Project Proposal  
SIMILAR: Submodular Information Measures Based Active Learning In Realistic Scenarios

Group name and members: boinet. Lisa Wimmer, Sven Lorenz, Andreas Klaß.

## Paper in a nutshell

* <https://arxiv.org/pdf/2107.00717v1.pdf>
* Use submodular information measures as acquisition functions in active learning
  + Handle different modes by using relationship between SIM

## **Define** the different **building blocks** that need to be implemented

1. Standard training loop for model
2. Similarity Matrix with hypothesized labels + data handling (simulating rare classes, ood etc.)
3. Use different modes of Submodular Functions on Similarity Matrix

## Clearly define what **requirements** the final **code** must fulfill

* Implement different general SIMILAR framework
* Different implemented SMI Functions
* Handle different situations (standard AL, rare classes, out of distribution data and data redundancy)

## How will the **framework** be **tested**

* Test SIMILAR framework against paper results
* Model can be tested against other pretrained models

## What do we start with? Which code parts already exist?

* Standard AL mode similar to BADGE: <https://github.com/JordanAsh/badge>
  + Some computational tricks are the same as well

## Particularly **challenging** parts:

* Make sure computationally feasible using tricks in paper
* Get different modes correctly
* Reduce code redundancies by using Appendix Table 1

## Main contribution

* Implementing unified active learning approach.

## Related works and references

* <https://github.com/JordanAsh/badge>
* <https://github.com/decile-team/distil>
* https://github.com/gudovskiy/al-fk-self-supervision

## Appendix

 