SAND2021-8315D

The Center for Cyber Defenders

Expanding computer security knowledge

Image Similarity for Large Scale Non-Invasive Inspection

Hugh Galloway, University of Texas at San Antonio

Project Mentor: Dan Krofcheck (8722)

Problem Statement:

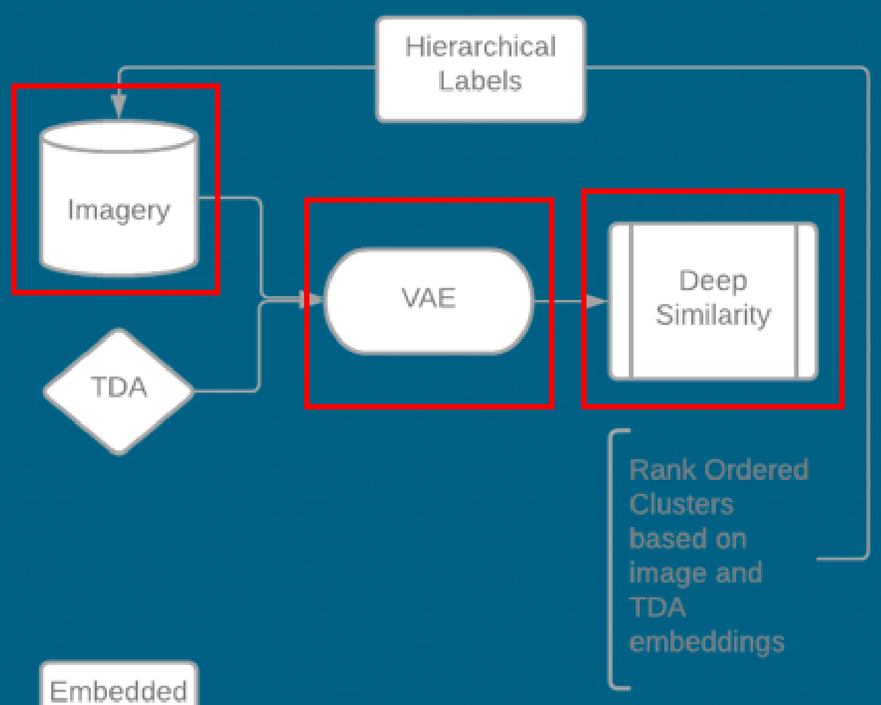
Customs and Border Protection (CBP) operators at US Land Ports of Entry are responsible for screening all cargo that enters the United States using combinations of vehicle metadata and on-site vehicular X-ray scans. The goal of this project is to produce an unsupervised approach to assist operators with X-ray adjudication.

Objectives and Approach:

We utilize deep similarity and change detection to characterize areas where anomalous change is likely to occur. The research presented here is the deep similarity workflow, shown in RED.

Our approach groups similar images by clustering the image embeddings from a pretrained convolutional variational autoencoder (CVAE) modified by the Sigma-VAE loss function [2].

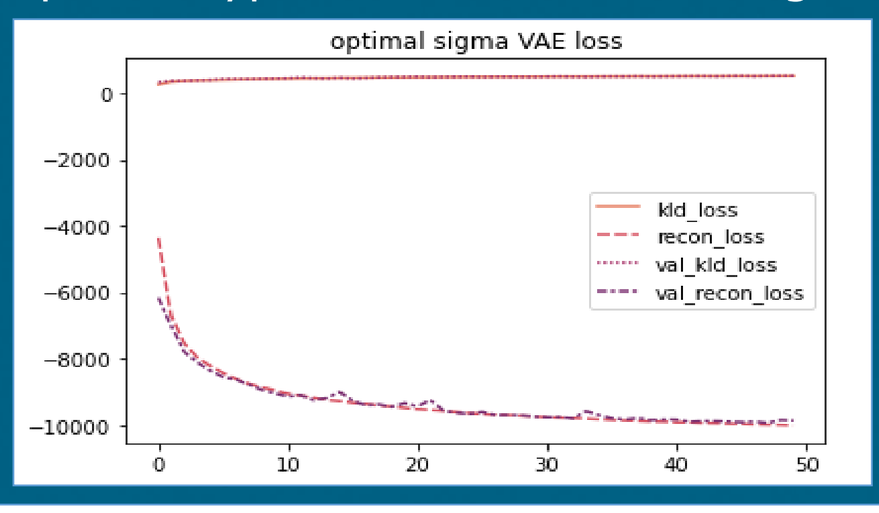
After pretraining, images are assigned a cluster probability from a clustering algorithm, HDBSCAN [1]. This property allows for selection of images that best represent a cluster, informing future change detection efforts.

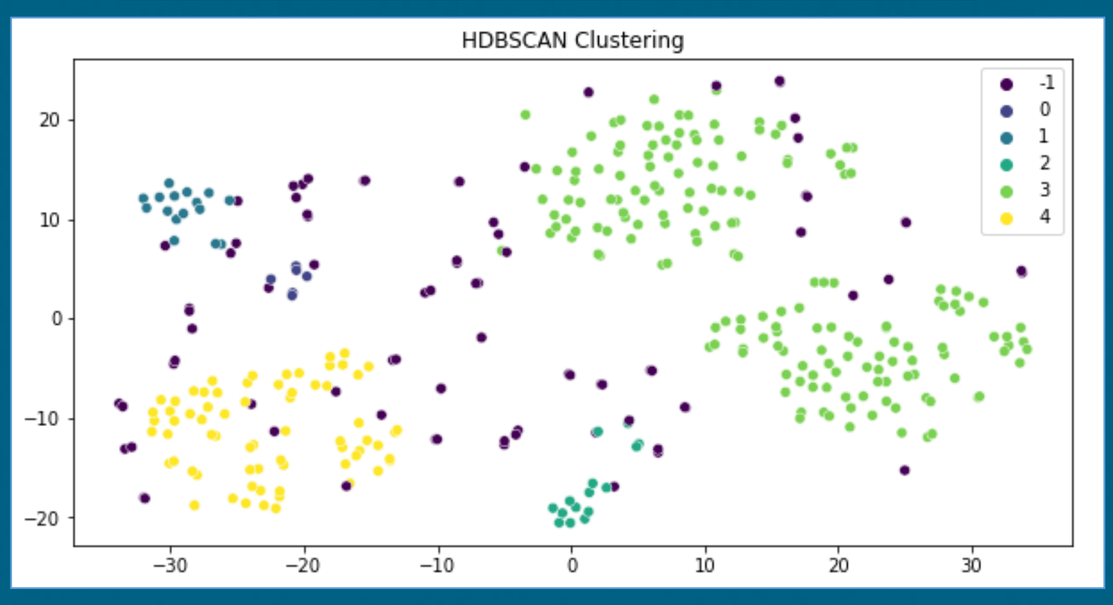


Results

The CVAE was pretrained on the Tiny ImageNet dataset [3] and evaluated on a stream of commerce CBP dataset made available to SNL. The CVAE embeddings captured image features which resulted in informative clusters. For example cluster 1 only holds truck cabs and cluster 2 consists of trucks with flatbed trailers with no cargo. Using k-nearest neighbors to rank similar images showed that repeat instances of a image were often identified as the most similar image.

Moving forward we plan to use the cluster probabilities from HDBSCAN to generate idealistic cluster centroids, allowing us to identify common areas of change belonging to specific types of vehicles and cargo.





Impact and Benefits:

The success of the deep similarity workflow enables us to apply change detection on unlabeled data to assist with vehicular X-ray adjudication.

References

[1] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *JOSS*, vol. 2, no. 11, p. 205, Mar. 2017, doi: 10.21105/joss.00205.

[2] O. Rybkin, K. Daniilidis, and S. Levine, "Simple and Effective VAE Training with Calibrated Decoders," *arXiv:2006.13202 [cs, eess, stat]*, Aug. 2020, Accessed: Jul. 07, 2021. [Online]. Available: http://arxiv.org/abs/2006.13202

[3] P. Chrabaszcz, I. Loshchilov, and F. Hutter, "A Downsampled Variant of ImageNet as an Alternative to the CIFAR datasets," *arXiv:1707.08819 [cs]*, Aug. 2017, Accessed: Jul. 07, 2021. [Online]. Available: http://arxiv.org/abs/1707.08819

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2017-XXXXC.

U.S. DEPARTMENT OF ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.