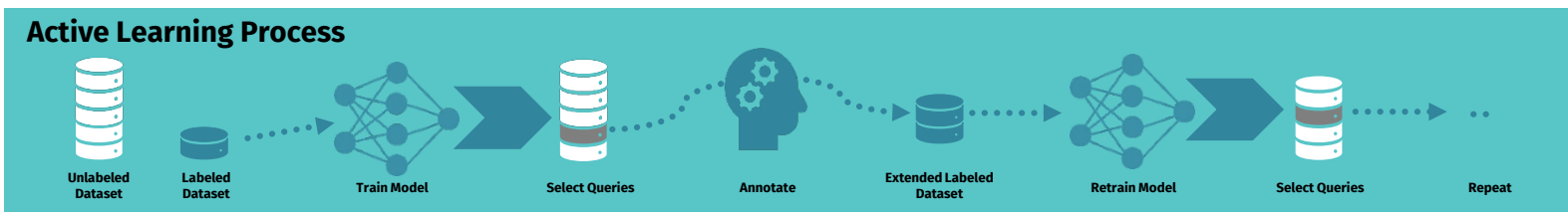
KI DELTA LEARNING
Scalable AI for Automated Driving

Methods for identification of training samples for the training of Neural Networks in the context of active learning

Alexander Wölker, Nicolas Thewes, Syed Saqib Bukhari, Georg Schneider

WP 3.3



Use-Case in Delta Learning

Delta Active Learning: Starting with an initially trained Network further training procedures are performed using actively sampled data in order to decrease labeling effort.

Technical Problem

Labeled data is an obligatory requirement for a lot of machine learning applications, at the same time getting labeled data is often time-consuming and/or expensive.
The field of active learning deals with the question, how the amount of labeled training data for a machine learning algorithm can be reduced by choosing "the right" samples. More specifically, this rises two important questions: Firstly, how to evaluate the informativeness of a given unlabeled sample (query or sampling strategy), where informativeness denotes the benefit of a sample regarding the generalization error reduction of a machine learning model. Secondly, how to find the right label of a chosen sample. Also, thereby two approaches exist, a human annotator (human in the loop), or algorithmic (model-based) solutions.

Technical Solution

We conduct experiments in order to evaluate the effectiveness of different approaches in the field of active learning.
Therefore the available training data is randomly split into a pool of labeled and pool of (pretended) unlabeled data. Initial training is performed using the pool of labeled data. The "unlabeled" data is added iteratively in portions. Thereby, the next set is chosen randomly, according to an uncertainty measure (here we use entropy calculated on the basis of dropout class probability or softmax output) or by using the LLAL [1] procedure, see Figure 1. The identified unlabeled set is then augmented by its correct label. This procedure is repeated until all unlabeled data consumed.

Evaluation

We conducted experiments using following parameters: We added 1000 samples from the pool of unlabeled samples on each iteration and further trained for 200 epochs. Additionally we investigated the effect of resetting the model on each iteration and of data augmentation. The work is under progress.

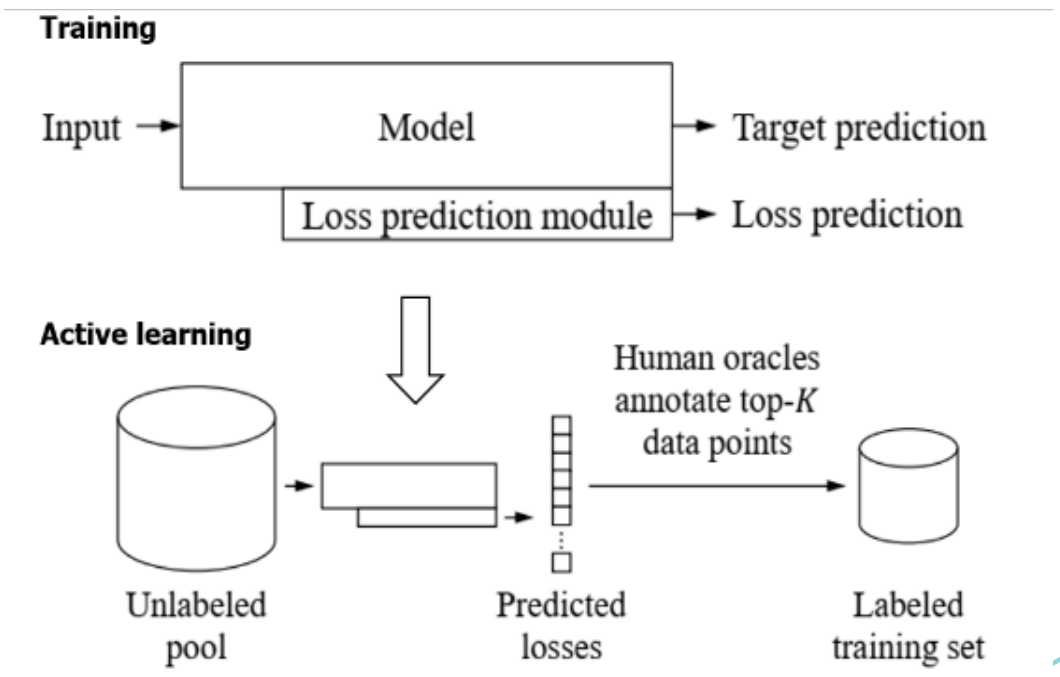


Figure 1: Sketch of the basic architecture of the LLAL [1] procedure.

References:
[1] "Learning Loss for Active Learning", Yoo, D., & Kweon, I. S., et al. 2019

For more information contact:
alexander.woelker@zf.com

Partners

External partners

KI Delta Learning is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.

Supported by: Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag

www.ki-deltalearning.de @KI_Familie KI Familie