Software Impacts 11 (2022) 100195

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

CVAD - An unsupervised image anomaly detector

Xiaoyuan Guo [a,*], Judy Wawira Gichoya [b], Saptarshi Purkayastha [c], Imon Banerjee [d]

[a] Department of Computer Science, Emory University, United States of America
[b] Department of Radiology and Imaging Sciences, Emory University, United States of America
[c] School of Informatics and Computing, Indiana University-Purdue University Indianapolis, United States of America
[d] Arizona State University, School of Computing, Informatics, and Decision Systems Engineering, United States of America

ARTICLE INFO

Keywords:
Anomaly detection
OOD detection
Variational autoencoder

ABSTRACT

Detecting out-of-distribution samples for image applications plays an important role in safeguarding the reliability of machine learning model deployment. In this article, we developed a software tool to support our OOD detector **CVAD** - a self-supervised **C**ascade **V**ariational autoencoder-based **A**nomaly **D**etector , which can be easily applied to various image applications without any assumptions. The corresponding open-source software is published for better public research and tool usage.

Code metadata

Current code version	V1.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2021-167
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/3191573/tree/v1
Legal Code License	MIT
Code versioning system used	git
Software code languages, tools, and services used	Pytorch, Python
Compilation requirements, operating environments & dependencies	Ubuntu 18.04.5
If available Link to developer documentation/manual	https://github.com/XiaoyuanGuo/CVAD/blob/main/README.md
Support email for questions	xiaoyuan.guo@emory.edu

1. Introduction

Despite the recent advances in deep learning that have contributed to solving various complex real-world problems, the safety and reliability of AI technologies remain a big concern in medical applications [1]. Deep learning models for medical tasks are often trained with data from known distributions, and fail to identify out-of-distribution(OOD) outputs and possibly assign high probabilities to the anomalies during inference because of the insensitivity to distribution shifting. To ensure the reliability of deep learning models' predictions, it is necessary to identify unknown types of data that are different from the training data distribution. However, the OOD data can be infinite to enumerate and unavailable during training. To train an anomaly detector with only in-distribution(ID) data available, learning high-quality "normality" features is the fundamental step to identify the OOD samples during inference [2]. Inspired by [3], we propose CVAD [4], which is built on top of a branch-cascaded VAE. With the architecture to model the in-distribution representations unsupervised, CVAD gains superior reconstructions and learns good-quality features to threshold out the OOD data. The ability of CVAD to detect anomalies is further enhanced through training a binary discriminator with the constructed data with random perturbations on aforementioned cascade VAE's latent parameters as OOD category.

2. Description

There are two main components of CVAD — the generator G and the discriminator D. The generator is composed of a branch cascade

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: xiaoyuan.guo@emory.edu (X. Guo), banerjee.imon@mayo.edu (I. Banerjee).

https://doi.org/10.1016/j.simpa.2021.100195
Received 24 November 2021; Received in revised form 29 November 2021; Accepted 29 November 2021

variational autoencoder. By inputting in-distribution(ID) images to it, the generator learns to reconstruct the original data and minimizes the reconstruction loss. During this process, the generator models the latent embeddings of in-distribution, which formulates the "normality" features and provides the basis of excluding the outliers that show dissimilarity. The reconstructed images are then treated as fake OOD data for the second-stage discriminator to distinguish from the original ID data. As a consequence, the discriminator learns to assign each data sample an anomaly probability.

2.1. Anomaly score computation

CVAD utilizes both the generator and discriminator to identify anomalies. To keep it unified, the reconstruction error output S^G in the first stage is scaled into [0, 1] with $S^{G'} = S^G - S^G_{min}/S^G_{max} - S^G_{min}$. As the anomaly probability S^D given by the discriminator is in range [0, 1], the final anomaly score is calculated via $S = 0.5 * (S^{G'} + S^D)$, which ensures the anomaly score is also in the range of [0, 1].

2.2. Usage

The CVAD architecture is implemented in Python and relies on the Pytorch [5] library. To train the CVAD model, two steps should be taken. First, get the image generator G trained. As there are some situations that images are in large and irregular sizes, we resize the images into a standard size of 256×256 before feeding the data into the model. We also provide an example for cifar10 dataset [6], and design a corresponding small architecture considering its image size $32 \times 32 \times 3$, please refer the code for more details. To adapt the model for various applications, we set the *channel* parameter of CVAD's generator and discriminator to be changeable. As medical images are usually in grayscale, the corresponding network *channel* is thus set to 1. When the input images are in RGB format, the *channel* parameter should be set to 3. The variation autoencoder are notorious for instability in training and probably will get nan loss. Giving this situation, the learning rate of generator should be set relatively small to avoid the bad loss, 0.00001 is one of the good values to start. As the learning rate is set small to get stable converge, the training epochs should be set relatively larger, 500 epochs for example. Given a dataset that has multiple classes, one or more classes can be identified as normal classes, a parameter *normal_class* helps provide the users with the option to select which classes to be normal and the left ones as abnormal data automatically. The generator will use normal data for training, and the normal data will be split into training (80%) and validation (20%) parts.

The discriminator D is trained after the generator, which will be used again during the discriminator training to generate reconstructions of the ID data but not optimized. The discriminator will be optimized to distinguish the generated data from the original one. This phase will no take too long as the difference between the reconstructions and the raw inputs is clear, it is relatively easy for the binary discriminator to learn it. Therefore, a few epochs will be enough, 10–20 epochs for example.

2.3. Evaluation

We also provide the evaluation part for reporting model performance. For evaluation, we mix up the ID data (i.e., the validation data aforementioned) and the OOD data (i.e., the abnormal classes) together. To calculate the model's performance, we assign the ID data with label 0s and the OOD data as 1s for the anomaly targets. With the trained generator and the discriminator, the final anomaly scores are obtained as predictions. By using scikit-learn package roc_curve function, three standard metrics — TPR, FPR, AUC score are calculated.

3. Impact overview

Since our CVAD anomaly detector pose no assumptions on the original input data, it can be easily applied to various situations. Its generalization can be helpful in solving various tasks. The implementation also has a few parameters to set up, which would be friendly to new users. We hope this work could facilitate the research in this field and provide new researchers a convenient tool to start from.

4. Limitations and developments

As CVAD takes advantages of VAE-based approach and Classifier-based approach, where the VAE performs as an image generator and the classifier acts as the discriminator. Users should first train the generator and then the discriminator model. The initial architecture of CVAD was designed for images with size $256 \times 256 \times$ *channel*. To use the model for large images, resizing or changing the architecture should be done.

5. Conclusions

We have developed an image OOD detector called CVAD, it has good generalizations on various application situations and can be easily reproducible. Researchers and developers can take advantage of this tool for test and experiments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was funded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), USA MIDRC grant of the National Institutes of Health under contracts 75N92020C00008 and 75N92020C00021 and the US National Science Foundation #1928481 from the Division of Electrical, Communication & Cyber Systems.

References

[1] Litjens Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, Clara I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[2] Guo Xiaoyuan, Judy Wawira Gichoya, Saptarshi Purkayastha, Imon Banerjee, Margin-aware intra-class novelty identification for medical images, 2021, arXiv preprint arXiv:2108.00117.

[3] Zimmerer David, Jens Petersen, Klaus Maier-Hein, High-and low-level image component decomposition using VAEs for improved reconstruction and anomaly detection, 2019, arXiv preprint arXiv:1911.12161.

[4] Guo Xiaoyuan, Judy Wawira Gichoya, Saptarshi Purkayastha, Imon Banerjee, CVAD: A Generic medical anomaly detector based on cascade VAE, 2021, arXiv preprint arXiv:2110.15811.

[5] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, et al., Pytorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019) 8026–8037.

[6] Krizhevsky Alex, Geoffrey Hinton, Learning multiple layers of features from tiny images, Vol. 7, 2009.