# Data-Efficient and Iterative Metric Learning for Open Set Classification in an Industrial Setting

Samuel Kees – HOCHSCHULE DARMSTADT

## MOTIVATION

Many Machine Learning (ML) algorithms are known to need a lot of curated and labeled data to be trained. The industry often could easily provide raw data, but lacks the knowledge needed to apply it to ML [1]. This poses the challenge of acquiring relevant data and processing it to a degree where it becomes usable for ML [2]. A common way to provide this information is through labeling. For the industry, labeling is much harder than just providing the raw data. The reason for this is, that labeling in the industry often requires human experts, who are always scarce and expensive. This gives an incentive to improve the labeling process so that it can be done by non-experts or machines, with minimal expert supervision.
This work proposes a concept, namely the Data Refinery (DR), to efficiently label data by exploring and exploiting semantic information, based on Deep Metric Learning (DMeL). The goal is to reduce the labeling effort from experts and non-experts alike. The proposed concept is first applied in an industrial setting to create Feature Extractors (FE), which use Deep Computer Vision (DCV) to detect semantic information in images with the aim to then discriminate objects via clustering. Hence, this thesis provides a scheme of experts, non-experts, and machines working successfully in tandem to generate curated data sets for ML tasks.
The industrial setting, in which this work was conducted, is part of the German “Pfandsystem”. Germany has a deposit-refund system called "Pfandsystem" for collecting and returning packaging materials, especially bottles. It is defined in the German law "Verpackungsgesetz (VerpackG)" (packaging act)[4]. One aspect is to reuse bottles, this comes with a logistic effort to sort, clean and refill the bottles. Schulz Systemtechnik GmbH developed the Crate Inspector (CrI) to detect the bottle-types of the bottles inside a crate without unpacking the bottle, to solve the sorting task. For this the CrI uses traditional Computer Vision. In this work, the concept of the DR was tested on the described task to make ML algortihms feasible.

## METHODOLOGY

Figure 1 shows a simplified version of the DR concept applied to the described use case. First a FE is trained via Metric Learning (Figure 1.C) by using Tidy Data (Figure 1.A). The trained FE maps data samples with similar features in a Feature Space (FS) close to each other. With the help of the provided Tidy Data, the FE is capable of finding interesting Untidy Data samples, which are worth showing to a human. A score function determines how interesting a data sample is (Figure 1.D). A select function is used as a filter to control the score. The goal of this process is to simplify the task of the human (Figure 1.E), which is to label only the interesting data, tidying it in the process (Figure 1.F). So that in the next iteration, a FE can be trained with more Tidy Data (Figure 1.C).
For training a loss function defines how much can be learned. The score aims to simulate the expected loss for a given Untidy Data sample. This is not easy, because some assumptions about the Untidy Data sample have to be made. In order to make these assumptions, the FS of the Tidy Data can be used. One way to exploit the FS is to put the feature embedding of the Untidy Data sample in the FS and look at the neighborhood (Figure 1.C).

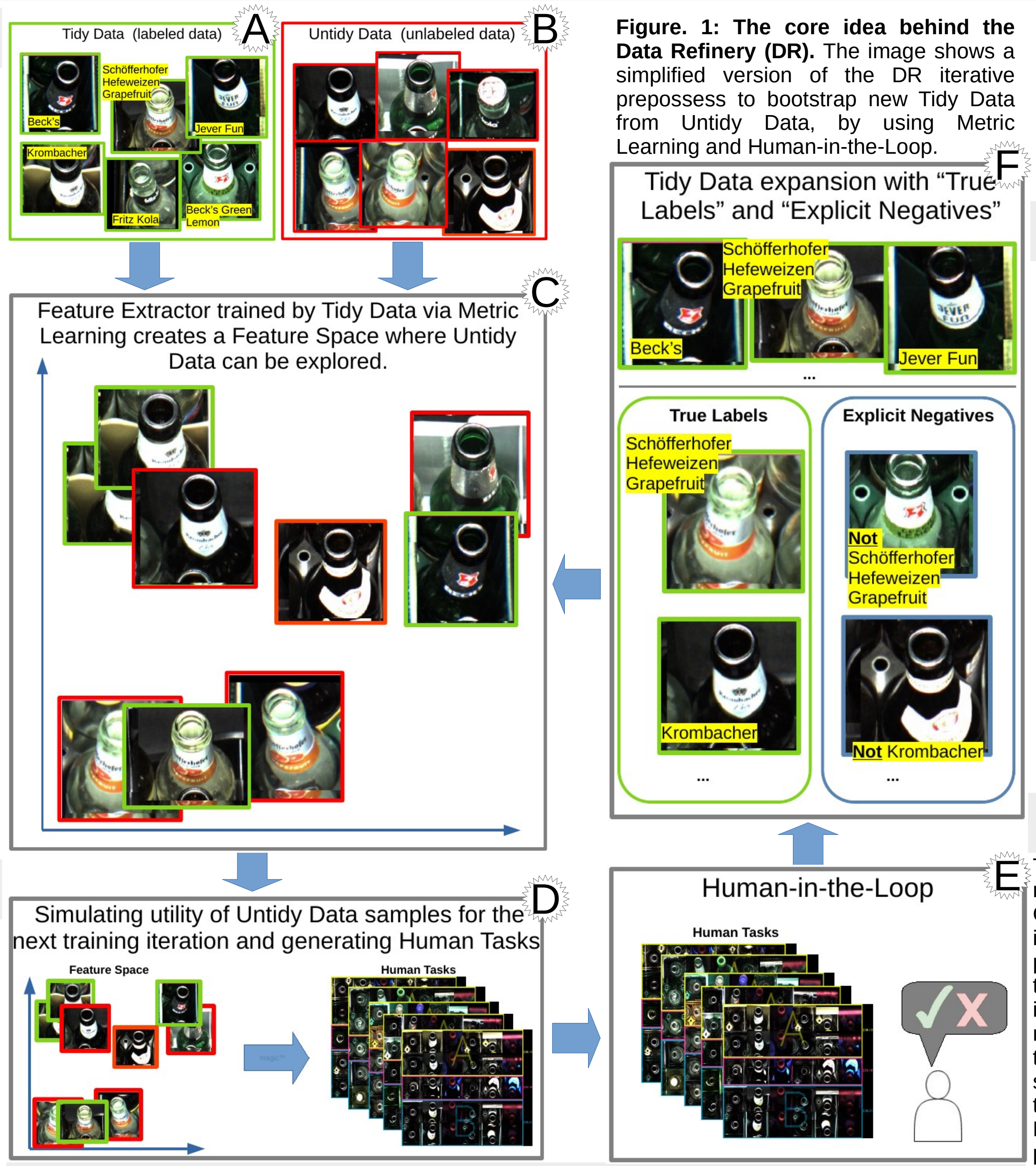


**Figure. 1: The core idea behind the Data Refinery (DR).** The image shows a simplified version of the DR iterative prepossess to bootstrap new Tidy Data from Untidy Data, by using Metric Learning and Human-in-the-Loop.

The select function is responsible for the selection of the Untidy Data samples shown to a human. The most straightforward way could be by selecting Untidy Data samples with the highest score. In this work two DRs were implemented which only differ in their selection. One DR selected Untidy Data samples to show to a human based on the score (Data Refinery Score Sampling (DR-S)) and the other randomly without considering the score at all (Data Refinery Random Sampling (DR-R)).

## RESULTS

In the overall performances per iteration DR-R and DR-S showed different performance values in the initial iteration for all measured metrics (e.g. DR-R: ~58% and DR-S: ~60% Micro Precision (MiP)). For the second iteration a decrease in performance for DR-R and DR-S was observed. DR-S was able to recover after one additional iteration in contrast to DR-R which needed two iterations. After iteration four, respectively five, not much of an improvement of the performances of DR-R or DR-S were observed. In the last (tenth) iteration DR-R ended with ~61% and DR-S with ~62% MiP.
In the end of every DR iteration new data samples were added for DR-R and DR-S. The classes that showed significant increases in relative class frequency over all iterations were examined. DR-R showed not much of an improvement in performance for either classes where DR-R or DR-S increased the relative class frequency. DR-S, on the other hand, showed an improvement in performance for classes for which it increased the relative class frequency.
It could also be observed that the time humans needed to answer a "question" (task) asked by DR-R was decreased over the iterations. This could not be shown for "questions" asked by DR-S where the annotation time fluctuated. In most cases, the median annotation time required to answer a "question" asked by DR-S was higher than the compared annotation time of DR-R.

## CONCLUSION

The results indicate that DR-S, in contrast to DR-R, specifically adds new data samples for training in classes where it underperformed. Contrary to the expectations, this did not lead to an overall improvement of DR-S when compared with DR-R, but to a similar performance of ~61% for Micro Precision (MiP). When the results of this case study are compared with other deep metric learning works, it underperformed [3]. The main reasons for this, could be that in the mentioned work, only experts labeled the human tasks and the training was up to 10 times longer. Regardless, the goal of the case study was not to train a state of the art classification model for a given task, but to efficiently mine data for such a model.
Even though the success of the case study is only moderate performance wise, the new approach of the DR as a concept proves to be a promising way to acquire specific data that can be used to create curated data sets. Such curated data sets, are not only valuable for the industry (i.e., bottle-type recognition task), but can be used for all kinds of ML tasks throughout all industries.
Before this concept can be applied, the parts of the case study that proved to be problematic should be revised. Also, certain aspects of the general concept, as an example: the score function, should be optimized or alternative options reviewed.

## References

[1] Stephen C-Y. Lu. "Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation." (Jan. 1, 1990)
[2] Thorsten Wuest, Daniel Weimer, Christopher Irgens, and Klaus-Dieter Thoben. "Machine learning in manufacturing: advantages, challenges, and applications." (Jan. 1, 2016).
[3] Yin Cui, Feng Zhou, Yuanqing Lin, and Serge Belongie. "Fine-Grained Categorization and Dataset Bootstrapping Using Deep Metric Learning with Humans in the Loop." (Jun. 1, 2016)
[4] Verpackungsgesetz § VerpackG

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

**Kontakt**
Samuel.Kees@Stud.h-da.de

Fachbereiche Mathematik und Naturwissenschaften & Informatik

**Referent**
Prof. Dr. Andreas Thümmel

**Korreferent**
Prof. Dr. Markus Döhring

SCHULZ SYSTEMTECHNIK