Thesis

1. Introduction & Motivation

Generalizationability is important for agents, Effect of Buffer underresearched.

2. Hypothesis

Hypothesises:

- "SpiderWeb Buffer" best option
- Generalization improved through higher variance of starting positions
- Way Buffer is filled has impact on generalizability of agent

3. Background

Explaining following concepts:

- CMDP:
  explaining formal part, as importat to understand what context, state, actions, state action pair means
- Online/Offline Learning:
  explaining replay buffer and what difference is between onlien/offline learning and why it matters for us
- (Zero-Shot) Generalization:
  What does generalization mean, why does it matter and what options exist to improve it so far,
- UVF's:
  What is an UVF, how does it work? why would you use it, what are the drawbacks (why not just use an uvf for the whole agent)
- RND:
  What is RND, how does it work, what is it used for?

4. Literature Study

Papers that will be mentioned:

- Max paper
- Go Explore
- HER
- Generalized Hindsight for Reinforcement Learning
- Efficient Self-Supervised Data Collection for Offline Robot Learning

5. Methodology

This chapter will explain how things were implemented and how they work and what they do, what their idea was and why they should work.

5.1 Architecture

actual agent is the same in all the implementations, DQN with CNN etc,

5.2 Four Room Minigrid

Explain the enviornment, how it works, what its based on, action & observation space, What context changes, explaining reachability of states etc.

5.3 Teleportation

Explaining how it works, RND inclusion, why it should work, proof of concept, idea taken from go explore, could be considered cheating but if this doesnt work nothing does, hyperparameter tp percentage,

5.4 Random-Walk

Explaining how it works, no RND, pure uniform action sampling, hyperparameter step length,

5.5 Intrinisic Exploration Random Walk

Smarter Extension of Random-Walk with intirinsic exploration, explain implementation of intrinsic exploration

5.6 UVF

Explain idea, explain architecture of uvf agent, explain how algorithm works, explain various tried methods (uvf value decreases) and downfall,

6. Experiments

This chapter will showcase various results of the previously explained methods with silly little graphs. Also will come to some conclusions on why something worked/hasn't worked.

6.1 Teleportation

6.2 Random-Walk

6.3 Intrinisic Exploration Random Walk

6.4 UVF

Explain that uvf doesn't work with small buffersize, even with big buffersize too sampleinefficient,

7. Conclusion

Drawing conclusion from the data, which method worked the best, which offers the biggest advantage with least drawback,

A.1 Appendix

All the tables of hyperparameters that were used etc.

Questions:

- Where do I mention stable-baselines? in Architectures?