Learning Agile, Vision-based Drone Flight: from Simulation to Reality

Davide Scaramuzza[1] and Elia Kaufmann[1]

University of Zurich

Abstract. We present our latest research in learning deep sensorimotor policies for agile, vision-based quadrotor flight. We show methodologies for the successful transfer of such policies from simulation to the real world. In addition, we discuss the open research questions that still need to be answered to improve the agility and robustness of autonomous drones toward human-pilot performance.

1 Introduction

Quadrotors are among the most agile and dynamic machines ever created. However, developing fully autonomous quadrotors that can approach or even outperform the agility of birds or human drone pilots with only onboard sensing and computing is very challenging and still unsolved. Prior work on autonomous quadrotor navigation has approached the challenge of vision-based autonomous navigation by separating the system into a sequence of independent compute modules [1–4]. While such modularization of the system is beneficial in terms of interpretability and allows to easily exchange modules, it results in a substantial increase in latency from perception to action and results in errors being propagated from one module to the next. Furthermore, prior work on autonomous quadrotor navigation heavily relies on control techniques such as PID control [5–7] or model predictive control (MPC) [8, 9]. While these methods have demonstrated impressive feats in controlled environments [5, 7, 9, 10], they require substantial amount of tuning, and are difficult, if not impossible, to scale to complex dynamics models without large penalties in computation time.

In this extended abstract, we summarize our latest research in learning deep sensorimotor policies for agile vision-based quadrotor flight. Learning sensorimotor controllers represents a holistic approach that is more resilient to noisy sensory observations and imperfect world models. Training robust policies requires however a large amount of data. Nevertheless, we will show that simulation data, combined with randomization and abstraction of sensory observations, is enough to train policies that generalize to the real world. Such policies enable autonomous quadrotors to fly faster and more agile than what was possible before with only onboard sensing and computation. In addition, we discuss the open research questions that still need to be answered to improve the agility and robustness of autonomous drones toward human-pilot performance.

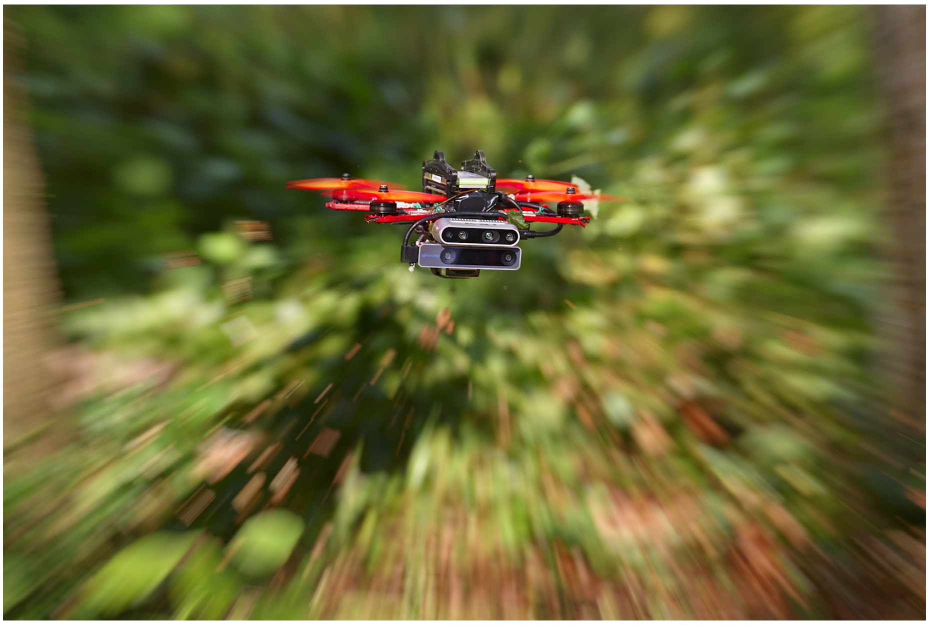

Fig. 1. Our drone flying at high speed (10 m/s) through a dense forest using only onboard sensing and computation.

2 High-Speed Flight in the Wild

In [11], we presented an end-to-end approach that can autonomously fly quadrotors through complex natural and human-made environments at high speeds (10 m/s), with purely onboard sensing and computation.[1] The key principle is to directly map noisy sensory observations to collision-free trajectories in a receding-horizon fashion. This direct mapping drastically reduces processing latency and increases robustness to noisy and incomplete perception. The sensorimotor mapping is performed by a convolutional network that is trained *exclusively* in simulation via privileged learning: imitating an expert with access to privileged information. We leverage abstraction of the input data to transfer the policy from simulation to reality [12, 13]. To this end, we utilize a stereo matching algorithm to provide depth images as input to the policy. We show that this representation is both rich enough to safely navigate through complex environments and abstract enough to bridge the gap between simulation and real world. By combining abstraction with simulating realistic sensor noise, our approach achieves zero-shot transfer from simulation to challenging real-world environments that were never experienced during training: dense forests, snow-covered terrain, derailed trains, and collapsed buildings. Our work demonstrates that

[1] A video of the results can be found here: https://youtu.be/m89bNn6RFoQ

end-to-end policies trained in simulation enable high-speed autonomous flight through challenging environments, outperforming traditional obstacle avoidance pipelines. A qualitative example of flight in the wild is shown in Figure 1.

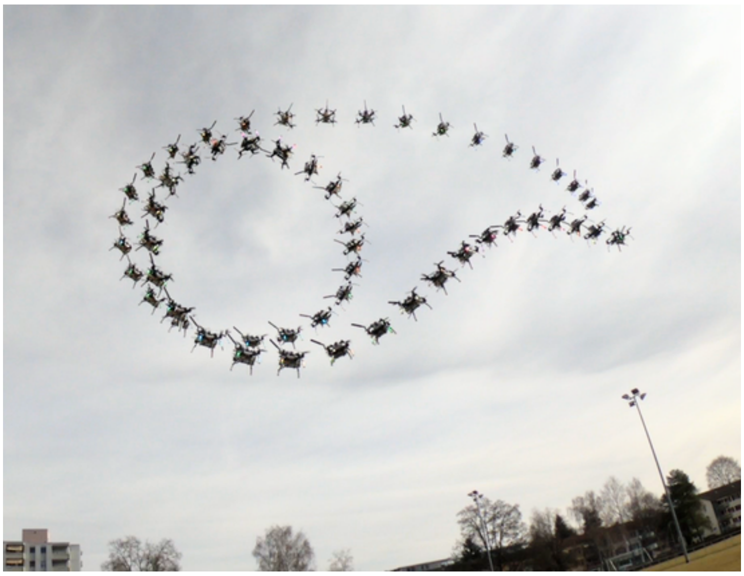

Fig. 2. Our quadrotor performs a Matty Flip. The drone is controlled by a deep sensorimotor policy and uses only onboard sensing and computation.

3 Acrobatic Flight

Acrobatic flight with quadrotors is extremely challenging. Human drone pilots require many years of practice to safely master maneuvers such as power loops and barrel rolls. For aerial vehicles that rely only on onboard sensing and computation, the high accelerations that are required for acrobatic maneuvers together with the unforgiving requirements on the control stack raise fundamental questions in both perception and control. For this reason, we challenged our drone with the task of performing acrobatic maneuvers [13].[2] In order to achieve this task, we trained a neural network to predict actions directly from visual and inertial sensor observations. Training is done by imitating an optimal controller with access to privileged information in the form of the exact robot state. Since such information is not available in the physical world, we trained the neural network to predict actions instead from inertial and visual observations.

[2] A video of the results can be found here: https://youtu.be/2N_wKXQ6MXA

Similarly to the work described in the previous section, all of the training is done in simulation, without the need of any data from the real world. We achieved this by using abstraction of sensor measurements, which reduces the simulation-to-reality gap compared to feeding raw observations. Both theoretically and experimentally, we have shown that abstraction strongly favours simulation-to-reality transfer. The learned policy allowed our drone to go faster than ever before and successfully fly maneuvers with accelerations of up to 3g, such as the Matty flip illustrated in Figure 2.

4 Autonomous Drone Racing

Drone racing is an emerging sport where pilots race against each other with remote-controlled quadrotors while being provided a first-person-view (FPV) video stream from a camera mounted on the drone. Drone pilots undergo years of training to master the skills involved in racing. In recent years, the task of autonomous drone racing has received substantial attention in the robotics community, which can mainly be attributed to two reasons: (i) The sensorimotor skills required for autonomous racing would also be valuable to autonomous systems in applications such as disaster response or structure inspection, where drones must be able to quickly and safely fly through complex dynamic environments. (ii) The task of autonomous racing provides an ideal test bed to objectively and quantifiably compare agile drone flight against human baselines. It also allows us to measure the research progress towards the ambitious goal of achieving super-human performance.

One approach to autonomous racing is to fly through the course by tracking a precomputed, potentially time-optimal, trajectory [10]. However, such an approach requires to know the race-track layout in advance, along with highly accurate state estimation, which current methods are still not able to provide. Indeed, visual inertial odometry is subject to drift in estimation over time. SLAM methods can reduce drift by relocalizing in a previously-generated, globally-consistent map. However, enforcing global consistency leads to increased computational demands that strain the limits of on-board processing.

Instead of relying on globally consistent state estimates, we deploy a convolutional neural network to identify the next location to fly to, also called waypoints. However, it is not clear a priori what should be the representation of the next waypoint. In our works, we have explored different solutions. In our preliminary work, the neural network predicts a fixed distance location from the drone [14]. Training was done by imitation learning on a globally optimal trajectory passing through all the gates. Despite being very efficient and easy to develop, this approach cannot efficiently generalize between different track layouts, given the fact that the training data depends on a track-dependent global trajectory representing the optimal path through all gates.

For this reason, a follow-up version of this work proposed to use as waypoint the location of the next gate [15]. As before, the prediction of the next gate is provided by a neural network. However, in contrast to the previous work, the neural network also predicts a measure of uncertainty.

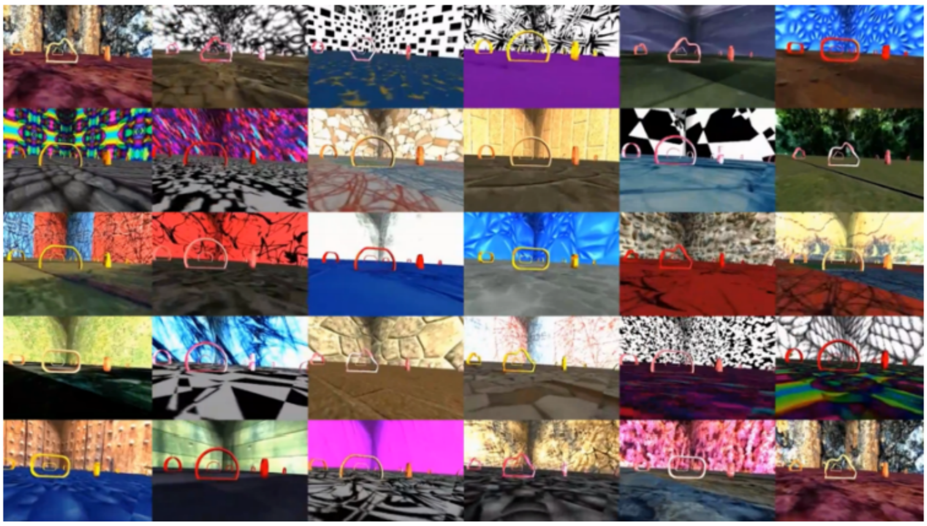

Fig. 3. The perception component of our system, represented by a convolutional neural network (CNN), is trained only with non-photorealistic simulation data.

Even though the waypoint representation proposed in [15] allowed for efficient transfer between track layouts, it still required substantial amount of real-world data to train. Collecting such data is generally a very tedious and time consuming process, which represents a limitation of the two previous works. In addition, when something changes in the environment, or the appearance of the gates changes substantially, the data collection process needs to be repeated from scratch. For this reason, in our recent work [16] we have proposed to collect data *exclusively* in simulation. To enable transfer between the real and the physical world, we randomized all the features which predictions should be robust against, *i.e.* illumination, gate shape, floor texture, and background. A sample of the training data generated by this process, called domain randomization, can be observed in Figure 3. Our approach was the first to demonstrate zero-shot sim-to-real transfer on the task of agile drone flight. A collection of the ideas presented in the above works has been used by our team to successfully compete in the Alpha-Pilot competition [17].[3]

[3] A video of the results can be found here: https://youtu.be/DGjwm5PZQT8

5 Future Work

Our work shows that neural networks have a strong potential to control agile platforms like quadrotors. In comparison to traditional methods, neural policies are more robust to noise in the observations and can deal with imperfection in sensing and actuation. However, the approaches presented in this extended abstract fall still short of the performance of professional drone pilots. To further push the capabilities of autonomous drones, a specialization to the task is required, potentially through real-world adaptation and online learning. Solving those challenges could potentially bring autonomous drones closer in agility to human pilots and birds. In this context, we present some limitations of the approaches proposed so far and interesting avenues for future work.

5.1 Generalization

One example of such a challenge is *generalization*: even though this extended abstract presents methods presents promising approaches to push the frontiers of autonomous, vision-based agile drone navigation in scenarios such as acrobatic flight, agile flight in cluttered environments, and drone racing, it is not clear how these sensorimotor policies can be efficiently transferred to novel task formulations, sensor configurations, or physical platforms. Following the methodologies presented in this paper, transferring to any of these new scenarios would require to retrain the policy in simulation, or perform adaptation using learned residual models. While the former suffers from the need to re-identify observation models and dynamics models, the latter is restricted to policy transfer between simulation and reality for the same task.

Possible avenues for future work to address this challenge include hierarchical learning [18] or approaches that optimize policy parameters on a learned manifold [19].

5.2 Continual Learning

The approaches to sensorimotor policy training presented in this paper are *static* in their nature: after the policy is trained in simulation via imitation learning or reinforcement learning, its parameters are frozen and applied on the real robot. In contrast, natural agents interact fundamentally different with their environment; they continually adapt to new experience and improve their performance in a much more dynamic fashion. Designing an artificial agent with similar capabilities would greatly increase the utility of robots in the real world and is an interesting direction for future work.

Recent work has proposed methods to enable artificial agents to perform few-shot learning of new tasks and scenarios using techniques such as adaptive learning [20,21] or meta learning [22,23]. These methods have shown promising results on simple manipulation and locomotion tasks but remain to be demonstrated on complex navigation tasks such as high-speed drone flight in the real world.

5.3 Autonomous Drone Racing

The approaches presented in this paper addressing the challenge of autonomous drone racing are limited to single-agent time trial races. To achieve true racing behaviour, these approaches need to be extended to a multi-agent setting, which raises novel challenges in perception, planning, and control. Regarding perception, multi-agent drone racing requires to detect opponent agents, which is a challenging task when navigating at high speeds and when onboard observations are subject to motion blur. The planning challenges arise from the need to design game-theoretic strategies [24] that involve maneuvers such as strategic blocking, which are potentially not time-optimal but still dominant approaches when competing in a multi-agent setting. Additionally, the limited field of view of the onboard camera renders opponents potentially unobservable, which requires a mental model of the trajectory of opponents. Finally, racing simultaneously with other drones through a race track poses new challenges in modeling and control, as aerodynamic effects induced by other agents need to be accounted for.

We envision that many of these challenges can be addressed using deep reinforcement learning in combination with *self-play*, where agents improve by competing against each other [25, 26].

References

1. T. Cieslewski, E. Kaufmann, and D. Scaramuzza, “Rapid exploration with multirotors: A frontier selection method for high speed flight,” in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2017, pp. 2135–2142.
2. M. Ryll, J. Ware, J. Carter, and N. Roy, “Efficient trajectory planning for high speed flight in unknown environments,” in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 732–738.
3. R. Mahony, V. Kumar, and P. Corke, “Multirotor aerial vehicles: Modeling, estimation, and control of quadrotor,” *IEEE Robot. Autom. Mag.*, 2012.
4. S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, “Vision-based state estimation and trajectory control towards high-speed flight with a quadrotor,” in *Robotics: Science and Systems (RSS)*, 2013.
5. D. Falanga, E. Mueggler, M. Faessler, and D. Scaramuzza, “Aggressive quadrotor flight through narrow gaps with onboard sensing and computing using active vision,” in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2017, pp. 5774–5781.
6. M. Faessler, A. Franchi, and D. Scaramuzza, “Differential flatness of quadrotor dynamics subject to rotor drag for accurate tracking of high-speed trajectories,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 620–626, Apr. 2018.
7. S. Sun, L. Sijbers, X. Wang, and C. de Visser, “High-speed flight of quadrotor despite loss of single rotor,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3201–3207, 2018.
8. D. Falanga, P. Foehn, P. Lu, and D. Scaramuzza, “PAMPC: Perception-aware model predictive control for quadrotors,” in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018.
9. F. Nan, S. Sun, P. Foehn, and D. Scaramuzza, “Nonlinear mpc for quadrotor fault-tolerant control,” in *IEEE Robot. Autom. Lett.*, 2022.

10. P. Foehn, A. Romero, and D. Scaramuzza, "Time-optimal planning for quadrotor waypoint flight," *Science Robotics*, vol. 6, no. 56, 2021. [Online]. Available: https://robotics.sciencemag.org/content/6/56/eabh1221
11. A. Loquercio, E. Kaufmann, R. Ranftl, M. Müller, V. Koltun, and D. Scaramuzza, "Learning high-speed flight in the wild," in *Science Robotics*, 2021.
12. M. Müller, A. Dosovitskiy, B. Ghanem, and V. Koltun, "Driving policy transfer via modularity and abstraction," in *Conference on Robot Learning*, 2018, pp. 1–15.
13. E. Kaufmann, A. Loquercio, R. Ranftl, M. Müller, V. Koltun, and D. Scaramuzza, "Deep drone acrobatics," in *Robotics: Science and Systems (RSS)*, 2020.
14. E. Kaufmann, A. Loquercio, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, "Deep drone racing: Learning agile flight in dynamic environments," in *Conf. on Robotics Learning (CoRL)*, 2018.
15. E. Kaufmann, M. Gehrig, P. Foehn, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, "Beauty and the beast: Optimal methods meet learning for drone racing," *IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 690–696, 2019. [Online]. Available: https://doi.org/10.1109/ICRA.2019.8793631
16. A. Loquercio, E. Kaufmann, R. Ranftl, A. Dosovitskiy, V. Koltun, and D. Scaramuzza, "Deep drone racing: From simulation to reality with domain randomization," *IEEE Trans. Robot.*, vol. 36, no. 1, pp. 1–14, 2019.
17. P. Foehn, D. Brescianini, E. Kaufmann, T. Cieslewski, M. Gehrig, M. Muglikar, and D. Scaramuzza, "Alphapilot: Autonomous drone racing," *Robotics: Science and Systems (RSS)*, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s11370-018-00271-6
18. T. Haarnoja, K. Hartikainen, P. Abbeel, and S. Levine, "Latent space policies for hierarchical reinforcement learning," in *International Conference on Machine Learning.* PMLR, 2018, pp. 1851–1860.
19. X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," in *Robotics: Science and Systems (RSS)*, 2020.
20. L. Smith, J. C. Kew, X. B. Peng, S. Ha, J. Tan, and S. Levine, "Legged robots that keep on learning: Fine-tuning locomotion policies in the real world," *arXiv preprint arXiv:2110.05457*, 2021.
21. A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv preprint arXiv:2107.04034*, 2021.
22. M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas, "Learning to learn by gradient descent by gradient descent," *Advances in neural information processing systems*, vol. 29, 2016.
23. C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning.* PMLR, 2017, pp. 1126–1135.
24. R. Spica, D. Falanga, E. Cristofalo, E. Montijano, D. Scaramuzza, and M. Schwager, "A real-time game theoretic planner for autonomous two-player drone racing," in *Robotics: Science and Systems (RSS)*, 2018.
25. B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch, "Emergent tool use from multi-agent autocurricula," *arXiv preprint arXiv:1909.07528*, 2019.
26. D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.