2021.
- [47] A. Nada and M. Chen, “Visual Question Answering,” *2024 International Conference on Computing, Networking and Communications, ICNC 2024*, pp. 6–10, 2024.
- [48] K. Schlegel, P. Neubert, and P. Protzel, “Building a Navigation System for a Shopping Assistant Robot from Off-the-Shelf Components,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12228 LNAI, pp. 103–115, 2020.

- [49] X. Yuan, T. Xu, X. Liu, Y. Wang, H. Qin, Y. Fang, and J. Li, “Multi-Step Temporal Modeling for UAV Tracking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 8, pp. 7216–7230, 2024.
- [50] K. Kelchtermans and T. Tuytelaars, “How hard is it to cross the room? – Training (Recurrent) Neural Networks to steer a UAV,” 2017. [Online]. Available: <http://arxiv.org/abs/1702.07600>
- [51] J. Lin, C. Gan, and S. Han, “TSM: Temporal shift module for efficient video understanding,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7082–7092, 2019.
- [52] 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)*, vol. 8689 LNCS, no. PART 1, pp. 818–833, 2014.
- [53] Z. Li, F. Li, W. Zhang, Z. Zheng, X. Liu, Y. Liu, and L. Zeng, “THUD++: Large-Scale Dynamic Indoor Scene Dataset and Benchmark for Mobile Robots,” pp. 1–11, 2024. [Online]. Available: <http://arxiv.org/abs/2412.08096>
- [54] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 2242–2251, 2017.