Unsupervised Lesion Detection with Locally Gaussian Approximation

Xiaoran Chen[1], Nick Pawlowski[2] Ben Glocker[2], and Ender Konukoglu[1]

[1] Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland
chenx@vision.ee.ethz.ch
[2] Biomedical Image Analysis Group, Imperial College London, London, UK

Abstract. Generative models have recently been applied to unsupervised lesion detection, where a distribution of normal data, i.e. the normative distribution, is learned and lesions are detected as out-of-distribution regions. However, directly calculating the probability for the lesion region using the normative distribution is intractable. In this work, we address this issue by approximating the normative distribution with local Gaussian approximation and evaluating the probability of test samples in an iterative manner. We show that the local Gaussian approximator can be applied to several auto-encoding models to perform image restoration and unsupervised lesion detection. The proposed method is evaluated on the BraTS dataset, where the proposed method shows improved detection and achieves state-of-the-art results.

1 Introduction

Automated lesion detection has been an active topic in medical imaging research. Conventionally, lesions are detected by visual inspection based on intensity characteristics in medical scans, such as Computed Tomography (CT) and Magnetic Resonance Images (MRI).

Algorithmic approaches have emerged as viable automatic alternatives to visual inspection, where lesion detection is often formulated as a classification or segmentation problem. Early approaches are based on supervised classification [1, 2]. More recently, impressive performance has been achieved by deep learning-based methods, in supervised [3] and weakly supervised [4] settings. Despite the success, methods with supervision require laborious image acquisition for annotated labels. On the other hand, those methods are lesion-specific as they learn a mapping between images and labels and may be limited in applications such as pre-screening for a large range of abnormalities. In contrast, unsupervised methods are not specific to lesions in the training data and enable critical applications such as automatic screening for radiological assessment.

The principle behind unsupervised detection is to learn a normative distribution of images from healthy individuals [5]. Lesions can then be detected as out-of-distribution areas in the images. Compared to supervised methods, unsupervised detection is arguably more difficult as the model cannot encode any lesion specific characteristics. Earlier attempts model the normative distribution either by assuming the expected tissue composition or via atlas registration [6–8]. More recent methods use deep learning models to estimate the normative distributions, for example, via Generative Adversarial Nets (GANs) [9] and Variational Auto-Encoders (VAEs) [10]. AnoGAN [5] uses GANs to model the normative distribution. Given an image with abnormalities, it estimates a "pseudo-normal" version of the image by determining the closest image the GAN can generate. Abnormal pixels are then detected using the absolute differences between the original and generated pseudo-normal images. VAE-based methods take a similar approach where pseudo-normal images are estimated by reconstructing the image with an encoder-decoder model trained only on healthy images. Lesions are then detected as regions with high reconstruction errors [11–13]. An advantage of VAEs is the explicit estimation of the normative distribution.

However, the generation of pseudo-normal images in previous works relies on mapping the image to the latent space and back, assuming the healthy regions would not change during this operation. This assumption may not hold as the mapping to the latent space may drift away from the ideal point due to the abnormality, leading to false positive detection caused by reconstruction errors in healthy regions. We also seek the "pseudo-normal" images to detect the lesions, but with a probabilistic formulation. To find the pseudo-healthy images more accurately, we use the image with lesion as an observation and the normative distribution as its prior distribution. The corresponding "pseudo-normal" image is then obtained by maximizing the probability of the observation in the normative distribution. However the normative distribution estimated by models such as VAE cannot be explicitly accessed, we propose a locally Gaussian approximation method to perform likelihood maximization with the normative distribution.

Likelihood maximization is performed with the local Gaussian approximator accessing the prior distribution by approximating its gradients. The most related work to ours is You et al. [14], where they perform Maximum-a-Posteriori to obtain "pseudo-normal" images and the normative distribution is approximated with the Evidence Lower Bound (ELBO) of GMVAE, which can be inaccurate and cannot be applied to models that do not optimize the ELBO. Unlike [14], our approximation constructs local Gaussian approximations to the prior for each gradient ascent step rather than taking derivatives of the ELBO. We investigate two variations for constructing the proposed approximator. We evaluate the detection performance on BraTS dataset and achieve state-of-the-art performance.

2 Methods

Define a latent space model with $z \in R^d$ and $X \in R^{m \times n}$. The latent space model can estimate $P(X) = \int P(X|z)P(z)dz$ with mapping $z = f(X)$ and $X = g(z)$. Direct computation of $P(X)$ is often very difficult. Depending on the purpose, such direct computation is not always required. Specifically, for the propose of maximum likelihood estimation (MLE), one calculates the gradients of $P(X)$ instead of $P(X)$ itself. We describe a local Gaussian approximation method to provide gradients for MLE with an indirect $P(X)$ as a prior distribution.

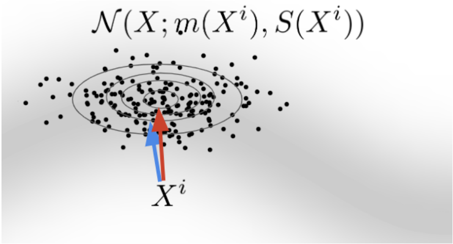

Fig. 1: Gaussian approximation illustration where grey shaded region represents the normative distribution for healthy images modelled by a VAE. The red arrow represents the derivative of the $(\log P(X))_{X^i}$, which is intractable to compute, while the blue arrow represents gradient computed using the proposed approach that locally approximates the image prior with a Gaussian close to X^i and uses its derivative.

2.1 Local Gaussian Approximation for Likelihood Estimation

To compute the gradients from $P(X)$, we assume that $P(X)$ locally follows a Gaussian distribution, $P_{local}(X) \sim \mathcal{N}(X; \mu_{local}, \Sigma_{local})$, where estimate μ_{local} and Σ_{local} can be easily estimated using the latent space model and MLE can be performed in an iterative manner until convergence.

For likelihood maximization, we start with the observation X^0 and perform gradient ascent with gradients computed from the approximated local Gaussian parameterized by $\mu_{local,X|z}$ and $\Sigma_{local,X|z}$ at each iteration i to obtain X^i (Eq.1).

$$\begin{aligned}(\nabla \log P(X))_{X^i} \approx& \left(\nabla \log \mathcal{N}\left(X; \mu_{local}(X^i), \Sigma_{local}(X^i)\right)\right)_{X^i}, \\ X^{i+1} =& X^i + \eta \nabla \log P(X))_{X^i}\end{aligned} \tag{1}$$

where $\mu_{local}(X^i)$ and $\Sigma_{local}(X^i)$ are the mean and covariance matrix of the approximation, respectively. The principal behind this approximation is pictorially depicted in Fig. 1. The role of $(\nabla \log P(X))_{X^i}$ is to direct the gradient ascent towards the image prior, which is depicted as the gray cloud in the figure with darker areas illustrating higher probability regions in the prior. At X^i, instead of taking the derivative of $\log P(X)$, which is intractable, the approach approximates the local region of the prior close to X^i with a local Gaussian distribution and uses its derivative, which is much easier to compute.

Assuming the mean and covariance are obtained, the local Gaussian approximation produces gradients w.r.t to X as $(\nabla \log P(X))_{X_i} \approx -\Sigma_{local}(X^i)^{-1}(X^i - \mu_{local}(X^i))$. We observe in this equation that while the gradient direction tries to move X towards $\mu_{local}(X^i)$, $\Sigma_{local}(X^i)$ provides additional knowledge on the expected variation in pixel intensities of healthy images at that local region of the prior. Next, we provide technical details on the estimation of $\mu_{local}(X^i)$ and $\Sigma_{local}(X^i)$.

For computational reasons $\Sigma_{local}(X^i)$ is modeled as a diagonal matrix. We compute the mean of the Gaussian as $\mu_{local}(X^i) = 1/L \sum_l g(z^l),\ z^l \sim Q(z|X^i)$ For $\Sigma_{local}(X^i)$ we provide two options. The first is to use the sample variance at every pixel as $\Sigma_{local}(X^i)_{jj} = 1/(L-1) \sum_l (g(z^l)_j - \mu_{local}(X^i)_j)^2,\ z^l \sim Q(z|X^i),\ \Sigma_{local}(X^i)_{jk} = 0$ for $j \neq k$. which we refer to as $\Sigma_{local,est}(X^i)$. The second option is to learn a deterministic mapping with a neural network that takes the image X^i and directly predicts the diagonal covariance matrix as illustrated in Fig. 2, which we refer to as $\Sigma_{local,neural}(X^i)$. The network predicting $\Sigma_{local,neural}$ can be trained on the healthy images by predicting the expected ℓ_2 loss between a given input X and its reconstruction, i.e. $E_{z \sim Q(z|X)}[(g(z) - X)^2]$.

2.2 Unsupervised Detection with Probabilistic Image Restoration

One of the task that our approximation can applied to is unsupervised lesion detection. To perform unsupervised detection, we learn the prior distribution $P(X)$ for normal samples and perform image restoration using the local Gaussian approximation to maximize the likelihood of test samples. The lesions can then be revealed by the absolute difference between the test image and its restoration.

Auto-Encoder (VAE) is a typical model to estimate the prior distribution from the healthy images without lesions. For detailed explanation of VAEs, we refer readers to [10]. Note that the proposed local Gaussian approximation is not limited to combination with VAE but can also be used together with other latent space models that model $P(X)$, *e.g.* Adversarial Auto-encoder (AAE) [15]. Here we use VAE as an example to demonstrate the unsupervised detection work-flow.

VAEs assume that the image distribution can be modelled with a lower-dimensional latent variable model as $P(X) = \int P(X \mid z)P(z)dz$, where $P(z)$ stands for the pre-specified prior distribution in the latent space and $P(X|z)$ is modeled with a *decoding* network as $P(X|z) = \mathcal{N}(X; \mu(z), \sigma(z)\mathbf{I})$. VAEs build an *encoding* network to approximate the true posterior $P(z|X) \approx Q(z|X) = \mathcal{N}(z; \mu_z, \sigma_z\mathbf{I})$. Overall, VAEs encode an image X in the latent space and then reconstruct it with information encoded in the latent space. With networks for $Q(z|X)$ and $P(X|z)$ defined, the evidence lower bound (ELBO) can be derived as $E_{z \sim Q(z|X)}[\log P(X \mid z)] - KL[Q(z \mid X)||P(z)]$ and used as a lower bound for $\log P(X)$ and is maximized to train VAE as a surrogate of $\log P(X)$.

An abnormal image Y is seen as a healthy image X with "errors" δ, $Y = X + \delta$, where δ corresponds to the lesion. To detect δ, X is restored from Y by maximizing $\log P(X|Y) \propto \log[P(Y|X)P(X)]$ with respect to X using gradient ascent with steps defined as,

$$L(X) = \log P(Y|X) + \log P(X), \;\; X^{i+1} = X^i + \eta(\nabla L)_{X^i} \text{ and } X^0 = Y, \tag{2}$$

where η is the step size, superscripts denote iterations and the subscript denotes where gradient is evaluated. $\log P(Y|X)$ integrates modelling assumptions on the abnormality. The restoration given in Eq.2 iteratively computes the gradient at each X^i,

$$(\nabla L)_{X^i} = (\nabla \log P(X))_{X^i} + (\nabla \log P(Y|X))_{X^i}. \tag{3}$$

Define a generic cost $\lambda R(Y, X)$ to calculate $\log P(Y|X)$, we obtain the final gradient ascent step with the proposed approximation,

$$X^{i+1} = X^i + \eta \left[-\Sigma_{local}(X^i)^{-1}(X^i - \mu_{local}(X^i)) + \lambda \left(\nabla R(Y, X)\right)_{X^i} \right], \tag{4}$$

with $X^0 = Y$ and for $i = 0, \ldots, \texttt{max\_iter}$. Here, λ controls the relative strength of R and is selected on the training data. We select the suitable λ by restoring

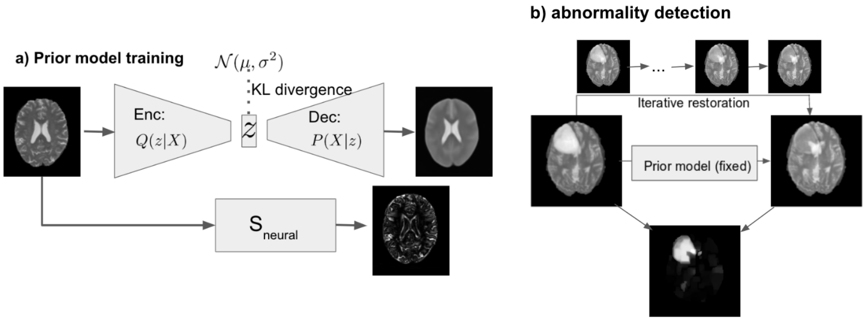

Fig. 2: Unsupervised lesion detection work-flow. a) Learning prior normative distribution, the network takes healthy images as input and outputs reconstructed images and predicted variance, b) likelihood maximization via local Gaussian approximation for test images using the prior distribution learned in a).

healthy images and quantifying the similarity between the restored and original images with Structural Similarity (SSIM). Specifically, we choose the smallest λ to obtain SSIM of at least 0.95. The final detection, i.e. an abnormality map, is computed at the end of the restoration as $\delta^* = |Y - X^{\texttt{max\_iter}}|$.

2.3 One-class Segmentation with Allowed False Positive Rate

To obtain the lesion segmentation, a threshold on δ^* is required. Using the same approach as in [16], we select the minimum τ that produces a false positive rate (FPR) of at most $x\%$ on the training data by assuming all detection is false positives as training images are lesion-free. The segmented lesion are defined as the set of pixels $\{r|\delta^*(r) > \tau\}$. Here, $x\%$ is a user-defined parameter.

3 Experiments

Datasets and pre-processing. We train and evaluate the proposed method on publicly available datasets: 1) Healthy individuals from the Cam-CAN study [17] which contains T2-weighted brain MRI of 652 subjects of age 18–87. All subjects have been confirmed to be normal by radiological assessment; 2) Abnormal data from BraTS 2018 [18] which contains T2-weighted images of 285 patients with high-grade (210) and low-grade (75) gliomas. All images are pre-processed with N4 bias correction, histogram matching between Cam-CAN and BraTS and intensity normalization to zero mean and unit variance within a brain mask, with background intensities set to -3.5.

Training details. The encoder network of the fully convolutional VAE has one `conv` layer followed by six residual blocks with 8, 16, 32, 64, 128 and 256 channels and 3x3 kernels, with a latent variable z size of $2 \times 2 \times 256$. The decoder network is symmetrical to the encoder. The variance prediction network has four `conv` layers with 8, 16, 8, 1 channels and 3x3 kernels without reducing the image size in hidden layers. We use `LeakyRelu` activation for hidden layers and `identity` activation for output layers. The network is trained on 128x128 images with a

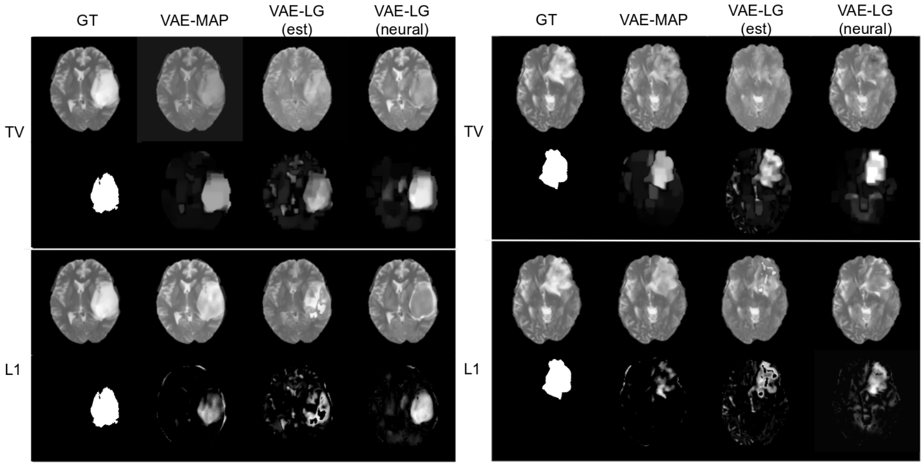

Fig. 3: Restoration of low-grade (left) and high-grade (right) gliomas. Abnormal images are restored using VAE-MAP, VAE-LG(est), VAE-LG(neural) (column 2-4) with TV- and L1-norms. Restoration results (top) and the corresponding absolute error maps δ (bottom) are shown along with the abnormal images (top) and reference segmentation (bottom) in the column GT.

batch size of 64 using the Adam optimizer and a learning rate of 2×10^{-5} for 193k iterations on one GPU Titan X for approximately 18 hrs. Abnormal images are restored with λ selected as in Sec. 2.3 using `Adam` and a learning rate of 5×10^{-3} for 800 iterations.

3.1 Results

We evaluate VAE and AAE with the proposed approximation, which we refer to as VAE-LG and AAE-LG, specifically VAE-LG(est) with estimated variance $\Sigma_{local,est}$ and VAE-LG(neural) with predicted variance $\Sigma_{local,neural}$, AAE-LG(neural) with predicted variance $\Sigma_{local,neural}$. We compare to the state-of-the-art method, VAE-MAP [14] as well as auto-encoding methods without image restoration as in [19].

When restored with the same constraint, local Gaussian approximation successfully restored the lesion area into a healthy-looking region compared to MAP as shown in Fig.3. For each method, restoration with TV-norm gave better visual results than with L1-norm. The advantage of TV may be attributed to the intensity characteristics of the gliomas. The best restoration was achieved by VAE-LG(neural). In contrast, the other methods either only restore the lesion partially (VAE-LG (est)) or naively lower the intensity values without restoring normal structures (VAE-MAP). This confirms the effectiveness of the proposed variance prediction network which preserves high-frequency details for meaningful restoration.

We quantify detection performance by Area-Under-Curve of ROC (AUC) and Dice computed at thresholds corresponding to different FPR limits in Table 1. The results are consistent with the visual inspection. VAE-LG(neural)

Table 1: Performance comparison of Dice for different thresholds and AUC. Best results are in bold. Run-time is reported per iteration. *na*: evaluation not available.

Methods	Constraints	.1%fpr	.5%fpr	1%fpr	5%fpr	AUC	Runtime (s)
GMM [7]	/	na	*na*	*na*	*na*	0.800	/
AnoGAN [5]	/	0.000±0.000	0.006±0.006	0.020±0.020	0.100±0.060	0.670	/
VAE [19]	/	0.000±0.000	0.030±0.030	0.090±0.060	0.200±0.140	0.690	/
AAE [19]	/	0.000±0.000	0.011±0.011	0.030±0.030	0.180±0.140	0.700	/
VAE-MAP [14]	TV	0.039±0.076	0.286±0.222	0.341±0.221	0.365±0.187	0.805	0.177
GMVAE-MAP [14]	TV	0.069±0.084	0.195±0.109	0.218±0.208	**0.455±0.225**	0.827	0.170
VAE-LG(est)	L1	0.063±0.025	0.213±0.183	0.269±0.208	0.347±0.216	0.772	0.127
VAE-LG(neural)	L1	0.133±0.143	0.309±0.288	0.360±0.276	0.315±0.224	0.824	**0.096**
VAE-LG(est)	TV	0.117±0.101	0.236±0.155	0.296±0.195	0.362±0.203	0.782	0.125
VAE-LG(neural)	TV	**0.259±0.246**	**0.407±0.252**	**0.448±0.209**	0.303 ±0.123	**0.828**	0.098
AAE-LG (neural)	TV	0.220±0.207	0.395±0.244	0.418±0.210	0.302±0.156	0.821	0.097

achieves the highest AUC of 0.828 with TV-norm, followed by a slightly lower AUC of 0.824 for VAE-LG(neural) with L1-norm. VAE-LG(neural) with TV norm reaches the highest Dice at all thresholds except for 5%fpr outperforming the other methods by a large margin. Notably, our detection methods using the predicted variance achieve high Dice at low FPR limits, such as .1%fpr, where the other methods are ineffective. VAE-LG(neural) with TV norm achieved +0.220, +0.142 and +0.190 improvement over VAE-MAP with TV norm, VAE-LG(est) with TV norm and GMVAE-MAP with TV norm at .1%fpr. The high AUC and low Dice for GMVAE-MAP with TV norm at low FPR limits indicate that it is difficult to identify the threshold of this method using the training set. Comparison between GMVAE-MAP and VAE-MAP indicates that a more complex prior distribution brings improvement in the detection while neither are effective at low FPR limits, such as .1%fpr and .5%fpr. We also combine the proposed local Gaussian approximation with AAE as AAE-LG(neural). AAE-LG(neural) gives similar results to VAE-LG(neural).

4 Conclusion

We have presented an unsupervised detection method with prior distribution learning and local Gaussian approximation with estimated pixel-wise variance. By restoring abnormal images, abnormalities are detected as the absolute difference resulted from the restoration. With restoration visualization and quantitative evaluation, we compared with previous works and observed significant improvement of detection accuracy with reduced false positives.

References

1. Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, Cagatay Demiralp, Jamie Shotton, Owen M Thomas, Tilak Das, Raj Jena, and Stephen J Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multichannel MR. *MICCAI*, 2012.
2. Chris A Cocosco, Alex P Zijdenbos, and Alan C Evans. A fully automatic and robust brain MRI tissue classification method. *MedIA*, 7(4):513–527, 2003.

3. Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *MedIA*, 36:61–78, 2017.
4. Simon Andermatt, Antal Horváth, Simon Pezold, and Philippe Cattin. Pathology segmentation using distributional differences to images of healthy origin. *arXiv:1805.10344*, 2018.
5. Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *IPMI*, 2017.
6. Marcel Prastawa, Elizabeth Bullitt, Sean Ho, and Guido Gerig. A brain tumor segmentation framework based on outlier detection. *MedIA*, 8(3):275–283, 2004.
7. Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester, and Paul Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE TMI*, 20(8):677–688, 2001.
8. Ke Zeng, Guray Erus, Aristeidis Sotiras, Russell T Shinohara, and Christos Davatzikos. Abnormality detection via iterative deformable registration and basis-pursuit decomposition. *IEEE TMI*, 35(8):1937–1951, 2016.
9. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NIPS*, 2014.
10. Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2014.
11. Zara Alaverdyan, Julien Jung, Romain Bouet, and Carole Lartizien. Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: application to epilepsy lesion screening. *MIDL*, 2018.
12. Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. *International MICCAI Brainlesion Workshop*, 2018.
13. David Zimmerer, Simon AA Kohl, Jens Petersen, Fabian Isensee, and Klaus H Maier-Hein. Context-encoding variational autoencoder for unsupervised anomaly detection. *arXiv:1812.05941*, 2018.
14. Suhang You, Kerem Tezcan, Xiaoran Chen, and Ender Konukoglu. Unsupervised Lesion Detection via Image Restoration with a Normative Prior. *MIDL*, 2019.
15. Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
16. Ender Konukoglu, Ben Glocker, Alzheimer's Disease Neuroimaging Initiative, et al. Reconstructing subject-specific effect maps. *NeuroImage*, 181:521–538, 2018.
17. Jason R Taylor, Nitin Williams, Rhodri Cusack, Tibor Auer, Meredith A Shafto, Marie Dixon, Lorraine K Tyler, Richard N Henson, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144:262–269, 2017.
18. Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.
19. Xiaoran Chen, Nick Pawlowski, Martin Rajchl, Ben Glocker, and Ender Konukoglu. Deep Generative Models in the Real-World: An Open Challenge from Medical Imaging. *arXiv:1806.05452*, 2018.