Supplementary Materials - Rethinking Autoencoders for Medical Anomaly Detection from A Theoretical Perspective

Yu Cai, Hao Chen[(✉)], and Kwang-Ting Cheng

The Hong Kong University of Science and Technology, Hong Kong, China
jhc@cse.ust.hk

1 Datasets

We use four datasets for evaluation, where three are from [2]:

RSNA Dataset[1]. The dataset contains 8851 normal and 6012 lung opacity CXRs. In experiments, we use 3851 normal images as the normal training dataset $\mathcal{D}_{train}$, and 1000 normal and 1000 abnormal (lung opacity) images as the test dataset $\mathcal{D}_{test}$.

VinDr-CXR Dataset[2] [3]. The dataset contains 10606 normal and 4394 abnormal CXRs that include 14 categories of anomalies in total. In experiments, we use 4000 normal images as $\mathcal{D}_{train}$, and 1000 normal and 1000 abnormal images as $\mathcal{D}_{test}$.

Brain Tumor Dataset[3]. The dataset contains 2000 MRI slices with no tumors, 1621 with glioma, and 1645 with meningioma. The glioma and meningioma are regarded as anomalies. In experiments, we use 1000 normal images (with no tumors) as $\mathcal{D}_{train}$, and 600 normal and 600 abnormal images (300 with glioma and 300 with meningioma) as $\mathcal{D}_{test}$.

Additionally, **BraTS2021** [1] is reorgnized for pixel-level anomaly segmentation in this paper. It provides 1251 MRI cases with the resolution $155 \times 240 \times 240$, as well as the corresponding voxel-level annotation for tumor regions. Each case has multiple modalities, including T1, T1ce, T2, and flair. We use only the flair in experiments as it is the most sensitive to tumor regions. In preprocessing, scans are cropped into $70 \times 208 \times 208$ to remove the empty corners. To construct $\mathcal{D}_{train}$, 4211 normal 2D axial slices are extracted from 1051 MRI scans, while the extracted slices are at least 5 slices apart from each other, ensuring enough different content between neighbors. For $\mathcal{D}_{test}$, similarly, 828 normal and 1948 tumor slices are extracted from the remaining 200 MRI scans.

Tab. 1 summarizes these datasets.

---

[1] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

[2] https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection

[3] https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Table 1. Summary of the datasets.

Dataset	Modality	Repartition	
		$\mathcal{D}_{train}$	$\mathcal{D}_{test}$ (Normal+Abnormal)
RSNA	Chest X-ray	3851	1000+1000
VinDr-CXR	Chest X-ray	4000	1000+1000
Brain Tumor	Brain MRI	1000	600+600
BraTS2021	Brain MRI	4211	828+1948

2   Implementation details

The AE in our experiments comprises an encoder and a decoder. The encoder consists of four convolutional layers with kernel size 4 and stride 2, whose channel sizes are 16-32-64-64. The decoder consists of four deconvolutional layers with the same kernel size and stride as the encoder, and the channel sizes are 64-32-16-1. The encoder and decoder are connected by four fully connected layers, where the dimensions are 1024-D-d-D-1024. D is always set to 1024 in our experiments, and d is the latent dimension of AE. All layers except the output layer are followed by batch normalization (BN) and ReLU. For fair comparison, other reconstruction methods are implemented based on this backbone consistently. All the input images are resized to 64×64, and all the reconstruction models are trained for 250 epochs using the Adam optimizer with a learning rate of 1e-3.

The performance on image-level anomaly detection is assessed with the area under the ROC curve (AUC) and average precision (AP). The performance on pixel-level anomaly segmentation is assessed with the pixel-level AP (AP_{pix}) and the best possible Dice score ($\lceil Dice \rceil$).

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
2. Cai, Y., Chen, H., Yang, X., Zhou, Y., Cheng, K.T.: Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. Medical Image Analysis **86**, 102794 (2023)
3. Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T., Dinh, D.H., et al.: Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. Scientific Data **9**(1), 1–7 (2022)