KI ABSICHERUNG
Safe AI for Automated Driving

Analysis and Comparison of Datasets by Leveraging Data Distributions in Latent Spaces

Hanno Stage, Lennart Ries, Eric Sax

DNN specific Safety Concern: Inadequate separation of test and training data

Domain Shift

The performance of deep neural networks suffers from several insufficiencies. One central insufficiency is the ability to generalize from given training data. If the underlying domain of the input data differs from the domain of the training dataset, the prediction accuracy may decrease. Therefore, it is important to identify the domain shift during the development. Domain shifts can be global shifts such as weather changes or new locations. Moreover, our method can detect an unknown domain shift between the training, validation, and test dataset, which could lead to wrong conclusions about the performance of the deep neural network.

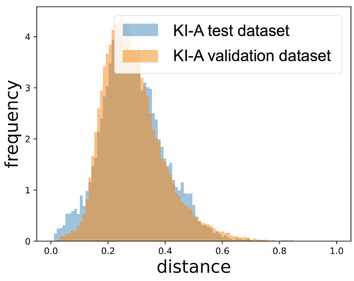


Figure 1: The histogram shows the result of a comparison of the KI-A data from tranche 2, 3, 4, 5 datasplit.

Fig. 1 shows the comparison of the data split from KI-A project. Fig. 2 shows a domain shift in the Oxford Robotcar dataset. The comparisons with both data splits were performed using the JointVAE and the z-score as distance metric in latent space, which performed best in experiments with a variety of VAE architectures, metrics and datasets.

Safety Hypothesis:
The method deals with the monitoring of the input data. It can prevent that there is a domain shift between different data splits. (SC-2.2-**Inadequat seperation of test and training data**)

Mechanism

Our method is based on the functionality of Variational Autoencoder (VAE). For our method, a VAE is trained to provide meaningful latent space representations. Based on the aggregated distances between single samples in latent space, a domain shift can be calculated. The goal was to detect domain shifts between two data sets in an unsupervised way,

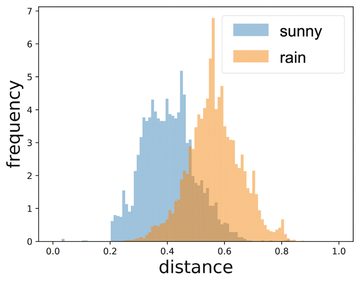
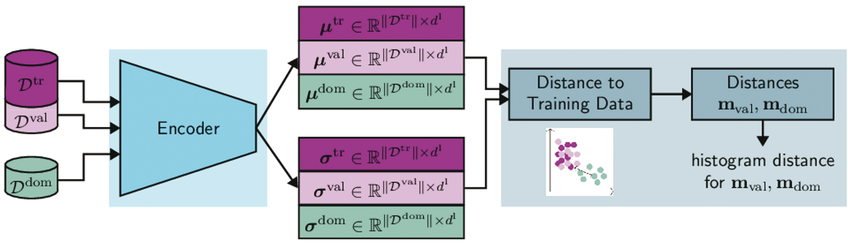


Figure 2: The histogram shows the result of a comparison of a data split from the Oxford RobotCar dataset where sunny and rain images are compared.

and without the need of a specific network or task. A process diagram of our method can be seen in Fig. 2.

References:

Stage, H. et al. (2022). Analysis and Comparison of Datasets by Leveraging Data Distributions in Latent Spaces. In T. Fingscheidt, H. Gottschalk, S. Houben et al.. Deep Neural Networks and Data for Automated Driving. Springer.
doi: 10.1007/978-3-031-01233-4_3



FZI

For more information contact:
stage@fzi.de
ries@fzi.de
sax@fzi.de

KI Absicherung 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 Climate Action.

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

KI FAMILIE

https://www.ki-absicherung-projekt.de @KI_Familie KI Familie

Image sources: FZI - Forschungszentrum Informatik