Performing Aerobatic Maneuver with Imitation Learning

Henrique Freitas[1[0000−0002−7201−5960]], Rui Camacho[1,2[0000−0003−0940−3554]], and Daniel Castro Silva[1,3[0000−0001−9293−0341]]

[1] Department of Informatics Engineering, Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, s/n., 4200-465, Porto, Portugal
up201707046@edu.fe.up.pt, {rcamacho, dcs}@fe.up.pt
[2] LIAAD/INESC TEC - Artificial Intelligence and Decision Support Laboratory / Institute for Systems and Computer Engineering, Technology and Science
[3] LIACC - Artificial Intelligence and Computer Science Laboratory

Abstract. The work reported in this article addresses the challenge of building models for non-trivial aerobatic aircraft maneuvers in an automated fashion. It is built using a Behavioural Cloning approach where human pilots provide a set of example maneuvers used by a Machine Learning algorithm to induce a control model for each maneuver. The best examples for each maneuver were selected using a set of objective evaluation metrics. Using those example sets, robust models were induced that could replicate (and in some cases outperform) the human pilots that provided the examples (the clean-up effect). Complete complex maneuvers were performed using a meta-controller capable of sequencing the basic ones learned by imitation. This endeavor was rewarded by the results that show several Machine Learning models capable of performing highly complex aircraft maneuvers.

Keywords: Imitation Learning · Behavioural Cloning · Aviation · Autopilot · Aerobatic maneuvers.

1 Introduction

With the increase of robot complexity, manually programming the behaviors and actions is becoming costly [1]. It requires excellent knowledge about the robot movement and many resources to develop [15].

In recent years, autonomous driving technologies are improving with Artificial Intelligence methods. One of the used techniques is Imitation Learning (IL) and Behaviour Cloning (BC) [8, 24, 25]; these have been used for a few decades, are based on Supervised Learning and learn the behaviors based on expert demonstrations through imitation [19, 20, 22], i.e., learning the mappings between the environment state (input) and actions (output).

The work reported in this paper concerns the application of such Behavioural Cloning methodology to an aviation domain. Currently used autopilots are not helpful in various complex events, being usually only used in specific parts of the flight [6].

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

The commercial aviation industry keeps growing, and according to historical data, the passenger count doubles roughly every 15 years [26]. In the 2004-2019 period remarkably, it increased by a factor of 2.4. As stated by the International Air Transport Association in their 20-Year Passenger Forecast, from 2019 to 2040, the global air passenger growth is predicted to be 3.3% annually, which results in a 90% increase in that interval [13].

Although the global airlines market size was over 800B USD in 2019 [14], the airlines anticipate problems in keeping up with the demand, such as pilot shortage [16, 27]. The Federal Aviation Administration (FAA) rule that limits airlines to hire pilots with a minimum of 1500 flight hours is one of the possible causes [17]. This rule aggravates the high training cost [17] because it requires pilot trainees to spend even more money to meet this requirement. The United States Air Force is also struggling with pilot training [5].

To solve these problems, an intelligent autopilot system can be built to be responsible for piloting through all phases of a flight or mission and/or help pilot apprentices improve their training with tips and cues on what they are doing wrong and how to improve. This system could also define an evaluation threshold to inform how close to the end one specific trainee is.

Focused on this issue, the first contribution of this work is the implementation of evaluation metrics for three aerobatic maneuvers: Immelmann, Split-S, and Half Cuban Eight. For instance, flight instructors can use these metrics to better understand the eventual difficulties their trainees are having.

These metrics are then used to develop Machine Learning controllers capable of performing the mentioned aerobatic maneuvers. Previous work uses Neural Networks (NN) based on Long Short-Term Memory (LSTM), mostly in autonomous driving [8,24,25] but also in the aerobatics context [18,21,23], where air vehicles are controlled solely by automated systems.

The final contribution is a high-level controller to perform an aerobatic performance show, similar to a Red Bull Air Race performance. This system sequentially invokes the respective controllers to execute the specific maneuvers.

This document is structured as follows: section 2 presents a study and review of the related work using this technology in other use cases. Section 3 briefly describes the data and evaluation metrics. Section 4 explains the controller training process and the results. Section 5 describes the high-level controller for the previously trained controllers. Finally, section 6 concludes this work with possible future work evolution.

2 Related Work

Regarding IL, there are many introductions, studies, and reviews on the subject, such as [2,12,22]. Despite being generic, it has been recently operated thoroughly in robots and similar contexts, such as autonomous driving [1, 11].

In particular to aerial vehicles, Müller et al. used a Convolutional Neural Network (CNN) to learn to control a racing drone [21]. The data was collected in a simulator, automatically generating stadium racing tracks and three levels

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

of expert pilots. Joystick controls and images from the Unmanned Aerial Vehicle (UAV) point-of-view were recorded. The control output is divided into throttle, elevator, aileron, and roll. The CNN could fly through the racing tracks at high speed and even outperform state-of-the-art methods. Several skill levels were used to compare how data quality influences the learning pipeline. The authors conclude that better data results in better models and lap times.

Rodriguez-Hernandez at el. also demonstrated how a BC approach could control a Micro Aerial Vehicle (MAV) [23]. The network consisted of a CNN to extract features from the input images and return actions to fly the MAV through gates.

One exciting concept was introduced in [9]: collect drone crashing data to enable better learning through negative examples. One of the real-world fears of autonomous UAV systems is that it hits objects due to the low generalization of the NNs. The authors collected many crashing examples to be used alongside the positive samples, resulting in a robust policy. The negative data were demonstrated to be very important, and the results were comparable to humans in some environments.

As referred above, IL techniques are very used in autonomous driving [8, 24, 25]. All authors tried end-to-end trainable models, with some variations of CNN architecture connected to LSTM units or Fully Connected Networks (FCN). Curiously, [24] used a UAV to test this approach; nonetheless, the output is context-agnostic as the steering angle is used in any vehicle navigating the environment. Here, the CNN is connected to an LSTM network to ensure a good fit for the temporal-dependent actions. The environment used in [8] and [25] was a simulator, and both architectures output values relative to the vehicle's steering angle, based on the front-road images. Also, in both essays, CNN's last layers were FCN.

On a slightly different effort, [7] focused on a new benchmark to test the scalability of BC. In an autonomous driving setting, they confirmed, besides some well-known limitations, such as dataset bias and overfitting, generalization issues, and training instabilities. The requirement was to further research BC before putting such models in real-world driving.

Regarding aerobatics, [18] compares a standard FCN and an LSTM-based NN. Two controllers were developed, one for the ailerons and another for the elevators. The maneuver performed was an Immelmann turn, and the testing results revealed that the LSTM-based NN is better generalized, with better parameterisable values.

A thorough and complete work is [3], presenting a compilation of four published works and a subsequent article published later [4]. The general goal is to have an intelligent autopilot system, composed of a set of small FCNs, divided into specific flight phases, and a Flight Manager responsible for controlling them and selecting the appropriate ones based on a behavior tree.

With all applications of such techniques, especially to aerial vehicles, it is possible to conclude that applying BC in an airplane-control system is feasible. Also, LSTM-based NNs seem to be a good method to solve this problem.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

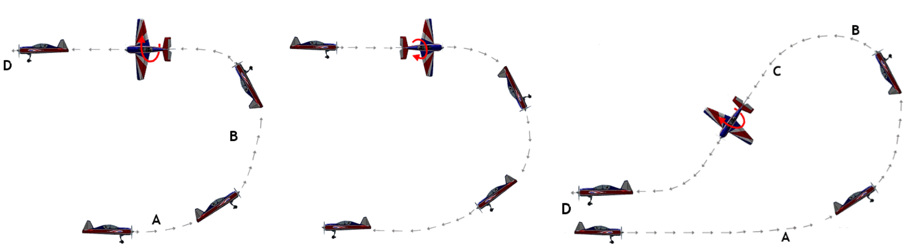
3 Data Analysis

The data collection phase counted on 25 amateur volunteers from the academic community (mainly BSc., MSc., and Ph.D. students, as well as professors associated with the department), with experience varying from first-time performances to practiced flight simulator enthusiasts. They were tasked with executing the aerobatic maneuvers to collect different examples. The idea was to gather a diverse dataset without the possible bias if collected from only 1 or 2 people. The collected values are relative to aircraft position and attitude, as well as environmental values such as atmospheric pressure and temperature; the airplane used was the aerobatics-capable Extra 300S. Microsoft Flight Simulator X (FSX) was the chosen flight simulator, used both to collect the data and test the controllers.

The complete dataset was published in Zenodo, on `https://zenodo.org/record/6803193`[4].

3.1 Maneuvers Description

The three maneuvers are composed of specific segments and must obey some 'rules' to be considered good; they are all composed of a main vertical component, upwards for Immelmann and Half Cuban Eight or downwards for Split-S. Figure 1 displays execution diagrams for each one. The Immelmann starts with an upwards semi-loop, followed by a semi-roll to stabilize the airplane. The Split-S comprises a semi-roll for the airplane to become upside down, followed by a semi-loop downwards until straight level flight is achieved. The Half Cuban Eight is similar to the Immelmann, with a 5/8 loop upwards, a semi-roll with a pitch of 45 degrees downwards, and a final 1/8 loop to level the airplane. A good Half Cuban Eight should have the exit point at the same altitude as the entry point; however, as the volunteers are amateurs, this component was not mandatory.



(a) Immelmann (b) Split-S (c) Half Cuban Eight

Fig. 1: Aerobatic maneuvers Diagrams

[4] Details on the dataset itself are in the *readme.me* file in the repository and specifics on the collection protocol in `https://hdl.handle.net/10216/143035`

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

3.2 Evaluation Metrics

Some metrics were developed to numerically quantify and then sort the collected examples according to their performance; some are maneuver-specific, while others apply to all maneuvers.

Initial and Final Heading (IFH) is the calculated error between the initial and final heading to establish vertical plane consistency; given the initial heading h_0 and the final heading h_f, both in degrees, the difference d is the smaller angle between both values $d = (h_f - h_0 + 180) \ \% \ 360 - 180$; normally d should be 180 degrees (opposite direction of entry and exit points), so $err = 180 - abs(d)$.

Heading Difference (HD) is the average error of every timeframe heading compared with initial and final values, a metric also related to vertical plane consistency that accounts for mid-execution drifts. Considering the same initial heading h_0 and final heading h_f, h_i is the heading at timeframe $i : 0..f$; d_{i0} and d_{if} are respectively the smaller angle between h_i and h_0 or h_f, calculated as d from the metric above. Then, from both d_{i0} and d_{if}, we consider the smallest (closer to one of the values), $sum = \sum_{i=0}^{f} min(\ abs(d_{i0}),\ abs(d_{if}))$ and finally $err = sum/s$, being s the size of the example.

Semi-Loop (S-L) measures how well the semi-loop trajectory fits a semi-circumference, as intended for the maneuvers; thus, it uses positional data points to find the best circumference curve[5]. Only the timeframes with the airplane pointing upwards (or downwards in Split-S) are considered. After this, the Euclidean distance $d_i, i : 0..f$ from each data point to the closest circumference point is calculated; the error for this metric is the average of the distances, making $sum = \sum_{i=0}^{f} d_i$ and $err = sum/s$.

Semi-Roll Overshoot (S-RO) is one of the metrics that evaluate how well the semi-roll is executed, focusing on whether or not there was some overshooting (more than 180 degrees of rotation). Similar to Semi-Loop error calculation, only the timeframes in specific conditions can be considered: pitch near zero and bank value far from straight level flight. This metric works by analyzing the evolution of bank values, looking for any eventual progression over the zero-angle bank value. The implementation uses the multiplication of successive bank values $b_i, i : 0..f-1$ to find a negative value (when bank goes from positive to negative or vice-versa). If found, future bank differences between consecutive frames d_{b_i} are summed. The formal definition is

$$err = \begin{cases} \sum_{n=i}^{f-1} d_{b_n} & \text{if } \exists i : b_i * b_{i+1} < 0 \\ 0 & \text{otherwise} \end{cases}$$

[5] Calculation based on the Coope method. Library API documentation available online at `https://scikit-guess.readthedocs.io/en/latest/generated/skg.nsphere_fit.html`.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

Semi-Roll Straightness (S-RS) measures the pitch variations during the semi-roll execution within the Half Cuban Eight maneuver, as it is important to keep the pitch constant, near 45 degrees downwards. With the same timestamp restriction as Semi-Roll Overshoot, the values considered in this set are relative to the pitch value and calculating standard deviation std, finally $error = std$.

Semi-Roll Altitude Consistency (S-RAC) measures altitude changes during the semi-roll part of a maneuver, as altitude should be kept constant. For this evaluation, all altitudes differences d_{a_i} between two consecutive values $a_i, i : 0..f$ are added in the form $err = \sum_{i=0}^{f-1} abs(d_{d_i})$.

Semi-Roll Completion (S-RC) When collecting Split-S examples, it was frequent for the volunteers to start the semi-loop before the semi-roll was complete, resulting in a less-than-180-degree semi-roll. This caused the airplane to have more significant side drift than supposed; to counteract on and reduce such wrongful actions, this metric was added, calculated as follows: the bank value b_i is correspondent to timeframe $i : 0..f$ from the semi-loop, d_{i0} and d_{i180} is the smaller angle between b_i and 0 (straight level flight) or 180 (upside down), respectively. Then, $sum = \sum_{i=0}^{f} min(\ abs(d_{i0}),\ abs(d_{i180}))$ and $error = sum/s$, s being the size of the loop timeframes set.

Total Evaluation is calculated as the sum of the singular metrics, using a set of weights to manipulate the individual error distribution in order to increase the impact of bigger errors (exponential – $e_i : i = 1..n$) and balance the distributions of the metrics (multiplication – $m_i : i = 1..n$). Thus, the final score is given by $evaluation = \sum_{i=1}^{n} err_i^{e_i} * m_i$.

3.3 Maneuvers Evaluation

Table 1 presents the metrics used for each maneuver, with the respective multiplicative weights; the exponential weights are all set to 1, except S-RS for Half Cuban Eight, which is set to 2.

Besides the two parts that compose the Immelmann (semi-loop and semi-roll), another thing to consider is the vertical plane consistency and the heading deviations from this initial plane, which are considered errors; this is common to the three maneuvers. Figure 2a shows the distribution of the metrics with the weights for all 162 collected examples.

Split-S is quite similar to the Immelmann maneuver, with the exception of some key details: the order of the operations and the direction. Figure 2b displays the metric distribution for the 153 collected examples. We note that Semi-Roll Completion is the component with the highest error, which is congruent with the difficulties felt during data collection, also negatively affecting other metrics.

Again, Half Cuban Eight is similar to Immelmann; the metric distribution for the 157 collected examples is depicted in Fig. 2c.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

Table 1: Multiplicative weights used for each maneuver

Metric	Immelmann	Split-S	Half Cuban Eight
IFH	2	0.6	1
HD	2	1	1
S-L	1	1	1
S-RO	1.4	-	1
S-RAC	0.2	0.5	-
S-RC	-	1	-
S-RS	-	-	0.5

4 Controllers Training

Throughout the development of this work, different feature combinations were tried until good results were obtained, with some feature engineering required. Due to space constraints, those experiments are not detailed here.

The used NN configuration is composed of a layer of LSTM cells, with the same amount of units as the number of features. A dropout layer was used with a value of 0.175, a learning rate of 0.005, and a batch size of 64. These values were obtained by a hyperparameter tuning step performed in [18]; also, 15 was the better window size to observe in the LSTM layer.

Since the examples were all executed with full throttle on the airplane and the maneuvers do not require rudder control, only controllers for the elevator and aileron were trained. The specific features for each one are listed in Table 2.

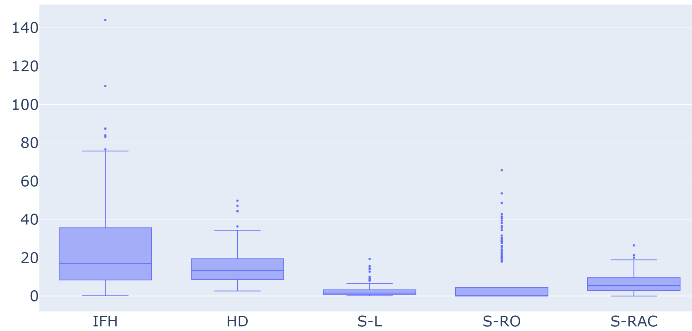
With the evaluation metrics developed, it is possible to select the best examples for training the models. While using only the best examples for training is likely to provide for better results, a balance must be obtained between the quality and the number of samples used for training, as a lower amount of samples is likely to lead to worst results. For this step, several subsets were chosen to train the controllers and compare the results. The chosen subsets are as follows:

- **100%** or **All**: the control group with all collected samples;
- **Best 90%**: removing the worst 10%, as those are likely to include outliers and bad examples;
- **Best 75%**: mid value between 90 and 60%;
- **Best 60%**: for the most restricted subset, we use 60% of the collected data; considering the variability necessary to train generalizable controllers, any less than that seemed like a considerable reduction in sample size.

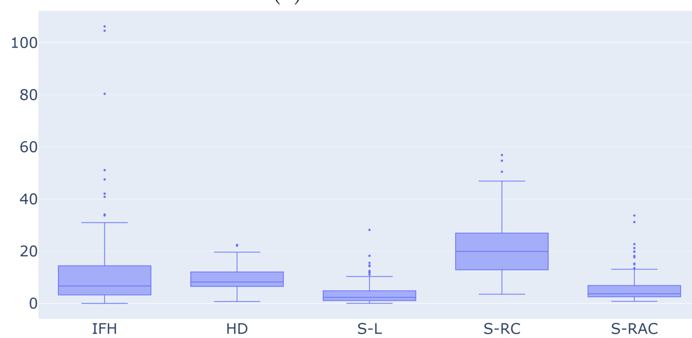
4.1 Results

To gather results, the test consisted in positioning the same aircraft in similar conditions in the simulator and having the controllers guide it through a full

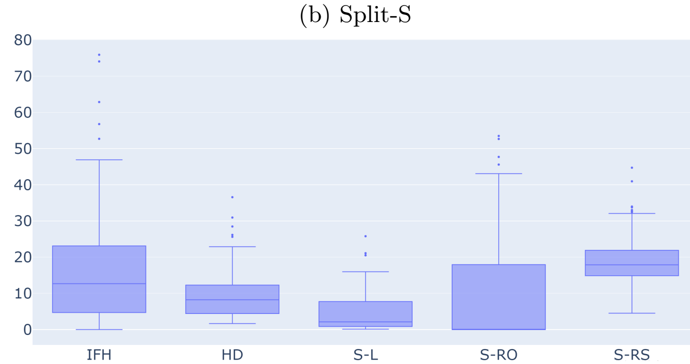
ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15



(a) Immelmann



(b) Split-S



(c) Half Cuban Eight

Fig. 2: Metric evaluation distribution for the three maneuvers

execution of the respective maneuver. Hence, five sequential maneuver executions for each previously mentioned version were collected and evaluated using the previously developed metrics; the evaluations for each version were compared with the remaining versions and the human-collected examples (shown as 'Examples' in the graphics).

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

Table 2: Features used in each controller

Maneuver	Elevator	Aileron
Immelmann	Angle of Attack Pitch Bank Velocity World y Velocity Body z	Angle of Attack Pitch Bank Velocity Rotation Body y Elevator
Split-S	Angle of Attack Pitch Bank Velocity World y Velocity Body z Aileron	Angle of Attack Pitch Bank Velocity Rotation Body y
Half Cuban Eight	Angle of Attack Pitch Bank Velocity World y Velocity Body z Aileron	Angle of Attack Pitch Bank Velocity Rotation Body y Elevator

Immelmann The results fulfilled the expectations. The graph presents a favorable evolution with the increase in data quality. The 75% model shows great maturity and consistency in the results. The best 60% model (less than 100 examples used in training) appears to have suffered from overfitting or low variability of input examples.

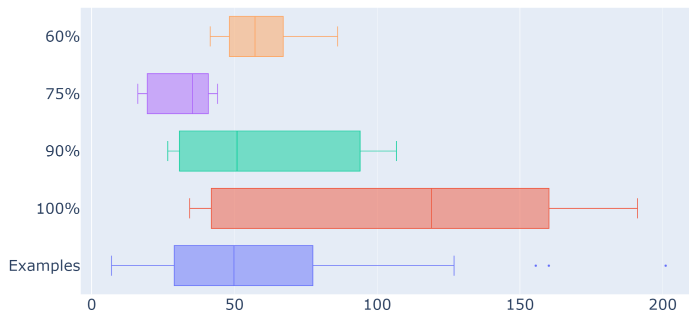
The five examples for each version were grouped and compared in Fig. 3a. The evaluations are visible in the five markings of each box: minimum value, first quartile, median, third quartile, and maximum value. The X axis is relative to the error evaluations – lower is better.

Split-S Similar to Immelmann, it is possible to notice the increase in performance, although less significant, when training with better data; however, it is most noticeable that the examples have more consistent evaluations, as depicted in Fig. 3b. It is clear the best model is the best 90%, as it gets all five executions with consistent quality relative to the others and even four within the best 25% of the collected data.

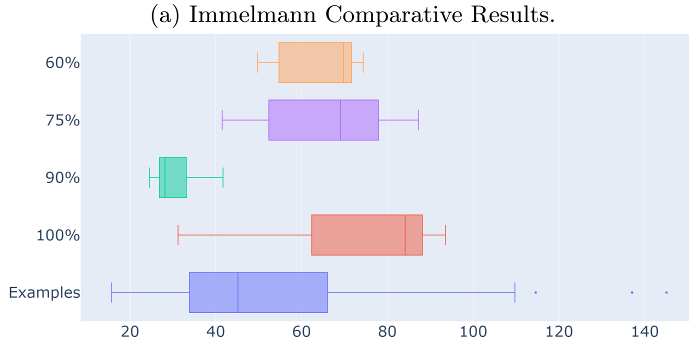
Half Cuban Eight Observing Fig. 3c, the model trained with the best 60% examples presented the best results. However, a strange phenomenon arose: it

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

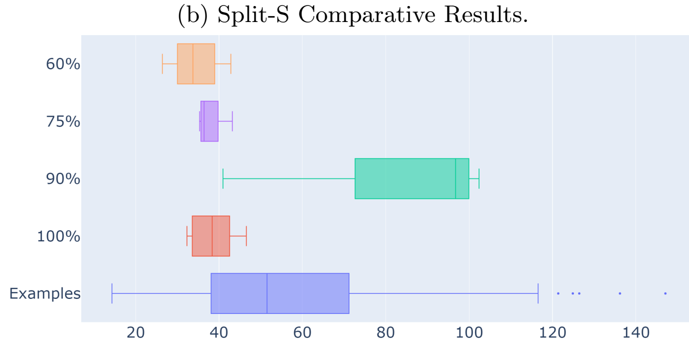
only started the maneuver with an airspeed close to 200 knots – it was found that the minimum speed to perform a complete semi-loop was 140 knots, as was advised to all volunteers. 200 knots is not an effortless speed to reach in a straight flight with this particular aircraft since its maximum speed is only 240 knots. Therefore, the best 75% model was the best; the base model (100%) was close, but the consistency was decisive.



(a) Immelmann Comparative Results.



(b) Split-S Comparative Results.



(c) Half Cuban Eight Comparative Results.

Fig. 3: Maneuvers comparative results. X-axis represents the total evaluation calculated by the metrics - lower is better

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

4.2 Discussion of Results

The results reflect an expected phenomenon, the difficult balance between using quality data to train the models and the amount and diversity of the data available for the training process. The Immelmann results showcase this phenomenon rather well, presenting a threshold when results stop improving with better data quality and start worsening due to poor amount/diversity.

Unfortunately, the results cannot be directly compared to those obtained by Medeiros [18], which is the closest work to the one conducted by this study. The evaluation metrics used are different, as is the aircraft model used. Medeiros measured the performance by stipulating a required altitude for the Immelmann controller to finish the action. This was possible because the aircraft used, a Boeing F/A-18, is capable of maintaining upwards flight without losing speed – it is very versatile and powerful since it is used in military forces. The Extra 300S, however, does not have enough motor thrust force to perform that ascent, which means it cannot perform the Immelmann maneuver with any specified target altitude. Therefore, in the training process, the target altitude was found not to be accomplished, probably functioning as a noise feature. Despite the different metrics, both works show promising results, considering that a similar network configuration was capable of learning piloting controls.

Regarding these results, five examples do not allow any advanced statistics; therefore, the conclusions may be slightly skewed since the boxes might be different when using 50 examples for each version. This was not possible due to heavily-manual data collection since it is not easy (or even possible) to fully automate this process. However, the trained controllers were capable of piloting the airplane, even though the data was from a non-expert sample. The metric evaluation and consequent examples selection also exhibited partial success, showing better controllers with better data.

5 Circuit Controller

The set of controllers proved to execute well the respective maneuvers, which is enough to develop a high-level controller. To perform an aerobatic performance, it needs to select the correct maneuver, detect when each execution is over, and proceed to the next one.

The idea behind this phase is to automate an aerobatic performance that could be performed by a real-world airplane. Although this is an experimental step, the focus is to reach an advanced high-level controller that can correctly choose the best maneuver suited for the specific flight.

The main execution of the Circuit Controller (CC) was the state machine, where it iterates through the maneuvers' controllers' actions until it finishes all of them. Between maneuvers, the airplane flies in a straight line for 10 seconds, using FSX built-in autopilot capabilities, so that the maneuvers can be easily isolated and identified, and the aircraft can gain some speed for the next maneuver. For the conducted tests, the order of execution was as follows:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

1. **Immelmann**: best 75% trained controller;
2. **Straight Level Flight**: FSX autopilot;
3. **Half Cuban Eight**: best 75% trained controller;
4. **Straight Level Flight**;
5. **Split-S**: best 90% trained controller;
6. **Straight Level Flight**.

Figure 4 shows the trajectory completed when performing the circuit in two different views. The circuit starts at point A, followed by B and C, marking the beginning and end of the Immelmann maneuver. Until point D, the FSX autopilot controlled the airplane in a straight level flight, where the Half Cuban Eight controller started piloting. With E indicating the end, a small deviation is noticeable in heading, both in the semi-loop (curve closer to D) and in the semi-roll (curve closer to E). The Split-S goes from F to G, and it is also showing some signs of deviations when executing the semi-roll, right after point F.

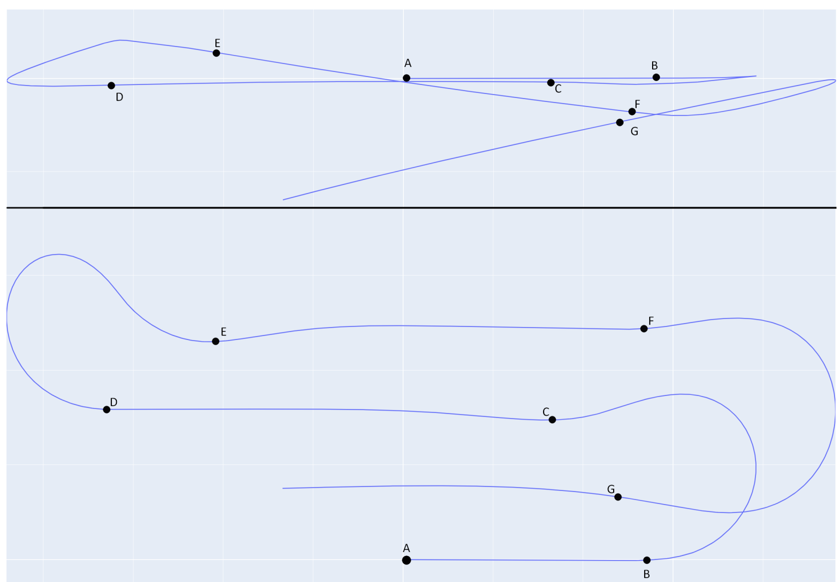


Fig. 4: Circuit views: The top part shows a top (YX) view, while the bottom part shows a side (YZ) view

This system could be evolved and used in automated air vehicle performance shows (something like Automated Red Bull Air Race or Automated Drone Racing League). From what is possible to understand of Baomar's Flight Manager [3], which is the most similar work found, CC is a similar system with an also similar use case, only differing in context: Flight Manager is used in a regular commercial flight, while CC is focused on more complex aerobatic maneuvers.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

6 Conclusion

With this work, it is possible to conclude that with better demonstration data the controllers' performance also improves with IL. The technique developed by Medeiros [18] was expanded into other aerobatic maneuvers and a different airplane, showing consistency in the aerial environment.

It was interesting to use non-expert data and still successfully train controllers; besides, after excluding the worst examples by evaluating the amateur data, the controllers' performance also benefited.

We also present the Circuit Controller, an automated system that iterates through the different maneuvers, composing an aerobatic performance show. This system selects between the trained controllers and the built-in autopilot functionalities present in FSX.

This work can be extended or tuned in several ways. Regarding the evaluation metrics, some deeper study about the maneuvers is advised, such as better implementations or relations between the metrics – the multiplicative and exponential weights. Also, normalizing the values used by the metrics is a great way to reduce the impact of example scales.

Another idea is to encapsulate each individual low-level controller in an agent responsible for detecting maneuver completion and bad execution, ensuring fail-safe actions, and guaranteeing operational security [10].

The extension or adaptation of CC usage is expected to be the main focus in future research, with the goal of extending its use cases; one possibility is to have a simple user interface that receives waypoint or mission information, and the system decides which is the better route or actions, be it in the context of military operations, or simply when using a more maneuverable aircraft.

Instead of training individual maneuvers, a more scalable solution might be to train a generalized controller that can act on any required change. The maneuvers can be divided into smaller steps, such as 100 consecutive waypoints with coordinates, heading, pitch, and bank target information.

Acknowledgements The authors would like to acknowledge all volunteers who provided us with the invaluable data that allowed for this work to be developed.

References

1. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robotics and Autonomous Systems **57**(5), 469–483 (May 2009), DOI: 10.1016/j.robot.2008.10.024
2. Attia, A., Dayan, S.: Global overview of Imitation Learning. arXiv preprint p. 9 pages (January 2018), DOI: 10.48550/arXiv.1801.06503
3. Baomar, H.: Using Learning from Demonstration to Enable Automated Flight Control Comparable with Experienced Human Pilots. Ph.D. thesis, University College London (2020), https://discovery.ucl.ac.uk/id/eprint/10108999/

4. Baomar, H., Bentley, P.J.: Autonomous Flight Cycles and Extreme Landings of Airliners Beyond the Current Limits and Capabilities Using Artificial Neural Networks. Applied Intelligence **51**(9), 6349–6375 (September 2021), DOI: 10.1007/s10489-021-02202-y
5. Caballero, W.N., Gaw, N., Jenkins, P.R., Johnstone, C.: Toward Automated Instructor Pilots in Legacy Air Force Systems: Physiology-Based Flight Difficulty Classification Via Machine Learning. SSRN (July 2022), DOI: 10.2139/ssrn.4170114
6. Claiborne, M.: How Does Autopilot Work on a Plane? AeroCorner: `https://aerocorner.com/blog/how-does-autopilot-work/` (NA), accessed: April 2023
7. Codevilla, F., Santana, E., Lopez, A., Gaidon, A.: Exploring the Limitations of Behavior Cloning for Autonomous Driving. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2019), October 27 - November 2 2019, Seoul, South Korea. vol. 2019-October, pp. 9328–9337 (2019), DOI: 10.1109/ICCV.2019.00942
8. Farag, W., Saleh, Z.: Behavior Cloning for Autonomous Driving using Convolutional Neural Networks. In: Proceedings of the 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), November 18-20 2018, Sakhier, Bahrain. pp. 189–195 (2018), DOI: 10.1109/3ICT.2018.8855753
9. Gandhi, D., Pinto, L., Gupta, A.: Learning to Fly by Crashing. In: Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017), September 24-28 2017, Vancouver, Canada. pp. 3948–3955 (2017), DOI: 10.1109/IROS.2017.8206247
10. Garrido, D., Ferreira, L., Jacob, J., Silva, D.C.: Fault Injection, Detection and Treatment in Simulated Autonomous Vehicles. In: Proceedings of the 20th International Conference on Computational Science (ICCS 2020), 3-5 June 2020, Amsterdam, The Netherlands. Lecture Notes in Computer Science, vol. 12137, pp. 471–485. Springer, Cham (2020), DOI: 10.1007/978-3-030-50371-0_35
11. Hua, J., Zeng, L., Li, G., Ju, Z.: Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning. Sensors **21**(4), 21 pages (2021), DOI: 10.3390/s21041278
12. Hussein, A., Elyan, E., Gaber, M.M., Jayne, C.: Deep Imitation Learning for 3D Navigation Tasks. Neural Computing and Applications **29**(7), 389–404 (April 2018), DOI: 10.1007/s00521-017-3241-z
13. IATA: 20 Year Passenger Forecast. IATA: `https://www.iata.org/pax-forecast/` (2021), accessed: April 2023
14. IBISWorld: Global Airlines - Market Size 2005–2027. IBISWorld: `https://www.ibisworld.com/global/market-size/global-airlines/` (2021), accessed: April 2023
15. Kober, J., Peters, J.: Learning Motor Primitives for Robotics. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA 2009), May 12-17 2009, Kobe, Japan. pp. 2112–2118 (2009), DOI: 10.1109/ROBOT.2009.5152577
16. Kotoky, A., Yap, C.: A Shortage of Pilots Looms as the Next Challenge for Airlines. Bloomberg: `https://www.bloomberg.com/news/articles/2021-09-21/a-shortage-of-pilots-looms-as-the-next-challenge-for-airlines` (2021), accessed: April 2023
17. Mburu, W.: An In-Depth Study of the Pilot Shortage and its Consequences. Oklahoma State University (2017)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15

18. Medeiros, C.: Learn to Fly: Cloning the Behavior of a Pilot. Master's thesis, Faculty of Engineering, University of Porto (July 2021)
19. Michie, D., Camacho, R.: Building Symbolic Representations of Intuitive Real-Time Skills from Performance Data. In: Furukawa, K., Michie, D., Muggleton, S. (eds.) Machine Intelligence 13: Machine Intelligence and Inductive Learning, chap. 15, pp. 385–418. Oxford University Press, Inc. (October 1994), ISBN: 978-0-19-853850-9
20. Morales, E.F., Sammut, C.: Learning to Fly by Combining Reinforcement Learning with Behavioural Cloning. In: Proceedings of the 21st International Conference on Machine Learning (ICML'04), July 4-8 2004, Banff, Alberta, Canada. p. 8 pages (July 2004), DOI: 10.1145/1015330.1015384
21. Müller, M., Casser, V., Smith, N., Michels, D.L., Ghanem, B.: Teaching UAVs to Race: End-to-End Regression of Agile Controls in Simulation. In: Proceedings of the European Conference on Computer Vision (ECCV 2018) Workshops, September 8-14 2018, Munich, Germany. pp. 11–29 (2018), DOI: 10.1007/978-3-030-11012-3_2
22. Osa, T., Pajarinen, J., Neumann, G., Bagnell, J.A., Abbeel, P., Peters, J.: An Algorithmic Perspective on Imitation Learning. Foundations and Trends in Robotics **7**(1-2), 1–179 (2018), DOI: 10.1561/2300000053
23. Rodriguez-Hernandez, E., Vasquez-Gomez, J.I., Herrera-Lozada, J.C.: Flying through gates using a behavioral cloning approach. In: Proceedings of the 2019 International Conference on Unmanned Aircraft Systems (ICUAS 2019), June 11-14 2019, Atlanta, GA, USA. pp. 1353–1358 (2019), DOI: 10.1109/ICUAS.2019.8798172
24. Saksena, S.K., Navaneethkrishnan, B., Hegde, S., Raja, P., Vishwanath, R.M.: Towards Behavioural Cloning for Autonomous Driving. In: Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC 2019), February 25-27 2019, Naples, Italy. pp. 560–567 (2019), DOI: 10.1109/IRC.2019.00115
25. Samak, T.V., Samak, C.V., Kandhasamy, S.: Robust Behavioral Cloning for Autonomous Vehicles Using End-to-End Imitation Learning. SAE International Journal of Connected and Automated Vehicles **4**(3) (2021), DOI: 10.4271/12-04-03-0023
26. The World Bank: Air Transport, Passengers Carried. The World Bank: `https://data.worldbank.org/indicator/IS.AIR.PSGR` (2021), accessed: April 2023
27. Wall, R., Tangel, A.: Facing a Critical Pilot Shortage, Airlines Scramble to Hire New Pilots. The Wall Street Journal: `https://www.wsj.com/articles/pilot-shortage-spurs-hiring-spree-1533720602` (2018), accessed: April 2023

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_15