Deep Semi-supervised Anomaly Detection Using VQ-VAE

Renuka Sharma[1,2,3,4], Hengcan Shi[1], Jianfei Cai[1], Suyash P. Awate[2] , and Nick Birbilis[5]

[1]Monash University, Melbourne.[2]Indian Institute of Technology (IIT) Bombay, Mumbai.

[3]IITB-Monash Research Academy, Mumbai. [4]CSIRO's Data61, Brisbane. [5]Deakin University, Melbourne.

Motivation

- Anomaly detection is a one-class classification (OCC) problem
- Previous methods used unsupervised learning, semi-supervised approach seems to help a lot
- Incorporating vector quantization (VQ) in autoencoders since images contain a lot of redundant information since most of the pixels are correlated.
- We aim to learn a VQ-dictionary containing the latent embeddings for normal data points.

Introduction

- We propose a novel semi-supervised variational learning based DNN framework leveraging vector quantization (VQ) models.
- Our semi-supervised framework ss-VQ-VAE (Semi-supervised Vector Quantized Variational Autoencoder) is an end-to-end one-class classification model for anomaly detection.
- ss-VQ-VAE can leverage some outlier data during training to improve performance.
- It leverages the VQ- VAE [1] model where the latent space embedding is discrete rather than continuous to generate high-fidelity reconstruction.
- We introduce an anomaly score to better adapt VQ-VAE to anomaly detection, which compares the encoded features of the input with the dictionary embeddings.
- More accurate anomaly detection results can be obtained by our anomaly score
- We show improvements on the real-world datasets.
- We also curate an industrial inspection corrosion dataset for semi-supervised anomaly detection and show our superior performance and applicability on it.

Semi-supervised VQ-VAE (ss-VQ-VAE)

- ss-VQ-VAE is an extension of VQ-VAE which takes in a fraction of outliers Y along with the inliers X as part of the training set which helps in refining the normal data points boundary. We have separate encoder-decoder based losses for the normal and abnormal data points and they are weighted by the hyperparameter η and we minimize the loss as follows: $\arg\min_{\theta^{\mathcal{EQD}}} \eta\mathcal{L}_{\text{normal}}(\{X_n\}_{n=1}^{N};\theta^{\mathcal{EQD}})$

$$+ (1-\eta)\mathcal{L}_{\text{anomalous}}(\{Y_m\}_{m=1}^{M};\theta^{\mathcal{EQD}})$$

Thanks to CSIRO's Autonomous Sensors Future Science Platform (AS FSP) for funding the travel and presentation of paper at DICTA 2023.

Learning Framework

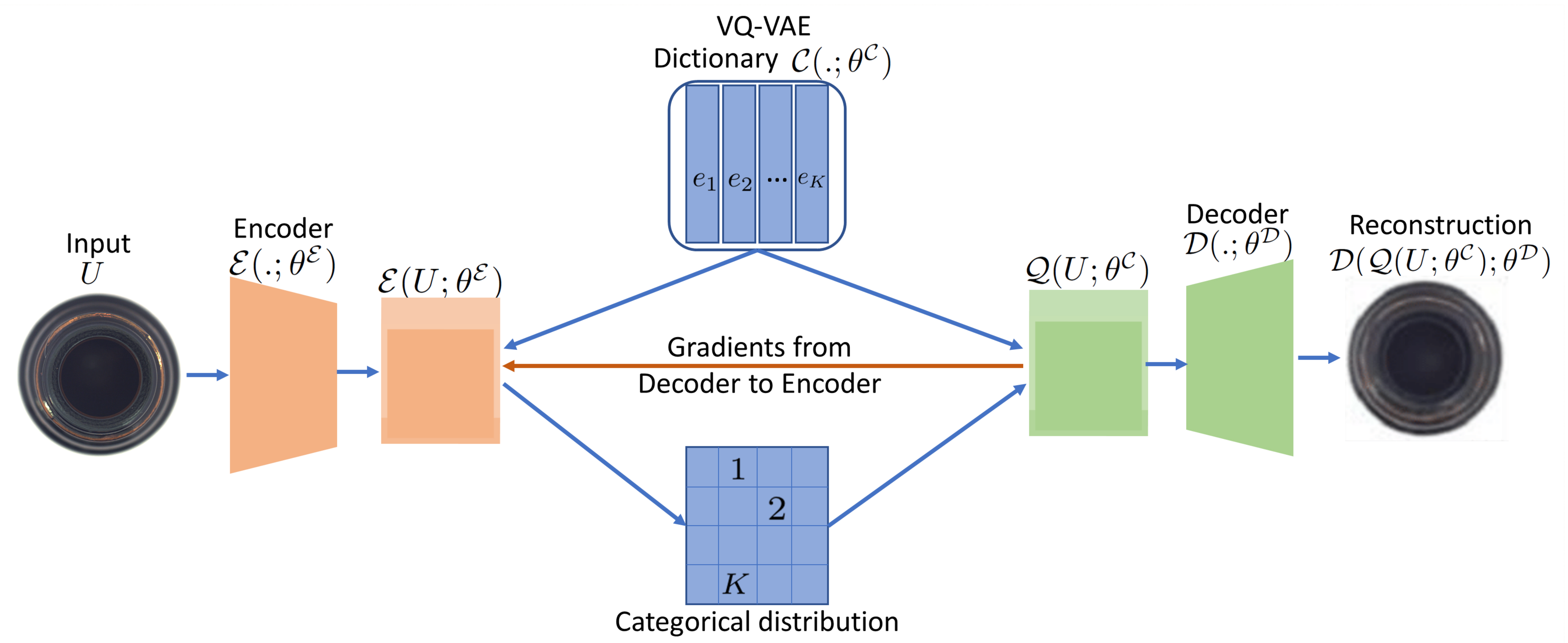
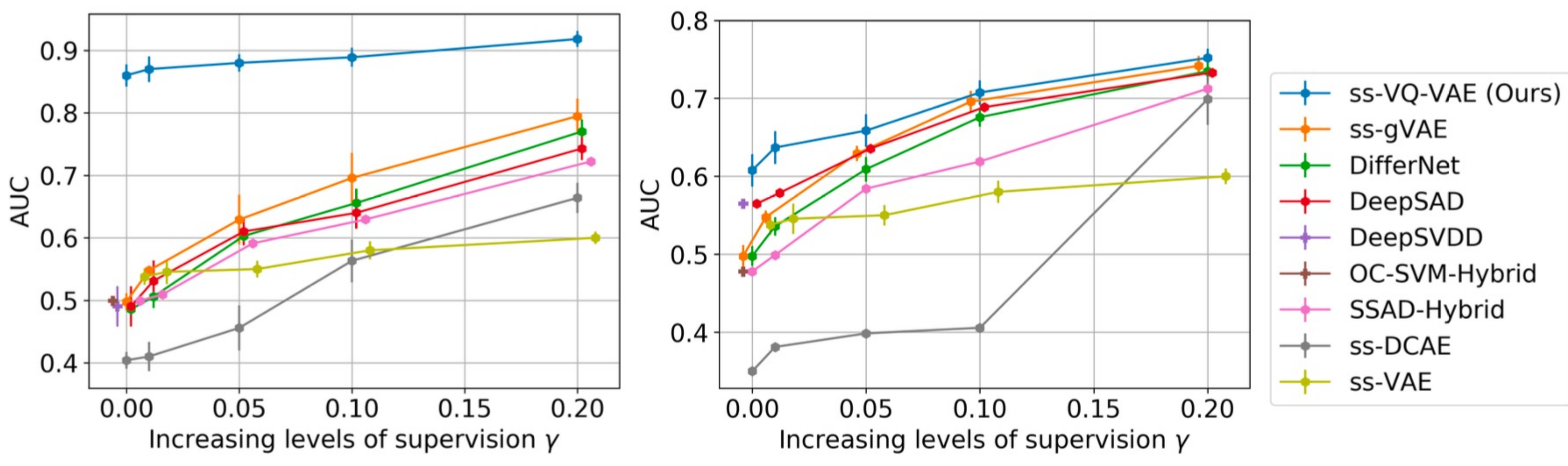


Fig: Semi-supervised VQ-VAE learning framework for one-class classification. In this example, we are using MVTec's bottle (object category) data.

Datasets used

- 15 datasets from MVTec Anomaly Detection [2] (consisting of texture and object categories, each has sub-categories of anomalies), taking patches of size 64x64.
- Corrosion dataset contains curated patches of size 512×512 from high-resolution industrial setup images. A patch is labelled anomalous if the corresponding anomaly mask contains at least 50% of anomalous pixels.



(a) MVTec's Carpet Dataset **(b) Corrosion Datsaset** **Baselines**

Fig: Quantitative Results on MVTec's Carpet and Corrosion dataset

References: [1] A Van Den Oord, O Vinyals, et al., "Neural Discrete Representation Learning," *Adv. in Neu. Inf. Proc. Sys.*, vol. 30, 2017.
[2] P Bergmann, M Fauser, D Sattlegger, and C Steger. MVTec AD–a comprehensive real-world dataset for unsupervised anomaly detection. In IEEE Comp. Vis. Pattern Recog., 2019.

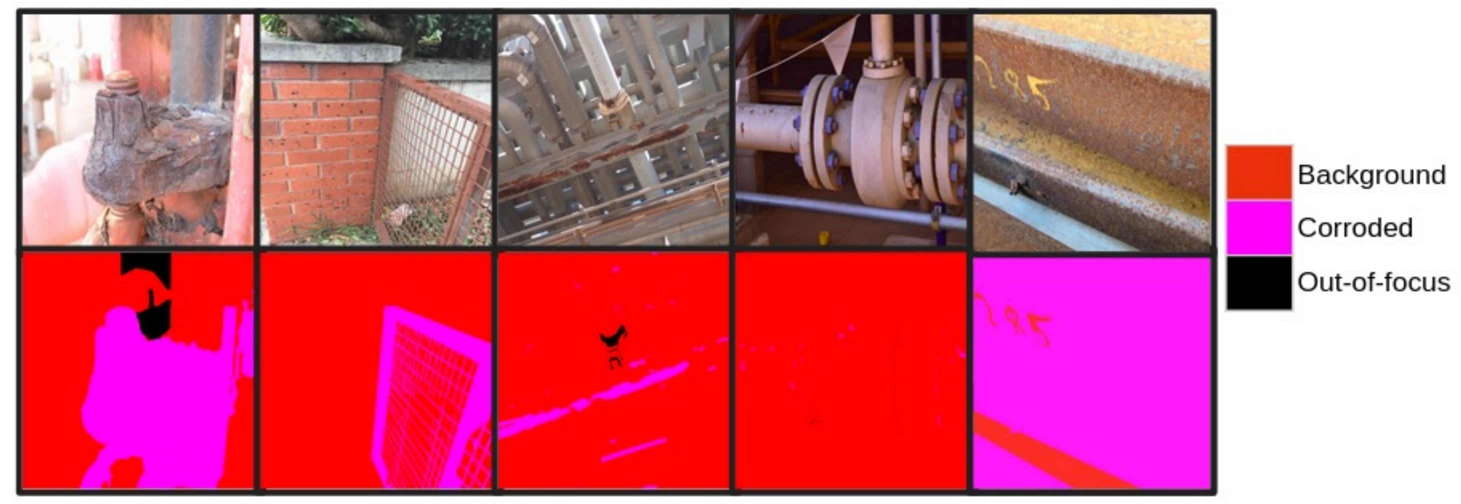


Fig: Corrosion Dataset

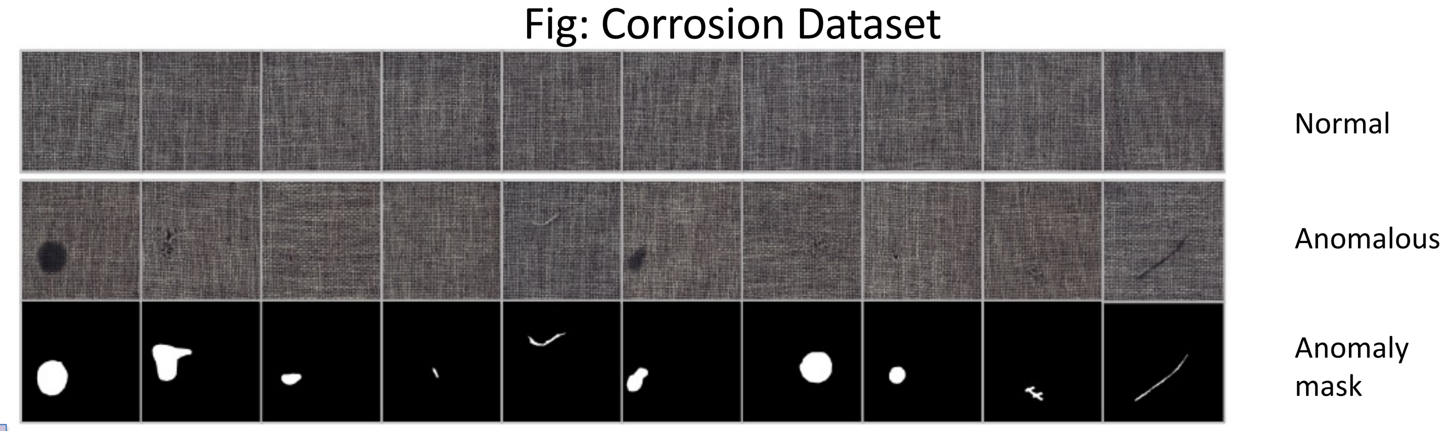


Fig: MVTec's Carpet Dataset

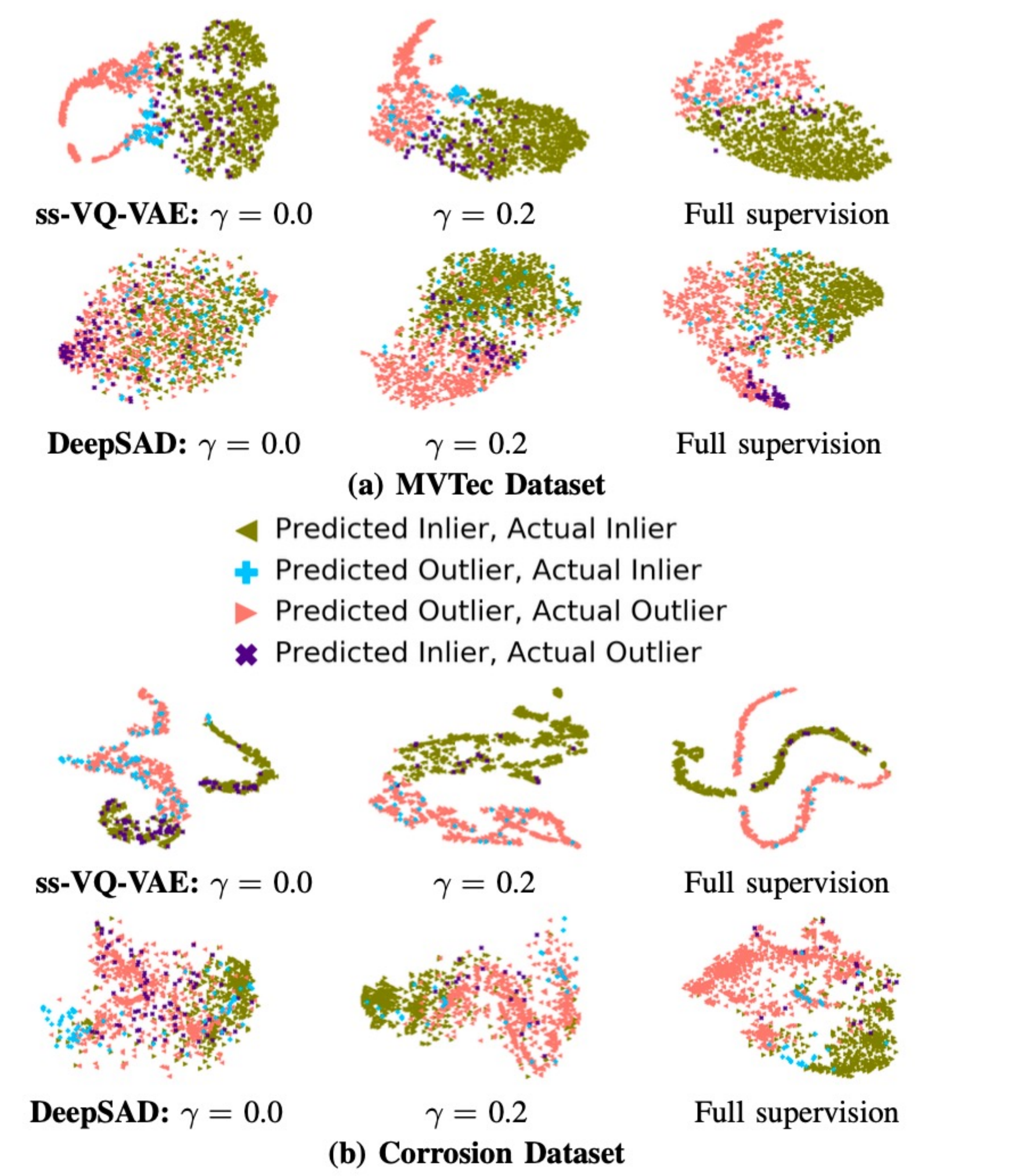


Fig: Qualitative Results on MVTec and Corrosion Datasets