THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	Learning High-Speed Flight of Unmanned Aerial Vehicle in Cluttered Environments
Author's name:	Vit Knobloch
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Karel Zimmermann
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — **challenging**

How demanding was the assigned project?

Please insert your comments here.

Fulfilment of assignment — **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

Please insert your comments here.

Methodology — **correct**

Comment on the correctness of the approach and/or the solution methods.

Please insert your comments here.

Technical level — **B - very good.**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

Please insert your comments here.

Formal and language level, scope of thesis — **A - excellent.**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

Please insert your comments here.

Selection of sources, citation correctness — **A - excellent.**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

Please insert your comments here.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

Please insert your comments here.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Student fulfilled the assignment of the bachelor thesis. Despite of PPO being reasonable choice for RL, the proposed solution suffers from (i) sample inefficiency and (ii) poor generalization (see question below for further discussion and suggestions). Nevertheless, the overall approach is correct, its experimental evaluation seems fair, and it opens the space for an interesting future research.

The grade that I award for the thesis is **A - excellent.**

Question to be discussed during the defense:

1. Table 5.3. shows comparisons with existing approaches. The caption claims: "The other methods don't account for perception awareness, which disadvantages the presented method since it has to keep a suitable yaw angle". Would not be reasonable to change the criterion of your method in order to match the criterion being evaluated (i.e. the time)?
2. Section 5.4 claims: "The policy was not able to learn to fly through the testing tracks but was able to fly through the training tracks. Given the small size of the policy network, it is unlikely that the policy was able overfit to the combination of five different tracks of this length. This shows a tendency of the RL approach toward generalization.". While first part of this claim says that the generalization is bad (typically due to overfitting or significant mismatch between training/testing distribution), the last sentence makes impression that the RL has tendency towards generalization. Can you disentangle these claims?
3. In conclusions you claim that "high durability against disturbances and model mismatch errors"? Does not the inability to generalize on testing tracks suggest, that it is the other way around? Would it be possible to test it explicitly?
4. What about simple baselines such as (i) PID controller that tries to follow collision-free trajectory with some kind of obstacle safety margin? (ii) DWA controller for dynamic obstacle avoidance? Or (iii) MPC controller that optimizes the progress reward?
5. The sample-inefficiency of RL as well as the poor generalization of neural networks is the main source of problems of the proposed solution. Consider addressing these issues in future research as follows:
   a. What about replacing the policy by something more explainable such as simple network that sets parameters (e.g. cost function, safety margins) of a controller, e.g. adaptive DWA [1]?
   b. What about supervised learning of the policy from globally optimal trajectories? I am not sure, but I guess that exhaustive search with some kind of LQR-RRT could provide reasonable trajectories.
   c. What about combining (a) and (b): Perform supervised learning on ground truth trajectories of the policy consisting from a simple network followed by the differentiable MPC [3].

[1] https://ras.papercept.net/images/temp/IROS/files/2250.pdf
[2] http://motion.cs.illinois.edu/RoboticSystems/GeometricMotionPlanning.html
[3] paper: https://arxiv.org/abs/1810.13400 NeurIPS, 2019
implementation: https://github.com/locuslab/differentiable-mpc

Date: **25.5.2023** Signature: