Unsupervised Brain Anomaly Detection and Segmentation with Transformers

Walter Hugo Lopez Pinaya[1], Petru-Daniel Tudosiu[1], Robert Gray[2], Geraint Rees[3], Parashkev Nachev[2], Sébastien Ourselin[1], M. Jorge Cardoso[1]

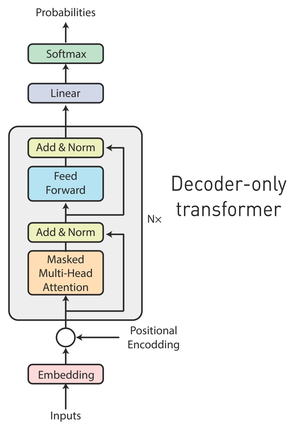
[1] Department of Biomedical Engineering, School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
[2] Institute of Neurology, University College London, London, UK
[3] Institute of Cognitive Neuroscience, University College London, London, UK

Introduction

Pathological brain appearances may be so heterogeneous as to be intelligible only as anomalies, defined by their deviation from normality rather than any specific pathological characteristic.

Amongst the hardest tasks in medical imaging, detecting such anomalies requires models of the normal brain that combine compactness with the expressivity of the complex, long-range interactions that characterise its structural organisation.

These are requirements transformers (Vaswani et al., 2017) have arguably greater potential to satisfy than other current candidate architectures.

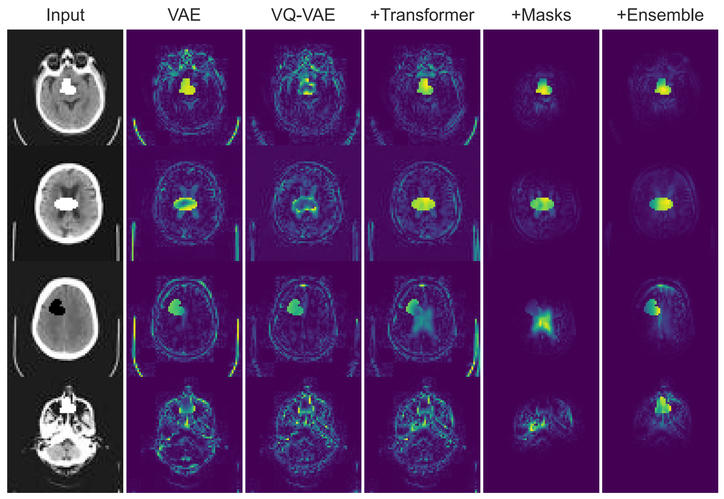


Anomaly Segmentation on Synthetic Data I

We trained our models on 8,000 Head CT images from the MedNIST dataset.

Then, we contaminated 100 images from the test set with sprites (i.e., synthetic anomalies).

Finally, we applied the several stages of our anomaly segmentation method and compared our results against state-of-the-art autoencoder models based on the architectures proposed in the (Baur et al., 2020a)



Anomaly Segmentation on Synthetic Data II

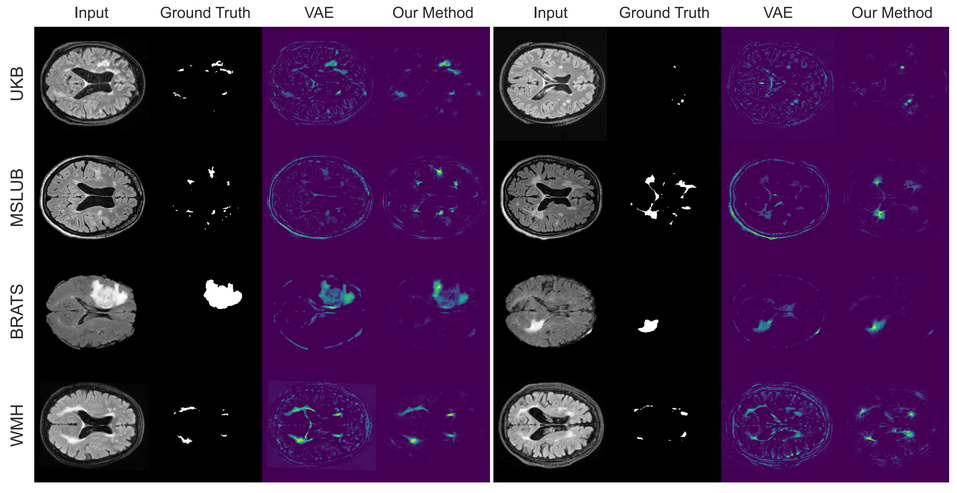
Method	$\lceil DICE \rceil$	AUPRC
AE (Dense) (Baur et al., 2020a)	0.213	0.129
AE (Spatial) (Baur et al., 2020a)	0.165	0.093
VAE (Dense) (Baur et al., 2020a)	0.533	0.464
VQ-VAE (reconstruction-based; van den Oord et al., 2017)	0.457	0.346
VQ-VAE + Transformer (Ours)	0.675	0.738
VQ-VAE + Transformer + Masked Residuals (Ours)	0.768	0.808
VQ-VAE + Transformer + Masked Residuals + Different Orderings (Ours)	**0.895**	**0.956**

The use of the transformer to "heal" the latent representation, the spatial information in the resampling mask, and the ensemble with different orderings, improved the segmentation performance by a large margin.

Anomaly Segmentation on Real Neuroimaging Data I

We used the FLAIR images (224x224 pixels) from the UK Biobank (UKB) dataset to evaluate our method's performance on real-world lesion data.

We evaluated our method generalisability on four different datasets, containing multiple sclerosis lesions, white matter hyperintensities.



Anomaly Segmentation on Real Neuroimaging Data II

UKB Dataset	*DICE*	AUPRC
AE (Dense) (Baur et al., 2020a)	0.016	0.005
AE (Spatial) (Baur et al., 2020a)	0.054	0.015
VAE (Dense) (Baur et al., 2020a)	0.016	0.006
VQ-VAE (reconstruction-based; van den Oord et al., 2017)	0.028	0.005
VQ-VAE + Transformer + Masked Residuals + Different Orderings (Ours)	**0.232**	**0.159**
MSLUB Dataset		
AE (Dense) (Baur et al., 2020a)	0.041	0.016
AE (Spatial) (Baur et al., 2020a)	0.061	0.026
VAE (Dense) (Baur et al., 2020a)	0.039	0.016
VQ-VAE (reconstruction-based; van den Oord et al., 2017)	0.040	0.016
VQ-VAE + Transformer + Masked Residuals + Different Orderings (Ours)	**0.378**	**0.272**
BRATS Dataset		
AE (Dense) (Baur et al., 2020a)	0.276	0.094
AE (Spatial) (Baur et al., 2020a)	0.531	0.215
VAE (Dense) (Baur et al., 2020a)	0.294	0.107
VQ-VAE (reconstruction-based; van den Oord et al., 2017)	0.331	0.125
VQ-VAE + Transformer + Masked Residuals + Different Orderings (Ours)	**0.759**	**0.555**
WMH Dataset		
AE (Dense) (Baur et al., 2020a)	0.073	0.024
AE (Spatial) (Baur et al., 2020a)	0.150	0.054
VAE (Dense) (Baur et al., 2020a)	0.068	0.022
VQ-VAE (reconstruction-based; van den Oord et al., 2017)	0.100	0.036
VQ-VAE + Transformer + Masked Residuals + Different Orderings (Ours)	**0.429**	**0.320**

Our method showed a better performance than the autoencoder approaches from the literature in all datasets.

Image-wise Anomaly Detection on Synthetic Data

We also evaluated our method to detect anomalous (out-of-distribution - OOD). We used 1,000 images from the HeadCT class as the in-distribution test set, the 100 HeadCT images contaminated by sprite anomalies as the near OOD set, and 1.000 images of each other MedNIST classes as the far OOD set images.

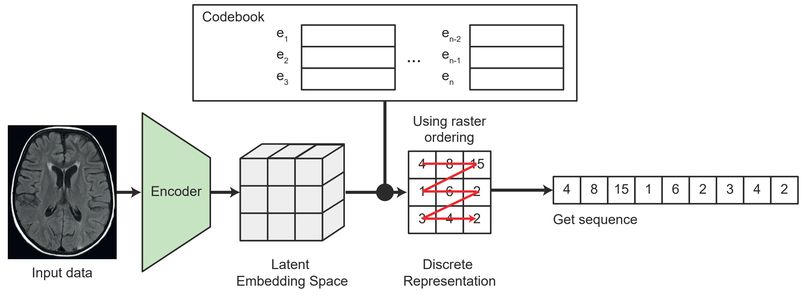
	AUCROC	**AUPRC In**	**AUPRC Out**	**FPR95**	**FPR99**	**FPR999**
vs. far OOD classes						
VAE (Dense) (Baur et al., 2020a)	0.298	0.855	0.060	0.986	0.996	0.996
Our approach	**1.000**	**1.000**	**1.000**	**0.000**	0.001	0.004
Our approach with general purpose VQ-VAE	**1.000**	**1.000**	**1.000**	**0.000**	**0.000**	**0.000**
vs. near OOD classes						
VAE (Dense) (Baur et al., 2020a)	0.111	0.094	0.672	1.000	1.000	1.000
Our approach	0.921	0.988	0.707	**0.409**	0.885	0.885
Our approach with general purpose VQ-VAE	**0.932**	**0.990**	**0.721**	0.482	**0.882**	**0.882**

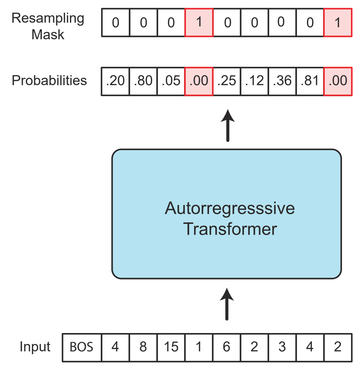
Our method obtained high scores when classifying an image as an in distribution image or an out-of-distribution image.

Method: Anomaly Segmentation

The core of our anomaly detector is an expressive transformer (Choromanski et al., 2020) that learns the probability density function of 2D images of healthy brains. This requires us to express each image's contents as a sequence of observations on which transformers-like models can operate. Instead of learning the distributions on individual pixels directly, we use the compact latent discrete representation of a VQ-VAE (van den Oord et al., 2017).

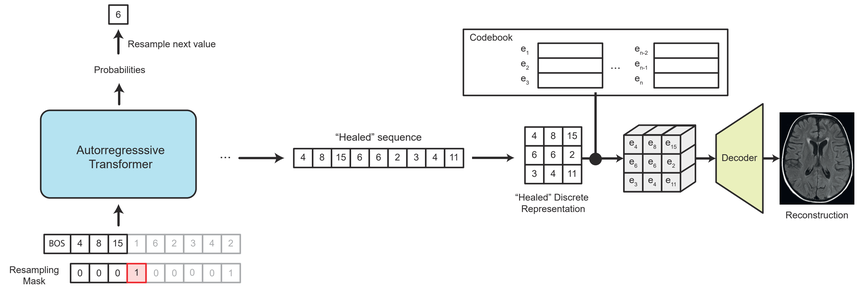
To segment an anomaly in an image, first, we obtain the 2D latent discrete representation from the VQ-VAE model. Next, we reshape the discrete representation into a 1D sequence.





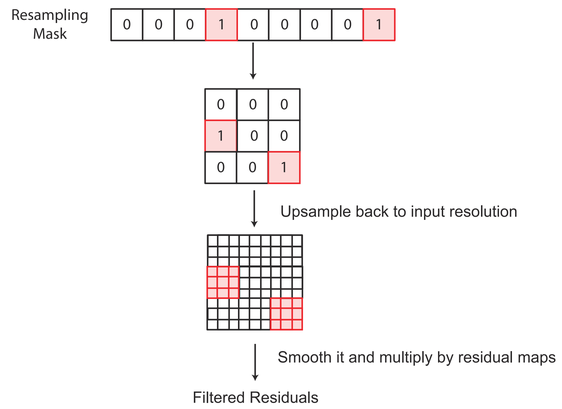
We use the autoregressive transformer to obtain the likelihood of each latent variable value. Using a threshold, we can select indices with the lowest likelihood values and create a "resampling mask" indicating which variables are abnormal.

Then, we replace these abnormal values with values sampled by the transformer. This approach attenuates the influence of the anomalies by replacing them with values that conform to the healthy distribution. This in-painted latent representation is then reconstructed to a brain image, and we obtain the pixel-wise residuals from the difference of the reconstruction and the inputted image.



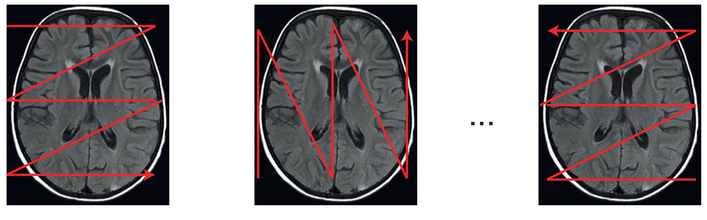
Method: Spatial information from the latent space

Autoencoders are known for creating blurry reconstructions. To avoid areas being mislabelled because of it, we used the spatial information present in the "resampling mask" since this mask indicates the spatial location of the latent values with anomalies according to the transformer model.



Method: Multiple views of the latent space through reordering

Using the same VQ-VAE model, we trained an ensemble of autoregressive transformers. However, each one of our transformers uses a different reordering of the 2D latent image to create a sequence. This compels each transformer to use a different context of the latent image when predicting the likelihood of an element.



Conclusions

Automatically determining the presence of lesion and delineating their boundaries is essential to the introduction complex models of rich neuroimaging features in clinical care

Novel transformer-based approach which achieves superior results in all tested tasks when compared to competing methods

Future studies
New network designs, model conditioning, and explore the performance in other medical data

References

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017.

Christoph Baur, Stefan Denner, Benedikt Wiestler, Nassir Navab, and Shadi Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain mr images: A comparative study. Medical Image Analysis, page 101952, 2020a

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. CoRR, abs/1711.00937, 2017. URL http://arxiv.org/abs/1711.00937.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, et al. Rethinking attention with performers. arXiv preprint arXiv:2009.14794, 2020

Acknowledgments

WHLP and MJC are supported by Wellcome Innovations [WT213038/Z/18/Z]. PTD is supported by the EPSRC Research Council, part of the EPSRC DTP, grant Ref: [EP/R513064/1]. PN is supported by Wellcome Innovations [WT213038/Z/18/Z] and the UCLH NIHR Biomedical Research Centre. This research has been conducted using the UK Biobank Resource (Project number: 58292).

wellcome trust

AMiGO Artificial Medical Intelligence Group

https://amigos.ai

email: walter.diaz_sanz@kcl.ac.uk
twitter: @Warvito