Learning to Identify Diseases From Healthy Medical Images

MUDANO
WASTE LESS. DO MORE.

Sue Liu, Boris Mitrovic

Mudano Ltd., UK.

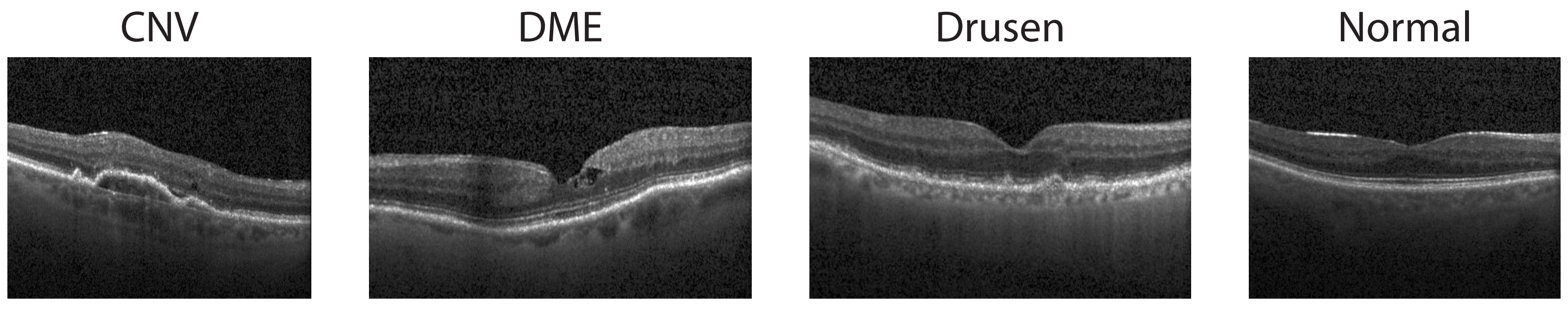
Correspondence to: **sue.liu@mudano.com**

Introduction

One major unsolved problem in the field of machine learning is the ability to learning efficiently from few data points, as humans are able to do. Generative models, a branch of unsupervised learning, try to overcome this limitation by learning meaningful features of the input while requiring little or no human supervision or labelling. In the case of disease detection from medical images, by only observing healthy data a generative model will able to identify anomalies for multiple diseases or even rare diseases. This can save time and money in data collection and labelling required for supervised learning methods.

Problem

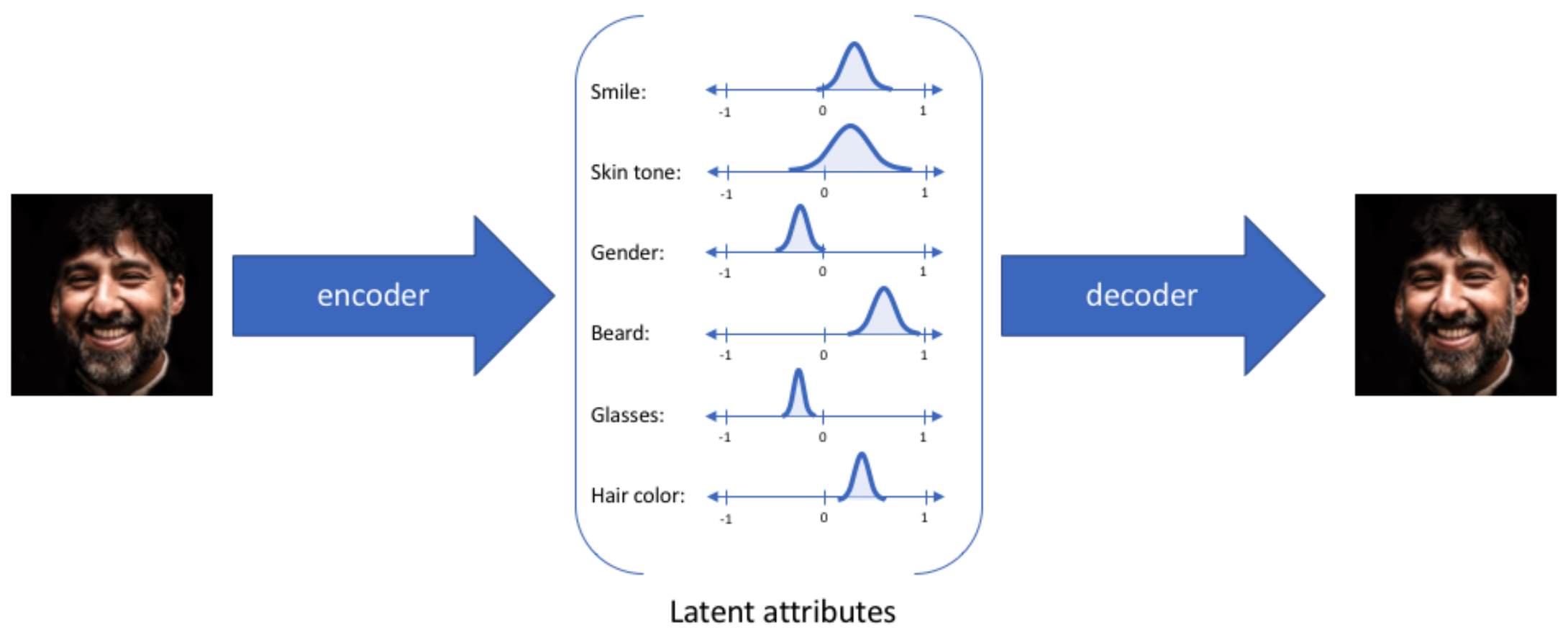
In this study we investigate the problem of unsupervised anomaly detection for retinal diseases. We used an optical coherence tomography OCT dataset [1] containing images of three types of retinal diseases and a set of healthy images.

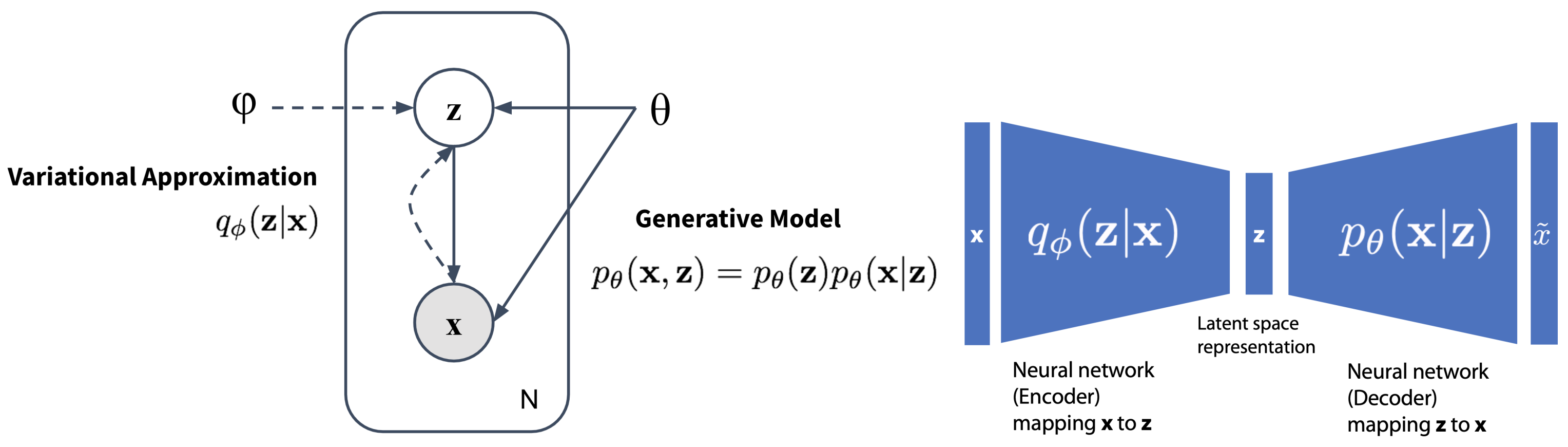


From left to right: Choroidal neovascularization (CNV), diabetic macular edema (DME), drusen and normal images. CNV and DME left untreated can result in total blindness, and once diagnosed are immediately referred for treatment. A total of 25236 normal images were used in training and 1329 for validation. The test set consists of 500 images, of which 300 are diseased images and 200 are normal images.

Model

Performing unsupervised learning over images is challenging due to the curse of dimensionality, but the recent development of deep generative models have addressed this issue. One such model is the Variational Autoencoder (VAE) [2, 3].




φ
θ
z
x
N
Variational Approximation
$q_\phi(\mathbf{z}|\mathbf{x})$
Generative Model
$p_\theta(\mathbf{x},\mathbf{z}) = p_\theta(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{z})$
x
$q_\phi(\mathbf{z}|\mathbf{x})$
z
$p_\theta(\mathbf{x}|\mathbf{z})$
$\tilde{x}$
Latent space representation
Neural network (Encoder) mapping **x** to **z**
Neural network (Decoder) mapping **z** to **x**

VAE can be viewed as a directed probabilistic graphical model that is constructed into a neural network architecture. The encoder is trained to learn a probabilitic distribution for each latent variable $\mathbf{z}$ from the input $\mathbf{x}$, and the decoder samples the latent distribution $\mathbf{z}$ and learn the mapping $p(\mathbf{x}|\mathbf{z})$ to reconstruct $\mathbf{x}$.

The learning is performed using variational inference by approximating the conditional distribution $p(\mathbf{z}|\mathbf{x})$ (usually intractable) that describes the mapping from the input to the latent variables by a variational approximation $q(\mathbf{z}|\mathbf{x})$ (defined to have a tractable distribution, usually Gaussian), and maximising the evidence lower bound (ELBO):

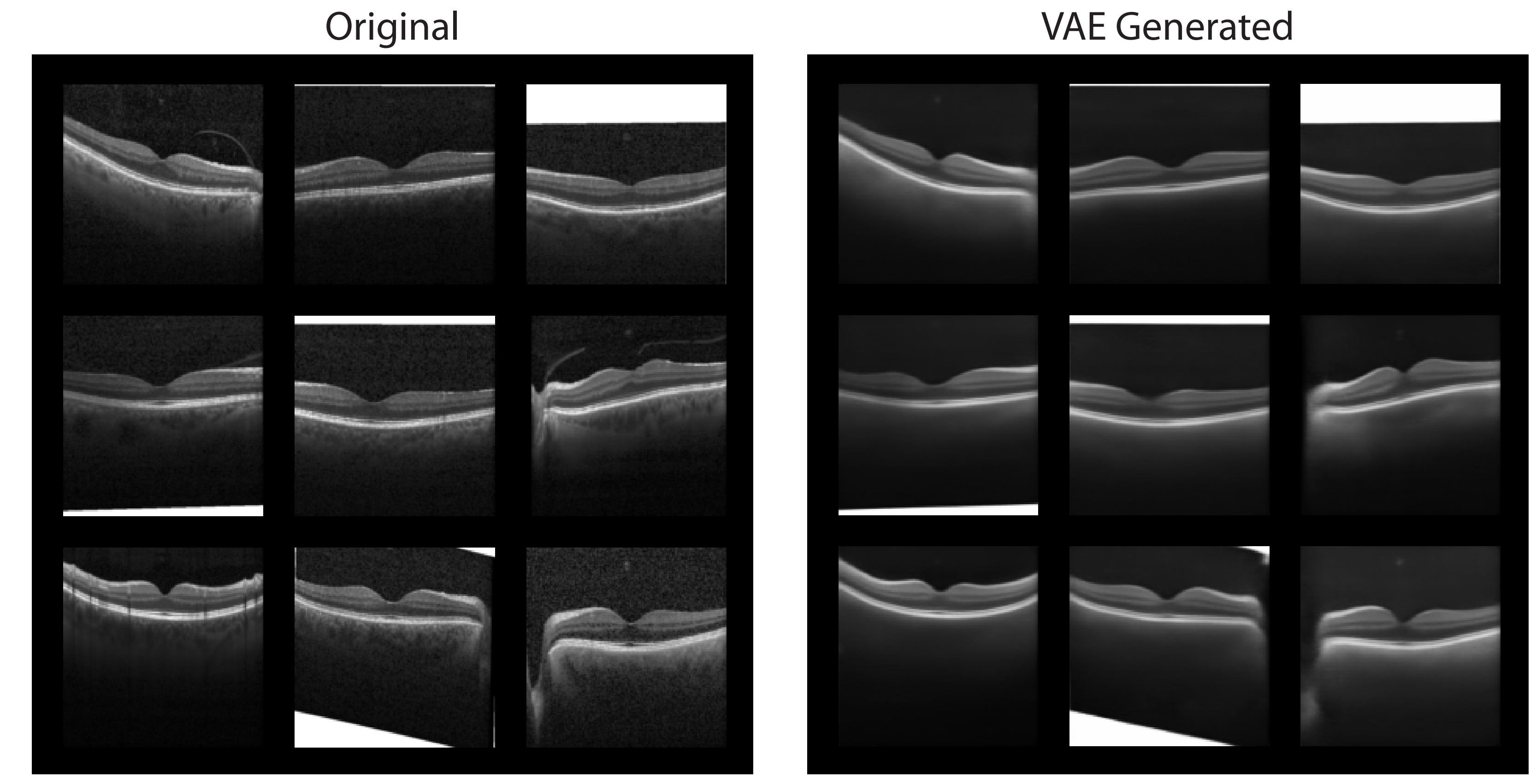
$$\log p(\mathbf{x}) \geq E_{q(\mathbf{z}|\mathbf{x})} \log p(\mathbf{x}|\mathbf{z}) - KL(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

The first term represents the reconstruction likelihood and the second term ensures that our learned distribution q is similar to the true prior distribution p.

Method

We use a fully convolutional architecture similar to that of DCGAN [4]. In the encoder, instead of using a linear layer to produce mean and log variance we use two separate convolutional layers. In this set up the images are normalized to have size 128 x 128, so the encoder consists of 5 convolutional blocks and the decoder consists of 5 deconvolutional blocks. The number of latent dimensions is 300, and the model was trained for 40 epochs using a learning rate of 10^{-4} and ADAM optimizer.

Results



Above: 9 randomly selected reconstructions on the validation set. *left:* original images, *right:* VAE reconstructed images. Note that while reconstruction using VAE has captured the significant features in the images, it has smoothed out the edges. This is typical of VAE and pixel based loss function.

The degree of anomaly can be characterised by the possibility of seeing $\mathbf{x}$ appear under the distribution $p(\mathbf{x})$, hence computing the anomaly score is essentially estimating $s(\mathbf{x}) = -\log p(\mathbf{x})$, i.e. the negative of the ELBO loss:

$$s_{\mathrm{VAE}} = KL(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) - \frac{1}{L}\sum_{i=1}^{L} \log p(\mathbf{x}|\mathbf{z}_i), \quad \mathbf{z}_i \sim q(\mathbf{z}|\mathbf{x})$$

The first term is the KL-divergence loss and the second the reconstruction loss.

Method used to compute AUC	CNV	DME	Drusen	All
Reconstruction Score	0.93	0.90	0.77	0.87
KL Score	0.56	0.51	0.48	0.52
VAE Score	0.92	0.89	0.76	0.86

Table: AUC values for detecting the three different types of retinal diseases separately and for all diseases. The values are computed using three separate anomaly scores based on reconstruction loss, KL-divergence loss and ELBO loss respectively. AUC values obtained using only KL-divergence loss are low, suggesting our prior is not expressive enough to approximate the true latent space distribution.

Summary

In this work we have demonstrated the potential of applying Variational Autoencoder (VAE) for anomaly detection in retinal disease images. When trained on only the normal data, the model is able to perform efficient inference and to determine whether a test image is diseased or not. Using the optical coherence tomography (OCT) images from [1], the model is able to detect all diseases with AUC of 0.87, and is able to detect age related macular degeneration and diabetic macular edema, which are diseases requiring immediate referral with AUC of 0.93 and 0.9 respectively.

VAEs are appealing because they have a solid and elegant theoretical framework, are built on top of standard function approximators (neural networks) and can be easily trained using stochastic gradient descent. In this work we have not fully explored the powers of VAE for anomaly detection: we could perform postprocessing on the anomaly scores and/or make use of the latent space (e.g. create distance metrics).

We argue that although supervised deep learning models have been successfully applied for retinal disease diagnosis [1], applying deep unsupervised learning in disease identification is a promising research direction when the normal images are in abundance, whilst labelled diseases are rare.

References

[1] D. S. Kermany *et al.* Identifying medical diagnoses and treatable diseases by image-based deep learning *Cell.*, Vol.172, 2018.

[2] D. P. Kingma and Welling, M. Auto-encoding variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations*, 2013.

[3] D. J. Rezende *et al.* Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-14)*, 2014.

[4] A. Radford *et al.* Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proceedings of the 5th International Conference on Learning Representations*, 2016.

Acknowledgements

The authors thank A. Grivas, J. Langr and E. Wielewski for helpful discussions.