Dietrichstein, Marc; Major, David; Trapp, Martin; Wimmer, Maria; Lenis, Dimitrios; Winter, Philip; Berg, Astrid; Neubauer, Theresa; Bühler, Katja

Anomaly Detection Using Generative Models and Sum-Product Networks in Mammography Scans

Published in:
Deep Generative Models - 2nd MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022, Proceedings

DOI:
10.1007/978-3-031-18576-2_8

Published: 01/01/2022

Document Version
Peer reviewed version

Please cite the original version:
Dietrichstein, M., Major, D., Trapp, M., Wimmer, M., Lenis, D., Winter, P., Berg, A., Neubauer, T., & Bühler, K. (2022). Anomaly Detection Using Generative Models and Sum-Product Networks in Mammography Scans. In A. Mukhopadhyay, I. Oksuz, S. Engelhardt, D. Zhu, & Y. Yuan (Eds.), *Deep Generative Models - 2nd MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022, Proceedings* (pp. 77-86). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 13609 LNCS). Springer. https://doi.org/10.1007/978-3-031-18576-2_8

Anomaly Detection Using Generative Models and Sum-Product Networks in Mammography Scans

Marc Dietrichstein[1⋆], David Major[1⋆], Martin Trapp[2], Maria Wimmer[1], Dimitrios Lenis[1], Philip Winter[1], Astrid Berg[1], Theresa Neubauer[1], and Katja Bühler[1]

[1] VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH, Vienna, Austria
david.major@vrvis.at
[2] Department of Computer Science, Aalto University, Espoo, Finland **

Abstract. Unsupervised anomaly detection models that are trained solely by healthy data, have gained importance in recent years, as the annotation of medical data is a tedious task. Autoencoders and generative adversarial networks are the standard anomaly detection methods that are utilized to learn the data distribution. However, they fall short when it comes to inference and evaluation of the likelihood of test samples. We propose a novel combination of generative models and a probabilistic graphical model. After encoding image samples by autoencoders, the distribution of data is modeled by Random and Tensorized Sum-Product Networks ensuring exact and efficient inference at test time. We evaluate different autoencoder architectures in combination with Random and Tensorized Sum-Product Networks on mammography images using patch-wise processing and observe superior performance over utilizing the models standalone and state-of-the-art in anomaly detection for medical data.

Keywords: Anomaly Detection · Generative Models · Sum-Product Networks · Mammography.

1 Introduction

Acceleration of the detection and segmentation of anomalous tissue by automated computer-aided approaches is a key to enhancing cancer screening programs. It is especially important for mammography screening, as breast cancer

⋆ Equal contribution

** This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-031-18576-2_8. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

is the most common cancer type and the leading cause of death in women worldwide [20]. Training an artificial neural network in a supervised way requires a high amount of pixel-wise annotated data. As data annotation is very costly, methods that involve as less annotation as possible are in high demand. Anomaly detection approaches are good representatives of this type, as they only utilize healthy cases for learning, and anomalous spots are detected as a deviation from the learned data distribution. The deviation is measured either by straightforward metrics such as reconstruction error of input and output samples or by more sophisticated constructs such as log-likelihood in probabilistic models.

Unsupervised anomaly detection methods have been evaluated on a plethora of different pathologies and medical imaging modalities. A state-of-the-art method in this area is f-AnoGAN [15], which leverages Generative Adversarial Networks (GANs) to model an implicit distribution of healthy images and detect outliers via a custom anomaly score based on reconstruction performance. f-AnoGAN has been utilized to detect anomalies in Optical Coherence Tomography (OCT) scans [15], Chest X-rays [1], and 3D Brain scans [17]. However, it requires the training of a separate encoder module to obtain latent codes of images, which are used by the generator for reconstruction. The autoencoder (AE) architecture, on the other hand, jointly trains an encoder and decoder and is thus able to directly map an input to its corresponding latent representation. AE variants have been applied to lesion detection in mammography images [19] and brain scans [8,21], as well as head [14] and abdomen [8] Computed Tomography scans. However, the practical applicability of all those models is limited by the fact that the respective anomaly scores are not easily interpretable by a human decision maker. Here, to remedy the situation, it would be desirable for the model to provide some degree of certainty for its decision. To this end, density estimation models can be employed. Such models learn an explicit probability density function from the training data and assume that anomalous samples are located within low-density regions. Examples are the application of Gaussian Mixture Models [2] for brain lesion detection as well as Bayesian U-Nets for OCT anomaly detection [16]. Although these approaches are similar to ours, they are tailored to specific image modalities and can thus not be directly applied to our domain.

In this work, we introduce a novel and general method for anomaly detection that combines AEs with probabilistic graphical models called Sum-Product Networks (SPNs). A recent powerful SPN architecture called Random and Tensorized SPN (RAT-SPN) [12] was chosen, as it is easy to integrate into deep learning frameworks and is trained by GPU-based optimization. More than that, standard and variational AEs do not allow to derive exact data likelihoods, they rather provide approximations that can be used for anomaly detection. SPNs solve this problem and allow *exact* and *efficient* likelihood inference by imposing special structural constraints on the model capturing the data distribution. We compare the performance of different standalone AEs to that of their combination with RAT-SPNs on unsupervised mass and calcification detection in public mammography scans and demonstrate improvements.

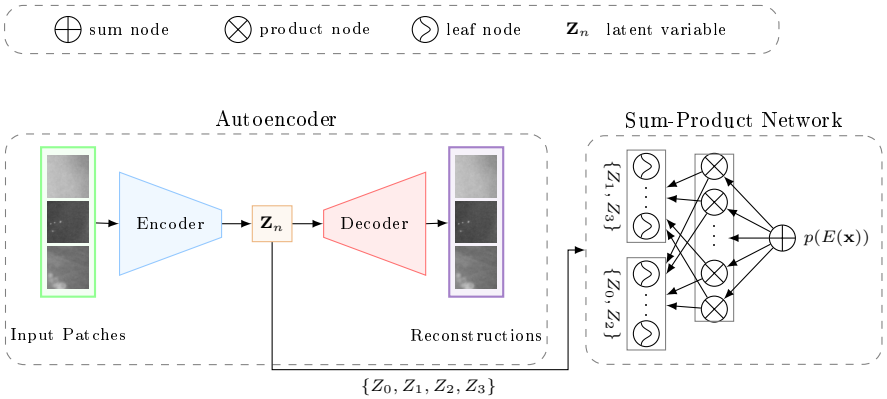


Fig. 1: The encoder of an AE outputs a low-dimensional latent representation $\mathbf{z}_n$ of healthy mammography patches. This representation serves then as input to a SPN that learns the corresponding probability distribution $p(E(\mathbf{x}))$. The likelihood of test samples is predicted over the same pipeline using trained models.

2 Methods

Our approach learns the healthy data distribution in a patch-wise fashion. First, the dimensionality of patch data is reduced by an AE, and the likelihood for membership to the data distribution is approximated by a RAT-SPN. During inference, the learned model is applied to test images patch by patch, where at every position, the likelihood yields the anomaly score. As the models capture the distribution of healthy data, this score should be significantly different at anomalous image positions. We compare the performance of our pipeline to that of standalone AE models. The different AEs, that we considered, are described in Section 2.1 and RAT-SPNs in Section 2.2. It is followed by our proposed combination of an AE with a RAT-SPN in Section 2.3. A system overview is provided in Fig. 1.

2.1 Autoencoders

Convolutional AEs (CAEs) [9] utilize convolutional blocks to map high dimensional image data $\mathbf{x} \in \mathbb{R}^{H \times W}$ into a lower dimensional latent space $\mathbf{z} \in \mathbb{R}^{M}$ using an encoder by $\mathbf{z} = E(\mathbf{x})$ and reconstruct it utilizing a decoder model by $\hat{\mathbf{x}} = D(E(\mathbf{x}))$. The compression and reconstruction process is learned by minimizing the reconstruction loss $\mathcal{L}_{\text{CAE}} = \ell_2(\mathbf{x}, \hat{\mathbf{x}})$ where ℓ_2 signalizes the mean squared error (MSE). Computation of $\mathcal{L}_{\text{CAE}}$ for test samples yields the anomaly score at inference.

Variational Autoencoders (VAEs) [6] are equipped with the same building blocks as CAEs when applied to images, but additionally, they aim to approximate the true posterior distribution $p(\mathbf{z}|\mathbf{x})$ in the encoder E by a simpler and more

tractable distribution $q(\mathbf{z}|\mathbf{x})$. This is achieved by minimizing the KL divergence $D_{KL}(q(\mathbf{z}|\mathbf{x})\,||\,p(\mathbf{z}|\mathbf{x}))$ between the two distributions. On the other hand, the decoder D learns the posterior $p(\mathbf{x}|\mathbf{z})$ and reconstructs $\mathbf{x}$ from a given $\mathbf{z}$ by maximizing the log-likelihood $\log p(\mathbf{x}|\mathbf{z})$. The overall objective to minimize is called the evidence lower bound (ELBO), and it can be formulated as follows:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log p(\mathbf{x}|\mathbf{z})\right] - \beta D_{KL}(q(\mathbf{z}|\mathbf{x})\,||\,p(\mathbf{z})). \tag{1}$$

The MSE was utilized as the reconstruction loss for $\log p(\mathbf{x}|\mathbf{z})$, and β was set to 0.1 following [5], which is a weighting factor between the two terms. Thus, our models are called βVAEs [5]. $\mathcal{L}_{VAE}$ is utilized as the anomaly score for a given test sample during inference.

Vector Quantised-Variational Autoencoder (VQVAE) [10] is a VAE variant that differs from the original in a crucial aspect: it uses discrete instead of continuous variables to represent the latent space. Discretization is realized by mapping the encoder output $E(\mathbf{x})$ to the index of the closest vector e_i in the latent embedding space $\mathbf{e} \in \mathbb{R}^{K\times B}$, where K is the number of distinct discrete values and B is the dimension of each embedding vector e_i. The posterior variational distribution $q(\mathbf{z}|\mathbf{x})$ is one-hot-encoded in such a way that $q(\mathbf{z}=k|\mathbf{x})=1$, with $k = \arg\min_i ||E(\mathbf{x}) - e_i||_2$. The mapping of $E(\mathbf{x})$ to the nearest embedding vector e_i is defined as $E_q(\mathbf{x}) = e_k$ with $k = \arg\min_i ||E(\mathbf{x}) - e_i||_2$. The loss formulation consists of three parts, each aiming to optimize a different aspect of the model:

$$\mathcal{L}_{\text{VQVAE}} = \log p(\mathbf{x}|E_q(\mathbf{x})) + ||sg[E(\mathbf{x})] - \mathbf{e}||_2^2 + \lambda||E(\mathbf{x}) - sg[\mathbf{e}]||_2^2. \tag{2}$$

The first term is the reconstruction loss, for which MSE was again chosen. The remaining terms are concerned with learning an optimal embedding space. The codebook loss, the second term, attempts to move the embedding vectors closer to the encoder output, whereas the third term, the commitment loss, attempts the inverse and forces the encoder output to be closer to the closest embedding vector $\mathbf{e}$. $sg[\cdot]$ is the stop-gradient operator and prevents its operand from being updated during back-propagation. λ is a weighting factor for the commitment loss, which we set to 0.25, following [10]. The anomaly score for a given test sample is determined by calculating its reconstruction loss.

2.2 Sum-Product Networks

SPNs [13] are tractable probabilistic models of the family of probabilistic circuits [3] and allow various probabilistic queries to be computed efficiently and exactly. For consistency with recent works, we will introduce SPNs based on the formalism in [18]. An SPN on a set of random variables $\mathbf{Z} = \{Z_j\}_{j=1}^{J}$ is a tuple $(\mathcal{G}, \psi)$ consisting of a computational graph $\mathcal{G}$, which is a directed acyclic graph, and a scope function ψ mapping from the set of nodes in $\mathcal{G}$ to the set of all subsets of $\mathbf{Z}$ including $\mathbf{Z}$. The computational graph of an SPN is typically composed of sum nodes, product nodes, and leaf nodes. Sum nodes compute a weighted sum of

their children, i.e., $(\mathrm{S}(\mathbf{z}) = \sum_{\mathrm{N}\in\mathrm{ch(S)}} \theta_{\mathrm{S,N}}\,\mathrm{N}(\mathbf{z}))$, product nodes compute a product of their children, i.e., $(\mathrm{P}(\mathbf{z}) = \prod_{\mathrm{N}\in\mathrm{ch(P)}} \mathrm{N}(\mathbf{z}))$, and leaf nodes are tractable multivariate or univariate probability distributions or indicator functions. The scope function assigns each node a scope (subset of $\mathbf{Z}$ or $\mathbf{Z}$) and ensures that the SPN fulfills certain structural properties, guaranteeing that specific probabilistic queries can be answered tractably. In this work, we will focus on SPNs that are *smooth* and *decomposable*, we refer to [3] for a detailed discussion. Moreover, we consider a representation of the SPN in the form of a random and tensorized region graph called RAT-SPNs [12] and employ the implementation based on Einstein summation as proposed in [11]. The region graph is parametrized by the number of root nodes C, input distributions I as well as the graph depth D, and the number of parallel SPN instances, or recursive splits, R. By choosing these parameters, RAT-SPNs with arbitrary complexity may be constructed. From a given region graph, it is possible to obtain the underlying SPN structure in terms of its computational graph and scope function exactly, for more details see [18,12]. A simple region graph with $C = 1$, $I = 2$, $D = 1$, and $R = 1$, and the underlying SPN is illustrated in Fig. 1. In a generative learning setting like ours, the optimal network parameters w are found by applying (stochastic) Expectation Maximization (EM) to maximize the log-likelihood LL of the training samples:

$$LL(w) = \frac{1}{N}\sum_{n=1}^{N} \log \mathrm{S}(\mathbf{z}_n). \tag{3}$$

2.3 Combining Autoencoders and Sum-Product Networks

We combine each AE type of Section 2.1 with a RAT-SPN by passing the learned latent representation of encoded samples $\mathbf{z} = E(\mathbf{x})$ as observed states for the random variables $\mathbf{Z}$ to a RAT-SPN (see Fig. 1). This way, after input images are mapped to a low-dimensional space, likelihoods can be obtained exactly and efficiently in an end-to-end fashion at inference. Extra computations, essential for reconstruction and ELBO-based scores of standalone AEs, are therefore not necessary. Two RAT-SPN setups are utilized, one with Gaussian input distributions for the continuous latent representations of CAEs and βVAEs, and the other with categorical inputs for the discrete features of VQVAEs. Training is done separately, first the AE models are trained followed by RAT-SPNs. The anomaly score is yielded by the likelihood of a trained AE and RAT-SPN combination for a given test sample.

3 Experimental Setup

3.1 Datasets

We train our models on the Digital Database for Screening Mammography (DDSM) [4], a collection of 2620 mammography exams, with each exam consisting of multiple images. The images in this dataset are categorized according

to the type of diagnosis, either into *healthy* or into a *cancer* type (i.e., malignant, benign). As we want to learn a healthy model, we selected only the 695 healthy exams containing 2798 images for training purposes. From each image, 120 patches of 64×64 pixels (px), the resolution also used by [15], were extracted; half of these containing internal breast tissue, the other half were sampled along the breast contour. We evaluated all methods against a selection of cancerous mammograms from the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [7], which provides improved annotations of masses and calcifications for images from DDSM. In order to filter out images with large-scale annotations, we set the restriction that the annotation mask area must be smaller than 4-times of our patch area for masses, and it should contain the whole calcifications. 79 scans were selected from the mass and 30 scans from the calcification test set that fulfilled these criteria. The healthy training images consisted of equally distributed dense and non-dense tissues, whereas the mass test cases had a ratio of 14%/86% and the calcification test samples a ratio of 40%/60% (dense/non-dense).

3.2 Training

The 2798 healthy images were split into 90% training and 10% validation images (with no patient overlap) for training both the AE (CAE, βVAE, VQVAE) and the RAT-SPN models. Following [9], all of our AE models had an architecture with 32-64-128 2D convolutional layers with 5×5 kernels and a stride of 2 in the encoder and 2D transposed convolutional layers in the decoder. The VQVAE model had additional 6 residual blocks with 128 filters, and the dimensionality of the embedding vector was 64. All models had 64 latent units and were trained with a batch size of 64. CAEs and βVAEs were trained for 100 epochs with a learning rate of 1e-5, whereas VQVAEs converged to an optimum after 20 epochs with a learning rate of 1e-4. The best-fit RAT-SPN parameters of $C = 1$, $I = 45$, $D = 1$, $R = 50$ were found utilizing the 10% validation images, possible values were taken from the supplement of [12]. The RAT-SPN setup was the same for all AE models, and it was trained by the EM algorithm for 50 epochs with a batch size of 64 and a learning rate of 1e-4.

3.3 Evaluation

We evaluated our methods on the 79 mass and 30 calcification test images. The anomaly score assignment was performed in a lower dimensional image space than the original resolution, and thus, patches were sampled around every 16^{th} pixel per image. Only breast tissue pixels were considered using pre-segmentations of the breast area in every image. In order to show the anomaly scores' discriminative power between healthy and anomalous positions, we derived the Area Under the ROC Curve (AUC) in two ways, either considering all pixels from all test images at once (pixel-wise) or doing it for each image separately and calculating the average over all test samples additionally (image-wise). In order to measure the capability of the methods for detection of the

anomalous regions, we apply the Hausdorff distance (H) image-wise to assess the pixel distance between masks generated by our models and the provided CBIS-DDSM ground-truth. It measures the maximum of the distances from any annotated point in one mask to the nearest point in the other mask, thus, the smaller it is, the closer the match between prediction and ground-truth.

Test Data	*Model*	**Pixel-wise**	**Image-wise**	
		AUC	*AUC*	*H*
Masses	CAE	0.53	0.58±0.25	30.80±12.57
	CAE-RATSPN	**0.88**	**0.88±0.10**	**30.10±14.50**
	βVAE	0.80	0.83±0.14	30.78±12.79
	βVAE-RATSPN	0.88	0.88±0.11	29.07±14.57
	VQVAE	0.67	0.67±0.19	33.37±12.53
	VQVAE-RATSPN	**0.82**	**0.84±0.13**	**32.21±14.01**
	f-AnoGAN	0.86	0.85±0.12	30.41±11.91
Calcifications	CAE	0.65	0.77±0.21	32.79±10.82
	CAE-RATSPN	0.72	0.78±0.16	33.03±13.74
	βVAE	0.73	0.80±0.17	29.73±12.56
	βVAE-RATSPN	0.66	0.73±0.17	30.72±13.07
	VQVAE	0.69	0.79±0.17	31.79±12.66
	VQVAE-RATSPN	0.68	0.75±0.19	33.34±10.61
	f-AnoGAN	0.67	0.74±0.20	34.95±6.98

Table 1: Anomaly detection results utilizing different models. Metrics are computed either over pixels or images. Next to AUC scores average Hausdorff (H) distances (px) between anomaly segmentations and ground-truth masks were computed. Segmentations are calculated after score thresholding by 99^{th}-percentile. Statistically significantly better performance (based on image-wise AUCs) between standalone and RAT-SPN extended models are depicted in bold ($p<0.01$).

4 Results and Conclusion

We compare the anomaly detection performance of the three AE models in their standalone configuration as well as with a RAT-SPN extension. Additionally, we trained and evaluated a state-of-the-art f-AnoGAN model in its default configuration. The results are illustrated in Table 1 and Fig. 2.

For the mass test set, the overall best performing model was the βVAE-RATSPN with 0.88 pixel-wise and average image-wise AUCs, and an average H-distance of 29.07 px (see Fig. 2). Statistically significant superior image-wise AUC performances over standalone models were achieved by CAE-RATSPNs and VQVAE-RATSPNs. Except for VQVAE-RATSPN, all RAT-SPN extended models performed better than f-AnoGAN in terms of image-wise AUC, although there were no statistically significant differences (cf. Table 1). It is also visible

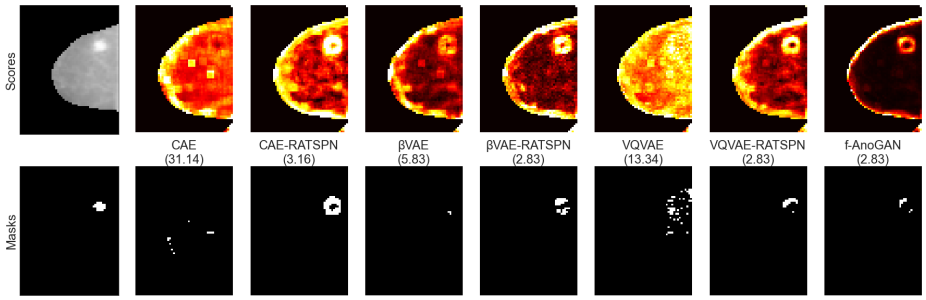
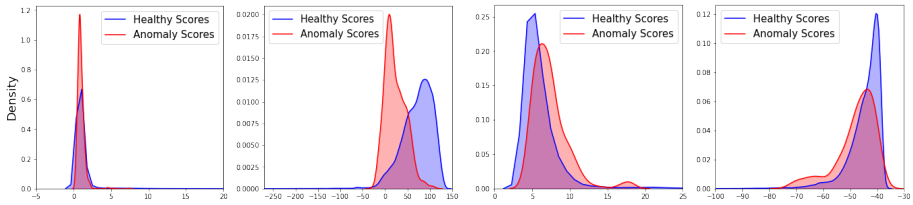


Fig. 2: Anomaly detection results of a mass sample. The first column shows the mammography scan (top) and anomaly ground-truth (bottom). The remaining columns depict an anomaly score heatmap (top) and a segmentation mask (bottom) for each model. Bright yellow pixels represent high and dark red pixels low anomaly scores in the heatmaps. The respective Hausdorff distance to the ground-truth (px) is displayed after each method name in brackets. Segmentations are calculated after score thresholding by 99^{th}-percentile.

in Table 1 that RAT-SPNs applied to continuous features yielded better results than the discrete version. Furthermore, it is depicted in Fig. 3 a) and b), that attaching RAT-SPN models to AE models facilitate a better discrimination between healthy and anomalous tissue by increasing the gap between their respective distributions.

Moreover, the standalone βVAE was the best performing model for the calcification test set with an 0.73 pixel-wise and 0.80 average image-wise AUC, and an average H-distance of 29.73 px. It is in general visible that all models reflect a consistently poorer performance for this data. This is due to the fact that this set contains a higher proportion of dense breasts than the mass collection (see Section 3.1), and most of the small calcifications were generally hard to detect accurately by all models in images dominated by dense tissue. On the other hand, the standalone versions performed here better than the ones with RAT-SPN extension except for the CAE setup, but no statistically significant differences were discovered based on the image-wise AUC scores (cf. Table 1). This behavior is well visualized by the score distribution plots of the best-performing standalone βVAE and βVAE-RATSPN versions in Fig. 3 c) and d). All models except for βVAE-RATSPN yielded better image-wise AUCs than f-AnoGAN, although no statistically significant differences were detected (cf. Table 1).

In summary, we have introduced a novel unsupervised anomaly detection method that extends various AE architectures with a RAT-SPN module. This approach is a promising avenue for generating exact likelihoods and incorporating them into the detection of different anomalies, such as masses and calcifications in mammography scans. Our experiments suggest that our method clearly outperforms standalone AE models on mass samples. Furthermore, it exhibits similar results to those of the state-of-the-art f-AnoGAN, however, with



(a) (b) (c) (d)

Fig. 3: Distribution of healthy and anomaly scores on the masses (a,b) and calcifications datasets (c,d) for CAE without (a) and with RAT-SPN extension (b), for βVAE without (c) and with RAT-SPN extension (d).

the advantages of a comparatively simpler training setup and exact likelihood inference. All of the investigated methods have difficulties when applied to calcification samples. We interpret that this is due to the presence of a larger proportion of dense tissue in the latter dataset. In future work, we plan to analyze how this problem can be eliminated and in particular whether increasing the input resolution has a positive effect on the performance.

Acknowledgements VRVis is funded by BMK, BMDW, Styria, SFG, Tyrol and Vienna Business Agency in the scope of COMET - Competence Centers for Excellent Technologies (879730) which is managed by FFG. Thanks go to AGFA HealthCare, project partner of VRVis, for providing valuable input. Martin Trapp acknowledges funding from the Academy of Finland (347279).

References

1. Bhatt, N., Prados, D.R., Hodzic, N., Karanassios, C., Tizhoosh, H.R.: Unsupervised Detection of Lung Nodules in Chest Radiography Using Generative Adversarial Networks. In: Proceedings of EMBC. pp. 3842–3845. IEEE (2021)
2. Bowles, C., Qin, C., Guerrero, R., Gunn, R., Hammers, A., Dickie, D.A., Hernández, M.V., Wardlaw, J., Rueckert, D.: Brain Lesion Segmentation through Image Synthesis and Outlier Detection. NeuroImage: Clinical **16**, 643–658 (2017)
3. Choi, Y., Vergari, A., Van den Broeck, G.: Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Models. Tech. rep., UCLA (2020)
4. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W.P.: The Digital Database for Screening Mammography. In: Proceedings of the International Workshop on Digital Mammography. pp. 212–218. Medical Physics Publishing (2000)
5. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: Beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. ICLR (2017)
6. Kingma, D.P., Welling, M.: Auto-encoding Variational Bayes. arXiv preprint arXiv:1312.6114 (2013)
7. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A Curated Mammography Data Set for Use in Computer-aided Detection and Diagnosis Research. Scientific data **4**(1), 1–9 (2017)

8. Marimont, S.N., Tarroni, G.: Anomaly Detection through Latent Space Restoration using Vector Quantized Variational Autoencoders. In: Proceedings of ISBI. pp. 1764–1767. IEEE (2021)
9. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked Convolutional Auto-encoders for Hierarchical Feature Extraction. In: Proceedings of ICANN. pp. 52–59. Springer (2011)
10. Van den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural Discrete Representation Learning. In: Proceedings of NIPS. pp. 6309–6318 (2017)
11. Peharz, R., Lang, S., Vergari, A., Stelzner, K., Molina, A., Trapp, M., Van den Broeck, G., Kersting, K., Ghahramani, Z.: Einsum Networks: Fast and Scalable Learning of Tractable Probabilistic Circuits. In: Proceedings of ICML. pp. 7563–7574. PMLR (2020)
12. Peharz, R., Vergari, A., Stelzner, K., Molina, A., Shao, X., Trapp, M., Kersting, K., Ghahramani, Z.: Random Sum-Product Networks: A Simple and Effective Approach to Probabilistic Deep Learning. In: Uncertainty in Artificial Intelligence. pp. 334–344. PMLR (2020)
13. Poon, H., Domingos, P.: Sum-Product Networks: A New Deep Architecture. In: Proceedings of ICCV Workshops. pp. 689–690. IEEE (2011)
14. Sato, D., Hanaoka, S., Nomura, Y., Takenaga, T., Miki, S., Yoshikawa, T., Hayashi, N., Abe, O.: A Primitive Study on Unsupervised Anomaly Detection with an Autoencoder in Emergency Head CT Volumes. In: Medical Imaging 2018: Computer-Aided Diagnosis. vol. 10575, p. 105751P. International Society for Optics and Photonics (2018)
15. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. Medical image analysis **54**, 30–44 (2019)
16. Seeböck, P., Orlando, J.I., Schlegl, T., Waldstein, S.M., Bogunović, H., Klimscha, S., Langs, G., Schmidt-Erfurth, U.: Exploiting Epistemic Uncertainty of Anatomy Segmentation for Anomaly Detection in Retinal OCT. IEEE Transactions on Medical Imaging **39**(1), 87–98 (2019)
17. Simarro Viana, J., de la Rosa, E., Vande Vyvere, T., Robben, D., Sima, D.M., et al.: Unsupervised 3D Brain Anomaly Detection. In: Proceedings of the MICCAI Brain Lesion Workshop. pp. 133–142. Springer (2020)
18. Trapp, M., Peharz, R., Ge, H., Pernkopf, F., Ghahramani, Z.: Bayesian Learning of Sum-Product Networks. In: Proceedings of NeurIPS. pp. 6347–6358 (2019)
19. Wei, Q., Ren, Y., Hou, R., Shi, B., Lo, J.Y., Carin, L.: Anomaly Detection for Medical Images based on a One-class Classification. In: Medical Imaging 2018: Computer-Aided Diagnosis. vol. 10575, pp. 375–380. SPIE (2018)
20. Wild, C., Weiderpass, E., Stewart, B.W.: World Cancer Report: Cancer Research for Cancer Prevention. IARC Press (2020)
21. Zimmerer, D., Isensee, F., Petersen, J., Kohl, S., Maier-Hein, K.: Unsupervised Anomaly Localization using Variational Auto-encoders. In: Proceedings of MICCAI. pp. 289–297. Springer (2019)