Unsupervised image anomaly detection via likelihood estimation on high-level representation

César A. Salcedo
Department of Computer Science
Universidad de Ingeniería y Tecnología
Lima, Peru 15063
cesar.salcedo@utec.edu.pe

Marco A. Alvarez
Department of Computer Science and Statistics
University of Rhode Island
Kingston, RI 02881
malvarez@uri.edu

Abstract

Unsupervised image anomaly detection is a challenging task in which a model is expected to discriminate between normal and anomalous images, given that it has only been exposed to normal images during training. A recent branch of development relies on generative models to learn the underlying distribution of normal, in-distribution samples. The trained model can then ideally be used to identify out-of-distribution samples [1, 8, 9]. One promising type of generative model for this setup are normalizing flows, since they allow for exact likelihood estimation of individual samples by applying a sequence of invertible transformations in reverse order. However, most of state-of-the-art normalizing flow approaches for anomaly detection obtain the likelihood of a sample from raw pixels, which does not give accurate results, and can be computationally expensive for some flow models (e.g. Glow [3]).

To alleviate this issue, we propose an encoder-flow pipeline for anomaly detection using an encoder and normalizing flow for likelihood estimation on high-level features. The encoder receives an image and outputs a latent space vector as the compact representation of the sample. The latent vector is then passed through the normalizing flow to compute its likelihood. The use of an encoder is supported by previous work in which likelihood estimation from image embeddings has yield better results on anomaly detection compared to likelihood estimation from raw pixels [4]. The notion of encoder can be generalized to any function that maps an image to latent space, ideally reflecting the most characteristic features of a sample with its corresponding latent vector. In particular, on this paper we explore the use of an encoder trained as part of an autoencoder.

Given the autoencoder and the encoder-flow pipeline, training is performed on two stages. On the first stage, the autoencoder attempts to improve the reconstruction quality of images passed through the latent space bottleneck by minimizing the L_1 distance from the original image to its reconstruction. On the second stage, the normalizing flow is trained over the latent space vectors obtained from the encoder (of fixed weights) to approach a normal distribution. After training, a clear line to define which samples are in- or out-of-distribution can be drawn by selecting a threshold according to a confidence interval from likelihood estimations of all training set samples.

We follow on the experimental setup employed by [1] to measure the performance of methods for anomaly detection. On this regard, we select labeled datasets MNIST [6] and CIFAR-10 [5] for training and validation. In this setup, nine out of ten classes are labeled normal and used for training, while the remaining class is labeled anomalous and kept for validation. Additionally, we expect any dataset different from the one used for training to be considered out-of-distribution. For

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

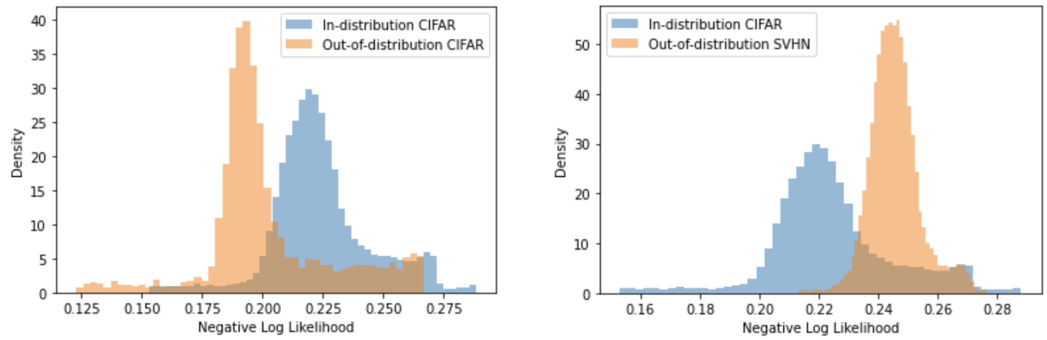

Figure 1: Distribution of likelihood estimations for samples from in-distribution CIFAR-10 dataset compared to *(left)* out-of-distribution CIFAR-10 class *plane*, and *(right)* out-of-distribution SVHN dataset. *(left)* The distribution of normal images has higher likelihood than the distribution of anomalous images, and shows a clear separation from it. *(right)* Note the distribution of normal images has *lower* likelihood than the distribution of anomalous images.

this reason, we also select SVHN as a validation dataset on our experiments. The decoder and encoder architectures are based on the generator and discriminator components of a DCGAN respectively [7]. The normalizing flow chosen for experimentation is Real-NVP [2].

The results of training and evaluating the model on CIFAR-10 with anomalous class *plane* are shown on Figure 1 *(left)*. As expected, there is a clear separation between the distribution of normal and anomalous images. However, as shown on Figure 1 *(right)*, the out-of-distribution dataset is assigned a higher likelihood than the dataset used for training, which results in low performance on anomaly detection. This results align to ones obtained on previous studies of normalizing flows for anomaly detection [4]. A comparison with Ganomaly [1] shown on Figure 2 confirms our proposal still requires further improvements to achieve state-of-the-art results.

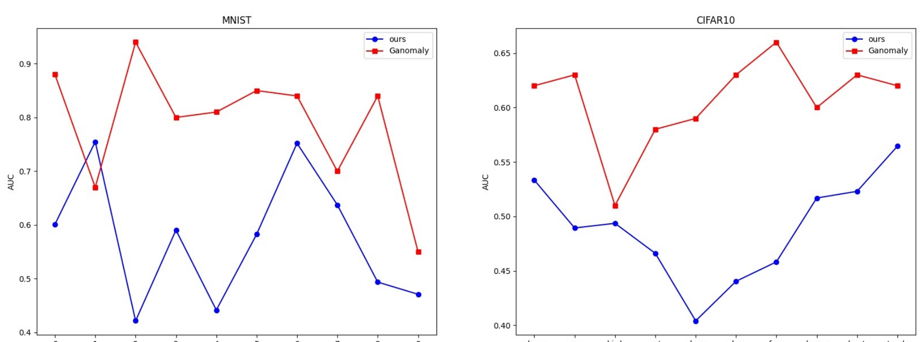

Figure 2: Comparison between our approach and Ganomaly [1] measured by the AUC score. Each value on the x-axis represents a class selected as anomaly and used for evaluation, from which an AUC score can be computed.

This work presents a novel encoder-flow architecture for unsupervised image anomaly detection. The results so far suggest more work must be done to understand the mechanics of normalizing flows in anomaly detection, and why they assign overall lower likelihood to training samples compared to samples taken from other datasets. Additionally, another future research path is replacing the encoder from the autoencoder by a modern encoder architecture as a means to obtain better sample representations for likelihood estimation.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018.

[2] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[3] Diederik P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *arXiv preprint arXiv:1807.03039*, 2018.

[4] Polina Kirichenko, Pavel Izmailov, and Andrew Gordon Wilson. Why normalizing flows fail to detect out-of-distribution data. *arXiv preprint arXiv:2006.08545*, 2020.

[5] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research).

[6] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.

[7] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[8] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.

[9] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar. Efficient gan-based anomaly detection. *arXiv preprint arXiv:1802.06222*, 2018.