Self-Supervised Anomaly Detection from Anomalous Training Data via Iterative Latent Token Masking

Ashay Patel, Petru-Daniel Tudosiu, Walter H.L. Pinaya, Mark S. Graham, Olusola Adeleke, Gary Cook, Vicky Goh, Sebastien Ourselin, M. Jorge Cardoso

King's College London

Introduction

- Unsupervised Anomaly detection methods predominantly require clean datasets without the presence of any anomalies
- Obtaining anomaly free datasets can be difficult in many practical scenarios, especially medical imaging
- Ethical issues for certain imaging modalities mean most scans are only performed given a strong prior suspicion of pathology
- By combining the use of Vector Quantised Variational Autoencoders (VQ-VAE)[1] and transformers we devise a mechanism for self-supervised anomaly detection when faced with anomalous training data
- Utilising the token masking capabilities of transformers we can iteratively mask out suspected anomalous areas to emulate training on normal data only
- We demonstrate the performance of our work on AutoPET[2] – a whole body PET dataset in addition to MVTec[3] – an industrial manufacturing dataset showcasing the efficacy of our approach on both 3D medical imaging and everyday 2D images

Iterative Token Masking for Anomalous Training Data

- Given a baseline anomaly detection approach we now seek to utilise the token masking capabilities of the transformer for anomaly detection when training on anomalous data
- The aim is to perform anomaly detection on the original training dataset to highlight suspected anomalies
- Given a binary anomaly map for an image we can downsample the image to that of the latent space dimension to represent which latent codes represent anomalous data
- These latent codes can then be masked within the transformer in the next series of training to remove the influence of said suspected anomalies
- This process can then be repeated until convergence is seen

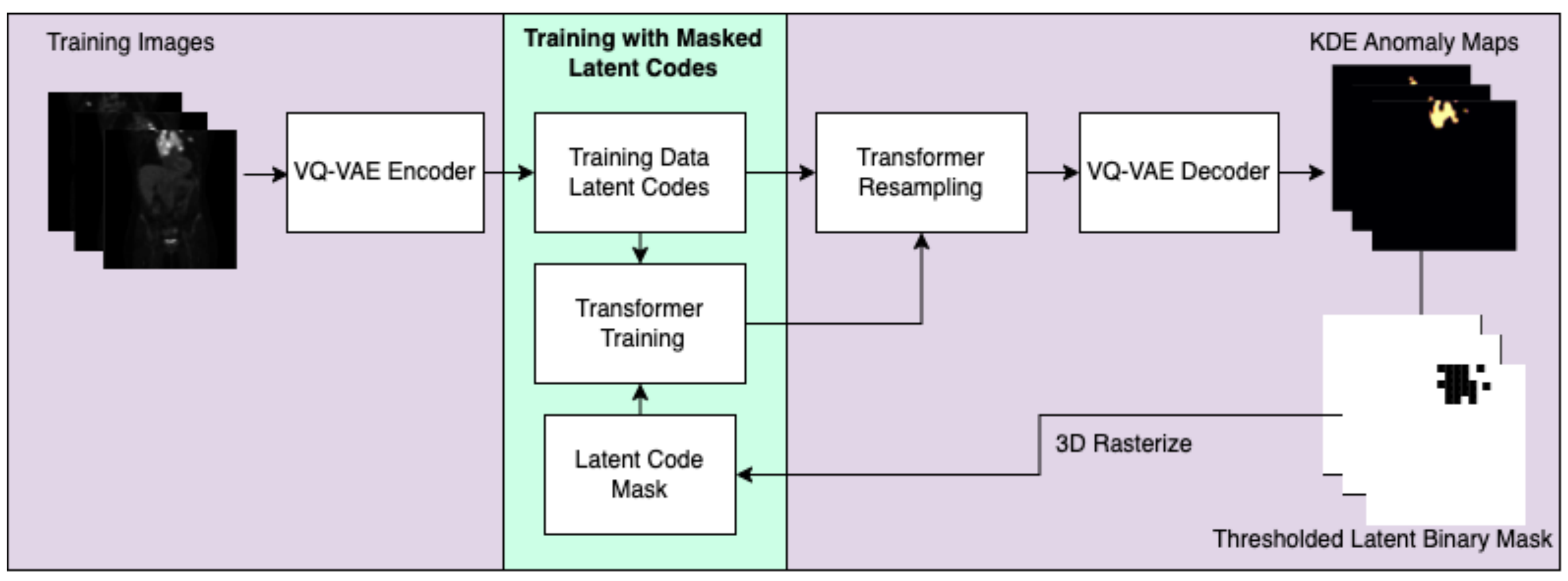

FIG 1: Iterative Training cycle with token masking. Images are encoded and used to train a transformer. The trained transformer is used to run anomaly detection on the training data's latent codes, that are then decoded to calculate KDE Anomaly Maps. These maps can then be used to generate a binary mask over the latent space to be used in the next round of transformer training. This process is then repeated until convergence in the anomaly maps is seen

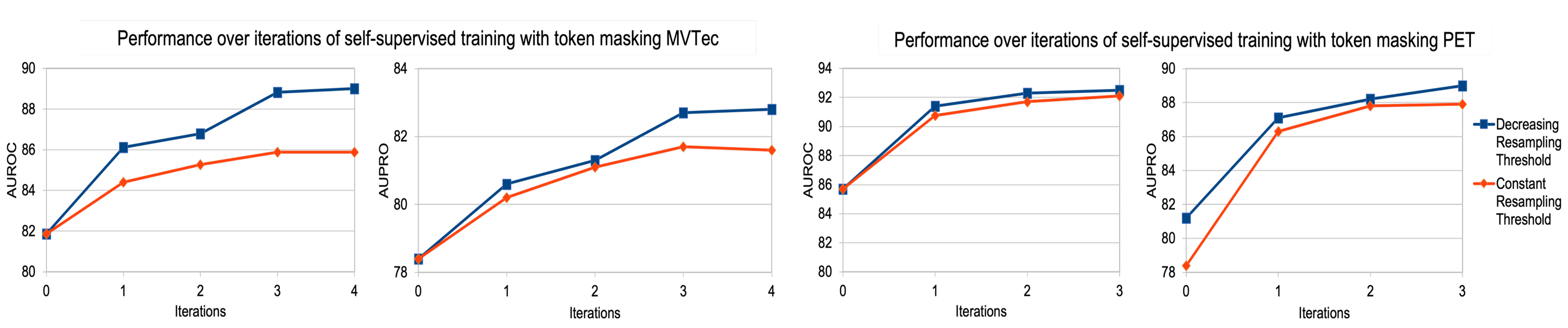

FIG 3: Plot of improvement in performance over training iterations for our proposed model when trained on 100% anomalous training data for PET and 80% Anomalous training data for MVTec

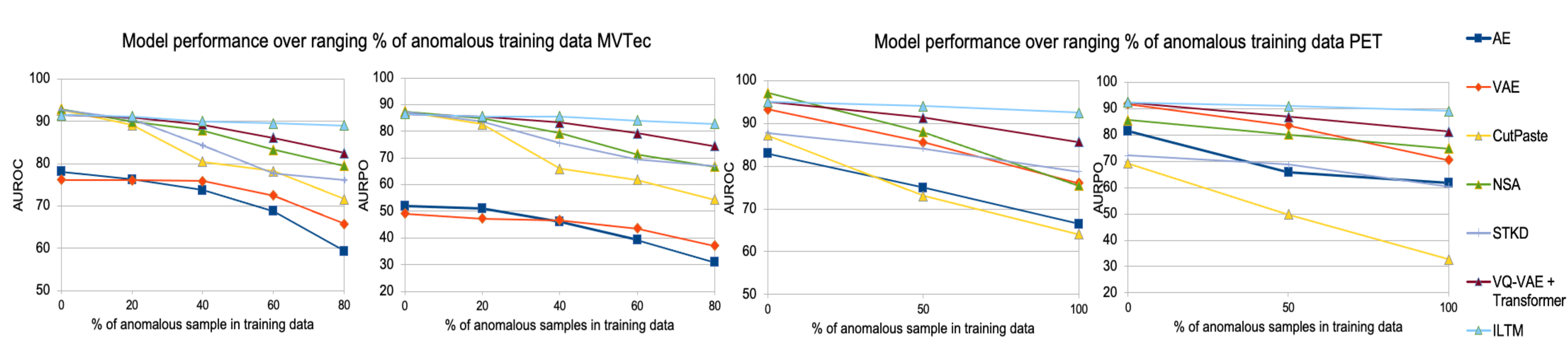

FIG 4: Comparison showing the variation of performance of our model against leading baselines with ranging number of samples containing anomalies

Models and Anomaly Detection

- The driving components behind the proposed method are the VQ-VAE and Transformer used to learn the probability density function of the encoded training images
- To use the Transformer model images must be presented in a 1D sequence ideally categorical
- Using the VQ-VAE we can encode images into a discrete latent space that can then be flattened and fed to the transformer for training and inference.
- Once trained the transformer can be used to highlight tokens in the discrete latent space that are low likelihood i.e. those that represent anomalous data
- By using the transformer to replace low likelihood/anomalous tokens with high likelihood tokens we can "heal" the anomalous latent space and then decode a new image with the suspected anomalies removed
- To improve on this following the work from Patel et al [4] we utilise the generative capabilities of the transformer by resampling low likelihood tokens multiple times, in addition to decoding the latent codes with dropout multiple times to realise a variety of healthy realisations of the original image. Through this instead of creating a residual anomaly map we can use Kernel Density Estimation (KDE) at each voxel position to model the intensity/uptake of specific voxels and then score the original intensities against the modelled distribution generating a KDE anomaly map.
- For medical imaging this is especially useful as often there are bi-modal or multi-modal distributions present for healthy representations

Results

- Experiments were set up creating training datasets with ranging anomalous contamination between 0-100% for the 3D PET data and 0–80% contamination on MVTec

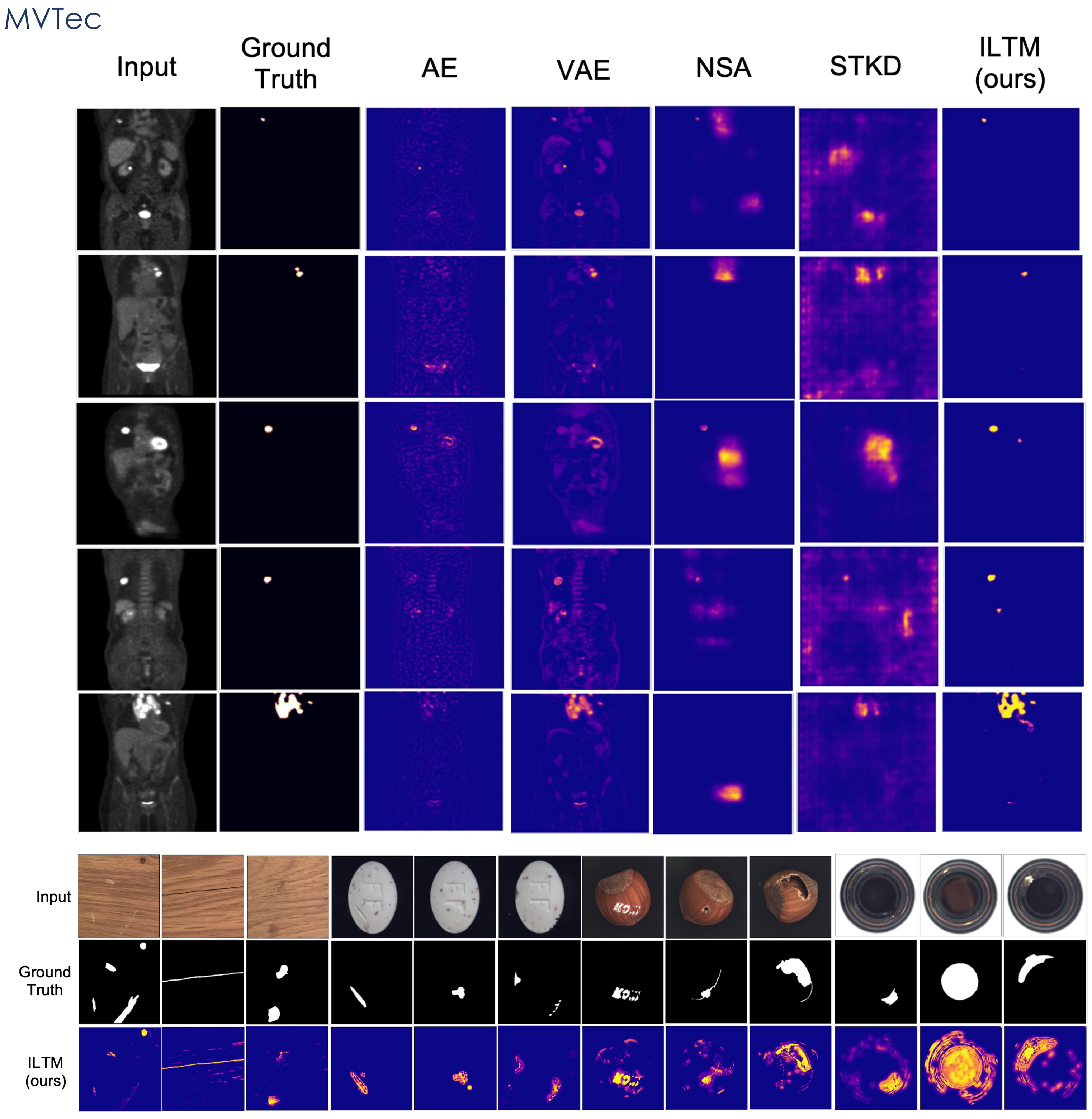

FIG 2: Results for models trained on 100% anomalous training data for PET and 80% Anomalous training data for MVTec

Conclusion

- In medical imaging scenarios it can often be difficult to obtain datasets free of anomalies, as such the use of unsupervised methods are often limited
- In this work we present an iterative method of training in a self-supervised way without any labelled data in a scenario where non-anomalous data is unavailable
- Not only does our model show little degradation in performance as the levels of anomalies in training increases, but it also far outperforms the competing baselines that show large drops in performance

References

[1] – Van Den Oord, A. and Vinyals, O., 2017. Neural discrete representation learning. Advances in neural information processing systems, 30.

[2] – Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C. and Rubin, D., 2022. A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. Scientific Data, 9(1), p.601.

[3] – Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C., 2019. MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9592-9600).

[4] - Patel, A., Tudosiu, P.D., Pinaya, W.H.L., Cook, G., Goh, V., Ourselin, S. and Cardoso, M.J., 2022, September. Cross attention transformers for multi-modal unsupervised whole-body PET anomaly detection. In MICCAI Workshop on Deep Generative Models (pp. 14-23). Cham: Springer Nature Switzerland.